最优化方法

最优化方法

教材：《最优化理论与算法》（第 2 版） 陈宝林

参考书：《Numerical Optimization》（《数值最优化》） Jorge Nocedal, Stephen J.Wright

《实用最优化方法》 唐焕文，秦学志

《线性规划》 张干宗

《非线性规划》 胡毓达

《最优化理论与方法》 袁亚湘

课件和作业：邮箱：opt_sjtu@163.com，密码(课件)：2018-2019-1(仅供下载，不要对邮箱设置进行任何修改，不要用此邮箱发邮件)

联系：邮箱：xmwang@sjtu.edu.cn，办公室：数学系 3 号楼 313 室（包图后面，即将搬家），电话 ：34202654 转 1526

答疑：1.时间：周一中午 12:10-12:45，地点：中院 108 休息室

2.时间：周二中午 11:00-13:40，地点：东中院 2-204 休息室（务必事先邮件联系）

作业：单周上课前交（若遇放假，则延续至下一周上课前交），作业本封面左上角注明本人序号（按课件邮箱中学生名单，序号 100-120 的注意按序补上最后一位数）

第一部分 准备工作

第 1 章 引言

§1 学科简述

最优化方法是一个重要的数学分支，它研究在众多的方案中什么样的方案是最优的以及怎样找出最优方案。例如：

工程设计中：怎样选择设计参数，使设计方案既满足设计要求又能降低成本；

资源分配中：怎样分配有限资源，使分配方案既满足各方面的基本要求，又获得好的经济效益；

生产计划安排中：选择怎样的计划方案，才能提高产值和利润；

城建规划中：怎样安排企业、机关、学校、商店、住户和其他单位的合理布局，才能方便群众，又有利于城市各行业的发展；

军事指挥中：怎样确定最佳作战方案，才能有效消灭敌人，保存自己，有利于战争全局；

。。。。。。

最优化正是为这些问题的解决提供理论基础和求解方法。它是一门应用广泛、实用性强的学科。

最优化是个古老的话题：

17 世纪： Newton 发明微积分，提出极值问题，Lagrange 乘数法；

1847 年：Cauchy 提出最速下降法；

1939 年：康德洛维奇提出下料问题和运输问题求解方法；

20 世纪 40 年代后：形成系统学科。

分支：线性规划，整数规划，非线性规划，几何规划，动态规划，随机规划，多目标规划，等等

在本课程中，若无特别说明，约定：

1. 向量和矩阵用粗体表示。
2. 向量均为实列向量，R^n 为 n 维列向量的集合，R^n 中的元素 $\boldsymbol{x}$ 表示为 $\boldsymbol{x}=(x_1,\cdots,x_n)^T$，其中 $x_j(j=1,\cdots,n)$

表示 $\boldsymbol{x}$ 的第 j 个分量。R=R^1。

3. 矩阵均为实矩阵，$R^{m\times n}$ 为 m 行 n 列矩阵的集合，$R^{m\times n}$ 中的元素 $\boldsymbol{A}$ 表示为 $\boldsymbol{A}=(a_{ij})_{m\times n}$，其中 a_{ij} 表示 $\boldsymbol{A}$ 的第 i 行第 j 列元素。R^m=$R^{m\times 1}$。
4. 带“*”内容为有一定难度，因时间关系无法在课内介绍，可课外自学。

§2　最优化问题

利用最优化的理论和方法解决具体问题，一般分为以下几个步骤：

1. 建模。对所要解决的问题进行分析研究，加以简化，形成最优化问题；
2. 加工。将所得问题进行整理变换，使其成为既反映实际又易于求解的形式；
3. 求解。选择或提出解决该问题的适当计算方法，编制计算程序上机计算；
4. 分析。分析计算结果，看其是否符合实际。

许多实际问题抽象成数学模型后，可归结为线性与非线性规划问题。

2.1　举例

建立优化模型三步曲：

定义决策变量，寻找限制条件，确定追求目标。

例2.1.1　生产计划问题 P2

已知数据：

产品 / 资源	$\cdots\ j\ \cdots$	限制（上限）
$\vdots$ i $\vdots$	a_{ij}	b_i
需求（上限）	d_j	
利润	c_j	

决策变量：设 x_j —第 j 种产品的生产量；

限制条件：$\sum_{j=1}^{3} a_{ij}x_j \le b_i, i=1,\cdots,4, 0\le x_j \le d_j, j=1,\cdots,3$；

追求目标：$\sum_{j=1}^{3} c_j x_j$ 尽可能大。

建立优化模型：

$$
\begin{aligned}
&\max \sum_{j=1}^{3} c_j x_j \\
&s.t.\ \sum_{j=1}^{3} a_{ij}x_j \le b_i, i=1,\cdots,4, \\
&\qquad 0\le x_j \le d_j, j=1,\cdots,3,
\end{aligned}
$$

其中 max：maximun 极大化，s.t.：subject to 受限制于。

例2.1.2　食谱问题 P2（自学）

已知数据：

食物	$\cdots\ j\ \cdots$	限制（下限）

营养		
⋮ i ⋮	a_{ij}	b_i
价格	c_j	

决策变量：设 x_j —第 j 种食物的数量；

限制条件：$\sum_{j=1}^{n} a_{ij}x_j \geq b_i, i=1,\cdots,m, x_j \geq 0, j=1,\cdots,n$；

追求目标：$\sum_{j=1}^{n} c_j x_j$ 尽可能小。

建立优化模型：

$$\begin{aligned} &\min \sum_{j=1}^{n} c_j x_j \\ &s.t.\ \sum_{j=1}^{n} a_{ij}x_j \geq b_i, i=1,\cdots,m \\ &\qquad x_j \geq 0, j=1,\cdots,n \end{aligned}$$

其中 min：minimun 极小化。

例2.1.3 结构设计问题 P3

决策变量：设 x_1 —钢管直径，x_2 —桁架高度；

限制条件：

空间限制：$x_2 \leq H$

压应力限制：压应力=单位面积上的压力

$$\text{压力} F = \frac{P}{\cos\theta} = \frac{P(L^2+x_2^2)^{1/2}}{x_2}，\text{面积} S \approx \pi T x_1，\text{压应力} \sigma = \frac{F}{S} = \frac{P(L^2+x_2^2)^{1/2}}{\pi T x_1 x_2}$$

不超过容许应力：$\sigma \leq \sigma_y$，即 $\frac{P(L^2+x_2^2)^{1/2}}{\pi T x_1 x_2} \leq \sigma_y$

不超过临界应力（避免钢管弯曲）：$\sigma \leq \sigma_l = \frac{\pi^2 E(x_1^2+T^2)}{8(L^2+x_2^2)}$，即 $\frac{P(L^2+x_2^2)^{1/2}}{\pi T x_1 x_2} \leq \frac{\pi^2 E(x_1^2+T^2)}{8(L^2+x_2^2)}$

追求目标：重量最小，即 $2\pi\rho T x_1 (L^2+x_2^2)^{1/2}$ 最小。

建立模型：

$$\begin{aligned} &\min 2\pi\rho T x_1 (L^2+x_2^2)^{1/2} \\ &s.t.\ x_1 \geq 0, 0 \leq x_2 \leq H \\ &\qquad \frac{P(L^2+x_2^2)^{1/2}}{\pi T x_1 x_2} \leq \sigma_y \\ &\qquad \frac{P(L^2+x_2^2)^{1/2}}{\pi T x_1 x_2} \leq \frac{\pi^2 E(x_1^2+T^2)}{8(L^2+x_2^2)} \end{aligned}$$

例2.1.4 选址问题 P4（自学）

设有 n 个市场，第 j 个市场的位置为 (a_j, b_j)，对某种货物的需求量为 q_j。现计划建立 m 个货栈，第 i

个货栈的容量为c_i。试确定货栈的位置，使各货栈到各市场的运输量与路程乘积之和最小。

设(x_i, y_i)：第 i 个货栈的位置；z_{ij}：第 i 个货栈供给第 j 个市场的货物量。

$$\begin{aligned}
\min\ & \sum_{i=1}^{m}\sum_{j=1}^{n} z_{ij}\sqrt{(x_i-a_j)^2+(y_i-b_j)^2} \\
s.t.\ & \sum_{j=1}^{n} z_{ij} \le c_i, i=1,\cdots,m \\
& \sum_{i=1}^{m} z_{ij} = q_j, j=1,\cdots,n \\
& z_{ij} \ge 0, i=1,\cdots,m, j=1,\cdots,n
\end{aligned}$$

例2.1.5 投资决策问题

设一企业有 B 万元可用于投资，现有 m 个项目可供选择。若对第 i 个项目投资，需花资金a_i万元，可获利c_i万元。试确定最佳的投资方案。

设 $x_i = \begin{cases} 0, \text{不投} i^{\#} \\ 1, \quad \text{投} i^{\#} \end{cases}$，则考虑收益最大时，有

$$\begin{aligned}
\max\ & \sum_{j=1}^{n} c_j x_j \\
s.t.\ & \sum_{j=1}^{n} a_j x_j \le B \\
& x_j \in \{0,1\}, j=1,\cdots,n
\end{aligned}$$

若同时要求总投资额尽可能小，则有

$$\begin{aligned}
\max\ & \sum_{j=1}^{n} c_j x_j \\
\min\ & \sum_{j=1}^{n} a_j x_j \\
s.t.\ & \sum_{j=1}^{n} a_j x_j \le B \\
& x_j \in \{0,1\}, j=1,\cdots,n
\end{aligned}$$

2.2 数学模型

最优化问题的一般模型为

$$\begin{aligned}
\min\ & f(\boldsymbol{x}) \\
s.t.\ & \boldsymbol{x} \in S
\end{aligned} \tag{P}$$

其中$\boldsymbol{x} \in R^n$是决策变量， $S \subset R^n$是约束集， $f: S \to R$是目标函数。

当$S = \{\boldsymbol{x} \in R^n \mid g_i(\boldsymbol{x}) \ge 0, i=1,\cdots,m, h_j(\boldsymbol{x}) = 0, j=1,\cdots,l\}$时，则(P)具体表示为：

$$\begin{aligned}
\min\ & f(\boldsymbol{x}) \\
s.t.\ & g_i(\boldsymbol{x}) \ge 0, i=1,\cdots,m \\
& h_j(\boldsymbol{x}) = 0, j=1,\cdots,l
\end{aligned}$$

