

第四章 约束优化算法

修贤超

<https://xianchaoxiu.github.io>

■ 4.1 罚函数法

■ 4.2 增广拉格朗日函数法

约束优化问题

■ 考虑约束优化问题

$$\begin{array}{ll}\min_{x \in \mathbb{R}^n} & f(x) \\ \text{s.t.} & x \in \mathcal{X}\end{array}$$

■ 相比于无约束问题的困难

- x 不能随便取值, 梯度下降法所得点不一定在可行域内
- 最优解处目标函数的梯度不一定为零向量

■ 将约束优化问题转化为无约束优化问题处理

- 罚函数法
- 增广拉格朗日函数法

等式约束的二次罚函数法

- 考虑仅包含等式约束的约束优化问题

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & c_i(x) = 0, \quad i \in \mathcal{E} \end{aligned}$$

- **定义** 定义二次罚函数为

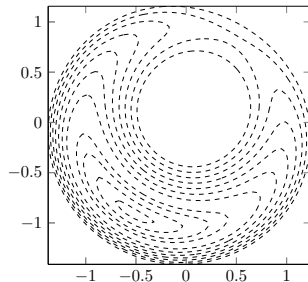
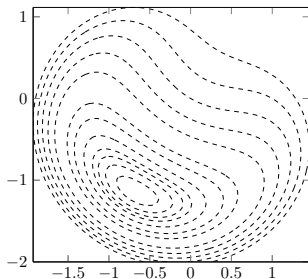
$$P_E(x, \sigma) = f(x) + \frac{1}{2}\sigma \sum_{i \in \mathcal{E}} c_i^2(x)$$

其中等式右端第二项称为**罚函数**, $\sigma > 0$ 称为**罚因子**

- 对不满足约束的点进行惩罚, 被称为**外点罚函数**

■ 考虑优化问题

$$\begin{array}{ll}\min & x + \sqrt{3}y \\ \text{s.t.} & x^2 + y^2 = 1\end{array}$$

■ 容易求得最优解为 $(-\frac{1}{2}, -\frac{\sqrt{3}}{2})^\top$ ■ 考虑二次罚函数 $P_E(x, y, \sigma) = x + \sqrt{3}y + \frac{\sigma}{2}(x^2 + y^2 - 1)^2$ 

- 考虑优化问题

$$\begin{array}{ll}\min & -x^2 + 2y^2 \\ \text{s.t.} & x = 1\end{array}$$

- 容易求得最优解为 $(1, 0)^\top$, 然而考虑罚函数

$$P_E(x, y, \sigma) = -x^2 + 2y^2 + \frac{\sigma}{2}(x - 1)^2$$

- 对任意的 $\sigma \leq 2$, 罚函数无下界

二次罚函数法

算法 二次罚函数法

- 1 给定 $\sigma_1 > 0, x_0, k \leftarrow 1$. 罚因子增长系数 $\rho > 1$
- 2 **while** 未达到收敛准则 **do**
- 3 以 x^k 为初始点, 求解 $x^{k+1} = \arg \min_x P_E(x, \sigma_k)$
- 4 选取 $\sigma_{k+1} = \rho \sigma_k$
- 5 $k \leftarrow k + 1$
- 6 **end while**

=====

- σ_k 增长过快会使子问题求解困难, σ_k 增长过慢则会增加迭代次数
- 检测到迭代点发散就应该立即终止迭代并增大罚因子
- 为保证收敛, 子问题求解误差需要趋于零

分析 KKT 条件

■ 原问题的 KKT 条件

$$\begin{aligned}\nabla f(x^*) - \sum_{i \in \mathcal{E}} \lambda_i^* \nabla c_i(x^*) &= 0 \\ c_i(x^*) &= 0, \quad \forall i \in \mathcal{E}\end{aligned}$$

■ 添加罚函数项问题的 KKT 条件

$$\nabla f(x) + \sum_{i \in \mathcal{E}} \sigma c_i(x) \nabla c_i(x) = 0$$

■ 假设两个问题收敛到同一点, 对比 KKT 条件式成立

$$\sigma c_i(x) \approx -\lambda_i^*, \quad \forall i \in \mathcal{E}$$

■ 为使约束 $c_i(x) = 0$ 成立, 需要 $\sigma \rightarrow \infty$

- 考虑罚函数 $P_E(x, \sigma)$ 的海瑟矩阵

$$\nabla_{xx}^2 P_E(x, \sigma) = \nabla^2 f(x) + \sum_{i \in \mathcal{E}} \sigma c_i(x) \nabla^2 c_i(x) + \sigma \nabla c(x) \nabla c(x)^\top$$

