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：

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

超几何体：

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

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

$\forall x_1, x_2 \in C, \forall \theta \in R$，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：

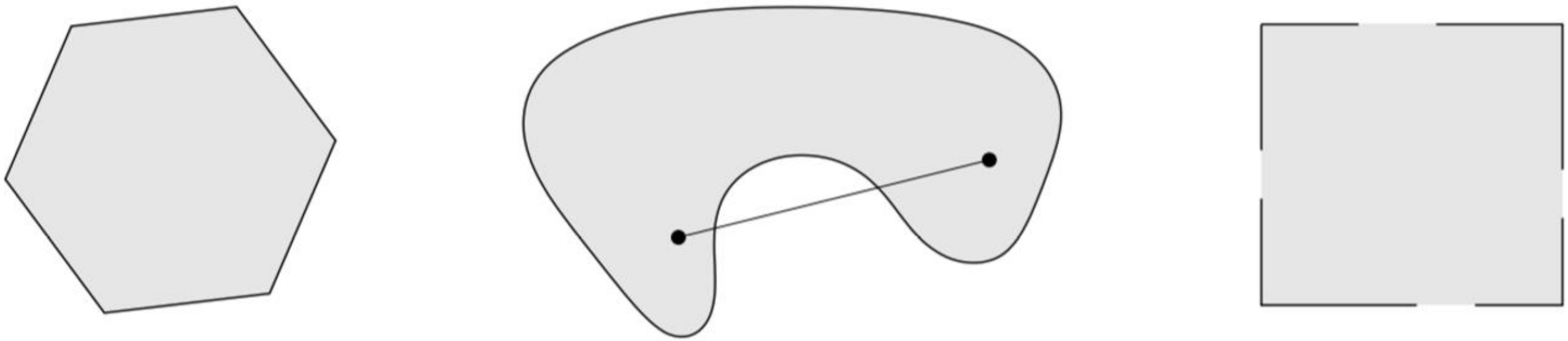
$\forall x_1, x_2 \in C, \forall \theta \in [0,1]$，有 $x = \theta \cdot x_1 + (1-\theta) \cdot x_2 \in C$

