目录

1　线性规划

1.1　基本概念

定义: 称

$$\begin{cases} \min & z = c_1x_1 + \cdots + c_nx_n \\ \text{s.t.} & a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i, \quad i = 1, \cdots, p \\ & a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \geqslant b_i, \quad i = p+1, \cdots, m \\ & x_j \geqslant 0, \quad j = 1, \cdots, q \\ & x_j \gtreqless 0, \quad j = q+1, \cdots, n \end{cases}$$

为线性规划的**一般形式**.

定义: 称

$$\begin{cases} \min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{b} \\ & \quad \boldsymbol{x} \geqslant \boldsymbol{0} \end{cases}$$

为线性规划的**规范形式**.

定义: 称

$$\begin{cases} \min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \quad \boldsymbol{x} \geqslant \boldsymbol{0} \end{cases}$$

为线性规划的**标准形式**, 其中 $\boldsymbol{A} \in \mathbf{R}^{m\times n}$ 称为**约束矩阵**, $\boldsymbol{b} \in \mathbf{R}^m$ 称为**右端向量**, $\boldsymbol{c} \in \mathbf{R}^n$ 称为**价值向量**.

对于线性规划的一般形式与规范形式, 要能够将其化为标准形式. 其中需要特别注意的是松弛变量与剩余变量的添加.

1.2　可行域与基本可行解

设线性规划

$$\begin{cases} \min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \quad \boldsymbol{x} \geqslant \boldsymbol{0} \end{cases}$$

的约束矩阵 $\boldsymbol{A} = (\boldsymbol{B}, \boldsymbol{N})$, 其中 $\boldsymbol{B}$ 为 $\boldsymbol{A}$ 的一个 m 阶满秩子方阵. 由 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 得

$$\boldsymbol{B}\boldsymbol{x}_{\boldsymbol{B}} + \boldsymbol{N}\boldsymbol{x}_{\boldsymbol{N}} = \boldsymbol{b}.$$

定义: 矩阵 $\boldsymbol{B}$ 称为一个**基**, 它的列向量称为**基向量**, $\boldsymbol{x}$ 与之对应的分量称为**基变量**, 其余分量称为**非基变量**. 令所有的非基变量取 0, 得到的解 $\overline{\boldsymbol{x}} = \begin{pmatrix} \boldsymbol{B}^{-1}\boldsymbol{b} \\ \boldsymbol{0} \end{pmatrix}$ 称为相应于 $\boldsymbol{B}$ 的**基本解**. 若 $\boldsymbol{B}^{-1}\boldsymbol{b} \geqslant \boldsymbol{0}$, 则

称 $\bar{\boldsymbol{x}}$ 为**基本可行解**, 对应的基 $\boldsymbol{B}$ 称为**可行基**.

定义: 设 $S \subset \mathbf{R}^n$, 若对任意的 $\boldsymbol{x}, \boldsymbol{y} \in S$, 对任意的 $\lambda \in [0,1]$, 都有

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

则称 S 为 $\mathbf{R}^n$ 中的**凸集**.

定义: $H = \{\boldsymbol{x} \in \mathbf{R}^n \mid \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} = b\}$ 称为 $\mathbf{R}^n$ 中的**超平面**, 其中 $\boldsymbol{a} \in \mathbf{R}^n$ 非零, $b \in \mathbf{R}$.

易知超平面是平面直线、三维空间的平面在高维空间的推广. 且由 H 产生的两个闭的半空间

$$H^{+} = \{\boldsymbol{x} \in \mathbf{R}^n \mid \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \geqslant b\}$$

$$H^{-} = \{\boldsymbol{x} \in \mathbf{R}^n \mid \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \leqslant b\}$$

也是凸集.

定义: 称

$$S = \left\{\boldsymbol{x} \in \mathbf{R}^n \mid \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} = b_i, i = 1, \cdots, p;\ \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} \geqslant b_i, i = p+1, \cdots, p+q\right\}$$

为 $\mathbf{R}^n$ 中的**多面凸集**. 非空有界多面凸集称为**多面体**.

定义: 设 S 为 $\mathbf{R}^n$ 中的凸集, $\boldsymbol{x} \in S$. 若对任意的 $\boldsymbol{y}, \boldsymbol{z} \in S, \boldsymbol{y} \neq \boldsymbol{z}$, 对任意的 $\lambda \in (0,1)$, 都有

$$\boldsymbol{x} \neq \lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{z},$$

则称 $\boldsymbol{x}$ 为凸集 S 的**顶点**.

定理: 可行域 $D = \{\boldsymbol{x} \in \mathbf{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geqslant \boldsymbol{0}\}$ 为凸集.

定理: 任意多个凸集的交仍是凸集.

定理: 可行解 $\bar{\boldsymbol{x}}$ 是基本可行解当且仅当 $\bar{\boldsymbol{x}}$ 正分量对应 $\boldsymbol{A}$ 的列向量线性无关.

定理: $\bar{\boldsymbol{x}}$ 是基本可行解当且仅当 $\bar{\boldsymbol{x}}$ 是可行域 D 的顶点.

例　证明: 若一个线性规划问题在两个顶点处达到最优值, 则此问题有无穷多个最优解.

证明　设 LP 问题在两个不同顶点 $\boldsymbol{x}^1, \boldsymbol{x}^2$ 处达到最优值, 即

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^1 = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^2 = \min \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}.$$

任取 $\lambda \in (0,1)$, 因可行域为凸集, 故 $\lambda\boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2$ 为可行解. 又

$$\boldsymbol{c}^{\mathrm{T}}[\lambda\boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2] = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^1 = \min \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},$$

故 $\lambda\boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2$ 为最优解, 由 λ 的任意性知该 LP 问题有无穷多个最优解.

定义: 若基本可行解 $\overline{\boldsymbol{x}}$ 的所有基变量均为正数, 则称该基本可行解是**非退化**的. 若一个线性规划问题的所有基本可行解都非退化, 则称该线性规划是非退化的.

定理 (线性规划基本定理):

(1) 若标准线性规划有可行解, 则有基本可行解;

(2) 若标准线性规划具有有限的最优值, 则有基本可行解为最优解.

1.3 单纯形法

设标准形式的 LP 问题有非退化的基本可行解 $\overline{\boldsymbol{x}} = \begin{pmatrix} \boldsymbol{B}^{-1}\boldsymbol{b} \\ \boldsymbol{0} \end{pmatrix} = \begin{pmatrix} \overline{\boldsymbol{b}} \\ \boldsymbol{0} \end{pmatrix}$, $\boldsymbol{A} = (\boldsymbol{B}, \boldsymbol{N}), \boldsymbol{c}^{\mathrm{T}} = (\boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}, \boldsymbol{c}_{\boldsymbol{N}}^{\mathrm{T}})$, $\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_{\boldsymbol{B}} \\ \boldsymbol{x}_{\boldsymbol{N}} \end{pmatrix}$, $\boldsymbol{x}_{\boldsymbol{B}} + \boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{x}_{\boldsymbol{N}} = \overline{\boldsymbol{b}}$, 则

$$
\begin{aligned}
z = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} &= \boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{x}_{\boldsymbol{B}} + \boldsymbol{c}_{\boldsymbol{N}}^{\mathrm{T}}\boldsymbol{x}_{\boldsymbol{N}} \\
&= \boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}(\overline{\boldsymbol{b}} - \boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{x}_{\boldsymbol{N}}) + \boldsymbol{c}_{\boldsymbol{N}}^{\mathrm{T}}\boldsymbol{x}_{\boldsymbol{N}} \\
&= \boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}\overline{\boldsymbol{b}} - (\boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{N} - \boldsymbol{c}_{\boldsymbol{N}}^{\mathrm{T}})\boldsymbol{x}_{\boldsymbol{N}}.
\end{aligned}
$$

记 $\xi_j = \boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{A}_j - c_j$, 称其为**检验数**, $\boldsymbol{\xi}^{\mathrm{T}} = \boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{A} - \boldsymbol{c}^{\mathrm{T}}$ 称为**检验数向量**. $z_0 = \boldsymbol{c}^{\mathrm{T}}\overline{\boldsymbol{x}} = \boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}\overline{\boldsymbol{b}}$. 易知 $\boldsymbol{\xi}^{\mathrm{T}} = (\boldsymbol{\xi}_{\boldsymbol{B}}^{\mathrm{T}}, \boldsymbol{\xi}_{\boldsymbol{N}}^{\mathrm{T}}) = (\boldsymbol{0}, \boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{N} - \boldsymbol{c}_{\boldsymbol{N}}^{\mathrm{T}})$ (与基变量对应的检验数为 0), 从而

$$
z = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} = z_0 - \boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{x} = z_0 - \sum_{j=m+1}^{n} \xi_j x_j.
$$

由上述讨论知, 对非退化的基本可行解 $\overline{\boldsymbol{x}} = \begin{pmatrix} \boldsymbol{B}^{-1}\boldsymbol{b} \\ \boldsymbol{0} \end{pmatrix} = \begin{pmatrix} \overline{\boldsymbol{b}} \\ \boldsymbol{0} \end{pmatrix}$, 原线性规划

$$
\begin{cases}
\min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\
\text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\
& \quad \boldsymbol{x} \geqslant \boldsymbol{0}
\end{cases}
$$

等价于

$$
\begin{cases}
\min & z = z_0 - \boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{x} \\
\text{s.t.} & \boldsymbol{x}_{\boldsymbol{B}} + \boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{x}_{\boldsymbol{N}} = \overline{\boldsymbol{b}} \\
& \quad \boldsymbol{x} \geqslant \boldsymbol{0}
\end{cases}
$$

定理 (最优性准则): 若 $\boldsymbol{\xi} \leqslant \boldsymbol{0}$, 则 $\overline{\boldsymbol{x}}$ 为原问题的最优解.

