Tutorial - 1

Chinmaya Udupa
SC19B081

Q1. Partial Derivatives

a) given $L = \frac{1}{2}(y - f(x))^2$ where $f(x) = \frac{1 + \tanh\left(\frac{wx+b}{2}\right)}{2}$ [x & y are constants, w & b are variable]

find $\frac{\partial L}{\partial w}$ and $\frac{\partial L}{\partial b}$

Sol: first let us find $\frac{d \tanh(x)}{dx}$; $\tanh(x) = \frac{\sinh(x)}{\cosh(x)}$ $\sinh(x) = \frac{e^x + e^{-x}}{2}$

$\cosh(x) = \frac{e^x - e^{-x}}{2}$

$\frac{d}{dx}\sinh(x) = \cosh(x)$ & $\frac{d}{dx}\cosh(x) = \sinh(x)$

$\Rightarrow \frac{d}{dx}\tanh(x) = \frac{\cosh^2(x) - \sinh^2(x)}{\cosh^2(x)}$ [using division rule]

which can be further simplified, as $\cosh^2 x - \sinh^2 x = 1$

$\Rightarrow \frac{d}{dx}\tanh(x) = \text{sech}^2(x)$

$\frac{\partial L}{\partial w} = \frac{\partial}{\partial w}\left(\frac{1}{2}(y - f(x))^2\right) = (y - f(x)) * -\frac{\partial f(x)}{\partial w}$; [chain rule]

$\frac{\partial f}{\partial w} = \text{sech}^2\left(\frac{wx+b}{2}\right) * \frac{x}{2}$ [chain rule]

$\frac{\partial f}{\partial b} = \text{sech}^2\left(\frac{wx+b}{2}\right) * \frac{1}{2}$ [chain rule]

$$\Rightarrow \frac{\partial L}{\partial w} = -(y - f(x)) \times \operatorname{sech}^2\left(\frac{wx+b}{2}\right)\frac{x}{2}$$

|||ly $$\frac{\partial L}{\partial b} = -(y - f(x)) \times \operatorname{sech}^2\left(\frac{wx+b}{2}\right) \times \frac{1}{2}$$ $$\left[\frac{\partial L}{\partial b} = -(y - f(x)) \times \frac{\partial f}{\partial b}\right]$$

$$\boxed{\begin{aligned} \frac{\partial L}{\partial w} &= \left[\left(\frac{1+\tanh\left(\frac{wx+b}{2}\right)}{2} - y\right)\right] \times \operatorname{sech}^2\left(\frac{wx+b}{2}\right) \times \frac{x}{2} \\ \frac{\partial L}{\partial b} &= \left[\left(\frac{1+\tanh\left(\frac{wx+b}{2}\right)}{2} - y\right)\right] \times \operatorname{sech}^2\left(\frac{wx+b}{2}\right) \times \frac{1}{2} \end{aligned}}$$

b) given $E = g(x, y, z) = \sigma\left(c(ax+by) + dz\right)$

& $\sigma(x) = \dfrac{1}{1+e^{-x}}$; x, y, z are constants
a, b, c, d are parameters

find $\dfrac{\partial E}{\partial a}$, $\dfrac{\partial E}{\partial b}$ & $\dfrac{\partial E}{\partial d}$

Soln: $$E = \frac{1}{1 + \exp(-c(ax+by) + dz)}$$ $$\left[\frac{d}{dx}\left(\frac{1}{f(x)}\right) = \frac{-f'(x)}{f(x)^2}\right]$$

$$\frac{\partial E}{\partial a} = \frac{-\partial\left(\exp(-c(ax+by)+dz)\right)/\partial a}{\left(1+\exp(-c(ax+by)+dz)\right)^2}$$

$$= \frac{-\exp(-c(ax+by)+dz) \cdot -cx}{\left(1+\exp(-c(ax+by)+dz)\right)^2}$$

$$\Rightarrow \boxed{\frac{\partial E}{\partial a} = \frac{cx \cdot \exp(-c(ax+by)+dz)}{(1+\exp(-c(ax+by)+dz))^2}}$$

|||ly

$$\boxed{\frac{\partial E}{\partial b} = \frac{cy \cdot \exp(-c(ax+by)+dz)}{(1+\exp(-c(ax+by)+dz))^2}}$$

&

$$\boxed{\frac{\partial E}{\partial d} = \frac{-z \cdot \exp(-c(ax+by)+dz)}{(1+\exp(-c(ax+by)+dz))^2}}$$

Q2 Erroneous estimates

a) $f(x) = x^2 - 2x + 1$ using lim defitnition . S.T $f'(x) = 2x-2$

Sol: $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$

$$= \lim_{h\to 0} \frac{(x+h)^2 - 2(x+h) + \cancel{1} - x^2 + \cancel{2x} - \cancel{1}}{h}$$

$$= \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} + \lim_{h \to 0} \frac{-2h}{h}$$

$$= \left(\lim_{h\to 0} \frac{\cancel{x^2} + h^2 + 2xh - \cancel{x^2}}{h} \right) - 2$$

$$= 2x - 2$$

M	T	W	T	F	S	S

Page No.:
Date:
YOUVA

b $f(x) = x^2 - 2x + 1$

approximate $f(1.01)$ & $f(1.5)$ given $f(1) = 0$

Sol: $f(x+h) \simeq f(x) + h f'(x)$

as computed earlier $f'(x) = 2x - 2$

$f(1.01) \simeq f(1) + 0.01 \times f'(1)$

$= 0 + 0.01 \times 2(\cancel{1-1})^{\to 0}$

$f(1.01) \simeq 0$

$f(1.5) \simeq f(1) + 0.5 \times f'(1)$

$= 0 + 0.5 \times 2(\cancel{1-1})^{\to 0}$

$f(1.5) \simeq 0$

c) actual value of $f(1.01) = 0.0001$ & $f(1.5) = 0.25$
estimate value of $f(1.01) = 0$ & $f(1.5) = 0$

$|\text{error in } f(1.01)| = 0.0001$
$|\text{error in } f(1.5)| = 0.25$

d) because, definition of $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$

as we are fixing x @ 1. we get better estimates "very close" to $x = 1$, as h in these cases are close to 0, hence ~~give~~ gets closer to the actual value of $f(x+h)$.

e) We could use a second degree approximation i.e

$$f(x+h) \simeq f(x) + h\frac{df(x)}{dx} + \frac{h^2}{2}\frac{d^2f(x)}{dx^2}$$

$$\boxed{f''(x) = 2}$$

from this $f(1.01) = 0 + \frac{(0.01)^2}{2} \times 2 = \underline{0.0001}$

& $f(1.5) = 0 + \frac{(0.5)^2}{2} \times 2 = \underline{0.25}$

error in both cases $= 0$

M T W T F S S
Page No.:
Date:
YOUVA

3 Differentiation W.R.T vectors and matrices

a) 1. $\nabla_x u^T x$

let $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ & $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$\Rightarrow u^T \cdot x = \begin{bmatrix} u_1, u_2, \ldots u_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$u^T \cdot x = u_1 x_1 + u_2 x_2 + \ldots + u_n x_n$

$\left(\nabla_x f = \begin{bmatrix} \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots \frac{\partial f}{\partial x_n} \end{bmatrix} \right)$

$\Rightarrow \boxed{\nabla_x u^T x = (u_1, u_2 \ldots u_n) = u^T}$

2. $\nabla_x x^T x$

$x^T \cdot x = \begin{bmatrix} x_1, x_2, \ldots x_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1^2 + x_2^2 + \ldots + x_n^2$

$\boxed{\nabla_x x^T x = \begin{bmatrix} 2x_1, 2x_2, \ldots 2x_n \end{bmatrix} = 2\vec{x}}$

3. $\nabla_x \, x^T A x$

let $A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix}$

~~Matrix~~

$$x^T A = [x_1 A_{11} + x_2 A_{21} + \dots + x_n A_{n1},$$
$$x_1 A_{12} + x_2 A_{22} + \dots + x_n A_{n2},$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$x_1 A_{1n} + x_2 A_{2n} + \dots + x_n A_{nn}]$$

$$x^T A x = x_1(x_1 A_{11} + x_2 A_{21} + \dots + x_n A_{n1}) +$$
$$x_2(x_1 A_{12} + x_2 A_{22} + \dots + x_n A_{n2}) +$$
$$\vdots$$
$$+$$
$$x_n(x_1 A_{1n} + x_2 A_{2n} + \dots + x_n A_{nn})$$

$$\nabla_x x^T A x = [(2x_1 A_{11}) + (x_2 A_{21} + x_2 A_{12}) + \dots + (x_n A_{n1} + x_n A_{1n}),$$
$$(x_1 A_{12} + x_1 A_{21}) + (2x_2 A_{22}) + \dots + (x_n A_{n1} + x_n A_{1n}),$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$(x_1 A_{1n} + x_1 A_{n1}) + (x_2 A_{2n} + x_2 A_{3n}) + (2x_n A_{nn})]$$

[

We can notice a trend of $\sum_{k=1}^{n} x_k (A_{ik} + A_{ki})$ in i^{th} element of $\nabla_x x^T A x$

$$\Rightarrow \boxed{\nabla_x x^T A x = (A + A^T)\, x}$$

4. $\nabla_A x^T A x$

$$\begin{aligned} x^T A x = {} & x_1 (x_1 A_{11} + x_2 A_{21} + \dots + x_n A_{n1}) + \\ & x_2 (x_1 A_{12} + x_2 A_{22} + \dots + x_n A_{n2}) + \\ & \vdots \\ & x_n (x_1 A_{1n} + x_2 A_{2n} + \dots + x_n A_{nn}) \end{aligned}$$

$$\nabla_A x^T A x = \begin{bmatrix} x_1 x_1 & x_2 x_1 & \dots & x_n x_1 \\ x_1 x_2 & x_2 x_2 & \dots & x_n x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1 x_n & x_2 x_n & \dots & x_n x_n \end{bmatrix}$$

$$\boxed{\nabla_A x^T A x = x^T . x}$$

5. $\nabla_x^2 x^T A x$

$$\nabla_x^2 x^T A x = \begin{bmatrix} 2A_{11} & A_{21} + A_{12} & \dots & A_{n1} + A_{1n} \\ A_{12} + A_{21} & 2A_{22} & \dots & A_{n2} + A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} + A_{n1} & A_{2n} + A_{n2} & \dots & 2A_{nn} \end{bmatrix}$$

$$\boxed{\nabla_x x^T A x = A + A^T}$$

M T W T F S S
Page No.:
Date:
YOUVA

4.

