Unit 2: Vector spaces, matrices, and linear functions

Dr.P.M.Bajracharya

October 9, 2021

Contents

1 Vector spaces, matrices, and linear functions

1.1 Vector spaces

Definition. A nonempty set V is a **vector space** over $\mathbb{R}$ if for every $x, y, z \in V$ and $a, b \in \mathbb{R}$

1. $x + y \in V$ with the following **addition** properties:
   (a) $x + y = y + x$
   (b) $(x + y) + z = y + (x + z)$
   (c) There is an element $0 \in V$ such that $x + 0 = x$
   (d) There is an element $-x \in V$ such that $x + (-x) = 0$
2. $ax \in V$ with the following **scalar multiplication** properties:
   (a) $a(x + y) = ax + ay$
   (b) $(a + b)x = ax + bx$
   (c) $a(bx) = (ab)x$
   (d) $1x = x$.

The elements of a vector space are called **vectors**.

Using only these axioms, one can show that

- The element 0, called the **zero vector** in Axiom 1 (c) is unique.
- The element $-x$, called the **negative** of x, in Axiom 1 (d) is unique for each x in V.

1.2 Subspaces of a vector space

In many problems, a vector space consists of an appropriate subset of vectors from some larger vector space. In this case, only three of the ten vector space axioms need to be checked; the rest are automatically satisfied.

Definition. Let S be a non-empty subset of a vector space V. The set S is called a **subspace** of V if S is closed under the same vector addition and scalar multiplication as V, that is, for all $x, y \in S$ and for all $a \in \mathbb{R}$

$$x + y \in S \text{ and } ax \in S.$$

1.3 Examples of subspaces and not-subspaces in two dimensions

Example 1.

The L_2 ball $B(0; 1) = \{x \in \mathbb{R}^2 : \ \|x\|_2 \leq 1\}$ is not a subspace. In fact,

- It is not closed under addition: for example,

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \in B(0; 1),$$

  but their sum

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \notin B(0; 1).$$

- Also, it is not closed under scalar multiplication: for example,

$$2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \notin B(0; 1).$$

The same is true for the L_2 sphere. Similarly, L_1 and L_∞ balls and spheres are not subspaces of $\mathbb{R}^2$.

Example 2.

The positive orthant is not a subspace. In fact, it is not closed under scalar multiplication, for example: the vector $\begin{pmatrix}1\\0\end{pmatrix}$ is in positive orthant (first quadrant), but

$$-2\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}-2\\0\end{pmatrix}$$

has negative entry and so, it is not in prositive orthant (first quadrant).

Remark 1. The positive orthant is, however, closed under multiplication of nonnegative scalars, a fact that is sometimes of interest.

Example 3.

The positive simplex is not a subspace. In fact, it is not closed under the addition of two vectors, and it is nor closed under scalar multiplication.

It is, however, closed under the special class of vector additions of the form $z = ax + by$, where x, y, z are vectors, and a, b are numbers such that $a, b \geq 0$ and $a + b = 1$. We will see later why this is of interest.

Example 4.

The line L given by the equation $x_1 + x_2 = 1$ is not a subspace. For example, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \in L$, but

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \notin L.$$

Examples of subspaces.

Here are examples of sets that are subspaces of $\mathbb{R}^2$.

Example 5.

The set $\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$ is a "trivial" subspace of $\mathbb{R}^2$ since: if we multiply it by any scalar or add it to itself, then the output is still the 0 vector

Example 6.

All of R2. We can see that all of $\mathbb{R}^2$ is a "trivial" subspace of $\mathbb{R}^2$ since: $\mathbb{R}^2$ is a vector space; and $\mathbb{R}^2$ is a subset of itself.

Example 7.

The line $x_2 = ax_1 + b$ (perhaps more familiar as $y = ax + b$) is not a subspace $\mathbb{R}^2$ for $b \neq 0$.

Proof.

Let u, v be points on the line $x_2 = ax_1 + b$ and

$$u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

Then

$$u_2 = au_1 + b$$
$$v_2 = av_1 + b.$$

Consider the point

$$w = u + v = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \end{pmatrix}.$$

Then we have

$$u_2 + v_2 = a(u_1 + v_1) + 2b.$$

This line is not on the line $x_2 = ax_1 + b$, the intercepts being different.

Example 8.

The line $x_2 = ax_1$ (in usual notations, $y = ax$) is a subspace $\mathbb{R}^2$.

Proof.

Let u, v be two points on the line $x_2 = ax_1$ and

$$u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

Then

$$u_2 = au_1$$
$$v_2 = av_1.$$

Consider the point

$$w = u + v = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \end{pmatrix}.$$

Then we have

$$u_2 + v_2 = a(u_1 + v_1).$$

This line is on the line $x_2 = ax_1$.

Example 9.

The set of points that is the union of two lines through the origin is not a subspace.

Proof.

Let

$$\Omega_1 = \{(x_1, x_2) : \ x_1 = ax_2)\},$$
$$\Omega_2 = \{(y_1, y_2) : \ y_1 = ay_2)\}.$$

Then $\Omega = \Omega_1 + \Omega_2$ is not a subspace of $\mathbb{R}^2$. In fact, the set Ω is not closed under addition of two vectors, since if you add two vectors from two different lines lead to a vector that is not on either of those lines (except in the degenerate case when the two lines are the same).

Remark 2. The last two examples indicate that the union of two subspaces may not be a subspace.

1.4 Subspaces in $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$, and beyond

There are two types of subspaces of $\mathbb{R}$:

$$\mathbb{R}, \quad \{0\}.$$

The one-dimensional $\mathbb{R}$ is itself a vector space; and it has a one-dimensional subspace (itself) and a zero dimensional subspace (the origin). Both of these are "trivial," in the sense that there is not too much interesting going on, and so one typically does not spend much time discussing the subspace aspects of $\mathbb{R}$, but it's good to understand such "extreme cases" in the definitions of vector spaces and subspaces.

The examples of the previous section suggest (correctly, as we discussed above) that there are three kinds of subspaces for $\mathbb{R}^2$:

- $\mathbb{R}^2$ itself is a two-dimensional subspace of $\mathbb{R}^2$.
- A line through the origin is a subspace of dimension 1, and it takes 1 number to specify a point on a line.
- The singleton set $\{0\}$ is a subspace of dimension 0.

We see that there are four kinds of subspaces for $\mathbb{R}^3$:

- $\mathbb{R}^3$ is a three-dimensional subspace of $\mathbb{R}^3$.
- A plane through the origin is a two-dimensional subspace of $\mathbb{R}^3$.
- A line through the origin is a one-dimensional subspace of $\mathbb{R}^3$.
- The set $\{0\}$ is a zero-dimensional subspace of $\mathbb{R}^3$.

