Linear Algebra (MA203/MA283): Lecture Notes
Semester 2 2024-2025

Dr Rachel Quinlan
School of Mathematical and Statistical Sciences
University of Galway

Contents

Chapter 1

Preview: what is linear algebra about?

1.1 The setup: vector spaces

This module is an introduction to the theory, methods, practices and applications of linear algebra. Like many other areas of mathematics, linear algebra is concerned with mathematical objects that have some properties in common, and with some functions between them that have some sort of good behaviour with respect to the properties that we are interested in. This is a fairly vague description of a way of organizing ideas that is very prominent throughout mathematics. It can be expressed in more detail within certain frameworks. For example, in calculus, we are generally interested in functions from (subsets of) $\mathbb{R}$ to $\mathbb{R}$ - that's what the subject deals with. We are not really interested in *all* such functions, but maybe in the continuous functions or the differentiable ones, basically those functions that are amenable to study via the objects and themes of calculus.

In linear algebra, the environments that we work in are called *vector spaces*, and the main (but not only) functions of interest are called *linear transformations*. The entire setup is closely linked to the algebra of *matrices*. The subject and its methods have extraordinary prevalence, importance and applicability in every area of the mathematical sciences. For example, linear algebra allows for a translation of many problems of geometry into a concrete or computational setting, thanks to the innovation of coordinate geometry in or before the 17th century. Because linear algebra is relatively well understood and well suited to computer implementations, methods for solving complex problems that cannot be handled analytically often involve a reduction to or approximation by a problem of linear algebra. Linear algebra is central to the study of statistics, particularly in any situation where multiple random variables need to be considered simultaneously. A robust knowledge of the basic principles and methods of linear algebra, which we will study in this course, is essential for the study of virtually every area of mathematics and its applications (further examples include analysis, abstract algebra, combinatorics, theoretical physics, the theory of graphs and networks and mathematical modelling in general). Despite its position as part of the basic fabric of mathematics, linear algebra also continues to be a subject of research activity in its own right. The University of Galway hosted the conference of the International Linear Algebra Society in 2022.

1.2 Systems of Linear Equations

Consider the equation

$$2x + y = 3.$$

This is an example of a *linear equation* in the variables x and y. As it stands, the statement "$2x + y = 3$" is neither true nor untrue : it is just a statement involving the abstract symbols x and y. However if we replace x and y with some particular pair of real numbers, the statement will become either true or false. For example

Putting $x = 1,\ y = 1$ gives $2x + y = 2(1) + (1) = 3$: True
$x = 1,\ y = 2$ gives $2x + y = 2(1) + (2) \neq 3$: False
$x = 0,\ y = 3$ gives $2x + y = 2(0) + (3) = 3$: True

Definition 1.2.1. *A pair* (x_0, y_0) *of real numbers is a* solution *to the equation* $2x+y = 3$ *if setting* $x = x_0$ *and* $y = y_0$ *makes the equation true; i,.e. if* $2x_0 + y_0 = 3$.

For example $(1, 1)$ and $(0, 3)$ are solutions - so are $(2, -1),\ (3, -3),\ (-1, 5)$ and $(-1/2, 4)$ (check these).

However $(1, 4)$ is not a solution since setting $x = 1,\ y = 4$ gives $2x + y = 2(1) + 4 \neq 3$.

The set of all solutions to the equation is called its *solution set*. In tis example, the solution set is a *line* in $\mathbb{R}^2$. In general, the solution set of the linear equation

$$a_1X_1 + \cdots + a_nX_n = b,$$

where b and the a_i are real numbers (and the a_i are not all zero) is an *affine hyperplane* in $\mathbb{R}^n$; geometrically it resembles a copy of $\mathbb{R}^{n-1}$ inside $\mathbb{R}^n$.

A collection of linear equations in the same n variables is referred to as a *linear system* or *system of linear equations*. The solution set of the system is the subset of $\mathbb{R}^n$ consisting of those elements that satify all of the equations in the system; it is the intersection of the solution sets of the individual equations. For small systems in few variables, like the one below, the solution set can be easily computed.

Example 1.2.2. *Solve the linear system*

$$\left.\begin{array}{rcrcrl} 2x & + & y & = & 3 & (A) \\ 4x & + & 3y & = & 4 & (B) \end{array}\right\}$$

Step 1: Multiply Equation (A) by 2 : $4x + 2y = 6$ (A2).
Any solution of (A2) is a solution of (A).

Step 2: Multiply Equation (B) by -1 : $-4x - 3y = -4$ (B2)
Any solution of (B2) is a solution of (B).

Step 3: Now add equations (A2) and (B2).

$$\begin{array}{rcrcr} 4x & + & 2y & = & 6 \\ -4x & - & 3y & = & -4 \\ \hline & & -y & = & 2 \end{array}$$

Step 4: So $y = -2$ and the value of y in any simultaneous solution of (A) and (B) is -2 : Now we can use (A) to find the value of x.

$$\begin{aligned} 2x + y = 3 \text{ and } y = -2 \Longrightarrow 2x + (-2) &= 3 \\ \Longrightarrow 2x &= 5 \\ \Longrightarrow x &= \frac{5}{2} \end{aligned}$$

So $x = 5/2,\ y = -2$ is the *unique* solution to this system of linear equations.

No surprises there, but this kind of "ad hoc" approach may not be so easy if we have a more complicated system, involving a greater number of variables, or more equations. We will devise a systematic approach, known as Gauss-Jordan elimination, for solving systems of linear equations.

1.2.1 Elementary Row Operations

Example 1.2.3. *Find all solutions of the following system :*

$$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 5 \\ 3x & + & y & - & 2z & = & 9 \\ -x & + & 4y & + & 2z & = & 0 \end{array}$$

In other (perhaps simpler) examples we were able to find solutions by simplifying the system (perhaps by eliminating certain variables) through operations of the following types :

1. We could multiply one equation by a non-zero constant.
2. We could add one equation to another (for example in the hope of eliminating a variable from the result).

A similar approach will work for Example 1.2.3 but with this and other harder examples it may not always be clear how to proceed. We now develop a new technique both for describing our system and for applying operations of the above types more systematically and with greater clarity.

Back to Example 1.2.3: We associate a *matrix* to our system of equations.

$$\begin{array}{rcrcrcl} x & + & 2y & - & z & = & 5 \\ 3x & + & y & - & 2z & = & 9 \\ -x & + & 4y & + & 2z & = & 0 \end{array}$$

$$\left(\begin{array}{rrrr} 1 & 2 & -1 & 5 \\ 3 & 1 & -2 & 9 \\ -1 & 4 & 2 & 0 \end{array}\right) \begin{array}{l} \text{Equation 1} \\ \text{Equation 2} \\ \text{Equation 3} \end{array}$$

Note that the first *row* of this matrix contains as its four entries the coefficients of the variables x, y, z, and the number appearing on the right-hand-side of Equation 1 of the system. Rows 2 and 3 correspond similarly to Equations 2 and 3. The *columns* of the matrix correspond (from left to right) to the variables $x,\ y,\ z$ and the right hand side of our system of equations.

