Advanced Linear Algebra MA500-1: Lecture Notes
Semester 1 2015-2016

Dr Rachel Quinlan
School of Mathematics, Statistics and Applied Mathematics, NUI Galway

March 14, 2017

Contents

1 Three ways to think about a matrix **2**
1.1 Linear transformations . . . 2
1.1.1 Interpreting a matrix as a linear transformation . . . 2
1.1.2 Interpreting a linear transformation as a matrix . . . 3
1.1.3 Change of Basis . . . 4
1.1.4 Similarity . . . 6
1.1.5 Rank . . . 8
1.2 Bilinear Forms . . . 11
1.2.1 Symmetric and alternating forms . . . 13
1.2.2 Duality . . . 18
1.3 Matrices and Graphs . . . 19

2 Spectral Properties **24**
2.1 The determinant . . . 24
2.2 The spectrum of a matrix . . . 28
2.3 Positive matrices - the Frobenius-Perron Theorem . . . 31
2.4 Supplement to Chapter 2: even and odd permutations . . . 38

Chapter 1

Three ways to think about a matrix

Matrices are ubiquitous in mathematics and arise centrally in many areas that are not necessarily closely related in an obvious way. Matrices are naturally equipped with lots of algebraic structure (for example the set of $n \times n$ matrices over a ring R or field $\mathbb{F}$ is itself a ring with lots of interesting properties). Which aspects of this extensive algebraic structure are of interest can depend a lot on the context. There are many ways of thinking about what a matrix is, and it is often helpful, even necessary, to have access to more than one of them. We discuss three different viewpoints (and the algebraic considerations that accompany them) in this chapter.

1.1 Linear transformations

For a field $\mathbb{F}$ and positive integer n, we will write $\mathbb{F}^n$ for the vector space consisting of all column vectors of length n with entries in $\mathbb{F}$, and $M_n(\mathbb{F})$ for the set of all $n \times n$ matrices with entries in $\mathbb{F}$.

1.1.1 Interpreting a matrix as a linear transformation

If $A \in M_n(\mathbb{F})$ and $v \in \mathbb{F}^n$, we can "multiply" A by v to get another element of $\mathbb{F}^n$.

Example 1.1.1. *In* $M_3(\mathbb{Q})$, *write* $A = \begin{pmatrix} -1 & 1 & 2 \\ -12 & 8 & 6 \\ 12 & -7 & -3 \end{pmatrix}$. *In* $\mathbb{Q}^3$, *write* $v = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$. *Then*

$$Av = \begin{pmatrix} -1 & 1 & 2 \\ -12 & 8 & 6 \\ 12 & -7 & -3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -1(1) + 1(2) + 2(3) \\ -12(1) + 8(2) + 6(3) \\ 12(1) - 7(2) - 3(3) \end{pmatrix} = \begin{pmatrix} 7 \\ 22 \\ -11 \end{pmatrix}.$$

Example 1.1.1 demonstrates the process of *matrix-vector* multiplication. Although this is already familiar it is worthwhile to consider what is going on in slightly more detail. For a matrix A and column vector v, you can calculate the product Av only if the number of columns of A is the same as the number of entries in v. What we are doing when we calculate Av is taking the linear combination of the columns of A that is determined by the entries of v. This means that if $v = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ and A has n columns, then the column vector Av is given by

$$a_1 \begin{bmatrix} \text{Col 1} \\ \text{of} \\ A \end{bmatrix} + a_2 \begin{bmatrix} \text{Col 2} \\ \text{of} \\ A \end{bmatrix} + \cdots + a_n \begin{bmatrix} \text{Col } n \\ \text{of} \\ A \end{bmatrix}.$$

Exercise 1.1.2. *If in doubt, confirm this from Example 1.1.1.*

Thus every matrix $A \in M_{n\times n}(\mathbb{F})$ defines a function $T_A : \mathbb{F}^n \to \mathbb{F}^n$ by

$$T_A(v) = Av.$$

This function is a *linear transformation*, which means that

- For all $u, v \in \mathbb{F}^n$, $T_A(u+v) = T_A(u) + T_A(v)$ (since $A(u+v) = Au + Av$), and
- For all $v \in \mathbb{F}^n$ and $\alpha \in \mathbb{F}$, $T_A(\alpha u) = \alpha T_A(u)$. (The field element α is referred to as a *scalar* in this context).

In general a linear transformation is a function that respects addition and scalar multiplication (in a context where that makes sense).

NOTES

1. More generally, a $p \times n$ matrix A (p rows, n columns) may be thought of as a linear transformation T_A from $\mathbb{F}^n$ to $\mathbb{F}^p$, via matrix-vector multiplication. If $v \in \mathbb{F}^n$, then $T_A(v) = Av$ is the linear combination of the columns of A in which the coefficient of Column i is the ith entry of v.
2. If $A \in M_{p\times n}(\mathbb{F})$ then Column i of A is the image under T_A of the vector e_i, which has entry 1 in the ith position and zero in all other positions. So different matrices correspond to different linear transformations.
3. If $T : \mathbb{F}^n \to \mathbb{F}^p$ is *any* linear transformation, let A be the matrix in $M_{p\times n}(\mathbb{F})$ whose ith column is $T(e_i)$. Then $T(v) = Av$ for all vectors $v \in \mathbb{F}^n$.

Exercise 1.1.3. *Prove the statement in Item 3 above.*

So we can think of the matrix space $M_{p\times n}(\mathbb{F})$ as being the set of linear transformations from $\mathbb{F}^n$ to $\mathbb{F}^p$. The presentation given here involves a choice to discuss matrices multiplied (on the right) by column vectors. It could equally well be presented as in terms of matrices being multiplied on the left by row vectors.

We let $(\mathbb{F}^p)^T$ denote the *transpose* of $\mathbb{F}^p$, i.e. the space of *row vectors* of length p with entries in $\mathbb{F}$. Then a $p \times n$ matrix A describes a linear transformation from $(\mathbb{F}^p)^T$ to $(\mathbb{F}^n)^T$ via

$$v \to vA.$$

Note that vA, which belongs to $\mathbb{F}^n$, is the linear combination of the rows of A in which the coefficient of Row i is entry i of v. So a (row-)vector-matrix product gives a linear combination of the rows of the matrix, with coefficients given by the vector entries.

1.1.2 Interpreting a linear transformation as a matrix

Suppose now that V and W are vector spaces of finite dimensions n and p respectively over $\mathbb{F}$, and let $f : V \to W$ be a linear transformation. Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ and $\mathcal{C} = \{c_1, \ldots, c_k\}$ be bases for V and W respectively. Write $M_{f,\mathcal{B},\mathcal{C}}$ for the $p \times n$ matrix whose ith column contains the $\mathcal{C}$-coordinates of the element $f(b_i)$ of W.

Theorem 1.1.4. *If $v \in V$, then the $\mathcal{C}$-coordinates of $f(v)$ are the entries of the matrix-vector product $M_{f,\mathcal{B},\mathcal{C}}[v]$, where $[v]$ is the vector in $\mathbb{F}^n$ whose entries are the $\mathcal{B}$-coordinates of v.*

Proof. Write $v = a_1b_1 + \cdots + a_nb_n$, so $[v] = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$. Then

$$f(v) = a_1f(b_1) + a_2f(b_2) + \cdots + a_nf(b_n).$$

Since the $\mathcal{C}$-coordinates of $f(b_i)$ are written into Column i of $M_{f,\mathcal{B},\mathcal{C}}$, it follows that the $\mathcal{C}$-coordinates of $f(v)$ are the entries of

$$a_1 \begin{bmatrix} \text{Col } 1 \\ \text{of} \\ M_{f,\mathcal{B},\mathcal{C}} \end{bmatrix} + a_2 \begin{bmatrix} \text{Col } 2 \\ \text{of} \\ M_{f,\mathcal{B},\mathcal{C}} \end{bmatrix} + \cdots + a_n \begin{bmatrix} \text{Col } n \\ \text{of} \\ M_{f,\mathcal{B},\mathcal{C}} \end{bmatrix} = M_{f,\mathcal{B},\mathcal{C}}[v].$$

□

Example 1.1.5. *Let $V = \mathbb{Q}_4[x]$, the space of polynomials in x of degree at most 4 over $\mathbb{Q}$, with basis $\mathcal{B} = \{1, x, x^2, x^3, x^4\}$. Let $W = \mathbb{Q}_3[x]$, the space of polynomials of degree at most 3 over $\mathbb{Q}$, with basis $\mathcal{C} = 1, x, x^2, x^3$. Let $D : V \to W$ be the differential operator, which maps a polynomial to its derivative. Then D is a linear transformation and*

$$M_{D,\mathcal{B},\mathcal{C}} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix}.$$

If we want to use this to calculate the derivative of $x^4 - 3x^3 + 2x^2 - x$, we can calculate the matrix-vector product

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \\ 2 \\ -3 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 4 \\ -9 \\ 4 \end{pmatrix}.$$

So $D(x^4 - 3x^3 + 2x^2 - x) = 4x^3 - 9x^2 + 4x - 1$.

Exercise 1.1.6. *In the above example, suppose we used the basis $\mathcal{C}' = \{1, 1+x, 1+x+x^2, 1+x+x^2+x^3\}$ for W instead of $\mathcal{C}$. How would the matrix change?*

The point here is that given a linear transformation f between two vector spaces of finite dimension, the choice of a basis for each space allows us consider f as a matrix. It is not exactly true to say that every transformation f corresponds to a matrix in some objective way, because it is not only f but also the choice of two bases that determine the matrix.

Question 1.1.7. *Suppose that $f : V \to W$ is a linear transformation between different vector spaces. How do the matrices that represent f with respect to different bases resemble each other?*

In order to answer this question we need some matrix machinery.

1.1.3 Change of Basis

Let V be a $\mathbb{F}$-vector space of dimension n and suppose that $\mathcal{B} = \{b_1, \ldots, b_n\}$ and $\mathcal{B}' = \{v_1, \ldots, v_n\}$ are bases of V. Then each b_i can be written in a unique way as a linear combination of $v_1, \ldots, v_n$. For $j = 1, \ldots, n$ write

$$b_j = \sum_{i=1}^{n} a_{ij} v_i, \quad a_{ij} \in \mathbb{F}.$$

Let P denote the $n \times n$ matrix whose entry in the (i, j) position is a_{ij}. Possibly a better way to think about the matrix P is that Column j of P is the vector whose entries are the $\mathcal{B}'$-coordinates of b_j. So the columns of P express the elements of the basis $\mathcal{B}$ in terms of their $\mathcal{B}'$-coordinates.

Lemma 1.1.8. *Let $x \in V$, and suppose that $x = \sum_{i=1}^{n} c_i b_i$, so that the $\mathcal{B}$-coordinates of x are $c_1, \ldots, c_n$. Then the $\mathcal{B}'$-coordinates of x are given by the entries of the matrix-vector product* $P \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$.

Proof. The matrix-vector product is

$$c_1 \begin{bmatrix} \text{Col } 1 \\ \text{of} \\ P \end{bmatrix} + c_2 \begin{bmatrix} \text{Col } 2 \\ \text{of} \\ P \end{bmatrix} + \cdots + c_n \begin{bmatrix} \text{Col } n \\ \text{of} \\ P \end{bmatrix}.$$

Since Column j of P expresses b_j in terms of its $\mathcal{B}'$-coordinates, the entries of this product are the coordinates of $c_1b_1 + \cdots + c_nb_n = x$ in terms of $\mathcal{B}'$. □

Definition 1.1.9. *In view of Lemma 1.1.8, we refer to* P *as the* change of basis matrix *from* $\mathcal{B}$ *to* $\mathcal{B}'$*. Its columns are the elements of* $\mathcal{B}$ *expressed in terms of* $\mathcal{B}'$*.*

Remark In the context of Section 1.1.2, you can consider P to be the matrix $M_{id,\mathcal{B},\mathcal{B}'}$, where $id : V \to V$ is the identity mapping. (Thanks to Ben for this nice observation).

Now let Q be the change of basis matrix from $\mathcal{B}'$ to $\mathcal{B}$, defined equivalently - the columns of Q are the $\mathcal{B}$-column representations of the elements $v_1, \ldots, v_n$ of $\mathcal{B}'$. As above, we can pass from the $\mathcal{B}'$-column representation of any element of V to its $\mathcal{B}$-column representation by multiplying on the left by Q.

Now let c be any vector in $\mathbb{F}^n$. Then c is the $\mathcal{B}$-column representation of some element v of V, and the $\mathcal{B}'$ column representation of v is Pc. But then the $\mathcal{B}$-column representation of v is given by $Q(Pc) = QPc$. However this must be equal to c Thus

$$QPc = c \text{ for all } c \in \mathbb{F}^n.$$

Hence $QP = I_n$, the $n \times n$ identity matrix. Similarly $PQ = I_n$, so P and Q are inverses of each other. In particular, every change of basis matrix is invertible, and its inverse is the reverse change of basis matrix.

On the other hand, *every* invertible $n \times n$ matrix determines a change of basis in $\mathbb{F}^n$. To see this let $P \in M_n(\mathbb{F})$ be invertible and let Q be its inverse. This means that $PQ = QP = I_n$.

We focus on the product $QP = I_n$. This means that e_1, the first column of I_n, is the linear combination of the columns of Q whose coefficients are the entries of Column 1 of P. Similarly $e_2, \ldots, e_n$ are linear combinations of the columns of Q with coefficients given by the entries of Columns $2, \ldots, n$ of P. Then in particular all of the standard basis vectors of $\mathbb{F}^n$ belong to the span of the Columns of Q, and so the columns of Q form a spanning set of $\mathbb{F}^n$. Since $\mathbb{F}^n$ has dimension n it cannot be spanned by fewer than n columns, and so the columns of Q form a *minimal* spanning set of $\mathbb{F}^n$, hence they must be linearly independent. Thus the columns of the invertible matrix Q form a basis $\mathcal{B}_Q$ of $\mathbb{F}^n$.

Moreover, for $j = 1, \ldots, n$, the entries of Column 1 of P are the coordinates of e_i with respect to the basis $\mathcal{B}_Q$. If v is any vector in $\mathbb{F}^n$, then the coordinates of v with respect to $\mathcal{B}_Q$ are given by the entries of Pv (or equivalently $Q^{-1}v$). So P or Q^{-1} is the change of basis matrix from the standard basis of $\mathbb{F}^n$ to $\mathcal{B}_Q$.

Remarks

1. As above, we could use the fact that $PQ = I_n$ to show that the columns of P form a basis for $\mathbb{F}^n$, and by thinking of the rows of PQ (or QP) as linear combinations of the rows of Q (or P), we can show that the rows of Q (or P) form a basis for $(\mathbb{F}^n)^T$.

2. The above argument shows that if $A \in M_n(\mathbb{F})$ has a right inverse (i.e. there exists $B \in M_n(\mathbb{F})$ with $AB = I_n$, then the columns of A form a basis of $\mathbb{F}^n$. It is true that if A has a right inverse then it also has a left inverse and these coincide, but we are not quite in a position to prove that yet. We will soon.

We are now in a position to answer Question 1.1.7.

Suppose that V and W are $\mathbb{F}$-vector spaces of dimensions p and n respectively and that $f : V \to W$ is a linear transformation. Let $\mathcal{B}$ and $\mathcal{B}'$ be two bases of V and let $\mathcal{C}$ and $\mathcal{C}'$ be two bases of W.

Let the change of basis matrices from $\mathcal{B}'$ to $\mathcal{B}$ and from $\mathcal{C}$ to $\mathcal{C}'$ be denoted by P and Q respectively (so P is a nonsingular $n \times n$ matrix and Q is a nonsingular $p \times p$ matrix). Then

$$M_{f,\mathcal{B}',\mathcal{C}'} = Q\, M_{f,\mathcal{B},\mathcal{C}}\, P.$$

Explanation: Suppose that c is the $\mathcal{B}'$-column representation of some element v of V. We want to know what matrix should multiply c on the left in order to give the $\mathcal{C}'$-column representation of $f(v)$. Multiplying c by P gives us the $\mathcal{B}$-column representation of v, multiplying that by the $p \times n$ matrix $M_{f,\mathcal{B},\mathcal{C}}$ gives us the $\mathcal{C}$-column representation of $f(v)$, and multiplying that by Q gives us the $\mathcal{C}'$-column representation of $f(v)$. Thus, overall we have

$$[f(v)]_{\mathcal{C}'} = Q\, M_{f,\mathcal{B},\mathcal{C}}\, P\, [v]_{\mathcal{B}'}.$$

This motivates the following definition.

Definition 1.1.10. *Let* A *and* B *be matrices in* $M_{p\times n}(\mathbb{F})$. *Then* A *and* B *are said to be* equivalent *if there exist nonsingular matrices* $P \in M_n(\mathbb{F})$ *and* $Q \in M_p(\mathbb{F})$ *for which*

$$B = QAP.$$

If two $p \times n$ matrices are equivalent, if means that they represent the same linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^p$, possibly with respect to different bases for both spaces.

1.1.4 Similarity

We now specialize the discussion to the case where $V = W$. Our set up now is that we have a single vector space V of dimension n, and a linear transformation $f : V \to V$. If $\mathcal{B}$ is a basis of V, we write $M_{f,\mathcal{B}}$ for the matrix that was called $M_{f,\mathcal{B},\mathcal{B}}$ in Section **??**. Suppose that $\mathcal{B}'$ is another basis of V, and let P be the change of basis matrix from $\mathcal{B}'$ to $\mathcal{B}$ (so the columns of P are the $\mathcal{B}$-representations of the elements of $\mathcal{B}'$. Then P^{-1} is the change of basis matrix from $\mathcal{B}$ to $\mathcal{B}'$ (and its columns are the $\mathcal{B}'$-representations of the elements of $\mathcal{B}$). Then

$$M_{f,\mathcal{B}'} = P^{-1} M_{f,\mathcal{B}} P.$$

Definition 1.1.11. *Suppose that* A *and* B *are matrices in* $M_n(\mathbb{F})$. *Then* A *and* B *are said to be* similar *if there exists an invertible matrix* $P \in M_n(\mathbb{F})$ *for which*

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

If two matrices are *similar*, it means that they describe the same linear transformation, with respect to different bases.

