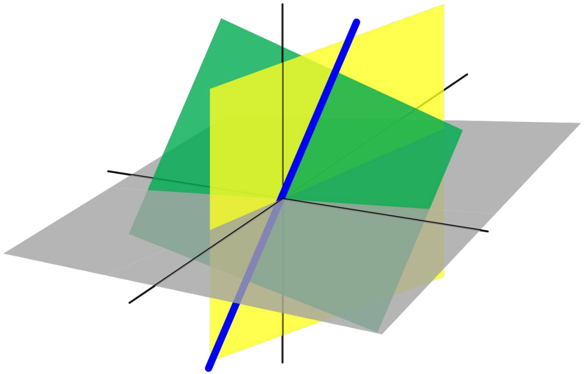
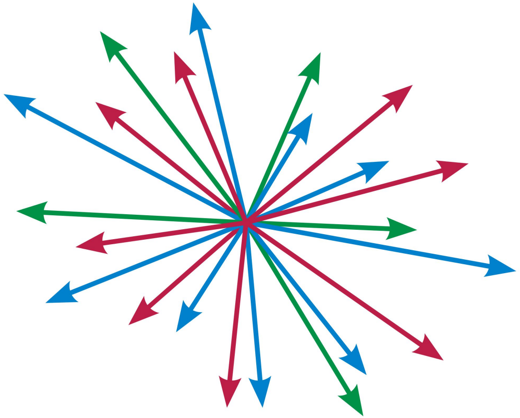
Foundations of Linear Algebra- PCA, SVD with Applications



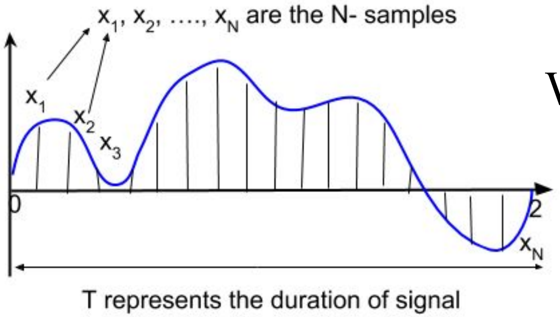
School of AI

Harikrishnan N B
Research Associate
Consciousness Studies Programme
National Institute of Advanced Studies



6th April 2019

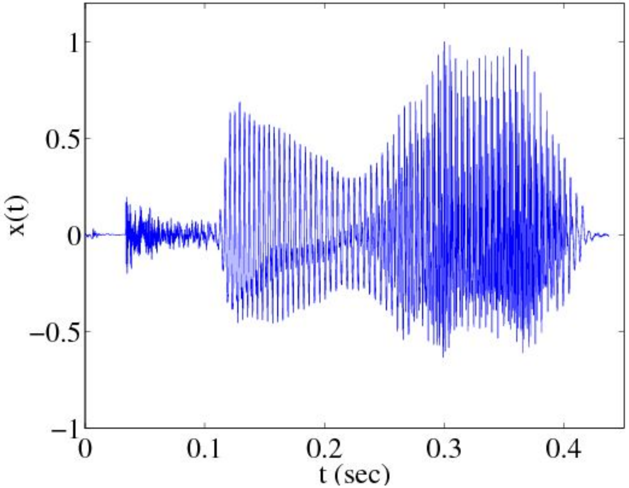
Why Should I Learn Linear Algebra?

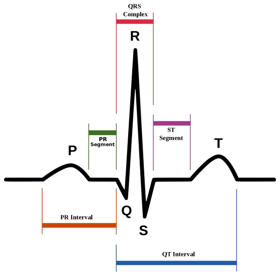
- Most of the Engineering problems boils down to Linear Algebra(LA).
- 

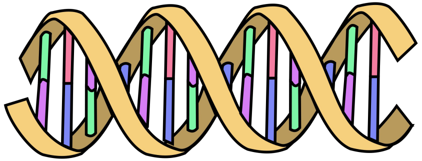
With what all frequency this signal is made up of ?
- LA plays an important role in extracting meaningful information from data. (data science)
- Image processing (jpeg and jpeg 2000)
- Quantum Mechanics and Quantum Computation

Data is the king - AI Community

What is **Data?**

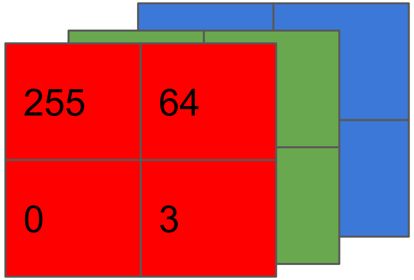






Data is everywhere!

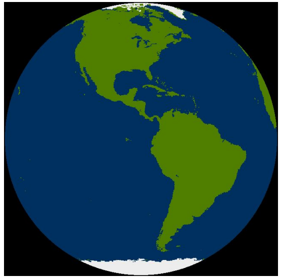
Data for Machine Learning Enthusiasts



(2 x 2 x 3)

[255 64 0 3 - - - - - - - - - -]

1 x 12



Sensors

DATA

Pre-processing

Vectors

Matrices

Tensors

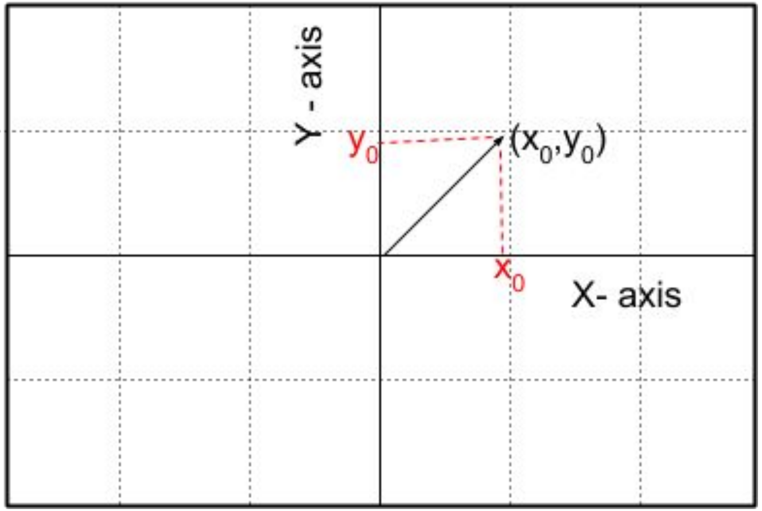
Feature Engineering

Machine Learning (ML)/ Deep Learning (DL)

Harikrishnan N B

Vectors - Different Understanding

Physicists



Computer Science

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

Mathematicians

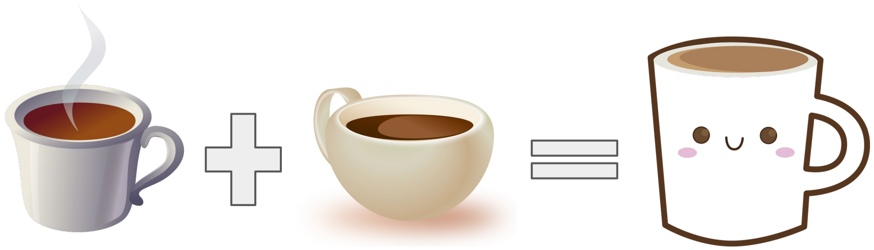
Vector space is a **collection of objects**(it can be anything) called vectors which satisfies mainly two important properties:

1. **closed under vector addition**
2. **closed under scalar multiplication**.

Vector Space - Coffee Space

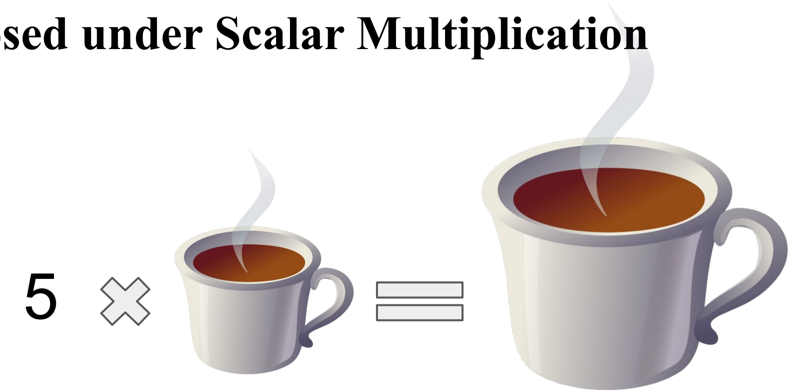
Coffee Space - In Coffee space we have different kinds of coffee with varying strength. Now we will understand the vector space properties with this anecdote.

Closed under Vector Addition



Adding two coffee's will give you another coffee which is in the coffee space

Closed under Scalar Multiplication



Scaling a coffee will give a coffee which is in the coffee space

Harikrishnan N B

Dimension and Basis of a Vector Space

Dimension of a Vector space - Every vector space has a dimension. Dimension is the number of basis vectors required to span the vector space.

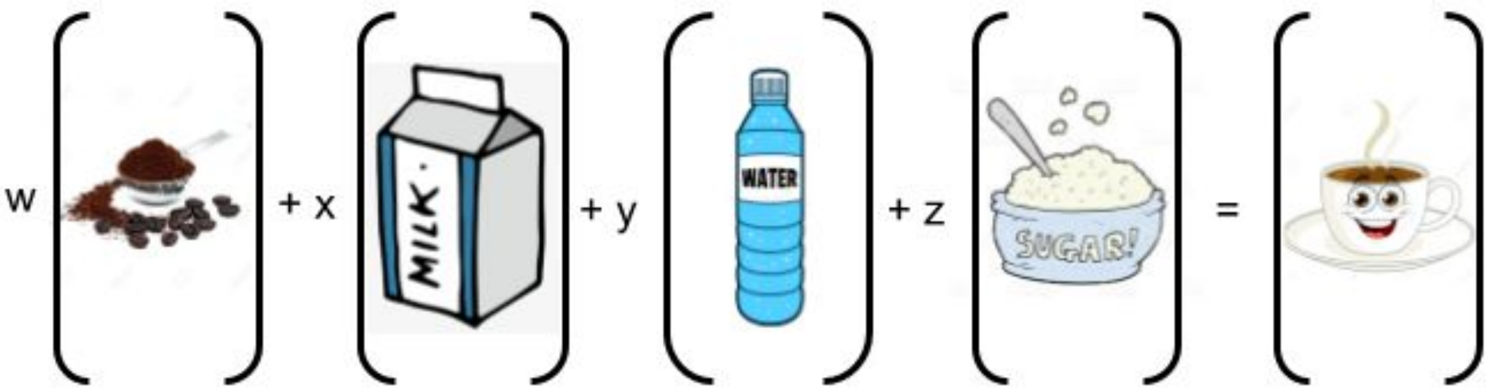
Properties of Basis Vectors -

- Basis vectors has to be linearly independent.
- Basis vectors should span the vector space.

