最优化理论与方法

讲课人：吕茂斌
模式识别与智能系统研究所
北京理工大学自动化学院
lumaobin@bit.edu.cn

凸分析

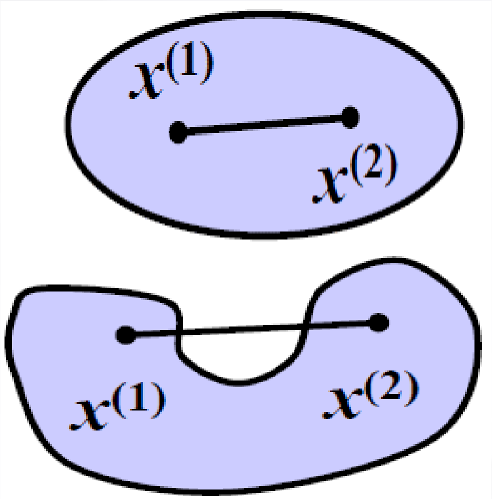
基本定义

◆ **定义** 设S为n维欧氏空间$\mathbf{R}^n$中一个集合。若对S中任意两点，联结它们的线段仍属于S；换言之，对S中任意两点$x^{(1)},x^{(2)}$及每个实数$\lambda\in[0,1]$, 都有

$$\lambda\mathbf{x}^{(1)}+(1-\lambda)\mathbf{x}^{(2)}\in S$$

则称S为**凸集**.

$x^{(1)},x^{(2)}$的凸组合



例子: 盒子、球、椭球、多面体

问题：

1 两（多）个凸集的并集是凸的吗？

2 两（多）个凸集的交集是凸集吗？

3 空集是凸集吗？

例 1: 验证集合$H=\left\{\boldsymbol{x}|\boldsymbol{p}^T\boldsymbol{x}=\alpha\right\}$为凸集，其中$\boldsymbol{p}$为$n$维列向量，$\alpha$为实数。

解 由于对任意两点$x^{(1)},x^{(2)}\in H$及每个实数$\lambda\in[0,1]$都有

$$p^{\mathrm{T}}\left[\lambda x^{(1)}+(1-\lambda)x^{(2)}\right]=\alpha$$

因此 $$\lambda x^{(1)}+(1-\lambda)x^{(2)}\in H$$

根据凸集定义知H为凸集.

例中定义的集合H称为$\mathbf{R}^n$中的**超平面**，故**超平面为凸集**.

例 2: 验证集合$H^- = \left\{\boldsymbol{x} | \boldsymbol{p}^T\boldsymbol{x} \leq \alpha\right\}$为凸集，其中$\boldsymbol{p}$为$\boldsymbol{n}$维列向量，$\alpha$为实数。

解 由于对任意两点$x^{(1)}, x^{(2)} \in H^-$及每个实数$\lambda \in [0,1]$都有

$$p^{\mathrm{T}}\left[\lambda x^{(1)} + (1-\lambda)x^{(2)}\right] = \lambda p^{\mathrm{T}} x^{(1)} + (1-\lambda) p^{\mathrm{T}} x^{(2)} \leqslant \alpha$$

因此
$$\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)} \in H^-$$

根据凸集定义知H^-为凸集.

例中定义的集合$H^- = \left\{\boldsymbol{x} | \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x} \leq \alpha\right\}$称为<u>半空间</u>，故<u>半空间为凸集</u>.

例 3: 验证集合$L=\{x|x=x^{(0)}+\lambda \boldsymbol{d},\lambda\geqslant 0\}$为凸集，其中$\boldsymbol{d}$是给定的非零向量, $\boldsymbol{x}^{(0)}$是定点.

解 由于对任意两点$\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}\in L$及每个实数$\lambda\in[0,1]$都有

$$\boldsymbol{x}^{(1)}=\boldsymbol{x}^{(0)}+\lambda_1\boldsymbol{d},\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(0)}+\lambda_2\boldsymbol{d}$$

λ_1和λ_2是两个非负数，以及

$$\begin{aligned}\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)}&=\lambda\left(\boldsymbol{x}^{(0)}+\lambda_1\boldsymbol{d}\right)+(1-\lambda)\left(\boldsymbol{x}^{(0)}+\lambda_2\boldsymbol{d}\right)\\&=\boldsymbol{x}^{(0)}+[\lambda\lambda_1+(1-\lambda)\lambda_2]\boldsymbol{d}\end{aligned}$$

由于$\lambda_1+(1-\lambda)\lambda_2\geqslant 0$, 因此有$\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)}\in L$，根据凸集定义知$L$为凸集.

集合$L=\{x|x=x^{(\delta)}+\lambda\boldsymbol{d},\lambda\geqslant 0\}$称为**射线**，$\boldsymbol{x}^{(0)}$为射线的顶点．故**射线为凸集**.

