

Linear Algebra + Stats + Probability Notes

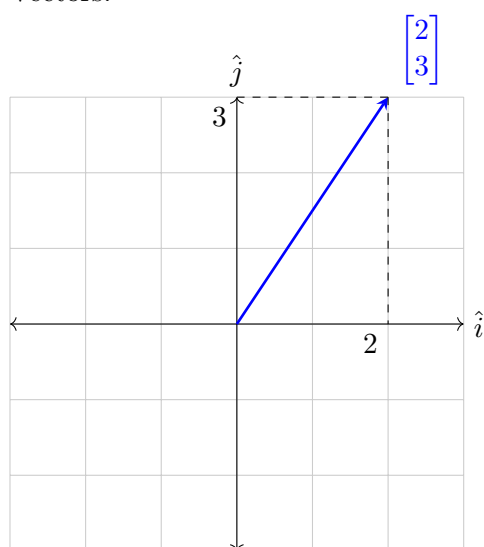
Bhargav github.com/brpy

February 27, 2021

0.1 Linear Algebra:

The following notes is primarily made for my revision. I might have skipped some topics that seemed obvious to me. I do not guarantee factual correctness of the notes. If you feel there are any errors, open a github issue/pr. Notes material is collected from various sources. Image credits are given in tex document and in src.csv

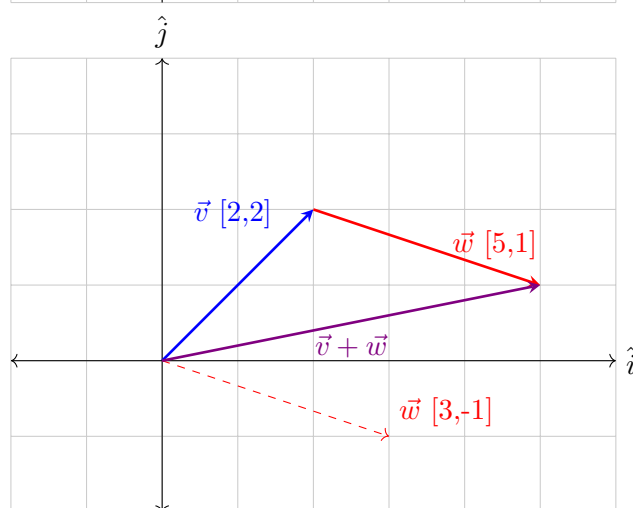
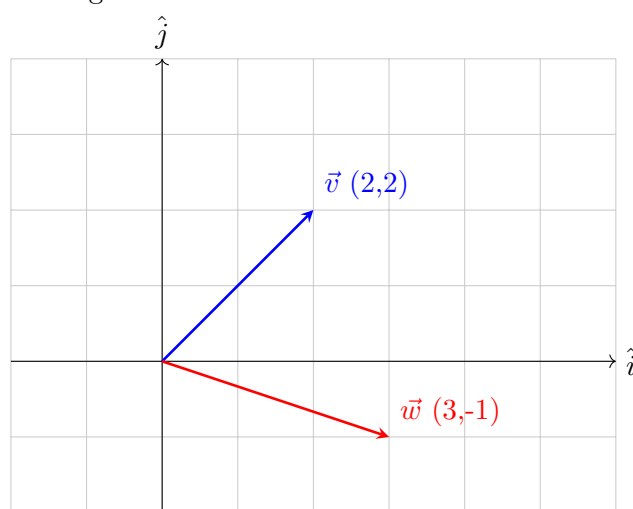
Vectors:



- This vector is represented as $2\hat{i} + 3\hat{j}$ where \hat{i} and \hat{j} are unit vectors perpendicular to each other also known as basis vectors (in 2d).
- It is also represented as a column matrix $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$
- A vector is represented by its length and direction wrto. basis vectors \hat{i}, \hat{j} .
- A vector can be freely moved without changing its length and the direction it is pointing to.
- So any vector in space can be represented using a

linear combination of \hat{i}, \hat{j} by moving the starting point to origin.

