# I. BACKGROUND

- Vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [\, x_1, x_2, \cdots, x_n \,]^T$$

The following notation may also be used

$$x = (x_1, x_2, \cdots, x_n)$$

where we used parentheses to construct col. vectors.
(Sometimes it helps; e.g., $(x,t) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ t \end{bmatrix}$)

- Matrix

$$X = \begin{bmatrix} \ddots & \vdots & \\ \cdots & X_{ij} & \\ & \vdots & \ddots \end{bmatrix}.$$

- Set of Real nos.

$\mathbb{R}$ — set of real nos. (~~does~~ $\infty$ too, $-\infty$ included in $\mathbb{R}$?)

$\mathbb{R}_+$ — set of nonnegative nos.

$\mathbb{R}_{++}$ — set of positive nos.

$\mathbb{R}^n$ — set of real n-vectors

$\mathbb{R}^{m\times n}$ — set of real $m\times n$ matrices

$S^k$ — set of $k\times k$ real symmetric matrices

$$S^k = \{ X \in \mathbb{R}^{k\times k} \mid X = X^T \}$$

$\mathbb{R} \cup \{\pm\infty\}$ — set of extended real nos.

- Inner Product, Euclidean norm, & Angle

- The standard inner product on $\mathbb{R}^n$:

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i, \quad x, y \in \mathbb{R}^n.$$

- The Euclidean norm, or the 2-norm

$$\|x\|_2 = (x^T x)^{1/2}$$

- Cauchy Schwartz inequality:

$$|x^T y| \leq \|x\|_2 \|y\|_2 .$$

- The (unsigned) angle between nonzero vectors $x, y \in \mathbb{R}^n$

$$\angle(x, y) = \cos^{-1}\left( \frac{x^T y}{\|x\|_2 \|y\|_2} \right)$$

Note $\angle(x, y) \in [0, \pi]$. The vectors $x$ & $y$ are orthogonal if $x^T y = 0$, in that case $\angle(x, y) = \pi$.

- The standard inner product on $\mathbb{R}^{m \times n}$:

$$\langle X, Y \rangle = tr(X^T Y) = \sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij} Y_{ij} .$$

- The Frobenius norm:

$$\|X\|_F = \left( tr(X^T X) \right)^{1/2}$$

• Norms, distance, & norm ball.

- A function $f: \mathbb{R}^n \to \mathbb{R}$ with $\text{dom} f = \mathbb{R}^n$ is a norm if
  1. $f(x) \geq 0 \quad \forall \; x \in \mathbb{R}^n$ (non-ve)
  2. $f(x) = 0 \iff x = 0$ (definite)
  3. $f(tx) = |t| f(x) \quad \forall \; x \in \mathbb{R}^n \; \& \; t \in \mathbb{R}$ (homogeneous)
  4. $f(x+y) \leq f(x) + f(y), \; \forall \; x, y \in \mathbb{R}^n$ ($\Delta$ inequality)

The notation $f(x) = \|x\|$ is often used to represent a norm. Norm is a measure of length of a vector.

- Distance between $x, y$:

$$\text{dist}(x,y) = \|x-y\|.$$

- A norm ball

$$\mathcal{B}(x,r) = \{ y \in \mathbb{R}^n \mid \|y-x\| \leq r \}$$
$$= \{ x+ru \in \mathbb{R}^n \mid \|u\| \leq 1 \}$$

← mention it later in analysis.

- Some examples of norm:

$p$-norm : $\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}$.

1-norm : $\|x\|_1 = \sum_{i=1}^n |x_i|$

$\infty$-norm : $\|x\|_\infty = \lim_{p\to\infty} \|x\|_p = \max\{|x_1|, \dots, |x_p|\}$.

the $P$-quadratic norm: $\|x\|_P = (x^T P x)^{1/2} = \|P^{1/2} x\|_2$

where $P$ is +ve ~~semi~~definite.