设S_1和S_2为$\mathbf{R}^n$中两个凸集,β是实数，则

(1) $\beta S_1=\{\beta\boldsymbol{x}|\boldsymbol{x}\in S_1\}$为凸集；

(2) $S_1\cap S_2$为凸集；

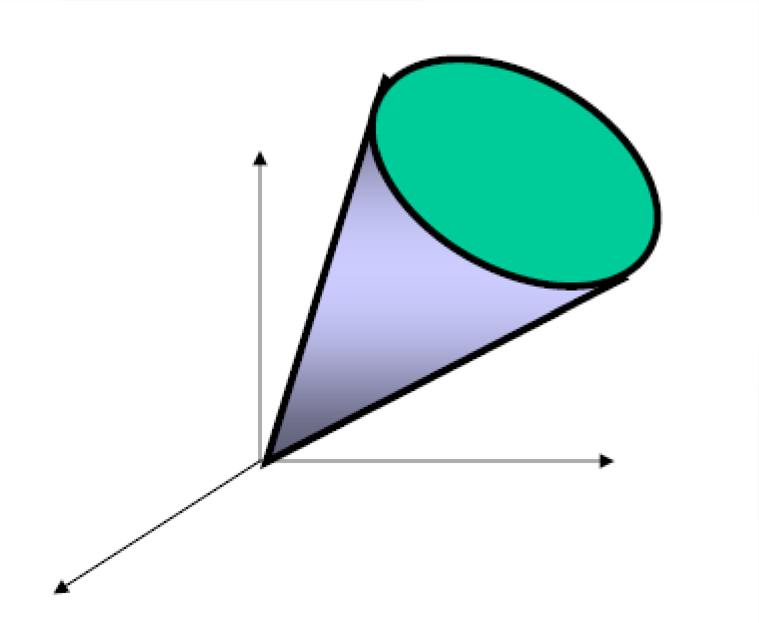
(3) $S_1+S_2=\left\{\boldsymbol{x}^{(1)}+\boldsymbol{x}^{(2)}|\boldsymbol{x}^{(1)}\in S_1,\boldsymbol{x}^{(2)}\in S_2\right\}$为凸集；

(4) $S_1-S_2=\left\{\boldsymbol{x}^{(1)}-\boldsymbol{x}^{(2)}|\boldsymbol{x}^{(1)}\in S_1,\boldsymbol{x}^{(2)}\in S_2\right\}$为凸集 .

● 锥、凸锥

定义 设有集合$C \subset \mathbf{R}^n$，若对C中每一点x，当λ取任何非负数时，都有$\lambda x \in C$，则称C为**锥**，又若C为凸集，则称C 为**<u>凸锥</u>**.

例 向量集$\alpha^{(1)}, \alpha^{(2)}, \cdots, \alpha^{(k)}$的所有非负线性组合构成的集合$\{\sum_{i=1}^{k} \lambda_i \alpha^{(i)} | \lambda_i \geqslant 0, i = 1, \cdots, k\}$为凸锥。



多面集

定义 有限个半空间的交

$$\{\boldsymbol{x} | \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}\}$$

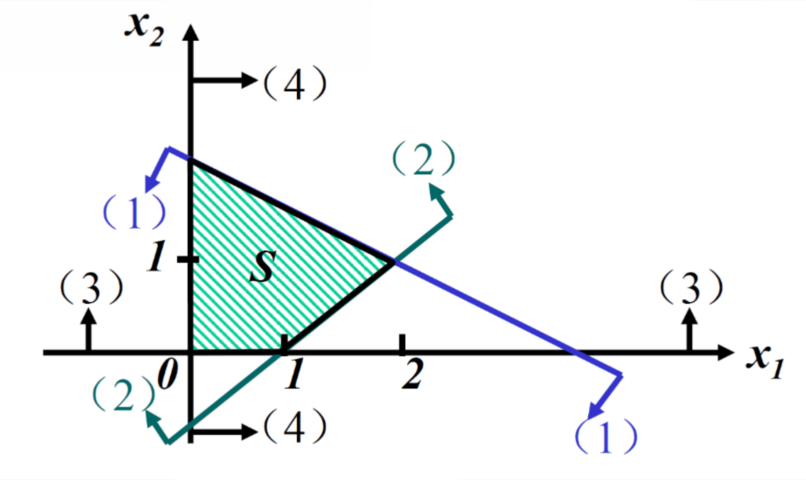
称为**多面集**，其中A为$m \times n$矩阵，$\boldsymbol{b}$为m维向量．

多面锥：$\{\boldsymbol{x} | \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{0}\}$ $\longrightarrow$ 凸锥

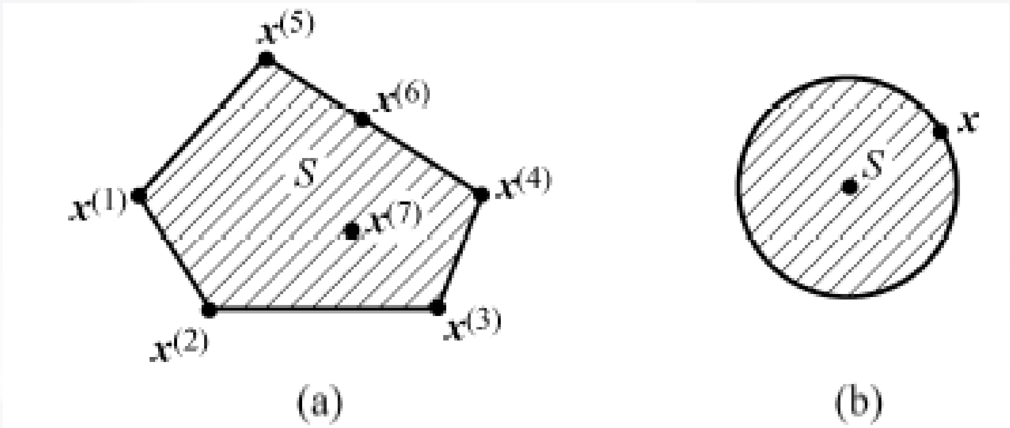
例 4 集合

$$S=\{x|x_1+2x_2\leqslant 4, x_1-x_2\leqslant 1, x_1\geqslant 0, x_2\geqslant 0\}$$

为多面集. 其几何表示如图画斜线部分 .