证明　对任意可行解 $\boldsymbol{x}$, 因为 $\boldsymbol{\xi} \leqslant 0$, $\boldsymbol{x} \geqslant 0$, 故 $\boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{x} \leqslant 0$, 从而

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} = z_0 - \boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{x} \geqslant z_0 = \boldsymbol{c}^{\mathrm{T}}\overline{\boldsymbol{x}},$$

故 $\overline{\boldsymbol{x}}$ 为最优解.

由上述证明容易得到, $\overline{\boldsymbol{x}}$ 为唯一最优解当且仅当非基变量对应的检验数 $\xi_j < 0$.

定理: 若 $\xi_k > 0, \overline{\boldsymbol{A}}_k = \boldsymbol{B}^{-1}\boldsymbol{A}_k \leqslant \boldsymbol{0}$, 则原问题无界.

证明　令 $\boldsymbol{d} = \begin{pmatrix} -\overline{\boldsymbol{A}}_k \\ \boldsymbol{0} \end{pmatrix} + \boldsymbol{e}_k$, 其中 $\boldsymbol{e}_k$ 表示第 k 个分量为 1, 其余分量为 0 的 n 维向量. 由于 $\overline{\boldsymbol{A}}_k \leqslant 0$, 故 $\boldsymbol{d} \geqslant 0$, 且

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{d} &= (\boldsymbol{B}, \boldsymbol{N}) \begin{pmatrix} -\overline{\boldsymbol{A}}_k \\ \boldsymbol{0} \end{pmatrix} + \boldsymbol{A}\boldsymbol{e}_k \\
&= (\boldsymbol{B}, \boldsymbol{N}) \begin{pmatrix} -\boldsymbol{B}^{-1}\boldsymbol{A}_k \\ \boldsymbol{0} \end{pmatrix} + (\boldsymbol{A}_1, \cdots, \boldsymbol{A}_k, \cdots \boldsymbol{A}_n) \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\
&= -\boldsymbol{A}_k + \boldsymbol{A}_k \\
&= \boldsymbol{0}.
\end{aligned}$$

任取正数 θ, 考察 $\overline{\boldsymbol{x}} + \theta\boldsymbol{d}$, 有

$$\boldsymbol{A}(\overline{\boldsymbol{x}} + \theta\boldsymbol{d}) = \boldsymbol{A}\overline{\boldsymbol{x}} + \theta\boldsymbol{A}\boldsymbol{d} = \boldsymbol{b},$$
$$\overline{\boldsymbol{x}} + \theta\boldsymbol{d} \geqslant 0,$$

故 $\overline{\boldsymbol{x}} + \theta\boldsymbol{d}$ 为原问题可行解. 但

$$\begin{aligned}
\boldsymbol{c}^{\mathrm{T}}(\overline{\boldsymbol{x}} + \theta\boldsymbol{d}) &= \boldsymbol{c}^{\mathrm{T}}\overline{\boldsymbol{x}} + \theta\boldsymbol{c}^{\mathrm{T}}\boldsymbol{d} \\
&= z_0 + \theta(\boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}, \boldsymbol{c}_{\boldsymbol{N}}^{\mathrm{T}}) \begin{pmatrix} -\overline{\boldsymbol{A}}_k \\ \boldsymbol{0} \end{pmatrix} + \theta\boldsymbol{c}^{\mathrm{T}}\boldsymbol{e}_k \\
&= z_0 - \theta(\boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{A}_k - c_k) \\
&= z_0 - \theta\xi_k.
\end{aligned}$$

因 $\xi_k > 0$, 令 $\theta \to +\infty$, 即得原问题无界.

定理: 对非退化的基本可行解 $\overline{\boldsymbol{x}}$, 若 $\xi_k > 0, \overline{\boldsymbol{A}}_k = \boldsymbol{B}^{-1}\boldsymbol{A}_k$ 存在正分量, 则有新的基本可行解 $\hat{\boldsymbol{x}}$, 使

得 $\boldsymbol{c}^{\mathrm{T}}\hat{\boldsymbol{x}} < \boldsymbol{c}^{\mathrm{T}}\overline{\boldsymbol{x}}$.

证明 仍然令 $\boldsymbol{d} = \begin{pmatrix} -\overline{\boldsymbol{A}}_k \\ \boldsymbol{0} \end{pmatrix} + \boldsymbol{e}_k$, 则有 $\boldsymbol{A}\boldsymbol{d} = \boldsymbol{0}$. 又令 $\hat{\boldsymbol{x}} = \overline{\boldsymbol{x}} + \theta\boldsymbol{d}$, 即

$$\hat{\boldsymbol{x}} = \begin{pmatrix} \overline{\boldsymbol{b}} - \theta\overline{\boldsymbol{A}}_k \\ \boldsymbol{0} \end{pmatrix} + \theta\boldsymbol{e}_k.$$

为使 $\hat{\boldsymbol{x}} \geqslant \boldsymbol{0}$, 取

$$\theta = \min\left\{ \frac{\overline{b}_i}{\overline{a}_{ik}} \middle| \overline{a}_{ik} > 0, i = 1, \cdots, m \right\} = \frac{\overline{b}_r}{\overline{a}_{rk}},$$

又因为

$$\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}(\overline{\boldsymbol{x}} + \theta\boldsymbol{d}) = \boldsymbol{A}\overline{\boldsymbol{x}} + \theta\boldsymbol{A}\boldsymbol{d} = \boldsymbol{b},$$

故 $\hat{\boldsymbol{x}}$ 为原问题可行解. 下证 $\hat{\boldsymbol{x}}$ 为基本解. $\hat{\boldsymbol{x}}$ 的各分量为

$$\begin{aligned}
&\hat{x}_i = \overline{b}_i - \frac{\overline{b}_r}{\overline{a}_{rk}}\overline{a}_{ik}, \quad i = 1, \cdots, m, \quad i \neq r, \\
&\hat{x}_r = 0, \\
&\hat{x}_k = \frac{\overline{b}_r}{\overline{a}_{rk}}, \\
&\hat{x}_j = 0, \quad j = m+1, \cdots, n, \quad j \neq k.
\end{aligned}$$

因此只需说明 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_{r-1}, \boldsymbol{A}_k, \boldsymbol{A}_{r+1}, \cdots, \boldsymbol{A}_m$ 线性无关即可. 假设它们线性相关, 由于原向量组 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_m$ 线性无关, 故可设

$$A_k = \sum_{\substack{i=1 \\ i \neq r}}^{m} y_i \boldsymbol{A}_i \tag{1}$$

又由 $\overline{\boldsymbol{A}}_k = \boldsymbol{B}^{-1}\boldsymbol{A}_k$ 知

$$\boldsymbol{A}_k = \boldsymbol{B}\overline{\boldsymbol{A}}_k = (\boldsymbol{A}_1, \cdots \boldsymbol{A}_m)\begin{pmatrix} \overline{a}_{1k} \\ \vdots \\ \overline{a}_{mk} \end{pmatrix} = \sum_{i=1}^{m} \overline{a}_{ik} A_i \tag{2}$$

上面两式相减, 得

$$\overline{a}_{rk}\boldsymbol{A}_r + \sum_{\substack{i=1 \\ i \neq r}}^{m} (\overline{a}_{ik} - y_i)\boldsymbol{A}_i = \boldsymbol{0}.$$

因为 $\overline{a}_{rk} \neq 0$, 故 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_m$ 线性相关, 矛盾. 从而 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_{r-1}, \boldsymbol{A}_k, \boldsymbol{A}_{r+1}, \cdots, \boldsymbol{A}_m$ 线性无关, $\hat{\boldsymbol{x}}$ 为基本可行解.

因为 $\overline{\boldsymbol{x}}$ 非退化, 故 $\overline{\boldsymbol{b}} > \boldsymbol{0}$, 从而 $\theta = \dfrac{\overline{b}_r}{\overline{a}_{rk}} > 0$, 故

$$\boldsymbol{c}^{\mathrm{T}}\hat{\boldsymbol{x}} = z_0 - \theta\xi_k < z_0 = \boldsymbol{c}^{\mathrm{T}}\overline{\boldsymbol{x}}.$$

单纯形法步骤:

(1) 列出初始单纯形表 (含典式);

(2) 求 $\xi_k = \max\{\xi_j\}$, 若 $\xi_k \leqslant 0$, 停止, 已找到最优解 $\bar{\boldsymbol{x}}$;

(3) $\overline{\boldsymbol{A}}_k = \boldsymbol{B}^{-1}\boldsymbol{A}_k \leqslant \boldsymbol{0}$, 停止, 原问题无界;

(4) 求 $\theta = \min\left\{\dfrac{\bar{b}_i}{\bar{a}_{ik}}\middle|\bar{a}_{ik} > 0, i = 1, \cdots, m\right\} = \dfrac{\bar{b}_r}{\bar{a}_{rk}}$;

(5) 以 x_k 为进基变量, x_r 为出基变量, $\bar{a}_{rk}$ 为转轴元进行转轴变换, 转 (2).

