# 最优化方法

Methods of Optimization

# 目录

# 第一章　基础知识

## 1.1　数学规划

规划论，又称数学规划，是运筹学的分支，研究在约束条件下通过分配资源寻求目标函数极值的数学方法，应用于经济管理、工程设计和过程控制等领域. 其核心是建立约束条件与目标函数的数学模型，这种数学模型的一般形式为

$$\begin{aligned} \text{Opt.} \quad & f(x_i, y_j, \xi_k) \\ \text{s.t.} \quad & g_h(x_i, y_j, \xi_k) \leq (=, \geq) 0 \\ & h = 1, 2, \cdots, m \end{aligned}$$

其中，$x_i$ 是控制变量，$y_j$ 是已知参数，$\xi_k$ 是随机因素，$f$ 是目标函数，$g_h$ 是约束函数.

按照目标函数和自变量取值范围的不同，数学规划可分为：

- 线性规划：$f, g$ 都是线性函数；
- 非线性规划：$f, g$ 中有非线性函数；
- 多目标规划：$f$ 是向量函数；
- 整数规划：决策变量 $x_i$ 为整数；
- 动态规划：含多阶段决策过程；
- 随机规划：含有随机因子.

若记

$$S = \{\boldsymbol{x} \mid g_h(x_i, y_j, \xi_k) \leq (=, \geq) 0\} \subseteq \mathbb{R}^n$$

为约束集合，则这种数学模型可进一步简记为

$$(f\ S) \begin{cases} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in S \end{cases}$$

其中 $\boldsymbol{x} \in S$ 称模型 $(f\ S)$ 的可行解.

**定义 1.1.（全局最优解、最优值）**

若对于 $\boldsymbol{x}^* \in S$，满足 $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$，$\forall \boldsymbol{x} \in S$，则称 $\boldsymbol{x}^*$ 为 $(f\ S)$ 的全局最优解（最优解），记作 g.opt.，简记为 opt.. 称此时的 $f(\boldsymbol{x}^*)$ 为 $(f\ S)$ 的最优值（最优目标函数值）.

**定义 1.2.（局部最优解）**

若对于 $\boldsymbol{x}^* \in S$，$\exists \boldsymbol{x}^*$ 的某邻域 $U(\boldsymbol{x}^*)$，使得 $f(\boldsymbol{x}^*) \le f(\boldsymbol{x})$，$\forall \boldsymbol{x} \in S$，则称 $\boldsymbol{x}^*$ 为 $(f\ S)$ 的局部最优解，记作 l.opt.，

在上述定义中，若当 $\boldsymbol{x} \neq \boldsymbol{x}^*$ 时有严格不等式成立，则分别称 $\boldsymbol{x}^*$ 为 $(f\ S)$ 的严格全局最优解和严格局部最优解.

## 1.2　基本概念和符号

接下来将约定一些术语的符号表示：

- $\mathbb{R}^n$——$n$ 维欧式空间；
- $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^\top\in\mathbb{R}^n$——表示 $\mathbb{R}^n$ 中的一个点或一个向量，其中分量 $x_i\in\mathbb{R}$；
- $\boldsymbol{d}=(d_1,d_2,\cdots,d_n)^\top\in\mathbb{R}^n$，其中 $\boldsymbol{d}\neq\boldsymbol{0}$——表示从 $\boldsymbol{0}$ 指向 $\boldsymbol{d}$ 的方向；
- $\boldsymbol{x}+\lambda\boldsymbol{d}$——表示从 $\boldsymbol{x}$ 点出发，沿 $d$ 方向移动 $\lambda\boldsymbol{d}$ 长度得到的点.

下面是对欧式空间中向量的运算的定义.

**定义 1.3.（向量的运算）**

设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^\top\in\mathbb{R}^n$，$\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^\top\in\mathbb{R}^n$，则

（1）**向量的内积运算**：$\boldsymbol{x}^\top\boldsymbol{y}=\boldsymbol{y}^\top\boldsymbol{x}=\sum\limits_{i=1}^n x_iy_i$；

（2）**向量长度**：$\|\boldsymbol{x}\|=\sqrt{\boldsymbol{x}^\top\boldsymbol{x}}=\sqrt{\sum\limits_{i=1}^n x_i^2}$；

（3）**两点间距离**：$\|\boldsymbol{x}-\boldsymbol{y}\|$.