2 Dot product between two vectors.

Definition. If $x = (x_1, x_2)$ and $y = (y_1, y_2))$ are points on the plane $\mathbb{R}^2$, then the dot product or inner product between those two vectors is

$$x \cdot y = x_1 y_1 + x_2 y_2 = \sum_1^2 x_i y_i.$$

The following relation establishes the relationship between the dot poduct and the L_2 norm:

$$\|x\|_2 = \sqrt{x \cdot x} = \left(\sum_1^2 x_i^2 \right)^{1/2}.$$

We can easily generalize the dot product on $\mathbb{R}^2$ to $\mathbb{R}^n$.

Definition. If $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n))$ are points in $\mathbb{R}^n$, then the dot product or inner product between those two vectors is

$$x \cdot y = x_1 y_1 + ... + x_n y_n = \sum_1^n x_i y_i.$$

The following relation establishes the relationship between the dot poduct and the L_2 norm on $\mathbb{R}^n$:

$$\|x\|_2 = \sqrt{x \cdot x} = \left(\sum_1^n x_i^2 \right)^{1/2}.$$

The dot product allows us to define the perpendicularity (or orthogonality).

Definition. Let x and y be two vectors in $\mathbb{R}^n$. We say that x is **perpendicular** (or **orthogonal**) to y, if

$$x \cdot y = 0.$$

Definition. Let x be vector in $\mathbb{R}^n$. The set of vectors perpendicular to a vector $x \in \mathbb{R}^n$ is denoted by $x^\perp$. Thus,

$$x^\perp = \{y \in \mathbb{R}^n : \ x \cdot y = 0\}.$$

Here is an example of how one tries to prove things more generally about subspaces.

Often we would like to prove that certain things we construct or define are or are not subspaces. Note that we have already done this: we defined the positive orthant, and we proved that it is not a subspace; and we defined lines through the origin on the plane, and we proved that they are subspaces. Moreover, we would like to do this assuming as little as possible. As a simple example of this, consider the following.

Theorem 2.1. *Given a vector $x \in \mathbb{R}^2$, such that $x \neq 0$, let $x^\perp$ be the set of vectors that are perpendicular to x. Then, $x^\perp$ is a subspace.*

Proof.

If $x = 0$, then $x^\perp = \mathbb{R}^2$, which we know is a subspace.

Suppose that $x \neq 0$. Let $u, v \in x^\perp$. Then

$$\begin{aligned} x \cdot (u + v) &= x_1(u_1 + v_1) + x_2(u_2 + v_2) \\ &= (x_1u_1 + x_2u_2) + (x_1v_1 + x_2v_2) = 0. \end{aligned}$$

Therefore,

$$u + v \in x^\perp.$$

Now, if $a \in \mathbb{R}$, then

$$\begin{aligned} x \cdot (au) &= x_1(au_1) + x_2(au_2) \\ &= a(x_1u_1 + x_2u_2) = 0. \end{aligned}$$

Therefore,

$$au \in x^{\perp}.$$

Thus, by definition, $x^{\perp}$ is a subspace of $\mathbb{R}^2$.

2.1 Basis vectors

Definition. Let $\{v_1, v_2, ..., v_k\}$ be a set of vectors in a vector space V. The span of $v_1, v_2, ..., v_k$ is denoted by $\operatorname{span}(\{v_1, v_2, ..., v_k\})$ and is defined by

$$\operatorname{span}(\{v_1, v_2, ..., v_k\}) = \left\{ \sum_{1}^{k} a_i v_i : \text{ each } a_i \in \mathbb{R} \right\}$$

Theorem 2.2. *The span of a set* $\operatorname{span}(\{v_1, v_2, ..., v_k\})$ *is a subspace of a vector space* V.

Proof.

It is clear that $\operatorname{span}(\{v_1, v_2, ..., v_k\}) \subseteq V$. Let us demonstrate that it is closed under addition and scalar multiplication.

Addition: Let $x, y \in \operatorname{span}(\{v_1, v_2, ..., v_k\})$. Then we can write

$$x = \sum_{1}^{k} a_i v_i, \quad y = \sum_{1}^{k} b_i v_i,$$

where $a_i, b_i \in \mathbb{R}$. Then

$$x + y = \sum_1^k (a_i + b_i) v_i \in \operatorname{span}(\{v_1, v_2, ..., v_k\}).$$

Scalar multiplication: Let $a \in \mathbb{R}$. Then

$$\begin{aligned} ax &= a \sum_1^k a_i v_i \\ &= \sum_1^k (a a_i) v_i \in \operatorname{span}(\{v_1, v_2, ..., v_k\}). \end{aligned}$$

Therefore, $\operatorname{span}(\{v_1, v_2, ..., v_k\})$ is a subspace of V.

Standard basis vectors

Definition. The **standard basis vectors** (or **standard unit vectors**) in $\mathbb{R}^n$, denoted e_k, have n entries, with a 1 in the kth position and a 0 in all the other positions.

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad k\text{th position}$$

Figure 1: kth standard basis vector

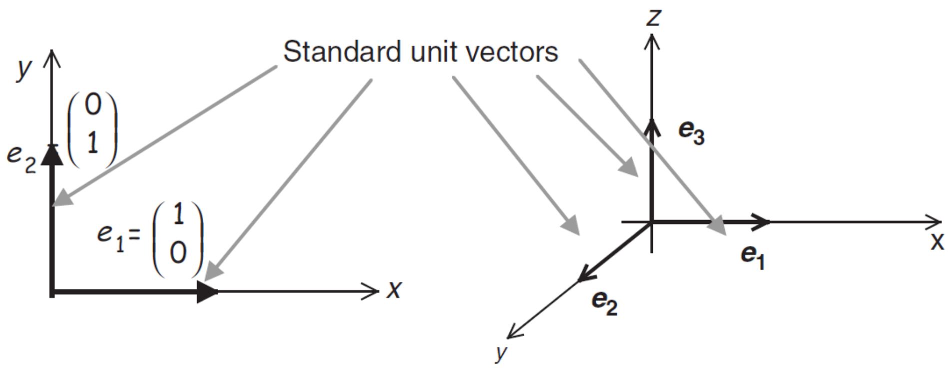


Figure 2: Standard basis vectors in the plane and in the space.

Why are these standard basis vectors important?

