第三章：凸优化基础

——优化论五部曲

Jason 博士

网易微专业 x 稀牛学院

人工智能数学基础微专业

一般优化问题

凸集和凸函数基础

凸优化问题

无约束优化问题

- 自变量为矢量的函数 $f: \mathbb{R}^n \to \mathbb{R}$

$$\min \ f(\mathbf{x}) \qquad \mathbf{x} \in \mathbb{R}^n$$

- 局部最优解常规解法

- 直接法
    - 梯度等于 0，求得驻点，必要时 Hessian 矩阵再进一步判断.

- 迭代法
    - 梯度下降法
    - 牛顿法
    - 拟牛顿法

一般约束优化问题

- 约束优化问题一般形式：

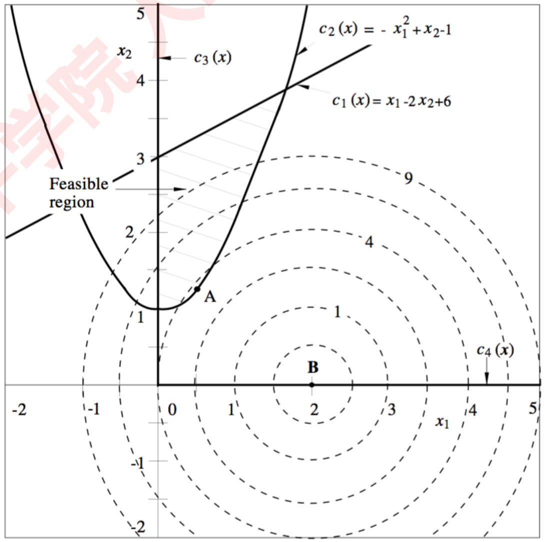
$$
\begin{aligned}
\text{minimize} \quad & f_0(\mathbf{x}) \\
\text{subject to} \quad & f_i(\mathbf{x}) \leq 0 \text{ for } i = 1, 2, \cdots, m \\
& h_i(\mathbf{x}) = 0 \text{ for } i = 1, 2, \cdots, p
\end{aligned} \tag{1}
$$

- 可行域：满足 $f(\mathbf{x})$ 定义域和约束条件的 $\mathbf{x}$ 的集合. $f_j(\mathbf{x}) = 0$ 表明不等式约束被激活 (active).

一般约束优化问题 (举例)

- 考虑以下约束优化问题

$$
\begin{aligned}
\text{minimize} \quad & f(\mathbf{x}) = x_1^2 + x_2^2 - 4x_1 + 4 = (x_1 - 2)^2 + x_2^2 \\
\text{subject to} \quad & c_1(\mathbf{x}) = x_1 - 2x_2 + 6 \geq 0 \\
& c_2(\mathbf{x}) = -x_1^2 + x_2 - 1 \geq 0,\ c_3(\mathbf{x}) = x_1 \geq 0,\ c_4(\mathbf{x}) = x_2 \geq 0
\end{aligned}
$$



补充知识 $\mathbf{Ax} = \mathbf{b}$

矩阵乘法

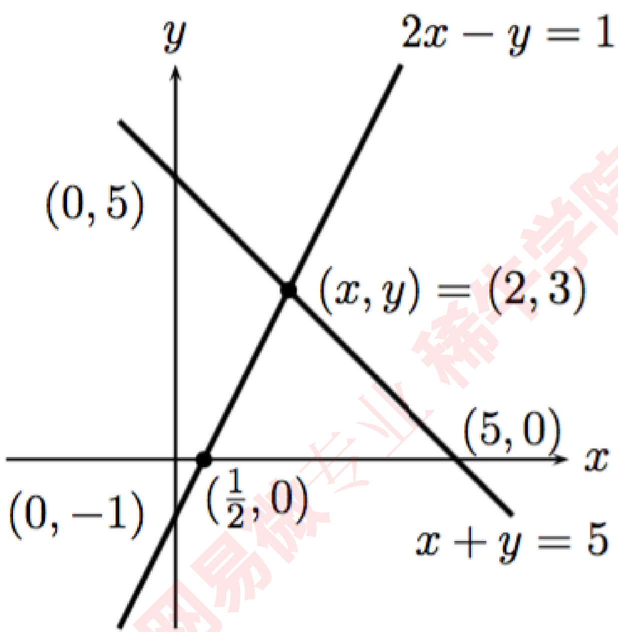
$$\underbrace{\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}}_{\mathbf{A}\in\mathbb{R}^{2\times 2}} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\mathbf{x}\in\mathbb{R}^{2}} = \underbrace{\begin{bmatrix} 1 \\ 5 \end{bmatrix}}_{\mathbf{b}\in\mathbb{R}^{2}} \tag{2}$$

