MLF Week 2 : Calculus

Lecture 1 : Sets and Functions

Outline

- **Sets and Functions**
  - **Notations**
  - **Logic**
  - **Graphs and visualisations.**
- Univariate Calculus
  - Continuity and differentiability
  - Derivatives and Linear approximations
  - Applications/Advanced rules
- Multivariate Calculus
  - Lines and planes in high dimensional space.
  - Partial derivatives
  - Gradients
  - Linear approximations and Alternate gradient interpretations
  - Applications/Advanced rules

Sets

$\mathbb{R}$ - set of real numbers

$\mathbb{R}_+$ - set of positive reals including 0

$\mathbb{Z}$ - set of Integers

$\mathbb{Z}_+$ - " " +ve Inte

$[a,b] = \{x \in \mathbb{R} : a \leq x \leq b\}$

$(a,b) : \{x \in \mathbb{R} : a < x < b\}$

$\mathbb{R}^d$: Set of d-dimensional vectors $= \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$

$$\begin{pmatrix} 1 \\ 2 \\ 3.3 \end{pmatrix} \in \mathbb{R}^3$$

$[a,b]^d : \{x \in \mathbb{R}^d : x_i \in [a,b] \quad i \in \{1,2,\ldots d\}\}$

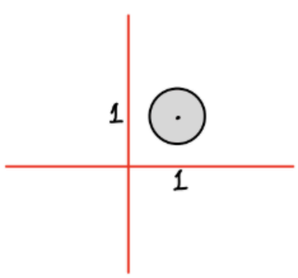
Matric Spaces

$\mathbb{R}^d$: $D(x,y) = \|x-y\| = \sqrt{(x_1-y_1)^2 + \cdots + (x_d-y_d)^2}$

$x \in \mathbb{R}^d$. $B(x,\epsilon) = \{y \in \mathbb{R}^d : D(x,y) < \epsilon\}$

$\bar{B}(x,\epsilon) = \{y \in \mathbb{R}^d : D(x,y) \leq \epsilon\}$

$d = 2$



$x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$B\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, 0.5\right)$

Sets and Logic

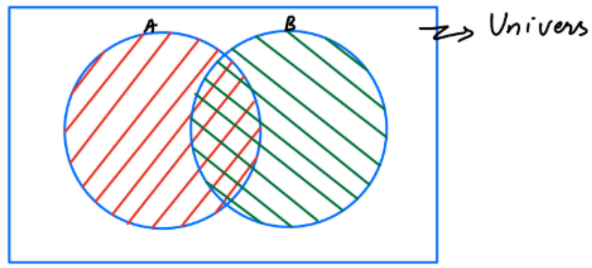
U = Universe

$A \cup B$, $A \cap B$, A^c, B^c

$A^c = U \setminus A$

$(A \cup B)^c = A^c \cap B^c$

$(A \cap B)^c = A^c \cup B^c$



$U = [0, 10]$

$A = [2, 5]$, $B = [4, 7]$; $A \cup B = [2, 7]$, $A \cap B = [4, 5]$

$(A \cup B)^c = [0, 2) \cup (7, 10] = A^c \cap B^c$

as $A^c = [0, 2) \cup (5, 10]$, $B^c = [0, 4) \cup (7, 10]$

$\forall$	For all	$\Longrightarrow$	Implies	$A \Longrightarrow B$
$\exists$	There exists	$\Longleftrightarrow$	Equivalent	$A \Longleftrightarrow B$

Sequences

$x_1, x_2, \dots$

where $x_i \in \mathbb{R}^d$

$\lim_{i \to \infty} x_i = x^*$ $\qquad x_i \to x^*$

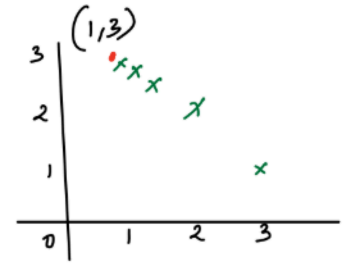
$\Updownarrow$

$\forall \epsilon > 0, \exists N$ s.t

$x_n \in B(x^*, \epsilon) \ \forall n \geq N$

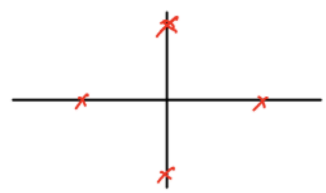
Example sequence 1

$x_n = \left(1 + \frac{4}{2^n}, 3 - \frac{4}{2^n}\right)$



Example sequence 2

$x_n = \left(\cos \frac{\pi}{2} n, \sin \frac{\pi}{2} n\right)$



Example:

(i) $x_i \in \mathbb{R}$. ; $x_n = 1 + n$

(ii) $x_i \in \mathbb{R}^2$; $x_n = \left(\frac{1}{2^n} \cos\left(\frac{\pi}{2} n\right), \frac{1}{2^n} \sin\left(\frac{\pi}{2} n\right)\right)$

iii) $x_i \in \mathbb{R}^2$; $x_n = \left(\frac{1}{2^n} \cos\left(\frac{\pi}{2} n\right), \sin\left(\frac{\pi}{2} n\right)\right)$

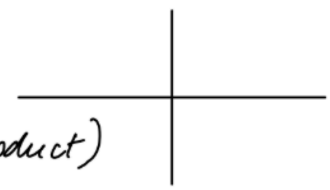
Vector Spaces

If V is a vector space

$u \in V, \; v \in V \qquad \alpha, \beta \in \mathbb{R}$

$\alpha u + \beta v \in V$

- $\mathbb{R}^d$ is a vector space.
- $x \cdot y = x^T y = \sum_{i=1}^{d} x_i y_i$ (dot product)



- $\|x\|^2 = x \cdot x = x^T x = \sum_{i=1}^{d} x_i^2$
- x & y are perpendicular/orthogonal

$x \cdot y = x^T y = \sum x_i y_i = 0$

Functions and Graphs

$$f : A \rightarrow B$$

A → Domain, B → Co-domain

1-dimensional function

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

d-dimensional functions

$$f : \mathbb{R}^d \rightarrow \mathbb{R}$$

$$G_f \subseteq \mathbb{R}^{d+1}$$

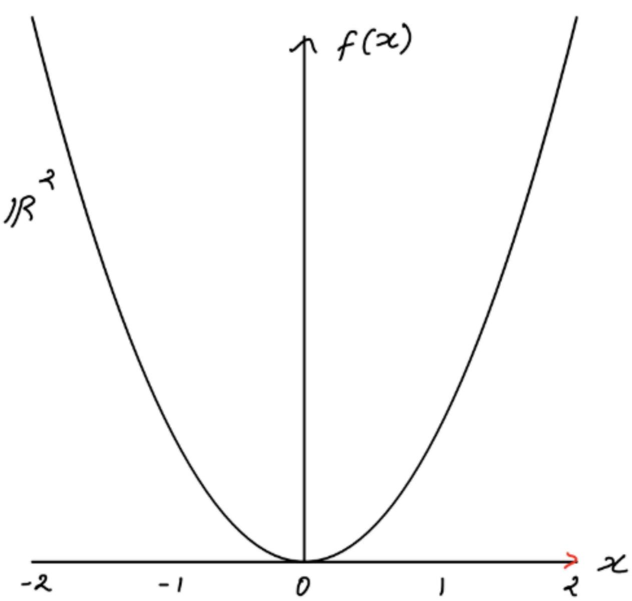
$$G_f = \{ (x, f(x)) : x \in \mathbb{R}^d \}$$

Plots of 1-dimensional Functions

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

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

$$G_f = \{ (x, x^2) : x \in \mathbb{R} \} \subseteq \mathbb{R}^2$$



Contour Plots of 2-dimensional Functions

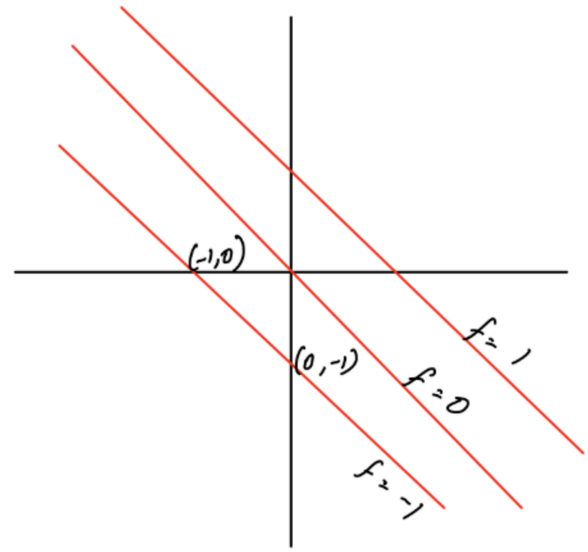
$$f : \mathbb{R}^2 \to \mathbb{R}$$

$$f(x) = x_1 + x_2$$

Values = $\{-1, 0, 1\}$

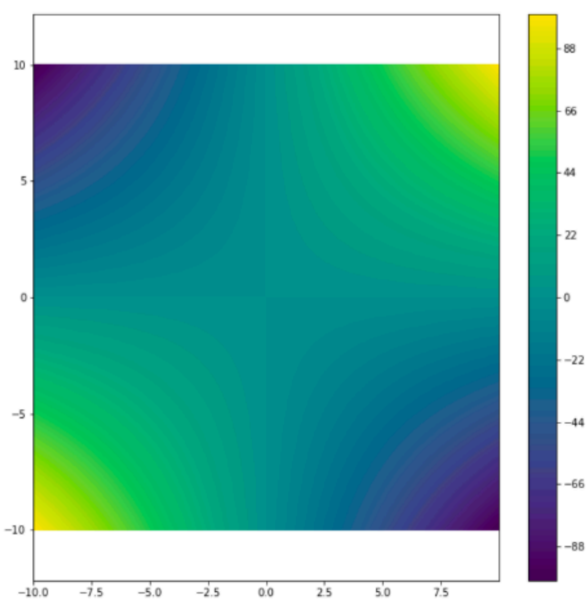
$$f(x) = -1 \Rightarrow x_1 = -x_2 - 1$$

$$f(x) = 0 \Rightarrow x_1 = -x_2$$

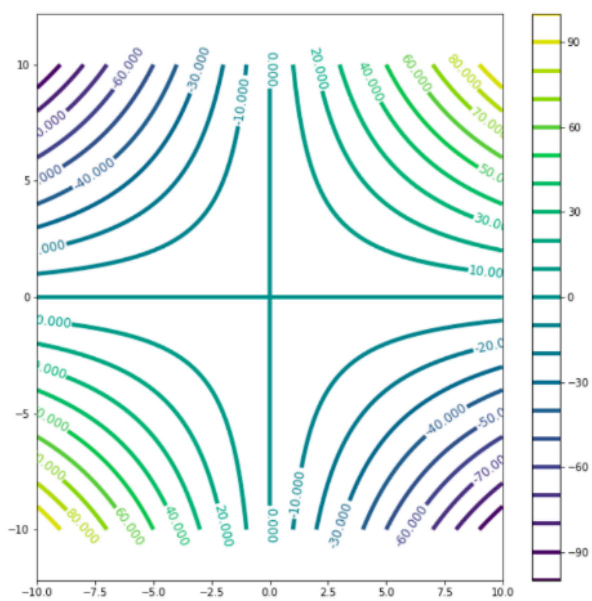


2.