Dimension and Basis of a Coffee Space

- **Linear Independence**
- **Span the space**

Coffee Space- Vector Space



Coffee powder, milk, water and sugar are the basis vectors. Since there are only 4 basis vectors then coffee space has a dimension of 4.

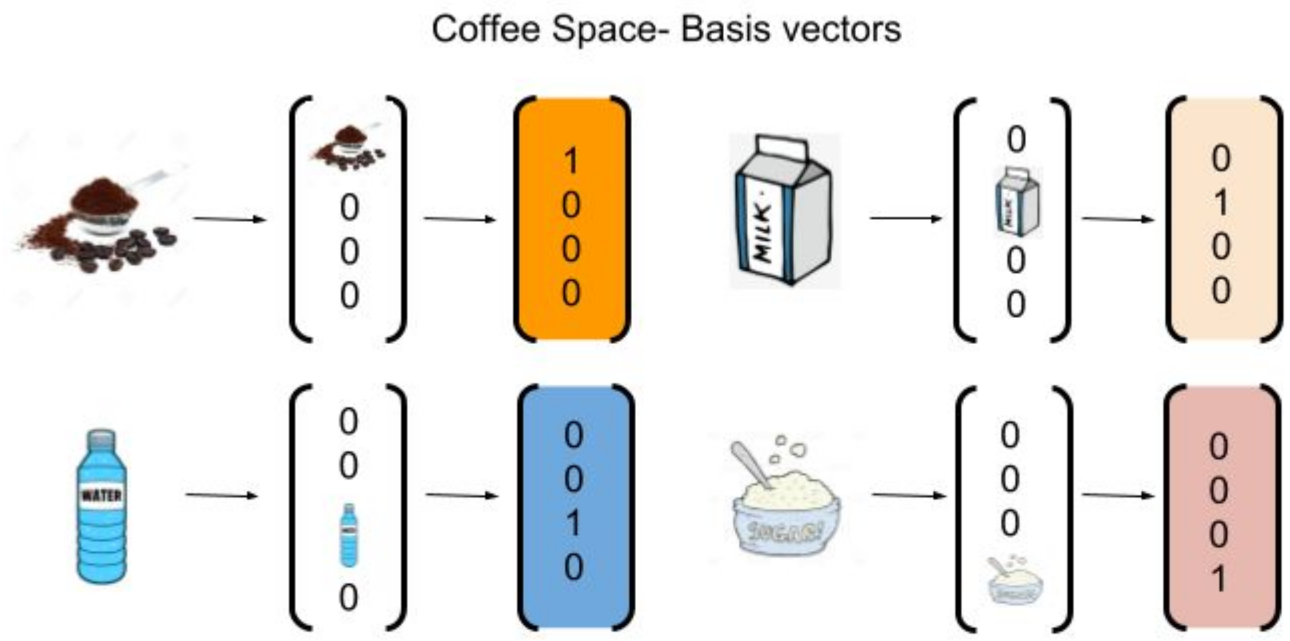
My Friend's Horrible Coffee

My Friend’s Horrible Coffee

My Friend’s Horrible Coffee

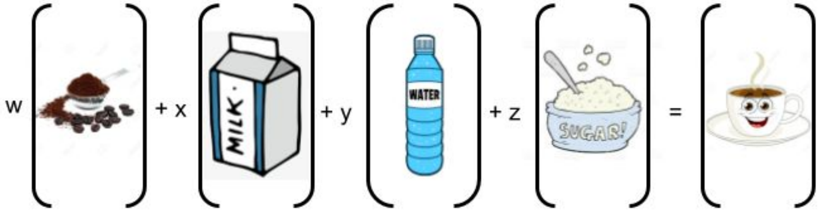
$$2\begin{bmatrix}1\\0\\0\\0\end{bmatrix} + 1\begin{bmatrix}0\\1\\0\\0\end{bmatrix} + 4\begin{bmatrix}0\\0\\1\\0\end{bmatrix} + 3\begin{bmatrix}0\\0\\0\\1\end{bmatrix} = \begin{bmatrix}2\\1\\4\\3\end{bmatrix}$$

Visualizing Coffee Space Basis Vectors

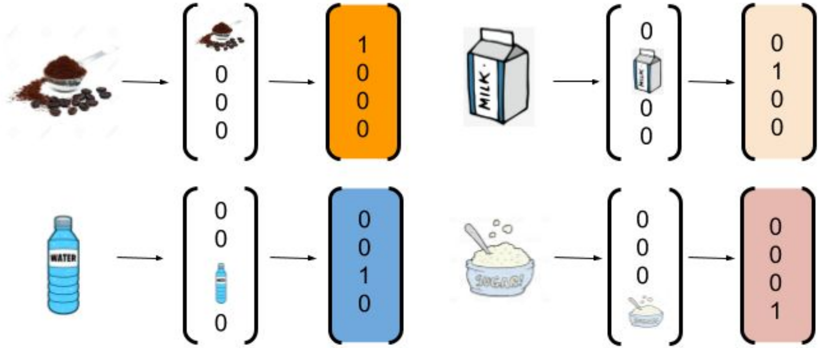


Matrix Multiplication - Visualization

Coffee Space- Vector Space



Coffee Space- Basis vectors



Ax = b

$$w\begin{bmatrix}1\\0\\0\\0\end{bmatrix} + x\begin{bmatrix}0\\1\\0\\0\end{bmatrix} + y\begin{bmatrix}0\\0\\1\\0\end{bmatrix} + z\begin{bmatrix}0\\0\\0\\1\end{bmatrix} = \begin{bmatrix}5\\6\\7\\8\end{bmatrix}$$

Ax = b

$$\begin{bmatrix}1 & 0 & 0 & 0\\0 & 1 & 0 & 0\\0 & 0 & 1 & 0\\0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}w\\x\\y\\z\end{bmatrix} = \begin{bmatrix}5\\6\\7\\8\end{bmatrix}$$

Column Space - Visualization

Ax = b

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \end{bmatrix}$$

$$\Downarrow$$

Ax = b

$$w \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \end{bmatrix}$$

Column Space of Matrix A - Column space of matrix A denoted as C(A) is the space spanned by the column vectors of A.

C(A)

$$span \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Dimension of C(A) = 4. Since 4 linearly independent vectors are there in the columns of matrix A. These vectors act as the basis and span the entire R^4 .

Thinking

C(A)

Why $span\left\{\begin{bmatrix}1\\0\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\\0\end{bmatrix},\begin{bmatrix}0\\0\\1\\0\end{bmatrix},\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right\}$ can represent any point in R^4 ?

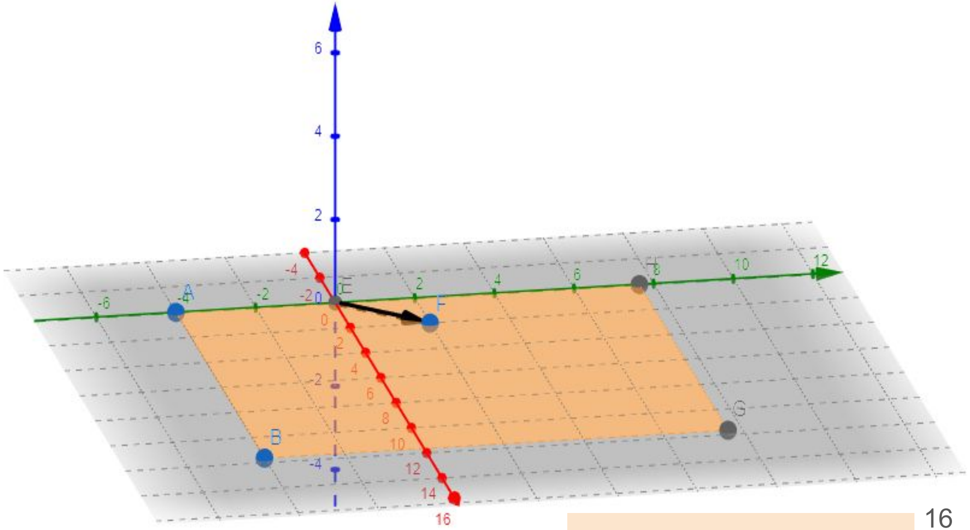
Can you see the Column Space?

$$\begin{bmatrix}1 & 0\\ 0 & 1\\ 0 & 0\end{bmatrix}\begin{pmatrix}w\\ x\end{pmatrix} = \begin{pmatrix}2\\ 2\\ 0\end{pmatrix}$$

$$w\begin{pmatrix}1\\ 0\\ 0\end{pmatrix} + x\begin{pmatrix}0\\ 1\\ 0\end{pmatrix} = \begin{pmatrix}2\\ 2\\ 0\end{pmatrix}$$

C(A)

$$span\left\{\begin{pmatrix}1\\ 0\\ 0\end{pmatrix}, \begin{pmatrix}0\\ 1\\ 0\end{pmatrix}\right\}$$



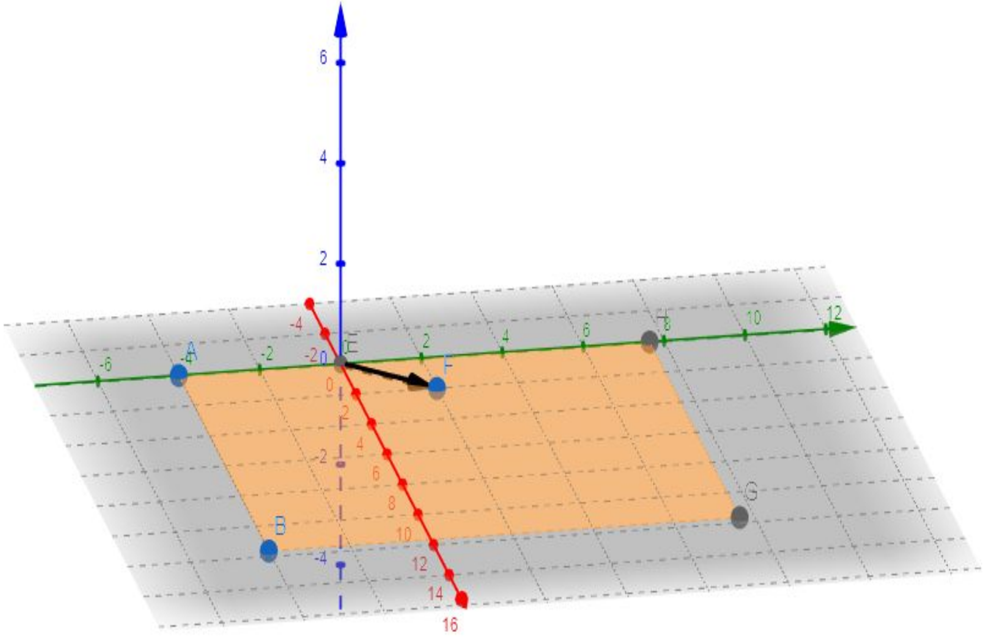
Some Observations !!!

