Chapter 2

Matrices, vectors, and vector spaces

Reading

There is no specific essential reading for this chapter. It is essential that you do some reading, but the topics discussed in this chapter are adequately covered in so many texts on ‘linear algebra’ that it would be artificial and unnecessarily limiting to specify precise passages from precise texts. The list below gives examples of relevant reading. (For full publication details, see Chapter 1.)

Ostaszewski, A. *Advanced Mathematical Methods*. Chapter 1, sections 1.1–1.7; Chapter 3, sections 3.1–3.2 and 3.4.

Leon, S.J., *Linear Algebra with Applications*. Chapter 1, sections 1.1–1.3; Chapter 2, section 2.1; Chapter 3, sections 3.1–3.6; Chapter 4, sections 4.1–4.2.

Simon, C.P. and Blume, L., *Mathematics for Economists*. Chapter 7, sections 7.1–7.4; Chapter 9, section 9.1; Chapter 11, and Chapter 27.


Introduction

In this chapter we first very briefly revise some basics about vectors and matrices, which will be familiar from the ‘Mathematics for economists’ subject. We then explore some new topics, in particular the important theoretical concept of a vector space.
Vectors and matrices

An \( n \)-vector \( \mathbf{v} \) is a list of \( n \) numbers, written either as a **row-vector**

\[
(v_1, v_2, \ldots, v_n),
\]

or a **column-vector**

\[
\begin{pmatrix}
  v_1 \\
v_2 \\
  \vdots \\
v_n
\end{pmatrix}.
\]

The numbers \( v_1, v_2, \) and so on are known as the **components** or **coordinates** of \( \mathbf{v} \). The **zero vector** is the vector with all of its entries equal to 0.

The set \( \mathbb{R}^n \) denotes the set of all vectors of length \( n \), and we usually think of these as column vectors.

We can define **addition** of two \( n \)-vectors by the rule

\[
(w_1, w_2, \ldots, w_n) + (v_1, v_2, \ldots, v_n) = (w_1 + v_1, w_2 + v_2, \ldots, w_n + v_n).
\]

(The rule is described here for row vectors but the obvious counterpart holds for column vectors.) Also, we can multiply a vector by any single number \( \alpha \) (usually called a **scalar** in this context) by the following rule:

\[
\alpha(v_1, v_2, \ldots, v_n) = (\alpha v_1, \alpha v_2, \ldots, \alpha v_n).
\]

For vectors \( \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \) and numbers \( \alpha_1, \alpha_2, \ldots, \alpha_k \), the vector \( \alpha_1 \mathbf{v}_1 + \cdots + \alpha_k \mathbf{v}_k \) is known as a **linear combination** of the vectors \( \mathbf{v}_1, \ldots, \mathbf{v}_k \).

A **matrix** is an array of numbers

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

We denote this array by the single letter \( A \), or by \((a_{ij})\), and we say that \( A \) has \( m \) **rows** and \( n \) **columns**, or that it is an \( m \times n \) matrix. We also say that \( A \) is a matrix of **size** \( m \times n \). If \( m = n \), the matrix is said to be **square**. The number \( a_{ij} \) is known as the \((i, j)\)th entry of \( A \). The row vector \((a_{i1}, a_{i2}, \ldots, a_{in})\) is **row** \( i \) of \( A \), or the \( i \)th row of \( A \), and the column vector

\[
\begin{pmatrix}
a_{1j} \\
a_{2j} \\
  \vdots \\
a_{nj}
\end{pmatrix}
\]

is **column** \( j \) of \( A \), or the \( j \)th column of \( A \).

It is useful to think of row and column vectors as matrices. For example, we may think of the row vector \((1, 2, 4)\) as being equal to the \( 1 \times 3 \) matrix \((1 \ 2 \ 4)\). (Indeed, the only visible difference is that the vector has commas and the matrix does not, and this is merely a notational difference.)
Recall that if \( A \) and \( B \) are two matrices of the same size then we define \( A + B \) to be the matrix whose elements are the sums of the corresponding elements in \( A \) and \( B \). Formally, the \((i,j)\)th entry of the matrix \( A + B \) is \( a_{ij} + b_{ij} \) where \( a_{ij} \) and \( b_{ij} \) are the \((i,j)\)th entries of \( A \) and \( B \), respectively. Also, if \( c \) is a number, we define \( cA \) to be the matrix whose elements are \( c \) times those of \( A \); that is, \( cA \) has \((i,j)\)th entry \( ca_{ij} \). If \( A \) and \( B \) are matrices such that the number (say \( n \)) of columns of \( A \) is equal to the number of rows of \( B \), we define the product \( C = AB \) to be the matrix whose elements are:

\[
c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.
\]

The transpose of a matrix

\[
A = (a_{ij}) = \begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{pmatrix}
\]

is the matrix

\[
A^T = (a_{ji}) = \begin{pmatrix}
a_{11} & a_{21} & \cdots & a_{n1} \\
a_{12} & a_{22} & \cdots & a_{n2} \\
\vdots & \vdots & \ddots & \vdots \\
a_{1m} & a_{2m} & \cdots & a_{nm}
\end{pmatrix}
\]

that is obtained from \( A \) by interchanging rows and columns.

**Example:** For example, if \( A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \) and \( B = \begin{pmatrix} 1 & 5 & 3 \end{pmatrix} \), then

\[
A^T = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \quad B^T = \begin{pmatrix} 1 \\ 5 \\ 3 \end{pmatrix}.
\]

**Determinants**

The determinant of a square matrix \( A \) is a particular number associated with \( A \), written \( \text{det} A \) or \( |A| \). When \( A \) is a \( 2 \times 2 \) matrix, the determinant is given by the formula

\[
\text{det} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.
\]

For example,

\[
\text{det} \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} = 1 \times 3 - 2 \times 2 = -1.
\]

For a \( 3 \times 3 \) matrix, the determinant is given as follows:

\[
\text{det} \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} = aei - afh + bfg - bdi + cdh - ceg.
\]

We have used expansion by first row to express the determinant in terms of \( 2 \times 2 \) determinants. Note how this works and note, particularly, the minus sign in front of the second \( 2 \times 2 \) determinant.
Activity 2.1 Calculate the determinant of
\[
\begin{pmatrix}
3 & 1 & 0 \\
-2 & -4 & 3 \\
5 & 4 & -2
\end{pmatrix}.
\]
(You should find that the answer is $-1$.)

Row operations

Recall from ‘Mathematics for economists’ (for it is not going to be reiterated here in all its gory detail!) that there are three types of elementary row operation one can perform on a matrix. These are:

- multiply a row by non-zero constant
- add a multiple of one row to another
- interchange (swap) two rows.

In ‘Mathematics for economists’, row operations were used to solve systems of linear equations by reducing the augmented matrix of the system to echelon form. (Yes, I am afraid you do have to remember this!) A matrix is an echelon matrix if it is of the following form:
\[
\begin{pmatrix}
1 & * & * & \ldots & * \\
0 & 0 & 1 & \ldots & * \\
0 & 0 & 0 & 1 & \ldots & * \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \ldots & 0
\end{pmatrix},
\]
where the $*$ entries denote any numbers. The 1s which have been indicated are called leading ones.

Rank, range, null space, and linear equations

The rank of a matrix

There are several ways of defining the rank of a matrix, and we shall meet some other (more sophisticated) ways later. For the moment, we use the following definition.

**Definition 2.1 (Rank of a matrix)** The rank, $\text{rank}(A)$, of a matrix $A$ is the number of non-zero rows in an echelon matrix obtained from $A$ by elementary row operations.

By a non-zero row, we simply mean one that contains entries other than 0.

**Example:** Consider the matrix
\[
A = \begin{pmatrix}
1 & 2 & 1 & 1 \\
2 & 2 & 0 & 2 \\
3 & 4 & 1 & 3
\end{pmatrix}.
\]
Reducing this to echelon form using elementary row operations, we have:

\[
\begin{pmatrix}
1 & 2 & 1 & 1 \\
2 & 2 & 0 & 2 \\
3 & 4 & 1 & 3 \\
\end{pmatrix}
\rightarrow
\begin{pmatrix}
1 & 2 & 1 & 1 \\
0 & -2 & -2 & 0 \\
0 & -2 & -2 & 0 \\
\end{pmatrix}
\rightarrow
\begin{pmatrix}
1 & 2 & 1 & 1 \\
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 \\
\end{pmatrix}.
\]

This last matrix is in echelon form and has two non-zero rows, so the matrix \(A\) has rank 2.

**Activity 2.2** Prove that matrix

\[
A = \begin{pmatrix}
3 & 1 & 1 & 3 \\
1 & -1 & -1 & 1 \\
1 & 2 & 2 & 1 \\
\end{pmatrix}
\]

has rank 2.

If a square matrix of size \(n \times n\) has rank \(n\) then it is invertible.

Generally, if \(A\) is an \(m \times n\) matrix, then the number of non-zero rows in a reduced, echelon, form of \(A\) can certainly be no more than the total number of rows, \(m\). Furthermore, since the leading ones must be in different columns, the number of leading ones—and hence non-zero rows—in the echelon form, can be no more than the total number, \(n\), of columns. Thus we have:

**Theorem 2.1** For an \(m \times n\) matrix \(A\), \(\text{rank}(A) \leq \min\{m, n\}\), where \(\min\{m, n\}\) denotes the smaller of the two integers \(m\) and \(n\).

**Rank and systems of linear equations**

Recall that to solve a system of linear equations, one forms the augmented matrix and reduces it to echelon form by using elementary row operations.