定义 设S为非空凸集，$\boldsymbol{x} \in S$，若$\boldsymbol{x}$不能表示成S中两个不同点的凸组合；换言之，若假设$\boldsymbol{x} = \lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)}$ $(\lambda \in (0,1)), \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in S$，必推得$\boldsymbol{x} = \boldsymbol{x}^{(1)} = \boldsymbol{x}^{(2)}$，则称$\boldsymbol{x}$是凸集$S$的**极点**。



对于**紧**凸集（$\mathbf{R}^n$中的凸**有界闭集**），任何一点都能表示成极点的凸组合。
但是对于**无界集**并不成立。

定义 (极方向)

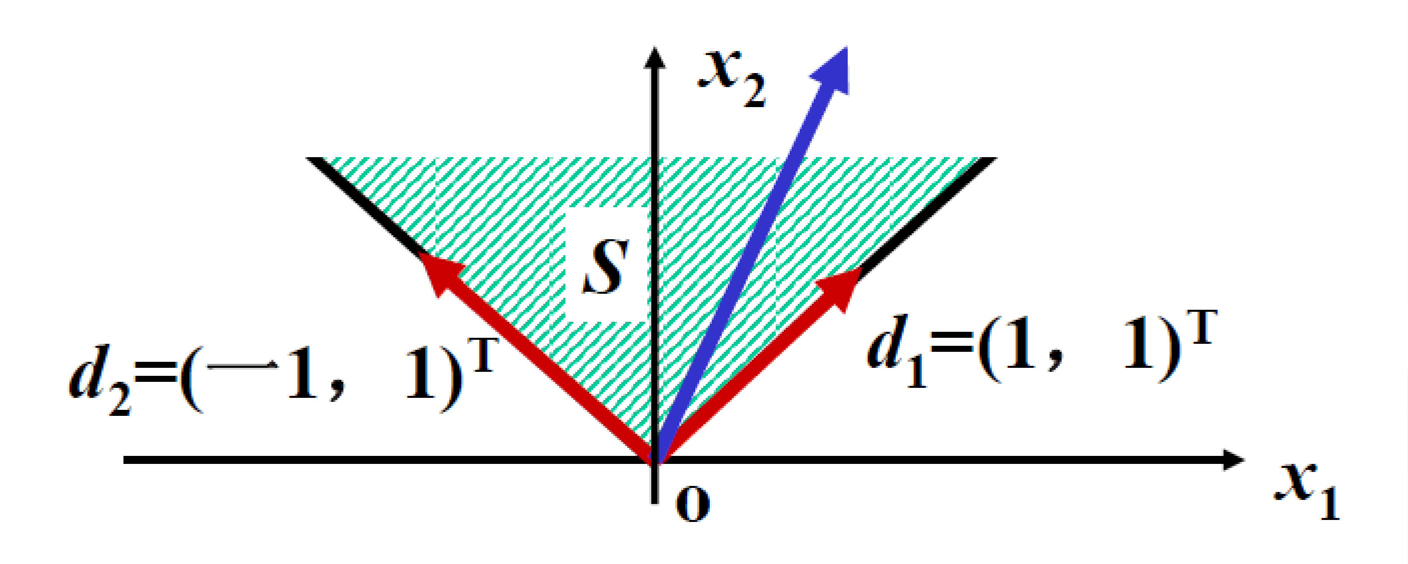
- 设S为$\mathbf{R}^n$中的闭凸集，$\boldsymbol{d}$为非零向量，如果对S中的每一个$\boldsymbol{x}$，都有射线$\{\boldsymbol{x}+\lambda\boldsymbol{d}|\lambda\geqslant 0\}\subset S$，则称向量$\boldsymbol{d}$为$S$的**方向**.
- 又设$\boldsymbol{d}^{(1)}$和$\boldsymbol{d}^{(2)}$是S的两个方向，若对任何正数λ，有$\boldsymbol{d}^{(1)}\neq\lambda\boldsymbol{d}^{(2)}$，则称$\boldsymbol{d}^{(1)}$和$\boldsymbol{d}^{(2)}$是两个**不同的方向**.
- 若S的方向$\boldsymbol{d}$不能表示成该集合的两个不同方向的正的线性组合，则称$\boldsymbol{d}$为S的**极方向**.

有界集不存在方向，因而也不存在极方向。对于无界集才有方向的概念。

例 5 集合

$$S=\{(x_1,x_2)|x_2 \geqslant |x_1|\},$$

凡是与向量$(0,1)^{\mathrm{T}}$夹角小于或等于45°的向量，都是它的方向．其中$(1,1)^{\mathrm{T}}$和$(-1,1)^{\mathrm{T}}$是S的两个极方向．S的其他方向都能表示成这两个极方向的正线性组合.



例 6 设

$$S=\{\boldsymbol{x}|\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq\boldsymbol{0}\}$$

为非空集合，$\boldsymbol{d}$是非零向量。证明d为S的方向的充要条件是$\boldsymbol{d}\geq 0$且$\boldsymbol{A}\boldsymbol{d}=\boldsymbol{0}$。

证明：按照定义，$\boldsymbol{d}$为S的方向的充要条件是：对每一个$x\in S$，有 $\{\boldsymbol{x}+\lambda\boldsymbol{d}|\lambda\geqslant 0\}\subset S.$

根据集合S的定义，即

$$\boldsymbol{A}(\boldsymbol{x}+\lambda\boldsymbol{d})=\boldsymbol{b}\ ,$$
$$\boldsymbol{x}+\lambda\boldsymbol{d}\geq\boldsymbol{0}$$

由于$\mathbf{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq\boldsymbol{0}$及$\lambda$可取任意非负数，

因此$\boldsymbol{A}\boldsymbol{d}=\boldsymbol{0},\boldsymbol{d}\geq\boldsymbol{0}$.

