Introduction to Optimization

Jun Moon
junmoon@hanyang.ac.kr

Note 3

Outline

- Basic real analyis and calculus

Sets

- Sets

$$\mathbb{R}, \mathbb{R}^n, \mathbb{Q}, \mathbb{N},$$

- n-dimensional Euclidean space

$$\mathbb{R}^n$$

Functions

- Mapping, operator, etc

- Let $X \subset \mathbb{R}^n$

  $Y \subset \mathbb{R}^m$

- A function is a mapping that assigns the element of X to the element of Y, which is written by

$$f : X \to Y$$

Functions

- A function f

$$f : X \to Y$$

- X: domain
- Y: range or codomain

Functions

- Examples

$$f(x) = \cos(x),\ \ f(x) = x^2$$

$$f(x_1, x_2) = x_1^2 + x_2^2,\ \ f(x_1, x_2, x_3) = \cos(x_1 + x_2) + x_3^2$$

$$f(x_1, x_2) = \begin{bmatrix} \cos(x_1 + x_2) + x_3^2 \\ x_1^2 + x_2^2 \end{bmatrix}$$

Function Differentiation

- Real-valued function f

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

- The differentiation of f

$$\nabla f(x) : \mathbb{R} \to \mathbb{R}$$

Function Differentiation

- Real-value function with n-dimensional domain

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

$$\nabla f(x) : \mathbb{R}^n \to \mathbb{R}^n$$

- Differentiation

$$\nabla f(x) : \begin{bmatrix} \partial_{x_1} f(x) \\ \partial_{x_2} f(x) \\ \vdots \\ \partial_{x_n} f(x) \end{bmatrix}$$

Function Differentiation

- Real-value function with n-dimensional domain (우리는 사용하지 않습니다)

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

- Differentiation

$$\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^{m \times n}$$

Function Differentiation

Jacobian matrix

$$\nabla f(x) = \begin{bmatrix} \nabla f_1(x)^\top \\ \vdots \\ \nabla f_m(x)^\top \end{bmatrix} = \begin{bmatrix} \partial_{x_1} f_1(x) & \cdots & \partial_{x_n} f_1(x) \\ \vdots & \ddots & \vdots \\ \partial_{x_1} f_m(x) & \cdots & \partial_{x_n} f_m(x) \end{bmatrix}$$

Functions

- Examples

$$f(x) = \cos(x),\ \ f(x) = x^2$$

- Differentiation?

Functions

- Examples

$$f(x_1, x_2) = x_1^2 + x_2^2, \;\; f(x_1, x_2, x_3) = \cos(x_1 + x_2) + x_3^2$$

- Differentiation?

Functions

- Examples

$$f(x_1, x_2) = \begin{bmatrix} \cos(x_1 + x_2) + x_3^2 \\ x_1^2 + x_2^2 \end{bmatrix}$$

- Differentiation?

Basic Mathematics

Basic Mathematics

- vector space = linear space = linear vector space
- A linear space over a field $\mathbb{F}$, $(\mathbb{V}, \mathbb{F})$, consists of a set $\mathbb{V}$ of vectors, a field $\mathbb{F}$, and two operations, vector addition and scalar multiplication
- The two operations satisfy

Vector addition and scalar multiplication

Basic Mathematics

- vector space = linear space = linear vector space
- A linear space over a field $\mathbb{F}$, $(\mathbb{V}, \mathbb{F})$, consists of a set $\mathbb{V}$ of vectors, a field $\mathbb{F}$, and two operations, vector addition and scalar multiplication
- The two operations satisfy

Multiplication

(a) multiplication: for any $\alpha \in \mathbb{F}$ and $x \in \mathbb{V}$, $\alpha x \in \mathbb{V}$

(b) associative: for any $\alpha, \beta \in \mathbb{F}$ and $x \in \mathbb{V}$, $\alpha(\beta x) = (\alpha\beta)x$

(c) distributive w.r.t. scalar addition:

for any $\alpha \in \mathbb{F}$ and $x, y \in \mathbb{V}$, $\alpha(x + y) = \alpha x + \alpha y$

