MATH1060 Introductory Linear Algebra

Cheatsheet

2022/23

This document collects together the important definitions and results presented throughout the lecture notes. The numbering of the sections will be consistent with that in the lecture notes.

Contents

1 General Systems of Linear Equations

1.1 Introduction to Linear Algebra?

Definition 1.1.1 The general system of m linear equations in n unknowns is

$$\begin{aligned}
a_{11}x_1 &+ a_{12}x_2 + \cdots + a_{1j}x_j + \cdots + a_{1n}x_n = b_1,\\
a_{21}x_1 &+ a_{22}x_2 + \cdots + a_{2j}x_j + \cdots + a_{2n}x_n = b_2,\\
&\vdots\\
a_{i1}x_1 &+ a_{i2}x_2 + \cdots + a_{ij}x_j + \cdots + a_{in}x_n = b_i,\\
&\vdots\\
a_{m1}x_1 &+ a_{m2}x_2 + \cdots + a_{mj}x_j + \cdots + a_{mn}x_n = b_m,
\end{aligned}$$

where a_{ij} and b_j are real numbers for all integers $1 \leq i \leq m$ and $1 \leq j \leq n$.

Definition A solution to the system in Definition 1.1.1 is an n-tuple $(x_1, ..., x_n) = (c_1, ..., c_n)$ of real numbers such that every one of the m equations is satisfied.

1.2 Different Solution Possibilities

Note: There are three possibilities which can happen in general for a system of equations:

- There are no solutions.
- There is one unique solution.
- There are infinitely-many solutions.

1.3 Introducing Matrices

Definition 1.3.1 An $m \times n$ matrix is an array of numbers in m rows and n columns, i.e.

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Remark It is important to note that the first number m is always the number of **rows**, and the second number n is always the number of **columns**. Furthermore, we use the shorthand notation $A = (a_{ij})$ to mean the matrix whose entries are a_{ij}.

Definition The coefficient matrix of a system of equations as in Definition 1.1.1 is the matrix $A = (a_{ij})$ whose entries are the coefficients of the x-variables. Then, the augmented matrix representing the system is the coefficient matrix with the column of constants $\mathbf{b}$:

$$(A \mid \mathbf{b}) = \left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1m} & b_1 \\ a_{21} & \cdots & a_{2m} & b_2 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_n \end{array}\right).$$

1.4 Reduction by Elementary Row Operations to Echelon Form

Definition 1.4.1 An elementary row operation on an $m \times n$ matrix is one of the following:

(i) Adding and subtracting one row to/from another, e.g. $R1 \mapsto R1 - 3R2$.
(ii) Multiplying a row by a constant, e.g. $R3 \mapsto 2R3$.
(iii) Swapping two rows, e.g. $R2 \leftrightarrow R4$.

Note: We write the symbol $\sim$ between matrices when we apply elementary row operations.

1.5 Row Echelon Form and Reduced Row Echelon Form

Definition 1.5.1 An $m \times n$ matrix is in row echelon form (REF) if these are true:

(i) Any all-zero rows are at the bottom of the matrix.
(ii) The first non-zero entry in each row is 1, called a leading one.
(iii) The leading one of a non-zero row is to the right of any leading one in rows above.

Definition An $m \times n$ matrix is in reduced row echelon form (RREF) if these are true:

(i) The matrix is in row echelon form (i.e. all conditions in Definition 1.5.1 are satisfied).
(ii) In each column containing a leading one, every other entry is zero.

1.6 An Algorithm for Reducing a Matrix to REF or RREF

Method – Gaussian Elimination: We can find a REF of a matrix A as follows:

(i) Find the first non-zero column.
   - If the first non-zero entry is in $R1$, move to Step (ii).
   - If the first non-zero entry is not in $R1$, swap the top row with the row containing the first non-zero entry.

(ii) Multiply $R1$ by a constant so that the non-zero entry becomes 1.
(iii) Clear below it by adding/subtracting multiples of $R1$ to/from the other rows.
(iv) Fix this row and repeat Steps (i)-(iii) with those below, until all rows are covered.

Note: Some mathematicians do **not** distinguish between *Gaussian* elimination and *Gauss-Jordan* elimination (e.g. in this module). However, be aware that some people use these names to mean different things: Gaussian for REF and Gauss-Jordan for RREF.

Method – Gauss-Jordan Elimination: We can find the RREF of a matrix A as follows:

(i) Transform A into a REF by applying the above method.
(ii) Starting at the rightmost column with a leading entry, clear above it.
(iii) Repeat Step (ii) by moving right-to-left through the leading entries.

Remark There isn't a unique row echelon form for a general matrix. However, the reduced row echelon form **is** unique. This is why we refer to them as "*a* REF" and "*the* RREF".

1.7 Summary of the Matrix Method of Solving Systems of Linear Equations

Method – Solving Systems of Linear Equations: We can solve a system of linear equations by applying the following procedure:

(i) Replace the equations by the augmented matrix $(A \mid \mathbf{b})$.
(ii) Transform the left part of the augmented matrix, A, into REF.
[**Note:** Optionally, we can continue on to transform A into RREF.]
(iii) Read the solutions, or lack thereof, directly from the (R)REF.

Note: Suppose we write our system of equations as an augmented matrix, in its RREF.

- If there are as many leading ones as there are columns, we have a **unique solution**.
- If there are less leading ones as there are columns, then the variables corresponding to the columns without said leading ones are free; this means they can be any real number. This means there are **infinitely-many solutions**.
- If there is a row with zeros to the left of the augmented column but with a non-zero in the augmented column, then there is **no solution**.

Remark We can do an identical process even if the coefficients of our system of linear equations themselves includes some variables, e.g. a system like $3x + ky = 4$ and $x - 2ky = -8$.

1.9 Common Mistakes

These are some common mistakes encountered when performing row operations:

- Doing too many row operations at once.
- Using the wrong row to put the matrix into RREF.
- Getting the sign wrong of the coefficient of free variables.

Note: The best way to avoid these mistakes is to take your time and check your answers.

2 Matrices and Matrix Algebras

2.1 Equality

Definition 2.1.1 Two matrices $A = (a_{ij})$ of size $m \times n$ and $B = (b_{ij})$ of size $r \times s$ are equal if $m = r$, $n = s$ and $a_{ij} = b_{ij}$ for all i and j.

2.2 Addition

Definition 2.2.1 Let $A = (a_{ij})$ and $B = (b_{ij})$ be matrices of the same size $m \times n$. Their sum is the $m \times n$ matrix $A + B = (a_{ij} + b_{ij})$, that is we add the entries component-wise.

Note: The addition of two matrices is not defined if they are of different sizes.

Lemma *Matrix addition is associative, that is for three matrices A, B, C of the same size,*

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

2.3 Scalar Multiplication

Definition 2.3.2 Let $A = (a_{ij})$ be an $m \times n$ matrix and $k \in \mathbb{R}$. The scalar multiplication of A by the number k is the matrix $kA = (ka_{ij})$, that is we multiply each entry by k.

2.4 Matrix Multiplication

Definition 2.4.2 Let $A = (a_{ik})$ be an $m \times n$ matrix and $B = (b_{kj})$ be an $n \times p$ matrix. The matrix multiplication of A and B is the matrix $AB = (c_{ij})$, where the entries are given by

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Remark So, c_{ij} is found by adding the multiplies of the i^{th} row of A and the j^{th} column of B.

Note: Notice in Definition 2.4.2 that matrix multiplication is only defined when the number of rows of the first matrix is equal to the number of columns of the second.

Remark 2.4.3 We use boldface letters to denote vectors when typed, and underlined letters to denote vectors when written by-hand, e.g. $\mathbf{x}$ and $\underline{x}$ are the same vector.

Note: It is helpful to think of a column vector with n entries as an $n \times 1$ matrix.

Proposition *Matrix multiplication is not commutative in general, that is for matrices A, B,*

$$AB \neq BA,$$

assuming that each of these matrix multiplications is well-defined.

Definition 2.4.5 An $m \times n$ matrix A is square if $m = n$, meaning it is of size $n \times n$.

Lemma *Let A and B be matrices such that $AB = BA$. Then, A and B are both square.*

Proof: Let A be an $m \times n$ matrix and B be an $r \times s$ matrix. Because AB is well-defined, we know that $n = r$. Similarly, because BA is well-defined, we know that $m = s$. However, AB is a matrix of size $m \times s$ and BA is a matrix of size $r \times n$. Therefore, the equality $AB = BA$ implies that these are the same size, and thus $m = n = r = s$. □

Definition 2.4.7 Let $A = (a_{ij})$ be an $m \times n$ matrix. The transpose is the matrix $A^T = (a_{ji})$, that is the $n \times m$ matrix obtained by interchanging the rows and columns.

Remark 2.4.9 Let $\mathbf{x}$ and $\mathbf{y}$ be vectors of the same size, say n. Recall that the dot product is the number $\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \cdots + x_n y_n$. We can re-write the dot product as this matrix multiplication:

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}.$$

Proposition 2.4.10 *For A an $m \times n$ matrix and B an $n \times p$ matrix, we have*

$$(AB)^T = B^T A^T.$$

Proof: Let $A = (a_{ik})$ and $B = (b_{kj})$. From Definition 2.4.2, we know that the ij^{th} entry of the product AB is $\sum_{k=1}^{n} a_{ik} b_{kj}$. Taking the transposition swaps rows and columns (i and j) to give

$$\sum_{k=1}^{n} a_{jk} b_{ki}.$$

Via similar reasoning, the ij^{th} entry of $B^T A^T$ is $\sum_{k=1}^{n} b'_{ik} a'_{kj}$, where b'_{ik} is the ik^{th} entry of B^T and a'_{kj} is the kj^{th} entry of A^T. From Definition 2.4.7, we conclude that $b'_{ik} = b_{ki}$ and $a'_{kj} = a_{jk}$. Substituting these into the formula for the ij^{th} entry of $B^T A^T$ produces the same sum as above:

$$\sum_{k=1}^{n} b_{ki} a_{jk} = \sum_{k=1}^{n} a_{jk} b_{ki}.$$ □

2.5 Identity Matrices and Inverses

Definition 2.5.1 The $n \times n$ identity matrix I_n is the matrix with ones on the main diagonal and zeros everywhere else, i.e. $I_n = (\delta_{ij})$ with δ_{ij} the so-called *Kronecker delta function*:

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}.$$

Lemma *Let A be an $m \times n$ matrix. Then, we have that*

$$AI_n = A = I_m A.$$

Definition 2.5.2 Let A and B be two $n \times n$ matrices. If $AB = I_n = BA$, then we say that A and B are multiplicative inverses of each other. We denote this by writing $B = A^{-1}$.

