ELL780 ELL 457 Indian Institute of Technology Delhi
Lecture 10 Metric Spaces and Norms
Scribed by: *Kshitij Alwadhi(2019EE10577)*, *Shivam Garg(2019EE10600)*
Instructors: Sandeep Kumar and Jayadeva

Disclaimer: These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Course Coordinator.

Contents

1 Introduction

The study of linear algebra has become more and more popular in the last few years. People are attracted to this subject because of its never ending application in domains like Machine learning. One theoretical aspect is the need to measure the length of vectors. For this purpose, norm functions are considered on a vector space. In this lecture we explore more about metric spaces and norms. Towards the end we also look at Open ball and Closed ball.

2 Metric Space

Consider a non-empty set X.

2.1 Definition

A metric or *distance-function*, on X is a function d : X $\times$ X $\rightarrow \mathbb{R}$ with the following properties:

1. d(x,y) $\geq 0\ \forall$ x,y $\in$ X and d(x,y) = 0 if and only if x=y.
2. d(x,y) = d(y,x) for all x,y $\in$ X.
3. d(x,z) $\leq$ d(x,y) + d(y,z) for all x,y,z $\in$ X.

2.2 Quasi-k-semimetric

A quasi-k-semimetric on an arbitrary nonempty set X is a function $\rho : \mathbb{X} \times \mathbb{X} \rightarrow [0, \infty)$ that satisfies:

1. $\rho(x, x) = 0$,
2. $\rho(x, z) \leq k(\rho(x, y) + \rho(y, z))$, for all x,y,z $\in$ X,
3. $\rho(x, y) = \rho(y, x) = 0$ implies x=y, for all x,y $\in$ X, then ρ is called a quasi-k-metric.

2.3 Examples

Example 1 Let d(x,y) = $|x - y|$. To prove that this is a metric space, the first two conditions are trivial and the third follows from the ordinary triangle inequality for real numbers:

$$d(x, y) = |x - y| = \big|(x - z) + (z - y)\big| \leq |x - z| + |z - y| = d(x, z) + d(z, y)$$

Example 2 Assume that you want to move from one point (x_1, x_2) to another (y_1, y_2) in the plane but are only allowed to move horizontally and vertically. Say you take a path from (x_1, x_2) to (y_1, x_2) and then from (y_1, x_2) to (y_1, y_2), the total distance is

$$d(x, y) = |y_1 - x_1| + |y_2 - x_2|$$

This gives us a metric on $\mathbb{R}^2$ which is different from our usual metric given in Example 1. It is often referred to as the *Manhattan metric* or the *taxi cab metric.*

Also in this case, the first two conditions are trivial. To prove the triangle inequality, observe that for any third point z = (z_1, z_2), we have

$$\begin{aligned} d(x,y) =& |y_1 - x_1| + |y_2 - x_2| = \\ =& \big|(y_1 - z_1) + (z_1 - x_1)\big| + \big|(y_2 - z_2) + (z_2 - x_2)\big| \leq \\ \leq& |y_1 - z_1| + |z_1 - x_1| + |y_2 - z_2| + |z_2 - x_2| = \\ \leq& |z_1 - x_1| + |z_2 - x_2| + |y_1 - z_1| + |y_2 - z_2| = \\ =& d(x,z) + d(z,y) \end{aligned}$$

2.4 Exercise

Exercise 1 A sequence x_n of real numbers is called *bounded* if there is a number M $\in \mathbb{R}$ such that $|x_n| \ \leq$ M for all n $\in \mathbb{N}$. Let X be the set of all bounded sequences. Show that

$$d(x_n, y_n) = sup(|x_n - y_n| : n \in \mathbb{N})$$

Exercise 2 Let X be a non empty set, and let ρ: X x X $\rightarrow \mathbb{R}$ be a function satisfying:

a. ρ(x,y) $\geq$ 0 with equality if and only if x = y.

b. ρ(x,y) $\leq\rho$(x,z) + ρ(z,y) for all x,y,z $\in$ X.

Define d : X x X $\rightarrow$ R by d(x,y) = max[ρ(x,y) , ρ(y,x)].
Show that d is a metric on X.

3 Norms

3.1 Definition

The norm of a vector x $\in$ V, where V is the vector space is a real non-negative value intuitively representing the length/magnitude of the vector. A norm should satisfy the following properties:

1. $\|x\| = 0 \implies$ x = 0

2. $\|\lambda.x\| = |\lambda|\ .\ x$

3. $\|x + y\| \leq\ \|x\|\ +\ \|y\|$

3.2 All norms induce metrics, but all metrics are not induced by norms

Proof: Let V be a vector space over the field F. A norm

$$\|.\| : V \to \mathbb{F}$$

on V satisfies the homogeneity condition:

$$\|ax\| = |a| \ . \ \|x\|$$

for all a $\in$ F and x $\in$ V. So the metric

$$d \ : \ \mathbb{V} \times \mathbb{V} \to \mathbb{F},$$

$$d(x, y) = \|x - y\|$$

defined by the norm is such that

$$d(ax, ay) = \|ax - ay\| = |a| \ . \ \|x - y\| = |a| \ . \ d(x, y)$$

for all a $\in$ V. This property is not satisfied by general metrics. For example, let δ by the discrete metric:

$$\delta(x, y) = \begin{cases} 1, & x \neq y, \\ 0, & x = y \end{cases}$$

Then δ clearly does not satisfy the homogeneity property of the metric induced by a norm. ∎

A metric is called **translation invariant** if

$$d(x + z, y + z) = d(x, y)$$

for all x,y,z $\in$ V. Then a homogeneous and translation invariant metric d can be used to define a norm $\| \ . \ \|$ by

$$\|x\| = d(x, 0)$$

for all x $\in$ V.