\Downarrow

$$\nabla_{xx}^2 P_E(x, \sigma) \approx \nabla_{xx}^2 L(x, \lambda^*) + \sigma \nabla c(x) \nabla c(x)^\top$$

- $\nabla_{xx}^2 P_E(x, \sigma)$ 条件数越来越大, 子问题的难度也会相应地增加
- 在实际应用中, 不可能令罚因子趋于正无穷

收敛性分析

■ **定理** 设 x^{k+1} 是 $P_E(x, \sigma_k)$ 的全局极小解, σ_k 单调上升趋于无穷, 则 x^k 的每个极限点 x^* 都是原问题的全局极小解

证明 设 \bar{x} 为原问题的极小解. 由 x^{k+1} 为 $P_E(x, \sigma_k)$ 的极小解, 得 $P_E(x^{k+1}, \sigma_k) \leq P_E(\bar{x}, \sigma_k)$, 即

$$f(x^{k+1}) + \frac{\sigma_k}{2} \sum_{i \in \mathcal{E}} c_i^2(x^{k+1}) \leq f(\bar{x}) + \frac{\sigma_k}{2} \sum_{i \in \mathcal{E}} c_i^2(\bar{x}) = f(\bar{x})$$

\Downarrow

$$\sum_{i \in \mathcal{E}} c_i^2(x^{k+1}) \leq \frac{2}{\sigma_k} (f(\bar{x}) - f(x^{k+1}))$$

设 x^* 是 x^k 的一个极限点, 令 $k \rightarrow \infty$, 得 $\sum_{i \in \mathcal{E}} c_i^2(x^*) = 0$. 易知 x^* 为原问题的可行解, 又 $f(x^{k+1}) \leq f(\bar{x})$, 取极限得 $f(x^*) \leq f(\bar{x})$, 故 x^* 为全局极小解

收敛性分析

- **定理** 设 $f(x)$ 与 $c_i(x)$ ($i \in \mathcal{E}$) 连续可微, 正数序列 $\varepsilon_k \rightarrow 0$, $\sigma_k \rightarrow +\infty$. 子问题的解 x^{k+1} 满足

$$\|\nabla_x P_E(x^{k+1}, \sigma_k)\| \leq \varepsilon_k$$

而对 x^k 的任何极限点 x^* , 都有 $\{\nabla c_i(x^*), i \in \mathcal{E}\}$ 线性无关, 则 x^* 是等式约束最优化问题的 KKT 点, 且

$$\lim_{k \rightarrow \infty} (-\sigma_k c_i(x^{k+1})) = \lambda_i^*, \quad \forall i \in \mathcal{E}$$

其中 λ_i^* 是约束 $c_i(x^*) = 0$ 对应的拉格朗日乘子

- 精确求解 \Rightarrow 精度需要越来越高

一般约束问题的二次罚函数法

■ 考虑一般约束问题

$$\begin{array}{ll}\min & f(x) \\ \text{s.t.} & c_i(x) = 0, \quad i \in \mathcal{E} \\ & c_i(x) \leq 0, \quad i \in \mathcal{I}\end{array}$$

■ 定义二次罚函数

$$P(x, \sigma) = f(x) + \frac{1}{2}\sigma \left[\sum_{i \in \mathcal{E}} c_i^2(x) + \sum_{i \in \mathcal{I}} \tilde{c}_i^2(x) \right]$$

其中 $\tilde{c}_i(x) = \max\{c_i(x), 0\}$

二次罚函数法的优缺点

■ 优点

- 将约束优化问题转化为无约束优化问题
- 二次罚函数形式简洁直观广泛使用

■ 缺点

- 需要 $\sigma \rightarrow \infty$, 导致海瑟矩阵条件数过大
- 对于不等式约束的问题可能不存在二次可微性质, 光滑性降低
- 不精确, 与原问题最优解存在距离

应用举例: LASSO 问题

- 考虑 LASSO 问题

$$\min_x \quad \frac{1}{2} \|Ax - b\|^2 + \mu \|x\|_1$$

以及基追踪 (BP) 问题

$$\begin{aligned} \min \quad & \|x\|_1 \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

