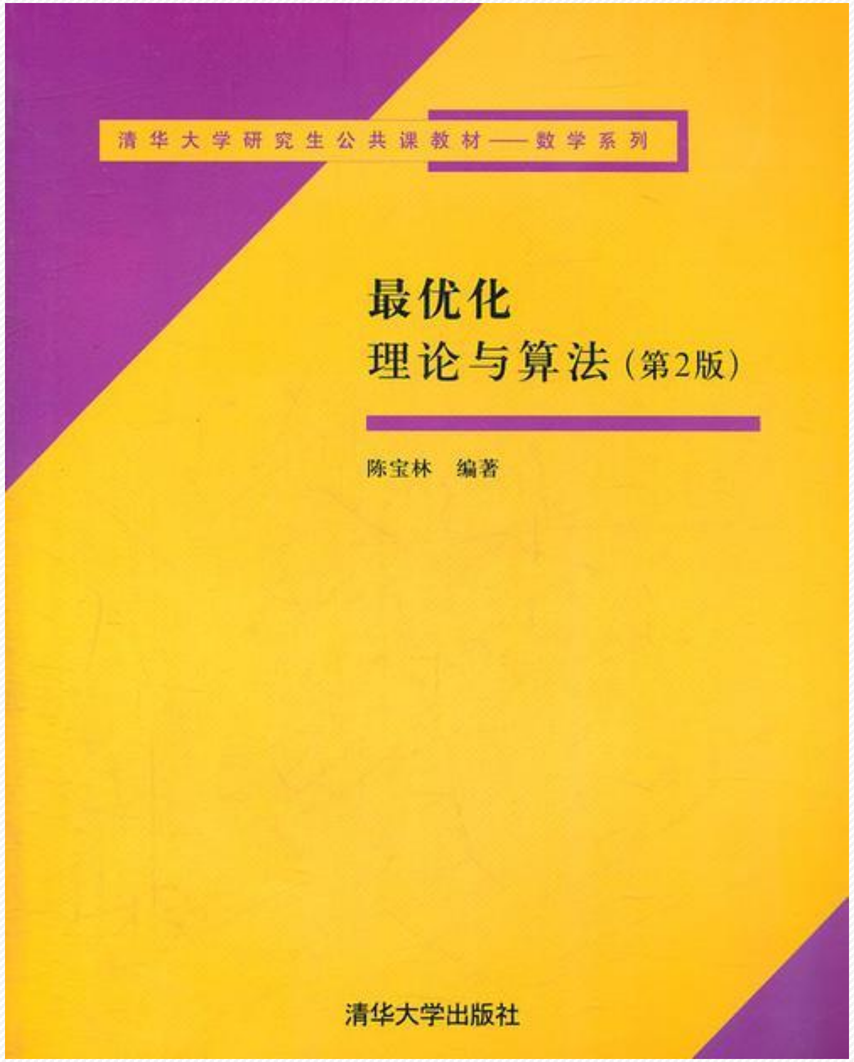
# 第1章 引言

李政伟

中国矿业大学 计算机科学与技术学院

# 参考书目

■陈宝林，最优化理论与算法（第2版），清华大学出版社



# 目 录

# 1.1 学科概述

- **最优化理论与算法**：是一个重要的数学分支。
- **最优化**：从所有可能的方案中选择最合理的一种方案，构造寻求最优解的计算方法。
- **最优方案**：达到最佳目标的方案。
- **最优化方法(算法)**：寻找最优方案的方法。
- **最优化理论**：最优化方法的数学理论。

# 最优化理论与算法简介

- 17世纪，在微积分时代已经出现了极值问题，后来又出现了Lagrange乘法;
- 1847年，法国数学家Cauchy提出了最速下降法;
- 19世纪40年代，第二次世界大战期间，优化问题在战争中得到了充分的运用和发展;
- 20世纪40年代以来，优化问题成为一门独立的学科，出现了线性规划、整数规划、非线性规划、几何规划等许多分支。

# 1.2 线性规划与非线性规划问题

## 例1 生产计划问题

设某工厂用4种资源生产3种产品，每单位第$j$种产品需要第$i$种资源的数量为$a_{ij}$，可获利润为$c_j$，第$i$种资源总消耗量不能超过$b_i$，由于市场限制，第$j$种产品的产量不超过$d_j$，试问如何安排生产才能使总利润最大?

**解析**: （1）决策变量 $x_j(j=1,2,3)$ 为第$j$种产品的产量

（2）目标函数：总利润 $f = c_1x_1 + c_2x_2 + c_3x_3 = (c_1, c_2, c_2)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \boldsymbol{c}^T\boldsymbol{x}$

（3）约束条件

资源限制

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \le b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \le b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \le b_3 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 \le b_4 \end{array}\right\} \quad \sum_{j=1}^{3} a_{ij}x_j \le b_i \quad i=1,2,3,4$$

# 例1 生产计划问题（续）

资源限制

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \boldsymbol{Ax} \leq \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

市场销量限制

$$x_j < d_j \qquad j = 1,2,3$$

产量非负限制

$$x_j \geq 0 \qquad j = 1,2,3$$

# 例1 生产计划问题（续）

在一组约束条件下，确定一个最优生产方案$x^* = (x_1^*, x_2^*, x_3^*)$，使目标函数值最大。数学模型如下：

$$\max \sum_{j=1}^{3} c_j x_j$$

$$\text{s.t.} \sum_{j=1}^{3} a_{ij} x_j \le b_j, \qquad i = 1,2,3,4,$$

$$x_j < d_j \qquad j = 1,2,3,$$

$$x_j \ge 0 \qquad j = 1,2,3.$$

# 例2 食谱问题

设市场上可买到$n$种不同的食品，第$j$种食品单位售价为$c_j$，每种食品含有$m$种基本营养成分，第$j$种食品每个单位含第$i$种营养成分为$a_{ij}$，又设每人每天对第$i$种营养成分的需要量不少于$b_i$，试确定在保证营养要求条件下的最经济食谱。

**解析**：（1）决策变量：$x_j(j=1,\dots,n)$为第$j$种产品的产量

（2）目标函数：

$$\min\sum_{j=1}^{n}c_jx_j$$

（3）约束条件：

$$\text{s.t.}\ \sum_{j=1}^{n}a_{ij}x_j\geq b_i,\qquad i=1,\dots,m$$

$$x_j\geq 0\qquad j=1,\dots,n$$

# 例3 结构设计问题

以两个构件组成的对称桁架为例。已知桁架的跨度$2L$，高度$x_2$的上限$H$，承受负荷$2P$，钢管的厚度$T$，材料比重$\rho$，纵向弹性模量$E$及容许应力$\sigma_y$ 。试确定钢管的平均直径$x_1$及桁架的高度$x_2$，使桁架的重量最小。

**解析** 对应的数学模型为：

$$\min 2\pi\rho T x_1\sqrt{L^2+x_2^2}$$

$$\text{s.t.}\quad x_2 \le H,$$

$$\frac{P\sqrt{L^2+x_2^2}}{\pi T x_1 x_2} \le \sigma_y,$$

$$\frac{P\sqrt{L^2+x_2^2}}{\pi T x_1 x_2} \le \frac{\pi^2 E(T^2+x_1^2)}{8(L^2+x_2^2)},$$

$$x_1, x_2 \ge 0.$$

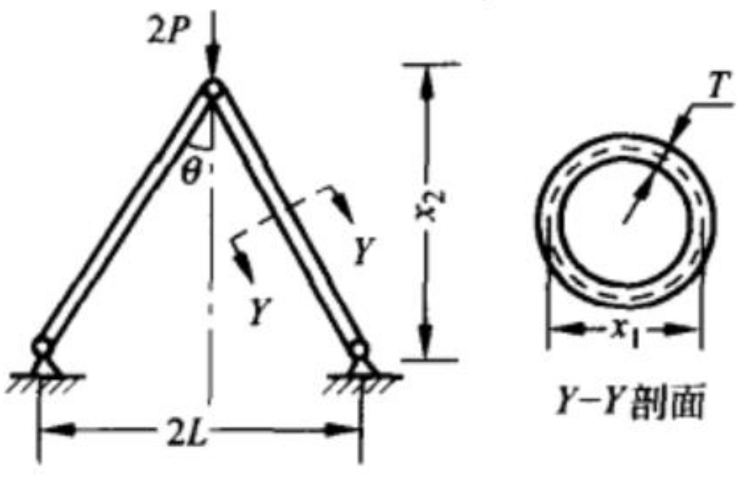



图 1.2.1

# 最优化问题的模型及分类

## 一、最优化问题的模型

$$\begin{cases} \min f(\boldsymbol{x}) \\ s.t.\quad g_i(\boldsymbol{x}) \le 0 \quad i = 1,2,\dots,m_i \\ \qquad\quad h_j(\boldsymbol{x}) = 0 \quad j = 1,2,\dots,m_j \\ \qquad\qquad x \in X \quad X \subset R^n \end{cases}$$

决策变量

目标函数

不等式约束

等式约束

箱子约束

**最优化问题分类**

- **线性规划问题**：模型中的目标函数和约束函数都是**线性**的；
- **非线性规划问题**：模型中含有**非线性**函数。

# 基本概念

**可行点**：满足约束条件的点。

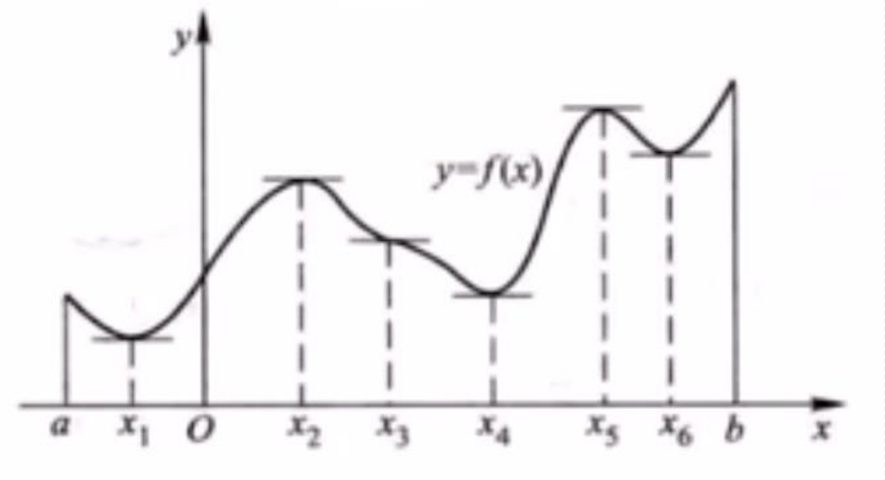
**可行集（可行域）**：全体可行点组成的集合。

$$F = \{\boldsymbol{x} | g_i(\boldsymbol{x}) \le 0 \quad i = 1,2,\dots,m_i, h_j(x) = 0 \quad j = 1,2,\dots,m_j, \boldsymbol{x} \in X\}$$