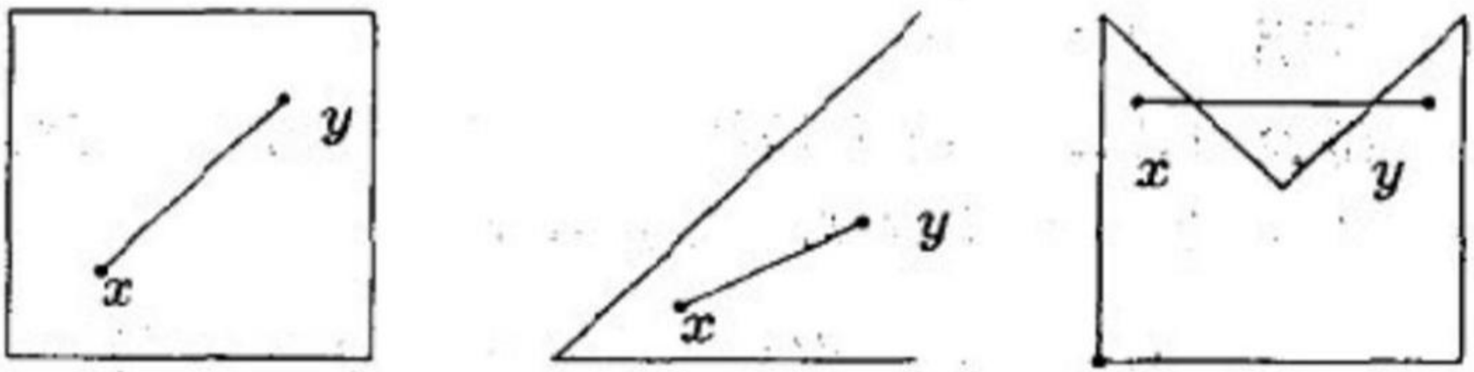
k个点的版本：

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

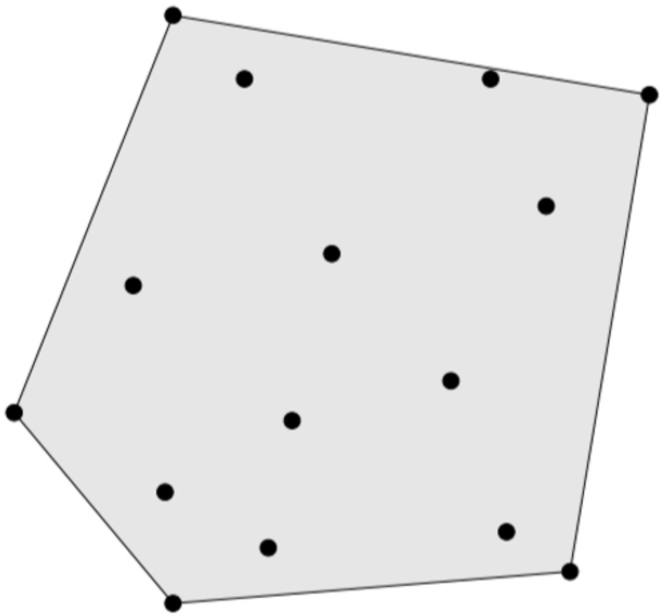
凸包

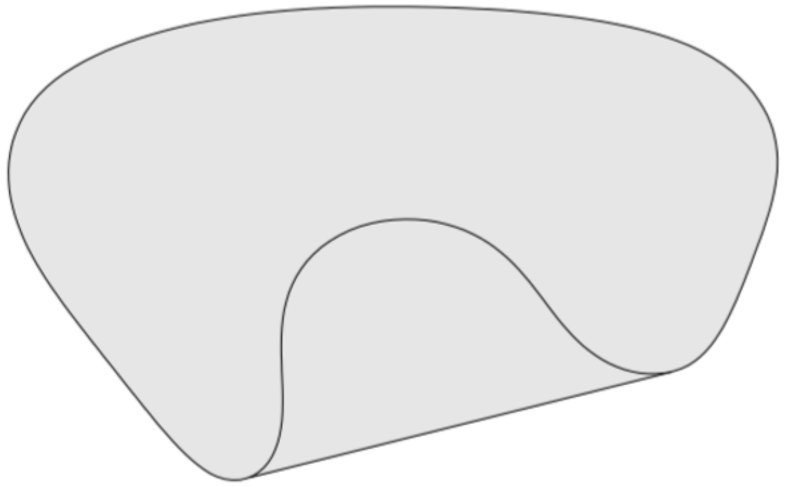
集合C的所有点的凸组合形成的集合，叫做集合C的凸包

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

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

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

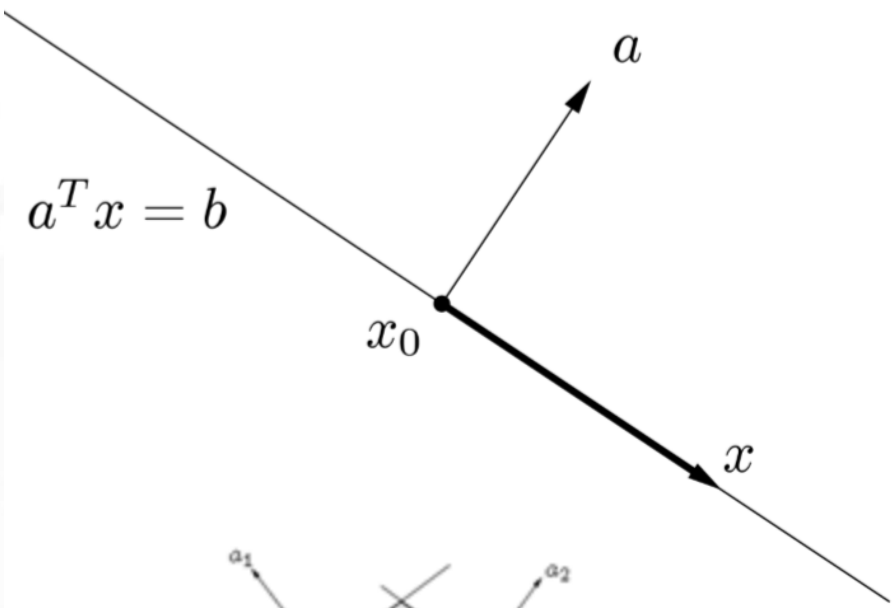
半平面halfspace

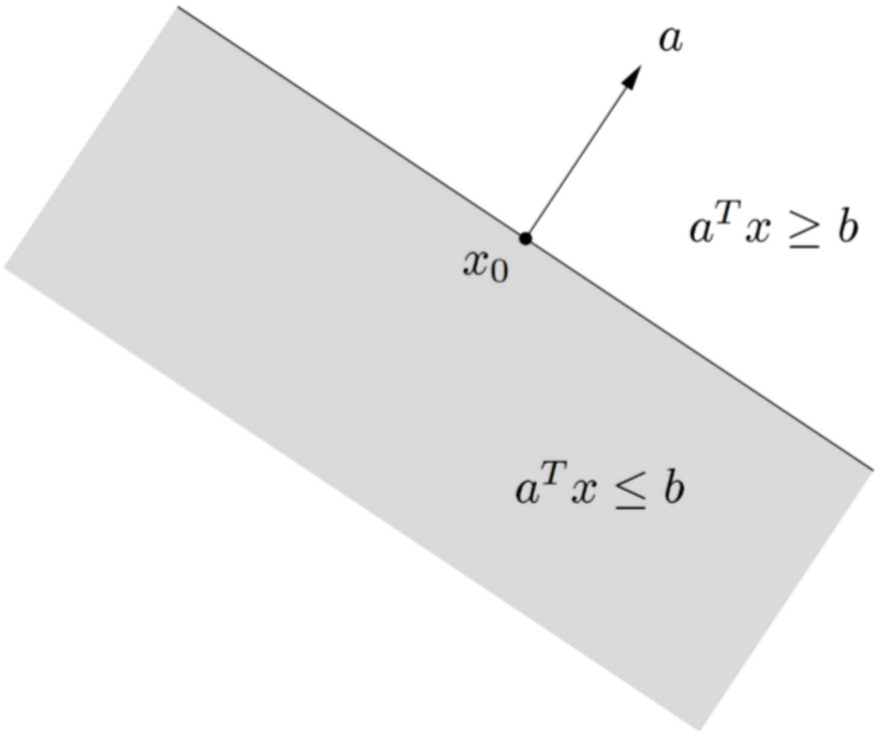
$$\{x \mid a^T x \le b\} \qquad \{x \mid a^T x \ge b\}$$

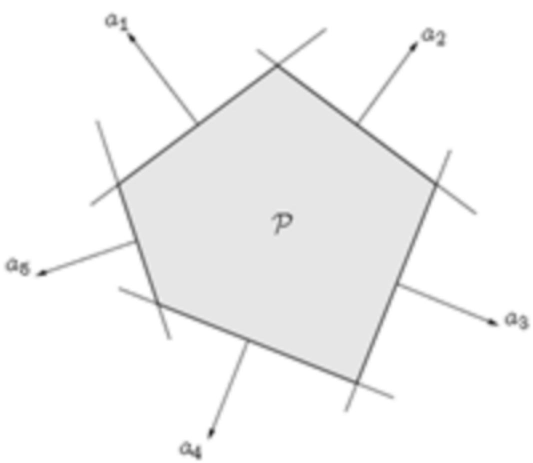
多面体Polyhedron

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

超平面和半空间







欧式球和椭球

欧式球

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \le r\}$$
$$= \{x \mid (x - x_c)^T (x - x_c) \le r^2\}$$

椭球

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

P为对称正定矩阵

范数球和范数锥

范数：

$$\|x\| \geq 0, \|x\| = 0 \;\; iff \;\; x = 0;$$
$$\|tx\| = |t|\|x\|, t \in R;$$
$$\|x + y\| \leq \|x\| + \|y\|$$