Adding 2 Vectors:



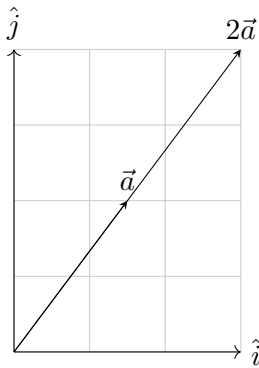
This is also known as Triangular law of vector addition.

- Can also be interpreted as,

$$\begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$
- The vector \vec{w} is moved along \vec{v} such that the starting point of \vec{w} meets the end point of \vec{v}

- Both \vec{v} and \vec{w} direction and lengths remain unchanged.

Scaling:



- 2 *scales* the vector \vec{a} . So, 2 is a Scalar.
- Similarly basis vectors \hat{i} and \hat{j} can be scaled to represent any vector in 2d plane.

Span:

- $a\vec{v} + b\vec{w} \implies$ Linear combination of \vec{v} , \vec{w}
- Set of all vectors of linear combination of \vec{v} , \vec{w} ; $a\vec{v} + b\vec{w}$ is called **span**.
- For most vectors span consists of all points on the plane.
- If \vec{v} , \vec{w} lie on same line, span is a line passing through origin.
- If both \vec{v} , \vec{w} are zero, span is zero.

Linear (in)dependent:

- If one vector can be represented as a linear combination of other, then the vectors are linearly dependent.
- A linearly dependent vector doesn't add to span of a vector.
- If one vector adds a dimension to a span of a vector, they are linearly independent.

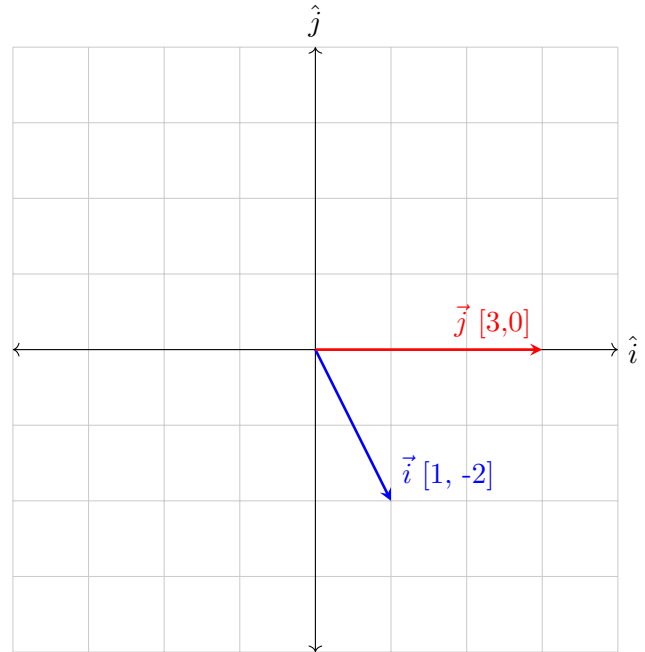
Transformation:

- Takes a vector and gives an output vector.

Linear Transformation:

The conditions for the transformation to be linear are:

- All lines must remain lines.
- Origin must remain in place.
- This makes all equidistant parallel lines remain equidistant and parallel.
- Matrices = Transformation of space.
- For example, the linear transform $\begin{bmatrix} 1 & 3 \\ -2 & 0 \end{bmatrix}$ transforms the 2d plane such a way that the new \hat{i} lands at $\begin{bmatrix} 1\hat{i} \\ -2\hat{j} \end{bmatrix}$ and \hat{j} lands at $\begin{bmatrix} 3\hat{i} \\ 0\hat{j} \end{bmatrix}$



The new basis vectors are blue \vec{i} and red \vec{j} .

Now in this new transformed space, every vector has to be represented as a linear combination of these two new basis vectors.

The vector $\vec{v} = -1\hat{i} + 2\hat{j}$ after transformation lands at the point $\begin{bmatrix} 5 \\ 2 \end{bmatrix}$ in the new vector space.

Even after the transformation, the linear combination doesn't change. So, \hat{i} and \hat{j} are replaced by $\begin{bmatrix} 1\hat{i} \\ -2\hat{j} \end{bmatrix}$ and $\begin{bmatrix} 3\hat{i} \\ 0\hat{j} \end{bmatrix}$ respectively.

