LECTURE 2: PRELIMINARIES

1. Basic terminologies
2. Review of background knowledge

Euclidean space E^n

- An *n*-dimensional Euclidean space is a space of elements specified by *n* coordinates in real numbers with emphasis on the structure of Euclidean geometry, such as distance and angle, using the standard “inner product” operation. It is a real vector space with an inner product operation.

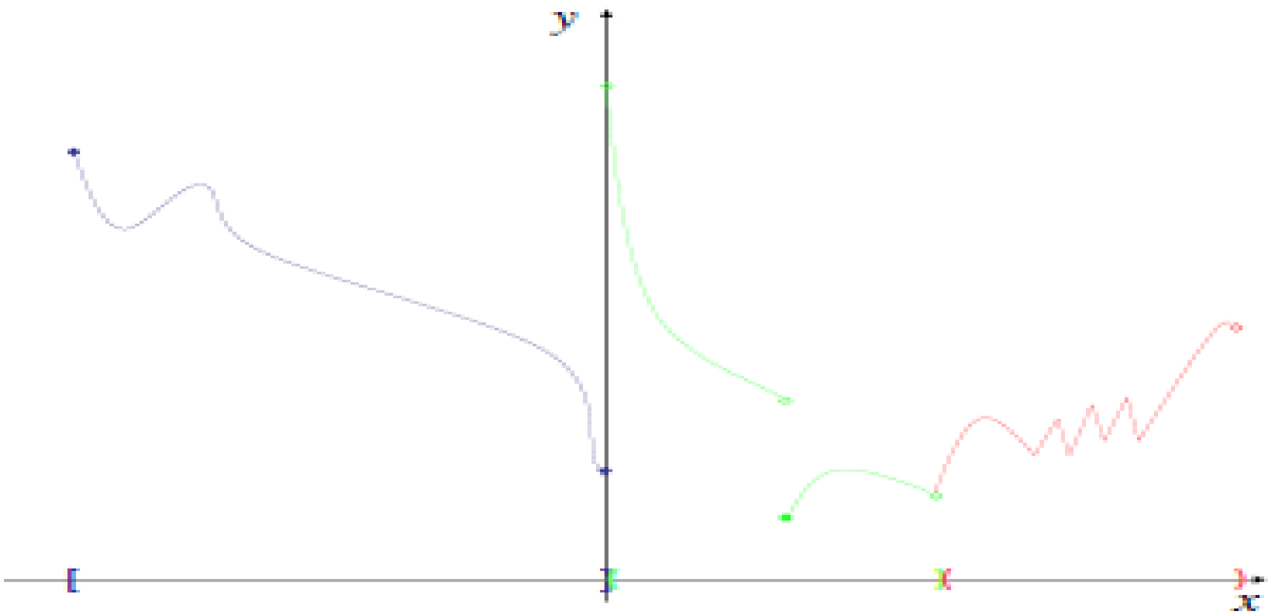
- Notation

  Commonly used 2-norm of a vector x in $\mathbf{E}^n$ is denoted by

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{\sum_{i=1}^{n} (x_i)^2}.$$

  Sometimes by |x| for convenience.

General aspects of sets and functions in E^n



- boundary / interior points
- closed / open sets
- bounded / compact sets
- convex sets

⋮

- continuous functions
- differentiable functions
- convex / concave functions
- Taylor Series

⋮

- sets functions

Unconstrained optimization problem

- Let $f : E^n \to R$ be a real-valued function. Consider

$$\begin{array}{ll} \text{Min} & f(x) \\ \text{s. t.} & x \in S \subset E^n \\ \text{where} & S \subset E^n \text{ is a "}\textit{simple set}\text{".} \end{array}$$

Basic terminologies

(i) $x^* \in S$ is a minimum solution if

$f(x) \geq f(x^*), \forall x \in S.$

We denote

$$f(x^*) = \min_{x \in S} f(x).$$

(ii) $x^* \in E^n$ is an infimum solution if

$f(x^*) =$ greatest lower bound of $f(x)$

over S.