3.3 Induced Norm

If a metric over a vector space satisfies the following properties:

1. d(x, y) = d(x+z, y+z)
2. d(ax, ay) = $|x|$ d(x, y)

then it can be turned into a norm via w(x) = d(x,0). This is called the **induced norm.**

Proof: Suppose d is a metric that satisfies the additional properties as given above; we show the "induced norm" is indeed a norm. We concentrate on the triangle inequality, as the other properties are trivial. Let x,y,z $\in$ V; then

$$\begin{aligned}
w(x+y) = d(x+y,0) &\leq d(x+y,(x+y)-y) + d((x+y)-y,0) && (1)\\
&= d(x+y,x) + d(x,0) && (2)\\
&= d(y+x,0+x) + d(x,0) && (3)\\
&= d(y,0) + d(x,0) && (4)\\
&= w(y) + w(x) && (5)
\end{aligned}$$

■

3.4 L_p Norm

Let x $\in \mathbb{R}^n$ such that x = $(x_1, x_2, ...x_n)$.

An L_p norm is defined as:

- $\|x\|_p = (\sum_{n=1}^{\infty} |x_i|^p)^{1/p}$ if 1 $\leq$ p $< \infty$.
- $\|x\|_p = \max(|x_1| \ , |x_2| \ , \ldots \ |x_n| \ )$ if p $= \infty$.

Proof: To prove that $\|x\|_\infty = \max(|x_k| \ : 1 \leq k \leq \infty)$ is a norm.
Proving $\|x\|_\infty = 0$ when x = 0 is trivial since

$$x_k = 0 \ \forall \ k = 0, 1..$$

Next we prove the second property:

$$\begin{aligned}
\|\lambda x\|_\infty &= max \ \ |\lambda x_k| \ \forall \ k && (6)\\
&= \lambda \ \cdot \ max \ \ (|x_k|) && (7)\\
&= \lambda \ \cdot \ \|x\|_\infty && (8)
\end{aligned}$$

For the triangle inequality we see that:

$$\begin{aligned}
\|x+y\|_\infty &= max \ \ |x_k + y_k| && (9)\\
&\leq max \ \ |x_k| + |y_k| && (10)\\
&\leq max \ \ |x_k| \quad + \quad max \ \ |y_k| && (11)\\
&\leq \|x\|_\infty + \|y\|_\infty && (12)
\end{aligned}$$

Hence, it is a norm. ■

3.5 Holder's Inequality

Suppose we have p,q $\in$ [1,∞] such that,

$$\frac{1}{p} + \frac{1}{q} = 1$$

According to Holder's inequality, if f and g are measurable functions,

$$\|fg\|_1 \leq \|f\|_p \|g\|_q$$

Proof: Dividing f and g by $\|f\|_p and \|g\|_q$, respectively, we can assume that:

$$\|f\|_p = \|g\|_q = 1$$

We now use Young's Inequality for products, which states that

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

for all non-negative a and b, where equality is achieved if and only if $a^p = b^q$. Hence

$$\left|f(s)g(s)\right| \leq \frac{\left|f(s)\right|^p}{p} + \frac{\left|g(s)\right|^q}{q}, \qquad s \in S.$$

Integrating both sides gives

$$\|fg\|_1 \leq \frac{\|f\|_p^p}{p} + \frac{\|g\|_q^q}{q} = \frac{1}{p} + \frac{1}{q} = 1,$$

which proves the claim.

Under the assumptions p $\in$ (1,∞) and $\|f\|_p = \|g\|_q$ almost everywhere. More generally, if $\|f\|_p$ and $\|g\|_q$ are in (0,∞), then Holder's inequality becomes an equality if and only if there exists real numbers $\alpha, \beta > 0$, namely

$$\alpha = \|g\|_q^q, \quad \beta = \|f\|_p^p$$

such that

$$\alpha |f|^p = \beta |g|^q \quad \mu - almost\ everywhere$$

$\blacksquare$

3.6 Minkowski's inequality

$\|x+y\|_p \leq \|x\|_p + \|y\|_p$
Proof:

$$\begin{aligned}
\|f+g\|_p^p &= \int |f+g|^p \, d\mu && (13)\\
&= \int |f+g| \cdot |f+g|^{p-1} \, d\mu && (14)\\
&\leq \int (|f| + |g|)|f+g|^{p-1} \, d\mu && (15)\\
&= \int |f||f+g|^{p-1} \, d\mu + \int |g||f+g|^{p-1} \, d\mu && (16)\\
&\leq \left(\left(\int |f|^p \, d\mu \right)^{\frac{1}{p}} + \left(\int |g|^p \, d\mu \right)^{\frac{1}{p}} \right) \left(\int |f+g|^{(p-1)(\frac{p}{p-1})} \, d\mu \right)^{1-\frac{1}{p}} && (17)\\
&= (\|f\|_p \ + \ \|g\|_p \ ) \frac{\|f+g\|_p^p}{\|f+g\|_p} && (18)
\end{aligned}$$

