1 Introduction

You are encouraged to look through appendix A before you start reading these notes.

Recall that all pairs (x, y) of real numbers are regarded as points in the xy-plane, where the set of all such points is denoted by

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}.$$

Here we will interpret the pair (x, y) in two ways: sometimes as the point (x, y) in the plane just as before, which we will call the *position* (x, y); other times as the *directed line segment* taking us from the *origin* $(0, 0)$ to the point (x, y), which we call the *vector* (x, y).

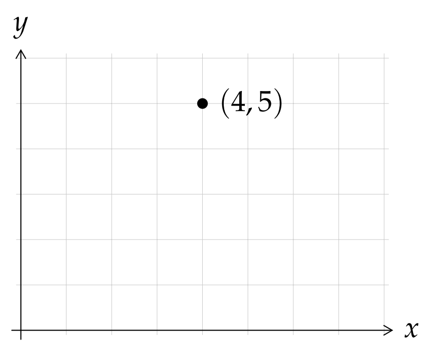

FIGURE 1: The position $(4, 5)$

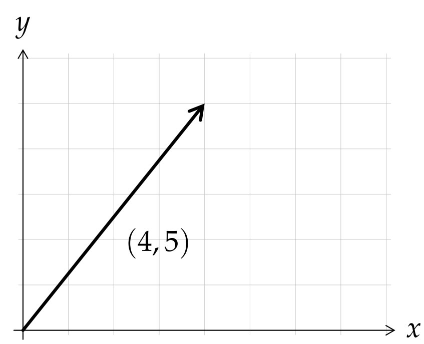

FIGURE 2: The vector $(4, 5)$

The distinction between the two interpretations is rarely important, and whenever the distinction is important, it is often clear from the context.

Note that vectors which are translated in the plane (that is, vectors which are moved so that their tails do not sit at the origin $(0, 0)$) correspond to the same pair of coordinates (x, y), since what the pair of numbers represent in this case is the *displacement* from the tip of the arrow to its head. Thus if a vector is translated, we treat the tip as the "new origin", and read off the coordinates at the head of the arrow, thus obtaining the same pair (x, y).

These ideas easily extend to the ordered triples (x, y, z) of real numbers, corresponding to points or vectors in three dimensional space

$$\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}.$$

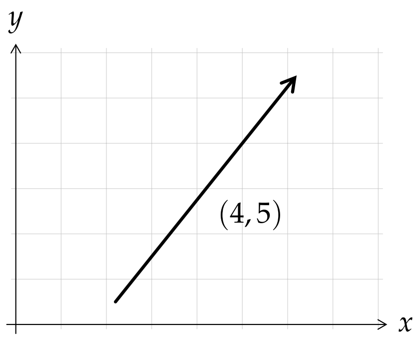

FIGURE 3: Still the vector $(4, 5)$

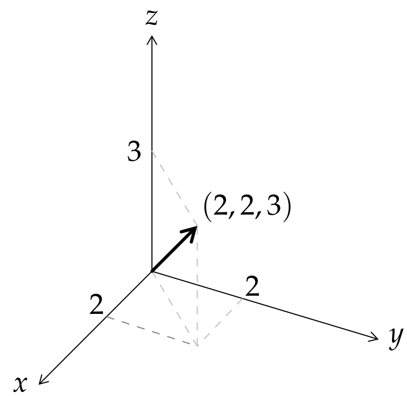

FIGURE 4: A vector in $\mathbb{R}^3$

Nothing stops us from considering the set $\mathbb{R}^n = \prod_{i=1}^n \mathbb{R}$ of ordered n-tuples. Although there is no geometric meaning for $n > 3$, it is convenient to use geometric *language*. Thus, we still call these tuples *points* or *vectors*, its entries are called *coordinates* or *components*, and the set as a whole we call *n-dimensional (Euclidean) space*.

1.1 Vector Operations

Here we introduce some operations on vectors. We denote vectors using single letters in bold typeface, and their coordinates are denoted using the same letter with corresponding subscripts. Thus we write

$$\boldsymbol{x} = (x_1, x_2, \dots, x_n)$$

for example. We sometimes write $\overrightarrow{O\boldsymbol{x}}$ when we want to emphasise that we consider $\boldsymbol{x}$ to be a directed line segment whose tail sits at the origin (as in figure 2). In this case, $\overrightarrow{O\boldsymbol{x}}$ is called the *position vector* of $\boldsymbol{x}$. In $\mathbb{R}^2$ and $\mathbb{R}^3$, we use the letters x, y, z to avoid subscripts, so we write $\boldsymbol{u} = (x, y) \in \mathbb{R}^2$ and $\boldsymbol{v} = (x, y, z) \in \mathbb{R}^3$ for example. In writing, you are encouraged to underline vectors to distinguish them from numbers, e.g., writing $\underset{\sim}{v}$, for $\boldsymbol{v}$.

Notation. We adopt the notation

$$\boldsymbol{v} = (v_i)_n$$

or simply $\boldsymbol{v} = (v_i)$ to stand for $\boldsymbol{v} = (v_1, v_2, \dots, v_n)$, where v_i is denoting the general ith component of the vector $\boldsymbol{v}$.

Definition 1.1 (Vector Addition)**.** Let $\boldsymbol{u} = (u_i)$ and $\boldsymbol{v} = (v_i)$ be two vectors in $\mathbb{R}^n$. Then the *sum* $\boldsymbol{u} + \boldsymbol{v}$ is defined by

$$\boldsymbol{u} + \boldsymbol{v} := (u_i + v_i).$$

Example 1.2. In $\mathbb{R}^3$, if $\boldsymbol{u} = (u_1, u_2, u_3)$ and $\boldsymbol{v} = (v_1, v_2, v_3)$, then

$$\boldsymbol{u} + \boldsymbol{v} = (u_i + v_i) = (u_1 + v_1, u_2 + v_2, u_3 + v_3).$$

Remark 1.3. Observe that the vector addition $\boldsymbol{u} + \boldsymbol{v}$ corresponds to the position obtained when translating the vector $\boldsymbol{v}$ such that its tail is at the head of the vector $\boldsymbol{u}$, or vice-versa; as shown in figure 5. This is a consequence of the fact that we think of vectors as representing only *relative displacement*, and not position. Think about it this way: first we travel from the tip of the vector $\boldsymbol{u}$ to the head, and then treating the head as if it were the "new origin", we travel along the vector $\boldsymbol{v}$. This is known as *the parallelogram law*.

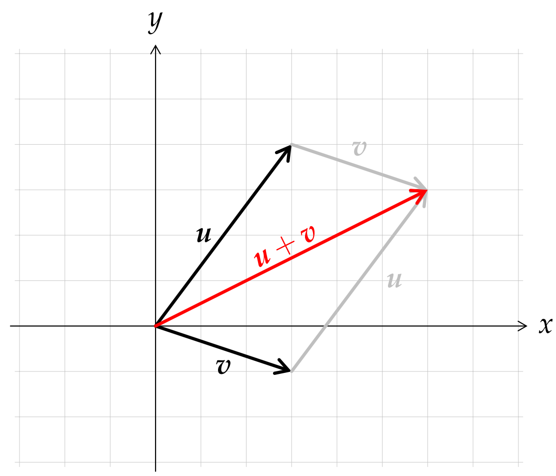

FIGURE 5: Illustration of the parallelogram law in $\mathbb{R}^2$

Definition 1.4 (Scalar Multiplication)**.** Let $\lambda \in \mathbb{R}$, and let $\boldsymbol{v} = (v_i)$ be a vector in $\mathbb{R}^n$. Then the *scalar multiplication* of $\boldsymbol{v}$ by λ, denoted $\lambda\boldsymbol{v}$, is the vector given by

$$\lambda\boldsymbol{v} = (\lambda v_i).$$

Example 1.5. In $\mathbb{R}^3$, if $\boldsymbol{u} = (x, y, z)$ then

$$\lambda\boldsymbol{u} = (\lambda x, \lambda y, \lambda z).$$

Remark 1.6. The reason we call this operation *scalar* multiplication is that the result of $\lambda \boldsymbol{v}$ is a *scaled* version of $\boldsymbol{v}$ by a factor of λ. When $\lambda < 0$, then the direction of $\boldsymbol{v}$ is reversed. In particular, $-1\boldsymbol{v}$, which we denote by $-\boldsymbol{v}$, corresponds to the vector with the arrow head and tail interchanged.

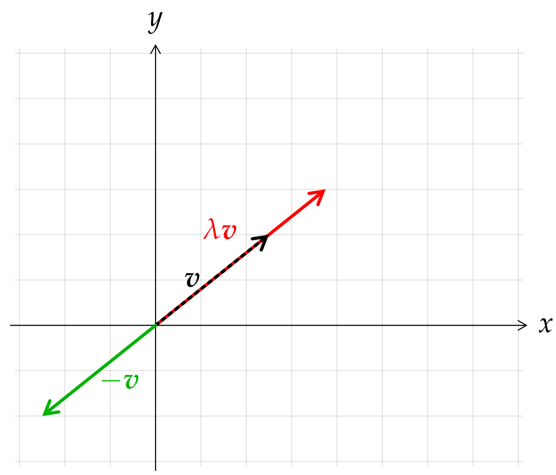

FIGURE 6: Illustration of scaling in $\mathbb{R}^2$

As a consequence of this scaling behaviour, we call single real numbers *scalars* instead of numbers throughout. Thus the entries in a vector are scalars, for example.

Notation. As mentioned in remark 1.6, we denote $-1\boldsymbol{v}$ by $-\boldsymbol{v}$, and we also introduce the *difference* between two vectors, denoted $\boldsymbol{u} - \boldsymbol{v}$, defined by

$$\boldsymbol{u} - \boldsymbol{v} = \boldsymbol{u} + (-\boldsymbol{v}).$$

Remark 1.7. A *relative vector* is a vector which takes us from a position $\boldsymbol{a}$ to a position $\boldsymbol{b}$, that is, another vector $\boldsymbol{v}$ such that $\boldsymbol{a} + \boldsymbol{v} = \boldsymbol{b}$. This vector $\boldsymbol{v}$ is given by $\boldsymbol{b} - \boldsymbol{a}$, as illustrated in figure 7.

Sometimes positions are denoted using upper case letters such as A or B. In this case, the vector from A to B is denoted by $\overrightarrow{AB}$, thus

$$\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA}.$$

Example 1.8. The vector from position $\boldsymbol{a} = (1, 3, 2)$ to $\boldsymbol{b} = (-1, 0, 1)$ is given by $\boldsymbol{b} - \boldsymbol{a} = (-1, 0, 1) - (1, 3, 2) = (-2, -3, -1)$. Indeed, if we add

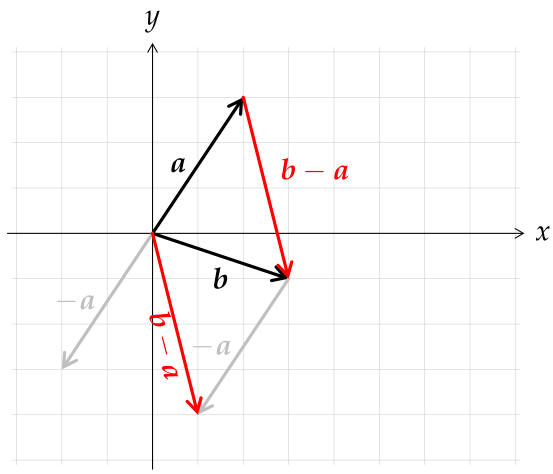

FIGURE 7: Relative vector from $\boldsymbol{a}$ to $\boldsymbol{b}$ in $\mathbb{R}^2$

$(-2, -3, -1)$ to $\boldsymbol{a}$, we get

$$\boldsymbol{a} + (-2, -3, -1) = (1, 3, 2) + (-2, -3, -1) = (-1, 0, 1) = \boldsymbol{b},$$

so $(-2, -3, -1)$ "takes us" from $\boldsymbol{a}$ to $\boldsymbol{b}$, as expected.

Definition 1.9 (Zero vector)**.** The vector $\mathbf{0} = (0) = (0, \ldots, 0)$ is called the *zero vector* or the *origin*.

Note. $\mathbf{0} \neq 0$. One is a vector with n entries, the other is a scalar.