Definition 1.2.4. *The above matrix is called the* augmented matrix *of the system of equations in Example 1.2.3.*

In solving systems of equations we are allowed to perform operations of the following types:

1. Multiply an equation by a non-zero constant.
2. Add one equation (or a non-zero constant multiple of one equation) to another equation.

These correspond to the following operations on the augmented matrix :

1. Multiply a *row* by a non-zero constant.
2. Add a multiple of one row to another row.
3. We also allow operations of the following type : Interchange two rows in the matrix (this only amounts to writing down the equations of the system in a different order).

Definition 1.2.5. *Operations of these three types are called* Elementary Row Operations (ERO's) *on a matrix.*

We now describe how ERO's on the augmented matrix can be used to solve the system of Example 1.2.3. The following table describes how an ERO is performed at each step to produce a new augmented matrix corresponding to a new (hopefully simpler) system.

ERO	Matrix	System
	$\begin{pmatrix} 1 & 2 & -1 & 5 \\ 3 & 1 & -2 & 9 \\ -1 & 4 & 2 & 0 \end{pmatrix}$	$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 5 \\ 3x & + & y & - & 2z & = & 9 \\ -x & + & 4y & + & 2z & = & 0 \end{array}$
1. $R3 \to R3 + R1$	$\begin{pmatrix} 1 & 2 & -1 & 5 \\ 3 & 1 & -2 & 9 \\ 0 & 6 & 1 & 5 \end{pmatrix}$	$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 5 \\ 3x & + & y & - & 2z & = & 9 \\ & & 6y & + & z & = & 5 \end{array}$
2. $R2 \to R2 - 3R1$	$\begin{pmatrix} 1 & 2 & -1 & 5 \\ 0 & -5 & 1 & -6 \\ 0 & 6 & 1 & 5 \end{pmatrix}$	$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 5 \\ & - & 5y & + & z & = & -6 \\ & & 6y & + & z & = & 5 \end{array}$
3. $R2 \to R2 + R3$	$\begin{pmatrix} 1 & 2 & -1 & 5 \\ 0 & 1 & 2 & -1 \\ 0 & 6 & 1 & 5 \end{pmatrix}$	$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 5 \\ & & y & + & 2z & = & -1 \\ & & 6y & + & z & = & 5 \end{array}$
4. $R3 \to R3 - 6R2$	$\begin{pmatrix} 1 & 2 & -1 & 5 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & -11 & 11 \end{pmatrix}$	$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 5 \\ & & y & + & 2z & = & -1 \\ & & & & -11z & = & 11 \end{array}$
5. $R3 \times \left(-\frac{1}{11}\right)$	$\begin{pmatrix} 1 & 2 & -1 & 5 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$	$\begin{array}{rcrcrcrl} x & + & 2y & - & z & = & 5 & (A) \\ & & y & + & 2z & = & -1 & (B) \\ & & & & z & = & -1 & (C) \end{array}$

We have produced a new system of equations. This is easily solved :

<u>Backsubstitution</u>

$$\begin{cases} (C) & z = -1 \\ (B) & y = -1 - 2z \quad \Longrightarrow \quad y = -1 - 2(-1) = 1 \\ (A) & x = 5 - 2y + z \quad \Longrightarrow \quad x = 5 - 2(1) + (-1) = 2 \end{cases}$$

Solution : $x = 2,\ y = 1, z = -1$

You should check that this is a solution of the original system. It is the only solution both of the final system and of the original one (and every intermediate one).

NOTE : The matrix obtained in Step 5 above is in *Row-Echelon Form*. This means :

1. The first non-zero entry in each row is a 1 (called a *Leading 1*).
2. If a column contains a leading 1, then every entry of the column below the leading 1 is a zero.
3. As we move downwards through the rows of the matrix, the leading 1's move from left to right.
4. Any rows consisting entirely of zeroes are grouped together at the bottom of the matrix.

NOTE : The process by which the augmented matrix of a system of equations is reduced to row-echelon form is called *Gaussian Elimination*. In Example 1.2.3 the solution of the system was found by Gaussian elimination with *Backsubstitution*.

General Strategy to Obtain a Row-Echelon Form

1. Get a 1 as the top left entry of the matrix.
2. Use this first leading 1 to "clear out" the rest of the first column, by adding suitable multiples of Row 1 to subsequent rows.
3. If column 2 contains non-zero entries (other than in the first row), use ERO's to get a 1 as the second entry of Row 2. After this step the matrix will look like the following (where the entries represented by stars may be anything):

$$\begin{pmatrix} 1 & * & * & \dots & \dots \\ 0 & 1 & \dots & \dots & \dots \\ 0 & * & \dots & \dots & \dots \\ 0 & * & \dots & \dots & \dots \\ \vdots & \vdots & & & \vdots \\ 0 & * & \dots & \dots & \dots \end{pmatrix}$$

4. Now use this second leading 1 to "clear out" the rest of column 2 (below Row 2) by adding suitable multiples of Row 2 to subsequent rows. After this step the matrix will look like the following :

$$\begin{pmatrix} 1 & * & * & \dots & \dots \\ 0 & 1 & * & \dots & \dots \\ 0 & 0 & * & \dots & \dots \\ 0 & 0 & * & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & * & \dots & \dots \end{pmatrix}$$

5. Now go to column 3. If it has non-zero entries (other than in the first two rows) get a 1 as the third entry of Row 3. Use this third leading 1 to clear out the rest of Column 3, then proceed to column 4. Continue until a row-echelon form is obtained.

Example 1.2.6. *Let* A *be the matrix*

$$\begin{pmatrix} 1 & -1 & -1 & 2 & 0 \\ 2 & 1 & -1 & 2 & 8 \\ 1 & -3 & 2 & 7 & 2 \end{pmatrix}$$

Reduce A *to row-echelon form.*

Solution:

1. Get a 1 as the first entry of Row 1. Done.

2. Use this first leading 1 to clear out column 1 as follows :

$$\begin{array}{rcl} & & \\ R2 & \to & R2 - 2R1 \\ R3 & \to & R3 - R1 \end{array} \quad \begin{pmatrix} 1 & -1 & -1 & 2 & 0 \\ 0 & 3 & 1 & -2 & 8 \\ 0 & -2 & 3 & 5 & 2 \end{pmatrix}$$

3. Get a leading 1 as the second entry of Row 2, for example as follows :

$$R2 \to R2 + R3 \quad \begin{pmatrix} 1 & -1 & -1 & 2 & 0 \\ 0 & 1 & 4 & 3 & 10 \\ 0 & -2 & 3 & 5 & 2 \end{pmatrix}$$

4. Use this leading 1 to clear out whatever appears below it in Column 2 :

$$R3 \to R3 + 2R2 \quad \begin{pmatrix} 1 & -1 & -1 & 2 & 0 \\ 0 & 1 & 4 & 3 & 10 \\ 0 & 0 & 11 & 11 & 22 \end{pmatrix}$$

5. Get a leading 1 in Row 3 :

$$R3 \times \frac{1}{11} \quad \begin{pmatrix} 1 & -1 & -1 & 2 & 0 \\ 0 & 1 & 4 & 3 & 10 \\ 0 & 0 & 1 & 1 & 2 \end{pmatrix}$$

This matrix is now in row-echelon form.
Remark Starting with a particular matrix, different sequences of ERO's can lead to different row-echelon forms. However, all have the same number of non-zero rows.

