# 经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

## 全书合并版

Chapter 0–31 + Appendix A–C

| | |
|---|---|
| 正文 | 导论、多元积分、向量分析、投影与 Fourier、动态系统、优化、对偶、固定点、动态优化、概率统计 |
| 附录 | 全书公式大全、综合习题册、论文阅读工具包 |
| 定位 | 为经济学、金融学、计量经济学和动态模型定制的数学分析桥梁课程 |

math4econ personalized notes
2026

# 合并版说明

本合并版把已完成的正文 Chapter 0–31 与 Appendix A–C 按学习顺序合并为一册。各章内部仍保留原章节目录、页眉和页码；本文件另加入书签，方便在 PDF 阅读器中按章节跳转。

这套讲义的主线不是“公共高数压缩版”, 而是把数学分析、线性代数、优化、动态系统、概率统计重组为经济金融模型语言。学习时建议把每章当作一个工具模块：先看直觉和 worked examples, 再做 A–E 组练习，最后回到综合附录复习公式与跨章题。

# 全书结构速览

| 部分 | 内容 |
|---|---|
| Part I | Chapter 0：导论，说明为什么经济金融需要这套数学。 |
| Part II | Chapter 1–4：二重/三重积分、Jacobian、含参积分与 Leibniz Rule。 |
| Part III | Chapter 5–7：梯度、全微分、曲线积分、Green/Gauss/Stokes 直觉。 |
| Part IV | Chapter 8–9：正交投影、最小二乘、Fourier 与频域直觉。 |
| Part V | Chapter 10–13：ODE、差分方程、线性系统、非线性动态。 |
| Part VI | Chapter 14–20：多元优化、Lagrange/KKT、凸性、比较静态、对偶、固定点。 |
| Part VII | Chapter 21–25：动态优化、Bellman、最优控制、HJB、数值方法。 |
| Part VIII | Chapter 26–31：概率、条件期望、极限定理、多元正态、模拟推断、高级课程地图。 |
| Appendices | Appendix A 公式大全，Appendix B 综合习题，Appendix C 论文阅读工具包。 |

# 章节目录


| 编号 | 标题 |
| --- | --- |
| Chapter 0 | 导论：为什么经济金融需要这套数学 |
| Chapter 1 | 二重积分与区域累加 |
| Chapter 2 | 三重积分与多维分布 |
| Chapter 3 | 变量替换与 Jacobian |
| Chapter 4 | 含参积分、Leibniz Rule 与积分下求导 |
| Chapter 5 | 向量场、梯度与全微分 |
| Chapter 6 | 曲线积分与路径依赖 |
| Chapter 7 | Green、Gauss、Stokes 的直观版 |
| Chapter 8 | 正交、投影与最小二乘 |
| Chapter 9 | 傅里叶级数与频域直觉 |
| Chapter 10 | 一阶 ODE 与连续时间动态 |
| Chapter 11 | 差分方程与离散时间经济学 |
| Chapter 12 | 线性系统、特征值与相图 |
| Chapter 13 | 非线性动态系统初步 |
| Chapter 14 | 多元优化与 Hessian |
| Chapter 15 | Lagrange 乘子与等式约束 |
| Chapter 16 | KKT 条件与不等式约束 |
| Chapter 17 | 凸性、凹性与经济最优化 |
| Chapter 18 | 隐函数定理、包络定理与比较静态 |
| Chapter 19 | 对偶性与经济函数 |
| Chapter 20 | 不动点定理与均衡存在 |
| Chapter 21 | 动态优化基础与 Euler 方程 |
| Chapter 22 | 动态规划与 Bellman 方程 |
| Chapter 23 | 连续时间最优控制 |
| Chapter 24 | HJB 方程与连续时间动态规划 |
| Chapter 25 | 数值方法与计算经济学入门 |
| Chapter 26 | 概率论基础与随机变量 |
| Chapter 27 | 条件期望与回归的数学基础 |
| Chapter 28 | 收敛概念与极限定理 |

| | |
|---|---|
| Chapter 29 | 多元正态、二次型与金融风险 |
| Chapter 30 | Monte Carlo、Bootstrap 与模拟推断 |
| Chapter 31 | 通向高级课程的数学地图 |
| Appendix A | 全书公式大全与速查手册 |
| Appendix B | 全书综合习题册与详细答案 |
| Appendix C | 论文阅读工具包与建模速查表 |

# 经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

---

# Chapter 0

## 导论：为什么经济金融需要这套数学

---

这不是公共高数的压缩版，也不是数学系数分三的替代品。
它的目标是把你已经学过的数学分析，重组为经济学、金融学、计量经济学和动态模型真正会使用的语言。

math4econ personalized notes
2026

# 目录

**Chapter 0 总览** **3**

**1 为什么不是继续机械学公共高数或数分三** **3**
1.1 三种微积分传统 . . . . . 3
1.2 你现在真正缺的是什么 . . . . . 4

**2 本讲义的定位** **4**
2.1 本讲义要解决的核心问题 . . . . . 4

**3 经济金融数学的六个核心动词** **5**
3.1 边际：导数是经济学最基本的语言 . . . . . 5
3.2 聚合：积分是异质性世界的总和 . . . . . 6
3.3 选择：优化是经济学的骨架 . . . . . 6
3.4 演化：ODE 和差分方程描述时间 . . . . . 7
3.5 均衡：固定点是“大家互相影响后自洽”的数学表达 . . . . . 7
3.6 推断：概率和统计让模型接触数据 . . . . . 7

**4 本书的 26 章地图** **8**
4.1 真正的优先级 . . . . . 9

**5 如何读本讲义：不要把它当普通教材** **9**
5.1 每章固定结构 . . . . . 9
5.2 为什么要先看 worked examples . . . . . 10
5.3 为什么习题要混合编排 . . . . . 10

**6 从经济问题到数学工具：一张翻译表** **10**

**7 Worked examples：如何把问题翻译成数学** **11**

**8 你后续应该怎么学：一个操作协议** **13**
8.1 每章学习的五步 . . . . . 13
8.2 如何做笔记 . . . . . 13

**9 常见误区** **13**

**10 练习题** **14**
练习题 A. 基础概念题 . . . . . 14
练习题 B. 概念判断题 . . . . . 14
练习题 C. 工具识别题 . . . . . 14

练习题 D. 经济金融建模题 . . . . . 15
练习题 E. 挑战/解释题 . . . . . 15

**11 练习题参考答案** **16**
A 组答案 . . . . . 16
B 组答案 . . . . . 16
C 组答案 . . . . . 17
D 组答案 . . . . . 17
E 组答案 . . . . . 18

**12 本章压缩总结** **19**

**下一章预告** **19**

# Chapter 0 总览

**本章目标**

学完本章，你不需要掌握新的计算技巧；你需要获得一张完整地图：

1. 明白为什么你现在不应该机械补“公共高数”或“数学系数分三”，而应该学习一套经济金融导向的数学分析课程；
2. 知道微积分、线性代数、优化、动态系统、概率统计在经济金融模型中的位置；
3. 能把一个经济金融问题初步翻译成数学对象：边际、聚合、选择、动态、均衡、风险、估计；
4. 知道后面 26 章应该如何学习、如何做题、如何复习、如何连接高级微观、宏观、计量和金融。

**直觉**

本章的核心判断可以压缩成一句话：

> 数学不是额外负担，而是经济金融模型的语法。

你后面要学的不是“更多公式”，而是学会用数学表达经济对象：消费者如何选择，厂商如何优化，市场如何均衡，资产如何定价，参数如何估计，系统如何动态演化。

## 1 为什么不是继续机械学公共高数或数分三

你现在最容易走错的路线有两条。第一条是退回公共高数，从二重积分、三重积分、曲线积分、ODE 开始，按普通理工科课表一章一章刷题。第二条是继续沿数学系数学分析的路径，追求更抽象、更严谨、更完整的实分析训练。

这两条路都有价值，但都不是此刻的最优路线。原因不是它们不好，而是它们的目标函数和你的目标函数不同。

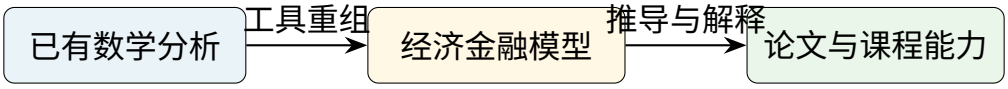
### 1.1 三种微积分传统

| 传统 | 核心目标 | 优点 | 对你的不足 |
| --- | --- | --- | --- |
| 工程计算型微积分 | 会算导数、积分、ODE、向量分析，服务物理和工程模型 | 计算训练充分，例题多，题型清楚 | 证明弱；经济解释弱；大量物理场论和几何技巧对经济金融短期收益不高 |

| | | | |
|---|---|---|---|
| 数学分析型训练 | 严格理解极限、连续、可微、积分、收敛和证明 | 训练严谨性；能理解定理条件；适合继续数学理论 | 经济建模连接弱；很多抽象细节短期不能转化为模型能力 |
| 经济金融数学型课程 | 把微积分、线代、优化、动态、概率重组为经济金融模型语言 | 工具和模型直接连接；适合高级微观、宏观、计量、金融 | 需要主动补足计算训练和必要的严谨条件 |

### 1.2 你现在真正缺的是什么

你已经接触过数学分析，所以你并不是“完全没有严谨性”。你真正缺的是一种转换能力：




比如你可能知道极限、连续、导数、积分的严格定义，但看到下面这些对象时仍然会卡住：

$$\mathbb{E}[u(W+X)], \qquad \dot{k} = sf(k) - (n+\delta)k, \qquad \max_{x} f(x,\theta), \qquad F(x,\theta) = 0, \qquad x \in \Phi(x).$$

这些式子不是简单的“数学题”，而是经济金融模型的基本句子。你需要做的，是把数分里学过的工具转成模型语言。

## 2 本讲义的定位

本讲义暂定名为：

《经济金融数学分析补课讲义：从多元积分到动态优化》

英文副标题是：

A Personalized Mathematical Methods Course for Economics and Finance.

它不是一本“高等数学讲义”，也不是一本“实分析讲义”，而是一本桥梁教材。它连接的是：

$$\text{数学分析基础} \longrightarrow \text{经济金融数学工具} \longrightarrow \text{高级课程与论文阅读}.$$

### 2.1 本讲义要解决的核心问题

这套讲义要解决五个问题。

1. **计算问题：**你要会算多重积分、变量替换、ODE、差分方程、最优化、条件期望。
2. **语言问题：**你要知道经济学中“边际、聚合、选择、动态、均衡、风险、识别”分别对应什么数学结构。

3. **严谨问题：**你要知道什么时候可以求导、什么时候最优解存在、什么时候局部最优是全局最优、什么时候可以交换极限和积分。
4. **建模问题：**你要能从一段经济语言写出数学模型，而不是只会解别人已经写好的题。
5. **迁移问题：**你要把这些工具迁移到高级微观、宏观、计量、金融、机器学习和论文阅读中。

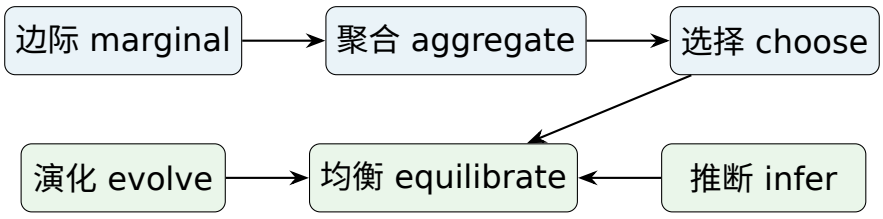
**方法**

本讲义每个概念都尽量使用三层讲法：

1. **直觉层：**一句话说明它在干什么；
2. **计算层：**用例题说明怎么用；
3. **严谨层：**说明定义、条件、定理和可能失败的地方。

# 3 经济金融数学的六个核心动词

经济金融模型看起来复杂，但很多数学工具可以按照六个“动词”来理解。




## 3.1 边际：导数是经济学最基本的语言

在经济学里，导数通常不是单纯的斜率，而是“一个变量小幅变化时，目标对象如何变化”。

**边际效应**

若某个经济结果可以写成函数

$$Y = f(x),$$

那么

$$f'(x)$$

表示在当前位置附近，$x$ 增加一个很小单位时，$Y$ 的局部变化率。经济学中常见的 marginal cost、marginal utility、marginal product、marginal effect 都是这个思想。

例如，厂商成本为 $C(q)$，边际成本是

$$MC(q) = C'(q).$$

消费者效用为 $u(x_1, x_2)$，边际效用向量是

$$\nabla u(x_1, x_2) = \begin{pmatrix} \frac{\partial u}{\partial x_1} \\ \frac{\partial u}{\partial x_2} \end{pmatrix}.$$

这就是为什么后面要学梯度、全微分、Hessian 和隐函数定理。经济学不是只问“函数值是多少”，而是不断问“一个变量变动时，另一个对象怎么变”。

## 3.2 聚合：积分是异质性世界的总和

一元积分最容易被理解成面积，但在经济金融里，积分更常见的含义是聚合：把不同个体、不同状态、不同收入、不同风险下的贡献加总。

例如，如果收入 $y$ 的密度是 $f(y)$，税收函数是 $T(y)$，总税收可以写成

$$\int T(y) f(y)\, \mathrm{d}y.$$

如果随机收益 $R$ 的密度是 $f(r)$，投资者的期望效用是

$$\mathbb{E}[u(W(1+R))] = \int u(W(1+r)) f(r)\, \mathrm{d}r.$$

后面的多重积分、Jacobian、含参积分、期望、条件期望，都在服务这个思想：

| 经济世界有异质性和不确定性，所以需要积分。 |
|---|

## 3.3 选择：优化是经济学的骨架

经济学最核心的建模动作是：一个主体在约束下做选择。

消费者：

$$\max_{x \geq 0} u(x) \quad \text{s.t.} \quad p \cdot x \leq w.$$

厂商：

$$\max_{y \in Y} p \cdot y.$$

投资者：

$$\max_{w} \ \mu_p - \frac{\gamma}{2}\sigma_p^2 \quad \text{s.t.} \quad \mathbf{1}'w = 1.$$

计量中的估计量：

$$\hat{\beta} = \arg\min_{b} \sum_{i=1}^{n} (y_i - x_i' b)^2.$$

它们看起来属于不同课程，但数学骨架是同一个：目标函数、选择变量、约束、最优条件、比较静态。

## 3.4 演化：ODE 和差分方程描述时间

经济变量不是只在一个时点存在。资本积累、债务、通胀、资产价格、技术扩散、财富分布都会随时间变化。

连续时间模型通常写成：

$$\dot{x}(t) = F(x(t), t, \theta).$$

离散时间模型通常写成：

$$x_{t+1} = G(x_t, \theta).$$

例如 Solow 模型的核心动态是

$$\dot{k} = sf(k) - (n + \delta)k.$$

资产价格递推可能写成

$$p_t = \frac{1}{1+r}\mathbb{E}_t[p_{t+1} + d_{t+1}].$$

所以 ODE 和差分方程不是额外的数学章节，而是宏观经济学和金融经济学的基本语法。

## 3.5 均衡：固定点是“大家互相影响后自洽”的数学表达

均衡不是“供给曲线和需求曲线交一下”那么简单。一般地，均衡是一个自洽状态：每个人根据环境最优选择，而这些选择合在一起又生成同一个环境。

非常抽象地说，均衡常常可以写成：

$$x \in \Phi(x).$$

或者在函数情形下写成：

$$x = F(x).$$

这就是固定点。后面学 Brouwer、Kakutani、对应和上半连续，不是为了拓扑本身，而是为了理解一般均衡和 Nash 均衡为什么存在。

## 3.6 推断：概率和统计让模型接触数据

经济金融不仅要写模型，还要面对数据。数据有噪声，个体有异质性，状态有随机性。于是我们需要概率、条件期望、极限定理、回归和渐近理论。

例如，计量经济学里的核心条件之一是：

$$\mathbb{E}[\varepsilon \mid X] = 0.$$

它不是一个“形式假设”，而是在说：给定解释变量后，误差项不再含有系统性可预测部分。

OLS 的总体形式可以写成：

$$\beta = (\mathbb{E}[X'X])^{-1}\mathbb{E}[X'Y].$$

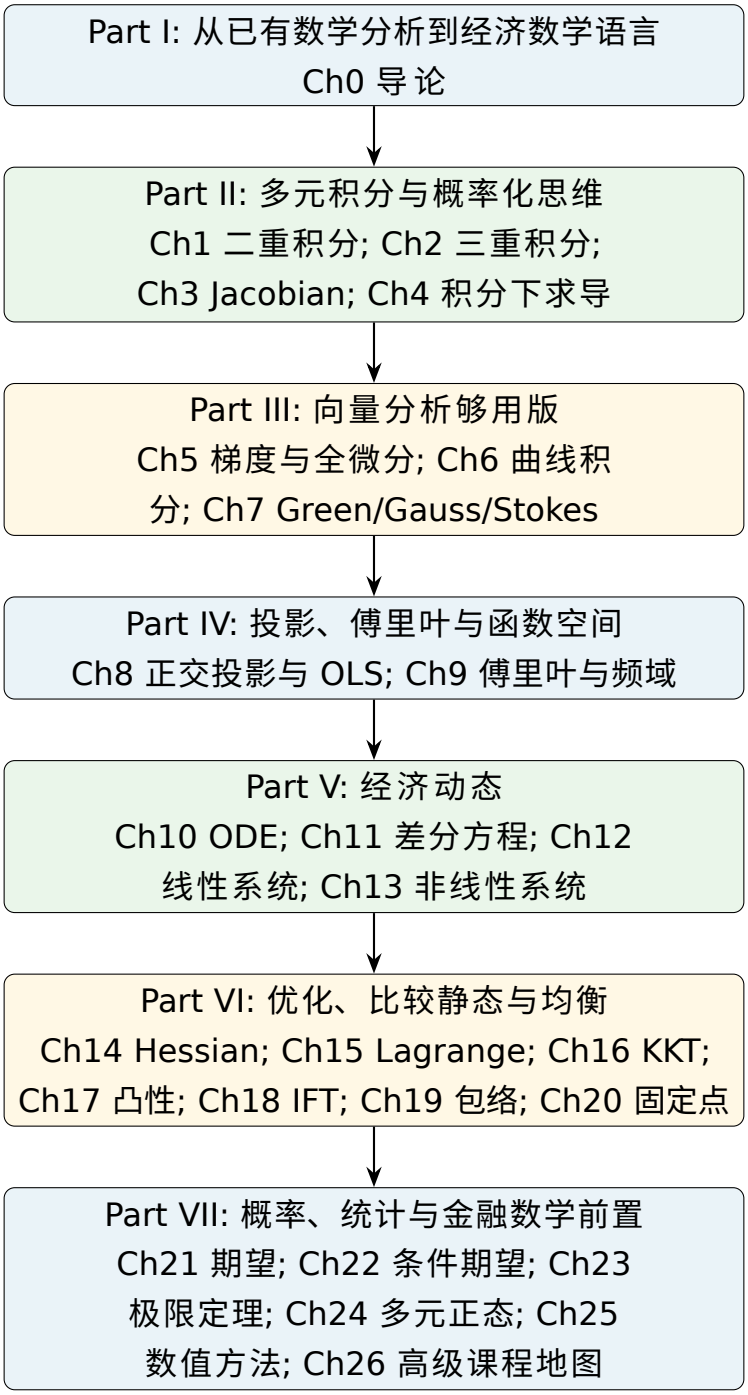
金融里的组合方差写成：

$$\sigma_p^2 = w'\Sigma w.$$

这些都要求你把积分、线性代数、概率和优化合在一起理解。

# 4 本书的 26 章地图

本书分为七个部分。为了后续不迷路，你应该把它看成一条主线，而不是 26 个散点。



### 4.1 真正的优先级

完整目录是 26 章，但你的实际学习不应该平均用力。最优推进顺序是：

**方法**

多重积分 → Jacobian → 积分下求导 → ODE/差分方程
→ 线性系统 → 优化/KKT → 隐函数/包络
→ 概率与条件期望 → 傅里叶/投影 → 金融与计量应用

**方法**

第一优先级：Ch1–Ch4, Ch10–Ch11, Ch14–Ch19。
第二优先级：Ch8–Ch9, Ch12–Ch13, Ch21–Ch24。
第三优先级：Ch5–Ch7, Ch25–Ch26。
这不是说第三优先级不重要，而是说它们对你当前经济金融学习的边际收益较低，可以学够用而不必深挖成数学系课程。

## 5 如何读本讲义：不要把它当普通教材

这本讲义的每章都会固定采用同一套结构。你读的时候也要按固定动作走。

### 5.1 每章固定结构

每章包括十个部分：

1. **本章为什么重要：**先说明这个工具在经济金融中的位置。
2. **最小前置知识：**告诉你需要回忆哪些旧知识。
3. **核心直觉：**用语言、图像、简单例子说明它在干什么。
4. **正式定义与定理：**给出必要的数学表述。
5. **计算方法：**训练标准套路、常见题型和易错点。
6. **经济金融应用：**至少一个经济学应用，一个金融或计量应用。
7. **Worked examples：**完整样板题，不跳步。
8. **练习题：**基础计算、概念判断、混合题、建模题、挑战题。
9. **详细解答：**不仅给答案，还解释为什么这样想。
10. **总结图谱：**用“概念 - 公式 - 应用 - 常见错误”收束。

## 5.2 为什么要先看 worked examples

很多数学学习失败，不是因为学生不聪明，而是因为一上来就被要求独立解决过复杂的问题。对于新工具，最好的顺序不是“先自己硬做”，而是：

**方法**

worked example → guided exercise → independent exercise
→ mixed review → model application

这也是本讲义的习题设计原则。你先看完整样板，再做半引导题，再做独立题，最后把不同章节混在一起训练判断能力。

## 5.3 为什么习题要混合编排

如果二重积分那章的题全部都是“改积分次序”，你会形成题型依赖：看到这一章就知道该怎么做，但换到计量、金融或宏观模型里就不会判断。真正的能力是看到问题后知道该用哪种工具。

所以本讲义每章习题分五档：

**A. 基础计算**

刚学完马上练，巩固工具。

**B. 概念判断**

判断条件、含义和错误说法。

**C. 混合题**

和前几章内容交错。

**D. 经济金融建模题**

把文字问题翻译成数学模型。

**E. 挑战/证明题**

接近研究生课程或论文阅读需要。

# 6 从经济问题到数学工具：一张翻译表

下面这张表非常重要。后面你遇到一个经济金融问题时，第一步不是马上算，而是先识别它属于哪类数学动作。

| 经济金融问题 | 数学对象 | 典型表达 | 对应章节 |
| --- | --- | --- | --- |
| 边际成本、边际效用、政策边际效应 | 导数、偏导、梯度、全微分 | $MC = C'(q)$, $dU = \nabla U \cdot dx$ | Ch5, Ch14, Ch18 |

| | | | |
|---|---|---|---|
| 异质性个体加总、期望效用、联合分布 | 多重积分、密度、期望 | $\mathbb{E}[g(X,Y)] = \iint gf$ | Ch1–Ch4, Ch21 |
| 消费者选择、厂商选择、投资组合选择 | 最优化、约束、KKT | $\max_x f(x)$, $\lambda g(x) = 0$ | Ch14–Ch17 |
| 价格、收入、税率变化下的需求变化 | 隐函数、比较静态、包络 | $dx/d\theta = -F_x^{-1}F_\theta$ | Ch18–Ch19 |
| 资本积累、债务演化、资产价格递推 | ODE、差分方程、动态系统 | $\dot{x} = F(x)$, $x_{t+1} = G(x_t)$ | Ch10–Ch13 |
| 一般均衡、Nash 均衡、匹配稳定性 | 固定点、对应 | $x \in \Phi(x)$ | Ch20 |
| 回归、估计、渐近推断 | 条件期望、投影、极限定理 | $\mathbb{E}[\varepsilon \mid X] = 0$ | Ch8, Ch21–Ch23 |
| 资产风险、因子模型、组合方差 | 矩阵、二次型、多元正态 | $\sigma_p^2 = w'\Sigma w$ | Ch24 |

# 7 Worked examples：如何把问题翻译成数学

**Worked Example 0.1：判断一个问题需要什么数学工具**

**问题**。政府提高所得税税率 $\tau$，不同收入 $y$ 的家庭受到不同影响。假设收入密度为 $f(y)$，家庭效用为 $u((1-\tau)y)$。问社会平均福利如何随税率变化?

**第一步：识别对象**。这里有异质性收入 $y$，所以需要积分聚合。税率 $\tau$ 是参数，所以还需要含参积分和积分下求导。

**第二步：写出平均福利**。

$$W(\tau) = \int u((1-\tau)y) f(y)\,\mathrm{d}y.$$

**第三步：对税率求边际影响**。若可以交换求导与积分，则

$$W'(\tau) = \int u'((1-\tau)y)(-y) f(y)\,\mathrm{d}y.$$

**解释**。每个收入类型的边际福利变化是 $-yu'((1-\tau)y)$，再用收入密度 $f(y)$ 加权平均。

**涉及章节**。 Ch1 二重/一重积分思想，Ch4 含参积分，Ch21 期望。

### Worked Example 0.2：从消费者问题看优化语言

**问题**。消费者选择两种商品 $x_1, x_2$，效用为 $u(x_1, x_2)$，价格为 $p_1, p_2$，财富为 $w$。如何写成数学问题?

**模型**。

$$\max_{x_1, x_2 \geq 0} u(x_1, x_2) \quad \text{s.t.} \quad p_1 x_1 + p_2 x_2 \leq w.$$

**如果内点解存在，**拉格朗日函数为

$$\mathcal{L}(x_1, x_2, \lambda) = u(x_1, x_2) + \lambda(w - p_1 x_1 - p_2 x_2).$$

一阶条件是

$$\frac{\partial u}{\partial x_1} = \lambda p_1, \qquad \frac{\partial u}{\partial x_2} = \lambda p_2, \qquad p_1 x_1 + p_2 x_2 = w.$$

因此

$$\frac{MU_1}{MU_2} = \frac{p_1}{p_2}.$$

**解释**。这不是死公式，而是说：最优点处，主观边际替代率等于市场价格比。你愿意交换的比例，必须等于市场允许你交换的比例。

**涉及章节**。 Ch14 多元优化，Ch15 Lagrange，Ch16 KKT，Ch19 包络定理。

### Worked Example 0.3：从动态方程看经济演化

**问题**。一个经济体的人均资本 $k(t)$ 满足

$$\dot{k} = sk^{\alpha} - (n + \delta)k, \qquad 0 < \alpha < 1.$$

这是什么类型的问题?

**识别**。这是连续时间一阶非线性 ODE。核心不是只求显式解，而是看稳态和稳定性。

**稳态**。稳态满足 $\dot{k} = 0$，所以

$$sk^{\alpha} = (n + \delta)k.$$

若 $k > 0$，则

$$k^* = \left(\frac{s}{n + \delta}\right)^{\frac{1}{1-\alpha}}.$$

**经济解释**。当 $k < k^*$ 时，投资项 $sk^{\alpha}$ 相对较大，资本上升；当 $k > k^*$ 时，折旧和人口稀释项 $(n + \delta)k$ 较大，资本下降。

**涉及章节**。 Ch10 ODE，Ch12 线性系统与相图，Ch13 非线性动态。

# 8 你后续应该怎么学：一个操作协议

## 8.1 每章学习的五步

你读每章时可以按下面五步走。

1. **先看本章为什么重要**。不要急着进入公式，先知道它服务什么模型。
2. **再看核心直觉**。用几何和经济语言理解它在干什么。
3. **然后看正式定义**。定义不是背诵材料，而是防止你在边界情况犯错。
4. **模仿 worked examples**。先复现样板题，再做 guided exercises。
5. **最后做 mixed review**。真正能力来自混合题，而不是同题型刷题。

## 8.2 如何做笔记

建议每章笔记固定分成四栏：

| 概念 | 数学表达 | 经济金融含义 | 常见错误 |
|---|---|---|---|
| 导数 | $f'(x)$ | 边际效应 | 忘记“局部”含义 |
| 积分 | $\int g(x)f(x)\,\mathrm{d}x$ | 加权平均、期望、聚合 | 把积分只理解成面积 |
| 优化 | $\arg\max f(x)$ | 选择行为 | 只写 FOC，不检查约束和二阶条件 |
| ODE | $\dot{x} = F(x)$ | 连续时间动态 | 只求解，不分析稳态和稳定性 |
| 固定点 | $x = F(x)$ | 均衡、自洽 | 把均衡理解成单一交点图 |
| 条件期望 | $\mathbb{E}[Y\|X]$ | 最优预测、回归基础 | 以为它只是一个数 |

# 9 常见误区

**常见误区**

1. **误区一：把经济金融数学理解成高数题型集合**。实际上，它是一套模型语言。会算只是第一层。
2. **误区二：认为严谨性和应用性冲突**。对经济学来说，很多严谨条件正是经济结论成立的边界，例如紧性保证最优解存在，凸性保证局部最优可以推出全局最优。

3. **误区三：只追求显式解**。经济动态和均衡问题中，经常没有漂亮显式解。稳态、稳定性、比较静态、数值解同样重要。
4. **误区四：学概率时只背分布**。对计量和金融来说，更重要的是条件期望、协方差结构、收敛概念和极限定理。
5. **误区五：忽视反例**。一个定理的假设去掉后会失败，这比会套定理更重要。

# 10 练习题

## A. 基础概念题

A1. 用一句话分别解释 derivative、integral、optimization、ODE、fixed point、conditional expectation 在经济金融中的含义。

A2. 判断下列对象分别属于哪类数学工具：

$$C'(q), \quad \mathbb{E}[u(W+X)], \quad \max_x u(x) \text{ s.t. } p\cdot x \le w, \quad x_{t+1} = 0.8x_t + 2, \quad p = D(p).$$

A3. 为什么“积分不只是面积”这句话对经济学很重要?

A4. 为什么“最优化不只是求导等于零”？至少给出两个原因。

## B. 概念判断题

判断正误，并说明理由。

B1. 只要会求导，就能解决所有消费者最优化问题。

B2. 如果一个经济变量随时间变化，那么它一定可以用 ODE 表示，而不需要差分方程。

B3. 固定点可以被理解为一种“自洽状态”。

B4. 条件期望 $\mathbb{E}[Y|X]$ 通常是一个关于 $X$ 的函数，而不只是一个常数。

B5. 凸性在经济学中重要，是因为它常常帮助我们从局部条件推出整体结论。

## C. 工具识别题

对下面每个问题，写出最可能需要的数学工具和对应章节。

C1. 给定两个随机收益 $R_1, R_2$ 的联合密度，计算 $\mathbb{E}[R_1R_2]$。

C2. 研究税率 $\tau$ 变化如何影响社会福利

$$W(\tau) = \int u((1-\tau)y) f(y)\,\mathrm{d}y.$$

C3. 求解投资组合问题

$$\min_{w} w'\Sigma w \quad \text{s.t.} \quad \mathbf{1}'w = 1.$$

C4. 分析资本动态

$$k_{t+1} = sf(k_t) + (1-\delta)k_t.$$

C5. 证明一个价格向量 $p$ 满足 $p = F(p)$ 的均衡存在。

## D. 经济金融建模题

D1. 一个家庭的收入为随机变量 $Y$，消费为 $c(Y) = a + bY$，效用为 $u(c) = \sqrt{c}$。写出期望效用表达式，并说明它需要什么数学工具。

D2. 一个厂商利润为 $\pi(q) = pq - C(q)$。写出最优产量的一阶条件，并解释其经济含义。

D3. 一个国家债务比率 $b(t)$ 满足 $\dot{b} = (r-g)b - d$，其中 $d$ 是 primary surplus ratio 的相反数。求稳态，并说明稳定性取决于什么。

D4. 一个回归模型写作 $Y = X\beta + \varepsilon$。如果总体最小二乘问题是

$$\min_{b} \mathbb{E}[(Y - Xb)^2],$$

写出一阶条件并解释为什么这和“投影”有关。

## E. 挑战/解释题

E1. 解释为什么“经济学需要 real analysis”不等于“经济学学生必须像数学系一样学习全部实分析细节”。

E2. 试说明：为什么一般均衡和 Nash 均衡都可以和 fixed point 联系起来?

E3. 选择一个你学过的中级微观模型，把它拆成 primitive、assumption、optimization/equilibrium、comparative statics 四个部分。

# 11 练习题参考答案

## A 组答案

**A1.** derivative 表示边际变化；integral 表示连续加总、加权平均或期望；optimization 表示在约束下选择最优行动；ODE 表示连续时间动态；fixed point 表示自洽均衡；conditional expectation 表示给定信息下的平均预测或投影。

**A2.**

$$C'(q):\text{导数/边际成本};\quad \mathbb{E}[u(W+X)]:\text{期望/积分};$$

$$\max_x u(x)\ \text{s.t.}\ p\cdot x\le w:\text{约束优化};$$

$$x_{t+1}=0.8x_t+2:\text{差分方程};\qquad p=D(p):\text{固定点/均衡}.$$

**A3.** 因为经济学经常面对连续类型的个体、收入分布、风险状态和随机变量。积分在这些场景下表示对不同状态或个体的加权聚合。例如平均福利、期望收益、总税收都可以写成积分形式。

**A4.** 第一，存在约束时最优点可能是角点，简单的 $f'(x)=0$ 不适用。第二，一阶条件通常只是必要条件，还需要二阶条件、凸性/凹性或全局性判断。第三，最优解是否存在也需要条件，如连续性和紧性。

## B 组答案

**B1.** 错。消费者问题可能有不等式约束、非负约束、角点解、不可微偏好或非凸偏好。只会求导不足以完整解决问题。

**B2.** 错。连续时间适合用 ODE，离散时间适合用差分方程。宏观和金融中大量模型是离散时间的，例如资产价格递推、AR(1)、离散 Solow 模型。

**B3.** 对。固定点 $x=F(x)$ 的含义是：由 $x$ 生成的反应或结果仍然是 $x$。这正是许多均衡的自洽结构。

**B4.** 对。$\mathbb{E}[Y|X]$ 通常随 $X$ 变化，因此是 $X$ 的函数。回归函数本质上就是条件期望或其线性近似。

**B5.** 对。比如最大化凹函数时，一阶条件不仅是局部条件，还能刻画全局最优。这就是凸分析在经济学中极其重要的原因。

## C 组答案

**C1.** 需要联合密度下的二重积分：

$$\mathbb{E}[R_1R_2] = \iint r_1r_2 f_{R_1,R_2}(r_1,r_2)\,\mathrm{d}r_1\,\mathrm{d}r_2.$$

对应 Ch1–Ch3 和 Ch21。

**C2.** 需要含参积分和积分下求导。若条件允许，

$$W'(\tau) = \int -yu'((1-\tau)y)f(y)\,\mathrm{d}y.$$

对应 Ch4。

**C3.** 这是带等式约束的二次优化问题，需要 Lagrange 乘子、Hessian/二次型和投资组合方差。对应 Ch14、Ch15、Ch24。

**C4.** 这是差分方程和动态系统问题。需要稳态、稳定性和相图。对应 Ch11–Ch13。

**C5.** 这是固定点问题。若 $F$ 满足适当条件，例如连续且把非空紧凸集映到自身，可以使用 Brouwer fixed point theorem。对应 Ch20。

## D 组答案

**D1.** 期望效用为

$$\mathbb{E}[u(c(Y))] = \mathbb{E}[\sqrt{a+bY}] = \int \sqrt{a+by} f_Y(y)\,\mathrm{d}y.$$

需要积分、期望、随机变量密度。如果 $Y$ 是离散变量，则相应地变成求和。

**D2.** 利润最大化：

$$\max_q \; pq - C(q).$$

若内点解存在，一阶条件为

$$p - C'(q) = 0, \quad \text{即} \quad p = MC(q).$$

经济含义是：最优产量处，额外卖一单位产品得到的价格等于额外生产一单位的边际成本。

**D3.** 稳态满足

$$0 = (r - g)b - d,$$

若 $r \neq g$，则

$$b^* = \frac{d}{r - g}.$$

稳定性取决于 $r - g$。如果 $r - g > 0$，偏离稳态后系统会远离稳态；如果 $r - g < 0$，系统倾向于回到稳态。这里还要结合 $d$ 的经济定义和政策反应函数理解。

**D4.** 总体目标为

$$Q(b) = \mathbb{E}[(Y - Xb)^2].$$

一阶条件为

$$\mathbb{E}[X(Y - X\beta)] = 0.$$

这说明残差 $Y - X\beta$ 与解释变量 $X$ 正交。因此 OLS 可以理解为把 $Y$ 投影到由 $X$ 张成的线性空间上。

## E 组答案

**E1.** 经济学需要 real analysis，是因为高级经济理论依赖连续性、紧性、凸性、收敛、极值存在、固定点和对应等概念。但经济学学生的目标不是研究所有分析学病态例子，而是掌握足以理解模型成立条件和定理边界的分析语言。因此最优路径是“够严谨的应用型分析”，而不是照搬数学系完整训练。

**E2.** 一般均衡中，价格决定个体需求，个体需求聚合后又决定市场是否出清；均衡价格必须在这个反馈系统中自洽。Nash 均衡中，每个玩家对他人策略作最优反应，而所有人的策略组合又必须等于这些最优反应的组合。因此两者都可以抽象成 $x \in \Phi(x)$ 或 $x = F(x)$ 的固定点结构。

**E3.** 以消费者问题为例。Primitive 包括消费集、价格、财富和偏好；assumption 包括偏好完备、传递、连续、局部非饱和、凸性等；optimization 是在预算约束下最大化效用；comparative statics 是研究价格或财富变化如何影响需求。这个拆解方式以后会反复用于高级微观、宏观和金融模型。

## 12 本章压缩总结

**本章压缩总结**

1. 你现在要学的不是公共高数压缩版，也不是数学系数分三复刻版，而是经济金融导向的数学分析桥梁课。
2. 经济金融数学的核心不是“公式更多”，而是六个动作：边际、聚合、选择、演化、均衡、推断。
3. 导数对应边际，积分对应聚合和期望，优化对应选择，ODE/差分方程对应动态，固定点对应均衡，条件期望对应预测和回归。
4. 后续 26 章分为七部分：导论、多元积分、向量分析、投影傅里叶、经济动态、优化均衡、概率金融前置。
5. 你实际应优先学习 Ch1–Ch4、Ch10–Ch11、Ch14–Ch19；不要平均用力。
6. 每章都要按“直觉 - 计算 - 严谨 - 应用 - 练习 - 总结”的方式学习。

## 下一章预告

Chapter 1 将正式进入多元积分。你要先把二重积分理解成“二维异质性世界中的累加”：不仅会算面积，更要会处理联合密度、总福利、平均税负、二维收入-消费分布，以及如何通过改变积分次序看清一个经济问题的结构。

# 经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

---

# Chapter 1

## 二重积分与区域累加

---

本章把“一元积分是沿线累加”推广为“二重积分是在区域上累加”。
对经济金融来说，二重积分不是孤立的计算技巧，而是异质性个体聚合、联合分布期望、福利总量、收入税负和二维风险因子的基本语言。

math4econ personalized notes
2026

# 目录

**Chapter 1 总览** **2**

**1 本章为什么重要** **2**
1.1 最小前置知识 . . . . . 3

**2 从一元积分到二重积分** **3**
2.1 一元积分：沿区间的累加 . . . . . 3
2.2 二重积分：在区域上的累加 . . . . . 4
2.3 几何图像：曲面下的体积 . . . . . 4

**3 矩形区域上的二重积分** **5**
3.1 Fubini 定理的直觉 . . . . . 5

**4 一般区域：Type I 与 Type II** **6**
4.1 Type I 区域：上下边界由 $y$ 表示 . . . . . 6
4.2 Type II 区域：左右边界由 $x$ 表示 . . . . . 7

**5 改变积分次序** **7**
5.1 改变次序的标准流程 . . . . . 8

**6 二重积分的基本解释** **10**
6.1 面积 . . . . . 10
6.2 总量 . . . . . 10
6.3 平均值 . . . . . 11

**7 二重积分与联合密度** **11**
7.1 期望就是加权平均 . . . . . 12
7.2 边际密度 . . . . . 13

**8 经济金融应用一：异质性消费者聚合** **13**

**9 经济金融应用二：收入分布上的平均税负** **14**

**10 金融应用：二维风险因子的期望收益** **15**

**11 计算策略与常见错误** **15**
11.1 计算策略 . . . . . 15
11.2 常见错误 . . . . . 16

**12 Worked Examples：完整样板题** **16**

**13 练习题** **19**
A 组：基础计算 . . . . . 19
B 组：概念判断 . . . . . 19
C 组：混合题 . . . . . 20
D 组：经济金融建模题 . . . . . 20
E 组：挑战/证明题 . . . . . 20

**14 练习题详细解答** **21**
A 组答案 . . . . . 21
B 组答案 . . . . . 22
C 组答案 . . . . . 22
D 组答案 . . . . . 23
E 组答案 . . . . . 23

**15 本章总结图谱** **25**
本章最重要的三句话 . . . . . 25
下一章连接 . . . . . 25

# Chapter 1 总览

**本章目标**

学完本章，你应该达到以下目标：

1. 能把一元积分的“沿区间累加”推广为二重积分的“在平面区域上累加”；
2. 能识别并画出 Type I 与 Type II 区域，熟练写出积分限；
3. 能根据区域图像改变积分次序；
4. 理解 Fubini 定理的经济含义：二维聚合可以按不同顺序进行；
5. 能把二重积分解释为面积、总量、平均值、联合密度下的期望；
6. 能用二重积分处理异质性消费者、收入分布、福利聚合和二维随机变量问题。

**直觉**

一元积分问的是：一条线上很多小段的贡献加起来是多少？

$$\int_a^b f(x)\,\mathrm{d}x.$$

二重积分问的是：一个区域里很多小块的贡献加起来是多少？

$$\iint_R f(x,y)\,\mathrm{d}A.$$

如果把经济体里的家庭按收入 $y$ 和财富 $a$ 排列在一个平面区域中，那么二重积分就是在问：所有类型家庭的消费、税负、效用或风险暴露，加总起来是多少。

## 1 本章为什么重要

在普通微积分课里，二重积分经常以“求体积”“求曲面下方体积”“换积分次序”的形式出现。这些当然重要，但对经济金融来说，二重积分更深层的意义是：**它是二维异质性、联合分布与聚合问题的基本语言**。

经济学中很多对象不是单个个体，而是一群人、一批企业、一组资产状态或一个二维分布。例如：

- 个体由收入 $y$ 和教育 $e$ 描述；
- 家庭由财富 $a$ 和生产率 $z$ 描述；
- 资产收益由市场因子 $m$ 和行业因子 $s$ 描述；
- 税负由收入 $y$ 和消费 $c$ 同时决定；

- 福利由类型 $(\theta_1, \theta_2)$ 的分布决定。

  如果某个经济量写作 $g(x, y)$，而类型 $(x, y)$ 在区域 $R$ 中分布，那么总量通常长得像：

$$\iint_R g(x, y)\,\mathrm{d}x\,\mathrm{d}y,$$

或者如果 $(X, Y)$ 有联合密度 $f_{X,Y}(x, y)$，期望就写作：

$$\mathbb{E}[g(X, Y)] = \iint_{\mathbb{R}^2} g(x, y) f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

所以本章不是“补一点高数题型”，而是把你带进一种概率化、分布化、聚合化的模型语言。

## 1.1 最小前置知识

你需要回忆以下内容：

1. 一元定积分 $\int_a^b f(x)\,\mathrm{d}x$ 的几何含义和累加含义；
2. 基本积分公式，如 $\int x^n\,\mathrm{d}x$、$\int e^x\,\mathrm{d}x$、$\int \ln x\,\mathrm{d}x$；
3. 简单平面区域，如直线、抛物线、圆、三角形；
4. 一元函数平均值：$\frac{1}{b-a}\int_a^b f(x)\,\mathrm{d}x$；
5. 基础概率直觉：密度、期望、联合分布的雏形。

   如果这些不完全熟，也没关系。本章会在需要处重新解释。

# 2 从一元积分到二重积分

## 2.1 一元积分：沿区间的累加

一元积分的核心不是“面积”，而是**累加**。如果 $f(x)$ 表示位置 $x$ 处单位长度的数量，那么：

$$\int_a^b f(x)\,\mathrm{d}x$$

就是从 $a$ 到 $b$ 把所有小段上的贡献加起来。

把区间 $[a, b]$ 切成小段，每段长度为 $\Delta x$，取样点 $x_i^*$，则近似为：

$$\sum_i f(x_i^*)\Delta x.$$

当切分越来越细，这个和式的极限就是积分。

## 2.2 二重积分：在区域上的累加

现在区域不再是一条线，而是平面上的一个集合 $R \subset \mathbb{R}^2$。函数 $f(x,y)$ 表示点 $(x,y)$ 处单位面积的贡献。把区域切成很多小矩形，每块面积是 $\Delta A_{ij} = \Delta x \Delta y$，则总贡献近似为：

$$\sum_i \sum_j f(x_i^*, y_j^*) \Delta x \Delta y.$$

如果切分越来越细，极限就是二重积分：

$$\iint_R f(x,y)\,\mathrm{d}A.$$

当用直角坐标时，常写作：

$$\iint_R f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

**定义 1.1：二重积分的直观定义**

设 $R \subset \mathbb{R}^2$ 是平面区域，$f : R \to \mathbb{R}$ 是函数。若把 $R$ 划分成许多小块后，Riemann 和

$$\sum_{k=1}^{n} f(x_k^*, y_k^*) \Delta A_k$$

在划分直径趋于 0 时趋向同一个极限，则称 $f$ 在 $R$ 上可积，并把极限记为：

$$\iint_R f(x,y)\,\mathrm{d}A.$$

**直觉**

$\mathrm{d}A$ 表示“一个很小的面积单位”。在直角坐标中，$\mathrm{d}A = \mathrm{d}x\,\mathrm{d}y$。所以二重积分的基本想法是：

$$\text{总量} \approx \sum \text{单位面积贡献} \times \text{小面积}.$$

这和经济学中“总量 = 各类型贡献的加总”完全同构。

## 2.3 几何图像：曲面下的体积

如果 $f(x,y) \geq 0$，那么 $z = f(x,y)$ 是三维空间中的一个曲面，$R$ 是它在 $xy$-平面上的底面。二重积分：

$$\iint_R f(x,y)\,\mathrm{d}A$$

就是曲面 $z = f(x,y)$ 与底面 $R$ 之间的体积。

但请注意：几何体积只是一个解释。对经济金融更重要的是“区域聚合”：

$$\text{aggregate} = \iint \text{individual contribution} \times \text{population density}.$$

# 3 矩形区域上的二重积分

最简单的区域是矩形：

$$R=[a,b]\times[c,d]=\{(x,y):a\le x\le b,\ c\le y\le d\}.$$

在矩形上，二重积分通常可以通过累次积分计算：

$$\iint_R f(x,y)\,\mathrm{d}A=\int_a^b\int_c^d f(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$

这句话的意思是：先固定 $x$，沿 $y$ 方向累加，再对所有 $x$ 累加。

同样，也可以先对 $x$ 积分，再对 $y$ 积分：

$$\iint_R f(x,y)\,\mathrm{d}A=\int_c^d\int_a^b f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

**定义 1.2：累次积分**

给定矩形区域 $R=[a,b]\times[c,d]$，表达式

$$\int_a^b\left[\int_c^d f(x,y)\,\mathrm{d}y\right]\mathrm{d}x$$

称为累次积分。内层积分把 $x$ 暂时看作常数，先对 $y$ 积分；外层积分再对 $x$ 积分。

**例 1.1：矩形区域上的直接计算**

计算

$$\int_0^2\int_1^3 xy\,\mathrm{d}y\,\mathrm{d}x.$$

**解：**先把 $x$ 当作常数，对 $y$ 积分：

$$\int_1^3 xy\,\mathrm{d}y=x\int_1^3 y\,\mathrm{d}y=x\left[\frac{y^2}{2}\right]_1^3=x\cdot\frac{9-1}{2}=4x.$$

再对 $x$ 积分：

$$\int_0^2 4x\,\mathrm{d}x=\left[2x^2\right]_0^2=8.$$

所以答案是：

$$\boxed{8}.$$

## 3.1 Fubini 定理的直觉

如果函数足够好，例如连续，或者更一般地可积，那么在矩形上可以交换积分顺序：

$$\int_a^b\int_c^d f(x,y)\,\mathrm{d}y\,\mathrm{d}x=\int_c^d\int_a^b f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

**定理 1.1：Fubini 定理的基础版**

若 $f(x,y)$ 在矩形区域 $[a,b]\times[c,d]$ 上连续，则

$$\iint_R f(x,y)\,\mathrm{d}A=\int_a^b\int_c^d f(x,y)\,\mathrm{d}y\,\mathrm{d}x=\int_c^d\int_a^b f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

**证明思路：为什么可以换顺序**

严格证明需要 Riemann 或 Lebesgue 积分理论。这里给出经济金融中最重要的直觉：矩形区域被分成许多小格子，每个小格子的贡献是

$$f(x_i,y_j)\Delta x\Delta y.$$

无论先按行相加再按列相加，还是先按列相加再按行相加，有限求和的结果都一样：

$$\sum_i\sum_j f(x_i,y_j)\Delta x\Delta y=\sum_j\sum_i f(x_i,y_j)\Delta x\Delta y.$$

Fubini 定理就是在适当条件下把这个“有限求和换顺序”的直觉推广到极限。

**直觉**

Fubini 定理在经济学中的含义是：聚合可以按不同维度进行。例如先按收入组内部聚合、再按教育水平聚合，或者先按教育水平、再按收入聚合。在条件合适时，最终总量相同。

# 4 一般区域：Type I 与 Type II

经济金融中的区域通常不是矩形。例如收入 $y$ 和消费 $c$ 可能满足 $0\le c\le y$，即消费不能超过收入。这个区域是三角形，不是矩形。

因此，我们需要学会把区域写成积分限。

## 4.1 Type I 区域：上下边界由 $y$ 表示

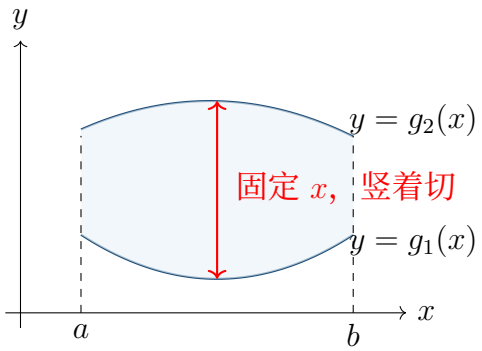
Type I 区域形如：

$$R=\{(x,y):a\le x\le b,\ g_1(x)\le y\le g_2(x)\}.$$

也就是说，先确定 $x$ 的范围；每个固定的 $x$ 下，$y$ 从下边界 $g_1(x)$ 到上边界 $g_2(x)$。

此时：

$$\iint_R f(x,y)\,\mathrm{d}A=\int_a^b\int_{g_1(x)}^{g_2(x)} f(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$



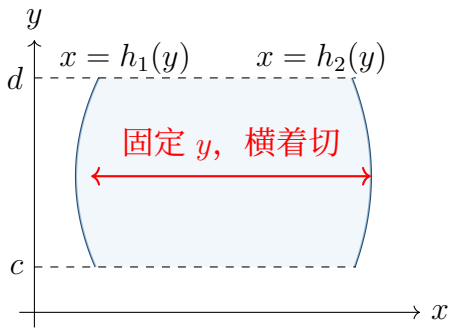

### 4.2 Type II 区域：左右边界由 $x$ 表示

Type II 区域形如：

$$R = \{(x, y) : c \le y \le d,\ h_1(y) \le x \le h_2(y)\}.$$

此时：

$$\iint_R f(x, y)\,\mathrm{d}A = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$




**方法**

判断 Type I 还是 Type II 的方法：

1. 画区域，不要先写积分限；
2. 如果每条竖线与区域相交成一个连续线段，优先用 Type I；
3. 如果每条横线与区域相交成一个连续线段，优先用 Type II；
4. 如果竖切或横切会得到多个断开的线段，就需要分割区域。

## 5 改变积分次序

改变积分次序是本章最重要的技术。它不是机械变形，而是区域重写。

给定：

$$\int_a^b \int_{g_1(x)}^{g_2(x)} f(x, y)\,\mathrm{d}y\,\mathrm{d}x.$$

这表示区域

$$R = \{(x, y) : a \le x \le b,\ g_1(x) \le y \le g_2(x)\}.$$

如果要改成先对 $x$ 积分，就必须把同一个区域写成：

$$R = \{(x, y) : c \le y \le d,\ h_1(y) \le x \le h_2(y)\}.$$

于是：

$$\int_a^b \int_{g_1(x)}^{g_2(x)} f(x, y)\, \mathrm{d}y\, \mathrm{d}x = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x, y)\, \mathrm{d}x\, \mathrm{d}y.$$

## 5.1 改变次序的标准流程

**方法**

改变积分次序的五步：

1. 从原积分限读出区域 $R$；
2. 画图，标出所有边界曲线；
3. 确定新外层变量的整体范围；
4. 固定新外层变量，用一条横线或竖线穿过区域，读出内层变量的左右或上下边界；
5. 检查是否需要分段。如果穿过区域时边界形式发生变化，就必须拆成多个积分。

**例 1.2：三角形区域改积分次序**

把下面积分改成先对 $x$ 积分：

$$\int_0^1 \int_0^x f(x, y)\, \mathrm{d}y\, \mathrm{d}x.$$

**解：** 原区域为：

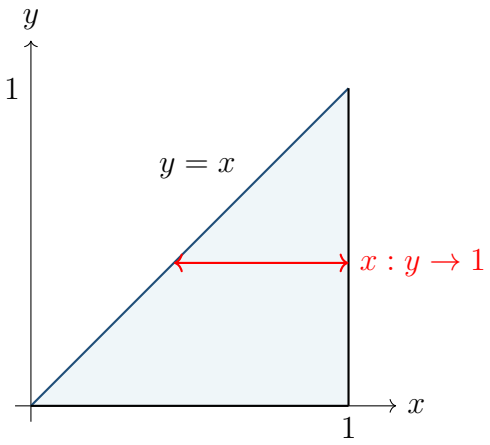
$$R = \{(x, y) : 0 \le x \le 1,\ 0 \le y \le x\}.$$

这表示单位正方形中直线 $y = x$ 下方的三角形。
画图后可见，$y$ 的整体范围是 $0 \le y \le 1$。固定一个 $y$，横着看区域，$x$ 从直线 $x = y$ 到右边界 $x = 1$。所以：

$$R = \{(x, y) : 0 \le y \le 1,\ y \le x \le 1\}.$$

因此：

$$\boxed{\int_0^1 \int_0^x f(x, y)\, \mathrm{d}y\, \mathrm{d}x = \int_0^1 \int_y^1 f(x, y)\, \mathrm{d}x\, \mathrm{d}y.}$$




**常见误区**

初学者最常见的错误是直接把 $\mathrm{d}y\,\mathrm{d}x$ 换成 $\mathrm{d}x\,\mathrm{d}y$，但积分限不变。这是错误的。改变次序时，区域不能变，变量边界必须重新读图。

**例 1.3：需要分段的区域**

将积分

$$\int_0^2 \int_{x^2}^{2x} f(x,y)\,\mathrm{d}y\,\mathrm{d}x$$

改为先对 $x$ 积分。

**解：** 区域为：

$$R = \{(x,y) : 0 \le x \le 2,\ x^2 \le y \le 2x\}.$$

边界是抛物线 $y = x^2$ 和直线 $y = 2x$。二者交点由

$$x^2 = 2x$$

给出，即 $x = 0, 2$，对应 $y = 0, 4$。所以 $y$ 的整体范围为 $0 \le y \le 4$。

固定 $y$，需要同时满足：

$$y \ge x^2 \quad \Longleftrightarrow \quad x \le \sqrt{y},$$

以及

$$y \le 2x \quad \Longleftrightarrow \quad x \ge \frac{y}{2}.$$

因此横切时，$x$ 从 $y/2$ 到 $\sqrt{y}$。所以：

$$\boxed{\int_0^2 \int_{x^2}^{2x} f(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_0^4 \int_{y/2}^{\sqrt{y}} f(x,y)\,\mathrm{d}x\,\mathrm{d}y.}$$

这个例子虽然没有分段，但它提醒你：改变次序的本质是把原来的边界方程反解成新变量的边界。

**例 1.4：真正需要分段的例子**

将

$$\int_0^1 \int_0^{\sqrt{x}} f(x,y)\,\mathrm{d}y\,\mathrm{d}x + \int_1^2 \int_0^{2-x} f(x,y)\,\mathrm{d}y\,\mathrm{d}x$$

合并为一个区域后，改为先对 $x$ 积分。

**解：**区域由两部分组成：

$$R_1 = \{0 \le x \le 1,\ 0 \le y \le \sqrt{x}\},$$

$$R_2 = \{1 \le x \le 2,\ 0 \le y \le 2 - x\}.$$

第一部分的上边界 $y = \sqrt{x}$ 等价于 $x = y^2$。第二部分的上边界 $y = 2 - x$ 等价于 $x = 2 - y$。

整体上 $y$ 的范围是 $0 \le y \le 1$。固定 $y \in [0,1]$，区域中的 $x$ 从左边界 $x = y^2$ 到右边界 $x = 2 - y$。所以：

$$\boxed{\int_0^1 \int_0^{\sqrt{x}} f(x,y)\,\mathrm{d}y\,\mathrm{d}x + \int_1^2 \int_0^{2-x} f(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 \int_{y^2}^{2-y} f(x,y)\,\mathrm{d}x\,\mathrm{d}y.}$$

# 6 二重积分的基本解释

同一个符号 $\iint_R f(x,y)\,\mathrm{d}A$ 在不同模型中有不同解释。你需要学会“看到语境，翻译含义”。

## 6.1 面积

如果 $f(x,y) = 1$，则

$$\iint_R 1\,\mathrm{d}A = \mathrm{Area}(R).$$

**例 1.5：用二重积分求三角形面积**

求区域

$$R = \{(x,y) : 0 \le x \le 1,\ 0 \le y \le x\}$$

的面积。

**解：**

$$\mathrm{Area}(R) = \iint_R 1\,\mathrm{d}A = \int_0^1 \int_0^x 1\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 x\,\mathrm{d}x = \frac{1}{2}.$$

这和几何上底高都为 1 的直角三角形面积一致。

## 6.2 总量

如果 $f(x,y)$ 是单位面积上的密度或单位类型上的贡献，那么积分就是总量。

例如 $x$ 表示收入，$y$ 表示消费倾向，$f(x,y)$ 表示类型 $(x,y)$ 人群的人数密度，$T(x,y)$ 表示该类型税负，那么总税收是：

$$\iint_R T(x,y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

## 6.3 平均值

区域 $R$ 上函数 $f$ 的平均值定义为：

$$\bar{f}_R = \frac{1}{\text{Area}(R)}\iint_R f(x,y)\,\mathrm{d}A.$$

这和一元函数平均值：

$$\frac{1}{b-a}\int_a^b f(x)\,\mathrm{d}x$$

完全对应。

**例 1.6：区域上的平均值**

求 $f(x,y)=x+y$ 在三角形区域

$$R=\{(x,y):0\le x\le 1,\ 0\le y\le x\}$$

上的平均值。

**解：**面积为 $1/2$。先算积分：

$$\iint_R (x+y)\,\mathrm{d}A=\int_0^1\int_0^x (x+y)\,\mathrm{d}y\,\mathrm{d}x.$$

内层积分：

$$\int_0^x (x+y)\,\mathrm{d}y=\left[xy+\frac{y^2}{2}\right]_0^x=x^2+\frac{x^2}{2}=\frac{3}{2}x^2.$$

外层积分：

$$\int_0^1 \frac{3}{2}x^2\,\mathrm{d}x=\frac{1}{2}.$$

因此平均值为：

$$\bar{f}_R=\frac{1}{1/2}\cdot\frac{1}{2}=1.$$

# 7 二重积分与联合密度

二重积分最重要的经济金融入口是概率。设 $(X,Y)$ 是二维连续随机变量，联合密度为 $f_{X,Y}(x,y)$。则某个区域 $R$ 中的概率为：

$$P((X,Y)\in R)=\iint_R f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

整个平面上的总概率必须是 1：

$$\iint_{\mathbb{R}^2} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1.$$

## 7.1 期望就是加权平均

若 $g(X,Y)$ 是随机变量的函数，则：

$$\mathbb{E}[g(X,Y)] = \iint_{\mathbb{R}^2} g(x,y) f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

这里的 $f_{X,Y}(x,y)$ 起到权重作用。也就是说，期望不是“简单平均”，而是按概率密度加权后的平均。

**定义 1.3：联合密度下的期望**

若 $(X,Y)$ 有联合密度 $f_{X,Y}$，且 $g(X,Y)$ 可积，则

$$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$

如果密度只在区域 $R$ 上非零，也可以写成：

$$\mathbb{E}[g(X,Y)] = \iint_R g(x,y) f_{X,Y}(x,y)\,\mathrm{d}A.$$

**例 1.7：联合密度归一化与期望**

设 $(X,Y)$ 在三角形区域

$$R = \{(x,y) : 0 \le y \le x \le 1\}$$

上的联合密度为常数 $c$，区域外为 0。求 $c$ 和 $\mathbb{E}[X+Y]$。

**解：** 因为总概率为 1：

$$1 = \iint_R c\,\mathrm{d}A = c \cdot \mathrm{Area}(R) = c \cdot \frac{1}{2}.$$

所以 $c = 2$。

再算期望：

$$\mathbb{E}[X+Y] = \iint_R (x+y)c\,\mathrm{d}A = 2\int_0^1 \int_0^x (x+y)\,\mathrm{d}y\,\mathrm{d}x.$$

前面算过：

$$\int_0^1 \int_0^x (x+y)\,\mathrm{d}y\,\mathrm{d}x = \frac{1}{2}.$$

所以：

$$\boxed{\mathbb{E}[X+Y] = 1.}$$

## 7.2 边际密度

从联合密度得到单个变量的密度，需要把另一个变量“积分掉”：

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y,$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x.$$

**直觉**

“边际密度”这个名字和经济学中的“边际”不完全一样。这里的 marginal 指的是从联合分布中只看一个维度。比如只看收入 $X$，就把教育 $Y$ 的所有可能值加总掉。

**例 1.8：从联合密度得到边际密度**

沿用例 1.7，联合密度为

$$f_{X,Y}(x,y) = 2, \quad 0 \le y \le x \le 1.$$

求 $f_X(x)$。

**解：** 固定 $x \in [0,1]$，$y$ 的范围是 $0 \le y \le x$。所以：

$$f_X(x) = \int_0^x 2\,\mathrm{d}y = 2x, \quad 0 \le x \le 1.$$

区域外 $f_X(x) = 0$。检查归一化：

$$\int_0^1 2x\,\mathrm{d}x = 1.$$

# 8 经济金融应用一：异质性消费者聚合

设经济体中消费者有两种异质性：收入 $y$ 和消费倾向 $\alpha$。假设类型 $(y,\alpha)$ 的密度为 $f(y,\alpha)$，消费函数为

$$c(y,\alpha) = \alpha y.$$

那么总消费为：

$$C = \iint_R \alpha y f(y,\alpha)\,\mathrm{d}y\,\mathrm{d}\alpha.$$

如果 $f(y,\alpha)$ 是概率密度，那么 $C$ 是平均消费；如果 $f$ 是人数密度，那么 $C$ 是总消费。

**例 1.9：一个简单的总消费模型**

设 $y \in [0,1]$，$\alpha \in [0,1]$，类型均匀分布，消费为 $c(y,\alpha) = \alpha y$。求平均消费。

**解：** 均匀分布在单位正方形上，密度为 1。平均消费为：

$$\mathbb{E}[c(Y,A)] = \int_0^1 \int_0^1 \alpha y \, \mathrm{d}\alpha \, \mathrm{d}y.$$

先对 $\alpha$ 积分：

$$\int_0^1 \alpha y \, \mathrm{d}\alpha = y \cdot \frac{1}{2}.$$

再对 $y$ 积分：

$$\int_0^1 \frac{y}{2} \, \mathrm{d}y = \frac{1}{4}.$$

所以平均消费为：

$$\boxed{\frac{1}{4}.}$$

如果总人口规模是 $N$，则总消费为 $N/4$。

# 9 经济金融应用二：收入分布上的平均税负

设收入为 $y$，扣除项或家庭规模变量为 $h$，税负函数为 $T(y,h)$，联合密度为 $f(y,h)$。平均税负为：

$$\bar{T} = \iint T(y,h) f(y,h) \, \mathrm{d}y \, \mathrm{d}h.$$

总税收为：

$$\text{Tax Revenue} = N \iint T(y,h) f(y,h) \, \mathrm{d}y \, \mathrm{d}h,$$

其中 $N$ 是人口规模。

**例 1.10：二维税负聚合**

假设收入 $y \in [0,1]$，家庭规模 $h \in [0,1]$，二者均匀分布。税负函数为：

$$T(y,h) = \tau y (1-h), \quad \tau \in (0,1).$$

求平均税负。

**解：**

$$\bar{T} = \int_0^1 \int_0^1 \tau y (1-h) \, \mathrm{d}h \, \mathrm{d}y.$$

内层：

$$\int_0^1 \tau y (1-h) \, \mathrm{d}h = \tau y \left[ h - \frac{h^2}{2} \right]_0^1 = \frac{\tau y}{2}.$$

外层：

$$\int_0^1 \frac{\tau y}{2}\,\mathrm{d}y = \frac{\tau}{4}.$$

所以平均税负是：

$$\boxed{\bar{T} = \frac{\tau}{4}.}$$

# 10 金融应用：二维风险因子的期望收益

设资产收益由两个风险因子 $(M, S)$ 决定，其中 $M$ 是市场因子，$S$ 是行业因子。假设资产收益函数为：

$$R(M,S) = r_f + \beta_M M + \beta_S S + \gamma MS.$$

如果 $(M, S)$ 的联合密度为 $f(m, s)$，那么期望收益为：

$$\mathbb{E}[R] = \iint (r_f + \beta_M m + \beta_S s + \gamma ms)\, f(m,s)\,\mathrm{d}m\,\mathrm{d}s.$$

利用期望符号：

$$\mathbb{E}[R] = r_f + \beta_M \mathbb{E}[M] + \beta_S \mathbb{E}[S] + \gamma \mathbb{E}[MS].$$

其中 $\mathbb{E}[MS]$ 由二重积分给出：

$$\mathbb{E}[MS] = \iint ms f(m,s)\,\mathrm{d}m\,\mathrm{d}s.$$

**直觉**

金融里二维积分经常不是显式写出来的，因为很多教材直接用期望符号 $\mathbb{E}[\cdot]$。但你要知道，连续型二维随机变量的期望符号背后就是二重积分。

# 11 计算策略与常见错误

## 11.1 计算策略

**方法**

二重积分题的标准流程：

1. 先判断区域：矩形、三角形、曲线围成的区域，还是需要分段？
2. 画图。二重积分题不画图，错误率会大幅上升。
3. 选择积分顺序。优先选择积分限简单、被积函数容易积分的顺序。
4. 计算内层积分时，把外层变量看成常数。