**无约束问题**：如果一个问题的可行集是整个空间。

**定义1.2.1** 设 $f(\boldsymbol{x})$为目标函数，$S$为可行域，$\overline{\boldsymbol{x}} \in S$，若对每个$\boldsymbol{x} \in S$，成立$f(\boldsymbol{x}) \ge f(\overline{\boldsymbol{x}})$，则称$\overline{\boldsymbol{x}}$为$f(\boldsymbol{x})$在$S$上的**全局极小点**。

**定义1.2.2** 设$f(\boldsymbol{x})$为目标函数，$S$为可行域，若存在$\overline{\boldsymbol{x}} \in S$的$\varepsilon > 0$邻域$N(\overline{\boldsymbol{x}}, \varepsilon) = \{\boldsymbol{x} | \, |\boldsymbol{x} - \overline{\boldsymbol{x}}| | < \varepsilon\}$，使得对每个$x \in S \cap N(\overline{\boldsymbol{x}}, \boldsymbol{\varepsilon})$成立$f(\boldsymbol{x}) \ge f(\overline{\boldsymbol{x}})$，则称$\overline{\boldsymbol{x}}$为$f(\boldsymbol{x})$在$S$上的一个**局部极小点**。




**注**：全局极小点也是局部极小点，而局部极小点不一定是全局极小点。

# 机器学习中的优化问题

（1） $L_2$范数正则

$$\min L(\boldsymbol{w}) = \frac{1}{N}\sum_{i=1}^{N}(f(x_i;\boldsymbol{w}) - y_i)^2 + \frac{\lambda}{2}\|\boldsymbol{w}\|_2^2$$

（2） $L_1$范数正则

$$\min L(\boldsymbol{w}) = \frac{1}{N}\sum_{i=1}^{N}(f(x_i;\boldsymbol{w}) - y_i)^2 + \frac{\lambda}{2}\|\boldsymbol{w}\|$$

（3） 感知机模型

$$\min L(\boldsymbol{w},\boldsymbol{b}) = -\sum_{i=1}^{N} y_i(\boldsymbol{w}x_i + \boldsymbol{b})$$

# 1.3 几个数学概念

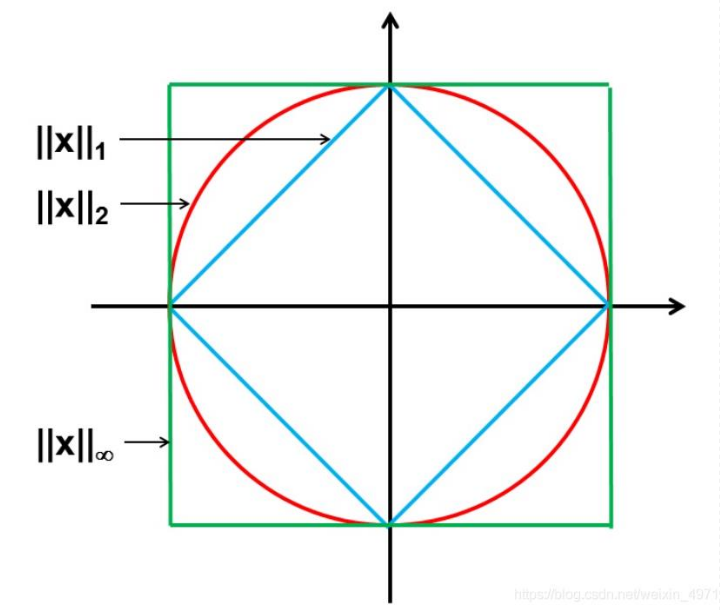
**定义1.3.1** 若实值函数$\|\cdot\|:\mathbb{R}^n \to \mathbb{R}$满足下面条件：

（1）正定型：$\|\boldsymbol{x}\| \geq 0, \forall \boldsymbol{x} \in \mathbb{R}^n; \|\boldsymbol{x}\| = 0$当且仅当$\boldsymbol{x} = 0$;

（2）齐次性：$\|\alpha \boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|, \forall \alpha \in \mathbb{R}, \boldsymbol{x} \in \mathbb{R}^n$

（2）三角不等式：$\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

则称 $\|\cdot\|$为$R^n$上的**向量范数**。




**常见范数**：设$\boldsymbol{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$

$L_1$范数 $$\|\boldsymbol{x}\|_1 = \sum_{j=1}^{n} |x_j|$$

$L_2$范数 $$\|\boldsymbol{x}\|_2 = \left(\sum_{j=1}^{n} x_j^2\right)^{\frac{1}{2}}$$

$L_\infty$范数 $$\|\boldsymbol{x}\|_\infty = \max_j |x_j|$$

$L_p$范数 $$\|\boldsymbol{x}\|_p = \left(\sum_{j=1}^{n} |x_j|^p\right)^{\frac{1}{p}}$$

# 1.3.1 向量范数和矩阵范数

**定义1.3.2** 设 $\|\cdot\|_\alpha$和$\|\cdot\|_\beta$是$\mathbb{R}^n$上任意两个范数，如果存在正数$c_1$和$c_2$，使得对每个$\boldsymbol{x} \in \mathbb{R}^n$成立$c_1\|\boldsymbol{x}\|_\alpha \le \|\boldsymbol{x}\|_\beta \le c_2\|\boldsymbol{x}\|_\alpha$，则称范数$\|\boldsymbol{x}\|_\alpha$和范数$\|\boldsymbol{x}\|_\beta$等价。
在$\mathbb{R}^n$中任何两种向量范数都是等价的。

**定义1.3.3** 设 $\boldsymbol{A}$是$\mathrm{m} \times n$矩阵，$\|\cdot\|_\alpha$是$\mathbb{R}^m$上的范数，$\|\cdot\|_\beta$是$\mathbb{R}^n$上的范数，则定义矩阵范数$\|\boldsymbol{A}\| = \max\limits_{\|\cdot\|_\beta=\mathbf{1}} \|\boldsymbol{Ax}\|_\alpha$。

若对于任意$\boldsymbol{A} \in \mathbb{R}^{n\times n}$，都有一个矩阵范数$\|\boldsymbol{A}\|$与之对应，且满足：
（1）非负性：$\|\boldsymbol{A}\| \ge \mathbf{0}$，当且仅当$\boldsymbol{A} = \mathbf{0}$时，$\|\boldsymbol{A}\| = \mathbf{0}$
（2）齐次性：$\|\boldsymbol{\lambda A}\| = |\boldsymbol{\lambda}|\|\boldsymbol{A}\|, \forall\boldsymbol{\lambda} \in \mathbb{R}$
（3）三角不等式：$\|\boldsymbol{A} + \boldsymbol{B}\| \le \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$
（4）相容性：$\|\boldsymbol{AD}\| \le \|\boldsymbol{A}\|\|\boldsymbol{D}\|, \boldsymbol{A}, \boldsymbol{D} \in \mathbb{R}^{n\times n}$

# 1.3.1 向量范数和矩阵范数

**常用矩阵范数：**

$$\|\boldsymbol{A}\|_1 = \max_j \sum_{i=1}^{m} |a_{ij}|$$

$$\|\boldsymbol{A}\|_2 = \sqrt{\boldsymbol{\lambda}_{\boldsymbol{A}^T\boldsymbol{A}}}$$

$\boldsymbol{\lambda}_{\boldsymbol{A}^T\boldsymbol{A}}$是$\boldsymbol{A}^T\boldsymbol{A}$的最大特征值。 $\|\boldsymbol{A}\|_2$又称为谱范数。

$$\|\boldsymbol{A}\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$$

# 1.3.2 序列的极限

**定义1.3.4** 设$\{\boldsymbol{x}^{(k)}\}$是$\mathbb{R}^n$中一个向量序列，$\bar{\boldsymbol{x}} \in \mathbb{R}^n$，如果对每个任给的$\varepsilon > 0$存在正整数$K_\varepsilon$，使得当$k > K_\varepsilon$，时就有$\|\boldsymbol{x}^{(k)} - \bar{\boldsymbol{x}}\| < \varepsilon$，则称序列收敛到$\bar{\boldsymbol{x}}$，或称序列以$\bar{\boldsymbol{x}}$为**极限**， 记作$\lim\limits_{k\to\infty} \boldsymbol{x}^{(k)} = \bar{\boldsymbol{x}}$。

**注意**：序列极限若存在，则必定唯一。

**定义1.3.4** 设$\{\boldsymbol{x}^{(k)}\}$是$\mathbb{R}^n$中一个向量序列，若存在一个子序列$\{\boldsymbol{x}^{(k_j)}\}$，使$\lim\limits_{k\to\infty} \boldsymbol{x}^{(k_j)} = \hat{\boldsymbol{x}}$，则称$\hat{\boldsymbol{x}}$是序列$\{\boldsymbol{x}^{(k)}\}$的一个**聚点**。

$\{1,-1,1,-1,1,-1,1,-1,\ldots.,\}$振荡，没有**极限**，但是有**聚点**。

**定义1.3.6** 设$\{\boldsymbol{x}^{(k)}\}$是$\mathbb{R}^n$中一个向量序列，如果对任意给定的$\varepsilon > 0$，总存在正整数$K_\varepsilon$，使得当$m, l > K_\varepsilon$时，就有$\|\boldsymbol{x}^{(m)} - \boldsymbol{x}^{(l)}\| < \varepsilon$，则$\{\boldsymbol{x}^{(k)}\}$称为**Cauchy序列**。

