SOMEBODY TRY TO EXPLAIN

LINEAR ALGEBRA

written by Yang: GuanZhong

CATALOGUE

Introduction of Vector

We shall firstly consider a vector as the sum of several basis vectors:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = x\vec{i} + y\vec{j} + z\vec{k} \qquad \text{for } \vec{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; \vec{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}; \vec{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Through this way, we could easy solve the problems about the vectors addition and the scale multiplication.

Definition of Linearity

Then we should explain what linear combination means in a funny and intuitive way.

Linear combination: for $\vec{r} = \alpha\vec{x} + \beta\vec{y}$, if you fix one of those scalars(α, β) and let the other one change its value freely, the tip of the resulting vector draws a straight line.

We'll come back to the exact explanation later on.

Understandably, combinations like $\vec{r} = \alpha\vec{v} + \beta\vec{w} + ...$ are linear.

As we can get any points that this combination could lie on by changing the value of scalars, we just define the space/volume/... that made up by these points.

Span: the set of all possible points that you can reach with a linear combination of given vectors is called the 'span' of those vectors.

But sometimes, when $\vec{v} = k\vec{w}$, the span of $\vec{r} = \alpha\vec{v} + \beta\vec{w}$ is just a line; or should I say, the span of the multiple vectors won't reduce if we remove one of the vector.

Linearly dependent: when you have multiple vectors and you could remove one without reducing the span, then we call these multiple vectors 'linearly dependent'.

Linearly independent: when you have multiple vectors and if you remove one, the span will reduce, then we call these multiple vectors 'linearly independent'.

So, having a deep understanding of linear combination and linearly independent vectors, now it's the time for us to define what basis vectors are.

The basis of a vector space is a set of linearly independent vectors that span the full space.

In above-mentioned case, in determinant, we'll say the outcome is 'rank 1' means that the span is one-dimensional(a line).

Rank n means the span is n-dimensional.

Rank is useful in determine whether det(A) = 0 or not.

If the number of rank deceases after translation, det(A) = 0.

If det(A) = 0, then a determinant won't have its inverse determinant.

But we won't focus on determinant now, it's enough to just know what it means when you see them as they have some relationships to some extent.

Introduction of Matrix

Now we're going to the most fundamental and important part of this whole book.

What's the relationship between matrix and linear transformation?

Linear transformation, to be honest, it's just another way to express 'linear function'.

The word 'transformation' is just to suggest you to think the process using movement.

We shall define the linear transformation(function) first:

If a transformation(function) satisfies following rules, then we can simply consider them as 'linear transformation'('linear function'):

(1) Lines remain lines without becoming curves.

(2) The origin must be fix.

[Or just remember: keeping grid lines parallel and evenly spaced]

So, if we want to know the result of a vector after transformation, what should we do?

Recall that:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = x\vec{i}+y\vec{j}+z\vec{k} \qquad \text{for } \vec{i}=\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix};\ \vec{j}=\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix};\ \vec{k}=\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

What we should do is to:

(1) Consider a vector as the sum of several basis vectors(aforementioned $\vec{i}/\vec{j}/\vec{k}$).

(2) Consider the result as the sum of several changed basis vectors which have been transformed from basis vectors(aforementioned $\vec{i}/\vec{j}/\vec{k}$) during the transformation.

(3) Turn the transformation of the vector as the transformation of several basis vectors.

Once we know what exactly the changed basis vectors are, we could express the result as:

$$L\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = x\vec{i}'+y\vec{j}'+z\vec{k}' \qquad \text{for } \vec{i}'=A\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right);\ \vec{j}'=B\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right);\ \vec{k}'=C\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right)$$

where L/A/B/C are all functions(the process of transformation).

Understandably, there's no need for us to know what L exactly is as long as we get what A/B/C exactly are.

In order to express A/B/C, we put them in a matrix.

In a matrix, the first column means where i' lands, the second column means where j' lands etc.

We now can regard this matrix as aforementioned L: after all, at the time we get this matrix and a vector, we'll know what the result is.

This is known as matrix-vector multiplication.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = x\begin{bmatrix} a \\ c \end{bmatrix} + y\begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} ax+by \\ cx+dy \end{bmatrix}$$

ATTENTION! You should calculate them from right to left.

Remember, every time you see a matrix, you can interpret it as certain transformation of space.

We now can express most of the transformation using matrix, right?

I mean, what we need to do is to just record where the result basis vectors land.

But just image, if we want to transform a vector twice, what the result is?

Or in another word, if we are given a matrix multiplied by a matrix multiplied by a vector, what will it be?

This kind of new linear transformation is commonly called the composition of the two separated transformation we applied.

Just like all linear transformation, it can be described with the matrix all of its own by following the transformation of these basis vectors i/j/k.