Given a linear transformation $f : V \to V$, it is reasonable to ask whether there is some basis of V with respect to which its matrix has a particulary nice form. The nicest form that you can hope for is a *diagonal* form, in which all entries away from the main diagonal (from upper left to lower right) are zeros. If there is a basis $\mathcal{B} = \{b_1, \ldots, b_n\}$ of V for which

$$M_{f,\mathcal{B}} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix},$$

it means that $f(b_i) = \lambda_i(b_i)$ for $i = 1, \ldots, n$. This means exactly that each b_i is an eigenvector of f.

Definition 1.1.12. *Let* V *be a* $\mathbb{F}$*-vector space and let* $f : V \to V$ *be a linear transformation. A non-zero element* v *of* V *is called an* eigenvector *of* f *if* $f(v) = \lambda v$ *for some* $\lambda \in \mathbb{F}$. *In this case* λ *is called the* eigenvalue *of* f *to which* v *corresponds.*

So if v is an eigenvector of f, it means that $f(v)$ is just a scalar multiple of v. The mapping $f : V \to V$ is called diagonalizable (or diagonable) if there exists a basis of V with respect to which the matrix of f is diagonal. This means that there exists a basis of V consisting entirely of eigenvalues of f.

The "matrix" versions of (some of) these definitions are given below. There are many concepts and statements in linear algebra that can be expressed either in terms of matrices or in terms of linear transformations.

Definition 1.1.13. *Let $A \in M_n(\mathbb{F})$. A non-zero column vector $v \in \mathbb{F}^n$ is a* right eigenvector *of A if $Av = \lambda v$ for some $\lambda \in \mathbb{F}$. A non-zero row vector $w \in \mathbb{F}^n$ is a* left eigenvector *of A if $wA = \lambda w$ for sme λ in $\mathbb{F}$. In each case the scalar λ is the corresponding* eigenvalue *of A.*

If is not always true for a linear transformation $f : V \to V$ or for a square matrix $A \in M_n(\mathbb{F})$ that there exists a basis of V (or $\mathbb{F}^n$) consisting of eigenvectors. For example let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and investigate right eigenvectors of A. Suppose that

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix} \Longrightarrow \begin{array}{rcl} x + y & = & \lambda x \\ y & = & \lambda y. \end{array}$$

For an eigenvector x and y cannot both be zero. From the second equation, either $y = 0$ or $\lambda = 1$. However if $y = 0$ then $\lambda = 1$ anyway from the first equation, so we must have $\lambda = 1$, and 1 is the only eigenvalue of A. Now

$$x + y = x \Longrightarrow y = 0,$$

and any vector of the form $\binom{x}{0}$ is an eigenvector of A corresponding to the eigenvalue 1. These are the only eigenvectors of A and they are all scalar multiples of $\binom{1}{0}$, they form a 1-dimensional subspace of $\mathbb{F}^2$. So $\mathbb{F}^2$ does not have a basis consisting of eigenvectors of A, and A is not similar to a diagonal matrix.

We do have the following theorem, but its converse is not true (since we can obviously easily write down examples of diagonal matrices that have repeated eigenvalues).

Theorem 1.1.14. *Suppose that $A \in M_n(\mathbb{F})$ has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, and that $v_1, \ldots, v_n$ are respective eigenvectors of A. Then $\mathcal{B} = \{v_1, \ldots, v_n\}$ is a basis of $\mathbb{F}^n$.*

Proof. We show that the set $\mathcal{B}$ is linearly independent. Suppose it's not, and seek a contradiction. Then k be the least index for which $\{v_1, v_2, \ldots, v_k\}$ is linearly dependent. Then $k \leqslant n$ since $\mathcal{B}$ is linearly dependent, and $k \geqslant 2$ since v_1 is not the zero vector. Then $v_1, \ldots, v_{k-1}$ are linearly independent and v_k is a linear combination of these, so there exist $a_1, \ldots, a_{k-1} \in \mathbb{F}$, not all zero, with

$$v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1}. \tag{1.1}$$

Multiplying 1.1 on the left by A gives

$$\lambda_k v_k = a_1 \lambda_1 v_1 + a_2 \lambda_2 v_2 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1},$$

and multiplying 1.1 by the scalar λ_k gives

$$\lambda_k v_k = a_1 \lambda_k v_1 + a_2 \lambda_k v_2 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1}.$$

Equating the right hand sides of these two expressions for $\lambda_k v_k$ gives

$$a_1(\lambda_k - \lambda_1) + a_2(\lambda_k - \lambda_2) + \cdots + a_{k-1}(\lambda_k - \lambda_{k-1}) v_{k-1}. \tag{1.2}$$

Since the eigenvalues $\lambda_1, \ldots, \lambda_k$ are distinct, the field elements $\lambda_k - \lambda_i$ in 1.2 above are all non-zero. Furthermore the a_i in this expression are not all zero, so 1.2 is a linear dependence relation among $v_1, \ldots, v_{k-1}$. This contradicts the choice of k as the least index for which $\{v_1, \ldots, v_k\}$ has such a relation. □

A matrix $A \in M_n(\mathbb{F})$ is *diagonalizable* (over $\mathbb{F}$) if there exists a basis $\mathcal{B}$ of $\mathbb{F}^n$ consisting of eigenvectors of A. In this case $P^{-1}AP$ is diagonal, where P is matrix whose columns are the elements of $\mathcal{B}$ (this is the change of basis matrix from $\mathcal{B}$ to the standard basis). So a matrix is *diagonalizable* if it is similar to a diagonal matrix.

There is an issue here that might not be immediately obvious. We demonstrate it with an example.

Example 1.1.15. *Let* $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ *in* $M_2(\mathbb{R})$. *To investigate eigenvectors of* A *we can write*

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \lambda\begin{pmatrix} x \\ y \end{pmatrix} \Longrightarrow \begin{array}{rcl} -y & = & \lambda x \\ x & = & \lambda y \end{array}$$

Now substituting the second equation into the first gives $-y = \lambda x = \lambda(\lambda y) = \lambda^2 y$. Taking $y = 0$ is not an option as this would force $x = 0$ also, and the zero vector cannot be an eigenvector. So from $-y = \lambda^2 y$ we must conclude $\lambda^2 = -1$. There is no real number λ with this property, so A has no eigenvalues in $\mathbb{R}$ and has no eigenvectors with real entries. However, if we allow ourselves to consider complex eigenvalues, we can try $\lambda_1 = i$ and $\lambda_2 = -i$. An eigenvector $\binom{x}{y}$ corresponding to λ_1 must satisfy $x = iy$, for example $\binom{i}{1}$, and an eigenvector corresponding to λ_2 must satisfy $x = -iy$, for example $\binom{-i}{1}$. Now $\binom{i}{1}$ and $\binom{-i}{1}$ form a basis of $\mathbb{C}^2$, so A is diagonalizable if we consider it as an element of $M_2(\mathbb{C})$, and in this case

$$P^{-1}AP = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \text{ where } P = \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix}. \quad \text{(Check this!)}$$

However, A is *not* diagonalizable within $M_2(\mathbb{R})$.

Given a matrix $A \in M_n(\mathbb{F})$ (or a linear transformation $f : V \to V$ of a n-dimensional $\mathbb{F}$-vector space V) we can ask about the existence of "nice" bases for describing A, or of "nice" matrices that are similar to A. The eigenvalues of A may or may not belong to the field $\mathbb{F}$. We can consider two cases:

- If we only want to consider similarity within $M_n(\mathbb{F})$, we can try to identify the "nicest" matrix of the form $P^{-1}AP$, where $P \in GL(n, \mathbb{F})$. This leads to the theory of the *rational canonical form*.
- If $\bar{\mathbb{F}}$ is a field that has $\mathbb{F}$ has a subfield and contains all the eigenvalues of $\mathbb{F}$, then we can consider A to be an element of $M_n(\bar{\mathbb{F}})$ and consider similarity over $\bar{\mathbb{F}}$. Then we would be looking for the "nicest" matrix of the form $P^{-1}AP$ where the entries of the invertible matrix P (and of $P^{-1}AP$) belong to $\bar{\mathbb{F}}$ but not necessarily to $\mathbb{F}$. This leads to the theory of the *Jordan canonical form*. The Jordan canonical form of the matrix in Example 1.1.15 above is the diagonal matrix $\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$.

We will revisit the concept of similarity later in the course.

1.1.5 Rank

Let $f : V \to W$ be a linear transformation of $\mathbb{F}$-vector spaces. The *kernel* and *image* of f are defined by

$$\ker f = \{x \in V : f(x) = 0\}; \quad \Im f = f(x) : x \in V.$$

Lemma 1.1.16. *The kernel and image of* f *are subspaces of* V *and* W *respectively.*

The proof is left as an exercise.

Definition 1.1.17. *The dimension of* $\Im f$ *is called the* rank *of* f.

Let $A \in M_{p\times n}(\mathbb{F})$. The *column space* of A is the subspace of $\mathbb{F}^p$ that is spanned by the columns of A. The dimension of this space is called the *column rank* of A. Since linear combinations of the columns of A are precisely equal to matrix-vector products of the form Av with $v \in \mathbb{F}^n$, the column space of A is the set of all such vectors. This is also the image of the linear transformation $T_A : \mathbb{F}^n \to \mathbb{F}^p$ defined as left-multiplication by A, and its dimension is the rank of T_A. Thus

The column rank of A is the rank of the linear transformation T_A.

Analagously, we can define the *row rank* of A to be the dimension of the subspace of $(\mathbb{F}_n)^T$ spanned by the rows of A. This space is called the row space of A and it is the image of the linear transformation from $L_A : (\mathbb{F}^p)^T \to (F_n)^T$ defined for a row vector $w \in (\mathbb{F}_p)^T$ by

$$L_A(w) = wA.$$

What precisely is the connection between T_A and L_A is not a particularly easy question to answer but we will come back to it in Section 1.2. For now we can prove the surprising and non-obvious fact that they have the same rank. Many different proofs of this theorem can be found in books, not all of them offer a lot of insight into why the statement is true - many rely on row and column operations that reduce the matrix to an echelon form.

Theorem 1.1.18. *Let $A \in M_{p\times n}(\mathbb{F})$. Then the row rank and column rank of A are equal.*

Proof. Write r for the row rank of A and c for the column rank. We want to show that $r = c$. Let $v_1, \ldots, v_c$ be column vectors in $\mathbb{F}^n$ that form a basis for the column space of A, and let C_A be the $p \times c$ matrix that has $v_1, \ldots, v_c$ as its columns. Then every column of A can be expressed as a linear combination of $v_1, \ldots, v_c$ in a unique way, and it follows that there exists a $c \times n$ matrix R for which

$$\underbrace{C_A}_{p\times c}\underbrace{R}_{c\times n} = A.$$

The first column in R contains the coordinates of Column 1 of A in terms of $v_1, \ldots, v_c$, and so on. Looking at the same product the other way round, we see that Row 1 of A is the linear combination of the rows of R whose coefficients are the entries of Row 1 of C_A and so on; each row of A is a linear combination of the c rows of R. Thus the c rows of R span the rowspace of A, and the row rank of A is at most c, so $r \leqslant c$.

We use essentially the same argument to show that $c \leqslant r$ and hence that $r = c$. The row rank of A is r, hence there exists a spanning set $w_1, \ldots, w_r$ for the row space of A. Let R_A be the $r \times n$ matrix whose rows are $w_1, \ldots, w_r$. Then every row of A has a unique expression as a linear combination of the rows of R_A, so there exists a $p \times r$ matrix C for which

$$\underbrace{C}_{p\times r}\underbrace{R_A}_{r\times n} = A.$$

The entries in Row i of C are the coefficients in the expression for Row i of A as a linear combination of the rows of R_A. But now every column of A is a linear combination of the r columns of C, and hence the dimension of the column space of A is at most r, so $c \leqslant r$.

Since $c \leqslant r$ and $r \leqslant c$, we conclude that $c = r$ and that the row rank and column rank of A are equal. □

In view of Theorem 1.1.18, we do not need to distinguish between the row rank and column rank of a matrix, and we can just refer to its *rank*. If A is a $p \times n$ matrix, then $\text{rank}(A)$ is an integer between 0 and $\min(p, n)$. Here are a few remarks about rank.

1. A matrix has rank 0 if and only if it is the zero matrix.
2. A matrix has rank 1 if and only if it is not the zero matrix and all of its non-zero rows are scalar multiples of each other (same for columns).

3. Suppose that $A \in M_n(\mathbb{F})$ (so A is square). Then A has rank n if and only if its columns form a basis of $\mathbb{F}^n$. By Theorem 1.1.18, this happens if and only if its rows form a basis of $(\mathbb{F}^n)^T$. From the discussion on page 5 we can now say that A has rank n if and only if it has both a right and a left inverse. Finally if A has both a right inverse B and a left inverse C these must coincide, since if AB and CA are equal to the identity matrix, then

$$C = CI_n = C(AB) = (CA)B = I_nB = B.$$

So A has rank n if and only if A is invertible.

4. It is interesting to consider how rank behaves under matrix addition and matrix multiplication. There is not that much that can be said about how the ranks of the sum and product of a pair of matrices depend on their individual rank. We have the following statements.

   - Suppose that A and B are matrices of the same size $p \times n$. Then

   $$|\text{rank}(A) - \text{rank}(B)| \leqslant \text{rank}(A + B) \leqslant \text{rank}(A) + \text{rank}(B).$$

   - Suppose that A is a $p \times q$ matrix and B is a $q \times n$ matrix. Then the rank of the product AB is at least equal to $\text{rank}(A) + \text{rank}(B) - q$ and at most equal to $\min\{\text{rank}(A), \text{rank}(B)\}$.

   These statements will feature in Assignment 2.

We finish this section now with a discussion of the *Rank-Nullity Theorem*.

Theorem 1.1.19. *Suppose that $f : V \to W$ is a linear transformation of finite-dimensional $\mathbb{F}$-vector spaces. Then*

$$\dim(\ker f) + \dim(\text{Im} f) = \dim V.$$

Proof. Write n for the dimension of V and k for the dimension of the kernel of f. Let $\{b_1, \ldots, b_k\}$ be a basis for $\ker f$. This can be extended to a basis $\mathcal{B} = \{b_1, \ldots, b_k, v_{k+1}, \ldots, v_n\}$ of V. We show now that $\mathcal{B}' = \{f(v_{k+1}), \ldots, f(v_n)\}$ is a basis of $\text{Im} f$.

- Let $w \in \text{Im} f$. Then $w = f(v)$ for some $v \in V$, and $v = a_1b_1 + \cdots + a_kb_k + c_{k+1}v_{k+1} + \cdots + c_nv_n$, for $a_1, \ldots, a_k$ and $c_{k+1}, \ldots, c_n$ in $\mathbb{F}$. Applying f to this description of v, we obtain

$$\begin{aligned} w = f(v)v &= a_1f(b_1) + \cdots + a_kf(b_k) + c_{k+1}f(v_{k+1}) + \cdots + c_nf(v_n) \\ &= c_{k+1}f(v_{k+1}) + \cdots + c_nf(v_n), \end{aligned}$$

  since $f(b_i) = 0$ for $i = 1, \ldots, k$. Thus w belongs to the linear span of the elements of $\mathcal{B}'$ and $\mathcal{B}'$ is a spanning set for $\text{Im} f$.

- To see that $\mathcal{B}'$ is linearly independent, note that a linear dependence relation amongst the elements of $\mathcal{B}'$ would imply that some linear combination of $v_{k+1}, \ldots, v_n$, with coefficients not all zero, belongs to the kernel of f. This is impossible since $\mathcal{B}$ is a basis of V and $\{b_1, \ldots, b_k\}$ is a basis of $\ker f$ - this means that $\ker f$ intersects the span of $\{v_{k+1}, \ldots, v_n\}$ only in the zero element. However the only way to express zero as a linear combination of $\{v_{k+1}, \ldots, v_n\}$ is by taking all the coefficients to be zero.

Thus $n - k$ is the dimension of $\text{Im} f$, which means exactly that $\dim(\ker f) + \dim(\text{Im} f) = n = \dim V$. □

A way of thinking informally about Theorem 1.1.19 is that if $f : V \to W$ is a linear transformation, then f carries the space V into the space W. The full dimension of V must be accounted for in this transformation - the elements of the kernel get "lost" under f and this accounts for some of the dimension, the rest must survive in the image.

Like most theorems about linear transformations, Theorem 1.1.19 has a "matrix version", stated below. If A is $p \times n$ matrix, the *right nullspace* of A is the subspace of $\mathbb{F}^n$ consisting of all those vectors v for which $Av = 0$. It is the kernel of the linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^p$ defined as left multiplication by A. The *left nullspace* of A consists of those vectors $w \in (\mathbb{F}^p)^T$ for which $wA = 0$. It is the kernel of the linear transformation from $(\mathbb{F}^p)^T$ to $(\mathbb{F}^n)^T$ defined as right multiplication by A.

Theorem 1.1.20. *(Rank-nullity theorem, matrix version) Let A be a $p \times n$ matrix. Then*

1. *The sum of the dimension of the right nullspace of A and the rank of A is n (the number of columns).*
2. *The sum of the dimension of the left nullspace of A and the rank of A is p (the number of rows).*

1.2 Bilinear Forms

Think about the ordinary scalar product on $\mathbb{F}^n$, defined by

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n \in \mathbb{F}.$$

This product can be considered to be a function from $V \times V$ to $\mathbb{F}$. It is an example of a *bilinear form*.

