Unit 2: Geometry: angles, spans, bases, and projections

Dr.P.M.Bajracharya

November 23, 2021

Contents

Now, we'll discuss geometric properties of $\mathbb{R}^n$. This includes angles and perpendicularity; linear combinations, spans, and linear dependence/independence; basis vectors, including orthogonal/orthonormal basis vectors; and projections onto basis vectors. Your are familiar with many of these ideas in $\mathbb{R}^2$ and $\mathbb{R}^3$. At present, we'll generalize them to $\mathbb{R}^n$.

As we will see, several of these ideas have a large algebraic component, in the sense that to implement them on problems of interest typically requires many many relatively-straightforward algebraic steps. This is the sort of thing at which computers excel. On the other hand, to understand why these linear algebra methods are so useful in data science (and beyond) requires a good understanding Euclidean geometry, and in particular how basic ideas of the Euclidean geometry of $\mathbb{R}^2$ or $\mathbb{R}^3$ are generalized to $\mathbb{R}^n$.

1 Geometry of $\mathbb{R}^n$: dot products, angles, and perpendicularity

1.1 Dot products

Recall that the dot product on $\mathbb{R}^n$ is a generalization of the notion of the dot product on the plane $\mathbb{R}^2$, and it gives us the geometric notions of lengths and angles for vectors in $\mathbb{R}^n$.

The dot product $x \cdot y$ of two vectors $x, y \in \mathbb{R}^n$ is defined as

$$x \cdot y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i.$$

Remark. As defined, the dot product is a multiplication between two vectors that returns as output a number.

If we view these two (column) vectors as matrices, then they are $n \times 1$ matrices. Then the dot product can be expressed in terms of matrix multiplication by taking transposes.

$$x^T y = \sum_{i=1}^{n} x_i y_i = y^T x.$$

So, we have

$$x \cdot y = x^T y = y^T x.$$

This is a very special case of a matrix multiplication. It is in fact commutative; and in this special case, it is also distributive, i.e.,

$$(x + z)^T y = x^T y + z^T y, \quad x^T (y + z) = x^T y + x^T z.$$

This is sometimes called an **inner product** or the **standard inner product** since the "inner dimension" (i.e., n) is "dotted out" and one ends up with a number.

Remark. Given two vectors $x, y \in \mathbb{R}^n$, if we view them as $n \times 1$ matrices and consider transposes, then we can compute the matrix product in two different ways,

- $x^T y$ – yielding a number (in $\mathbb{R} = \mathbb{R}^1$)
- xy^T – yielding an $n \times n$ matrix (in $\mathbb{R}^n$)

We saw the former above, it is called an *inner product*. The latter is also of interest, and it is called the **outer product** of x and y since the higher dimension (i.e., n) "sticks out" to be multiplied by a vector in $\mathbb{R}^n$.

Remark. Let A be an $m \times n$ matrix and B an $n \times p$ matrix. Then, we know, the (ij) entry of the matrix product AB is given by

$$(AB)_{ij} = A_{i:}B_{:j} = A_{i:}^T B_{:j} = B_{:j}^T A_{i:},$$

i.e., it is given by the inner product between the ith row of A and the jth column of B. There are mp such entries in the product matrix. On the other hand, then entire matrix product can be expressed as

$$AB = \sum_{k=1}^{n} A_{:k} B_{k:}^T,$$

i.e., as the sum of the outer products between the kth column of A and the corresponding kth row of B. Observe that each $A_{:k}B_{k:}^T$ is an $m \times p$ matrix, for each $k \in \{1, ..., n\}$, and we can add them up elementwise to get the matrix AB which is also an $m \times p$ matrix.

As we will see, the inner product is important for many reasons. One reason is that it has close connections with a particular vector norm.

Definition. The length or norm or Euclidean norm of a vector $x \in \mathbb{R}^n$ is given by

$$\|x\|_2 = (x \cdot x)^{1/2} = (x^T x)^{1/2} = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$$

Remark 1.

> As we have discussed, there are other notions of norms, e.g, L_1 and L_∞, but the L_2 or Euclidean norm is so ubiquitous since it gives the most useful geometry. That geometry is directly related to the fact that it can be expressed as an inner product, or equivalently as a matrix multiplication. That is the reason it of central interest in linear algebra, machine learning, data science, etc.

1.2 Angles

Recall that for $x, y in \mathbb{R}^2$ or $\mathbb{R}^3$, we have a notion of an angle between x and y:

$$x \cdot y = \|x\|_2 \|y\|_2 \cos\theta.$$

We want to generalize this to $\mathbb{R}^n$.

This generalization is straightforward: simply use the above equation, where the vectors are now in $\mathbb{R}^n$. The equations are the same, and the generalization works. Doing so, however, does require establishing one slightly subtle thing, i.e., establishing that

$$\frac{x \cdot y}{\|x\|_2 \|y\|_2} \in [-1, 1],$$

that is,

$$\frac{|x \cdot y|}{\|x\|_2 \|y\|_2} = \left| \frac{x \cdot y}{\|x\|_2 \|y\|_2} \right| \leq 1.$$

for arbitrary vectors x and y. The reason we need to establish this is so that we can take the arccos of this expression, i.e., so that there is in fact an angle θ, the cosine of which equals it. See the figure given below for an illustration.

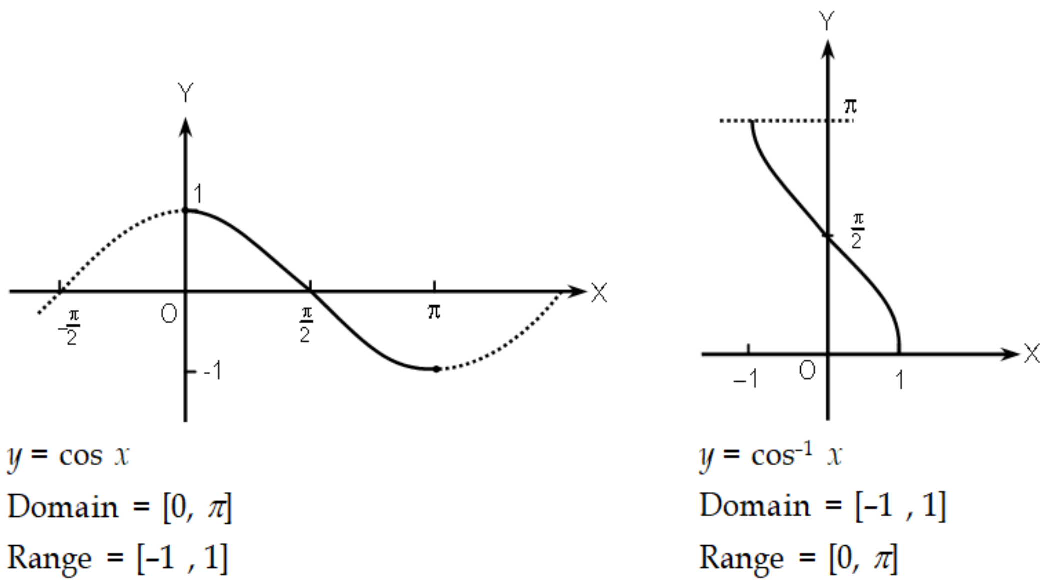


Figure 1: Illustration of cosine and arccosine functions.

The last expression is true, for arbitrary vectors x and y in $\mathbb{R}^n$, and it is known as the Cauchy-Schwartz Inequality. It is an extremely important result, largely because, although we show it for vectors in $\mathbb{R}^n$, it actually holds much more generally.

We will actually prove this special case using elementary methods.

Theorem 1.1. *If x and y are vectors in $\mathbb{R}^n$, then*

(a) $|x \cdot y| \leq \|x\|_2 \|y\|_2$.