5. 完成后检查量纲和符号：面积应非负，概率应在 $[0,1]$，密度积分应为 1。

## 11.2 常见错误

**常见误区**

常见错误包括：

1. 不画区域，直接猜积分限；
2. 把内层变量和外层变量混淆；
3. 改积分次序时只交换 $\mathrm{d}x\,\mathrm{d}y$，不重写边界；
4. 忘记联合密度需要归一化；
5. 把 $\iint_R f(x,y)\,\mathrm{d}A$ 的值误认为平均值，忘记除以面积或密度权重；
6. 遇到非矩形区域时仍然套矩形积分限；
7. 需要分段时强行写成一个积分。

# 12 Worked Examples：完整样板题

**样板题 1：从区域到积分限**

设 $R$ 是由 $y=x^2$ 和 $y=x$ 围成的区域。写出 $\iint_R f(x,y)\,\mathrm{d}A$ 的两种累次积分形式。

**解：** 交点由 $x^2=x$ 给出：$x=0,1$。在 $[0,1]$ 上，$x^2\le x$。所以 Type I 写法为：

$$\iint_R f(x,y)\,\mathrm{d}A=\int_0^1\int_{x^2}^{x} f(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$

要写 Type II，先找 $y$ 的范围。由于交点对应 $y=0,1$，所以 $0\le y\le 1$。固定 $y$，边界为：

$$y=x\Rightarrow x=y,\quad y=x^2\Rightarrow x=\sqrt{y}.$$

在区域内，$x$ 从 $y$ 到 $\sqrt{y}$。所以：

$$\iint_R f(x,y)\,\mathrm{d}A=\int_0^1\int_{y}^{\sqrt{y}} f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

### 样板题 2：选择更容易的积分顺序

计算：

$$I=\int_0^1\int_x^1 e^{y^2}\,\mathrm{d}y\,\mathrm{d}x.$$

**解：**直接对 $y$ 积分会遇到 $e^{y^2}$，没有初等原函数。因此考虑换次序。原区域是：

$$0\le x\le 1,\quad x\le y\le 1.$$

画图可知，这是单位正方形中直线 $y=x$ 上方的三角形。换成先对 $x$ 积分：

$$0\le y\le 1,\quad 0\le x\le y.$$

于是：

$$I=\int_0^1\int_0^y e^{y^2}\,\mathrm{d}x\,\mathrm{d}y.$$

内层对 $x$ 积分时，$e^{y^2}$ 是常数：

$$\int_0^y e^{y^2}\,\mathrm{d}x=ye^{y^2}.$$

所以：

$$I=\int_0^1 ye^{y^2}\,\mathrm{d}y.$$

令 $u=y^2$，$\mathrm{d}u=2y\,\mathrm{d}y$，得到：

$$I=\frac{1}{2}\int_0^1 e^u\,\mathrm{d}u=\frac{1}{2}(e-1).$$

答案：

$$\boxed{I=\frac{e-1}{2}.}$$

### 样板题 3：联合密度中的区域概率

设 $(X,Y)$ 的联合密度为

$$f(x,y)=6x,\quad 0\le y\le x\le 1,$$

区域外为 0。求 $P(Y\le 1/2)$。

**解：**首先可以检查密度归一化：

$$\int_0^1\int_0^x 6x\,\mathrm{d}y\,\mathrm{d}x=\int_0^1 6x^2\,\mathrm{d}x=2,$$

这不是 1，所以题目中 $6x$ 不是合法联合密度。若要成为合法密度，常数应为 3，因为：

$$\int_0^1 \int_0^x 3x \,\mathrm{d}y \,\mathrm{d}x = \int_0^1 3x^2 \,\mathrm{d}x = 1.$$

因此这个题目的第一个关键是发现归一化错误。

若把密度修正为 $f(x,y) = 3x$，则

$$P(Y \le 1/2) = \iint_{R \cap \{y \le 1/2\}} 3x \,\mathrm{d}y \,\mathrm{d}x.$$

对固定 $y \in [0, 1/2]$，原区域要求 $y \le x \le 1$。所以：

$$P(Y \le 1/2) = \int_0^{1/2} \int_y^1 3x \,\mathrm{d}x \,\mathrm{d}y.$$

内层：

$$\int_y^1 3x \,\mathrm{d}x = \frac{3}{2}(1 - y^2).$$

外层：

$$\int_0^{1/2} \frac{3}{2}(1-y^2) \,\mathrm{d}y = \frac{3}{2}\left[y - \frac{y^3}{3}\right]_0^{1/2} = \frac{3}{2}\left(\frac{1}{2} - \frac{1}{24}\right) = \frac{11}{16}.$$

所以修正后答案为：

$$\boxed{\frac{11}{16}.}$$

**常见误区**

这个样板题故意放了一个“密度不合法”的陷阱。经济金融题里遇到 density 时，第一反应应该是检查它在支持集上的积分是否等于 1。

**样板题 4：收入与税负的政策解释**

设收入 $Y$ 和避税能力 $A$ 在单位正方形 $[0,1]^2$ 上均匀分布。政府税负函数为

$$T(y,a) = \tau y(1 - a^2).$$

求平均税负，并解释 $a$ 的作用。

**解：** 平均税负为：

$$\mathbb{E}[T(Y,A)] = \int_0^1 \int_0^1 \tau y(1-a^2) \,\mathrm{d}a \,\mathrm{d}y.$$

内层：

$$\int_0^1 \tau y(1-a^2) \,\mathrm{d}a = \tau y \left[a - \frac{a^3}{3}\right]_0^1 = \frac{2\tau y}{3}.$$

外层：

$$\int_0^1 \frac{2\tau y}{3}\,\mathrm{d}y = \frac{\tau}{3}.$$

所以平均税负为：

$$\boxed{\frac{\tau}{3}.}$$

经济解释：$a$ 越大，$1-a^2$ 越小，税负越低。它可以理解为避税能力、扣除能力或制度漏洞利用能力。二重积分把不同收入与不同避税能力的人群全部聚合起来，得到平均税负。

# 13 练习题

练习分五组。A 组训练基本计算；B 组训练概念判断；C 组混合本章技能；D 组连接经济金融建模；E 组挑战题。

## A 组：基础计算

A1. 计算 $\displaystyle\int_0^2\int_0^1 (x+2y)\,\mathrm{d}y\,\mathrm{d}x$。

A2. 计算 $\displaystyle\int_1^3\int_0^2 xy^2\,\mathrm{d}y\,\mathrm{d}x$。

A3. 求区域 $R=\{(x,y):0\le x\le 2,\ 0\le y\le x\}$ 的面积。

A4. 计算 $\displaystyle\int_0^1\int_0^x (x-y)\,\mathrm{d}y\,\mathrm{d}x$。

A5. 计算 $\displaystyle\int_0^1\int_x^1 y\,\mathrm{d}y\,\mathrm{d}x$。

A6. 求 $f(x,y)=x^2+y^2$ 在单位正方形 $[0,1]^2$ 上的平均值。

A7. 求 $f(x,y)=xy$ 在三角形区域 $0\le y\le x\le 1$ 上的平均值。

A8. 计算 $\displaystyle\int_0^1\int_0^{1-y} (x+y)\,\mathrm{d}x\,\mathrm{d}y$。

## B 组：概念判断

B1. 判断：$\int_0^1\int_0^x f(x,y)\,\mathrm{d}y\,\mathrm{d}x$ 的区域是单位正方形。对还是错？说明理由。

B2. 判断：若 $f(x,y)\ge 0$，则 $\iint_R f(x,y)\,\mathrm{d}A\ge 0$。对还是错？

B3. 判断：$\iint_R f(x,y)\,\mathrm{d}A$ 一定等于 $f$ 在 $R$ 上的平均值。对还是错？

B4. 设 $f_{X,Y}$ 是联合密度。为什么必须有 $\iint f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y=1$？

B5. 解释为什么改变积分次序时必须重新画区域。

## C 组：混合题

C1. 把 $\displaystyle\int_0^1\int_0^{1-x} f(x,y)\,\mathrm{d}y\,\mathrm{d}x$ 改成先对 $x$ 积分。

C2. 把 $\displaystyle\int_0^4\int_{\sqrt{y}}^2 f(x,y)\,\mathrm{d}x\,\mathrm{d}y$ 改成先对 $y$ 积分。

C3. 计算 $\displaystyle\int_0^1\int_y^1 e^{x^2}\,\mathrm{d}x\,\mathrm{d}y$。提示：换次序。

C4. 设 $R$ 由 $y=x^2$ 和 $y=2x$ 围成，写出 $\iint_R f(x,y)\,\mathrm{d}A$ 的两种积分形式。

C5. 计算 $\displaystyle\int_0^1\int_0^{\sqrt{1-y^2}} 1\,\mathrm{d}x\,\mathrm{d}y$，并解释几何含义。

## D 组：经济金融建模题

D1. 设收入 $y\in[0,1]$、消费倾向 $\alpha\in[0,1]$ 均匀分布，消费函数为 $c(y,\alpha)=\alpha y^2$。求平均消费。

D2. 设税负函数 $T(y,h)=\tau(y-h)^2$，其中 $(y,h)$ 在 $[0,1]^2$ 上均匀分布。求平均税负。

D3. 设二维风险因子 $(M,S)$ 在 $[-1,1]^2$ 上均匀分布，资产收益为 $R=1+2M+S+MS$。求 $\mathbb{E}[R]$。

D4. 设 $(X,Y)$ 在三角形区域 $0\le y\le x\le 1$ 上均匀分布，求 $\mathbb{E}[X]$ 和 $\mathbb{E}[Y]$。

D5. 设 $(X,Y)$ 的联合密度为 $f(x,y)=c(x+y)$，定义在 $[0,1]^2$ 上。求 $c$，并求 $\mathbb{E}[X]$。

## E 组：挑战/证明题

E1. 用 Riemann 和的直觉解释为什么在矩形区域上可以交换积分次序。

E2. 证明若 $f(x,y)=g(x)h(y)$，且区域为矩形 $[a,b]\times[c,d]$，则

$$\iint_R f(x,y)\,\mathrm{d}A=\left(\int_a^b g(x)\,\mathrm{d}x\right)\left(\int_c^d h(y)\,\mathrm{d}y\right).$$

E3. 设 $(X,Y)$ 在单位正方形上均匀分布。证明 $X$ 与 $Y$ 的边际密度都是 1。

E4. 设 $R$ 是有限面积区域。证明常数函数 $f(x,y)=k$ 在 $R$ 上的平均值等于 $k$。

E5. 构造一个经济解释：为什么 $\mathbb{E}[g(X,Y)]$ 可以被看作“按类型分布加权的总平均”。

# 14 练习题详细解答

## A 组答案

A1.
$$\int_0^2 \int_0^1 (x+2y)\,\mathrm{d}y\,\mathrm{d}x = \int_0^2 [x+y^2]_0^1\,\mathrm{d}x = \int_0^2 (x+1)\,\mathrm{d}x = 4.$$

A2.
$$\int_1^3 \int_0^2 xy^2\,\mathrm{d}y\,\mathrm{d}x = \int_1^3 x\left[\frac{y^3}{3}\right]_0^2\,\mathrm{d}x = \int_1^3 \frac{8x}{3}\,\mathrm{d}x = \frac{32}{3}.$$

A3.
$$\text{Area}(R) = \int_0^2 \int_0^x 1\,\mathrm{d}y\,\mathrm{d}x = \int_0^2 x\,\mathrm{d}x = 2.$$

A4.
$$\int_0^1 \int_0^x (x-y)\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 \left[xy - \frac{y^2}{2}\right]_0^x\,\mathrm{d}x = \int_0^1 \frac{x^2}{2}\,\mathrm{d}x = \frac{1}{6}.$$

A5.
$$\int_0^1 \int_x^1 y\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 \frac{1-x^2}{2}\,\mathrm{d}x = \frac{1}{2}\left(1-\frac{1}{3}\right) = \frac{1}{3}.$$

A6. 单位正方形面积为 1，因此平均值等于积分：
$$\int_0^1 \int_0^1 (x^2+y^2)\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 \left(x^2+\frac{1}{3}\right)\,\mathrm{d}x = \frac{2}{3}.$$

A7. 先算积分：
$$\int_0^1 \int_0^x xy\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 x\cdot\frac{x^2}{2}\,\mathrm{d}x = \frac{1}{2}\int_0^1 x^3\,\mathrm{d}x = \frac{1}{8}.$$
三角形面积为 $1/2$，所以平均值为：
$$\frac{1/8}{1/2} = \frac{1}{4}.$$

A8.
$$\int_0^1 \int_0^{1-y} (x+y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 \left[\frac{x^2}{2}+yx\right]_0^{1-y}\,\mathrm{d}y.$$
内层结果为：
$$\frac{(1-y)^2}{2} + y(1-y) = \frac{1}{2} - \frac{y^2}{2}.$$
外层积分：
$$\int_0^1 \left(\frac{1}{2}-\frac{y^2}{2}\right)\,\mathrm{d}y = \frac{1}{3}.$$

## B 组答案

B1. 错。区域是 $0 \le x \le 1,\ 0 \le y \le x$，即单位正方形中直线 $y = x$ 下方的三角形。

B2. 对。因为每个小块的贡献 $f(x,y)\Delta A \ge 0$，极限也非负。

B3. 错。平均值是

$$\frac{1}{\text{Area}(R)} \iint_R f(x,y)\,\mathrm{d}A.$$

积分本身是总量，不是平均值。

B4. 因为联合密度描述所有可能状态的概率分布，整个样本空间中的概率总和必须等于 1。连续情形下“概率总和”由积分表示。

B5. 因为积分限描述的是区域。改变积分次序等于换一种方式扫描同一个区域，如果不重新画区域，就容易改变了实际积分范围。

## C 组答案

C1. 原区域：$0 \le x \le 1,\ 0 \le y \le 1-x$。这是单位三角形。换次序：$0 \le y \le 1,\ 0 \le x \le 1-y$。所以

$$\int_0^1 \int_0^{1-x} f(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 \int_0^{1-y} f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

C2. 原区域：$0 \le y \le 4,\ \sqrt{y} \le x \le 2$。由 $x \ge \sqrt{y}$ 得 $y \le x^2$。所以 $0 \le x \le 2,\ 0 \le y \le x^2$。因此

$$\int_0^4 \int_{\sqrt{y}}^2 f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^2 \int_0^{x^2} f(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$

C3. 原区域：$0 \le y \le 1,\ y \le x \le 1$。换次序：$0 \le x \le 1,\ 0 \le y \le x$。所以

$$\int_0^1 \int_y^1 e^{x^2}\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 \int_0^x e^{x^2}\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 x e^{x^2}\,\mathrm{d}x = \frac{e-1}{2}.$$

C4. 交点：$x^2 = 2x \Rightarrow x = 0, 2$。在 $[0,2]$ 上，$x^2 \le 2x$，所以

$$\iint_R f\,\mathrm{d}A = \int_0^2 \int_{x^2}^{2x} f(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$

换成 $y$ 外层：$0 \le y \le 4$，且 $y/2 \le x \le \sqrt{y}$，所以

$$\iint_R f\,\mathrm{d}A = \int_0^4 \int_{y/2}^{\sqrt{y}} f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

C5. 区域为第一象限单位圆的四分之一：$x^2 + y^2 \le 1,\ x, y \ge 0$。因此积分为四分之一圆面积：

$$\frac{\pi}{4}.$$

## D 组答案

D1.

$$\mathbb{E}[c] = \int_0^1 \int_0^1 \alpha y^2 \,\mathrm{d}\alpha \,\mathrm{d}y = \left(\int_0^1 \alpha \,\mathrm{d}\alpha\right)\left(\int_0^1 y^2 \,\mathrm{d}y\right) = \frac{1}{2}\cdot\frac{1}{3} = \frac{1}{6}.$$

D2.

$$\mathbb{E}[T] = \tau \int_0^1 \int_0^1 (y-h)^2 \,\mathrm{d}h \,\mathrm{d}y.$$

展开：

$$(y-h)^2 = y^2 - 2yh + h^2.$$

因为 $Y, H$ 独立均匀：

$$\mathbb{E}[(Y-H)^2] = \mathbb{E}[Y^2] - 2\mathbb{E}[Y]\mathbb{E}[H] + \mathbb{E}[H^2] = \frac{1}{3} - 2\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{3} = \frac{1}{6}.$$

所以平均税负为：

$$\boxed{\frac{\tau}{6}.}$$

D3. 在 $[-1,1]^2$ 上均匀分布，$\mathbb{E}[M] = \mathbb{E}[S] = 0$。独立且对称，所以 $\mathbb{E}[MS] = \mathbb{E}[M]\mathbb{E}[S] = 0$。因此

$$\mathbb{E}[R] = 1.$$

D4. 三角形面积为 $1/2$，均匀密度为 2。于是

$$\mathbb{E}[X] = 2\int_0^1 \int_0^x x \,\mathrm{d}y \,\mathrm{d}x = 2\int_0^1 x^2 \,\mathrm{d}x = \frac{2}{3}.$$

$$\mathbb{E}[Y] = 2\int_0^1 \int_0^x y \,\mathrm{d}y \,\mathrm{d}x = 2\int_0^1 \frac{x^2}{2} \,\mathrm{d}x = \frac{1}{3}.$$

D5. 归一化要求：

$$1 = \int_0^1 \int_0^1 c(x+y) \,\mathrm{d}y \,\mathrm{d}x = c\int_0^1 \left(x + \frac{1}{2}\right) \mathrm{d}x = c.$$

所以 $c = 1$。

$$\mathbb{E}[X] = \int_0^1 \int_0^1 x(x+y) \,\mathrm{d}y \,\mathrm{d}x = \int_0^1 \left(x^2 + \frac{x}{2}\right) \mathrm{d}x = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

## E 组答案

E1. 把矩形切成小格子，每个格子的贡献是 $f(x_i, y_j)\Delta x \Delta y$。先按行加再按列加，或先按列加再按行加，有限和相同。在连续函数等合适条件下，取极限后仍然相同，这就是 Fubini 定理的直觉。

E2.

$$\iint_R g(x)h(y)\,\mathrm{d}A = \int_a^b \int_c^d g(x)h(y)\,\mathrm{d}y\,\mathrm{d}x.$$

内层对 $y$ 积分时，$g(x)$ 是常数：

$$= \int_a^b g(x)\left(\int_c^d h(y)\,\mathrm{d}y\right)\mathrm{d}x.$$

括号内与 $x$ 无关，可提出：

$$= \left(\int_c^d h(y)\,\mathrm{d}y\right)\left(\int_a^b g(x)\,\mathrm{d}x\right).$$

E3. 单位正方形上联合密度为 $f_{X,Y}(x,y)=1$。因此

$$f_X(x) = \int_0^1 1\,\mathrm{d}y = 1, \quad 0 \le x \le 1.$$

同理：

$$f_Y(y) = \int_0^1 1\,\mathrm{d}x = 1, \quad 0 \le y \le 1.$$

E4. 平均值定义为：

$$\bar{f}_R = \frac{1}{\mathrm{Area}(R)} \iint_R k\,\mathrm{d}A.$$

因为 $k$ 是常数，

$$\iint_R k\,\mathrm{d}A = k\iint_R 1\,\mathrm{d}A = k\,\mathrm{Area}(R).$$

所以：

$$\bar{f}_R = \frac{k\,\mathrm{Area}(R)}{\mathrm{Area}(R)} = k.$$

E5. 可以把 $(X,Y)$ 看作个体类型，例如 $X$ 是收入，$Y$ 是教育；$g(X,Y)$ 是该类型个体的消费、税负或效用。联合密度 $f_{X,Y}(x,y)$ 描述每种类型出现的相对频率。因此

$$\mathbb{E}[g(X,Y)] = \iint g(x,y)f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

就是把所有类型的贡献按其人口权重加权平均。

# 15 本章总结图谱

**本章压缩总结**

| 模块 | 核心公式 | 经济金融含义 |
| --- | --- | --- |
| 二重积分 | $\iint_R f(x,y)\,\mathrm{d}A$ | 在二维类型空间中聚合贡献 |
| 矩形区域 | $\int_a^b \int_c^d f(x,y)\,\mathrm{d}y\,\mathrm{d}x$ | 先按一个维度聚合，再按另一个维度聚合 |
| Type I 区域 | $\int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,\mathrm{d}y\,\mathrm{d}x$ | 固定 $x$，竖着切区域 |
| Type II 区域 | $\int_c^d \int_{h_1(y)}^{h_2(y)} f(x,y)\,\mathrm{d}x\,\mathrm{d}y$ | 固定 $y$，横着切区域 |
| 面积 | $\iint_R 1\,\mathrm{d}A$ | 类型空间大小、可行集合面积 |
| 平均值 | $\frac{1}{\mathrm{Area}(R)} \iint_R f\,\mathrm{d}A$ | 均匀类型分布下的平均贡献 |
| 联合密度 | $P((X,Y)\in R) = \iint_R f_{X,Y}\,\mathrm{d}A$ | 二维随机变量落在某区域的概率 |
| 期望 | $\mathbb{E}[g(X,Y)] = \iint g(x,y) f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$ | 按类型分布加权的平均经济量 |

## 本章最重要的三句话

1. 二重积分的本质不是“体积题”，而是“二维区域上的累加”。
2. 改变积分次序的本质不是交换符号，而是重新描述同一个区域。
3. 在经济金融中，二重积分最常见的身份是联合分布下的期望、异质性个体的聚合、二维风险因子的加权平均。

## 下一章连接

下一章将把二重积分推广到三重积分。你会看到：当个体不仅有收入和教育，还有财富；当资产不仅有两个风险因子，还有第三个宏观状态；当联合分布从二维变成三维时，积分语言如何自然扩展。核心问题仍然不变：

在多维类型空间中，把局部贡献聚合成总体对象。

# 经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

---

# Chapter 2

## 三重积分与多维分布

---

本章把二重积分推广到三维空间：从“平面区域上的累加”走向“空间区域上的累加”。
对经济金融来说，三重积分不是只为求体积，而是多维异质性、联合密度、边际分布、条件分布和多风险因子模型的基本语言。

math4econ personalized notes
2026

# 目录

**Chapter 2 总览** **2**

**1 本章为什么重要** **2**
1.1 经济金融中的典型三维对象 . . . . . 3
1.2 最小前置知识 . . . . . 3

**2 从二重积分到三重积分** **3**
2.1 长方体上的三重积分 . . . . . 3
2.2 为什么积分次序可以交换 . . . . . 4

**3 一般三维区域：投影与截面** **5**
3.1 第一种常用写法：先投影到 $xy$ 平面 . . . . . 5
3.2 另外两种投影 . . . . . 6
3.3 经济解释：先按哪个维度聚合? . . . . . 6

**4 截面法：把三维问题切成二维问题** **7**

**5 三重积分的几种核心含义** **8**
5.1 含义一：体积 . . . . . 8
5.2 含义二：总量 . . . . . 8
5.3 含义三：平均值 . . . . . 8
5.4 含义四：期望 . . . . . 9

**6 联合密度、边际密度与条件分布** **9**
6.1 三维联合密度 . . . . . 9
6.2 从联合密度到边际密度 . . . . . 9
6.3 条件密度 . . . . . 10

**7 经济金融应用：从三维积分到模型语言** **11**
7.1 应用一：三维家庭异质性与总体消费 . . . . . 11
7.2 应用二：三资产收益与组合风险 . . . . . 11
7.3 应用三：计量中的联合分布和条件期望 . . . . . 12

**8 Worked Examples：完整样板题** **12**

**9 计算方法总结** **14**

**10 练习题** **14**
A 组：基础计算 . . . . . 14
B 组：概念判断与积分限 . . . . . 15
C 组：混合题 . . . . . 15
D 组：经济金融建模题 . . . . . 16
E 组：挑战与证明题 . . . . . 16

**11 练习题参考解答** **17**
A 组解答 . . . . . 17
B 组解答 . . . . . 17
C 组解答 . . . . . 18
D 组解答 . . . . . 18
E 组解答 . . . . . 19

**12 本章总结图谱** **20**

**下一章预告：变量替换与 Jacobian** **20**

# Chapter 2 总览

**本章目标**

学完本章，你应该达到以下目标：

1. 能把二重积分的“区域累加”推广为三重积分的“空间累加”；
2. 能识别长方体区域、柱状区域、四面体区域以及一般三维区域；
3. 能用投影区域和截面法写出三重积分限；
4. 能理解 Fubini 定理在三维中的意义：多维聚合可以按不同顺序完成；
5. 能把三重积分解释为体积、总量、平均值和联合密度下的期望；
6. 能从三维联合密度中推出二维或一维边际密度；
7. 能把三重积分用于多维异质性家庭、三资产风险、三因素模型和计量中的联合分布问题。

**直觉**

二重积分是在问：平面区域里每个小面积块的贡献加起来是多少？

$$\iint_D f(x,y)\,\mathrm{d}A.$$

三重积分是在问：空间区域里每个小体积块的贡献加起来是多少？

$$\iiint_E f(x,y,z)\,\mathrm{d}V.$$

如果 $(y,e,a)$ 分别表示收入、教育和财富，那么三重积分就是在整个“家庭类型空间”中加总消费、税负、福利或风险暴露。

## 1 本章为什么重要

上一章的二重积分已经让我们从“一维类型”进入“二维类型”。但是现代经济金融模型很少只有两个维度。一个家庭可能由收入、财富、教育、年龄、风险偏好、生产率共同描述；一个资产组合可能暴露于市场因子、行业因子、利率因子、汇率因子；一个计量模型可能需要处理多个随机变量的联合分布。

三重积分是进入高维分析之前最好的训练场。它不会一开始就把你扔到抽象测度论里，而是用三维空间帮助你形成几个重要能力：

- 看懂一个多维区域由哪些约束定义；

- 知道先固定哪些变量，再对剩下的变量积分；
- 能把联合密度压缩成边际密度；
- 能把一个多维模型中的总量、均值和条件对象写成积分；
- 能理解后面概率、计量、金融风险中反复出现的“integrate out”的含义。

## 1.1 经济金融中的典型三维对象

| 三维变量 | 经济金融含义 |
|---|---|
| $(y, e, a)$ | 收入、教育、财富构成家庭类型；可以研究总体消费、税负、福利分布。 |
| $(r_1, r_2, r_3)$ | 三只资产收益率；可以研究组合收益、协方差结构和尾部风险。 |
| $(m, s, i)$ | 市场因子、行业因子、利率因子；可以研究资产收益的多因子暴露。 |
| $(x, p, \theta)$ | 消费量、价格、偏好参数；可以研究异质消费者需求聚合。 |
| $(Y, D, X)$ | 结果变量、处理变量、协变量；可以研究计量模型中的联合分布和条件期望。 |

## 1.2 最小前置知识

你需要回忆：

1. 二重积分的区域写法：

$$\iint_D f(x,y)\,\mathrm{d}A = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$

2. Type I / Type II 区域的判断方法；
3. 基本平面图像：直线、三角形、圆、抛物线；
4. 基础概率语言：联合密度、边际密度、期望；
5. 简单线性代数直觉：多个变量可以写成向量。

# 2 从二重积分到三重积分

## 2.1 长方体上的三重积分

先从最简单的空间区域开始：

$$B = [a,b] \times [c,d] \times [p,q].$$

如果函数 $f(x,y,z)$ 表示点 $(x,y,z)$ 处单位体积的贡献，那么在长方体 $B$ 上的总贡献是：

$$\iiint_B f(x,y,z)\,\mathrm{d}V.$$

用直角坐标写作：

$$\int_a^b \int_c^d \int_p^q f(x,y,z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x.$$

**定义 2.1：三重积分的直观定义**

设 $E \subset \mathbb{R}^3$ 是空间区域，$f : E \to \mathbb{R}$ 是函数。把 $E$ 分成许多小体积块 $\Delta V_k$，在每块中取样点 $(x_k^*, y_k^*, z_k^*)$。如果 Riemann 和

$$\sum_{k=1}^{n} f(x_k^*, y_k^*, z_k^*)\Delta V_k$$

在切分越来越细时收敛到同一个极限，就称这个极限为 $f$ 在 $E$ 上的三重积分，记作

$$\iiint_E f(x,y,z)\,\mathrm{d}V.$$

**直觉**

三重积分中的 $\mathrm{d}V$ 可以想成一个很小的体积单位。若用直角坐标小盒子近似，则

$$\mathrm{d}V = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z.$$

所以三重积分就是：每个小盒子的“单位体积贡献”乘以“小盒子体积”，然后全部加总。

## 2.2 为什么积分次序可以交换

如果 $f$ 在长方体上足够良好，例如连续，那么三重积分可以按不同顺序计算：

$$\int_a^b \int_c^d \int_p^q f(x,y,z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x = \int_p^q \int_c^d \int_a^b f(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z.$$

三维中一共有 $3! = 6$ 种变量顺序：

$$\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z,\quad \mathrm{d}x\,\mathrm{d}z\,\mathrm{d}y,\quad \mathrm{d}y\,\mathrm{d}x\,\mathrm{d}z,\quad \mathrm{d}y\,\mathrm{d}z\,\mathrm{d}x,\quad \mathrm{d}z\,\mathrm{d}x\,\mathrm{d}y,\quad \mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x.$$

**定理 2.1：三维 Fubini 定理的实用版本**

若 $f$ 在长方体 $B = [a,b] \times [c,d] \times [p,q]$ 上连续，则

$$\iiint_B f(x,y,z)\,\mathrm{d}V$$

可以写成任意一种迭代积分，且结果相同。

**证明思路**。连续函数在紧长方体上可积。先把三维区域切成小盒子，对所有小盒子的贡献求和。无论先按 $z$ 层加、再按 $y$ 条加、最后按 $x$ 方向加，还是换其他顺序，本质上都是在加同一批小盒子。极限存在且唯一时，累加顺序不改变结果。这就是 Fubini 定理的直觉。严格证明需要 Riemann 积分或 Lebesgue 积分理论。

**例 2.1：长方体上的简单三重积分**

计算

$$I=\int_0^1\int_0^2\int_0^3 (x+y+z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x.$$

先对 $z$ 积分：

$$\int_0^3 (x+y+z)\,\mathrm{d}z = 3x+3y+\frac{9}{2}.$$

再对 $y$ 积分：

$$\int_0^2 \left(3x+3y+\frac{9}{2}\right)\mathrm{d}y = 6x+6+9 = 6x+15.$$

最后对 $x$ 积分：

$$I=\int_0^1 (6x+15)\,\mathrm{d}x = 3+15 = 18.$$

# 3 一般三维区域：投影与截面

三重积分真正困难的地方，不是算积分，而是写积分限。经济模型中也类似：难的不是对给定公式求导，而是知道约束集合到底是什么。

## 3.1 第一种常用写法：先投影到 $xy$ 平面

很多三维区域可以写成：

$$E=\{(x,y,z):(x,y)\in D,\ u(x,y)\le z\le v(x,y)\}.$$

这里 $D$ 是 $E$ 在 $xy$ 平面上的投影区域。此时：

$$\iiint_E f(x,y,z)\,\mathrm{d}V=\iint_D\int_{u(x,y)}^{v(x,y)} f(x,y,z)\,\mathrm{d}z\,\mathrm{d}A.$$

如果 $D$ 又是 Type I 区域：

$$D=\{(x,y):a\le x\le b,\ g_1(x)\le y\le g_2(x)\},$$

则

$$\iiint_E f\,\mathrm{d}V=\int_a^b\int_{g_1(x)}^{g_2(x)}\int_{u(x,y)}^{v(x,y)} f(x,y,z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x.$$

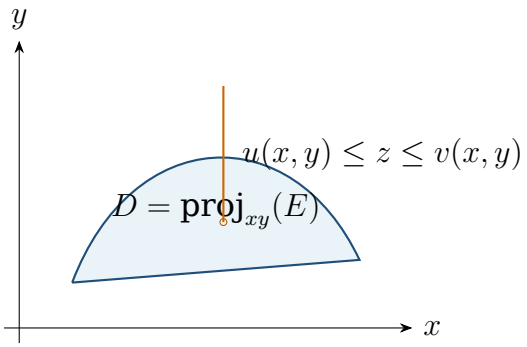



图 1：把三维区域投影到 $xy$ 平面，再对每个 $(x,y)$ 上方的竖直线段积分。

**方法**

写三重积分限的标准流程：

1. 先决定最里面积分变量，例如先对 $z$ 积分；
2. 固定外层变量，例如固定 $(x,y)$；
3. 找 $z$ 的下界和上界；
4. 把三维区域投影到剩下两个变量所在平面；
5. 把这个二维投影区域再写成二重积分限。

## 3.2 另外两种投影

除了投影到 $xy$ 平面，也可以投影到 $xz$ 或 $yz$ 平面：

$$E = \{(x,y,z) : (x,z) \in D_{xz},\ a(x,z) \le y \le b(x,z)\},$$

则

$$\iiint_E f \,\mathrm{d}V = \iint_{D_{xz}} \int_{a(x,z)}^{b(x,z)} f(x,y,z)\,\mathrm{d}y\,\mathrm{d}A_{xz}.$$

或者

$$E = \{(x,y,z) : (y,z) \in D_{yz},\ c(y,z) \le x \le d(y,z)\},$$

则

$$\iiint_E f \,\mathrm{d}V = \iint_{D_{yz}} \int_{c(y,z)}^{d(y,z)} f(x,y,z)\,\mathrm{d}x\,\mathrm{d}A_{yz}.$$

## 3.3 经济解释：先按哪个维度聚合？

假设家庭由收入 $y$、教育 $e$、财富 $a$ 描述，消费函数为 $c(y,e,a)$，联合密度为 $f(y,e,a)$。总消费为：

$$C = \iiint c(y,e,a) f(y,e,a)\,\mathrm{d}y\,\mathrm{d}e\,\mathrm{d}a.$$

如果先对财富 $a$ 积分，就是先在每个 $(y,e)$ 组内把不同财富家庭聚合：

$$C=\iint\left[\int c(y,e,a)f(y,e,a)\,\mathrm{d}a\right]\,\mathrm{d}y\,\mathrm{d}e.$$

如果先对教育 $e$ 积分，就是先在每个 $(y,a)$ 组内聚合不同教育水平。数学上的积分次序，对应经济上的“先按哪个维度分组”。

# 4 截面法：把三维问题切成二维问题

三重积分还可以用截面法理解。固定一个变量，比如固定 $z$，三维区域 $E$ 与平面 $z=t$ 的交集是一个二维截面：

$$E_t=\{(x,y):(x,y,t)\in E\}.$$

那么体积可以写成：

$$\mathrm{Vol}(E)=\int \mathrm{Area}(E_t)\,\mathrm{d}t.$$

更一般地：

$$\iiint_E f(x,y,z)\,\mathrm{d}V=\int\left[\iint_{E_t} f(x,y,t)\,\mathrm{d}x\,\mathrm{d}y\right]\,\mathrm{d}t.$$

**直觉**

截面法就是把一个三维对象切成很多薄片。每一片是二维问题；所有薄片叠起来，就是三维问题。宏观经济中的时间截面、金融中的状态截面、计量中的条件分布，都可以用类似思想理解。

**例 2.2：四面体的体积**

设

$$E=\{(x,y,z):x\ge 0,y\ge 0,z\ge 0,x+y+z\le 1\}.$$

求 $E$ 的体积。

这个区域由坐标平面和斜平面 $x+y+z=1$ 围成。先投影到 $xy$ 平面：

$$D=\{(x,y):x\ge 0,y\ge 0,x+y\le 1\}.$$

对给定 $(x,y)\in D$，有

$$0\le z\le 1-x-y.$$

所以

$$\mathrm{Vol}(E)=\int_0^1\int_0^{1-x}\int_0^{1-x-y}1\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x.$$

先对 $z$ 积分：

$$\int_0^{1-x-y}1\,\mathrm{d}z=1-x-y.$$

于是

$$\mathrm{Vol}(E)=\int_0^1\int_0^{1-x}(1-x-y)\,\mathrm{d}y\,\mathrm{d}x.$$

计算内层:

$$\int_0^{1-x}(1-x-y)\,\mathrm{d}y=(1-x)^2-\frac{(1-x)^2}{2}=\frac{(1-x)^2}{2}.$$

所以

$$\mathrm{Vol}(E)=\frac{1}{2}\int_0^1(1-x)^2\,\mathrm{d}x=\frac{1}{6}.$$

# 5 三重积分的几种核心含义

## 5.1 含义一：体积

如果 $f(x,y,z)=1$，则

$$\iiint_E 1\,\mathrm{d}V=\mathrm{Vol}(E).$$

这和二重积分中

$$\iint_D 1\,\mathrm{d}A=\mathrm{Area}(D)$$

完全对应。

## 5.2 含义二：总量

如果 $\rho(x,y,z)$ 表示单位体积的密度，那么总量为

$$M=\iiint_E \rho(x,y,z)\,\mathrm{d}V.$$

经济学中，$\rho$ 可以不是物理密度，而是“类型密度”“家庭密度”“企业密度”。如果 $g(x,y,z)$ 表示某类个体的贡献，$h(x,y,z)$ 表示该类型的密度，则总贡献为:

$$G=\iiint_E g(x,y,z)h(x,y,z)\,\mathrm{d}V.$$

## 5.3 含义三：平均值

在区域 $E$ 上，函数 $f$ 的平均值为:

$$\bar{f}=\frac{1}{\mathrm{Vol}(E)}\iiint_E f(x,y,z)\,\mathrm{d}V.$$

如果 $E$ 是家庭类型空间，$f$ 是消费或效用，这就是类型均值。

### 5.4 含义四：期望

如果随机向量 $(X,Y,Z)$ 的联合密度是 $f_{X,Y,Z}(x,y,z)$，那么任意函数 $g(X,Y,Z)$ 的期望为：

$$\mathbb{E}[g(X,Y,Z)] = \iiint_{\mathbb{R}^3} g(x,y,z) f_{X,Y,Z}(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z.$$

这是本章最重要的概率化解释。

**常见误区**

不要把三重积分只理解为几何体积。对经济金融来说，更常见的解释是：

三重积分 = 在三维类型或状态空间上的加权平均 / 加总。

其中权重通常就是联合密度。

## 6 联合密度、边际密度与条件分布

这一节是本章通向概率、计量和金融的关键。很多同学学三重积分时只会算体积，却没有意识到：边际密度本质上就是把不关心的变量积分掉。

### 6.1 三维联合密度

**定义 2.2：三维联合密度**

若随机向量 $(X,Y,Z)$ 的联合密度为 $f_{X,Y,Z}(x,y,z)$，则它满足：

$$f_{X,Y,Z}(x,y,z) \geq 0,$$

并且

$$\iiint_{\mathbb{R}^3} f_{X,Y,Z}(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = 1.$$

对任意区域 $E \subset \mathbb{R}^3$，有

$$\Pr((X,Y,Z)\in E) = \iiint_E f_{X,Y,Z}(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z.$$

### 6.2 从联合密度到边际密度

如果只关心 $X$，就把 $Y,Z$ 积分掉：

$$f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\,\mathrm{d}y\,\mathrm{d}z.$$

如果关心 $(X,Y)$，就把 $Z$ 积分掉：

$$f_{X,Y}(x,y) = \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\,\mathrm{d}z.$$

**方法**

边际化 marginalization 的核心动作就是：

保留关心的变量，积分掉不关心的变量。

计量里说的 control variables、金融里说的 integrating over states、宏观里说的 aggregate over heterogeneity，本质上都与这个动作有关。

## 6.3 条件密度

如果 $f_{Y,Z}(y,z)>0$，条件密度为：

$$f_{X\mid Y,Z}(x\mid y,z)=\frac{f_{X,Y,Z}(x,y,z)}{f_{Y,Z}(y,z)}.$$

于是条件期望可以写成：

$$\mathbb{E}[X\mid Y=y,Z=z]=\int_{-\infty}^{\infty} x f_{X\mid Y,Z}(x\mid y,z)\,\mathrm{d}x.$$

这个式子在计量经济学里非常重要。回归函数 $\mathbb{E}[Y\mid X=x]$ 就是条件期望的一个特例。后面 Chapter 22 会专门讲。

**例 2.3：三维联合密度的归一化与边际密度**

设 $(X,Y,Z)$ 的联合密度在区域

$$0\le x\le 1,\quad 0\le y\le 1,\quad 0\le z\le 1$$

上为

$$f(x,y,z)=C(x+y+z),$$

其他地方为 0。求常数 $C$，并求 $f_X(x)$。
因为密度总积分为 1：

$$1=\int_0^1\int_0^1\int_0^1 C(x+y+z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x.$$

注意

$$\int_0^1\int_0^1\int_0^1 x\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x=\frac{1}{2},$$

同理 $y$ 和 $z$ 的积分也各为 $1/2$。所以

$$1=C\left(\frac{1}{2}+\frac{1}{2}+\frac{1}{2}\right)=\frac{3}{2}C,$$

因此

$$C=\frac{2}{3}.$$

边际密度为

$$f_X(x)=\int_0^1\int_0^1 \frac{2}{3}(x+y+z)\,\mathrm{d}z\,\mathrm{d}y.$$

计算：

$$\int_0^1\int_0^1 (x+y+z)\,\mathrm{d}z\,\mathrm{d}y = x+\frac{1}{2}+\frac{1}{2}=x+1.$$

所以

$$f_X(x)=\frac{2}{3}(x+1),\quad 0\le x\le 1.$$

# 7 经济金融应用：从三维积分到模型语言

## 7.1 应用一：三维家庭异质性与总体消费

假设家庭类型由收入 $y$、教育 $e$、财富 $a$ 描述。家庭消费函数为

$$c(y,e,a)=\alpha y+\beta a+\gamma e,$$

联合密度为 $f(y,e,a)$。总体平均消费为

$$\bar{C}=\iiint c(y,e,a)f(y,e,a)\,\mathrm{d}y\,\mathrm{d}e\,\mathrm{d}a.$$

如果政策只影响收入 $y$，例如税后收入变为 $(1-\tau)y$，那么政策后的平均消费为

$$\bar{C}(\tau)=\iiint [\alpha(1-\tau)y+\beta a+\gamma e]f(y,e,a)\,\mathrm{d}y\,\mathrm{d}e\,\mathrm{d}a.$$

税率变化的边际影响为

$$\frac{\mathrm{d}\bar{C}}{\mathrm{d}\tau}=-\alpha\iiint yf(y,e,a)\,\mathrm{d}y\,\mathrm{d}e\,\mathrm{d}a=-\alpha\mathbb{E}[Y].$$

这说明三重积分不仅能算总量，还能帮助做政策比较静态。

## 7.2 应用二：三资产收益与组合风险

设三只资产收益为 $(R_1,R_2,R_3)$，联合密度为 $f(r_1,r_2,r_3)$。组合权重为 $w=(w_1,w_2,w_3)$，组合收益为

$$R_p=w_1R_1+w_2R_2+w_3R_3.$$

组合期望收益为

$$\mathbb{E}[R_p]=\iiint (w_1r_1+w_2r_2+w_3r_3)f(r_1,r_2,r_3)\,\mathrm{d}r_1\,\mathrm{d}r_2\,\mathrm{d}r_3.$$

组合方差为

$$\mathrm{Var}(R_p)=\iiint (w_1r_1+w_2r_2+w_3r_3-\mathbb{E}[R_p])^2\,f(r_1,r_2,r_3)\,\mathrm{d}r_1\,\mathrm{d}r_2\,\mathrm{d}r_3.$$

后面 Chapter 24 会把这个表达压缩成更简洁的矩阵形式：

$$\mathrm{Var}(R_p)=w'\Sigma w.$$

### 7.3 应用三：计量中的联合分布和条件期望

设 $Y$ 是结果变量，$D$ 是处理变量，$X$ 是协变量。联合密度为 $f_{Y,D,X}(y,d,x)$。如果想知道在 $D=d, X=x$ 条件下的平均结果，则：

$$\mathbb{E}[Y \mid D=d, X=x] = \int y f_{Y|D,X}(y \mid d,x)\,\mathrm{d}y.$$

其中

$$f_{Y|D,X}(y \mid d,x) = \frac{f_{Y,D,X}(y,d,x)}{f_{D,X}(d,x)}.$$

而

$$f_{D,X}(d,x) = \int f_{Y,D,X}(y,d,x)\,\mathrm{d}y.$$

这就是“先从联合分布积分得到边际分布，再形成条件分布”的基本逻辑。

## 8 Worked Examples：完整样板题

**Worked Example 1：写出空间区域的三重积分**

设

$$E = \{(x,y,z) : 0 \le x \le 1, 0 \le y \le x, 0 \le z \le x+y\}.$$

写出 $\iiint_E (x+2y+3z)\,\mathrm{d}V$，并计算其值。

**解**。题目已经给出了很自然的积分顺序：先固定 $x,y$，再让 $z$ 从 0 到 $x+y$。因此：

$$I = \int_0^1 \int_0^x \int_0^{x+y} (x+2y+3z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x.$$

先对 $z$ 积分：

$$\int_0^{x+y} (x+2y+3z)\,\mathrm{d}z = (x+2y)(x+y) + \frac{3}{2}(x+y)^2.$$

展开：

$$(x+2y)(x+y) = x^2 + 3xy + 2y^2,$$
$$\frac{3}{2}(x+y)^2 = \frac{3}{2}x^2 + 3xy + \frac{3}{2}y^2.$$

所以内层积分为：

$$\frac{5}{2}x^2 + 6xy + \frac{7}{2}y^2.$$

再对 $y$ 从 0 到 $x$ 积分：

$$\int_0^x \left(\frac{5}{2}x^2 + 6xy + \frac{7}{2}y^2\right)\mathrm{d}y = \frac{5}{2}x^3 + 3x^3 + \frac{7}{6}x^3 = \frac{20}{3}x^3.$$

最后：

$$I = \int_0^1 \frac{20}{3}x^3\,\mathrm{d}x = \frac{5}{3}.$$

### Worked Example 2：用截面法求体积

设

$$E=\{(x,y,z):x^2+y^2\le z\le 1\}.$$

求 $E$ 的体积。

**解**。条件 $x^2+y^2\le z\le 1$ 表示区域位于抛物面 $z=x^2+y^2$ 上方、平面 $z=1$ 下方。投影到 $xy$ 平面，需要满足

$$x^2+y^2\le 1.$$

于是

$$\mathrm{Vol}(E)=\iint_{x^2+y^2\le 1}\int_{x^2+y^2}^{1}1\,\mathrm{d}z\,\mathrm{d}A.$$

先对 $z$ 积分：

$$\int_{x^2+y^2}^{1}1\,\mathrm{d}z=1-x^2-y^2.$$

由于投影是单位圆，使用极坐标更方便：

$$\mathrm{Vol}(E)=\int_0^{2\pi}\int_0^1(1-r^2)r\,\mathrm{d}r\,\mathrm{d}\theta.$$

注意这里的 $r$ 来自 Jacobian，下一章会系统讲。计算：

$$\int_0^1(1-r^2)r\,\mathrm{d}r=\int_0^1(r-r^3)\,\mathrm{d}r=\frac12-\frac14=\frac14.$$

所以

$$\mathrm{Vol}(E)=2\pi\cdot\frac14=\frac{\pi}{2}.$$

### Worked Example 3：从联合密度求边际与期望

设 $(X,Y,Z)$ 在单位立方体 $[0,1]^3$ 上的联合密度为

$$f(x,y,z)=8xyz.$$

求 $\mathbb{E}[X]$ 与 $f_{X,Y}(x,y)$。

**解**。先检查它确实是密度：

$$\int_0^1\int_0^1\int_0^1 8xyz\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x=8\cdot\frac12\cdot\frac12\cdot\frac12=1.$$

求 $\mathbb{E}[X]$：

$$\mathbb{E}[X]=\int_0^1\int_0^1\int_0^1 x\cdot 8xyz\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x=8\int_0^1x^2\,\mathrm{d}x\int_0^1y\,\mathrm{d}y\int_0^1z\,\mathrm{d}z.$$

所以

$$\mathbb{E}[X] = 8 \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{2}{3}.$$

二维边际密度为：

$$f_{X,Y}(x,y) = \int_0^1 8xyz \,\mathrm{d}z = 4xy, \quad 0 \le x, y \le 1.$$

# 9 计算方法总结

**方法**

处理三重积分时，按以下顺序思考：

1. **先判断含义**：是在算体积、总量、平均值，还是概率/期望？
2. **看区域约束**：区域是盒子、四面体、柱状体，还是曲面围成？
3. **选择投影平面**：投影到 $xy$、$xz$ 还是 $yz$？
4. **确定最内层变量上下界**：固定外层变量后，最内层变量从哪里到哪里？
5. **写二维投影区域**：把剩下两个变量的范围写成二重积分限；
6. **计算或解释**：能算就算，不能算也要能解释它表示什么。

**常见误区**

三重积分最常见的错误有四类：

1. 忘记投影区域的约束，只写出最里面变量的上下界；
2. 积分次序改变后，积分限没有同步改变；
3. 把边际密度写成“直接删掉变量”，而不是积分掉变量；
4. 在概率题中忘记先检查密度是否归一化。

# 10 练习题

## A 组：基础计算

A1. 计算

$$\int_0^1 \int_0^1 \int_0^1 (x+y+z) \,\mathrm{d}z \,\mathrm{d}y \,\mathrm{d}x.$$

A2. 计算

$$\int_0^2 \int_0^x \int_0^y xyz \,\mathrm{d}z \,\mathrm{d}y \,\mathrm{d}x.$$

A3. 求区域

$$E = \{(x, y, z) : 0 \le x \le 1, 0 \le y \le 2, 0 \le z \le 3\}$$

的体积。

A4. 求四面体

$$E = \{(x, y, z) : x, y, z \ge 0, x + y + z \le 2\}$$

的体积。

A5. 设

$$E = \{(x, y, z) : 0 \le x \le 1, 0 \le y \le x^2, 0 \le z \le y\}.$$

写出并计算 $\iiint_E 1 \,\mathrm{d}V$。

## B 组：概念判断与积分限

B1. 用文字解释为什么

$$\iiint_E 1 \,\mathrm{d}V$$

等于区域 $E$ 的体积。

B2. 设

$$E = \{(x, y, z) : 0 \le z \le 1 - x - y, x \ge 0, y \ge 0, x + y \le 1\}.$$

说明 $E$ 投影到 $xy$ 平面是什么区域。

B3. 设

$$E = \{(x, y, z) : x^2 + y^2 \le 1, 0 \le z \le x^2 + y^2\}.$$

写出 $\iiint_E f(x, y, z) \,\mathrm{d}V$ 的迭代积分形式。

B4. 判断以下说法是否正确：如果 $f_{X,Y,Z}$ 是联合密度，那么

$$f_X(x) = f_{X,Y,Z}(x, y, z)$$

只要把 $y, z$ 去掉即可。

B5. 为什么三维 Fubini 定理在经济学上可以理解为“分组聚合顺序不改变总量”？

## C 组：混合题

C1. 设

$$E = \{(x, y, z) : 0 \le x \le 1, 0 \le z \le 1, 0 \le y \le x + z\}.$$

计算 $\iiint_E y \,\mathrm{d}V$。

C2. 设 $(X, Y, Z)$ 在 $[0,1]^3$ 上的联合密度为

$$f(x, y, z) = Cxyz.$$

求 $C$。

C3. 在上一题中，求 $f_X(x)$。

C4. 设

$$f(x, y, z) = 6z$$

在区域 $0 \le x \le 1, 0 \le y \le 1, 0 \le z \le 1$ 上声称为联合密度。先判断是否合法；再求 `E[Z]`。

C5. 设三只资产收益 $(R_1, R_2, R_3)$ 的联合密度为 $f$，组合收益为 $R_p = w_1R_1 + w_2R_2 + w_3R_3$。写出 $\mathbb{E}[R_p]$ 的三重积分形式。

## D 组：经济金融建模题

D1. 家庭类型为 $(y, e, a)$，联合密度为 $f(y, e, a)$，消费函数为 $c(y, e, a)$。写出平均消费。

D2. 若政府征收收入税 $\tau y$，家庭税负函数为 $T(y, e, a) = \tau y$。写出平均税负，并说明它等于什么期望。

D3. 设企业类型为 $(k, l, z)$，分别表示资本、劳动和生产率。产出为 $F(k, l, z) = zk^{\alpha}l^{1-\alpha}$，联合密度为 $h(k, l, z)$。写出总产出。

D4. 设 $(Y, D, X)$ 的联合密度为 $f_{Y,D,X}$。写出 $f_{D,X}(d, x)$，并解释它的含义。

D5. 设投资者效用依赖于三种状态收益 $(r_1, r_2, r_3)$，效用函数为 $u(r_1, r_2, r_3)$，联合密度为 $f$。写出期望效用。

## E 组：挑战与证明题

E1. 证明若 $(X, Y, Z)$ 独立，且边际密度分别为 $f_X, f_Y, f_Z$，则

$$\mathbb{E}[g(X)h(Y)q(Z)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]\mathbb{E}[q(Z)].$$

E2. 设

$$E = \{(x, y, z) : 0 \le z \le 1, x^2 + y^2 \le z\}.$$

用截面法求 $E$ 的体积。

E3. 设 $(X, Y, Z)$ 的联合密度为 $f$。证明二维边际密度 $f_{X,Y}(x, y) = \int f(x, y, z)\,\mathrm{d}z$ 满足

$$\iint f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1.$$

E4. 给出一个经济学例子，说明为什么“先对收入聚合再对教育聚合”和“先对教育聚合再对收入聚合”虽然顺序不同，但总量应相同。

E5. 思考题：当函数不绝对可积时，积分次序可能影响结果。这对经济建模有什么警示?

# 11 练习题参考解答

## A 组解答

A1.
$$\int_0^1\int_0^1\int_0^1 (x+y+z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x = \frac{1}{2}+\frac{1}{2}+\frac{1}{2} = \frac{3}{2}.$$

A2. 先对 $z$ 积分：
$$\int_0^y xyz\,\mathrm{d}z = xy\frac{y^2}{2} = \frac{1}{2}xy^3.$$
再对 $y$：
$$\int_0^x \frac{1}{2}xy^3\,\mathrm{d}y = \frac{1}{2}x\frac{x^4}{4} = \frac{1}{8}x^5.$$
最后：
$$\int_0^2 \frac{1}{8}x^5\,\mathrm{d}x = \frac{1}{8}\cdot\frac{64}{6} = \frac{4}{3}.$$

A3. 长方体体积为
$$1\cdot 2\cdot 3 = 6.$$

A4. 区域为边长截距为 2 的三维四面体。直接积分：
$$\mathrm{Vol}(E) = \int_0^2\int_0^{2-x}\int_0^{2-x-y} 1\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x = \frac{2^3}{6} = \frac{4}{3}.$$

A5.
$$\iiint_E 1\,\mathrm{d}V = \int_0^1\int_0^{x^2}\int_0^{y} 1\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x = \int_0^1\int_0^{x^2} y\,\mathrm{d}y\,\mathrm{d}x.$$
因此
$$= \int_0^1 \frac{x^4}{2}\,\mathrm{d}x = \frac{1}{10}.$$

## B 组解答

B1. 因为 $1\,\mathrm{d}V$ 表示每个小体积块的体积贡献。把所有小体积块加起来，就是整个区域的体积。

B2. 投影到 $xy$ 平面后，去掉 $z$ 变量，留下约束：
$$x\ge 0,\quad y\ge 0,\quad x+y\le 1.$$
这是第一象限内由坐标轴和直线 $x+y=1$ 围成的三角形。

B3. 可写为
$$\iiint_E f\,\mathrm{d}V = \iint_{x^2+y^2\le 1}\int_0^{x^2+y^2} f(x,y,z)\,\mathrm{d}z\,\mathrm{d}A.$$
若下一章使用极坐标，则也可写成
$$\int_0^{2\pi}\int_0^1\int_0^{r^2} f(r\cos\theta, r\sin\theta, z)\,r\,\mathrm{d}z\,\mathrm{d}r\,\mathrm{d}\theta.$$

B4. 错。边际密度不是“删掉变量”，而是把不关心的变量积分掉：

$$f_X(x) = \int\int f_{X,Y,Z}(x,y,z)\,\mathrm{d}y\,\mathrm{d}z.$$

B5. 因为如果函数足够良好，总量是由所有微小类型块的贡献加总得到的。先按收入组求和再按教育组求和，和先按教育组求和再按收入组求和，都是对同一批个体加总。

## C 组解答

C1.

$$I = \int_0^1\int_0^1\int_0^{x+z} y\,\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}x = \int_0^1\int_0^1 \frac{(x+z)^2}{2}\,\mathrm{d}z\,\mathrm{d}x.$$

展开：

$$\frac{1}{2}\int_0^1\int_0^1 (x^2+2xz+z^2)\,\mathrm{d}z\,\mathrm{d}x = \frac{1}{2}\left(\frac{1}{3}+\frac{1}{2}+\frac{1}{3}\right) = \frac{7}{12}.$$

C2.

$$1 = \int_0^1\int_0^1\int_0^1 Cxyz\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x = C\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{C}{8}.$$

所以 $C=8$。

C3.

$$f_X(x) = \int_0^1\int_0^1 8xyz\,\mathrm{d}y\,\mathrm{d}z = 8x\cdot\frac{1}{2}\cdot\frac{1}{2} = 2x.$$

C4.

$$\mathbb{E}[Z] = \int_0^1\int_0^1\int_0^1 z\cdot 6z\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x = 6\int_0^1 z^2\,\mathrm{d}z = 2.$$

但这里出现问题：$f=6z$ 在单位立方体上的积分为

$$\int_0^1\int_0^1\int_0^1 6z\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x = 3,$$

不是 1，所以它不是合法密度。若把密度改为 $f=2z$，则 $\mathbb{E}[Z]=2/3$。本题的目的就是提醒你先检查归一化。

C5.

$$\mathbb{E}[R_p] = \iiint (w_1r_1+w_2r_2+w_3r_3)f(r_1,r_2,r_3)\,\mathrm{d}r_1\,\mathrm{d}r_2\,\mathrm{d}r_3.$$

## D 组解答

D1. 平均消费为

$$\bar{C} = \iiint c(y,e,a)f(y,e,a)\,\mathrm{d}y\,\mathrm{d}e\,\mathrm{d}a.$$

D2. 平均税负为

$$\bar{T} = \iiint \tau y f(y,e,a)\,\mathrm{d}y\,\mathrm{d}e\,\mathrm{d}a = \tau\mathbb{E}[Y].$$

D3. 总产出为

$$Q = \iiint z k^{\alpha} l^{1-\alpha} h(k,l,z)\,\mathrm{d}k\,\mathrm{d}l\,\mathrm{d}z.$$

D4.

$$f_{D,X}(d,x) = \int f_{Y,D,X}(y,d,x)\,\mathrm{d}y.$$

它表示 $(D,X)$ 的二维边际密度，即不关心 $Y$ 后，处理变量和协变量本身的联合分布。

D5. 期望效用为

$$\mathbb{E}[u] = \iiint u(r_1,r_2,r_3) f(r_1,r_2,r_3)\,\mathrm{d}r_1\,\mathrm{d}r_2\,\mathrm{d}r_3.$$

## E 组解答

E1. 若独立，则

$$f_{X,Y,Z}(x,y,z) = f_X(x) f_Y(y) f_Z(z).$$

所以

$$\mathbb{E}[g(X)h(Y)q(Z)] = \iiint g(x)h(y)q(z) f_X(x) f_Y(y) f_Z(z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z.$$

由 Fubini 定理可分解为

$$\left(\int g(x) f_X(x)\,\mathrm{d}x\right)\left(\int h(y) f_Y(y)\,\mathrm{d}y\right)\left(\int q(z) f_Z(z)\,\mathrm{d}z\right),$$

即

$$\mathbb{E}[g(X)]\mathbb{E}[h(Y)]\mathbb{E}[q(Z)].$$

E2. 固定 $z$，截面为

$$x^2 + y^2 \le z,$$

这是半径 $\sqrt{z}$ 的圆，面积为 $\pi z$。因此

$$\mathrm{Vol}(E) = \int_0^1 \pi z\,\mathrm{d}z = \frac{\pi}{2}.$$

E3.

$$\iint f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \iint \left[\int f(x,y,z)\,\mathrm{d}z\right]\mathrm{d}x\,\mathrm{d}y.$$

由 Fubini 定理，等于

$$\iiint f(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = 1.$$

E4. 例如全国平均消费可写为

$$\iint c(y,e) f(y,e)\,\mathrm{d}y\,\mathrm{d}e.$$

先对收入 $y$ 聚合，是在每个教育组内先加总收入异质性；先对教育 $e$ 聚合，是在每个收入组内先加总教育异质性。只要函数和密度足够良好，两者都是对所有家庭的同一批贡献加总，所以总量相同。

E5. 警示是：经济模型中的“总量”表达需要良好的可积性条件。如果尾部太厚、福利函数无界或某些状态下损失极端巨大，那么随意交换积分、期望、极限或求导顺序可能导致错误结论。这也是后面学习 dominated convergence、uniform integrability 和渐近理论时的重要动机。

## 12 本章总结图谱

**本章压缩总结**

| 核心概念 | 你需要记住的内容 |
|---|---|
| **三重积分** | $\iiint_E f(x,y,z)\,\mathrm{d}V$，表示空间区域上的累加。 |
| **投影区域** | 写积分限时常先把三维区域投影到 $xy$、$xz$ 或 $yz$ 平面。 |
| **截面法** | 固定一个变量，把三维问题切成二维薄片再加总。 |
| **体积** | $\iiint_E 1\,\mathrm{d}V = \mathrm{Vol}(E)$。 |
| **期望** | $\mathbb{E}[g(X,Y,Z)] = \iiint g(x,y,z)f(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$。 |
| **边际密度** | 保留关心变量，积分掉不关心变量。 |
| **条件密度** | $f_{X\mid Y,Z} = f_{X,Y,Z}/f_{Y,Z}$。 |
| **经济含义** | 多维异质性、多风险因子、联合分布、分组聚合。 |
| **常见错误** | 忘记投影区域；换序不换限；把边际化误解为删除变量；不检查密度归一化。 |

## 下一章预告：变量替换与 Jacobian

本章中我们已经看到，三维区域有时用直角坐标写积分很麻烦。例如圆柱、球、椭球、多元正态密度、资产收益的线性组合，都更适合换变量。下一章的核心问题是：

换变量以后，面积或体积单位如何变化？

答案就是 Jacobian。你将看到：Jacobian 不只是一个行列式，而是从数学积分通向随机变量变换、极坐标、多元正态和金融收益分布的桥梁。

# 经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

---

# Chapter 3

## 变量替换与 Jacobian

---

本章解释多元积分换元公式为什么必须乘上 Jacobian。对经济金融来说，Jacobian 不只是积分技巧，而是坐标变换、随机变量密度变换、正态标准化、投资组合收益分布和对数收益率建模的共同语言。

math4econ personalized notes
2026

# 目录

**Chapter 3 总览** **3**

**1 本章为什么重要** **4**
1.1 经济金融里为什么经常换变量 . . . . . 4
1.2 从本章开始的难度变化 . . . . . 4

**2 最小前置知识** **4**

**3 从一元换元重新理解 Jacobian** **5**
3.1 一元换元不是魔法 . . . . . 5

**4 二维变量替换：Jacobian 的定义和几何意义** **6**
4.1 二维变换 . . . . . 6
4.2 几何意义：小矩形变成小平行四边形 . . . . . 7
4.3 为什么要取绝对值 . . . . . 8

**5 二维换元定理与标准做题模板** **9**

**6 极坐标：最重要的二维换元** **9**
6.1 极坐标公式 . . . . . 9
6.2 为什么极坐标面积元是 $r\,\mathrm{d}r\,\mathrm{d}\theta$ . . . . . 10

**7 线性变换：椭圆、矩阵和面积缩放** **11**
7.1 线性变换的 Jacobian . . . . . 11
7.2 椭圆区域的标准化 . . . . . 12
7.3 金融中的线性变换：组合收益 . . . . . 13

**8 随机变量变换：密度为什么要除以 Jacobian** **13**
8.1 一维密度变换 . . . . . 14
8.2 二维密度变换公式 . . . . . 14

**9 经济金融中的典型变量变换** **16**
9.1 价格到对数收益：一维 Jacobian . . . . . 16
9.2 Cobb-Douglas 中的比例变换 . . . . . 16
9.3 风险因子标准化 . . . . . 16

**10 三维换元：柱坐标与球坐标** **17**
10.1 三维 Jacobian . . . . . 18
10.2 柱坐标 . . . . . 18

10.3 球坐标 . . . . . 18

**11 Worked Examples：完整样板题 19**

**12 常见错误专题 23**

**13 练习题 24**

练习题 A 组：基础计算 . . . . . 24
练习题 B 组：概念判断 . . . . . 24
练习题 C 组：混合题 . . . . . 24
练习题 D 组：经济金融建模题 . . . . . 25
练习题 E 组：挑战和证明题 . . . . . 25

**14 练习题详细解答 25**

A 组答案 . . . . . 26
B 组答案 . . . . . 27
C 组答案 . . . . . 27
D 组答案 . . . . . 29
E 组答案 . . . . . 30

**15 本章总结图谱 32**

**下一章预告 33**

# Chapter 3 总览

**本章目标**

学完本章，你应该达到以下目标：

1. 从一元换元公式理解二维和三维换元公式的来源；
2. 能解释 Jacobian determinant 的几何意义：局部面积或体积缩放比例；
3. 能熟练使用极坐标、柱坐标、球坐标和线性变换；
4. 能判断什么时候需要取绝对值，什么时候不能忘记区域也要跟着变；
5. 能把换元公式用于椭圆区域、多元正态、lognormal 变量和 Cobb-Douglas 型变量变换；
6. 能处理二维随机变量函数变换，并写出新变量的联合密度；
7. 能在经济金融模型中识别“变量变换”的位置：价格到收益、收益到对数收益、个体类型到充分统计量、风险因子到组合收益。

**直觉**

一句话：**变量替换就是换一个更适合问题的坐标系统；Jacobian 就是新坐标中的一个小格子在原坐标里变成多大。**

如果你只记住公式

$$\iint_R f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \iint_S f(x(u,v),y(u,v))\left|\frac{\partial(x,y)}{\partial(u,v)}\right|\,\mathrm{d}u\,\mathrm{d}v,$$

却不理解 Jacobian 是面积缩放，那么后面学联合密度变换时会非常容易错。比如从价格 $P$ 变成对数收益 $R=\log(P_1/P_0)$，从两个资产收益 $(R_1,R_2)$ 变成组合收益和相对价差 $(W,S)$，本质都是变量替换。

**本章压缩总结**

本章比前两章更详细，因为 Jacobian 是一个“枢纽工具”。前两章的二重、三重积分告诉你如何在多维空间里累加；本章告诉你：如果我们换了坐标，累加单位本身会改变。这个小小的改变，正是极坐标里的 $r$、球坐标里的 $\rho^2\sin\varphi$、多元正态标准化里的 $|\det A|$、随机变量密度变换里的 $|J|^{-1}$ 的共同来源。

# 1 本章为什么重要

在普通高数里，变量替换常常表现为“为了把积分算出来”。比如圆盘区域用极坐标，椭圆区域做线性变换，三维球体用球坐标。这当然重要，但对经济金融来说，它的意义更深。

## 1.1 经济金融里为什么经常换变量

经济金融模型里，原始变量往往不是最方便分析的变量。例如：

- 金融中，价格 $P_t$ 经常换成收益率 $R_t = P_{t+1}/P_t - 1$，或者对数收益 $r_t = \log P_{t+1} - \log P_t$；
- 资产组合中，原变量是多个资产收益 $(R_1, R_2, \dots, R_n)$，但你可能更关心组合收益 $R_p = w'R$ 和风险因子；
- 计量中，原变量可能是 $(Y, X, Z)$，但识别时更关心残差、投影、充分统计量；
- 消费者理论中，原变量是商品数量 $(x_1, x_2)$，但 Cobb-Douglas 或 CES 问题里常常通过比例、总支出、份额来变换；
- 概率论中，原变量 $(X, Y)$ 的联合密度已知，但研究对象可能是 $U = X + Y$、$V = X/Y$ 或 $R = \sqrt{X^2 + Y^2}$。

