NUS National University of Singapore | College of Humanities and Sciences

OPEN HOUSE
13TH MAY 2023

Math Masterclass

Linear Algebra and an Application

Jonathon Teo
jonathonteo@nus.edu.sg

Scan me now!

Welcome to NUS, CHS (College of Humanities and Sciences)

Introduction to Linear Algebra

- Why study Linear Algebra
- What is Linear Algebra
- Who is Linear Algebra

OPEN HOUSE

Introduction to Linear Algebra

- Why study Linear Algebra
- What is Linear Algebra
- ~~Who is Linear Algebra~~

Why study Linear Algebra

- Compulsory for Math, Data Science, Quantitative finance, Computer Science, and Engineering students.
- It is ubiquitous.
- Data are organized as matrices
  - Data analytics, machine learning, statistics, finance, economics
- Computer information and physical system are coded as vectors
  - This entire presentation slide is encoded as a (binary) vector
  - A 720 x 576 pixels coloured picture can be represented as 3 times 720 by 576 matrix, with entries from 1 to 256.

Why study Linear Algebra

Who are they?

Founders of Google.

- Left: Larry Page, net worth of $84 billion
- Right: Sergey Brin, net worth of $78 billion

Enrolled in Stanford University for their Ph.D. but did not graduate.

Developed the Google page-rank algorithm.

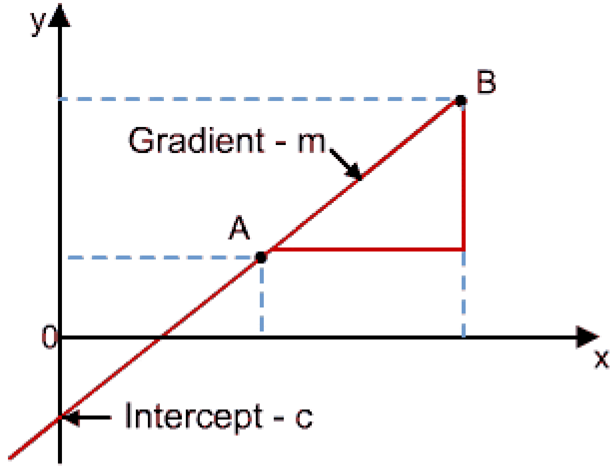
What is Linear Algebra

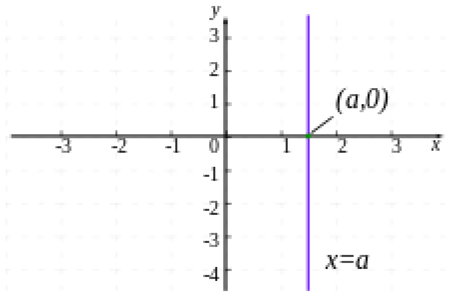
In the x,y-plane, a straight line can be expressed as

- $y = mx + c$
- $x = a$
- More generally

$$ax + by = c$$

- This is known as a linear equation with 2 variables.





What is Linear Algebra

- A linear equation with n variables (in standard form) can be written as

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

- $a_1, a_2, \dots, a_n$ are called the coefficients.
- b is known as the constant.
- May solve m linear equations simultaneously,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

- This is known as a linear system.

What is Linear Algebra

- Example: 2 equations, 2 variables

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

Provide a solution to the linear system

- $x = 1, y = 1$ is the only solution to the system.
- Example: 2 equations, 3 variables

$$x + y + z = 1$$
$$x - y + z = 2$$

- $y = -\frac{1}{2}, x = \frac{3}{2} - z$ for any real number z.

What is Linear Algebra

- Example: 5 equations, 5 variables

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 &= -1 \\ x_1 + 2x_2 + 3x_3 + x_4 + x_5 &= 0 \\ x_1 + x_2 + x_3 + x_4 + 2x_5 &= 0 \\ 2x_1 + 2x_2 + 2x_3 + x_4 + 3x_5 &= 0 \\ x_1 + 3x_2 + x_3 + x_4 + 3x_5 &= -1 \end{aligned}$$

Provide a solution to the linear system

- Solution: $x_1 = -1, x_2 = -1, x_3 = 1, x_4 = -1, x_5 = 1.$

How?

