拉格朗日法, 对偶和 Kuhn-Tucker 条件

- 信息经济学
- 湖南大学课程

1. 一般情形下的最优激励合同

1. 委托人 (公司) 提供工资合同 $w(q)$, 代理人 (打工人张三) 选择接受或拒绝.
   - 若张三拒绝, 博弈结束; 否则, 博弈进入下一阶段.
2. 张三选择行动 $a \in A$, 行动集 $A \subseteq \mathbb{R}$ 是有界闭集.
3. 给定张三的行动 a, 产出 q 从分布 $f(\cdot \mid a)$ 中抽样得到.
4. 公司和张三观测到产出 q 后, 公司按照合同 $w(q)$ 支付工资. 博弈结束.

效用函数:

- 张三的效用函数为 $u(w) - c(a)$.
  - 张三是风险厌恶的: $u(w)$ 为严格递增的凹函数.
  - 成本函数 $c(a)$ 为严格递增的凸函数.
  - 张三的保留效用 $\underline{u} \in \mathbb{R}$ 外生给定.
- 公司是风险中性的, 其最终利润为 $q - w(q)$.

和上一讲 (线性激励合同) 的情形相比, 主要区别在于:

- 工资合同 $w(q)$ 不需要是线性的
- 分布族 $f(\cdot \mid a)$ 不需要是正态分布族
  - 之后需要引入其它假设, 来保证张三的高努力行动会导致 “更好的” 产出分布

2. 引入求解最优化问题的数学工具

描述委托人的最优化问题不难, 写出 IC 和 IR 条件即可. 难点在于如何求解该问题.

引入必要的数学工具:

- 拉格朗日法和 Kuhn-Tucker 条件.

3. 包含等式约束的最优化问题

考虑如下包含 k 个决策变量和 n 个等式约束的最优化问题:

$$\begin{aligned} \max_{x_1, x_2, \dots, x_k} \quad & f(x_1, x_2, ..., x_k) \\ \text{subject to} \quad & g_1(x_1, x_2, ..., x_k) = 0 \\ & g_2(x_1, x_2, ..., x_k) = 0 \\ & \vdots \\ & g_n(x_1, x_2, ..., x_k) = 0 \end{aligned}$$

令 $x = (x_1, x_2, ..., x_k)$ 且 $\lambda = (\lambda_1, ..., \lambda_n)$, 写出拉格朗日函数:

$$L(x, \lambda) = f(x_1, x_2, ..., x_k) - \sum_{j=1}^{n} \lambda_j g_j(x_1, x_2, ..., x_k)$$

一阶条件:

- 对任意 $i = 1, 2, ..., k$, 有 $\partial L(x, \lambda)/\partial x_i = 0 \Longrightarrow \partial f(x)/\partial x_i - \sum_{j=1}^{n} \lambda_j \partial g_j(x)/\partial x_i = 0$

- 几何解释: 上面这 k 个等式描述了均衡中 $f(x)$ 和 $\left\{g_j(x)\right\}_{j=1}^n$ 梯度之间的关系:

$$\nabla f(x)=\sum_{j=1}^{n}\lambda_j\nabla g_j(x)$$

- 对任意 $j=1,2,...,n$, 有 $\partial L(x,\lambda)/\partial\lambda_j=0\Longrightarrow g_j(x)=0$

一共有 $k+n$ 个一阶条件.

命题: 若原最优化问题的目标函数 $f(x)$ 和约束函数 $\left\{g_j(x)\right\}_{j=1}^n$ 满足适当的要求 (光滑性, 凹凸性等), 则关于 $\{x_i\}$ 和 $\left\{\lambda_j\right\}$ 的 $k+n$ 个一阶条件是必要的.

- 说明: 上述前提条件的表述较为马虎. 在这门课中, 除非特别说明, 我们一般不用验证这些条件, 直接默认一阶条件的必要性成立即可.