So the new transformed \vec{v} becomes,

$$\vec{v}_{new} = -1\hat{i}_{new} + 2\hat{j}_{new} \quad (1)$$

$$\vec{v}_{new} = -1 \begin{bmatrix} 1\hat{i} \\ -2\hat{j} \end{bmatrix} + 2 \begin{bmatrix} 3\hat{i} \\ 0\hat{j} \end{bmatrix} \quad (2)$$

$$\vec{v}_{new} = \begin{bmatrix} -1\hat{i} \\ 2\hat{j} \end{bmatrix} + \begin{bmatrix} 6\hat{i} \\ 0\hat{j} \end{bmatrix} \quad (3)$$

$$\vec{v}_{new} = \begin{bmatrix} 5\hat{i} \\ 2\hat{j} \end{bmatrix} \quad (4)$$

For any vector $x\hat{i} + y\hat{j}$, after applying the transformation $\begin{bmatrix} 1 & 3 \\ -2 & 0 \end{bmatrix}$ becomes,

$$\begin{bmatrix} x\hat{i} \\ y\hat{j} \end{bmatrix} \xrightarrow{\text{lr. transform}} x \begin{bmatrix} 1\hat{i} \\ -2\hat{j} \end{bmatrix} + y \begin{bmatrix} 3\hat{i} \\ 0\hat{j} \end{bmatrix} = \begin{bmatrix} 1x + 3y\hat{i} \\ -2x + 0y\hat{j} \end{bmatrix} \quad (5)$$

Removing \hat{i}, \hat{j} for legibility,

$$\begin{bmatrix} x \\ y \end{bmatrix} \xrightarrow{\text{lr. transform}} x \begin{bmatrix} 1 \\ -2 \end{bmatrix} + y \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 1x + 3y \\ -2x + 0y \end{bmatrix} \quad (6)$$

This is the basis for Matrix multiplication.

- $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ is where \hat{i} lands and $\begin{bmatrix} 3 \\ 0 \end{bmatrix}$ is where \hat{j} (basis vectors) lands after the transformation.
- So the 2X2 matrix $\begin{bmatrix} 1 & 3 \\ -2 & 0 \end{bmatrix}$ itself can represent the transformation.
- This explains the *rules* for multiplication and why matrix multiplication is not commutative.
- This also explains why the matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is an identity matrix. Since, this *transform* actually does nothing, \hat{i} and \hat{j} remain unchanged.

For any transformation $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$,

$$\underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\text{Transformation}} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\text{old vector where } \hat{i} \text{ lands}} = x \underbrace{\begin{bmatrix} a \\ c \end{bmatrix}}_{\text{where } \hat{i} \text{ lands}} + y \underbrace{\begin{bmatrix} b \\ d \end{bmatrix}}_{\text{where } \hat{j} \text{ lands}} = \underbrace{\begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}}_{\text{Transformed matrix}} \quad (7)$$

- $\begin{bmatrix} a \\ c \end{bmatrix}$ and $\begin{bmatrix} b \\ d \end{bmatrix}$ are the new basis vectors; where old \hat{i} and \hat{j} land. These are the new transformed basis vectors.
- Now these new basis vectors have to be used to represent all the vectors in its span. In other words the linear combination of these two basis vectors.
- From equation 7 it can be observed that the scalars x and y scale the corresponding new basis vectors.

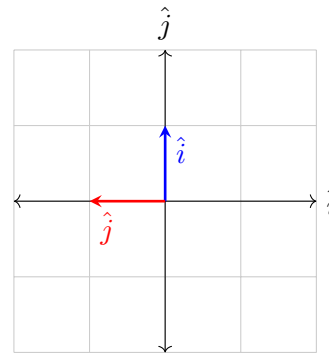
- For no transformation (or) Identity transform / Multiplication,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \quad (8)$$

Which makes sense !!