**定理 1.1.（三角不等式）**

$\forall\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$，有

$$\|\boldsymbol{x}+\boldsymbol{y}\|\leq\|\boldsymbol{x}\|+\|\boldsymbol{y}\| \tag{1.1}$$

**【证明】**对左式和右式分别平方后相减，有

$$\begin{aligned}
\text{左式}^2-\text{右式}^2&=(\boldsymbol{x}+\boldsymbol{y})^\top(\boldsymbol{x}+\boldsymbol{y})-(\boldsymbol{x}^\top\boldsymbol{x}+\boldsymbol{y}^\top\boldsymbol{y}+2\|\boldsymbol{x}\|\|\boldsymbol{y}\|)\\
&=(\boldsymbol{x}^\top\boldsymbol{x}+\boldsymbol{y}^\top\boldsymbol{y}+2\boldsymbol{x}^\top\boldsymbol{y})-(\boldsymbol{x}^\top\boldsymbol{x}+\boldsymbol{y}^\top\boldsymbol{y}+2\|\boldsymbol{x}\|\|\boldsymbol{y}\|)\\
&=2(\boldsymbol{x}^\top\boldsymbol{y}-\|\boldsymbol{x}\|\|\boldsymbol{y}\|)
\end{aligned}$$

也即只需证明柯西不等式

$$\boldsymbol{x}^\top\boldsymbol{y}\leq\|\boldsymbol{x}\|\|\boldsymbol{y}\|$$

成立即可. 于是我们构造关于 $t$ 的方程

$$\|\boldsymbol{x}+t\boldsymbol{y}\|^2=\boldsymbol{y}^\top\boldsymbol{y}t^2+2\boldsymbol{x}^\top\boldsymbol{y}t+\boldsymbol{x}^\top\boldsymbol{x}=0$$

显然由于向量的模长一定非负，故该方程至多有一个解，也即该二次方程的判别式

$$\Delta=4(\boldsymbol{x}^\top\boldsymbol{y})^2-4(\boldsymbol{x}^\top\boldsymbol{x})(\boldsymbol{y}^\top\boldsymbol{y})\leq 0$$

从而柯西不等式得证，从而原不等式得证. 【证毕】

**定义 1.4.（点列的收敛）**

设 $\{\boldsymbol{x}^{(n)}\}$ 是定义在 $\mathbb{R}^n$ 上的点列，若 $\exists\boldsymbol{x}\in\mathbb{R}^n$，使得 $\lim\limits_{n\to\infty}\|\boldsymbol{x}^{(n)}-\boldsymbol{x}\|=0$，则称点列 $\{\boldsymbol{x}^{(n)}\}$ 收敛到 $\boldsymbol{x}$，记作

$$\lim_{n\to\infty}\boldsymbol{x}^{(n)}=\boldsymbol{x}$$

**定义 1.5.（向量的大小关系）**

设 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$，若 $\forall i \in \{1, 2, \cdots, n\}$，都有 $x_i \leq y_i$，则称 $\boldsymbol{x} \leq \boldsymbol{y}$. 类似可规定 $\boldsymbol{x} \geq \boldsymbol{y}$，$\boldsymbol{x} < \boldsymbol{y}$，$\boldsymbol{x} > \boldsymbol{y}$.

显然，在如上定义中，$(\mathbb{R}^n, \leq)$ 并不是全序集（证明略）. 特别地，$\boldsymbol{x} \leq \boldsymbol{0} \Leftrightarrow \forall i \in \{1, 2, \cdots, n\}, x_i \leq 0$.

**例 1.1** 设 $\boldsymbol{x}, \boldsymbol{y} \in R^n$，若 $\forall \boldsymbol{y} \geq \boldsymbol{0}$，都有 $\boldsymbol{x}^\top \boldsymbol{y} \leq \alpha$，则有 $\boldsymbol{x} \leq \boldsymbol{0}$ 且 $\alpha \geq 0$.

**【证明】**取 $\boldsymbol{y} = \boldsymbol{0}$，则 $\boldsymbol{x}^\top \boldsymbol{y} = 0$，从而 $\boldsymbol{x}^\top \boldsymbol{y}$ 的上界 $\alpha \geq 0$. 假设 $\boldsymbol{x} \leq 0$ 不成立，则 $\exists i \in \{1, 2, \cdots, n\}$，使得 $x_i > 0$. 令 $y_i \to +\infty$，则有 $\boldsymbol{x}^\top \boldsymbol{y} \to +\infty$，从而与 $\boldsymbol{x}^\top \boldsymbol{y}$ 有上界 $\alpha$ 矛盾，故必有 $\boldsymbol{x} \leq \boldsymbol{0}$. 【证毕】

该命题还有以下等价表述：

- 若 $\forall \boldsymbol{y} \leq \boldsymbol{0}$，都有 $\boldsymbol{x}^\top \boldsymbol{y} \leq \alpha$，则有 $\boldsymbol{x} \geq \boldsymbol{0}$ 且 $\alpha \geq 0$;
- 若 $\forall \boldsymbol{y} \geq \boldsymbol{0}$，都有 $\boldsymbol{x}^\top \boldsymbol{y} \geq \alpha$，则有 $\boldsymbol{x} \geq \boldsymbol{0}$ 且 $\alpha \leq 0$;
- 若 $\forall \boldsymbol{y} \leq \boldsymbol{0}$，都有 $\boldsymbol{x}^\top \boldsymbol{y} \geq \alpha$，则有 $\boldsymbol{x} \leq \boldsymbol{0}$ 且 $\alpha \leq 0$.

