



華東師範大學  
EAST CHINA NORMAL UNIVERSITY

# 凸优化基础

---

□ 凸集基本概念

□ 凸函数基本概念

□ 凸优化问题

□ 对偶性

□ 优化算法

# 凸函数

## □ 凸集的基本概念

# 几何图形的向量表达

已知二维平面上两定点A(5,1)、B(2,3),试给出经过点A、B的直线方程。

直线:  $\vec{x} = \theta \cdot \vec{a} + (1 - \theta) \cdot \vec{b}$

线段:  $\vec{x} = \theta \cdot \vec{a} + (1 - \theta) \cdot \vec{b}, \theta \in [0, 1]$

三维平面:

$$\vec{x} = \theta_1 \cdot \vec{a}_1 + \theta_2 \cdot \vec{a}_2 + \theta_3 \cdot \vec{a}_3, \theta_1, \theta_2, \theta_3 \in R, \theta_1 + \theta_2 + \theta_3 = 1$$

三角形:

$$\vec{x} = \theta_1 \cdot \vec{a}_1 + \theta_2 \cdot \vec{a}_2 + \theta_3 \cdot \vec{a}_3, \theta_1, \theta_2, \theta_3 \in [0, 1], \theta_1 + \theta_2 + \theta_3 = 1$$

# 几何图形的向量表达

超平面:

$$\vec{x} = \theta_1 \cdot \vec{a}_1 + \theta_2 \cdot \vec{a}_2 + \dots + \theta_k \cdot \vec{a}_k, \theta_1, \theta_2, \dots, \theta_k \in R, \theta_1 + \theta_2 + \dots + \theta_k = 1$$

超几何体 :

$$\vec{x} = \theta_1 \cdot \vec{a}_1 + \theta_2 \cdot \vec{a}_2 + \dots + \theta_k \cdot \vec{a}_k, \theta_1, \theta_2, \dots, \theta_k \in [0, 1], \theta_1 + \theta_2 + \dots + \theta_k = 1$$

# 仿射集 (Affine set)

定义：通过集合C中任意两个不同点的直线仍然在集合C内，则称集合C为仿射集

$$\forall x_1, x_2 \in C, \forall \theta \in R, \text{ 有 } x = \theta \cdot x_1 + (1 - \theta) \cdot x_2 \in C$$

例子：直线、平面、超平面

- 超平面：  $Ax=b$
- $f(x)=0$ 表示定义域在 $R^n$ 的超平面：令 $f(x)=Ax-b$ ,则 $f(x)=0$ 表示截距为b的超平面

# 凸集

集合C内任意两点间的线段均在集合C内，则称集合C为凸集，即：

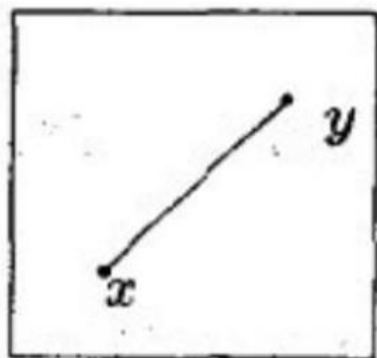
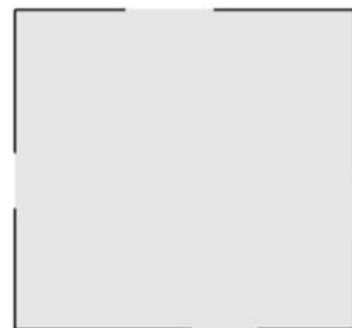
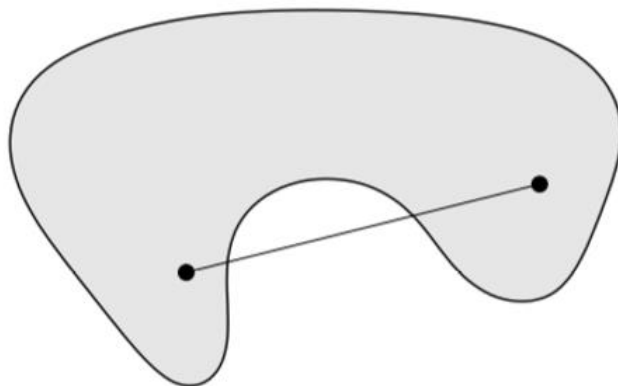
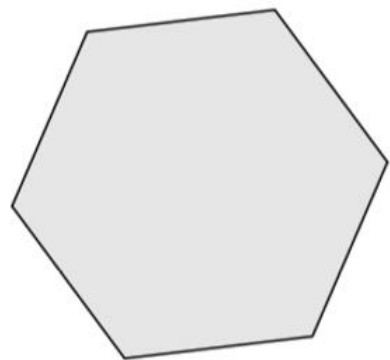
$$\forall x_1, x_2 \in C, \forall \theta \in [0, 1], \text{ 有 } x = \theta \cdot x_1 + (1 - \theta) \cdot x_2 \in C$$

k个点的版本：

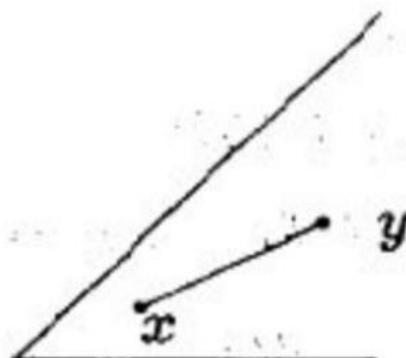
$$\forall x_1, x_2, \dots, x_k \in C, \theta_i \in [0, 1] \text{ 且 } \sum_{i=1}^k \theta_i = 1, \text{ 有 } x = \sum_{i=1}^k \theta_i x_i \in C$$

■ 仿射集也是凸集

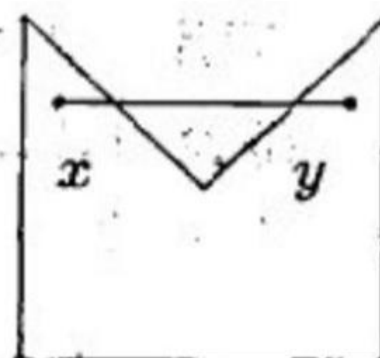
# 凸集



(a) 有界凸集



(b) 无界凸集



(c) 非凸集



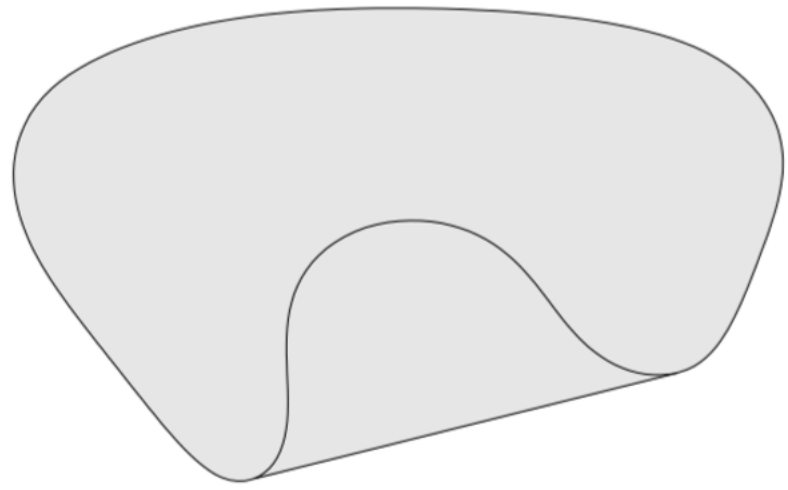
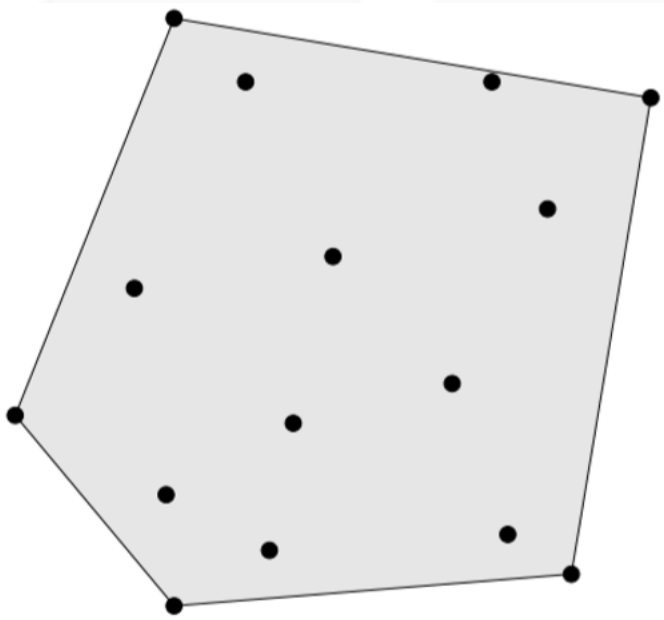
# 凸包

集合C的所有点的凸组合形成的集合，叫做集合C的凸包

$$\text{conv } C = \left\{ \sum_{i=1}^k \theta_i x_i \mid x_i \in C, \theta_i \geq 0, \sum_{i=1}^k \theta_i = 1 \right\}$$

集合C的凸包时能够包含C的最小的凸集

# 凸包



# 超平面和半空间

超平面hyperplane

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

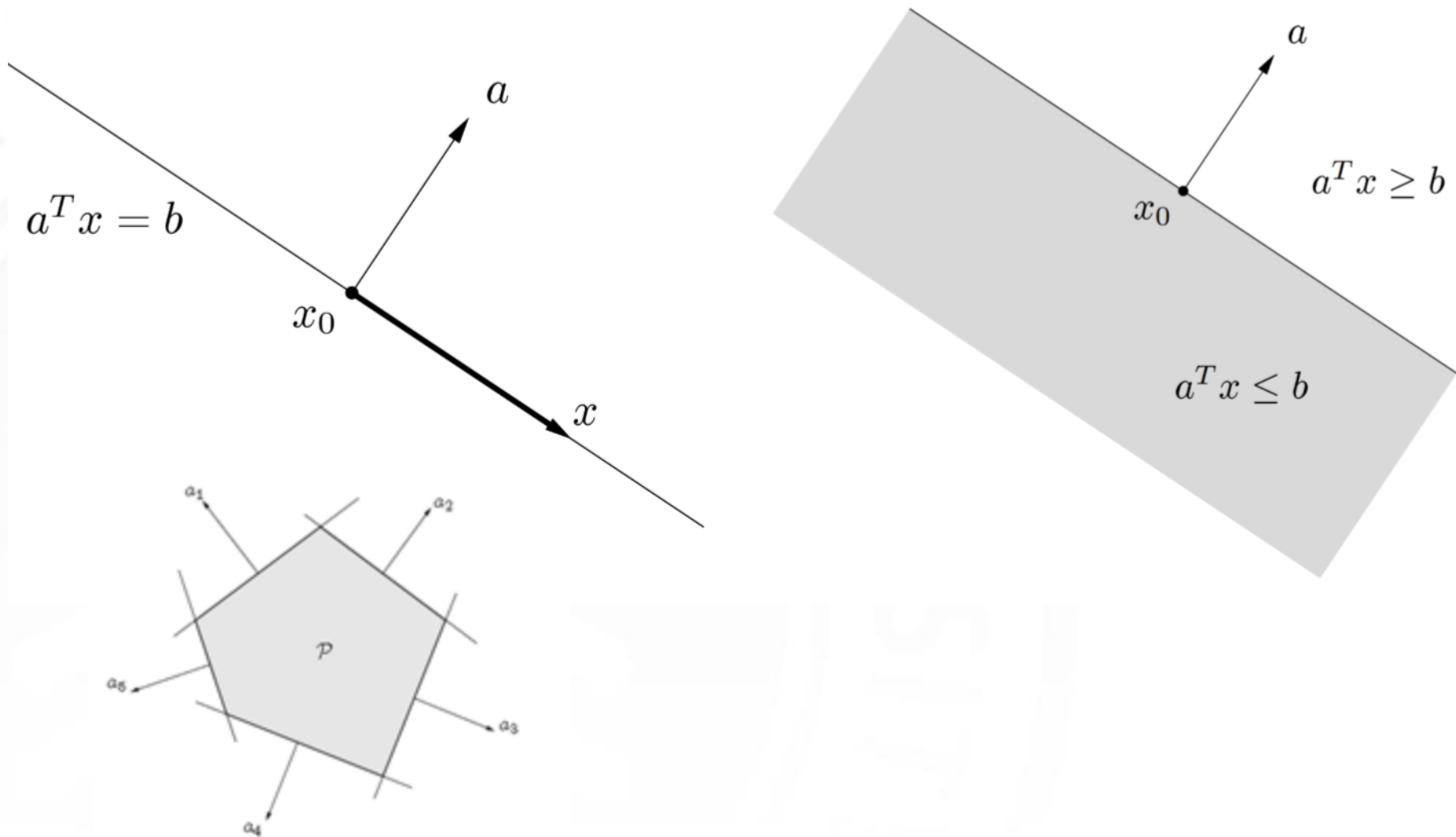
半平面halfspace

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

多面体Polyhedron

$$\{x \mid Ax \leq b\}$$

# 超平面和半空间



# 欧式球和椭球

欧式球

$$\begin{aligned} B(x_c, r) &= \{x \mid \|x - x_c\|_2 \leq r\} \\ &= \{x \mid (x - x_c)^T (x - x_c) \leq r^2\} \end{aligned}$$