- Equivalence of norms : Suppose $\|x\|_a$ and $\|x\|_b$ are norms on $\mathbb{R}^n$.

$$\alpha\|x\|_a \leq \|x\|_b \leq \beta\|x\|_a.$$

- Operator Norms : (May talk about it after Analysis, & Matrix Theory)

$$\|X\|_{a,b} = \sup\{\|Xu\|_a \mid \|u\|_b \leq 1\}$$

$$\|X\|_2 = \sup\{\|Xu\|_2 \mid \|x\|_2 \leq 1\}$$

$$= \sigma_{max}(X) = (\lambda_{max}(X^T X))^{1/2}.$$

$$\|X\|_\infty = \max_{i=1,\dots,m} \sum_{j=1}^{n} |X_{ij}| \quad \text{(max-row-sum)}$$

$$\|X\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^{m} |X_{ij}| \quad \text{(max-col.-sum)}$$

- Dual Norm :

  - Let $\|.\|$ be a norm on $\mathbb{R}^n$.

$$\|z\|_* = \sup\{z^T x \mid \|x\| \leq 1\}$$

  - The dual norm can be shown to be a norm, as well.

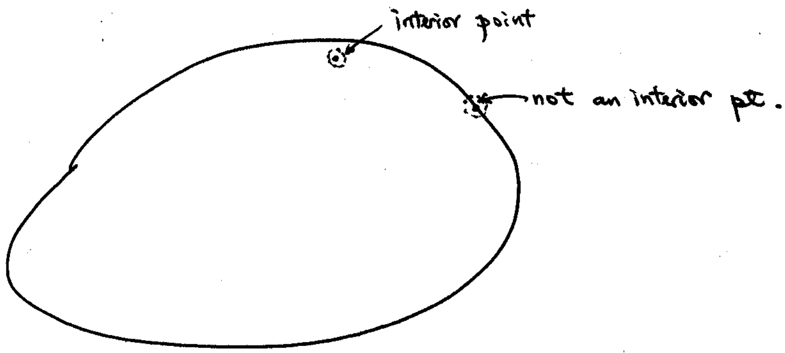
- Analysis (it helps to talk analysis before operator norms)

- Open & Closed Sets

- An element $x \in C \subseteq \mathbb{R}^n$ is an interior point of $C$ if there exists an $\epsilon > 0$ for which

$$B(x, \epsilon) \subseteq C$$

or $\{y \mid \|y - x\|_2 \le \epsilon\} \subseteq C$




- int $C$ – the set of all interior pts. of $C$, called the interior of $C$.

- $C$ is open if int $C = C$.

e.g., $C = (a, b)$ is open.

$C = \{1, 2, 3\}$ is not open.

$C = (a, b]$ is not open.

- note: A set that is not open is NOT called closed.

- $C$ is closed if its complement $\mathbb{R}^n \setminus C = \{x \in \mathbb{R}^n \mid x \notin C\}$ is open.

e.g., $C = [a, b]$.
$\mathbb{R}^n \setminus C = (-\infty, a) \cup (b, \infty)$.
Its complement is open so $C$ is closed.

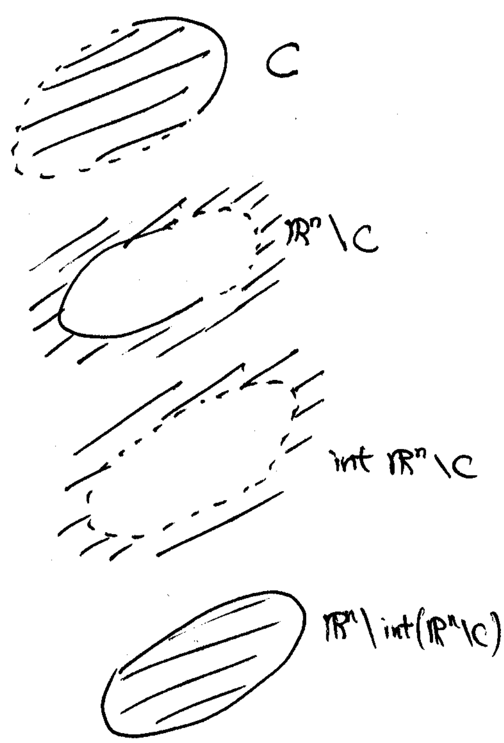
e.g., $C = (a, b]$ is not closed.
$C = \{1, 2, 3\}$ is closed (await verification)