Remark 2.5.3 Strictly speaking, we must show that $AB = I_n$ **and** $BA = I_n$ in order to conclude that the two matrices are inverse to one another. However, it turns out that showing only one of these is enough to guarantee the other.

Definition 2.5.6 Let A be a square matrix.

(i) If A has a multiplicative inverse, we call it invertible or non-singular.

(ii) If A doesn't have a multiplicative inverse, we call it non-invertible or singular.

Theorem 2.5.9 *Let A be an $n \times n$ matrix. If A is invertible, then its inverse is unique.*

Proof: Assume that B and C are both inverses of A. By definition, we have that $AB = I_n = BA$ and that $AC = I_n = CA$. The goal is to show that $B = C$ (that is the two inverses are actually one in the same). Notice we have $AB = AC$ $(= I_n)$. If we multiply on the left by B, we obtain

$$BAB = BAC \qquad \Rightarrow \qquad I_n B = I_n C \qquad \Rightarrow \qquad B = C,$$

using the fact $BA = I_n$ and that multiplying by I_n changes nothing via the above lemma. □

Theorem 2.5.10 *Let A and B be invertible $n \times n$ matrices. Then, the matrix product AB is invertible, and has inverse $(AB)^{-1} = B^{-1}A^{-1}$.*

Proof: It suffices to check that $B^{-1}A^{-1}$ is the inverse of AB (and we check both equalities):

(i) $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n$.

(ii) $(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_nB = B^{-1}B = I_n$.

Consequently, the result follows from Definition 2.5.2. □

Corollary 2.5.11 *Let $A_1, A_2, ..., A_t$ be invertible $n \times n$ matrices. The product is such that*

$$(A_1 A_2 \cdots A_t)^{-1} = A_t^{-1} \cdots A_2^{-1} A_1^{-1}.$$

Proof: This is an induction argument following from Theorem 2.5.10. Indeed, the base case $t = 1$ is just Definition 2.5.2 and the less-trivial 'second base' case $t = 2$ is simply Theorem 2.5.10. Now, assume the result holds for $t = k$. We must prove that, under this hypothesis, the result holds for $t = k + 1$. Well,

$$\begin{aligned}(A_1 A_2 \cdots A_k A_{k+1})^{-1} &= \big((A_1 A_2 \cdots A_k) A_{k+1}\big)^{-1} \\ &= A_{k+1}^{-1} (A_1 A_2 \cdots A_k)^{-1} \\ &= A_{k+1}^{-1} A_k^{-1} \cdots A_2^{-1} A_1^{-1},\end{aligned}$$

using the previous theorem to get the second equality and using the induction hypothesis to get the third equality. Hence, by mathematical induction, the result holds for every $t \geq 1$. □

Notation Let A be a matrix. We use the following notation, where $k \geq 1$ is an integer:

- $A^k = AA \cdots A$, with A appearing k times in the product.
- $A^{-k} = A^{-1} A^{-1} \cdots A^{-1}$, with A^{-1} appearing k times in the product.

Note: If A is invertible, so too is A^k by Corollary 2.5.11. Therefore, we can write this:

$$(A^k)^{-1} = A^{-k} = (A^{-1})^k.$$

Lemma 2.5.12 *Let A be an invertible $n \times n$ matrix. Then, the transpose A^T is invertible, and has inverse $(A^T)^{-1} = (A^{-1})^T$.*

Proof: By Proposition 2.4.10, we know that $(AB)^T = B^T A^T$. Applying this result with $B = A^{-1}$ tells us that $(AA^{-1})^T = (A^{-1})^T A^T$, but we clearly see that the left-hand side of this is

$$(AA^{-1})^T = I_n.$$

Combining these immediately implies that the inverse of A^T is $(A^{-1})^T$, as required. □

2.7 Summary of the Method of Finding Matrix Inverses

Method – Finding the Inverse of a Matrix: To find the inverse of the $n \times n$ matrix A, if it exists, we form the augmented matrix $(A \mid I_n)$ and apply Gauss-Jordan elimination; this will get it into the form $(I_n \mid A^{-1})$, from which we can read-off the inverse matrix. In the case that we get a zero row on the left of the vertical line during this process, we conclude that A is **not** invertible.

Theorem 2.7.3 *If an $n \times n$ matrix A has an all-zero row/column, then it is non-invertible.*

Proof: Suppose $A = (a_{ij})$ has every entry in the m^{th} row zero, i.e. $a_{mj} = 0$ for all $1 \leq j \leq n$. For any $n \times n$ matrix $B = (b_{ij})$, we know that the mj^{th} entry in the product AB is the following:

$$\sum_{k=1}^{n} a_{mk} b_{kj} = 0,$$

because of the fact that $a_{mk} = 0$. Because this is true for any j, we know now that the entire m^{th} row of AB is zero. Consequently, AB can never equal the identity matrix I_n, a requirement for A to have an inverse B. A similar argument works if instead the m^{th} column of A is all-zero; the m^{th} column of BA will also be all-zero in this case. □

2.8 Elementary Matrices

Definition 2.8.1 An elementary matrix is a matrix obtained by applying **one** single elementary row operation to the identity matrix I_n.

Proposition *Let E be the elementary matrix corresponding to the elementary row operation $\mathcal{R}$. Applying $\mathcal{R}$ to some matrix A is equivalent to doing the matrix multiplication EA.*

Note: Note that the elementary matrix appears on the **left** in the above proposition; if we instead multiplied on the right, we would see that this actually corresponds to an elementary column operation instead of a row operation.

Remark 2.8.7 Every elementary matrix is invertible. In fact, the inverse matrix is the elementary matrix corresponding to doing the **opposite** row operation.

2.9 Why Our Method of Finding Matrix Inverses Works

Theorem 2.9.1 *Let A be an $n \times n$ matrix. If its reduced row echelon form is the identity matrix, then A is invertible. Furthermore, this implies that both A and A^{-1} are products of elementary matrices.*

Proof: We begin by forming the augmented matrix $(A \mid I_n)$. Since the RREF of A is the identity, we know that there is a finite sequence of row operations (corresponding to elementary matrices) taking the left part to the identity. Indeed, suppose it takes k row operations to do this. Then,

Start:	$(A \mid I_n)$,
Row Operation 1:	$(E_1 A \mid E_1 I_n)$,
Row Operation 2:	$(E_2 E_1 A \mid E_2 E_1 I_n)$,

$$\vdots \qquad\qquad\qquad \vdots$$

$$\text{Row Operation } k: \qquad (E_k \cdots E_2 E_1 A \mid E_k \cdots E_2 E_1 I_n),$$

where we assumed that $E_k \cdots E_2 E_1 A = I_n$. In other words, this tells us that $A = (E_k \cdots E_2 E_1)^{-1}$. Because each elementary matrix E_i is invertible, we know that their product is invertible by Corollary 2.5.11. Consequently, we see that

$$A = E_1^{-1} E_2^{-1} \cdots E_k^{-1}.$$

Ergo, we know that A is invertible by the same corollary and that its inverse is

$$A^{-1} = E_k \cdots E_2 E_1.$$ □

In fact, the converse to Theorem 2.9.1 is true and we will state and prove this below.

Theorem 2.9.2 *If A is an invertible $n \times n$ matrix, then its reduced row echelon form is I_n.*

Proof: Assume to the contrary $J \neq I_n$ is the RREF of A and suppose $E_1, E_2, ..., E_k$ is the sequence of elementary matrices corresponding to the row operations taking A to its RREF, i.e.

$$E_k \cdots E_2 E_1 A = J.$$

Because J is in RREF which is not the identity, and as it is an $n \times n$ matrix, we know that there are strictly less than n leading ones. As such, it contains an all-zero row. But by Theorem 2.7.3, this means J is non-invertible. However, the left-hand side of the equation above is a product of invertible matrices, so J is invertible; this is a contradiction. □

Note: If we start with an augmented matrix $(A \mid I_n)$ and, when doing row operations to find the RREF of the left part A, if we end up with a non-identity matrix, we now know from the contrapositive of Theorem 2.9.2 that A is non-invertible.

Theorem 2.9.4 *Let A be an $n \times n$ matrix and $\mathbf{b} \in \mathbb{R}^n$ be a column vector. Then, the linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution if and only if A is invertible.*

Proof: ($\Rightarrow$) Assume that $A\mathbf{x} = \mathbf{b}$ has a unique solution. It must be that the RREF of A is the identity matrix. If it was not, then the RREF would have strictly less than n leading ones, implying either infinitely-many or no solutions. Hence, A is invertible by Theorem 2.9.1.

($\Leftarrow$) Assume that A is invertible. Then, we can re-write the linear system $A\mathbf{x} = \mathbf{b}$ as follows: $A^{-1}A\mathbf{x} = A^{-1}\mathbf{b}$, which clearly tells us that a solution is $\mathbf{x} = A^{-1}\mathbf{b}$. Suppose the system has another solution $A\mathbf{y} = \mathbf{b}$. The same trick implies $\mathbf{y} = A^{-1}\mathbf{b} = \mathbf{x}$, so uniqueness is proven. □

Theorem 2.9.5 *Let A be a non-invertible $n \times n$ matrix. Then, the matrix product AB is non-invertible for* **any** *$n \times n$ matrix B.*

Proof: Let $E_1, E_2, ..., E_k$ be the sequence of elementary matrices corresponding to the row operations taking A to its RREF J, that is $E_k \cdots E_2 E_1 A = J$. Because A is non-invertible, we know from the contrapositive of Theorem 2.9.1 that $J \neq I_n$. In particular, J contains at least one all-zero row.