椭球

$$E = \{x \mid (x - x_c)^T P (x - x_c) \leq r^2\},$$

$P$ 为对称正定矩阵

# 范数球和范数锥

范数:

$$\|x\| \geq 0, \|x\| = 0 \text{ iff } x = 0;$$

$$\|tx\| = |t|\|x\|, t \in R;$$

$$\|x + y\| \leq \|x\| + \|y\|$$

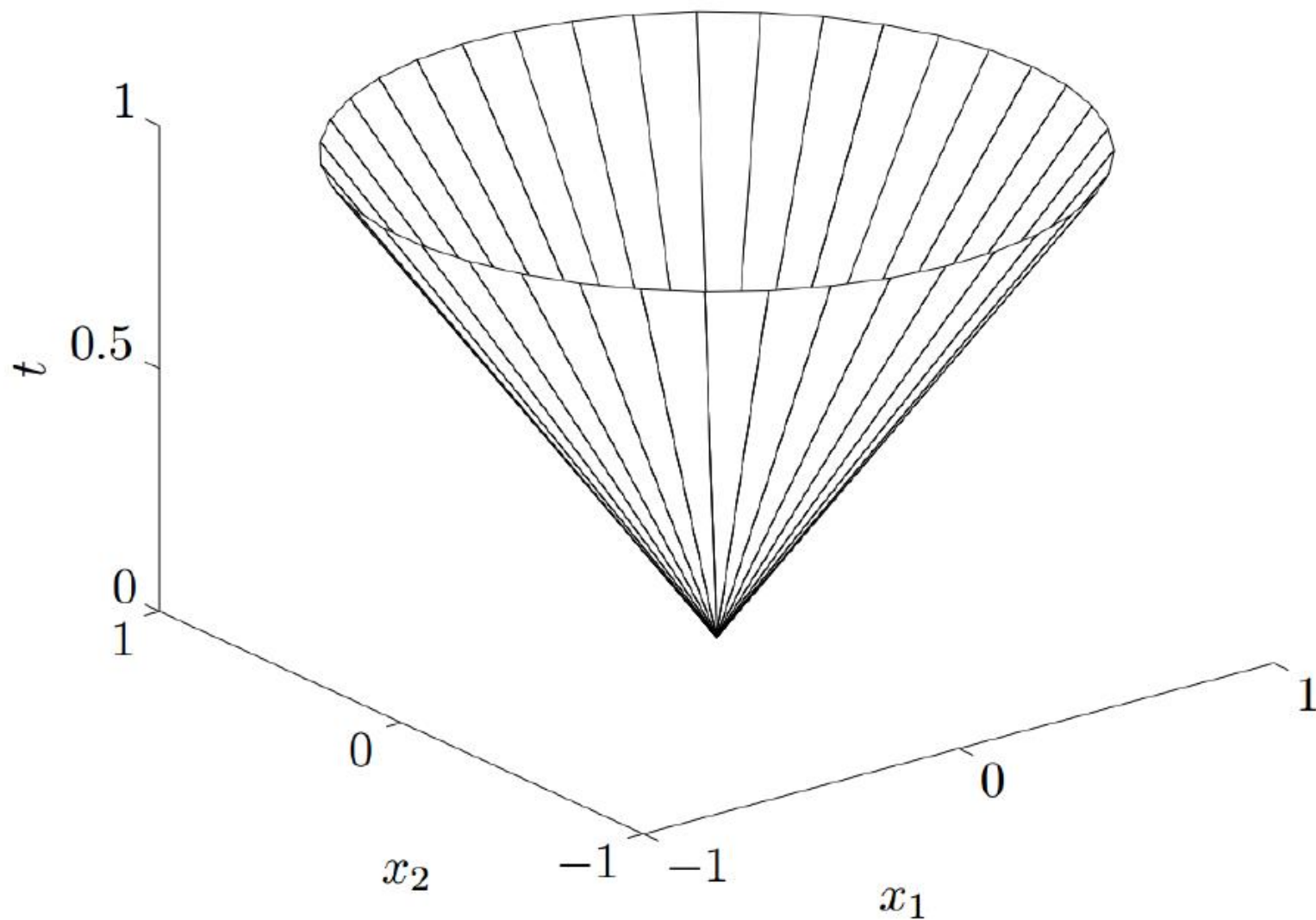
范数球:

$$B(x_c, r) = \{x \mid \|x - x_c\| \leq r\}$$

范数锥:

