

香 港 大 學
THE UNIVERSITY OF HONG KONG

# Classification of Optimization Problems

**Prof. Li Jiayang**

**DASE7113 Optimization Methods for Intelligent Systems**

# About Optimization

## What is optimization?

**A quest for the best solution to achieve certain objective given certain constraints.**

| | | | |
|---|---|---|---|
| **1. Decision variable:** | What you are optimizing? | → | $x$ |
| **2. Objective function:** | What you are optimizing for? | → | $f(x)$ |
| **3. Constraints:** | Rules to follow. | → | $g_i(x) \geq 0, i = 1, \ldots, m$<br>$h_i(x) = 0, i = 1, \ldots, p$ |

# About Optimization

## What is optimization?

**A quest for the best solution to achieve certain objective given certain constraints.**

**1. Decision variable:**

**2. Objective function:**

**3. Constraints:**

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \ \ i = 1, \ldots, m \\ & h_i(x) = 0, \ i = 1, \ldots, p \end{aligned}$$

# Feasible Region

## Definition

**The feasible region of an optimization problem is the set of decision variables that satisfy all constraints:** $\mathcal{A} = \{x: g_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p\}$

**Note that** $\mathcal{A} = \mathcal{G} \cap \mathcal{H}$, **where** $\mathcal{G} = \cap_{i=1}^{m} \mathcal{G}_i$ **and** $\mathcal{H} = \cap_{i=1}^{p} \mathcal{H}_i$

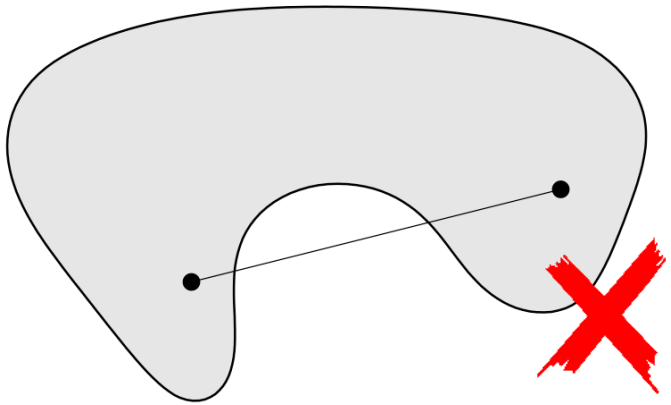
$\mathcal{G}_i = \{x: g_i(x) \leq 0\}$ $\mathcal{H}_i = \{x: h_i(x) = 0\}$

# Convex Set

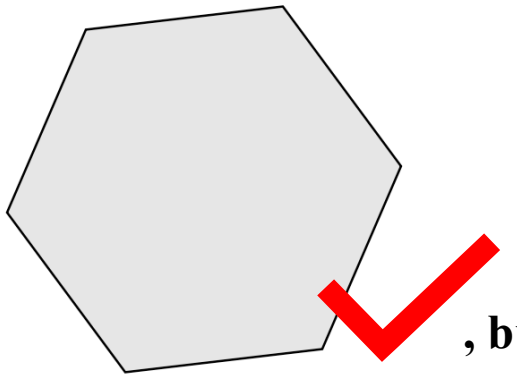
## Definition

A set $\mathcal{A} \in \mathbb{R}^n$ is convex if the line segment between any two points in $\mathcal{A}$ lies in $\mathcal{A}$,

i.e., for any $x, y \in \mathcal{A}$ and $\lambda \in (0, 1)$, we have $\lambda \cdot x + (1 - \lambda) \cdot y \in \mathcal{A}$.



Polyhedron



, but how to prove this?

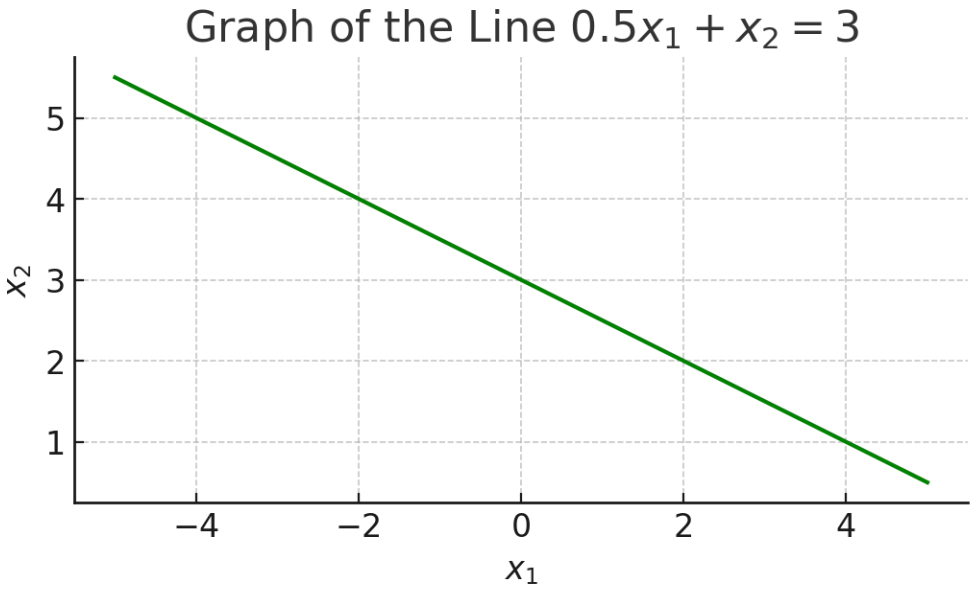
# Convex Set

## Example

**Hyperplane:** $\mathcal{H} = \{x \in \mathbb{R}^n : a^T x = b\}$, **where** $a \in \mathbb{R}^n$, $a \neq 0$, **and** $b \in \mathbb{R}$.

- **In 2-dimensional space, a hyperplane is a line.**

  $\{x \in \mathbb{R}^2 : a_1 x_1 + a_2 x_2 = b\}$.



