Mathematics for Machine Learning and AI

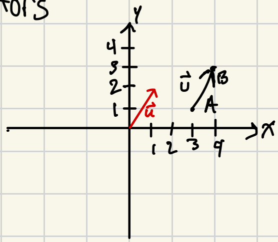
MAT 2215

2.1 Vector: a quantity defined by both a magnitude and a direction. (geometric definition)

"arrow vectors"

coordinates

A(3,1)
B(4,3)



Notation!

$\vec{u}$ ← "vector $\vec{u}$"

$\vec{u} = \langle 4-3, 3-1 \rangle = \boxed{\langle 1,2 \rangle}$

components

component form

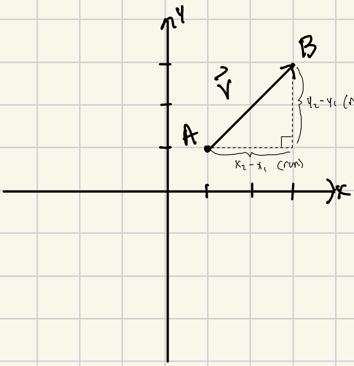
$\|\vec{u}\|$ ← "magnitude of $\vec{u}$"

$\|\vec{u}\|^2 = x^2 + y^2$

$\|\vec{u}\|^2 = (1)^2 + (2)^2$

ℓ_1-norm: $\|\vec{u}\|_1 = \sqrt{|x| + |y|}$
$= \sqrt{|1| + |2|} = \sqrt{3}$

ℓ_2-norm, → $\|\vec{u}\| = \sqrt{1+4} = \boxed{\sqrt{5}}$
length, or magnitude



$A(x_1, y_1)$, $B(x_2, y_2)$

$\overrightarrow{AB} = \vec{v} = \langle x_2 - x_1, y_2 - y_1 \rangle$

$\sqrt{\|\vec{v}\|^2} = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$

$\|\vec{v}\| = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$

Convention: a vector is a column vector.

$\vec{v} = \langle 2,3 \rangle$ is written as $\boxed{\vec{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} = (2,3)}$ space saving notation

e.g. $(1,2,3) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

$[2,3] = \vec{v}^T$ ← transpose of vector $\vec{v}$

$\vec{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \in \mathbb{R}^2$ (vector in the xy-plane)

belongs to

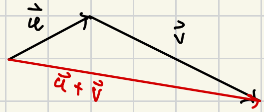
$\vec{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \mathbb{R}^3$ (vector in 3D space)

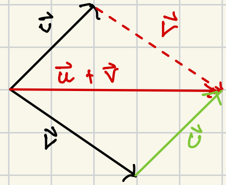
$\mathbb{R}^n$ and the "Curse of Dimensionality"

In general, $\vec{u} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$

Vector addition:

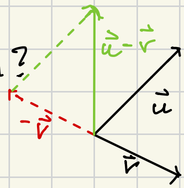
The Triangle Law





The Parallelogram Law

Subtraction?



Graph $\vec{u}-\vec{v}$

$\vec{u}+(-1)\vec{v}$

Adding and subtracting component wise: Let $\vec{u}=\begin{bmatrix}u_1\\u_2\end{bmatrix}\in\mathbb{R}^2$, $\vec{v}=\begin{bmatrix}v_1\\v_2\end{bmatrix}\in\mathbb{R}^2$

$$\vec{u}+\vec{v}=\begin{bmatrix}u_1\\u_2\end{bmatrix}+\begin{bmatrix}v_1\\v_2\end{bmatrix}=\begin{bmatrix}u_1+v_1\\u_2+v_2\end{bmatrix}$$

$$\vec{u}-\vec{v}=\begin{bmatrix}u_1\\u_2\end{bmatrix}-\begin{bmatrix}v_1\\v_2\end{bmatrix}=\begin{bmatrix}u_1-v_1\\u_2-v_2\end{bmatrix}$$

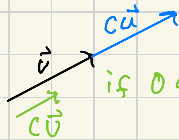
scalar

Scalar Multiplication

Let $\vec{v}=\begin{bmatrix}v_1\\v_2\end{bmatrix}\in\mathbb{R}^2$, Let $c\in\mathbb{R}$

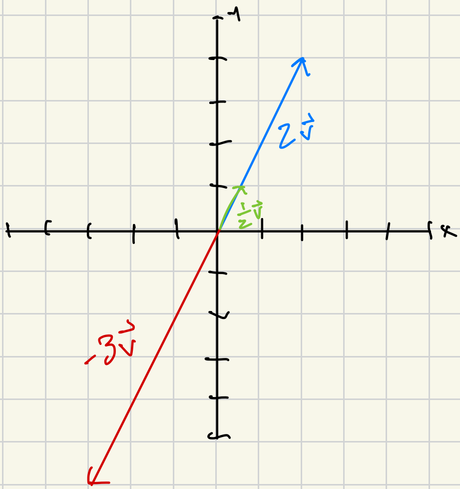
$$c\vec{v}=c\begin{bmatrix}v_1\\v_2\end{bmatrix}=\begin{bmatrix}cv_1\\cv_2\end{bmatrix}$$

if $c<1$



if $0<c<1$

Ex 1: $\vec{v}=\begin{bmatrix}1\\2\end{bmatrix}$, find and graph $2\vec{v}$, $-3\vec{v}$, $\frac{1}{2}\vec{v}$



$$2\vec{v}=2\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}2\\4\end{bmatrix}$$

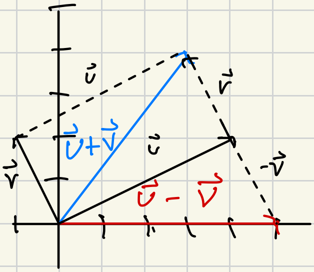
$$-3\vec{v}=-3\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}-3\\-6\end{bmatrix}$$

$$\frac{1}{2}\vec{v}=\frac{1}{2}\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}1/2\\1\end{bmatrix}$$

Ex 2: Let $\vec{u} = (4,2)$, $\vec{v} = (-1,2)$. Find and graph $\vec{u}+\vec{v}$ and $\vec{u}-\vec{v}$

$\vec{u}+\vec{v} = (4,2)+(-1,2) = (4-1, 2+2) = (3,4) = \begin{bmatrix}3\\4\end{bmatrix}$

$\vec{u}-\vec{v} = (4,2)-(-1,2) = (4+1, 2-2) = (5,0) = \begin{bmatrix}5\\0\end{bmatrix}$



Linear Combinations $(\mathbb{R}^2)$

Let's combine scalar multiplication with addition.

Let $\vec{u}, \vec{v} \in \mathbb{R}^2$, let $c, d \in \mathbb{R}$
vectors scalars

The expression $c\vec{u} + d\vec{v}$ represents all possible linear combinations of $\vec{u}$ and $\vec{v}$

Ex: $\vec{u} = \begin{bmatrix}1\\2\end{bmatrix}$ $\vec{v} = \begin{bmatrix}2\\-1\end{bmatrix}$

$1\vec{u} + 1\vec{v} = 1\begin{bmatrix}1\\2\end{bmatrix} + 1\begin{bmatrix}2\\-1\end{bmatrix} = \begin{bmatrix}1+2\\2-1\end{bmatrix} = \begin{bmatrix}3\\1\end{bmatrix}$

$1\vec{u} - 1\vec{v} = 1\begin{bmatrix}1\\2\end{bmatrix} - 1\begin{bmatrix}2\\-1\end{bmatrix} = \begin{bmatrix}1-2\\2+1\end{bmatrix} = \begin{bmatrix}-1\\3\end{bmatrix}$

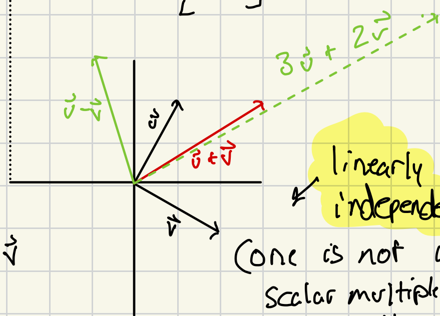
$3\vec{u} + 2\vec{v} = 3\begin{bmatrix}1\\2\end{bmatrix} + 2\begin{bmatrix}2\\-1\end{bmatrix} = \begin{bmatrix}3\\6\end{bmatrix} + \begin{bmatrix}4\\-2\end{bmatrix} = \begin{bmatrix}3+4\\6-2\end{bmatrix} = \begin{bmatrix}7\\4\end{bmatrix}$

$5\vec{u} - 3\vec{v} = 5\begin{bmatrix}1\\2\end{bmatrix} - 3\begin{bmatrix}2\\-1\end{bmatrix} = \begin{bmatrix}5\\10\end{bmatrix} + \begin{bmatrix}-6\\3\end{bmatrix} = \begin{bmatrix}5-6\\10+3\end{bmatrix} = \begin{bmatrix}-1\\13\end{bmatrix}$

$0\vec{u} - 5\vec{v} = 0\begin{bmatrix}1\\2\end{bmatrix} - 5\begin{bmatrix}2\\-1\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} + \begin{bmatrix}-10\\5\end{bmatrix} = \begin{bmatrix}0-10\\0+5\end{bmatrix} = \begin{bmatrix}-10\\5\end{bmatrix}$

$0\vec{u} + 0\vec{v} = 0\begin{bmatrix}1\\2\end{bmatrix} + 0\begin{bmatrix}2\\-1\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} + \begin{bmatrix}0\\0\end{bmatrix} = \begin{bmatrix}0+0\\0+0\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$

$c\vec{u} + d\vec{v} = c\begin{bmatrix}1\\2\end{bmatrix} + d\begin{bmatrix}2\\-1\end{bmatrix}$
$= \begin{bmatrix}c\\2c\end{bmatrix} + \begin{bmatrix}2d\\-d\end{bmatrix}$
$= \begin{bmatrix}c+2d\\2c-d\end{bmatrix}$



The set of linear combinations of $\vec{u}$ and $\vec{v}$ fill the plane.

"$\vec{u}$ and $\vec{v}$ span $\mathbb{R}^2$"

linearly independent
(One is not a scalar multiple of the other)

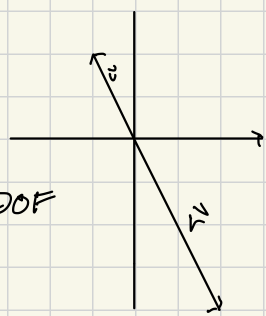
Ex: $\vec{u} = \begin{bmatrix}-1\\2\end{bmatrix}$, $\vec{v} = \begin{bmatrix}2\\-4\end{bmatrix}$. What does the set of all linear combinations of $\vec{u}$ & $\vec{v}$ fill? (What do $\vec{u}$ & $\vec{v}$ span?)

The line that holds $\vec{u}$ and $\vec{v}$

"degrees of freedom" (2)

$\vec{u}$ and $\vec{v}$ **span** a line

1 DOF



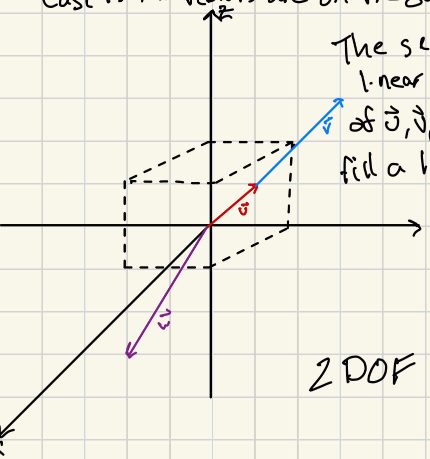
linearly dependent
(One is a scalar multiple of the other)

Linear Combinations of 3 vectors in $\mathbb{R}^3$

Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^3$ $c, d, e \in \mathbb{R}$

Linear combinations: $c\vec{u} + d\vec{v} + e\vec{w}$

Case 1: All vectors are on the same line. (all vectors are scalar multiples of one of them)



The set of linear combinations of $\vec{u}, \vec{v}$, and $\vec{w}$ fill a line

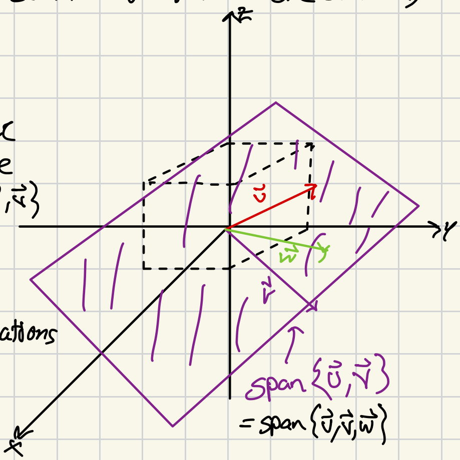
Case 2:

If $\vec{u}$ & $\vec{v}$ are not on the same line

& $\vec{w} \in \text{span}\{\vec{u}, \vec{v}\}$

span$\{\vec{u}, \vec{v}, \vec{w}\}$ is a plane

The linear combinations of $\vec{u}, \vec{v}$, & $\vec{w}$ fill a plane.

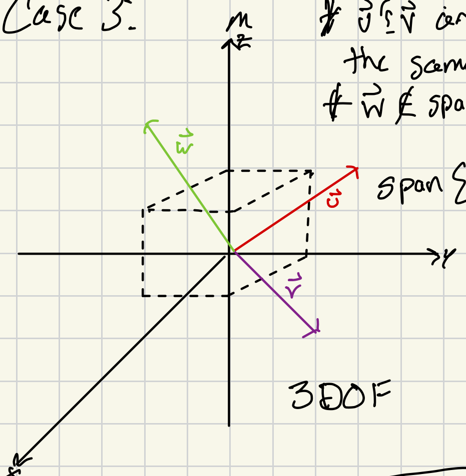


Case 3:

If $\vec{u}$ & $\vec{v}$ are not on the same line

& $\vec{w} \notin \text{span}\{\vec{u}, \vec{v}\}$

(span$\{\vec{u}, \vec{v}\}$ is a plane)



span$\{\vec{u}, \vec{v}, \vec{w}\} \in \mathbb{R}^3$

Linear combinations of $\vec{u}, \vec{v}$, & $\vec{w}$ fill $\mathbb{R}^3$

Ex. (a) let $\vec{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} -4 \\ -8 \\ -12 \end{bmatrix}$. The linear combinations of $\vec{u}$ & $\vec{v}$ fill a line

← check the ratios to see if they're consistent.

$\vec{v} = -4\vec{u}$

(b) let $\vec{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} 0 \\ 1 \\ 5 \end{bmatrix}$. The linear combinations of $\vec{u}$ & $\vec{v}$ fill a plane

$\vec{v} \neq c\vec{u}$

(c) let $\vec{u} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$, $\vec{w} = \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}$. The linear combinations of $\vec{u}, \vec{v}$, & $\vec{w}$ fill a plane

$\vec{w} = 2\vec{u} - \vec{v}$

(d) let $\vec{u} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$, $\vec{w} = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}$. The linear combinations of $\vec{u}, \vec{v}$, & $\vec{w}$ fill a plane

$\vec{w} = 2\vec{u} + \vec{v}$

$\vec{u}$ & $\vec{v}$ span a plane $\quad$ $\vec{w} \in \text{span}\{\vec{u}, \vec{v}\}$

Lengths & Dot Products

Def: The dot product (inner product) of $\vec{u}=(u_1,u_2)$ and $\vec{v}=(v_1,v_2)$ is given by $\vec{u}\cdot\vec{v}=(u_1,u_2)\cdot(v_1,v_2)$ $\mathbb{R}^2$

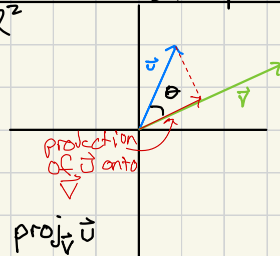
$= u_1v_1+u_2v_2$ ← outcome is a scalar

$\mathbb{R}^3$: $\vec{u}=\begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix}$, $\vec{v}=\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix}$, $\vec{u}\cdot\vec{v}=u_1v_1+u_2v_2+u_3v_3$

$\mathbb{R}^n$: $\vec{u}=\begin{bmatrix}u_1\\u_2\\\vdots\\u_n\end{bmatrix}$, $\vec{v}=\begin{bmatrix}v_1\\v_2\\\vdots\\v_n\end{bmatrix}$, $\vec{u}\cdot\vec{v}=u_1v_1+u_2v_2+\cdots+u_nv_n$

Geometric Interpretation

$\vec{u},\vec{v}\in\mathbb{R}^2$



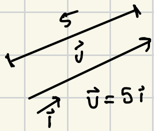
$\text{proj}_{\vec{v}}\vec{u}$

$\underbrace{\text{length of } \text{proj}_{\vec{v}}\vec{u}}_{\text{comp}_{\vec{v}}\vec{u}} = \|\text{proj}_{\vec{v}}\vec{u}\|$

- component of $\vec{u}$ along direction of $\vec{v}$

$\cos\theta=\dfrac{\text{comp}_{\vec{v}}\vec{u}}{\|\vec{u}\|}$

$\text{comp}_{\vec{v}}\vec{u}=\|\vec{u}\|\cos\theta=\dfrac{\vec{u}\cdot\vec{v}}{\|\vec{v}\|}$



$\vec{u}\cdot\vec{v}=\|\vec{u}\|\|\vec{v}\|\cos\theta$

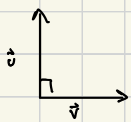
$=\text{comp}_{\vec{v}}\vec{u}\,\|\vec{v}\|$

$\dfrac{\vec{v}}{\|\vec{v}\|}$ ← unit vector in the direction of $\vec{v}$ (unit vector means it has a length of 1)

$\text{proj}_{\vec{v}}\vec{u}=\left(\text{comp}_{\vec{v}}\vec{u}\right)\dfrac{\vec{v}}{\|\vec{v}\|}$

$=\dfrac{\vec{u}\cdot\vec{v}}{\|\vec{v}\|}\dfrac{\vec{v}}{\|\vec{v}\|}$

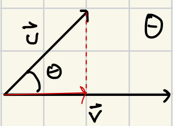
$=\dfrac{\vec{u}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v}$



$\text{comp}_{\vec{v}}\vec{u}=0$ $\quad\vec{u}\cdot\vec{v}=0$

$\text{proj}_{\vec{v}}\vec{u}=\vec{0}$

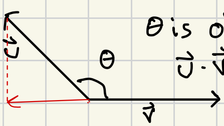
orthogonal



θ is acute $(\theta<90°)$

$\vec{u}\cdot\vec{v}>0$

"Information Compression"



θ is obtuse $(\theta>90°)$

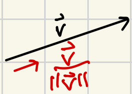
$\vec{u}\cdot\vec{v}<0$

$\cos\theta=\dfrac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|}$

$0\le\theta\le\pi$

180°

Unit vector: A vector with length or magnitude of 1 unit. We can manufacture a unit vector in the direction of $\vec{v}$ by dividing $\vec{v}$ by its magnitude



Ex: Let $\vec{u}=\begin{bmatrix}4\\2\end{bmatrix}$, $\vec{v}=\begin{bmatrix}-1\\2\end{bmatrix}$, find $\vec{u}\cdot\vec{v}$.

$$\vec{u}\cdot\vec{v}=\begin{bmatrix}4\\2\end{bmatrix}\cdot\begin{bmatrix}-1\\2\end{bmatrix}=4(-1)+2(2)=-4+4=0 \quad \text{(Orthogonal)}$$

Ex: Let $\vec{u}=(2,3)$, $\vec{v}=(1,5)$. Find the cosine of the angle between $\vec{u}$ & $\vec{v}$ then deduce the angle (to nearest degree)

$$\vec{u}\cdot\vec{v}=(2,3)\cdot(1,-5)=2(1)+3(-5)=2-15=-13$$

$$\|\vec{u}\|=\sqrt{(2)^2+(3)^2}=\sqrt{4+9}=\sqrt{13}$$

$$\|\vec{v}\|=\sqrt{(1)^2+(5)^2}=\sqrt{1+25}=\sqrt{26}$$

$$\cos\theta=\frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|}=\frac{-13}{\sqrt{13}\sqrt{26}}=\frac{-13}{\sqrt{13^2\cdot 2}}=\frac{-13}{13\sqrt{2}}=-\frac{1}{\sqrt{2}}=-\frac{\sqrt{2}}{2}$$

$$\theta=135^\circ$$

Ed Loenz
chaos theory

Matrices and Operations

Matrix: An array of numbers or quantities arranged into rows and columns.

Ex: $\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}=A$ Dimension (size): rows x columns

3 x 3

main diagonal Square matrix

$B=\begin{bmatrix}-1&15&0\\\sqrt{2}&\frac{1}{1000}&-0.3\end{bmatrix}$ Dimension: 2x3

rectangular matrix

elements are called "entries"

entry in row 1, column 2 is denoted by a_{12}

$a_{12}=2$

Matrix Addition/Subtraction

Let A, B be mxn matrices

$$A=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}\end{bmatrix}_{m\times n} \quad B=\begin{bmatrix}b_{11}&b_{12}&\cdots&b_{1n}\\b_{21}&b_{22}&\cdots&b_{2n}\\\vdots&\vdots&&\vdots\\b_{m1}&b_{m2}&\cdots&b_{mn}\end{bmatrix}_{m\times n}$$

$$A+B=\begin{bmatrix}a_{11}+b_{11}&a_{12}+b_{12}&\cdots&a_{1n}+b_{1n}\\a_{21}+b_{21}&a_{22}+b_{22}&\cdots&a_{2n}+b_{2n}\\\vdots&\vdots&&\vdots\\a_{m1}+b_{m1}&a_{m2}+b_{m2}&\cdots&a_{mn}+b_{mn}\end{bmatrix}$$

$$A=\left[a_{ij}\right]_{1\le j\le n}^{1\le i\le m} \quad B=\left[b_{ij}\right]_{1\le j\le n}^{1\le i\le m} \quad A+B=\left[a_{ij}+b_{ij}\right]_{1\le j\le n}^{1\le i\le m}$$

Ex: Find A+B and A-B. $A=\begin{bmatrix}1&-2&3\\5&0&7\end{bmatrix}$, $B=\begin{bmatrix}-2&7&5\\11&0&-1\end{bmatrix}$

$$A+B=\begin{bmatrix}1-2&-2+7&3+5\\5+11&0+0&7-1\end{bmatrix}=\begin{bmatrix}-1&5&8\\16&0&6\end{bmatrix}$$

$$A-B=\begin{bmatrix}1+2&-2-7&3-5\\5-11&0-0&7+1\end{bmatrix}=\begin{bmatrix}3&-9&-2\\-6&0&8\end{bmatrix}$$

Scalar Multiplication

Let A be an $m \times n$ matrix $(\in \mathbb{R}^{m \times n}$ or $\mathbb{M}_{m\times n})$. Let $c \in \mathbb{R}$
scalar

$$A = [a_{ij}]_{\substack{1\le i\le m \\ 1 \le j \le n}} \qquad cA = c[a_{ij}]_{\substack{1\le i\le m \\ 1 \le j \le n}} = [ca_{ij}]_{\substack{1\le i\le m \\ 1 \le j \le n}}$$

Ex: $A = \begin{bmatrix} 5 & -1 \\ 3 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 7 & 0 & 1 \\ -3 & 2 & 5 \end{bmatrix}$. Find $2A$, $-3B$.

$$2A = 2\begin{bmatrix} 5 & -1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 10 & -2 \\ 6 & 4 \end{bmatrix} \qquad -3B = -3\begin{bmatrix} 7 & 0 & 1 \\ -3 & 2 & 5 \end{bmatrix} = \begin{bmatrix} -21 & 0 & -3 \\ 9 & -6 & -15 \end{bmatrix}$$

Matrix Multiplication

Fundamental operation is the dot (inner) product of two vectors.

Recall: $\vec{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_{3\times 1}$ $\vec{v} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}_{3\times 1}$ $\vec{u}\cdot\vec{v} = 1(4) + 2(5) + 3(6) = 4 + 10 + 18 = 32 \leftarrow$ scalar

$$\underbrace{\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}}_{\vec{u}^T} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = 1(4) + 2(5) + 3(6) \qquad \vec{u}\cdot\vec{v} = \vec{u}^T\vec{v}$$

transpose of $\vec{u}$

product of a row by a column vector

We can only multiply matrices, say matrix A by matrix B, if the number of columns of the first matrix matches the number of rows in the second matrix.

$\underset{3\times 2}{A}$, $\underset{2\times 2}{B}$ Find AB, BA.