这些问题都在问同一件事：**换了变量以后，原来的面积、体积、概率质量如何重新表达？**

## 1.2 从本章开始的难度变化

前两章主要是“在给定坐标系中积分”。从本章开始，你要习惯一件事：**坐标本身可以被选择**。选择好的坐标，很多问题会直接变简单；选择不好的坐标，积分区域和密度都会很难处理。

一个典型例子是单位圆盘：

$$R = \{(x, y) : x^2 + y^2 \le 1\}.$$

如果用直角坐标，区域写成

$$-1 \le x \le 1, \qquad -\sqrt{1 - x^2} \le y \le \sqrt{1 - x^2}.$$

但如果用极坐标，区域就是

$$0 \le r \le 1, \qquad 0 \le \theta \le 2\pi.$$

这就是变量替换的力量。

# 2 最小前置知识

本章默认你已经掌握：

1. 一元函数换元积分；
2. 偏导数与全微分；
3. 二重积分和三重积分的基本定义；
4. 矩阵、行列式的基本计算；
5. 二维线性变换对面积的缩放直觉。

其中最关键的是行列式。二维中，两个向量

$$\boldsymbol{a} = (a_1, a_2), \qquad \boldsymbol{b} = (b_1, b_2)$$

张成的平行四边形面积为

$$\left| \det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \right|.$$

三维中，三个向量张成的平行六面体体积为

$$\left| \det \begin{pmatrix} | & | & | \\ \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \\ | & | & | \end{pmatrix} \right|.$$

Jacobian determinant 本质上就是把这个思想应用到“局部线性化”的结果。

# 3 从一元换元重新理解 Jacobian

## 3.1 一元换元不是魔法

先回到最熟悉的一元换元。设

$$x = g(u).$$

则

$$\int_a^b f(x)\,\mathrm{d}x = \int_\alpha^\beta f(g(u))g'(u)\,\mathrm{d}u,$$

其中 $g(\alpha) = a$、$g(\beta) = b$，并且假设 $g$ 单调递增。若不关心方向，长度缩放比例是 $|g'(u)|$。

**直觉**

一元换元中的 $g'(u)$ 不是凭空冒出来的。它表示：$u$ 轴上的一个小长度 $\mathrm{d}u$，经过 $x = g(u)$ 变换后，在 $x$ 轴上大约变成

$$\mathrm{d}x \approx g'(u)\,\mathrm{d}u.$$

所以一维中的导数就是局部长度缩放率。

如果 $x = 2u$，那么 $u$ 方向每走 1 单位，$x$ 方向走 2 单位，所以

$$\mathrm{d}x = 2\,\mathrm{d}u.$$

如果 $x = u^2$，在 $u = 3$ 附近，$g'(3) = 6$，所以一个很短的 $\mathrm{d}u$ 在 $x$ 空间里大约被放大 6 倍。

**一元换元的局部推导**

在 $u$ 附近取一个小区间 $[u, u + \Delta u]$。它在 $x$ 空间中的长度大约为

$$g(u + \Delta u) - g(u) \approx g'(u)\Delta u.$$

如果原积分是把很多小段上的贡献 $f(x)\Delta x$ 加起来，那么替换后每个小段贡献变成

$$f(g(u)) \cdot g'(u)\Delta u.$$

把所有小段加起来并取极限，就得到

$$\int f(x)\,\mathrm{d}x = \int f(g(u))g'(u)\,\mathrm{d}u.$$

二维 Jacobian 的推导完全类似，只是“小区间”变成“小矩形”，“长度缩放”变成“面积缩放”。

# 4 二维变量替换：Jacobian 的定义和几何意义

## 4.1 二维变换

设新变量为 $(u, v)$，旧变量为 $(x, y)$，二者关系为

$$T(u, v) = (x(u, v), y(u, v)).$$

也可以写成

$$\begin{cases} x = x(u, v), \\ y = y(u, v). \end{cases}$$

这表示：$(u, v)$ 平面中的每一个点，被映射到 $(x, y)$ 平面中的一个点。

**Jacobian matrix 与 Jacobian determinant**

二维变换

$$T(u, v) = (x(u, v), y(u, v))$$

的 Jacobian matrix 为

$$DT(u,v)=\begin{pmatrix}\dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v}\\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v}\end{pmatrix}.$$

其行列式

$$J(u,v)=\frac{\partial(x,y)}{\partial(u,v)}=\det DT(u,v)=\frac{\partial x}{\partial u}\frac{\partial y}{\partial v}-\frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

称为 Jacobian determinant。

注意顺序：

$$\frac{\partial(x,y)}{\partial(u,v)}$$

表示“从 $(u,v)$ 变到 $(x,y)$”。如果你反过来写

$$\frac{\partial(u,v)}{\partial(x,y)},$$

它通常是倒数关系：

$$\frac{\partial(u,v)}{\partial(x,y)}=\left(\frac{\partial(x,y)}{\partial(u,v)}\right)^{-1},$$

前提是变换可逆且 Jacobian 非零。

## 4.2 几何意义：小矩形变成小平行四边形

在 $(u,v)$ 平面中，取一个很小的矩形：

$$[u,u+\Delta u]\times[v,v+\Delta v].$$

沿 $u$ 方向移动 $\Delta u$ 时，$(x,y)$ 中的变化近似为

$$\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix}\approx\begin{pmatrix}x_u\\ y_u\end{pmatrix}\Delta u.$$

沿 $v$ 方向移动 $\Delta v$ 时，变化近似为

$$\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix}\approx\begin{pmatrix}x_v\\ y_v\end{pmatrix}\Delta v.$$

因此小矩形在 $(x,y)$ 平面里近似变成由两个向量

$$\boldsymbol{a}=\begin{pmatrix}x_u\\ y_u\end{pmatrix}\Delta u,\qquad \boldsymbol{b}=\begin{pmatrix}x_v\\ y_v\end{pmatrix}\Delta v$$

张成的小平行四边形。它的面积为

$$|\det(\boldsymbol{a},\boldsymbol{b})|=\left|\det\begin{pmatrix}x_u\Delta u & x_v\Delta v\\ y_u\Delta u & y_v\Delta v\end{pmatrix}\right|=|J(u,v)|\Delta u\Delta v.$$

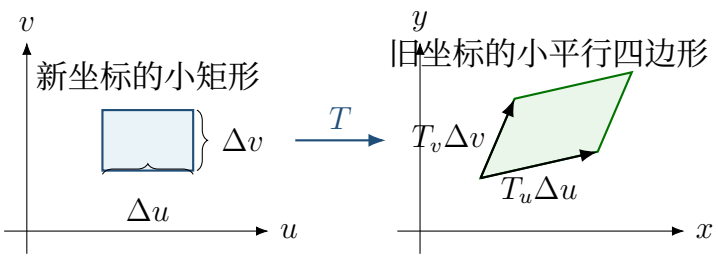



图 1: Jacobian 的几何意义：新坐标的小矩形经过变换后，在旧坐标中近似成为小平行四边形。

**直觉**

你可以把 $|J(u,v)|$ 读成：

$$\text{在}(u,v)\text{ 附近，一个单位面积在}(x,y)\text{ 中变成多大。}$$

如果 $|J|>1$，局部面积被放大；如果 $|J|<1$，局部面积被压缩；如果 $|J|=0$，变换局部塌缩，比如二维区域被压成一条曲线，不能作为正常的二维换元。

## 4.3 为什么要取绝对值

行列式有正负。正负表示方向 orientation 是否被翻转。例如

$$T(u,v)=(v,u)$$

对应矩阵

$$\begin{pmatrix}0&1\\1&0\end{pmatrix},$$

行列式为 $-1$。它没有改变面积大小，只是把方向翻转了。因此积分中的面积元素必须使用

$$\mathrm{d}x\,\mathrm{d}y=|J(u,v)|\,\mathrm{d}u\,\mathrm{d}v.$$

**常见误区**

常见错误：把 $J$ 直接代入，而不是代入 $|J|$。面积、体积、概率质量不能为负，所以换元公式中一定是 Jacobian determinant 的绝对值。只有在有向曲线积分、微分形式等更高阶语境里，方向符号才会被保留。

# 5 二维换元定理与标准做题模板

**二维换元公式**

设 $T : S \to R$ 为一一、连续可微变换,

$$T(u,v) = (x(u,v), y(u,v)),$$

且在区域内部

$$\frac{\partial(x,y)}{\partial(u,v)} \neq 0.$$

在适当正则条件下，有

$$\iint_R f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \iint_S f(x(u,v),y(u,v)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right| \mathrm{d}u\,\mathrm{d}v.$$

这个定理中的条件有三个关键点:

1. **区域也要变**：不是只把 integrand 中的 $x, y$ 换掉，积分区域 $R$ 也要变成 $S$;
2. **Jacobian 非零**：否则局部面积塌缩;
3. **基本一一对应**：否则同一个 $(x, y)$ 可能被重复计算。

**方法**

二维换元的六步模板:

1. 看区域形状：圆、椭圆、三角形、双曲线边界，还是线性约束?
2. 选变换：极坐标、线性变换，或自定义 $u, v$;
3. 写反向或正向关系：明确 $x = x(u,v), y = y(u,v)$;
4. 变区域：把 $R$ 中的边界转成 $S$ 中的边界;
5. 算 Jacobian:
$$J = \frac{\partial(x,y)}{\partial(u,v)};$$
6. 换 integrand 和面积元，最后积分并做合理性检查。

# 6 极坐标：最重要的二维换元

## 6.1 极坐标公式

极坐标定义为

$$x = r\cos\theta, \qquad y = r\sin\theta.$$

Jacobian 为

$$\frac{\partial(x,y)}{\partial(r,\theta)} = \det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r(\cos^2\theta + \sin^2\theta) = r.$$

所以

$$\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\theta.$$

**常见误区**

极坐标里最常见的错误是忘记乘 $r$。如果你写成 $\mathrm{d}x\,\mathrm{d}y = \mathrm{d}r\,\mathrm{d}\theta$，那么圆盘面积会被算成矩形面积。比如单位圆盘面积本该是

$$\int_0^{2\pi}\int_0^1 r\,\mathrm{d}r\,\mathrm{d}\theta = \pi,$$

若忘记 $r$，就会得到 $2\pi$，显然不对。

## 6.2 为什么极坐标面积元是 $r\,\mathrm{d}r\,\mathrm{d}\theta$

在半径 $r$ 附近，径向厚度是 $\mathrm{d}r$，角向弧长约为 $r\,\mathrm{d}\theta$。所以小扇环面积近似为

$$\mathrm{d}A \approx \mathrm{d}r \cdot r\,\mathrm{d}\theta = r\,\mathrm{d}r\,\mathrm{d}\theta.$$

这和 Jacobian 计算得到的结果一致。

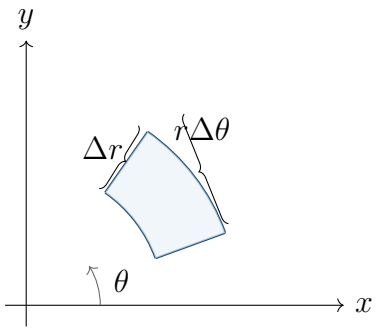



图 2：极坐标的小扇环面积近似为 $r\Delta\theta\cdot\Delta r$。

**例 3.1：用极坐标计算单位圆盘上的积分**

计算

$$I = \iint_{x^2+y^2\le 1} (x^2+y^2)\,\mathrm{d}x\,\mathrm{d}y.$$

**第一步：看区域。**区域是单位圆盘，极坐标最自然。

**第二步：换 integrand。**

$$x^2+y^2 = r^2.$$

**第三步：换面积元。**

$$\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\theta.$$

**第四步：换区域。**

$$0 \le r \le 1, \qquad 0 \le \theta \le 2\pi.$$

所以

$$I = \int_0^{2\pi}\int_0^1 r^2 \cdot r\,\mathrm{d}r\,\mathrm{d}\theta = \int_0^{2\pi}\int_0^1 r^3\,\mathrm{d}r\,\mathrm{d}\theta = 2\pi\cdot\frac{1}{4} = \frac{\pi}{2}.$$

**例 3.2：二维正态归一化常数**

计算

$$I = \iint_{\mathbb{R}^2} e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y.$$

用极坐标，$x^2+y^2=r^2$，区域为

$$0 \le r < \infty, \qquad 0 \le \theta \le 2\pi.$$

于是

$$I = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2} r\,\mathrm{d}r\,\mathrm{d}\theta.$$

令 $s = r^2/2$，则 $\mathrm{d}s = r\,\mathrm{d}r$，所以

$$\int_0^{\infty} e^{-r^2/2} r\,\mathrm{d}r = \int_0^{\infty} e^{-s}\,\mathrm{d}s = 1.$$

因此

$$I = 2\pi.$$

这说明二维标准正态密度为

$$f(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}.$$

# 7 线性变换：椭圆、矩阵和面积缩放

## 7.1 线性变换的 Jacobian

若

$$\begin{pmatrix} x \\ y \end{pmatrix} = A\begin{pmatrix} u \\ v \end{pmatrix} + \boldsymbol{b}, \qquad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

则

$$\frac{\partial(x,y)}{\partial(u,v)} = \det A = ad - bc.$$

平移 $\boldsymbol{b}$ 不改变面积，只有矩阵 $A$ 决定面积缩放。

**直觉**

线性变换是 Jacobian 最简单的情形，因为缩放比例处处一样。非线性变换可以被理解为“每一点附近都有一个不同的线性近似”，所以 Jacobian 会随 $(u,v)$ 改变。

## 7.2 椭圆区域的标准化

椭圆

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}\le 1$$

可以通过

$$x=au,\qquad y=bv$$

变成单位圆盘

$$u^2+v^2\le 1.$$

Jacobian 为

$$\frac{\partial(x,y)}{\partial(u,v)}=\det\begin{pmatrix}a & 0\\ 0 & b\end{pmatrix}=ab.$$

所以椭圆面积为

$$\iint_{x^2/a^2+y^2/b^2\le 1}1\,\mathrm{d}x\,\mathrm{d}y=\iint_{u^2+v^2\le 1}ab\,\mathrm{d}u\,\mathrm{d}v=ab\pi.$$

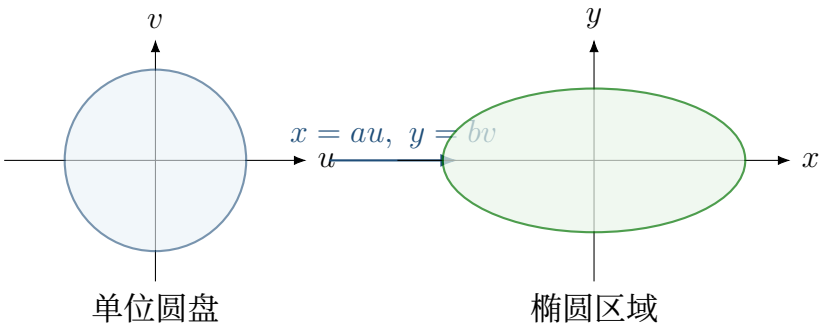



图 3: 线性变换把单位圆盘拉伸为椭圆，面积缩放为 $|ab|$。

**例 3.3：椭圆区域上的二次积分**

计算

$$I=\iint_{x^2/4+y^2/9\le 1}(x^2+y^2)\,\mathrm{d}x\,\mathrm{d}y.$$

令

$$x=2u,\qquad y=3v.$$

则区域变成 $u^2+v^2\le 1$，Jacobian 为

$$\left|\frac{\partial(x,y)}{\partial(u,v)}\right| = 6.$$

integrand 变为

$$x^2+y^2 = 4u^2+9v^2.$$

所以

$$I = \iint_{u^2+v^2\le 1} (4u^2+9v^2)6\,\mathrm{d}u\,\mathrm{d}v.$$

再用极坐标 $u = r\cos\theta, v = r\sin\theta$：

$$I = 6\int_0^{2\pi}\int_0^1 r^2(4\cos^2\theta+9\sin^2\theta)r\,\mathrm{d}r\,\mathrm{d}\theta.$$

分离变量：

$$\int_0^1 r^3\,\mathrm{d}r = \frac{1}{4},$$

并且

$$\int_0^{2\pi}\cos^2\theta\,\mathrm{d}\theta = \int_0^{2\pi}\sin^2\theta\,\mathrm{d}\theta = \pi.$$

因此

$$I = 6\cdot\frac{1}{4}\cdot(4\pi+9\pi) = \frac{39\pi}{2}.$$

## 7.3 金融中的线性变换：组合收益

设两个资产收益向量为

$$\boldsymbol{R} = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix}.$$

有时我们不直接研究 $(R_1,R_2)$，而研究

$$U = w_1R_1 + w_2R_2, \qquad V = R_1 - R_2.$$

这里 $U$ 是组合收益，$V$ 是两个资产相对表现。写成矩阵形式：

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} w_1 & w_2 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} R_1 \\ R_2 \end{pmatrix}.$$

如果你知道 $(R_1,R_2)$ 的联合密度，想推 $(U,V)$ 的联合密度，就必须使用 Jacobian。这里的核心不是积分技巧，而是“概率质量守恒”：同一块概率质量换坐标之后仍然是同一块概率质量。

# 8 随机变量变换：密度为什么要除以 Jacobian

## 8.1 一维密度变换

设随机变量 $X$ 有密度 $f_X(x)$，令

$$Y = g(X),$$

其中 $g$ 可逆，反函数为 $x = h(y)$。那么

$$f_Y(y) = f_X(h(y))|h'(y)|.$$

也可以写成

$$f_Y(y) = f_X(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|.$$

注意这里乘的是“从新变量回到旧变量”的导数。这一点在二维中特别容易混淆。

## 8.2 二维密度变换公式

设

$$(U, V) = G(X, Y)$$

可逆，反变换为

$$(X, Y) = T(U, V) = (x(u, v), y(u, v)).$$

如果 $(X, Y)$ 的联合密度是 $f_{X,Y}(x, y)$，那么

$$f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v))\left|\frac{\partial(x, y)}{\partial(u, v)}\right|.$$

**常见误区**

如果题目给的是 $(u, v) = G(x, y)$，而你直接算了

$$\frac{\partial(u, v)}{\partial(x, y)},$$

那么密度公式中需要的是它的倒数：

$$\left|\frac{\partial(x, y)}{\partial(u, v)}\right| = \frac{1}{\left|\frac{\partial(u,v)}{\partial(x,y)}\right|}.$$

很多概率题错就错在这里。

**密度变换的面积守恒推导**

取 $(u, v)$ 平面中的一个小区域 $A$。它在 $(x, y)$ 平面中的对应区域是 $T(A)$。同一件事件可以写成

$$(U, V) \in A \quad \Longleftrightarrow \quad (X, Y) \in T(A).$$

所以概率相等：

$$\Pr((U,V)\in A)=\Pr((X,Y)\in T(A)).$$

左边约为

$$f_{U,V}(u,v)\Delta u\Delta v.$$

右边约为

$$f_{X,Y}(x(u,v),y(u,v))\cdot\left|\frac{\partial(x,y)}{\partial(u,v)}\right|\Delta u\Delta v.$$

比较两边，得到

$$f_{U,V}(u,v)=f_{X,Y}(x(u,v),y(u,v))\left|\frac{\partial(x,y)}{\partial(u,v)}\right|.$$

这就是密度变换公式。

**例 3.4：从两个独立变量到和与差**

设 $(X,Y)$ 在单位正方形 $[0,1]^2$ 上均匀分布。令

$$U=X+Y,\qquad V=X-Y.$$

求 $(U,V)$ 的联合密度形式。

先求反变换：

$$X=\frac{U+V}{2},\qquad Y=\frac{U-V}{2}.$$

Jacobian 为

$$\frac{\partial(x,y)}{\partial(u,v)}=\det\begin{pmatrix}1/2 & 1/2\\ 1/2 & -1/2\end{pmatrix}=-\frac{1}{2}.$$

所以绝对值为 $1/2$。原密度在 $[0,1]^2$ 上等于 1，因此

$$f_{U,V}(u,v)=\frac{1}{2}$$

在满足

$$0\le\frac{u+v}{2}\le 1,\qquad 0\le\frac{u-v}{2}\le 1$$

的区域上成立，其他地方为 0。

这些不等式等价于

$$-u\le v\le 2-u,\qquad u-2\le v\le u.$$

所以支持集是一个菱形区域。这个例子说明：随机变量变换不仅要算 Jacobian，还要认真变换支持集。

# 9 经济金融中的典型变量变换

## 9.1 价格到对数收益：一维 Jacobian

设资产未来价格 $P_1 > 0$，当前价格 $P_0 > 0$ 给定。定义对数收益

$$R = \log \frac{P_1}{P_0}.$$

则

$$P_1 = P_0 e^R, \qquad \frac{\mathrm{d}P_1}{\mathrm{d}R} = P_0 e^R.$$

如果 $P_1$ 的密度为 $f_P(p)$，则 $R$ 的密度为

$$f_R(r) = f_P(P_0 e^r) P_0 e^r.$$

**直觉**

为什么对数收益的密度要乘 $P_0 e^r$？因为同样长度的 $\mathrm{d}r$ 在价格空间里对应的 $\mathrm{d}p$ 并不一样。价格越高，$r$ 上一个小区间对应的价格区间越大。

## 9.2 Cobb-Douglas 中的比例变换

考虑两个投入 $(K, L)$，有时比起绝对数量，经济学家更关心总规模和资本份额：

$$S = K + L, \qquad \alpha = \frac{K}{K + L}.$$

反变换为

$$K = \alpha S, \qquad L = (1 - \alpha) S.$$

Jacobian 为

$$\frac{\partial(K, L)}{\partial(S, \alpha)} = \det \begin{pmatrix} \alpha & S \\ 1 - \alpha & -S \end{pmatrix} = -S.$$

所以

$$\mathrm{d}K \, \mathrm{d}L = S \, \mathrm{d}S \, \mathrm{d}\alpha.$$

这个结果非常有解释力：当总规模 $S$ 越大时，同样一个份额变化 $\mathrm{d}\alpha$ 对应的实际投入变化越大。

## 9.3 风险因子标准化

设

$$\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma),$$

其中 $\Sigma$ 正定。若 $\Sigma = AA'$，定义

$$\boldsymbol{Z} = A^{-1}(\boldsymbol{X} - \boldsymbol{\mu}).$$

则

$$\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Z}.$$

这个线性变换的体积缩放为

$$|\det A|.$$

因此多元正态密度中会出现

$$\frac{1}{\sqrt{\det \Sigma}},$$

因为

$$|\det A| = \sqrt{\det \Sigma}.$$

**多元正态常数中的 Jacobian 来源**

若 $\boldsymbol{Z} \in \mathbb{R}^n$ 是标准正态，密度为

$$f_Z(z) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2} z'z\right).$$

设

$$\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Z}, \qquad \Sigma = AA'.$$

反变换为

$$\boldsymbol{Z} = A^{-1}(\boldsymbol{X} - \boldsymbol{\mu}).$$

根据密度变换,

$$f_X(x) = f_Z(A^{-1}(x-\mu)) \left|\det A^{-1}\right|.$$

所以

$$f_X(x) = (2\pi)^{-n/2} \exp\left[-\frac{1}{2}(x-\mu)'(A^{-1})'A^{-1}(x-\mu)\right] \frac{1}{|\det A|}.$$

因为

$$(A^{-1})'A^{-1} = (AA')^{-1} = \Sigma^{-1},$$

且

$$|\det A| = \sqrt{\det(AA')} = \sqrt{\det \Sigma},$$

得到

$$f_X(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right].$$

这说明多元正态密度的归一化常数不是背出来的，而是标准正态经过线性变换后的体积缩放。

# 10 三维换元：柱坐标与球坐标

## 10.1 三维 Jacobian

设

$$(x,y,z)=T(u,v,w).$$

Jacobian determinant 为

$$\frac{\partial(x,y,z)}{\partial(u,v,w)}=\det\begin{pmatrix}x_u & x_v & x_w\\ y_u & y_v & y_w\\ z_u & z_v & z_w\end{pmatrix}.$$

三维换元公式为

$$\iiint_E f(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z=\iiint_D f(T(u,v,w))\left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right|\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w.$$

**直觉**

三维中，Jacobian 是局部体积缩放比例。$(u,v,w)$ 中的小长方体，经过变换后在 $(x,y,z)$ 中近似成为一个小平行六面体；它的体积由三个切向量组成的行列式给出。

## 10.2 柱坐标

柱坐标为

$$x=r\cos\theta,\qquad y=r\sin\theta,\qquad z=z.$$

Jacobian 为

$$\frac{\partial(x,y,z)}{\partial(r,\theta,z)}=r.$$

所以

$$\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z=r\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}z.$$

柱坐标适合圆柱、圆锥、绕 $z$ 轴旋转的区域。

## 10.3 球坐标

本讲义采用物理和多元微积分中常见约定：

$$x=\rho\sin\varphi\cos\theta,\qquad y=\rho\sin\varphi\sin\theta,\qquad z=\rho\cos\varphi,$$

其中

$$\rho\ge 0,\qquad 0\le\varphi\le\pi,\qquad 0\le\theta<2\pi.$$

Jacobian 为

$$\left|\frac{\partial(x,y,z)}{\partial(\rho,\varphi,\theta)}\right|=\rho^2\sin\varphi.$$

所以

$$\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z=\rho^2\sin\varphi\,\mathrm{d}\rho\,\mathrm{d}\varphi\,\mathrm{d}\theta.$$

**常见误区**

不同教材对球坐标的符号可能不同。有的教材用 $\theta$ 表示极角，有的用 $\varphi$ 表示极角。不要死记字母，要记住面积元中一定有：半径平方 $\rho^2$ 和纬圈半径比例 $\sin\varphi$。

**例 3.5：用球坐标计算球体体积**

计算半径 $R$ 的球体体积：

$$B_R = \{(x,y,z) : x^2+y^2+z^2 \le R^2\}.$$

球坐标下：

$$0 \le \rho \le R, \quad 0 \le \varphi \le \pi, \quad 0 \le \theta \le 2\pi.$$

所以

$$\mathrm{Vol}(B_R) = \int_0^{2\pi}\int_0^{\pi}\int_0^{R} \rho^2 \sin\varphi \,\mathrm{d}\rho\,\mathrm{d}\varphi\,\mathrm{d}\theta.$$

分别计算：

$$\int_0^R \rho^2\,\mathrm{d}\rho = \frac{R^3}{3}, \qquad \int_0^{\pi}\sin\varphi\,\mathrm{d}\varphi = 2, \qquad \int_0^{2\pi}\mathrm{d}\theta = 2\pi.$$

因此

$$\mathrm{Vol}(B_R) = \frac{4\pi R^3}{3}.$$

# 11 Worked Examples：完整样板题

**Worked Example 1：三角形区域的自定义线性变换**

计算

$$I = \iint_R (x+y)\,\mathrm{d}x\,\mathrm{d}y,$$

其中 $R$ 是由

$$x \ge 0, \qquad y \ge 0, \qquad x+y \le 1$$

围成的三角形。令

$$u = x+y, \qquad v = x.$$

求积分。

**思路**。这个区域的边界中出现 $x+y$，所以把 $u=x+y$ 作为新变量会让边界变简单。

**第一步：反变换**。由

$$u = x+y, \qquad v = x,$$

得到

$$x = v, \qquad y = u - v.$$

**第二步：变区域。**原条件：

$$x \geq 0 \Rightarrow v \geq 0,$$

$$y \geq 0 \Rightarrow u - v \geq 0 \Rightarrow v \leq u,$$

$$x + y \leq 1 \Rightarrow u \leq 1.$$

同时 $u = x + y \geq 0$。所以新区域为

$$0 \leq u \leq 1, \qquad 0 \leq v \leq u.$$

**第三步：算 Jacobian。**

$$\frac{\partial(x, y)}{\partial(u, v)} = \det \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} = -1.$$

绝对值为 1。

**第四步：换 integrand。**

$$x + y = u.$$

所以

$$I = \int_0^1 \int_0^u u \, \mathrm{d}v \, \mathrm{d}u = \int_0^1 u^2 \, \mathrm{d}u = \frac{1}{3}.$$

## Worked Example 2：二维随机变量乘积与比值变换

设 $X, Y > 0$，联合密度为 $f_{X,Y}(x, y)$。定义

$$U = XY, \qquad V = \frac{X}{Y}.$$

求 $(U, V)$ 的联合密度表达式。

**第一步：反解** $x, y$。由

$$u = xy, \qquad v = x/y,$$

得

$$x = vy.$$

代入 $u = xy$：

$$u = vy^2 \Rightarrow y = \sqrt{\frac{u}{v}}.$$

于是

$$x = \sqrt{uv}, \qquad y = \sqrt{\frac{u}{v}}.$$

定义域为

$$u > 0, \qquad v > 0.$$

**第二步：算 Jacobian。**

$$x = u^{1/2}v^{1/2}, \qquad y = u^{1/2}v^{-1/2}.$$

所以

$$x_u = \frac{1}{2}u^{-1/2}v^{1/2}, \quad x_v = \frac{1}{2}u^{1/2}v^{-1/2},$$
$$y_u = \frac{1}{2}u^{-1/2}v^{-1/2}, \quad y_v = -\frac{1}{2}u^{1/2}v^{-3/2}.$$

因此

$$J = \frac{\partial(x,y)}{\partial(u,v)} = x_u y_v - x_v y_u.$$

代入：

$$J = \left(\frac{1}{2}u^{-1/2}v^{1/2}\right)\left(-\frac{1}{2}u^{1/2}v^{-3/2}\right) - \left(\frac{1}{2}u^{1/2}v^{-1/2}\right)\left(\frac{1}{2}u^{-1/2}v^{-1/2}\right).$$

第一项为

$$-\frac{1}{4}v^{-1},$$

第二项为

$$\frac{1}{4}v^{-1}.$$

所以

$$J = -\frac{1}{2v}, \qquad |J| = \frac{1}{2v}.$$

**第三步：写密度。**

$$f_{U,V}(u,v) = f_{X,Y}\left(\sqrt{uv}, \sqrt{\frac{u}{v}}\right)\frac{1}{2v}, \qquad u > 0, v > 0.$$

## Worked Example 3：两资产收益的线性变换

设 $(R_1, R_2)$ 的联合密度为 $f(r_1, r_2)$。定义

$$U = wR_1 + (1-w)R_2, \qquad V = R_1 - R_2,$$

其中 $0 < w < 1$。求 $(U, V)$ 的联合密度。

**第一步：反解。**由 $V = R_1 - R_2$，得

$$R_1 = R_2 + V.$$

代入 $U = wR_1 + (1-w)R_2$：

$$U = w(R_2 + V) + (1-w)R_2 = R_2 + wV.$$

所以

$$R_2 = U - wV, \qquad R_1 = U + (1-w)V.$$

**第二步：算 Jacobian。**

$$\frac{\partial(r_1, r_2)}{\partial(u, v)} = \det \begin{pmatrix} 1 & 1-w \\ 1 & -w \end{pmatrix} = -w - (1-w) = -1.$$

所以绝对值为 1。

**第三步：写密度。**

$$f_{U,V}(u, v) = f\left(u + (1-w)v, u - wv\right).$$

这个例子说明：并非所有变量变换都会带来非 1 的 Jacobian。很多金融线性组合只是旋转、剪切或某种面积保持变换。

## Worked Example 4：从联合密度推出总量的分布

设 $X, Y$ 独立且都服从 $[0, 1]$ 上的均匀分布。令

$$S = X + Y.$$

求 $S$ 的密度。

**方法一：用二维变换。**取

$$S = X + Y, \qquad V = X.$$

则反变换为

$$X = V, \qquad Y = S - V.$$

Jacobian：

$$\frac{\partial(x, y)}{\partial(s, v)} = \det \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} = -1.$$

所以 $|J| = 1$。原密度在单位正方形内为 1。

支持条件为

$$0 \le V \le 1, \qquad 0 \le S - V \le 1.$$

也就是

$$\max(0, S-1) \le V \le \min(1, S).$$

因此

$$f_S(s) = \int f_{S,V}(s, v)\, \mathrm{d}v = \int_{\max(0,s-1)}^{\min(1,s)} 1\, \mathrm{d}v.$$

当 $0 \le s \le 1$ 时，

$$f_S(s) = s.$$

当 $1 \le s \le 2$ 时，

$$f_S(s) = 2 - s.$$

所以

$$f_S(s) = \begin{cases} s, & 0 \le s \le 1, \\ 2 - s, & 1 \le s \le 2, \\ 0, & \text{otherwise}. \end{cases}$$

**经济解释**。如果 $X, Y$ 是两个独立收入来源且都在 $[0,1]$ 均匀分布，那么总收入 $S$ 最可能靠近 1，而不太可能靠近 0 或 2，因为产生中间总量的组合更多。

## 12 常见错误专题

**常见误区**

**错误 1：只换 integrand，不换区域。**
例如用极坐标时，把 $x^2 + y^2$ 换成 $r^2$，但区域仍写成 $-1 \le x \le 1$，这是坐标混用。

**常见误区**

**错误 2：忘记 Jacobian。**
尤其是极坐标忘记 $r$，球坐标忘记 $\rho^2 \sin\varphi$，随机变量变换忘记 $|J|$。这类错误会让结果数量级完全不对。

**常见误区**

**错误 3：用错 Jacobian 方向。**
如果你写的是反变换 $(x, y) = T(u, v)$，公式乘

$$\left| \frac{\partial(x, y)}{\partial(u, v)} \right|.$$

如果你先算的是正向变换 $(u, v) = G(x, y)$，则需要取倒数。

**常见误区**

**错误 4：没有检查一一性。**
例如极坐标若让 $0 \le \theta \le 4\pi$，同一个点会被计算两次。概率密度变换中如果变换不是一一的，需要分支求和。

**常见误区**

**错误 5：把 Jacobian 当成纯公式，不理解它是局部缩放。**
一旦遇到没见过的新变量，比如 $U = X/Y, V = Y$，只背公式就很容易错。理解局部面积缩放后，你就能自己推。

# 13 练习题

## A 组：基础计算

A1. 计算

$$\iint_{x^2+y^2\le 4} (x^2+y^2)\,\mathrm{d}x\,\mathrm{d}y.$$

A2. 计算椭圆

$$\frac{x^2}{9}+\frac{y^2}{4}\le 1$$

的面积。

A3. 求变换 $x = u + v, y = u - v$ 的 Jacobian，并解释面积缩放。

A4. 求极坐标变换 $x = r\cos\theta, y = r\sin\theta$ 的 Jacobian。

A5. 用球坐标计算半径为 2 的球体体积。

## B 组：概念判断

B1. 为什么换元公式里是 $|J|$ 而不是 $J$?

B2. 若 $J(u_0, v_0) = 0$，这说明变换在该点附近可能发生了什么?

B3. 极坐标中的 $r$ 为什么不是额外的“经验修正项”?

B4. 在线性变换 $x = 2u, y = 3v$ 下，一个面积为 0.01 的小区域大约变成多大?

B5. 为什么随机变量变换时支持集比 Jacobian 更容易出错?

## C 组：混合题

C1. 计算

$$\iint_R e^{x^2+y^2}\,\mathrm{d}x\,\mathrm{d}y,$$

其中 $R = \{(x, y) : 1 \le x^2 + y^2 \le 4\}$。

C2. 设 $R$ 是由 $x \ge 0, y \ge 0, x + y \le 2$ 围成的三角形。用 $u = x + y, v = x - y$ 变换计算 $R$ 的面积。

C3. 设 $X, Y$ 在单位正方形上均匀分布，令 $U = X + Y, V = Y$。求 $(U, V)$ 的联合密度和支持集。

C4. 设 $X > 0$ 的密度为 $f_X(x)$，令 $Y = \log X$。求 $Y$ 的密度。

C5. 计算

$$\iiint_{x^2+y^2\le 1, 0\le z\le 3} (x^2 + y^2)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z.$$

### D 组：经济金融建模题

D1. 设未来价格 $P_1$ 的密度为 $f_P(p)$，当前价格 $P_0$ 固定，定义简单收益 $R = P_1/P_0 - 1$。求 $R$ 的密度。

D2. 设 $(R_1, R_2)$ 为两个资产收益，定义组合收益 $U = wR_1 + (1-w)R_2$ 和价差 $V = R_1 - R_2$。推导反变换和 Jacobian。

D3. 设家庭收入 $Y > 0$，令 $Z = \log Y$。若 $Z \sim N(\mu, \sigma^2)$，用变量变换推出 $Y$ 的密度。

D4. 设投入 $(K, L)$ 用总规模 $S = K + L$ 和资本份额 $\alpha = K/(K + L)$ 表示。求 $\mathrm{d}K\,\mathrm{d}L$ 与 $\mathrm{d}S\,\mathrm{d}\alpha$ 的关系，并解释经济含义。

D5. 若二维风险因子 $Z \sim N(0, I_2)$，资产收益 $R = \mu + AZ$。说明为什么 $R$ 的密度中会出现 $1/|\det A|$。

### E 组：挑战和证明题

E1. 推导球坐标 Jacobian $\rho^2 \sin\varphi$。

E2. 设 $U = X/Y, V = Y$，其中 $Y \neq 0$。求反变换和 Jacobian，并写出 $(U, V)$ 的密度公式。

E3. 设 $(X, Y)$ 的联合密度为 $f(x, y)$，令 $R = \sqrt{X^2 + Y^2}$, $\Theta = \arctan(Y/X)$。在合适象限范围内写出 $(R, \Theta)$ 的密度。

E4. 证明若 $(x, y) = A(u, v)$，则任何区域面积都被放大 $|\det A|$ 倍。

E5. 设 $X, Y$ 独立指数分布，参数均为 1。令 $S = X + Y, T = X/(X + Y)$。求 $(S, T)$ 的联合密度，并判断 $S, T$ 是否独立。

# 14 练习题详细解答

# A 组答案

## A1

用极坐标：区域为 $0 \le r \le 2, 0 \le \theta \le 2\pi$，integrand 为 $r^2$，面积元为 $r\,\mathrm{d}r\,\mathrm{d}\theta$。

$$I = \int_0^{2\pi} \int_0^2 r^3 \,\mathrm{d}r\,\mathrm{d}\theta = 2\pi \cdot \frac{2^4}{4} = 8\pi.$$

## A2

椭圆半轴为 $a = 3, b = 2$，面积为

$$\pi ab = 6\pi.$$

也可令 $x = 3u, y = 2v$，Jacobian 为 6，单位圆盘面积为 $\pi$，所以总面积为 $6\pi$。

## A3

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \det \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = -2.$$

所以面积缩放比例为 $|J| = 2$。一个很小的 $(u, v)$ 面积元在 $(x, y)$ 中变成原来的 2 倍。

## A4

$$J = \det \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r.$$

所以 $\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\theta$。

## A5

球坐标下：

$$0 \le \rho \le 2, \quad 0 \le \varphi \le \pi, \quad 0 \le \theta \le 2\pi.$$

体积为

$$\int_0^{2\pi} \int_0^{\pi} \int_0^2 \rho^2 \sin\varphi \,\mathrm{d}\rho\,\mathrm{d}\varphi\,\mathrm{d}\theta = 2\pi \cdot 2 \cdot \frac{8}{3} = \frac{32\pi}{3}.$$

## B 组答案

### B1

因为行列式的符号表示方向是否翻转，而面积和概率质量只关心大小。若变换把坐标方向翻转，$J<0$，但面积不能变成负数，所以换元公式用 $|J|$。

### B2

$J=0$ 表明局部面积缩放比例为 0，变换可能把二维小区域压成一条线或一个点，也可能局部不可逆。因此标准换元公式的条件失效。

### B3

极坐标中的 $r$ 来自局部面积缩放。半径越大，同样的角度变化 $\mathrm{d}\theta$ 对应的弧长 $r\,\mathrm{d}\theta$ 越大，所以面积元为 $r\,\mathrm{d}r\,\mathrm{d}\theta$。

### B4

变换矩阵为 $\mathrm{diag}(2,3)$，行列式为 6。面积 $0.01$ 的小区域大约变成 $0.06$。

### B5

因为 Jacobian 只是局部缩放，而支持集决定哪些 $(u,v)$ 真的可能出现。很多变换下，原来的矩形或三角形会变成斜的、多段的区域。若支持集写错，即使 Jacobian 正确，密度也会在错误区域上取值。

## C 组答案

### C1

区域是圆环：$1\le r\le 2, 0\le\theta\le 2\pi$。所以

$$I=\int_0^{2\pi}\int_1^2 e^{r^2} r\,\mathrm{d}r\,\mathrm{d}\theta.$$

令 $s=r^2$，$\mathrm{d}s=2r\,\mathrm{d}r$：

$$\int_1^2 e^{r^2} r\,\mathrm{d}r=\frac{1}{2}\int_1^4 e^s\,\mathrm{d}s=\frac{1}{2}(e^4-e).$$

故

$$I=\pi(e^4-e).$$

## C2

由

$$u = x + y, \qquad v = x - y$$

得

$$x = \frac{u+v}{2}, \qquad y = \frac{u-v}{2}.$$

Jacobian:

$$\frac{\partial(x,y)}{\partial(u,v)} = \det \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix} = -\frac{1}{2}.$$

原区域条件变为:

$$0 \le u \le 2, \qquad -u \le v \le u.$$

面积为

$$\int_0^2 \int_{-u}^{u} \frac{1}{2} \, \mathrm{d}v \, \mathrm{d}u = \int_0^2 u \, \mathrm{d}u = 2.$$

这与直角三角形面积 $2 \cdot 2/2 = 2$ 一致。

## C3

反变换为

$$X = V, \qquad Y = U - V.$$

Jacobian 绝对值为 1。支持条件:

$$0 \le V \le 1, \qquad 0 \le U - V \le 1.$$

因此

$$f_{U,V}(u,v) = 1$$

在满足上述条件的区域上成立，其他地方为 0。

## C4

由 $Y = \log X$ 得 $X = e^Y$，且

$$\frac{\mathrm{d}x}{\mathrm{d}y} = e^y.$$

所以

$$f_Y(y) = f_X(e^y) e^y.$$

**C5**

使用柱坐标:

$$x^2+y^2=r^2,\quad \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z=r\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}z.$$

区域:

$$0\le r\le 1,\quad 0\le\theta\le 2\pi,\quad 0\le z\le 3.$$

所以

$$I=\int_0^3\int_0^{2\pi}\int_0^1 r^2\cdot r\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}z=3\cdot 2\pi\cdot\frac{1}{4}=\frac{3\pi}{2}.$$

## D 组答案

**D1**

$$R=\frac{P_1}{P_0}-1\Rightarrow P_1=P_0(1+R).$$

因此

$$\frac{\mathrm{d}P_1}{\mathrm{d}R}=P_0.$$

所以

$$f_R(r)=f_P(P_0(1+r))P_0,$$

支持集由 $P_0(1+r)>0$ 决定，即 $r>-1$。

**D2**

由

$$U=wR_1+(1-w)R_2,\qquad V=R_1-R_2,$$

得到

$$R_2=U-wV,\qquad R_1=U+(1-w)V.$$

Jacobian 为

$$\det\begin{pmatrix}1 & 1-w\\ 1 & -w\end{pmatrix}=-1.$$

绝对值为 1。

**D3**

若 $Z=\log Y\sim N(\mu,\sigma^2)$，则 $Y=e^Z$。反过来 $z=\log y$，且

$$\frac{\mathrm{d}z}{\mathrm{d}y}=\frac{1}{y}.$$

所以

$$f_Y(y) = \frac{1}{y}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(\log y-\mu)^2}{2\sigma^2}\right], \qquad y>0.$$

这就是 lognormal 分布。

**D4**

反变换为

$$K=\alpha S, \qquad L=(1-\alpha)S.$$

Jacobian:

$$\frac{\partial(K,L)}{\partial(S,\alpha)} = \det\begin{pmatrix}\alpha & S\\ 1-\alpha & -S\end{pmatrix} = -S.$$

所以

$$\mathrm{d}K\,\mathrm{d}L = S\,\mathrm{d}S\,\mathrm{d}\alpha.$$

经济含义：总规模越大，同样的份额变化对应的资本和劳动绝对变化越大。

**D5**

若 $R=\mu+AZ$，则反变换为

$$Z=A^{-1}(R-\mu).$$

密度变换中要乘

$$\left|\det A^{-1}\right| = \frac{1}{|\det A|}.$$

所以 $R$ 的密度中出现 $1/|\det A|$。这表示标准正态空间中的单位体积经过 $A$ 变换后被放大 $|\det A|$ 倍，密度高度相应降低。

## E 组答案

**E1**

球坐标:

$$x=\rho\sin\varphi\cos\theta, \quad y=\rho\sin\varphi\sin\theta, \quad z=\rho\cos\varphi.$$

计算 Jacobian determinant 可得

$$\left|\frac{\partial(x,y,z)}{\partial(\rho,\varphi,\theta)}\right| = \rho^2\sin\varphi.$$

几何上看，小体积由三个方向构成：径向长度 $\mathrm{d}\rho$，极角方向弧长 $\rho\,\mathrm{d}\varphi$，方位角方

向弧长 $\rho\sin\varphi\,\mathrm{d}\theta$。乘起来就是

$$\rho^2\sin\varphi\,\mathrm{d}\rho\,\mathrm{d}\varphi\,\mathrm{d}\theta.$$

**E2**

$$U = X/Y, \qquad V = Y.$$

反变换：

$$X = UV, \qquad Y = V.$$

Jacobian：

$$\frac{\partial(x,y)}{\partial(u,v)} = \det\begin{pmatrix} v & u \\ 0 & 1 \end{pmatrix} = v.$$

所以

$$f_{U,V}(u,v) = f_{X,Y}(uv,v)|v|.$$

如果 $Y > 0$，则 $|v| = v$。

**E3**

在一个不重复覆盖的角度范围内，例如 $0 \le \theta < 2\pi$，反变换为

$$X = R\cos\Theta, \qquad Y = R\sin\Theta.$$

Jacobian 为 $r$。所以

$$f_{R,\Theta}(r,\theta) = f_{X,Y}(r\cos\theta, r\sin\theta)r, \qquad r \ge 0.$$

**E4**

线性变换把单位正方形的两条边向量 $e_1, e_2$ 变为 $Ae_1, Ae_2$。它们张成的平行四边形面积为

$$|\det(Ae_1, Ae_2)| = |\det A|.$$

任意区域可通过小网格近似，每个小格子的面积都被放大 $|\det A|$ 倍，求和取极限后，整体面积也被放大 $|\det A|$ 倍。

**E5**

原密度为

$$f_{X,Y}(x,y) = e^{-x}e^{-y} = e^{-(x+y)}, \qquad x, y > 0.$$

令

$$S = X + Y, \qquad T = \frac{X}{X+Y}.$$

反变换：

$$X = ST, \qquad Y = S(1-T).$$

支持集：

$$S > 0, \qquad 0 < T < 1.$$

Jacobian：

$$\frac{\partial(x,y)}{\partial(s,t)} = \det \begin{pmatrix} t & s \\ 1-t & -s \end{pmatrix} = -s.$$

所以 $|J| = s$。因此

$$f_{S,T}(s,t) = e^{-s}s, \qquad s > 0, 0 < t < 1.$$

它可以分解为

$$f_S(s) = se^{-s}, \qquad f_T(t) = 1,$$

所以 $S$ 与 $T$ 独立，且 $T \sim U(0,1)$。

## 15 本章总结图谱

| 主题 | 核心公式 | | 经济金融含义 |
|---|---|---|---|
| 一元换元 | $\mathrm{d}x = \lvert g'(u)\rvert \,\mathrm{d}u$ | | 变量尺度改变, 密度和积分单位也改变 |
| 二维 Jacobian | $\mathrm{d}x\,\mathrm{d}y = \left\lvert \frac{\partial(x,y)}{\partial(u,v)} \right\rvert \mathrm{d}u\,\mathrm{d}v$ | | 坐标变换下的面积缩放 |
| 极坐标 | $\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\theta$ | | 圆形风险区域、二维正态、半径型模型 |
| 线性变换 | $\mathrm{d}x\,\mathrm{d}y = \lvert\det A\rvert \,\mathrm{d}u\,\mathrm{d}v$ | | 投资组合、因子模型、多元正态标准化 |
| 密度变换 | $f_{U,V}$ | = | 随机变量函数的分布推导 |
| | $f_{X,Y}(x(u,v), y(u,v))\lvert J\rvert$ | | |
| 柱坐标 | $\mathrm{d}V = r\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}z$ | | 圆柱型状态空间、旋转对称分布 |
| 球坐标 | $\mathrm{d}V = \rho^2 \sin\varphi\,\mathrm{d}\rho\,\mathrm{d}\varphi\,\mathrm{d}\theta$ | | 三维风险因子、球形约束、范数型模型 |

**本章压缩总结**

Chapter 3 的核心结论：

1. Jacobian determinant 是局部面积或体积缩放比例；
2. 换元时 integrand、区域、面积元都要换；
3. 极坐标、柱坐标、球坐标只是 Jacobian 公式的特殊例子；
4. 随机变量密度变换本质上是概率质量守恒；
5. 金融中的价格到收益、收益到组合、风险因子标准化，都是变量变换；
6. 多元正态密度里的 $\sqrt{\det \Sigma}$ 来自线性变换的体积缩放。

## 下一章预告

下一章是 **Chapter 4：含参积分、Leibniz Rule 与积分下求导**。本章处理的是“换变量时积分怎么变”；下一章处理的是“参数变化时积分怎么变”。如果说 Jacobian 是变量坐标改变下的面积修正，那么 Leibniz Rule 就是模型参数改变下的边界和 integrand 修正。它会直接连接到期望效用、税收政策、likelihood score、Bayesian normalization constant 和比较静态。

# 经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

---

# Chapter 4

## 含参积分、Leibniz Rule 与积分下求导

---

本章讲清楚一个高频但容易误用的操作：什么时候可以把导数推进积分号里面，什么时候边界也会贡献一项。在经济金融中，它对应期望效用的参数变化、福利函数的税率效应、似然函数的 score、需求边界变化和贝叶斯归一化常数的导数。

math4econ personalized notes

2026

# 目录

**Chapter 4 总览** **3**

**1 本章为什么重要** **5**

**2 最小前置知识** **5**

**3 从最简单的含参积分开始** **6**
3.1 定义：含参积分 . . . . . 6
3.2 第一种核心公式：固定区间 . . . . . 7

**4 Leibniz Rule：上下限也会动** **8**
4.1 先看一个会出错的例子 . . . . . 8
4.2 Leibniz Rule 的正式公式 . . . . . 9
4.3 Leibniz Rule 的推导 . . . . . 9

**5 计算模板：看到题目怎么做** **10**

**6 期望函数也是含参积分** **11**
6.1 从积分到期望 . . . . . 11
6.2 经济应用：风险厌恶参数 . . . . . 12

**7 密度也含参数：score function 的入口** **12**
7.1 score function trick . . . . . 13
7.2 一个重要恒等式：score 的期望为零 . . . . . 13

**8 经济应用一：税率变化与社会福利** **14**
8.1 带政策阈值的福利：边界项出现 . . . . . 15

**9 经济应用二：需求、阈值与市场份额** **15**

**10 金融应用：从收益分布到风险度量** **16**
10.1 VaR 和 Expected Shortfall 中的边界 . . . . . 17

**11 贝叶斯归一化常数与 log-sum-exp** **17**

**12 更严谨一点：什么时候可以积分下求导** **18**
12.1 紧区间上的连续性条件 . . . . . 18
12.2 无限区间或概率空间上的支配条件 . . . . . 18

**13 Worked Examples：完整样板题** **19**

**14 常见错误清单** **22**

**15 练习题** **22**

A 组：基础计算 . . . . . 22
B 组：概念判断 . . . . . 23
C 组：混合题 . . . . . 23
D 组：经济金融建模题 . . . . . 23
E 组：挑战与证明题 . . . . . 24

**16 练习题详细解答** **25**

A 组答案 . . . . . 25
B 组答案 . . . . . 26
C 组答案 . . . . . 27
D 组答案 . . . . . 28
E 组答案 . . . . . 29

**17 本章总结图谱** **31**

**18 下一章连接** **31**

# Chapter 4 总览

**本章目标**

学完本章，你应该达到以下目标：

1. 理解“含参积分”不是一种技巧，而是一类函数：参数变了，整个积分值跟着变；
2. 能区分三种情况：固定积分区间、上下限依赖参数、被积函数和积分区域同时依赖参数；
3. 能从差商和面积变化的角度推导 Leibniz rule，而不是只背公式；
4. 能判断什么时候可以积分下求导，即什么时候

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\int f(x,\theta)\,\mathrm{d}x=\int \frac{\partial f}{\partial \theta}(x,\theta)\,\mathrm{d}x$$

成立；
5. 能把这个工具用于期望、概率、似然、福利、需求和金融风险模型；
6. 能识别常见错误：漏掉边界项、忘记支撑集随参数变化、把点态可导误当成可交换极限、在密度参数问题中漏掉归一化约束。

**直觉**

一句话：**含参积分就是把一整片面积、概率质量或福利总量看成参数的函数；积分下求导就是问“当参数轻轻动一下，这片总量怎么变”**。
有两种变化来源：

1. **内部变化**：同一个点上的高度变了，即 $f(x,\theta)$ 对 $\theta$ 变了；
2. **边界变化**：积分区间本身变了，新加入或被移出的那一小条区域也会改变总量。

因此 Leibniz rule 的结构一定是：

$$\text{总变化}=\text{右边界贡献}-\text{左边界贡献}+\text{内部高度变化}.$$

**本章压缩总结**

本章是从“积分计算”走向“经济模型比较静态”的关键桥梁。前几章你学会了怎样把多维对象加总；本章开始，你要学会怎样研究这个加总结果对参数的反应。以后你看到

$$\mathbb{E}_\theta[g(X)],\qquad \int u(c(x,\tau))f(x)\,\mathrm{d}x,\qquad \log\int \exp(v(x,\theta))\,\mathrm{d}x,$$

都要立刻意识到：这些都是含参积分，比较静态离不开本章。

# 1 本章为什么重要

在经济金融模型中，我们很少只关心一个静态积分的数值。更常见的问题是：参数改变以后，某个总量如何变化。例如：

- 税率 $\tau$ 提高以后，社会福利

  $$W(\tau)=\int u(c(y,\tau))f(y)\,\mathrm{d}y$$

  怎么变？

- 风险厌恶参数 $\gamma$ 提高以后，期望效用

  $$V(\gamma)=\mathbb{E}\left[\frac{C^{1-\gamma}}{1-\gamma}\right]$$

  怎么变？

- 概率模型的参数 $\theta$ 改变以后，似然函数

  $$L(\theta)=\prod_{i=1}^{n} f(X_i\mid\theta)$$

  的导数，也就是 score，是什么？

- 某个政策只影响收入低于阈值 $b(\theta)$ 的人，那么福利积分的上限本身也随政策改变，这时边界项不能漏掉。

  这些问题有共同结构：

  $$F(\theta)=\int_{a(\theta)}^{b(\theta)} f(x,\theta)\,\mathrm{d}x.$$

你要研究的是 $F'(\theta)$。

**常见误区**

很多同学会机械写成

$$F'(\theta)=\int_{a(\theta)}^{b(\theta)} f_\theta(x,\theta)\,\mathrm{d}x.$$

这通常是错的。只要上下限依赖 $\theta$，就一定要检查边界项：

$$f(b(\theta),\theta)b'(\theta)-f(a(\theta),\theta)a'(\theta).$$

如果被积区域是多维集合 $D(\theta)$，问题还会更复杂，边界项会变成“边界流量”的形式。我们本章主要掌握一维积分区间和经济金融常用情形。

# 2 最小前置知识

本章默认你已经掌握：

1. 一元微积分基本定理；
2. 链式法则；
3. 偏导数；
4. 积分的线性性；
5. 二重积分和概率密度的基本解释；
6. 第 1–3 章关于积分区域、密度和变量替换的直觉。

最核心的旧知识是链式法则。若

$$G(t) = \int_c^t f(x)\,\mathrm{d}x,$$

则

$$G'(t) = f(t).$$

如果 $t = b(\theta)$，那么

$$\frac{\mathrm{d}}{\mathrm{d}\theta} G(b(\theta)) = G'(b(\theta))b'(\theta) = f(b(\theta))b'(\theta).$$

这就是 Leibniz rule 中右边界项的来源。

# 3 从最简单的含参积分开始

## 3.1 定义：含参积分

**定义 4.1：含参积分**

设 $x$ 是积分变量，$\theta$ 是参数。若

$$F(\theta) = \int_A f(x,\theta)\,\mathrm{d}x,$$

则 $F$ 称为一个含参积分函数。这里 $A$ 可以是固定区间，也可以是依赖参数的区间 $[a(\theta), b(\theta)]$。

注意：积分变量 $x$ 会被积分“消掉”，所以 $F$ 是 $\theta$ 的函数，而不是 $x$ 的函数。

**例 4.1：固定区间的含参积分**

令

$$F(\theta) = \int_0^1 x^\theta\,\mathrm{d}x, \qquad \theta > -1.$$

直接积分得

$$F(\theta)=\frac{1}{\theta+1}.$$

因此

$$F'(\theta)=-\frac{1}{(\theta+1)^2}.$$

但也可以先对被积函数求偏导：

$$\frac{\partial}{\partial\theta}x^\theta=x^\theta\log x.$$

若允许积分下求导，则

$$F'(\theta)=\int_0^1 x^\theta\log x\,\mathrm{d}x.$$

于是得到一个有用积分：

$$\int_0^1 x^\theta\log x\,\mathrm{d}x=-\frac{1}{(\theta+1)^2}.$$

这个例子说明，积分下求导不仅能求导，还能反过来计算一些难积分。

## 3.2 第一种核心公式：固定区间

**命题 4.1：固定区间上的积分下求导**

设 $f(x,\theta)$ 和 $f_\theta(x,\theta)$ 在矩形区域 $[a,b]\times(\theta_0-\epsilon,\theta_0+\epsilon)$ 上连续，则

$$F(\theta)=\int_a^b f(x,\theta)\,\mathrm{d}x$$

在 $\theta_0$ 可导，并且

$$F'(\theta_0)=\int_a^b \frac{\partial f}{\partial\theta}(x,\theta_0)\,\mathrm{d}x.$$

若上述连续性对区间中所有 $\theta$ 成立，则

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}\theta}\int_a^b f(x,\theta)\,\mathrm{d}x=\int_a^b f_\theta(x,\theta)\,\mathrm{d}x.}$$

**证明思路：为什么可以把导数放进去**

从差商开始：

$$\frac{F(\theta+h)-F(\theta)}{h}=\int_a^b\frac{f(x,\theta+h)-f(x,\theta)}{h}\,\mathrm{d}x.$$

对每个固定 $x$，当 $h \to 0$ 时，差商趋于 $f_\theta(x,\theta)$。所以我们想写：

$$\lim_{h\to 0}\int_a^b \frac{f(x,\theta+h)-f(x,\theta)}{h}\,\mathrm{d}x = \int_a^b \lim_{h\to 0}\frac{f(x,\theta+h)-f(x,\theta)}{h}\,\mathrm{d}x.$$

关键问题是：**极限能否和积分交换?**
在闭矩形上，如果 $f_\theta$ 连续，则一致连续且有界。由一元均值定理，

$$\frac{f(x,\theta+h)-f(x,\theta)}{h} = f_\theta(x,\theta+\lambda h)$$

对某个 $\lambda \in (0,1)$ 成立。由于 $f_\theta$ 在闭矩形上一致连续，上式会一致收敛到 $f_\theta(x,\theta)$。一致收敛允许积分与极限交换，于是得到结论。

**直觉**

固定区间时，积分区域没有动，所以总变化只来自每一点高度的变化：

$$\text{总高度变化的累加} = \int \text{局部高度变化}\,\mathrm{d}x.$$

这就是

$$F'(\theta) = \int f_\theta(x,\theta)\,\mathrm{d}x.$$

# 4 Leibniz Rule：上下限也会动

## 4.1 先看一个会出错的例子

**例 4.2：只对 integrand 求导会漏掉全部变化**

令

$$F(\theta) = \int_0^\theta x^2\,\mathrm{d}x.$$

这里被积函数 $x^2$ 不含 $\theta$。如果错误地只看 $f_\theta = 0$，你会得到 $F'(\theta) = 0$。
但直接算：

$$F(\theta) = \frac{\theta^3}{3}, \qquad F'(\theta) = \theta^2.$$

变化来自哪里? 来自上限 $\theta$ 向右移动时新增加的一小条面积。新增加的宽度约为 $\mathrm{d}\theta$，高度约为 $f(\theta) = \theta^2$，所以面积变化约为

$$\theta^2\,\mathrm{d}\theta.$$

因此导数是 $\theta^2$。

## 4.2 Leibniz Rule 的正式公式

**定理 4.2：一维 Leibniz Rule**

设

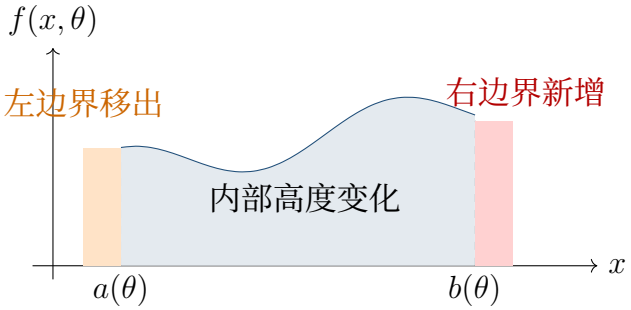
$$F(\theta)=\int_{a(\theta)}^{b(\theta)} f(x,\theta)\,\mathrm{d}x,$$

其中 $a,b$ 可导，$f$ 与 $f_\theta$ 在相关区域上连续。则

$$\boxed{\frac{\mathrm{d}F}{\mathrm{d}\theta}=f(b(\theta),\theta)b'(\theta)-f(a(\theta),\theta)a'(\theta)+\int_{a(\theta)}^{b(\theta)} f_\theta(x,\theta)\,\mathrm{d}x.}$$

这个公式非常重要，建议你把它读成三项：

$$\boxed{F'=\text{右边界项}-\text{左边界项}+\text{内部项}.}$$




## 4.3 Leibniz Rule 的推导

**推导：把移动上下限转成固定区间**

定义原函数型对象

$$H(t,\theta)=\int_{c}^{t} f(x,\theta)\,\mathrm{d}x,$$

其中 $c$ 是任意固定点。则

$$F(\theta)=\int_{a(\theta)}^{b(\theta)} f(x,\theta)\,\mathrm{d}x=H(b(\theta),\theta)-H(a(\theta),\theta).$$

对第一项用链式法则：

$$\frac{\mathrm{d}}{\mathrm{d}\theta}H(b(\theta),\theta)=H_t(b(\theta),\theta)b'(\theta)+H_\theta(b(\theta),\theta).$$

由微积分基本定理，

$$H_t(t,\theta)=f(t,\theta).$$

又由固定区间积分下求导，

$$H_\theta(t,\theta)=\int_{c}^{t} f_\theta(x,\theta)\,\mathrm{d}x.$$

因此

$$\frac{\mathrm{d}}{\mathrm{d}\theta}H(b(\theta),\theta)=f(b(\theta),\theta)b'(\theta)+\int_c^{b(\theta)}f_\theta(x,\theta)\,\mathrm{d}x.$$

同理，

$$\frac{\mathrm{d}}{\mathrm{d}\theta}H(a(\theta),\theta)=f(a(\theta),\theta)a'(\theta)+\int_c^{a(\theta)}f_\theta(x,\theta)\,\mathrm{d}x.$$

两式相减：

$$F'(\theta)=f(b(\theta),\theta)b'(\theta)-f(a(\theta),\theta)a'(\theta)+\int_{a(\theta)}^{b(\theta)}f_\theta(x,\theta)\,\mathrm{d}x.$$

**常见误区**

符号方向很重要。右边界扩大使积分增加，所以是加号；左边界向右移动会切掉左侧一段，所以是减号：

$$+f(b)b'-f(a)a'.$$

如果 $a'(\theta)<0$，左边界向左移动，虽然公式仍是 $-f(a)a'$，但因为 $a'<0$，它实际是正贡献。

# 5 计算模板：看到题目怎么做

**方法**

处理

$$F(\theta)=\int_{a(\theta)}^{b(\theta)}f(x,\theta)\,\mathrm{d}x$$

时，按四步走：

1. 识别下限 $a(\theta)$、上限 $b(\theta)$、被积函数 $f(x,\theta)$；
2. 写边界项：
$$f(b(\theta),\theta)b'(\theta)-f(a(\theta),\theta)a'(\theta);$$
3. 写内部项：
$$\int_{a(\theta)}^{b(\theta)}f_\theta(x,\theta)\,\mathrm{d}x;$$
4. 最后再代入、化简。不要一上来就直接求完积分，因为很多经济模型中的积分根本没有闭式解。

**例 4.3：上下限和被积函数都含参数**

求

$$F(\theta)=\int_{\theta}^{\theta^2} e^{-\theta x}\,\mathrm{d}x$$

的导数。

这里

$$a(\theta)=\theta,\qquad b(\theta)=\theta^2,\qquad f(x,\theta)=e^{-\theta x}.$$

所以

$$a'(\theta)=1,\qquad b'(\theta)=2\theta,\qquad f_\theta(x,\theta)=-xe^{-\theta x}.$$

由 Leibniz rule：

$$F'(\theta)=e^{-\theta\cdot\theta^2}(2\theta)-e^{-\theta\cdot\theta}(1)+\int_{\theta}^{\theta^2}(-xe^{-\theta x})\,\mathrm{d}x.$$

即

$$\boxed{F'(\theta)=2\theta e^{-\theta^3}-e^{-\theta^2}-\int_{\theta}^{\theta^2} xe^{-\theta x}\,\mathrm{d}x.}$$

如果题目只要求导数，这个形式已经很好；如果要求完全化简，再对最后一项做积分。

# 6 期望函数也是含参积分

## 6.1 从积分到期望

若随机变量 $X$ 的密度为 $p(x)$，那么

$$\mathbb{E}[g(X,\theta)]=\int g(x,\theta)p(x)\,\mathrm{d}x.$$

这是最常见的固定区间含参积分。若 $p$ 不依赖 $\theta$，则在条件允许时，

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}[g(X,\theta)]=\mathbb{E}[g_\theta(X,\theta)].$$

**命题 4.3：期望下求导**

若 $g(x,\theta)$ 对 $\theta$ 可导，并且存在一个可积函数 $M(x)$ 使得在 $\theta$ 的邻域内

$$|g_\theta(x,\theta)|\le M(x),$$

则

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}[g(X,\theta)]=\mathbb{E}[g_\theta(X,\theta)].}$$