# Convex Set

## Example

**Hyperplane:** $\mathcal{H} = \{x \in \mathbb{R}^n : a^T x = b\}$, **where** $a \in \mathbb{R}^n$, $a \neq 0$, **and** $b \in \mathbb{R}$.

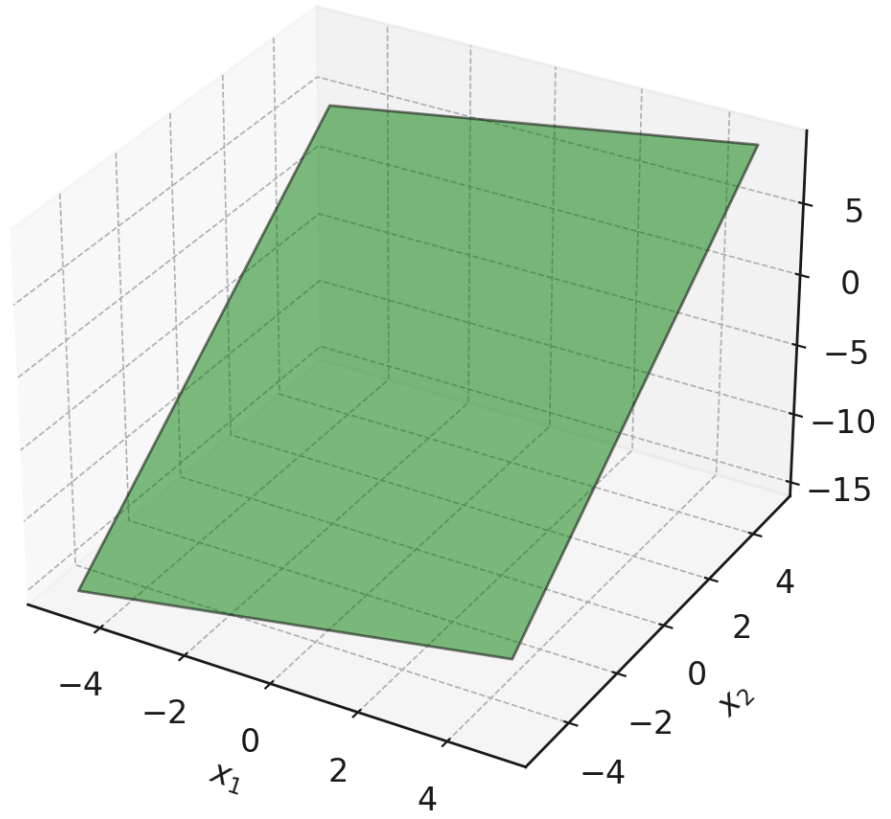
- In 2-dimensional space, a hyperplane is a line.

  $\{x \in \mathbb{R}^2 : a_1 x_1 + a_2 x_2 = b\}$.

- **In 3-dimensional space, a hyperplane is a plane.**

  $\{x \in \mathbb{R}^3 : a_1 x_1 + a_2 x_2 + a_3 x_3 = b\}$.

Graph of the Plane $0.5x_1 + 2x_2 - x_3 = -3$



# Convex Set

## Question

**How to prove any $\mathcal{H} = \{x \in \mathbb{R}^n : a^{\mathrm{T}}x = b\}$, where $a \in \mathbb{R}^n$, $a \neq 0$, and $b \in \mathbb{R}$, is a convex set?**

**Proof. Given any two points $x, y \in \mathcal{H}$ and $\lambda \in (0, 1)$, we have**

$$a^{\mathrm{T}}(\lambda \cdot x + (1 - \lambda) \cdot y) = \lambda \cdot a^{\mathrm{T}}x + (1 - \lambda) \cdot a^{\mathrm{T}}y = \lambda \cdot b + (1 - \lambda) \cdot b = b.$$

**Thus, $\lambda \cdot x + (1 - \lambda) \cdot y \in \mathcal{H}$.**

# Convex Set

## Example

**Halfspace:** $\mathcal{L} = \{x \in \mathbb{R}^n : a^T x \leq b\}$, **where** $a \in \mathbb{R}^n$, $a \neq 0$, **and** $b \in \mathbb{R}$.

# Convex Set

## Question

**How to prove any $\mathcal{L} = \{x \in \mathbb{R}^n : a^T x \leq b\}$, where $a \in \mathbb{R}^n$, $a \neq 0$, and $b \in \mathbb{R}$, is a convex set.**