$$f(x) = x_1 x_2$$



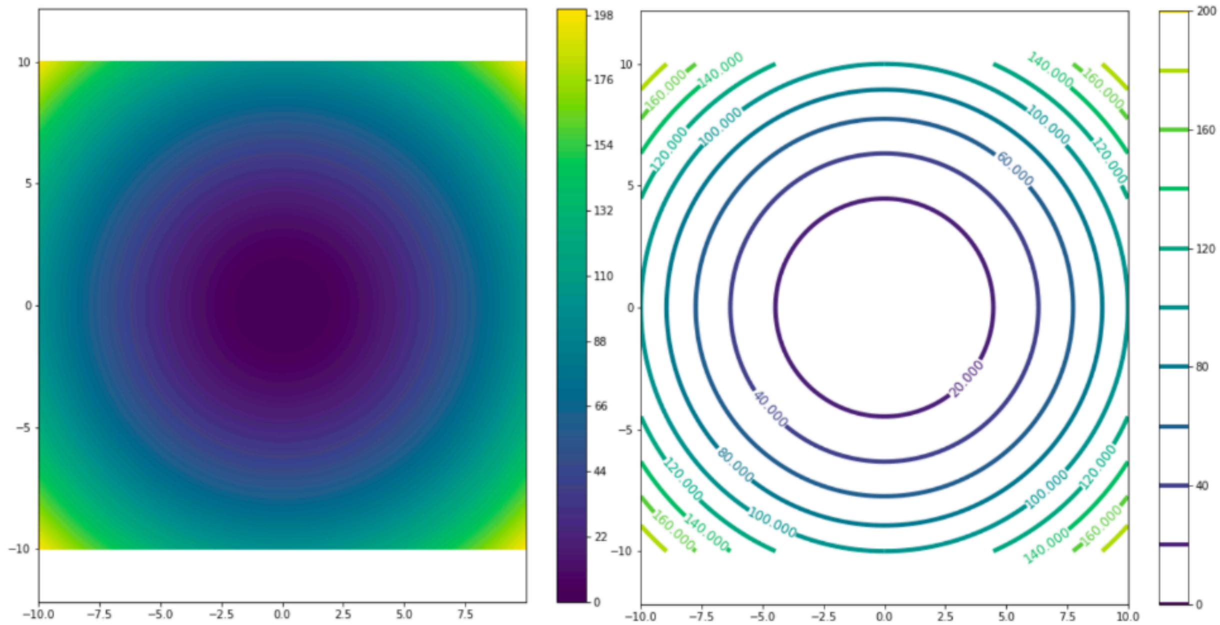
Heat map



Contour map

3.

$$f(x_1, x_2) = x_1^2 + x_2^2$$



Heat map

contour map.

QN : 2,6

Lecture 2 : Univariate Calculus: Continuity and Differentiability

Continuity of Functions

$$f : \mathbb{R} \to \mathbb{R}$$

is continuous at $x^* \in \mathbb{R}$ if for all sequences $x_1, x_2 \cdots$ converging to x^* we have that $f(x_i)$ converges to $f(x^*)$

$$\lim_{i \to \infty} x_i = x^* \implies \lim_{i \to \infty} f(x_i) = f(x^*)$$

$$\lim_{x \to x^*} f(x) = f(x^*)$$

E.g 1 : $f(x) = x^2$, $x^* = 2$

x_i : $3, 2.5, 2.25, \cdots \to 2$

$f(x_i)$: $9, 6.25, 2.25^2, \cdots \to 4$

$e.g\ 2 \quad : \quad f(x) = sign(x) = \begin{cases} -1 & \text{if } x < 0 \\ +1 & \text{if } x = 0 \\ +1 & \text{if } x > 0 \end{cases}$

$x^* = 0$

$x_i : 1, \frac{1}{2}, \frac{1}{4}, \cdots \quad \rightarrow 0$

$f(x_i) : 1, 1, 1, \cdots \quad \rightarrow 1$

$x_i : -1, -\frac{1}{2}, -\frac{1}{4}, \cdots \quad \rightarrow 0$

$f(x_i) : -1, -1, -1, \quad \rightarrow -1$

$E.g.3 : \quad f(x) = \begin{cases} 2x+1 & \text{if } x > 1 \\ 3 & \text{if } x \leq 1 \end{cases}$

$E.g.3 \quad f(x) = \frac{1}{x}$,

$x_i : 1, \frac{1}{2}, \frac{1}{4}, \cdots$

$f(x_i) : 1, 2, 4, 8$

$E.g.3 \quad f(x) = \cos(\frac{1}{x})$

$x_i : 1, \frac{1}{2}, \frac{1}{4}, \cdots$

$f(x_i) : \cos(1), \cos(2), \cos(4),$

Differentiability of Functions

A function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $x^* \in \mathbb{R}$

if $\underbrace{\lim_{x \to x^*} \frac{f(x) - f(x^*)}{x - x^*}}_{\text{is } f'(x^*)}$ exists.

f is NOT continuous at x^* $\Rightarrow$ f is NOT differentiable at x^*.

E.g. 1 $f(x) = |x|$

x_i : $1, \frac{1}{2}, \frac{1}{4}, \ldots \to 0$

$\frac{f(x_i) - f(0)}{x_i}$: $1, 1, 1 \to 1$

x_i : $-1, -\frac{1}{2}, -\frac{1}{4} \ldots$

: $-1, -1, -1$

E.g : 2

$$f(x) = \begin{cases} 4x+2 & \text{if } x \geq 2 \\ 2x+8 & \text{if } x < 2 \end{cases}$$

E.g : 3

$$f(x) = \begin{cases} 4x+2 & \text{if } x \geq 2 \\ 2x+6 & \text{if } x < 2 \end{cases}$$

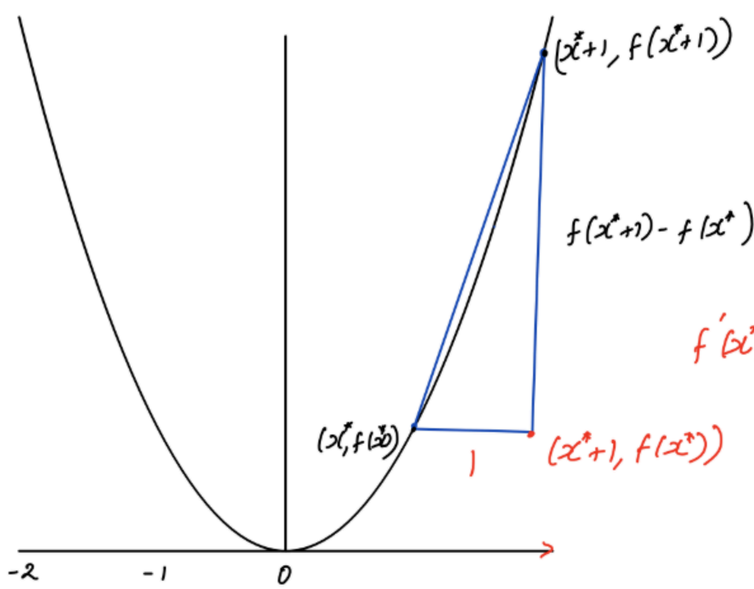
$$\lim_{x \to 2_+} \frac{f(x) - f(2)}{x-2} = 4$$

$$\lim_{x \to 2_-} \frac{f(x) - f(2)}{x-2} = 2$$

E.g. 4:

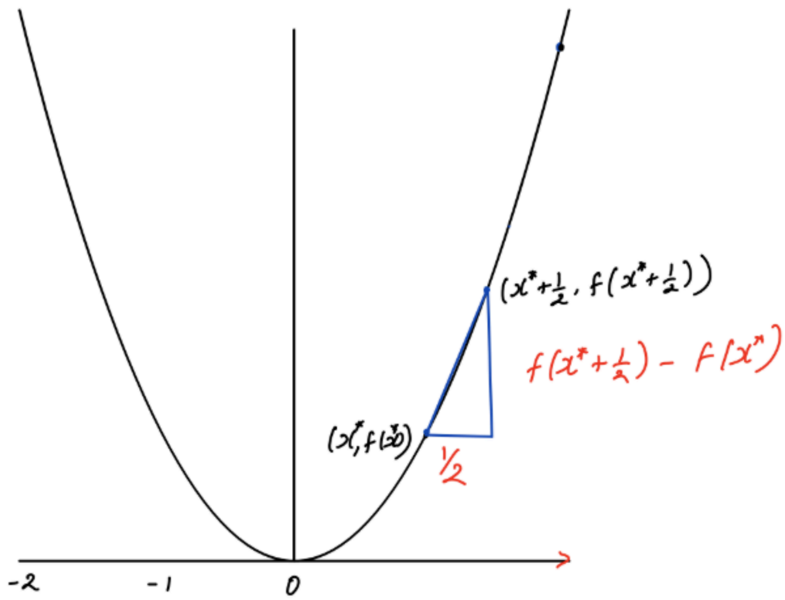
$$f(x) = \begin{cases} 4x+2 & \text{if } x \geq 2 \\ x^2+6 & \text{if } x < 2 \end{cases}$$