AB ← outcome is a 3×2 matrix
3×2 2×2
must match for product to be possible

BA
2×2 3×2
different, product does not exist.

$$A = \begin{bmatrix} 2 & -1 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 \\ -1 & 7 \end{bmatrix}$$

$$AB = \begin{array}{c} r_1 \\ r_2 \\ r_3 \end{array}\begin{bmatrix} 2 & -1 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} \begin{array}{cc} c_1 & c_2 \\ \begin{bmatrix} 1 & 2 \\ -1 & 7 \end{bmatrix} \end{array} = \begin{bmatrix} 3 & -3 \\ -1 & 7 \\ 3 & 24 \end{bmatrix}$$

$\underset{r_1}{\begin{bmatrix} 2 & -1 \end{bmatrix}} \underset{c_1}{\begin{bmatrix} 1 \\ -1 \end{bmatrix}} = 2(1) + (-1)(-1) = 2 + 1 = 3$

$\underset{r_2}{\begin{bmatrix} 0 & 1 \end{bmatrix}} \underset{c_1}{\begin{bmatrix} 1 \\ -1 \end{bmatrix}} = 0(1) + (1)(-1) = 0 - 1 = -1$

$\underset{r_3}{\begin{bmatrix} 5 & 2 \end{bmatrix}} \underset{c_1}{\begin{bmatrix} 1 \\ -1 \end{bmatrix}} = 5(1) + (2)(-1) = 5 - 2 = 3$

$\underset{r_1}{\begin{bmatrix} 2 & -1 \end{bmatrix}} \underset{c_2}{\begin{bmatrix} 2 \\ 7 \end{bmatrix}} = 2(2) + (-1)(7) = 4 - 7 = -3$

$\underset{r_2}{\begin{bmatrix} 0 & 1 \end{bmatrix}} \underset{c_2}{\begin{bmatrix} 2 \\ 7 \end{bmatrix}} = 0(2) + 1(7) = 0 + 7 = 7$

$\underset{r_3}{\begin{bmatrix} 5 & 2 \end{bmatrix}} \underset{c_2}{\begin{bmatrix} 2 \\ 7 \end{bmatrix}} = 5(2) + 2(7) = 10 + 14 = 24$

Ex: Let $A = \begin{bmatrix} 3 & 2 & 1 \\ -1 & 5 & 0 \\ 0 & -1 & 11 \end{bmatrix}_{3\times 3}$, $\vec{x} = \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix}_{3\times 1}$. Find $A\vec{x}$.

$$A\vec{x} = \begin{bmatrix} 3 & 2 & 1 \\ -1 & 5 & 0 \\ 0 & -1 & 11 \end{bmatrix}\begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 15 \\ -10 \\ 23 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 2 & 1 \end{bmatrix}\begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix} = 3(5) + 2(-1) + 1(2) = 15 - 2 + 2 = 15$$

$$\begin{bmatrix} -1 & 5 & 0 \end{bmatrix}\begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix} = -1(5) + 5(-1) + 0(2) = -5 - 5 + 0 = -10$$

$$\begin{bmatrix} 0 & -1 & 11 \end{bmatrix}\begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix} = 0(5) - 1(-1) + 11(2) = 0 + 1 + 22 = 23$$

Other ways to multiply matrices

② Multiply matrix A by each column of matrix B

$$\underset{m\times n}{A}\ \underset{n\times p}{B} = A\begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_p \end{bmatrix} = \begin{bmatrix} A\vec{b}_1 & A\vec{b}_2 & \cdots & A\vec{b}_p \end{bmatrix}$$

Ex: $A = \begin{bmatrix} 2 & 3 & 5 \\ 5 & 1 & 0 \end{bmatrix}_{2\times 3}$ $B = \begin{bmatrix} -1 & 0 & 5 \\ 2 & 1 & 7 \\ -3 & 2 & 1 \end{bmatrix}_{3\times 3}$ with columns $\vec{b}_1, \vec{b}_2, \vec{b}_3$

$$A\vec{b}_1 = \begin{bmatrix} 2 & 3 & 5 \\ 5 & 1 & 0 \end{bmatrix}\begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} -11 \\ -3 \end{bmatrix} \qquad A\vec{b}_2 = \begin{bmatrix} 2 & 3 & 5 \\ 5 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 13 \\ 1 \end{bmatrix}$$

$$A\vec{b}_3 = \begin{bmatrix} 2 & 3 & 5 \\ 5 & 1 & 0 \end{bmatrix}\begin{bmatrix} 5 \\ 7 \\ 1 \end{bmatrix} = \begin{bmatrix} 36 \\ 32 \end{bmatrix}$$

$$AB = \begin{bmatrix} -11 & 13 & 36 \\ -3 & 1 & 32 \end{bmatrix}$$

③ Multiply rows of A by B

$$\underset{1\times n}{\begin{bmatrix} \text{row } i \text{ of } A \end{bmatrix}}\ \underset{n\times p}{\begin{bmatrix} b_{np} \end{bmatrix}} = \begin{bmatrix} \text{row } i \text{ of } AB \end{bmatrix}$$

Ex: $A = \begin{bmatrix} 2 & 3 & 5 \\ 5 & 1 & 0 \end{bmatrix}_{2\times 3}$ $B = \begin{bmatrix} -1 & 0 & 5 \\ 2 & 1 & 7 \\ -3 & 2 & 1 \end{bmatrix}_{3\times 3}$

$$\underset{1\times 3}{\begin{bmatrix} 2 & 3 & 5 \end{bmatrix}}\ \underset{3\times 3}{\begin{bmatrix} -1 & 0 & 5 \\ 2 & 1 & 7 \\ -3 & 2 & 1 \end{bmatrix}} = \begin{bmatrix} -11 & 13 & 36 \end{bmatrix}$$

$$AB = \begin{bmatrix} -11 & 13 & 36 \\ -3 & 1 & 32 \end{bmatrix} \qquad \begin{bmatrix} 5 & 1 & 0 \end{bmatrix}\begin{bmatrix} -1 & 0 & 5 \\ 2 & 1 & 7 \\ -3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} -3 & 1 & 32 \end{bmatrix}$$

④ Multiply columns of A by rows of B (outer product)

$$\underset{A}{\begin{bmatrix} | & | & | \\ col^1 & col^2 & col^3 \\ | & | & | \end{bmatrix}} \underset{B}{\begin{bmatrix} -row\,1- \\ -row\,2- \\ -row\,3- \end{bmatrix}} = (col\,1)(row\,1) + (col\,2)(row\,2) + (col\,3)(row\,3)$$

Note: $\underset{\vec{u}^T}{\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}} \underset{\vec{v}}{\begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}} = 1(4)+2(5)+3(6) = 4+10+18 = 32$

inner product (dot product)

$$\overset{3\times 1}{\underset{\vec{u}}{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}} \overset{1\times 3}{\underset{\vec{v}^T}{\begin{bmatrix} 4 & 5 & 6 \end{bmatrix}}} = \overset{3\times 3}{\begin{bmatrix} 4 & 5 & 6 \\ 8 & 10 & 12 \\ 12 & 15 & 18 \end{bmatrix}}$$ outer product

Ex: $A = \begin{bmatrix} 2 & 3 & 5 \\ 5 & 1 & 0 \end{bmatrix}$ $B = \begin{bmatrix} -1 & 0 & 5 \\ 2 & 1 & 7 \\ -3 & 2 & 1 \end{bmatrix}$ $AB = \begin{bmatrix} 2 \\ 5 \end{bmatrix}\begin{bmatrix} -1 & 0 & 5 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \end{bmatrix}\begin{bmatrix} 2 & 1 & 7 \end{bmatrix} + \begin{bmatrix} 5 \\ 0 \end{bmatrix}\begin{bmatrix} -3 & 2 & 1 \end{bmatrix}$

$$= \begin{bmatrix} -2 & 0 & 10 \\ -5 & 0 & 25 \end{bmatrix} + \begin{bmatrix} 6 & 3 & 21 \\ 2 & 1 & 7 \end{bmatrix} + \begin{bmatrix} -15 & 10 & 5 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} -11 & 13 & 36 \\ -3 & 1 & 32 \end{bmatrix}$$

Note: $\underset{n\times n}{A} \; \underset{n\times n}{B} \to \underset{n\times n}{AB} \quad O(n^3)$

for 1000, 1000^3 computations are required to get the product

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$ sparse matrix

Special Case

Product of a matrix by a vector

Ex: $A = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 0 & 7 \\ 1 & -1 & 2 \end{bmatrix}$ $\vec{x} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$ $A\vec{x} = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 0 & 7 \\ 1 & -1 & 2 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$ We can reframe the product as a linear combination of the column of A

$$A\vec{x} = 2\begin{bmatrix} 1 \\ 5 \\ 1 \end{bmatrix} + (-1)\begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} + 0\begin{bmatrix} 3 \\ 7 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 10 \\ 2 \end{bmatrix} + \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 10 \\ 3 \end{bmatrix}$$

Laws of Matrix Operations

Let A, B, C be $m\times n$ matrices; $c, d \in \mathbb{R}$

- $A + B = B + A$ commutativity of addition
- $c(A+B) = cA + cB$ distributivity I
- $(c+d)A = cA + dA$ distributivity II
- $(A+B)+C = A+(B+C)$ associativity of addition

In general, regardless of the sizes of A & B, $AB \neq BA$. noncommutativity of matrix addition

$\underset{m\times n}{A}(\underset{n\times p}{B}+\underset{n\times r}{C}) = AB+AC$ distributivity from the left

$(\underset{m\times n}{A}+\underset{p\times n}{B})\underset{n\times r}{C} = AC+BC$ distributivity from the right

Special Matrices

The zero matrix is a matrix where all entries are 0.

$$O_{m\times n} = \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}}_{n \text{ columns}} \Big\} m \text{ rows} \qquad O_{2\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Let A be an $m\times n$ matrix, $A + O_{m\times n} = A$

The identity matrix is a square matrix with all zeroes except for 1's in the main diagonal

$$\underset{n\times n}{I_n} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}}_{m \text{ columns}} \Big\} n \text{ rows} \qquad I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\underset{m\times n}{A}\,\underset{n\times n}{I_n} = \underset{m\times n}{A}$$

Transpose of a matrix

Let A be an $m\times n$ matrix, $A = [a_{ij}]_{1\le j\le n}^{1\le i\le m}$. The transpose of A, denoted by A^T, is given by $A^T = [a_{ji}]_{1\le j\le n}^{1\le i\le m}$. ("$A$ transpose")

Ex. $\underset{2\times 3}{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, $\underset{3\times 2}{A^T} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

Properties of Transpose operation

$(A^T)^T = A$

$(A+B)^T = A^T + B^T$

$(cA)^T = cA^T$

$(AB)^T = B^T A^T$

$(\underset{2\times 3}{A}\ \underset{3\times 3}{B})^T$ $\qquad \underset{3\times 2}{A^T}\,\underset{3\times 3}{B^T}$ invalid

$\underset{2\times 3}{AB}$ $\qquad \underset{3\times 3}{B^T}\,\underset{3\times 2}{A^T}$

$\underset{3\times 2}{(AB)^T}$ $\qquad 3\times 2$

Inverse of a matrix

Let A be an $n \times n$ matrix. A is invertible if there exists a matrix $(n \times n)$ denoted by A^{-1}, such that $AA^{-1} = A^{-1}A = I_n$

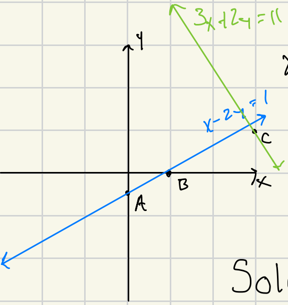
A^{-1}: The inverse matrix of matrix A ("A inverse")

Solving Systems of Linear Equations

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases}$$

System of two linear equations with two unknowns (variables)

* graphical
* Algebraic
  - Substitution
  - Elimination
  - Hybrid



$x - 2y = 1$

$x=0, \ -2y = 1 \Rightarrow y = -\frac{1}{2}$

$A(0, -\frac{1}{2})$

$y=0, \ x=1$

$B(1,0)$

$3x + 2y = 11$

$x=0, \ 2y = 11 \Rightarrow \frac{11}{2}$

$\cancel{C(0, \frac{11}{2})}$ $C(3,1)$

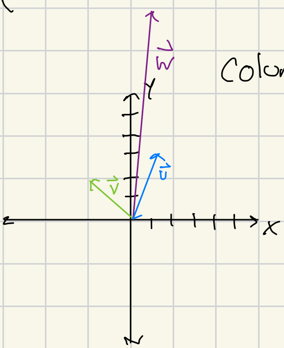
$y=0, \ 3x = 11 \Rightarrow x = \frac{11}{3}$

$D(0, \frac{11}{3})$

Solution $\{(3,1)\}$

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases}$$

Row picture: each row represents a line. Solving means finding intersection of the two lines.



Column picture: $\begin{bmatrix} x \\ 3x \end{bmatrix} + \begin{bmatrix} -2y \\ 2y \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}$

$x\begin{bmatrix} 1 \\ 3 \end{bmatrix} + y\begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}$ – finding the linear combination of $\underset{\vec{u}}{\begin{bmatrix} 1 \\ 3 \end{bmatrix}}$ & $\underset{\vec{v}}{\begin{bmatrix} -2 \\ 2 \end{bmatrix}}$ that produces $\underset{\vec{w}}{\begin{bmatrix} 1 \\ 11 \end{bmatrix}}$

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 1 \end{cases} \Leftrightarrow \left[\begin{array}{cc|c} 1 & -2 & 1 \\ 3 & 2 & 11 \end{array}\right]$$ Augmented matrix of the system

A: $\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}$ coefficient matrix

$\vec{b}$: $\begin{bmatrix} 1 \\ 11 \end{bmatrix}$ constant vector

Target: $\begin{cases} x = 3 \\ y = 1 \end{cases}$ or $\left[\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & 1 \end{array}\right]$ reduced row-echelon form (RREF)

$\vec{x}$: $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

$$\underset{A}{\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}} \underset{\vec{x}}{\begin{bmatrix} x \\ y \end{bmatrix}} = \underset{\vec{b}}{\begin{bmatrix} 1 \\ 11 \end{bmatrix}} \quad A\vec{x} = \vec{b} \Rightarrow \vec{x} = ?$$

matrix-vector equation

Solving is equivalent to finding a linear combination of the columns of A that produces $\vec{b}$.

Solving by backward substitution

Solve $\begin{cases} x - 2y + z = 11 \\ y - 3z = 2 \\ z = 10 \end{cases}$ triangular system

$\boxed{z = 10}$

$y - 3z = 2$

$y - 30 = 2$

$\boxed{y = 32}$

$x - 2y + z = 11$

$x - 64 + 10 = 11$

$x - 54 = 11$

$\boxed{x = 64}$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 64 \\ 32 \\ 10 \end{bmatrix} \qquad \left[\begin{array}{ccc|c} 1 & -2 & 1 & 11 \\ 0 & 1 & -3 & 2 \\ 0 & 0 & 1 & 10 \end{array}\right]$$

row-echelon form (REF)

Gaussian Elimination

1. Form augmented matrix of the system $[A|\vec{b}]$
2. Apply row operations to reduce the augmented matrix to row-echelon form.
3. Convert REF to triangular system and solve by backwards substitution

Elementary Row Operations

- Interchange any two rows
- Multiply/Divide any row by a non-zero constant

$\}$ useful for creating a "1" (ones)

- Add a non-zero multiple of one row to another to change it $\}$ create "0s" (zeroes)

Row Echelon Form (REF)

A matrix is in row echelon form if it satisfies the following criteria:

1. If a row consists entirely of zeroes, it must appear at the bottom of the matrix.
2. For non-zero rows only, the first non zero number (pivot or leading coefficient) that must equal 1.
3. The pivot or 1 in one row should be to the left of a pivot in a row below it.

Reduced Row Echelon Form (RREF)

A matrix is in reduced row echelon form if it is in row echelon form and satisfies the following additional criteria:

4. The pivot is the only non-zero entry in its column.

Ex: Identify which of the following matrices are in REF or RREF

a) $\begin{bmatrix} 2 & -1 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ no, #2

b) $\begin{bmatrix} 1 & -2 & 3 & 4 \\ 0 & 0 & 1 & 5 \end{bmatrix}$ REF, no, #4

c) $\begin{bmatrix} 0 & 0 & 1 & 2 & 3 \\ 1 & 0 & 5 & 0 & 0 \\ 0 & 1 & 3 & -1 & 7 \end{bmatrix}$ no, #3

d) $\begin{bmatrix} 1 & 0 & 5 & 7 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 11 \end{bmatrix}$ no, #2

e) $\begin{bmatrix} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 100 \end{bmatrix}$ REF, RREF

f) $\begin{bmatrix} 1 & 3 & 2 & 1 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ REF, no, #4

g) $\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 0 & -1 & 5 \end{bmatrix}$ no, #2

Gauss-Jordan Reduction

① Form augmented matrix $[A \mid \vec{b}]$

② Use row reduction to convert augmented matrix to RREF

Remarks:

* If all columns of A lead to pivot columns, then the system has a unique solution.
* If some columns of A turn into non-pivot columns, then the system has infinitely many solutions.
* If the REF or RREF has a row of the form $[0\,0\,0 \cdots 0 \mid k]$ where $k \neq 0$, then the system has no solution. $\Rightarrow \{\ \}$

Consistent: the system has at least one solution

inconsistent: the system has no solution.

Ex: a) $\begin{cases} x+y+z=0 \\ \quad\ y+z=1 \end{cases}$

$$\begin{array}{c} \text{non-pivot column} \\ \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{array}\right] \end{array}$$
REF

z is a non-pivot variable (arbitrary variable)

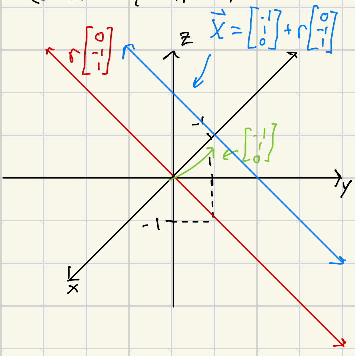
x, y: pivot variables

Let $z = r$, $r \in \mathbb{R}$
(r is a real parameter)

$y = 1 - z = 1 - r$ $\qquad x = -y - z = -(1-r) - r = -1 + r - r = -1$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -1 \\ 1-r \\ r \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -r \\ r \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + r\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

$$\left\{ \underbrace{\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}}_{\substack{\vec{x}_p \\ \text{particular} \\ \text{solution}}} + r \underbrace{\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}}_{\substack{\vec{x}_h \\ \text{homogeneous} \\ \text{solution}}} \middle| \ r \in \mathbb{R} \right\}$$



$\begin{cases} x+y+z=0 \\ \quad\ y+z=1 \end{cases}$
non-homogeneous system, because $\vec{b} \neq 0$

associated homogeneous system $\begin{cases} x+y+z=0 \\ \quad\ y+z=0 \end{cases}$

$\vec{x}_p$: particular solution to the non-homogeneous system

$\vec{x}_h$: General solution to the homogeneous system

$\vec{x} = \vec{x}_p + \vec{x}_h$: general solution to the non-homogeneous system.

Ex: $\begin{cases} x - y - 2z = 1 \\ 2x + 3y + z = 2 \\ 5x + 4y + 2z = 4 \end{cases}$ Solve by Gaussian Elimination

$$\left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \\ 2 & 3 & 1 & 2 \\ 5 & 4 & 2 & 4 \end{array}\right] \begin{array}{l} \\ \leftarrow -2\left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \end{array}\right] \\ \leftarrow -5\left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \end{array}\right] \end{array} \xrightarrow[R_3 \leftarrow (-5)R_1 + R]{R_2 \leftarrow (-2)R_1 + R_2} \left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \\ 0 & 5 & 5 & 0 \\ 0 & 9 & 12 & -1 \end{array}\right] \begin{array}{l} \\ R_3 \leftarrow -1R_2 + R_3 \\ \leftarrow -1\left[\begin{array}{ccc|c} 0 & 5 & 5 & 0 \end{array}\right] \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \\ 0 & 5 & 5 & 0 \\ 0 & 4 & 7 & -1 \end{array}\right] \begin{array}{l} R_2 \leftarrow -1R_3 + R_2 \\ \leftarrow -1\left[\begin{array}{ccc|c} 0 & 4 & 7 & -1 \end{array}\right] \\ \\ \end{array} \rightsquigarrow \left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \\ 0 & 1 & -2 & 1 \\ 0 & 4 & 7 & -1 \end{array}\right] \begin{array}{l} \\ R_3 \leftarrow -4R_2 + R_3 \\ \leftarrow -4\left[\begin{array}{ccc|c} 0 & 1 & -2 & 1 \end{array}\right] \end{array}$$

$$\left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 15 & -4 \end{array}\right] \leftarrow \frac{1}{15} \longrightarrow \left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -\frac{4}{15} \end{array}\right] \qquad \begin{cases} x - y - 2z = 1 \\ y - 2z = 1 \\ z = -\frac{4}{15} \end{cases}$$

$y + 2\left(\frac{4}{15}\right) = 1$

$y = 1 - \frac{8}{15} = \frac{7}{15}$

Doesn't work. Try again.

$$\left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \\ 2 & 3 & 1 & 2 \\ 5 & 4 & 2 & 4 \end{array}\right] \begin{array}{l} \\ \leftarrow -2\left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \end{array}\right] \\ \leftarrow -5\left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \end{array}\right] \end{array} \xrightarrow[R_3 \leftarrow (-5)R_1 + R]{R_2 \leftarrow (-2)R_1 + R_2} \left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \\ 0 & 5 & 5 & 0 \\ 0 & 9 & 12 & -1 \end{array}\right] \begin{array}{l} \\ R_3 \leftarrow -1R_2 + R_3 \\ \leftarrow -1\left[\begin{array}{ccc|c} 0 & 5 & 5 & 0 \end{array}\right] \end{array}$$

$R_2 \leftarrow \frac{1}{5}R_2$ $\qquad$ $R_3 \leftarrow -9R_2 + R_3$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 9 & 12 & -1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 3 & -1 \end{array}\right] \qquad \begin{cases} x - y - 2z = 1 \\ y + z = 0 \\ z = -\frac{1}{3} \end{cases}$$

$z = -\frac{1}{3}$

$y - \frac{1}{3} = 0 \Rightarrow y = \frac{1}{3}$

$x - \frac{1}{3} - 2\left(\frac{-1}{3}\right) = 1$

$x - \frac{1}{3} + \frac{2}{3} = 1$

$x + \frac{1}{3} = 1$

$x = \frac{2}{3}$

$R_3 \leftarrow \frac{1}{3}R_3$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & -\frac{1}{3} \end{array}\right]$$

all pivot columns

So we expect

a unique solution

$2 - \frac{1}{3} - 2\left(-\frac{1}{3}\right) = 1$

$2 - \frac{1}{3} + \frac{2}{3} = 1$

$$\vec{x}: \left\{ \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} \right\}$$

Ex. $A=\begin{bmatrix}2&1\\-3&5\end{bmatrix}$ $\vec{b}=\begin{bmatrix}1\\2\end{bmatrix}$ Solve by Gauss-Jordan Reduction

$$A|\vec{b}=\left[\begin{array}{cc|c}2&1&1\\-3&5&2\end{array}\right]\xrightarrow{R_1\leftarrow 1R_2+R_1}\left[\begin{array}{cc|c}-1&6&3\\-3&5&2\end{array}\right]\xrightarrow{R_1\leftarrow -1R_1}\left[\begin{array}{cc|c}1&-6&-3\\-3&5&2\end{array}\right]\xrightarrow{R_2\leftarrow 3R_1+R_2}\left[\begin{array}{cc|c}1&-6&-3\\0&-13&-7\end{array}\right]$$

$$\xrightarrow{R_2\leftarrow -\frac{1}{13}R_2}\left[\begin{array}{cc|c}1&-6&-3\\0&1&7/13\end{array}\right]\xrightarrow{R_1\leftarrow 6R_2}\underbrace{\left[\begin{array}{cc|c}1&0&3/13\\0&1&7/13\end{array}\right]}_{\text{RREF}}$$

$-3+\frac{42}{13}$

$-\frac{39}{13}+\frac{42}{13}=3/13$

$$\left\{\begin{bmatrix}3/13\\7/13\end{bmatrix}\right\}$$

Ex. $\begin{cases} x_2+x_3-3x_4=2\\ x_1\quad +2x_3-4x_4=1\end{cases}$ Solve by Gauss-Jordan Reduction

$$\left[\begin{array}{cccc|c}0&1&1&-3&2\\1&0&2&-4&1\end{array}\right]\xrightarrow{R_1\leftrightarrow R_2}\overset{\text{RREF}}{\left[\begin{array}{cccc|c}1&0&2&-4&1\\0&1&1&-3&2\end{array}\right]}$$

pivot columns — non-pivot columns

only pivot columns need to have 0's and 1's

let $x_3=r$, $r\in\mathbb{R}$

let $x_4=s$, $s\in\mathbb{R}$

x_1,x_2: pivot (basic) variable

x_3,x_4: nonpivot (free) variables

$$\begin{cases}x_1+\quad 2x_3-4x_4=1\\ \quad x_2+x_3-3x_4=2\end{cases}$$

$x_2=2-x_3+3x_4$ $\quad$ $x_1=1-2x_3+4x_4$

$x_2=2-r+3s$ $\quad$ $x_1=1-2r+4s$

$$\vec{x}=\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}1-2r+4s\\2-r+3s\\r\\s\end{bmatrix}=\underbrace{\begin{bmatrix}1\\2\\0\\0\end{bmatrix}}_{\vec{x}_p}+\underbrace{r\begin{bmatrix}-2\\-1\\1\\0\end{bmatrix}+s\begin{bmatrix}4\\3\\0\\1\end{bmatrix}}_{\vec{x}_h}$$

$$\left\{\begin{bmatrix}1\\2\\0\\0\end{bmatrix}+r\begin{bmatrix}-2\\-1\\1\\0\end{bmatrix}+s\begin{bmatrix}4\\3\\0\\1\end{bmatrix}\;\middle|\; r,s\in\mathbb{R}\right\}$$

No Solution Case!

$$\ldots\left[\begin{array}{ccc|c}1&0&0&7\\0&1&0&-3\\0&0&0&5\end{array}\right]\leftarrow 0=5\text{, false}\qquad \{\;\}\text{ or no solution}$$

Inverse of a Matrix

A: square matrix, $n\times n$ matrix

If A has an inverse, we denote it by A^{-1} and it satisfies $AA^{-1}=A^{-1}A=\underset{n\times n\text{ identity}}{I_n}$

Finding the inverse of a Matrix A

① Form the augmented matrix $[A \mid I_n]$, where A is $n \times n$

② Use Gauss-Jordan reduction to row reduce to get $[I_n \mid A^{-1}]$

If Gauss-Jordan reduction fails to produce I_n, then A^{-1} does not exist.
In that case, we say A is noninvertible or singular.

Ex: Find A^{-1}, if it exists. $A = \begin{bmatrix} 1 & 2 & 1 \\ 4 & 4 & 5 \\ 6 & 7 & 7 \end{bmatrix}$

$$[A \mid I_n] = \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 4 & 4 & 5 & 0 & 1 & 0 \\ 6 & 7 & 7 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R_2 \leftarrow (-4)R_1 + R_2} \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -4 & 1 & -4 & 1 & 0 \\ 6 & 7 & 7 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R_3 \leftarrow (-6)R_1 + R_3} \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -4 & 1 & -4 & 1 & 0 \\ 0 & -5 & 1 & -6 & 0 & 1 \end{array}\right] \xrightarrow{R_2 \leftarrow (-\frac{1}{4})R_2} \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & -\frac{1}{4} & 1 & -\frac{1}{4} & 0 \\ 0 & -5 & 1 & -6 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{R_3 \leftarrow 5R_2 + R_3} \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1/4 & 1 & -1/4 & 0 \\ 0 & 0 & -\frac{1}{4} & -1 & -\frac{5}{4} & 1 \end{array}\right] \xrightarrow{R_3 \leftarrow -4R_3} \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1/4 & 1 & -1/4 & 0 \\ 0 & 0 & 1 & 4 & 5 & -4 \end{array}\right] \xrightarrow{R_2 \leftarrow \frac{1}{4}R_3 + R_2} \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2 & \frac{4}{4} & -1 \\ 0 & 0 & 1 & 4 & 5 & -4 \end{array}\right]$$

$$\xrightarrow{R_1 \leftarrow (-2)R_2 + R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & -3 & -2 & 2 \\ 0 & 1 & 0 & 2 & 1 & -1 \\ 0 & 0 & 1 & 4 & 5 & -4 \end{array}\right] \xrightarrow{R_1 \leftarrow (-1)R_3 + R_1} \underbrace{\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -7 & -7 & 6 \\ 0 & 1 & 0 & 2 & 1 & -1 \\ 0 & 0 & 1 & 4 & 5 & -4 \end{array}\right]}_{I_3 \qquad A^{-1}}$$

$$\therefore A^{-1} = \begin{bmatrix} -7 & -7 & 6 \\ 2 & 1 & -1 \\ 4 & 5 & -4 \end{bmatrix}$$

Ex: Find A^{-1}, if it exists

ⓐ $A = \begin{bmatrix} 2 & 3 \\ -4 & -6 \end{bmatrix}$ ⓑ $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

a) $$\left[\begin{array}{cc|cc} 2 & 3 & 1 & 0 \\ -4 & -6 & 0 & 1 \end{array}\right] \xrightarrow{R_1 \leftarrow \frac{1}{2}R_1} \left[\begin{array}{cc|cc} 1 & 3/2 & 1/2 & 0 \\ -4 & -6 & 0 & 1 \end{array}\right] \xrightarrow{R_2 \leftarrow 4R_1 + R_2} \left[\begin{array}{cc|cc} 1 & 3/2 & 1/2 & 0 \\ 0 & 0 & 2 & 1 \end{array}\right] \quad \therefore A^{-1} \text{ does not exist}$$

b) $$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right] \xrightarrow{R_2 \leftarrow (-3)R_1 + R_2} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{array}\right] \xrightarrow{R_2 \leftarrow (-\frac{1}{2})R_2} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array}\right]$$

$$\xrightarrow{R_1 \leftarrow (-2)R_2 + R_1} \left[\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array}\right] \quad \therefore A^{-1} = \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix}$$

Shortcut Formula for finding Inverse of a 2x2 Matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \quad ad - bc \neq 0$$

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad A^{-1} = \frac{1}{(1)(4) - (3)(2)}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = \frac{1}{4 - 6}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix}$$

ad - bc is the determinant. A is only invertible if $\det(A) \neq 0$.

