第 7 章：不等式约束

(2)：原对偶内点法

授课教师：曹语

课程主页：`https://yucaoyc.github.io/math3806`

问题背景：从障碍法到原对偶法

$$\begin{array}{ll} \text{minimize} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}) \leqslant 0, \quad i = 1, \cdots, m \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

沿用上一节假设：$f_0, \cdots, f_m : \mathbb{R}^n \to \mathbb{R}$ 二阶可微且凸，$\boldsymbol{A} \in \mathbb{R}^{p \times n}$ 且 $\operatorname{rank} \boldsymbol{A} = p$，Slater 条件成立，原问题存在最优解。

上一节的障碍方法考虑近似问题

$$\begin{array}{ll} \text{minimize} & t f_0(\boldsymbol{x}) + \phi(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

其中 $t > 0$，$\phi(\boldsymbol{x}) = -\sum_{i=1}^{m} \log(-f_i(\boldsymbol{x}))$。

本节想法：不必把每个障碍子问题都精确解完，而是直接对中心路径的 KKT 方程做牛顿修正。

两种方法的区别：解子问题 vs 修正 KKT

第 7-1 节：障碍方法

固定 t_k
解障碍子问题

得到中心点
$\boldsymbol{x}^\star(t_k)$

增大 t
$t_{k+1} = \mu t_k$

逐站停靠：每一轮重
新解一个障碍子问题

相同：中心路径思想
不同：解子问题 vs 修正 KKT

第 7-2 节：原对偶内点法

当前 $(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$
不必是中心点

计算残差
$r_{\text{dual}}, r_{\text{cent}}, r_{\text{pri}}$

一次牛顿修正
同时更新原变量和乘子

边走边校准：每一
轮用 KKT 残差修正

学习目标

本节结束后，大家应当能够：

1. 从障碍问题的 KKT 条件推出扰动 KKT 方程组；
2. 写出原对偶残差，并由线性化得到原对偶牛顿方向；
3. 复述原对偶内点法的基本流程、停止准则和线搜索要求。

课前回顾

上一节得到中心点 $\boldsymbol{x}^\star(t)$ 满足

$$\begin{aligned}
&\boldsymbol{A}\boldsymbol{x}^\star(t)=\boldsymbol{b}, \qquad f_i\left(\boldsymbol{x}^\star(t)\right)<0, \quad i=1,\cdots,m,\\
&t\nabla f_0\left(\boldsymbol{x}^\star(t)\right)+\sum_{i=1}^m \frac{1}{-f_i\left(\boldsymbol{x}^\star(t)\right)}\nabla f_i\left(\boldsymbol{x}^\star(t)\right)+\boldsymbol{A}^\top\widehat{\boldsymbol{\nu}}=\mathbf{0}.
\end{aligned}$$

问题：如果两边同除以 t，应该如何定义新的乘子 $\lambda_i(t)$?

课前回顾

上一节得到中心点 $\boldsymbol{x}^\star(t)$ 满足

$$\boldsymbol{A}\boldsymbol{x}^\star(t) = \boldsymbol{b}, \qquad f_i\left(\boldsymbol{x}^\star(t)\right) < 0, \quad i = 1, \cdots, m,$$
$$t\nabla f_0\left(\boldsymbol{x}^\star(t)\right) + \sum_{i=1}^{m} \frac{1}{-f_i\left(\boldsymbol{x}^\star(t)\right)} \nabla f_i\left(\boldsymbol{x}^\star(t)\right) + \boldsymbol{A}^\top \widehat{\boldsymbol{\nu}} = \boldsymbol{0}.$$

问题：如果两边同除以 t，应该如何定义新的乘子 $\lambda_i(t)$?