1.4 初始解

对线性规划

$$\begin{cases} \min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \quad (\boldsymbol{b} \geqslant \boldsymbol{0}) \\ & \quad \boldsymbol{x} \geqslant \boldsymbol{0} \end{cases}$$

添加人工变量 $\boldsymbol{x}_a = (x_{n+1}, \cdots, x_{n+m})^{\mathrm{T}}$, 得到**辅助问题**

$$\begin{cases} \min & g = \sum\limits_{i=n+1}^{n+m} x_i \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}_a = \boldsymbol{b} \\ & \boldsymbol{x} \geqslant \boldsymbol{0}, \boldsymbol{x}_a \geqslant \boldsymbol{0} \end{cases}$$

以 $\boldsymbol{x}_a$ 为基变量, 利用单纯形法对辅助问题求解, 此时有三种情况:

(1) 辅助问题最优值 $g = 0$, 人工变量均为非基变量, 此时已经得到原问题的一个基本可行解;

(2) 辅助问题最优值 $g = 0$, 人工变量含有基变量, 此时仍然得到原问题的一个基本可行解, 只是该基本可行解是退化的 (对应的基不惟一);

(3) 辅助问题最优值 $g > 0$, 此时原问题无解.

1.5 对偶理论

定义: 给定一个 LP 问题, 它的对偶问题定义如下:

原始 (P)			对偶 (D)	
min	$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$		max	$\boldsymbol{b}^{\mathrm{T}}\boldsymbol{w}$
s.t.	$\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} = b_i$	$i = 1, \cdots, p$	s.t.	$w_i \gtrless 0$
	$\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} \geqslant b_i$	$i = p+1, \cdots, m$		$w_i \geqslant 0$
	$x_j \geqslant 0$	$j = 1, \cdots, q$		$\boldsymbol{A}_j^{\mathrm{T}}\boldsymbol{w} \leqslant c_j$
	$x_j \gtrless 0$	$j = q+1, \cdots, n$		$\boldsymbol{A}_j^{\mathrm{T}}\boldsymbol{w} = c_j$

由此定义, 易得标准形式的对偶规划为

$$\begin{cases} \max & \boldsymbol{b}^{\mathrm{T}}\boldsymbol{w} \\ \text{s.t.} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{w} \leqslant \boldsymbol{c} \\ & \boldsymbol{w} \gtrless \boldsymbol{0} \end{cases}$$

规范形式的对偶问题为

$$\begin{cases} \max & \boldsymbol{b}^{\mathrm{T}}\boldsymbol{w} \\ \text{s.t.} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{w} \leqslant \boldsymbol{c} \\ & \boldsymbol{w} \geqslant \boldsymbol{0} \end{cases}$$

定理 (弱对偶定理): 设有标准形式原问题

$$\begin{cases} \min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geqslant \boldsymbol{0} \end{cases} \tag{P}$$

和其对偶问题

$$\begin{cases} \max & \boldsymbol{b}^{\mathrm{T}}\boldsymbol{w} \\ \text{s.t.} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{w} \leqslant \boldsymbol{c} \\ & \boldsymbol{w} \gtrless \boldsymbol{0} \end{cases} \tag{D}$$

则对原问题任一可行解 $\boldsymbol{x}$ 与对偶问题任一可行解 $\boldsymbol{w}$, 有

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \geqslant \boldsymbol{b}^{\mathrm{T}}\boldsymbol{w}.$$

证明

$$
\begin{aligned}
\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{c} \\
&\geqslant \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{w} \quad (\text{因为}\boldsymbol{x} \geqslant \boldsymbol{0}) \\
&= \boldsymbol{b}^{\mathrm{T}}\boldsymbol{w},
\end{aligned}
$$

其中第二行是因为 $\boldsymbol{w}$ 为对偶问题可行解, 第三行是因为 $\boldsymbol{x}$ 是原问题可行解.

定理 (强对偶定理): 若原问题有最优解, 则对偶问题有最优解, 且它们的最优值相等.

证明 设 $\overline{\boldsymbol{x}}$ 为原问题最优基本可行解, 对应的可行基为 $\boldsymbol{B}$. 取

$$\boldsymbol{w}^{\mathrm{T}} = \boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{B}^{-1},$$

则

$$(\boldsymbol{\xi}^{\mathrm{T}} =)\, \boldsymbol{w}^{\mathrm{T}}\boldsymbol{A} - \boldsymbol{c}^{\mathrm{T}} \leqslant 0,$$

从而 $\boldsymbol{w}$ 为对偶问题可行解. 且

$$\boldsymbol{c}^{\mathrm{T}}\overline{\boldsymbol{x}} = \boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{x}_{\boldsymbol{B}} = \boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{b} = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{w}.$$

故由弱对偶定理知 $\boldsymbol{w}$ 为对偶问题最优解.

定理: 若原线性规划问题有最优解, 则其对偶问题有最优解, 且它们的最优值相等.

推论: 设 $\boldsymbol{x}, \boldsymbol{w}$ 为原问题与对偶问题的可行解, 则 $\boldsymbol{x}, \boldsymbol{w}$ 为最优解当且仅当 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{w}$.

定理 (对称性): 对偶问题的对偶为原问题.

定理: 对于给定的线性规划及其对偶问题, 下述三种情况恰有一种发生.

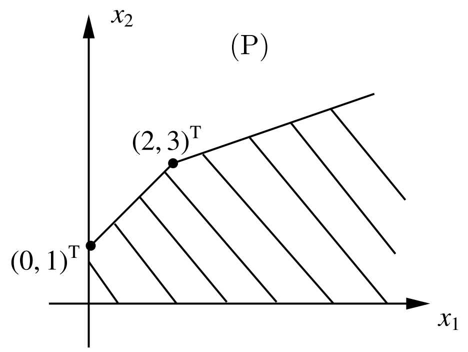
原始	对偶		
	有最优解	问题无界	无可行解
有最优解	(1)	×	×
问题无界	×	×	(3)
无可行解	×	(3)	(2)

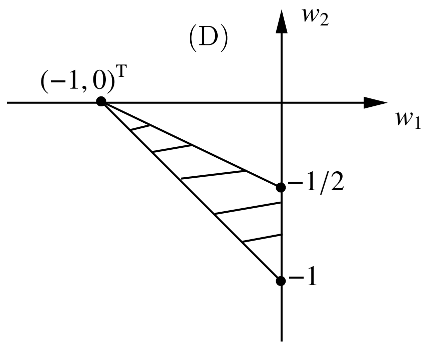
注意: 互为对偶的两个 LP 问题, 它们的最优解个数之间没有必然联系.

例　考察互为对偶的问题

$$(\mathrm{P})\left\{\begin{array}{ll}\min & x_1 - x_2 \\ \text{s.t.} & -x_1 + x_2 \leqslant 1 \\ & -x_1 + 2x_2 \leqslant 4 \\ & x_1, x_2 \geqslant 0\end{array}\right. \qquad (\mathrm{D})\left\{\begin{array}{ll}\max & w_1 + 4w_2 \\ \text{s.t.} & -w_1 - w_2 \leqslant 1 \\ & w_1 + 2w_2 \leqslant -1 \\ & w_1, w_2 \leqslant 0\end{array}\right.$$

它们的可行域如下:





问题 (P) 有无穷多个最优解 (事实上, 从 $(0,1)^{\mathrm{T}}$ 到 $(2,3)^{\mathrm{T}}$ 的这条线段上每一点均为问题 (P) 的最优解), 问题 (D) 有唯一最优解 $(-1,0)^{\mathrm{T}}$. 由此可见, 一个 LP 问题有唯一最优解, 但其对偶问题的最优解不一定唯一.

定理 (互补松紧性): 设 $\boldsymbol{x}, \boldsymbol{w}$ 为原问题与对偶问题的可行解, 则 $\boldsymbol{x}, \boldsymbol{w}$ 为最优解当且仅当

$$(\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} - b_i)\, w_i = 0,\ i = 1, 2, \cdots, m;$$
$$(c_j - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{A}_j) x_j = 0,\ j = 1, 2, \cdots, n.$$

互补松紧条件说明, 对于最优解 $\boldsymbol{x}$ 与 $\boldsymbol{w}$, 如果一个问题中的约束取严格不等式, 则另一个问题的对应变量必取值为 0; 如果一个问题中的非负变量取正值, 则另一个问题对应的约束必取等式.

定义: 使得检验数向量 $\boldsymbol{\xi} \leqslant \boldsymbol{0}$ 的基本解称为**正则解**.

对偶单纯形法步骤:

(1) 列出包含正则解的初始单纯形表;

(2) 求 $\bar{b}_r = \min\{\bar{b}_i\}$;

(3) 若 $\bar{b}_r \geqslant 0$, 停止, 已找到最优解;

(4) 若 $\bar{a}_{rj} \geqslant 0,\ j = 1, 2, \cdots, n$, 停止, 原问题无可行解;

(5) 求 $\min\left\{\dfrac{\xi_j}{\bar{a}_{rj}} \,\middle|\, \bar{a}_{rj} < 0,\ j = 1, 2, \cdots, n\right\} = \dfrac{\xi_k}{\bar{a}_{rk}}$;

(6) 以 $\bar{a}_{rk}$ 为转轴元进行转轴变换, 转 (2).