Ex: Find the transpose of A.

ⓐ $\underset{3\times3}{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ ⓑ $\underset{2\times3}{A} = \begin{bmatrix} 1 & 2 & -5 \\ 11 & 0 & 32 \end{bmatrix}$

ⓐ $\underset{3\times3}{A^T} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$ ⓑ $\underset{3\times2}{A^T} = \begin{bmatrix} 1 & 11 \\ 2 & 0 \\ -5 & 32 \end{bmatrix}$

Vector Spaces and Subspaces

Vectors: 2 notions

* Geometric notion: arrow vectors, used in Calculus, physics, etc
* Linear algebra: broader notion that applies to data science, ML/AI, etc.
  (Transcends the geometric arrow vector idea)

Vector: object that satisfies the following:

- If you add two instances of this object, it produces an object of the same type (closure under addition).
- If you multiply the object by a scalar, you still get an object of the same type (closure under scalar multiplication)

Examples: arrow vectors, column vectors, matrices, polynomials, functions, etc

Vector Space

A vector space V is a nonempty collection of objects called vectors for which are defined the operations:

- Vector addition, denoted $\vec{x} + \vec{y}$
- Scalar multiplication (multiplication by a real constant), denoted $c\vec{x}$ $\begin{pmatrix} c \in \mathbb{R} \\ \vec{x} \in V \end{pmatrix}$

That satisfy the following properties for all $\vec{x}, \vec{y}, \vec{z} \in V$; $c, d \in \mathbb{R}$:

Closure properties:

① $\vec{x} + \vec{y} \in V$ (closure under addition)

② $c\vec{x} \in V$ (closure under scalar multiplication)

Addition properties:

③ $\exists$ a zero vector, $\vec{0}$, in V such that $\vec{x} + \vec{0} = \vec{x}$ (additive identity)

④ $\forall \vec{x} \in V$, $\exists -\vec{x} \in V$ such that $\vec{x} + (-\vec{x}) = \vec{0}$

⑤ $(\vec{x}+\vec{y})+\vec{z} = \vec{x}+(\vec{y}+\vec{z})$ (associativity of addition)
⑥ $\vec{x}+\vec{y} = \vec{y}+\vec{x}$ (commutativity of addition)

Scalar multiplication properties:

⑦ $1\vec{x} = \vec{x}$ (1 is scalar multiplication identity)
⑧ $c(\vec{x}+\vec{y}) = c\vec{x}+c\vec{y}$ (1st distributive property)
⑨ $(c+d)\vec{x} = c\vec{x}+d\vec{x}$ (2nd distributive property)
⑩ $c(d\vec{x}) = (cd)\vec{x}$ (associativity of scalar multiplication)

Note: The closure properties can be checked at once by verifying the following property:

$$c\vec{x}+d\vec{y} \in V \text{ whenever } \vec{x},\vec{y} \in V \text{ and } c,d \in \mathbb{R}$$

(closure under linear combination)

Examples of Vector Spaces:

$\mathbb{R}^n$: The familiar n-dimensional coordinate space is a vector space. We can designate the vectors in $\mathbb{R}^n$ as points.

$$(x_1, x_2, x_3, \dots, x_n) \quad \text{or} \quad \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$\mathbb{R}^2$: The space of all ordered pairs. (2-vectors)

$\mathbb{R}^3$: The space of all ordered triples. (3-vectors)

M_{mn}: The space of all mxn matrices. $(\mathbb{R}^{m\times n})$.
(e.g. M_{22} or $\mathbf{M}$ or $\mathbb{R}^{2\times 2}$: space of 2x2 matrices)

$\mathbb{P}$: The space of all polynomials.

$\mathbb{P}_n$: The space of all polynomials of degree $\leq n$.

Vector Subspace Theorem

A nonempty subset W of a vector space V is a subspace of V if it is closed under addition and scalar multiplication.

ie (i) For any $\vec{u},\vec{v} \in W$, $\vec{u}+\vec{v} \in W$
(ii) For any $\vec{u} \in W$, $c \in \mathbb{R}$, $c\vec{u} \in W$

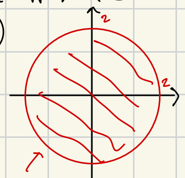
A subspace is a subset that is a vector space.

Note: The zero vector of V must belong to a subspace W.
(If a subset does not include $\vec{0}$, it cannot be a subspace)

Trivial subspaces: $\mathbf{Z} = \{\vec{0}\}$ and the vector space itself
the zero subspace

Ex: Which of the following subsets are subspaces of $\mathbb{R}^2$?

(a)

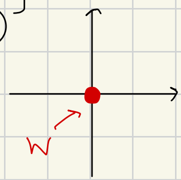


$(0,2)+(2,0)=(2,2)$

no, $(2,2) \notin W$

not closed under addition or scalar multiplication

(b)



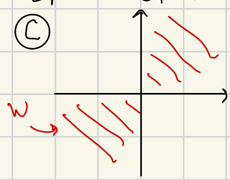
$(0,0)+(0,0)=(0,0)$

$3(0,0)=(0,0)$

yes. $(0,0) \in W$

$W = \mathbf{Z} = \{\vec{0}\} = \{(0,0)\}$

(c)



$(1,1)+(-1,-1)=(0,0)$

$3(3,3)=(9,9)$

$(0,0),(9,9) \in W$

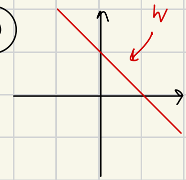
$(1,10) \in W$

$(-10,-1) \in W$

$(-9,9) \notin W$

No, not closed under addition

(d)



no, $(0,0) \notin W$

$\vec{0}_{\mathbb{R}^2} \notin W$

(e)

$(2,2)+(-1,-1)=(1,1)$

$5(-2,-2)=(-10,-10)$

yes, $(1,1),(-10,-10) \in W$

Ex: Which of the following subsets of $\mathbb{R}^2$ are subspaces?

(a) $W = \{(x,y) \mid x = y\}$ (b) The set of (a,b) such that $a=1$.

(c) The set of vectors (x_1,x_2) such that $x_2 = 0$

(d) $A = \{(a,b) \mid a^2+b^2=4\}$ (e) A linear combinations of $\vec{u}=(1,2)$ and $\vec{v}=(2,4)$

(f) All linear combinations of $\vec{u}=(1,2)$ and $\vec{v}=(-1,2)$

a) proof

b) No, because $0_{\mathbb{R}^2} \notin A$ since $0^2+0^2 \neq 4$

c) proof

d) No, because $0_{\mathbb{R}^2} \notin$ the set since $(0,0)$ does not satisfy $a=1$.
a

e) proof

f) proof

a) $W = \{(x,y) \mid x = y\}$

$0_{\mathbb{R}^2} = (0,0) \in W$ because $0 = 0$

Closure under addition?

Let $\vec{u}_1 = (x_1, y_1)$, $\vec{u}_2 = (x_2, y_2) \in W$

$\vec{u}_1 = (x_1, x_1)$ because $x_1 = y_1$

$\vec{u}_2 = (x_2, x_2)$ because $x_2 = x_2$

$\vec{u}_1 + \vec{u}_2 = (x_1 + x_2, x_1 + x_2) \in W$ because $x = y$

$\therefore$ W is closed under addition I

Closure under scalar multiplication?

Let $\vec{u}_1 = (x_1, y_1) \in W$, $c \in \mathbb{R}$

$\vec{u}_1 \in W$ means $\vec{u}_1 = (x_1, x_1)$

$c\vec{u}_1 = c(x_1, x_1)$

$= (cx_1, cx_1) \in W$ because $x = y$

$\therefore$ W is closed under scalar multiplication II

I & II $\Rightarrow$ W is a subspace of $\mathbb{R}^2$

e) W: set of linear combinations of $\vec{u} = (1,2)$ and $\vec{v} = (2,4)$

$0_{\mathbb{R}^2} \in W$ because $(0,0) = 0(1,2) + 0(2,4)$

Closure under addition?

Let $\vec{u}_1 = (x_1, y_1)$, $\vec{u}_2 = (x_2, y_2) \in W$, $c_1, c_2, d_1, d_2 \in \mathbb{R}$

$\vec{u}_1 \in W$ means $\vec{u}_1 = c_1\vec{u} + d_1\vec{v}$

$\vec{u}_2 \in W$ means $\vec{u}_2 = c_2\vec{u} + d_2\vec{v}$

$\vec{u}_1 + \vec{u}_2 = c_1\vec{u} + c_2\vec{u} + d_1\vec{v} + d_2\vec{v}$

$= (c_1 + c_2)\vec{u} + (d_1 + d_2)\vec{v}$

$\vec{u}_1 + \vec{u}_2$ is a linear combination of $\vec{u}$ & $\vec{v}$

$\therefore$ W is closed under addition I

Closure under scalar multiplication?

Let $\vec{u}_1 = (x_1, y_1) \in W$, $c, c_1, d_1 \in \mathbb{R}$

$\vec{u}_1 \in W$ means $\vec{u}_1 = c_1\vec{u} + d_1\vec{v}$

$c\vec{u}_1 = \underbrace{(cc_1)}_{c}\vec{u} + \underbrace{(cd_1)}_{d}\vec{v}$

$c\vec{u}_1 \in W$

$c\vec{u}_1$ is a linear combination of $\vec{u}$ & $\vec{v}$

$\therefore$ W is closed under scalar multiplication II

I & II $\Rightarrow$ W is a subspace of $\mathbb{R}$

Ex Which of the following subsets of $\mathbb{R}^3$ are subspaces?

a) all vectors (a,b,c) with $a = -b$

b) The plane of vectors (b_1, b_2, b_3) with $b_3 = 5$

c) all linear combinations of $\vec{u}_1 = (1,1,1)$ and $\vec{v} = (-1,2,4)$

d) all vectors of the form (a,b,c) satisfying $a - b - c = 0$

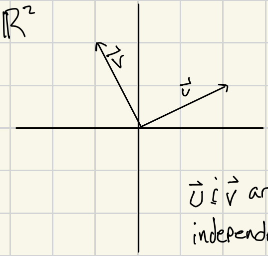
a) Subspace

b) No, because $0_{\mathbb{R}^3} \notin W$ since $(0,0,0)$ does not satisfy $b_3 = 5$

c) Subspace ← planes through the origin → d) Subspace

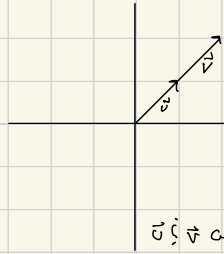
Linear Dependence / Independence

$\mathbb{R}^2$



The linear combinations of $\vec{u}$ & $\vec{v}$ fill the plane.
The span of $\vec{u}$ & $\vec{v}$, $\operatorname{span}\{\vec{u},\vec{v}\} = \mathbb{R}^2$

$\vec{u}$ & $\vec{v}$ are linearly independent

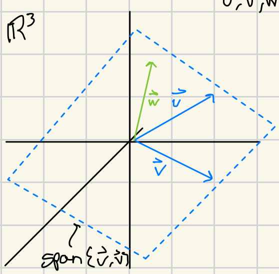


The linear combinations of $\vec{u}$ & $\vec{v}$ fill a line.

$\vec{u}$ & $\vec{v}$ are linearly dependent.

$\mathbb{R}^3$

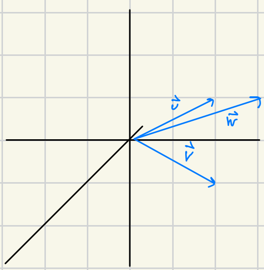
$\vec{u},\vec{v},\vec{w}$



$\vec{u}$ and $\vec{v}$ are linearly independent

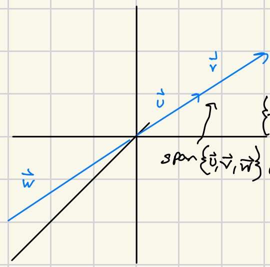
$\vec{u},\vec{v}$, and $\vec{w}$ are linearly independent
$\{\vec{u},\vec{v},\vec{w}\}$ is linearly independent

$\operatorname{span}\{\vec{u},\vec{v},\vec{w}\} = \mathbb{R}^3$



$\vec{w}$ is a linear combination of $\vec{u}$ & $\vec{v}$. (There is a non-trivial linear combination of $\vec{u},\vec{v}$, & $\vec{w}$ that produces $\vec{0}$)

$\vec{u}$ & $\vec{v}$ are linearly independent.
$\vec{u},\vec{v}$, and $\vec{w}$ are linearly dependent
$\{\vec{u},\vec{v},\vec{w}\}$ is linearly dependent.



$\vec{u},\vec{v}$, and $\vec{w}$ are linearly dependent.
$\{\vec{u},\vec{v},\vec{w}\}$ is linearly dependent

$\operatorname{span}\{\vec{u},\vec{v},\vec{w}\}$

$\mathbb{R}^2$ (two vectors)

Dependent: one vector is a scalar multiple of the other (they span a line)

Independent: one vector is not a scalar multiple of the other (they span a line)

$\mathbb{R}^3$ (three vectors)

Dependent: They span a line (pairwise each vector is a scalar multiple of another)

They span a plane

Independent: No vector is a linear combination of the other two. (They span $\mathbb{R}^3$)

$$c\vec{u} + d\vec{v} \overset{?}{=} \vec{w} \qquad c_1\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} + c_2\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + c_3\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$c\vec{u} + d\vec{v} - \vec{w} = \vec{0}$$

In general, a set of vectors is linearly independent if the only linear combination of the vectors that produces the zero vector is the trivial linear combination (all scalars are zero)

Ex: Are the following vectors linearly independent? Support your answer.

a) $\vec{x}_1 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}, \vec{x}_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \vec{x}_3 = \begin{bmatrix} 3 \\ -3 \\ 8 \end{bmatrix}$

Plan A: Try to find a non-trivial linear combination that produces $\vec{0}$. If successful, then the vectors ar linearly dependent

$$(2)\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} + (-1)\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} + (-1)\begin{bmatrix} 3 \\ -3 \\ 8 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$\therefore$ The vectors are linearly dependent.

b) $\vec{u} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}, \vec{v} = \begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix}, \vec{w} = \begin{bmatrix} 5 \\ 0 \\ 3 \end{bmatrix}$

$$\begin{bmatrix} 1 & -1 & 5 \\ 0 & 3 & 0 \\ -2 & 0 & 3 \end{bmatrix} \xrightarrow{R_3 \leftarrow 2R_1 + R_3} \begin{bmatrix} 1 & -1 & 5 \\ 0 & 3 & 0 \\ 0 & -2 & 13 \end{bmatrix} \xrightarrow{R_2 \leftarrow \frac{1}{3}R_2} \begin{bmatrix} 1 & -1 & 5 \\ 0 & 1 & 0 \\ 0 & -2 & 13 \end{bmatrix}$$

$$\xrightarrow{R_3 \leftarrow 2R_2} \begin{bmatrix} 1 & -1 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 13 \end{bmatrix} \xrightarrow{R_3 \leftarrow \frac{1}{13}R_3} \begin{bmatrix} 1 & -1 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$\therefore$ The vectors are linearly independent

c) $\vec{x} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}, \vec{y} = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix}$

$$\frac{2}{-2} = \frac{-1}{1} \neq \frac{3}{3}$$

$\vec{x}$ is not a scalar multiple of $\vec{y}$

$\therefore$ The vectors are linearly independent.

Plan B: Group the vectors into a matrix and convert to REF. If all columns are pivot columns, then the vectors are linearly independent.

$$\begin{bmatrix} 2 & 1 & 3 \\ -1 & 1 & -3 \\ 3 & -2 & 8 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow (-1)R_2} \begin{bmatrix} 1 & -1 & 3 \\ 2 & 1 & 3 \\ 3 & -2 & 8 \end{bmatrix} \xrightarrow{R_2 \leftarrow (-2)R_1 + R_2} \begin{bmatrix} 1 & -1 & 3 \\ 0 & 3 & -3 \\ 3 & -2 & 8 \end{bmatrix}$$

$$\xrightarrow{R_3 \leftarrow R_3 + (-3)R_1} \begin{bmatrix} 1 & -1 & 3 \\ 0 & 3 & -3 \\ 0 & 1 & -1 \end{bmatrix} \xrightarrow{R_2 \leftarrow (-3)R_3 + R_2} \begin{bmatrix} 1 & -1 & 3 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix} 1 & -1 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

$\therefore$ The vectors are linearly dependent

Determinant and Trace

Determinant of a 2x2 matrix

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, 2x2 matrix

The determinant of A, denoted by det(A) or |A|, is given by:

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

e.g. $\vec{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$, $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$, Find det(A)

$$|A| = \begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix} = 1(4) - 2(2) = 0$$ linearly dependent

$\vec{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$

$$|A| = \begin{vmatrix} 1 & 1 \\ 2 & -2 \end{vmatrix} = -2(1) - (2)(1) = -2 - 2 = -4$$ linearly independent

Determinant: signed area of the parallelogram formed by $\vec{u}$ & $\vec{v}$ (columns of A)

Minors and Cofactors of a Matrix

Every element a_{ij} of an $m \times n$ matrix has an associated minor and cofactor.

- Minor M_{ij} of a_{ij} is the determinant of the matrix obtained after deleting the i^{th} row and j^{th} column of the matrix
- Cofactor C_{ij} of a_{ij} is the scalar given by $C_{ij} = (-1)^{i+j} M_{ij}$
  - if $i+j$ are even, $C_{ij} = M_{ij}$
  - if $i+j$ are odd, $C_{ij} = -M_{ij}$

3x3 $\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}$ signs of cofactors

Ex: $A = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 1 & 3 \\ 0 & 1 & 2 \end{bmatrix}$ (a) Find M_{12}, M_{22}, M_{31} (b) Find C_{12}, C_{22}, C_{31}

a) $M_{12} = \begin{vmatrix} 2 & 3 \\ 0 & 2 \end{vmatrix} = 4 - 0 = 4$, $M_{22} = \begin{vmatrix} 3 & -1 \\ 0 & 2 \end{vmatrix} = 6 - 0 = 6$, $M_{31} = \begin{vmatrix} 1 & -1 \\ 1 & 3 \end{vmatrix} = 3 + 1 = 4$

b) $1+2=3$ $C_{12} = -M_{12} = -4$, $2+2=4$ $C_{22} = M_{22} = 6$, $3+1=4$ $C_{31} = M_{31} = 4$

Finding Determinants by Expansion by Cofactors (Laplace Expansion)

Let A be a $n \times n$ matrix. Expansion by the i^{th} row

$$A = \begin{bmatrix} \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \\ \end{bmatrix} \leftarrow i^{th} \text{ row}$$

$$|A| = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} = \sum_{k=1}^{n} a_{ik}C_{ik}$$

We can expand using any row or column

$$= a_{i1}(-1)^{i+1}M_{i1} + a_{i2}(-1)^{i+2}M_{i2} + \cdots + a_{in}(-1)^{i+n}M_{in}$$

Ex: Find $\det(A)$, expand by ⓐ row 1, ⓑ row 3, and ⓒ column 1. $A=\begin{bmatrix} 3 & 1 & -1 \\ 2 & 1 & 3 \\ 0 & 1 & 2 \end{bmatrix}$

a) $|A| = 3\begin{vmatrix} 1 & 3 \\ 1 & 2 \end{vmatrix} - 1\begin{vmatrix} 2 & 3 \\ 0 & 2 \end{vmatrix} - 1\begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} = 3[2-3] - 1[4-0] - [2-0] = 3(-1) - (4) - 2 = -9$

b) $|A| = 0\begin{vmatrix} 1 & -1 \\ 1 & 3 \end{vmatrix} - 1\begin{vmatrix} 3 & -1 \\ 2 & 3 \end{vmatrix} + 2\begin{vmatrix} 3 & 1 \\ 2 & 1 \end{vmatrix} = 0(3+1) - 1(9+2) + 2(3-2) = 0 - 11 + 2 = -9$

c) $|A| = 3\begin{vmatrix} 1 & 3 \\ 1 & 2 \end{vmatrix} - 2\begin{vmatrix} 1 & -1 \\ 1 & 2 \end{vmatrix} + 0\begin{vmatrix} 1 & -1 \\ 1 & 3 \end{vmatrix} = 3[2-3] - 2[2+1] + 0[3+1] = 3(-1) - 2(3) + 0$
$= -3 - 6 + 0 = -9$

Ex: Find the determinant.

$A = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 5 & 0 & 1 & -2 \\ 2 & 0 & 3 & 1 \\ 1 & 0 & 2 & -3 \end{bmatrix}$

$|A| = 2\begin{vmatrix} 5 & 1 & -2 \\ 2 & 3 & 1 \\ 1 & 2 & -3 \end{vmatrix} = -2\left[\begin{vmatrix} 1 & -2 \\ 3 & 1 \end{vmatrix} - 2\begin{vmatrix} 5 & -2 \\ 2 & 1 \end{vmatrix} - 3\begin{vmatrix} 5 & 1 \\ 2 & 3 \end{vmatrix}\right] = 2[(1+6) - 2(5+4) - 3(15-4)]$

$= 2[-14 - 18 - 33] = -2[-57] = 114$

Determinant is 0 if: matching row pairs, matching column pairs, rows/columns are multiples of others.

Properties of Determinants

- $|AB| = |A||B|$
- If $|A| \neq 0$, $|A^{-1}| = \dfrac{1}{|A|}$
- If one row (column) consists entirely of zeroes then $\det(A) = 0$
- If two rows (or columns) are identical, then $\det(A) = 0$
- If two rows (or columns) are linearly dependent, then $\det(A) = 0$

Determinant of a Triangular or Diagonal Matrix

$A = \begin{bmatrix} 4 & 5 & 3 & 1 \\ 0 & 1 & -3 & 2 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 5 \end{bmatrix}$ upper triangular matrix

$B = \begin{bmatrix} 7 & 0 & 0 \\ -3 & 1 & 0 \\ 2 & -11 & 10 \end{bmatrix}$ lower triangular matrix

$C = \begin{bmatrix} 5 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ diagonal matrix

Determinant is simply the product of the main diagonal entries

$|A| = 4(1)(-2)(5) = -40$

$|B| = 7(1)(10) = 70$

$|C| = 5(-2)(3) = -30$

For an $n \times n$ matrix A, the following statements are equivalent:

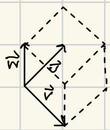
- A is invertible
- $|A| \neq 0$
- columns of A are linearly independent
- every column of A is a pivot column after converting to REF
- The system $A\vec{x} = \vec{0}$ has a unique solution $\vec{x} = \vec{0}$
- the system $A\vec{x} = \vec{b}$ has a unique solution $\vec{x} = A^{-1}\vec{b}$

Trace: The sum of the main diagonal entries

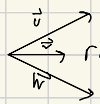
e.g. $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ $tr(A) = 1+5+9 = 15$ or $trace(A)$

Remark: The determinant of a matrix A is the signed volume of the parallelepiped formed by the column vectors or row vectors of A

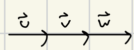
$$A = \begin{array}{c} \begin{matrix} \vec{u} & \vec{v} & \vec{w} \end{matrix} \\ \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \end{array}$$



rank 3



rank 2



rank 1

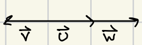
Rank: The rank of an $n \times n$ matrix A is the number of pivot columns in the REF of A. This number corresponds to the number of linearly independent columns/rows of A. Denoted by $rk(A)$ or $rank(A)$

e.g. $A = \begin{bmatrix} 5 & 1 & 9 \\ -1 & 1 & -3 \\ 4 & -2 & 10 \end{bmatrix} \xrightarrow{} \overset{REF}{\begin{bmatrix} 1 & -1 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}}$ $rk(A) = 2$

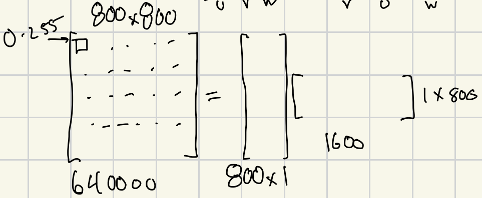
only 2 columns independent

$B = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$ $rk(B) = 3$ All columns are independent

Rank 1 $A = \begin{array}{c} \begin{bmatrix} 1 & -1 & 2 \\ 2 & -2 & 4 \\ 3 & -3 & 6 \end{bmatrix} \\ \begin{matrix} \vec{u} & \vec{v} & \vec{w} \end{matrix} \end{array}$



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



Span: The span of a set $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$ of vectors in a vector space V denoted by $span\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$, is the set of all linear combinations of these vectors.

Note: if $span\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\} = V$, we say that the set span (or is a generating set of) the vector space V.

Basis: The set $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$ is a basis for the vector space V provided that:

(i) $\text{span}\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\} = V$

and

(ii) $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$ is linearly independent

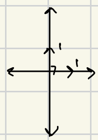
Basis: smallest set of vectors that span the entire vector space (no redundancy)

Dimension: The dimension of a vector space is the number of vectors in any given basis for that vector space.