答案：$\lambda_i(t) = -1/\big(tf_i(\boldsymbol{x}^\star(t))\big) > 0$，于是 $-\lambda_i(t)f_i(\boldsymbol{x}^\star(t)) = 1/t$。

目录

从中心路径到扰动 KKT

在中心点 $\boldsymbol{x}^\star(t)$ 上，如果令

$$\lambda_i(t) = -\frac{1}{t f_i(\boldsymbol{x}^\star(t))}, \quad i = 1, \cdots, m, \qquad \boldsymbol{\nu}(t) = \frac{\widehat{\boldsymbol{\nu}}}{t},$$

省略 (t) 后，障碍问题的 KKT 条件可等价地写成如下扰动 KKT 方程组：

$$\begin{aligned} \nabla f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\boldsymbol{x}) + \boldsymbol{A}^\top \boldsymbol{\nu} = \boldsymbol{0}, \\ -\lambda_i f_i(\boldsymbol{x}) = \frac{1}{t}, \quad i = 1, \cdots, m, \\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \\ \lambda_i > 0, \quad i = 1, \cdots, m. \end{aligned} \tag{1}$$

由于 $\lambda_i > 0$，第二行也隐含 $f_i(\boldsymbol{x}) < 0$。当 $t \to \infty$ 时，它逐渐接近原问题 KKT 条件中的互补松弛 $\lambda_i f_i(\boldsymbol{x}) = 0$。

一个一维例子：中心路径在做什么？

先看一个最小模型：

$$
\begin{array}{ll}
\text{minimize} & x \\
\text{subject to} & x \geq 0.
\end{array}
\qquad \Longleftrightarrow \qquad f_0(x) = x, \quad f_1(x) = -x \leq 0.
$$

没有等式约束时，扰动 KKT 条件变成

$$
1 - \lambda = 0, \qquad -\lambda(-x) = \frac{1}{t}, \qquad \lambda > 0.
$$

一个一维例子：中心路径在做什么？

先看一个最小模型：

$$\begin{array}{ll} \text{minimize} & x \\ \text{subject to} & x \geq 0. \end{array} \quad \Longleftrightarrow \quad f_0(x) = x, \quad f_1(x) = -x \leq 0.$$

没有等式约束时，扰动 KKT 条件变成

$$1 - \lambda = 0, \qquad -\lambda(-x) = \frac{1}{t}, \qquad \lambda > 0.$$

因此

$$\lambda(t) = 1, \qquad x(t) = \frac{1}{t}.$$

图像理解： t 越大，$x(t)$ 越靠近真正最优解 $x^\star = 0$，但始终留在可行域内部 $x > 0$。

原对偶残差

记 $\boldsymbol{y} = (\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$，并定义

$$\boldsymbol{r}_t(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{bmatrix} r_{\mathsf{dual}} \\ r_{\mathsf{cent}} \\ r_{\mathsf{pri}} \end{bmatrix} = \begin{bmatrix} \nabla f_0(\boldsymbol{x}) + D\boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{\lambda} + \boldsymbol{A}^\top \boldsymbol{\nu} \\ -\operatorname{diag}(\boldsymbol{\lambda})\boldsymbol{f}(\boldsymbol{x}) - \frac{1}{t}\mathbf{1} \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \end{bmatrix}.$$

其中

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{bmatrix}, \qquad D\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} \nabla f_1(\boldsymbol{x})^\top \\ \vdots \\ \nabla f_m(\boldsymbol{x})^\top \end{bmatrix}.$$

三个残差分别在检查什么？

可以把三个残差想成算法的三个“仪表盘”：

- r_{dual}：检查 Lagrange 函数的驻点条件是否接近成立；
- r_{cent}：检查当前点是否接近参数 t 对应的中心路径；
- r_{pri}：检查等式约束是否接近满足。

问题：若 $\boldsymbol{r}_t(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \mathbf{0}$，那么 $\boldsymbol{x}$ 和 $\boldsymbol{\lambda}$ 满足什么关系？

三个残差分别在检查什么？

可以把三个残差想成算法的三个“仪表盘”：

- r_{dual}：检查 Lagrange 函数的驻点条件是否接近成立；
- r_{cent}：检查当前点是否接近参数 t 对应的中心路径；
- r_{pri}：检查等式约束是否接近满足。

问题：若 $\boldsymbol{r}_t(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \mathbf{0}$，那么 $\boldsymbol{x}$ 和 $\boldsymbol{\lambda}$ 满足什么关系？

答案：在算法维护的区域 $f_i(\boldsymbol{x}) < 0$、$\lambda_i > 0$ 内，此时 $\boldsymbol{x}$ 在等式约束上可行，且 $-\lambda_i f_i(\boldsymbol{x}) = 1/t$，所以它位于参数 t 对应的中心路径上。

