第六章 约束优化算法

修贤超

https://xianchaoxiu.github.io

目录

二次罚函数法的数值困难

- 对于等式约束问题

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

- 二次罚函数法需要求解最小化罚函数的子问题

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

- 定义 增广拉格朗日函数为

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

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, $L_{\sigma_k}(x, \lambda^k)$ 的最小值点 x^{k+1} 应满足梯度条件

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

- 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, 应满足

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

等式约束问题的增广拉格朗日函数法

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 $k = 0$

for $k = 0, 1, 2, \cdots$ **do**

以 x^k 为初始点，求解 $\min_x \quad L_{\sigma_k}(x, \lambda^k)$ 得到满足需求的精度条件 $\|\nabla_x L_{\sigma_k}(x, \lambda^k)\| \leqslant \eta_k$ 的解 x^{k+1}

if $\|c(x^{k+1})\| \leqslant \varepsilon$ **then**

返回近似解 (x^{k+1}, λ_k)，终止迭代

end if

更新乘子 $\lambda^{k+1} = \lambda^k + \sigma_k c(x^{k+1})$

更新罚因子 $\sigma_{k+1} = \rho\sigma_k$

end for

例 6.4

- 考虑优化问题

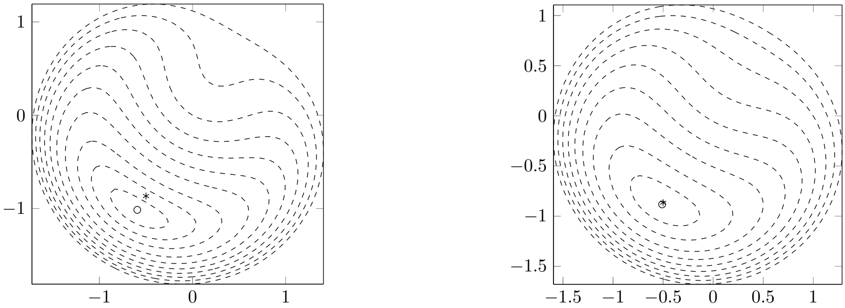
$$\begin{aligned} \min \quad & x + \sqrt{3}y \\ \text{s.t.} \quad & x^2 + y^2 = 1 \end{aligned}$$

- 容易求得最优解为 $x^* = (-\frac{1}{2}, -\frac{\sqrt{3}}{2})^{\mathrm{T}}$, 相应的拉格朗日乘子 $\lambda^* = 1$
- 根据增广拉格朗日函数的形式, 写出本问题的增广拉格朗日函数:

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

例 6.4

- 绘制 $L_2(x, y, 0.9)$ 的等高线
  - “$*$” 为原问题的最优解 x^*
  - “$\circ$” 为二次罚函数或增广拉格朗日函数的最优解



- 增广拉格朗日函数法比二次罚函数法更精确的寻优能力，且约束违反度更低

ρ 与 σ_k 的取值指导

- σ_k 不应增长过快
  - 随着罚因子 σ_k 的增大，可见 $L_{\sigma_k}(x, \lambda^k)$ 关于 x 的海瑟矩阵的条件数也将增大，这将导致数值困难
  - 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 (惩罚不足)
- 一个经验的取法 $\rho \in [2, 10]$

- **定理 6.5** 假设乘子列 $\{\lambda^k\}$ 是有界的，罚因子 $\sigma_k \to +\infty$, $k \to \infty$, 增广拉格朗日方法中精度 $\mu_k \to 0$, 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

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

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

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

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

收敛性分析

由于点 x^* 处 LICQ 成立, 故 $\mathrm{rank}(\nabla c(x^{k_j+1})) = |\mathcal{E}|$ 成立（当 x^{k_j+1} 充分接近 x^* 时），从而下式成立

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

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

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

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

收敛性分析

- **定理 6.6** 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

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

则

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

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

- 一般的约束优化问题可以写成

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

- 引入松弛变量，得到如下等价形式

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

构造增广拉格朗日函数

- **构造拉格朗日函数**

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

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

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

- 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

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

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

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

- $\inf \phi_k$ 是未知的, 直接验证数值上不可行. 但如果 ϕ_k 是 α-强凸函数, 则有

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

- 构造如下数值可验证的不精确条件

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

凸问题的增广拉格朗日函数法的收敛性

- **定理 6.7** 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^∞ (λ^∞ 为对偶问题的一个最优解).
- 如果存在一个 γ, 使得下水平集 $\{x \in \mathcal{X} \mid f(x) \leqslant \gamma\}$ 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

基追踪问题 (BP)

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

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

- 考虑其对偶问题

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

- 通过引入变量 s, 对偶问题可以等价地写成

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

原始问题的增广拉格朗日函数法

- 引入罚因子 σ 和乘子 λ, 其增广拉格朗日函数为

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

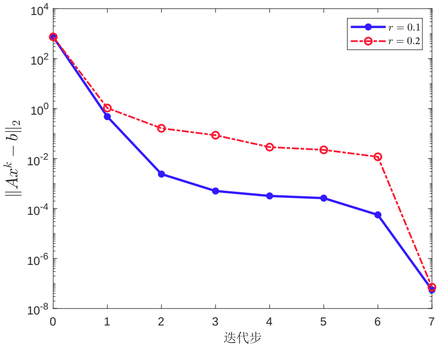
- 固定 σ, 第 k 步迭代更新格式为

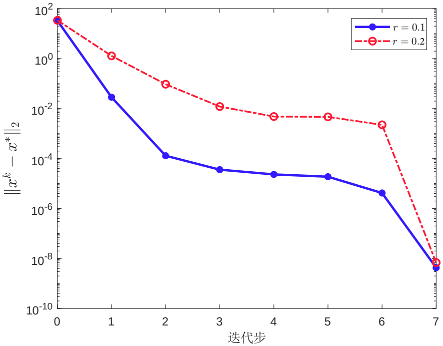
$$\begin{cases} x^{k+1} = \underset{x\in\mathbb{R}^n}{\arg\min}\{\|x\|_1 + \frac{\sigma}{2}\|Ax-b+\frac{\lambda^k}{\sigma}\|_2^2\} \\ \lambda^{k+1} = \lambda^k + \sigma(Ax^{k+1}-b) \end{cases}$$

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

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

- 考虑$b = Au$, 其中 $u \in \mathbb{R}^{1024}$ 是服从正态分布随机稀疏向量, 设其稀疏度 $r = 0.1$ 或 0.2





对偶问题的增广拉格朗日函数法

- 现在考虑对偶问题

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

- 引入拉格朗日乘子 λ 和罚因子 σ, 作增广拉格朗日函数

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

- 增广拉格朗日函数法的迭代格式为

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

消元法求解子问题

- 关于s 的极小化问题为

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

- 问题的解为

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

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

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

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

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

$$\Downarrow$$

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

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

对偶问题的收敛性定理

- **定理 6.8** 假设$\{y^k\},\{\lambda^k\}$ 是由迭代产生的序列，并且 y^{k+1} 的求解精度满足不精确条件，而矩阵 A 是行满秩的. 那么
  - 序列 $\{y^k\}$ 是有界的，且其任一聚点均为对偶问题的最优解
  - 序列 $\{\lambda^k\}$ 有界且收敛，其极限为原始问题的某个最优解
- 基于原问题和对偶问题的算法比较

Q&A

Thank you!

感谢您的聆听和反馈