$\dim \mathbb{R}^2 = 2$, $\dim \mathbb{R}^3 = 3$, $\dim \mathbb{R}^n = n$

Standard basis (canonical)

$\mathbb{R}^2$: $S = \left\{ \underset{\vec{e}_1}{\begin{bmatrix}1\\0\end{bmatrix}}, \underset{\vec{e}_2}{\begin{bmatrix}0\\1\end{bmatrix}} \right\}$ $\quad \vec{e}_1 = \hat{i}$, $\vec{e}_2 = \hat{j}$



$\mathbb{R}^3$: $S = \left\{ \underset{\vec{e}_1}{\overset{\hat{i}}{\begin{bmatrix}1\\0\\0\end{bmatrix}}}, \underset{\vec{e}_2}{\overset{\hat{j}}{\begin{bmatrix}0\\1\\0\end{bmatrix}}}, \underset{\vec{e}_3}{\overset{\hat{k}}{\begin{bmatrix}0\\0\\1\end{bmatrix}}} \right\}$

- vectors are orthogonal
- unit vectors
- form a basis for $\mathbb{R}^2$

$\mathbb{R}^n$: $S = \left\{ \underset{\vec{e}_1}{\begin{bmatrix}1\\0\\0\\\vdots\\0\end{bmatrix}}, \underset{\vec{e}_2}{\begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix}}, \cdots, \underset{\vec{e}_n}{\begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}} \right\}$

orthonormal

orthogonal $\quad$ normalized (unit)

Ex: Determine whether each of the following sets forms a basis for $\mathbb{R}^2$.

a) $\left\{\begin{bmatrix}1\\1\end{bmatrix}\right\}$ $\quad$ b) $\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}\right\}$ $\quad$ c) $\left\{\begin{bmatrix}-1\\2\end{bmatrix}, \begin{bmatrix}3\\-6\end{bmatrix}\right\}$ $\quad$ d) $\left\{\begin{bmatrix}1\\-1\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}0\\-3\end{bmatrix}\right\}$

a) No, $\text{span}\left\{\begin{bmatrix}1\\1\end{bmatrix}\right\} \neq \mathbb{R}^2$ $\quad$ b) yes, $\text{span}\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}\right\} = \mathbb{R}^2$, $\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}\right\}$ is linearly independent

c) No, $\text{span}\left\{\begin{bmatrix}-1\\2\end{bmatrix}, \begin{bmatrix}3\\-6\end{bmatrix}\right\} \neq \mathbb{R}^2$, $\left\{\begin{bmatrix}-1\\2\end{bmatrix}, \begin{bmatrix}3\\-6\end{bmatrix}\right\}$ is linearly dependent

d) no, $\text{span}\left\{\begin{bmatrix}1\\-1\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}0\\-3\end{bmatrix}\right\} = \mathbb{R}^2$, $\left\{\begin{bmatrix}1\\-1\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}0\\-3\end{bmatrix}\right\}$ is linearly dependent

Ex: Determine whether each of the following sets forms a basis for $\mathbb{R}^3$

a) $\left\{\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\-1\\1\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}\right\}$ $\quad$ b) $\left\{\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}-1\\1\\-1\end{bmatrix}, \begin{bmatrix}0\\3\\2\end{bmatrix}\right\}$ $\quad$ c) $\left\{\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}\right\}$

a) $\dim \mathbb{R}^3 = 3$, the set has 3 vectors (I)

$$\begin{vmatrix}1&1&1\\1&-1&0\\0&1&1\end{vmatrix} = 0\begin{vmatrix}1&1\\-1&0\end{vmatrix} - \begin{vmatrix}1&1\\1&0\end{vmatrix} + \begin{vmatrix}1&1\\1&-1\end{vmatrix}$$

$$= -\underbrace{(0-1)}_{1} + \underbrace{(-1-1)}_{-2} = -1$$

$\therefore$ The vectors are linearly independent (II)

(I) & (II) $\Rightarrow$ the set forms a basis for $\mathbb{R}^3$

b) $\dim \mathbb{R}^3 = 3$ the set has 3 vectors (I)

$$\begin{vmatrix}1&-1&0\\2&1&3\\3&-1&2\end{vmatrix} = 1\begin{vmatrix}1&3\\-1&2\end{vmatrix} - (-1)\begin{vmatrix}2&3\\3&2\end{vmatrix} + 0\begin{vmatrix}2&1\\3&-1\end{vmatrix}$$

$$= (2-(-3)) + (4-9) = 5 + -5 = 0$$

$\therefore$ The set is linearly dependent

$\therefore$ The set does not form a basis for $\mathbb{R}^3$

c) $\dim \mathbb{R}^3 = 3$, the set has 2 vectors

$\begin{vmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{vmatrix}$ does not exist

$\therefore$ The set does not form a basis

Ex: Fill in the blanks to complete the given rank 1 matrix.

$$\begin{bmatrix} 5 & -3 & 2 \\ 10 & -6 & 4 \\ -15 & 9 & -6 \end{bmatrix}$$

Low Rank Approximation

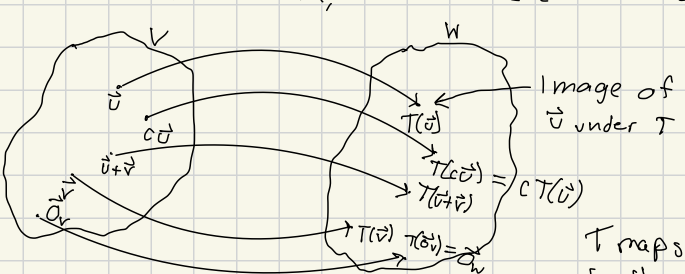
Linear Transformation

Def: A linear transformation (or linear map/transform) T on a vector space V to a vector space W is a function, $T: V \to W$, that preserves vector addition and scalar multiplication:

(i) $T(\vec{u}+\vec{v}) = T(\vec{u}) + T(\vec{v})$, $\forall \vec{u}, \vec{v} \in V$

(ii) $T(c\vec{u}) = c\,T(\vec{u})$, $\forall \vec{u} \in V, c \in \mathbb{R}$

V is called the domain, W is called the codomain (or target)



The image of the sum is the sum of the image.

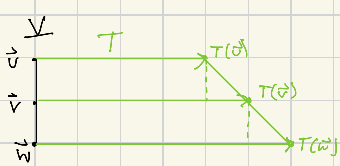
T maps the zero vector of V to the zero vector of W

Remarks

- A linear transformation transforms (or maps) lines to lines, equally spaced points to equally spaced points, and the zero vector to the zero vector.

$\vec{v}$ is the midpoint of $\vec{u}$ & $\vec{v}$

distance between points $= a$



distance between points is clearly more $(a\sqrt{2})$ but spacing remains equal

- Suppose $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ are basis vectors of V and $\vec{u} \in V$, so $\vec{u} = c_1\vec{v}_1 + c_2\vec{v}_2 + \dots + c_n\vec{v}_n$, where $c_1, c_2, \dots, c_n$ are unique.

$$T(\vec{u}) = T(c_1\vec{v}_1 + c_2\vec{v}_2 + \dots + c_n\vec{v}_n) = T(c_1\vec{v}_1) + T(c_2\vec{v}_2) + \dots + T(c_n\vec{v}_n)$$
$$= c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \dots + c_nT(\vec{v}_n) \leftarrow \text{image of basis}$$

$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \vec{u}_B$ coordinates of $\vec{u}$ relative to basis B

$B = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$

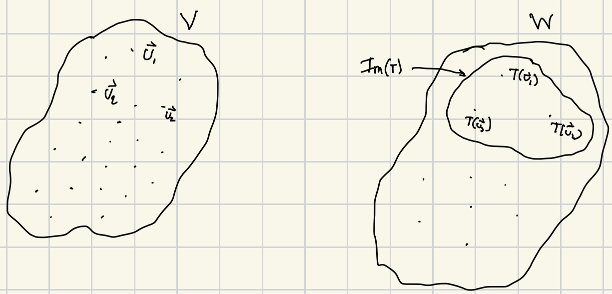
If we know how to transform the basis vectors, then we know how to transform any vector in the domain V.

Image of a Linear Transformation

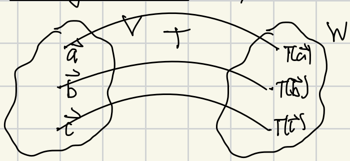
The image (or range) of a linear transformation $T: V \to W$ is the set of all vectors in W which are images of vectors in V

$$Im(T) = \{\vec{w} \in W \mid \vec{w} = T(\vec{u}), \vec{u} \in V\}$$

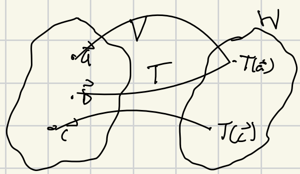
$Im(T)$ is a subset of W



Injective: If for all $\vec{u}, \vec{v} \in V$, $T(\vec{u}) = T(\vec{v})$ iff $\vec{u} = \vec{v}$



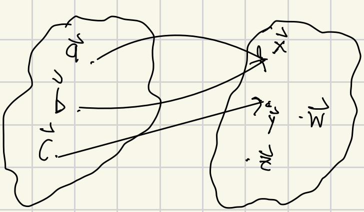
injective
surjective
bijective



not injective
surjective
not bijective

Surjective: If $Im(T) = W$. Every element of W is an image of an element in V

Bijective: Both injective and surjective.



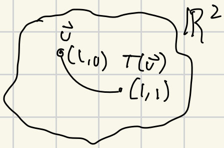
not injective
not surjective
not bijective

A bijective T can be reversed. We can find a linear mapping $T^{-1}: V \to V$ such that $T(T^{-1}(\vec{v})) = T^{-1}(T(\vec{v})) = \vec{v}$

Ex: Compute the image of $\vec{v}$ under T and find the vectors, if any, that are mapped to $\vec{w}$. $T(x,y)=(x+y,x)$, $\vec{v}=(1,0)$, $\vec{w}=(3,1)$

$T(\vec{v})=T(1,0)=(1+0,1)=\boxed{(1,1)}$

$T:\mathbb{R}^2\to\mathbb{R}^2$



$T(x,y)=(3,1)$

$(x+y,\ x)=(3,1) \Rightarrow x+y=3,\ x=1 \Rightarrow y=2$

$\boxed{\{(1,2)\}}$

$A=\begin{bmatrix}1&1\\1&0\end{bmatrix}$, $|A|\neq 0$, $|A|=0-1=-1$ $\qquad T\left(\begin{bmatrix}x\\y\end{bmatrix}\right)=\begin{bmatrix}x+y\\x\end{bmatrix}$

Every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be represented as $T(\vec{v})=A\vec{v}$, where $\vec{v}\in\mathbb{R}^n$ and A is called the standard matrix associated with T and is defined by: $\quad T:\mathbb{R}^n\to\mathbb{R}^m$

$$A_{m\times n}=\begin{bmatrix}|&|&&|\\T(\vec{e}_1)&T(\vec{e}_2)&\cdots&T(\vec{e}_n)\\|&|&&|\end{bmatrix}$$

Standard basis of $\mathbb{R}^n$ $\quad S=\{\vec{e}_1,\vec{e}_2,\dots,\vec{e}_n\}$

The columns are the images of the standard basis vectors of $\mathbb{R}^n$

Ex: $T(x,y)=(x+y,x)$ $\quad T:\mathbb{R}^2\to\mathbb{R}^2$ $\quad S=\{(1,0),(0,1)\}=\left\{\overset{\vec{e}_1}{\begin{bmatrix}1\\0\end{bmatrix}},\overset{\vec{e}_2}{\begin{bmatrix}0\\1\end{bmatrix}}\right\}$ $\quad \vec{v}=\begin{bmatrix}1\\0\end{bmatrix}$

$T(\vec{e}_1)=\begin{bmatrix}1\\1\end{bmatrix}$, $T(\vec{e}_2)=\begin{bmatrix}1\\0\end{bmatrix}$

$A=\begin{bmatrix}1&1\\1&0\end{bmatrix}$ $\qquad T(\vec{v})=A\vec{v}=\underset{2\times2}{\begin{bmatrix}1&1\\1&0\end{bmatrix}}\underset{2\times1}{\begin{bmatrix}1\\0\end{bmatrix}}=\begin{bmatrix}1\\1\end{bmatrix}$

$T\left(\begin{bmatrix}x\\y\end{bmatrix}\right)=\begin{bmatrix}x+y\\y\end{bmatrix}$

$=\begin{bmatrix}1&1\\1&0\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}$ $\quad A\vec{x}$

$=\begin{bmatrix}x+y\\y\end{bmatrix}$

Matrix becomes a change object.

Ex: For each linear transformation $T:\mathbb{R}^m\to\mathbb{R}^n$. Determine the standard matrix such that $T(\vec{v})=A\vec{v}$

a) $T(x,y)=(x+2y,-y)$ $\quad n\times m$

b) $T(\underbrace{x_1,x_2,x_3}_{n})=(\underbrace{x_1-3x_2,\ x_1+x_2+x_3,\ 2x_1-x_3,\ -4x_3}_{m})$

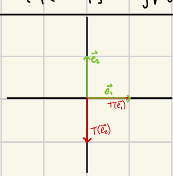
a) $A=\begin{bmatrix}1&2\\0&-1\end{bmatrix}$ $\quad T:\mathbb{R}^2\to\mathbb{R}^2$

b) $A=\begin{bmatrix}1&-3&0\\1&1&1\\2&0&-1\\0&0&-4\end{bmatrix}$ $\quad T:\mathbb{R}^3\to\mathbb{R}^4$

Linear Transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$

a) $\begin{bmatrix}1&0\\0&-1\end{bmatrix}$ $\quad \vec{e}_1=\begin{bmatrix}1\\0\end{bmatrix}$ $\quad \vec{e}_2=\begin{bmatrix}0\\1\end{bmatrix}$

$T(\vec{e}_1)\ T(\vec{e}_2)$



Reflection across x-axis

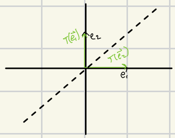
b) $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ reflection across the y-axis

c) $\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ clockwise rotation by θ around the origin

e.g. $\theta = \frac{\pi}{4}$ $\begin{bmatrix} \cos\frac{\pi}{4} & \sin\frac{\pi}{4} \\ -\sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ counterclockwise

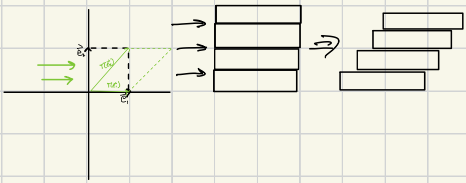
d) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$



reflection across the line $y = x$

e) Shear of 1 in x-direction

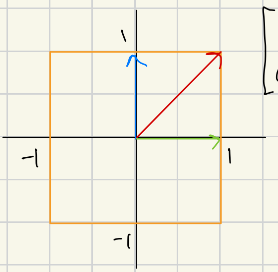
$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$



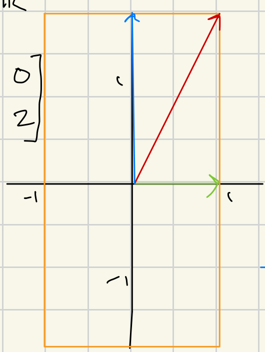
Eigen Values and Eigen vectors

- Geometric intuition
- Computation

Every LT $T: \mathbb{R}^n \to \mathbb{R}^m$



$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$



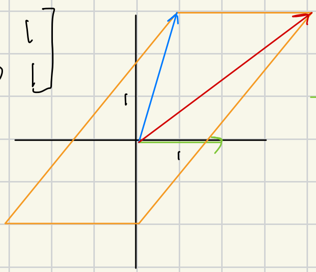
eigenvectors - same location after LT

eigenvalue - magnitude of eigenvector

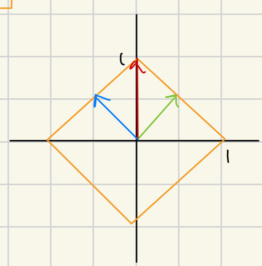
← eigenvector, eigenvalue = 1

← eigenvector, eigenvalue = 2

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



eigenvector



No eigenvectors

$\begin{bmatrix} \cos\frac{\pi}{4} & \sin\frac{\pi}{4} \\ -\sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix}$ -

Def: Let $T: V \to V$ be a linear transformation.

A nonzero vector $\vec{v} \in V$ is an eigenvector for T if there is a Scalar $\lambda \in \mathbb{R}$ such that $T(\vec{v}) = \lambda \vec{v}$

If $V = \mathbb{R}^n$, then T is represented by an $n \times n$ Matrix A and $T(\vec{v}) = Av$.

So, λ and v satisfy the equation $A\vec{v} = \lambda \vec{v}$

$V = \mathbb{R}^n$, $T: \mathbb{R}^n \to \mathbb{R}^n \Rightarrow A_{n \times n}$

λ is the eigenvalue corresponding to the eigenvector.

Note: Once an eigen vector is located, all vectors on the same line are also eigenvectors sharing the same eigenvalue.

Eigen, from German, means proper or characteristic.

<u>Finding eigen values and eigenvectors</u>

1) Write characteristic equation:

$$\det(A - \lambda I_n) = 0$$

2) Solve characteristic equation for the eigen values λ_i

3) For each λ_i, find the eigenvectors $\vec{v}_i$ by solving $(A - \lambda_i I_n)\vec{v} = \vec{0}$

$$A\vec{v} = \lambda \vec{v} \Rightarrow \underset{n \times n\, n \times 1}{A\vec{v}} - \underset{1 \times 1\, n \times 1}{\lambda \vec{v}} = \vec{0} \Rightarrow A\vec{v} - \lambda I_n \vec{v} = \vec{0} \Rightarrow (A - \lambda I_n)\vec{v} = \vec{0}$$

$\det(A - \lambda I_n) = \vec{0}$ because the system has many Solutions

Ex: Find the eigen values and eigenvectors of the given matrix.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix} \quad A - \lambda I_n = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 1-\lambda & 2 \\ 2 & -2-\lambda \end{bmatrix}$$

$$\det(A - \lambda I_n) = \begin{vmatrix} 1-\lambda & 2 \\ 2 & -2-\lambda \end{vmatrix} = 0 \Rightarrow (1-\lambda)(-2-\lambda) - 2(2) = 0$$

$$\Rightarrow -2 - \lambda + 2\lambda + \lambda^2 - 4 = 0 \Rightarrow \lambda^2 + \lambda - 6 = 0 \Rightarrow (\lambda + 3)(\lambda - 2) = 0$$

$\lambda = \{-3, 2\} \Rightarrow$

$\therefore$ Eigenvalues: $\lambda_1 = -3$, $\lambda_2 = 2$

characteristic polynomial, $p(\lambda) = \lambda^2 + \lambda - 6$

characteristic equation, $p(\lambda) = 0 \Rightarrow \lambda^2 + \lambda - 6 = 0$

$\underline{\lambda_1 = -3}$

$(A + 3I_2)\vec{v} = 0$

$\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\left[\begin{array}{cc|c} 4 & 2 & 0 \\ 2 & 1 & 0 \end{array}\right] \xrightarrow{R_1 \leftarrow \frac{1}{2}R_1} \left[\begin{array}{cc|c} 1 & 1/2 & 0 \\ 2 & 1 & 0 \end{array}\right]$

$\xrightarrow{R_2 \leftarrow (-2)R_1 + R_2} \begin{bmatrix} 1 & 1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

v_2: free / nonpivot

let $v_2 = s$, $s \in \mathbb{R}$, $s \neq 0$

$\Leftrightarrow \begin{cases} v_1 + \frac{1}{2}v_2 = 0 \\ 0 = 0 \end{cases}$

$v_1 = -\frac{1}{2}v_2 = -\frac{1}{2}s$

$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} -1/2\, s \\ s \end{bmatrix} = s\begin{bmatrix} -1/2 \\ 1 \end{bmatrix}$

let $s = 2$

$\vec{v}_1 = 2\begin{bmatrix} -1/2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$

$\underline{\lambda_2 = 2}$

$(A - 2I_2)\vec{v} = 0$

$\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\left[\begin{array}{cc|c} -1 & 2 & 0 \\ 2 & -4 & 0 \end{array}\right] \xrightarrow{R_1 \leftarrow (-1)R_1} \left[\begin{array}{cc|c} 1 & -2 & 0 \\ 2 & -4 & 0 \end{array}\right]$

$\xrightarrow{R_2 \leftarrow (-2)R_1 + R_2} \left[\begin{array}{cc|c} 1 & -2 & 0 \\ 0 & 0 & 0 \end{array}\right]$

v_1: free / non-pivot

let $v_2 = s$, $s \in \mathbb{R}$, $s \neq 0$

$\Leftrightarrow \begin{cases} v_1 - 2v_2 = 0 \\ 0 = 0 \end{cases}$

$v_1 = 2v_2 = 2s$

$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 2s \\ s \end{bmatrix} = s\begin{bmatrix} 2 \\ 1 \end{bmatrix}$

let $s = 1$

$v_2 = 1\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

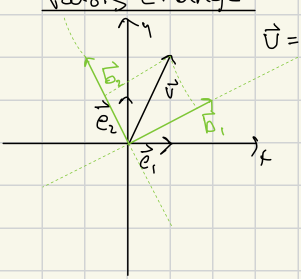
Summary:

$\lambda_1 = -3$, $\vec{v}_1 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$

$\lambda_2 = 2$, $\vec{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

Basis Change and Diagonalization

Basis change



$\vec{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$S = \left\{ \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{e_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{e_2} \right\}$

$B = \{\vec{b}_1, \vec{b}_2\}$

$\vec{u}_B = ?$

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

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

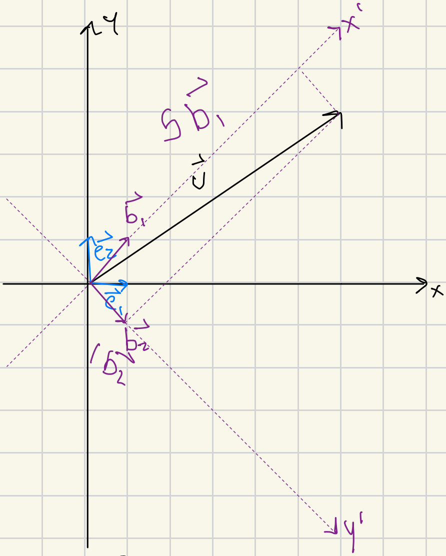
$= 1\vec{e}_1 + 2\vec{e}_2 = \vec{u}_S$

Coordinates

Let $\vec{v}$ be a vector in the finite-dimensional vector space V, with a basis $B = \{\vec{b}_1, \vec{b}_2, \cdots, \vec{b}_n\}$, then the coordinates of $\vec{v}$ relative to basis B are the unique numbers $\beta_1, \beta_2, \cdots \beta_n$ such that $\vec{V} = \beta_1\vec{b}_1 + \beta_2\vec{b}_2 + \cdots + \beta_n\vec{b}_n$.

$\vec{V}_B = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \in \mathbb{R}^n$ coordinate vector of $\vec{V}$ relative to B.

Ex: Let $S=\left\{\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix}\right\}$ standard basis in $\mathbb{R}^2$. Let $\vec{u}_S=\begin{bmatrix}6\\4\end{bmatrix}$. We seek the coordinates of $\vec{u}$ relative to the basis $B=\left\{\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}1\\-1\end{bmatrix}\right\}$ in $\mathbb{R}^2$.



$\vec{u}_B = 5\vec{b}_1 + \vec{b}_2$?

$\vec{u}_S = \beta_1\vec{b}_1 + \beta_2\vec{b}_2 = \beta_1\begin{bmatrix}1\\1\end{bmatrix} + \beta_2\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}6\\4\end{bmatrix}$

$\underbrace{\begin{bmatrix}6\\4\end{bmatrix}}_{\vec{u}_S} = \underbrace{\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}}_{M_B}\underbrace{\begin{bmatrix}\beta_1\\\beta_2\end{bmatrix}}_{\vec{u}_B} \Rightarrow \left[\begin{array}{cc|c}1 & 1 & 6\\1 & -1 & 4\end{array}\right] \xrightarrow{R_2 \leftarrow (-1)R_1 + R_2} \left[\begin{array}{cc|c}1 & 1 & 6\\0 & -2 & -2\end{array}\right]$

$\xrightarrow{R_2 \leftarrow -\frac{1}{2}R_2} \left[\begin{array}{cc|c}1 & 1 & 6\\0 & 1 & 1\end{array}\right] \xrightarrow{R_1 \leftarrow (-1)R_2 + R_1} \left[\begin{array}{cc|c}1 & 0 & 5\\0 & 1 & 1\end{array}\right]$

OR: $\vec{u}_S = M_B\vec{u}_B$

$M_B^{-1}\vec{u}_S = \underbrace{M_B^{-1}M_B}_{I_2}\vec{u}_B$

$M_B^{-1}\vec{u}_S = \vec{u}_B$

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

$M_B^{-1} = \left[\begin{array}{cc|cc}1 & 1 & 1 & 0\\1 & -1 & 0 & 1\end{array}\right] \xrightarrow{R_2 \leftarrow -1(R_1)+R_2} \left[\begin{array}{cc|cc}1 & 1 & 1 & 0\\0 & -2 & -1 & 1\end{array}\right] \xrightarrow{R_2 \leftarrow \frac{1}{2}R_2} \left[\begin{array}{cc|cc}1 & 1 & 1 & 0\\0 & 1 & \frac{1}{2} & -\frac{1}{2}\end{array}\right] \xrightarrow{R_1 \leftarrow -R_2+R_1} \left[\begin{array}{cc|cc}1 & 0 & \frac{1}{2} & \frac{1}{2}\\0 & 1 & \frac{1}{2} & -\frac{1}{2}\end{array}\right] \Rightarrow M_B^{-1}\begin{bmatrix}1/2 & 1/2\\1/2 & -1/2\end{bmatrix}$

$M_B^{-1}\vec{u}_S = \begin{bmatrix}1/2 & 1/2\\1/2 & -1/2\end{bmatrix}\begin{bmatrix}6\\4\end{bmatrix} = \begin{bmatrix}3+2\\3-2\end{bmatrix} = \begin{bmatrix}5\\1\end{bmatrix}$

$\boxed{\vec{u}_S = M_B\vec{u}_B}$ to convert from basis B to basis S, multiply by M_B

$\boxed{M_B^{-1}\vec{u}_S = \vec{u}_B}$ to convert from standard basis to basis B, multiply by M_B^{-1}

$M_S = M_B^{-1}$

Two separate bases

$B_1 \xrightarrow{M_{B_1}} S \xrightarrow{M_{B_2}^{-1}} B_2$ $\qquad \vec{v}_{B_1} \longrightarrow \vec{v}_{B_2} = M_{B_2}^{-1} M_{B_1} \vec{v}_{B_1}$

Diagonalizing a Matrix

Eigenvectors form a "good" basis (Eigenbasis). They provide a way to replace the original matrix with one that is diagonal

Diagonalization: An $n \times n$ matrix is diagonalizable iff
(i) it has n linearly independent, real, eigenvectors.
(ii) the sum of the dimensions of its unique eigenvectors is n.