## 1.3　子空间、正交子空间

**定义 1.6.（子空间）**

设 $\boldsymbol{d}^{(i)} \in \mathbb{R}^n, i = 1, 2, \cdots, m$ 是线性无关的向量组，称集合 $\{\boldsymbol{x} \mid \boldsymbol{x} = \sum_{j=1}^{m} \alpha_j \boldsymbol{d}^{(j)}, \alpha_J \in \mathbb{R}\}$ 是由向量 $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(m)}$ 生成的**子空间**，记作 $L(\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(m)})$，简记作 $L$.

可以看到，$\forall L \subseteq \mathbb{R}^n$，都有 $\boldsymbol{0} \in L$. 若令 $\boldsymbol{A} = (\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(m)}) \in \mathbb{R}^{n\times m}$，$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_m)^\top \in \mathbb{R}^m$，则

$$\boldsymbol{x} \in L \Leftrightarrow \exists \boldsymbol{\alpha} \in \mathbb{R}^m, \boldsymbol{x} = \boldsymbol{A}\boldsymbol{\alpha}$$

并且 $\mathrm{Rank}(\boldsymbol{A}) = m$，也就是说 $\boldsymbol{A}$ 是**列满秩**的. 我们将会在接下来的证明中用到这个性质.

**定义 1.7.（正交子空间）**

称集合 $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x}^\top \boldsymbol{y} = 0, \forall \boldsymbol{y} \in L\}$ 为子空间 $L$ 的**正交子空间**，记作 $L^\perp$.

不难看出 $\forall L \subseteq \mathbb{R}^n$，都有 $\boldsymbol{0} \in L^\top$. 且有

$$\begin{aligned} \boldsymbol{x} \in L^\perp &\Leftrightarrow \forall \boldsymbol{y} \in L, \boldsymbol{x}^\top \boldsymbol{y} = 0 \\ &\Leftrightarrow \forall \boldsymbol{\alpha} \in \mathbb{R}^m, \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{\alpha} = 0 \end{aligned}$$

**定理 1.2.（子空间投影定理）**

设 $L$ 为 $\mathbb{R}^n$ 的子空间，则 $\forall \boldsymbol{z} \in \mathbb{R}^n$，则必有以下两个命题成立：

（1）$\exists ! \boldsymbol{x} \in L, \boldsymbol{y} \in L^\perp$，使得 $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$；

（2）在（1）中的 $\boldsymbol{x}$ 为问题

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

的唯一解，且该问题的最优值为（1）中的 $\|\boldsymbol{y}\|$.

**【证明】**（1）先证明存在性. 我们从线性方程近似解的求法得到灵感. （2）中问题的目标实质上是求线性方程 $\boldsymbol{A}\boldsymbol{\alpha} = \boldsymbol{z}$ 在 $\mathbb{R}^m$ 中的近似解，为此我们将左右两边左乘 $\boldsymbol{A}^\top$，于是得到正规方程

$$\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{\alpha} = \boldsymbol{A}^\top \boldsymbol{z}$$

由于 $\boldsymbol{A}$ 是列满秩的，所以 $\boldsymbol{A}^\top \boldsymbol{A}$ 是可逆的，所以 $\boldsymbol{\alpha} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \boldsymbol{z}$，所以取

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{\alpha} = \boldsymbol{A}(\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \boldsymbol{z}$$

接下来只需证 $\boldsymbol{z} - \boldsymbol{A}(\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \boldsymbol{z} \in L^\perp$. 由于

$$\begin{aligned} & (\boldsymbol{z} - \boldsymbol{A}(\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \boldsymbol{z})^\top \boldsymbol{A}\boldsymbol{\beta} \\ =& \boldsymbol{z}^\top \boldsymbol{A}\boldsymbol{\beta} - \boldsymbol{z}^\top \boldsymbol{A}(\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{\beta} \\ =& \boldsymbol{z}^\top \boldsymbol{A}\boldsymbol{\beta} - \boldsymbol{z}^\top \boldsymbol{A}\boldsymbol{E}\boldsymbol{\beta} \\ =& \boldsymbol{0} \end{aligned}$$

$\forall \boldsymbol{\beta} \in \mathbb{R}^m$ 都成立，所以 $\boldsymbol{z}-\boldsymbol{A}(\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top\boldsymbol{z} \in L^\top$，也即取 $y=\boldsymbol{z}-\boldsymbol{A}(\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top\boldsymbol{z} \in L^\top$ 即可. 存在性证毕.

