运筹学与优化方法

晁国清
计算机科学与技术学院

课程回顾

- 运筹学定义

运筹学是把现实世界的问题抽象成数学模型然后用最优化方法求解使目标最优

- 最优化问题的分类
- 最优化问题的数学建模

解题步骤:
1) 设决策变量
2) 列目标函数
3) 列约束条件

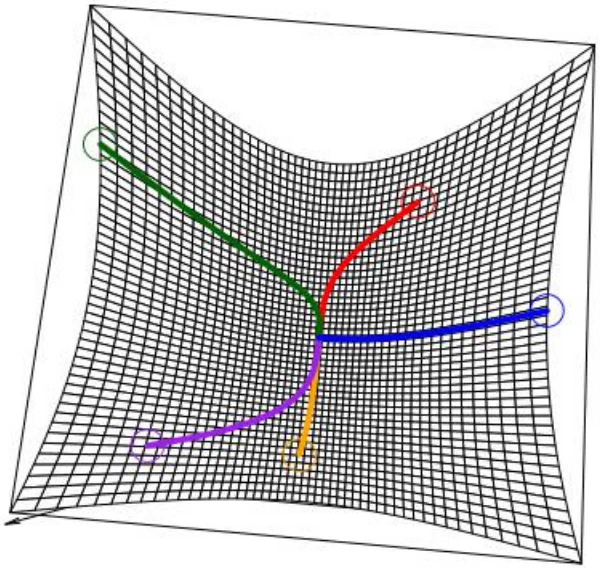
- 无约束优化与约束优化
- 线性优化与非线性优化
- 凸优化与非凸优化
- 连续优化与离散优化（运输问题）
- 单目标优化与多目标优化（行车路线规划）
- 确定性优化与随机优化/鲁棒性优化

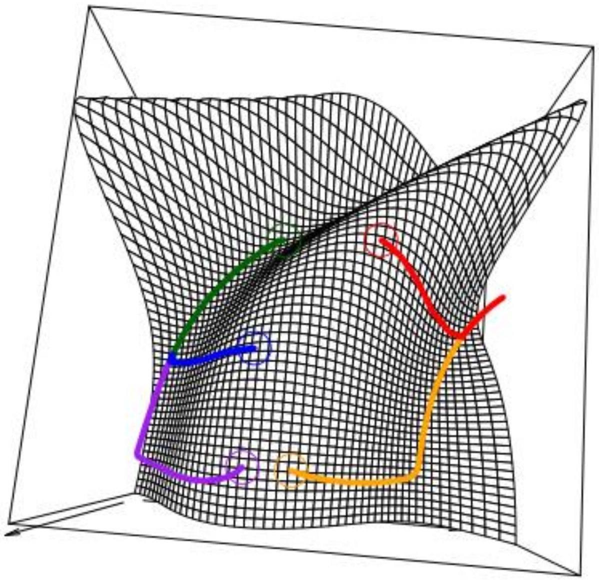
为什么凸优化问题重要？

凸优化问题形式：$\begin{cases} \min\limits_{x} f(x) \\ s.t.\ g_i(x) \le 0, i = 1, \cdots, m \\ \quad h_j(x) = 0, j = 1, \cdots, l \end{cases}$

f是凸函数。可行集$S = \{x \in X | g_i(x) \le 0, i = 1, \cdots, m, h_i(x) = 0, i = 1, \cdots, l\}$是凸集合。

特性：凸优化问题的任意局部最优解就是一个全局最优解





局部最优解与全局最优解

凸优化问题形式： $\begin{cases} \min\limits_{x} f(x) \\ s.t.\, g_i(x) \le 0, i = 1, \cdots, m \\ \quad h_j(x) = 0, j = 1, \cdots, l \end{cases}$

- 局部最优解：指对于一个问题的解在一定范围或区域内最优，或者说解决问题或达成目标的手段在一定范围或限制内最优。

数学表示：对于可行集S的一个子集S_D，存在$\exists x^* \in S_D$使得$f(x^*) \le f(x), \forall x \in S_D$

- 全局最优解：指针对一定条件或环境下的一个问题或目标，若一项决策和所有解决该问题的决策相比是最优的。

数学表示：对于可行集S，存在$\exists x^* \in S$使得$f(x^*) \le f(x), \forall x \in S$

课程内容

- 基本概念和符号
- 凸集的定义
- 常见凸集
- 保持凸性的运算
- 凸集合的性质

基本概念和符号

1. 向量和子空间投影定理

（1） n维欧氏空间：R^n

点（向量）：$\boldsymbol{x} \in \mathrm{R}^n$，$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$