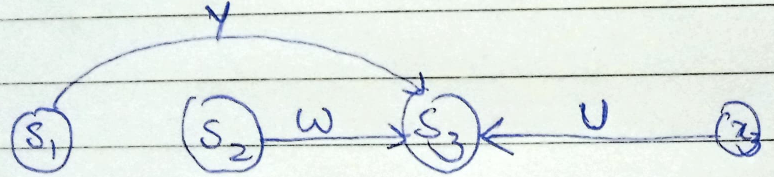
c) find $\nabla_T f$ where $T \in \mathbb{R}^{n \times n \times n}$

$$\nabla_T f = \left[\begin{bmatrix} \frac{\partial f}{\partial T_{111}} & \frac{\partial f}{\partial T_{112}} & \cdots & \frac{\partial f}{\partial T_{11n}} \\ \frac{\partial f}{\partial T_{121}} & \vdots & \ddots & \\ \vdots & \vdots & & \\ \frac{\partial f}{\partial T_{1n1}} & & & \frac{\partial f}{\partial T_{1nn}} \end{bmatrix}, \begin{bmatrix} \frac{\partial f}{\partial T_{211}} & \frac{\partial f}{\partial_{212}} & \cdots & \frac{\partial f}{\partial T_{21n}} \\ \frac{\partial f}{\partial T_{221}} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial f}{\partial T_{2n1}} & \cdots & & \frac{\partial f}{\partial T_{2nn}} \end{bmatrix}, \cdots \begin{bmatrix} \frac{\partial f}{\partial T_{n11}} & \frac{\partial f}{\partial_{n12}} & \cdots & \frac{\partial f}{\partial T_{n1n}} \\ \frac{\partial f}{\partial T_{n21}} & \ddots & & \\ \vdots & & \ddots & \\ \frac{\partial f}{\partial T_{nn1}} & \cdots & & \frac{\partial f}{\partial T_{nnn}} \end{bmatrix} \right]$$

let T_i be a matrix such that $\begin{bmatrix} T_{i11} & T_{i12} & \cdots & T_{i1n} \\ T_{i21} & T_{i22} & \cdots & T_{i2n} \\ \vdots & \vdots & \ddots & \vdots \\ T_{in1} & T_{in2} & \cdots & T_{inn} \end{bmatrix}$

$$\boxed{\nabla_T f = \left[\nabla_{T_1} f, \nabla_{T_2} f, \ldots \nabla_{T_n} f \right]}$$

c)

b) 

c) $\dfrac{\partial S_3}{\partial w} = \left.\dfrac{\partial \sigma(x)}{\partial w}\right|_{x = (wS_2 + YS_1 + Ux_3 + b)}$

$\cancel{\sigma(x)} = \dfrac{d\sigma(x)}{\partial x} \times \dfrac{dx}{dw}$ $\quad$ $\boxed{\sigma(x) = \dfrac{1}{1+e^{-x}} \text{ ; previously defined}}$

s_0 & s_{-1} are base cases

$$s_1 = \sigma(ws_0 + Ys_{-1} + Ux_1 + b)$$
$$s_2 = \sigma(ws_1 + Ys_0 + Ux_2 + b)$$

$$\frac{\partial s_3}{\partial w} = \left.\frac{\partial \sigma(x)}{\partial w}\right|_{x = ws_2 + Ys_1 + Ux_3 + b}$$

$$= \left.\sigma(x)(1-\sigma(x))\right|_{x = ws_2 + Ys_1 + Ux_3 + b} \times \underbrace{\frac{\partial(ws_2 + Ys_1 + Ux_3 + b)}{\partial w}}_{\parallel}$$

$$\left(\underbrace{w\frac{\partial s_2}{\partial w} + s_2}\right) + \underbrace{Y\frac{\partial s_1}{\partial w}}$$

$$s_2 + w\left(\left.\sigma(x)(1-\sigma(x))\right|_{x = ws_1 + Ys_0 + Ux_2 + b} \cdot \frac{\partial s_1}{\partial w}\right) \qquad \left.Y\sigma(x)(1-x)s_0\right|_{x = ws_0 + Ys_{-1} + Ux_1 + b}$$

$$= \left.\sigma(x)(1-\sigma(x))\right|_{x=\sigma^{-1}(s_3)} \left[\left.Y\sigma(x)(1-\sigma(x))s_0\right|_{x=\sigma^{-1}(s_1)} + s_2 + w\left(\left.\sigma(x)(1-\sigma(x))\right|_{x=\sigma^{-1}(s_2)} \cdot \left.\sigma(x)(1-\sigma(x))\right|_{x=\sigma^{-1}(s_1)}\right)\right.$$

$$\boxed{\frac{\partial s_3}{\partial w} = s_3(1-s_3)\left[Ys_1(1-s_1)s_0 + s_2 + w\left[s_2(1-s_2) \cdot (s_1(1-s_1)s_0)\right]\right]}$$

IIIly
$$\frac{\partial s_3}{\partial Y} = s_3(1-s_3) \cdot \frac{\partial(ws_2 + Ys_1 + Ux_3 + b)}{\partial Y}$$

$$= s_3(1-s_3) \cdot \left[w\frac{\partial s_2}{\partial Y} + s_1 + Y\frac{\partial s_1}{\partial Y}\right]$$

$$\frac{\partial s_1}{\partial Y} = s_1(1-s_1) \cdot s_{-1}$$

$$\frac{\partial s_2}{\partial Y} = s_2(1-s_2) \cdot \left[w\frac{\partial s_1}{\partial Y} + s_0\right] = s_2(1-s_2) \cdot \left[w\left[s_1(1-s_1) \cdot s_{-1}\right] + s_0\right]$$

b)

$$\Rightarrow \boxed{\frac{\partial s_3}{\partial Y} = s_3(1-s_3)\left[W\left(s_2(1-s_2)\cdot\left(W(s_1(1-s_1)s_{-1}) + s_0\right)\right) + s_1 + Y\left(s_1(1-s_1)s_{-1}\right)\right]}$$

c)

$$\frac{\partial s_3}{\partial U} = s_3(1-s_3)\,\frac{\partial\left[Ws_2 + Ys_1 + Ux_3 + b\right]}{\partial U}$$

$$\parallel$$

$$W\frac{\partial s_2}{\partial U} + Y\frac{\partial s_1}{\partial U} + x_3$$

$$\frac{\partial s_1}{\partial U} = x_1 \quad ; \quad \frac{\partial s_2}{\partial U} = Wx_1 + x_2$$

d)

$$\Rightarrow \boxed{\frac{\partial s_3}{\partial U} = s_3(1-s_3)\left[W\left(Wx_1 + x_2\right) + Yx_1 + x_3\right]}$$

5. a) gradient of a function at any point gives the vector ~~pointing to~~ that points to fastest increase from that point.

$\Rightarrow$ negative of gradient will give the slope and direction of maximum decrease

hence, if we were to minimize a function / find minima of a function from an arbitrary start point, taking negative of a gradient as a feedback for the next iteration will converge to a local minima (also depends on ~~para~~ other parameters)

e)

$$\Rightarrow x_{k+1} = x_k - \alpha\nabla f(x_k) \;; \quad k \in \mathbb{N},\ x_0 \text{ is arbitrarily chosen}$$

f)

$$\alpha \in \mathbb{R}^+, \alpha \neq 0$$

α is a parameter chosen according to the function

b) The above mentioned method applies to multidimensional functions too.

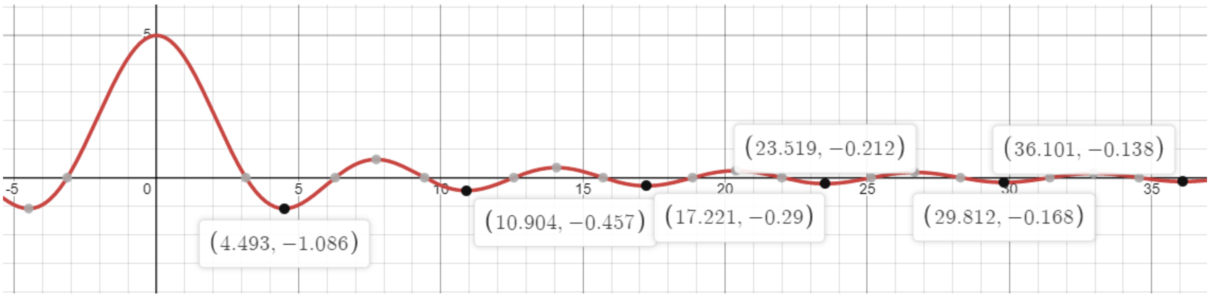
c) No, ~~[illegible]~~ global minima cannot be found out always, ~~because~~ even after tuning ~~[illegible]~~ α to ~~[illegible]~~ all possible values, minima can ~~[illegible]~~ be missed by this algorithm.

