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，ξ_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 $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$，$\forall \boldsymbol{x} \in S$，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 $\boldsymbol{d}\neq 0$——表示从 0 指向 $\boldsymbol{d}$ 的方向. 此外，设 $\boldsymbol{d}^{(1)},\boldsymbol{d}^{(2)}\in\mathbb{R}^n$，若 $\exists\lambda>0$，使得 $\boldsymbol{d}^{(1)}=\lambda\boldsymbol{d}^{(2)}$，则称 $\boldsymbol{d}^{(1)},\boldsymbol{d}^{(2)}$ 是**同方向**的.
- $\boldsymbol{x}+\lambda\boldsymbol{d}$——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

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

【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

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

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

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

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 $x_i \le y_i$，则称 $\boldsymbol{x} \le \boldsymbol{y}$. 类似可规定 $\boldsymbol{x} \ge \boldsymbol{y}$，$\boldsymbol{x} < \boldsymbol{y}$，$\boldsymbol{x} > \boldsymbol{y}$.

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

定理 1.2

设 $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$，若 $\forall \boldsymbol{y} \ge 0$，都有 $\boldsymbol{x}^\top \boldsymbol{y} \le \alpha$，则有 $\boldsymbol{x} \le 0$ 且 $\alpha \ge 0$.

【证明】取 $\boldsymbol{y} = 0$, 则 $\boldsymbol{x}^\top \boldsymbol{y} = 0$, 从而 $\boldsymbol{x}^\top \boldsymbol{y}$ 的上界 $\alpha \ge 0$. 假设 $\boldsymbol{x} \le 0$ 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 α 矛盾，故必有 $\boldsymbol{x} \le 0$. **【证毕】**

该命题还有以下等价表述:

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

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 $0 \in L$. 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 $0 \in L^\top$. 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 1.3 （子空间投影定理）

设 L 为 $\mathbb{R}^n$ 的子空间，则 $\forall z \in \mathbb{R}^n$，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

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

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 $z-A(A^\top A)^{-1}A^\top z\in L^\perp$. 由于

$$(z-A(A^\top A)^{-1}A^\top z)^\top A\beta$$

$$=z^\top A\beta-z^\top A(A^\top A)^{-1}A^\top A\beta$$

$$=z^\top A\beta-z^\top AE\beta$$

$$=0$$

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

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

$$x^\top q=x^\top(x+y-p)$$

$$=x^\top x-x^\top q$$

$$\neq 0$$

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

（2）考查函数 $f(t)=\|z-x-tv\|^2$，$\forall v\in L$ 且 $v\neq 0$，则要证明目标问题的唯一解是 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

$$f'(t)=-2y^\top v+2tv^\top v$$

$$=2tv^\top v$$

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

定理 1.4

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

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

$$x^\top y=(x^{(1)}+x^{(2)})^\top y$$

$$=x^{(1)\top}y+x^{(2)\top}y$$

$$=x^{(1)\top}y$$

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

该命题还有以下等价表述:

- 若 $\forall y\in L$，都有 $x^\top y\geq\alpha$，则有 $x\in L^\perp$ 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 1.5　（一些常见函数的梯度）

- 线性函数：$f(\boldsymbol{x})=\boldsymbol{c}^\top\boldsymbol{x}+\boldsymbol{d}$，$\nabla f(\boldsymbol{x})=\boldsymbol{c}$，$\nabla^2 f(x)=0$；
- 二次函数：$f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^\top\boldsymbol{Q}\boldsymbol{x}+\boldsymbol{c}^\top\boldsymbol{x}+\boldsymbol{d}$，$\nabla f(\boldsymbol{x})=\boldsymbol{Q}\boldsymbol{x}+\boldsymbol{c}$，$\nabla^2 f(x)=\boldsymbol{Q}$，其中 $\boldsymbol{Q}$ 为对称阵.
- 向量值线性函数：$\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{d}\in\mathbb{R}^m$，$\frac{\mathrm{d}\boldsymbol{f}}{\mathrm{d}\boldsymbol{x}}=\boldsymbol{A}^\top$.

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 1.6　（Taylor 展开式）

设 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $\boldsymbol{x}^*$ 的某邻域内分别一阶、二阶可微，则分别有：

- 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：$f(\boldsymbol{x})=f(\boldsymbol{x}^*)+\nabla f^\top(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*)+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^\top\nabla^2 f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*)+o(\|\boldsymbol{x}-\boldsymbol{x}^*\|^2)$.

定理 1.7　（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
$$f(\boldsymbol{x})=f(\boldsymbol{x}^*)+\nabla f^\top(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*)+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^\top\nabla^2 f(\boldsymbol{x}_\mu)(\boldsymbol{x}-\boldsymbol{x}^*)$$

定义 1.8　（方向导数）

设 $S\subseteq\mathbb{R}^n$ 为非空凸集，函数 $f:S\to\mathbb{R}$，$\boldsymbol{x}^*\in S$，方向 $\boldsymbol{d}\in\mathbb{R}^n$，若 $\exists\lambda>0$，使得 $\boldsymbol{x}^*+\lambda\boldsymbol{d}\in S$（通常令 λ 充分小，此时称 $\boldsymbol{d}$ 为**可行方向**），且此时极限
$$\lim_{\lambda\to0^+}\frac{f(\boldsymbol{x}^*+\lambda\boldsymbol{d})-f(\boldsymbol{x}^*)}{\lambda}$$
存在，则称 $f(\boldsymbol{x})$ 在点 $\boldsymbol{x}^*$ 沿方向 $\boldsymbol{d}$ 的**方向导数**存在，记作 $f'(\boldsymbol{x}^*;\boldsymbol{d})$.