1.6 灵敏度分析

设线性规划

$$\begin{cases} \min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geqslant \boldsymbol{0} \end{cases}$$

的最优单纯形表为

	$\boldsymbol{x}_B$	$\boldsymbol{x}_N$	RHS
z	$\boldsymbol{0}$	$\boldsymbol{c}_B^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{N} - \boldsymbol{c}_N^{\mathrm{T}}$	$\boldsymbol{c}_B^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{b}$
$\boldsymbol{x}_B$	$\boldsymbol{I}$	$\boldsymbol{B}^{-1}\boldsymbol{N}$	$\boldsymbol{B}^{-1}\boldsymbol{b}$

一、改变价值向量 $\boldsymbol{c}$

因为 $\boldsymbol{A}, \boldsymbol{b}$ 不变, 故可行域不变. 由最优单纯形表知, 只有第 0 行发生改变. 此时

$$\boldsymbol{\xi}_N'^{\mathrm{T}} = \boldsymbol{c}_B'^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{N} - \boldsymbol{c}_N'^{\mathrm{T}},$$
$$z_0' = \boldsymbol{c}_B'^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{b}.$$

当只有一个分量发生改变时, 情况更为简单. 设 c_k 变为 c_k'.

(1) c_k 对应非基变量, 此时

$$\xi_k' = \xi_k + (c_k - c_k').$$

若 $\xi_k' \leqslant 0$, 即 $c_k' \geqslant \xi_k + c_k$, 则原问题最优解仍为新问题最优解; 若 $\xi_k' > 0$, 则用单纯形法继续求解.

(2) c_k 对应基变量, 且对应单纯形表的第 l 行. 此时将单纯形表的第 l 行乘以 $(c_k' - c_k)$ 倍加到第 0 行上去, 再令 $\xi_k' = 0$ 即得新问题的单纯形表.

二、改变右端向量 $\boldsymbol{b}$

由最优单纯形表知, 只有 RHS 列发生改变. 此时

$$\overline{\boldsymbol{b}}' = \boldsymbol{B}^{-1}\boldsymbol{b}',\ z_0' = \boldsymbol{c}_B^{\mathrm{T}}\overline{\boldsymbol{b}}'.$$

因为 $\boldsymbol{\xi}' = \boldsymbol{\xi} \leqslant \boldsymbol{0}$, 从而若 $\overline{\boldsymbol{b}}' \geqslant \boldsymbol{0}$, 则已找到新问题最优解 (但与原最优解不一定相同, 因为最优值发生了改变, 而价值向量 $\boldsymbol{c}$ 不变), 否则用对偶单纯形法继续求解.

若只改变一个分量, 设 b_s 变为 b_s', 此时

$$\overline{\boldsymbol{b}}' = \overline{\boldsymbol{b}} + (\boldsymbol{b}_s' - \boldsymbol{b}_s)\boldsymbol{B}_s^{-1},$$

其中 $\boldsymbol{B}_s^{-1}$ 表示 $\boldsymbol{B}^{-1}$ 的第 s 列.

2 整数线性规划

2.1 Gomory 割平面法

考虑纯整数线性规划 (ILP)

$$\begin{cases} \min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geqslant \boldsymbol{0} \\ & \boldsymbol{x}\text{为整数向量} \end{cases} \tag{P}$$

对应的松弛问题

$$\begin{cases} \min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geqslant \boldsymbol{0} \end{cases} \tag{P$_0$}$$

设问题 (P_0) 最优单纯形表中的典式为

$$z + \sum_{j\in\overline{S}} \zeta_j x_j = z_0$$
$$x_{B_i} + \sum_{j\in\overline{S}} \overline{a}_{ij} x_j = \overline{b}_i, \quad i = 1, \cdots, m$$

其中 $\overline{S}$ 为非基变量的下标集合. 若 $\overline{b}_i, i = 1, \cdots, m$ 均为整数, 则已经得到原 ILP 问题的最优解; 若 $\overline{b}_l$ 不是整数, 设其对应方程为

$$x_{B_l} + \sum_{j\in\overline{S}} \overline{a}_{lj} x_j = \overline{b}_l.$$

我们用 $[a]$ 表示不超过 a 的最大整数, 则

$$\begin{cases} \overline{a}_{lj} = [\overline{a}_{lj}] + f_{lj}, & 0 \leqslant f_{lj} < 1, j \in \overline{S} \\ \overline{b}_l = [\overline{b}_l] + f_l, & 0 < f_l < 1 \end{cases}$$

易知

$$x_{B_l} + \sum_{j\in\overline{S}} [\overline{a}_{lj}] x_j \leqslant \overline{b}_l$$

从而

$$x_{B_l} + \sum_{j\in\overline{S}} [\overline{a}_{lj}] x_j \leqslant [\overline{b}_l]$$

故

$$x_{B_l} + \sum_{j\in\overline{S}} (\overline{a}_{lj} - [\overline{a}_{lj}]) x_j = \overline{b}_l - [\overline{b}_l]$$

即

$$\sum_{j\in\overline{S}} f_{lj}x_j \geqslant f_l.$$

上式称为**生成行 l 的** Gomory **割平面条件**. 添加松弛变量 s 得超平面方程

$$-\sum_{j\in\overline{S}} f_{lj}x_j + s = -f_l,$$

称其为**割平面**.

当 $\overline{b}_l$ 不是整数时, 因为 $-f_l < 0$ 且表中第 0 行没有改变, 因此将割平面添加到松弛问题的最优单纯形表中得到的是基本不可行解与对偶可行解.

定理: 将割平面添加到松弛问题的最优单纯形表中, 不会割掉原 ILP 问题的任何可行点. 当 $\overline{b}_l$ 不为整数时, 新表得到的是对偶可行解.

当松弛问题 (P_0) 的最优基本可行解有多个非整数分量时, 任意选取其中一个分量开始 Gomory 割平面法即可. 若添加割平面求解后仍未得到整数解, 则继续对非整数分量添加割平面, 直到得到整数解或者判定原 ILP 问题不可行 (无界或无可行解).

2.2 分枝定界法

分枝定界法可用于求解纯整数规划与混合整数规划.

一、分枝

设整数线性规划问题

$$\begin{cases} \min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b} \\ & \boldsymbol{x} \geqslant \boldsymbol{0} \\ & \boldsymbol{x}\text{为整数向量} \end{cases} \tag{P}$$

其中 $\boldsymbol{A}$ 为整数矩阵, $\boldsymbol{b}, \boldsymbol{c}$ 为整数向量. 若 (P) 的松弛问题 (P_0) 的最优解 $\boldsymbol{x}^0$ 的某个分量 x_i^0 不满足整数要求, 则 (P) 的整数最优解的第 i 个分量必定落在区域 $x_i \leqslant [x_i^0]$ 或 $x_i \geqslant [x_i^0]+1$ 中, 因此可将原问题 (P) 分为两个子问题

$$\begin{cases} \min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b} \\ & \boldsymbol{x} \geqslant \boldsymbol{0},\ \boldsymbol{x}\text{为整数向量} \\ & x_i \leqslant [x_i^0] \end{cases} \tag{P_1}$$

与

$$
\begin{cases}
\min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\
\text{s.t.} & \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b} \\
& \boldsymbol{x} \geqslant \boldsymbol{0},\ \boldsymbol{x}\text{为整数向量} \\
& x_i \geqslant [x_i^0] + 1
\end{cases}
\tag{P_2}
$$

这两个子问题将 (P) 的可行域分成两部分, 且把不满足整数要求的 (P_0) 的最优解 $\boldsymbol{x}^0$ 排除在外. 对于 (P_1) 与 (P_2), 仍是求解相应的松弛问题. 若 (P_1) 的松弛问题的解 $\boldsymbol{x}^1$ 满足整数要求, 则这个子域就查清了, 无需再分支; 若 (P_2) 的松弛问题的解 $\boldsymbol{x}^2$ 不满足整数要求, 则又根据 $\boldsymbol{x}^2$ 的某个不满足整数要求的分量 x_k^2, 按 $x_k \leqslant [x_k^2]$与$x \geqslant [x_k^2] + 1$ 将问题 (P_2) 分解为问题 (P_3) 与 (P_4). 如此继续下去即可.

若求解混合整数规划, 则只需对具有整数要求的分量进行分枝.

二、定界

若仅用分枝过程, 计算量太大, 下面介绍定界过程, 它可以巧妙地删去一些不必要的分枝过程, 使计算量大大减少.

假设在某一时刻, 到当时为止所得到的最好的整数解对应的目标函数值为 z_m, 而且我们正打算由某一点 $\boldsymbol{x}^k$ 分枝, 该子域对应的 ILP 的下界为 $z_k = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^k$. 若 $z_k \geqslant z_m$, 无需由 $\boldsymbol{x}^k$ 继续分枝了, 因为其后代的最优目标值函数不可能优于 z_m.

3 非线性规划

3.1 基本概念

定义: 称如下的数学模型

$$
\begin{cases}
\min & f(\boldsymbol{x}) \\
\text{s.t.} & g_i(\boldsymbol{x}) \leqslant 0, i = 1, 2, \cdots, p \\
& h_j(\boldsymbol{x}) = 0, j = 1, 2, \cdots, q
\end{cases}
$$

为**数学规划** (MP). 当目标函数或至少一个约束条件非线性时, 称为**非线性规划**.