Now, the equation $AB\mathbf{x} = \mathbf{0}$ has the same set of solutions as the equation $E_1 AB\mathbf{x} = E_1\mathbf{0} = \mathbf{0}$, which in turn has the same set of solutions as the equation $E_2 E_1 AB\mathbf{x} = E_2\mathbf{0} = \mathbf{0}$, and so forth. We can conclude that this equation has the same set of solutions as

$$E_k \cdots E_2 E_1 AB\mathbf{x} = \mathbf{0}.$$

However, the left-hand side is precisely $JB\mathbf{x}$; because J has an all-zero row, it follows that JB has an all-zero row (and is therefore non-invertible) and thus there are infinitely-many solutions, by Theorem 2.9.4. Hence, $AB\mathbf{x} = \mathbf{0}$ has infinitely-many solutions, so AB is also non-invertible. □

Note: In practice, one can find the inverse and get the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$ (assuming the matrix is invertible). However, it is often quicker to do the row reduction method.

2.10 Extra Material

We discuss when $AB = AC$ implies $B = C$. Of course, if A is square and invertible, this holds rather straightforwardly. However, if A is non-invertible, this may not be true. Well, the next result tells us precisely when this cancellation is possible.

Theorem 2.10.2 *Let A be an $m \times n$ matrix and B, C be two $n \times p$ matrices. Then, $AB = AC$ implies $B = C$ if and only if $A\mathbf{x} = \mathbf{0}$ has a unique solution, namely $\mathbf{x} = \mathbf{0}$.*

Proof: Note first that $\mathbf{x} = \mathbf{0}$ is always a solution to the system $A\mathbf{x} = \mathbf{0}$, meaning that it is never inconsistent. Hence, there are either infinitely-many solutions or a unique solution.

($\Rightarrow$) Suppose that $AB = AC$ implies $B = C$. We take $B = \mathbf{x} \in \mathbb{R}^n$ a column vector ($n \times 1$ matrix) and $C = \mathbf{0} \in \mathbb{R}^n$ the zero column vector. Then, substituting says that $A\mathbf{x} = A\mathbf{0} = \mathbf{0}$ implies that $\mathbf{x} = \mathbf{0}$, as required.

($\Leftarrow$) Suppose that $A\mathbf{x} = \mathbf{0}$ has the unique solution $\mathbf{x} = \mathbf{0}$. Well, $AB = AC$ is equivalent to $AB - AC = 0_{m\times p}$, where $0_{m\times p}$ is the $m \times p$ matrix containing all-zeros. We can factorise the left-hand side: $A(B - C) = 0_{m\times p}$. The matrix $B - C$ is an $n \times p$ matrix with columns $\mathbf{v}_1, ..., \mathbf{v}_p$ (so the $\mathbf{v}_i \in \mathbb{R}^n$ are column vectors in their own right). Therefore, for each $1 \leq i \leq p$, we have

$$\begin{aligned} A(B - C) = 0_{m\times p} &\Rightarrow A\mathbf{v}_i = \mathbf{0} \\ &\Rightarrow \mathbf{v}_i = \mathbf{0} \\ &\Rightarrow B - C = 0_{n\times p} \\ &\Rightarrow B = C. \end{aligned}$$

□

3 Determinants

3.1 An Inductive Definition of the Determinant

Definition Let A be an $m \times n$ matrix. The ij^{th} minor A_{ij} is the $(m-1) \times (n-1)$ matrix we obtain by deleting the i^{th} row and the j^{th} column from A.

Definition 3.1.1 Let $A = (a_{ij})$ be an $n \times n$ matrix. The determinant is the number

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

where we assume that the determinant of a 1×1 matrix is simply its entry.

Remark We just need to pick a row and move along it, computing the minors at each step. We then multiply each minor by the number in that position of the matrix and then do an alternating sum ("plus, minus, plus, minus, etc." **or** " minus, plus, minus, plus, etc."). In fact, the formula above is the expansion along the first row, but really we could do it along **any** row, or indeed down **any** column.

Note: Be careful with the definition of a minor if you read elsewhere; some authors use the minor A_{ij} to mean the **determinant** of a minor, not just the sub-matrix itself.

Notation 3.1.2 We can write the determinant of a matrix A either as $\det(A)$ or as $|A|$. Thus,

$$\det \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}.$$

Note: Geometrically, the determinant of the matrix A tells us the volume of the parallelepiped (the generalisation of a parallelogram to any number of dimensions) whose sides are the vectors given by the rows of A.

3.2 Evaluating Determinants

Definition 3.2.1 Let $A = (a_{ij})$ be an $n \times n$ matrix.

(i) The diagonal entries a_{ii} form the so-called main diagonal.

(ii) A is upper triangular if all entries **below** the main diagonal are zero.

(iii) A is lower triangular if all entries **above** the main diagonal are zero.

(iv) A is triangular if it is either upper triangular or lower triangular.

(v) A is diagonal if all entries **not on** the main diagonal are zero.

Lemma *We have the following properties of the transpose:*

(i) *The transpose of an upper triangular matrix is lower triangular.*

(ii) *The transpose of a lower triangular matrix is upper triangular.*

(iii) *The transpose of a diagonal matrix is diagonal.*

Theorem 3.2.5 (Properties of Determinants) *Let A be an $n \times n$ matrix.*

(i) *Applying the row operation $R_i \leftrightarrow R_j$ changes the determinant to $-\det(A)$.*

(ii) *Applying the row operation $R_i \mapsto kR_i$ changes the determinant to $k\det(A)$.*

(iii) *Applying the row operation $R_i \mapsto R_i + kR_j$ does* **not** *change the determinant.*

(iv) *If A is triangular, the determinant is the product of the main diagonal entries, i.e.*

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

Note: Using (ii) in Theorem 3.2.5, we see that for A an $n \times n$ matrix and a number $k \in \mathbb{R}$,

$$\det(kA) = k^n \det(A).$$

3.3 An Algorithm for Computing Determinants

Method – Finding the Determinant: Let A be an $n \times n$ matrix.

(i) If A is triangular, it is obvious. Otherwise, go to Step (ii).

(ii) If A is not triangular, perform a sequence of row operations to transform it to an triangular matrix, which we now call A'.

(iii) Determine $\det(A')$ by using Theorem 3.2.5(iv).

(iv) Compute the changes made to the determinant of A by performing the sequence of row operations in Step (ii) by using Theorem 3.2.5(i)–(iii).

(v) Finally, we can write $\det(A)$ by **dividing** $\det(A')$ by the number we find in Step (iii).

Remark 3.3.2 If, during the above process, it is clear that the triangular matrix we obtain will have a zero on the main diagonal, we know immediately that $\det(A) = 0$.

Theorem 3.3.5 *Let A be an $n \times n$ matrix. If $\det(A) \neq 0$, then A is invertible.*

Proof: If $\det(A) \neq 0$, then the (upper) triangular matrix formed from A during the above method has no zeros on the main diagonal. Hence, a REF of A has leading ones down the main diagonal, so its RREF will be the identity matrix. We know from Theorem 2.9.1 that this means A is invertible. □

Note: In fact, the converse to Theorem 3.3.5 is also true, and we will prove it a bit later.

3.4 Proofs of Some of the Properties of Determinants

This section is used to prove Theorem 3.2.5(ii) and (iv). The proofs are non-examinable and therefore are omitted, but there are some useful corollaries that should be known.

Corollary 3.4.2 *If a matrix A has an all-zero row, then* $\det(A) = 0$.

Proof: If A has an all-zero row, then we can take zero as a common factor from that row. Hence, Theorem 3.2.5(ii) implies that $\det(A) = 0\det(A) = 0$. □

Corollary 3.4.4 *The determinant of the identity matrix* $\det(I_n) = 1$.

Proof: The identity matrix is diagonal. In particular, it is triangular. Hence, Theorem 3.2.5(iv) applies, but each entry on the diagonal is one, so $\det(I_n) = 1^n = 1$. □

3.5 Extra Material

This section is used to prove Theorem 3.2.5(i) and (iii). The proofs are non-examinable and therefore are omitted, but is a useful corollary that should be known.

Corollary 3.5.3 *If a matrix A has two identical rows, then* $\det(A) = 0$.

Proof: If we swap the two identical rows, A remains unchanged. Now, Theorem 3.2.5(iii) implies that $\det(A) = -\det(A)$, and the only way this equation is satisfied is if $\det(A) = 0$. □

3.6 The Determinant is Multiplicative

Theorem 3.6.1 *For any matrix A and E an elementary matrix,* $\det(EA) = \det(E)\det(A)$.

Proof: Recall that EA is the matrix obtained from A by applying the row operation encoded by E. Therefore, the statement is proven if we simply consider the three different types of row operation that E can describe. Indeed, this is a case-by-case proof.

(i) Suppose that E corresponds to $R_i \leftrightarrow R_j$. Then, we can conclude that

$$\begin{aligned} \det(E) &= -\det(I), && \text{by Theorem 3.2.5(i)}, \\ &= -1, && \text{by Corollary 3.4.4.} \end{aligned}$$

Theorem 3.2.5(i) also implies $\det(EA) = -\det(A)$, and thus $\det(EA) = \det(E)\det(A)$.

(ii) Suppose that E corresponds to $R_i \mapsto kR_i$. Then, we can conclude that

$$\begin{aligned} \det(E) &= k\det(I), && \text{by Theorem 3.2.5(ii)}, \\ &= k, && \text{by Corollary 3.4.4.} \end{aligned}$$

Theorem 3.2.5(ii) also implies $\det(EA) = k\det(A)$, and thus $\det(EA) = \det(E)\det(A)$.