~~[illegible]~~ Here is a graphical explanation ; ~~[illegible]~~

In this example, we will take function f(x)= sinc(x) =sin(x)/x;

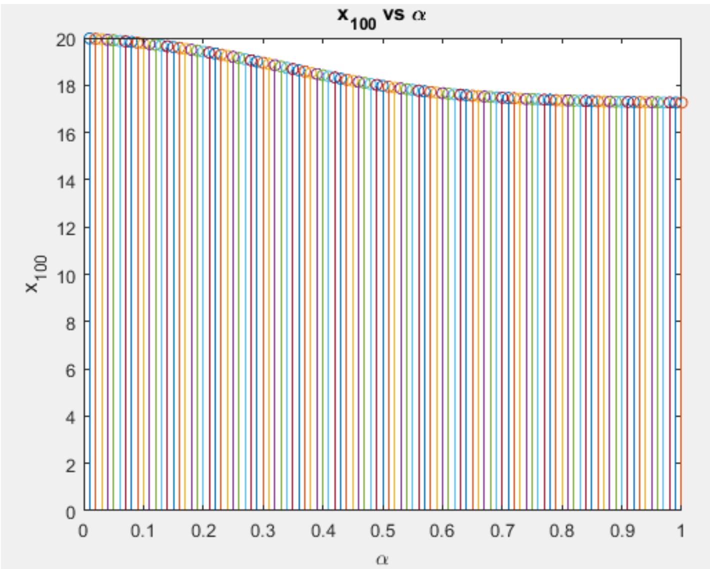
Given below are first few minima of the function, as we can see clearly at 4.493 we get a global minima (also at -4.493, but from here on we will only consider positive values, however, negative counterpart is also a valid answer).

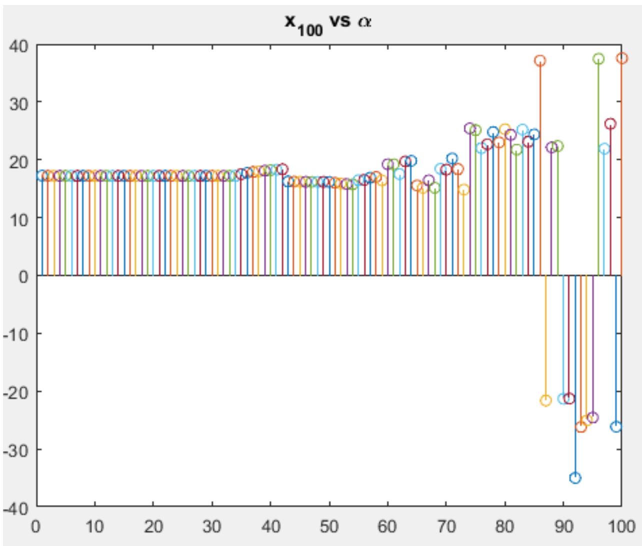
10.904, 17.221, 23.519, 29.812, 36.101, are some local minima.



Given a certain initial value (let us assume it to be 20 for this example), our goal is to show that the global minima cannot be reached for any value of alpha.

For every value of alpha, 100 iterations are performed to find minima (local or global). And alpha is being varied from 0.01 to 1 in first case and 1 to 100 in second case.





From this we can observe that the x_{100} converges to its local minima of 17.221 only for 0.5<alpha<40. Alpha less than 0.5, requires more than 100 iterations to converge, and for alpha greater than ~40, x_{100} doesn't converge to any minima (this has been verified by changing the number of iterations, the plot of which diverges exponentially and leads to no meaningful data).

We were never able to get a convergent value of global minima for any combination of iterations and alpha.

d) No this does not always work for convex function,

Ex: $f(x) = x^2$

$\nabla f = 2x$

Say $x_0 = 10$

$x_1 = x_0 - 2x_0 = -10$

$\Rightarrow x_2 = x_1 - 2x_1 = 10$

and so on, $x_{even\_no} = 10$ & $x_{odd\_no} = -10$

we never truly reach 0 even after multiple iterations

e) Say we want error to be under ε,

i.e $|x_k - x_{min}| < \varepsilon$

from which k can be calculated

f) we can manipulate x_k by changing α accordingly.

M T W T F S S
Page No.:
Date:
YOUVA

6. a) $\frac{\partial f}{\partial x} = 0$:- gives a locus / curve ~~where~~ f(x,y)= of minima of f(x,y) for every possible value of y

~~$\frac{\partial f}{\partial y} = 0$~~ ; & ||ly $\frac{\partial f}{\partial y} = 0$ gives locus of minima for every possibl value of x

locus of both curves intersects at minima.

c) gives ~~b~~ vector perpendicular to tangent at that point

d) gives vector perpendicular to tangent at that point

e) we want the magnitude of both ∇f & ∇g to be the same, as the intersection ~~curve~~ point's gradients would point in the same direction, where we ~~want~~ can use a simple multiplier

$$\nabla f = \lambda \nabla g$$

~~$\nabla g \leq b y$~~ gives the $g(x, y) = c$ also gives the next constraint needed to solve for x, y & λ

f) $f = x^a y^b z^c$ over $g(x, y, z) = x + y + z = 1$

$$\nabla f = \left(a x^{a-1} y^b z^c,\ x^a b y^{b-1} z^c,\ x^a y^b c z^{c-1}\right) = \lambda \nabla g = \lambda(1, 1, 1)$$

$\Rightarrow a = \lambda x,\ b = \lambda y,\ c = \lambda z$; $\frac{x}{a} = \frac{y}{b} = \frac{z}{c} = \frac{1}{\lambda}$

$\Rightarrow \frac{a}{\lambda} + \frac{b}{\lambda} + \frac{c}{\lambda} = 1 \Rightarrow \lambda = (a+b+c) \Rightarrow \max(x, y, z) = \left(\frac{a}{a+b+c}, \frac{b}{a+b+c}, \frac{c}{a+b+c}\right)$

7. a) $k = 10^9$

k_1, k_2, k_3 are unknown but $k_1 + k_2 + k_3 = k$

$\hat{k}_1, \hat{k}_2, \hat{k}_3$ are estimates for 1000 balloons

Error in estimate of Red ballon $= \left| \frac{k_1}{k} - \frac{\hat{k}_1}{1000} \right|$

|||ly for blue $= \left| \frac{k_2}{k} - \frac{\hat{k}_2}{1000} \right|$

green $= \left| \frac{k_3}{k} - \frac{\hat{k}_3}{1000} \right|$

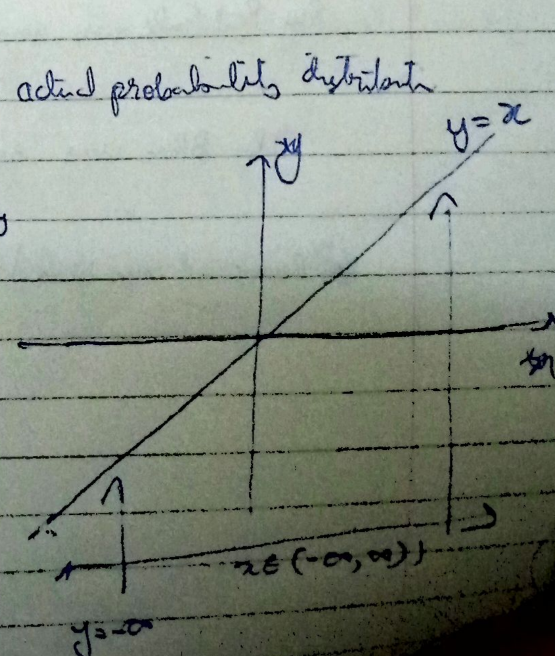
b) p is prefered over q distribution, as, we dont want to scale our estimate exponentially,

Ex: Say $k_1 = 5$, $k_2 = 3$, $k_3 = 2$, $k = 10$
& $\hat{k}_1 = 5$, $\hat{k}_2 = 3$, $\hat{k}_3 = 2$

using this we get $p = [0.5, 0.3, 0.2]$
$q = [0.8438, 0.1142, 0.042]$

q is far from ideal and p is the actual probability distribution

8 $E_x(F(x)) = \int_{-\infty}^{\infty} x F(x) dx = \int_{-\infty}^{\infty} x \int_{y=-\infty}^{y=x} P(y) dy$



area of the space $= \frac{1}{2}$

[graph is cut in half]

9a) According to intuition, Alice has stronger evidence that the ~~probability~~ of drawing blue ball was drawn. because they didn't draw a blue ball even once.

b) $P\left(\text{Removed was blue} \mid \text{Alice observation}\right)$

Say Red ball was removed then $P_{\text{Alice case}} = \left(\frac{2}{5}\right)^6$

111ly Blue was removed then $P_{\text{Alice}} = \left(\frac{3}{5}\right)^6$

$$P(\text{Removed was blue}) = \frac{P\left(\substack{\text{case after} \\ \text{Blue was removed}}\right)}{P(\text{case after B was removed}) + P(\text{case after R was removed})}$$

$$= \frac{(3/5)^6}{(2/5)^6 + (3/5)^6} = 0.912938$$

M T W T F S S

Page No.:

Date:

YOUVA

c) Bobs observation

Say red was removed then the prob of given case = $^{600}C_{303}\left(\frac{2}{5}\right)^{303}\left(\frac{3}{5}\right)^{297}$

blue was removed then, $^{600}C_{303}\left(\frac{3}{5}\right)^{303}\times\left(\frac{2}{5}\right)^{297}$