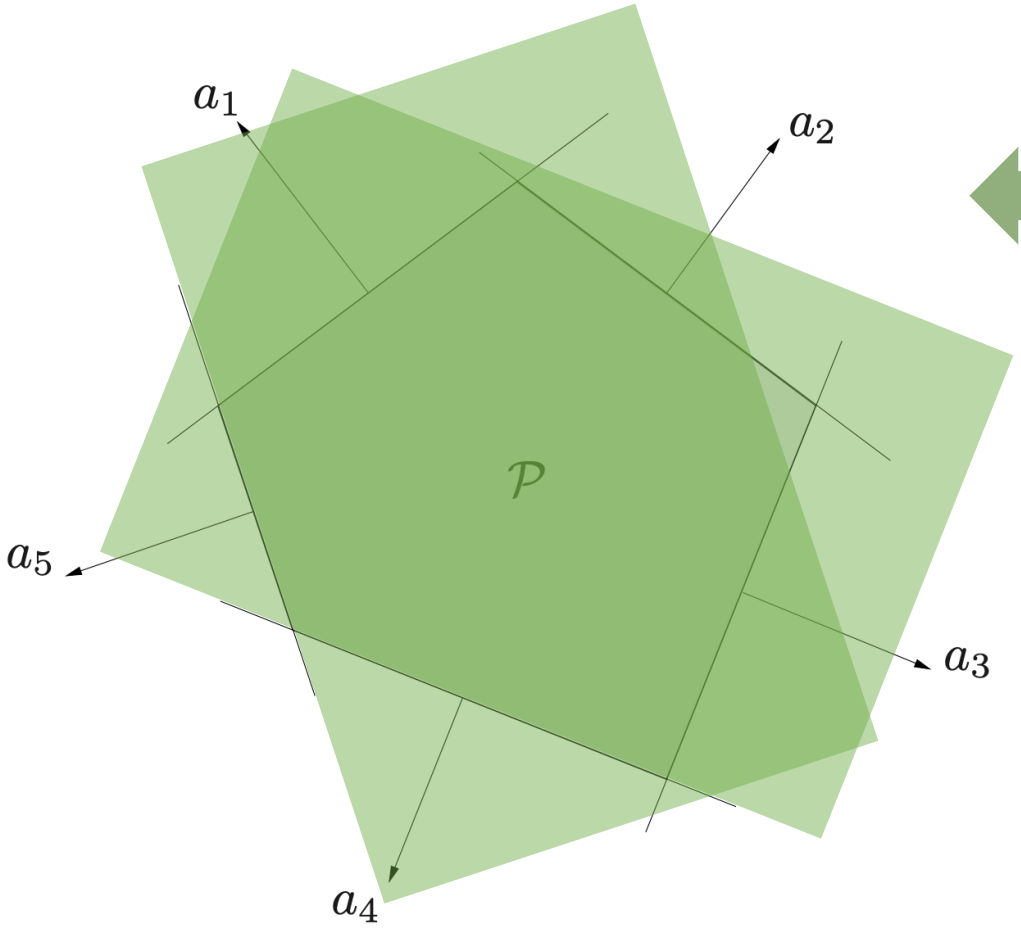
**Proof. Given any two points $x, y \in \mathcal{L}$ and $\lambda \in (0, 1)$, we have**

$$a^T(\lambda \cdot x + (1 - \lambda) \cdot y) = \lambda \cdot a^T x + (1 - \lambda) \cdot a^T y \leq \lambda \cdot b + (1 - \lambda) \cdot b = b.$$

**Thus, $\lambda \cdot x + (1 - \lambda) \cdot y \in \mathcal{L}$.**

# Convex Set

**What is a polyhedron?**




**A polyhedron is the intersection of several halfspaces.**

**Example**

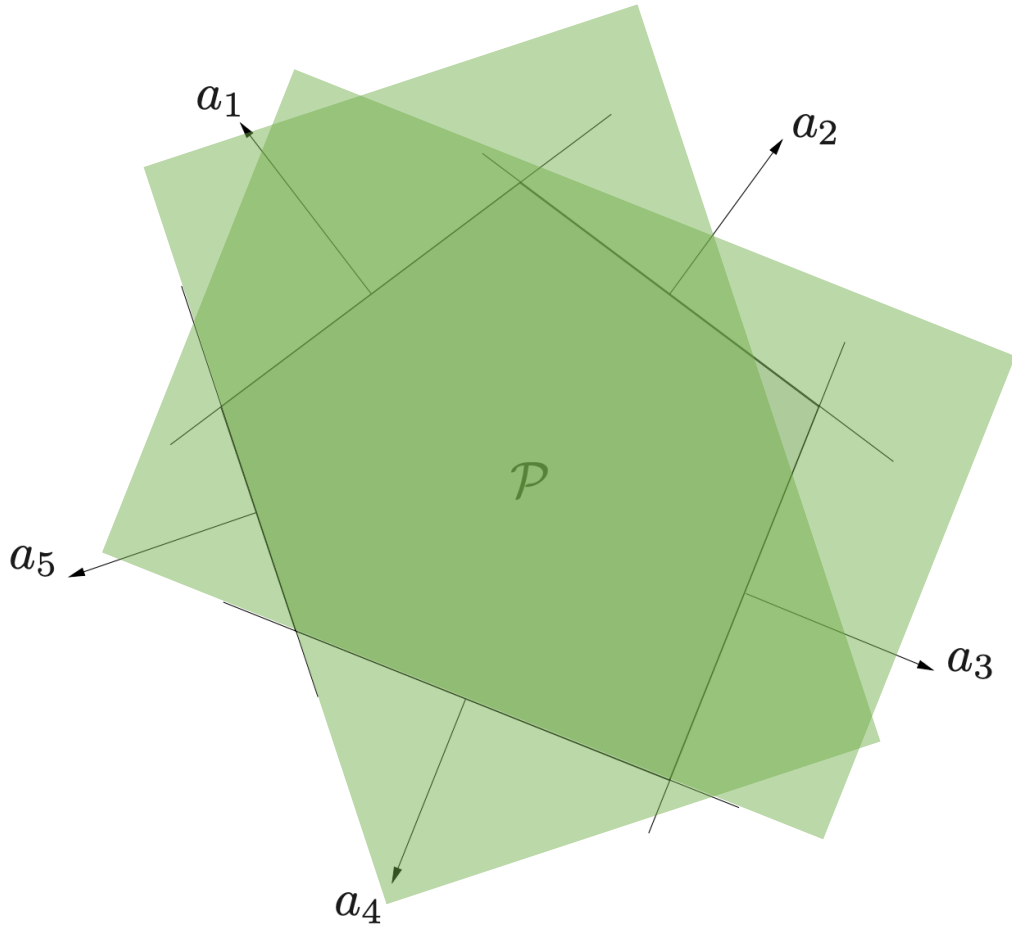
**Polyhedron:** $\mathcal{P} = \{x \in \mathbb{R}^n : a_j^T x \leq b_j, j = 1, \ldots, m\}$.