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w \\ x \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}$$

C(A)

$$span \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

What is the dimension of Column space of Matrix A?
Will the basis vectors of C(A) span the entire 3-D space?



What can you say about this?

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix} \begin{pmatrix} w \\ x \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

What can you say about this?

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix} \begin{pmatrix} w \\ x \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

$$w \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + 2x \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

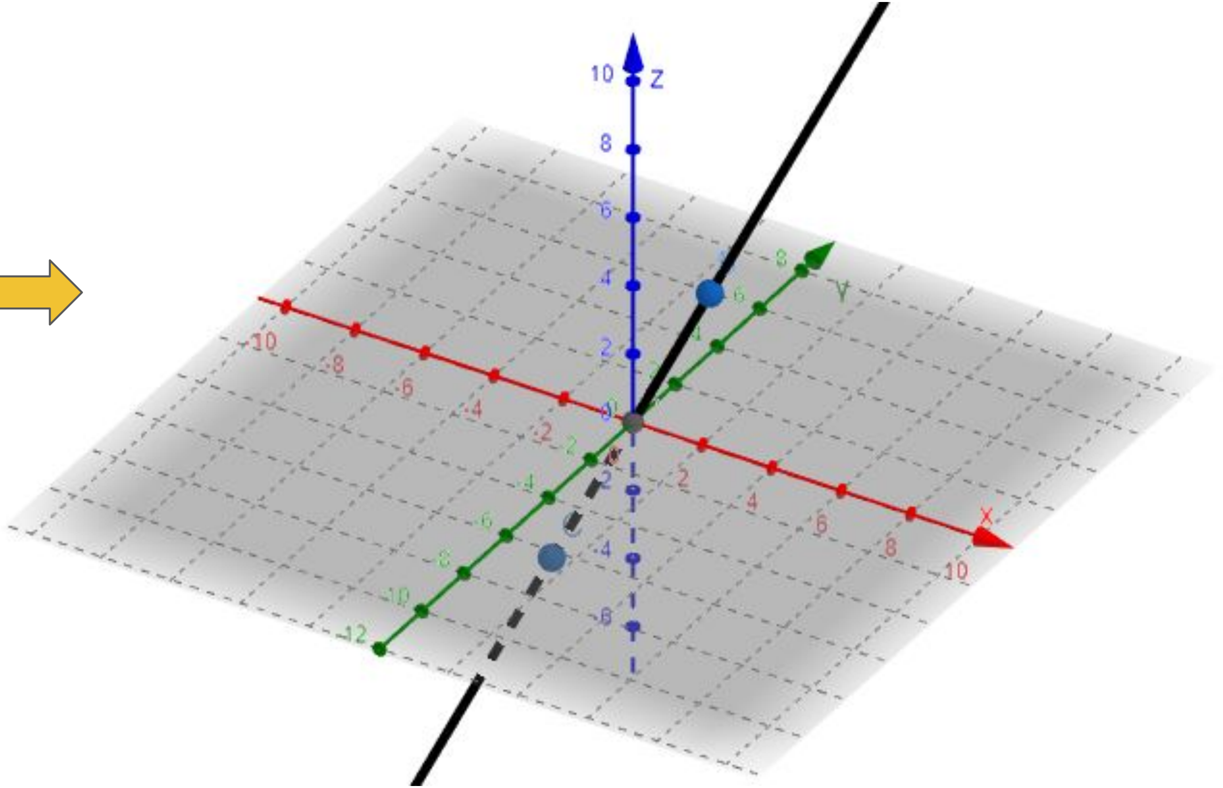
Column Space of Matrix A - Column space of matrix A denoted as C(A) is the space spanned by the column vectors of A.

$$C(A) = \mathit{span} \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\}$$

Dimension of C(A) = 1. Here Column space is a line passing through origin.

Do you see a Subspace ?

$$C(A) = span\left\{\begin{bmatrix}1\\2\\3\end{bmatrix}\right\}$$



Harikrishnan N B

Is there anything Mysterious ?

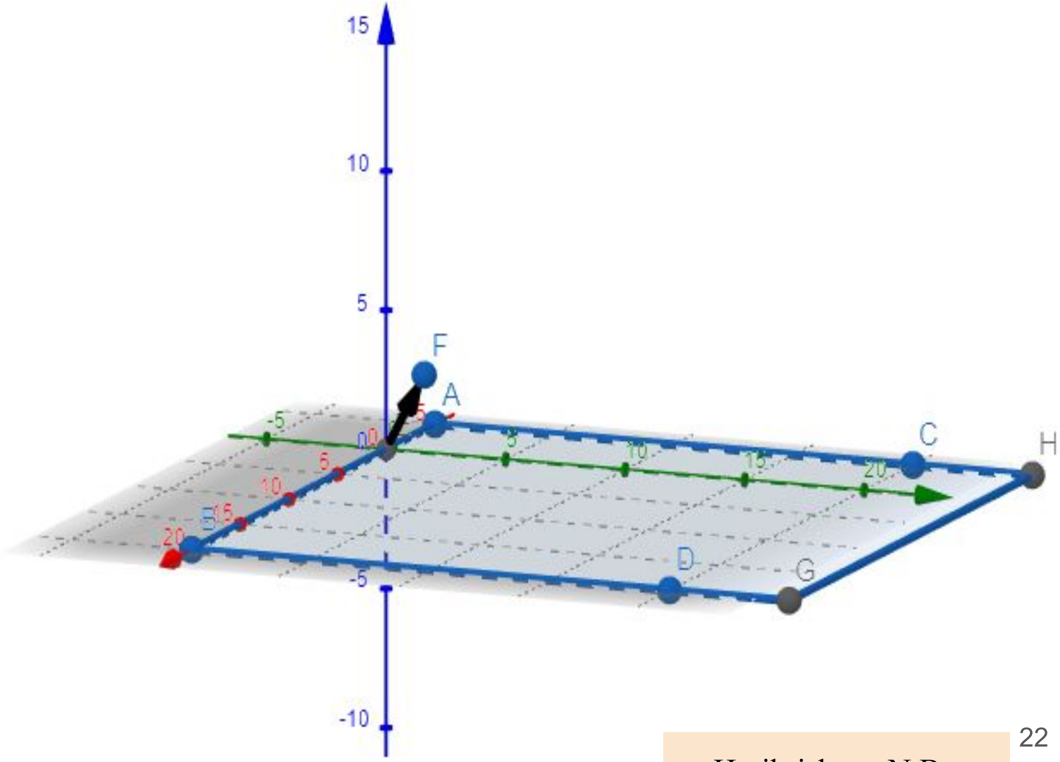
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} w \\ x \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Is there anything Mysterious ?

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} w \\ x \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

$$\Downarrow$$

$$w \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$



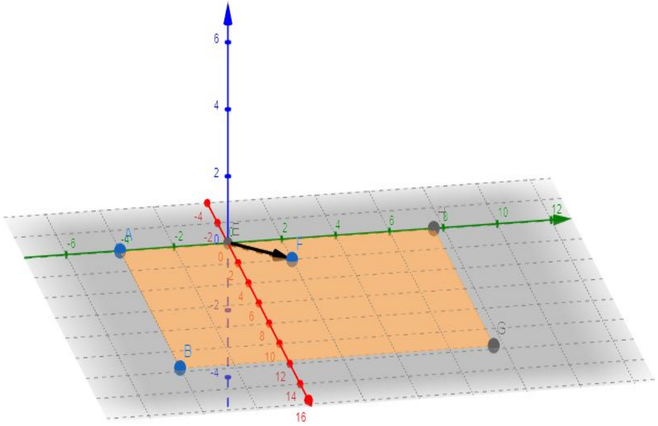
Harikrishnan N B

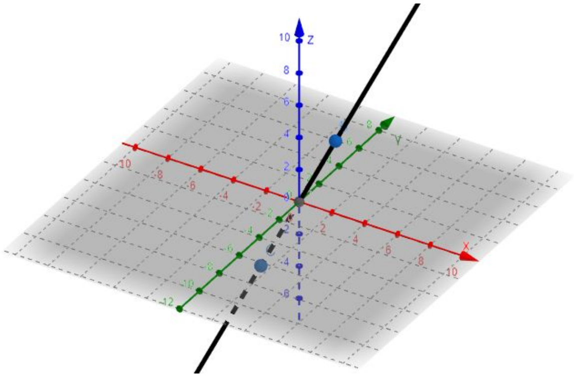
Solution to Ax = b

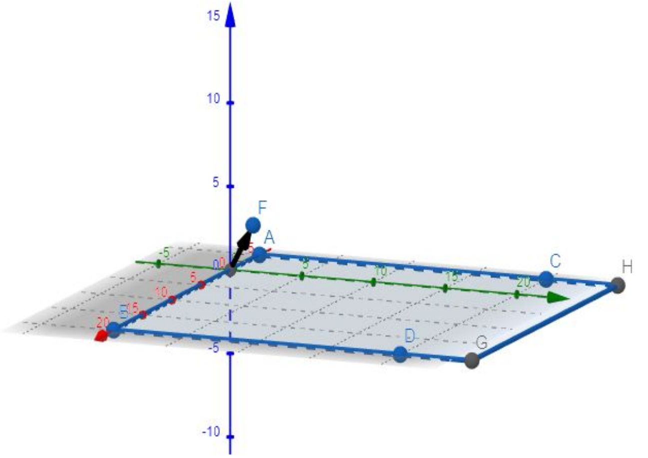
$$\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}\begin{bmatrix}w\\x\end{bmatrix}=\begin{bmatrix}2\\2\\0\end{bmatrix}$$

$$\begin{bmatrix}1&2\\2&4\\3&6\end{bmatrix}\begin{bmatrix}w\\x\end{bmatrix}=\begin{bmatrix}1\\2\\3\end{bmatrix}$$

$$\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}\begin{bmatrix}w\\x\end{bmatrix}=\begin{bmatrix}1\\2\\3\end{bmatrix}$$