(d) distributive w.r.t. scalar multiplication

for any $\alpha, \beta \in \mathbb{F}$ and $x \in \mathbb{V}$, $(\alpha + \beta)x = \alpha x + \beta x$

(e) there exists a unique $1 \in \mathbb{F}$ such that for any $x \in \mathbb{V}$, $1x = x$

(f) there exists a unique $0 \in \mathbb{F}$ such that for any $x \in \mathbb{V}$, $0x = 0$

Basic Mathematics

- vector space = linear space = linear vector space
- A linear space over a field $\mathbb{F}$, $(\mathbb{V}, \mathbb{F})$, consists of a set $\mathbb{V}$ of vectors, a field $\mathbb{F}$, and two operations, vector addition and scalar multiplication

Example: $(\mathbb{F}^n, \mathbb{F})$ where $\mathbb{F}^n = \mathbb{F} \times \cdots \times \mathbb{F}$

Example: $(\mathbb{R}^n, \mathbb{R})$, $(\mathbb{C}^n, \mathbb{C})$, $(\mathbb{C}^n, \mathbb{R})$

Example: $(\mathbb{R}, \mathbb{C})$ is not a vector space! (why?) $(1 + i)1 = 1 + i \notin \mathbb{R}$

Example: a continuous function $f : [t_0, t_1] \to \mathbb{R}^n$, the set of such functions, $(C([t_0, t_1], \mathbb{R}^n), \mathbb{R})$, is a linear space

In this course, the n-dimensional real vector space, $(\mathbb{R}^n, \mathbb{R})$, will be considered

Basic Mathematics

Normed vector space (Normed linear space): Length of the vector

A function $\|x\| : \mathbb{R}^n \to \mathbb{R}$ is said to be a norm if the following properties hold

- $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$ (separate points)
- $\|\alpha x\| = |\alpha|\|x\|$ (absolute homogeneity)
- $\|x + y\| \leq \|x\| + \|y\|$ (triangular inequality)

Basic Mathematics

Example: The norm can be chosen as

$$\|x\|_1 := \sum_{i=1}^{n} |x_i|, \ \|x\|_2 := \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}, \ \|x\|_\infty := \max_i |x_i|$$

Example: signal norm for the real-valued continuous function $f(t)$

$$\|f\|_p = \left(\int_0^t |f(t)|^p dt\right)^{1/p}$$

where $1 \leq p < \infty$

Basic Mathematics

Inner Product: measure angle of two vectors
An inner product between two vectors, $\langle x, y\rangle$, on the vector space $(\mathbb{R}^n, \mathbb{R})$ is a function that maps from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$ such that the following properties hold

- $\langle x, y\rangle = \langle y, x\rangle$
- $\langle x, \alpha_1 y_1 + \alpha_2 y_2\rangle = \alpha_1\langle x, y_1\rangle + \alpha_2\langle x, y_2\rangle$
- $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ if and only if $x = 0$

Basis Mathematics

Example: Let $(\mathbb{R}^n, \mathbb{R})$. Then the inner product is

$$\|x\|_2^2 = \langle x, x\rangle = \sum_{i=1}^{n} |x_i|^2, \quad \langle x, y\rangle = x^T y = \sum_{i=1}^{n} x_i y_i$$

Example: signal inner product for the real-valued continuous function $f(t)$

$$\|f\|_2^2 = \int_0^t |f(t)|^2 dt, \ \langle f, g\rangle = \int_0^t f(t)g(t)dt$$

where $1 \leq p < \infty$

Basis Mathematics

Fact (no proof): If $\|\cdot\|$ and $\|\cdot\|'$ are two norms on $\mathbb{R}^n$, then there exists constants $c_1, c_2 > 0$ such that

$$c_1\|x\|' \leq \|x\| \leq c_2\|x\|' \quad \forall x \in \mathbb{R}^n$$

- The ratio between two norms is bounded below and above, independent of x
- The above fact implies that in the finite-dimensional space, we can use any norm to define convergence and continuity on $\mathbb{R}^n$

