基本概念

最优化问题的数学模型的一般形式 (优化模型):(第七章又重新定义了一遍, 只是符号有所不同, 这里定义为极小化问题, 极大化问题可以转化为极小化问题)

$$\begin{cases} \min\ \ f(x) \\ \ \text{s.t.}\ \ c_i(x)=0,\ i=1,2,\cdots,m, \\ \qquad\ c_i(x)\geqslant 0,\ i=m+1,\cdots,p, \end{cases}$$

其中 $x=(x_1,x_2,\cdots,x_n)^T\in\mathbb{R}^n$, $f:\mathbb{R}^n\to\mathbb{R}^1$, $c_i=\mathbb{R}^n\to\mathbb{R}^1(i=1,2,\cdots,p)$ 为连续函数, 通常还要求连续可微. x 称为**决策变量**, $f(x)$ 为**目标函数**, $c_i(x),i=1,2,\cdots,p$ 为**约束函数 (约束条件)**, $c_i(x)=0,\ i=1,2,\cdots,m$ 为**等式约束**, $c_i(x)\geqslant 0,\ i=m+1,\cdots,p$ 为**不等式约束**, 并记等式约束指标集为 $E=\{1,2,\cdots,m\}$, 不等式约束指标集为 $I=\{m+1,\cdots,p\}$.

可行点: 若点 $x\in\mathbb{R}^n$ 满足优化模型中的所有约束条件, 则称 x 为**可行点**.

可行域: 全体可行点所成之集称为**可行域**, 即

$$\mathcal{F}=\{x:c_i(x)=0,\ i=1,2,\cdots,m,\ c_i(x)\geqslant 0,i=m+1,\cdots,p\}.$$

有效约束 (起作用约束): 可行点 $\bar{x}\in\mathcal{F}$, 若 $c_i(\bar{x})=0$, 则称不等式约束 $c_i(x)\geqslant 0$ 在点 $\bar{x}$ 是**有效约束**, 并称可行点 $\bar{x}$ 位于约束 $c_i(x)\geqslant 0$ 的边界.

有效约束集: 记 $I(x)=\{i:c_i(x)=0,\ i\in I\}$, 对任何 $x\in\mathbb{R}^n$, 我们称集合

$$A(x)=E\cup I(x)$$

是在 x 点处的**有效约束指标集 (或积极约束指标集)**, 简称**有效约束集或有效集**.

可行方向: 设 $x^*\in\mathcal{F}, 0\neq d\in\mathbb{R}^n$, 若存在 $\delta>0$ 使得

$$x^*+\alpha d\in\mathcal{F},\quad \forall\alpha\in[0,\delta],$$

则称 d 是 $\mathcal{F}$ 在 x^* 处的**可行方向**. $\mathcal{F}$ 在 x^* 处的全体可行方向所成之集记为 $\mathcal{FD}(x^*,\mathcal{F})$.

下降方向: 设 $f(x)$ 为 $\mathbb{R}^n$ 上的连续函数, 点 $\bar{x}\in\mathbb{R}^n$, 若对于方向 $s\in\mathbb{R}^n$ 存在 $\delta>0$ 使得下式成立

$$f(\bar{x}+\alpha s)<f(\bar{x}),\quad \forall\alpha\in(0,\delta),$$

则称 s 为 $f(x)$ 在 $\bar{x}$ 处的一个**下降方向**, 将点 $\bar{x}$ 处的所有下降方向所成之集记为 $\mathcal{D}(\bar{x})$.

无效约束 (不起作用约束): 可行点 $\bar{x}\in\mathcal{F}$, 若 $c_i(\bar{x})>0$, 则称不等式约束 $c_i(x)\geqslant 0$ 在点 $\bar{x}$ 是**无效约束**, 并称可行点 $\bar{x}$ 位于约束 $c_i(x)\geqslant 0$ 的内部.