1.2.2 The Reduced Row-Echelon Form (RREF)

Definition 1.2.7. *A matrix is in* reduced row-echelon form (RREF) *if*

1. *It is in row-echelon form, and*
2. *If a particular column contains a leading 1, then* all *other entries of that column are zeroes.*

If we have a row-echelon form, we can use ERO's to obtain a reduced row-echelon form (using ERO's to obtain a RREF is called *Gauss-Jordan elimination*).

Example 1.2.8. *In Example 1.2.6, we obtained the following row-echelon form :*

$$\begin{pmatrix} 1 & -1 & -1 & 2 & 0 \\ 0 & 1 & 4 & 3 & 10 \\ 0 & 0 & 1 & 1 & 2 \end{pmatrix} \quad \text{(REF, not } \textit{reduced} \text{ REF)}$$

To get a RREF from this REF :

1. Look for the leading 1 in Row 2 - it is in column 2. Eliminate the non-zero entry *above* this leading 1 by adding a suitable multiple of Row 2 to Row 1.

$$R1 \to R1 + R2 \quad \begin{pmatrix} 1 & 0 & 3 & 5 & 10 \\ 0 & 1 & 4 & 3 & 10 \\ 0 & 0 & 1 & 1 & 2 \end{pmatrix}$$

2. Look for the leading 1 in Row 3 - it is in column 3. Eliminate the non-zero entries *above* this leading 1 by adding suitable multiples of Row 3 to Rows 1 and 2.

$$\begin{array}{rcl} & & \\ R1 & \to & R1 - 3R3 \\ R2 & \to & R2 - 4R3 \end{array} \quad \begin{pmatrix} 1 & 0 & 0 & 2 & 4 \\ 0 & 1 & 0 & -1 & 2 \\ 0 & 0 & 1 & 1 & 2 \end{pmatrix}$$

This matrix is in *reduced* row-echelon form. The technique outlined in this example will work in general to obtain a RREF from a REF: you should practise with similar examples.
Remark: Different sequences of ERO's on a matrix can lead to different row-echelon forms. However, the *reduced* row-echelon form of any matrix is unique.

1.2.3 Leading Variables and Free Variables

Example 1.2.9. *Find the general solution of the following system :*

$$\begin{array}{rcrcrcrcrr} x_1 & - & x_2 & - & x_3 & + & 2x_4 & = & 0 & \text{I} \\ 2x_1 & + & x_2 & - & x_3 & + & 2x_4 & = & 8 & \text{II} \\ x_1 & - & 3x_2 & + & 2x_3 & + & 7x_4 & = & 2 & \text{III} \end{array}$$

SOLUTION :

1. Write down the augmented matrix of the system :

$$\begin{array}{l} \text{Eqn I} \\ \text{Eqn II} \\ \text{Eqn III} \end{array} \quad \begin{pmatrix} 1 & -1 & -1 & 2 & 0 \\ 2 & 1 & -1 & 2 & 8 \\ 1 & -3 & 2 & 7 & 2 \end{pmatrix}$$
$$\quad x_1 \quad x_2 \quad x_3 \quad x_4$$

   Note : This is the matrix of Example 1.2.6

2. Use Gauss-Jordan elimination to find a reduced row-echelon form from this augmented matrix. We have already done this in Examples 1.2.6 and 1.2.8 :-

$$\text{RREF :} \quad \begin{pmatrix} 1 & 0 & 0 & 2 & 4 \\ 0 & 1 & 0 & -1 & 2 \\ 0 & 0 & 1 & 1 & 2 \end{pmatrix}$$
$$\quad x_1 \quad x_2 \quad x_3 \quad x_4$$

   This matrix corresponds to a new system of equations:

$$\begin{array}{rcrcrl} x_1 & + & 2x_4 & = & 4 & \text{(A)} \\ x_2 & - & x_4 & = & 2 & \text{(B)} \\ x_3 & + & x_4 & = & 2 & \text{(C)} \end{array}$$

   Remark : The RREF involves 3 leading 1's, one in each of the columns corresponding to the variables x_1, x_2 and x_3. The column corresponding to x_4 contains no leading 1.

   Definition 1.2.10. *The variables whose columns in the RREF contain leading 1's are called* leading variables*. A variable whose column in the RREF does not contain a leading 1 is called a* free variable*.*

   So in this example the leading variables are x_1, x_2 and x_3, and the variable x_4 is free. What does this distinction mean in terms of solutions of the system? The system corresponding to the RREF can be rewritten as follows :

$$\begin{array}{rcrcrl} x_1 & = & 4 & - & 2x_4 & \text{(A)} \\ x_2 & = & 2 & + & x_4 & \text{(B)} \\ x_3 & = & 2 & - & x_4 & \text{(C)} \end{array}$$

   i.e. this RREF tells us how the values of the leading variables x_1, x_2 and x_3 *depend* on that of the free variable x_4 in a solution of the system. In a solution, the free variable x_4 may assume the value of *any* real number. However, once a value for x_4 is chosen, values are immediately assigned to x_1, x_2 and x_3 by equations A, B and C above. For example

(a) Choosing $x_4 = 0$ gives $x_1 = 4 - 2(0) = 4,\ x_2 = 2 + (0) = 2,\ x_3 = 2 - (0) = 2$. Check that $x_1 = 4,\ x_2 = 2,\ x_3 = 2,\ x_4 = 0$ is a solution of the (original) system.

(b) Choosing $x_4 = 3$ gives $x_1 = 4 - 2(3) = -2,\ x_2 = 2 + (3) = 5,\ x_3 = 2 - (3) = -1$. Check that $x_1 = -2,\ x_2 = 5,\ x_3 = -1,\ x_4 = 3$ is a solution of the (original) system.

Different choices of value for x_4 will give different solutions of the system. The number of solutions is *infinite*.

The *general solution* is usually described by the following type of notation. We assign the *parameter* name t to the value of the variable x_4 in a solution (so t may assume any real number as its value). We then have

$$x_1 = 4 - 2t,\ x_2 = 2 + t,\ x_3 = 2 - t,\ x_4 = t;\ t \in \mathbb{R}$$

or

$$\textbf{General Solution :}\ (x_1, x_2, x_3, x_4) = (4 - 2t, 2 + t, 2 - t, t);\ t \in \mathbb{R}$$

This general solution describes the infinitely many solutions of the system : we get a *particular* solution by choosing a specific numerical value for t : this determines specific values for x_1, x_2, x_3 and x_4.

Example 1.2.11. *Solve the following system of linear equations :*

$$\begin{array}{rcrcrcrcrr} x_1 & - & x_2 & - & x_3 & + & 2x_4 & = & 0 & \text{I} \\ 2x_1 & + & x_2 & - & x_3 & + & 2x_4 & = & 8 & \text{II} \\ x_1 & - & 3x_2 & + & 2x_3 & + & 7x_4 & = & 2 & \text{III} \\ x_1 & - & x_2 & + & x_3 & - & x_4 & = & -6 & \text{IV} \end{array}$$

Remark : The first three equations of this system comprise the system of equations of Example 1.2.9. The problem becomes : Can we find a solution of the system of Example 1.2.9 which is in addition a solution of the equation $x_1 - x_2 + x_3 - x_4 = -6$?