We denote

$$f(x^*) = \inf_{x \in S} f(x).$$

Notations

- E^n: n-dimensional Euclidean space

  (Sometimes we use R for E^1)
- $S(x^0, r) \triangleq \{x \in E^n \big| \; |x - x^0| < r\}$

  (open sphere with center $x^0 \in E^n$ and radius $r > 0$)
- $\overline{S}(x^0, r) \triangleq \{x \in E^n \big| \; |x - x^0| \leq r\}$

  (closed sphere)
- $\text{bdry}\, S(x^0, r) \triangleq \{x \in E^n \big| \; |x - x^0| = r\}$

  (boundary of sphere)

Neighborhood

- $N(x^0)$: neighborhood of x^0 is an open sphere with center at x^0
- $N(x^0; r) \triangleq S(x^0, r)$
- $N'(x^0) \triangleq N(x^0) - \{x^0\}$ (deleted neighborhood)

Open sets

- Let $S \subset E^n$ and $x \in S$, then x is an <u>interior point</u> of S, if $\exists\ N(x) \subset S$.
- $int(S) \triangleq \{\ x \in E^n \mid x$ is an interior point of $S\}$
- $S \subset E^n$ is <u>open</u>, if $S = int(S)$.

Closed sets

- Let $S \subset E^n$ and $x \in E^n$, then x is an accumulation point of S, if

$$N'(x) \cap S \neq \phi, \quad \forall \; N(x).$$

- $acc(S) \triangleq \{x \in E^n \mid x$ is an accumulation point of S $\}$
- $S \subset E^n$ is closed, if $acc(S) \subset S$.
- $\bar{S} \triangleq S \cup acc(S)$ is called the closure of S.

Open and closed sets

- $\mathrm{brdy}(S) \triangleq \bar{S} - \mathrm{int}(S)$

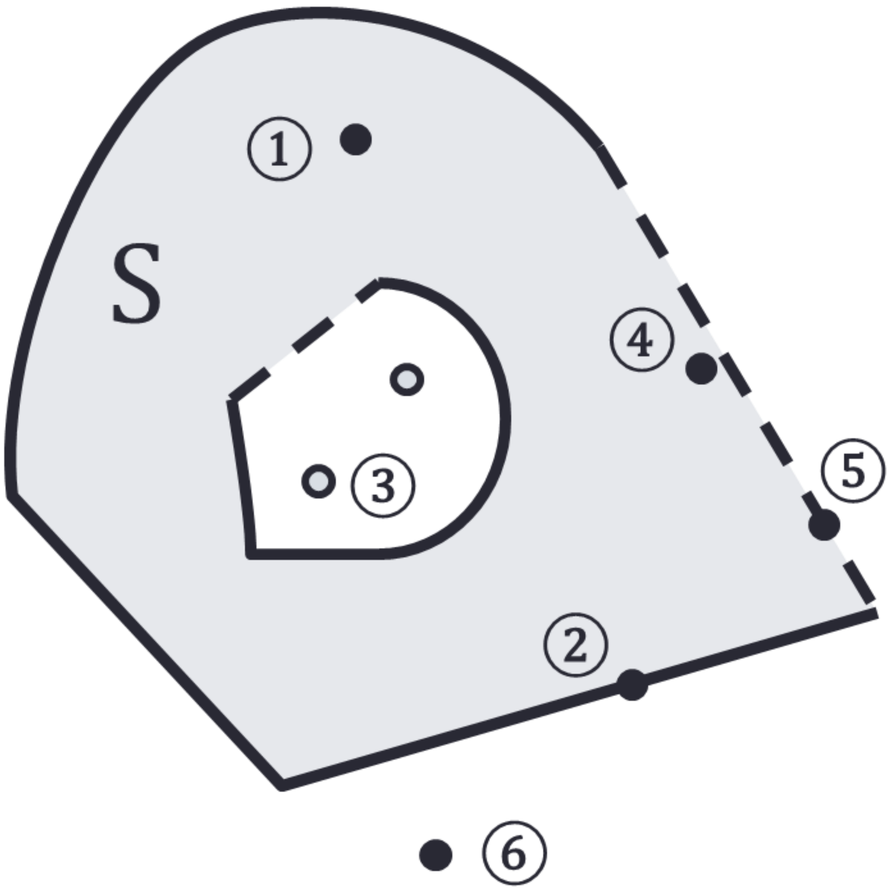
$$= \{\, \mathbf{x} \in E^n \mid \forall N(x), \exists\, \mathbf{y}, \mathbf{z} \in E^n \text{ s.t. } y \in N(x) \cap S, z \notin N(x) \cap S\}$$

- S is closed if and only if $S = \bar{S}$.