这里的 $M(x)$ 是“支配函数”。直觉上，它保证导数不会在某些区域爆掉，所以极限和积分可以交换。

## 6.2 经济应用：风险厌恶参数

**例 4.4：CRRA 期望效用对风险厌恶参数求导**

设消费 $C>0$，效用函数为

$$u(C;\gamma)=\frac{C^{1-\gamma}}{1-\gamma},\qquad \gamma\neq 1.$$

定义

$$V(\gamma)=\mathbb{E}\left[\frac{C^{1-\gamma}}{1-\gamma}\right].$$

求 $V'(\gamma)$。

把 $C$ 看成随机变量，先对效用函数求 $\gamma$ 的偏导。记

$$u(C;\gamma)=\frac{e^{(1-\gamma)\log C}}{1-\gamma}.$$

令 $A=C^{1-\gamma}$，$B=1-\gamma$。则

$$\frac{\partial A}{\partial \gamma}=-C^{1-\gamma}\log C,\qquad \frac{\partial}{\partial\gamma}\frac{1}{B}=\frac{1}{(1-\gamma)^2}.$$

所以

$$\frac{\partial u}{\partial\gamma}=-\frac{C^{1-\gamma}\log C}{1-\gamma}+\frac{C^{1-\gamma}}{(1-\gamma)^2}.$$

若可交换求导与期望，则

$$\boxed{V'(\gamma)=\mathbb{E}\left[-\frac{C^{1-\gamma}\log C}{1-\gamma}+\frac{C^{1-\gamma}}{(1-\gamma)^2}\right].}$$

**直觉**

这个导数不只是数学计算。它描述的是：风险厌恶参数 $\gamma$ 增大一点时，随机消费所带来的期望效用怎样变化。由于 $\log C$ 出现，消费分布中高消费和低消费状态对导数的贡献不一样。这就是为什么研究风险偏好时，经常需要对期望效用做参数求导。

# 7 密度也含参数：score function 的入口

上一节默认密度 $p(x)$ 不随 $\theta$ 变。但在统计和计量中，我们经常有

$$X\sim f(x\mid\theta).$$

此时

$$\mathbb{E}_\theta[g(X)] = \int g(x) f(x \mid \theta)\, \mathrm{d}x.$$

若 $g$ 本身不含 $\theta$，导数来自密度的变化：

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[g(X)] = \int g(x) \frac{\partial f(x \mid \theta)}{\partial \theta}\, \mathrm{d}x.$$

## 7.1 score function trick

只要 $f(x \mid \theta) > 0$，就有

$$\frac{\partial f(x \mid \theta)}{\partial \theta} = f(x \mid \theta) \frac{\partial}{\partial \theta} \log f(x \mid \theta).$$

因此

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[g(X)] = \mathbb{E}_\theta\left[g(X) \frac{\partial}{\partial \theta} \log f(X \mid \theta)\right].}$$

其中

$$s_\theta(X) = \frac{\partial}{\partial \theta} \log f(X \mid \theta)$$

称为 score。

**例 4.5：正态均值参数的 score**

设

$$X \sim N(\mu, \sigma^2), \qquad \sigma^2 \text{ 已知}.$$

密度为

$$f(x \mid \mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

对数密度为

$$\log f(x \mid \mu) = C - \frac{(x-\mu)^2}{2\sigma^2},$$

其中 $C$ 与 $\mu$ 无关。因此

$$\frac{\partial}{\partial \mu} \log f(x \mid \mu) = \frac{x-\mu}{\sigma^2}.$$

所以

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\mathbb{E}_\mu[g(X)] = \mathbb{E}_\mu\left[g(X) \frac{X-\mu}{\sigma^2}\right].$$

这在统计、计量、机器学习里的似然梯度和策略梯度中都会出现。

## 7.2 一个重要恒等式：score 的期望为零

因为密度积分恒等于 1：

$$\int f(x \mid \theta)\, \mathrm{d}x = 1.$$

对 $\theta$ 求导，如果可以积分下求导，则

$$\int f_\theta(x \mid \theta)\,\mathrm{d}x = 0.$$

又由于

$$f_\theta = f\frac{\partial}{\partial\theta}\log f,$$

所以

$$\boxed{\mathbb{E}_\theta[s_\theta(X)] = 0.}$$

**常见误区**

这个结论依赖支撑集不随参数变化，或者边界项能够正确处理。如果支撑集随参数变化，例如 $X \sim U[0,\theta]$，直接使用 $\mathbb{E}[s]=0$ 会出问题，因为密度的支撑 $[0,\theta]$ 本身在动。

**例 4.6：支撑随参数变化时的陷阱**

设

$$X \sim U[0,\theta], \qquad \theta > 0.$$

密度为

$$f(x \mid \theta) = \frac{1}{\theta}, \qquad 0 \le x \le \theta.$$

若只看对数密度，

$$\log f = -\log\theta, \qquad s_\theta = -\frac{1}{\theta}.$$

于是似乎

$$\mathbb{E}[s_\theta] = -\frac{1}{\theta} \neq 0,$$

这和“score 期望为零”矛盾。
矛盾来自哪里？来自支撑集 $[0,\theta]$ 随 $\theta$ 变化。归一化条件应该写为

$$\int_0^\theta \frac{1}{\theta}\,\mathrm{d}x = 1.$$

用 Leibniz rule 求导：

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\int_0^\theta \frac{1}{\theta}\,\mathrm{d}x = \frac{1}{\theta}\cdot 1 + \int_0^\theta \left(-\frac{1}{\theta^2}\right)\mathrm{d}x = \frac{1}{\theta} - \frac{1}{\theta} = 0.$$

边界项抵消了内部项。若忽略边界项，就会错误地得到 $-1/\theta$。

# 8 经济应用一：税率变化与社会福利

考虑收入 $y$ 的分布密度 $f(y)$。政府设置税率 $\tau$，税后消费为

$$c(y,\tau) = (1-\tau)y + T(\tau),$$

其中 $T(\tau)$ 是人均转移支付。社会福利为

$$W(\tau)=\int_0^{\infty} u(c(y,\tau))f(y)\,\mathrm{d}y.$$

如果积分下求导成立，则

$$W'(\tau)=\int_0^{\infty} u'(c(y,\tau))\frac{\partial c(y,\tau)}{\partial \tau}f(y)\,\mathrm{d}y.$$

由于

$$\frac{\partial c(y,\tau)}{\partial \tau}=-y+T'(\tau),$$

所以

$$\boxed{W'(\tau)=\int_0^{\infty} u'(c(y,\tau))[-y+T'(\tau)]f(y)\,\mathrm{d}y.}$$

**直觉**

这条式子非常有经济含义。对收入为 $y$ 的个体来说，税率上升的直接损失是 $-y$，但如果税收被均等返还，转移支付增加 $T'(\tau)$。两者相加是消费变化；再乘以边际效用 $u'(c)$，就是福利变化的权重。由于低收入者通常边际效用更高，福利效应不是简单的平均收入效应。

## 8.1 带政策阈值的福利：边界项出现

设某政策只补贴收入低于 $b(\tau)$ 的人，福利增量为

$$G(\tau)=\int_0^{b(\tau)} \Delta u(y,\tau)f(y)\,\mathrm{d}y.$$

由 Leibniz rule：

$$\boxed{G'(\tau)=\Delta u(b(\tau),\tau)f(b(\tau))b'(\tau)+\int_0^{b(\tau)}\frac{\partial \Delta u}{\partial \tau}(y,\tau)f(y)\,\mathrm{d}y.}$$

第一项是资格边界移动带来的新增或减少受益人；第二项是原本受益群体内部福利变化。

**常见误区**

做政策分析时，很多错误都来自漏掉“extensive margin”。在上式里，内部项是 intensive margin：已经被覆盖的人福利变多少；边界项是 extensive margin：覆盖范围扩大或缩小带来多少新人进入或退出。

# 9 经济应用二：需求、阈值与市场份额

离散选择、市场需求和平台经济中，经常有一个阈值规则：消费者类型 $v$ 大于价格 $p$ 时购买。若 $v$ 的密度为 $f(v)$，需求为

$$D(p)=\int_p^{\infty} f(v)\,\mathrm{d}v.$$

则

$$D'(p) = -f(p).$$

这就是“边际消费者”的密度。价格提高一点，流失的消费者正是估值接近 $p$ 的那一小群。

**例 4.7：利润最大化中的边界项**

垄断者边际成本为 $c$，消费者估值 $v \sim F$，密度为 $f$。需求

$$D(p) = 1 - F(p) = \int_p^\infty f(v)\,\mathrm{d}v.$$

利润

$$\pi(p) = (p-c)D(p).$$

求一阶条件。
首先

$$D'(p) = -f(p).$$

所以

$$\pi'(p) = D(p) + (p-c)D'(p) = 1 - F(p) - (p-c)f(p).$$

一阶条件为

$$\boxed{1 - F(p) = (p-c)f(p).}$$

或者

$$\boxed{p - c = \frac{1-F(p)}{f(p)}.}$$

右边是 inverse hazard rate。这个式子本质上来自对需求积分下限 $p$ 求导产生的边界项。

# 10 金融应用：从收益分布到风险度量

设资产收益 $R$ 的密度为 $f(r;\theta)$，损失函数为 $\ell(r)$。风险度量可能写成

$$\mathcal{L}(\theta) = \int \ell(r) f(r;\theta)\,\mathrm{d}r.$$

若 $\theta$ 表示波动率、尾部厚度或某个风险因子暴露，则

$$\mathcal{L}'(\theta) = \int \ell(r) f_\theta(r;\theta)\,\mathrm{d}r.$$

也可以写成 score 形式：

$$\mathcal{L}'(\theta) = \mathbb{E}_\theta[\ell(R) s_\theta(R)].$$

### 10.1 VaR 和 Expected Shortfall 中的边界

令损失 $L$ 的分布函数为 $F_L$，$\alpha$-VaR 定义为

$$q_\alpha = F_L^{-1}(\alpha).$$

Expected Shortfall 的一个连续型表达是

$$ES_\alpha = \mathbb{E}[L \mid L \geq q_\alpha] = \frac{1}{1-\alpha}\int_{q_\alpha}^{\infty} \ell f_L(\ell)\,\mathrm{d}\ell.$$

如果参数改变导致 $q_\alpha$ 改变，就会出现下限移动项。虽然本讲义暂时不深入 VaR 的严格可微性，但你要形成意识：**凡是分位数作为边界，求导时边界项必须被检查**。

## 11 贝叶斯归一化常数与 log-sum-exp

很多模型中会出现

$$Z(\theta) = \int \exp(v(x,\theta))\,\mathrm{d}x.$$

这叫归一化常数、配分函数或 partition function。定义

$$A(\theta) = \log Z(\theta).$$

则

$$A'(\theta) = \frac{Z'(\theta)}{Z(\theta)}.$$

如果可以积分下求导，

$$Z'(\theta) = \int \exp(v(x,\theta)) v_\theta(x,\theta)\,\mathrm{d}x.$$

因此

$$A'(\theta) = \frac{\int \exp(v(x,\theta)) v_\theta(x,\theta)\,\mathrm{d}x}{\int \exp(v(x,\theta))\,\mathrm{d}x}.$$

定义归一化后的密度

$$p_\theta(x) = \frac{\exp(v(x,\theta))}{Z(\theta)},$$

则

$$\boxed{A'(\theta) = \mathbb{E}_\theta[v_\theta(X,\theta)].}$$

**直觉**

这条式子在离散选择、贝叶斯统计、最大熵模型、机器学习和经济学中的 logit 模型中都非常常见。它说：log-normalizer 的导数等于某个充分统计量的模型期望。

**例 4.8：logit 中的 log-sum-exp 导数**

离散情形中，令

$$A(\theta) = \log\left(\sum_{j=1}^{J} e^{v_j(\theta)}\right).$$

则

$$A'(\theta) = \frac{\sum_{j=1}^{J} e^{v_j(\theta)} v_j'(\theta)}{\sum_{k=1}^{J} e^{v_k(\theta)}} = \sum_{j=1}^{J} s_j(\theta) v_j'(\theta),$$

其中

$$s_j(\theta) = \frac{e^{v_j(\theta)}}{\sum_{k=1}^{J} e^{v_k(\theta)}}.$$

这就是 logit choice probability。连续版本只是把求和换成积分。

# 12 更严谨一点：什么时候可以积分下求导

你不需要现在完整学习测度论，但必须知道“把导数推进积分号”不是永远合法。常见充分条件有两个层级。

## 12.1 紧区间上的连续性条件

如果积分区间是有限闭区间 $[a,b]$，且 $f$ 和 $f_\theta$ 连续，那么通常可以安全交换：

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\int_a^b f(x,\theta)\,\mathrm{d}x = \int_a^b f_\theta(x,\theta)\,\mathrm{d}x.$$

这是本科微积分中最常用的条件。

## 12.2 无限区间或概率空间上的支配条件

如果区间是无限的，如 $(-\infty,\infty)$，仅有点态可导不够。一个常用条件是存在可积函数 $M(x)$，使得

$$|f_\theta(x,\theta)| \le M(x)$$

在 $\theta$ 的邻域内成立。这样可以用 dominated convergence theorem 的思想来交换极限和积分。

**常见误区**

错误直觉：

$$\text{每个}x\text{ 上都可导} \quad \Longrightarrow \quad \text{积分后一定可导。}$$

这是错的。点态可导只控制每一个点，不控制整体积分。真正需要的是一种“整体控制”，例如一致收敛或支配收敛。

**例 4.9：一个危险信号**

考虑

$$F(\theta)=\int_0^{\infty} e^{-\theta x}\,\mathrm{d}x=\frac{1}{\theta},\qquad \theta>0.$$

对被积函数求偏导：

$$f_\theta(x,\theta)=-xe^{-\theta x}.$$

如果固定在 $\theta_0>0$ 附近，例如 $\theta\in[\theta_0/2,3\theta_0/2]$，则

$$|xe^{-\theta x}|\le xe^{-(\theta_0/2)x},$$

右边在 $[0,\infty)$ 上可积，因此可交换求导和积分。于是

$$F'(\theta)=\int_0^{\infty}-xe^{-\theta x}\,\mathrm{d}x=-\frac{1}{\theta^2}.$$

但如果 $\theta\to 0^+$，支配函数会失效，因为积分本身发散。这说明参数空间边界附近需要格外小心。

# 13 Worked Examples：完整样板题

**Worked Example 1：用积分下求导计算对数积分**

计算

$$I(a)=\int_0^1 \frac{x^a-1}{\log x}\,\mathrm{d}x,\qquad a>-1.$$

求 $I'(a)$，并由 $I(0)=0$ 求 $I(a)$。

**Step 1：先对参数求导。**

$$\frac{\partial}{\partial a}\left(\frac{x^a-1}{\log x}\right)=x^a.$$

所以

$$I'(a)=\int_0^1 x^a\,\mathrm{d}x=\frac{1}{a+1}.$$

**Step 2：对 $a$ 积分。**

$$I(a)=\int I'(a)\,\mathrm{d}a=\log(a+1)+C.$$

**Step 3：用初始值。**当 $a=0$ 时，

$$I(0)=\int_0^1 \frac{1-1}{\log x}\,\mathrm{d}x=0.$$

因此

$$0=\log 1+C,\qquad C=0.$$

所以

$$\boxed{I(a) = \log(a+1).}$$

**理解：**这个积分直接看很奇怪，因为分母有 $\log x$。但把 $a$ 当参数后，求导会消掉分母。这是含参积分的典型威力。

## Worked Example 2：政策阈值与边界效应

设收入 $y$ 在 $[0,1]$ 上均匀分布。政府给收入低于 $b(\tau) = \sqrt{\tau}$ 的人补贴 $\tau y$，福利增量为

$$G(\tau) = \int_0^{\sqrt{\tau}} \tau y \, \mathrm{d}y, \qquad 0 < \tau < 1.$$

用两种方法求 $G'(\tau)$。

**方法一：先积分。**

$$G(\tau) = \tau \cdot \frac{(\sqrt{\tau})^2}{2} = \frac{\tau^2}{2}.$$

所以

$$G'(\tau) = \tau.$$

**方法二：Leibniz rule。**这里

$$f(y,\tau) = \tau y, \qquad b(\tau) = \sqrt{\tau}, \qquad b'(\tau) = \frac{1}{2\sqrt{\tau}}.$$

因此

$$G'(\tau) = f(\sqrt{\tau},\tau) b'(\tau) + \int_0^{\sqrt{\tau}} f_\tau(y,\tau)\, \mathrm{d}y.$$

边界项为

$$f(\sqrt{\tau},\tau) b'(\tau) = \tau\sqrt{\tau} \cdot \frac{1}{2\sqrt{\tau}} = \frac{\tau}{2}.$$

内部项为

$$\int_0^{\sqrt{\tau}} y \, \mathrm{d}y = \frac{\tau}{2}.$$

所以

$$G'(\tau) = \frac{\tau}{2} + \frac{\tau}{2} = \tau.$$

**经济解释：**一半变化来自已经获得补贴的人补贴强度增加；另一半来自资格门槛扩大，新增受益人进入。

## Worked Example 3：似然 score 与一阶条件

设 $X_1, \dots, X_n$ 独立同分布于 $N(\mu, \sigma^2)$，$\sigma^2$ 已知。求 log-likelihood 对 $\mu$ 的导数，并得到 MLE。

单个观测的 log density 为

$$\ell_i(\mu) = C - \frac{(X_i - \mu)^2}{2\sigma^2}.$$

总体 log-likelihood：

$$\ell(\mu) = \sum_{i=1}^{n} \ell_i(\mu) = Cn - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2.$$

求导：

$$\ell'(\mu) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu).$$

一阶条件 $\ell'(\mu) = 0$：

$$\sum_{i=1}^{n} X_i - n\mu = 0.$$

所以

$$\boxed{\hat{\mu} = \bar{X}.}$$

**和本章的关系：** score 是对 log density 的参数导数。虽然这里没有显式积分，但 score 的基本性质，如期望为零、Fisher information，都来自“密度积分为 1 后对参数求导”。

## Worked Example 4：log-normalizer 的二阶导数

令

$$Z(\theta) = \int e^{\theta T(x)} q(x)\, \mathrm{d}x, \qquad A(\theta) = \log Z(\theta),$$

其中 $q(x) \geq 0$。证明

$$A'(\theta) = \mathbb{E}_\theta[T(X)], \qquad A''(\theta) = \mathrm{Var}_\theta(T(X)),$$

其中

$$p_\theta(x) = \frac{e^{\theta T(x)} q(x)}{Z(\theta)}.$$

**Step 1：一阶导数。**

$$Z'(\theta) = \int T(x) e^{\theta T(x)} q(x)\, \mathrm{d}x.$$

所以

$$A'(\theta) = \frac{Z'(\theta)}{Z(\theta)} = \int T(x) \frac{e^{\theta T(x)} q(x)}{Z(\theta)}\, \mathrm{d}x = \mathbb{E}_\theta[T(X)].$$

**Step 2：二阶导数。**

$$A''(\theta) = \frac{Z''Z - (Z')^2}{Z^2}.$$

而

$$Z''(\theta) = \int T(x)^2 e^{\theta T(x)} q(x)\,\mathrm{d}x.$$

因此

$$\frac{Z''}{Z} = \mathbb{E}_\theta[T(X)^2], \qquad \left(\frac{Z'}{Z}\right)^2 = \mathbb{E}_\theta[T(X)]^2.$$

所以

$$\boxed{A''(\theta) = \mathbb{E}_\theta[T(X)^2] - \mathbb{E}_\theta[T(X)]^2 = \mathrm{Var}_\theta(T(X)).}$$

**经济金融意义：** log-normalizer 的凸性来自方差非负。这是 logit、指数族、熵正则化、风险敏感控制中的基本数学结构。

# 14 常见错误清单

**常见误区**

1. **漏边界项**。只要上下限含参数，就先写 Leibniz rule。
2. **把积分变量和参数混淆**。$x$ 被积分消掉，$\theta$ 才是外部变量。
3. **忘记支撑集随参数变化**。Uniform $[0,\theta]$ 是典型陷阱。
4. **无条件交换导数与积分**。至少要检查连续性、紧区间，或支配条件。
5. **密度求导漏掉归一化约束**。密度积分永远等于 1，这个恒等式本身会给出 score 的限制。
6. **把“内部项”和“边界项”的经济含义混在一起**。政策分析中，前者常对应 intensive margin，后者常对应 extensive margin。

# 15 练习题

## A 组：基础计算

A1. 设 $F(\theta) = \int_0^1 e^{\theta x}\,\mathrm{d}x$。用积分下求导求 $F'(\theta)$。

A2. 设 $F(\theta) = \int_1^2 \log(x+\theta)\,\mathrm{d}x$。求 $F'(\theta)$。

A3. 设 $F(\theta) = \int_0^\theta \sin x\,\mathrm{d}x$。求 $F'(\theta)$。

A4. 设 $F(\theta) = \int_\theta^1 x^\theta\,\mathrm{d}x$，$0<\theta<1$。求 $F'(\theta)$。

A5. 设 $F(\theta) = \int_{\theta^2}^{e^\theta} (x+\theta^2)\,\mathrm{d}x$。求 $F'(\theta)$。

## B 组：概念判断

B1. 判断：如果 $f(x,\theta)$ 对 $\theta$ 可导，则一定有

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\int f(x,\theta)\,\mathrm{d}x = \int f_\theta(x,\theta)\,\mathrm{d}x.$$

对还是错？说明原因。

B2. 解释 Leibniz rule 中

$$f(b(\theta),\theta)b'(\theta)$$

的几何含义。

B3. 为什么 $X\sim U[0,\theta]$ 时，score 的期望不可以只通过 $\mathbb{E}[-1/\theta]$ 来判断？

B4. 在政策阈值模型

$$G(\tau)=\int_0^{b(\tau)} g(y,\tau)f(y)\,\mathrm{d}y$$

中，边界项对应什么经济含义？

## C 组：混合题

C1. 设

$$F(\theta)=\int_0^1\int_0^{\theta x} e^{xy}\,\mathrm{d}y\,\mathrm{d}x.$$

求 $F'(\theta)$。提示：先把内层积分看成上限含参数。

C2. 设

$$H(\theta)=\int_0^\infty xe^{-\theta x}\,\mathrm{d}x,\qquad \theta>0.$$

先求 $H(\theta)$，再用积分下求导求 $H'(\theta)$。

C3. 设 $D(p)=\int_p^\infty f(v)\,\mathrm{d}v$，$\pi(p)=(p-c)D(p)$。推导 $\pi'(p)$ 并解释每一项。

C4. 设

$$A(\theta)=\log\int_0^1 e^{\theta x}\,\mathrm{d}x.$$

求 $A'(\theta)$，并把答案写成某个期望。

## D 组：经济金融建模题

D1. 收入 $Y$ 密度为 $f(y)$。税后消费为 $c(y,\tau)=(1-\tau)y+T(\tau)$。福利

$$W(\tau)=\int u(c(y,\tau))f(y)\,\mathrm{d}y.$$

推导 $W'(\tau)$，并解释边际效用权重。

D2. 消费者估值 $v$ 的分布为 $F$，平台价格为 $p$，市场份额

$$s(p) = \int_p^{\infty} f(v)\,\mathrm{d}v.$$

平台收益 $R(p) = ps(p)$。推导一阶条件。

D3. 设资产损失 $L$ 密度为 $f(\ell;\theta)$，风险损失函数

$$\mathcal{L}(\theta) = \int \ell^2 f(\ell;\theta)\,\mathrm{d}\ell.$$

用 score 写出 $\mathcal{L}'(\theta)$。

D4. 设某补贴只给 $y \le b(\tau)$ 的人，补贴效用增量为 $h(y,\tau)$。写出总福利增量

$$G(\tau) = \int_0^{b(\tau)} h(y,\tau) f(y)\,\mathrm{d}y$$

的导数，并解释 intensive margin 和 extensive margin。

## E 组：挑战与证明题

E1. 证明：若

$$Z(\theta) = \int e^{\theta T(x)} q(x)\,\mathrm{d}x, \qquad A(\theta) = \log Z(\theta),$$

则 $A''(\theta) = \mathrm{Var}_\theta(T(X))$。

E2. 设

$$I(a,b) = \int_0^{\infty} e^{-ax}\cos(bx)\,\mathrm{d}x, \qquad a > 0.$$

已知 $I(a,b) = a/(a^2+b^2)$。通过对 $b$ 求导，计算

$$\int_0^{\infty} x e^{-ax}\sin(bx)\,\mathrm{d}x.$$

E3. 设 $f(x \mid \theta)$ 的支撑不随 $\theta$ 变化，并且可交换积分和求导。证明

$$\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right] = 0.$$

# 16 练习题详细解答

## A 组答案

### A1

$$F'(\theta)=\int_0^1 xe^{\theta x}\,\mathrm{d}x.$$

若完全算出，分部积分：

$$\int_0^1 xe^{\theta x}\,\mathrm{d}x=\left[\frac{xe^{\theta x}}{\theta}\right]_0^1-\frac{1}{\theta}\int_0^1 e^{\theta x}\,\mathrm{d}x=\frac{e^\theta}{\theta}-\frac{e^\theta-1}{\theta^2}.$$

所以

$$F'(\theta)=\frac{(\theta-1)e^\theta+1}{\theta^2},\quad \theta\neq 0.$$

### A2

固定区间，直接积分下求导：

$$F'(\theta)=\int_1^2\frac{1}{x+\theta}\,\mathrm{d}x=\log(2+\theta)-\log(1+\theta).$$

### A3

上限含参数：

$$F'(\theta)=\sin\theta.$$

也可先积分：$F(\theta)=1-\cos\theta$，导数为 $\sin\theta$。

### A4

这里

$$a(\theta)=\theta,\qquad b(\theta)=1,\qquad f(x,\theta)=x^\theta.$$

所以

$$f_\theta=x^\theta\log x.$$

Leibniz rule 给出

$$F'(\theta)=0-\theta^\theta\cdot 1+\int_\theta^1 x^\theta\log x\,\mathrm{d}x.$$

即

$$\boxed{F'(\theta)=-\theta^\theta+\int_\theta^1 x^\theta\log x\,\mathrm{d}x.}$$

**A5**

$$a(\theta)=\theta^2,\quad b(\theta)=e^{\theta},\quad f(x,\theta)=x+\theta^2.$$

因此

$$a'=2\theta,\quad b'=e^{\theta},\quad f_\theta=2\theta.$$

所以

$$F'(\theta)=(e^{\theta}+\theta^2)e^{\theta}-(\theta^2+\theta^2)(2\theta)+\int_{\theta^2}^{e^{\theta}}2\theta\,\mathrm{d}x.$$

化简：

$$F'(\theta)=e^{2\theta}+\theta^2e^{\theta}-4\theta^3+2\theta(e^{\theta}-\theta^2).$$

## B 组答案

**B1**

错。点态可导不够，还要能交换极限和积分。有限闭区间上，若 $f_\theta$ 连续，通常可以；无限区间或概率空间上，需要支配条件等。否则可能出现导数在局部爆炸，导致积分与极限不能交换。

**B2**

$f(b(\theta),\theta)b'(\theta)$ 是右边界移动带来的面积变化率。若 $b'(\theta)>0$，上限向右扩张，新增加一条宽度约 $b'(\theta)\,\mathrm{d}\theta$、高度约 $f(b(\theta),\theta)$ 的窄条，所以贡献为 $f(b(\theta),\theta)b'(\theta)$。

**B3**

因为 $U[0,\theta]$ 的支撑集随 $\theta$ 变化。归一化条件是

$$\int_0^{\theta}\frac{1}{\theta}\,\mathrm{d}x=1.$$

求导时有边界项：

$$\frac{1}{\theta}-\int_0^{\theta}\frac{1}{\theta^2}\,\mathrm{d}x=0.$$

如果只对 log density 求导，会漏掉边界项。

**B4**

边界项对应 extensive margin，即资格线变化导致进入或退出政策覆盖范围的人群。内部积分项对应 intensive margin，即已经在覆盖范围内的人，其福利强度如何随政策变化。

# C 组答案

## C1

把

$$F(\theta)=\int_0^1\left[\int_0^{\theta x}e^{xy}\,\mathrm{d}y\right]\mathrm{d}x.$$

对内层上限求导：

$$\frac{\partial}{\partial\theta}\int_0^{\theta x}e^{xy}\,\mathrm{d}y=e^{x(\theta x)}\cdot x=xe^{\theta x^2}.$$

所以

$$\boxed{F'(\theta)=\int_0^1 xe^{\theta x^2}\,\mathrm{d}x.}$$

进一步令 $u=\theta x^2$，可得

$$F'(\theta)=\frac{e^\theta-1}{2\theta},\qquad \theta\neq 0.$$

## C2

先计算：

$$H(\theta)=\int_0^\infty xe^{-\theta x}\,\mathrm{d}x=\frac{1}{\theta^2}.$$

因此

$$H'(\theta)=-\frac{2}{\theta^3}.$$

用积分下求导：

$$H'(\theta)=\int_0^\infty -x^2e^{-\theta x}\,\mathrm{d}x=-\frac{2}{\theta^3}.$$

二者一致。

## C3

需求导数：

$$D'(p)=-f(p).$$

利润：

$$\pi'(p)=D(p)+(p-c)D'(p)=D(p)-(p-c)f(p).$$

第一项 $D(p)$ 是价格升高对每个仍购买者多收的钱；第二项 $(p-c)f(p)$ 是边际消费者流失造成的利润损失。

**C4**

令

$$Z(\theta) = \int_0^1 e^{\theta x}\,\mathrm{d}x.$$

则

$$A'(\theta) = \frac{Z'(\theta)}{Z(\theta)} = \frac{\int_0^1 x e^{\theta x}\,\mathrm{d}x}{\int_0^1 e^{\theta x}\,\mathrm{d}x}.$$

定义

$$p_\theta(x) = \frac{e^{\theta x}}{\int_0^1 e^{\theta t}\,\mathrm{d}t}, \qquad x \in [0,1].$$

则

$$\boxed{A'(\theta) = \mathbb{E}_\theta[X].}$$

## D 组答案

**D1**

$$W'(\tau) = \int u'(c(y,\tau))c_\tau(y,\tau)f(y)\,\mathrm{d}y.$$

因为

$$c_\tau(y,\tau) = -y + T'(au),$$

所以

$$\boxed{W'(\tau) = \int u'(c(y,\tau))[-y + T'(\tau)]f(y)\,\mathrm{d}y.}$$

边际效用 $u'(c)$ 是福利权重。低消费群体通常边际效用更高，因此同样一单位消费变化对社会福利的贡献更大。

**D2**

市场份额

$$s(p) = 1 - F(p), \qquad s'(p) = -f(p).$$

收益

$$R(p) = ps(p).$$

所以

$$R'(p) = s(p) + ps'(p) = 1 - F(p) - pf(p).$$

一阶条件为

$$\boxed{1 - F(p) = pf(p).}$$

左边是当前购买概率，右边是提价导致边际消费者流失造成的收益损失。

**D3**

$$\mathcal{L}'(\theta) = \int \ell^2 f_\theta(\ell;\theta)\,\mathrm{d}\ell.$$

使用 score

$$s_\theta(\ell) = \frac{\partial}{\partial\theta}\log f(\ell;\theta),$$

则

$$f_\theta = f s_\theta.$$

所以

$$\boxed{\mathcal{L}'(\theta) = \mathbb{E}_\theta[L^2 s_\theta(L)].}$$

**D4**

由 Leibniz rule：

$$G'(\tau) = h(b(\tau),\tau)f(b(\tau))b'(\tau) + \int_0^{b(\tau)} h_\tau(y,\tau)f(y)\,\mathrm{d}y.$$

第一项是 extensive margin：资格门槛变化导致新个体进入或退出。第二项是 intensive margin：原有受益者的补贴强度或效用增量改变。

## E 组答案

**E1**

与 Worked Example 4 相同。首先

$$A'(\theta) = \frac{Z'(\theta)}{Z(\theta)} = \mathbb{E}_\theta[T(X)].$$

再求导：

$$A''(\theta) = \frac{Z''Z - (Z')^2}{Z^2}.$$

由于

$$\frac{Z''}{Z} = \mathbb{E}_\theta[T^2], \qquad \frac{Z'}{Z} = \mathbb{E}_\theta[T],$$

因此

$$A''(\theta) = \mathbb{E}_\theta[T^2] - \mathbb{E}_\theta[T]^2 = \mathrm{Var}_\theta(T).$$

**E2**

已知

$$I(a,b) = \int_0^\infty e^{-ax}\cos(bx)\,\mathrm{d}x = \frac{a}{a^2+b^2}.$$

对 $b$ 求导。左边：

$$\frac{\partial I}{\partial b} = \int_0^\infty e^{-ax}[-x\sin(bx)]\,\mathrm{d}x = -\int_0^\infty xe^{-ax}\sin(bx)\,\mathrm{d}x.$$

右边：

$$\frac{\partial}{\partial b}\frac{a}{a^2+b^2} = -\frac{2ab}{(a^2+b^2)^2}.$$

所以

$$-\int_0^\infty xe^{-ax}\sin(bx)\,\mathrm{d}x = -\frac{2ab}{(a^2+b^2)^2}.$$

因此

$$\boxed{\int_0^\infty xe^{-ax}\sin(bx)\,\mathrm{d}x = \frac{2ab}{(a^2+b^2)^2}.}$$

## E3

密度归一化：

$$\int f(x\mid\theta)\,\mathrm{d}x = 1.$$

对 $\theta$ 求导：

$$\int f_\theta(x\mid\theta)\,\mathrm{d}x = 0.$$

因为

$$f_\theta(x\mid\theta) = f(x\mid\theta)\frac{\partial}{\partial\theta}\log f(x\mid\theta),$$

所以

$$\int f(x\mid\theta)\frac{\partial}{\partial\theta}\log f(x\mid\theta)\,\mathrm{d}x = 0.$$

这正是

$$\boxed{\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X\mid\theta)\right] = 0.}$$

# 17 本章总结图谱

| 概念 | 核心公式 | 经济金融解释 |
| --- | --- | --- |
| 固定区间积分下求导 | $\frac{\mathrm{d}}{\mathrm{d}\theta}\int_a^b f(x,\theta)\,\mathrm{d}x = \int_a^b f_\theta(x,\theta)\,\mathrm{d}x$ | 同一群体或同一区域内，每一点的高度/效用/密度随参数变化。 |
| Leibniz rule | $F' = f(b)b' - f(a)a' + \int_a^b f_\theta\,\mathrm{d}x$ | 总变化 = 右边界新增 - 左边界移出 + 内部强度变化。 |
| 期望下求导 | $\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}[g(X,\theta)] = \mathbb{E}[g_\theta(X,\theta)]$ | 期望效用、期望损失、风险指标对参数变化的反应。 |
| Score function | $\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[g(X)] = \mathbb{E}_\theta[g(X)s_\theta(X)]$ | 似然梯度、计量估计、模拟梯度和风险敏感分析。 |
| Score 期望为零 | $\mathbb{E}_\theta[s_\theta(X)] = 0$ | 来自密度积分等于 1；但支撑随参数变化时要小心边界项。 |
| 政策阈值 | $G' = h(b)b' + \int_0^b h_\tau\,\mathrm{d}y$ | 边界项是 extensive margin，内部项是 intensive margin。 |
| Log-normalizer | $A'(\theta) = \mathbb{E}_\theta[v_\theta(X,\theta)]$ | logit、贝叶斯归一化、指数族和熵正则化的基础。 |

**本章压缩总结**

本章要带走的核心不是某一个难积分，而是一种思维：

$$\text{一个总量} = \int \text{局部对象}\,\mathrm{d}\,\text{分布或区域}.$$

当参数变化时，总量的变化可能来自三处：局部对象变了、边界变了、分布或密度变了。固定区间积分下求导、Leibniz rule 和 score function，分别处理这三类变化。以后学优化、包络定理、极大似然、GMM、风险度量、动态规划时，这套语言会反复出现。

# 18 下一章连接

下一章进入向量分析：**向量场、梯度与全微分**。如果说本章研究的是“一维参数变化如何影响积分总量”，下一章则研究多维变量一起变化时，函数如何被局部线性化：

$$\mathrm{d}U = U_x\,\mathrm{d}x + U_y\,\mathrm{d}y.$$

# 经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

---

# Chapter 5

## 向量场、梯度与全微分

---

本章开始进入向量分析，但目标不是照搬数学系的曲线积分技巧，而是建立经济金融模型里最常见的几何语言：梯度是边际效应向量，全微分是局部线性近似，exact differential 是路径无关的福利变化。

math4econ personalized notes
2026

# 目录

**Chapter 5 总览** **2**

**1 本章为什么重要** **3**

**2 最小前置知识** **3**

**3 从向量到场：经济对象的几何化** **4**
3.1 向量：一组有方向的数 . . . . . 4
3.2 标量场和向量场 . . . . . 4

**4 梯度：边际效应向量** **5**
4.1 梯度的定义 . . . . . 5
4.2 梯度和等值线 . . . . . 6
4.3 经济解释：MRS 是梯度分量的比值 . . . . . 7

**5 方向导数：沿任意方向的边际变化** **7**
5.1 为什么梯度是最陡上升方向 . . . . . 8

**6 全微分：多变量同时变化的局部总效应** **9**
6.1 从一元线性近似到多元线性近似 . . . . . 9
6.2 经济解释：全微分就是局部比较静态 . . . . . 10

**7 全微分和弹性、半弹性** **10**

**8 Exact differential：什么时候总变化路径无关** **11**
8.1 二维 exactness 判别 . . . . . 12

**9 路径依赖和福利变化** **13**

**10 经济金融中的四个核心应用** **14**
10.1 应用一：消费者理论中的边际效用向量 . . . . . 14
10.2 应用二：成本函数和 Shephard lemma 的预备 . . . . . 14
10.3 应用三：投资组合方差的梯度 . . . . . 15
10.4 应用四：梯度下降与最小化损失 . . . . . 16

**11 Worked Examples：完整样板题** **16**

**12 常见误区整理** **19**

**13 本章总结图谱** **20**

**14 练习题** **21**
A 组：基础计算 . . . . . 21
B 组：概念判断 . . . . . 21
C 组：混合题 . . . . . 21
D 组：经济金融应用题 . . . . . 21
E 组：挑战/证明题 . . . . . 22

**15 练习题详细解答** **23**
A 组答案 . . . . . 23
B 组答案 . . . . . 24
C 组答案 . . . . . 25
D 组答案 . . . . . 26
E 组答案 . . . . . 28

**本章结束语** **31**

# Chapter 5 总览

**本章目标**

学完本章，你应该达到以下目标：

1. 能区分标量场 scalar field 与向量场 vector field，并知道经济学里哪些对象属于哪一类；
2. 理解梯度 $\nabla f$ 的三重含义：偏导数组合、最陡上升方向、等值线的法向量；
3. 能把效用函数、成本函数、利润函数、投资组合风险函数的梯度解释为边际效应向量；
4. 掌握方向导数的公式
$$D_{\boldsymbol{u}} f(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) \cdot \boldsymbol{u},$$
并能说明为什么梯度方向给出最大上升率；
5. 理解全微分
$$\mathrm{d}f = f_{x_1}\,\mathrm{d}x_1 + \cdots + f_{x_n}\,\mathrm{d}x_n$$
不是形式记号，而是多元函数的一阶局部近似；
6. 掌握 exact differential 与 potential function 的判别，理解路径无关和福利测度之间的关系；
7. 能完成本章的经济金融应用：MRS、成本变化、投资组合方差梯度、福利变化路径依赖。

**直觉**

一句话：**梯度把“很多个方向上的边际变化”压缩成一个向量；全微分把“多变量同时小幅变化的总效果”压缩成一个线性近似。**
在一元函数里，你只有一个方向，所以导数 $f'(x)$ 就够了。在经济金融模型里，一个对象通常同时依赖很多变量：价格、收入、利率、风险、税率、资本、劳动、消费束。于是你需要一个向量把所有边际效应放在一起，这就是梯度。

**本章压缩总结**

前四章解决的是“怎样累加”：二重积分、三重积分、变量替换、含参积分。第 5 章转向“怎样描述局部变化”。后面 Lagrange 乘子、KKT、隐函数定理、包络定理、动态系统、最优化，都建立在本章的语言上。

# 1 本章为什么重要

很多经济学模型看起来不复杂，但只要写成多变量形式，就已经需要向量分析语言。例如消费者问题：

$$\max_{x_1,x_2} u(x_1,x_2) \quad \text{s.t.} \quad p_1x_1+p_2x_2=w.$$

中微里你可能会说：一阶条件是

$$\frac{u_1}{u_2}=\frac{p_1}{p_2}.$$

高一点的视角是：

$$\nabla u(x_1,x_2)=\lambda(p_1,p_2).$$

这句话的含义非常丰富：效用上升最快的方向和预算约束的法向量平行；在最优点，不能沿着预算线方向继续提高效用。

再比如投资组合方差：

$$\sigma_p^2=\boldsymbol{w}'\Sigma\boldsymbol{w}.$$

你要研究某个资产权重 $w_i$ 小幅增加时风险如何变化，不能只盯着单个偏导，而要看整个梯度：

$$\nabla_{\boldsymbol{w}}(\boldsymbol{w}'\Sigma\boldsymbol{w})=2\Sigma\boldsymbol{w}.$$

这就是每个资产对组合风险的边际贡献。

**方法**

本章的关键词可以记成三组：

| 数学语言 | 经济金融解释 | 后续章节用途 |
|---|---|---|
| $\nabla f$ | 边际效应向量 | 最优化、MRS、FOC |
| $\mathrm{d}f$ | 多变量同时变化的总效应 | 比较静态、福利变化 |
| exact differential | 路径无关的总变化 | 曲线积分、福利测度、势函数 |
| vector field | 每个状态点上的方向/力量 | 动态系统、梯度下降、均衡调整 |
| level set | 等值线/无差异曲线/等成本线 | 几何最优化 |

# 2 最小前置知识

本章默认你知道：

1. 向量加法、数乘、内积：$\boldsymbol{a}\cdot\boldsymbol{b}=\sum_i a_ib_i$；
2. 欧氏范数：$\|\boldsymbol{a}\|=\sqrt{\boldsymbol{a}\cdot\boldsymbol{a}}$；
3. 多元函数偏导数：$f_{x_i}=\partial f/\partial x_i$；

4. 一元函数线性近似：$f(x+\Delta x)\approx f(x)+f'(x)\Delta x$；
5. 等值线、无差异曲线的基本直觉。

如果这些知识不熟，也不影响阅读本章；本章会在使用时顺手补足。

# 3 从向量到场：经济对象的几何化

## 3.1 向量：一组有方向的数

**定义 5.1：向量、内积与范数**

在 $\mathbb{R}^n$ 中，一个向量写成

$$\boldsymbol{x}=(x_1,x_2,\dots,x_n)'.$$

两个向量的内积是

$$\boldsymbol{x}\cdot\boldsymbol{y}=x_1y_1+\cdots+x_ny_n.$$

向量的长度或范数是

$$\|\boldsymbol{x}\|=\sqrt{x_1^2+\cdots+x_n^2}.$$

若 $\|\boldsymbol{u}\|=1$，则 $\boldsymbol{u}$ 称为单位向量。

经济学里很多对象自然就是向量：

$$\boldsymbol{x}=(x_1,\dots,x_n)\quad \text{消费束},$$

$$\boldsymbol{p}=(p_1,\dots,p_n)\quad \text{价格向量},$$

$$\boldsymbol{w}=(w_1,\dots,w_n)\quad \text{投资组合权重},$$

$$\boldsymbol{\beta}=(\beta_1,\dots,\beta_k)\quad \text{回归系数向量}.$$

## 3.2 标量场和向量场

**定义 5.2：标量场 scalar field**

一个标量场是从空间中的点到一个实数的函数：

$$f:D\subseteq\mathbb{R}^n\to\mathbb{R}.$$

也就是说，每个点 $\boldsymbol{x}$ 都对应一个数 $f(\boldsymbol{x})$。

典型经济金融例子包括：

- 效用函数：$u(x_1,x_2)$；
- 成本函数：$C(q,w_1,w_2)$；

- 利润函数：$\pi(p, w)$；
- 损失函数：$L(\boldsymbol{\theta})$；
- 投资组合方差：$\boldsymbol{w}'\Sigma\boldsymbol{w}$。

**定义 5.3：向量场 vector field**

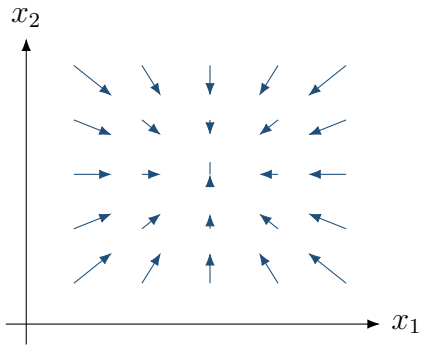
一个向量场是从空间中的点到一个向量的函数：

$$\boldsymbol{F} : D \subseteq \mathbb{R}^n \to \mathbb{R}^n, \quad \boldsymbol{F}(\boldsymbol{x}) = \big(F_1(\boldsymbol{x}), \ldots, F_n(\boldsymbol{x})\big).$$

也就是说，每个点 $\boldsymbol{x}$ 上都有一个箭头。

经济金融中常见的向量场包括：

- 动态系统的速度场：$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x})$；
- 梯度场：$\nabla f(\boldsymbol{x})$；
- excess demand field：$\boldsymbol{z}(\boldsymbol{p})$；
- 梯度下降中的更新方向：$-\nabla L(\boldsymbol{\theta})$。




一个简单向量场：每个点都有一个箭头

图 1：向量场的直观图像：它不是一个数值表，而是整个空间上的方向结构。

# 4 梯度：边际效应向量

## 4.1 梯度的定义

**定义 5.4：梯度 gradient**

若 $f:\mathbb{R}^n\to\mathbb{R}$ 在点 $\boldsymbol{x}$ 的各偏导数存在，则梯度定义为

$$\nabla f(\boldsymbol{x})=\begin{pmatrix}\frac{\partial f}{\partial x_1}(\boldsymbol{x})\\ \vdots\\ \frac{\partial f}{\partial x_n}(\boldsymbol{x})\end{pmatrix}.$$

如果把 $f$ 看成经济结果，那么 $\nabla f$ 就是所有输入变量的边际影响合在一起。

例如效用函数 $u(x_1,x_2)=x_1^{\alpha}x_2^{1-\alpha}$ 的梯度为

$$\nabla u(x_1,x_2)=\begin{pmatrix}\alpha x_1^{\alpha-1}x_2^{1-\alpha}\\ (1-\alpha)x_1^{\alpha}x_2^{-\alpha}\end{pmatrix}.$$

这里第一个分量是商品 1 的边际效用，第二个分量是商品 2 的边际效用。

**直觉**

一元导数告诉你：“往右走一点，函数变多少。”
梯度告诉你：“在每一个坐标方向上各走一点，函数分别变多少；如果你允许向任意方向走，哪个方向上升最快。”

## 4.2 梯度和等值线

等值线 level set 是满足

$$f(x_1,x_2)=c$$

的点的集合。经济学中的无差异曲线就是效用函数的等值线；等成本线、等产量线也是类似对象。

梯度和等值线有一个极重要的关系：**梯度垂直于等值线**。

**证明 5.1：梯度垂直于等值线**

设一条等值线可以写成参数曲线

$$\boldsymbol{r}(t)=(x_1(t),x_2(t)),$$

且沿着这条线有

$$f(\boldsymbol{r}(t))=c.$$

对 $t$ 求导，利用链式法则：

$$\frac{\mathrm{d}}{\mathrm{d}t}f(\boldsymbol{r}(t))=\nabla f(\boldsymbol{r}(t))\cdot\boldsymbol{r}'(t)=0.$$

这里 $\boldsymbol{r}'(t)$ 是等值线的切向量。所以梯度和等值线切向量内积为 0，即梯度垂直于等值线。

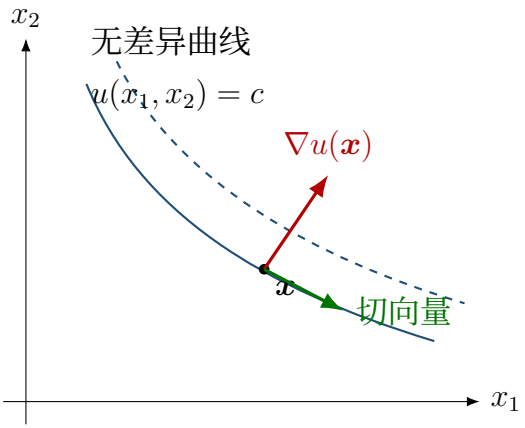



图 2: 梯度垂直于等值线。消费者最优的一阶条件可以被理解为：预算线切向方向上已经没有提高效用的空间。

### 4.3 经济解释：MRS 是梯度分量的比值

设效用函数为 $u(x_1, x_2)$。沿着同一条无差异曲线，效用不变，所以

$$\mathrm{d}u = u_1\,\mathrm{d}x_1 + u_2\,\mathrm{d}x_2 = 0.$$

于是

$$\frac{\mathrm{d}x_2}{\mathrm{d}x_1} = -\frac{u_1}{u_2}.$$

边际替代率通常定义为愿意放弃多少商品 2 来换一单位商品 1：

$$\mathrm{MRS}_{12} = \frac{u_1}{u_2}.$$

这说明 MRS 本质上是效用梯度两个分量的比值。

**常见误区**

不要把 $u_1/u_2$ 只理解为“一个公式”。它表达的是：在效用不变的局部条件下，两个商品的边际效用如何换算成彼此的补偿量。它来自全微分 $\mathrm{d}u = 0$，不是凭空出现的。

## 5 方向导数：沿任意方向的边际变化

偏导数只看坐标轴方向。但在经济问题中，变量往往不是只沿某个坐标轴变化。例如价格向量整体上升、投资组合权重沿某个方向调整、税率和补贴同时变化。这时需要方向导数。

**定义 5.5：方向导数**

设 $\boldsymbol{u}$ 是单位向量。函数 $f$ 在点 $\boldsymbol{x}$ 沿方向 $\boldsymbol{u}$ 的方向导数定义为

$$D_{\boldsymbol{u}} f(\boldsymbol{x}) = \lim_{h \to 0} \frac{f(\boldsymbol{x} + h\boldsymbol{u}) - f(\boldsymbol{x})}{h},$$

若极限存在。

如果 $f$ 可微，则方向导数有简单公式：

$$D_{\boldsymbol{u}} f(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) \cdot \boldsymbol{u}.$$

**证明 5.2：方向导数公式**

令一元函数

$$g(h) = f(\boldsymbol{x} + h\boldsymbol{u}).$$

那么

$$D_{\boldsymbol{u}} f(\boldsymbol{x}) = g'(0).$$

由多元链式法则，

$$g'(h) = \nabla f(\boldsymbol{x} + h\boldsymbol{u}) \cdot \boldsymbol{u}.$$

令 $h = 0$，得到

$$D_{\boldsymbol{u}} f(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) \cdot \boldsymbol{u}.$$

## 5.1 为什么梯度是最陡上升方向

由 Cauchy-Schwarz 不等式，若 $\|\boldsymbol{u}\| = 1$，则

$$D_{\boldsymbol{u}} f(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) \cdot \boldsymbol{u} \le \|\nabla f(\boldsymbol{x})\| \, \|\boldsymbol{u}\| = \|\nabla f(\boldsymbol{x})\| \, .$$

等号当且仅当 $\boldsymbol{u}$ 与 $\nabla f(\boldsymbol{x})$ 同方向时成立。因此：

$$\max_{\|\boldsymbol{u}\|=1} D_{\boldsymbol{u}} f(\boldsymbol{x}) = \|\nabla f(\boldsymbol{x})\| \, ,$$

并且最大上升方向是

$$\boldsymbol{u}^* = \frac{\nabla f(\boldsymbol{x})}{\|\nabla f(\boldsymbol{x})\|}.$$

**直觉**

梯度不是随便把偏导数堆成一列。它是一个最有效的“方向总结器”：

- 梯度方向：函数上升最快的方向；
- 梯度长度：最快上升的速度；
- 负梯度方向：函数下降最快的方向。

这就是为什么机器学习和数值优化中会出现 gradient descent。

# 6 全微分：多变量同时变化的局部总效应

## 6.1 从一元线性近似到多元线性近似

一元函数的线性近似是

$$f(x+\Delta x)-f(x)\approx f'(x)\Delta x.$$

多元函数的对应形式是

$$f(\boldsymbol{x}+\Delta\boldsymbol{x})-f(\boldsymbol{x})\approx \nabla f(\boldsymbol{x})\cdot\Delta\boldsymbol{x}.$$

如果 $\boldsymbol{x}=(x_1,$
$ldots,x_n)$，就写成

$$\Delta f\approx f_{x_1}\Delta x_1+\cdots+f_{x_n}\Delta x_n.$$

把 $\Delta x_i$ 写成 $\mathrm{d}x_i$，得到全微分：

$$\mathrm{d}f=f_{x_1}\,\mathrm{d}x_1+\cdots+f_{x_n}\,\mathrm{d}x_n.$$

**定义 5.6：全微分 total differential**

若 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $\boldsymbol{x}$ 可微，则其全微分为线性映射

$$\mathrm{d}f_{\boldsymbol{x}}(\,\mathrm{d}\boldsymbol{x})=\nabla f(\boldsymbol{x})\cdot\,\mathrm{d}\boldsymbol{x}.$$

在坐标形式下，写成

$$\mathrm{d}f=f_{x_1}\,\mathrm{d}x_1+\cdots+f_{x_n}\,\mathrm{d}x_n.$$

**证明 5.3：全微分来自可微性的定义**

函数 $f$ 在 $\boldsymbol{x}$ 可微的意思是存在一个线性映射 $L$，使得

$$f(\boldsymbol{x}+\boldsymbol{h})-f(\boldsymbol{x})=L(\boldsymbol{h})+r(\boldsymbol{h}),$$

且

$$\frac{r(\boldsymbol{h})}{\|\boldsymbol{h}\|}\to 0\quad(\boldsymbol{h}\to 0).$$

对于 $\mathbb{R}^n\to\mathbb{R}$ 的可微函数，这个线性映射正是

$$L(\boldsymbol{h})=\nabla f(\boldsymbol{x})\cdot\boldsymbol{h}.$$

所以

$$f(\boldsymbol{x}+\boldsymbol{h})-f(\boldsymbol{x})=\nabla f(\boldsymbol{x})\cdot\boldsymbol{h}+o(\|\boldsymbol{h}\|).$$

将 $\boldsymbol{h}$ 理解为小变化 $\mathrm{d}\boldsymbol{x}$，就得到

$$\mathrm{d}f=\nabla f(\boldsymbol{x})\cdot\,\mathrm{d}\boldsymbol{x}.$$

## 6.2 经济解释：全微分就是局部比较静态

设需求函数为

$$q = D(p, y),$$

其中 $p$ 是价格，$y$ 是收入。若价格和收入同时变化，需求变化近似为

$$\mathrm{d}q = D_p\,\mathrm{d}p + D_y\,\mathrm{d}y.$$

这句话比单独看 $D_p$ 或 $D_y$ 更接近现实，因为政策和市场冲击往往不是单变量冲击。

设成本函数为 $C(q, w)$，其中 $q$ 是产量，$w$ 是工资。则

$$\mathrm{d}C = C_q\,\mathrm{d}q + C_w\,\mathrm{d}w.$$

这可以解释为：总成本变化来自产量变化造成的边际成本效应，以及工资变化造成的要素价格效应。

**例 5.1：Cobb-Douglas 效用的全微分**

设

$$u(x, y) = x^{1/2}y^{1/2}.$$

则

$$u_x = \frac{1}{2}x^{-1/2}y^{1/2}, \qquad u_y = \frac{1}{2}x^{1/2}y^{-1/2}.$$

所以

$$\mathrm{d}u = \frac{1}{2}\sqrt{\frac{y}{x}}\,\mathrm{d}x + \frac{1}{2}\sqrt{\frac{x}{y}}\,\mathrm{d}y.$$

在点 $(x, y) = (4, 9)$，

$$\mathrm{d}u = \frac{3}{4}\,\mathrm{d}x + \frac{1}{3}\,\mathrm{d}y.$$

若 $\mathrm{d}x = 0.2$, $\mathrm{d}y = -0.3$，则

$$\mathrm{d}u \approx \frac{3}{4}(0.2) + \frac{1}{3}(-0.3) = 0.15 - 0.10 = 0.05.$$

这说明在该点附近，增加少量商品 $x$ 带来的效用上升略大于减少商品 $y$ 带来的效用损失。

# 7 全微分和弹性、半弹性

全微分不仅用于水平变量，也可以用于对数变量。经济学里大量比较静态其实是对数微分。

如果 $y = f(x)$，则

$$\mathrm{d}\ln y = \frac{\mathrm{d}y}{y}, \qquad \mathrm{d}\ln x = \frac{\mathrm{d}x}{x}.$$

弹性为

$$\varepsilon_{y,x} = \frac{\partial y}{\partial x}\frac{x}{y} = \frac{\partial \ln y}{\partial \ln x}.$$

对于多变量函数 $y=f(x_1,\dots,x_n)$，有

$$\mathrm{d}\ln y \approx \sum_{i=1}^{n} \frac{\partial \ln y}{\partial \ln x_i}\,\mathrm{d}\ln x_i.$$

**例 5.2：Cobb-Douglas 生产函数的对数微分**

设

$$Y = AK^{\alpha}L^{1-\alpha}.$$

取对数：

$$\ln Y = \ln A + \alpha \ln K + (1-\alpha)\ln L.$$

全微分得到

$$\mathrm{d}\ln Y = \mathrm{d}\ln A + \alpha\,\mathrm{d}\ln K + (1-\alpha)\,\mathrm{d}\ln L.$$

这句话在增长核算中非常重要：产出增长率近似等于技术增长率、资本增长率和劳动增长率的加权和。

# 8 Exact differential：什么时候总变化路径无关

前面我们从标量函数 $f$ 出发得到全微分：

$$\mathrm{d}f = f_x\,\mathrm{d}x + f_y\,\mathrm{d}y.$$

反过来，如果有人直接给你一个表达式

$$M(x,y)\,\mathrm{d}x + N(x,y)\,\mathrm{d}y,$$

你要问：它是不是某个函数 $F(x,y)$ 的全微分？如果是，那么它就叫 exact differential。

**定义 5.7：exact differential 与势函数**

若存在函数 $F(x,y)$，使得

$$\mathrm{d}F = F_x\,\mathrm{d}x + F_y\,\mathrm{d}y = M(x,y)\,\mathrm{d}x + N(x,y)\,\mathrm{d}y,$$

也就是

$$F_x = M, \qquad F_y = N,$$

则称

$$M(x,y)\,\mathrm{d}x + N(x,y)\,\mathrm{d}y$$

为 exact differential，称 $F$ 为它的势函数 potential function。

为什么这对经济学重要？因为如果某个局部变化可以写成 $\mathrm{d}F$，那么从状态 $A$ 到状态 $B$ 的总变化只等于

$$F(B) - F(A),$$

与走哪条路径无关。福利变化、效用变化、成本变化如果是 exact，就可以用端点差来度量；如果不是 exact，就必须说明路径。

## 8.1 二维 exactness 判别

**命题 5.1：二维 exact differential 的必要条件**

如果

$$M(x,y)\,\mathrm{d}x + N(x,y)\,\mathrm{d}y = \mathrm{d}F$$

且 $F$ 有连续二阶偏导，则

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}.$$

**证明**

由 exactness,

$$M = F_x, \qquad N = F_y.$$

所以

$$M_y = F_{xy}, \qquad N_x = F_{yx}.$$

若 $F$ 的二阶偏导连续，由 Clairaut 定理,

$$F_{xy} = F_{yx}.$$

因此

$$M_y = N_x.$$

在单连通区域上，如果 $M, N$ 一阶偏导连续，条件

$$M_y = N_x$$

也足以保证 exact。这里的“单连通”可以先理解为区域里没有洞；本章不展开拓扑细节。

**方法**

判别 $M\,\mathrm{d}x + N\,\mathrm{d}y$ 是否 exact 的实操步骤:

1. 先认出 $M$ 是 $\mathrm{d}x$ 前面的系数，$N$ 是 $\mathrm{d}y$ 前面的系数;
2. 计算 $M_y$ 和 $N_x$;
3. 如果不相等，则不是 exact;
4. 如果相等，并且区域没有洞，通常可以认为是 exact;
5. 要找势函数 $F$，先对 $M$ 关于 $x$ 积分，再用 $F_y = N$ 补出遗漏的 $y$ 函数。

**例 5.3：判断并找出势函数**

判断

$$(2xy+3)\,\mathrm{d}x + x^2\,\mathrm{d}y$$

是否 exact。
这里

$$M(x,y) = 2xy+3, \qquad N(x,y) = x^2.$$

计算

$$M_y = 2x, \qquad N_x = 2x.$$

两者相等，所以 exact。
找势函数。由 $F_x = 2xy+3$，对 $x$ 积分：

$$F(x,y) = x^2y + 3x + g(y).$$

再求 $F_y$：

$$F_y = x^2 + g'(y).$$

要求 $F_y = N = x^2$，所以

$$g'(y) = 0.$$

因此可以取

$$F(x,y) = x^2y + 3x.$$

**常见误区**

当你对 $M$ 关于 $x$ 积分时，常数不能写成普通常数，而要写成 $g(y)$。因为对 $x$ 求偏导时，任何只依赖 $y$ 的函数都会消失。这是 exact differential 题最常见的丢分点。

# 9 路径依赖和福利变化

设一个局部福利变化公式为

$$\mathrm{d}W = M(x,y)\,\mathrm{d}x + N(x,y)\,\mathrm{d}y.$$

如果它是 exact，那么从 $(x_0,y_0)$ 到 $(x_1,y_1)$ 的福利变化就是

$$W(x_1,y_1) - W(x_0,y_0),$$

与政策执行路径无关。

如果它不是 exact，那么沿不同路径积分会得到不同数值。经济含义是：同样的起点和终点，如果调整顺序不同，总福利变化可能不同。

**例 5.4：一个路径依赖的局部变化**

考虑

$$\omega = y\,\mathrm{d}x.$$

这里

$$M = y, \qquad N = 0.$$

所以

$$M_y = 1, \qquad N_x = 0.$$

不相等，因此 $\omega$ 不是 exact。

从 $(0,0)$ 到 $(1,1)$：

- 路径 A：先沿 $x$ 方向从 $(0,0)$ 到 $(1,0)$，再沿 $y$ 方向到 $(1,1)$。第一段 $y=0$，所以 $\int y\,\mathrm{d}x = 0$；第二段 $\mathrm{d}x = 0$，所以也是 0。总积分为 0。
- 路径 B：先沿 $y$ 方向从 $(0,0)$ 到 $(0,1)$，再沿 $x$ 方向到 $(1,1)$。第一段 $\mathrm{d}x = 0$，积分为 0；第二段 $y=1$，所以 $\int_0^1 1\,\mathrm{d}x = 1$。总积分为 1。

同样起点和终点，不同路径给出不同结果。这就是路径依赖。

# 10 经济金融中的四个核心应用

## 10.1 应用一：消费者理论中的边际效用向量

设 $u(\boldsymbol{x})$ 是效用函数。梯度

$$\nabla u(\boldsymbol{x}) = \begin{pmatrix} MU_1 \\ \vdots \\ MU_n \end{pmatrix}$$

是边际效用向量。

在内点最优下，消费者问题的一阶条件写成

$$\nabla u(\boldsymbol{x}^*) = \lambda \boldsymbol{p}.$$

含义是：效用上升最快方向与预算约束法向量平行；或者说每一单位货币带来的边际效用相等：

$$\frac{MU_i}{p_i} = \lambda, \qquad i = 1, \dots, n.$$

## 10.2 应用二：成本函数和 Shephard lemma 的预备

设成本函数

$$C(w_1, w_2, q)$$

取决于要素价格和产量。全微分为

$$\mathrm{d}C = C_{w_1}\,\mathrm{d}w_1 + C_{w_2}\,\mathrm{d}w_2 + C_q\,\mathrm{d}q.$$

后面学对偶理论时会看到：在成本最小化问题中，

$$C_{w_i} = z_i^*(\boldsymbol{w}, q),$$

即成本函数对要素价格的偏导等于条件要素需求。现在先记住几何语言：**成本函数的梯度记录了每个外生变量的边际成本效应**。

## 10.3 应用三：投资组合方差的梯度

设投资组合权重为 $\boldsymbol{w}$，资产协方差矩阵为 $\Sigma$。组合方差为

$$V(\boldsymbol{w}) = \boldsymbol{w}'\Sigma\boldsymbol{w}.$$

若 $\Sigma$ 对称，则

$$\nabla_{\boldsymbol{w}} V(\boldsymbol{w}) = 2\Sigma\boldsymbol{w}.$$

第 $i$ 个分量为

$$\frac{\partial V}{\partial w_i} = 2(\Sigma\boldsymbol{w})_i.$$

它衡量第 $i$ 个资产权重增加一点时，组合方差如何变化。

**推导：二次型梯度**

写开：

$$V(\boldsymbol{w}) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i \Sigma_{ij} w_j.$$

对 $w_k$ 求偏导：

$$\frac{\partial V}{\partial w_k} = \sum_{j=1}^{n} \Sigma_{kj} w_j + \sum_{i=1}^{n} w_i \Sigma_{ik}.$$

如果 $\Sigma$ 对称，即 $\Sigma_{ik} = \Sigma_{ki}$，第二项等于

$$\sum_{i=1}^{n} \Sigma_{ki} w_i.$$

所以

$$\frac{\partial V}{\partial w_k} = 2\sum_{j=1}^{n} \Sigma_{kj} w_j.$$

矩阵形式就是

$$\nabla V(\boldsymbol{w}) = 2\Sigma\boldsymbol{w}.$$

### 10.4 应用四：梯度下降与最小化损失

在计量、机器学习和数值优化中，经常要最小化损失函数

$$L(\boldsymbol{\theta}).$$

负梯度方向是局部下降最快方向，因此最基本的梯度下降更新为

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \nabla L(\boldsymbol{\theta}_t),$$

其中 $\eta > 0$ 是步长 learning rate。

这并不是一个神秘算法，而是方向导数结论的直接应用：沿 $-\nabla L$ 方向走，局部下降最快。

## 11 Worked Examples：完整样板题

**样板题 1：效用梯度、MRS 与无差异曲线斜率**

设

$$u(x,y) = x^{0.4}y^{0.6}, \quad x,y>0.$$

求 $\nabla u$、$\mathrm{MRS}_{xy}$，并解释无差异曲线斜率。

**解：**偏导为

$$u_x = 0.4x^{-0.6}y^{0.6}, \qquad u_y = 0.6x^{0.4}y^{-0.4}.$$

所以

$$\nabla u(x,y) = \begin{pmatrix} 0.4x^{-0.6}y^{0.6} \\ 0.6x^{0.4}y^{-0.4} \end{pmatrix}.$$

边际替代率为

$$\mathrm{MRS}_{xy} = \frac{u_x}{u_y} = \frac{0.4}{0.6}\frac{y}{x} = \frac{2y}{3x}.$$

沿无差异曲线，$\mathrm{d}u = 0$，所以

$$u_x\,\mathrm{d}x + u_y\,\mathrm{d}y = 0 \quad \Rightarrow \quad \frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{u_x}{u_y} = -\frac{2y}{3x}.$$

解释：如果增加一点 $x$，为了保持效用不变，需要减少约 $2y/(3x)$ 单位的 $y$。

**样板题 2：方向导数与最大上升方向**

设

$$f(x,y) = x^2y + 3y^2.$$

在点 $(1,2)$，求沿方向 $(3,4)$ 的方向导数，并求最大上升率。

**解：**先求梯度：

$$f_x = 2xy, \qquad f_y = x^2 + 6y.$$

在 $(1,2)$，

$$\nabla f(1,2) = (4, 13).$$

方向 $(3,4)$ 的单位向量为

$$\boldsymbol{u} = \left(\frac{3}{5}, \frac{4}{5}\right).$$

方向导数为

$$D_{\boldsymbol{u}} f(1,2) = \nabla f(1,2) \cdot \boldsymbol{u} = 4 \cdot \frac{3}{5} + 13 \cdot \frac{4}{5} = \frac{64}{5}.$$

最大上升率为

$$\|\nabla f(1,2)\| = \sqrt{4^2 + 13^2} = \sqrt{185}.$$

最大上升方向为

$$\frac{(4,13)}{\sqrt{185}}.$$

## 样板题 3：成本函数的全微分与近似变化

设企业成本函数为

$$C(q, w) = wq^2 + 10q.$$

当前 $(q,w) = (5,2)$。若产量增加 $\mathrm{d}q = 0.1$，工资下降 $\mathrm{d}w = -0.05$，近似成本变化是多少？

**解：**计算偏导：

$$C_q = 2wq + 10, \qquad C_w = q^2.$$

在 $(5,2)$，

$$C_q = 2 \cdot 2 \cdot 5 + 10 = 30, \qquad C_w = 25.$$

全微分：

$$\mathrm{d}C = C_q \, \mathrm{d}q + C_w \, \mathrm{d}w.$$

代入：

$$\mathrm{d}C \approx 30(0.1) + 25(-0.05) = 3 - 1.25 = 1.75.$$

所以成本大约增加 $1.75$。

## 样板题 4：判断 exact differential 并计算端点变化

判断

$$\omega = (3x^2y + 2)\,\mathrm{d}x + (x^3 + 4y)\,\mathrm{d}y$$

是否 exact。若 exact，求势函数，并计算从 $(0,0)$ 到 $(1,2)$ 的总变化。

**解：** 这里

$$M = 3x^2y + 2, \qquad N = x^3 + 4y.$$

计算

$$M_y = 3x^2, \qquad N_x = 3x^2.$$

相等，因此 exact。

由 $F_x = M$，对 $x$ 积分：

$$F(x,y) = x^3y + 2x + g(y).$$

求 $F_y$：

$$F_y = x^3 + g'(y).$$

要求等于 $N = x^3 + 4y$，故

$$g'(y) = 4y, \qquad g(y) = 2y^2 + C.$$

可取

$$F(x,y) = x^3y + 2x + 2y^2.$$

从 $(0,0)$ 到 $(1,2)$ 的总变化为

$$F(1,2) - F(0,0) = 1^3 \cdot 2 + 2 \cdot 1 + 2 \cdot 2^2 = 2 + 2 + 8 = 12.$$

## 样板题 5：投资组合方差梯度

设两个资产的协方差矩阵为

$$\Sigma = \begin{pmatrix} 0.04 & 0.01 \\ 0.01 & 0.09 \end{pmatrix},$$

投资组合权重为

$$\boldsymbol{w} = \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix}.$$

求组合方差梯度，并解释其含义。

**解：** 组合方差

$$V(\boldsymbol{w}) = \boldsymbol{w}'\Sigma\boldsymbol{w}.$$

梯度为

$$\nabla V = 2\Sigma\boldsymbol{w}.$$

先算

$$\Sigma\boldsymbol{w} = \begin{pmatrix} 0.04 & 0.01 \\ 0.01 & 0.09 \end{pmatrix} \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix} = \begin{pmatrix} 0.024 + 0.004 \\ 0.006 + 0.036 \end{pmatrix} = \begin{pmatrix} 0.028 \\ 0.042 \end{pmatrix}.$$

所以

$$\nabla V = \begin{pmatrix} 0.056 \\ 0.084 \end{pmatrix}.$$

解释：在当前组合附近，如果单独增加资产 1 的权重，方差的边际上升约为 $0.056$；如果单独增加资产 2 的权重，方差的边际上升约为 $0.084$。资产 2 的风险边际贡献更大。

# 12 常见误区整理

**常见误区**

1. **把梯度当成普通导数**。一元导数是一个数，多元梯度是一个向量；方向导数才是沿某个方向得到的数。
2. **忘记方向向量必须单位化**。方向导数公式 $D_{\boldsymbol{u}} f = \nabla f \cdot \boldsymbol{u}$ 中的 $\boldsymbol{u}$ 应该是单位向量。
3. **把全微分当成精确等式**。对有限变化，$\Delta f \approx \mathrm{d}f$ 通常只是近似；只有线性函数时才完全相等。
4. **混淆 $\mathrm{d}f = 0$ 和 $\nabla f = 0$**。沿等值线有 $\mathrm{d}f = 0$，但这只说明沿切向方向不变，不说明梯度为零。
5. **exactness 判别时漏掉区域条件**。$M_y = N_x$ 在普通无洞区域上足够好用，但有洞区域可能出现更细的拓扑问题。
6. **找势函数时漏掉 $g(y)$ 或 $h(x)$**。对某个变量积分后，积分“常数”可以是其他变量的函数。

# 13 本章总结图谱

| 概念 | 公式 | 经济金融含义 |
|---|---|---|
| 梯度 | $\nabla f = (f_{x_1}, \dots, f_{x_n})'$ | 所有变量的边际效应向量 |
| 方向导数 | $D_{\boldsymbol{u}} f = \nabla f \cdot \boldsymbol{u}$ | 沿某个政策/投资方向的边际变化 |
| 最陡上升 | $\max_{\|\boldsymbol{u}\|=1} D_{\boldsymbol{u}} f = \|\nabla f\|$ | 梯度方向是局部最快改善方向 |
| 全微分 | $\mathrm{d}f = \sum_i f_{x_i}\,\mathrm{d}x_i$ | 多个变量同时变化的总效应 |
| 等值线 | $f(x,y) = c$ | 无差异曲线、等成本线、等产量线 |
| 梯度与等值线 | $\nabla f \perp$ level set | 最优化几何条件的基础 |
| exact differential | $M\,\mathrm{d}x + N\,\mathrm{d}y = \mathrm{d}F$ | 路径无关的福利/成本/效用变化 |
| exact 判别 | $M_y = N_x$ | 局部边际变化能否整合成总量函数 |
| 投资组合方差梯度 | $\nabla_{\boldsymbol{w}}(\boldsymbol{w}'\Sigma\boldsymbol{w}) = 2\Sigma\boldsymbol{w}$ | 资产权重的风险边际贡献 |

# 14 练习题

## A 组：基础计算

A1. 求 $f(x,y)=x^3+2xy+y^2$ 的梯度，并计算 $\nabla f(1,2)$。

A2. 求 $u(x,y)=\ln x+2\ln y$ 的全微分，并求 $\mathrm{MRS}_{xy}$。

A3. 设 $f(x,y,z)=xy+yz+z^2$，求 $\nabla f$。

A4. 设 $f(x,y)=e^{xy}$。求 $\mathrm{d}f$。

A5. 设 $F(K,L)=K^{0.3}L^{0.7}$。对数微分，写出 $\mathrm{d}\ln F$。

## B 组：概念判断

B1. 判断：梯度是函数沿任意方向的导数。这个说法准确吗？说明理由。

B2. 判断：如果沿一条无差异曲线 $\mathrm{d}u=0$，那么 $\nabla u=0$。这个说法对吗？

B3. 判断：方向导数中的方向向量不需要单位化。这个说法对吗？

B4. 判断：如果 $M_y\neq N_x$，则 $M\,\mathrm{d}x+N\,\mathrm{d}y$ 不可能是 exact differential。这个说法在通常光滑条件下对吗？

B5. 用一句话解释为什么梯度垂直于等值线。

## C 组：混合题

C1. 设 $f(x,y)=x^2+4xy+y^2$。在点 $(1,1)$，求沿方向 $(1,-1)$ 的方向导数。

C2. 判断 $(2xy+y^2)\,\mathrm{d}x+(x^2+2xy)\,\mathrm{d}y$ 是否 exact。若是，求势函数。

C3. 设 $q=D(p,y)=y/p$。求 $\mathrm{d}q$，并解释价格和收入同时变化对需求的局部影响。

C4. 设 $C(q,w)=w^{1/2}q^2$。在 $(q,w)=(3,4)$ 处，若 $\mathrm{d}q=0.05$、$\mathrm{d}w=0.2$，近似 $\mathrm{d}C$。

C5. 设 $L(\theta_1,\theta_2)=(\theta_1-1)^2+4(\theta_2+2)^2$。求负梯度方向，并解释梯度下降的下一步方向。

## D 组：经济金融应用题

D1. 消费者效用为 $u(x,y)=x^{1/3}y^{2/3}$，价格为 $(p_x,p_y)=(2,3)$。写出内点最优的一阶条件 $\nabla u=\lambda\boldsymbol{p}$，并推出 $\mathrm{MRS}_{xy}=p_x/p_y$。

D2. 设生产函数 $Y=AK^{\alpha}L^{1-\alpha}$。若 $A$ 增长 2%，$K$ 增长 5%，$L$ 增长 1%，$\alpha=0.4$，近似产出增长率是多少？

D3. 两资产协方差矩阵为

$$\Sigma=\begin{pmatrix}0.09 & 0.03\\ 0.03 & 0.04\end{pmatrix},\qquad \boldsymbol{w}=(0.5,0.5)'.$$

求 $\nabla_{\boldsymbol{w}}(\boldsymbol{w}'\Sigma\boldsymbol{w})$，并解释风险边际贡献。

D4. 一个局部福利变化为 $\mathrm{d}W = (2xy)\,\mathrm{d}x + x^2\,\mathrm{d}y$。判断它是否路径无关；若是，求 $W$。

D5. 设损失函数 $L(\beta) = \frac{1}{2}\|y - X\beta\|^2$。证明其梯度为 $\nabla_\beta L = -X'(y - X\beta)$。

## E 组：挑战/证明题

E1. 证明：若 $f$ 可微，且 $\nabla f(\boldsymbol{x}) = \mathbf{0}$，则 $f$ 在 $\boldsymbol{x}$ 沿任意方向的方向导数都为 0。反过来，如果沿任意方向的方向导数都为 0，是否一定有 $\nabla f(\boldsymbol{x}) = 0$?