分量 $x_i \in \mathrm{R}$（实数集）

方向（自由向量）：$\boldsymbol{d} \in \mathrm{R}^n$，$\boldsymbol{d} \neq 0$

$\boldsymbol{d} = (d_1, d_2, \cdots, d_n)^{\mathrm{T}}$ 表示从0指向$\boldsymbol{d}$ 的方向

实际使用中，常用 $\boldsymbol{x} + \lambda \boldsymbol{d}$ 表示从$\boldsymbol{x}$ 点出发沿$\boldsymbol{d}$ 方向移动$\lambda \boldsymbol{d}$ 长度得到的点。

基本概念和符号

（2） 向量运算：x，$y \in \mathrm{R}^n$

x，y 的内积：$x^T y = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$

x，y 的距离： $\| x-y \| = [(x-y)^{\mathrm{T}}(x-y)]^{(1/2)}$

x 的长度： $\| x \| = [x^T x]^{(1/2)}$

三角不等式： $\| x+y \| \leqslant \| x \| + \| y \|$

点列的收敛：设点列 $\{x^{(k)}\} \subset \mathrm{R}^n$，$x \in \mathrm{R}^n$

点列 $\{x^{(k)}\}$ 收敛到 x，记

$$\lim_{k\to\infty} x^{(k)} = x \Leftrightarrow \lim_{k\to\infty} \| x^{(k)} - x \| = 0 \Leftrightarrow \lim_{k\to\infty} x_i^{(k)} = x_i, \forall i$$

基本概念和符号

(3) 子空间：设$\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(m)} \in \boldsymbol{R}^n, \boldsymbol{d}^{(k)} \neq \boldsymbol{0}$，记

$$L\left(\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(m)}\right) = \{x = \sum_{j=1}^{m} a_j d^j | a_j \in R\}$$

为由向量$\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(m)}$生成的子空间，简记为L。

正交子空间：设L为$\boldsymbol{R}^n$的子空间，其正交子空间为

$$L^{\perp} = \{x \in \boldsymbol{R}^n | x^T y = 0, \forall y \in L\}$$

子空间投影定理：设L为$\boldsymbol{R}^n$的子空间，那么$\forall z \in \boldsymbol{R}^n$，$\exists$唯一$x \in L, y \in L^{\perp}$，使$z = x + y$，且$x$为问题

$$\begin{aligned} &\min \quad \| z - u \| \\ &\text{s.t.} \quad u \in L \end{aligned}$$

的唯一解，最优值为$\| y \|$。

特别地，$L = \boldsymbol{R}^n$时，正交子空间 $L^{\perp} = \{0\}$（零空间）。

基本概念和符号

- 规定：$x, y \in \mathrm{R}^n$，$x \leqslant y \Leftrightarrow x_i \le y_i, \forall i$；
类似地规定 $x \geqslant y$，$x = y$，$x < y$，$x > y$。

- 一个有用的定理

  设$x \in \mathrm{R}^n$，$\alpha \in \mathrm{R}$，L为R^n 的线性子空间。

  若 $x^{\mathrm{T}}y \leqslant \alpha$，$\forall\ y \in \mathrm{R}^n$ 且 $y \geqslant 0$，则 $x \leqslant 0$，$\alpha \geqslant 0$

  若 $x^{\mathrm{T}}y \leqslant \alpha$，$\forall\ y \in L \subseteq \mathrm{R}^n$，则

  $x \in L^{\perp}$，$\alpha \geqslant 0$（特别地，当$L = \mathrm{R}^n$时，$x = 0$）

- 定理的其他形式：

若 $x^{\mathrm{T}}y \leqslant \alpha$，$\forall y \in \mathrm{R}^n$ 且 $y \leqslant 0$，则 $x \geqslant 0$，$\alpha \geqslant 0$。

若 $x^{\mathrm{T}}y \geqslant \alpha$，$\forall y \in \mathrm{R}^n$ 且 $y \geqslant 0$，则 $x \geqslant 0$，$\alpha \leqslant 0$。

