Some basics of topology and real analysis

Upper and lower bounds; sup and max; inf and min

$S \subseteq \mathbb{R}$

* We say that x is an upper bound for S if

$$y \le x \quad \forall y \in S$$

lower bound if $\quad y \ge x \quad \forall y \in S$

* We say that x is the supremum of S if

least upper bound.

① x is an upper bound for S

② if y is an upper bound for S, $y \ge x$

* Similarly infimum is the greatest lower bound.

$\sup [1,2] = 2$ $\max [1,2] = 2$

$\sup (1,2) = 2$ $\max [1,2)$ does not exist

* If $\sup S$ lies in S, we call it the maximum
  $\inf S \in S$, we call it the minimum.

* If S is not bounded from above, $\sup S = \infty$
  below, $\inf S = -\infty$

* Every nonempty $S \subseteq \mathbb{R}$ has sup & inf.

Consider $\underset{x \in (0,1)}{\text{maximize}} \quad x^2$

$$\sup_{x \in (0,1)} x^2 = 1$$

Countable and uncountable sets

A set S is countable if $\exists$ a one-one map from S to $\mathbb{N}$

Eg: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q} \equiv (\mathbb{Z}, \mathbb{Z})$ $\mathbb{Q}^k$

$2\mathbb{Z}$, $2\mathbb{Z}+1$

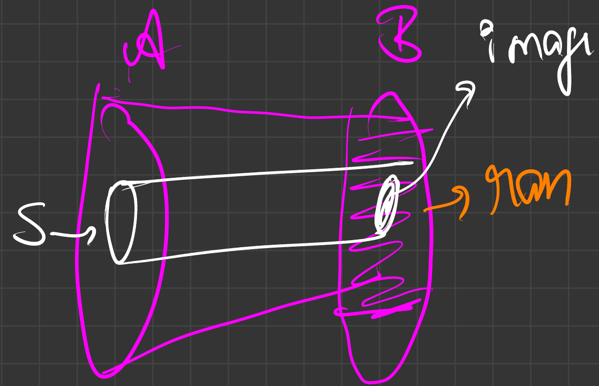
Eg: $\mathbb{R}$, $\mathbb{C}$, $[0,1]$, etc.

Functions: domain, co-domain, range, image, inverse image

$$f : A \to B$$

$A \to$ domain, $B \to$ codomain

$$f(A) = \{ f(x) : x \in A \} \longrightarrow \text{range}$$



For $S \subseteq A$,

$$f(S) = \{ f(x) : x \in S \} \longrightarrow \text{image of } S \text{ under } f$$

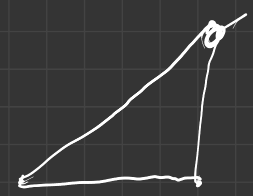
$y \in B$,

$$f^{-1}(y) = \{ x \in A : f(x) = y \} \quad \text{inverse image}$$

Metric : $\mathcal{A}$

A metric $d : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$

1. $d(x,y) \geq 0 \quad \forall x,y \in \mathcal{A}$
2. $d(x,y) = 0$ if & only if $x = y$
3. $d(x,y) = d(y,x)$
4. $d(x,z) \leq d(x,y) + d(y,z)$



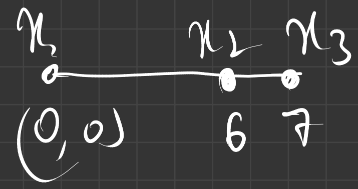
Eg : $d(\underline{x}, \underline{y}) = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2}$ is a metric

$$d^2(\underline{x}, \underline{y}) = \sum_{i=1}^{n} (x_i - y_i)^2$$

$$(x_1 - x_3)^2 = 49$$

$$(x_1 - x_2)^2 = 36$$

$$(x_2 - x_3)^2 = 1$$



$$d_p(\underline{x}, \underline{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p} \rightarrow L_p \text{ metric}$$

Norm and inner product

Norm : $f : A \to \mathbb{R}$ satisfying (A is a vector space over $\mathbb{R}$)

① $f(\underline{x}) \geq 0 \quad \forall x \in A$

② $f(\underline{x}) = 0$ iff $x = 0$

③ $f(\alpha \underline{x}) = |\alpha| f(\underline{x}) \quad \forall \underline{x} \in A, \ \alpha \in \mathbb{R}$

④ $f(x+y) \leq f(x) + f(y)$

If f is a norm, then $d(x,y) = f(x-y)$ is a metric.