Theorem 1.10 (Vector space properties in $\mathbb{R}^n$)**.** *Let $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ be three vectors in $\mathbb{R}^n$, and let $\lambda, \mu \in \mathbb{R}$ be scalars. Then the following properties hold:*

I) $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$

II) $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$

III) $\boldsymbol{u} + \mathbf{0} = \boldsymbol{u}$

IV) $\boldsymbol{v} + (-\boldsymbol{v}) = \mathbf{0}$

V) $\lambda(\mu\boldsymbol{v}) = (\lambda\mu)\boldsymbol{v}$

VI) $1\boldsymbol{v} = \boldsymbol{v}$

VII) $\lambda(\boldsymbol{u} + \boldsymbol{v}) = \lambda\boldsymbol{u} + \lambda\boldsymbol{v}$

VIII) $0\boldsymbol{v} = \mathbf{0}$

IX) $(\lambda + \mu)\boldsymbol{v} = \lambda\boldsymbol{v} + \mu\boldsymbol{v}$

Proof. These results all easily follow from the definitions, and properties of real numbers, e.g. for I, we have

$$\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (u_i)_n + ((v_i) + (w_i))_n$$

$$
\begin{aligned}
&= (u_i)_n + (v_i + w_i)_n && \text{(by definition 1.1)} \\
&= (u_i + (v_i + w_i))_n && \text{(by definition 1.1)} \\
&= ((u_i + v_i) + w_i)_n && \text{(by associativity of addition in } \mathbb{R}) \\
&= (u_i + v_i)_n + (w_i)_n && \text{(by definition 1.1)} \\
&= ((u_i)_n + (v_i)_n) + (w_i)_n && \text{(by definition 1.1)} \\
&= (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w},
\end{aligned}
$$

as required. Similarly for VIII, we have

$$0\boldsymbol{v} = (0v_i) = (0) = \mathbf{0}.$$

The proofs of the remaining properties are left as an exercise. □

Exercise 1.11. Try to visualise each of the "laws" in theorem 1.10 in terms of scaling and translation, as we illustrated in the various figures (figures 5 to 7). Construct figures which show the equalities of each. Then, provide a proof for each of them.

Definition 1.12 (Basic Unit Vectors)**.** Let δ_{ik}, known as the *Kronecker delta*, be defined by

$$\delta_{ik} = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{otherwise.} \end{cases}$$

Then we define the *basic unit vectors* $\boldsymbol{e}_k$ for $k = 1, \ldots, n$ in $\mathbb{R}^n$ by

$$\boldsymbol{e}_k = (\delta_{ik})_n,$$

that is, $\boldsymbol{e}_k$ has a k entry in the ith position, and zeros everywhere else:

$$
\begin{aligned}
\boldsymbol{e}_1 &= (1, 0, 0, \ldots, 0) \\
\boldsymbol{e}_2 &= (0, 1, 0, \ldots, 0) \\
&\vdots \\
\boldsymbol{e}_n &= (0, 0, 0, \ldots, 1).
\end{aligned}
$$

In $\mathbb{R}^2$, we denote $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ by $\boldsymbol{i}$ and $\boldsymbol{j}$, so $\boldsymbol{i} = (1, 0)$ and $\boldsymbol{j} = (0, 1)$. Similarly in $\mathbb{R}^3$, we denote $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ by $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$, so $\boldsymbol{i} = (1, 0, 0)$, $\boldsymbol{j} = (0, 1, 0)$, and $\boldsymbol{k} = (0, 0, 1)$.

Definition 1.13 (Linear combination of vectors)**.** Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$, and let $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ be scalars. Then the vector

$$\boldsymbol{v} = \sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i = \lambda_1 \boldsymbol{v}_1 + \lambda_2 \boldsymbol{v}_2 + \cdots + \lambda_k \boldsymbol{v}_k$$

is said to be a *linear combination* of the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$.

Example 1.14. Let $\boldsymbol{u} = (1, 2, 3)$, $\boldsymbol{v} = (3, 2, -1)$ and $\boldsymbol{w} = (1, 0, 7)$ be vectors in $\mathbb{R}^3$. Then the vectors

$$\boldsymbol{u} + 3\boldsymbol{v} = \boldsymbol{u} + 3\boldsymbol{v} + 0\boldsymbol{w} = (10, 8, 0) \qquad \text{and} \qquad 3\boldsymbol{u} + \boldsymbol{v} - 5\boldsymbol{w} = (1, 8, -27)$$

are examples of linear combinations of $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$.

Theorem 1.15. *Every vector $\boldsymbol{v} \in \mathbb{R}^n$ can be written as a linear combination of the vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$.*

Proof. Let $\boldsymbol{v} = (v_i)$ be any vector in $\mathbb{R}^n$. Then

$$\begin{aligned}\boldsymbol{v} &= (v_1, v_2, \ldots, v_n) \\ &= (v_1, 0, \ldots, 0) + (0, v_2, \ldots, 0) + \cdots + (0, 0, \ldots, v_n) \\ &= v_1(1, 0, \ldots, 0) + v_2(0, 1, \ldots, 0) + \cdots + v_n(0, 0, \ldots, 1) \\ &= v_1\boldsymbol{e}_1 + v_2\boldsymbol{e}_2 + \cdots + v_n\boldsymbol{e}_n = \sum_{k=1}^{n} v_k \boldsymbol{e}_k,\end{aligned}$$

as required. □

Remark 1.16. In particular, any vector $\boldsymbol{u} = (a, b) \in \mathbb{R}^2$ can be expressed in terms of $\boldsymbol{i}$ and $\boldsymbol{j}$ as

$$a\boldsymbol{i} + b\boldsymbol{j},$$

and similarly any vector $\boldsymbol{v} = (a, b, c) \in \mathbb{R}^3$ can be written as $a\boldsymbol{i} + b\boldsymbol{j} + c\boldsymbol{k}$.

1.2 Distances and Angles

So far we have encoded positions in $\mathbb{R}^n$, as well as operations we can carry out when interpreting them as directed line segments (vectors), but we have not yet described the notion of distance between positions; i.e., we do not yet have a way to express that $(1, 0)$ is closer to $(1, 2)$ than to $(5, 5)$, for example.

In the chapter on geometry, we defined the distance function (or *metric*) $d\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, where for two points $A = (a_1, a_2)$ and $B = (b_1, b_2)$ in $\mathbb{R}^2$, we have

$$d(A, B) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

which is inspired by Pythagoras. Here we do things in a different but equivalent way which nicely generalises to $\mathbb{R}^n$. We first define the length of a vector in terms of the dot product.

Definition 1.17 (Dot Product)**.** Let $\boldsymbol{u} = (u_i)$, $\boldsymbol{v} = (v_i)$ be vectors in $\mathbb{R}^n$. Then the *dot product* or *scalar product of* $\boldsymbol{u}$ *and* $\boldsymbol{v}$, denoted $\boldsymbol{u} \cdot \boldsymbol{v}$ or $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$, is the scalar defined by

$$\boldsymbol{u} \cdot \boldsymbol{v} = \sum_{i=1}^{n} u_i v_i = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Example 1.18. $(1, 2, 3) \cdot (4, 5, 6) = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 32.$

Definition 1.19. The *length* (or *magnitude* or *norm*) of a vector $\boldsymbol{v} \in \mathbb{R}^n$, denoted $\|\boldsymbol{v}\|$ or $|\boldsymbol{v}|$, is the scalar defined by

$$\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}.$$

When a vector $\boldsymbol{v}$ has $\|\boldsymbol{v}\| = 1$, then $\boldsymbol{v}$ is called a *unit vector* or a *direction*.

Example 1.20. $\|(1, 2, 3)\| = \sqrt{(1, 2, 3) \cdot (1, 2, 3)} = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}.$

For $\boldsymbol{v} = (x, y) \in \mathbb{R}^2$, we have $\|\boldsymbol{v}\| = \sqrt{x^2 + y^2} = \sqrt{(x - 0)^2 + (y - 0)^2}$. With our old definition, this turns out to be the distance $d((0, 0), (x, y))$ from the tip of the vector $\boldsymbol{v}$ to its head.

Definition 1.21 (Distance)**.** Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two positions in $\mathbb{R}^n$. Then the *distance* between $\boldsymbol{a}$ and $\boldsymbol{b}$, denoted $d(\boldsymbol{a}, \boldsymbol{b})$, is the length of their relative vector; i.e.,

$$d(\boldsymbol{a}, \boldsymbol{b}) = \|\boldsymbol{b} - \boldsymbol{a}\|.$$

Remark 1.22. For $A = (a_1, a_2)$ and $B = (b_1, b_2)$ in $\mathbb{R}^2$, we have

$$\begin{aligned} d(A, B) = \|\overrightarrow{AB}\| = \|\overrightarrow{OB} - \overrightarrow{OA}\| &= \|(b_1, b_2) - (a_1, a_2)\| \\ &= \|(b_1 - a_1, b_2 - a_2)\| \\ &= \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}, \end{aligned}$$

which shows that definition 1.21 agrees with our old definition for $d(A, B)$.

Exercise 1.23. 1. Find the distance between the following pairs of vectors.

a) $(1,2)$ and $(3,4)$ b) $(1,2,3)$ and $(-1,0,1)$

c) $\boldsymbol{i}$ and $\boldsymbol{j}$ in $\mathbb{R}^2$ d) $\boldsymbol{i}$ and $\boldsymbol{j}$ in $\mathbb{R}^3$

2. Show that in general for $\boldsymbol{a},\boldsymbol{b} \in \mathbb{R}^n$,

$$d(\boldsymbol{a},\boldsymbol{b}) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}.$$

3. Let $\boldsymbol{c} = (a,b) \in \mathbb{R}^2$ and $r \in \mathbb{R}$ where $r \geqslant 0$. Show that the set of points $\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^2 : d(\boldsymbol{c},\boldsymbol{x}) = r\}$ correspond to a circle, centred at (a,b) with radius r.

Definition 1.24 (Midpoint)**.** Let $\boldsymbol{u} = (u_i), \boldsymbol{v} = (v_i) \in \mathbb{R}^n$. The *midpoint* of $\boldsymbol{u}$ and $\boldsymbol{v}$ is the position $\boldsymbol{m}$ with coordinates

$$\boldsymbol{m} = \left(\frac{u_i + v_i}{2}\right),$$

i.e., the coordinates of $\boldsymbol{m}$ are the averages of the corresponding coordinates of $\boldsymbol{u}$ and $\boldsymbol{v}$.

Example 1.25. The midpoint of $(1,3,-5)$ and $(5,-3,2)$ is

$$\boldsymbol{m} = (\tfrac{1+5}{2}, \tfrac{3-3}{2}, \tfrac{-5+2}{2}) = (3,0,-\tfrac{3}{2}).$$

Proposition 1.26. *Let $\boldsymbol{u},\boldsymbol{v} \in \mathbb{R}^n$ and let $\boldsymbol{m}$ be their midpoint. Then*

$$d(\boldsymbol{u},\boldsymbol{m}) = d(\boldsymbol{m},\boldsymbol{v}),$$

i.e., the midpoint $\boldsymbol{m}$ lies "in the middle" of $\boldsymbol{u}$ and $\boldsymbol{v}$.

Proof. This goes similarly to the proof for $\mathbb{R}^2$ in geometry:

$$\begin{aligned} d(\boldsymbol{u},\boldsymbol{m}) &= d\left((u_i), \left(\frac{u_i+v_i}{2}\right)\right) \\ &= \left\|(u_i) - \left(\frac{u_i+v_i}{2}\right)\right\| = \left\|\left(u_i - \frac{u_i+v_i}{2}\right)\right\| = \left\|\left(\frac{u_i-v_i}{2}\right)\right\| \end{aligned}$$

$$= \left\| \left(\frac{u_i + v_i}{2} - v_i \right) \right\| = \left\| \left(\frac{u_i + v_i}{2} \right) - (v_i) \right\|$$
$$= d\left(\left(\frac{u_i + v_i}{2} \right), (v_i) \right) = d(\boldsymbol{m}, \boldsymbol{v}),$$

as required. □

Notation (Scalar division)**.** Let $\lambda \in \mathbb{R}$, $\lambda \neq 0$, and $\boldsymbol{v} \in \mathbb{R}^n$. We adopt the notation

$$\frac{\boldsymbol{v}}{\lambda}$$

or $\boldsymbol{v}/\lambda$ in-line, to stand for $\frac{1}{\lambda}\boldsymbol{v}$.

Definition 1.27 (Normalised vector)**.** Let $\boldsymbol{v} \in \mathbb{R}^n$ be a non-zero vector. The *normalised version* or *direction* of $\boldsymbol{v}$, denoted $\hat{\boldsymbol{v}}$, is the vector $\boldsymbol{v}/\|\boldsymbol{v}\|$.

Proposition 1.28. *Every normalised (non-zero) vector is a unit vector.*

Proof. Let $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$. Then

$$\|\hat{\boldsymbol{v}}\|^2 = \left\| \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \right\|^2 = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \cdot \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} = \frac{\boldsymbol{v} \cdot \boldsymbol{v}}{\|\boldsymbol{v}\|^2} = \frac{\boldsymbol{v} \cdot \boldsymbol{v}}{\boldsymbol{v} \cdot \boldsymbol{v}} = 1,$$

and so $\|\hat{\boldsymbol{v}}\| = 1$, as required. □

Exercise 1.29. Do you believe the proof presented for proposition 1.28? Are there any unjustified steps? Yes! The step $\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \cdot \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} = \frac{\boldsymbol{v}\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}$ is not obvious for the dot product (remember this is not ordinary multiplication). Prove using the definition of the dot product that if $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ and $\lambda, \mu \in \mathbb{R}$, then $(\lambda\boldsymbol{u}) \cdot (\mu\boldsymbol{v}) = (\lambda\mu)(\boldsymbol{u} \cdot \boldsymbol{v})$. (In the case of the proof, this was applied with $\lambda = \mu = \frac{1}{\|\boldsymbol{v}\|}$ and $\boldsymbol{u} = \boldsymbol{v}$.)