E2. 设 $u(x, y)$ 是严格递增效用函数。解释为什么在预算约束内点最优处，$\nabla u$ 必须与价格向量平行。

E3. 设 $\omega = M\,\mathrm{d}x + N\,\mathrm{d}y$。在光滑且单连通区域内，若 $M_y = N_x$，说明如何构造势函数 $F$。

E4. 给出一个 $\mathrm{d}W = M\,\mathrm{d}x + N\,\mathrm{d}y$ 的例子，使得从 $(0,0)$ 到 $(1,1)$ 的积分依赖路径，并计算两条路径的不同结果。

E5. 设 $V(\boldsymbol{w}) = \boldsymbol{w}'\Sigma\boldsymbol{w}$，其中 $\Sigma$ 不一定对称。证明 $\nabla V = (\Sigma + \Sigma')\boldsymbol{w}$。为什么在金融中通常写成 $2\Sigma\boldsymbol{w}$?

# 15 练习题详细解答

## A 组答案

### A1

$$f_x = 3x^2 + 2y, \qquad f_y = 2x + 2y.$$

所以

$$\nabla f(x,y) = \begin{pmatrix} 3x^2 + 2y \\ 2x + 2y \end{pmatrix}.$$

在 $(1,2)$,

$$\nabla f(1,2) = \begin{pmatrix} 3+4 \\ 2+4 \end{pmatrix} = \begin{pmatrix} 7 \\ 6 \end{pmatrix}.$$

### A2

$$u_x = \frac{1}{x}, \qquad u_y = \frac{2}{y}.$$

全微分为

$$\mathrm{d}u = \frac{1}{x}\,\mathrm{d}x + \frac{2}{y}\,\mathrm{d}y.$$

边际替代率

$$\mathrm{MRS}_{xy} = \frac{u_x}{u_y} = \frac{1/x}{2/y} = \frac{y}{2x}.$$

### A3

$$f_x = y, \qquad f_y = x + z, \qquad f_z = y + 2z.$$

因此

$$\nabla f = (y, x+z, y+2z)'.$$

### A4

$$f_x = ye^{xy}, \qquad f_y = xe^{xy}.$$

所以

$$\mathrm{d}f = ye^{xy}\,\mathrm{d}x + xe^{xy}\,\mathrm{d}y.$$

**A5**

$$\ln F = 0.3 \ln K + 0.7 \ln L.$$

因此

$$\mathrm{d} \ln F = 0.3 \, \mathrm{d} \ln K + 0.7 \, \mathrm{d} \ln L.$$

含义是资本增长率和劳动增长率按产出弹性加权。

## B 组答案

**B1**

不准确。梯度是一个向量，它本身不是沿任意方向的导数。沿单位方向 $\boldsymbol{u}$ 的方向导数是

$$D_{\boldsymbol{u}} f = \nabla f \cdot \boldsymbol{u}.$$

梯度通过内积生成各个方向上的导数。

**B2**

不对。沿无差异曲线 $\mathrm{d}u = 0$ 只说明沿该曲线的切向小变化不改变效用。梯度通常不为零，而是垂直于无差异曲线。只有在驻点处才可能 $\nabla u = 0$。

**B3**

不对。方向导数通常定义为沿单位方向的单位长度变化率。如果用非单位向量，计算出来会混入方向向量长度，不再是“每单位距离”的变化率。

**B4**

对。在通常光滑条件下，如果 $M \, \mathrm{d}x + N \, \mathrm{d}y = \mathrm{d}F$，则一定有 $M_y = N_x$。所以若二者不相等，就不可能 exact。

**B5**

沿等值线有 $f(\boldsymbol{r}(t)) = c$。对 $t$ 求导得到

$$\nabla f(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t) = 0.$$

所以梯度与等值线切向量垂直。

# C 组答案

## C1

$$\nabla f = (2x + 4y, 4x + 2y).$$

在 $(1, 1)$，

$$\nabla f(1, 1) = (6, 6).$$

方向 $(1, -1)$ 的单位向量为

$$\boldsymbol{u} = \frac{1}{\sqrt{2}}(1, -1).$$

方向导数：

$$D_{\boldsymbol{u}} f = (6, 6) \cdot \frac{1}{\sqrt{2}}(1, -1) = 0.$$

这说明该方向在此点是等值线切向方向。

## C2

$$M = 2xy + y^2, \qquad N = x^2 + 2xy.$$

$$M_y = 2x + 2y, \qquad N_x = 2x + 2y.$$

所以 exact。由 $F_x = M$，对 $x$ 积分：

$$F = x^2 y + xy^2 + g(y).$$

于是

$$F_y = x^2 + 2xy + g'(y).$$

令其等于 $N = x^2 + 2xy$，得 $g'(y) = 0$。可取

$$F = x^2 y + xy^2.$$

## C3

$$q = \frac{y}{p}.$$

$$q_p = -\frac{y}{p^2}, \qquad q_y = \frac{1}{p}.$$

所以

$$\mathrm{d}q = -\frac{y}{p^2}\,\mathrm{d}p + \frac{1}{p}\,\mathrm{d}y.$$

价格上升降低需求，收入上升提高需求；同时变化时总效应是二者的线性加总。

**C4**

$$C(q,w)=w^{1/2}q^2.$$

$$C_q=2qw^{1/2},\qquad C_w=\frac{1}{2}w^{-1/2}q^2.$$

在 $(3,4)$：

$$C_q=2\cdot 3\cdot 2=12,\qquad C_w=\frac{1}{2}\cdot\frac{1}{2}\cdot 9=\frac{9}{4}=2.25.$$

因此

$$\mathrm{d}C\approx 12(0.05)+2.25(0.2)=0.6+0.45=1.05.$$

**C5**

$$\nabla L=(2(\theta_1-1),8(\theta_2+2)).$$

负梯度方向为

$$-\nabla L=(-2(\theta_1-1),-8(\theta_2+2)).$$

梯度下降会沿这个方向调整参数，使 $\theta_1$ 向 1 靠近，使 $\theta_2$ 向 -2 靠近。

## D 组答案

**D1**

$$u_x=\frac{1}{3}x^{-2/3}y^{2/3},\qquad u_y=\frac{2}{3}x^{1/3}y^{-1/3}.$$

内点最优条件：

$$\begin{pmatrix}u_x\\u_y\end{pmatrix}=\lambda\begin{pmatrix}2\\3\end{pmatrix}.$$

因此

$$\frac{u_x}{u_y}=\frac{2}{3}.$$

而

$$\frac{u_x}{u_y}=\frac{1}{2}\frac{y}{x}.$$

所以最优时

$$\mathrm{MRS}_{xy}=\frac{p_x}{p_y}=\frac{2}{3}.$$

**D2**

由

$$Y = AK^{\alpha}L^{1-\alpha}$$

得

$$\mathrm{d}\ln Y = \mathrm{d}\ln A + \alpha\,\mathrm{d}\ln K + (1-\alpha)\,\mathrm{d}\ln L.$$

代入：

$$\mathrm{d}\ln Y \approx 2\% + 0.4\times 5\% + 0.6\times 1\% = 2\% + 2\% + 0.6\% = 4.6\%.$$

产出近似增长 $4.6\%$。

**D3**

$$\Sigma\boldsymbol{w} = \begin{pmatrix} 0.09 & 0.03 \\ 0.03 & 0.04 \end{pmatrix}\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 0.06 \\ 0.035 \end{pmatrix}.$$

所以

$$\nabla V = 2\Sigma\boldsymbol{w} = \begin{pmatrix} 0.12 \\ 0.07 \end{pmatrix}.$$

资产 1 的边际风险贡献高于资产 2；在当前权重附近，增加资产 1 权重对方差的边际增加更大。

**D4**

$$M = 2xy, \qquad N = x^2.$$

$$M_y = 2x, \qquad N_x = 2x.$$

所以 exact。由 $W_x = 2xy$，对 $x$ 积分：

$$W = x^2y + g(y).$$

求 $W_y = x^2 + g'(y)$。令其等于 $x^2$，得 $g'(y) = 0$。故可取

$$W = x^2y.$$

福利变化路径无关。

**D5**

$$L(\beta) = \frac{1}{2}(y - X\beta)'(y - X\beta).$$

展开:

$$L(\beta) = \frac{1}{2}\left(y'y - 2\beta'X'y + \beta'X'X\beta\right).$$

对 $\beta$ 求梯度:

$$\nabla_\beta L = -X'y + X'X\beta.$$

整理:

$$\nabla_\beta L = -X'(y - X\beta).$$

这就是 OLS 一阶条件的核心：令梯度为零得到

$$X'(y - X\beta) = 0.$$

## E 组答案

**E1**

若 $f$ 可微，则沿单位方向 $\boldsymbol{u}$ 的方向导数为

$$D_{\boldsymbol{u}} f(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) \cdot \boldsymbol{u}.$$

如果 $\nabla f(\boldsymbol{x}) = \mathbf{0}$，则对任意 $\boldsymbol{u}$，

$$D_{\boldsymbol{u}} f(\boldsymbol{x}) = 0.$$

反过来，如果沿任意方向方向导数都为 0，特别地沿坐标方向 $\boldsymbol{e}_i$ 的方向导数为 0，所以所有偏导数都为 0。因此在可微且方向导数按梯度表示的条件下，$\nabla f(\boldsymbol{x}) = 0$。

**E2**

预算约束为

$$\boldsymbol{p} \cdot \boldsymbol{x} = w.$$

它的切向方向 $\boldsymbol{v}$ 满足

$$\boldsymbol{p} \cdot \boldsymbol{v} = 0.$$

如果在最优点存在一个可行切向方向 $\boldsymbol{v}$，使得

$$\nabla u(\boldsymbol{x}^*) \cdot \boldsymbol{v} > 0,$$

那么消费者沿预算线小幅移动就能提高效用，矛盾。因此对所有满足 $\boldsymbol{p} \cdot \boldsymbol{v} = 0$ 的切向方向，都必须有

$$\nabla u(\boldsymbol{x}^*) \cdot \boldsymbol{v} = 0.$$

这说明 $\nabla u(\boldsymbol{x}^*)$ 垂直于预算线切空间，因此与预算线法向量 $\boldsymbol{p}$ 平行：

$$\nabla u(\boldsymbol{x}^*) = \lambda \boldsymbol{p}.$$

**E3**

构造方法如下。给定

$$\omega = M(x,y)\,\mathrm{d}x + N(x,y)\,\mathrm{d}y,$$

若 $M_y = N_x$，先令

$$F(x,y) = \int_{x_0}^{x} M(s,y)\,\mathrm{d}s + g(y).$$

这样自动有 $F_x = M$。再计算

$$F_y = \int_{x_0}^{x} M_y(s,y)\,\mathrm{d}s + g'(y).$$

利用 $M_y = N_x$，上式可写为

$$F_y = \int_{x_0}^{x} N_x(s,y)\,\mathrm{d}s + g'(y) = N(x,y) - N(x_0,y) + g'(y).$$

选择

$$g'(y) = N(x_0,y)$$

即可使 $F_y = N(x,y)$。因此构造出势函数。

**E4**

例子：

$$\mathrm{d}W = y\,\mathrm{d}x.$$

从 $(0,0)$ 到 $(1,1)$。路径 A：先横后竖。横向时 $y=0$，积分为 0；竖向时 $\mathrm{d}x = 0$，积分为 0。总积分为 0。
路径 B：先竖后横。竖向时 $\mathrm{d}x = 0$，积分为 0；横向时 $y = 1$，积分为

$$\int_0^1 1\,\mathrm{d}x = 1.$$

所以同样起点终点，路径 A 得 0，路径 B 得 1，说明路径依赖。

**E5**

写开：

$$V(\boldsymbol{w}) = \boldsymbol{w}'\Sigma\boldsymbol{w}.$$

对微小变化 $\mathrm{d}\boldsymbol{w}$：

$$\mathrm{d}V = (\mathrm{d}\boldsymbol{w})'\Sigma\boldsymbol{w} + \boldsymbol{w}'\Sigma(\mathrm{d}\boldsymbol{w}).$$

第一项可写为

$$(\mathrm{d}\boldsymbol{w})'\Sigma\boldsymbol{w}.$$

第二项是标量，等于其转置：

$$\boldsymbol{w}'\Sigma\,\mathrm{d}\boldsymbol{w} = ((\boldsymbol{w}'\Sigma\,\mathrm{d}\boldsymbol{w})') = (\mathrm{d}\boldsymbol{w})'\Sigma'\boldsymbol{w}.$$

所以

$$\mathrm{d}V = (\mathrm{d}\boldsymbol{w})'(\Sigma + \Sigma')\boldsymbol{w}.$$

因此

$$\nabla V = (\Sigma + \Sigma')\boldsymbol{w}.$$

金融中的协方差矩阵满足 $\Sigma = \Sigma'$，所以

$$\nabla V = 2\Sigma\boldsymbol{w}.$$

## 本章结束语

本章真正要你带走的不是几条公式，而是一套新的语言：

$$\text{局部变化} \longrightarrow \text{梯度} \longrightarrow \text{全微分} \longrightarrow \text{路径与总量}.$$

以后在经济金融模型里看到 $u(x)$、$C(q,w)$、$V(w)=w'\Sigma w$、$L(\theta)$、$F(x,\theta)=0$，你都应该立刻问：

- 它的梯度是什么？
- 哪些变量同时变化？
- 全微分如何写？
- 是否存在路径无关的势函数？
- 这个局部变化能不能用于比较静态或最优化？

下一章进入曲线积分与路径依赖。你会看到本章的 exact differential 不只是一个判别题，而是曲线积分是否只取决于端点的核心条件。

# 经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

---

# Chapter 6

## 曲线积分与路径依赖

---

本章的目标不是把曲线积分学成物理场论技巧，而是理解一个经济金融问题：当局部边际变化沿着一条路径不断累积时，最终结果是否只由起点和终点决定？如果不是，模型就存在路径依赖。

math4econ personalized notes

2026

# 目录

**Chapter 6 总览** **2**

**1 本章为什么重要** **4**

**2 最小前置知识** **4**
2.1 参数曲线 . . . . . 5
2.2 梯度与全微分 . . . . . 5
2.3 链式法则 . . . . . 5

**3 第一类曲线积分：沿路径累加密度** **5**
3.1 定义与直觉 . . . . . 6
3.2 为什么要乘长度因子 . . . . . 6

**4 第二类曲线积分：沿变化方向累加边际场** **7**
4.1 定义 . . . . . 7
4.2 经济学解释：局部边际变化的累积 . . . . . 8
4.3 Riemann 和推导 . . . . . 8

**5 路径依赖：同样端点，不同结果** **8**
5.1 核心例子 . . . . . 9
5.2 经济解释 . . . . . 10

**6 保守场、势函数与 exact differential** **10**
6.1 线积分基本定理 . . . . . 11

**7 怎样判断一个场是不是保守场** **12**
7.1 二维局部判别条件 . . . . . 12
7.2 找势函数的实际步骤 . . . . . 12

**8 Green 定理：边界循环与内部旋转** **13**
8.1 为什么 Green 定理和路径依赖有关 . . . . . 14
8.2 一个重要反例：区域有洞时要小心 . . . . . 15

**9 经济金融应用** **15**
9.1 应用一：福利变化是否可以由终点差值表示 . . . . . 15
9.2 应用二：从边际效用到效用函数 . . . . . 16
9.3 应用三：调整成本与路径依赖 . . . . . 17
9.4 应用四：投资组合再平衡的路径成本 . . . . . 17

**10 Worked Examples：完整样板题** **17**

**11 常见错误与纠正** **20**

**12 练习题** **21**

A 组：基础计算 . . . . . 21
B 组：概念判断 . . . . . 21
C 组：混合题 . . . . . 21
D 组：经济金融应用题 . . . . . 21
E 组：挑战与证明题 . . . . . 22

**13 练习题详细答案** **22**

A 组答案 . . . . . 22
B 组答案 . . . . . 24
C 组答案 . . . . . 25
D 组答案 . . . . . 27
E 组答案 . . . . . 29

**14 本章总结图谱** **31**

**下一章连接** **31**

# Chapter 6 总览

**本章目标**

学完本章，你应该达到以下目标：

1. 能把平面或空间中的路径写成参数形式 $\boldsymbol{r}(t)$，并理解速度 $\|\boldsymbol{r}'(t)\|$ 与弧长元素 $\mathrm{d}s$ 的关系；
2. 掌握第一类曲线积分

$$\int_C f \,\mathrm{d}s = \int_a^b f(\boldsymbol{r}(t))\,\|\boldsymbol{r}'(t)\|\,\mathrm{d}t,$$

并能解释它是沿路径累加密度、成本、风险暴露或调整强度；
3. 掌握第二类曲线积分

$$\int_C \boldsymbol{F}\cdot \mathrm{d}\boldsymbol{r} = \int_C P\,\mathrm{d}x + Q\,\mathrm{d}y,$$

并能解释它是把局部边际场沿变化方向累积；
4. 理解方向、参数化、路径和端点的区别，知道为什么同样的起点与终点可能产生不同积分；
5. 掌握保守场 conservative field、势函数 potential function、exact differential 和路径无关 path independence 的等价关系；
6. 会用二维判别式 $P_y = Q_x$ 判断局部 exactness，并知道它需要单连通区域这一条件；
7. 理解 Green 定理的直观含义：边界上的总循环等于区域内部局部旋转的总和；
8. 能把本章方法用于福利变化、政策调整路径、revealed preference 的可积性直觉、交易成本和动态调整成本。

**直觉**

一句话：**普通积分是在一条数轴上累加；曲线积分是在一条路径上累加。**
一元积分问：从 $a$ 到 $b$ 沿直线走，累积多少？曲线积分问：如果从 $A$ 到 $B$ 可以走很多条路，那么沿某一条具体路径走，局部成本、局部福利、局部风险暴露、局部边际收益加起来是多少？

**本章压缩总结**

第 5 章讲了梯度与全微分。本章接着问：如果一个局部微分形式

$$P(x,y)\,\mathrm{d}x + Q(x,y)\,\mathrm{d}y$$

沿路径累加，得到的总变化是否等于某个全局函数的差值？如果是，它像效用函数、价值函数、势函数；如果不是，它体现路径依赖、制度摩擦、调整成本或不可积性。

# 1 本章为什么重要

在经济学里，我们经常把一个复杂变化拆成许多小步，然后把每一步的边际影响加起来。例如：

- 消费束从 $(x_1, x_2)$ 调整到 $(x_1', x_2')$，效用改变多少？
- 税率和补贴率同时变化，社会福利的总变化是否只取决于最终政策？
- 企业从一种生产技术逐步转向另一种技术，总调整成本是否取决于调整顺序？
- 投资组合从旧权重调到新权重，交易成本是否取决于再平衡路径？
- 如果我们只观察到局部边际替代率，是否能恢复一个全局效用函数？

  这些问题背后都有同一个数学结构：沿路径累加。

**本章的核心问题**

给定一个局部边际场

$$\boldsymbol{F}(x, y) = (P(x, y), Q(x, y)),$$

沿路径 $C$ 从 $A$ 到 $B$ 的累积量是

$$\int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \int_C P(x, y)\,\mathrm{d}x + Q(x, y)\,\mathrm{d}y.$$

核心问题是：这个积分是否只依赖起点 $A$ 和终点 $B$，还是依赖中间走了哪条路径？

如果只依赖起点和终点，那么存在某个函数 $\phi$，使得

$$\boldsymbol{F} = \nabla \phi,$$

于是

$$\int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \phi(B) - \phi(A).$$

这时 $\phi$ 就像一个全局福利函数、效用函数、势函数或价值函数。反过来，如果积分依赖路径，那么局部边际量不能简单整合成一个全局对象。

**常见误区**

不要把“曲线积分”理解成只服务物理学的概念。经济学里大量问题虽然不写成 $\int_C P\,\mathrm{d}x + Q\,\mathrm{d}y$，但本质上就是在问：**局部边际变化能否被积成全局函数？路径是否重要？**

# 2 最小前置知识

本章需要你回忆三件事。

## 2.1 参数曲线

一条平面曲线可以写成

$$\boldsymbol{r}(t) = (x(t), y(t)), \qquad t \in [a, b].$$

这里 $t$ 不一定是时间。它只是一个参数，表示你沿着路径前进的位置。

速度向量是

$$\boldsymbol{r}'(t) = (x'(t), y'(t)).$$

单位时间走过的长度是

$$\|\boldsymbol{r}'(t)\| = \sqrt{(x'(t))^2 + (y'(t))^2}.$$

所以弧长元素是

$$\mathrm{d}s = \|\boldsymbol{r}'(t)\|\,\mathrm{d}t.$$

## 2.2 梯度与全微分

若 $\phi(x, y)$ 可微，则

$$\nabla\phi = (\phi_x, \phi_y),$$

全微分为

$$\mathrm{d}\phi = \phi_x\,\mathrm{d}x + \phi_y\,\mathrm{d}y.$$

这说明如果 $\boldsymbol{F} = \nabla\phi$，那么

$$\boldsymbol{F}\cdot\,\mathrm{d}\boldsymbol{r} = \phi_x\,\mathrm{d}x + \phi_y\,\mathrm{d}y = \,\mathrm{d}\phi.$$

这就是 exact differential 的基本来源。

## 2.3 链式法则

如果沿路径 $\boldsymbol{r}(t) = (x(t), y(t))$ 观察函数 $\phi(x, y)$，那么

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi(x(t), y(t)) = \phi_x(x(t), y(t))x'(t) + \phi_y(x(t), y(t))y'(t).$$

向量写法是

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi(\boldsymbol{r}(t)) = \nabla\phi(\boldsymbol{r}(t))\cdot\boldsymbol{r}'(t).$$

这个公式会直接推出线积分基本定理。

# 3 第一类曲线积分：沿路径累加密度

## 3.1 定义与直觉

**第一类曲线积分**

设 $C$ 是由 $\boldsymbol{r}(t)=(x(t),y(t))$, $t\in[a,b]$ 参数化的光滑曲线。若 $f$ 是定义在曲线附近的标量函数，则第一类曲线积分定义为

$$\int_C f(x,y)\,\mathrm{d}s=\int_a^b f(x(t),y(t))\,\|\boldsymbol{r}'(t)\|\,\mathrm{d}t.$$

这里 $f(x,y)$ 是路径上每一点的“单位长度密度”。乘上 $\mathrm{d}s$ 后，就是那一小段路径贡献的累积量。

**直觉**

把一条路径想象成一条路。$f(x,y)$ 是这条路上每单位长度的成本、污染暴露、风险强度、信息采集成本或调整摩擦。沿整条路加起来，就是总成本。

## 3.2 为什么要乘长度因子

这是初学最容易忽略的地方。参数 $t$ 本身不是长度。如果你用不同速度走同一条曲线，$\mathrm{d}t$ 的大小没有几何意义；真正的一小段弧长是

$$\mathrm{d}s\approx\sqrt{(\Delta x)^2+(\Delta y)^2}.$$

当 $\Delta t$ 很小时，

$$\Delta x\approx x'(t)\Delta t,\qquad \Delta y\approx y'(t)\Delta t,$$

所以

$$\Delta s\approx\sqrt{(x'(t))^2+(y'(t))^2}\,\Delta t.$$

这就是

$$\mathrm{d}s=\|\boldsymbol{r}'(t)\|\,\mathrm{d}t.$$

**方法**

第一类曲线积分计算步骤：

1. 写出路径参数化 $\boldsymbol{r}(t)=(x(t),y(t))$；
2. 计算速度 $\boldsymbol{r}'(t)$ 和长度因子 $\|\boldsymbol{r}'(t)\|$；
3. 把 $f(x,y)$ 代成 $f(x(t),y(t))$；
4. 计算普通一元积分

$$\int_a^b f(\boldsymbol{r}(t))\,\|\boldsymbol{r}'(t)\|\,\mathrm{d}t.$$

**例 6.1：沿线段累加调整成本**

一项政策从状态 $A=(0,0)$ 调整到状态 $B=(1,2)$。设状态空间中单位距离调整成本为

$$f(x,y)=1+x+y.$$

若政策沿直线移动，计算总调整成本

$$\int_C f\,\mathrm{d}s.$$

**解：**直线可参数化为

$$\boldsymbol{r}(t)=(t,2t),\qquad 0\le t\le 1.$$

于是

$$\boldsymbol{r}'(t)=(1,2),\qquad \|\boldsymbol{r}'(t)\|=\sqrt{5}.$$

并且

$$f(\boldsymbol{r}(t))=1+t+2t=1+3t.$$

因此

$$\int_C f\,\mathrm{d}s=\int_0^1(1+3t)\sqrt{5}\,\mathrm{d}t=\sqrt{5}\left[t+\frac{3}{2}t^2\right]_0^1=\frac{5\sqrt{5}}{2}.$$

# 4 第二类曲线积分：沿变化方向累加边际场

第一类曲线积分累加的是“单位长度密度”。第二类曲线积分累加的是“向量场在前进方向上的分量”。

## 4.1 定义

**第二类曲线积分**

设向量场

$$\boldsymbol{F}(x,y)=(P(x,y),Q(x,y)),$$

曲线 $C$ 由 $\boldsymbol{r}(t)=(x(t),y(t))$，$t\in[a,b]$ 参数化。第二类曲线积分定义为

$$\int_C \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}=\int_a^b \boldsymbol{F}(\boldsymbol{r}(t))\cdot\boldsymbol{r}'(t)\,\mathrm{d}t.$$

在平面中也常写成

$$\int_C P\,\mathrm{d}x+Q\,\mathrm{d}y=\int_a^b\left[P(x(t),y(t))x'(t)+Q(x(t),y(t))y'(t)\right]\,\mathrm{d}t.$$

## 4.2 经济学解释：局部边际变化的累积

如果 $P(x,y)$ 是状态 $x$ 的边际福利，$Q(x,y)$ 是状态 $y$ 的边际福利，那么当状态从 $(x,y)$ 变动一个小量 $(\mathrm{d}x,\,\mathrm{d}y)$ 时，局部福利变化近似为

$$P(x,y)\,\mathrm{d}x + Q(x,y)\,\mathrm{d}y.$$

沿整条路径累加，就是

$$\int_C P\,\mathrm{d}x + Q\,\mathrm{d}y.$$

这里最关键的是：$\mathrm{d}x$ 和 $\mathrm{d}y$ 带有方向。如果你反向走同一条路径，第二类曲线积分会变号；第一类曲线积分一般不会。

**常见误区**

第一类曲线积分 $\int_C f\,\mathrm{d}s$ 对路径方向不敏感，因为 $\mathrm{d}s$ 是长度，总是非负。第二类曲线积分 $\int_C \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}$ 对方向敏感，因为 $\mathrm{d}\boldsymbol{r}$ 是有方向的小位移。

## 4.3 Riemann 和推导

把路径切成很多小段。第 $i$ 小段的位移为

$$\Delta\boldsymbol{r}_i = (\Delta x_i, \Delta y_i).$$

在这一小段上，向量场近似为

$$\boldsymbol{F}(\boldsymbol{r}_i) = (P(\boldsymbol{r}_i), Q(\boldsymbol{r}_i)).$$

局部贡献是点积

$$\boldsymbol{F}(\boldsymbol{r}_i)\cdot\Delta\boldsymbol{r}_i = P(\boldsymbol{r}_i)\Delta x_i + Q(\boldsymbol{r}_i)\Delta y_i.$$

累加并取极限：

$$\sum_i \boldsymbol{F}(\boldsymbol{r}_i)\cdot\Delta\boldsymbol{r}_i \longrightarrow \int_C \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}.$$

如果用参数 $t$，那么

$$\Delta\boldsymbol{r}_i \approx \boldsymbol{r}'(t_i)\Delta t_i,$$

所以

$$\sum_i \boldsymbol{F}(\boldsymbol{r}_i)\cdot\Delta\boldsymbol{r}_i \approx \sum_i \boldsymbol{F}(\boldsymbol{r}(t_i))\cdot\boldsymbol{r}'(t_i)\Delta t_i,$$

极限就是

$$\int_a^b \boldsymbol{F}(\boldsymbol{r}(t))\cdot\boldsymbol{r}'(t)\,\mathrm{d}t.$$

# 5 路径依赖：同样端点，不同结果

这是本章最重要的概念。

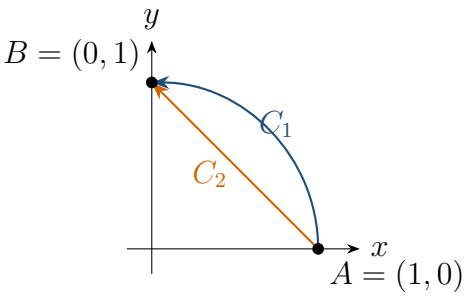
## 5.1 核心例子

考虑向量场

$$\boldsymbol{F}(x,y) = (-y, x).$$

从 $A = (1,0)$ 到 $B = (0,1)$，我们比较两条路径：

- 路径 $C_1$：单位圆第一象限圆弧；
- 路径 $C_2$：连接 $(1,0)$ 和 $(0,1)$ 的直线段。




**路径 $C_1$：圆弧**

参数化为

$$\boldsymbol{r}(t) = (\cos t, \sin t), \qquad 0 \le t \le \frac{\pi}{2}.$$

于是

$$\boldsymbol{r}'(t) = (-\sin t, \cos t),$$

并且

$$\boldsymbol{F}(\boldsymbol{r}(t)) = (-\sin t, \cos t).$$

所以

$$\boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t) = (-\sin t, \cos t) \cdot (-\sin t, \cos t) = 1.$$

因此

$$\int_{C_1} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \int_0^{\pi/2} 1 \, \mathrm{d}t = \frac{\pi}{2}.$$

**路径 $C_2$：直线**

参数化为

$$\boldsymbol{r}(t) = (1 - t, t), \qquad 0 \le t \le 1.$$

于是

$$\boldsymbol{r}'(t) = (-1, 1),$$

并且

$$\boldsymbol{F}(\boldsymbol{r}(t)) = (-t, 1 - t).$$

因此

$$\boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t) = (-t, 1 - t) \cdot (-1, 1) = t + 1 - t = 1.$$

所以

$$\int_{C_2} \boldsymbol{F}\cdot \mathrm{d}\boldsymbol{r} = \int_0^1 1\,\mathrm{d}t = 1.$$

**直觉**

同样从 $(1,0)$ 到 $(0,1)$，沿圆弧得到 $\pi/2$，沿直线得到 $1$。所以这个积分依赖路径。也就是说，$\boldsymbol{F}=(-y,x)$ 不能被看作某个全局函数 $\phi$ 的梯度。

## 5.2 经济解释

如果 $\boldsymbol{F}=(P,Q)$ 是一个政策空间中的局部福利边际场，那么路径依赖意味着：

同样的最终政策组合，不同改革顺序会产生不同累计福利。

例如先改革税制再改革补贴，与先改革补贴再改革税制，虽然最终政策相同，但过渡期成本、预期调整和行为反应可能不同。这类现象不能用一个简单的“终点福利函数差值”完全概括。

# 6 保守场、势函数与 exact differential

路径依赖的反面是路径无关。路径无关是经济学里非常重要的结构，因为它意味着局部边际量可以整合成一个全局目标函数。

**保守场与势函数**

若存在一个可微函数 $\phi(x,y)$，使得

$$\boldsymbol{F}(x,y)=\nabla\phi(x,y)=(\phi_x(x,y),\phi_y(x,y)),$$

则称 $\boldsymbol{F}$ 是保守场 conservative field，称 $\phi$ 是 $\boldsymbol{F}$ 的势函数 potential function。

**exact differential**

如果一个微分形式

$$P(x,y)\,\mathrm{d}x+Q(x,y)\,\mathrm{d}y$$

可以写成某个函数的全微分

$$\mathrm{d}\phi=\phi_x\,\mathrm{d}x+\phi_y\,\mathrm{d}y,$$

也就是

$$P=\phi_x,\qquad Q=\phi_y,$$

则称它是 exact differential。

## 6.1 线积分基本定理

**定理 6.1：线积分基本定理**

若 $\boldsymbol{F}=\nabla\phi$，且路径 $C$ 从 $A$ 到 $B$，则

$$\int_C \boldsymbol{F}\cdot \mathrm{d}\boldsymbol{r}=\phi(B)-\phi(A).$$

**证明：**设 $C$ 的参数化为 $\boldsymbol{r}(t)$, $a\le t\le b$，其中 $\boldsymbol{r}(a)=A$, $\boldsymbol{r}(b)=B$。因为 $\boldsymbol{F}=\nabla\phi$,

$$\int_C \boldsymbol{F}\cdot \mathrm{d}\boldsymbol{r}=\int_a^b \nabla\phi(\boldsymbol{r}(t))\cdot \boldsymbol{r}'(t)\,\mathrm{d}t.$$

由链式法则,

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi(\boldsymbol{r}(t))=\nabla\phi(\boldsymbol{r}(t))\cdot\boldsymbol{r}'(t).$$

所以

$$\int_C \boldsymbol{F}\cdot \mathrm{d}\boldsymbol{r}=\int_a^b \frac{\mathrm{d}}{\mathrm{d}t}\phi(\boldsymbol{r}(t))\,\mathrm{d}t=\phi(\boldsymbol{r}(b))-\phi(\boldsymbol{r}(a))=\phi(B)-\phi(A).$$

证毕。

这个定理告诉我们：如果局部边际场来自一个全局势函数，那么沿任何路径累加，结果都只是终点函数值的差。

**例 6.2：效用变化为什么路径无关**

设

$$U(x,y)=x^2y+y^3.$$

则

$$\mathrm{d}U=2xy\,\mathrm{d}x+(x^2+3y^2)\,\mathrm{d}y.$$

若消费束从 $A=(1,1)$ 变到 $B=(2,1)$，沿任何路径都有

$$\int_C 2xy\,\mathrm{d}x+(x^2+3y^2)\,\mathrm{d}y=U(2,1)-U(1,1).$$

计算得

$$U(2,1)=4\cdot1+1=5,\qquad U(1,1)=1+1=2,$$

所以积分为

3.

这就是“效用函数存在”时的路径无关。

# 7 怎样判断一个场是不是保守场

## 7.1 二维局部判别条件

假设

$$\boldsymbol{F}(x,y) = (P(x,y), Q(x,y)).$$

如果 $\boldsymbol{F} = \nabla\phi$，那么

$$P = \phi_x, \qquad Q = \phi_y.$$

于是

$$P_y = \phi_{xy}, \qquad Q_x = \phi_{yx}.$$

如果 $\phi$ 二阶连续可微，由 Clairaut 定理，

$$\phi_{xy} = \phi_{yx}.$$

因此必要条件是

$$P_y = Q_x.$$

**二维保守场判别**

在一个没有洞的单连通区域上，如果 $P, Q$ 有连续一阶偏导，并且

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x},$$

则向量场 $\boldsymbol{F} = (P, Q)$ 是保守场，微分形式 $P\,\mathrm{d}x + Q\,\mathrm{d}y$ 是 exact differential。

**常见误区**

条件 $P_y = Q_x$ 不是在任何区域上都自动充分。它还需要区域没有洞，即单连通 simply connected。最常见反例是去掉原点的平面 $\mathbb{R}^2 \setminus \{(0,0)\}$。

## 7.2 找势函数的实际步骤

**方法**

给定 $P\,\mathrm{d}x + Q\,\mathrm{d}y$，找势函数 $\phi$ 的步骤：

1. 先检查 $P_y = Q_x$。若不相等，通常不是 exact；
2. 由 $\phi_x = P$，对 $x$ 积分：

$$\phi(x,y) = \int P(x,y)\,\mathrm{d}x + g(y),$$

这里 $g(y)$ 是对 $x$ 积分后的“常数”，但它可以依赖 $y$；

3. 对得到的 $\phi$ 求 $\phi_y$，并令 $\phi_y = Q$；
4. 解出 $g'(y)$，再积分得到 $g(y)$；
5. 写出势函数 $\phi$，差一个常数无所谓。

**例 6.3：找势函数**

判断

$$\boldsymbol{F}(x,y) = (2xy+3,\ x^2+4y)$$

是否保守，并求势函数。

**解：** 这里

$$P(x,y) = 2xy+3, \qquad Q(x,y) = x^2+4y.$$

先检查交叉偏导：

$$P_y = 2x, \qquad Q_x = 2x.$$

相等，所以在 $\mathbb{R}^2$ 上它是保守场。

由 $\phi_x = P = 2xy+3$，对 $x$ 积分：

$$\phi(x,y) = \int (2xy+3)\,\mathrm{d}x = x^2y + 3x + g(y).$$

求 $y$ 偏导：

$$\phi_y = x^2 + g'(y).$$

令其等于 $Q = x^2+4y$，得到

$$g'(y) = 4y,$$

所以

$$g(y) = 2y^2 + C.$$

因此势函数可以取为

$$\phi(x,y) = x^2y + 3x + 2y^2.$$

# 8 Green 定理：边界循环与内部旋转

Green 定理是向量分析中最有思想性的定理之一。本讲义不把它作为技巧堆叠，而是把它理解为“局部不一致如何累积成整体路径依赖”。

**Green 定理**

设 $D$ 是平面上的一个适当光滑区域，边界 $\partial D$ 取正向，即逆时针方向。若 $P, Q$ 在区域附近有连续偏导，则

$$\oint_{\partial D} P\,\mathrm{d}x + Q\,\mathrm{d}y = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}A.$$

右边的

$$Q_x - P_y$$

称为二维旋度 scalar curl。

**直觉**

Green 定理说：沿边界走一圈得到的总循环，等于区域内部每一点局部旋转倾向的总和。如果每一点都没有局部旋转，即 $Q_x - P_y = 0$，那么闭合路径积分为 0；这正是路径无关的核心。

## 8.1 为什么 Green 定理和路径依赖有关

如果 $\boldsymbol{F}$ 是保守场，那么沿任何闭合曲线 $C$ 都有

$$\oint_C \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r} = 0.$$

因为闭合曲线起点等于终点，势函数差值为 0。

Green 定理告诉我们：

$$\oint_C P\,\mathrm{d}x + Q\,\mathrm{d}y = \iint_D (Q_x - P_y)\,\mathrm{d}A.$$

所以如果 $Q_x - P_y \neq 0$，闭合路径积分一般就不会为 0，路径依赖就会出现。

**例 6.4：用 Green 定理算闭合路径积分**

计算

$$\oint_C (-y)\,\mathrm{d}x + x\,\mathrm{d}y,$$

其中 $C$ 是单位圆 $x^2 + y^2 = 1$ 的逆时针方向。

**解：** 这里

$$P = -y, \qquad Q = x.$$

所以

$$Q_x = 1, \qquad P_y = -1.$$

于是

$$Q_x - P_y = 1 - (-1) = 2.$$

由 Green 定理，

$$\oint_C (-y)\,\mathrm{d}x + x\,\mathrm{d}y = \iint_D 2\,\mathrm{d}A = 2\cdot\pi = 2\pi.$$

这说明向量场 $(-y, x)$ 有非零旋转，不可能是某个势函数的梯度。

## 8.2 一个重要反例：区域有洞时要小心

考虑

$$\boldsymbol{F}(x,y) = \left(-\frac{y}{x^2+y^2}, \frac{x}{x^2+y^2}\right),$$

定义域为

$$\mathbb{R}^2 \setminus \{(0,0)\}.$$

可以验证在定义域内

$$P_y = Q_x.$$

但是沿单位圆逆时针一圈：

$$x = \cos t, \qquad y = \sin t, \qquad 0 \le t \le 2\pi,$$

有

$$\boldsymbol{F}(\boldsymbol{r}(t)) = (-\sin t, \cos t), \qquad \boldsymbol{r}'(t) = (-\sin t, \cos t).$$

所以

$$\oint_C \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r} = \int_0^{2\pi} 1\,\mathrm{d}t = 2\pi \neq 0.$$

它不是保守场。

为什么没有矛盾？因为定义域去掉了原点，有一个洞，不是单连通区域。

# 9 经济金融应用

## 9.1 应用一：福利变化是否可以由终点差值表示

假设政策状态为 $(\tau, s)$，其中 $\tau$ 是税率，$s$ 是补贴率。局部福利变化可以写成

$$\mathrm{d}W = M_\tau(\tau,s)\,\mathrm{d}\tau + M_s(\tau,s)\,\mathrm{d}s.$$

如果存在一个全局福利函数 $W(\tau,s)$，那么

$$M_\tau = \frac{\partial W}{\partial \tau}, \qquad M_s = \frac{\partial W}{\partial s}.$$

于是必须满足

$$\frac{\partial M_\tau}{\partial s} = \frac{\partial M_s}{\partial \tau}.$$

如果这个条件不成立，说明你给出的两个“局部边际福利”并不能整合为同一个全局福利函数；政策路径会影响总福利测度。

**例 6.5：一个政策福利场**

设

$$\mathrm{d}W = (2\tau + s)\,\mathrm{d}\tau + (\tau + 4s)\,\mathrm{d}s.$$

判断它是否 exact，并求 $W$。

**解：** 令

$$P = 2\tau + s, \qquad Q = \tau + 4s.$$

有

$$P_s = 1, \qquad Q_\tau = 1.$$

相等，所以 exact。

由 $W_\tau = 2\tau + s$，对 $\tau$ 积分：

$$W(\tau, s) = \tau^2 + s\tau + g(s).$$

求 $s$ 偏导：

$$W_s = \tau + g'(s).$$

令其等于 $Q = \tau + 4s$，得

$$g'(s) = 4s, \qquad g(s) = 2s^2.$$

所以

$$W(\tau, s) = \tau^2 + s\tau + 2s^2.$$

## 9.2 应用二：从边际效用到效用函数

如果我们知道某消费者在每一点的边际效用向量

$$\boldsymbol{F}(x, y) = (MU_x(x, y), MU_y(x, y)),$$

那么要问：是否存在一个效用函数 $U(x, y)$，使得

$$U_x = MU_x, \qquad U_y = MU_y?$$

这就是 exact differential 问题。

如果存在，就有

$$\mathrm{d}U = MU_x\,\mathrm{d}x + MU_y\,\mathrm{d}y.$$

如果不存在，则这些局部边际效用不是来自同一个效用函数，存在内在不一致。

这和消费者理论里的 integrability condition 有深层联系。高级微观里，从需求函数反推出偏好，需要满足一组交叉对称性和负半定性条件；本章的 $P_y = Q_x$ 是最简单的二维版本。

### 9.3 应用三：调整成本与路径依赖

企业把生产状态从 $(K, L)$ 调整到 $(K', L')$。如果局部调整成本是

$$\mathrm{d}C = A(K, L)\,\mathrm{d}K + B(K, L)\,\mathrm{d}L,$$

并且

$$A_L \neq B_K,$$

则调整成本依赖路径。经济含义是：先调资本后调劳动，与先调劳动后调资本，成本不同。

这在动态模型里很自然。因为当企业先改变资本时，后续劳动调整的边际成本可能改变；反过来也一样。

### 9.4 应用四：投资组合再平衡的路径成本

设投资组合权重为 $(w_1, w_2)$。从旧权重调到新权重时，交易成本可能不仅取决于最终权重差，还取决于调仓路径。局部成本形式可以写成

$$c_1(w_1, w_2)\,\mathrm{d}w_1 + c_2(w_1, w_2)\,\mathrm{d}w_2.$$

若

$$\frac{\partial c_1}{\partial w_2} eq \frac{\partial c_2}{\partial w_1},$$

则再平衡成本路径依赖。这对应现实中市场冲击、流动性、交易顺序和价格反馈等因素。

## 10 Worked Examples：完整样板题

**Worked Example 1：第一类曲线积分**

求

$$\int_C (x + 2y)\,\mathrm{d}s,$$

其中 $C$ 是从 $(0, 0)$ 到 $(1, 1)$ 的直线段。

**Step 1：参数化路径。**

$$\boldsymbol{r}(t) = (t, t), \qquad 0 \le t \le 1.$$

**Step 2：计算弧长因子。**

$$\boldsymbol{r}'(t) = (1, 1), \qquad \|\boldsymbol{r}'(t)\| = \sqrt{2}.$$

**Step 3：代入密度。**

$$x + 2y = t + 2t = 3t.$$

**Step 4：积分。**

$$\int_C (x + 2y)\,\mathrm{d}s = \int_0^1 3t\sqrt{2}\,\mathrm{d}t = \frac{3\sqrt{2}}{2}.$$

### Worked Example 2：第二类曲线积分

求

$$\int_C (x+y)\,\mathrm{d}x + (x-y)\,\mathrm{d}y,$$

其中 $C$ 为抛物线 $y=x^2$ 从 $(0,0)$ 到 $(1,1)$。

**Step 1：参数化。**

$$x = t, \qquad y = t^2, \qquad 0 \le t \le 1.$$

于是

$$\mathrm{d}x = \mathrm{d}t, \qquad \mathrm{d}y = 2t\,\mathrm{d}t.$$

**Step 2：代入。**

$$x+y = t+t^2, \qquad x-y = t-t^2.$$

所以

$$\int_C (x+y)\,\mathrm{d}x + (x-y)\,\mathrm{d}y = \int_0^1 \left[(t+t^2) + (t-t^2)2t\right]\,\mathrm{d}t.$$

**Step 3：整理。**

$$(t+t^2) + 2t(t-t^2) = t + t^2 + 2t^2 - 2t^3 = t + 3t^2 - 2t^3.$$

**Step 4：积分。**

$$\int_0^1 (t+3t^2-2t^3)\,\mathrm{d}t = \frac{1}{2} + 1 - \frac{1}{2} = 1.$$

### Worked Example 3：证明路径依赖

设

$$\boldsymbol{F}(x,y) = (y, 0).$$

从 $(0,0)$ 到 $(1,1)$，比较两条路径：

$$C_1 : (0,0) \to (1,0) \to (1,1),$$

$$C_2 : (0,0) \to (0,1) \to (1,1).$$

计算 $\int_C y\,\mathrm{d}x$。

**路径 $C_1$：**第一段 $y=0$，所以

$$\int y\,\mathrm{d}x = 0.$$

第二段 $x=1$，所以 $\mathrm{d}x = 0$，积分仍为 0。因此

$$\int_{C_1} y\,\mathrm{d}x = 0.$$

**路径 $C_2$：**第一段 $x=0$，所以 $\mathrm{d}x=0$，积分为 0。第二段 $y=1$，$x:0\to 1$，所以

$$\int y\,\mathrm{d}x=\int_0^1 1\,\mathrm{d}x=1.$$

因此

$$\int_{C_2} y\,\mathrm{d}x=1.$$

同样端点，不同路径，积分不同，所以路径依赖。

## Worked Example 4：用势函数快速算线积分

计算

$$\int_C (3x^2y+2)\,\mathrm{d}x+(x^3+2y)\,\mathrm{d}y,$$

其中 $C$ 是任意从 $(0,1)$ 到 $(1,2)$ 的路径。

**Step 1：判断是否 exact。**

$$P=3x^2y+2,\qquad Q=x^3+2y.$$

于是

$$P_y=3x^2,\qquad Q_x=3x^2.$$

相等，区域为 $\mathbb{R}^2$，所以 exact。

**Step 2：找势函数。**由 $\phi_x=3x^2y+2$，对 $x$ 积分：

$$\phi=x^3y+2x+g(y).$$

求 $y$ 偏导：

$$\phi_y=x^3+g'(y).$$

令其等于 $Q=x^3+2y$，得

$$g'(y)=2y,\qquad g(y)=y^2.$$

所以

$$\phi(x,y)=x^3y+2x+y^2.$$

**Step 3：用终点差。**

$$\int_C \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}=\phi(1,2)-\phi(0,1).$$

$$\phi(1,2)=1^3\cdot 2+2\cdot 1+4=8,$$

$$\phi(0,1)=0+0+1=1.$$

所以积分为

7.

**Worked Example 5：用 Green 定理计算矩形边界积分**

令 $C$ 为矩形 $[0,2]\times[0,1]$ 的逆时针边界，计算

$$\oint_C y^2\,\mathrm{d}x + x^2\,\mathrm{d}y.$$

**解：**这里

$$P = y^2, \qquad Q = x^2.$$

因此

$$Q_x = 2x, \qquad P_y = 2y.$$

由 Green 定理，

$$\oint_C y^2\,\mathrm{d}x + x^2\,\mathrm{d}y = \int_0^2 \int_0^1 (2x - 2y)\,\mathrm{d}y\,\mathrm{d}x.$$

先对 $y$ 积分：

$$\int_0^1 (2x - 2y)\,\mathrm{d}y = 2x - 1.$$

再对 $x$ 积分：

$$\int_0^2 (2x - 1)\,\mathrm{d}x = [x^2 - x]_0^2 = 4 - 2 = 2.$$

所以

$$\oint_C y^2\,\mathrm{d}x + x^2\,\mathrm{d}y = 2.$$

# 11 常见错误与纠正

**常见误区**

1. **把 $\mathrm{d}s$ 写成 $\mathrm{d}t$。**第一类曲线积分必须乘 $\|\boldsymbol{r}'(t)\|$。
2. **忘记方向。**第二类曲线积分反向走会变号。
3. **把路径无关当作默认。**只有保守场或者 exact differential 才路径无关。
4. **只检查 $P_y = Q_x$，忘记区域条件。**有洞区域上可能失败。
5. **找势函数时把积分常数当成常数。**对 $x$ 积分后的“常数”是 $g(y)$，对 $y$ 积分后的“常数”是 $h(x)$。
6. **Green 定理方向弄反。**逆时针是正向；顺时针会多一个负号。

# 12 练习题

## A 组：基础计算

A1. 计算 $\int_C (x+y)\,\mathrm{d}s$，其中 $C$ 为线段 $(0,0)\to(2,1)$。

A2. 计算 $\int_C x\,\mathrm{d}x + y\,\mathrm{d}y$，其中 $C$ 为 $\boldsymbol{r}(t)=(t,t^2)$, $0\le t\le 1$。

A3. 计算 $\int_C y\,\mathrm{d}x$，其中 $C$ 为单位圆上半圆 $(1,0)\to(-1,0)$，逆时针方向。

A4. 判断 $(2xy, y^2+x^2)$ 是否为保守场。

A5. 判断 $(e^x\cos y, -e^x\sin y)$ 是否为保守场，并求势函数。

## B 组：概念判断

B1. 为什么 $\int_C f\,\mathrm{d}s$ 通常不随路径方向改变，而 $\int_C \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}$ 会随方向改变?

B2. 若 $\boldsymbol{F}=\nabla\phi$，为什么任意闭合路径积分都等于 0?

B3. 若 $P_y\neq Q_x$，能否存在势函数? 说明理由。

B4. 如果 $P_y=Q_x$，一定保守吗? 说明区域条件。

B5. Green 定理中的 $Q_x-P_y$ 可以怎样理解?

## C 组：混合题

C1. 计算 $\int_C (x^2+y)\,\mathrm{d}x + (x+y^2)\,\mathrm{d}y$，其中 $C$ 是从 $(0,0)$ 到 $(1,1)$ 的直线。

C2. 对上题，改用路径 $y=x^2$，比较结果。这个场是否保守?

C3. 令 $C$ 为单位圆逆时针方向，计算 $\oint_C (x-y)\,\mathrm{d}x + (x+y)\,\mathrm{d}y$。

C4. 求 $\int_C (y+2x)\,\mathrm{d}x + (x+2y)\,\mathrm{d}y$，其中 $C$ 为任意从 $(1,0)$ 到 $(2,3)$ 的路径。

C5. 若 $\boldsymbol{F}=(ay, bx)$，在什么条件下保守? 若不保守，单位正方形边界上的闭合积分是多少?