$$\{(x, t) \mid \|x\| \leq t\}$$

# 例子：二阶锥



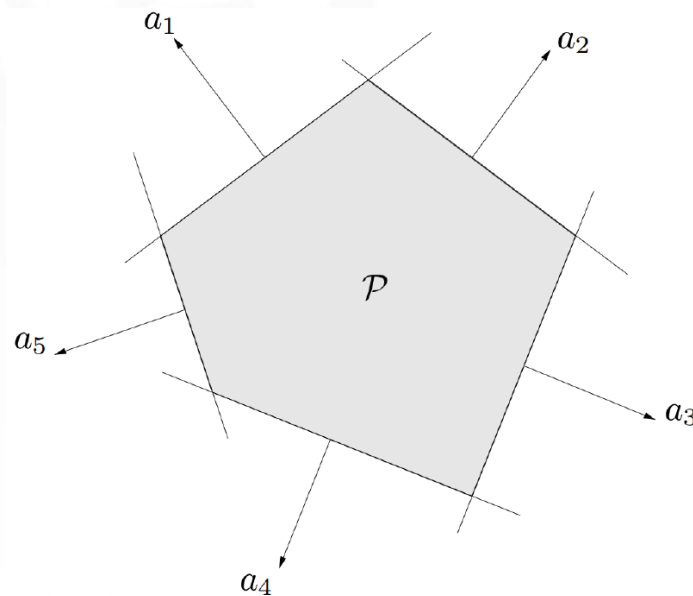
# 多面体

多面体有限个半空间和超平面的交集：

$$P = \{x \mid a_j^T x \leq b_j, c_i^T x = d_i\}$$

仿射集（如超平面、直线）、射线、线段、半空间都是多面体

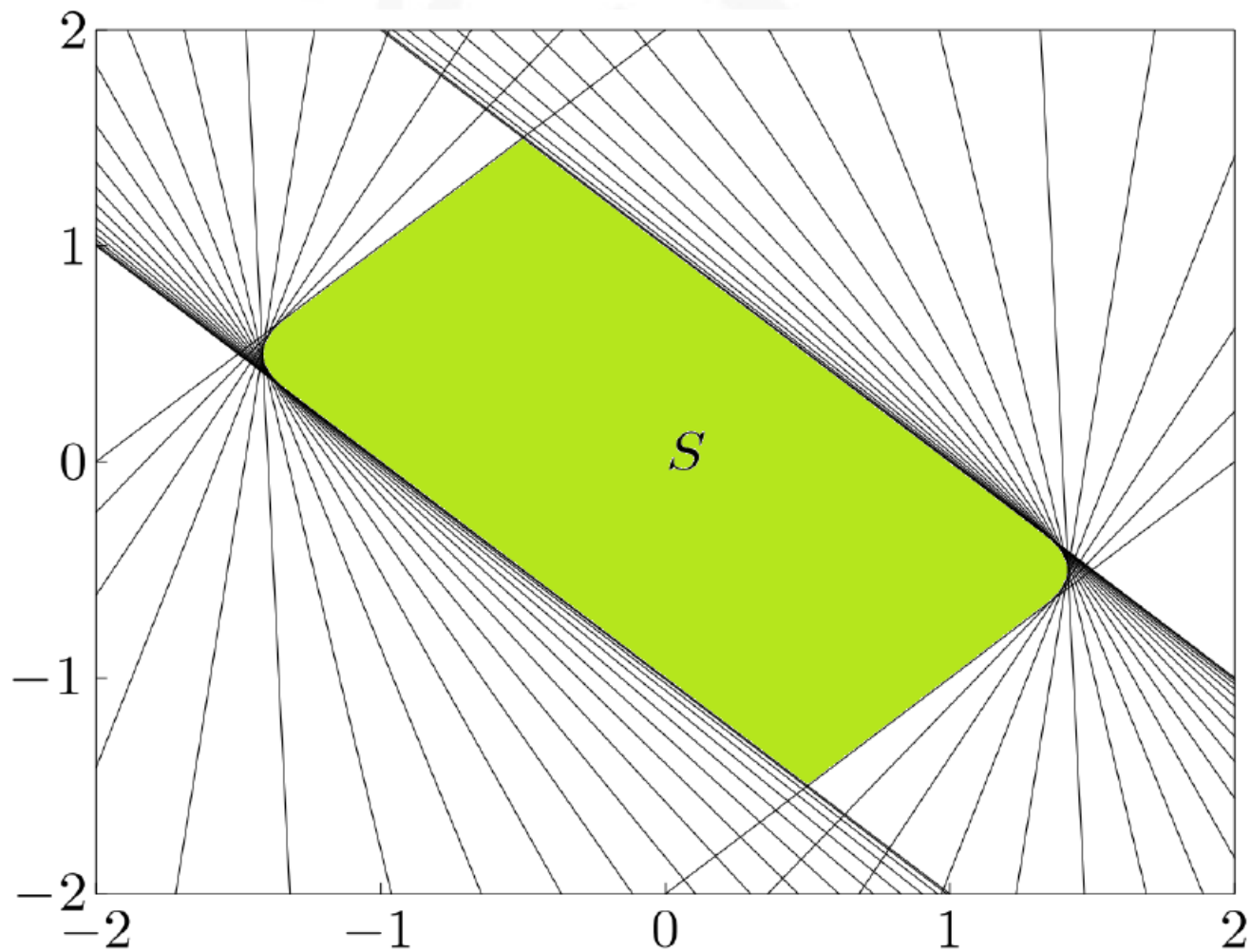
多面体是凸集





# 保凸运算

## 集合交运算



# 保凸运算

仿射变换  $f(x) = Ax + b$  ,  $A \in R^{m \times n}$  ,  $b \in R^m$

若 $f$ 是仿射变换,  $f: \mathbf{R}^n \rightarrow R^m$  ,  $f(S) = \{f(x) | x \in S\}$

- 若 $S$ 为凸集, 则 $f(S)$ 为凸集
- 若 $f(S)$ 为凸集, 则 $S$ 为凸集

# 保凸运算

两个凸集的和为凸集

$$S_1 + S_2 = \{x + y \mid x \in S_1, y \in S_2\}$$

两个凸集的笛卡尔积（直积）为凸集

$$S_1 \times S_2 = \{(x_1, x_2) \mid x_1 \in S_1, x_2 \in S_2\}$$

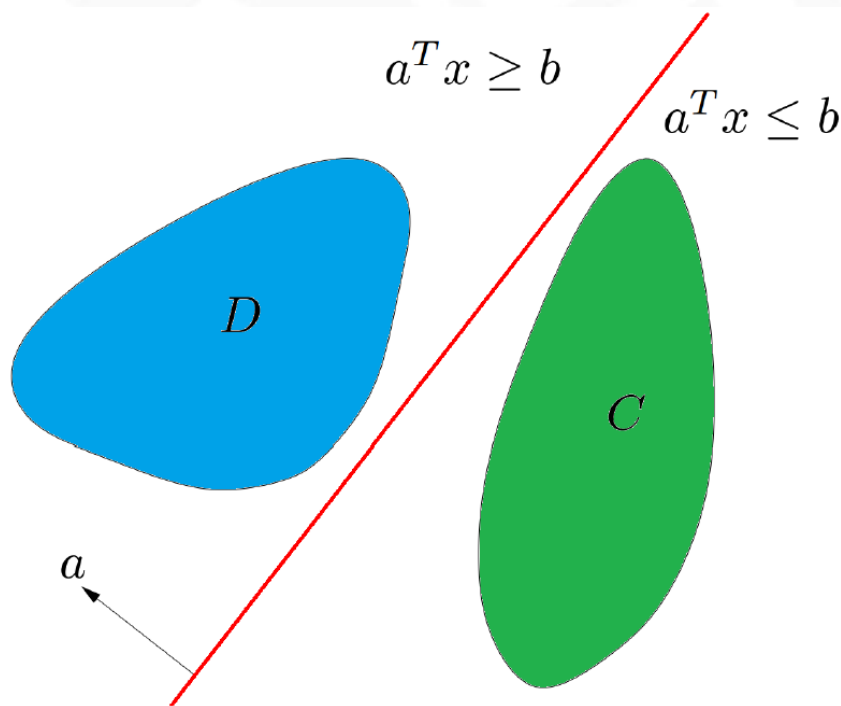
两个凸集的部分和为凸集

$$S = \{(x, y_1 + y_2) \mid (x, y_1) \in S_1, (x, y_2) \in S_2\}$$

# 分割超平面

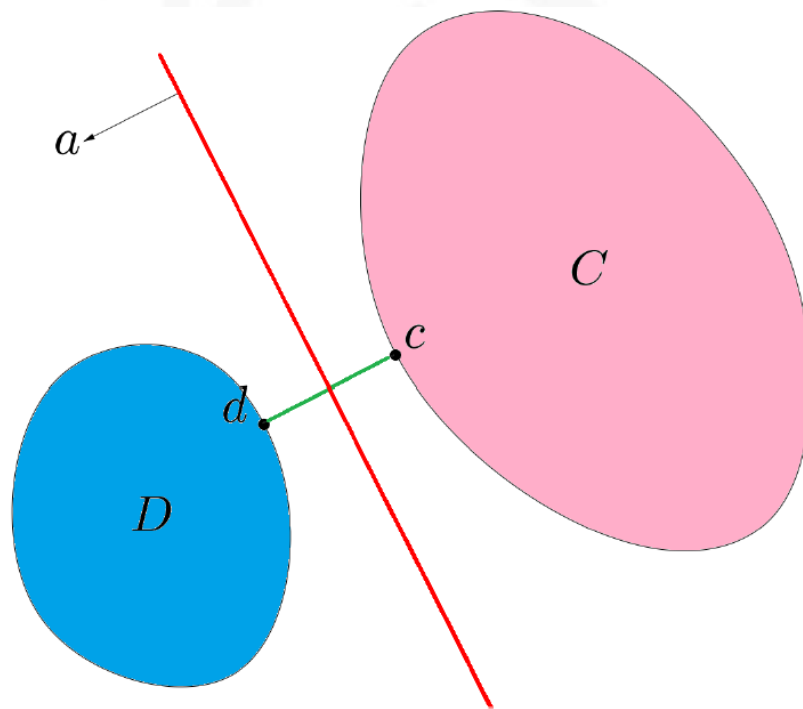
设C和D为两不相交的凸集，则存在超平面P，P可以将C和D分离

$$\forall x \in C, a^T x \leq b \text{ 且 } \forall x \in D, a^T x \geq b$$



# 分割超平面的构造

两个集合的距离定义为两个集合间元素的最短距离



做集合 $C$ 和集合 $D$ 的最短线段的垂直平分线

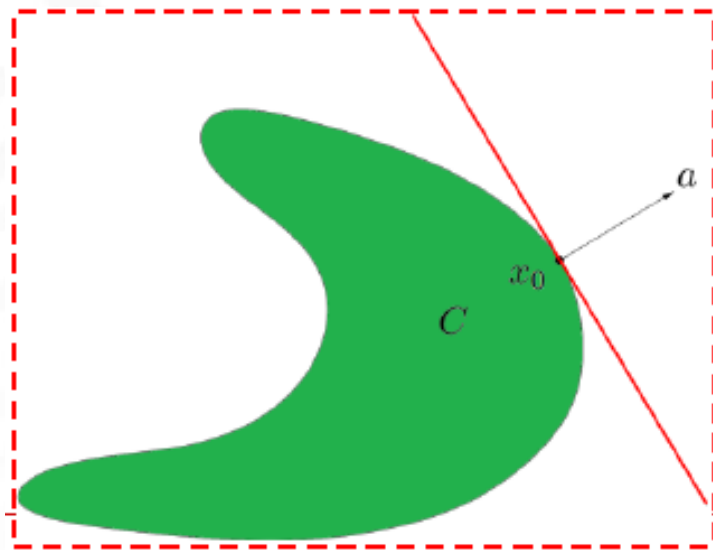
# 支撑超平面

设集合 $C$ ， $x_0$ 为 $C$ 边界上的点。若存在 $a \neq 0$  满足对任意 $x \in C$  都有  $a^T x \leq a^T x_0$  成立，则称超平面  $\{x | a^T x = a^T x_0\}$  为集合 $C$ 在点 $x_0$ 处的支撑超平面

- 凸集边界上任意一点，均存在支撑超平面

若一个闭的非中空（内部点不为

空）集合，在边界上的任意一点均存在支撑超平面，则该集合为凸集。



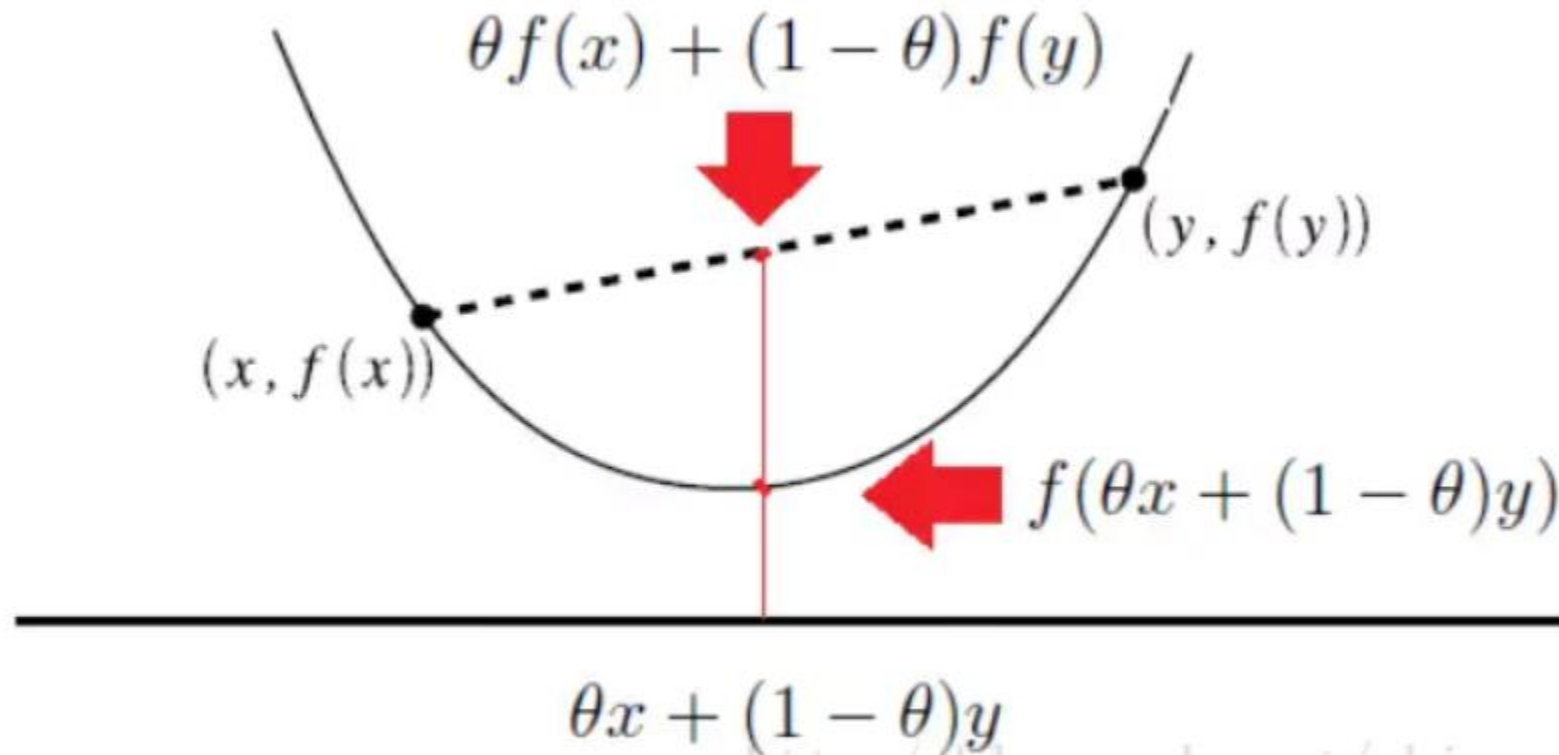
# 凸函数

## □ 凸函数基本概念

# 凸函数定义1

定义1:  $f: \mathbf{R}^n \rightarrow \mathbf{R}$  为凸函数当且仅当(1)  $\text{dom} f$  为凸集

(2)  $\forall x, y \in \text{dom} f, \forall 0 \leq \theta \leq 1$ , 有  $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$



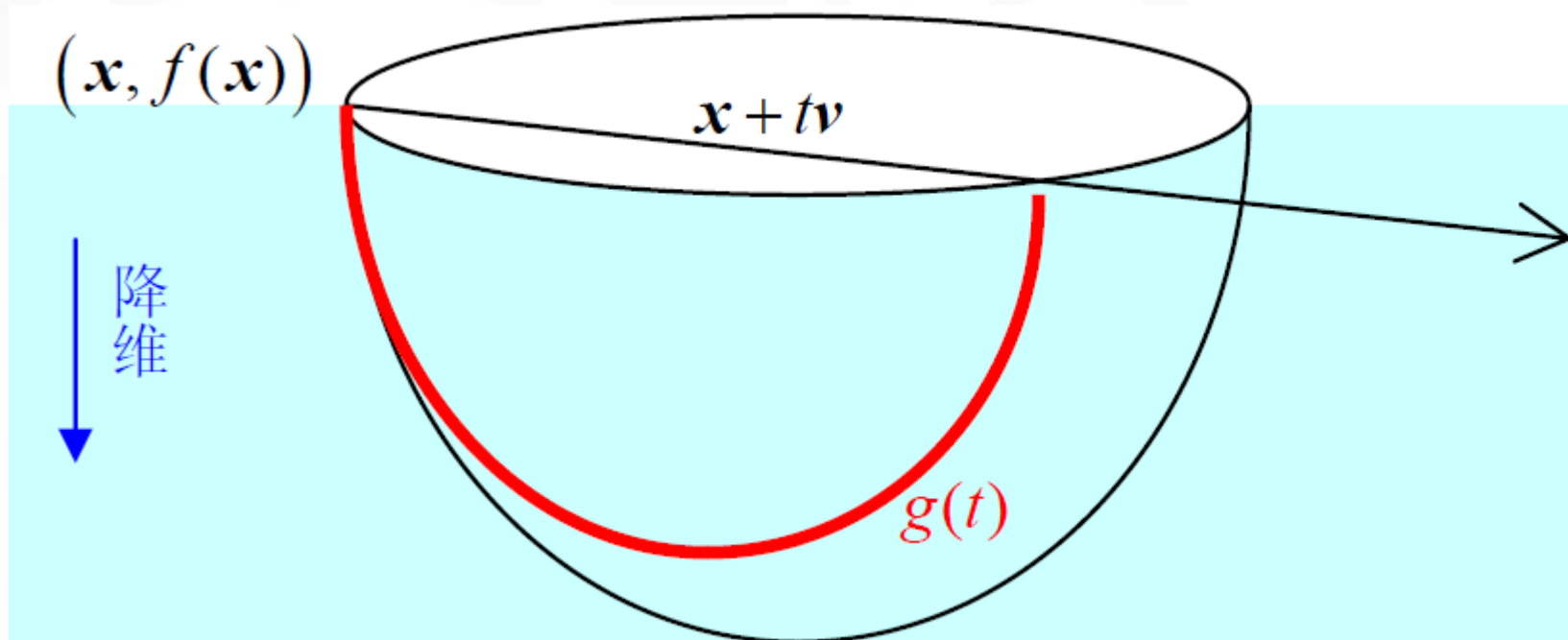


# 凸函数定义2

定义2:  $f: \mathbf{R}^n \rightarrow \mathbf{R}$  为凸函数当且仅当(1)  $\text{dom } f$  为凸集