4. 例: 拉格朗日法

求解如下最优化问题

- 目标函数: $f(x_1,x_2)=x_1+x_2$
- 约束: $g(x_1,x_2)=x_1^2+x_2^2-1=0$

  $L(x,\lambda)=x_1+x_2-\lambda(x_1^2+x_2^2-1)$

一阶条件

- 关于 x_1 的一阶条件: $1-2\lambda x_1=0$
- 关于 x_2 的一阶条件: $1-2\lambda x_2=0$
- 关于 λ 的一阶条件: $x_1^2+x_2^2-1=0$

$$\Longrightarrow x_1=x_2=\frac{\sqrt{2}}{2}\quad\text{或}\quad x_1=x_2=-\frac{\sqrt{2}}{2}$$

5. 从等式约束到不等式约束

考虑如下包含 n 个**不等式约束**的最优化问题 $\mathcal{M}$:

$$\begin{aligned}\max_{x_1,x_2,...,x_k}\ & f(x_1,x_2,...,x_k)\\ \text{subject to}\quad & g_1(x_1,x_2,...,x_k)\le 0\\ & g_2(x_1,x_2,...,x_k)\le 0\\ & \vdots\\ & g_n(x_1,x_2,...,x_k)\le 0\end{aligned}$$

关于不等式约束形式的两点补充说明:

1. 所有不等式约束的不等号方向, 都是约束的左边小于等于 0.
   - 这只是形式上的要求. 如果出现了 $g(x)\ge 0$, 将它等价改写为 $-g(x)\le 0$ 即可.
2. 不等式约束的情形比等式约束的情形**更一般**:
   - 对任意等式约束 $g(x)=0$, 都可以把它等价地描述为两个不等式约束: $g(x)\le 0$ 且 $-g(x)\le 0$.

对问题 $\mathcal{M}$ 做出如下假设:

1. 目标函数 $f(x_1,x_2,...,x_k)$ 是凹的,

2. 每个约束函数 $g_i(x_1, x_2, ..., x_k)$ $(i = 1, ..., n)$ 都是凸的,
3. 问题 $\mathcal{M}$ 的可行域是非空的, 并且该问题有解.

说明: **可行域**为所有满足 n 个不等式约束的 $x \in \mathbb{R}^k$ 所构成的集合.

- 我们用字母 D 表示 n 个不等式约束构成的可行域. 若 $x \in D$, 则称 x 是**可行的**.
- 问题 $\mathcal{M}$ 可等价地描述为 $\max_{x \in D} f(x)$.

6. 可行域的例子

1. 圆盘

$$D = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1\}$$

2. 三角形区域

$$D = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 + x_2 \leq 1, x_1 \geq 0, x_2 \geq 0\}$$

3. 三维单纯形 (3-simplex)

$$D = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 + x_2 + x_3 \leq 1, x_1 \geq 0, x_2 \geq 0, x_3 \geq 0\}$$

7. 拉格朗日函数

对于问题 $\mathcal{M}$, 其拉格朗日函数的形式如下:

$$L(x, \lambda) = f(x_1, x_2, ..., x_k) - \sum_{j=1}^{n} \lambda_j g_j(x_1, x_2, ..., x_k)$$

- 不同于等式约束的情形, 问题 $\mathcal{M}$ 的拉格朗日乘子 $\lambda_1, ..., \lambda_n$ 都是**非负的**
  - 乘子是非负的 $\Longrightarrow$ 对任意可行的 $x \in D$, 拉格朗日函数的值 $L(x, \lambda)$ 总是大于等于 $f(x)$.
- 每个乘子 λ_j 对应约束 $g_j(x) \leq 0$.
  - 问题 $\mathcal{M}$ 的拉格朗日乘子还有一个经济学名字——影子价格. 影子价格这个叫法在一般均衡模型中很常用.

8. 对偶

问题 $\mathcal{M}$ 的对偶函数: $d(\lambda) \equiv \max_x L(x, \lambda)$.

- **注意:** 定义对偶 $d(\lambda)$ 时的最优化计算 $(\max_x)$ 里不要求 x 是可行的.