Diagonalizing a Matrix: For an $n \times n$ matrix A with n linearly independent eigenvectors.
① Construct an $n \times n$ diagonal matrix D of the eigenvalues λ_i, $1 \le i \le n$
② Construct an $n \times n$ matrix P with the eigenvectors $\vec{v}_i$, $1 \le i \le n$, as columns, listed in the order corresponding to the eigenvalues in D

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

$$P = \begin{bmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ | & | & | & | \end{bmatrix}$$

$$AP = PD$$
$$APP^{-1} = PDP^{-1} \leftarrow \text{Eigendecompisition of } A$$
$$A = PDP^{-1}$$

Ex: $A = \begin{bmatrix} 4 & 1 \\ 3 & 2 \end{bmatrix}$

a) Find eigenvalues and eigenvectors
b) Find a matrix P that diagonalizes
c) Find D and P^{-1}
d) Verify $A = PDP^{-1}$

ⓐ $A - \lambda I = \begin{bmatrix} 4 & 1 \\ 3 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 4-\lambda & 1 \\ 3 & 2-\lambda \end{bmatrix}$

$\det(A - \lambda I) = \begin{vmatrix} 4-\lambda & 1 \\ 3 & 2-\lambda \end{vmatrix} = (4-\lambda)(2-\lambda) - 3$ $\quad tr(A) = 6 \quad |A| = 5$

$= 8 - 6\lambda + \lambda^2 - 3 = \lambda^2 - 6\lambda + 5 \Rightarrow 0 = \lambda^2 - 6\lambda + 5 \Rightarrow 0 = (\lambda - 5)(\lambda - 1) \Rightarrow \lambda_1 = 5, \lambda_2 = 1$

$\lambda_1 = 5$

$\left[\begin{array}{cc|c} 4-5 & 1 & 0 \\ 3 & 2-5 & 0 \end{array}\right] = \left[\begin{array}{cc|c} -1 & 1 & 0 \\ 3 & -3 & 0 \end{array}\right] \xrightarrow{R_2 \leftarrow 3R_1 + R_2} \left[\begin{array}{cc|c} -1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$

$\xrightarrow{R_1 \leftarrow -1R_1} \underbrace{\left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]}_{RREF} \Rightarrow \begin{cases} v_1 - v_2 = 0 \\ 0 = 0 \end{cases}$

$\lambda_2 = 1$

$\left[\begin{array}{cc|c} 4-1 & 1 & 0 \\ 3 & 2-1 & 0 \end{array}\right] = \left[\begin{array}{cc|c} 3 & 1 & 0 \\ 3 & 1 & 0 \end{array}\right] \xrightarrow{R_2 \leftarrow (-1)R_1 + R_2} \left[\begin{array}{cc|c} 3 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$

$\xrightarrow{R_1 \leftarrow \frac{1}{3}R_1} \underbrace{\left[\begin{array}{cc|c} 1 & 1/3 & 0 \\ 0 & 0 & 0 \end{array}\right]}_{RREF} \Rightarrow \begin{cases} v_1 + \frac{1}{3}v_2 = 0 \\ 0 = 0 \end{cases}$

let $v_2 = s,\ s \in \mathbb{R},\ s \neq 0$

$v_1 - s = 0 \Rightarrow v_1 = s$

$\begin{bmatrix} s \\ s \end{bmatrix} = s\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ let $s = 1$

$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

let $v_2 = s,\ s \in \mathbb{R},\ s \neq 0$

$v_1 = -\frac{1}{3}s$

$\begin{bmatrix} -\frac{1}{3}s \\ s \end{bmatrix} = s\begin{bmatrix} -1/3 \\ 1 \end{bmatrix}$ let $s = 3$

$\vec{v}_2 = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$

Summary:

$\lambda_1 = 5,\ \vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$\lambda_2 = 1,\ \vec{v}_2 = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$

(b) $P = \begin{bmatrix} 1 & -1 \\ 1 & 3 \end{bmatrix}$

(c) $D = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$

$$P^{-1} = \left[\begin{array}{cc|cc} 1 & -1 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{array}\right] \xrightarrow{R_2 \leftarrow (-1)R_1 + R_2} \left[\begin{array}{cc|cc} 1 & -1 & 1 & 0 \\ 0 & 4 & -1 & 1 \end{array}\right] \xrightarrow{R_2 = \frac{1}{4}R_2} \left[\begin{array}{cc|cc} 1 & -1 & 1 & 0 \\ 0 & 1 & -1/4 & 1/4 \end{array}\right] \xrightarrow{R_1 \leftarrow R_2 + R_1} \left[\begin{array}{cc|cc} 1 & 0 & 3/4 & 1/4 \\ 0 & 1 & -1/4 & 1/4 \end{array}\right]$$

$$\Rightarrow P^{-1} = \begin{bmatrix} 3/4 & 1/4 \\ -1/4 & 1/4 \end{bmatrix}$$

(d) $A = PDP^{-1}$

$$\text{RHS} = \begin{bmatrix} 1 & -1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 3/4 & 1/4 \\ -1/4 & 1/4 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & -1 \\ 5 & 3 \end{bmatrix}\begin{bmatrix} 3/4 & 1/4 \\ -1/4 & 1/4 \end{bmatrix} = \begin{bmatrix} (15/4 + 1/4) & (5/4 - 1/4) \\ (\frac{15}{4} - 3/4) & (\frac{5}{4} + 3/4) \end{bmatrix} = \begin{bmatrix} 16/4 & 4/4 \\ 12/4 & 8/4 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 3 & 2 \end{bmatrix} = \text{LHS} \quad \text{QED}$$

Benefits of Eigendecompisition

- Producing a suitable basis to express the transformation in terms of.
- Computing powers of a matrix (reduces computational overhead)

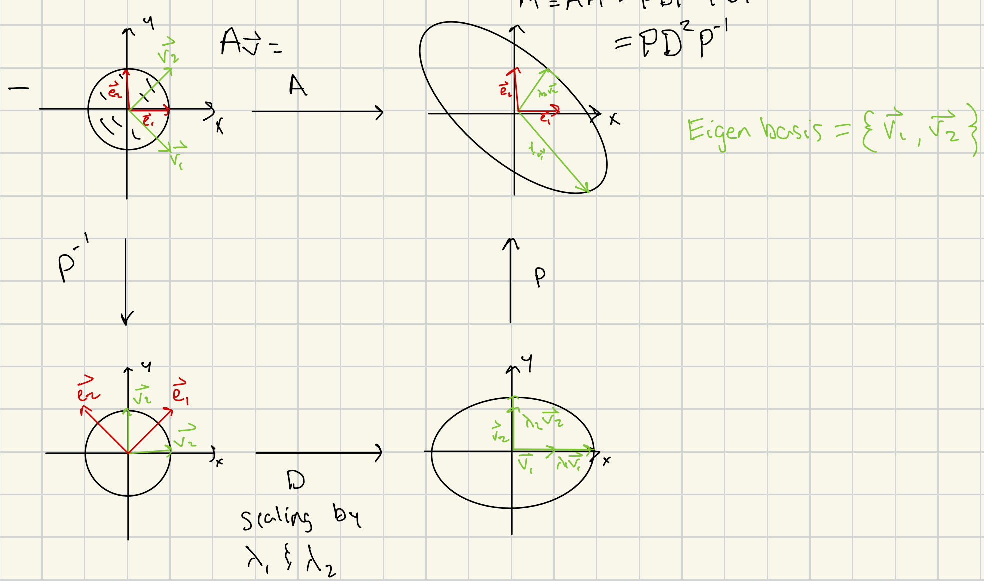
$A^2 = AA$, $A^{100} = AAA \cdots A$

$A = PDP^{-1}$

$A^2 = AA = PD\overbrace{P^{-1}P}^{I}DP^{-1}$

$= PD^2P^{-1}$

$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}^2 = \begin{bmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{bmatrix}$

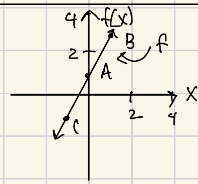


Differentiation of Univariate Functions

$\underset{\uparrow \text{ input}}{f(x)} = \underbrace{2x+1}_{\text{output}}$

$x \longrightarrow \boxed{f} \longrightarrow 2x+1$

Domain $\qquad$ Range

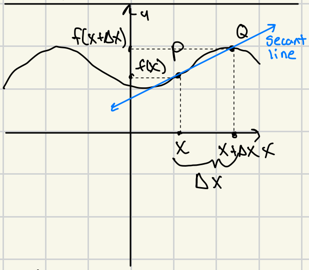


x	f(x)	
0	1	A
1	3	B
-1	-1	C

for each x, f assigns exactly one y-value.

one-to-one function

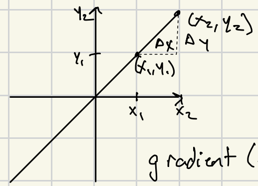
Difference Quotient

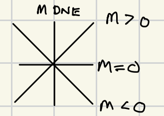


$h = \Delta x$

$P(x, f(x))$

$Q(x+h, f(x+h))$





gradient (slope) $= \dfrac{y_2 - y_1}{x_2 - x_1} = m$

Slope of secant line

$m_{secant} = \dfrac{f(x+h) - f(x)}{x+h - x} = \dfrac{f(x+h) - f(x)}{h}$ ← difference quotient

Slope of tangent line?

$\lim_{h \to 0} \dfrac{f(x+h) - f(x)}{h} = m_{tangent} = f'(x)$ ← Derivative of f at x

Ex: Compute the average rate of change of $f(x) = x^2$ over the following intervals when $x = 2$. a) $[2, 2.1]$, $[2, 2.01]$, $[2, 2.001]$, $[1.999, 2]$. Approximate the derivative.
b) Find the derivative and evaluate at $x = 2$

a) $\dfrac{f(2.1) - f(2)}{2.1 - 2} = \dfrac{(2.1)^2 - 2^2}{0.1} = 4.1$, $\dfrac{f(2.01) - f(2)}{2.01 - 2} = \dfrac{(2.01)^2 - 2^2}{0.01} = \dfrac{0.0401}{0.01} = 4.01$, $\dfrac{(2.001)^2 - 4}{0.001} = 4.001$

$\dfrac{f(1.999) - f(2)}{-0.001} = \dfrac{(1.999)^2 - 4}{-0.001} = \dfrac{3.996001 - 4}{-0.001} = \dfrac{-0.00399}{-0.001} = 3.999$

≈ 4

b) $\dfrac{d}{dx} f(x) = \dfrac{d}{dx}(x^2) = 2x$, $f'(2) = 2(2) = 4$

Power rule: $f(x) = x^n$, $f'(x) = n x^{n-1}$

Notation

y', $f'(x)$, $\dfrac{df(x)}{dx}$, $\dfrac{dy}{dx}$, $\underbrace{\dot{f}(x), \dot{y}}_{\text{typically derivatives with respect to time}}$

Rules of Differentiation

Constant rule: $f(x) = c$, $f'(x) = 0$, $\dfrac{dc}{dx} = 0$

Constant multiple: $\dfrac{d}{dx}[cf(x)] = c \dfrac{d}{dx}[f(x)]$

Sum / difference: $\dfrac{d}{dx}[f(x) + g(x)] = \dfrac{d}{dx}[f(x)] + \dfrac{d}{dx}[g(x$

Power: $\dfrac{d}{dx}[x^n] = n x^{n-1}$

Product: $\dfrac{d}{dx}[f(x)g(x)] = f'(x)g(x) + f(x)g'(x)$

Quotient: $\dfrac{d}{dx}\left[\dfrac{f(x)}{g(x)}\right] = \dfrac{f'(x)g(x) - f(x)g'(x)}{[g'(x)]^2}$

Ex: Find the derivatives.

a) $f(t) = 101$ — a) $f'(t) = 0$

b) $f(x) = 5x^2$ — b) $f'(t) = 10x$

c) $g(x) = 5x^2 - 3x + 7$ — c) $g'(t) = 10x - 3$

d) $h(t) = 17\sqrt{x} - x^{3/2} + 8x$ — d) $h(x) = 17x^{1/2} - x^{3/2} + 8x \Rightarrow h'(t) = \frac{17}{2\sqrt{x}} - \frac{3}{2}\sqrt{x} + 8$

Derivatives of Trigonometric Functions

$\frac{d}{dx}\sin x = \cos x$ — $\frac{d}{dx}\csc x = -\csc x \cot x$

$\frac{d}{dx}\sec x = \sec x \tan x$ — $\frac{d}{dx}\cos x = -\sin x$

$\frac{d}{dx}\tan x = \sec^2 x$ — $\frac{d}{dx}\cot x = -\csc^2 x$

$\tan x = \frac{\sin x}{\cos x}$

$\csc x = \frac{1}{\sin x}$

$\sec x = \frac{1}{\cos x}$

$\cot x = \frac{\cos x}{\sin x} = \frac{1}{\tan x}$

Ex: Differentiate

a) $y = 12\cos x - 20\sin x$ — b) $f(x) = \tan x - 3\cot x$ — c) $g(x) = \sec x - \csc x + 5\sin x$

a) $y' = -12\sin x - 20\cos x$ — b) $f'(t) = \sec^2 x + 3\csc^2 x$ — c) $g'(t) = \sec x \tan x + \csc x \cot x + 5\cos x$

Derivatives of Exponential and Logarithmic Functions

$\frac{d}{dx}[e^x] = e^x$ — $\frac{d}{dx}[\ln x] = \frac{1}{x}$

$\frac{d}{dx}[b^x] = b^x \ln b$ — $\frac{d}{dx}[\log_b x] = \frac{1}{x}\cdot\frac{1}{\ln b} = \frac{1}{x\ln b}$

Product Rule: $\frac{d}{dx}[f(x)g(x)] = f'(x)g(x) + f(x)g'(x)$

Quotient Rule: $\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$

Ex: Find derivatives.

a) $f(t) = t^3\cos t$

$f'(x) = 3t^2\cos t - t^3\sin t$

b) $g(x) = \frac{\sin x}{\cos x}$

$g'(t) = \frac{(\sin x)'\cos x - \sin x(\cos x)'}{\cos^2 x}$

$= \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = (\cos^2 x + \sin^2 x)\cdot\sec^2 x = 1\sec^2 x$

Ex: Differentiate

a) $f(x) = x^2e^x$

$f'(x) = 2xe^x + x^2e^x$

$= e^x(x^2 + 2x)$

$= xe^x(x+2)$

b) $g(x) = \ln x \tan x$

$g'(x) = \frac{\tan x}{x} + \ln x \sec^2 x$

c) $f(x) = \frac{2x+1}{x-1}$

$f'(x) = \frac{2(x-1) - (2x+1)(1)}{(x-1)^2}$

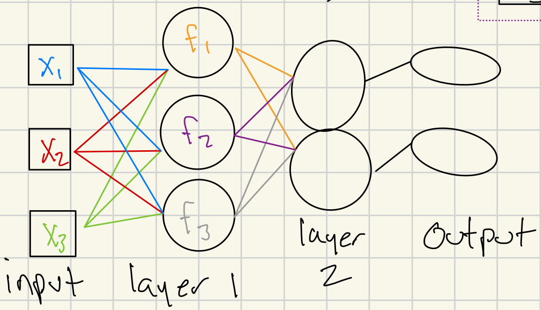
$= \frac{2x - 2 - 2x - 1}{x^2 - 2x + 1}$

$= \frac{-3}{x^2 - 2x + 1}$

Chain Rule

chain of functions $\quad x \rightarrow \boxed{g} \rightarrow g(x) \rightarrow \boxed{f} \rightarrow f(g(x))$

$f \circ g$ composition of two functions $\quad (f\circ g)(x) = f(g(x))$



$$(f\circ g)'(x) = (f(g(x)))' = f'(g(x))\,g'(x) \qquad \frac{d}{dx}\left[f(g(x))\right] = \frac{d}{dx}\left[f(g(x))\right]\frac{d}{dx}\left[g(x)\right]$$

e.g.: $f(x) = \sqrt{x} \quad g(x) = x^2 - 1 \Rightarrow h(x) = (f\circ g)(x) = \sqrt{x^2-1}$

$$\frac{d}{dx}h(x) = \frac{1}{2\sqrt{x^2-1}}(2x) = \frac{x}{\sqrt{x^2-1}}$$

Ex. Differentiate

a) $f(x) = \sqrt[3]{x^2+1}$ $\quad$ b) $g(x) = \left(\frac{2x-1}{5x+1}\right)^4$ $\quad$ c) $h(\theta) = e^{\sec(3\theta)}$ $\quad$ d) $y = x^2 e^{-3x}$

e) $g(x) = \cos^2(x)$ $\quad$ f) $F(x) = \sin(\cos(\tan(x^2)))$

a) $f(x) = (x^2+1)^{1/3}$

$$f'(x) = \frac{1}{3}(x^2+1)^{-2/3}(x^2+1)'$$
$$= \frac{1}{3(x^2+1)^{2/3}}(2x)$$
$$= \frac{2x}{3(x^2+1)^{2/3}}$$

b) $g(x) = \left(\frac{2x-1}{5x+1}\right)^4$

$$g'(x) = 4\left(\frac{2x-1}{5x+1}\right)^3 \frac{d}{dx}\left[\frac{2x-1}{5x+1}\right]$$
$$g'(x) = 4\left(\frac{2x-1}{5x+1}\right)^3 \frac{1}{(5x+1)^2}\left[2(5x+1) - 5(2x-1)\right]$$
$$= 4\left(\frac{2x-1}{5x+1}\right)^3 \frac{1}{(5x+1)^2}\, 10x+2-10x+5$$
$$= 4\left(\frac{2x-1}{5x+1}\right)^3 \frac{7}{(5x+1)^2} = \frac{28(2x-1)^3}{(5x+1)^5}$$

c) $h(\theta) = e^{\sec(3\theta)}$

$$h'(\theta) = e^{\sec(3\theta)}(\sec(3\theta))'$$
$$= e^{\sec(3\theta)}\left[\sec(3\theta)\tan(3\theta)(3\theta)'\right]$$
$$= 3e^{\sec(3\theta)}\sec(3\theta)\tan(3\theta)$$

d) $y = x^2 e^{-3x}$

$$y' = 2xe^{-3x} + x^2(-3e^{-3x})$$
$$= \left[2x - 3x^2\right]e^{-3x}$$

e) $g(x) = \cos(x)\cos(x) = \cos^2(x)$

$$g'(x) = -\sin(x)\cos(x) + \cos(x)\left[-\sin(x)\right]$$
$$= -2\sin(x)\cos(x)$$
$$= -\sin(2x)$$

f) $F(x) = \sin(\cos(\tan(x^2)))$

$$F'(x) = \left[\cos(\tan(x^2))\right]'\cos(\cos(\tan(x^2)))$$
$$= \cos(\cos(\tan(x^2)))\left[-\sin(\tan(x^2))\cdot(\tan(x^2))'\right]$$
$$= \cos(\cos(\tan(x^2)))(-\sin(\tan(x^2)))(\sec^2(x^2))(x^2)'$$
$$= \cos(\cos(\tan(x^2)))(-\sin(\tan(x^2)))\sec^2(x^2)\,2x$$
$$= -2x\sec^2(x^2)\sin(\tan(x^2))\cos(\cos(\tan(x^2)))$$

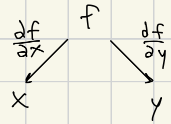
Partial Differentiation (Multivariate Differentiation)

Ex: $f(x,y) = x^2 + y^2$ $\quad (x,y) \longrightarrow \boxed{f} \longrightarrow x^2+y^2$

$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \mathbb{R}^2 \quad\quad\quad\quad \mathbb{R}$

$f_x = \frac{\partial f}{\partial x} = 2x$

$f_y = \frac{\partial f}{\partial y} = 2x$



1st order partial derivatives

In practice, when finding a partial derivative of f with respect to one variable, treat the other variables as constants (fixed) and differentiate as if it were as if in univariate differentiation.

Ex: Find the 1st order partial derivatives.

a) $f(x,y) = x^2 + y^2$

$f_x(x,y) = 2x$

$f_y(x,y) = 2y$

b) $g(x,y) = x^3 \sin x - 3x^2 y - e^{xy^2}$

$g_x(x,y) = [3x^2 \sin x + x^3 \cos x] - 6xy - y^2 e^{xy^2}$

$g_y(x,y) = -3x^2 - 2xy e^{xy^2}$

c) $f(x_1, x_2, x_3) = \ln(x_1^2 + x_2^2 + x_3^2)$

$f_{x_1}(x_1, x_2, x_3) = \frac{2x_1}{x_1^2 + x_2^2 + x_3^2}$

$f_{x_2}(x_1, x_2, x_3) = \frac{2x_2}{x_1^2 + x_2^2 + x_3^2}$

$f_{x_3}(x_1, x_2, x_3) = \frac{2x_3}{x_1^2 + x_2^2 + x_3^2}$

Ex: Find the 1st order partials

a) $f(x_1, x_2) = 5x_1 e^{3x_2} - (x_1 + 2x_2)^3$

$f_{x_1}(x_1, x_2) = 5e^{3x_2} - 3(x_1 + 2x_2)^2$

$f_{x_2}(x_1, x_2) = 15x_1 e^{3x_2} - 6(x_1 + 2x_2)^2$

b) $f(x_1, x_2, x_3) = x_1 x_2^3 \ln(x_1 x_3)$

$f_{x_1}(x_1, x_2, x_3) = x_2^3 \ln(x_1 x_3) + \frac{x_1 x_2^4}{x_1 x_3} = x_2^3 \ln(x_1 x_3) + x_2^3$

$f_{x_2}(x_1, x_2, x_3) = 3x_1 x_2^2 \ln(x_1 x_3)$

$f_{x_3}(x_1, x_2, x_3) = \frac{x_1 x_2^3}{x_1 x_3} = \frac{x_1 x_2^3}{x_3}$

$J = \begin{bmatrix} f_x(x,y) & f_y(x,y) \end{bmatrix}$ Gradient (Jacobian)

$H_f = \begin{bmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{bmatrix}$ Hessian of f

In general, for $f: \mathbb{R}^n \to \mathbb{R}$, $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \longrightarrow f(\vec{x})$, we have n first order partial derivatives of f: $\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}$. We group them into a row vector that we call the gradient, or Jacobian of f.

$$J = \vec{\nabla} f = \vec{\text{grad}}\, f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} \longrightarrow \frac{\partial f}{\partial(x_1, x_2, \cdots, x_n)}$$

Dimension $1 \times n$

$$\vec{\nabla} \cdot = \begin{bmatrix} \frac{\partial \cdot}{\partial x_1} & \frac{\partial \cdot}{\partial x_2} & \cdots & \frac{\partial \cdot}{\partial n} \end{bmatrix}$$

del operator or nabla

Ex: $f(x_1, x_2, x_3) = x_1 x_2^2 x_3^3$, $f: \mathbb{R}^3 \longrightarrow \mathbb{R}$ $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \longrightarrow x_1 x_2^2 x_3^3$ $\quad$ J is 1×3.

$J = \vec{\nabla} f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} x_2^2 x_3^3 & 2x_1 x_2 x_3^3 & 3x_1 x_2^2 x_3^2 \end{bmatrix}$

Ex: For the function $f(x,y,z) = e^{-2x}\sin y - 5\sqrt{x}\,y^2z^3$, find the Jacobian row vector J

$$J = \left[\left(-2e^{-2x}\sin y - \frac{5y^2z^3}{2\sqrt{x}}\right)\ \left(e^{-2x}\cos y - 10\sqrt{x}\,yz^3\right)\ \left(-15\sqrt{x}\,y^2z^2\right)\right]$$

Gradients of Vector-Valued Functions

$\vec{f}: \mathbb{R}^n \longrightarrow \mathbb{R}^m$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \longrightarrow \vec{f} = \begin{bmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{bmatrix}, \quad J_{m\times n} = \underbrace{\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & & & \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}}_{n \text{ columns}} \Big\} m \text{ rows}$$

$\frac{\partial f_7}{\partial x_5}$ ← row, ← column

$f_j(\vec{x})\ 1 \le j \le m$, $x_i\ 1 \le i \le n$

$f_1 \quad f_2 \cdots f_m$

$x_1\, x_2 \cdots x_n \quad x_1\, x_2 \cdots x_n \quad x_1\, x_2 \cdots x_n$

Ex: For the functions $f_1(x,y) = x^2 + y^2$ and $f_2(x,y) = 5xy^3$, calculate the Jacobian matrix then evaluate it at $(1,1)$.

$\vec{f}: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$

$\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix} \longrightarrow \begin{bmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \end{bmatrix}$

$$J = \begin{bmatrix} (2x) & (2y) \\ (5y^3) & (15xy^2) \end{bmatrix} \qquad J(1,1) = \begin{bmatrix} (2(1)) & (2(1)) \\ (5(1)^3) & (15(1)(1)^2) \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 5 & 15 \end{bmatrix}$$

or $J|_{(1,1)}$

Higher Order Derivatives

Univariate

$f(x)$ → $f'(x)$ 1st order derivative → $f''(x)$ 2nd order derivative → $f'''(x)$ 3rd order derivative → $f^{(n)}(x)$ n^{th} order derivative

multivariate

Notation: $\frac{\partial f}{\partial x}\frac{\partial f}{\partial x} = \frac{\partial^2 f(x)}{\partial x^2} = f_{xx}(x,y)$

$\frac{\partial f}{\partial y}\frac{\partial f}{\partial x} = \frac{\partial^2 f(x,y)}{\partial y x} = f_{xy}(x,y)$

$f(x,y)$

1st order partials: $f_x(x,y)$, $f_y(x,y)$

2nd order partials: $f_{xx}(x,y)$, $f_{xy}(x,y)$, $f_{yx}(x,y)$, $f_{yy}(x,y)$

3rd order partials: $f_{xxx}(x,y)$, $f_{xxy}(x,y)$, $f_{xyx}(x,y)$, $f_{xyy}(x,y)$, $f_{yxx}(x,y)$, $f_{yxy}(x,y)$, $f_{yyx}(x,y)$, $f_{yyy}(x,y)$

Hessian of f

$$H_f = \begin{bmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{bmatrix}$$

symmetric matrix

$H_f^T = H_f$

Clairaut's Theorem: If the second order partial derivatives are continuous, then $f_{xy}(x,y) = f_{yx}(x,y)$

Ex: Find the 2^{nd} order partial derivatives. $f(x,y) = 5x^2y - 3x + 5xy^2 - y^3$

$f_x(x,y) = 10xy - 3 + 5y^2$

$f_y(x,y) = 5x^2 + 10xy - 3y^2$

$f_{xx}(x,y) = 10y$

$f_{yx}(x,y) = 10x + 10y$

$f_{xy}(x,y) = 10x + 10y$

$f_{yy}(x,y) = 10x - 6y$

$$H_f = \begin{bmatrix} 10y & (10x+10y) \\ (10x+10y) & (10x - 6y) \end{bmatrix}$$

(look up Kaggle - data science / M.L.)