定理 1.8　（方向导数和梯度的关系）

若 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}^*$ 处可微，则
$$f'(\boldsymbol{x}^*;\boldsymbol{d})=\nabla f^\top(\boldsymbol{x}^*)\boldsymbol{d}$$

【证明】 应用 $f(\boldsymbol{x})$ 的一阶展开式，则

$$\begin{aligned}
\lim_{\lambda\to 0^+}\frac{f(\boldsymbol{x}^*+\lambda\boldsymbol{d})-f(\boldsymbol{x}^*)}{\lambda} &= \lim_{\lambda\to 0^+}\frac{f(\boldsymbol{x}^*)+\nabla f^\top(\boldsymbol{x}^*)(\lambda\boldsymbol{d})+o(\lambda\|\boldsymbol{d}\|)}{\lambda}\\
&= \nabla f^\top(\boldsymbol{x}^*)\boldsymbol{d}+\lim_{\lambda\to 0^+}\frac{o(\lambda\|\boldsymbol{d}\|)}{\lambda}\\
&= \nabla f^\top(\boldsymbol{x}^*)\boldsymbol{d}+\lim_{\lambda\to 0^+}\frac{o(\lambda\|\boldsymbol{d}\|)}{\lambda\|\boldsymbol{d}\|}\cdot\|\boldsymbol{d}\|\\
&= \nabla f^\top(\boldsymbol{x}^*)\boldsymbol{d}
\end{aligned}$$

【证毕】

1.5　凸集、凸函数和凸规划

1.5.1　凸集

1. 凸集、凸组合

定义 1.9　（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 $\emptyset$ 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 1.1　设 $A \in \mathbb{R}^{m\times n}$，$\boldsymbol{b} \in \mathbb{R}^n$，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 $S = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}\}$ 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

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

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

在介绍凸集的性质之前，我们先对一些将要用到的名词进行必要的说明.

定义 1.10　（内点、内点集、开集）

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 $\mathrm{int}(S)$.

进一步地，如果 $\mathrm{int}(S) = S$，则称 S 是**开集**

所以 $\boldsymbol{x} \in \mathrm{int}(S) \Leftrightarrow \exists \varepsilon > 0, (\|\boldsymbol{y} - \boldsymbol{x}\| < \varepsilon \Rightarrow \boldsymbol{y} \in S)$.

定义 1.11 （聚点、闭包）

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

定义 1.12 （超平面）

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

定义 1.13 （支撑超平面）

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**.

定义 1.14 （分离超平面）

设 $S_1, S_2 \subseteq \mathbb{R}^n$，且 $S_1 \cap S_2 = \emptyset$. 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 $\forall \boldsymbol{x}^{(2)} \in S_2$，$\boldsymbol{\alpha}^\top \boldsymbol{x}^{(2)} \leq \beta$，则称 H 是 S_1, S_2 的**分离超平面**.

定理 1.9 （凸集的性质）

凸集有以下性质：

（1）凸集的交是凸集；

（2）凸集的内点集是凸集；

（3）凸集的闭包是凸集；

（4）凸集边界上任意点存在支撑超平面；

（5）两个互相不交的凸集之间存在分离超平面.

【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：

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

其中 $\boldsymbol{y}^{(1)},\boldsymbol{y}^{(2)}$ 满足 $\|\boldsymbol{y}^{(i)}-\boldsymbol{x}^{(i)}\|\le r, i=1,2$, 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 $z\in S$，那么只需要控制 $\varepsilon=r$，于是就有

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

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

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

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）

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 $B(\boldsymbol{x},\varepsilon)\cap S\neq\emptyset$，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 1.15 （凸组合）

设 $\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)},\cdots,\boldsymbol{x}^{(m)}\in\mathbb{R}^n$. 若 $\forall j\in\{1,2,\cdots,m\}$，$\lambda_j\ge 0$ 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 1.10 （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】先证 $\Rightarrow$. 使用数学归纳法. 即证明 S 中任意 m 个点的凸组合属于 S. $m=1,2$ 时，由于 S 是凸集，所以显然成立. 假设 $m=k$ 时，结论成立，往证 $m=k+1$ 时结论也成立，即证明当 $\sum\limits_{j=1}^{k+1}\lambda_j=1$ 时 $\sum\limits_{j=1}^{k+1}\lambda_j\boldsymbol{x}_j\in S$. 由于 $m=k$ 时结论成立，则有：

$$\frac{\sum\limits_{j=1}^{k}\lambda_j\boldsymbol{x}_j}{\sum\limits_{j=1}^{k}\lambda_j}=\sum_{j=1}^{k}\frac{\lambda_j}{\sum\limits_{i=1}^{k}\lambda_i}\boldsymbol{x}_j\in S$$

从而

$$\sum_{j=1}^{k+1}\lambda_j\boldsymbol{x}^{(j)}=\sum_{j=1}^{k}\lambda_j\boldsymbol{x}^{(j)}+\lambda_{k+1}\boldsymbol{x}^{(k+1)}=\sum_{j=1}^{k}\lambda_j\cdot\frac{\sum\limits_{j=1}^{k}\lambda_j\boldsymbol{x}^{(j)}}{\sum\limits_{j=1}^{k}\lambda_j}+\lambda_{k+1}\boldsymbol{x}^{(k+1)}\in S$$

结论得证.