定义: 称满足约束条件的点的集合

$$
X = \left\{ \boldsymbol{x} \in \mathbf{R}^n \;\middle|\; \begin{array}{ll} g_i(\boldsymbol{x}) \leqslant 0, & i = 1, \cdots, p \\ h_j(\boldsymbol{x}) = 0, & j = 1, \cdots, q \end{array} \right\}
$$

为对应数学规划的**可行域**.

定义: 若存在 $\delta > 0$, 对任意的 $t \in (0, \delta)$, 都有

$$f(\overline{\boldsymbol{x}} + t\boldsymbol{p}) < f(\overline{\boldsymbol{x}}),$$

则称向量 $\boldsymbol{p}$ 为函数 f 在 $\overline{\boldsymbol{x}}$ 处的**下降方向**.

定义: 设 $\overline{\boldsymbol{x}} \in X$, 若存在 $t > 0$, 使得

$$\overline{\boldsymbol{x}} + t\boldsymbol{p} \in X,$$

则称向量 $\boldsymbol{p}$ 为点 $\overline{\boldsymbol{x}}$ 处关于 X 的**可行方向**.

若 $\boldsymbol{p}$ 既是可行方向, 又是下降方向, 则称 $\boldsymbol{p}$ 为可行下降方向.

3.2 凸规划

定义: 设 $S \subset \mathbf{R}^n$ 为凸集. 若对任意的 $\boldsymbol{x}^1, \boldsymbol{x}^2 \in S$, 对任意的 $\alpha \in (0,1)$, 都有

$$f(\alpha \boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2) \leqslant \alpha f(\boldsymbol{x}^1) + (1-\alpha) f(\boldsymbol{x}^2),$$

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

定义: 设 $S \subset \mathbf{R}^n$ 为凸集. 若对任意的 $\boldsymbol{x}^1, \boldsymbol{x}^2 \in S, \boldsymbol{x}^1 \neq \boldsymbol{x}^2$, 对任意的 $\alpha \in (0,1)$, 都有

$$f(\alpha \boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2) < \alpha f(\boldsymbol{x}^1) + (1-\alpha) f(\boldsymbol{x}^2),$$

则称函数 $f(\boldsymbol{x})$ 为 (S 上的) **严格凸函数**.

性质: (1) f 为凸函数, $\alpha \geqslant 0$, 则 αf 为凸函数;

(2) f_1, f_2 为凸函数, 则 $f_1 + f_2$ 为凸函数 (但 $f_1 f_2$ 不一定为凸函数).

定理: 若函数 f 是 S 上的凸函数, 则

$$H_s(f, c) = \{\, \boldsymbol{x} \in S \mid f(\boldsymbol{x}) \leqslant c \,\}$$

是凸集.

证明 任取 $\boldsymbol{x}^1, \boldsymbol{x}^2 \in H_s(f,c)$, 则 $\boldsymbol{x}^1, \boldsymbol{x}^2 \in S$ 且 $f(\boldsymbol{x}^1) \leqslant c, f(\boldsymbol{x}^2) \leqslant c$. 对任意的 $\alpha \in (0,1)$, 由 f 是凸集 S 上的凸函数知 $\alpha \boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2 \in S$ 且

$$\begin{aligned} f(\alpha \boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2) &\leqslant \alpha f(\boldsymbol{x}^1) + (1-\alpha) f(\boldsymbol{x}^2) \\ &\leqslant \alpha c + (1-\alpha) c \\ &= c, \end{aligned}$$

故 $\alpha \boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2 \in H_s(f,c)$. 由 $\boldsymbol{x}^1, \boldsymbol{x}^2, \alpha$ 的任意性知 $H_s(f,c)$ 为凸集.

上述定理的逆并不成立 (事实上很容易举出反例). 我们称集合 $H_s(f,c)$ 为函数 f 在集合 S 上关于数 c 的**水平集**.

定义: 称

$$\mathrm{epi}\, f = \{(\boldsymbol{x},c) \in S \times \mathbf{R} \mid f(\boldsymbol{x}) \leqslant c\}$$

为函数 f 的**上方图**.

定理: f 为凸函数当且仅当 f 的上方图 $\mathrm{epi}\, f$ 为凸集.

证明　必要性 $(\Rightarrow)$　任取 $(\boldsymbol{x}^1, c_1), (\boldsymbol{x}^2, c_2) \in \mathrm{epi}\, f$, 则 $\boldsymbol{x}^1, \boldsymbol{x}^2 \in S$ 且 $f(\boldsymbol{x}^1) \leqslant c_1, f(\boldsymbol{x}^2) \leqslant c_2$, 从而由 S 为凸集且 f 为凸函数知对任意的 $\alpha \in (0,1)$, 有 $\alpha\boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2 \in S$ 且

$$\begin{aligned} f(\alpha\boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2) &\leqslant \alpha f(\boldsymbol{x}^1) + (1-\alpha) f(\boldsymbol{x}^2) \\ &\leqslant \alpha c_1 + (1-\alpha) c_2. \end{aligned}$$

由 $(\boldsymbol{x}^1, c_1), (\boldsymbol{x}^2, c_2), \alpha$ 的任意性知 $\mathrm{epi}\, f$ 为凸集.

充分性 $(\Leftarrow)$　任取 $\boldsymbol{x}^1, \boldsymbol{x}^2 \in S$, 则 $(\boldsymbol{x}^1, f(\boldsymbol{x}^1)), (\boldsymbol{x}^2, f(\boldsymbol{x}^2)) \in \mathrm{epi}\, f$. 由 $\mathrm{epi}\, f$ 为凸集知对任意的 $\alpha \in (0,1)$, 有

$$\alpha(\boldsymbol{x}^1, f(\boldsymbol{x}^1)) + (1-\alpha)(\boldsymbol{x}^2, f(\boldsymbol{x}^2)) = (\alpha\boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2, \alpha f(\boldsymbol{x}^1) + (1-\alpha) f(\boldsymbol{x}^2)) \in \mathrm{epi}\, f,$$

从而由 $\mathrm{epi}\, f$ 的定义知

$$f(\alpha\boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2) \leqslant \alpha f(\boldsymbol{x}^1) + (1-\alpha) f(\boldsymbol{x}^2).$$

由 $\boldsymbol{x}^1, \boldsymbol{x}^2, \alpha$ 的任意性知 f 为 S 上的凸函数.

定理: 设 S 为 $\mathbf{R}^n$ 中的凸集, 则

(1) f 为凸函数当且仅当对任意的 $\boldsymbol{x}^1, \boldsymbol{x}^2 \in S$, 都有

$$\nabla f(\boldsymbol{x}^1)^{\mathrm{T}}(\boldsymbol{x}^2 - \boldsymbol{x}^1) \leqslant f(\boldsymbol{x}^2) - f(\boldsymbol{x}^1).$$

(2) f 为严格凸函数当且仅当对任意的 $\boldsymbol{x}^1, \boldsymbol{x}^2 \in S, \boldsymbol{x}^1 \neq \boldsymbol{x}^2$, 都有

$$\nabla f(\boldsymbol{x}^1)^{\mathrm{T}}(\boldsymbol{x}^2 - \boldsymbol{x}^1) < f(\boldsymbol{x}^2) - f(\boldsymbol{x}^1).$$

其中 $\nabla f(\boldsymbol{x})$ 为函数 f 的梯度.

证明　必要性 $(\Rightarrow)$　因 f 为凸函数, 从而对任意的 $\boldsymbol{x}^1, \boldsymbol{x}^2 \in S$, 对任意的 $\alpha \in (0,1)$, 有

$$f(\alpha\boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2) \leqslant \alpha f(\boldsymbol{x}^1) + (1-\alpha) f(\boldsymbol{x}^2)$$

即

$$\frac{f(\boldsymbol{x}^2+\alpha(\boldsymbol{x}^1-\boldsymbol{x}^2))-f(\boldsymbol{x}^2)}{\alpha}\leqslant f(\boldsymbol{x}^1)-f(\boldsymbol{x}^2). \tag{*}$$

由 f 在 $\boldsymbol{x}^2$ 处的 Taylor 展式知

$$f(\boldsymbol{x}^2+\alpha(\boldsymbol{x}^1-\boldsymbol{x}^2))-f(\boldsymbol{x}^2)=\alpha\nabla f(\boldsymbol{x}^2)^{\mathrm{T}}(\boldsymbol{x}^1-\boldsymbol{x}^2)+o(\|\alpha(\boldsymbol{x}^1-\boldsymbol{x}^2)\|)$$

带入 (*) 式得

$$\nabla f(\boldsymbol{x}^2)^{\mathrm{T}}(\boldsymbol{x}^1-\boldsymbol{x}^2)+\frac{o(\|\alpha(\boldsymbol{x}^1-\boldsymbol{x}^2)\|)}{\alpha}\leqslant f(\boldsymbol{x}^1)-f(\boldsymbol{x}^2).$$

令 $\alpha\to 0^+$, 即得

$$\nabla f(\boldsymbol{x}^2)^{\mathrm{T}}(\boldsymbol{x}^1-\boldsymbol{x}^2)\leqslant f(\boldsymbol{x}^1)-f(\boldsymbol{x}^2).$$