接下来证明唯一性. 假设在子空间 $L$ 中，$\exists \boldsymbol{p} \neq \boldsymbol{x}$，使得 $\exists \boldsymbol{q} \in L^\perp$，$\boldsymbol{z}=\boldsymbol{p}+\boldsymbol{q}$，则有 $\boldsymbol{q}=\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{p}$，但是

$$\begin{aligned}\boldsymbol{x}^\top\boldsymbol{q} &= \boldsymbol{x}^\top(\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{p}) \\ &= \boldsymbol{x}^\top\boldsymbol{x}-\boldsymbol{x}^\top\boldsymbol{q} \\ &\neq 0\end{aligned}$$

与 $\exists \boldsymbol{q} \in L^\perp$ 矛盾，所以假设不成立，唯一性得证.

（2）考查函数 $f(t)=\|\boldsymbol{z}-\boldsymbol{x}-t\boldsymbol{v}\|^2$，$\forall \boldsymbol{v} \in L$ 且 $\boldsymbol{v} \neq \boldsymbol{0}$，则要证明目标问题的唯一解是 $\boldsymbol{x}$，即证明 $f(t)$ 取最小值当且仅当 $t=0$，也即证明 $\forall v \in L$，$f'(0)=0$，且 $\forall t \neq 0$，$f'(t) \neq 0$，而

$$\begin{aligned}f'(t) &= -2\boldsymbol{y}^\top\boldsymbol{v}+2t\boldsymbol{v}^\top\boldsymbol{v} \\ &= 2t\boldsymbol{v}^\top\boldsymbol{v}\end{aligned}$$

所以上述命题成立，从而目标问题的唯一解就是（1）中的 $\boldsymbol{x}$，代入 $\boldsymbol{u}=\boldsymbol{x}$ 即得到目标问题最优值为 $\|\boldsymbol{y}\|$. 【证毕】

**例 1.2** 设 $\boldsymbol{x},\boldsymbol{y} \in R^n$，若 $\forall \boldsymbol{y} \in L$，都有 $\boldsymbol{x}^\top\boldsymbol{y} \leq \alpha$，则有 $\boldsymbol{x} \in L^\perp$ 且 $\alpha \geq 0$.

**【证明】**取 $\boldsymbol{y}=\boldsymbol{0}$，则 $\boldsymbol{x}^\top\boldsymbol{y}=0$，从而 $\boldsymbol{x}^\top\boldsymbol{y}$ 的上界 $\alpha \geq 0$. 假设 $\boldsymbol{x} \notin L^\perp$，则由子空间投影定理的（1），$\exists \boldsymbol{x}^{(1)} \in L$，$\boldsymbol{x}^{(2)} \in L^\perp$，使得 $\boldsymbol{x}=\boldsymbol{x}^{(1)}+\boldsymbol{x}^{(2)}$，那么

$$\begin{aligned}\boldsymbol{x}^\top\boldsymbol{y} &= (\boldsymbol{x}^{(1)}+\boldsymbol{x}^{(2)})^\top\boldsymbol{y} \\ &= \boldsymbol{x}^{(1)\top}\boldsymbol{y}+\boldsymbol{x}^{(2)\top}\boldsymbol{y} \\ &= \boldsymbol{x}^{(1)\top}\boldsymbol{y}\end{aligned}$$

令 $\boldsymbol{y}=\lambda\boldsymbol{x}^{(1)}$ 且 $\lambda \to +\infty$，则 $\boldsymbol{x}^{(1)\top}\boldsymbol{y} \to +\infty$，与 $\boldsymbol{x}^\top\boldsymbol{y}$ 有上界 $\alpha$ 矛盾，故必有 $\boldsymbol{x} \in L^\perp$.

【证毕】

该命题还有以下等价表述：

- 若 $\forall \boldsymbol{y} \in L$，都有 $\boldsymbol{x}^\top\boldsymbol{y} \geq \alpha$，则有 $\boldsymbol{x} \in L^\perp$ 且 $\alpha \leq 0$.