注意，极大化问题可转化为极小化问题。

当 $S=R^n(m=l=0)$ 时，(P)实际上是一个无约束优化问题，记作

$$\min f(\boldsymbol{x}) \tag{UNP}$$

否则(P)是一个约束优化问题。

（1）当 $f, g_i(i=1,\cdots,m), h_j(j=1,\cdots,l)$ 为线性函数时，(P)称为线性规划问题；

（2）当 $f, g_i(i=1,\cdots,m), h_j(j=1,\cdots,l)$ 中至少有一个非线性函数时，(P)称为非线性规划问题；

（3）当 f 为二次函数，$g_i(i=1,\cdots,m), h_j(j=1,\cdots,l)$ 为线性函数时，(P)称为二次规划问题；

（4）当 S 中要求某些变量为非负整数时，(P)称为整数规划问题。

（5）当 $f(\boldsymbol{x})$ 是向量函数时，(P)称为多目标规划问题。

例如，例 2.1.1 和例 2.1.2 是线性规划问题，例 2.1.3 和例 2.1.4 是非线性规划问题，例 2.1.5 是整数规划问题和多目标规划问题。

2.3 解的定义

对一般最优化问题(P)，我们给出如下解的定义。

定义 2.3.1 若 $\boldsymbol{x}\in S$，则称 $\boldsymbol{x}$ 是(P)的可行解。S 称为(P)的可行集或可行域或约束集。

定义 2.3.2 设 $\boldsymbol{x}^*\in S$。若

$$f(\boldsymbol{x}^*)\le f(\boldsymbol{x}),\forall \boldsymbol{x}\in S$$

则称 $\boldsymbol{x}^*$ 是(P)的(整体或全局)最优解，$f(\boldsymbol{x}^*)$ 是(P)的最优值。若

$$f(\boldsymbol{x}^*)< f(\boldsymbol{x}),\forall \boldsymbol{x}\in S\setminus\{\boldsymbol{x}^*\}$$ $x\in S \cup x\notin\{x^*\}$

则称 $\boldsymbol{x}^*$ 是(P)的严格(整体或全局)最优解。

例 2.3.1 求解无约束优化问题

$$\min f(x_1,x_2)=(x_1-1.5)^2+(x_2-1)^2$$

由于该问题只含 2 个变量，我们可用图解法求解(见图 1)，得最优解 $\boldsymbol{x}^*=(x_1^*,x_2^*)^T=(1.5,1)^T$。

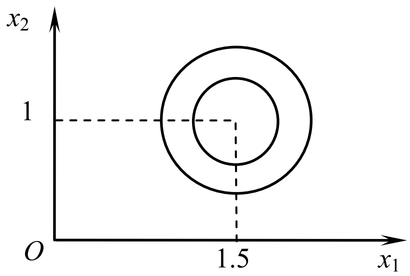


图 1

例 2.3.2 求解约束优化问题

$$\begin{aligned}\min\ & f(x_1,x_2)=(x_1-1.5)^2+(x_2-1)^2\\ s.t.\ & x_1^2+x_2^2\le 1\\ & 2x_1+x_2\ge 1\\ & x_1\ge 0, x_2\ge 0\end{aligned}$$

由于该问题只含 2 个变量，我们可用图解法求解(见图 2)，得最优解 $\boldsymbol{x}^*=(x_1^*,x_2^*)^T=(\frac{3}{\sqrt{13}},\frac{2}{\sqrt{13}})^T$。

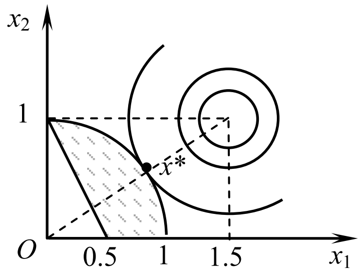


图 2

定义 2.3.3 设 $\boldsymbol{x}^* \in S$。若存在 $\boldsymbol{x}^*$ 的 δ-邻域 $N(\boldsymbol{x}^*,\delta)$，使

$$f(\boldsymbol{x}^*) \le f(\boldsymbol{x}), \forall \boldsymbol{x} \in S \cap N(\boldsymbol{x}^*,\delta)$$

则称 $\boldsymbol{x}^*$ 是(P)的局部最优解，$f(\boldsymbol{x}^*)$ 是(P)的局部最优值。若存在 $\boldsymbol{x}^*$ 的 δ-邻域 $N(\boldsymbol{x}^*,\delta)$，使

$$f(\boldsymbol{x}^*) < f(\boldsymbol{x}), \forall \boldsymbol{x} \in S \cap N(\boldsymbol{x}^*,\delta) \setminus \{\boldsymbol{x}^*\}$$

则称 $\boldsymbol{x}^*$ 是(P)的严格局部最优解。

显然，整体最优解一定是局部最优解，整体最优解不一定唯一，但严格整体最优解一定唯一。

实际问题通常是求全局最优解，但求解最优化问题的绝大多数方法只能求得问题的局部最优解。在实际中，希望一个局部最优解是全局的，可以用一系列初始点利用最优化方法得出若干个局部最优解，然后通过比较得到全局最优解。

§3 几个数学概念

本节复习多元可微函数的梯度、Hesse 矩阵及其 Taylor 展开式。

3.1 向量的范数

设 $\boldsymbol{x} = (x_1,\cdots,x_n)^T$，$\boldsymbol{y} = (y_1,\cdots,y_n)^T$，则

$\vec{x} \ge \vec{y} \cup \vec{x} \ne \vec{y} \not\Rightarrow \vec{x} > \vec{y}$

$$\boldsymbol{x} = \boldsymbol{y} \Leftrightarrow x_i = y_i, i = 1,\cdots,n,\quad \boldsymbol{x} \ge \boldsymbol{y} \Leftrightarrow x_i \ge y_i, i = 1,\cdots,n,\quad \boldsymbol{x} > \boldsymbol{y} \Leftrightarrow x_i > y_i, i = 1,\cdots,n$$

$\boldsymbol{x}^T\boldsymbol{y} = \sum_{i=1}^n x_i y_i = \boldsymbol{y}^T\boldsymbol{x}$ 为 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的内积，$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^T\boldsymbol{x}} = \sqrt{\sum_{i=1}^n x_i^2}$ 为 $\boldsymbol{x}$ 的 Euclid 范数，$\|\boldsymbol{x}-\boldsymbol{y}\|$ 称为 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的距离。

$\|\vec{x}-\vec{y}\| = \sqrt{\sum (x_i-y_i)^2}$

$\boldsymbol{x}^T\boldsymbol{y} = \|\boldsymbol{x}\|\|\boldsymbol{y}\|\cos\theta$，其中 $\theta \in [0,\pi]$ 是 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的夹角。因此，

$$\boldsymbol{x}^T\boldsymbol{y}\begin{cases} > 0, \theta \in \left[0,\dfrac{\pi}{2}\right) \\ = 0, \theta = \dfrac{\pi}{2} \\ < 0, \theta \in \left(\dfrac{\pi}{2},\pi\right] \end{cases}$$

两个重要不等式：

三角不等式：$\|\boldsymbol{x}+\boldsymbol{y}\| \le \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$，等式成立当且仅当 $\boldsymbol{x} = \lambda\boldsymbol{y}$ ($\lambda \ge 0$ 为实数) 同方向

Cauchy 不等式：$|\boldsymbol{x}^T\boldsymbol{y}| \le \|\boldsymbol{x}\|\|\boldsymbol{y}\|$，等式成立当且仅当 $\boldsymbol{x} = \lambda\boldsymbol{y}$ (λ 为实数) 平行

3.2 开集和闭集

设 $S \subset R^n$， $\boldsymbol{x} \in R^n$ 。

如果 $\boldsymbol{x}$ 存在一个属于 S 的邻域，则称 $\boldsymbol{x}$ 是 S 的内点。S 的所有内点组成的集合称为 S 的内部，记作 int$S$。如果 S 中每一点都为内点，即 int$S$=S，则称 S 为开集。

如果 $\boldsymbol{x}$ 的任意邻域内含有 S 中的点和非 S 中的点，则称 $\boldsymbol{x}$ 是 S 的边界点。S 的所有边界点组成的集合称为 S 的边界，记作 rb$S$，并称 $clS = \mathrm{int}\, S \cup \mathrm{rb}S$ 为 S 的闭包。如果 S 的所有边界点都属于 S，即 cl$S$=S，则称 S 为闭集。有界闭集称为紧集。

注 3.2.1 若 $S \subset R^n$ 为闭集，则 S 中每个收敛序列的极限均属于 S；若 $\boldsymbol{x}$ 是闭集 $S \subset R^n$ 的边界点，则存在 $\{\boldsymbol{x}^k\} \subset S^C$，使 $\boldsymbol{x}^k \to \boldsymbol{x}$，其中 S^C=$R^n \setminus S$ 为 S 的补集。

3.3 多元函数的梯度、Hesse 矩阵和 Taylor 公式

设集合 $S \subset R^n$ 非空，函数 $f: S \to R^1$。在 $\boldsymbol{x} \in S$ 处，记 f 对 x_j 的偏导数为 $\dfrac{\partial f(\boldsymbol{x})}{\partial x_j}$，$\dfrac{\partial f(\boldsymbol{x})}{\partial x_j}$ 对 x_i 的偏导数为 $\dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_j \partial x_i} = \dfrac{\partial}{\partial x_i}\left(\dfrac{\partial f(\boldsymbol{x})}{\partial x_j}\right)$（书上 $\dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = \dfrac{\partial}{\partial x_i}\left(\dfrac{\partial f(\boldsymbol{x})}{\partial x_j}\right)$）。