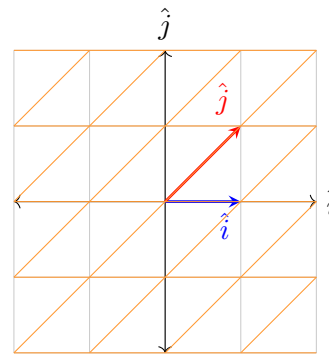
Counterclock Transform:

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$



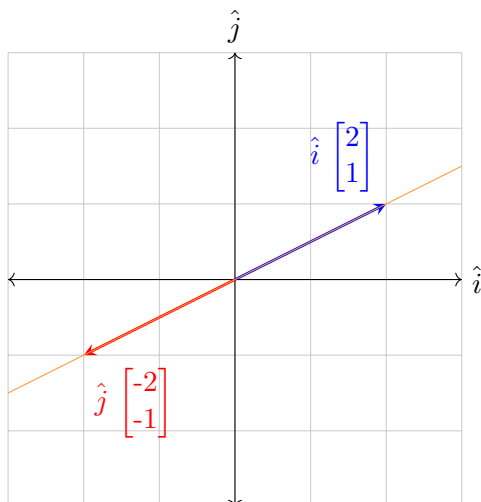
Shear Transform:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$



Few notable points:

- In transformations, order matters. $M_1 M_2 \neq M_2 M_1$ because $f(g(x)) \neq g(f(x))$
- The associative property $(AB)C = A(BC)$ holds true because the order is C, B then A regardless.
- For a $\begin{bmatrix} 2 & -2 \\ 1 & -1 \end{bmatrix}$ transformation all 2d space is squished into a line.



These two vectors $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -2 \\ -1 \end{bmatrix}$ are linearly dependent vectors.

- Two transform for ex. shear and rotation can be composed into a single transform. This composition transform is nothing but a multiplication of shear and rotation matrix.

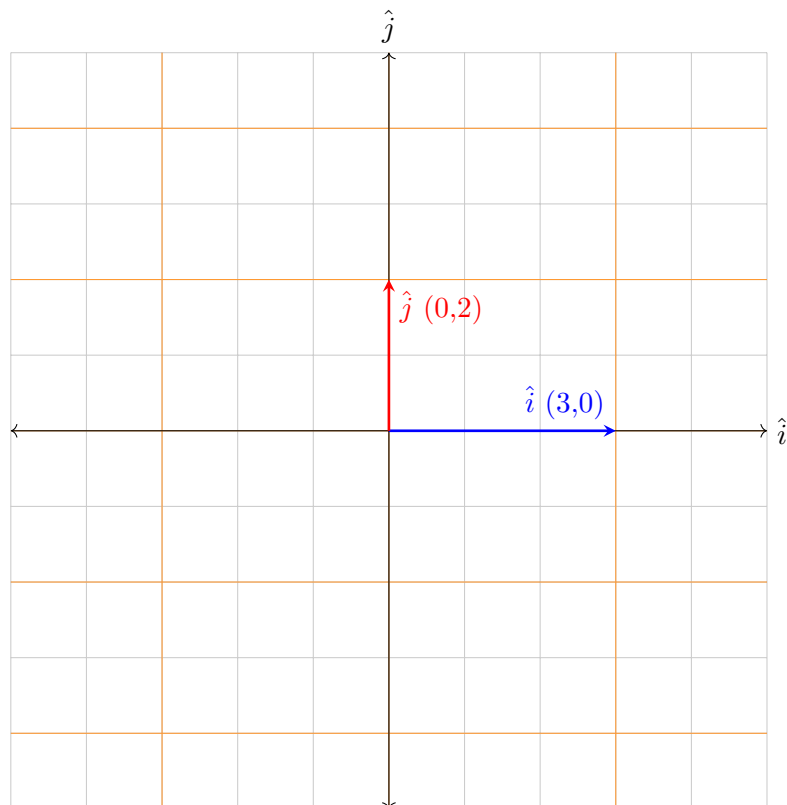
$$\underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}_{\text{Shear}} \underbrace{\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}}_{\text{Rotation}} \begin{bmatrix} x \\ y \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}}_{\text{Composition}} \begin{bmatrix} x \\ y \end{bmatrix} \quad (9)$$