$\mathcal{V}$ vector space.

$f : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ is called an inner product if

① $f(\underline{x}, \underline{y}) = f(\underline{y}, \underline{x}) \quad \forall \underline{x}, \underline{y} \in \mathcal{V}$

② $f(\underline{x}, \underline{x}) \geq 0 \quad \forall \underline{x}$

equality iff $\underline{x} = \underline{0}$

③ $f(\alpha \underline{x}_1 + \beta \underline{x}_2, \underline{y}) = \alpha f(\underline{x}_1, \underline{y}) + \beta f(\underline{x}_2, \underline{y})$

Eg: $f(\underline{x}, \underline{y}) = \sum_{i=1}^{n} x_i y_i$

(2) Consider $\mathcal{Y} = R^2$

$$f(\underline{x}, \underline{y}) = 2x_1 y_1 + 3x_2 y_2$$

$$= \begin{bmatrix} 2y_1 & 3y_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

This is an inner product

(3) $f(\underline{x}, \underline{y}) = 2x_1 y_1 - 3x_2 y_2$

Not an inner product

(3) $f(\underline{x}, \underline{y}) = \underline{x}^T A \underline{y}$ is an inner product if A is symm. P.D.

Sequences and limits

A sequence on $\mathbb{R}$ is $f : \{1, 2, 3, \ldots\} \to \mathbb{R}$

Limit: A sequence on a metric space (A, d) $\{x_1, x_2 \ldots\}$ converges to a limit x if for every $\epsilon > 0$, we can find N s.t.

$$d(x_n, x) < \epsilon \quad \forall n \geq N$$

Eg: $(0, 1)$ $\quad x_n = 1/2^n$, $n = 1, 2, \ldots$

Closed sets: A set A is closed if every convergent sequence of elements from A, converges to some element in A

① $[0, 1]$

② $A = \{1, 2, 5, 10\}$ is closed

$$x_n = \begin{cases} 1 & \text{if } 100 \nmid n \\ 2 & \text{if } 100 \mid n \end{cases}$$

③ Every finite subset of $\mathbb{R}^n$ is closed

④ $\mathbb{Q}$

is not closed $x_n = \left(1 + \frac{1}{n}\right)^n \rightarrow e \notin \mathbb{Q}$

Open set

Open neighborhood: $B(\underline{x}, \epsilon) = \{ y \in A : d(\underline{x}, y) < \epsilon \}$

$\downarrow$

open neighborhood of radius ϵ around $\underline{x}$

A set S is open if for every $\underline{x} \in S$, $\exists\ \epsilon$ s.t. $B(\underline{x}, \epsilon) \subseteq S$

Eg: ① $(0, 1)$

② Every open neighbourhood is open

③ $\mathbb{R}$ is Open

(4) The compliment of a closed set is open.

$$[0,1]^c = (-\infty, 0) \cup (1, \infty)$$

$\{1, 2, 3\}^c$ is open

$$= (-\infty, 1) \cup (1, 2) \cup (2, 3) \cup (3, \infty)$$

(5) Union of open sets is open

(6) $\mathbb{R}$ is open & closed

(7) Compliment of an open set is closed

$$(0,1)^c = (-\infty, 0] \cup [1, \infty)$$

$(0, 1]$

(8) ϕ is open & closed

* All these definitions assume you have a fixed metric space.