(2)  $\forall x \in \text{dom } f, \forall v \in \mathbf{R}^n, g(t) = f(x + tv)$  在  $\text{dom } g = \{t \mid x + tv \in \text{dom } f\}$

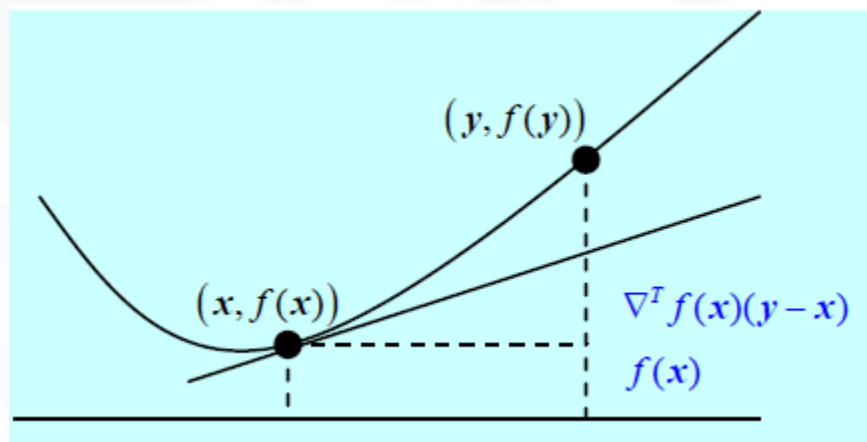
上是凸的,即:  $\forall 0 \leq \theta \leq 1, g(\theta t_1 + (1 - \theta)t_2) \leq \theta g(t_1) + (1 - \theta)g(t_2)$



# 凸函数定义3

定义3:  $f: \mathbf{R}^n \rightarrow \mathbf{R}$  为凸函数当且仅当(1)  $\text{dom } f$  为凸集

(2) 若  $f$  可微,  $\forall x, y \in \text{dom } f, f(y) \geq f(x) + \nabla f^T(x)(y - x)$



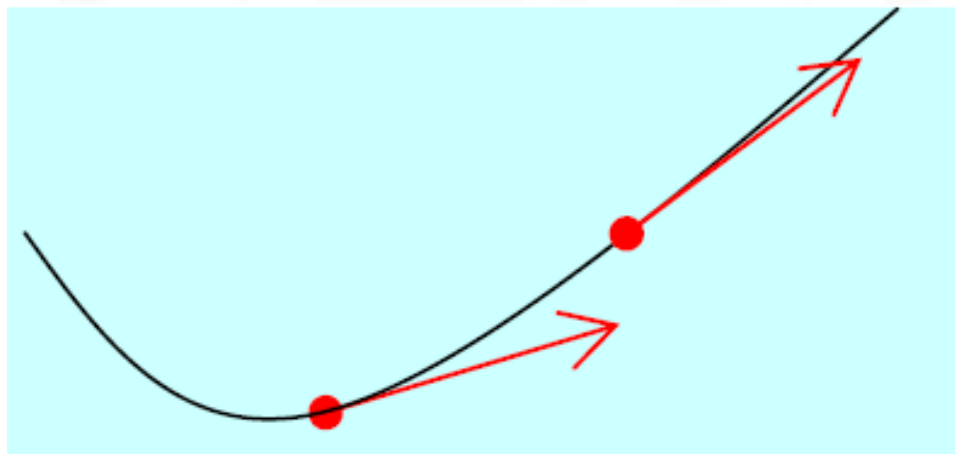
性质: 若  $f$  为凸函数,  $\exists x_0 \in \text{dom } f$  使  $\nabla f(x_0) = 0$ , 对

$$\forall y \in \text{dom } f, f(y) \geq f(x_0) + \nabla^T f(x_0)(y - x_0) = f(x_0)$$

则  $f(x_0)$  是  $f$  的最小值

# 凸函数定义4

定义4: 设函数  $f: \mathbf{R}^n \rightarrow \mathbf{R}$  二阶可微, 即  $\nabla^2 f$  在  $\text{dom} f$  均存在, 则  $f$  为凸函数当且仅当  $\text{dom} f$  为凸集, 且对  $\forall x \in \text{dom} f$  有  $\nabla^2 f$  半正定



# 凸函数

例：二次函数  $f: \mathbf{R}^n \rightarrow \mathbf{R}$ ,  $\text{dom}f = \mathbf{R}^n$ ,  $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{P} \mathbf{x} + \mathbf{q}^T \mathbf{x} + r$  ( $\mathbf{P} \in \mathbf{S}^n$ ,  $\mathbf{q} \in \mathbf{R}^n$ ,  $r \in \mathbf{R}$ )

$\nabla^2 f(\mathbf{x}) = \mathbf{P}$ ,  
if  $\mathbf{P} \succeq 0$ , convex function  
if  $\mathbf{P} \succ 0$ , strictly convex function  
if  $\mathbf{P} \preceq 0$ , concave function

例：仿射函数  $f: \mathbf{R}^n \rightarrow \mathbf{R}^m$ ,  $\text{dom}f = \mathbf{R}^n$ ,  $f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$  ( $\mathbf{A} \in \mathbf{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbf{R}^m$ )

$\nabla^2 f(\mathbf{x}) = \mathbf{0}$ , 故  $f(\mathbf{x})$  即是凸函数又是凹函数

例：指数函数

$$f(x) = e^{ax}, x \in \mathbf{R}$$

$$f'(x) = e^{ax} a$$

$$f''(x) = e^{ax} a^2 \geq 0$$

故指数函数是凸函数

例：幂函数

$$f(x) = x^a, x \in \mathbf{R}_{++}$$

$$f'(x) = ax^{a-1}$$

$$f''(x) = a(a-1)x^{a-2} \begin{cases} \geq 0 & a \geq 1 \text{ or } a \leq 0 \\ \leq 0 & 0 \leq a \leq 1 \end{cases}$$

# 凸函数

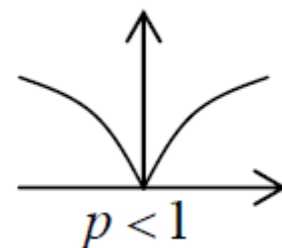
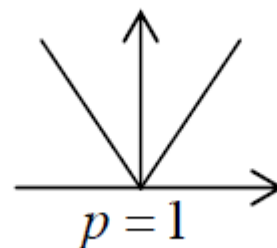
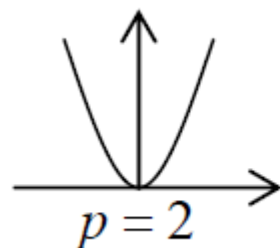
例：绝对值的幂函数

$$f(x) = |x|^p, x \in \mathbb{R}, p \geq 0 (\text{avoid singularity})$$

$$f'(x) = \begin{cases} px^{p-1} & x \geq 0 \\ -p(-x)^{p-1} & x < 0 \end{cases}$$

$$f''(x) = \begin{cases} p(p-1)x^{p-2} & x \geq 0 \\ p(p-1)(-x)^{p-2} & x < 0 \end{cases}$$

$$\text{故 } f(x) \text{ is } \begin{cases} \text{convex} & p \geq 2 \\ \text{convex} & p \in [1, 2) \\ \text{not convex, not concave} & p \in (0, 1) \end{cases}$$



# 凸函数

例：对数函数

$$f(x) = \log x, x \in \mathbf{R}_{++}$$

$$f'(x) = \frac{1}{x}$$

$$f''(x) = -\frac{1}{x^2} < 0$$

故对数函数是凹函数

例：负熵

$$f(x) = x \log x, x \in \mathbf{R}_{++}$$

$$f'(x) = 1 + \log x$$

$$f''(x) = \frac{1}{x} > 0$$

故负熵是凸函数

例：范数函数是凸函数

极大值函数是凸函数

log-sum-exp函数是凸函数： $f(\mathbf{x}) = \log(e^{x_1} + \cdots + e^{x_n}) \quad \mathbf{x} \in \mathbf{R}^n$

# 保凸操作

保持凸函数凸性的操作

➤ 非负加权和 (Non-negative weighted sum)

先求值，再做线性变换  
变换的是值域

1) 若  $f_1, \dots, f_m$  为凸函数，则  $f \triangleq \sum_{i=1}^m \omega_i f_i$ ,  $\omega_i \geq 0$  为凸函数。

2) 若  $f(\mathbf{x}, y)$  对  $\forall y \in A$  均为  $\mathbf{x}$  的凸函数，则  $g(\mathbf{x}) \triangleq \int_{y \in A} \omega(y) f(\mathbf{x}, y) dy$ ,  $\omega(y) \geq 0$  为凸函数。

➤ 仿射映射的复合 (Composition with an affine mapping)

先做线性变换，再求值  
变换的是定义域

若  $f(\mathbf{x}): \mathbf{R}^n \rightarrow \mathbf{R}$  为凸函数，则  $g(\mathbf{x}) \triangleq f(A\mathbf{x} + \mathbf{b})$ ,  $A \in \mathbf{R}^{n \times n}, \mathbf{b} \in \mathbf{R}^n, A\mathbf{x} + \mathbf{b} \in \text{dom}f$  为凸函数。

➤ 两个函数的极大值 (Point-wise Maximum)

若  $f_1, f_2$  为凸函数，则  $f(\mathbf{x}) \triangleq \max\{f_1(\mathbf{x}), f_2(\mathbf{x})\}$ ,  $\text{dom}f = \text{dom}f_1 \cap \text{dom}f_2$  为凸函数。

# 保凸操作

## ➤ 函数的复合 (Composition of functions)

$$\begin{aligned} h: \mathbf{R}^k &\rightarrow \mathbf{R} \\ g: \mathbf{R}^n &\rightarrow \mathbf{R}^k, \quad f = h \circ g: \mathbf{R}^n \rightarrow \mathbf{R}, \quad \text{dom} f = \{\mathbf{x} \in \text{dom} g \mid g(\mathbf{x}) \in \text{dom} h\} \end{aligned}$$

考虑一维情况, 假定  $h, g$  的定义域为全空间, 且它们均二阶可微

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

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

$$f''(x) = h''(g(x))(g'(x))^2 + h'(g(x))g''(x)$$

若  $g$  为凸,  $h$  为凸且单增, 则  $f$  为凸

若  $g$  为凹,  $h$  为凸且单减, 则  $f$  为凸

若  $g$  为凹,  $h$  为凹且单增, 则  $f$  为凹

若  $g$  为凸,  $h$  为凹且单减, 则  $f$  为凹



# □ 凸优化问题

# 凸优化问题

## 优化问题的标准形式

$$\begin{array}{ll}\min & f_o(\mathbf{x}) \\ \text{s.t.} & f_i(\mathbf{x}) \leq 0 \quad i = 1, \dots, m \\ & h_j(\mathbf{x}) = 0 \quad j = 1, \dots, p\end{array}$$

$\mathbf{x}$ : 优化变量 (optimization variable)

$f_o$ : 目标函数/损失函数 (objective function / cost function)

$f_i$ : 不等式约束 (inequality constraint)

$h_i$ : 等式约束 (equality constraint)

# 凸优化问题

域 (Domain)

$$\mathcal{D} = \bigcap_{i=0}^m \text{dom } f_i \cap \bigcap_{i=1}^p \text{dom } h_i,$$