Definition 1.2.1. *A* bilinear form *on a $\mathbb{F}$-vector space V is a function $\tau : V \times V \to \mathbb{F}$ that satisfies the following conditions.*

1. $\tau(u + v, w) = \tau(u, w) + \tau(v, w)\ \forall\ u, v, w \in V$.
2. $\tau(\lambda u, v) = \lambda\tau(u, v)\ \forall\ u, v \in V,\ \lambda \in \mathbb{F}$.
3. $\tau(u, v + w) = \tau(u, v) + \tau(u, w)\ \forall\ u, v, w \in V$.
4. $\tau(u, \lambda v) = \lambda\tau(u, v)\ \forall\ u, v \in V,\ \lambda \in \mathbb{F}$.

Suppose that V has dimension n and that $\mathcal{B} = \{b_1, \ldots, b_n\}$ is a basis of V. If τ is a bilinear form on V, let $A_{\mathcal{B}}$ denote the $n \times n$ matrix whose (i, j) entry is $\tau(b_i, b_j)$. Then $A_{\mathcal{B}}$ is called the *Gram matrix* of τ with respect to the basis $\mathcal{B}$.

Lemma 1.2.2. *Let $u, v \in V$ and let $[u]_{\mathcal{B}}$ and $[v]_{\mathcal{B}}$ denote their respective $\mathcal{B}$-column representations. Then*

$$\tau(u, v) = [u]_{\mathcal{B}}^T A_{\mathcal{B}} [v]_{\mathcal{B}}.$$

Proof. Write $u = \sum_{i=1}^n p_i b_i$ and $v = \sum_{i=1}^n q_i b_i$ where the p_i and q_p belong to $\mathbb{F}$. Then

$$\begin{aligned}
\tau(u, v) &= \tau(p_1b_1 + \cdots + p_nb_n, q_1b_1 + \cdots + p_nb_n) \\
&= \sum_{i=1}^n p_i\tau(b_i, q_1b_1 + \cdots + p_nb_n) \\
&= \sum_{i=1}^n p_i \sum_{j=1}^n q_j\tau(b_i, b_j) \\
&= \sum_{i=1}^n p_i \sum_{j=1}^n a_{ij}q_j \\
&= \begin{pmatrix} p_1 & \ldots & p_n \end{pmatrix} A_{n\times n} \begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix} \\
&= [u]_{\mathcal{B}}^T A_{\mathcal{B}} [v]_{\mathcal{B}}.
\end{aligned}$$

□

On the other hand, if $A \in M_n(\mathbb{F})$, we can use A to define a bilinear form τ on $\mathbb{F}^n$ by

$$\tau(u, v) = u^T A v, \text{ for } u, v \in \mathbb{F}^n.$$

The set of all bilinear forms on a $\mathbb{F}$-vector space V is itself a $\mathbb{F}$-vector space, with addition given by $(\tau_1 + \tau_2)(u, v) = \tau_1(u, v) + \tau_2(u, v)$ and scalar multiplication given by $(\lambda\tau)(u, v) = \lambda\tau(u, v)$. If V is finite-dimensional, then we can choose a basis for V and as above associate every bilinear form to its Gram matrix, and every matrix to a different bilinear form. This correspondence respects addition and scalar multiplication, so we can say that the choice of a basis for V determines an explicit isomorphism between the vector space $B(V)$ of all $\mathbb{F}$-bilinear forms on V and the space $M_n(\mathbb{F})$ of all $n \times n$ matrices over $\mathbb{F}$. In particular the dimension of $B(V)$ is n^2.

Now that we have a matrix description of bilinear forms, we can play the same game as in Section 1.1 and ask how the matrices that describe the form with respect to different bases are related. The answer is a bit different.

Theorem 1.2.3. *Let f be a bilinear form on the $\mathbb{F}$-vector space V. Let $\mathcal{B}$ and $\mathcal{B}'$ be bases of V, and let P be the change of basis matrix from $\mathcal{B}'$ to $\mathcal{B}$. Let $A_{\mathcal{B}}$ and $A_{\mathcal{B}'}$ be the Gram matrices of τ with respect to the two bases. Then*

$$A_{\mathcal{B}'} = P^T A_{\mathcal{B}} P.$$

Proof. Let $u, v \in V$. Then, by definition of $A_{\mathcal{B}}$,

$$\tau(u, v) = ([u]_{\mathcal{B}})^T A_{\mathcal{B}} [v]_{\mathcal{B}}.$$

However, $[v]_{\mathcal{B}} = P[v]_{\mathcal{B}'}$, and $[u]_{\mathcal{B}} = P[u]_{\mathcal{B}'}$. So

$$\begin{aligned} \tau(u, v) &= (P[u]_{\mathcal{B}'})^T A_{\mathcal{B}} P[v]_{\mathcal{B}'} \\ &= ([u]_{\mathcal{B}'})^T P^T A_{\mathcal{B}} P[v]_{\mathcal{B}'} \end{aligned}$$

Thus for all column vectors u and v in $\mathbb{F}^n$ we have

$$u^T P^T A_{\mathcal{B}} P v = u^T A_{\mathcal{B}'} v.$$

For each i and j, choosing $u = e_i$ and $v = e_j$ confirms that $P^T A_{\mathcal{B}} P$ and $A_{\mathcal{B}'}$ have the same entry in the (i, j) position, and hence that these matrices are equal. □

Note on the transpose of a matrix product: The transpose of a $p \times n$ matrix A is the matrix $n \times p$ matrix A^T (read "A transpose") defined by $(A^T)_{ij} = A_{ji}$. The rows of A^T are the columns of A. Suppose now that A is $p \times n$ and B is $n \times q$, so AB is $p \times n$. We can ask how $(AB)^T$ depends on the transposes of A and B.

$$\begin{aligned} (AB)^T_{ij} &= (AB)_{ji} \\ &= \sum_{k=1}^{q} A_{jk} B_{ki} \\ &= \sum_{k=1}^{q} A^T_{kj} B^T_{ik} \\ &= \sum_{k=1}^{q} B^T_{ik} A^T_{kj} \\ &= (B^T A^T)_{ij}. \end{aligned}$$

So $(AB)^T = B^T A^T$: the transpose of the product of two matrices is the product of their transposes, in the opposite order. This fact is used in the proof of Theorem 1.2.3 above.

Theorem 1.2.3 motivates the following definition.

Definition 1.2.4. *Two square matrices* A *and* B *in* $M_n(\mathbb{F})$ *are* congruent *to each other if there exists an invertible matrix* $P \in GL(n, \mathbb{F})$ *with*

$$B = P^T A P.$$

Congruence is an equivalence relation on $M_n(\mathbb{F})$, and two matrices are congruent if and only if they describe the same bilinear form on $\mathbb{F}^n$, with respect to different bases.

Now we consider the meaning of the *rank* of the Gram matrix.

Definition 1.2.5. *Let* τ *be a bilinear form on a* $\mathbb{F}$*-vector space* V *of dimension* n*.*

- *The* left radical L *of* τ *is the subset of* V *consisting of all those elements* v *with the property that* $\tau(v, x) = 0$ *for all* $x \in V$*.*
- *The* right radical R *of* τ *is the subset of* V *consisting of all those elements* w *with the property that* $\tau(x, w) = 0$ *for all* $x \in V$*.*

It is straightforward to verify from the definition of bilinear form that both L and R are *subspaces* of V.

Example 1.2.6. *Let* τ *be the bilinear form on* $\mathbb{Q}^2$ *with Gram matrix* $\begin{pmatrix} 1 & -2 \\ -3 & 6 \end{pmatrix}$ *(with respect to the standard basis).*

The left radical of τ is given by

$$\begin{aligned} L &= \left\{ \begin{pmatrix} a \\ b \end{pmatrix} : \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} 1 & -2 \\ -3 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \, \forall \, x, y \in \mathbb{Q} \right\} \\ &= \left\{ \begin{pmatrix} a \\ b \end{pmatrix} : \begin{pmatrix} a - 3b & -2a + 6b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \, \forall \, x, y \in \mathbb{Q} \right\}. \end{aligned}$$

Now $\begin{pmatrix} a - 3b & -2a + 6b \end{pmatrix}$ is a row vector - the only way that it can have zero product with *every* column vector in $\mathbb{Q}^2$ is if it is the zero vector. Thus

$$L = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} : a - 3b = -2a + 6b = 0 \right\} = \left\{ \lambda \begin{pmatrix} 3 \\ 1 \end{pmatrix} : \lambda \in \mathbb{Q} \right\}.$$

In particular L is the left nullspace of the Gram matrix, and similarly R is the right nullspace. Since the Gram matrix is square, its left and right nullspaces have the same dimension by Theorem 1.1.20, and so the left and right radicals of τ have the same dimension. This example demonstrates the general principle - for any bilinear form τ, the left and right radicals of τ correspond (with respect to particular basis) to the left and right nullspaces of the Gram matrix of τ with respect to that basis. A bilinear form is called *nondegenerate* if its left and right radicals consist only of the zero element. The following is an alternative version of that definition.

Definition 1.2.7. *Let* τ *be a bilinear form on a vector space* V*. Then* τ *is* nondegenerate *if for every non-zero element* x *of* V*, there exist* $y, z \in V$ *such that* $\tau(x, y) = 0$ *and* $\tau(z, x) = 0$*.*

The *rank* of a bilinear form is defined to be the rank of its Gram matrix (with respect to any basis, they all have the same rank).

1.2.1 Symmetric and alternating forms

Throughout this section we assume that the characteristic of our field $\mathbb{F}$ is different from 2. This essentially means that the element $2 (= 1 + 1)$ is not equal to the zero element of our field and so 2 has an inverse for multiplication.

Definition 1.2.8. *A bilinear form* τ *on* V *is* symmetric *if and only if* $\tau(u, v) = \tau(v, u)$ *for all* $u, v \in V$*.*
A bilinear form τ *on* V *is* skew-symmetric *or* alternating *if and only if* $\tau(u, v) = -\tau(v, u)$ *for all* $u, v \in V$*.*

It is a straightforward consequence of this definition that a bilinear form is symmetric if and only if its Gram matrices with respect to all bases are symmetric, and that a bilinear form is skew-symmetric if and only if its Gram matrices are all skew-symmetric. A square matrix A is *symmetric* if $A^T = A$ and *skew-symmetric* if $A^T = -A$. Thus we can observe that the properties of symmetry and skew symmetry must be preserved under congruence. This is something that is easily observed at the matrix level anyway, for suppose that A is a symmetric matrix, C is skew-symmetric, and P is non-singular. Then

$$(P^TAP)^T = P^TA^T(P^T)^T = P^TAP \quad \text{and} \quad (P^TCP)^T = P^TC^T(P^T)^T = P^T(-C)P = -(P^TCP).$$

So any matrix that is congruent to A is symmetric and any matrix that is congruent to C is skew-symmetric.

Example 1.2.9. *(A symmetric form) For $p \times q$ matrices A and B, define $\tau(A, B)$ by*

$$\tau(A, B) = \text{trace}A^TB.$$

Then τ is a non-degenerate symmetric bilinear form on $M_{p\times n}(\mathbb{F})$.

The proof of the assertion in Example 1.2.9 is an exercise. The symmetry of τ means that A^TB and B^TA always have the same trace, even though they may not have the same size. The nondegeneracy of τ says that if A is a non-zero $p \times q$ matrix, then there exists a $q \times p$ matrix B for which AB has non-zero trace.

Example 1.2.10. *(An alternating form) Define a form d on $\mathbb{F}^2$ by*

$$d\left(\begin{pmatrix}a\\b\end{pmatrix}, \begin{pmatrix}c\\d\end{pmatrix}\right) = \det\begin{pmatrix}a & c\\ b & d\end{pmatrix}.$$

Then d is an alternating bilinear form.

Elements of $\mathbb{F}^2$ are column vectors of length 2 with entries in $\mathbb{F}$. The form d takes a pair of such vectors and returns the determinant of the 2×2 matrix with these columns.

IF τ is an alternating or symmetric bilinear form, it is easily checked that the left and right radicals of τ coincide, so we can just talk about the *radical*, denoted $\text{rad}\tau$. The radical corresponds to the left and right nullspaces of a Gram matrix that represents the form; the left and right nullspaces coincide for symmetric and skew-symmetric matrices. Two elements u and v of V are said to be *orthogonal* with respect to τ (written $u \perp v$ or $u \perp_\tau v$) if $\tau(u, v) = 0$.

Definition 1.2.11. *Let τ be a symmetric or alternating bilinear form on V, and let S be a subset of V. The* orthogonal complement *of S with respect to τ, denoted by $S^\perp$ or $S^{\perp_\tau}$ is the subset of V consisting of all elements that are τ-orthogonal to every element of S.*

$$S^\perp = \{x \in V : \tau(x, u) = 0 \; \forall \, u \in U\}.$$

NOTES

1. Any element of $S^\perp$ is orthogonal to all linear combinations of elements of S as well as to elements of S themselves. So $S^\perp = \langle S\rangle^\perp$, where $\langle S\rangle$ is the subspace of V spanned by S. It is probably more usual to define the orthogonal complement for subspaces instead of for subsets.

2. If $u \in V$, we write $u^\perp$ for the orthogonal complement of the set $\{u\}$ (or the 1-dimensional subspace $\langle u\rangle$).

3. If S is a subset or subspace of U, then $S^\perp$ is a subspace of U (not just a subset). This follows in a straightforward way from the bilinearity of τ.

4. We could define orthogonal complements for forms that are not symmetric or alternating, but we would have to distinguish between left and right complements.

5. $S^\perp \supseteq \mathrm{rad}\tau$ for all $S \subseteq V$.

6. If U_1 and U_2 are subspaces of V with $U_1 \subseteq U_2$, then $U_2^\perp \supseteq U_1^\perp$ (complementation is *inclusion reversing*).

Example 1.2.12. Define a bilinear form τ of $M_3(\mathbb{F})$ by

$$\tau(A, B) = \mathrm{trace}(AB).$$

Let $S_3(\mathbb{F})$ be the subspace of $M_3(\mathbb{F})$ consisting of all symmetric matrices. Then $S_3(\mathbb{F})$ has dimension 6 and has the set

$$\{E_{11}, E_{22}, E_{33}, E_{12} + E_{21}, E_{13} + E_{31}, E_{23} + E_{32}\}$$

as a basis. Here E_{ij} denotes the matrix that has entry 1 in the (i, j) position and zeros in all other positions. We can use the definition of τ to confirm that

$$\begin{aligned}
E_{11}^\perp &= \{A \in M_3(\mathbb{F}) : A_{11} = 0\} \\
E_{22}^\perp &= \{A \in M_3(\mathbb{F}) : A_{22} = 0\} \\
E_{33}^\perp &= \{A \in M_3(\mathbb{F}) : A_{33} = 0\} \\
(E_{12} + E_{21})^\perp &= \{A \in M_3(\mathbb{F}) : A_{12} = -A_{21}\} \\
(E_{13} + E_{31})^\perp &= \{A \in M_3(\mathbb{F}) : A_{13} = -A_{31}\} \\
(E_{23} + E_{32})^\perp &= \{A \in M_3(\mathbb{F}) : A_{23} = -A_{32}\}
\end{aligned}$$

Thus a matrix A is orthogonal to each of our basis elements of $S_3(\mathbb{F})$ and hence to every symmetric matrix, if and only if A has zero entries on the diagonal and $A_{ij} = -A_{ji}$ for $i \neq j$. This can be summarized by saying that $A = -A^T$. Thus the orthogonal complement of $S_3(\mathbb{F})$ is the space $A_3(\mathbb{F})$ of *skew-symmetric* matrices in $M_3(\mathbb{F})$.

Note that $\dim S_3(\mathbb{F}) = 6$ and $\dim A_3(\mathbb{F}) = 3$, these dimensions are complementary in the sense that their sum is $\dim M_3(\mathbb{F})$. In this particular example $S_3(\mathbb{F})$ and its orthogonal complement have trivial intersection, but that does not always happen.

Lemma 1.2.13. *Let τ be a nondegenerate symmetric or alternating form on a vector space V of dimension n over $\mathbb{F}$. Let $v \in V$, $v \neq 0$. Then $v^\perp$ has dimension $n - 1$ in V.*

Proof. The function $f_v : V \to \mathbb{F}$ defined for $x \in V$ by

$$f_v(x) = \tau(v, x)$$

is a linear mapping. It does not map every element of V to zero, since τ is nondegenerate. Thus the image of f_v is at least one-dimensional, and since it is contained in the one-dimensional space $\mathbb{F}$, it is equal to $\mathbb{F}$. The kernel of f_v is exactly $v^\perp$, and by the rank-nullity theorem this has dimension $n - 1$. □

Our next theorem states that for a nondegenerate symmetric or alternating form on V, the dimensions of any subspace of V and its orthogonal complement add up to $\dim V$. The idea of the proof is similar to that of Lemma 1.2.13, which is the special case of a one-dimensional subspace. We need a preliminary lemma.

Lemma 1.2.14. *Suppose that τ is a symmetric or alternating bilinear form on a vector space V of dimension n, and that U is a subspace of V of dimension k. Then $U^\perp$ has dimension at least $n - k$ in V.*

Proof. Let $\{b_1, \ldots, b_k\}$ be a basis of U. For each i define f_{b_i} as in the proof of Lemma 1.2.13. Define a function $f : V \to \mathbb{F}^k$ by

$$f(x) = \begin{pmatrix} f_{b_1}(x) \\ f_{b_2}(x) \\ \vdots \\ f_{b_k}(x) \end{pmatrix} = \begin{pmatrix} \tau(b_1, x) \\ \tau(b_2, x) \\ \vdots \\ \tau(b_k, x) \end{pmatrix}.$$

Then f is a linear mapping from a n-dimensional space to a k-dimensional space. Its kernel is $U^\perp$. Since the image of f has dimension at most k, it is immediate from the rank-nullity theorem that $U^\perp = \ker f$ has dimension at least $n - k$. □

In particular it follows from Lemma 1.2.14 that if U is a proper subspace of V, then $U^\perp$ is not the zero subspace.