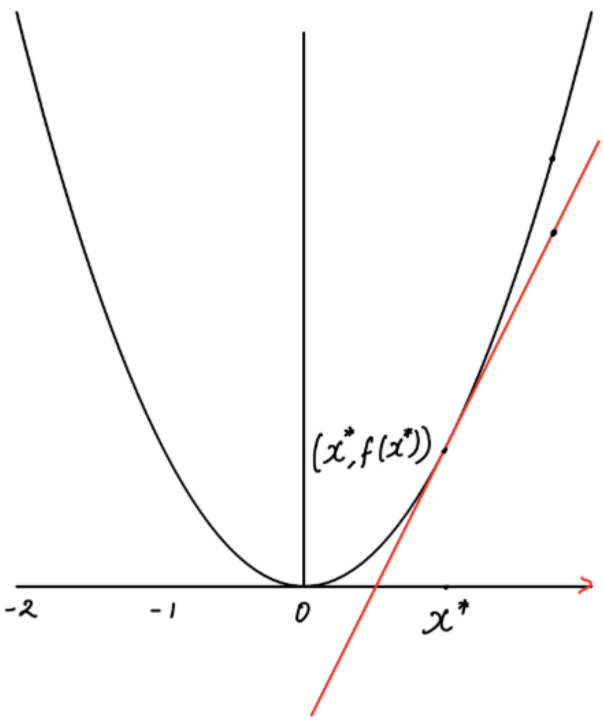
$$\lim_{x \to 2^+} = \lim_{x \to 2^-} = \frac{f(x) - f(2)}{x-2} = 4$$



$$f'(x^*) = \lim_{x \to x^*} \frac{f(x) - f(x^*)}{x - x^*}$$

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


$(x^*+\frac{1}{2}, f(x^*+\frac{1}{2}))$
$f(x^*+\frac{1}{2}) - f(x^*)$
$(x^*, f(x^*))$
$1/2$
-2
-1
0


$(x^*, f(x^*))$
-2
-1
0
x^*

Lecture 3 : Univariate Calculus: Derivatives and Linear Approximations

Derivatives and Linear Approximation

Let $f : \mathbb{R} \to \mathbb{R}$ be a diff function

$$f'(x^*) = \lim_{x \to x^*} \frac{f(x) - f(x^*)}{x - x^*}$$

$$f'(x^*) \approx \frac{f(x) - f(x^*)}{x - x^*} \quad (\text{around } x = x^*)$$

$$f(x) \approx \underbrace{f(x^*) + f'(x^*)(x - x^*)}_{L_{x^*}[f](x)} \quad (\text{around } x = x^*)$$

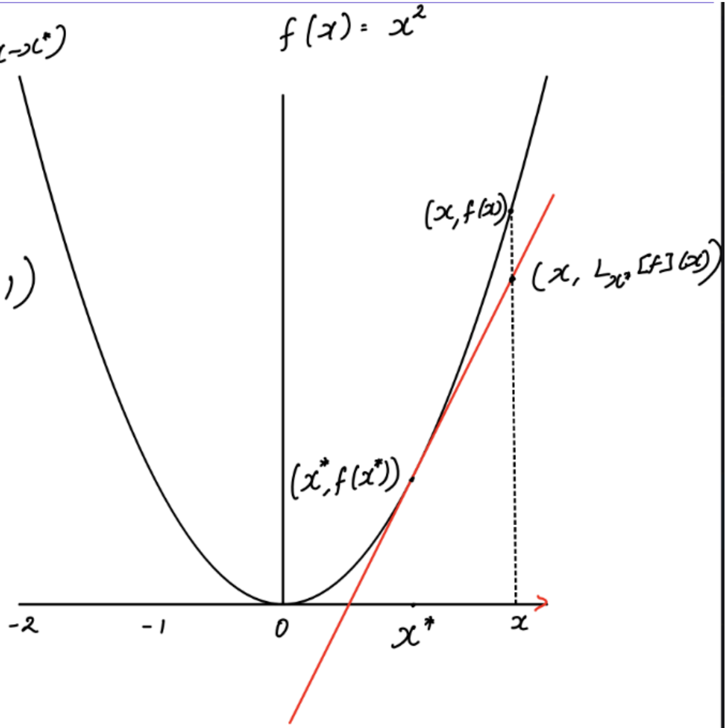
$$f(x) \approx L_{x^*}[f](x) \quad \text{around } x = x^*$$

$$L_{x^*}[f] = f(x^*) + f'(x^*)(x - x^*)$$

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

$$= 1 + 2x - 2$$

$$= 2x - 1 \quad (\text{around } x = 1)$$



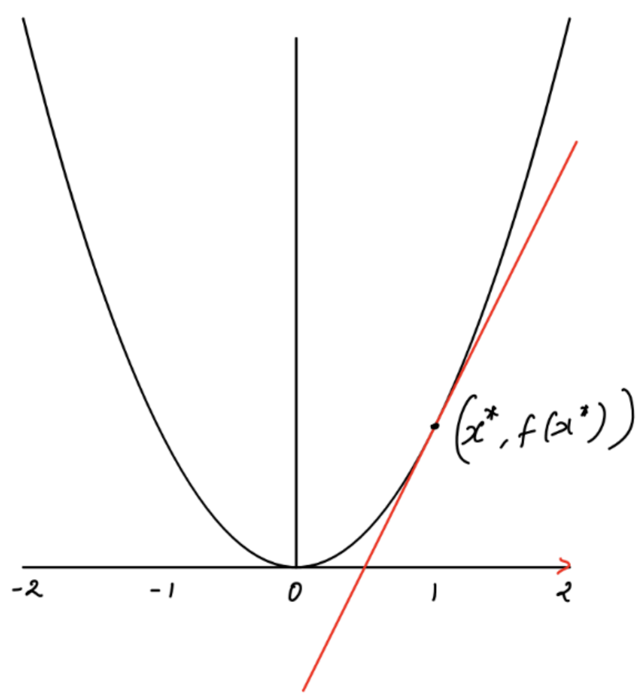
Linear Approximations and Tangent Lines

$G_{L_{x^*}[f]}$ is a

$\cap$

$\mathbb{R}^2$

tangent to the graph of f G_f at the point $x^*, f(x^*)$



Derivatives and Linear Approximation Ex :

i) Linear approximation of $f(x) = \sin(x)$ around $x^* = 0$

$$f(x) \approx f(x^*) + f'(x^*)(x - x^*)$$

$f'(x) = \cos(x)$

$f'(x^*) = 1$

$f(x^*) = 0$

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

$$= x \quad \text{around } x = 0$$

$\sin x \approx x$ if $x \approx 0$

(ii) $f(x) = e^x$ around $x^* = 0$

$$e^x \approx e^0 + (x - 0) \cdot 1$$

$$\approx 1 + x \quad \text{around } x = 0$$

---

iii) $\ln(1+x)$ around $x^* = 0$

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

$f'(x^*) = 1$

$f(x^*) = 0$

$$\ln(1+x) \approx \ln(1) + 1(x - 0)$$

$$\approx x \quad \text{around } x = 0$$

iv) $f(x) = (1+x)^r$ around $x^* = 0$

$f(x^*) = 1$

$f'(x) = r(1+x)^{r-1}$

$f'(x^*) = r$

$$(1+x)^r \approx 1 + r(x)$$

$$= 1 + rx \quad \text{around } x = 0.$$

(v) $(0.99)^7$

(a) 0.95 (b) 0.93

(c) 0.91 (d) 0.9

Lecture 4 : Univariate Calculus: applications and advanced rules

Higher Order Approximations

$$f(x) \approx f(x^*) + f'(x^*)(x - x^*) \quad \text{(Linear apx)}$$

$$f(x) \approx f(x^*) + f'(x^*)(x - x^*) + \frac{1}{2} f''(x^*)(x - x^*)^2 \quad \text{(quadratic)}$$

e.g: 1 $\quad f(x) = x^2$

$f'(x) = 2x$

$f''(x) = 2$

$$x^2 \approx (x^*)^2 + 2x^*(x - x^*) + \frac{1}{2} \cdot 2 \cdot (x - x^*)^2$$

$$= x^2$$

Approx e^x around $x^* = 0$

$$e^x \approx e^0 + e^0 (x - x^*) + e^0 \cdot \frac{1}{2} \cdot (x - x^*)$$

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

Ex: Which is closest to $(1.1)^7$

(a) 1.7 (b) 1.9 (c) 2.1 (d) 2.3

$f(x) = (1+x)^7$, $f'(x) = 7(1+x)^6$, $f''(x) = 42(1+x)^5$

$f'(0) = 7$ $f''(0) = 42$

$$f(0.1) \approx 1 + 7(0.1) + \frac{1}{2} \cdot 42 (0.01)$$

$$= 1.91$$

Product Rule

$$f(x) = g(x) \cdot h(x)$$

$f'(x) = ?$ $x^* = 0$

$$f(x) \approx (g(0) + x\, g'(0))(h(0) + x\, h'(0))$$

$$= g(0)h(0) + x\left[g'(0)h(0) + h'(0)g(0)\right] + x^2 g'(0) h'(0)$$

$$L_x[f] = f(0) + x\, f'(0)$$

$$f'(0) = g'(0)h(0) + h'(0)g(0)$$

Chain Rule

$$f(x) = g(h(x))$$

$$\approx g(h(0) + h'(0)x)$$

$$\approx g(h(0)) + g'(h(0))\left[h(0) + h'(0)x - h(0)\right]$$

$$= g(h(0)) + g'(h(0))\, h'(0)\, x$$

$$f(x) \approx f(0) + f'(0) \cdot x$$

$$f'(0) = g'(h(0))\, h'(0)$$

(i) $\dfrac{e^{3x}}{\sqrt{1+x}}$ Give LA around $x = 0$

$$\frac{e^{3x}}{\sqrt{1+x}} \approx (1+3x)\left(1 - \frac{x}{2}\right)$$

$$\approx 1 + \frac{5}{2}x \qquad (\text{around } x = 0)$$

(ii) Give Lin. Apx.