全局 (或总体) 最优解 (极小点): 可行点 $x^*\in\mathcal{F}$ 称为优化模型的**全局最优解**, 当且仅当

$$f(x^*)\leqslant f(x),\quad \forall x\in\mathcal{F}.$$

严格全局 (或总体) 最优解 (极小点): 可行点 $x^* \in \mathcal{F}$ 称为优化模型的**严格全局最优解**, 当且仅当

$$f(x^*) < f(x), \quad \forall x \in \mathcal{F},\ x \neq x^*.$$

局部最优解 (极小点): 可行点 $x^* \in \mathcal{F}$ 称为优化模型的**局部最优解**, 当且仅当, 存在 x^* 的一个邻域

$$\mathcal{N}(x^*) = \{x : ||x - x^*|| \leqslant \delta\}$$

使得下式成立

$$f(x^*) \leqslant f(x), \quad \forall x \in \mathcal{N}(x^*) \cap \mathcal{F}.$$

同理也可以定义严格局部最优解.

严格局部最优解 (极小点): 可行点 $x^* \in \mathcal{F}$ 称为优化模型的**严格局部最优解**, 当且仅当, 存在 x^* 的一个邻域

$$\mathcal{N}(x^*) = \{x : ||x - x^*|| \leqslant \delta\}$$

使得下式成立

$$f(x^*) < f(x), \quad \forall x \in \mathcal{N}(x^*) \cap \mathcal{F},\ x \neq x^*.$$

凸集: 集合 $D \subset \mathbb{R}^n$ 称为**凸集**, 当且仅当, 对于任意 $x, y \in D$ 有

$$\lambda x + (1-\lambda)y \in D, \quad \forall \lambda \in [0,1].$$

即连接 x, y 的直线段上的所有点均在集合 D 内.

凸函数: 设函数 $f(x)$ 在凸集 D 上有定义, 如果对于任意 $x, y \in D$ 有

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y), \quad \forall \lambda \in [0,1],$$

则称 $f(x)$ 是凸集 D 上的**凸函数**.

严格凸函数: 设函数 $f(x)$ 在凸集 D 上有定义, 如果对于任意 $x, y \in D,\ x \neq y$ 有

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y), \quad \forall \lambda \in (0,1),$$

则称 $f(x)$ 是凸集 D 上的**严格凸函数**.

凸规划问题: 可行域 $\mathcal{F}$ 是凸集, 目标函数 $f(x)$ 是凸函数的最优化问题称为**凸规划问题**. (可以证明, 凸规划问题的局部最优解必是全局最优解)

判断凸函数方法: 设 $f(x)$ 为非空开凸集 $D \subset \mathbb{R}^n$ 上的二阶可微函数, 则

(1) $f(x)$ 的 Hesse 矩阵 $\nabla^2 f(x)$ 在 D 上**半正定** (所有主子式均 $\geqslant 0$) $\Longleftrightarrow$ $f(x)$ 是 D 上的**凸函数**.

(2) $f(x)$ 的 Hesse 矩阵 $\nabla^2 f(x)$ 在 D 上**正定** (顺序主子式均 > 0)$\Rightarrow$$f(x)$ 是 D 上的**严格凸函数**.

最优性条件: 最优化问题的最优解 (局部的或者全局的) 所必须满足的条件.

KKT 点: 利用迭代的方法产生一个逐步改善的序列 $\{x^{(k)}\}$, 在 $\{x^{(k)}\}$ 为**有限**点列时, 它最后一个点是 **KKT 点**; 当 $\{x^{(k)}\}$ 是**无限点列**时, 其任意一个聚点为 **KKT 点**.

