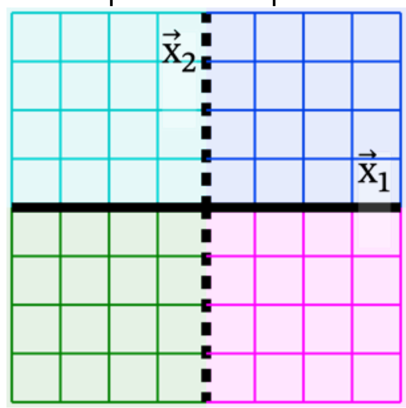
Space, vector, matrix

① One simple example of a space is a 2D grid or $\mathbb{R}^2$



Every point on the grid (denoted as $\vec{v}$) can be described by a set of 2 coordinates s & t along directions $\vec{x}_1$ & $\vec{x}_2$

$$\vec{v} = s\,\vec{x}_1 + t\,\vec{x}_2$$

② Coordinates 's' & 't' written as a column constitute a vector

$$\vec{v} = \begin{bmatrix} s \\ t \end{bmatrix}$$

Suppose we want to move $\vec{v}$ within the grid: we can accomplish it by providing a set of instructions for each component of $\vec{v}$ in the form of

③ a structure called matrix $A = [\,\vec{a}_1 \mid \vec{a}_2\,]$ where

$$\vec{a}_1 = \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} \quad \vec{a}_2 = \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix}$$

- $\vec{a}_1$ has instructions for where 's' component moves
- $\vec{a}_2$ has instructions for where 't' component moves

- the number of components $\vec{a}_1$ & $\vec{a}_2$ has to match the number of scalars s & t

Matrix-vector multiplication

On the previous page, we defined
- Vector as a set of coordinates
- Matrix as a set of instructions for each coordinate

This leads us to the concept of matrix-vector multiplication
or equivalently,
instructions A applied to $\vec{v}$:

$$A\,\vec{v} = [\,\vec{a}_1 \mid \vec{a}_2\,]\begin{bmatrix} s \\ t \end{bmatrix} = \vec{a}_1\, s + \vec{a}_2\, t$$

This translates as: we add together
- 's' amount of column $\vec{a}_1$
- 't' amount of column $\vec{a}_2$

Now suppose we have particular instances of $\vec{v}$, called basis vectors

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

notice that $\vec{e}_1$ & $\vec{e}_2$ correspond to standard coordinate directions

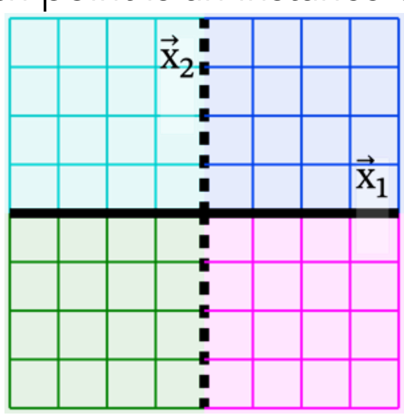
Applying the above multiplication rule:

- $A\,\vec{e}_1 = \vec{a}_1$
- $A\,\vec{e}_2 = \vec{a}_2$

This is another way to view matrix A:
each column of A is a set of directions for the corresponding coordinate
↓
the count of columns must match the dimension of the space
(this is a restatement of the dimension rule from previous page)

Matrix-vector multiplication: two perspectives

Will now go back to the grid of $\mathbb{R}^2$
where each point is an instance of vector $\vec{v}$

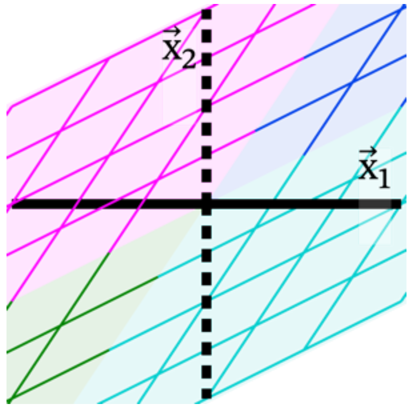


and use an arbitrary example of 'instructions'