SOLUTION We know that every simultaneous solution of the first three equations has the form

$$x_1 = 4 - 2t,\ x_2 = 2 + t,\ x_3 = 2 - t,\ x_4 = t,$$

where t can be any real number . Is there some choice of t for which the solution of the first three equations is also a solution of the fourth? i.e. for which

$$x_1 - x_2 + x_3 - x_4 = -6 \text{ i.e. } (4 - 2t) - (2 + t) + (2 - t) - t = -6$$

Solving for t gives

$$\begin{aligned} 4 - 5t &= -6 \\ \Longrightarrow -5t &= 10 \\ \Longrightarrow t &= 2 \end{aligned}$$

$$t = 2\ :\ x_1 = 4 - 2t = 4 - 2(2) = 0;\ x_2 = 2 + t = 2 + 2 = 4;\ x_3 = 2 - t = 2 - 2 = 0;\ x_4 = t = 2$$

SOLUTION : $x_1 = 0,\ x_2 = 4,\ x_3 = 0,\ x_4 = 2$ (or $(x_1, x_2, x_3, x_4) = (0, 4, 0, 2)$).
This is the *unique* solution to the system in Example 1.2.11.

REMARKS:

1. To solve the system of Example 1.2.11 directly (without 1.2.9) we would write down the augmented matrix :

$$\begin{pmatrix} 1 & -1 & -1 & 2 & 0 \\ 2 & 1 & -1 & 2 & 8 \\ 1 & -3 & 2 & 7 & 2 \\ 1 & -1 & 1 & -1 & -6 \end{pmatrix}$$

Check: Gauss-Jordan elimination gives the reduced row-echelon form :

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

which corresponds to the system

$$x_1 = 0;\ x_2 = 4;\ x_3 = 0;\ x_4 = 2$$

i.e. the unique solution is given exactly by the RREF. In this system, all four variables are leading variables. This is always the case for a system which has a unique solution : that each variable is a leading variable, i.e. corresponds in the RREF of the augmented matrix to a column which contains a leading 1.

2. The system of Example 1.2.9, consisting of Equations 1,2 and 3 of that in Example 1.2.11, had an infinite number of solutions. Adding the fourth equation in Example 1.2.11 pinpointed exactly one of these infinitely many solutions.

1.2.4 Consistent and Inconsistent Systems

Example 1.2.12. *Consider the following system :*

$$\begin{array}{rcrcrcr} 3x & + & 2y & - & 5z & = & 4 \\ x & + & y & - & 2z & = & 1 \\ 5x & + & 3y & - & 8z & = & 6 \end{array}$$

To find solutions, obtain a row-echelon form from the augmented matrix :

$$\begin{pmatrix} 3 & 2 & -5 & 4 \\ 1 & 1 & -2 & 1 \\ 5 & 3 & -8 & 6 \end{pmatrix} \quad \overset{R1 \leftrightarrow R2}{\longrightarrow} \quad \begin{pmatrix} 1 & 1 & -2 & 1 \\ 3 & 2 & -5 & 4 \\ 5 & 3 & -8 & 6 \end{pmatrix}$$

$$\begin{matrix} R2 \to R2 - 3R1 \\ \longrightarrow \\ R3 \to R3 - 5R1 \end{matrix} \quad \begin{pmatrix} 1 & 1 & -2 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & -2 & 2 & 1 \end{pmatrix} \quad \overset{R2 \times (-1)}{\longrightarrow} \quad \begin{pmatrix} 1 & 1 & -2 & 1 \\ 0 & 1 & -1 & -1 \\ 0 & -2 & 2 & 1 \end{pmatrix}$$

$$\overset{R3 \to R3 + 2R2}{\longrightarrow} \quad \begin{pmatrix} 1 & 1 & -2 & 1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad \overset{R3 \times (-1)}{\longrightarrow} \quad \underset{\text{(Row-Echelon Form)}}{\begin{pmatrix} 1 & 1 & -2 & 1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}}$$

The system of equations corresponding to this REF has as its third equation

$$0x + 0y + 0z = 1 \quad \text{i.e.} \quad 0 = 1$$

This equation clearly has no solutions - no assignment of numerical values to x, y and z will make the value of the expression $0x + 0y + 0z$ equal to anything but zero. Hence the system has no solutions.

Definition 1.2.13. *A system of linear equations is called* inconsistent *if it has no solutions. A system which has a solution is called* consistent.

If a system is inconsistent, a REF obtained from its augmented matrix will include a row of the form 0 0 0 ... 0 1, i.e. will have a leading 1 in its rightmost column. Such a row corresponds to an equation of the form $0x_1 + 0x_2 + \cdots + 0x_n = 1$, which certainly has no solution.

Example 1.2.14.

(a) Find the unique value of t for which the following system has a solution.

$$\begin{array}{rcrcrcrcl} -x_1 & & & + & x_3 & - & x_4 & = & 3 \\ 2x_1 & + & 2x_2 & - & x_3 & - & 7x_4 & = & 1 \\ 4x_1 & - & x_2 & - & 9x_3 & - & 5x_4 & = & t \\ 3x_1 & - & x_2 & - & 8x_3 & - & 6x_4 & = & 1 \end{array}$$

SOLUTION: First write down the augmented matrix and begin Gauss-Jordan elimination.

$$\begin{pmatrix} -1 & 0 & 1 & -1 & 3 \\ 2 & 2 & -1 & -7 & 1 \\ 4 & -1 & -9 & -5 & t \\ 3 & -1 & -8 & -6 & 1 \end{pmatrix} \begin{array}{c} R1 \times (-1) \\ \longrightarrow \end{array} \begin{pmatrix} 1 & 0 & -1 & 1 & -3 \\ 2 & 2 & -1 & -7 & 1 \\ 4 & -1 & -9 & -5 & t \\ 3 & -1 & -8 & -6 & 1 \end{pmatrix}$$

$$\begin{array}{c} R2 \to R2 - 2R1 \\ R3 \to R3 - 4R1 \\ \longrightarrow \\ R4 \to R4 - 3R1 \end{array} \begin{pmatrix} 1 & 0 & -1 & 1 & -3 \\ 0 & 2 & 1 & -9 & 7 \\ 0 & -1 & -5 & -9 & t+12 \\ 0 & -1 & -5 & -9 & 10 \end{pmatrix} \begin{array}{c} R3 \to R3 - R4 \\ \longrightarrow \end{array} \begin{pmatrix} 1 & 0 & -1 & 1 & -3 \\ 0 & 2 & 1 & -9 & 7 \\ 0 & 0 & 0 & 0 & t+2 \\ 0 & -1 & -5 & -9 & 10 \end{pmatrix}$$

From the third row of this matrix we can see that the system can be consistent only if $t + 2 = 0$. i.e. only if $t = -2$.

