Math Background Review

Basic Analysis:

A. Norm: A fn $f: \mathbb{R}^n \to \mathbb{R}$ is called norm if:

* (non-neg.): $f(\underline{x}) \geq 0$, $\forall \underline{x} \in \mathbb{R}^n$, $f(\underline{x}) = 0$ iff $\underline{x} = \underline{0}$
* (homogeneity): $f(t\underline{x}) = |t| f(\underline{x})$, $\forall \underline{x} \in \mathbb{R}^n$, $t \in \mathbb{R}$.
* (triangle ineq.): $f(\underline{x}+\underline{y}) \leq f(\underline{x}) + f(\underline{y})$, $\forall \underline{x}, \underline{y} \in \mathbb{R}^n$

If $f(\underline{x})$ is a norm, denote it as $\|\underline{x}\|$.

2. Norm $\|\underline{x}\|$'s meaning:

* $\|\underline{x}\|$: length of $\underline{x}$.
* $\|\underline{x}-\underline{y}\|$: dist. btwn $\underline{x}$ & $\underline{y}$.

3. Unit ball: Set of vectors with $\|\underline{x}\| \leq 1$.

$$B = \{\underline{x} \in \mathbb{R}^n : \|\underline{x}\| \leq 1\}.$$

Ex: * l_2-norm (Euclidean Norm): $\|\underline{x}\|_2 \triangleq (\underline{x}^T\underline{x})^{\frac{1}{2}} = (x_1^2 + \cdots + x_n^2)^{\frac{1}{2}}$.

* l_1-norm (sum-abs-val.): $\|\underline{x}\|_1 \triangleq |x_1| + \cdots + |x_n|$ (Manhaten dist.)

* l_∞-norm (chebyshev): $\|\underline{x}\|_\infty \triangleq \max\{|x_1|, \ldots, |x_n|\}$.

* l_p-norm: $\|\underline{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}}$ Q: $\|\underline{x}\|_\infty \overset{?}{=} \lim_{p\to\infty} \|\underline{x}\|_p$

Prof. $\|\underline{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}} = \left(\frac{|x_1|^p}{\|\underline{x}\|_\infty^p} + \cdots + \frac{|x_n|^p}{\|\underline{x}\|_\infty^p}\right)^{\frac{1}{p}} \|\underline{x}\|_\infty$

$\leq (1 + \cdots + 1)^{\frac{1}{p}} \|\underline{x}\|_\infty = \sqrt[p]{n}\,\|\underline{x}\|_\infty$

Let $i^* \in \arg\max_{\forall i} \{|x_i|\}$

$\geq (|x_{i^*}|^p)^{\frac{1}{p}} = |x_{i^*}| = \|\underline{x}\|_\infty$

Let $p \to \infty$, $\sqrt[p]{n} \to 1$ (squeez thm) ∎

Equivalence of Norm:

Suppose $\|\cdot\|_a$ and $\|\cdot\|_b$ are norms on $\mathbb{R}^n$, then $\exists\, \alpha, \beta > 0$

s.t. $\forall\, \underline{x} \in \mathbb{R}^n$, $\alpha \|\underline{x}\|_a \le \|\underline{x}\|_b \le \beta \|\underline{x}\|_a$.

Ex: $\|\underline{x}\|_2 \le \|\underline{x}\|_1 \le \sqrt{n}\, \|\underline{x}\|_2$.

$\|\underline{x}\|_\infty \le \|\underline{x}\|_2 \le \sqrt{n}\, \|\underline{x}\|_\infty$

$\|\underline{x}\|_\infty \le \|\underline{x}\|_1 \le n \|\underline{x}\|_\infty$

4. Convergen Sequence & Limits

1° Def (Convergence): A seq. of vectors $\underline{x}_1, \underline{x}_2 \cdots$ are said to be convergent to a limit pt. $\bar{\underline{x}}$, if $\forall \varepsilon > 0$, $\exists\, N_\varepsilon \in \mathbb{N}$

s.t. $\|\underline{x}_k - \bar{\underline{x}}\| < \varepsilon$, $\forall\, k \ge N_\varepsilon$ ($\{\underline{x}_k\} \to \bar{\underline{x}}$ as $k \to \infty$, $\lim_{k\to\infty} \underline{x}_k = \bar{\underline{x}}$).