We obtain Minkowski's inequality by multiplying both sides by

$$\frac{\|f+g\|_p}{\|f+g\|_p^p}$$

■

4 Inner Product

4.1 Definition

Let V be a vector space over $\mathbb{R}$. We define an inner product $< . >$ from $V \times V \to \mathbb{R}$ such that it satisfies the following conditions:

1. $< v, v > \geq 0 \forall$ u,v $\in V$.
2. $< v, v >= 0 \Leftrightarrow$ v $= 0 \ \forall$ v $\in V$.
3. $< v, u >=< u, v > \ \forall$ u,v $\in V$.
4. $< x + z, y >=< x, y > + < z, y > \ \forall$ x,y,z $\in V$.
5. $< \lambda x, y >= \lambda < x, y > \ \forall$ x,y $\in$ V and $\lambda \in \mathbb{R}$.

4.2 Properties

Inner Product $< . >$ on V $\times$ V $\to \mathbb{R}$ has following properties:

1. Orthogonality: If $< x, y > = 0$ then x and y are orthogonal.
2. Cauchy-Schwarz Inequality: $| < x, y > | \leq \|x\| * \|y\|$.

4.3 Examples

Example 1: Find an inner product on R^2 such that $<e_1,e_1> = 2$, $<e_2,e_2> = 3$, and $<e_1,e_2> =$ -1, where e_1 = (1, 0), e_2 = (0, 1). Let x = (x_1, x_2), y = (y_1, y_2). Then x = $x_1 e_1$ + $x_2 e_2$, y = $y_1 e_1$ + $y_2 e_2$.

$$< x, y >=< x_1e_1 + x_2e_2, y_1e_1 + y_2e_2 >$$

$$< x, y >= x_1 < e_1, y_1e_1 + y_2e_2 > +x_2 < e_2, y_1e_1 + y_2e_2 >$$

$$< x, y >= x_1y_1 < e_1, e_1 > +x_1y_2 < e_1, e_2 > +x_2y_1 < e_2, e_1 > +x_2y_2 < e_2, e_2 >$$

$$< x, y >= 2x_1y_1 - x_1y_2 - x_2y_1 + 3x_2y_2$$

We need to check that $<x, x> > 0$ for $x \neq 0$.

$$< x, x >= 2x_1^2 - 2x_1x_2 + 3x_2^2 = (x_1 - x_2)^2 + x_1^2 + x_2^2$$

5 Matrix Norm

5.1 Definition

It is a function $\mathbb{R}^{m*n} \rightarrow \mathbb{R}$ that satisfies the same properties of vector norms. Some examples of Matrix norms are:

- Frobenius Norm: It is defined as $\|A\|_F = \sqrt{Tr(A^TA)}$.
- Sum Absolute Norm: It is defined as $\|A\|_{SAN} = \sum \mid a_{ij} \mid$.
- Max Absolute Norm: It is defined as $\|A\|_{MAN} = \max \mid a_{ij} \mid$.

5.2 Equivalence Norm

Two norms p_1, p_2 are said to be equivalent if $\exists$ a,b $\in \mathbb{R}$ such that a$\|.\|_{p1} \leq \|.\|_{p2} \leq$ b$\|.\|_{p1}$

5.3 Examples

Matrix norms are used to calculate perturbations. The perturbation theory is the study of a small change in a system. For example, to compare two images which look similar to naked eye we calculate L_2 and L_∞ norms.

Example1: Calculate L_2 distance between these 2 Black-and-white images of pixel size 3.

A_1 = [[1,2,3],[2,3,4],[0,1,2]] A_2 = [[0,0,2],[4,2,4],[0,3,2]]

$$L_2 = \sqrt{(1-0)^2 + (2-0)^2 + (3-2)^2 + (2-4)^2 + (3-2)^2 + (4-4)^2 + (0-0)^2 + (1-3)^2 + (2-2)^2}$$

$$L_2 = \sqrt{15} \approx 3.873$$

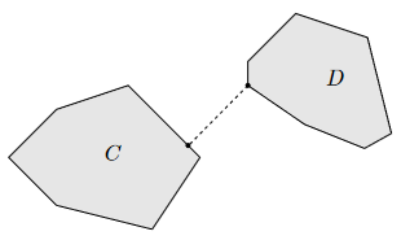
This can be executed using python with the help of the numpy library. The sample code is given below:

```python
import numpy as np

vec_a = np.array([[1,2,3],[2,3,4],[0,1,2]])
vec_b = np.array([[0,0,2],[4,2,4],[0,3,2]])

l2_dist = np.linalg.norm(vec_a - vec_b)

print(l2_dist)
```

Listing 1: Calculating L_2 distance using python

6 Distance Between Sets

Let (X,d) be a metric space.