由对偶 $d(\lambda)$ 的定义可知:

$$d(\lambda) \geq L(x, \lambda) \quad \text{对任意} \quad x \in \mathbb{R}^k \quad \text{和} \quad \lambda \geq 0 \quad \text{成立}.$$

进一步可知,

$$d(\lambda) \geq f(x) \quad \text{对任意} \quad x \in D \quad \text{和} \quad \lambda \geq 0 \quad \text{成立}.$$

因此, 对偶 $d(\lambda)$ 是问题 $\mathcal{M}$ 中目标函数的上界.

对偶性的思想: 可以通过最小化上界 (即 $\min_{\lambda \geq 0} d(\lambda)$) 来求解原问题 $\mathcal{M}$.

9. Kuhn-Tucker 定理

定理 (Kuhn-Tucker). 若存在 $x^* = (x_1^*, ..., x_k^*)$ 和 $\lambda^* = (\lambda_1^*, ..., \lambda_n^*)$ 使得

$$\min_{\lambda\geq 0} d(\lambda) = \min_{\lambda\geq 0}\max_{x} L(x,\lambda) = L(x^*,\lambda^*)$$

则 x^* 为原最优化问题的解.

重要的内容说两次: 定义对偶 $d(\lambda)$ 时没有要求 x 必须落在可行域内

- 不过, Kuhn-Tucker 定理保证了问题 $\mathcal{M}$ 的解, 可通过求解如下 (几乎无约束) 最优化问题得到:

$$\min_{\lambda\geq 0}\max_{x} L(x,\lambda)$$

- **直观解释:** 在 $\min_\lambda$ 的过程中, 乘子 λ 实际上是在扮演一个"惩罚者". 如果 x 试图违反约束 (即 $g_j(x) > 0$), 惩罚者通过最小化 $-\lambda_j g(x)$ 会把目标函数拉到负无穷 (严格论述见 Kuhn-Tucker 定理的证明)

10. 互补松弛条件

1. 如果 $\lambda_j^* > 0$, 则约束 $g_j(x)$ 一定是紧的 (即 $g_j(x^*) = 0$).
2. 如果 $\lambda_j^* = 0$, 则约束 $g_j(x)$ 可以是不紧的 (即 $g_j(x^*) < 0$).

- 上面两点通常被归纳为所谓的**互补松弛条件** $\lambda_j^* g_j(x^*) = 0$

11. Kuhn-Tucker 条件

实际应用中, 我们通常不需要每次都去计算对偶 $d(\lambda)$ 或鞍点 (x^*,λ^*), 而是直接检查一组被称为 **Kuhn-Tucker 条件**的一阶必要条件.

- 这组一阶条件是 Kuhn-Tucker 定理的直接推论.

历史趣闻:

- Kuhn-Tucker 条件由运筹学研究者 Kuhn 和 Tucker 于 1951 年分别独立提出. 在那个年代, 许多研究者都在探索如何求解一般的不等式约束优化问题.
- 大约三十年后, 人们发现几乎完全相同的一阶必要条件, 早在 1939 年就已经由 Karush 在他的硕士论文中提出. 当时的 Karush 还只是二十二岁的年轻学生.
- 由于 Karush 的发现要早于 Kuhn 和 Tucker, 今天的很多资料称 Kuhn-Tucker 条件为 KKT 条件 (Karush–Kuhn–Tucker).

针对优化问题 $\max f(x)$ 和约束 $g_j(x) \leq 0,\ j = 1, ..., n$.
写出拉格朗日函数 $L(x,\lambda) = f(x) - \sum \lambda_j g_j(x)$ 后, 其 Kuhn-Tucker 条件由以下四个部分组成:

1. 一阶条件:

$$\frac{\partial f(x^*)}{\partial x_i} = \sum_{j=1}^{n} \lambda_j^* \frac{\partial g_j(x^*)}{\partial x_i} \quad (i = 1, ..., k)$$

   - 说明: 拉格朗日函数对原变量 x 的梯度在最优解处必须为零. 这意味着目标函数的梯度必须是约束函数梯度的线性组合.