# Convex Set

**Question**

**How to prove any**

$\mathcal{P} = \{x \in \mathbb{R}^n : a_j^T x \leq b_j, j = 1, \ldots, m\}$ **is a convex set?**




**Theorem: Given two convex sets $\mathcal{A}_1$ and $\mathcal{A}_2$,**
**their intersection $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2$ is also convex.**

# Convex Set

## Question

**How to prove given two convex sets $\mathcal{A}_1$ and $\mathcal{A}_2$, their intersection $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2$ is also convex.**

**Proof. Given any two points $x, y \in \mathcal{A}$, we must have $x, y \in \mathcal{A}_1$ and $x, y \in \mathcal{A}_2$.**

**Thus, for any $\lambda \in (0, 1)$, it must hold that**

$$\lambda \cdot x + (1 - \lambda) \cdot y \in \mathcal{A}_1,$$

$$\lambda \cdot x + (1 - \lambda) \cdot y \in \mathcal{A}_2,$$

**which then implies that $\lambda \cdot x + (1 - \lambda) \cdot y \in \mathcal{A}$.**

# Convex Set

## Question

Is $\{x \in \mathbb{R}^n : a_j^T x \leq b_j, j = 1, \ldots, m, c_j^T x = d_j, j = 1, \ldots, n\}$ a polyhedron?

Yes. Because it can be equivalently written as

$$\{x \in \mathbb{R}^n : a_j^T x \leq b_j, j = 1, \ldots, m, c_j^T x \leq d_j, -c_j^T x \leq -d_j, j = 1, \ldots, n\}$$

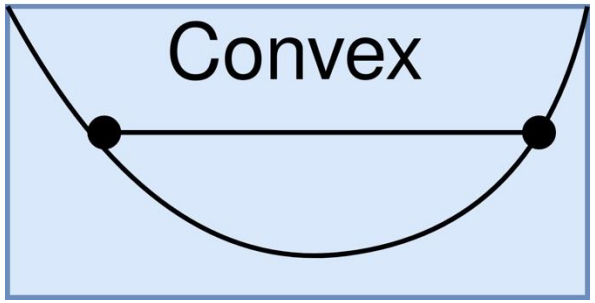
This is an important trick!

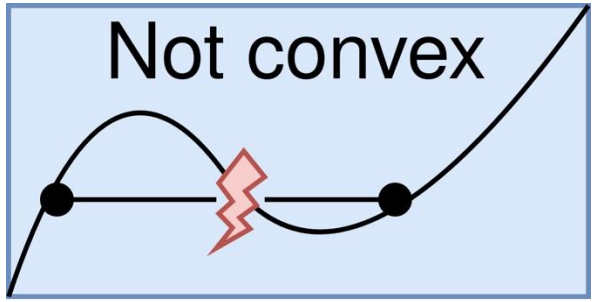
# Convex Function

## Definition

A function $f: \mathcal{A} \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex if $\mathcal{A}$ is a convex set and for all $x, y \in \mathcal{A}$ and $\lambda \in (0, 1)$, we have

$$f(\lambda \cdot x + (1 - \lambda) \cdot y) \leq \lambda \cdot f(x) + (1 - \lambda) \cdot f(y).$$



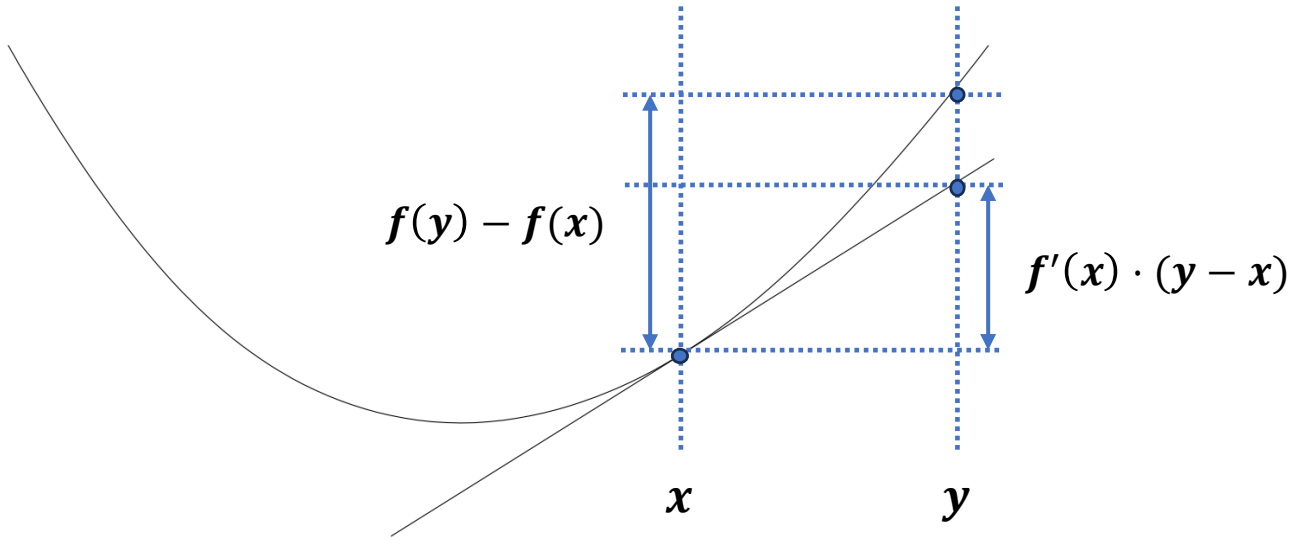

How to check convexity? Let's first consider a simple scenario when $n = 1$.