Theorem 1.2.15. *Let τ be a nondegenerate symmetric or alternating form on a vector space V of dimension n over $\mathbb{F}$. Let U be a subspace of V of dimension k. Then $U^\perp$ has dimension $n - k$ in V.*

Proof. Let $\{b_1, \ldots, b_k\}$ be a basis of U. For each i define f_{b_i} as in the proof of Lemma 1.2.13. Define a function $f : V \to \mathbb{F}^k$ by

$$f(x) = \begin{pmatrix} f_{b_1}(x) \\ f_{b_2}(x) \\ \vdots \\ f_{b_k}(x) \end{pmatrix} = \begin{pmatrix} \tau(b_1, x) \\ \tau(b_2, x) \\ \vdots \\ \tau(b_k, x) \end{pmatrix}.$$

Then f is a linear mapping and its kernel is $U^\perp$. The dimension of $U^\perp$ is $n - \dim(\mathrm{Im}V)$ by the rank-nullity theorem. Since $\mathrm{Im}f$ is a subspace of $\mathbb{F}^k$ its dimension is at most k. What we need to do to complete the proof is show that it is exactly k.

Now suppose that $\mathrm{Im}f$ is not equal to $\mathbb{F}^k$. Then it is a proper subspace of $\mathbb{F}^k$, and we can consider its orthogonal complement with respect to the ordinary scalar product on $\mathbb{F}^k$, which is a symmetric bilinear form. By Lemma 1.2.14 and the comment following it, there exists a nonzero element of $\mathbb{F}^k$ which is orthogonal to every element of $\mathrm{Im}f$ with respect to the ordinary scalar product. This means there exist $a_1, \ldots, a_k$ in $\mathbb{F}$, not all zero, for which

$$a_1\tau(b_1, x) + a_2\tau(b_2, x) + \cdots + a_k\tau(b_k, x) = 0 \text{ for all } x \in V.$$

Then $\tau(a_1b_1 + a_2b_2 + \cdots + a_kb_k, x) = 0$ for all $x \in V$, which means that $a_1b_1 + a_2b_2 + \cdots + a_kb_k \in \mathrm{rad}\tau$. This contradicts the hypothesis that τ is nondegenerate, since $a_1b_1 + a_2b_2 + \cdots + a_kb_k$ is a nonzero element of V.

We conclude that the image of f is equal to $\mathbb{F}^k$ and hence that $U^\perp$, the kernel of f, has dimension $n - k$. □

We show now that every symmetric bilinear form can be represented by a Gram matrix that is diagonal, or equivalently that every symmetric matrix (over any field) is congruent to a diagonal matrix. A basis $\{b_1, \ldots, b_n\}$ that satisfies $\tau(b_i, b_j) = 0$ for $i \neq j$ is said to be τ-orthogonal - this usage of the word "orthogonal" arises from the fact that different basis elements are orthogonal to each other. A τ-orthogonal basis of V is one with respect to which the Gram matrix of τ is diagonal.

Theorem 1.2.16. *Let τ be a symmetric bilinear form defined on a vector space V of dimension n over a field $\mathbb{F}$. Then there exists a basis of V with respect to which the Gram matrix of τ is diagonal.*

Proof. First we assume that τ is nondegenerate. Choose $x, y \in V$ for which $\tau(x, y) \neq 0$. Then

$$\tau(x + y, x + y) = \tau(x, x) + 2\tau(x, y) + \tau(y, y).$$

Since $\tau(x, y) \neq 0$, it follows that at least one of $\tau(x, x), \tau(y, y)$ and $\tau(x + y, x + y) \neq 0$. The point of this is that it shows that there exists an element b_1 of V for which $\tau(b_1, b_1) = d_1 \neq 0$.
Now write V_1 for the τ-orthogonal complement of $\langle b_1 \rangle$ in V. Then V_1 has dimension $n - 1$ by Lemma 1.2.13 and $b_1 \notin V_1$ (since b_1 is not self-orthogonal under τ). It follows that $V = \langle b_1 \rangle \oplus V_1$.
Now τ restricts to a nondegenerate bilinear form on V_1, for let $x \in V_1$ be a non-zero element. Then there exists $y \in V$ with $\tau(x, y) \neq 0$, and since $\tau(x, b_1) = 0$, such a y exists in V_1.
Now the same argument as above says that there exists an element b_2 of V_1 with $\tau(b_2, b_2) = d_2 \neq 0$. The set $\{b_1, b_2\}$ is linearly independent since $\tau(b_2, b_1) = 0$ but $\tau(b_2, b_2) \neq 0$. Repeating this

argument leads to a list $b_1, b_2, \ldots, b_n$ of linearly independent elements of V with the following properties.

$$\tau(b_i, b_i) = d_i \neq 0, \quad \tau(b_i, b_j) = 0 \text{ for } i \neq j.$$

Note that by the time we get to the introduction of b_n, the orthogonal complement of $\langle b_1, \ldots, b_{n-1} \rangle$ is one-dimensional, so there is no choice about b_n and we cannot use the argument above to show that b_n is not self-orthogonal. However, we know that b_n *is* orthogonal to all of $b_1, \ldots, b_{n-1}$ but is not in the radical of τ (because τ is nondegenerate). Thus it must be that $\tau(b_n, b_n) \neq 0$.

Now $\mathcal{B}$ is a linearly independent set of n elements and is therefore a basis of V, with respect to which the Gram matrix of τ is $\mathrm{diag}(d_1, d_2, \ldots, d_n)$.

If τ is degenerate, let W be a subspace of V for which $V = \mathrm{rad}\tau \oplus W$. Then τ restricts to a nondegenerate symmetric form on W, and so W has a τ-orthogonal basis. Extending this by elements of $\mathrm{rad}\tau$ leads to a basis of V that is τ-orthogonal. □

Note: The statement $V = \mathrm{rad}\tau \oplus W$ above says that V is the *direct sum* of $\mathrm{rad}\tau$ and W. This means that $\mathrm{rad}\tau$ and W are subspaces of V that span V together and have trivial intersection. Thus the union of a basis of $\mathrm{rad}\tau$ and a basis of W is a basis of V.

Whether much more can be said about diagonal matrices that represent a particular bilinear form often depends on properties of the field $\mathbb{F}$. The number of zeros on the main diagonal of a diagonal matrix representing a symmetric bilinear form τ is the dimension of the radical of τ - this is the same for different τ-orthogonal bases.

Suppose that $\mathbb{F}$ has the property that every element of $\mathbb{F}$ arise as a square in $\mathbb{F}$ - that is, for every $x \in \mathbb{F}$ there exists $a \in \mathbb{F}$ with $a^2 = x$. Such fields are called *quadratically closed*; examples include the field $\mathbb{C}$ of complex numbers. Over a quadratically closed field, if the elements b_i of a τ-orthogonal basis satisfy $\tau(b_i, b_i) = d_i$, then for those i with $d_i \neq 0$, we can let $\sqrt{d_i}$ be a square root of d_i, and replace b_i with $\frac{1}{\sqrt{d_i}} b_i$ to obtain a basis with respect to which the Gram matrix is diagonal with only 1s and zeros on the diagonal. In particular every nonsingular symmetric matrix over a quadratically closed field $\mathbb{F}$ is congruent to the identity matrix.

The field $\mathbb{R}$ of real numbers is not quadratically closed, but it has the property that every nonzero element is either a square or the negative of a square. Using the same idea as above, any orthogonal basis for a symmetric bilinear form on a $\mathbb{R}$-vector space can be adapted to one in which the entries on the diagonal of the Gram matrix are all equal to 1, -1 or 0.

In the case of alternating forms, we can again show the existence of special bases.

Theorem 1.2.17. *Let $\tau : V \times V \to \mathbb{F}$ be a nondegenerate alternating form on a vector space V over a field $\mathbb{F}$, where $\dim V = n$. Then $n = 2k$ for some integer k, and there exists a basis $\mathcal{B} = \{b_1, \ldots, b_k, c_1, \ldots, c_k\}$ of V for which*

- $\tau(b_i, c_i) = 1$ *for* $i = 1, \ldots, k$, *and*
- $\tau(b_i, b_j) = 0 = \tau(c_i, c_j)$ *for all* i, j.
- $\tau(b_i, c_j) = 0$ *for* $i \neq j$.

Proof Outline: This is Problem 7 in Problem Sheet 2 so a full proof is not included here. The basic idea is similar to that of the proof of Theorem 1.2.16. Start with a non-zero element b_1 of V. Then, by the nondegeneracy of τ, there exists c_1' with $\tau(b_1, c_1') \neq 0$, and c_1' can be adjusted by a scalar to obtain an element c_1 with $\tau(b_1, c_1) = 1$. Now move to the τ-orthogonal complement of $\langle b_1, c_1 \rangle$, show that it has trivial intersection with $\langle b_1, c_1 \rangle$, and repeat the process there.

A basis of the type described in Theorem 1.2.17 is referred to as a *symplectic* basis for the form τ. Note that it follows from Theorem 1.2.17 that a nondegenerate alternating form can be defined only on a vector space of even dimension, and also that every skew-symmetric matrix has even rank.

1.2.2 Duality

Definition 1.2.18. *Let V be a vector space of dimension n over a field $\mathbb{F}$. The* dual space *of V, denoted $\hat{V}$, is the space of all linear mappings from V to $\mathbb{F}$.*

Note that $\hat{V}$ is itself a vector space over $\mathbb{F}$, with addition and scalar multiplication defined in the obvious way. If $\mathcal{B} = \{b_1, \ldots, b_n\}$ is a basis of V, then for i from 1 to n we can define an element $\hat{b_i}$ of $\hat{V}$ by

$$\hat{b_i}(b_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Once $\hat{b_i}$ has been defined on the basis elements, it extends by linearity to all of V. Since every element of $\hat{V}$ is defined by the images of $b_1, \ldots, b_n$, every element of $\hat{V}$ can be expressed in a unique way as a $\mathbb{F}$-linear combination of $\hat{b_1}, \ldots, \hat{b_n}$, and hence $\{\hat{b_1}, \ldots, \hat{b_n}\}$ is a *basis* of $\hat{V}$, referred to as the *dual basis* of $\mathcal{B}$ and denoted $\hat{\mathcal{B}}$. In particular $\hat{V}$ has dimesion n also, and V and $\hat{V}$ are isomorphic as vector spaces.

A bilinear form τ on V defines a linear mapping $\theta_\tau : V \to \hat{V}$ as follows. For $x \in V$, $\theta_\tau(x)$ is the function from V to $\mathbb{F}$ defined by

$$\theta_\tau(x)(y) = \tau(x, y).$$

On the other hand, suppose that we start with a linear mapping $\theta : V \to \hat{V}$. We can associate to this a bilinear form τ_θ on V defined for $x, y \in V$ by

$$\tau_\theta(x, y) = \theta(x)(y).$$

Recall that $\theta(x)$ is a function from V to $\mathbb{F}$, so this makes sense.
With this interpretation, we can think of a bilinear form as a linear mapping from V to $\hat{V}$. This mapping is a bijection if and only if the bilinear form is nondegenerate.

So far we have two ways of thinking about a matrix - as a linear transformation and as a bilinear form. If we think of a square matrix A as the Gram matrix of some bilinear form τ, then it is fairly easy to interpret the meaning of the transpose of A - it represents the bilinear form τ' defined by

$$\tau'(x, y) = \tau(y, x).$$

The forms τ and τ' coincide if and only if τ is a symmetric form, which means exactly that A is a symmetric matrix. If we are thinking of matrices as linear transformations though, it is a bit harder to interpret the meaning of the transpose.

Suppose that $T : V \to V$ is a linear transformation. Associated with T is a linear transformation $\hat{T}$ of $\hat{V}$, defined for $f \in \hat{V}$ by

$$\hat{T}(f) = f \circ T.$$

Question: What is the connection between the matrix of T with respect to $\mathcal{B}$ and the matrix of $\hat{T}$ with respect to $\hat{\mathcal{B}}$?

Let A be the matrix of T with respect to $\mathcal{B}$. This means that $T(b_j) = \sum_{i=1}^{n} a_{ij} b_i$. We need to figure out how to express $\hat{T}(\hat{b_j})$ as a linear combination of the $\hat{b_i}$. Let $x \in V$ and write $x = \sum_i c_i b_i$. Then

$$\begin{aligned}
\hat{T}(\hat{b_j})(x) &= \hat{b_j}(T(x)) \\
&= \hat{b_j}(T(\sum_i c_i b_i)) \\
&= \hat{b_j}(\sum_i c_i T(b_i)) \\
&= \hat{b_j}(\sum_i c_i \sum_k a_{ki} b_k) \\
&= \sum_i a_{ji} c_i \\
&= \sum_i a_{ji} \hat{b_i}(x).
\end{aligned}$$

Thus

$$\hat{T}(\hat{b_j}) = \sum_i a_{ji}\hat{b_i},$$

and the matrix of $\hat{T}$ with respect to $\hat{\mathcal{B}}$ is A^T, the transpose of the matrix of T with respect to $\mathcal{B}$.

Remark: *Natural isomorphism between* V *and its double dual* $\hat{\hat{V}}$.
We have seen above that every finite dimensional vector space V has the same dimension as its dual $\hat{V}$, and hence they are isomorphic as vector spaces. Once we choose a basis for V we also define a dual basis for $\hat{V}$ and the correspondence between the two bases gives us an explicit isomorphism between V and $\hat{V}$. However this isomorphism is not intrinsic to the space V, in the sense that it depends upon a choice of basis and cannot be described independently of a choice of basis.

The dual space of $\hat{V}$ is denoted $\hat{\hat{V}}$; it is the space of all linear mappings from $\hat{V}$ to $\mathbb{F}$. By all of the above discussion it has the same dimension as $\hat{V}$ and V. The reason for mentioning this however is that there *is* a "natural" isomorphism θ from V to $\hat{\hat{V}}$. It is defined in the following way, for $x \in V$ and $f \in \hat{V}$ - note that $\theta(x)$ belongs to $\hat{\hat{V}}$, so $\theta(x)(f)$ should be an element of $\mathbb{F}$.

$$\theta(x)(f) = f(x).$$

To see that θ is an isomorphism, suppose that $x_1, \ldots, x_k$ are independent elements of V. Then $\theta(x_1), \ldots, \theta(x_k)$ are independent elements of $\hat{\hat{V}}$. To see this let $a_1, \ldots, a_k$ be element of $\mathbb{F}$ for which $a_1\theta(x_1) + \cdots + a_k\theta(x_k) = 0$. This means that $f(a_1x_1 + \cdots + a_kx_k) = 0$ for all $f \in \hat{V}$, which means that $a_1x_1 + \cdots + a_kx_k = 0$, which means that each $a_i = 0$.

1.3 Matrices and Graphs

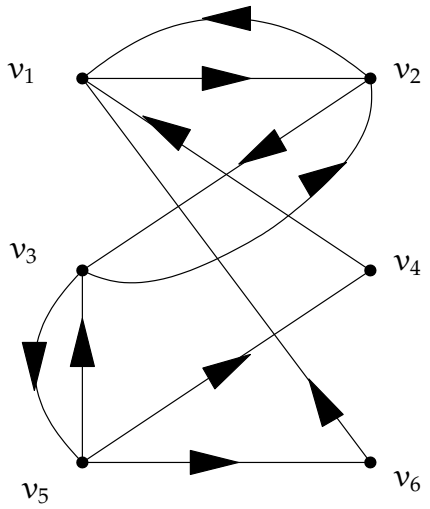
A *directed graph* (or *digraph*) G consists of a non-empty set V of vertices and a set E of ordered pairs of these vertices, called edges. Each edge is directed from one vertex of G to another. An *undirected graph* is similar, except that the edges are not considered to have a direction.

A number of square matrices are typcially associated to a graph, the most elementary of which is the *adjacency matrix*.

Definition 1.3.1. *Let* G *be a (directed) graph with* n *vertices labelled* $v_1, \ldots, v_n$. *The* adjacency matrix A *of* G *is the* $n \times n$ *matrix whose entries are given by*

$$A_{ij} = \begin{cases} 1 & \text{there is an edge directed from } v_i \text{ to } v_j \text{ in } G \\ 0 & \text{otherwise} \end{cases}$$

Example 1.3.2. *A directed graph and its adjacency matrix.*



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

If u and v are vertices in a directed graph, a *(directed) walk* from u to v is a sequence of vertices that starts at u and finishes at v, with the property that every pair of consecutive entries is a directed edge. The length of a path is the number of edges in it. In the example above, $v_5, v_6, v_1, v_2, v_1, v_2, v_3$ is a directed walk of length 6 from v_5 to v_3. The adjacency matrix has the interesting property that its powers count directed walks.

Theorem 1.3.3. *Let G be a directed graph with adjacency matrix A (with respect to the ordering $v_1, v_2, \ldots, v_n$ of the vertices). For every positive integer k, the (i, j) entry of A^k is the number of walks of length k from v_i to v_j in G.*

Proof. By induction on k. The case $k = 1$ is just the definition of the adjacency matrix.
So assume that the theorem is true for $k = m - 1$ and consider $k = m$. Then

$$(A^m)_{ij} = \sum_l A_{il} A_{lj}^{m-1}.$$

Every path of length m from v_i to v_j must start with an edge from v_i to some v_l and follow that with a path of length $m - 1$ from v_l to v_j. For each l, the entry A_{il} is 1 if (v_i, v_l) is an edge and 0 otherwise. By the induction hypothesis A_{lj}^{m-1} is the number of directed walks of length $m - 1$ from v_l to v_j in G. Thus for each vertex l, the integer $A_{il}A_{lj}^{m-1}$ is the number of walks of length m from v_i to v_j that have v_l as their second vertex. The sum over l of these is the total number of walks of length m from v_i to v_j in G. This completes the induction proof. □

An *undirected* graph resembles a directed graph except that the edges are *unordered* pairs of vertices. The adjacency matrix of an undirected graph is symmetric.

