Linear Algebra Week 2

Felix Breuer

September 2023

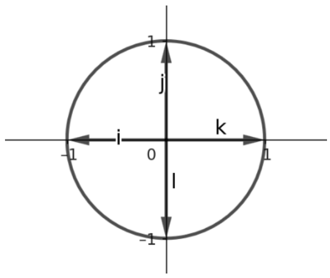
1 Recap

1.1

$$\begin{bmatrix}5\\6\end{bmatrix} \in \mathsf{span}(\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}4\\3\end{bmatrix})?$$

True: We can solve a linear system of equations to get the scalars such that the linear combination of the two given vectors equals $\begin{bmatrix}5\\6\end{bmatrix}$

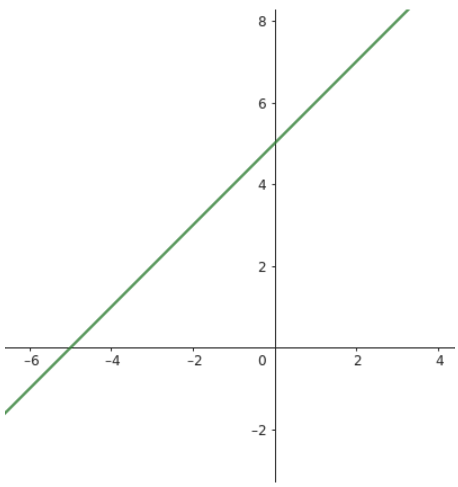
1.2



Consider the four vectors i, j, k, l pictured above. What is the sum $i + j + k + l$ equal to? What happens if we subtract k?
Answer: $i + j + k + l = 0$. Leaving k out gives us $i + j + l = -k = i$.

1.3

There is some vector $v \in \mathbb{R}^2$ that spans (i.e. $\mathsf{span}(v)$ is equal to) the line given below:



False: A line spanned by a vector always contains the 0 vector ($0 \in \mathsf{span}(\dots)$ always holds). We call what the graph shows an affine subspace.

2 Linear independence

Let's consider a sequence of vectors $v_1, \dots, v_k \in \mathbb{R}^n$ and a vector $v \in \mathsf{span}(v_1, \dots, v_k)$. Per the definition of span we have

$$v = \alpha_1 v_1 + \dots + \alpha_k v_k$$

for some $\alpha_1, \dots, \alpha_k \in \mathbb{R}$.
We might ask ourselves: Was this choice of scalars unique? Is there another way to express v as a linear combination of $v_1, \dots, v_k$? Assume there are also scalars $\beta_1, \dots, \beta_k \in \mathbb{R}$ that also produce v:

$$v = \alpha_1 v_1 + \dots + \alpha_k v_k$$
$$v = \beta_1 v_1 + \dots + \beta_k v_k$$

Subtracting the second from the first line gives us

$$0 = (\alpha_1 - \beta_1)v_1 + \dots + (\alpha_k - \beta_k)v_k$$

Now if the choice of scalars to get v was unique we have $\alpha_i = \beta_i$ and thus $\alpha_i - \beta_i = 0$ for all $i \in \{1, \dots, k\}$. Hence the only linear combination that gives us the zero vector is choosing zero for **all** scalars. We then call $v_1, \dots, v_k$ *linearly independent*. Conversely, if there is some linear combination that gives the zero vector where not all scalars are zero, this sequence of vectors is *linearly dependent*.

The following definitions are all equivalent:

Definition A sequence of vectors $v_1, \dots, v_k \in \mathbb{R}^n$ is *linearly independent* if the only way to express the zero vector as a linear combination $\alpha_1 v_1 + \dots + \alpha_k v_k = 0$ is by choosing $\alpha_1 = \dots = \alpha_k = 0$

$$\iff$$

Every vector in $\mathsf{span}(v_1, \dots, v_k)$ has exactly one (unique) representation as a linear combination of $v_1, \dots, v_k$

$$\iff$$

No vector $v_i \in \{v_1, \dots, v_k\}$ can be expressed as a linear combination of $v_1, \dots, v_{i-1}, v_{i+1}, \dots v_k$ (the other vectors)

If a sequence of vectors is not linearly independent it is linearly dependent: Thus we negate all the statements above for this definition.