设集合$S$是$\mathbb{R}^n$中的一个集合，如果$S$中的每个收敛序列的极限均属于$S$，则$S$为**闭集**。
如果对每个点$\hat{x} \in S$，存在正数$\varepsilon$，使得$\hat{x}$的$\varepsilon$邻域$N(\hat{x}, \varepsilon) = \{x \mid \|x - \hat{x}\| < \varepsilon\} \subset S$，则称$S$为**开集**。
如果$S$是**有界闭集**，则称S为**紧集**。

## 1.3.3 梯度、Hesse矩阵、Taylor展开式

设集合$S \subset \mathbb{R}^n$非空，$f(\boldsymbol{x})$为定义在$S$上的实函数。如果$f$在每一点$\boldsymbol{x} \in S$连续，则称$f$在$S$上连续，记作$f \in C(S)$。再设$S$为开集，如果在每一点$\boldsymbol{x} \in S$，对所有$j = 1, \dots, n$，偏导数$\frac{\partial f(x)}{\partial x_j}$存在且连续，则称$f$在开集$S$上连续可微，记作$f \in C^1(S)$。

如果在每一点$\boldsymbol{x} \in S$，对所有$i = 1, \dots, n$和$j = 1, \dots, n$，二阶偏导数$\frac{\partial f(x)}{\partial x_i \partial x_j}$存在且连续，则称$f$在开集$S$上二次连续可微，记作$f \in C^2(S)$。

函数$f$在$\boldsymbol{x}$处的梯度为$n$维列向量:

$$\boldsymbol{\nabla} f(\boldsymbol{x}) = \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \dots, \frac{\partial f(\boldsymbol{x})}{\partial x_n}\right]^T,$$

$f$在$\boldsymbol{x}$处的Hesse矩阵为$n \times n$矩阵$\boldsymbol{\nabla}^2 f(\boldsymbol{x})$，第$i$行第$j$列元素为

$$[\boldsymbol{\nabla}^2 f(\boldsymbol{x})]_{ij} = \frac{\partial f(\boldsymbol{x})}{\partial x_i \partial x_j}, \quad 1 \le i, j \le n$$

# 1.3.3 梯度、Hesse矩阵、Taylor展开式

**例4**，$f(\boldsymbol{x}) = \boldsymbol{c}^T\boldsymbol{x}, \boldsymbol{c} \in \mathbb{R}^n, \boldsymbol{x} \in \mathbb{R}^n$，计算$\nabla f(\boldsymbol{x})$。

**解：** $f(\boldsymbol{x}) = c_1x_1 + c_2x_2 + \cdots + c_nx_n$

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \boldsymbol{c}$$

**例5**，$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c, \boldsymbol{A}^T = \boldsymbol{A}, b \in \mathbb{R}^n, c \in \mathbb{R}$，计算$\nabla f(\boldsymbol{x})$、$\nabla^2 f(\boldsymbol{x})$。

**解：**

$$\nabla f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$$

$$\nabla^2 f(\boldsymbol{x}) = \boldsymbol{A}$$

## 1.3.3 梯度、Hesse矩阵、Taylor展开式

$f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x}$，其中$\boldsymbol{x} \in \mathbb{R}^n$， $\boldsymbol{A} \in \mathbb{R}^{n\times n}$，则$\nabla f(\boldsymbol{x}) = (\boldsymbol{A}^T + \boldsymbol{A})\boldsymbol{x}$

推导1：

$$f(\boldsymbol{x}) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$$

$$\frac{\partial f(\boldsymbol{x})}{\partial x_k} = \sum_{j=1}^{n} a_{kj}x_j + \sum_{i=1}^{n} a_{ik}x_i$$

$$\begin{aligned}\nabla f(\boldsymbol{x}) &= \boldsymbol{x}^T\boldsymbol{A}^T + \boldsymbol{x}^T\boldsymbol{A} \\ &= (\boldsymbol{A}^T + \boldsymbol{A})\boldsymbol{x}\end{aligned}$$

推导2：

$$\begin{aligned}d(f(\boldsymbol{x})) &= (\boldsymbol{x}^T\boldsymbol{A})d\boldsymbol{x} + d(\boldsymbol{x}^T\boldsymbol{A})\boldsymbol{x} \\ &= \boldsymbol{x}^T\boldsymbol{A}d\boldsymbol{x} + \boldsymbol{x}^T d(\boldsymbol{A}^T\boldsymbol{x}) \\ &= \boldsymbol{x}^T\boldsymbol{A}d\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{A}^T d\boldsymbol{x} \\ &= \boldsymbol{x}^T(\boldsymbol{A}^T + \boldsymbol{A})\boldsymbol{dx} \\ &= (\boldsymbol{A}^T + \boldsymbol{A})\boldsymbol{x}\boldsymbol{dx}\end{aligned}$$

$$\nabla f(\boldsymbol{x}) = (\boldsymbol{A}^T + \boldsymbol{A})\boldsymbol{x}$$

# 1.3.3 梯度、Hesse矩阵、Taylor展开式

$f: \mathbb{R} \to \mathbb{R}$

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0)$$

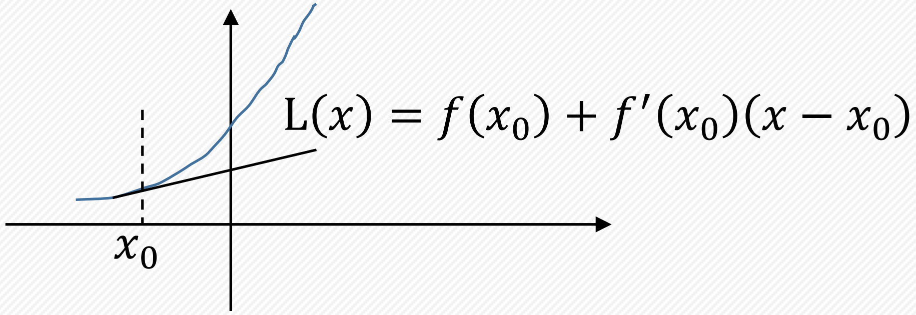
$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + o((x - x_0)^2)$$

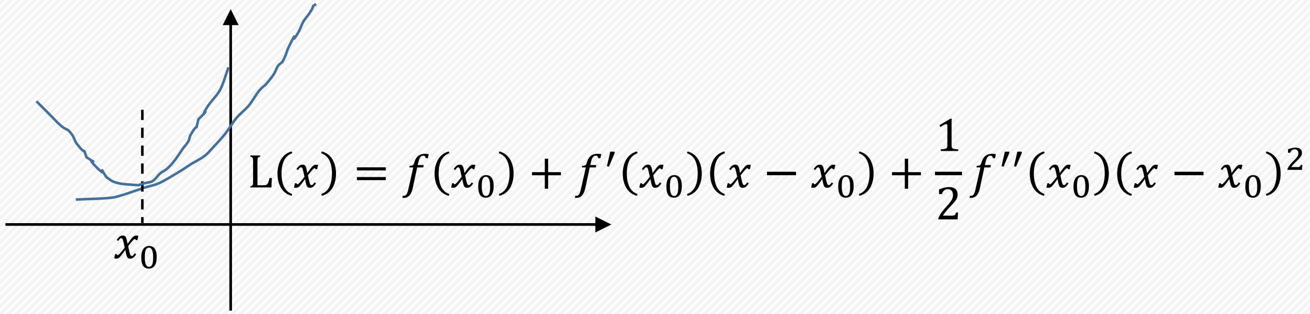
假设在开集$S \subset \mathbb{R}^n$上$f \in C^1(S)$，给定点$\boldsymbol{x} \in S$，则$f$在点$\bar{\boldsymbol{x}}$的一阶Taylor展开式为

$$f(\boldsymbol{x}) \ = \ f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T (\boldsymbol{x} - \bar{\boldsymbol{x}}) + o(\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|),$$

其中$o(\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|)$当$\|\boldsymbol{x} - \bar{\boldsymbol{x}}\| \to \mathbf{0}$时，关于$\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|$是高阶无穷小。
假设在开集$S \subset R^n$上$f \in C^2(S)$，则$f$在$\boldsymbol{x} \in S$的二阶Taylor展开式为

$$f(\boldsymbol{x}) \ = \ f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T (\boldsymbol{x} - \bar{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^T \nabla^2 f(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}) + o(\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|^2),$$



# 1.3.4Jacobi矩阵、链式法则和隐函数存在定理

## 1. Jacobi 矩阵

考虑向量值函数$\boldsymbol{h}(\boldsymbol{x}) = (h_1(\boldsymbol{x}),\ h_2(\boldsymbol{x}),\ \ldots, h_m(\boldsymbol{x}))^T$，

其中每个分量$h_i(\boldsymbol{x})$为$n$元实值函数，假设对所有$i,j$偏导数$\frac{\partial h_i(\boldsymbol{x})}{\partial x_j}$存在。$\boldsymbol{h}$在点$\boldsymbol{x}$的Jacobi矩阵为

$$\begin{bmatrix} \frac{\partial h_1(\boldsymbol{x})}{\partial x_1} & \frac{\partial h_1(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial h_1(\boldsymbol{x})}{\partial x_n} \\ \frac{\partial h_2(\boldsymbol{x})}{\partial x_1} & \frac{\partial h_2(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial h_2(\boldsymbol{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_m(\boldsymbol{x})}{\partial x_1} & \frac{\partial h_m(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial h_m(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$$

这个矩阵称为向量值函数$\boldsymbol{h}$在$\boldsymbol{x}$的导数，记作$\boldsymbol{h}'(\boldsymbol{x})$或$\nabla h(\boldsymbol{x})^T$，其中$\boldsymbol{\nabla h}(\boldsymbol{x}) = (\boldsymbol{\nabla} h_1(\boldsymbol{x}), \boldsymbol{\nabla} h_2(\boldsymbol{x}), \ldots, \boldsymbol{\nabla} h_m(\boldsymbol{x}))$

# 1. Jacobi 矩阵

**例6** 设有向量值函数

$$f(\boldsymbol{x}) = f(x_1, x_2) = \begin{bmatrix} sinx_1 + cosx_2 \\ e^{2x_1+x_2} \\ 2x_1^2 + x_1x_2 \end{bmatrix}$$