## 1.4　多元函数及其导数

多元函数 $f:\mathbb{R}^n\to\mathbb{R}$ 可看作一个 $n$ 维向量 $\boldsymbol{x}$ 的函数 $f(\boldsymbol{x})$. 受篇幅所限，我们并不给出矩阵导数的定义，具体请参见《矩阵分析》课程相关内容. 在此我们只不加证明地给出一些常用函数的矩阵导数（即其梯度）.

**定理 1.3.（一些常见函数的梯度）**

- 线性函数：$f(\boldsymbol{x})=\boldsymbol{c}^\top\boldsymbol{x}+\boldsymbol{d}$，$\nabla f(\boldsymbol{x})=\boldsymbol{c}$，$\nabla^2 f(x)=\boldsymbol{0}$；
- 二次函数：$f(\boldsymbol{x})=\dfrac{1}{2}\boldsymbol{x}^\top\boldsymbol{Q}\boldsymbol{x}+\boldsymbol{c}^\top\boldsymbol{x}+\boldsymbol{d}$，$\nabla f(\boldsymbol{x})=\boldsymbol{Q}\boldsymbol{x}+\boldsymbol{c}$，$\nabla^2 f(x)=\boldsymbol{Q}$，其中 $\boldsymbol{Q}$ 为对称阵.

接下来依旧不加证明地介绍多元函数的 Taylor 展开式及中值定理. 此处不加证明是因为此内容可以被读者在任何一本《数学分析》教材中找到.

**定理 1.4.（Taylor 展开式）**

设 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $\boldsymbol{x}^*$ 的某邻域内二阶可导，则有：

- 一阶 Taylor 展开式：$f(\boldsymbol{x})=f(\boldsymbol{x}^*)+\nabla f^\top(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*)+o(\|\boldsymbol{x}-\boldsymbol{x}^*\|)$；
- 二阶 Taylor 展开式：$f(\boldsymbol{x})=f(\boldsymbol{x}^*)+\nabla f^\top(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*)+\dfrac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^\top\nabla^2 f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*)+o(\|\boldsymbol{x}-\boldsymbol{x}^*\|^2)$.

**定理 1.5.（中值定理）**

设 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $\boldsymbol{x}^*$ 的某邻域 $U(\boldsymbol{x}^*)$ 内二阶可导，则 $\forall x\in U(\boldsymbol{x}^*)$，都有：

- Lagrange 中值定理：$\exists\lambda\in(0,1)$，记 $\boldsymbol{x}_\lambda=\boldsymbol{x}^*+\lambda(\boldsymbol{x}-\boldsymbol{x}^*)$，则有 $f(\boldsymbol{x})=f(\boldsymbol{x}^*)+\nabla f^\top(\boldsymbol{x}_\lambda)(\boldsymbol{x}-\boldsymbol{x}^*)$；
- Taylor 中值定理：$\exists\mu\in(0,1)$，记 $\boldsymbol{x}_\mu=\boldsymbol{x}^*+\mu(\boldsymbol{x}-\boldsymbol{x}^*)$，则有 $f(\boldsymbol{x})=f(\boldsymbol{x}^*)+\nabla f^\top(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*)+\dfrac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^\top\nabla^2 f(\boldsymbol{x}_\mu)(\boldsymbol{x}-\boldsymbol{x}^*)$.

## 1.5　凸集、凸函数和凸规划

### 1.5.1　凸集

#### 1. 凸集、凸组合

**定义 1.8.（凸集）**

设 $S\subseteq\mathbb{R}^n$，若 $\forall\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}\in S$，$\lambda\in[0,1]$，必有

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

则称 $S$ 为**凸集**.

规定单点集 $\{\boldsymbol{x}\}$ 和空集 $\varnothing$ 都是空集. 需要注意的是，方程 $\boldsymbol{x}=\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(2)}+\lambda(\boldsymbol{x}^{(1)}-\boldsymbol{x}^{(2)})$，$\lambda\in[0,1]$ 表示连接 $\boldsymbol{x}^{(1)}$ 和 $\boldsymbol{x}^{(2)}$ 的线段.

**例 1.3**　设 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$，$\boldsymbol{b}\in\mathbb{R}^n$，证明以下命题成立：

（1）集合 $S=\{\boldsymbol{x}\mid\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}\}$ 是凸集；

（2）集合 $S=\{\boldsymbol{x}\mid\boldsymbol{A}\boldsymbol{x}\leq\boldsymbol{b}\}$ 是凸集.

**【证明】**（1）任取 $\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}\in S$，$\lambda\in[0,1]$，则都有

$$\begin{aligned}&\boldsymbol{A}(\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)})\\=&\lambda\boldsymbol{A}\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{A}\boldsymbol{x}^{(2)}\\=&\lambda\boldsymbol{b}+(1-\lambda)\boldsymbol{b}\\=&\boldsymbol{b}\end{aligned}$$

从而 $\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)}\in S$，故 $S$ 是凸集.

