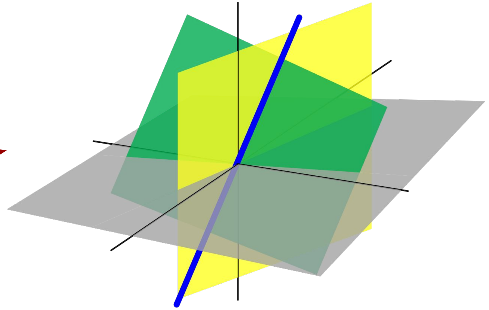
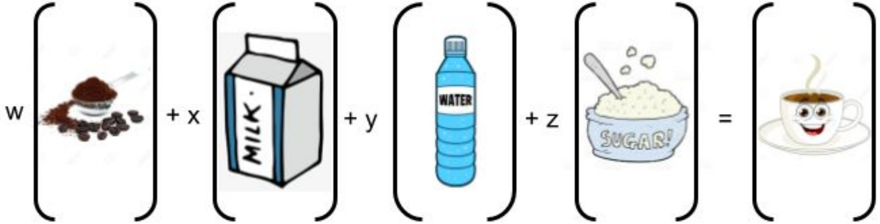
From Coffee Space to Vector Space: Foundations of Linear Algebra

IISc SIAM Student Chapter
Indian Institute of Science, Bengaluru



Coffee Space- Vector Space

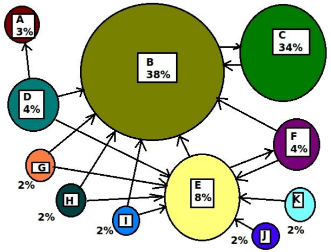


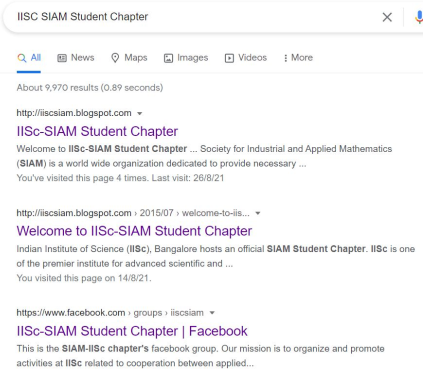
Harikrishnan N B
Research Associate
Consciousness Studies Programme
National Institute of Advanced Studies

28 August 2021

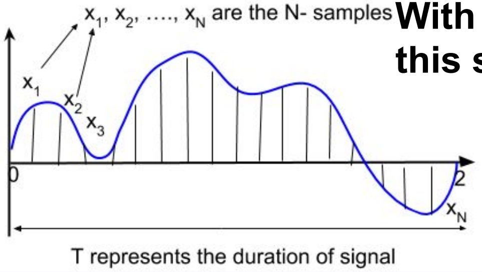
Why should I learn Linear Algebra?

Mathematics of Google Search - Page Rank

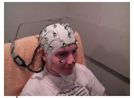




With what all frequency this signal is made up of ?



Common EEG Artifacts



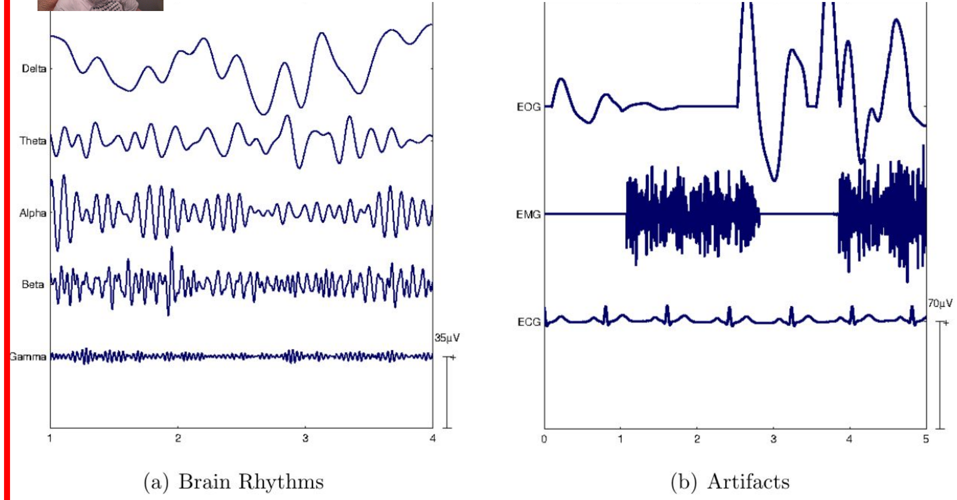


Figure 1. (a) *Five normal brain rhythms, from low to high frequencies.* Delta, Theta, Alpha, Beta and Gamma rhythms comprise the background EEG spectrum. (b) *Three different types of artifacts.* Ocular, muscular and cardiac artifacts are the most frequent physiological contaminants in the literature on EEG artifact removal.

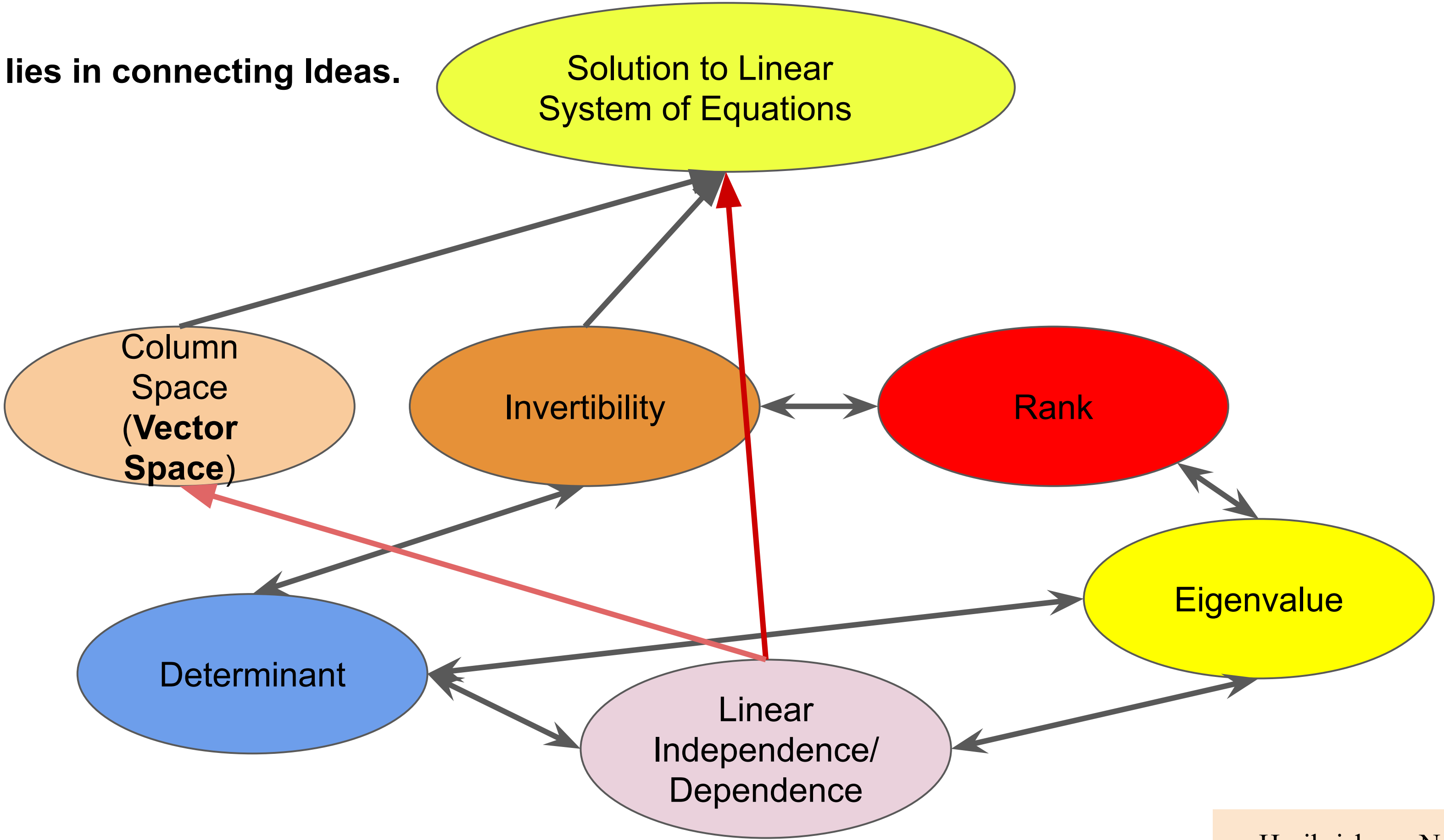
1. http://pi.math.cornell.edu/~mec/Winter2009/RalucaRemus/index.html (Page Rank)
2. Reference: Urigüen, J. A., & Garcia-Zapirain, B. (2015). EEG artifact removal—state-of-the-art and guidelines. *Journal of neural engineering*, *12*(3), 031001.

Goal of Linear Algebra

- The central problem of Linear Algebra is to **understand** a system of linear equations.
- **Understanding involves**
  - Insights about row picture and column picture.
  - Explore the existence of solution to the system of linear equations.
  - Insights about column space, row space, right null space, left null space.
  - **What new can we say about the system?**

OUR APPROACH

Beauty lies in connecting Ideas.



Two Equations and Two Unknowns-
Algebraic Interpretation

$$-x + y = 0$$
$$2x + y = 3$$

What is the value of the **unknown variables *x* and *y*** that satisfies this system of linear equations?

Elimination

$$\left[\begin{array}{cc|c} -1 & 1 & 0 \\ 2 & 1 & 3 \end{array}\right] R_1 \to 2R_1 \left[\begin{array}{cc|c} -2 & 2 & 0 \\ 2 & 1 & 3 \end{array}\right] \quad R_2 \to R_1 + R_2 \quad \left[\begin{array}{cc|c} -2 & 2 & 0 \\ 0 & 3 & 3 \end{array}\right]$$

$$3y = 3$$
$$y = 1$$

Sub. *y* = *1* in Equation 1, we get: ***x* = *1***

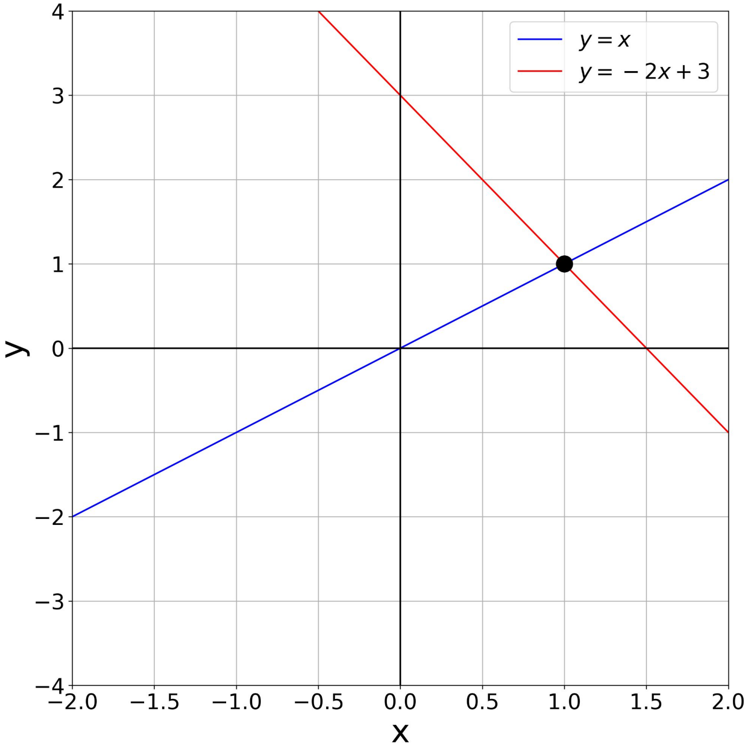
$$\text{Solution: } x = 1, \text{and } y = 1$$