可行解集 (feasible set)

$$X = \left\{ \mathbf{x} \in D \left| \begin{array}{ll} f_i(\mathbf{x}) \leq 0 & i = 1, \dots, m \\ h_j(\mathbf{x}) = 0 & j = 1, \dots, p \end{array} \right. \right\}$$

局部最优值 (local optimal value)

$$\exists R > 0, \quad p_{\text{loc}} = \inf \{ f_o(\mathbf{z}) \mid \|\mathbf{z} - \mathbf{x}\|_2 \leq R, \mathbf{x} \in X, \mathbf{z} \in X \}$$

局部最优解集 (local optimal set)

$$X_{\text{loc}} = \{ \mathbf{x}_{\text{loc}} \in X \mid f_o(\mathbf{x}_{\text{loc}}) = p_{\text{loc}} \}$$

最优值 (optimal value)

$$p^* = \inf \{ f_o(\mathbf{x}) \mid \mathbf{x} \in X \}$$

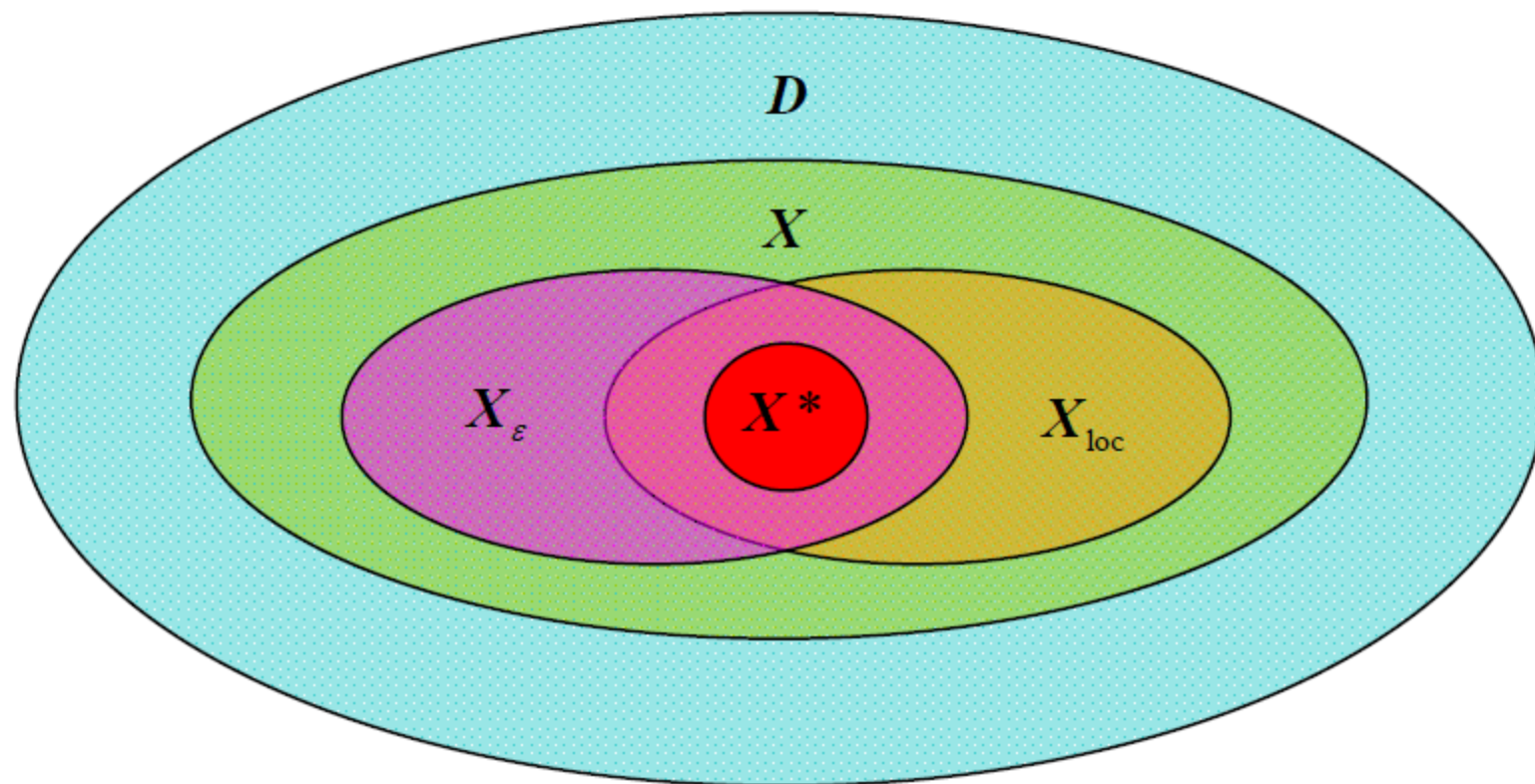
最优解集 (optimal set)

$$X^* = \{ \mathbf{x}^* \in X \mid f_o(\mathbf{x}^*) = p^* \}$$

$\varepsilon$  次优解 ( $\varepsilon$ -suboptimal set)

$$X_\varepsilon = \{ \mathbf{x} \in X \mid f_o(\mathbf{x}) = p^* + \varepsilon \}$$

# 凸优化问题



# 凸优化问题

凸优化问题 (Convex Optimization Problems)

$$\min \quad f_o(\mathbf{x})$$

$$\text{s.t.} \quad f_i(\mathbf{x}) \leq 0 \quad i = 1, \dots, m$$

$$A_j^T \mathbf{x} = b_j \quad j = 1, \dots, p$$

1) 目标函数  $f_o(\mathbf{x})$  为凸函数

2) 不等式约束  $f_1(\mathbf{x}), \dots, f_m(\mathbf{x})$  为凸函数 (Note: 凸函数  $f_i(\mathbf{x}) \leq 0$  的解集一定是凸集)

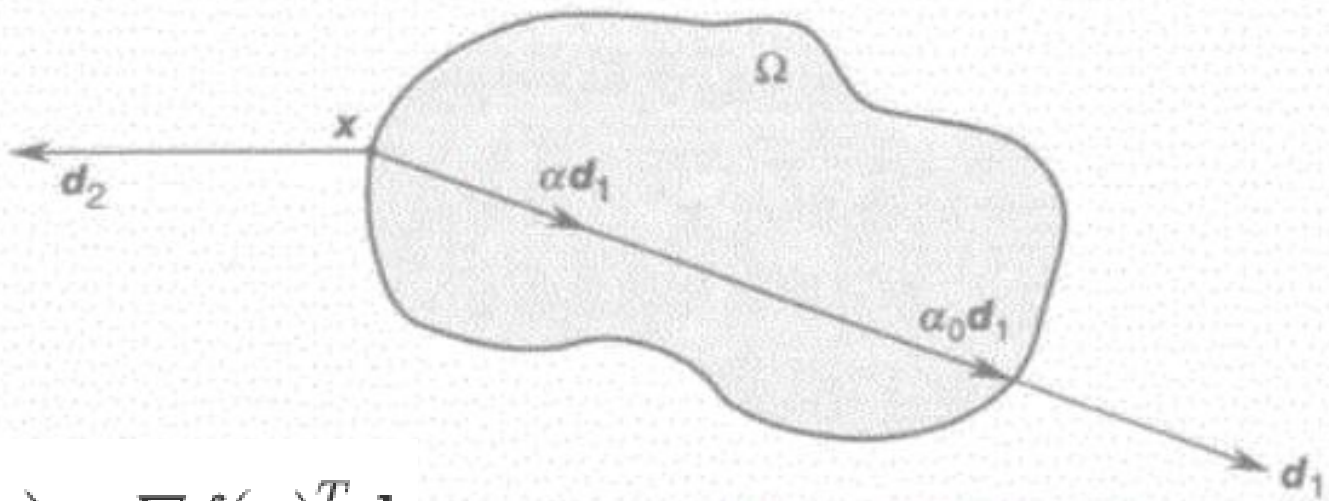
3) 等式约束  $A_1^T \mathbf{x} = b_1, \dots, A_p^T \mathbf{x} = b_p$  为仿射函数 (Note: 仿射函数  $A_j^T \mathbf{x} = b_j$  的解集一定是凸集)

凸优化问题的重要性质:

局部最优=全局最优!

# 凸优化问题

定义 对于向量  $\mathbf{d} \in \mathbf{R}^n$ ,  $\mathbf{d} \neq 0$  和约束集中的某个点  $\mathbf{x} \in \Omega$ , 如果存在一个实数  $a_0 > 0$ , 使得对于所有的  $a \in [0, a_0]$ ,  $\mathbf{x} + a\mathbf{d}$  仍然在约束集内, 即  $\mathbf{x} + a\mathbf{d} \in \Omega$ , 则称  $\mathbf{d}$  为  $\mathbf{x}$  处的可行方向.

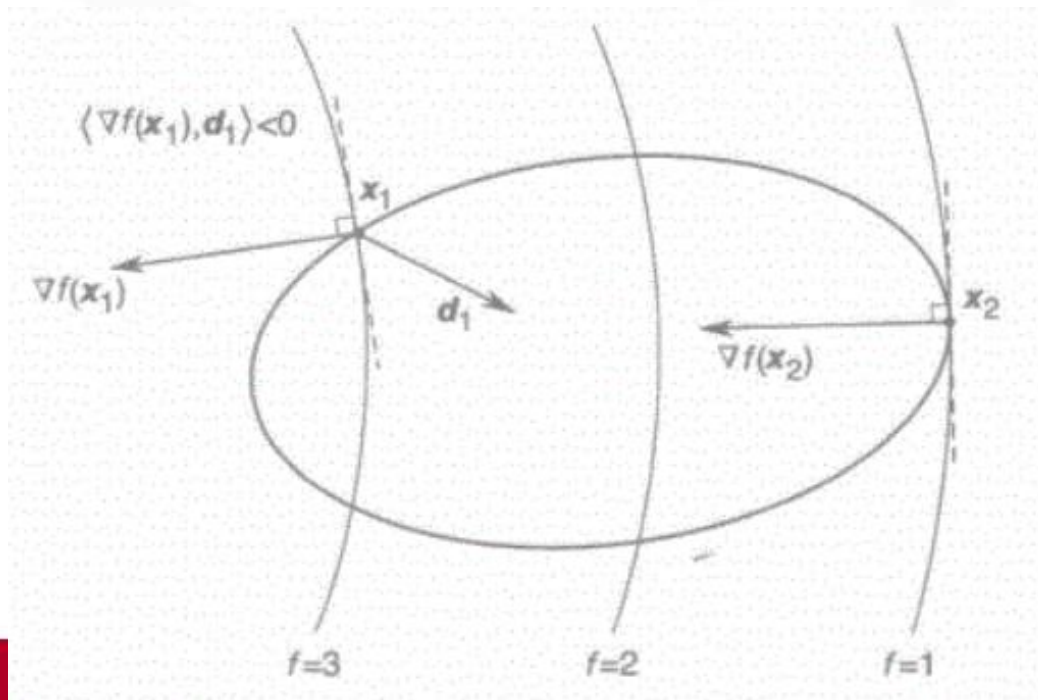


$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \nabla f(\mathbf{x})^T \mathbf{d}$$

# 凸优化问题

定理（一阶必要条件）多元实值函数  $f$  在约束集  $\Omega$  上一阶连续可微，约束集  $\Omega$  是  $\mathbf{R}^n$  的子集。如果  $\mathbf{x}^*$  是函数  $f$  在  $\Omega$  上的局部极小点，则对于  $\mathbf{x}^*$  处的任意可行方向  $\mathbf{d}$ ，都有

$$\nabla f(\mathbf{x})^T \mathbf{d} \geq 0$$



# 凸优化问题

可微目标函数下最优解的性质 (optimality criterion for differentiable  $f_o$ )

若  $f_o$  为可微且为凸函数, 则  $f_o(y) \geq f_o(x) + \nabla^T f_o(x)(y - x)$ ,  $\forall x, y \in \text{dom} f_o$

对无约束问题  $\begin{cases} \min & f_o(x) \\ \text{s.t.} & x \in \text{dom} f_o \end{cases}$  其最优解  $x^*$  满足  $\nabla^T f_o(x^*)(y - x^*) \geq 0, \forall y \in \text{dom} f_o$