Basic Mathematics

Definition: A real-valued sequence $x^1, x^2, \ldots \in \mathbb{R}$ converges to x^* if for each $\epsilon > 0$, there exists $K = K(\epsilon)$ such that

$$|x^k - x^*| < \epsilon \quad \forall k \geq K,$$

and we write

$$\lim_{k \to \infty} x^k = x^*$$

or

$$x^k \to x^* \text{ as } k \to \infty$$

Basic Mathematics

We say that $\{x^k\}$ is a sequence in $\mathbb{R}^n$ if $x^k \in \mathbb{R}^n$ for all k.

Definition: A sequence $\{x^k\}$ in $\mathbb{R}^n$ converges to x^* if and only if every component of x^k converges, that is

$$\lim_{k\to\infty} x_i^k = x_i^*, \; i = 1, 2, \ldots, n$$

$$x^k = \begin{pmatrix} x_1^k \\ x_2^k \\ \vdots \\ x_n^k \end{pmatrix}, \; x^* = \begin{pmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{pmatrix}$$

We write $\lim_{k\to\infty} x^k = x^*$.

Basic Mathematics

Equivalently, $\lim_{k\to\infty} x^k = x^* \Leftrightarrow$ for each $\epsilon > 0$, there exists K such that

$$\|x^k - x^*\| < \epsilon \quad \forall k \geq K$$

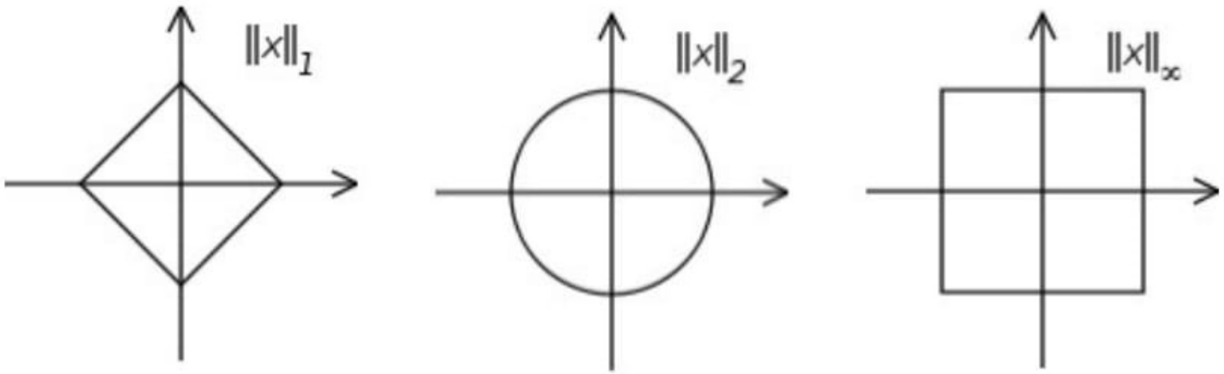
We can use any norm $\|\cdot\|$ due to the previous fact

Basic Mathematics

A ball $B_\delta(x)$ with a radius of $\delta > 0$ is a subset of $\mathbb{R}^n$, defined by

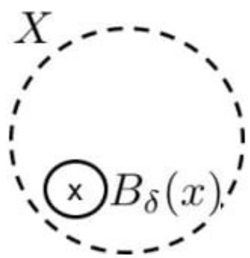
$$B_\delta(x) = \{s \in \mathbb{R}^n \ : \ \|x - s\| \leq \delta\}$$

Note that the set $B_\delta(x)$ depends on the norm $\|\cdot\|$ used



Basic Mathematics

Definition: A set $X \in \mathbb{R}^n$ is open if, for each $x \in X$, there exists $\delta > 0$ such that there is a ball $B_\delta(x)$ around x such that $B_\delta(x) \subseteq X$.



Basic Mathematics

Definition: A set X is closed if X^c is open. Equivalently, X is closed if and only if it contains the limit of every convergent sequence in X.