(b) *Equality holds iff $x = \alpha y$ for $\alpha \in \mathbb{R}$.*

Proof. Consider the function

$$f(t) = |x + ty|^2,$$

where $t \in \mathbb{R}$. Then

$$\begin{aligned}
f(t) &= |x + ty|^2 \\
&= (x + ty)^T(x + ty) \\
&= x^T x + 2tx^T y + t^2 y^T y \\
&= \|y\|_2^2 t^2 + 2x^T yt + \|x\|_2^2 \\
&= at^2 + bt + c \\
&\geq 0,
\end{aligned}$$

setting $a = \|y\|_2^2$, $b = 2x^T y$ and $c = \|x\|_2^2$ We see that the function $f(t)$ is quadratic in t and $f(t) \geq 0$. That means, the graph of $f(t)$ lies above the x-axis. Hence the discriminant of the quadratic function

$$\begin{aligned}
& b^2 - 4ac \leq 0 \\
\Rightarrow \quad & 4(x^T y)^2 - 4\|y\|_2^2\|x\|_2^2 \leq 0 \\
\Rightarrow \quad & (x \cdot y)^2 \leq \|y\|_2^2\|x\|_2^2.
\end{aligned}$$

Therefore,

$$|x \cdot y| \leq \|x\|_2\|y\|_2.$$

This establishes Part (a). For Part (b), assume that $y = \alpha x$. Then we have

$$\begin{aligned}
|x \cdot y| &= |x \cdot (\alpha x)| \\
&= |\alpha||x \cdot x| \\
&= |\alpha|\|x\|_2^2 \\
&= (\alpha\|x\|_2)\|x\|_2 \\
&= \|y\|_2\|x\|_2.
\end{aligned}$$

Conversely, assume that

$$|x \cdot y| = \|x\|_2\|y\|_2$$

in which case the discriminant

$$4(x^T y)^2 - 4\|y\|_2^2\|x\|_2^2 = 0$$

So, the quadratic equation $\|y\|_2^2 t^2 + 2x^T yt + \|x\|_2^2$ has a single root t_0 as $|x + t_0 y|^2 = 0$, from which it follows that

$$x = -t_0 y.$$

$\square$

This theorem shows that

$$\begin{aligned} & -\|x\|_2\|y\|_2 \leq x \cdot y \leq \|x\|_2\|y\|_2 \\ \Rightarrow \qquad & -1 \leq \frac{x \cdot y}{\|x\|_2\|y\|_2} \leq 1 \end{aligned}$$

for all vectors x and y. So, the arccosine of this expression exists, for all vectors x and y. This means that there is number, which we can interpret as an angle θ, such that this is the cosine of it. That is, we can define this as the angle between two vectors in $\mathbb{R}^n$.

Definition. Let $x, y \in \mathbb{R}^n$. The angle θ between these two vectors is

$$\theta = \arccos\left(\frac{x \cdot y}{\|x\|_2\|y\|_2}\right),$$

which is an angle θ such that $0 \leq \theta \leq \pi$.

Given this definition, we might wonder how similar or different are vectors in $\mathbb{R}^2$ and $\mathbb{R}^n$. There are a number of ways to answer this. Here is one.

Question. What is the angle between the diagonal of the unit cube in the positive orthant and the vector e_1?

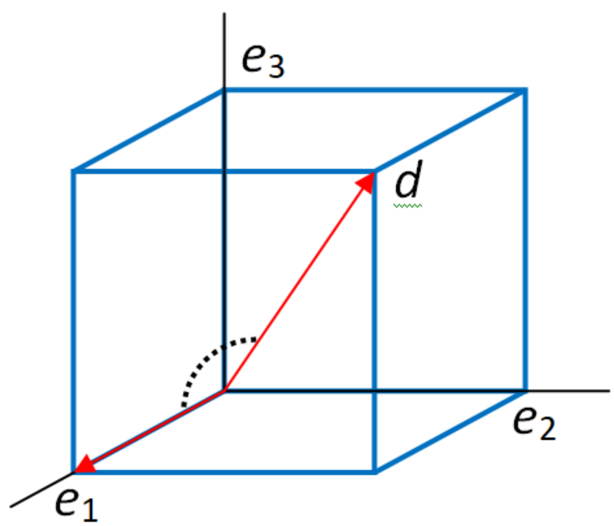
Answer: For $\mathbb{R}^2$, it is 45 degrees or $\arccos(1/\sqrt{2}) = \pi/4$ radians.

What is the answer in the case of $\mathbb{R}^3$ or $\mathbb{R}^n$ in general ? Let's see how to come to the answer for $\mathbb{R}^3$ and then we can easily get the answer for $\mathbb{R}^n$.

So, consider the unit cube whose edges are the standard basis vectors:

$$e_1 = (1, 0, 0), e_2 = (0, 1, 0), e_3 = (0, 0, 1).$$

Let d be the diagonal vector. Then we have to find the angle between d and e_1. See the figure given below. We



know that the diagonal vector d is given by

$$d = e_1 + e_2 + e_3 = (1, 0, 0) + (0, 1, 0) + (0, 0, 1) = (1, 1, 1).$$

Then its length is

$$\|d\|_2 = \sqrt{(d \cdot d)} = \sqrt{1^2 + 1^2 + 1^2} = \sqrt{3}.$$

The length of the vector e_1 is

$$\|e_1\|_2 = \sqrt{(e_1 \cdot e_1)} = \sqrt{1^2 + 0^2 + 0^2} = 1.$$

Moreover,

$$(e_1 \cdot d) = ((1,0,0) \cdot (1,1,1)) = 1.$$

If α is the angle between the diagonal d and the edge e_1, then

$$\cos\alpha = \frac{e_1 \cdot d}{\|e_1\|_2 \|d\|_2} = \frac{1}{\sqrt{3}}.$$

Therefore, the angle between the diagonal d and the edge e_1 is

$$\alpha = \arccos(1/\sqrt{3}) \approx 54.7°.$$

We observe that the angle between the diagonal of the unit cube in the positive orthant and the vector e_1 increases with increasing n. Therefore, the answer depends strongly on the value of n in $\mathbb{R}^n$.

1.3 Orthogonality between two vectors in $\mathbb{R}^n$

Using the notion of dot product between two vectors, we can define what it means for two vectors to be orthogonal, which generalizes the notion of perpendicularity defined on $\mathbb{R}^2$. Here is the definition.

Definition. Two vectors $x, y \in \mathbb{R}^n$ are **orthogonal** or **perpendicular** if

$$x \cdot y = 0.$$

That is, in $\mathbb{R}^n$, two vectors are orthogonal/perpendicular if the angle between them is $90° = \pi/2$ radians.

Remark.

- Clearly, the defining condition in this definition is the same as

$$x^T y = y^T x = 0.$$

- This definition does not depend on whether or not the vectors are of unit-length. Thus, the following should also be clear.

  Two vectors $x, y \in \mathbb{R}^n$ are orthogonal if the unit-length vector in the direction of x, call it u, and the unit-length vector in the direction of y, call it v, are orthogonal.

Problem.

Recall the definition of the set of vectors perpendicular to a vector $x \in \mathbb{R}^n$ denoted by $x^\perp$. Thus,

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

We have shown that for $x \in \mathbb{R}^2$, the set $x^\perp$ is a subspace of $\mathbb{R}^2$. This is just a one-dimensional line perpendicular to x.

For $x \in \mathbb{R}^3$, the set $x^\perp$ is a two-dimensional plane perpendicular to x; and for $x \in \mathbb{R}^n$, the set $x^\perp$ is a subsapce of dimension $n-1$ that is oriented to be perpendicular to x. (Conversely, we could have asked for the set of vectors perpendicular to two or more vectors. This is algebraically more complex, but the ideas are similar, and we will get to this later.)

1.4 Spans, and linear dependence/independence

We can't visualize the data in $\mathbb{R}^n$. However, we would like to be able to to say that data points are roughly the same from the perspective of linear algebra in the sense that one could generate each data point from the others by performing vector addition as well as scalar multiplication.

A linear algebraic idea that permit one to do that is the idea of linear dependence/independence.

Linear independence captures the idea that the vectors do not contain redundant information in the sense that you can compute one from the others with the linear operations of scalar multiplication and vector addition. On the other hand, linear dependence captures the idea that the vectors contain redundant information in the same sense.

To develop this, let's start with the following notion of linear combination.

Definition. If $a_1, ..., a_k \in \mathbb{R}$, then a linear combination of the vectors $v_1, ..., v_k \in \mathbb{R}^n$ is a vector $w \in \mathbb{R}^n$ s.t.

$$w = \sum_{i=1}^{k} a_i v_i.$$

This definition expresses the idea that a vector can be expressed in terms of given vectors.

Example.

- Let $\begin{pmatrix} 0 \\ 3 \\ 4 \end{pmatrix} \in \mathbb{R}^3$. Then,

$$v = 3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 4 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

  Thus v is a linear combination of e_2 and e_3.

- Let A be an $m \times n$ matrix, and let x be an n-dimensional column vector. Then Ax is a linear combination of the columns of A. For example,

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 \\ 4 \end{pmatrix} = 5 \begin{pmatrix} 1 \\ 3 \end{pmatrix} + 6 \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 5+12 \\ 15+24 \end{pmatrix} = \begin{pmatrix} 17 \\ 39 \end{pmatrix}.$$

- Let A be an $m \times n$ matrix, and let x be an m-dimensional column vector. Then $x^T A$ is a linear combination of the rows of A. Consider the following example:

$$\begin{pmatrix} 7 & 8 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = 7 \begin{pmatrix} 1 & 2 \end{pmatrix} + 8 \begin{pmatrix} 3 & 4 \end{pmatrix} = \begin{pmatrix} 31 & 46 \end{pmatrix}.$$

Remark. The last two examples show that matrix multiplication can be understood in terms of linear combinations,

1.5 More about span

The notion of linear combination had to do with whether a given vector could be described by a set of given vectors with the operations of scalar multiplications and vector additions. We often want to go "in the other direction" and ask: **if we have a**

set of vectors, then which vectors can be computed from them with the operations of scalar multiplications and vector additions. This gets us to the notion of *span*.