- Product or composition of two matrices/ transforms,

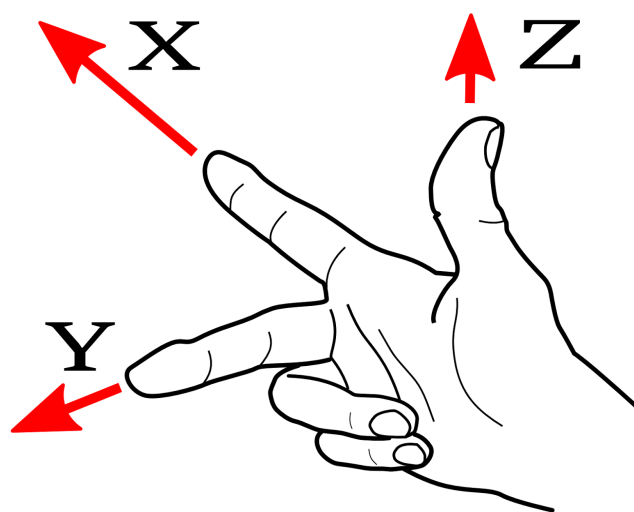
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix} \quad (10)$$

Determinant:

- Transformation $\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ scales \hat{i} by factor 3 \hat{j} by a factor 2.



- We observe that the unit square in transformed space is scaled from 1 sq. unit to 3 sq. units. So the determinant of the transform/matrix $\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ is 3. In case of 3d, volume is scaled.
- For a matrix/transform $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the determinant is $ad - bc$.
- The determinant can be $-ve$ if the axis cross each other during the transformation. It has the effect of *flipping* or *inverting* the space.



- This can be checked using right hand rule to check if the axis still lie in the same orientation after transformation. If not, the determinant is $-ve$.
- For 3d space, determinant is measured by unit volume instead of area.
- Determinant can be zero if the space is squished. For ex. if transformed to a line or point in 2d or to a plane, line or a point in 3d. This has the effect of reduction in number of dimensions.

$$\det(M_1 M_2) = \det(M_1) \det(M_2) \quad (11)$$

System of equations:

$$A\vec{X} = \vec{V}$$

$$\begin{aligned} 3x + 1y + 4z &= 1 \\ 3x + 9y + 2z &= 6 \\ 3x + 3y + 3z &= 8 \end{aligned}$$

- Solving these system of equations imply, finding the vector $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ that if applied the transformation $\begin{bmatrix} 3 & 1 & 4 \\ 3 & 9 & 2 \\ 3 & 3 & 3 \end{bmatrix}$, would land on the vector $\begin{bmatrix} 1 \\ 6 \\ 8 \end{bmatrix}$.
- This can only be solved if A^{-1} exists. Since space cannot be unpacked since there is/are lost dimension(s) if $\det(A) = 0$. So $\det(A) \neq 0$ has to be true for the system of equations to be solved using,

$$\vec{X} = A^{-1}\vec{V} \quad (12)$$

Rank:

- Rank is the number of dimensions of the transformed space.
- The Rank of a matrix ; output space.
 Rank : 1 \implies output transformation : Line
 Rank : 2 \implies output transformation : Plane
 Rank : 0 \implies output transformation : Point

Column space:

- Set of all possible linear combinations or span of column vectors of A . Or,
- Set of all possible $A\vec{V}$

Null space:

- Set of vectors that get squished into origin after transformation.

Dot Product:

- \vec{a} and \vec{b} are two vectors and angle between them is θ .

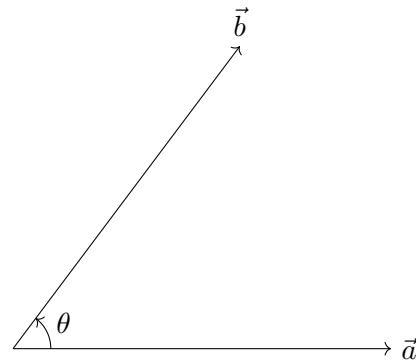
$$\vec{a} = [a_1, a_2, \dots, a_n] \quad (13)$$