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}_{\substack{\text{Shear} \\ \text{剪切矩阵}}}\left(\underbrace{\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}}_{\substack{\text{Rotation} \\ \text{旋转矩阵}}}\begin{bmatrix} x \\ y \end{bmatrix}\right) = \underbrace{\begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}}_{\substack{\text{Composition} \\ \text{复合矩阵}}}\begin{bmatrix} x \\ y \end{bmatrix} \qquad \underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}_{\substack{\text{Shear} \\ \text{剪切矩阵}}}\underbrace{\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}}_{\substack{\text{Rotation} \\ \text{旋转矩阵}}} = \underbrace{\begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}}_{\substack{\text{Composition} \\ \text{复合矩阵}}}$$

Given that several matrices $M_1/M_2/.../M_n$ and M_{result}:

If $$M_n ... M_2 M_1 \begin{bmatrix} x \\ y \\ z \end{bmatrix} = M_{result}\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Then it means that the overall effect of $M_n ... M_2 M_1$ is just M_{result}:

So $$M_n ... M_2 M_1 = M_{result}$$

Just remember: multiplying several matrices has geometric meaning of applying one transformation to another.

ATTENTION! The order of matrices is really important.

In fact, generally, $M_2 M_1 \neq M_1 M_2$.

Speaking of this, what about $(M_3 M_2) M_1$ and $M_3 (M_2 M_1)$?

Well, $(M_3 M_2) M_1 = M_3 (M_2 M_1)$ without doubt, for both of them we calculate the matrices from right to left.

Now we shall calculate the result of multiplication of matrices.

Just like what I above-mentioned, we calculate the result by following the transformation of these basis vectors i/j/k.

Given two matrices: $M_1=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $M_2=\begin{bmatrix} e & f \\ g & h \end{bmatrix}$

So $M_2M_1=\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

we shall find where the result basis vectors lie:

Where i lies=$\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} a \\ c \end{bmatrix}=\begin{bmatrix} ae+cf \\ ag+ch \end{bmatrix}$

Where j lies=$\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} b \\ d \end{bmatrix}=\begin{bmatrix} be+df \\ bg+dh \end{bmatrix}$

$$\overbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}^{M_2}\overbrace{\begin{bmatrix} e & f \\ g & h \end{bmatrix}}^{M_1}=\begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$$

So we finally get:

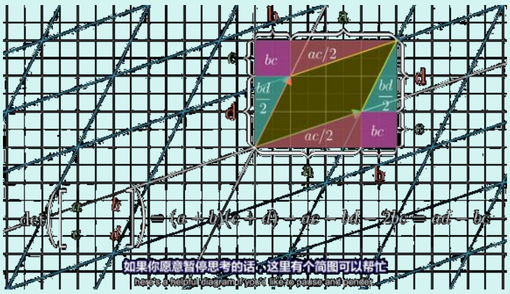
$M_2M_1=\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}=\begin{bmatrix} ae+cf & be+df \\ ag+ch & bg+dh \end{bmatrix}$

Introduction of Determinant

During the transformation, we find that sometimes the space is squeezed in and sometimes the space is stretched out.

How much exactly are things being stretched?

To solve this, we just need to find how much is area/volume/... enclosed by basis vectors scaled.



So if a matrix(transformation) is $\begin{bmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{bmatrix}$**,**

compared to original area enclosed by basis vectors i/j, its area is stretched out by factor (ad-bc).

And that's how we define determinant:

The factor by which a linear transformation changes the area/volume/... is called determinant of that matrix(transformation).

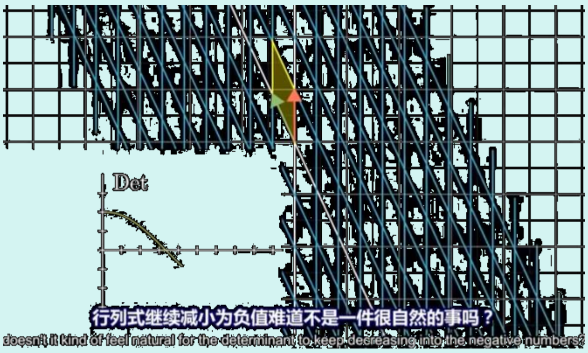
$$\det\left(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\right) = a \det\left(\begin{bmatrix} e & f \\ h & i \end{bmatrix}\right) - b \det\left(\begin{bmatrix} d & f \\ g & i \end{bmatrix}\right) + c \det\left(\begin{bmatrix} d & e \\ g & h \end{bmatrix}\right)$$

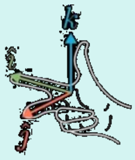
So $\det\left(\begin{bmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{bmatrix}\right)$**=ad-bc**

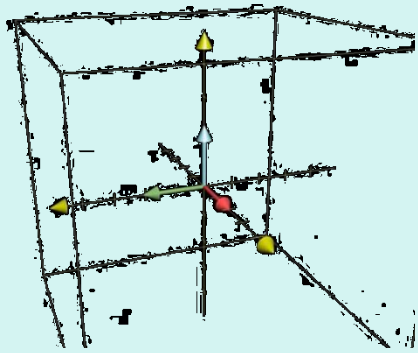
If the result is positive, the result basis vectors satisfies right-hand rule.

If the result is negative, the transformation invert the orientation of the space, the result basis vectors doesn't satisfies right-hand rule.

If the result is 0, the transformation squishes everything into a smaller dimension.







Just remember: the determinant of a 2×2 matrix(transformation) is the factor by which a linear transformation changes the area. Aiming to solve how much exactly are things being stretched, we just need to focus on the changed of the area enclosed by basis vectors.

So, what does the determination of $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ represent?

As we know the meaning of the determination of a 2×2 matrix. Understandably, the determination of a 3×3 matrix shall be represent the factor by which a linear transformation changes the volume.