e.g., $C = \emptyset$ is open & closed.
$\because$ $\operatorname{int} C = \emptyset \Rightarrow \operatorname{int} C = C$
$\because$ $\mathbb{R}^n \setminus C = \mathbb{R}^n \Rightarrow \operatorname{int} \mathbb{R}^n = \mathbb{R}^n \Rightarrow \mathbb{R}^n$ is open.

e.g., likewise $C = \mathbb{R}^n$ is open & closed.

klouzə'
- Closure of $C$

$$\operatorname{cl} C = \mathbb{R}^n \setminus \operatorname{int}(\mathbb{R}^n \setminus C)$$




A pt. $x$ is in the closure of $C$ if for every $\epsilon > 0$, there is a $y \in C$ with $\|x - y\|_2 \leq \epsilon$.

Sep. 9, 2008:

$x \in \operatorname{cl} C \Leftarrow$ For every $\epsilon > 0$,
$\exists\, y \in C$ s.t. $\|x - y\|_2 \leq \epsilon$.

Trivial for $x \in \operatorname{int} C$. Meaningful when $x$ is on the boundary.

- boundary of $C$

$$\operatorname{bd} C = \operatorname{cl} C \setminus \operatorname{int} C.$$

A boundary pt. $x \in C$ has the property that for every $\epsilon > 0$, there exist $y \in C$ and $z \notin C$ with

$$\|y - x\|_2 \leq \epsilon, \quad \|z - y\|_2 \leq \epsilon.$$

- $C$ is closed if $\operatorname{bd} C \subseteq C$.
- $C$ is open if $C \cap \operatorname{bd} C = \emptyset$.

• Supremum and Infimum

- Suppose $C \subseteq \mathbb{R}$.
- A no. $a$ is an upper bound on $C$ if for each $x \in C$, $x \leq a$.
  Note that $a \in C$ is not necessary.
- An ub on $C$ is not necessarily existent; e.g., for $C = \mathbb{R}$ there is no ub on $C$.
  In this case $C$ is unbounded above.
- When $C = \emptyset$, the set of ub on $C$ is $\mathbb{R}^n$.
- A no. $b$ is the least ub, or the supremum of $C$ if
  1. $b$ is an ub on $C$.
  2. $\cancel{a \leq b}$ $b \leq a$ for every ub $a$ on $C$.

- (In some literature supremum is not defined for $C = \emptyset$ or $C$ that is unbounded above).

We take

$$\sup C = +\infty, \quad \text{if } C \text{ is unbounded above}$$
$$\sup C = -\infty, \quad \text{if } C = \emptyset.$$

- We define lower bound in a similar manner.
- A no. $b$ is the greatest lb or infimum if
  1. $b$ is a lb on $C$.
  2. $a \leq b$ for each lb $a$ on $C$.

• FUNCTIONS

$$f: A \to B$$

- $\mathrm{dom}\, f \subseteq A$ is the set (domain) of proper inputs.
- In ~~general~~ many cases we look into $f: \mathbb{R}^n \to \mathbb{R}$.