$$A = \begin{bmatrix} 1 & 2 \\ 1.5 & 1 \end{bmatrix}$$

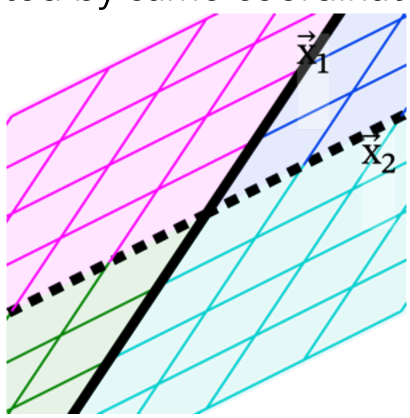
① Perspective 1 shows this 'instructions' applied to every point on the grid
or a transformation by A
'viewed' by the same $\vec{x}_1$ & $\vec{x}_2$ directions:

- $\vec{x}_1$ & $\vec{x}_2$ remain fixed
- every $\vec{v}$ with coordinates 's' & 't' moves to $(\vec{a}_1\, s + \vec{a}_2\, t)$



② Perspective 2 shows the 'instructions' applied to the coordinate system
as shown on the previous page:

- $\vec{x}_1$ moves to $\vec{a}_1$
- $\vec{x}_2$ moves to $\vec{a}_2$
- every $\vec{v}$ is represented by same coordinates in the new system



The same multiplication rule $A\,\vec{v} = \vec{a}_1\, s + \vec{a}_2\, t$ can be viewed in two ways:
① the vectors move while the coordinate directions stay fixed
② the coordinate directions move while the vectors keep the same coordinates

Linearity of matrix-vector multiplication

From the previous page:

$$A\,\vec{v} = \vec{a}_1\, s + \vec{a}_2\, t, \quad \text{where } A = \begin{bmatrix} 1 & 2 \\ 1.5 & 1 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} s \\ t \end{bmatrix}$$

This gives two linearity rules:

① $A\,(\vec{v} + \vec{w}) = A\,\vec{v} + A\,\vec{w}$

② $A\,(c\,\vec{v}) = c\,(A\,\vec{v})$

$$\vec{v} = \begin{bmatrix} s \\ t \end{bmatrix} \quad \vec{w} = \begin{bmatrix} p \\ q \end{bmatrix} \quad \vec{v} + \vec{w} = \begin{bmatrix} s+p \\ t+q \end{bmatrix}$$

① Additivity

Using $A\,\vec{v} = \vec{a}_1\, s + \vec{a}_2\, t$

$A\,(\vec{v} + \vec{w}) = \vec{a}_1\,(s+p) + \vec{a}_2\,(t+q)$

$= (\vec{a}_1\, s + \vec{a}_2\, t) + (\vec{a}_1\, p + \vec{a}_2\, q)$

$= A\,\vec{v} + A\,\vec{w}$

② Scaling

Let c be any scalar

$c\,\vec{v}$ has coordinates [c s ; c t]

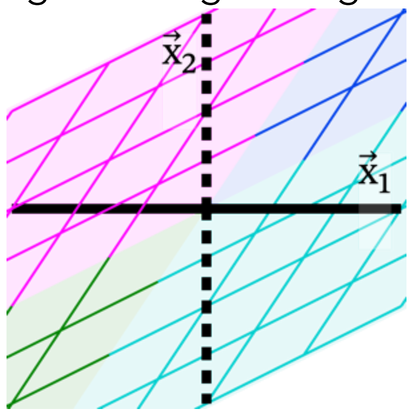
$A\,(c\,\vec{v}) = \vec{a}_1\,(c\, s) + \vec{a}_2\,(c\, t)$

$= c\,(\vec{a}_1\, s + \vec{a}_2\, t)$

$= c\,(A\,\vec{v})$

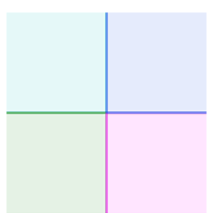
The square grid becomes a skewed grid with preserved addition and scaling:

- straight lines stay straight
- parallel lines stay parallel
- lines going through the origin still go through the origin