(iii) Suppose that E corresponds to $R_i \mapsto R_i + kR_j$. Then, we can conclude that

$$\begin{aligned} \det(E) &= \det(I), && \text{by Theorem 3.2.5(iii)}, \\ &= 1, && \text{by Corollary 3.4.4.} \end{aligned}$$

Theorem 3.2.5(iii) also implies $\det(EA) = \det(A)$, and thus $\det(EA) = \det(E)\det(A)$. ☐

Theorem 3.6.2 *Let A be an $n \times n$ matrix. Then, A is invertible if and only if* $\det(A) \neq 0$.

Proof: ($\Rightarrow$) Suppose A is invertible. Then, we know there is a sequence of elementary matrices taking it to its RREF, namely $E_k \cdots E_2 E_1 A = I_n$ by Theorem 2.9.2. We appeal to Corollary 3.4.4 to conclude that $\det(E_k \cdots E_2 E_1 A) = \det(I_n) = 1$. We then use Theorem 3.6.1 repeatedly to conclude that $\det(E_k) \cdots \det(E_2)\det(E_1)\det(A) = 1$. Since both sides are real numbers, we must have that $\det(A) \neq 0$.

($\Leftarrow$) This is Theorem 3.3.5. ☐

Theorem 3.6.3 *Let A and B be $n \times n$ matrices. Then,* $\det(AB) = \det(A)\det(B)$.

Proof: If A is non-invertible, it follows that AB is non-invertible by Theorem 2.9.5. Hence, by Theorem 3.6.2, we know that $\det(A) = 0$ and $\det(AB) = 0$, so the result holds true. On the other hand, if A is invertible, we can write it as a product of elementary matrices $A = E_1 E_2 \cdots E_k$. Thus, we see can repeatedly apply Theorem 3.6.1 to see that

$$\begin{aligned} \det(AB) &= \det(E_1 E_2 \cdots E_k B) \\ &= \det(E_1)\det(E_2 \cdots E_k B) \\ &\vdots \\ &= \det(E_1)\det(E_2)\cdots\det(E_k)\det(B). \end{aligned}$$

But we can also see that $\det(A) = \det(E_1 E_2 \cdots E_k) = \det(E_1)\det(E_2)\cdots\det(E_k)$. Substituting this into the final line above indeed shows that $\det(AB) = \det(A)\det(B)$. ☐

Corollary 3.6.4 *Let A be an invertible $n \times n$ matrix. Then,* $\det(A^{-1}) = 1/\det(A)$.

Proof: Since $AA^{-1} = I_n$, taking determinants tells us that $\det(AA^{-1}) = 1$, by Corollary 3.4.4. Then, by Theorem 3.6.3, this becomes $\det(A)\det(A^{-1}) = 1$, and the result follows. ☐

3.7 Column Operations and Determinants

Theorem 3.7.1 *Let A be an $n \times n$ matrix. Then,* $\det(A^T) = \det(A)$.

Proof: By Lemma 2.5.12, the inverse of the transpose is the transpose of the inverse. Hence, A^T is invertible if and only if A is invertible. In particular, if A is non-invertible, $\det(A) = 0 = \det(A^T)$.

Suppose now that A is invertible. As usual, we can write it as a product of elementary matrices: $A = E_1 E_2 \cdots E_k$. By repeatedly applying Proposition 2.4.10, we see that

$$A^T = (E_1 E_2 \cdots E_k)^T = E_k^T \cdots E_2^T E_1^T.$$

Thus, by repeatedly applying Theorem 3.6.3, we can conclude that

$$\det(A^T) = \det(E_k^T) \cdots \det(E_2^T) \det(E_1^T).$$

But also, the same theorem tells us that $\det(A) = \det(E_1)\det(E_2) \cdots \det(E_k)$. Therefore, it is sufficient to prove that $\det(E^T) = \det(E)$ for any elementary matrix E. Indeed, if E corresponds to either $R_i \leftrightarrow R_j$ or $R_i \mapsto kR_i$, then $E^T = E$ and this result is trivial. Otherwise, if E corresponds to $R_i \mapsto R_i + kR_j$, then both E and E^T are triangular matrices with all-ones on the main diagonal. Thus, Theorem 3.2.5(iv) tells us that $\det(E^T) = 1 = \det(E)$. □

Note: This result now allows us to translate between row and column operations, since taking the transpose interchanges rows and columns. Moreover, it justifies why we talk about *triangular* matrices when discussing determinants (making no distinction between upper and lower): the transpose doesn't affect the determinant and taking the transpose simply interchanges upper for lower (and lower for upper).

Corollary 3.7.2 *Let A be an $n \times n$ matrix. Applying the column operation $C_i \mapsto kC_i$ changes the determinant to $k\det(A)$.*

Proof: Suppose B is the matrix obtained from A via the column operation $C_i \mapsto kC_i$. Then, B^T will be the matrix obtained from A^T via the corresponding row operation $R_i \mapsto kR_i$. Hence, we see that $\det(B) = \det(B^T) = k\det(A^T) = k\det(A)$. □

Corollary 3.7.3 *If a matrix A has an all-zero column, then* $\det(A) = 0$.

Proof: If A has an all-zero column, then A^T has an all-zero row. So, $\det(A) = \det(A^T) = 0$ by Corollary 3.4.2. □

Corollary 3.7.4 *Let A be an $n \times n$ matrix. Applying the column operation $C_i \leftrightarrow C_j$ changes the determinant to $-\det(A)$.*

Proof: Suppose B is the matrix obtained from A via the column operation $C_i \leftrightarrow C_j$. Then, B^T will be the matrix obtained from A^T via the corresponding row operation $R_i \leftrightarrow R_j$. Hence, we see that $\det(B) = \det(B^T) = -\det(A^T) = -\det(A)$. □

Corollary 3.7.5 *If a matrix A has two identical columns, then* $\det(A) = 0$.

Proof: If A has two identical columns, then A^T has two identical rows. So, $\det(A) = \det(A^T) = 0$ by Corollary 3.5.3. □

Corollary 3.7.6 *Let A be an $n \times n$ matrix. Applying the column operation $C_i \mapsto C_i + kC_j$ does* **not** *change the determinant.*

Proof: Suppose B is the matrix obtained from A via the column operation $C_i \mapsto C_i + kC_j$. Then, B^T will be the matrix obtained from A^T via the corresponding row operation $R_i \mapsto R_i + kR_j$. Hence, we see that $\det(B) = \det(B^T) = \det(A^T) = \det(A)$. □

3.8 General Inductive Formula for the Determinant

When we introduced the determinant in Definition 3.1.1, the sum was specific to expanding minors along the first row (hence why A_{1j} has a fixed number one). However, we can be more general and this still gives us the same notion of determinant as we have in our definition.

Proposition 3.8.1 *Let A be an $n \times n$ matrix. Then, the determinant satisfies the following:*

$$\det(A) = \sum_{j=1}^{n} (-1)^{\ell+j} a_{\ell j} \det(A_{\ell j}), \qquad \text{expanding along the } \ell^{th} \text{ row},$$

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+\ell} a_{i\ell} \det(A_{i\ell}), \qquad \text{expanding down the } \ell^{th} \text{ column}.$$

Sketch of Proof: (**Non-examinable**) The idea is to use the formula from Definition 3.1.1 in order to express the determinant of B, the matrix obtained from A via the row operation $R_1 \leftrightarrow R_\ell$. We then transform the minor B_{1j} into the minor $A_{\ell j}$ via a sequence of row operations, just swaps in fact. This gives us the first formula.

The second formula is a consequence of the first; this boils down to taking the transpose and using the fact that this will not affect the determinant (Theorem 3.7.1). □

4 Linear Transformations on $\mathbb{R}^n$ and Matrices

4.1 Real Linear Transformations

Recall that the Cartesian plane $\mathbb{R}^2$ is given by pairs of real numbers:

$$\mathbb{R}^2 = \{(x_1, x_2) : x_1, x_2 \in \mathbb{R}\}.$$

Similarly, the Euclidean space $\mathbb{R}^3$ is given by triples of real numbers:

$$\mathbb{R}^3 = \{(x_1, x_2, x_3) : x_1, x_2, x_3 \in \mathbb{R}\}.$$

In general, we have an n-dimensional Euclidean coordinate system as follows:

$$\mathbb{R}^n = \{(x_1, x_2, ..., x_n) : x_1, x_2, ..., x_n \in \mathbb{R}\}.$$

Note: If we take the set $\mathbb{R}^n$ and define upon it an addition rule (i.e. how to add two elements) and a scalar multiplication rule (i.e. how to multiply an element by a real number), then what we get is a real vector space. Indeed, here is an example of each rule:

(i) $(x_1, x_2, ..., x_n) + (y_1, y_2, ..., y_n) \coloneqq (x_1 + y_1, x_2 + y_2, ..., x_n y_n)$.

(ii) $\lambda(x_1, x_2, ..., x_n) \coloneqq (\lambda x_1, \lambda x_2, ..., \lambda x_n)$.

Definition 4.1.1 A real linear transformation is a map $T : \mathbb{R}^n \to \mathbb{R}^m$ satisfying the following:

(i) $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. (**Additivity**)

(ii) $T(\lambda \mathbf{x}) = \lambda T(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. (**Homogeneity**)

Remark Geometrically, a linear transformation takes straight lines to straight lines, and parallelograms to parallelograms. Recall that matrices had something to do with parallelograms, so is there a link here...? Spoiler alert: yes.

Lemma *A linear map $T : \mathbb{R}^2 \to \mathbb{R}^m$ is completely determined by its values on the vectors*

$$\mathbf{e}_1 = (1, 0) \qquad \textit{and} \qquad \mathbf{e}_2 = (0, 1).$$

Proof: First, note that any vector $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ can be written as a linear combination of the vectors in the result, namely $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$. Consequently, we see that

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2) \\ &= T(x_1\mathbf{e}_1) + T(x_2\mathbf{e}_2) \\ &= x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2), \end{aligned}$$

where we use additivity for the second equality and homogeneity for the third inequality. □

Note: This result actually generalises to $T : \mathbb{R}^n \to \mathbb{R}^m$ on any so-called *basis* of $\mathbb{R}^n$.