$e^{\sqrt{1+x}}$ around $x = 1$

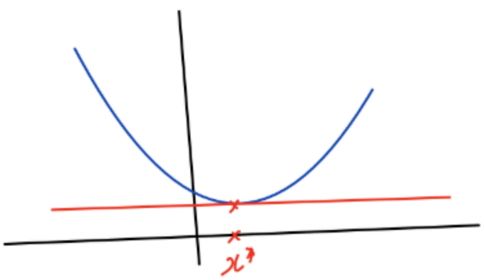
$$e^{\sqrt{1+x}} \approx e^{\sqrt{2}} + \frac{e^{\sqrt{2}}}{2\sqrt{2}}(x-1) \qquad (\text{around } x = 1)$$

Maxima, minima and saddle points

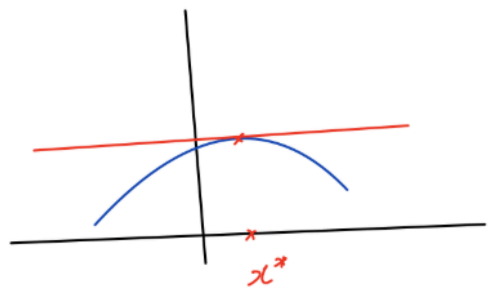
$$L_{x^*}[f] = f(x^*) + f'(x^*)(x - x^*)$$

$f'(x^*) = 0 \iff x^*$ is a critical point of f

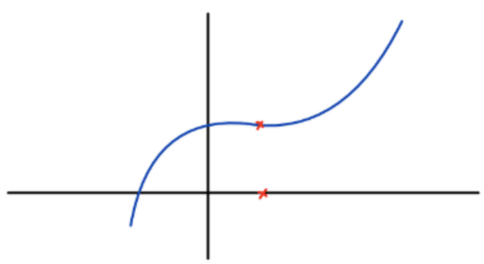
$$L_{x^*}[f] = 7$$



Minima



Maxima



saddle point.

QN 1,2

Lecture 5 : Multivariate Calculus: Lines and planes in higher dimensional space

Geometry of Lines

(i) A line in $\mathbb{R}^d \subseteq \mathbb{R}^d$

(ii) (a) A line through the point $u \in \mathbb{R}^d$ along the vector $v \in \mathbb{R}^d$

$$= \{ x \in \mathbb{R}^d : x = u + \alpha v \quad \text{for } \alpha \in \mathbb{R} \}$$

(b) Line through $u, u' \in \mathbb{R}^d$

$$= \{ x \in \mathbb{R}^d : x = u + \alpha (u' - u) \quad \text{for } \alpha \in \mathbb{R} \}$$

$$= \{ x \in \mathbb{R}^d : x = (1-\alpha) u + \alpha u' \quad \text{for } \alpha \in \mathbb{R} \}$$

Line through u along $u' - u$

Line through u' along $u - u'$

Geometry of (Hyper)planes

A $(d-1)$ dimensional hyperplane $\subseteq \mathbb{R}^d$

A hyperplane normal to the vector $w \in \mathbb{R}^d$ with value $b \in \mathbb{R}$

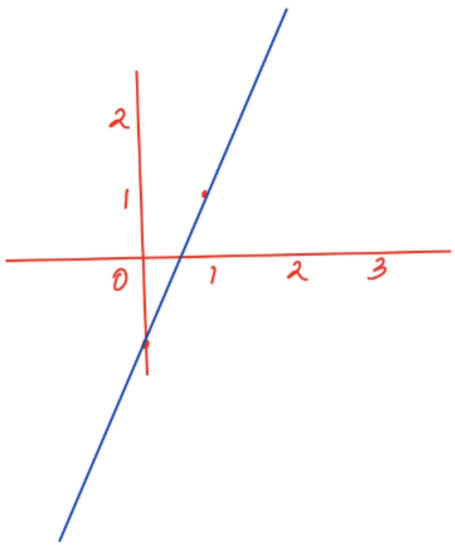
$$= \{ x \in \mathbb{R}^d : w^T x = b \}$$

$$= \{ x \in \mathbb{R}^d : \sum_{i=1}^{d} w_i x_i = b \}$$

Example Lines

Line through $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ along $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$

$$\left\{ x \in \mathbb{R}^2 : x = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ 2 \end{bmatrix} ; \alpha \in \mathbb{R} \right\}$$

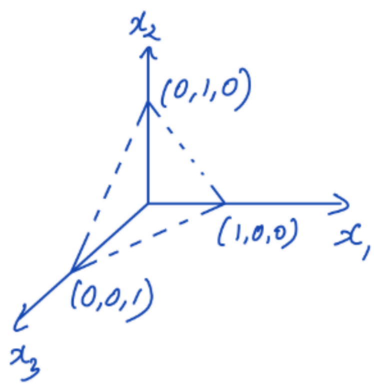


Example Planes

$d = 3$

Hyperplane normal to $\begin{bmatrix}1\\1\\1\end{bmatrix}$ with value 1

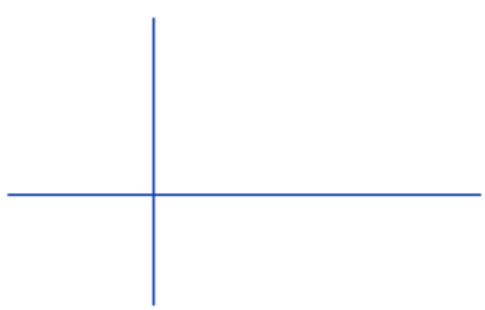
$$T = \{ x \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 1 \}$$



(0,1,0) lies on T which is perpendicular to the (1,1,1)

Tuples vs Points vs Vectors

$\mathbb{R}^d$



Partial Derivatives

$$f : \mathbb{R}^2 \to \mathbb{R} \qquad f(x_1, x_2) = x_1^2 + x_2^2$$

$$\frac{\partial f}{\partial x_1}(v) = \lim_{\alpha \to 0} \frac{f\left(v + \begin{bmatrix} \alpha \\ 0 \end{bmatrix}\right) - f(v)}{\alpha}$$

$$= \lim_{\alpha \to 0} \frac{f(v_1 + \alpha, v_2) - f(v_1, v_2)}{\alpha}$$

$$\frac{\partial f}{\partial x_2}(v) = \lim_{\alpha \to 0} \frac{f(v_1, v_2 + \alpha) - f(v_1, v_2)}{\alpha}$$

$$\frac{\partial f}{\partial x_i}(v) = \lim_{\alpha \to 0} \frac{f(v + \alpha e_i) - f(v)}{\alpha} \qquad e_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Gradients

$$f : \mathbb{R}^d \to \mathbb{R}$$

$$\frac{\partial f}{\partial x}(v) = \left[\frac{\partial f}{\partial x_1}(v), \frac{\partial f}{\partial x_2}(v) \cdots \frac{\partial f}{\partial x_d}(v) \right]$$

$$\nabla f(v) = \left[\frac{\partial f}{\partial x} \right]^T$$

e.g 1 $\quad d=2 \qquad f(x) = x_1^2 + x_2^2$;

$$\frac{\partial f}{\partial x_1}(v) = 2v_1 \quad ; \quad \frac{\partial f}{\partial x_2}(v) = 2v_2$$

$$\nabla f(v) = \begin{bmatrix} 2v_1 \\ 2v_2 \end{bmatrix}$$

---

e.g 1

$f : \mathbb{R}^3 \rightarrow \mathbb{R} \qquad f(x) = x_1 + 2x_2 + 3x_3$

$$\nabla f(v) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

QN : 2,3

Lecture 6 : Multivariate Calculus: Linear approximation and applications

Gradients and Linear Approximations

$$f: \mathbb{R} \to \mathbb{R}$$

$$f(x) \approx \underbrace{f(x^*) + f'(x^*)(x - x^*)}_{L_{x^*}[f](x)} \qquad \text{around } x = x^*$$

$$v \in \mathbb{R}^d, \quad x \in \mathbb{R}^d$$

$$f(x) \approx f(v) + \nabla f(v)^T (x - v)$$

$$= \underbrace{f(v) + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(v) \cdot (x_i - v_i)}_{L_v[f](x)} \qquad \text{around } x = v$$

2.