UNIQUE SOLUTION

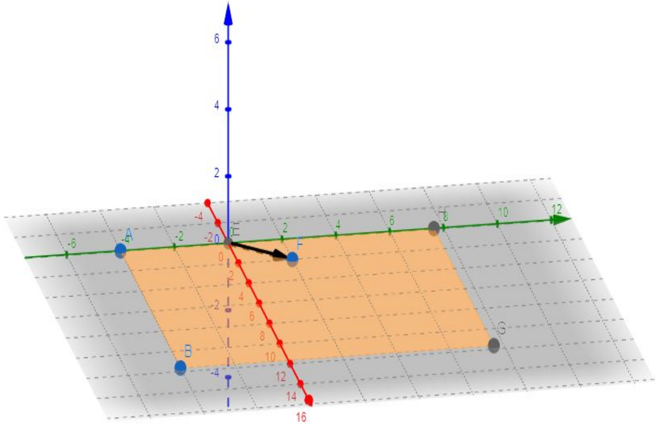
INFINITELY MANY SOLUTIONS

NO SOLUTION

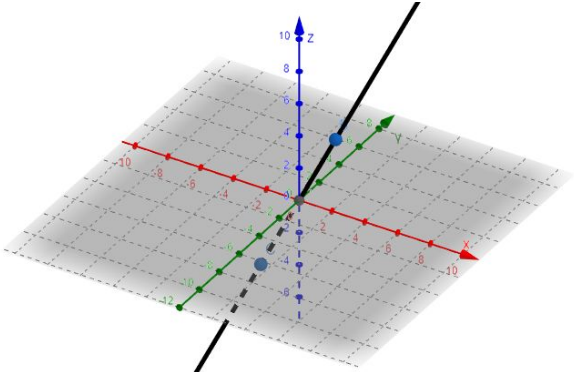
Harikrishnan N B

So when does Ax = b have a Solution

Ax = b has solution when b lies in the column space of A or in other words b is a linear combination of column vectors of A.



UNIQUE SOLUTION



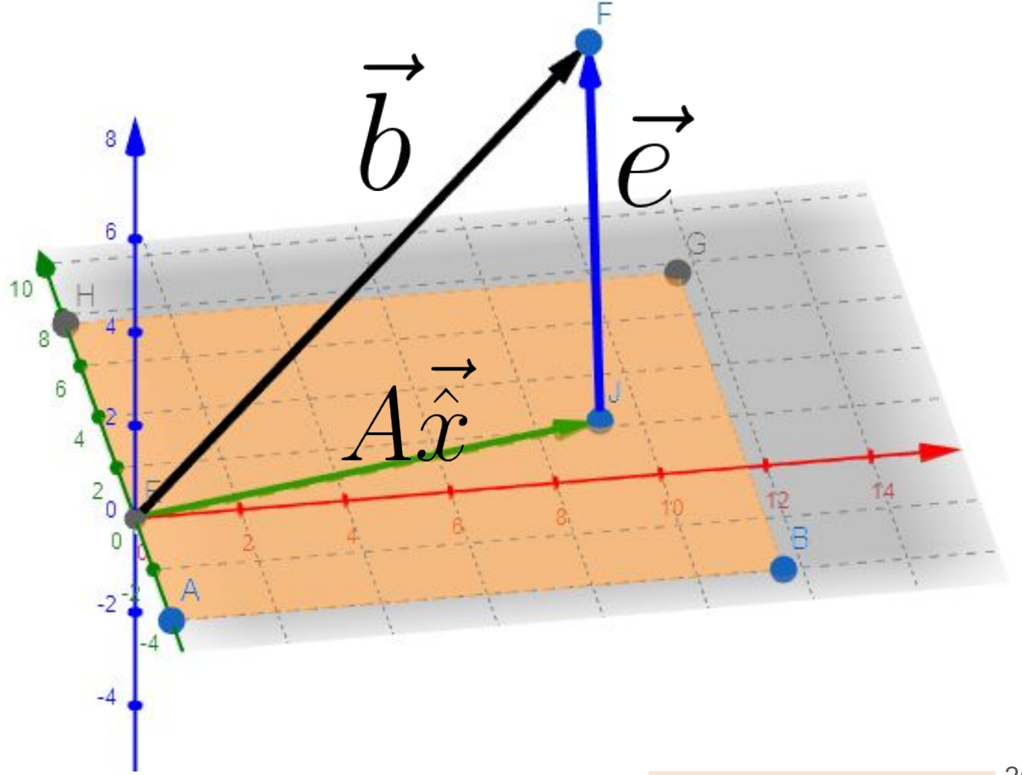
INFINITELY MANY SOLUTIONS

- For unique solution and infinitely many solutions b lies in the column space of A.
- In the case of NO solution b does not lie in the column space of A.

NO SOLUTION CASE

Can I find the best approximate solution ?

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w \\ x \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$



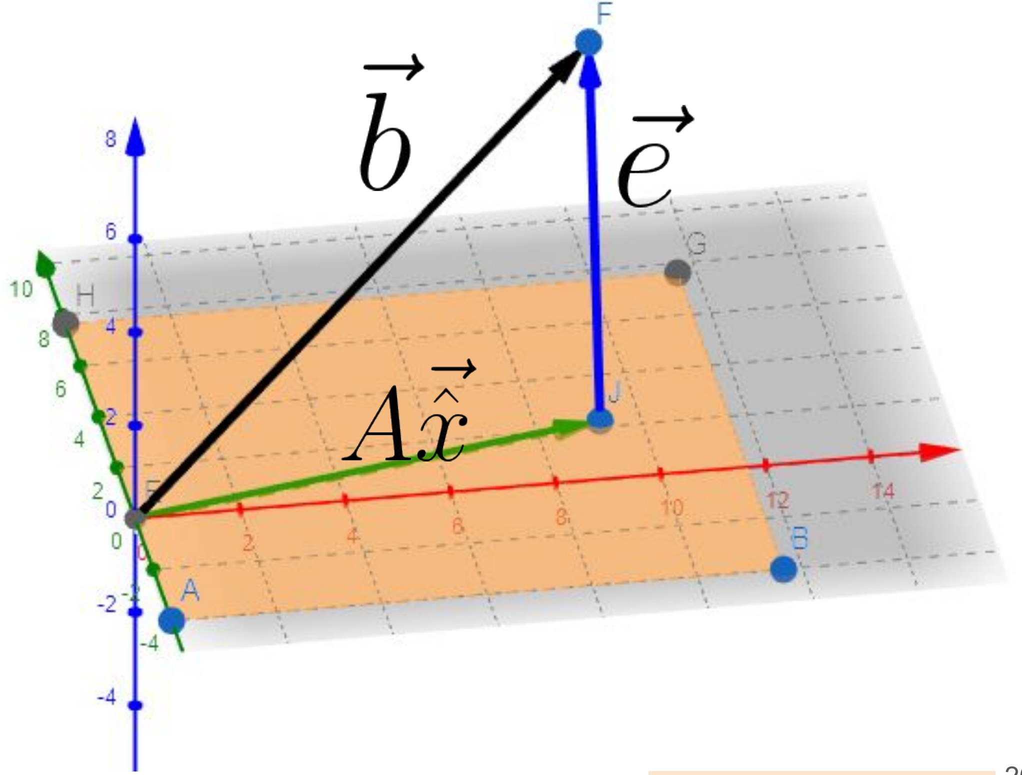
Harikrishnan N B

NO SOLUTION CASE

Can I find the best approximate solution ?

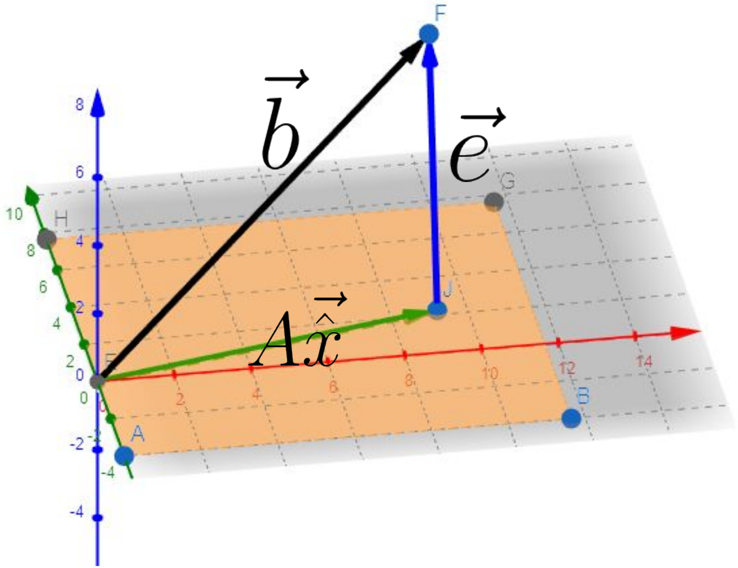
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w \\ x \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$A\vec{\hat{x}} + \vec{e} = \vec{b}$$
$$\vec{e} = A\vec{\hat{x}} - \vec{b}$$
$$A^T\vec{e} = \vec{0}$$
$$A^T(A\vec{\hat{x}} - \vec{b}) = \vec{0}$$
$$A^TA\vec{\hat{x}} - A^T\vec{b} = \vec{0}$$
$$A^TA\vec{\hat{x}} = A^T\vec{b}$$
$$\vec{\hat{x}} = (A^TA)^{-1}A^T\vec{b}$$

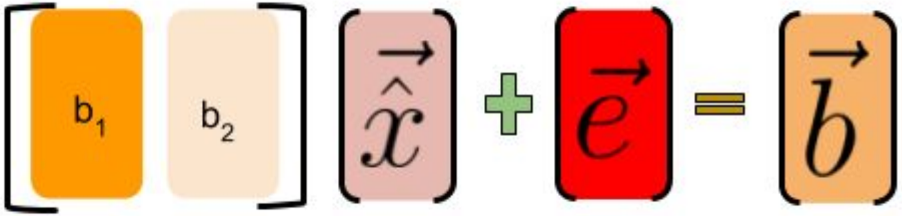


Harikrishnan N B

No solution case - Visualization



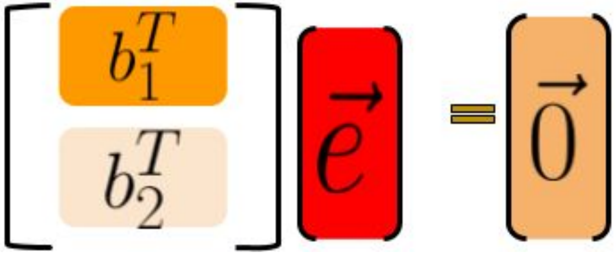
$$A\hat{\vec{x}} + \vec{e} = \vec{b}$$



$$A\hat{\vec{x}} + \vec{e} = \vec{b}$$
$$\vec{e} = A\hat{\vec{x}} - \vec{b}$$
$$A^T\vec{e} = \vec{0}$$
$$A^T(A\hat{\vec{x}} - \vec{b}) = \vec{0}$$
$$A^T A\hat{\vec{x}} - A^T\vec{b} = \vec{0}$$
$$A^T A\hat{\vec{x}} = A^T\vec{b}$$
$$\hat{\vec{x}} = (A^T A)^{-1}A^T\vec{b}$$