表示定理

- 表示定理是多面集的一个重要性质。
- 对于有界多面集，它的每一个点可用极点的凸组合来表示。反之，由极点的凸组合表示的点一定属于这个多面集。

定理（表示定理）

设$S=\{\boldsymbol{x}|\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq\boldsymbol{0}\}$为非空多面集，则有：

（1）极点集非空，且存在有限个极点$\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(k)}$.

（2）极方向集合为空集的充要条件是S有界．若S无界，则存在有限个极方向$\boldsymbol{d}^{(1)},\cdots,\boldsymbol{d}^{(l)}$.

（3）$\boldsymbol{x}\in S$的充要条件是：

$$\boldsymbol{x}=\sum_{j=1}^{k}\lambda_j\boldsymbol{x}^{(j)}+\sum_{j=1}^{l}\mu_j\boldsymbol{d}^{(j)}$$

$$\sum_{j=1}^{k}\lambda_j=1$$

$$\lambda_j\geqslant 0,\quad j=1,\cdots,k$$

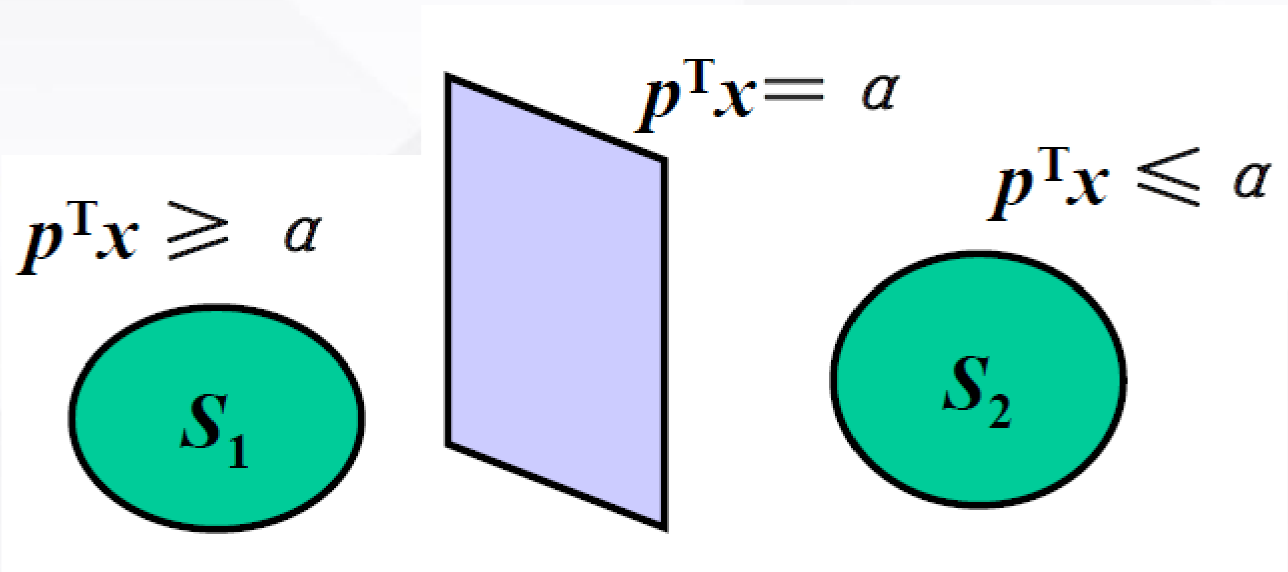
$$\mu_j\geqslant 0,\quad j=1,\cdots,l$$

凸组合

正线性组合

凸集分离定理

定义 设S_1和S_2是$\mathbf{R}^n$中两个非空集合，$H=\left\{\boldsymbol{x}|\boldsymbol{p}^{\mathrm{T}}\boldsymbol{x}=\alpha\right\}$为超平面．如果对每个$\boldsymbol{x}\in S_1$,都有$p^{\mathrm{T}}\boldsymbol{x}\geq\alpha$，对于每个$\boldsymbol{x}\in S_2$,都有$p^{\mathrm{T}}\boldsymbol{x}\leq\alpha$（或情形恰好相反), 则称超平面H分离集合$S_1$和$S_2$　.