OPEN HOUSE

What is Linear Algebra

A linear system (in standard form)

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

can be converted to a matrix equation

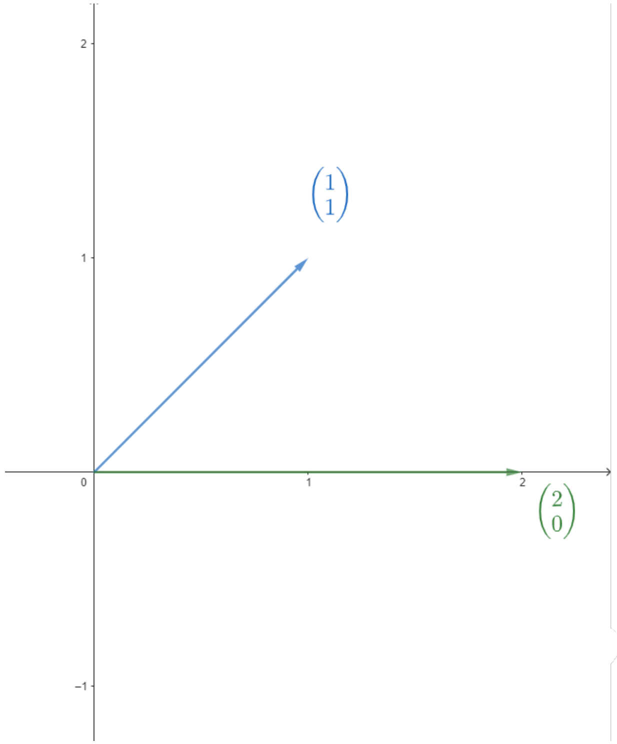
$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

Matrix coefficient Variable vector Constant vector

The is obtained via matrix multiplication.

What is Linear Algebra

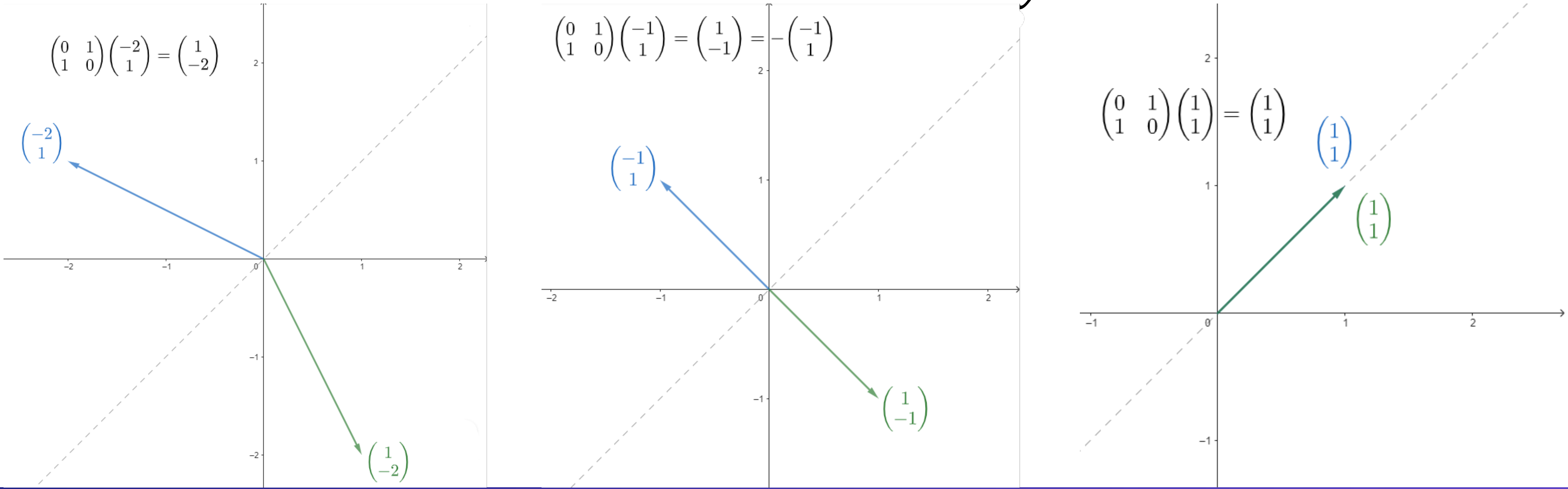
- A solution to a linear system can be represented as a vector
- Example: $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$
- Geometrical interpretation:



What is Linear Algebra

Example: $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

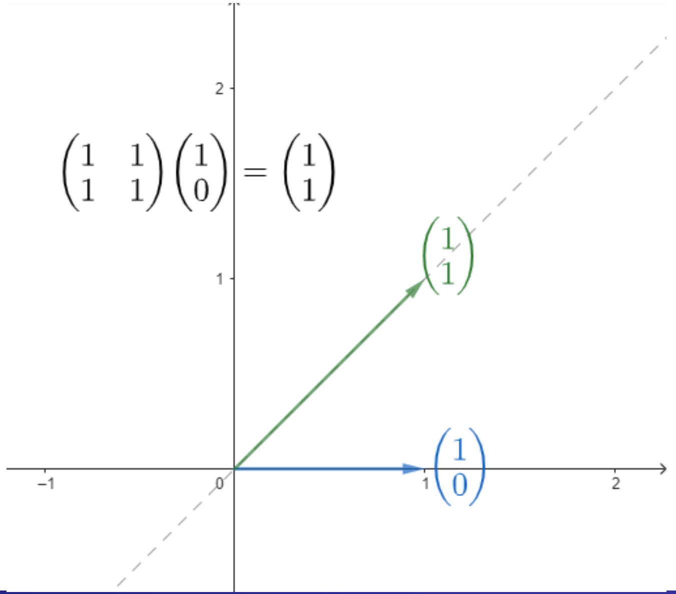
The matrix reflects a vector alone the line $x = y$.

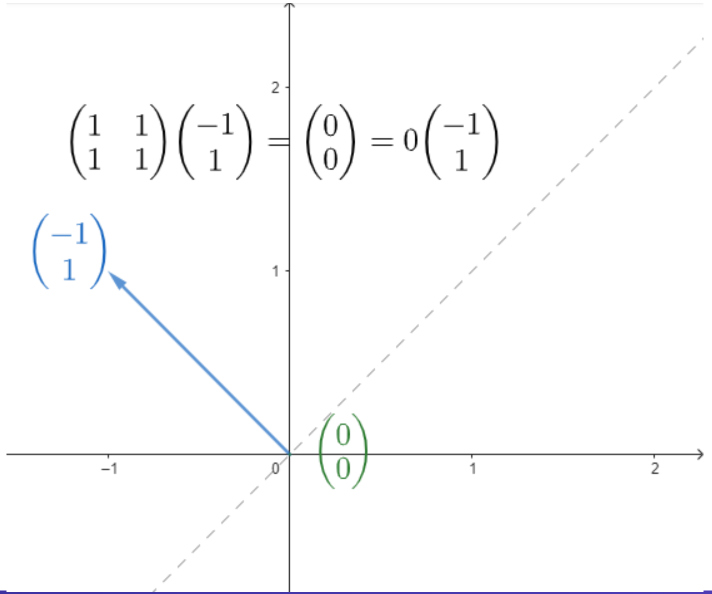


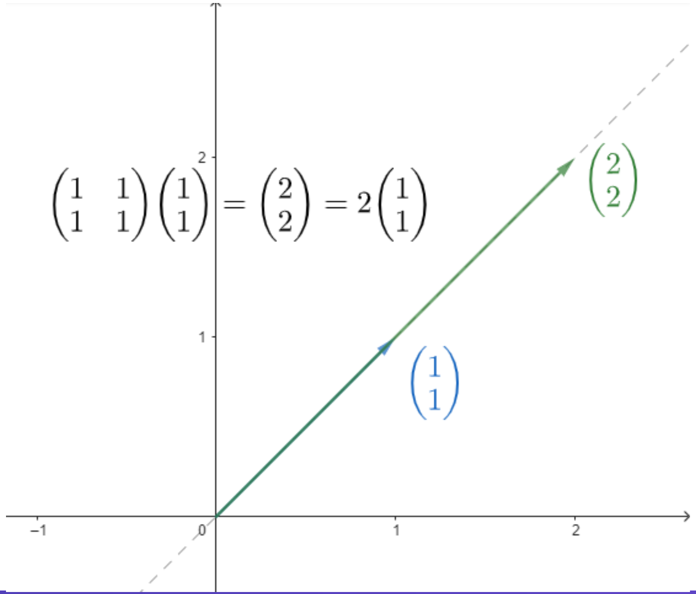
What is Linear Algebra

Example: $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

The matrix takes a vector and maps it to the line $x = y$ such that both coordinates are the sum of the original coordinates.







What is Linear Algebra

A vector $\boldsymbol{v}$ is an eigenvector of a matrix $\boldsymbol{A}$ if $A\boldsymbol{v} = \lambda\boldsymbol{v}$ for some real number λ.