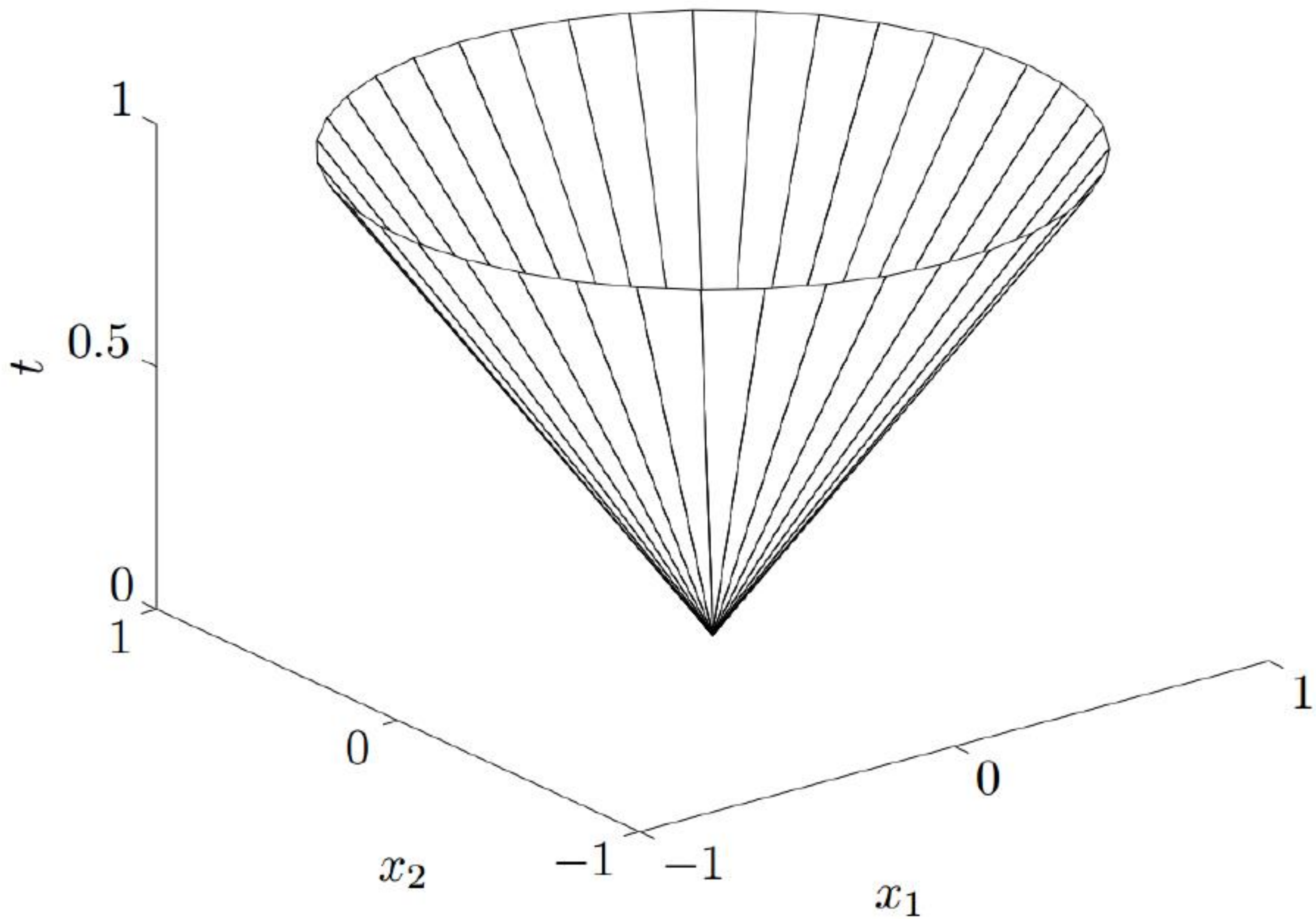
范数球：

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

范数锥：

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

例子：二阶锥



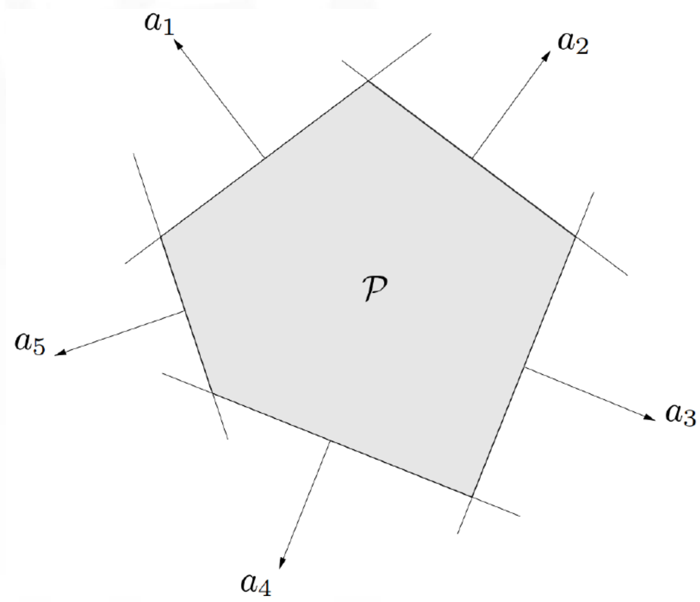
多面体

多面体有限个半空间和超平面的交集：

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

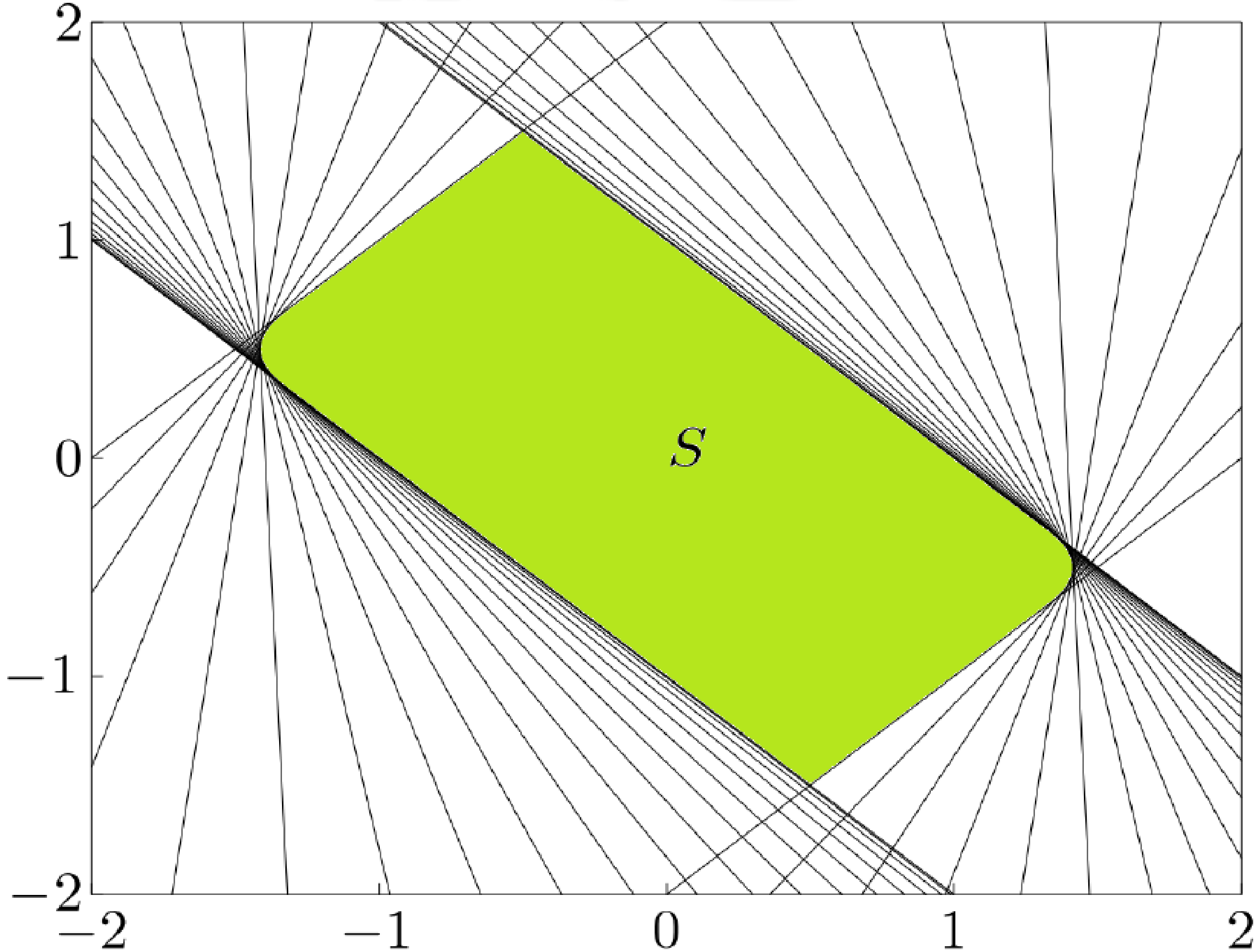
仿射集（如超平面、直线）、射线、线段、半空间都是多面体

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，$f : \mathbf{R}^n \to R^m$ ，$f(S) = \{f(x) | x \in S\}$

- 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

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

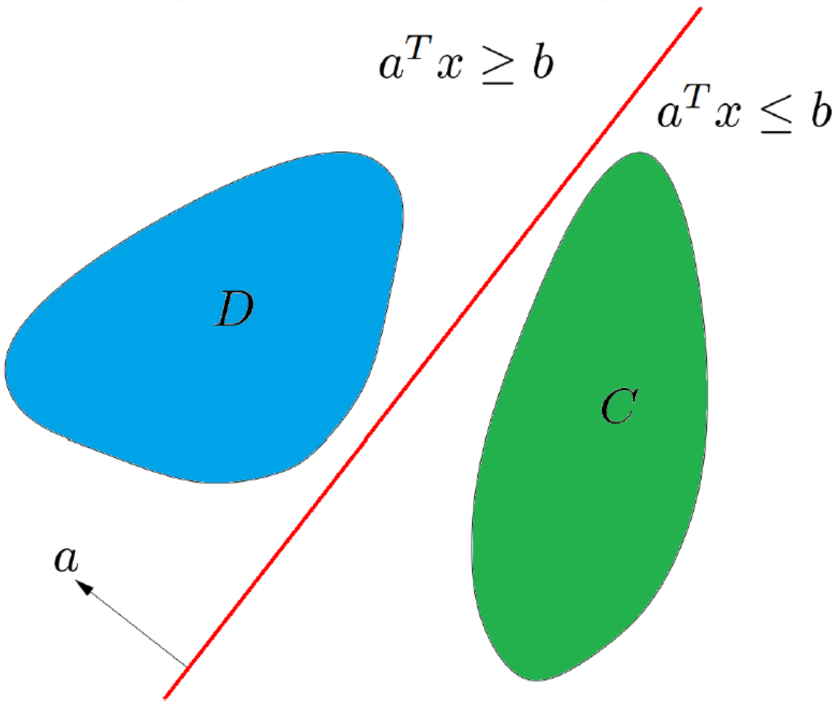
两个凸集的部分和为凸集

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

分割超平面

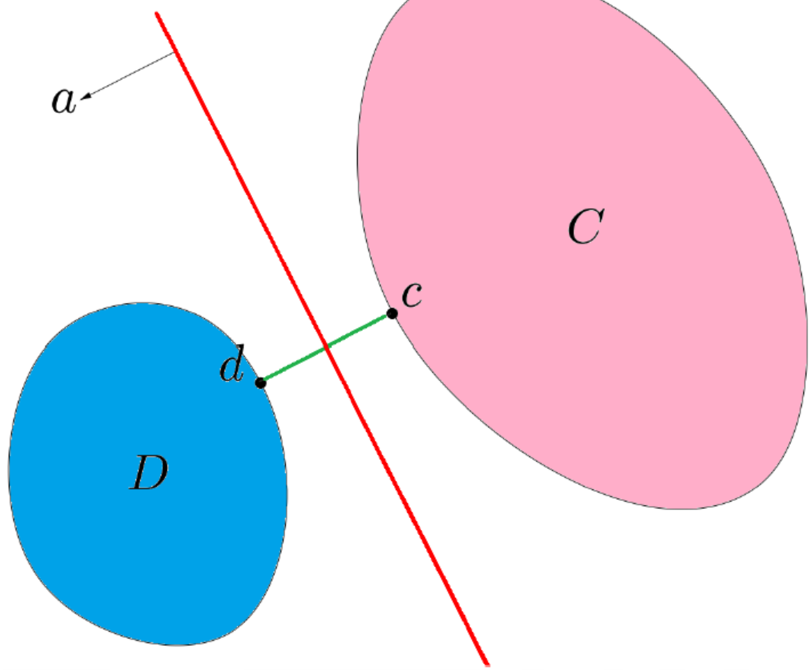
设C和D为两不相交的凸集，则存在超平面P，P可以将C和D分离

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



分割超平面的构造

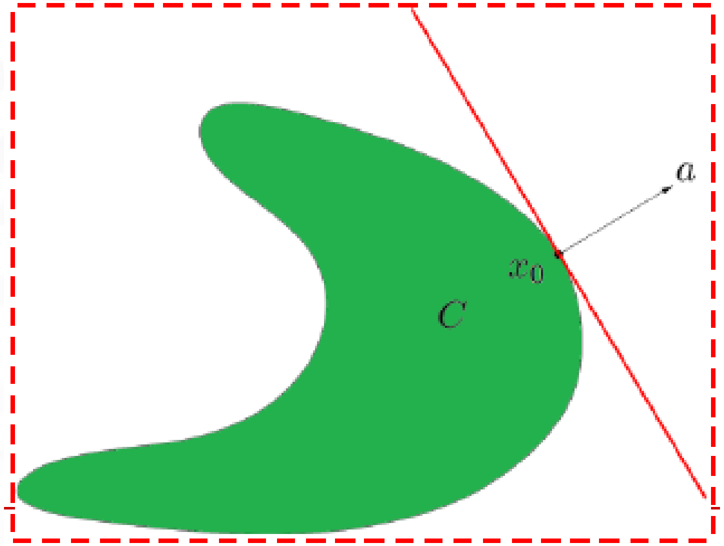
两个集合的距离定义为两个集合间元素的最短距离



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，x0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。