再证 $\Leftarrow$. 取 $m=2$，则若 $\lambda_1,\lambda_2>0$ 且 $\lambda_1+\lambda_2=1$，都有 $\lambda_1\boldsymbol{x}^{(1)}+\lambda_2\boldsymbol{x}^{(2)}\in S$，这正是凸集的定义，从而结论成立. 【证毕】

定义 1.16 （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 1.11 （凸集和凸包的关系）

如果 S 是凸集，那么 $\mathrm{cov}(S)=S$.

【证明】由定理 1.8 知，$\mathrm{cov}(S)\subseteq S$，下证 $S\subseteq\mathrm{cov}(S)$. 一个点的凸组合就是其本身，所以只要 $\boldsymbol{x}\in S$，那么这个点本身的凸组合 $\boldsymbol{x}\in\mathrm{cov}(S)$，所以 $S\subseteq\mathrm{cov}(S)$. 【证毕】

定义 1.17 （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 1.18 （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 1.19 （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 0 为顶点的**锥**. 若 S 还是凸集，则称 S 为**凸锥**.

规定 $\{0\}, \mathbb{R}^n$ 都是凸锥. 锥不一定包含 0.

定义 1.20 （半正组合）

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 $\sum\limits_{j=1}^{m} \lambda_j > 0$，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 1.12 （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】 先证 $\Rightarrow$. 即证若 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(m)} \in S$. 且 $\forall j \in \{1, 2, \cdots, m\}$，$\lambda_j \geq 0$ 且 $\sum\limits_{j=1}^{m} \lambda_j > 0$，则称 $\sum\limits_{j=1}^{m} \lambda_j \boldsymbol{x}^{(j)} \in S$. 注意到：

$$\sum_{j=1}^{m} \lambda_j \boldsymbol{x}^{(j)} = \sum_{j=1}^{m} \lambda_j \cdot \sum_{j=1}^{m} \frac{\lambda_j}{\sum\limits_{i=1}^{m} \lambda_i} \boldsymbol{x}^{(j)}$$

其中：$\sum\limits_{j=1}^{m} \frac{\lambda_j}{\sum\limits_{i=1}^{m} \lambda_i} \boldsymbol{x}^{(j)}$ 正是 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(m)}$ 的凸组合，由定理 1.8 知，$\sum\limits_{j=1}^{m} \frac{\lambda_j}{\sum\limits_{i=1}^{m} \lambda_i} \boldsymbol{x}^{(j)} \in S$. 此外，由 $\sum\limits_{j=1}^{m} \lambda_j > 0$ 以及锥的定义，有

$$\sum_{j=1}^{m} \lambda_j \cdot \sum_{j=1}^{m} \frac{\lambda_j}{\sum\limits_{i=1}^{m} \lambda_i} \boldsymbol{x}^{(j)} \in S$$

即原命题得证.

再证 $\Leftarrow$. 显然任意有限点的凸组合一定是其半正组合，从而由定理 1.8 的 $\Leftarrow$，S 是凸集. 我们仿照定理 1.8 的证明方法，一个点 $\boldsymbol{x}$ 的半正组合正是 $\lambda \boldsymbol{x}$，$\lambda > 0$，而 “若 $\forall \boldsymbol{x} \in S$，$\lambda > 0$，则 $\lambda \boldsymbol{x} \in S$” 正是锥的定义. 因此原命题得证. 【证毕】

1.5.2　凸函数

1. 凸函数

定义 1.21 （凸函数）

设集合 $S \subseteq \mathbb{R}^n$ 为凸集，函数 $f: S \to R$. 11122 若 $\forall \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in S$，$\lambda \in (0,1)$，均有

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

则称 $f(\boldsymbol{x})$ 为凸集 S 上的**凸函数**.

若进一步有上面不等式以严格不等式成立, 则称 $f(\boldsymbol{x})$ 为凸集 S 上的**严格凸函数**. 当 $-f(\boldsymbol{x})$ 为凸函数（严格凸函数时），则称 $f(\boldsymbol{x})$ 为**凹函数（严格凹函数）**.