最优化方法的基本迭代格式:

1. 给出初始点 $x^{(0)}$, 令 $k := 0$; (初始化)

2. 如果 $x^{(k)}$ 满足对最优解估计的终止条件, 停止迭代; (结束条件)

3. 确定一个改善 $x^{(k)}$ 的修正量 $s^{(k)}$; (计算)

4. 令 $x^{(k+1)} = x^{(k)} + s^{(k)},\ k := k + 1$, 转到第 2 步. (更新)

算法 (方法) 的收敛性: 一个算法是收敛的, 当且仅当, 算法产生的序列 $\{x^{(k)}\}$ 满足

$$\lim_{k\to\infty} ||x^{(k)} - x^*|| = 0,$$

其中 x^* 为该问题的 KKT 点, 即该序列 $\{x^{(k)}\}$ 收敛.

全局收敛 (总体收敛): 如果一个算法对于任意给定的初始点都能能够收敛, 则称该算法**全局收敛**.

局部收敛: 如果一个算法只有当初始点接近或充分接近最优解时才具有收敛性, 则称该算法**局部收敛**.

收敛速度: 设向量序列 $\{x^{(k)}\} \subset \mathbb{R}^n$ 收敛于 x^*, 定义误差序列

$$e_k = x^{(k)} - x^*.$$

若存在正常数 C 和 r 使得下式成立

$$\lim_{k\to\infty} \frac{||x^{(k+1)} - x^*||}{||x^{(k)} - x^*||^r} = \lim_{k\to\infty} \frac{||e_{k+1}||}{||e_k||^r} = C,$$

则称序列 $\{x^{(k)}\}$ $\boldsymbol{r}$ **阶**收敛于 x^*(以 C 为因子, 有时也称 C 为**收敛速度的值**).

当 $r = 1,\ 0 < C < 1$ 时称为**线性收敛**, $r = 2$ 时称为**二次收敛**, $r > 1$ 时称为**超线性收敛**.

精确线性搜索的无约束最优化算法的一般形式:

1. 给出初始点 $x_0 \in \mathbb{R}^n, \varepsilon \geqslant 0$, 令 $k := 0$.

2. 计算 $\nabla f(x_k)$, 若 $||\nabla f(x_k)|| \leqslant \varepsilon$, 停止迭代.

3. 计算下降方向 d_k, 计算步长因子 α_k, 使得

$$f(x_k + \alpha_k d_k) = \min_{\alpha\geqslant 0} f(x_k + \alpha d_k).$$

4. 令 $x_{k+1} = x_k + \alpha_k d_k, k := k + 1$, 转到第 2 步.

5 种无约束最优化方法

最速下降法: 设目标函数 $f(x)$ 在 x_k 附近连续可微, 且 $g_k := \nabla f(x_k) \neq 0$.

1. 给出初始点 $x_0 \in \mathbb{R}^n, \varepsilon \geqslant 0$, 令 $k := 0$.

2. 计算 $d_k = -g_k$; 若 $||g_k|| \leqslant \varepsilon$, 停止迭代.

3. 计算步长因子 α_k, 使得

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geqslant 0} f(x_k + \alpha d_k).$$

4. 令 $x_{k+1} = x_k + \alpha_k d_k, k = k + 1$, 转到第 2 步.

带步长因子的牛顿法: 设 $f(x)$ 二次连续可微, $x_k \in \mathbb{R}^n$, 令 $G_k := \nabla^2 f(x_k),\ g_k := \nabla f(x_k)$.

1. 给出初始点 $x_0, \varepsilon \geqslant 0$, 令 $k := 0$.

2. 计算 g_k. 若 $||g_k|| \leqslant \varepsilon$, 停止迭代.

3. 解方程组 $G_k d = -g_k$ 得 d_k 为牛顿方向, 计算步长因子 α_k, 使得

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geqslant 0} f(x_k + \alpha d_k).$$

4. 令 $x_{k+1} = x_k + \alpha_k d_k, k := k + 1$, 转到第 2 步.

牛顿法具有 **2 阶**收敛速度.

向量的共轭: 设 G 是 $n \times n$ 对称正定矩阵, d_1, d_2 为 n 为非零向量, 若

$$d_1^T G d_2 = 0,$$

则称向量 d_1 和 d_2 是 G-共轭的, 简称**共轭的**.