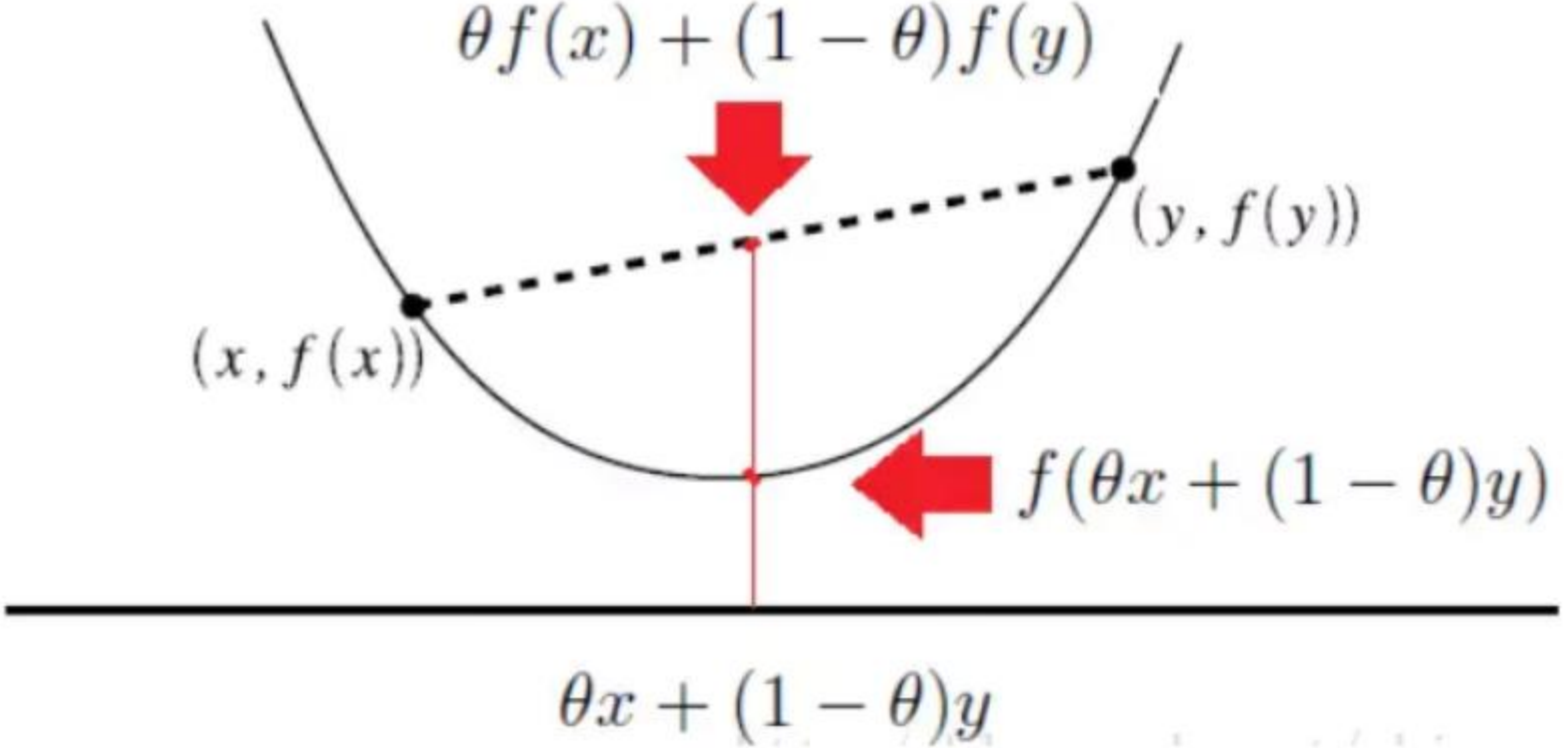
凸函数

- 凸函数基本概念

凸函数定义1

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

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

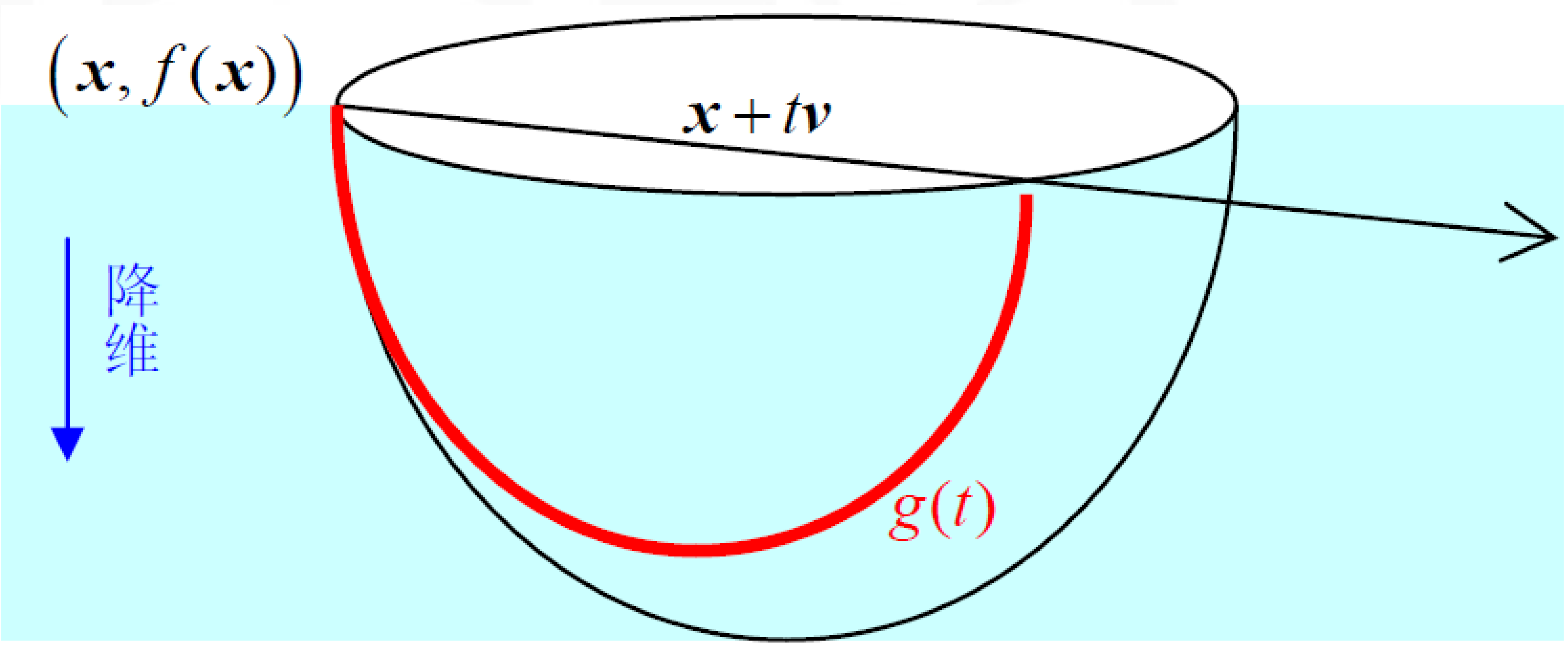


凸函数定义2

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

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

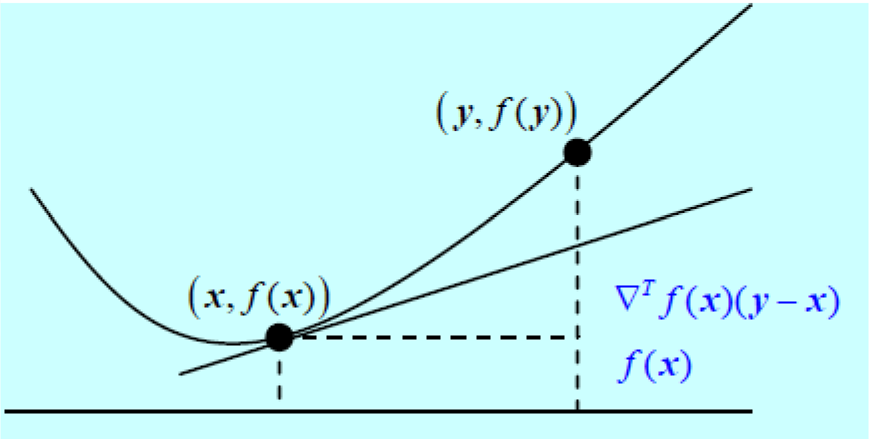
上是凸的



凸函数定义3

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

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



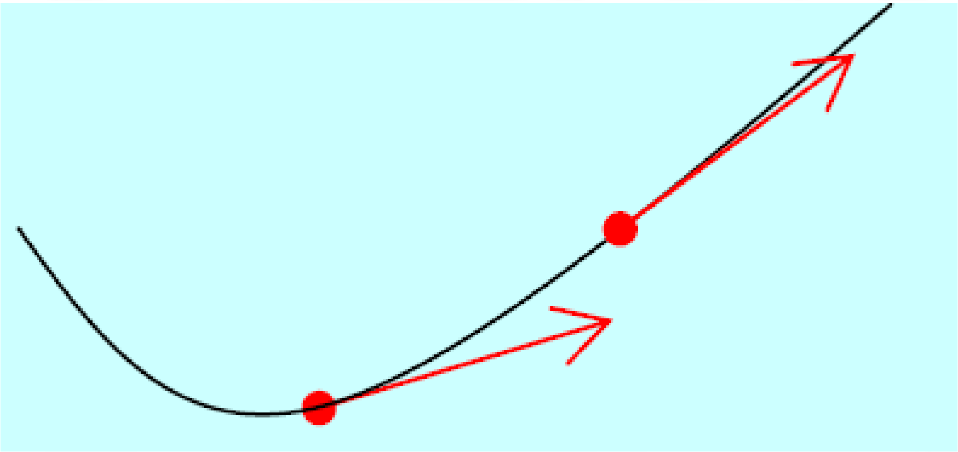
性质：若f为凸函数，$\exists x_0 \in \mathbf{dom} f$ 使 $\nabla f(x_0) = 0$ ，对