若 $x^{\mathrm{T}}y \geqslant \alpha$，$\forall y \in \mathrm{R}^n$ 且 $y \leqslant 0$，则 $x \leqslant 0$，$\alpha \leqslant 0$。

若 $x^{\mathrm{T}}y \geqslant \alpha$，$\forall y \in L \subseteq \mathrm{R}^n$，则 $x \in L^{\perp}$，$\alpha \leqslant 0$。

基本概念和符号

2. 多元函数及其导数

(1) n元函数：$f(x)$：$R^n \to R$

线性函数：$f(x) = c^T x + b = \sum c_i x_i + b$

二次函数：$f(x) = (1/2)\ x^T Q x + c^T x + b$

$= (1/2) \sum \sum a_{ij} x_i x_j + \sum c_i x_i + b$

向量值线性函数：$F(x) = Ax + d \in R^m$

其中， A为 $m \times n$矩阵，d为m维向量

$F(x) = (\ f_1(x),\ f_2(x),\ \cdots,\ f_m(x)\ )^T$

记 a_i^T为A的第i行向量，$f(x) = a_i^T x + d_i$

哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

基本概念和符号

（2）梯度（一阶偏导数向量）：

$$\nabla f(x)=(\partial f/\partial x_1,\ \partial f/\partial x_2,\ \cdots,\ \partial f/\partial x_n)^{T}\in R^{n}$$

线性函数：$f(x)=c^{T}x+b$，$\nabla f(x)=c$

二次函数：$f(x)=(1/2)\,x^{T}Qx+c^{T}x+b$

$$\nabla f(x)=Qx+c$$

向量值线性函数：$F(x)=Ax+d\in R^{m}$

$$\partial F/\partial x=A^{T}$$

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

基本概念和符号

（3） Hesse 矩阵（二阶偏导数矩阵）：

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}, & \frac{\partial^2 f}{\partial x_2 \partial x_1}, & \cdots, & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}, & \frac{\partial^2 f}{\partial x_2^2}, & \cdots, & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \cdots, & \cdots, & \cdots, & \cdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}, & \frac{\partial^2 f}{\partial x_2 \partial x_n}, & \cdots, & \frac{\partial^2 f}{\partial^2 x_n}, \end{bmatrix}$$

线性函数： $f(x) = c^T x + b$, $\nabla^2 f(x) = 0$

二次函数： $f(x) = (1/2)\, x^T Q x + c^T x + b$, $\nabla^2 f(x) = Q$

基本概念和符号

(4) n元函数的Taylor展开式及中值公式

设 $f(x): R^n \to R$，二阶可导。在x^* 的邻域内，有

➢一阶Taylor展开式：

$$f(x) = f(x^*) + \nabla f^T(x^*)(x-x^*) + o\|x - x^*\|$$

➢二阶Taylor展开式：

$$f(x) = f(x^*) + \nabla f^T(x^*)(x-x^*) + (1/2)(x-x^*)^T \nabla^2 f(x^*)(x-x^*) + o\|x - x^*\|^2$$

➢一阶中值公式：对x, $\exists\ \lambda \in (0,1)$，使

$$f(x) = f(x^*) + [\nabla f(x^*+\lambda(x - x^*))]^T(x - x^*)$$

➢Lagrange余项：对x, $\exists\ \mu \in (0,1)$，记$x_\mu = x^* + \mu(x - x^*)$

$$f(x) = f(x^*) + \nabla f^T(x)(x - x^*) + (1/2)(x - x^*)^T \nabla^2 f(x_\mu)(x - x^*)$$

基本概念和符号

➢复习下列知识：

线性代数的有关概念：向量与矩阵的运算，向量的线性相关和线性无关，矩阵的秩，正定、半正定矩阵，线性空间等。集合的有关概念：开集、闭集，集合运算，内点、边界点等。

凸集

凸集：对任何$x_1 \in C$，$x_2 \in C$, 有$\lambda x_1 + (1-\lambda)x_2 \in C(0 \le \lambda \le 1)$, 则称$C$为凸集。如下图(a)是凸集，(b)不是凸集。

通俗的讲：如果对于集合C中任意两个点x_1, x_2，它们连线上的所有点也都是集合C中的点，则C为凸集。

(a) (b)