Note that the adjacency matrix of a (directed or undirected) graph G depends not only on the graph itself but also on the choice of an ordering of the vertices. Suppose that σ is a permutation of the set $\{1, \ldots, n\}$. Let A be the adjacency matrix of G with respect to the ordering $v_1, \mathtt{dots}, v_n$ of the vertices, and let A' be the adjacency matrix with respect to the ordering $v_{\sigma(1)}, \ldots, v_{\sigma(n)}$. Then A' is obtained from A by

- first reordering the columns by replacing Column 1 with Column $\sigma(1)$, Column 2 with Column $\sigma(2)$, and so on. This means multiplying on the right by the matrix P_σ, in which each Column j (for each j) has a 1 as its $\sigma(j)$-th entry and is otherwise full of zeros.

- then reordering the rows by replacing Row 1 with Row $\sigma(1)$, Row 2 with Row $\sigma(2)$, etc. This means multiplying A on the left by the matrix $(P_\sigma)^T$, which is also equal to P_σ^{-1}.

- A *permutation matrix* is a matrix that has exactly one 1 in each row and column and is otherwise full of zeros. Every permutation matrix has the property that its inverse is equal to its transpose. We have shown that adjacency matrices A and A' represent the same graph if and only if
$$A' = P^T A P,$$
for some permutation matrix P. This relation is known as *permutation equivalence*; it is a special case of both similarity and congruence.

The adjacency matrix is one of a number of matrices often associated with a graph. We mention a few more.

Definition 1.3.4. *Let* G *be an undirected graph with* n *vertices* $v_1, \ldots, v_n$ *and* m *edges* $e_1, \ldots, e_m$.

- *The* incidence matrix *of* G *is the* $n \times n$ *matrix* C *that has a* 1 *in position* (i, j) *if the vertex* v_i *is incident with the edge* e_j*, and zeros elsewhere.*

- *An* oriented incidence matrix *of* G *is the* $n \times m$ *matrix* B *defined by first assigning a direction to every edge of* G *and then setting*
$$B_{ij} = \begin{cases} 1 & \text{if } v_i \text{ is the start vertex of } e_j \\ -1 & \text{if } v_i \text{ is the end vertex of } e_j \\ 0 & \text{otherwise} \end{cases}$$

The oriented incidence matrix depends on a choice of ordering of both the vertices and edges, and on a choice of orientation of the edges.

Given matrices that are associated with graphs, a general philosophy is to consider how the properties of the matrix and the graph are related to each other. In the case of an oriented incidence matrix, the rank of the matrix tells us about the number of connected components in the graph.

Theorem 1.3.5. *Let* G *be a simple graph with* n *vertices and* m *edges, and let* B *be an oriented incidence matrix of* G*. Then the rank of* B *is* $n - t$*, where* t *is the number of connected components of* G*.*

Proof. First suppose that G is connected. This means that for any pair of vertices v_i and v_j in G, there exists a walk in G from v_i to v_j. We consider the left nullspace N of the matrix B. Note that every column of B has one entry equal to 1, one equal to -1, and is otherwise full of zeros. This means that the vector $(1\ 1\ \ldots\ 1)$ belongs to the left nullspace of V, so this nullspace is at least 1-dimensional.

Suppose that $(a_1\ a_2 \ldots a_n)$ is a non-zero element of N, and choose k for which $a_k \neq 0$, write $a_k = \alpha$. Then $(a_1\ a_2 \ldots a_n)$ must satisfy $(a_1\ a_2 \ldots a_n)v = 0$ for every column v of B, and in particular for those columns corresponding to edges that are incident with the vertex v_k. It follows that $a_i = a_k = \alpha$ whenever v_i is adjacent to v_k. Now by the same reasoning applied to the neighbours of v_k, we must have $a_j = \alpha$ whenever v_j is adjacent to a neighbour of v_k. Since G is connected, repetition of this step reaches all vertices of G and we conclude that $a_i = \alpha$ for all i and that N is a 1-dimensional space. Thus $n = 1 + \text{rank}(B)$ and $\text{rank}(B) = n - 1$.

Now suppose that G has t connected components and let their numbers of vertices be $n_1, n_2, \ldots, n_t$, with $m_1, m_2, \ldots, m_t$ edges respectively. By ordering the vertices component by component, we can arrange that B has a rectangular block diagonal structure with a $n_1 \times n_1$ block in the upper left, etc. Each block is an oriented incidence matrix of a connected component of G, and so a block with n_i vertices has rank $n_i - 1$. It follows that the total rank is
$$(n_1 - 1) + (n_2 - 1) + \cdots + (n_t - 1) = n - t.$$

□

Theorem 1.3.6. *Let* B *be an oriented incidence matrix of an undirected simple graph* G*. Then*

$$BB^T = D - A,$$

where D *is the diagonal matrix whose diagonal entries are the total degrees of the vertices, and* A *is the adjacency matrix of* A*.*

Proof. For $i = 1, \ldots, n$, the entry in the (i,i)-position of BB^T is just the ordinary scalar product of Row i of B with itself. Since every entry of this row is 1 or -1 or 0, this scalar product is the number of non-zero entries in Row i, which is the total degree of the vertex v_i.

Note that each Column of B has exactly two non-zero entries, which are equal to 1 and -1. For $i \neq j$, the entry in the (i,j) position of BB^T is the scalar product of Row i and Row j of B. If this is not zero it means that there is a Column whose only two non-zero entries occur in positions i and j, which means exactly that $v_i v_j$ is an edge in G. There can be at most one such column since there are no multiple edges in G. So the (i,j) entry of BB^T is -1 if v_i and v_j are adjacent in G and by 0 otherwise. We conclude that $BB^T = D - A$. □

Note that a consequence of Theorem 1.3.6 is that the matrix BB^T does not depend on the choice of orientation of the edges.

Definition 1.3.7. *Let* G *be an undirected graph with adjacency matrix* A*. The matrix* $L = D - A$ *is called the Laplacian matrix of* G*. Its entries on the main diagonal are the degrees of the vertices of* G*. Away from the main diagonal, the entry in position* (i,j) *is* -1 *or* 0 *according to whether* v_i *and* v_j *are adjacent or not.*

Properties of the Laplacian matrix of a graph G carry extensive information about properties of G itself. Moreover, as a real symmetric matrix it enjoys various special properties. For instance, it is a consequence of the following lemma that the rank of L tells us the number of connected components of G.

Lemma 1.3.8. *Suppose that* $A \in M_{n\times p}(\mathbb{R})$*. Then the rank of the* $n \times n$ *matrix* AA^T *is equal to the rank of* A*.*

Proof. That $\text{rank}(AA^T) \leqslant \text{rank}(A)$ is clear, since every columnn of AA^T is a real linear combination of the columns of A. We show now that in this special case, the right nullspace of AA^T is equal to the right nullspace of A^T. Suppose that $A^T v \neq 0$ for some $v \in \mathbb{R}^n$ Then $A^T v$ belongs to the columnspace of A^T, and since $A^T v$ is a non-zero vector in $\mathbb{R}^p$, it follows that $(A^T v)^T A^T v = v^T AA^T v \neq 0$. Thus $AA^T v \neq 0$, and v does not belong to the right nullspace of AA^T. Then the right nullspaces of A^T and AA^T coincide and have the same dimension d, and the ranks of AA^T and A^T (and A) are all equal to $n - d$. □

As a real symmetric matrix, L has the property that its eigenvalues are all real.

Lemma 1.3.9. *Let* A *be a complex Hermitian* $n \times n$ *matrix and let* λ *be a complex eigenvalue of* A*. Then* $\lambda \in \mathbb{R}$*.*

Note That A is Hermitian means that $A = A^*$, where A^* denotes the *Hermitian conjugate* of A, whose entries are the complex conjugates of the entries of A^T. A real symmetric matrix is a special case of a complex Hermitian matrix.

Proof. Since A is Hermitian we have for any vector $v \in \mathbb{C}^n$ that $v^* Av \in \mathbb{R}$, since

$$(v^* Av)^* = v^* A^* v = v^* Av.$$

Thus $v^* Av$ is a complex number that is equal to its own complex conjugate, hence it is real.
Now let $u \in \mathbb{C}^n$ be an eigenvector corresponding to λ. Then

$$u^* Au \in \mathbb{R} \Longrightarrow u^* \lambda u \in \mathbb{R} \Longrightarrow \lambda u^* u \in \mathbb{R}.$$

Since $v^* v \in \mathbb{R}$ (since it is the sum of the entries of u each multiplied by its own complex conjugate) it follows that $\lambda \in \mathbb{R}$ also. □

If G is a graph, it is a consequence of Lemma 1.3.9 that the eigenvalues of the Laplacian matrix L of G are real numbers. In fact they are all non-negative, for let v be an eigenvector of L corresponding to the eigenvalue λ, and let B be an oriented incidence matrix of G. Then

$$v^T L v = v^T \lambda v = \lambda v^T v.$$

On the other hand

$$v^T L v = v^T B B^T v = (v^T B)(B^T v) = (B^T v)^T (B^T v).$$

Since $v^T v$ is a positive real number and $(B^T v)^T (B^T v)$ is a non-negative real number, it follows that $\lambda \geqslant 0$.

How the eigenvalues of L are related to properties of G is one of the themes of *spectral graph theory*. We will be able to prove the following statements.

1. We know that 0 occurs at least once as an eigenvalue of L. We will show that it occurs exactly once if and only if G is connected.

2. If G is connected, let μ be the least positive eigenvalue of L. This number is called the *algebraic connectivity* of G. We will show that it can be considered as a measure of how robustly connected the graph is. It is bounded above by the *vertex connectivity*, which is the least number of vertices whose removal disconects G.

3. The determinant of any $(n-1) \times (n-1)$ submatrix of L is the number of *spanning trees* in G. A subgraph of G is a spanning tree if it involves all the vertices of G, is connected, and has no cycles.

We will return to these statements later after some investigations of determinants and eigenvalues, in general and for the special case of symmetric matrices.

Chapter 2

Spectral Properties

2.1 The determinant

Definition 2.1.1. *Let A be a $n \times n$ matrix with entries in a field $\mathbb{F}$. The* determinant *of A, written* $\det(A)$ *or* $|A|$*, is the element of $\mathbb{F}$ given by*

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \dots A_{n\sigma(n)},$$

where S_n is the group of all permutations of the set $\{1, 2, \dots, n\}$.

The *sign* of a permutation σ is defined by its parity; the sign is 1 if σ is even and -1 if σ is odd. A permutation is even if it can be written as the product of an even number of transpositions and odd if it can be written as the product of an odd number of transpositions; no permutation is both even and odd. We will write A_n for the subgroup of S_n consisting of all even permutations.

An way of thinking about the content of Definition 2.1.1 is by looking at all possible ways of taking the product of one entry from each row and column of A. The number of such products is $n!$, and each one of them determines a permutation of σ of $\{1, \dots, n\}$, where $\sigma(i)$ is the index of the column from which the contributing entry from Row i is taken. The determinant is obtained by subtracting the sum of the products corresponding to odd permutations from the sum of those corresponding to even permutations.

Example 2.1.2. *1.* $n = 2$

In this case there are two permutations, the identity which is even and the transposition $(1\ 2)$*, which is odd. Thus*

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}.$$

2. $n = 3$

In this case there are six permutations, three even and three odd. The even permutations are the identity and the two 3-cycles $(1\ 2\ 3)$ *and* $(1\ 3\ 2)$*. The odd permutations are the three transpositions* $(1\ 2)$*,* $(1\ 3)$ *and* $(2\ 3)$*. The determinant is given by*

$$\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33}.$$

OBSERVATIONS ON THE DEFINITION

1. Suppose that Rows i and j of the square matrix A are identical. Then $\det(A) = 0$.
   To see this, let τ denote the transposition that switches i and j and leaves all other points fixed. If σ is an even permutation, then $\sigma\tau$ (which means $\sigma \circ \tau$) is odd, and every odd

permutation in S_n has the form $\sigma\tau$ for some $\sigma \in A_n$. Now

$$\begin{aligned}\det(A) &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) A_{1\sigma(1)}A_{2\sigma(2)}\dots A_{n\sigma(n)} \\ &= \sum_{\sigma\in A_n} A_{1\sigma(1)}A_{2\sigma(2)}\dots A_{n\sigma(n)} - \sum_{\sigma\in A_n} A_{1\sigma\tau(1)}A_{2\sigma\tau(2)}\dots A_{n\sigma\tau(n)} \\ &= \sum_{\sigma\in A_n} (A_{1\sigma(1)}A_{2\sigma(2)}\dots A_{n\sigma(n)} - A_{1\sigma\tau(1)}A_{2\sigma\tau(2)}\dots A_{n\sigma\tau(n)}).\end{aligned}$$

Now $\sigma\tau(i) = \sigma(j)$ and $\sigma\tau(j) = \sigma(i)$, and $\sigma\tau(k) = \sigma(k)$ for $k \neq i, j$. Furthermore, since Rows i and j of A are identical, $A_{i\sigma\tau(i)} = A_{i\sigma(j)} = A_{j\sigma(j)}$, and $A_{j\sigma\tau(j)} = A_{j\sigma(i)} = A_{i\sigma(i)}$. Hence, for each $\sigma \in S_n$,

$$A_{1\sigma\tau(1)}A_{2\sigma\tau(2)}\dots A_{n\sigma\tau(n)} = A_{i\sigma(j)}A_{j\sigma(i)}\prod_{k\neq i,j} A_{k\sigma(k)} = A_{j\sigma(j)}A_{i\sigma(i)}\prod_{k\neq i,j} A_{k\sigma(k)} = A_{1\sigma(1)}A_{2\sigma(2)}\dots A_{n\sigma(n)}.$$

Thus each term in the last sum above is zero, and $\det(A) = 0$.

2. If the matrix A' is obtained from A by multiplying a single row by the scalar λ, then $\det(A') = \lambda \det(A)$. This follows from Definition 2.1.1, since the effect is to multiply each term in the sum once by λ. In particular, it follows from 1 that $\det(A) = 0$ if one row of A is a scalar multiple of another.

3. Suppose that u and v are row vectors of length n. Let $A(u)$, $A(v)$ and $A(u+v)$ be square matrices that respectively have u, v and $u+v$ as their ith rows and are otherwise identical. Then
$$\det A(u+v) = \det A(u) + \det A(v).$$
This observation is a straightforward deduction from Definition 2.1.1 but it has important consequences. Because of 1 above, it means that the act of adding a scalar multiple of one row to another row in a square matrix does not change the determinant. Applying this observation repeatedly means that adding any linear combination of Rws k (with $k \neq i$) to Row i in a square matrix does not change the determinant. Thus if a matrix has the property that one of its rows is a linear combination of the other $n-1$ rows, then its determinant is zero.

Our first main theorem in this section is the well-known (but not obvious) fact that a matrix is invertible if and only if its determinant is not zero.

Theorem 2.1.3. *Let $A \in M_n(\mathbb{F})$. Then A has an inverse in $M_n(\mathbb{F})$ if and only if $\det(A) \neq 0$.*

Proof. From Item 3. on page 10 we know that A has an inverse if and only if A has rank n, which occurs if and only if the rows of A form a basis for $(\mathbb{F}^n)^T$. Thus if A is not invertible, then one of its rows is a linear combination of the others, and it follows from 3 above that $\det A = 0$. On the other hand, suppose that A *is* invertible. Then each of the standard row vectors $e_1^T, \dots, e_n^T$ has a unique expression as a linear combination of the rows of A. At least one of these, say $e_{\sigma(1)}^T$, involves Row 1. This means that by adding a linear combination of Rows $2, \dots, n$ to Row 1 in A, we can obtain a new matrix A_1 whose determinant is the same as that of A, and whose only non-zero entry in Row 1 is in the $(1, \sigma(1))$ position. Furthermore A_1 has rank n and is invertible, since it has the same rowspace as A. Now there is an $e_{\sigma(2)}^T$ (different from $e_{\sigma(1)}$) whose unique expression as a combination of the rows of A_1 involves Row 2. By adding a combination of the other rows of A_1 to Row 2, we can obtain a matrix A_2 whose only non-zero entry in Row 2 is in the $(2, \sigma(2))$ position, and which has the same determinant as A_1 and A. Continuing in this manner, after n steps we produce a matrix A_n that has exactly one non-zero entry in each row and column (such a matrix is called *monomial*), and has the same determinant as A. According to Definition 2.1.1, the determinant of A_n is the product $\operatorname{sign}(\sigma)A_{1\sigma(1)}A_{2\sigma(2}\dots A_{n\sigma(n)}$, which is not zero. □

Our next goal will be to establish the famous Laplace expansion formula (or cofactor expansion formula) for determinants, and then use it to describe the inverse of a matrix A in terms of its entries.

We begin with the following lemma, of which item 1 in the observations above is a special case.

Lemma 2.1.4. *Let A be a square matrix and let A' be obtained from A by swapping two rows (or two columns). Then* $\det(A') = -\det(A)$.