If we know $T(1,0) = (a_{11}, a_{21})$ and $T(0,1) = (a_{12}, a_{22})$, our linear transformation is given by

$$T(\mathbf{x}) = x_1(a_{11}, a_{21}) + x_2(a_{12}, a_{22}) = (a_{11}x_1 + a_{12}x_2, a_{21}x_1 + a_{22}x_2).$$

Writing this up using column vectors (instead of row vectors) will show us that

$$T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Definition The vector $\mathbf{e}_i = (0, ..., 0, 1, 0, ..., 0) \in \mathbb{R}^n$ with one in the i^{th} place and zeros elsewhere is called a standard basis vector for $\mathbb{R}^n$. This concept is discussed further later.

Proposition 4.1.5 *If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then $T(\mathbf{x}) = A\mathbf{x}$, where A is an $m \times n$ matrix whose columns are the vectors $T(\mathbf{e}_i)$. We then refer to the matrix A as the* matrix of T with respect to the basis $\mathbf{e}_1, ..., \mathbf{e}_n$.

Sketch of Proof: Follow a near-identical argument for what we did in the previous lemma and what we did immediately above for $T(1,0)$ and $T(0,1)$. □

Proposition 4.1.6 *If $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by $T(\mathbf{x}) = A\mathbf{x}$ for an $m \times n$ matrix A, then T is a linear transformation.*

Proof: We need to check the linearity conditions in Definition 4.1.1 for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$:

(i) $T(\mathbf{x} + \mathbf{y}) = A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = T(\mathbf{x}) + T(\mathbf{y})$, so additivity is satisfied.

(ii) $T(\lambda\mathbf{x}) = A(\lambda\mathbf{x}) = \lambda A\mathbf{X} = \lambda T(\mathbf{X})$, so homogeneity is satisfied. □

Note: In words, all linear transformations can be expressed by matrices (Proposition 4.1.5) and all matrices give rise to linear transformations (Proposition 4.1.6).

4.2 Types of Linear Transformation in $\mathbb{R}^2$

Definition We define the following linear transformations $T : \mathbb{R}^2 \to \mathbb{R}^2$ where $T(\mathbf{x}) = A\mathbf{x}$:

(i) T is a rotation of angle θ anti-clockwise about the origin if

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

(ii) T is a reflection in the line through the origin at angle θ with the x-axis if

$$A = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}.$$

(iii) T is a projection if $A^2 = A$, for example

$$A = \begin{pmatrix} 0 & 0 \\ 7 & 1 \end{pmatrix}.$$

(iv) T is a shear of scale factor k in the x-direction if

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

(v) T is a shear of scale factor v in the y-direction if

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

(vi) T is a scaling of scale factors k in the x-direction and v in the y-direction if

$$A = \begin{pmatrix} k & 0 \\ 0 & v \end{pmatrix}.$$

Theorem *Every linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a combination of those listed above.*

4.3 Composing Linear Transformations

Proposition 4.3.1 *Let $T_1 : \mathbb{R}^p \to \mathbb{R}^n$ and $T_2 : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations. Then, their composition $T_2 \circ T_1 : \mathbb{R}^p \to \mathbb{R}^m$ is a linear transformation. Also, if A_1 is the $n \times p$ matrix for T_1 and A_2 is the $m \times n$ matrix for T_2, then $A_2 A_1$ is the $m \times p$ matrix for $T_2 \circ T_1$.*

Proof: The first part is a simple case of checking linearity, for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ and $\lambda \in \mathbb{R}$.

$$\begin{aligned}(T_2 \circ T_1)(\mathbf{x}+\mathbf{y}) &= T_2(T_1(\mathbf{x}+\mathbf{y})) \\ &= T_2(T_1(\mathbf{x}) + T_1(\mathbf{y})) \\ &= T_2(T_1(\mathbf{x})) + T_2(T_1(\mathbf{y})) \\ &= (T_2 \circ T_1)(\mathbf{x}) + (T_2 \circ T_1)(\mathbf{y}),\end{aligned}$$

and

$$\begin{aligned}(T_2 \circ T_1)(\lambda\mathbf{x}) &= T_2(T_1(\lambda\mathbf{x})) \\ &= T_2(\lambda T_1(\mathbf{x})) \\ &= \lambda T_2(T_1(\mathbf{x})) \\ &= \lambda(T_2 \circ T_1)(\mathbf{x}).\end{aligned}$$

Hence, the composition is linear. Finally, assume that $T_1(\mathbf{x}) = A_1\mathbf{x}$ and $T_2(\mathbf{x}) = A_2\mathbf{x}$. Then,

$$\begin{aligned}(T_2 \circ T_1)(\mathbf{x}) &= T_2(T_1(\mathbf{x})) \\ &= T_2(A_1\mathbf{x}) \\ &= A_2(A_1\mathbf{x}) \\ &= (A_2A_1)\mathbf{x}.\end{aligned}$$

□

4.4 Eigenvalues and Eigenvectors

Definition 4.4.1 Let A be an $n \times n$ matrix. If $A\mathbf{v} = \lambda\mathbf{v}$ for some **non-zero** vector $\mathbf{v} \in \mathbb{R}^n$ and a scalar $\lambda \in \mathbb{R}$, then we call $\mathbf{v}$ an eigenvector of A with corresponding eigenvalue λ.

Remark 4.4.2 The prefix *eigen* is German and translates to "own" or "particular to".

Note: Geometrically, the idea is that the linear transformation defined by the matrix A has fixed a line in $\mathbb{R}^n$. Indeed, $A\mathbf{v} = \lambda\mathbf{v}$ tells us that, when we apply the transformation to $\mathbf{v}$, it is just a scaling of $\mathbf{v}$. This works for the **line** through $\mathbf{v}$ (i.e. $t\mathbf{v}$ for $t \in \mathbb{R}$) because

$$A(t\mathbf{v}) = t(A\mathbf{v}) = t\lambda\mathbf{v} = \lambda(t\mathbf{v}).$$

Hence, applying A to the line just scales it up by a factor of λ.

4.5 How to Find Eigenvalues and Eigenvectors

Definition 4.5.1 Let A be an $n \times n$ matrix. The characteristic polynomial is the degree n polynomial in one variable λ defined as $C_A(t) = \det(A - tI_n)$.

Lemma *The roots of $C_A(t)$ are precisely the eigenvalues of A.*

Proof: The defining equation for an eigenvalue is $A\mathbf{v} = \lambda\mathbf{v}$, that is $A\mathbf{v} - \lambda\mathbf{v} = (A - \lambda I_n)\mathbf{v} = \mathbf{0}$. Because $A - \lambda I_n$ is a square matrix, the equation has non-trivial solutions for $\mathbf{v}$ if and only if $\det(A - \lambda I_n) = 0$, by Theorems 2.9.4 and 3.6.2. But this is precisely the condition $C_A(\lambda) = 0$. □

Method – Finding the Eigenvectors of a Matrix: Let A be an $n \times n$ matrix.

(i) Find the eigenvalues $\lambda_1, ..., \lambda_k$ of A by solving $\det(A - \lambda I_n) = 0$.
(ii) Choose one of the eigenvalues, say λ_1.
(iii) Solve the system $(A - \lambda_1 I_n)\mathbf{v} = \mathbf{0}$ to get a family of solutions for $\mathbf{v}$.
(iv) Repeat this for each of the other eigenvalues $\lambda_2, ..., \lambda_k$ in Step (i).

Note: Surprisingly, the eigenvalues of a triangular matrix are just the diagonal entries!

4.6 Determinants and Eigenvalues

Lemma 4.6.1 *Let A be an $n \times n$ matrix with* **complex** *eigenvalues $\lambda_1, ..., \lambda_n \in \mathbb{C}$ counted with multiplicities (i.e. some may be listed more than once). Then, $\det(A) = \lambda_1\lambda_2 \cdots \lambda_n$. Moreover, $\det(A)$ is the constant term in the characteristic polynomial $C_A(t)$.*

Proof: The only reason we work over $\mathbb{C}$ instead of $\mathbb{R}$ is that we can use the Fundamental Theorem of Algebra to factorise the characteristic polynomial as

$$C_A(t) = (\lambda_1 - t)(\lambda_2 - t) \cdots (\lambda_n - t).$$

We can find the constant term by substituting $t = 0$ into the characteristic polynomial, that is $C_A(0) = \det(A - 0I_n) = \det(A)$ from Definition 4.5.1. On the other hand, if we substitute $t = 0$ into the above factorised form, we obtain precisely that which we want:

$$\det(A) = C_A(0) = \lambda_1\lambda_2 \cdots \lambda_n.$$ □

Corollary 4.6.2 *A matrix A is invertible if and only if all its eigenvalues are non-zero.*

4.7 Polynomial Equations and Eigenvectors

Lemma 4.7.1 *Let A be an $n \times n$ matrix with eigenvalue λ paired with the eigenvector $\mathbf{v}$. Then, $A^m\mathbf{v} = \lambda^m\mathbf{v}$ for every integer $m \geq 1$.*

Proof: The base case $m = 1$ is trivial. We next assume the result is true for $m = k$, that is $A^k\mathbf{v} = \lambda^k\mathbf{v}$. We must prove it holds true for $m = k + 1$ under this hypothesis. To that end,

$$\begin{aligned} A^{k+1}\mathbf{v} &= A(A^k\mathbf{v}) \\ &= A(\lambda^k\mathbf{v}) \end{aligned}$$

$$
\begin{aligned}
&= \lambda^k (A\mathbf{v}) \\
&= \lambda^k (\lambda\mathbf{v}) \\
&= \lambda^{k+1}\mathbf{v},
\end{aligned}
$$

where we use the inductive hypothesis for the second equality and the fact that $A\mathbf{v} = \lambda\mathbf{v}$ fir the fourth equality. By the principal of mathematical induction, the result is true. □

5 Real Vector Spaces and Subspaces

5.1 Real Vector Spaces

Definition A vector space is a set V with binary operations $+$ and $\cdot$ called vector addition and scalar multiplication, respectively, satisfying these for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $\lambda, \mu \in \mathbb{R}$:

(i) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$. (**Associativity**)
(ii) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$. (**Commutativity**)
(iii) There exists $\mathbf{0} \in V$ such that $\mathbf{v} + 0 = \mathbf{v}$ for all $\mathbf{v} \in V$. (**Additive Identity**)
(iv) For all $\mathbf{v} \in V$, there exists $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$. (**Additive Inverses**)
(v) $\lambda(\mu\mathbf{v}) = (\lambda\mu)\mathbf{v}$. (**Compatability**)
(vi) $1\mathbf{v} = \mathbf{v}$. (**Scalar Identity**)
(vii) $\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v}$. (**Distributivity**)
(viii) $(\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}$. (**Distributivity**)

Recall $\mathbb{R}^n$ is a real vector space with defined addition and scalar multiplication operations.

