

# 第 7 章：不等式约束

## (1): 障碍方法

---

授课教师：曹语

课程主页：<https://yucaoyc.github.io/math3806>

## 问题背景

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & \mathbf{Ax} = \mathbf{b} \end{array} \quad (\text{新增部分})$$

假设：

- $f_0, f_1, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}$  是二次可微的凸函数；
- $\mathbf{A} \in \mathbb{R}^{p \times n}$ ,  $\text{rank } \mathbf{A} = p$ ;
- Slater 条件成立，即存在严格可行点  $\mathbf{x}$  满足  $f_i(\mathbf{x}) < 0$  且  $\mathbf{Ax} = \mathbf{b}$ ;
- 原问题存在最优解，最优值记为  $p^*$ 。

本节结束后，大家应当能够：

- 1 解释为什么对数障碍函数可以近似不等式约束；
- 2 写出中心点  $\mathbf{x}^*(t)$  对应的 KKT 条件，并说明误差界  $m/t$  的来源；
- 3 复述障碍方法的路径跟踪流程，并说明如何构造严格可行初始点。

考虑原问题

$$\begin{aligned} & \text{minimize} && f_0(\mathbf{x}) \\ & \text{subject to} && f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & && \mathbf{Ax} = \mathbf{b}. \end{aligned}$$

问题：请回忆：若 KKT 条件成立，乘子  $\lambda_i$  应满足哪些条件？

考虑原问题

$$\begin{aligned} & \text{minimize} && f_0(\mathbf{x}) \\ & \text{subject to} && f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & && \mathbf{Ax} = \mathbf{b}. \end{aligned}$$

问题：请回忆：若 KKT 条件成立，乘子  $\lambda_i$  应满足哪些条件？

答案：  $\lambda_i \geq 0$ ,  $\lambda_i f_i(\mathbf{x}^*) = 0$ , 且  $\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla f_i(\mathbf{x}^*) + \mathbf{A}^\top \boldsymbol{\nu}^* = \mathbf{0}$ 。

障碍方法

初始阶段

总结

## 思路:

- 等式约束问题  $\rightarrow$  无约束问题
- 不等式约束问题  $\rightarrow$  仅含等式约束的近似问题

# 障碍方法

思路:

- 等式约束问题  $\rightarrow$  无约束问题
- 不等式约束问题  $\rightarrow$  仅含等式约束的近似问题

构造想法:

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) + \sum_{i=1}^m I_-(f_i(\mathbf{x})) \\ \text{subject to} \quad & \mathbf{Ax} = \mathbf{b} \end{aligned}$$

其中

$$I_-(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0 \end{cases}$$

## 近似问题

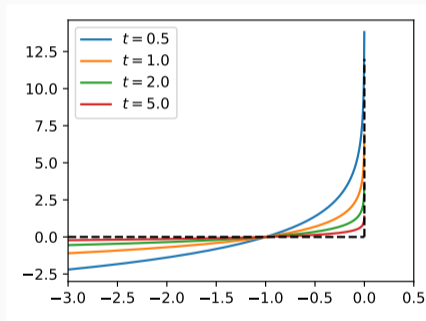
为了应用梯度法、牛顿法等，我们希望目标函数二阶可微，因此可以用一个光滑函数  $\hat{I}_-(u)$  近似  $I_-(u)$ ：

$$\hat{I}_-(u) = -\frac{1}{t} \log(-u),$$

定义域 =  $(-\infty, 0)$

当  $t$  增大时，该函数在可行域内部逐渐接近 0，并且仍会在边界附近迅速增大。

直观上：可行域边界像一堵“软墙”。点越靠近边界，障碍项推回来的力量越强。



# 障碍子问题

因此，原问题可以近似为

$$\begin{aligned} & \text{minimize} && f_0(\mathbf{x}) + \sum_{i=1}^m -\frac{1}{t} \log(-f_i(\mathbf{x})) \\ & \text{subject to} && \mathbf{Ax} = \mathbf{b}. \end{aligned}$$

两边乘以正数  $t$ ，等价于

$$\begin{aligned} & \text{minimize} && tf_0(\mathbf{x}) + \phi(\mathbf{x}) \\ & \text{subject to} && \mathbf{Ax} = \mathbf{b} \end{aligned} \tag{1}$$