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

$$f(y_1, v_2) \approx f(v_1, v_2) + \frac{\partial f}{\partial x_1}(v) \cdot (y_1 - v_1)$$

$$f(y_1, v_2) - f(v_1, v_2) \approx \frac{\partial f}{\partial x_1}(v)(y_1 - v_1)$$

$$f(v_1, y_2) - f(v_1, v_2) \approx \frac{\partial f}{\partial x_2}(v)(y_2 - v_2)$$

$$f(y_1, y_2) - f(v_1, v_2) \approx \frac{\partial f}{\partial x_1}(v)(y_1 - v_1) + \frac{\partial f}{\partial x_2}(v)(y_2 - v_2)$$

$$f(y_1, y_2) \approx f(v_1, v_2) + \nabla f(v)^T (y - v)$$

Example :

$$f(x_1, x_2) = x_1^2 + x_2^2 \qquad \nabla f(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

(i) Approximate f around $(6,2)$

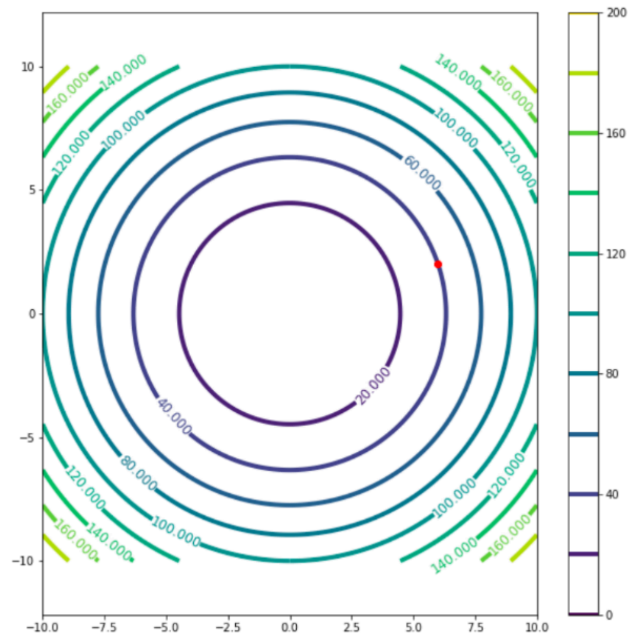
$$f(v) = 40, \quad \nabla f(v) = \begin{bmatrix} 12 \\ 4 \end{bmatrix}$$

$$f(x) \approx 40 + [12, 4] \begin{bmatrix} x_1 - 6 \\ x_2 - 2 \end{bmatrix}$$

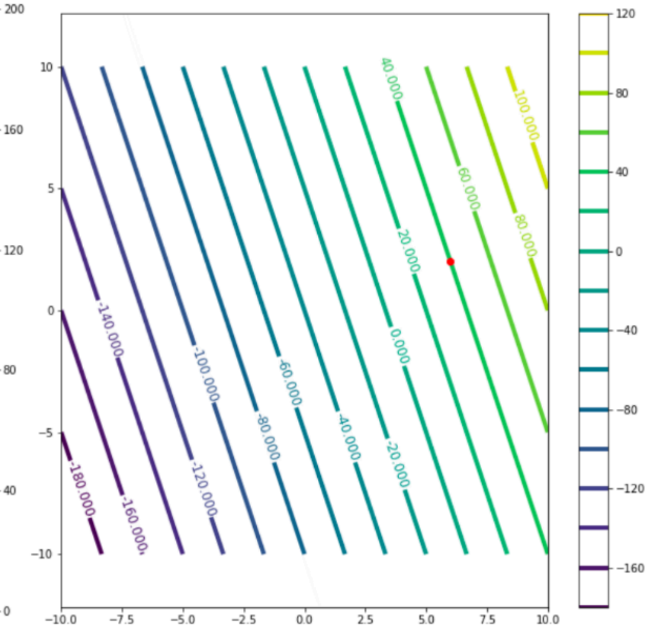
$$= 40 + 12(x_1 - 6) + 4(x_2 - 2)$$

$$= 40 + 12x_1 + 4x_2 - 72 - 8$$

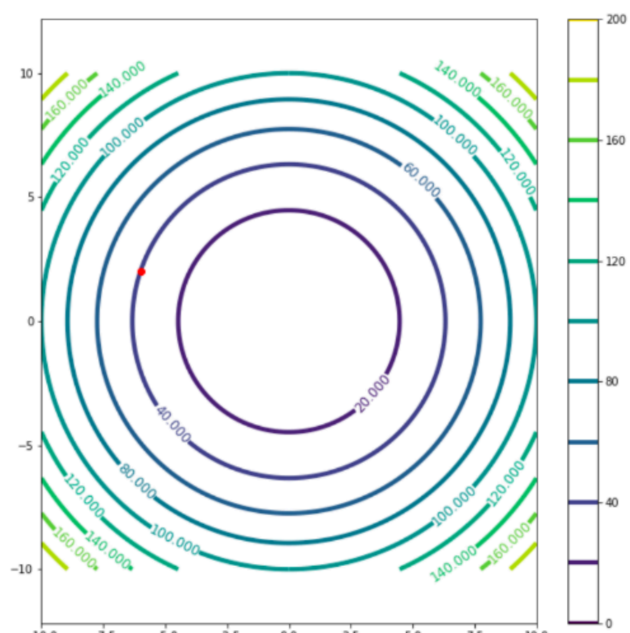
$$= 12x_1 + 4x_2 - 40 \qquad (x_1, x_2) \approx (6,2)$$

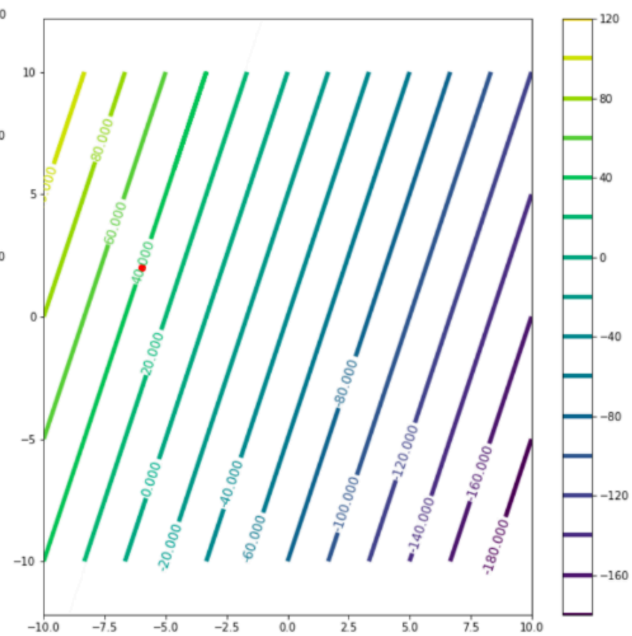


$f(x) = x_1^2 + x_2^2$



$L_v[f]$



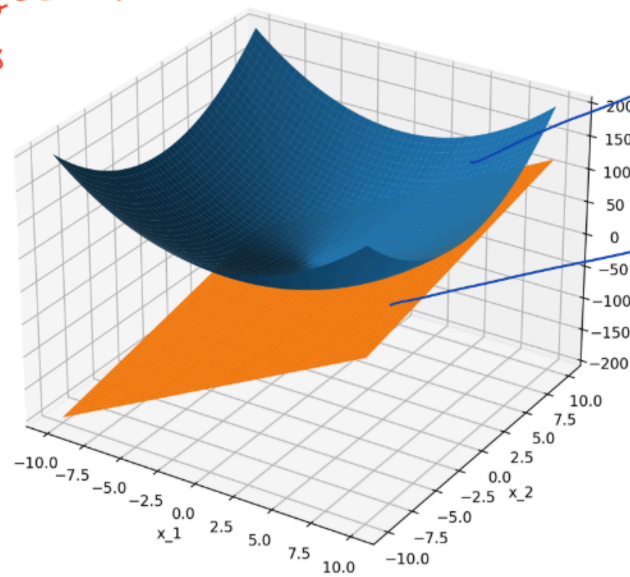


$f(x) \approx L_v[f](x) \quad (x \approx v)$

Gradients and Tangent Planes

$f(x) = x_1^2 + x_2^2$

The graph of $L_v[f]$ is a plane that is tangent to the graph of f at the point $(v, f(v))$

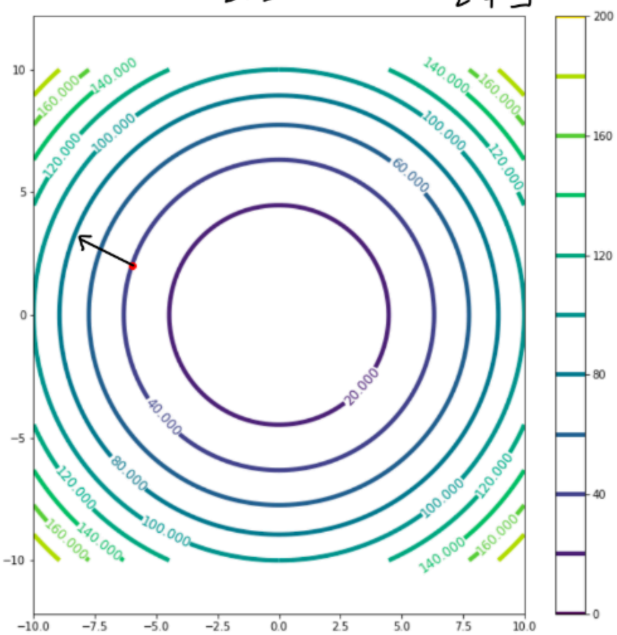


Graph of f

Graph of $L_v[f]$

Gradients and Contours

$$v = \begin{bmatrix} -6 \\ 2 \end{bmatrix} \quad \nabla f(v) = \begin{bmatrix} -12 \\ 4 \end{bmatrix}$$



$$\nabla f(v) \perp \{x \in \mathbb{R}^d : f(x) = f(v)\}$$

$$\nabla f(v) \perp \{x \in \mathbb{R}^d : L_v[f](x) = f(v)\}$$

$$\|$$

$$\{x \in \mathbb{R}^d : f(v) + \nabla f(v)^T (x - v) = f(v)\}$$

$$\|$$

$$\{x \in \mathbb{R}^d : \nabla f(v)^T x = \nabla f(v)^T v\}$$

$$\{x \in \mathbb{R}^d : w^T x = b\}$$

Directional Derivative

$$D_u[f](v) = \lim_{\alpha \to 0} \frac{f(v + \alpha u) - f(v)}{\alpha}$$

Directional derivative of f at the point v, along u.

$$= \lim_{\alpha \to 0} \frac{f(v) + \nabla f(v)^T \alpha u - f(v)}{\alpha}$$

$$= \nabla f(v)^T u$$

Cauchy-Schwarz Inequality

$$a_1, a_2 \cdots a_d \qquad \|a\| = \sqrt{a_1^2 + \cdots + a_d^2}$$
$$b_1, b_2 \cdots b_d$$

$$-\|a\| \cdot \|b\| \leq a^T b \leq \|a\| \|b\|$$

$a = \alpha b$, $\alpha < 0$ $\qquad$ $a = \alpha b$, $\alpha > 0$

Direction of Steepest Ascent

f

Find a direction u, that maximises the rate of change of f as you move from v along u.