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

则 $f(x_0)$ 是 f 的最小值

凸函数定义4

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



凸函数

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

$$\nabla^2 f(\boldsymbol{x})=\boldsymbol{P}，\quad \begin{array}{l}\text{if } \boldsymbol{P}\succeq 0\text{, convex function}\\ \text{if } \boldsymbol{P}\succ 0\text{, strictly convex function}\\ \text{if } \boldsymbol{P}\preceq 0\text{, concave function}\end{array}$$

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

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

例：指数函数

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

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

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

故指数函数是凸函数

例：幂函数

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

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

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

凸函数

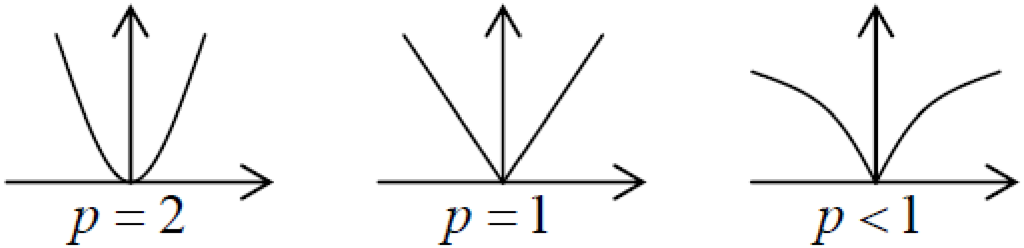
例：绝对值的幂函数

$f(x)=|x|^p, x\in \mathrm{R}, p\geq 0$(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

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

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

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

故对数函数是凹函数

例：负熵

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

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

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

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：$f(\boldsymbol{x})=\log(e^{x_1}+\cdots+e^{n}) \qquad \boldsymbol{x}\in \mathbf{R}^n$

保凸操作

保持凸函数凸性的操作

先求值，再做线性变换
变换的是值域

- 非负加权和（**Non-negative weighted sum**）

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

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

先做线性变换，再求值
变换的是定义域

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

若 $f(\boldsymbol{x}):\mathbf{R}^n \to \mathrm{R}$ 为凸函数，则 $g(\boldsymbol{x}) \triangleq f(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}), \quad \boldsymbol{A} \in \mathbf{R}^{n\times n}, \boldsymbol{b} \in \mathbf{R}^n, \boldsymbol{A}\boldsymbol{x}+\boldsymbol{b} \in \mathrm{dom}f$ 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 $f(\boldsymbol{x}) \triangleq \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}, \quad \mathrm{dom}f = \mathrm{dom}f_1 \cap \mathrm{dom}f_2$ 为凸函数。

保凸操作

➢ 函数的复合（**Composition of functions**）

$$\begin{array}{l} h:\mathbf{R}^k \to \mathrm{R} \\ g:\mathbf{R}^n \to \mathbf{R}^k \end{array},\quad f = h \circ g : \mathbf{R}^n \to \mathrm{R},\quad \mathrm{dom}f = \{\boldsymbol{x} \in \mathrm{dom}g \mid g(\boldsymbol{x}) \in \mathrm{dom}h\}$$

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

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

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

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

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

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

$\boldsymbol{x}$： 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）

$$\boldsymbol{D} \triangleq \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{j=0}^{p} \mathrm{dom} h_j$$

可行解集（feasible set）

$$\boldsymbol{X} = \left\{ \boldsymbol{x} \in \boldsymbol{D} \,\middle|\, \begin{array}{ll} f_i(\boldsymbol{x}) \le 0 & i = 1, \cdots, m \\ h_j(\boldsymbol{x}) = 0 & j = 1, \cdots, p \end{array} \right\}$$

最优值（optimal value）

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

最优解集（optimal set）

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

ε 次优解（ε -suboptimal set）

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

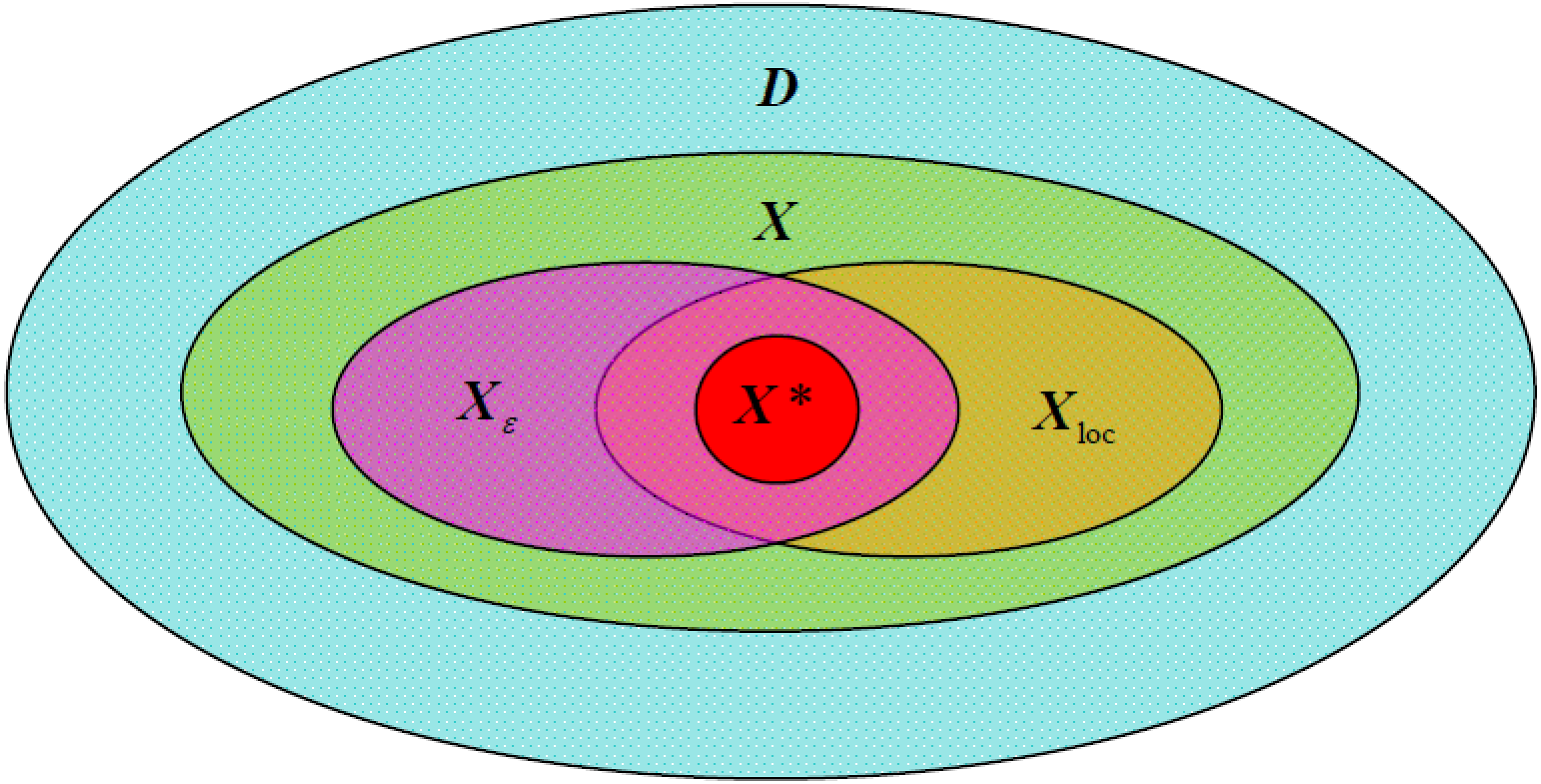
局部最优值（local optimal value）

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

局部最优解集（local optimal set）

$$\boldsymbol{X}_{\mathrm{loc}} = \{ \boldsymbol{x}_{\mathrm{loc}} \in \boldsymbol{X} \mid f_o(\boldsymbol{x}_{\mathrm{loc}}) = p_{\mathrm{loc}} \}$$

凸优化问题



凸优化问题

凸优化问题（**Convex Optimization Problems**）

$$
\begin{aligned}
\min \quad & f_o(\boldsymbol{x}) \\
\text{s.t.} \quad & f_i(\boldsymbol{x}) \le 0 \qquad i = 1, \cdots, m \\
& \boldsymbol{A}_j^T \boldsymbol{x} = \boldsymbol{b}_j \qquad j = 1, \cdots, p
\end{aligned}
$$