定理

设S为$\mathbf{R}^{\mathrm{n}}$中的闭凸集，$\boldsymbol{y} \notin S$,则存在惟一的点$\bar{\boldsymbol{x}} \in S$, 使得

$$\|\boldsymbol{y}-\bar{\boldsymbol{x}}\| = \inf_{x \in S}\|\boldsymbol{y}-\boldsymbol{x}\|$$

(inf: greatest lower bound)

- (点与闭凸集分离)设S为$\mathbf{R}^{\mathrm{n}}$中的闭凸集，$\boldsymbol{y} \notin S$，$H = \{\boldsymbol{x} \mid \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x} = \alpha\}$为超平面．根据定义，$H$分离点$\boldsymbol{y}$与集合$S$意味着，若$\boldsymbol{p}^{\mathrm{T}}\boldsymbol{y} > \alpha$，则$\boldsymbol{p}^{\mathrm{T}}\boldsymbol{x} \le \alpha$，$\forall \boldsymbol{x} \in S$．令$\boldsymbol{p}^{\mathrm{T}}\boldsymbol{y} - \alpha = \varepsilon$, 于是$\boldsymbol{y}$与$S$的分离可表示为

$$\boldsymbol{p}^{\mathrm{T}}\boldsymbol{y} \ge \varepsilon + \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x}, \quad \forall \boldsymbol{x} \in S.$$

■ 定理

设S为$\mathbf{R}^{\mathrm{n}}$中的非空闭凸集，$\boldsymbol{y} \notin S$,则存在非零向量$\boldsymbol{p}$及数$\varepsilon>0$，使得对每个点$\boldsymbol{x} \in S$，成立$\boldsymbol{p}^{T}\boldsymbol{y} \geqslant \boldsymbol{\varepsilon} + \boldsymbol{p}^{T}\boldsymbol{x}$.

■ 定理

设S为$\mathbf{R}^{\mathrm{n}}$中的非空凸集，$\boldsymbol{y} \in \partial S$,则存在非零向量$\boldsymbol{p}$，使得对每个点$\boldsymbol{x} \in clS$，成立$\boldsymbol{p}^{T}\boldsymbol{y} \geqslant \boldsymbol{p}^{T}\boldsymbol{x}$.

- $cl\,S$表示集合S的闭包（closure，由S的内点和边界点组成的集合）。
- ∂y表示边界。

■ 推论

设S为$\mathbf{R}^{\mathbf{n}}$中的非空凸集, $\boldsymbol{y} \notin S$,则存在非零向量$\boldsymbol{p}$，使得对每个点$\boldsymbol{x} \in clS$, 成立$\boldsymbol{p}^{\boldsymbol{T}}(\boldsymbol{x}-\boldsymbol{y}) \leqslant \mathbf{0}$.

■ 定理

设S_1、S_2为$\mathbf{R}^{\mathbf{n}}$中的两个非空凸集, $S_1 \cap S_2=\emptyset$,则存在非零向量$\boldsymbol{p}$, 使得

$$\inf\left\{\mathbf{p}^{T}\mathbf{x} \middle| \mathbf{x} \in S_1\right\} \geqslant \sup\left\{\mathbf{p}^{T}\mathbf{x} \middle| \mathbf{x} \in S_2\right\}$$

两个非空凸集的分离定理

■ 定理（Farkas定理）

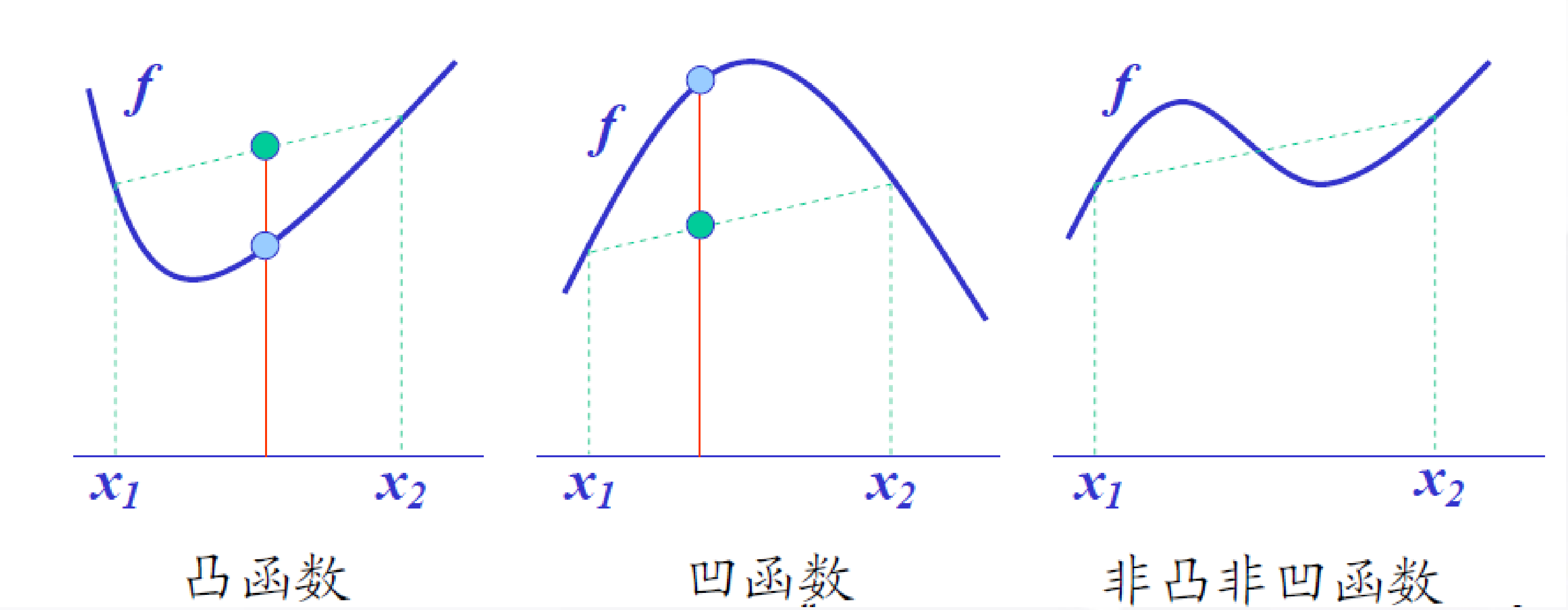
设A为$m \times n$矩阵, $\boldsymbol{c}$为n维向量, 则$\boldsymbol{Ax} \leq \mathbf{0}$, $\boldsymbol{c}^T\boldsymbol{x} > 0$有解的充要条件是$\boldsymbol{A}^T\boldsymbol{y} = \boldsymbol{c}$，$\boldsymbol{y} \geq \mathbf{0}$ 无解.