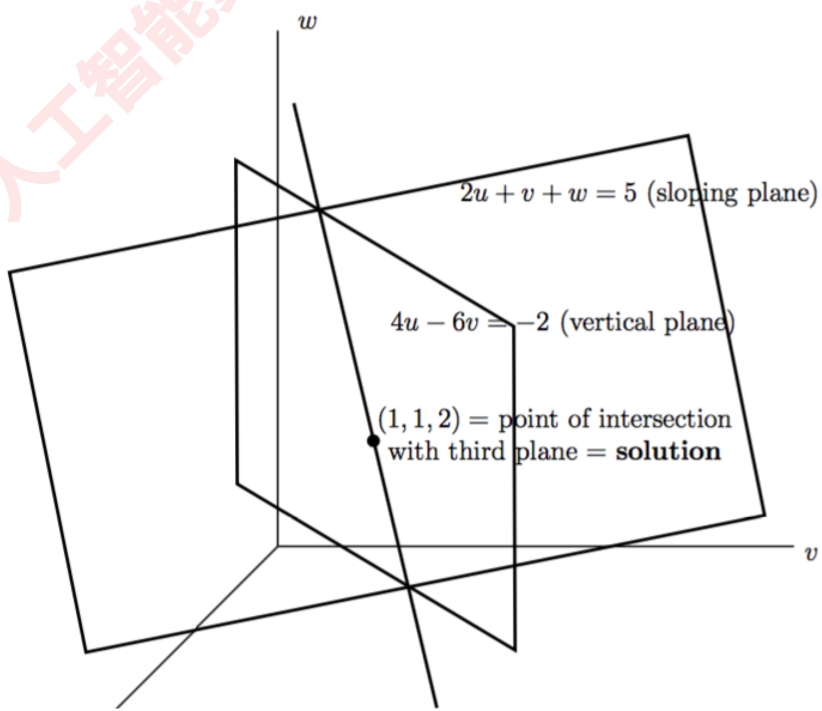
$$\underbrace{\begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}}_{\mathbf{A}\in\mathbb{R}^{3\times 3}} \underbrace{\begin{bmatrix} u \\ v \\ w \end{bmatrix}}_{\mathbf{x}\in\mathbb{R}^{3}} = \underbrace{\begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}}_{\mathbf{b}\in\mathbb{R}^{3}} \tag{3}$$

$\mathrm{Ax} = \mathrm{b}$ 的行视图

- 行视图—超平面

$$\begin{aligned} 2x - y &= 1 \\ x + y &= 5 \end{aligned} \quad (4)$$





一般优化问题

凸集和凸函数基础

凸优化问题

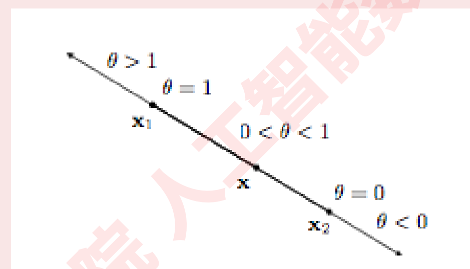
为什么要凸优化?

- 思考：对于无约束优化问题，梯度等于 0 这一条件是否可以成为充要条件?
- 思考：什么样的情况局部最小解可以成为全局最小解?
- 思考：约束优化问题怎么办?（下次课）
- 思考：研究凸优化问题对于非凸问题又有什么帮助? (下次课)

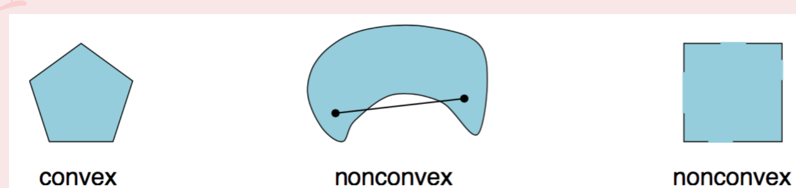
凸集 (Convex Sets)(Solid, no holes, and curve outward)

仿射集 (Affine Sets) 和凸集

- 如果一个集合 $C \in \mathbb{R}^n$ 是仿射的，则 C 中两点间的直线也在 C 中，例如 $\mathbf{x} = \theta \mathbf{x}_1 + (1-\theta)\,\mathbf{x}_2 \in C,\ \theta \in \mathbb{R}$，即 $\mathbf{Ax} = \mathbf{b}$ 的解