（2）任取 $\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}\in S$，$\lambda\in[0,1]$，则都有

$$\begin{aligned}&\boldsymbol{A}(\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)})\\=&\lambda\boldsymbol{A}\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{A}\boldsymbol{x}^{(2)}\\\leq&\lambda\boldsymbol{b}+(1-\lambda)\boldsymbol{b}\\=&\boldsymbol{b}\end{aligned}$$

从而 $\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)}\in S$，故 $S$ 是凸集.　【证毕】

**定理 1.6.（凸集的性质）**

（1）凸集的交是凸集；

（2）凸集的内点集 ① 是凸集；

（3）凸集的闭包 ② 是凸集；

---

①设 $S\subseteq\mathbb{R}^n$，$\boldsymbol{x}\in S$，如果存在一个开球 $B(\boldsymbol{x},\varepsilon)=\{\boldsymbol{y}\in\mathbb{R}^n\mid\|\boldsymbol{y}-\boldsymbol{x}\|<\varepsilon\}$，使得 $B(\boldsymbol{x},\varepsilon)\subseteq S$，则称 $\boldsymbol{x}$ 是 $S$ 的**内点**，$S$ 的内点的全体称为 $S$ 的**内点集**，记作 $\mathrm{int}(S)$. 所以 $\boldsymbol{x}\in\mathrm{int}(S)\Leftrightarrow\exists\varepsilon>0,(\|\boldsymbol{y}-\boldsymbol{x}\|<\varepsilon\Rightarrow\boldsymbol{y}\in S)$.

②设 $S\subseteq\mathbb{R}^n$，$\boldsymbol{x}\in\mathbb{R}^n$，如果 $\forall\varepsilon>0$ 都有 $B(\boldsymbol{x},\varepsilon)\cap S\neq\varnothing$，则称 $\boldsymbol{x}$ 是 $S$ 的**聚点**，$S$ 的聚点的全体称为 $S$ 的**闭包**，记作 $\overline{S}$.

（4）凸集边界上任意点存在支撑超平面 ③；

（5）两个互相不交的凸集之间存在分离超平面 ④.

**【证明】**（1）设 $S_1, S_2$ 都是凸集，即证明 $S_1 \cap S_2$ 是凸集，也即证明 $\forall \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in (S_1 \cap S_2)$，$\forall \lambda \in [0,1]$，都有 $\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)} \in (S_1 \cap S_2)$. 由于 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in (S_1 \cap S_2)$，则有 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in S_1$ 且 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in S_2$，故 $\forall \lambda \in [0,1]$，都有 $\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)} \in S_1$ 且 $\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)} \in S_2$，所以 $\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)} \in (S_1 \cap S_2)$，原命题得证.

（2）设 $S$ 是一凸集，$\mathrm{int}(S)$ 是其内点集，任取 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in \mathrm{int}(S)$，$\lambda \in [0,1]$，则 $\exists r > 0$，使得

$$\|\boldsymbol{y}^{(1)} - \boldsymbol{x}^{(1)}\| \leq r \Rightarrow \boldsymbol{y}^{(1)} \in S,$$

$$\|\boldsymbol{y}^{(2)} - \boldsymbol{x}^{(2)}\| \leq r \Rightarrow \boldsymbol{y}^{(2)} \in S$$

可以证明内点定义中使用开球和使用闭球是等价的，这里为了讨论方便，我们使用有闭球的定义. 要证明 $\mathrm{int}(S)$ 是凸集，即证明 $\boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)} \in \mathrm{int}(S)$，也即证明 $\exists \varepsilon > 0$，使得

$$\|\boldsymbol{z} - (\boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)})\| \leq \varepsilon \Rightarrow \boldsymbol{z} \in S$$

我们由三角不等式得到启发，于是上述逻辑式的前件的左式可以作如下放缩：

$$\begin{aligned}
&\|\boldsymbol{z} - (\boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)})\| \\
=&\|\boldsymbol{z} - \lambda\boldsymbol{y}^{(1)} - (1-\lambda)\boldsymbol{y}^{(2)} + \lambda(\boldsymbol{y}^{(1)} - \boldsymbol{x}^{(1)}) + (1-\lambda)(\boldsymbol{y}^{(2)} - \boldsymbol{x}^{(2)})\| \\
\leq&\|\boldsymbol{z} - \lambda\boldsymbol{y}^{(1)} - (1-\lambda)\boldsymbol{y}^{(2)}\| + \|\lambda(\boldsymbol{y}^{(1)} - \boldsymbol{x}^{(1)})\| + \|(1-\lambda)(\boldsymbol{y}^{(2)} - \boldsymbol{x}^{(2)})\| \\
\leq&\|\boldsymbol{z} - \lambda\boldsymbol{y}^{(1)} - (1-\lambda)\boldsymbol{y}^{(2)}\| + \lambda r + (1-\lambda) r \\
=&\|\boldsymbol{z} - \lambda\boldsymbol{y}^{(1)} - (1-\lambda)\boldsymbol{y}^{(2)}\| + r
\end{aligned}$$