Maximise $D_u[f](v)$

Find $u \in \mathbb{R}^d$, $\|u\| = 1$ and which maximises

$$D_u[f](v)$$
$$= \nabla f(v)^T u$$

$$u = \alpha \cdot \nabla f(v)$$

Descent Directions

$$f: \mathbb{R}^d \to \mathbb{R}$$

$$v \in \mathbb{R}^d.$$

What are the valid directions, such that f decreases

For what values of u: $D_u[f](v) < 0$

$$\Updownarrow$$

$$\nabla f(v)^T u < 0$$

Descent directions: $\{u \in \mathbb{R}^d : \nabla f(v)^T u < 0\}$

Higher Order Approximations

$$f: \mathbb{R}^d \to \mathbb{R}$$

$$f(x) \approx f(v) + \nabla f(v)^T (x - v)$$ (Valid around $x = v$)

$$f(x) \approx f(v) + \nabla f(v)^T (x - v) + \frac{1}{2} (x - v)^T \underbrace{\nabla^2 f(v)}_{\substack{d \times d \text{ matrix} \\ \text{Hessian}}} (x - v)$$

Maxima, minima and saddle points

If $f(x)$ is minimised at v

$$\Downarrow$$

$$\nabla f(v) = 0$$

$$\{v : \nabla f(v) = 0\} \Rightarrow \text{Critical point}$$

QN : 3

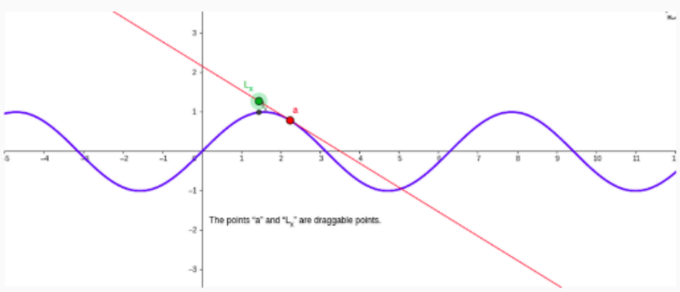
MACHINE LEARNING - FOUNDATIONS

Tutorial Week 2

1. Linear approximation
2. Higher order approximations
3. Multivariate linear approximation
4. Directional derivatives

Linear approximation (Linearization)

Def:
Approximation of any function using a linear function .



Need:

- Linear functions are easier to work with.
- Finding approximate values of functions at certain points when exact values are not known.

The equation

The linear approximation $L(x)$ of a function $f(x)$ at point a is given by:

$$\boxed{L(x) = f(a) + f^{'}(a)(x - a)}$$

This is indeed the equation of a tangent line:

$$y - y_1 = m(x - x_1)$$
$$y = y_1 + m(x - x_1)$$

If $x_1 = a$, $y_1 = f(a)$ and $m = f'(a)$, we get,

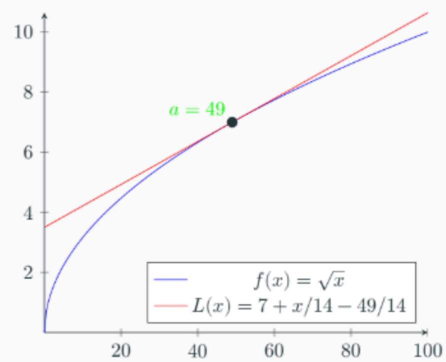
$$y = f(a) + f'(a)(x - a)$$

Problem 1

Compute the approximate value of $\sqrt{50}$.

The closest known value to $\sqrt{50}$ is $\sqrt{49}$, so we set $f(x) = \sqrt{x}$ and $a = 49$.

$$\begin{aligned} f^{'}(x) &= \frac{1}{2}x^{-\frac{1}{2}} \\ f(49) &= \sqrt{49} = 7 \\ f^{'}(49) &= \frac{1}{(2)(\sqrt{49})} = \frac{1}{14} \\ L(x) &= f(49) + f^{'}(49)(x - 49) \\ L(x) &= 7 + \frac{1}{14}(x - 49) \end{aligned}$$



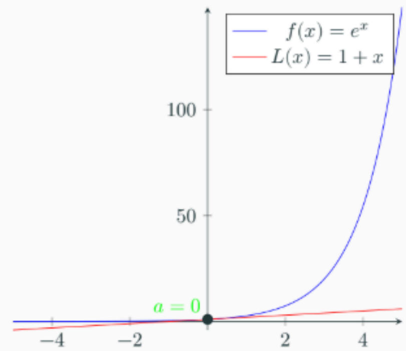
Approximate value of $\sqrt{50} = L(50) = 7 + \frac{1}{14}(50 - 49) = 7 + \frac{1}{14} = 7.071$

- **Note 1:** Actual value of $\sqrt{50}$ (up to 3 decimal places) is 7.071.
- **Note 2:** $L(100)$ gives 10.64 while the actual value of $\sqrt{100}$ is 10.

Problem 2

Compute the approximate value of $e^{0.017}$.

The closest known value to $e^{0.017}$ is e^0, so we set $f(x) = e^x$ and $a = 0$.



$$\begin{aligned} f'(x) &= e^x \\ f(0) &= 1 \\ f'(0) &= 1 \\ L(x) &= f(0) + f'(0)(x-0) \\ L(x) &= 1 + 1(x) = 1 + x \end{aligned}$$

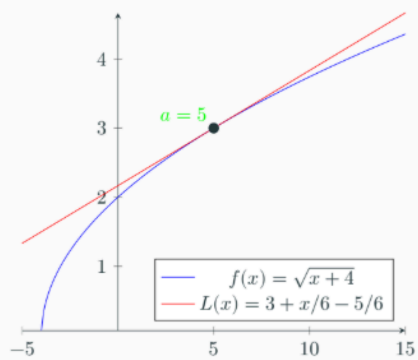
Approximate value of $e^{0.017} = L(0.017) = 1 + 0.017 = 1.017$

- **Note 1:** Actual value of $e^{0.017}$ is also 1.017.
- **Note 2:** $L(1)$ gives 2 while the actual value of e is 2.718.

Problem 3

Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

The closest known value to $\sqrt{x+4}$ is $\sqrt{9}$ with $x = 5$, so we set $f(x) = \sqrt{x+4}$ and $a = 5$.



$$\begin{aligned} f'(x) &= \frac{1}{2}(x+4)^{-\frac{1}{2}} \\ f(5) &= \sqrt{5+4} = 3 \\ f'(5) &= \frac{1}{(2)(\sqrt{9})} = \frac{1}{6} \\ L(x) &= f(5) + f'(5)(x-5) \\ L(x) &= 3 + \frac{1}{6}(x-5) = 3 + \frac{x-5}{6} \end{aligned}$$

Approximate value of $f(6) = L(6) = 3 + \frac{6-5}{6} = 3 + \frac{1}{6} = \frac{19}{6} = 3.1666$

- **Note:** Actual value of $f(6) = \sqrt{10}$ is 3.1622. **Why?**

HIGHER ORDER APPROXIMATIONS

Linear Approximation

$$\boxed{L(x) = f(a) + f'(a)(x-a)}$$

Quadratic Approximation

$$\boxed{L(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2}$$

Higher-order Approximations

$$\boxed{\begin{aligned} L(x) = f(a) + f^{(1)}(a)(x-a) + \frac{f^{(2)}(a)}{2}(x-a)^2 + \\ + \frac{f^{(3)}(a)}{3\cdot 2}(x-a)^3 + \frac{f^{(4)}(a)}{4\cdot 3\cdot 2}(x-a)^4 \ldots \end{aligned}}$$

Problem 4

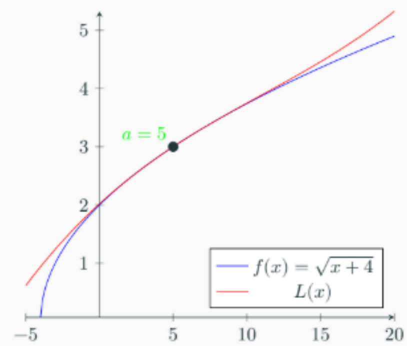
Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

The closest known value to $\sqrt{x+4}$ is $\sqrt{9}$ with $x = 5$, so we set $f(x) = \sqrt{x+4}$ and $a = 5$.

$$\begin{aligned} f'(x) &= \frac{1}{2}(x+4)^{-\frac{1}{2}} \\ f''(x) &= -\frac{1}{4}(x+4)^{-\frac{3}{2}} \\ f'''(x) &= \frac{3}{8}(x+4)^{-\frac{5}{2}} \\ f(5) &= \sqrt{5+4} = 3 \\ f'(5) &= \frac{1}{(2)(\sqrt{9})} = \frac{1}{6} \\ f''(5) &= -\frac{1}{108} \\ f'''(5) &= \frac{1}{(24)(27)} \end{aligned}$$

$$L(x) = f(5) + f'(5)(x-5) + \frac{f''(5)}{2}(x-5)^2 + \frac{f'''(5)}{(3)(2)}(x-5)^3 + \ldots$$

$$L(x) = 3 + \frac{1}{6}(x-5) - \frac{1}{(108)(2)}(x-5)^2 + \frac{1}{(24)(27)(6)}(x-5)^3$$



Approximate value of
$f(6) = L(6) = 3 + \frac{1}{6}(6-5) - \frac{1}{(108)(2)}(6-5)^2 + \frac{1}{(24)(27)(6)}(6-5)^3 = 3.1622$