**Example:** Consider the system of equations

\[
x_1 + 2x_2 + x_3 = 1 \\
2x_1 + 2x_2 = 2 \\
3x_1 + 4x_2 + x_3 = 2.
\]

Using row operations to reduce the augmented matrix to echelon form, we obtain

\[
\begin{pmatrix}
1 & 2 & 1 & 1 \\
2 & 2 & 0 & 2 \\
3 & 4 & 1 & 2 \\
\end{pmatrix}
\rightarrow
\begin{pmatrix}
1 & 2 & 1 & 1 \\
0 & -2 & -2 & 0 \\
0 & -2 & -2 & 1 \\
\end{pmatrix}
\rightarrow
\begin{pmatrix}
1 & 2 & 1 & 1 \\
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 \\
\end{pmatrix}.
\]
Thus the original system of equations is equivalent to the system
\[
\begin{align*}
x_1 + 2x_2 + x_3 &= 1 \\
x_2 + x_3 &= 0 \\
0x_1 + 0x_2 + 0x_3 &= -1.
\end{align*}
\]
But this system has no solutions, since there are no values of \(x_1, x_2, x_3\) that satisfy the last equation. It reduces to the false statement ‘0 = -1’, whatever values we give the unknowns. We deduce, therefore, that the original system has no solutions.

In the example just given, it turned out that the echelon form has a row of the kind 
\((0 \ 0 \ldots \ 0 a)\), with \(a \neq 0\). This means that the original system is equivalent to one in which one there is an equation
\[0x_1 + 0x_2 + \cdots + 0x_n = a \ (a \neq 0).\]
Clearly this equation cannot be satisfied by any values of the \(x_i\)’s, and we say that such a system is inconsistent.

If there is no row of this kind, the system has at least one solution, and we say that it is consistent.

In general, the rows of the echelon form of a consistent system are of two types. The first type, which we call a zero row, consists entirely of zeros: 
\((0 \ 0 \ldots \ 0)\). The other type is a row in which at least one of the components, not the last one, is non-zero: 
\((* \ * \ldots \ *b)\), where at least one of the *’s is not zero. We shall call this a non-zero row. The standard method of reduction to echelon form ensures that the zero rows (if any) come below the non-zero rows.

The number of non-zero rows in the echelon form is known as the rank of the system. Thus the rank of the system is the rank of the augmented matrix. Note, however, that if the system is consistent then there can be no leading one in the last column of the reduced augmented matrix, for that would mean there was a row of the form \((0 \ 0 \ldots \ 0 1)\). Thus, in fact, the rank of a consistent system \(Ax = b\) is precisely the same as the rank of the matrix \(A\).

Suppose we have a consistent system, and suppose first that the rank \(r\) is strictly less than \(n\), the number of unknowns. Then the system in echelon form (and hence the original one) does not provide enough information to specify the values of \(x_1, x_2, \ldots, x_n\) uniquely.

**Example:** Suppose we are given a system for which the augmented matrix reduces to the echelon form
\[
\begin{pmatrix}
1 & 3 & -2 & 0 & 2 & 0 & 0 \\
0 & 0 & 1 & 2 & 0 & 3 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 5 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}.
\]
Here the rank (number of non-zero rows) is \(r = 3\) which is strictly less than the number of unknowns, \(n = 6\). The corresponding system is
\[
\begin{align*}
x_1 + 3x_2 - 2x_3 + 2x_5 &= 0 \\
x_3 + 2x_4 + 3x_6 &= 1 \\
x_6 &= 5.
\end{align*}
\]
These equations can be rearranged to give \(x_1, x_3\) and \(x_6\):
\[
x_1 = -3x_2 + 2x_3 - 2x_5, \quad x_3 = 1 - 2x_4 - 3x_6, \quad x_6 = 5.
\]
Using back-substitution to solve for $x_1$, $x_3$ and $x_6$ in terms of $x_2$, $x_4$ and $x_5$ we get
\[ x_6 = 5, \quad x_3 = -14 - 2x_4, \quad x_1 = -28 - 3x_2 - 4x_4 - 2x_5. \]
The form of these equations tells us that we can assign any values to $x_2$, $x_4$ and $x_5$, and then the other variables will be determined. Explicitly, if we give $x_2$, $x_4$, $x_5$ the arbitrary values $s, t, u$, the solution is given by
\[ x_1 = -28 - 3s - 4t - 2u, \quad x_2 = s, \quad x_3 = -14 - 2t, \quad x_4 = t, \quad x_5 = u, \quad x_6 = 5. \]

Observe that there are infinitely many solutions, because the so-called ‘free unknowns’ $x_2, x_4, x_5$ can take any values $s, t, u$.

Generally, we can describe what happens when the echelon form has $r < n$ non-zero rows (0 0 ... 0 1 * * ... *). If the leading 1 is in the $k$th column it is the coefficient of the unknown $x_k$. So if the rank is $r$ and the leading 1’s occur in columns $c_1, c_2, \ldots, c_r$ then the general solution to the system can be expressed in a form where the unknowns $x_{c_1}, x_{c_2}, \ldots, x_{c_r}$ are given in terms of the other $n - r$ unknowns, and those $n - r$ unknowns are free to take any values. In the preceding example, we have $n = 6$ and $r = 3$, and the 3 unknowns $x_1, x_3, x_6$ can be expressed in terms of the $6 - 3 = 3$ free unknowns $x_2, x_4, x_5$.

In the case $r = n$, where the number of non-zero rows $r$ in the echelon form is equal to the number of unknowns $n$, there is only one solution to the system — for, the echelon form has no zero rows, and the leading 1’s move one step to the right as we go down the rows, and in this case there is a unique solution obtained by back-substitution from the echelon form. In fact, this can be thought of as a special case of the more general one discussed above: since $r = n$ there are $n - r = 0$ free unknowns, and the solution is therefore unique.

We can now summarise our conclusions concerning a general linear system.

- If the echelon form has a row (0 0 ... 0 0 a), with $a \neq 0$, the original system is inconsistent; it has no solutions.
- If the echelon form has no rows of the above type it is consistent, and the general solution involves $n - r$ free unknowns, where $r$ is the rank. When $r < n$ there are infinitely many solutions, but when $r = n$ there are no free unknowns and so there is a unique solution.

General solution of a linear system in vector notation

Consider the system solved above. We found that the general solution in terms of three free unknowns, or parameters, $s, t, u$ is
\[ x_1 = -28 - 3s - 4t - 2u, \quad x_2 = s, \quad x_3 = -14 - 2t, \quad x_4 = t, \quad x_5 = u, \quad x_6 = 5. \]
If we write $\mathbf{x}$ as a column vector,
\[ \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}, \]
then
\[
x = \begin{pmatrix} -28 - 3s - 4t - 2u \\ s \\ -14 - 2t \\ u \\ 5 \end{pmatrix} = \begin{pmatrix} -28 \\ 0 \\ -14 \\ 0 \\ 5 \end{pmatrix} + \begin{pmatrix} -3s \\ s \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -4t \\ 0 \\ -2t \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -2u \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.
\]

That is, the general solution is
\[
x = v + s u_1 + t u_2 + u u_3,
\]
where
\[
v = \begin{pmatrix} -28 \\ -14 \\ 0 \\ 0 \\ 5 \end{pmatrix}, \quad u_1 = \begin{pmatrix} -3 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} -4 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad u_3 = \begin{pmatrix} -2 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.
\]

Applying the same method generally to a consistent system of rank \(r\) with \(n\) unknowns, we can express the general solution of a consistent system \(A x = b\) in the form
\[
x = v + s_1 u_1 + s_2 u_2 + \cdots + s_{n-r} u_{n-r}.
\]

Note that, if we put all the \(s_i\)’s equal to 0, we get a solution \(x = v\), which means that \(A v = b\), so \(v\) is a particular solution of the system. Putting \(s_1 = 1\) and the remaining \(s_i\)’s equal to zero, we get a solution \(x = v + u_1\), which means that \(A(v + u_1) = b\). Thus
\[
b = A(v + u_1) = Av + Au_1 = b + Au_1.
\]

Comparing the first and last expressions, we see that \(Au_1\) is the zero vector \(0\). Clearly, the same equation holds for \(u_2, \ldots, u_{n-r}\). So we have proved the following.

The general solution of \(A x = b\) is the sum of:

- a particular solution \(v\) of the system \(A x = b\) and
- a linear combination \(s_1 u_1 + s_2 u_2 + \cdots + s_{n-r} u_{n-r}\) of solutions \(u_1, u_2, \ldots, u_{n-r}\) of the system \(A x = 0\).

**Null space**

It’s clear from what we’ve just seen that the general solution to a consistent linear system involves solutions to the system \(A x = 0\). This set of solutions is given a special name: the null space or kernel of a matrix \(A\). This null space, denoted \(N(A)\), is the set of all solutions \(x\) to \(A x = 0\), where \(0\) is the all-zero vector. That is,

**Definition 2.2 (Null space)** For an \(m \times n\) matrix \(A\), the null space of \(A\) is the subset
\[
N(A) = \{ x \in \mathbb{R}^n : A x = 0 \}
\]
of \(\mathbb{R}^n\), where \(0 = (0, 0, \ldots, 0)^T\) is the all-0 vector of length \(m\).
Note that the linear system \( Ax = 0 \) is always consistent since, for example, \( 0 \) is a solution. It follows from the discussion above that the general solution to \( Ax = 0 \) involves \( n - r \) free unknowns, where \( r \) is the rank of \( A \) (and that, if \( r = n \), then there is just one solution, namely \( x = 0 \)).

We now formalise the connection between the solution set of a general consistent linear system, and the null space of the matrix determining the system.