2. 可行性: x^* 必须是可行的, 即满足所有不等式约束:

$$g_j(x^*) \leq 0, \quad j = 1, ..., n$$

3. 拉格朗日乘子非负条件 (也叫对偶可行性条件):

$$\lambda \geq 0$$

4. 互补松弛条件:

$$\lambda_j^* g_j(x^*) = 0 \quad (j = 1, ..., n)$$

12. Kuhn-Tucker 条件: 算例

目标函数

$$\max f(x_1, x_2, x_3) = x_1 + x_2 + x_3$$

不等式约束:

$$g(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + 3x_3^2 - 1 \leq 0.$$

- 几何直观: 可行域为椭球内部(含边界), 目标函数是线性的, 故最优解必在边界 $g(x) = 0$ 上.

拉格朗日函数:

$$L(x_1, x_2, x_3, \lambda) = x_1 + x_2 + x_3 - \lambda(x_1^2 + 2x_2^2 + 3x_3^2 - 1),$$

KKT 条件:

- 一阶条件 (梯度为零):

$$\frac{\partial L}{\partial x_1} = 0 \quad \Rightarrow \quad 2\lambda x_1 = 1,$$

$$\frac{\partial L}{\partial x_2} = 0 \quad \Rightarrow \quad 4\lambda x_2 = 1,$$

$$\frac{\partial L}{\partial x_3} = 0 \quad \Rightarrow \quad 6\lambda x_3 = 1.$$

- 可行性:

$$x_1^2 + 2x_2^2 + 3x_3^2 - 1 \leq 0.$$

- 乘子非负: $\lambda \geq 0$
- 互补松弛:

$$\lambda(x_1^2 + 2x_2^2 + 3x_3^2 - 1) = 0.$$

由梯度为零可知 $\lambda \neq 0$, 因此 $\lambda > 0$ 且 $x_1^2 + 2x_2^2 + 3x_3^2 - 1 = 0$ (边界解).

将 x_1, x_2, x_3 的表达式代入边界方程:

$$\left(\frac{1}{2\lambda}\right)^2 + 2\left(\frac{1}{4\lambda}\right)^2 + 3\left(\frac{1}{6\lambda}\right)^2 = 1.$$

$$\Rightarrow \quad \lambda = \sqrt{\frac{11}{24}}.$$

代回得:

$$x_1^* = \frac{1}{2\lambda} = \sqrt{\frac{6}{11}}, x_2^* = \frac{1}{4\lambda} = \sqrt{\frac{3}{22}}, x_3^* = \frac{1}{6\lambda} = \sqrt{\frac{2}{33}}.$$

最优值:

$$f(x^*) = -\sqrt{\frac{11}{6}}.$$

13. 拉格朗日法: 无穷维情形

本讲介绍的拉格朗日法 (以及 Kuhn-Tucker 条件) 只考虑了**有限维**情形的最优化问题.

- 最优化问题的控制变量只有 k 个, 且 k 为有限数.

经济学模型中的最优化问题经常是**无穷维**的.

- 比如, 求解公司的最优工资方案 $w(q)$ 就是典型的无穷维最优化问题: 函数 $w(q)$ 可以看成是一个无穷维向量, 它规定了每个产出 q 对应的工资水平 $w(q)$
- 线性工资假设下的工资方案不是无穷维的, 因为它只包含两个参数: w_0 和 b.

对于无穷维问题, 我们仍然可以通过构造拉格朗日函数, 并仿照有限维情形写出一阶条件 (或 Kuhn-Tucker 条件) 来求解.

- 同学们不必过于纠结此时拉格朗日法的严格性. 对于无穷维最优化问题, 其一阶条件必要性的严格证明, 不可避免地会涉及到**无穷维分析**的基本语言和工具.

尽管无穷维分析本身是一个很有意思的话题 (它在求解最优合同问题时也会很自然地出现), 我们暂时选择回避关于它的讨论.

- 这门课很短, 但同学们未来的人生还很长, 总会有机会去学习它的 :)