(b) Find the general solution of this system for this value of t.
SOLUTION: Set $t = -2$ and continue with the Gaussian elimination. We omit the third row, which consists fully of zeroes and carries no information.

$$\begin{pmatrix} 1 & 0 & -1 & 1 & -3 \\ 0 & 2 & 1 & -9 & 7 \\ 0 & -1 & -5 & -9 & 10 \end{pmatrix} \begin{array}{c} R4 \times (-1) \\ \longrightarrow \\ R3 \leftrightarrow R4 \end{array} \begin{pmatrix} 1 & 0 & -1 & 1 & -3 \\ 0 & 1 & 5 & 9 & -10 \\ 0 & 2 & 1 & -9 & 7 \end{pmatrix}$$

$$\begin{array}{c} R3 \to R3 - 2R2 \\ \longrightarrow \end{array} \begin{pmatrix} 1 & 0 & -1 & 1 & -3 \\ 0 & 1 & 5 & 9 & -10 \\ 0 & 0 & -9 & -27 & 27 \end{pmatrix} \begin{array}{c} R3 \times (-\frac{1}{9}) \\ \longrightarrow \end{array} \begin{pmatrix} 1 & 0 & -1 & 1 & -3 \\ 0 & 1 & 5 & 9 & -10 \\ 0 & 0 & 1 & 3 & -3 \end{pmatrix}$$

$$\begin{array}{c} R1 \to R1 + R3 \\ \longrightarrow \\ R2 \to R2 + 5R3 \end{array} \begin{pmatrix} 1 & 0 & 0 & 4 & -6 \\ 0 & 1 & 0 & -6 & 5 \\ 0 & 0 & 1 & 3 & -3 \end{pmatrix}$$

Having reached a reduced row-echelon form, we can see that the variables x_1, x_2 and x_3 are leading variables, and the variable x_4 is free. We have from the RREF

$$x_1 = -6 - 4x_4, \; x_2 = 5 + 6x_4, \; x_3 = -3 - 3x_4.$$

If we assign the parameter name s to the value of the free variable x_4 in a solution of the system, we can write the general solution as

$$(x_1, x_2, x_3, x_4) = (-6 - 4s, 5 + 6s, -3 - 3s, s), \; s \in \mathbb{R}.$$

Summary of Possible Outcomes when Solving a System of Linear Equations:

1. The system may be inconsistent. This happens if a REF obtained from the augmented matrix has a leading 1 in its rightmost column.

2. The system may be consistent. In this case one of the following occurs :

(a) There may be a unique solution. This will happen if all variables are leading variables, i.e. every column except the rightmost one in a REF obtained from the augmented matrix has a leading 1. In the case the *reduced* row-echelon form obtained from the augmented matrix will have the following form :

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 & * \\ 0 & 1 & 0 & \dots & 0 & * \\ 0 & 0 & 1 & \dots & 0 & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & * \end{pmatrix}$$

with possibly some additional rows full of zeroes at the bottom. The unique solution can be read from the right-hand column.

NOTE: If a system of equations has a unique solution, the number of equations must be at least equal to the number of variables (since the augmented matrix must have enough rows to accommodate a leading 1 for every variable).

(b) There may be infinitely many solutions. This happens if the system is consistent but at least one of the variables is free. In this case the number of leading 1s in the row echelon form is less than the number of variables in the system.

1.3 Gaussian Elimination and Matrix Algebra

We finish Chapter 1 with two observations about connections between the process of Gaussian (or Gauss-Jordan) elimination and the algebra of matrices as objects that can be added, multipied, inverted etc.

The first is that elementary row operations may themselves be interpreted as matrix multiplication exercises. We write I_m for the $m \times m$ identity matrix and $E_{i,j}$ for the matrix that has 1 in the (i, j)-position and zeros everywhere else. So for example $I_3 + 4E_{1,2}$ is the 3×3 matrix with entries 1 on the main diagonal, 4 in the $(1, 2)$ position, and zeros everywhere else.

$$I_3 + 4E_{1,2} = \begin{bmatrix} 1 & 4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Theorem 1.3.1. *Let A be a $m \times m$ matrix. Then elementary row operations on A amount to multiplying A on the left by $m \times m$ matrices, as follows:*

1. *Mutiplying Row i by the non-zero scalar α is equivalent to multiplying A on the left by the matrix $I_m + (\alpha - 1)E_{i,i}$, which has entries α in Position (i, i), 1 in all other positions on the main diagonal, and zeros in all off-diagonal positions.*

2. *Switching Rows i and k amounts to multiplying A on the left by the matrix $I_m + E_{i,k} + E_{k,i} - E_{i,i} - E_{k,i}$. This matrix has entries 1 in the (i, k) and (k, i) positions, and in the (j, j) position for all $j \notin \{i, k\}$, and zeros elsewhere. It has exactly one 1 in each row and column, and is otherwise full of zeros.*

3. *Adding $\alpha\times$ Row i to Row k amounts to multiplying A on the left by the matrix $I_m + \alpha E_{k,i}$, which has α in the (k, i) position, entries 1 on the main diagonal, and zeros elswhere.*

Here are a couple of examples.

1. $$\begin{bmatrix} 1 & 0 & 2 & 4 \\ 2 & 3 & -1 & 1 \\ 2 & 2 & 3 & 2 \end{bmatrix} \xrightarrow{R2 \to R2 - 2R1} \begin{bmatrix} 1 & 0 & 2 & 4 \\ 0 & 3 & -5 & -7 \\ 2 & 2 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 4 \\ 2 & 3 & -1 & 1 \\ 2 & 2 & 3 & 2 \end{bmatrix}$$

2. $$\begin{bmatrix} 1 & 0 & 2 & 4 \\ 2 & 3 & -1 & 1 \\ 2 & 2 & 3 & 2 \end{bmatrix} \xrightarrow{R1 \leftrightarrow R3} \begin{bmatrix} 2 & 2 & 3 & 2 \\ 2 & 3 & -1 & 1 \\ 1 & 0 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 4 \\ 2 & 3 & -1 & 1 \\ 2 & 2 & 3 & 2 \end{bmatrix}$$

Matrices of the three types described in Theorem 1.3.1 are sometimes referred to as *elementary matrices*. They are always invertible, and their inverses are also elementary matrices. The statement that every matrix can be reduced to RREF through a sequence of EROs is equivalent to saying that for every matrix A with m rows, there exists a $m \times m$ matrix B, which is a product of elementary matrices, with the property that BA is in RREF.

Exercise 1.3.2. *Write down the inverse of an elementary matrix of each of the three types, and show that it is also an elementary matrix.*
(Hint: Think about how to reverse an elementary row operation, with another elementary row operation).