**Theorem 2.2** Suppose that \( A \) is an \( m \times n \) matrix, that \( b \in \mathbb{R}^n \), and that the system \( Ax = b \) is consistent. Suppose that \( x_0 \) is any solution of \( Ax = b \). Then the set of all solutions of \( Ax = b \) consists precisely of the vectors \( x_0 + z \) for \( z \in N(A) \); that is,

\[
\{ x : Ax = b \} = \{ x_0 + z : z \in N(A) \}.
\]

**Proof** To show the two sets are equal, we show that each is a subset of the other. Suppose, first, that \( x \) is any solution of \( Ax = b \). Because \( x_0 \) is also a solution, we have

\[
A(x - x_0) = Ax - Ax_0 = b - b = 0,
\]

so the vector \( z = x - x_0 \) is a solution of the system \( Az = 0 \); in other words, \( z \in N(A) \). But then \( x = x_0 + (x - x_0) = x_0 + z \) where \( z \in N(A) \). This shows that

\[
\{ x : Ax = b \} \subseteq \{ x_0 + z : z \in N(A) \}.
\]

Conversely, if \( z \in N(A) \) then

\[
A(x_0 + z) = Ax_0 + Az = b + 0 = b,
\]

so \( x_0 + z \in \{ x : Ax = b \} \). This shows that

\[
\{ x_0 + z : z \in N(A) \} \subseteq \{ x : Ax = b \}.
\]

So the two sets are equal, as required.

**Range**

The **range** of a matrix \( A \) is defined as follows.

**Definition 2.3 (Range of a matrix)** Suppose that \( A \) is an \( m \times n \) matrix. Then the **range** of \( A \), denoted by \( R(A) \), is the subset

\[
\{ Ax : x \in \mathbb{R}^n \}
\]

of \( \mathbb{R}^m \). That is, the range is the set of all vectors \( y \in \mathbb{R}^m \) of the form \( y = Ax \) for some \( x \in \mathbb{R}^n \).

Suppose that the columns of \( A \) are \( c_1, c_2, \ldots, c_n \). Then we may write \( A = (c_1 c_2 \ldots c_n) \). If \( x = (\alpha_1, \alpha_2, \ldots, \alpha_n)^T \in \mathbb{R}^n \), then the product \( Ax \) is equal to

\[
\alpha_1 c_1 + \alpha_2 c_2 + \cdots + \alpha_n c_n.
\]
Activity 2.3 Convince yourself of this last statement, that

\[ Ax = \alpha_1 c_1 + \alpha_2 c_2 + \cdots + \alpha_n c_n. \]

So, \( R(A) \), as the set of all such products, is the set of all linear combinations of the columns of \( A \). For this reason \( R(A) \) is also called the column space of \( A \). (More on this later in this chapter.)

Example: Suppose that \( A = \begin{pmatrix} 1 & 2 \\ -1 & 3 \\ 2 & 1 \end{pmatrix} \). Then for \( x = (\alpha_1, \alpha_2)^T \),

\[ Ax = \begin{pmatrix} 1 & 2 \\ -1 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \alpha_1 + 2\alpha_2 \\ -\alpha_1 + 3\alpha_2 \\ 2\alpha_1 + \alpha_2 \end{pmatrix}, \]

so

\[ R(A) = \left\{ \begin{pmatrix} \alpha_1 + 2\alpha_2 \\ -\alpha_1 + 3\alpha_2 \\ 2\alpha_1 + \alpha_2 \end{pmatrix} : \alpha_1, \alpha_2 \in \mathbb{R} \right\}. \]

This may also be written as

\[ R(A) = \{ \alpha_1 c_1 + \alpha_2 c_2 : \alpha_1, \alpha_2 \in \mathbb{R} \}, \]

where

\[ c_1 = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \quad c_2 = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} \]

are the columns of \( A \).

Vector spaces

Definition of a vector space

We know that vectors of \( \mathbb{R}^n \) can be added together and that they can be ‘scaled’ by real numbers. That is, for every \( x, y \in \mathbb{R}^n \) and every \( \alpha \in \mathbb{R} \), it makes sense to talk about \( x + y \) and \( \alpha x \). Furthermore, these operations of addition and multiplication by a scalar (that is, multiplication by a real number) behave and interact ‘sensibly’, in that, for example,

\[ \alpha(x + y) = \alpha x + \alpha y, \]

\[ \alpha(\beta x) = (\alpha\beta)x, \]

\[ x + y = y + x, \]

and so on.

But it is not only vectors in \( \mathbb{R}^n \) that can be added and multiplied by scalars. There are other sets of objects for which this is possible. Consider the set \( V \) of all functions from \( \mathbb{R} \) to \( \mathbb{R} \). Then any two of these functions can be added: given \( f, g \in V \) we simply define the function \( f + g \) by

\[ (f + g)(x) = f(x) + g(x). \]
Also, for any $\alpha \in \mathbb{R}$, the function $\alpha f$ is given by

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

These operations of addition and scalar multiplication are sometimes said to be **pointwise addition** and **pointwise scalar multiplication**. This might seem a bit abstract, but think about what the functions $x + x^2$ and $2x$ represent: the former is the function $x$ plus the function $x^2$, and the latter is the function $x$ multiplied by the scalar 2. So this is just a different way of looking at something you are already familiar with. It turns out that $V$ and its rules for addition and multiplication by a scalar satisfy the same key properties as does the set of vectors in $\mathbb{R}^n$ with its addition and scalar multiplication. We refer to a set with an addition and scalar multiplication which behave appropriately as a vector space. We now give the formal definition of a vector space.

**Definition 2.4 (Vector space)** A vector space $V$ is a set equipped with an addition operation and a scalar multiplication operation such that for all $\alpha, \beta \in \mathbb{R}$ and all $u, v, w \in V$,

1. $u + v \in V$ (closure under addition)
2. $u + v = v + u$ (the commutative law for addition)
3. $u + (v + w) = (u + v) + w$ (the associative law for addition)
4. There is a single member $0$ of $V$, called the **zero vector**, such that for all $v \in V$, $v + 0 = v$
5. For every $v \in V$ there is an element $w \in V$ (usually written as $-v$), called the **negative** of $v$, such that $v + w = 0$
6. $\alpha v \in V$ (closure under scalar multiplication)
7. $\alpha(u + v) = \alpha u + \alpha v$ (distributive under scalar multiplication)
8. $(\alpha + \beta)v = \alpha v + \beta v$
9. $\alpha(\beta v) = (\alpha \beta)v$
10. $1v = v$.

Other properties follow from those listed in the definition. For instance, we can see that $0x = 0$ for all $x$, as follows:

$$0x = (0 + 0)x = 0x + 0x,$$

so, adding the negative $-0x$ of $0x$ to each side,

$$0 = 0x + (-0x) = (0x + 0x) + (-0x) = 0x + (0x + (-0x)) = 0x + 0 = 0x.$$

(A bit sneaky, but just remember the result: $0x = 0$.)

(Note that this definition says nothing at all about ‘multiplying’ together two vectors: the only operations with which the definition is concerned are addition and scalar multiplication.)

A vector space as we have defined it is often called a **real vector space**, to emphasise that the ‘scalars’ $\alpha, \beta$ and so on are real numbers rather than complex numbers. There is a notion of **complex vector space**, but this will not concern us in this subject.
Examples

Example: The set $\mathbb{R}^n$ is a vector space with the usual way of adding and scalar multiplying vectors.

Example: The set $V$ of functions from $\mathbb{R}$ to $\mathbb{R}$ with pointwise addition and scalar multiplication (described earlier in this section) is a vector space. Note that the zero vector in this space is the function that maps every real number to 0—that is, the identically-zero function. Similarly, if $S$ is any set then the set $V$ of functions $f : S \rightarrow \mathbb{R}$ is not a vector space. Why? Well, let $f$ be any function other than the identically-zero function. Then for some $x$, $f(x) \neq 0$. Now, for some $\alpha$, the real number $\alpha f(x)$ will not belong to $S$. (Since $S$ is not the whole of $\mathbb{R}$, there is $z \in \mathbb{R}$ such that $z \notin S$. Since $f(x) \neq 0$, we may write $z = (z/f(x)) f(x)$ and we may therefore take $\alpha = z/f(x)$.) This all shows that the function $\alpha f$ is not in $V$, since it is not a function from $\mathbb{R}$ to $S$.

Example: The set of $m \times n$ matrices is a vector space, with the usual addition and scalar multiplication of matrices. The ‘zero vector’ in this vector space is the zero $m \times n$ matrix with all entries equal to 0.

Example: Let $V$ be the set of all 3-vectors with third entry equal to 0: that is,

$$V = \left\{ \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} : x, y \in \mathbb{R} \right\}.$$

Then $V$ is a vector space with the usual addition and scalar multiplication. To verify this, we need only check that $V$ is closed under addition and scalar multiplication: associativity and all the other required properties hold because they hold for $\mathbb{R}^3$ (and hence for this subset of $\mathbb{R}^3$). Furthermore, if we can show that $V$ is closed under scalar multiplication and addition, then for any particular $v \in V$, $0v = 0 \in V$. So we simply need to check that $V \neq \emptyset$, that if $u, v \in V$ then $u + v \in V$, and if $\alpha \in \mathbb{R}$ and $v \in V$ then $\alpha v \in V$. Each of these is easy to check.

**Activity 2.4** Verify that for $u, v \in V$ and $\alpha \in \mathbb{R}$, $u + v \in V$ and $\alpha v \in V$.

Subspaces

The last example above is informative. Arguing as we did there, if $V$ is a vector space and $W \subseteq V$ is non-empty and closed under scalar multiplication and addition, then $W$ too is a vector space (and we do not need to verify that all the other properties hold). The formal definition of a subspace is as follows.

**Definition 2.5 (Subspace)** A subspace $W$ of a vector space $V$ is a non-empty subset of $V$ that is itself a vector space (under the same operations of addition and scalar multiplication as $V$).

The discussion given justifies the following important result.
Theorem 2.3 Suppose $V$ is a vector space. Then a non-empty subset $W$ of $V$ is a subspace if and only if:

- for all $u, v \in W$, $u + v \in W$ ($W$ is closed under addition), and
- for all $v \in W$ and $\alpha \in \mathbb{R}$, $\alpha v \in W$ ($W$ is closed under scalar multiplication).