充分性 ($\Leftarrow$)　对任意的 $\boldsymbol{x}^1,\boldsymbol{x}^2\in S$, 对任意的 $\alpha\in(0,1)$, 令 $\boldsymbol{x}=\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2$, 由条件知

$$\nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{x}^1-\boldsymbol{x})\leqslant f(\boldsymbol{x}^1)-f(\boldsymbol{x}) \tag{①}$$

及

$$\nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{x}^2-\boldsymbol{x})\leqslant f(\boldsymbol{x}^2)-f(\boldsymbol{x}). \tag{②}$$

将 ① 式乘以 α, ② 式乘以 $(1-\alpha)$, 再将所得不等式相加, 利用 $\boldsymbol{x}=\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2$ 即得

$$f(\alpha\boldsymbol{x}^1+(1-\alpha)\boldsymbol{x}^2)\leqslant \alpha f(\boldsymbol{x}^1)+(1-\alpha)f(\boldsymbol{x}^2).$$

定理: 设 S 为 $\mathbf{R}^n$ 中的凸集, 则

(1) $f(\boldsymbol{x})$ 凸当且仅当 $\nabla^2 f(\boldsymbol{x})$ 半正定;

(2) $\nabla^2 f(\boldsymbol{x})$ 正定 $\Rightarrow f(\boldsymbol{x})$ 严格凸.

其中 $\nabla^2 f(\boldsymbol{x})$ 为函数 f 的 Hessian 矩阵.

例　设二次函数 $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}+c$, 其中 $\boldsymbol{A}$ 为 n 阶实正定矩阵. 则

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

从而

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

因为 $\boldsymbol{A}$ 正定, 故 $f(\boldsymbol{x})$ 为 $\mathbf{R}^n$ 上的凸函数 (且严格凸).

定义: 若可行域 X 是凸集, 目标函数 $f(\boldsymbol{x})$ 是 X 上的凸函数, 则称对应的数学规划为**凸规划**.

定理: 若 $g_i(\boldsymbol{x})\,(i=1,\cdots,p)$ 为 $\mathbf{R}^n$ 上的凸函数, $h_j(\boldsymbol{x})\,(j=1,\cdots,q)$ 为线性函数, $f(\boldsymbol{x})$ 为 X 上的凸函数, 则对应的数学规划为凸规划.

定理: 凸规划的任一局部最优解为整体最优解.

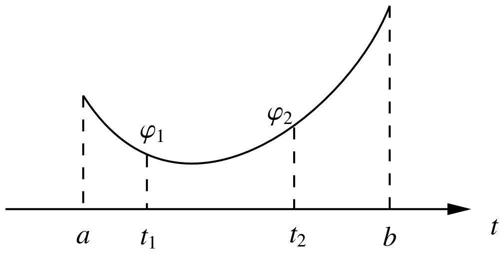
3.3 一维搜索问题

定义: 称 $\min\limits_{t\geqslant 0}\varphi(t)$ 为**一维搜索问题**, $\min\limits_{0\leqslant t\leqslant t_{\max}}\varphi(t)$ 为**有效一维搜索问题**.

一、**0.618 法**

定义: 若在区间 $[a,b]$ 上存在点 t^*, 使得函数 $\varphi(t)$ 在 $[a,t^*]$ 上严格递减, 在 $[t^*,b]$ 上严格递增, 则称 $\varphi(t)$ 在 $[a,b]$ 上是**单谷的**, $[a,b]$ 称为**单谷区间**.

设初始搜索区间为 $[a,b]$, 取 0.618 作为黄金分割比 ω 的近似值. 0.618 法的主要思想就是不断缩小搜索区间, 让后一次与前一次的搜索区间长度之比为 0.618, 从而得到惟一极小点 t^* 的近似值.



在区间 $[a,b]$ 上任取两点 t_1,t_2, 且$t_1 < t_2$. 若 $\varphi(t_1)\leqslant\varphi(t_2)$, 则 $t^*\in[a,t_2]$, 若 $\varphi(t_1)\geqslant\varphi(t_2)$, 则 $t^*\in[t_1,b]$. 为了满足搜索区间长度之比为 0.618 的要求, 我们取初始探索点

$$t_1 = b-\omega(b-a)$$

$$t_2 = a+\omega(b-a)$$

此时分为两种情况:

(1) 若 $\varphi(t_1)\leqslant\varphi(t_2)$, 则 $t^*\in[a,t_2]$, 令 $b\leftarrow t_2,\quad t_2\leftarrow t_1,\quad t_1\leftarrow b-\omega(b-a),\quad \varphi_2\leftarrow\varphi_1$.

(2) 若 $\varphi(t_1)\geqslant\varphi(t_2)$, 则 $t^*\in[t_1,b]$, 令 $a\leftarrow t_1,\quad t_1\leftarrow t_2,\quad t_2\leftarrow a+\omega(b-a),\quad \varphi_1\leftarrow\varphi_2$.

不断重复上述步骤, 当搜索区间达到精度要求后停止即可.

二、Newton **法**

对问题 $\min\varphi(t)$, 设 $\varphi''(t)\neq 0$, 按公式

$$t_{k+1} = t_k - \frac{\varphi'(t_k)}{\varphi''(t_k)}$$

迭代的方法称为 Newton **法**, 它具有局部收敛性. Newton 法具体步骤如下:

(1) 给初始点 t_1 与精度 ε,并令 $k=1$;

(2) 若 $|\varphi'(t_k)|<\varepsilon$, 停止, 输出 t_k. 否则, 若 $\varphi''(t_k)=0$, 停止, 解题失败; 若 $\varphi''(t_k)\neq 0$, 转 (3);

(3) 令 $t_{k+1}=t_k-\dfrac{\varphi'(t_k)}{\varphi''(t_k)}$, 若 $|t_{k+1}-t_k|<\varepsilon$, 停止, 输出 t_{k+1}; 否则, 令 $k\leftarrow k+1$, 转 (2);

3.4 无约束规划

对于无约束规划 (UMP), 有如下一些定理:

定理: 设 $f(\boldsymbol{x})$ 在 $\overline{\boldsymbol{x}}$ 处可微. 若 $\nabla f(\overline{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{p} < 0$, 则 $\boldsymbol{p}$ 为 f 在 $\overline{\boldsymbol{x}}$ 处的下降方向.

证明 由 f 在 $\overline{\boldsymbol{x}}$ 处的泰勒展式知, 对任意的 $t > 0$, 有

$$f(\overline{\boldsymbol{x}} + t\boldsymbol{p}) = f(\overline{\boldsymbol{x}}) + t\nabla f(\overline{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{p} + o(\|t\boldsymbol{p}\|).$$

因为 $\nabla f(\overline{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{p} < 0, t > 0$, 故存在 $\delta > 0$, 当 $t \in (0,\delta)$ 时,

$$t\nabla f(\overline{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{p} + o(\|t\boldsymbol{p}\|) < 0,$$

从而

$$f(\overline{\boldsymbol{x}} + t\boldsymbol{p}) < f(\overline{\boldsymbol{x}}).$$

由下降方向的定义知 $\boldsymbol{p}$ 为 f 在 $\overline{\boldsymbol{x}}$ 处的下降方向.

定理: 设 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}^*$ 处可微. 若 $\boldsymbol{x}^*$ 为局部最优解, 则 $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$.

证明 假设 $\nabla f(\boldsymbol{x}^*) \neq \boldsymbol{0}$, 令 $\boldsymbol{p} = -\nabla f(\boldsymbol{x}^*)$, 则

$$\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{p} < 0.$$

由前一定理知 $\boldsymbol{p}$ 为 f 在 $\boldsymbol{x}^*$ 处的下降方向, 与 $\boldsymbol{x}^*$ 为局部最优解矛盾. 故假设错误, $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$.

定理: 若 $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}, \nabla^2 f(\boldsymbol{x}^*)$ 正定, 则 $\boldsymbol{x}^*$ 为严格局部最优解.

证明 任取 $\boldsymbol{p} \in \mathbf{R}^n, \boldsymbol{p} \neq \boldsymbol{0}$, 由 f 在 $\boldsymbol{x}^*$ 处的 Taylor 展式知对任意的 $t > 0$, 有

$$\begin{aligned} f(\boldsymbol{x}^* + t\boldsymbol{p}) &= f(\boldsymbol{x}^*) + t\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{p} + \frac{1}{2}t^2\boldsymbol{p}^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{p} + o(\|t\boldsymbol{p}\|^2) \\ &= f(\boldsymbol{x}^*) + \frac{1}{2}t^2\boldsymbol{p}^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{p} + o(\|t\boldsymbol{p}\|^2). \end{aligned}$$

由 $\nabla^2 f(\boldsymbol{x}^*)$ 正定及 $\boldsymbol{p} \neq \boldsymbol{0}$ 知 $\boldsymbol{p}^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{p} > 0$, 从而存在 $\delta > 0$, 对任意的 $t \in (0,\delta)$, 有

$$\frac{1}{2}t^2\boldsymbol{p}^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{p} + o(\|t\boldsymbol{p}\|^2) > 0$$

从而

$$f(\boldsymbol{x}^* + t\boldsymbol{p}) > f(\boldsymbol{x}^*).$$

由 $\boldsymbol{p}$ 的任意性知 $\boldsymbol{x}^*$ 为严格局部最优解.