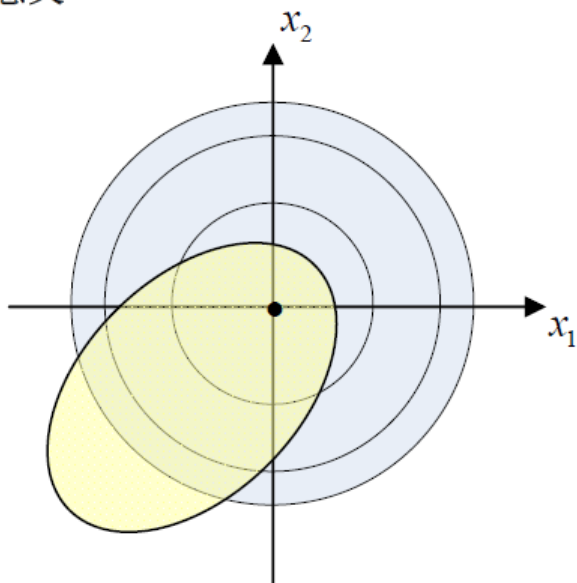
同理

对有约束问题  $\begin{cases} \min & f_o(x) \\ \text{s.t.} & x \in \text{dom} f_o \cap \underbrace{X}_{\text{可行解集}} \end{cases}$  其最优解  $x^*$  满足  $\nabla^T f_o(x^*)(y - x^*) \geq 0, \forall y \in \text{dom} f_o \cap X$



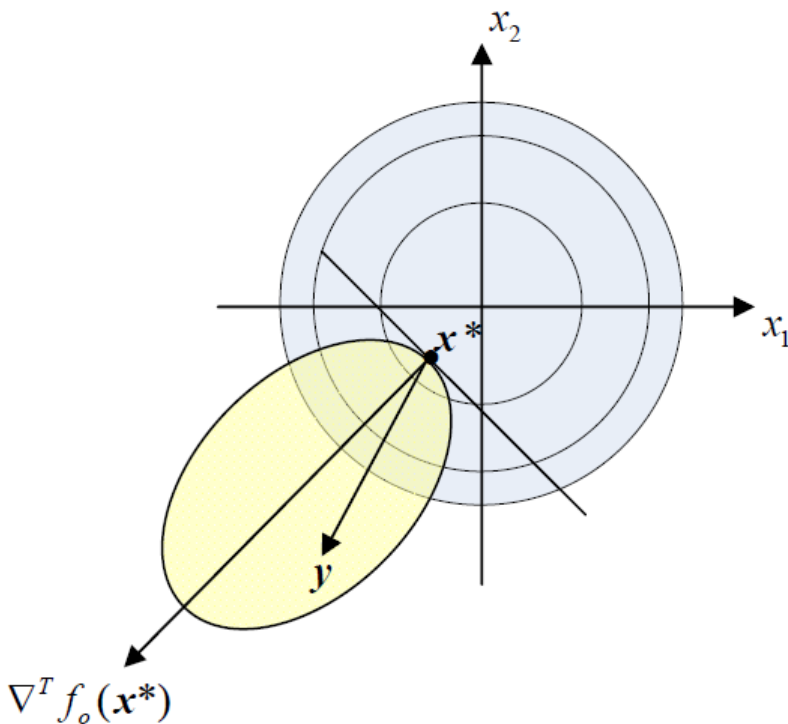
# 凸优化问题

几何意义



$$\nabla^T f_o(\mathbf{x}^*) = 0$$

$$\nabla^T f_o(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) = 0, \forall \mathbf{y} \in \text{dom} f_o$$



The angle between  $\nabla^T f_o(\mathbf{x}^*)$  and  $(\mathbf{y} - \mathbf{x}^*)$  is acute

$$\nabla^T f_o(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) > 0, \forall \mathbf{y} \in \text{dom} f_o$$

# 凸优化问题

线性规划问题 (Linear Programming/LP)

$$\left\{ \begin{array}{ll} \min & \mathbf{c}^T \mathbf{x} + d \\ \text{s.t.} & \mathbf{G}\mathbf{x} \preceq \mathbf{h} \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} \right\} \longrightarrow P : \text{多面体}$$

二次规划 (Quadratic Programming / QP)

$$\left\{ \begin{array}{ll} \min & \frac{1}{2} \mathbf{x}^T \mathbf{P}\mathbf{x} + \mathbf{q}^T \mathbf{x} + r \\ \text{s.t.} & \mathbf{G}\mathbf{x} \preceq \mathbf{h} \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} \right.$$

# 凸优化问题

二次约束二次规划 (Quadratically Constrained Quadratic Programming / QCQP)

$$\begin{cases} \min & \frac{1}{2} \mathbf{x}^T \mathbf{P} \mathbf{x} + \mathbf{q}^T \mathbf{x} + r \\ \text{s.t.} & \frac{1}{2} \mathbf{x}^T \mathbf{P}_i \mathbf{x} + \mathbf{q}_i^T \mathbf{x} + r_i \leq 0, i = 1, \dots, m \\ & \mathbf{A} \mathbf{x} = \mathbf{b} \end{cases}$$

这里要求  $\mathbf{P} \succ \mathbf{0}$ ,  $\mathbf{P}_i \succ \mathbf{0}, i = 1, \dots, m$

# 凸优化问题

例：线性测量方程—— $\mathbf{b} = \mathbf{A}\mathbf{x} + \rho$ （ $\rho$  为误差项）

1) 选择合适的  $\mathbf{x}^*$  使误差项最小，即

$$\min \quad \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2$$

$$\min \quad \mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x} - 2\mathbf{b}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{b}$$

——QP

2-1) 一范数规范化优化问题

$$\min \quad \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_1 \|\mathbf{x}\|_1$$

2-2) 一范数约束优化问题

$$\begin{cases} \min & \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 \\ \text{s.t.} & \|\mathbf{x}\|_1 \leq \varepsilon_1 \end{cases}$$

# 凸优化问题

3-1) 二范数规范化优化问题

$$\min \quad \| \mathbf{b} - \mathbf{A}\mathbf{x} \|_2^2 + \lambda_2 \| \mathbf{x} \|_2^2$$

3-2) 二范数约束优化问题

$$\begin{cases} \min & \| \mathbf{b} - \mathbf{A}\mathbf{x} \|_2^2 \\ \text{s.t.} & \| \mathbf{x} \|_2^2 \leq \varepsilon_2 \end{cases}$$

# 凸优化问题

□ 对偶性

# 凸优化问题

## 标准最优化问题

$$\begin{array}{ll} \min & f_o(\mathbf{x}) \\ \text{s.t.} & f_i(\mathbf{x}) \leq 0 \quad i = 1, \dots, m \\ & h_j(\mathbf{x}) = 0 \quad j = 1, \dots, p \end{array}$$

域 (Domain):  $\mathbf{D} \triangleq \bigcap_{i=1}^m \text{dom} f_i \cap \bigcap_{j=1}^p \text{dom} h_j$

可行解集 (feasible set):  $\mathbf{X} = \left\{ \mathbf{x} \in \mathbf{D} \left| \begin{array}{ll} f_i(\mathbf{x}) \leq 0 & i = 1, \dots, m \\ h_j(\mathbf{x}) = 0 & j = 1, \dots, p \end{array} \right. \right\}$

最优值 (optimal value):  $p^* = \inf \{ f_o(\mathbf{x}) \mid \mathbf{x} \in \mathbf{X} \}$

最优解 (optimal solution)  $\mathbf{x}^*$ :  $f_o(\mathbf{x}^*) = p^*$

## 拉格朗日函数 (Lagrangian Function)

$$L(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) = f_o(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^p v_j h_j(\mathbf{x})$$

$$\text{dom} L = \mathbf{D} \times \mathbf{R}^m \times \mathbf{R}^p$$

$\mathbf{x}$ : 原变量 (primal variable)

$\boldsymbol{\lambda}, \mathbf{v}$ : 对偶变量 (dual variable)

$\lambda_i, v_i$ : 拉格朗日乘子 (Lagrange Multiplier)

# 凸优化问题

## 对偶函数 (Lagrange Dual Function)

$$g(\lambda, \nu) = \inf_{\mathbf{x} \in D} L(\mathbf{x}, \lambda, \nu) = \inf_{\mathbf{x} \in D} \left( f_o(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^p \nu_j h_j(\mathbf{x}) \right)$$

(1) 对偶函数一定为凹函数

(2)  $\forall \lambda \succeq \mathbf{0}, \forall \nu, \quad g(\lambda, \nu) \leq p^*$

例: 
$$\begin{cases} \min & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} & \mathbf{Ax} = \mathbf{b} \\ & \mathbf{x} \succeq \mathbf{0} \end{cases}$$



# 凸优化问题

## 对偶问题 (Lagrange Dual Problem)

$$\begin{cases} \max & g(\lambda, v) \\ \text{s.t.} & \lambda \succeq \mathbf{0} \end{cases}$$

最优值 (optimal value:)  $d^* = \sup\{g(\lambda, v) \mid \lambda, v \in \text{dom } g \& \lambda \succeq 0\}$

## 原问题 (Primal Problem)

$$\begin{aligned} \min & \quad f_o(\mathbf{x}) \\ \text{s.t.} & \quad f_i(\mathbf{x}) \leq 0 & i = 1, \dots, m \\ & \quad h_j(\mathbf{x}) = 0 & j = 1, \dots, p \end{aligned}$$

最优值 (optimal value):  $p^* = \inf \{f_o(\mathbf{x}) \mid \mathbf{x} \in X\}$

$$d^* \leq p^* \quad \text{弱对偶性 (Weak Duality)}$$

# 凸优化问题

$d^* = p^*$ ——强对偶性 (Strong Duality)

$p^* - d^*$ ——对偶间隙 (Duality Gap)

定义:  $D$  的 Relative Interior

$\text{relint}D = \{ \mathbf{x} \in D \mid \exists r > 0, (B(\mathbf{x}, r) \cap \text{aff}D) \in D \}$  (即  $D$  去掉边界后的集合)

**Slater's Condition**

$\exists \mathbf{x} \in \text{relint}D$

$$\begin{cases} f_i(\mathbf{x}) < 0 & i = 1, \dots, m \\ A_j^T \mathbf{x} = \mathbf{b}_j & j = 1, \dots, p \end{cases}$$

# 凸优化问题

## Refined Slater's Condition

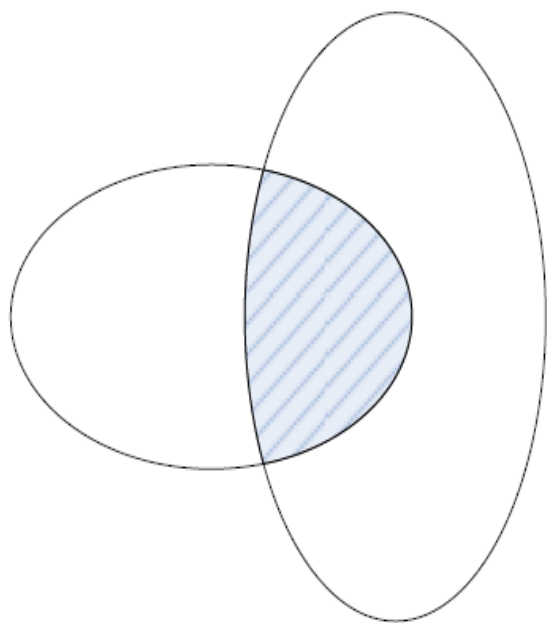
$$\exists \mathbf{x} \in \text{relint} D$$

$$\begin{cases} f_i(\mathbf{x}) \leq 0 & i = 1, \dots, k & \text{when } f_i \text{ is affine} \\ f_i(\mathbf{x}) < 0 & i = k + 1, \dots, m \\ \mathbf{A}_j^T \mathbf{x} = \mathbf{b}_j & j = 1, \dots, p \end{cases}$$