Matrix-vector multiplication $\mathbb{R}^2 \to \mathbb{R}^3$

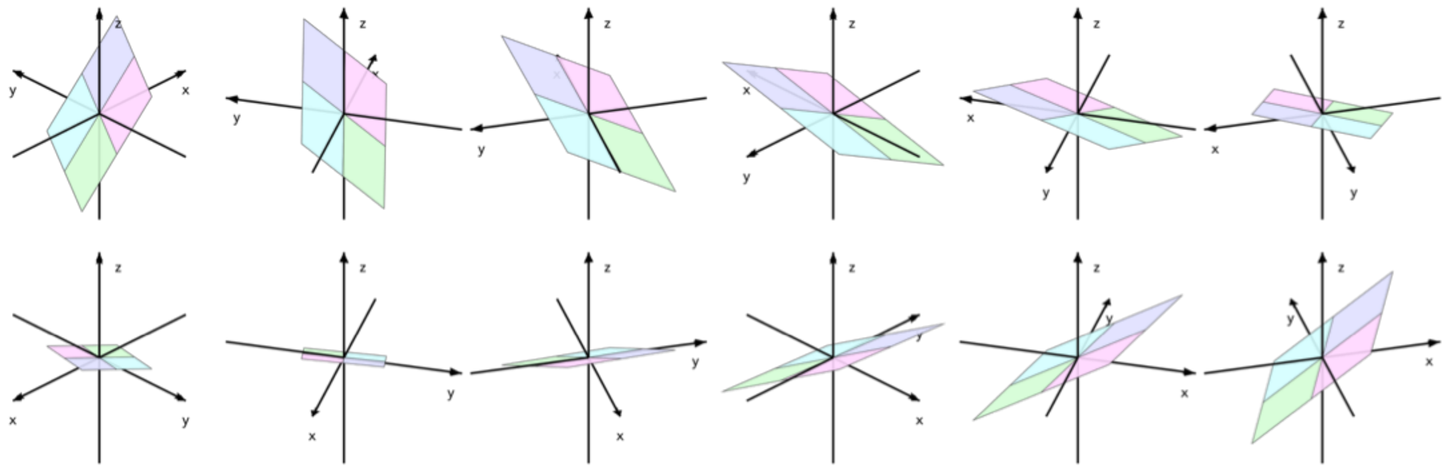
On the previous page, we saw how a 2×2 matrix transforms a square grid



Now we will look at the transformation of the same square grid by a 3×2 matrix

$$A = [\,\vec{a}_1 \mid \vec{a}_2\,] = \begin{bmatrix} 4 & 1.5 \\ 1 & 3 \\ 2 & 1 \end{bmatrix}$$

This maps vectors from $\mathbb{R}^2$ into $\mathbb{R}^3$, as shown



Here, we see the following geometric effects:

1. Scaling: the grid is stretched more in the direction of one column than the other
2. Skew: $\vec{a}_1$ and $\vec{a}_2$ are not orthogonal, so the image is not a rectangle

As a result of 1 and 2, the unit square becomes a slanted parallelogram

3. The parallelogram is encoded in 3D coordinates as

$$x_1 \vec{a}_1 + x_2 \vec{a}_2$$

and is embedded in 3D

Matrix-matrix multiplication

Previously, we defined matrix-vector multiplication $A\vec{b}$ as each column of A providing instructions for each element of $\vec{b}$

We can view the product AB in a similar fashion:

- B provides the vectors we are moving as columns $\vec{b}_1$ & $\vec{b}_2$
- A provides the instructions for how vectors of B should move

$\Leftrightarrow$

the product A B takes each column of B and applies A to it

$A\,B = [\,A\,\vec{b}_1 \mid A\,\vec{b}_2\,]$

From here, you see why order matters

- AB: the vectors come from B and the instructions come from A
- BA: the vectors come from A and the instructions come from B

Later we will see how the above definitions lead to algebraic identities

- $(A\,B)\,\vec{v} = A\,(B\,\vec{v})$
- $(A\,B)\,V = A\,(B\,V)$

Size compatibility:

- For $A\,\vec{b}$ to be defined, A has to have one column for each element of $\vec{b}$
- Similarly, A must have one column for each scalar in every column of B

$\downarrow$

The number of columns of A must equal the number of rows of B

Consider examples of matrices A & B:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad B = [\,\vec{b}_1 \mid \vec{b}_2\,] = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}$$

$$A\,\vec{b}_1 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 10 \end{bmatrix} \quad A\,\vec{b}_2 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 9 \end{bmatrix}$$

$$A\,B = [\,A\,\vec{b}_1 \mid A\,\vec{b}_2\,] = \begin{bmatrix} 4 & 5 \\ 10 & 9 \end{bmatrix}$$