- 一个集合 $C \in \mathbb{R}^n$ 是凸的，则对于任意的 $\mathbf{x}, \mathbf{y} \in C$，有 $\theta \mathbf{x} + (1-\theta)\,\mathbf{y} \in C \quad 0 \le \theta \le 1$

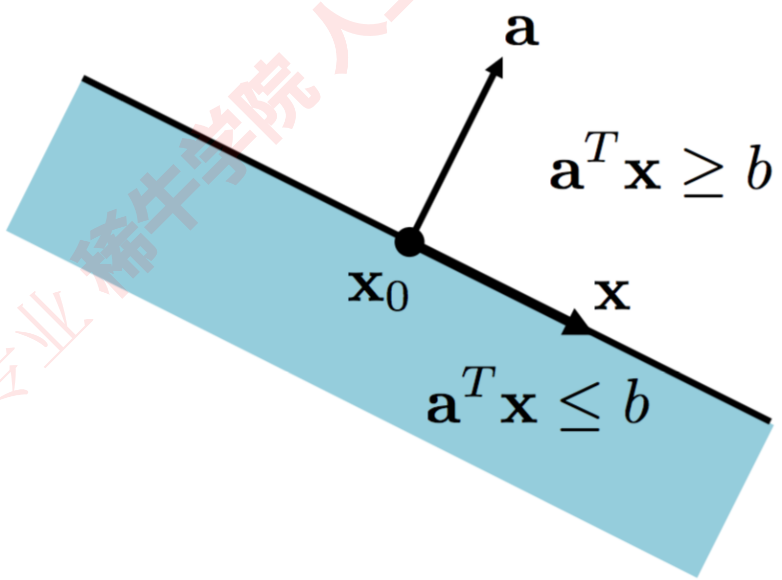


常见的凸集 (1/3)

- 所有 $\mathbb{R}^n$，给定任意 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, 则有 $\theta \mathbf{x} + (1-\theta)\,\mathbf{y} \in \mathbb{R}^n$
- 所有 $\mathbb{R}^n_+$
- 超平面 (Hyperplane): 既是仿射又是凸

$$C = \{\mathbf{x} \mid \mathbf{a}^T \mathbf{x} = b\}$$

- 半空间 (Halfspace)：只是凸



向量范数 (补充知识)

- **2-norm:**

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} = (\mathbf{x}^T\mathbf{x})^{1/2}$$

- **1-norm:**

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

- ∞**-norm:**

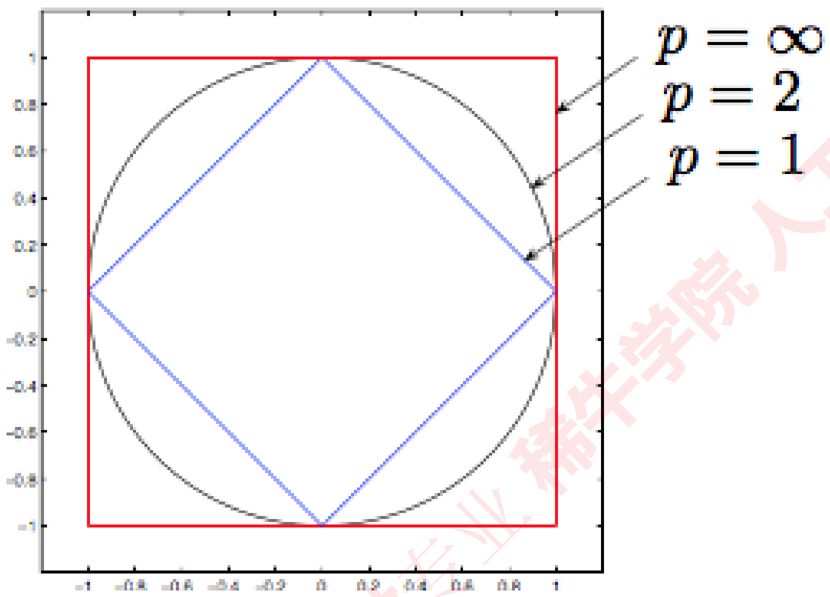
$$\|\mathbf{x}\|_\infty = \max_{i=1,\ldots,n} |x_i|$$

- p**-norm,** $p \geq 1$:

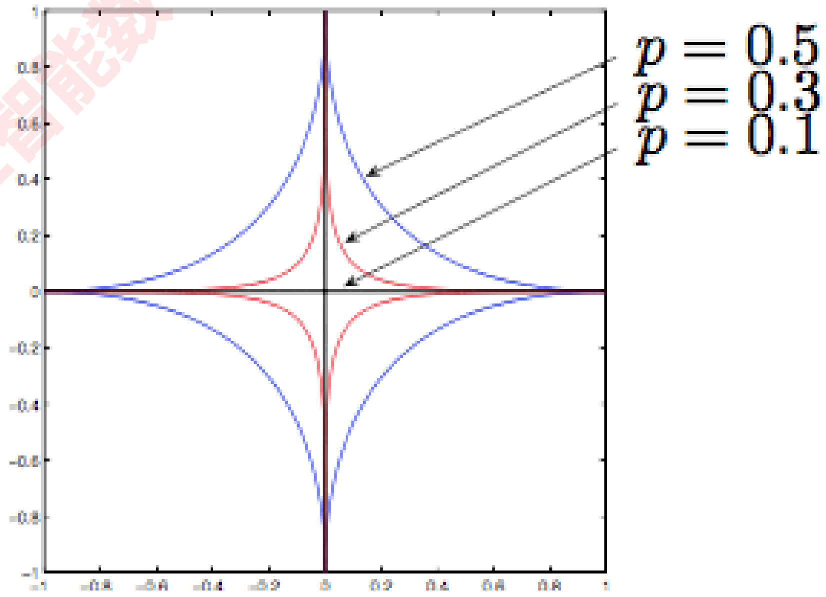
$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}.$$

常见的凸集 (2/3)

- 范数球，例如 $||\mathbf{x}||_2 \leq 1$。给定任意 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ 且 $||\mathbf{x}||_2 \leq 1$，$||\mathbf{y}||_2 \leq 1$，则有 $||\theta \mathbf{x} + (1-\theta)\,\mathbf{y}||_2 \leq \theta ||\mathbf{x}||_2 + (1-\theta)\,||\mathbf{y}||_2 \leq 1$



(a) Region of $\|x\|_p = 1$, $p \geq 1$.



(b) Region of $\|x\|_p = 1$, $p \leq 1$.

凸集的性质

- **凸集的交集是凸集，** 例如：$S = \{\|\mathbf{x}\| \leq 1, \mathbf{x} \geq \mathbf{0}\}$
- 证明：假定 $S_1, \cdots, S_k$ 是凸集，给定 $\mathbf{x}, \mathbf{y} \in \cap_{i=1}^{k} S_i$，则有

$$\theta \mathbf{x} + (1 - \theta) \mathbf{y} \in S_i, \quad i = 1, \cdots, k$$

因此，

$$\theta \mathbf{x} + (1 - \theta) \mathbf{y} \in \cap_{i=1}^{k} S_i$$

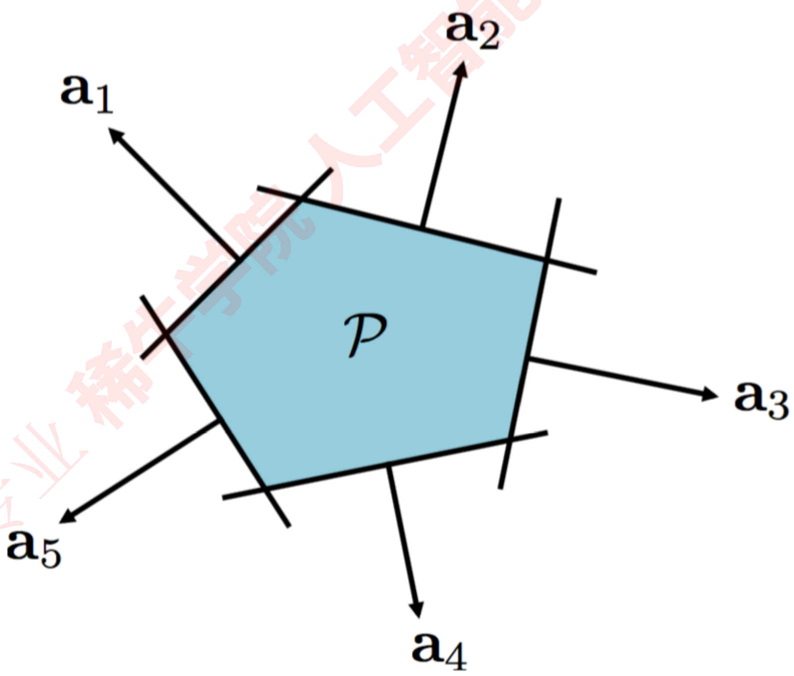
- 注意：凸集的并集不一定是凸集

常见的凸集 (3/3)