# Convex Function

## First-Order Condition

**A differentiable function $f: \mathcal{A} \subseteq \mathbb{R} \to \mathbb{R}$ is convex if and only if**

**$\mathcal{A}$ is a convex set and $f(y) - f(x) \geq f'(x) \cdot (y - x)$ for all $x, y \in \mathcal{A}$.**



# Convex Function

## Question

**How to approve a differentiable function $f: \mathcal{A} \subseteq \mathbb{R} \to \mathbb{R}$ is convex if and only if $\mathcal{A}$ is a convex set and $f(y) - f(x) \geq f'(x) \cdot (y - x)$ for all $x, y \in \mathcal{A}$?**

$\Rightarrow$: Since $f(x)$ is convex, for all $\lambda \in [0, 1]$, we have $x + \lambda \cdot (y - x) \in \mathcal{A}$ and

$$f\big(x + \lambda \cdot (y - x)\big) \leq (1 - \lambda) \cdot f(x) + \lambda \cdot f(y).$$

Thus, we have

$$f(y) - f(x) \geq \frac{f\big(x + \lambda \cdot (y - x)\big) - f(x)}{\lambda} = \frac{f\big(x + \lambda \cdot (y - x)\big) - f(x)}{\lambda \cdot (y - x)} \cdot (y - x).$$

# Convex Function

## Question

**How to prove a differentiable function $f\colon \mathcal{A} \subseteq \mathbb{R} \to \mathbb{R}$ is convex if and only if $\mathcal{A}$ is a convex set and $f(y) - f(x) \geq f'(x) \cdot (y - x)$ for all $x, y \in \mathcal{A}$?**

**$\Leftarrow$: For all $x, y \in \mathcal{A}$ and all $\lambda \in [0, 1]$, let $z = \lambda \cdot x + (1 - \lambda) \cdot y$. Since $z \in \mathcal{A}$, we have**

$$f(x) \geq f(z) + f'(z) \cdot (x - z), \quad \text{and} \quad f(y) \geq f(z) + f'(z) \cdot (y - z).$$

**Thus, we have**

$$\lambda \cdot f(x) \geq \lambda \cdot f(z) + \lambda \cdot f'(z) \cdot (x - z), \quad \text{and} \quad (1 - \lambda) \cdot f(y) \geq (1 - \lambda) \cdot f(z) + (1 - \lambda) \cdot f'(z) \cdot (y - z).$$

**Adding them up, we then have $\lambda \cdot f(x) + (1 - \lambda) \cdot f(y) \geq f(z)$, which concludes the proof.**

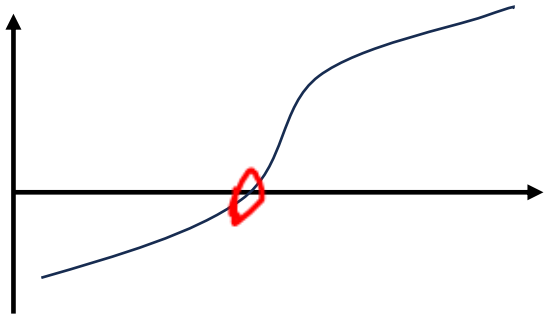
# Convex Function

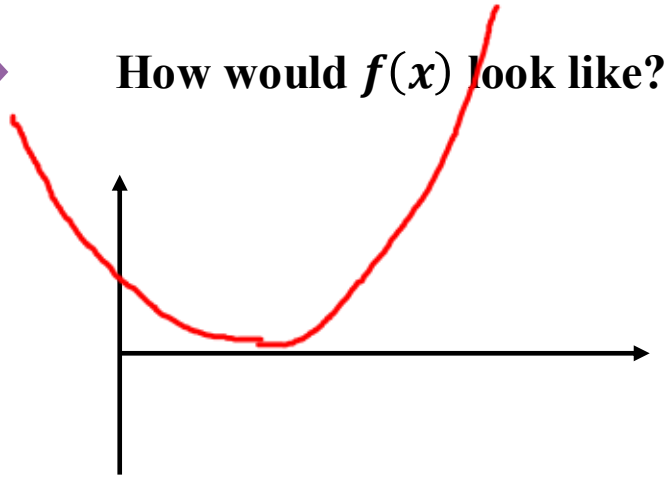
## Second-Order Condition

A *twice differentiable* function $f: \mathcal{A} \subseteq \mathbb{R} \to \mathbb{R}$ is convex if and only if

$\mathcal{A}$ is a convex set and $f''(x) \geq 0$ for all $x \in \mathcal{A}$.

$f''(x) \geq 0$ ➡ $f'(x)$ is an increasing function ➡ How would $f(x)$ look like?





# Convex Function

## Question

**How to prove a twice differentiable function $f: \mathcal{A} \subseteq \mathbb{R} \to \mathbb{R}$ is convex if and only if**

**$\mathcal{A}$ is a convex set and $f''(x) \geq 0$ for all $x \in \mathcal{A}$.**

**$\Rightarrow$: Since $f(x)$ is convex, for any $x, y \in \mathcal{A}$, we have**

$$f'(x) \cdot (y - x) \leq f(y) - f(x) \leq f'(y) \cdot (y - x).$$

**Thus, we have**

$$\frac{f'(y) - f'(x)}{y - x} \geq 0.$$

**Letting $y \to x$ then concludes the proof.**

# Convex Function

## Question

**How to prove a twice differentiable function $f: \mathcal{A} \subseteq \mathbb{R} \to \mathbb{R}$ is convex if and only if**

**$\mathcal{A}$ is a convex set and $f''(x) \geq 0$ for all $x \in \mathcal{A}$.**

**$\Leftarrow$: For any $x, y \in \mathcal{A}$ with $x < y$, we have**

$$0 \leq \int_x^y f''(z) \cdot (y - z) \mathrm{d}z$$

$$= f'^{(z)} \cdot (y - z) \Big|_{z=x}^{z=y} - \int_x^y f'(z) \cdot (-1) \mathrm{d}z = -f'(x) \cdot (y - x) + f(y) - f(x),$$