- **Theorem:**

  $S \subset E^n$ is closed if and only if $E^n - S$ is open.

Interior, accumulation, boundary points



F : ①②③④

I : ①④

A : ①②④⑤

B : ②③⑤

Relatively open and close

Relatively open and close sets

- Let $A \subset B \subset E^n$ we say A is closed relative to B if $acc(A) \cap B \subset A$. We also say that A is open relative to B if B-A is closed relative to B.

- **Theorem:**

  Let $A \subset B \subset E^n$. Then A is open relative to B, if and only if, $A = B \cap C$ for some C open in E^n.

Various cones

Cones

- Definition:

Let $X \subset E^n$. Then X is cone, if $\lambda x \in X$,

$\forall\ x \in X$ and $\lambda \geq 0$.

Alternative definition:

- Let $X \subset E^n$. Then X is cone, if $\lambda x \in X$, $\forall\ x \in X$ and $\lambda > 0$. X is a pointed cone if $0 \in X$.

Boundedness and compactness

Bounded sets

- $S \subset E^n$ is <u>bounded</u> if $\exists\, r > 0$ such that $S \subset N(0, r)$.
- **Theorem** (Bolzano-Weierstrass)
  If $S \subset E^n$ is bounded and S contains infinitely many points, then $acc(S) \neq \phi$.

- https://en.wikipedia.org/wiki/Bolzano%E2%80%93Weierstrass_theorem

Compact sets

- A collection F of sets is said to be a covering of a given set S, if $S \subset U_{T \in F} T$. When T is open, $\forall\, T \in F$, then F is called an open covering of S.
- **Theorem**(Heine-Borel)
  Let $S \subset E^n$ be closed and bounded and F be an open covering of S. Then $\exists$ a finite subcollection of F that covers S.
- $S \subset E^n$ is compact if, and only if, every open covering of S contains a finite subcollection that also covers S.

Compact sets

- **Theorem:**

  Let $S \subset E^n$. Then the following statements are equivalent:

  (i) S is compact;

  (ii) S is closed and bounded;

  (iii) Every infinite subset of S has an accumulation point in S.

Convex sets and convex hulls

Convex sets

- Let $x^1,\ x^2 \in E^n$, the <u>line segment</u> formed by $x^1$and x^2 is

  $$L(x^1, x^2) \triangleq \{x \in E^n \big| x = \theta x^1 + (1-\theta)x^2 \quad \text{for some} \quad \theta \in [0,1]\}.$$

- $S \subset E^n$ is <u>convex</u> if

  $$L(x^1, x^2) \subset S, \quad \forall x^1, x^2 \in S.$$

Convex hulls

- Let $S \subset E^n$. The <u>convex hull</u> of S is the intersection of all convex sets containing S, i.e.,

$$co(S) = \bigcap_{S \subset T:convex} T$$

- The closure of $co(S)$ is called the <u>closed convex hull</u> of S.

Extreme points

Extreme points

- Let $S \subset E^n$ be convex and $x \in S$. Then x is an <u>extreme point</u> of S, if
  $x \notin L^i(x^1, x^2), \; \forall x^1, x^2 \in S$ and $x^1 \neq x^2$
  where
  $$L^i(x^1, x^2) \triangleq \{x \in E^n \mid x = \theta x^1 + (1-\theta)x^2 \quad \text{for some} \quad \theta \in (0,1)\}$$

Characterization of convex sets

- **Theorem**
  A closed bounded convex set $S \subset E^n$ is equal to the closed convex hull of its extreme points.