- 多面体 (Polyhedron): 有限个半空间和半平面的交集

$$\mathcal{P} = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} \leq \mathbf{b},\ \mathbf{C}\mathbf{x} = \mathbf{d}\}$$

其中 $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^{m}$, $\mathbf{C} \in \mathbb{R}^{p\times n}$, $\mathbf{d} \in \mathbb{R}^{p}$

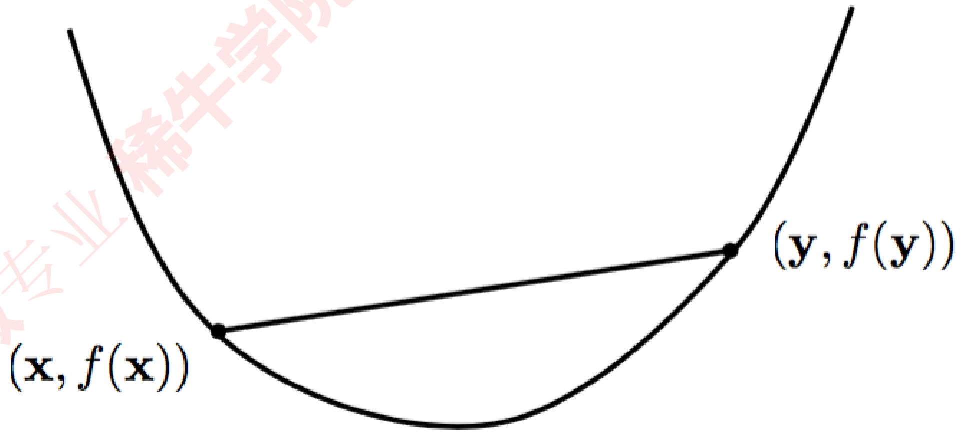


凸函数

- 一个函数 $f\colon \mathbb{R}^n \to \mathbb{R}$ 被称为凸函数，如果
  1. $\mathrm{dom}\,(f)$(f 的定义域) 是凸集
  2. 对于任何 $\mathbf{x}, \mathbf{y} \in \mathrm{dom}\,(f)$ 和 $0 \le \theta \le 1$，有

$$f(\theta\mathbf{x} + (1-\theta)\,\mathbf{y}) \le \theta f(\mathbf{x}) + (1-\theta)\,f(\mathbf{y})$$

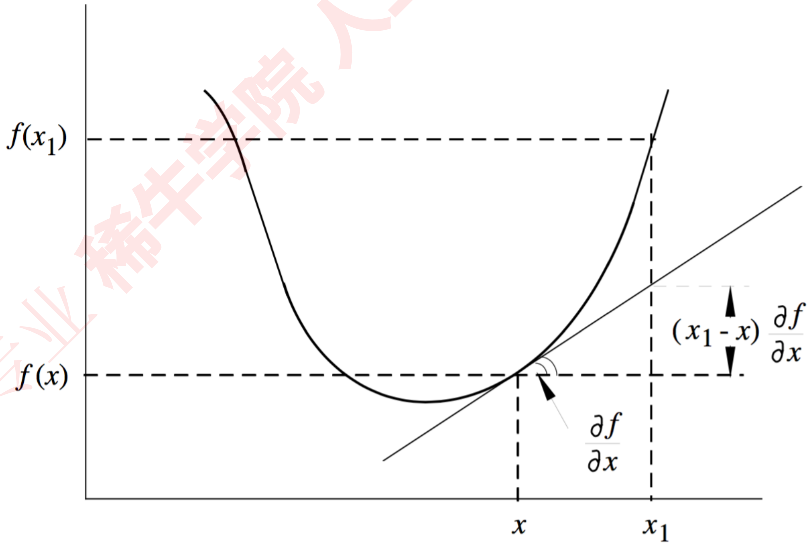
- 几何解释



凸函数的一阶二阶条件

- 一阶充要条件 (不好用)：$f(\mathbf{x}_1) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_1 - \mathbf{x})$ 对于所有 $\mathbf{x}_1, \mathbf{x}$ 均成立．作业：证明凸函数局部最优解就是全局最优解 (提示：用凸函数定义)
- 二阶充要条件 (好用)：如果函数 f 二阶可导，则凸函数充要条件：