Proposition 1.30. *Every non-zero vector $\boldsymbol{v} \in \mathbb{R}^n$ can be written uniquely as $\lambda\boldsymbol{u}$ for some $\lambda > 0$ and unit vector $\boldsymbol{u} \in \mathbb{R}^n$, where $\lambda = \|\boldsymbol{v}\|$ and $\boldsymbol{u} = \hat{\boldsymbol{v}}$.*

Proof. Clearly $\boldsymbol{v} = \|\boldsymbol{v}\|\,\hat{\boldsymbol{v}}$ by definition of $\hat{\boldsymbol{v}}$.

Now for uniqueness, suppose $\boldsymbol{v} = \lambda\boldsymbol{u}$ where $\boldsymbol{u}$ is unit, and $0 < \lambda \neq \|\boldsymbol{v}\|$. But then

$$\|\boldsymbol{v}\| = \|\lambda\boldsymbol{u}\| \overset{\text{by exercise 1.31 1(c)}}{=\!=\!=\!=\!=} |\lambda|\|\boldsymbol{u}\| = |\lambda| = \lambda \neq \|\boldsymbol{v}\|$$

since $\boldsymbol{u}$ is unit, a contradiction. Therefore we must have $\lambda = \|\boldsymbol{v}\|$, and so suppose $\boldsymbol{v} = \lambda\boldsymbol{u}$ where $\boldsymbol{u}$ is unit but this time $\boldsymbol{u} \neq \hat{\boldsymbol{v}}$. Then

$$\boldsymbol{v} = \lambda\boldsymbol{u} \ = \|\boldsymbol{v}\|\boldsymbol{u}$$

since λ must be $\|\boldsymbol{v}\|$. But it follows that $\boldsymbol{u} = \boldsymbol{v}/\|\boldsymbol{v}\| = \hat{\boldsymbol{v}}$, a contradiction. Therefore we must have $\lambda = \|\boldsymbol{v}\|$ and $\boldsymbol{u} = \hat{\boldsymbol{v}}$, so uniqueness follows. □

Exercise 1.31. We assume that the easy results proved in these exercises are known throughout the rest of the notes.

1. Prove the following for any $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

   a) $\mathbf{0} \cdot \boldsymbol{v} = 0$
   b) $\boldsymbol{v} \cdot \boldsymbol{v} = \|\boldsymbol{v}\|^2$
   c) $\|\lambda\boldsymbol{v}\| = |\lambda|\|\boldsymbol{v}\|$
   d) $\boldsymbol{u} \cdot \boldsymbol{v} = \|\boldsymbol{u}\|\|\boldsymbol{v}\|(\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{v}})$

2. Prove the following for any $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

   a) $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{w}$
   b) $\boldsymbol{u} \cdot \boldsymbol{u} = 0 \iff \boldsymbol{u} = \mathbf{0}$
   c) $\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \cdot \boldsymbol{w}$
   d) $\boldsymbol{u} \cdot (\lambda\boldsymbol{v}) = \lambda(\boldsymbol{u} \cdot \boldsymbol{v}) = (\lambda\boldsymbol{u}) \cdot \boldsymbol{v}$
   e) $\|\boldsymbol{u} + \boldsymbol{v}\| = \|\boldsymbol{u}\|^2 + 2\boldsymbol{u} \cdot \boldsymbol{v} + \|\boldsymbol{v}\|^2$
   f) $(\boldsymbol{u} + \boldsymbol{v}) \cdot (\boldsymbol{u} - \boldsymbol{v}) = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$

3. Show, diagrammatically, that any unit vector $\boldsymbol{u} \in \mathbb{R}^2$ has the form
$$\boldsymbol{u} = (\cos\theta, \sin\theta)$$
where $\theta \in [-\pi, \pi]$ is the angle $\boldsymbol{u}$ makes with the x-axis.

4. Show that for any vector $\boldsymbol{v} \in \mathbb{R}^n$, the ith component of $\boldsymbol{v}$ is given by $v_i = \langle \boldsymbol{v}, \boldsymbol{e}_i \rangle$, and
$$\boldsymbol{v} = \sum_{k=1}^{n} \langle \boldsymbol{v}, \boldsymbol{e}_i \rangle \, \boldsymbol{e}_i.$$

Proposition 1.32 (Cauchy–Schwarz Inequality)**.** *Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. Then*

$$|\boldsymbol{u} \cdot \boldsymbol{v}| \leqslant \|\boldsymbol{u}\|\|\boldsymbol{v}\|.$$

Proof. We prove that $\boldsymbol{u} \cdot \boldsymbol{v} \leqslant \|\boldsymbol{u}\|\|\boldsymbol{v}\|$, because then replacing $\boldsymbol{u}$ with $-\boldsymbol{u}$ yields $\|\boldsymbol{u}\|\|\boldsymbol{v}\| = \|-\boldsymbol{u}\|\|\boldsymbol{v}\| \geqslant (-\boldsymbol{u}) \cdot \boldsymbol{v} = \sum_{i=1}^n(-u_i v_i) = -\sum_{i=1}^n u_i v_i = -\boldsymbol{u} \cdot \boldsymbol{v}$, so that $-\|\boldsymbol{u}\|\|\boldsymbol{v}\| \leqslant \boldsymbol{u} \cdot \boldsymbol{v}$.

Clearly for any $x, y \in \mathbb{R}$, we have $(x-y)^2 \geqslant 0$, which expands to give $x^2 + y^2 \geqslant 2xy$. If we suppose for now that $\boldsymbol{u}$ and $\boldsymbol{v}$ are unit vectors, by definition we get

$$\begin{aligned} \boldsymbol{u} \cdot \boldsymbol{v} = \sum_{i=1}^n u_i v_i \leqslant \sum_{i=1}^n \frac{{u_i}^2 + {v_i}^2}{2} &= \frac{1}{2}\left(\sum_{i=1}^n {u_i}^2 + \sum_{i=1}^n {v_i}^2\right) \\ &= \frac{1}{2}(\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2) \\ &= \frac{1}{2}(1+1) = 1 = \|\boldsymbol{u}\|\|\boldsymbol{v}\|. \end{aligned}$$

So the result holds for unit vectors. If $\boldsymbol{u}$ and $\boldsymbol{v}$ are not unit (and not zero), then their normalised versions are unit by proposition 1.28, so $\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{v}} \leqslant 1$, and thus $(\boldsymbol{u} \cdot \boldsymbol{v})/\|\boldsymbol{u}\|\|\boldsymbol{v}\| \leqslant 1$. Finally if any of $\boldsymbol{u}, \boldsymbol{v}$ are zero, the result is immediate. □

The Cauchy–Schwarz inequality is a very useful inequality. We need it here for the following definition, because it ensures that the inverse cosine of a dot product of two unit vectors is always defined.

Definition 1.33 (Angle between two vectors)**.** Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. The *angle* $\measuredangle(\boldsymbol{u}, \boldsymbol{v})$ between the vectors $\boldsymbol{u}, \boldsymbol{v}$ is the real number $\measuredangle(\boldsymbol{u}, \boldsymbol{v})$ in $[0, \pi]$ defined by

$$\measuredangle(\boldsymbol{u}, \boldsymbol{v}) = \cos^{-1}(\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{v}}).$$

Remark 1.34. Note that there is no geometric meaning for angles in $\mathbb{R}^n$ for $n > 3$, that is why we take this definition. Let us show that it agrees with our usual understanding of an angle in $\mathbb{R}^2$.

Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$. By proposition 1.30 and by exercise 1.31.3, we can write

$$\boldsymbol{u} = \|\boldsymbol{u}\|(\cos\theta, \sin\theta) \qquad \text{and} \qquad \boldsymbol{v} = \|\boldsymbol{v}\|(\cos\phi, \sin\phi)$$

where $\theta, \phi \in [-\pi, \pi]$ are the angles $\boldsymbol{u}$ and $\boldsymbol{v}$ make with the x-axis.

Sketching a quick diagram, we see that the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ is $\theta - \phi$. Indeed, we have

$$\boldsymbol{u} \cdot \boldsymbol{v} = \|\boldsymbol{u}\|\|\boldsymbol{v}\|(\cos\theta\cos\phi + \sin\theta\sin\phi) = \|\boldsymbol{u}\|\|\boldsymbol{v}\|\cos(\theta - \phi),$$

so that $\cos(\theta - \phi) = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|} = \hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{v}}$, as required.

Definitions 1.35. Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ be two vectors.

(i) $\boldsymbol{u}$ and $\boldsymbol{v}$ are said to be *in the same direction* if $\hat{\boldsymbol{u}} = \hat{\boldsymbol{v}}$, and *in opposite directions* if $\hat{\boldsymbol{u}} = -\hat{\boldsymbol{v}}$. In either case, $\boldsymbol{u}$ and $\boldsymbol{v}$ are said to be *parallel*, denoted $\boldsymbol{u} \parallel \boldsymbol{v}$.

(ii) $\boldsymbol{u}$ and $\boldsymbol{v}$ are said to be *perpendicular* or *orthogonal* if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.

Proposition 1.36. *Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ be two vectors. Then*

(i) If $\boldsymbol{u}$ and $\boldsymbol{v}$ are in the same direction, the angle between them is 0.

(ii) If $\boldsymbol{u}$ and $\boldsymbol{v}$ are in opposite directions, the angle between them is π.

(iii) If $\boldsymbol{u}$ and $\boldsymbol{v}$ are perpendicular, the angle between them is $\frac{\pi}{2}$.

Proof. These easily follow from definition 1.33. For (i), $\hat{\boldsymbol{u}} = \hat{\boldsymbol{v}}$ gives that $\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{v}} = \hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{u}} = \|\hat{\boldsymbol{u}}\|^2 = 1$, so in this case $\measuredangle(\boldsymbol{u}, \boldsymbol{v}) = \cos^{-1}(1) = 0$. For (ii), we have $\hat{\boldsymbol{u}} = -\hat{\boldsymbol{v}}$, so now $\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{v}} = \hat{\boldsymbol{u}} \cdot (-\hat{\boldsymbol{u}}) = -(\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{u}}) = -\|\hat{\boldsymbol{u}}\|^2 = -1$, and therefore $\measuredangle(\boldsymbol{u}, \boldsymbol{v}) = \cos^{-1}(1) = \pi$. Finally for (iii), we have

$$\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{v}} = \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|} \cdot \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} = \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|} = \frac{0}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|} = 0,$$

so $\measuredangle(\boldsymbol{u}, \boldsymbol{v}) = \cos^{-1}(0) = \frac{\pi}{2}$. □

Exercise 1.37.

1. A triangle ABC has vertices $A(0, -1, 1)$, $B(2, 3, -2)$ and $C(3, 1, 0)$. Express the vectors $\overrightarrow{AB}$, $\overrightarrow{BC}$ and $\overrightarrow{CA}$ in terms of $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$, and hence find the lengths of the three sides.
2. Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal vectors in $\mathbb{R}^n$. Show that
$$\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$$
and explain why this is equivalent to Pythagoras' theorem.
3. In triangle ABC, $A = (3, 3, -2)$, $B = (-2, 0, 5)$ and $C = (1, -2, 1)$. If L and M are the midpoints of AB and AC respectively, show

that LM is parallel to BC.

4. Simplify the expression $\|\boldsymbol{b}\|^2 + \|\boldsymbol{c}\|^2 - (\boldsymbol{b}-\boldsymbol{c})\cdot(\boldsymbol{b}-\boldsymbol{c})$. By taking $\boldsymbol{b} = \overrightarrow{AC}$ and $\boldsymbol{c} = \overrightarrow{AB}$, deduce the cosine formula

$$a^2 = b^2 + c^2 - 2bc\cos(B\widehat{A}C)$$

for triangle ABC shown in figure 8.

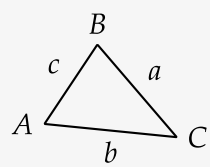

FIGURE 8: Triangle ABC

5. Suppose $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ and $\boldsymbol{d}$ are positions of the vertices of a parallelogram. Express $\boldsymbol{d}$ in terms of the other three vectors.

6. Suppose $(1,2)$ and $(4,1)$ are opposite vertices of a square in $\mathbb{R}^2$. Find the coordinates of the other two vertices.

7. The vectors $\mathbf{0}$, $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{x}+\boldsymbol{y}$ are the position vectors of vertices of a parallelogram. Show that the sum of the squares of the diagonals is equal to the sum of the squares of the sides, i.e.

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 = 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2.$$

Deduce Euclid's median formula: in a triangle ABC where M is the midpoint of BC,

$$AM^2 = \frac{AB^2 + AC^2}{2} - \frac{BC^2}{4}.$$

8. Points on the perpendicular bisector of the line segment $\overrightarrow{AB}$ satisfy the equation $\|\boldsymbol{x} - \overrightarrow{OA}\| = \|\boldsymbol{x} - \overrightarrow{OB}\|$. Show that for $\mathbb{R}^2$, this expands out to

$$2(a_1 - b_1)x + 2(a_2 - b_2)y = {a_1}^2 + {a_2}^2 - {b_1}^2 - {b_2}^2$$

where $\boldsymbol{x} = (x,y)$, $A = (a_1, a_2)$ and $B = (b_1, b_2)$.