- 写成二次罚函数法形式

$$\min_x \quad \|x\|_1 + \frac{\sigma}{2} \|Ax - b\|^2$$

- 仅在 μ 趋于 0 时, LASSO 问题的解收敛于 BP 问题的解
- 当 μ 较小时问题病态, 收敛较慢, 可逐渐缩小 μ 的值求解子问题逼近

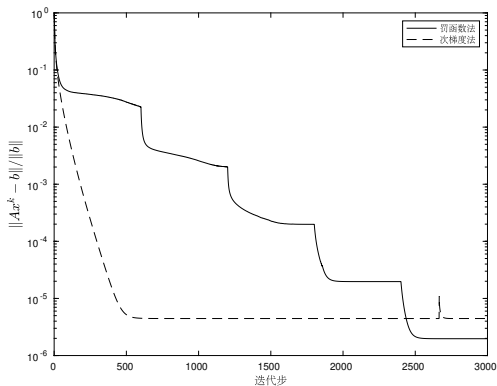
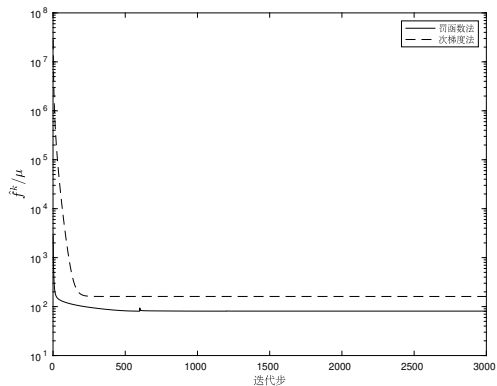
LASSO 问题罚函数法算法

```
1 给定初值  $x_0$ , 最终参数  $\mu$ , 初始参数  $\mu_0$ , 因子  $\gamma \in (0, 1), k \leftarrow 0$   
2 while  $\mu_k \geq \mu$  do  
3 以  $x^k$  为初值, 求解问题  $x^{k+1} = \arg \min_x \{ \frac{1}{2} \|Ax - b\|^2 + \mu_k \|x\|_1 \}$   
4 if  $\mu_k = \mu$  then  
5 停止迭代, 输出  $x^{k+1}$   
6 else  
7 更新罚因子  $\mu_{k+1} = \max\{\mu, \gamma\mu_k\}$   
8  $k \leftarrow k + 1$   
9 end if  
10 end while
```

LASSO 问题——对比罚函数法和次梯度法

■ 次梯度法 $\mu = 10^{-3}$

■ 罚函数法 $\mu^0 = 10, \quad \gamma = 0.1, \quad \alpha = 0.0002$



其他类型的罚函数法：内点罚函数法

- 考虑不等式约束问题

$$\begin{array}{ll}\min & f(x) \\ \text{s.t.} & c_i(x) \leq 0, \quad i \in \mathcal{I}\end{array}$$

- **定义 6.4** 定义对数罚函数

$$P_I(x, \sigma) = f(x) - \sigma \sum_{i \in \mathcal{I}} \ln(-c_i(x))$$

- 始终要求自变量 x 不能违反约束，适用于不等式约束优化问题
- 当 x 趋于可行域边界时， $P_I(x, \sigma)$ 会趋于正无穷，这说明对数罚函数的极小值严格位于可行域内部，应调整罚因子 σ 使其趋于 0

对数罚函数法算法

算法 对数罚函数法

```
1 给定  $\sigma_0 > 0$ , 可行解  $x^0$ ,  $k \leftarrow 0$ . 罚因子缩小系数  $\rho \in (0, 1)$ 
2 while 未达到收敛准则 do
3 以  $x^k$  为初始点, 求解  $x^{k+1} = \arg \min_x P_I(x, \sigma_k)$ 
4 选取  $\sigma_{k+1} = \rho \sigma_k$ 
5  $k \leftarrow k + 1$ 
6 end while
=====
```