Image 1 shows how every vector $\vec{v}$ on the square grid is transformed by B standard basis $\vec{x}_1$ and $\vec{x}_2$ are kept fixed

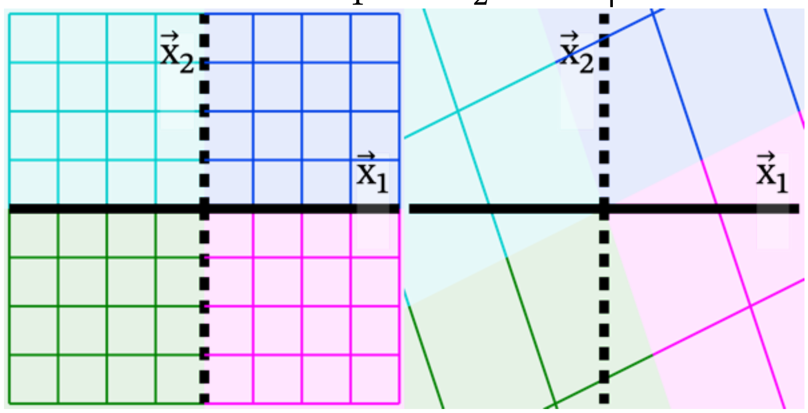
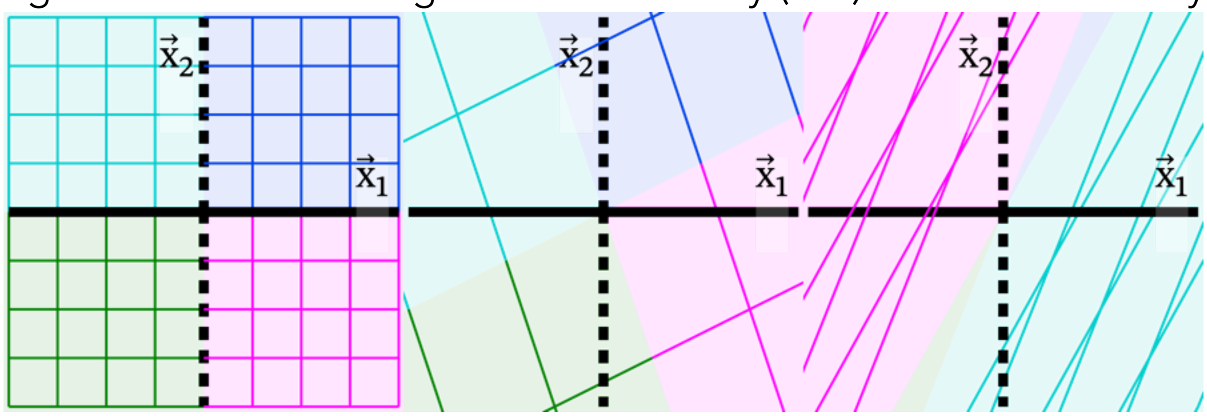


Image 2 shows the same grid transformed by (A B): B first followed by A



Sequence view

- B acts first on the square grid
- A acts on the result

Concept summary

Space	A coordinate grid such as $\mathbb{R}^2$ or $\mathbb{R}^3$ A collection of directions in which vectors live
Vector	A set of coordinates written as a column $\vec{v}$ = (s ; t) represents a point in the space
Matrix A (for A $\vec{v}$)	A set of instructions for each coordinate $A = [\ \vec{a}_1 \mid \vec{a}_2\ ]$
Matrix B (for A B)	Provides the vectors that will be moved Columns $\vec{b}_1$, $\vec{b}_2$ are inputs to A
Matrix A (for A B)	Provides the instructions for columns of B $A\,B = [\ A\,\vec{b}_1 \mid A\,\vec{b}_2\ ]$
Basis vectors	Coordinate directions of the space $\vec{e}_1$, $\vec{e}_2$ determine how coordinates are measured
Canonical basis	Standard orthogonal unit directions $\vec{e}_1 = (1,0,0)$, $\vec{e}_2 = (0,1,0)$, ...