Recall:

$$\det\left(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\right)=a\begin{bmatrix} e & f \\ h & i \end{bmatrix}-b\begin{bmatrix} d & f \\ g & i \end{bmatrix}+c\begin{bmatrix} d & e \\ g & h \end{bmatrix}$$

$$\text{where } V=a\begin{bmatrix} e & f \\ h & i \end{bmatrix}=b\begin{bmatrix} d & f \\ g & i \end{bmatrix}=c\begin{bmatrix} d & e \\ g & h \end{bmatrix}$$

We could get the above-mentioned relation by Schmidt orthogonalization.

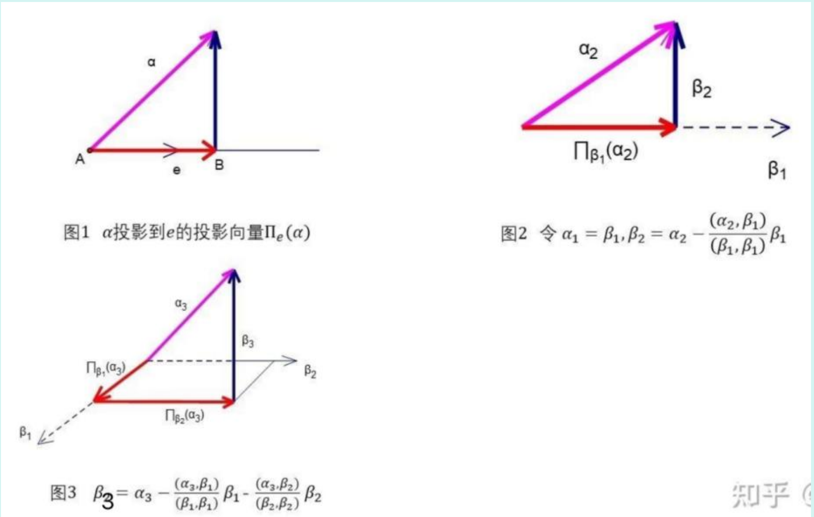


图1 α投影到e的投影向量$\Pi_e(\alpha)$

图2 令 $\alpha_1=\beta_1, \beta_2=\alpha_2-\frac{(\alpha_2,\beta_1)}{(\beta_1,\beta_1)}\beta_1$

图3 $\beta_3=\alpha_3-\frac{(\alpha_3,\beta_1)}{(\beta_1,\beta_1)}\beta_1-\frac{(\alpha_3,\beta_2)}{(\beta_2,\beta_2)}\beta_2$

Putting everything into a nutshell:

the determinant of a matrix(transformation) is the factor by which a linear transformation changes the area/volume/....

Aiming to solve how much exactly are things being stretched, we just need to focus on the changed of the area/volume/... enclosed by basis vectors.

So now, we can deduce that:

det($\mathbf{M_1M_2}$)=det($\mathbf{M_1}$) · det($\mathbf{M_2}$)

It's easy to image that:

det($\mathbf{M_1M_2}$) means the factor by which we stretch the area at once.

det($\mathbf{M_1}$) · det($\mathbf{M_2}$) means that we firstly stretch the area thought transformation $\mathbf{M_1}$, then stretch this area through transformation $\mathbf{M_2}$, finally we find out how much area are exactly scaled.

Solving Linear System Using Matrix

Given a system of equations, which consist of a list of variables and a list of equations related to them, they are in form of 'ax+by+cz+...=n'

Within each equation, the only thing happening to each variable is that it's scaled by some constant, and the only thing happening to each of those scaled variables is that they're added to each other.

We throw all the variables on the left, and any lingering constant on the right.

Vertically line up all the same variable and use 0 as coefficient if necessary.

Now, we get a linear system of equations.

This really looks like a matrix-vector application, right?

$$\begin{aligned}2x+5y+3z&=-3\\4x+0y+8z&=0\\1x+3y+0z&=2\end{aligned} \;\rightarrow\; \overbrace{\begin{bmatrix}2&5&3\\4&0&8\\1&3&0\end{bmatrix}}^{A}\overbrace{\begin{bmatrix}x\\y\\z\end{bmatrix}}^{\vec{x}}=\overbrace{\begin{bmatrix}-3\\0\\2\end{bmatrix}}^{\vec{v}}$$

That should be said, we can turn this linear system of equations into a matrix-vector application. And as long as we solve what the vector is, we'll get the variables.

As $A\vec{x}=\vec{v}$

So $A^{-1}A\vec{x}=A^{-1}\vec{v}$ **where** A^{-1} **is the inverse transformation of A**

As $A^{-1}A\vec{x}$ **means** $\vec{x}$ **returns to its original position:** $A^{-1}A=\begin{bmatrix}1&0\\0&1\end{bmatrix}$

where $\begin{bmatrix}1&0\\0&1\end{bmatrix}$ **is the identity transformation,I,transformation of doing nothing**

So $\vec{x}=A^{-1}\vec{v}$

But we shall deliberate this conversion for a moment:

Does equation $\vec{x}=A^{-1}\vec{v}$ always satisfy?