Basic Rules of Partial Differentiation

$\left[\frac{\partial}{\partial x_1}\ \frac{\partial}{\partial x_2} \cdots \frac{\partial}{\partial x_n}\right]$

Product: $\frac{\partial}{\partial \vec{x}} f(\vec{x})g(\vec{x}) = \frac{\partial}{\partial \vec{x}} f(\vec{x}) g(\vec{x}) + f(\vec{x})\frac{\partial}{\partial \vec{x}} g(x)$

Sum: $\frac{\partial}{\partial \vec{x}}\left[f(\vec{x})+g(\vec{x})\right] = \frac{\partial}{\partial \vec{x}} f(\vec{x}) + \frac{\partial}{\partial \vec{x}} g(\vec{x})$

Chain: $\frac{\partial}{\partial \vec{x}} f(g(\vec{x})) = \frac{\partial}{\partial \vec{x}} f(g(\vec{x})) \cdot \frac{\partial}{\partial \vec{x}} g(\vec{x})$

Ex: Find $\frac{df}{dt}$, $f(x,y) = x^4 - 2xy^3 + x^2y - y^5$, $x = e^{-3t}$, $y = \sin(5t)$

$$\begin{tikzcd}
& f \arrow[dl, dash, "\frac{\partial f}{\partial x}"'] \arrow[dr, dash, "\frac{\partial f}{\partial y}"] & \\
x \arrow[d, dash, "\frac{dx}{dt}"'] & & y \arrow[d, dash, "\frac{dy}{dt}"] \\
t & & t
\end{tikzcd}$$

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

$$= \left[4x^3 - 2y^3 + 2xy\right]\left(-3e^{-3t}\right) + \left[-6xy^2 + x^2 - 5y^4\right]\left(5\cos(5t)\right)$$

matrix form

$f(x,y) = f(\vec{x})$ $\quad \vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}$

$$\underset{1\times1}{\frac{df}{dt}} = \underset{1\times2}{\frac{\partial f}{\partial \vec{x}}}\ \underset{2\times1}{\frac{d\vec{x}}{dt}} = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix}$$

top bottom

$$= \begin{bmatrix} (4x^3 - 2y^3 + 2xy) & (-6xy + x^2 - 5y^4) \end{bmatrix} \begin{bmatrix} -3e^{-3t} \\ 5\cos(5t) \end{bmatrix}$$

Ex: Find $\frac{\partial f}{\partial s}$ and $\frac{\partial f}{\partial t}$ for $f(x_1, x_2) = 5x_1 \ln x_2$, $x_1 = te^{2s}$, $x_2 = t^2\cos(s)$

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial s} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial s}$$

$$= (5\ln x_2)(2te^{2s}) + \left(\frac{5x_1}{x_2}\right)(-t^2\sin(s))$$

$$= 10te^{2s}\ln x_2 - \frac{5x_1t^2\sin(s)}{x_2}$$

$$= \begin{bmatrix} (5\ln x_2) & \left(\frac{5x_1}{x_2}\right) \end{bmatrix} \begin{bmatrix} 2te^{2s} \\ -t^2\sin(s) \end{bmatrix}$$

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t}$$

$$= (5\ln x_2)(e^{2s}) + \left(\frac{5x_1}{x_2}\right)(2t\cos(s))$$

$$= 5e^{2s}\ln x_2 + \frac{10x_1t\cos(s)}{x_2}$$

$$= \begin{bmatrix} (5\ln x_2) & \left(\frac{5x_1}{x_2}\right) \end{bmatrix} \begin{bmatrix} e^{2s} \\ 2t\cos(s) \end{bmatrix}$$

$\vec{x} = (x_1, x_2)$, $\vec{u} = (s, t)$

$$\underset{1\times2}{\frac{\partial f}{\partial \vec{u}}} = \frac{\partial f}{\partial (s,t)} = \underset{1\times2}{\frac{\partial f}{\partial \vec{x}}} \cdot \underset{2\times2}{\frac{\partial \vec{x}}{\partial \vec{u}}}$$

$$\begin{bmatrix} \frac{\partial f}{\partial s} & \frac{\partial f}{\partial t} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1}{\partial s} & \frac{\partial x_1}{\partial t} \\ \frac{\partial x_2}{\partial s} & \frac{\partial x_2}{\partial t} \end{bmatrix}$$

Ex: For the following functions, calculate the expression $\frac{df}{dt} = \frac{\partial f}{\partial \vec{x}}\frac{d\vec{x}}{dt}$ in matrix form, where

a) $\vec{x} = (x_1, x_2)$, $f(\vec{x}) = x_1^2 + x_2^2 + x_1x_2$, $x_1(t) = 1 - t^2$, $x_2(t) = 1 + t^2$

$$\begin{bmatrix} (2x_1x_2^2 + x_2) & (2x_1^2x_2 + x_1) \end{bmatrix} \begin{bmatrix} -2t \\ 2t \end{bmatrix}$$

b) Calculate $\frac{df}{dt} = \frac{\partial f}{\partial \vec{x}}\frac{\partial \vec{x}}{\partial \vec{u}}\frac{d\vec{u}}{dt}$, where $\vec{x} = (x_1, x_2)$, $\vec{u} = (u_1, u_2)$, $f(\vec{x}) = x_1\sin x_2 - x_2\cos x_1$,
$x_1(\vec{u}) = 2u_1 - 3u_2^2$, $x_2(\vec{u}) = 2u_1^2 + 3u_2$, $u_1(t) = e^{5t}$, $u_2(t) = \cos(2t)$

$$\underset{1\times2}{\begin{bmatrix} (\sin x_2 + x_2\sin x_1) & (x_1\cos x_2 - \cos x_1) \end{bmatrix}} \underset{2\times2}{\begin{bmatrix} 2 & -6u_2 \\ 4u_1 & 3 \end{bmatrix}} \underset{2\times1}{\begin{bmatrix} 5e^{5t} \\ -2\sin(2t) \end{bmatrix}}$$

$\rightarrow 1\times2 \quad \rightarrow 1\times1$

More Hessians & Gradients

Ex: Calculate the Hession of $f(x,y,z) = x^2 z - xye^z + z\sin y$

$f_x(x,y,z) = 2xz - ye^z$ $\quad$ $f_y(x,y,z) = -xe^z + z\cos y$ $\quad$ $f_z(x,y,z) = x^2 - xye^z + \sin y$

$f_{xx}(x,y,z) = 2z$ $\quad$ $f_{yx}(x,y,z) = -e^z$ $\quad$ $f_{zx}(x,y,z) = 2x - ye^z$

$f_{xy}(x,y,z) = -e^z$ $\quad$ $f_{yy}(x,y,z) = -z\sin y$ $\quad$ $f_{zy}(x,y,z) = -xe^z + \cos y$

$f_{xz}(x,y,z) = 2x - ye^z$ $\quad$ $f_{yz}(x,y,z) = -xe^z + \cos y$ $\quad$ $f_{zz}(x,y,z) = -xye^z$

$$H_f = \begin{bmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{bmatrix} = \begin{bmatrix} 2z & -e^z & (2x - ye^z) \\ -e^z & -z\sin y & (-xe^z + \cos y) \\ (2x - ye^z) & (-xe^z + \cos y) & -xye^z \end{bmatrix}$$

Bonus Material: Gradients of Vector-valued Functions

Ex 5.9 (pg 153): $\vec{f}(\vec{x}) = A\vec{x}$, $A \in \mathbb{R}^{m \times n}$, $\vec{x} \in \mathbb{R}^n$ $\quad \vec{f}: \mathbb{R}^n \to \mathbb{R}^m$

Find $\frac{d\vec{f}}{d\vec{x}} \begin{smallmatrix} \leftarrow m \\ \leftarrow n \end{smallmatrix}$ $\quad \dim\left(\frac{d\vec{f}}{d\vec{x}}\right) = m \times n$

$$\vec{f}(\vec{x}) = \begin{bmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{bmatrix} \qquad \frac{d\vec{f}}{d\vec{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \qquad A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & & A_{mn} \end{bmatrix}$$

1st row

$$f_1(\vec{x}) = A\vec{x} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_m$$

$$\vec{f}'(\vec{x}) = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & & A_{mn} \end{bmatrix} = A$$

$$\boxed{\frac{dA\vec{x}}{d\vec{x}} = A \quad \frac{d}{d\vec{x}}\left[\vec{x}^T\vec{x}\right] = 2\vec{x}^T \quad \frac{d\vec{x}^T\vec{a}}{d\vec{x}} = \frac{d\vec{a}^T\vec{x}}{d\vec{x}} = \vec{a}^T}$$

Ex 5.12 (pg 156): Find $\frac{d\vec{f}}{dA}$. Dimension? $m \times (m \times N)$ 3D Tensor

$$\frac{df_1(\vec{x})}{dA} = ? \qquad \frac{df_1(\vec{x})}{dA_{1,:}} = \begin{bmatrix} \frac{\partial f_1}{\partial A_{11}} & \frac{\partial f_1}{\partial A_{12}} & \cdots & \frac{\partial f_1}{\partial A_{1n}} \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \vec{x}^T$$

first row of A

$$\frac{df_1(\vec{x})}{dA_{2,:}} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \end{bmatrix} = \vec{0}^T \qquad \frac{df_1(\vec{x})}{dA_{i,:}} = \vec{0}^T \text{ for } i \neq 1$$

$$\underset{1 \times (m \times N)}{\frac{\partial f_1(\vec{x})}{\partial A}} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Ex 5.13: $f(R) = R^T R$, $R \in \mathbb{R}^{m \times n}$. Find dimension of $\frac{df}{dR}$.

$$\underset{n \times m}{R^T} \underset{m \times n}{R} = \underset{n \times n}{?} \qquad \underset{n \times n}{f(R)} \qquad n \times n \times m \times n$$

Vectorize (Flatten)

$$\underset{5\times 2}{A}=\begin{bmatrix}1&2\\3&4\\5&6\\7&8\\9&10\end{bmatrix}\xrightarrow{\text{Vectorize}}\underset{10\times 1}{\begin{bmatrix}1\\3\\5\\7\\9\\2\\4\\6\\8\\10\end{bmatrix}}\qquad \vec{x}=\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$$

$\dfrac{dA}{d\vec{x}} \leftarrow 5\times 2\times 3$ 3D tensor $\qquad \dfrac{d\tilde{A}}{d\vec{x}}$ 10×3

Ex: Compute the derivatives of the following functions by using the chain rule. Provide the dimension of every single partial derivative. Describe your steps in detail

a) $f(z)=\cos(z)$, $z=\vec{x}^T\vec{x}$, $\vec{x}\in\mathbb{R}^n$

b) $f(\vec{z})=\ln(\vec{z})$, $\vec{z}=A\vec{x}+\vec{b}$, $A\in\mathbb{R}^{m\times n}$, $\vec{x}\in\mathbb{R}^n$, $\vec{b}\in\mathbb{R}^m$, $\ln(\cdot)$ is applied to every element of $\vec{z}$

a) $\vec{x}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}\in\mathbb{R}^n$, $z=\underset{1\times n\ n\times 1}{\vec{x}^T\vec{x}}\Rightarrow 1\times 1$, $f(z)=\underset{1\times 1}{\cos(z)}$ $\qquad z=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=x_1^2+x_2^2+\cdots+x_n^2$

$$\frac{\partial f}{\partial\vec{x}}=\frac{df}{dz}\frac{\partial z}{\partial\vec{x}}=-\sin(z)\begin{bmatrix}\frac{\partial z}{\partial x_1}&\frac{\partial z}{\partial x_2}&\cdots&\frac{\partial z}{\partial x_n}\end{bmatrix}=-\sin(z)\begin{bmatrix}2x_1&2x_2&\cdots&2x_n\end{bmatrix}=-2\sin(z)\begin{bmatrix}x_1&x_2&\cdots&x_n\end{bmatrix}=-2\sin(z)\vec{x}^T$$

b) $\vec{x}=\underset{n\times 1}{\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}}\in\mathbb{R}^n$, $A=\underset{m\times n}{\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\\vdots&\vdots&&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}\end{bmatrix}}\in\mathbb{R}^{m\times n}$, $\vec{b}=\underset{m\times 1}{\begin{bmatrix}b_1\\b_2\\\vdots\\b_m\end{bmatrix}}$, $A\vec{x}=m\times 1$ $\qquad \vec{z}=A\vec{x}+\vec{b}=\begin{bmatrix}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n+b_1\\a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n+b_2\\\vdots\\a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n+b_m\end{bmatrix}$

$$\underset{n\times 1}{\overset{m\times 1}{\frac{\partial f}{\partial\vec{x}}}}=\underset{m\times 1}{\overset{m\times 1}{\frac{\partial f}{\partial\vec{z}}}}\underset{n\times 1}{\overset{m\times 1}{\frac{\partial\vec{z}}{\partial\vec{x}}}}=\underset{m\times m}{D}\ \underset{m\times n}{A}=\begin{bmatrix}\frac{\partial f_1}{\partial z_1}&\frac{\partial f_1}{\partial z_2}&\cdots&\frac{\partial f_1}{\partial z_m}\\\vdots&\vdots&&\vdots\\\frac{\partial f_m}{\partial z_1}&\frac{\partial f_m}{\partial z_2}&\cdots&\frac{\partial f_m}{\partial z_m}\end{bmatrix}A=\begin{bmatrix}\frac{1}{z_1}&0&\cdots&0\\0&\frac{1}{z_2}&0\cdots&0\\\vdots&&\ddots&\\0&\cdots&0&\frac{1}{z_m}\end{bmatrix}$$

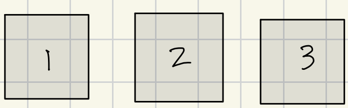
Probability and Statistics

Why probability?

Machine learning must always deal with uncertain quantities and sometimes nondeterministic quantities.

Uncertainty and stochasticity can arise from many sources:

1. Inherent stochasticity in the system being modeled
   e.g. a game of cards where cards are truly shuffled into a random order
2. Incomplete observability:
   even deterministic systems can appear stochastic where we cannot observe all the variables that drive the behavior of the system.
   e.g. Monty Hall Problem



③ Incomplete Modeling:

when the model must discard data we have observed. This creates uncertainty in the model's prediction

e.g. robot that can exactly observe the location of every object around it. If space is discretized when predicting future location. Each object could be anywhere in the cell.

Philosophical Approaches

Frequentist: probability theory was originally developed to analyze the frequencies of events. Estimation comes from experiments only.

Limitations:
- requires a large amount of data
- can't inject prior knowledge of our estimates
- These kinds of events are often repeatable

e.g. Diagnosing a patient with the flu.

Frequentist Definition: when we say an outcome has probability of some number p, it means that if we repeated the experiment infinitely many times, the proportion p of the repetitions would result in that outcome.

Bayesian: We use probability to quantify our degree of belief, with 1 indicating absolute certainty of the event occurring and 0 indicating absolute certainty of the event not occurring. We combine our prior beliefs with observation to find the probability.

Def: If there are several equally likely, mutually exclusive, and collectively exhaustive outcomes of an experiment, the probability of an event E is given by:

$$P(E) = \frac{\text{number of outcomes favorable to event } E}{\text{total number of possible outcomes}}$$

"probability of event E"

Probability: likelihood of an event occurring

Axioms:
- $0 \le p(A) \le 1$
- $\Sigma(\text{probabilities of all outcomes}) = 1$

flipping a coin $\{H, T\}$

$p(H) = \frac{1}{2}$, $p(T) = \frac{1}{2}$, $p(H) + p(T) = 1$

Trial: observing an event occur & recording the outcome

Experiment: a collection of one or more trials.

Experimental probability: the probability assigned to an event based on an experiment.

Sample space: the set of all possible outcomes of an experiment, denoted by S.

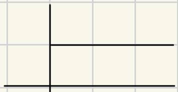
e.g. $S=\{H,T\}$

(each individual outcome is called a point in the sample space)

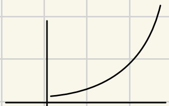
Expected Value: specific outcome we expect to occur when we run an experiment

Distributions: the main way we like to classify data sets. If a data set complies with certain characteristics, we can usually attribute the likelihood of its values to a specific distribution

Gaussian (or normal) Distribution



uniform Distribution



Exponential Distribution

Event: a subset of the sample space S. We say an event A occurs if the actual outcome $\in A$

rolling a d6: $S=\{1,2,3,4,5,6\}$

event: rolling an even number. $A=\{2,4,6\}$

$p(A)=\frac{3}{6}=\frac{1}{2}$

Ex: Find the probability that a single card drawn from a shuffled deck of cards will be either a diamond or King (or both)

$A=\{K_S, K_C, K_H, K_D, Q_D, J_D, 10_D, 9_D, 8_D, 7_D, 6_D, 5_D, 4_D, 3_D, 2_D, A_D\}$

$|A|=16$, $|S|=52$

$P(A)=\frac{16}{52}=\frac{8}{26}=\frac{4}{13}$

Ex: A 3 digit number (that is a number from 100-999) is selected at "random" ("at random" means we assume all numbers to have the same probability to be selected). What is the probability that all 3 digits are the same?

$A=\{111,222,333,444,555,666,777,888,999\}$ final − initial + 1 = # of numbers

$S=\{100,101,\cdots,999\}$

$P(A)=\frac{|A|}{|S|}=\frac{9}{900}=\frac{1}{100}$

Ex: A single card is drawn at random from a shuffled deck. What is the probability it is red? That it is the Ace of Hearts? That it is either an Ace or red or both?

red, Ace of ♡, Ace or red, Deck

$|A_1| = 26$, $|A_2| = 1$, $|A_3| = 26+2 = 28$, $|S| = 52$

$p(A_1) = \frac{|A_1|}{|S|} = \frac{26}{52} = \frac{1}{2}$, $p(A_2) = \frac{|A_2|}{|S|} = \frac{1}{52}$, $P(A_3) = \frac{|A_3|}{|S|} = \frac{28}{52} = \frac{7}{13}$

Ex: A letter is selected from the alphabet. What is the probability it is one of the letters in the word "probability"? What is the probability that it occurs in the first half of the alphabet? What is the probability that it is a letter after x

$A_1 = \{a, b, i, l, o, p, r, t, y\}$, $|A_1| = 9$ $\quad |S| = 26$

$p(A_1) = \frac{|A_1|}{|S|} = \frac{9}{26}$.

$A_2 = \{a, b, c, d, e, f, g, \overset{8}{h}, i, j, k, l, m\}$, $|A_2| = 13$

$p(A_2) = \frac{|A_2|}{|S|} = \frac{13}{26} = \frac{1}{2}$

$A_3 = \{y, z\}$, $|A_3| = 2$

$p(A_3) = \frac{|A_3|}{|S|} = \frac{2}{26} = \frac{1}{13}$

Sample Space

There are many different sample spaces for any given probability problem.
For example, here are some sample spaces for the experiment of tossing a fair coin twice and recording the outcomes.

$S = \{\underset{\frac{1}{4}}{HH}, \underset{\frac{1}{4}}{HT}, \underset{\frac{1}{4}}{TH}, \underset{\frac{1}{4}}{TT}\}$ uniform sample space (four equally likely, mutually exclusive outcomes)

or

$S = \{\underset{\frac{1}{4}}{HH}, \underset{\frac{1}{2}}{H}, \underset{\frac{1}{4}}{0}\}$ nonuniform sample space ← counting # of heads

or

$S = \{\underset{\frac{1}{4}}{0H}, \underset{\frac{3}{4}}{H \geq 1}\}$

Note: $\{\underset{\frac{1}{4}}{2H}, \underset{\frac{3}{4}}{H \geq 1}, \underset{\frac{1}{2}}{T = 1}\}$ not collectively exhaustive (no TT)
not a valid sample space

In order to use a sample space to solve a problem, we need to have the probabilities corresponding to the different sample points in the sample space.
If a sample space n equally likely outcomes, we assign a probability of $\frac{1}{n}$ to each.
We call such a space a uniform sample space.

e.g.: $S = \{\underset{\frac{1}{4}}{HH}, \underset{\frac{1}{4}}{HT}, \underset{\frac{1}{4}}{TH}, \underset{\frac{1}{4}}{TT}\}$

Suppose the outcomes are not equally likely: $S = \{2H, 1H, 0H\}$ or $S = \{0H, H \geq 1\}$

$\frac{1}{4}$ $\frac{1}{2}$ $\frac{1}{4}$ ← non uniform sample spaces → $\frac{1}{4}$ $\frac{3}{4}$

\# For some problems, there may be both uniform & non-uniform sample spaces. But sometimes, there is no uniform sample space. For example, consider the experiment of tossing a weighted coin which has a probability $\frac{1}{3}$ for heads and $\frac{2}{3}$ for tails. In such a case, we can't use our definition of probability and need the following more general definition.

Def: Given any sample space (uniform or non-uniform) and the probability associated with each outcome, we find the probability of an event by adding the probabilities associated with all the sample points favorable to that event.

Ex: A coin is tossed three times. A uniform sample space for this problem contains eight points: $S = \{hhh, ttt, htt, tht, tth, hht, hth, thh\}$ and we attach a probability of $\frac{1}{8}$ to each point. a) What is the probability of at least 2 tails in succession? b) What is the probability that two consecutive coins fall the same? c) If we know there was at least 1 tail, what is the probability of all tails?

a) Let A = "at least 2 tails in succession"

$S_a = \{ttt, tth, htt\}$

$p(A) = \frac{|S_a|}{|S|} = \frac{3}{8} = \frac{1}{8} + \frac{1}{8} + \frac{1}{8}$

b) Let B = "2 consecutive coins fall the same"

$S_b = \{hhh, ttt, htt, tth, hht, thh\}$

$p(B) = \frac{|S_b|}{|S|} = \frac{6}{8} = \frac{3}{4}$

c) Let C = "all tails"

$S_c = \{ttt\}$, $S' = \{ttt, htt, tht, tth, hht, hth, thh\}$

$p(C) = \frac{|S_c|}{|S'|} = \frac{1}{7}$

Ex: Let two dice be thrown, the first die can show any number between 1 and 6 and similarly for the second die. There are 36 possible outcomes or points in a uniform sample space for this problem. Each point has an associated probability of $\frac{1}{36}$. We can indicate a 3 in the first and 2 in the second by (3,2). Then the sample space is the following:

1st \ 2nd	⚀	⚁	⚂	⚃	⚄	⚅
⚀	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
⚁	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
⚂	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
⚃	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
⚄	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
⚅	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

a) What is the probability the sum of the numbers is 5?

b) What is the probability the sum is divisible by 5?

c) Set up a sample space in which the points correspond to the possible sums and find the probability associated with the points of this non-uniform sample space.

a) Let A = "the sum is 5"

$S_a = \{(1,4),(2,3),(3,2),(4,1)\}$

$p(A) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{4}{36} = \frac{2}{18} = \frac{1}{9}$

d) What is the most probable sum in a toss of 2 dice?

e) What is the probability the sum is greater than 9?

b) Let B = "the sum is divisible by 5"

$S_B = \{\underbrace{(1,4),(2,3),(3,2),(4,1)}_{5}, \underbrace{(4,6),(5,5),(6,4)}_{10}\}$

$p(B) = \frac{7}{36}$

c)

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

$2: \frac{1}{36}$, $3: \frac{2}{36}$, $4: \frac{3}{36}$, $5: \frac{4}{36}$, $6: \frac{5}{36}$, $7: \frac{6}{36}$, $8: \frac{5}{36}$, $9: \frac{4}{36}$, $10: \frac{3}{36}$, $11: \frac{2}{36}$, $12: \frac{1}{36}$

d) 7, $p(7) = \frac{6}{36}$

e) $p(10) + p(11) + p(12) = \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{6}{36}$

Probability Theorems

S: sample space (set of all possible outcomes)

A: event (subset of S)

We say A occurred if the actual outcome S_0 is in A.

Translating Between Probability and Sets

Probability	Sets
Sample space	S
a possible outcome s	$s \in S$
event A	$A \subseteq S$
A or B occurred (inclusive)	$A \cup B$
A and B	$A \cap B$
Not A (complement of A)	A^c
A and B are mutually disjoint (exclusive)	$A \cap B = \emptyset$
A implies B	$A \subseteq B$
probability of A	$p(A)$
A and B are independent	$p(A \cap B) = p(A)p(B)$

Properties of Probability

1) If two events A & B are mutually exclusive (disjoint), i.e. $A \cap B = \emptyset$, then
$p(A \text{ or } B) = p(A \cup B) = p(A) + p(B)$
2) In general, $p(A \text{ or } B) = p(A \cup B) = p(A) + p(B) - p(A \cap B)$
3) $p(A^c) = 1 - p(A)$
4) If $A \subseteq B$, $p(A) \leq p(B)$
5) A & B are independent, $p(A \cap B) = p(A)p(B)$

Ex: Two students are working seporately on the same problem. If the first student has probability 1/2 of solving it and the second student has probability 3/4 of solving it, what is the probability that at least one of them solves it?

$p(1)=\frac{1}{2}$, $p(2)=\frac{3}{4}$

$p(1 \text{ or } 2)=p(1\cup 2)=p(1)+p(2)-p(1\cap 2)=p(1)+p(2)-p(1)p(2)$

$=\frac{1}{2}+\frac{3}{4}-\frac{1}{2}\frac{3}{4}=\frac{2}{4}+\frac{3}{4}-\frac{3}{8}=\frac{10}{8}-\frac{3}{8}=\frac{7}{8}$

Ex: a) In three tosses of a coin, what is the probability all three are heads?
b) What is the probability of all heads when a coin is tossed ten times?
c) What is the probability that we get at least one tails in ten tosses?

a) Let $A=$ "all 3 heads"

$A=\{hhh\}$, $|A|=1$

$|S|=2^3=8$

$p(A)=\frac{|A|}{|S|}=\frac{1}{8}$

or

let $A=$ "heads on first toss"

let $B=$ "heads on second toss"

let $C=$ "heads on third toss"

$p(A\cap B\cap C)=p(A)\,p(B)\,p(C)$

$=\frac{1}{2}\frac{1}{2}\frac{1}{2}=\frac{1}{8}$

b) Let $B=$ "all 10 heads"

$B=\{hhhhhhhhhh\}$

$|S|=2^{10}=1024$

$p(B)=\frac{|B|}{|S|}=\frac{1}{1024}$

or

$p(B)=\underbrace{\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdots\frac{1}{2}}_{10 \text{ times}}=\left(\frac{1}{2}\right)^{10}=\frac{1}{1024}$

c) Let $C=$ "at least 1 tails"

$C^c=\{hhhhhhhhhh\}$

$p(C^c)=p(B)=\frac{1}{1024}$

$p(C)=1-p(C^c)=1-\frac{1}{1024}=\frac{1023}{1024}$

Conditional Probability

If A & B are events with $p(B)>0$, then the probability of A given B, denoted by $p(A|B)$, is defined as: $p(A|B)=\frac{p(A\cap B)}{p(B)}$

Similarly, $p(B|A)=\frac{p(A\cap B)}{p(A)}$

$p(A|B)$ – "posterior probability"

$\frac{p(A\cap B)}{p(B)}$ – "prior probability"

B – "evidence"

Note: If A and B are independent, $p(A|B)=\frac{p(A\cap B)}{p(B)}=\frac{p(A)p(B)}{p(B)}=p(A)$

Note: In general, $p(B|A)\neq p(A|B)$

Prosecutor's Fallacy

Ex: A standard deck of 52 cards is shuffled well. Two cards are drawn separately one at a time without replacement. Let A be the event that the first card is a heart and B the event that the second card is red. Find $p(A|B)$ and $p(B|A)$.

$|A| = 13$ $\quad |S| = 52$ $\quad$ Note: A & B are not independent

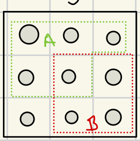
$|B| = 26 - 1 = 25$

$$p(B|A) = \frac{25}{51} \qquad p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{p(B|A)\,p(A)}{p(B)} = \frac{\frac{25}{51} \cdot \frac{13}{52}}{26/51} = \frac{25}{51} \cdot \frac{13}{52} \cdot \frac{52}{26} = \frac{25}{102}$$

Intuition using Pebbles (Blizstein-Hwang)

Consider a finite sample space S with the outcomes visualized as pebbles with a total mass equal to 1 kg.

S



A: event, B: event (sets of pebbles)

A & B are subsets of S

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

restrict sample space to event B

$$p(A|B) = \frac{1}{4}$$

Bayes' Rule & Law of Total Probability

The definition of conditional probability is simple, just the ratio of 2 probabilities, but it has far-reaching consequences.

(i) The first consequence is obtained by moving the denominator to the other side

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \iff p(A|B)\,p(B) = p(A \cap B)$$

Theorem: For any events A & B with positive probabilities ($p(A) > 0$ & $p(B) > 0$), we have

$$p(A \cap B) = p(A|B)\,p(B) = p(B|A)\,p(A).$$

We are ready now to introduce the two main theorems about probability:

Bayes' Rule & the Law of Total Probability (LoTP)

Bayes' Rule: $p(A|B) = \dfrac{p(B|A)\,p(A)}{p(B)}$

$p(A)$ - prior probability of A $\quad$ $p(B)$ - evidence

$p(A|B)$ - posterior probability of A $\quad$ $p(B|A)$ - likelihood

Bayes' Rule has important applications and implications in probability & statistics, since it is often necessary to find conditional probabilities, and often $p(B|A)$ is much easier to find directly than $p(A|B)$, and vice versa.

("sum rule")

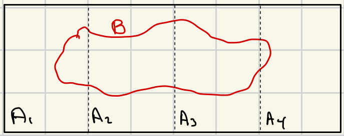
Law of Total Probability: Relates conditional probability to unconditional probability. It is essential for fulfilling the promise that conditional probability can be used to decompose complicated probability problems into simpler pieces. It is often used in tandem with Bayes' Rule.

Def: Let $A_1, A_2, \cdots, A_n$ be a partition of the sample space S (i.e., A_i's are disjoint events and their union is the set S), with $p(A_i) > 0$, $1 \le i \le n$.

Then $p(B) = p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + \cdots + p(B|A_n)p(A_n)$
$= \sum_{i=1}^{n} p(B|A_i)p(A_i)$

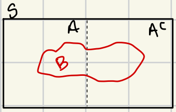
Visual Intuition:

S



$p(B) = p(A_1 \cap B) + p(A_2 \cap B) + p(A_3 \cap B) + p(A_4 \cap B)$
$= p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + p(B|A_3)p(A_3) + p(B|A_4)p(A_4)$

S



$p(B) = p(B|A)p(A) + p(B|A^c)p(A^c)$

Ex: Suppose we have 2 boxes: red & blue. Red: 2 apples, 6 oranges   Blue: 3 apples, 1 orange
Pick a box at random and pick a fruit from it at random. Red box is picked 40% of the time and blue box is picked 60% of the time. a) What is the probability the fruit is an apple?
b) If the fruit is an apple, what is the probability it came from the red box?

a) Let A = "fruit is an apple", B = "box is red", $R = B^c$ = "box is blue"

$R = \frac{1}{4}$, $B = \frac{3}{4}$ $A = \frac{5}{12}$

$p(A) = p(A|R)p(R) + p(A|B)p(B)$
$= \frac{1}{4}\cdot\frac{4}{10} + \frac{3}{4}\cdot\frac{6}{10} = \frac{1}{10} + \frac{9}{20} = \frac{2}{20} + \frac{9}{20} = \frac{11}{20}$

b) $p(R|A) = \frac{P(A|R)\,P(R)}{P(A)} = \frac{20}{11}\left[\frac{1}{4}\,\frac{4}{10}\right] = \frac{2}{11}$

Ex: You have one fair coin, and one biased coin which lands heads with probability $\frac{3}{4}$. You pick one of the coins at random and flip it three times. It lands heads all 3 times. Given this information, what is the probability that the coin you picked is the fair one?