■ 定理（Gordan定理）

设A为$m \times n$矩阵, $\boldsymbol{c}$为n维向量, 则$\boldsymbol{Ax} < \mathbf{0}$有解的充要条件是存在非零向量$\boldsymbol{y} \geq \mathbf{0}$,使$\boldsymbol{A}^T\boldsymbol{y} = \mathbf{0}$.

凸函数

称定义于凸集S上的函数$f(x)$是凸函数如果对于任意$x_1,x_2\in S$和$\lambda\in(0, 1)$，都有$f(\lambda x_1 + (1 - \lambda)x_2)\leq\lambda f(x_1)+ (1 - \lambda) f(x_2)$。称其为严格凸的如果不等式取严格小于号。另外，称函数$f(x)$是一个(严格) 凹函数如果 - $f(x)$是一个(严格) 凸函数。



Q2.

凸函数性质

凸函数性质：若 f 为定义在凸集 S 上的凸函数，则：

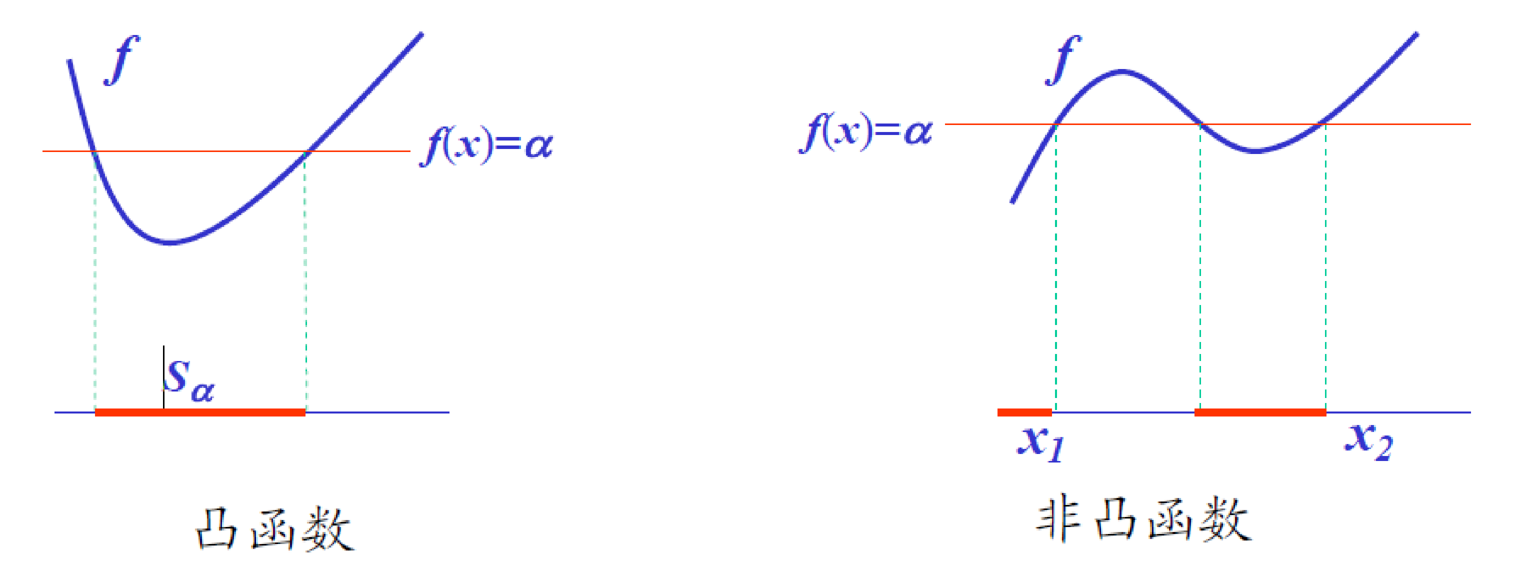
- 局部最小值就是全局最小值。
- 对任意实数 $\beta \geq 0$，函数 βf 也是定义在 S 上的凸函数;
- 两(多)个凸函数的和仍为凸函数；