则$\boldsymbol{f}(\boldsymbol{x})$在任一点$(x_1, x_2)$的Jacobi矩阵，即导数为

$$f'(\boldsymbol{x}) = \begin{bmatrix} cosx_1 & -sinx_2 \\ 2e^{2x_1+x_2} & e^{2x_1+x_2} \\ 4x_1 + x_2 & x_1 \end{bmatrix}$$

# 2.链式法则

设有复合函数$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x}))$，其中向量值函数$\boldsymbol{f}(\boldsymbol{g})$和$\boldsymbol{g}(\boldsymbol{x})$均可微，$\boldsymbol{x} \in D^n \subset \mathbb{R}^n$，$\boldsymbol{g}: D^n \to D_1^m, \boldsymbol{f}: D_2^m \to \mathbb{R}^k$，其中$D_1^m \subset D_2^m$，$\boldsymbol{h}: D^n \to \mathbb{R}^k$。

根据复合函数求导数的链式法则，必有

$$\boldsymbol{h}'(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{g}(\boldsymbol{x}))\boldsymbol{g}'(\boldsymbol{x}), \boldsymbol{x} \in D^n \qquad (1.3.4)$$

其中$\boldsymbol{f}'$和$\boldsymbol{g}'$分别为$k \times m$和$m \times n$矩阵，$\boldsymbol{h}'$为$k \times n$矩阵。

若记$\nabla \boldsymbol{f} = (\nabla \boldsymbol{f}, \nabla \boldsymbol{f}_2, \dots, \nabla \boldsymbol{f}_k)$，$\nabla \boldsymbol{g} = (\nabla \boldsymbol{g}_1, \nabla \boldsymbol{g}_2, \dots, \nabla \boldsymbol{g}_m)$，

由于$\boldsymbol{h}' = \nabla \boldsymbol{h}^T$，$\boldsymbol{f}' = \nabla \boldsymbol{f}^T$和$\boldsymbol{g}' = \nabla \boldsymbol{g}^T$，可将(1.3.4)式改写为

$$\nabla \boldsymbol{h}(\boldsymbol{x}) = \nabla \boldsymbol{g}(\boldsymbol{x}) \nabla \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x})) \qquad (1.3.5)$$

其中，$\nabla \boldsymbol{h}$为$n \times k$矩阵，第$j$列为$\nabla \boldsymbol{f}_j$的$h_j(\boldsymbol{x})$梯度。

# 2.链式法则

**例7** 设有复合函数$\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{f}(\boldsymbol{u}(\boldsymbol{x}))$，其中

$$\boldsymbol{f}(\boldsymbol{u})=\begin{bmatrix}\boldsymbol{f_1}(\boldsymbol{u})\\ \boldsymbol{f_2}(\boldsymbol{u})\end{bmatrix}=\begin{bmatrix}u_1^2-u_2\\ u_1+u_2^2\end{bmatrix}\qquad \mathbf{u}(\boldsymbol{x})=\begin{bmatrix}\boldsymbol{u_1}(\boldsymbol{x})\\ \boldsymbol{u_2}(\boldsymbol{x})\end{bmatrix}=\begin{bmatrix}x_1+x_2\\ x_2^2-x_3\end{bmatrix}$$

试求复合函数$\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{f}(\boldsymbol{u}(\boldsymbol{x}))$的导数.

**解**：$\boldsymbol{h}'(\boldsymbol{x})=\boldsymbol{f}'(\boldsymbol{u}(\boldsymbol{x}))\,\boldsymbol{u}'(\boldsymbol{x})$

$$=\begin{bmatrix}2u_1 & -1\\ 1 & 2u_2\end{bmatrix}\begin{bmatrix}1 & 0 & 1\\ 0 & 2x_2 & -1\end{bmatrix}=\begin{bmatrix}2u_1 & -2x_2 & 2u_1+1\\ 1 & 4u_2x_2 & 1-2u_2\end{bmatrix}$$

将$\boldsymbol{u}$代替$\boldsymbol{x}$，得到

$$\begin{bmatrix}\dfrac{\partial \boldsymbol{h}_1(\boldsymbol{x})}{\partial x_1} & \dfrac{\partial \boldsymbol{h}_1(\boldsymbol{x})}{\partial x_2} & \dfrac{\partial \boldsymbol{h}_1(\boldsymbol{x})}{\partial x_3}\\ \dfrac{\partial \boldsymbol{h}_2(\boldsymbol{x})}{\partial x_1} & \dfrac{\partial \boldsymbol{h}_2(\boldsymbol{x})}{\partial x_2} & \dfrac{\partial \boldsymbol{h}_2(\boldsymbol{x})}{\partial x_3}\end{bmatrix}=\begin{bmatrix}2(x_1+x_2) & -2x_2 & 2(x_1+x_2)+1\\ 1 & 4(x_2^2-x_3)x_2 & 1-2(x_2^2-x_3)\end{bmatrix}$$

# 1.4 凸集和凸函数

## 1.4.1 凸集

**定义1.4. 1** 设$S$为$n$维欧氏空间$\mathbb{R}^n$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$\boldsymbol{x}^{(1)}$，$\boldsymbol{x}^{(2)}$及每个实数$\lambda \in [0,1]$，都有

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

则称$S$为凸集。

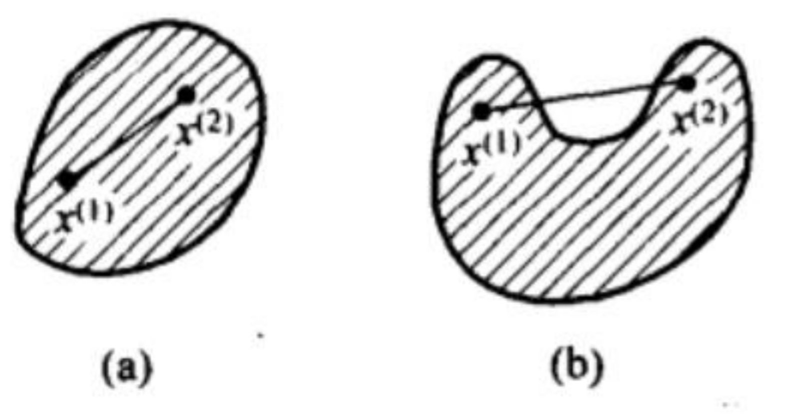
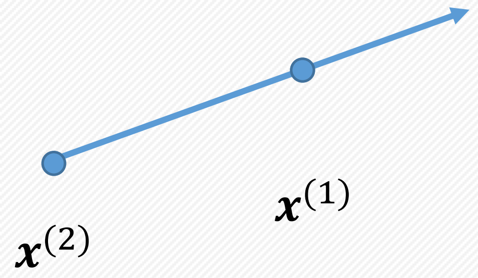


图 1.4.1




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

# 1.4.1 凸集

**例1.4.1** 验证集合$H=\{\boldsymbol{x}|\boldsymbol{p}^T\boldsymbol{x}=\alpha\}$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$\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}\in H$，有

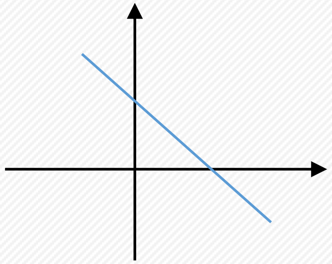
$$\boldsymbol{p}^T\boldsymbol{x}^{(1)}=\alpha,\ \boldsymbol{p}^T\boldsymbol{x}^{(2)}=\alpha$$

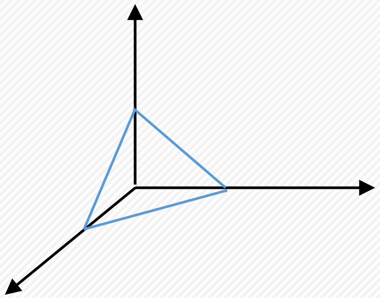
$\forall\lambda\in[0,1]$，有

$$\boldsymbol{p}^T[\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)}]=\lambda\boldsymbol{p}^T\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{p}^T\boldsymbol{x}^{(2)}=\alpha,$$

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

根据定义1.4.1知$H$为凸集。





集合$H$称为$\mathbb{R}^n$中的超平面，故**超平面**为凸集。

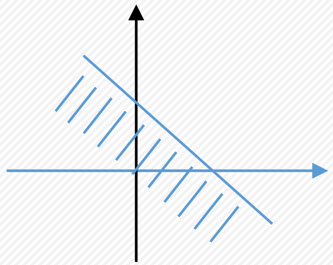
**例1.4.2** 验证集合$H^-=\{\boldsymbol{x}|\boldsymbol{p}^T\boldsymbol{x}\le\alpha\}$为凸集。

**解** 对任意的 $\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}\in H$及每一个实数$\lambda\in[0,1]$，都有

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

所以$\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)}\in H$。根据定义1.4.1知$H$为凸集。

集合$H=\{\boldsymbol{x}|\boldsymbol{p}^T\boldsymbol{x}\le \mathrm{a}\}$称为半空间，故**半空间**为凸集。



# 1.4.1 凸集

**例1.4.3** 验证集合$L=\{\boldsymbol{x}|\boldsymbol{x}=\boldsymbol{x}^{(0)}+\lambda\boldsymbol{d},\lambda\geq 0\}$为凸集，其中$\boldsymbol{d}$是给定的非零向量，$\boldsymbol{x}^{(0)}$是定点。

**解** 因为对$\forall\, \boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}\in L$及$\lambda\in[0,1]$，必有$\boldsymbol{x}^{(1)}=\boldsymbol{x}^{(0)}+\lambda_1 d$， $\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(0)}+\lambda_2 d$，$\lambda_1,\lambda_2\geq 0$，以及

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

由于$[\lambda_1+(1-\lambda)\lambda_2]\geq 0$，因此有$\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)}\in L$，根据定义1.4.1知$L$为凸集。

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

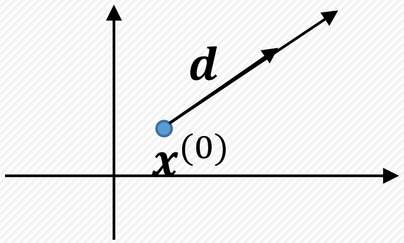
设$S_1$和$S_2$为$\mathbb{R}^n$中两个凸集，$\beta$是实数，则