定理 1.13 （Jesen 不等式）

$f(\boldsymbol{x})$ 为凸集 S 上的凸函数的充要条件是 S 上任意有限点的凸组合的函数值不大于各点函数值的凸组合, 即 $\forall \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(m)} \in S$, 对于任意的序列 $\{\lambda_j\}_{j=1}^m$ 满足 $\forall j \in \{1, 2, \cdots, m\}$, $\lambda_j \ge 0$ 且 $\sum_{j=1}^m \lambda_j = 1$，都有

$$f\left(\sum_{j=1}^m \lambda_j \boldsymbol{x}^{(j)}\right) \le \sum_{j=1}^m \lambda_j f(\boldsymbol{x}^{(j)})$$

【证明】我们仿照定理 1.8 进行证明. 先证 $\Rightarrow$. 使用数学归纳法. $m=1$ 时，待证命题显然成立. $m=2$ 时，待证命题与凸函数定义等价，故其也成立. 设 $m=k$ 时，待证命题成立，往证 $m=k+1$ 时命题也成立. 这里为了讨论方便，令 $\sum_{j=1}^k \lambda_j = \theta$，则

$$\begin{aligned}
f\left(\sum_{j=1}^{k+1} \lambda_j \boldsymbol{x}^{(j)}\right) &= f\left(\sum_{j=1}^{k} \lambda_j \boldsymbol{x}^{(j)} + \lambda_{k+1}\boldsymbol{x}^{(k+1)}\right) \\
&= f\left(\theta \sum_{j=1}^{k} \frac{\lambda_j}{\theta} x^{(j)} + \lambda_{k+1}\boldsymbol{x}^{(k+1)}\right) \\
&\le \theta f\left(\frac{\lambda_j}{\theta} x^{(j)}\right) + \lambda_{k+1} f(\boldsymbol{x}^{(k+1)}) \\
&\le \theta \left(\sum_{j=1}^{k} \frac{\lambda_j}{\theta} f(\boldsymbol{x}^{(j)})\right) + \lambda_{k+1} f(\boldsymbol{x}^{(k+1)}) \\
&= \sum_{j=1}^{k+1} \lambda_j f(\boldsymbol{x}^{(j)})
\end{aligned}$$

命题得证.

再证 $\Leftarrow$. 取 $m=2$ 时即可，因为这正是凸函数的定义. 命题得证. 【证毕】

例 1.2 设 f_1, f_2 是凸函数，判断：

（1）设 $\lambda_1, \lambda_2 > 0$，$\lambda_1 f_1 + \lambda_2 f_2$ 是否为凸函数？$\lambda_1 f_1 - \lambda_2 f_2$ 是否为凸函数？

（2）$f(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}$ 是否为凸函数？$g(\boldsymbol{x}) = \min\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}$ 是否为凸函数？

【解】（1）$\lambda_1 f_1 + \lambda_2 f_2$ 是凸函数. 设 f_1, f_2 定义域的交集为 S，则 $\forall \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in S$，$\forall \lambda > 0$，

都有

$$\begin{aligned}
&\lambda_1 f_1(\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)}) + \lambda_2 f_2(\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)}) \\
\le& \lambda_1(\lambda f_1(\boldsymbol{x}^{(1)}) + (1-\lambda) f_1(\boldsymbol{x}^{(2)})) + \lambda_2(\lambda f_2(\boldsymbol{x}^{(1)}) + (1-\lambda) f_2(\boldsymbol{x}^{(2)})) \\
=& \lambda(\lambda_1 f_1(\boldsymbol{x}^{(1)}) + \lambda_2 f_2(\boldsymbol{x}^{(1)})) + (1-\lambda)(\lambda_1 f_1(\boldsymbol{x}^{(2)}) + \lambda_2 f_2(\boldsymbol{x}^{(2)}))
\end{aligned}$$

$\lambda_1 f_1 - \lambda_2 f_2$ 不一定是凸函数，构造反例如下：令 $\lambda_1 = 1$，$\lambda_2 = 2$，$f_1(x) = f_2(x) = x^2$，从而 $\lambda_1 f_1(x) - \lambda_2 f_2(x) = -x^2$ 在 $\mathbb{R}$ 上是凹函数. 若要构造正例，只需要调整上述反例中的 λ_1, λ_2，使得 $\lambda_1 > \lambda_2$ 即可.

（2）$f(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}$ 是凸函数. 设 f_1, f_2 定义域的交集为 S，则 $\forall \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in S$，$\forall \lambda > 0$，都有

$$f_1(\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)}) \le \lambda f_1(\boldsymbol{x}^{(1)}) + (1-\lambda) f_1(\boldsymbol{x}^{(2)}) \le \lambda f(\boldsymbol{x}^{(1)}) + (1-\lambda) f(\boldsymbol{x}^{(2)})$$

同理，有

$$f_2(\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)}) \le \lambda f(\boldsymbol{x}^{(1)}) + (1-\lambda) f(\boldsymbol{x}^{(2)})$$

从而

$$\begin{aligned}
&f(\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)}) \\
=& \max\{f_1(\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)}), f_2(\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)})\} \\
\le& \lambda f(\boldsymbol{x}^{(1)}) + (1-\lambda) f(\boldsymbol{x}^{(2)})
\end{aligned}$$

$g(\boldsymbol{x}) = \min\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}$ 不一定是凸函数. 考虑如下反例：令 $f_1(x) = x^2$，$f_2(x) = (x-1)^2$，则 $g(0) = \min\{0, 1\} = 0$，$g(1) = \min\{1, 0\} = 0$，$g\left(\frac{1}{2}\right) = \min\left\{\frac{1}{4}, \frac{1}{4}\right\} = \frac{1}{4}$，然而：

$$g\left(\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1\right) = g\left(\frac{1}{2}\right) = \frac{1}{4} > \frac{1}{2} g(0) + \frac{1}{2} g(1)$$