Two Equations and Two Unknowns-
Geometric Interpretation

$$-x + y = 0$$
$$2x + y = 3$$

Row Picture

$$y = x$$
$$y = -2x + 3$$

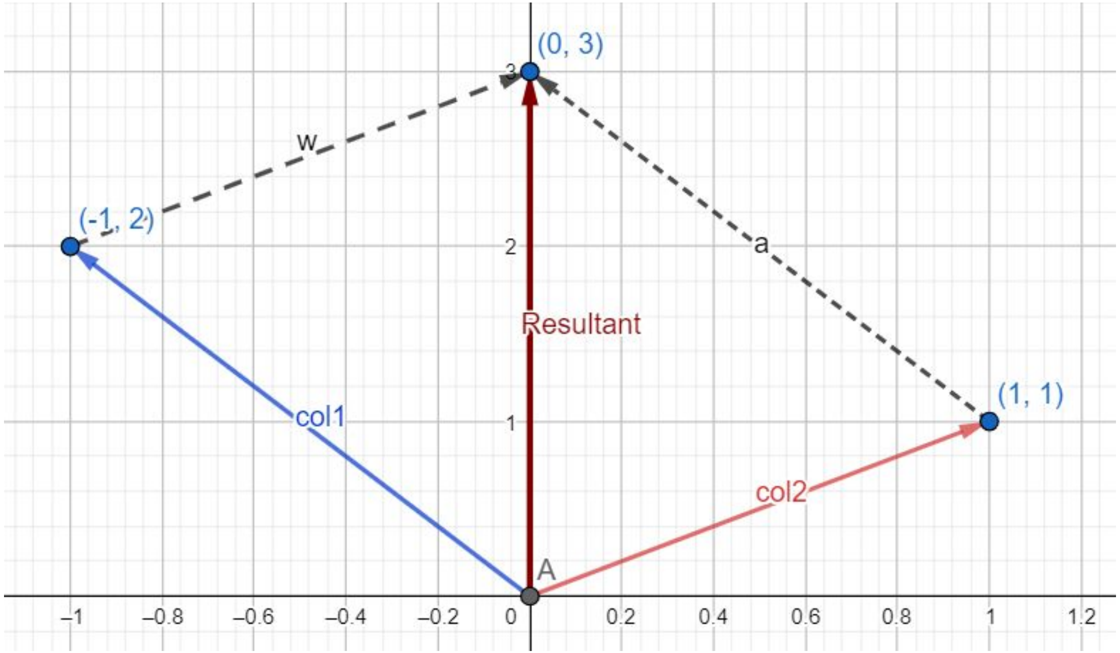


Two Equations and Two Unknowns-
Geometric Interpretation

$$-x + y = 0$$
$$2x + y = 3$$

Column Picture

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



Two Equations and Two Unknowns-
Some Observations

$$-x + y = 0$$
$$2x + y = 3$$

$$\begin{bmatrix} -1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix} \Leftrightarrow \boxed{A\vec{x} = b}$$

$$\Updownarrow$$

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

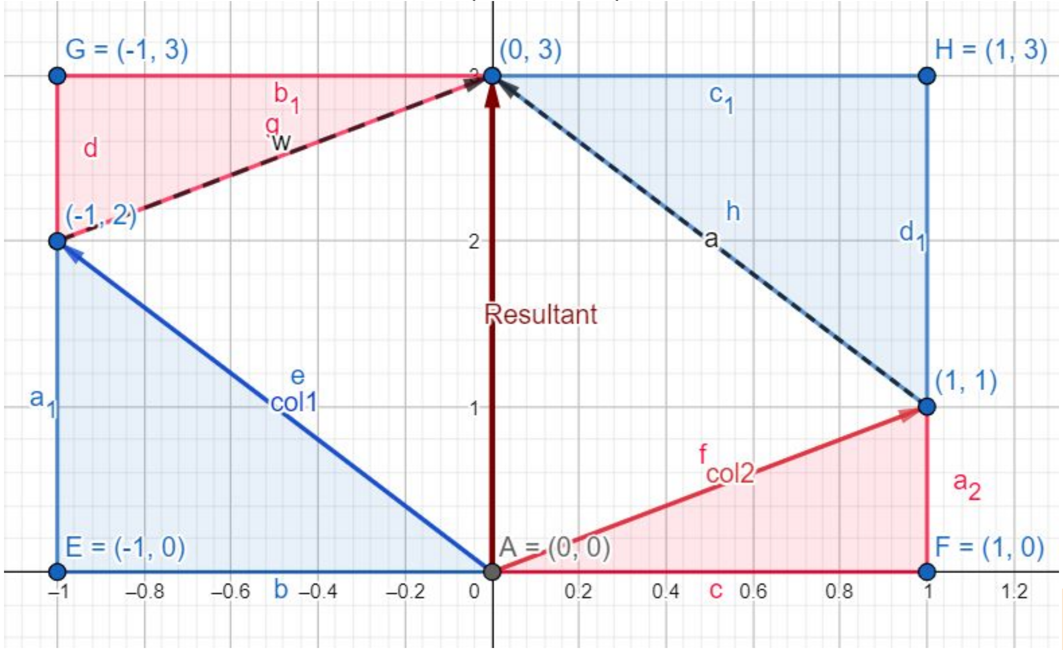
$A\vec{x} = b$ is the weighted linear combinations of columns of A

Two Equations and Two Unknowns-
Determinant

$$-x + y = 0$$
$$2x + y = 3$$

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = (ad - bc)$$

$$|A| = -3$$



Two Equations and Two Unknowns- Invertibility

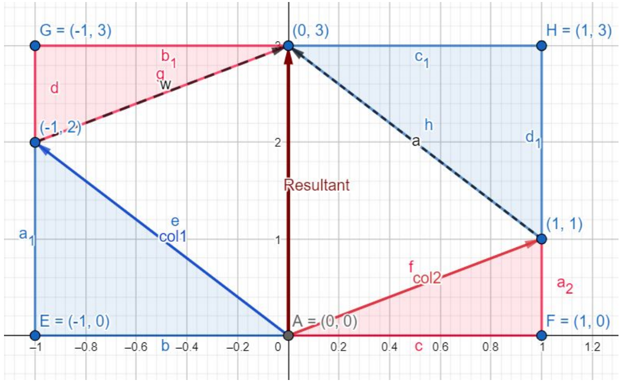
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad A^{-1} = \frac{1}{|A|}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \qquad AA^{-1} = A^{-1}A = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = (ad - bc)$$

$$-x + y = 0$$
$$2x + y = 3$$

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

$$|A| = -3$$

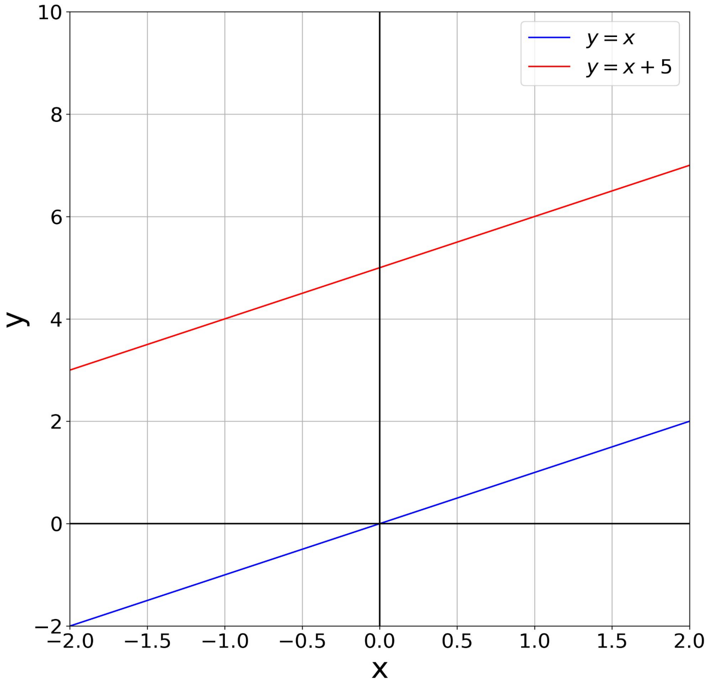


Two Equations and Two Unknowns-

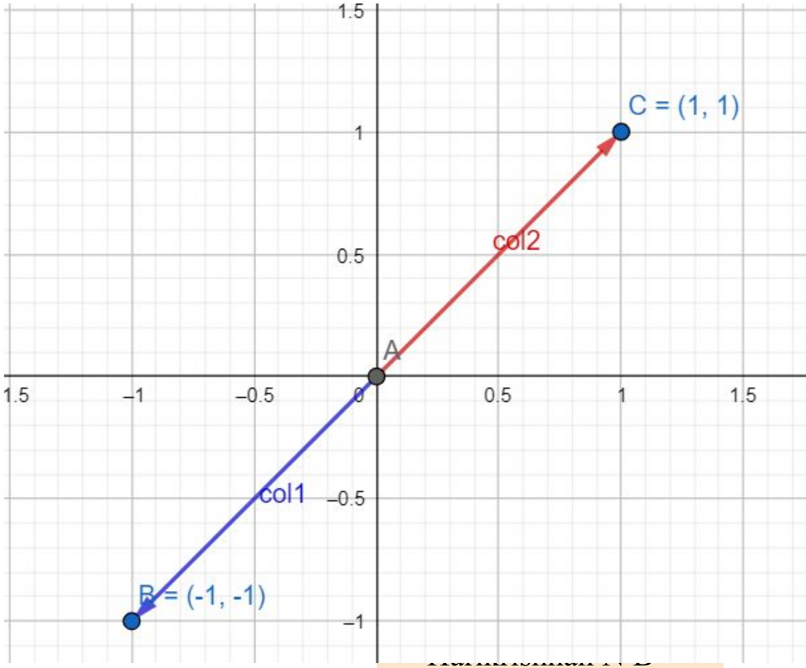
Contd..

$$-x + y = 0$$
$$-x + y = 5$$

Row Picture



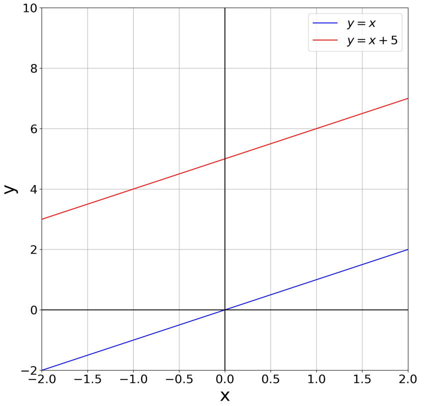
Column Picture



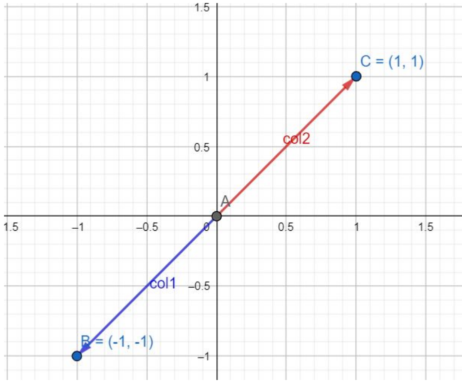
Permanent Breakdown of Elimination

$$\begin{aligned} -x + y &= 0 \\ -x + y &= 5 \end{aligned}$$

$$\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$$



$$\left[\begin{array}{cc|c} -1 & 1 & 0 \\ -1 & 1 & 5 \end{array}\right] \quad R_2 \to R_2 - R_1 \quad \left[\begin{array}{cc|c} -1 & 1 & 0 \\ 0 & 0 & 5 \end{array}\right]$$



$$0y = 5$$

Permanent Breakdown of elimination (NO SOLUTION)

Harikrishnan N B

Two Equations and Two Unknowns-
Invertibility

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad A^{-1} = \frac{1}{|A|}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \qquad AA^{-1} = A^{-1}A = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

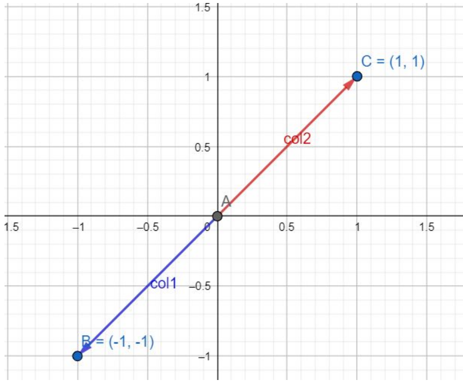
$$-x + y = 0$$
$$-x + y = 5$$

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = (ad - bc)$$

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

NOT INVERTIBLE

$$|A| = 0$$



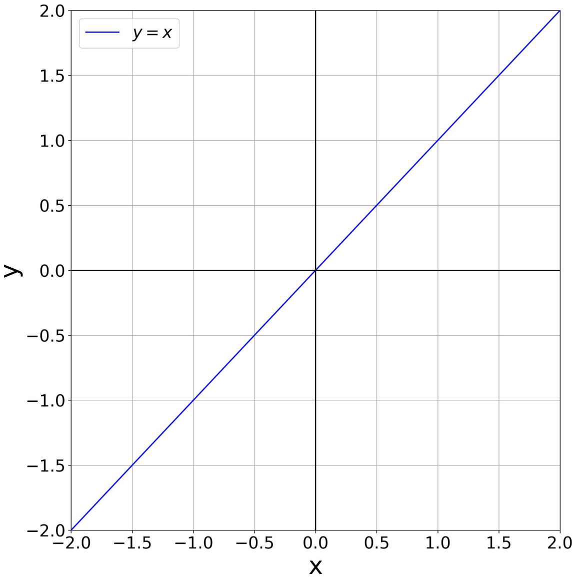
Harikrishnan N B

Two Equations and Two Unknowns- Contd..