$d_1, d_2, \cdots, d_m$ 为一组 n 为非零向量, 若

$$d_i^T G d_j = 0, \quad (i \neq j),$$

则称向量 $d_1, d_2, \cdots, d_m$ 是 G-共轭的, 简称**共轭的**.

一般共轭方向法:

1. 给出初始点 $x_0, \varepsilon \geqslant 0$, 令 $k := 0$.

   计算 $g_0 = \nabla f(x_0)$ 和初始下降方向 d_0, 使得 $d_0^T g_0 < 0$.

2. 计算 g_k. 如果 $||g_k|| \leqslant \varepsilon$, 停止迭代.

3. 计算步长因子 α_k, 使得

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geqslant 0} f(x_k + \alpha d_k).$$

4. 采用某种共轭方向法计算 d_{k+1} 使得

$$d_{k+1}^T G d_j = 0,\ j = 0, 1, \cdots, k.$$

令 $x_{k+1} = x_k + \alpha_k d_k, k := k+1$, 转到第 2 步.

二次终止性: 对于正定二次函数, 算法是有限步终止的. (可以证明若搜索方向是相互共轭的, 则算法具有二次终止性, **共轭梯度法和拟牛顿法**都就具有**二次终止性**)

拟牛顿法: B_k 为牛顿法迭代中的 Hesse 矩阵 G_k 的近似值, 通常取 $B_0 = E_n$ 单位阵.

1. 给出初始点 $x_0 \in \mathbb{R}^n$, $B_0 \in \mathbb{R}^{n\times n}$, $\varepsilon \geqslant 0$, $k := 0$.

2. 计算 $g_k = \nabla f(x_k)$. 如果 $||g_k|| \leqslant \varepsilon$, 停止迭代.

3. 解方程组 $B_k d = -g_k$ 得到搜索方向 d_k, 计算步长因子 α_k, 使得

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geqslant 0} f(x_k + \alpha d_k).$$

4. 令 $x_{k+1} = x_k + \alpha_k d_k$. 校正 B_k 产生 B_{k+1}, 使得拟牛顿条件

$$B_k(x_{k+1} - x_k) = g_{k+1} - g_k,$$

成立, 其中 $g_{k+1} = \nabla f(x_{k+1})$. 令 $k := k+1$, 转到第 2 步.

最小二乘问题:

$$\min_{x\in\mathbb{R}^n} \quad f(x) = \frac{1}{2}||r(x)||_2^2 = \frac{1}{2}\sum_{i=1}^{m}[r_i(x)]^2, \quad m \geqslant n. \tag{1}$$

其中 $r(x) = (r_1(x), r_2(x), \cdots, r_m(x))^T$, $r_i(x)$ $(i = 1, 2, \cdots, m)$ 称为**残量函数**, 有无 $\frac{1}{2}$ 对最优解没有影响. 当 $r_i(x)$ 为线性函数时, 称 (1) 为**线性最小二乘问题**; 当 $r_i(x)$ 为非线性函数时, 称 (1) 为**非线性最小二乘问题**.

Gauss-Newton 法求解非线性最小二乘问题: 记向量函数 $r(x)$ 的 $m \times n$ 阶 Jacobian 矩阵为

$$A(x) = \nabla r(x) = [\nabla r_1(x), \cdots, \nabla r_m(x)]^T = \begin{bmatrix} \frac{\partial r_1}{\partial x_1}(x) & \frac{\partial r_1}{\partial x_2}(x) & \cdots & \frac{\partial r_1}{\partial x_n}(x) \\ \frac{\partial r_2}{\partial x_1}(x) & \frac{\partial r_2}{\partial x_2}(x) & \cdots & \frac{\partial r_2}{\partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r_m}{\partial x_1}(x) & \frac{\partial r_m}{\partial x_2}(x) & \cdots & \frac{\partial r_m}{\partial x_n}(x) \end{bmatrix}$$

记 $A_k = A(x_k),\ r_k = r(x_k)$, 则 Gauss-Newton 法如下

1. 给定初始值 x_k, $\varepsilon \geqslant 0$, 令 $k := 0$.