6.1 Distance Between a point and a Set

Let x $\in$ X and A $\subset$ X where A $\neq \phi$, then distance between a point and a set is defined by:

$$d(x, A) = inf < d(x, a) : a \in A >$$

NOTE : If x $\in$ A $\rightarrow$ d(x,A) = 0 but its converse that is if d(x,A) = 0 $\rightarrow$ x $\in$ A is not true.

6.2 Distance Between Two Sets

Let A $\subset$ X and B $\subset$ X where A,B $\neq \phi$, then distance between two set is defined by:

$$d(A, B) = inf < d(x, y) : x \in A, y \in B >$$

7 Diameter of a set

It is defined as the biggest possible distance between two points in the set A.

$$Diameter(A) = sup < d(x, y) : x, y \in A >$$

If it is not bounded above, then we say that the diameter is ∞. If diameter is less than ∞, then we call the set as bounded.

8 Open ball

8.1 Definition

For a metric space (X,d), Let a be a given point in metric space (X,d) and let r be a positive number. The set of all points x in metric space (X,d) such that:

$$U_r(x) = y \in X : d(y, x) < r$$

We use two parameters here: Centre(x) and Radius(r). The ball consists of all the points whose distance from x is less than r. We have the following interpretations:

- In 1-D space, the points lie on a line with centre at x (open interval).
- In 2-D space, the points lie on a circular disk (excluding boundary).
- In 3-D space, the points lie in a spherical solid with centre at x and radius r.

8.2 Examples

Example 8.1 $U_1[(0,0)] = \{ \ x \in \mathbb{R}^2 \ ; \ x_1^2 + x_2^2 < 1 \ \}$ *This is same as our Euclidean metric where the sum of the squares of our co-ordinates is less than 1.*

9 Closed ball

9.1 Definition

A closed ball is defined as:

$$B_r(x) = y \in X : d(y, x) \leq r$$

We use two parameters here: Centre(x) and Radius(r). The ball consists of all the points whose distance from x is less than or equal to r. We have the following interpretations:

- In 1-D space, the points lie on a line with centre at x (closed interval).
- In 2-D space, the points lie on a circular disk (including boundary).
- In 3-D space, the points lie in (also on) a spherical solid with centre at x and radius r.

9.2 Examples

Example 9.1 *For the set* $(x_1, x_2) \in \mathbb{R}^2 \ : x_1^2 + x_2^2 \leq r^2$*, a closed ball of radius* r *centered at* x *is the set defined as:*

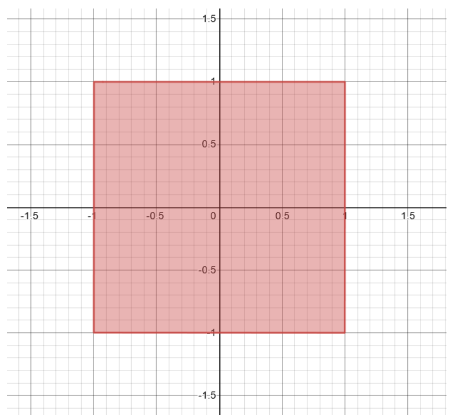
$$B_r(x) = \{ \ y \in X : \ d(x, y) \leq r \ \}$$

Example 9.2 *Suppose we have the metric* $d((x_1, y_1), (x_2, y_2)) = max(|x_1 - x_2|, |y_1 - y_2|)$*. Setting our second point* (x_2, y_2) *as origin, we get the following:*

$$B_1(x) = \{ \ y \in X : \ d(0, y) \leq 1 \ \}$$

$$d(0, y) = max(|x_1|, |y_1|) \leq 1$$

Now graphing this, we get the following plot:



9.3 Exercise

Exercise 1 Let d : $R^2 \times R^2 \rightarrow$ R is defined as:

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2|$$

were $x = (x_1, x_2) andy = (y_1, y_2)$, is defined as a metric on R^2 called the taxicab metric as discussed before. Practice the proving that it satisfies the metric axioms. Also sketch the unit ball $B_d(0, 1)$.

References

[1] Tom M. Apostol, *Mathematical Analysis*. Addison-Wesley,

[2] Walter Rudin, *Principles of Mathematical Analysis*. McGraw Hill Education, Third edition

[3] Desmos Graphing Calculator for plots, https://www.desmos.com/calculator

[4] Stackexchange, https://math.stackexchange.com

[5] Holders Inequality and Minkowski's Inequality, https://services.math.duke.edu/~wka/math204/conv.12.4.pdf

[6] Norm to/from metric, https://www.math.usm.edu/perry/old_classes/mat681sp14/norm_and_metric.pdf

[7] Wikipedia (Proof of inequalities), https://en.wikipedia.org/wiki/

HANDWRITTEN NOTES

- KSHITIJ ALWADHI

2019EE10577

- Limit, Continuity & Differentiability
- Linear Algebra
- Optimization Theory

LIMIT & CONT (2019EE10577)

(*) A sequence $\langle x_n \rangle$ cgs in (X,d) if $\exists\, p \in X$ st. $\forall\, \varepsilon > 0$ $\exists\, n_0 \in \mathbb{N}$ st. $\forall\, n \geq n_0$,
$$d(x_n, p) \leq \varepsilon$$

$x_n \to p \iff d(x_n, p) \to 0$

(*) Cauchy Sequence

Sequence $\langle x_n \rangle$ is cauchy if $\forall\, \varepsilon > 0$, $\exists\, n_0 \in \mathbb{N}$ such that
$$d(x_n, x_m) < \varepsilon \quad \forall\, m, n \geq n_0.$$

(*) Every cgt. sequence is cauchy (but reverse not true).