其中  $\phi(\mathbf{x}) = -\sum_{i=1}^m \log(-f_i(\mathbf{x}))$  称为**对数障碍函数**。若该问题存在最优解，记为  $\mathbf{x}^*(t)$ ，称为**中心点**；当  $t > 0$  变化时，中心点构成**中心路径**。

注意： $\phi$  的定义域要求  $f_i(\mathbf{x}) < 0$ ，因此障碍方法始终在严格可行域内部工作。

## 练习：障碍子问题的 KKT 条件

对于障碍子问题

$$\begin{aligned} & \text{minimize} && t f_0(\mathbf{x}) + \phi(\mathbf{x}) \\ & \text{subject to} && \mathbf{Ax} = \mathbf{b}, \end{aligned}$$

其中

$$\phi(\mathbf{x}) = - \sum_{i=1}^m \log(-f_i(\mathbf{x})),$$

问题：它的 KKT 条件是什么？

## 练习：障碍子问题的 KKT 条件

对于障碍子问题

$$\begin{aligned} & \text{minimize} && t f_0(\mathbf{x}) + \phi(\mathbf{x}) \\ & \text{subject to} && \mathbf{Ax} = \mathbf{b}, \end{aligned}$$

其中

$$\phi(\mathbf{x}) = - \sum_{i=1}^m \log(-f_i(\mathbf{x})),$$

问题：它的 KKT 条件是什么？

答案：

$$\begin{aligned} \mathbf{Ax}^*(t) &= \mathbf{b}, & f_i(\mathbf{x}^*(t)) &< 0, & i = 1, \dots, m, \\ t \nabla f_0(\mathbf{x}^*(t)) &+ \sum_{i=1}^m \frac{1}{-f_i(\mathbf{x}^*(t))} \nabla f_i(\mathbf{x}^*(t)) \\ &+ \mathbf{A}^\top \hat{\nu} &= \mathbf{0}. \end{aligned}$$

# 先看一个一维例子

问题：如何选择  $t$ ?

先用最简单的问题看清楚中心点怎样靠近边界，以及为什么“一个约束”会带来  $1/t$  量级的误差。

考虑最简单的不等式约束问题：

$$\begin{array}{ll} \text{minimize} & x \\ \text{subject to} & 1 - x \leq 0 \end{array} \quad \implies \quad x^* = 1, \quad p^* = 1.$$

## 先看一个一维例子

考虑最简单的不等式约束问题：

$$\begin{array}{ll} \text{minimize} & x \\ \text{subject to} & 1 - x \leq 0 \end{array} \quad \Longrightarrow \quad x^* = 1, \quad p^* = 1.$$

它的障碍子问题是

$$\min_{x>1} tx - \log(x - 1).$$

一阶条件给出

$$t - \frac{1}{x - 1} = 0 \quad \Longrightarrow \quad x^*(t) = 1 + \frac{1}{t}.$$

因此  $f_0(x^*(t)) - p^* = \frac{1}{t}$ 。

## 先看一个一维例子

考虑最简单的不等式约束问题：

$$\begin{array}{ll} \text{minimize} & x \\ \text{subject to} & 1 - x \leq 0 \end{array} \quad \Longrightarrow \quad x^* = 1, \quad p^* = 1.$$

它的障碍子问题是

$$\min_{x>1} tx - \log(x - 1).$$

一阶条件给出

$$t - \frac{1}{x - 1} = 0 \quad \Longrightarrow \quad x^*(t) = 1 + \frac{1}{t}.$$

因此  $f_0(x^*(t)) - p^* = \frac{1}{t}$ 。

一个约束贡献约  $1/t$  的误差； $m$  个约束就会出现  $m/t$ 。

## 中心点与近似互补

把上一页的驻点条件除以  $t$ , 并记

$$\tilde{\lambda}_i(t) = -\frac{1}{t f_i(\mathbf{x}^*(t))}, \quad i = 1, \dots, m, \quad \tilde{\nu}(t) = \frac{\hat{\nu}}{t}$$

则中心点满足

$$\nabla f_0(\mathbf{x}^*(t)) + \sum_{i=1}^m \tilde{\lambda}_i(t) \nabla f_i(\mathbf{x}^*(t)) + \mathbf{A}^\top \tilde{\nu}(t) = \mathbf{0}.$$

同时

$$\tilde{\lambda}_i(t) > 0, \quad \tilde{\lambda}_i(t) (-f_i(\mathbf{x}^*(t))) = 1/t.$$