- Consider a vector $x = \begin{pmatrix}2\\3\end{pmatrix} \in \mathbb{R}^2$. It can be written as

$$\begin{pmatrix}2\\3\end{pmatrix} = 2\begin{pmatrix}1\\0\end{pmatrix} + 3\begin{pmatrix}0\\1\end{pmatrix}.$$

  This gives

$$x = x_1e_1 + x_2e_2,$$

  where $x_1 = 2$ and $x_2 = 3$. Thus,

  > We have expressed x as a *linear combination* of the basis vectors e_1 and e_2, and the coefficients involved in the linear combination are the components of the vector x.

- Further, consider a linear combination:

$$x_1e_1 + x_2e_2,$$

  Now the question is:

  Under what conditions is this linear combination zero?

that is,

When does the equality $x_1e_1 + x_2e_2 = 0$ hold?

To answer to this question, put

$$x = x_1e_1 + x_2e_2.$$

Then

$$x = x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x_1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

If $x_1e_1 + x_2e_2 = 0$, then

$$x = 0 \Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0 \Rightarrow x_1 = x_2 = 0.$$

On the otherhand,

$$x_1 = x_2 = 0 \Rightarrow x_1e_1 + x_2e_2 = 0.$$

Therefore,

$$\boxed{x_1e_1 + x_2e_2 = 0 \Leftrightarrow x_1 = x_2 = 0.}$$

3 Home Work

1. Show that any vector in $\mathbb{R}^3$ can be expressed as a linear combination of the three unit basis vectors in $\mathbb{R}^3$. Also, show that a linear combination of the three unit basis vectors in $\mathbb{R}^3$ equals to 0 if and only if all coefficients in the linear combination are zeros.

2. Do the above problem for $\mathbb{R}^n$.

The above discussion motivates the following definitions:

Definition. Vectors $v_1, v_2, ..., v_n$ in a vector space V are called **linearly independent vector**s if

$$\sum_1^k a_i v_i = 0 \Rightarrow a_i = 0 \text{ for all } i \in \{1, ..., k\}.$$

Now, we come to the following definition:

Definition. A set $\{v_1, v_2, ..., v_n\}$ of vectors in a vector space V is called a **basis** for V if

1. The vectors $v_1, v_2, ..., v_n$ are linearly independent.
2. Every vector in V can be expressed as a linear combination of the vectors $v_1, v_2, ..., v_n$.

Not-standard basis vectors.

If we rotate e_1 by the angle θ, then

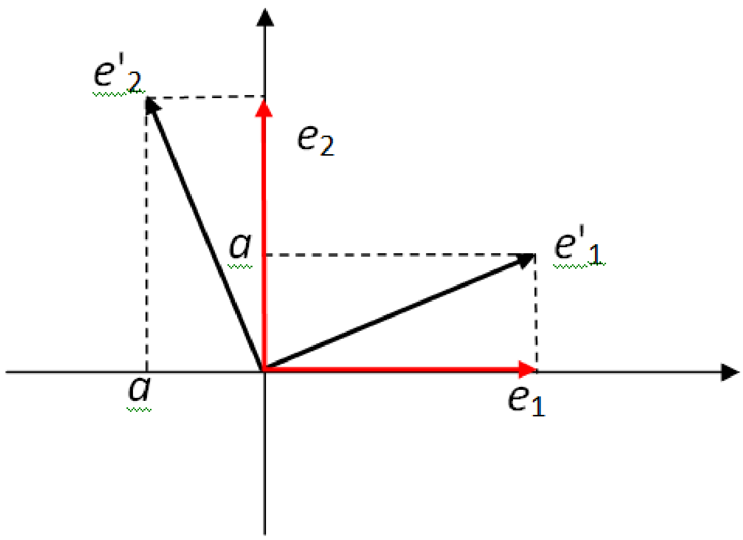
$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \to \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \to \begin{pmatrix} 1 \\ \tan\theta \end{pmatrix} = \begin{pmatrix} 1 \\ a \end{pmatrix} = e_1'.$$

If we rotate e_2 by the same angle θ, then

$$e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \to \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} \to \begin{pmatrix} -\tan\theta \\ 1 \end{pmatrix} = \begin{pmatrix} -a \\ 1 \end{pmatrix} = e_2'.$$

Note that just as $e_1 e_2 = 0$, since we have rotated both vectors by the same angle, so too $e_1' e_2' = 0$, i.e., e_1' and e_2' are also perpendicular.

With respect to these new basis vectors: any point on the line can be described by one number (the magnitude along e_1', and in the same way we might be able to consider only one coordinate axis, here we might be able to consider only one of the new coordinate axes e_1' and ignore the other e_2' and still be able to do something useful with the data.

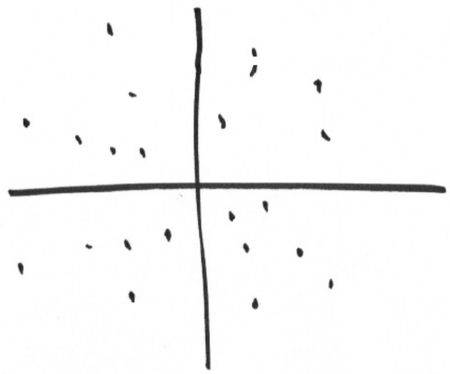


Summarizing this discussion, there are several points to note.

- The vector $\begin{pmatrix} 1 \\ a \end{pmatrix}$ seems more natural to describe the data in Figure (c) given below.
- Using this more natural description it takes 1 number rather than 2 numbers.
- The line through the origin defined by $\begin{pmatrix} 1 \\ a \end{pmatrix}$ – as well as the line through the origin defined by the $\begin{pmatrix} -a \\ 1 \end{pmatrix}$ perpendicular to it – is a one-dimensional subspace of $\mathbb{R}^2$.

In the same way that any point on the plane can be expressed in terms of the standard basis vectors, so too any point on the plane can be expressed in terms of the two vectors $\begin{pmatrix} 1 \\ a \end{pmatrix}$ and $\begin{pmatrix} -a \\ 1 \end{pmatrix}$. We will study this later in detail.

Usefulness of basis in data science.

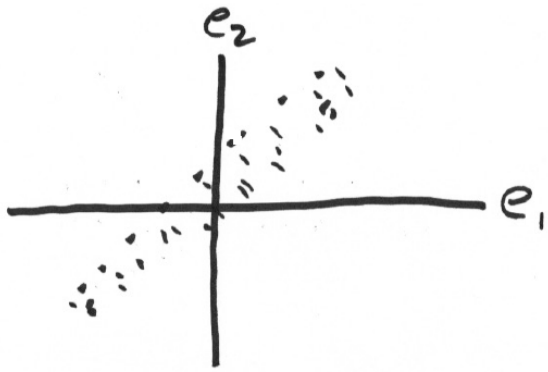


(a) Data points on a two dimensional plane, scattered in a round manner.