因此 $g(x)$ 不是凸函数. 【解毕】

定理 1.14 （凸函数的性质）

设 $S \subseteq \mathbb{R}^n$ 为非空凸集，函数 $f : S \to \mathbb{R}$，则有：

（1）若 f 是凸函数，则 f 在 $\mathrm{int}(S)$ 上连续；

（2）若 f 是凸函数，则其对任意方向的方向导数（若方向可行）存在；

（3）设 S 是开集，f 在 S 上可微，则 f 是凸函数的充要条件是 $\forall \boldsymbol{x}^* \in S$，都有

$$f(\boldsymbol{x}) \ge f(\boldsymbol{x}^*) + \nabla f^{\top}(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*)$$

$\forall \boldsymbol{x} \in S$ 都成立. 进一步地，如果 f 是严格凸函数，则其充要条件是上述不等式的严格不等式成立；

（4）设 S 是开集，f 在 S 上二次可微，则：

（i）f 是凸函数的充要条件是 $\forall \boldsymbol{x} \in S$，$\nabla^2 f(\boldsymbol{x})$ 是半正定的；

（ii）若 $\forall \boldsymbol{x} \in S$，$\nabla^2 f(\boldsymbol{x})$ 是正定的，则 f 是严格凸函数.

【证明】（1）以下基于估计 $f(\boldsymbol{x}) - f(\boldsymbol{x}^{(0)})$ 的上下界的证明来自知乎用户 @cvgmt.

我们先证明：凸函数 $f(\boldsymbol{x})$ 在其定义域的内点集上处处有上界. 设凸函数 $f(\boldsymbol{x})$ 及其定义域为 S，则 $\forall \boldsymbol{x}^{(0)} \in \text{int}(S)$，都一定存在一个充分小的 $r > 0$，使得以 $\boldsymbol{x}^{(0)}$ 为中心的正方体包含在 $\text{int}(S)$ 中. 由凸函数的定义

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

容易得到凸函数的最大值一定在线段端点上取得. 反复应用上述定义，最终能够证明凸函数的最大值一定在正方形的端点处取得. [1]

下面考虑以 $\boldsymbol{x}^{(0)}$ 为圆心，r 为半径的闭球 $\overline{B}(\boldsymbol{x}^{(0)}, r)$（包含于上述正方形），则由上面的证明，知道 $f(\boldsymbol{x})$ 在该闭球上一定有上界，从而一定有上确界，记作 M_r. 我们先来估计 $f(\boldsymbol{x}) - f(\boldsymbol{x}^{(0)})$ 的上界. 在 $\overline{B}(\boldsymbol{x}^{(0)}, r)$ 内任取一点 $\boldsymbol{x}$，连接 $\boldsymbol{x}^{(0)}$ 和 $\boldsymbol{x}$ 并延长到球面上，交于一点 $\boldsymbol{y}$，从而 $\boldsymbol{x}$ 在以 $\boldsymbol{x}^{(0)}$ 和 $\boldsymbol{y}$ 为端点的线段上，则由凸函数的定义，有

$$f(\boldsymbol{x}) \leq \frac{\|\boldsymbol{x}-\boldsymbol{y}\|}{r} f(\boldsymbol{x}^{(0)}) + \frac{\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\|}{r} f(\boldsymbol{y})$$

其中 $\dfrac{\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\|}{r} + \dfrac{\|\boldsymbol{x}-\boldsymbol{y}\|}{r} = 1$，从而有

$$\begin{aligned} f(\boldsymbol{x}) - f(\boldsymbol{x}^{(0)}) &\leq -\frac{\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\|}{r} f(\boldsymbol{x}^{(0)}) + \frac{\|\boldsymbol{x}-\boldsymbol{y}\|}{r} f(\boldsymbol{y}) \\ &= \frac{\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\|}{r}(f(\boldsymbol{y}) - f(\boldsymbol{x}^{(0)})) \\ &\leq \|\boldsymbol{x}-\boldsymbol{x}^{(0)}\| \cdot \frac{M_r - f(\boldsymbol{x}^{(0)})}{r} \end{aligned}$$

接下来估计 $f(\boldsymbol{x}) - f(\boldsymbol{x}^{(0)})$ 的下界. 类似地，在 $\overline{B}(\boldsymbol{x}^{(0)}, r)$ 内任取一点 $\boldsymbol{x}'$，连接 $\boldsymbol{x}'$ 和 $\boldsymbol{x}^{(0)}$ 并延长到球面上，交于一点 $\boldsymbol{y}'$，从而 $\boldsymbol{x}^{(0)}$ 在以 $\boldsymbol{x}'$ 和 $\boldsymbol{y}'$ 为端点的线段上，则由凸函数的定义，有

$$f(\boldsymbol{x}^{(0)}) \leq \frac{\|\boldsymbol{y}-\boldsymbol{x}^{(0)}\|}{\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\| + r} f(\boldsymbol{x}) + \frac{\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\|}{\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\| + r} f(\boldsymbol{y})$$