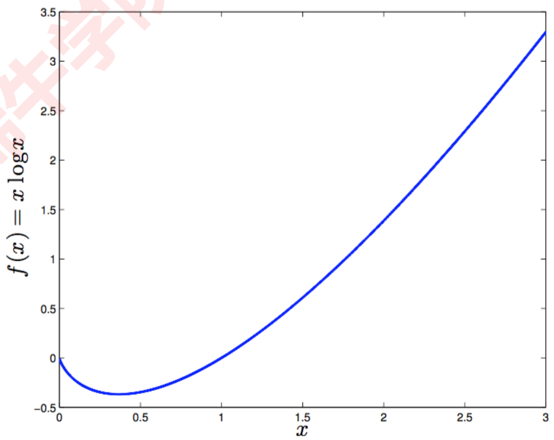
$$\mathbf{H}(\mathbf{x}) \succeq \mathbf{0}$$



常见的凸函数 (1/2)

一元函数举例

- $ax + b$ convex and also concave
- x^2 convex
- $e^{\alpha x}$ convex
- $-\log x$ convex on $x > 0$
- $x \log x$ convex on $x \geq 0$

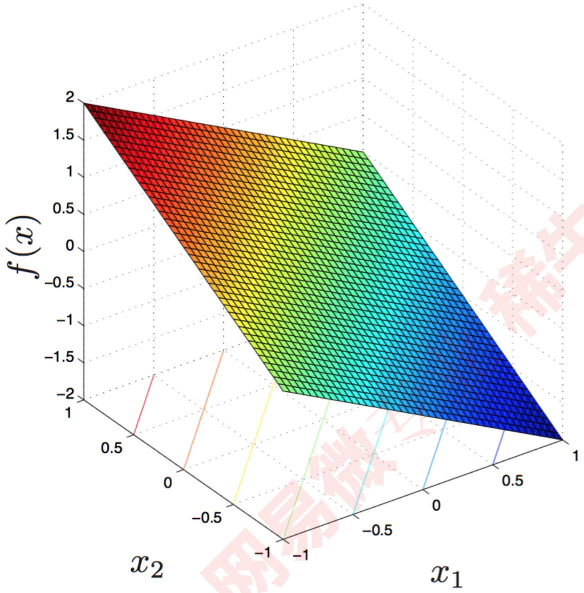


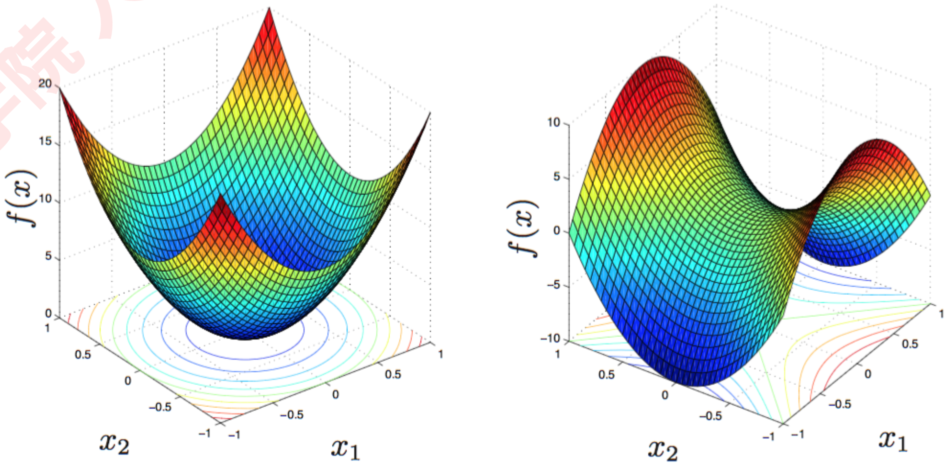
常见的凸函数 (2/2)

二元函数举例

- $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + \mathbf{b}$ convex and also concave
- $f(\mathbf{x}) = \mathbf{x}^T\mathbf{P}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + \mathbf{r}$ 是 convex 当且仅当 $\mathbf{P} \succeq \mathbf{0}$

$$f(\mathbf{x}) = ||\mathbf{x}||_2^2 = \mathbf{x}^T\mathbf{x}$$





(a) $P \succeq 0$. **(b)** $P \not\succeq 0$.