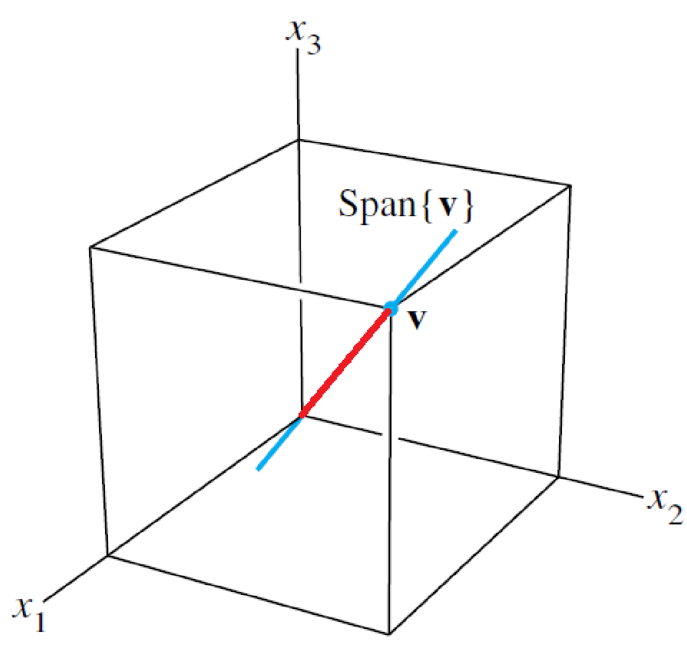
Definition. Let $\{v_1, v_2, ..., v_k\}$ be a set of vectors in a vector space $\mathbb{R}^n$. The span of $v_1, v_2, ..., v_k$ is defined by

$$\operatorname{span}(v_1, v_2, ..., v_k) = \left\{ \sum_1^k a_i v_i : \text{ each } a_i \in \mathbb{R} \right\}.$$

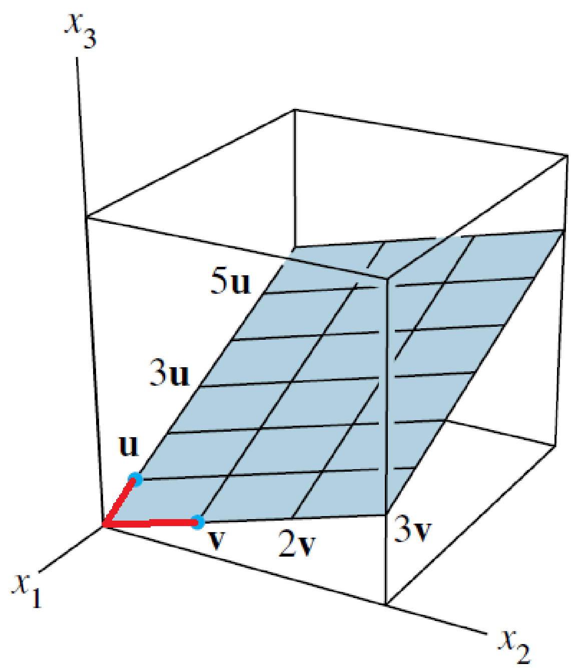
If V is a vector space and $\operatorname{span}(v_1, v_2, ..., v_k) = V$, we say that the vectors $v_1, v_2, ..., v_k$ span V.

A Geometric Description of $\operatorname{span}(v)$ and $\operatorname{span}(u, v)$

Let v be a nonzero vector in $\mathbb{R}^3$. Then $\operatorname{span}(v)$ is the set of all scalar multiples of v, and we visualize it as the set of points on the line in $\mathbb{R}^3$ through v and 0. See the figure given below.



If u and v are nonzero vectors in $\mathbb{R}^3$, with v not a multiple of u, then $\operatorname{span}(u, v)$ is the plane in $\mathbb{R}^3$ that contains u, v, and 0. In particular, $\operatorname{span}(u, v)$ contains the line in $\mathbb{R}^3$ through u and 0 and the line through v and 0. See the figure given below.



Examples.

- The span of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is the set of all vectors of the form $\begin{pmatrix} \alpha \\ 0 \end{pmatrix}$, for $\alpha \in \mathbb{R}$ and the span of $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is the set of all vectors of the form $\begin{pmatrix} \alpha \\ 2\alpha \end{pmatrix}$, for $\alpha \in \mathbb{R}$. Similarly, the span of $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} -3 \\ -6 \end{pmatrix}$ is the set of all vectors of the form $\begin{pmatrix} \alpha \\ 2\alpha \end{pmatrix}$, for $\alpha \in \mathbb{R}$. These sets are all lines through the origin on $\mathbb{R}^2$, and thus subspaces of $\mathbb{R}^2$.

- Let e_1 and e_2 be the coordinate vectors for $\mathbb{R}^2$, i.e.,$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Then
$$\operatorname{span}(e_1, e_2) = \mathbb{R}^2.$$
- Similarly, $\operatorname{span}(e_1, e_2, e_3) = \mathbb{R}^3$..
- Let
$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$
Then $\operatorname{span}(v_1, v_2) = \mathbb{R}^2$.

Proof.

We have
$$v_1 + v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \sqrt{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
and
$$v_1 - v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \sqrt{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$
If $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$, then
$$\begin{aligned} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= \frac{x_1}{\sqrt{2}}(v_1 + v_2) + \frac{x_2}{\sqrt{2}}(v_1 - v_2) \\ &= \frac{x_1 + x_2}{\sqrt{2}} v_1 + \frac{x_1 - x_2}{\sqrt{2}} v_2 \end{aligned}$$
Therefore, $\operatorname{span}(v_1, v_2) = \mathbb{R}^2$.

- If $x_1 = \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix}$ and $x_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$, then

  $$\operatorname{span}(x_1, x_2) \neq \mathbb{R}^3.$$

  The reason is that the vector $x = \begin{pmatrix} 0 \\ a \\ 0 \end{pmatrix}$ with $a \neq 0$ cannot be expressed as a linear combination of x_1 and x_2.

- If $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, then

  $$\operatorname{span}(e_1, e_2) \neq \mathbb{R}^3.$$

  The reason is that the vector $e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ cannot be expressed as a linear combination of e_1 and e_2.

- Let A be an $m \times n$ matrix, and let x vary over all possible n-dimensional column vectors. Then, the span of the columns of A is given by

  $$\{Ax : \quad x \in \mathbb{R}^n\}.$$

  In particular, if $A_{:j}$ denotes the jth column of A, then this set is all vectors of the form

  $$\sum_{i=1}^{n} x_j A_{:j},$$

  as x is varied over all of $\mathbb{R}^n$.

- Let A be an $m \times n$ matrix, and let y vary over all possible m-dimensional column vectors. Then, the span of the rows of A is given by

  $$\{y^T A : \quad y \in \mathbb{R}^m\}.$$

  In particular, if $A_{i:}$ denotes the ith row of A, then this set is all vectors of the form

  $$\sum_{i=1}^{m} y_i A_{i:},$$

  as x is varied over all of $\mathbb{R}^m$.

Problem. Consider the vectors $u = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $v = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

(a) Write the vector $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ in terms of the vectors u and v.

(b) Show that the vectors u and v span $\mathbb{R}^2$.

Problem. If $v_1, ..., v_k \in \mathbb{R}^n$ and $V = \operatorname{span}(v_1, ..., v_k)$, then V is a subspace of $\mathbb{R}^n$.

Remark. It is a fact that $V = \operatorname{span}(v_1, ..., v_k)$ is the smallest subspace of $\mathbb{R}^n$ that contains $v_1, ..., v_k$. We will not prove it.

Definition. Let $S \subseteq \mathbb{R}^n$. Put

$$S^{\perp} = \{u \in \mathbb{R}^n : \quad \forall\, v \in S \; u \cdot v = 0\}.$$

This set is called the **orthogonal complement** of S.

Problem. Prove that if $v_1, ..., v_k \in \mathbb{R}^n$ and $V = \operatorname{span}(v_1, ..., v_k)$, then $V^{\perp}$ is a subspace of $\mathbb{R}^n$.