凸组合

- 凸组合convex combination of $x_1, \cdots, x_k$:

$$\lambda_1 x_1 + \cdots + \lambda_k x_k$$

其中，$\lambda_1, \cdots, \lambda_k \geq 0, \sum_{i=1}^{k} \lambda_i = 1$

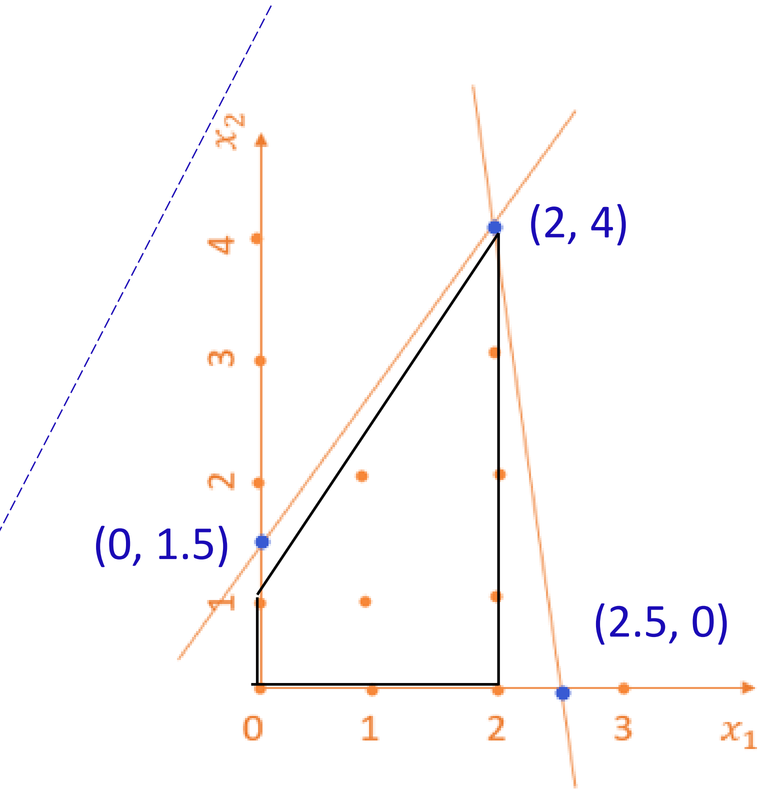
凸包convex hull of set C : 由C中点的凸组合构成。

例如：

凸包的用途

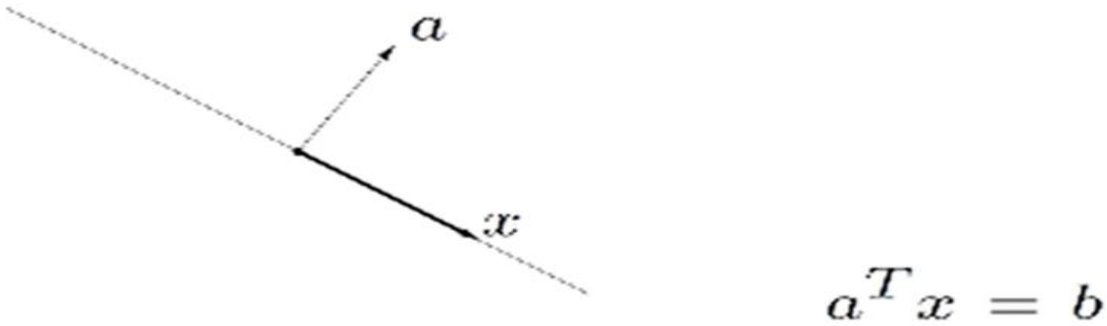
凸包可以将非凸优化
转化为凸优化问题

$$\begin{aligned} &\min 2x_1 - x_2 \\ &s.t.\ 8x_1 + x_2 \le 20 \\ &\qquad 1.25x_1 - x_2 \ge -1.5 \\ &\qquad x_i \ge 0, i = 1,2 \\ &\qquad x_i \in Z, i = 1,2 \end{aligned}$$