Definition A sequence of vectors $v_1, \dots, v_k \in \mathbb{R}^n$ is *linearly dependent* if there is a way to express the zero vector as a linear combination $\alpha_1 v_1 + \dots + \alpha_k v_k = 0$ where not all $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ are equal to zero

$$\iff$$

There is some vector in $\mathsf{span}(v_1, \dots, v_k)$ that can be expressed as a linear combination of $v_1, \dots, v_k$ in more than one unique way

$$\iff$$

Some vector $v_i \in \{v_1, \dots, v_k\}$ can be expressed as a linear combination of $v_1, \dots, v_{i-1}, v_{i+1}, \dots v_k$ (the other vectors)

You can find proofs on the equivalence of similar statements in the blackboard notes.

- $v_1, \dots, v_k$ linearly dependent $\implies$ we can remove some vector without affecting $\mathsf{span}(v_1, \dots, v_k)$. In fact, we can make any sequence of linearly dependent vectors linearly independent by removing vectors. We will see a systematic method on how to do this (Gaussian Elimination) in the coming lectures.

- Vectors chosen at random are linearly independent with very high probability (you could verify this with numpy)

3 Dot product, norm

You have seen two different definition of the (euclidean) dot product in the lecture. First, an algebraic one:

$$\begin{bmatrix} v_1 \\ \vdots \\ v_k \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = v_1 w_1 + \cdots + v_k w_k = \sum_{i=1}^{k} v_i w_i$$

3.1 The euclidean norm

Using this definition we can compute a vectors length which is given by its (euclidean) norm:

Definition A vectors *euclidean norm* denoted $||v||$ is a real number given by

$$||v|| = \sqrt{v \cdot v} = \sqrt{v_1^2 + \cdots + v_k^2}$$

and assigns to a vector $v \in \mathbb{R}^n$ its magnitude.

We can see that this actually corresponds to a vectors length by visualizing a vector in the xy plane and applying the pythagorean theorem:

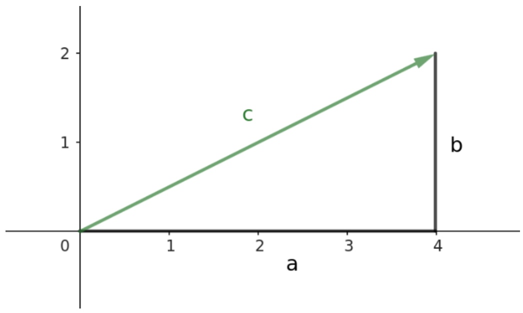


Figure 1: The pythagorean theorem gives us $c = \sqrt{a^2 + b^2}$. This also extends to higher dimensions.

Example What is the length of the vector $v = \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix}$ in $\mathbb{R}^{49}$?

We get $||v|| = \sqrt{1^2 + \cdots + 1^2} = \sqrt{49} = 7$.

The norm of a vector v can be used to get a vector u pointing the same direction as v but with length one (we call this normalizing): $u = \frac{v}{||v||}$.

The lecture also showed a geometric definition of the dot product (cosine formula):

$$v \cdot w = ||v||\,||w||\cos(\alpha)$$

where α is the angle between v and w.

How do the two definitions (algebraic and geometric)relate? What is the geometric meaning of the dot product and what does it tell us (aside from a vectors length)? Read ahead to find out.