**which concludes the proof.**

# Convex Function

## Second-Order Condition

A twice differentiable function $f: \mathcal{A} \subseteq \mathbb{R} \to \mathbb{R}$

is convex if and only if

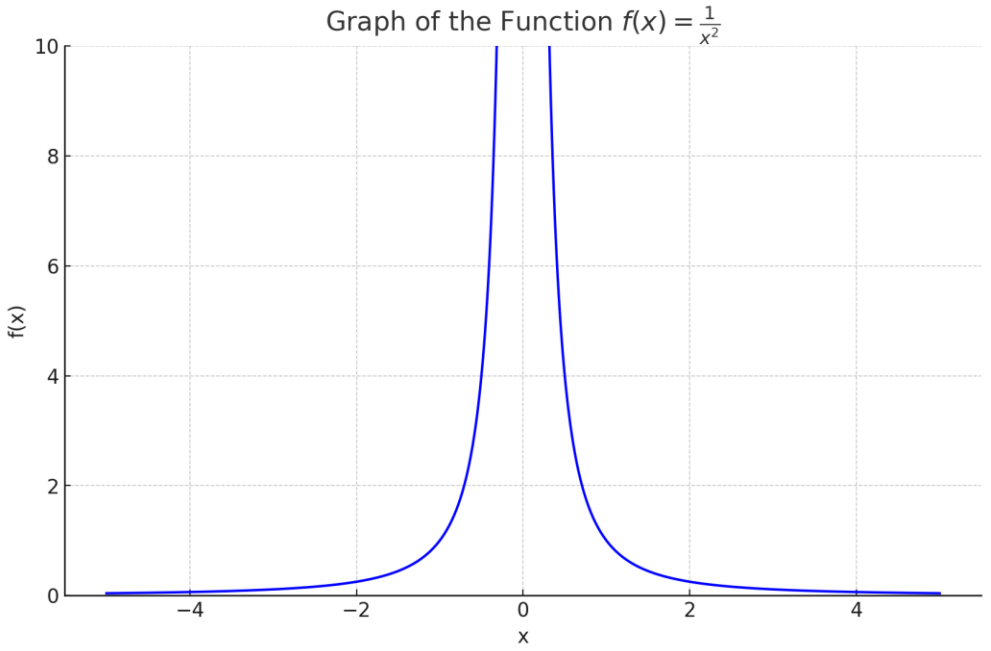
$\mathcal{A}$ is a convex set and $f''(x) \geq 0$ for all $x \in \mathcal{A}$.

This is very important!

Let's then move on to the general case when $\mathcal{A} \subseteq \mathbb{R}^n$.

Consider $f(x) = \frac{1}{x^2}$,

defined on $[-\infty, 0) \cup (0, \infty]$



# Review

## Gradient and Jacobian

Given a function $f: \mathbb{R}^n \to \mathbb{R}$, $\nabla f(x) = [\partial f(x)/\partial x_1, \dots, \partial f(x)/\partial x_n]^{\mathrm{T}}$ ➡ gradient

$$\frac{\partial f(x)}{\partial x} = [\partial f(x)/\partial x_1, \dots, \partial f(x)/\partial x_n] \quad \Rightarrow \text{Jacobian}$$

The first-order Taylor approximation of $f(x)$ at $x^0$ can be written as

either $\quad f(x^0) + \langle \nabla f(x^0), x - x^0 \rangle$

or $\quad f(x^0) + \frac{\partial f(x^0)}{\partial x^0} \cdot (x - x^0)$

# Review

## Gradient and Jacobian

Given a function $f: \mathbb{R}^n \to \mathbb{R}^m$, $\nabla f(x)$ is a $n \times m$ matrix ➡ gradient

$$\begin{bmatrix} \partial f_1(x)/\partial x_1, \dots, \partial f_m(x)/\partial x_1 \\ \dots \\ \partial f_1(x)/\partial x_n, \dots, \partial f_m(x)/\partial x_n \end{bmatrix}$$

$\partial f(x)/\partial x$ is a $m \times n$ matrix ➡ Jacobian

$$\begin{bmatrix} \partial f_1(x)/\partial x_1, \dots, \partial f_1(x)/\partial x_n \\ \dots \\ \partial f_m(x)/\partial x_1, \dots, \partial f_m(x)/\partial x_n \end{bmatrix}$$

# Review

## Hessian

**Given a function $f$: $\mathbb{R}^n \to \mathbb{R}$, its Hessian at $x$ is the Jacobian of its gradient at $x$**

$$\nabla^2 f(x) = \frac{\partial}{\partial x} \nabla f(x)$$

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

# Review

## Positive Semidefinite (PSD) Matrix

**Definition: A symmetric matrix $\boldsymbol{H} \in \mathbb{R}^{n \times n}$ is PSD if $\boldsymbol{x}^T \boldsymbol{H} \boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$.**

**A symmetric matrix $\boldsymbol{H} \in \mathbb{R}^{n \times n}$ is PSD**

- **If and only if all of its eigenvalues are non-negative.**
- **If and only if there exists $\boldsymbol{L} \in \mathbb{R}^{n \times m}$ such that $\boldsymbol{H} = \boldsymbol{L}\boldsymbol{L}^T$.**
- **If and only if all of its principal minors (主子式) are non-negative**

$$\begin{bmatrix} 4 & 5 & 1 \\ 3 & 3 & 0 \\ 1 & 8 & 2 \end{bmatrix}$$

**If the leading principal minors of $\boldsymbol{H}$ are all positive, then $\boldsymbol{H}$ is PSD.**

# Review

## Positive Semidefinite (PSD) Matrix

**Definition: A symmetric matrix $H \in \mathbb{R}^{n \times n}$ is PSD if $x^T H x \geq 0$ for all $x \in \mathbb{R}^n$.**

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & -3 \\ -3 & 10 \end{pmatrix}$$

$$\begin{pmatrix} -1 & -3 \\ -3 & 10 \end{pmatrix}$$

# Convex Function

## First-Order Condition

A differentiable function $f$: $\mathcal{A} \subseteq \mathbb{R} \to \mathbb{R}$ is convex if and only if

$\mathcal{A}$ is a convex set and $f(y) - f(x) \geq f'(x) \cdot (y - x)$ for all $x, y \in \mathcal{A}$.

A differentiable function $f$: $\mathcal{A} \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex if and only if

$\mathcal{A}$ is a convex set and $f(y) - f(x) \geq \langle \nabla f(x) \cdot y - x \rangle$ for all $x, y \in \mathcal{A}$.