$$\vec{b} = [b_1, b_2, \dots, b_n] \quad (14)$$

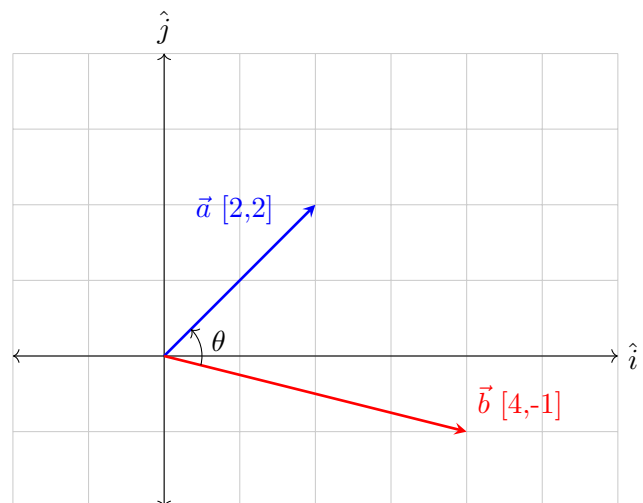
$$\vec{a} \cdot \vec{b} = (a_1) * (b_1) + (a_2) * (b_2) + \dots + (a_n) * (b_n) \quad (15)$$

- Geometric representation:

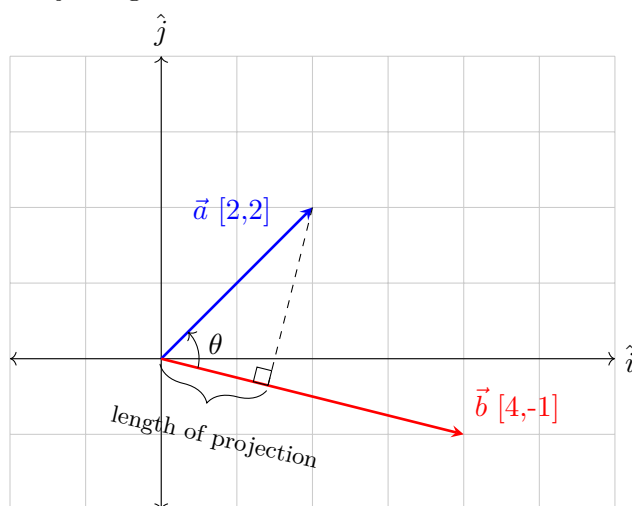
$$\vec{a} \cdot \vec{b} = ||a|| * ||b|| * \cos\theta \quad (16)$$



dot product of two vectors = (length of any vector) x (length of projection made on the vector)



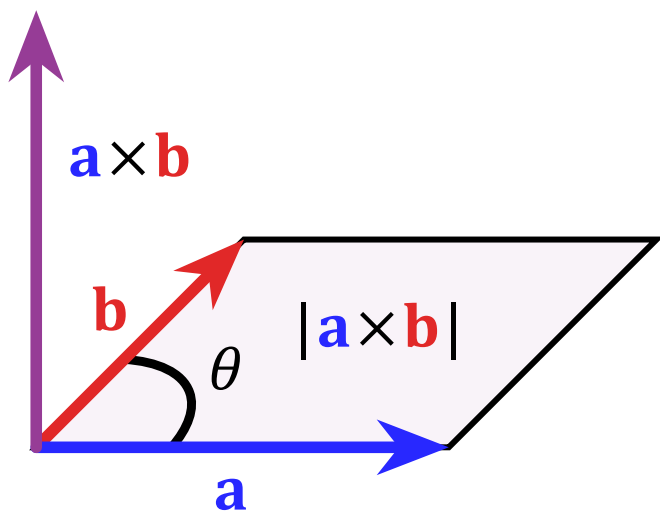
Projecting \vec{a} on \vec{b} :