What if $A\vec{x}$ **squishes** $\vec{x}$ **into a smaller dimension?**

In this case, if we want find A^{-1}**, which means we want a point on line(one input) to be several points in space(several output).**

It's no a function as a function has one output for each input. That should be said, if $A\vec{x}$ **squishes** $\vec{x}$ **into a smaller dimension, or det(A)=0, there's no** A^{-1}**.**

Recall:

Rank n means the span is n-dimensional.
Rank is useful in determine whether det(A)=0 or not.
If the number of rank deceases after translation, det(A)=0.
If det(A)=0, then a determinant won't have its inverse determinant.

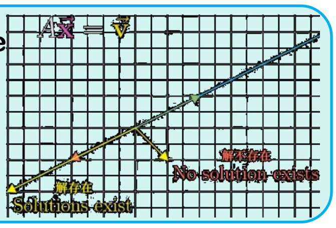
See? Now you know what they mean.

Here's something new:

Set of all possible outputs $A\vec{x}$ is called column space of A.
If the number of rank is equal to the number of column space,
we say that the transformation is full rank.
No matter what transformation is, the origin is always in the column space.
The set of vectors that land on origin after transformation is called the null space or the kernel of the matrix.

To be honest, there terminologies are too theoretical, just knowing what it is is enough.

But maybe you just notice that: when sometimes det(A)=0, if we don't focus on matrix but the linear system of equation, we may find the result.



ATTENTION! det(A)=0 only means there's no $\mathbf{A}^{-1}$.

So now,here's the question: how do we find A^{-1}?

We shall use Gaussian elimination to find $\mathbf{A}^{-1}$.

As $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\begin{bmatrix} A & B & C \\ D & E & F \\ G & H & I \end{bmatrix}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **where** $\begin{bmatrix} A & B & C \\ D & E & F \\ G & H & I \end{bmatrix}$**(A) is given**

As we have already known $\mathbf{A}$, and we want to find $\mathbf{A}^{-1}$:
We firstly put A and I into a common space.
[As there we'll use I twice for different purposes, we'll use $\mathbf{I}_1$ and $\mathbf{I}_2$ for different purposes]
At the time we try to turn A into $\mathbf{I}_1$, the basis vectors $\mathbf{I}_2$ will change, right?
During this time, the transformation is $\mathbf{A}^{-1}$, so the basis vectors $\mathbf{I}_2$ will turn into $\mathbf{A}^{-1}$.
That should been said: if we can transform A and $\mathbf{I}_2$ at the same time, the result of $\mathbf{I}_2$ will be $\mathbf{A}^{-1}$ --- the answer we want!
Now we use a specific form to link them:

A B C \ 1 0 0
[D E F \ 0 1 0]
G H I \ 0 0 1

Eventually, we'd like this thing turning into:

1 0 0 \ a b c
[0 1 0 \ d e f]
0 0 1 \ g h i

So, how to turn the first form into the second one?

a b c A B C 1 0 0
Now look to the original equation: [d e f][D E F]=[0 1 0]
g h i G H I 0 0 1

If we just focus on each basis vector:

a b c A 1
[d e f][D]=[0]
g h i G 0

a b c B 0
[d e f][E]=[1]
g h i H 0

a b c C 0
[d e f][F]=[0]
g h i I 1

As they're similar, we'll just focus on the first one.

aA+bD+cG=1 or (1)=1

dA+eD+fG=0 or (2)=0

gA+hD+iG=0 or (3)=0

So they must satisfy: $n\cdot(1) = n\cdot 1$ $(2)-n\cdot(3)=0-n\cdot 0$ etc.

A B C \ 1 0 0 (1) \ 1 0 0
Similarly, if: [D E F \ 0 1 0] is [(2) \ 0 1 0]
G H I \ 0 0 1 (3) \ 0 0 1

So it must satisfy:

(1) \ 1 0 0 (1) \ 1 0 0
[(2) \ 0 1 0] [(2) + n·(1) \ 0 + n·1 1 + n·0 0+
n·(3) \ n·0 n·0 n·1 (3) \ 0 0 1
n·0] etc.

Generally, we start this kind of translation from the left lower corner to the right higher one, as during this process, we'll get more and more 0 in our left hand and thus the process of calculation will be easier.

You may ask: what about nonsquare matrices?

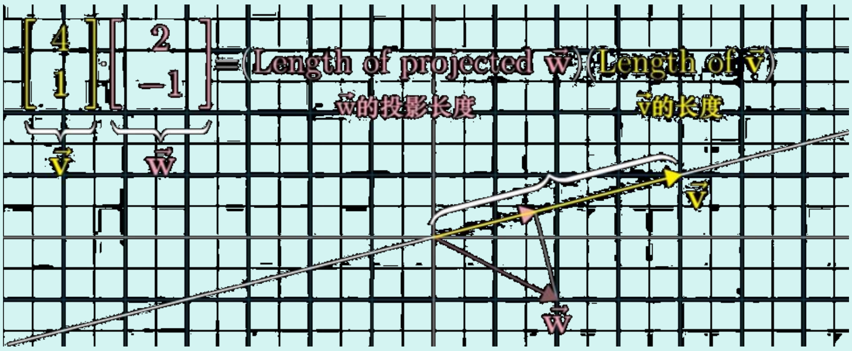
Firstly, we'll know that it's accept for us to transform one vector to another different-dimensional vector.
The number of columns is the number of dimension of the original vector, and the number of the roles is the number of dimension of the result vector.