# Convex Function

## Second-Order Condition

A twice differentiable function $f: \mathcal{A} \subseteq \mathbb{R} \to \mathbb{R}$ is convex if and only if

$\mathcal{A}$ is a convex set and $f''(x) \geq 0$ for all $x \in \mathcal{A}$.

A twice differentiable function $f: \mathcal{A} \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex if and only if

$\mathcal{A}$ is a convex set and $\nabla^2 f(x)$ is positive semidefinite (PSD) for all $x \in \mathcal{A}$.

# Convex Function

## Example

- **Linear function:** $f(x) = c^{\mathrm{T}}x$
- **Quadratic function:** $f(x) = \frac{1}{2}x^{\mathrm{T}}Px + q^{\mathrm{T}}x + r$, **where** $P$ **is PSD.**
- **Negative entropy function:** $f(x) = \langle x, \log(x) \rangle$

**Check their Hessian.**

- **Norms:**
  - $f(x) = ||x||_1 = \sum_{i=1}^{n} |x_i|,$
  - $f(x) = ||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$
  - $f(x) = ||x||_\infty = \max_{i=1}^{n} |x_i|$

**Properties of norms:**

1. $||\lambda x|| = \lambda \cdot ||x||$
2. $||x + y|| \leq ||x|| + ||y||$

# Exam Questions

## Question

**Prove that a twice differentiable function $f$: $\mathcal{A} \subseteq \mathbb{R} \to \mathbb{R}$ is convex if and only if**

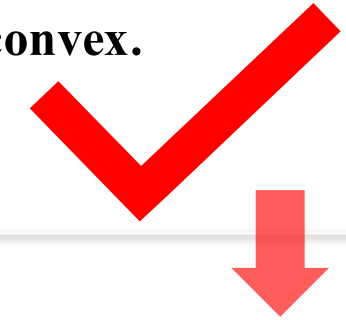
**$\mathcal{A}$ is a convex set and $f''(x) \geq 0$ for all $x \in \mathcal{A}$.**



## Question

**Prove that $f$: $\mathcal{A} \subseteq \mathbb{R}^n \to \mathbb{R}$ defined as xxx is convex.**

**Prove that $\mathcal{A} \subseteq \mathbb{R}^n$ defined as xxx is convex.**



***This is an important skill for optimization!***

***Why…… Let's move to the next lecture.***

# Decision Variable

**1. Decision variable $x$**

→ **$x$ is a real variable**
**e.g., $x \in \mathbb{R}^n$**

→ **$x$ is a discrete variable**
**e.g., $x \in \{0,1\}^n$**

→ **$x$ is partially real, partially discrete**
**e.g., $x \in \mathbb{R}^n \times \{0,1\}^n$**

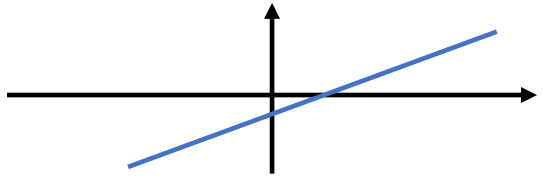



**Lane number: discrete or continuous?**

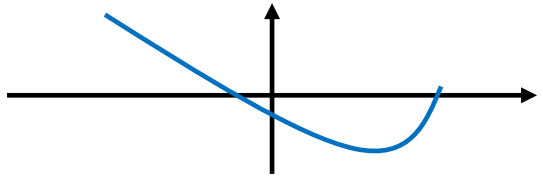
# Objective Function

**2. Objective function $f(x)$**

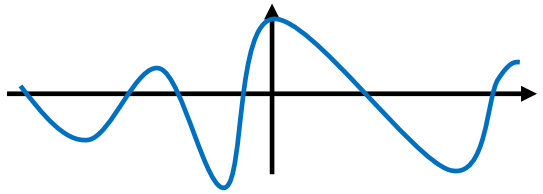
$f(x)$ **is linear**



$f(x)$ **is convex**



$f(x)$ **is nonconvex**



# Constraint

**3. Inequality constraint $g_i(x) \leq 0$**

$g_i(x)$ is linear → $\mathcal{G}_i = \{x: g_i(x) \leq 0\}$ is a halfspace

$g_i(x)$ is convex → $\mathcal{G}_i = \{x: g_i(x) \leq 0\}$ is convex

$g_i(x)$ is nonconvex

**Proof.** Given any $x, y \in \mathcal{G}_i$ and $\lambda \in (0, 1)$, we have

$$g_i(\lambda \cdot x + (1-\lambda) \cdot y) \leq \lambda \cdot g_i(x) + (1-\lambda) \cdot g_i(y) \leq 0,$$

which implies that $\lambda \cdot x + (1-\lambda) \cdot y \in \mathcal{G}_i$.

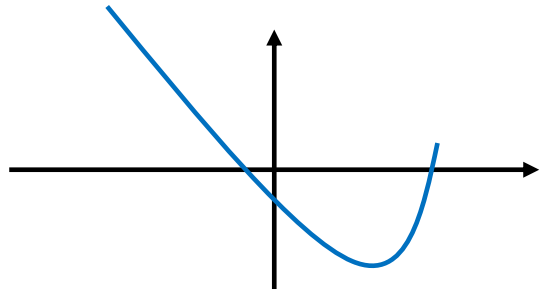
# Constraint

**4. Equality constraint $h_i(x) = 0$**

→ $h_i(x)$ is linear → $\mathcal{H}_i = \{x: h_i(x) = 0\}$ is a hyperplane

→ $h_i(x)$ is convex → $\mathcal{H}_i = \{x: h_i(x) = 0\}$ is *not* necessarily convex

→ $h_i(x)$ is nonconvex



# Optimization Problems

## Integer Program

$$\begin{aligned} \underset{x \in \{0,1\}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0, \ i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{aligned}$$

## Mixed-Integer Program

$$\begin{aligned} \underset{x \in \mathbb{R}^n \times \{0,1\}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0, \ i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{aligned}$$