定理: 若 f 为 $\mathbf{R}^n$ 上的可微凸函数, 且 $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$, 则 $\boldsymbol{x}^*$ 为整体最优解.

证明 由 f 是 $\mathbf{R}^n$ 上的凸函数知, 对任意的 $\boldsymbol{x} \in \mathbf{R}^n$, 有

$$0 = \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \leqslant f(\boldsymbol{x}) - f(\boldsymbol{x}^*),$$

故 $\boldsymbol{x}^*$ 为整体最优解.

一、最速下降法

最速下降法是求解无约束规划的一种经典方法, 其他的解析方法或是它的变形, 或是受它的启发而得到, 因此它是最优化方法的基础. 最速下降法具有全局收敛性, 它的具体步骤如下:

(1) 给初始点 $\boldsymbol{x}^0$ 与精度 $\varepsilon, k \leftarrow 0$;

(2) 若 $\|\nabla f(\boldsymbol{x}^k)\| \leqslant \varepsilon$, 停止, 输出 $\boldsymbol{x}^k$;

(3) 令 $\boldsymbol{p}^k = -\nabla f(\boldsymbol{x}^k)$;

(4) 求 t_k, 使得 $f(\boldsymbol{x}^k + t_k\boldsymbol{p}^k) = \min\limits_{t \geqslant 0} f(\boldsymbol{x}^k + t\boldsymbol{p}^k)$, 令 $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + t_k\boldsymbol{p}^k$;

(5) 若 $\|\nabla f(\boldsymbol{x}^{k+1})\| \leqslant \varepsilon$, 停止; 否则, 令 $k \leftarrow k + 1$, 转 (2);

定理: 最速下降法的相邻两次迭代方向彼此正交, 即

$$\nabla f(\boldsymbol{x}^{k+1})^{\mathrm{T}}\,\nabla f(\boldsymbol{x}^k) = 0.$$

证明 设 $\varphi(t) = f(\boldsymbol{x}^k + t\boldsymbol{p}^k)$, 则

$$\varphi'(t_k) = \nabla f(\boldsymbol{x}^k + t_k\boldsymbol{p}^k)^{\mathrm{T}}\boldsymbol{p}^k = -\nabla f(\boldsymbol{x}^{k+1})^{\mathrm{T}}\,\nabla f(\boldsymbol{x}^k) = 0,$$

故最速下降法的相邻两次迭代方向彼此正交.

二、共轭方向法

定义: 设 $\boldsymbol{A}$ 为 n 阶实对称矩阵. 若非零向量 $\boldsymbol{p}, \boldsymbol{q}$ 满足

$$\boldsymbol{p}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{q} = 0,$$

则称 $\boldsymbol{p}, \boldsymbol{q}$ 为 $\boldsymbol{A}$ 的**共轭方向**. 若

$$(\boldsymbol{p}^i)^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^j = 0, \quad i, j = 0, 1, \cdots, n-1, \quad i \neq j,$$

则称 $\boldsymbol{p}^0, \boldsymbol{p}^1, \cdots, \boldsymbol{p}^{n-1}$ 为 $\boldsymbol{A}$ 的一组共轭方向.

正定矩阵的共轭方向组有如下重要性质:

定理: 设 $\boldsymbol{A}$ 为 n 阶实对称正定矩阵, $\boldsymbol{p}^0, \boldsymbol{p}^1, \cdots, \boldsymbol{p}^{n-1}$ 为 $\boldsymbol{A}$ 的一组共轭方向, 则 $\boldsymbol{p}^0, \boldsymbol{p}^1, \cdots, \boldsymbol{p}^{n-1}$ 线性无关.

证明 设

$$\sum_{j=0}^{n-1}\alpha_j\boldsymbol{p}^j=\mathbf{0}.$$

分别用 $\boldsymbol{p}^i, i=0,1,\cdots,n-1$ 左乘上式, 得

$$\sum_{j=0}^{n-1}\alpha_j(\boldsymbol{p}^i)^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^j=0,\quad i=0,1,\cdots,n-1.$$

因为 $\boldsymbol{p}^0,\boldsymbol{p}^1,\cdots,\boldsymbol{p}^{n-1}$ 为 $\boldsymbol{A}$ 的共轭方向, 故

$$(\boldsymbol{p}^i)^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^j=0,\quad i,j=0,1,\cdots,n-1,\quad i\neq j,$$

从而

$$\alpha_j(\boldsymbol{p}^i)^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^i=0,\quad i=0,1,\cdots,n-1.$$

因为 $\boldsymbol{A}$ 正定, $\boldsymbol{p}^i\neq\mathbf{0}$, 故 $(\boldsymbol{p}^i)^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^i>0,\quad i=0,1,\cdots,n-1$, 从而

$$\alpha_i=0,\quad i=0,1,\cdots,n-1,$$

因此 $\boldsymbol{p}^0,\boldsymbol{p}^1,\cdots,\boldsymbol{p}^{n-1}$ 线性无关.

定理: 对二次严格凸函数的无约束规划

$$\min f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}+c \tag{AP}$$

其中 $\boldsymbol{A}$ 为 n 阶实对称正定矩阵, 从任意 $\boldsymbol{x}^0\in\mathbf{R}^n$ 出发, 依次沿共轭方向组 $\boldsymbol{p}^0,\boldsymbol{p}^1,\cdots,\boldsymbol{p}^{n-1}$ 迭代, 至多 n 次可达整体最优解.

定义: 沿共轭方向组进行迭代的方法称为**共轭方向法**.

定义: 若用某种方法求解问题 (AP), 在有限轮内能达到最优解, 则称该方法为**具有二次终止性的方法**.

由前述定理易知, 共轭方向法是具有二次终止性的方法.

利用负梯度向量产生共轭方向的方法称为共轭梯度法. 下面给出构造共轭梯度方向的一般方法. 任取初始点 $\boldsymbol{x}^0\in\mathbf{R}^n$, 若 $\nabla f(\boldsymbol{x}^0)\neq\mathbf{0}$, 则第一个搜索方向取 $\boldsymbol{p}^0=-\nabla f(\boldsymbol{x}^0)$, 设

$$\boldsymbol{p}^1=-\nabla f(\boldsymbol{x}^1)+\lambda_0\boldsymbol{p}^0,$$

上式两边同时左乘 $(\boldsymbol{p}^0)^{\mathrm{T}}\boldsymbol{A}$ 得

$$\lambda_0=\frac{(\boldsymbol{p}^0)^{\mathrm{T}}\boldsymbol{A}\nabla f(\boldsymbol{x}^1)}{(\boldsymbol{p}^0)^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^0}.$$

假设已得共轭方向 $\boldsymbol{p}^0,\boldsymbol{p}^1,\cdots,\boldsymbol{p}^{k-1}$, 设 $\boldsymbol{p}^{k+1}=-\nabla f(\boldsymbol{x}^{k+1})+\lambda_k\boldsymbol{p}^k$, 则

$$\boldsymbol{p}^{k+1}=-\nabla f(\boldsymbol{x}^{k+1})+\sum_{i=0}^{k}\alpha_i\boldsymbol{p}^i.$$

上式两边同时左乘 $(\boldsymbol{p}^k)^{\mathrm{T}}\boldsymbol{A}$ 得

$$\lambda_k=\frac{(\boldsymbol{p}^k)^{\mathrm{T}}\boldsymbol{A}\nabla f(\boldsymbol{x}^{k+1})}{(\boldsymbol{p}^k)^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^k}.$$

即

$$\begin{cases}\boldsymbol{p}^0=-\nabla f(\boldsymbol{x}^0)\\ \boldsymbol{p}^{k+1}=-\nabla f(\boldsymbol{x}^{k+1})+\lambda_k\boldsymbol{p}^k\quad(k=0,1,\cdots,n-2)\\ \lambda_k=\dfrac{(\boldsymbol{p}^k)^{\mathrm{T}}\boldsymbol{A}\nabla f(\boldsymbol{x}^{k+1})}{(\boldsymbol{p}^k)^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^k}\end{cases}$$

可以证明, 上式的 λ_k 可以化简, 从而得到

$$\begin{cases}\boldsymbol{p}^0=-\nabla f(\boldsymbol{x}^0)\\ \boldsymbol{p}^{k+1}=-\nabla f(\boldsymbol{x}^{k+1})+\lambda_k\boldsymbol{p}^k\quad(k=0,1,\cdots,n-2)\\ \lambda_k=\dfrac{\|\nabla f(\boldsymbol{x}^{k+1})\|^2}{\|\nabla f(\boldsymbol{x}^k)\|^2}\end{cases}$$

此公式称为 F-R 公式, 由此公式求解无约束规划的方法称为 Fletcher-Reeves 法, 简称 F-R 法. 带有再开始技巧的 F-R 法步骤如下:

(1) 给初始点 $\boldsymbol{x}^0$, 精度 ε;

(2) 若 $\|\nabla f(\boldsymbol{x}^0)\|\leqslant\varepsilon$, 停止, 输出 $\boldsymbol{x}^0$;

(3) 令 $\boldsymbol{p}^0=-\nabla f(\boldsymbol{x}^0), k\leftarrow 0$;

(4) 求 t_k, 使得 $f(\boldsymbol{x}^k+t_k\boldsymbol{p}^k)=\min\limits_{t\geqslant 0}f(\boldsymbol{x}^k+t\boldsymbol{p}^k)$, 令 $\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+t_k\boldsymbol{p}^k$;

(5) 若 $\|\nabla f(\boldsymbol{x}^{k+1})\|\leqslant\varepsilon$, 停止, 输出 $\boldsymbol{x}^{k+1}$;

(6) 若 $k+1=n$, 令 $\boldsymbol{x}^0\leftarrow\boldsymbol{x}^n$, 转 (3);

(7) 令 $\boldsymbol{p}^{k+1}=-\nabla f(\boldsymbol{x}^{k+1})+\lambda_k\boldsymbol{p}^k$, 其中 $\lambda_k=\dfrac{\|\nabla f(\boldsymbol{x}^{k+1})\|^2}{\|\nabla f(\boldsymbol{x}^k)\|^2}$, 令 $k\leftarrow k+1$, 转 (4).