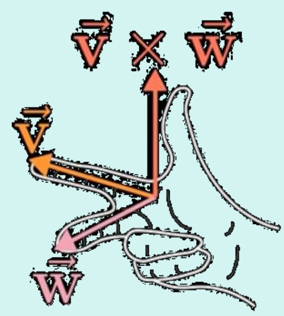
ATTENTION! This part is really important if you want to understand the real meaning of dot products and cross products.

Dot Product and Cross Product

If we want to understand dot products and cross products, then the first thing we need to do is to know why we need them.

The dot product of $\vec{v}$ and $\vec{w}$ is to calculate the product between the length of one of the vectors($\vec{v}$ e.g.) and the length of the line that another vector($\vec{w}$ there) projecting onto the line that passes through the origin and the tip of the first vector($\vec{v}$ there).





The cross product of two three-dimensional vectors is that: if there's another three-dimensional vector and we want to find the volume enclosed by these three vectors, we can get the result by just finding out the dot product of the new vector and the result of cross product of these two three-dimensional vectors.
That should be said, the cross product of two three-dimensional vectors is to find a vector which is perpendicular to these two vectors and the length of the new vector is equal in magnitude to the area enclosed by these two vectors. Of course, the new vector obey the right-hand rule: if the length is positive, then the order of three vectors is just like the order of our right-hand fingers; if it's negative, just like our left-hand ones.

First,we'll focus on dot product.

Given $\begin{bmatrix}a\\b\end{bmatrix}$ and $\begin{bmatrix}c\\d\end{bmatrix}$, where the length of $\begin{bmatrix}a\\b\end{bmatrix}$ is 1(unit vector):

$\begin{bmatrix}a\\b\end{bmatrix} \cdot \begin{bmatrix}c\\d\end{bmatrix}$=(1×)the length of the line that $\begin{bmatrix}c\\d\end{bmatrix}$ projecting

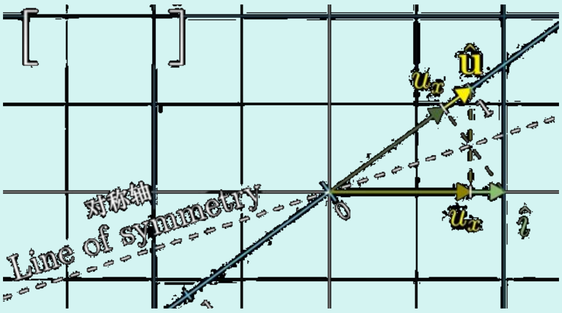
onto the line that passes through the origin and the tip of the $\begin{bmatrix}a\\b\end{bmatrix}$

So we just need to turn $\begin{bmatrix}c\\d\end{bmatrix}$ into one-dimensional line that passes through the

origin and the tip of the $\begin{bmatrix}a\\b\end{bmatrix}$.

During this time, there's a linear transformation from multiplied dimensions to

one dimension(number line).
It's a linear transformation, which means that if we have a line of evenly spaced dots and the apply this transformation, those dots will keep evenly spaced in the number line.
So we just need to find the exact projecting transformation---or should I say, the positions of basis vectors lie after projecting transformation---in these case, we'll get the true projecting transformation.



It's obvious that each vector(column) will be a number.
Then we shall find out the exact numbers: Just like right-hand graph. If we make an angular bisector between a straight line which passes through $\begin{bmatrix} a \\ b \end{bmatrix}$ and origin, and the x-axis. Then we'll find that projecting i onto the line is symmetric to project $\begin{bmatrix} a \\ b \end{bmatrix}$ onto x-axis, as the angular bisector is just the line of symmetry(try using ASA).

So the length after i projecting onto the $\begin{bmatrix} a \\ b \end{bmatrix}$ is just equal to the length after $\begin{bmatrix} a \\ b \end{bmatrix}$ projecting onto i. In that case, it's just a.

The same, the length after j projecting onto $\begin{bmatrix} a \\ b \end{bmatrix}$ is b.

So $\begin{bmatrix} a \\ b \end{bmatrix} \cdot \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix}$ where $\begin{bmatrix} a \\ b \end{bmatrix}$ is an unit vector

But what if $\begin{bmatrix} a \\ b \end{bmatrix}$ isn't an unit vector?

If $\begin{bmatrix} A \\ B \end{bmatrix} = k \begin{bmatrix} a \\ b \end{bmatrix}$, where k is a constant:

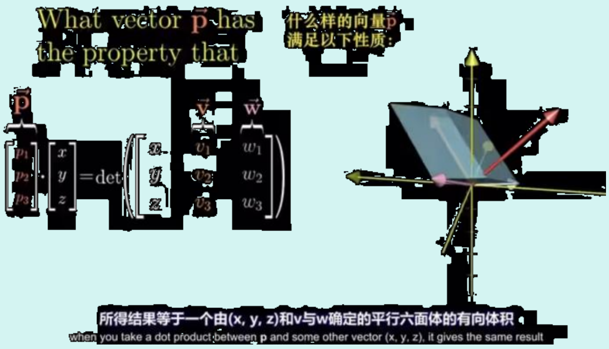
$$\begin{bmatrix} A \\ B \end{bmatrix} \cdot \begin{bmatrix} c \\ d \end{bmatrix} = k \begin{bmatrix} a \\ b \end{bmatrix} \cdot \begin{bmatrix} c \\ d \end{bmatrix} = k \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} \quad \text{as } A=ka,\ B=kb$$

So in conclusion:

$$\begin{bmatrix} a \\ b \end{bmatrix} \cdot \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = ac+bd \quad \text{for all } \begin{bmatrix} a \\ b \end{bmatrix} \text{ and } \begin{bmatrix} c \\ d \end{bmatrix}$$

Remember: anytime we have a linear transformation whose output is the number line, no matter how it defines, there's going to be an unique vector corresponding to that transformation. In the sense that applying transformation is the same thing as taking a dot product with this vector.
Remember: applying transformation is the same thing as taking a dot product with this vector!!!

