第六章 最优性理论

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
  - 约束优化问题中 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

等式约束的二次罚函数法

- 首先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**

- 由于这种罚函数对不满足约束的点进行惩罚，在迭代过程中点列一般处于可行域之外，因此被称为**外点罚函数**

例 6.1

- 考虑优化问题

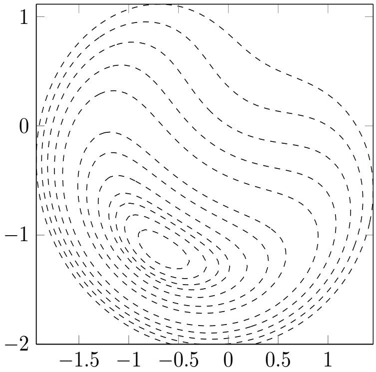
$$\begin{aligned} \min \quad & x + \sqrt{3}y \\ \text{s.t.} \quad & x^2 + y^2 = 1 \end{aligned}$$

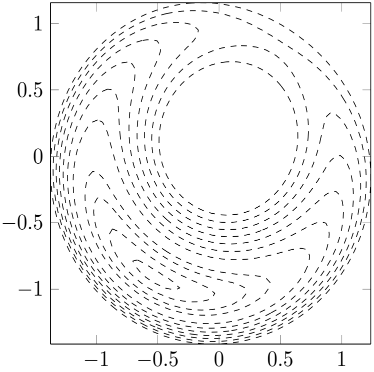
- 容易求得最优解为 $\left(-\frac{1}{2}, -\frac{\sqrt{3}}{2}\right)^\top$
- 考虑二次罚函数

$$P_E(x, y, \sigma) = x + \sqrt{3}y + \frac{\sigma}{2}\left(x^2 + y^2 - 1\right)^2$$

例 6.1

- 绘制出 $\sigma = 1$ 和 $\sigma = 10$ 对应的罚函数的等高线





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 $\sigma\leq 2$，罚函数无下界

二次罚函数法算法

===========

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 $x^{k+1} = \arg\min_x P_E(x, \sigma_k)$

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 (梯度式)，应有下式成立

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

- 最优点处乘子 λ^* 固定，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

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

- 等号右边的前两项可以使用拉格朗日函数 $L(x,\lambda^*)$ 来近似，即

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

- $\nabla_{xx}^2 P_E(x,\sigma)$ 条件数越来越大, 求解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 $P_E\left(x^{k+1},\sigma_k\right)\leqslant P_E(\bar{x},\sigma_k)$，即

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

$$\Downarrow$$

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

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

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

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

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

其中 λ_i^* 是约束 $c_i(x^*)=0$ 对应的拉格朗日乘子

- 虽然不要求每一个子问题精确求解，但要获得原问题的解精度需要越来越高

一般约束问题的二次罚函数法

- 现在考虑一般约束问题

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

- 定义该问题的二次罚函数为

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

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

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，当 $c_i(x)$ 光滑时可以调用一般的无约束光滑优化问题算法求解
  - 二次罚函数形式简洁直观而在实际中广泛使用
- 缺点
  - 需要 $\sigma \to \infty$，此时海瑟矩阵条件数过大，对于无约束优化问题的数值方法拟牛顿法与共轭梯度法存在数值困难，且需要多次迭代求解子问题
  - 对于存在不等式约束的 $P_E(x, \sigma)$ 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
- 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 $x^{k+1} = \arg\min\limits_{x} \left\{ \frac{1}{2}\|Ax - b\|^2 + \mu_k \|x\|_1 \right\}$

if μ_k = μ **then**

停止迭代，输出 x^{k+1}

else

更新罚因子 $\mu_{k+1} = \max\{\mu, \gamma\mu_k\}$

$k \leftarrow k+1$

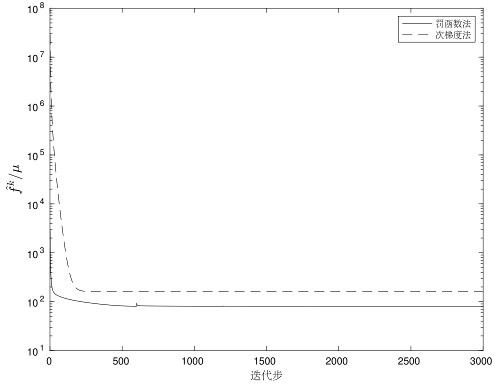
end if

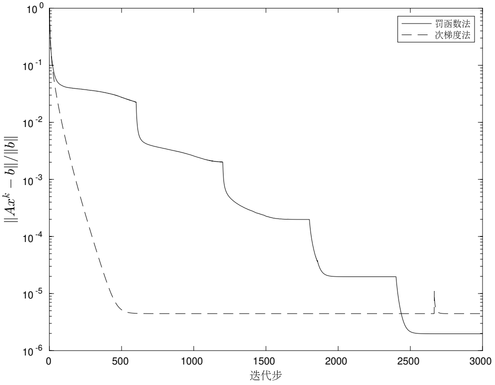
end while

===========

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** μ 从 10 开始，因子 $\gamma = 0.1$，步长 $\alpha = 0.0002$





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

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

- 内点罚函数在迭代时始终要求自变量 x 不能违反约束，故主要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