Remark 5.1.1 Because our vector space operations for $\mathbb{R}^n$ are component-wise, there is no practical difference (here) as to whether we use column vectors or row vectors. However, when we talk about linear transformations and matrix multiplication, we will always assume that the elements of $\mathbb{R}^n$ are **column** vectors.

5.2 Subspaces

Definition 5.2.1 Let V be a vector space. We call $U \subseteq V$ a vector subspace of V if the following are satisfied:

(i) $U \neq \emptyset$.
(ii) For all $\mathbf{u}, \mathbf{v} \in U$, we have $\mathbf{u} + \mathbf{v} \in U$. (**Closed under Vector Addition**)
(iii) For all $\mathbf{u} \in U$ and $\lambda \in \mathbb{R}$, we have $\lambda\mathbf{u} \in U$. (**Closed under Scalar Multiplication**)

Note: Hence, $U \subseteq V$ is a vector space in its own right with operations inherited from V.

Lemma 5.2.3 *If U is a subspace of V, then $\mathbf{0} \in U$.*

Proof: As $U \neq \emptyset$, there exists $\mathbf{u} \in U$. Let $\lambda = 0$ in Definition 5.2.1(iii), so $0\mathbf{u} = \mathbf{0} \in U$. □

Method – Proving a Subset is a Subspace: Let V be a vector space and $U \subseteq V$ a subset.

(i) Show that $\mathbf{0} \in U$ (this is sufficient for non-emptiness).
(ii) Show that U is closed under vector addition.
(iii) Show that U is closed under scalar multiplication.

Proposition *Let A be an $n \times n$ matrix. The solution set to $A\mathbf{x} = \mathbf{0}$ is a subspace of $\mathbb{R}^n$.*

Proof: We proceed with showing the required properties.

(i) Clearly, $A\mathbf{0} = \mathbf{0}$, so $\mathbf{0}$ lives inside the set of solutions and thus the set is non-empty.

(ii) Let $\mathbf{x}$ and $\mathbf{y}$ be solutions to the equation, meaning $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{y} = \mathbf{0}$. Then, we see that $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \mathbf{0} + \mathbf{0} = \mathbf{0}$, so $\mathbf{x} + \mathbf{u}$ lives inside the set of solutions and thus the set is closed under vector addition.

(iii) Let $\mathbf{x}$ be a solution to the equation, meaning $A\mathbf{x} = \mathbf{0}$, and take some scalar $\lambda \in \mathbb{R}$. Then, $A(\lambda\mathbf{x}) = \lambda(A\mathbf{x}) = \lambda\mathbf{0} = \mathbf{0}$, so $\lambda\mathbf{x}$ lies inside the set of solutions and thus the set is closed under scalar multiplication. □

Note: The trivial subspace of V is $\{\mathbf{0}\}$ and the non-proper subspace of V is V itself.

Remark We can describe some subspaces geometrically.

- The only subspaces of $\mathbb{R}$ are the trivial and non-proper subspaces.
- The non-trivial proper subspaces of $\mathbb{R}^2$ are lines through the origin.
- The non-trivial proper subspaces of $\mathbb{R}^3$ are planes and lines through the origin.

5.3 Spans and Linear Combinations

Definition 5.3.1 Let $\mathbf{v}_1, ..., \mathbf{v}_k \in V$ be a collection of vectors. A linear combination of them is a vector of the form $\lambda_1\mathbf{v}_1 + \cdots + \lambda_k\mathbf{v}_k$, where $\lambda_1, ..., \lambda_k \in \mathbb{R}$ are scalars.

Definition 5.3.3 The set of all vectors in a vector space V that are linear combinations of $\mathbf{v}_1, ..., \mathbf{v}_k \in V$ is called the (linaer) span of $\mathbf{v}_1, ..., \mathbf{v}_k$ and is denoted $\text{span}\{\mathbf{v}_1, ..., \mathbf{v}_k\}$.

Note: If $U \subseteq V$ such that $U = \text{span}\{\mathbf{v}_1, ..., \mathbf{v}_k\}$, we say U is spanned by $\mathbf{v}_1, ..., \mathbf{v}_k$.

Remark The vector $(a, b, c) \in \mathbb{R}^3$ can be written as $a(1, 0, 0) + b(0, 1, 0) + c(0, 0, 1)$. Therefore, every vector in $\mathbb{R}^3$ lives in the set $\text{span}\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$. Moreover, this linear span is actually **equal** to the whole of $\mathbb{R}^3$. This generalises to $\mathbb{R}^n = \text{span}\{\mathbf{e}_1, ..., \mathbf{e}_n\}$.

Method – Vectors Belong to a Span: We want to see if $\mathbf{x} \in \text{span}\{\mathbf{v}_1, ..., \mathbf{v}_k\}$.

(i) In other words, we want to find $\lambda_1, ..., \lambda_k \in \mathbb{R}$ satisfying $\mathbf{x} = \lambda_1\mathbf{v}_1 + \cdots + \lambda_k\mathbf{v}_k$.

(ii) Transform this into a system of equations where the $\lambda_1, ..., \lambda_k$ are the unknowns.

(iii) Solve this system; there being no solutions means $\mathbf{x}$ is **not** in the spanning set.

Note: Specifically for $\mathbb{R}^3$, the span of two vectors corresponds to a plane through the origin containing each of those vectors. Moreover, we can get the equation of this plane by computing the cross product of the vectors. Indeed, for $\text{span}\{\mathbf{v}, \mathbf{w}\}$, the cross product will have the form $\mathbf{v} \times \mathbf{w} = (a, b, c)$, corresponding to the plane $ax + by + cz = 0$.

5.4 Linear Dependence and Independence

Definition 5.4.1 A set of vectors $\{\mathbf{v}_1, ..., \mathbf{v}_k\} \subseteq V$ is called linearly independent if

$$\lambda_1 \mathbf{v}_1 + \cdots \lambda_k \mathbf{v}_k = \mathbf{0} \qquad \Rightarrow \qquad \lambda_1 = \cdots = \lambda_k = 0,$$

that is the only way that a linear combination of them can be the zero vector is if each scalar is itself zero. If this is not the case, the we call the set of vectors linearly dependent.

Remark As a matter of convention, we define the empty set $\emptyset$ to be linearly <u>in</u>dependent.

Method – Linear Independence: We want to see if $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ are linearly independent.

(i) In other words, we want to find $\lambda_1, ..., \lambda_k \in \mathbb{R}$ satisfying $\lambda_1 \mathbf{v}_1 + \cdots + \lambda_k \mathbf{v}_k = \mathbf{0}$.
(ii) Transform this into a system of equations where the $\lambda_1, ..., \lambda_k$ are the unknowns.
(iii) Solve this system; we have <u>in</u>dependence if the **unique** solution is $\lambda_1 = \cdots = \lambda_k = 0$.

Lemma *Any set of vectors containing* $\mathbf{0}$ *is automatically linearly dependent.*

Proof: Suppose we have $\{\mathbf{v}_1, ..., \mathbf{v}_{k-1}, \mathbf{0}\}$. The linear combination we are interested in is

$$\lambda_1 \mathbf{v}_1 + \cdots \lambda_{k-1} \mathbf{v}_{k-1} + \lambda_k \mathbf{0} = \mathbf{0}.$$

Note that this is solved by $\lambda_1 = \cdots = \lambda_{k-1} = 0$ and $\lambda_k = 24$, for example. Because there is not a unique only-zero solution for the λ_i (since the final scalar can be any number we want), we know that we have linear dependence. □

Theorem 5.4.5 *The set* $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ *is linearly dependent if and only if at least one vector* $\mathbf{v}_i$ *is a linear combination of its predecessors, that is* $\mathbf{v}_i \in \text{span}\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}\}$.