- $\mathbb{R}$ is open & closed : $\mathbb{R}$, d = absolute value of diff

- $\underbrace{\mathbb{C}, d} = |x-y|$

If this is our metric space, then

$\mathbb{R}$ is closed but not open

Compact sets:

A is bounded if $\exists\ \alpha \in \mathbb{R}$ st

$$d(x,y) \leq \alpha \quad \forall x,y \in A.$$

If $A \subseteq \mathbb{R}^n$ is closed & bounded, we say that it is <u>compact</u>

Theorem: If $S \subseteq \mathbb{R}^n$ is compact, $f: S \to \mathbb{R}$ is continuous

* $\max\limits_{x \in S} f(x)$ exists

* $\min\limits_{x \in S} f(x)$ exists

Eg: ① $f: \mathbb{R} \to \mathbb{R}$ $\quad$ $\mathbb{R}$ is not bounded

$f(x) = x^2$ $\quad$ Max does not exist

② $S = (0,1) \to$ open

$f(x) = x^2$ $\quad$ (sup & inf exist)

min & max do not exist

③ $f: [0,1] \to \mathbb{R}$ not continuous.

$$f(x) = \begin{cases} x^2 & x \in (0,1) \\ \frac{1}{2} & x \in \{0,1\} \end{cases}$$

Continuity :

$f: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is continuous at x if for every $\epsilon > 0$,

$\exists\ \delta > 0$ st

$$d(x,y) < \delta \Rightarrow d(f(x), f(y)) < \epsilon$$

$$|||$$

$$x_n \to x \Rightarrow f(x_n) \to f(x)$$

Derivative

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

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

Definition $f : \mathbb{R}^n \to \mathbb{R}^m$

We say that $D_{f(x)} \in \mathbb{R}^{m \times n}$ is derivative of f at x

$$\lim_{y \to x} \frac{\| f(y) - f(x) - D_{f(x)}(y - x) \|}{\|y - x\|} = 0$$

Check that both defns are same in $\mathbb{R}$

If ~~the~~ derivative exists, then

$D_{f(x)}$ is equal to the Jacobian

If all partial derivatives are continuous then

$D_{f(x)}$ = Jacobian.

If $f: \mathbb{R}^n \to \mathbb{R}$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = D^T_{f(x)}$$

① $f(\underline{x}) = \underline{x}^T \underline{x} = \sum_{i=1}^{n} x_i^2$

$D_{f(x)} \in \mathbb{R}^{1 \times n}$

$$\nabla f(\underline{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix} = 2\underline{x}$$

② $f(\underline{x}) = \underline{b}^T \underline{x}$

$\nabla f(\underline{x}) = \underline{b}$

③ $f(\underline{x}) = \underline{x}^T A \underline{x}$

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

$$x^T A x = [x_1 \quad x_2] \begin{bmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{bmatrix}$$

$$= x_1^2 + 2x_1x_2 + 3x_2x_1 + 4x_2^2$$

$$= x_1^2 + 5x_1x_2 + 4x_2^2$$

$$\frac{\partial f}{\partial x_1} = 2x_1 + 5x_2$$

$$\nabla f(x) = \begin{bmatrix} 2 & 5 \\ 5 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\frac{\partial f}{\partial x_2} = 5x_1 + 8x_2$$

$$\nabla f(\underset{\sim}{x}) = (A + A^T)\underset{\sim}{x}$$

Ex : Prove this

Recap

1. Metric, norm and inner product

1. Limits of sequences

2. Open, closed, compact sets

3. Limits and continuity of functions

4. Derivative and gradient

Derivative: examples

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

$$f(\underline{x}) = \|\underline{x}\|_2^2 = \underline{x}^T \underline{x}$$

$$D_{f(x)} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

$$= \begin{bmatrix} 2x_1 & 2x_2 & \cdots & 2x_n \end{bmatrix} = 2\underline{x}^T$$

Inner product for matrices

$D_{f(x)}$ is the derivative if

$$\lim_{z \to x} \frac{\| f(z) - f(x) - D_{f(x)}(z - x) \|_2}{\| z - x \|} = 0$$

$$\lim_{z \to x} \frac{\| f(z) - f(x) - \langle D_{f(x)}, z - x \rangle \|}{\| z - x \|} = 0$$