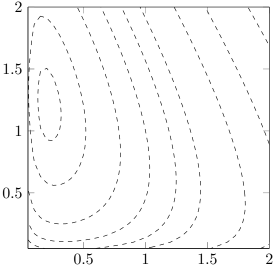
例 6.3

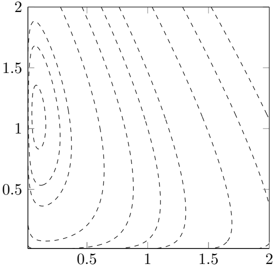
- 考虑优化问题

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

- 容易求得最优解为 $(0,1)$，考虑对数罚函数

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





对数罚函数法算法

===========

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**
- **常用的收敛准则可以包含** $\left|\sigma_k \sum_{i\in\mathcal{I}} \ln(-c_i(x^{k+1}))\right| \le \varepsilon$
- **当 σ 趋于 0 的时候，同样存在数值困难**

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

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

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
- 罚因子过大，子问题求解困难
- 子问题求解的初始点取法不唯一

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

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

- 对比等式约束问题的 KKT 条件

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

- 对充分大的 k，应满足

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

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

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

返回近似解 (x^{k+1}, λ_k)，终止迭代

end if

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

更新罚因子 $\sigma_{k+1} = \rho\sigma_k$

end for

===========

例 6.4

- 考虑优化问题

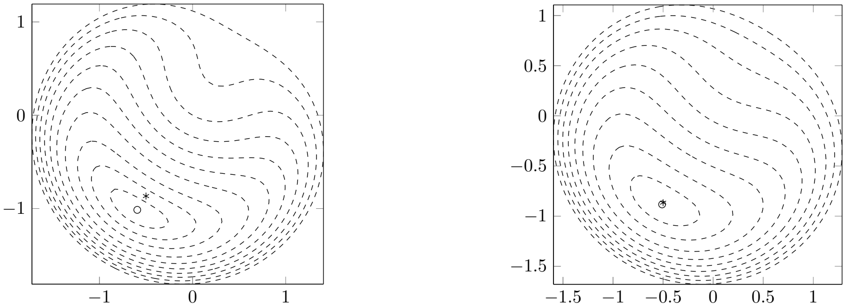
$$\begin{aligned} \min \quad & x+\sqrt{3}y \\ \text{s.t.} \quad & x^2+y^2=1 \end{aligned}$$

- 容易求得最优解为 $x^*=\left(-\frac{1}{2},-\frac{\sqrt{3}}{2}\right)^{\mathrm{T}}$, 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:

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

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 $\rho \in [2, 10]$

收敛性分析

- **定理 6.5** 假设乘子列 $\left\{\lambda^k\right\}$ 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 $\left\{x^k\right\}$ 的一个子序列 $\left\{x^{k_j+1}\right\}$ 收敛到 x^* , 并且在点 x^* 处 LICQ 成立. 那么存在 λ^*, 满足

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

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

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

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

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

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

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

$$\lambda^{k_j+1} \to \lambda^* \overset{\text{def}}{=} -\left(\nabla c\left(x^*\right)^{\mathrm{T}} \nabla c\left(x^*\right)\right)^{-1} \nabla c\left(x^*\right)^{\mathrm{T}} \nabla f\left(x^*\right)$$

$$\nabla_x L\left(x^*, \lambda^*\right) = 0$$

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

- **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

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

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

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

其中

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

- 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

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

- 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

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

不精确条件

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

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

- $\inf \phi_k$ 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

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

- 构造如下数值可验证的不精确条件

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

凸问题的增广拉格朗日函数法的收敛性

- **定理 6.7** 假设$\left\{x^k\right\},\left\{\lambda^k\right\}$ 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 $\left\{\lambda^k\right\}$ 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 $\left\{x^k\right\}$ 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

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

- 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

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

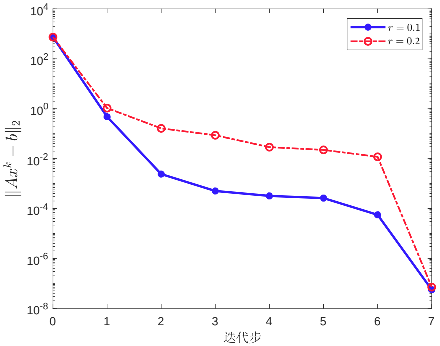
- 固定 σ, 第 k 步迭代更新格式为

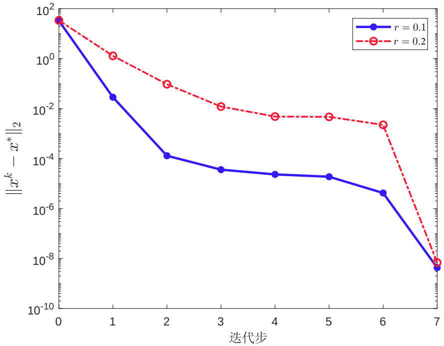
$$\begin{cases} x^{k+1} = \underset{x\in\mathbb{R}^n}{\arg\min} \left\{ \|x\|_1 + \frac{\sigma}{2}\left\| Ax - b + \frac{\lambda^k}{\sigma} \right\|_2^2 \right\} \\ \lambda^{k+1} = \lambda^k + \sigma\left(Ax^{k+1} - b\right) \end{cases}$$

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

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

- 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

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

- 增广拉格朗日函数法的迭代格式为

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

消元法求解子问题

- 关于s 的极小化问题为

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

- 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

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

$$\Downarrow$$

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

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

对偶问题的收敛性定理

- 定理 6.8 假设$\left\{y^k\right\}, \left\{\lambda^k\right\}$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 $\left\{y^k\right\}$ 是有界的，且其任一聚点均为对偶问题的最优解
  - 序列 $\left\{\lambda^k\right\}$ 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