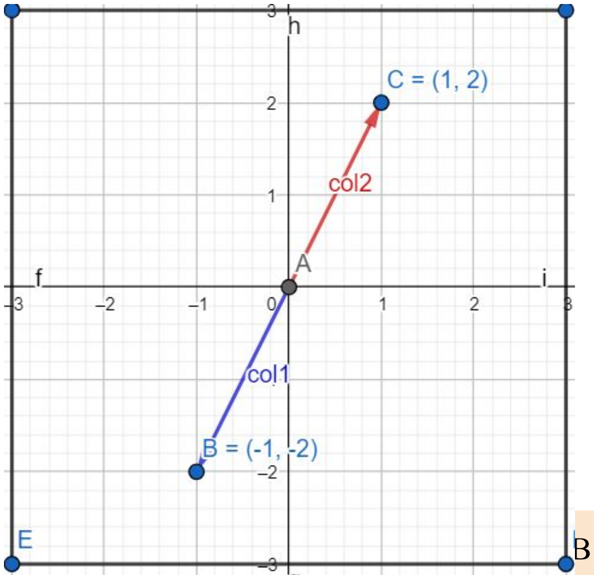
$$-x + y = 0$$
$$-2x + 2y = 0$$

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

Row Picture



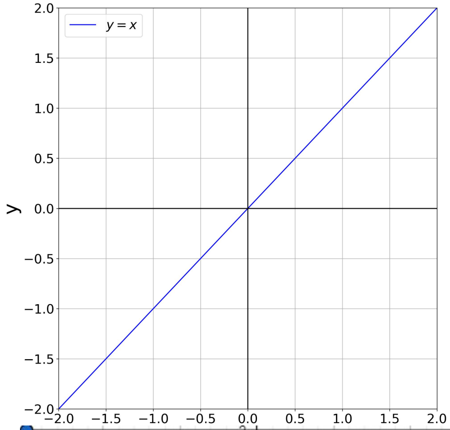
Column Picture



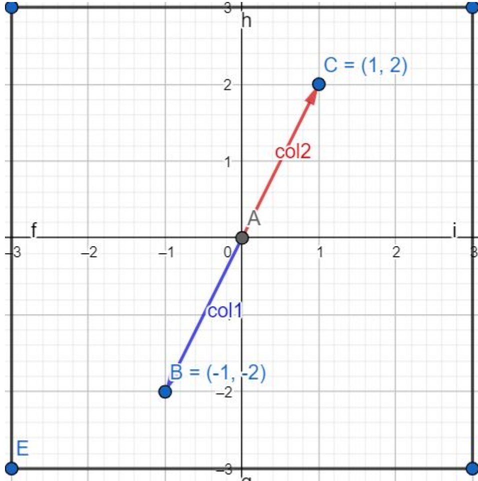
Temporary Breakdown of Elimination

$$-x + y = 0$$
$$-2x + 2y = 0$$

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



$$\left[\begin{array}{cc|c} -1 & 1 & 0 \\ -2 & 2 & 0 \end{array}\right] R_2 \to R_2 - R_1 \left[\begin{array}{cc|c} -1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$



$$0y = 0$$

***y* can take any value (Infinitely many solutions)**

Two Equations and Two Unknowns-
Invertibility

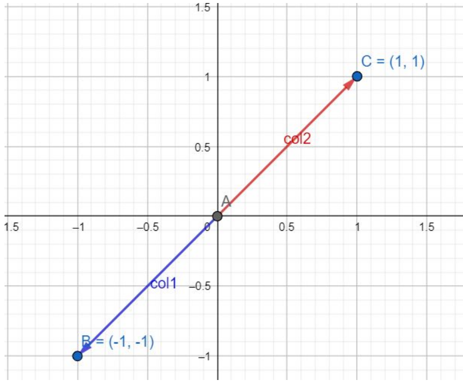
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad A^{-1} = \frac{1}{|A|}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \qquad AA^{-1} = A^{-1}A = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$-x + y = 0$$
$$-2x + 2y = 0$$

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = (ad - bc)$$

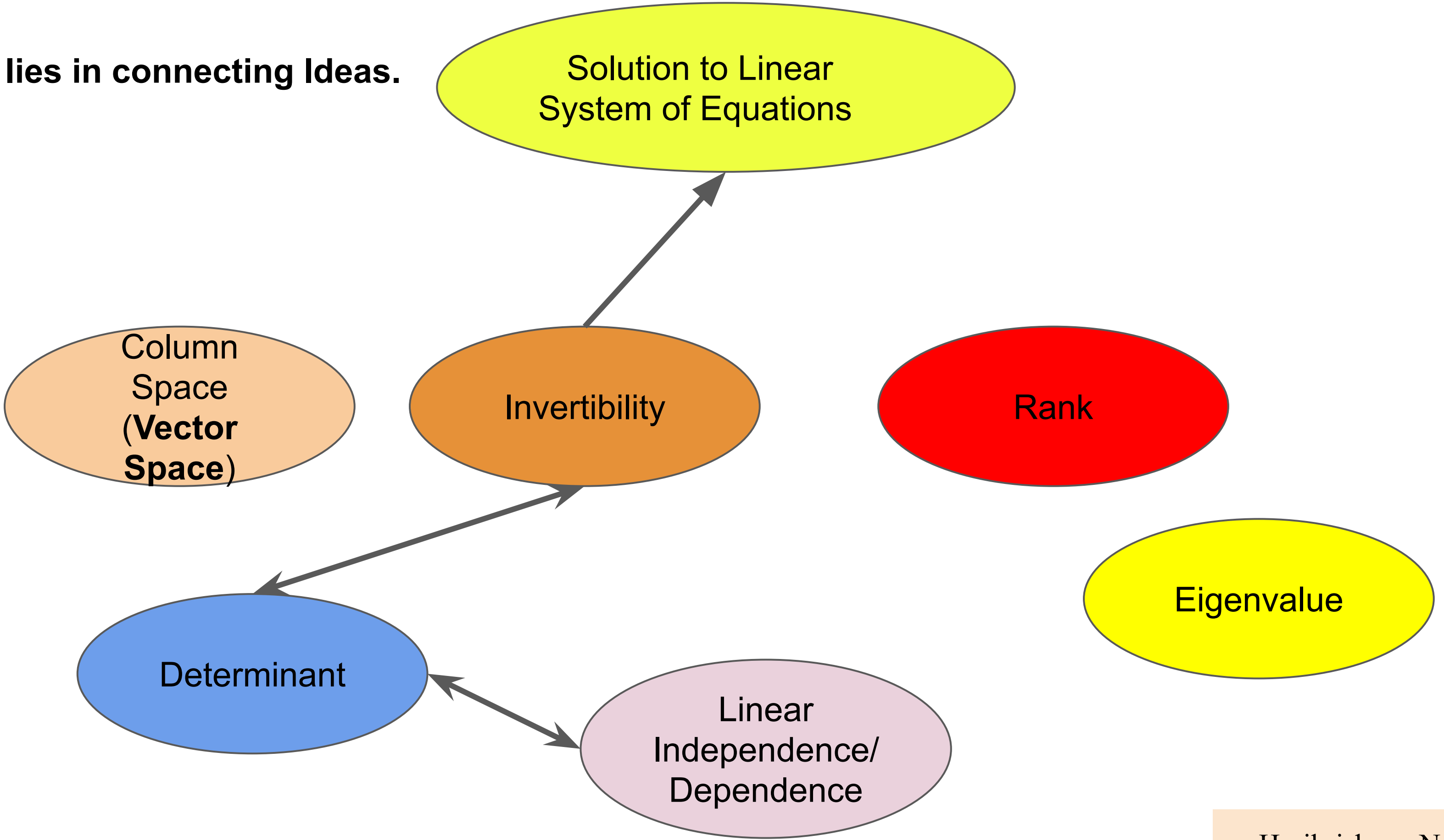
$$A = \begin{bmatrix} -1 & 1 \\ -2 & 2 \end{bmatrix}$$

NOT INVERTIBLE!!!



$$|A| = 0$$

Beauty lies in connecting Ideas.



Harikrishnan N B

Dot Product/ Matrix Multiplication/ Inverse

$$\vec{x}.\vec{y} = ||x||||y||\cos\theta = x^T y$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} row_1.col \\ row_2.col \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

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

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

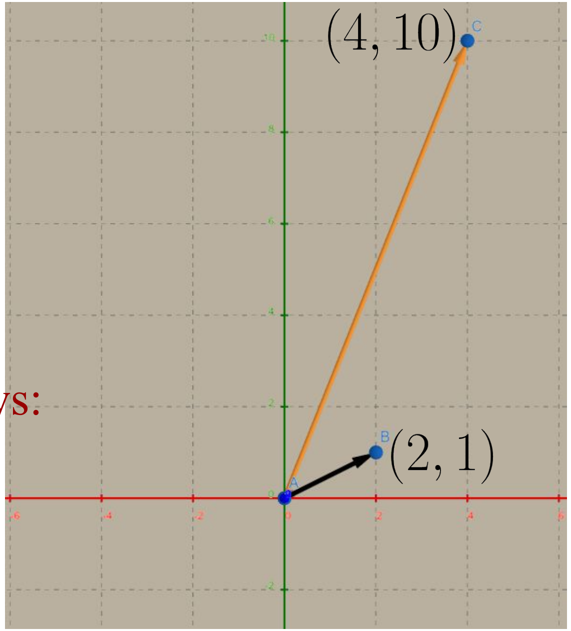
Orthogonal Matrix or Orthonormal Matrix

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

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

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

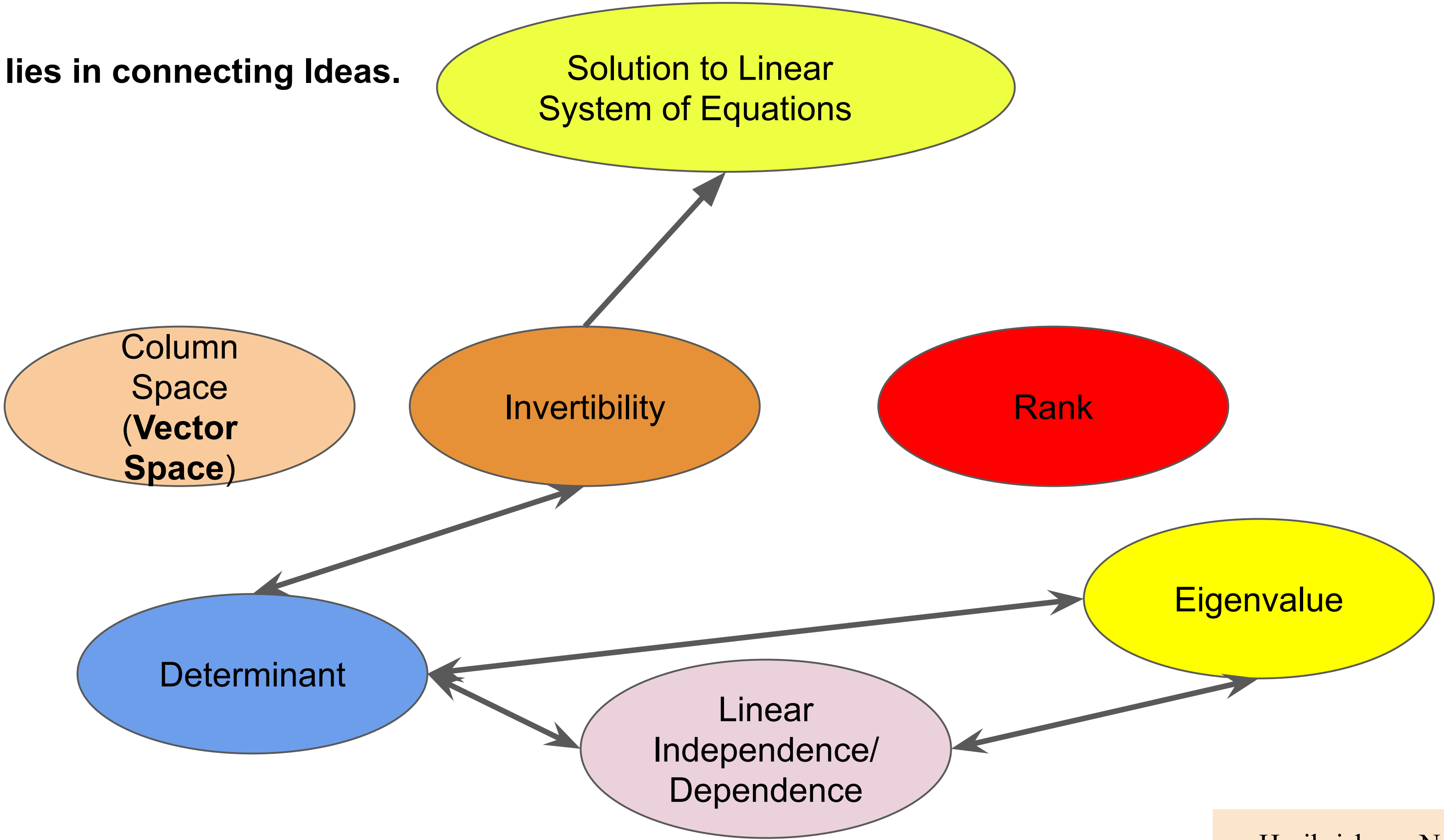
$$A\vec{x} \qquad \lambda\vec{x}$$

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

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**

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

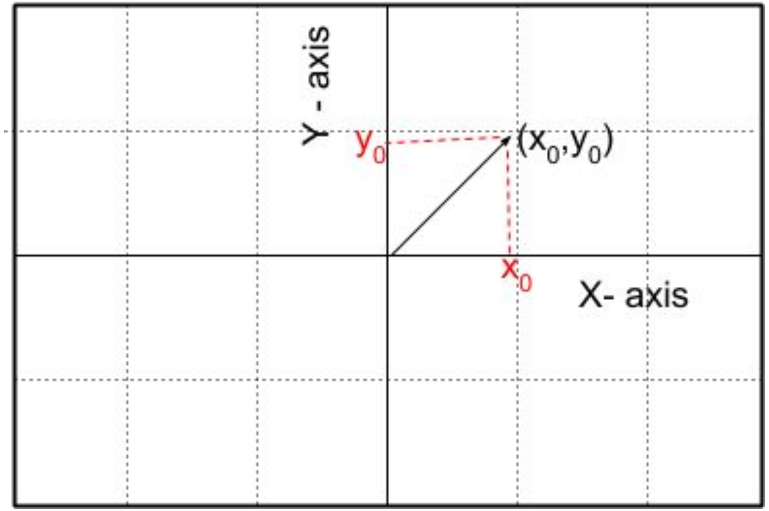
Beauty lies in connecting Ideas.