2° Def (Cauchy Seq.): A seq. $\{\underline{x}_k\}$ is Cauchy if

$\forall\, \varepsilon > 0$, $\exists\, N \in \mathbb{N}$, s.t. $\|\underline{x}_m - \underline{x}_n\| < \varepsilon$, $\forall\, m, n > N$

Thm: A seq. in $\mathbb{R}^n$ has a limit iff it's Cauchy.

Ex: (p-series): $a_n = \frac{1}{n^p}$. Show $\{b_n\} = \left\{\sum_{k=1}^{n} a_k\right\}$ has a limit for $p=2$.

Also, $\{b_n\}$ doesn't converge for $p=1$.

Proof. w.l.o.g let $m, n \in \mathbb{N}$ and $m < n$.

1. $p=2$: $b_n - b_m = \sum_{k=1}^{n} \frac{1}{k^2} - \sum_{k=1}^{m} \frac{1}{k^2} = \sum_{k=m+1}^{n} \frac{1}{k^2} < \sum_{k=m+1}^{n} \frac{1}{k(k-1)}$

$= \sum_{k=m+1}^{n} \left(\frac{1}{k-1} - \frac{1}{k}\right) = \frac{1}{m} - \frac{1}{m+1} + \frac{1}{m+1} - \frac{1}{m+2} \cdots - \frac{1}{n} = \frac{1}{m} - \frac{1}{n}$

$< \frac{1}{m} < \varepsilon$ I can always find suff. large m s.t. $b_n - b_m < \varepsilon$

2. $(p=1)$ ← harmonic series

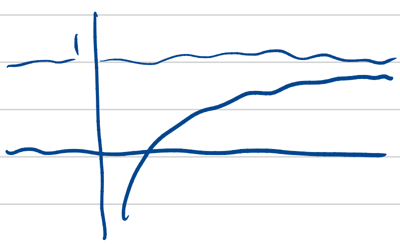
$b_n - b_m = \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=1}^{m} \frac{1}{k} = \sum_{k=m+1}^{n} \frac{1}{k} = \underbrace{\frac{1}{m+1}}_{>\frac{1}{n}} + \cdots + \underbrace{\frac{1}{n}}_{\geq\frac{1}{n}}$

$> \frac{n-m}{n} = 1 - \frac{m}{n}$

$\Rightarrow$ For any $\varepsilon > 0$, for any m (no matter how large m is).

can choose $n \geq \left\lceil \frac{m}{1-\varepsilon} \right\rceil$, s.t. $|b_n - b_m| > \varepsilon$.

5. Supremum of S (least UB): Smallest possible α : $\alpha \geq x, \forall \in S$.



$1 - \frac{1}{x}, \forall x > 0$

Infimum of S. (largest LB): largest possible value $\alpha \leq x, \forall x \in S$.

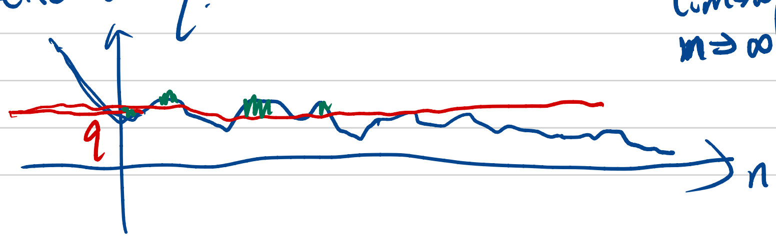
6. Maximum, Minimum: (achievable).

* The limit supremum $\limsup_{k\to\infty} x_k$ is the infimum of all

$q \in \mathbb{R}$ for which all but a finite # of elements in $\{x_k\}$

exceed q:

$\limsup_{n\to\infty} x_n \triangleq \lim_{n\to\infty} \{\sup_{m\geq n} x_m\}$



* The limit infimum $\liminf_{k\to\infty} x_k$ is the supremum of all $q \in \mathbb{R}$ for which all but a finite # of elements in $\{x_k\}$ less than q. $\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \{\inf_{m\geq n} x_m\}$

* limsup & liminf <u>always</u> exist.

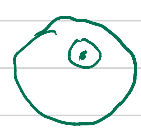
* $\{x_n\}$ converge iff $\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n$.