定义 3.3.1 设集合 $S \subset R^n$ 非空，函数 $f: S \to R^1$。

（1）若 f 在每一点 $\boldsymbol{x} \in S$ 处连续，则称 f 在 S 上连续，记 $f \in C(S)$。 $C^0(S)$

（2）再设 S 为开集，$\bar{\boldsymbol{x}} \in S$，若 $\dfrac{\partial f(\bar{\boldsymbol{x}})}{\partial x_j}, j = 1, \cdots, n$ 存在，则称 f 在 $\bar{\boldsymbol{x}}$ 处(一阶)可导，称 $f'(\bar{\boldsymbol{x}}) = \left(\dfrac{\partial f(\bar{\boldsymbol{x}})}{\partial x_1}, \cdots, \dfrac{\partial f(\bar{\boldsymbol{x}})}{\partial x_n}\right)$ 是 f 在 $\bar{\boldsymbol{x}}$ 处(一阶)导数，并且称向量

$$\nabla f(\bar{\boldsymbol{x}}) = f'(\bar{\boldsymbol{x}})^T = \left(\frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_1}, \cdots, \frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_n}\right)^T$$

是 f 在 $\bar{\boldsymbol{x}}$ 处的梯度。若 $\dfrac{\partial f(\boldsymbol{x})}{\partial x_j}, j = 1, \cdots, n$ 在每一点 $\boldsymbol{x} \in S$ 处都存在且连续，则称 f 在 S 上连续可微，记 $f \in C^1(S)$。 $C^1(S)$

（3）再设 S 为开集，$\bar{\boldsymbol{x}} \in S$，若 $\dfrac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_i \partial x_j}, i, j = 1, \cdots, n$ 存在，则称 f 在 $\bar{\boldsymbol{x}}$ 处二阶可导，并且称矩阵

$$\nabla^2 f(\bar{\boldsymbol{x}}) = \left(\nabla \frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_1}, \cdots, \nabla \frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_n}\right) = \begin{pmatrix} \dfrac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_n \partial x_1} \\ & \cdots & \\ \dfrac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_1 \partial x_n} & & \dfrac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_n^2} \end{pmatrix}$$

对称矩阵.

是 f 在 $\bar{\boldsymbol{x}}$ 处的 Hesse 阵。若 $\dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}, i, j = 1, \cdots, n$ 在每一点 $\boldsymbol{x} \in S$ 处都存在且连续，则称 f 在 S 上二阶连续

可微，记 $f\in C^2(S)$（这时 $\nabla^2 f(\boldsymbol{x})$ 为对称矩阵）。

当 $S=R^n$ 时，简记 $C(S)$ 为 C，$C^1(S)$ 为 C^1，$C^2(S)$ 为 C^2。

$= a_1x_1 + a_2x_2 + \cdots + a_nx_n + b$

注 3.2.1 $\nabla f(\bar{\boldsymbol{x}})$ 是等值面 $f(\boldsymbol{x})=f(\bar{\boldsymbol{x}})$ 在点 $\bar{\boldsymbol{x}}$ 处的法方向，并且指向 f 增加的方向。

若 $f(\boldsymbol{x})=\boldsymbol{a}^T\boldsymbol{x}+b$，则 $\nabla f(\boldsymbol{x})=\boldsymbol{a}$，$\nabla^2 f(\boldsymbol{x})=0_{n\times n}$。 $f(x)=\sum ax+b = \cancel{\begin{pmatrix} ax_1 \\ a_2x_2 \\ \vdots \\ a_nx_n \end{pmatrix}} + b$ $\nabla f(x)=\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ $\nabla^2 f(x)=0_{n\times n}$

若 $f(\boldsymbol{x})=\dfrac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}+\boldsymbol{b}^T\boldsymbol{x}+c$，其中 $\boldsymbol{Q}$ 为对称矩阵，则 $\nabla f(\boldsymbol{x})=\boldsymbol{Q}\boldsymbol{x}+\boldsymbol{b}$，$\nabla^2 f(\boldsymbol{x})=\boldsymbol{Q}$。

$f(x)=\frac{1}{2}\sum_{i=1}\sum_{j=1} Q_{ij}x_ix_j+\sum_{i=1} b_ix_i + c$

若 $f:R^n\to R$，$\bar{\boldsymbol{x}}\in R^n, \boldsymbol{d}\in R^n\setminus\{0\}$，$\varphi(\lambda)=f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d}), \lambda\in R$，则 $f\in C^1$ 时 $\varphi'(\lambda)=\nabla f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})^T\boldsymbol{d}$，$f\in C^2$ 时 $\varphi''(\lambda)=\boldsymbol{d}^T\nabla^2 f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})\boldsymbol{d}$。因此，$f\in C^1$ 时 $\varphi'(0)=\nabla f(\bar{\boldsymbol{x}})^T\boldsymbol{d}$，$f\in C^2$ 时 $\varphi''(0)=\boldsymbol{d}^T\nabla^2 f(\bar{\boldsymbol{x}})\boldsymbol{d}$。

对称矩阵.

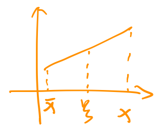
定理 3.3.1 (Taylor 展开式)设开集 $S\subset R^n$，函数 $f: S\to R$，$\bar{\boldsymbol{x}}\in S$，$\boldsymbol{x}\in S$。

（1）若 $f\in C^1(S)$，则

$f(x)=f(x_1)+f'(x_1)(x-x_1)+o(\|x-x_1\|)$

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

中值形式



$\xi=\bar{x}+\theta(x-\bar{x})$

其中 $\boldsymbol{\xi}=\bar{\boldsymbol{x}}+\theta(\boldsymbol{x}-\bar{\boldsymbol{x}})\in S, 0<\theta<1$；

（2）若 $f\in C^2(S)$，则

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

$$\cancel{=f(\bar{\boldsymbol{x}})+\nabla f(\boldsymbol{\xi}\bar{\boldsymbol{x}})^T(\boldsymbol{x}-\bar{\boldsymbol{x}})+\frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{x}})^T\nabla^2 f(\ )(\ -\ )}$$

其中 $\boldsymbol{\xi}=\bar{\boldsymbol{x}}+\theta(\boldsymbol{x}-\bar{\boldsymbol{x}})\in S, 0<\theta<1$。 $=f(\bar{x})+\nabla f(\bar{x})^T(x-\bar{x})+\frac{1}{2}(x-\bar{x})^T\nabla^2 f(\xi)(x-\bar{x})$

根据定理 3.3.1 知，若 $f\in C^1$，则 注：在 $\lambda=0$ 的 Taylor 展开

$$f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})=f(\bar{\boldsymbol{x}})+\lambda\nabla f(\bar{\boldsymbol{x}})^T\boldsymbol{d}+o(\lambda)=f(\bar{\boldsymbol{x}})+\lambda\nabla f(\bar{\boldsymbol{x}}+\theta\boldsymbol{d})^T\boldsymbol{d}$$

其中 $0<\theta<\lambda$；若 $f\in C^2$，则

$$f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})=f(\bar{\boldsymbol{x}})+\lambda\nabla f(\bar{\boldsymbol{x}})^T\boldsymbol{d}+\frac{\lambda^2}{2}\boldsymbol{d}^T\nabla^2 f(\bar{\boldsymbol{x}})\boldsymbol{d}+o\left(\lambda^2\right)$$

$$=f(\bar{\boldsymbol{x}})+\lambda\nabla f(\bar{\boldsymbol{x}})^T\boldsymbol{d}+\frac{\lambda^2}{2}\boldsymbol{d}^T\nabla^2 f(\bar{\boldsymbol{x}}+\theta\boldsymbol{d})\boldsymbol{d}$$

其中 $0<\theta<\lambda$。 θ 相当于 ξ.

定义 3.3.2 设集合 $S\subset R^n$ 非空，函数 $\boldsymbol{h}: S\to R^m$，$\boldsymbol{h}(\boldsymbol{x})=(h_1(\boldsymbol{x}),\cdots,h_m(\boldsymbol{x}))^T$。若对 $i=1,\cdots,m$，$\dfrac{\partial h_i(\boldsymbol{x})}{\partial x_j}, j=1,\cdots,n$ 在 $\bar{\boldsymbol{x}}\in S$ 处存在，则称 $\boldsymbol{h}$ 在 $\bar{\boldsymbol{x}}$ 处一阶可导，并且称矩阵

$$\boldsymbol{h}'(\bar{\boldsymbol{x}})=J_{\boldsymbol{h}}(\bar{\boldsymbol{x}})=\left(\frac{\partial h_i(\bar{\boldsymbol{x}})}{\partial x_j}\right)_{m\times n}$$

$m=1$，即为单函数梯度.

~~[illegible]~~ 导数和梯度是转置关系

是 $\boldsymbol{h}$ 在 $\bar{\boldsymbol{x}}$ 处的一阶导数或 Jacobi 矩阵，$\nabla\boldsymbol{h}(\bar{\boldsymbol{x}})=\boldsymbol{h}'(\bar{\boldsymbol{x}})^T=(\nabla h_1(\bar{\boldsymbol{x}}),\cdots,\nabla h_m(\bar{\boldsymbol{x}}))$ 是 $\boldsymbol{h}$ 在 $\bar{\boldsymbol{x}}$ 处的梯度。若对 $i=1,\cdots,m$，$\dfrac{\partial h_i(\boldsymbol{x})}{\partial x_j}, j=1,\cdots,n$ 在所有 $\boldsymbol{x}\in S$ 处存在并且连续，则称 $\boldsymbol{h}$ 在 S 上连续可微，记 $\boldsymbol{h}\in C^1(S)$。

显然，对函数 $f:R^n\to R^1$，$\nabla^2 f(\boldsymbol{x})=\nabla(\nabla f(\boldsymbol{x}))$。

确实可以分两次梯度过程.

$h'(x) = J_h(x) = A$　　$\nabla h(x) = h(x)^T = A^T$

若 $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$，则 $J_{\boldsymbol{h}}(\boldsymbol{x}) = \boldsymbol{A}$，$\nabla \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{A}^T$。

定理 3.3.2　设 $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x}))$，其中 $\boldsymbol{g}: R^n \to R^m$ 和 $\boldsymbol{f}: R^m \to R^k$ 均可微。则 $\boldsymbol{h}$ 也可微，并且 $\boldsymbol{h}'(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{g}(\boldsymbol{x}))\boldsymbol{g}'(\boldsymbol{x})$，即 $\nabla \boldsymbol{h}(\boldsymbol{x}) = \nabla \boldsymbol{g}(\boldsymbol{x}) \nabla \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x}))$。

先 $f(g(x))$ 导再 $g(x)$　　梯度反之，也符合了 $\nabla h(x) = h'(x)^T$ 的原理.