Separation and supporting hyperplanes

Theorem (Separating):

Let $S \subset E^n$ be convex and $y \notin \bar{S}$. Then $\exists$ a vector $a \in E^n$

s.t.

$$a^T y < \inf_{x \in S} a^T x$$

Separation and supporting hyperplanes

Theorem (Supporting): Let $S \subset E^n$ be convex and $y \in bdry(S)$. Then $\exists$ a hyperplane containing y and containing S in one of its closed half spaces, i.e., $\exists\, a \in E^n$ and $b \in R$ s.t.

$$a^T y = b \;\; and \;\; a^T x \geq b, \;\; \forall x \in S.$$

Feasible directions

Feasible directions

- Let $S \in E^n$ and $x^0 \in S$. Then $d \in E^n$ is a feasible direction of x^0 in S if $\exists\ \bar{\theta} > 0$ s.t.

$$x^0 + \theta d \ \in \ S, \quad \forall\ \theta \in [0, \bar{\theta}].$$

- **Theorem**

  Let $S \in E^n$ be closed. Then S is convex if and only if, for all $x^0 \in S$, $d \triangleq x - x^0$ is a feasible direction of x^0 in S, $\forall\ x \in S$ and $x \neq x^0$.

Continuous functions

- Let $f : S \subset E^n \to E^m$ and $x^0 \in acc(S)$. Then f is continuous at x^0 provided that

  (i) f is defined at x^0

  (ii) $\lim_{x \to x^0} f(x) = f(x^0)$

  If f is continuous on every point of S, we say f is continuous on S, i.e., $f \in C(S)$.

Characterization of continuous functions

- **Theorem**

  Let $f : S \subset E^n \to E^m$ and

  $T = f(S) = \{y \in E^m \mid y = f(x) \text{ for some } x \in S\}$.

  Then the following statements are equivalent:

  (i) $f \in C(S)$;

  (ii) If Y is open relative to T, then $f^{-1}(Y)$ is open relative to S ;

  (iii) If Y is closed relative to T, then $f^{-1}(Y)$ is closed relative to S.

Optimization of continuous functions

- **Theorem(Bolzano)**

  Let $f : E^n \rightarrow R$ be continuous and $S \subset E^n$be

  compact. Then f achieves both its maximum

  and minimum on S, and $f(S)$ is compact.

Differentiable functions

- Let $f: S \subset E^n \to R$ and S is open . Then $f \in C^1(S)$ means its first partial derivatives are continuous at each point of S. We denote its gradient as

$$\nabla f(x) = [\frac{\partial f(x)}{\partial x_1}, \cdots, \frac{\partial f(x)}{\partial x_n}]_{1\times n}$$

- Similarly, $f \in C^p(S)$ means its p-th partial derivatives are continuous at each point of S.

Differentiable functions

- If $f \in C^2(S)$, we denote its Hessian as

$$F(x) = [\frac{\partial^2 f(x)}{\partial x_i x_j}]_{n \times n}$$

which is a symmetric square matrix of dimensionality n.

Differentiable functions

- Let $f = (f_1, f_2, \cdots, f_m)$ and $f_i : E^n \to R$ be real-valued function. If $f_i \in C^p$, $\forall\ i = 1, \ldots,$ m, then we say $f \in C^p$.
- If $f \in C^1$, we define

$$\nabla f(x) \triangleq [\frac{\partial f_i}{\partial x_j}]_{m \times n}$$

- If $f \in C^2$, its Hessian

$$F(x) \triangleq (F_1(x), F_2(x), \cdots, F_m(x))$$

is a third-order tensor.