3. Functions.

1° <u>Cont. fn:</u> A fn $f: S \to \mathbb{R}$ is cont. a $\bar{x} \in S$ if $\forall \varepsilon > 0$, $\exists$ a $\delta > 0$, s.t. $\forall x \in S$, with $\|x - \bar{x}\| < \delta \to |f(x) - f(\bar{x})| < \varepsilon$.

write: $f(x) \to f(\bar{x})$, as $x \to \bar{x}$.

<u>Fact:</u> Cont. fn achieves both maximum & minimum over a non-empty compact set.
closed & bounded



2. Diff'ble fn:

a) S non-empty set in $\mathbb{R}^n$, $\bar{x} \in \text{int } S$. Given $f: S \to \mathbb{R}$.
f is diff'ble at $\bar{x}$ if $\exists$ a vector (called gradient)

$$\nabla f(\bar{x}) \triangleq \left[\frac{\partial f(\bar{x})}{\partial x_1} \cdots \frac{\partial f(\bar{x})}{\partial x_n}\right]^T$$

at $\bar{x}$ and $\exists$ fn $\beta(x, \bar{x}) \to 0$ as $x \to \bar{x}$, s.t.

$$f(x) = \underbrace{f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})}_{\text{FO - approx. (linear approx)}} + \underbrace{\|x - \bar{x}\| \beta(x, \bar{x})}_{o(\|x - \bar{x}\|)}, \quad \forall x \in S.$$

(2) f is called twice diff'ble at $\bar{x}$ if, in addition to grad, $\exists$ a symmetric $n\times n$ matrix $\underline{H}(\bar{x})$ (called Hessian matrix) of f at $\bar{x}$, and $\beta(\underline{x},\bar{x}) \to 0$ as $\underline{x}\to\bar{x}$, s.t.

$$f(\underline{x}) = \underbrace{f(\bar{x}) + \nabla f(\bar{x})^T(\underline{x}-\bar{x}) + \frac{1}{2}(\underline{x}-\bar{x})^T \underline{H}(\bar{x})(\underline{x}-\bar{x})}_{\text{SO - approx.}} + \underbrace{\|\underline{x}-\bar{x}\|^2\beta(\underline{x},\bar{x})}_{o(\|\underline{x}-\bar{x}\|^2)}$$

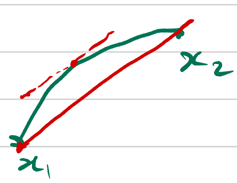
$$\underline{H}(\underline{x}) \triangleq \begin{bmatrix} \frac{\partial^2 f(\underline{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\underline{x})}{\partial x_1 \partial x_n} \\ \vdots & & \\ \frac{\partial^2 f(\underline{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\underline{x})}{\partial x_n^2} \end{bmatrix}$$

3° A vector-valued fn f is diff'ble (twice) if each component is diff'ble (twice).

A diff'ble vector-valued fn: $h: \mathbb{R}^m \to \mathbb{R}^n$, The Jacobian $\underline{J}(\underline{x}) = \nabla \underline{h}(\underline{x})$ is a $n\times m$ matrix:

$$\underline{J}(\underline{x}) = \nabla h(\underline{x}) = \begin{bmatrix} \nabla h_1(\underline{x})^T \\ \vdots \\ \nabla h_n(\underline{x})^T \end{bmatrix}_{n\times m}.$$

Hessian is a special case of Jacobian.



4° (MVT): S nonempty open convex set in $\mathbb{R}^n$. Let $f: S\to\mathbb{R}$ be diff'ble. For every $x_1, x_2 \in S$, we have

$$f(x_2) = f(x_1) + \nabla f(\underline{x})^T(x_2 - x_1), \text{ where } \underline{x} = \lambda x_1 + (1-\lambda)x_2$$

for some $\lambda_2 \in (0,1)$.