§4 凸集和凸函数

凸集和凸函数的理论一般称为凸分析，是最优化的理论基础。

4.1 凸集

设 $\boldsymbol{x}^1,\cdots,\boldsymbol{x}^m\in R^n$，则称 $\bar{\boldsymbol{x}}=\lambda_1\boldsymbol{x}^1+\cdots+\lambda_m\boldsymbol{x}^m$ 为 $\boldsymbol{x}^1,\cdots,\boldsymbol{x}^m$ 的线性组合；若 $\lambda_1,\cdots,\lambda_m\ge(>)0$，则称 $\bar{\boldsymbol{x}}$ 为 $\boldsymbol{x}^1,\cdots,\boldsymbol{x}^m$ 的非负(正)线性组合；若 $\lambda_1+\cdots+\lambda_m=1$，则称 $\bar{\boldsymbol{x}}$ 为 $\boldsymbol{x}^1,\cdots,\boldsymbol{x}^m$ 的仿射组合；若 $\lambda_1,\cdots,\lambda_m\ge(>)0$，并且 $\lambda_1+\cdots+\lambda_m=1$，则称 $\bar{\boldsymbol{x}}$ 为 $\boldsymbol{x}^1,\cdots,\boldsymbol{x}^m$ 的（严格）凸组合。

特别，当 m=2 时，$\boldsymbol{x}^1,\boldsymbol{x}^2$ 的（严格）凸组合为 $\bar{\boldsymbol{x}}=\lambda\boldsymbol{x}^1+(1-\lambda)\boldsymbol{x}^2$，其中 $\lambda\in[0,1]$（$\lambda\in(0,1)$）。

我们称

$$[\boldsymbol{x}^1,\boldsymbol{x}^2]\triangleq\{\lambda\boldsymbol{x}^1+(1-\lambda)\boldsymbol{x}^2\mid\lambda\in[0,1]\}$$

为连结 $\boldsymbol{x}^1,\boldsymbol{x}^2$ 的闭线段；$\boldsymbol{x}^1\ne\boldsymbol{x}^2$ 时，我们称

$$(\boldsymbol{x}^1,\boldsymbol{x}^2)\triangleq\{\lambda\boldsymbol{x}^1+(1-\lambda)\boldsymbol{x}^2\mid\lambda\in(0,1)\}$$

为连结 $\boldsymbol{x}^1,\boldsymbol{x}^2$ 的开线段；同样可定义 $[\boldsymbol{x}^1,\boldsymbol{x}^2),(\boldsymbol{x}^1,\boldsymbol{x}^2]$。

定义 4.1.1 设 $S\subset R^n$。若对任意的 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S$，有 $[\boldsymbol{x}^1,\boldsymbol{x}^2]\subset S$，则称 S 为凸集。

规定空集为凸集。

例 4.1.1 单点集 $\{\boldsymbol{x}\}$，全空间 R^n，超平面 $H(\boldsymbol{p},\alpha)=\{\boldsymbol{x}\in R^n\mid\boldsymbol{p}^T\boldsymbol{x}=\alpha\}$（其中 $\boldsymbol{p}\in R^n\setminus\{\boldsymbol{0}\}$ 为超平面的法向量），半空间

$$H^+(\boldsymbol{p},\alpha)=\{\boldsymbol{x}\in R^n\mid\boldsymbol{p}^T\boldsymbol{x}\ge\alpha\},\quad H^-(\boldsymbol{p},\alpha)=\{\boldsymbol{x}\in R^n\mid\boldsymbol{p}^T\boldsymbol{x}\le\alpha\}\ \text{（闭半空间）}$$

$$\operatorname{int}H^+(\boldsymbol{p},\alpha)=\{\boldsymbol{x}\in R^n\mid\boldsymbol{p}^T\boldsymbol{x}>\alpha\},\quad \operatorname{int}H^-(\boldsymbol{p},\alpha)=\{\boldsymbol{x}\in R^n\mid\boldsymbol{p}^T\boldsymbol{x}<\alpha\}\ \text{（开半空间）}$$

为凸集；超球 $B(\boldsymbol{x}^0,\alpha)=\{\boldsymbol{x}\in R^n\,|\,\|\boldsymbol{x}-\boldsymbol{x}^0\|\le\alpha\}$ 也是凸集。

定理 4.1.1(习题 4) 设 $S\subset R^n$，则 S 为凸集的充要条件是，S 中任意 m 个点的凸组合都在 S 中，即

$$S=\left\{\sum_{i=1}^m\lambda_i\boldsymbol{x}^i\mid\sum_{i=1}^m\lambda_i=1,\lambda_i\ge0,i=1,\cdots,m,m\ge1,\text{int}\right\}$$

定理 4.1.2 任意一族凸集的交集是凸集(自己推导)。

例 4.1.2 $S=\{\boldsymbol{x}\in R^n\mid\boldsymbol{A}\boldsymbol{x}\ge\boldsymbol{b}\}$，$S=\{\boldsymbol{x}\in R^n\mid\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\ge\boldsymbol{0}\}$ 是凸集。

有限个半空间的交集称为多面凸集，有界且非空的多面凸集称为凸多面体。

例 4.1.2 中的集合均为多面凸集。$S=\{\boldsymbol{x}\in R^n\mid\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\ge\boldsymbol{0}\}$ 称为标准多面凸集。

定义 4.1.2 设 $S\subset R^n$ 是凸集，$\boldsymbol{x}\in S$。若不存在 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S,\boldsymbol{x}^1\ne\boldsymbol{x}^2$，使 $\boldsymbol{x}\in(\boldsymbol{x}^1,\boldsymbol{x}^2)$，则称 $\boldsymbol{x}$ 是 S 的极点。多面凸集的极点又可称为顶点。

从图中可以看出，有界凸集中任意点可以表示成有限个极点的凸组合。但是，对于无界凸集来说，没有此结论，还需要利用方向和极向的概念。

定义 4.1.3 设 $S\subset R^n$ 是凸集，$\boldsymbol{d}\in R^n$ 是非零向量。

（1）若对任意的 $\boldsymbol{x}\in S$，有

$$\boldsymbol{x}+\lambda\boldsymbol{d}\in S,\quad\forall\lambda>0$$

则称 $\boldsymbol{d}$ 是 S 的方向。S 的所有方向组成的集合记作 $D(S)$。

（2）若$\boldsymbol{d}^1,\boldsymbol{d}^2 \in D(S)$，并且存在$\lambda>0$，使$\boldsymbol{d}^1=\lambda\boldsymbol{d}^2$，则称$\boldsymbol{d}^1,\boldsymbol{d}^2$是相同方向，否则称为不同方向。

（3）若$\boldsymbol{d}\in D(S)$，并且不存在S的两个不同方向$\boldsymbol{d}^1,\boldsymbol{d}^2$，使$\boldsymbol{d}$表示成$\boldsymbol{d}^1,\boldsymbol{d}^2$的正线性组合，则称$\boldsymbol{d}$是$S$的极向。

显然，有界凸集不存在方向和极向。

例 4.1.3　设$S=\{(x_1,x_2)^T \mid x_2-1\ge |x_1|\}$，则$D(S)=\{(d_1,d_2)^T\ne 0 \mid d_2\ge |d_1|\}$。（画图）

例4.1.4　设$S=\{\boldsymbol{x}\mid \boldsymbol{Ax}=\boldsymbol{b},\boldsymbol{x}\ge\boldsymbol{0}\}\ne\varnothing$，则$D(S)=\{\boldsymbol{d}\ne 0\mid \boldsymbol{Ad}=\boldsymbol{0},\boldsymbol{d}\ge\boldsymbol{0}\}$。

定理 4.1.3(标准多面凸集表示定理)　设$S=\{\boldsymbol{x}\mid \boldsymbol{Ax}=\boldsymbol{b},\boldsymbol{x}\ge\boldsymbol{0}\}\ne\varnothing$，则

（1）S有并且有有限个不同极点，设为$\boldsymbol{x}^1,\boldsymbol{x}^2,\cdots,\boldsymbol{x}^k\,(k\ge 1)$；

（2）S有极向当且仅当S无界；若S无界，则S有有限个不同极向，设为$\boldsymbol{d}^1,\boldsymbol{d}^2,\cdots,\boldsymbol{d}^l\,(l\ge 1)$；

（3）$\boldsymbol{x}\in S$当且仅当$\boldsymbol{x}=\sum_{i=1}^{k}\lambda_i\boldsymbol{x}^i+\sum_{j=1}^{l}\mu_j\boldsymbol{d}^j$，其中$\lambda_i\ge 0,i=1,\cdots,k,\sum_{i=1}^{k}\lambda_i=1,\mu_j\ge 0,j=1,\cdots,l$，其中$l$=0若$S$有界或$l\ge 1$若$S$无界。

证明见参考文献。

注 4.1.2　一般的非空多面凸集不一定存在极点，例$S=\{(x_1,x_2)^T\mid x_1\ge 0\}$。

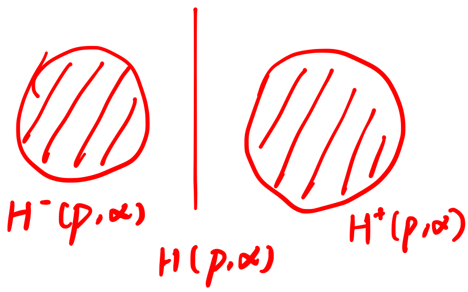
定义 4.1.4　设$S\subset R^n$。包含S的所有凸集的交集，即包含S的最小凸集，称为S的凸包，记*conv*(*S*)。可以证明，*conv*(*S*)是由S中任意有限个点的凸组合的全体构成的集合，即

$$conv(S)=\left\{\sum_{i=1}^{m}\lambda_i\boldsymbol{x}^i \mid \sum_{i=1}^{m}\lambda_i=1,\lambda_i\ge 0,i=1,\cdots,m,m\ge 1,\mathrm{int}\right\}$$