Proof: ($\Rightarrow$) Suppose that the vectors are linearly dependent, meaning there exist $\lambda_j \in \mathbb{R}$ **not all** zero such that $\lambda_1 \mathbf{v}_1 + \cdots + \lambda_k \mathbf{v}_k = \mathbf{0}$. Suppose that i is the largest index such that $\lambda_i \neq 0$ (meaning that $\lambda_j = 0$ if $j > i$). Then, we can re-write this linear combination as

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_i \mathbf{v}_i = \mathbf{0}.$$

Because $\lambda_i \neq 0$, we can divide by this number and rearrange to obtain a linear combination expressing $\mathbf{v}_i$ in terms of its predecessors:

$$\mathbf{v}_i = -\frac{\lambda_1}{\lambda_i} \mathbf{v}_1 - \cdots - \frac{\lambda_{i-1}}{\lambda_i} \mathbf{v}_{i-1}.$$

($\Leftarrow$) Assume there is an i such that $\mathbf{v}_i$ is a linear combination of its predecessors of the form

$$\mathbf{v}_i = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_{i-1}\mathbf{v}_{i-1},$$

with some of the $\lambda_j \in \mathbb{R}$ non-zero. But now, we see that

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_{i-1}\mathbf{v}_{i-1} - \mathbf{v}_i + 0\mathbf{v}_{i+1} + \cdots + 0\mathbf{v}_k = \mathbf{0},$$

which shows linear dependence because not all the scalars are zero. □

Lemma 5.4.6 *Let $\mathbf{v}_1, ..., \mathbf{v}_k \in V$ be vectors. Then, we have the following:*

(i) $\operatorname{span}\{\alpha\mathbf{v}_1, ..., \mathbf{v}_k\} = \operatorname{span}\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ *for any non-zero $\alpha \in \mathbb{R}$.*

(ii) $\operatorname{span}\{\mathbf{v}_1 + \alpha\mathbf{v}_2, ..., \mathbf{v}_k\} = \operatorname{span}\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ *for any $\alpha \in \mathbb{R}$.*

Proof: (i) Let $\mathbf{x} \in \operatorname{span}\{\alpha\mathbf{v}_1, ..., \mathbf{v}_k\}$, meaning that for some $\lambda_i \in \mathbb{R}$,

$$\mathbf{x} = \lambda_1 \alpha \mathbf{v}_1 + \cdots + \lambda_k \mathbf{v}_k.$$

But we can define new scalars as follows: $\mu_1 = \lambda_1\alpha$ and $\mu_i = \lambda_i$ for $2 \le i \le k$. Thus, we see that

$$\mathbf{x} = \mu_1 \mathbf{v}_1 + \cdots + \mu_k \mathbf{v}_k,$$

which is precisely that $\mathbf{x} \in \operatorname{span}\{\mathbf{v}_1, ..., \mathbf{v}_k\}$. This gives us $\operatorname{span}\{\alpha\mathbf{v}_1, ..., \mathbf{v}_k\} \subseteq \operatorname{span}\{\mathbf{v}_1, ..., \mathbf{v}_k\}$. Conversely, let $\mathbf{x} \in \operatorname{span}\{\mathbf{v}_1, ..., \mathbf{v}_k\}$, meaning that for some $\lambda_i \in \mathbb{R}$,

$$\mathbf{x} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_k \mathbf{v}_k.$$

We again define new scalars as follows: $\nu_1 = \lambda_1/\alpha$ and $\nu_i = \lambda_i$ for $2 \le i \le k$. Thus, we see that

$$\mathbf{x} = \nu_1 \alpha \mathbf{v_1} + \cdots + \nu_k \mathbf{v}_k$$

which is precisely that $\mathbf{x} \in \operatorname{span}\{\alpha\mathbf{v}_1, ..., \mathbf{v}_k\}$. This gives us $\operatorname{span}\{\mathbf{v}_1, ..., \mathbf{v}_k\} \subseteq \operatorname{span}\{\alpha\mathbf{v}_1, ..., \mathbf{v}_k\}$. Because we have both inclusions, we have equality.

(ii) Let $\mathbf{x} \in \operatorname{span}\{\mathbf{v}_1 + \alpha\mathbf{v}_2, ..., \mathbf{v}_k\}$, meaning that form some $\lambda_i \in \mathbb{R}$,

$$\begin{aligned}\mathbf{x} &= \lambda_1(\mathbf{v}_1 + \alpha\mathbf{v}_2) + \lambda_2\mathbf{v}_2 + \cdots + \lambda_k\mathbf{v}_k \\ &= \lambda_1\mathbf{v}_1 + (\lambda_1\alpha + \lambda_2)\mathbf{v}_2 + \cdots + \lambda_k\mathbf{v}_k \\ &= \mu_1\mathbf{v}_1 + \mu_2\mathbf{v}_2 + \cdots + \mu_k\mathbf{v}_k,\end{aligned}$$

where we define the new scalars as $\mu_2 = \lambda_1\alpha + \lambda_2$ and $\mu_i = \lambda_i$ for all $i \neq 2$. This implies that $\mathbf{x} \in \operatorname{span}\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ and gives us the inclusion $\operatorname{span}\{\mathbf{v}_1 + \alpha\mathbf{v}_2, ..., \mathbf{v}_k\} \subseteq \operatorname{span}\{\mathbf{v}_1, ..., \mathbf{v}_k\}$. Conversely, let $\mathbf{x} \in \operatorname{span}\{\mathbf{v}_1, ..., \mathbf{v}_k\}$, meaning that for some $\lambda_i \in \mathbb{R}$,

$$\begin{aligned}\mathbf{x} &= \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_k\mathbf{v}_k \\ &= \lambda_1(\mathbf{v}_1 + \alpha\mathbf{v}_2) + (\lambda_2 - \lambda_1\alpha)\mathbf{v}_2 + \cdots + \lambda_k\mathbf{v}_k \\ &= \nu_1(\mathbf{v}_1 + \alpha\mathbf{v}_2) + \nu_2\mathbf{v}_2 + \cdots + \nu_k\mathbf{v}_k,\end{aligned}$$

where we define the new scalars as $\nu_2 = \lambda_2 - \lambda_1\alpha$ and $\nu_i = \lambda_i$ for all $i \neq 2$. This implies that $\mathbf{x} \in \operatorname{span}\{\mathbf{v}_1 + \alpha\mathbf{v}_2, ..., \mathbf{v}_k\}$ and gives us the inclusion $\operatorname{span}\{\mathbf{v}_1, ..., \mathbf{v}_k\} \subseteq \operatorname{span}\{\mathbf{v}_1 + \alpha\mathbf{v}_2, ..., \mathbf{v}_k\}$. Because we have both inclusions, we have equality. □

Definition 5.4.7 We say two matrices are row-equivalent if each can be obtained from the other via a sequence of elementary row operations.

Definition 5.4.8 Let A be an $m \times n$ matrix. The row space is the vector subspace of $\mathbb{R}^n$ spanned by the vectors whose entries are the rows of A.

Theorem 5.4.9 *Let A and B be row-equivalent $m \times n$ matrices.*
- (i) *The set of rows of A is linearly independent if and only if the set of rows of B is linearly independent.*
- (ii) *The set of rows of A is linearly dependent if and only if the set of rows of B is linearly dependent.*
- (iii) *The row space of A is the same as the row space of B.*

Proof: (of (iii) only) It suffices to prove the result in the case that B is obtained from A via **one** elementary row operation, that is $B = EA$ for some elementary matrix E. Suppose the set of rows of A is $\{\mathbf{r}_1, ..., \mathbf{r}_n\}$, meaning that the row space of A is $\text{span}\{\mathbf{r}_1, ..., \mathbf{r}_n\}$.

(a) If E corresponds to $R_i \leftrightarrow R_j$, it is clear that the row space of B is precisely that of A.

(b) If E corresponds to $R_i \mapsto kR_i$, then the set of rows of B is $\{\mathbf{r}_1, ..., \mathbf{r}_{i-1}, k\mathbf{r}_i, \mathbf{r}_{i+1}, ..., \mathbf{r}_n\}$, meaning that the row space of B is $\text{span}\{\mathbf{r}_1, ..., \mathbf{r}_{i-1}, k\mathbf{r}_i, \mathbf{r}_{i+1}, ..., \mathbf{r}_n\}$. But by Lemma 5.4.6(i), we know that this is precisely $\text{span}\{\mathbf{r}_1, ..., \mathbf{r}_n\}$.

(c) If E corresponds to $R_i \mapsto R_i + kR_j$, the set of rows of B is $\{\mathbf{r}_1, ..., \mathbf{r}_{i-1}, \mathbf{r}_i + k\mathbf{r}_j, \mathbf{r}_{i+1}, ..., \mathbf{r}_n\}$, meaning that the row space of B is $\text{span}\{\mathbf{r}_1, ..., \mathbf{r}_{i-1}, \mathbf{r}_i + k\mathbf{r}_j, \mathbf{r}_{i+1}, ..., \mathbf{r}_n\}$. But by Lemma 5.4.6(ii), we know that this is precisely $\text{span}\{\mathbf{r}_1, ..., \mathbf{r}_n\}$. □

Method – Linear Independence Redux: We now have another method for testing if a set of vectors is linearly independent. Namely, construct the matrix whose rows are those vectors and reduce it to REF. If there is at least one all-zero row, then we have linear dependence. If not, then we have linear independence.

5.5 Bases and Dimensions of Vector Spaces

Definition 5.5.1 A basis for a vector space V is a set $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ satisfying these:
- (i) B is linearly independent.
- (ii) B spans V.

Note: This means that not only can every element of V be written as a linear combination of the elements of B (via the spanning property), but that B is the smallest-sized set where this occurs (by linear independence).

Method – Proving a Set is a Basis: To prove that a set B is a basis for V, it suffices to use the previous methods to show each part of Definition 5.5.1 in turn. Indeed, show that the elements of B are linearly independent via the previous method(s) and show that **any** $\mathbf{v} \in V$ is such that $\mathbf{v} \in \text{span}(B)$.

Theorem 5.5.6 *Let V be a vector space with basis $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$. Every vector $\mathbf{v} \in V$ can be written uniquely as a linear combination of the basis vectors:*

$$\mathbf{v} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n, \qquad \lambda_i \in \mathbb{R}.$$

Proof: Because B is a basis, we have $V = \text{span}(B)$ by Definition 5.5.1(ii). Hence, every $\mathbf{v} \in V$ can be written in the form $\mathbf{v} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n$ for some $\lambda_i \in \mathbb{R}$. It remains to prove that this is unique. Suppose we also have $\mathbf{v} = \mu_1 \mathbf{v}_1 + \cdots + \mu_n \mathbf{v}_n$ for some $\mu_i \in \mathbb{R}$. It is now clear that

$$\begin{aligned} \mathbf{0} &= \mathbf{v} - \mathbf{v} \\ &= (\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n) - (\mu_1 \mathbf{v}_1 + \cdots + \mu_n \mathbf{v}_n) \\ &= (\lambda_1 - \mu_1)\mathbf{v}_1 + \cdots + (\lambda_n - \mu_n)\mathbf{v}_n. \end{aligned}$$

By linear independence of B, it must be that each of these scalars is zero, that is $\lambda_i - \mu_i = 0$. Consequently, $\lambda_i = \mu_i$ and so the linear combination is unique. □

Theorem 5.5.9 *Let V be a vector space with basis $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$.*
(i) *If a subset of V has strictly* **more** *than n elements, then it is linearly dependent.*
(ii) *If a subset of V has strictly* **less** *than n elements, then it* **can't** *span V.*

Proof: (**Non-examinable**) The proof can be found in the lecture notes for the course. □

Corollary 5.5.10 *Let V be a vector space with basis $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ and $S \subseteq V$.*
(i) *If S is linearly independent, then $|S| \leq n$.*
(ii) *If S spans V, then $|S| \geq n$.*
(iii) *If S is a basis for V, then $|S| = n$.*

Proof: This is just the contrapositive of Theorem 5.5.9. □

Theorem 5.5.7 *Every basis for a vector space V has the same size.*

Definition 5.5.8 The dimension $\dim(V)$ of a vector space V is the size of a basis for V.