(*) If $\langle x_n \rangle$ is cauchy and it has a cgt subsequence $\langle x_{n(F)} \rangle$, then $\langle x_n \rangle$ itself is cgt.

(*) $f(x) \to l$ as $x \to p$. limits of a function.

$\exists$ a, $\delta > 0$ s.t. $d(x,p) < \delta$
$$\Downarrow$$
$$e(f(x), l) < \delta.$$

(*) Continuous function (X,d), $(Y,e) \to$ metric spaces

$f: A \subseteq X \to Y$

f is cont. at $p \in A$ if:

$\forall\, x \in A$, $\forall\, \varepsilon > 0$ $\exists\, \delta > 0$ st.
$$d(x,p) < \delta \implies e(f(x), f(p)) < \delta.$$

(*) f is cont. iff
$x_n \to x_0 \Rightarrow f(x_n) \to f(x_0)$

(*) $f : A \to R$, f cont.
A is compact subset of X.
then $\exists$ points p & q in X s.t.
$f(p) = \inf(f(x))$ and $f(q) = \sup(f(x))$

(*) Uniformly continuous
$\forall \varepsilon > 0$, $\exists \delta > 0$ (depending only on ε)
such that if $x \in A$ and $p \in A$ then
$$d(x,p) < \delta \Rightarrow \rho(f(x), f(p)) < \varepsilon$$

(*) Lipschitz Continuity
(X, d) (Y, ρ)
$f : X \to Y$ satisfies lipschitz condition
at x_0 if,
$\exists R > 0$ s.t. $\forall x \in X$
$$d(x, x_0) < \delta \Rightarrow \rho(f(x), f(x_0)) < R \cdot d(x, x_0)$$
R: lipschitz const.

Fixed Point

$f : X \to X$ $\qquad (x, d)$

A point $p \in X$ is called fixed point of f if
$$f(p) = p.$$

A function is called contraction of X if $\exists\ \alpha$, $0 < \alpha < 1$ st. $\qquad f : X \to X$
$$d(f(x), f(y)) \leq \alpha\, d(x, y) \quad \forall\ x, y \in X$$

$$\begin{array}{ccc} \text{Lipschitz} & \Longrightarrow & \text{locally Lipschitz} \\ \Downarrow & & \Downarrow \\ \text{Uniformly cont.} & \Longrightarrow & \text{Continuous} \end{array}$$

Rolles theorem

$f : [a, b] \to R$

$f(a) = f(b)$, then

$\exists\ c \in (a, b)$ st. $f'(c) = 0$

MVT

f : cont. and diff in (a, b)

then $\exists\ c \in (a, b)$ st. $f'(c) = \dfrac{f(b) - f(a)}{b - a}$

(2019EE10577)

* Row space of A = Col. space of A^T
  $= C(A^T)$

* $[A][B] = [C]$

  Columns of C are combinations of columns of A.

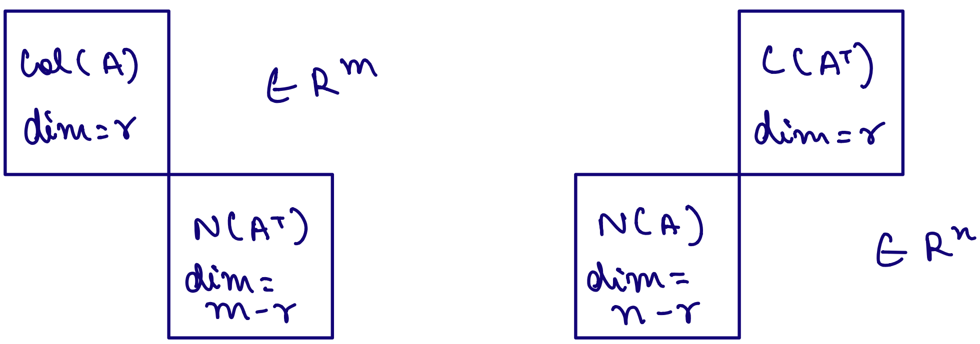
  Rows of C are combinations of rows of B.

* $Col(A)$: $Ax \quad \forall x$

  $N(A)$: $\{ x \mid Ax = 0 \}$

  $Col(A^T)$: $A^T x$

  $N(A^T)$: $\{ x \mid A^T x = 0 \}$

* 

  $A \in R^{m \times n}$

* Orthogonal vectors

  $\langle x, y \rangle = 0$

  $x^T y = 0$

* Orthogonal complement

  $S^\perp = \langle x \mid \langle x, y \rangle = 0 \quad \forall y \in S \rangle$

(*) A vector x is orthogonal to a subspace S if,

$$\langle x, b_i \rangle = 0 \quad \forall\ b_1, b_2 \ldots b_n$$

where b_i is any basis.