定义 4.1.5　设$C\subset R^n$。若$\forall\boldsymbol{x}\in C$和实数$\lambda\ge 0$，有$\lambda\boldsymbol{x}\in C$，则称$C$是锥。此外，若$C$又是凸集，则称$C$是凸锥。

例 4.1.5　$R_+^n=\{\boldsymbol{x}\in R^n\mid \boldsymbol{x}\ge\boldsymbol{0}\}$是凸锥，称为非负锥。$S=\{\boldsymbol{A}^T\boldsymbol{y}\mid \boldsymbol{y}\ge\boldsymbol{0}\}$是凸锥，其中$\boldsymbol{A}\in R^{m\times n}$。

4.2　凸集分离定理



凸集的许多应用都涉及到凸集的分离性质，为此先讨论凸集的分离问题。

定义 4.2.1　设$S_1,S_2\subset R^n$是非空集合。若存在$\boldsymbol{p}\in R^n\setminus\{\boldsymbol{0}\}$，$\alpha\in R$，

（1）使$S_1\subset H^-(\boldsymbol{p},\alpha)$，$S_2\subset H^+(\boldsymbol{p},\alpha)$，即

$$\boldsymbol{p}^T\boldsymbol{x}^1\le(\alpha\le)\boldsymbol{p}^T\boldsymbol{x}^2,\ \forall\boldsymbol{x}^1\in S_1,\boldsymbol{x}^2\in S_2$$

则称超平面$H(\boldsymbol{p},\alpha)$分离S_1,S_2，或称S_1,S_2是可分离的；特别地，$S_1\cup S_2\not\subset H(\boldsymbol{p},\alpha)$，则称超平面$H(\boldsymbol{p},\alpha)$正常分离$S_1,S_2$，或称$S_1,S_2$是可正常分离的。

(2)使$S_1\subset H^-(\boldsymbol{p},\alpha)$，$S_2\subset \mathrm{int}\,H^+(\boldsymbol{p},\alpha)$，或$S_1\subset \mathrm{int}\,H^-(\boldsymbol{p},\alpha)$，$S_2\subset H^+(\boldsymbol{p},\alpha)$，即

$$\boldsymbol{p}^T\boldsymbol{x}^1\le\alpha<\boldsymbol{p}^T\boldsymbol{x}^2,\ \forall\boldsymbol{x}^1\in S_1,\boldsymbol{x}^2\in S_2\text{ 或 }\boldsymbol{p}^T\boldsymbol{x}^1<\alpha\le\boldsymbol{p}^T\boldsymbol{x}^2,\ \forall\boldsymbol{x}^1\in S_1,\boldsymbol{x}^2\in S_2$$

则称超平面$H(\boldsymbol{p},\alpha)$严格分离S_1,S_2，或称S_1,S_2是可严格分离的。

例 4.2.1　$S_1=\{(x_1,x_2)\mid x_2>\frac{1}{x_1},x_1>0\}$，$S_2=\{(x_1,x_2)\mid x_2<0\}$，超平面$H=\{(x_1,x_2)\mid x_2=0\}$正常分离和严格分离$S_1$和$S_2$。

注 4.2.1 若 S_1,S_2 可分离，则分离 S_1,S_2 的超平面不一定唯一。分离的关键在于超平面的法向量 $\boldsymbol{p}\in R^n\setminus\{\boldsymbol{0}\}$ 的存在。

先建立闭凸集的一个性质。

定理 4.2.1 设 $S\subset R^n$ 是非空闭凸集，$\boldsymbol{y}\in R^n$。若 $\boldsymbol{y}\notin S$，则存在唯一的 $\bar{\boldsymbol{x}}\in S$，使

$$\|\bar{\boldsymbol{x}}-\boldsymbol{y}\|=\min_{\boldsymbol{x}\in S}\|\boldsymbol{x}-\boldsymbol{y}\|>0$$

离 S外一点 y 最近点。

***证明：** 存在性。设超球 $B(\boldsymbol{y},\beta)=\{\boldsymbol{x}\in R^n \mid \|\boldsymbol{x}-\boldsymbol{y}\|\le\beta\}$。取 $\beta>0$ 充分大，使 $D=S\bigcap B(\boldsymbol{y},\beta)\neq\varnothing$。显然，$D$ 是有界闭集，并且 $\min\limits_{\boldsymbol{x}\in S}\|\boldsymbol{x}-\boldsymbol{y}\|=\min\limits_{\boldsymbol{x}\in D}\|\boldsymbol{x}-\boldsymbol{y}\|$。由于 $f(\boldsymbol{x})=\|\boldsymbol{x}-\boldsymbol{y}\|$ 是连续函数，因此 $f(\boldsymbol{x})$ 在有界闭集 D 上存在最优解，即存在 $\bar{\boldsymbol{x}}\in D\subset S$，使 $\|\bar{\boldsymbol{x}}-\boldsymbol{y}\|=\min\limits_{\boldsymbol{x}\in D}\|\boldsymbol{x}-\boldsymbol{y}\|=\min\limits_{\boldsymbol{x}\in S}\|\boldsymbol{x}-\boldsymbol{y}\|$。由于 $\bar{\boldsymbol{x}}\in S$，$\boldsymbol{y}\notin S$，因此 $\|\bar{\boldsymbol{x}}-\boldsymbol{y}\|>0$。存在性得证。

唯一性。设 $\tilde{\boldsymbol{x}}\in S$，使 $\|\bar{\boldsymbol{x}}-\boldsymbol{y}\|=\|\tilde{\boldsymbol{x}}-\boldsymbol{y}\|=r>0$，记 $\hat{\boldsymbol{x}}=\frac{1}{2}(\tilde{\boldsymbol{x}}+\bar{\boldsymbol{x}})$，则 $\hat{\boldsymbol{x}}\in S$，因此

$$r\le\|\hat{\boldsymbol{x}}-\boldsymbol{y}\|\le\frac{1}{2}\|\bar{\boldsymbol{x}}-\boldsymbol{y}\|+\frac{1}{2}\|\tilde{\boldsymbol{x}}-\boldsymbol{y}\|=r$$

$$=\|\tfrac{1}{2}(\hat{x}+\bar{x})-y\|$$
$$=\|\tfrac{1}{2}\hat{x}-\tfrac{1}{2}y+\tfrac{1}{2}\bar{x}-\tfrac{1}{2}y\|$$
$$\le\tfrac{1}{2}\|\hat{x}-y\|+\tfrac{1}{2}\|\bar{x}-y\|$$
$$=r$$

由此得 $\bar{\boldsymbol{x}}-\boldsymbol{y}=\pm(\tilde{\boldsymbol{x}}-\boldsymbol{y})$。若 $\bar{\boldsymbol{x}}-\boldsymbol{y}=-(\tilde{\boldsymbol{x}}-\boldsymbol{y})$，则 $\boldsymbol{y}=\frac{1}{2}(\tilde{\boldsymbol{x}}+\bar{\boldsymbol{x}})\in S$，矛盾，因此 $\bar{\boldsymbol{x}}-\boldsymbol{y}=\tilde{\boldsymbol{x}}-\boldsymbol{y}$，即 $\bar{\boldsymbol{x}}=\tilde{\boldsymbol{x}}$。唯一性得证。

证毕。

根据定理 4.2.1 可得到点与闭凸集的分离定理。

定理 4.2.2 设 $S\subset R^n$ 是非空闭凸集，$\boldsymbol{y}\in R^n$。若 $\boldsymbol{y}\notin S$，则存在 $\boldsymbol{p}\in R^n\setminus\{\boldsymbol{0}\}$，$\alpha\in R$，使

$$\boldsymbol{p}^T\boldsymbol{x}<\alpha<\boldsymbol{p}^T\boldsymbol{y},\forall\boldsymbol{x}\in S$$

即 $H(\boldsymbol{p},\alpha)$ 严格分离 S 和 $\boldsymbol{y}$。

（考虑几何意义）

***证明：** 由定理 4.2.1 知，存在唯一的 $\bar{\boldsymbol{x}}\in S$，使

$$\|\bar{\boldsymbol{x}}-\boldsymbol{y}\|=\min_{\boldsymbol{x}\in S}\|\boldsymbol{x}-\boldsymbol{y}\|>0$$

令 $\boldsymbol{p}=\boldsymbol{y}-\bar{\boldsymbol{x}}$，则 $\boldsymbol{p}\neq\boldsymbol{0}$。对任意的 $\boldsymbol{x}\in S$，$\lambda\in(0,1)$，有 $\lambda\boldsymbol{x}+(1-\lambda)\bar{\boldsymbol{x}}\in S$，于是有

$$\|\bar{\boldsymbol{x}}-\boldsymbol{y}\|^2\le\|\lambda\boldsymbol{x}+(1-\lambda)\bar{\boldsymbol{x}}-\boldsymbol{y}\|^2=\|(\bar{\boldsymbol{x}}-\boldsymbol{y})+\lambda(\boldsymbol{x}-\bar{\boldsymbol{x}})\|^2$$

$$=\|\bar{\boldsymbol{x}}-\boldsymbol{y}\|^2+\lambda^2\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^2+2\lambda(\bar{\boldsymbol{x}}-\boldsymbol{y})^T(\boldsymbol{x}-\bar{\boldsymbol{x}})$$

$$\lambda\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^2+2(\bar{\boldsymbol{x}}-\boldsymbol{y})^T(\boldsymbol{x}-\bar{\boldsymbol{x}})\ge0$$