Proof. Suppose that A' is obtained from A by swapping Row i and Row j. Let τ be the transposition $(i\,j)$, and note that $\text{sign}(\sigma\tau) = -\text{sign}(\sigma)$ for every $\sigma \in S_n$. Then

$$\begin{aligned}
\det A' &= \sum_{\sigma\in S_n} \text{sign}(\sigma) A'_{1\sigma(1)} A'_{2\sigma(2)} \dots A'_{n\sigma(n)} \\
&= \sum_{\sigma\in S_n} \text{sign}(\sigma) A_{i\sigma(j)} A_{j\sigma(i)} \prod_{k\neq i,j} A_{k\sigma(k)} \\
&= \sum_{\sigma\in S_n} \text{sign}(\sigma) A_{i\sigma\tau(i)} A_{j\sigma\tau(j)} \prod_{k\neq i,j} A_{k\sigma\tau(k)} \\
&= \sum_{\sigma\tau\in S_n} -\text{sign}(\sigma\tau) A_{1\sigma\tau(1)} A_{2\sigma\tau(2)} \dots A_{n\sigma\tau(n)} \\
&= -\det A.
\end{aligned}$$

□

Definition 2.1.5. *Let $A \in M_n(\mathbb{F})$. The* minor *of the entry in the (i,j) position of A, denoted M_{ij}, is the determinant of the $(n-1)\times(n-1)$ matrix that remains when Row i and Column j are deleted from A.*

Note that M_{ij} depends only on those entries of A that belong neither to Row i nor to Column j.

For $A \in M_n(\mathbb{F})$, the expression for $\det A$ in Definition 2.1.1 is a sum of $n!$ different terms, one for each permutation of n objects. The number of these terms that involve A_{11}, or A_{12}, or any particular A_{ij}, is $(n-1)!$. We can observe that the sum of all the terms that involve A_{11} is $A_{11}M_{11}$.

Choose a row (or column) of A - for example choose Row i. For j from 1 to n, let $A(j)$ denote the matrix that has A_{ij} in the (i,j)-position, has zeros otherwise throughout Row i, and has the same entries as A in the other rows. By item 3. in the observations above, we have

$$\det A = \sum_{j=1}^{n} \det A(j).$$

So we just need to figure out what $\det A(j)$ is. It seems to be "something like" $A_{ij}M_{ij}$ but we need to be careful about the signs on the permutations. We can use Lemma 2.1.4. Adjust the matrix $A(j)$ as follows.

- Move Row i into Row 1, and move Rows $1,\dots,i-1$ into Rows $2,\dots,i$. This means swapping Row i with Row $i-1$, then Row $i-1$ with Row $i-2$ and so on, eventually Row 2 with Row 1, to get the original Row i into the first row. This involves $i-1$ swaps of pairs of rows.
- Now move Column j into Column 1, and move Columns $1,\dots,j-1$ into Columns $2,\dots,j$. As above, this involves $j-1$ swaps of pairs of columns.

After all of these swaps, what we are left with is a matrix that has the original A_{ij} in the $(1,1)$ position, has zeros otherwise in Row 1, and in the lower right $n\times n$ region has the matrix that results from deleting Row i and Column j from the original A. This has been obtained from $A(j)$ by a total of $(i-1)+(j-1)$ swaps of pairs of rows or columns, and its determinant is $A_{ij}M_{ij}$. Each row or column swap changes the sign of the determinant, so by Lemma 2.1.4 we have

$$\det A(j) = (-1)^{i+j-2} A_{ij}M_{ij} = (-1)^{i+j} A_{ij}M_{ij}.$$

Then we obtain the *Laplace expansion formula* or *cofactor expansion formula* for Row i:

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

Definition 2.1.6. *For any position* (i, j) *in the matrix* A*, the* cofactor C_{ij} *is defined by* $C_{ij} = (-1)^{i+j} M_{ij}$.

So the cofactor expansion formulae are given for rows and columns of A by

- for Row i: $\det A = \sum_{j=1}^{n} A_{ij} C_{ij}$
- for Column j: $\det A = \sum_{i=1}^{n} A_{ij} C_{ij}$

Finally we describe how to use these formulae and our other observations on determinants, to write down the inverse of a (nonsingular) matrix A. Define the *adjugate* of A to be the $n \times n$ matrix given by

$$(\mathrm{adj} A)_{ij} = C_{ji}.$$

The adjugate is the transpose of the matrix of cofactors of A - its entry in the (i, j)-position is the cofactor of the entry in the (j, i) position of A.

Theorem 2.1.7. *Let* $A \in M_n(\mathbb{F})$*. Then*

$$A \times \mathrm{adj} A = \mathrm{adj} A \times A = \det A I_n.$$

Proof. We will prove the theorem for $A \times \mathrm{adj} A$ - the other part is similar.
Look at the diagonal entries first. For each i, the entry in the (i, i) position of $A \times \mathrm{adj} A$ is

$$\sum_{j=1}^{n} A_{ij} (\mathrm{adj} A)_{ji} = \sum_{j=1}^{n} A_{ij} C_{ij} = \det A,$$

by the Laplace expansion formula for Row i.

Now, the off-diagonal entries. Suppose that $i \neq k$. Then the (i, k) entry of $A \times \mathrm{adj} A$ is

$$\sum_{j=1}^{n} A_{ij} (\mathrm{adj} A)_{jk} = \sum_{j=1}^{n} A_{ij} C_{kj}.$$

By the Laplace expansion formula again, $\sum_{j=1}^{n} A_{ij} C_{kj}$ is exactly the determinant of the matrix that has entries $A_{i1}, A_{i2}, \ldots, A_{in}$ in Row k, and otherwise coincides with A. But this matrix has the same entries in Row i and Row k, hence its determinant is zero by Lemma 1. □

From Theorem 2.1.7, we conclude that if $\det A \neq 0$, then the inverse of A is given by

$$A^{-1} = \frac{1}{\det A} \mathrm{adj} A.$$

Finally, we note that the determinant is a *multiplicative function* from $M_n(\mathbb{F})$ to $\mathbb{F}$. For a proof of this statement, refer to Problem Sheet 3.

Theorem 2.1.8. *Let* $A, B \in M_n(\mathbb{F})$ *Then* $\det(AB) = \det(A) \det(B)$.

2.2 The spectrum of a matrix

Suppose that $A \in M_n(\mathbb{F})$ and that $v \in \mathbb{F}^n$ is an eigenvector of A with corresponding eigenvalue x. Then

$$Av = xv = xI_n v \iff xI_n v - Av = 0 \iff (xI_n - A)v = 0.$$

So x is an eigenvalue of A if and only if the right nullspace of the matrix $xI_n - A$ contains a non-zero vector, i.e. if and only if this matrix is singular. By Theorem **??**, this happens if and only if $\det(xI - A) = 0$.

Definition 2.2.1. *The equation* $\det(xI_n - A) = 0$ *is called the* characteristic equation *of* A*.*

Note that

$$xI_n - A = \begin{pmatrix} x - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & x - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \dots & x - a_{nn} \end{pmatrix}.$$

From the definition of the determinant, we can observe that $\det(xI - A)$ is a polynomial of degree n in x, whose leading term is x^n, this comes from the term in the determinant that is just the product of the entries on the main diagonal. All other terms have lower degree in x, since the do not involve all of the entries on the main diagonal - these are the only ones in which x appears.

Definition 2.2.2. *The expression* $\det(xI_n - A)$ *is called the* characteristic polynomial *of* A*. Its roots are the eigenvalues of* A*.*

We can identify how the coefficients in the characteristic polynomial depend on the entries of A.

1. The coefficient of x^n is 1.
2. All terms involving x^{n-1} come from the product of all n entries on the main diagonal. The coefficient of x^{n-1} in
$$(x - a_{11})(x - a_{22}) \dots (x - a_{nn})$$
is $-(a_{11} + a_{22} + \dots + a_{nn})$, which is $-\text{trace}(A)$.
3. The constant term is the determinant of $-A$, which is $(-1)^n \det A$.
4. In general, the coefficient of x^{n-k} is $(-1)^k P_k$, where P_k is the sum of the $k \times k$ *principal minors* of A. A $k \times k$ principal minor is the determinant of a $k \times k$ submatrix of A whose main diagonal is part of the main diagonal of A itself. Such a submatrix is called *principal* it consists of the intersection of Rows $i_1, i_2, \dots, i_k$ with Columns $i_1, i_2, \dots, ik$, for some subset $\{i_1, i_2, \dots, i_k\}$ of $\{1, \dots, n\}$. The number of $k \times k$ principal submatrices is $\binom{n}{k}$.

The eigenvalues of A are the roots of the characteristic polynomial. These may not belong to $\mathbb{F}$ but possibly to some extension of $\mathbb{F}$, such as the field of complex numbers. Let $p(x)$ denote the characteristic polynomial of A, and let $\lambda_1, \dots, \lambda_n$ denote its roots. Then

$$p(x) = (x - \lambda_1)(x - \lambda_2) \dots (x - \lambda_n).$$

Looking at the coefficients of $p(x)$ in terms of the λ_i gives

1. Constant term: $(-1)^n \lambda_1 \lambda_2 \dots \lambda_n = (-1)^n \det A$, so the determinant of A is the product of all the eigenvalues of A (counting multiplicity).
2. Coefficient of x^{n-1} : $-(\lambda_1 + \lambda_2 + \dots + \lambda_n) = -\text{trace}(A)$, so the trace of A (the sum of the 1×1 principal minors) is the sum of the eigenvalues.
3. Coefficient of x^{n-2} : $\prod_{i<j} \lambda_i \lambda_j$ is the sum of the 2×2 principal minors of A - this is the sum of the products in pairs of the eigenvalues.

4. In general, the coefficient of x^{n-k} is the sum of all products of k eigenvalues of A, multiplied by $(-1)^k$. So the sum of all products of k eigenvalues (there are $\binom{n}{k}$ of these) is the sum of the principal $k \times k$ minors (there are $\binom{n}{k}$ of these too).

The list of all eigenvalues of a matrix A is called the *spectrum* of A, denoted spec$A$. The number of times that a particular eigenvalue occurs as a root of the characteristic polynomial is called its *algebraic multiplicity*. If $\lambda \in \text{spec}A$, the *eigenspace* of A corresponding to λ is

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

This is a subspace of $\mathbb{F}^n$ whose dimension is called the *geometric multiplicity* of λ as an eigenvalue of A. The geometric multiplicity is $n - \text{rank}(\lambda I - A)$.

Lemma 2.2.3. *Similar matrices have the same eigenvalues, and the same algebraic and geometric multiplicities for each one.*

Proof. Suppose that A and B are similar matrices in $M_n(\mathbb{F})$ and write $B = P^{-1}AP$ for some $P \in \text{GL}(n, \mathbb{F})$. Then

$$\begin{aligned} \det(xI - B) &= \det(xI - P^{-1}AP) \\ &= \det(P^{-1}(xI - A)P) \\ &= \det P^{-1} \det(xI - A) \det(P) \\ &= \det(xI - A). \end{aligned}$$

Thus A and B have the same characteristic polynomial, and hence have the same eigenvalues with the same algebraic multiplicities.

To see that they have the same geometric multiplicities also, let λ be an eigenvalue of A (and B), and let U_A and U_B respectively denote the eigenspaces of A and B corresponding to λ. Let $u \in U_A$. Then

$$B(P^{-1}u) = P^{-1}APP^{-1}u = P^{-1}Au = P^{-1}\lambda u = \lambda(P^{-1}u).$$

Thus $P^{-1}u$ is an eigenvector of B corresponding to λ, and $P^{-1}U_A \subseteq U_B$. Since $P^{-1}U_A$ has the same dimension as U_A, it follows that $\dim U_A \leqslant \dim U_B$. On the other hand $A = PBP^{-1}$, and the same argument shows that $PU_B \subseteq U_A$, hence that $\dim U_B \leqslant \dim U_A$. Thus $\dim U_A = \dim U_B$, and the geometric multiplicities of λ for A and B coincide. □

We now turn our attention, for a while, to real and complex matrices. The real and complex fields are somewhat special because they are equipped with the notion of modulus or *absolute value*, which allows us to say whether one field element has greater magnitude than another. Recall that the *modulus* or *absolute value* of a complex number $z = a + bi$ is the non-negative real number given by

$$|z| = |a + bi| = \sqrt{a^2 + b^2}.$$

The spectrum of a complex $n \times n$ matrix A is a list of n complex numbers. These may be interpreted as points in the complex plane.

Definition 2.2.4. *The* spectral radius $\rho(A)$ *of a matrix $A \in M_n(\mathbb{C})$ is the maximum of the moduli of the eigenvalues of A.*

Thus all of the eigenvalues of A are located in the closed disc of radius $\rho(A)$ centred at 0 in the complex plane, and at least one of them is located on the boundary of this disc.

Example 2.2.5. *Let* $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -2 & 1 \end{pmatrix}$.
The characteristic polynomial of A is $x^3 - x^2 + 2x - 2 = (x-1)(x^2+2)$. The eigenvalues of A are $1, 1+\sqrt{2}i$ and $1-\sqrt{2}i$ and the spectral radius is $\sqrt{3}$.

For certain matrices with specified properties, we can say a bit more.

1. If A is an upper or lower triangular matrix (over any field), then its eigenvalues are just the entries on its main diagonal.
2. The spectral radius of A is 0 if and only if A is nilpotent, i.e. $A^n = 0$.
3. If A is a complex Hermitian matrix (or a real symmetric matrix) then its eigenvalues are *real numbers* - we saw this in Lemma 1.3.9.
4. If A is a complex Hermitian positive definite matrix (this means that $x^*Ax > 0$ for all $x \in \mathbb{C}^n$, then the eigenvalues of A are positive real numbers.
5. If $A \in M_n(\mathbb{R})$, then the spectrum of A is a list of complex numbers that contains the complex conjugate of each of its entries (because it is the list of roots of a polynomial with real coefficients). If n is odd, A has at least one real eigenvalue.
6. If A is a positive real matrix (this means that all of the entries of A) are positive, then the spectral radius of A is actually an eigenvalue of A, with algebraic multiplicity 1. This is (part of) the Frobenius-Perron Theorem, which we will discuss later.

If A is diagonal (or triangular) its eigenvalues are its diagonal entries. For any A, the sum of its eigenvalues is the sum of its diagonal entries. The product of the diagonal entries is one of the $n!$ different terms that contributes to the determinant of A, which is the product of the eigenvalues. So maybe there is some (loose, tenuous) connection between diagonal entries and eigenvalues. Our next theorem, which dates back to 1931, makes this connection a bit more precise in the case of complex matrices.

Theorem 2.2.6. *(Geršgorin's Circle Theorem, part 1) Let $A \in M_n(\mathbb{C})$. For $i = 1, \dots n$, let r_i be the sum of the moduli of the off-diagonal entries of Row i of A. Let D_i be the closed disc in the complex plane that has centre A_{ii} (the diagonal entry in Row i of A) and radius r_i. Then every eigenvalue of A belongs to some D_i.*

Example 2.2.7. *In $M_2(\mathbb{C})$, let $A = \begin{pmatrix} 0 & 2 \\ 3 & -1 \end{pmatrix}$. The eigenvalues of A are -3 and 2. The Geršgorin Circle Theorem says that each of these numbers lies either in the disc of radius 2 centred at 0 or the disc of radius 3 centred at -1. In fact they both lie in the latter of these discs in this particular example (one lies in the intersection of both discs).*

Proof. (of Theorem 2.2.6) Let $\lambda \in \mathbb{C}$ be an eigenvalue of A and let $v \in \mathbb{C}^n$ be a corresponding eigenvector, with entries $v_1, \dots, v_n$. Choose some i for which the modulus of v_i is maximal amongst the entries of v. Then looking at entry i of Av gives

$$\sum_{j=1}^{n} A_{ij}v_j = \lambda v_i, \text{and } \sum_{j \neq i} A_{ij}v_j = (\lambda - A_{ii})v_i.$$

Since $v_i \neq 0$ (as $v \neq 0$ and v_i is an entry of greatest modulus in v), we may rearrange this to get

$$\lambda - A_{ii} = \frac{1}{v_i} \sum_{j \neq i} A_{ij}v_j.$$

Taking moduli and using the triangle inequality, we get

$$\begin{aligned} |\lambda - A_{ii}| &= \left| \frac{1}{v_i} \sum_{j \neq i} A_{ij}v_j \right| \\ &= \left| \sum_{j \neq i} A_{ij} \frac{v_j}{v_i} \right| \\ &\leqslant \sum_{j \neq i} |A_{ij}| \frac{|v_j|}{|v_i|}. \end{aligned}$$

From the choice of i it follows that $\frac{|v_j|}{|v_i|} \leqslant 1$ for for all $j \neq i$, hence

$$|\lambda - A_{ii}| \leqslant \sum_{j \neq i} |A_{ij}|,$$

which is saying exactly that λ is an element of the disc D_i. □

Note: The discs D_i are referred to as Geršgorin discs.

Example 2.2.8. *Let* $A = \begin{pmatrix} 2 & -5 \\ 1 & -2 \end{pmatrix}$. *The eigenvalues of* A *are* i *and* $-i$ *(since* $\det A = 1$ *and* $\text{trace} A = 0$*). The Geršgorin discs for* A *are*

- D_1*: centre* 2*, radius 5*
- D_2*: centre* -2*, radius 1.*

In this example, neither of the eigenvalues belongs to D_2*, they both belong only to* D_1*.*

So it is not necessarily true that each Geršgorin disc contains an eigenvalue. The following extension to the theorem gives some more detail about the connection between the distribution of eigenvalues and the geometry of the union of discs.

Theorem 2.2.9. *Let* $A \in M_n(\mathbb{C})$*, with Geršgorin discs* $D_1, \ldots, D_n$*, centred at* $A_{11}, A_{22}, \ldots, A_{nn}$ *respectively. Suppose that the union* R_1 *of* k *of these discs is disjoint from the union* R_2 *of the other* $n - k$*. Then* R_1 *contains* k *eigenvalues of* A *and* R_2 *contains* $n - k$*.*

Proof. Let D be the diagonal matrix whose diagonal entries are the same as those of A, and let t be a continuous real variable. Define

$$A(t) = (1 - t)D + t(A),$$

so $A(t)$ is the matrix whose main diagonal is the same as that of A, and whopse entries otherwise coincide with those of tA. In particular $A(0) = D$ and $A(1) = A$ Let the Geršgorin disc of $A(t)$ correpsonding to Row i be $D_i(t)$. Then $D_i(t)$ has centre A_{ii} for all t, and the radius of $D_i(t)$ is $t \times \text{radius}(D_i)$.