Example: $(0, 1)$ is open

Example: $[0, 1]$ is closed

Example: $(0, 1]$ is neither open nor closed. Why? think about a sequence $\{\frac{1}{k}\}$

Basic Mathematics

A set $X \subset \mathbb{R}^n$ is said to be compact if X is closed and bounded

- Known as Heine-Borel theorem

- It is also necessary condition, i.e., If S is compact, then it is closed and bounded.

Hence $X \subset \mathbb{R}^n$ is compact if and only if X is closed and bounded

Basic Mathematics

Examples

- $[0, 1]$: closed and bounded
- $[0, 1)$: bounded but not closed
- $[0, \infty)$: closed but not bounded

Basic Mathematics

Definition: A set $X \subseteq \mathbb{R}^n$ is bounded if there exists $M < \infty$ such that

$$\|x\| \leq M, \ \forall x \in X.$$

Basic Mathematics

Definition: A function $f : X \to \mathbb{R}$ where $X \subseteq \mathbb{R}^n$ is said to be continuous at $x^* \in X$ if for every sequence $\{y^k\}$ with $\lim_{k\to\infty} y^k = x^*$, we have

$$\lim_{k\to\infty} f(y^k) = f(\lim_{k\to\infty} y^k) = f(x^*)$$

Equivalently, for each $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in X$,

$$\|x - x^*\| < \delta \to |f(x) - f(x^*)| < \epsilon$$

Function of Several Variables

A function f on $\mathbb{R}^n$

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

f is said to be differentiable at a if the partial derivative of f exists for all coordinates, i.e.,

$$\nabla f(a) = \begin{bmatrix} \partial_{x_1} f(a) \\ \vdots \\ \partial_{x_n} f(a) \end{bmatrix}$$

Note that

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

Function of Several Variables

Example

$$f(x_1, x_2) = 5x_1^2 + 6x_1x_2 + 10x_2^2$$

Note that $f : \mathbb{R}^2 \to \mathbb{R}$. Compute

$$\nabla f(x_1, x_2)$$

Function of Several Variables

A function f on $\mathbb{R}^n$

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

f is said to be twice differentiable at a if $\nabla f(a)$ exists and the second-order partial derivative of f exists for all coordinates, i.e.,

$$H(f(a)) = \nabla^2 f(a) = \begin{bmatrix} \partial_{x_1 x_1} f(a) & \cdots & \partial_{x_1 x_n} f(a) \\ \vdots & \ddots & \vdots \\ \partial_{x_n x_1} f(a) & \cdots & \partial_{x_n x_n} f(a) \end{bmatrix}$$

Note that

$$H = \nabla^2 f : \mathbb{R}^n \to \mathbb{R}^{n \times n}$$

Function of Several Variables

Example

$$f(x_1, x_2) = 5x_1^2 + 6x_1x_2 + 10x_2^2$$

Note that $f : \mathbb{R}^2 \to \mathbb{R}$. Compute

$$H = \nabla^2 f(x_1, x_2)$$

Matrix

A square $n \times n$ matrix A is said be symmetric if $A = A^T$, where A^T is transpose of A.

- Hessian H is always symmetric

A symmetric matrix A is said to be positive semi-definite if for any $v \in \mathbb{R}^n$