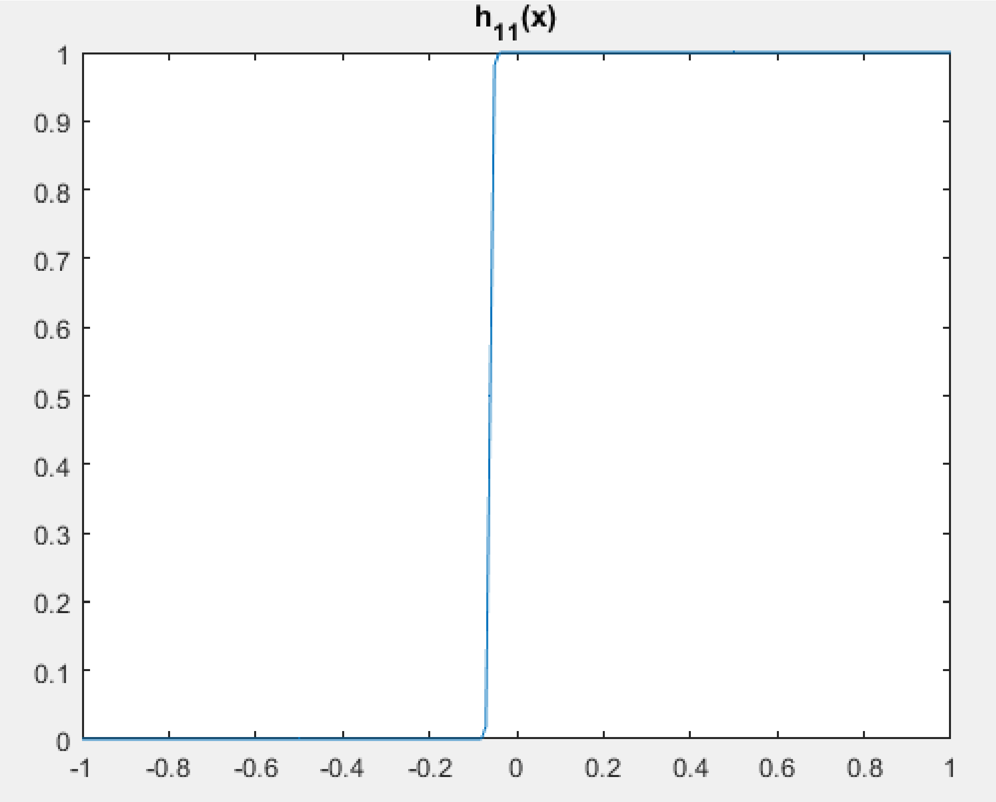
P that blue was removed $= \dfrac{^{600}C_{303}\left(\frac{3}{5}\right)^{303}\times\left(\frac{2}{5}\right)^{297}}{^{600}C_{303}\left(\frac{3}{5}\right)^{303}\times\left(\frac{2}{5}\right)^{297} + {^{600}C_{303}}\left(\frac{2}{5}\right)^{303}\left(\frac{3}{5}\right)^{297}}$

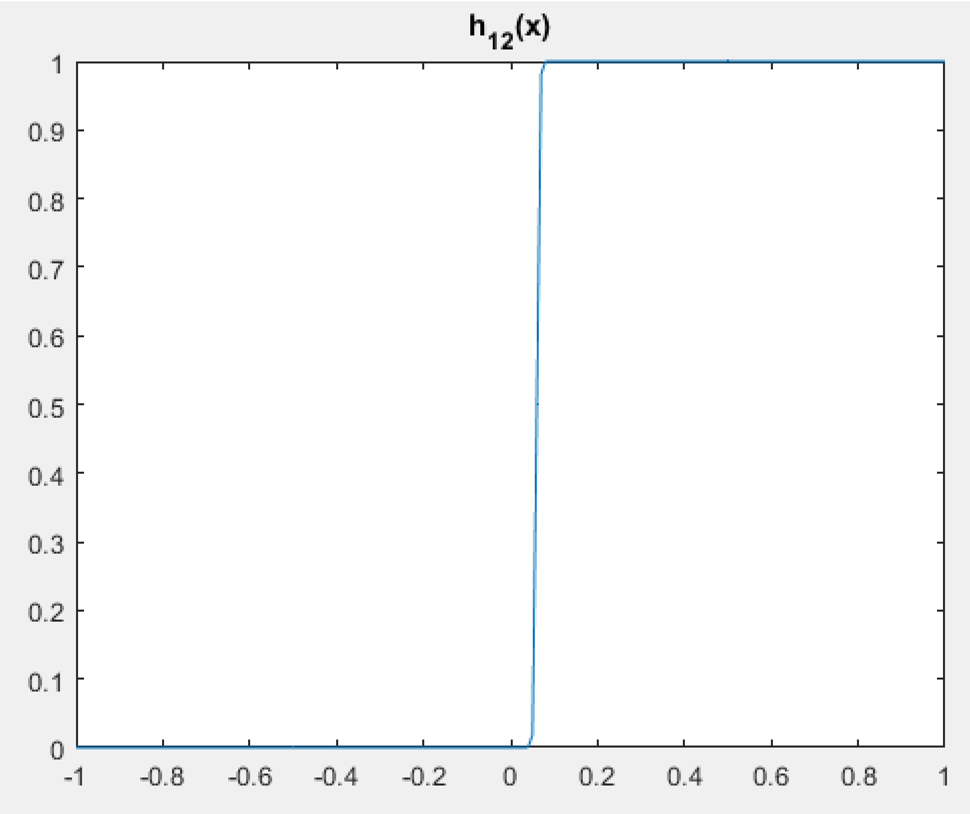
$= 0.9192938$

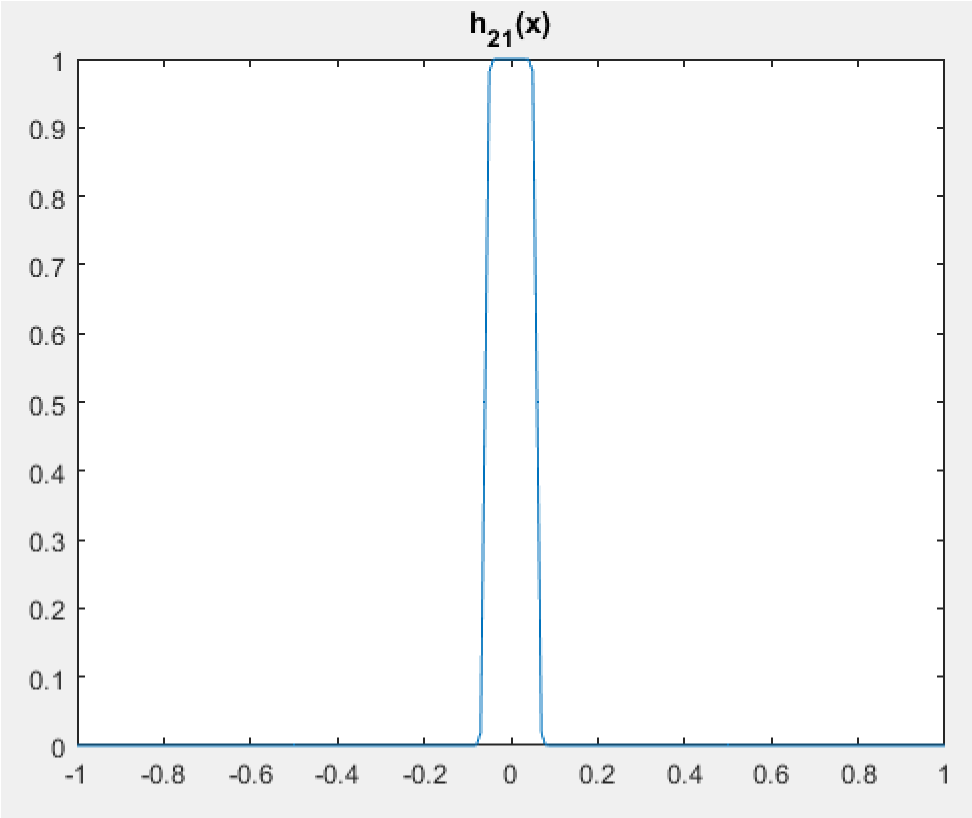
d) both have credible evidence for blue ball being removed.

Q10)

a)