令 $\lambda\to0^+$，得 $(\bar{\boldsymbol{x}}-\boldsymbol{y})^T(\boldsymbol{x}-\bar{\boldsymbol{x}})\ge0$，即 $\boldsymbol{p}^T(\boldsymbol{x}-\bar{\boldsymbol{x}})\le0$。令 $\alpha=\boldsymbol{p}^T\frac{\boldsymbol{y}+\bar{\boldsymbol{x}}}{2}$（即让超平面经过 $\boldsymbol{y},\bar{\boldsymbol{x}}$ 的中点 $\frac{\boldsymbol{y}+\bar{\boldsymbol{x}}}{2}$），则对任意的 $\boldsymbol{x}\in S$，

$$\boldsymbol{p}^T\boldsymbol{x}-\alpha\le\boldsymbol{p}^T\bar{\boldsymbol{x}}-\boldsymbol{p}^T\frac{\boldsymbol{y}+\bar{\boldsymbol{x}}}{2}=\boldsymbol{p}^T\frac{\bar{\boldsymbol{x}}-\boldsymbol{y}}{2}=-\frac{1}{2}\boldsymbol{p}^T\boldsymbol{p}<0$$

又

$$\boldsymbol{p}^T\boldsymbol{y}-\alpha=\boldsymbol{p}^T\boldsymbol{y}-\boldsymbol{p}^T\frac{\boldsymbol{y}+\overline{\boldsymbol{x}}}{2}=\boldsymbol{p}^T\frac{\boldsymbol{y}-\overline{\boldsymbol{x}}}{2}=\frac{1}{2}\boldsymbol{p}^T\boldsymbol{p}>0$$

由此得结论。证毕。

推论 4.2.1 设 $C\subset R^n$ 是非空闭凸锥，$\boldsymbol{y}\in R^n$。若 $\boldsymbol{y}\notin C$，则存在 $\boldsymbol{p}\in R^n\setminus\{\boldsymbol{0}\}$，使

$$\boldsymbol{p}^T\boldsymbol{x}\le 0<\boldsymbol{p}^T\boldsymbol{y},\forall\boldsymbol{x}\in C$$

即 $H(\boldsymbol{p},0)$ 严格分离 C 和 $\boldsymbol{y}$。

***证明：** 根据定理 4.2.2，存在 $\boldsymbol{p}\in R^n\setminus\{\boldsymbol{0}\}$，$\alpha\in R$，使

$$\boldsymbol{p}^T\boldsymbol{x}<\alpha<\boldsymbol{p}^T\boldsymbol{y},\forall\boldsymbol{x}\in C$$

由 C 是锥知，对任意的 $\lambda>0$ 和 $\boldsymbol{x}\in C$，有 $\lambda\boldsymbol{x}\in C$，故有

$$\lambda\boldsymbol{p}^T\boldsymbol{x}<\alpha<\boldsymbol{p}^T\boldsymbol{y},\forall\lambda>0,\boldsymbol{x}\in C$$

若 $p^Tx>0$，则 $\lambda p^Tx\to+\infty$ ∴不符合 因此，$p^Tx\le0$.

由 $\lambda>0$ 可任意大得 $\boldsymbol{p}^T\boldsymbol{x}\le 0,\forall\boldsymbol{x}\in C$，再由 $\lambda\to 0^+$ 得 $0<\boldsymbol{p}^T\boldsymbol{y}$，由此得结论。证毕。

再建立一般凸集与点的分离定理。

定理 4.2.3 设 $S\subset R^n$ 是非空凸集，$\boldsymbol{y}\in R^n$。若 $\boldsymbol{y}\notin S$，则存在 $\boldsymbol{p}\in R^n\setminus\{\boldsymbol{0}\}$，使

$$\boldsymbol{p}^T\boldsymbol{x}\le\boldsymbol{p}^T\boldsymbol{y},\forall\boldsymbol{x}\in S$$

即 S 和 $\boldsymbol{y}$ 是可分离的。

***证明：** 若 $\boldsymbol{y}\notin clS$，由 clS 是非空闭凸集和定理 4.2.2 得结论。

若 $\boldsymbol{y}\in rbS$，S 凸知 $\boldsymbol{y}\in rbS=rb(clS)$，则由注 3.2.1 知，存在 $\boldsymbol{y}^k\notin clS,k=1,2,\cdots$，使 $\boldsymbol{y}^k\to\boldsymbol{y}$。对任意的 $k=1,2,\cdots$，由 $\boldsymbol{y}^k\notin clS$ 和定理 4.2.2 得，存在 $\boldsymbol{p}^k\in R^n:\left\|\boldsymbol{p}^k\right\|=1$，使

$$(\boldsymbol{p}^k)^T\boldsymbol{x}<(\boldsymbol{p}^k)^T\boldsymbol{y},\forall\boldsymbol{x}\in S$$

因为 $\{\boldsymbol{p}^k\}$ 有界，不妨设 $\boldsymbol{p}^k\to\boldsymbol{p}$，则 $\boldsymbol{p}\in R^n\setminus\{\boldsymbol{0}\}$，并且 $\boldsymbol{p}^T\boldsymbol{x}\le\boldsymbol{p}^T\boldsymbol{y},\forall\boldsymbol{x}\in S$。证毕。

现在建立凸集与凸集的分离定理。

定理 4.2.4 设 $S_1,S_2\subset R^n$ 是非空凸集。若 $S_1\cap S_2=\varnothing$，则存在 $\boldsymbol{p}\in R^n\setminus\{\boldsymbol{0}\}$，使

$$\boldsymbol{p}^T\boldsymbol{x}^1\le\boldsymbol{p}^T\boldsymbol{x}^2,\forall\boldsymbol{x}^1\in S_1,\boldsymbol{x}^2\in S_2$$

证明： 令 $S=S_1-S_2$，则 S 是非空凸集，并且 $\boldsymbol{0}\notin S$。由定理 4.2.3 知，存在 $\boldsymbol{p}\in R^n\setminus\{\boldsymbol{0}\}$，使

$$\boldsymbol{p}^T\boldsymbol{x}\le 0,\forall\boldsymbol{x}\in S=S_1-S_2$$

因为 S_1、S_2 不相交 因此不会有相同的点 即 $0\notin S$

由此得结论。证毕。

4.3 凸集分离定理应用

利用凸集分离定理，可以得到最优化理论中非常重要的几个择一性定理。

定理 4.3.1(Farkas 定理) 设 $A\in R^{m\times n}$，$\boldsymbol{c}\in R^n$，则下列两个系统有且仅有一个有解：

有且仅有一个 步骤不等式

(I) $\boldsymbol{Ax}\le\boldsymbol{0},\boldsymbol{c}^T\boldsymbol{x}>0$

(II) $\boldsymbol{A}^T\boldsymbol{y}=\boldsymbol{c},\boldsymbol{y}\ge\boldsymbol{0}$ ⇒ C可表示为 $a_1\cdots a_m$ 的非负线性组合

两个系统同时有解 ⇒ 矛盾
一个系统无解 ⇒ 另一个系统有解

（考虑几何意义！）

证明： 若(I)有解 $\overline{\boldsymbol{x}}$，即 $\boldsymbol{A}\overline{\boldsymbol{x}}\le\boldsymbol{0},\boldsymbol{c}^T\overline{\boldsymbol{x}}>0$，(II)有解 $\overline{\boldsymbol{y}}$，即 $\boldsymbol{A}^T\overline{\boldsymbol{y}}=\boldsymbol{c},\overline{\boldsymbol{y}}\ge\boldsymbol{0}$，则

$$0<\boldsymbol{c}^T\overline{\boldsymbol{x}}=\overline{\boldsymbol{x}}^TA^T\overline{\boldsymbol{y}}=\overline{\boldsymbol{y}}^TA\overline{\boldsymbol{x}}\le 0$$

$\bar{y}^TA\bar{x}=(\bar{x}^TA^Ty)^T$

矛盾，所以(I)和(II)不同时有解。

设(II)无解。记 $S=\{A^T y \mid y\geq 0\}$，则 $c\notin S$，并且 S 是闭凸锥。由推论 4.2.1 知，存在 $p\in R^n\setminus\{0\}$，使

$$(Ap)^T y = p^T A^T y \leq 0 < p^T c = c^T p, \forall y\geq 0$$

得 $c^T p>0$，并且由 $y\geq 0$ 的分量可任意大，得到 $Ap\leq 0$（否则，假若 Ap 的第 i 个分量 $(Ap)_i>0$，则取 y 的第 i 个分量 $y_i>0$ 充分大，那么由上式 $c^T p>(Ap)^T y=(Ap)_i y_i$=充分大，矛盾）。由此知 p 是(I)的解，即(I)存在解。

证毕。

① 假设两个都有解，矛盾
② (1)无解，证(2)有解
③ (2)无解，证(1)有解

定理 4.3.2(Gordan 定理) 设 $B\in R^{m\times n}$，则下列两个系统有且仅有一个有解：

（I） $Bx<0$

（II） $B^T y=0, y\geq 0, y\neq 0$

证明：若(I)有解 $\bar{x}$，即 $B\bar{x}<0$，(II)有解 $\bar{y}$，即 $B^T\bar{y}=0,\bar{y}\geq 0,\bar{y}\neq 0$，则 $0=\bar{x}^T B^T\bar{y}=(B\bar{x})^T\bar{y}<0$，矛盾，所以(I)和(II)不同时有解。

设(I)无解，要证(II)有解。

方法 1（利用凸集分离定理）：因(I)无解，则 $S_1\triangleq\{Bx\mid x\in R^n\}$ 与 $S_2\triangleq\{z\in R^m\mid z<0\}$ 不相交。显然，S_1,S_2 为凸集，则由定理 4.2.4，存在 $p\in R^m\setminus\{0\}$，使

$$p^T Bx\geq p^T z,\forall x\in R^n, z\in R^m, z<0$$

在上式中令 $x=0$ 得 $p^T z\leq 0,\forall z<0$，由 z 的每个分量可任意小知 $p\geq 0$。在上式中令 $z\to 0^-$ 得 $p^T Bx = x^T(B^T p)\geq 0,\forall x\in R^n$，则由 x 的任意性，得到 $B^T p=0$。由此知 p 是(II)的解，即(II)存在解。

方法 2（利用 Farkas 定理）：因(I)无解，则不存在 $x\in R^n$ 和 $z<0$，使 $Bx\leq ze$，其中 $e=(1,\cdots,1)^T\in R^n$，因此系统