Taylor theorem – 1 dimensional case

Assume that

$$f : R \rightarrow R, \quad f \in C^n([a,b]), \quad x_0 \in [a,b].$$

Then $\forall\ x \in [a,b]$ and $x \neq x_0$,

$$\exists \quad x_1 = \theta x_0 + (1-\theta)x \ \ with \ \ \theta \in (0,1)$$

s.t.

$$f(x) = f(x_0) + \sum_{k=1}^{n-1} \frac{f^k(x_0)}{k!}(x-x_0)^k + \frac{f^n(x_1)}{n!}(x-x_0)^n$$

Approximation

- $f \in C^1$

$$f(x) = f(x_0) + f'(x_1)(x - x_0)$$

- $f \in C^2$

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_1)(x - x_0)^2$$

- When $x \approx x_0$

$$f(x) = f(x_0) + \sum_{k=1}^{n-1} \frac{f^k(x_0)}{k!}(x - x_0)^k + o(\cdot)$$

$$\approx f(x_0) + \sum_{k=1}^{n-1} \frac{f^k(x_0)}{k!}(x - x_0)^k$$

Taylor Theorem – *n* dimensional case

Let $f : E^n \to R$, $\quad S \subset E^n$ be open, $\quad f \in C^m(S)$,
$x^1, x^2 \in S$, $\quad x^1 \neq x^2$ and $L(x^1, x^2) \subset S$.

Then $\exists \ \bar{x} = \theta x^1 + (1-\theta)x^2 \in L^i(x^1, x^2)$ s.t.

$$f(x^2) = f(x^1) + \sum_{k=1}^{m-1} \frac{1}{k!} d^k f(x^1, x^2 - x^1)$$

$$+ \frac{1}{m!} d^m f(\bar{x}; x^2 - x^1)$$

where $d^k f(x; t)$ is the k-th order differential of function f along t.

Taylor Theorem

- $f \in C^1$

$$f(x^2) = f(x^1) + \nabla f(\bar{x})(x^2 - x^1)$$

- $f \in C^2$

$$f(x^2) = f(x^1) + \nabla f(x^1)(x^2 - x^1)$$

$$+\frac{1}{2}(x^2 - x^1)^T F(\bar{x})(x^2 - x^1)$$

Approximation

When $x \approx x^1$

$$f(x) \approx f(x^1) + \sum_{k=1}^{m-1} \frac{1}{k!} d^k f(x^1; x - x^1)$$

Take $m = 2$

$$f(x) \approx f(x^1) + \nabla f(x^1)(x - x^1)$$

Assume $\nabla f(x^1) \neq 0$.

- Take $x - x^1 = \nabla f(x^1)$, i.e., moving from x^1 in the gradient direction at x^1

$$f(x) \approx f(x^1) + \|\nabla f(x^1)\|^2 > f(x^1)$$

Approximation

- For $x - x^1 = -[\nabla f(x^1)]$, i.e., moving from x^1 in the negative gradient direction

$$f(x) \approx f(x^1) - \|\nabla f(x^1)\|^2 < f(x^1)$$

- For any $d \triangleq x - x^1$

$$\nabla f(x^1)(x - x^1) = \|d\| \underbrace{\|\nabla f(x^1)\| \cos\theta}_{\text{projection of } \nabla f(x^1) \text{ onto } d}$$

Approximation

- Take $m = 3$

$$f(x) \approx f(x') + \nabla f(x')(x - x') + \frac{1}{2}(x - x')^T F(x')(x - x')$$

  - If $\nabla f(x') = 0$ and $F(x')$ is positive definite, then $f(x) \approx f(x') + \frac{1}{2}(x - x')^T F(x')(x - x') > f(x')$
  - If $\nabla f(x') = 0$ and $F(x')$ is negative definite, then $f(x) \approx f(x') + \frac{1}{2}(x - x')^T F(x')(x - x') < f(x')$

Big O and small o

Let $g(\cdot)$ be a real-valued function on R.

(1) If $g(x)$ goes to zero at least as fast as x does, i.e., $\exists\ c \geq 0$ such that

$$\left|\frac{g(x)}{x}\right| \leq c \quad \text{as } x \longrightarrow 0,$$

then we say $g(x) = O(x)$.

(2) If $g(x)$ goes to zero faster than x does, i.e.,

$$\left|\frac{g(x)}{x}\right| = 0 \quad \text{as } x \longrightarrow 0,$$

then we say $g(x) = o(x)$.