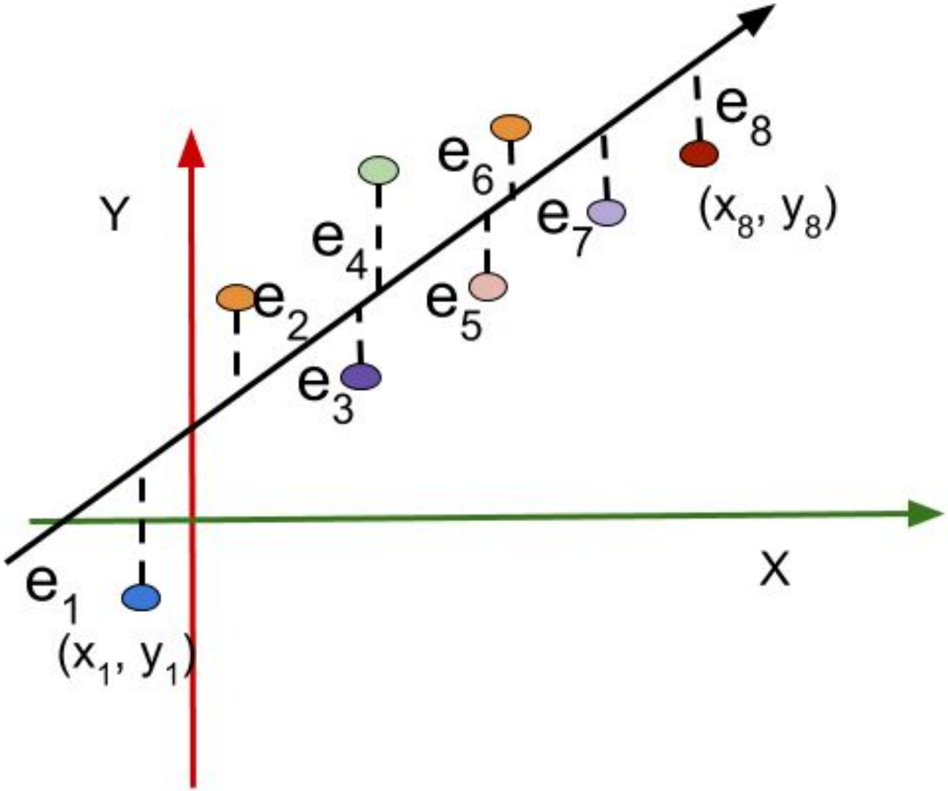
$$A^T\vec{e} = \vec{0}$$

ORTHOGONAL

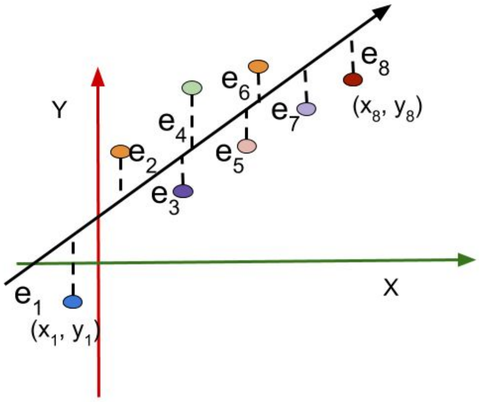


Harikrishnan N B

Linear Least Square Regression



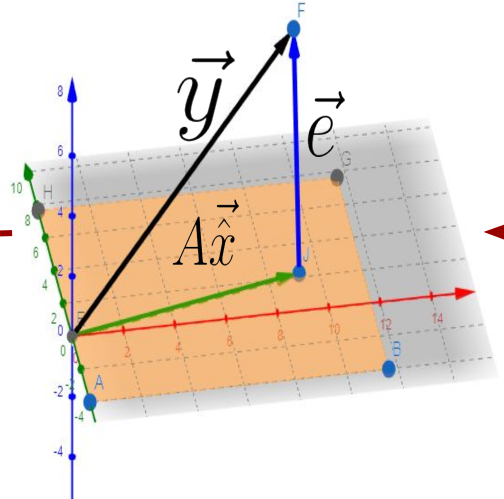
$$y_1 = mx_1 + c + e_1$$
$$y_2 = mx_2 + c + e_2$$
$$y_3 = mx_3 + c + e_3$$
$$y_4 = mx_4 + c + e_4$$
$$y_5 = mx_5 + c + e_5$$
$$y_6 = mx_6 + c + e_6$$
$$y_7 = mx_7 + c + e_7$$
$$y_8 = mx_8 + c + e_8$$



$$y_1 = mx_1 + c + e_1$$
$$y_2 = mx_2 + c + e_2$$
$$y_3 = mx_3 + c + e_3$$
$$y_4 = mx_4 + c + e_4$$
$$y_5 = mx_5 + c + e_5$$
$$y_6 = mx_6 + c + e_6$$
$$y_7 = mx_7 + c + e_7$$
$$y_8 = mx_8 + c + e_8$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \\ x_5 & 1 \\ x_6 & 1 \\ x_7 & 1 \\ x_8 & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \end{bmatrix}$$

$$\vec{y} = A\vec{x} + \vec{e}$$



$$\boxed{\vec{x} = (A^T A)^{-1} A^T \vec{y}}$$

What happens in this case ? - Code it

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \begin{bmatrix} x_1 & x_1^2 & 1 \\ x_2 & x_2^2 & 1 \\ x_3 & x_3^2 & 1 \\ x_4 & x_4^2 & 1 \\ x_5 & x_5^2 & 1 \\ x_6 & x_6^2 & 1 \\ x_7 & x_7^2 & 1 \\ x_8 & x_8^2 & 1 \end{bmatrix} \begin{bmatrix} m \\ p \\ c \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \end{bmatrix}$$

What change you will observe in the graph?

What happens when you add more higher order terms like x^3, x^4.. x^n?

$$\boxed{\vec{x} = (A^T A)^{-1} A^T \vec{y}}$$

$$\vec{y} = A\vec{x} + \vec{e}$$

Practical Challenges

$$\boxed{\vec{x} = (A^T A)^{-1} A^T \vec{y}}$$

- Curse of Dimensionality
  - Computing the inverse of a matrix has a complexity of order O(N^3).
  - In the case of high dimensional data, we go for a matrix free implementation of linear least square regression.
- When to use linear regression and non-linear regression depends on the problem.

Dimensionality Reduction Techniques

- PCA - (Principal Component Analysis)
- SVD - (Singular Value Decomposition)
- ICA - (Independent Component Analysis)
- DMD - (Dynamic Mode Decomposition)

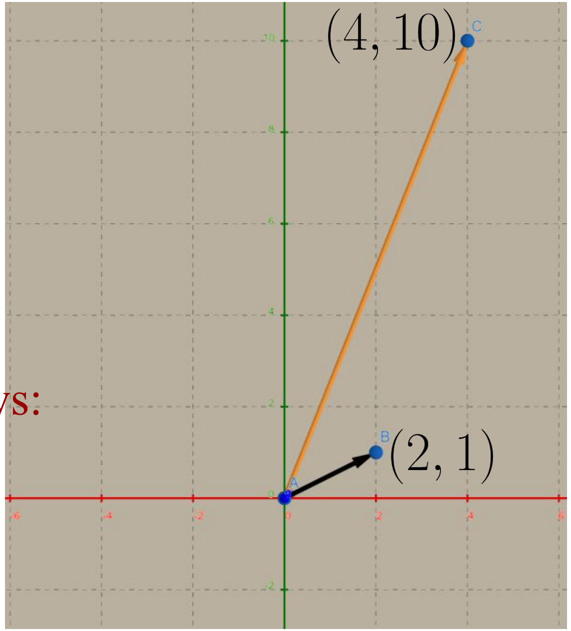
Things to Know - Before running into PCA

- Eigenvectors/Eigenvalues
- Spectral Decomposition
- Variance- Covariance Matrix
- Change of Basis

Matrix Vector Multiplication as a Transformation

Intuition for Matrix vector multiplication for Square Matrices

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 10 \end{bmatrix}$$



Matrix(Square Matrix) vector multiplication can be seen as follows:

- Rotation
- Stretching or Shrinking

Special Vectors

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -0.415 \\ -0.909 \end{bmatrix} = \begin{bmatrix} -2.234 \\ -4.885 \end{bmatrix} = 5.372 \begin{bmatrix} -0.415 \\ -0.909 \end{bmatrix}$$

$$A\vec{x} \qquad \lambda\vec{x}$$

$$A\vec{x} = \lambda\vec{x}$$

1. Direction of $\vec{x}$ is unchanged. (No rotation)
2. Only the magnitude is scaled by a factor λ
3. $\vec{x}$ - **eigenvector of matrix A**
4. λ - **eigenvalue of matrix A**

Eigenvalues and Eigenvectors

- For an n x n square matrix A, there are 'n' eigenvalues and 'n' eigenvectors. Let x_1, x_2, x_3,....x_n be the 'n' eigenvectors and $\lambda_1, \lambda_2, \lambda_3, \ldots \lambda_n$,are the corresponding eigenvalues.

$$A\vec{x_1} = \lambda_1 \vec{x_1}$$
$$A\vec{x_2} = \lambda_2 \vec{x_2}$$
$$A\vec{x_3} = \lambda_3 \vec{x_3}$$
$$.$$
$$.$$
$$A\vec{x_n} = \lambda_n \vec{x_n}$$

$$X = \begin{bmatrix} | & | & | & & | \\ \vec{x_1} & \vec{x_2} & \vec{x_3} & \cdots & \vec{x_n} \\ | & | & | & & | \end{bmatrix}$$

Very Very Important Part

- For an n x n square matrix A, there are ‘n’ eigenvalues and ‘n’ eigenvectors. Let x_1, x_2, x_3,....x_n be the ‘n’ eigenvectors and $\lambda_1, \lambda_2, \lambda_3, \ldots \lambda_n$,are the corresponding eigenvalues.

$$AX = \begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} | & | & | & & | \\ \vec{x_1} & \vec{x_2} & \vec{x_3} & \cdots & \vec{x_n} \\ | & | & | & & | \end{bmatrix}$$

Spectral Decomposition

$$\begin{bmatrix} & & & \\ & A & & \\ & & & \end{bmatrix} \begin{bmatrix} | & | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \vec{x}_3 & \cdots & \vec{x}_n \\ | & | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & | & & | \\ A\vec{x}_1 & A\vec{x}_2 & A\vec{x}_3 & \cdots & A\vec{x}_n \\ | & | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & | & & | \\ \lambda_1\vec{x}_1 & \lambda_2\vec{x}_2 & \lambda_3\vec{x}_3 & \cdots & \lambda_n\vec{x}_n \\ | & | & | & & | \end{bmatrix}$$

$$\begin{bmatrix} | & | & | & & | \\ \lambda_1\vec{x}_1 & \lambda_2\vec{x}_2 & \lambda_3\vec{x}_3 & \cdots & \lambda_n\vec{x}_n \\ | & | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \vec{x}_3 & \cdots & \vec{x}_n \\ | & | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

Harikrishnan N B

Spectral Decomposition

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} | & | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \vec{x}_3 & \cdots & \vec{x}_n \\ | & | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \vec{x}_3 & \cdots & \vec{x}_n \\ | & | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$AX = X\Lambda$$

$$AXX^{-1} = X\Lambda X^{-1}$$

$$AI = X\Lambda X^{-1}$$

$$A = X\Lambda X^{-1}$$

Practical Challenges and Important Points

When can we apply $A = X\Lambda X^{-1}$?