The eigenvalues of the diagonal matrix $A(0)$ are just the diagonal entries of A, and the discs $D_i(0)$ all have radius 0. In particular k of these eigenvalues are in R_1 and the other k are in R_2. As t increases from 0 to 1, the discs $D_i(t)$ expand but their centres do not move. The union of discs corresponding to the k values of i that contribute to R_1 reamins in R_1 as t increases from 0 to 1, and the same is true of the values of i contributing to R_2. The eigenvalues of $A(t)$ vary continuously with t, and so k of them must remain in R_1 and $n - k$ in R_2 throughout this process.

□

2.3 Positive matrices - the Frobenius-Perron Theorem

In this section we are considering matrices whose entries belong to the field $\mathbb{R}$ of real numbers. In $\mathbb{R}$ (but not in other fields such as $\mathbb{C}$ for example), non-zero elements are either positive or negative, in fact $\mathbb{R}$ is an example of an ordered field. This is a very familiar property but actually it is quite special. It allows us to define the notion of a *positive matrix* and to investigate what special properties positive matrices might have.

Definition 2.3.1. *A matrix in* $M_{p \times q}(\mathbb{R})$ *is* positive *if all of its entries are positive (and* non-negative *if all of its entries are non-negative). We write* $A > 0$ *and* $A \geqslant 0$ *to indicate that a matrix* A *is positive or non-negative. The difference is that some entries of a non-negative matrix may be zero.*

Positive square matrices (and certain classes of non-negative matrices) have some special spectral properties that are often collected together into the statement of the Frobenius-Perron Theorem.

Theorem 2.3.2. *Let A be a $n \times n$ positive matrix with spectral radius ρ. Then $\rho > 0$ and*

1. *ρ is an eigenvalue of A.*
2. *ρ has algebraic multiplicity 1 as an eigenvalue of A.*
3. *There is an eigenvector v of A corresponding to ρ that has all of its entries positive.*
4. *If λ is an eigenvalue of A and $\lambda \neq \rho$, then $|\lambda| < \rho$.*
5. *If u is an eigenvector of A (corresponding to any eigenvalue) whose entries are all positive, then u is a scalar multiple of v (from 3. above).*

The key point of Theorem 2.3.2 is Item 1, which is more significant than it might look at first glance. In general, for a matrix A in $M_n(\mathbb{R})$ or $M_n(\mathbb{C})$, the spectral radius is just the maximum of the moduli of the eigenvalues. In general there is no reason to expect that the spectral radius is itself an eigenvalue, since the eigenvalue of greatest modulus need not be real, and if it is real it need not be positive. So the situation for positive matrices is really special - there is an eigenvalue of greatest modulus that is a positive real number (this is sometimes called the *Perron root*. Not only that but every other eigenvalue has modulus strictly less than that of the Perron root, there is only one eigenvalue on the circle of radius ρ. Not only that, but this eigenvalue has a corresponding eigenvector in which all entries are positive, and no other eigenvalue has this property.

Example 2.3.3. *Let* $A = \begin{pmatrix} 1 & 3 \\ 4 & 5 \end{pmatrix}$.
The characteristic polynomial of A is $x^2 - 6x - 7$ or $(x-7)(x+1)$; the eigenvalues are 7 and -1. The spectral radius is 7. Eigenvectors corresponding to $\lambda = 7$ and $\lambda = 1$ respectively are $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ *and* $\begin{pmatrix} 3 \\ -2 \end{pmatrix}$.

Before proving this theorem in general, we consider what it says about the 2×2 case. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a 2×2 matrix with positive real entries. Let the eigenvalues of A be λ and μ. Then

- $\mu + \lambda = a + d$.
- $\mu\lambda = ad - bc$

Either μ and λ are both real or they are complex numbers that are complex conjugates of each other. We first show that they must be real.

Lemma 2.3.4. *$\mu \in \mathbb{R}$ and $\lambda \in \mathbb{R}$.*

Proof. Suppose not, and write $\mu = x + yi$, $\lambda = x - iy$. Then $2x = a + d$ and $x^2 + y^2 = ad - bc$. Now

$$\begin{aligned} x^2 + y^2 &= \frac{1}{4}(a^2 + d^2 + 2ad) + y^2 = ad - bc \\ &\implies \frac{1}{4}(a-d)^2 + y^2 = -bc \end{aligned}$$

Since $-bc$ is negative, this is impossible. □

Thus the eigenvalues of A are both real, and since their sum is positive at least one of them is positive. If one is positive and one negative, then the positive one must have the greater absolute value. Note that A cannot have a repeated real eigenvalue, thus is ruled out by the above argument with $y = 0$. This proves items 1, 2 and 4 of the Theorem for the 2×2 case.

Now suppose that $\mu < \lambda$. Then

$$\begin{aligned} \lambda + \mu &= a + d \\ \implies 2\lambda &> a + d \\ \implies \lambda &> \frac{a+d}{2}. \end{aligned}$$

This means that either $\lambda > a$ or $\lambda > d$ (or both). Let $v \in \mathbb{R}^2$ be an eigenvector of A, with $v = \binom{v_1}{v_2}$. Then neither v_1 nor v_2 can be equal to zero. To see this note that if $v_1 = 0$ and $v_2 \neq 0$, then the first component of Av is not zero. Thus we can choose v with $v_1 = 1$. Then

$$Av = \lambda v \implies a + bv_2 = \lambda, \;\; c + dv_2 = \lambda v_2.$$

If $\lambda > a$ then since b is positive it follows from the first equation that v_2 is positive. If $\lambda \not> a$ then $\lambda > d$ and the second equation says $c = (\lambda - d)v_2$. Since c and $\lambda - d$ are positive, it follows that v_2 is positive. Thus both entries of v are positive, which proves part 3. of the theorem in the 2×2 case.

Finally, let $u = \binom{u_1}{u_2}$ be an eigenvector of A corresponding to μ, and as above suppose that $u_1 = 1$. As above we have

$$\mu + \lambda = a + d \implies 2\mu < a + d \implies \mu < a \text{ or } \mu \leqslant d$$

and

$$Au = \mu u \implies a + bu_2 = \mu \text{ and } c + du_2 = \mu u_2.$$

If $\mu < a$ then the first equation implies that $u_2 < 0$. If $\mu \not< a$ then $\mu < d$ and the second equation says $c = (\mu - d)u_2$ which means that u_2 is negative since c is positive. So the entries of u are of opposite sign. This completes the proof of the Frobenius-Perron Theorem for $n = 2$.

We now embark on the proof of the Frobenius-Perron Theorem, starting with the following lemma.

Lemma 2.3.5. *Let A be a positive $n \times n$ matrix. Then A has a positive eigenvalue with a positive corresponding eigenvector.*

Proof. Let A be a positive matrix in $M_n(\mathbb{R})$.
Let S denote the set of vectors x in $\mathbb{R}^n$ that have all entries non-negative and satisfy $\|x\| = 1$. This is the intersection of the sphere S^{n-1} with the set of non-negative vectors in $\mathbb{R}^n$. It is a compact subset of $\mathbb{R}^n$.

If $x \in S$, then all entries of Ax are positive. To see this, just think about any entry of Ax. The non-zero entries of x all contribute positively to this, and there are no negative contributions. We define a function $L : S \to \mathbb{R}_{>0}$ as follows. For $x \in S$,

$$L(x) = \min \left\{ \frac{(Ax)_i}{x_i} : x_i \neq 0 \right\}.$$

To understand what the function L does, start with $x \in S$. Compare the vectors x and Ax entry by entry. Look at those positions i in which x has a positive entry x_i. For each of these i, the ith entry of Ax is also positive, so it is x_i multiplied by some positive scaling factor α_i. The least of these α_i is what we are calling $L(x)$. It is a positive real number.

That is how $L(x)$ is defined for a particular $x \in S$, and L is a continuous function from S to the set of positive real numbers. Since S is compact, this means that L has a maximum value on S. Call this ρ, and let $v \in S$ be a vector for which $L(v) = \rho$.

We will show that ρ is an eigenvalue of A and that v is a corresponding eigenvector, and that $v > 0$. There are two steps.

1. First we show that $Av = \rho v$. We know that $Av \geqslant \rho v$ since $L(v) = \rho$, this means that $(Av)_i \geqslant \rho v_i$ for all i. Thus $Av - \rho v \geqslant 0$. This means that $A(Av - \rho v)$ is a positive vector

and so we can choose $\epsilon > 0$ small enough that $A(Av - \rho v) > \epsilon Av$. The vector Av may not belong to S, but there is a positive real number c for which $cAv \in S$.

$$A(Av) > (\rho + \epsilon)Av \Longrightarrow A(cAv) \geqslant (\rho + \epsilon)cAv \Longrightarrow L(cAv) \geqslant (\rho + \epsilon).$$

This contradicts the choice of ρ as the maximum value of L on S, and we conclude that $Av = \rho v$.

2. Secondly, we know that $v \geqslant 0$ since $v \in S$. It follows that $Av > 0$ - no entry of Av can be equal to zero since v is a non-negative non-zero vector and A is positive. Hence ρv is strictly positive and so v is strictly positive also.

□

Lemma 2.3.6. *The spectral radius of A is ρ.*

Proof. Let μ be any eigenvalue of A, and let y be a corresponding eigenvector, with $\|y\| = 1$. Bear in mind that μ, and the entries of y, need not be real numbers. Now look at entry i of Ay and μy.

$$\begin{aligned} \mu y &= Ay \\ \Longrightarrow \mu y_i &= \sum_{j=1}^{n} A_{ij} y_j \\ \Longrightarrow |\mu y_i| &\leqslant \sum_{j=1}^{n} |A_{ij} y_j| \\ \Longrightarrow |\mu|\,|y_i| &\leqslant \sum_{j=1}^{n} A_{ij} |y_j|. \end{aligned}$$

Let $|y|$ denote the vector whose entries are the moduli of the entries of y. Then $|y| \in S$, and the last statement above says that each entry of the vector $A|y|$ is at least equal to $|\mu|$ multiplied by the corresponding element of $|y|$. This means exactly that $L(|y|) \geqslant |\mu|$. Since ρ is the maximum value of L on S, it follows that $|\mu| \leqslant \rho$. Thus ρ is the spectral radius of A. □

We have now proved parts 1. and 3. of Theorem 2.3.2, but we have not yet fully proved any of the other parts.

Lemma 2.3.7. *ρ has geometric multiplicity 1 as an eigenvalue of A.*

Proof. We know that v is a positive eigenvector of A corresponding to ρ. Suppose, anticipating contradiction, that u is an eigenvector of A corresponding to ρ, and that u is independent of v over $\mathbb{C}$.

We may assume that the entries of u are real, since ρ is real. If u has entries that are non-real complex numbers, then the real and imaginary part of u would separately be eigenvectors of A and at least one of them would be independent of v.

Now, according to this hypothesis, every element of the 2-dimensional space spanned by u and v (over $\mathbb{C}$ or $\mathbb{R}$) is an eigenvector of A corresponding to ρ. Since $v > 0$, there is a real number ϵ with the property that $u' = v + \epsilon u$ is a non-negative vector with at least one entry equal to zero. However $u' \neq 0$ since u and v are independent.

This is the required contradiction, since Au' would be positive in this case and could not be a scalar multiple of u'. □

Lemma 2.3.8. *The algebraic multiplicity of ρ as an eigenvalue of A is 1.*

Proof. The key to this step is to show that A is similar to a (real) matrix A' that has the entry ρ in the $(1,1)$ position and zeros throughout the rest of Row 1 and Column 1.

Since A and its transpose have the same characteristic polynomial and hence the same spectrum, the spectral radius of A^T is ρ. Our proof of Lemma 2.3.5 shows that there is a positive

column vector w that is an eigenvector of A^T corresponding to ρ. Thus $A^T w = \rho w$ and the row vector w^T satisifies

$$w^T A = \rho w^T.$$

Now let U be the $(n-1)$-dimensional orthogonal complement of w with respect to the ordinary scalar product on $\mathbb{R}^n$:

$$U = \{u \in \mathbb{R}^n : w^T u = 0\}.$$

Let $u \in U$, and consider the vector $Au \in \mathbb{R}^n$. Note that

$$w^T Au = \rho w^T u = 0,$$

so $Au \in U$ whenever $u \in U$. This means that the subspace U of $\mathbb{R}^n$ is A-*invariant*. This is because U is the orthogonal complement in $\mathbb{R}^n$ of a left eigenvector of A, it has nothing to do with the positivity of A or the special properties of ρ and w. However these special properties give us an important extra piece of information.

Let v be the positive eigenvector of A corresponding to ρ, whose existence was shown in Lemma 2.3.5. Then $v \notin U$ since $w \cdot v = w^T v > 0$, because w and v are both positive. Let $\{b_1, \ldots, b_{n-1}\}$ be a basis of U. Then $\mathcal{B} = \{v, b_1, \ldots, b_{n-1}\}$ is a basis of $\mathbb{R}^n$.

Now the matrix A' that describes the linear transformation of $\mathbb{R}^n$ determined by left multiplication by A, with respect to the basis $\mathcal{B}$, has the following form:

$$A' = \left(\begin{array}{c|ccc} \rho & 0 & \ldots & 0 \\ \hline 0 & & & \\ \vdots & & B_{(n-1)\times(n-1)} & \\ 0 & & & \end{array}\right)$$

Here B is $n \times n$ matrix with real entries. Since A and A' are similar, ρ occurs as an eigenvalue of both, with the same algebraic multiplicity and with geometric multiplicity 1 in each case (by Lemmas 2.2.3 and 2.3.7). The characteristic polynomial of A' is $(x-\rho)p_B(x)$, where $p_B(x)$ is the characteristic polynomial of B. If the algebraic multiplicity of ρ as an eigenvalue of A' exceeds 1, then ρ is an eigenvalue of B with a corresponding eigenvector $v_B \in \mathbb{R}^{n-1}$. This means that the vector in $\mathbb{R}^n$ obtained by preceding v with a zero entry is an eigenvector of A' corresponding to ρ. Since e_1 is also an eigenvector of A' corresponding to ρ, this means that ρ has geometric multiplicity at least 2 as an eigenvector of A', and hence also as an eigenvector of A. This contradiction to Lemma 2.3.7 completes the proof, and we conclude that ρ occurs once as a root of the characteristic polynomial of A. □

Now we come to part 5., which is easy at this stage.

Lemma 2.3.9. *Let u be a positive eigenvector of A. Then u is a real positive scalar multiple of v.*

Proof. Let μ be the eigenvalue of A to which u corresponds. Then, μ is real and $\mu > 0$, since A and u are positive and $Au = \mu u$. Thus $0 < \mu \leqslant \rho$. Choose ϵ small enough that $u' = v - \epsilon u$ is positive. For each i we have

$$(Au')_i = \rho v_i - \mu\epsilon u_i \geqslant \rho(v_i - \epsilon u_i) = \rho u'_i.$$

Thus $Au' \geqslant \rho u'$, which means that $Au' = \rho u'$ by the maximality of ρ as a value of the function L. This means that u' is a ρ-eigenvector of A, which means that u', hence u, is a scalar multiple of v and $\mu = \rho$. □

The only item remaining is Item 4.

Lemma 2.3.10. *Suppose that μ is an eigenvalue of A, $\mu \neq \rho$. Then $|\mu| < \rho$.*

Proof. Suppose, anticipating contradiction, that $|\mu| = \rho$, and let y be an eigenvector of A corresponding to μ, with $\|y\| = 1$. Let $|y|$ denote the vector in $\mathbb{C}^n$ whose entries are the moduli of the entries of y. Then $|y| \in S$ and for each i we have

$$(A|y|)_i = \sum_j A_{ij}|y_j| = \sum |A_{ij}y_j| \geqslant \left|\sum_j A_{ij}y_j\right| = |\mu y_i| = \rho|y_i|.$$

Thus $A|y| \geqslant \rho|y|$ and by Lemmas 2.3.5 and 2.3.7 this means that $|y|$ is a ρ-eigenvector of A and $|y| = v$. Then equality holds in the triangle inequality above and we have for each i that

$$\sum_j |A_{ij}y_j| = \left|\sum_j A_{ij}y_j\right|.$$

So $A_{i1}y_1,\ A_{i2}y_2, \ldots, A_{in}y_n$ are complex numbers with the property that the sum of their moduli is the modulus of their sum. This means that they all lie on the same ray in the complex plane (a ray is a half-line with its endpoint at 0). Since the numbers A_{ij} are all real and positive, this means that $y_1, \ldots, y_n$ all lie on the same ray. Hence there is some θ for which $e^{i\theta}y$ is a positive vector. Thus y is a (complex) scalar multiple of a positive vector, and since ρ is the only eigenvalue of A to have a positive corresponding eigenvector, it follows that $\mu = \rho$. Thus the only eigenvalue of A to have modulus ρ is ρ itself, and every other eigenvalue has modulus strictly less than the spectral radius. □

This completes the proof of the Frobenius-Perron theorem.

The theorem was proved by Perron for positive matrices in 1907, and extended to a slightly broader class of non-negative matrices by Frobenius in 1912. The proof in our lecture notes is mostly due to Wielandt (1950).

We conclude with some slight extensions of the theorem. A non-negative square matrix is one whose entries are all non-negative real numbers (they can be zero).