线性化证明：残差的一阶变化

记 $\Delta \boldsymbol{y} = (\Delta \boldsymbol{x}, \Delta \boldsymbol{\lambda}, \Delta \boldsymbol{\nu})$。在当前 t 固定时，牛顿法使用一阶近似：

$$\boldsymbol{r}_t(\boldsymbol{y} + \Delta \boldsymbol{y}) \approx \boldsymbol{r}_t(\boldsymbol{y}) + D\boldsymbol{r}_t(\boldsymbol{y})\Delta \boldsymbol{y} = \boldsymbol{0} \quad \Longrightarrow \quad D\boldsymbol{r}_t(\boldsymbol{y})\Delta \boldsymbol{y} = -\boldsymbol{r}_t(\boldsymbol{y}).$$

三类残差的微分分别为

$$\begin{aligned}
dr_{\mathsf{dual}} &= \left(\nabla^2 f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \nabla^2 f_i(\boldsymbol{x})\right) \Delta \boldsymbol{x} + D\boldsymbol{f}(\boldsymbol{x})^\top \Delta \boldsymbol{\lambda} + \boldsymbol{A}^\top \Delta \boldsymbol{\nu}, \\
dr_{\mathsf{cent},i} &= d\left(-\lambda_i f_i(\boldsymbol{x}) - \frac{1}{t}\right) = -\lambda_i \nabla f_i(\boldsymbol{x})^\top \Delta \boldsymbol{x} - f_i(\boldsymbol{x}) \Delta \lambda_i, \\
dr_{\mathsf{pri}} &= \boldsymbol{A} \Delta \boldsymbol{x}.
\end{aligned}$$

注意：$1/t$ 在本次牛顿步中视为常数，所以其微分为零。

从残差到牛顿系统：逐行看

令

$$\boldsymbol{H} = \nabla^2 f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \nabla^2 f_i(\boldsymbol{x}).$$

把上一页的微分写成向量形式，并令一阶近似残差为零：

$$\begin{aligned}
r_{\mathsf{dual}}\text{行：} \quad & \boldsymbol{H}\Delta\boldsymbol{x} + D\boldsymbol{f}(\boldsymbol{x})^\top \Delta\boldsymbol{\lambda} + \boldsymbol{A}^\top \Delta\boldsymbol{\nu} = -r_{\mathsf{dual}}, \\
r_{\mathsf{cent}}\text{行：} \quad & -\operatorname{diag}(\boldsymbol{\lambda}) D\boldsymbol{f}(\boldsymbol{x})\Delta\boldsymbol{x} - \operatorname{diag}(\boldsymbol{f}(\boldsymbol{x}))\Delta\boldsymbol{\lambda} = -r_{\mathsf{cent}}, \\
r_{\mathsf{pri}}\text{行：} \quad & \boldsymbol{A}\Delta\boldsymbol{x} = -r_{\mathsf{pri}}.
\end{aligned}$$

下一页只是把这三行拼成一个大的线性方程组。

原对偶牛顿方向

将残差方程 $\boldsymbol{r}_t(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \mathbf{0}$ 线性化，得到牛顿方向：

$$\begin{bmatrix} \nabla^2 f_0(\boldsymbol{x}) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(\boldsymbol{x}) & D\boldsymbol{f}(\boldsymbol{x})^\top & \boldsymbol{A}^\top \\ -\operatorname{diag}(\boldsymbol{\lambda}) D\boldsymbol{f}(\boldsymbol{x}) & -\operatorname{diag}(\boldsymbol{f}(\boldsymbol{x})) & 0 \\ \boldsymbol{A} & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{x} \\ \Delta \boldsymbol{\lambda} \\ \Delta \boldsymbol{\nu} \end{bmatrix} = - \begin{bmatrix} r_{\text{dual}} \\ r_{\text{cent}} \\ r_{\text{pri}} \end{bmatrix}.$$

这个方向同时更新原变量 $\boldsymbol{x}$、不等式乘子 $\boldsymbol{\lambda}$ 和等式乘子 $\boldsymbol{\nu}$，因此称为原对偶内点法。