Second Iteration

Vectors - Different Understanding

Physicists



Computer Scientist

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

Mathematicians

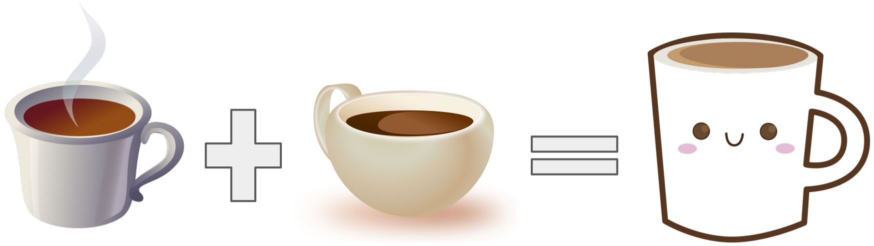
Vector space is a **collection of objects**(it can be anything) called vectors which satisfies mainly two important properties:

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

Vector Space - Coffee Space

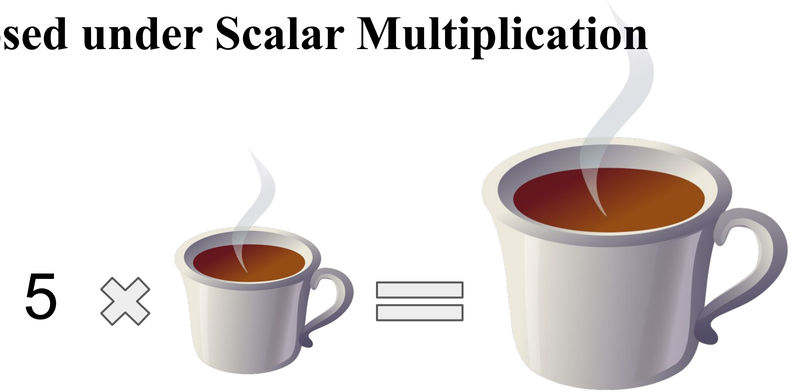
Coffee Space - In Coffee space we have different kinds of coffee with varying strength. Now we will understand the vector space properties with this metaphor.

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

Vector Space

- A real vector space is a set/collection of "*vectors*" together with the rules for vector addition and multiplication by real numbers.*

*Strang, Gilbert. *Linear Algebra and Its Applications*. Cengage Learning, 2017.

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.

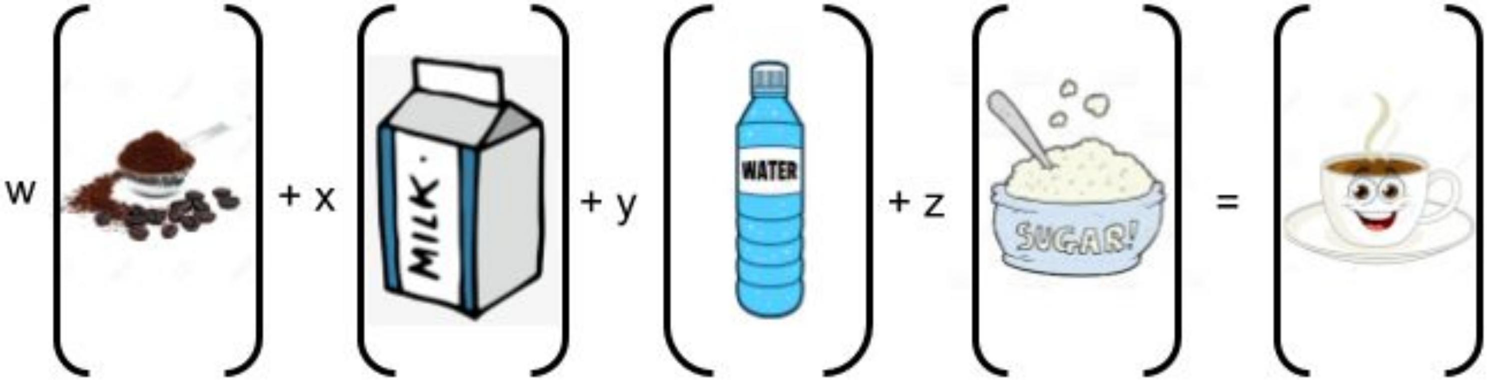
Properties of Basis Vectors -

- 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

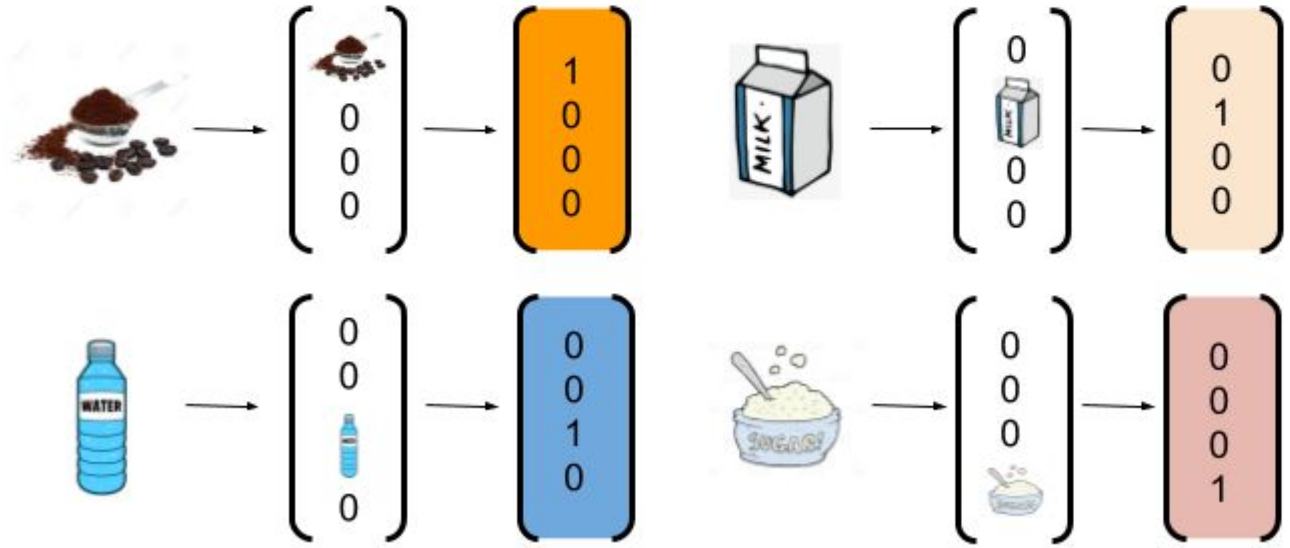
My Friend's Horrible Coffee

My Friend’s Horrible Coffee

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

Visualizing Coffee Space Basis Vectors

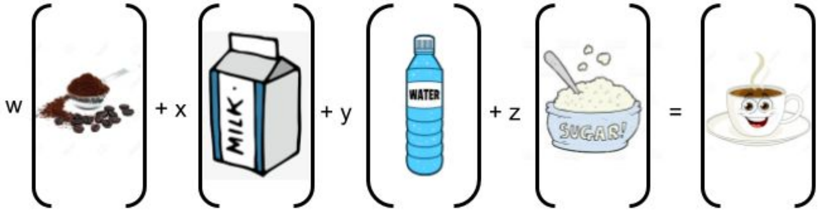
Coffee Space- Basis vectors



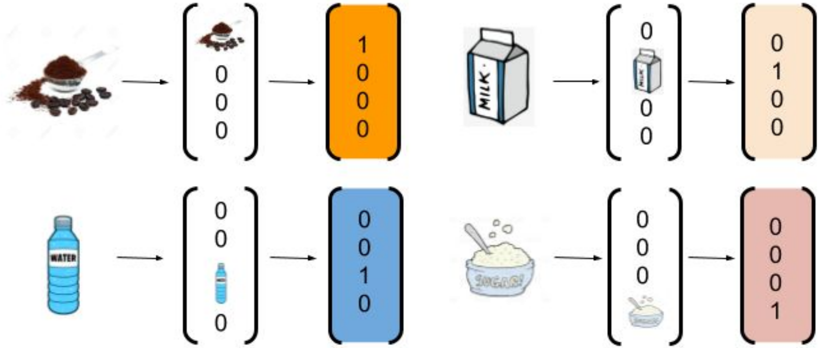
Harikrishnan N B

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

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

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)

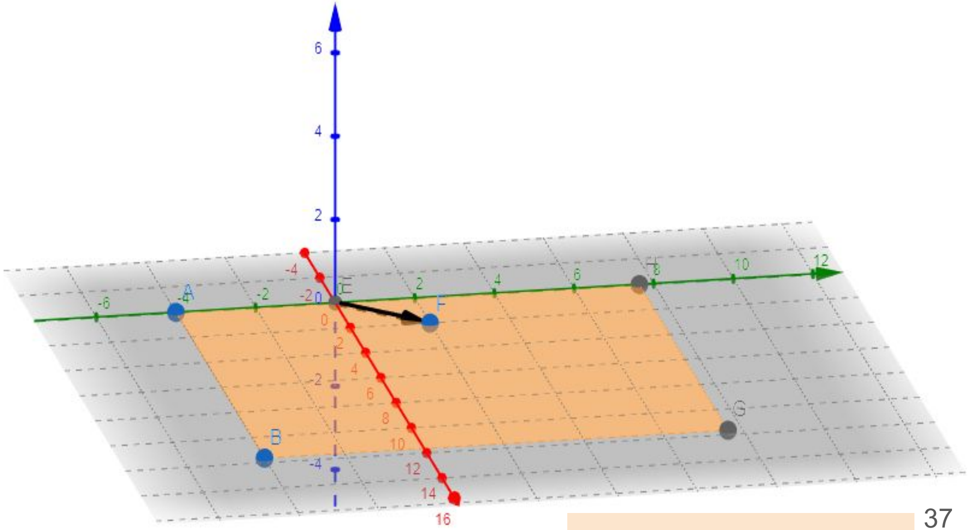
Why $span\left\{ \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix} \right\}$ can represent any point in R^4 ?

Can you see the Column Space?

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

C(A)

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



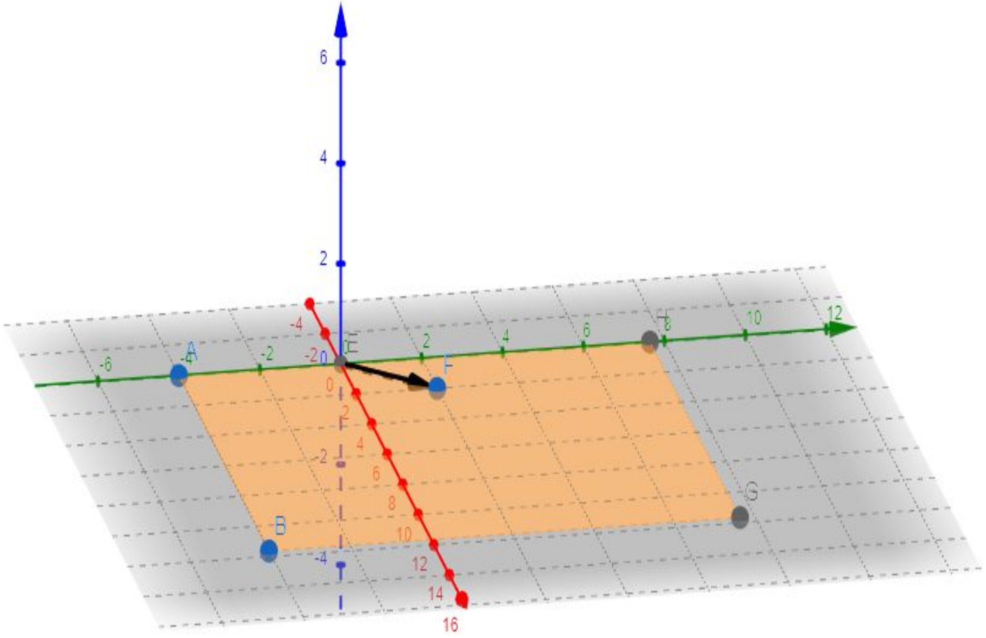
Harikrishnan N B

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?

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

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

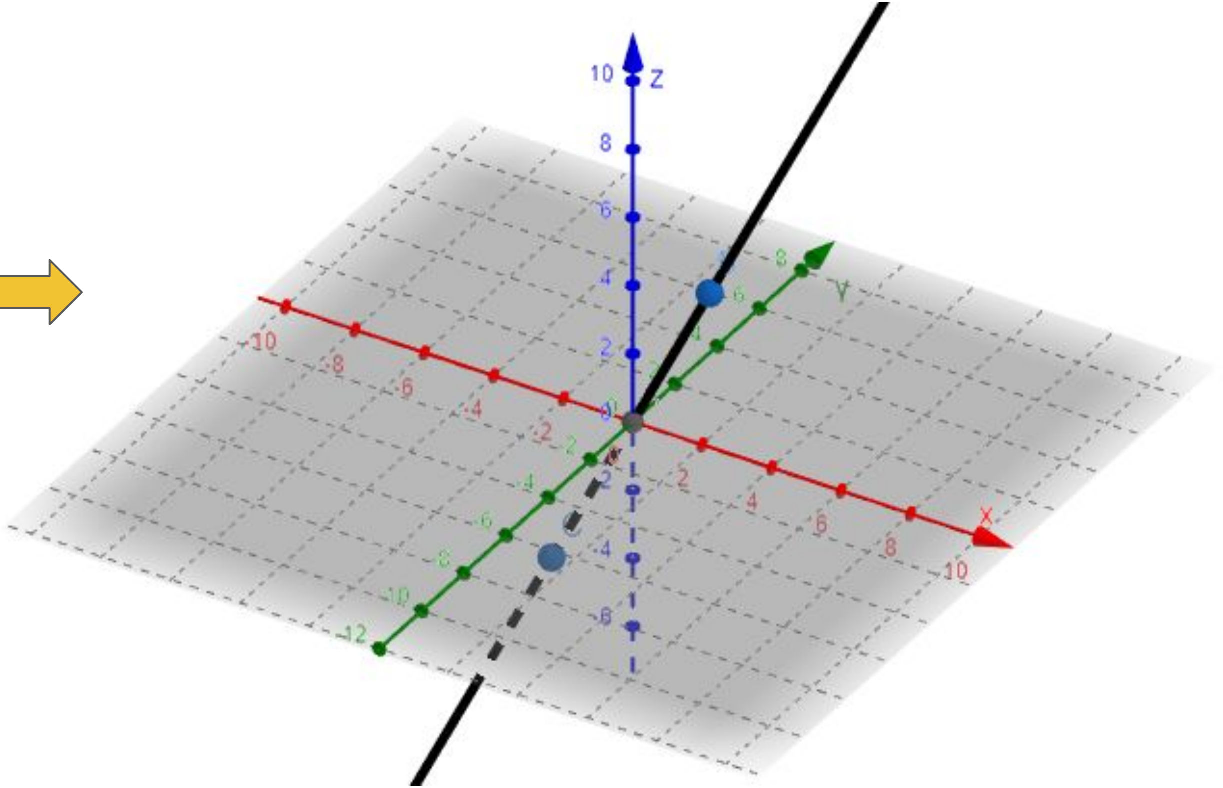
Column Space of Matrix A - Column space of matrix A denoted as C(A) is the space spanned by the column vectors of A.