这说明中心点满足一种“近似互补松弛”条件：不是完全贴在边界上，而是离边界保持一个由  $t$  控制的距离。

## 误差界：两个关键事实

先注意： $\mathbf{x}^*(t)$  是原问题的严格可行点，所以

$$p^* \leq f_0(\mathbf{x}^*(t)).$$

由近似互补关系可得

$$\sum_{i=1}^m \tilde{\lambda}_i(t) f_i(\mathbf{x}^*(t)) = -\frac{m}{t}, \quad \tilde{\nu}(t)^\top (\mathbf{A}\mathbf{x}^*(t) - \mathbf{b}) = 0.$$

每个约束给出  $-1/t$ ，所以  $m$  个约束给出  $-m/t$ 。

## 误差界：从对偶函数得到结论

因为  $\tilde{\lambda}_i(t) > 0$ ，原问题的拉格朗日函数关于  $\mathbf{x}$  仍是凸函数。由驻点条件可知， $\mathbf{x}^*(t)$  是它在  $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$  下的全局最小点。因此对偶函数满足

$$\begin{aligned} g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) &= f_0(\mathbf{x}^*(t)) + \sum_{i=1}^m \tilde{\lambda}_i(t) f_i(\mathbf{x}^*(t)) + \tilde{\boldsymbol{\nu}}(t)^\top (\mathbf{A}\mathbf{x}^*(t) - \mathbf{b}) \\ &= f_0(\mathbf{x}^*(t)) - \frac{m}{t}. \end{aligned}$$

再由弱对偶性  $g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \leq p^*$ ，以及  $p^* \leq f_0(\mathbf{x}^*(t))$  可得

$$0 \leq f_0(\mathbf{x}^*(t)) - p^* \leq \frac{m}{t}$$

具体实验验证请见对应代码。

因此，对于误差阈值  $\epsilon$ ，令  $t = m/\epsilon$ ，并精确求解上述问题得到  $x^*(t)$ ，即可保证目标函数值的误差不超过  $\epsilon$

即通过对数障碍函数来近似原问题的误差是可控的

**挑战：**对于很小的误差阈值  $\epsilon$ ，直接设定很大的  $t$  会使近似问题更难求解。

**思路：**对于困难的问题，分多步走

# 障碍方法/路径跟踪方法

20 世纪 60 年代, Fiacco 和 McCormick 提出了如下路径跟踪思想:

---

## Algorithm 1: 障碍方法/路径跟踪方法

---

**Input:** 给定严格可行点  $x$ , 阈值  $\epsilon$ ,  $t > 0$ ,  $\mu > 1$

```
1 while True do
2     以  $x$  为初始值, 求解优化问题(1)得到  $x^*(t)$ 
3     更新:  $x \leftarrow x^*(t)$ 
4     停止准则: 如果  $m/t \leq \epsilon$ , 则退出
5     令  $t \leftarrow t\mu$ ,
6 end
```

---

一般而言, 可以取  $\mu$  在  $3 \sim 100$ ; 这里的  $t$  是障碍参数, 不是回溯线搜索中的步长。

## 参数 $\mu$ 的影响

问题:  $\mu$  取得过大或者过小时, 算法会出现什么现象?

问题:  $\mu$  取得过大或者过小时, 算法会出现什么现象?

- $\mu$  较小: 相邻中心点距离近, 内层问题较容易, 但外层迭代次数较多;
- $\mu$  较大: 外层迭代次数较少, 但每次中心点变化大, 内层求解可能更困难。

# 牛顿步径

**练习：**如果使用牛顿法来求解问题(1)，牛顿步径如何求解呢？

# 牛顿步径

**练习:** 如果使用牛顿法来求解问题(1), 牛顿步径如何求解呢?

**答案:** 对于  $\tilde{f}(\mathbf{x}) = tf_0(\mathbf{x}) + \phi(\mathbf{x})$

$$\begin{bmatrix} \nabla^2 \tilde{f}(\mathbf{x}) & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_{\text{nt}} \\ \boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} -\nabla \tilde{f}(\mathbf{x}) \\ \mathbf{b} - \mathbf{A}\mathbf{x} \end{bmatrix}$$

变成了

$$\begin{bmatrix} t\nabla^2 f_0(\mathbf{x}) + \nabla^2 \phi(\mathbf{x}) & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_{\text{nt}} \\ \boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} -t\nabla f_0(\mathbf{x}) - \nabla \phi(\mathbf{x}) \\ \mathbf{b} - \mathbf{A}\mathbf{x} \end{bmatrix}$$