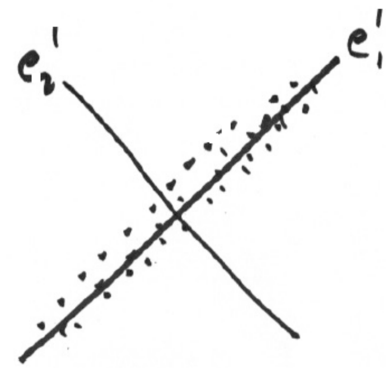
(b) Data points on a two dimensional plane, scattered in an elongated manner.

In Figure (a), different data points have different values for the two variables, x_1 and x_2, but both variables seem important to capture properties of the data.

In Figure (b), the elongation is along the x-axis. It could be due to one feature being more "important" in some sense. That means, most of the information of interest is captured by x_1, while x_2 might be less important or simply random noise. In this case, we might hope or expect get very similar results by considering only (x_1), rather than (x_1, x_2), for each data point.



(c) Data points on a two dimensional plane, scattered in a different elongated manner.



(d) Same data points on a two dimensional plane, scattered in an elongated manner, but with rotated axes.

In Figure (c), on the other hand, the data are elongated along some other direction. As in Figure (b), there is one direction on the plane that seems more "important" than the other perpendicular direction on the plane. However, this direction lies, for example, on the line $x_2 = ax_1$.

By changing the axes, the data set plotted in Figure (c) can be visualized as in Figure(d), Now, the information from the data set can be obtained as in Figure (b) with respect to new axes e_1' and e_2'.

Summarizing this discussion, there are several points here.

- The vector $e_1' = \begin{pmatrix} 1 \\ a \end{pmatrix}$ seems more natural to describe the data in Figure (c).
- Using this more natural description it takes 1 number rather than 2 numbers.

- The line through the origin defined by $\begin{pmatrix} 1 \\ a \end{pmatrix}$ as well as the line through the origin defined by the $\begin{pmatrix} -a \\ 1 \end{pmatrix}$ perpendicular to it – is a one dimensional subspace of $\mathbb{R}^2$.

4 Matrices and operations on matrices

Definition 1. An $m \times n$ matrix is a rectangular array of entries, m high and n wide, i.e., with m horizontal rows and n vertical columns.

Of greatest interest is when the elements of a matrix are real numbers, but they could be other things, e.g., *Boolean values*, *integers*, *complex numbers*, *polynomials*, *other matrices*, etc.

Note that, from this perspective, vectors and numbers are simple matrices.

- A vector $x \in \mathbb{R}^m$, viewed as a column vector, is an $m \times 1$ matrix.
- Alternatively, a vector $x \in \mathbb{R}^n$, viewed as a row vector, is a $1 \times n$ matrix.
- A number $x \in \mathbb{R}$ is a 1×1 matrix.

4.1 Two operations similar to vector operations

Addition of two matrices: Let A and B be two matrices of the same size, say, two $m \times n$ matrices. The sum matrix

$C = A + B$ is a matrix of the same size that has entries

$$C_{ij} = A_{ij} + B_{ij}$$

for $i \in \{1, 2, ..., m\}$ and $j \in \{1, 2, ..., n\}$.

Multiplication of a matrix by a scalar: Let A be an $m \times n$ matrix and $\alpha \in \mathbb{R}$. The product matrix $B = \alpha A$ is a matrix of the same size that has entries

$$B_{ij} = \alpha A_{ij}$$

for $i \in \{1, 2, ..., m\}$ and $j \in \{1, 2, ..., n\}$.

With these two operations on matrices, the set of all $m \times n$ matrices,

$$V = \{A : \ A \text{ is an } m \times n \text{ matrix}\},$$

is a vector space.

4.2 Two more operations on matrices

Multiplication Here we present two ways one can define the product of two matrices.

1. **Hadamard product of matrices.** This is a matrix product that is defined for any two matrices of the same size. Let A and B be two $m \times n$ matrices. In this case, the Hadamard matrix product $C = A \circ B$ is a matrix of the same size as matrices A and B that has entries

   $$C_{ij} = B_{ij} A_{ij}$$

   for $i \in \{1, 2, ..., m\}$ and $j \in \{1, 2, ..., n\}$.

Example.

If $A = \begin{pmatrix} 1 & 2 & 3 \\ -2 & 4 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 2 & 1 \\ 4 & 1 & 7 \end{pmatrix}$, then

$$\begin{aligned} A \circ B &= \begin{pmatrix} 1 & 2 & 3 \\ -2 & 4 & 0 \end{pmatrix} \circ \begin{pmatrix} 3 & 2 & 1 \\ 4 & 1 & 7 \end{pmatrix} \\ &= \begin{pmatrix} 1 \times 3 & 2 \times 2 & 3 \times 1 \\ -2 \times 4 & 4 \times 1 & 0 \times 7 \end{pmatrix} \\ &= \begin{pmatrix} 3 & 4 & 3 \\ -8 & 4 & 0 \end{pmatrix}. \end{aligned}$$

2. **Matrix multiplication.** This is a much more useful notion of the product of two matrices. Let A be an $m \times n$ matrix and B an $n \times p$ matrix. Then the product $C = AB$ is an $m \times p$ matrix with elements

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

for $i \in \{1, 2, ..., m\}$ and $j \in \{1, 2, ..., n\}$.

Row and column names. Let us adopt the following notations for row and column names.

$$\begin{array}{c} \\ A_{1:} \\ A_{2:} \\ A_{3:} \end{array} \begin{array}{c} \begin{array}{cccc} A_{:1} & A_{:2} & A_{:3} & A_{:4} \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} \end{array}$$

Remark:

Requirement: To exists the product matrix AB, the number of columns in the first matrix A must be the same as the number of rows in the second matrix B.

Each element C_{ij} of the product matrix AB is determined as the dot product of $A_{i:}$ and $B_{:j}$:

$$C_{ij} = A_{i:} \cdot B_{:j} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

for $i \in \{1, 2, ..., m\}$ and $j \in \{1, 2, ..., n\}$.

Transposition: Given an $m \times n$ matrix A, the transpose A^T of A is the matrix of the size $n \times m$ obtained by interchanging the rows and columns, i.e.,

$$A = (A_{ij}) \Rightarrow A^T = (A_{ji}).$$

Here are two things that are good to know about transposes.

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

Examples of matrix multiplication

- If $A = \begin{pmatrix} 2 & -1 \\ 3 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 4 & -2 \\ 3 & 0 & 2 \end{pmatrix}$, then find AB.
  In this case, BA is not defined. In fact, the number of columns in B is 3. So, it is not equal to the number of rows in the second matrix A which is equals to 2.
- If $A = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}, B = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, C = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix}$ and $D = \begin{pmatrix} 1 & 0 \\ 2 & 2 \\ 1 & 1 \end{pmatrix}$, then find

$$AB,\ BA,\ AC,\ CD,\ DC.$$

CA is not defined. Why ?