(*) Rowspace is orthogonal to Nullspace.

(*) Orthonormal

$$\langle x, y \rangle = 0$$

$$\|x\| = 1, \quad \|y\| = 1$$

(*) $Q = [q_1\ q_2\ \ldots\ q_n]$

$$q_i^T q_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

$Q^T Q = I$ → General matrix

$Q^T = Q^{-1}$ → Square matrix

(*) Let x be any vector, Q be orthonormal matrix.

$\|Qx\| = \|x\|$ (length unchanged)

$$\begin{aligned} \langle Qx, Qx \rangle &= (Qx)^T (Qx) \\ &= x^T Q^T Q x \\ &= x^T x = \|x\| \end{aligned}$$

Also preserves the inner product and angle.

(*) Grahm - Schmidt Orthogonalization

Set of linearly independent vectors $x_1, x_2 \dots x_n$. To obtain orthonormal basis of the span of the vectors.

1. $$U_1 = \frac{x_1}{\|x_1\|_2}$$

2. For $i = 1, 2 \dots m$

$$V_i = x_i - \sum_{j=1}^{i-1} \langle v_j, x_i \rangle v_j$$

$$u_i = \frac{V_i}{\|V_i\|_2}$$

(*) Grahm - Matrix

$A \in R^{m \times n}$

$(A^T A)$ is said to be grahm matrix.

$(A^T A)$ will be invertible if columns of A are linearly independent.

(*) If A is left invertible then columns of A are linearly independent.

If A is right invertible then rows of A are L.I.

Pseudo Inverse

$[A]_{m\times n}$ $\quad\quad Ax = b$

$$A^+ = (A^T A)^{-1} \cdot A^T$$

$m > n$	$m < n$
$XA = I$	$AX = I$

$$X = A^+$$

LU decomposition

$A = LU$

$L \rightarrow$ lower triangular

$U \rightarrow$ upper triangular.

QR decomposition

$A \in R^{m\times n}$ $\quad$ A having linearly independent columns.

$$\underset{m\times n}{\overset{A}{\begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}}} = \underset{m\times n}{\overset{Q}{\begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix}}} \underset{n\times n}{\overset{R}{\begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ 0 & R_{22} & & R_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & R_{nn} \end{bmatrix}}}$$

$[A] = [Q]\,[R] \rightarrow$ upper triangular.

$[A] \downarrow$ n independent cols.

$[Q] \downarrow$ n orthonormal cols

(*) Eigen Value decomposition

(A needs to be symmetric)
$(A = A^T)$

$$A_{n\times n} = Q \Lambda Q^T$$

Q: orthonormal matrix

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$ diagonal matrix

$$A q_i = \lambda_i q_i \quad \forall\, i$$

$$Q \Lambda Q^T = \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T \; - - - .$$

$$Q \Lambda Q^T = \sum_{i=1}^{n} \lambda_i q_i q_i^T$$

(*) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$

$$\operatorname{tr}(A) = \operatorname{tr}(Q \Lambda Q^T) = \operatorname{tr}(Q^T Q \Lambda)$$
$$= \operatorname{tr}(\Lambda) = \sum_{i=1}^{n} \lambda_i$$

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

④ Single Value Decomposition (SVD)

(works for non symmetric too)

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

diagonal matrix

orthogonal matrix

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r & & \\ & & & & 0 & \\ & & & & & \ddots \; 0 \end{bmatrix}$$

diagonal matrix

$$A = \sum_{i=1}^{r} \sigma_i V_i U_i^T$$

④ $$(A^T A) = (V \Sigma U^T)^T (V \Sigma U^T)$$

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

$$= U \Sigma^2 U^T$$

like EVD

$$\lambda_i(A^T A) = (\sigma_i(A))^2 \quad \forall\, i = 1 \ldots r$$

Positive Definite

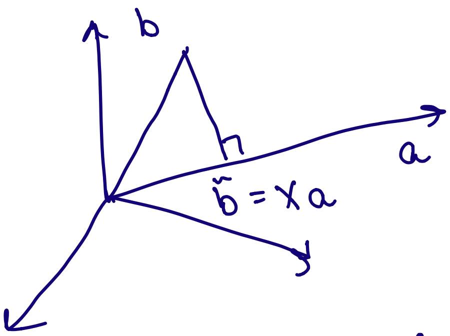
$\|v\|_P^2 = v^T P v > 0 \quad \forall\ v$ only if

$v^T P v > 0 \quad \forall\ v \neq 0$

1) $v^T P v > 0 \quad \forall\ v \neq 0$
2) $\lambda_i(P) = 0 \quad \forall\ i$
3) $P = LL^T$

$$A x = b$$

if b not in col. space of A i.e. Ax.



$$\langle a, \tilde{b} - b \rangle = 0$$

$$a^T (b - xa) = 0$$

$$x = \frac{a^T b}{a^T a}$$

$$\tilde{b} = \left(\frac{a^T b}{a^T a}\right) a = \left(\frac{a a^T}{a^T a}\right) b$$

ask how?