1.6 Linear dependence and independence

We obderve that

$$\operatorname{span}(e_1, e_2, e_3) = \mathbb{R}^3$$
$$\operatorname{span}\left(e_1, e_2, e_3, \begin{pmatrix}3\\4\\0\end{pmatrix}\right) = \mathbb{R}^3$$
$$\operatorname{span}\left(e_1, e_2, \begin{pmatrix}3\\4\\0\end{pmatrix}\right) \neq \mathbb{R}^3.$$

These examples show that not all sets of vectors may span a vector space. A similar statement holds for linear combinations of rows or columns of a matrix that are harder to visualize. So, we face with the following queston:

Which set of vectors span a vector space ?

To answer it, we need the following notion.

Definition. The vectors $v_1, ..., v_k$ are **linearly independent** if

$$\sum_{i=1}^{k} \alpha_i v_i = 0 \Rightarrow \alpha_i = 0 \text{ for all } i.$$

If a set of vectors is not linearly independent, then they are **linearly dependent**.

Nature of linerly dependent/independent vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$

Let u, v be two nonzero vectors in $\mathbb{R}^n$. Then they are linearly dependent if one is the scalar multiple of the other. For example, the vectors (3,1) and (6, 2) are linearly dependent, $(6, 2) = 2(3, 1)$. See the figure given below.

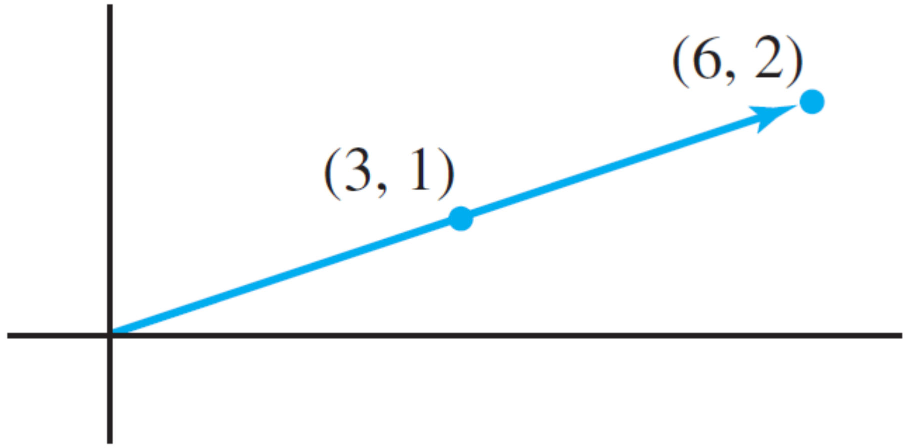


Figure 2: Linearly dependent

Let u, v be two nonzero vectors in $\mathbb{R}^n$. Then they are linearly independent if one cannot be the scalar multiple of the other. For example, the vectors (3,2) and (6, 2) are linearly independent. See the figure given below.

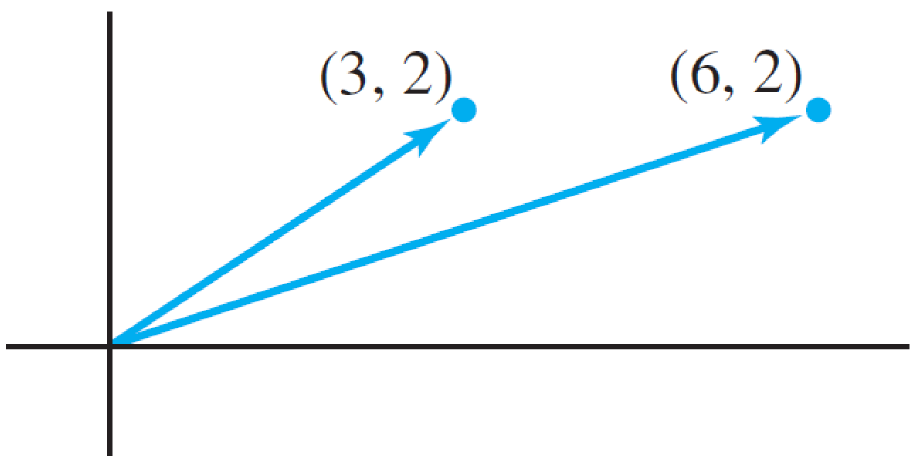


Figure 3: Linearly independent

Let u, v, w be three nonzero vectors in $\mathbb{R}^n$. Then they are linearly dependent if all three vectors lie on the same plare. See the figure given below.

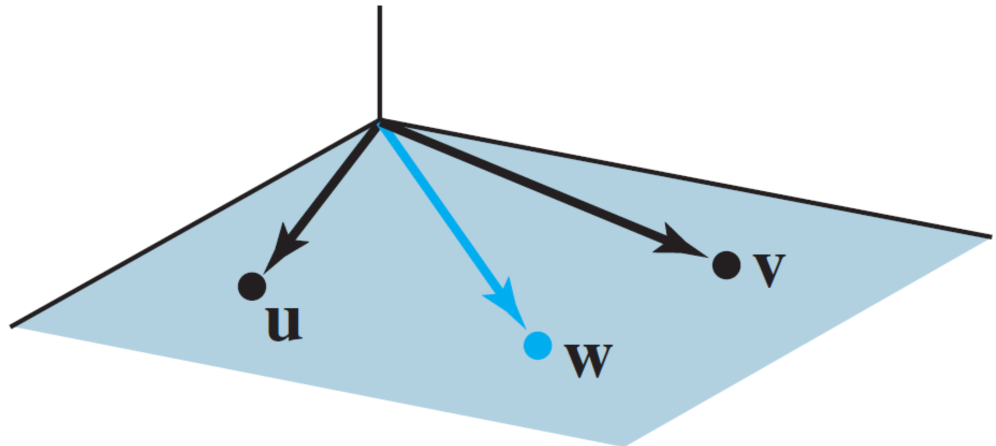


Figure 4: Linearly dependent
$w \in \text{span}\,(u, v)$

Let u, v, w be three nonzero vectors in $\mathbb{R}^n$. Then they are linearly independent if one of the vectors does not lie on the plane containing the other two. See the figure given below.

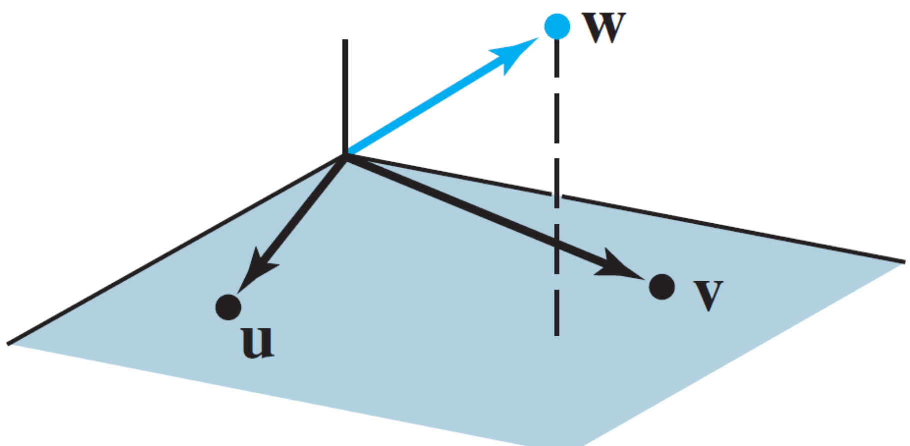


Figure 5: Linearly independent
$w \notin \text{span}\,(u, v)$

Let's see an important question: How many vectors in $\mathbb{R}^n$ can be linearly independent?

To get an idea, recall that

$$\text{span}\,(e_1) = \mathbb{R},\ \text{span}\,(e_1, e_2) = \mathbb{R}^2,\ \text{span}\,(e_1, e_2, e_3) = \mathbb{R}^3.$$

Here is the theorem that generalizes that.

Theorem 1.2. *In $\mathbb{R}^n$,*

(a) Any set of $n+1$ vectors are never linearly independent.

(b) Any set of $n-1$ vectors never span all of $\mathbb{R}^n$.

Examples.

1. Vectors $\begin{pmatrix}1\\0\\0\end{pmatrix}$ and $\begin{pmatrix}3\\4\\0\end{pmatrix}$ don't span $\mathbb{R}^3$. Similarly, Vectors $\begin{pmatrix}1\\1\\0\end{pmatrix}$ and $\begin{pmatrix}1\\-1\\0\end{pmatrix}$ don't span $\mathbb{R}^3$. In both cases, the span of these two linearly independent vectors is a two-dimensional plane corresponding to $x_3 = 0$, and so the span of these two vectors is a two-dimensional subspace of $\mathbb{R}^3$.

2. Vectors $\begin{pmatrix}1\\1\\0\end{pmatrix}$, $\begin{pmatrix}1\\-1\\0\end{pmatrix}$ $\begin{pmatrix}0\\0\\1\end{pmatrix}$ and $\begin{pmatrix}17\\12\\-2\end{pmatrix}$ are not linearly independent, but their span is all of $\mathbb{R}^3$.