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

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

Do you see a Subspace ?

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



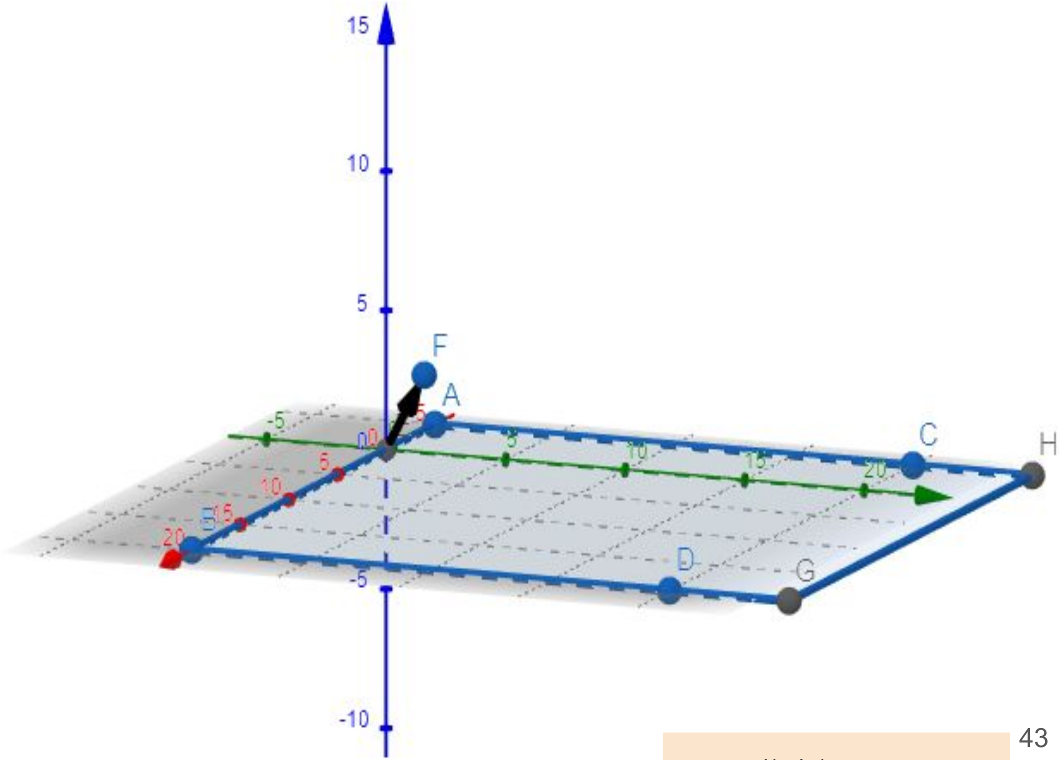
Is there anything Mysterious ?

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

Is there anything Mysterious ?

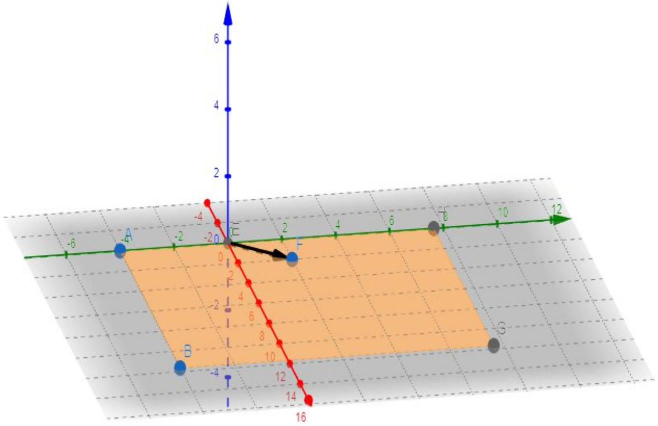
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} w \\ x \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

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



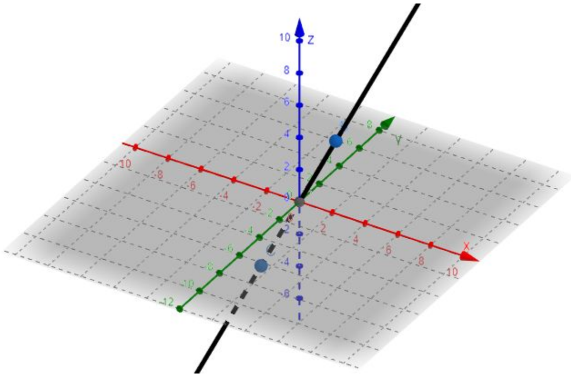
Solution to Ax = b

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



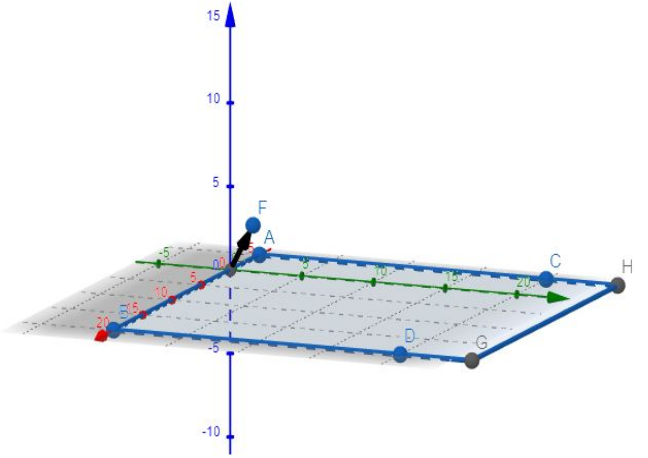
UNIQUE SOLUTION

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



INFINITELY MANY SOLUTIONS

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

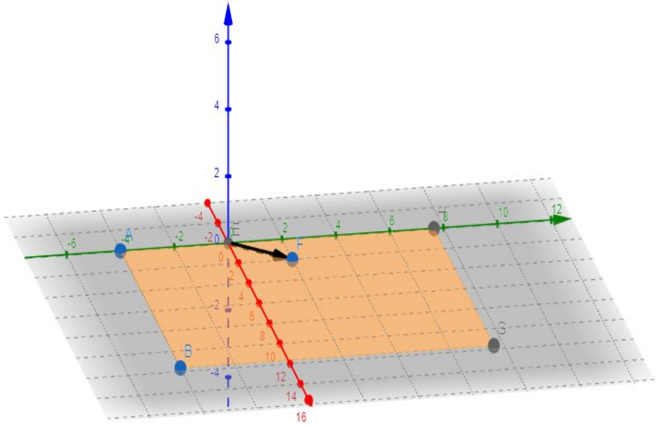


NO SOLUTION

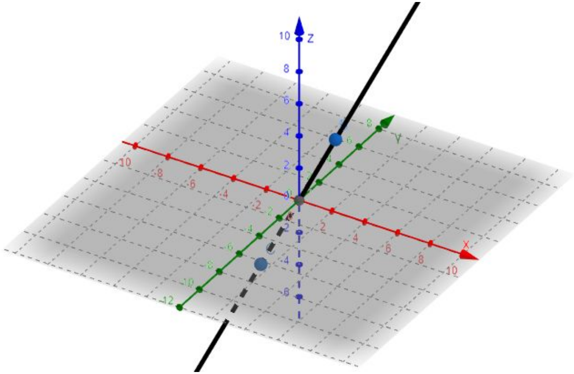
Harikrishnan N B

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

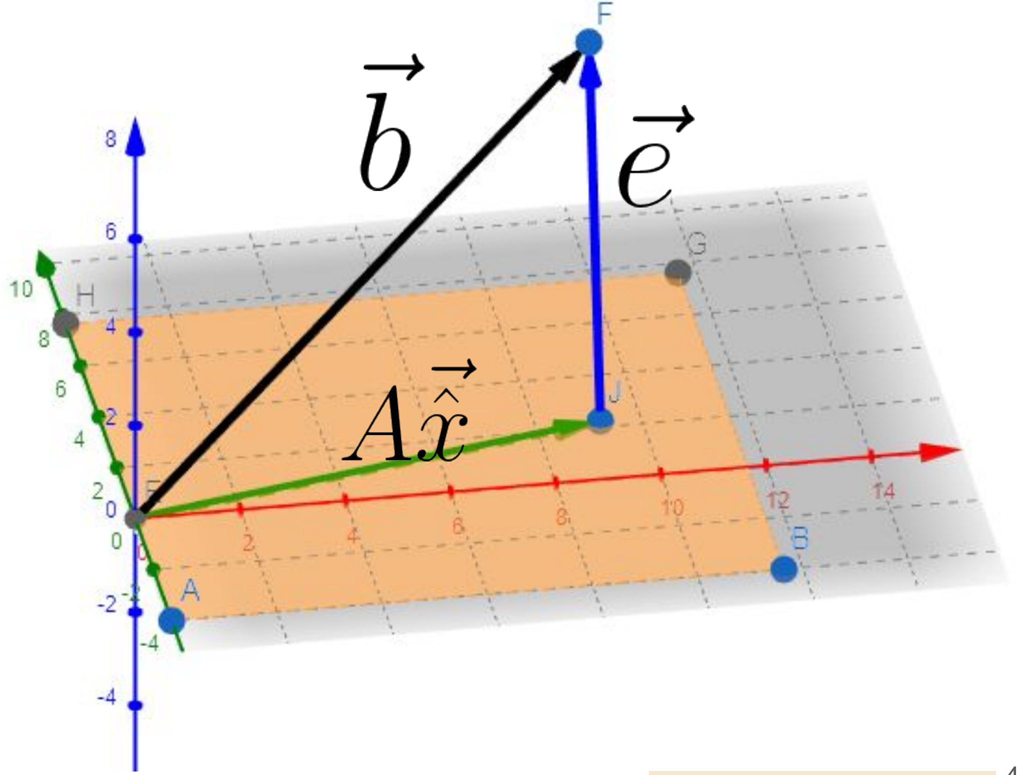
- For unique solution and infinitely many solutions b lies in the column space of A.
- In the case of NO solution b does not lie in the column space of A.

NO SOLUTION CASE 🙁

Can I find the best approximate solution ?

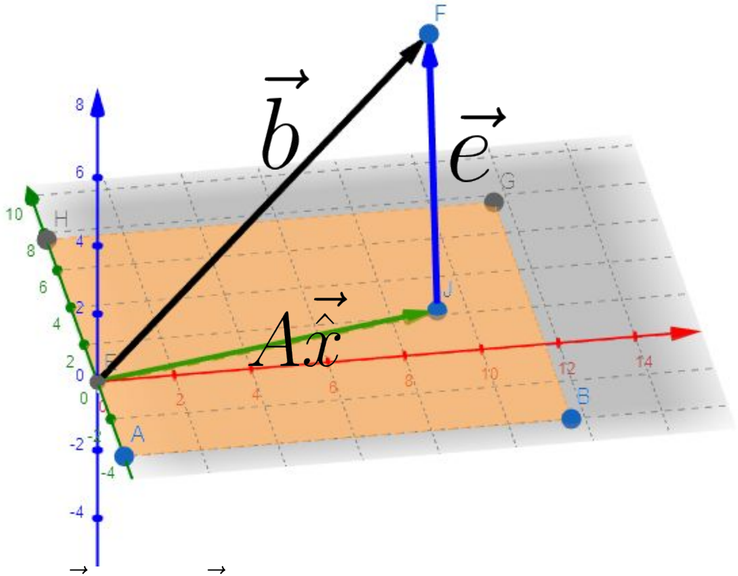
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w \\ x \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$A\vec{x} \neq \vec{b}$$



Harikrishnan N B

No solution case - Visualization



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

$$\begin{bmatrix} b_1 & b_2 \end{bmatrix} \left(\vec{\hat{x}}\right) + \left(\vec{e}\right) = \left(\vec{b}\right)$$

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

$$\begin{bmatrix} b_1^T \\ b_2^T \end{bmatrix} \left(\vec{e}\right) = \left(\vec{0}\right)$$

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

ORTHOGONAL

NO SOLUTION CASE

Can I find the best approximate solution ?