2. 计算 $g_k = \nabla f(x_k)$. 若 $||g_k|| \leqslant \varepsilon$, 停止迭代.

3. 解方程组 $A_k^T A_k \delta = -A_k^T r_k$ 得到搜索方向 δ_k.

4. 令 $x_{k+1} = x_k + \delta_k$, $k := k+1$, 转到第 2 步.

KKT 条件

KKT 定理: 设 $f(x), c_i(x)$ $(i = 1, 2, \cdots, p)$ 在 x^* 的邻域内一阶连续可微, 约束规范条件 (应该不是重点)

$$\mathcal{SFD}(x^*, \mathcal{F}) = \mathcal{LFD}(x^*, \mathcal{F})$$

成立, 则存在 λ_i^* $(i = 1, 2, \cdots, m)$ 使得 (下面五个条件就是 **KKT 条件**)

$$\begin{aligned}
&\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*), && \text{驻点条件} \qquad (2)\\
&c_i(x^*) = 0, \quad i \in E, && \text{等式约束条件}\\
&c_i(x^*) \geqslant 0, i \in I, && \text{不等式约束条件}\\
&\lambda_i^* \geqslant 0, \quad i \in I, && \text{非负乘子条件}\\
&\lambda_i^* c_i(x^*) = 0, \quad i \in I. && \text{互补松弛条件}
\end{aligned}$$

若记 **Lagrange 函数** $L(x, \lambda^*) = f(x) - \sum_{i=1}^{m} \lambda_i^* c_i(x)$, 则 KKT 条件中第一个条件 (即 (2) 式) 可以视为

$$\nabla_x L(x^*, \lambda^*) = \frac{\partial L}{\partial x}(x^*) = \begin{bmatrix} \frac{\partial L}{\partial x_1}(x^*) & \cdots & \frac{\partial L}{\partial x_n}(x^*) \end{bmatrix}^T = \mathbf{0}$$

注意: 使用上述 KKT 条件, 必须满足为极小化问题, 且**不等式约束为** $\geqslant 0$ **的形式**.

二次罚函数和内点障碍法

只需会构造即可.

二次罚函数: 关于优化模型我们定义二次罚函数 $Q(x; \mu)$ 如下

$$Q(x : \mu) := f(x) + \frac{1}{2\mu} \sum_{i \in E} c_i^2(x) + \frac{1}{2\mu} \sum_{i \in I} ([c_i(x)]^-)^2$$

其中 $[c_i(x)]^- := \min\{c_i(x), 0\}, (i \in I)$, $\frac{1}{2\mu}$ 称为罚系数.

当罚系数 $\to \infty$ 时, $Q(x; \mu)$ 的极小值趋近优化模型的极小值.

内点障碍法: 考虑不等式约束最优化问题

$$\begin{cases} \min \ f(x) \\ \text{s.t. } c_i(x) \geqslant 0, \ i = I. \end{cases}$$

两种常用的障碍函数:

对数障碍函数:

$$P(x; \mu) = f(x) - \mu \sum_{i \in I} \log c_i(x).$$

分数障碍函数:

$$P(x;\mu) = f(x) - \mu \sum_{i \in I} \frac{1}{[c_i(x)]^+}.$$

其中 $[c_i(x)]^+ = \max\{0, c_i(x)\}$.

当 $\mu \to 0$ 时, $P(x;\mu)$ 的极小值趋近优化模型的极小值.

内点障碍法的优势在于, $P(x,\mu)$ 每一步所求出的的极小值, 一定符合优化模型的所有不等式约束条件, 而罚函数方法并不能保证.