其中 $\dfrac{\|\boldsymbol{y}-\boldsymbol{x}^{(0)}\|}{\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\| + r} + \dfrac{\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\|}{\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\| + r} = 1$，从而有：

$$\begin{aligned} f(\boldsymbol{x}) - f(\boldsymbol{x}^{(0)}) &\geq \frac{\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\|}{r} f(\boldsymbol{x}^{(0)}) - \frac{\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\|}{r} f(\boldsymbol{y}) \\ &= \frac{\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\|}{r}(f(\boldsymbol{x}^{(0)}) - f(\boldsymbol{y})) \\ &\geq \|\boldsymbol{x}-\boldsymbol{x}^{(0)}\| \cdot \frac{f(\boldsymbol{x}^{(0)}) - M_r}{r} \end{aligned}$$

综上，有

$$|f(\boldsymbol{x}) - f(\boldsymbol{x}^{(0)})| \leq \|\boldsymbol{x}-\boldsymbol{x}^{(0)}\| \cdot \frac{|M_r - f(\boldsymbol{x}^{(0)})|}{r}$$

将 r 保持为不是无穷小的固定值，从而 $\dfrac{|M_r - f(\boldsymbol{x}^{(0)}|}{r}$ 为一有限值，并且令 $\boldsymbol{x} \to \boldsymbol{x}^{(0)}$，从而 $\|\boldsymbol{x}-\boldsymbol{x}^{(0)}\| \to 0$，于是由夹挤准则，$|f(\boldsymbol{x}) - f(\boldsymbol{x}^{(0)})| \to 0$，于是连续性得证.

（2）设凸函数 $f(\boldsymbol{x})$ 的定义域为 S，$\forall \boldsymbol{x} \in S$ 及其可行方向 $\boldsymbol{d}$，即证明函数

$$g(\lambda) = \frac{f(\boldsymbol{x} + \lambda \boldsymbol{d}) - f(\boldsymbol{x})}{\lambda}$$

---

[1] 这个条件还有另一种基于 Minkowski-Weyl 定理（见定理**??**）的证明方法. 由上述定理，正方体（或任意有界凸多面体）实际上是其顶点集的凸包. 之后应用 Jesen 不等式进行证明即可.

在 0^+ 处有定义且极限存在. 取充分小的 $\delta > 0$，由于 $\boldsymbol{d}$ 是可行方向，故其在 $(0, \delta)$ 上有定义. 下证其在 0^+ 处极限存在. 取 $0 < \lambda_1 < \lambda_2 < \delta$，由 f 的凸性，有：

$$f(\boldsymbol{x} + \lambda_1 \boldsymbol{d}) \leq \frac{\lambda_1}{\lambda_2} f(\boldsymbol{x} + \lambda_2 \boldsymbol{d}) + \left(1 - \frac{\lambda_1}{\lambda_2}\right) f(\boldsymbol{x})$$

整理得

$$g(\lambda_1) = \frac{f(\boldsymbol{x} + \lambda_1 \boldsymbol{d}) - f(\boldsymbol{x})}{\lambda_1} \leq \frac{f(\boldsymbol{x} + \lambda_2 \boldsymbol{d}) - f(\boldsymbol{x})}{\lambda_2} = g(\lambda_2)$$

由 λ_1, λ_2 及 δ 的任意性，有 $g(\lambda)$ 在 0 的任意右半去心邻域内单调递增，从而极限 $\lim\limits_{\lambda \to 0^+} g(\lambda)$ 为有限数或 $-\infty$.

（3）

（4）

【证毕】

例 1.3 若 $f(x) = \begin{cases} 2, & x = \pm 1, \\ x^2, & -1 < x < 1 \end{cases}$，容易验证 f 是凸函数，但是 f 在边界点上不连续.

2. 水平集

定义 1.22 （水平集）

设集合 $S \subseteq \mathbb{R}^n$，函数 $f: S \to \mathbb{R}$，$\alpha \in \mathbb{R}$，称 $S_\alpha = \{\boldsymbol{x} \in S \mid f(\boldsymbol{x}) \leq \alpha\}$ 为 $f(\boldsymbol{x})$ 在 S 上的 α **水平集**.

水平集的概念相当于在地形图中，海拔高度不高于某一数值的区域.

定理 1.15 （凸函数的水平集是凸集）

设集合 $S \subseteq \mathbb{R}^n$ 是凸集，函数 $f: S \to \mathbb{R}$ 是凸函数，则对 $\forall \alpha \in \mathbb{R}$，$S_\alpha$ 是凸集.

【证明】

【证毕】

上述定理的逆不真.

例 1.4 考虑分段函数 $f(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0 \end{cases}$，函数非凸，但其任意水平集是凸集.

1.5.3　凸规划

定义 1.23 （凸规划）

（1）若 $(f\ S)$ 中的 S 为凸集，f 是 S 上的凸函数，优化目标为 min 时，则称 $(f\ S)$ 为凸规划；

（2）若 (fgh) 中的 f, g_i 为凸函数，h_j 为线性函数，则称 (fgh) 为凸规划.