Exercise 1.3.3. *Prove that every invertible matrix in $M_n(\mathbb{R})$ is a product of elementary matrices.*

The second point of this section is that not only can elementary row operations be interpreted as matrix products, but they can also be used for calculations in matrix algebra beyond the context of solving systems of linear equations. As an example of this, we note that Gauss-Jordan elimination can be used to calculate the inverse of a square matrices (and this is a much more efficient method than calculating cofactors as we often do for 3×3 matrices). Suppose that $A \in M_n(\mathbb{F})$, for some field $\mathbb{F}$. If A is invertible, let $v_1, v_2, \ldots, v_n$ be the columns of its inverse. Then

$$AA^{-1} = A \begin{bmatrix} | & | & \cdots & | \\ v_1 & v_2 & \ldots & v_n \\ | & | & \cdots & | \end{bmatrix} = A \begin{bmatrix} | & | & \cdots & | \\ Av_1 & Av_2 & \ldots & Av_n \\ | & | & \cdots & | \end{bmatrix} = I_n.$$

It follows that for each i, Av_i is the ith column of the identity matrix, which has 1 in position i and zeros elsewhere. This means that v_i is the solution of the linear system $Av_i = e_i$, where e_i is column i of the identity matrix, and the variables are the unknown entries of v_i. Since A is invertible, this system has the unique solution $v_i = A^{-1}e_i$, and this unique solution cn be found by applying Gauss-Jordan elimination to the augmented matrix of the system, which is $[A|e_i]$. We need to this for each column, but we can combine this into a single process by writing $e_1, e_2, \ldots, e_n$ as n distinct columns in the "right hand side" of a $n \times 2n$ augmented matrix whose coefficient matrix is A. If A is invertible, the RREF obtained from this augmented matrix has leading 1s in the first n columns, which form a copy of I_n, and the inverse of A is written in the last n columns of the RREF.

Example 1.3.4. *Find* A^{-1} *if* $A = \begin{bmatrix} 3 & 4 & -1 \\ 1 & 0 & 3 \\ 2 & 5 & -4 \end{bmatrix}$.

To calculate A^{-1}, We start with the 3×6 matrix

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

Reduce A' to RREF. If the RREF has I_3 in its first three columns, then columns 4,5,6 contain A^{-1}. If the RREF does not have leading 1s in its first three columns, we conclude that A is not invertible (more later on the justification for this). We proceed as follows.

$$\begin{bmatrix} 3 & 4 & -1 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 2 & 5 & -4 & 0 & 0 & 1 \end{bmatrix} \quad \overset{R1 \leftrightarrow R2}{\longrightarrow} \quad \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 3 & 4 & -1 & 1 & 0 & 0 \\ 2 & 5 & -4 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{array}{c} R2 \to R2 - 3R1 \\ \longrightarrow \\ R3 \to R3 - 2R1 \end{array} \quad \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 4 & -10 & 1 & -3 & 0 \\ 0 & 5 & -10 & 0 & -2 & 1 \end{bmatrix} \quad \overset{R3 \to R3 - R2}{\longrightarrow} \quad \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 4 & -10 & 1 & -3 & 0 \\ 0 & 1 & 0 & -1 & 1 & 1 \end{bmatrix}$$

$$\overset{R3 \leftrightarrow R2}{\longrightarrow} \quad \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 1 & 1 \\ 0 & 4 & -10 & 1 & -3 & 0 \end{bmatrix} \quad \overset{R3 \leftrightarrow R3 - 4R2}{\longrightarrow} \quad \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 1 & 1 \\ 0 & 0 & -10 & 5 & -7 & -4 \end{bmatrix}$$

$$\overset{R3 \times (-\frac{1}{10})}{\longrightarrow} \quad \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & -\frac{1}{2} & \frac{7}{10} & \frac{2}{5} \end{bmatrix} \quad \overset{R1 \to R1 - 3R3}{\longrightarrow} \quad \begin{bmatrix} 1 & 0 & 0 & \frac{3}{2} & -\frac{11}{10} & -\frac{6}{5} \\ 0 & 1 & 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & -\frac{1}{2} & \frac{7}{10} & \frac{2}{5} \end{bmatrix}$$

The above matrix is in RREF and its first three columns comprise I_3. We conclude that the matrix A^{-1} is written in the last three columns, i.e.

$$A^{-1} = \begin{bmatrix} \frac{3}{2} & -\frac{11}{10} & -\frac{6}{5} \\ -1 & 1 & 1 \\ -\frac{1}{2} & \frac{7}{10} & \frac{2}{5} \end{bmatrix}.$$

It is easily checked that $AA^{-1} = I_3$.

1.4 Review of Matrix Algebra

A $m \times n$ matrix over a field $\mathbb{F}$ is an an array of m rows and n columns, whose entries are elements of $\mathbb{F}$. We can take $\mathbb{F}$ to be the field of real numbers. The expression $m \times n$ is referred to as the *size* of a matrix (even though what it really describes is the *shape*). Two matrices can be added together if they have the same size; in this case their sum is obtained by just adding the entries in each position. The $m \times n$ zero matrix is the $m \times n$ matrix whose entries are all zeros. It is the *identity element* for addition of $m \times n$ matrices - this means that addition it to another $m \times n$ matrix has no effect. A matrix can be multiplied by a scalar; this means multiplying each of its entries by that scalar. With these operations of addition and scalar multiplication, the set of $m \times n$ matrices over a field $\mathbb{F}$ is a *vector space* over $\mathbb{F}$.
Notation: We use the notation $M_{m \times n}(\mathbb{F})$ for the vector space of all $m \times n$ matrices over $\mathbb{F}$. When $m = n$, we abbreviate this to $M_n(\mathbb{F})$.

Example 1.4.1. *In* $M_{2\times 3}(\mathbb{R})$,

$$2\begin{pmatrix} 1 & 0 & -1 \\ 2 & -5 & 1 \end{pmatrix} - 3\begin{pmatrix} 2 & 4 & -1 \\ 0 & 1 & -3 \end{pmatrix} = \begin{pmatrix} 2(1)-3(2) & 2(0)-3(4) & 2(-1)-3(-1) \\ 2(2)-3(0) & 2(-5)-3(1) & 2(-3)-3(-3) \end{pmatrix} = \begin{pmatrix} -4 & -12 & 1 \\ 4 & -13 & 3 \end{pmatrix}.$$

We can sometimes also *multiply* matrices, but the way to do this is not as obvious. We begin with a few definitions.

Definition 1.4.2. *Suppose that* $v_1, v_2, \ldots, v_k$ *are elements of a vector space* V *over a field* $\mathbb{F}$. *A* $\mathbb{F}$*-linear combination (or just linear combination) of* $v_1, \ldots, v_k$ *is an element of* V *that has the form* $a_1v_1 + a_2v_2 + \cdots + a_kv_k$, *wher the* a_i *are elements of* $\mathbb{F}$. *In this situation the* a_i *are called the* coefficients *in the linear combination.*

The example above shows a linear combination of two matrices in $M_{2\times 3}(\mathbb{R})$, with coefficients 2 and -3.

Definition 1.4.3. *A* column vector *is a matrix with one column. A* row vector *is a matrix with one row.*

Before defining matrix multiplication in general, we define the product of a matrix with a column vector (when that exists).