Sometimes,
$$f: \mathbb{R}^n \to \mathbb{R}^m.$$
And in some cases,
$$f: \mathbb{R}^{n \times n} \to \mathbb{R}.$$

e.g., $f(x) = \log x$. We have $\mathrm{dom}\, f = \mathbb{R}_{++}$.

e.g., $f: S^n \to \mathbb{R}$

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

It is appropriate that $\mathrm{dom}\, f = S^n_{++}$. (set of PD matrices)

- Continuity

A function $f: \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $x \in \mathrm{dom}\, f$ if for all $\epsilon > 0$ there exists a $\delta > 0$ such that

$$y \in \mathrm{dom}\, f \cap B(x, \underset{\sim}{\delta}) \Rightarrow \| f(y) - f(x) \|_2 \le \epsilon .$$

to be precise
$\delta(\epsilon)$

sign fn. think about it!

Continuity can be described ~~as~~ in terms of limits:
whenever $X = \{x_1, x_2, \cdots\}$, $x_i \in \mathrm{dom}\, f$ converges to $x \in \mathrm{dom}\, f$, the sequence $f(x_1), f(x_2), \cdots$ converges to $f(x)$:

$$\lim_{i \to \infty} f(x_i) = f(x), \quad x = \lim_{i \to \infty} x_i$$

(very similar to limit of a function, isn't it?)

- A function is continuous if it is continuous at every pt. in its domain.

- Boundness (await verification)

$C \subseteq \mathbb{R}^n$ is bounded if there exist $r \in \mathbb{R}$ & $x_c \in \mathbb{R}^n$, such that

$$C \subseteq B(x_c, r).$$

# Derivatives & Gradient

- Suppose $f: \mathbb{R}^n \to \mathbb{R}$.

- Gradient

or assume f is open.

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}, \qquad x \in \text{int dom} f.$$

Recall taylor series for $f: \mathbb{R} \to \mathbb{R}$:

$$f(z) = f(x) + \frac{df}{dx}(z-x) + \frac{1}{2}\frac{d^2 f}{dx^2}(z-x)^2 + \frac{1}{3!}\frac{d^3 f}{dx^3}(z-x)^3$$
$$\vdots$$

- First order approximation:

$$f(z) \approx f(x) + \nabla^T f(\cancel{z}x)(z-x)$$

$$\lim_{\substack{z \in \text{dom} f, \\ z \neq x \\ z \to x}} \frac{\| f(z) - f(x) - \nabla^T f(x)(z-x) \|_2}{\| z - x \|_2} = 0.$$

Compare the derivative of a scalar function

$$\lim_{\substack{z \in \text{dom} f \\ z \to x}} \frac{f(z) - f(x)}{z - x} = \frac{df}{dx}$$

e.g. $f(x) = \frac{1}{2} x^T P x + q^T x + r$

$\nabla f(x) = Px + q.$

( if have time, consider $f(X) = \log \det X$

$\nabla f(X) = X^{-1}$ )

- note that derivative/gradient doesn't always exist; e.g. $f(x) = |x|$.

• Second derivative or Hessian matrix of $f$ at $x \in \mathrm{int\,dom} f$, denoted by $\nabla^2 f(x) \in \mathbb{R}^{n \times n}$, is given by

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

2nd order approximation of $f$

$$\hat{f}(z) = f(x) + \nabla f(x)^T (z-x) + \frac{1}{2}(z-x)^T \nabla^2 f(x)(z-x).$$

It satisfies

$$\lim_{\substack{z \in \mathrm{dom} f \\ z \neq x \\ z \to x}} \frac{|f(z) - \hat{f}(z)|}{\|z-x\|_2^2} = 0.$$

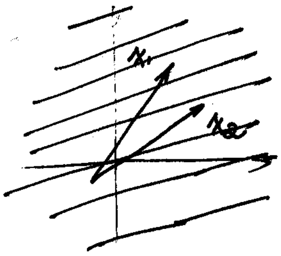
e.g., $f(x) = \frac{1}{2} x^T P x + q^T x + r$