定义 1.24 （凸优化的最优解）

设 $(f\ S)$ 为凸优化. 若 $\boldsymbol{x}^*$ 为问题 $(f\ S)$ 的 l.opt.，则 $\boldsymbol{x}^*$ 为 g.opt. 进一步地，若 f 是严格凸函数，则 $\boldsymbol{x}^*$ 是 $(f\ S)$ 的唯一 g.opt.

【证明】

【证毕】

1.6　多面体、极点、极方向

定义 1.25（半空间）

设超平面 $H:\boldsymbol{\alpha}^\top \boldsymbol{x}=\beta\subseteq\mathbb{R}^n$，则这个超平面把 $\mathbb{R}^n$ 分成两部分 $\{\boldsymbol{x}\in\mathbb{R}^n\mid\boldsymbol{\alpha}^\top\boldsymbol{x}\geq\beta\}$ 和 $\{\boldsymbol{x}\in\mathbb{R}^n\mid\boldsymbol{\alpha}^\top\boldsymbol{x}\leq\beta\}$，这样的两部分称为**闭半空间**，若有严格不等式成立，则称其为**开半空间**.

定义 1.26（多面体）

有限个闭半空间的交称为**多面体**.

由定义看出，多面体实际上由一系列线性不等式组定义，故一般可以将其写作 $\{\boldsymbol{x}\in\mathbb{R}^n\mid \boldsymbol{A}\boldsymbol{x}\leq\boldsymbol{b}\}$，其中 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$，$\boldsymbol{b}\in\mathbb{R}^m$. 其中我们选取一种特殊的多面体 $\{\boldsymbol{x}\in\mathbb{R}^n\mid\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq 0\}$，称其为**标准型多面体**.

定义 1.27（极点）

设 $S\subseteq\mathbb{R}^n$，$\boldsymbol{x}\in S$，若不存在 $\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}\in S$，$\boldsymbol{x}^{(1)}\neq\boldsymbol{x}^{(2)}$，使得 $\boldsymbol{x}=\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)}$，则称 x 是 S 的**极点**（或**顶点**）.

定义 1.28（极方向）

设 $S\subseteq\mathbb{R}^n$，$\boldsymbol{d}\in\mathbb{R}^n$，若 $\boldsymbol{d}$ 是 S 的可行方向（见**方向导数**的定义），且不能被表示为 S 的两个不同可行方向的非负组合，则称 $\boldsymbol{d}$ 是 S 的**极方向**.

定理 1.16（极点特征）

设多面体 $S=\{\boldsymbol{x}\in\mathbb{R}^n\mid\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq 0\}$，其中 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ 行满秩，则 $\boldsymbol{x}$ 是多面体 S 的极点的充要条件是存在分解 $\boldsymbol{A}=(\boldsymbol{B},\boldsymbol{N})$，使得

（i）B 为 m 阶可逆阵；

（ii）$\boldsymbol{x}=(\boldsymbol{x}_B^\top,\boldsymbol{x}_N^\top)^\top$（分解方式与 $\boldsymbol{A}$ 相同），其中 $\boldsymbol{x}_B=\boldsymbol{B}^{-1}\boldsymbol{b}\geq 0$，$\boldsymbol{x}_N=0$.

【证明】

【证毕】

例 1.5 证明：多面体 $S=\{\boldsymbol{x}\in\mathbb{R}^n\mid\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq 0\}$ 中必存在有限多的极点.

【证明】

【证毕】

例 1.6 设多面体 $S=\{\boldsymbol{x}\in\mathbb{R}^n\mid\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq 0\}$，在以下条件下，求该多面体的极点：

（1）$A=\begin{pmatrix}3&1&2&0\\0&4&5&1\end{pmatrix},b=\begin{pmatrix}3\\6\end{pmatrix}$;

（2）$A = \begin{pmatrix} 3 & 2 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \\ 0 & 3 & 0 & 0 & 1 \end{pmatrix}, b = \begin{pmatrix} 65 \\ 40 \\ 75 \end{pmatrix}$.

【解】

【解毕】

定理 1.17 （极方向特征）

设多面体 $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq 0\}$，其中 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 行满秩，则

（1）$\boldsymbol{d}$ 是 S 的方向的充要条件是 $\boldsymbol{A}\boldsymbol{d} = 0$ 且 $\boldsymbol{d} \geq 0$;

（2）$\boldsymbol{d}$ 是 S 的极方向的充要条件是存在分解 $\boldsymbol{A} = (\boldsymbol{B}, \boldsymbol{N})$，使得

（i）$\boldsymbol{B}$ 为 m 阶可逆阵；

（ii）$\boldsymbol{d} = (\boldsymbol{d}_{\boldsymbol{B}}^\top, \boldsymbol{d}_{\boldsymbol{N}}^\top)^\top$（分解方式与 $\boldsymbol{A}$ 相同），其中 $\boldsymbol{d}_{\boldsymbol{B}} = -\boldsymbol{B}^{-1}\boldsymbol{a}_j \geq 0$，$\boldsymbol{d}_{\boldsymbol{N}} = \boldsymbol{e}_j$. 这里 $\boldsymbol{e}_j$ 是第 j 个元素为 1，其余元素为 0 的单位向量. j 是使得 $\boldsymbol{N}$ 中的列向量 $\boldsymbol{a}_j$ 满足 $\boldsymbol{B}^{-1}\boldsymbol{a}_j \leq 0$ 的下标.