1.3 Matrix Operations

A matrix is, in some sense, a generalisation of a vector. Consider the set

$$(\mathbb{R}^2)^3 = (\mathbb{R}\times\mathbb{R})^3 = (\mathbb{R}\times\mathbb{R})\times(\mathbb{R}\times\mathbb{R})\times(\mathbb{R}\times\mathbb{R}).$$

This contains elements of the form $((a,b),(c,d),(e,f))$. We will write these elements instead as

$$\begin{pmatrix} a & c & e \\ b & d & f \end{pmatrix}.$$

Such an object is what we call a 2×3 matrix, and the set of such matrices is denoted $\mathbb{R}^{2\times 3}$. Let us give a general definition.

Definition 1.38 (Matrix)**.** An $m\times n$ *matrix* is a rectangular array of real numbers, called *entries*, arranged in m *rows* and n *columns*. The expression $m\times n$ is called the *size* of the matrix, and the set of all $m\times n$ matrices is denoted $\mathbb{R}^{m\times n}$.

An $m\times n$ matrix can be expressed in general as

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

or, similarly to vectors, concisely as

$$\mathbf{A} = (a_{ij})_{m\times n}$$

where a_{ij} denotes the entry in the ith row and jth column, i and j being called the row and column indices, respectively. As before, we relax this to (a_{ij}) when the size is clear in context.

As we have seen already, matrices are denoted by single capital letters in bold typeface, and their entries are denoted using the corresponding small letter with two subscripts ranging over the rows and columns ($i = 1,\dots,m$ and $j = 1,\dots,n$).

When $m = n$ (i.e., size $n\times n$) the matrix is said to be *square*, and the entries a_{ii} (that is, a_{11}, a_{22}, …, a_{nn}) make up the *diagonal* of the matrix. If a matrix is not square, it is called *rectangular*.

We identify n-vectors (i.e., vectors in $\mathbb{R}^n$) in the world of matrices with $n\times 1$ matrices. So for example, the vector (x,y) corresponds to the matrix $\binom{x}{y}$.

Examples 1.39. Consider the following matrices.

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 2 & -5 & 7 & 12 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 4 \end{pmatrix},$$

$$\mathbf{D} = \begin{pmatrix} 0 & \pi & e \\ \frac{1}{2} & -\frac{1}{3} & \frac{1}{4} \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} 2 & 3 \\ 5 & 1 \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

These matrices are, in order, 3×2, 1×5, 1×1, 3×3, 2×2 and 3×1. Matrices **C**, **D** and **E** are square, whereas **A**, **B** and $\boldsymbol{x}$ are rectangular.

The matrix $\boldsymbol{x}$ is equivalent to the vector $(0, 1, 0)$, whereas the matrix **B** is *not* a vector ($1 \times n$ matrices are sometimes called *row vectors* or *covectors*, but they are not considered vectors). Notice that we still use lowercase letters for vectors here.

The matrix **C** is a number (or scalar). We do not distinguish between scalars and 1×1 matrices.

Definition 1.40 (Matrix Equality)**.** Let $\mathbf{A} = (a_{ij})_{m\times n}$ and $\mathbf{B} = (b_{ij})_{\ell\times k}$. Then the matrices **A** and **B** are *equal*, denoted

$$\mathbf{A} = \mathbf{B},$$

if $m = \ell$, $n = k$ and $a_{ij} = b_{ij}$ for all $i = 1, \dots, m$ and $j = 1, \dots, n$.

Example 1.41. None of the matrices below are equal to each other.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 0 \\ 3 & 4 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}$$

Definition 1.42 (Matrix Addition)**.** Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be two $m \times n$ matrices. Then the *sum* $\mathbf{A} + \mathbf{B}$ is the $m \times n$ matrix given by

$$\mathbf{A} + \mathbf{B} = (a_{ij} + b_{ij}).$$

Matrices of different size cannot be added.

Example 1.43. Consider the matrices

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & -4 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 4 & -1 & 0 \\ 6 & 11 & -13 \end{pmatrix}.$$

Then their sum is

$$\mathbf{A}+\mathbf{B}=\begin{pmatrix}1+4 & 2-1 & 3+0\\ 0+6 & 5+11 & -4-13\end{pmatrix}=\begin{pmatrix}5 & 1 & 3\\ 6 & 16 & -17\end{pmatrix}.$$

Definition 1.44 (Scalar Multiplication)**.** Let $\lambda \in \mathbb{R}$ be a scalar, and let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix. Then the *scalar multiplication of* $\mathbf{A}$ *with* λ, denoted $\lambda\mathbf{A}$, is the matrix given by

$$\lambda\mathbf{A} = (\lambda a_{ij}).$$

Example 1.45. Consider the matrix

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

Then the matrix $5\mathbf{A}$ is

$$5\mathbf{A}=\begin{pmatrix}10 & 0\\ -5 & 25\end{pmatrix}.$$

Remark 1.46. The operations of addition and scalar multiplication which we defined here coincide with those defined earlier for vectors. What this means is that if we treat a vector $\boldsymbol{v} \in \mathbb{R}^n$ as an $n \times 1$ matrix, and apply the definitions given here for addition and scalar multiplication, then the result will be identical to what we expect using the original definitions.

Notation. Just as we did with vectors, we denote the matrix $-1\mathbf{A}$ by $-\mathbf{A}$, and introduce the *difference* $\mathbf{A}-\mathbf{B}$ between two matrices (of the same size), defined by

$$\mathbf{A}-\mathbf{B}=\mathbf{A}+(-\mathbf{B}).$$

Now we introduce something which we has no vector analogue.

Definition 1.47 (Matrix Multiplication)**.** Let $\mathbf{A} = (a_{ij})_{m\times d}$ and $\mathbf{B} = (b_{ij})_{d\times n}$ be two matrices. Then we define the *product of* $\mathbf{A}$ *and* $\mathbf{B}$, denoted $\mathbf{AB}$, to be the $m \times n$ matrix given by

$$\mathbf{AB}=\left(\sum_{k=1}^{d} a_{ik}b_{kj}\right)_{m\times n}.$$

Remark 1.48. This definition seems rather complicated, so let's break it down. Notice that the index of summation, k, varies the column index j of a_{ij}, and the row index i of b_{ij}. Thus the ijth entry of $\mathbf{AB}$ is

$$\sum_{k=1}^{d} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{id}b_{dj}.$$

If we consider the rows of the matrix $\mathbf{A}$ to be the vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ and the columns of the matrix $\mathbf{B}$ to be the vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n$, then the ijth entry turns out to be the dot product $\boldsymbol{a}_i \cdot \boldsymbol{b}_j$, that is,

$$\mathbf{AB} = \begin{pmatrix} — & \boldsymbol{a}_1 & — \\ — & \boldsymbol{a}_2 & — \\ & \vdots & \\ — & \boldsymbol{a}_m & — \end{pmatrix} \begin{pmatrix} | & | & & | \\ \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_n \\ | & | & & | \end{pmatrix}$$

$$= \begin{pmatrix} \boldsymbol{a}_1 \cdot \boldsymbol{b}_1 & \boldsymbol{a}_1 \cdot \boldsymbol{b}_2 & \cdots & \boldsymbol{a}_1 \cdot \boldsymbol{b}_n \\ \boldsymbol{a}_2 \cdot \boldsymbol{b}_1 & \boldsymbol{a}_2 \cdot \boldsymbol{b}_2 & \cdots & \boldsymbol{a}_2 \cdot \boldsymbol{b}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_m \cdot \boldsymbol{b}_1 & \boldsymbol{a}_m \cdot \boldsymbol{b}_2 & \cdots & \boldsymbol{a}_m \cdot \boldsymbol{b}_n \end{pmatrix}$$

Thus we say that the each entry is obtained by doing "*row times column*". Notice the restriction that this places on the dimensions of the matrices $\mathbf{A}$ and $\mathbf{B}$: they need to be of size $m \times d$ and $d \times n$, so that the dot product is between two vectors of dimension d (that is, both have d entries).

When two matrices $\mathbf{A}$ and $\mathbf{B}$ are of size $m \times n$ and $\ell \times k$ with $n = \ell$ as required for multiplication, we say that they are *compatible*, *conformal* or that *their inner dimensions match*. Otherwise, the product $\mathbf{AB}$ does not exist.

Let us give some examples.

Examples 1.49. Take the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}, \boldsymbol{x}$ from examples 1.39:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 2 & -5 & 7 & 12 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 4 \end{pmatrix},$$

$$\mathbf{D} = \begin{pmatrix} 0 & \pi & e \\ \frac{1}{2} & -\frac{1}{3} & \frac{1}{4} \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} 2 & 3 \\ 5 & 1 \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

The product **AD** does not exist, since **A** is of size 3×2 and **D** is of size 3×3. **DA** on the other hand does exist, since **D** has size 3×3 and **A** has size 3×2:

$$\mathbf{D}_{3\times 3}\mathbf{A}_{3\times 2} = \begin{pmatrix} 0 & \pi & e \\ \frac{1}{2} & -\frac{1}{3} & \frac{1}{4} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

We know that the resulting product will have size 3×2 by definition ($\mathbf{D}_{3\times 3}\mathbf{A}_{3\times 2}$). Now each ijth entry is the dot product of the ith row with the jth column:

$$\mathbf{DA} = \begin{pmatrix} (0,\pi,e)\cdot(1,3,5) & (0,\pi,e)\cdot(2,4,6) \\ (\frac{1}{2},-\frac{1}{3},\frac{1}{4})\cdot(1,3,5) & (\frac{1}{2},-\frac{1}{3},\frac{1}{4})\cdot(2,4,6) \\ (0,0,0)\cdot(1,3,5) & (0,0,0)\cdot(2,4,6) \end{pmatrix}$$

$$= \begin{pmatrix} 3\pi + 5e & 4\pi + 6e \\ \frac{3}{4} & \frac{7}{6} \\ 0 & 0 \end{pmatrix}$$

What follows immediately from this example is that matrix multiplication is not commutative; i.e., in general, $\mathbf{AB} \neq \mathbf{BA}$. In fact, in this case, only one of these products exists.

As another example, let's find the product $\mathbf{D}\boldsymbol{x}$. The product exists because **D** is 3×3 and **x** is 3×1. The result is 3×1.

$$\mathbf{D}\boldsymbol{x} = \begin{pmatrix} 0 & \pi & e \\ \frac{1}{2} & -\frac{1}{3} & \frac{1}{4} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \pi \\ -\frac{1}{3} \\ 0 \end{pmatrix}.$$

Exercise 1.50. Find, if they exist, the products

a) **AB** b) **CB** c) **BC** d) **AE**

e) $\boldsymbol{x}$**B** f) $\boldsymbol{x}$**C** g) (**DA**)**E** h) **D**(**AE**)

Remark 1.51. Observe that the 1×1 matrix $\mathbf{C}$, when compatible with other matrices, behaves as a scalar multiple. Thus in general 1×1 matrices are treated as scalars and are considered "compatible" with all matrices (in the sense of definition 1.44).

Definition 1.52 (Zero matrix). The matrix $\mathbf{O} = (0) = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$ is called the *zero matrix*.

Theorem 1.53 (Ring properties for matrices). *Let* $\mathbf{A}, \mathbf{B}, \mathbf{C}$ *be three matrices. Assuming that the matrix dimensions are such that the operations can be performed, we have the following:*

I) $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$

II) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$

III) $\mathbf{A} + \mathbf{O} = \mathbf{A}$

IV) $\mathbf{A} + (-\mathbf{A}) = \mathbf{O}$

V) $\mathbf{A}(\mathbf{B}\mathbf{C}) = \mathbf{A}(\mathbf{B}\mathbf{C})$

VI) $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{A}\mathbf{B} + \mathbf{B}\mathbf{C}$

VII) $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{A}\mathbf{C} + \mathbf{B}\mathbf{C}$

Proof. Just as in theorem 1.10, these results easily follow from the definitions. The only hard one is V, where we have

$$\begin{aligned}
\mathbf{A}(\mathbf{B}\mathbf{C}) &= (a_{ij})\big((b_{ij})(c_{ij})\big) \\
&= (a_{ij}) \left(\sum_{k=1}^{d} b_{ik} c_{kj} \right) && \text{(by definition 1.47)} \\
&= \left(\sum_{\ell=1}^{\delta} \left(a_{i\ell} \sum_{k=1}^{d} b_{\ell k} c_{kj} \right) \right) && \text{(by definition 1.47)} \\
&= \left(\sum_{\ell=1}^{\delta} \sum_{k=1}^{d} a_{i\ell} b_{\ell k} c_{kj} \right) && \text{(by linearity of } \Sigma\text{)} \\
&= \left(\sum_{k=1}^{d} \sum_{\ell=1}^{\delta} a_{i\ell} b_{\ell k} c_{kj} \right) && \text{(finite sums can be interchanged[1])} \\
&= \left(\sum_{k=1}^{d} \left(\sum_{\ell=1}^{\delta} a_{i\ell} b_{\ell k} \right) c_{kj} \right) && \text{(by linearity of } \Sigma\text{)} \\
&= \left(\sum_{\ell=1}^{\delta} a_{i\ell} b_{\ell j} \right) (c_{ij}) && \text{(by definition 1.47)}
\end{aligned}$$

$$= \big((a_{ij})(b_{ij})\big)(c_{ij}) = (\mathbf{AB})\mathbf{C},$$

as required. Proofs of the remaining properties are left as an exercise. □

Definition 1.54 (Matrix Transpose)**.** Let $\mathbf{A} = (a_{ij})_{m\times n}$. The *transpose of* $\mathbf{A}$, denoted $\mathbf{A}^{\mathsf{T}}$, is the $n \times m$ matrix given by

$$\mathbf{A}^{\mathsf{T}} = (a_{ji})_{n\times m}.$$

Example 1.55. If $\mathbf{A}$, $\mathbf{B}$ and $\boldsymbol{x}$ are given by

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 4 & \frac{1}{2} \end{pmatrix} \qquad \boldsymbol{x} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix},$$

then we have

$$\mathbf{A}^{\mathsf{T}} = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \qquad \mathbf{B}^{\mathsf{T}} = \begin{pmatrix} 1 & 4 \\ 0 & \frac{1}{2} \end{pmatrix} \qquad \boldsymbol{x}^{\mathsf{T}} = \begin{pmatrix} -1 & 0 & 1 \end{pmatrix}.$$

Now we introduce a special matrix which behaves analogously to the number 1 in the set of real numbers. The number 1 is called the *multiplicative identity* in $\mathbb{R}$, because it does nothing to numbers under multiplication (preserving their "identity"):

$$x \cdot 1 = 1 \cdot x = x.$$

Analogously we have the *additive identity* $0 \in \mathbb{R}$, since 0 does not change numbers under addition:

$$x + 0 = 0 + x = x.$$

For matrices, the additive identity is simply the zero matrix $\mathbf{O}$, which exhibits the desired behaviour

$$\mathbf{A} + \mathbf{O} = \mathbf{O} + \mathbf{A} = \mathbf{A}$$

for any matrix $\mathbf{A}$ (of compatible size), as seen in theorem 1.53 III. What we would like to try and obtain here is a multiplicative identity for matrices. Note that this is not as simple a task as determining an additive identity,

[1]See proposition A.12.

firstly because multiplication of matrices is not defined as simply as addition, and secondly because multiplication is not commutative; so even if we find some identity matrix $\mathbf{I}$ which satisfies $\mathbf{AI} = \mathbf{A}$, it needn't satisfy $\mathbf{IA} = \mathbf{A}$.