例 设 $\boldsymbol{A}$ 是 $n\times n$ 对称正定矩阵, 并设 $\boldsymbol{v}^{(i)}(i=0,1,\cdots,n-1)$ 为线性无关的一组向量. 令 $\boldsymbol{p}^{(k)}(k=0,1,\cdots,n-1)$ 按如下方式生成:

$$\begin{aligned}&\boldsymbol{p}^{(0)}=\boldsymbol{v}^{(0)}\\ &\boldsymbol{p}^{(k+1)}=\boldsymbol{v}^{(k+1)}-\sum_{i=0}^{k}\frac{\boldsymbol{v}^{(k+1)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(i)}}{\boldsymbol{p}^{(i)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(i)}}\boldsymbol{p}^{(i)},\quad k=0,1,\cdots,n-2\end{aligned}$$

证明方向 $\boldsymbol{p}^{(k)}(k=0,1,\cdots,n-1)$ 是 $\boldsymbol{A}$ 共轭的. 上述过程称为共轭化, 它从一组线性无关方向出发,

产生一组 $\boldsymbol{A}$ 共轭方向.

证明　我们使用数学归纳法进行证明.

(1) $\boldsymbol{p}^{(0)}, \boldsymbol{p}^{(1)}$ 共轭, 因为

$$\begin{aligned}\boldsymbol{p}^{(0)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(1)} &= \boldsymbol{p}^{(0)\mathrm{T}}\boldsymbol{A}\left[\boldsymbol{v}^{(1)} - \frac{\boldsymbol{v}^{(1)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(0)}}{\boldsymbol{p}^{(0)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(0)}}\boldsymbol{p}^{(0)}\right] \\ &= \boldsymbol{p}^{(0)\mathrm{T}}\boldsymbol{A}\boldsymbol{v}^{(1)} - \boldsymbol{v}^{(1)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(0)} \\ &= 0,\end{aligned}$$

(2) 假设 $\boldsymbol{p}^{(0)}, \boldsymbol{p}^{(1)}, \cdots, \boldsymbol{p}^{(j)}\ (1 \leqslant j \leqslant n-2)$ 共轭, 即

$$\boldsymbol{p}^{(i)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(l)} = 0, \quad i, l = 0, 1, \cdots, j, \quad i \neq l.$$

下面考察向量组 $\boldsymbol{p}^{(0)}, \boldsymbol{p}^{(1)}, \cdots, \boldsymbol{p}^{(j)}, \boldsymbol{p}^{(j+1)}$. 由上述假设, 我们只需验证 $\boldsymbol{p}^{(i)}, i = 0, 1, \cdots, j$ 分别与 $\boldsymbol{p}^{(j+1)}$ 共轭即可. 直接计算, 我们有

$$\begin{aligned}\boldsymbol{p}^{(i)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(j+1)} &= \boldsymbol{p}^{(i)\mathrm{T}}\boldsymbol{A}\left[\boldsymbol{v}^{(j+1)} - \sum_{l=0}^{j}\frac{\boldsymbol{v}^{(j+1)T}\boldsymbol{A}\boldsymbol{p}^{(l)}}{\boldsymbol{p}^{(l)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(l)}}\boldsymbol{p}^{(l)}\right] \\ &= \boldsymbol{p}^{(i)\mathrm{T}}\boldsymbol{A}\boldsymbol{v}^{(j+1)} - \sum_{l=0}^{j}\frac{\boldsymbol{v}^{(j+1)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(l)}}{\boldsymbol{p}^{(l)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(l)}}\boldsymbol{p}^{(i)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(l)} \\ &= \boldsymbol{p}^{(i)\mathrm{T}}\boldsymbol{A}\boldsymbol{v}^{(j+1)} - \boldsymbol{v}^{(j+1)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(i)} - \sum_{\substack{l=0\\ l\neq i}}^{j}\frac{\boldsymbol{v}^{(j+1)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(l)}}{\boldsymbol{p}^{(l)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(l)}}\boldsymbol{p}^{(i)\mathrm{T}}\boldsymbol{A}\boldsymbol{p}^{(l)} \\ &= 0, \quad i = 0, 1, \cdots, j,\end{aligned}$$

因此 $\boldsymbol{p}^{(0)}, \boldsymbol{p}^{(1)}, \cdots, \boldsymbol{p}^{(j)}, \boldsymbol{p}^{(j+1)}$ 共轭.

由数学归纳法知 $\boldsymbol{p}^{(0)}, \boldsymbol{p}^{(1)}, \cdots, \boldsymbol{p}^{(n-1)}$ 共轭.

3.5　约束最优化方法

设约束非线性规划

$$\begin{cases}\min & f(\boldsymbol{x}) \\ \text{s.t.} & g_i(\boldsymbol{x}) \leqslant 0, i = 1, 2, \cdots, p \\ & h_j(\boldsymbol{x}) = 0, j = 1, 2, \cdots, q\end{cases}$$

记 $I = \{1, 2, \cdots, p\}, J = \{1, 2, \cdots, q\}$.

定义: 称 $I(\boldsymbol{x}^*) = \{i \in I \mid g_i(\boldsymbol{x}^*) = 0\}$ 为 $\boldsymbol{x}^*$ 的**积极指标集**.

定理: 设 $\boldsymbol{x}^*$ 为 (MP) 的局部最优解, $\nabla g_i(\boldsymbol{x}^*), i \in I(\boldsymbol{x}^*)$ 与 $\nabla h_j(\boldsymbol{x}^*)$, $j \in J$ 线性无关, 则存在两组实

数 $\lambda_i^*, i\in I(\boldsymbol{x}^*)$与 $\mu_j^*, j\in J$, 使得

$$\begin{cases} \nabla f(\boldsymbol{x}^*)+\sum\limits_{i\in I(x^*)}\lambda_i^*\nabla g_i(\boldsymbol{x}^*)+\sum\limits_{j\in J}\mu_j^*\nabla h_j(\boldsymbol{x}^*)=\mathbf{0} \\ \lambda_i^*\geqslant 0,\quad i\in I(x^*) \end{cases} \tag{*}$$

定义: $\nabla g_i(\boldsymbol{x}^*), i\in I(\boldsymbol{x}^*)$ 与 $\nabla h_j(\boldsymbol{x}^*)$, $j\in J$ 线性无关这一条件称为**约束规范条件**; 条件 $(*)$ 称为 Kuhn-Tucker 条件, 简称 K-T **条件**. 满足 K-T 条件的可行点称为 K-T **点**.

若各 $g_i(\boldsymbol{x}^*)$ 在 $\boldsymbol{x}^*$ 处均可微, 则 K-T 条件可改写为

$$\begin{cases} \nabla f(\boldsymbol{x}^*)+\sum\limits_{i\in I}\lambda_i^*\nabla g_i(\boldsymbol{x}^*)+\sum\limits_{j\in J}\mu_j^*\nabla h_j(\boldsymbol{x}^*)=\mathbf{0} \\ \lambda_i^* g_i(\boldsymbol{x}^*)=0,\quad i\in I \\ \lambda_i^*\geqslant 0,\quad i\in I \end{cases}$$

其中第二个条件 $\lambda_i^* g_i(\boldsymbol{x}^*)=0, i\in I$ 称为**互补松紧条件**.

定义 (MP) 的 Lagrange 函数为

$$L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})=f(\boldsymbol{x})+\sum_{i\in I}\lambda_i g_i(\boldsymbol{x})+\sum_{j\in J}\mu_j h_j(\boldsymbol{x}),$$

其中 $\boldsymbol{\lambda}=(\lambda_1,\cdots,\lambda_p)^{\mathrm{T}}, \boldsymbol{\mu}=(\mu_1,\cdots,\mu_q)^{\mathrm{T}}$ 称为 Lagrange 乘子.

利用 Lagrange 函数, K-T 条件可改写为

$$\begin{cases} \nabla_x L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)=\mathbf{0} \\ \lambda_i^* g_i(\boldsymbol{x}^*)=0,\quad i\in I \\ \lambda_i^*\geqslant 0,\quad i\in I \end{cases}$$

定理: 若可行点 $\boldsymbol{x}^*$ 为 K-T 点, f 与 $g_i, i\in I(\boldsymbol{x}^*)$ 为凸函数, $h_j, j\in J$ 为线性函数, 则 $\boldsymbol{x}^*$ 为整体最优解.