其中 $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}$ 满足 $\|\boldsymbol{y}^{(i)} - \boldsymbol{x}^{(i)}\| \leq r$，$i = 1, 2$，所以 $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)} \in S$，故 $\forall \lambda \in [0,1]$，$\boldsymbol{y}^{(1)} + (1-\lambda)\boldsymbol{y}^{(2)} \in S$. 我们要保证 $\boldsymbol{z} \in S$，那么只需要控制 $\varepsilon = r$，于是就有

$$0 \leq \|\boldsymbol{z} - \lambda\boldsymbol{y}^{(1)} - (1-\lambda)\boldsymbol{y}^{(2)}\| \leq r - r = 0$$

从而 $\boldsymbol{z} = \lambda\boldsymbol{y}^{(1)} + (1-\lambda)\boldsymbol{y}^{(2)} \in S$，也即待证命题：$\exists \varepsilon = r > 0$，使得

$$\|\boldsymbol{z} - (\boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)})\| \leq \varepsilon \Rightarrow \boldsymbol{z} \in S$$

的前件可推出后件，从而 $\boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)} \in \mathrm{int}(S)$，待证命题得证.

（3）我们先证明一个引理：$\boldsymbol{x} \in \overline{S}$ 当且仅当存在一个点列 $\{\boldsymbol{x}^{(n)}\} \subset S$ 满足 $\lim\limits_{n\to\infty} \boldsymbol{x}^{(n)} = \boldsymbol{x}$.（实际上这正是聚点的另一等价定义）

---

③设 $\boldsymbol{\alpha} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$，$\beta \in \mathbb{R}$，称集合 $H = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{\alpha}^\top \boldsymbol{x} = \beta\}$ 是**超平面**，简记作 $H : \boldsymbol{\alpha}^\top \boldsymbol{x} = \beta$. 设 $S \subseteq \mathbb{R}^n$，$\boldsymbol{x}_0 \in S$，若超平面 $H : \boldsymbol{\alpha}^\top \boldsymbol{x} = \beta$ 满足：

（i）$\exists \boldsymbol{x}^{(0)} \in S$，使得 $\boldsymbol{x}^{(0)} \in H$；

（ii）$\forall \boldsymbol{x} \in S$，$\boldsymbol{\alpha}^\top \boldsymbol{x} \leq \beta$

则称 $H$ 是 $S$ 的**支撑超平面**.

④设 $S_1, S_2 \subseteq \mathbb{R}^n$，且 $S_1 \cap S_2 = \varnothing$. 若存在超平面 $H : \boldsymbol{\alpha}^\top \boldsymbol{x} = \beta$，使得 $\forall \boldsymbol{x}^{(1)} \in S_1$，$\boldsymbol{\alpha}^\top \boldsymbol{x}^{(1)} \geq \beta$，且 $\forall \boldsymbol{x}^{(2)} \in S_1$，$\boldsymbol{\alpha}^\top \boldsymbol{x}^{(2)} \leq \beta$，则称 $H$ 是 $S_1, S_2$ 的**分离超平面**.

先证 $\Rightarrow$. 对每一个整数 $n$，取 $\varepsilon_n = \dfrac{1}{n}$，则能得到一个开球列 $\{B(\boldsymbol{x}, \varepsilon_n)\}$. 应用选择公理，从这个开球列中依次取一个点，排成一个点列 $\{\boldsymbol{x}^{(n)}\}$，则必然有

$$\|\boldsymbol{x}^{(n)} - \boldsymbol{x}\| = \frac{1}{n} \to 0$$

再证 $\Leftarrow$. 由于 $\lim\limits_{n\to\infty} \boldsymbol{x}^{(n)} = \boldsymbol{x}$，则 $\forall\varepsilon > 0$，都 $\exists N > 0$，使得 $\|\boldsymbol{x}^{(n)} - \boldsymbol{x}\| < \varepsilon$ 对于 $\forall n > N$ 成立，从而 $\forall\varepsilon > 0$，都一定存在一个 $n$，使得 $\boldsymbol{x}^{(n)} \in B(\boldsymbol{x}, \varepsilon)$ 并且 $\boldsymbol{x}^{(n)} \in S$，从而 $\boldsymbol{x}^{(n)} \in B(\boldsymbol{x}, \varepsilon) \cap S$，故 $B(\boldsymbol{x}, \varepsilon) \cap S \neq \varnothing$，所以 $\boldsymbol{x} \in \overline{S}$. 从而引理得证.