It is worthwhile that this example illustrates several things about the product of two matrices:

1. $AB \neq BA$. Hence matrix multiplication is not **commutative** in general.
2. AC is defined, but CA is not defined, so both need not be defined, and in particular both are not defined unless $m = p$ in the definition of matrix multiplication.
3. if both are defined, then their dimensions need not be the same and are not unless $m = n = p$.

Theorem 4.1. *Let A be an $m \times n$ matrix, B an $n \times p$ matrix, and C a $p \times q$ matrix, so that $(AB)C$ and $A(BC)$ are defined. Then $(AB)C = A(BC)$.*

Proof. Recall that $(AB)C$ is a matrix with $m \times q$ elements, indexed as $\alpha \in \{1, 2, ..., m\}$ and $\beta \in \{1, 2, ..., q\}$. Let's consider the $(\alpha\beta)$ element:

$$
\begin{aligned}
((AB)C)_{\alpha\beta} &= \sum_{l=1}^{p} (AB)_{\alpha l} C_{l\beta} && (1)\\
&= \sum_{l=1}^{p} \left(\sum_{k=1}^{n} A_{\alpha k} B_{kl} \right) C_{l\beta} && (2)\\
&= \sum_{l=1}^{p} \sum_{k=1}^{n} A_{\alpha k} B_{kl} C_{l\beta} && (3)\\
&= \sum_{k=1}^{n} \sum_{l=1}^{p} A_{\alpha k} B_{kl} C_{l\beta} && (4)\\
&= \sum_{k=1}^{n} A_{\alpha k} \left(\sum_{l=1}^{p} B_{kl} C_{l\beta} \right) && (5)\\
&= \sum_{l=1}^{p} A_{\alpha k} (BC)_{k\beta} && (6)\\
&- A(BC). && (7)
\end{aligned}
$$

$\square$

5 Functions, linear functions, and linear transformations

5.1 Functions and transformations

Recall the following definition.

Definition. A function from a set X into a set Y is a rule f that associates each element of X with a unique element of Y.

We express this by writing

$$f:\ X \to Y.$$

We are going to be interested in functions whose domains and ranges are $\mathbb{R}^m$ or $\mathbb{R}^n$.

For example, if we fix $a \in \mathbb{R}^n$, then the dot product $a \cdot x = f(x)$ transforms a vector $x \in \mathbb{R}^n$ into a number $a \cdot x \in \mathbb{R}$. There are lots of functions we could have between vector spaces, but the following will be a particularly important class of functions.

Definition. Let V be a vector subspace of a vector space over $\mathbb{R}$ (such as $\mathbb{R}^n$). Then a linear function is a mapping f such that

1. $\forall\, x, y \in V \quad f(x+y) = f(x) + f(y),$
2. $\forall\, x \in V,\ \forall\, \alpha \in \mathbb{R} \quad f(\alpha x) = \alpha f(x).$

Examples of nonlinear functions.

- A function given by the equation $y_1 = (x_1 - 2)^2 + 3$ is not linear.
- A function given by the equation $y_1 = \sin(x_1)$ is not a linear.
- A function given by the equation $y_1 = ax_1 + b$, where $a, b \in \mathbb{R}$ is not linear, in general.

  The class of functions such as $y_1 = ax_1 + b$ for $b \neq 0$ is a sufficiently important class of functions, however, that it has a special name. It is called an **affine function**, meaning basically a linear function plus a constant. Note that b could be anything, but in data science b often corresponds to something like the mean, and thus subtracting it corresponds to preprocessing that is mean-centering. In this case, linear functions are mean centered, while affine functions are not mean centered. Both are of interest, and they are clearly related, but they are different.

Examples (linear).

1. A function given by the equation $y_1 = ax_1$ is a linear function.

2. The function that takes as input the vector $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and returns as output the vector

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix}.$$

Note that by writing the RHS of this equation as a matrix-vector product, we obtain

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

3. The composition of two (and thus more than two) linear functions is a linear function. For example, if $f(x) = ax$ and $g(x) = bx$ (where $x, a, b, f(x), g(x)$ are all real numbers in $\mathbb{R}$), then

$$f(g(x)) = a(bx) = (ab)x,$$

which is a linear function.

4. Consider the linear function that takes as input the vector $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ and returns as output the vector

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

If we consider the function that first applies the linear function defined in (2) and then applies the linear function just defined, then we have

$$\begin{aligned} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} &= \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ \Rightarrow \qquad \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} &= \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \end{aligned}$$

Clearly, the last equation defines a linear function.

Notice that a linear function can be represented as a matrix. Let us return to the equation considered above:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{8}$$

Put

$$\begin{aligned} x &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \\ y &= \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \\ f &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \end{aligned}$$

We now observe that Equation (8) can be rewritten as follows:

$$y = f(x).$$

In general, if we have a linear function $f : \mathbb{R}^m \to \mathbb{R}^n$, then it can be fully described by an $m \times n$ matrix.

5.2 Inverse matrices

Recall that the multiplicative inverse of a number such as 5 is $1/5$ or 5^{-1}. This inverse satisfies the equations

$$5^{-1} \cdot 5 = 1 \text{ and } 5 \cdot 5^{-1} = 1.$$

The matrix generalization requires both equations and avoids the slanted-line notation (for division) because matrix multiplication is not commutative. Furthermore, a full generalization is possible only if the matrices involved are square matrices.

We need the identity matrix in order to define and explain the inverse matrix. So, we begin with the identity matrix.

What does the term identity matrix mean?

The **identity matrix** is a matrix denoted by I such that

$$AI = A \text{ for any matrix } A$$

What does the identity matrix look like?

An identity matrix is a square matrix defined by

$$I = (I_{kj}) = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases}$$

This means, all the diagonal elements of a matrix I are 1:

$$i_{11} = i_{22} = i_{33} = ... = 1$$

and all the other entries are zero.

Example.

For a 2×2 matrix we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where a, b, c and d are real numbers.

Definition. A square matrix A is said to be **invertibl**e or **non-singular** if there is a matrix B of the same size such that

$$AB = BA = I.$$

Matrix B is called the (multiplicative) **inverse** of A and is denoted by A^{-1}.

Here is a basic result about inverses.