MULTIVARIATE LINEAR APPROXIMATION

Linear approximation of functions involving multiple variables

The linear approximation of a function f of two variables x and y in the neighborhood of (a, b) is:

$$\boxed{L(x,y) = f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b)}$$

Problem 5

Find the linearization of $f(x,y) = xe^{xy}$ at $(1,0)$. Use it to approximate $f(1.1,-0.1)$.

$$\begin{aligned}\frac{\partial f}{\partial x}(x,y) &= xe^{xy}y + e^{xy} = xye^{xy} + e^{xy}\\ \frac{\partial f}{\partial y}(x,y) &= xe^{xy}x = x^2e^{xy}\end{aligned}$$

Here $(a,b) = (1,0)$.

$$\begin{aligned}f(1,0) &= e^0 = 1\\ \frac{\partial f}{\partial x}(a,b) &= \frac{\partial f}{\partial x}(1,0) = e^0 = 1\\ \frac{\partial f}{\partial y}(a,b) &= \frac{\partial f}{\partial y}(1,0) = e^0 = 1\end{aligned}$$

$$\begin{aligned}L(x,y) &= f(1,0) + \frac{\partial f}{\partial x}(1,0)(x-1) + \frac{\partial f}{\partial y}(1,0)(y-0)\\ &= 1 + 1(x-1) + 1(y)\\ &= x+y\\ f(1.1,-0.1) &= L(1.1,-0.1)\\ &= 1.1 - 0.1 = 1\end{aligned}$$

The actual value of f(1.1, -0.1) = 1.1e$^{-0.11} = \frac{1.1}{1.11628} = 0.98542$

DIRECTIONAL DERIVATIVES

- $f_x(x,y) = \frac{\partial f}{\partial x}(x,y)$ = Rate of change of f as we vary x (keeping y fixed).
- $f_y(x,y) = \frac{\partial f}{\partial y}(x,y)$ = Rate of change of f as we vary y (keeping x fixed).
- Directional derivative of $f(x,y)$ = Rate of change of f if we allow both x and y to change simultaneously (in some direction (u)).

$$\begin{aligned} D_{\vec{u}} f(x,y) &= \nabla f \cdot u \\ &= \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right] \cdot [u_1, u_2] \\ &= u_1 \frac{\partial f}{\partial x} + u_2 \frac{\partial f}{\partial y} \end{aligned}$$

Directional derivative can be considered to be a weighted sum of partial derivatives.

Problem 6

Find the derivative of $f(x,y) = xcos(y)$ in the direction of $\vec{u} = [2,1]$.

$$\frac{\partial f}{\partial x} = \cos(y)$$
$$\frac{\partial f}{\partial y} = -x\sin(y)$$

Unit vector in the direction of $\vec{u}$ = $\left[\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right]$

$$\begin{aligned} D_{\vec{u}} f(x,y) &= u_1 \frac{\partial f}{\partial x} + u_2 \frac{\partial f}{\partial y} \\ &= \frac{2}{\sqrt{5}} \cos(y) - \frac{1}{\sqrt{5}} x \sin(y) \end{aligned}$$

Problem 7

Find the derivative of $f(x,y) = x^2 - xy$ in the direction of $\vec{u} = 0.6i + 0.8j$ at the point $(2,-3)$.

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

$\vec{u}$ is already a unit vector. $u_1 = 0.6, u_2 = 0.8$.

$$\begin{aligned} D_{\vec{u}} f(x,y) &= 0.6(2x-y) + 0.8(-x) \\ &= 0.6(2x-y) - 0.8x \end{aligned}$$

$$\begin{aligned} D_{\vec{u}} f(2,-3) &= 0.6(2(2)+3) - 0.8(2) \\ &= 0.6(7) - 1.6 \\ &= 4.2 - 1.6 \\ &= 2.6 \end{aligned}$$

Live Session 2 - Recup of Linear algebra .

$\mathbb{R}^n$

$\mathbb{R}$

line

$\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$

plane

$\mathbb{R}^n = \{(x_1, \cdots, x_n) \mid x_1, \cdots, x_n \in \mathbb{R}\}$

$\mathbb{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$

space

$M_{m\times n}(\mathbb{R})$

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}$$

$M_{3\times 4}(\mathbb{R})$

set of all 3×4 real matrices

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \end{bmatrix}$$

3×4

$M_{m\times n}(\mathbb{R})$

set of all $m \times n$ real matrices

Matrix-Vector Multiplication

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} = 3 \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} + (-1)\begin{bmatrix} 3 \\ 4 \end{bmatrix} + 2 \cdot \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

Linear combination of the columns

$m \times n$ $\quad n \times 1$ $\quad m \times 1$

$$\begin{bmatrix} | & & | \\ c_1 & \cdots & c_n \\ | & & | \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 c_1 + \cdots + x_n c_n$$

$M_{m\times n}(\mathbb{R})$ $\quad \mathbb{R}^n$ $\quad \mathbb{R}^m$

Vector-Matrix Multiplication

$$\begin{bmatrix} 3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} = 3 \cdot \begin{bmatrix} 1 & 3 & 5 \end{bmatrix} + (-1) \begin{bmatrix} 2 & 4 & 6 \end{bmatrix}$$

Linear combination of the rows

$$\underset{\mathbb{R}^m}{\overset{1 \times m}{\begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}}} \underset{M_{m\times n}(\mathbb{R})}{\overset{m \times n}{\begin{bmatrix} - & r_1^T & - \\ & \vdots & \\ - & r_m^T & - \end{bmatrix}}} = \underset{\mathbb{R}^n}{\overset{1 \times n}{x_1 r_1^T + \cdots + x_m r_m^T}}$$

Vector-Vector Multiplication (Inner Product)

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

Dot product

$$\underset{\mathbb{R}^n}{\overset{1 \times n}{\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}}} \underset{\mathbb{R}^n}{\overset{n \times 1}{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}} = \underset{\mathbb{R}}{\overset{1 \times 1}{x_1 y_1 + \cdots + x_n y_n}}$$

Vector-Vector Multiplication (Outer Product)

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 5 & 6 & 7 \end{bmatrix} = \begin{bmatrix} 5 & 6 & 7 \\ 10 & 12 & 14 \\ 15 & 18 & 21 \end{bmatrix}$$

Outer Product

$$\begin{array}{ccccc} m \times 1 & 1 \times n & & m \times n \\ \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} & \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix} & = & \begin{bmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & & \vdots \\ x_m y_1 & \cdots & x_m y_n \end{bmatrix} \\ \mathbb{R}^m & \mathbb{R}^n & & M_{m \times n}(\mathbb{R}) \end{array}$$

Matrix-Matrix Multiplication

$$\boxed{AB = C}$$

$$m \times n \qquad n \times p \qquad m \times p$$

- Only matrices of compatible dimensions can be multiplied
  - # columns of A = # rows of B
- Matrix multiplication is not commutative
  - In general $AB \neq BA$
  - If $AB = BA$, we say that A and B commute

Matrix-Matrix Multiplication

$$AB = C$$

$$m \times n \qquad n \times p \qquad m \times p$$

Matrix-Vector

$$A\begin{bmatrix} | & & | \\ b_1 & \cdots & b_p \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ Ab_1 & \cdots & Ab_p \\ | & & | \end{bmatrix}$$

Vector-Vector (Inner Product)

$$\begin{bmatrix} — & a_1^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix}\begin{bmatrix} | & & | \\ b_1 & \cdots & b_p \\ | & & | \end{bmatrix} = \begin{bmatrix} & \vdots & \\ \cdots & a_i^T b_j & \cdots \\ & \vdots & \end{bmatrix}$$

Vector-Matrix

$$\begin{bmatrix} — & a_1^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} B = \begin{bmatrix} — & a_1^T B & — \\ & \vdots & \\ — & a_m^T B & — \end{bmatrix}$$

Vector-Vector (Outer Product)

$$\begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix}\begin{bmatrix} — & b_1^T & — \\ & \vdots & \\ — & b_n^T & — \end{bmatrix} = a_1 b_1^T + \cdots + a_n b_n^T$$

Special Matrices

square matrix

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

$$A \to n \times n$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$ Diagonal $$D = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & d_n \end{bmatrix}$$

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$ Scalar $$S = \begin{bmatrix} c & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & c \end{bmatrix}$$

$$S = cI$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$ Identity $$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Special Matrices

$$\begin{bmatrix} 1 & 1 & -1 \\ 0 & -2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$$ Upper Triangular $$U = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ & \ddots & \vdots \\ \mathbf{0} & & a_{nn} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & -2 & 0 \\ -1 & 2 & 3 \end{bmatrix}$$ Lower Triangular $$L = \begin{bmatrix} a_{11} & & \mathbf{0} \\ \vdots & \ddots & \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

Transpose

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

$$A \to m \times n$$

$$A^T \to n \times m$$

$$\left(A^T\right)_{ij} = A_{ji}$$

$$\left(A^T\right)^T = A$$

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

Symmetric and Skew-symmetric

Symmetric

$$A^T = A$$

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

$$A^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 4 \\ 3 & 4 & 1 \end{bmatrix}$$

Skew-Symmetric

$$A^T = -A$$

$$A = \begin{bmatrix} 0 & 2 & 3 \\ -2 & 0 & 4 \\ -3 & -4 & 0 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 0 & -2 & -3 \\ 2 & 0 & -4 \\ 3 & 4 & 0 \end{bmatrix}$$

For any square matrix A

$$A = \underbrace{\frac{A + A^T}{2}}_{\text{symmetric}} + \underbrace{\frac{A - A^T}{2}}_{\text{skew-symmetric}}$$