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w \\ x \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad A\vec{x} \neq \vec{b}$$

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

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

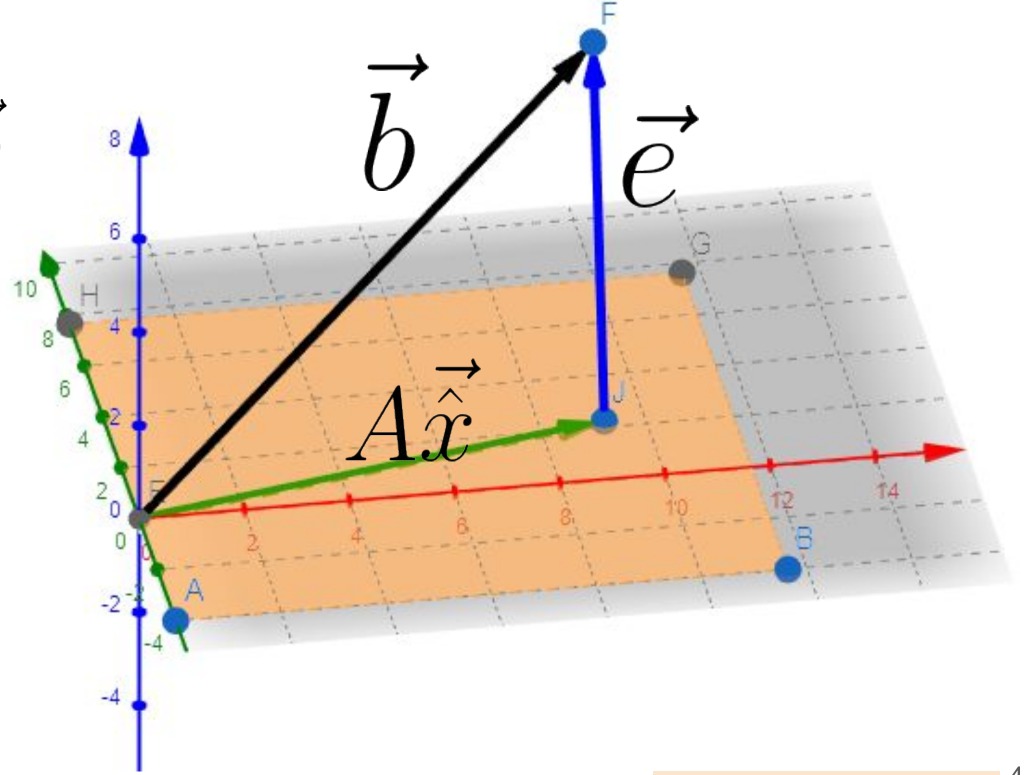
$$A^T\vec{e} = \vec{0}$$

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

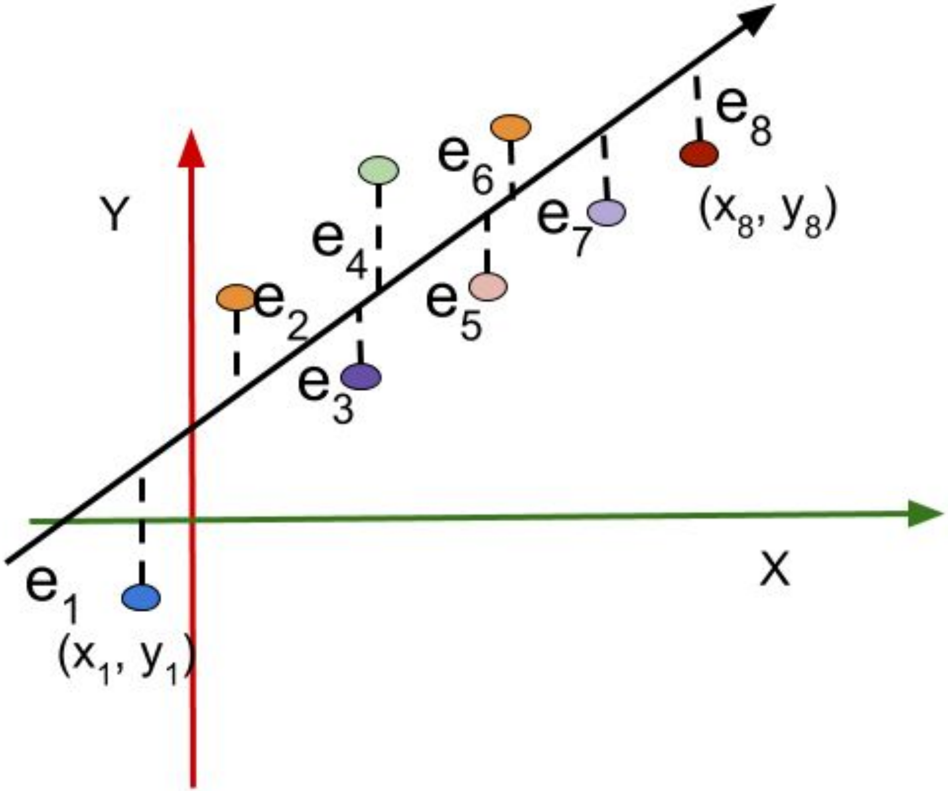
$$A^T\vec{b} - A^TA\vec{\hat{x}} = \vec{0}$$

$$A^TA\vec{\hat{x}} = A^T\vec{b}$$

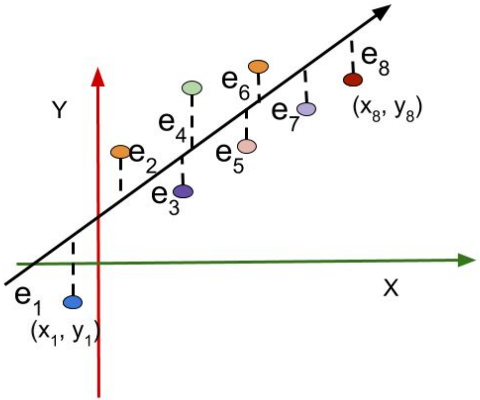
$$\vec{x} = (A^TA)^{-1}A^T\vec{b}$$



Linear Least Square Regression



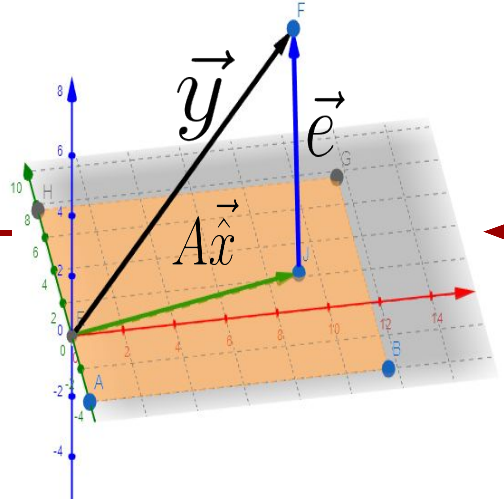
$$y_1 = mx_1 + c + e_1$$
$$y_2 = mx_2 + c + e_2$$
$$y_3 = mx_3 + c + e_3$$
$$y_4 = mx_4 + c + e_4$$
$$y_5 = mx_5 + c + e_5$$
$$y_6 = mx_6 + c + e_6$$
$$y_7 = mx_7 + c + e_7$$
$$y_8 = mx_8 + c + e_8$$



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

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

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



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

What happens in this case ? - Code it

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

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

What change you will observe in the graph?

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

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

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.

Applications of Least Squares in Signal Processing

- Linear/ Non-linear Prediction
- Denoising
- Deconvolution
- System Identification
- Estimating Missing Data

Link to Ivan Selesnick's Tutorial

LEAST SQUARES WITH EXAMPLES IN SIGNAL PROCESSING*

Ivan Selesnick

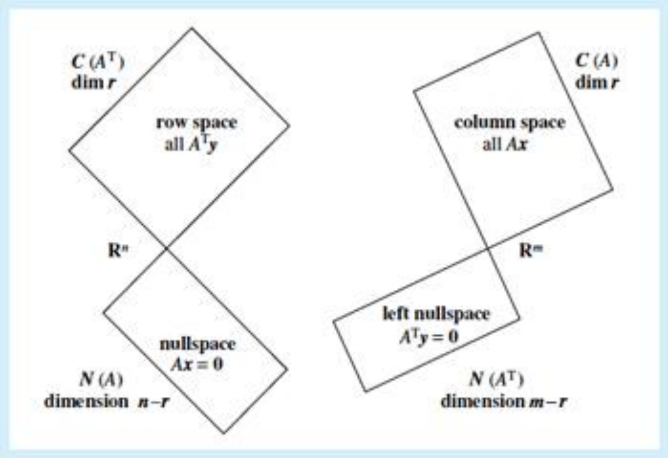
This work is produced by OpenStax-CNX and licensed under the Creative Commons Attribution License 3.0†

Four Fundamental Subspaces

- **Column Space**
- **Left Null Space**
- **Row Space**
- **Right Null Space**

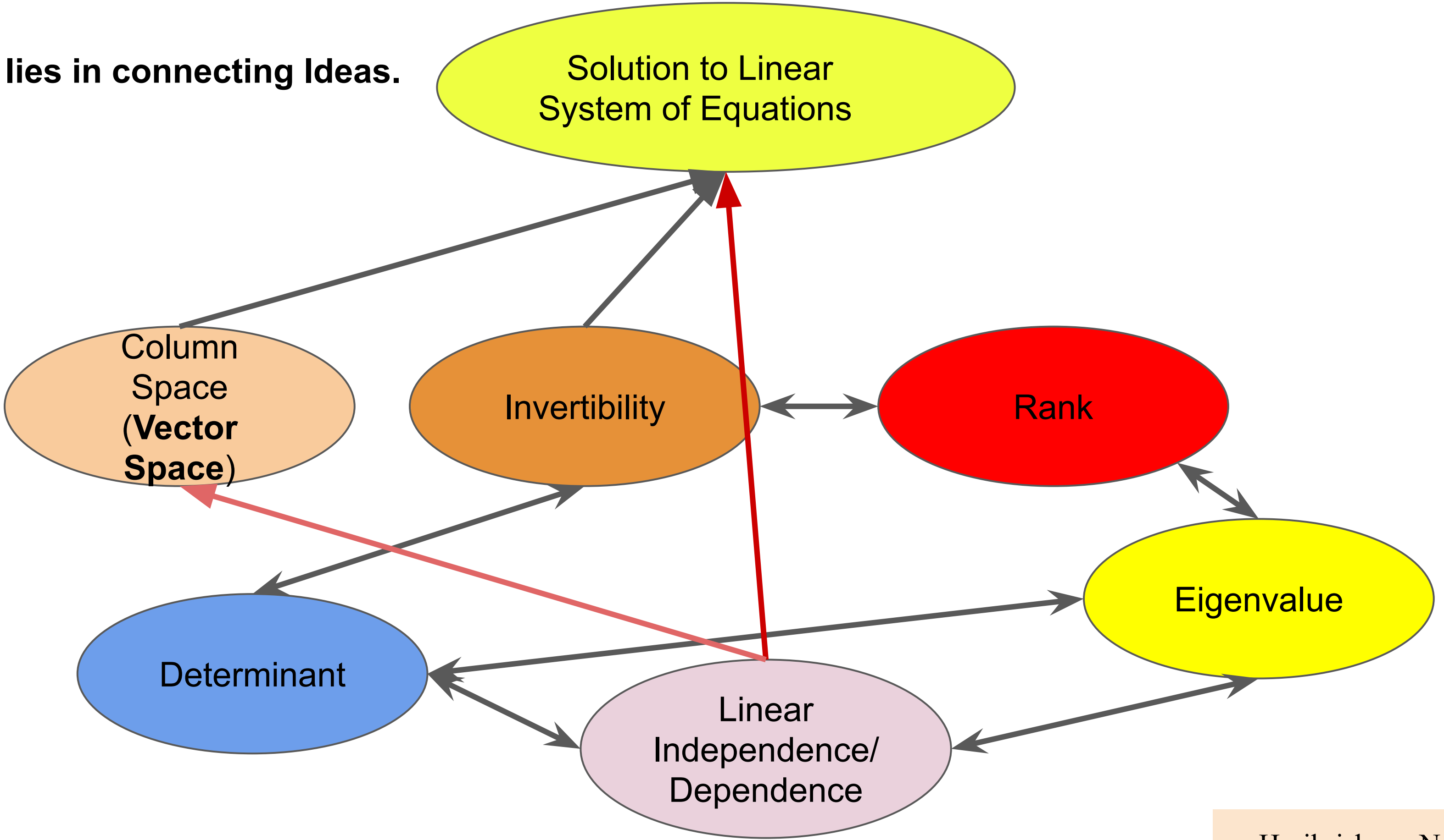
Fundamental Theorem of Linear Algebra

- Column space and Row space both have dimension r (rank).
- The Right Null Space have dimension *n-r* and the left null space has dimension *m-r*.
- Right Null Space is the orthogonal complement of the row space.
- Left Null Space is the orthogonal complement of the column space



Source:
https://ocw.aprende.org/courses/mathematics/18-06sc-linear-algebra-fall-2011/ax-b-and-the-four-subspaces/

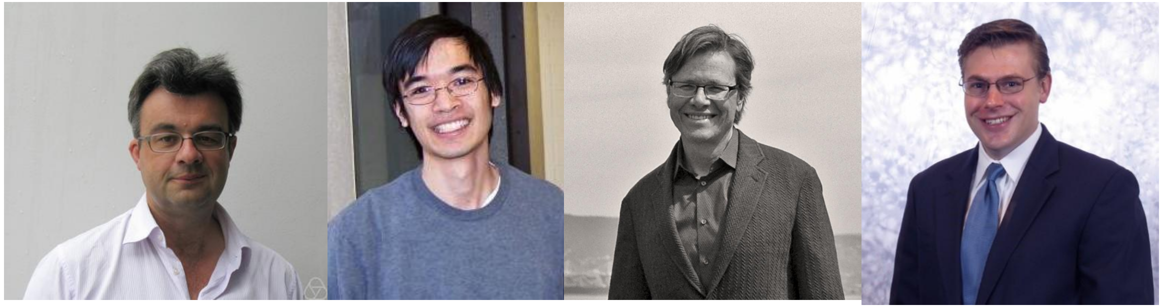
Beauty lies in connecting Ideas.