Theorem 5.1. *If A and B are invertible, then AB is invertible, and*

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

Proof. Let's show that $B^{-1}A^{-1}$ gives the identity when acting on the left and right.

$$\begin{aligned} AB(AB)^{-1} &= ABB^{-1}A^{-1} = AIA^{-1} = AA^{-1} = I \\ (AB)^{-1}AB &= B^{-1}A^{-1}AB = B^{-1}IB = B^{-1}B = I, \end{aligned}$$

□

From a linear equation perspective, inverse matrices are important. For example, a system of linear equations can generally be written as $Ax = b$ where x is the vector of unknowns that we need to find. If we multiply both sides of this $Ax = b$ by the inverse matrix A^{-1} we obtain:

$$\begin{aligned} & A^{-1}Ax = A^{-1}b \\ \Rightarrow \qquad & x = A^{-1}b. \end{aligned}$$

Hence we can find the unknowns by finding the inverse matrix.

Question. When does such an inverse matrix exist, i.e., when is a matrix invertible?

Answer. Here is the partial answer.

- Case $m = n = 1$: We can write a 1×1 matrix as $A = (a)$. In this case, $A^{-1} = 1/a$, which is defined for all $a \neq 0$. So, for 1×1 matrices, i.e., real numbers, such a number exists for every $a \in \mathbb{R}$ such that $a \neq 0$. This is particularly simple, and the situation is considerably more subtle for matrices of larger size.

- Case $m = n = 2$: We can write a 2×2 matrix as

  $$\begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

  in which case

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

  which is defined only when $ad - bc \neq 0$, and otherwise the inverse is not defined.

  Check that the above matrix A^{-1} is the inverse of the matrix A, and vice versa

  A 2×2 matrix A doesn't have an inverse in two cases:

1. When it is the all-zeros matrix, in which case it doesn't have any non-trivial information and it sends all input vectors to the zero vector; and
2. when it's two columns are the same, up to scaling, in which case you might imagine that it is missing some information. Note that this degeneracy corresponds to multiplication by a scalar, i.e., one column is a scalar multiple of the other column. Note also that if $\alpha = 0$, then the second column is the all-zeros column, and it still holds that the matrix is not invertible.

The quantity $ad - bc$ is called the **determinant** of $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, and we write

$$\det A = ad - bc.$$

We also write

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

For the case of 3×3 matrices, if

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix},$$

the determinant of this 3×3 matrix is defined by

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

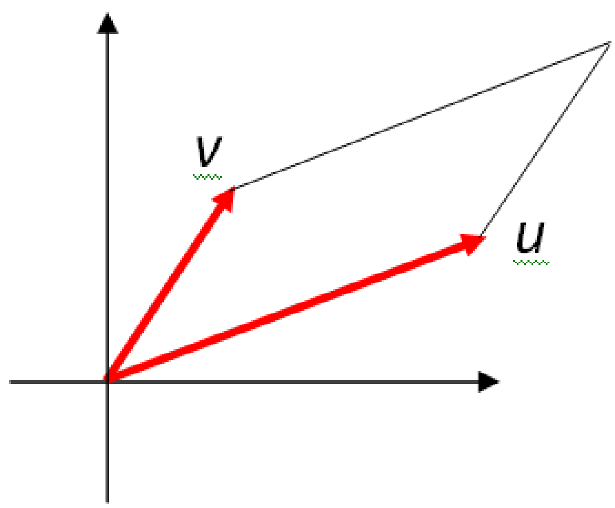
In this way, we can also define the determinant of an $n \times n$ matrix with $n \geq 4$.

From the above discussion, we observe that

det$(\cdot)$ is a real-valued function defined on the set of all square matrices.

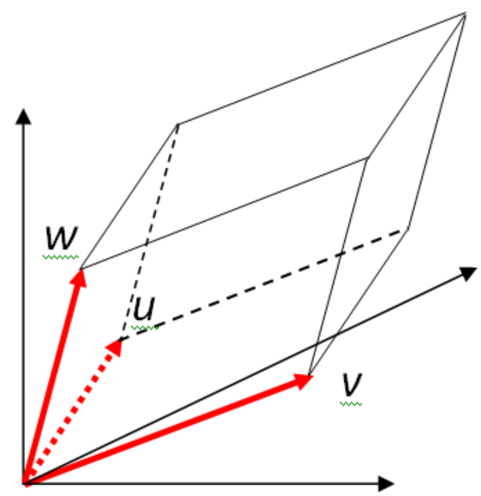
Geometrical interpretation of determinants.

- For a 2×2 matrix, if we consider the parallelogram defined by the two columns (or rows) of the matrix, for example, $x = \begin{pmatrix} a \\ c \end{pmatrix}$ and $y = \begin{pmatrix} b \\ d \end{pmatrix}$, then the determinant equals the area of the parallelogram. Clearly, this equals zero if the



two columns are linearly dependent, i.e., if one is a scalar multiple of the other, and thus they point in the same direction, in which case the parallelogram has no height.

- For a 3×3 matrix, if we consider the parallelepiped defined by the three columns (or rows) of the matrix, then the determinant equals the volume of the parallelepiped. Clearly, this equals zero if the three columns are linearly dependent, i.e., if one is a scalar multiple of another or if



one can be obtained by a linear combination of two others, in which case the parallelepiped has no height.

5.3 Transformations and matrices as transformations

We know that the set of matrices forms a vector space, and so we can add them, multiply them by a scalar, etc.; and we also have the operation of matrix multiplication, which gives the set of matrices additional structure.

Now, let's turn to the question:

What is this additional structure?

Relatedly, what is this matrix multiplication doing?

Relatedly, what does it "mean"?

Relatedly, why is it so useful?

We will consider the answers to these questions.

Dot product. We have viewed this:

- as computing a function of two vectors; and
- as doing a matrix-matrix multiplication.
- as a function that fixes y and takes x as an input. Then $f(x) = f_y(x) = y^T x$, and this is a function that takes as input a vector $x \in \mathbb{R}^n$ and returns as output a number that is an element of $\mathbb{R}$ that equals $\sum_{i=1}^n x_i y_i$, where recall y is assumed to be fixed.

Random walk transition matrix. Here, we take as input a vector $x \in \mathbb{R}^n$ and return a vector $y \in \mathbb{R}^n$, where $\sum_{j=1}^n A_{ij} x_j$. Of course, we could send y in as input and iterate the process, but regardless it is a function that transforms input vector in $\mathbb{R}^n$ and returns a vector in $\mathbb{R}^n$.

This perspective as viewing a matrix in terms of transformations or as a function is true more generally:

> An $m \times n$ matrix can be used to characterize a certain class of functions (**linear transformations**) from $\mathbb{R}^m$ to $\mathbb{R}^n$, and, conversely, every linear transformation can be written as a matrix.