目录

替代精确中心化

障碍方法的每一步需要近似求解

$$\begin{array}{ll} \text{minimize} & t f_0(\boldsymbol{x}) + \phi(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{array}$$

原对偶内点法改为维护三个量：

- 原变量 $\boldsymbol{x}$：保持在不等式约束内部，即 $f_i(\boldsymbol{x}) < 0$；
- 对偶变量 $\boldsymbol{\lambda} \succ \boldsymbol{0}$ 和 $\boldsymbol{\nu}$；
- 代理对偶间隙 $\hat{\eta} = -\boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{\lambda}$。

若中心残差 r_{cent} 为零，则 $\hat{\eta} = m/t$。

原对偶内点法

Algorithm 1: 原对偶内点法

Input: 给定满足 $f_i(\boldsymbol{x}) < 0,\ i = 1, \cdots, m$ 的 $\boldsymbol{x}$，$\boldsymbol{\lambda} \succ \mathbf{0}$，$\boldsymbol{\nu}$，阈值 $\epsilon, \epsilon_{\text{feas}}$，参数 $\mu > 1$

while *True* **do**

  计算 $\hat{\eta} := -\boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{\lambda}$;

  停止准则：如果 $\|r_{\text{pri}}\| \leq \epsilon_{\text{feas}}$，$\|r_{\text{dual}}\| \leq \epsilon_{\text{feas}}$，$\hat{\eta} \leq \epsilon$，则退出;

  令 $t = \mu m / \hat{\eta}$;

  求解牛顿系统，得到原对偶搜索方向 $\Delta \boldsymbol{y} = (\Delta \boldsymbol{x}, \Delta \boldsymbol{\lambda}, \Delta \boldsymbol{\nu})$;

  通过回溯线搜索选择步长 $s > 0$，并更新 $\boldsymbol{y} \leftarrow \boldsymbol{y} + s \Delta \boldsymbol{y}$;

end

这是常见的“等式约束可暂时不可行”的版本；若初始点满足 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 且方向保持 $\boldsymbol{A}\Delta\boldsymbol{x} = 0$，则可得到可行初始点版本。

线搜索要检查什么？

- 保持不等式严格可行：

$$f_i(\boldsymbol{x} + s\Delta\boldsymbol{x}) < 0, \quad i = 1, \cdots, m;$$

- 保持乘子为正：

$$\boldsymbol{\lambda} + s\Delta\boldsymbol{\lambda} \succ \mathbf{0};$$

- 使当前 t 下的残差足够下降，例如检查 $\|\boldsymbol{r}_t(\boldsymbol{y} + s\Delta\boldsymbol{y})\|_2$ 相对于当前残差有充分下降。

这里 t 是障碍参数，s 是线搜索步长，二者含义不同。

课堂快速检验

假设某次迭代已经达到

$$r_{\text{dual}} = \mathbf{0}, \qquad r_{\text{pri}} = \mathbf{0}, \qquad r_{\text{cent}} = \mathbf{0}.$$

请回答:

1. 此时 $\lambda_i f_i(\boldsymbol{x})$ 等于多少?
2. 代理对偶间隙 $\hat{\eta} = -\boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{\lambda}$ 等于多少?
3. 当 t 越来越大时，这和原问题 KKT 条件中的哪一条越来越接近?

答案提示: $\lambda_i f_i(\boldsymbol{x}) = -1/t$，$\hat{\eta} = m/t$，并且逐渐接近互补松弛条件。

目录

总结

主要需要掌握的知识：

- 原对偶内点法来自中心路径的扰动 KKT 方程组；
- 三个残差 r_{dual}，r_{cent}，r_{pri} 分别对应驻点、中心性和原可行性；
- 牛顿系统来自对残差方程 $\boldsymbol{r}_t = \mathbf{0}$ 的线性化；
- 算法通过 $\hat{\eta}$ 选择新的障碍参数 t，并用线搜索保持 $f_i(\boldsymbol{x}) < 0$ 与 $\boldsymbol{\lambda} \succ \mathbf{0}$。

阅读作业 & 参考资料：

- 课本第 11.7 章