Definition 1.4.4. *Let* A *be a* $m \times n$ *matrix and let* v *be a column vector with* n *entries. Then the matrix-vector product* Av *is the column vector obtained by taking the linear combination of the columns of* A *whose coefficients are the entries of* v*. It is a column vector with* m *entries.*

Example 1.4.5. $\begin{pmatrix} 2 & 4 & -1 \\ 0 & 1 & -3 \end{pmatrix} \begin{pmatrix} 3 \\ -2 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 2 \\ 0 \end{pmatrix} + (-2) \begin{pmatrix} 4 \\ 1 \end{pmatrix} + 1 \begin{pmatrix} -1 \\ -3 \end{pmatrix} = \begin{pmatrix} -3 \\ -5 \end{pmatrix}$

Note that the product Av is only defined when the number of entries in a row of A (i.e. the number of columns in A) is equal to the number of entries in v. In the same way, if u is a row vector with m entries, and A is a $m \times n$ matrix, then the vector-matrix product uA is the linear combination of the rows of A whose coefficients are the entries of u. It is a row vector with n entries (a $1 \times n$ matrix).

Definition 1.4.6. *Let* A *and* B *be matrices of size* $m \times p$ *and* $p \times n$ *respectively. Write* $v_1, \ldots, v_n$ *for the columns of* B*. Then the product* AB *is the* $m \times n$ *matrices whose columns are* $Av_1, \ldots, Av_n$*.*

Remarks

1. This version of the definition of the matrix product AB emphasizes that we can think about the matrix B as being an arrangement of n column vectors placed side by side. This viewpoint can be quite useful, but maybe not as the only way to think about matrices. But in this situation it allows use to understand matrix multiplication as a straightforward extension of matrix-vector multiplication.

2. Exercise: write down an alternative version of Definition 1.4.6, that emphasises vector-matrix products of the rows of A with the matrix B.

3. If the number of entries in a row of A (the number of columns of A) is not equal to the number of entries in a column of B, then the product AB is not defined.

Matrix products are often presented and explained just in terms of their individual entries. This viewpoint is sometimes convenient and it is quite standard, and it gives us an opportunity to introduce some notation that essential for linear algebra. Suppose that A is a $m \times p$ matrix and B is a $p \times n$ matrix, with entries in a field $\mathbb{F}$. The rows of A are labelled Row 1 through Row m, from top to bottom, and the columns of A are labelled Column 1 through Column p, from left to right (similar story for B). The entry in Row i and Column j of A is denoted A_{ij}. So A_{11} is the entry in the upper left corner of A. Now AB is the product of a $m \times p$ and a $p \times m$ matrix: it is a $m \times n$ matrix. According to Definition 1.4.6, the entry in the the (i, j) position of AB (i.e. Row i and Column j) is the ith entry of the vector Av_j, where the vector v_j is Column j of B. Again according to Definition 1.4.6, this is the ith entry of the linear combination of the columns of A with coefficients from the jth column of B. This is the linear combination of the ith entries of the columns of A (i.e. the entries of Row i of A, with coefficients from Column j of B). It is given by

$$(AB)_{ij} = A_{i1}B_{ij} + A_{i2}B_{2j} + \cdots + A_{ip}B_{pj} = \sum_{k=1}^{p} A_{ik}B_{kj}.$$

It is worth taking some time to get used to the notation in the above line if it is not already familiar.

We note that the expression for $(AB)_{ij}$ above involves the *scalar product* of two vectors with p entries. For a field $\mathbb{F}$, we write $\mathbb{F}^p$ for the vector space of all vectors with p entries from $\mathbb{F}$. Sometimes we will need to specify whether we mean row vectors or column vectors, but for now we will cheat and allow every ordered list of p elements of $\mathbb{F}$ to be considered as an element of $\mathbb{F}^p$, regardless of how it is written. For example we might consider elements of $\mathbb{R}^3$ to be written as coordinates of a point, like $(1, 1, 3)$, or as column vectors with three real entries.

Definition 1.4.7. *Let* $u = (a_1, \ldots, a_p)$ *and* $v = (b_1, \ldots, b_p)$ *be vectors in* $\mathbb{F}^p$*. Then the* ordinary scalar product *or* dot product *of* u *and* v *is the element of* $\mathbb{F}$ *defined by*

$$u \cdot v = a_1 b_1 + a_2 b_2 + \cdots + a_p b_p = \sum_{k=1}^{p} a_k b_k.$$

If $u \cdot v = 0$, we say that u and v are *orthogonal* with respect to the scalar product. If $\mathbb{F} = \mathbb{R}$, this means that the vectors u and v are perpendicular in Euclidean space.

We may now observe that if A and B are respectively a $m \times p$ and a $p \times n$ matrix, then the entry in the (i, j)-position of the product AB is the scalar product of Row i of A and Column j of B, both regarded as vectors in $\mathbb{F}^p$. The product AB itself is a table of values of scalar products of Rows of A with Columns of B. If we write $u_1, \dots, u_m$ for the rows of A (vectors in $\mathbb{F}^p$) and $v_1, \dots, v_n$ for the columns of B (vectors in $\mathbb{F}^p$), then

$$AB = \begin{pmatrix} u_1 \cdot v_1 & u_1 \cdot v_2 & \dots & u_1 \cdot v_n \\ u_2 \cdot v_1 & u_2 \cdot v_2 & \dots & u_2 \cdot v_n \\ \vdots & \vdots & & \vdots \\ u_m \cdot v_1 & u_m \cdot v_2 & \dots & u_m \cdot v_n \end{pmatrix}$$

1.4.1 Two ways to think about a matrix

The definition of matrix multiplication can look a bit obscure, if it is presented purely in terms of how the entries of A and B are combined to produce the entries of AB. It does make sense however, even in very practical contents as in the following example.

Example 1.4.8. One way for a matrix to arise is as a table of data from some "real" (i.e. not just mathematical) process. Remarkably, the operations of matrix algebra can have a meaning even in this context. As an example, let A be the 3×3 matrix formed by the table that gives the numbers of first year Humanities (H), Engineering (E) and Science (S) students in first year at Eigen University, in 2015, 2016 and 2017.

	H	E	S
2015	50	100	70
2016	60	80	80
2017	80	70	70

$$A = \begin{pmatrix} 50 & 100 & 70 \\ 60 & 80 & 80 \\ 80 & 70 & 70 \end{pmatrix}$$

Every fiirst year student at Eigen University takes either Linear Algebra (LA) or Calculus (C) or both. The table below shows the numbers of ECTS credits completed annually in each, by students in each of the three subject areas.

	LA	C
H	10	0
E	15	15
S	20	10

$$B = \begin{pmatrix} 10 & 0 \\ 15 & 15 \\ 20 & 10 \end{pmatrix}$$

Now look at the meaning of the entries of the product AB.

$$\begin{aligned} AB &= \begin{pmatrix} 50 & 100 & 70 \\ 60 & 80 & 80 \\ 80 & 70 & 70 \end{pmatrix} \begin{pmatrix} 10 & 0 \\ 15 & 15 \\ 20 & 10 \end{pmatrix} = \begin{pmatrix} 50(10) + 100(15) + 70(20) & 50(0) + 100(15) + 70(10) \\ 60(10) + 80(15) + 80(20) & 60(0) + 80(15) + 80(10) \\ 80(10) + 70(15) + 70(20) & 80(0) + 70(15) + 70(10) \end{pmatrix} \\ &= \begin{pmatrix} 3400 & 2200 \\ 3400 & 2000 \\ 3250 & 1750 \end{pmatrix}. \end{aligned}$$

The entries in the first column of AB are the total numbers of linear algebra credits taken by first year students in 2015, 2016 and 2017. In the second column are the total numbers of calculus credits in each of the three years. The matrix product AB represents the following table of data

	LA credits	C credits
2015	3400	2200
2016	3400	2000
2017	3250	1750

Another way to interpret matrix multiplication is in terms of *linear transformations*, which are the primary functions between vector spaces that are of interest in linear algebra. For now we will stick to linear transformations between spaces of real column vectors.