## Continuous Optimization Problem

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0, \ i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{aligned}$$

# Optimization Problems

## Unconstrained Program

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

## Constrained Program

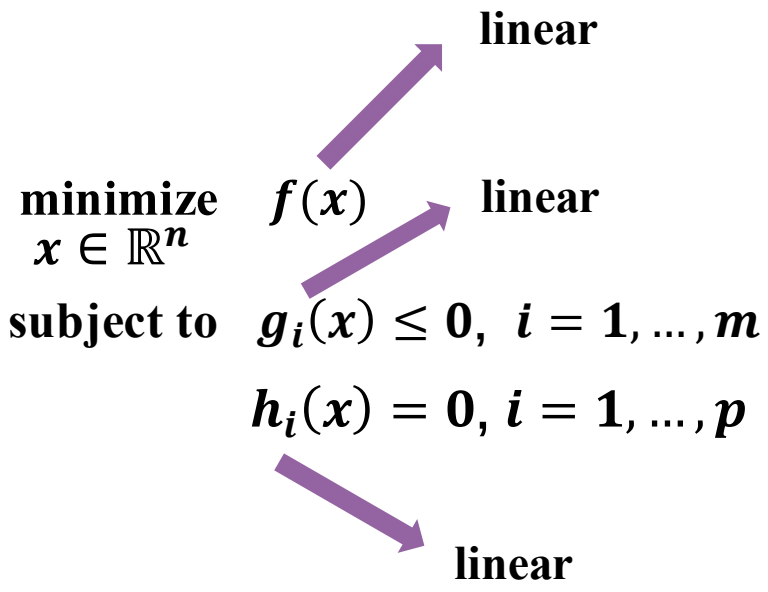
$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

$$\text{subject to} \quad g_i(x) \leq 0, \ \ i = 1, \ldots, m$$

$$h_i(x) = 0, \ i = 1, \ldots, p$$

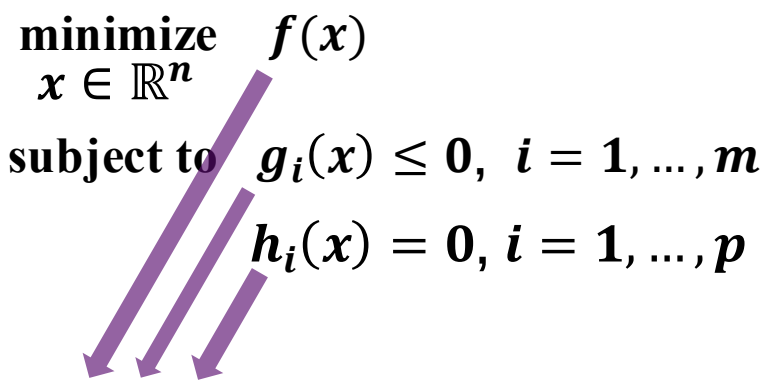
# Optimization Problems

## Linear Program




The feasible region is a polyhedron

## Nonlinear Program




As long as *one* of these is nonlinear

# Optimization Problems

## Convex Program

$$\begin{aligned} \text{minimize}_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0, \ i = 1, \dots, m \\ & h_i(x) = 0, \ i = 1, \dots, p \end{aligned}$$

$f(x)$: convex

$g_i(x) \le 0$: convex

$h_i(x) = 0$: linear!

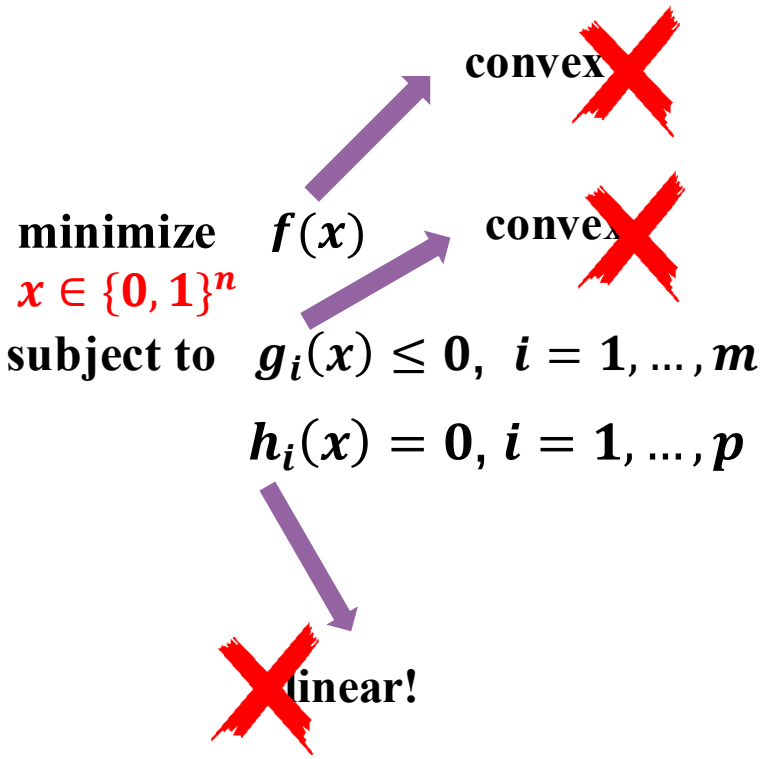
The feasible region is a convex set

$\mathcal{G}_i = \{x: g_i(x) \le 0\}$ is a convex set

$\mathcal{H}_i = \{x: h_i(x) = 0\}$ is a hyperplane

$\Rightarrow$ $\mathcal{A} = (\cap_{i=1}^{m} \mathcal{G}_i) \cap (\cap_{i=1}^{p} \mathcal{H}_i)$ is convex

# Optimization Problems




**Convex programs** can usually be solved very accurately.

→ Many efficient algorithms exist

**Nonconvex programs** are mostly *intractable.*

**Integer programs** are nightmares.

→ Metaheuristic methods often become helpful

→ You must know how to check convexity