Now look at the cross product.



Recall: if there's another three-dimensional vector and we want to find the volume enclosed by these three vectors, we can get the result by just finding out the dot product of the new vector and the result of cross product of these two three-dimensional vectors.

So $\begin{bmatrix}A\\B\\C\end{bmatrix}\cdot\begin{bmatrix}x\\y\\z\end{bmatrix}=\det\left(\begin{bmatrix}x&a&d\\y&b&e\\z&c&f\end{bmatrix}\right)$ where $\begin{bmatrix}x\\y\\z\end{bmatrix}$ is a vector, $\begin{bmatrix}a\\b\\c\end{bmatrix}\times\begin{bmatrix}d\\e\\f\end{bmatrix}=\begin{bmatrix}A\\B\\C\end{bmatrix}$

As $\det\left(\begin{bmatrix}x&a&d\\y&b&e\\z&c&f\end{bmatrix}\right) = x(bf-ce)+y(cd-af)+z(ae-bd)$

So $\begin{bmatrix}A\\B\\C\end{bmatrix}=\begin{bmatrix}bf-ce\\cd-af\\ae-bd\end{bmatrix}$

In order to get the answer easily, just remember:

$\begin{bmatrix}a\\b\\c\end{bmatrix}\times\begin{bmatrix}d\\e\\f\end{bmatrix}=\begin{bmatrix}i&a&d\\j&b&e\\k&c&f\end{bmatrix}$ where i/j/k are basis vectors

Introduction of Change in Basis

The vector that represents a point in a space not only depends on its position, but also the basis vectors.
That should been said: although the space is certain, the vector expressing a point depends on the system we used.

Understandably, if the basis vectors are different, chances are that we'll use different vectors to represent the same point.

Only if the point is at the origin, the vectors always the same, $\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

So, given that a vector$\begin{bmatrix} X \\ Y \\ \vdots \\ N \end{bmatrix}$, based on system(let call it 'system C') whose the basis

vectors are $\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix} \dots \begin{bmatrix} N_1 \\ N_2 \\ \vdots \\ N_n \end{bmatrix}$, how do we represent a point, according to the system

whose basis vectors are $\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \dots \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$(let call it 'system N')?

[ATTENTION!

Those basis vectors $\begin{bmatrix} \mathbf{A_1} \\ \mathbf{A_2} \\ \vdots \\ \mathbf{A_n} \end{bmatrix} \begin{bmatrix} \mathbf{B_1} \\ \mathbf{B_2} \\ \vdots \\ \mathbf{B_n} \end{bmatrix} \dots \begin{bmatrix} \mathbf{N_1} \\ \mathbf{N_2} \\ \vdots \\ \mathbf{N_n} \end{bmatrix}$, according to their own system C, are so-

called ' $\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \\ \vdots \\ \mathbf{0} \end{bmatrix} \dots \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{1} \end{bmatrix}$ '.

Hence, we could use x$\begin{bmatrix} \mathbf{A_1} \\ \mathbf{A_2} \\ \vdots \\ \mathbf{A_n} \end{bmatrix}$+y $\begin{bmatrix} \mathbf{B_1} \\ \mathbf{B_2} \\ \vdots \\ \mathbf{B_n} \end{bmatrix}$+...+n$\begin{bmatrix} \mathbf{N_1} \\ \mathbf{N_2} \\ \vdots \\ \mathbf{N_n} \end{bmatrix}$ to find the vector of the point in the system N.
Since we firstly merely write down a linear equation; then, just express the basis vectors of system C in the way of corresponding vectors in system N]
To some extent, it is just a transformation that we have already discussed before, right?

$\begin{bmatrix} X \\ Y \\ \vdots \\ N \end{bmatrix}$ based on the system C

$$= X\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix} + Y\begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix} + \ldots + N\begin{bmatrix} N_1 \\ N_2 \\ \vdots \\ N_n \end{bmatrix} = \begin{bmatrix} x \\ y \\ \vdots \\ n \end{bmatrix}$$ according to the system N

Honestly, as we can see, it's just a matrix-vector multiplication.

So: $$\begin{bmatrix} A_1 & B_1 & \cdots & N_1 \\ A_2 & B_2 & \cdots & N_2 \\ \vdots & \vdots & \ddots & \vdots \\ A_n & B_n & \cdots & N_n \end{bmatrix}\begin{bmatrix} X \\ Y \\ \vdots \\ N \end{bmatrix} = \begin{bmatrix} x \\ y \\ \vdots \\ n \end{bmatrix}$$

[In below example, $\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$]

You may find this matrix-vector multiplication quiet interesting:

If you take the transformation into consideration, then you'll find that you are transforming your 'real' space(N) into another space(C).