- A should be a square matrix
- When A has 'n' distinct eigenvalues, then X^{-1} always exist.

What happens when A is Symmetric ($A^T = A$)?

- The eigenvectors of a symmetric matrix A can be chosen as ORTHONORMAL. So in this case X is orthonormal.
- For an ORTHONORMAL matrix X, the inverse is its transpose $X^{-1} = X^T$
- $A = X\Lambda X^{-1}$
  $A = X\Lambda X^T$

Orthogonal and Orthonormal vectors

Orthogonal vectors

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

L2 - norm = $\sqrt{2}$

Orthonormal vectors

$$\begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = 0$$

L2 - norm = 1

Orthonormal Matrix

$$XX^T = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Practical Challenges

What if A is not a square matrix?

- We cannot apply Spectral Decomposition.

Don't Worry!!!

Singular Value Decomposition works for any Matrix.

A Few more steps to PCA

What all minimum can we say about this data?

X	1	2	3	4	5
Y	1	5	4	6	7

X, Y are the features

What all minimum can we say about this data?

Mean(X) = $$\frac{1}{N}\sum_{i=1}^{N} x_i$$

Variance(X) = $$\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2$$

Cov(X,Y) = $$\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)$$

X	Y	var(X)	var(Y)	cov(X,Y)	cov(Y,X)
1	1				
2	5				
3	4				
4	6				
5	7				
Mean (X)	Mean(Y)	var(X)	var(Y)	cov(X,Y)	cov(Y,X)
3.0	**4.6**	**2.5**	**5.3**	**2.25**	**2.25**

Variance- Covariance Matrix

$$\begin{array}{cc} & \begin{array}{cc} X & Y \end{array} \\ \begin{array}{c} X \\ Y \end{array} & \begin{bmatrix} var(X) & cov(X,Y) \\ cov(Y,X) & var(Y) \end{bmatrix} \end{array}$$

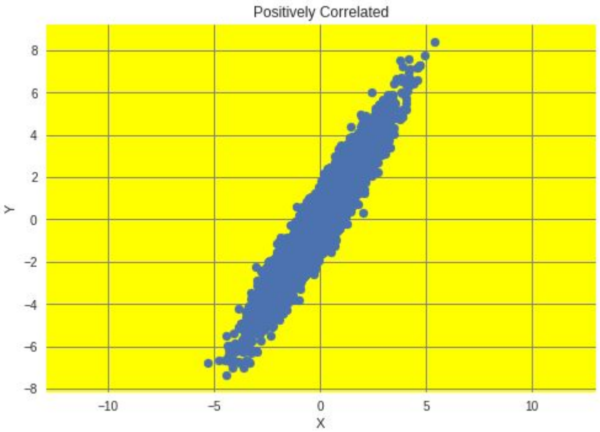
Recall the properties of a symmetric matrix!!!

- **Variance - Covariance Matrix is symmetric. cov(X,Y) = cov(Y,X)**
- **The diagonal entries represents variance**
- **The off- diagonal entries represents the correlation of X and Y**

What does Variance - Covariance Matrix signifies?

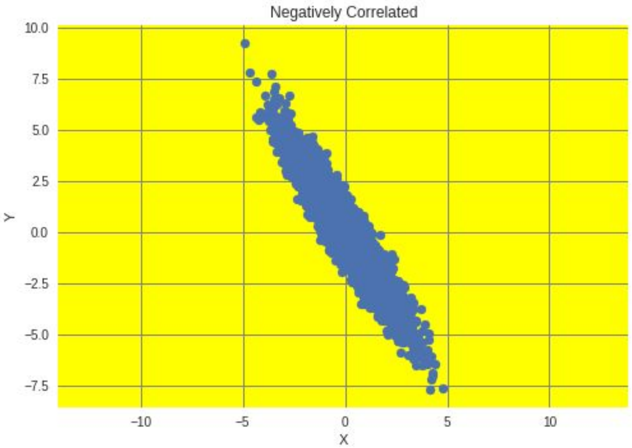
Case I

$$\begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}$$



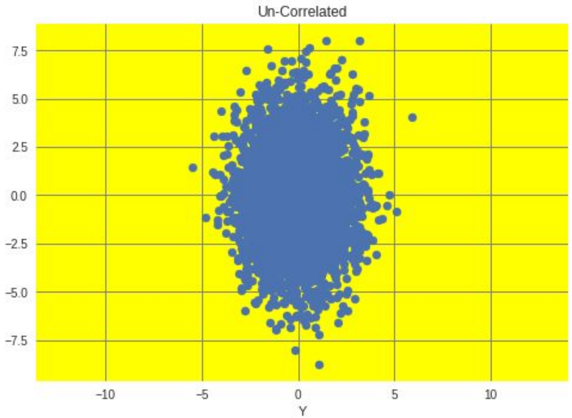
Case II

$$\begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$$



Case III

$$\begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}$$

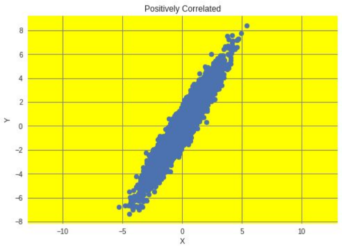


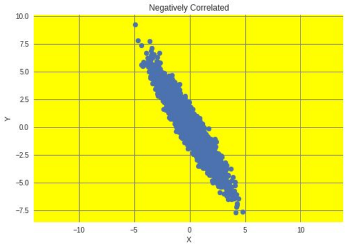
Note: In all cases mean is (0,0)

Harikrishnan N B

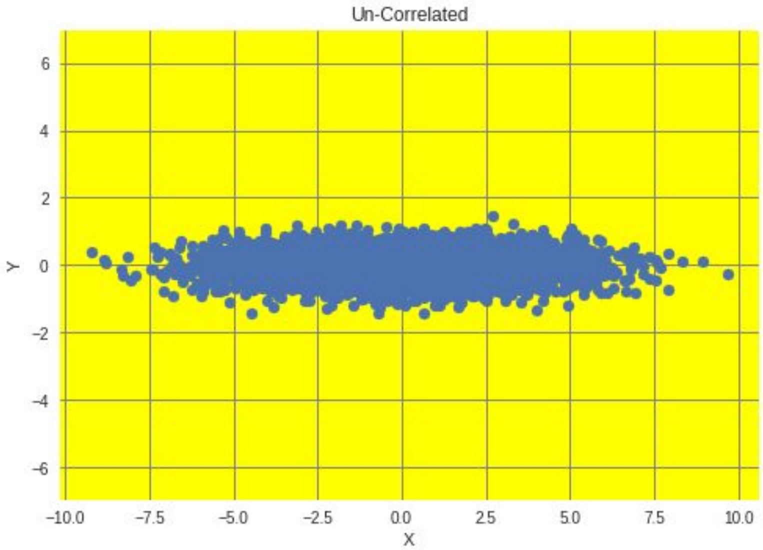
So what does PCA do ?

- Principal Component Analysis (PCA) makes the data **UNCORRELATED.**



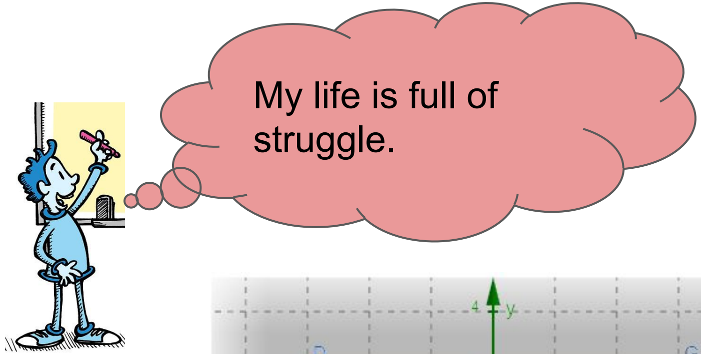


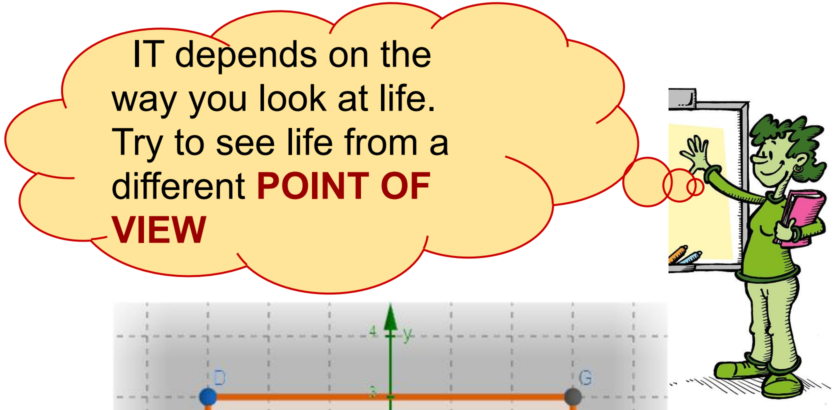
PCA achieves this by **Change of Basis**