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

(2) $S_1\cap S_2$为凸集；

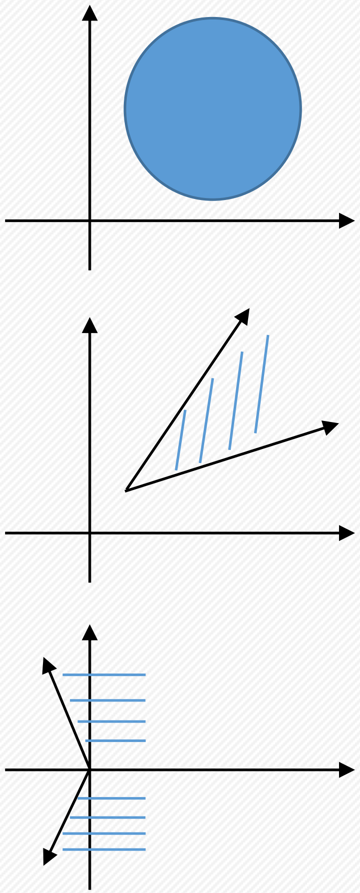
(3) $S_1+S_2=\{\boldsymbol{x}^{(1)}+\boldsymbol{x}^{(2)})|\boldsymbol{x}^{(1)}\in S_1,\boldsymbol{x}^{(2)}\in S_2\}$为凸集；

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



# 1.4.1 凸集

**定义1.4.2** 设有集合$C \subset \mathbb{R}^n$，若对$C$中每一点$\boldsymbol{x}$，当$\lambda$取任何非负数时，都有$\lambda\boldsymbol{x} \in C$，称$C$为**锥**，又若$C$为凸集，则称$C$为**凸锥**。



**例1.4.4** 向量集$\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \dots, \boldsymbol{\alpha}^{(k)}$的所有非负线性组合构成的集合

$$\{\sum_{i=1}^{k} \lambda_i \boldsymbol{\alpha}^{(i)} \mid \lambda_i \geq 0, i = 1, \dots, k\}$$

为凸锥.

# 1.4.1 凸集

**定义1.4.3** 有限个半空间的交

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

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

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \le b_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \le b_2$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \le b_m$$

**例1.4.5** 集合

$$S = \{\boldsymbol{x}|x_1 + 2x_2 \le 4, x_1 - x_2 \le 1, x_1 \ge 0, x_2 \ge 0\}$$

为多面集。

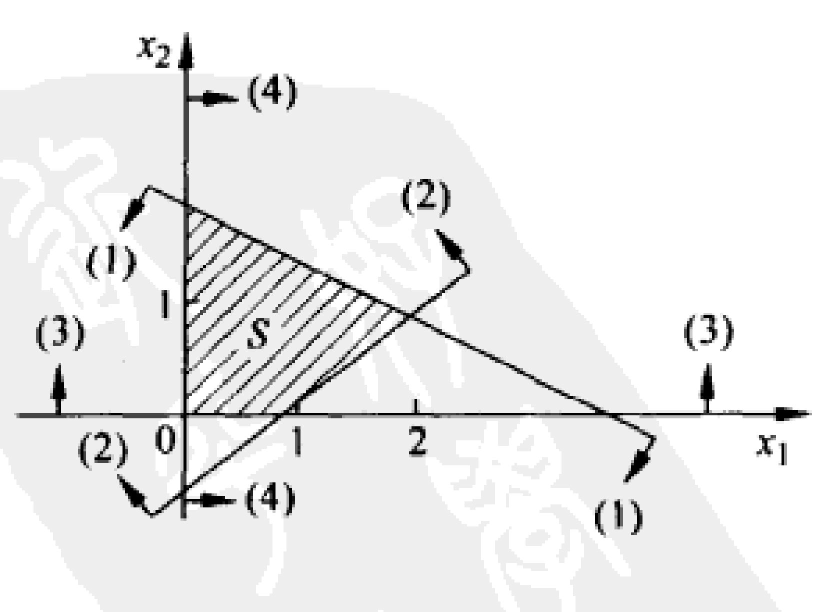
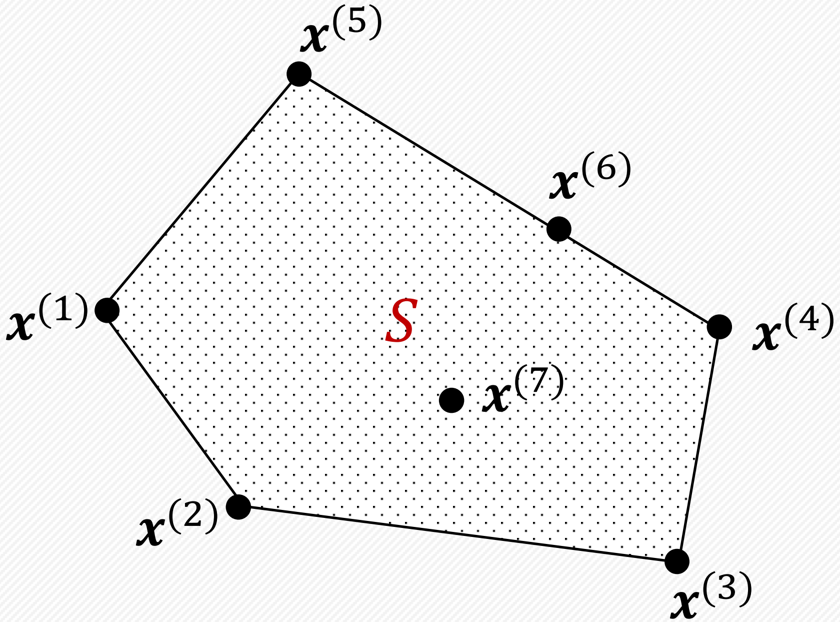


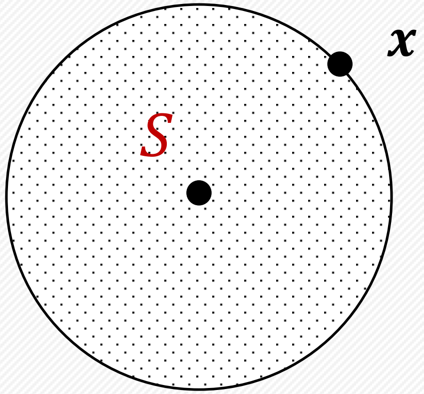
图 1.4.2

# 1.4.1 凸集

**定义1.4.4** 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$\boldsymbol{x} = \lambda \boldsymbol{x}^{(1)} + (1 - \lambda)\boldsymbol{x}^{(2)} (\lambda \in (0,1)), \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(1)} \in S$，必推得$\boldsymbol{x} = \boldsymbol{x}^{(1)} = \boldsymbol{x}^{(2)}$，则称$\boldsymbol{x}$是凸集$S$的**极点**。




(a)




(b)

**定理**：若$S$是紧凸集，则$S$中的任一点都可以表示成极点的凸组合。

# 1.4.1 凸集

**定义1.4.5** 设$S$为$\mathbb{R}^n$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
$\{\boldsymbol{x}+\lambda\boldsymbol{d} \mid \lambda \geq 0\} \subset S$，
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$\lambda$，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**。
显然，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1.4.6** 对于集合 $S=\{(x_1,x_2) \mid x_2 \geq |x_1|\}$，凡是与向量$(0,1)^T$夹角小于或等于45°的向量，都是它的方向。其中$(1,1)^T$和$(-1,1)^T$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。

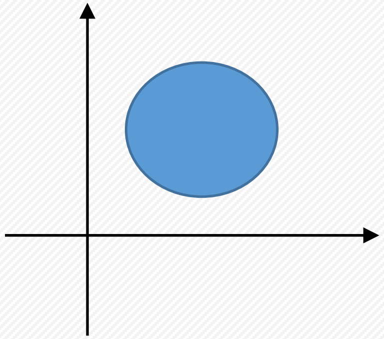
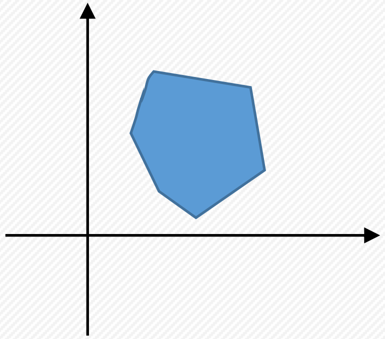
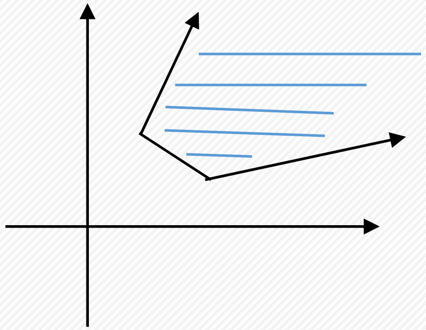
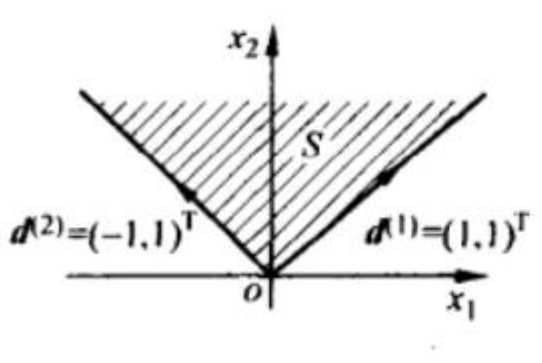





图 1.4.4

## 1.4.1 凸集

**例1.4.7** 设$S=\{\boldsymbol{x}|\ \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq 0\}$为非空集合，$\boldsymbol{d}$是非零向量。证明$\boldsymbol{d}$为$S$的方向的充要条件是$\boldsymbol{d}\geq 0$且$\boldsymbol{A}\boldsymbol{d}=0$.