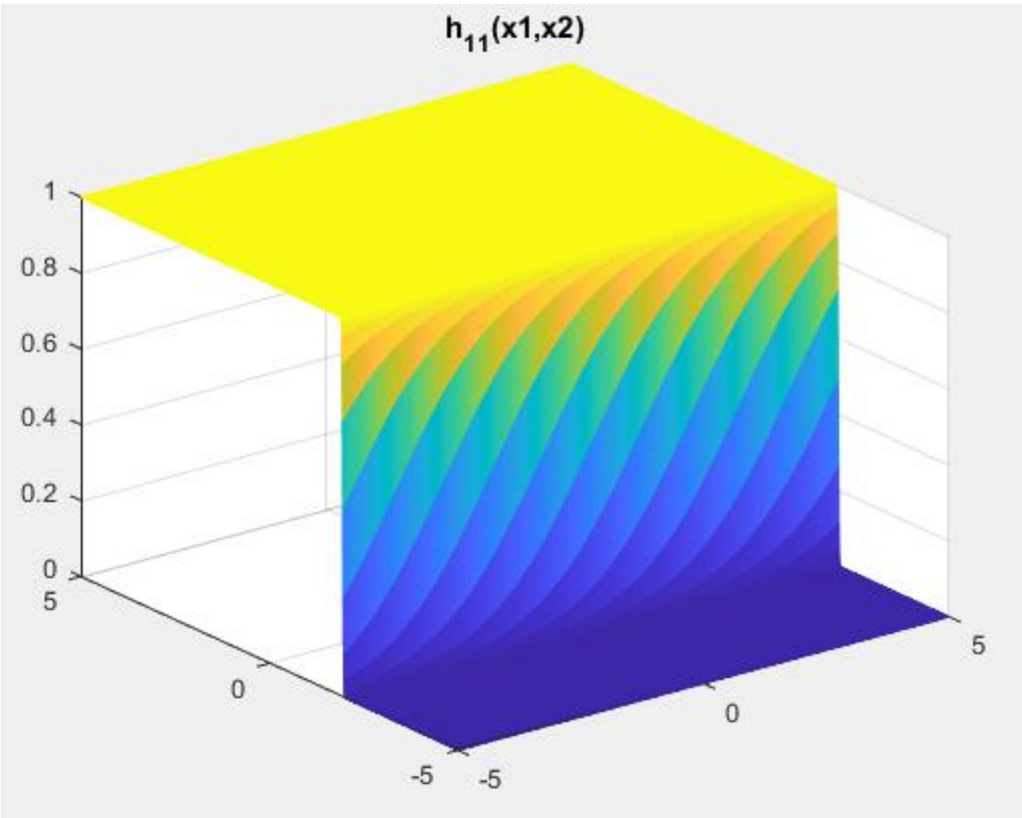

$h_{11}(x)$
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1


$h_{12}(x)$
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

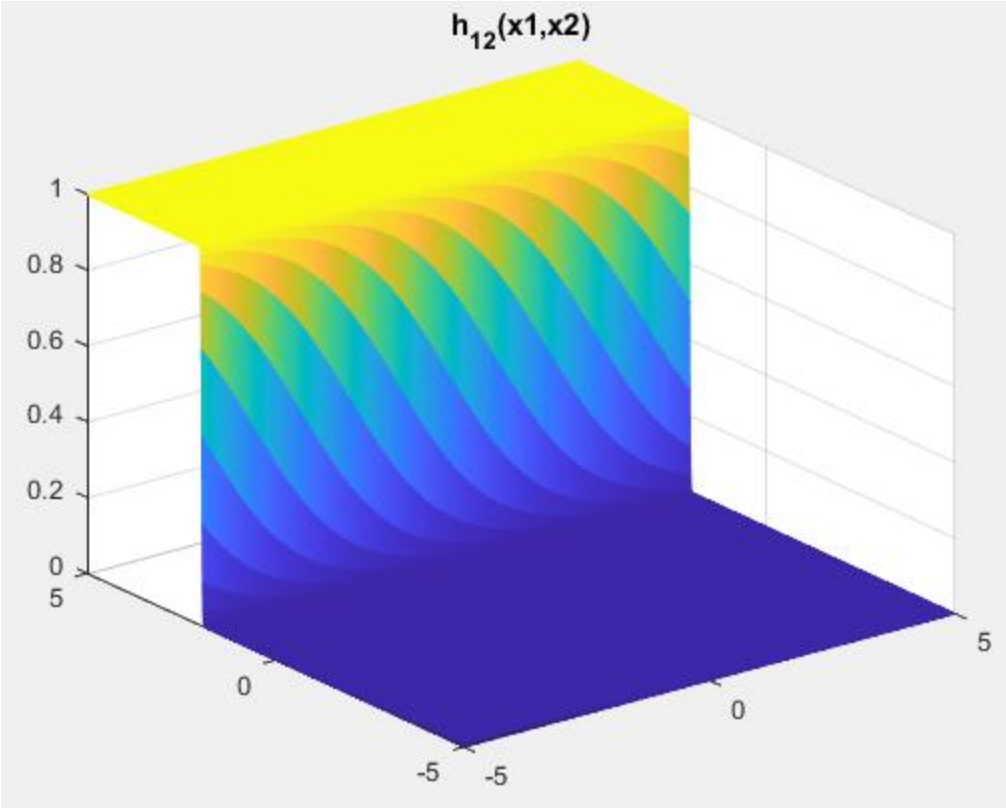


$h_{21}(x)$
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

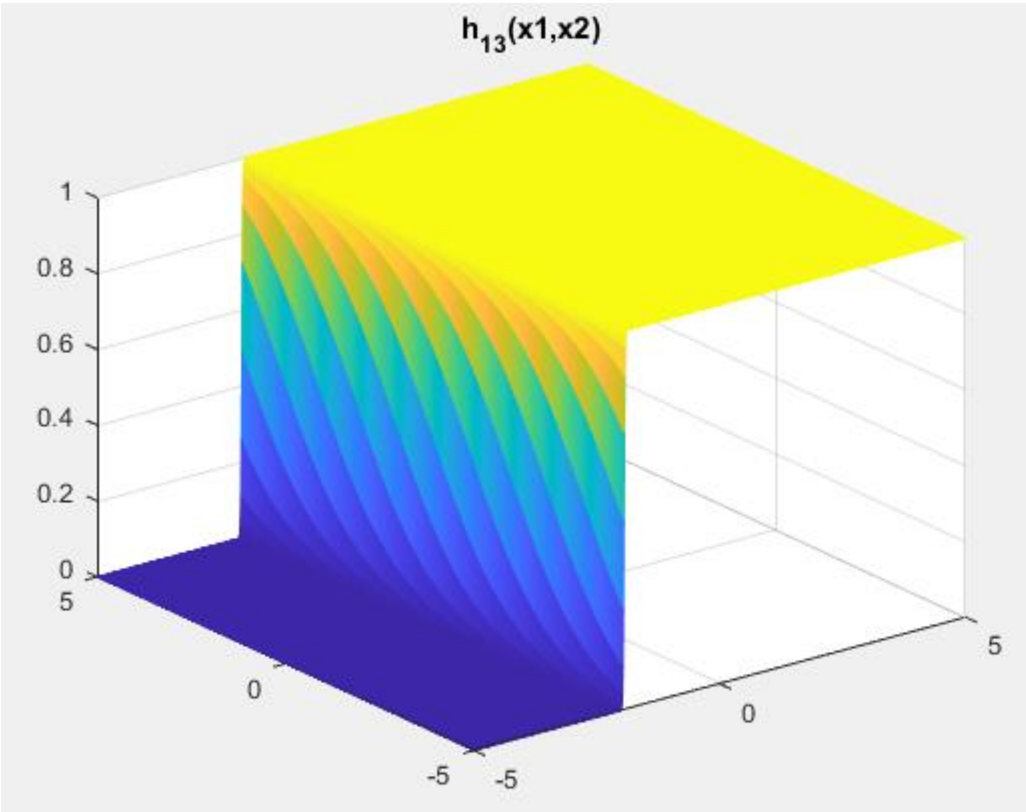
b)



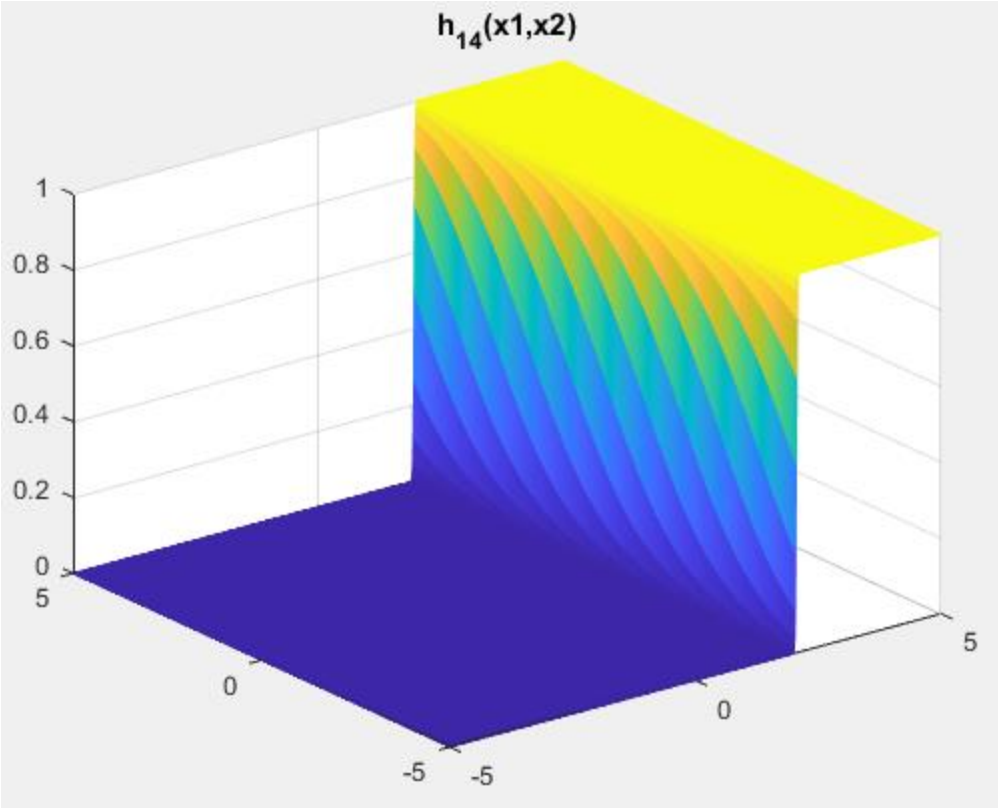
$h_{11}(x1,x2)$
1
0.8
0.6
0.4
0.2
0
5
0
-5
-5
0
5