为什么凸函数和凸集中都叫凸（convex）？

它们有什么联系吗？

凸函数性质

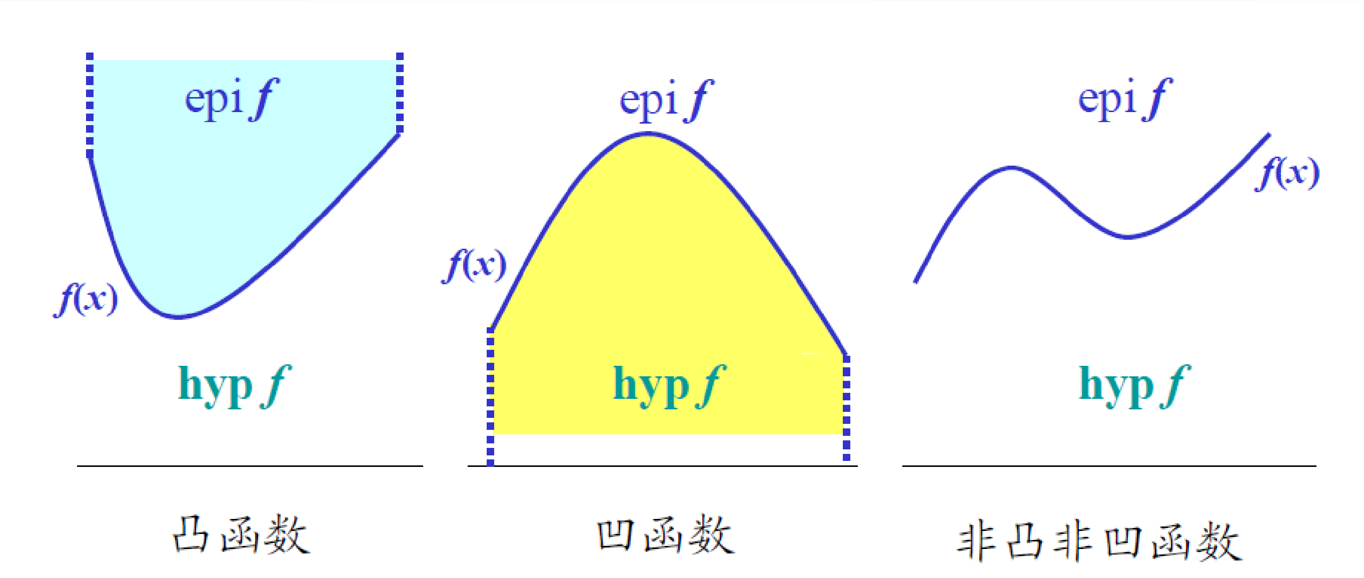
练习1：若 f 为定义在凸集 S 上的凸函数，对每一实数c，水平集$\{x|x\in S, f(x)\le c\}$都是凸集吗？



Q3.证明

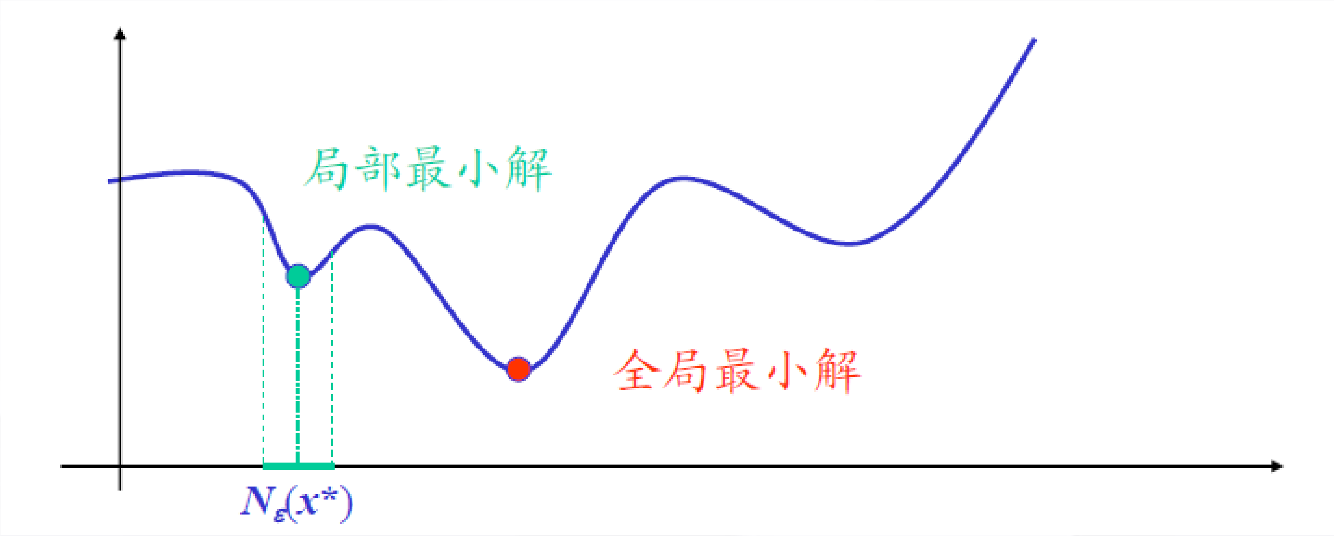
凸函数性质

练习2：凸集 $S\subset R^n$ ，f 为凸函数当且仅当其上图 epi $f=\{(x,y)|x\in S, f(x)\le y\}$ 是凸集？



最优解概念

考虑求最小值问题$\min\{f(x)|\ x\in S\subset R^n\}$。设 $x^*\in S$。若对所有 $x\in S$ 都有 $f(x^*)\le f(x)$，则称 $x^*\in S$ 是一个全局最小解或者全局最优解。若存在x^*的一个邻域$N_\varepsilon(x^*)$，使得对任意 $x\in S\cap N_\varepsilon(x^*)$，都有 $f(x^*)\le f(x)$，则称 x^* 为一个局部最小解或者局部最优解。



凸函数最优条件

■ **定理1.4.13（凸函数的重要性质）**

设S是R^n中非空**凸集**，f是定义在凸集S上的**凸函数**，则f在S上的局部极小点是**全局极小点**，且极小点的集合为**凸集**.