**证明** 按照定 义，$\boldsymbol{d}$为$S$的方向的充要条件是：对每一个$\boldsymbol{x}\in S$,有

$\{\boldsymbol{x}+\lambda\boldsymbol{d}|\lambda\geq 0\}\subset S$.　(1.4.1)

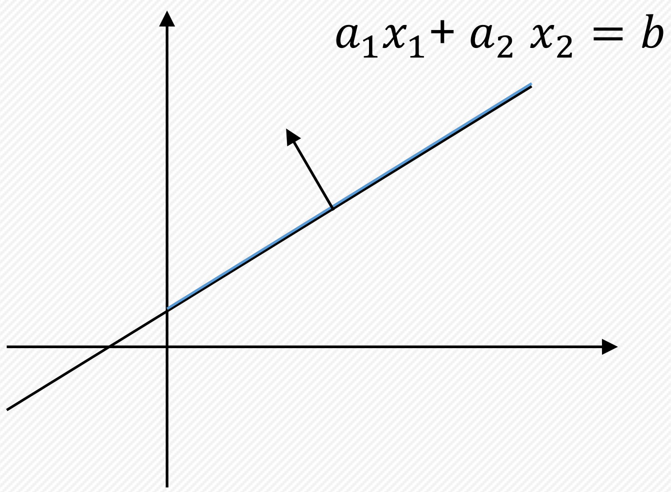
根据集合$S$的定义，(1.4.1)式即

$\boldsymbol{A}(\boldsymbol{x}+\lambda\boldsymbol{d})\ =\ \boldsymbol{b}$,　　(1.4.2).

$\boldsymbol{x}+\lambda\boldsymbol{d}\geq\boldsymbol{0}$.　　(1.4.3)

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

因此由(1.4.2)式和(1.4.3)式知$\boldsymbol{A}\boldsymbol{d}=0$及$\boldsymbol{d}\geq 0$。



# 1.4.1 凸集

**定理1.4.1 (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$\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)},\cdots,\boldsymbol{x}^{(k)}$.

(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$\boldsymbol{d}^{(1)},\boldsymbol{d}^{(2)},\cdots,\boldsymbol{d}^{(k)}$。

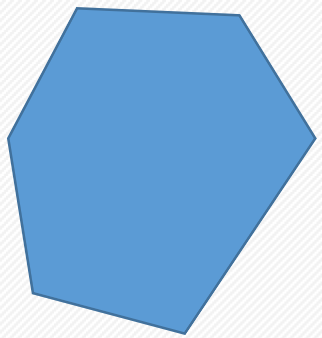
(3)$\boldsymbol{x}\in S$的充要条件是:

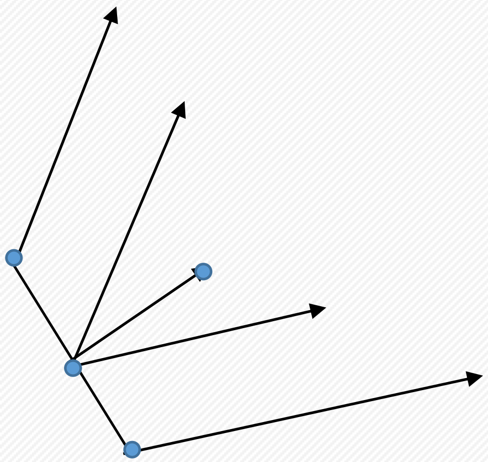
$$\boldsymbol{x}=\sum_{j=1}^{k}\lambda_j\boldsymbol{x}^{(j)}+\sum_{j=1}^{l}\mu_j\boldsymbol{d}^{(j)}$$

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

$$\lambda_j\geq 0,\qquad j=1,\dots,k$$

$$\mu_j\geq 0,\qquad j=1,\dots,l$$





# 1.4.2 凸集分离定理

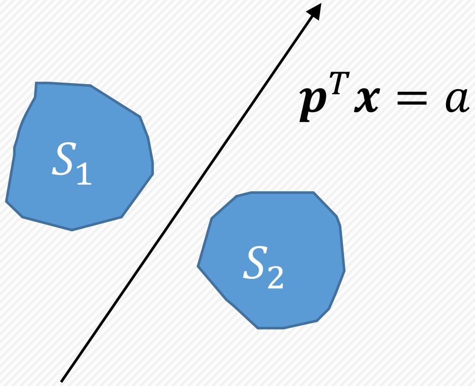
**定义1.4.6** 设$S_1$和$S_2$是$\mathbb{R}^n$中两个非空集合，$H=\{\boldsymbol{x}|\boldsymbol{p}^T\boldsymbol{x}=a\}$为超平面。如果对每个$\boldsymbol{x}\in S_1$，都有$\boldsymbol{p}^T\boldsymbol{x}\geq a$，对于每个$\boldsymbol{x}\in S_2$，都有$\boldsymbol{p}^T\boldsymbol{x}\leq a$(或情形恰好相反)，则称超平面$H$**分离**集合$S_1$和$S_2$。

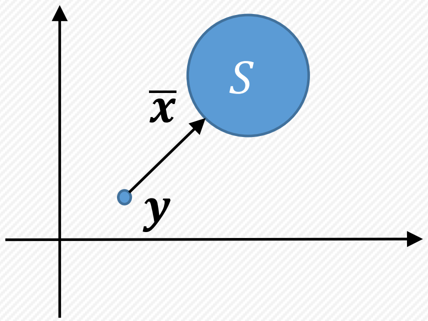
**定理1.4.2** 设 $S$为$\mathbb{R}^n$中的**闭凸集**，$\boldsymbol{y}\notin S$，则存在**惟一的点**$\bar{\boldsymbol{x}}\in S$，使得

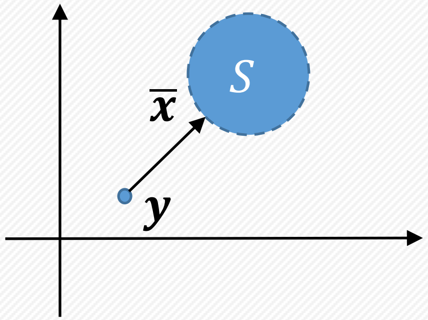
$$\|\boldsymbol{y}-\bar{\boldsymbol{x}}\|=\inf_{x\in s}\|\boldsymbol{y}-\boldsymbol{x}\|$$

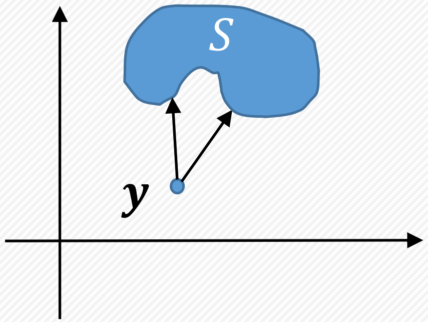
$$\inf_{x\in s} x \quad x\in(0,1)$$

$$\min_{x\in s} x \quad x\in(0,1)$$



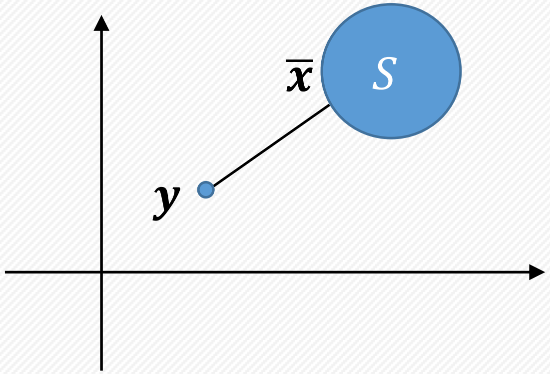



# 1.4.2 凸集分离定理

**定理1.4.3** 设 $S$是$\mathbb{R}^n$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$\boldsymbol{p}^T\boldsymbol{y} \geq \varepsilon + \boldsymbol{p}^T\boldsymbol{x}$。




**证明**：由于$S$是闭凸集，$\boldsymbol{y} \notin S$，则由定理1.4.2知，存在$\overline{\boldsymbol{x}} \in S$，使

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

令$\boldsymbol{p} = \boldsymbol{y} - \overline{\boldsymbol{x}}$,$\varepsilon = \boldsymbol{p}^T(\boldsymbol{y} - \overline{\boldsymbol{x}})$

由于 
$$\begin{aligned}\boldsymbol{p}^T(\boldsymbol{y} - \boldsymbol{x}) &= \boldsymbol{p}^T(\boldsymbol{y} - \overline{\boldsymbol{x}} + \overline{\boldsymbol{x}} - \boldsymbol{x}) \\ &= \boldsymbol{p}^T(\boldsymbol{y} - \overline{\boldsymbol{x}}) + \boldsymbol{p}^T(\overline{\boldsymbol{x}} - \boldsymbol{x}) \\ &= \varepsilon + (\boldsymbol{y} - \overline{\boldsymbol{x}})^T(\overline{\boldsymbol{x}} - \boldsymbol{x})\end{aligned}$$

$$\begin{aligned}\|\boldsymbol{y} - \overline{\boldsymbol{x}}\|^2 &\leq \|\boldsymbol{y} - [\lambda\boldsymbol{x} + (1-\lambda)\overline{\boldsymbol{x}}]\|^2 \\ &= \|(\boldsymbol{y} - \overline{\boldsymbol{x}}) + \lambda(\boldsymbol{x} - \overline{\boldsymbol{x}})]\|^2 \\ &= \|\boldsymbol{y} - \overline{\boldsymbol{x}}\|^2 + \lambda^2\|\boldsymbol{x} - \overline{\boldsymbol{x}}\|^2 + 2\lambda(\boldsymbol{y} - \overline{\boldsymbol{x}})^T(\overline{\boldsymbol{x}} - \boldsymbol{x})\end{aligned}$$

所以 $(\boldsymbol{y} - \overline{\boldsymbol{x}})^T(\overline{\boldsymbol{x}} - \boldsymbol{x}) + \dfrac{\lambda}{2}\|\boldsymbol{x} - \overline{\boldsymbol{x}}\|^2 \geq 0$

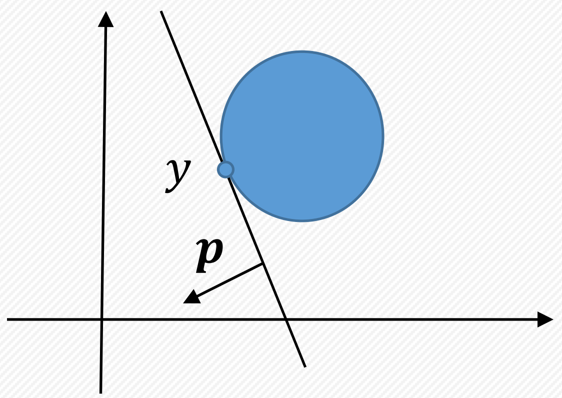
令$\lambda \to 0$，则 $(\boldsymbol{y} - \overline{\boldsymbol{x}})^T(\overline{\boldsymbol{x}} - \boldsymbol{x}) \geq \boldsymbol{0}$