$P = \frac{A A^T}{A^T A}$ → projection matrix.

$$P = A (A^T A)^{-1} A^T$$

$P^2 = P$ and $P^T = P$

(2019EE10577)

* Optimization Problem

$$x^* = \operatorname{argmin} f_0(x) \rightarrow \text{Optimization function.}$$

$$\text{s.t.} \quad f_i(x) \leq 0 \rightarrow \text{inequality constraint}$$

$$h_i(x) = 0 . \rightarrow \text{equality constraint}$$

* domain $\Rightarrow$ implicit constraints

* Affine set : C is affine if the line through any 2 points in C, also lies in C. (C must contain ∞ pts).

$x_1, x_2 \in C$

$$C \text{ affine} \Leftrightarrow \theta x_1 + (1-\theta) x_2 \in C \quad \forall \theta \in R$$

* Affine Hull $\{x_i\}_{i=1}^{k}$

$$\text{Aff}(C) = \left\{ \sum_{i=1}^{k} \theta_i x_i \;\middle|\; \begin{array}{ll} k \geq 2 & \sum \theta_i = 1 \\ x_i \in C & \theta_i \in R \end{array} \right\}$$

* Convex Set

line between any 2 points also lie in C.

$x_1, x_2 \in C$ then C is convex set

$$\theta x_1 + (1-\theta) x_2 \in C$$

$$\forall \theta \in [0,1]$$

* If A is affine, then it is also convex.
~~If~~ A is convex, then it is not necessary to be affine.

* **Convex Hull**

$\{x_i\}_{i=1}^{k}$ are points given

$$\operatorname{Conv}(C) = \left\{ \sum_{i=1}^{k} \theta_i x_i \;\middle|\; \begin{array}{l} x_i \in C \\ \sum \theta_i = 1 \;;\; 0 \le \theta_i \le 1 \end{array} \right\}$$

* **Convex Cone**

$x_1, x_2 \in C$, $\quad C$ is a convex cone

$$\iff \theta_1 x_1 + \theta_2 x_2 \in C$$
$$\forall\, \theta_1, \theta_2 \geq 0$$

* **Convex Conic Hull** $\quad x_i \in C$

$$C(C) = \left\{ \sum_{i=1}^{k} \theta_i x_i \;\middle|\; \begin{array}{l} k \geq 2 \\ \theta_i \geq 0 \end{array} \quad x_i \in C \right\}$$

* $$B(x_c, r) = \left\{ x \in \mathbb{R}^n \;\middle|\; \|x - x_c\| \le r \right\}$$

$$\updownarrow$$

$$= \left\{ x_c + r u \;\middle|\; \|u\| \le 1 \right\}$$

Ellipsoid

$$\mathcal{E}(x_c, P) = \{ x \in \mathbb{R}^n \mid (x - x_c)^T P^{-1} (x - x_c) \le 1 \}$$

- If $P = r I$, then l_2-norm ball.
- $P \succ 0$, $P \in S^n$.

$$P_{n \times n} = \mathcal{E} \Pi \mathcal{E}^T,$$

$$\sqrt{P} = \mathcal{E} \sqrt{\Pi}\, \mathcal{E}^T, \qquad \sqrt{\Pi} = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix}$$

$$\mathcal{E}_c = \{ x_c + \sqrt{P}\, u \mid \|u\| \le 1 \}$$

Norm Cone

$$C = \left\{ \begin{bmatrix} x \in \mathbb{R}^n \\ t \end{bmatrix} \in \mathbb{R}^{n+1} \;\middle|\; \|x\|_2 \le t \right\}$$

Affine transformation

$$a(x) = Ax + b, \qquad \begin{array}{l} x \in \mathbb{R}^n \\ A \in \mathbb{R}^{m \times n} \\ b \in \mathbb{R}^m \end{array}$$

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

$$A(C) = B = \{ a(x) \mid x \in C \}$$

$$A^{-1}(B) = \{ x \in \mathbb{R}^n \mid a(x) \in B \}$$

if C is convex then $A(C)$ is convex.

if B is convex then $A^{-1}(B)$ is also convex.

(*) Product of two convex sets is also convex.

$$C_1 \times C_2 = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \;\middle|\; \begin{matrix} x_1 \in C_1 \\ x_2 \in C_2 \end{matrix} \right\}$$

(*) <u>Minkowski Sum</u> (Sum also convex)

$$C_1 + C_2 = \{ x_1 + x_2 \mid x_1 \in C, \ x_2 \in C \}$$

It is also an affine transformation of product set.

Q. Hyperbolic cone: (check if convex or not).

$$H = \left\{ u \in R^n \mid u^T P u \leq (c^T u)^2, \ \begin{matrix} c^T u \geq 0 \\ P > 0 \end{matrix} \right\}$$

Also convex ✓.