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

Claim: $f(A, B) = Tr(A^T B)$ is an inner product

① Symmetric ✓

② $Tr(A^T A) \geq 0$

$$\begin{bmatrix} \underline{a}_1 & \underline{a}_2 & \cdots & \underline{a}_n \end{bmatrix}^T \begin{bmatrix} \underline{b}_1 & \underline{b}_2 & \cdots & \underline{b}_n \end{bmatrix}$$

$$= \begin{bmatrix} \underline{a}_1^T \underline{b}_1 & & & \\ & \underline{a}_2^T \underline{b}_2 & & \\ & & \ddots & \\ & & & \underline{a}_n^T \underline{b}_n \end{bmatrix}$$

$$Tr(A^T B) = \sum_{i=1}^{n} a_i^T b_i = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ij}$$

Derivative

$f : P.D_{n\times n} \text{ matrices} \longrightarrow \mathbb{R}$

$f(A) = \log \det(A)$

$Z = X + \Delta X$ $\qquad$ Z, X are $P.D$

* $y^T Z y > 0$, $y^T X y > 0$ $\quad \forall\, y \neq \underline{0}$
* All eigenvalues of X, Z are real & +ve.
* $Z^{1/2} Z^{1/2} = Z$

$(Z^{1/2})^{-1} = Z^{-1/2}$

$$\frac{|f(Z) - f(X) - \langle D, Z-X \rangle|}{\|Z-X\|} \to 0$$

$$f(Z) = f(X) + \langle D, Z-X \rangle + \underbrace{o(\|Z-X\|)}_{\text{decays faster than } Z-X}$$

$$= f(X) + \langle D, \Delta X \rangle + o(\|\Delta X\|)$$

$$f(Z) = f(X + \Delta X) = \log\det(X + \Delta X)$$

$$= \log\det\left(X^{1/2} X^{1/2} + \overset{X^{1/2}X^{-1/2}}{\wedge}\,\Delta X\, \overset{X^{-1/2}X^{1/2}}{\wedge}\right)$$

$$= \log\det\left[X^{1/2}\left(I + X^{-1/2} \Delta X X^{-1/2}\right) \times X^{1/2}\right]$$

$$X + \Delta X = X^{1/2} X^{1/2} + X^{1/2} X^{-1/2} \Delta X \, X^{-1/2} X^{1/2}$$

$$= X^{1/2} \left(I \, X^{1/2} + X^{-1/2} \Delta X \, X^{-1/2} X^{1/2} \right)$$

$$= X^{1/2} \left(I + X^{-1/2} \Delta X \, X^{-1/2} \right) X^{1/2}$$

$$f(Z) = \log \left[\left(\det X^{1/2} \right) \times \det \left(I + X^{-1/2} \Delta X \, X^{-1/2} \right) \times \det \left(X^{1/2} \right) \right]$$

$$= \log \left[\det(X) \times \det \left(I + X^{-1/2} \Delta X \, X^{-1/2} \right) \right]$$

$$= f(X) + \log \det \left(I + X^{-1/2} \Delta X \, X^{-1/2} \right)$$

Suppose $\lambda_1, \lambda_2 - - \lambda_n$ are eigenvalues of $X^{-1/2} \Delta X X^{-1/2}$

$$f(Z) = f(X) + \log \prod_{i=1}^{n} (1+\lambda_i)$$

$$= f(X) + \sum_{i=1}^{n} \log(1+\lambda_i)$$

$$= f(x) + \sum_{i=1}^{n} \left(\lambda_i + o(\lambda_i) \right)$$

$$= f(X) + \sum_{i=1}^{n} \lambda_i + o\left(\sum_{i=1}^{n} \lambda_i\right)$$

$$= f(X) + \mathrm{Tr}\left(X^{-1/2} \Delta X X^{-1/2}\right) + o\left(\mathrm{Tr}(X^{-1/2} \Delta X X^{-1/2})\right)$$