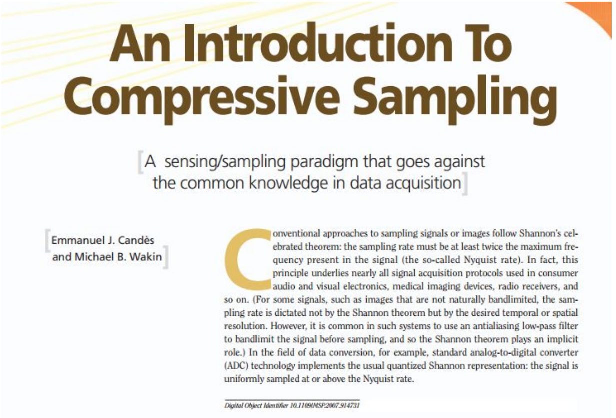


Harikrishnan N B

Do we know everything about $y=Ax$?

Emmanuel Candès   Terence Tao   David Donoho   Justin Romberg





An Introduction To Compressive Sampling

[A sensing/sampling paradigm that goes against the common knowledge in data acquisition]

[Emmanuel J. Candès and Michael B. Wakin]

Conventional approaches to sampling signals or images follow Shannon's celebrated theorem: the sampling rate must be at least twice the maximum frequency present in the signal (the so-called Nyquist rate). In fact, this principle underlies nearly all signal acquisition protocols used in consumer audio and visual electronics, medical imaging devices, radio receivers, and so on. (For some signals, such as images that are not naturally bandlimited, the sampling rate is dictated not by the Shannon theorem but by the desired temporal or spatial resolution. However, it is common in such systems to use an antialiasing low-pass filter to bandlimit the signal before sampling, and so the Shannon theorem plays an implicit role.) In the field of data conversion, for example, standard analog-to-digital converter (ADC) technology implements the usual quantized Shannon representation: the signal is uniformly sampled at or above the Nyquist rate.

Digital Object Identifier 10.1109/MSP.2007.914731

Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?

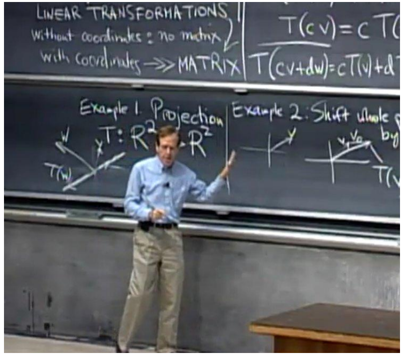
Emmanuel Candes† and Terence Tao♯

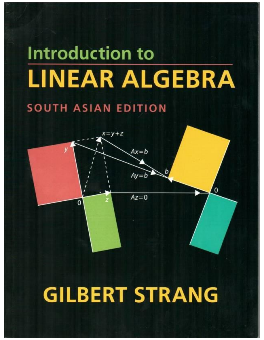
† Applied and Computational Mathematics, Caltech, Pasadena, CA 91125

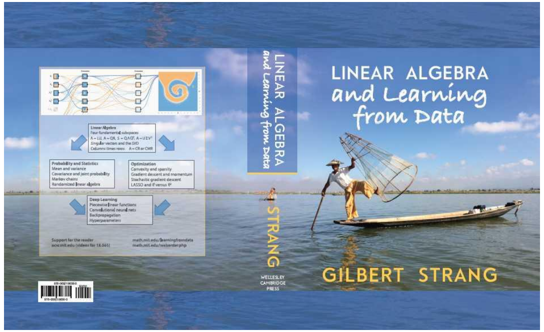
♯ Department of Mathematics, University of California, Los Angeles, CA 90095

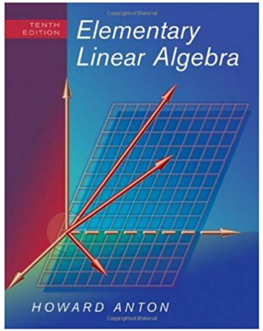
Interesting Materials

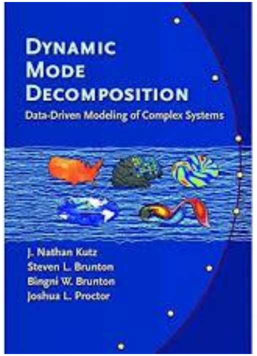
Prof. Gilbert Strang











Tutorial on PCA - [(Click here)]

MAY YOU FIND THE
EIGENVECTOR OF
YOUR LIFE :)

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

Extras

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

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

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

Spectral Decomposition

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

$$AX = X\Lambda$$
$$AXX^{-1} = X\Lambda X^{-1}$$
$$AI = X\Lambda X^{-1}$$
$$A = X\Lambda X^{-1}$$

Harikrishnan N B

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

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

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

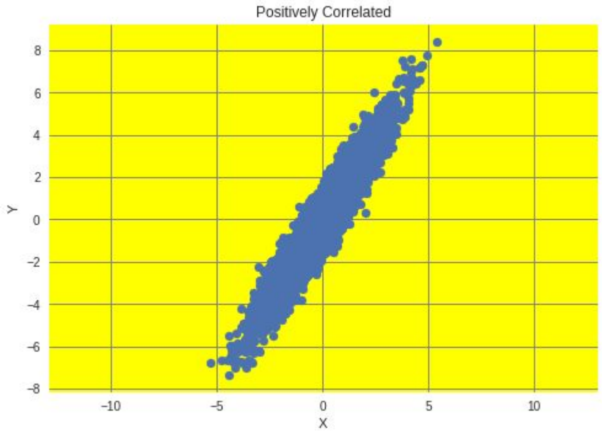
Recall the properties of a symmetric matrix!!!

- **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?

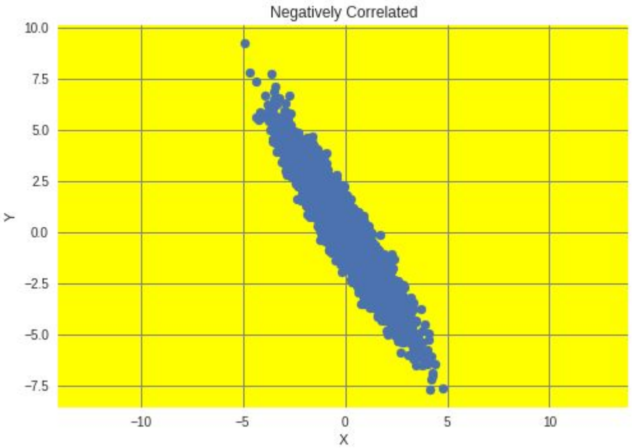
Case I

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



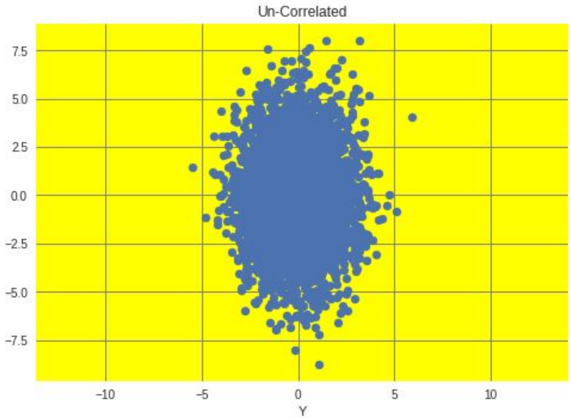
Case II

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



Case III

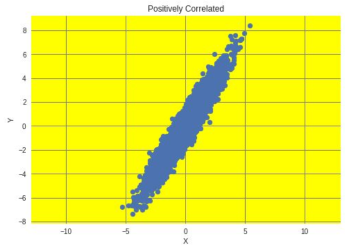
$$\begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}$$

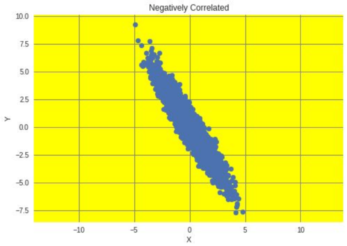


Note: In all cases mean is (0,0)

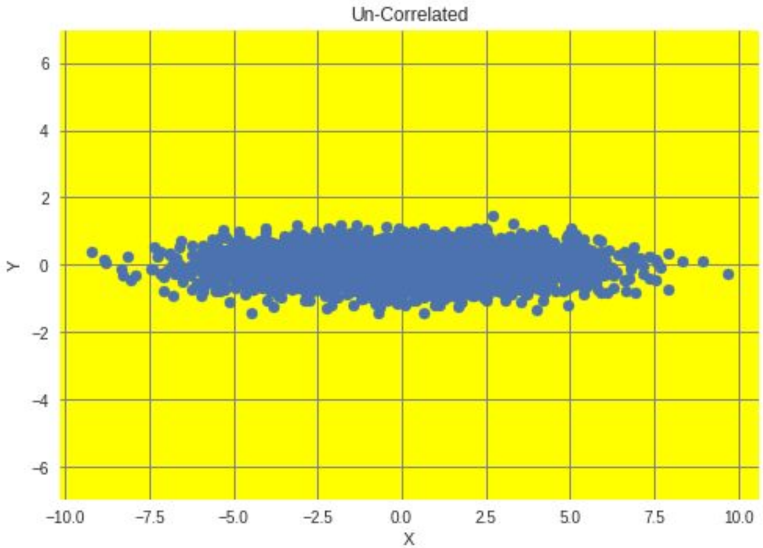
So what does PCA do ?

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





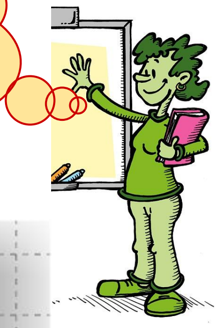
PCA achieves this by **Change of Basis**

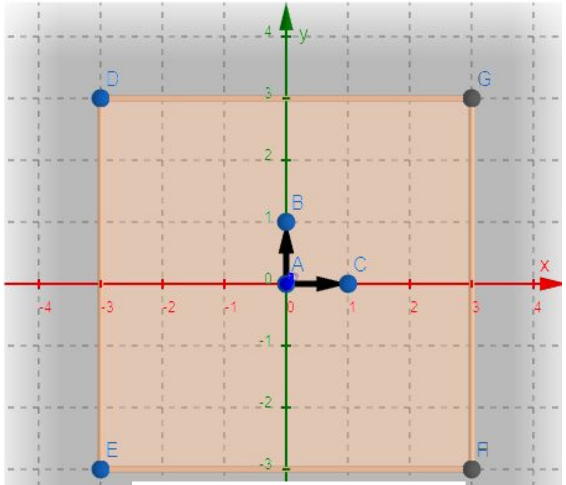


Change of Basis

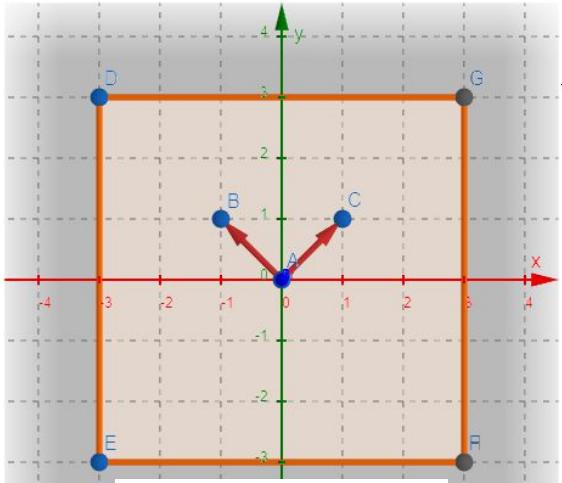
My life is full of struggle.

IT depends on the way you look at life. Try to see life from a different **POINT OF VIEW**





Old Basis



New Basis

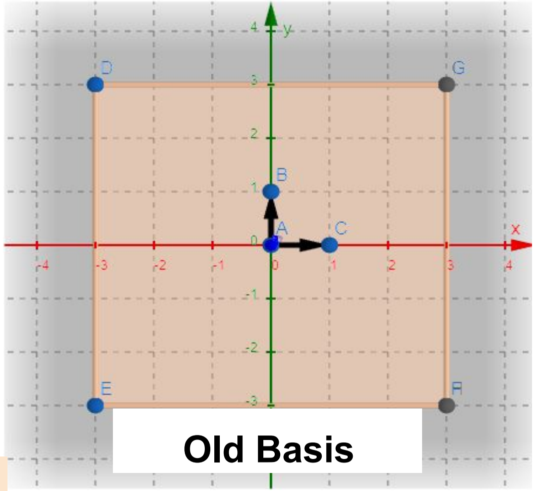
Harikrishnan N B

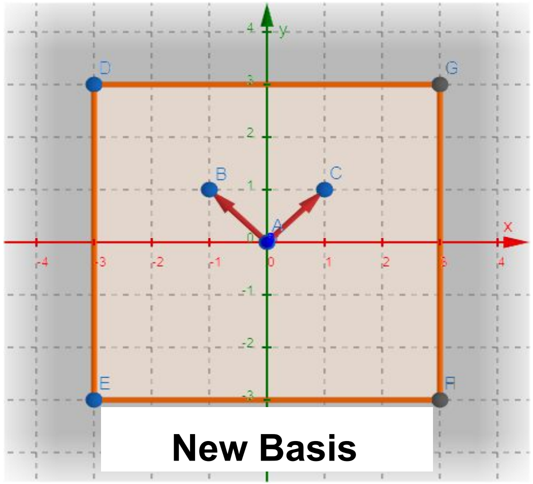
Recall

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 -

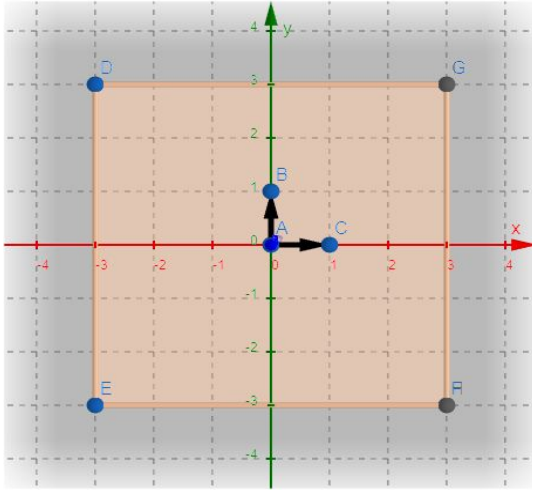
- Basis vectors has to be linearly independent.
- Basis vectors should span the vector space.