Note: These results show that the dimension is encoded into **every** basis for a vector space, and that this number is well-defined because all bases have the same size.

Remark 5.5.11 A natural question to ask is the following: does the trivial space $\{\mathbf{0}\}$ have a basis? For the notion of dimension to be reasonable, we would expect that $\dim(\{\mathbf{0}\}) < 1$, i.e. the zero vector space has dimension less than that of a line. Hence, the basis of $\{\mathbf{0}\}$ should be of size zero and this is true if the basis is $\emptyset$.

Corollary 5.5.13 (to Theorem 5.5.7) *Let U be a subspace of V. Then,* $\dim(U) \leq \dim(V)$ *with equality if and only if $U = V$.*

Definition 5.5.15 The row rank of a matrix A is the dimension of its row space.

Corollary 5.5.16 (to Theorem 5.4.9(iii)) *Let A and B be row-equivalent matrices. Then, the row rank of A is the same as the row rank of B.*

Remark We now know that the dimension of the row space of a matrix A is precisely the number of linearly independent vectors required to span it. Therefore, we have the following:

- If the number of rows is **equal to** the row rank, then the rows are linearly independent.
- If the number of rows is **greater than** the row rank, then the rows are linearly dependent.

Corollary 5.5.17 *Let A be an $n \times n$ matrix. Then, the rows of A form a basis for $\mathbb{R}^n$ if and only if the reduced row echelon form of A is the identity matrix I_n.*

Proof: Well, the n rows form a basis for $\mathbb{R}^n$ if and only if the row rank of A is n, which is equivalent to saying that the RREF is I_n. □

Corollary 5.5.18 *Let A be an $n \times n$ matrix. Then, the rows of A are linearly independent if and only if* $\det(A) \neq 0$.

Proof: Per Corollary 5.5.17, the n rows are linearly independent if and only if the RREF is I_n, which is equivalent to $\det(A) \neq 0$. □

Method – Basis for the Row Space: Suppose we want a basis for the row space of an $m \times n$ matrix A. We simply perform row reduction on A and the non-zero rows are precisely the vectors in the basis for the row space.

6 Eigenspaces and Diagonalising Matrices

6.1 Diagonalising Matrices

Definition 6.1.1 Let A be an $n \times n$ matrix. We call it diagonalisable if there exists an invertible matrix P such that $D := P^{-1}AP$ is a diagonal matrix.

Method – Diagonalising a Matrix: We want to diagonalise A (assuming it possible).

(i) Find the eigenvalues $\lambda_1, ..., \lambda_k$ of A.
(ii) Find the corresponding eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_k$ using the previous method.
(iii) The diagonal matrix D is that with the eigenvalues on the diagonal.
(iv) The invertible matrix P is that with the eigenvectors as its columns.

We need to have the eigenvalues and corresponding eigenvectors in the same order, so

$$D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_k \end{pmatrix} \qquad \text{means that} \qquad P = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}.$$

Remark 6.1.2 If a matrix has too few linearly independent eigenvectors, then it is not possible to diagonalise the matrix. Indeed, the method above tells us that the columns of P are precisely the eigenvectors, and notice that

$$\begin{aligned} P \text{ is invertible } &\Leftrightarrow P^T \text{ is invertible} && \text{(Lemma 2.5.12)} \\ &\Leftrightarrow \text{ the rows of } P^T \text{ are linearly independent} && \text{(Corollary 5.5.18)} \\ &\Leftrightarrow \text{ the columns of } P \text{ are linearly independent.} \end{aligned}$$

6.2 Some Properties of Eigenvectors

Definition 6.2.1 Let A be an $n \times n$ matrix with eigenvalue λ. The eigenspace of λ is

$$\mathcal{U}_\lambda := \{\mathbf{v} \in \mathbb{R}^n : A\mathbf{v} = \lambda\mathbf{v}\}.$$

Note: The eigenspace is the set of eigenvectors associated to λ together with $\mathbf{0}$.

Lemma 6.2.2 *The eigenspace $\mathcal{U}_\lambda$ is a subspace of $\mathbb{R}^n$.*

Proof: We proceed with showing the required properties.

(i) Clearly, $A\mathbf{0} = 0 = \lambda\mathbf{0}$, which is to say $\mathbf{0} \in \mathcal{U}_\lambda$.

(ii) Let $\mathbf{v}, \mathbf{w} \in \mathcal{U}_\lambda$, meaning $A\mathbf{v} = \lambda\mathbf{v}$ and $A\mathbf{w} = \lambda\mathbf{w}$. Then, we see that

$$A(\mathbf{v} + \mathbf{w}) = A\mathbf{v} + A\mathbf{w} = \lambda\mathbf{v} + \lambda\mathbf{w} = \lambda(\mathbf{v} + \mathbf{w}),$$

which is to say $\mathbf{v} + \mathbf{w} \in \mathcal{U}_\lambda$.

(iii) Let $\mathbf{v} \in \mathcal{U}_\lambda$, meaning $A\mathbf{v} = \lambda\mathbf{v}$, and take some scalar $\mu \in \mathbb{R}$. Then, we see that

$$A(\mu\mathbf{v}) = \mu(A\mathbf{v}) = \mu\lambda\mathbf{v} = \lambda(\mu\mathbf{v}),$$

which is to say $\mu\mathbf{v} \in \mathcal{U}_\lambda$. □

Theorem 6.2.3 *Let A be an $n \times n$ matrix with at least m distinct eigenvalues $\lambda_1, ..., \lambda_m$ with corresponding eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_m$. Then, the set $\{\mathbf{v}_1, ..., \mathbf{v}_m\}$ is linearly independent.*

Proof: We proceed by induction. Indeed, if $m = 1$, the set $\{\mathbf{v}_1\}$ is trivially linearly independent, so the base case holds true. Next, assume that the result holds for $m \leq k$; we will prove that the result holds for $m = k + 1$ under this hypothesis. Indeed, let $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ be linearly independent and assume to the contrary that $\{\mathbf{v}_1, ..., \mathbf{v}_k, \mathbf{v}_{k+1}\}$ is **not** linearly independent. As the first k eigenvectors are linearly independent, we know from Theorem 5.4.5 that this means $\mathbf{v}_{k+1}$ is a linear combination of its predecessors:

$$\mathbf{v}_{k+1} = \mu_1\mathbf{v}_1 + \cdots + \mu_k\mathbf{v}_k, \qquad \mu_i \in \mathbb{R} \textbf{ not all } \text{zero}. \tag{1}$$

If we multiply the above on the left by the matrix A, we obtain the following expression:

$$A\mathbf{v}_{k+1} = A(\mu_1\mathbf{v}_1 + \cdots + \mu_k\mathbf{v}_k) = \mu_1 A\mathbf{v}_1 + \cdots + \mu_k A\mathbf{v}_k.$$

Taking into account the eigenvalue equation, we see that this is precisely the same as

$$\lambda_{k+1}\mathbf{v}_{k+1} = \mu_1\lambda_1\mathbf{v}_1 + \cdots + \mu_k\lambda_k\mathbf{v}_k. \tag{2}$$

If $\lambda_{k+1} = 0$, then we know that the other $\lambda_i \neq 0$ since we assume the eigenvalues are distinct. But this would mean that the left-hand side and, therefore, right-hand side of equation (2) is zero. This is a contradiction because we assume that the set $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ are linearly independent. Hence, $\lambda_{k+1} \neq 0$. We can therefore divide equation (2) by this eigenvalue:

$$\lambda_{k+1}\mathbf{v}_{k+1} = \mu_1\frac{\lambda_1}{\lambda_{k+1}}\mathbf{v}_1 + \cdots + \mu_k\frac{\lambda_k}{\lambda_{k+1}}\mathbf{v}_k. \tag{3}$$

If we subtract equation (3) from equation (1), we obtain the following:

$$\mu_1\left(1 - \frac{\lambda_1}{\lambda_{k+1}}\right)\mathbf{v}_1 + \cdots + \mu_k\left(1 - \frac{\lambda_k}{\lambda_{k+1}}\right)\mathbf{v}_k = \mathbf{0}.$$

Again, by linear independence of $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$, we know that $\mu_i(1 - \lambda_i/\lambda_{k+1}) = 0$ for each i. Because the μ_i are not all zero, there is at least one index such that $1 - \lambda_i/\lambda_{k+1} = 0$, that is $\lambda_i = \lambda_{k+1}$. This is a contradiction to distinctiveness. Therefore, the inductive step is proven (by contradiction). Hence, by the principal of mathematical induction, the result follows. □

Theorem 6.2.4 *Let A be an $n \times n$ matrix with at least m distinct eigenvalues $\lambda_1, ..., \lambda_m$. Let B_i be a basis for the eigenspace $\mathcal{U}_{\lambda_i}$. Then, $S := B_1 \cup \cdots \cup B_m$ is linearly independent.*

6.3 Calculating Powers of Matrices

Method – Powers of a Matrix: Suppose we wish to find A^m for any integer $m \in \mathbb{Z}$.

(i) Diagonalise the matrix, so that $P^{-1}AP = D$.

(ii) Rearrange the equation in Step (i) so that we have $A = PDP^{-1}$.

(iii) Notice that $A^m = (PDP^{-1})^m = PD^mP^{-1}$.

(iv) Finally, the entries of D^m are just the entries of D raised to the power of m.