保凸运算

保凸

- f 凸，则 $f(\mathbf{A}\mathbf{x}+\mathbf{b})$ 凸，例如 $\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2$

- g 凸，h 凸，扩展的 h 非递减，则 $f(\mathbf{x})=h(g(\mathbf{x}))$ 凸．例如 $f(\mathbf{x})=\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2$ 凸， $g(\mathbf{x})=\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2, h(x)=x^2$ 在 $x\geq 0$ 的部分非递减.

- $f_1,\cdots,f_m$ 凸， $w_1,\cdots,w_m\geq 0$，则 $\sum_{i=1}^m w_i f_i$ 凸．例如 $f(\mathbf{x})=\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2+\gamma\|\mathbf{x}\|_2^2$ 凸， $\gamma\geq 0$

- 逐点最大： $f_1,\cdots,f_m$ 凸，则 $f(\mathbf{x})=\max\{f_1(\mathbf{x}),\cdots,f_m(\mathbf{x})\}$ 凸． $f(\mathbf{x},\mathbf{y})$ 对于每个 $\mathbf{y}\in\mathcal{A}$ 凸，则 $\sup_{\mathbf{y}\in\mathcal{A}} f(\mathbf{x},\mathbf{y})$ 凸.

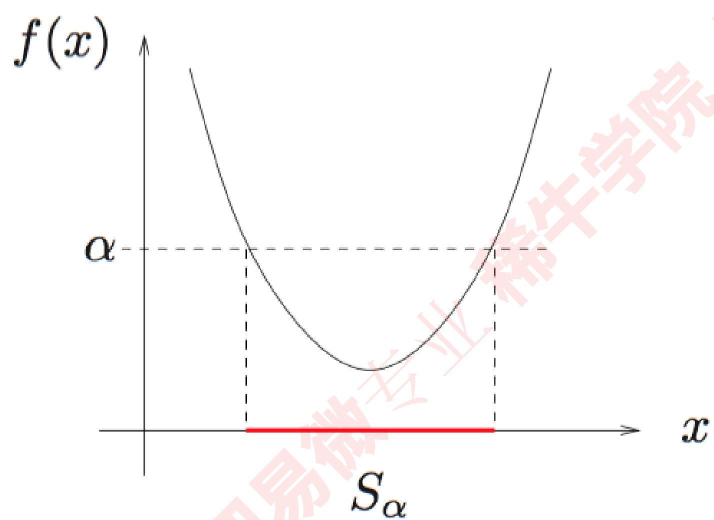
凸函数和凸集的关系

α水平集 (sublevel set)

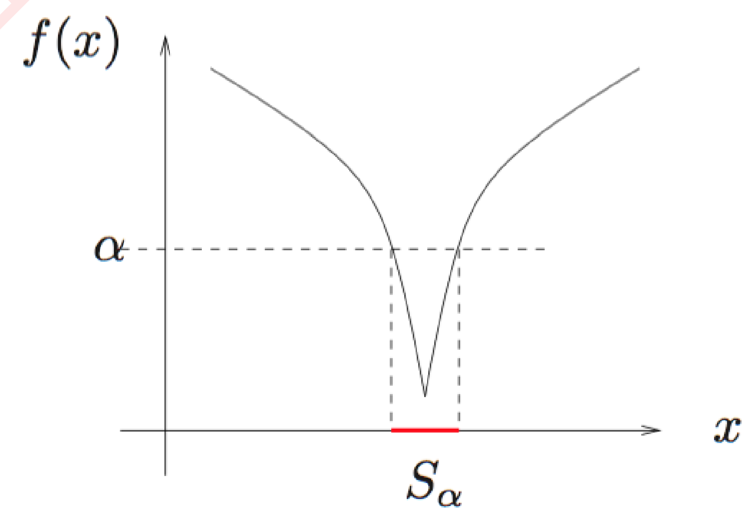
- 一元函数 f 的 α 水平集为

$$S_\alpha = \{x \mid f(x) \leq \alpha\}$$

- 则有 f 凸函数 $\Rightarrow S_\alpha$ 对于每个 α 是凸集，反之不成立.



convex f and convex S_α



non-convex f but convex S_α

一般优化问题

凸集和凸函数基础

凸优化问题

凸优化问题标准形式 (Game Over)

- 凸优化问题

$$
\begin{aligned}
\text{minimize} \quad & f_0(\mathbf{x}) \\
\text{subject to} \quad & f_i(\mathbf{x}) \leq 0 \text{ for } i = 1, 2, \cdots, m \\
& h_i(\mathbf{x}) = 0 \text{ for } i = 1, 2, \cdots, p
\end{aligned} \tag{5}
$$