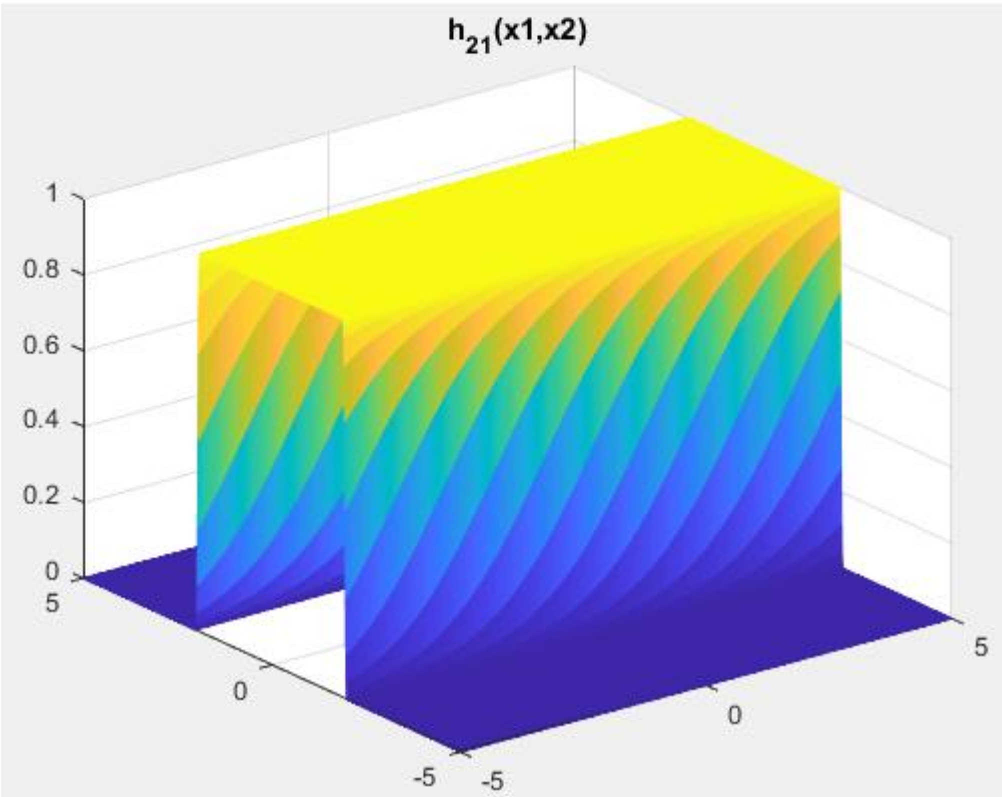


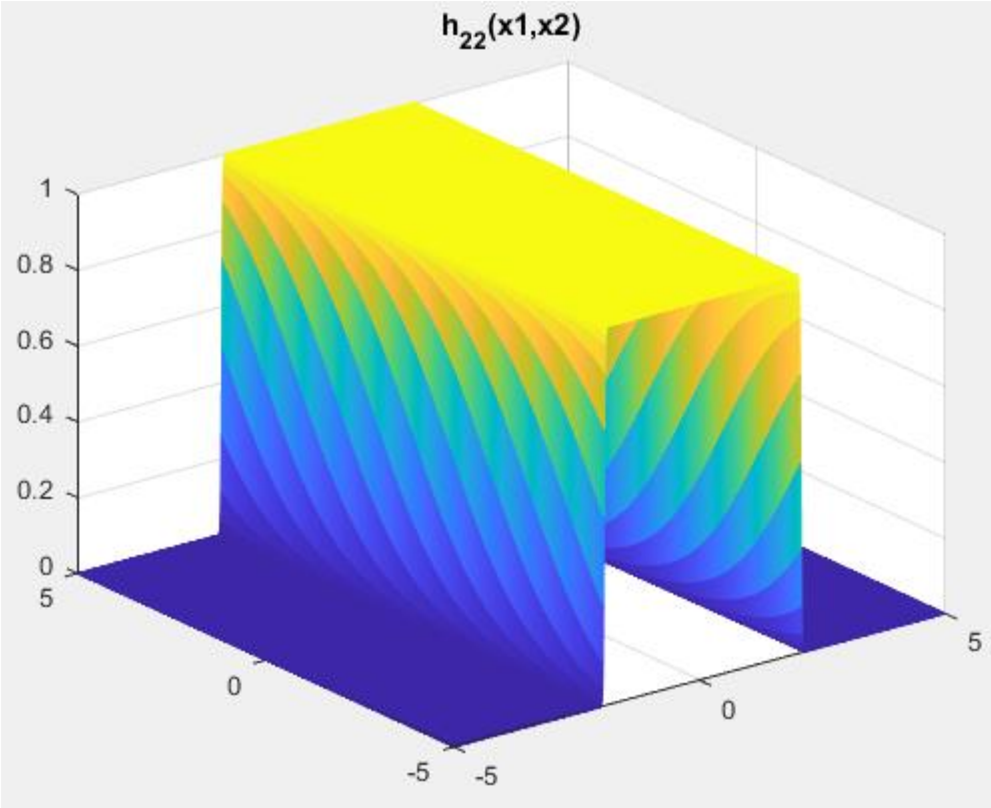
$h_{12}(x1,x2)$
1
0.8
0.6
0.4
0.2
0
5
0
-5
-5
0
5

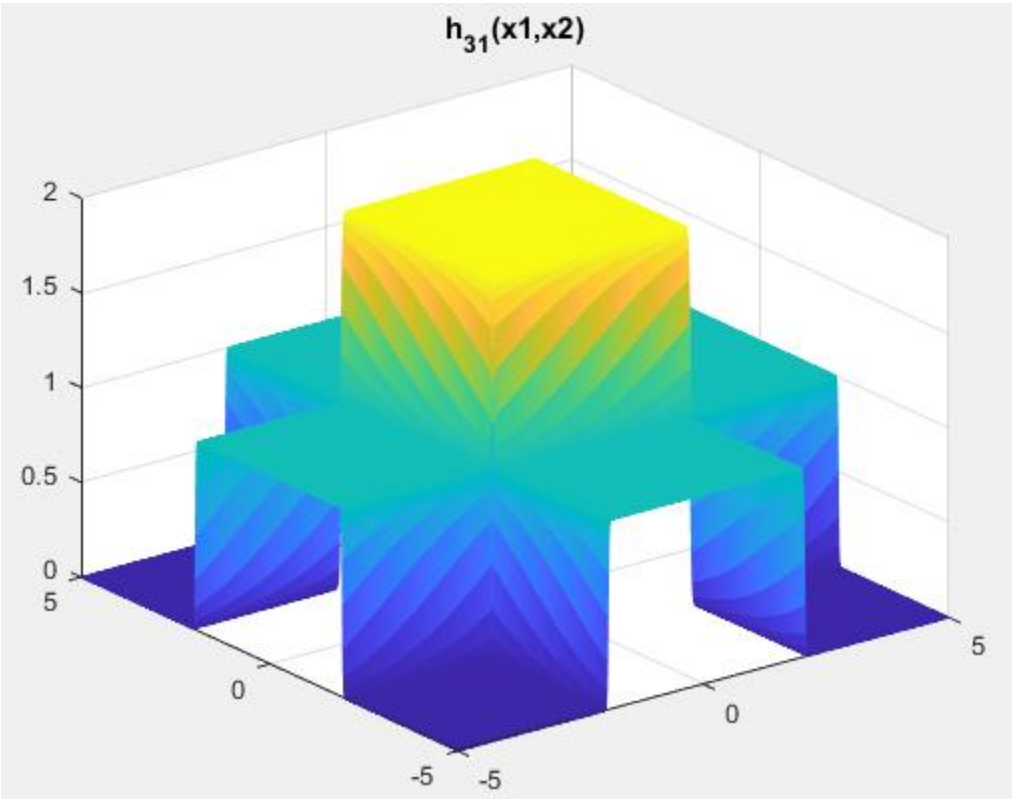


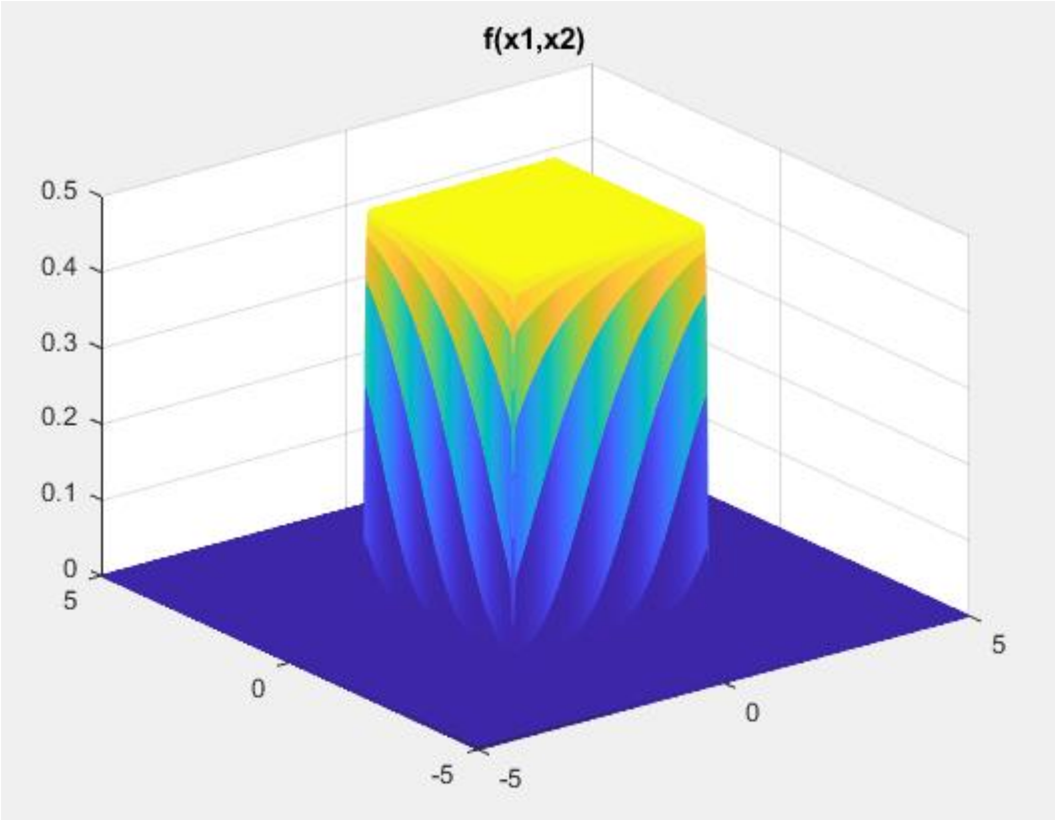
$h_{13}(x1,x2)$
1
0.8
0.6
0.4
0.2
0
5
0
-5
-5
0
5


h_14(x1,x2)
1
0.8
0.6
0.4
0.2
0
5
0
-5
-5
0
5


h_21(x1,x2)
1
0.8
0.6
0.4
0.2
0
5
0
-5
-5
0
5


h_{22}(x1,x2)
1
0.8
0.6
0.4
0.2
0
5
0
-5
-5
0
5


h_{31}(x1,x2)
2
1.5
1
0.5
0
5
0
-5
-5
0
5



f(x1,x2)
0.5
0.4
0.3
0.2
0.1
0
5
0
-5
-5
0
5

Page No.
Date

Tutorial 2

1.a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 6 & 4 & 2 \\ 1 & 2 & 8 \\ 1 & 5 & 7 \end{bmatrix}$