Linear independence & linear dependence

Set of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$

- is linearly independent if each vector contributes a new direction

$\Leftrightarrow$

removing any one vector reduces the span

• is linearly dependent if
at least one of the vectors does not contribute a new direction
⇔
removing that vector does not change the span

Algebraic equivalents:

• Linearly independent:

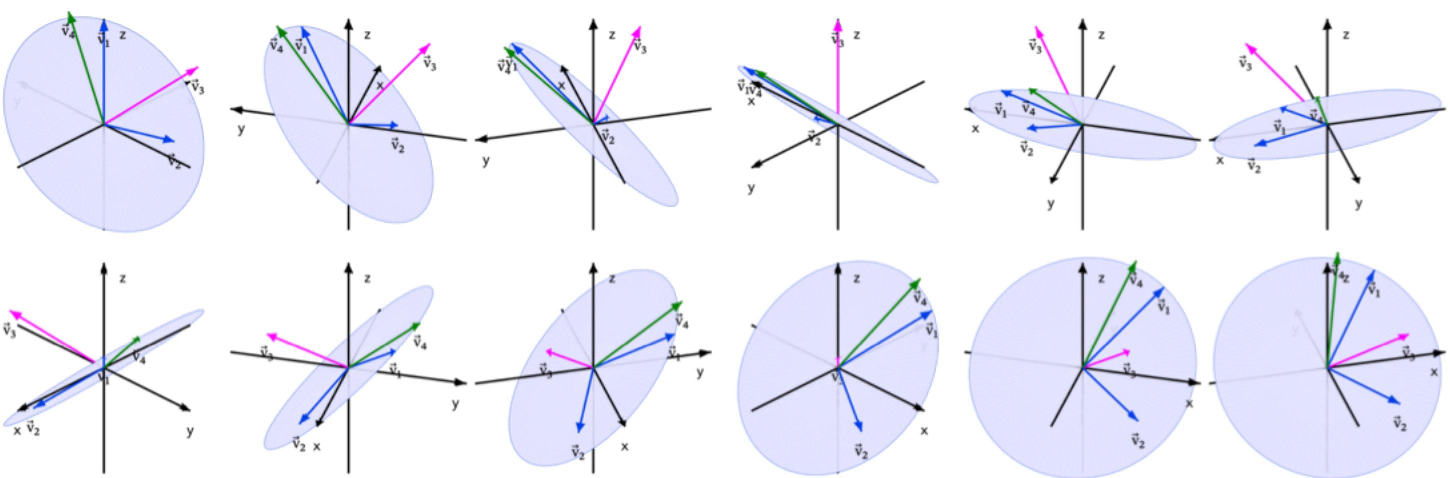
$$s_1\vec{v}_1+s_2\vec{v}_2+\ldots+s_n\vec{v}_n = \vec{0}$$

implies all coefficients s = 0

• Linearly dependent:

$$s_1\vec{v}_1+s_2\vec{v}_2+\ldots+s_n\vec{v}_n = \vec{0}$$

there exists a non-zero choice of coefficients (not all s are 0)



The image above shows

• linearly independent $\vec{v}_1$ & $\vec{v}_2$ (blue arrows)

• all vectors that can be obtained as $s_1\vec{v}_1 + s_2\vec{v}_2$ (plane)

• linearly dependent $\vec{v}_4 = \vec{v}_1 - 0.5\vec{v}_2$ (green arrow)

Vector span & basis of a space

① Vector span describes the space generated by the vectors
Given original vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$,
their span is the set of all vectors
that can be built from the original set:
$s_1\vec{v}_1 + s_2\vec{v}_2 + \ldots + s_n\vec{v}_n$
or all linear combinations of original vectors

Geometric intuition in $\mathbb{R}^3$:
- one vector: line through origin
- collinear or linearly dependent vectors: same line
- two linearly independent vectors: plane through origin
- three vectors where $\vec{v}_3 = a\,\vec{v}_1 + b\,\vec{v}_2$: the $\vec{v}_3$ adds no new dimension
  ↓
  span remains a plane
- three linearly independent vectors
  ↓
  span is now the entire $\mathbb{R}^3$