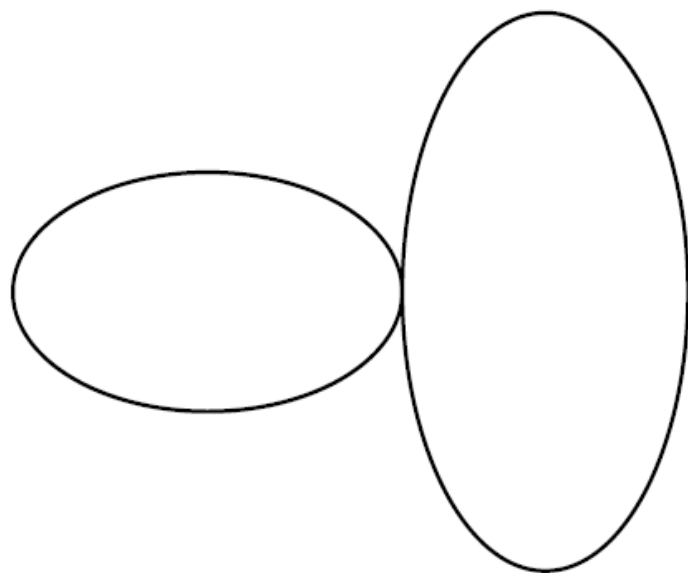
定理:

- ①对于非凸问题, 其对偶问题通常没有强对偶性
- ②对于凸问题, 有充分条件 (**Slater's Condition**), 使得该问题的对偶问题满足强对偶性

# 凸优化问题



1) 原问题满足 Slater's Condition  
故对偶问题满足强对偶性



2) 原问题不满足 Slater's Condition  
但因为可行解集只有一个解（必为最优解）  
故对偶问题也满足强对偶性

# 凸优化问题

几何解释:

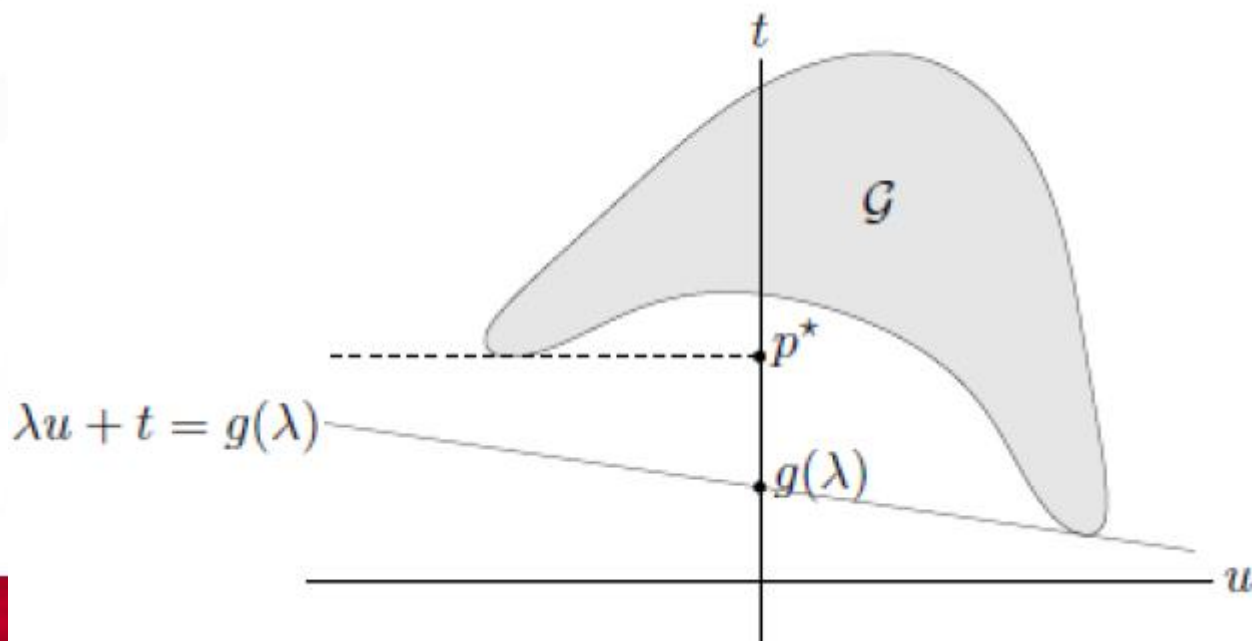
考虑如下单约束单目标优化问题

$$\begin{cases} \min & f_0(x) \\ \text{s.t.} & f_1(x) \leq 0 \end{cases}$$

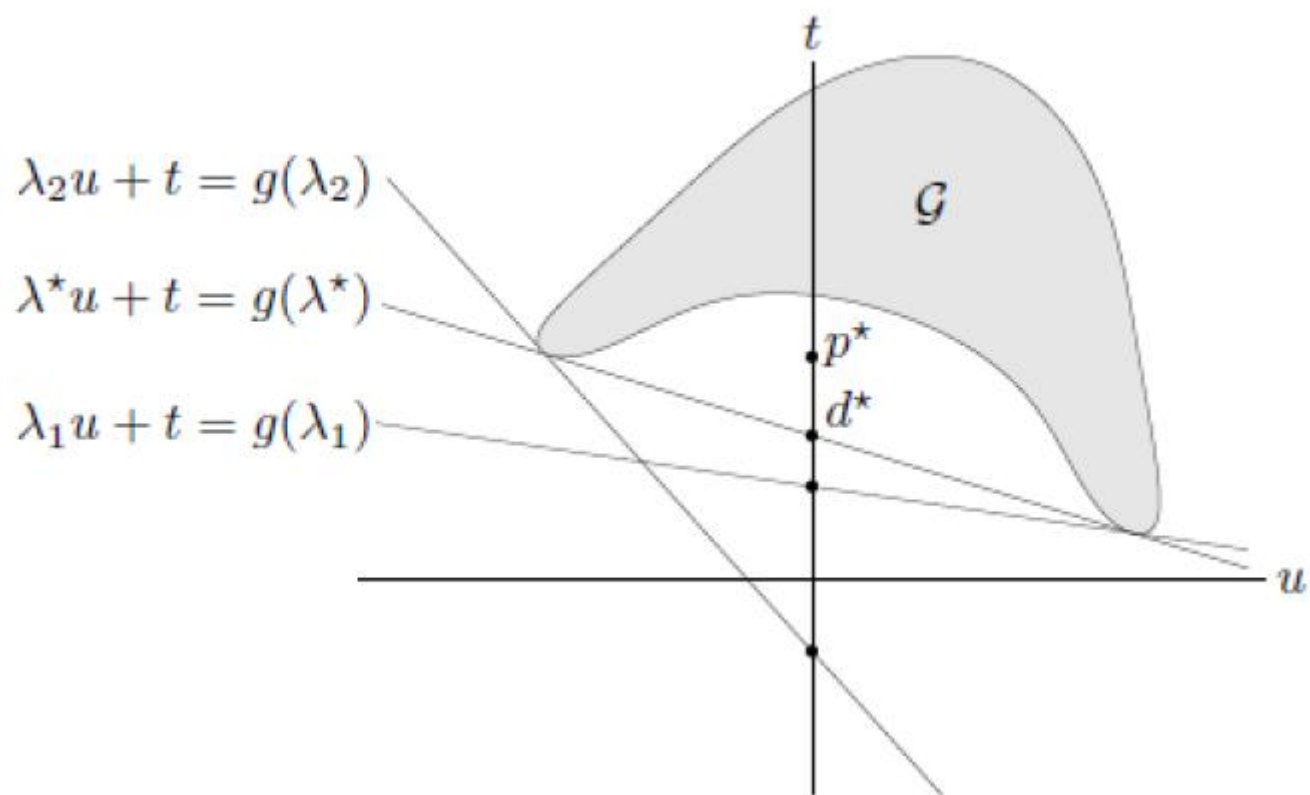
$$\text{设 } \mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in D\}$$

$$\text{则 } g(\lambda) = \inf \{t + \lambda u \mid (u, t) \in \mathcal{G}\}$$

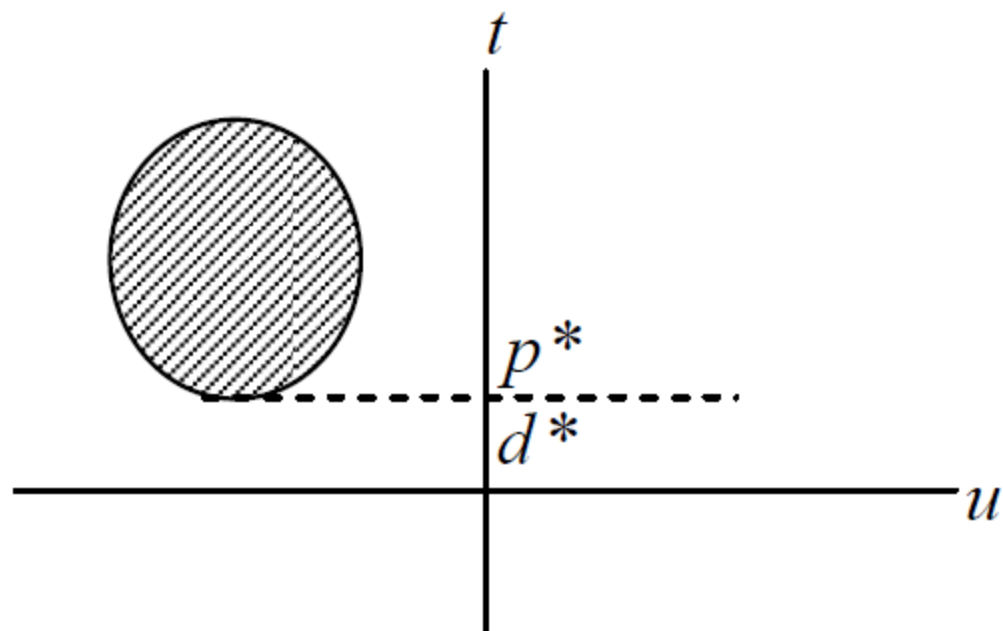
$$p^* = \inf \{t \mid (u, t) \in \mathcal{G}, u \leq 0\}$$



# 凸优化问题



# 凸优化问题



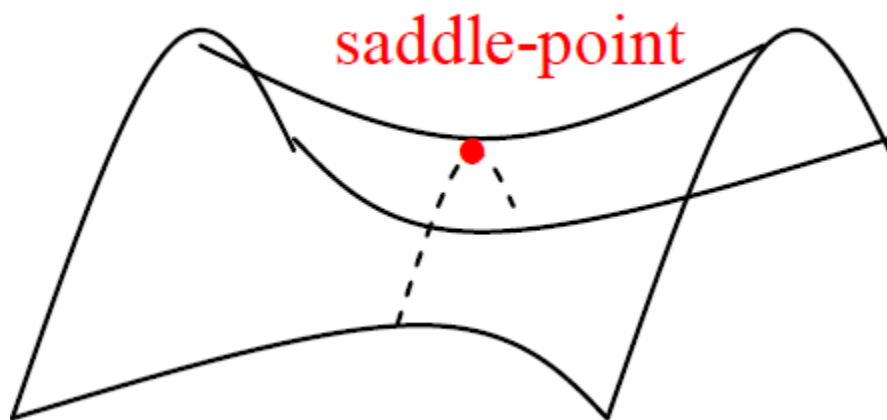
# 凸优化问题

## 鞍点解释

定理：考虑函数  $f(w, z)$ ， $w \in S_w$ ， $z \in S_z$ ，有  $\sup_{z \in S_z} \inf_{w \in S_w} f(w, z) \leq \inf_{w \in S_w} \sup_{z \in S_z} f(w, z)$

定义：若  $\exists (\tilde{w}, \tilde{z}) \in \text{dom} f$ ， $\sup_{z \in S_z} \inf_{w \in S_w} f(\tilde{w}, \tilde{z}) = \inf_{w \in S_w} \sup_{z \in S_z} f(\tilde{w}, \tilde{z})$ ，则  $(\tilde{w}, \tilde{z})$  称为鞍点

性质：若  $(\tilde{w}, \tilde{z})$  是鞍点，则有  $f(\tilde{w}, z) \leq f(\tilde{w}, \tilde{z}) \leq f(w, \tilde{z})$ ， $\forall z \in S_z, \forall w \in S_w$