This representation also explains the definition of matrix multiplication: **applying a function twice (as in iterating the random walk matrix) corresponds exactly to doing a multiplication of the two matrices**.

As an example, consider $A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix}$. It can be viewed as a transformation with domain $\mathbb{R}^3$ and range $\mathbb{R}^2$. For example, if

$v = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \in \mathbb{R}^3$, then

$$w = Av = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix} \in \mathbb{R}^2.$$

Thus, the $m \times n$ matrix A transforms vectors from $\mathbb{R}^n$ to $\mathbb{R}^m$ according to the rule

$$w_i = \sum_{j=1}^{n} A_{ij} v_j \qquad \text{for } i = 1, 2, ..., m.$$

This is a linear function.

More abstractly, here is the de

nition of a linear transformation.

Definition. A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a linear transformation if

1. $\forall\, x, y \in \mathbb{R}^n \quad T(x + y) = T(x) + T(y)$,
2. $\forall\, x \in \mathbb{R}^n,\ \forall\, \alpha \in \mathbb{R} \quad T(\alpha x) = \alpha T(x)$.

Observe that $T(0) = 0$. This indicates that a linear transformation goes through the origin.

Definition. An affine transformation is a transformation that is a linear transform plus a constant offset.

In particular, this means that the function would be a linear transformation if it were modified by taking the image of the origin and transforming it back to the origin.

Note that affine transformations give outputs that are first degree polynomials in the inputs, but they could have a term that is a zero order polynomial, i.e., a constant, while linear transformation have only first order terms, e.g., the origin has been shifted and/or variables redefined. In $/bbR$, this is just another way of saying that

$$y_1 = ax_1 + b, \quad \text{for } b \neq 0,$$

is not a linear function, but $y_1 = ax_1$ is a linear function. This holds true more generally.

Here is an example:

$$\begin{aligned}
\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} x_1 - x_2 + 2 \\ 2x_1 + x_2 \end{pmatrix} & \quad \text{is affine} \\
= \begin{pmatrix} x_1 - x_2 \\ 2x_1 + x_2 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix} & \\
= \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix} &
\end{aligned}$$

i.e., $x \to Ax+b$, where the first term is the linear transformation and the second term is the affine offset. Here the constant offset $b = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \in \mathbb{R}^2$ is a vector offset and not a scalar offset.

Here are two important theorems that we won't prove but that the above discussion suggests.

Theorem 5.2. *Let A be an $m \times n$ matrix. Then, A defines a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ by matrix multiplication: $T(v) = Av$, where v is a column vector.*

Theorem 5.3. *Every linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by an $m \times n$ matrix, call it A. The functional form* *is given by $T(v) = Av$, i.e., a matrix -vector multiplication, where th jth column of A is $T(e_j)$..*

The two theorems given above relate matrices and linear transformations more precisely. Due to them, we will sometimes, but not always, distinguish between a linear transformation T and its associated matrix A.

The second theorem is powerful and surprising. It says not just that every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is given by a matrix; it also says that one can construct the matrix by seeing how the transformation acts on the standard basis vectors. This is rather remarkable.

For completeness, we note the following results, which states that if we have a linear transformation corresponding to a matrix, then the inverse linear transformation corresponds to the inverse matrix.

Theorem 5.4. *A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is invertible iff the $m \times n$ matrix A associated with it is invertible, and $T^{-1} = A^{-1}$.*

Note that only square matrices can have inverses. Thus for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ to be invertible, we must have $m = n$.

Multiplying a matrix by a standard basis vector

Multiplying a matrix A by the standard basis vector e_i selects out the ith column of A, as shown in the following example.

Example. (The jth column of A is Ae_j)

Below, we show that the second column of A is Ae_2:

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

$$Ae_2 = \begin{pmatrix} 3 & -2 & 0 \\ 2 & 1 & 2 \\ 0 & 4 & 4 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \\ 4 \\ 0 \end{pmatrix} = A_{:2}.$$

Generalizing it we get

> The jth column of a matrix A is Ae_j, that is,
>
> $$Ae_j = A_{:j}.$$

Example(The jth column of AB is Ab_j). The second column of the product AB is the product of A and the second column of B.

To show it, let

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

Then

$$AB = \begin{pmatrix} 2 & -1 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 4 & -2 \\ 3 & 0 & 2 \end{pmatrix} = \begin{pmatrix} -1 & 8 & -6 \\ 9 & 12 & -6 \end{pmatrix}.$$

Now, we find the product $AB_{:2}$.

$$AB_{:2} = \begin{pmatrix} 2 & -1 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 4 \\ 0 \end{pmatrix} = \begin{pmatrix} 8 \\ 12 \end{pmatrix} = (AB)_{:2}.$$

The next theorem is a key result, in the linear transformation theory underlying matrix operations used in data science and more generally.

Theorem 5.5. *Let* $S : \mathbb{R}^\ell \to \mathbb{R}^m$ *and* $T : \mathbb{R}^n \to \mathbb{R}^\ell$ *be linear transformations, given by matrices* A *and* B*, respectively. Then, the composition* $S \circ T : \mathbb{R}^n \to \mathbb{R}^m$ *is a linear transformation and is given by* AB*.*

Proof.

Let $v, w \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$. Then

$$\begin{aligned}(S \circ T)(av + bw) &= S(T(av + bw)) \\ &= S(aT(v) + bT(w)) \\ &= aS(T(v)) + bS(T(w)) \\ &= a(S \circ T)(v) + b(S \circ T)(w).\end{aligned}$$

This shows that $S \circ T$ is linear. Then it can be represented by a matrix, say, C. For any basis vector e_i of $\mathbb{R}^n$ we then have

$$Ce_i = (S \circ T)(e_i) = S(T(e_i)) = S(Be_i) = A(Be_i) = (AB)e_i.$$

This implies that each column of C is equal to the corresponding column of AB and so,

$$C = AB.$$

Therefore, $S \circ T$ is given by AB.

Symmetric, triangular and diagonal matrices

Symmetric matrices. Symmetric matrices are important in general but are very important in data science. Here is the definition.

Definition. A symmetric matrix is a matrix that equals its transpose, i.e.,

$$A_{ij} = A^T_{ij} = A_{ji}.$$

For example, the matrix

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

is a symmetric matrix, while the matrix

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

is not a symmetric matrix.

Triangular matrices.

Definition. An **upper triangular matrix** is a square matrix with non-zero entries only on or above the diagonal.

An **lower triangular matrix** is a square matrix with non-zero entries only on or below the diagonal.

For example, the matrix

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

is neither upper-triangular nor lower-triangular, the matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 9 \end{pmatrix}$$

is upper triangular but not lower triangular, and the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

is both upper triangular and lower triangular.