Example:

- $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ is an eigenvector of $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = (-1)\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.
- $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector of $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = (1)\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.
- $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ is an eigenvector of $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.
- $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector of $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Google Page-rank Algorithm

- Assume a set S consisting of 4 of sites containing the key words. Name the sites s_1, s_2, s_3, s_4. Suppose
  - s_1 references s_2, s_3 and s_4.
  - s_2 references s_4 only.
  - s_3 references s_1 and s_4.
  - s_4 references s_1 and s_3.
- Define the adjacency matrix as such

$$\begin{array}{cccc} s_1 & s_2 & s_3 & s_4 \end{array}$$

$$\begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 1/2 & 0 \end{pmatrix} \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \end{array}$$

Google Page-rank Algorithm

- s_1 references s_2, s_3 and s_4.
- s_2 references s_4 only.
- s_3 references s_1 and s_4.
- s_4 references s_1 and s_3.

Suppose a person is at site 3, his state vector is represented by $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$.

His next state vector will be $\begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 0 \\ 0 \\ 1/2 \end{pmatrix}$.

This means that he will be either in site 1 or 4, with probability ½ each.

Google Page-rank Algorithm

If he continues surfing, the next state vector will be

$$\begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} 1/2 \\ 0 \\ 0 \\ 1/2 \end{pmatrix} = \begin{pmatrix} 1/4 \\ 1/6 \\ 5/12 \\ 1/6 \end{pmatrix}$$

This means that after 2 clicks, he will end up in either of the 4 sites, with the respective probabilities.

What happens if he keeps clicking randomly?

Google Page-rank Algorithm

$$1 \text{ click: } \begin{pmatrix} 1/2 \\ 0 \\ 0 \\ 1/2 \end{pmatrix}, \text{ 2 clicks: } \begin{pmatrix} 1/4 \\ 1/6 \\ 5/12 \\ 1/6 \end{pmatrix}, \text{ 5 clicks: } \begin{pmatrix} 0.2951 \\ 0.1042 \\ 0.2361 \\ 0.3646 \end{pmatrix}$$

$$7 \text{ clicks: } \begin{pmatrix} 0.3006 \\ 0.1001 \\ 0.2604 \\ 0.3388 \end{pmatrix}, \text{ 10 clicks: } \begin{pmatrix} 0.3000 \\ 0.1000 \\ 0.2675 \\ 0.3325 \end{pmatrix}, \text{ 12 clicks: } \begin{pmatrix} 0.3000 \\ 0.1000 \\ 0.2669 \\ 0.3331 \end{pmatrix}$$

$$15 \text{ clicks: } \begin{pmatrix} 0.3000 \\ 0.1000 \\ 0.2666 \\ 0.3335 \end{pmatrix}, \text{ 20 clicks: } \begin{pmatrix} 0.3000 \\ 0.1000 \\ 0.2667 \\ 0.3333 \end{pmatrix}, \text{ 100 clicks: } \begin{pmatrix} 0.3000 \\ 0.1000 \\ 0.2667 \\ 0.3333 \end{pmatrix}$$

Google Page-rank Algorithm

Observe that

- The state vectors are approximately the same after 20 clicks.
- In the long run, the resultant state vector is $\begin{pmatrix} 0.3000 \\ 0.1000 \\ 0.2667 \\ 0.3333 \end{pmatrix}$.
- This means that in the long run, he will most probably end up in site 4, followed by site 1, then 3, and least likely be in site 2.

Google Page-rank Algorithm

Fact: the resultant state vector is the same regardless of which site one begins on.

- Initial state vector: $\begin{pmatrix}1\\0\\0\\0\end{pmatrix}$, after 100 clicks: $\begin{pmatrix}0.3000\\0.1000\\0.2667\\0.3333\end{pmatrix}$.
- Initial state vector: $\begin{pmatrix}0\\1\\0\\0\end{pmatrix}$, after 100 clicks: $\begin{pmatrix}0.3000\\0.1000\\0.2667\\0.3333\end{pmatrix}$.
- Initial state vector: $\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$, after 100 clicks: $\begin{pmatrix}0.3000\\0.1000\\0.2667\\0.3333\end{pmatrix}$.

This means that the probability of a user ending up in each site is independent of where he starts. The probability can thus be used to rank the websites.

Google Page-rank Algorithm

$$\begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} 0.3000 \\ 0.1000 \\ 0.2667 \\ 0.3333 \end{pmatrix} = \begin{pmatrix} 0.3000 \\ 0.1000 \\ 0.2667 \\ 0.3333 \end{pmatrix}$$

Observe that the resultant vector is an eigenvector of the adjacency matrix!

Google Page-rank algorithm

- After a user types in the key words, the adjacency matrix is formed.
- The eigenvector of the adjacency matrix is computed.
- The sites are rank according to their probabilities given in the resultant state vector.

Thank You for your Attention.

Email: AskMathUG@nus.edu.sg

Department webpage

https://www.math.nus.edu.sg/