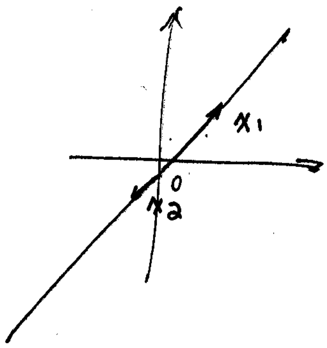
$\nabla^2 f(x) = P.$

March 12, 2007

- Subspace: $C$ is a subspace if

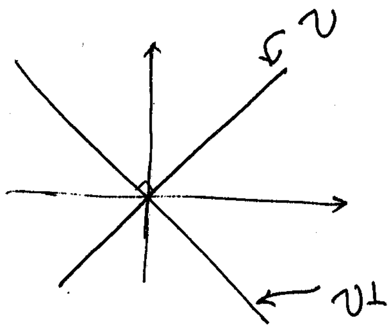
$$x_1, x_2 \in C \Rightarrow \alpha x_1 + \beta x_2 \in C \quad \forall \alpha, \beta \in \mathbb{R}$$




- Orthogonal Complement. Let $\mathcal{V}$ be a subspace of $\mathbb{R}^n$.

$$\mathcal{V}^\perp = \{ x \mid z^T x = 0 \quad \forall z \in \mathcal{V} \}$$




- Let $\oplus$ denote the operation of direct sum of two sets:

$$\mathcal{V} \oplus \mathcal{U} = \{ x + y \mid x \in \mathcal{V}, y \in \mathcal{U} \}$$

Then, $\mathcal{V} \oplus \mathcal{V}^\perp = \mathbb{R}^n$.

- Note that i) $\mathcal{V} \cup \mathcal{U}$ is not a subspace

ii) $\mathcal{V} \cap \mathcal{U}$ is a subspace.

- Range space of $A \in \mathbb{R}^{m \times n}$

$$\mathcal{R}(A) = \{ y \in \mathbb{R}^m \mid y = Ax, \ x \in \mathbb{R}^n \} \subseteq \mathbb{R}^m$$

A range space is a subspace.
The dim. of $\mathcal{R}(A)$ is the rank of $A$.

- Nullspace of $A$

$$\mathcal{N}(A) = \{ x \in \mathbb{R}^n \mid Ax = 0 \} \subseteq \mathbb{R}^n$$

e.g., $A = [a \ a]$. Then $\mathcal{N}(A) = \{ x \in \mathbb{R}^2 \mid x_1 = -x_2 \} = \mathcal{R}(\left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right])$.

Nullspace is a subspace.

- Positive definite matrices:

  $A \in S^n$ is pos. def. (PD) if $x^T A x > 0$ for all $x \in \mathbb{R}^n, x \neq 0$.

  We denote this as $A \succ 0$.

  We use the notation $S^n_{++}$ to denote the set of PD matrices.

- Positive semidefinite (PSD) matrices

  $A \in S^n$ is PSD if $x^T A x \geq 0$ for all $x \in \mathbb{R}^n, x \neq 0$.

  We denote this as $A \succeq 0$.

  We use the notation $S^n_+$ to denote the set of PSD matrices.

(Think about $n=1$ case. Then $A = a$, & we need $a > 0$ for PD & $a \geq 0$ for PSD).

- Symmetric Eigendecomposition: Let $A \in S^n$. Then $A$ can be factored as

$$A = Q \Lambda Q^T$$

where $Q \in \mathbb{R}^{m \times n}$ is orthogonal (i.e., $QQ^T = I$), and

$$\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n).$$

We assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, so that

$$\lambda_{max}(A) = \lambda_1, \quad \lambda_{min}(A) = \lambda_n.$$

- $A$ PD $\iff$ $\lambda_i > 0 \quad \forall i$

  $A$ PSD $\iff$ $\lambda_i \geq 0 \quad \forall i$

- Properties:

$$\det(A) = \prod_{i=1}^{n} \lambda_i$$

$$\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i$$

$$\|A\|_F^2 = \sum_{i=1}^{n} \lambda_i^2$$

These properties can be proven by using the basic property $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

- $$\lambda_{max}(A) = \sup_{x \neq 0} \frac{x^T A x}{x^T x}, \quad \lambda_{min}(A) = \inf_{x \neq 0} \frac{x^T A x}{x^T x}$$

Thus,

$$\lambda_{min}(A)\, x^T x \leq x^T A x \leq \lambda_{max}(A)\, x^T x.$$

- Square root factorizations: Let $A \in S^n_{++}$. Then $A$ can be factored as

$$A = B^T B$$

where $B \in \mathbb{R}^{n\times n}$ is invertible.