# 凸优化问题

若  $(\tilde{\mathbf{x}}, \tilde{\lambda})$  为  $L(\mathbf{x}, \lambda)$  的鞍点  $\Leftrightarrow$

对偶问题满足强对偶性, 且  $(\tilde{\mathbf{x}}, \tilde{\lambda})$  为 Primal-Dual 最优解  
也就是

$$(1) \quad \inf_{\mathbf{x}} \sup_{\lambda \succeq 0} L(\mathbf{x}, \lambda) = \sup_{\lambda \succeq 0} \inf_{\mathbf{x}} L(\mathbf{x}, \lambda)$$

$$(2) \quad \begin{cases} \tilde{\mathbf{x}} = \arg \inf_{\mathbf{x}} \sup_{\lambda \succeq 0} L(\mathbf{x}, \lambda) & \text{primal optimal point} \\ \tilde{\lambda} = \arg \sup_{\lambda \succeq 0} \inf_{\mathbf{x}} L(\mathbf{x}, \lambda) & \text{dual optimal point} \end{cases}$$

# 凸优化问题

下面讨论最优解  $\tilde{\mathbf{x}}$  与  $\tilde{\boldsymbol{\lambda}}$  有哪些性质

$$(1) \begin{cases} f_i(\tilde{\mathbf{x}}) \leq 0, i = 1, \dots, m \\ \tilde{\boldsymbol{\lambda}} \succeq \mathbf{0} \end{cases}$$

$$(2) f_o(\tilde{\mathbf{x}}) = g(\tilde{\boldsymbol{\lambda}}) = \inf_{\mathbf{x}} \left\{ f_o(\mathbf{x}) + \sum_{i=1}^m \tilde{\lambda}_i f_i(\mathbf{x}) \right\}$$

$$\leq f_o(\tilde{\mathbf{x}}) + \sum_{i=1}^m \tilde{\lambda}_i f_i(\tilde{\mathbf{x}})$$

$$\leq f_o(\tilde{\mathbf{x}}) \quad \underline{\leq 0}$$

# 凸优化问题

上式若成立，显然两个不等式必须取等号

$$(2-1) \quad \inf_{\mathbf{x}} \left\{ f_o(\mathbf{x}) + \sum_{i=1}^m \tilde{\lambda}_i f_i(\mathbf{x}) \right\} = f_o(\tilde{\mathbf{x}}) + \sum_{i=1}^m \tilde{\lambda}_i f_i(\tilde{\mathbf{x}})$$

$$\inf_{\mathbf{x}} L(\mathbf{x}, \tilde{\lambda}) = L(\tilde{\mathbf{x}}, \tilde{\lambda})$$

$$(2-2) \quad f_o(\tilde{\mathbf{x}}) + \sum_{i=1}^m \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = f_o(\tilde{\mathbf{x}}) = \sup_{\lambda \succeq 0} \left\{ f_o(\tilde{\mathbf{x}}) + \sum_{i=1}^m \lambda_i f_i(\tilde{\mathbf{x}}) \right\}$$

$$L(\tilde{\mathbf{x}}, \tilde{\lambda}) = \sup_{\lambda \succeq 0} L(\tilde{\mathbf{x}}, \lambda)$$

# 凸优化问题

$$(2-3) \quad \sum_{i=1}^m \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = 0$$

$$\begin{cases} f_i(\tilde{\mathbf{x}}) \neq 0 \\ \tilde{\lambda}_i = 0 \end{cases} \quad \text{或} \quad \begin{cases} f_i(\tilde{\mathbf{x}}) = 0 \\ \tilde{\lambda}_i \neq 0 \end{cases}$$

# 凸优化问题

一般优化问题（可以是非凸问题）

$$\begin{cases} \min & f_o(\mathbf{x}) \\ \text{s.t.} & f_i(\mathbf{x}) \leq 0 & i = 1, \dots, m \\ & h_j(\mathbf{x}) = 0 & j = 1, \dots, p \end{cases}$$

$$g(\lambda, \mathbf{v}) = \inf_{\mathbf{x} \in D} L(\mathbf{x}, \lambda, \mathbf{v}) = \inf_{\mathbf{x} \in D} \left( f_o(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^p \nu_j h_j(\mathbf{x}) \right)$$

$$\begin{cases} \max & g(\lambda, \mathbf{v}) \\ \text{s.t.} & \lambda \succeq \mathbf{0} \end{cases}$$

# 凸优化问题

假设原问题与对偶问题的对偶间隙为 0，则该问题存在 Primal-Dual 最优解

设  $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \mathbf{v}^*)$  是该问题的 Primal-Dual 最优解，下面讨论最优解的性质

$$(1) \quad \begin{cases} f_i(\mathbf{x}^*) \leq 0 & i = 1, \dots, m \\ h_j(\mathbf{x}^*) = 0 & j = 1, \dots, p \\ \boldsymbol{\lambda}^* \succeq \mathbf{0} \end{cases} \quad (\text{约束条件/Constraint})$$

# 凸优化问题

$$\begin{aligned}(2) \quad f_o(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \mathbf{v}^*) = \inf_{\mathbf{x}} \left\{ f_o(\mathbf{x}) + \sum_{i=1}^m \lambda_i^* f_i(\mathbf{x}) + \sum_{j=1}^p \nu_j^* h_j(\mathbf{x}) \right\} \\ &\leq f_o(\mathbf{x}^*) + \underbrace{\sum_{i=1}^m \lambda_i^* f_i(\mathbf{x}^*)}_{\leq 0} + \underbrace{\sum_{j=1}^p \nu_j^* h_j(\mathbf{x}^*)}_{= 0} \\ &\leq f_o(\mathbf{x}^*)\end{aligned}$$

# 凸优化问题

上式若成立，显然两个不等式必须取等号

$$(2-1) \quad \inf_{\mathbf{x}} \left\{ f_o(\mathbf{x}) + \sum_{i=1}^m \lambda_i^* f_i(\mathbf{x}) + \sum_{j=1}^p \nu_j^* h_j(\mathbf{x}) \right\} = f_o(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* f_i(\mathbf{x}^*) + \sum_{j=1}^p \nu_j^* h_j(\mathbf{x}^*)$$

假设  $f_o, f_i, h_j$  均可微，则  $\left. \frac{\partial L(\mathbf{x}, \lambda^*, \nu^*)}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}^*} = 0$

即  $\nabla f_o(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{j=1}^p \nu_j^* \nabla h_j(\mathbf{x}^*) = 0$  (稳定性条件/Stationarity)



# 凸优化问题

$$(2-2) \quad f_o(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* f_i(\mathbf{x}^*) + \sum_{j=1}^p \nu_j^* h_j(\mathbf{x}^*) = f_o(\mathbf{x}^*)$$

$$\sum_{i=1}^m \lambda_i^* f_i(\mathbf{x}^*) = 0$$

$$\lambda_i^* f_i(\mathbf{x}^*) = 0, \quad \forall i = 1, \dots, m \quad (\text{互补松弛条件/Complementary Slackness})$$

# 凸优化问题

## KKT 条件 (Karush-Kuhn-Tucker Conditions)

对于可微无对偶间隙优化问题，其最优解的必要条件是

$$f_i(\mathbf{x}^*) \leq 0 \quad i = 1, \dots, m \quad \text{——Primal feasibility}$$

$$h_j(\mathbf{x}^*) = 0 \quad j = 1, \dots, p$$

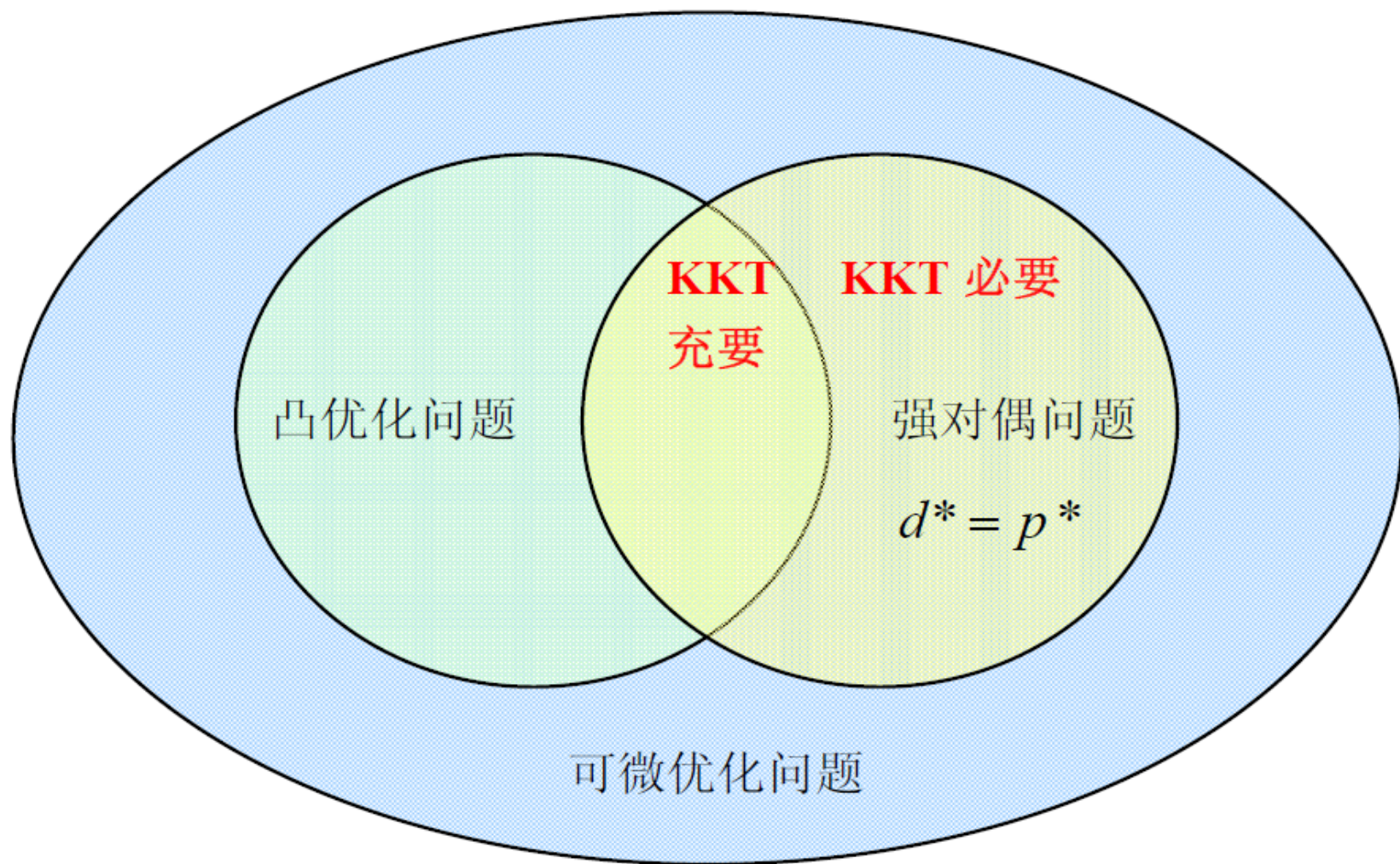
$$\lambda_i^* \geq 0 \quad i = 1, \dots, m \quad \text{——Dual feasibility}$$

$$\lambda_i^* f_i(\mathbf{x}^*) = 0 \quad i = 1, \dots, m \quad \text{——Complementary Slackness}$$

$$\nabla f_o(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{j=1}^p \nu_j^* \nabla h_j(\mathbf{x}^*) = 0 \quad \text{——Stationarity}$$

# 凸优化问题

对于可微无对偶间隙凸优化问题，KKT 条件等价于 Primal-Dual 最优解



# 凸函数

□ 优化算法

# 凸优化问题

优化算法都是迭代的

基本结构:  $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$

↓      ↘  
步长    方向

步长选择:

确定性步长

固定步长  $\alpha^k = \alpha$

递减步长  $\alpha^k = \frac{1}{k+1}, \frac{1}{\sqrt{k+1}}$

自适应步长  $\alpha^k = \arg \min f_o(\mathbf{x}^k + \alpha \mathbf{d}^k), 0 \leq \alpha \leq \alpha_{\max}$

↓  
选定一个方向 (局部最优方向), 使目标函数最小的步长



# THE END