Let F = "the coin is fair", H = "lands heads three times"

$p(H) = p(H|F)p(F) + p(H|F^c)p(F^c)$
$= \frac{1}{8}\cdot\frac{1}{2} + \frac{3}{4}\cdot\frac{1}{2} = \frac{1}{16} + \frac{3}{8} = \frac{1}{16} + \frac{6}{16} = \frac{7}{16}$

Solution $p(F|H) = p(H|F)p(F)\frac{1}{p(H)} = \frac{1}{8}\cdot\frac{1}{2}\cdot\frac{1}{p(H)} = \frac{1}{16}\cdot\frac{1}{p(H)}$

$p(H) = \left[p(H|F)p(F) + p(H|F^c)p(F^c)\right] = \frac{1}{8}\cdot\frac{1}{2} + \frac{27}{64}\cdot\frac{1}{2} = \frac{1}{16} + \frac{27}{128} = \frac{8}{128} + \frac{27}{128} = \frac{35}{128}$

$p(F|H) = \frac{1}{16}\cdot\frac{1}{p(H)} = \frac{1}{16}\cdot\frac{128}{35} = \frac{8}{35}$

Methods of Counting

How many 2 digit numbers have either a 5 or a 7 for the tens digit and either a 3, 4, or 6 for the units digit?

	3	4	6
5	53	54	56
7	73	74	76

Fundamental Counting Principle

If one thing can be done n_1 ways, and after that a second thing can be done n_2 ways, the two things can be done in succession $n_1 n_2$ ways.
This can be extended to doing any numbers of things one after the other. The 1^{st} n_1 ways, 2nd n_2 ways, ..., the n^{th} n_n ways. The total # of ways to perform the succession of acts is the product $n_1 n_2 \cdots n_n$

Consider a set of n things lined up in a row, we ask how many ways can we arrange (permute) them?
The result is called the number of permutations of n things n at a time and is denoted by ${}_nP_n$, $P(n,n)$, or P_n^n

To find this number, we think of seating n people in a row of n chairs.
We can seat anyone in the first chair, $n-1$ people in the second chair, $n-2$ people in the third chair, etc.
By the fundamental counting principle, the number of ways we can seat n people in a row of n chairs is $P(n,n) = n(n-1)(n-2)(n-3)\cdots 2\cdot 1 = n!$

Suppose there are n people but only $r < n$ and we ask how many ways can we select a group of r people and seat them in the r chairs.
The result is called the number of permutations of n things r at a time and is denoted by ${}_nP_r$, $P(n,r)$, or P_n^r.
Arguing as before, we find that there are n ways to fill the first, $(n-1)$ to fill the second, $(n-2)$ to fill the third, etc and finally $(n-r+1)$ to fill the r^{th}.

$$P(n,r) = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

(order matters for permutations, no repetition)

So far, we have been talking about arranging things in a definite order
Suppose instead that we ask how many committees of r people can be chosen from n people $(n \geq r)$?
Here the number of combinations of n things r at a time is denoted by C_n^r, $C(n,r)$, or $\binom{n}{r}$. To find C_n^r, we take P_n^r and divide by $r!$ (# of ways to permute r items)

$$C(n,r) = \frac{P(n,r)}{r!} = \frac{n!}{(n-r)!\,r!}$$

(order does not matter, no repetition)

Repeats Allowed		Repeats not Allowed	
order matters	order doesn't matter	order matters	order doesn't matter
$\underbrace{n \cdot n \cdot n \cdots n}_{r \text{ times}} = n^r$	$\frac{n^r}{r!}$	$P(n,r) = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$	$C(n,r) = \frac{P(n,r)}{r!}$

Ex. A club consists of 50 members.

a) How many ways can a president, vice-president, secretary, and treasurer be chosen?

b) How many ways can a committee of 4 members be chosen?

a) $P(50,4) = \frac{50!}{46!} = 5527200$

b) $C(50,4) = \frac{P(50,4)}{4!} = 230300$

Ex. How many outcomes are there if you toss a coin 10 times?

$n=2, \ r=10$

$2^{10} = 1024$

Ex. How many five-letter "words" are there?

$n=26 \quad r=5$

$n^r = 26^5 = 11881376$

Ex. How many ways to order a pizza with 3 toppings from a menu of 10 toppings?

$n=10 \quad r=3$

$\frac{10^3}{3!} = \frac{1000}{3!} = \frac{500}{3}$

Ex. How many 4 digit PINs are possible if no repeats are allowed?

$10 \cdot 9 \cdot 8 \cdot 7 = 5040 \qquad \frac{10!}{(10-4)!}$

Ex. How many ways to pick 6 numbers between 1 and 50 (like in a lotto game)? What is the probability of guessing the correct 6 numbers?

$n=50 \quad r=6$

order doesn't matter

repeats not allowed

$C(50,6) = \frac{50!}{(50-6)!\,6!} = 15890700$

$p(A) = \frac{1}{C(50,6)} = \frac{1}{15890700}$

Random Variables

Consider the problem of tossing two fair dice: $S = \{(1,1),(1,2),\dots,(6,5),(6,6)\}$ 36 points

We may be more interested in the value of the sum of the two numbers.

Let X be the sum. Then $\forall$ point $\in X$, X has a value of x.

$\underset{x=2}{(1,1)}, \underset{x=3}{(1,2)}, \dots, \underset{x=12}{(6,6)}$. Such a variable X, which has definite value for each point in S, is called a random variable.

We can easily construct many more examples of random variables, such as:

X = number on first die minus number on the second die

X = number on second die

X = probability associated with the sample point

$X = \begin{cases} 1, & \text{the sum is 7 or 11} \\ 0, & \text{otherwise} \end{cases}$

For each of these, we can setup a table listing all the sample points and next to each point the corresponding value of X.

So, X becomes like a function of the sample points.

Given a point $\in S$, we can find a corresponding value of X, if we are given a description of X.

Probability Functions

Let X = "sum of numbers on the two dice."

There are several points in S for which $x=5$, similarly for other x-values

It is then convenient to lump together all the sample points corresponding to a given x-value. We consider a new sample space in which each point corresponds to one value of x

X	2	3	4	5	6	7	8	9	10	11	12
$p=f(x)$	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	$5/36$	$4/36$	$3/36$	$2/36$	$1/36$

probability density function (PDF)

associated probability to x-value

Each X_i has a probability p_i, $p_i = f(x_i) = P(X=x_i)$

e.g. $f(3) = P(X=3) = 2/36$

f is called the probability density function. Also called probability function for the random variable X.

What is the probability that the sum is between 3 and 5?

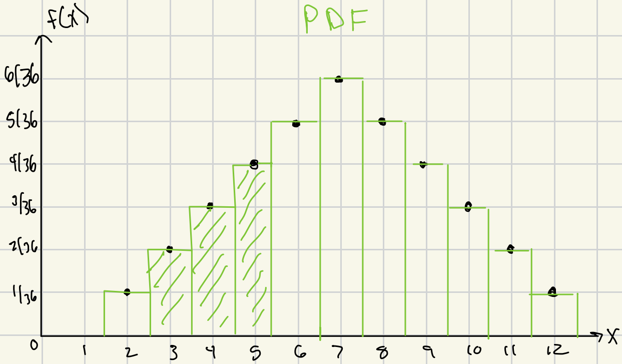
$P(3 \le X \le 5) = P(X=3) + P(x=4) + P(x=5) = f(3) + f(4) + f(5) = 2/36 + 3/36 + 4/36 = 9/36$

What is the probability the sum is greater than 10?

$$P(X>10) = P(X=11) + P(X=12) = f(11) + f(12) = 2/36 + 1/36 = 3/36$$

Discrete Random Variable - finite number of discrete values

Continuous Random Variable -



$P(3 \leq X \leq 5)$ = sum of the areas of the rectangles shaded

$$= 1\left(\frac{2}{36}\right) + 1\left(\frac{3}{36}\right) + 1\left(\frac{4}{36}\right)$$

$$= \frac{9}{36}$$

Cumulative Distribution Function (CDF)

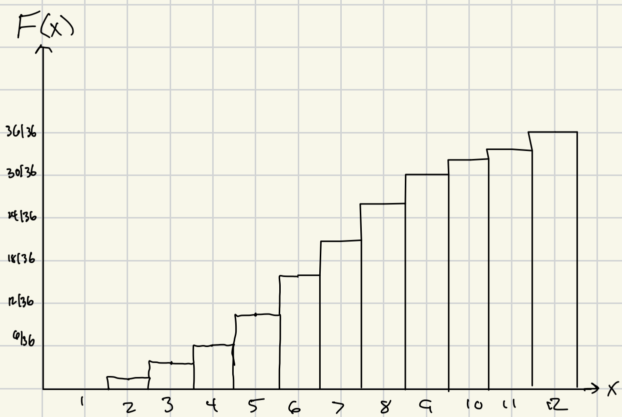
$$F(x) = P(X \leq x)$$

$$F(3) = P(X \leq 3) = p(X=2) + p(X=3) = f(2) + f(3) = \frac{1}{36} + \frac{2}{36} = \frac{3}{36}$$

$$F(5) = P(X \leq 5) = F(3) + f(4) + f(5) = \frac{3}{36} + \frac{3}{36} + \frac{4}{36} = \frac{10}{36}$$

CDF

X	2	3	4	5	6	7	8	9	10	11	12
$F(X) = P(X \leq x)$	1/36	3/36	6/36	10/36	15/36	21/36	26/36	30/36	33/36	35/36	36/36



Summary Statistics

Mean or Expected Value: $\mu = E(x) = \sum_i x_i \, p(X = x_i)$

$$= x_1 \, p(X = x_1) + x_2 \, p(X = x_2) + \cdots$$

$$= x_1 f(x_1) + x_2 f(x_2) + \cdots$$

For our example, $E(x) = 2\left(\frac{1}{36}\right) + 3\left(\frac{2}{36}\right) + 4\left(\frac{3}{36}\right) + 5\left(\frac{4}{36}\right) + 6\left(\frac{5}{36}\right) + 7\left(\frac{6}{36}\right) + 8\left(\frac{5}{36}\right) + 9\left(\frac{4}{36}\right) + 10\left(\frac{3}{36}\right)$
$+ 11\left(\frac{2}{36}\right) + 12\left(\frac{1}{36}\right) = 7$

Variance: $Var(x) = \sum_i (x_i - \mu)^2 f(x_i)$

$$= (x_1 - \mu)^2 f(x_1) + (x_2 - \mu)^2 f(x_2) + \cdots$$

Note: $(x_i - \mu)^2$ is deviation from the mean

For our example, $var(x) = (2-7)^2\left(\frac{1}{36}\right) + (3-7)^2\left(\frac{2}{36}\right) + (4-7)^2\left(\frac{3}{36}\right) + (5-7)^2\left(\frac{4}{36}\right) + (6-7)^2\left(\frac{5}{36}\right) + \underbrace{(7-7)^2\left(\frac{6}{36}\right)}_{0}$
$+ (8-7)^2\left(\frac{5}{36}\right) + (9-7)^2\left(\frac{4}{36}\right) + (10-7)^2\left(\frac{3}{36}\right) + (11-7)^2\left(\frac{2}{36}\right) + (12-7)^2\left(\frac{1}{36}\right)$

Standard deviation: $\sigma = \sqrt{var(x)}$

Ex: Suppose you will be paid \$5 if a die shows a 5, \$2 if it shows 2 or 3, and nothing otherwise. Find the expected gain.
Let X = "gain in playing game"

X	0	2	5
$f(x)$	$3/6$	$2/6$	$1/6$

$E(x) = 0\left(\frac{3}{6}\right) + 2\left(\frac{2}{6}\right) + 5\left(\frac{1}{6}\right) = \frac{9}{6} = \$1.50$

Probability Distributions

Random Variable

Discrete
Binomial Distribution
(Directly or using tables)

$f(x)$ PDF
$F(x)$ CDF

Gaussian (Normal) Distribution
standard normal distribution ← Empirical Rule
(use tables)

Probability Distributions

M. BEN-AZZOUZ

MAT 2215

Overview

Discrete Random Variables

- Probability Density Function (PDF) for a Discrete Random Variable
- Mean or Expected Value and Standard Deviation
- Discrete Probability Distributions
  - Binomial Distribution

Continuous Random Variables

- Continuous Probability Functions
- Examples of Continuous Distributions
- The Normal (Gaussian) Distribution
- Z-Scores and the Empirical Rule
- The Standard Normal Distribution

Random Variable Notation

A **random variable** describes the outcomes of a statistical experiment in words. The values of a random variable can vary with each repetition of an experiment.

Upper case letters such as X or Y denote a random variable. Lower case letters like x or y denote the value of a random variable. If **X is a random variable, then X is written in words, and x is given as a number.**

For example, let X = the number of heads you get when you toss three fair coins. The sample space for the toss of three fair coins is S={*TTT, THH, HTH, HHT, HTT, THT, TTH, HHH*}. Then, x = 0, 1, 2, 3. X is in words and x is a number. Notice that for this example, the x values are countable outcomes. Because you can count the possible values that X can take on and the outcomes are random (the x values 0, 1, 2, 3), X is a discrete random variable.

PDF for a Discrete Random Variable

A discrete **probability distribution (or density) function** has two characteristics:

- 1. Each probability is between zero and one, inclusive.
- 2. The sum of the probabilities is one.

Example

A hospital researcher is interested in the number of times the average post-op patient will ring the nurse during a 12-hour shift. For a random sample of 50 patients, the following information was obtained. Let *X* = the number of times a patient rings the nurse during a 12-hour shift. For this exercise, *x* = 0, 1, 2, 3, 4, 5. f(x)=*P*(X = *x*) = the probability that *X* takes on value *x*. Why is this a discrete probability distribution function (two reasons)?

– the # of values of x is countable

–

X	*P*(*x*)
0	$P(x = 0) = \frac{4}{50}$
1	$P(x = 1) = \frac{8}{50}$
2	$P(x = 2) = \frac{16}{50}$
3	$P(x = 3) = \frac{14}{50}$
4	$P(x = 4) = \frac{6}{50}$
5	$P(x = 5) = \frac{2}{50}$

Mean or Expected Value and Standard Deviation

The **expected value** is often referred to as the **"long-term" average or mean**. This means that over the long term of doing an experiment over and over, you would **expect** this average.

The Law of Large Numbers states that, as the number of trials in a probability experiment increases, the difference between the theoretical probability of an event and the relative frequency approaches zero **(the theoretical probability and the relative frequency get closer and closer together)**.

When evaluating the long-term results of statistical experiments, we often want to know the “average” outcome. This “long-term average” is known as the **mean** or **expected value** of the experiment and is denoted by the Greek letter μ. In other words, after conducting many trials of an experiment, you would expect this average value.

Expected Value—Example

A men's soccer team plays soccer zero, one, or two days a week. The probability that they play zero days is 0.2, the probability that they play one day is 0.5, and the probability that they play two days is 0.3. Find the long-term average or expected value, μ, of the number of days per week the men's soccer team plays soccer.

- To do the problem, first let the random variable X = the number of days the men's soccer team plays soccer per week. X takes on the values 0, 1, 2. Construct a PDF table adding a column $x*P(x)$. In this column, you will multiply each x value by its probability.

Add the last column $x*P(x)$ to find the long-term average or expected value: (0)(0.2) + (1)(0.5) + (2)(0.3) = 0 + 0.5 + 0.6 = 1.1.

The expected value is 1.1. The men's soccer team would, on the average, expect to play soccer 1.1 days per week. The number 1.1 is the long-term average or expected value if the men's soccer team plays soccer week after week after week. We say μ = 1.1.

x	***P*(*x*)**	***x**P*(*x*)**
0	0.2	(0)(0.2) = 0
1	0.5	(1)(0.5) = 0.5
2	0.3	(2)(0.3) = 0.6

Standard Deviation

Like data, probability distributions have standard deviations. To calculate the standard deviation (σ) of a probability distribution, find each deviation from its expected value, square it, multiply it by its probability, add the products, and take the square root. To understand how to do the calculation, look at the table for the number of days per week a men's soccer team plays soccer. To find the standard deviation, add the entries in the column labeled $(x - \mu)^2 P(x)$ and take the square root.

Add the last column in the table. 0.242 + 0.005 + 0.243 = 0.490. The standard deviation is the square root of 0.49, or σ = sqrt(0.49) = 0.7

Note: the variance is the square of the standard deviation.

x	$P(x)$	$x*P(x)$	$(x - \mu)^2 P(x)$
0	0.2	(0)(0.2) = 0	$(0 - 1.1)^2(0.2) = 0.242$
1	0.5	(1)(0.5) = 0.5	$(1 - 1.1)^2(0.5) = 0.005$
2	0.3	(2)(0.3) = 0.6	$(2 - 1.1)^2(0.3) = 0.243$

Example

A hospital researcher is interested in the number of times the average post-op patient will ring the nurse during a 12-hour shift. For a random sample of 50 patients, the following information was obtained.

x	$P(x)$
0	$P(x = 0) = \frac{4}{50}$
1	$P(x = 1) = \frac{8}{50}$
2	$P(x = 2) = \frac{16}{50}$
3	$P(x = 3) = \frac{14}{50}$
4	$P(x = 4) = \frac{6}{50}$
5	$P(x = 5) = \frac{2}{50}$

- What is the expected value?

$$\mu = 0\left(\frac{4}{50}\right) + 1\left(\frac{8}{50}\right) + 2\left(\frac{16}{50}\right) + 3\left(\frac{14}{50}\right) + 4\left(\frac{6}{50}\right) + 5\left(\frac{2}{50}\right) = \frac{0}{50} + \frac{32}{50} + \frac{42}{50} + \frac{24}{50} + \frac{10}{50} = \frac{116}{50} = \frac{58}{25}$$

- What is the standard deviation?

$$Var(x) = \left(0 - \frac{116}{50}\right)^2\left(\frac{4}{50}\right) + \left(1 - \frac{116}{50}\right)^2\left(\frac{8}{50}\right) + \left(2 - \frac{116}{50}\right)^2\left(\frac{16}{50}\right) + \left(3 - \frac{116}{50}\right)^2\left(\frac{14}{50}\right) + \left(4 - \frac{116}{50}\right)^2\left(\frac{6}{50}\right) + \left(5 - \frac{116}{50}\right)^2\left(\frac{2}{50}\right)$$

$$= 0.431 + 0.279 + 0.033 + 0.129 + 0.339 + 0.287 = 1.498$$

$$\sigma = \sqrt{Var(x)} = \sqrt{1.498} = 1.224$$

Binomial Distribution

There are three characteristics of a binomial experiment.

- 1. There are a fixed number of trials. Think of trials as repetitions of an experiment. The letter n denotes the number of trials.
- 2. There are only two possible outcomes, called "success" and "failure," for each trial. The letter p denotes the probability of a success on one trial, and q denotes the probability of a failure on one trial. $p + q = 1$.
- 3. The n trials are independent and are repeated using identical conditions. Because the n trials are independent, the outcome of one trial does not help in predicting the outcome of another trial.

Binomial Distribution—Introduction

The outcomes of a binomial experiment fit a **binomial probability distribution**. The random variable X = the number of successes obtained in the n independent trials.

The mean, μ, and variance, Var(x)=σ^2, for the binomial probability distribution are

- $\mu = np$ and
- $\sigma^2 = npq$.
- The standard deviation, σ, is then

$$\sigma = \sqrt{npq}$$

Any experiment that has characteristics two and three and where $n = 1$ is called a **Bernoulli Trial**. A binomial experiment takes place when the number of successes is counted in one or more Bernoulli Trials.

n: # of trials
p: probability of success
q: Probability of failure

Notation for the Binomial: B = Binomial Distribution Function—1

***X* ~ *B*(*n*, *p*)**

Read this as "*X* is a random variable with a binomial distribution." The parameters are *n* and *p*; *n* = number of trials, *p* = probability of a success on each trial. Formula:

$$PDF: \; f(x) = C(n,x)\,p^x q^{n-x}$$

Toss a coin 5 times:
$n=5 \quad p=q=\frac{1}{2}$
Probability of landing heads 3 times:
$f(3) = C(5,3)\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{5-3} = 0.3125$

$$CDF:$$

$$F(x) = P(X \le x) = f(0) + f(1) + f(2) + \ldots + f(x) =$$

$$= C(n,0)\,p^0 q^{n-0} + C(n,1)\,p^1 q^{n-1} + C(n,\underset{2}{\cancel{0}})\,p^2 q^{n-2} + \ldots + C(n,x)\,p^x q^{n-x}$$

Binomial Distribution--Example

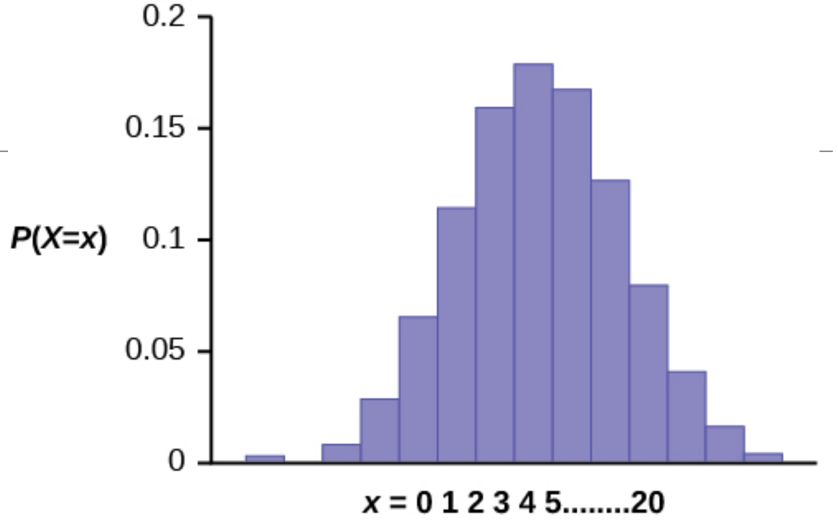
- It has been stated that about 41% of adult workers have a high school diploma but do not pursue any further education. If 20 adult workers are randomly selected, find the probability that at most 3 of them have a high school diploma but do not pursue any further education. How many adult workers do you expect to have a high school diploma but do not pursue any further education?
  - Let *X* = the number of workers who have a high school diploma but do not pursue any further education.
  - *X* takes on the values 0, 1, 2, ..., 20 where *n* = 20, *p* = 0.41, and *q* = 1 – 0.41 = 0.59. *X* ~ *B*(20, 0.41)

$$P(X \le 3) = f(0) + f(1) + f(2) + f(3)$$

$$= \binom{20}{0}(0.41)^0(0.59)^{20} + \binom{20}{1}(0.41)^1(0.59)^{19} + \binom{20}{2}(0.41)^2(0.59)^{18} + \binom{20}{3}(0.41)^3(0.59)^{17}$$

$$= 0 + 0 + 0.003 + 0.012 = 0.015$$

Notation for the Binomial: B = Binomial Distribution Function—2



The graph of $X \sim B(20, 0.41)$

is as follows:

The *y*-axis contains the probability of *x*, where *X* = the number of workers who have only a high school diploma.

The number of adult workers that you expect to have a high school diploma but not pursue any further education

is the mean, $\mu = np = (20)(0.41) = 8.2$.

The formula for the variance is $\sigma^2 = npq$. The standard deviation is

$$\sigma = \sqrt{npq} = \sqrt{(20)(0.41)(0.59)} = 2.20$$

Technology use to get P(x<=12)

Using the TI-83, 83+, 84, 84+ Calculator

Go into 2nd DISTR. The syntax for the instructions are as follows:

To calculate (x = value): binompdf(n, p, number) if "number" is left out, the result is the binomial probability table.
To calculate $P(x \le \text{value})$: binomcdf(n, p, number) if "number" is left out, the result is the cumulative binomial probability table.

For this problem: After you are in 2nd DISTR, arrow down to binomcdf. Press ENTER. Enter 20,0.41,12). The result is $P(x \le 12) = 0.9738$.

NOTE

If you want to find $P(x = 12)$, use the pdf (binompdf). If you want to find $P(x > 12)$, use 1 - binomcdf(20,0.41,12).

The probability that at most 12 workers have a high school diploma but do not pursue any further education is 0.9738.

Binomal Distribution Table (PDF)

Numbers in the table represent $p(X=x)$ for a binomial distribution with n trials and probability of success p.