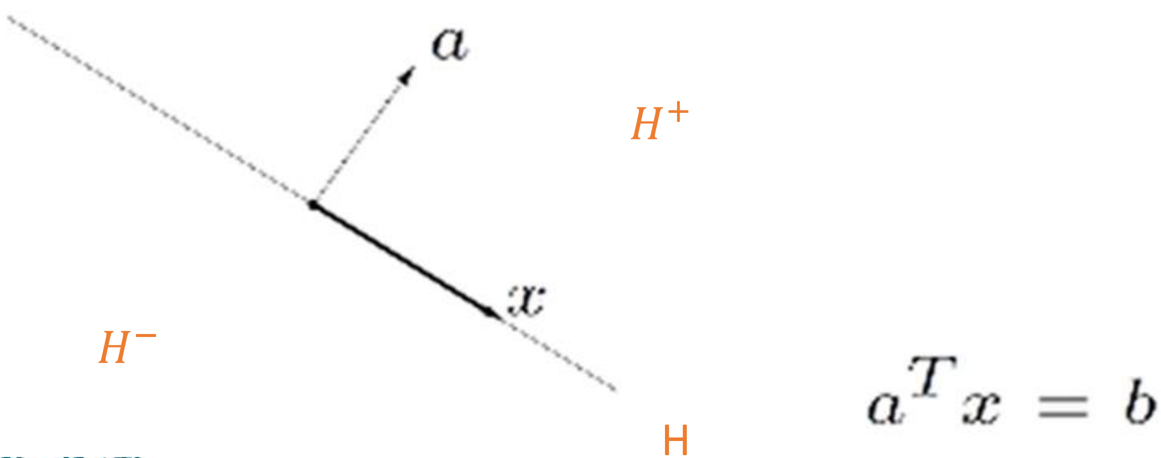


常见凸集

- 超平面hyperplane： $H=\{x|a^Tx=b\}(a\neq 0);$



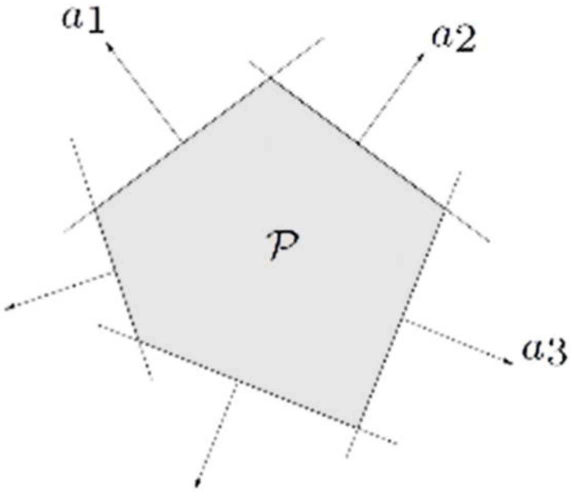
- 半空间halfspace： $H^+=\{x|a^Tx\geq b\}(a\neq 0)$
$H^-=\{x|a^Tx\leq b\}(a\neq 0);$



常见凸集

- 多面体Polyhedra:多个线性不等式所刻画的集合

$\{x|a_i^T x \leq b_i, i = 1,\cdots,m\}$



思考题：由线性等式刻画的集合是多面体么?

$$a_i^T x = b_i \Leftrightarrow \begin{cases} a_i^T x \geq b \\ a_i^T x \leq b \end{cases}$$

常见凸集

- 球体 Ball with center x_c and radius r:

$$B(x_c, r) = \{x \,|\, \|x - x_c\|_2 \le r\} = \{x_c + ru \,|\, \|u\|_2 \le 1\}$$

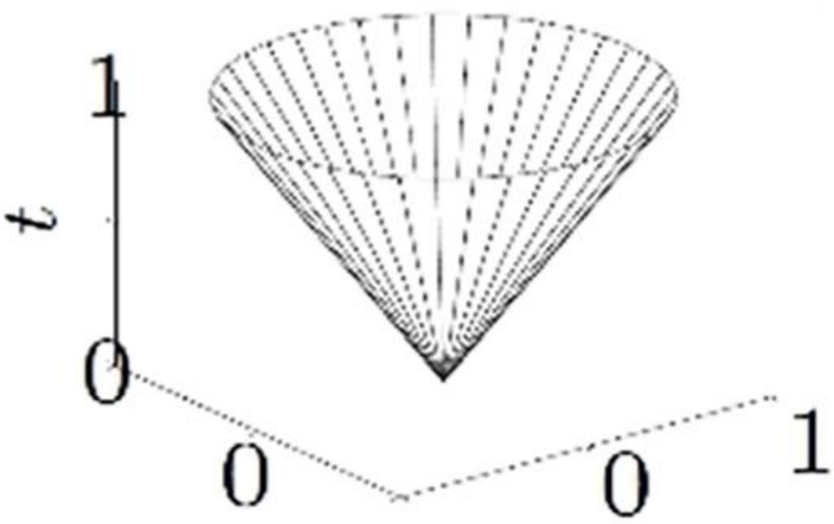
- 椭球Ellipsoid:

$$\{x \,|\, (x - x_c)^T P^{-1} (x - x_c) \le 1\}$$

其中P为正定矩阵

- 二阶锥Second-Order Cone:

$$\{(x, t) \,|\, \|x\|_2 \le t\}$$



常见凸集

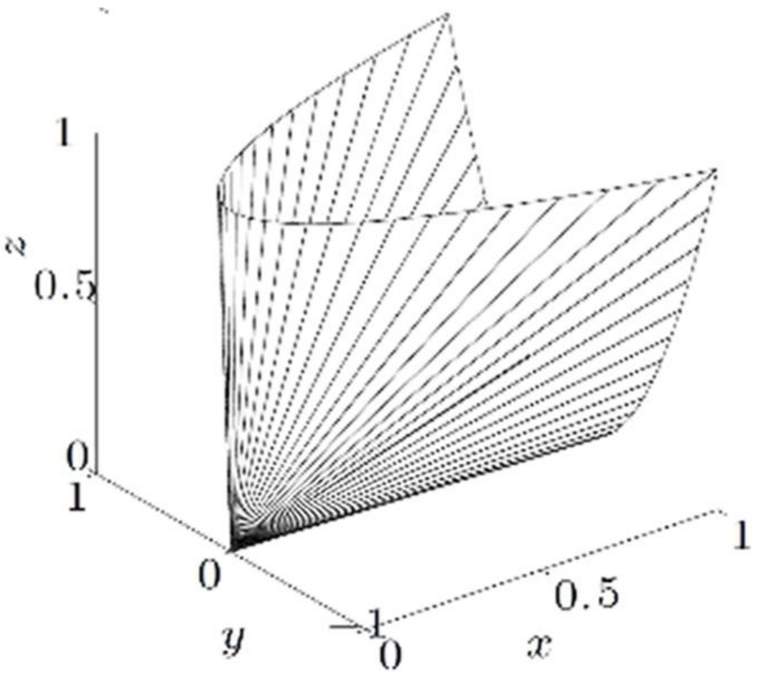
半定矩阵锥

- S^n: 所有n阶对称矩阵组成的集合
- $S_+^n=\{X\in S^n|X\succcurlyeq 0\}$:所有半正定矩阵组成的集合，这里

$$X\succcurlyeq 0 \Leftrightarrow z^TXz\geq 0,\forall z$$

- $\mathrm{S}_{++}^{\mathrm{n}}=\{\mathrm{X}\in \mathrm{S}^{\mathrm{n}}|\mathrm{X}\succ 0\}$:所有正定矩阵组成的集合

例子: $\begin{bmatrix} x & y \\ y & z \end{bmatrix}\in S_+^2$



常见凸集

- 记线性规划$\min\{c^Tx|Ax=b, x\geq 0\}$的最优解组成的集合为S，问S是凸集合么？

$$\forall x_1, x_2 \in S, \qquad c^Tx_1 = c^Tx_2 = d$$

$$\forall\, \lambda \in [0,1], \qquad \lambda x_1 + (1-\lambda)x_2 ?\in S$$

$$c^T(\lambda x_1 + (1-\lambda)x_2)$$

$$= \lambda c^Tx_1 + (1-\lambda)c^Tx_2$$

$$= d$$

保持凸性的运算

- 设$C_1, C_2 \subset R^n$是凸集，$\alpha \in R$, 则

（1）$C_1 \cap C_2 = \{x | x \in C_1, x \in C_2\}$是凸集

（2）$C_1 \pm C_2 = \{x \pm y | x \in C_1, y \in C_2\}$是凸集

问题：$S = \{\mathrm{x} \in R^n || p(t)| \le 1, |t| \le \pi/3\}$, 其中

$$p(t) = \sum_{i=1}^{n} x_i cosit$$

是否为凸集?

HARBIN INSTITUTE OF TECHNOLOGY

保持凸性的运算

仿射变换

假设$f: R^n \to R^m$是仿射函数，即$f(x) = Ax + b, A \in R^{m\times n}, b \in R^m$.