Subspaces connected with matrices

Null space

Suppose that $A$ is an $m \times n$ matrix. Then the null space $N(A)$, the set of solutions to the linear system $Ax = 0$, is a subspace of $\mathbb{R}^n$.

Theorem 2.4 For any $m \times n$ matrix $A$, $N(A)$ is a subspace of $\mathbb{R}^n$.

Proof To prove this we have to verify that $N(A) \neq \emptyset$, and that if $u, v \in N(A)$ and $\alpha \in \mathbb{R}$, then $u + v \in N(A)$ and $\alpha u \in N(A)$. Since $A0 = 0$, $0 \in N(A)$ and hence $N(A) \neq \emptyset$. Suppose $u, v \in N(A)$. Then to show $u + v \in N(A)$ and $\alpha u \in N(A)$, we must show that $u + v$ and $\alpha u$ are solutions of $Ax = 0$. We have

$$A(u + v) = Au + Av = 0 + 0 = 0$$

and

$$A(\alpha u) = \alpha(Au) = \alpha 0 = 0,$$

so we have shown what we needed.

Note that the null space is the set of solutions to the homogeneous linear system. If we instead consider the set of solutions $S$ to a general system $Ax = b$, $S$ is not a subspace of $\mathbb{R}^n$ if $b \neq 0$ (that is, if the system is not homogeneous). This is because $0$ does not belong to $S$. However, as we indicated above, there is a relationship between $S$ and $N(A)$: if $x_0$ is any solution of $Ax = b$ then $S = \{x_0 + z : z \in N(A)\}$, which we may write as $x_0 + N(A)$. Generally, if $W$ is a subspace of a vector space $V$ and $x \in V$ then the set $x + W$ defined by

$$x + W = \{x + w : w \in W\}$$

is called an affine subspace of $V$. An affine subspace is not generally a subspace (although every subspace is an affine subspace, as we can see by taking $x = 0$).

Range

Recall that the range of an $m \times n$ matrix is

$$R(A) = \{Ax : x \in \mathbb{R}^n\}.$$

Theorem 2.5 For any $m \times n$ matrix $A$, $R(A)$ is a subspace of $\mathbb{R}^m$. 

18
Proof We need to show that if \( u, v \in R(A) \) then \( u + v \in R(A) \) and, for any \( \alpha \in \mathbb{R} \), \( \alpha v \in R(A) \). So suppose \( u, v \in R(A) \). Then for some \( y_1, y_2 \in \mathbb{R}^n \), \( u = Ay_1, v = Ay_2 \). We need to show that \( u + v = Ay \) for some \( y \). Well,
\[
    u + v = Ay_1 + Ay_2 = A(y_1 + y_2),
\]
so we may take \( y = y_1 + y_2 \) to see that, indeed, \( u + v \in R(A) \). Next,
\[
    \alpha v = \alpha (Ay_1) = A(\alpha y_1),
\]
so \( \alpha v = Ay \) for some \( y \) (namely \( y = \alpha y_1 \)) and hence \( \alpha v \in R(A) \).

Linear independence

Linear independence is a central idea in the theory of vector spaces. We say that vectors \( x_1, x_2, \ldots, x_m \) in \( \mathbb{R}^n \) are linearly dependent (LD) if there are numbers \( \alpha_1, \alpha_2, \ldots, \alpha_m \), not all zero, such that
\[
    \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m = 0,
\]
the zero vector. The left-hand side is termed a non-trivial linear combination. This condition is entirely equivalent to saying that one of the vectors may be expressed as a linear combination of the others. The vectors are linearly independent (LI) if they are not linearly dependent; that is, if no non-trivial linear combination of them is the zero vector or, equivalently, whenever
\[
    \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m = 0,
\]
then, necessarily, \( \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0 \). We have been talking about \( \mathbb{R}^n \), but the same definitions can be used for any vector space. We state them formally now.

Definition 2.6 (Linear independence) Let \( V \) be a vector space and \( v_1, \ldots, v_m \in V \). Then \( v_1, v_2, \ldots, v_m \) form a linearly independent set or are linearly independent if and only if
\[
    \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m = 0 \implies \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0;
\]
that is, if and only if no non-trivial linear combination of \( v_1, v_2, \ldots, v_m \) equals the zero vector.

Definition 2.7 (Linear dependence) Let \( V \) be a vector space and \( v_1, v_2, \ldots, v_m \in V \). Then \( v_1, v_2, \ldots, v_m \) form a linearly dependent set or are linearly dependent if and only if there are real numbers \( \alpha_1, \alpha_2, \ldots, \alpha_m \), not all zero, such that
\[
    \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m = 0;
\]
that is, if and only if some non-trivial linear combination of the vectors is the zero vector.

Example: In \( \mathbb{R}^3 \), the following vectors are linearly dependent:
\[
    v_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 4 \\ 5 \\ 11 \end{pmatrix}.
\]
This is because
\[
    2v_1 + v_2 - v_3 = 0.
\]
(Note that this can also be written as \( v_3 = 2v_1 + v_2 \).)
Testing linear independence in $\mathbb{R}^n$

Given $m$ vectors $x_1, \ldots, x_k \in \mathbb{R}^n$, let $A$ be the $n \times k$ matrix $A = (x_1 \ x_2 \ \cdots \ x_k)$ with the vectors as its columns. Then for a $k \times 1$ matrix (or column vector)

$$c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix},$$

the product $Ac$ is exactly the linear combination

$$c_1x_1 + c_2x_2 + \cdots + c_kx_k.$$

**Theorem 2.6** The vectors $x_1, x_2, \ldots, x_k$ are linearly dependent if and only if the linear system $Ac = 0$ has a solution other than $c = 0$, where $A$ is the matrix $A = (x_1 \ x_2 \ \cdots \ x_k)$. Equivalently, the vectors are linearly independent precisely when the only solution to the system is $c = 0$.

If the vectors are linearly dependent, then any solution $c \neq 0$ of the system $Ac = 0$ will directly give a linear combination of the vectors that equals the zero vector.

Now, we know from our experience of solving linear systems with row operations that the system $Ac = 0$ will have precisely the one solution $c = 0$ if and only if we obtain from $A$ an echelon matrix in which there are $k$ leading ones. That is, if and only if $	ext{rank}(A) = k$. (Think about this!) But (as we argued earlier when considering rank), the number of leading ones is always at most the number of rows, so we certainly need to have $k \leq n$. Thus, we have the following result.

**Theorem 2.7** Suppose that $x_1, \ldots, x_k \in \mathbb{R}^n$. Then the set $\{x_1, \ldots, x_k\}$ is linearly independent if and only if the matrix $(x_1 \ x_2 \ \cdots \ x_k)$ has rank $k$.

But the rank is always at most the number of rows, so we certainly need to have $k \leq n$. Also, there is a set of $n$ linearly independent vectors in $\mathbb{R}^n$. In fact, there are infinitely many such sets, but an obvious one is

$$\{e_1, e_2, \ldots, e_n\},$$

where $e_i$ is the vector with every entry equal to 0 except for the $i$th entry, which is 1. Thus, we have the following result.

**Theorem 2.8** The maximum size of a linearly independent set in $\mathbb{R}^n$ is $n$.

So any set of more than $n$ vectors in $\mathbb{R}^n$ is linearly dependent. On the other hand, it should not be imagined that any set of $n$ or fewer is linearly independent: that isn’t true.

**Linear span**

Recall that by a linear combination of vectors $v_1, v_2, \ldots, v_k$ we mean a vector of the form

$$v = \alpha_1v_1 + \alpha_2v_2 + \cdots + \alpha_kv_k.$$
The set of all linear combinations of a given set of vectors forms a vector space, and we give it a special name.

**Definition 2.8** Suppose that $V$ is a vector space and that $v_1, v_2, \ldots, v_k \in V$. Let $S$ be the set of all linear combinations of the vectors $v_1, \ldots, v_k$. That is,

$$S = \{ \alpha_1 v_1 + \cdots + \alpha_k v_k : \alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R} \}.$$  

Then $S$ is a subspace of $V$, and is known as the subspace spanned by the set $X = \{v_1, \ldots, v_k\}$ (or, the linear span or, simply, span, of $v_1, v_2, \ldots, v_k$). This subspace is denoted by

$$S = \text{Lin}\{v_1, v_2, \ldots, v_k\}$$

or $S = \text{Lin}(X)$.

Different texts use different notations. For example, Simon and Blume use $\mathcal{L}[v_1, v_2, \ldots, v_n]$. Notation is important, but it is nothing to get anxious about: just always make it clear what you mean by your notation: use words as well as symbols!

We have already observed that the range $R(A)$ of an $m \times n$ matrix $A$ is equal to the set of all linear combinations of its columns. In other words, $R(A)$ is the span of the columns of $A$ and is often called the column space. It is also possible to consider the row space $\text{RS}(A)$ of a matrix: this is the span of the rows of $A$. If $A$ is an $m \times n$ matrix the row space will be a subspace of $\mathbb{R}^n$.

**Bases and dimension**

The following result is very important in the theory of vector spaces.

**Theorem 2.9** If $x_1, x_2, \ldots, x_n$ are linearly independent vectors in $\mathbb{R}^n$, then for any $x$ in $\mathbb{R}^n$, $x$ can be written as a linear combination of $x_1, \ldots, x_n$. We say that $x_1, x_2, \ldots, x_n$ span $\mathbb{R}^n$.