3. Vectors $\begin{pmatrix}1\\0\\0\end{pmatrix}$, $\begin{pmatrix}2\\0\\0\end{pmatrix}$ $\begin{pmatrix}3\\0\\0\end{pmatrix}$ and $\begin{pmatrix}17\\12\\-2\end{pmatrix}$ are not linearly independent, and their span is not all of $\mathbb{R}^3$, but instead a two-dimensional subspace of $\mathbb{R}^3$.

1.7 Testing for linear dependence and independence

1. A set containing only one vector, say, v, is linearly independent if and only if v is not the zero vector. This is because the vector equation $x_1 v = 0$ has only the trivial solution when $v \neq 0$. The zero vector is linearly dependent because any nonzero value of x_1 satisfies the equation $x_1 0 = 0$.

2. Two vectors v_1, v_2 are linearly dependent if one of the vectors is a multiple of the other. The set is linearly independent if and only if neither of the vectors is a multiple of the other.

3. The vectors $v_1, ..., v_k \in \mathbb{R}^n$ are linearly dependent if one of the vectors is zero.

4. The vectors $v_1, ..., v_k \in \mathbb{R}^n$ are linearly dependent if $k \geq n$. For example, if $p > n$, the columns a $n \times p$ matrix are linearly dependent.

$$\begin{array}{cc} & p \\ n & \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} \end{array}$$

Figure 6: The columns are Linearly independent.

Recall that the span of a set of elements in a vector space is the set of all their linear combinations. The following proposition says, roughly, that linearly dependent vectors do not contribute to the span.

Theorem 1.3. *Let $v_1, ..., v_k$ be vectors in a vector space V and v_{k+1} is a linear combination of the vectors $v_1, ..., v_k$. Then*

$$\operatorname{span}(v_1, ..., v_k) = \operatorname{span}(v_1, ..., v_k, v_{k+1}).$$

Proof. Let

$$\begin{aligned} A &= \operatorname{span}(v_1, ..., v_k) \\ B &= \operatorname{span}(v_1, ..., v_k, v_{k+1}) \end{aligned}$$

If $x \in A$, then there are scalars $c_1, c_2, ..., c_k$ such that

$$\begin{aligned} x &= c_1 + v_1 + ... + c_k v_k \\ &= c_1 + v_1 + ... + c_k v_k + 0 v_{k+1} \in B. \end{aligned}$$

This shows that $A \subseteq B$. On the other hand, by assumption, $v_{k+1} \in A$. So, then there are scalars $c_1, c_2, ..., c_k$ such that

$$v_{k+1} = c_1 v_1 + ... + c_k v_k.$$

Now, if $y \in B$, then there are scalars $b_1, b_2, ..., b_{k+1}$ such that

$$\begin{aligned} y &= b_1 v_1 + ... + b_k v_k + b_{k+1} v_{k+1} \\ &= b_1 v_1 + ... + b_k v_k + b_{k+1}(c_1 v_1 + ... + c_k v_k) \\ . &= (b_1 + b_{k+1} c_1) v_1 + ... + (b_k + b_{k+1} c_k) v_k \in A. \end{aligned}$$

This shows that $B \subseteq A$. Therefore, $A = B$. $\square$

1.8 Bases

Definition. Let V be a subspace of $\mathbb{R}^n$. A set of vectors $v_1, ..., v_k \in V$ is called a **basis** for V if the following conditions are satisfied

(a) Vectors $v_1, ..., v_k \in V$ are linearly independent.

(b) $\operatorname{span}(v_1, ..., v_k) = V$.

The number of vectors in any basis for V is called the **dimension** of V, and is written $\dim V$.

Examples.

- The set $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ is a basis for $\mathbb{R}^2$. Hence the plane $\mathbb{R}^2$ has dimension 2.

- The columns of an invertible $n \times n$ matrix form a basis for all of $\mathbb{R}^n$ because they are linearly independent and span $\mathbb{R}^n$. One such matrix is the $n \times n$ identity matrix. Its columns are denoted by $e_1, ..., e_n$:

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \cdots \quad , \quad e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

  The set $\{e_1, ..., e_n\}$ is the standard basis for $\mathbb{R}^n$.

- Vectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ span $\mathbb{R}^2$ and provide a basis for $\mathbb{R}^2$. Moreover, each pair of vectors that are not scalar multiples of each other provides a basis for $\mathbb{R}^2$.

- Consider the solution to the following linear equation

$$x_1 + x_2 + x_3 = 0,$$

  which we can write as $x_3 = -x_1 - x_2$. We claim that this is a subspace of $\mathbb{R}^3$, call it V . Then each of the following sets:

$$\left\{\begin{pmatrix}1\\-1\\0\end{pmatrix},\begin{pmatrix}1\\0\\-1\end{pmatrix}\right\},\left\{\begin{pmatrix}-1\\0\\1\end{pmatrix},\begin{pmatrix}0\\1\\-1\end{pmatrix}\right\},$$
$$\left\{\begin{pmatrix}-1/2\\-1/2\\1\end{pmatrix},\begin{pmatrix}1\\-1\\0\end{pmatrix}\right\}$$

is a basis for V. Also, any set of 2 vectors that are linearly independent, i.e., s.t. the sum of the entries is 0, is a basis for V.

Note that if we add any of the following vectors:

$$\begin{pmatrix}1\\1\\1\end{pmatrix},\quad\begin{pmatrix}1\\1\\2\end{pmatrix},\quad\begin{pmatrix}1\\1\\3\end{pmatrix}$$

then we get a basis for $\mathbb{R}^3$. But if we donnot consider any of these vectors, and we only consider linear combinations of the first two, then we have a basis for the subspace orthogonal to the line we started with.

Problem. Let

$$V=\left\{\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}\in\mathbb{R}^3:\ x_1+x_2+x_3=0\right\},$$

$$B=\left\{\begin{pmatrix}1\\-1\\0\end{pmatrix},\begin{pmatrix}1\\0\\-1\end{pmatrix}\right\}$$

Show that V is a subspace of $\mathbb{R}^3$, and B is a basis for V.

Solution.

1. V is a subspace of $\mathbb{R}^3$, because
   (a) V is closed under vector addition.
   (b) V is closed under scalar multiplication.
2. B is a basis for V, because
   (a) Both vectors are in V, since

$$\begin{aligned} 1 - 1 + 0 &= 0 \\ 1 + 0 - 1 &= 0. \end{aligned}$$

   (b) $\operatorname{span}(B) = V$. In fact, let $\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \in V$ Then

$$v_1 + v_2 + v_3 = 0 \Rightarrow \qquad v_1 = -v_2 - v_3,$$

   and so

$$\begin{aligned} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} &= \begin{pmatrix} -v_2 - v_3 \\ v_2 \\ v_3 \end{pmatrix} \\ &= \begin{pmatrix} -v_2 \\ v_2 \\ 0 \end{pmatrix} + \begin{pmatrix} -v_3 \\ 0 \\ v_3 \end{pmatrix} \\ &= -v_2 \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - v_3 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \end{aligned}$$

   (c) Vectors in B are linearly independent. In fact,

$$\begin{aligned} a_1 \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = 0 \Rightarrow \begin{pmatrix} a_1 + a_2 \\ -a_1 \\ -a_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ \Rightarrow a_1 = a_2 &= 0. \end{aligned}$$

Problem. Let

$$B' = \left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}$$

Show that B' is a basis for V.

Problem. Let

$$B" = \left\{ \begin{pmatrix} -1/2 \\ -1/2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}$$

Show that $B"$ is a basis for V.

The definition of a basis for a vector space is designed so that

- the basis contains enough vectors
- the basis does not contain too many vectors.

In other words, a basis contains neither less nor more than the necessary number of vectors. If a basis B for a vector space V contains less than the necessary number of vectors, then B will not span the vector space V and If a basis B for a vector space V contains too many vectors, then the vectors in B will not be linearly independent.

Thus, a set $B = \{v_1, ..., v_k\}$ is a basis for V, iff

1. the set B is a maximally linearly independent set, i.e., it is linearly independent, and if we add one more vector from V to it, then it will not be linearly independent.

2. or the set B is a minimal spanning set, i.e.,B spans V and if we remove one vector from it, then it will no longer span V .

Theorem 1.4. *Let $B = \{v_1, ..., v_n\}$ be a basis for $\mathbb{R}^n$. Then every vector $w \in \mathbb{R}^n$ can be written uniquely as a linear combination of vectors in the basis B:*

$$w = a_1 v_1 + ... + a_n v_n,$$

where $a_1, ..., a_n$ are real numbers.