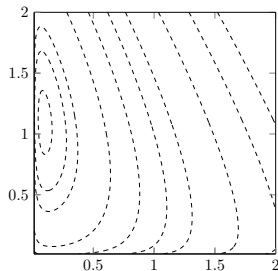
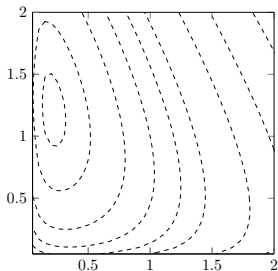
- 初始点 x^0 必须是一个可行点
- 当 σ 趋于 0 时存在数值困难
- 常用的收敛准则 $|\sigma_k \sum_{i \in \mathcal{I}} \ln(-c_i(x^{k+1}))| \leq \varepsilon$

■ 考虑优化问题

$$\begin{aligned} \min \quad & x^2 + 2xy + y^2 + 2x - 2y \\ \text{s.t.} \quad & x \geq 0, y \geq 0 \end{aligned}$$

■ 容易求得最优解为 $(0, 1)$, 考虑对数罚函数 $(\sigma = 1, \sigma = 0.1)$

$$P_I(x, y, \sigma) = x^2 + 2xy + y^2 + 2x - 2y - \sigma(\ln x + \ln y)$$



其他类型的罚函数法：精确罚函数法

- 二次罚函数存在数值困难，并与原问题的解存在误差
- 精确罚函数是一种问题求解时不需要令罚因子趋于正无穷（或零）的罚函数
- **定义** 一般约束优化问题的 ℓ_1 罚函数

$$P(x, \sigma) = f(x) + \sigma \left[\sum_{i \in \mathcal{E}} |c_i(x)| + \sum_{i \in \mathcal{I}} \tilde{c}_i(x) \right]$$

- **定理** 设 x^* 是一般约束优化问题的一个严格局部极小解，且满足 KKT 条件，其对应的拉格朗日乘子为 $\lambda_i^*, i \in \mathcal{E} \cup \mathcal{I}$ ，则**当罚因子 $\sigma > \sigma^*$ 时**， x^* 也为 $P(x, \sigma)$ 的一个局部极小解，其中

$$\sigma^* = \|\lambda^*\|_\infty \stackrel{\text{def}}{=} \max_i |\lambda_i^*|$$

精确罚函数算法

- 1 给定 $\sigma_1 > 0, x_0, k \leftarrow 1$. 罚因子增长系数 $\rho > 1$
- 2 **while** 未达到收敛准则 **do**
- 3 以 x^k 为初始点, 求解
$$x^{k+1} = \arg \min_x \{f(x) + \sigma [\sum_{i \in \mathcal{E}} |c_i(x)| + \sum_{i \in \mathcal{I}} \tilde{c}_i(x)]\}$$
- 4 选取 $\sigma_{k+1} = \rho \sigma_k$
- 5 $k \leftarrow k + 1$
- 6 **end while**

=====

- 初始罚因子过小, 迭代次数增加
- 初始罚因子过大, 子问题求解困难

- 4.1 罚函数法

- 4.2 增广拉格朗日函数法

二次罚函数法的数值困难

■ 对于等式约束问题

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & c_i(x) = 0, \quad i \in \mathcal{E} \end{aligned}$$

■ 二次罚函数

$$\min_x \quad P_E(x, \sigma) = f(x) + \frac{1}{2}\sigma \sum_{i \in \mathcal{E}} c_i^2(x)$$

■ 增广拉格朗日函数

$$L_\sigma(x, \lambda) = f(x) + \sum_{i \in \mathcal{E}} \lambda_i c_i(x) + \frac{1}{2}\sigma \sum_{i \in \mathcal{E}} c_i^2(x)$$

等式约束问题的增广拉格朗日函数法

- 在第 k 步迭代, 给定罚因子 σ_k 和乘子 λ^k , 最小值点 x^{k+1} 满足梯度条件

$$\nabla_x L_{\sigma_k}(x^{k+1}, \lambda^k) = \nabla f(x^{k+1}) + \sum_{i \in \mathcal{E}} (\lambda_i^k + \sigma_k c_i(x^{k+1})) \nabla c_i(x^{k+1}) = 0$$

- 对比等式约束问题的 KKT 条件

$$\nabla f(x^*) + \sum_{i \in \mathcal{E}} \lambda_i^* \nabla c_i(x^*) = 0$$

- 对充分大的 k , 有

$$\lambda_i^* \approx \lambda_i^k + \sigma_k c_i(x^{k+1}) \quad \Rightarrow \quad c_i(x^{k+1}) \approx \frac{1}{\sigma_k} (\lambda_i^* - \lambda_i^k)$$