5° Taylor's Thm: S non-empty, open, convex in $\mathbb{R}^n$.
$f: S \to \mathbb{R}$, twice diff'ble. For every $x_1, x_2 \in S$, we have:
$$f(\underline{x}_2) = f(\underline{x}_1) + \nabla f(\underline{x}_1)^T(\underline{x}_2 - \underline{x}_1) + \frac{1}{2}(\underline{x}_2 - \underline{x}_1)^T \underline{\underline{H}}(\underline{x})(\underline{x}_2 - \underline{x}_1).$$
Hessian

Linear Algebra:

1. Linear indep: $\underline{x}_1, \dots, \underline{x}_k \in \mathbb{R}^n$ are lin. indep. if
$$\sum_{i=1}^{k} \lambda_i x_i = 0 \Rightarrow \lambda_i = 0, \forall i = 1, \dots, k.$$

2. Linear comb: $y \in \mathbb{R}^n$ is lin. comb. of $\underline{x}_1 \dots \underline{x}_k \in \mathbb{R}^n$ if
$$y = \sum_{i=1}^{k} \lambda_i \underline{x}_i \text{ for some } \lambda_1 \dots \lambda_k \in \mathbb{R}.$$
   * $\sum_{i=1}^{k} \lambda_i = 1$: y is an affine comb. of $\underline{x}_1 \dots \underline{x}_k$.
   * $\sum_{i=1}^{k} \lambda_i = 1$, $\lambda_i \geq 0, \forall i$: y is a convex comb. of $\underline{x}_1 \dots \underline{x}_k$

The linear, affine, convex hulls of $S \subseteq \mathbb{R}^n$ are, resp. the sets of all lin. affine, convex comb. of pts. in S.

3. Spanning vectors: $\underline{x}_1 \dots \underline{x}_k \in \mathbb{R}^n$, $k \geq n$ are said to be spanning $\mathbb{R}^n$ if any vector in $\mathbb{R}^n$ can be represented as a lin. comb. of $\underline{x}_1, \dots, \underline{x}_k$.

The cone spanned by $\underline{x}_1, \dots, \underline{x}_k$ is set of non-neg. lin. comb.

4. Basis: A minimal set of $\underline{x}_1, \ldots, \underline{x}_k \in \mathbb{R}^n$ spans $\mathbb{R}^n$: if the deletion of any of $\underline{x}_1, \ldots, \underline{x}_k$ prevents remaining vectors from spanning $\mathbb{R}^n$. (Basis $\underline{x}_1 \cdots \underline{x}_k$ spans $\mathbb{R}^n$ iff $k=n$).

5. Cauchy-Schwartz Ineq: $|\langle \underline{x}, \underline{y} \rangle| = |\underline{x}^T \underline{y}| \leq \|\underline{x}\|_2 \cdot \|\underline{y}\|_2$.

(with eq. achieved iff $\underline{x}, \underline{y}$ are lin. dep.)

(unsigned angle btwn $\underline{x}, \underline{y} \in \mathbb{R}^n$.

$$\angle(\underline{x}, \underline{y}) \triangleq \cos^{-1}\left(\frac{\underline{x}^T \underline{y}}{\|\underline{x}\|_2 \cdot \|\underline{y}\|_2}\right) \in [0, \pi]$$

(cos. sim.)

$\underline{x}$ and $\underline{y}$ are othogonal, i.e., $\underline{x} \perp \underline{y}$ if $\langle \underline{x}, \underline{y} \rangle = 0$.

Young's Ineq: For $a \geq 0$, $b > 0$, and any $p, q > 0$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$ (conjugate) we have $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$, with eq. achieved iff $a^p = b^q$.

(special cases: $p=q=2$, p (or q) $=1$, q (or p) $= \infty$)

Holder's Ineq: For any pair of vec. $\underline{x}, \underline{y} \in \mathbb{R}^n$, and for p, q s.t. $\frac{1}{p} + \frac{1}{q} = 1$, we have:

$$\sum_{i=1}^{n} |x_i y_i| \leq \|\underline{x}\|_p \cdot \|\underline{y}\|_q$$

$= \|\underline{x} \odot \underline{y}\|_1$ (Holder's $\Rightarrow$ Cauchy-Schwartz)

6. Othogonal matrix: $\underline{Q} \in \mathbb{R}^{m \times n}$. $\underline{Q}^T \underline{Q} = \underline{I}_n$ $(m \geq n)$ or $\underline{Q}\underline{Q}^T = \underline{I}_m$ $(n \geq m)$

If $\underline{Q}$ is square: $\underline{Q}^{-1} = \underline{Q}^T$.