Proof. By the definition of a basis,

$$\text{span}\,(B) = \mathbb{R}^n.$$

If $w \in \mathbb{R}^n$, then

$$w = a_1 v_1 + ... + a_n v_n,$$

where $a_1, ..., a_n$ are real numbers. To prove the uniqueness, suppose that we have one more representation for w:

$$w = b_1 v_1 + ... + b_n v_n,$$

where $b_1, ..., b_n$ are real numbers. Then

$$(a_1 - b_1)v_1 + ... + (a_n - b_n)v_n = 0.$$

Since the vectors $v_1, ..., v_n$ are linearly independent,

$$\begin{aligned} &(a_1 - b_1) = ... = (a_n - b_n) = 0 \\ \Rightarrow \quad &a_1 = b_1, ..., a_n = b_n, \end{aligned}$$

This shows that there is one and only one way to write w as a linear combination of the basis vectors $v_1, ..., v_n$. $\square$

Theorem 1.5. *Let V be an n-dimensional vector space. Then*

1. *Any set of n elements of V that spans V must be linearly independent and thus is a basis.*

2. *Any linearly independent set of n elements of V must span V and thus is a basis.*

Proof. 1. Suppose that $B = \{v_1, ..., v_n\}$ spans V. If vectors $v_1, ..., v_n$ linearly dependent, then we know that some proper subset of B still spans V, making the dimension less than n, which is impossible. This proves Statement (1).

2. Let $B = \{v_1, ..., v_n\}$ be a set of n linearly independent vectors in V. Suppose that $\text{span}(B) \neq V$. We take a vector $u \in V - \text{span}(B)$. Then the set $B' = B \cup \{u\}$ is linearly independent. This is a contradiction, since on an n-dimensional vector space $n+1$ vectors cannot be linearly independent. Therefore, $\text{span}(B) = V$.

□

1.9 Orthogonal and orthonormal bases

Definition. A set S of vectors in $\mathbb{R}^n$ is called **orthogonal** if every pair of distinct vectors in S are orthogonal.

A set S of vectors in $\mathbb{R}^n$ is called **orthonormal** if S is orthogonal and every vector in S is a unit vector.

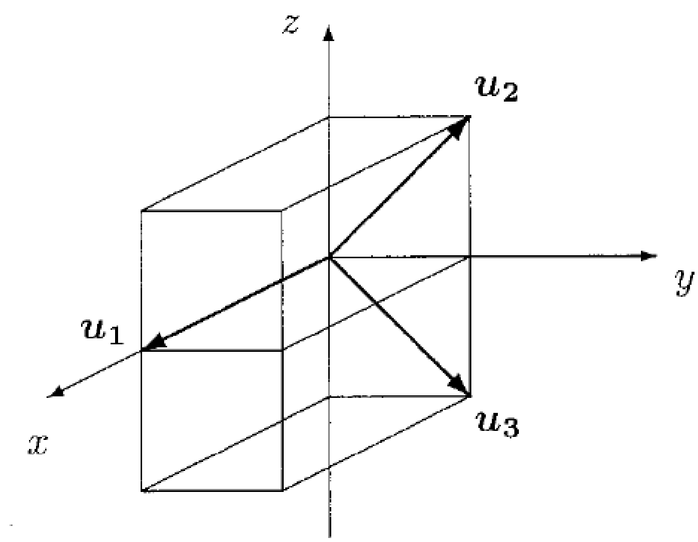
Example. It is easy to check that the standard basis $E = \{e_1, e_2, ..., e_n\}$ for $\mathbb{R}^n$ is an orthonormal set.

Example. Let $u_1 = (2\ 0\ 0)^T$, $u_2 = (0\ 1\ 1)^T$ and $u_3 = (0\ 1\ -1)^T$. Find the orthonormal set associated with the set $S = \{u_1, u_2, u_3\}$.

Solution. We have

$$u_1 \cdot u_2 = 2 \cdot 0 + 0 \cdot 1 + 0 \cdot 0 = 0$$
$$u_1 \cdot u_3 = 2 \cdot 0 + 0 \cdot 1 + 0 \cdot (-1) = 0$$
$$u_2 \cdot u_3 = 0 \cdot 0 + 1 \cdot 1 + 1 \cdot (-1) = 0.$$

Hence the set S is orthogonal. Let



$$v_1 = \frac{1}{\|u_1\|}u_1 = \frac{1}{2}(2\ 0\ 0)^T = (1\ 0\ 0)^T$$
$$v_2 = \frac{1}{\|u_2\|}u_2 = \frac{1}{\sqrt{2}}(0\ 1\ 1)^T = (0\ 1/\sqrt{2}\ 1/\sqrt{2})^T$$
$$v_3 = \frac{1}{\|u_3\|}u_3 = \frac{1}{\sqrt{2}}(0\ 1\ -1)^T = (0\ 1/\sqrt{2}\ -1/\sqrt{2})^T$$

Then for $i = 1, 2, 3$

$$\|v + i\| = 1$$

and if $i \neq j$, then

$$v_i \cdot v_j = \frac{1}{\|u_i\|\|u_j\|}(u_i \cdot u_j) = 0.$$

Therefore, the set $\{v_1, v_2, v_3\}$ is orthonormal.

Note that The process of converting an orthogonal set to an orthonormal set by multiplying each vector u by $\frac{1}{\|u\|}$ is called **normalizing**.

Theorem 1.6. *An orthogonal set of nonzero vectors in a vector space is linearly independent.*

Proof. Let $S = \{v_1, v_2, ..., v_k\}$. Consider the vector equation

$$c_1v_1 + c_2v_2 + ... + c_kv_k = 0.$$

Since S is orthogonal, $v_i \cdot v_j = 0$ for $i \neq j$. For $i = 1, 2, ..., k$,

$$\begin{aligned} 0 &= (c_1v_1 + c_2v_2 + ... + c_kv_k) \cdot v_i \\ &= (c_1v_1) \cdot v_i + (c_2v_2) \cdot v_i + ... + (c_kv_k) \cdot v_i \\ &= c_1(v_1 \cdot v_i) + c_2(v_2 \cdot v_i) + ... + c_k(v_k \cdot v_i) \\ &= c_i(v_i \cdot v_i) \end{aligned}$$

Given that $v_i \neq 0$ for each i, we have $v_i \cdot v_i \neq 0$. This means, $c_i = 0$ for each i. Therefore, S is linearly independent. □

Definition. A basis B for a vector space is called an **orthogonal basis** if B is orthogonal.

A basis B for a vector space is called an **orthonormal basis** if B is orthonormal.

Example. Let $u_1 = (2\ 0\ 0)^T$, $u_2 = (0\ 1\ 1)^T$ and $u_3 = (0\ 1\ -1)^T$. We have shown that the set $S = \{u_1, u_2, u_3\}$ is orthogonal. It is clear that each vector in S is nonzero. Hence by the previous theorem, S is linearly independent. We also know that any linearly independent set of n elements of an n-dimensional vector space V must span V and thus is a basis. Therefore, S is an orthogonal basis for $\mathbb{R}^3$. Moreover, the orthonormal set associated with S is an orthonormal basis $\mathbb{R}^3$.

The proof of the last theorem motivates the following result:

Theorem 1.7. *Let $B = \{v_1, v_2, ..., v_k\}$ be an orthogonal basis for a vector space V. Then for any vector $w \in V$,*

$$w = \frac{w \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{w \cdot v_2}{v_2 \cdot v_2} v_2 + ... + \frac{w \cdot v_k}{v_k \cdot v_k} v_k,$$

that is,

$$w = \left(\frac{w \cdot v_1}{v_1 \cdot v_1}, \frac{w \cdot v_2}{v_2 \cdot v_2}, ..., \frac{w \cdot v_k}{v_k \cdot v_k} \right)^T$$

with respect to Basis B.

Moreover, if $B = \{v_1, v_2, ..., v_k\}$ is an orthonormal basis for a vector space V, then for any vector $w \in V$,

$$w = (w \cdot v_1)v_1 + (w \cdot v_2)v_2 + ... + (w \cdot v_k)v_k,$$

that is,

$$w = (w \cdot v_1, w \cdot v_2, ..., w \cdot v_k)^T$$

with respect to Basis B.