【证明】

【证毕】

例 1.7 证明：多面体 $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq 0\}$ 中必有极点，且个数不超过 $(n-m)C_n^m$.

【证明】

【证毕】

例 1.8 设多面体 $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq 0\}$，其中 $A = \begin{pmatrix} 3 & 1 & 2 & -6 & 0 \\ 0 & 4 & 5 & -5 & 1 \end{pmatrix}, b = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$，求该多面体的极方向.

【解】

【解毕】

定理 1.18 （表示定理，Minkowski-Weyl）

设多面体 $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq 0\}$，其中 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 行满秩，$\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(k)}$ 为所有极点，$\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(l)}$ 为所有极方向，则 $\forall \boldsymbol{x} \in S$，存在序列 $\{\lambda_i\}_{i=1}^k, \{\mu_j\}_{j=1}^l$，其中：

（1）$\lambda_i > 0$，$i = 1, 2, \cdots, k$，且 $\sum\limits_{i=1}^k \lambda_i = 1$;

（2）$\mu_j \geq 0$，$j = 1, 2, \cdots, l$，使得

$$\begin{aligned} \boldsymbol{x} = & \lambda_1 \boldsymbol{x}^{(1)} + \lambda_2 \boldsymbol{x}^{(2)} + \cdots + \lambda_k \boldsymbol{x}^{(k)} && \text{（极点的凸组合）} \\ & + \mu_1 \boldsymbol{d}^{(1)} + \mu_2 \boldsymbol{d}^{(2)} + \cdots + \mu_l \boldsymbol{d}^{(l)} && \text{（极方向的半正组合）} \end{aligned}$$

特别地，如果上述多面体 S 是有界的，则有 $\mu_j = 0$.

【证明】

【证毕】

第二章 线性规划

2.1 线性规划模型

2.1.1 线性规划的定义

定义 2.1 （线性规划的一般形式）

$$
\begin{aligned}
\max(\min) \quad & z = c_1x_1 + c_2x_2 + \cdots + c_nx_n \\
\text{s.t.} \quad & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \le (=, \ge) b_1 \\
& a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \le (=, \ge) b_2 \\
& \vdots \\
& a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \le (=, \ge) b_m \\
& x_1, x_2, \cdots, x_n \in \mathbb{R}
\end{aligned}
$$

我们为了方便解决问题，常把上述一般形式转化为标准形式：

定义 2.2 （线性规划的标准形式）

$$
\begin{aligned}
\max \quad & z = c_1x_1 + c_2x_2 + \cdots + c_nx_n \\
\text{s.t.} \quad & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\
& a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\
& \vdots \\
& a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \\
& x_1, x_2, \cdots, x_n \ge 0
\end{aligned}
$$

为了研究问题的方便，我们将其写成矩阵形式

$$
\begin{aligned}
\max \quad & z = \boldsymbol{c}^\top \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\
& \boldsymbol{x} \ge 0
\end{aligned}
$$

其中，$\boldsymbol{c}, \boldsymbol{x} \in \mathbb{R}^n$，$\boldsymbol{b} \in \mathbb{R}^m$，$\boldsymbol{A} \in \mathbb{R}^{m\times n}$.

不难发现，标准形式有区别于一般形式的以下四个特点：

- 目标最大化；
- 约束为等式；
- 决策变量均非负；
- 右端项非负.

接下来我们将会讨论如何将一般形式一步步转化为标准形式.

2.1.2　将一般形式转化为标准形式

1. 极小目标极大化

在标准形式中，要求优化目标为极大化目标函数值. 若目标函数为

$$\min f = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

则令 $z = -f$，从而上述极小化问题与下面的极大化问题

$$\max z = -c_1x_1 - c_2x_2 - \cdots - c_nx_n$$

有相同的最优解，但它们的最优解的目标函数值相差一个符号，即

$$\min f = -\max z$$

2. 约束条件不为等式

在标准形式中，要求每约束条件均为等式. 若约束条件为小于等于号，即设第 i 个约束条件为

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \le b_i$$

则可以引入一个松弛变量 s，使得

$$s = b_i - (a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n)$$

从而 $s \ge 0$，也具有非负约束，这时新的约束条件成为

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n + s = b_i$$

若约束条件为大于等于号，即设第 i 个约束条件为

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \ge b_i$$

相似地，也可以引入一个松弛变量 s，使得

$$s = (a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n) - b_i$$

从而 $s \ge 0$，也具有非负约束，这时新的约束条件成为

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - s = b_i$$

3. 变量无符号限制

在标准形式中，要求每一个变量均有非负约束. 当某一个变量 x_j 没有非负约束时，可以令

$$x_j = x_j' - x_j''$$

其中 $x_j', x_j'' \ge 0$，即用两个非负数之差表示一个无符号限制的变量.

4. 右端项有负值

在标准形式中，要求每一个约束条件右端均为正值. 当某一个右端项系数为负时，假设是第 i 个，即

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i \leq 0$$

则把该等式两端同时乘以 -1，得到

$$-a_{i1}x_1 - a_{i2}x_2 - \cdots - a_{in}x_n = -b_i \geq 0$$