7. Rank of matrix: For $\underline{A} \in \mathbb{R}^{m \times n}$, $\text{rank}(\underline{A})$ = max # of lin. indep. rows (or equivalently, cols) of $\underline{A}$.

If $\text{rank}(\underline{A}) = \min\{m, n\}$, $\underline{A}$ is full row/col rank.

8. Eigenvalues and Eigenvectors: $\underline{A} \in \mathbb{R}^{n \times n}$. If λ and $\underline{x} \neq 0$ satisfy: $\underline{A}\underline{x} = \lambda \underline{x}$, then λ, $\underline{x}$ are eigenvalues & eigenvector.

* λ can be computed by solving $\det(\underline{A} - \lambda \underline{I}) = 0$.

* A is symmetric $\Rightarrow$ n (possibly non-distinct) real eigenvalues
multiplicity.

* Eigenvectors assoc. with distinct eigenvalues are othogonal.

* Given symmetric $\underline{A}$ $\Rightarrow$ can construct basis $\underline{B} \in \mathbb{R}^{n \times r}$. where each col in $\underline{B}$ is an eigenvector of $\underline{A}$

* Normalize $\underline{B}$ to have unit l_2-norm: s.t. $\underline{B}^T \underline{B} = \underline{I}$ ($\underline{B}^T = \underline{B}^{-1}$).

Then $\underline{B}$ is called othonormal matrix.

Eigen-decomp: $\underline{A} = \underline{B} \underline{\Lambda} \underline{B}^T$.

9. Singular-Value Decomp. (SVD):

Let $\underline{A} \in \mathbb{R}^{m \times n}$. Then $\underline{A} = \underline{U} \underline{\Sigma} \underline{V}^T$, where $\underline{U} \in \mathbb{R}^{m \times m}$ othonormal, $\underline{V} \in \mathbb{R}^{n \times n}$ othonormal, $\underline{\Sigma} \in \mathbb{R}^{m \times n}$, $(\underline{\Sigma})_{ij} = 0$, for $\forall i \neq j$.

$(\underline{\Sigma})_{ii} \geq 0$
$\in \mathbb{R}$.

* Cols of $\underline{\underline{U}}$: Normalized eigenvectors of $\underline{\underline{A}}\,\underline{\underline{A}}^T$.

* —— $\underline{\underline{V}}$: - - ~ —— —— ~ of $\underline{\underline{A}}^T\underline{\underline{A}}$

* $(\underline{\underline{\Sigma}})_{ii}$: Abs. square root of eigenvalues of $\underline{\underline{A}}\,\underline{\underline{A}}^T$ if $m \le n$, or $\underline{\underline{A}}^T\underline{\underline{A}}$ if $m \ge n$.

12. Definite & Semidefinite Matrices : $\underline{\underline{A}} \in \mathbb{R}^{n\times n}$ symmetric.

$\underline{\underline{A}}$ is PD if $\underline{x}^T\underline{\underline{A}}\,\underline{x} > 0$, $\forall\, \underline{x} \neq 0,\ \underline{x} \in \mathbb{R}^n$.

PSD: ≥ 0, $\forall\, \underline{x} \in \mathbb{R}^n$

ND: < 0, $\forall\, \underline{x} \neq 0,\ \underline{x} \in \mathbb{R}^n$.

NSD: ≤ 0, $\forall\, \underline{x} \in \mathbb{R}^n$

$\underline{\underline{A}}$ is indef. if neither PSD nor NSD.

$\underline{\underline{A}}$ is PD / PSD / ND / NSD if eigenvalues are pos. / non-neg. / neg. / non-pos.

If $\underline{\underline{A}}$ is PSD, then $\underline{\underline{A}}^{\frac{1}{2}}$ is the matrix satisfying

$$\underline{\underline{A}}^{\frac{1}{2}}\underline{\underline{A}}^{\frac{1}{2}} = \underline{\underline{A}} \quad \text{and} \quad \underline{\underline{A}}^{\frac{1}{2}} = \underline{\underline{B}}\,\underline{\underline{\Lambda}}^{\frac{1}{2}}\underline{\underline{B}}^T$$

$\uparrow \begin{bmatrix} \lambda_1^{\frac{1}{2}} & & \\ & \ddots & \\ & & \lambda_n^{\frac{1}{2}} \end{bmatrix}$