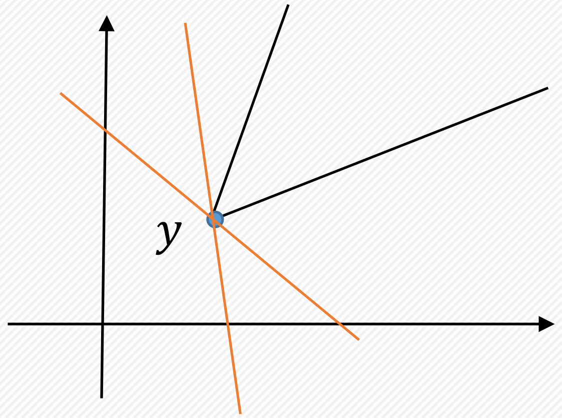
所以 $\boldsymbol{p}^T(\boldsymbol{y} - \boldsymbol{x}) \geq \varepsilon$

即 $\boldsymbol{p}^T\boldsymbol{y} \geq \varepsilon + \boldsymbol{p}^T\boldsymbol{x}$

# 1.4.2 凸集分离定理

**定理1.4.4** 设$S$是$\mathbb{R}^n$中一个非空凸集，$\boldsymbol{y} \in \partial S$，则存在非零向量$\boldsymbol{p}$，使得对每一点$\boldsymbol{x} \in clS$，有$\boldsymbol{p}^T\boldsymbol{y} \geq \boldsymbol{p}^T\boldsymbol{x}$成立。




**推论** 设$S$是$\mathbb{R}^n$中的非空凸集，$\boldsymbol{y} \notin S$, 则存在非零向量$\boldsymbol{p}$，使得对每一点$\boldsymbol{x} \in clS$，有$\boldsymbol{p}^T(\boldsymbol{x}-\boldsymbol{y}) \leq 0$。

# 1.4.2 凸集分离定理

**定理1.4.5** 设$S_1$和$S_2$是$\mathbb{R}^n$中两个非空凸集，$S_1 \cap S_2 = \emptyset$，则存在非零向量$\boldsymbol{p}$，使

$$\inf\{\boldsymbol{p}^T\boldsymbol{x} | \boldsymbol{x} \in S_1\} \geq \sup\{\boldsymbol{p}^T\boldsymbol{x} | \boldsymbol{x} \in S_2\}$$

**证明** 令

$$S = S_1 - S_2 = \{\boldsymbol{z} | \boldsymbol{z} = \boldsymbol{x}^{(2)} - \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(1)} \in S_1, \boldsymbol{x}^{(2)} \in S_2\}$$

由于$S_1$和$S_2$为非空凸集，因此$S$是非空凸集。

根据定理1.4.4的推论，存在非零向量$\boldsymbol{p}$，使得对于每一个$\boldsymbol{z} \in \boldsymbol{S}$，成立$\boldsymbol{p}^T\boldsymbol{z} \leq \boldsymbol{0}$。

$$\boldsymbol{p}^T(\boldsymbol{x}^{(2)} - \boldsymbol{x}^{(1)}) \leq 0$$

$$\boldsymbol{p}^T\boldsymbol{x}^{(1)} \geq \boldsymbol{p}^T\boldsymbol{x}^{(2)}$$

# 1.4.2 凸集分离定理

## (1) Farkas定理

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

**证明** 先证必要性

设$\boldsymbol{A}\boldsymbol{x} \leq \mathbf{0}, \boldsymbol{c}^T\boldsymbol{x} > 0$有解，即存在$\overline{\boldsymbol{x}}$，使$\boldsymbol{A}\overline{\boldsymbol{x}} \leq \mathbf{0}, \boldsymbol{c}^T\overline{\boldsymbol{x}} > 0$。现在证明$\boldsymbol{A}^T\boldsymbol{x} = \boldsymbol{c}, \boldsymbol{y} \geq \mathbf{0}$无解。

用反证法，设存在$\boldsymbol{y} \geq \mathbf{0}$，使

$$\boxed{\boldsymbol{A}^T\boldsymbol{y} = \boldsymbol{c}}$$

两端转置，并右乘$\overline{\boldsymbol{x}}$，得到

$$\boxed{\boldsymbol{y}^T\boldsymbol{A}\overline{\boldsymbol{x}} = \boldsymbol{c}^T\overline{\boldsymbol{x}}}$$

由于$\boldsymbol{y} \geq \mathbf{0}, \boldsymbol{A}\overline{\boldsymbol{x}} \leq \mathbf{0}$，因此$\boldsymbol{y}^T\boldsymbol{A}\overline{\boldsymbol{x}} \leq 0$，因此$\boldsymbol{y}^T\boldsymbol{A}\overline{\boldsymbol{x}} \leq 0$

得到$\boldsymbol{c}^T\overline{\boldsymbol{x}} \leq 0$，与$\boldsymbol{c}^T\overline{\boldsymbol{x}} > 0$的假设矛盾。

# 1.4.2 凸集分离定理

## (1) Farkas定理

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

再证明充分性 $A^T y = c, y \geq 0$无解，证明$Ax \leq 0, c^T x > 0$有解，令

$$S = \{z = A^T y = c, y \geq 0\}$$

则$S$为闭凸集。由假设$c \notin S$，根据定理1.4.3，存在非零向量$x$及数$\varepsilon > 0$，使得对每一个点$z \in S$，有

$$x^T c \geq \varepsilon + x^T z$$

由于$\varepsilon > 0$，必有 $x^T c > x^T z$

两端转置，并考虑到集合$S$的定义，有 $c^T x > z^T x = y^T A x$

令$y = 0$, 得 $c^T x > 0$

由于$c^T x$为某个确定的数，$y \geq 0$，$y$的分量可取任意大，因此得出 $Ax \leq 0$

综上，非零向量$x$是$Ax \leq 0, c^T x > 0$的解。

# 1.4.2 凸集分离定理

## (2) Gordan 定理

设$A$为$m \times n$矩阵，那么，$\boldsymbol{Ax} < \mathbf{0}$有解的充要条件是不存在非零向量$\boldsymbol{y} \geq \mathbf{0}$，使$\boldsymbol{A}^T\boldsymbol{y} = \mathbf{0}$。

$\boldsymbol{A}^T$是$\mathbb{R}^n$中一组基，由m个n维列向量组成，只存在两种情况：

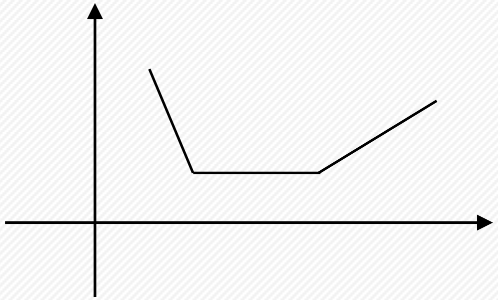
（1）存在一个方向，与$\boldsymbol{A}$中所有向量都呈钝角。

（2）$\boldsymbol{A}^T$这组基的非负、非零线性组合可以得到原点。

# 1.4.3 凸函数

**定义1.4.7** 设$S$为$\mathbb{R}^n$中的非空凸集，$f$是定义在$S$上的实函数，如果对任意的$\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}\in S$及每个数$\lambda\in(0,1)$，都有

$$f\left(\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)}\right)\le\lambda f\left(\boldsymbol{x}^{(1)}\right)+(1-\lambda)f\left(\boldsymbol{x}^{(2)}\right)$$

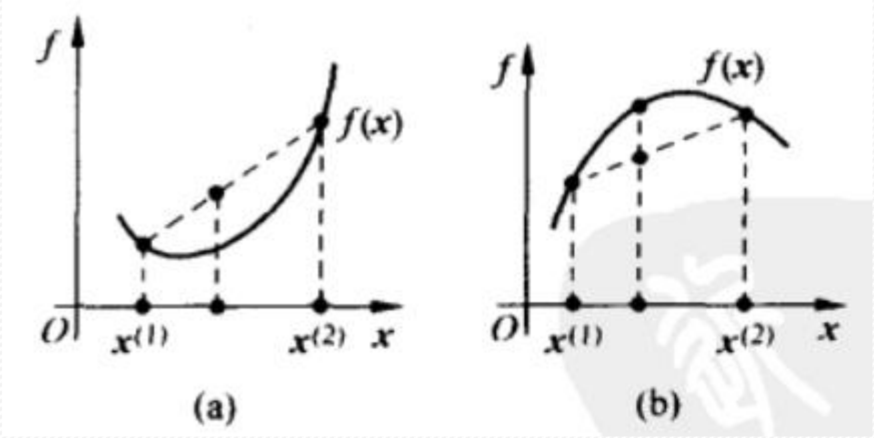


则称$f$为$S$上的**凸函数**。

如果对任意互不相同的$\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}\in S$，及每一个数$\lambda\in(0,1)$，都有

$$f\left(\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)}\right)<\lambda f\left(\boldsymbol{x}^{(1)}\right)+(1-\lambda)f\left(\boldsymbol{x}^{(2)}\right)$$

则称$f$为$S$上的**严格凸函数**。
如果$-f$为$S$上的凸函数，则称$f$为$S$上的凹函数。

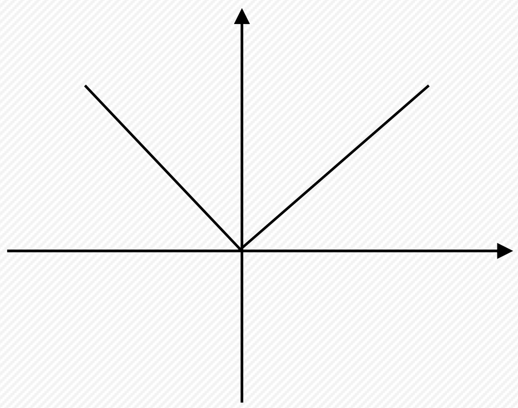


## 1.4.3 凸函数

**例1.4.8** 一元函数$f(x) = |x|$是$\mathbb{R}^1$上的凸函数.

**解** 对任意的 $x^{(1)}, x^{(2)} \in \mathbb{R}^1$及每个数$\lambda \in (0,1)$，均有

$$
\begin{aligned}
f\left(\lambda x^{(1)} + (1-\lambda)x^{(2)}\right) &= \left|\lambda x^{(1)} + (1-\lambda)x^{(2)}\right| \\
&\leq \lambda\left|x^{(1)}\right| + (1-\lambda)\left|x^{(2)}\right| \\
&= \lambda f(x^{(1)}) + (1-\lambda) f(x^{(2)})
\end{aligned}
$$