等式约束问题的增广拉格朗日函数法

算法 增广拉格朗日函数法

- 1 给定坐标 $x^0 \in \mathbb{R}^n$, 乘子 λ^0 , 罚因子 $\sigma_0 > 0$, 约束违反度常数 $\varepsilon > 0$, 精度 $\eta_k > 0$, 迭代步 $k = 0$
- 2 **for** $k = 0, 1, 2, \dots$ **do**
- 3 以 x^k 为初始点, 求解 $\min_x L_{\sigma_k}(x, \lambda^k)$ 得到满足需求的精度条件 $\|\nabla_x L_{\sigma_k}(x, \lambda^k)\| \leq \eta_k$ 的解 x^{k+1}
- 4 **if** $\|c(x^{k+1})\| \leq \varepsilon$ **then**
- 5 返回近似解 (x^{k+1}, λ_k) , 终止迭代
- 6 **end if**
- 7 更新乘子 $\lambda^{k+1} = \lambda^k + \sigma_k c(x^{k+1})$
- 8 更新罚因子 $\sigma_{k+1} = \rho \sigma_k$
- 9 **end for**

ρ 与 σ_k 的取值指导

■ 增广拉格朗日函数

$$L_{\sigma}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E}} \lambda_i c_i(x) + \frac{1}{2} \sigma \sum_{i \in \mathcal{E}} c_i^2(x)$$

■ σ_k 不应增长过快

- 随着 σ_k 的增大, $L_{\sigma_k}(x, \lambda^k)$ 海瑟矩阵的条件数也将增大, 导致数值困难
- σ_k 与 σ_{k+1} 接近时, x^k 可以作为求解 x^{k+1} 的初始点, 以加快收敛

■ σ_k 不应增长过慢

- 算法整体的收敛速度将变慢

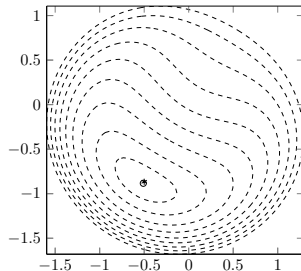
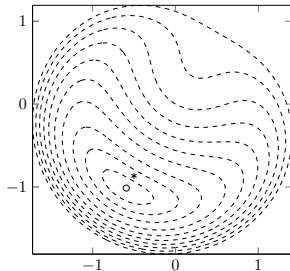
■ 一个经验的取法 $\rho \in [2, 10]$

■ 考虑优化问题

$$\begin{array}{ll}\min & x + \sqrt{3}y \\ \text{s.t.} & x^2 + y^2 = 1\end{array}$$

■ 增广拉格朗日函数

$$L_{\sigma}(x, y, \lambda) = x + \sqrt{3}y + \lambda(x^2 + y^2 - 1) + \frac{\sigma}{2}(x^2 + y^2 - 1)^2$$



收敛性分析

- **定理** 假设乘子列 $\{\lambda^k\}$ 是有界的, 罚因子 $\sigma_k \rightarrow +\infty, k \rightarrow \infty$, 增广拉格朗日方法中精度 $\eta_k \rightarrow 0$, 迭代点列 $\{x^k\}$ 的一个子序列 $\{x^{k_j+1}\}$ 收敛到 x^* , 并且在点 x^* 处 LICQ 成立. 那么存在 λ^* , 满足

$$\lambda^{k_j+1} \rightarrow \lambda^*, \quad j \rightarrow \infty$$

$$\nabla f(x^*) + \nabla c(x^*)\lambda^* = 0, \quad c(x^*) = 0$$

证明 对于增广拉格朗日函数 $L_{\sigma_k}(x, \lambda^k)$, 有

$$\begin{aligned}\nabla_x L_{\sigma_k}(x^{k+1}, \lambda^k) &= \nabla f(x^{k+1}) + \nabla c(x^{k+1})(\lambda^k + \sigma_k c(x^{k+1})) \\ &= \nabla f(x^{k+1}) + \nabla c(x^{k+1})\lambda^{k+1} \\ &= \nabla_x L(x^{k+1}, \lambda^{k+1})\end{aligned}$$

收敛性分析

由于点 x^* 处 LICQ 成立, 故 $\text{rank}(\nabla c(x^{k_j+1})) = |\mathcal{E}|$, 从而成立

$$\lambda^{k_j+1} = (\nabla c(x^{k_j+1})^\top \nabla c(x^{k_j+1}))^{-1} \nabla c(x^{k_j+1})^\top (\nabla_x L_{\sigma_k}(x^{k_j+1}, \lambda^{k_j}) - \nabla f(x^{k_j+1}))$$