1）目标函数 $f_o(\boldsymbol{x})$ 为凸函数

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

3）等式约束 $\boldsymbol{A}_1^T \boldsymbol{x} = \boldsymbol{b}_1, \cdots, \boldsymbol{A}_p^T \boldsymbol{x} = \boldsymbol{b}_p$ 为仿射函数（Note：仿射函数 $\boldsymbol{A}_j^T \boldsymbol{x} = \boldsymbol{b}_j$ 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

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 $f_o(\boldsymbol{y}) \geq f_o(\boldsymbol{x}) + \nabla^T f_o(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x})$， $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom} f$

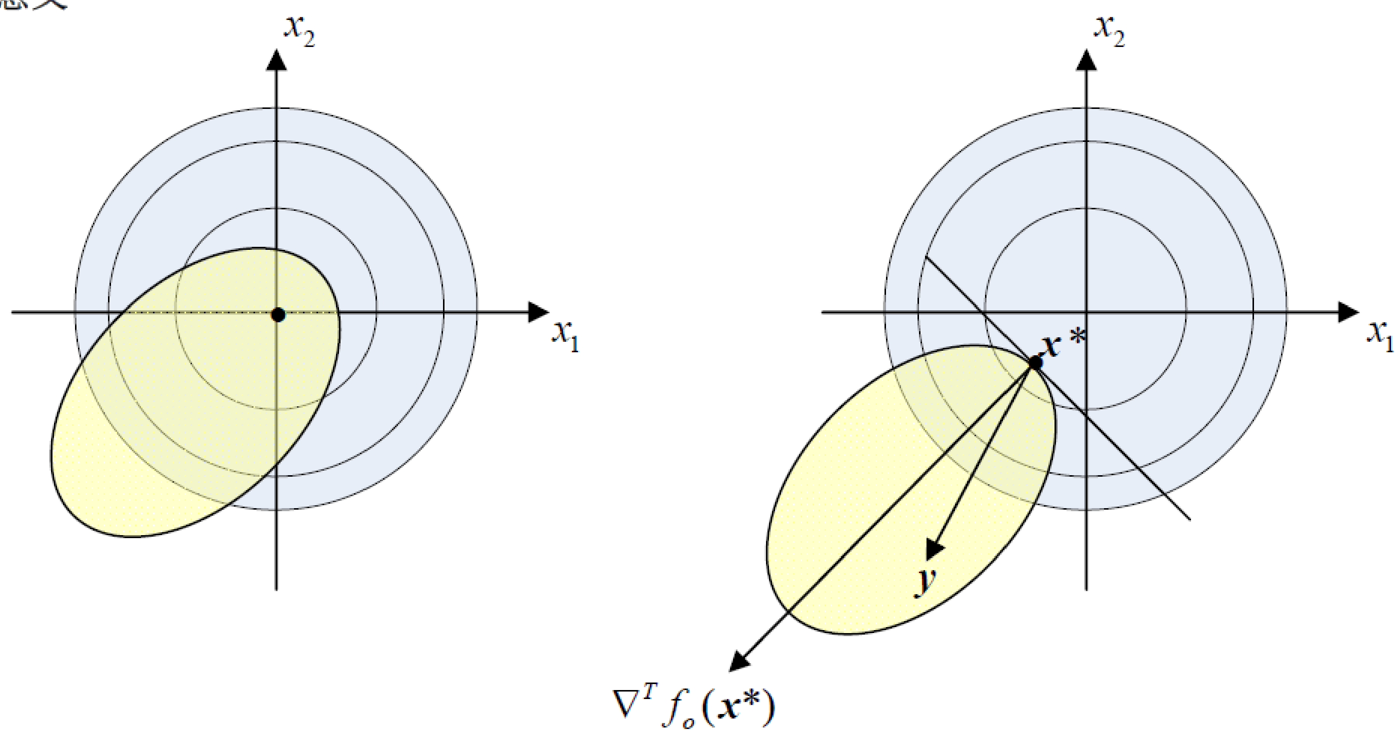
对无约束问题 $\begin{cases} \min & f_o(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in \mathrm{dom} f_o \end{cases}$ 其最优解 $\boldsymbol{x}^*$ 满足 $\nabla^T f_o(\boldsymbol{x}^*)(\boldsymbol{y}-\boldsymbol{x}^*) \geq 0,\ \forall \boldsymbol{y} \in \mathrm{dom} f_o$

同理

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

凸优化问题

几何意义



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

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

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

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

凸优化问题

线性规划问题（Linear Programming/LP）

$$\begin{cases} \min & c^T x + d \\ \text{s.t.} & Gx \preceq h \\ & Ax = b \end{cases}$$

$\left.\begin{matrix} Gx \preceq h \\ Ax = b \end{matrix}\right\} \longrightarrow P$：多面体

二次规划（**Quadratic Programming / QP**）

$$\begin{cases} \min & \frac{1}{2} x^T P x + q^T x + r \\ \text{s.t.} & Gx \preceq h \\ & Ax = b \end{cases}$$

凸优化问题

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

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

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

凸优化问题

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

1）选择合适的 $\boldsymbol{x}^*$ 使误差项最小，即

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

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

——QP

2-1）一范数规范化优化问题

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

2-2）一范数约束优化问题

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

凸优化问题

3-1）二范数规范化优化问题

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

3-2）二范数约束优化问题

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

凸优化问题

- 对偶性

凸优化问题

标准最优化问题

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

域（Domain）：$\boldsymbol{D} \triangleq \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{j=0}^{p} \mathrm{dom} h_j$

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

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

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

拉格朗日函数（Lagrangian Function）

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

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

$\boldsymbol{x}$： 原变量（primal variable）

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

λ_i, ν_i： 拉格朗日乘子（Lagrange Multiplier）

凸优化问题

对偶函数（Lagrange Dual Function）

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

（1）对偶函数一定为凹函数

（2）$\forall\boldsymbol{\lambda}\succeq\mathbf{0},\forall\boldsymbol{v}$， $g(\boldsymbol{\lambda},\boldsymbol{v})\le p^*$

例：

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

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:） $d^* = \sup\{g(\lambda, v) \mid \lambda, v \in \mathbf{dom}\ g \& \lambda \geq 0)\}$

原问题（Primal Problem）

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

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

$d^* \leq p^*$ 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： $\boldsymbol{D}$ 的 Relative Interior

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

Slater's Condition

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

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

凸优化问题

Refined Slater's Condition

$$\exists x \in \operatorname{relint} D$$

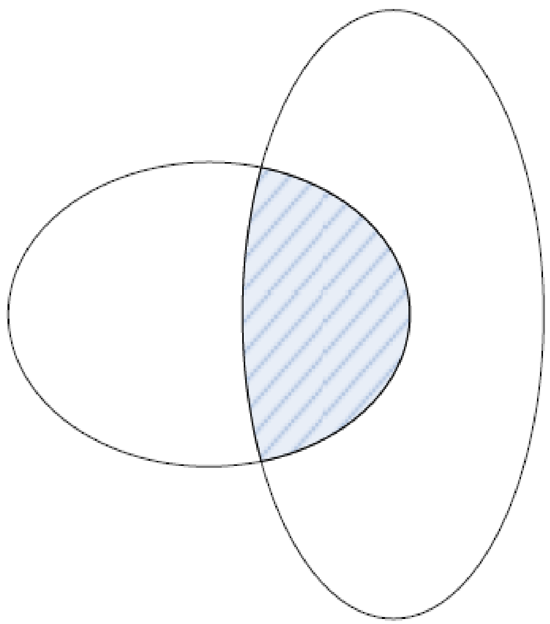
$$\begin{cases} f_i(x) \le 0 & i = 1, \cdots, k \quad \text{when } f_i \text{ is affine} \\ f_i(x) < 0 & i = k+1, \cdots, m \\ A_j^T x = b_j & j = 1, \cdots, p \end{cases}$$