Definition 1.4.9. *Let* m *and* n *be positive integers. A linear tranformation* T *from* $\mathbb{R}^n$ *to* $\mathbb{R}^m$ *is a function* $T : \mathbb{R}^n \to \mathbb{R}^m$ *that satisfies*

- $T(u + v) = T(u) + T(v)$*, and*
- $T(\lambda v) = \lambda T(v)$*,*

for all u *and* v *in* $\mathbb{R}^n$*, and all scalars* $\lambda \in \mathbb{R}$*.*

Suppose that $T : \mathbb{R}^3 \to \mathbb{R}^2$ is a linear transformation. Then we can calculate the image under T of any vector $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$, if we know the images under T of the *standard basis vectors* $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. From the definition, we have

$$T\begin{pmatrix} a \\ b \\ c \end{pmatrix} = aT\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + bT\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + cT\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = A\begin{pmatrix} a \\ b \\ c \end{pmatrix},$$

where A is the 2×3 matrix that has the images of the three standard basis vectors as its three columns.

Example 1.4.10. *Suppose that the images of the three standard basis vectors under* T *are respectively* $\binom{1}{2}$*,* $\binom{1}{4}$ *and* $\binom{-2}{3}$*. Then the matrix* A *of* T *is*

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

For any vector $v \in \mathbb{R}^3$*, its image under* T *is the matrix-vector product* Av*. For example*

$$T\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & -2 \\ 2 & 4 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} = \begin{pmatrix} -3 \\ 8 \end{pmatrix}$$

The matrix A may be considered as a representation of the linear transformation T. Now suppose that $S : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation whose matrix is $B = \begin{pmatrix} 1 & -1 \\ 2 & 0 \end{pmatrix}$. This means that the images under S of $\binom{1}{0}$ and $\binom{0}{1}$ are respectively the two columns of S. Now the composition $S \circ T$ is a linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$, so is is represented by a matrix. How dos this matrix depend on A and B. To answer this, think about calculating the image of a vector v under the composition $S \circ T$ (S *after* T).

$$S \circ T(v) = S(Tv) = S(Av) = B(Av) = (BA)v. \tag{1.1}$$

This is saying that the matrix of the transformation $S \circ T$ is the matrix product BA, where B and A are respectively the matrices of S and T. Thus matrix multiplcation may be interpreted as corresponding to composition of linear transformations.

Two things to note about 1.1:

1. It uses the fact that matrix multiplication is *associative*, i.e. $(AB)C = A(BC)$, whenever A, B, C are matrices for which the products AB and BC are defined. It is true but not entirely obvious that this property holds. Something to think about.

2. We have also used the fact that the composition of two linear transformations (when it is defined) is a linear transformation. If $f : \mathbb{R}^n \to R^p$ and $g : \mathbb{R}^q \to \mathbb{R}^m$ are linear transformations, then the composition $f \circ g$ is defined only if $p = q$, and in this case $f \circ g$ is a linear transformation from R^n to $\mathbb{R}^m$. This is equivalent to the statment that the product of a $m \times q$ with a $p \times n$ matrix is defined only if $p = q$, and in this case it is a $m \times n$ matrix.

1.4.2 Some more concepts from matrix algebra

This short section notes some objects and notation that we will need throughout the course.
The $n \times n$ *Identity Matrix*
For a positive integer n, the $n \times n$ identity matrix, denoted I_n, is the $n \times n$ matrix whose entries in the $(1,1), (2,2), \ldots, (n,n)$ positions (the positions on the main diagonal) are all 1, and whose entries in all other positions (all off-diagonal positions) are 0. For example

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The special property that I_n has is that is an *identity element* or *neutral element* for matrix multplication. Multiplying another matrix by it has no effect. This means

- If A is any matrix with n rows, then $I_nA = A$, and
- If B is any matrix with n columns, then $BI_n = B$.
- In particular, if C is a $n \times n$ matrix, then $CI_n = I_nC = C$.

Exercise 1.4.11. *Using our intepretations of matrix multiplication so far, explain why the matrix* I_n *has this neutral property. For example, use Definition 1.4.4 to describe what happens when a column vector with* n *entries is multiplied on the left by* I_n.
What is the linear transformation that is represented by I_n*? Why is it that composing this linear transformation with any other has no effect?*

The Inverse of a Matrix
Let A be a square matrix of size $n \times n$. If there exists a $n \times n$ matrix B for which $AB = I_n$ and $BA = I_n$, then A and B are called *inverses* (or *multiplicative inverses*) of each other. If it does not already have another name, the inverse of A is denoted A^{-1}. The relationship between A and A^1 resembles that of two rational numbers that are reciprocals of each other, such as $\frac{5}{31}$ and $\frac{31}{5}$. Their product is the identity element for multiplication (1 in the case of the rational numbers) and so multiplying by one of them reverses the work of multiplying by the other. When applying this general principle in the case of matrices, we need to remember that matrix multiplication is not commutative.

Example 1.4.12. *In* $M_2(\mathbb{Q})$*, the matrices* $\begin{pmatrix} 3 & 2 \\ -5 & -4 \end{pmatrix}$ *and* $\begin{pmatrix} 2 & 1 \\ -\frac{5}{2} & -\frac{3}{2} \end{pmatrix}$ *are inverses of each other.*

Not every square matrix has an inverse. For example the 2×2 matrix $\begin{pmatrix} 3 & 2 \\ -6 & -4 \end{pmatrix}$ does not. (Challenge: prove this without using any knowledge about the determinant of a 2×2 matrix).

Exercise 1.4.13. *Prove that a square matrix can have only one inverse.*
(Hint: If both B *and* C *are inverses of the square matrix* A*, think about the product* BAC*.)*

Another important ingredient of matrix algebra that we will need is the *determinant* of a square matrix. The determinant of a matrix in $M_n(\mathbb{F})$ is an element of $\mathbb{F}$ that is defined in a complicated way in terms of the matrix entries (it is not too bad if $n = 2$ or $n = 3$, but in general it is complicated to describe and to calculate). A square matrix has an inverse if and only if its determinant is not zero. We will define the determinant later.

The Transpose

Definition 1.4.14. *The transpose of the* $m \times n$ *matrix* A*, which is denoted* A^T*, is defined to be the* $n \times m$ *matrix which has the entries of Row 1 of* A *in its first column, the entries of Row 2 of* A *in its second column, and so on.*