(think about $a \geq 0$. Then $b = \sqrt{a}$).

- The matrix $B$ is not unique. An illustration:

$$\begin{aligned} A &= Q \Lambda Q^T \\ &= Q \Lambda^{1/2} U^T U \Lambda^{1/2} Q^T \end{aligned}$$

where $\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_n})$, & $U \in \mathbb{R}^{n\times n}$ is (any) orthogonal matrix.
We then have $B = U \Lambda^{1/2} Q^T$.

- The symmetric square root factorization is unique though:

$$B = Q \Lambda^{1/2} Q^T.$$

- Singular Value Decomposition (SVD): Let $A \in \mathbb{R}^{m\times n}$ with $\text{rank} A = r$.

$$A = U \Sigma V^T$$

where $U \in \mathbb{R}^{m\times m}$ & $V \in \mathbb{R}^{n\times n}$ are orthogonal, and

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{bmatrix} \in \mathbb{R}^{m\times n}$$

where $\sigma_i > 0$.

We assume $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$.

- Thin SVD

$$A = \begin{array}{c} \overset{r}{\longleftrightarrow} \\ [\, U_1 \;\; U_2 \,] \end{array} \begin{bmatrix} \tilde{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \updownarrow r$$

$$= U_1 \tilde{\Sigma} V_1^T$$

where $\tilde{\Sigma} = \operatorname{diag}(\sigma_1, \dots, \sigma_r)$.

- Recall $\|A\|_2 = \sup\{\|Ax\|_2 \mid \|x\|_2 \le 1\}$

It can be shown that $\|A\|_2 = \sigma_{max}(A) \; (= \sigma_1)$.

-

$$\sigma_{max}(A) = \sigma_1, \quad \sigma_{min}(A) = \begin{cases} \sigma_r, & r = \min\{m, n\} \\ 0, & \text{otherwise} \end{cases}$$

- Inverse: Let $A \in \mathbb{R}^{n \times n}$. Its inverse, $A^{-1}$, satisfies $A^{-1}A = I$ & $AA^{-1} = I$.

- Pseudo-inverse: Let $A \in \mathbb{R}^{m \times n}$.

$$A^+ = V_1 \tilde{\Sigma}^{-1} U_1^T$$

If $\operatorname{rank} A = n$ (full col. rank), then $A^+ = (A^TA)^{-1}A^T$.
If $\operatorname{rank} A = m$ (full row rank), then $A^+ = A^T(AA^T)^{-1}$.

- Least squares (LS)

$$\min \|Ax - b\|_2^2$$

An LS solution is $x^* = A^+b$.

Moreover, $A^+b + v$, $v \in \mathcal{N}(A)$, is a solution of LS.

- Schur Complement

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \in S^n.$$

We also have $A \in S^k$. If $\det A \neq 0$,

$$S = C - B^T A^{-1} B$$

is called the Schur complement.

Property:

$$\det X = \det A \det S.$$

- compare the $2\times 2$ case

$$X = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

$$\det X = ac - b^2$$
$$= a\underbrace{\left(c - b^2/a\right)}_{S}.$$

Important Property: (will be proved later on)

- $X \succ 0$ if and only if $A \succ 0$ and $S \succ 0$.
- If ~~$A \succeq 0$~~ $A \succ 0$, then $X \succeq 0$ if and only if $S \succeq 0$.