Regarding of the expression of a vector, we are just trying to explain a vector based on the system C in the way of system N.

Introduction of Diagonalisation

So, what if I want to explain a vector based on system N in the way of system C?

Think about the meaning of inverse matrix as well as its application, you'll soon find out the answer, and then have a deeper understanding of the change of basis.

Now, consider we want to make a transformation in the system C, given the corresponding matrix-multiple expression in the system N.

Supposed the basis vectors in system C are $\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix} \ldots \begin{bmatrix} N_1 \\ N_2 \\ \vdots \\ N_n \end{bmatrix}$.

Subsequently, we call $\begin{bmatrix} A_1 & B_1 & \cdots & N_1 \\ A_2 & B_2 & \cdots & N_2 \\ \vdots & \vdots & \ddots & \vdots \\ A_n & B_n & \cdots & N_n \end{bmatrix}$ as 'T'.

The corresponding matrix-multiple(for instance, M) expression in the system N is M.

And there is a vector in the system C, named $\vec{v}$.

Then, we shall begin.

You may think that we could simply express this transformation as: M$\vec{v}$.

But attention! The transformation matrix M is based on the system N instead of C.

In fact, most of the time, the transformation matrix M has different forms, depending on which system you choose.

[In above example, you rotate a vector based on system C. From left to right:T^{-1}MT$\vec{v}$]

Hey! Here's a way to deliberate!

We first express $\vec{v}$ in systems N: $\mathbf{T}\vec{\mathbf{v}}$

Then, apply the transformation in the system N: **MT**$\vec{\mathbf{v}}$

Finally, since we want a result based on system C, we transit the form: $\mathbf{T}^{-1}$**MT**$\vec{\mathbf{v}}$

During the first step, although we 'interpret' $\vec{\mathbf{v}}$, we don't change the point which on the space expresses the $\vec{\mathbf{v}}$ before the transformation.

And for the next two steps, we don't change the point which on the space expresses the $\vec{\mathbf{v}}$ before the transformation.

Remember:

An expression like $\mathbf{T}^{-1}$MT suggests a mathematical sort of empathy,where T is a process to transform one system A to another system B, and M is the transformation matrix expressed in system B.

This process is called diagonalisable.

The middle matrix represents a transformation, the the other two explain a change in basis vectors.

You may ask me, why we don’t just directly find the transformation matrix based on the system C?

The answer is: you, most of the time, won’t know it!

Just like the above-mentioned example: do you know the matrix repressing a rotation?

Another factor is that: solving the transformation in this way might be easily.

Why? That’s the next part.

Introduction of Eigenvalue and Eigenvector

Just image: if we want to apply a single transformation to a vector for multiple times, this process will definitely be difficult unless the matrix is $\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$.

Why it is easy for use to apply $\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$ to a vector for multiple times?

Cuz: $\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}^n = \begin{bmatrix} a_{11}^n & 0 & \cdots & 0 \\ 0 & a_{22}^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn}^n \end{bmatrix}$

This kind of matrix is called: diagonal matrix.

So, let simplify this question:

Interpret the transformation matrix to the diagonal matrix, then apply it.

↓

Change your system N into a suitable system C to interpret the transformation matrix to the diagonal matrix.

↓

Find the suitable basis vectors to ‘built’ a suitable system C.

↓

Those suitable basis vectors you find must only be stretched along their original lines during the transformation.

↓

Find out, in any cases, during the transformation, which lines don’t deviate their original paths.

Now, let’s go.

Firstly, we want to know, given a transformation matrix, which lines don’t deviate their original paths.

There are two ways to solve this question: invariant line and eigenvalue & eigenvector.

For invariant line:

We could express a vector as $\begin{bmatrix} m \\ km \end{bmatrix}$**, and after transformation** $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$**, if the vector doesn’t deviate its original path, it will be** $\begin{bmatrix} M \\ kM \end{bmatrix}$**.**

Then: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} m \\ km \end{bmatrix}=\begin{bmatrix} M \\ kM \end{bmatrix}$

$$am + bkm = M$$

$$cm + dkm = kM$$

$$\frac{am + bkm}{cm + dkm} = \frac{M}{kM}$$

$$\frac{a + bk}{c + dk} = \frac{1}{k}$$

And then, we may find the values of k(the number of roots depends on Δ), where k is the gradient of the line which doesn't deviate its original paths after transformation.

These lines y=kx are called invariant lines.

Whereas, as you can find, you can use this measure only when the vector is in a two-dimensional space.

The other measure, could efficiently tackle with this problem.

Given a transformation M $\begin{bmatrix} a_1 & b_1 & \cdots & n_1 \\ a_2 & b_2 & \cdots & n_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_n & b_n & \cdots & n_n \end{bmatrix}$ and the vector $\vec{v}$:

$M\vec{v} = k\vec{v}$ since vector must only be stretched along its original lines

Then: $M\vec{v} = kI\vec{v}$

$$M(kI)\vec{v} = \vec{0}$$

We don't want $\vec{v} = \vec{0}$ cuz at that time, the equation is always satisfied. Instead, we want to find a general solution of k. So we shall let $\vec{v} \neq \vec{0}$.