$A \longrightarrow A'$

$$|A'| = \begin{vmatrix} 6 & 4 & 2 \\ 1 & 2 & 8 \\ 1 & 5 & 7 \end{vmatrix} = 6.(14-40) - 4(7-8) + 2(5-2)$$
$$= -146$$

100 volume in A , volume in $A' = |A'| \times 100 = 14600$ unit^3

b) $A \rightarrow A'$

$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$ $\quad |A'| = 1.(45-48) - 4(18-24) + 7(12-15)$
$= 0$

$\Rightarrow$ volume in $A' = 0$ unit^3

c) The second transformation is invalid, as the basis vectors aren't linearly independent, let alone orthogonal. They are related as follows

$$2y' - x' = z' \text{ i.e, } 2\begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} - \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$$

2. a) x - component of $\begin{bmatrix}2\\-1\\6\end{bmatrix}$ in transformed space $= \begin{bmatrix}2\\-1\\6\end{bmatrix} \cdot \begin{bmatrix}4\\2\\1\end{bmatrix}$

$= 8 - 2 + 6 = 12$

IIIly

y - component $= \begin{bmatrix}2\\-1\\6\end{bmatrix}\begin{bmatrix}-3\\-1\\5\end{bmatrix} = -6 + 1 + 30 = 25$

& z - component $= \begin{bmatrix}2\\-1\\6\end{bmatrix}\begin{bmatrix}0\\2\\7\end{bmatrix} = -2 + 42 = 40$

$[2, -1, 6]^T \longrightarrow (12, 25, 40)^T$

b)

~~tait~~ standardized expression,

$||x|| = \sqrt{4^2 + 2^2 + 1^2} = \sqrt{21}$

$||y|| = \sqrt{3^2 + 1^2 + 5^2} = \sqrt{35}$

$||z|| = \sqrt{2^2 + 7^2} = \sqrt{51}$

$[2, -1, 6]^T = \left[\frac{12}{\sqrt{21}}, \frac{25}{\sqrt{35}}, \frac{40}{\sqrt{51}}\right]^T$

5)

b) $x \to x'$

$$x' = \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} = 4 - 2 + 6 = 8$$

$y \to y'$

$$y' = \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix} = -4 - 1 + 12 = 7$$

$z \to z'$

$$z' = \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix} = 2 + 2 + 12 = 16$$

$$\|x\| = \|y\| = \|z\| = \sqrt{9} = 3$$

$\Rightarrow$ standardized expression $= \begin{bmatrix} 8/3 \\ 7/3 \\ 16/3 \end{bmatrix}$

5. $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$

given, $\sum_{i}^{n} a_{ij} = c \quad \forall j \in \{1, 2 \ldots n\}$

say, we choose an eigen vector $V = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

then, we get

$$A^T V = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} c \\ c \\ \vdots \\ c \end{bmatrix} = cV \quad ; \quad A^T V = cV$$

$\Rightarrow$ c is eigen value for A^T with v as its eigen vector.

4. $C = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ given $a_{11} + a_{22} = 0$

we are asked to compute C^n, lets start with C^2, C^3 & so on.

$$C^2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11}^2 + a_{12}a_{21} & a_{11}a_{12} + a_{12}a_{22} \\ a_{21}a_{11} + a_{22}a_{21} & a_{21}a_{12} + a_{22}^2 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}^2 + a_{12}a_{21} & 0 \\ 0 & a_{21}a_{12} + a_{22}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a_{11}^2 + a_{12}a_{21} \\ a_{21}a_{12} + a_{22}^2 \end{bmatrix}$$

$$C^3 = \begin{bmatrix} a_{11}^2 + a_{12}a_{21} & 0 \\ 0 & a_{21}a_{12} + a_{22}^2 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \cancel{\left[a_{11}(a_{11}^2 + a_{12}a_{21}) + a_{12} f a_{11}^2 \right.}$$

$$C^3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a_{11}^2 + a_{12}a_{21} \\ a_{22}^2 + a_{21}a_{12} \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}(a_{11}^2 + a_{12}a_{21}) & a_{12}(a_{11}^2 + a_{12}a_{21}) \\ a_{21}(a_{22}^2 + a_{12}a_{21}) & a_{22}(a_{22}^2 + a_{21}a_{12}) \end{bmatrix}$$

(or)

$$C^n = I \cdot \begin{bmatrix} a_{11}^2 + a_{12}a_{21} \\ a_{22}^2 + a_{21}a_{12} \end{bmatrix} \cdot C^{n-2} \quad ; \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Say $A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$

then $A^2 = \begin{bmatrix} a^2 & 0 \\ 0 & b^2 \end{bmatrix}$

$\parallel$ly $A^n = \begin{bmatrix} a^n & 0 \\ 0 & b^n \end{bmatrix}$

if n is even:

$$\boxed{C^n = I \cdot \begin{bmatrix} (a_{11}^2 + a_{12}a_{21})^{n/2} \\ (a_{22}^2 + a_{21}a_{12})^{n/2} \end{bmatrix}}$$

if n is odd

$$C^n = I \cdot \begin{bmatrix} (a_{11}^2 + a_{12}a_{21})^{\frac{(n-1)}{2}} \\ (a_{22}^2 + a_{21}a_{12})^{\frac{(n-1)}{2}} \end{bmatrix} C$$

5. (b)

$$e^C e^D = e^{C+D} \quad \text{iff} \quad CD = DC$$

Say $C = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ & $D = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$

$$CD = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} \neq DC = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}$$

$$e^C = \begin{bmatrix} 4.8942 & 5.4755 \\ 5.4755 & 10.3247 \end{bmatrix}$$

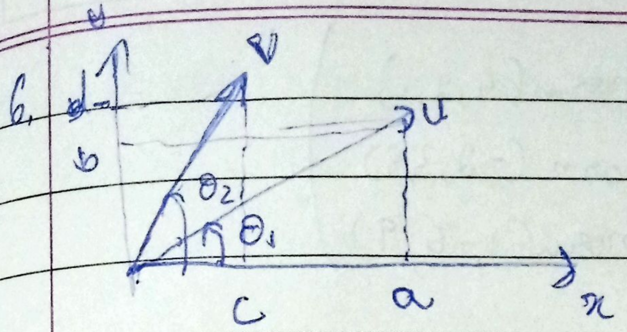
$$e^D = \begin{bmatrix} e & 0 \\ e & e \end{bmatrix}$$

$$e^{C+D} = \begin{bmatrix} 20.0116 & 17.2933 \\ 34.5866 & 37.3049 \end{bmatrix}$$

$$e^C \cdot e^D = \begin{bmatrix} 13.1815 & 0 \\ 14.8839 & 28.0655 \end{bmatrix}$$

as we can see $e^C \cdot e^D \neq e^{C+D}$

6.



$$\|\vec{u}\| = \sqrt{a^2+b^2}$$
$$\|\vec{v}\| = \sqrt{c^2+d^2}$$

$$\cos\theta_1 = \frac{a}{\sqrt{a^2+b^2}} \qquad \cos\theta_2 = \frac{c}{\sqrt{c^2+d^2}}$$

$$\cos(\theta_2 - \theta_1) = \cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2$$
$$= \frac{ac + bd}{\sqrt{a^2+b^2}\sqrt{c^2+d^2}}$$

$$\langle u, v\rangle = \begin{bmatrix} a \\ b \end{bmatrix} \cdot \begin{bmatrix} c \\ d \end{bmatrix} = ac + bd$$

$$\Rightarrow \langle u, v\rangle = \|\vec{u}\|\,\|\vec{v}\| \cos\theta$$

7 $\{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n\}$ set of linearly dependent vectors,

if this set is dependent, then

$$a_1\vec{u} + a_2\vec{u}_2 + \dots a_n\vec{u}_n = 0$$

where $a_i \in R$ and $a_j \neq 0$ $\forall i \in \{1 \dots n\}$

say we find j vectors that are independent (we can assume ~~it is first~~ they are first j vectors without loss of generality).

$$a_1\vec{u}_1 + a_2\vec{u}_2 + \dots a_j\vec{u}_j = 0$$

$$\vec{u}_j = -\frac{a_1\vec{u}}{a_j} - \frac{a_2\vec{u}_2}{a_j} \dots -\frac{a_{j-1}\vec{u}_{j-1}}{a_j}$$

$\therefore$ $(u_1, v_2 \dots v_{j-1})$ spans the entire space and $u_j \in$ span of $(v_1 \dots v_{j-1})$