**Proof** Because $x_1, \ldots, x_n$ are linearly independent, the $n \times n$ matrix

$$A = (x_1 x_2 \ldots x_n)$$

is such that $\text{rank}(A) = n$. (See above.) In other words, $A$ reduces to an echelon matrix with exactly $n$ leading ones. Suppose now that $x$ is any vector in $\mathbb{R}^n$ and consider the system $Ax = x$. By the discussion above about the rank of linear systems, this system has a (unique) solution. But let’s spell it out. Because $A$ has rank $n$, it can be reduced (by row operations) to an echelon matrix with $n$ leading ones. The augmented matrix $(Ax)$ can therefore be reduced to a matrix $(Es)$ where $E$ is an $n \times n$ echelon matrix with $n$ leading ones. Solving by back-substitution in the usual manner, we can find a solution to $Ax = x$. This shows that any vector $x$ can be expressed in the form

$$x = A \mathbf{z} = (x_1 x_2 \ldots x_n) \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix},$$

where we have written $\mathbf{z}$ as $(\alpha_1, \alpha_2, \ldots, \alpha_n)^T$. Expanding this matrix product, we have that any $x \in \mathbb{R}^n$ can be expressed as a linear combination

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n,$$
as required.

There is another important property of linearly independent sets of vectors.

**Theorem 2.10** If \(x_1, x_2, \ldots, x_m\) are linearly independent in \(\mathbb{R}^n\) and
\[
  c_1 x_1 + c_2 x_2 + \cdots + c_m x_m = c'_1 x_1 + c'_2 x_2 + \cdots + c'_m x_m
\]
then
\[
  c_1 = c'_1, c_2 = c'_2, \ldots, c_m = c'_m.
\]

**Activity 2.5** Prove this. Use the fact that
\[
  c_1 x_1 + c_2 x_2 + \cdots + c_m x_m = c'_1 x_1 + c'_2 x_2 + \cdots + c'_m x_m
\]
if and only if
\[
  (c_1 - c'_1)x_1 + (c_2 - c'_2)x_2 + \cdots + (c_m - c'_m)x_m = 0.
\]

It follows from these two results that if we have \(n\) linearly independent vectors in \(\mathbb{R}^n\), then any vector in \(\mathbb{R}^n\) can be expressed in exactly one way as a linear combination of the \(n\) vectors. We say that the \(n\) vectors form a basis of \(\mathbb{R}^n\). The formal definition of a (finite) basis for a vector space is as follows.

**Definition 2.9** (Finite Basis) Let \(V\) be a vector space. Then the subset \(B = \{v_1, v_2, \ldots, v_n\}\) of \(V\) is said to be a basis for (or of) \(V\) if:

- \(B\) is a linearly independent set of vectors, and
- \(V = \text{Lin}(B)\).

An alternative characterisation of a basis can be given: \(B\) is a basis of \(V\) if every vector in \(V\) can be expressed in exactly one way as a linear combination of the vectors in \(B\).

**Example:** The vector space \(\mathbb{R}^n\) has the basis \(\{e_1, e_2, \ldots, e_n\}\) where \(e_i\) is (as earlier) the vector with every entry equal to 0 except for the \(i\)th entry, which is 1. It’s clear that the vectors are linearly independent, and there are \(n\) of them, so we know straight away that they form a basis. In fact, it’s easy to see that they span the whole of \(\mathbb{R}^n\), since for any \(x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n\),
\[
  x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n.
\]
The basis \(\{e_1, e_2, \ldots, e_n\}\) is called the standard basis of \(\mathbb{R}^n\).

**Activity 2.6** Convince yourself that the vectors are linearly independent and that they span the whole of \(\mathbb{R}^n\).
Dimensions

A fundamental result is that if a vector space has a finite basis, then all bases are of the same size.

**Theorem 2.11** Suppose that the vector space $V$ has a finite basis consisting of $d$ vectors. Then any basis of $V$ consists of exactly $d$ vectors. The number $d$ is known as the **dimension** of $V$, and is denoted $\dim(V)$.

A vector space which has a finite basis is said to be **finite-dimensional**. Not all vector spaces are finite-dimensional. (For example, the vector space of real functions with pointwise addition and scalar multiplication has no finite basis.)

**Example:** We already know $\mathbb{R}^n$ has a basis of size $n$. (For example, the standard basis consists of $n$ vectors.) So $\mathbb{R}^n$ has dimension $n$ (which is reassuring, since it is often referred to as $n$-dimensional Euclidean space).

Other useful characterisations of bases and dimension can be given.

**Theorem 2.12** Let $V$ be a finite-dimensional vector space of dimension $d$. Then:

- $d$ is the largest size of a linearly independent set of vectors in $V$. Furthermore, any set of $d$ linearly independent vectors is necessarily a basis of $V$;
- $d$ is the smallest size of a subset of vectors that span $V$ (so, if $X$ is a finite subset of $V$ and $V = \text{Lin}(X)$, then $|X| \geq d$). Furthermore, any finite set of $d$ vectors that spans $V$ is necessarily a basis.

Thus, $d = \dim(V)$ is the largest possible size of a linearly independent set of vectors in $V$, and the smallest possible size of a spanning set of vectors (a set of vectors whose linear span is $V$).

**Dimension and bases of subspaces**

Suppose that $W$ is a subspace of the finite-dimensional vector space $V$. Any set of linearly independent vectors in $W$ is also a linearly independent set in $V$.

**Activity 2.7** *Convince yourself of this last statement.*

Now, the dimension of $W$ is the largest size of a linearly independent set of vectors in $W$, so there is a set of $\dim(W)$ linearly independent vectors in $V$. But then this means that $\dim(W) \leq \dim(V)$, since the largest possible size of a linearly independent set in $V$ is $\dim(V)$. There is another important relationship between bases of $W$ and $V$: this is that any basis of $W$ can be extended to one of $V$. The following result states this precisely.\(^1\)

**Theorem 2.13** Suppose that $V$ is a finite-dimensional vector space and that $W$ is a subspace of $V$. Then $\dim(W) \leq \dim(V)$. Furthermore, if $\{w_1, w_2, \ldots, w_r\}$ is a

\(^1\) See Section 27.7 of Simon and Blume for a proof.
basis of $W$ then there are $s = \dim(V) - \dim(W)$ vectors $v_1, v_2, \ldots, v_s \in V$ such that
\{w_1, w_2, \ldots, w_r, v_1, v_2, \ldots, v_s\} is a basis of $V$. (In the case $W = V$, the basis of $W$
 is already a basis of $V$.) That is, we can obtain a basis of the whole space $V$ by adding
vectors of $V$ to any basis of $W$. 

Finding a basis for a linear span in $\mathbb{R}^n$

Suppose we are given $m$ vectors $x_1, x_2, \ldots, x_m$ in $\mathbb{R}^n$, and we want to find a basis for
the linear span $\text{Lin}\{x_1, \ldots, x_m\}$. The point is that the $m$ vectors themselves might
not form a linearly independent set (and hence not a basis). A useful technique is to
form a matrix with the $x_i^T$ as rows, and to perform row operations until the resulting
matrix is in echelon form. Then a basis of the linear span is given by the transposed
non-zero rows of the echelon matrix (which, it should be noted, will not generally be
among the initial given vectors). The reason this works is that: (i) row operations
are such that at any stage in the resulting procedure, the row space of the matrix is
equal to the row space of the original matrix, which is precisely the linear span of the
original set of vectors (if we ignore the difference between row and column vectors),
and (ii) the non-zero rows of an echelon matrix are linearly independent (which is
clear, since each has a one in a position where the others all have zero).

**Example:** We find a basis for the subspace of $\mathbb{R}^5$ spanned by the vectors

$$
x_1 = \begin{pmatrix} 1 \\ -1 \\ 2 \\ -2 \\ -1 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 2 \\ 1 \\ -2 \\ -2 \\ -2 \end{pmatrix}, \quad x_3 = \begin{pmatrix} -1 \\ 2 \\ 1 \\ -2 \\ -2 \end{pmatrix}, \quad x_4 = \begin{pmatrix} 3 \\ 0 \\ 1 \\ -2 \\ -3 \end{pmatrix}.
$$

The matrix

$$
\begin{pmatrix}
\begin{bmatrix} x_1^T \\ x_2^T \\ x_3^T \\ x_4^T \end{bmatrix}
\end{pmatrix}
$$

is

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

Reducing this to echelon form by elementary row operations,

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

The echelon matrix at the end of this tells us that a basis for $\text{Lin}\{x_1, x_2, x_3, x_4\}$ is
formed from the first two rows, transposed, of the echelon matrix:

$$
\left\{ \begin{pmatrix} 1 \\ -1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -2 \\ 0 \end{pmatrix} \right\}.
$$
If we really want a basis that consists of some of the original vectors, then all we need to do is take those vectors that ‘correspond’ to the final non-zero rows in the echelon matrix. By this, we mean the rows of the original matrix that have ended up as non-zero in the echelon matrix. For instance, in the example just given, the first and second rows of the original matrix correspond to the non-zero rows of the echelon matrix, so a basis of the span is \( \{x_1, x_2\} \). On the other hand, if we interchange rows, the correspondence won’t be so obvious. If, for example, in reduction to echelon form, we end up with the top two rows of the echelon matrix being non-zero, but have at some stage performed a single ‘interchange’ operation, swapping rows 2 and 3 (without swapping any others), then it is the first and third rows of the original matrix that we should take as our basis.

### Dimensions of range and nullspace

As we have seen, the range and null space of an \( m \times n \) matrix are subspaces of \( \mathbb{R}^m \) and \( \mathbb{R}^n \) (respectively). Their dimensions are so important that they are given special names.

**Definition 2.10 (Rank and Nullity)** The **rank** of a matrix \( A \) is

\[
\text{rank}(A) = \text{dim}(\mathbb{R}(A))
\]

and the **nullity** is

\[
\text{nullity}(A) = \text{dim}(N(A)).
\]

We have, of course, already used the word ‘rank’, so it had better be the case that the usage just given coincides with the earlier one. Fortunately it does. In fact, we have the following connection.

