第六章 约束优化算法

修贤超

https://xianchaoxiu.github.io

目录

约束优化问题

- 考虑约束优化问题

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

- 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 6.1 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

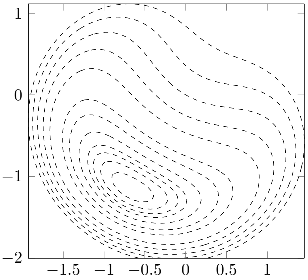
- 对不满足约束的点进行惩罚，被称为外点罚函数

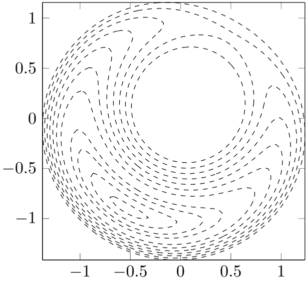
例 6.1

- 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 $(-\frac{1}{2}, -\frac{\sqrt{3}}{2})^\top$
- 考虑二次罚函数 $P_E(x, y, \sigma) = x + \sqrt{3} y + \frac{\sigma}{2}(x^2 + y^2 - 1)^2$





例 6.2

- 考虑优化问题

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

- 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 $\sigma \le 2$，罚函数无下界

二次罚函数法算法

算法 6.1 二次罚函数法

给定 $\sigma_1 > 0, x_0, k \leftarrow 1.$ 罚因子增长系数 $\rho > 1$

while 未达到收敛准则 **do**

以 x^k 为初始点，求解 $x^{k+1} = \arg\min_{x} P_E(x, \sigma_k)$

选取 $\sigma^{k+1} = \rho\sigma_k$

$k \leftarrow k + 1$

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

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

- 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 $\sigma \to \infty$

分析数值困难

- 考虑罚函数 $P_E(x,\sigma)$ 的海瑟矩阵

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

$$\Downarrow$$

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

- $\nabla^2_{xx} P_E(x,\sigma)$ 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

- **定理 6.1** 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 $P_E(x^{k+1},\sigma_k) \leqslant P_E(\bar{x},\sigma_k)$，即

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

$$\Downarrow$$

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

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

收敛性分析

- 定理 6.2 设 $f(x)$ 与 $c_i(x)$ $(i \in \mathcal{E})$ 连续可微，正数序列 $\varepsilon_k \to 0$，$\sigma_k \to +\infty$. 子问题的解 x^{k+1} 满足
$$\|\nabla_x P_E(x^{k+1}, \sigma_k)\| \le \varepsilon_k$$
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
$$\lim_{k\to\infty} (-\sigma_k c_i(x^{k+1})) = \lambda_i^*, \quad \forall i \in \mathcal{E}$$
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**

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

- **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 $\sigma \to \infty$，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

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

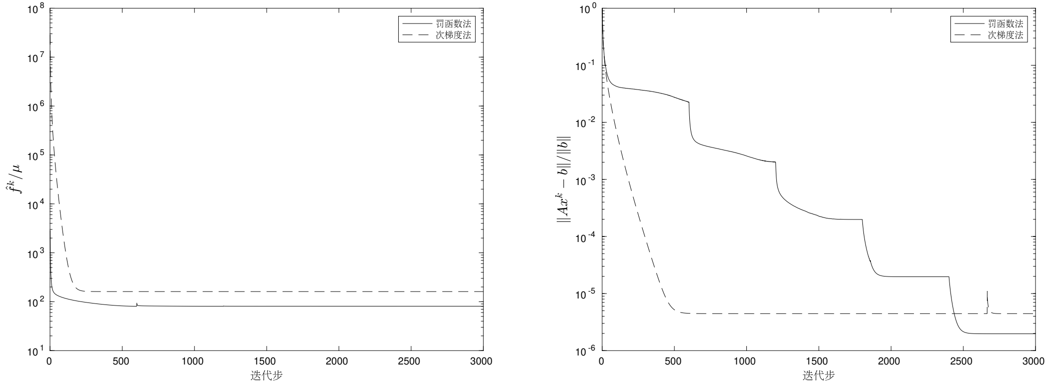
- 仅在 μ 趋于 0 时，LASSO 问题的解收敛于 BP 问题的解
- 当 μ 较小时问题病态，收敛较慢，可逐渐缩小 μ 的值求解子问题逼近

LASSO 问题罚函数法算法

给定 初值 x_0，最终参数 μ，初始参数 μ_0，因子 $\gamma \in (0,1), k \leftarrow 0$
while $\mu_k \geq \mu$ **do**
以 x^k 为初值，求解问题 $x^{k+1} = \arg\min_x \{\frac{1}{2}\|Ax - b\|^2 + \mu_k \|x\|_1\}$
if $\mu_k = \mu$ **then**
停止迭代，输出 x^{k+1}
else
更新罚因子 $\mu_{k+1} = \max\{\mu, \gamma\mu_k\}$
$k \leftarrow k + 1$
end if
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** $\mu^0 = 10$，$\gamma = 0.1$，$\alpha = 0.0002$



其他类型的罚函数法：内点罚函数法

- 考虑不等式约束问题

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

- 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 6.3　对数罚函数法

给定$\sigma_0 > 0$，可行解 x^0，$k \leftarrow 0$. 罚因子缩小系数 $\rho \in (0,1)$
while 未达到收敛准则 **do**
以 x^k 为初始点，求解 $x^{k+1} = \arg\min_x P_I(x, \sigma_k)$
选取 $\sigma_{k+1} = \rho\sigma_k$
$k \leftarrow k+1$
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** $\left|\sigma_k \sum_{i\in\mathcal{I}} \ln(-c_i(x^{k+1}))\right| \leq \varepsilon$

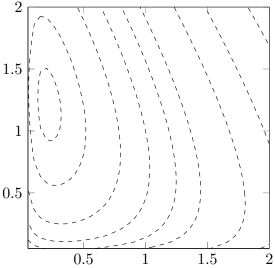
例 6.3

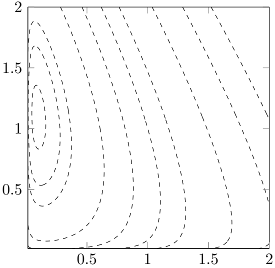
- **考虑优化问题**

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

- **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 6.5　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 6.3　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

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

精确罚函数法算法

给定 $\sigma_1 > 0, x_0, k \leftarrow 1$. 罚因子增长系数 $\rho > 1$
while 未达到收敛准则 **do**
以 x^k 为初始点，求解

$$x^{k+1} = \arg\min_x \{f(x) + \sigma[\sum_{i\in\mathcal{E}} |c_i(x)| + \sum_{i\in\mathcal{I}} \tilde{c}_i(x)]\}$$

选取 $\sigma_{k+1} = \rho\sigma_k$
$k \leftarrow k + 1$
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**

Q&A

Thank you!

感谢您的聆听和反馈