## D 组：经济金融应用题

D1. 设政策福利微分为

$$\mathrm{d}W = (3\tau+2s)\,\mathrm{d}\tau + (2\tau+s)\,\mathrm{d}s.$$

判断是否 exact，并求 $W(\tau,s)$。

D2. 设局部调整成本为

$$\mathrm{d}C = (K+L)\,\mathrm{d}K + (2L)\,\mathrm{d}L.$$

判断资本和劳动调整是否路径依赖。

D3. 设边际效用场为

$$(MU_x, MU_y) = \left(\frac{1}{x}, \frac{1}{y}\right), \qquad x, y > 0.$$

是否存在效用函数？若存在，求一个。

D4. 设再平衡成本场为

$$\boldsymbol{F}(w_1, w_2) = (w_2, w_1 + w_2).$$

判断投资组合再平衡成本是否路径无关。

D5. 设局部福利场为 $\boldsymbol{F} = (y, x + y)$。比较从 $(0,0)$ 到 $(1,1)$ 沿直线与沿折线路径 $(0,0) \to (1,0) \to (1,1)$ 的积分。

### E 组：挑战与证明题

E1. 证明：如果 $\boldsymbol{F} = \nabla\phi$，则线积分与路径无关。

E2. 证明：如果任意闭合路径上的积分为 0，则积分与路径无关。

E3. 设 $\boldsymbol{F} = (P, Q)$ 保守且二阶连续可微，证明 $P_y = Q_x$。

E4. 对 $\boldsymbol{F} = (-y/(x^2+y^2), x/(x^2+y^2))$，直接验证在定义域上 $P_y = Q_x$，并解释为什么它仍不是保守场。

E5. 用 Green 定理解释：为什么 $Q_x - P_y$ 可以看作路径依赖的局部来源。

## 13 练习题详细答案

### A 组答案

**A1**

线段参数化为

$$\boldsymbol{r}(t) = (2t, t), \qquad 0 \le t \le 1.$$

则

$$\|\boldsymbol{r}'(t)\| = \sqrt{2^2 + 1^2} = \sqrt{5},$$

且

$$x + y = 2t + t = 3t.$$

所以

$$\int_C (x + y)\,\mathrm{d}s = \int_0^1 3t\sqrt{5}\,\mathrm{d}t = \frac{3\sqrt{5}}{2}.$$

**A2**

路径 $x=t, y=t^2$。于是

$$\mathrm{d}x=\mathrm{d}t, \qquad \mathrm{d}y=2t\,\mathrm{d}t.$$

所以

$$\int_C x\,\mathrm{d}x+y\,\mathrm{d}y=\int_0^1 \left[t+t^2(2t)\right]\,\mathrm{d}t=\int_0^1 (t+2t^3)\,\mathrm{d}t=\frac{1}{2}+\frac{1}{2}=1.$$

也可注意到

$$x\,\mathrm{d}x+y\,\mathrm{d}y=\mathrm{d}\left(\frac{x^2+y^2}{2}\right),$$

所以结果等于

$$\frac{1^2+1^2}{2}-0=1.$$

**A3**

上半圆逆时针可参数化为

$$x=\cos t, \qquad y=\sin t, \qquad 0\le t\le \pi.$$

于是

$$\mathrm{d}x=-\sin t\,\mathrm{d}t.$$

所以

$$\int_C y\,\mathrm{d}x=\int_0^{\pi}\sin t(-\sin t)\,\mathrm{d}t=-\int_0^{\pi}\sin^2 t\,\mathrm{d}t=-\frac{\pi}{2}.$$

**A4**

这里

$$P=2xy, \qquad Q=y^2+x^2.$$

有

$$P_y=2x, \qquad Q_x=2x.$$

相等，在 $\mathbb{R}^2$ 上保守。求势函数：由 $\phi_x=2xy$,

$$\phi=x^2y+g(y).$$

于是

$$\phi_y=x^2+g'(y)=x^2+y^2,$$

所以 $g'(y)=y^2$,

$$g(y)=\frac{y^3}{3}.$$

势函数为

$$\phi = x^2 y + \frac{y^3}{3}.$$

### A5

这里

$$P = e^x \cos y, \qquad Q = -e^x \sin y.$$

有

$$P_y = -e^x \sin y, \qquad Q_x = -e^x \sin y.$$

相等，所以保守。由 $\phi_x = e^x \cos y$，得

$$\phi = e^x \cos y + g(y).$$

求 $y$ 偏导:

$$\phi_y = -e^x \sin y + g'(y).$$

令其等于 $Q = -e^x \sin y$，得 $g'(y) = 0$。故可取

$$\phi = e^x \cos y.$$

## B 组答案

### B1

$\int_C f \,\mathrm{d}s$ 中 $\mathrm{d}s$ 是弧长元素，始终非负。反向走同一路径，长度不变，所以结果通常不变。$\int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$ 中 $\mathrm{d}\boldsymbol{r}$ 是有方向的小位移，反向走时 $\mathrm{d}\boldsymbol{r}$ 变成 $-\mathrm{d}\boldsymbol{r}$，所以积分变号。

### B2

若 $\boldsymbol{F} = \nabla\phi$，线积分基本定理给出

$$\int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \phi(B) - \phi(A).$$

闭合路径满足 $A = B$，所以

$$\phi(B) - \phi(A) = 0.$$

因此任意闭合路径积分为 0。

**B3**

如果存在势函数 $\phi$，则 $P=\phi_x, Q=\phi_y$。若 $\phi$ 二阶连续可微，则

$$P_y=\phi_{xy}=\phi_{yx}=Q_x.$$

因此 $P_y=Q_x$ 是必要条件。若它不成立，就不能存在这样的二阶良好势函数。

**B4**

不一定。$P_y=Q_x$ 在单连通区域上通常是充分条件，但如果区域有洞，可能失败。经典例子是

$$\left(-\frac{y}{x^2+y^2}, \frac{x}{x^2+y^2}\right)$$

在 $\mathbb{R}^2\setminus\{0\}$ 上满足交叉偏导相等，但绕原点一圈积分为 $2\pi$，所以不是保守场。

**B5**

$Q_x-P_y$ 是二维旋度，表示向量场在某点附近的局部循环倾向。Green 定理说明，边界上的总循环等于区域内部局部旋转的累积。若它处处为 0，则闭合路径积分倾向于为 0，路径无关成为可能。

## C 组答案

**C1**

直线路径 $x=t, y=t$, $0\le t\le 1$。有 $\mathrm{d}x=\mathrm{d}t$, $\mathrm{d}y=\mathrm{d}t$。积分为

$$\int_0^1 [(t^2+t)+(t+t^2)]\,\mathrm{d}t=\int_0^1 (2t^2+2t)\,\mathrm{d}t=\frac{2}{3}+1=\frac{5}{3}.$$

**C2**

路径 $x=t, y=t^2$, $0\le t\le 1$。有

$$\mathrm{d}x=\mathrm{d}t, \qquad \mathrm{d}y=2t\,\mathrm{d}t.$$

于是

$$P=x^2+y=t^2+t^2=2t^2,$$

$$Q=x+y^2=t+t^4.$$

积分为

$$\int_0^1 \left[2t^2+(t+t^4)2t\right]\,\mathrm{d}t=\int_0^1 (2t^2+2t^2+2t^5)\,\mathrm{d}t=\frac{4}{3}+\frac{1}{3}=\frac{5}{3}.$$

结果相同。检查保守性:

$$P_y = 1, \qquad Q_x = 1.$$

在 $\mathbb{R}^2$ 上保守，所以路径无关。

## C3

这里

$$P = x - y, \qquad Q = x + y.$$

由 Green 定理,

$$\oint_C P\,\mathrm{d}x + Q\,\mathrm{d}y = \iint_D (Q_x - P_y)\,\mathrm{d}A.$$

有

$$Q_x = 1, \qquad P_y = -1,$$

所以

$$Q_x - P_y = 2.$$

单位圆面积为 $\pi$，故积分为

$$2\pi.$$

## C4

$$P = y + 2x, \qquad Q = x + 2y.$$

检查

$$P_y = 1, \qquad Q_x = 1.$$

保守。由 $\phi_x = y + 2x$，得

$$\phi = xy + x^2 + g(y).$$

于是

$$\phi_y = x + g'(y) = x + 2y,$$

所以 $g'(y) = 2y$，$g(y) = y^2$。势函数为

$$\phi = x^2 + xy + y^2.$$

积分等于

$$\phi(2,3) - \phi(1,0) = (4 + 6 + 9) - 1 = 18.$$

**C5**

$$P = ay, \qquad Q = bx.$$

有

$$P_y = a, \qquad Q_x = b.$$

保守条件为

$$a = b.$$

若单位正方形边界逆时针，Green 定理给出

$$\oint_C ay\,\mathrm{d}x + bx\,\mathrm{d}y = \iint_{[0,1]^2} (b-a)\,\mathrm{d}A = b - a.$$

## D 组答案

**D1**

$$P = 3\tau + 2s, \qquad Q = 2\tau + s.$$

有

$$P_s = 2, \qquad Q_\tau = 2.$$

所以 exact。由 $W_\tau = 3\tau + 2s$，得

$$W = \frac{3}{2}\tau^2 + 2s\tau + g(s).$$

求 $s$ 偏导:

$$W_s = 2\tau + g'(s) = 2\tau + s,$$

故 $g'(s) = s$,

$$g(s) = \frac{1}{2}s^2.$$

所以

$$W(\tau, s) = \frac{3}{2}\tau^2 + 2s\tau + \frac{1}{2}s^2.$$

**D2**

这里

$$P = K + L, \qquad Q = 2L.$$

检查

$$P_L = 1, \qquad Q_K = 0.$$

不相等，因此不是 exact，调整成本路径依赖。经济含义是：资本变化对后续劳动

调整的局部成本影响，与劳动变化对资本调整的局部成本影响不对称。

**D3**

$$P = \frac{1}{x}, \qquad Q = \frac{1}{y}.$$

在 $x, y > 0$ 上，

$$P_y = 0, \qquad Q_x = 0.$$

相等，所以存在效用函数。由 $U_x = 1/x$，得

$$U = \ln x + g(y).$$

再由 $U_y = g'(y) = 1/y$，得

$$g(y) = \ln y.$$

因此可取

$$U(x, y) = \ln x + \ln y.$$

**D4**

$$P = w_2, \qquad Q = w_1 + w_2.$$

有

$$P_{w_2} = 1, \qquad Q_{w_1} = 1.$$

在普通平面区域上保守。势函数：由 $\phi_{w_1} = w_2$，

$$\phi = w_1 w_2 + g(w_2).$$

再由

$$\phi_{w_2} = w_1 + g'(w_2) = w_1 + w_2,$$

得

$$g(w_2) = \frac{1}{2} w_2^2.$$

所以

$$\phi = w_1 w_2 + \frac{1}{2} w_2^2.$$

再平衡成本路径无关。

**D5**

$$P = y, \qquad Q = x + y.$$

直线路径 $x = t, y = t$，有

$$\mathrm{d}x = \mathrm{d}t, \qquad \mathrm{d}y = \mathrm{d}t.$$

积分为

$$\int_0^1 [t + (t + t)]\,\mathrm{d}t = \int_0^1 3t\,\mathrm{d}t = \frac{3}{2}.$$

折线路径第一段 $(0,0) \to (1,0)$：$y = 0$, $\mathrm{d}y = 0$，积分为 0。第二段 $(1,0) \to (1,1)$：$x = 1$, $\mathrm{d}x = 0, y : 0 \to 1$，积分为

$$\int_0^1 (1 + y)\,\mathrm{d}y = \frac{3}{2}.$$

结果仍为 $3/2$。其实

$$P_y = 1, \qquad Q_x = 1,$$

该场保守，所以路径无关。势函数可取

$$\phi = xy + \frac{1}{2}y^2.$$

## E 组答案

**E1**

这正是线积分基本定理。若 $\boldsymbol{F} = \nabla\phi$，路径 $C$ 从 $A$ 到 $B$，则

$$\int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \phi(B) - \phi(A).$$

右边只依赖起点和终点，不依赖路径，所以路径无关。

**E2**

设 $C_1, C_2$ 是从 $A$ 到 $B$ 的两条路径。令 $-C_2$ 表示沿 $C_2$ 反向从 $B$ 到 $A$。那么 $C_1 \cup (-C_2)$ 是闭合路径。若任意闭合路径积分为 0，则

$$\int_{C_1} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} + \int_{-C_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = 0.$$

而

$$\int_{-C_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = -\int_{C_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}.$$

因此

$$\int_{C_1} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \int_{C_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}.$$

所以路径无关。

## E3

若 $\boldsymbol{F} = (P, Q)$ 保守，则存在 $\phi$，使

$$P = \phi_x, \qquad Q = \phi_y.$$

因为 $\phi$ 二阶连续可微，由混合偏导相等，

$$P_y = (\phi_x)_y = \phi_{xy} = \phi_{yx} = (\phi_y)_x = Q_x.$$

所以 $P_y = Q_x$。

## E4

设

$$P = -\frac{y}{x^2+y^2}, \qquad Q = \frac{x}{x^2+y^2}.$$

直接计算可得

$$P_y = \frac{y^2-x^2}{(x^2+y^2)^2}, \qquad Q_x = \frac{y^2-x^2}{(x^2+y^2)^2}.$$

所以在去掉原点的区域上 $P_y = Q_x$。但是沿单位圆逆时针积分为

$$\int_0^{2\pi} 1 \,\mathrm{d}t = 2\pi \neq 0.$$

因此不是保守场。原因是定义域 $\mathbb{R}^2 \setminus \{0\}$ 有洞，不是单连通。

## E5

Green 定理给出

$$\oint_{\partial D} P \,\mathrm{d}x + Q \,\mathrm{d}y = \iint_D (Q_x - P_y) \,\mathrm{d}A.$$

左边是沿闭合路径走一圈的净累积。若它不为 0，说明从某点出发绕一圈回来之后，局部边际量仍累积出非零净变化，这就意味着无法用一个单值势函数描述总变化。右边显示这种非零循环来自区域内部每一点的局部旋转 $Q_x - P_y$。因此 $Q_x - P_y$ 可以看作路径依赖的局部来源。

# 14 本章总结图谱

**本章压缩总结**

| 概念 | 要点 |
| --- | --- |
| **第一类曲线积分** | $\int_C f\,\mathrm{d}s=\int_a^b f(\boldsymbol{r}(t))\,\|\boldsymbol{r}'(t)\|\,\mathrm{d}t$，沿路径按长度累加密度。 |
| **第二类曲线积分** | $\int_C \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}=\int_C P\,\mathrm{d}x+Q\,\mathrm{d}y$，沿路径累加向量场在前进方向上的分量。 |
| **路径依赖** | 同样起点终点，不同路径积分不同。经济上对应改革顺序、调整路径、交易顺序影响总结果。 |
| **保守场** | $\boldsymbol{F}=\nabla\phi$。此时线积分只等于 $\phi(B)-\phi(A)$。 |
| **exact differential** | $P\,\mathrm{d}x+Q\,\mathrm{d}y=\mathrm{d}\phi$。局部边际量可以整合成全局函数。 |
| **判别式** | 在单连通区域上，若 $P_y=Q_x$，则通常保守。 |
| **Green 定理** | $\oint_{\partial D}P\,\mathrm{d}x+Q\,\mathrm{d}y=\iint_D(Q_x-P_y)\,\mathrm{d}A$。边界循环等于内部旋度总和。 |
| **经济含义** | 局部边际福利、局部成本、边际效用、调仓成本能否整合成全局函数，取决于 exactness 和路径无关。 |

**直觉**

本章最重要的一句话：**如果局部边际场可以写成某个函数的梯度，那么沿路径累加就只看起点和终点；如果不能，路径本身就是模型的一部分。**

# 下一章连接

下一章将讲 Green、Gauss、Stokes 的直观版。它会把本章的思想进一步推广：

- Green 定理：平面区域的边界循环与内部旋度；
- Gauss 定理：空间区域边界通量与内部散度；
- Stokes 定理：曲面边界循环与曲面旋度。

对经济金融来说，下一章不是为了训练复杂曲面积分，而是为了建立一个更成熟的数学直觉：**局部条件如何推出整体条件，整体约束如何反映局部结构。**

# 经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

---

# Chapter 7

## Green、Gauss、Stokes 的直观版

---

本章的目标不是训练复杂曲面积分技巧，而是理解一个统一思想：边界上的累计量，等于内部局部量的总和。Green、Gauss、Stokes 是同一个思想在平面、空间体和曲面上的三个版本。

math4econ personalized notes

2026

# 目录

**Chapter 7 总览** **2**

**1 本章为什么重要** **3**
1.1 为什么经济金融学生需要它 . . . . . 3

**2 最小前置知识** **4**
2.1 第一，上一章的线积分 . . . . . 4
2.2 第二，偏导和局部线性近似 . . . . . 4
2.3 第三，面积、体积和曲面上的积分 . . . . . 4

**3 Green 定理：边界循环等于内部旋度总和** **5**
3.1 从上一章的路径积分出发 . . . . . 5
3.2 Green 定理的直觉推导：小矩形相互抵消 . . . . . 6
3.3 计算例题 1：用 Green 定理算闭合路径积分 . . . . . 7
3.4 经济金融解释：闭合调整路径的净收益 . . . . . 7

**4 散度与通量：进入 Gauss 定理之前** **8**
4.1 通量：穿过边界的净流动 . . . . . 8
4.2 散度：局部源泉和汇点 . . . . . 9
4.3 二维散度定理：Gauss 的平面版本 . . . . . 9

**5 Gauss 定理：封闭曲面通量等于体内散度总和** **10**
5.1 定理表述 . . . . . 10
5.2 Gauss 定理的直觉推导：小盒子的内部面抵消 . . . . . 10
5.3 计算例题 4：球面通量 . . . . . 11
5.4 经济金融解释：状态空间里的守恒 . . . . . 11

**6 旋度与 Stokes 定理** **12**
6.1 三维旋度 . . . . . 12
6.2 Stokes 定理 . . . . . 12
6.3 Green 定理是 Stokes 定理的特例 . . . . . 13
6.4 计算例题 6：Stokes 比直接算更简单 . . . . . 13
6.5 经济金融解释：局部无旋和全局可积性 . . . . . 14

**7 三个定理的统一图谱** **14**
7.1 统一形式 . . . . . 15
7.2 旋度和散度的区别 . . . . . 15

**8 经济金融中的三个模型化入口** **15**
8.1 入口一：守恒方程和分布动态 . . . . . 15
8.2 入口二：无套利和闭合交易路径 . . . . . 16
8.3 入口三：PDE 与连续时间资产定价 . . . . . 16

**9 Worked Examples：完整样板题** **16**

**10 练习题** **19**
A 组：基础计算 . . . . . 19
B 组：概念判断 . . . . . 19
C 组：混合题 . . . . . 19
D 组：经济金融应用题 . . . . . 19
E 组：挑战/证明题 . . . . . 20

**11 练习题详细解答** **21**
A 组答案 . . . . . 21
B 组答案 . . . . . 22
C 组答案 . . . . . 23
D 组答案 . . . . . 24
E 组答案 . . . . . 25

**12 本章总结图谱** **27**

**下一章连接** **28**

# Chapter 7 总览

**本章目标**

学完本章，你应该达到以下目标：

1. 能把 Green 定理理解成：平面区域边界上的总循环等于区域内部旋度的总和；
2. 能把 Gauss 定理理解成：封闭曲面上的净流出量等于体内部源泉强度的总和；
3. 能把 Stokes 定理理解成：曲面边界上的总循环等于曲面上旋度通量的总和；
4. 掌握散度 divergence、旋度 curl、通量 flux、法向量 normal vector、正向 orientation 的基本含义；
5. 会做经济金融需要的基本计算：矩形/圆盘上的 Green 定理、球面/盒子上的 Gauss 定理、简单曲面上的 Stokes 定理；
6. 理解三个定理的共同结构：局部量经过内部累加，转化为边界上的整体量；
7. 能解释它们和连续时间模型、守恒方程、PDE、状态空间中的财富流、风险暴露流、套利条件之间的关系；
8. 知道本章在经济金融里是数学成熟度章节，而不是短期刷题主战场。

**直觉**

一句话：**Green、Gauss、Stokes 都在说“内部的局部结构，会在边界上留下整体痕迹”**。

上一章讲曲线积分，重点是沿一条路径累加。本章进一步问：如果路径是某个区域或曲面的边界，那么边界上的累加量能不能直接由内部每一点的局部性质算出来？答案就是三大定理。

**本章压缩总结**

三个定理可以压缩成一张表：

| 定理 | 边界量 | 内部局部量 |
|---|---|---|
| **Green** | 平面闭曲线上的循环 | 区域内部的二维旋度 $Q_x - P_y$ |
| **Gauss** | 封闭曲面上的通量 | 体内部的散度 $\nabla \cdot \boldsymbol{F}$ |
| **Stokes** | 曲面边界上的循环 | 曲面上的旋度通量 $(\nabla \times \boldsymbol{F}) \cdot \boldsymbol{n}$ |

# 1 本章为什么重要

上一章你已经学过第二类曲线积分：

$$\int_C \boldsymbol{F}\cdot \mathrm{d}\boldsymbol{r} = \int_C P\,\mathrm{d}x + Q\,\mathrm{d}y.$$

它表示沿路径累加局部边际场。现在我们考虑一种特殊但非常重要的情形：路径 $C$ 是某个区域 $D$ 的边界 $\partial D$。

这时我们想问：

> 沿边界走一圈得到的总循环，能不能不用逐段算边界，而是直接看区域内部每一点的局部旋转？

Green 定理回答了这个问题。进一步，在三维空间中，如果向量场表示一种流动，例如人口流、财富流、概率质量流、资本流、风险暴露流，我们想问：

> 穿过封闭边界流出去的净量，是否等于内部所有源泉和汇点的总和？

Gauss 定理回答了这个问题。再进一步，如果边界不再是平面区域边界，而是空间中某张曲面的边界，我们想问：

> 曲面边界上的总循环，是否等于曲面内部局部旋转在法向方向上的总和？

Stokes 定理回答了这个问题。

**本章的统一问题**

给定一个区域、体或曲面，边界上的整体积分是否可以转换成内部的局部积分？

$$\boxed{\text{boundary integral} = \text{interior integral of local structure.}}$$

这就是本章的中心思想。

## 1.1 为什么经济金融学生需要它

你短期内不一定会在经济学作业里大量计算 Stokes 定理，但这章仍然有价值，因为它训练的是一种模型语言：从局部规律推出整体约束。

经济金融中的典型对应包括：

- **守恒方程**：财富、人口、概率质量在状态空间中流动，边界流出量和内部增长率相关；
- **PDE 直觉**：资产定价、连续时间宏观、最优控制和随机过程经常出现“局部变化率 + 流量 = 0”的结构；
- **无套利条件**：某些局部价格场若沿闭合交易路径产生非零循环，就暗示可能存在套利或摩擦；

- **福利与路径依赖**：如果局部福利变化无法整合成全局福利函数，闭合路径积分可能不为零；
- **分布动态**：异质性主体分布随时间移动时，分布质量穿过边界的通量很重要。

**常见误区**

本章容易学偏。你不需要把它学成数学系向量分析的计算竞赛。对经济金融来说，更重要的是：知道散度和旋度分别刻画什么，知道三大定理为什么把边界和内部联系起来，知道什么时候可以用它们简化计算和理解模型。

# 2 最小前置知识

本章需要三类前置知识。

## 2.1 第一，上一章的线积分

对平面向量场

$$\boldsymbol{F}(x,y)=(P(x,y),Q(x,y)),$$

沿曲线 $C$ 的线积分是

$$\int_C \boldsymbol{F}\cdot \mathrm{d}\boldsymbol{r}=\int_C P\,\mathrm{d}x+Q\,\mathrm{d}y.$$

若路径用参数 $\boldsymbol{r}(t)=(x(t),y(t))$，$t\in[a,b]$ 表示，则

$$\int_C \boldsymbol{F}\cdot \mathrm{d}\boldsymbol{r}=\int_a^b \left[P(x(t),y(t))x'(t)+Q(x(t),y(t))y'(t)\right]\,\mathrm{d}t.$$

## 2.2 第二，偏导和局部线性近似

你需要会理解：

$$P_y=\frac{\partial P}{\partial y},\qquad Q_x=\frac{\partial Q}{\partial x}.$$

在 Green 定理里，组合

$$Q_x-P_y$$

衡量向量场在平面上的局部旋转趋势。

## 2.3 第三，面积、体积和曲面上的积分

本章会出现三种内部积分：

$$\iint_D f(x,y)\,\mathrm{d}A,\qquad \iiint_V f(x,y,z)\,\mathrm{d}V,\qquad \iint_S \boldsymbol{F}\cdot\boldsymbol{n}\,\mathrm{d}S.$$

你暂时不需要掌握复杂曲面参数化，但需要理解：

- $\mathrm{d}A$：面积小块；

- $\mathrm{d}V$：体积小块；
- $\boldsymbol{n}\,\mathrm{d}S$：带方向的面积小块。

# 3 Green 定理：边界循环等于内部旋度总和

## 3.1 从上一章的路径积分出发

设 $D \subset \mathbb{R}^2$ 是一个平面区域，边界记作 $\partial D$。如果 $\partial D$ 按逆时针方向走，这叫正向 orientation。直觉是：沿边界走的时候，区域始终在你的左手边。

**二维旋度的标量形式**

对平面向量场

$$\boldsymbol{F}(x,y) = (P(x,y), Q(x,y)),$$

它的二维旋度可以写成

$$\operatorname{curl}_{2D} \boldsymbol{F} = Q_x - P_y.$$

它衡量的是向量场在点 $(x,y)$ 附近推动一个微小小轮子逆时针旋转的强度。

**Green 定理**

如果 $D$ 是平面上足够规则的区域，边界 $\partial D$ 按正向取向，且 $P, Q$ 有连续偏导数，则

$$\boxed{\oint_{\partial D} P\,\mathrm{d}x + Q\,\mathrm{d}y = \iint_D (Q_x - P_y)\,\mathrm{d}A.}$$

左边是边界上的总循环，右边是区域内部每一点局部旋转的总和。

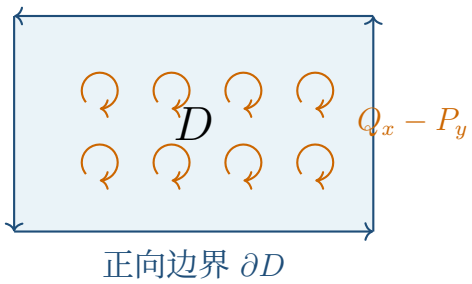



图 1：Green 定理：边界上的总循环等于内部局部旋度的面积积分。

## 3.2 Green 定理的直觉推导：小矩形相互抵消

把区域 $D$ 切成很多小矩形。对每个小矩形来说，沿边界逆时针走一圈的线积分近似等于该点的局部旋度乘以面积：

$$\oint_{\partial R_{ij}} P\,\mathrm{d}x + Q\,\mathrm{d}y \approx (Q_x - P_y)\Delta x \Delta y.$$

为什么把所有小矩形加起来以后，只剩下大区域外边界？因为相邻小矩形共享的内部边界会被走两次，而且方向相反。一次从左到右，另一次从右到左；一次向上，另一次向下。所以内部边界贡献相互抵消，最后只剩下外边界。

**小矩形上的计算骨架**

考虑以 $(x, y)$ 为左下角的小矩形，宽 $\Delta x$，高 $\Delta y$。逆时针绕一圈：

$$\begin{aligned}
\text{下边}: &\int_{x}^{x+\Delta x} P(s, y)\,\mathrm{d}s \approx P(x, y)\Delta x, \\
\text{右边}: &\int_{y}^{y+\Delta y} Q(x+\Delta x, t)\,\mathrm{d}t \approx Q(x+\Delta x, y)\Delta y, \\
\text{上边}: &\int_{x+\Delta x}^{x} P(s, y+\Delta y)\,\mathrm{d}s \approx -P(x, y+\Delta y)\Delta x, \\
\text{左边}: &\int_{y+\Delta y}^{y} Q(x, t)\,\mathrm{d}t \approx -Q(x, y)\Delta y.
\end{aligned}$$

加起来：

$$\begin{aligned}
&[P(x, y) - P(x, y+\Delta y)]\Delta x + [Q(x+\Delta x, y) - Q(x, y)]\Delta y \\
&\approx [-P_y(x, y)\Delta y]\Delta x + [Q_x(x, y)\Delta x]\Delta y \\
&= (Q_x - P_y)\Delta x \Delta y.
\end{aligned}$$

这就是 Green 定理的局部来源。

**直觉**

Green 定理不是魔法。它就是“很多小格子的内部边界互相抵消”。这个抵消思想非常重要，以后你会在守恒方程、离散到连续极限、概率流、有限体积法、积分形式 PDE 中反复见到。

## 3.3 计算例题 1：用 Green 定理算闭合路径积分

**例 1：矩形上的边界积分**

设

$$\boldsymbol{F}(x,y)=(-y,x),$$

即 $P=-y, Q=x$。令 $D=[0,a]\times[0,b]$。求

$$\oint_{\partial D} -y\,\mathrm{d}x + x\,\mathrm{d}y.$$

**解：**直接沿四条边算当然可以，但 Green 定理更快。因为

$$Q_x = 1, \qquad P_y = -1,$$

所以

$$Q_x - P_y = 1-(-1) = 2.$$

因此

$$\oint_{\partial D} -y\,\mathrm{d}x + x\,\mathrm{d}y = \iint_D 2\,\mathrm{d}A = 2ab.$$

**例 2：用闭合路径积分计算面积**

取

$$P=-\frac{y}{2}, \qquad Q=\frac{x}{2}.$$

则

$$Q_x - P_y = \frac{1}{2} - \left(-\frac{1}{2}\right) = 1.$$

所以对任意正向简单闭曲线 $C=\partial D$,

$$\mathrm{Area}(D) = \iint_D 1\,\mathrm{d}A = \oint_C -\frac{y}{2}\,\mathrm{d}x + \frac{x}{2}\,\mathrm{d}y.$$

等价写法是

$$\boxed{\mathrm{Area}(D) = \frac{1}{2}\oint_C x\,\mathrm{d}y - y\,\mathrm{d}x.}$$

这说明面积也可以由边界路径计算。

## 3.4 经济金融解释：闭合调整路径的净收益

假设 $(x,y)$ 是两个政策工具，例如税率和补贴率。某个局部福利变化近似为

$$\mathrm{d}W = P(x,y)\,\mathrm{d}x + Q(x,y)\,\mathrm{d}y.$$

如果这是某个全局福利函数 $W(x,y)$ 的全微分，那么沿任何闭合政策路径走一圈应该回到原福利：

$$\oint_C P\,\mathrm{d}x + Q\,\mathrm{d}y = 0.$$

如果对某个闭合路径非零，则说明局部福利场不可积，政策调整存在路径依赖。Green 定理进一步告诉我们，闭合路径上的净福利循环来自区域内部的

$$Q_x - P_y.$$

**方法**

判断二维局部边际场是否可能没有路径依赖：先看

$$P_y = Q_x.$$

如果区域没有洞且偏导连续，那么这通常说明存在势函数。如果

$$Q_x - P_y \neq 0,$$

则闭合路径积分往往不为零，路径依赖来自局部旋度。

# 4 散度与通量：进入 Gauss 定理之前

Green 定理讨论的是平面循环。Gauss 定理讨论的是流量。

## 4.1 通量：穿过边界的净流动

设 $\boldsymbol{F}(x,y,z)$ 是三维向量场。你可以把它想成一个流速场：在点 $(x,y,z)$，流体、人口、财富、概率质量或风险暴露正以方向和速度 $\boldsymbol{F}$ 移动。

给定一个曲面 $S$，单位法向量是 $\boldsymbol{n}$。穿过曲面的一小块面积 $\mathrm{d}S$ 的流量近似为

$$\boldsymbol{F}\cdot\boldsymbol{n}\,\mathrm{d}S.$$

如果 $\boldsymbol{F}$ 与法向量方向一致，流出为正；相反则为负；如果切着边界走，则点积为零。

**通量 flux**

向量场 $\boldsymbol{F}$ 穿过有向曲面 $S$ 的通量定义为

$$\iint_S \boldsymbol{F}\cdot\boldsymbol{n}\,\mathrm{d}S.$$

如果 $S=\partial V$ 是某个体 $V$ 的封闭边界，通常 $\boldsymbol{n}$ 取外法向量，此时通量表示净流出量。

## 4.2 散度：局部源泉和汇点

**散度 divergence**

对三维向量场

$$\boldsymbol{F} = (F_1, F_2, F_3),$$

散度定义为

$$\operatorname{div} \boldsymbol{F} = \nabla \cdot \boldsymbol{F} = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}.$$

它衡量一个点附近是净流出源还是净流入汇。

如果 $\operatorname{div} \boldsymbol{F} > 0$，该点附近像“源泉”，小球边界上净流出为正。如果 $\operatorname{div} \boldsymbol{F} < 0$，该点附近像“汇点”，净流入为正。如果 $\operatorname{div} \boldsymbol{F} = 0$，则局部没有净创造或消失，只是搬运。

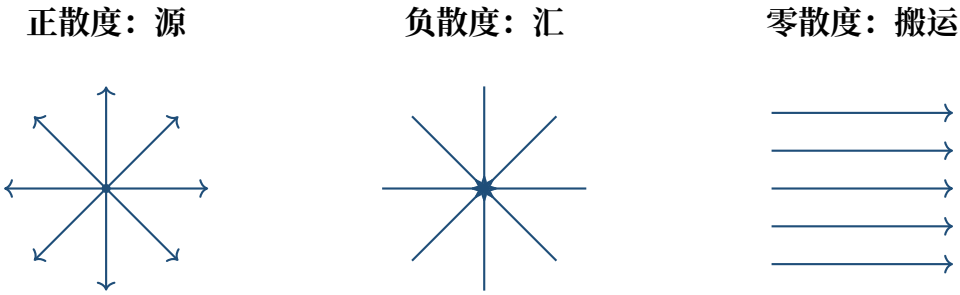



图 2: 散度的直觉：源、汇和无净创造的搬运。

## 4.3 二维散度定理：Gauss 的平面版本

在平面中，如果 $\boldsymbol{F} = (P, Q)$，区域 $D$ 的外法向量为 $\boldsymbol{n}$，则二维散度定理是

$$\boxed{\oint_{\partial D} \boldsymbol{F} \cdot \boldsymbol{n} \, \mathrm{d}s = \iint_D (P_x + Q_y) \, \mathrm{d}A.}$$

左边是穿出平面区域边界的净流量，右边是区域内部散度的总和。

**例 3：二维通量**

设 $\boldsymbol{F} = (x, y)$，$D$ 是半径为 $R$ 的圆盘。求穿出边界圆的通量。

**方法一：直接算**。边界上 $(x, y) = R(\cos t, \sin t)$，外单位法向量

$$\boldsymbol{n} = (\cos t, \sin t), \qquad \mathrm{d}s = R \, \mathrm{d}t.$$

所以

$$\boldsymbol{F} \cdot \boldsymbol{n} = R(\cos t, \sin t) \cdot (\cos t, \sin t) = R.$$

通量为

$$\int_0^{2\pi} R \cdot R \, \mathrm{d}t = 2\pi R^2.$$

**方法二：用散度定理**。因为

$$\nabla \cdot \boldsymbol{F} = P_x + Q_y = 1 + 1 = 2,$$

所以

$$\oint_{\partial D} \boldsymbol{F} \cdot \boldsymbol{n}\, \mathrm{d}s = \iint_D 2\, \mathrm{d}A = 2\pi R^2.$$

# 5 Gauss 定理：封闭曲面通量等于体内散度总和

## 5.1 定理表述

**Gauss 定理 / Divergence Theorem**

设 $V \subset \mathbb{R}^3$ 是一个足够规则的三维体，边界为封闭曲面 $\partial V$，外法向量为 $\boldsymbol{n}$。若 $\boldsymbol{F} = (F_1, F_2, F_3)$ 有连续偏导数，则

$$\boxed{\iint_{\partial V} \boldsymbol{F} \cdot \boldsymbol{n}\, \mathrm{d}S = \iiint_V \nabla \cdot \boldsymbol{F}\, \mathrm{d}V.}$$

左边是穿过封闭边界的净流出量，右边是体内部所有局部源泉强度的总和。

## 5.2 Gauss 定理的直觉推导：小盒子的内部面抵消

把体 $V$ 切成很多小盒子。对每个小盒子，净流出量近似等于散度乘以体积：

$$\text{flux out of small box} \approx (\nabla \cdot \boldsymbol{F})\Delta x \Delta y \Delta z.$$

相邻小盒子的共享面上，一个盒子的流出正好是另一个盒子的流入，所以内部面通量抵消。所有小盒子加起来，最后只剩大体外边界的净通量。

**小盒子上的计算骨架**

只看 $x$ 方向的两个面。左面在 $x$，外法向量约为 $(-1, 0, 0)$，右面在 $x + \Delta x$，外法向量约为 $(1, 0, 0)$。它们贡献近似为

$$-F_1(x, y, z)\Delta y \Delta z + F_1(x + \Delta x, y, z)\Delta y \Delta z.$$

差为

$$[F_1(x + \Delta x, y, z) - F_1(x, y, z)]\Delta y \Delta z \approx F_{1x} \Delta x \Delta y \Delta z.$$

同理，$y$ 方向给出 $F_{2y}\Delta V$，$z$ 方向给出 $F_{3z}\Delta V$。合计为

$$(F_{1x} + F_{2y} + F_{3z})\Delta V = (\nabla \cdot \boldsymbol{F})\Delta V.$$

## 5.3 计算例题 4：球面通量

**例 4：径向场穿出球面的通量**

设

$$\boldsymbol{F}(x,y,z)=(x,y,z),$$

$V$ 是半径为 $R$ 的球体。求

$$\iint_{\partial V}\boldsymbol{F}\cdot\boldsymbol{n}\,\mathrm{d}S.$$

**解：**用 Gauss 定理。因为

$$\nabla\cdot\boldsymbol{F}=1+1+1=3,$$

所以

$$\iint_{\partial V}\boldsymbol{F}\cdot\boldsymbol{n}\,\mathrm{d}S=\iiint_{V}3\,\mathrm{d}V=3\cdot\frac{4}{3}\pi R^3=4\pi R^3.$$

也可以直接算：球面上外法向量 $\boldsymbol{n}=(x,y,z)/R$，而 $\boldsymbol{F}=(x,y,z)$，所以

$$\boldsymbol{F}\cdot\boldsymbol{n}=\frac{x^2+y^2+z^2}{R}=R.$$

球面积为 $4\pi R^2$，通量为

$$R\cdot 4\pi R^2=4\pi R^3.$$

**例 5：常向量场穿过封闭曲面的通量**

设 $\boldsymbol{F}=(a,b,c)$ 是常向量场，$V$ 为任意封闭体。求净通量。

**解：**因为

$$\nabla\cdot\boldsymbol{F}=0+0+0=0,$$

所以

$$\iint_{\partial V}\boldsymbol{F}\cdot\boldsymbol{n}\,\mathrm{d}S=0.$$

直觉上，常向量场只是把东西从一边搬到另一边。流入多少，流出多少，没有内部创造或消失。

## 5.4 经济金融解释：状态空间里的守恒

设 $m(x,t)$ 是状态 $x$ 处的主体密度，例如财富分布、资本分布、生产率分布。若主体以速度场 $v(x,t)$ 在状态空间中移动，那么概率质量流可以写作

$$\boldsymbol{F}(x,t)=m(x,t)v(x,t).$$

对某个状态区域 $D$，区域内总质量变化取决于穿过边界的净流出和内部出生死亡项。一个典型守恒形式是

$$\frac{\partial m}{\partial t} + \nabla \cdot (mv) = s,$$

其中 $s$ 是源项。这个式子背后的积分直觉就是 Gauss 定理：内部散度控制边界净流量。

**直觉**

在连续时间经济模型里，很多 PDE 的核心不是“偏微分很难”，而是一个守恒逻辑：如果某区域的质量减少，要么流出边界，要么内部有负源项。Gauss 定理就是把这个逻辑数学化。

# 6 旋度与 Stokes 定理

Green 定理讨论平面区域的边界循环；Stokes 定理把这个思想推广到空间中的曲面。

## 6.1 三维旋度

**三维旋度 curl**

对三维向量场

$$\boldsymbol{F} = (F_1, F_2, F_3),$$

旋度定义为

$$\nabla \times \boldsymbol{F} = \begin{pmatrix} F_{3y} - F_{2z} \\ F_{1z} - F_{3x} \\ F_{2x} - F_{1y} \end{pmatrix}.$$

它衡量向量场在三维空间中的局部旋转轴和旋转强度。

如果 $\nabla \times \boldsymbol{F}$ 指向某个方向，表示一个小桨轮放在该点时，最强旋转轴沿那个方向。旋度与法向量 $\boldsymbol{n}$ 的点积

$$(\nabla \times \boldsymbol{F}) \cdot \boldsymbol{n}$$

表示沿法向方向观察到的局部旋转强度。

## 6.2 Stokes 定理

**Stokes 定理**

设 $S$ 是一个足够光滑的有向曲面，边界为 $\partial S$。边界方向与曲面法向量满足右手定则。若 $\boldsymbol{F}$ 有连续偏导数，则

$$\boxed{\oint_{\partial S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \iint_S (\nabla \times \boldsymbol{F}) \cdot \boldsymbol{n} \, \mathrm{d}S.}$$

左边是曲面边界上的总循环，右边是曲面内部旋度穿过曲面的通量。

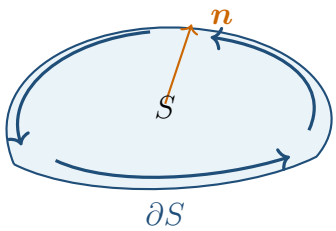



图 3: Stokes 定理：曲面边界循环等于曲面上旋度通量。

## 6.3 Green 定理是 Stokes 定理的特例

如果曲面 $S$ 就是 $xy$ 平面上的区域 $D$，法向量取 $\boldsymbol{n}=(0,0,1)$，且

$$\boldsymbol{F}=(P,Q,0),$$

那么

$$\nabla\times\boldsymbol{F}=(0,0,Q_x-P_y).$$

于是

$$(\nabla\times\boldsymbol{F})\cdot\boldsymbol{n}=Q_x-P_y.$$

Stokes 定理变成

$$\oint_{\partial D}P\,\mathrm{d}x+Q\,\mathrm{d}y=\iint_D(Q_x-P_y)\,\mathrm{d}A,$$

这正是 Green 定理。

**本章压缩总结**

Green 定理不是孤立公式，而是 Stokes 定理在平面上的版本。你可以把 Green 看成二维 Stokes。

## 6.4 计算例题 6：Stokes 比直接算更简单

**例 6：单位圆边界上的循环**

设

$$\boldsymbol{F}(x,y,z)=(-y,x,0),$$

$S$ 是 $xy$ 平面中的单位圆盘，法向量取 $(0,0,1)$。求

$$\oint_{\partial S}\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}.$$

**解：**旋度为

$$\nabla \times \boldsymbol{F} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}.$$

所以

$$(\nabla \times \boldsymbol{F}) \cdot \boldsymbol{n} = 2.$$

因此

$$\oint_{\partial S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \iint_S 2 \, \mathrm{d}S = 2\pi.$$

这和直接参数化单位圆 $\boldsymbol{r}(t) = (\cos t, \sin t, 0)$ 得到的结果一致。

**例 7：边界不变，曲面可换**

Stokes 定理的一个强大含义是：如果两个曲面 $S_1, S_2$ 有同一个边界 $C$，且取向一致，那么

$$\iint_{S_1} (\nabla \times \boldsymbol{F}) \cdot \boldsymbol{n} \, \mathrm{d}S = \iint_{S_2} (\nabla \times \boldsymbol{F}) \cdot \boldsymbol{n} \, \mathrm{d}S = \oint_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}.$$

因此有时可以把一个难曲面换成同边界的简单曲面来算。

## 6.5 经济金融解释：局部无旋和全局可积性

上一章说过，如果一个二维局部边际场是某个势函数的梯度，那么闭合路径积分为零。三维里也有类似思想：如果

$$\boldsymbol{F} = \nabla \phi,$$

则

$$\nabla \times \boldsymbol{F} = \nabla \times (\nabla \phi) = 0.$$

也就是说，梯度场没有旋度。反过来，在合适的区域条件下，如果旋度为零，场可能是保守的。

这对经济金融的直觉是：如果局部价格变化、边际福利变化或局部调仓收益可以来自某个全局价值函数，那么沿闭合路径走一圈不应产生净收益。若存在非零循环，就意味着模型中有摩擦、套利机会、路径依赖或状态变量遗漏。

# 7 三个定理的统一图谱

## 7.1 统一形式

三大定理可以写成同一种语言：

$$\boxed{\int_{\partial\Omega} \text{boundary object} = \int_{\Omega} \text{local derivative object.}}$$

更具体地：

| 定理 | 左边：边界 | 右边：内部 |
|---|---|---|
| Green | $\oint_{\partial D} P\,\mathrm{d}x + Q\,\mathrm{d}y$ | $\iint_D (Q_x - P_y)\,\mathrm{d}A$ |
| Gauss | $\iint_{\partial V} \boldsymbol{F}\cdot\boldsymbol{n}\,\mathrm{d}S$ | $\iiint_V \nabla\cdot\boldsymbol{F}\,\mathrm{d}V$ |
| Stokes | $\oint_{\partial S} \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}$ | $\iint_S (\nabla\times\boldsymbol{F})\cdot\boldsymbol{n}\,\mathrm{d}S$ |

## 7.2 旋度和散度的区别

**本章压缩总结**

- **散度**问：这里是不是在产生或吸收流量？关注净流出。
- **旋度**问：这里是不是在推动局部旋转？关注循环。
- **通量**问：有多少东西穿过边界？
- **循环**问：沿边界走一圈积累了多少切向分量？

**常见误区**

不要混淆两类积分：

$$\oint_C \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}$$

是沿曲线的切向积分，衡量循环；

$$\iint_S \boldsymbol{F}\cdot\boldsymbol{n}\,\mathrm{d}S$$

是穿过曲面的法向积分，衡量通量。一个看“沿着边界走”，一个看“穿过边界流”。

# 8 经济金融中的三个模型化入口

## 8.1 入口一：守恒方程和分布动态

设 $m(x,t)$ 是某种经济主体密度，$v(x,t)$ 是状态移动速度。若无出生死亡，质量守恒可以写成

$$\frac{\partial m}{\partial t} + \nabla\cdot(mv) = 0.$$

这句话的积分含义是：某区域内质量的下降，等于穿过边界流出去的净质量。Gauss 定理把“边界流出”转换成“内部散度”。

这类结构会出现在：

- heterogeneous-agent macro；
- wealth distribution dynamics；
- firm size distribution；
- probability density evolution；
- optimal transport 和匹配模型的初步直觉。

## 8.2 入口二：无套利和闭合交易路径

假设 $\boldsymbol{F}$ 表示某种局部价格收益场。若沿闭合交易路径 $C$ 有

$$\oint_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} \neq 0,$$

那么从同一状态出发又回到同一状态，却获得非零收益或成本。这在无摩擦模型里通常不应出现。于是无套利常常要求某些闭合路径积分为零，或者要求某个价格场可以写成价值函数的梯度。

这不是说金融课里会直接写 Green 定理，而是说本章给你一个几何语言：**无套利本质上常常排除闭合路径上的非零循环**。

## 8.3 入口三：PDE 与连续时间资产定价

连续时间资产定价、随机控制和宏观动态中经常出现 PDE。很多 PDE 都能被理解成局部守恒、局部均衡或局部无套利条件。

例如，一个状态变量 $x$ 的价值函数 $V(x,t)$ 常常满足类似

$$\rho V = u + \mathcal{L}V,$$

其中 $\mathcal{L}$ 是描述状态局部漂移和扩散的算子。本章不会进入随机微积分，但 Green/Gauss/Stokes 训练的正是你理解这类方程所需的“局部到整体”直觉。

# 9 Worked Examples：完整样板题

**样板题 1：直接用 Green 定理避免分段积分**

计算

$$\oint_C (x^2 - y)\,\mathrm{d}x + (x + y^2)\,\mathrm{d}y,$$

其中 $C$ 是三角形 $(0,0),(1,0),(0,1)$ 的正向边界。

**解：** 令

$$P = x^2 - y, \qquad Q = x + y^2.$$

则

$$Q_x = 1, \qquad P_y = -1.$$

所以

$$Q_x - P_y = 2.$$

由 Green 定理，积分等于

$$\iint_D 2 \,\mathrm{d}A = 2\,\mathrm{Area}(D).$$

三角形面积为 $1/2$，所以结果为

$$\boxed{1}.$$

**样板题 2：二维通量和散度定理**

设 $\boldsymbol{F} = (x^2, y^2)$，$D = [0,1] \times [0,2]$。求穿出 $\partial D$ 的通量

$$\oint_{\partial D} \boldsymbol{F} \cdot \boldsymbol{n} \,\mathrm{d}s.$$

**解：** 用二维散度定理：

$$\nabla \cdot \boldsymbol{F} = 2x + 2y.$$

所以

$$\begin{aligned} \oint_{\partial D} \boldsymbol{F} \cdot \boldsymbol{n} \,\mathrm{d}s &= \iint_D (2x + 2y) \,\mathrm{d}A \\ &= \int_0^1 \int_0^2 (2x + 2y) \,\mathrm{d}y \,\mathrm{d}x. \end{aligned}$$

先对 $y$ 积分：

$$\int_0^2 (2x + 2y) \,\mathrm{d}y = 4x + 4.$$

再对 $x$ 积分：

$$\int_0^1 (4x + 4) \,\mathrm{d}x = 2 + 4 = 6.$$

因此通量为

$$\boxed{6}.$$

### 样板题 3：Gauss 定理算盒子通量

设 $\boldsymbol{F}=(x,y,z^2)$，$V=[0,1]\times[0,1]\times[0,2]$。求外向通量。

**解：** 散度为

$$\nabla\cdot\boldsymbol{F}=1+1+2z=2+2z.$$

由 Gauss 定理，

$$\begin{aligned}\iint_{\partial V}\boldsymbol{F}\cdot\boldsymbol{n}\,\mathrm{d}S&=\iiint_V(2+2z)\,\mathrm{d}V\\&=\int_0^1\int_0^1\int_0^2(2+2z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x.\end{aligned}$$

内层积分：

$$\int_0^2(2+2z)\,\mathrm{d}z=4+4=8.$$

其余两个方向长度都是 1，所以结果为

$$\boxed{8}.$$

### 样板题 4：Stokes 定理把空间曲线换成平面圆盘

设

$$\boldsymbol{F}=(-y,x,z),$$

$C$ 是平面 $z=1$ 上半径为 2 的圆

$$x^2+y^2=4,$$

方向从上往下看为逆时针。求

$$\oint_C\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}.$$

**解：** 选取曲面 $S$ 为平面 $z=1$ 内的圆盘，法向量 $\boldsymbol{n}=(0,0,1)$。

计算旋度：

$$\nabla\times\boldsymbol{F}=\begin{pmatrix}F_{3y}-F_{2z}\\F_{1z}-F_{3x}\\F_{2x}-F_{1y}\end{pmatrix}=\begin{pmatrix}0-0\\0-0\\1-(-1)\end{pmatrix}=\begin{pmatrix}0\\0\\2\end{pmatrix}.$$

所以

$$(\nabla\times\boldsymbol{F})\cdot\boldsymbol{n}=2.$$

圆盘面积为 $4\pi$，因此

$$\oint_C\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}=\iint_S2\,\mathrm{d}S=8\pi.$$

# 10 练习题

## A 组：基础计算

A1. 设 $P = x - y, Q = x + y$。计算 $Q_x - P_y$。

A2. 设 $\boldsymbol{F} = (x^2, y^2, z^2)$。计算 $\nabla \cdot \boldsymbol{F}$。

A3. 设 $\boldsymbol{F} = (-y, x, 0)$。计算 $\nabla \times \boldsymbol{F}$。

A4. 用 Green 定理计算 $\oint_C -y\,\mathrm{d}x + x\,\mathrm{d}y$，其中 $C$ 是单位正方形 $[0,1]^2$ 的正向边界。

A5. 用二维散度定理计算 $\boldsymbol{F} = (x, y)$ 穿出单位圆边界的通量。

## B 组：概念判断

B1. Green 定理左边为什么必须注意边界方向？如果方向反了，结果如何变化？

B2. 散度为零是否意味着向量场处处为零？请解释。

B3. 旋度为零是否一定意味着闭合路径积分为零？需要什么区域条件？

B4. 通量积分和循环积分分别使用向量场的哪个方向分量？

B5. 为什么 Gauss 定理中封闭曲面通常取外法向量？

## C 组：混合题

C1. 设 $C$ 为单位圆正向边界。计算

$$\oint_C (x^2 - y)\,\mathrm{d}x + (x + y^2)\,\mathrm{d}y.$$

C2. 设 $D = [0,2] \times [0,1]$，计算

$$\oint_{\partial D} y\,\mathrm{d}x + x^2\,\mathrm{d}y.$$

C3. 设 $\boldsymbol{F} = (xy, yz, zx)$，计算 $\nabla \cdot \boldsymbol{F}$。

C4. 设 $\boldsymbol{F} = (y, z, x)$，计算 $\nabla \times \boldsymbol{F}$。

C5. 设 $\boldsymbol{F} = (x, y, z)$，$V$ 为单位立方体 $[0,1]^3$。用 Gauss 定理求外向通量。

## D 组：经济金融应用题

D1. 设局部福利变化为

$$\mathrm{d}W = (2x + y)\,\mathrm{d}x + (x + 3y)\,\mathrm{d}y.$$

判断该局部福利场是否存在路径依赖。

D2. 设政策场为

$$\mathrm{d}W = (-y)\,\mathrm{d}x + x\,\mathrm{d}y.$$

对单位圆闭合路径计算净福利循环，并解释含义。

D3. 某状态空间中的概率质量流为 $\boldsymbol{F} = (x, y)$。单位圆盘内散度总和是多少？解释它表示什么。

D4. 若财富流场 $\boldsymbol{F}$ 满足 $\nabla \cdot \boldsymbol{F} = 0$，对任意封闭区域意味着什么?

D5. 如果一个价格变化场沿闭合交易路径积分不为零，从无套利角度应如何理解?

## E 组：挑战/证明题

E1. 用小矩形思想推导 Green 定理中 $Q_x - P_y$ 的来源。

E2. 用小盒子思想推导 Gauss 定理中 $\nabla \cdot \boldsymbol{F}$ 的来源。

E3. 证明如果 $\boldsymbol{F} = \nabla\phi$，则 $\nabla \times \boldsymbol{F} = 0$。

E4. 说明为什么 Green 定理可以看成 Stokes 定理的特例。

E5. 设 $\nabla \cdot \boldsymbol{F} = c$ 是常数，$V$ 体积为 $|V|$。用 Gauss 定理解释为什么净外向通量为 $c|V|$。

# 11 练习题详细解答

## A 组答案

### A1

$$P = x - y, \qquad Q = x + y.$$

所以

$$Q_x = 1, \qquad P_y = -1.$$

因此

$$Q_x - P_y = 2.$$

### A2

$$\boldsymbol{F} = (x^2, y^2, z^2).$$

所以

$$\nabla \cdot \boldsymbol{F} = 2x + 2y + 2z.$$

### A3

$$\boldsymbol{F} = (-y, x, 0).$$

按公式:

$$\nabla \times \boldsymbol{F} = \begin{pmatrix} 0_y - x_z \\ (-y)_z - 0_x \\ x_x - (-y)_y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}.$$

### A4

这里 $P = -y, Q = x$，所以

$$Q_x - P_y = 1 - (-1) = 2.$$

单位正方形面积为 1，所以

$$\oint_C -y\,\mathrm{d}x + x\,\mathrm{d}y = \iint_D 2\,\mathrm{d}A = 2.$$

**A5**

二维散度为

$$\nabla \cdot \boldsymbol{F} = 1 + 1 = 2.$$

单位圆盘面积为 $\pi$，所以通量为

$$\iint_D 2 \,\mathrm{d}A = 2\pi.$$

## B 组答案

**B1**

Green 定理默认边界正向，即逆时针方向。若方向反过来，线积分变号：

$$\oint_{-\partial D} P \,\mathrm{d}x + Q \,\mathrm{d}y = -\oint_{\partial D} P \,\mathrm{d}x + Q \,\mathrm{d}y.$$

右边面积积分不自动变号，所以使用公式时必须配合正向边界。

**B2**

不意味着。散度为零只说明局部没有净源泉或汇点，不说明没有流动。例如常向量场 $\boldsymbol{F} = (1, 0, 0)$ 散度为零，但场不为零。它表示均匀搬运。

**B3**

不一定。若区域有洞，旋度为零仍可能存在非零闭合路径积分。典型例子是

$$\boldsymbol{F} = \left( -\frac{y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right)$$

在去掉原点的区域上旋度为零，但沿单位圆积分为 $2\pi$。通常还需要区域单连通等条件。

**B4**

循环积分

$$\oint_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$$

使用向量场沿路径切向方向的分量。通量积分

$$\iint_S \boldsymbol{F} \cdot \boldsymbol{n} \,\mathrm{d}S$$

使用向量场沿曲面法向方向的分量。

**B5**

因为 Gauss 定理解释的是封闭体的净流出量。外法向量使穿出体外的流为正，流入体内的流为负。这样左边才能自然对应右边内部散度的总和。

## C 组答案

**C1**

$$P = x^2 - y, \qquad Q = x + y^2.$$

所以

$$Q_x = 1, \qquad P_y = -1, \qquad Q_x - P_y = 2.$$

单位圆盘面积为 $\pi$，所以

$$\oint_C (x^2 - y)\,\mathrm{d}x + (x + y^2)\,\mathrm{d}y = 2\pi.$$

**C2**

$$P = y, \qquad Q = x^2.$$

所以

$$Q_x - P_y = 2x - 1.$$

于是

$$\begin{aligned}\oint_{\partial D} y\,\mathrm{d}x + x^2\,\mathrm{d}y &= \int_0^2 \int_0^1 (2x - 1)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_0^2 (2x - 1)\,\mathrm{d}x \\ &= [x^2 - x]_0^2 = 2.\end{aligned}$$

**C3**

$$\boldsymbol{F} = (xy, yz, zx).$$

散度为

$$\nabla \cdot \boldsymbol{F} = \frac{\partial(xy)}{\partial x} + \frac{\partial(yz)}{\partial y} + \frac{\partial(zx)}{\partial z} = y + z + x.$$

**C4**

$$\boldsymbol{F} = (y, z, x).$$

于是

$$\nabla \times \boldsymbol{F} = \begin{pmatrix} F_{3y} - F_{2z} \\ F_{1z} - F_{3x} \\ F_{2x} - F_{1y} \end{pmatrix} = \begin{pmatrix} 0 - 1 \\ 0 - 1 \\ 0 - 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}.$$

**C5**

$$\boldsymbol{F} = (x, y, z), \qquad \nabla \cdot \boldsymbol{F} = 3.$$

单位立方体体积为 1，所以外向通量为

$$\iiint_V 3 \, \mathrm{d}V = 3.$$

## D 组答案

**D1**

这里

$$P = 2x + y, \qquad Q = x + 3y.$$

计算：

$$P_y = 1, \qquad Q_x = 1.$$

所以 $P_y = Q_x$。在没有洞的区域上，该局部福利场通常可以整合成全局福利函数，没有路径依赖。事实上势函数可取

$$W = x^2 + xy + \frac{3}{2}y^2.$$

**D2**

$$P = -y, \qquad Q = x.$$

所以

$$Q_x - P_y = 1 - (-1) = 2.$$

单位圆盘面积为 $\pi$，净福利循环为

$$\oint_C -y \, \mathrm{d}x + x \, \mathrm{d}y = 2\pi.$$

经济含义是：沿闭合政策路径走一圈回到同一政策点，却产生非零净福利变化。这说明该局部福利场不能来自单值全局福利函数，存在路径依赖或遗漏状态变量。

**D3**

$$\boldsymbol{F} = (x, y), \qquad \nabla \cdot \boldsymbol{F} = 2.$$

单位圆盘内散度总和为

$$\iint_D 2 \, \mathrm{d}A = 2\pi.$$

它表示该区域内部像一个整体源泉，产生了 $2\pi$ 的净外向流量。若这是概率质量流，则必须有源项或边界外流来解释质量变化。

**D4**

若 $\nabla \cdot \boldsymbol{F} = 0$，则对任意封闭区域 $V$，Gauss 定理给出

$$\iint_{\partial V} \boldsymbol{F} \cdot \boldsymbol{n} \, \mathrm{d}S = 0.$$

这表示流入量等于流出量，区域内部没有净创造或消失。经济上可以解释为财富、人口、概率质量只是重新分配或搬运，而非在内部净生成。

**D5**

如果沿闭合交易路径积分不为零，意味着从同一资产配置或同一状态出发，经过一系列交易又回到原状态，却得到非零净收益或成本。在无摩擦无套利模型中，这通常不应存在。因此它提示可能有套利机会、交易摩擦、价格冲击、路径依赖或状态变量没有被完整建模。

## E 组答案

**E1**

对小矩形，四边逆时针积分合计近似为

$$[P(x, y) - P(x, y + \Delta y)]\Delta x + [Q(x + \Delta x, y) - Q(x, y)]\Delta y.$$

用一阶近似：

$$P(x, y + \Delta y) - P(x, y) \approx P_y \Delta y,$$

$$Q(x + \Delta x, y) - Q(x, y) \approx Q_x \Delta x.$$

因此小矩形边界循环近似为

$$(Q_x - P_y)\Delta x \Delta y.$$

把所有小矩形加总，内部边界相互抵消，只剩外边界，于是得到 Green 定理。

**E2**

对小盒子，$x$ 方向的净流出近似为

$$F_{1x}\Delta x\Delta y\Delta z.$$

$y$ 方向净流出近似为

$$F_{2y}\Delta x\Delta y\Delta z.$$

$z$ 方向净流出近似为

$$F_{3z}\Delta x\Delta y\Delta z.$$

合计为

$$(F_{1x}+F_{2y}+F_{3z})\Delta V=(\nabla\cdot\boldsymbol{F})\Delta V.$$

把小盒子加总，内部公共面通量相互抵消，剩下外边界通量，于是得到 Gauss 定理。

**E3**

设 $\boldsymbol{F}=\nabla\phi=(\phi_x,\phi_y,\phi_z)$。则

$$\nabla\times\boldsymbol{F}=\begin{pmatrix}\phi_{zy}-\phi_{yz}\\ \phi_{xz}-\phi_{zx}\\ \phi_{yx}-\phi_{xy}\end{pmatrix}.$$

若 $\phi$ 二阶连续可微，则混合偏导相等，所以每一项都是 0。因此

$$\nabla\times(\nabla\phi)=0.$$

**E4**

令曲面 $S$ 是 $xy$ 平面上的区域 $D$，法向量取 $(0,0,1)$。令三维向量场为

$$\boldsymbol{F}=(P,Q,0).$$

则

$$\nabla\times\boldsymbol{F}=(0,0,Q_x-P_y).$$

因此 Stokes 定理右边为

$$\iint_D(Q_x-P_y)\,\mathrm{d}A,$$

左边为

$$\oint_{\partial D}P\,\mathrm{d}x+Q\,\mathrm{d}y.$$

这正是 Green 定理。

**E5**

若 $\nabla \cdot \boldsymbol{F} = c$，则由 Gauss 定理

$$\iint_{\partial V} \boldsymbol{F} \cdot \boldsymbol{n} \, \mathrm{d}S = \iiint_{V} c \, \mathrm{d}V = c|V|.$$

这表示如果每单位体积都有相同的净源泉强度 $c$，那么整个区域的净外向通量就是源泉强度乘以体积。

# 12 本章总结图谱

**本章压缩总结**

| 概念 | 要点 |
| --- | --- |
| **Green 定理** | $\oint_{\partial D} P \, \mathrm{d}x + Q \, \mathrm{d}y = \iint_D (Q_x - P_y) \, \mathrm{d}A$。边界循环等于内部旋度总和。 |
| **二维散度定理** | $\oint_{\partial D} \boldsymbol{F} \cdot \boldsymbol{n} \, \mathrm{d}s = \iint_D \nabla \cdot \boldsymbol{F} \, \mathrm{d}A$。边界净流出等于内部散度总和。 |
| **Gauss 定理** | $\iint_{\partial V} \boldsymbol{F} \cdot \boldsymbol{n} \, \mathrm{d}S = \iiint_V \nabla \cdot \boldsymbol{F} \, \mathrm{d}V$。封闭曲面通量等于体内源泉总和。 |
| **Stokes 定理** | $\oint_{\partial S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \iint_S (\nabla \times \boldsymbol{F}) \cdot \boldsymbol{n} \, \mathrm{d}S$。曲面边界循环等于曲面旋度通量。 |
| **散度** | 衡量局部净流出或净创造。正散度像源，负散度像汇。 |
| **旋度** | 衡量局部旋转趋势。二维中是 $Q_x - P_y$，三维中是 $\nabla \times \boldsymbol{F}$。 |
| **通量** | 穿过边界的法向流量。用于守恒、分布动态、状态空间流。 |
| **循环** | 沿边界的切向累积。用于路径依赖、无套利、可积性。 |
| **经济含义** | 局部条件如何推出整体约束；边界流量如何反映内部源项；闭合路径积分如何揭示路径依赖或套利结构。 |

**直觉**

本章最重要的一句话：**边界不是孤立的，边界上的总循环或总流量，往往是内部局部旋转、源泉和约束累加出来的结果。**

## 下一章连接

下一章进入 Part IV：傅里叶级数、投影与函数空间。我们会从向量分析转向另一个对经济金融极其重要的思想：**把一个对象投影到一个更简单的空间上**。

这会直接连接：

- OLS 回归的几何解释；
- 正交投影；
- 函数空间中的最小二乘逼近；
- Fourier 展开；
- 因子模型和非参数回归的基础。

前几章的积分、路径和局部到整体关系，是连续模型的语言；下一章的投影和正交，是计量与金融建模的语言。

# 经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

---

# Chapter 8

## 正交、投影与最小二乘

---

本章把三个看似分开的对象统一起来：向量空间中的投影、回归中的最小二乘、函数空间中的 Fourier 展开。它们本质上都是：在一个可选空间里，找离目标最近的近似对象。

math4econ personalized notes

2026

# 目录

**Chapter 8 总览** **3**

**1 本章为什么重要** **4**

**2 最小前置知识** **4**

**3 内积、范数与正交** **5**
3.1 内积：把“相似程度”变成数字 . . . . . 5
3.2 范数：用内积定义距离 . . . . . 6
3.3 正交：信息已经用尽的几何表达 . . . . . 6

**4 一维投影：从最简单的几何开始** **7**
4.1 推导投影系数 . . . . . 7
4.2 金融解释：单因子 beta 是投影系数 . . . . . 8

**5 子空间投影与最小二乘** **9**
5.1 Normal equations 的推导 . . . . . 9
5.2 投影定理 . . . . . 10

**6 投影矩阵：把几何变成线性算子** **11**
6.1 投影矩阵的性质 . . . . . 11
6.2 残差生成矩阵 . . . . . 12

**7 OLS 的经济含义：不是公式，而是正交条件** **12**
7.1 样本回归的正交条件 . . . . . 12
7.2 Population projection . . . . . 13
7.3 条件期望：更强的投影对象 . . . . . 14

**8 Worked Examples：完整样板题** **14**

**9 函数空间投影：从向量到函数** **16**
9.1 函数投影的一般形式 . . . . . 17
9.2 若基函数两两正交，系数会大大简化 . . . . . 17

**10 Fourier 展开的入口：正交基函数** **19**

**11 从投影到计量：OLS、矩条件与识别** **21**
11.1 OLS 作为样本矩条件 . . . . . 21
11.2 工具变量和 GMM 的影子 . . . . . 21

**12 从投影到金融：因子模型与风险分解** **22**
12.1 多因子模型 . . . . . 22
12.2 风险分解的几何含义 . . . . . 22

**13 非参数回归和机器学习的入口：基函数展开** **23**

**14 常见错误与概念辨析** **24**

**15 本章概念图谱** **24**

**16 练习题** **24**
A 组：基础计算 . . . . . 24
B 组：概念判断 . . . . . 25
C 组：混合题 . . . . . 25
D 组：经济金融建模题 . . . . . 25
E 组：挑战/证明题 . . . . . 26

**17 练习题参考答案** **27**

**18 本章总结图谱：概念–公式–应用–错误** **31**

# Chapter 8 总览

**本章目标**

学完本章，你应该达到以下目标：

1. 理解内积、范数、正交、投影的几何意义；
2. 能推导向量投影公式和子空间投影公式；
3. 能把最小二乘 OLS 看成把 $\boldsymbol{y}$ 投影到 $\boldsymbol{X}$ 的列空间 $\mathrm{Col}(\boldsymbol{X})$ 上；
4. 能解释 normal equations $\boldsymbol{X}'\hat{\boldsymbol{\varepsilon}} = 0$ 的含义：残差和所有解释变量正交；
5. 会推导 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$，并知道什么时候不能这样写；
6. 理解投影矩阵 $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ 和残差矩阵 $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}$ 的性质；
7. 能把 population regression 写成 $\mathbb{E}[X(Y - X'\beta)] = 0$ 的正交条件；
8. 理解函数空间投影：Fourier 展开、非参数回归、基函数近似，本质上仍是最小平方近似；
9. 能把经济金融里的 beta、因子模型、风险分解、矩条件和投影语言联系起来。

**直觉**

一句话：**投影就是“在我允许使用的解释空间里，找一个最接近目标的影子”**。
如果目标是一个向量 $\boldsymbol{y}$，解释空间是 $\mathrm{span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$，那么最小二乘就是找

$$\hat{\boldsymbol{y}} = \hat{\beta}_1\boldsymbol{x}_1 + \cdots + \hat{\beta}_k\boldsymbol{x}_k$$

使得 $\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2$ 最小。最优时，误差 $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \hat{\boldsymbol{y}}$ 必须和整个解释空间垂直。回归里最核心的正交条件就是从这里来的。

**本章压缩总结**

本章主线可以压缩成：

$$\boxed{\text{内积} \Rightarrow \text{距离} \Rightarrow \text{正交} \Rightarrow \text{投影} \Rightarrow \text{最小二乘} \Rightarrow \text{回归/傅里叶/因子模型}}$$

它不是纯线性代数章节，而是计量经济学、金融因子模型和函数逼近的共同数学入口。

# 1 本章为什么重要

在前面的章节里，我们已经学习了积分、变量替换、向量场和曲线积分。这些工具大多回答这样的问题：**如何累加、如何变化、如何沿路径计算？**从本章开始，我们进入另一类非常核心的问题：**如何近似？**

经济金融中几乎所有实证和量化问题都离不开近似：

- 收入 $Y$ 不能被教育年限 $S$ 完全解释，但我们想找一个线性近似；
- 股票收益 $R_i$ 不能被市场收益 $R_m$ 完全解释，但 CAPM 用一个 beta 描述系统性风险；
- 一个复杂函数 $f(x)$ 不能被有限几个简单函数完全表示，但我们可以用多项式、样条或者 Fourier 基函数逼近；
- 一个高维收益矩阵不能被少数几个因子完全解释，但因子模型希望用几个共同风险因子近似。

这些问题的共同结构是：

$$\text{复杂对象} \approx \text{简单对象的线性组合}.$$

如果我们用平方误差衡量近似好坏，那么最优近似就是投影。于是本章就是在给你建立一套统一语言：

$$\text{OLS} = \text{投影}, \qquad \text{Fourier 展开} = \text{函数空间投影}, \qquad \text{因子模型} = \text{收益向量在因子空间中的投影}.$$

**常见误区**

很多同学第一次学回归时，会把 OLS 记成一个公式：

$$\hat{\beta} = (X'X)^{-1}X'y.$$

这非常危险。公式只是结果，真正重要的是三个问题：

1. 为什么要最小化平方误差?
2. 为什么一阶条件等价于残差和解释变量正交?
3. 为什么这个几何结构可以推广到条件期望、GMM、Fourier 和因子模型?