Definition 2.3.11. *A non-negative square matrix A is called* primitive *if A^k is positive for some positive integer k. The Frobenius-Perron theorem as we have stated it holds for primitive non-negative matrices as well as positive matrices.*

A slightly weaker version of the Perron-Frobenius Theorem holds for *irreducible non-negative matrices*. The concept of irreducibility is most easily explained by reference to a graph. Associated to a non-negative $n \times n$ matrix A is the directed graph G on n vertices in which there is an arc from vertex i to vertex j if and only if the entry A_{ij} is non-zero (i.e. positive). The graph G is *strongly connected* if every vertex can be reached from every other by a path that follows the direction of the arcs. An example of a non-negative matrix that is *not* irreducible is $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

An example of a non-negative matrix that is irreducible but not primitive is

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

The corresponding directed graph has arcs $1 \to 2,\ 2 \to 3$ and $3 \to 1$, but no power of the matrix is positive.

We have the following version of the Frobenius Perron theorem for irreducible matrices.

Theorem 2.3.12. *Let $A \in M_n(R), A \geqslant 0$. If A is irreducible, then*

1. *The spectral radiusρ of A is an eigenvalue, with a corresponding positive eigenvector v.*
2. *ρ has algebraic (and geometric) multiplicity 1 as an eigenvalue of A.*
3. *Every positive eigenvector of A is a scalar multiple of ρ.*

4. *It is not necessarily true that* ρ *is the only eigenvalue whose modulus is equal to* ρ*. The number of such eigenvalues is the greatest common divisor of the lengths of all closed paths in the directed graph of* A*, and they are evenly spaced around the circle* $r = \rho$*.*

2.4 Supplement to Chapter 2: even and odd permutations

Definition 2.4.1. *The group consisting of all permutations of a set of* n *elements is called the* symmetric group of degree n *and denoted* S_n.

REMARKS

1. The order of S_n is $n!$, the number of permutations of n objects (read this as "n factorial").
2. We often think of the n elements being permuted as the first n positive integers $1, 2, \ldots, n$, but this is not intrinsic to the definition of S_n. It doesn't really matter what these elements are called as long as they have distinct labels.
3. Although the terminology is potentially problematic, it is important not to confuse the term "symmetric group" with groups of symmetries of (for example) regular polygons.

This section is mostly about how to represent permutations and how to do calculations with them. Later in the chapter we will use this information to deduce some nice properties of the symmetric groups.

An element of S_4 is a permutation of the set $\{1, 2, 3, 4\}$; this means a function from that set to itself that sends each element to a different image, and hence shuffles the four elements. In S_4, a basic way to represent the permutation $1 \to 1,\ 2 \to 4,\ 3 \to 2,\ 4 \to 3$ is by the array

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 2 & 3 \end{pmatrix}.$$

Representing permutations like this we can practise multiplying (or composing) them. In these notes we will use the convention that for permutations σ and τ, the product $\sigma\tau$ means "σ after τ or $\sigma \circ \tau$, i.e. that the factor that is written on the right is applied first. This is not a universally agreed convention and people use both possible interpretations. For this course it is probably a good idea that we all share the same interpretation to avoid confusion, but in general all that is important is that you state in which order you are considering the composition to take place and that you are consistent.

Example 2.4.2. *In* S_5*, suppose that*

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 5 & 4 & 1 \end{pmatrix},\ \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 3 & 5 & 1 \end{pmatrix}.$$

Calculate the products $\sigma\tau$ *and* $\tau\sigma$.

Solution: To calculate $\sigma\tau$, we apply τ first and then σ. Remember that this is just a composition of functions.

- τ sends 1 to 4, then σ sends 4 to 4. So $\sigma\tau$ sends 1 to 4.
- τ sends 2 to 2, then σ sends 2 to 3. So $\sigma\tau$ sends 2 to 3.
- τ sends 3 to 3, then σ sends 3 to 5. So $\sigma\tau$ sends 3 to 5.
- τ sends 4 to 5, then σ sends 5 to 1. So $\sigma\tau$ sends 4 to 1.
- τ sends 5 to 1, then σ sends 1 to 2. So $\sigma\tau$ sends 5 to 2.

We conclude that

$$\sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 5 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 3 & 5 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 5 & 1 & 2 \end{pmatrix}$$

$$\tau\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 3 & 5 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 5 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 1 & 5 & 4 \end{pmatrix}$$

This array format is not the only way of representing a permutation and not always the most useful way. Another way of thinking about a permutation π is by thinking about how it moves the elements of the set around, by starting with a single element and looking at the sequence of images when you repeatedly apply π to it. Eventually you will have to get back to the original element. Consider the following example in S_{14}.

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\ 11 & 9 & 8 & 2 & 5 & 1 & 12 & 14 & 6 & 7 & 3 & 13 & 10 & 4 \end{pmatrix}$$

Start with the element 1 and look at what happens to it when you repeatedly apply π.

- First you get $1 \to 11$;
- Then $11 \to 3$;
- Then $3 \to 8$;
- Then $8 \to 14$;
- Then $14 \to 4$;
- Then $4 \to 2$;
- Then $2 \to 9$;
- Then $9 \to 6$;
- Then $6 \to 1$.

After nine applications of π we arrive back at 1 and this is the first time we have a repetition in the list. This will happen every time: the list can't continue indefinitely without repetition because there are only finitely many elements being permuted. Suppose that after starting at 1 the first repetition occurs at Step k, after k applications of π. Then we have

$$1 \to a_1 \to a_2 \to \cdots \to a_{k-1} \to$$

where $1, a_1, \ldots, a_{k-1}$ are distinct. The next element (a_k) is a repeat of one of these. However it can't be a repeat of a_1, because 1 is the only element whose image under π is a_1, and $a_{k-1} \neq 1$. The same applies to $a_2, \ldots, a_{k-1}$. So it must be that 1 (the element where we started) is the first element to be repeated, and that we close the circle that started with 1. In our example above there were nine distinct elements in the sequence that started at 1. So the permutation π produces the following *cycle*:

$$1 \to 11 \to 3 \to 8 \to 14 \to 4 \to 2 \to 9 \to 6 \to 1$$

This cycle is often written using the following notation:

$$(1\ 11\ 3\ 8\ 14\ 4\ 2\ 9\ 6).$$

Note that 1 is not written at the end here. The above notation means the permutation (of 14 elements in this case) that sends 1 to 11, 11 to 3, etc, and sends 6 back to 1. There is nothing in the notation to indicate that we are talking about an element of S_{14} - this has to be clear from the context. Also, it is understood that elements that are not mentioned in the above notation are fixed by the permutation that it denotes. The permutation (1 11 3 8 14 4 2 9 6) is an example of a *cycle of length 9* in S_{14}. It is not the same as the permutation π that we started with, but it does coincide with π on the set of nine elements that can be obtained by starting at 1 and repeatedly applying π. This set is called the *orbit* of 1 under π.

The point of this discussion is that π can be written as a product (or composition) of *disjoint* cycles in S_{14}. The next step towards doing so is to look for the first element (in the natural order)

of our set that is not involved in the first cycle. This is 5. Go back to π and see what happens to 5 under repeated application of π. We find that

$$5 \to 5,$$

so 5 is fixed by π. We could think of this as a cycle of length 1.

There are still some elements unaccounted for. The first one is 7. Looking at the orbit of 7 under π, we find

$$7 \to 12 \to 13 \to 10 \to 7$$

so we get the cycle $(7\ \ 12\ \ 13\ \ 10)$ of length 4. Note that this has no intersection with the previous cycles.

Our conclusion is that π can be written as the product of these disjoint cycles:

$$\pi = (1\ \ 11\ \ 3\ \ 8\ \ 14\ \ 4\ \ 2\ \ 9\ \ 6)(7\ \ 12\ \ 13\ \ 10).$$

If you like you can explicitly include (5) as a third factor, but the usual convention is not to bother including elements that are fixed in expressions of this nature, if an element does not appear it is understood to be fixed.

Notes

1. The representation of π in "array" format can easily be read from its representation as a product of disjoint cycles. For example if you want to know the image of 8 under π, just look at the cycle where 8 appears - its image under π is the next element that appears after it in that cycle, 14 in this example. If your element is written at the end of a cycle, like 10 in this example, then its image under π is the number that is written in the first position of that same cycle (so $10 \to 7$ here). An element that does not appear in any of the cycles is fixed by the permutation.

2. The statement above says that π can be effected by first applying the cycle $(7\ \ 12\ \ 13\ \ 10)$ (which only moves the elements $7, 12, 13, 10$) and then applying the cycle $(1\ 11\ 3\ 8\ 14\ 4\ 2\ 9\ 6)$ (which only moves the elements $1, 11, 3, 8, 14, 4, 2, 9, 6$). Since these two cycles operate on disjoint sets of elements and do not interfere with each other, they commute with each other under composition - it does not matter which is written first in the expression for π as a product of the two of them. So we could equally well write

   $$\pi = (7\ \ 12\ \ 13\ \ 10)(1\ \ 11\ \ 3\ \ 8\ \ 14\ \ 4\ \ 2\ \ 9\ \ 6).$$

3. The expression for a permutation as a product of disjoint cycles is unique up to the order in which the cycles are written. This means that the same cycles must appear in any such expression for a given permutation, but they can be written in different orders.

   It might also be worth mentioning that a given cycle can be written in slightly different ways, since it doesn't matter which element is taken as the "starting point". For example $(7\ \ 12\ \ 13\ \ 10)$ and $(13\ \ 10\ \ 7\ \ 12)$ represent the same cycle.

Definition 2.4.3. *The expression of an element of S_n as a product of disjoint cycles partitions the set $\{1, 2, \ldots, n\}$ into disjoint* orbits*. In the above example there are three orbits:*

$$\{1, 2, 3, 4, 6, 8, 9, 11, 14\},\ \{5\},\ \{7, 10, 12, 13\}.$$

If two elements belong to the same orbit for a permutation π, it means that some power of π takes one of those elements to the other. Note that fixed points *do* count as orbits. So the identity element of S_n has n orbits each consisting of a single element. A permutation in S_n has just one orbit if it is a single cycle involving all n elements.

It is good idea to practise moving between the "array representation" and "disjoint cycle representation" of a permutation. There is another way of representing permutations that is sometimes

useful. We could think of the "simplest" type of non-identity permutation as being one that just swaps two elements and leaves the rest fixed. Such a permutation is called a transposition. The transposition that (for example) interchanges 1 and 2 and leaves all the other elements fixed is denoted, in typical cycle notation, as $(1\ 2)$.

Theorem 2.4.4. *Every element of S_n can be expressed as a product of transpositions.*

Rather than giving a formal general proof of Theorem 2.4.4, we will look at a way of expressing a given permutation as a product of transpositions. This contains all that would be required for a fully detailed proof, without having to worry about setting up cumbersome general notation.

Example 2.4.5. *In S_8 (for example), the cycle $(2\ 4\ 7\ 6\ 8)$ can be written as the product*

$$(2\ 8)(2\ 6)(2\ 7)(2\ 4)$$

of four transpositions.
To see this, just look at what happens to each element under the proposed composition of transpositions. Start with 2. We have:

$$2 \to 4.$$

Move on to 4:

$$4 \to 2 \to 7.$$

Then 7:

$$7 \to 2 \to 6.$$

Then 6:

$$6 \to 2 \to 8.$$

Finally 8:

$$8 \to 2.$$

So overall our composition of transpositions amounts to the cycle

$$2 \to 4 \to 7 \to 6 \to 8 \to 2,$$

as we wanted.

Note: The expression for a given cycle (or permutation) as a product of transpositions is *not unique*. For example we could write the 4-cycle above equally well as $(4\ 7\ 6\ 8\ 2)$, then using the same technique to write it as a product of transpositions would result in

$$(4\ 2)(4\ 8)(4\ 6)(4\ 7),$$

which does not involve the same transpositions as our example above, although it is the same permutation.

Example 2.4.6. *In S_{12}, write the element*

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 11 & 4 & 6 & 1 & 10 & 7 & 8 & 12 & 9 & 3 & 2 & 5 \end{pmatrix}$$

as a product of transpositions.

Solution: First write it as a product of disjoint cycles.

$$(1\ 11\ 2\ 4)(3\ 6\ 7\ 8\ 12\ 5\ 10).$$

Then as a product of transpositions:

$$(1\ 4)(1\ 2)(1\ 11)(3\ 10)(3\ 5)(3\ 12)(3\ 8)(3\ 7)(3\ 6).$$

This expression involves nine transpositions.

Exercise 2.4.7. *How many of the* $n!$ *elements of* S_n *are transpositions? How many are* 3*-cycles? (i.e. cycles of length 3, like* $(1\ 2\ 3)$*).*

The number of transpositions involved in an expression for a permutation as a product of transpositions is not uniquely determined either, since for example $(2\ 3)$ and $(1\ 3)(2\ 3)(1\ 2)$ are the same permutation (check this). However, it *is* true that no permutation can be written both as the product of an even number and an odd number of permutations. To prove this is not difficult but involves a bit of fussing. This is our next task.

Theorem 2.4.8. *A permutation in* S_n *cannot be expressed both as the product of an even number and an odd number of transpositions.*

Proof. Let $\pi \in S_n$, and suppose that π can be written as a product of s transpositions, i.e.

$$\pi = \tau_s \tau_{s-1} \dots \tau_2 \tau_1,$$

where each τ_i is a transposition. Let r be the number of orbits of π (i.e. the number of cycles in the expression for π as a product of disjoint cycles, including fixed points). Then r is fully determined by π and so is $n - r$ (this means that the numbers r and $n - r$ do not depend on any choice about how π is represented). We will show that the numbers s and $n - r$ are either both even or both odd.

We will do this by induction on s, the starting point being $s = 0$. If $s = 0$ then π is the identity permutation, $r = n$ and $n - r = 0$. So in this case s and $n - r$ are both zero, they are both even.

The case $s = 1$ is also manageable. If $s = 1$, then π is a single transposition, so it has one cycle of length 2 and $n - 2$ fixed points. In this case $r = n - 1$ and $n - r = 1$, so s and $n - r$ are both equal to 1, they are both odd.

Now suppose that s and $n-r$ have the same parity for all values of s up to $s = k$, and consider the case $s = k + 1$. This means

$$\pi = \tau_{k+1} \tau_k \dots \tau_2 \tau_1,$$

where each τ_i is a transposition. Let $\tau_{k+1} = (1\ 2)$ (there is no loss of generality here since we can relabel the elements that are being permuted if necessary), let π' be the element of S_n given by

$$\pi' = \tau_k \dots \tau_2 \tau_1,$$

and let r' be the number of orbits of π'. We will show that the number r of orbits of π differs from r' by 1.

Case 1: Suppose first that 1 and 2 belong to the same orbit in π', and write the cycle corresponding to this orbit as $(1\ a_2 \dots a_l\ 2\ a_{l+m} \dots a_m)$. Then we have (check this)

$$(1\ 2)(1\ a_1 \dots\ a_l\ 2\ a_{l+1} \dots a_m) = (1\ a_1 \dots a_1)(2\ a_{l+1} \dots a_m).$$

So the orbit of π' that contained the elements 1 and 2 is split into two separate orbits by the multiplication by τ_{k+1}. Other orbits of π' are unaffected since they do not involve 1 or 2. So in the case where 1 and 2 belong to the same orbit of π', we have $r = r' + 1$.

Case 2: Suppose that 1 and 2 belong to different orbits of π', and write the cycles corresponding to these orbits as

$$(1\ a_1 \dots a_l),\ \ (2\ b_1 \dots b_m)$$

where none of the a_i is equal to any of the b_j. Then (check that)

$$(1\ 2)(1\ a_1 \dots a_l)(2\ b_1 \dots b_m) = (1\ a_1 \dots a_l\ 2\ b_1 \dots b_m),$$

so the effect of the multiplication by $(1\ 2)$ is to combine these two orbits into one. As in Case 1 there is no effect on the other orbits of π'. So in the case where 1 and 2 belong to different orbits of π', we have $r = r' - 1$.

By our induction hypothesis, $n-r'$ has the same parity as k. The above argument above shows that $n - r$ differs from $n - r'$ by 1, and hence it must have the same parity as $k + 1$ which is the number of transpositions in π.

We have proved that the parity (oddness or evenness) of the number of transpositions in any expression for π as a product of transpositions is the same as the parity of $n - r$. In particular, for a given π, this number of transpositions is always even or always odd. □

Definition 2.4.9. *An element of S_n is called* even *if it can be written as the product of an even number of transpositions, and* odd *if it can be written as the product of an odd number of transpositions. Every element of S_n is either even or odd (not both).*

Note that the inverse of an even permutation is again even (it involves the same transpositions listed in the opposite order), and that the product of two even permutations is even. Moreover, the identity permutation is even, since it can be written as the "product of zero transpositions" or as the square of any transposition. Thus the set of *even permutations* of n objects is a subgroup of S_n. This is known as the *alternating group* of degree n and denoted by A_n. Directly counting the even permutations of a set of n elements is a more difficult task than counting *all* the permutations. However, by showing that the even permutations can be put in one-to-one correspondence with the odd permutations, we can show that exactly half of all the elements of S_n are even.

Theorem 2.4.10. *The order of the alternating group A_n is $\frac{n!}{2}$.*

Proof. Let the numbers of even and odd permutations in S_n be k_1 and k_2 respectively, and let τ denote the transposition $(1\ 2)$. For every even permutation π, we have a corresponding odd permutation $\pi\tau$. Thus there are at least as many odd permutations as even permutations, $k_1 \leqslant k_2$.

On the other hand, for every *odd* permutation σ we have the corresponding *even* permutation $\sigma\tau$. So there are at least as many even permutations as odd permutations, $k_2 \leqslant k_1$.

It follows that $k_1 = k_2$ and hence that the even permutations and odd permutations each account for half of all permutations. Thus

$$|A_n| = \frac{n!}{2}.$$

□