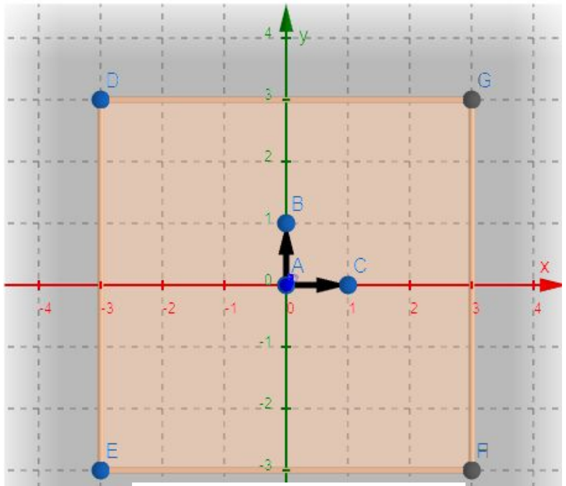


Harikrishnan N B

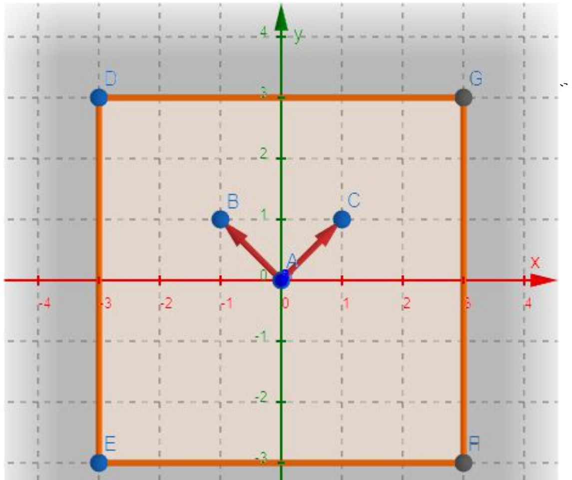
Change of Basis







Old Basis



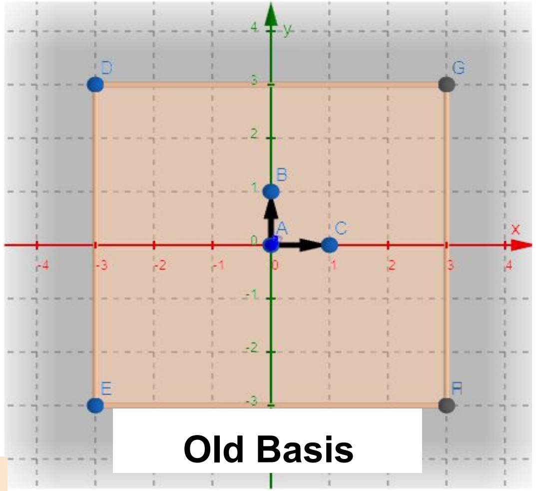
New Basis

Recall

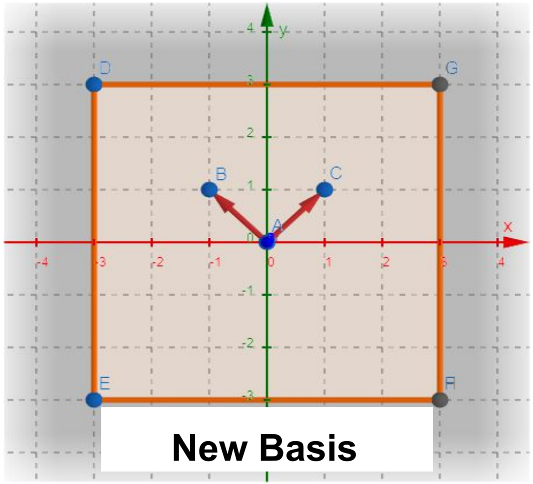
Dimension of a Vector space - Every vector space has a dimension. Dimension is the number of basis vectors required to span the vector space.

Properties of Basis Vectors -

- Basis vectors has to be linearly independent.
- Basis vectors should span the vector space.



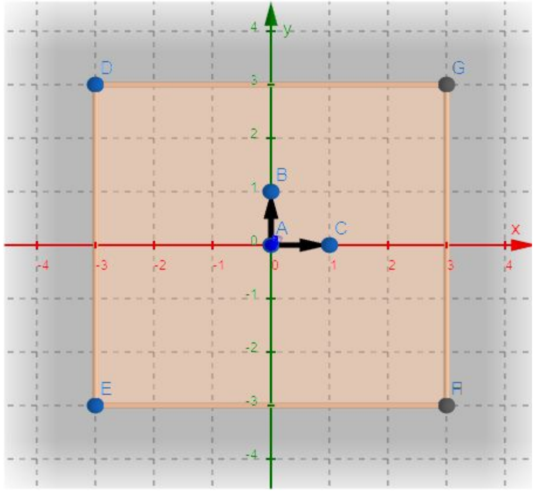
Old Basis



New Basis

Example of Change of Basis

To represent a point (2,3) in old basis and new basis- How to understand this?



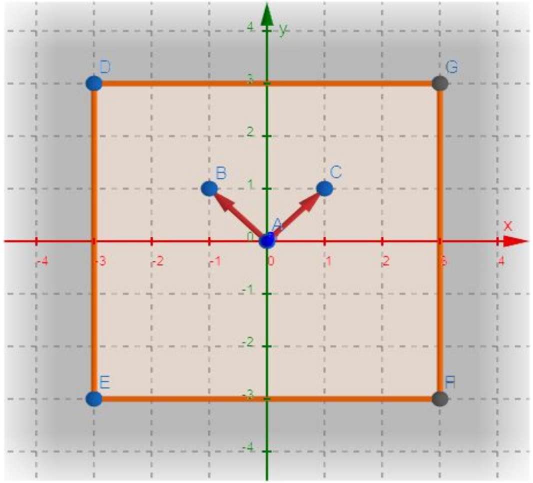
Old basis = $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$

$$2\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Harikrishnan N B

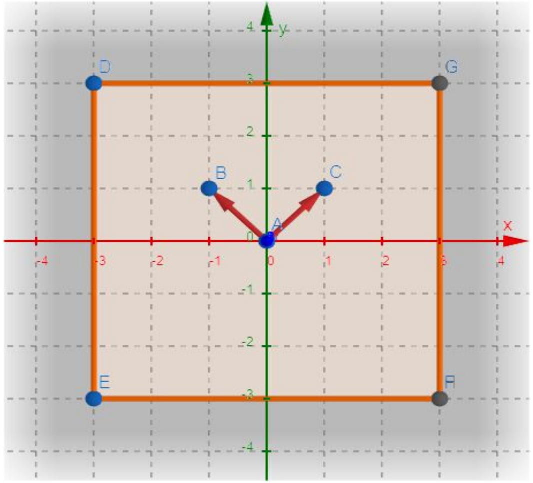
New Basis Representation



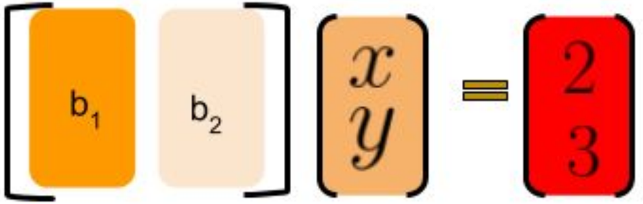
New Basis $= \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right\}$

$$x \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} + y \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

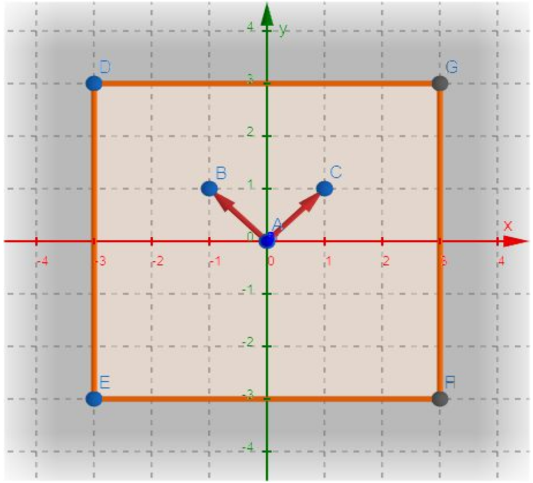
Finding x and y for representing (2,3) using new basis



$$x\begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{bmatrix}+y\begin{bmatrix}\frac{-1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{bmatrix}=\begin{bmatrix}2\\3\end{bmatrix}$$



$$\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}}\\\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}2\\3\end{bmatrix}$$



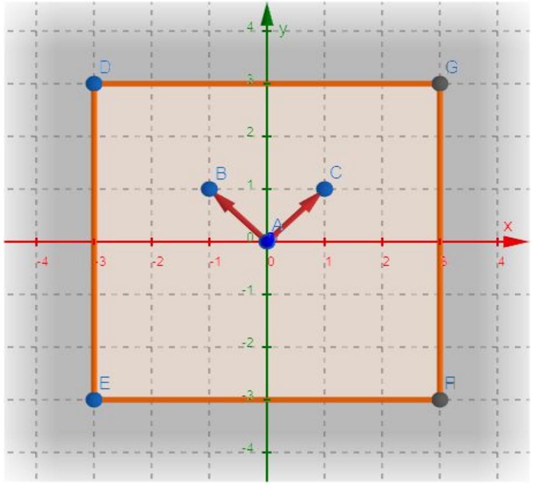
$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$P\vec{x} = \vec{y}$$