Let $u \in \text{span}\{v_1, u_2 \dots v_m\}$

$$v = b_1v_1 + b_2v_2 + \dots + b_mu_m$$

$$u_j = -\frac{a_1}{a_j}u_1 - \frac{a_2}{a_j}u_2 \dots -\frac{a_{j-1}u_{j-1}}{a_j}$$

$$V = b_1v_1 + b_2u_2 + \dots + b_{j-1}u_{j-1} + \left(-\frac{a_1u_1}{a_j} - \frac{a_2u_2}{a_j} \dots -\frac{a_{j-1}v_{j-1}}{a_j}\right) + b_{j+1}u_{j+1} + \dots + b_mu_m$$

$$V = \left(b_1 - \frac{a_1}{a_j}\right)v_1 + \left(b_2 - \frac{a_2}{a_j}\right)u_2 + \dots + \left(b_{j-1} - \frac{a_{j-1}}{a_j}\right)u_{j+1} + b_{j+1}u_{j+1} + b_{j+2}v_{j+2} + b_mv_m$$

$\Rightarrow v \in \text{span}\{v_1, v_2 \dots v_{j-1}, u_{j+1} \dots u_m\}$

span of $\{v_1 \dots v_m\}$ = span of $\{u_1, v_2 \dots v_{j-1}, v_{j+1} \dots U_m\}$

$\Rightarrow$ list is unchanged

9.a) A is linearly independent ~~then~~ only when $a_i \neq 0$

$$a_1 u_1 + a_2 u_2 + \dots + a_n u_n = 0$$ where $a_i \neq 0$ & $a_i \in \mathbb{R}$ $\forall i \in \{1 \dots n\}$

B is linearly independent only when $b_i \neq 0$

in $$b_1(u_1 - u_2) + b_2(u_2 - u_3) + \dots b_{n-1}(u_{n-1} - u_n) + b_n u_n = 0$$

given A & B are linearly independent vectors $\Rightarrow a_i = b_i = 0$ $\forall i \in \{1 \dots n\}$
~~regrouping~~ ~~but~~ this also implies the other way around, so, if we can prove $b_i = 0$, then B is independent

$$b_1 u_1 + (b_2 - b_1) u_2 + (b_3 - b_2) u_3 + \dots + (b_n - b_{n-1}) u_n = 0$$

$$b_1 = a_1 \Rightarrow b_1 = 0$$

$$\Rightarrow b_2 - b_1 = 0 \Rightarrow b_2 = 0$$

and so on,

$\Rightarrow$ B is linearly independent

b) A is spanning set, & $v \in \mathrm{span}(A)$

$$v = a_1 u_1 + a_2 u_2 + \dots a_n u_n \quad \text{for } a_i \in \mathbb{R}$$

$$v = b_1(u_1 - u_2) + b_2(u_2 - u_3) + \dots b_{n-1}(u_{n-1} - u_n) + b_n u_n$$

$$v = b_1 u_1 + (b_2 - b_1) u_2 + \dots + (b_n - b_{n-1}) u_n$$

WKT $v \in \mathrm{span}(A) \Rightarrow v \in B$

$\Rightarrow v$ $\Rightarrow$ B ~~also~~ spans vector space

10. $$\text{L1 norm}\left(\begin{bmatrix} 4 & -2 & 1 \\ 2 & 3 & -6 \\ 1 & 8 & 9 \end{bmatrix}\right) = \max\begin{pmatrix} \text{L1 norm}(4,2,1) \\ \text{L1 norm}(-2,3,8) \\ \text{L1 norm}(1,-6,9) \end{pmatrix}$$

$$= \max\begin{pmatrix} 7 \\ 13 \\ 16 \end{pmatrix} = 16$$

$$\text{L1 norm}\left(\begin{bmatrix} 4 & -2 & 1 \\ 8 & 3 & -6 \\ 9 & 2 & 1 \end{bmatrix}\right) = \max\begin{pmatrix} \text{L1 norm}(4,8,9) \\ \text{L1 norm}(8,3,-6) \\ \text{L1 norm}(9,2,1) \end{pmatrix}$$

$$= \max\begin{pmatrix} 17 \\ 17 \\ 12 \end{pmatrix} = 17$$

$$\text{Frobenius norm}\left(\begin{bmatrix} 4 & -2 & 1 \\ 2 & 3 & -6 \\ 1 & 8 & 9 \end{bmatrix}\right) = \sqrt{4^2+2^2+1^2+2^2+3^2+6^2+1^2+8^2+9^2}$$

$$= 14.6969$$

$$\text{Frobenius norm}\left(\begin{bmatrix} 4 & -2 & 1 \\ 8 & 3 & -6 \\ 9 & 2 & 1 \end{bmatrix}\right) = \sqrt{4^2+2^2+1^2+8^2+3^2+6^2+9^2+2^2+1^2}$$

$$= 14.6969$$

11a) Given: $\|A\| = \sup_x \{\|Ax\| : \|x\| = 1\}$ & by definition $\|x\| \geq 0 \ \forall x$

$\Rightarrow \|Ax\| \geq 0 \ \forall A \,\&\, x$

$\Rightarrow \|A\| \geq 0$

b) $\|A\| = \sup_x \{\|Ax\| : \|x\| = 1\}$; given

$\|\alpha x\| = |\alpha| \, \|x\|$

$\Rightarrow \|\alpha Ax\| = |\alpha| \, \|Ax\|$

$\|\alpha A\| = \sup_x \{\|\alpha Ax\| : \|x\| = 1\}$

$= \sup_x \{|\alpha| \, \|Ax\| : \|x\| = 1\}$

$= |\alpha| \cdot \|A\|$

c) $\|A+B\| = \sup_x \{\|(A+B)x\| : \|x\| = 1\}$

$\leq \sup_x \{\|Ax\| + \|Bx\| : \|x\| = 1\}$ $\quad\quad \|a+b\| \leq \|a\| + \|b\|$

$\leq \sup_x \{\|Ax\| : \|x\| = 1\} + \sup_x \{\|Bx\| : \|x\| = 1\}$

$= \|A\| + \|B\|$

d) given $\|A\| = 0 \rightarrow \sup_x \{\|Ax\| : \|x\| = 1\} = 0$

$\Rightarrow \|Ax\| = 0 \quad \forall \|x\| = 1$

$\Rightarrow Ax = 0 \quad \forall \|x\| = 1$

$\Rightarrow A = 0$

e) $\|A\| = \sup_x \{\|Ax\| : \|x\| = 1\}$

$\Rightarrow \|Ax\| \leq \|A\|$

$\|AB\| = \sup_x \{\|ABx\| : \|x\| = 1\} \leq \|A B\| \leq \|A\| \|B\|$

b) $\|Ax\|_2^2 = x^T A^T A x$

A^TA is diagonalizable given it is symmetric

$$\|Ax\|_2^2 = y^T D y$$
$$= \lambda_1^2 y_1^2 + \lambda_2^2 y_2^2$$

Since $\|x\|_2 = 1 \Rightarrow \|y\|_2 = 1 \Rightarrow \|Ax\|_2^2 \leq \sigma_{max}(A)$.

12. let x & y be eigenvectors of the symmetric matrix A, whose eigen values are λ_1 & λ_2

WKT. $\lambda_1 \langle x, y \rangle = \langle \lambda_1 x, y \rangle$
$= \langle A x, y \rangle$
$= \langle x, A^T y \rangle$ [if matrix is real $\langle Ax, y \rangle = \langle x, A^T y \rangle$]
$= \langle x, A y \rangle$ (symmetric matrix)
$= \langle x, \lambda_2 y \rangle$
$= \lambda_2 \langle x, y \rangle$

$(\lambda_1 - \lambda_2)\langle x, y \rangle = 0$

as $\lambda_1 \neq \lambda_2$, $\langle x, y \rangle = 0 \Rightarrow x \perp y$

13. $Ax = \lambda x$

$$\max_x \{x^T A x\} = \max_x \{x_1^T A x_1, x_2^T A x_2 \ldots x_n^T A x_n\}$$
$$= \max \{x_1^T A_1 x, x_2^T \lambda_2 x_2, \ldots x_n^T \lambda_n x_n\}$$

$$= \max \{\lambda_1 \|x_1\|^2, \lambda_2 \|x_2\|^2, \ldots \lambda_n \|x_n\|^2\} \text{ given } \|x_i\| = 1$$
$$= \max \{\lambda_1, \lambda_2 \ldots \lambda_n\}$$

$\Rightarrow$ Solution is given by largest eigen value

14 Since the matrix is full rank square matrix $A_{n \times n}$ of size n it has n eigen values

$$V = [x_1 \ldots x_n]$$
$$AV = A[x_1 \ldots x_n]$$
$$AV = [\lambda_1 x_1, \lambda_2 x_2 \ldots \lambda_n x_n]$$

$$= \begin{bmatrix} \lambda_1 x_{11} & \lambda_2 x_{12} & \cdots & \lambda_n x_{1n} \\ \lambda_1 x_{21} & \lambda_2 x_{22} & \cdots & \lambda_n x_{2n} \\ \vdots & \vdots & & \vdots \\ \lambda_1 x_{n1} & \lambda_2 x_{n2} & & \lambda_n x_{nn} \end{bmatrix}$$

$$AV = \begin{bmatrix} x_{11} & x_{12} \cdots & x_{1n} \\ x_{21} & x_{22} & x_{2n} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n2} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 & \cdot \\ 0 & \lambda_2 & & \\ \vdots & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

M	T	W	T	F	S	S

Page No.:

Date:

YOUVA

15 $x.y = \|x\|\|y\| \cos\theta$

$x' = Ax \quad , \quad y' = Ay$

$x'y' = \|x'\|\|y'\| \cos\phi = \|Ax\|\|Ay\| \cos\phi$

$= |A|^2 \|x\|\|y\| \cos\phi$

$\Rightarrow \|x\|\|y\| \cos\theta = \|x\|\|y\| \cos\phi \quad \Rightarrow \cos\theta = \cos\phi$

$\Rightarrow \theta = \phi$