Proof. Let $B = \{v_1, v_2, ..., v_k\}$ be an orthogonal basis for a vector space V. Then for any vector $w \in V$,

$$w = c_1 v_1 + c_2 v_2 + ... + c_k v_k$$

for some real numbers $c_1, c_2, ..., c_k$. Since S is orthogonal, $v_i \cdot v_j = 0$ for $i \neq j$. For $i = 1, 2, ..., k$,

$$\begin{aligned} w \cdot v_i &= (c_1 v_1 + c_2 v_2 + ... + c_k v_k) \cdot v_i \\ &= (c_1 v_1) \cdot v_i + (c_2 v_2) \cdot v_i + ... + (c_k v_k) \cdot v_i \\ &= c_1 (v_1 \cdot v_i) + c_2 (v_2 \cdot v_i) + ... + c_k (v_k \cdot v_i) \\ &= c_i (v_i \cdot v_i) \end{aligned}$$

Given that $v_i \neq 0$ for each i, we have $v_i \cdot v_i \neq 0$. Therefore,

$$c_i = \frac{w \cdot v_i}{v_i \cdot v_i}$$

for each i. That means,

$$w = \frac{w \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{w \cdot v_2}{v_2 \cdot v_2} v_2 + ... + \frac{w \cdot v_k}{v_k \cdot v_k} v_k,$$

that is,

$$w = \left(\frac{w \cdot v_1}{v_1 \cdot v_1}, \frac{w \cdot v_2}{v_2 \cdot v_2}, ..., \frac{w \cdot v_k}{v_k \cdot v_k} \right)^T$$

with respect to Base B.

If $B = \{v_1, v_2, ..., v_k\}$ is an orthonormal basis for a vector space V, then the result follows from the previous result, because $v_i \cdot v_i = \|v_i\|_2 = 1$ for all i. $\square$

Example. Let

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Show that $B = \{v_1, v_2\}$ is an orthonormal basis for $\mathbb{R}^2$. Find a vectyor $x \in \mathbb{R}^2$ with respect to the basis B.

Solution. We have

$$v_1 \cdot v_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix} = \frac{1}{2}(1 \cdot 1 + 1 \cdot (-1)) = 0.$$

Also, we have

$$v_1 \cdot v_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} = \frac{1}{2}(1^2 + 1^2) = 1.$$

Similarly,

$$v_2 \cdot v_2 = 1.$$

Therefore, $B = \{v_1, v_2\}$ is an orthonormal basis for $\mathbb{R}^2$. To find a vectyor $x \in \mathbb{R}^2$ with respect to the basis B, assume that $x = \begin{pmatrix}x_1\\x_2\end{pmatrix}$ with respect to the standard basis for $\mathbb{R}^2$. Then

$$v_1 \cdot x = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} \cdot \begin{pmatrix}x_1\\x_2\end{pmatrix} = \frac{1}{\sqrt{2}}(x_1 + x_2).$$

and

$$v_2 \cdot x = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix} \cdot \begin{pmatrix}x_1\\x_2\end{pmatrix} = \frac{1}{\sqrt{2}}(x_1 - x_2).$$

Therefore,

$$x = \left(\frac{1}{\sqrt{2}}(x_1 + x_2), \frac{1}{\sqrt{2}}(x_1 - x_2)\right)$$

with respect to the basis B.

Example. Let V be a plane in $\mathbb{R}^3$ defined by the equation $ax + by + cz = 0$. Let $n = \begin{pmatrix}a\\b\\c\end{pmatrix}$. For any vector $u =$

$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in V$, we have

$$n \cdot u = ax + by + cz = 0.$$

Thus n is orthogonal to V. In fact,

$V = \{(x\,y\,z)^T \in \mathbb{R}^3 : \; ax{+}by{+}cz = 0\} = \{u \in \mathbb{R}^3 |\; n{\cdot}u = 0\}.$

The vector n is called a normal vector of V.

Example. Let $V = \operatorname{span}(u_1, u_2)$ be a subspace of $\mathbb{R}^4$, where $u_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, u_2 = \begin{pmatrix} 0 \\ -1 \\ -1 \\ 1 \end{pmatrix}$. Find all vectors that are orthogonal to V.

Solution. Let $v = (w\ x\ y\ z)^T$ be a vector in $\mathbb{R}^4$. Then

$$\begin{aligned} & v \cdot (au_1 + bu_2) = 0 \text{ for all } a, b \in \mathbb{R} \\ \Leftrightarrow\quad & v \cdot u_1 = 0,\ v \cdot u_2 = 0 \\ \Leftrightarrow\quad & \begin{cases} w + x + y = 0 \\ -x - y + z = 0 \end{cases} \\ \Leftrightarrow\quad & (w\ x\ y\ z)^T = (-t,\ -s + t\ s\ t)^T \text{ for some } s, t \in \mathbb{R}. \end{aligned}$$

So a vector v is orthogonal to V if and only if

$$v = (-t,\ -s + t\ s\ t)^T = s(0\ -1\ 1\ 0)^T + t(-1\ 1\ 0\ 1)^T$$

for some $s, t \in \mathbb{R}$, i.e.

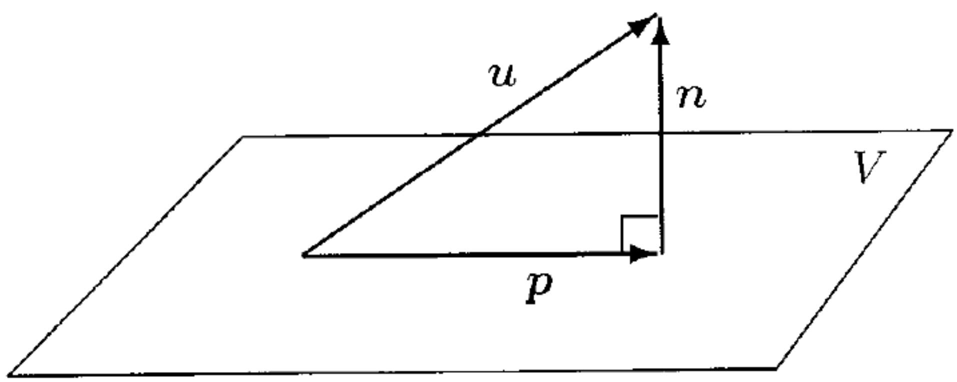
$$v \in \operatorname{span}\left((0\ -1\ 1\ 0)^T, (-1\ 1\ 0\ 1)^T\right).$$

Therefore, span $\left((0\ \ -1\ 1\ 0)^T, (-1\ 1\ 0\ 1)^T\right)$ is orthogonal to V.

Definition. Let V be a subspace of $\mathbb{R}^n$. Every vector $u \in \mathbb{R}^n$ can be written uniquely as

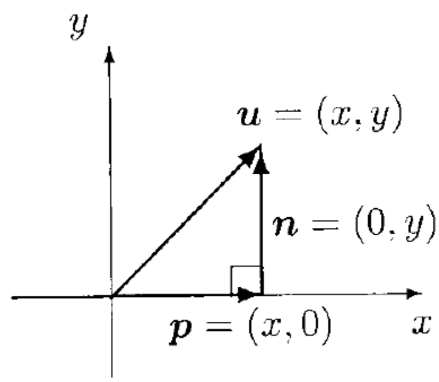
$$u = n + p$$

such that n is a vector orthogonal to V and p is a vector in V. The vector p is called the (**orthogonal**) **projection** of u onto V.



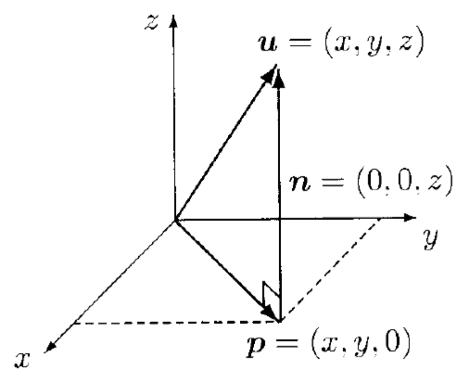
Example 1.

The projection of $u = \begin{pmatrix} x \\ y \end{pmatrix}$ onto the x-axis is $p = \begin{pmatrix} x \\ 0 \end{pmatrix}$. Here, $n = \begin{pmatrix} 0 \\ y \end{pmatrix}$.



Example 2.

The projection of $u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ onto the xy-plane is $p = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$. Here, $n = \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix}$.



Theorem 1.8. *Let V be a subspace of $\mathbb{R}^n$ and w a vector in $\mathbb{R}^n$.*

1. *If $\{v_1, v_2, ..., v_k\}$ be an orthogonal basis for V, then*
$$\frac{w \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{w \cdot v_2}{v_2 \cdot v_2} v_2 + ... + \frac{w \cdot v_k}{v_k \cdot v_k} v_k,$$
is the projection of w onto V.

2. *If $\{v_1, v_2, ..., v_k\}$ is an orthonormal basis for V, then*
$$(w \cdot v_1) v_1 + (w \cdot v_2) v_2 + ... + (w \cdot v_k) v_k,$$
is the projection of w onto V.