定理：

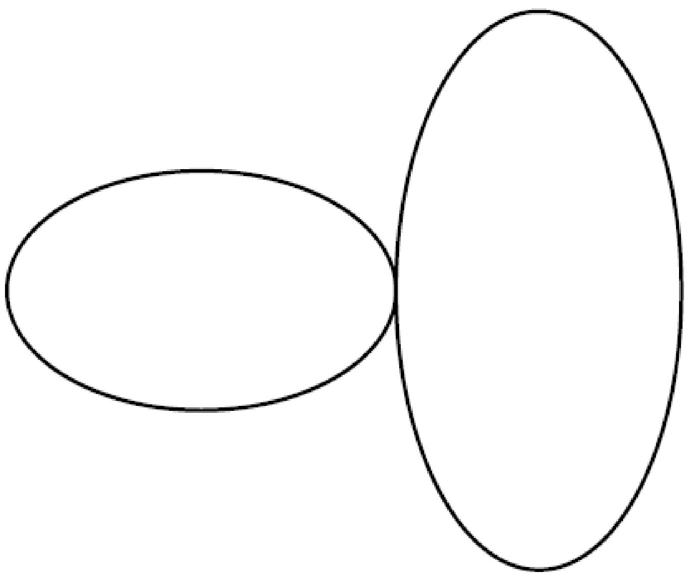
①对于**非凸问题**，其对偶问题通常没有强对偶性

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：

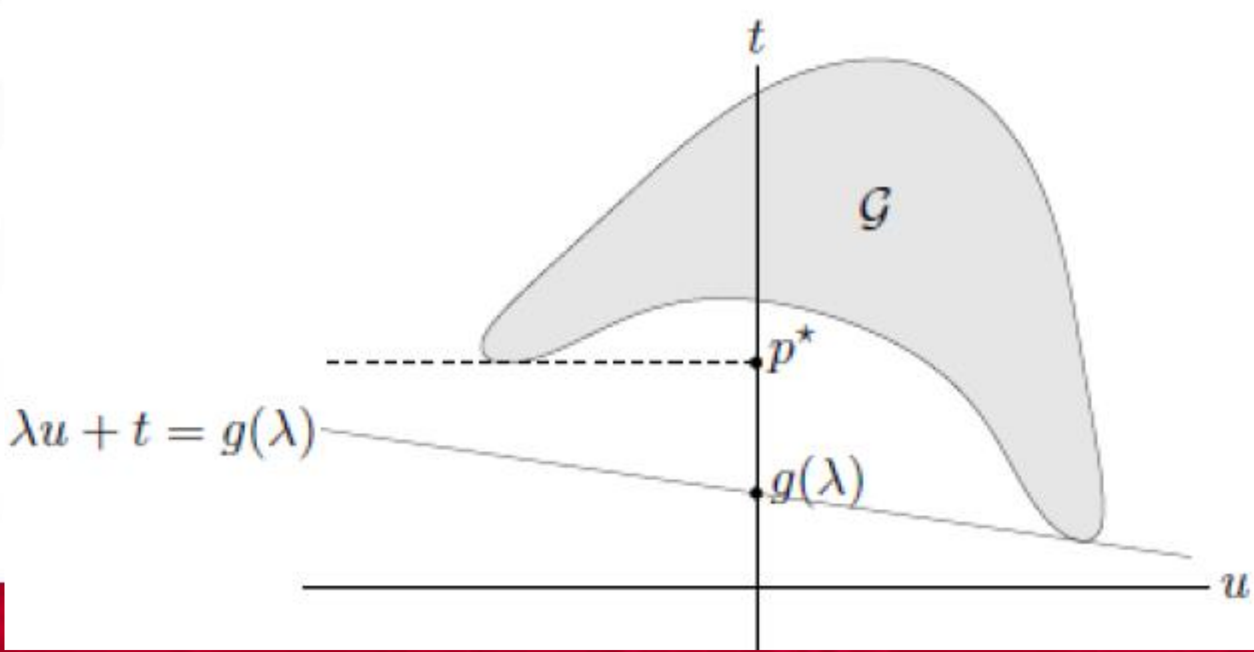
考虑如下单约束单目标优化问题

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

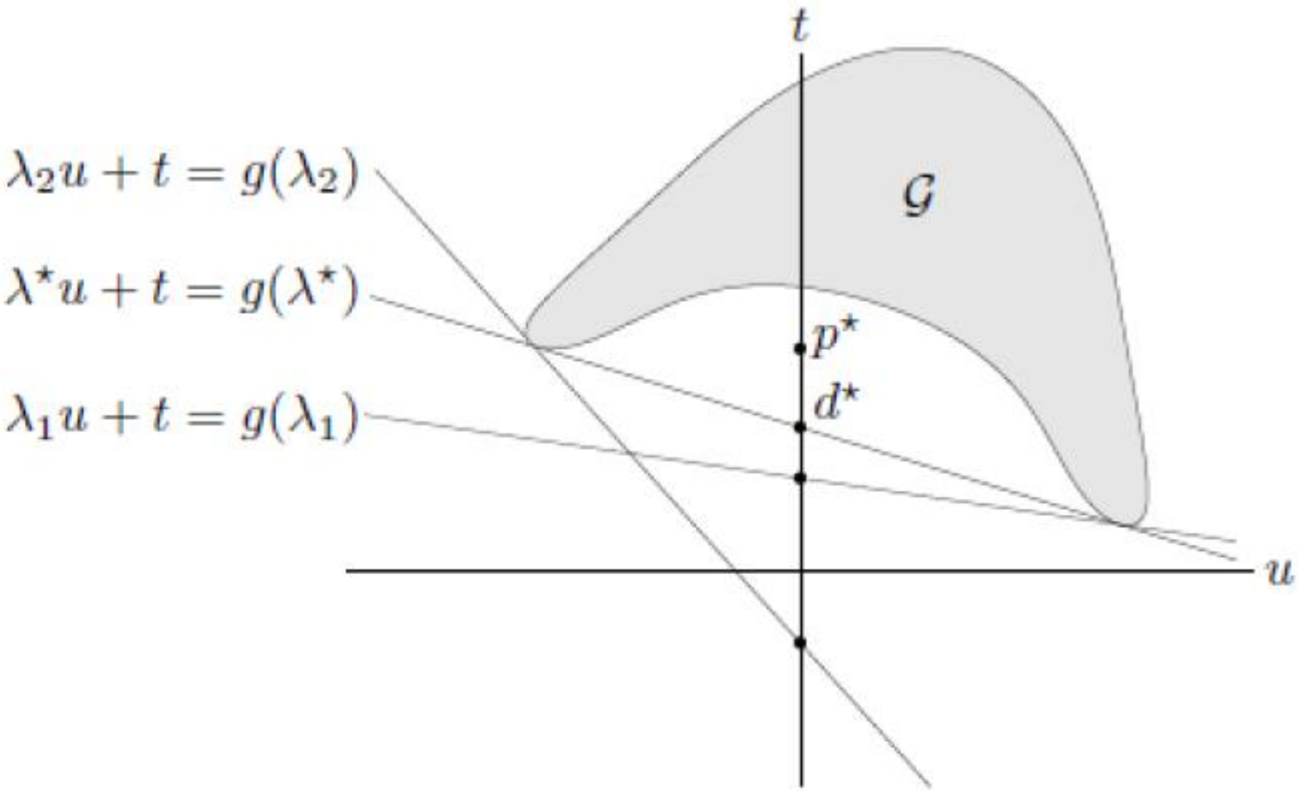
设 $\mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in D\}$

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

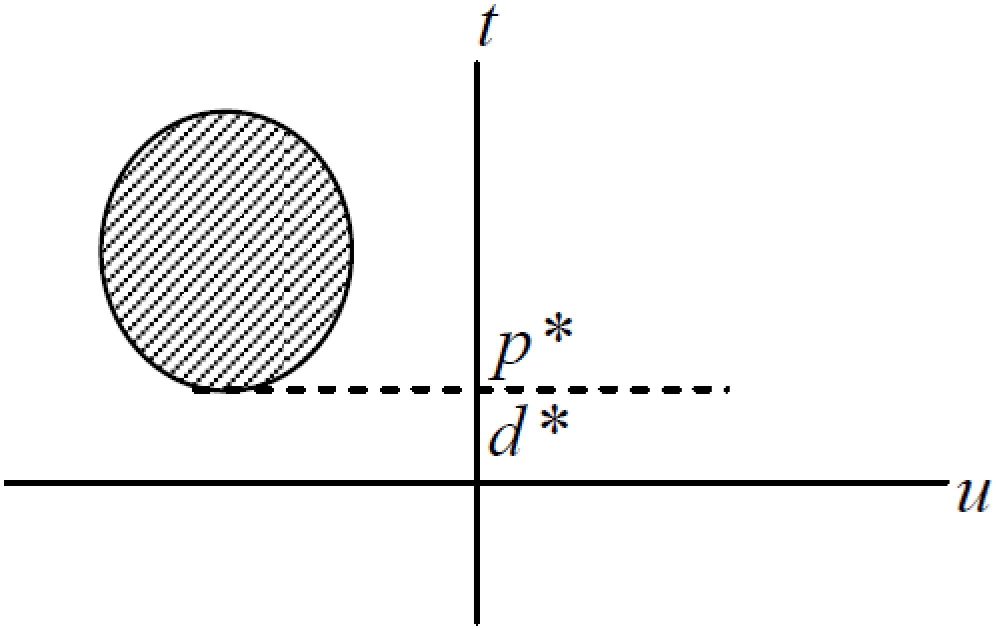
$$p^* = \inf\{t \mid (u,t) \in \mathcal{G}, u \le 0\}$$



凸优化问题


t
$\lambda_2 u + t = g(\lambda_2)$
$\mathcal{G}$
$\lambda^\star u + t = g(\lambda^\star)$
$p^\star$
$d^\star$
$\lambda_1 u + t = g(\lambda_1)$
u