(*) <u>Perspective function</u>

$$P : R^{n+1} \longrightarrow R$$

$$P(\overset{\in R^n}{z}, \overset{\in R}{t}) = \frac{z}{t}, \quad t > 0$$

$$\text{dom}(P) = R^n \times (R_{++})$$

$$\tilde{z} \in R^{n+1}$$

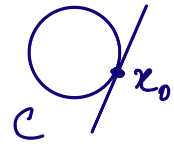
$$\tilde{z}_i = z_i / t$$

$$\tilde{z}_{n+1} = t$$

this transformation also convex sets.

⊛ Supporting Hyperplane

$\{ y \mid a^T y = a^T x_0 \}$ is a supporting H.P.



if $a \neq 0$,

$$a^T x \leq a^T x_0 \quad \forall x \in C.$$

⊛ Separating Hyperplane

C_1, C_2 convex, $C_1 \cap C_2 = \phi$

$a^T y = b. \longrightarrow$ Separating hyperplane

⊛ Convex Function

(zeroth order).

$$\theta f(x) + (1-\theta) \cdot f(y) \geq f(\theta x + (1-\theta) y)$$

$\forall x, y$ in domain.

⊛ Concave function

f concave $\Longleftrightarrow$ $-f$ is convex.

$$f(\theta x + (1-\theta) y) \geq \theta \cdot f(x) + (1-\theta) \cdot f(y)$$

⊛ First order condition

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle$$

$\forall y \in$ domain.

(*) Second order condition

$$[\nabla^2 f]_{i,j} = \frac{\partial^2 F}{\partial x_i \cdot \partial x_j} \qquad i,j = 1,2 \ldots n$$

$$\nabla^2 f \geq 0 \quad \Rightarrow \quad \text{convex function.}$$

(*) $S_\alpha = \{ x \mid f(x) \leq \alpha \} \longrightarrow$ Sublevel set

$S_\alpha = \{ x \mid f(x) \geq \alpha \} \longrightarrow$ Superlevel set

if $f(x)$ is convex, then S_α is convex set.
but converse is not true.

(*) Quasi - Convexity

Functions for which the Sublevel set is a convex set.

Quasi - Concave

Functions for which the Superlevel set is a convex set.

(*) Operations / Notes

1) f_i's are convex then,

$$f(x) = \sum_{i=1}^{m} w_i \cdot f_i \quad \text{is convex}, \quad w_i \geq 0.$$

2) If f is convex fun, then
the affine Transform

$f(Ax+b) \longrightarrow$ will also be convex.

3) If f_i, $i = 1, 2 \dots n$ are convex.

$g(x) = \max_i \{ f_i(x) \}$ is also convex.

domain $g = \bigcap_{i=1}^{m} \text{dom } f_i$ also convex.

4) <u>Support</u> : A set C is given,

$S_C(x) = \sup_{y \in C} \{ x^T y \}$ also convex.

The set C need not be convex.

5) Max. eigen value of a symmetric matrix

$f(x) = \lambda_{max}(X)$, dom $f = S^m$.

$f(x) = \sup_y \{ y^T X y \mid \|y\|_2 = 1 \}$ also convex.

(*) Composition of function

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

$f(x)$ is convex if :

sufficient but not necessary.

1) g convex, h convex and non-decreasing.

2) g concave, h convex and non-increasing.

$(h''(x) \geq 0)$ and $(g''(x) \cdot h'(x) \geq 0)$

(*) Extended real line

$R : \{ -\infty < x < \infty \}$

$\bar{R} = R \cup \{ -\infty, \infty \}$

∞ is part of it

* Extended convex function:

$$\check{f}(x) = \begin{cases} f(x) & x \in \text{dom } f \\ \infty & x \notin \text{dom } f \end{cases}$$

* <u>Perspective transformation</u>

$$g\left[\left(\frac{x}{t}\right)\right] = t \cdot f\left(\frac{x}{t}\right)$$

If f is convex $\Rightarrow$ g convex $(t > 0)$

* Let C be a convex set $\subseteq \mathbb{R}^n$.

$$g(x) = \min_{y \in C} f(x, y) \rightarrow \text{convex.}$$

* Generalized zeroth order definition.

$$f\left(\sum_{i=1} \theta_i x_i\right) \leq \sum_i \theta_i f_i(x_i)$$

$$0 \leq \theta_i \leq 1, \quad \sum \theta_i = 1$$

* <u>Jensens inequality</u> (generalization of zeroth order)

$$f(E[x]) \leq E[f(x)] \rightarrow f \text{ is convex}$$

Convex Optimization

$$x^* = \arg\min_x f_0(x)$$

s.t. $f_i(x) \leq 0$, $i = 1, 2 \ldots p$ → convex inequalities

$h_i(x) = 0$, $i = 1, 2 \ldots m$ → affine equality constraint.

Equivalent Problem

$\min f_0(x)$

$f_i(x) \leq 0 \quad i = 1 \ldots p$

$a_i^T x_i - b_i = 0$, $i = 1 \ldots m$

Jayadeva sir's Optimization

$$\text{Min} \sum_{i=1}^{m} q_i$$

such that,

$$y_i(w^T x^i + b) + q_i \geq 1 , \quad q_i \geq 0$$