Here is the definition.

Definition 1.56 (Identity matrix)**.** The *identity matrix* is the $n \times n$ matrix denoted $\mathbf{I}_n$ or simply $\mathbf{I}$, defined by

$$\mathbf{I}_n = (\delta_{ij})$$

where δ_{ij} denotes the Kronecker delta (definition 1.12).

Thus the first few identity matrices are

$$\mathbf{I}_1 = (1), \quad \mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{I}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{I}_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

with ones on the diagonal and zeros everywhere else.

Observe that if $\mathbf{A}$ is an $m \times n$ matrix where $m \neq n$, then it is impossible to have $\mathbf{I}_n\mathbf{A} = \mathbf{AI}_n = \mathbf{A}$ for the same n, simply because only one of these products can exist (because of their size). Refer to example 1.57, and verify the computation yourself.

Example 1.57. Suppose $\mathbf{A} = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{smallmatrix}\right)$. Then we have

$$\mathbf{AI}_3 = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \mathbf{I}_2\mathbf{A}.$$

Let us now prove that $\mathbf{I}$ behaves as desired in the general case.

Theorem 1.58. *Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix. Then*

$$\mathbf{AI}_n = \mathbf{I}_m\mathbf{A} = \mathbf{A}.$$

Proof. Recall $\mathbf{I}_n = (\delta_{ij})$. Thus

$$\mathbf{AI}_n = \left(\sum_{k=1}^{n} a_{ik}\delta_{kj} \right)$$

$$\begin{aligned}
&= (a_{i1}\delta_{1j} + \cdots + a_{i(j-1)}\delta_{(j-1)j} + a_{ij}\delta_{jj} + a_{i(j+1)}\delta_{(j+1)j} + \cdots + a_{in}\delta_{nj}) \\
&= (a_{i1}0 + \cdots + a_{i(j-1)}0 + a_{ij}1 + a_{i(j+1)}0 + \cdots + a_{in}0) \\
&= (a_{ij}) = \mathbf{A},
\end{aligned}$$

and by a similar reasoning we get $\mathbf{I}_m\mathbf{A} = \mathbf{A}$. □

Note that since in the general case, different **I**'s are required on the left and the right of a matrix **A** to act as a multiplicative identity, we call **I** such that $\mathbf{IA} = \mathbf{A}$ the *left identity*, and **I** such that $\mathbf{AI} = \mathbf{A}$ the *right identity*. It is easy to see that a matrix **A** has the same left and right identity if and only if the matrix **A** is square.

The final operation we introduce is the analogue of division for matrices. Before we give a definition however, let us again consider the real numbers first. What does it mean to *divide*? In an infantile treatment of arithmetic, division is introduced as a distinct operation from multiplication, just as subtraction is thought of being distinct from addition. But the way we actually treat subtraction in more formal considerations is the addition of some "inverse element", i.e., $x - y$ is shorthand for $x + (-y)$, where $-y$ is a number such that

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

We call $-y$ the *additive inverse* of y. (Note that 0 on the right-hand side is the additive identity).

Likewise, $x \div y$ or $\frac{x}{y}$ denotes $x \cdot y^{-1}$, where y^{-1} is a number such that

$$y \cdot y^{-1} = y^{-1} \cdot y = 1.$$

We call y^{-1} the *multiplicative inverse*, and this time we have $= 1$ on the right hand side, since 1 is the multiplicative identity. In the general case, when we have some binary operation $*$ defined on a set X, an element $i \in X$ is an *identity* if

$$x * i = i * x = x$$

for all $x \in X$, and the inverse element $x^{-1} \in X$ of x is an element such that

$$x * x^{-1} = x^{-1} * x = i.$$

In the context of matrices, the additive inverse of **A** is $-\mathbf{A} = (-1)\mathbf{A}$, since $\mathbf{A} + (-\mathbf{A}) = (-\mathbf{A}) + \mathbf{A} = \mathbf{O}$. Again, multiplication will prove to be the more challenging case. In fact, we will focus solely on 2×2 matrices for now.

First of all, if $\mathbf{A}$ is $m \times n$ with $m \neq n$, we cannot have one matrix $\mathbf{A}^{-1}$ such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I},$$

simply because of the sizes: if $\mathbf{A}^{-1}$ exists it will have to be $n \times m$ and on the left we get $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$, whereas on the right we get $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_m$. There is a study of *left* and *right inverses*, however we will not get into it here, and focus solely on square matrices where everything is $n \times n$ (everything meaning the matrix, its inverse, and $\mathbf{I}$). Thus, we have the following definition.

Definition 1.59 (Matrix Inverse)**.** Let $\mathbf{A}$ be an $n \times n$ (square) matrix. An $n \times n$ matrix $\mathbf{A}^{-1}$ such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$$

is said to be an *inverse* of $\mathbf{A}$. If $\mathbf{A}$ has an inverse, then it is said to be *invertible*. Otherwise, $\mathbf{A}$ is said to be *singular*.

Exercise 1.60. Suppose $\mathbf{A} = \begin{pmatrix} -5 & 1 & -2 \\ 2 & 1 & 2 \\ 2 & 0 & 1 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & -1 & 4 \\ 2 & -1 & 6 \\ -2 & 2 & -7 \end{pmatrix}$.

Work out $\mathbf{AB}$ and $\mathbf{BA}$. What do you conclude?

Now clearly for 1×1 matrices, being essentially numbers, the inverse of the matrix $\mathbf{A} = (a)$ is simply $\mathbf{A}^{-1} = \left(\frac{1}{a}\right)$, as long as $a \neq 0$. Indeed,

$$\mathbf{A}\mathbf{A}^{-1} = (a)\left(\frac{1}{a}\right) = (1) = \mathbf{I}_1,$$

and similarly we get $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_1$.

Theorem 1.61. *Let $\mathbf{A}$ be a square matrix. Then if $\mathbf{A}^{-1}$ exists, it is unique.*

Proof. Suppose $\mathbf{B}$ and $\mathbf{C}$ are two inverses of $\mathbf{A}$. Then

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I} \qquad \text{and} \qquad \mathbf{AC} = \mathbf{CA} = \mathbf{I}.$$

In particular, $\mathbf{AB} = \mathbf{AC}$, so

$$\mathbf{B}(\mathbf{AB}) = \mathbf{B}(\mathbf{AC}) \implies (\mathbf{BA})\mathbf{B} = (\mathbf{BA})\mathbf{C} \implies \mathbf{IB} = \mathbf{IC} \implies \mathbf{B} = \mathbf{C},$$

thus any two inverses of $\mathbf{A}$ are equal, proving that $\mathbf{A}^{-1}$ is unique. □

The following theorem is helpful because it saves us having to check that both $\mathbf{AA}^{-1} = \mathbf{I}$ and $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$; one of them is enough to prove inverse.

Theorem 1.62. *Suppose* $\mathbf{A}$ *and* $\mathbf{B}$ *are square matrices such that* $\mathbf{AB} = \mathbf{I}$*. Then* $\mathbf{BA} = \mathbf{I}$*.*

Proof. We give an incomplete proof, because we assume $\mathbf{B}^{-1}$ exists.[2] Indeed,

$$\begin{aligned} \mathbf{AB} = \mathbf{I} &\implies (\mathbf{AB})\mathbf{B}^{-1} = \mathbf{IB}^{-1} \\ &\implies \mathbf{A}(\mathbf{BB}^{-1}) = \mathbf{B}^{-1} \\ &\implies \mathbf{AI} = \mathbf{B}^{-1} \\ &\implies \mathbf{A} = \mathbf{B}^{-1} \\ &\implies \mathbf{BA} = \mathbf{BB}^{-1} \\ &\implies \mathbf{BA} = \mathbf{I}, \end{aligned}$$

as required. □

Combining theorems 1.61 and 1.62, we have the following.

Corollary 1.63. *If* $\mathbf{A}$ *and* $\mathbf{B}$ *satisfy* $\mathbf{AB} = \mathbf{I}$*, then* $\mathbf{A}$ *and* $\mathbf{B}$ *are each other's unique inverse.*

Proof. Indeed, if $\mathbf{AB} = \mathbf{I}$, then $\mathbf{BA} = \mathbf{I}$ by theorem 1.62, and thus

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}.$$

This gives us that $\mathbf{A}$ is an inverse of $\mathbf{B}$, and that $\mathbf{B}$ is an inverse of $\mathbf{A}$. Uniqueness then follows from theorem 1.61. □

Now let us go to 2×2 matrices. We have the following.

Theorem 1.64. *Suppose* $ad - bc \neq 0$*. Then the* 2×2 *matrix* $\mathbf{A} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ *is invertible. Moreover,*

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

[2] It can be shown that $\mathbf{B}^{-1}$ always exists whenever $\mathbf{AB} = \mathbf{I}$, but we need more tools before we can do so. It is given as an exercise in future sections.

Proof. We have

$$\begin{aligned}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\left(\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\right) &= \frac{1}{ad-bc}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \\ &= \frac{1}{ad-bc}\begin{pmatrix} ad-bc & -ab+ab \\ cd-cd & -bc+ad \end{pmatrix} \\ &= \frac{1}{ad-bc}\begin{pmatrix} ad-bc & 0 \\ 0 & -bc+ad \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I},\end{aligned}$$

which by corollary 1.63 completes the proof. □

Note. Just as we required $a \neq 0$ for 1×1 matrices to be invertible, here we require $ad - bc \neq 0$. This special number, $ad - bc$, is called the *determinant* of the matrix $\mathbf{A}$, which we denote by $|\mathbf{A}|$ or by $\det \mathbf{A}$. There are analogous numbers for general $n \times n$ matrices which we explore later, together with their algebraic and geometric significance.

Example 1.65. Suppose $\mathbf{X} = \left(\begin{smallmatrix} 2 & -1 \\ 5 & 3 \end{smallmatrix}\right)$. Then

$$\mathbf{X}^{-1} = \frac{1}{|\mathbf{X}|}\begin{pmatrix} 3 & -(-1) \\ -5 & 2 \end{pmatrix} = \frac{1}{2 \cdot 3 - (-1) \cdot 5}\begin{pmatrix} 3 & 1 \\ -5 & 2 \end{pmatrix} = \frac{1}{11}\begin{pmatrix} 3 & 1 \\ -5 & 2 \end{pmatrix}.$$

Example 1.66. We solve the equation $\mathbf{AX} + \mathbf{B} = \mathbf{C}$ for $\mathbf{X}$, where

$$\mathbf{A} = \begin{pmatrix} 2 & 7 \\ 4 & -3 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 4 \\ 3 & 1 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 10 & 29 \\ 4 & 0 \end{pmatrix}.$$

Indeed,

$$\begin{aligned}\mathbf{AX} + \mathbf{B} = \mathbf{C} &\implies \mathbf{AX} = \mathbf{C} - \mathbf{B} \\ &\implies \mathbf{A}^{-1}\mathbf{AX} = \mathbf{A}^{-1}(\mathbf{C} - \mathbf{B}) \\ &\implies \mathbf{X} = \mathbf{A}^{-1}(\mathbf{C} - \mathbf{B}).\end{aligned}$$

Therefore

$$\mathbf{X} = \frac{1}{|\mathbf{A}|}\begin{pmatrix} -3 & -7 \\ -4 & 2 \end{pmatrix}\left[\begin{pmatrix} 10 & 29 \\ 4 & 0 \end{pmatrix} - \begin{pmatrix} 1 & 4 \\ 3 & 1 \end{pmatrix}\right]$$

$$= -\frac{1}{34}\begin{pmatrix}-3 & -7\\ -4 & 2\end{pmatrix}\begin{pmatrix}9 & 25\\ 1 & -1\end{pmatrix}$$

$$= -\frac{1}{34}\begin{pmatrix}-34 & -68\\ -34 & -102\end{pmatrix}$$

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

Remark 1.67. Most of the algebra we carry out to solve matrix equations is analogous to the algebra with real numbers which we are used to. The two main differences are the lack of division, where we instead use inverses, and the lack of commutativity. Notice that in the second implication of example 1.66, we multiplied both sides by $\mathbf{A}^{-1}$. In particular, we multiplied on the left. This is different than multiplying on the right! So in general, if we have the equation $\text{LHS} = \text{RHS}$, we can *premultiply* by a matrix to get $\mathbf{A}\,\text{LHS} = \mathbf{A}\,\text{RHS}$, or *postmultiply* to get $\text{LHS}\,\mathbf{A} = \text{RHS}\,\mathbf{A}$. But we <u>cannot</u> do $\mathbf{A}\,\text{LHS} = \text{RHS}\,\mathbf{A}$ or $\text{LHS}\,\mathbf{A} = \mathbf{A}\,\text{RHS}$.