因为 $\|\nabla_x L_{\sigma_k}(x^{k_j+1}, \lambda^{k_j})\| \leq \eta_{k_j} \rightarrow 0$, 有

$$\begin{aligned} \lambda^{k_j+1} &\rightarrow \lambda^* \stackrel{\text{def}}{=} -(\nabla c(x^*)^\top \nabla c(x^*))^{-1} \nabla c(x^*)^\top \nabla f(x^*) \\ \nabla_x L(x^*, \lambda^*) &= 0 \end{aligned}$$

而乘子列 $\{\lambda^k\}$ 是有界的, 且 $\lambda^{k_j} + \sigma_{k_j} c(x^{k_j+1}) \rightarrow \lambda^*$, 故 $\{\sigma_{k_j} c(x^{k_j+1})\}$ 有界.
又 $\sigma_k \rightarrow +\infty$, 则 $c(x^*) = 0$

收敛性分析

- **定理** 假设 x^*, λ^* 分别是等式约束优化问题的严格局部极小解和相应的乘子, 则存在充分大的常数 $\bar{\sigma} > 0$ 和充分小的常数 $\delta > 0$, 如果对某个 k , 有

$$\frac{1}{\sigma_k} \|\lambda^k - \lambda^*\| < \delta, \quad \sigma_k \geq \bar{\sigma}$$

则

$$\lambda^k \rightarrow \lambda^*, \quad x^k \rightarrow x^*$$

同时, 如果

- $\limsup \sigma_k < +\infty$ 且 $\lambda^k \neq \lambda^*, \forall k$, 则 $\{\lambda^k\}$ 收敛的速度是 **Q-线性**
- $\limsup \sigma_k = +\infty$ 且 $\lambda^k \neq \lambda^*, \forall k$, 则 $\{\lambda^k\}$ 收敛的速度是 **Q-超线性**

一般约束问题的增广拉格朗日函数法

■ 一般约束优化问题

$$\begin{array}{ll}\min & f(x) \\ \text{s.t.} & c_i(x) = 0, i \in \mathcal{E} \\ & c_i(x) \leq 0, i \in \mathcal{I}\end{array}$$

■ 引入松弛变量, 得到如下等价形式

$$\begin{array}{ll}\min_{x,s} & f(x) \\ \text{s.t.} & c_i(x) = 0, i \in \mathcal{E} \\ & c_i(x) + s_i = 0, i \in \mathcal{I} \\ & s_i \geq 0, i \in \mathcal{I}\end{array}$$

构造增广拉格朗日函数

■ 构造拉格朗日函数

$$L_{\sigma}(x, s, \lambda, \mu) = f(x) + \sum_{i \in \mathcal{E}} \lambda_i c_i(x) + \sum_{i \in \mathcal{I}} \mu_i (c_i(x) + s_i) + \frac{\sigma}{2} p(x, s)$$
$$s_i \geq 0, i \in \mathcal{I}$$

其中

$$p(x, s) = \sum_{i \in \mathcal{E}} c_i^2(x) + \sum_{i \in \mathcal{I}} (c_i(x) + s_i)^2$$

■ 投影梯度法（第七章）

■ 消元法

凸优化问题的增广拉格朗日函数法

■ 考虑凸优化问题

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) \leq 0, i = 1, 2, \dots, m \end{aligned}$$

■ 增广拉格朗日函数

$$L_\sigma(x, \lambda) = f(x) + \frac{\sigma}{2} \sum_{i=1}^m (\max\{\frac{\lambda_i}{\sigma} + c_i(x), 0\}^2 - \frac{\lambda_i^2}{\sigma^2})$$

■ 给定一系列单调递增的乘子 $\sigma_k \uparrow \sigma_\infty$ 和初始乘子 λ^0 , 增广拉格朗日函数法为

$$\begin{cases} x^{k+1} \approx \arg \min_{x \in \mathbb{R}^n} L_{\sigma_k}(x, \lambda^k) \\ \lambda^{k+1} = \max\{0, \lambda^k + \sigma_k c(x^{k+1})\} \end{cases}$$