$$= f(X) + \mathrm{Tr}\left(X^{-1/2} X^{-1/2} \Delta X\right) + o\left(\mathrm{Tr}\left(X^{-1/2} \Delta X\, X^{-1/2}\right)\right)$$

$$= f(X) + \mathrm{Tr}\left(X^{-1} \Delta X\right) + o\left(\mathrm{Tr}\left(X^{-1} \Delta X\right)\right)$$

$$f(Z) = f(X) + \langle X^{-1}, \Delta X \rangle + o\left(\mathrm{Tr}\left(X^{-1} \Delta X\right)\right)$$

$$D_{f(X)} = X^{-1}$$

Chain rule for gradients

Suppose $\quad h(\underline{x}) = g(f(\underline{x}))$

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

$$g : \mathbb{R}^m \to \mathbb{R}^k$$

$$D_{h(\underline{x})} = D_{g(f(\underline{x}))} \; D_{f(\underline{x})}$$

$$D_h(\underline{x}) = D_g(f(\underline{x})) \; D_f(\underline{x})$$

$$f(\underline{x}) = \| A\underline{x} + \underline{b} \|_2^2 = g(f(\underline{x}))$$

$$g = \| \cdot \|_2^2 , \quad f = Ax + b.$$

$$D_f(\underline{x}) = D_g(f(x)) \times D_f(x)$$

$$= 2 \times (A\underline{x} + b)^T A^T$$

$$= 2 (A\underline{x} + b)^T A^T$$

$$\nabla f(x) = 2A(Ax + b)$$

Second derivative and hessian

$$D^2_{f(x)} = D_{D_{f(x)}}$$

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

$$D_{f(x)} \in \mathbb{R}^{1\times n}$$

$$D_f : \mathbb{R}^n \to \mathbb{R}^n$$

$$D^2_f : \mathbb{R}^n \to \mathbb{R}^{n\times n}$$

$$D_f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

$$D_f^2(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Transpose of this matrix is called the Hessian

$\nabla^2 f$

Eg:

$$f(\underline{x}) = \underline{x}^T A \underline{x} + \underline{b}$$

$$D_f(\underline{x}) = (A + A^T)\,\underline{x}$$

$$D_f^2(\underline{x}) = A + A^T$$

Review of Linear algebra

Vector space $(\mathbb{V}, +, \cdot)$

① $\underline{x} + \underline{y} = \underline{y} + \underline{x} \quad \forall \underline{x}, \underline{y} \in \mathbb{V}$

② $\underline{x} + (\underline{y} + \underline{z}) = (\underline{x} + \underline{y}) + \underline{z}$

③ $\exists\ \underline{0} \in \mathbb{V}$ s.t. $\underline{0} + \underline{x} = \underline{x} \quad \forall \underline{x} \in \mathbb{V}$.

④ For each $\underline{x}$, $\exists\ (-\underline{x})$ s.t. $\underline{x} + (-\underline{x}) = \underline{0}$

⑤ $\alpha(\beta \underline{x}) = (\alpha\beta)\underline{x} \quad \forall\ \alpha, \beta \in \mathbb{R}$ & $\underline{x} \in \mathbb{V}$

⑥ $\alpha(\underline{v}_1 + \underline{v}_2) = \alpha \underline{v}_1 + \alpha \underline{v}_2$

$(\alpha + \beta)\underline{v}_1 = \alpha \underline{v}_1 + \beta \underline{v}_1$

⑦ $1 \cdot \underline{v} = \underline{v}$

Eg : $\mathbb{R}^n$, $\mathbb{R}^k$, $\mathbb{R}^{n \times k}$ } All vector spaces

$\mathbb{R}^{n \times k} \longrightarrow$ dim =

$\mathbb{Q}^k$ Not a vector space over $\mathbb{R}$

S_+^n : set of all $n \times n$ symmetric PSD matrices

A, B PSD

$x^T A x \geq 0$ $x^T B x \geq 0$ $\forall x$

$x^T (A+B) x \geq 0$ $\forall x$

Not a vector space

S^n : Set of $n \times n$ symmetric matrices