任取 $\boldsymbol{x}, \boldsymbol{y} \in \overline{S}$，$\lambda \in [0, 1]$，由引理，$\exists\{\boldsymbol{x}^{(n)}\}, \{\boldsymbol{y}^{(n)}\} \subset S$，使得 $\boldsymbol{x}^{(n)} \to \boldsymbol{x}$，$\boldsymbol{y}^{(n)} \to \boldsymbol{y}$. 构造点列 $\{\lambda\boldsymbol{x}^{(n)} + (1-\lambda)\boldsymbol{y}^{(n)}\}$，则由 $\exists\{\boldsymbol{x}^{(n)}\}, \{\boldsymbol{y}^{(n)}\} \subset S$ 以及 $S$ 是凸集，有 $\{\lambda\boldsymbol{x}^{(n)} + (1-\lambda)\boldsymbol{y}^{(n)}\} \in S$. 由 $\boldsymbol{x}^{(n)} \to \boldsymbol{x}$，$\boldsymbol{y}^{(n)} \to \boldsymbol{y}$，知 $\{\lambda\boldsymbol{x}^{(n)} + (1-\lambda)\boldsymbol{y}^{(n)}\} \to \lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}$，所以根据引理，$\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in \overline{S}$. 原命题得证.

（4）暂时懒得写.

（5）暂时懒得写. 【证毕】

**定义 1.9.（凸组合）**

设 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(m)} \in \mathbb{R}^n$. 若 $\forall j \in \{1, 2, \cdots, m\}$，$\lambda_j \geq 0$ 且 $\sum\limits_{j=1}^{m} \lambda_j = 1$，则称 $\sum\limits_{j=1}^{m} \lambda_j \boldsymbol{x}^{(j)}$ 为 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(m)}$ 的**凸组合**.

**定理 1.7.（凸集和凸组合的关系）**

$S$ 是凸集 $\Leftrightarrow$ $S$ 中任意有限点的凸组合属于 $S$.

**【证明】**

【证毕】

**定义 1.10.（凸包）**

设非空集合 $S \subseteq \mathbb{R}^n$，由 $S$ 中所有有限点的凸组合所构成的集合，被称为 $S$ 的**凸包**，记作 $\mathrm{cov}(S)$.

**定理 1.8.（凸集和凸包的关系）**

如果 $S$ 是凸集，那么 $\mathrm{cov}(S) = S$.

**【证明】**

【证毕】

**定义 1.11.（多胞形）**

设 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(m)} \in \mathbb{R}^n$. 由 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(m)}$ 的所有凸组合构成的集合称为多胞形，记作 $H(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(m)})$.

**定义 1.12.（单纯形）**

设 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(m)} \in \mathbb{R}^n$. 若多胞形 $H(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(m)})$ 满足 $\boldsymbol{x}^{(2)} - \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(3)} - \boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(m)} - \boldsymbol{x}^{(1)}$ 线性无关，则称该多胞形是单纯形.

**2. 凸锥、半正组合**

**定义 1.13.** （锥、凸锥）

设 $S \subseteq \mathbb{R}^n$，若 $\forall \boldsymbol{x} \in S$，$\lambda > 0$，必有 $\lambda \boldsymbol{x} \in S$，则称 $S$ 为以 $\mathbf{0}$ 为顶点的**锥**. 若 $S$ 还是凸集，则称 $S$ 为**凸锥**.

规定 $\{\mathbf{0}\}, \mathbb{R}^n$ 都是凸锥.

**定义 1.14.** （半正组合）

设 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(m)} \in \mathbb{R}^n$. 若 $\sum\limits_{j=1}^{m} \lambda_j > 0$，则称 $\sum\limits_{j=1}^{m} \lambda_j \boldsymbol{x}^{(j)}$ 为 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(m)}$ 的**半正组合**.

**定理 1.9.** （凸锥和半正组合的关系）

$S$ 是凸锥 $\Leftrightarrow$ $S$ 中任意有限点的半正组合属于 $S$.

**【证明】**

【证毕】

## 1.5.2 凸函数

**1. 凸函数**

**2. 水平集**

## 1.5.3 凸规划