Notation (Matrix Power)**.** For any square matrix $\mathbf{A}$, we define the matrix power $\mathbf{A}^n$ by the following recursive definition.

$$\mathbf{A}^n = \begin{cases} \mathbf{A}\mathbf{A}^{n-1} & \text{if } n > 0\\ \mathbf{I} & \text{if } n = 0\\ (\mathbf{A}^{-1})^{-n} & \text{if } n < 0, \end{cases}$$

so for example, $\mathbf{A}^3 = \mathbf{AAA}$, $\mathbf{A}^0 = \mathbf{I}$ and $\mathbf{A}^{-2} = \mathbf{A}^{-1}\mathbf{A}^{-1}$.

Exercise 1.68.

1. Calculate

$$\begin{pmatrix}1 & 8 & 4\\ 2 & 9 & 5\\ -3 & 13 & 10\end{pmatrix} + \begin{pmatrix}3 & 0 & 12\\ 2 & 7 & 3\\ 3 & 3 & 14\end{pmatrix},$$

giving your answer in the form $k\mathbf{A}, k \in \mathbb{N}$.

2. Consider the following matrices with their sizes given below:

$\mathbf{A}$	$\mathbf{B}$	$\mathbf{C}$	$\mathbf{D}$	$\mathbf{E}$	$\mathbf{F}$
4×3	3×3	4×4	2×5	3×2	4×1

Determine the size of:

a) $\mathbf{AB}$ b) $\mathbf{BA}^\intercal$ c) $\mathbf{D}^\intercal\mathbf{E}$

d) $\mathbf{F}^\intercal\mathbf{CAB}$ e) $(\mathbf{BE}+\mathbf{E})^\intercal$ f) $\mathbf{F}^\intercal\mathbf{CF}$

3. Calculate:

a) $\begin{pmatrix} 0 & -8 & 2 & 1 \\ 1 & -6 & 1 & 9 \\ 3 & -2 & 4 & 5 \\ 8 & 2 & 5 & 1 \\ 7 & 2 & 7 & 4 \end{pmatrix} \begin{pmatrix} 1 & 8 & 7 & 3 & 2 \\ 1 & 1 & 0 & 1 & 2 \\ 3 & 1 & -4 & -3 & 5 \\ -6 & 2 & 3 & 5 & -1 \end{pmatrix}$

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

4. Given the matrices $\mathbf{A} = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{smallmatrix}\right)$, $\mathbf{B} = \left(\begin{smallmatrix} 2 & 0 \\ 7 & 5 \\ -3 & 8 \end{smallmatrix}\right)$, $\mathbf{C} = \left(\begin{smallmatrix} 2 & 1 \\ 0 & 5 \end{smallmatrix}\right)$ and $\mathbf{D} = \left(\begin{smallmatrix} 1 \\ 4 \\ 9 \end{smallmatrix}\right)$, find where possible:

a) $\mathbf{AB}$ b) $\mathbf{BA}$ c) $\mathbf{CB}$

d) $\mathbf{ABC}-\mathbf{B}$ e) $\mathbf{AD}+\mathbf{D}$ f) $\mathbf{C}^{-1}$

g) $\mathbf{BC}+\mathbf{BC}^{-1}$ h) $\mathbf{DC}$ i) $4\mathbf{A}^3$

j) $\mathbf{D}^\intercal\mathbf{A}$ k) $\mathbf{B}^\intercal\mathbf{A}$ l) $(\mathbf{A}^\intercal-\mathbf{A})^\intercal$

m) $\mathbf{AA}^\intercal$ n) $\mathbf{B}^\intercal\mathbf{A}^\intercal-\mathbf{AB}^\intercal$ o) $\mathbf{D}^\intercal\mathbf{D}$

p) $(\mathbf{AA}^\intercal)^\intercal$ q) $\frac{1}{2}(\mathbf{A}-\mathbf{A}^\intercal)$ r) $\mathbf{B}^\intercal\mathbf{AB}$

5. A matrix $\mathbf{A}$ is said to be *symmetric* if $\mathbf{A}^\intercal = \mathbf{A}$ and *skew-symmetric* if $\mathbf{A}^\intercal = -\mathbf{A}$. Let $\mathbf{A}$ be any $n \times n$ square matrix. Prove that:

a) The matrix $\mathbf{S} = \frac{1}{2}(\mathbf{A}+\mathbf{A}^\intercal)$ is always symmetric.

b) The matrix $\mathbf{V} = \frac{1}{2}(\mathbf{A}-\mathbf{A}^\intercal)$ is always skew-symmetric.

c) Any square matrix can be split as the sum of a symmetric and a skew-symmetric matrix. (Hint: use $\mathbf{S}$ and $\mathbf{V}$).

6. Given that

$$\mathbf{A} = \begin{pmatrix} -1 & 4 & 1 \\ 2 & -4 & 7 \\ -3 & 6 & -9 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 6 & -42 & -32 \\ 3 & -12 & -9 \\ 0 & 6 & 4 \end{pmatrix},$$

find $\mathbf{AB}$ and deduce $\mathbf{A}^{-1}$.

7. Theorem 1.64 gives us the implication

$$ad - bc \neq 0 \implies \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \text{ has an inverse.}$$

Prove that the converse is also true, that is, if the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ has an inverse, then it must be that $ad - bc \neq 0$. This way, we get that for any 2×2 matrix $\mathbf{A}$, "$\mathbf{A}$ is invertible" is equivalent to $|\mathbf{A}| \neq 0$. [i.e., we get $ad - bc \neq 0 \iff \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ has an inverse.]

8. Prove that for any two invertible square matrices $\mathbf{A}$ and $\mathbf{B}$, if the product $\mathbf{AB}$ exists, it is invertible, and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

9. Invert the following matrices.

   a) $\frac{1}{2}\left(\begin{smallmatrix} -4 & 2 \\ 3 & -1 \end{smallmatrix}\right)$ b) $\left(\begin{smallmatrix} 9 & -2 \\ -4 & 1 \end{smallmatrix}\right)$ c) $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ d) $\left(\begin{smallmatrix} 2 & 2 \\ 2 & -1 \end{smallmatrix}\right)$

10. Explain why $\left(\begin{smallmatrix} 1 & 2 \\ 4 & 8 \end{smallmatrix}\right)$ is singular.

11. Given the matrices $\mathbf{A} = \left(\begin{smallmatrix} 2 & 1 \\ 3 & -1 \end{smallmatrix}\right)$, $\mathbf{B} = \left(\begin{smallmatrix} 3 \\ 2 \end{smallmatrix}\right)$ and $\mathbf{C} = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$, solve the equation $\mathbf{A}\mathbf{x} + \mathbf{B} = 10\,\mathbf{C}$.

12. Consider the matrices $\mathbf{A} = \left(\begin{smallmatrix} 1 & 2 \\ 5 & 0 \end{smallmatrix}\right)$, $\mathbf{B} = \left(\begin{smallmatrix} 7 & -3 \\ 1 & 4 \end{smallmatrix}\right)$ and $\mathbf{C} = \left(\begin{smallmatrix} 6 & -8 \\ 2 & 3 \end{smallmatrix}\right)$. Find the inverses $\mathbf{A}^{-1}$, $\mathbf{B}^{-1}$, $\mathbf{C}^{-1}$ and the product $\mathbf{ABC}$. Hence verify that $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$, and prove that the result holds in general for any three invertible matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ (of the same size).

13. A diagonal matrix is a square matrix with entries in the diagonal, and zeros everywhere else. Prove that, in general, the product of two diagonal matrices (where it exists) is another diagonal matrix. Hence invert the matrix $\left(\begin{smallmatrix} \frac{1}{3} & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$.

14. The matrix $\mathbf{A}$ is given by $\left(\begin{smallmatrix} 1 & -1 & 2 \\ 3 & 2 & 5 \\ 5 & 1 & -3 \end{smallmatrix}\right)$.

   a) Find $\mathbf{A}^2$ and $\mathbf{A}^3$. Express $\mathbf{A}^3 + \lambda\mathbf{A} + \mu\mathbf{I}$ as a single 3×3 matrix.

b) Find values of λ and μ such that $\mathbf{A}^3 + \lambda\mathbf{A} + \mu\mathbf{I} = \mathbf{O}$, where $\mathbf{O}$ is the 3×3 zero matrix. Hence, express $\mathbf{A}^{-1}$ as a single 3×3 matrix.

15. A matrix $\mathbf{P}(k)$ is given by $\mathbf{P}(k) = \left(\begin{smallmatrix} k & 2 \\ k-6 & k-5 \end{smallmatrix}\right)$ for $k \in \mathbb{R}$.

    a) Determine the values of k for which $\mathbf{P}(k)$ has no inverse. What is such a matrix called?

    b) Find, in terms of k, the inverse matrix $\mathbf{P}^{-1}(k)$ for when k is not equal to any of the values found in part (a).

2 Computational Methods

In this section we demonstrate how matrices can help us to solve systems of linear equations in multiple variables, and then explore the notions of determinants and inverses for matrices larger than 2×2.

2.1 Systems of Linear Equations

A *system of linear equations* (or a *linear system*) is a collection of linear equations involving the same set of variables. For example,

$$\begin{cases} 5x - 2y = 7 \\ 7x + 3y = 4 \end{cases}$$

is a linear system involving the variables x and y. The use of the word "system" indicates that the equations are to be considered collectively, rather than individually. This is also why we use a curly bracket ({) to group the equations together notationally.

A *solution* to a linear system is an assignment of the variables, x and y in this case, such that all the equations are simultaneously satisfied. A solution to the system above is given by the assignment $x = 1 = -y$.

Notice that the system above can be written as

$$\begin{cases} (5, -2) \cdot (x, y) = 7 \\ (7, 3) \cdot (x, y) = 4 \end{cases}$$

where $\cdot$ denotes the dot product of vectors. Even more concisely, the definition of matrix multiplication (and matrix equality) gives us that the system is equivalent to the matrix equation

$$\begin{pmatrix} 5 & -2 \\ 7 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 7 \\ 4 \end{pmatrix}.$$

We call the matrix $\mathbf{A} = \left(\begin{smallmatrix} 5 & -2 \\ 7 & 3 \end{smallmatrix}\right)$ the *matrix of coefficients*, the vector $\boldsymbol{x} = (x, y)$ the *solution vector*, and $\boldsymbol{b} = (7, 4)$ the *vector of constant terms*. Thus this system is simply a matrix equation of the form $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$.

Indeed, this is true in general for any linear system of equations:

$$\begin{aligned}
\text{System} &\iff \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \\
\begin{cases} ax + by = \alpha \\ cx + dy = \beta \end{cases} &\iff \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \\
\begin{cases} ax + by + cz = \alpha \\ dx + ey + fz = \beta \\ gx + hy + iz = \gamma \end{cases} &\iff \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}, \\
\begin{cases} ax + by + cz + dw = \alpha \\ ex + fy + gz + hw = \beta \\ ix + jy + kz + lw = \gamma \\ mx + ny + oz + pw = \delta \end{cases} &\iff \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix}, \\
&\vdots
\end{aligned}$$

We can also have a different number of equations from variables, e.g.:

$$\begin{cases} au + bv + cw + dx = \alpha \\ eu + fv + gw + hx = \beta \end{cases} \iff \begin{pmatrix} a & b & c & d \\ e & f & g & h \end{pmatrix} \begin{pmatrix} u \\ v \\ w \\ x \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix},$$

$$\begin{cases} ax + by = \alpha \\ cx + dy = \beta \\ ex + fy = \gamma \end{cases} \iff \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix},$$

and in general,

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_n = b_2 \\ \qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \iff (a_{ij})_{m\times n}(x_i)_n = (b_i)_m.$$

However for now we focus on the case where the number of equations is equal to the number of variables (and therefore we have a square $n \times n$ matrix of coefficients).