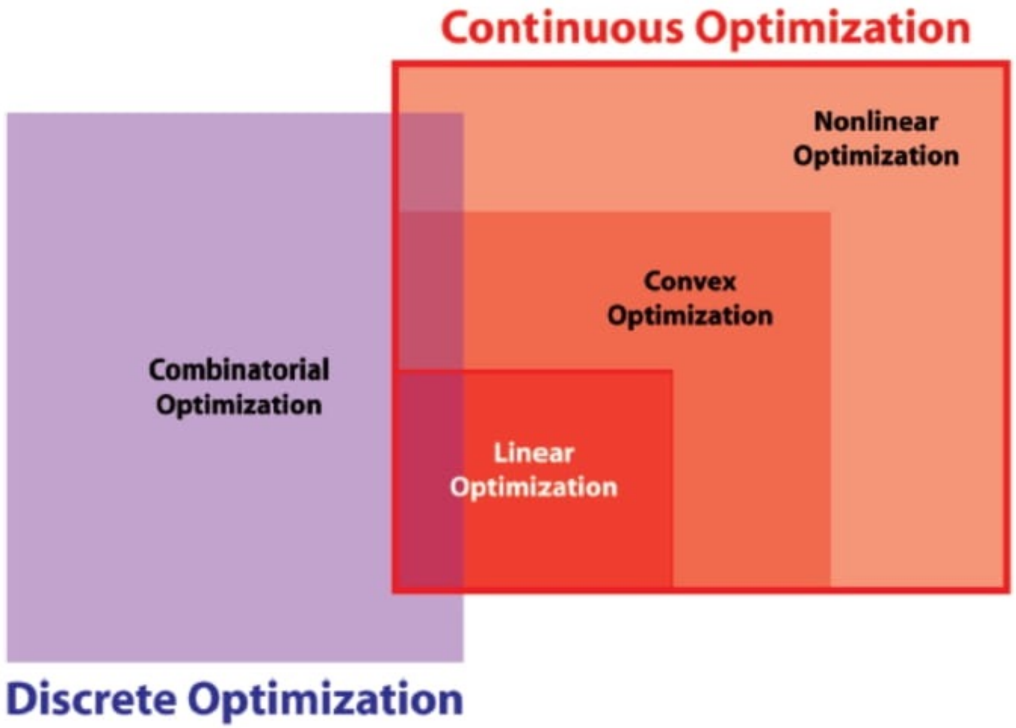
$$v^T A v \geq 0$$

- If eigenvalues (real part) of A are nonnegative, then A is positive semidefinite.
- Other conditions...: (principal eigenvalue... etc) see the textbook on linear algebra

Convex Sets and Convex Functions

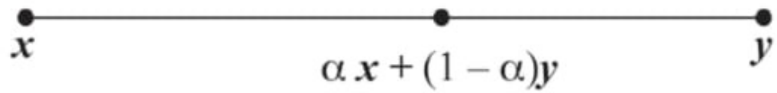
Convex Optimization?

Convex optimization: A special class of nonlinear optimization that includes *Linear Programming*

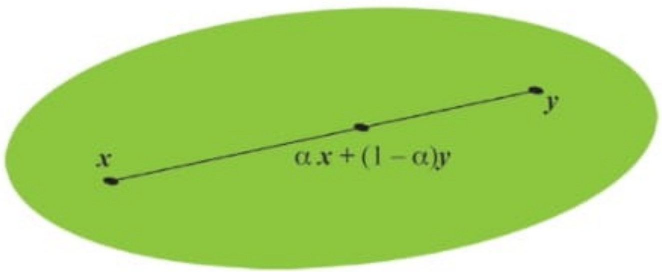


Convex Set

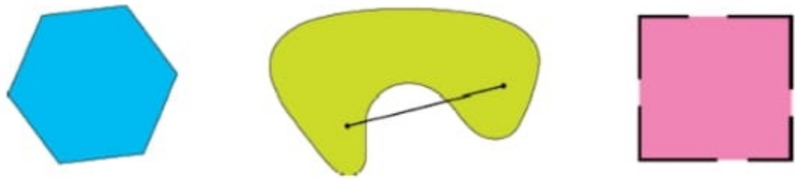
- A line segment defined by vectors $x, y \in \mathbb{R}^n$ is
  the set of points of the form $\alpha x + (1 - \alpha y)$ for $\alpha \in [0, 1]$



- A set $C \subset \mathbb{R}^n$ is convex when, with any two vectors x and y that belongs to the set C, the ling segment connecting x and y also belongs to C



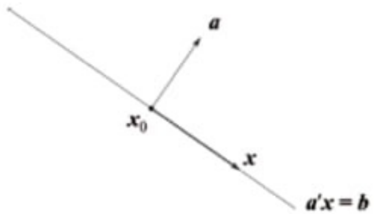
Examples of Convex Sets

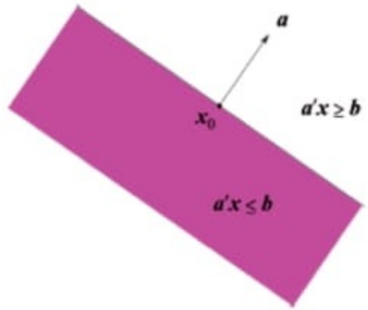


Which of the following sets are convex?