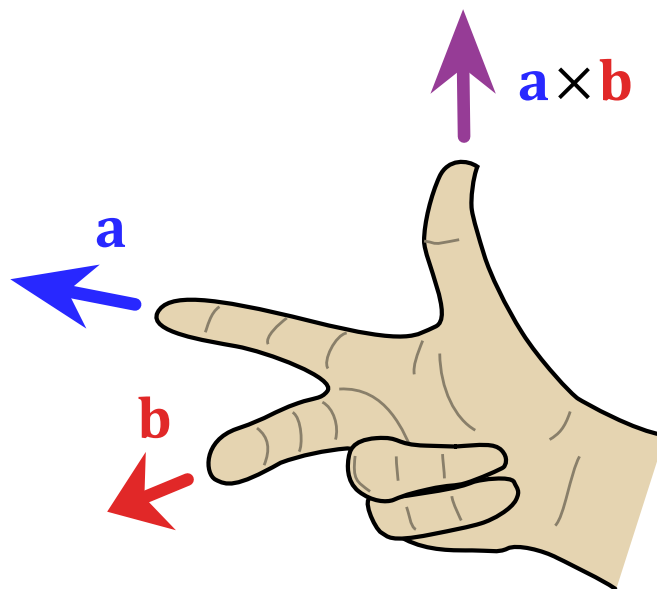
- Length of projection = $\|a\|\cos\theta$
- Dot product = $\|b\| \cdot \|a\|\cos\theta$
- The length of projection does not depend on the length of \vec{b} . It only depends on length of \vec{a} and θ .
- Projecting \vec{b} on \vec{a} will result in the same dot product result. The order does not matter.
- If projection does not lie between origin and the end point of the other vector, you can extend the other vector since length of projection is not affected by it.
- Dot product is -ve if projection is on -ve side.

Cross Product:

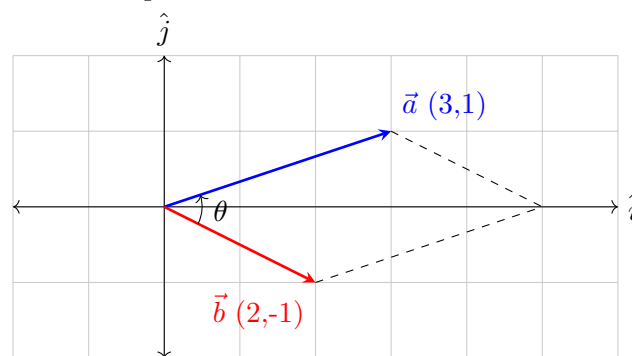
- Directed area product.



- $\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}$



- Length of $\vec{a} \times \vec{b} = \det\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}$
- The cross product vector is normal to the area.



- The area of this parallelogram is $\det\begin{pmatrix} 3 & 2 \\ 1 & -1 \end{pmatrix}$
- The direction of the cross product is normal to the area, i.e., in the direction of \hat{k}
- The magnitude of this vector is the area of this parallelogram i.e., determinant.

$$\vec{a} \times \vec{b} = \|a\| \cdot \|b\| \sin\theta \quad (17)$$

Eigen:

- For a few transformations, vectors only scale. i.e., only stretch or compress.
- Such vectors are called eigen vectors of the transformation.
- Such vectors are transformed into it's own span.

- The scaled value of such vectors (scalar) is the eigen value. It can be +ve or -ve.

$$A\vec{V} = \lambda\vec{V} \quad (18)$$

where:

A is the transformation,

\vec{V} is the eigen vector and,

λ is the eigen value.

Vector from origin:

- As discussed previously, any vector can be moved to origin for reference, without changing it's direction it is pointing to and it's length.

- To make a vector starting from $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$ and ending

at $\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ as a reference vector, we can apply triangular law of addition to get a vector which starts from origin and is equivalent with the vector starting and ending at the above two points.

- We need to remember that the above vectors themselves start at origin. So, we can get the

vector by doing $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$

Matrices:

Diagonal Matrix:

- A matrix with all 0's except diagonal elements.

Identity Matrix:

- A Matrix with all diagonal elements as 1.

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Scalar Matrix:

- A Matrix with only equal diagonal elements.

- A scalar matrix can be represented by a single scalar.

$$K.I = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Upper Triangular Matrix:

- A Matrix with all elements below diagonal are 0

$$\begin{bmatrix} 3 & 5 & 2 \\ 0 & 2 & 7 \\ 0 & 0 & 1 \end{bmatrix}$$

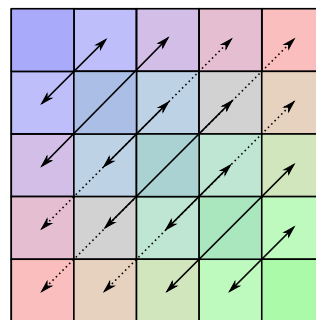
Lower Triangular Matrix:

- A Matrix with all elements above diagonal are 0

$$\begin{bmatrix} 3 & 0 & 0 \\ 7 & 2 & 0 \\ 5 & 3 & 1 \end{bmatrix}$$

Symmetric Matrix:

- A square matrix which is equal to its transpose.
- $A = A^T$
- $a_{ij} = a_{ji}$ for every i and j.