What is the advantage of representing linear systems in this way? The answer is simple: it reduces the problem of solving the system to finding the matrix inverse, since

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{b} \implies \boldsymbol{x} = \mathbf{A}^{-1}\boldsymbol{b},$$

assuming that the matrix of coefficients $\mathbf{A}$ is invertible. For 2×2 matrices, this is equivalent to requiring that $|\mathbf{A}| \neq 0$ (by exercise 1.68.7). Indeed, let us try and solve the system

$$\begin{cases} 5x - 2y = 7 \\ 7x + 3y = 4 \end{cases}$$

which we gave initially, by inverting the matrix of coefficients. Since this can be written as

$$\begin{pmatrix} 5 & -2 \\ 7 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 7 \\ 4 \end{pmatrix},$$

we have

$$\begin{aligned}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5 & -2 \\ 7 & 3 \end{pmatrix}^{-1}\begin{pmatrix} 7 \\ 4 \end{pmatrix} &= \frac{1}{5\cdot 3 - (-2)\cdot 7}\begin{pmatrix} 3 & 2 \\ -7 & 5 \end{pmatrix}\begin{pmatrix} 7 \\ 4 \end{pmatrix} \\ &= \frac{1}{29}\begin{pmatrix} 3\cdot 7 + 2\cdot 4 \\ -7\cdot 7 + 5\cdot 4 \end{pmatrix} \\ &= \frac{1}{29}\begin{pmatrix} 29 \\ -29 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix},\end{aligned}$$

and indeed the solution is $x = 1 = -y$.

Remark 2.1. We will prove later that each system of n linear equations in n variables (thus $n \times n$ matrix of coefficients) has a unique solution if and only if the determinant of its matrix of coefficients is non-zero.

Example 2.2. Consider the matrices $\mathbf{A} = \begin{pmatrix} -3 & 2 & 0 \\ -1 & 1 & -3 \\ 7 & 0 & 7 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} -7 & -14 & -6 \\ 14 & 21 & 9 \\ 7 & 14 & 5 \end{pmatrix}$. We find the product $\mathbf{AB}$ and hence solve the system

$$\begin{cases} -3x + 2y \qquad = 3 \\ -x + \ y - 3z = -1 \\ 7x \qquad + 7z = 0 \end{cases}.$$

Indeed, we have

$$\mathbf{AB} = \begin{pmatrix} -3 & 2 & 0 \\ -1 & 1 & -3 \\ 7 & 0 & 7 \end{pmatrix} \begin{pmatrix} -7 & -14 & -6 \\ 14 & 21 & 9 \\ 7 & 14 & 5 \end{pmatrix} = \begin{pmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 7 \end{pmatrix} = 7\mathbf{I}.$$

Now observe that the system we have is $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{x} = (x, y, z)$ and $\boldsymbol{b} = (3, -1, 0)$. Thus we can find the solution vector $\boldsymbol{x}$ since $\boldsymbol{x} = \mathbf{A}^{-1}\boldsymbol{b}$, and

$$\mathbf{AB} = 7\mathbf{I} \implies \tfrac{1}{7}\mathbf{AB} = \mathbf{I} \implies \mathbf{A}(\tfrac{1}{7}\mathbf{B}) = \mathbf{I}$$

and thus $\mathbf{A}^{-1} = \frac{1}{7}\mathbf{B}$. Therefore

$$\boldsymbol{x} = \mathbf{A}^{-1}\boldsymbol{b} = (\tfrac{1}{7}\mathbf{B})\boldsymbol{b} = \frac{1}{7}\begin{pmatrix} -7 & -14 & -6 \\ 14 & 21 & 9 \\ 7 & 14 & 5 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \\ -1 \end{pmatrix},$$

and thus the solution is $x = 1$, $y = -3$ and $z = -1$.

Exercise 2.3. 1. Solve the following systems of equations by inverting the matrix of coefficients.

a) $\begin{cases} 2x - 5y = -21 \\ 4x + 3y = 23 \end{cases}$ b) $\begin{cases} 5x + 4y = 40 \\ 3x - 9y = -33 \end{cases}$

c) $\begin{cases} 3x - 6y = -3 \\ 5x - 6y = 7 \end{cases}$ d) $\begin{cases} x + 9y = 34 \\ 4x - 5y = 13 \end{cases}$

e) $\begin{cases} 6x & -3y = 3 \\ 4x & -3y = -5 \end{cases}$ f) $\begin{cases} 6x & +4y = 65 \\ 6x & +8y = 86 \end{cases}$

g) $\begin{cases} 9x & +8y = 42 \\ 3x & -2y = 0 \end{cases}$ h) $\begin{cases} -9x & +8y = 4 \\ 3x & +5y = 37 \end{cases}$

2. The matrix $\mathbf{A} = \begin{pmatrix} -3 & 0 & 3 \\ -6 & 4 & 1 \\ 1 & -2 & -1 \end{pmatrix}$ has inverse $\frac{1}{15k}\begin{pmatrix} -k & 2\ell & -12 \\ -5 & k+\ell+1 & 5\ell \\ 4k & -6 & -12 \end{pmatrix}$. Find the values of k and ℓ, and hence or otherwise, solve the system of equations

$$\begin{cases} -u & & +w = 2 \\ -6u & +4v & +w = 5 \\ -u & +2v & +w = 6 \end{cases}.$$

Ans: x=1, y=2, z=3

3. Consider the matrix $\mathbf{M} = \begin{pmatrix} 1 & 4 & 0 \\ 5 & -2 & -1 \\ -5 & 0 & 1 \end{pmatrix}$.

   a) Determine constants λ, μ such that $\mathbf{M}^3 = \lambda\mathbf{M} + \mu\mathbf{I}$.

   b) Hence, determine $\mathbf{M}^{-1}$ and solve the system of equations

$$\begin{cases} x & +4y & & = 5 \\ 5x & -2y & -z & = 3 \\ -5x & & +z & = -5 \end{cases}.$$

   Ans: x=y=1, z=0

Notation. $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ is shorthand for $\left|\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\right| = \det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$.

4. Consider the lines $\ell_1, \ell_2 \subseteq \mathbb{R}^2$ whose respective equations are $ax + by = k$ and $cx + dy = m$. Show that ℓ_1 and ℓ_2 are parallel if and only if

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = 0.$$

5. (Cramer's Rule). Let $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\boldsymbol{b} = (\alpha, \beta)$. Prove that the solution $\boldsymbol{x} = (x, y)$ of the system of equations corresponding to

the equation $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ is given by the equations

$$x = \begin{vmatrix} \alpha & b \\ \beta & d \end{vmatrix} \Big/ \det(\mathbf{A}) \qquad \text{and} \qquad y = \begin{vmatrix} a & \alpha \\ c & \beta \end{vmatrix} \Big/ \det(\mathbf{A}).$$

2.2 Elementary Row Operations

Elementary row operations are simple operations one can carry out on the rows of a given matrix. There are three such operations.

I. **Row switching**.
Interchanging row i with row j, denoted by writing $R_i \leftrightarrow R_j$.

For example,

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{pmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 7 & 8 & 9 \end{pmatrix}.$$

II. **Row scaling**.
Multiplying row i by a non-zero scalar $\lambda \in \mathbb{R}$, denoted by writing $\lambda R_i \to R_i$.

For example,

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \xrightarrow{5R_2 \to R_2} \begin{pmatrix} 1 & 2 & 3 \\ 20 & 25 & 30 \\ 7 & 8 & 9 \end{pmatrix}.$$

III. **Row adding**.
Replacing row i by the sum of itself with a scalar multiple of another row j, where $j \neq i$, denoted by writing $R_i + \lambda R_j \to R_i$.

For example,

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \xrightarrow{R_3 + (-3)R_1 \to R_3} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 4 & 2 & 0 \end{pmatrix}.$$

We focus on these three operations in particular because it turns out that carrying out an elementary row operation on a matrix $\mathbf{A}$ can be achieved simply by pre-multiplying by some other matrix $\mathbf{E}$. Such matrices, that is, matrices which carry out elementary row operations when multiplied on the left, are called *elementary* matrices.

Definitions 2.4 (Elementary Matrices)**.** An *elementary matrix* is an $n \times n$ matrix $\mathbf{E}$ which falls under one of the following definitions.

(i) The *swap matrix* $\mathbf{S}_{k\ell}$ is an $n \times n$ matrix defined by

$$\mathbf{S}_{k\ell} = \big(\delta_{ij}(1-\delta_{ik})(1-\delta_{i\ell}) + \delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}\big),$$

where δ_{ij} denotes the Kronecker-delta.

(ii) The *row-scaling matrix* is an $n \times n$ matrix $\mathbf{L}_k(\lambda)$ defined by

$$\mathbf{L}_k(\lambda) = \big(\delta_{ij}(1 + (\lambda - 1)\delta_{ik})\big).$$

(iii) The *row-adding matrix* is an $n \times n$ matrix $\mathbf{R}_{k\ell}(\lambda)$, $k \neq \ell$, defined by

$$\mathbf{R}_{k\ell}(\lambda) = \big(\delta_{ij} + \lambda\delta_{ik}\delta_{j\ell}\big).$$

We then have that these matrices behave as we wish them to:

Theorem 2.5 (Elementary Row Operations)**.** *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. Then*

(i) The resulting matrix after applying the row operation $R_i \leftrightarrow R_j$ *to* $\mathbf{A}$ *is given by* $\mathbf{S}_{ij}\mathbf{A}$*.*

(ii) Let $\lambda \in \mathbb{R}$ *be a non-zero scalar. The resulting matrix after applying the row operation* $\lambda R_i \to R_i$ *to* $\mathbf{A}$ *is given by* $\mathbf{L}_i(\lambda)\mathbf{A}$*.*

(iii) Let $\lambda \in \mathbb{R}$ *be a scalar. The resulting matrix after applying the row operation* $R_i + \lambda R_j \to R_i$ *to* $\mathbf{A}$ *is given by* $\mathbf{R}_{ij}(\lambda)\mathbf{A}$*.*

The proof of this fact is a straightforward expansion of the definitions each of the matrices, and the definition of matrix multiplication (1.47), similar to that of theorem 1.58. We leave it as an exercise.

Remark 2.6. Even though the matrices given in definitions 2.4 may seem complicated when expressed in terms of δ's, they are actually equivalent to the matrices obtained by applying the corresponding elementary row operation to the identity matrix.

For example, a 4×4 $\mathbf{S}_{24}$ matrix (which corresponds to $R_2 \leftrightarrow R_4$) is simply the 4×4 identity matrix with rows 2 and 4 interchanged:

$$\mathbf{S}_{24} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

a 3×3 $\mathbf{L}_2(6)$ matrix (which corresponds to $6R_2 \to R_2$) is simply the 3×3 identity matrix with row 2 multiplied by 6:

$$\mathbf{L}_2(6) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and a 5×5 $\mathbf{R}_{53}(-2)$ matrix (which corresponds to $R_5 + (-2)R_3 \to R_5$) is simply the 5×5 identity matrix with -2 times row 3 added to row 5:

$$\mathbf{R}_{53}(-2) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -2 & 0 & 1 \end{pmatrix}.$$

Exercise 2.7. Prove, using the definitions in definitions 2.4 and the definition of matrix multiplication (1.47), show that

a) $\mathbf{S}_{ij}{}^{-1} = \mathbf{S}_{ij}$ b) $\mathbf{L}_i(\lambda)^{-1} = \mathbf{L}_i(1/\lambda)$

2.3 Determinants

We have already seen that the determinant of the 2×2 matrix $\mathbf{A} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is given by

$$|\mathbf{A}| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

In exercise 1.68.7, we have seen that the determinant of a matrix being non-zero is equivalent to the matrix being invertible. Moreover, in exercise 2.3.4, we arrived to this conclusion with geometric intuition about lines in two dimensions ($\mathbb{R}^2$).

For 3×3 matrices, the determinant also exists, but before we get to introducing it, we need some definitions.

Definition 2.8 (Submatrix)**.** Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ matrix. We associate an $(n-1) \times (n-1)$ matrix with each entry a_{ij} of $\mathbf{A}$, called the *submatrix* of a_{ij} in $\mathbf{A}$, denoted by $\mathbf{A}_{ij}$, obtained simply by deleting row i and column j from $\mathbf{A}$. In other words, we have

$$\mathbf{A}_{k\ell} = \big(a_{(i+[i\geqslant k])(j+[j\geqslant \ell])}\big)_{(n-1)\times(n-1)},$$

where $[\phi]$ denotes the *Iverson bracket*.[3]

Example 2.9. Suppose $\mathbf{A} = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{smallmatrix}\right)$. Then

$$\mathbf{A}_{12} = \begin{pmatrix} \square & \square & \square \\ 4 & \square & 6 \\ 7 & \square & 9 \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix}, \qquad \mathbf{A}_{22} = \begin{pmatrix} 1 & \square & 3 \\ \square & \square & \square \\ 7 & \square & 9 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix},$$
$$\text{and } \mathbf{A}_{33} = \begin{pmatrix} 1 & 2 & \square \\ 4 & 5 & \square \\ \square & \square & \square \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}.$$

In addition to the submatrix, we associate a sign ($+$ or $-$) with each entry in a matrix $\mathbf{A}$. These signs follow a chequerboard-like pattern, starting from $+$ in the top-left corner:

$$\begin{pmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \qquad \begin{pmatrix} \square & \blacksquare & \square & \cdots \\ \blacksquare & \square & \blacksquare & \cdots \\ \square & \blacksquare & \square & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Indeed, it is not hard to see that the sign corresponding to the entry a_{ij} is $+$ if $i+j$ is even, and $-$ if $i+j$ is odd. So we can nicely express the sign corresponding to the entry a_{ij} simply as $(-1)^{i+j}$. Now we are ready to give the most important definition before going on to introduce the determinant:

[3] The Iverson bracket is a notation defined by

$$[\phi] = \begin{cases} 1 & \text{if } \phi \text{ is true} \\ 0 & \text{otherwise,} \end{cases}$$

where ϕ is a statement which can be true or false. In this case, we as using it to add 1 to the matrix indices i and j so that we "skip over" the column/row we are deleting.