- A line through two given vectors x and y:
  $l(x, y) = \{z \mid z = x + t(y - x),\ t \in \mathbb{R}\}$
- A ray defined by a vector $\{z \mid z = \lambda x,\ \lambda \geq 0\}$
- The positive orthant $\{x \in \mathbb{R}^n \mid x \geq 0\}$ (componentwise inequality)
- Convex cone C: for any $x_1, x_2 \in C$ and $\theta_1, \theta_2 \geq 0$, $\theta_1 x_1 + \theta_2 x_2 \in C$
- Any convex set is connected but not vice versa
- Any subspace is affine, and a convex cone (hence convex)

Examples of Convex Sets: Hyperplanes and Half-spaces



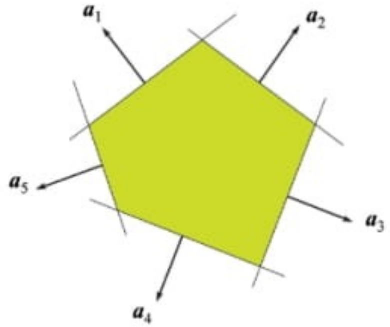


- Hyperplane is a set of the form $\{x \mid a^\top x = b\}$ for a nonzero vector a
- Half-space is a set of the form $\{x \mid a^\top x \leq b\}$ for a nonzero vector a
- A hyperplane in $\mathbb{R}^n$ divides the space into two half spaces:

$$\{x \mid a^\top x \leq b\} \quad \{x \mid a^\top x \geq b\}$$

  It is known as the separating hyperplane (related to duality in optimization)
- Half spaces are convex
- Hyperplanes are convex and affine

Examples of Convex Sets: Polyhedral Sets



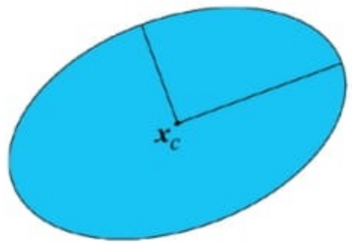
- A polyhedral set is given by finitely many linear inequalities

$$C = \{x \mid Ax \leq b\}, \ A \in \mathbb{R}^{m \times n}$$

- The polyhedral set is intersection of a finite number of half spaces and hyperplane
- Every polyhedral set is convex
- Bounded polyhedral is called polytope
- Linear program

$$\min_x c^\top x \quad \text{subject to } Bx \leq b, \ Dx = d$$

Examples of Convex Sets: Ellipsoid

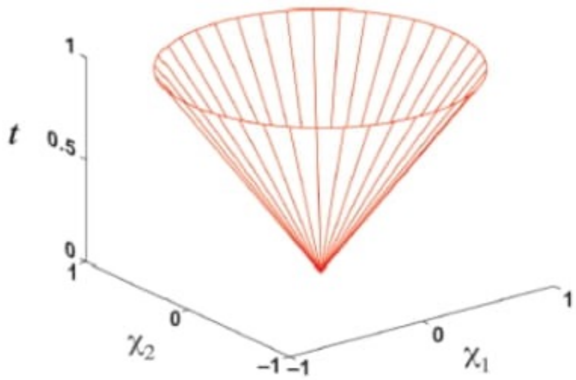


- An ellipsoid is a set of the form

$$\{x \mid (x - x_o)^\top P^{-1}(x - x_o) \leq 1\}, \; P = P^\top > 0$$

- x_o: center of the ellipsoid
- A ball is the case when $P = I$
- Ellipsoids are convex