## 对数障碍的 Hessian 矩阵

为了实际计算牛顿步径，还需要  $\nabla\phi$  和  $\nabla^2\phi$ ：

$$\nabla\phi(\mathbf{x}) = \sum_{i=1}^m \frac{1}{-f_i(\mathbf{x})} \nabla f_i(\mathbf{x}),$$

$$\begin{aligned} \nabla^2\phi(\mathbf{x}) &= \sum_{i=1}^m \frac{1}{f_i(\mathbf{x})^2} \nabla f_i(\mathbf{x}) \nabla f_i(\mathbf{x})^\top \\ &\quad + \sum_{i=1}^m \frac{1}{-f_i(\mathbf{x})} \nabla^2 f_i(\mathbf{x}) \end{aligned}$$

障碍方法

初始阶段

总结

# 初始点构造 1

理论上我们已经假设严格可行点存在；实际运行障碍法前，还需要先找到某个  $\mathbf{x}$  满足

$$f_i(\mathbf{x}) < 0, \quad i = 1, 2, \dots, m, \quad \mathbf{Ax} = \mathbf{b}$$

为了找到这样的初值，我们可以求解如下的优化问题：

$$\begin{aligned} & \text{minimize}_{\mathbf{x}, s} && s \\ & \text{subject to} && f_i(\mathbf{x}) \leq s, \quad i = 1, \dots, m \\ & && \mathbf{Ax} = \mathbf{b} \end{aligned}$$

我们标记  $\bar{p}^*$  = 该问题的最优值。

若等式约束本身可行，只要先找到满足  $\mathbf{Ax} = \mathbf{b}$  的点，再取足够大的  $s$ ，就容易得到该辅助问题的严格可行初值。直观上， $s$  像一个“临时放宽额度”：先把门开大，再逐渐把  $s$  压小。

## 初始点构造 1: 如何判断

- ☺ 若  $\bar{p}^* < 0$ , 任取满足  $s < 0$  的解, 其  $x$  就是原问题的严格可行解。
- ☹ 若  $\bar{p}^* > 0$ , 则原问题无可行解。
- ☹ 若  $\bar{p}^* = 0$  且能达到, 则原问题有可行解, 但没有严格可行解。
- ☹ 若  $\bar{p}^* = 0$  且不能达到, 则原问题无可行解。

因此, 初始点构造 1 不仅能帮助寻找严格可行点, 也能判断原问题是否违反 Slater 条件或不可行。

## 初始点构造 2

若只需要找到一个可行解，也可以为每个不等式约束引入一个松弛量：

$$\begin{aligned} & \text{minimize} && \mathbf{1}^\top \mathbf{s} \\ & \text{subject to} && f_i(\mathbf{x}) \leq s_i, \quad i = 1, \dots, m \\ & && \mathbf{Ax} = \mathbf{b} \\ & && \mathbf{s} \succeq \mathbf{0} \end{aligned}$$

若该问题的最优值达到 0，则达到最优值的  $\mathbf{x}$  就是原问题的可行解。

注意：这只能保证  $f_i(\mathbf{x}) \leq 0$ ，不一定保证  $f_i(\mathbf{x}) < 0$ 。因此它适合判断可行性；若要启动障碍方法，通常还需要严格可行点。

- 若  $m = 20$  且希望目标函数值误差不超过  $10^{-3}$ , 障碍参数  $t$  至少应取多大?
- 若初始点构造 1 得到  $\bar{p}^* < 0$ , 这说明什么?

## 快速检查

- 若  $m = 20$  且希望目标函数值误差不超过  $10^{-3}$ , 障碍参数  $t$  至少应取多大?
- 若初始点构造 1 得到  $\bar{p}^* < 0$ , 这说明什么?

### 答案:

- $t \geq 2 \times 10^4$ ;
- 原问题存在严格可行解。

障碍方法

初始阶段

总结

# 总结

主要需要掌握的知识:

- 将不等式约束问题转化为仅含等式约束的近似问题，即公式(1)；
- 理解误差界  $0 \leq f_0(\mathbf{x}^*(t)) - p^* \leq m/t$ ，并能够复述中心路径跟踪流程；
- 若没有严格可行初始点，可以先用初始点构造问题寻找或判断可行性。

阅读作业 & 参考资料:

- 课本第 11.1 - 11.4 章