- C为凸集$\Rightarrow f(C) = \{f(x) | x \in C\}$是凸集
- C为凸集$\Rightarrow f^{-1}(C) = \{x | f(x) \in C\}$是凸集

特殊仿射变换

- 放缩scaling:$\alpha C = \{\alpha x | x \in C\}$
- 平移translation:$x_0 + C = \{x_0 + x | x \in C\}$
- 投影projection:$\{x^1 | \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \in C\}$

凸集基本性质：投影定理

- 设$C \subset R^n$是非空闭凸集，$y \in R^n$但$y \notin C$, 则

（1）存在唯一的一点$\bar{x} \in C$, 使得$\bar{x}$是y到C的距离最小的点，即

$$\|\bar{x} - y\| = \inf\{\|x - y\| \,|x \in C\} > 0$$

（2）$\bar{x}$是y到C的距离最小的点的充要条件是

$$(x - \bar{x})^T(\bar{x} - y) \geq 0, \forall x \in C$$

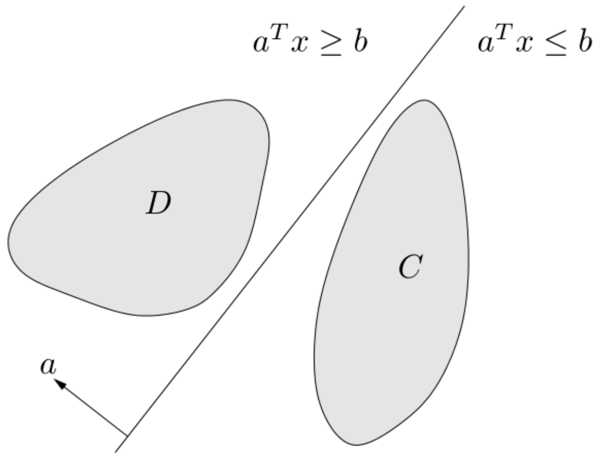
分割超平面定理

➢超平面分割定理Separating hyperplane theorem:

设C, D是两个非空凸集，且$C \cap D = \emptyset$, 则存在非零$\alpha \in R^n$和b使

$$\alpha^T x \leq b, \forall x \in C, \alpha^T y \geq b, \forall y \in D$$

则称超平面$H = \{x|\alpha^T x = b\}$分割集合C和 D



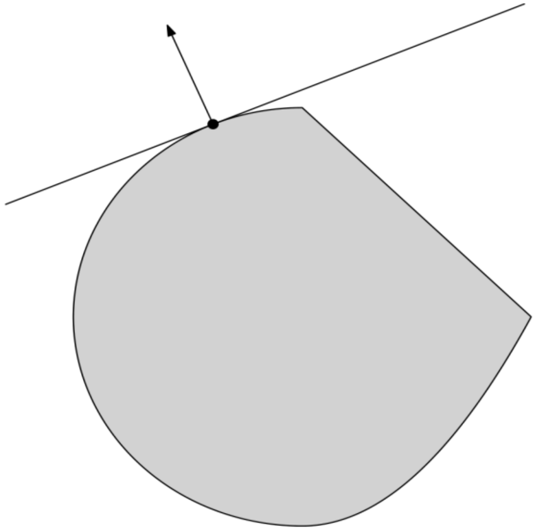
支撑超平面定理

- 支撑超平面定理Supporting Hyperplane Theorem:
- 设$C \in R^n$是非空凸集，$\bar{x} \in \partial C$, 则存在非零向量$\alpha \in R^n$使得

$$\alpha^T x \le \alpha^T \bar{x}, \forall x \in cl\ C$$

此时，称超平面$H = \{x \in R^n | \alpha^T x = \alpha^T \bar{x}\}$是集合$C$在$\bar{x}$处的支撑超平面。

∂C表示边界点集合；$intC$表示内部点集合；$cl\ C$ 表示包括内部点和边界点的集合。



参考文献

- 《运筹学与最优化问题》第二章
- Convex Optimization, Stephen Boyd and Lieven Vandenberghe, Cambridge University Press, 2016. Chapters 2 and 3
- J.P. Hiriart-Urruty and C. Lemarechal (1993), "Fundamentals of convex analysis", Chapters A and B
- R. T. Rockafellar (1970), "Convex analysis", Chapters 1 - 10,
- 最优化基础理论与方法 章节1.4
- 网课：最优化基础理论与方法