$$P^{-1}P\vec{x} = P^{-1}\vec{y}$$

$$\vec{x} = P^{-1}\vec{y}$$

For **ORTHONORMAL MATRIX, $P^{-1} = P^{T}$**



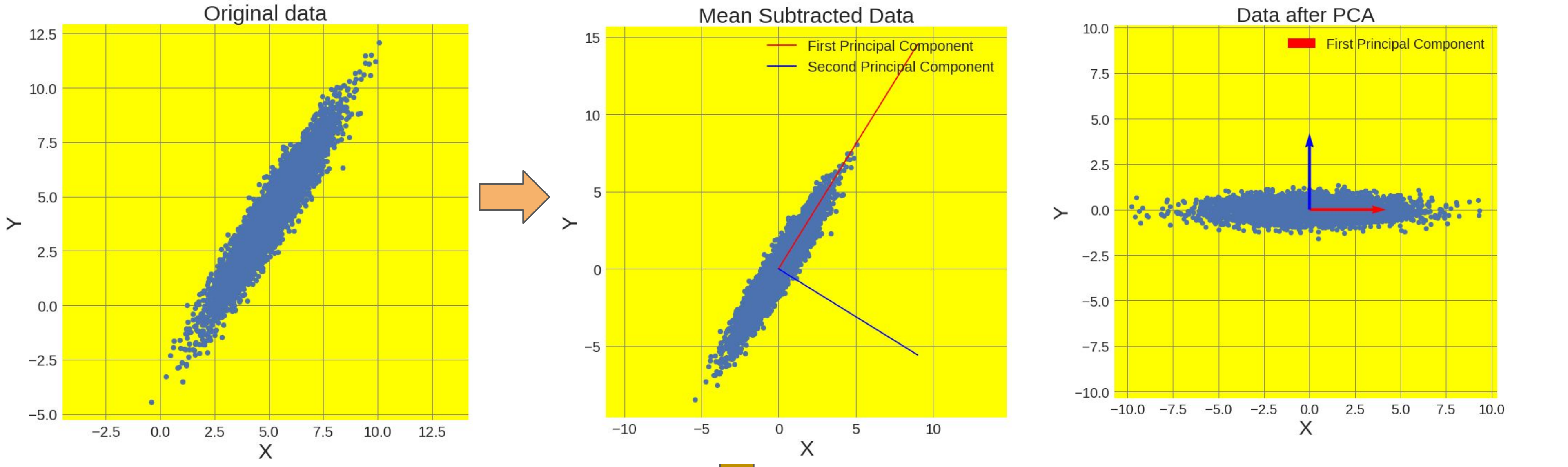
In our case the matrix P is ORTHONORMAL

$$\vec{x} = P^{-1}\vec{y} = P^{T}\vec{y}$$

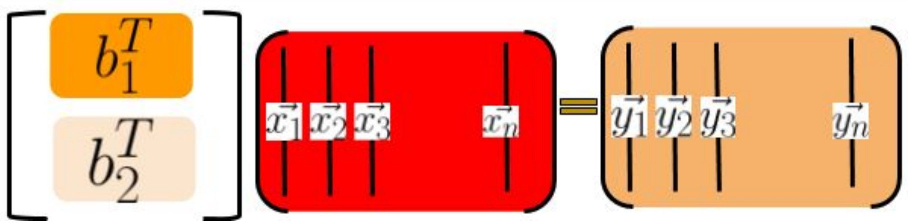
$$\begin{bmatrix} b_1^T \\ b_2^T \end{bmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \quad \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{5}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\frac{5}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Steps in PCA

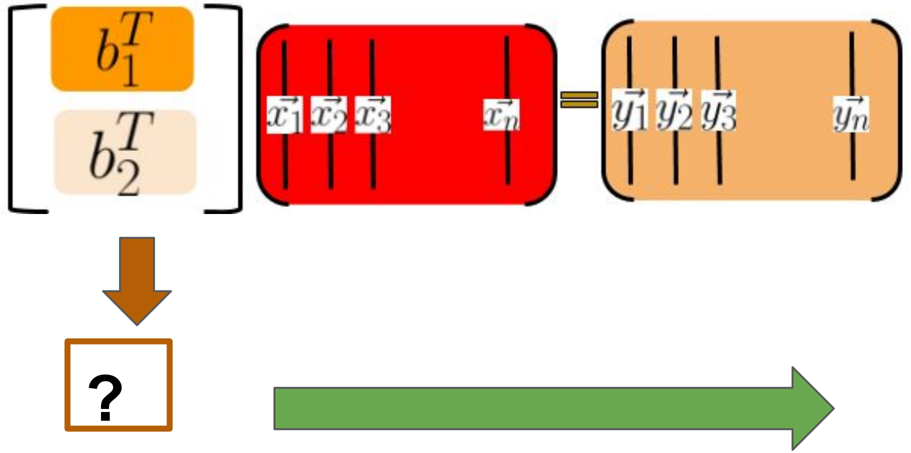


Change of Basis



Harikrishnan N B

What should be the NEW BASIS so that DATA is UNCORRELATED?



Rows of matrix P are the **eigenvectors** of the **variance -covariance matrix** of the **mean subtracted data**

$$PX = Y$$
$$cov(Y) = cov(PX)$$
$$cov(PX) = \frac{1}{N-1}(PX)(PX)^T$$
$$cov(PX) = \frac{1}{N-1}PXX^TP^T$$
$$cov(PX) = P(\frac{1}{N-1}XX^T)P^T$$
$$cov(PX) = Pcov(X)P^T$$
$$cov(PX) = P(V\Lambda V^T)P^T$$
$$P = V^T$$
$$cov(PX) = \Lambda$$

Some words about PCA

- PCA is "an orthogonal linear transformation that transfers the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (*first principal component*), the second greatest variance lies on the second coordinate (*second principal component*), and so on."

Applications of PCA

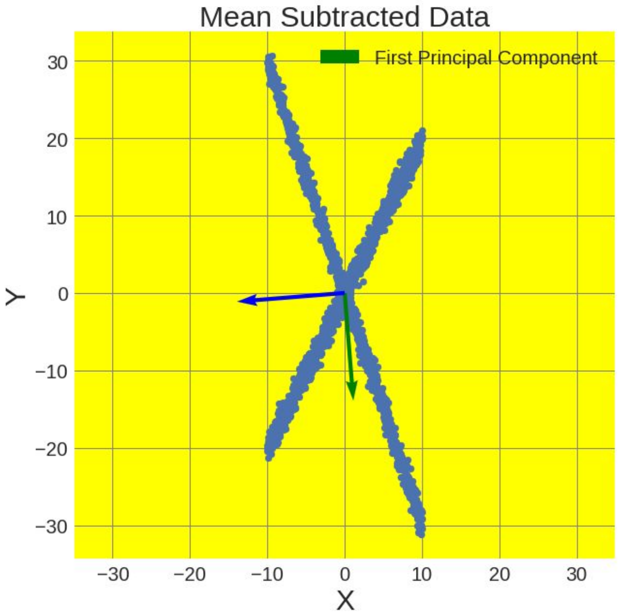
- Dimensionality Reduction
- Denoising
- Feature Extraction
- Image Compression
- EEG Analysis

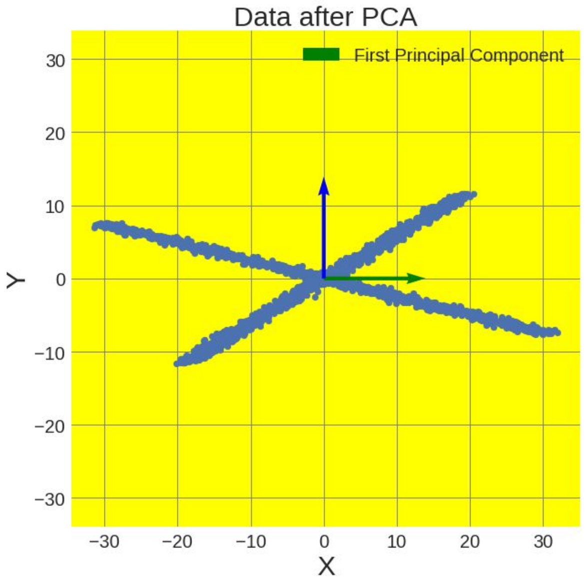
Assumptions in PCA

- Linearity
- Large variance have important structure
- Principal components are orthogonal

When does PCA fail?

- Non-linearity
- Non-Gaussian
- Non-orthogonality



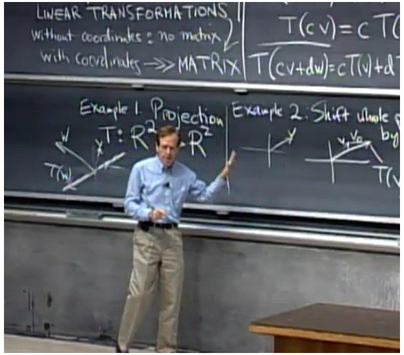


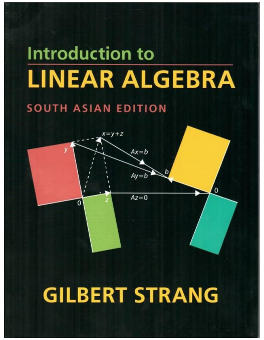
Ref: https://arxiv.org/abs/1404.1100

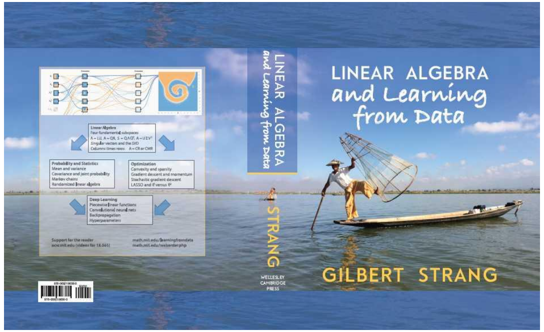
Harikrishnan N B

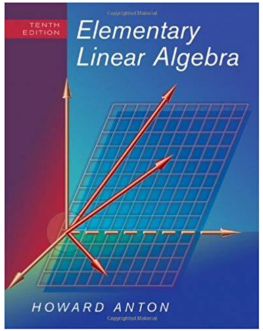
Interesting Materials

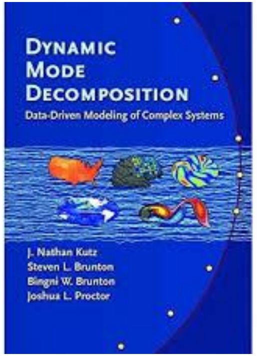
Prof. Gilbert Strang



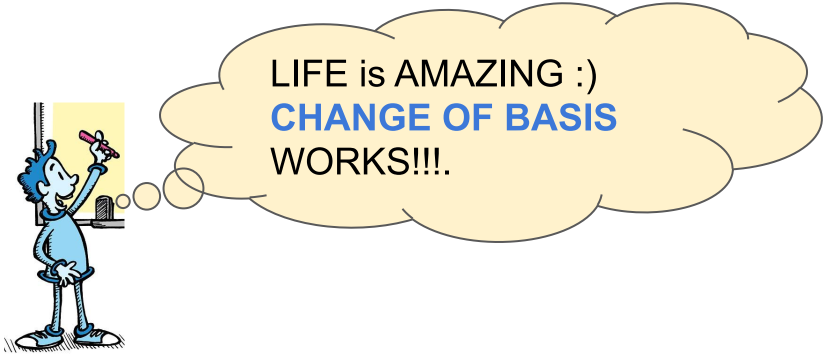








Tutorial on PCA - (Click here)



Harikrishnan N B
Research Associate
Consciousness Studies Programme
National Institute of Advanced Studies,
Indian Institute of Science Campus, Bengaluru, India

Email: harikrishnannb07@gmail.com
Web: https://sites.google.com/site/harikrishnannb8/home