本章重点不是背公式，而是让你以后看到 $E[X\varepsilon] = 0$、moment condition、orthogonality、projection 时知道它们其实是一条线。

# 2 最小前置知识

你需要回忆以下内容：

1. 向量点积：若 $\boldsymbol{u}=(u_1,\dots,u_n)'$、$\boldsymbol{v}=(v_1,\dots,v_n)'$，则

$$\boldsymbol{u}'\boldsymbol{v}=\sum_{i=1}^{n}u_iv_i.$$

2. 矩阵乘法：$\boldsymbol{X}'\boldsymbol{X}$、$\boldsymbol{X}'\boldsymbol{y}$ 的维度要会判断。

3. 一阶最优化：最小化二次函数时，对参数求导并令其为零。

4. 平方误差：$\|\boldsymbol{y}-\hat{\boldsymbol{y}}\|^2=(\boldsymbol{y}-\hat{\boldsymbol{y}})'(\boldsymbol{y}-\hat{\boldsymbol{y}})$。

5. 基本积分：函数空间内积常写成 $\int f(x)g(x)\,\mathrm{d}x$。

**方法**

本章你要反复检查维度。比如 $\boldsymbol{X}\in\mathbb{R}^{n\times k}$、$\boldsymbol{y}\in\mathbb{R}^n$、$\boldsymbol{\beta}\in\mathbb{R}^k$，那么

$$\boldsymbol{X}\boldsymbol{\beta}\in\mathbb{R}^n,\qquad \boldsymbol{X}'\boldsymbol{X}\in\mathbb{R}^{k\times k},\qquad \boldsymbol{X}'\boldsymbol{y}\in\mathbb{R}^k.$$

维度检查是避免回归推导出错的最有效方法。

# 3 内积、范数与正交

## 3.1 内积：把“相似程度”变成数字

**定义 8.1：欧氏内积**

在 $\mathbb{R}^n$ 中，两个向量 $\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^n$ 的标准内积定义为

$$\langle\boldsymbol{u},\boldsymbol{v}\rangle=\boldsymbol{u}'\boldsymbol{v}=\sum_{i=1}^{n}u_iv_i.$$

如果 $\langle\boldsymbol{u},\boldsymbol{v}\rangle>0$，说明两个向量大体同向；如果 $\langle\boldsymbol{u},\boldsymbol{v}\rangle<0$，说明大体反向；如果 $\langle\boldsymbol{u},\boldsymbol{v}\rangle=0$，说明二者正交。

内积的几何解释来自公式：

$$\langle\boldsymbol{u},\boldsymbol{v}\rangle=\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\cos\theta,$$

其中 $\theta$ 是两个向量之间的夹角。因此 $\langle\boldsymbol{u},\boldsymbol{v}\rangle=0$ 等价于 $\cos\theta=0$，也就是夹角为 $90^\circ$。

**直觉**

内积不是简单的乘法，它是在问：**一个向量在另一个向量方向上有多少成分？**
这和经济金融中的“暴露 exposure”很像。比如一个资产收益向量和市场因子收益向量内积很大，说明这个资产和市场因子同涨同跌程度较高；如果正交，说明它

在线性意义上不暴露于这个因子。

## 3.2 范数：用内积定义距离

**定义 8.2：范数与距离**

向量 $\boldsymbol{u} \in \mathbb{R}^n$ 的欧氏范数定义为

$$\|\boldsymbol{u}\| = \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle} = \sqrt{\sum_{i=1}^{n} u_i^2}.$$

两个向量之间的距离定义为

$$d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\|.$$

当我们说“找最接近 $\boldsymbol{y}$ 的线性组合”时，默认就是用平方距离

$$\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2$$

来衡量误差。

平方误差有两个优势：第一，数学上可微且凸；第二，在几何上可以由勾股定理分解。最小二乘的“二乘”就是平方误差。

## 3.3 正交：信息已经用尽的几何表达

**定义 8.3：正交**

若 $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$，则称 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 正交，记为 $\boldsymbol{u} \perp \boldsymbol{v}$。若一个向量 $\boldsymbol{e}$ 与子空间 $S$ 中所有向量都正交，则记为

$$\boldsymbol{e} \perp S.$$

正交在回归中的含义非常重要。假设 $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ 是回归残差。如果

$$\boldsymbol{X}'\hat{\boldsymbol{\varepsilon}} = 0,$$

那么它表示：残差和 $\boldsymbol{X}$ 的每一列都正交。也就是说，在解释变量所张成的线性空间中，已经没有可以进一步解释残差的线性信息。

**常见误区**

“残差和解释变量正交”不是说残差为零。OLS 通常不能完美拟合 $\boldsymbol{y}$，但它要求未解释部分和解释空间没有线性相关。直觉是：如果残差还和某个解释变量相关，那么你还可以沿那个方向调整预测值，从而降低平方误差；这就不是最优。

# 4 一维投影：从最简单的几何开始

先从最简单的问题开始。给定非零向量 $\boldsymbol{x} \in \mathbb{R}^n$，我们想用 $\boldsymbol{x}$ 的倍数 $b\boldsymbol{x}$ 去近似 $\boldsymbol{y}$：

$$\min_{b\in\mathbb{R}} \|\boldsymbol{y} - b\boldsymbol{x}\|^2 .$$

这就是把 $\boldsymbol{y}$ 投影到直线 $\mathrm{span}\{\boldsymbol{x}\}$ 上。

## 4.1 推导投影系数

目标函数为

$$Q(b) = \|\boldsymbol{y} - b\boldsymbol{x}\|^2 = (\boldsymbol{y} - b\boldsymbol{x})'(\boldsymbol{y} - b\boldsymbol{x}).$$

展开：

$$Q(b) = \boldsymbol{y}'\boldsymbol{y} - 2b\boldsymbol{x}'\boldsymbol{y} + b^2\boldsymbol{x}'\boldsymbol{x}.$$

对 $b$ 求导：

$$Q'(b) = -2\boldsymbol{x}'\boldsymbol{y} + 2b\boldsymbol{x}'\boldsymbol{x}.$$

令 $Q'(b) = 0$，得到

$$\hat{b} = \frac{\boldsymbol{x}'\boldsymbol{y}}{\boldsymbol{x}'\boldsymbol{x}}.$$

于是投影向量为

$$\mathrm{Proj}_{\boldsymbol{x}}(\boldsymbol{y}) = \hat{b}\boldsymbol{x} = \frac{\boldsymbol{x}'\boldsymbol{y}}{\boldsymbol{x}'\boldsymbol{x}}\boldsymbol{x}.$$

**证明骨架：投影误差为什么垂直于 $x$**

令残差 $\hat{\boldsymbol{e}} = \boldsymbol{y} - \hat{b}\boldsymbol{x}$。由 $\hat{b} = \frac{\boldsymbol{x}'\boldsymbol{y}}{\boldsymbol{x}'\boldsymbol{x}}$，有

$$\boldsymbol{x}'\hat{\boldsymbol{e}} = \boldsymbol{x}'(\boldsymbol{y} - \hat{b}\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{y} - \hat{b}\boldsymbol{x}'\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{y} - \frac{\boldsymbol{x}'\boldsymbol{y}}{\boldsymbol{x}'\boldsymbol{x}}\boldsymbol{x}'\boldsymbol{x} = 0.$$

所以最优残差和解释方向 $\boldsymbol{x}$ 正交。

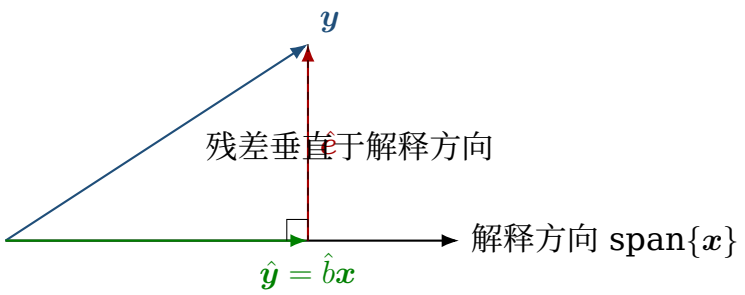



图 1：一维投影：最接近 $\boldsymbol{y}$ 的点是垂足，误差与解释方向正交。

**例 8.1：把一个向量投影到另一向量上**

设

$$\boldsymbol{y} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \qquad \boldsymbol{x} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

求 $\boldsymbol{y}$ 在 $\mathrm{span}\{\boldsymbol{x}\}$ 上的投影。

**解：**

$$\boldsymbol{x}'\boldsymbol{y} = 1 \cdot 2 + 1 \cdot 3 + 0 \cdot 1 = 5, \qquad \boldsymbol{x}'\boldsymbol{x} = 1^2 + 1^2 + 0^2 = 2.$$

所以

$$\hat{b} = \frac{5}{2}, \qquad \hat{\boldsymbol{y}} = \hat{b}\boldsymbol{x} = \begin{pmatrix} 5/2 \\ 5/2 \\ 0 \end{pmatrix}.$$

残差为

$$\hat{\boldsymbol{e}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \begin{pmatrix} -1/2 \\ 1/2 \\ 1 \end{pmatrix}.$$

检查正交：

$$\boldsymbol{x}'\hat{\boldsymbol{e}} = 1 \cdot (-1/2) + 1 \cdot (1/2) + 0 \cdot 1 = 0.$$

## 4.2 金融解释：单因子 beta 是投影系数

设资产超额收益为 $R_i$，市场超额收益为 $R_m$。CAPM 或单因子模型中常写：

$$R_i = \alpha + \beta R_m + \varepsilon.$$

如果先忽略截距，$\beta$ 的 population projection 系数就是

$$\beta = \frac{\mathbb{E}[R_m R_i]}{\mathbb{E}[R_m^2]}.$$

如果 $R_m$ 和 $R_i$ 都是均值为零的超额收益，则

$$\beta = \frac{\mathrm{Cov}(R_i, R_m)}{\mathrm{Var}(R_m)}.$$

这和刚才一维投影公式完全一样，只是把有限样本内积 $\boldsymbol{x}'\boldsymbol{y}$ 换成了概率内积 $\mathbb{E}[XY]$。

**直觉**

金融 beta 的几何含义：**资产收益在市场因子方向上的投影强度**。
如果 $\beta > 1$，说明资产在市场方向上的暴露比市场自身更大；如果 $0 < \beta < 1$，说明暴露较弱；如果 $\beta < 0$，说明它和市场方向反向移动，可能具有对冲属性。

# 5 子空间投影与最小二乘

一维投影只允许用一个方向 $\boldsymbol{x}$ 来解释 $\boldsymbol{y}$。回归中我们通常有多个解释变量，比如教育、经验、城市、家庭背景。于是解释空间不再是一条直线，而是一个由多列向量张成的子空间。

设

$$\boldsymbol{X} = \begin{pmatrix} | & | & & | \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_k \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{n\times k},$$

其中每一列 $\boldsymbol{x}_j \in \mathbb{R}^n$ 是一个解释变量在 $n$ 个观测上的取值。我们希望用

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta}$$

近似 $\boldsymbol{y}$。这就是最小二乘问题：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^k} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 .$$

## 5.1 Normal equations 的推导

设目标函数为

$$Q(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}).$$

展开：

$$Q(\boldsymbol{\beta}) = \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}.$$

对 $\boldsymbol{\beta}$ 求梯度：

$$\nabla_{\boldsymbol{\beta}} Q(\boldsymbol{\beta}) = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}.$$

令一阶条件为零：

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}.$$

这就是 normal equations。若 $\boldsymbol{X}'\boldsymbol{X}$ 可逆，则

$$\boxed{\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.}$$

**为什么 normal equations 表示残差与解释空间正交?**

令

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}.$$

normal equations 为

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}.$$

移项得到

$$\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = 0,$$

即

$$\boldsymbol{X}'\hat{\boldsymbol{\varepsilon}} = 0.$$

因为 $\boldsymbol{X}'\hat{\boldsymbol{\varepsilon}}$ 的第 $j$ 个元素是 $\boldsymbol{x}_j'\hat{\boldsymbol{\varepsilon}}$，所以它表示残差和每一个解释变量列向量都正交。由于 $\mathrm{Col}(\boldsymbol{X})$ 中任何向量都可以写成 $\boldsymbol{X}a$，有

$$(\boldsymbol{X}a)'\hat{\boldsymbol{\varepsilon}} = a'\boldsymbol{X}'\hat{\boldsymbol{\varepsilon}} = 0.$$

因此

$$\hat{\boldsymbol{\varepsilon}} \perp \mathrm{Col}(\boldsymbol{X}).$$

**方法**

OLS 的三个等价说法：

1. $\hat{\boldsymbol{\beta}}$ 最小化平方误差 $\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2$；
2. $\hat{\boldsymbol{\beta}}$ 满足 normal equations：$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$；
3. 残差与解释空间正交：$\boldsymbol{X}'\hat{\boldsymbol{\varepsilon}} = 0$。

以后学计量时，你会看到这三个版本不断出现。

## 5.2 投影定理

**定义 8.4：子空间投影**

设 $S \subset \mathbb{R}^n$ 是一个线性子空间。向量 $\hat{\boldsymbol{y}} \in S$ 若满足

$$\|\boldsymbol{y} - \hat{\boldsymbol{y}}\| \le \|\boldsymbol{y} - \boldsymbol{s}\| \quad \text{for all } \boldsymbol{s} \in S,$$

则称 $\hat{\boldsymbol{y}}$ 是 $\boldsymbol{y}$ 在 $S$ 上的投影，记为 $\mathrm{Proj}_S(\boldsymbol{y})$。

**定理 8.1：投影定理 Projection Theorem**

设 $S$ 是 $\mathbb{R}^n$ 的线性子空间。对任意 $\boldsymbol{y} \in \mathbb{R}^n$，存在唯一分解

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \hat{\boldsymbol{\varepsilon}}, \qquad \hat{\boldsymbol{y}} \in S, \qquad \hat{\boldsymbol{\varepsilon}} \perp S.$$

其中 $\hat{\boldsymbol{y}}$ 是 $\boldsymbol{y}$ 在 $S$ 上的唯一最优平方距离近似。

**证明思路：正交分解为什么给出最短距离**

严格证明需要线性代数中的基和 Gram-Schmidt 过程。这里给出经济数学中最常用的证明思路。

假设存在 $\hat{\boldsymbol{y}} \in S$，使得残差 $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \hat{\boldsymbol{y}} \perp S$。任取另一个 $\boldsymbol{s} \in S$，因为 $\hat{\boldsymbol{y}}, \boldsymbol{s} \in S$，所以 $\hat{\boldsymbol{y}} - \boldsymbol{s} \in S$。于是

$$\boldsymbol{y} - \boldsymbol{s} = (\boldsymbol{y} - \hat{\boldsymbol{y}}) + (\hat{\boldsymbol{y}} - \boldsymbol{s}) = \hat{\boldsymbol{\varepsilon}} + (\hat{\boldsymbol{y}} - \boldsymbol{s}).$$

由于 $\hat{\boldsymbol{\varepsilon}} \perp (\hat{\boldsymbol{y}} - \boldsymbol{s})$，由勾股定理，

$$\|\boldsymbol{y} - \boldsymbol{s}\|^2 = \|\hat{\boldsymbol{\varepsilon}}\|^2 + \|\hat{\boldsymbol{y}} - \boldsymbol{s}\|^2 \geq \|\hat{\boldsymbol{\varepsilon}}\|^2 = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2 .$$

所以 $\hat{\boldsymbol{y}}$ 确实是最接近 $\boldsymbol{y}$ 的点。
唯一性来自：若有两个不同的最优投影 $\hat{\boldsymbol{y}}_1, \hat{\boldsymbol{y}}_2 \in S$，则它们的差属于 $S$，而两个残差都与 $S$ 正交，经过代数整理会推出 $\|\hat{\boldsymbol{y}}_1 - \hat{\boldsymbol{y}}_2\|^2 = 0$，故二者相同。

# 6 投影矩阵：把几何变成线性算子

如果 $\boldsymbol{X}$ 满列秩，即 $\operatorname{rank}(\boldsymbol{X}) = k$，那么 $\boldsymbol{X}'\boldsymbol{X}$ 可逆，投影点是

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$

定义

$$\boxed{\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'.}$$

则

$$\hat{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{y}.$$

这说明投影不是每次重新发明一次几何，而是一个矩阵算子：输入 $\boldsymbol{y}$，输出它在 $\operatorname{Col}(\boldsymbol{X})$ 上的影子。

## 6.1 投影矩阵的性质

**命题 8.2：投影矩阵的基本性质**

若 $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$，且 $\boldsymbol{X}$ 满列秩，则：

1. $\boldsymbol{P}' = \boldsymbol{P}$，即 $\boldsymbol{P}$ 对称；
2. $\boldsymbol{P}^2 = \boldsymbol{P}$，即 $\boldsymbol{P}$ 幂等；
3. $\boldsymbol{P}\boldsymbol{y} \in \operatorname{Col}(\boldsymbol{X})$；
4. $\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y} \perp \operatorname{Col}(\boldsymbol{X})$。

**证明**

对称性：

$$\boldsymbol{P}' = \left[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right]' = \boldsymbol{X}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\right]'\boldsymbol{X}'.$$

因为 $\boldsymbol{X}'\boldsymbol{X}$ 是对称矩阵，它的逆也对称，所以

$$\boldsymbol{P}' = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \boldsymbol{P}.$$

幂等性：

$$\boldsymbol{P}^2 = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \boldsymbol{P}.$$

第三点显然，因为 $\boldsymbol{P}\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{X}$ 列向量的线性组合。
第四点：

$$\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y}) = \boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = 0.$$

所以残差与 $\mathrm{Col}(\boldsymbol{X})$ 正交。

## 6.2 残差生成矩阵

定义

$$\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}.$$

则残差为

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{y} = \boldsymbol{M}\boldsymbol{y}.$$

$\boldsymbol{M}$ 也有两个重要性质：

$$\boldsymbol{M}' = \boldsymbol{M}, \qquad \boldsymbol{M}^2 = \boldsymbol{M}.$$

它把任意 $\boldsymbol{y}$ 投影到 $\mathrm{Col}(\boldsymbol{X})$ 的正交补上。

**直觉**

$\boldsymbol{P}$ 提取“能被 $X$ 解释的部分”，$\boldsymbol{M}$ 提取“不能被 $X$ 线性解释的部分”。
这就是计量里 partialling out、Frisch-Waugh-Lovell 定理、固定效应残差化的底层语言。

# 7 OLS 的经济含义：不是公式，而是正交条件

## 7.1 样本回归的正交条件

假设有 $n$ 个观测，模型写成

$$y_i = x_i'\beta + \varepsilon_i,$$

其中 $x_i \in \mathbb{R}^k$。把所有观测堆叠起来就是

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

OLS 估计量满足

$$\boldsymbol{X}'\hat{\boldsymbol{\varepsilon}} = 0.$$

展开成求和形式：

$$\sum_{i=1}^{n} x_i \hat{\varepsilon}_i = 0.$$

如果模型包含截距项，即 $x_i$ 的第一维总是 1，那么第一条正交条件就是

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = 0.$$

这表示包含截距的 OLS 残差均值为零。其他解释变量对应的条件为

$$\sum_{i=1}^{n} x_{ij} \hat{\varepsilon}_i = 0,$$

表示残差与第 $j$ 个解释变量样本协动为零。

**常见误区**

$\sum x_i \hat{\varepsilon}_i = 0$ 是 OLS 算法机械保证的样本性质，不等于经济意义上的外生性假设 $\mathbb{E}[X\varepsilon] = 0$。前者总是在 OLS 解处成立；后者是关于真实数据生成过程的识别条件。二者长得像，但含义不同。

## 7.2 Population projection

在总体层面，我们不再使用样本内积 $\sum_i x_i y_i$，而使用概率内积 $\mathbb{E}[XY]$。设 $Y$ 是因变量，$X \in \mathbb{R}^k$ 是解释变量。population linear projection 定义为

$$\beta^* = \arg\min_{\beta \in \mathbb{R}^k} \mathbb{E}[(Y - X'\beta)^2].$$

目标函数为

$$Q(\beta) = \mathbb{E}[(Y - X'\beta)^2].$$

对 $\beta$ 求一阶条件：

$$\mathbb{E}[X(Y - X'\beta^*)] = 0.$$

即

$$\mathbb{E}[XY] = \mathbb{E}[XX']\beta^*.$$

若 $\mathbb{E}[XX']$ 可逆，则

$$\boxed{\beta^* = \mathbb{E}[XX']^{-1}\mathbb{E}[XY].}$$

**直觉**

样本 OLS 是

$$\boldsymbol{X}'(\boldsymbol{y}-\boldsymbol{X}\hat{\beta})=0.$$

总体线性投影是

$$\mathbb{E}[X(Y-X'\beta^*)]=0.$$

二者形式完全平行。前者使用样本内积，后者使用概率内积。

## 7.3 条件期望：更强的投影对象

线性投影只允许使用 $X$ 的线性函数 $X'\beta$。如果允许使用任意可测函数 $g(X)$，那么最佳平方预测是条件期望：

$$\mathbb{E}[Y\mid X]=\arg\min_{g(X)}\mathbb{E}[(Y-g(X))^2].$$

这个结论将在 Chapter 22 详细证明。现在你只需要知道：

> OLS 是在线性函数空间中的投影，条件期望是在所有$X$ 函数空间中的投影。

这就是为什么回归理论和函数空间投影联系极深。

# 8 Worked Examples：完整样板题

**例 8.2：手算一个简单 OLS**

有三个观测点：

$$(x_i,y_i)=(1,2),(2,2),(3,4).$$

考虑含截距的一元回归

$$y_i=\alpha+\beta x_i+\varepsilon_i.$$

求 OLS 估计量。

**解：**矩阵写法为

$$\boldsymbol{X}=\begin{pmatrix}1&1\\1&2\\1&3\end{pmatrix},\qquad \boldsymbol{y}=\begin{pmatrix}2\\2\\4\end{pmatrix}.$$

先算

$$\boldsymbol{X}'\boldsymbol{X}=\begin{pmatrix}1&1&1\\1&2&3\end{pmatrix}\begin{pmatrix}1&1\\1&2\\1&3\end{pmatrix}=\begin{pmatrix}3&6\\6&14\end{pmatrix}.$$

再算

$$\boldsymbol{X}'\boldsymbol{y} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 8 \\ 18 \end{pmatrix}.$$

因此 normal equations 是

$$\begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} 8 \\ 18 \end{pmatrix}.$$

解得

$$\hat{\beta} = 1, \qquad \hat{\alpha} = \frac{2}{3}.$$

所以拟合值为

$$\hat{\boldsymbol{y}} = \begin{pmatrix} 5/3 \\ 8/3 \\ 11/3 \end{pmatrix},$$

残差为

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \begin{pmatrix} 1/3 \\ -2/3 \\ 1/3 \end{pmatrix}.$$

检查正交条件：

$$\sum_i \hat{\varepsilon}_i = 1/3 - 2/3 + 1/3 = 0,$$

$$\sum_i x_i \hat{\varepsilon}_i = 1 \cdot \frac{1}{3} + 2 \cdot \left(-\frac{2}{3}\right) + 3 \cdot \frac{1}{3} = 0.$$

这说明残差和截距列、$x$ 列都正交。

**例 8.3：用投影理解 CAPM beta**

假设市场超额收益 $R_m$ 的方差为 $0.04$，某资产超额收益 $R_i$ 与市场超额收益的协方差为 $0.06$。求该资产对市场因子的 beta。

**解：** 单因子投影系数为

$$\beta_i = \frac{\mathrm{Cov}(R_i, R_m)}{\mathrm{Var}(R_m)}.$$

代入得到

$$\beta_i = \frac{0.06}{0.04} = 1.5.$$

解释：该资产在市场因子方向上的暴露为 1.5。若市场超额收益增加 1 个百分点，线性投影意义下该资产超额收益平均增加约 1.5 个百分点。

注意：这不是说每一次市场上涨 1% 时该资产必然上涨 1.5%，而是说在均方误差

意义下，$1.5R_m$ 是用市场因子解释 $R_i$ 的最佳线性部分。

**例 8.4：用两个因子解释资产收益**

设资产超额收益 $R$ 用两个因子 $F_1, F_2$ 线性投影：

$$R = \beta_1 F_1 + \beta_2 F_2 + \varepsilon.$$

假设

$$\mathbb{E}[FF'] = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}, \qquad \mathbb{E}[FR] = \begin{pmatrix} 5 \\ 3 \end{pmatrix},$$

其中 $F = (F_1, F_2)'$。求 $\beta^*$。

**解：**总体投影满足

$$\mathbb{E}[FF']\beta^* = \mathbb{E}[FR].$$

所以

$$\beta^* = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 5 \\ 3 \end{pmatrix}.$$

矩阵逆为

$$\begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}^{-1} = \frac{1}{8-1} \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix} = \frac{1}{7} \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix}.$$

因此

$$\beta^* = \frac{1}{7} \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} 5 \\ 3 \end{pmatrix} = \frac{1}{7} \begin{pmatrix} 7 \\ 7 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

所以两个因子暴露均为 1。

# 9 函数空间投影：从向量到函数

前面我们在 $\mathbb{R}^n$ 中讨论投影。现在把对象从有限维向量换成函数。设 $f, g$ 是定义在 $[a, b]$ 上的函数。我们可以定义函数内积：

$$\langle f, g \rangle = \int_a^b f(x) g(x)\, \mathrm{d}x.$$

函数的长度为

$$\|f\| = \sqrt{\int_a^b f(x)^2\, \mathrm{d}x}.$$

函数之间的平方距离为

$$\|f - g\|^2 = \int_a^b [f(x) - g(x)]^2\, \mathrm{d}x.$$

**定义 8.5：函数空间中的正交**

若

$$\int_a^b f(x)g(x)\,\mathrm{d}x = 0,$$

则称函数 $f$ 与 $g$ 在 $[a,b]$ 上正交。

这和向量正交完全平行。向量内积是求和，函数内积是积分。求和对应有限样本，积分对应连续空间或总体平均。

## 9.1 函数投影的一般形式

假设我们想用一组基函数 $\phi_1(x),\dots,\phi_k(x)$ 的线性组合近似函数 $f(x)$：

$$\hat{f}(x) = a_1\phi_1(x) + \cdots + a_k\phi_k(x).$$

目标是最小化积分平方误差：

$$\min_{a_1,dots,a_k} \int_a^b \left[ f(x) - \sum_{j=1}^{k} a_j\phi_j(x) \right]^2 \mathrm{d}x.$$

最优条件是残差函数

$$r(x) = f(x) - \hat{f}(x)$$

与每一个基函数正交：

$$\int_a^b r(x)\phi_j(x)\,\mathrm{d}x = 0, \qquad j = 1,\dots,k.$$

这就是函数空间版本的 normal equations。

**方法**

向量 OLS 与函数投影的对应关系：

| 向量回归 | 函数投影 |
|---|---|
| $\boldsymbol{y} \in \mathbb{R}^n$ | $f(x)$ |
| 解释变量列 $\boldsymbol{x}_j$ | 基函数 $\phi_j(x)$ |
| 拟合值 $\boldsymbol{X}\hat{\beta}$ | 近似函数 $\sum_j \hat{a}_j\phi_j(x)$ |
| 平方误差 $\sum_i \hat{\varepsilon}_i^2$ | 积分平方误差 $\int r(x)^2\,\mathrm{d}x$ |
| 正交条件 $\boldsymbol{x}_j'\hat{\boldsymbol{\varepsilon}} = 0$ | $\int r(x)\phi_j(x)\,\mathrm{d}x = 0$ |

## 9.2 若基函数两两正交，系数会大大简化

如果 $\phi_1,\dots,\phi_k$ 两两正交，即

$$\langle \phi_i, \phi_j \rangle = 0 \quad (i \neq j),$$

那么投影系数可以逐个计算：

$$\boxed{\hat{a}_j = \frac{\langle f, \phi_j \rangle}{\langle \phi_j, \phi_j \rangle}.}$$

这就是 Fourier 系数公式的底层来源。

**为什么正交基下系数可以逐个算？**

设

$$\hat{f} = \sum_{j=1}^{k} a_j \phi_j, \qquad r = f - \hat{f}.$$

最优条件为 $\langle r, \phi_m \rangle = 0$，即

$$\left\langle f - \sum_{j=1}^{k} a_j \phi_j, \phi_m \right\rangle = 0.$$

展开：

$$\langle f, \phi_m \rangle - \sum_{j=1}^{k} a_j \langle \phi_j, \phi_m \rangle = 0.$$

由于正交性，当 $j \neq m$ 时 $\langle \phi_j, \phi_m \rangle = 0$，只剩

$$\langle f, \phi_m \rangle - a_m \langle \phi_m, \phi_m \rangle = 0.$$

所以

$$a_m = \frac{\langle f, \phi_m \rangle}{\langle \phi_m, \phi_m \rangle}.$$

**例 8.5：把 $f(x) = x$ 投影到常数函数空间**

在区间 $[0, 1]$ 上，用常数 $a$ 近似 $f(x) = x$，即求

$$\min_{a} \int_0^1 (x - a)^2 \, \mathrm{d}x.$$

**解法一：直接求导。**

$$Q(a) = \int_0^1 (x - a)^2 \, \mathrm{d}x.$$

对 $a$ 求导：

$$Q'(a) = \int_0^1 -2(x - a) \, \mathrm{d}x = -2\left(\frac{1}{2} - a\right).$$

令 $Q'(a) = 0$，得

$$\hat{a} = \frac{1}{2}.$$

**解法二：投影公式**。常数函数空间由 $\phi_1(x)=1$ 张成。投影系数为

$$\hat{a}=\frac{\langle f,1\rangle}{\langle 1,1\rangle}=\frac{\int_0^1 x\,\mathrm{d}x}{\int_0^1 1\,\mathrm{d}x}=\frac{1}{2}.$$

解释：在 $[0,1]$ 上，最接近函数 $x$ 的常数就是它的平均值 $1/2$。

**例 8.6：用 $1,x$ 近似 $x^2$**

在 $[0,1]$ 上，用线性函数 $a+bx$ 近似 $f(x)=x^2$，求最小化

$$\int_0^1 (x^2-a-bx)^2\,\mathrm{d}x$$

的 $a,b$。

**解：**残差为

$$r(x)=x^2-a-bx.$$

最优条件是残差与基函数 $1,x$ 正交：

$$\int_0^1 r(x)\,\mathrm{d}x=0,\qquad \int_0^1 xr(x)\,\mathrm{d}x=0.$$

第一条：

$$\int_0^1 (x^2-a-bx)\,\mathrm{d}x=\frac{1}{3}-a-\frac{b}{2}=0.$$

第二条：

$$\int_0^1 x(x^2-a-bx)\,\mathrm{d}x=\frac{1}{4}-\frac{a}{2}-\frac{b}{3}=0.$$

所以

$$a+\frac{b}{2}=\frac{1}{3},\qquad \frac{a}{2}+\frac{b}{3}=\frac{1}{4}.$$

解得

$$b=1,\qquad a=-\frac{1}{6}.$$

因此最佳线性近似为

$$\hat{f}(x)=-\frac{1}{6}+x.$$

注意：这不是 Taylor 展开。Taylor 展开关注某一点附近的局部近似，而这里关注整个区间 $[0,1]$ 上的平均平方误差最小。

# 10 Fourier 展开的入口：正交基函数

Chapter 9 会系统讲 Fourier 级数，本章只讲它和投影的关系。

在区间 $[-\pi,\pi]$ 上，函数

$$1,\quad \cos x,\quad \sin x,\quad \cos 2x,\quad \sin 2x,\dots$$

具有正交性。例如：

$$\int_{-\pi}^{\pi}\sin(mx)\cos(nx)\,\mathrm{d}x=0,$$

$$\int_{-\pi}^{\pi}\cos(mx)\cos(nx)\,\mathrm{d}x=0\quad(m\neq n),$$

$$\int_{-\pi}^{\pi}\sin(mx)\sin(nx)\,\mathrm{d}x=0\quad(m\neq n).$$

因此，用这些函数逼近一个周期函数 $f$ 时，系数可以通过投影公式逐项计算。

**Fourier 系数的投影形式**

在 $[-\pi,\pi]$ 上，若

$$f(x)\sim\frac{a_0}{2}+\sum_{n=1}^{\infty}\left(a_n\cos nx+b_n\sin nx\right),$$

则系数为

$$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos(nx)\,\mathrm{d}x,$$

$$b_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin(nx)\,\mathrm{d}x,$$

$$a_0=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\,\mathrm{d}x.$$

**直觉**

Fourier 系数不是神秘公式，它就是问：**函数 $f$ 在频率 $n$ 的正弦/余弦方向上有多少投影?**

这和金融因子模型很像：资产收益在市场因子、规模因子、价值因子上的 beta，是收益在不同风险因子方向上的投影。Fourier 只是把“因子”换成了不同频率的波。

**例 8.7：计算一个简单 Fourier 投影系数**

在 $[-\pi,\pi]$ 上，令 $f(x)=x$。求它在 $\sin x$ 方向上的投影系数。

**解：**因为 $\sin x$ 的范数平方为

$$\langle\sin x,\sin x\rangle=\int_{-\pi}^{\pi}\sin^2 x\,\mathrm{d}x=\pi,$$

所以投影系数为

$$b_1=\frac{\langle x,\sin x\rangle}{\langle\sin x,\sin x\rangle}=\frac{1}{\pi}\int_{-\pi}^{\pi}x\sin x\,\mathrm{d}x.$$

由于 $x\sin x$ 是偶函数，

$$\int_{-\pi}^{\pi} x\sin x\,\mathrm{d}x = 2\int_{0}^{\pi} x\sin x\,\mathrm{d}x.$$

分部积分：

$$\int_{0}^{\pi} x\sin x\,\mathrm{d}x = [-x\cos x]_{0}^{\pi} + \int_{0}^{\pi}\cos x\,\mathrm{d}x = \pi + 0 = \pi.$$

因此

$$b_1 = \frac{1}{\pi}\cdot 2\pi = 2.$$

这说明 $x$ 在 $\sin x$ 方向上的 Fourier 投影系数为 2。

# 11 从投影到计量：OLS、矩条件与识别

本章虽然是数学章节，但它直接通向计量经济学。你后面学计量时会反复看到下面这些形式：

$$\mathbb{E}[X\varepsilon] = 0, \qquad \frac{1}{n}\sum_{i=1}^{n} X_i\hat{\varepsilon}_i = 0, \qquad \mathbb{E}[m(W,\theta)] = 0.$$

这些都是“正交条件”的变体。

## 11.1 OLS 作为样本矩条件

OLS 的一阶条件可以写为

$$\frac{1}{n}\sum_{i=1}^{n} x_i(y_i - x_i'\hat{\beta}) = 0.$$

这就是样本矩条件。它说参数 $\hat{\beta}$ 被选择为让样本中的解释变量和残差正交。

如果真实参数 $\beta_0$ 满足

$$\mathbb{E}[X(Y - X'\beta_0)] = 0,$$

那么样本矩条件就是总体矩条件的样本版本。大数定律保证样本平均接近总体期望，这就是 OLS 一致性的直觉入口。

## 11.2 工具变量和 GMM 的影子

如果解释变量 $X$ 和误差 $\varepsilon$ 不正交，OLS 就不再有因果解释。工具变量方法会寻找另一个变量 $Z$，要求

$$\mathbb{E}[Z\varepsilon] = 0.$$

GMM 更一般地写成

$$\mathbb{E}[m(W,\theta_0)] = 0.$$

虽然这些内容属于后续计量课程，但本章的投影语言会让你更容易理解：现代计量很大程度上就是在寻找、使用和估计各种正交条件。

**常见误区**

正交条件有两个层次：

1. **几何正交**：给定样本数据和 OLS 算法，残差一定和解释变量列空间正交；
2. **经济识别正交**：真实误差项与某些变量在总体上正交，这是一个关于经济机制和数据生成过程的假设。

前者是数学事实，后者是经济假设。不要混淆。

# 12 从投影到金融：因子模型与风险分解

## 12.1 多因子模型

金融中的多因子模型常写为

$$R_i = \alpha_i + \beta_{i1} F_1 + \cdots + \beta_{ik} F_k + \varepsilon_i.$$

这里 $F_1, \ldots, F_k$ 是共同风险因子，例如市场、规模、价值、动量、利率、汇率、波动率等。估计 $\beta_{ij}$ 的过程就是把资产收益投影到因子空间上。

若因子均值为零，且不考虑截距，则总体 beta 满足

$$\mathbb{E}[FF']\beta_i = \mathbb{E}[FR_i].$$

其中 $F = (F_1, \ldots, F_k)'$。

## 12.2 风险分解的几何含义

如果

$$R_i = \hat{R}_i + \varepsilon_i, \qquad \hat{R}_i = \beta_i' F, \qquad \varepsilon_i \perp F,$$

那么收益被分解成两部分：

- 可由共同因子解释的系统性部分 $\hat{R}_i$；
- 与共同因子正交的特质性部分 $\varepsilon_i$。

  如果所有变量均值为零，则由正交性可得方差分解：

$$\mathrm{Var}(R_i) = \mathrm{Var}(\hat{R}_i) + \mathrm{Var}(\varepsilon_i).$$

这正是勾股定理在概率内积空间中的版本。

**方差分解为什么成立?**

因为

$$R_i = \hat{R}_i + \varepsilon_i,$$

所以

$$\mathrm{Var}(R_i) = \mathrm{Var}(\hat{R}_i) + \mathrm{Var}(\varepsilon_i) + 2\,\mathrm{Cov}(\hat{R}_i, \varepsilon_i).$$

而 $\hat{R}_i$ 是因子 $F$ 的线性组合，$\varepsilon_i$ 与所有因子正交，因此

$$\mathrm{Cov}(\hat{R}_i, \varepsilon_i) = 0.$$

于是

$$\mathrm{Var}(R_i) = \mathrm{Var}(\hat{R}_i) + \mathrm{Var}(\varepsilon_i).$$

# 13 非参数回归和机器学习的入口：基函数展开

本章的投影思想还可以直接通向非参数回归和机器学习。假设真实关系不是线性的:

$$Y = f(X) + \varepsilon.$$

我们可以选一组基函数:

$$1, \quad X, \quad X^2, \quad X^3, \quad \sin X, \quad \cos X, \quad \text{spline basis}, \ldots$$

然后用

$$f_K(X) = \sum_{j=1}^{K} a_j \phi_j(X)$$

近似 $f(X)$。估计系数 $a_j$ 的过程本质上仍然是最小二乘投影。

这就是为什么:

- polynomial regression 是投影到多项式空间;
- Fourier regression 是投影到三角函数空间;
- spline regression 是投影到样条函数空间;
- series estimation 是投影到不断扩张的基函数空间;
- 一些机器学习方法可以看作在高维特征空间中做正则化投影。

**直觉**

你以后看到 “feature engineering” 时，可以把它理解成：人为扩张解释空间。原来只允许用 $X$，现在允许用 $X^2$、$X^3$、交互项、分段函数、神经网络隐层特征。解释空间越大，拟合能力越强，但过拟合风险也越大。

# 14 常见错误与概念辨析

**常见误区**

**误区 1：把投影理解为因果。** OLS 投影只说明最佳线性预测，不自动说明 $X$ 对 $Y$ 有因果作用。因果解释需要额外识别假设。

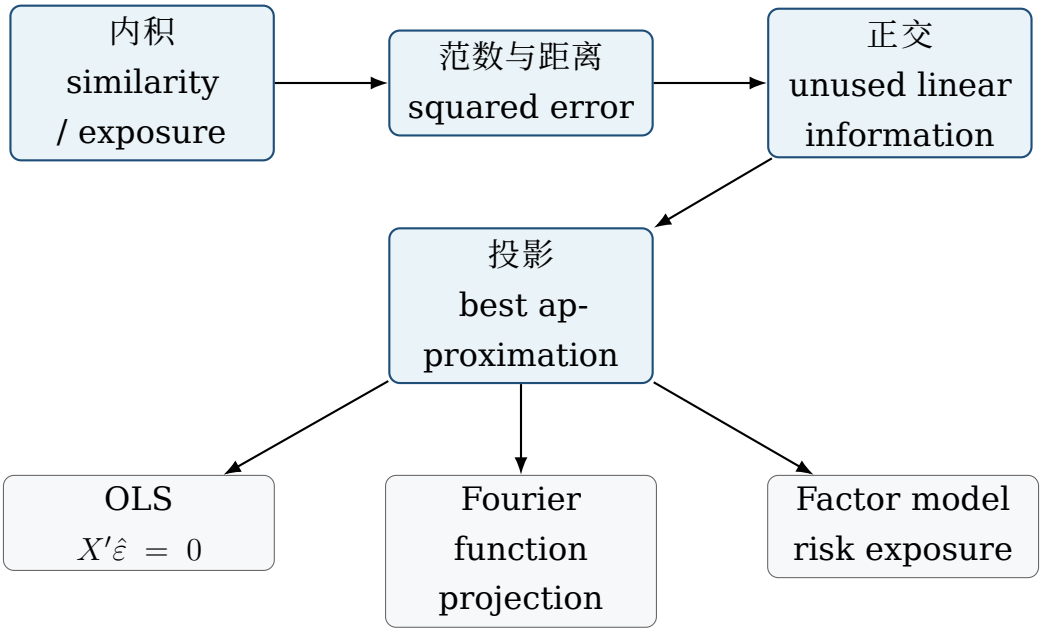
**误区 2：以为残差正交等于真实误差外生。**样本残差正交是 OLS 算法保证的；真实误差外生是模型假设。前者不能推出后者。

**误区 3：忘记满列秩条件。** $\hat{\beta} = (X'X)^{-1}X'y$ 需要 $X'X$ 可逆。如果解释变量完全共线，逆不存在。

**误区 4：认为 Fourier 和回归无关。**它们都是投影。区别只是一个在有限维样本空间里投影，一个在函数空间里投影。

**误区 5：混淆局部近似和全局平方近似。** Taylor 展开通常是某一点附近的局部近似；最小二乘投影是某个区间或样本上的整体平方误差最小近似。

# 15 本章概念图谱




# 16 练习题

## A 组：基础计算

A1. 设 $\boldsymbol{u} = (1, 2, 2)'$、$\boldsymbol{v} = (2, -1, 0)'$。计算 $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$、$\|\boldsymbol{u}\|$、$\|\boldsymbol{v}\|$，并判断二者是否正交。

A2. 设 $\boldsymbol{y} = (3, 1)'$、$\boldsymbol{x} = (1, 1)'$。求 $\boldsymbol{y}$ 在 $\operatorname{span}\{\boldsymbol{x}\}$ 上的投影和残差。

A3. 设 $\boldsymbol{y} = (1, 2, 3)'$、$\boldsymbol{x} = (1, 0, 1)'$。求 $\hat{b} = \arg\min_b \|\boldsymbol{y} - b\boldsymbol{x}\|^2$。

A4. 设

$$\boldsymbol{X}=\begin{pmatrix}1 & 0\\ 1 & 1\\ 1 & 2\end{pmatrix},\qquad \boldsymbol{y}=\begin{pmatrix}1\\ 2\\ 2\end{pmatrix}.$$

计算 $\boldsymbol{X}'\boldsymbol{X}$ 和 $\boldsymbol{X}'\boldsymbol{y}$。

A5. 在 $[0,1]$ 上，计算函数 $f(x)=x$ 和 $g(x)=1-x$ 的内积 $\int_0^1 f(x)g(x)\,\mathrm{d}x$。

A6. 在 $[-\pi,\pi]$ 上，验证 $\sin x$ 与 $\cos x$ 正交。

## B 组：概念判断

B1. 判断：OLS 残差与解释变量正交，所以解释变量一定外生。说明理由。

B2. 判断：如果 $\boldsymbol{X}'\boldsymbol{X}$ 不可逆，OLS 最小化问题一定没有解。说明理由。

B3. 判断：Fourier 系数可以理解为函数在正弦和余弦方向上的投影。说明理由。

B4. 判断：Taylor 展开和最小二乘投影都是近似，所以二者没有本质区别。说明理由。

B5. 判断：投影矩阵 $\boldsymbol{P}$ 满足 $\boldsymbol{P}^2=\boldsymbol{P}$，所以投影一次和投影两次结果相同。说明理由。

## C 组：混合题

C1. 给定 $\boldsymbol{X}\in\mathbb{R}^{n\times k}$ 满列秩，证明 $\boldsymbol{M}=\boldsymbol{I}-\boldsymbol{P}$ 对称且幂等。

C2. 设 $\hat{\boldsymbol{y}}=\boldsymbol{P}\boldsymbol{y}$、$\hat{\boldsymbol{\varepsilon}}=\boldsymbol{M}\boldsymbol{y}$。证明 $\hat{\boldsymbol{y}}'\hat{\boldsymbol{\varepsilon}}=0$，并推出 $\|\boldsymbol{y}\|^2=\|\hat{\boldsymbol{y}}\|^2+\|\hat{\boldsymbol{\varepsilon}}\|^2$。

C3. 在 $[0,1]$ 上，用 $ax$ 近似 $f(x)=1$，求使 $\int_0^1(1-ax)^2\,\mathrm{d}x$ 最小的 $a$。

C4. 在 $[0,1]$ 上，用 $a+bx$ 近似 $f(x)=x^3$。写出 $a,b$ 满足的 normal equations，不必求解。

C5. 若 $X,Y$ 为均值零随机变量，证明 $\beta^*=\arg\min_b \mathbb{E}[(Y-bX)^2]$ 满足 $\beta^*=\mathrm{Cov}(X,Y)/\mathrm{Var}(X)$，假设 $\mathrm{Var}(X)>0$。

## D 组：经济金融建模题

D1. 一个资产超额收益 $R_i$ 与市场超额收益 $R_m$ 的相关系数为 $0.6$，标准差分别为 $\sigma_i=0.30$、$\sigma_m=0.20$。求市场 beta，并解释含义。

D2. 设工资模型为 $w_i=\alpha+\beta s_i+\varepsilon_i$，其中 $s_i$ 是教育年限。解释 OLS 中 $\sum_i s_i\hat{\varepsilon}_i=0$ 的几何含义，并说明它为什么不等于教育的因果效应已经被识别。

D3. 在两因子模型 $R=\beta_1F_1+\beta_2F_2+\varepsilon$ 中，若 $F_1,F_2$ 正交且方差分别为 $4,9$，并且 $\mathrm{Cov}(R,F_1)=8$、$\mathrm{Cov}(R,F_2)=3$，求 $\beta_1,\beta_2$。

D4. 一个研究者用 $1,x,x^2$ 近似消费函数 $c(x)$。解释这在投影语言下意味着什么，并说明为什么增加 $x^3,x^4$ 会提高拟合能力但可能带来过拟合。

D5. 设一个基金收益被市场因子解释后的残差仍与动量因子高度相关。用投影语言解释这意味着什么。

## E 组：挑战/证明题

E1. 证明：若 $\boldsymbol{P}$ 是对称且幂等矩阵，则对任意 $\boldsymbol{y}$，$\boldsymbol{Py}$ 与 $(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{y}$ 正交。

E2. 设 $\phi_1,\dots,\phi_k$ 不一定正交。推导函数投影的一般 normal equations：

$$\sum_{j=1}^{k} a_j \langle \phi_j, \phi_m \rangle = \langle f, \phi_m \rangle, \qquad m = 1,\dots,k.$$

E3. 设 $Y = m(X) + u$，且 $\mathbb{E}[u \mid X] = 0$。证明对任意函数 $g(X)$，有 $\mathbb{E}[g(X)u] = 0$。解释这和条件期望作为投影的关系。

E4. 证明：若 $\hat{\beta}$ 是 OLS 解，则 $\hat{\boldsymbol{y}}$ 是 $\boldsymbol{y}$ 在 $\mathrm{Col}(\boldsymbol{X})$ 上的投影，即使 $\boldsymbol{X}'\boldsymbol{X}$ 不可逆，只要最小化问题有解。

E5. 思考：如果把平方损失换成绝对值损失 $\sum_i |y_i - x_i'\beta|$，为什么“残差正交”结构不再成立？这会通向什么类型的回归？

# 17 练习题参考答案

## A 组答案

**A1.**

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 1 \cdot 2 + 2 \cdot (-1) + 2 \cdot 0 = 0.$$

$$\|\boldsymbol{u}\| = \sqrt{1^2 + 2^2 + 2^2} = 3, \qquad \|\boldsymbol{v}\| = \sqrt{2^2 + (-1)^2} = \sqrt{5}.$$

内积为零，所以正交。

**A2.**

$$\hat{b} = \frac{\boldsymbol{x}'\boldsymbol{y}}{\boldsymbol{x}'\boldsymbol{x}} = \frac{1 \cdot 3 + 1 \cdot 1}{1^2 + 1^2} = 2.$$

投影为

$$\hat{\boldsymbol{y}} = 2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix},$$

残差为

$$\hat{\boldsymbol{e}} = \begin{pmatrix} 3 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

检查：$(1,1)(1,-1)' = 0$。

**A3.**

$$\hat{b} = \frac{\boldsymbol{x}'\boldsymbol{y}}{\boldsymbol{x}'\boldsymbol{x}} = \frac{1 \cdot 1 + 0 \cdot 2 + 1 \cdot 3}{1^2 + 0^2 + 1^2} = \frac{4}{2} = 2.$$

**A4.**

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}.$$

$$\boldsymbol{X}'\boldsymbol{y} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \end{pmatrix}.$$

**A5.**

$$\int_0^1 x(1-x)\,\mathrm{d}x = \int_0^1 (x - x^2)\,\mathrm{d}x = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.$$

**A6.**

$$\int_{-\pi}^{\pi} \sin x \cos x\,\mathrm{d}x = \frac{1}{2}\int_{-\pi}^{\pi} \sin(2x)\,\mathrm{d}x = 0.$$

也可以说 $\sin x \cos x$ 是奇函数，在对称区间积分为零。

## B 组答案

**B1.** 错。OLS 残差与解释变量正交是样本一阶条件，是算法保证的；解释变量外生是总体层面的经济假设，如 $\mathbb{E}[X\varepsilon]=0$。样本正交不能推出因果识别。

**B2.** 错。若 $\boldsymbol{X}'\boldsymbol{X}$ 不可逆，$\hat{\beta}$ 可能不唯一，但拟合值 $\hat{\boldsymbol{y}}$ 作为 $\boldsymbol{y}$ 在 $\mathrm{Col}(\boldsymbol{X})$ 上的投影仍然存在且唯一。

**B3.** 对。Fourier 的正弦、余弦函数构成正交系统，Fourier 系数就是函数在这些频率方向上的投影系数。

**B4.** 错。Taylor 主要是局部近似，围绕某一点匹配函数值和导数；最小二乘投影是整体平方误差最小，依赖选定区间或样本分布。

**B5.** 对。$\boldsymbol{P}^2=\boldsymbol{P}$ 表示已经落在投影空间里的对象再次投影不会改变。

## C 组答案

**C1.** 因为 $\boldsymbol{P}'=\boldsymbol{P}$，所以

$$\boldsymbol{M}'=(\boldsymbol{I}-\boldsymbol{P})'=\boldsymbol{I}-\boldsymbol{P}=\boldsymbol{M}.$$

又因为 $\boldsymbol{P}^2=\boldsymbol{P}$,

$$\boldsymbol{M}^2=(\boldsymbol{I}-\boldsymbol{P})^2=\boldsymbol{I}-2\boldsymbol{P}+\boldsymbol{P}^2=\boldsymbol{I}-\boldsymbol{P}=\boldsymbol{M}.$$

**C2.**

$$\hat{\boldsymbol{y}}'\hat{\boldsymbol{\varepsilon}}=(\boldsymbol{P}\boldsymbol{y})'(\boldsymbol{M}\boldsymbol{y})=\boldsymbol{y}'\boldsymbol{P}\boldsymbol{M}\boldsymbol{y}.$$

而

$$\boldsymbol{P}\boldsymbol{M}=\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{P})=\boldsymbol{P}-\boldsymbol{P}^2=0.$$

所以 $\hat{\boldsymbol{y}}'\hat{\boldsymbol{\varepsilon}}=0$。由于 $\boldsymbol{y}=\hat{\boldsymbol{y}}+\hat{\boldsymbol{\varepsilon}}$ 且二者正交，由勾股定理,

$$\|\boldsymbol{y}\|^2=\|\hat{\boldsymbol{y}}\|^2+\|\hat{\boldsymbol{\varepsilon}}\|^2.$$

**C3.**

$$Q(a)=\int_0^1(1-ax)^2\,\mathrm{d}x.$$

一阶条件:

$$\int_0^1 x(1-ax)\,\mathrm{d}x=0.$$

所以

$$\frac{1}{2}-a\frac{1}{3}=0,\qquad a=\frac{3}{2}.$$

**C4.** 残差为 $r(x)=x^3-a-bx$。正交条件:

$$\int_0^1(x^3-a-bx)\,\mathrm{d}x=0,$$

$$\int_0^1 x(x^3 - a - bx)\,\mathrm{d}x = 0.$$

即

$$\frac{1}{4} - a - \frac{b}{2} = 0, \qquad \frac{1}{5} - \frac{a}{2} - \frac{b}{3} = 0.$$

**C5.** 目标函数：

$$Q(b) = \mathbb{E}[(Y - bX)^2].$$

一阶条件：

$$\frac{dQ}{db} = \mathbb{E}[-2X(Y - bX)] = 0.$$

故

$$\mathbb{E}[XY] = b\mathbb{E}[X^2].$$

因均值为零，$\mathbb{E}[XY] = \mathrm{Cov}(X, Y)$，$\mathbb{E}[X^2] = \mathrm{Var}(X)$，所以

$$b = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}.$$

## D 组答案

**D1.**

$$\beta = \frac{\mathrm{Cov}(R_i, R_m)}{\mathrm{Var}(R_m)} = \frac{\mathrm{Corr}(R_i, R_m)\sigma_i\sigma_m}{\sigma_m^2} = \mathrm{Corr}(R_i, R_m)\frac{\sigma_i}{\sigma_m}.$$

代入：

$$\beta = 0.6 \cdot \frac{0.30}{0.20} = 0.9.$$

含义：该资产对市场因子的线性暴露为 0.9，系统性风险略低于市场。

**D2.** $\sum_i s_i\hat{\varepsilon}_i = 0$ 表示 OLS 残差向量和教育年限向量正交，即在样本线性投影后，残差中没有可以由教育年限继续线性解释的部分。但这不是因果识别，因为真实误差可能包含能力、家庭背景等遗漏变量，而这些变量可能与教育年限相关。

**D3.** 因子正交时，系数逐项计算：

$$\beta_1 = \frac{\mathrm{Cov}(R, F_1)}{\mathrm{Var}(F_1)} = \frac{8}{4} = 2,$$

$$\beta_2 = \frac{\mathrm{Cov}(R, F_2)}{\mathrm{Var}(F_2)} = \frac{3}{9} = \frac{1}{3}.$$

**D4.** 用 $1, x, x^2$ 近似消费函数，就是把真实消费函数投影到二次多项式空间。增加 $x^3, x^4$ 会扩张近似空间，因此训练样本内平方误差不会上升，通常会下降。但空间越大越可能拟合样本噪声，导致样本外预测变差，这就是过拟合。

**D5.** 如果市场因子回归后的残差仍与动量因子相关，说明市场因子空间没有用尽收益中的线性可解释信息。加入动量因子会扩张解释空间，使原残差的一部分被重新投影到新因子方向上。

## E 组答案

**E1.** 因为 $\boldsymbol{P}' = \boldsymbol{P}$、$\boldsymbol{P}^2 = \boldsymbol{P}$,

$$(\boldsymbol{P}\boldsymbol{y})'(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{P}'(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{P}-\boldsymbol{P}^2)\boldsymbol{y} = 0.$$

所以二者正交。

**E2.** 令 $\hat{f} = \sum_{j=1}^{k} a_j\phi_j$，残差 $r = f - \hat{f}$。最优要求残差与每个基方向正交:

$$\langle r, \phi_m\rangle = 0.$$

即

$$\left\langle f - \sum_{j=1}^{k} a_j\phi_j, \phi_m \right\rangle = 0.$$

线性展开得到

$$\langle f, \phi_m\rangle - \sum_{j=1}^{k} a_j \langle \phi_j, \phi_m\rangle = 0,$$

故

$$\sum_{j=1}^{k} a_j \langle \phi_j, \phi_m\rangle = \langle f, \phi_m\rangle, \quad m = 1, \ldots, k.$$

**E3.** 由迭代期望法则,

$$\mathbb{E}[g(X)u] = \mathbb{E}\left[\mathbb{E}[g(X)u \mid X]\right].$$

给定 $X$ 后，$g(X)$ 是已知量，所以

$$\mathbb{E}[g(X)u \mid X] = g(X)\mathbb{E}[u \mid X] = 0.$$

因此 $\mathbb{E}[g(X)u] = 0$。这说明若 $m(X) = \mathbb{E}[Y \mid X]$，则误差 $u = Y - m(X)$ 与所有 $X$ 的函数正交。条件期望正是 $Y$ 在所有 $X$ 函数构成的空间上的投影。

**E4.** 即使 $\boldsymbol{X}'\boldsymbol{X}$ 不可逆，$\mathrm{Col}(\boldsymbol{X})$ 仍是 $\mathbb{R}^n$ 的一个闭线性子空间。由投影定理，$\boldsymbol{y}$ 在 $\mathrm{Col}(\boldsymbol{X})$ 上存在唯一投影 $\hat{\boldsymbol{y}}$。可能有多个 $\hat{\beta}$ 使 $\boldsymbol{X}\hat{\beta} = \hat{\boldsymbol{y}}$，但拟合值唯一。任何使平方误差最小的 $\hat{\beta}$ 都对应这个投影点。

**E5.** 平方损失可微，导数会给出线性的一阶条件 $X'(y - X\beta) = 0$，因此产生正交结构。绝对值损失在零点不可微，一阶条件涉及符号函数或次梯度:

$$\sum_i x_i \,\mathrm{sign}(y_i - x_i'\beta) = 0$$

在合适解释下成立，但这不是残差本身与解释变量正交。绝对值损失通向 LAD regression 和 median regression，也与 quantile regression 密切相关。

# 18 本章总结图谱：概念–公式–应用–错误

| 概念 | 核心公式 | 应用 | 常见错误 |
|---|---|---|---|
| 内积 | $\langle u, v\rangle = u'v$ 或 $\int fg$ | 相似性、暴露、相关方向 | 把内积当成普通乘法 |
| 范数 | $\|u\| = \sqrt{u'u}$ | 距离、误差大小 | 忘记平方误差的几何含义 |
| 一维投影 | $\hat{b} = \frac{x'y}{x'x}$ | 单因子 beta | 忽略分母为零问题 |
| OLS | $\hat{\beta} = (X'X)^{-1}X'y$ | 回归、预测、拟合 | 只背公式不理解正交 |
| 正交条件 | $X'\hat{\varepsilon} = 0$ | 样本矩条件 | 误当作外生性证明 |
| 投影矩阵 | $P = X(X'X)^{-1}X'$ | fitted values | 忘记满列秩条件 |
| 残差矩阵 | $M = I - P$ | residualization, FWL | 不知道它也是投影矩阵 |
| 总体投影 | $\beta^* = E[XX']^{-1}E[XY]$ | population regression | 混淆样本与总体 |
| 函数投影 | $\int r(x)\phi_j(x)dx = 0$ | Fourier、非参数回归 | 以为和 OLS 无关 |
| Fourier 系数 | $a_n = \frac{1}{\pi}\int f\cos nx$, $b_n = \frac{1}{\pi}\int f\sin nx$ | 周期、频域、时间序列 | 只记公式不知投影含义 |

**本章压缩总结**

本章最重要的结论有四个：

1. 投影是最小平方误差近似；
2. 最优投影的残差与解释空间正交；
3. OLS 是 $\boldsymbol{y}$ 到 $\mathrm{Col}(\boldsymbol{X})$ 上的投影；
4. Fourier 展开、因子模型、条件期望和非参数回归都可以用投影语言理解。

如果你真正理解这四点，后面学计量经济学的 OLS、IV、GMM、条件期望、非参数估计，以及金融里的因子模型和风险分解，都会轻松很多。

# 经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

---

# Chapter 9

## 傅里叶级数与频域直觉

---

本章把 Chapter 8 的“投影”思想推进到周期函数和时间序列：Fourier 展开不是神秘公式，而是在正弦、余弦构成的函数空间里做最小平方投影。对经济金融来说，它提供了理解周期、趋势、噪声、商业周期和频域分析的入口。

math4econ personalized notes

2026

# 目录

**Chapter 9 总览** **3**

**1 本章为什么重要** **4**

**2 最小前置知识** **5**

**3 从投影到 Fourier 系数** **6**
3.1 函数空间里的内积 . . . . . 6
3.2 三角函数族的正交性 . . . . . 6
3.3 Fourier 级数的形式 . . . . . 7
3.4 系数公式来自投影 . . . . . 8

**4 奇偶性：最快的系数判断工具** **9**
4.1 偶函数只含余弦项 . . . . . 9
4.2 奇函数只含正弦项 . . . . . 9

**5 Worked Example 1：$f(x)=x$ 的 Fourier 级数** **10**
5.1 第一步：判断奇偶性 . . . . . 10
5.2 第二步：分部积分 . . . . . 10
5.3 第三步：写出级数 . . . . . 11

**6 Worked Example 2：$f(x)=x^2$ 的 Fourier 级数** **11**
6.1 第一步：奇偶性 . . . . . 12
6.2 第二步：计算常数项 . . . . . 12
6.3 第三步：计算余弦系数 . . . . . 12
6.4 第四步：写出级数 . . . . . 13

**7 一般区间 $[-L,L]$ 上的 Fourier 级数** **13**

**8 正弦级数与余弦级数：半区间展开** **15**
8.1 偶延拓与余弦级数 . . . . . 15
8.2 奇延拓与正弦级数 . . . . . 15

**9 复指数形式：把正弦余弦写得更统一** **15**

**10 收敛：Fourier 级数到底收敛到什么** **16**
10.1 连续点与跳跃点 . . . . . 16
10.2 均方收敛 . . . . . 17

**11 Gibbs phenomenon：跳跃点附近为什么会过冲** **17**

**12 Parseval identity：时域能量与频域能量** **18**
12.1 定理表述 . . . . . 18
12.2 从投影看 Parseval . . . . . 18
12.3 经济金融直觉 . . . . . 19

**13 频率、周期、振幅与相位** **19**
13.1 基本波的四个参数 . . . . . 19
13.2 相位如何进入正弦和余弦系数 . . . . . 20

**14 经济金融应用 1：商业周期、季节性与频率** **20**

**15 经济金融应用 2：金融高频数据中的噪声** **21**

**16 经济金融应用 3：因子模型与频率分解的类比** **21**

**17 计算模板：拿到一个 Fourier 题怎么做** **23**

**18 常见误区与纠正** **25**

**19 本章小结图谱** **26**

**20 练习题** **27**
A 组：基础计算 . . . . . 27
B 组：概念判断 . . . . . 27
C 组：混合题 . . . . . 27
D 组：经济金融建模题 . . . . . 28
E 组：挑战/证明题 . . . . . 28

**21 练习题详细解答** **29**
A 组答案 . . . . . 29
B 组答案 . . . . . 31
C 组答案 . . . . . 32
D 组答案 . . . . . 34
E 组答案 . . . . . 35

**22 本章到下一章的连接** **38**

# Chapter 9 总览

**本章目标**

学完本章，你应该达到以下目标：

1. 理解 Fourier 级数是 Chapter 8 投影思想在函数空间中的具体实现；
2. 能写出在 $[-\pi, \pi]$ 上的 Fourier 系数公式，并知道它们为什么来自正交投影；
3. 能区分一般 Fourier 级数、正弦级数、余弦级数、复指数形式；
4. 会利用奇偶性快速判断哪些系数为零；
5. 能计算几个标准函数的 Fourier 展开，例如 $f(x) = x$、$f(x) = x^2$、方波、三角波的基本版本；
6. 理解 Parseval identity：时域平方能量等于频域系数平方和；
7. 理解 Gibbs phenomenon 的含义：跳跃点附近会出现稳定的过冲，不是计算错误；
8. 能把频率、周期、振幅、相位与经济金融里的商业周期、季节性、噪声和高频数据联系起来；
9. 为后续时间序列、spectral density、滤波、PDE 和金融信号处理建立直觉入口。

**直觉**

一句话：**Fourier 级数就是把复杂周期函数分解成很多个简单波动。**
Chapter 8 告诉我们，回归是把 $y$ 投影到解释变量张成的空间。Fourier 级数做的是同一件事，只是解释变量换成了函数：

$$1, \quad \cos x, \quad \sin x, \quad \cos 2x, \quad \sin 2x, \quad \dots$$

它问的问题是：如果我只能使用这些不同频率的正弦、余弦波来近似 $f(x)$，最好的系数是什么？答案就是 Fourier coefficients。

**本章压缩总结**

本章主线可以压缩成：

正交投影 $\Rightarrow$ 三角基函数 $\Rightarrow$ Fourier 系数 $\Rightarrow$ 频率分解 $\Rightarrow$ 周期、噪声、滤波、时间序列

你需要记住的不是一堆孤立公式，而是一个思想：**从时域看函数怎样随时间变化；从频域看它由哪些周期成分组成**。

# 1 本章为什么重要

在前一章，我们把投影、最小二乘、OLS 和函数空间近似统一起来。本章进一步讨论一个特殊但极其重要的函数空间：由正弦和余弦构成的周期函数空间。它看上去像数学分析里的技巧章节，但对经济金融并不边缘。

经济金融中大量对象本身就是时间序列：GDP、通胀、利率、股票收益率、成交量、汇率、失业率、消费、投资、财政收入。我们经常想问：

- 一个序列里有没有周期?
- 周期是短周期还是长周期?
- 商业周期和季节性怎样区分?
- 股票高频收益里哪些是低频趋势，哪些是高频噪声?
- 平滑、滤波、去趋势到底在数学上做了什么?

Fourier 的基本思想就是：不要只在时间轴上看 $x_t$ 怎么上下波动，还要问它是由哪些频率的波动叠加出来的。用一句话说：

time domain asks “when”; frequency domain asks “how fast”.