Last example of three vectors spanning $\mathbb{R}^3$ helps us define

② Basis of a space:
A set of vectors forms a basis for $\mathbb{R}^n$ if it contains n independent directions
↓
any vector in $\mathbb{R}^n$ can be written uniquely as linear combination of the set

For the same reason, any new vector added to the set
will be linearly dependent

③ Canonical basis is a basis with a particular form:

In $\mathbb{R}^3$ the canonical basis is

- $\vec{e}_1 = (1, 0, 0)$
- $\vec{e}_2 = (0, 1, 0)$
- $\vec{e}_3 = (0, 0, 1)$

same pattern extends to $\mathbb{R}^n$

Matrix multiplication vs matrix addition

- Matrix multiplication represents composition of transformations

AB means: apply B first, then apply A

$$(AB)\vec{x} = A(B\vec{x})$$

For multiplication to make sense,
the output space of B must match the input space of A
(inner dimensions must agree)

- Matrix addition represents a linear combination of transformations

$$(A + B)\vec{x} = A\vec{x} + B\vec{x}$$

For addition to make sense,
A and B must have the same domain and the same codomain
Both take vectors from the same space and return vectors in the same space

- Multiplication combines transformations sequentially
- Addition creates a linear combination of transformations

(this will be discussed further in 'Rank and matrix composition' chapter)

Dot product $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

- A function that takes two vectors and returns a scalar
- Defined to satisfy the following rules

① Linearity

- $(a\,\vec{v}_1 + b\,\vec{v}_2) \cdot \vec{w} = a\,(\vec{v}_1 \cdot \vec{w}) + b\,(\vec{v}_2 \cdot \vec{w})$
- $\vec{v} \cdot (a\,\vec{w}_1 + b\,\vec{w}_2) = a\,(\vec{v} \cdot \vec{w}_1) + b\,(\vec{v} \cdot \vec{w}_2)$

② Symmetry

$\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v}$

③ Non-negativity

$\vec{v} \cdot \vec{v} \geq 0$

$\vec{v} \cdot \vec{v} = 0 \Leftrightarrow \vec{v} = \vec{0}$

④ Compatibility with Euclidean length

For the standard dot product in $\mathbb{R}^n$, the length is

$$|\vec{v}| = \sqrt{\vec{v} \cdot \vec{v}}$$

Algebraic definition that satisfies the above rules is

$$\vec{v} \cdot \vec{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$$

If $|\vec{w}| = 1$

Then $\vec{v} \cdot \vec{w}$ equals the signed length of the projection of $\vec{v}$ onto $\vec{w}$

For nonzero vectors,

$\vec{v} \cdot \vec{w} = |\vec{v}|\,|\vec{w}| \cos\theta$

where θ is the angle between them

(this identity can be derived using the law of cosines)

Definition of orthogonality

Vectors $\vec{v}$ and $\vec{w}$ are orthogonal if $\vec{v} \cdot \vec{w} = 0$

The dot product concept will soon help us

- understand the concept of linear equations
- introduce the concept of the four subspaces

Geometry of dot product

Consider an example of two vectors:

$$\vec{v} = \begin{bmatrix} 0.5 \\ 1.5 \end{bmatrix} \quad \vec{w} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Dot product is symmetric:

$\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v}$

$\vec{v} \cdot \vec{w}$ measures how much $\vec{v}$ lies in the direction of $\vec{w}$

It equals the length of the projection of $\vec{v}$ onto the direction of $\vec{w}$ multiplied by $|\vec{w}|$

$\vec{w} \cdot \vec{v}$ measures how much $\vec{w}$ lies in the direction of $\vec{v}$

It equals the length of the projection of $\vec{w}$ onto the direction of $\vec{v}$ multiplied by $|\vec{v}|$