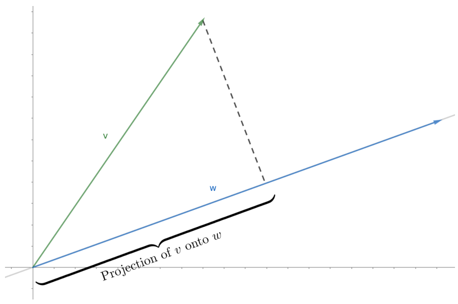
3.2 Geometric interpretation of the dot product

Geometrically $v \cdot w$ corresponds to the following:

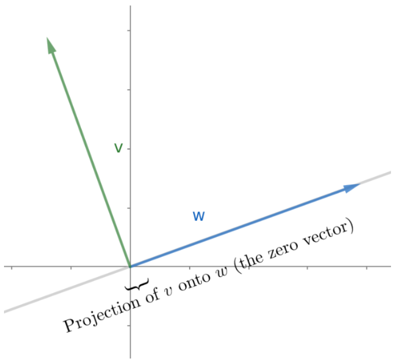
1. Project v onto w (the line that w spans)
2. Multiply (signed) length of projection with $||w||$

$$\Rightarrow v \cdot w = {}^s||proj(v \to w)||\ ||w||$$

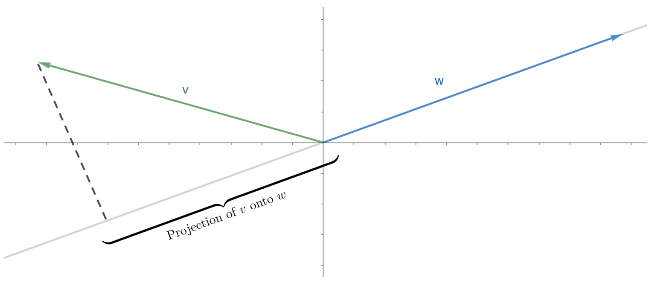
where signed means that if the projection of v onto w points in the opposite direction as w, we multiply the projection's length by (-1). ${}^s||proj(v \to w)||$ denotes the signed length of the projection of v onto w.
The following graphs illustrate the 3 main cases:



$v \cdot w > 0$
v and w point in approximately the same direction



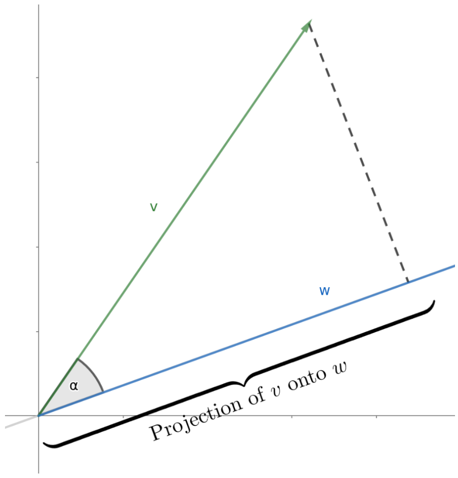
$v \cdot w = 0$
v and w are perpendicular/orthogonal to each other



$v \cdot w < 0$
v and w point in approximately the opposite direction

We have now seen that the dot product shows us the **length** of one vector (by taking its norm $||v|| = \sqrt{v \cdot v}$) and relative **direction** of two vectors (by checking the sign of $v \cdot w$).

With the geometric interpretation the geometric definition (cosine formula) becomes apparent:



Per the definition of cosine we get $\cos(\alpha) = \frac{\text{adjacent}}{\text{hypotenuse}} = \frac{^s||proj(v \to w)||}{||v||} \iff {^s||proj(v \to w)||} = \cos(\alpha)\ ||v||$. Plugging this into $v \cdot w = {^s||proj(v \to w)||}\ ||w||$ gives us $v \cdot w = \cos(\alpha)\ ||v||\ ||w||$, the cosine formula.

3.3 Proving the algebraic from the geometric definition

Here we try to find out why one can compute the dot product as a sum of the product of the vector's respective entries.

Claim: $v \cdot w = \sum_{i=1}^n v_i w_i$

First, we need a few ingredients:

(1) Any vector $w \in \mathbb{R}^n$ can be written as a linear combination with standard unit vectors e_i:

$$w = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} w_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ w_2 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ w_n \end{bmatrix} = \sum_{i=1}^n w_i e_i$$

(2) Commutativity of scalar-vector multiplication

(3) Distributivity of the dot product: One can confirm this with either definition of the dot product and also geometrically. The abstract scalar product is actually defined as being distributive (usually called linearity in first/second argument).