**Theorem 2.14** Suppose that \( A \) is an \( m \times n \) matrix with columns \( c_1, c_2, \ldots, c_n \), and that an echelon form obtained from \( A \) has leading ones in columns \( i_1, i_2, \ldots, i_r \). Then a basis for \( \mathbb{R}(A) \) is

\[
B = \{c_{i_1}, c_{i_2}, \ldots, c_{i_r}\}.
\]

Note that the basis is formed from columns of \( A \), **not** columns of the echelon matrix: the basis consists of those columns corresponding to the leading ones in the echelon matrix.

**Example:** Suppose that (as in an earlier example in this chapter),

\[
A = \begin{pmatrix}
1 & 2 & 1 & 1 \\
2 & 2 & 0 & 2 \\
3 & 4 & 1 & 3
\end{pmatrix}.
\]

Earlier, we reduced this to echelon form using elementary row operations, obtaining the echelon matrix

\[
\begin{pmatrix}
1 & 2 & 1 & 1 \\
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}.
\]

The leading ones in this echelon matrix are in the first and second columns, so a basis for \( \mathbb{R}(A) \) can be obtained by taking the first and second columns of \( A \). (Note:
columns of } \mathbf{A} \text{, not of the echelon matrix!} \text{) Therefore a basis for } \mathbb{R}^{\mathbf{A}} \text{ is}

\begin{align*}
\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 2 \\ 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \\ 3 \end{pmatrix} \right\}.
\end{align*}

There is a very important relationship between the rank and nullity of a matrix. We have already seen some indication of it in our considerations of linear systems. Recall that if an } m \times n \text{ matrix } \mathbf{A} \text{ has rank } r \text{ then the general solution to the (consistent) system } \mathbf{A}\mathbf{x} = \mathbf{0} \text{ involves } n - r \text{ ‘free parameters’. Specifically (noting that } \mathbf{0} \text{ is a particular solution, and using a characterisation obtained earlier in this chapter), the general solution takes the form}

\[ \mathbf{x} = s_1 \mathbf{u}_1 + s_2 \mathbf{u}_2 + \cdots + s_{n-r} \mathbf{u}_{n-r}, \]

where } \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-r} \text{ are themselves solutions of the system } \mathbf{A}\mathbf{x} = \mathbf{0}. \text{ But the set of solutions of } \mathbf{A}\mathbf{x} = \mathbf{0} \text{ is precisely the null space } \mathcal{N}(\mathbf{A}). \text{ Thus, the null space is spanned by the } n - r \text{ vectors } \mathbf{u}_1, \ldots, \mathbf{u}_{n-r}, \text{ and so its dimension is at most } n - r. \text{ In fact, it turns out that its dimension is precisely } n - r. \text{ That is,}

\[ \text{nullity}(\mathbf{A}) = n - \text{rank}(\mathbf{A}). \]

To see this, we need to show that the vectors } \mathbf{u}_1, \ldots, \mathbf{u}_{n-r} \text{ are linearly independent. Because of the way in which these vectors arise (look at the example we worked through), it will be the case that for each of them, there is some position where that vector will have entry equal to 1 and all the other ones will have entry 0. From this we can see that no non-trivial linear combination of them can be the zero vector, so they are linearly independent. We have therefore proved the following central result.

**Theorem 2.15 (Rank-nullity theorem)** For an } m \times n \text{ matrix } \mathbf{A},

\[ \text{rank}(\mathbf{A}) + \text{nullity}(\mathbf{A}) = n. \]

We have seen that dimension of the column space of a matrix is rank(\mathbf{A}). What is the dimension of the row space? Well, it is also rank(\mathbf{A}). This is easy to see if we think about using row operations. A basis of RS(\mathbf{A}) is given by the non-zero rows of the echelon matrix, and there are rank(\mathbf{A}) of these.

**Linear transformations**

We now turn attention to linear mappings (or linear transformations as they are usually called) between vector spaces.

**Definition of linear transformation**

**Definition 2.11 (Linear transformation)** Suppose that } \mathbf{V} \text{ and } \mathbf{W} \text{ are (real) vector spaces. A function } T : \mathbf{V} \rightarrow \mathbf{W} \text{ is linear if for all } \mathbf{u}, \mathbf{v} \in \mathbf{V} \text{ and all } \alpha \in \mathbf{R},

1. \( T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \) and
2. \( T(au) = \alpha T(u) \).

\( T \) is said to be a \textit{linear transformation} (or \textit{linear mapping} or \textit{linear function}).

Equivalently, \( T \) is linear if for all \( u, v \in V \) and \( \alpha, \beta \in \mathbb{R} \),

\[ T(\alpha u + \beta v) = \alpha T(u) + \beta T(v). \]

(This single condition implies the two in the definition, and is implied by them.)

\textbf{Activity 2.8} Prove that this single condition is equivalent to the two of the definition.

Sometimes you will see \( T(u) \) written simply as \( Tu \).

\textbf{Examples}

The most important examples of linear transformations arise from matrices.

\textbf{Example:} Let \( V = \mathbb{R}^n \) and \( W = \mathbb{R}^m \) and suppose that \( A \) is an \( m \times n \) matrix. Let \( T_A \) be the function given by \( T_A(x) = Ax \) for \( x \in \mathbb{R}^n \). That is, \( T_A \) is simply multiplication by \( A \). Then \( T_A \) is a linear transformation. This is easily checked, as follows: first,

\[ T_A(u + v) = A(u + v) = Au + Av = T_A(u) + T_A(v). \]

Next,

\[ T_A(\alpha u) = A(\alpha u) = \alpha Au = \alpha T_A(u). \]

So the two ‘linearity’ conditions are satisfied. We call \( T_A \) the \textit{linear transformation} corresponding to \( A \).

\textbf{Example:} (More complicated) Let us take \( V = \mathbb{R}^n \) and take \( W \) to be the vector space of all functions \( f : \mathbb{R} \to \mathbb{R} \) (with pointwise addition and scalar multiplication). Define a function \( T : \mathbb{R}^n \to W \) as follows:

\[ T(u) = T \left( \begin{array}{c} u_1 \\ u_2 \\ \vdots \\ u_n \end{array} \right) = pu_{u_1, u_2, \ldots, u_n} = pu, \]

where \( pu = pu_{u_1, u_2, \ldots, u_n} \) is the polynomial function given by

\[ pu_{u_1, u_2, \ldots, u_n} = u_1 x + u_2 x^2 + u_3 x^3 + \cdots + u_n x^n. \]

Then \( T \) is a linear transformation. To check this we need to verify that

\[ T(u + v) = T(u) + T(v), \quad T(\alpha u) = \alpha T(u). \]

Now, \( T(u + v) = pu_{u+v}, \quad T(u) = pu \), and \( T(v) = pv \), so we need to check that \( pu_{u+v} = pu + pv \). This is in fact true, since, for all \( x \),

\[ pu_{u+v}(x) = pu_{u_1+v_1, \ldots, u_n+v_n}(x) = (u_1 + v_1)x + (u_2 + v_2)x^2 + \cdots + (u_n + v_n)x^n \]

\[ = (u_1 x + u_2 x^2 + \cdots + u_n x^n) + (v_1 x + v_2 x^2 + \cdots + v_n x^n) \]

\[ = pu(x) + pv(x) \]

\[ = (pu + pv)(x). \]
The fact that for all $x$, $p_\mathbf{u} + p_\mathbf{v}(x) = (p_\mathbf{u} + p_\mathbf{v})(x)$ means that the functions $p_\mathbf{u} + p_\mathbf{v}$ and $p_\mathbf{u} + p_\mathbf{v}$ are identical. The fact that $T(\alpha \mathbf{u}) = \alpha T(\mathbf{u})$ is similarly proved, and you should try it!

**Activity 2.9** Prove that $T(\alpha \mathbf{u}) = \alpha T(\mathbf{u})$.

### Range and null space

Just as we have the range and null space of a matrix, so we have the range and null space of a linear transformation, defined as follows.

**Definition 2.12 (Range and null space of a linear transformation)** Suppose that $T$ is a linear transformation from vector space $V$ to vector space $W$. Then the **range**, $R(T)$, of $T$ is

$$R(T) = \{T(v) : v \in V\},$$

and the **null space**, $N(T)$, of $T$ is

$$N(T) = \{v \in V : T(v) = 0\},$$

where $0$ denotes the zero vector of $W$.

The null space is often called the kernel, and may be denoted $\text{ker}(T)$ in some texts.

Of course, for any matrix, $R(T_A) = R(A)$ and $N(A_T) = N(A)$.

**Activity 2.10** Check this last statement.

The range and null space of a linear transformation $T : V \to W$ are subspaces of $W$ and $V$, respectively.

### Rank and nullity

If $V$ and $W$ are both finite dimensional, then so are $R(T)$ and $N(T)$. We define the **rank** of $T$, $\text{rank}(T)$ to be $\dim(R(T))$ and the **nullity** of $T$, $\text{nullity}(T)$, to be $\dim(N(T))$. As for matrices, there is a strong link between these two dimensions:

**Theorem 2.16 (Rank-nullity theorem for transformations)** Suppose that $T$ is a linear transformation from the finite-dimensional vector space $V$ to the vector space $W$. Then

$$\text{rank}(T) + \text{nullity}(T) = \dim(V).$$

(Note that this result holds even if $W$ is not finite-dimensional.)

For an $m \times n$ matrix $A$, if $T = T_A$, then $T$ is a linear transformation from $V = \mathbb{R}^n$ to $W = \mathbb{R}^m$, and $\text{rank}(T) = \text{rank}(A)$, $\text{nullity}(T) = \text{nullity}(A)$, so this theorem states the earlier result that

$$\text{rank}(A) + \text{nullity}(A) = n.$$
Linear transformations and matrices

In what follows, we consider only linear transformations from \( \mathbb{R}^n \) to \( \mathbb{R}^m \) (for some \( m \) and \( n \)). But much of what we say can be extended to linear transformations mapping from any finite-dimensional vector space to any other finite-dimensional vector space.