Diagonal matrices.

Definition. A diagonal matrix is a square matrix with nonzero entries (if any) only on the main diagonal.

For example, the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

is a diagonal matrix.

Note that a diagonal matrix is both upper triangular and lower triangular.

Diagonal matrices have many nice properties. For example,

$$\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}^k = \begin{pmatrix} A_{11}^k & 0 \\ 0 & A_{22}^k \end{pmatrix}.$$

Due to this property, if $0 < A_{22} < 1$, then we obtain

$$\begin{pmatrix} 1 & 0 \\ 0 & A_{22} \end{pmatrix}^k = \begin{pmatrix} 1 & 0 \\ 0 & A_{22}^k \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{as } k \to \infty.$$

5.4 Examples of matrices as transformations

We saw that every matrix represents a linear transformation, and vice versa. In spite of that, it can be difficult to tell what exactly an arbitrary matrix is "doing" when it is just presented as an array of numbers.

Fortunately, there are some basic examples of matrices that are relatively simple to think about, in the sense that their associated transformations are relatively simple to understand. In some cases, these basic examples form the "building blocks" of arbitrary matrices. When that happens, one can express the arbitrary matrix in terms of those building blocks; these are often called **matrix decompositions**.

Matrix decompositions are useful for two things: They can help

- To understand structure in data (by describing what a matrix is "doing," and thus what is happening in data that are being represented by the matrix, in terms of simpler operations).
- To perform computations faster (by describing a difficult-to-work with matrix in terms of simpler easier-to-implement parts).

Now, we will give several examples.

- **Identity:** The identity transformation $I : \mathbb{R}^n \to \mathbb{R}^n$ is linear and is given by the matrix

$$I_n = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

This is the trivial transformation that doesn't do anything.

- **Scaling:** The transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ that takes any input vector and multiplies it by $a \in \mathbb{R}$ is given by

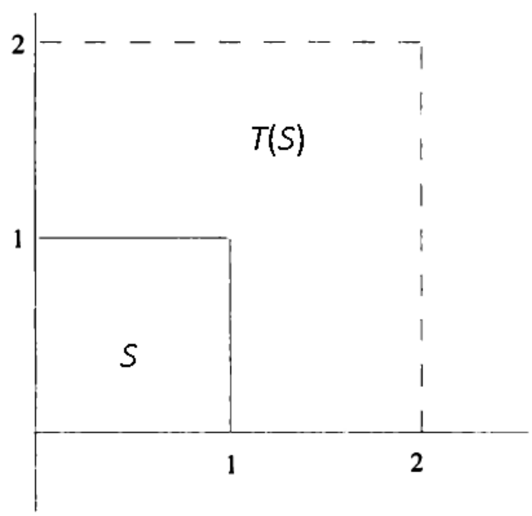
$$A = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$$

For example,

$$Ae_1 = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix},$$

$$Ae_2 = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ a \end{pmatrix}.$$

The following figure shows the result of applying T represented by the matrix $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ to the unit square A.



Note that

$$Ax = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 \\ ax_2 \end{pmatrix} = a \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = ax.$$

- **Stretching:** The transformation T given by $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ with $a \neq b$ is like scaling, except that the scaling is by a different amount in each direction. For example, since

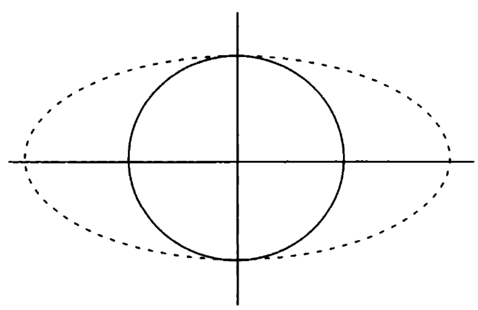
$$Ae_1 = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix},$$

$$Ae_2 = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ b \end{pmatrix},$$

we have that

$$Ax = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 \\ bx_2 \end{pmatrix}.$$

The linear transformation given by the matrix $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ stretches the unit circle into an ellipse.



<u>Note that</u> although we are calling this stretching, if one of the entries is negative, it might be a reflection. For example, if $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, then

$$Ax = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix}.$$

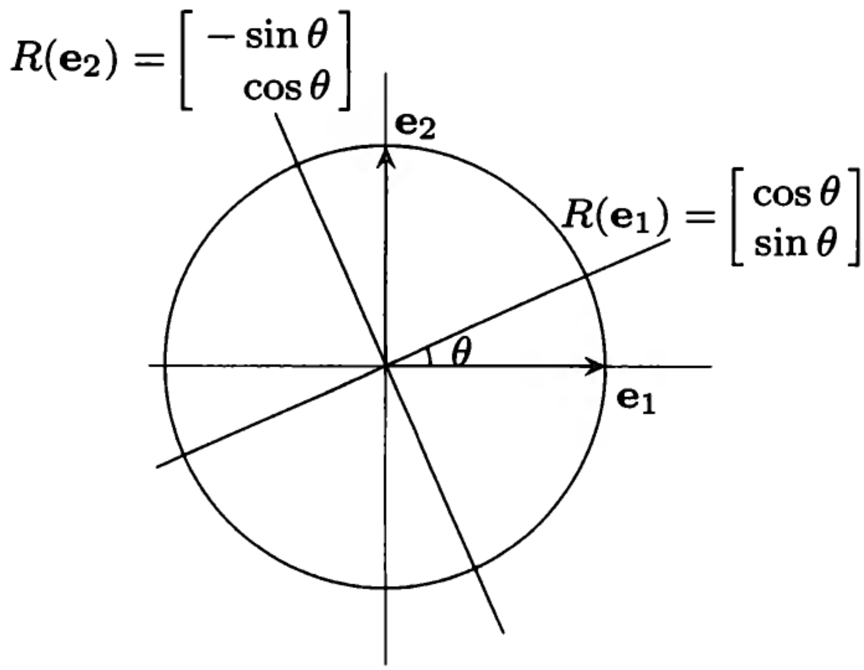
This is a reflection about the line $x_1 = 0$.

- **Rotation:** The transformation given by the matrix $R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ involves rotating by an angle of θ. In particular,

$$Re_1 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix},$$

$$Re_2 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix},$$

See the figure given below.



5.5 Random Walk Matrix

Consider an adjacency matrix:

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

Divide each (vertical) column by the sum of the entries in that column.

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

The resulting matrix is called the **random walk matrix**.

All in all, there will be two complementary interpretation to a lot of what we do.

1. **Linear algebra**: relate to linear functions and linear transformations.
2. **Probability theory**: relate to random walks and random variables.