Skew Symmetric Matrix:

- A square matrix which is equal to negative of its transpose.
- $A = -A^T$
- $a_{ij} = -a_{ji}$ for every i and j.

Orthogonal Matrix:

- A square matrix whose column vectors and row vectors are orthonormal.
- $AA^T = A^T A = I$ or,
- $A^T = A^{-1}$
- $a_{ij} = -a_{ji}$ for every i and j.

Lines, Planes and Hyperplanes:

- **Line** - A line has 2 degrees of freedom. You need a point and a slope or two points to uniquely identify a line.
- Equation of a line
 1. $y = mx + c$; slope = m ; y - intercept = c ; passes through (0, c)
 2. $ax + by + c = 0$ is a simple linear form with 2 unknowns
 3. $y - y_1 = m(x - x_1) + c$; slope = m ; passes through x_1 and y_1
- Slope is the *tan* of angle x axis makes with the line in anti-clock wise direction.
- Any 2 points on a line can uniquely identify the line.
- An object on a line can move in 1 dimension.

- **Plane** - A plane is a flat 2d surface.
- A xy plane is an example of a plane.
- To represent a plane you need 3 non collinear points or a point and a normal.
- A normal is a vector that is perpendicular to all vectors on the surface.
 1. So, by definition a plane is set all (x, y, z) points that satisfy,

$$\text{dot}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}, \vec{n}\right) = 0 \quad (19)$$

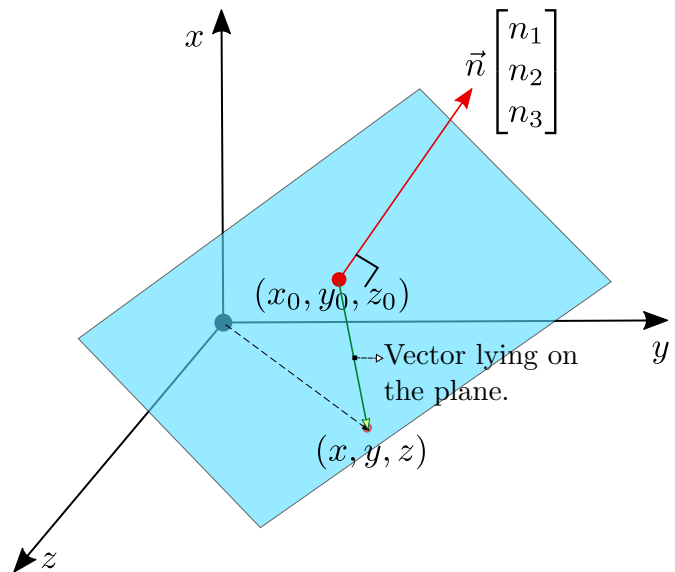
Where:

$\vec{n} = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}$ is the normal to the plane and

starts at point $\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}$ which is located on the plane.

- The equation subsequently solves to,

$$n_1(x - x_0) + n_2(y - y_0) + n_3(z - z_0) = 0 \quad (20)$$



- **Hyperplane** - Is an extension to Line and a plane in higher dimensions.
- A **Linear regression** model can successfully represent a line, plane or a hyperplane.

Miscellaneous:

- Distance between two points (x_1, y_1) and (x_2, y_2) is

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

- Length of a vector $a\hat{i} + b\hat{j} + c\hat{k}$ is

$$\sqrt{a^2 + b^2 + c^2}$$

also known as L2 Norm. (Check knn notes for info on norms.)

- Distance from a plane $ax + by + cz = 0$ to a point (x_1, y_1) is

$$\frac{|ax_1 + by_1 + cz_1|}{\sqrt{a^2 + b^2 + c^2}}$$

Credits:

Most of this notes is taken from 3Blue1Brown's Essence of Linear Algebra playlist. Some images are taken from wikipedia.