Binomial probabilities:

$\binom{n}{x} p^x (1-p)^{n-x}$

n	x	p = 0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9
15	0	0.206	0.035	0.013	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.343	0.132	0.067	0.031	0.005	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.267	0.231	0.156	0.092	0.022	0.003	0.000	0.000	0.000	0.000	0.000
	3	0.129	0.250	0.225	0.170	0.063	0.014	0.002	0.000	0.000	0.000	0.000
	4	0.043	0.188	0.225	0.219	0.127	0.042	0.007	0.001	0.000	0.000	0.000
	5	0.010	0.103	0.165	0.206	0.186	0.092	0.024	0.003	0.001	0.000	0.000
	6	0.002	0.043	0.092	0.147	0.207	0.153	0.061	0.012	0.003	0.001	0.000
	7	0.000	0.014	0.039	0.081	0.177	0.196	0.118	0.035	0.013	0.003	0.000
	8	0.000	0.003	0.013	0.035	0.118	0.196	0.177	0.081	0.039	0.014	0.000
	9	0.000	0.001	0.003	0.012	0.061	0.153	0.207	0.147	0.092	0.043	0.002
	10	0.000	0.000	0.001	0.003	0.024	0.092	0.186	0.206	0.165	0.103	0.010
	11	0.000	0.000	0.000	0.001	0.007	0.042	0.127	0.219	0.225	0.188	0.043
	12	0.000	0.000	0.000	0.000	0.002	0.014	0.063	0.170	0.225	0.250	0.129
	13	0.000	0.000	0.000	0.000	0.000	0.003	0.022	0.092	0.156	0.231	0.267
	14	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.031	0.067	0.132	0.343
	15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.013	0.035	0.206
20	0	0.122	0.012	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.270	0.058	0.021	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.285	0.137	0.067	0.028	0.003	0.000	0.000	0.000	0.000	0.000	0.000
	3	0.190	0.205	0.134	0.072	$\boxed{0.012}$	0.001	0.000	0.000	0.000	0.000	0.000
	4	0.090	0.218	0.190	0.130	$\boxed{0.035}$	0.005	0.000	0.000	0.000	0.000	0.000
	5	0.032	0.175	0.202	0.179	$\boxed{0.075}$	0.015	0.001	0.000	0.000	0.000	0.000
	6	0.009	0.109	0.169	0.192	0.124	0.037	0.005	0.000	0.000	0.000	0.000
	7	0.002	0.055	0.112	0.164	0.166	0.074	0.015	0.001	0.000	0.000	0.000
	8	0.000	0.022	0.061	0.114	0.180	0.120	0.035	0.004	0.001	0.000	0.000
	9	0.000	0.007	0.027	0.065	0.160	0.160	0.071	0.012	0.003	0.000	0.000
	10	0.000	0.002	0.010	0.031	0.117	0.176	0.117	0.031	0.010	0.002	0.000
	11	0.000	0.000	0.003	0.012	0.071	0.160	0.160	0.065	0.027	0.007	0.000
	12	0.000	0.000	0.001	0.004	0.035	0.120	0.180	0.114	0.061	0.022	0.000
	13	0.000	0.000	0.000	0.001	0.015	0.074	0.166	0.164	0.112	0.055	0.002
	14	0.000	0.000	0.000	0.000	0.005	0.037	0.124	0.192	0.169	0.109	0.009
	15	0.000	0.000	0.000	0.000	0.001	0.015	0.075	0.179	0.202	0.175	0.032
	16	0.000	0.000	0.000	0.000	0.000	0.005	0.035	0.130	0.190	0.218	0.090
	17	0.000	0.000	0.000	0.000	0.000	0.001	0.012	0.072	0.134	0.205	0.190
	18	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.028	0.067	0.137	0.285
	19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.021	0.058	0.270
	20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.012	0.122

n=20

								p											
x	**0.05**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.35**	**0.4**	**0.45**	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**	**0.95**
0	0.358	0.122	0.039	0.012	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.736	0.392	0.176	0.069	0.024	0.008	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.925	0.677	0.405	0.206	0.091	0.035	0.012	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.984	0.867	0.648	0.411	0.225	0.107	0.044	0.016	0.005	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.997	0.957	0.830	0.630	0.415	0.238	0.118	0.051	0.019	0.006	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	1.000	0.989	0.933	0.804	0.617	0.416	0.245	0.126	0.055	0.021	0.006	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	1.000	0.998	0.978	0.913	0.786	0.608	0.417	0.250	0.130	0.058	0.021	0.006	0.002	0.000	0.000	0.000	0.000	0.000	0.000
7	1.000	1.000	0.994	0.968	0.898	0.772	0.601	0.416	0.252	0.132	0.058	0.021	0.006	0.001	0.000	0.000	0.000	0.000	0.000
8	1.000	1.000	0.999	0.990	0.959	0.887	0.762	0.596	0.414	0.252	0.131	0.057	0.020	0.005	0.001	0.000	0.000	0.000	0.000
9	1.000	1.000	1.000	0.997	0.986	0.952	0.878	0.755	0.591	0.412	0.249	0.128	0.053	0.017	0.004	0.001	0.000	0.000	0.000
10	1.000	1.000	1.000	0.999	0.996	0.983	0.947	0.872	0.751	0.588	0.409	0.245	0.122	0.048	0.014	0.003	0.000	0.000	0.000
11	1.000	1.000	1.000	1.000	0.999	0.995	0.980	0.943	0.869	0.748	0.586	0.404	0.238	0.113	0.041	0.010	0.001	0.000	0.000
12	1.000	1.000	1.000	1.000	1.000	0.999	0.994	0.979	0.942	0.868	0.748	0.584	0.399	0.228	0.102	0.032	0.006	0.000	0.000
13	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.994	0.979	0.942	0.870	0.750	0.583	0.392	0.214	0.087	0.022	0.002	0.000

Binomal Distribution Table (CDF)

Continuous Probability Distributions

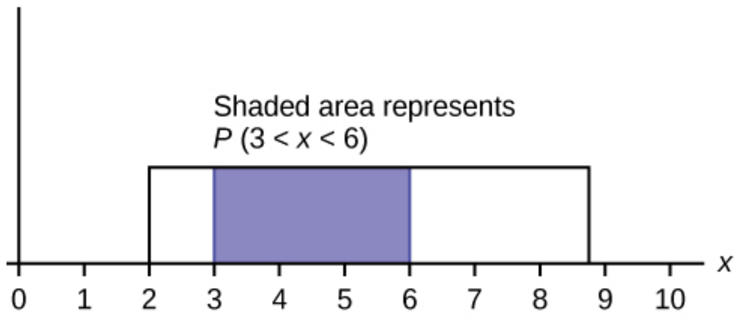
The graph of a continuous probability distribution is a curve. Probability is represented by area under the curve.

The curve is called the **probability density function** (abbreviated as **pdf or PDF**). We use the symbol $f(x)$ to represent the curve. $f(x)$ is the function that corresponds to the graph; we use the density function $f(x)$ to draw the graph of the probability distribution.

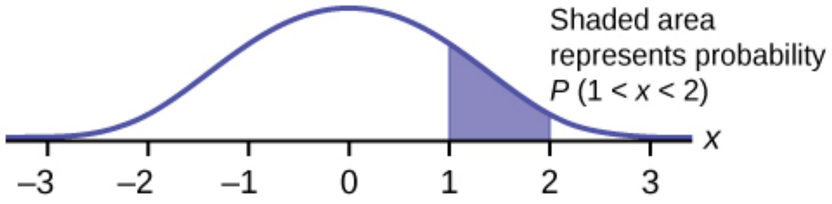
Cumulative Distribution Function F(x)

Area under the curve is given by a different function called the **cumulative distribution function** (abbreviated as **cdf or CDF**) and denoted F(x). The cumulative distribution function is used to evaluate probability as area.

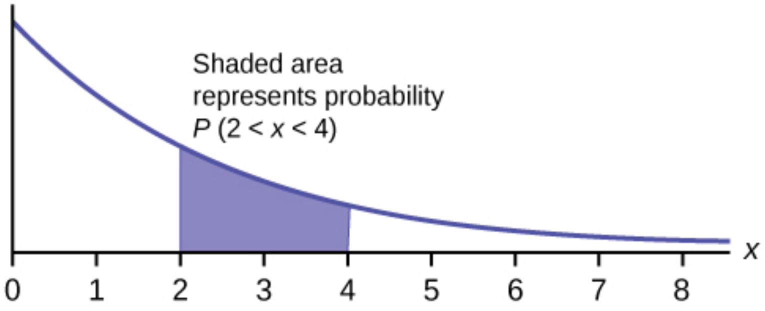
- The outcomes are measured, not counted.
- The entire area under the curve and above the x-axis is equal to one.
- Probability is found for intervals of *x*-values rather than for individual *x*-values.
- $P(c < X < d)$ is the probability that the random variable X is in the interval between the values c and d. $P(c < X < d)$ is the area under the curve, above the *x*-axis, to the right of c and the left of d.
- $P(X = c) = 0$ The probability that X takes on any single individual value is zero. The area below the curve, above the *x*-axis, and between $X = c$ and $X = c$ has no width, and therefore no area (area = 0). Since the probability is equal to the area, the probability is also zero.
- $P(c < X < d)$ is the same as $P(c \leq X \leq d)$ because probability is equal to area.



The uniform distribution



The normal distribution



The exponential distribution

Uniform, Exponential, and Gaussian Distributions

The normal, a continuous distribution, is the most important of all the distributions. Its graph is bell-shaped. You see the bell curve in almost all disciplines. Some of these include psychology, business, economics, the sciences, nursing, and, of course, mathematics. Some of your instructors may use the normal distribution to help determine your grade. Most IQ scores are normally distributed. Often real-estate prices fit a normal distribution. The normal distribution is extremely important, but it cannot be applied to everything in the real world.

We will focus on the normal distribution and the standard normal distribution.

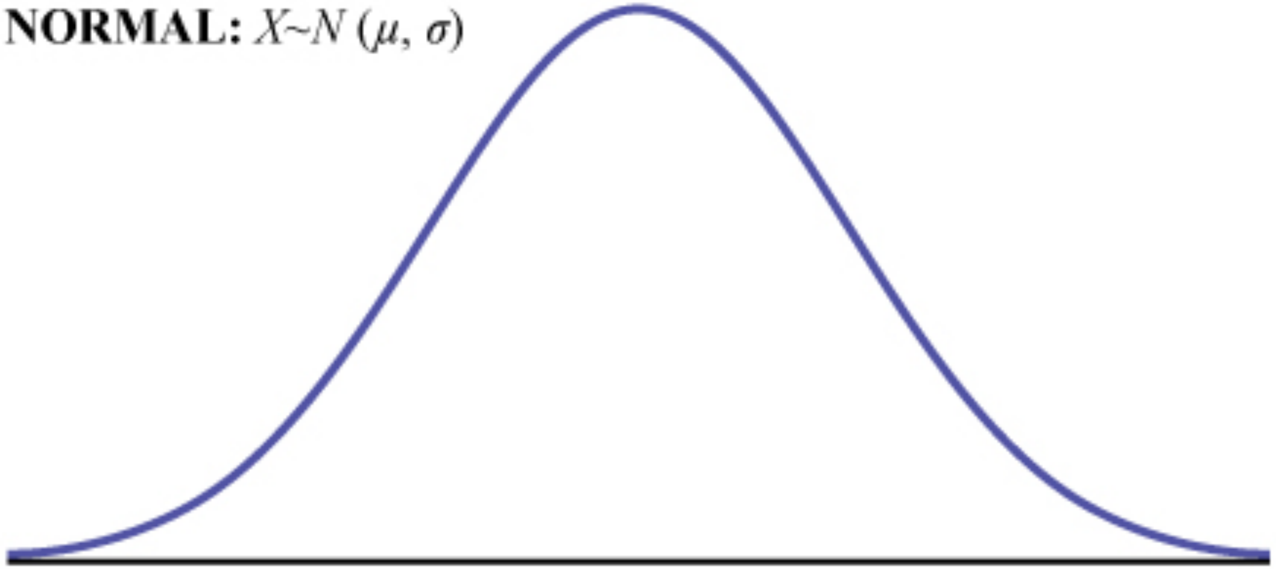
The Gaussian (Normal) Distribution

Mean and Standard Deviation—1

The normal distribution has two parameters (two numerical descriptive measures), the mean (μ) and the standard deviation (σ). If X is a quantity to be measured that has a normal distribution with mean (μ) and standard deviation (σ), we designate this by writing

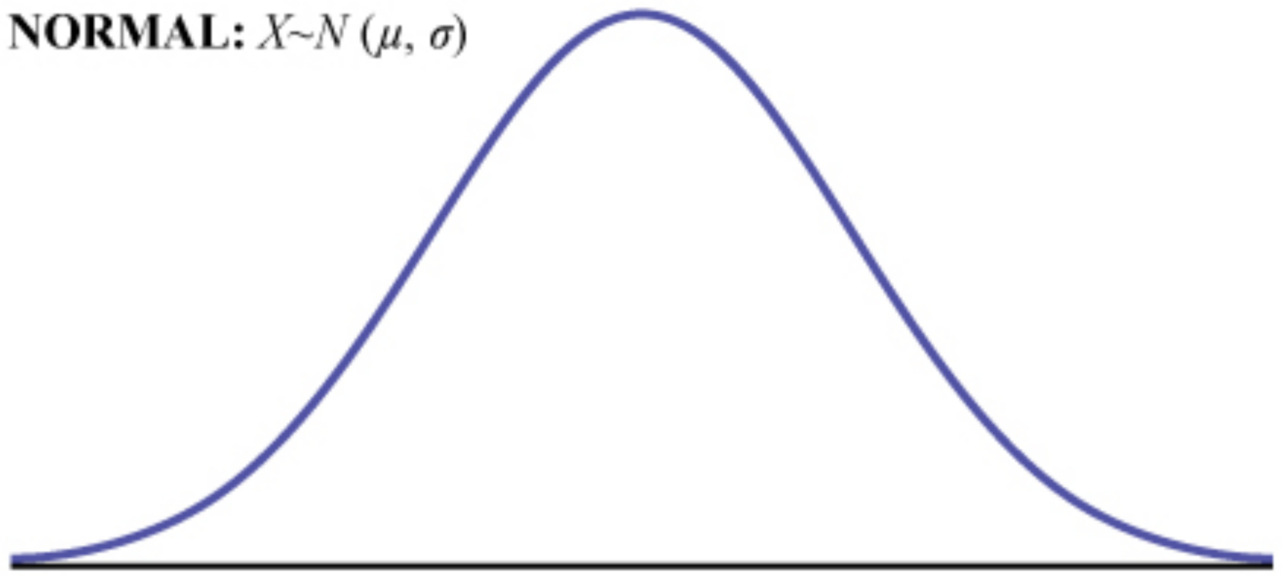
$X \sim N(\mu, \sigma)$

The cumulative distribution function is $P(X < x)$. It is calculated either by a calculator or a computer, or it is looked up in a table.



Mean and Standard Deviation—2

The curve is symmetrical about a vertical line drawn through the mean, μ. In theory, the mean is the same as the median, because the graph is symmetric about μ. As the notation indicates, the normal distribution depends only on the mean and the standard deviation. Since the area under the curve must equal one, a change in the standard deviation, σ, causes a change in the shape of the curve; the curve becomes fatter or skinnier depending on σ. A change in μ causes the graph to shift to the left or right. This means there are an infinite number of normal probability distributions. One of special interest is called the **standard normal distribution**.



The Normal Distribution: Formulas

$$X \sim N(\mu, \sigma)$$

$$PDF: \; f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$CDF: \; F(x) = P(X < x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt$$

The Standard Normal Distribution

The **standard normal distribution** is a normal distribution of **standardized values called *z*-scores**. **A *z*-score is measured in units of the standard deviation.**

- For example, if the mean of a normal distribution is five and the standard deviation is two, the value 11 is three standard deviations above (or to the right of) the mean. The calculation is as follows:

$$x = \mu + (z)(\sigma) = 5 + (3)(2) = 11$$

  - The *z*-score is three.
- The mean for the standard normal distribution is zero, and the standard deviation is one. The transformation

$$z = \frac{x - \mu}{\sigma}$$

  - produces the distribution ***Z* ~ *N*(0, 1)**. The value x comes from a normal distribution with mean μ and standard deviation σ.

Z-Scores

$$z = \frac{x - \mu}{\sigma}$$

If X is a normally distributed random variable and $X \sim N(\mu, \sigma)$, then the *z*-score is:

The *z*-score tells you how many standard deviations the value *x* is above (to the right of) or below (to the left of) the mean, μ. Values of *x* that are larger than the mean have positive *z*-scores, and values of *x* that are smaller than the mean have negative *z*-scores. If *x* equals the mean, then *x* has a *z*-score of zero.

- **The *z*-score allows us to compare data that are scaled differently.**

Suppose $X \sim N(5, 6)$. This says that x is a normally distributed random variable with mean $\mu = 5$ and standard deviation $\sigma = 6$. Suppose $x = 17$. Then:

$$z = \frac{x - \mu}{\sigma} = \frac{17 - 5}{6} = 2$$

This means that $x = 17$ is **two standard deviations** (2σ) above or to the right of the mean $\mu = 5$. The standard deviation is $\sigma = 6$.

Notice that: $5 + (2)(6) = 17$ (The pattern is $\mu + z\sigma = x$)

Now suppose $x = 1$. Then: $z = \dfrac{x - \mu}{\sigma} = \dfrac{1 - 5}{6} = -0.67$ (rounded to two decimal places)

This means that $x = 1$ is 0.67 standard deviations (-0.67σ) below or to the left of the mean $\mu = 5$. Notice that: $5 + (-0.67)(6)$ is approximately equal to one (This has the pattern $\mu + (-0.67)\sigma = 1$)

Summarizing, when z is positive, x is above or to the right of μ and when z is negative, x is to the left of or below μ. Or, when z is positive, x is greater than μ, and when z is negative x is less than μ.

Example (Z-Score)

The Empirical Rule

If X is a random variable and has a normal distribution with mean μ and standard deviation σ, then the **Empirical Rule** says the following:

- About 68% of the x values lie between -1σ and $+1\sigma$ of the mean μ (within one standard deviation of the mean).
- About 95% of the x values lie between -2σ and $+2\sigma$ of the mean μ (within two standard deviations of the mean).
- About 99.7% of the x values lie between -3σ and $+3\sigma$ of the mean μ (within three standard deviations of the mean).
  - Notice that almost all the x values lie within three standard deviations of the mean.

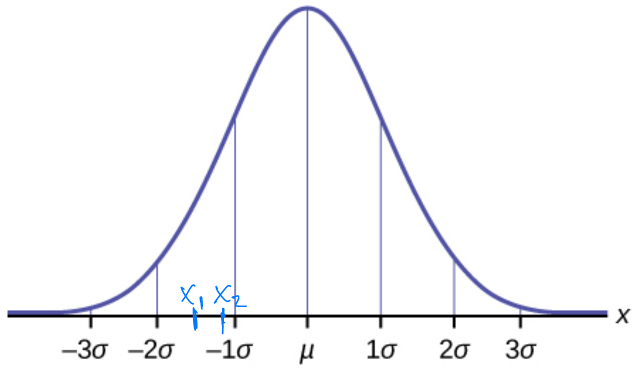
The empirical rule is also known as the 68-95-99.7 rule.

The Empirical Rule—Example

The empirical rule is also known as the 68-95-99.7 rule.

In 2012, 1,664,479 students took the SAT exam. The distribution of scores in the verbal section of the SAT had a mean 496 and standard deviation 114. Let X = a SAT exam verbal section score in 2012. Then $X \sim N(496, 114)$

Find the z-scores for x1 = 325 and x2 = 366.21. Interpret each z-score. What can you say about x1=325 and x2 = 366.21?



$\mu = 496 \quad \sigma = 114$

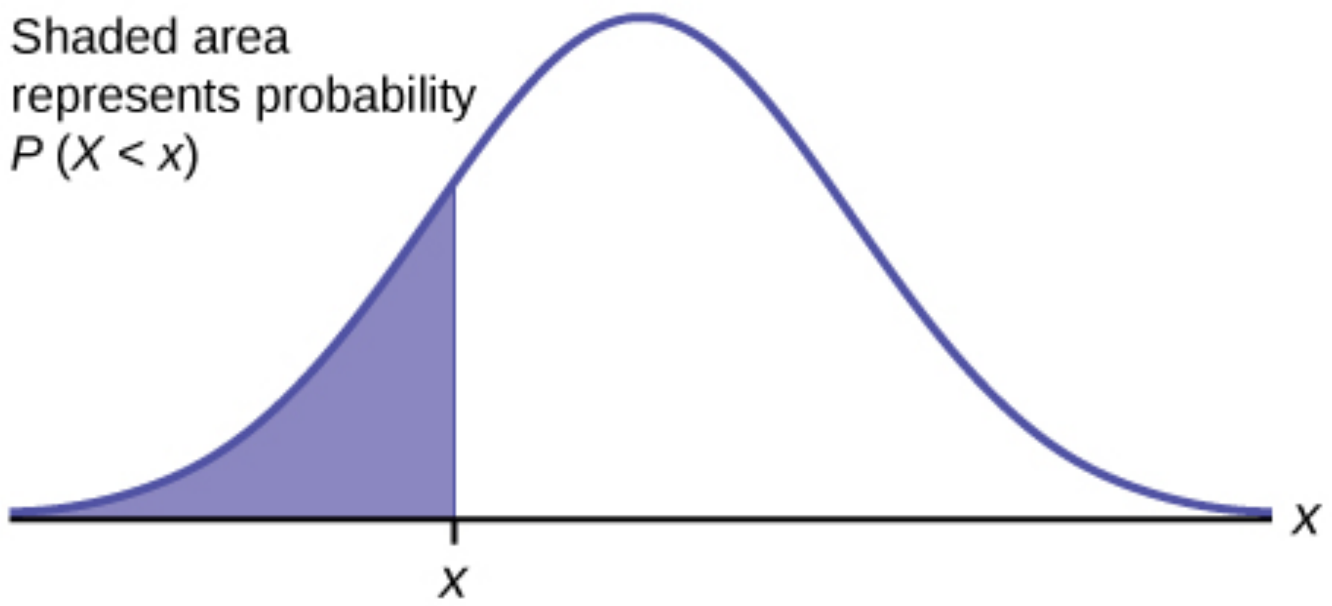
$z_1 = \dfrac{x_1 - \mu}{\sigma} = \dfrac{325 - 496}{114} = \dfrac{-171}{114} = -1.5$

$z_2 = \dfrac{366.21 - 496}{114} = -1.14$

Using the Normal Distribution

The shaded area in the following graph indicates the area to the left of *x*. This area is represented by the probability *P*(*X* < *x*). Normal tables, computers, and calculators provide or calculate the probability *P*(*X* < *x*).

- Remember, *P*(*X* < *x*) is the **Area to the left** of the vertical line through *x*.
- *P*(*X* > *x*) = 1 – *P*(*X* < *x*) is the **Area to the right** of the vertical line through *x*.
- *P*(*X* < *x*) is the same as *P*(*X* ≤ *x*) and *P*(*X* > *x*) is the same as *P*(*X* ≥ *x*) for continuous distributions.



Calculations of Probabilities

$\mu = 63 \quad \sigma\ 5$

Example

The final scores in a statistics class were normally distributed with a mean of 63 and a standard deviation of 5.

A. Find the probability that a randomly selected student scored more than 68. $X_a > 68 = 1 - P(X < 68)$
$= 1 - 0.34 - 0.5 = 0.16$

B. Find the probability that a randomly selected student scored between 58 and 68. $P(X_B) = 68\%$

C. Find the probability that a randomly selected student scored less than 53. $P(X_c) = 0.0015 + 0.0235$
$= 0.025$

D. Find the probability that a randomly selected student scored more than 65 (Technology or Z-Scores Table required)

$P(X_D > 65) = 1 - P(X < 65)$

$Z = \frac{65-63}{5} = \frac{2}{5} = 0.4$

$\Rightarrow 1 - 0.6554 = 0.3446$

Z-Scores Table

Standard Normal Probabilities

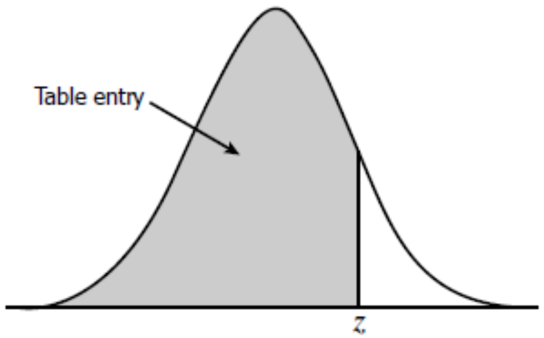


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177

Z-Score Table

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9031	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Using the TI-83, 83+, 84, 84+ Calculator

Go into `2nd DISTR`.
After pressing `2nd DISTR`, press `2:normalcdf`.

The syntax for the instructions are as follows:

normalcdf(lower value, upper value, mean, standard deviation) For this problem: normalcdf(65,1E99,63,5) = 0.3446. You get 1E99 (= 10^{99}) by pressing **1**, the **EE** key (a 2nd key) and then **99**. Or, you can enter **10^99** instead. The number 10^{99} is way out in the right tail of the normal curve. We are calculating the area between 65 and 10^{99}. In some instances, the lower number of the area might be –1E99 (= -10^{99}). The number -10^{99} is way out in the left tail of the normal curve.

Technology Use for Part D

Final Exam Objectives:

1) Matrix Operations

- Product
- Scalar multiplication
- Transpose
- Trace
- Determinant
- Inverse

Vector Operations

- Inner/outer product

2) Subspaces of R^2

3) Basis

4) Linear transformations

- Image of a vector
- Standard matrix

5) Eigenvalue and Eigenvectors

6) Jacobian matrix

7) Multivariate chain rule

8) Hessian

9) Probability (card problem)

10) Probability rules

11) Discrete random variables

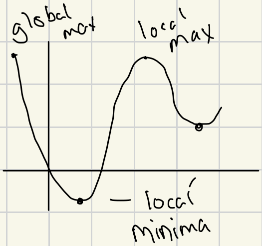
12) Normal distribution (z-score table)

13) Probability (lotto problem)

14) Bonus - surprise probability problem

Optimization

global: occurs over the domain of the function
local: occurs over a restricted domain of the function
Note: a local max/min cannot occur on the edge of the graph



critical number: an x value where the function is defined but its derivative is either 0, $f'(c)=0$ or horizontally tangent, or undefined

- a local max/min must occur at critical points

Caution: Not every critical number leads to a max/min. It could lead to a saddle point
We determine that a critical number yields a local max/min using either the First or Second Derivative Test

Second Derivative Test (univariate)

Let $f(x)$ be a smooth function with a critical number $x=c$.

if f

Multivariate:
A critical point of a 2 variable function is an input (a,b) such that both $f_x(a,b)=0$ and $f_y(a,b)=0$.
In other words, ∇f at (a,b) is 0.

Let $f(x,y)$ be a smooth two variable function, let (a,b) be a critical point of $f(x,y)$ and let $D(a,b) = |H_f| = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = f_{xx}(a,b)f_{yy}(a,b) - [f_{xy}(a,b)]^2$

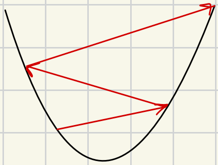
If	Then
$D(a,b) > 0$ and $f_{xx}(a,b) > 0$	$f(a,b)$ is a local minimum
$D(a,b) > 0$ and $f_{xx}(a,b) < 0$	$f(a,b)$ is a local maximum
$D(a,b) < 0$	$f(a,b)$ is a saddle point

Gradient Descent Algorithm

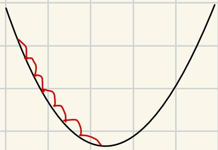
1) Pick an initial random vector $\vec{x}_0$

2) Compute $\vec{x}_{i+1} = \vec{x}_i - \eta \nabla f(\vec{x}_i)$ $\qquad$ η is learning rate ("step size")

3) Repeat for i iterations $\qquad\qquad$ eta



large learning rate



small learning rate

Ex: $3x^4 - 4x^3 - 12x^2 + 14 = g(x)$

$g'(x) = 12x^3 - 12x^2 - 24x$

critical numbers: set $g'(x) = 0$

$0 = 12x^3 - 12x^2 - 24x$

$0 = 12x(x^2 - x - 2)$

$0 = 12x(x-2)(x+1)$

$12x = 0$ or $x = 2$ or $x = -1$

$x = 0$

$g''(x) = 36x^2 - 24x - 24$

$g''(0) = -24 < 0$ concave down

$g''(2) = 144 - 48 - 24 = 72 > 0$ concave up

$g''(-1) = 36 + 24 - 24 = 36 > 0$ concave up

Summary:

local max at $x = 0$, local minima at $x = -1$ and $x = 2$

Ex: $f(x,y) = 1 + (x-2)^2 + (y-1)^2$

$\vec{\nabla} f(x,y) = \begin{bmatrix} 2x-4 \\ 2y-2 \end{bmatrix}$

$f_x(x,y) = 2(x-2) = 2x - 4$ $\qquad$ $f_y(x,y) = 2(y-1) = 2y - 2$ $\qquad$ $\Rightarrow$ critical point $(2,1)$

$$\begin{bmatrix} 2x-4 \\ 2y-2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} -4 \\ -2 \end{bmatrix}$$

```
np.array([[2,0],[0,2]]) @ X_opt + np.array([-4,-2]).reshape(2,1)
```

$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \qquad \begin{bmatrix} x \\ y \end{bmatrix} \qquad \begin{bmatrix} -4 \\ -2 \end{bmatrix}$

$f_{xx}(x,y) = 2$ $\qquad$ $f_{yy}(x,y) = 2$

$f_{xy}(x,y) = 0$ $\qquad$ $f_{yx}(x,y) = 0$

$H_f = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \Rightarrow |H_f| = 4 > 0$ ✓

$f_{xx}(2,1) = 2 > 0$ ✓ $\qquad$ local minimum