Image 1: $\vec{v} \cdot \vec{w}$ as component of $\vec{v}$ along $\vec{w}$

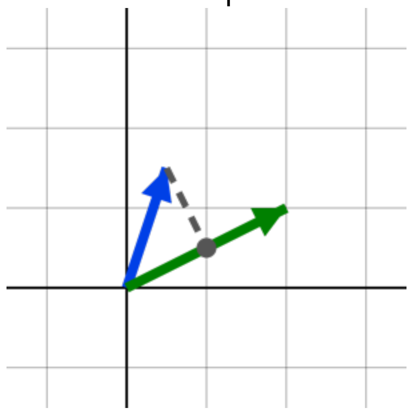
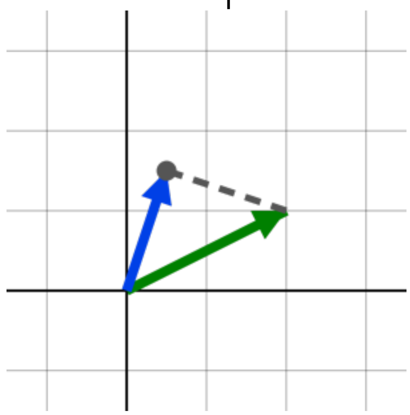


Image 2: $\vec{w} \cdot \vec{v}$ as component of $\vec{w}$ along $\vec{v}$



Sign and angle

- $\vec{v} \cdot \vec{w} > 0$ when the angle between $\vec{v}$ and $\vec{w}$ is less than 90°
- $\vec{v} \cdot \vec{w} = 0$ when the angle is 90° (orthogonal)
- $\vec{v} \cdot \vec{w} < 0$ when the angle is greater than 90°

Orthogonality

① Definition

Two vectors $\vec{v}$ and $\vec{w}$ are called orthogonal if

$$\vec{v} \cdot \vec{w} = 0$$

In $\mathbb{R}^2$ and $\mathbb{R}^3$ this corresponds to a right angle
but the definition via dot product works in $\mathbb{R}^n$

② Connection to linear independence

Orthogonal vectors have zero projection onto each other
They share no directional component
This is independence in its strongest geometric form:

- linearly dependent $\vec{v}$ & $\vec{w}$ completely overlap in direction
- linearly independent non-orthogonal $\vec{v}$ & $\vec{w}$ partially overlap in direction
- orthogonal $\vec{v}$ & $\vec{w}$ have zero overlap

③ Connection to canonical basis

Recall the canonical basis in $\mathbb{R}^3$:

$$\vec{e}_1 = (1, 0, 0)$$
$$\vec{e}_2 = (0, 1, 0)$$
$$\vec{e}_3 = (0, 0, 1)$$

- Canonical basis is orthogonal by construction:

$$\vec{e}_i \cdot \vec{e}_j = 0 \text{ whenever } i \neq j$$

- all vectors $\vec{e}_i$ have length 1 (unit length)

④ Looking ahead

- we will use this definition to derive the formula for projection of $\vec{v}$ onto $\vec{w}$
- the projection formula will help us solve many common problems

Matrix transpose

We previously defined matrix A as a collection of vectors $\vec{a}$:

$$A = [\,\vec{a}_1 \mid \vec{a}_2\,] = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

Transpose of A, or A^T, is obtained by turning rows into columns

$$A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{bmatrix}$$

- Columns of A become rows of A^T
- Rows of A become columns of A^T
- If A is $m \times n$, then A^T is $n \times m$

① If $\vec{a}$ and $\vec{b}$ are column vectors in $\mathbb{R}^n$, $\vec{a} \cdot \vec{b}$ can also be written as $\vec{a}^T \times \vec{b}$
Since $\vec{a}^T$ is a $1 \times n$ row vector and $\vec{b}$ is an $n \times 1$ column vector, their matrix product is a 1×1 scalar

② $\vec{a}^T$ notation allows us to present A as

$$[\,\vec{a}_1 \mid \vec{a}_2\,] = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} = \begin{bmatrix} (\vec{r}_1)^T \\ (\vec{r}_2)^T \\ (\vec{r}_3)^T \end{bmatrix}$$

(where $\vec{r}_i^{\,T}$ are rows of A)

③ $\vec{a}^T$ notation also allows us to present the product AB as

$$A \times B = \begin{bmatrix} (\vec{r}_1)^T \\ (\vec{r}_2)^T \\ (\vec{r}_3)^T \end{bmatrix} \begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \vec{b}_3 \end{bmatrix}$$