凸优化问题



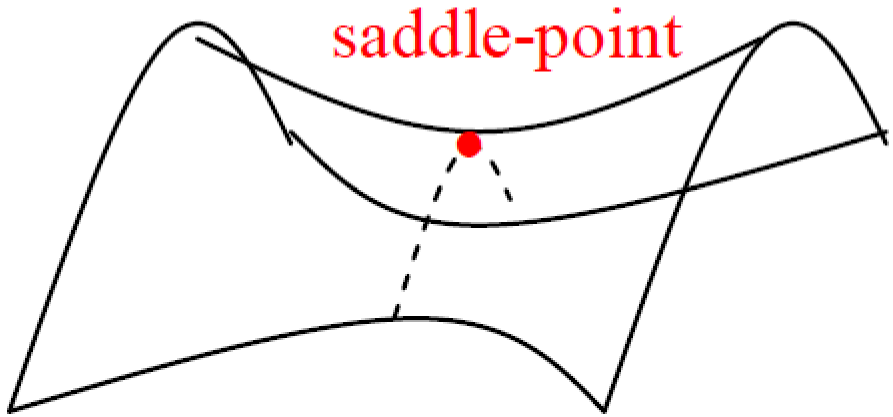
凸优化问题

鞍点解释

定理：考虑函数 $f(w,z)$， $w \in S_w$， $z \in S_z$，有 $\sup\limits_{z\in S_z}\inf\limits_{w\in S_w} f(w,z) \le \inf\limits_{w\in S_w}\sup\limits_{z\in S_z} f(w,z)$

定义：若 $\exists(\tilde{w},\tilde{z}) \in \mathrm{dom} f$， $\sup\limits_{z\in S_z}\inf\limits_{w\in S_w} f(\tilde{w},\tilde{z}) = \inf\limits_{w\in S_w}\sup\limits_{z\in S_z} f(\tilde{w},\tilde{z})$，则 $(\tilde{w},\tilde{z})$ 称为鞍点

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



凸优化问题

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

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

也就是

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

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

凸优化问题

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

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

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

$$\le f_o(\tilde{\boldsymbol{x}}) + \underbrace{\sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{\boldsymbol{x}})}_{\le 0}$$

$$\le f_o(\tilde{\boldsymbol{x}})$$

凸优化问题

上式若成立，显然两个不等式必须取等号

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

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

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

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

凸优化问题

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

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

凸优化问题

一般优化问题（可以是非凸问题）

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

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

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

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

$$
(1)\quad \begin{cases} f_i(\boldsymbol{x}^*) \le 0 & i = 1, \cdots, m \\ h_j(\boldsymbol{x}^*) = 0 & j = 1, \cdots, p \\ \boldsymbol{\lambda}^* \succeq \boldsymbol{0} & \end{cases}
$$

（约束条件/Constraint）

凸优化问题

（2） $f_o(\boldsymbol{x}^*) = g(\boldsymbol{\lambda}^*, \boldsymbol{v}^*) = \inf_{\boldsymbol{x}} \left\{ f_o(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{j=1}^{p} v_j^* h_j(\boldsymbol{x}) \right\}$

$$\le f_o(\boldsymbol{x}^*) + \underbrace{\sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}^*)}_{\le 0} + \underbrace{\sum_{j=1}^{p} v_j^* h_j(\boldsymbol{x}^*)}_{=0}$$

$$\le f_o(\boldsymbol{x}^*)$$

凸优化问题

上式若成立，显然两个不等式必须取等号

$$(2\text{—}1)\quad \inf_{\boldsymbol{x}}\left\{f_o(\boldsymbol{x})+\sum_{i=1}^{m}\lambda_i^* f_i(\boldsymbol{x})+\sum_{j=1}^{p}v_j^* h_j(\boldsymbol{x})\right\}=f_o(\boldsymbol{x}^*)+\sum_{i=1}^{m}\lambda_i^* f_i(\boldsymbol{x}^*)+\sum_{j=1}^{p}v_j^* h_j(\boldsymbol{x}^*)$$

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

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

凸优化问题

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

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

$\lambda_i^* f_i(\boldsymbol{x}^*) = 0, \quad \forall i = 1, \cdots, m$ （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

$$f_i(\boldsymbol{x}^*) \le 0 \qquad i = 1, \cdots, m$$

$$h_j(\boldsymbol{x}^*) = 0 \qquad j = 1, \cdots, p$$

——Primal feasibility

$$\lambda_i^* \ge 0 \qquad i = 1, \cdots, m$$

——Dual feasibility

$$\lambda_i^* f_i(\boldsymbol{x}^*) = 0 \qquad i = 1, \cdots, m$$

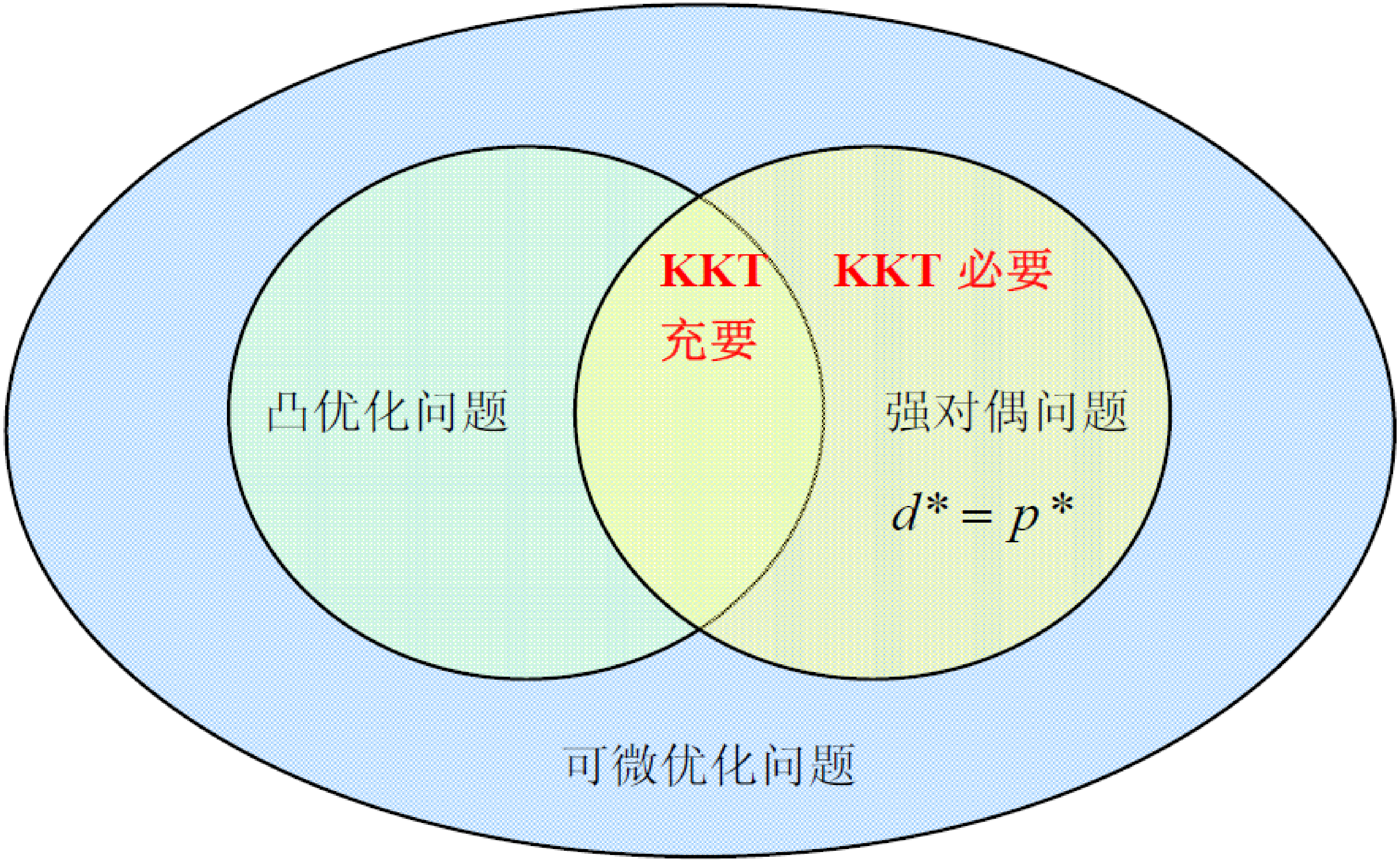
——Complementary Slackness

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

——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： $\boldsymbol{x}^{k+1} = \boldsymbol{x}^{k} + \underbrace{\alpha^{k}}_{\text{步长}}\,\underbrace{\boldsymbol{d}^{k}}_{\text{方向}}$

步长选择：

- 确定性步长
  - 固定步长 $\alpha^{k} = \alpha$
  - 递减步长 $\alpha^{k} = \dfrac{1}{k+1}, \dfrac{1}{\sqrt{k+1}}$
- 自适应步长 $\alpha^{k} = \underbrace{\arg\min}_{\text{选定一个方向（局部最优方向），使目标函数最小的步长}} f_{o}(\boldsymbol{x}^{k} + \alpha \boldsymbol{d}^{k}), 0 \le \alpha \le \alpha_{\max}$

THE END