不精确条件

- 为保证收敛性, $\phi_k(x) = L_{\sigma_k}(x, \lambda^k)$ 的近似解至少满足不精确条件. 例如

$$\phi_k(x^{k+1}) - \inf \phi_k \leq \frac{\varepsilon_k^2}{2\sigma_k}, \quad \varepsilon_k \geq 0, \quad \sum_{k=1}^{\infty} \varepsilon_k < +\infty$$

- 由于 $\inf \phi_k$ 是未知的, 直接验证不可行. 假设 ϕ_k 是 α -强凸函数, 存在

$$\phi_k(x) - \inf \phi_k \leq \frac{1}{2\alpha} \text{dist}^2(0, \partial\phi_k(x))$$

- 构造如下数值可验证的不精确条件

$$\text{dist}(0, \partial\phi_k(x^{k+1})) \leq \sqrt{\frac{\alpha}{\sigma_k}} \varepsilon_k, \quad \varepsilon_k \geq 0, \quad \sum_{k=1}^{\infty} \varepsilon_k < +\infty$$

凸问题的增广拉格朗日函数法的收敛性

- **定理** 假设 $\{x^k\}, \{\lambda^k\}$ 为生成的序列, x^{k+1} 满足不精确条件. 如果 Slater 约束品性成立, 那么序列 $\{\lambda^k\}$ 是有界序列且收敛到 λ^∞ . 进一步, 如果存在一个 γ , 使得下水平集 $\{x \in \mathcal{X} \mid f(x) \leq \gamma\}$ 是非空有界的, 那么序列 $\{x^k\}$ 也是有界的, 并且所有的聚点都是最优解

定理 假设乘子列 $\{\lambda^k\}$ 是有界的, 罚因子 $\sigma_k \rightarrow +\infty, k \rightarrow \infty$, 增广拉格朗日方法中精度 $\eta_k \rightarrow 0$, 迭代点列 $\{x^k\}$ 的一个子序列 $\{x^{k_j+1}\}$ 收敛到 x^* , 并且在点 x^* 处 LICQ 成立. 那么存在 λ^* , 满足

$$\lambda^{k_j+1} \rightarrow \lambda^*, \quad j \rightarrow \infty$$

$$\nabla f(x^*) + \nabla c(x^*)\lambda^* = 0, \quad c(x^*) = 0$$

基追踪问题 (BP)

- 设 $A \in \mathbb{R}^{m \times n} (m \leq n)$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, 基追踪问题被描述为

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad Ax = b$$

- 考虑其对偶问题

$$\min_{y \in \mathbb{R}^m} b^\top y \quad \text{s.t.} \quad \|A^\top y\|_\infty \leq 1$$

\Downarrow

$$\min_{y \in \mathbb{R}^m, s \in \mathbb{R}^n} b^\top y \quad \text{s.t.} \quad A^\top y - s = 0, \|s\|_\infty \leq 1$$

- 对比原始问题和对偶问题的增广拉格朗日函数法

原始问题的增广拉格朗日函数法

- 引入罚因子 σ 和乘子 λ , 原始问题的增广拉格朗日函数为

$$L_\sigma(x, \lambda) = \|x\|_1 + \lambda^\top (Ax - b) + \frac{\sigma}{2} \|Ax - b\|_2^2$$

- 固定 σ , 第 k 步迭代更新格式

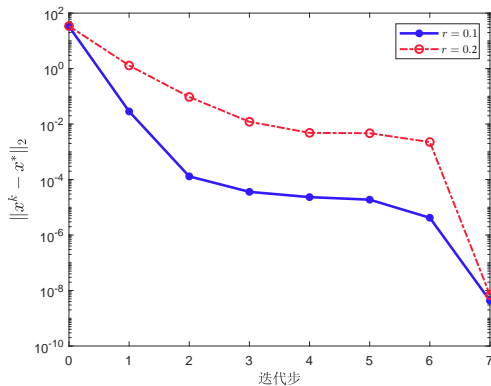
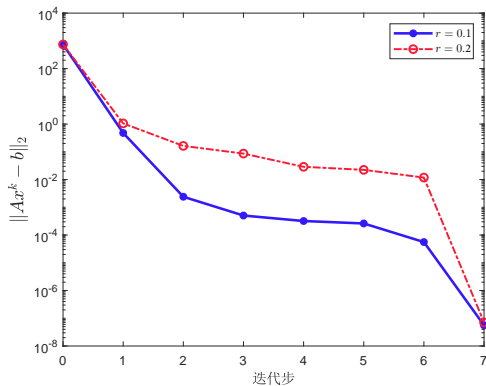
$$\begin{cases} x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \{ \|x\|_1 + \frac{\sigma}{2} \|Ax - b + \frac{\lambda^k}{\sigma}\|_2^2 \} \\ \lambda^{k+1} = \lambda^k + \sigma (Ax^{k+1} - b) \end{cases}$$