因此，由定义1.4.7知，$f(x) = |x|$为凸函数。

# 1.4.3 凸函数

**2.凸函数的性质**

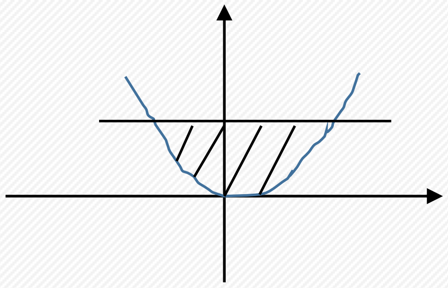
**定理1.4.8** 设$f$是定义在凸集$S$上的凸函数，实数$\lambda \geq 0$，则$\lambda f$也是定义在$S$上的凸函数。

**定理1.4.9** 设$f_1$和$f_2$是定义在凸集$S$上的凸函数，则$f_1 + f_2$也是定义在$S$上的凸函数。

**推论** 设$f_1, f_2, \dots, f_k$定义在凸集$S$上的凸函数，实数$\lambda_1, \lambda_2, \dots, \lambda_k \geq 0$，则$\sum_{i=1}^{k} \lambda_i f_i$也是定义在$S$上的凸函数。

**定理1.4.10** 设$S$是$\mathbb{R}^n$中一个非空凸集，$f$是定义在$S$上的凸函数，$\alpha$是一个实数，则**水平集**$S_\alpha = \{x | x \in S, f(x) \leq \alpha\}$是凸集。

**定理1.4.11** 设$S$是$\mathbb{R}^n$中一个凸集，$f$是定义在$S$上的凸函数，则$f$在$S$的内部连续。



# 1.4.3 凸函数

**定义1.4.8** 设$S$是$\mathbb{R}^n$中一个集合，$f$是定义在$S$上的实函数，$x \in int\ S$，$\boldsymbol{d}$是非零向量，$f$在$\bar{x}$处沿方向$\boldsymbol{d}$的**方向导数**$Df(\bar{\boldsymbol{x}};\boldsymbol{d})$定义为下列极限：

$$Df(\bar{\boldsymbol{x}};\boldsymbol{d}) = \lim_{\lambda\to 0}\frac{f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})-f(\bar{\boldsymbol{x}})}{\lambda}, \quad (1.4.25)$$

这里假设上述极限存在。$int\ S$表示集合$S$的内部。
$f$在$\bar{x}$处沿方向$\boldsymbol{d}$的右侧导数定义为

$$D^+f(\bar{\boldsymbol{x}};\boldsymbol{d}) = \lim_{\lambda\to 0^+}\frac{f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})-f(\bar{\boldsymbol{x}})}{\lambda}$$

假设，上述极限存在$f$在$\bar{x}$处沿方向$\boldsymbol{d}$的左侧导数定义为

$$D^-f(\bar{\boldsymbol{x}};\boldsymbol{d}) = \lim_{\lambda\to 0^-}\frac{f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})-f(\bar{\boldsymbol{x}})}{\lambda}$$

在(1.4.25)式中，若方向$\boldsymbol{d}=(0,\cdots,0,1,0,\cdots,0)^T$，其中第$j$个分量是1，其余$n-1$个分量全是零，则$f$在$\bar{x}$处沿方向$\boldsymbol{d}$的方向导数正好等于$f$对$x_j$的偏导数，即

$$Df(\bar{\boldsymbol{x}};\boldsymbol{d}) = \frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_j}$$

$$Df(\bar{\boldsymbol{x}};\boldsymbol{d}) = \boldsymbol{d}^T\nabla f(\bar{\boldsymbol{x}})$$

# 1.4.3 凸函数

**定理1.4.13** 设$S$是$\mathbb{R}^n$中非空凸集，$f$是定义在$S$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。

证明 设$\bar{\boldsymbol{x}}$是$f$在$S$上的局部极小点，即存在$\bar{\boldsymbol{x}}$的$\varepsilon > 0$邻域$N_\varepsilon(\bar{\boldsymbol{x}})$，使得对每一点$\boldsymbol{x} \in S \cap N_\varepsilon(\bar{\boldsymbol{x}})$，成立$f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}})$.
假设$\bar{\boldsymbol{x}}$不是全局极小点，则存在$\hat{\boldsymbol{x}} \in S$，使$f(\hat{\boldsymbol{x}}) < f(\bar{\boldsymbol{x}})$。由于$S$是凸集，因此对每一个数$\lambda \in [0,1]$，有$\lambda\hat{\boldsymbol{x}} + (1-\lambda)\bar{\boldsymbol{x}} \in S$。由于$\hat{\boldsymbol{x}}$与$\bar{\boldsymbol{x}}$是不同的两点，可取$\lambda \in (0,1)$.
又由于$f$是$S$上的凸函数，因此有

$$\begin{aligned} f(\lambda\hat{\boldsymbol{x}} + (1-\lambda)\bar{\boldsymbol{x}}) &\leq \lambda f(\hat{\boldsymbol{x}}) + (1-\lambda) f(\bar{\boldsymbol{x}}) \\ &< \lambda f(\bar{\boldsymbol{x}}) + (1-\lambda) f(\bar{\boldsymbol{x}}) \\ &< f(\bar{\boldsymbol{x}}) \end{aligned}$$

当$\lambda$取得充分小时,可使

$$\lambda\hat{\boldsymbol{x}} + (1-\lambda)\bar{\boldsymbol{x}} \in S \cap N_\varepsilon(\bar{\boldsymbol{x}}),$$

这与$\bar{\boldsymbol{x}}$为局部极小点矛盾。故$\bar{\boldsymbol{x}}$是$f$在$S$上的全局极小点。

# 1.4.4 凸函数的判别

**定理1.4.14** 设$S$是$\mathbb{R}^n$中非空开凸集，$f(\boldsymbol{x})$是定义在$S$上的可微函数，则$f(\boldsymbol{x})$为凸函数的充要条件是对任意两点$\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in S$，都有

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

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

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

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

$$f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}})$$

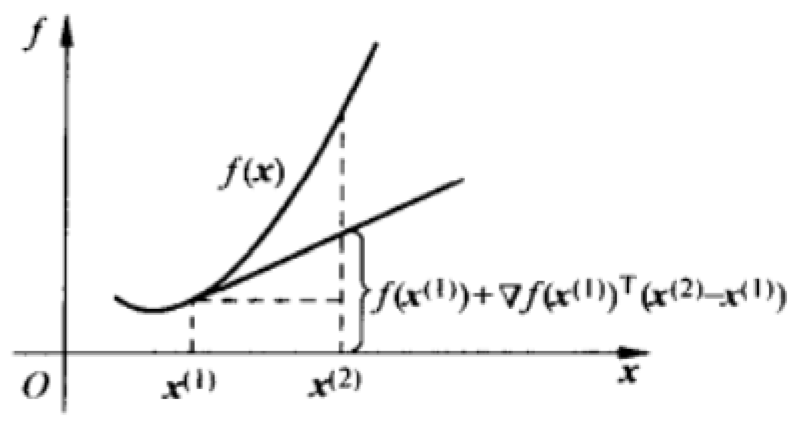



图 1.4.6

## 1.4.4 凸函数的判别

**定理1.4.15** 设$S$是$\mathbb{R}^n$中非空开凸集，$f(\boldsymbol{x})$是定义在$S$上的二次可微函数，则$f(\boldsymbol{x})$为凸函数的充要条件是在每一点$\boldsymbol{x} \in S$处Hesse矩阵半正定。

**定理1.4.16** 设$S$是$\mathbb{R}^n$中非空开凸集，$f(\boldsymbol{x})$是定义在$S$上的二次可微函数，如果在每一点$\boldsymbol{x} \in S$，Hesse矩阵正定，则$f(\boldsymbol{x})$为严格凸函数。

**注意**：逆定理并不成立。若$f(\boldsymbol{x})$是定义在$S$上的严格凸函数，则在每一点$\boldsymbol{x} \in S$处，Hesse矩阵是半正定的。

# 1.4.4 凸函数的判别

**例1.4.9** 二次函数

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

是严格凸函数。

因为

$$\nabla f(\boldsymbol{x})=\begin{pmatrix}\dfrac{\partial f}{\partial x_1}\\ \dfrac{\partial f}{\partial x_2}\end{pmatrix}=\begin{pmatrix}4x_1-2x_2+1\\ 2x_2-2x_1\end{pmatrix}$$

$$\nabla^2 f(\boldsymbol{x})=\begin{pmatrix}\dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\partial x_2}\\ \dfrac{\partial^2 f}{\partial x_2\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2}\end{pmatrix}=\begin{pmatrix}4 & -2\\ -2 & 2\end{pmatrix}$$

是正定的，因此$f(\boldsymbol{x})$是严格凸函数。

# 1.4.5 凸规划

极小化问题

$$\min f(\boldsymbol{x})$$
$$s.t.\ g_i(\boldsymbol{x}) \geq 0, i = 1, \dots, m,$$
$$h_j(\boldsymbol{x}) = 0, j = 1, \dots, l,$$

设$f(\boldsymbol{x})$是凸函数， $g_i(\boldsymbol{x})$是凹函数， $h_j(\boldsymbol{x})$是线性函数，问题的可行域是

$$S = \{\boldsymbol{x} | g_i(\boldsymbol{x}) \geq 0, i = 1, \dots, m, h_j(\boldsymbol{x}) = 0, j = 1, \dots, l\}$$

**凸规划**：求凸函数在凸集上的极小点。
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，那么它的极小点是唯一的。

# The end