Example of Change of Basis

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

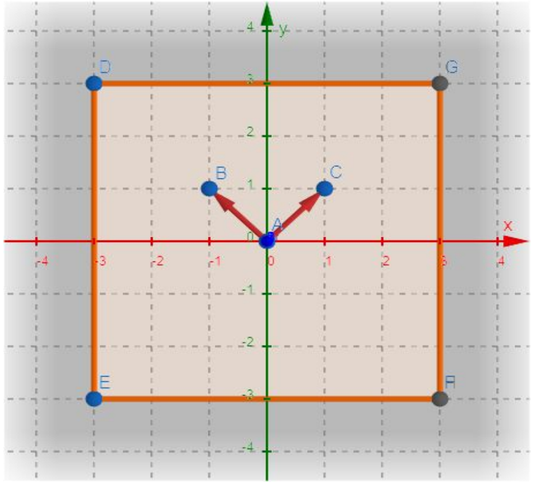


Old basis = $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$

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

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

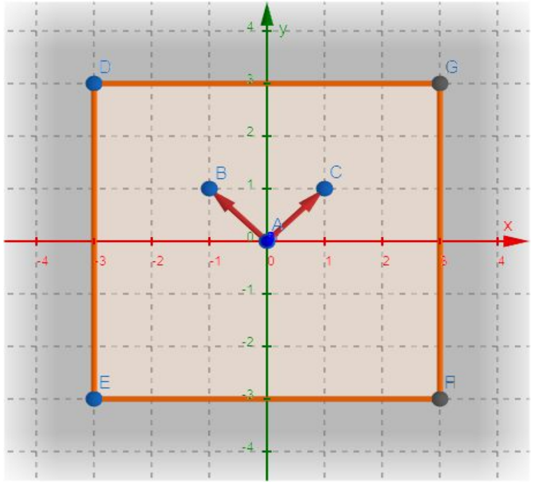
New Basis Representation



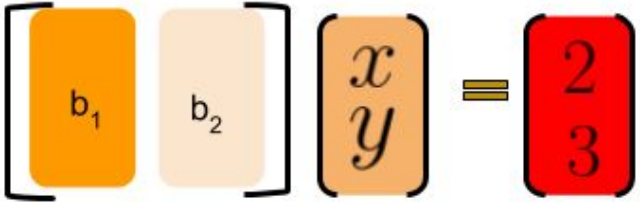
New Basis $= \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right\}$

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

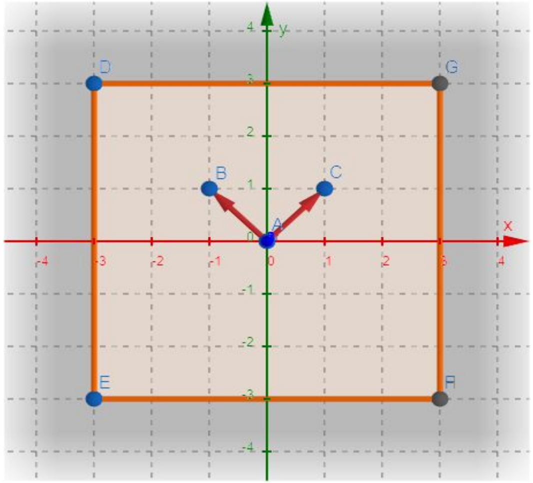
Finding x and y for representing (2,3) using new basis



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



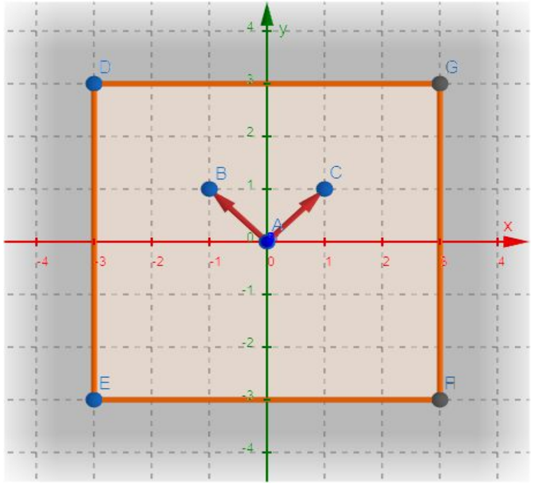
$$\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}=\begin{bmatrix}2\\ 3\end{bmatrix}$$



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

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

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



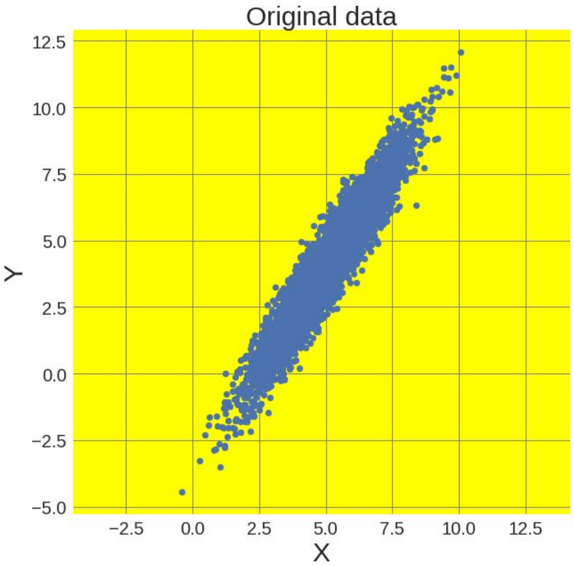
In our case the matrix P is ORTHONORMAL

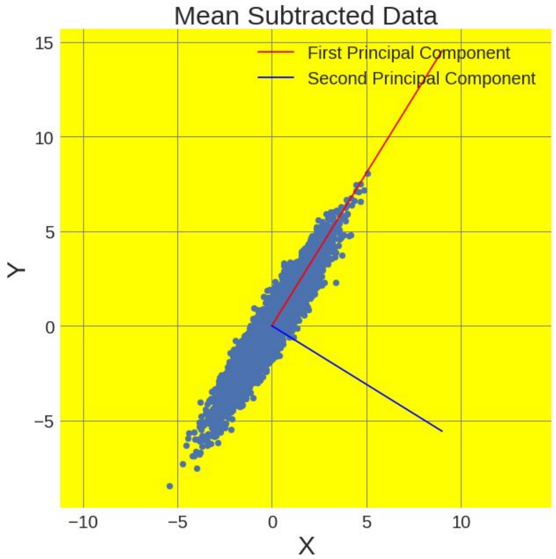
$$\vec{x} = P^{-1}\vec{y} = P^{T}\vec{y}$$

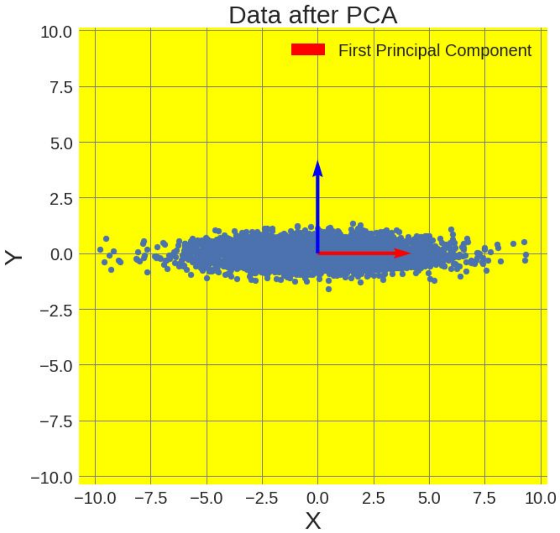
$$\begin{bmatrix} b_1^T \\ b_2^T \end{bmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \quad \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{5}{\sqrt{2}}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

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

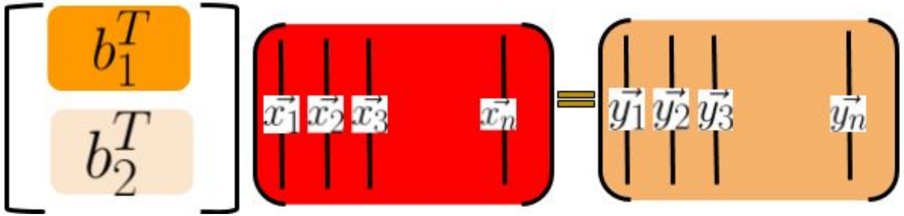
Steps in PCA





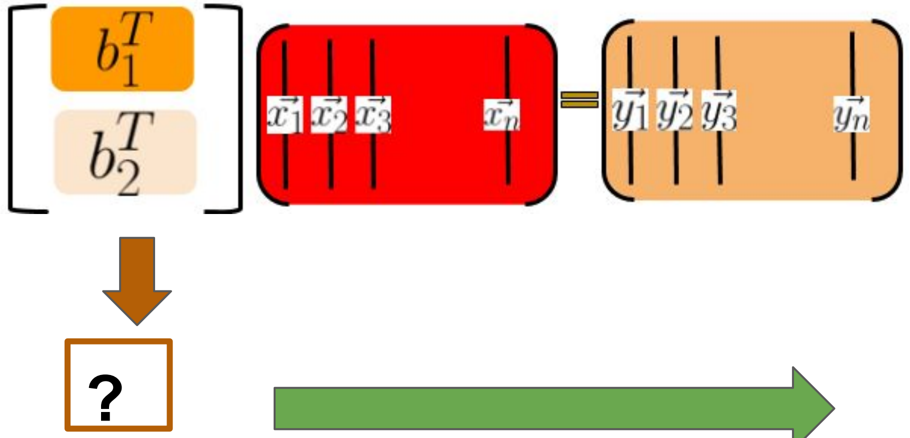


Change of Basis



$$\begin{bmatrix} b_1^T \\ b_2^T \end{bmatrix} \begin{bmatrix} \vec{x_1} & \vec{x_2} & \vec{x_3} & \cdots & \vec{x_n} \end{bmatrix} = \begin{bmatrix} \vec{y_1} & \vec{y_2} & \vec{y_3} & \cdots & \vec{y_n} \end{bmatrix}$$

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



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

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

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

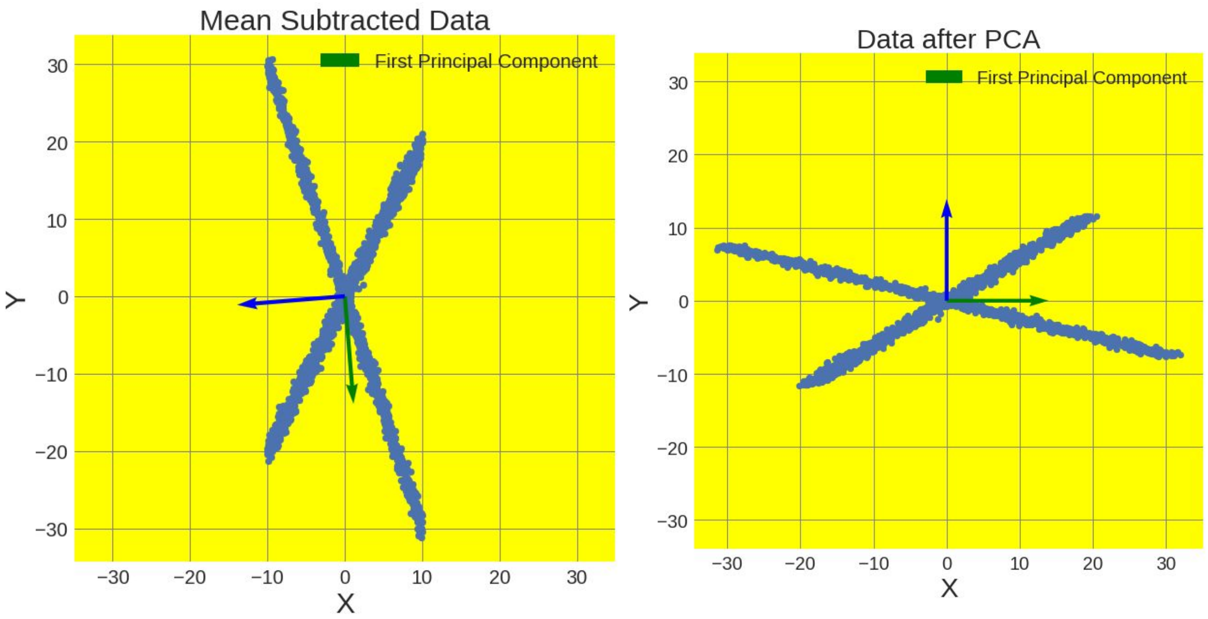
- Dimensionality Reduction
- Denoising
- Feature Extraction
- Image Compression
- EEG Analysis

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