Q4.证明

设$S \subset R^n$是一个非空凸集，函数 $f(x)$：$S \to R$。另设$x^* \in S$是求最小值问题$\min \{f(x) | x \in S \subset R^n\}$的一个局部最优解。

i) 若 $f(x)$是一个凸函数，则x^*是一个全局最优解。

ii) 若 $f(x)$是一个严格凸函数，则x^*是惟一一个全局最优解。

凸函数的判别

◆ 定义 （一阶充要条件）

设S是R^n中非空开凸集，f(x)是定义在S上的可微函数，则f(x)为凸函数的充要条件是对任意两点$x^{(1)}, x^{(2)} \in S$都有

$$f\left(x^{(2)}\right) \geqslant f\left(x^{(1)}\right) + \nabla f\left(x^{(1)}\right)^T\left(x^{(2)} - x^{(1)}\right)$$

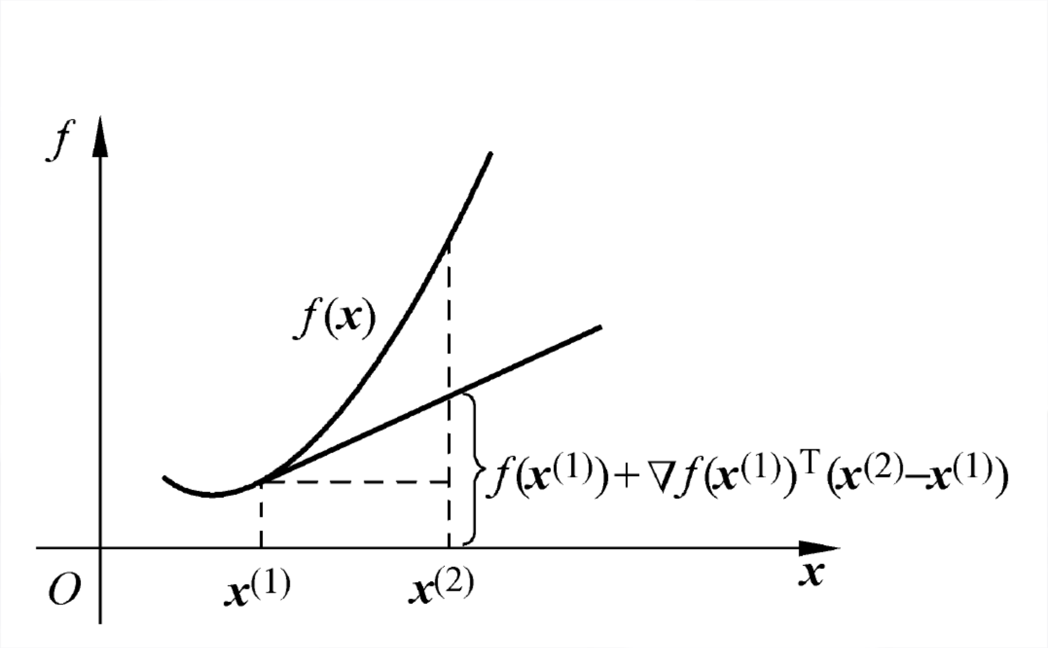
而f(x) 为严格凸函数的充要条件是对任意的**互不相同**的$x^{(1)}, x^{(2)} \in S$，成立

$$f\left(x^{(2)}\right) > f\left(x^{(1)}\right) + \nabla f\left(x^{(1)}\right)^T\left(x^{(2)} - x^{(1)}\right)$$

凸函数的判别

- **推论** 设S是R^n中的凸集，$\bar{\boldsymbol{x}} \in S$, f(x)是定义在$R^n$上的凸函数，则对任意x ∈ S，有

$$f(\mathbf{x}) \geqslant f(\bar{\mathbf{x}}) + \Delta f(\bar{\mathbf{x}})^T(\mathbf{x} - \bar{\mathbf{x}})$$



凸函数的判别

- **定理（二阶条件）** 设S是R^n中的非空开凸集，f(x)是定义在S上的二次可微函数，则f(x)为凸函数的充要条件是在每一点$x \in S$处Hesse矩阵半正定。

- **定理（严格凸函数判别条件）** 设S是R^n中的非空开凸集，f(x)是定义在S上的二次可微函数，如果在每一点$x \in S$处Hesse矩阵正定，则f(x)为严格凸函数。

注意：逆定理并不成立．若f(x)是定义在S上的严格凸函数，则在每一点$x \in S$处Hesse矩阵半正定．

凸函数的判别

- **练习题：** 判断二次函数是否为凸函数。

$$f(x_1,x_2)=2x_1^2+x_2^2-2x_1x_2+x_1+1$$

凸规划

极小化问题：

$$\min f(\mathbf{x})$$

$$\text{s.t. } g_i(\mathbf{x}) \geqslant 0 \quad i = 1, \ldots, m,$$

$$h_j(\mathbf{x}) = 0 \quad j = 1, \ldots, l$$

$f(\mathbf{x})$：凸函数

g_i：凹函数

h_j：线性函数

问题的可行域

$S = \{x | g_i(x) \geqslant 0 \quad i = 1, \ldots, m, \; h_j(x) = 0 \quad j = 1, \ldots, l\}$ ：<u>凸集</u>

（$-g_i$为凸函数，凸函数的水平集为凸集）

即求：凸函数在凸集上的极小点

- 凸规划是非线性规划中一种重要的特殊情形，它具有很好的性质
- 凸规划的局部极小点就是全局极小点，且极小点的集合是凸集
- 如果凸规划的目标函数是严格凸函数，又存在极小点，那么它的极小点是惟一的.