Examples of Convex Sets: Norm Cones



- A norm cone is the set of the form

$$C = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid \|x\| \le t\}$$

- The norm $\|\cdot\|$ can be any norm in the finite-dimensional space
- The norm cone for Euclidean norm is also known as ice-cream cone
- Any norm cone is convex

Examples of Convex Sets: Simplex

- A simplex is a set given as a convex combination of a finite collection of vectors $x_0, x_1, \ldots, x_m$:

$$C = \text{conv}\{x_0, x_1, \ldots, x_m\}$$

- Examples
  - Unit simplex: $\{x \in \mathbb{R}^n \mid x \geq 0,\ e^\top x \leq 1\}$, $e = (1, \ldots, 1)^\top$
  - Probability simple

Convex Functions

- Let $f : \mathbb{R}^n \to \mathbb{R}$. The domain of f is a set defined by

$$X = \mathrm{dom}(f) = \{x \in \mathbb{R}^n \mid f(x) \text{ is well defined (finite)}\} \subset \mathbb{R}^n$$

- Def: A function f is a convex function if
  - X is a convex set in $\mathbb{R}^n$
  - For any $x_1, x_2 \in X$ and $\alpha \in (0, 1)$

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2)$$

- Strict inequality: strictly convex, i.e.,

$$f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha)f(x_2)$$

Examples of Convex Functions

Convex functions

- Affine: $ax + b$ over $\mathbb{R}$ for any $a, b \in \mathbb{R}^n$
- Any norms in the finite-dimensional space are convex
- Exponential: e^{ax} over $\mathbb{R}$ for any $a \in \mathbb{R}$
- Power: x^p over $(0, \infty)$ for $p \geq 1$ or $p \leq 0$
- Powers of absolute values: $|x|^p$ over $\mathbb{R}$ for $p \geq 1$
- Negative entropy: $x \ln x$ over $(0, \infty)$

Concave

- Affine: $ax + b$ over $\mathbb{R}$ for any $a, b \in \mathbb{R}^n$
- Powers: x^p over $(0, \infty)$ for $0 \leq p \leq 1$
- Logarithm: $\ln x$ over $(0, \infty)$

Verifying Convexity of a Function

We can verify that a given function f is convex by

- Using the definition of the convex function

- Applying some special criteria provided that the function has some nice properties
  - Second-order conditions
  - First-order conditions

Second-Order Conditions

- Assume that f is twice differentiable on $\text{dom}(f)$
- The Hessian $\nabla^2 f(x)$ is a symmetric $n \times n$ matrix whose entries are the second-order partial derivatives of f at x:

$$\left[\nabla^2 f(x)\right]_{ij} = \frac{\partial^2 f(x)}{\partial x_i x_j} \qquad \text{for } i, j = 1, \dots, n$$

- 2nd-order condition: For a twice differentiable function f with the convex domain
  - f is convex if and only if

$$\nabla^2 f(x) \geq 0 \quad \forall x \in X$$

    That is, the Hessian is positive semi-definite
  - f is strictly convex when we have strict inequality, i.e., the Hessian is positive definite

Second-Order Conditions: Examples

- Quadratic function: $f(x) = \frac{1}{2}x^\top Px + q^\top x + r$ with $P = P^\top$. Note

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

  Hence f is convex if and only if $P \geq 0$

- Least-square: $f(x) = \|Ax - b\|^2$ with $A \in \mathbb{R}^{m \times n}$

$$\nabla^2 f(x) = 2A^\top A$$

  Note $A^\top A \geq 0$; hence, it is a convex function

First-Order Condition

- f is differentiable if $\text{dom}(f)$ is open and the gradient of f

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

  exists at each $x \in \text{dom}(f)$

- 1st-order condition: f is convex if and only if its domain is convex and

$$f(x) + \nabla f(x)^\top (z - x) \leq f(z) \quad \forall x, z \in X$$

- A first-order approximation is a global underestimate of f
- Very important property used in algorithm designs and performance analysis

Conclusions

- Basic mathematics