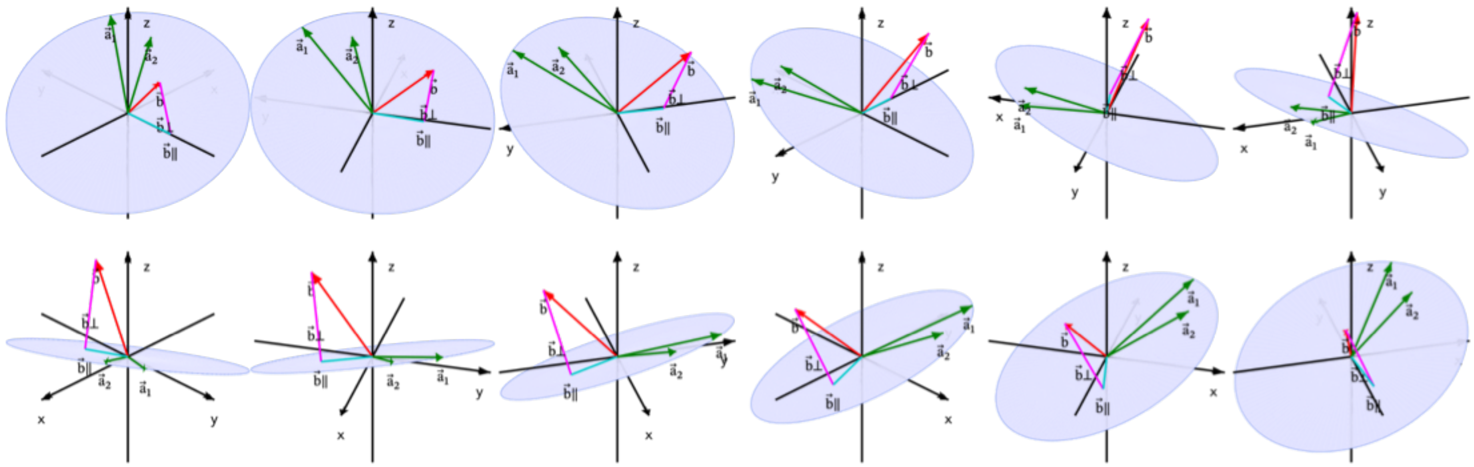
In particular, $A^T A$ can be written as $\begin{bmatrix} \vec{a}_1 \\ \vec{a}_2 \end{bmatrix} \begin{bmatrix} \vec{a}_1 & \vec{a}_2 \end{bmatrix}$

So far, we saw how transpose assists in arranging vector or matrix entries

It also helps bridge matrix domain and co-domain

④ A maps vectors from its domain ($\mathbb{R}^n$) into its codomain ($\mathbb{R}^m$)
A^T maps vectors from the codomain ($\mathbb{R}^m$) back into the domain ($\mathbb{R}^n$)
(This is not an inverse process, distortion usually occurs)

⑤ Later, we will see how A^T allows us to compute the closest reconstruction of $\vec{b}$ from $\vec{a}_1$ & $\vec{a}_2$



④ & ⑤ will be continued in 'Four Subspaces' chapter

Linear equations, first encounter

We showed how a matrix acts as an operator on vectors:
given a vector $\vec{x}$, we compute $A\vec{x}$ and obtain a new vector $\vec{b}$

When solving linear equations, we instead know A and $\vec{b}$ and want to compute $\vec{x}$

Suppose we have an algebraic system of equations

$$\left\{\begin{array}{cccc|cc} a_{11}x_1 & + & a_{12}x_2 & & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & & = & b_2 \end{array}\right.$$

This system can be written in several equivalent forms

① Each equation above is a dot product between a row of A and the vector $\vec{x}$
Equivalently, these rows are the columns of A^T
If $\vec{r}_1$ and $\vec{r}_2$ denote the rows of A, then the system is equivalently written as

$$\vec{r}_1 \cdot \vec{x} = b_1$$
$$\vec{r}_2 \cdot \vec{x} = b_2$$

② From the column viewpoint, the same system can be written in vector form:

$$x_1 \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

This matches the matrix–vector multiplication definition used earlier, therefore,

③ We can assemble it into the compact equation

$$A\vec{x} = \vec{b},$$

$$\text{where } A = [\ \vec{a}1 \mid \vec{a}2\ ] = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

In the next chapter, we describe an algorithm that simplifies equation systems without changing their solution sets