Definition 2.10 (Cofactor)**.** Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ matrix. The *cofactor* corresponding to the entry a_{ij}, denoted $\operatorname{co}(a_{ij})$, is defined by

$$\operatorname{co}(a_{ij}) = (-1)^{i+j} \det(\mathbf{A}_{ij}).$$

In other words, the cofactor of a_{ij} is the determinant of the submatrix $\mathbf{A}_{ij}$ paired with the entry's corresponding sign (as in the chequerboard-like pattern above).

Example 2.11. Again, suppose $\mathbf{A} = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{smallmatrix}\right)$. Then

$$\operatorname{co}(a_{23}) = (-1)^{2+3} \begin{vmatrix} 1 & 2 & \square \\ \square & \square & \square \\ 7 & 8 & \square \end{vmatrix} = -\begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} = -(1 \cdot 8 - 2 \cdot 7) = -(-6) = 6,$$

and

$$\operatorname{co}(a_{31}) = (-1)^{3+1} \begin{vmatrix} \square & 2 & 3 \\ \square & 5 & 6 \\ \square & \square & \square \end{vmatrix} = +\begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} = 2 \cdot 6 - 3 \cdot 5 = -3.$$

Now we are ready to introduce the determinant for 3×3 matrices, and, shortly after, for any $n \times n$ matrix. The definition we give is due to Laplace.

Definition 2.12 (3×3 Determinant)**.** The *determinant* of the 3×3 matrix $\mathbf{A} = (a_{ij})$ is defined by

$$|\mathbf{A}| = \sum_{k=1}^{3} a_{1k} \operatorname{co}(a_{1k}) = a_{11} \operatorname{co}(a_{11}) + a_{12} \operatorname{co}(a_{12}) + a_{13} \operatorname{co}(a_{13}).$$

In other words, we are defining the determinant of a 3×3 as the sum of entries of the first row, each multiplied by their corresponding cofactor.

Example 2.13. Again take $\mathbf{A} = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{smallmatrix}\right)$. We find $|\mathbf{A}|$.

$$\begin{aligned} |\mathbf{A}| &= \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} \\ &= 1 \cdot \left(+ \begin{vmatrix} \square & \square & \square \\ \square & 5 & 6 \\ \square & 8 & 9 \end{vmatrix} \right) + 2 \cdot \left(- \begin{vmatrix} \square & \square & \square \\ 4 & \square & 6 \\ 7 & \square & 9 \end{vmatrix} \right) + 3 \cdot \left(+ \begin{vmatrix} \square & \square & \square \\ 4 & 5 & \square \\ 7 & 8 & \square \end{vmatrix} \right) \end{aligned}$$

$$= \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix}$$

$$= 5 \cdot 9 - 6 \cdot 8 - 2(4 \cdot 9 - 6 \cdot 7) + 3(4 \cdot 8 - 5 \cdot 7)$$
$$= -3 - 2(-6) + 3(-3) = 0.$$

Although we have not yet proved that it is the case for 3×3 matrices, a zero determinant is in fact equivalent to having a non-invertible matrix. Thus we have that **A** is not invertible.

Example 2.14 (General formula). Here we will derive a general formula for the 3×3 determinant, in the style of $\left|\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right| = ad - bc$ for 2×2 matrices. Memorising it is not recommended!

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \left(+ \begin{vmatrix} \square & \square & \square \\ \square & e & f \\ \square & h & i \end{vmatrix} \right) + b \left(- \begin{vmatrix} \square & \square & \square \\ d & \square & f \\ g & \square & i \end{vmatrix} \right) + c \left(+ \begin{vmatrix} \square & \square & \square \\ d & e & \square \\ g & h & \square \end{vmatrix} \right)$$

$$= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

$$= aei - afh - bdi + bfg + cdh - ceg.$$

We can now give the general definition of the determinant.

Definition 2.15 (Determinant). Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ matrix. Then the *determinant of* **A**, denoted $\det(\mathbf{A})$ or $|\mathbf{A}|$, is the number defined by

$$\det(\mathbf{A}) = \begin{cases} a_{11} & \text{if } n = 1 \\ \displaystyle\sum_{j=1}^{n} a_{1j} \operatorname{co}(a_{1j}) & \text{otherwise.} \end{cases}$$

Notice that is a straightforward generalisation of the 3×3 case, which works for both smaller and larger matrices. Indeed, for $n = 1, 2$ and 4, the definition gives

$$|a| = a$$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = a \left(+ \begin{vmatrix} \square & \square \\ \square & d \end{vmatrix} \right) + b \left(- \begin{vmatrix} \square & \square \\ c & \square \end{vmatrix} \right)$$

$$= a\,|d| - b\,|c| = ad - bc$$

$$\begin{aligned}
\begin{vmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{vmatrix} &= a\left(+ \begin{vmatrix} \square & \square & \square & \square \\ \square & f & g & h \\ \square & j & k & l \\ \square & n & o & p \end{vmatrix} \right) + b\left(- \begin{vmatrix} \square & \square & \square & \square \\ e & \square & g & h \\ i & \square & k & l \\ m & \square & o & p \end{vmatrix} \right) \\
&\quad + c\left(+ \begin{vmatrix} \square & \square & \square & \square \\ e & f & \square & h \\ i & j & \square & l \\ m & n & \square & p \end{vmatrix} \right) + d\left(- \begin{vmatrix} \square & \square & \square & \square \\ e & f & g & \square \\ i & j & k & \square \\ m & n & o & \square \end{vmatrix} \right) \\
&= a\begin{vmatrix} f & g & h \\ j & k & l \\ n & o & p \end{vmatrix} - b\begin{vmatrix} e & g & h \\ i & k & l \\ m & o & p \end{vmatrix} + c\begin{vmatrix} e & f & h \\ i & j & l \\ m & n & p \end{vmatrix} - d\begin{vmatrix} e & f & g \\ i & j & k \\ m & n & o \end{vmatrix} \\
&= a\left(f\begin{vmatrix} \square & \square & \square \\ \square & k & l \\ \square & o & p \end{vmatrix} + \cdots \right) + \cdots - d\left(\cdots + g\begin{vmatrix} \square & \square & \square \\ i & j & \square \\ m & n & \square \end{vmatrix} \right) \\
&= dgjm - chjm - dfkm + bhkm + cflm - bglm - dgin + chin \\
&\quad + dekn - ahkn - celn + agln + dfio - bhio - dejo + ahjo \\
&\quad + belo - aflo - cfip + bgip + cejp - agjp - bekp + afkp.
\end{aligned}$$

As is clear from the 4×4 case, large determinants become too laborious to work out by hand. Indeed, as is illustrated, a 4×4 determinant requires four 3×3 determinants to be worked out, each of which in turn require three 2×2 determinants to work out, meaning that a 4×4 determinant requires twelve 2×2 determinants to be worked out. (It's not hard to see that in general, an $n \times n$ determinant requires the computation of $\frac{n!}{2}$ 2×2 determinants.)

It seems strange that the definition of the determinant specifically involves the first row of the matrix $\mathbf{A}$ (notice only a_{1j} appears in the definition). Is there something inherently important about the first row of a matrix? It turns out that the answer is no: we can perform the sum of entries times their cofactor in any row and the result will be the same determinant! This is stated below.

Theorem 2.16 (Laplace Expansion)**.** *Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ matrix, $n \geqslant 2$, and pick any row $i \in \{1, \ldots, n\}$. Then*

$$\det(\mathbf{A}) = \sum_{j=1}^{n} a_{ij} \operatorname{co}(a_{ij}).$$

The proof of this result requires elementary row operations, so we revisit it later.

Example 2.17. Here we illustrate the advantage of this result. Suppose we wish to evaluate the determinant

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 2 & 0 \\ 5 & 9 & 2 & 1 \\ 3 & 2 & 1 & 1 \end{vmatrix}.$$

It would be a lot easier if we were allowed to expand the determinant along the second row instead of the first row as we have been doing so far, because that would give

$$\begin{aligned} \begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 2 & 0 \\ 5 & 9 & 2 & 1 \\ 3 & 2 & 1 & 1 \end{vmatrix} &= 0\left(-\begin{vmatrix} \square & 2 & 3 & 4 \\ \square & \square & \square & \square \\ \square & 9 & 2 & 1 \\ \square & 2 & 1 & 1 \end{vmatrix}\right) + 0\left(+\begin{vmatrix} 1 & \square & 3 & 4 \\ \square & \square & \square & \square \\ 5 & \square & 2 & 1 \\ 3 & \square & 1 & 1 \end{vmatrix}\right) \\ &\quad + 2\left(-\begin{vmatrix} 1 & 2 & \square & 4 \\ \square & \square & \square & \square \\ 5 & 9 & \square & 1 \\ 3 & 2 & \square & 1 \end{vmatrix}\right) + 0\left(+\begin{vmatrix} 1 & 2 & 3 & \square \\ \square & \square & \square & \square \\ 5 & 9 & 2 & \square \\ 3 & 2 & 1 & \square \end{vmatrix}\right) \\ &= -2\begin{vmatrix} 1 & 2 & 4 \\ 5 & 9 & 1 \\ 3 & 2 & 1 \end{vmatrix} = \cdots = 130. \end{aligned}$$

Thus thanks to theorem 2.16, we can evaluate this 4×4 determinant by working out one 3×3 determinant instead of four!

Remark 2.18. In general, by expanding along the row with the most zeros we optimise the amount of computations. Always be aware to allocate the correct signs to the cofactors—remember the chequerboard pattern!

Another immediate consequence of this result is the following:

Proposition 2.19. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. If a row of* $\mathbf{A}$ *consists entirely of zeros, then* $|\mathbf{A}| = 0$.

Proof. (the idea: expand along the row consisting solely zeros.) Suppose row i is the row consisting entirely of zeros, i.e. $a_{ij} = 0$ for all $j = 1, \ldots, n$.

If $n = 1$, then there is only one row/column and $a_{ij} = a_{11} = 0$, so that $\det(\mathbf{A}) = \det(a_{11}) = a_{11} = 0$. If $n \geqslant 2$, then by theorem 2.16

$$\det(\mathbf{A}) = \sum_{j=1}^{n} a_{ij}\operatorname{co}(a_{ij}) = \sum_{j=1}^{n} 0 \cdot \operatorname{co}(a_{ij}) = 0,$$

as required. □

It also turns out that we have the following result which we prove later using elementary row operations.

Theorem 2.20. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. Then*

$$\det(\mathbf{A}) = \det(\mathbf{A}^{\intercal}).$$

Remark 2.21. Consequently, to find the determinant of a matrix $\mathbf{A}$, we can also choose to find the determinant of the transpose, which involves expanding along some row in $\mathbf{A}^{\intercal}$. But this is equivalent to expanding along some *column* of $\mathbf{A}$. In other words, we can also find the determinant of $\mathbf{A}$ by expanding along a column, which might be useful if some column contains more zeros than any row.

Example 2.22. We evaluate the following determinant by expanding along the third column, since it contains two zeros:

$$\begin{vmatrix} 1 & 2 & -1 & 7 \\ 2 & -4 & 0 & 5 \\ 1 & 9 & 0 & 6 \\ 2 & -6 & 9 & 7 \end{vmatrix}.$$

Indeed,

$$\begin{vmatrix} 1 & 2 & -1 & 7 \\ 2 & -4 & 0 & 5 \\ 1 & 9 & 0 & 6 \\ 2 & -6 & 9 & 7 \end{vmatrix} = -1\left(+\begin{vmatrix} \square & \square & \square & \square \\ 2 & -4 & \square & 5 \\ 1 & 9 & \square & 6 \\ 2 & -6 & \square & 7 \end{vmatrix}\right) + 9\left(-\begin{vmatrix} 1 & 2 & \square & 7 \\ 2 & -4 & \square & 5 \\ 1 & 9 & \square & 6 \\ \square & \square & \square & \square \end{vmatrix}\right)$$

$$= -1\begin{vmatrix} 2 & -4 & 5 \\ 1 & 9 & 6 \\ 2 & -6 & 7 \end{vmatrix} - 9\begin{vmatrix} 1 & 2 & 7 \\ 2 & -4 & 5 \\ 1 & 9 & 6 \end{vmatrix}$$

$$= -1\left(2\begin{vmatrix} 9 & 6 \\ -6 & 7 \end{vmatrix} + 4\begin{vmatrix} 1 & 6 \\ 2 & 7 \end{vmatrix} + 5\begin{vmatrix} 1 & 9 \\ 2 & -6 \end{vmatrix}\right)$$