Inverse

$$A \to n \times n \qquad\qquad B \to n \times n$$

$$AB = BA = I \implies B = A^{-1} \text{ and } A = B^{-1}$$

$$\left(A^{-1}\right)^{-1} = A$$

$$(cA)^{-1} = \frac{1}{c} \cdot A^{-1} \quad (c \neq 0)$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

$$\left(A^{T}\right)^{-1} = \left(A^{-1}\right)^{T}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \text{ if } ad - bc \neq 0$$

Determinants

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$|A| = (-1)^{(1+1)}a_{11}M_{11} + (-1)^{(1+2)}a_{12}M_{12} + (-1)^{(1+3)}a_{13}M_{13} = a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13}$$

$$M_{11} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{22}a_{33} - a_{23}a_{32} \qquad M_{12} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{21}a_{33} - a_{23}a_{31}$$

$$M_{13} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{21}a_{32} - a_{22}a_{31}$$

Determinants

$M_{ij} \to$ minor

$$|A| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij}$$

Expanding along row i Expanding along column j

$M_{ij} =$ determinant of the matrix formed by deleting row i, column j

Determinants

$C \to$ co-factor matrix

$\text{adj}(A) \to$ adjugate

$C_{ij} = (-1)^{i+j} M_{ij}$

$\text{adj}(A) = C^T$

$A \cdot \text{adj}(A) = \text{adj}(A) \cdot A = \det(A) \cdot I$

A is invertible if and only if $\det(A) \neq 0 \longrightarrow A^{-1} = \dfrac{1}{\det(A)} \text{adj}(A)$

Determinants and Row Operations

- Swapping two rows changes the sign of the determinant.
- Scaling a row by a constant scales the determinant by the same constant.
- Adding a constant times one row to another row leaves the determinant unchanged.

Consequences:

- If a matrix has a zero row, its determinant is zero
- If two rows of a matrix are the same, its determinant is zero.
- If a row of a matrix is a linear combination of other rows, its determinant is zero.

Determinants

- $|AB| = |A| \cdot |B|$
- $|A^T| = |A|$
- $|A^{-1}| = \dfrac{1}{|A|}$, if $|A| \neq 0$
- $|cA| = c^n|A|$

- If A is upper triangular or lower triangular its determinant is the product of the diagonal entries.
- Specifically, if A is diagonal, its determinant is the product of its diagonal entries.

Elementary Row Operations

- Swap two rows
- Scale a row by a **non-zero** constant
- Add a constant times one row to **another** row

Swap rows | Elementary Matrices

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix}$$

$$E_1 A = R$$

Reversible operation $\implies$ Invertible Matrix

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

$$E_1 E_2 = I$$

$$E_2 R = A$$

$$\begin{bmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Scale a row | Elementary Matrices

$$\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}\xrightarrow{R_1\to 2R_1}\begin{bmatrix}2&4&6\\4&5&6\\7&8&9\end{bmatrix}$$

$$\begin{bmatrix}2&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}=\begin{bmatrix}2&4&6\\4&5&6\\7&8&9\end{bmatrix}$$

$$E_1A=R$$

Reversible operation $\implies$ Invertible Matrix

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

$$E_1E_2=I$$

$$E_2R=A$$

$$\begin{bmatrix}2&4&6\\4&5&6\\7&8&9\end{bmatrix}\xrightarrow{R_1\to\frac{1}{2}R_1}\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}$$

$$\begin{bmatrix}1/2&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}2&4&6\\4&5&6\\7&8&9\end{bmatrix}=\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}$$

Add multiple of row to another | Elementary Matrices

$$\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}\xrightarrow{R_2\to R_2-2R_1}\begin{bmatrix}1&2&3\\2&1&0\\7&8&9\end{bmatrix}$$

$$\begin{bmatrix}1&0&0\\-2&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}=\begin{bmatrix}1&2&3\\2&1&0\\7&8&9\end{bmatrix}$$

$$E_1A=R$$

Reversible operation $\implies$ Invertible Matrix

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

$$E_1E_2=I$$

$$E_2R=A$$

$$\begin{bmatrix}1&2&3\\2&1&0\\7&8&9\end{bmatrix}\xrightarrow{R_2\to R_2+2R_1}\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}$$

$$\begin{bmatrix}1&0&0\\2&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}1&2&3\\2&1&0\\7&8&9\end{bmatrix}=\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}$$

Row Echelon Form (REF)

Leading entry (OR) **Pivot**: First non-zero entry in a row

<u>Conditions</u>

- All zero-rows are at the bottom
- The pivot in every non-zero row is 1.
- The pivot in a non-zero row is to the right of the pivot in the previous row.

$$\begin{bmatrix} \mathbf{1} & 0 & 1 & 2 \\ 0 & \mathbf{1} & 2 & 3 \\ 0 & 0 & 0 & \mathbf{1} \end{bmatrix} \quad \begin{bmatrix} \mathbf{1} & -3 \\ 0 & \mathbf{1} \end{bmatrix} \quad \begin{bmatrix} \mathbf{1} & 0 & 1 & 2 \\ 0 & 0 & \mathbf{1} & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} \mathbf{1} & 0 \\ 0 & \mathbf{1} \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 & 1 & 2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Reduced Row Echelon Form (RREF)

Pivot column: A column that contains a pivot

<u>Conditions</u>

- Matrix is in row echelon form.
- In every pivot column, the pivot is the only non-zero entry.

- REF(A) is not unique
- RREF(A) is unique

$$\begin{bmatrix} \mathbf{1} & 0 & -1 & 0 \\ 0 & \mathbf{1} & 1 & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{bmatrix} \quad \begin{bmatrix} \mathbf{1} & 0 \\ 0 & \mathbf{1} \end{bmatrix} \quad \begin{bmatrix} \mathbf{1} & 0 & 0 & 2 \\ 0 & 0 & \mathbf{1} & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} \mathbf{1} & 0 \\ 0 & \mathbf{1} \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{1} & -3 \\ 0 & \mathbf{1} \end{bmatrix} \quad \begin{bmatrix} \mathbf{1} & 2 \\ 0 & \mathbf{1} \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$A \rightarrow R$: Example

$$\begin{bmatrix} 1 & 2 & -1 & 3 \\ 3 & 1 & 2 & -6 \\ 4 & 3 & 1 & -3 \end{bmatrix} \xrightarrow{R_2 \rightarrow R_2 - 3R_1} \begin{bmatrix} 1 & 2 & -1 & 3 \\ 0 & -5 & 5 & -15 \\ 4 & 3 & 1 & -3 \end{bmatrix} \xrightarrow{R_3 \rightarrow R_3 - 4R_1} \begin{bmatrix} 1 & 2 & -1 & 3 \\ 0 & -5 & 5 & -15 \\ 0 & -5 & 5 & -15 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & -1 & 3 \\ 0 & -5 & 5 & -15 \\ 0 & -5 & 5 & -15 \end{bmatrix} \xrightarrow{R_3 \rightarrow R_3 - R_2} \begin{bmatrix} 1 & 2 & -1 & 3 \\ 0 & -5 & 5 & -15 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{R_2 \rightarrow \frac{-1}{5} R_2} \begin{bmatrix} 1 & 2 & -1 & 3 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\underset{\text{REF}}{\begin{bmatrix} 1 & 2 & -1 & 3 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}} \xrightarrow{R_1 \rightarrow R_1 - 2R_2} \underset{\text{RREF}}{\begin{bmatrix} 1 & 0 & 1 & -3 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}}$$

$A \rightarrow R$

- m row operations
- m elementary matrices
- $E_1 \rightarrow \cdots \rightarrow E_m$
- $E = E_m \cdots E_1$
- Since E_i is invertible, E is invertible

$$R = E_m \cdots E_1 A$$

$$R = EA$$

Elementary row operations $\rightarrow$ left-multiplication by an invertible matrix

REF, RREF and Invertibility

$A \to n \times n$

- A is invertible if and only if $REF(A)$ has n pivots
- A is invertible if and only if $RREF(A) = I$

$$R = EA \;\to\; I = EA$$

$$A^{-1} = E$$

System of Linear Equations

$$2x - 3y + 4z = 5$$

$$x + y - z = 2$$

$$3x - y + 5z = -1$$

$$\boxed{Ax = b}$$

$$\begin{bmatrix} 2 & -3 & 4 \\ 1 & 1 & -1 \\ 3 & -1 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$$

$b = 0 \Longrightarrow$ homogeneous system

Gaussian Elimination

$$\boxed{Rx = c}$$

Pivot columns $\rightarrow$ dependent variables
Non-pivot columns $\rightarrow$ independent variables

$$\begin{tikzcd}
Ax = b \arrow[d, leftrightarrow] \arrow[dd, bend left=60, "\text{Forward Elimination (using row reduction)}"] \\
EAx = Eb \arrow[d, leftrightarrow] \\
Rx = c
\end{tikzcd}$$

Backward Substitution

- Set aribitrary values for independent variables.
- Solve for dependent variables.

Augmented matrix $\quad [A|b] \rightarrow [R|c]$

Gaussian Elimination

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

$y, w \rightarrow$ independent
$x, z \rightarrow$ dependent

$$\begin{tikzcd}
\begin{array}{l} y = t_1 \\ w = t_2 \end{array} \arrow[rr] \arrow[dr] & & \begin{array}{l} x = 1 + t_1 + t_2 \\ z = -1 - 2t_2 \end{array} \arrow[dl] \\
& S = \{(1 + t_1 + t_2, t_1, \ -1 - 2t_2, t_2) : \ t_1, t_2 \in \mathbb{R}\} &
\end{tikzcd}$$

Gaussian Elimination

$Ax = b,\ A \rightarrow m \times n, \quad [A|b] \rightarrow [R|c]$

- \# pivots = \# dependent variables
- \# dependent variables + \# independent variables = n
- If $[R|c]$ has a pivot in the last column, $Ax = b$ has no solution
- If not, it has at least one solution:
  - If there are no independent variables, $Ax = b$ has a unique solution
  - If there is at least one independent variable, $Ax = b$ has infinitely many solutions.

PA : 2,7,10,14
GA : 4,5,7,11,12,13