We have seen that any \( m \times n \) matrix \( A \) gives a linear transformation \( T_A : \mathbb{R}^n \to \mathbb{R}^m \) (the linear transformation ‘corresponding to \( A \)’), given by \( T_A(u) = Au \). There is a reverse connection: for every linear transformation \( T : \mathbb{R}^n \to \mathbb{R}^m \) there is a matrix \( A \) such that \( T = T_A \).

**Theorem 2.17** Suppose that \( T : \mathbb{R}^n \to \mathbb{R}^m \) is a linear transformation and let \( \{e_1, e_2, \ldots, e_n\} \) denote the standard basis of \( \mathbb{R}^n \). If \( A = A_T \) is the matrix

\[
A = (T(e_1) \ T(e_2) \ \ldots \ T(e_n)),
\]

then \( T = T_A \): that is, for every \( u \in \mathbb{R}^n \), \( T(u) = Au \).

**Proof** Let \( u = (u_1, u_2, \ldots, u_n)^T \) be any vector in \( \mathbb{R}^n \). Then

\[
A_T u = \begin{pmatrix}
1 \\
0 \\
\vdots \\
0
\end{pmatrix}
\begin{pmatrix}
u_1 \\
u_2 \\
\vdots \\
u_n
\end{pmatrix} + \begin{pmatrix}
0 \\
1 \\
\vdots \\
0
\end{pmatrix}
\begin{pmatrix}
u_1 \\
u_2 \\
\vdots \\
u_n
\end{pmatrix} + \cdots + \begin{pmatrix}
0 \\
0 \\
\vdots \\
1
\end{pmatrix}
\begin{pmatrix}
u_1 \\
u_2 \\
\vdots \\
u_n
\end{pmatrix}.
\]

But

\[
u_1 e_1 + u_2 e_2 + \cdots + u_n e_n = u_1 \begin{pmatrix}
1 \\
0 \\
\vdots \\
0
\end{pmatrix} + u_2 \begin{pmatrix}
0 \\
1 \\
\vdots \\
0
\end{pmatrix} + \cdots + u_n \begin{pmatrix}
0 \\
0 \\
\vdots \\
1
\end{pmatrix},
\]

so we have (exactly as we wanted) \( A_T u = T(u) \).

Thus, to each matrix \( A \) there corresponds a linear transformation \( T_A \), and to each linear transformation \( T \) there corresponds a matrix \( A_T \).

**Coordinates**

Suppose that the vectors \( v_1, v_2, \ldots, v_n \) form a basis \( B \) for \( \mathbb{R}^n \). Then any \( x \in \mathbb{R}^n \) can be written in exactly one way as a linear combination,

\[
x = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n,
\]

29
of the vectors in the basis. The vector \((c_1, c_2, \ldots, c_n)^T\) is called the coordinate vector of \(x\) with respect to the basis \(B = \{v_1, v_2, \ldots, v_n\}\). (Note that we assume the vectors of \(B\) to be listed in a particular order, otherwise this definition makes no sense; that is, \(B\) is taken to be an ordered basis.) The coordinate vector is denoted \([x]_B\), and we sometimes refer to it as the coordinates of \(x\) with respect to \(B\).

One very straightforward observation is that the coordinate vector of any \(x \in \mathbb{R}^n\) with respect to the standard basis is just \(x\) itself. This is because

\[
x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n.
\]

What is less immediately obvious is how to find the coordinates of a vector \(x\) with respect to a basis other than the standard one.

**Example:** Suppose that we let \(B\) be the following basis of \(\mathbb{R}^3\):

\[
B = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} -2 \\ -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ -2 \\ -1 \end{pmatrix} \right\}.
\]

If \(x\) is the vector \((5, 7, -2)^T\), then the coordinate vector of \(x\) with respect to \(B\) is

\[
[x]_B = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix},
\]

because

\[
x = 1 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + (-1) \begin{pmatrix} -2 \\ -1 \\ 2 \end{pmatrix} + 2 \begin{pmatrix} 3 \\ -2 \\ -1 \end{pmatrix}.
\]

To find the coordinates of a vector with respect to a basis \(\{x_1, x_2, \ldots, x_n\}\), we need to solve the system of linear equations

\[
c_1 x_1 + c_2 x_2 + \cdots + c_n x_n = x,
\]

which in matrix form is

\[
(x_1 x_2 \ldots x_n) c = x.
\]

In other words, if we let \(P_B\) be the matrix whose columns are the basis vectors (in order),

\[
P_B = (x_1 \ x_2 \ \ldots \ x_n).
\]

Then for any \(x \in \mathbb{R}^n\),

\[
x = P_B^{-1} x.
\]

The matrix \(P_B\) is invertible (because its columns are linearly independent, and hence its rank is \(n\)). So we can also write

\[
[x]_B = P_B^{-1} x.
\]

**Matrix representation and change of basis**

We have already seen that if \(T\) is a linear transformation from \(\mathbb{R}^n\) to \(\mathbb{R}^m\), then there is a corresponding matrix \(A_T\) such that \(T(x) = A_T x\) for all \(x\). The matrix \(A_T\) is given by

\[
(T(e_1) \ T(e_2) \ \ldots \ T(e_n)).
\]
Now suppose that $B$ is a basis of $\mathbb{R}^n$ and $B'$ a basis of $\mathbb{R}^m$, and suppose we want to know the coordinates $[T(x)]_{B'}$ of $T(x)$ with respect to $B'$, given the coordinates $[x]_B$ of $x$ with respect to $B$. Is there a matrix $M$ such that

$$[T(x)]_{B'} = M[x]_B$$

for all $x$? Indeed there is, as the following result shows.

**Theorem 2.18** Suppose that $B = \{x_1, \ldots, x_n\}$ and $B' = \{x'_1, \ldots, x'_m\}$ are (ordered) bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ and that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Let $M = A_T[B, B']$ be the $m \times n$ matrix with $i$th column equal to $[T(x_i)]_{B'}$, the coordinate vector of $T(x_i)$ with respect to the basis $B'$. Then for all $x$, $[T(x)]_{B'} = M[x]_B$. The matrix $A_T[B, B']$ is called the **matrix representing $T$ with respect to bases $B$ and $B'$**.

Can we express $M$ in terms of $A_T$? Let $P_B$ and $P'_B$ be, respectively, the matrices having the basis vectors of $B$ and $B'$ as columns. (So $P_B$ is an $n \times n$ matrix and $P'_B$ is an $m \times m$ matrix.) Then we know that for any $v \in \mathbb{R}^m$, $v = P_B[v]_B$. Similarly, for any $u \in \mathbb{R}^m$, $u = P'_B[u]_{B'}$, so $[u]_{B'} = P^{-1}_B u$. We therefore have (taking $v = x$ and $u = T(x)$)

$$[T(x)]_{B'} = P^{-1}_B T(x).$$

Now, for any $v \in \mathbb{R}^m$, $T(v) = A_T z$ where

$$A_T = (T(e_1) T(e_2) \ldots T(e_n))$$

is the matrix corresponding to $T$. So we have

$$[T(x)]_{B'} = P^{-1}_B T(x) = P^{-1}_B A_T x = P^{-1}_B A_T P_B [x]_B = (P^{-1}_B A_T P_B) [x]_B.$$

Since this is true for all $x$, we have therefore obtained the following result:

**Theorem 2.19** Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, that $B$ is a basis of $\mathbb{R}^n$ and $B'$ is a basis of $\mathbb{R}^m$. Let $P_B$ and $P'_B$ be the matrices whose columns are, respectively, the vectors of $B$ and $B'$. Then the matrix representing $T$ with respect to $B$ and $B'$ is given by

$$A_T[B, B'] = P'^{-1}_B A_T P_B,$$

where

$$A_T = (T(e_1) T(e_2) \ldots T(e_n)).$$

So, for all $x$,

$$[T(x)]_{B'} = P'^{-1}_B A_T P_B [x]_B.$$

Thus, if we change the bases from the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, the matrix representation of the linear transformation changes.

A particular case of this theorem is worth stating separately since it is often useful. (It corresponds to the case in which $m = n$ and $B' = B$.)

**Theorem 2.20** Suppose that $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation and that $B = \{x_1, x_2, \ldots, x_n\}$ is some basis of $\mathbb{R}^n$. Let

$$P = (x_1 \ x_2 \ldots x_n)$$

31
be the matrix whose columns are the vectors of \( B \). Then for all \( x \in \mathbb{R}^n \),

\[
[T(x)]_B = P^{-1}A_T P[x]_B,
\]

where \( A_T \) is the matrix corresponding to \( T \),

\[
A_T = (T(e_1) \ T(e_2) \ldots T(e_n)).
\]

In other words,

\[
A_T[B, B] = P^{-1} A_T P.
\]

**Learning outcomes**

At the end of this chapter and the relevant reading, you should be able to:

- solve systems of linear equations and calculate determinants (all of this being revision)
- compute the rank of a matrix and understand how the rank of a matrix relates to the solution set of a corresponding system of linear equations
- determine the solution to a linear system using vector notation
- explain what is meant by the range and null space of a matrix and be able to determine these for a given matrix
- explain what is meant by a vector space and a subspace
- prove that a given set is a vector space, or a subspace of a given vector space
- explain what is meant by linear independence and linear dependence
- determine whether a given set of vectors is linearly independent or linearly dependent, and in the latter case, find a non-trivial linear combination of the vectors which equals the zero vector
- explain what is meant by the linear span of a set of vectors
- explain what is meant by a basis, and by the dimension of a finite-dimensional vector space
- find a basis for a linear span
- explain how rank and nullity are defined, and the relationship between them (the rank-nullity theorem)
- explain what is meant by a linear transformation and be able to prove a given mapping is linear
- explain what is meant by the range and null space, and rank and nullity of a linear transformation
- comprehend the two-way relationship between matrices and linear transformations
- find the matrix representation of a transformation with respect to two given bases
Sample examination questions

The following are typical exam questions, or parts of questions.