That should be said: det(M-kI) = 0

You may ask: why not (M-kI) = $\vec{0}$?

Cuz if this case is only one of many situation.

As long as we want (M-kI)$\vec{v} = \vec{0}$, we can not only let (M-kI) = $\vec{0}$, but also squash $\vec{v}$ into a lower-dimensional space(det(M-kI) = 0).

Just think this statement for a little bit: what's the volume of a two-dimensional square?

And when (M-kI) = $\vec{0}$, det(M-kI) = 0. Hence, we just need to let det(M-kI) = 0

Get it? So now we'll just find out the answer.

$$\det\left(\begin{bmatrix} a_1 - k & b_1 & \cdots & n_1 \\ a_2 & b_2 - k & \cdots & n_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_n & b_n & \cdots & n_n - k \end{bmatrix}\right) = 0$$

In this way, we sometimes could find n different k, and for each k, you can find the corresponding $\vec{v}$.

Here, k is called eigenvalue and $\vec{v}$ is known as eigenvector.

ATTENTION! if you fail to find n different k(some roots are the same), then you could never change the system N to system C, since the system C must be the same-dimensional as system N is.

As one matrix may have plenty of eigenvectors and eigenvalues, we use α_i(i=1,2,..,n) to represent each eigenvalue, and use e_i to represent corresponding eigenvector.

Here are some brilliant derivations about eigenvalue and eigenvector.

I won't explain them deeply cuz I want you to 'built' your own mind systems or 'draw' your mind map to try to figure out the truth.

Given matrix A has eigenvalue α_i and eigenvector e_i, matrix B satisfies B = mA + nI.

Then matrix B has eigenvalue mα_i+n with eigenvector e_i.

Since Be_i = mAe_i + nIe_i , the direction of the vectors in both sides of equation are the same(the vector doesn't change the direction, take some time to thing about it).

Given matrix A has eigenvalue α_i and eigenvector e_i, matrix B has eigenvalue β_i and eigenvector e_i.

Then: $(A+B)e_i = (\alpha_i + \beta_i)e_i$

$A^n e_i = \alpha_i^n e_i$

$(AB)e_i = (\alpha_i \beta_i)e_i$

Try to prove them one by one, their order are in logic.

Remember what we said just now?

We now could use these suitable basis vectors(whether using invariant lines or using eigenvector)to 'built' a suitable system C to interpret the transformation.

Recall:

An expression like T^{-1}MT suggests a mathematical sort of empathy,where T is a process to transform one system A to another system B, and M is the transformation matrix expressed in system B.

This process is called diagonalisable.

Hence, if we want to apply a single transformation to a vector for multiple times, we just need to use something like T^{-1}MT to simplify the process.

Introduction of Characteristic Equation

Now, after learning so many things, here are some new ways to find a inverse matrix. Yeah, I know, honestly, Gaussian elimination is enough for most of the situations; whereas, at the time you trying to understand and apply them, you will have a deeper understanding of linear transformation and linear algebra.

During the process of finding eigenvalue, we will inevitably use the characteristic equation:

If $A = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix}$, $B = \begin{bmatrix} b_{11} & 0 & \cdots & 0 \\ 0 & b_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_{nn} \end{bmatrix}$, $T = \begin{bmatrix} T_{11} & T_{21} & \cdots & T_{n1} \\ T_{12} & T_{22} & \cdots & T_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ T_{1n} & T_{2n} & \cdots & T_{nn} \end{bmatrix}$,

as well as $B = T^{-1}AT$, $A = TBT^{-1}$:

$P_A(k) = |A - kI| = 0$ where $|A - kI|$ means det($A - kI$)

So: $|TBT^{-1} - T(kI)T^{-1}| = 0$ since $T(kI)T^{-1} = kI$ (deliberate it!)

$$|T||B - (kI)||T^{-1}| = 0$$

$$|B - (kI)| = 0 = P_B(k) = (b_{11} - k)(b_{22} - k)\ldots(b_{nn} - k)$$

Then: $P_A(A) = (b_{11}I - A)(b_{22}I - A)\ldots(b_{nn}I - A)$

$$P_A(A) = (T(b_{11}I)T^{-1}\text{-}TBT^{-1})(T(b_{22}I)T^{-1}\text{-}TBT^{-1})\ldots(T(b_{nn}I)T^{-1}\text{-}TBT^{-1})$$

$$P_A(A)=T((b_{11}I - B)(b_{22}I - B)\ldots(b_{nn}I - B))T^{-1}$$

$P_A(A) = \vec{0}$ since $(b_{11}I - B)(b_{22}I - B)\ldots(b_{nn}I - B) = \vec{0}$

That should be said:

$$P_A(A) = m_nA^n + m_{n-1}A^{n-1} + \ldots + m_2A^2 + m_1A + m_0I = \vec{0}$$

where m_i is a constant

$$m_nA^n + m_{n-1}A^{n-1} + \ldots + m_2A^2 + m_1A = -m_0I$$

$$m_nA^{n-1} + m_{n-1}A^{n-2} + \ldots + m_2A + m_1 = -m_0A^{-1}$$

$$A^{-1} = \frac{m_nA^{n-1} + m_{n-1}A^{n-2} + \ldots + m_2A + m_1}{-m_0}$$

ATTENTION! A/B/P must have the same-dimensional space! Or they can not be diagonalised!