(4) $v \cdot e_i = v_i$: Applying the algebraic definition directly gives us this. However, we cannot reason with this definition as it is what we want to show. The illustration below shows that $\cos(\theta_i) = \frac{a_i}{||a||}$. The cosine formula gives us $a \cdot e_i = ||a||\ ||e_i||\ \cos(\theta_i) = a_i ||e_i|| = a_i$ using $||e_i|| = 1$ in the last step. Taking the dot product with e_i leaves only the part of v in the direction of e_i (here the x coordinate of a).

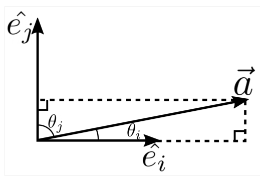


Figure 2: Source: [1]

Now onto the proof:

$$\begin{aligned}
v \cdot w &= v \cdot (\sum_{i=1}^{n} w_i e_i) && (w = \sum_{i=1}^{n} w_i e_i) && (1)\\
&= v \cdot (\sum_{i=1}^{n} e_i w_i) && \text{(commutativity scalar-vector multiplication)} && (2)\\
&= \sum_{i=1}^{n} (v \cdot e_i) w_i && \text{(distributivity dot product)} && (3)\\
&= \sum_{i=1}^{n} v_i w_i && (v \cdot e_i = v_i) && (4)
\end{aligned}$$

We summarize some key ideas surrounding the dot product:

Definition The *dot product* (euclidean scalar product) is a function $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ that takes two vectors as input and gives back a real number.
It gives us information the **length** of one vector (see *norm*) and **angle** between two vectors.
Geometrically $v \cdot w$ corresponds to projecting v onto w and multiplying the (signed) length of that projection by the length of w:

$$v \cdot w > 0 \implies v \text{ and } w \text{ point in approximately the same direction}$$
$$v \cdot w = 0 \implies v \text{ and } w \text{ are perpendicular/orthogonal to each other}$$
$$v \cdot w < 0 \implies v \text{ and } w \text{ point in approximately the opposite direction}$$

Algebraic definition:

$$v \cdot w = v_1 w_1 + \cdots + v_n w_n = \sum_{i=1}^{n} v_i w_i$$

Geometric definition (cosine formula):

$$v \cdot w = ||v|| \, ||w|| \cos(\alpha)$$

- Remark: It's also possible to define a scalar product and norm in a more abstract way through describing its required properties (see e.g. Linear Algebra Done Right - Axler, Definition 6.3)

4 Cauchy-Bunjakowski-Schwarz and triangle inequality

These two inequalities occur very often in mathematics. The Cauchy-Bunjakowski-Schwarz (CBS) inequality is stated as follows:

$$|v \cdot w| \leq ||v|| \, ||w||$$

Applying the geometric interpretation $v \cdot w = {}^{s}||proj(v \to w)|| \, ||w||$ gives us (assuming $w \neq 0$):

$$\begin{aligned}
&| \, {}^{s}||proj(v \to w)|| \, ||w|| \, | \leq ||v|| \, ||w|| \\
&\iff ||proj(v \to w)|| \, ||w|| \leq ||v|| \, ||w|| && \text{(lifting the absolute value)}\\
&\iff ||proj(v \to w)|| \leq ||v|| && \text{(dividing by } ||w||)
\end{aligned}$$

where the last statement can be read as "the length of v projected onto w can never exceed the length of v" which makes a lot of sense geometrically!

References

[1] By Igadjev3 - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=97260519

[2] https://en.wikipedia.org/wiki/Dot_product

[3] Sheldon Axler (2015) *Linear Algebra Done Right*, Springer International Publishing.

[4] Gilbert Strang, *Introduction to Linear Algebra*, 6th Edition, Wellesley - Cambridge Press.

[5] Geogebra for illustrations