Proof. 1. Let
$$p = \frac{w \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{w \cdot v_2}{v_2 \cdot v_2} v_2 + ... + \frac{w \cdot v_k}{v_k \cdot v_k} v_k, \qquad n = w - p.$$
Then for $i = 1, 2, ..., k$,
$$\begin{aligned} n \cdot v_i &= w \cdot v_i - p \cdot v_i \\ &= w \cdot v_i - \frac{w \cdot v_1}{v_1 \cdot v_1}(v_1 \cdot v_i) - ... - \frac{w \cdot v_k}{v_k \cdot v_k}(v_k \cdot v_i) \\ &= w \cdot v_i - \frac{w \cdot v_i}{v_i \cdot v_i}(v_i \cdot v_i) = 0. \end{aligned}$$

This implies that n is orthogonal to V. Since $w = n + p$ where n is orthogonal to V and p is a vector in V, p is the projection of w onto V.

2. This is a consequence of part (1), becasuse

$$v_i \cdot v_i = \|v_i\|_2^2 = 1.$$

$\square$

Example 3.

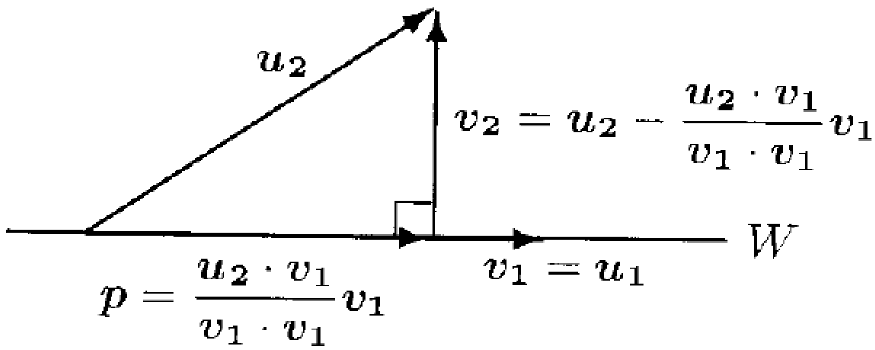
Let V be a subspace of $\mathbb{R}^3$ spanned by the orthogonal vectors $v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$. Then the projection of $w = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ onto V is equal to

$$\frac{w \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{w \cdot v_2}{v_2 \cdot v_2} v_2 = \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Discussion

(a) Let $\{u_1, u_2\}$ be a basis for a vector space V where V is either $\mathbb{R}^2$ or a plane in $\mathbb{R}^3$ containing the origin. Let W be the subspace of V spanned by u_1. (W is a line through the origin.) Then the projection of u_2 onto W is

$$p = \frac{u_2 \cdot u_1}{u_1 \cdot u_1} u_1$$



Define $v_1 = u_1$ and $v_2 = u_2 - p = u_2 - \dfrac{u_2 \cdot v_1}{v_1 \cdot v_1} v_1$. Then $\{v_1, v_2\}$ is an orthogonal basis for V.

(b) Let $\{u_1, u_2, u_3\}$ be a basis for $\mathbb{R}^3$ and let V be the subspace of $\mathbb{R}^3$ spanned by u_1, u_2. Let W be the subspace of V spanned by u_1. (V is a plane containing the origin.) Define

$$v_1 = u_1, \quad v_2 = u_2 - \frac{u_2 \cdot v_1}{v_1 \cdot v_1} v_1.$$

By Part 1, $\{v_1, v_2\}$ is an orthogonal basis for V. Then the projection of u_3 onto V is

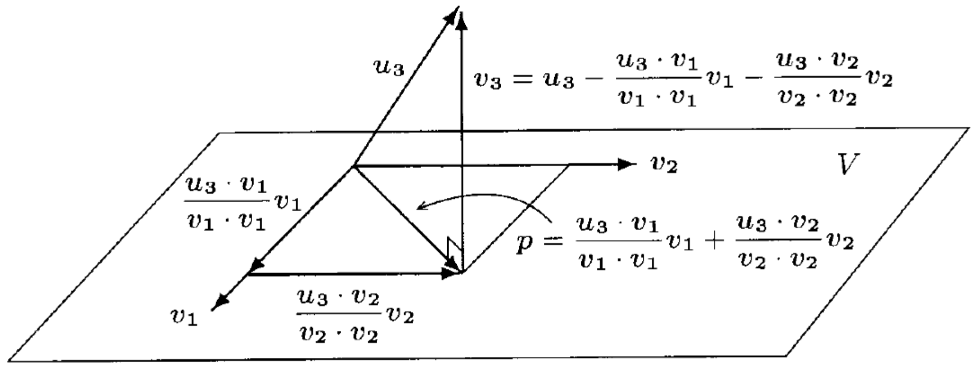
$$p = \frac{u_3 \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{u_3 \cdot v_2}{v_2 \cdot v_2} v_2.$$

Define

$$v_3 = u_3 - p = u_3 - \frac{u_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{u_3 \cdot v_2}{v_2 \cdot v_2} v_2.$$

Then $\{v_1, v_2, v_3\}$ is an orthogonal basis for $\mathbb{R}^3$.

In general, we have the following process, known as **Gram-Schmidt Process**:



Gram-Schmidt Process: Let $\{u_1, u_2, ..., u_k\}$ be a basis for a vector space V. Define

$$\begin{aligned}
v_1 &= u_1, \\
v_2 &= u_2 - \frac{u_2 \cdot v_1}{v_1 \cdot v_1} v_1, \\
v_3 &= u_3 - \frac{u_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{u_3 \cdot v_2}{v_2 \cdot v_2} v_2, \\
&\vdots \\
v_k &= u_k - \frac{u_k \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{u_k \cdot v_2}{v_2 \cdot v_2} v_2 - ... - \frac{u_k \cdot v_{k-1}}{v_{k-1} \cdot v_{k-1}} v_{k-1}
\end{aligned}$$

Then $\{v_1, v_2, v_3\}$ is an orthogonal basis for V.

Furthermore, let

$$w_1 = \frac{1}{\|v_1\|} v_1, w_2 = \frac{1}{\|v_2\|} v_2, ..., w_k = \frac{1}{\|v_k\|} v_k.$$

Then $\{w_1, w_2, w_3\}$ is an orthonormal basis for V.

1.10 Orthonormality and matrices

Let $\{v_1, v_2, ..., v_k\}$ be an orthonormal basis for vector space V. Then

$$
\begin{aligned}
v_i \cdot v_j = v_i^T v_j &= 0 \text{ if } i \neq j \\
\|v_i\| = \sqrt{v_i^T v_i} &= 1 \text{ for all } i.
\end{aligned}
$$

If we use the relatively common notation that

$$
\delta_{ij} = \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{if } i = j. \end{cases}
$$

then the two normality conditions can be written compactly as

$$
v_i \cdot v_j = \delta_{ij} \quad \text{for all } i \in \{1, 2, ..., k\}.
$$

To express the two orthonormality conditions in terms of conditions on matrices, let's define an $n \times k$ matrix A to be

$$
A = (v_1 \quad v_2 \quad ... \quad v_k),
$$

where $v_i = A_{:i}$ is the ith column of A.

Next, let's consider the matrix $A^T A$. Observe that $A^T A$ is a $k \times k$ matrix. Let's ask what information is contained in the (ij) element of this matrix, i.e., in the element $(A^T A)_{ij}$? That is,

$$
(A^T A)_{ij} = v_i^T v_j,
$$

which, in this case, equals 1 or 0, depending on whether or not $i = j$. That is, in this case,

$$
(A^T A)_{ij} = \delta_{ij}
$$

and we can write the matrix product as

$$
A^T A = I_k,
$$

where I_k is the identity matrix of dimension k. This is the matrix way to express that a matrix has columns that form an orthonormal basis.

Given an $n \times n$ matrix A, recall that the inverse matrix A^{-1} is the matrix such that

$$A^{-1}A = AA^{-1} = I_n.$$

So, we have the following definition:

Definition. A square matrix A is called orthogonal if

$$A^{-1} = A^T.$$

From the above discussion, we have the following result:

Theorem 1.9. *Let A be a square matrix of order n. The following statements are equivalent:*

1. *A is orthogonal.*
2. *The columns of A form an orthonormal basis for $\mathbb{R}^n$.*
3. *The rows of A form an orthonormal basis for $\mathbb{R}^n$.*

Suppose that we have a set of vectors $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$, such that any two different vectors v_i and v_j in that set are orthogonal to each other. To organize this information into a matrix, let's define an $n \times k$ matrix A to be

$$A = (v_1 \quad v_2 \quad \ldots \quad v_k).$$

The requirement that the basis vectors be orthogonal to each other means that

$$A^T A = D,$$

where D is a $k \times k$ diagonal matrix, all the diagonal entries of which are positive. The matrix D is diagonal since the off-diagonal elements are the dot products between different basis vectors, which equal zero, since they are orthogonal; and the diagonal entries are all non-zero since each vector in the basis is non-zero and thus has a non-zero non-negative norm.