- 假设 x^{k+1} 为 $L_\sigma(x, \lambda^k)$ 的一个全局极小解, 则

$$0 \in \partial \|x^{k+1}\|_1 + \sigma A^\top (Ax^{k+1} - b + \frac{\lambda^k}{\sigma}) \quad \Rightarrow \quad -A^\top \lambda^{k+1} \in \partial \|x^{k+1}\|_1$$

BP 问题的实例与解

- 考虑 $b = Au$, 其中 $u \in \mathbb{R}^{1024}$ 服从正态分布, 稀疏度 $r = 0.1$ 或 0.2



对偶问题的增广拉格朗日函数法

■ 考虑对偶问题

$$\min_{y \in \mathbb{R}^m, s \in \mathbb{R}^n} b^\top y \quad \text{s.t.} \quad A^\top y - s = 0, \quad \|s\|_\infty \leq 1$$

■ 引入拉格朗日乘子 λ 和罚因子 σ , 作增广拉格朗日函数

$$L_\sigma(y, s, \lambda) = b^\top y + \lambda^\top (A^\top y - s) + \frac{\sigma}{2} \|A^\top y - s\|_2^2, \quad \|s\|_\infty \leq 1$$

■ 增广拉格朗日函数法的迭代格式为

$$\begin{cases} (y^{k+1}, s^{k+1}) = \arg \min_{y, \|s\|_\infty \leq 1} \{b^\top y + \frac{\sigma_k}{2} \|A^\top y - s\|_2^2 + \frac{\lambda}{\sigma_k}\} \\ \lambda^{k+1} = \lambda^k + \sigma_k (A^\top y^{k+1} - s^{k+1}) \\ \sigma_{k+1} = \min\{\rho \sigma_k, \bar{\sigma}\} \end{cases}$$

消元法求解子问题

- 关于 s 的极小化问题为

$$\min_s \quad \frac{\sigma}{2} \|A^\top y - s + \frac{\lambda}{\sigma}\|_2^2 \quad \text{s.t.} \quad \|s\|_\infty \leq 1$$

- 问题的解为

$$s = \mathcal{P}_{\|s\|_\infty \leq 1} \left(A^\top y + \frac{\lambda}{\sigma} \right)$$

其中 $\mathcal{P}_{\|s\|_\infty \leq 1}(z)$ 为集合 $\{s \mid \|s\|_\infty \leq 1\}$ 的投影算子, 即

$$\mathcal{P}_{\|s\|_\infty \leq 1}(z) = \max\{\min\{z, 1\}, -1\}$$

消元法求解子问题

- 将上述 s 的表达式代入的增广拉格朗日函数法的迭代格式, 得

$$\begin{cases} (y^{k+1}, s^{k+1}) = \arg \min_{y, \|s\|_{\infty} \leq 1} \{b^{\top} y + \frac{\sigma_k}{2} \|A^{\top} y - s + \frac{\lambda}{\sigma_k}\|_2^2\} \\ \lambda^{k+1} = \lambda^k + \sigma_k (A^{\top} y^{k+1} - s^{k+1}) \\ \sigma_{k+1} = \min\{\rho \sigma_k, \bar{\sigma}\} \end{cases}$$

\Downarrow

$$\begin{cases} y^{k+1} = \arg \min_y \{b^{\top} y + \frac{\sigma}{2} \|\psi(A^{\top} y + \frac{\lambda}{\sigma})\|_2^2\} \\ \lambda^{k+1} = \sigma_k \psi(A^{\top} y^{k+1} + \frac{\lambda^k}{\sigma_k}) \\ \sigma_{k+1} = \min\{\rho \sigma_k, \bar{\sigma}\} \end{cases}$$

其中 $\psi(x) = \text{sign}(x) \max\{|x| - 1, 0\}$

Q&A

Thank you!

感谢您的聆听和反馈