**Question 2.1** Find the general solution of the following system of linear equations. Express your answer in the form $x = v + su$, where $v$ and $u$ are in $\mathbb{R}^3$.

\[
\begin{align*}
3x_1 + x_2 + x_3 &= 3 \\
x_1 - x_2 - x_3 &= 1 \\
x_1 + 2x_2 + 2x_3 &= 1.
\end{align*}
\]

**Question 2.2** Find the rank of the matrix

\[
A = \begin{pmatrix}
1 & -1 & 2 & -1 \\
2 & 1 & -2 & -2 \\
-1 & 2 & -4 & 1 \\
3 & 0 & 0 & -3
\end{pmatrix}.
\]

**Question 2.3** Prove that the following three vectors form a linearly independent set.

\[
x_1 = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}, \quad x_2 = \begin{pmatrix} -3 \\ 7 \\ 10 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 5 \\ 5 \\ 15 \end{pmatrix}.
\]

**Question 2.4** Show that the following vectors are linearly independent.

\[
\begin{align*}
\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}, & \quad \begin{pmatrix} 3 \\ 4 \\ 6 \end{pmatrix}, & \quad \begin{pmatrix} -2 \\ 3 \\ 2 \end{pmatrix}.
\end{align*}
\]

Express the vector

\[
\begin{pmatrix} 4 \\ 7 \\ -3 \end{pmatrix}
\]
as a linear combination of these three vectors.

**Question 2.5** Let $x_1 = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix}, x_2 = \begin{pmatrix} 1 \\ 9 \\ 2 \end{pmatrix}$. Find a vector $x_3$ such that $\{x_1, x_2, x_3\}$ is a linearly independent set of vectors.

**Question 2.6** Show that the following vectors are linearly dependent by finding a linear combination of the vectors that equals the zero vector.

\[
\begin{align*}
\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, & \quad \begin{pmatrix} 0 \\ -1 \\ -3 \end{pmatrix}, & \quad \begin{pmatrix} 4 \\ -11 \\ 5 \end{pmatrix}, & \quad \begin{pmatrix} 9 \\ 2 \\ -1 \end{pmatrix}.
\end{align*}
\]

**Question 2.7** Let

\[
S = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_2 = 3x_1 \right\}.
\]

Prove that $S$ is a subspace of $\mathbb{R}^2$ (where the operations of addition and scalar multiplication are the usual ones).
Question 2.8 Let $V$ be the vector space of all functions from $\mathbb{R} \to \mathbb{R}$ with pointwise addition and scalar multiplication. Let $n$ be a fixed positive integer and let $W$ be the set of all real polynomial functions of degree at most $n$; that is, $W$ consists of all functions of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

where $a_0, a_1, \ldots, a_n \in \mathbb{R}$.

Prove that $W$ is a subspace of $V$, under the usual pointwise addition and scalar multiplication for real functions. Show that $W$ is finite-dimensional by presenting a finite set of functions which spans $W$.

Question 2.9 Find a basis for the null space of the matrix $\begin{pmatrix} 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix}$.

Question 2.10 Let

$$A = \begin{pmatrix} 1 & -2 & 1 & 1 & 2 \\ -1 & 3 & 0 & 2 & -2 \\ 0 & 1 & 1 & 3 & 4 \\ 1 & 2 & 5 & 15 & 5 \end{pmatrix}.$$

Find a basis for the column space of $A$.

Question 2.11 Find the matrix corresponding to the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_1 \\ x_1 + x_2 \end{pmatrix}.$$

Question 2.12 Find bases for the null space and range of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 + 2x_3 \\ x_1 + x_3 \\ 2x_1 + x_2 + 3x_3 \end{pmatrix}.$$

Question 2.13 Let $B$ be the (ordered) basis $\{v_1, v_2, v_3\}$, where $v_1 = (0, 1, 0)^T$, $v_2 = (-4/5, 0, 3/5)^T$, $v_3 = (3/5, 0, 4/5)^T$, and let $x = (1, -1/5, 7/5)^T$. Find the coordinates of $x$ with respect to $B$.

Question 2.14 Suppose that $T : \mathbb{R}^2 \to \mathbb{R}^3$ is the linear transformation given by

$$T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ -5x_1 + 13x_2 \\ -7x_1 + 16x_2 \end{pmatrix}.$$

Find the matrix of $T$ with respect to the bases $B = \{(3, 1)^T, (5, 2)^T\}$ and $B' = \{(1, 0, -1)^T, (-1, 2, 2)^T, (0, 1, 2)^T\}$.

**Sketch answers or comments on selected questions**

Question 2.1 The general solution takes the form $v + su$ where $v = (1, 0, 0)^T$ and $u = (0, -1, 1)^T$. 

34
Question 2.2 The rank is 2.

Question 2.3 Show that the matrix \((x_1x_2x_3)\) has rank 3, using row operations.

Question 2.4 Call the vectors \(x_1, x_2, x_3\). To show linear independence, show that the matrix \(A = (x_1x_2x_3)\) has rank 3, using row operations. To express the fourth vector \(x_4\) as a linear combination of the first 3, solve \(Ax = x_4\). You should obtain the solution is \(x = (217/55, -18/55, 16/11)^T\), so

\[
\begin{pmatrix}
  4/7 \\
  -3
\end{pmatrix}
= \frac{217}{55}
\begin{pmatrix}
  2 \\
  -1
\end{pmatrix}
- \frac{18}{55}
\begin{pmatrix}
  3 \\
  4
\end{pmatrix}
+ \frac{16}{11}
\begin{pmatrix}
  -2 \\
  3
\end{pmatrix}.
\]

Question 2.5 There are many ways of solving this problem, and there are infinitely many possible \(x_i\)'s. We can solve it by trying to find a vector \(x_3\) such that the matrix \((x_1, x_2, x_3)\) has rank 3. Another approach is to take the \(2 \times 3\) matrix whose rows are \(x_1, x_2\) and reduce this to echelon form. The resulting echelon matrix will have two leading ones, so there will be one of the three columns, say column \(i\), in which neither row contains a leading one. Then all we need take for \(x_3\) is the vector \(e_i\), with a 1 in position \(i\) and 0 elsewhere.

Question 2.6 Let

\[
A = \begin{pmatrix}
  1 & 0 & 4 & 9 \\
  2 & -1 & -11 & 2 \\
  1 & 3 & 5 & 1 \\
  2 & 4 & -1 & -3
\end{pmatrix},
\]

the matrix with columns equal to the given vectors. If we only needed to show that the vectors were linearly dependent, it would suffice to show, using row operations, that \(\text{rank}(A) < 4\). But we’re asked for more: we have to find an explicit linear combination that equals the zero vector. So we need to find a solution of \(Ax = 0\). One solution is \(x = (5, -3, 1, -1)^T\). This means that

\[
5
\begin{pmatrix}
  1 \\
  2 \\
  1 \\
  2
\end{pmatrix}
- 3
\begin{pmatrix}
  0 \\
  -1 \\
  3 \\
  4
\end{pmatrix}
+ \frac{4}{5}
\begin{pmatrix}
  4 \\
  -11 \\
  5 \\
  -1
\end{pmatrix}
- \frac{9}{2}
\begin{pmatrix}
  9 \\
  2 \\
  1 \\
  -3
\end{pmatrix}
= \begin{pmatrix}
  0 \\
  0 \\
  0 \\
  0
\end{pmatrix}.
\]

Question 2.7 You need to show that for any \(\alpha \in \mathbb{R}\) and \(u, v \in S\), \(\alpha u \in S\) and \(u + v \in S\). Both are reasonably straightforward. Also, \(0 \in S\), so \(S \neq \emptyset\).

Question 2.8 Clearly, \(W \neq \emptyset\). We need to show that for any \(\alpha \in \mathbb{R}\) and \(f, g \in W\), \(\alpha f \in W\) and \(f + g \in W\). Suppose that

\[
f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n, \quad g(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n.
\]

Then

\[
(f + g)(x) = f(x) + g(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n,
\]

so \(f + g\) is also a polynomial of degree at most \(n\) and therefore \(f + g \in W\). Similarly,

\[
(\alpha f)(x) = \alpha f(x) = (\alpha a_0) + (\alpha a_1)x + \cdots + (\alpha a_n)x^n,
\]

so \(\alpha f \in W\) also.
Question 2.9 A basis for the null space is \{(1, -2, 1, 0)^T, (-1, 1, 0, 1)^T\}. There are many other possible answers.

Question 2.10 A basis for the column space is
\[
\begin{align*}
\begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix}, & \begin{bmatrix} -2 \\ 3 \\ 1 \\ 2 \end{bmatrix}, & \begin{bmatrix} 2 \\ -2 \\ 1 \\ 4 \end{bmatrix}
\end{align*}
\]

Question 2.11 The matrix \(A_T\) is
\[
\begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}.
\]

Question 2.12 A basis for the null space is \{(-1, -1, 1)^T\}, and a basis for the range is \{(1, 0, 1)^T, (0, 1, 1)^T\}. There are other possible answers.

Question 2.13 The coordinates are \([x]_B = (1, -1/5, 7/5)^T\).

Question 2.14 \(A_T[B, B']\) is
\[
\begin{bmatrix} 1 & 3 \\ 0 & 1 \\ -2 & -1 \end{bmatrix}.
\]