因为 Farkas 条件中只有一个多项不等式
因此需将 Gordan 中的一个多项转化为非多项，
即引入 $z<0$，使 $Bx\leq ze$ 无解

$$\text{(III)}\quad (B,-e)\begin{pmatrix}x\\ z\end{pmatrix}\leq 0,\ \begin{pmatrix}0\\ -1\end{pmatrix}^T\begin{pmatrix}x\\ z\end{pmatrix}>0$$

无解。根据定理 4.3.1，系统 Farkas

$$\text{(IV)}\quad \begin{pmatrix}B^T\\ -e^T\end{pmatrix} y=\begin{pmatrix}0\\ -1\end{pmatrix}, y\geq 0$$

有解，即存在 $\bar{y}$，使 $B^T\bar{y}=0, e^T\bar{y}=1,\bar{y}\geq 0$，因此 $\bar{y}$ 是系统(II)的解，即系统(II)存在解。

证毕。

定理 4.3.3(Mortzkin 定理) 设 $A\in R^{m\times n}(m\geq 0)$，$B\in R^{p\times n}(p\geq 1)$，$C\in R^{l\times n}(l\geq 0)$，则下列两个系统有且仅有一个有解：

（I） $Ax\leq 0, Bx<0, Cx=0$

（II） $A^T u+B^T v+C^T w=0, u\geq 0, v\geq 0, v\neq 0$

证明：若(I)有解 $\bar{x}$，即 $A\bar{x}\leq 0, B\bar{x}<0, C\bar{x}=0$，(II)有解 $(\bar{u},\bar{v},\bar{w})$，即 $A^T\bar{u}+B^T\bar{v}+C^T\bar{w}=0,\bar{u}\geq 0,\bar{v}\geq 0,\bar{v}\neq 0$，则

$$0=\bar{x}^T(A^T\bar{u}+B^T\bar{v}+C^T\bar{w})=(A\bar{x})^T\bar{u}+(B\bar{x})^T\bar{v}+(C\bar{x})^T\bar{w}<0$$

矛盾，所以(I)和(II)不同时有解。

设(I)无解，则不存在 $x\in R^n$ 和 $z<0$，使 $\boldsymbol{A}\boldsymbol{x}\le\boldsymbol{0}$，$\boldsymbol{B}\boldsymbol{x}\le z\boldsymbol{e},\boldsymbol{C}\boldsymbol{x}\le\boldsymbol{0},-\boldsymbol{C}\boldsymbol{x}\le\boldsymbol{0}$，其中 $\boldsymbol{e}=(1,\cdots,1)^T\in R^n$，因此系统

$$\text{(III)}\quad \begin{pmatrix}\boldsymbol{A},\ \boldsymbol{0}\\ \boldsymbol{B},-\boldsymbol{e}\\ \boldsymbol{C},\ \boldsymbol{0}\\ -\boldsymbol{C},\boldsymbol{0}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\ z\end{pmatrix}\le\boldsymbol{0},\begin{pmatrix}0\\ -1\end{pmatrix}^T\begin{pmatrix}\boldsymbol{x}\\ z\end{pmatrix}>0$$

无解。根据定理 4.3.1，系统

$$\text{(IV)}\quad \begin{pmatrix}\boldsymbol{A}^T,\boldsymbol{B}^T,\boldsymbol{C}^T,-\boldsymbol{C}^T\\ \boldsymbol{0}^T,-\boldsymbol{e}^T,\boldsymbol{0}^T,\ \ \boldsymbol{0}^T\end{pmatrix}\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{v}\\ \boldsymbol{w}_1\\ \boldsymbol{w}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{0}\\ -1\end{pmatrix},\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{v}\\ \boldsymbol{w}_1\\ \boldsymbol{w}_2\end{pmatrix}\ge\boldsymbol{0}$$

有解，即存在 $(\overline{\boldsymbol{u}},\overline{\boldsymbol{v}},\overline{\boldsymbol{w}}_1,\overline{\boldsymbol{w}}_2)$，使 $\boldsymbol{A}^T\overline{\boldsymbol{u}}+\boldsymbol{B}^T\overline{\boldsymbol{v}}+\boldsymbol{C}^T(\overline{\boldsymbol{w}}_1-\overline{\boldsymbol{w}}_2)=\boldsymbol{0},\boldsymbol{e}^T\overline{\boldsymbol{v}}=1,\overline{\boldsymbol{u}}\ge\boldsymbol{0},\overline{\boldsymbol{v}}\ge\boldsymbol{0},\overline{\boldsymbol{w}}_1,\overline{\boldsymbol{w}}_2\ge\boldsymbol{0}$，因此 $(\overline{\boldsymbol{u}},\overline{\boldsymbol{v}},\overline{\boldsymbol{w}}_1-\overline{\boldsymbol{w}}_2)$ 是系统(II)的解，即系统(II)存在解。

证毕。

4.4 凸函数

定义 4.4.1 设集合 $S\subset R^n$ 是非空凸集，函数 $f:S\to R^1$。若对任意的 $\lambda\in(0,1)$，有

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

则称 f 是 S 上的凸函数。若对任意的 $\lambda\in(0,1)$，有

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

则称 f 是 S 上的严格凸函数。

若 $-f$ 是 S 上的(严格)凸函数，则称 f 是 S 上的(严格)凹函数。

一元凸函数有明显的几何意义：过函数图像上任何两点的直线段在函数图像段的上方或重合。

显然，严格凸函数一定是凸函数。

定理 4.4.1 设 f 是非空凸集 $S\subset R^n$ 上的凸函数，则对任意的自然数 k，有

$$f(\sum_{i=1}^k\lambda_i\boldsymbol{x}^i)\le\sum_{i=1}^k\lambda_i f(\boldsymbol{x}^i),\forall\boldsymbol{x}^i\in S,\lambda_i\ge 0,i=1,\cdots,k,\sum_{i=1}^k\lambda_i=1$$

定理 4.4.2 设 f_1 和 f_2 是非空凸集 $S\subset R^n$ 上的凸函数，$\alpha_1,\alpha_2\ge 0$，则 $\alpha_1 f_1(\boldsymbol{x})+\alpha_2 f_2(\boldsymbol{x})$，$\max\{f_1(\boldsymbol{x}),f_2(\boldsymbol{x})\}$ 也是 S 上的凸函数。

定理 4.4.3 设 f 是非空凸集 $S\subset R^n$ 上的凸函数，则对任意实数 c，水平集 $H_c=\{\boldsymbol{x}\in S\mid f(\boldsymbol{x})\le c\}$ 是凸集。

推论 4.4.1 设 $g_i(\boldsymbol{x}),i=1,\cdots,m$ 为 R^n 上凹函数，$h_j(\boldsymbol{x})=0,j=1,\cdots,l$ 为 R^n 是线性函数，则集合

$$S=\{\boldsymbol{x}\in R^n\mid g_i(\boldsymbol{x})\ge 0,i=1,\cdots,m,h_j(\boldsymbol{x})=0,j=1,\cdots,l\}$$

为凸集。

定理 4.4.4 设 $S\subset R^n$ 是非空开凸集，$f\in C^1(S)$。则

（1）f 是 S 上的凸函数的充要条件是，

$$f(\boldsymbol{x}^2)-f(\boldsymbol{x}^1)\geq\nabla f(\boldsymbol{x}^1)^T(\boldsymbol{x}^2-\boldsymbol{x}^1),\forall\boldsymbol{x}^1,\boldsymbol{x}^2\in S$$

（2）f 是 S 上的严格凸函数的充要条件是，

$$f(\boldsymbol{x}^2)-f(\boldsymbol{x}^1)>\nabla f(\boldsymbol{x}^1)^T(\boldsymbol{x}^2-\boldsymbol{x}^1),\forall\boldsymbol{x}^1,\boldsymbol{x}^2\in S,\boldsymbol{x}^1\neq\boldsymbol{x}^2$$

证明：（1）必要性。设 f 是 S 上的凸函数，则对任意的 $\lambda\in(0,1)$ 和 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S$，有

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

因此再由 Taylor 展开即定理 3.3.1，

$$\begin{aligned}f(\boldsymbol{x}^2)-f(\boldsymbol{x}^1)&\geq\frac{f(\lambda\boldsymbol{x}^2+(1-\lambda)\boldsymbol{x}^1)-f(\boldsymbol{x}^1)}{\lambda}\\&=\frac{f(\boldsymbol{x}^1+\lambda(\boldsymbol{x}^2-\boldsymbol{x}^1))-f(\boldsymbol{x}^1)}{\lambda}=\nabla f(\boldsymbol{x}^1)^T(\boldsymbol{x}^2-\boldsymbol{x}^1)+\frac{o(\lambda)}{\lambda}\end{aligned}$$

令 $\lambda\to0^+$，得 $f(\boldsymbol{x}^2)-f(\boldsymbol{x}^1)\geq\nabla f(\boldsymbol{x}^1)^T(\boldsymbol{x}^2-\boldsymbol{x}^1)$。

充分性。任取 $\lambda\in(0,1)$ 和 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S$，令 $\boldsymbol{x}=\lambda\boldsymbol{x}^1+(1-\lambda)\boldsymbol{x}^2$，则 $\boldsymbol{x}\in S$，并且由条件，

$$f(\boldsymbol{x}^1)\geq f(\boldsymbol{x})+\nabla f(\boldsymbol{x})^T(\boldsymbol{x}^1-\boldsymbol{x}),\quad f(\boldsymbol{x}^2)\geq f(\boldsymbol{x})+\nabla f(\boldsymbol{x})^T(\boldsymbol{x}^2-\boldsymbol{x})$$

由此得

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

即 f 是 S 上的凸函数。

（2）必要性。任取 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S,\boldsymbol{x}^1\neq\boldsymbol{x}^2$，令 $\boldsymbol{x}=\frac{1}{2}(\boldsymbol{x}^1+\boldsymbol{x}^2)$，则 $\boldsymbol{x}\in S$，并且由(1)知，

$$f(\boldsymbol{x})\geq f(\boldsymbol{x}^1)+\nabla f(\boldsymbol{x}^1)^T(\boldsymbol{x}-\boldsymbol{x}^1)=f(\boldsymbol{x}^1)+\frac{1}{2}\nabla f(\boldsymbol{x}^1)^T(\boldsymbol{x}^2-\boldsymbol{x}^1)$$