- 则有 $f_0(\mathbf{x})$ 是凸函数，可行域是凸集
  - 目标函数是凸的
  - 不等式约束函数必须是凸的
  - 等式约束函数必须是仿射的
- 在一个凸集上极小化一个凸的目标函数
- 最优值 (目标函数在可行域上的最小值):

$p* = \min\{f_0(\mathbf{x}) : f_i(\mathbf{x}) \leq 0, h_i(\mathbf{x}) = 0\}$
  - $p* = +\infty$ 不可行（可行域是空集）
  - $p* = -\infty$，unbounded below （存在可行点使得 $f_0(x) \rightarrow -\infty$）
  - $f_0(\mathbf{x}^*) = p*$

凸优化问题的重要结论

凸优化问题局部最优 = 全局最优

- 局部最优 $\mathbf{x}$，存在 $R>0$，对于所有可行点 $\mathbf{z}$，且有 $||\mathbf{x}-\mathbf{z}||_2 \leq R$，满足 $f_0(\mathbf{x}) \leq f_0(\mathbf{z})$
- 全局最优 $\mathbf{x}$，对所有可行点 $\mathbf{z}$，有 $f_0(\mathbf{x}) \leq f_0(\mathbf{z})$
- 反证法：作业
- 特别对于无约束有问题
    - 梯度等于 0
    - 迭代法

典型的凸优化问题

- 线性规划 (Linear Programming)

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^T\mathbf{x} + d \\ \text{subject to} \quad & \mathbf{G}\mathbf{x} \leq \mathbf{h} \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \tag{6}$$

- 二次规划 (Quadratic Programming)($\mathbf{P}$ 半正定)

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{c}^T\mathbf{x} + d \\ \text{subject to} \quad & \mathbf{G}\mathbf{x} \leq \mathbf{h} \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \tag{7}$$

- QCQP($\mathbf{P}$ 和 $\mathbf{Q}_i$ 均半正定)

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{c}^T\mathbf{x} + d \\ \text{subject to} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{Q}_i\mathbf{x} + \mathbf{r}_i^T\mathbf{x} + s_i \leq 0, \quad i = 1, \ldots, m, \ \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

案例 1：凸优化问题转成标准型

- 给定下列问题

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{2}||\mathbf{w}||_2^2 + C\sum_{i=1}^{m}\xi_i \\
\text{subject to} \quad & y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1 - \xi_i,\ i = 1, \cdots, m \\
& \xi_i \geq 0
\end{aligned} \tag{8}
$$

其中 $\mathbf{w} \in \mathbb{R}^n, \boldsymbol{\xi} = [\xi_1, \cdots, \xi_m]^T \in \mathbb{R}^m, b \in \mathbb{R}$．定义 $k = m + n + 1$

- 变量

$$\mathbf{x} \in \mathbb{R}^k = \begin{bmatrix} \mathbf{w} \\ \boldsymbol{\xi} \\ b \end{bmatrix}$$

- 定义 $\mathbf{X} \in \mathbb{R}^{m\times n} = \begin{bmatrix} \mathbf{x}_1^T \\ \\ \mathbf{x}_m^T \end{bmatrix}, \mathbf{y} \in \mathbb{R}^m = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$

案例 1：凸优化问题转成标准型

- QP

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{c}^T\mathbf{x} + d \\
\text{subject to} \quad & \mathbf{G}\mathbf{x} \leq \mathbf{h} \\
& \mathbf{A}\mathbf{x} = \mathbf{b}
\end{aligned} \tag{9}
$$

- 定义

$$
\mathbf{P} \in \mathbb{R}^{k\times k} = \begin{bmatrix} \mathbf{I} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{c} \in \mathbb{R}^k = \begin{bmatrix} 0 \\ C \cdot \mathbf{1} \\ 0 \end{bmatrix}
$$

$$
\mathbf{G} \in \mathbb{R}^{2m\times k} = \begin{bmatrix} -\mathrm{diag}\left(\mathbf{y}\right)\mathbf{X} & -\mathbf{I} & -\mathbf{y} \\ 0 & -\mathbf{I} & 0 \end{bmatrix}, \mathbf{h} \in \mathbb{R}^{2m} = \begin{bmatrix} -\mathbf{1} \\ 0 \end{bmatrix}
$$

案例 2：CVX 演示

本章总结

本章参考资料

Boyd, Stephen, and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.