又由定义知，$f(\boldsymbol{x})<\frac{1}{2}f(\boldsymbol{x}^1)+\frac{1}{2}f(\boldsymbol{x}^2)$，于是

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

即 $f(\boldsymbol{x}^2)-f(\boldsymbol{x}^1)>\nabla f(\boldsymbol{x}^1)^T(\boldsymbol{x}^2-\boldsymbol{x}^1)$。

充分性。与(1)同理。证毕。

定理 4.4.5　设 $S\subset R^n$ 是非空开凸集，$f\in C^2(S)$。则

（1）f 是 S 上的凸函数的充要条件是，对任意的 $\boldsymbol{x}\in S$，$\nabla^2f(\boldsymbol{x})$ 半正定；

（2）当对任意的 $\boldsymbol{x}\in S$，$\nabla^2f(\boldsymbol{x})$ 正定时，f 是 S 上的严格凸函数；

（3）当 f 是二次函数时，f 是 S 上的严格凸函数的充要条件是，对任意的 $\boldsymbol{x}\in S$，$\nabla^2f(\boldsymbol{x})$ 正定。

证明：（1）必要性。任取 $\boldsymbol{x}\in S$，$\boldsymbol{d}\in R^n$。由 S 是开集知，λ>0 充分小时，$\boldsymbol{x}+\lambda\boldsymbol{d}\in S$，并且由定理 4.4.4 知，

$$f(\boldsymbol{x}+\lambda\boldsymbol{d})-f(\boldsymbol{x})\geq\lambda\nabla f(\boldsymbol{x})^T\boldsymbol{d}，\lambda\text{ 充分小时}$$

又由 Taylor 展开即定理 3.3.1，$f(\boldsymbol{x}+\lambda\boldsymbol{d})-f(\boldsymbol{x})=\lambda\nabla f(\boldsymbol{x})^T\boldsymbol{d}+\frac{1}{2}\lambda^2\boldsymbol{d}^T\nabla^2f(\boldsymbol{x})\boldsymbol{d}+o(\lambda^2)$，于是

$$\frac{1}{2}\lambda^2\boldsymbol{d}^T\nabla^2f(\boldsymbol{x})\boldsymbol{d}+o(\lambda^2)\geq0，\lambda\text{ 充分小时}$$

即λ充分小时$\boldsymbol{d}^T\nabla^2 f(\boldsymbol{x})\boldsymbol{d}+\frac{o(\lambda^2)}{\lambda^2}\ge 0$，令$\lambda\to 0$得，$\boldsymbol{d}^T\nabla^2 f(\boldsymbol{x})\boldsymbol{d}\ge 0$。由此知$\nabla^2 f(\boldsymbol{x})$半正定。

充分性。任取$\boldsymbol{x}^1,\boldsymbol{x}^2\in S$。由 Taylor 展开，

$$f(\boldsymbol{x}^2)=f(\boldsymbol{x}^1)+\nabla f(\boldsymbol{x}^1)^T(\boldsymbol{x}^2-\boldsymbol{x}^1)+\frac{1}{2}(\boldsymbol{x}^2-\boldsymbol{x}^1)^T\nabla^2 f(\tilde{\boldsymbol{x}})(\boldsymbol{x}^2-\boldsymbol{x}^1)$$

其中$\tilde{\boldsymbol{x}}=\boldsymbol{x}^1+\theta(\boldsymbol{x}^2-\boldsymbol{x}^1)\in S$，$0<\theta<1$，由条件知$(\boldsymbol{x}^2-\boldsymbol{x}^1)^T\nabla^2 f(\tilde{\boldsymbol{x}})(\boldsymbol{x}^2-\boldsymbol{x}^1)\ge 0$，因此

$$f(\boldsymbol{x}^2)\ge f(\boldsymbol{x}^1)+\nabla f(\boldsymbol{x}^1)^T(\boldsymbol{x}^2-\boldsymbol{x}^1)$$

由定理 4.4.4 知f是S上的凸函数。

（2）同理（注意反之不成立）。

（3）由（2）知只需证明必要性。同(1)必要性证明，任取$\boldsymbol{x}\in S$，$\boldsymbol{d}\in R^n\setminus\{0\}$。由$S$是开集知，$\lambda>0$充分小时$\boldsymbol{x}+\lambda\boldsymbol{d}\in S$，并且由定理 4.4.4 知，

$$f(\boldsymbol{x}+\lambda\boldsymbol{d})-f(\boldsymbol{x})>\lambda\nabla f(\boldsymbol{x})^T\boldsymbol{d}\text{，}\lambda\text{ 充分小时}$$

又由 Taylor 展开即定理 3.3.1 和f是二次函数，

$$f(\boldsymbol{x}+\lambda\boldsymbol{d})-f(\boldsymbol{x})=\lambda\nabla f(\boldsymbol{x})^T\boldsymbol{d}+\frac{1}{2}\lambda^2\boldsymbol{d}^T\nabla^2 f(\boldsymbol{x})\boldsymbol{d}$$

于是$\frac{1}{2}\lambda^2\boldsymbol{d}^T\nabla^2 f(\boldsymbol{x})\boldsymbol{d}>0$，即$\boldsymbol{d}^T\nabla^2 f(\boldsymbol{x})\boldsymbol{d}>0$，由此知$\nabla^2 f(\boldsymbol{x})$正定。

证毕。

例 4.4.1　$f(x_1,x_2)=2x_1^2+x_2^2-2x_1x_2+x_1+1$

解：$\nabla f(x_1,x_2)=\begin{pmatrix}4x_1-2x_2+1\\2x_2-2x_1\end{pmatrix}$，$\nabla^2 f(x_1,x_2)=\begin{pmatrix}4&-2\\-2&2\end{pmatrix}$正定，故$f(x_1,x_2)$严格凸函数。

4.5　凸规划

考虑数学规划问题：

$$\begin{aligned}&\min f(\boldsymbol{x})\\&s.t.\ \boldsymbol{x}\in S\end{aligned}\tag{P}$$

其中约束集$S\subset R^n$，目标函数$f:S\to R$。

定义 4.5.1　设集合$X\subset R^n$是非空凸集，(P)的约束集$S\subset X$是非空凸集，目标函数$f:X\to R$是X上的凸函数，则称问题(P)为凸规划问题。

对于约束优化问题：

$$\begin{aligned}&\min f(\boldsymbol{x})\\&s.t.\ g_i(\boldsymbol{x})\ge 0,i=1,\cdots,m\\&\qquad h_j(\boldsymbol{x})=0,j=1,\cdots,l\end{aligned}\tag{NP}$$

由定理推论 4.4.1 知，当$X\subset R^n$是非空凸集，$f(\boldsymbol{x})$为X上的凸函数，$g_i(\boldsymbol{x}),i=1,\cdots,m$为$X$上的凹函数，$h_j(\boldsymbol{x}),j=1,\cdots,l$为$X$上的线性函数，则(NP)为凸规划问题。

定理 4.5.1　设(P)是凸规划，$\boldsymbol{x}^*\in S$。若$\boldsymbol{x}^*$是(P)的局部最优解，则$\boldsymbol{x}^*$是(P)的整体最优解，并且(P)的最优解集是凸集。

证明：由$\boldsymbol{x}^*$是(P)的局部最优解知，存在$\boldsymbol{x}^*$的δ-邻域$N(\boldsymbol{x}^*,\delta)$，使

$$f(\boldsymbol{x}^*) \le f(\boldsymbol{x}), \forall \boldsymbol{x} \in S \bigcap N(\boldsymbol{x}^*, \delta) \tag{4.5.1}$$

假设 $\boldsymbol{x}^*$ 不是(P)的整体最优解，则存在 $\tilde{\boldsymbol{x}} \in S$，使 $f(\tilde{\boldsymbol{x}}) < f(\boldsymbol{x}^*)$。因为 f 是 S 上的凸函数，所以对任意的 $\lambda \in (0,1)$，有

$$f(\lambda \tilde{\boldsymbol{x}} + (1-\lambda)\boldsymbol{x}^*) \le \lambda f(\tilde{\boldsymbol{x}}) + (1-\lambda) f(\boldsymbol{x}^*) < f(\boldsymbol{x}^*) \tag{4.5.2}$$

由于当 $\lambda > 0$ 充分小时，$\lambda \tilde{\boldsymbol{x}} + (1-\lambda)\boldsymbol{x}^* \in N(\boldsymbol{x}^*, \delta)$，由此知(4.5.2)与(4.5.1)矛盾。

由于(P)的最优解集可表示为水平集 $\{\boldsymbol{x} \in S \mid f(\boldsymbol{x}) \le f^*\}$，其中 f^* 是(P)的最优值，又定理 4.4.3 得结论。

证毕。

定理 4.5.2　设(P)是凸规划，并且 f 是 S 上的严格凸函数，$\boldsymbol{x}^* \in S$。若 $\boldsymbol{x}^*$ 是(P)的局部最优解，则 $\boldsymbol{x}^*$ 是(P)的严格整体最优解。

证明：同样有(4.5.1)成立。假设 $\boldsymbol{x}^*$ 不是(P)的严格整体最优解，则存在 $\tilde{\boldsymbol{x}} \in S \setminus \{\boldsymbol{x}^*\}$，使 $f(\tilde{\boldsymbol{x}}) \le f(\boldsymbol{x}^*)$。因为 f 是 S 上的严格凸函数，所以对任意的 $\lambda \in (0,1)$，有

$$f(\lambda \tilde{\boldsymbol{x}} + (1-\lambda)\boldsymbol{x}^*) < \lambda f(\tilde{\boldsymbol{x}}) + (1-\lambda) f(\boldsymbol{x}^*) \le f(\boldsymbol{x}^*) \tag{4.5.3}$$

由于当 $\lambda > 0$ 充分小时，$\lambda \tilde{\boldsymbol{x}} + (1-\lambda)\boldsymbol{x}^* \in N(\boldsymbol{x}^*, \delta)$，由此知(4.5.3)与(4.5.1)矛盾。证毕。