在宏观经济学里，商业周期不是每天震荡，而是低频波动；季节性可能是一年一次或者一季度一次；金融市场的微观结构噪声常常表现为高频成分。Fourier 不是让你立刻做复杂频谱估计，而是给你一种语言：**复杂变化可以被拆成不同频率的基本波**。

**常见误区**

这章容易学歪。常见错误是把 Fourier 级数当成“背公式然后算积分”的章节。真正重要的是：

1. 系数公式来自正交投影；
2. 不同频率的正弦、余弦彼此正交；
3. 低频对应慢变化，高频对应快变化；
4. 近似项数增加时，不连续点附近会出现 Gibbs 过冲；

5. 频域分析是时间序列和金融信号处理的入口。

## 2 最小前置知识

本章需要你回忆四类知识。

1. **积分与内积**。在函数空间里，内积通常定义为

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,\mathrm{d}x.$$

   它和向量点积 $\boldsymbol{u}'\boldsymbol{v}$ 是同一思想。

2. **正交投影**。如果 $f$ 要投影到一个基函数 $\phi$ 张成的一维空间，最优系数是

$$c = \frac{\langle f, \phi\rangle}{\langle \phi, \phi\rangle}.$$

3. **三角函数基本性质**。你需要熟悉 $\sin(nx)$、$\cos(nx)$ 的周期、奇偶性和简单积分。

4. **级数直觉**。 Fourier 级数是一种函数级数。它的收敛比普通数字级数更复杂：可能逐点收敛，也可能均方收敛，也可能在跳跃点收敛到左右极限平均值。

**方法**

本章的学习方式：先把 Fourier 看成投影问题，再练系数计算，最后再理解频率和经济金融解释。不要先背一堆特殊展开式。

# 3 从投影到 Fourier 系数

## 3.1 函数空间里的内积

在 $[-\pi,\pi]$ 上，我们把两个函数的内积定义为：

$$\langle f,g\rangle=\int_{-\pi}^{\pi} f(x)g(x)\,\mathrm{d}x.$$

对应的范数是：

$$\|f\|=\sqrt{\langle f,f\rangle}=\left(\int_{-\pi}^{\pi} f(x)^2\,\mathrm{d}x\right)^{1/2}.$$

如果 $\langle f,g\rangle=0$，就说 $f$ 和 $g$ 正交。

这和向量空间完全平行。有限维向量里，$\boldsymbol{u}'\boldsymbol{v}=0$ 表示两个方向垂直；函数空间里，$\int fg=0$ 表示两个函数在整体上互不相关，或者说没有共同方向。

**定义 9.1：函数空间内积与均方距离**

在 $[-\pi,\pi]$ 上，两个平方可积函数 $f,g$ 的均方距离可以写成

$$\|f-g\|^2=\int_{-\pi}^{\pi}[f(x)-g(x)]^2\,\mathrm{d}x.$$

Fourier 部分和 OLS 一样，默认使用平方误差作为近似标准。

## 3.2 三角函数族的正交性

Fourier 级数之所以有效，是因为以下函数族在 $[-\pi,\pi]$ 上彼此正交：

$$1,\quad \cos x,\quad \sin x,\quad \cos 2x,\quad \sin 2x,\quad \dots$$

它们满足：

$$\int_{-\pi}^{\pi}\cos(nx)\cos(mx)\,\mathrm{d}x=\begin{cases}0, & n\neq m,\\ \pi, & n=m\geq 1,\end{cases}$$

$$\int_{-\pi}^{\pi}\sin(nx)\sin(mx)\,\mathrm{d}x=\begin{cases}0, & n\neq m,\\ \pi, & n=m\geq 1,\end{cases}$$

$$\int_{-\pi}^{\pi}\sin(nx)\cos(mx)\,\mathrm{d}x=0.$$

此外，

$$\int_{-\pi}^{\pi}1^2\,\mathrm{d}x=2\pi,\qquad \int_{-\pi}^{\pi}\cos(nx)\,\mathrm{d}x=0,\qquad \int_{-\pi}^{\pi}\sin(nx)\,\mathrm{d}x=0.$$

**证明骨架：为什么不同频率正交**

利用积化和差公式：

$$\cos(nx)\cos(mx)=\frac{1}{2}\{\cos[(n-m)x]+\cos[(n+m)x]\},$$

$$\sin(nx)\sin(mx)=\frac{1}{2}\{\cos[(n-m)x]-\cos[(n+m)x]\},$$

$$\sin(nx)\cos(mx)=\frac{1}{2}\{\sin[(n+m)x]+\sin[(n-m)x]\}.$$

当 $k\neq 0$ 时，

$$\int_{-\pi}^{\pi}\cos(kx)\,\mathrm{d}x=0,\qquad \int_{-\pi}^{\pi}\sin(kx)\,\mathrm{d}x=0.$$

因此，不同频率的项积分后抵消。只有 $n=m$ 时，$\cos[(n-m)x]=\cos 0=1$，才留下非零部分。

## 3.3 Fourier 级数的形式

如果一个周期为 $2\pi$ 的函数 $f$ 足够好，我们希望把它写成：

$$f(x)\sim\frac{a_0}{2}+\sum_{n=1}^{\infty}\left(a_n\cos nx+b_n\sin nx\right).$$

这里的 $a_n,b_n$ 就是 Fourier 系数。

为什么常数项写成 $a_0/2$，而不是 $a_0$？因为这样系数公式会统一为：

$$a_0=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\,\mathrm{d}x,$$

$$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos(nx)\,\mathrm{d}x,\qquad n\geq 1,$$

$$b_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin(nx)\,\mathrm{d}x,\qquad n\geq 1.$$

**定义 9.2：Fourier 系数**

对 $f\in L^2[-\pi,\pi]$，其 Fourier 系数定义为

$$a_0=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\,\mathrm{d}x,$$

$$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos(nx)\,\mathrm{d}x,\qquad b_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin(nx)\,\mathrm{d}x.$$

形式上的 Fourier 级数为

$$f(x)\sim\frac{a_0}{2}+\sum_{n=1}^{\infty}\left(a_n\cos nx+b_n\sin nx\right).$$

## 3.4 系数公式来自投影

假设我们只用前 $N$ 个频率近似 $f$：

$$S_N(x) = \frac{a_0}{2} + \sum_{n=1}^{N} (a_n \cos nx + b_n \sin nx)\,.$$

最好的近似应该最小化

$$\int_{-\pi}^{\pi} [f(x) - S_N(x)]^2 \,\mathrm{d}x.$$

这就是函数空间中的最小二乘。

由于 $1, \cos nx, \sin nx$ 彼此正交，最优条件是残差 $f - S_N$ 与所有基函数正交：

$$\langle f - S_N, 1 \rangle = 0,$$

$$\langle f - S_N, \cos kx \rangle = 0, \qquad \langle f - S_N, \sin kx \rangle = 0, \quad k = 1, \dots, N.$$

例如对 $\cos kx$：

$$0 = \int_{-\pi}^{\pi} [f(x) - S_N(x)] \cos kx \,\mathrm{d}x.$$

展开后，所有与 $\cos kx$ 正交的项都消失，只剩下

$$\int_{-\pi}^{\pi} f(x) \cos kx \,\mathrm{d}x - a_k \int_{-\pi}^{\pi} \cos^2 kx \,\mathrm{d}x = 0.$$

因为 $\int_{-\pi}^{\pi} \cos^2 kx \,\mathrm{d}x = \pi$，所以

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx \,\mathrm{d}x.$$

$b_k$ 的推导同理。

**直觉**

Fourier 系数不是凭空给出的积分公式，而是 OLS 的函数空间版本：

$$\text{残差和解释变量正交} \quad \Longleftrightarrow \quad \text{残差和每个基函数正交.}$$

回归里的 $X'\hat{\varepsilon} = 0$，在 Fourier 里变成

$$\int_{-\pi}^{\pi} (f - S_N) \cos nx \,\mathrm{d}x = 0, \qquad \int_{-\pi}^{\pi} (f - S_N) \sin nx \,\mathrm{d}x = 0.$$

# 4 奇偶性：最快的系数判断工具

## 4.1 偶函数只含余弦项

如果 $f(-x)=f(x)$，则 $f$ 是偶函数。因为 $\cos nx$ 是偶函数，$\sin nx$ 是奇函数，所以

$$f(x)\sin nx$$

是奇函数。奇函数在对称区间上的积分为零，因此

$$b_n=0.$$

所以偶函数的 Fourier 级数只含余弦项：

$$f(x)\sim\frac{a_0}{2}+\sum_{n=1}^{\infty}a_n\cos nx.$$

## 4.2 奇函数只含正弦项

如果 $f(-x)=-f(x)$，则 $f$ 是奇函数。此时 $f(x)\cos nx$ 是奇函数，所以

$$a_0=0,\qquad a_n=0.$$

所以奇函数的 Fourier 级数只含正弦项：

$$f(x)\sim\sum_{n=1}^{\infty}b_n\sin nx.$$

**方法**

计算 Fourier 级数前，先做三步：

1. 看区间是不是 $[-\pi,\pi]$ 或者可以缩放到标准区间；
2. 判断 $f$ 是奇函数、偶函数还是都不是；
3. 决定只算 $a_n$、只算 $b_n$，还是都要算。

这一步可以省掉一半以上计算。

**例 9.1：判断哪些系数为零**

判断以下函数在 $[-\pi,\pi]$ 上的 Fourier 级数形式。

(1) $f(x)=x$；

(2) $f(x)=x^2$；

(3) $f(x)=x+1$；

(4) $f(x)=|x|$。

**解。**

(1) $x$ 是奇函数，所以只含正弦项。

$$x \sim \sum_{n=1}^{\infty} b_n \sin nx.$$

(2) $x^2$ 是偶函数，所以只含常数项和余弦项。

$$x^2 \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx.$$

(3) $x+1$ 既有奇部分 $x$，又有偶部分 $1$，所以既可能有常数项，也可能有正弦项；余弦项除常数外为零。

(4) $|x|$ 是偶函数，所以只含常数项和余弦项。

# 5 Worked Example 1：$f(x)=x$ 的 Fourier 级数

在 $(-\pi,\pi)$ 上考虑 $f(x)=x$，并周期延拓。

## 5.1 第一步：判断奇偶性

$f(x)=x$ 是奇函数，因此

$$a_0=0, \qquad a_n=0.$$

只需要算 $b_n$：

$$b_n=\frac{1}{\pi}\int_{-\pi}^{\pi} x\sin(nx)\,\mathrm{d}x.$$

由于 $x\sin(nx)$ 是偶函数，

$$b_n=\frac{2}{\pi}\int_{0}^{\pi} x\sin(nx)\,\mathrm{d}x.$$

## 5.2 第二步：分部积分

令

$$u=x, \qquad \mathrm{d}v=\sin(nx)\,\mathrm{d}x.$$

则

$$\mathrm{d}u=\mathrm{d}x, \qquad v=-\frac{\cos(nx)}{n}.$$

所以

$$\int_0^{\pi} x\sin(nx)\,\mathrm{d}x=\left[-\frac{x\cos(nx)}{n}\right]_0^{\pi}+\frac{1}{n}\int_0^{\pi}\cos(nx)\,\mathrm{d}x.$$

第二项为

$$\frac{1}{n}\left[\frac{\sin(nx)}{n}\right]_0^{\pi} = 0.$$

第一项为

$$-\frac{\pi\cos(n\pi)}{n} = -\frac{\pi(-1)^n}{n}.$$

因此

$$\int_0^{\pi} x\sin(nx)\,\mathrm{d}x = -\frac{\pi(-1)^n}{n}.$$

于是

$$b_n = \frac{2}{\pi}\left[-\frac{\pi(-1)^n}{n}\right] = \frac{2(-1)^{n+1}}{n}.$$

### 5.3 第三步：写出级数

所以

$$\boxed{x \sim 2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\sin(nx), \qquad -\pi < x < \pi.}$$

即

$$x \sim 2\left(\sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x - \frac{1}{4}\sin 4x + \cdots\right).$$

**直觉**

这个例子很重要：一个看似“直线”的函数，在周期延拓后变成锯齿波。锯齿波可以由很多正弦波叠加出来。低频项给出大致形状，高频项负责修正尖锐跳跃。

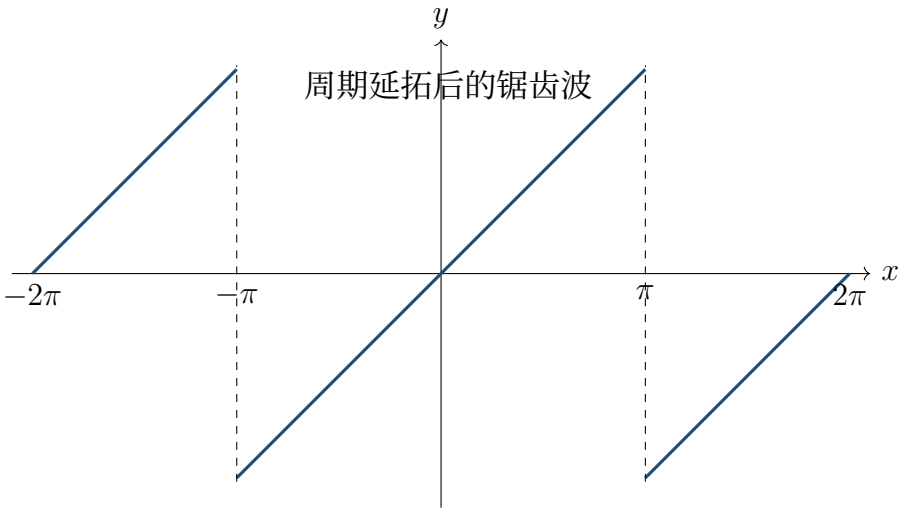



图 1：函数 $f(x) = x$ 在 $(-\pi, \pi)$ 上的周期延拓

## 6 Worked Example 2：$f(x) = x^2$ 的 Fourier 级数

考虑 $f(x) = x^2$，$-\pi < x < \pi$。

## 6.1 第一步：奇偶性

$x^2$ 是偶函数，所以

$$b_n = 0.$$

只需要计算 $a_0$ 和 $a_n$。

## 6.2 第二步：计算常数项

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} x^2\,\mathrm{d}x = \frac{2}{\pi}\int_{0}^{\pi} x^2\,\mathrm{d}x = \frac{2}{\pi}\cdot\frac{\pi^3}{3} = \frac{2\pi^2}{3}.$$

所以常数项为

$$\frac{a_0}{2} = \frac{\pi^2}{3}.$$

## 6.3 第三步：计算余弦系数

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x^2\cos(nx)\,\mathrm{d}x = \frac{2}{\pi}\int_{0}^{\pi} x^2\cos(nx)\,\mathrm{d}x.$$

对

$$I_n = \int_{0}^{\pi} x^2\cos(nx)\,\mathrm{d}x$$

做分部积分。令

$$u = x^2, \qquad \mathrm{d}v = \cos(nx)\,\mathrm{d}x.$$

则

$$\mathrm{d}u = 2x\,\mathrm{d}x, \qquad v = \frac{\sin(nx)}{n}.$$

于是

$$I_n = \left[\frac{x^2\sin(nx)}{n}\right]_0^{\pi} - \frac{2}{n}\int_{0}^{\pi} x\sin(nx)\,\mathrm{d}x.$$

第一项为零，因为 $\sin(n\pi) = 0$。而前面已经算过

$$\int_{0}^{\pi} x\sin(nx)\,\mathrm{d}x = -\frac{\pi(-1)^n}{n}.$$

所以

$$I_n = -\frac{2}{n}\left[-\frac{\pi(-1)^n}{n}\right] = \frac{2\pi(-1)^n}{n^2}.$$

因此

$$a_n = \frac{2}{\pi}I_n = \frac{2}{\pi}\cdot\frac{2\pi(-1)^n}{n^2} = \frac{4(-1)^n}{n^2}.$$

### 6.4 第四步：写出级数

$$\boxed{x^2 \sim \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}\cos(nx).}$$

把 $x=\pi$ 代入，可以得到一个经典级数恒等式。由于周期延拓在 $x=\pi$ 连续地取 $\pi^2$，有

$$\pi^2 = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty} \frac{1}{n^2},$$

因此

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

**常见误区**

这里的代入不是单纯“把任意点塞进去”。它依赖 Fourier 级数在该点收敛到函数值。一般而言，在跳跃点处，Fourier 级数收敛到左右极限的平均值。

## 7 一般区间 $[-L, L]$ 上的 Fourier 级数

经济金融中的周期变量未必以 $2\pi$ 为周期。比如季度数据的季节性可能周期为 $4$，月度数据的年度季节性周期为 $12$。这时需要把区间从 $[-\pi,\pi]$ 推广到 $[-L, L]$。

如果 $f$ 定义在 $[-L, L]$，周期延拓的周期是 $2L$，则 Fourier 级数写成

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[ a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \sin\left(\frac{n\pi x}{L}\right) \right].$$

系数为

$$a_0 = \frac{1}{L}\int_{-L}^{L} f(x)\,\mathrm{d}x,$$

$$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)\,\mathrm{d}x,$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right)\,\mathrm{d}x.$$

**直觉**

在 $[-\pi,\pi]$ 上，基函数是 $\sin nx$、$\cos nx$。在 $[-L, L]$ 上，基函数变成

$$\sin\left(\frac{n\pi x}{L}\right), \qquad \cos\left(\frac{n\pi x}{L}\right).$$

原因是：这些函数的周期为 $2L/n$，刚好能在长度 $2L$ 的区间里放下整数个完整波动。

**例 9.2：月度季节性如何写成 Fourier 项**

假设 $t$ 是月份编号，一个年度周期为 12 的季节性可以用哪些基函数表达?

**解**。周期为 12，对应基本角频率为

$$\omega_1 = \frac{2\pi}{12} = \frac{\pi}{6}.$$

因此可用

$$\cos\left(\frac{2\pi nt}{12}\right), \qquad \sin\left(\frac{2\pi nt}{12}\right), \qquad n = 1, 2, \ldots$$

例如只保留前两个频率，可以写成

$$s_t = \alpha_1 \cos\left(\frac{2\pi t}{12}\right) + \beta_1 \sin\left(\frac{2\pi t}{12}\right) + \alpha_2 \cos\left(\frac{4\pi t}{12}\right) + \beta_2 \sin\left(\frac{4\pi t}{12}\right).$$

这就是把季节性建模为若干个周期波动的叠加。

# 8 正弦级数与余弦级数：半区间展开

很多时候，函数只定义在 $[0,L]$ 上。比如年龄、收入、财富、时间长度、利率期限等变量自然是非负的。Fourier 方法仍然可以用，但需要先选择一种延拓方式。

## 8.1 偶延拓与余弦级数

如果把 $[0,L]$ 上的 $f$ 偶延拓到 $[-L,L]$，就得到余弦级数：

$$f(x)\sim\frac{a_0}{2}+\sum_{n=1}^{\infty}a_n\cos\left(\frac{n\pi x}{L}\right),\qquad 0<x<L,$$

其中

$$a_0=\frac{2}{L}\int_0^L f(x)\,\mathrm{d}x,$$

$$a_n=\frac{2}{L}\int_0^L f(x)\cos\left(\frac{n\pi x}{L}\right)\,\mathrm{d}x.$$

## 8.2 奇延拓与正弦级数

如果把 $[0,L]$ 上的 $f$ 奇延拓到 $[-L,L]$，就得到正弦级数：

$$f(x)\sim\sum_{n=1}^{\infty}b_n\sin\left(\frac{n\pi x}{L}\right),\qquad 0<x<L,$$

其中

$$b_n=\frac{2}{L}\int_0^L f(x)\sin\left(\frac{n\pi x}{L}\right)\,\mathrm{d}x.$$

**方法**

半区间展开的选择直觉：

- 如果你想让端点处像“无斜率”那样平滑，常用余弦展开；
- 如果你希望端点处取零或符合边界吸收条件，常用正弦展开；
- 在 PDE 里，正弦/余弦展开往往对应不同边界条件；
- 在经济金融中，它可以作为用周期基函数近似非周期函数的一种技术。

# 9 复指数形式：把正弦余弦写得更统一

利用 Euler 公式：

$$\mathrm{e}^{\mathrm{i}nx}=\cos nx+\mathrm{i}\sin nx.$$

Fourier 级数也可以写成复指数形式：

$$f(x)\sim\sum_{n=-\infty}^{\infty}c_n\mathrm{e}^{\mathrm{i}nx},$$

其中

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\mathrm{e}^{-\mathrm{i}nx}\,\mathrm{d}x.$$

这和实数形式等价。二者关系为：

$$c_0 = \frac{a_0}{2},$$

$$c_n = \frac{a_n - \mathrm{i}b_n}{2}, \qquad c_{-n} = \frac{a_n + \mathrm{i}b_n}{2}, \qquad n \geq 1.$$

**直觉**

复指数形式在数学金融、信号处理和时间序列里很常用，因为指数函数在微分、积分、平移、卷积下都很方便。你现在不需要把复分析学完，但要知道：复指数形式不是另一套理论，而是正弦余弦形式的更紧凑写法。

# 10 收敛：Fourier 级数到底收敛到什么

## 10.1 连续点与跳跃点

Fourier 级数不是在所有情况下都逐点收敛到原函数，但在常见情形下有一个非常重要的结论。

**定理 9.1：Dirichlet 型收敛结论，直观版**

如果 $f$ 在 $[-\pi,\pi]$ 上分段光滑，并作 $2\pi$ 周期延拓，那么它的 Fourier 级数在点 $x$ 收敛到

$$\frac{f(x+) + f(x-)}{2}.$$

如果 $f$ 在 $x$ 连续，则这个值就是 $f(x)$。如果 $x$ 是跳跃点，则收敛到左右极限的平均值。

**例 9.3：方波在跳跃点收敛到哪里**

定义方波

$$f(x) = \begin{cases} 1, & 0 < x < \pi, \\ -1, & -\pi < x < 0. \end{cases}$$

并作 $2\pi$ 周期延拓。问：Fourier 级数在 $x = 0$ 收敛到哪里？

**解**。在 $x = 0$，左极限为

$$f(0-) = -1,$$

右极限为

$$f(0+) = 1.$$

所以 Fourier 级数收敛到

$$\frac{f(0-)+f(0+)}{2}=0.$$

这并不意味着原函数在 $0$ 处必须等于零，而是 Fourier 级数在跳跃点采用左右平均值。

## 10.2 均方收敛

Fourier 级数更稳健的解释是均方收敛。也就是说，部分和 $S_N$ 可以使

$$\int_{-\pi}^{\pi}|f(x)-S_N(x)|^2\,\mathrm{d}x$$

越来越小。即使在个别点附近看起来有误差，从整体平方误差角度，Fourier 近似仍然是最优投影。

这和 OLS 非常相似：回归并不要求每一个样本点都拟合完美，而是使平方残差总和最小。

# 11 Gibbs phenomenon：跳跃点附近为什么会过冲

对于带跳跃的函数，比如方波，Fourier 部分和在跳跃点附近会出现振荡和过冲。随着 $N$ 增大，振荡区域变窄，但过冲高度不会完全消失。这叫 Gibbs phenomenon。

**直觉**

Gibbs 现象的直觉：正弦和余弦都是连续光滑的波。你试图用很多光滑波叠加出一个突然跳跃的函数，光滑波只能在跳跃附近“努力转弯”，于是会出现过冲和振荡。

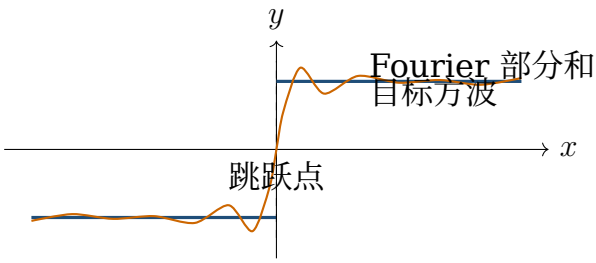



图 2: Gibbs phenomenon 的示意图：过冲区域变窄，但过冲比例不会完全消失

**常见误区**

如果你用 Fourier 级数近似带跳跃的经济金融信号，比如政策冲击、交易停牌后的价格跳变、制度断点，那么跳跃点附近的振荡未必代表真实经济机制，可能是数学近似造成的。

# 12 Parseval identity：时域能量与频域能量

## 12.1 定理表述

如果

$$f(x)\sim\frac{a_0}{2}+\sum_{n=1}^{\infty}(a_n\cos nx+b_n\sin nx),$$

则 Parseval identity 表示：

$$\boxed{\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)^2\,\mathrm{d}x=\frac{a_0^2}{2}+\sum_{n=1}^{\infty}(a_n^2+b_n^2).}$$

这就是函数空间里的 Pythagorean theorem。左边是函数在时域中的总平方大小；右边是它在各个正交频率方向上的平方系数之和。

## 12.2 从投影看 Parseval

在有限维空间里，如果 $\{e_1,\dots,e_k\}$ 是正交基，那么

$$\|x\|^2=\sum_j\frac{\langle x,e_j\rangle^2}{\|e_j\|^2}.$$

Fourier 里，基函数 $1,\cos nx,\sin nx$ 是正交的，所以函数平方范数可以拆成各频率贡献。

**例 9.4：用 Parseval 得到一个级数和**

已知

$$x\sim2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\sin nx.$$

用 Parseval 求 $\sum_{n=1}^{\infty}1/n^2$。

**解。**这里 $a_0=0$、$a_n=0$、$b_n=2(-1)^{n+1}/n$。Parseval 给出

$$\frac{1}{\pi}\int_{-\pi}^{\pi}x^2\,\mathrm{d}x=\sum_{n=1}^{\infty}b_n^2.$$

左边为

$$\frac{1}{\pi}\cdot\frac{2\pi^3}{3}=\frac{2\pi^2}{3}.$$

右边为

$$\sum_{n=1}^{\infty}\frac{4}{n^2}=4\sum_{n=1}^{\infty}\frac{1}{n^2}.$$

因此

$$\frac{2\pi^2}{3}=4\sum_{n=1}^{\infty}\frac{1}{n^2},$$

所以

$$\boxed{\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{\pi^2}{6}.}$$

## 12.3 经济金融直觉

Parseval identity 对经济金融的直觉是：一个时间序列的总波动可以分解到不同频率上。低频成分可能对应长期趋势或商业周期，高频成分可能对应噪声、短期冲击、微观结构扰动。

更高级的时间序列课程会用 spectral density 来描述不同频率对方差的贡献。本章只需要记住一个入口：

$$\text{variance over time} \quad \approx \quad \text{energy distributed over frequencies.}$$

# 13 频率、周期、振幅与相位

## 13.1 基本波的四个参数

一个简单周期波可以写成

$$A\cos(\omega t+\phi).$$

其中：

- $A$：振幅，表示波动幅度；
- $\omega$：角频率，表示单位时间转过多少弧度；
- $T=2\pi/\omega$：周期，表示完成一次完整波动需要多久；
- $\phi$：相位，表示波动整体向左或向右平移。

  在 Fourier 级数里，$\cos nt$、$\sin nt$ 对应角频率 $n$。$n$ 越大，波动越快，周期越短。

**例 9.5：季度数据中的周期项**

假设 $t$ 以季度为单位。一个周期为 4 个季度的季节性项可以写成什么？一个周期为 8 个季度的商业周期项可以写成什么？

**解**。周期 $T$ 和角频率 $\omega$ 的关系是

$$\omega=\frac{2\pi}{T}.$$

周期为 4 的季节性项可以写成

$$A\cos\left(\frac{2\pi}{4}t+\phi\right)=A\cos\left(\frac{\pi}{2}t+\phi\right).$$

周期为 8 的项可以写成

$$B\cos\left(\frac{2\pi}{8}t+\psi\right)=B\cos\left(\frac{\pi}{4}t+\psi\right).$$

频率越低，周期越长；频率越高，周期越短。

## 13.2 相位如何进入正弦和余弦系数

利用三角恒等式:

$$A\cos(nt-\phi)=A\cos\phi\cos nt+A\sin\phi\sin nt.$$

所以 Fourier 形式中的

$$a_n\cos nt+b_n\sin nt$$

可以写成

$$R_n\cos(nt-\phi_n),$$

其中

$$R_n=\sqrt{a_n^2+b_n^2},$$

$$\cos\phi_n=\frac{a_n}{R_n},\qquad \sin\phi_n=\frac{b_n}{R_n}.$$

这说明 $a_n,b_n$ 两个系数合起来描述了该频率成分的振幅和相位。

# 14 经济金融应用 1：商业周期、季节性与频率

假设一个宏观变量 $y_t$ 可以粗略写成

$$y_t=\text{trend}_t+\text{cycle}_t+\text{seasonal}_t+\text{noise}_t.$$

从频率角度:

- trend 是非常低频的变化;
- business cycle 是中低频变化;
- seasonal component 是固定周期的频率成分;
- noise 往往包含大量高频成分。

  如果我们用 Fourier 项近似季节性，可以写成

$$y_t=\alpha+\beta t+hoy_{t-1}+\sum_{n=1}^{N}\left[a_n\cos\left(\frac{2\pi nt}{s}\right)+b_n\sin\left(\frac{2\pi nt}{s}\right)\right]+u_t,$$

其中 $s=4$ 表示季度季节性，$s=12$ 表示月度年度季节性。

**直觉**

为什么实际回归中常用 sine/cosine 季节项？因为它比单纯月份 dummy 更平滑，也能把季节性看成周期信号。它不是必须替代 dummy，而是提供了另一种可解释的周期建模方式。

# 15 经济金融应用 2：金融高频数据中的噪声

金融高频数据常常包含不同尺度的信息：

- 低频部分：宏观趋势、基本面变化、风险溢价；
- 中频部分：交易日内模式、流动性周期；
- 高频部分：bid-ask bounce、订单簿微观结构噪声、短期冲击。

Fourier 思想告诉我们，可以把价格或收益信号分解为不同频率成分。虽然严格的金融频域分析需要离散 Fourier 变换和谱分析，但本章已经给出了核心直觉：

$$\text{观察到的价格变化} = \text{低频信息成分} + \text{高频噪声成分}.$$

这也是很多滤波和平滑方法的思想来源：保留低频成分，削弱高频噪声。

**常见误区**

频域滤波不是免费的。削弱高频噪声时，可能也会削弱真实的短期信息；保留低频趋势时，可能引入滞后。因此经济解释必须和数据生成机制结合，不能把数学滤波结果直接当成经济事实。

# 16 经济金融应用 3：因子模型与频率分解的类比

Chapter 8 里，我们把资产收益写成因子模型：

$$R_i = \alpha_i + \beta_{i1}F_1 + \cdots + \beta_{ik}F_k + \varepsilon_i.$$

这是把资产收益投影到因子空间。Fourier 级数则把函数投影到频率空间：

$$f(t) \approx \frac{a_0}{2} + \sum_{n=1}^{N} (a_n \cos nt + b_n \sin nt).$$

二者结构非常像：

| 对象 | 因子模型 | Fourier 近似 |
|---|---|---|
| 被解释对象 | 资产收益 $R_i$ | 时间函数 $f(t)$ |
| 解释基底 | 风险因子 $F_j$ | 正弦余弦基函数 |
| 系数 | factor loading $\beta_{ij}$ | Fourier coefficients $a_n, b_n$ |
| 误差 | idiosyncratic error | truncation error |
| 目标 | 解释横截面/时间收益 | 分解频率成分 |

这个类比能帮你理解：Fourier 不是孤立技巧，它是“选择一组基底来解释复杂对象”的又一个例子。

# 17 计算模板：拿到一个 Fourier 题怎么做

**方法**

标准流程：

1. **确认区间和周期**。是 $[-\pi,\pi]$、$[-L,L]$，还是 $[0,L]$ 半区间？
2. **判断奇偶性**。奇函数只算 $b_n$，偶函数只算 $a_0,a_n$。
3. **写系数公式**。不要直接凭记忆写结果。
4. **计算积分**。常用分部积分、积化和差、对称性。
5. **写出级数**。注意常数项是 $a_0/2$。
6. **判断收敛点**。连续点收敛到函数值，跳跃点收敛到左右平均。
7. **解释频率**。低频负责大轮廓，高频负责细节和跳跃修正。

**例 9.6：方波的 Fourier 级数**

设

$$f(x)=\begin{cases}1, & 0<x<\pi,\\ -1, & -\pi<x<0,\end{cases}$$

周期为 $2\pi$。求其 Fourier 级数。

**解**。这是奇函数，所以

$$a_0=0,\qquad a_n=0.$$

只算

$$b_n=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,\mathrm{d}x.$$

因为 $f(x)\sin nx$ 是偶函数，

$$b_n=\frac{2}{\pi}\int_0^{\pi} 1\cdot\sin nx\,\mathrm{d}x.$$

所以

$$b_n=\frac{2}{\pi}\left[-\frac{\cos nx}{n}\right]_0^{\pi}=\frac{2}{\pi}\cdot\frac{1-(-1)^n}{n}.$$

当 $n$ 为偶数时，$1-(-1)^n=0$；当 $n$ 为奇数时，$1-(-1)^n=2$。因此

$$b_n=\begin{cases}\dfrac{4}{\pi n}, & n\ \text{odd},\\ 0, & n\ \text{even}.\end{cases}$$

所以

$$\boxed{f(x) \sim \frac{4}{\pi}\left(\sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x + \cdots\right).}$$

**例 9.7：三角波的余弦展开**

令 $f(x) = |x|$，$-\pi < x < \pi$。求 Fourier 级数。

**解。** $f(x) = |x|$ 是偶函数，所以 $b_n = 0$。

先算

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} |x|\,\mathrm{d}x = \frac{2}{\pi}\int_0^{\pi} x\,\mathrm{d}x = \pi.$$

所以常数项是 $\pi/2$。

再算

$$a_n = \frac{2}{\pi}\int_0^{\pi} x\cos nx\,\mathrm{d}x.$$

分部积分：

$$\int_0^{\pi} x\cos nx\,\mathrm{d}x = \left[\frac{x\sin nx}{n}\right]_0^{\pi} - \frac{1}{n}\int_0^{\pi}\sin nx\,\mathrm{d}x.$$

第一项为零，第二项为

$$-\frac{1}{n}\left[-\frac{\cos nx}{n}\right]_0^{\pi} = \frac{\cos(n\pi) - 1}{n^2} = \frac{(-1)^n - 1}{n^2}.$$

因此

$$a_n = \frac{2}{\pi}\cdot\frac{(-1)^n - 1}{n^2}.$$

当 $n$ 为偶数时，$a_n = 0$；当 $n$ 为奇数时，

$$a_n = -\frac{4}{\pi n^2}.$$

所以

$$\boxed{|x| \sim \frac{\pi}{2} - \frac{4}{\pi}\left(\cos x + \frac{1}{3^2}\cos 3x + \frac{1}{5^2}\cos 5x + \cdots\right).}$$

# 18 常见误区与纠正

**常见误区**

**误区 1：看到 Fourier 就只背系数公式。**
纠正：系数公式来自投影。你要先想：基函数是什么？内积是什么？残差和哪些方向正交？

**常见误区**

**误区 2：忘记常数项是 $a_0/2$。**
纠正：系数 $a_0$ 的定义是 $a_0 = (1/\pi)\int f$，但级数里放的是 $a_0/2$。这是因为 1 的范数平方是 $2\pi$，而 $\cos nx$、$\sin nx$ 的范数平方是 $\pi$。

**常见误区**

**误区 3：在跳跃点强行认为 Fourier 级数等于原函数值。**
纠正：在跳跃点，常见收敛结果是左右极限平均：

$$\frac{f(x+) + f(x-)}{2}.$$

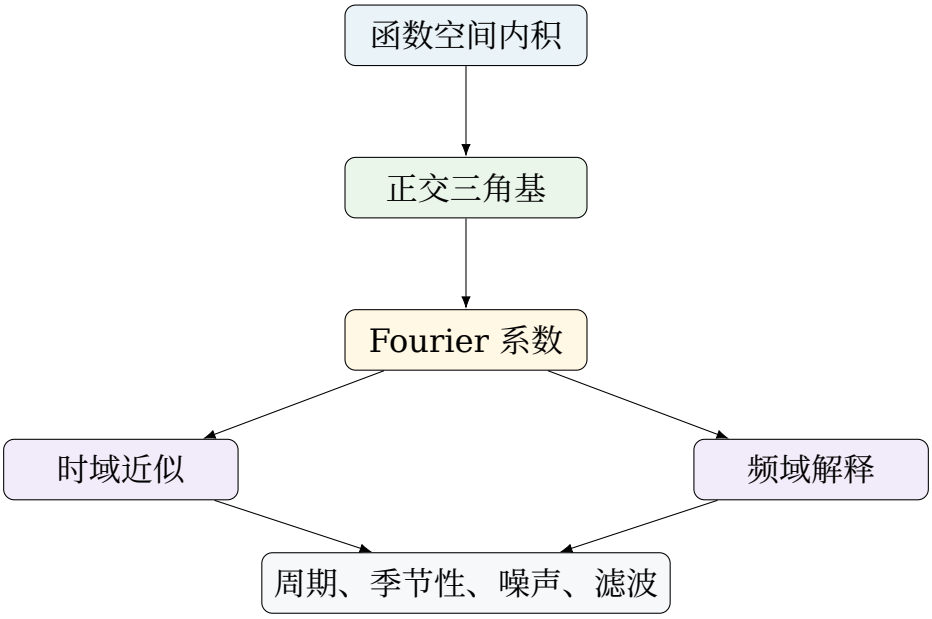
**常见误区**

**误区 4：把高频成分都解释成噪声。**
纠正：高频成分可能是噪声，也可能是真实短期信息。数学分解本身不自动给出经济解释。

**常见误区**

**误区 5：觉得 Fourier 和回归无关。**
纠正：二者都是投影。区别只是回归常用解释变量作为基底，Fourier 用正弦余弦作为基底。

# 19 本章小结图谱




**本章压缩总结**

本章最核心的 8 个结论：

1. Fourier 级数是函数空间中的正交投影；
2. $\sin nx$、$\cos nx$ 在 $[-\pi,\pi]$ 上彼此正交；
3. $a_n, b_n$ 是函数在第 $n$ 个频率方向上的投影系数；
4. 偶函数只含余弦项，奇函数只含正弦项；
5. 跳跃点处 Fourier 级数通常收敛到左右极限平均值；
6. Gibbs 现象说明用光滑波逼近跳跃会产生稳定过冲；
7. Parseval identity 把函数平方大小分解到各频率上；
8. 频域直觉是理解季节性、商业周期、高频噪声和滤波的入口。

# 20 练习题

## A 组：基础计算

A1. 证明 $\int_{-\pi}^{\pi} \sin(nx)\cos(mx)\,\mathrm{d}x = 0$。

A2. 证明 $\int_{-\pi}^{\pi} \cos(nx)\cos(mx)\,\mathrm{d}x = 0$ 当 $n \neq m$。

A3. 判断以下函数在 $[-\pi,\pi]$ 上 Fourier 级数只含哪些项：$x^3$、$\cos x + x$、$x^2+1$、$\sin x + x^2$。

A4. 求 $f(x)=1$ 的 Fourier 级数。

A5. 求 $f(x)=\cos 3x$ 的 Fourier 系数。

A6. 对 $f(x)=x$，重新计算 $b_n$。

A7. 对 $f(x)=x^2$，重新计算 $a_0, a_n$。

A8. 在 $[-L,L]$ 上写出 $f(x)=x$ 的 Fourier 级数形式，不必完全化简。

## B 组：概念判断

B1. 为什么 Fourier 系数可以被称为“投影系数”？

B2. 为什么常数项是 $a_0/2$ 而不是 $a_0$？

B3. 如果一个函数有跳跃，Fourier 级数在跳跃点通常收敛到哪里？

B4. Gibbs phenomenon 是否会随着项数增加而完全消失？解释。

B5. 低频和高频在经济时间序列中分别可能代表什么？

B6. 为什么不能把所有高频成分都解释为噪声？

## C 组：混合题

C1. 求 $f(x)=|x|$ 在 $[-\pi,\pi]$ 上的 Fourier 级数。

C2. 求方波

$$f(x)=\begin{cases} 1, & 0<x<\pi, \\ -1, & -\pi<x<0 \end{cases}$$

的 Fourier 级数。

C3. 利用 $f(x)=x$ 的 Fourier 级数和 Parseval identity 推出 $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$。

C4. 设 $f(x)=x^2$，把 $x=0$ 代入其 Fourier 级数，推出 $\sum_{n=1}^{\infty}(-1)^{n+1}/n^2 = \pi^2/12$。

C5. 求 $f(x)=x$ 在 $(0,\pi)$ 上的正弦级数。

C6. 求 $f(x)=x$ 在 $(0,\pi)$ 上的余弦级数。

## D 组：经济金融建模题

D1. 月度消费数据存在年度季节性。写出一个包含前两个 Fourier 季节频率的回归模型。

D2. 季度 GDP 增长率可能存在 8 个季度左右的周期。写出一个对应的周期项，并解释其频率。

D3. 假设某股票高频收益由低频信息成分和高频微观结构噪声组成。用 Fourier 语言解释“低通滤波”的经济含义和风险。

D4. 把 Fourier 近似和资产因子模型作类比，指出二者的被解释对象、基底、系数和误差分别是什么。

D5. 如果一个政策冲击造成变量在某个时点突然跳跃，用 Fourier 级数近似该变量时可能出现什么问题?

## E 组：挑战/证明题

E1. 从最小化 $\int_{-\pi}^{\pi}[f(x)-S_N(x)]^2\,\mathrm{d}x$ 出发，完整推导 Fourier 系数公式。

E2. 证明 Parseval identity 的有限维投影版本，并解释它如何推广到 Fourier 级数。

E3. 对复指数形式，证明 $c_n=(a_n-\mathrm{i}b_n)/2$、$c_{-n}=(a_n+\mathrm{i}b_n)/2$。

E4. 设 $f$ 是分段光滑函数。解释为什么高频系数的衰减速度和函数的光滑性有关。只要求直觉，不要求严格证明。

# 21 练习题详细解答

## A 组答案

### A1

利用积化和差公式：

$$\sin(nx)\cos(mx)=\frac{1}{2}\{\sin[(n+m)x]+\sin[(n-m)x]\}.$$

对称区间 $[-\pi,\pi]$ 上，$\sin(kx)$ 是奇函数，所以积分为零。因此

$$\int_{-\pi}^{\pi}\sin(nx)\cos(mx)\,\mathrm{d}x=0.$$

### A2

当 $n\neq m$ 时，

$$\cos(nx)\cos(mx)=\frac{1}{2}\{\cos[(n-m)x]+\cos[(n+m)x]\}.$$

由于 $n-m\neq 0$、$n+m\neq 0$，两个余弦项在 $[-\pi,\pi]$ 上积分都为零，所以整体积分为零。

### A3

- $x^3$ 是奇函数，只含正弦项；
- $\cos x+x$ 包含偶函数 $\cos x$ 和奇函数 $x$，所以含余弦项和正弦项；
- $x^2+1$ 是偶函数，只含常数项和余弦项；
- $\sin x+x^2$ 包含奇函数和偶函数，所以含正弦项、常数项和余弦项。

### A4

$f(x)=1$。

$$a_0=\frac{1}{\pi}\int_{-\pi}^{\pi}1\,\mathrm{d}x=2.$$

对 $n\geq 1$，

$$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}\cos nx\,\mathrm{d}x=0,\qquad b_n=\frac{1}{\pi}\int_{-\pi}^{\pi}\sin nx\,\mathrm{d}x=0.$$

所以 Fourier 级数是

$$\frac{a_0}{2}=1.$$

**A5**

因为 $f(x)=\cos 3x$ 本身就是一个基函数，所以

$$a_3 = 1,$$

其他 $a_n = 0$，$b_n = 0$，且 $a_0 = 0$。这也可以由正交性直接推出。

**A6**

见 Worked Example 1。结果为

$$b_n = \frac{2(-1)^{n+1}}{n}.$$

因此

$$x \sim 2\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin nx.$$

**A7**

见 Worked Example 2。结果为

$$a_0 = \frac{2\pi^2}{3}, \qquad a_n = \frac{4(-1)^n}{n^2}, \qquad b_n = 0.$$

所以

$$x^2 \sim \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \cos nx.$$

**A8**

在 $[-L, L]$ 上，$f(x)=x$ 是奇函数，所以只含正弦项：

$$x \sim \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi x}{L}\right),$$

其中

$$b_n = \frac{1}{L}\int_{-L}^{L} x \sin\left(\frac{n\pi x}{L}\right) \mathrm{d}x.$$

由于 integrand 是偶函数，

$$b_n = \frac{2}{L}\int_{0}^{L} x \sin\left(\frac{n\pi x}{L}\right) \mathrm{d}x.$$

## B 组答案

### B1

因为 Fourier 系数正是把 $f$ 投影到基函数方向上的系数。例如

$$a_n = \frac{\langle f, \cos nx \rangle}{\langle \cos nx, \cos nx \rangle}.$$

由于 $\langle \cos nx, \cos nx \rangle = \pi$，所以得到

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, \mathrm{d}x.$$

### B2

因为常数基函数 $1$ 的范数平方是

$$\int_{-\pi}^{\pi} 1^2 \, \mathrm{d}x = 2\pi,$$

所以常数方向上的投影系数是

$$\frac{\langle f, 1 \rangle}{\langle 1, 1 \rangle} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \, \mathrm{d}x.$$

而 $a_0 = (1/\pi) \int f$，所以真正放在级数中的常数项是 $a_0/2$。

### B3

通常收敛到左右极限的平均值：

$$\frac{f(x+) + f(x-)}{2}.$$

### B4

不会完全消失。随着项数增加，振荡集中在更窄的跳跃邻域，但过冲比例不会完全趋于零。这是用连续光滑的三角函数逼近不连续跳跃时产生的结构性现象。

### B5

低频成分通常对应慢变化，例如长期趋势、商业周期、持久性冲击；高频成分通常对应快速变化，例如季节性、短期冲击、市场微观结构噪声。但具体解释必须结合数据生成机制。

**B6**

因为高频成分也可能包含真实经济信息。例如高频金融数据中的消息冲击、流动性变化、订单流信息，都可能是高频但并非无意义噪声。频率高低是数学属性，不自动决定经济含义。

## C 组答案

**C1**

见例 9.7：

$$|x| \sim \frac{\pi}{2} - \frac{4}{\pi}\left(\cos x + \frac{1}{3^2}\cos 3x + \frac{1}{5^2}\cos 5x + \cdots\right).$$

**C2**

见例 9.6：

$$f(x) \sim \frac{4}{\pi}\left(\sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x + \cdots\right).$$

**C3**

由

$$x \sim 2\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}\sin nx$$

和 Parseval：

$$\frac{1}{\pi}\int_{-\pi}^{\pi} x^2\,\mathrm{d}x = \sum_{n=1}^{\infty}\frac{4}{n^2}.$$

左边为 $2\pi^2/3$，所以

$$\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}.$$

**C4**

由

$$x^2 = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\cos nx.$$

令 $x = 0$，得

$$0 = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}.$$

所以

$$\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}=-\frac{\pi^2}{12},$$

即

$$\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}=\frac{\pi^2}{12}.$$

## C5

在 $(0,\pi)$ 上做正弦级数:

$$x\sim\sum_{n=1}^{\infty}b_n\sin nx,\qquad b_n=\frac{2}{\pi}\int_0^{\pi}x\sin nx\,\mathrm{d}x.$$

前面算过

$$\int_0^{\pi}x\sin nx\,\mathrm{d}x=-\frac{\pi(-1)^n}{n}.$$

所以

$$b_n=\frac{2(-1)^{n+1}}{n}.$$

因此

$$x\sim 2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\sin nx.$$

## C6

在 $(0,\pi)$ 上做余弦级数:

$$x\sim\frac{a_0}{2}+\sum_{n=1}^{\infty}a_n\cos nx.$$

$$a_0=\frac{2}{\pi}\int_0^{\pi}x\,\mathrm{d}x=\pi,$$

所以常数项是 $\pi/2$。

$$a_n=\frac{2}{\pi}\int_0^{\pi}x\cos nx\,\mathrm{d}x.$$

例 9.7 中已经算过

$$\int_0^{\pi}x\cos nx\,\mathrm{d}x=\frac{(-1)^n-1}{n^2}.$$

因此

$$a_n=\frac{2}{\pi}\frac{(-1)^n-1}{n^2}.$$

所以偶数 $n$ 的 $a_n = 0$，奇数 $n$ 的

$$a_n = -\frac{4}{\pi n^2}.$$

于是

$$x \sim \frac{\pi}{2} - \frac{4}{\pi}\left(\cos x + \frac{1}{3^2}\cos 3x + \frac{1}{5^2}\cos 5x + \cdots\right), \qquad 0 < x < \pi.$$

注意这和 $|x|$ 在 $[-\pi, \pi]$ 上的展开一样，因为余弦展开对应偶延拓。

## D 组答案

### D1

月度年度季节性周期为 $s = 12$。包含前两个 Fourier 季节频率的模型可写成

$$C_t = \alpha + \beta X_t + a_1 \cos\left(\frac{2\pi t}{12}\right) + b_1 \sin\left(\frac{2\pi t}{12}\right) + a_2 \cos\left(\frac{4\pi t}{12}\right) + b_2 \sin\left(\frac{4\pi t}{12}\right) + u_t.$$

第一组频率捕捉一年一次的周期，第二组频率捕捉一年两次的周期。

### D2

周期为 8 个季度，则角频率为

$$\omega = \frac{2\pi}{8} = \frac{\pi}{4}.$$

对应周期项可以写成

$$A\cos\left(\frac{\pi}{4}t + \phi\right),$$

或者等价地写成

$$a\cos\left(\frac{\pi}{4}t\right) + b\sin\left(\frac{\pi}{4}t\right).$$

### D3

低通滤波的含义是保留低频成分、削弱高频成分。经济上，这可能被解释为保留较持久的信息变化、去除高频交易噪声。风险在于：真实短期信息也可能位于高频部分，因此滤波可能把有用信息一并删掉，并造成估计滞后。

### D4

因子模型中,被解释对象是资产收益 $R_i$,基底是风险因子 $F_j$,系数是 factor loadings $\beta_{ij}$，误差是 idiosyncratic component。Fourier 近似中，被解释对象是函数 $f(t)$，基底是 $\sin nt, \cos nt$，系数是 $a_n, b_n$，误差是截断误差 $f - S_N$。二者都是把复杂对

象投影到一个较简单的解释空间。

**D5**

可能出现 Gibbs phenomenon：在跳跃点附近，Fourier 近似会产生振荡和过冲。经济解释上，这些过冲可能是数学近似产生的，不一定代表真实政策效果。因此对制度断点、政策冲击、价格跳跃等变量使用 Fourier 近似时需要谨慎。

## E 组答案

**E1**

令

$$S_N(x) = \frac{a_0}{2} + \sum_{n=1}^{N}(a_n \cos nx + b_n \sin nx).$$

最小化

$$Q = \int_{-\pi}^{\pi} [f(x) - S_N(x)]^2 \,\mathrm{d}x.$$

对 $a_k$ 求一阶条件：

$$\frac{\partial Q}{\partial a_k} = -2\int_{-\pi}^{\pi} [f(x) - S_N(x)] \cos kx \,\mathrm{d}x = 0.$$

所以残差与 $\cos kx$ 正交：

$$\int_{-\pi}^{\pi} [f(x) - S_N(x)] \cos kx \,\mathrm{d}x = 0.$$

利用正交性，只有 $a_k \cos kx$ 项留下：

$$\int f(x) \cos kx \,\mathrm{d}x - a_k \int \cos^2 kx \,\mathrm{d}x = 0.$$

因为 $\int_{-\pi}^{\pi} \cos^2 kx \,\mathrm{d}x = \pi$，得

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \cos kx \,\mathrm{d}x.$$

对 $b_k$ 同理可得

$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \sin kx \,\mathrm{d}x.$$

对常数项也做一阶条件，可得

$$\frac{a_0}{2} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x) \,\mathrm{d}x,$$

即

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \,\mathrm{d}x.$$

**E2**

有限维中，若 $\{e_1, \ldots, e_k\}$ 是正交基，则

$$x = \sum_{j=1}^{k} c_j e_j, \qquad c_j = \frac{\langle x, e_j\rangle}{\langle e_j, e_j\rangle}.$$

于是

$$\|x\|^2 = \langle x, x\rangle = \sum_{j=1}^{k} c_j^2 \langle e_j, e_j\rangle = \sum_{j=1}^{k} \frac{\langle x, e_j\rangle^2}{\langle e_j, e_j\rangle}.$$

Fourier 级数把这个思想推广到无限维正交基 $1, \cos nx, \sin nx$，从而得到

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f^2 = \frac{a_0^2}{2} + \sum_{n=1}^{\infty}(a_n^2 + b_n^2).$$

## E3

由 Euler 公式:

$$\cos nx = \frac{\mathrm{e}^{\mathrm{i}nx} + \mathrm{e}^{-\mathrm{i}nx}}{2}, \qquad \sin nx = \frac{\mathrm{e}^{\mathrm{i}nx} - \mathrm{e}^{-\mathrm{i}nx}}{2\mathrm{i}}.$$

于是

$$a_n \cos nx + b_n \sin nx = \frac{a_n}{2}(\mathrm{e}^{\mathrm{i}nx} + \mathrm{e}^{-\mathrm{i}nx}) + \frac{b_n}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}nx} - \mathrm{e}^{-\mathrm{i}nx}).$$

因为 $1/\mathrm{i} = -\mathrm{i}$，所以 $\mathrm{e}^{\mathrm{i}nx}$ 的系数为

$$\frac{a_n}{2} - \frac{\mathrm{i}b_n}{2} = \frac{a_n - \mathrm{i}b_n}{2},$$

$\mathrm{e}^{-\mathrm{i}nx}$ 的系数为

$$\frac{a_n}{2} + \frac{\mathrm{i}b_n}{2} = \frac{a_n + \mathrm{i}b_n}{2}.$$

因此

$$c_n = \frac{a_n - \mathrm{i}b_n}{2}, \qquad c_{-n} = \frac{a_n + \mathrm{i}b_n}{2}.$$

## E4

直觉是：函数越光滑，高频成分越少，Fourier 系数衰减越快。因为高频正弦余弦负责表达快速变化和尖锐结构。如果函数有跳跃，需要很多高频项才能逼近跳跃，系数衰减较慢；如果函数连续且可微很多次，高频波与函数相乘积分时会大量抵消，系数衰减更快。分部积分可以看出这种关系：每多一次光滑性，往往可以多做一次分部积分，从而多得到一个 $1/n$ 的衰减因子。

# 22 本章到下一章的连接

本章结束后，Part IV 也结束了。我们已经完成了从投影到 Fourier 的一条线：

finite-dimensional projection $\rightarrow$ OLS $\rightarrow$ function-space projection $\rightarrow$ Fourier frequency d

下一章开始进入 Part V：ODE、差分方程与经济动态。那里我们将从“把函数分解成波”转向“一个变量如何随时间演化”。Fourier 关心的是周期结构，ODE 关心的是动态法则：

$$\text{Fourier: } f(t) \text{ contains which frequencies?}$$

$$\text{ODE: } \dot{x}(t) = F(x(t), t) \text{ implies what path?}$$

宏观、金融和动态优化会同时需要这两种眼光：一方面理解数据中有哪些波动成分，另一方面理解模型中状态变量为什么这样运动。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 10
# 一阶 ODE 与连续时间动态

本章开始进入动态系统。静态微积分告诉你“边际变化是多少”，一阶 ODE 则告诉你“变量会沿着什么路径变化”。对经济金融来说，ODE 是连续时间增长模型、债务动态、价格调整、资本积累、通胀调整、资产价格动态和后续动态优化的入口。本章的目标不是把微分方程学成工程技巧，而是建立经济学生最需要的三种能力：会解简单方程、会看相图和稳态、会解释参数如何影响动态路径。

math4econ personalized notes
2026

# 目录

**Chapter 10 总览** **4**

**1 本章为什么重要** **4**
1.1 从静态最优化到动态路径 . . . . . 5
1.2 经济金融中 ODE 的三个层次 . . . . . 5

**2 最小前置知识** **5**

**3 从微分方程到动态模型** **5**
3.1 什么叫 ODE . . . . . 6
3.2 初始条件为什么重要 . . . . . 6
3.3 自治方程 autonomous equation . . . . . 6

**4 最简单的 ODE: 指数增长与指数衰减** **7**
4.1 模型 . . . . . 7
4.2 详细推导 . . . . . 7
4.3 增长率的解释 . . . . . 8

**5 可分离变量方程** **8**
5.1 一般形式 . . . . . 8
5.2 Logistic 方程 . . . . . 9
5.3 稳态和经济解释 . . . . . 9

**6 自治方程、相线图与稳定性** **9**
6.1 稳态 . . . . . 9
6.2 稳定性 . . . . . 10
6.3 为什么 $f'(x^*) < 0$ 表示稳定 . . . . . 10

**7 一阶线性 ODE** **11**
7.1 标准形式 . . . . . 11
7.2 常系数线性方程的解 . . . . . 11
7.3 积分因子法的详细推导 . . . . . 12

**8 收敛速度与 half-life** **13**
8.1 半衰期 half-life . . . . . 13

**9 经济应用 1: Solow 模型** **13**
9.1 模型方程 . . . . . 13
9.2 稳态 . . . . . 14
9.3 稳定性 . . . . . 14

9.4 收敛速度 . . . . . 15

**10 经济应用 2：债务/GDP 动态 15**
10.1 基本方程 . . . . . 15
10.2 解的形式 . . . . . 15
10.3 经济解释 . . . . . 16

**11 经济应用 3：价格调整与市场稳定 16**
11.1 模型 . . . . . 16
11.2 局部稳定条件 . . . . . 17
11.3 经济解释 . . . . . 17

**12 经济应用 4：均值回复 17**

**13 计算模板：拿到一阶 ODE 怎么做 18**

**14 常见误区与纠正 19**

**15 本章小结图谱 19**

**16 存在唯一性与“路径不能交叉”的直觉 19**
16.1 为什么一维自治方程的路径不能交叉 . . . . . 20

**17 斜率场 slope field：不求解也能看路径 20**

**18 资本存量方程：流量与存量的最基本关系 21**

**19 比较静态：稳态如何随参数变化 22**
19.1 为什么这个公式重要 . . . . . 22

**20 符号、术语与经济翻译 23**

**21 练习题 23**
A 组：基础计算 . . . . . 23
B 组：概念判断 . . . . . 24
C 组：混合题 . . . . . 24
D 组：经济金融建模题 . . . . . 24
E 组：挑战 / 证明题 . . . . . 24

**22 练习题详细解答 25**
A 组答案 . . . . . 25
B 组答案 . . . . . 25
C 组答案 . . . . . 26
D 组答案 . . . . . 27

E 组答案 . . . . . 28

**23 本章到下一章的连接** **28**

# Chapter 10 总览

| | |
|---|---|
| **本章主题** | 一阶常微分方程 ordinary differential equations, ODE, 与连续时间经济动态。 |
| **核心问题** | 如果一个变量的变化率由当前状态决定，例如 $\dot{x} = f(x)$, 那么它会走向哪里, 走得多快, 稳态是否稳定, 参数改变后长期结果如何变化? |
| **数学工具** | 可分离变量、一阶线性 ODE、积分因子、自治方程、相线图、局部线性化、稳定性、收敛速度。 |
| **经济应用** | Solow 增长模型、债务/GDP 动态、价格调整、资本积累、通胀调整、连续复利、风险资产均值回复。 |
| **后续连接** | Chapter 11 差分方程, Chapter 12 线性系统, Chapter 13 非线性动态, Chapter 18 比较静态, 以及动态优化和宏观金融模型。 |

**直觉盒子**

微积分中你熟悉的是

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f'(x),$$

这里的 $x$ 常常是一个自变量。动态系统中更常见的是

$$\dot{x}(t) = \frac{\mathrm{d}x(t)}{\mathrm{d}t} = F(x(t), t),$$

也就是说，未知对象不再是一个数，而是一整条时间路径 $x(t)$。ODE 的本质问题是：**给定一个变化规律和初始状态，求整个未来路径**。

## 1 本章为什么重要

经济学里大量模型不是只问“最优点在哪里”，而是问“变量怎样从现在走向未来”。例如:

- 资本存量怎样积累: $\dot{K}(t) = I(t) - \delta K(t)$;
- 人均资本怎样走向稳态: $\dot{k} = sf(k) - (n + \delta)k$;
- 政府债务占 GDP 的比例是否会爆炸: $\dot{b} = (r - g)b + d$;
- 价格是否会向均衡价格调整: $\dot{p} = \lambda[D(p) - S(p)]$;
- 通胀或预期是否存在均值回复: $\dot{x} = -a(x - \bar{x})$。

这些式子看起来很短, 但它们表达的是一套动态逻辑: **状态变量的变化率等于某种净流入、净压力或调整力量**。

### 1.1 从静态最优化到动态路径

前面几章中，积分、梯度、全微分大多处理静态对象。例如消费者选择中，我们可能解

$$\max_x u(x) \quad \text{s.t.} \quad p \cdot x \le w.$$

这类问题的答案是一个最优点或最优需求函数。可是宏观和金融中常常需要研究的是

$$\text{state today} \longrightarrow \text{state tomorrow} \longrightarrow \text{long-run state}.$$

ODE 给我们的正是这种语言。

### 1.2 经济金融中 ODE 的三个层次

1. **会解：** 给定方程，求 $x(t)$。例如 $\dot{x} = ax$ 的解是 $x(t) = x_0 \mathrm{e}^{at}$。
2. **会看：** 不求显式解，也能从 $\dot{x} = f(x)$ 判断变量往哪里走，稳态是否稳定。
3. **会解释：** 把数学结论翻译成经济语言：增长、收敛、发散、均值回复、债务可持续性、调整速度。

本章最重视第二和第三层，因为经济模型常常没有漂亮的解析解。

## 2 最小前置知识

学习本章之前，你只需要回忆以下内容。

| 前置知识 | 本章如何使用 |
|---|---|
| 导数 | $\dot{x}(t) = \mathrm{d}x(t)/\mathrm{d}t$ 表示变量随时间的瞬时变化率。 |
| 不定积分 | 解可分离变量和线性方程时需要积分。 |
| 指数函数 | 连续增长、衰减、均值回复、收敛速度都依赖 $\mathrm{e}^{at}$。 |
| 多元微分 | 稳定性和线性化会用到导数 $f'(x^*)$。 |
| 图像直觉 | 相线图和 Solow 图是理解动态模型的重要工具。 |

**常见误区**

本章不是数学系 ODE 完整课程。我们暂时不处理高阶 ODE、Laplace transform、复杂边值问题、PDE。对经济金融建模而言，第一步是熟练掌握一阶动态方程的含义、求解和稳定性。

## 3 从微分方程到动态模型

## 3.1 什么叫 ODE

**定义 / 定理**

**常微分方程.** 如果一个方程包含未知函数 $x(t)$ 及其关于单一变量 $t$ 的导数，例如

$$F(t, x(t), \dot{x}(t)) = 0,$$

则称它为常微分方程 ordinary differential equation, ODE。
若方程只含一阶导数 $\dot{x}(t)$, 称为一阶 ODE。例如

$$\dot{x}(t) = ax(t), \qquad \dot{x}(t) = b - ax(t), \qquad \dot{x}(t) = x(t)(1 - x(t)).$$

在经济学中, $t$ 通常代表时间, $x(t)$ 是状态变量 state variable。例如 $k(t)$ 是人均资本, $b(t)$ 是债务/GDP 比例，$p(t)$ 是价格，$\pi(t)$ 是通胀率。

## 3.2 初始条件为什么重要

ODE 通常不是只给一个函数，而是给一族函数。例如

$$\dot{x} = ax$$

的解是

$$x(t) = C\mathrm{e}^{at},$$

其中 $C$ 是常数。要确定唯一轨迹，需要初始条件

$$x(0) = x_0.$$

于是 $C = x_0$, 解变成

$$x(t) = x_0\mathrm{e}^{at}.$$

经济解释是：同一个经济规律下，不同国家、企业或投资组合的起点不同, 后续路径也不同。

## 3.3 自治方程 autonomous equation

很多经济模型可写成

$$\dot{x} = f(x),$$

右边不显式依赖时间 $t$。这种方程叫自治方程 autonomous equation。

它的好处是：只要知道当前状态 $x$, 就知道下一瞬间往哪儿动。例如

$$\dot{k} = sf(k) - (n + \delta)k.$$

给定当前资本 $k$, 右边为正则资本上升, 为负则资本下降。

# 4 最简单的 ODE: 指数增长与指数衰减

## 4.1 模型

最基础的一阶 ODE 是

$$\dot{x} = ax.$$

它表示: 变量的增长速度与自身规模成正比。

- 若 $a > 0$, $x$ 指数增长。
- 若 $a < 0$, $x$ 指数衰减。
- 若 $a = 0$, $x$ 保持不变。

## 4.2 详细推导

假设 $x(t) > 0$, 方程为

$$\frac{\mathrm{d}x}{\mathrm{d}t} = ax.$$

把含 $x$ 的部分放到左边，含 $t$ 的部分放到右边:

$$\frac{1}{x}\,\mathrm{d}x = a\,\mathrm{d}t.$$

两边积分:

$$\int \frac{1}{x}\,\mathrm{d}x = \int a\,\mathrm{d}t.$$

因此

$$\ln x = at + C.$$

两边取指数:

$$x(t) = \mathrm{e}^{C}\mathrm{e}^{at}.$$

令 $A = \mathrm{e}^{C}$, 得

$$x(t) = A\mathrm{e}^{at}.$$

使用初始条件 $x(0) = x_0$:

$$x_0 = A\mathrm{e}^{0} = A.$$

所以

$$\boxed{x(t) = x_0\mathrm{e}^{at}.}$$

**直觉盒子**

连续复利、人口增长、技术增长、资本折旧、放射性衰减、均值回复的线性化形式，都可以看成指数函数的变形。指数函数是连续时间动态模型的“乘法骨架”。

### 4.3 增长率的解释

若 $x(t)>0$, 则

$$\frac{\dot{x}(t)}{x(t)}=a.$$

左边是瞬时增长率 instantaneous growth rate。因此 $\dot{x}=ax$ 的含义不是“每期加 $a$”, 而是“以连续复利增长率 $a$ 增长”。

**Worked Example**

[title=Worked Example 1: 连续复利] 假设初始财富 $W(0)=100$, 连续复利利率 $r=5\%$, 则

$$\dot{W}=0.05W.$$

解为

$$W(t)=100\mathrm{e}^{0.05t}.$$

10 年后财富为

$$W(10)=100\mathrm{e}^{0.5}\approx 164.87.$$

如果用离散年复利, 则是 $100(1.05)^{10}\approx 162.89$。连续复利略高, 因为计息频率趋于无穷。

## 5 可分离变量方程

### 5.1 一般形式

**定义 / 定理**

若一阶 ODE 可以写为

$$\frac{\mathrm{d}x}{\mathrm{d}t}=g(t)h(x),$$

则称它为可分离变量方程 separable equation。只要 $h(x)\neq 0$, 可以改写为

$$\frac{1}{h(x)}\,\mathrm{d}x=g(t)\,\mathrm{d}t.$$

两边积分即可。

**方法模板**

**可分离变量四步法**

1. 写成 $\mathrm{d}x/\mathrm{d}t=g(t)h(x)$。
2. 移项: $\frac{1}{h(x)}\,\mathrm{d}x=g(t)\,\mathrm{d}t$。
3. 两边积分。

4. 使用初始条件确定常数。

## 5.2 Logistic 方程

Logistic 方程是

$$\dot{x} = rx\left(1 - \frac{x}{K}\right), \qquad r > 0, K > 0.$$

它可以表示人口、市场扩散、技术采用或平台用户增长。$K$ 是 carrying capacity, 即长期容量上限。

## 5.3 稳态和经济解释

稳态 steady state 是满足 $\dot{x} = 0$ 的状态。对 logistic 方程,

$$rx\left(1 - \frac{x}{K}\right) = 0,$$

所以

$$x^* = 0, \qquad x^* = K.$$

若 $0 < x < K$, 则 $\dot{x} > 0$, 变量上升。若 $x > K$, 则 $\dot{x} < 0$, 变量下降。因此 $K$ 是稳定稳态, 0 在正半轴上是不稳定边界点。

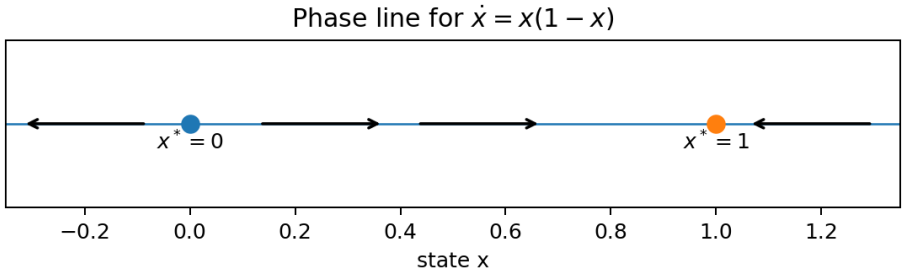



图 1: 相线图: 箭头显示状态变量的运动方向。

# 6 自治方程、相线图与稳定性

## 6.1 稳态

对自治方程

$$\dot{x} = f(x),$$

稳态 $x^*$ 满足

$$f(x^*) = 0.$$

经济上，稳态表示“净变化为零”。例如在 Solow 模型中，投资正好抵消折旧和人口增长; 在价格调整模型中，需求等于供给; 在债务动态中，债务比率不再变化。

## 6.2 稳定性

**定义 / 定理**

**局部稳定.** 如果从 $x^*$ 附近出发的路径最终回到 $x^*$, 则称 $x^*$ 是局部稳定 steady state。对一维自治方程 $\dot{x} = f(x)$, 若 $f$ 可微且

$$f'(x^*) < 0,$$

则 $x^*$ 局部稳定; 若

$$f'(x^*) > 0,$$

则 $x^*$ 局部不稳定。

## 6.3 为什么 $f'(x^*) < 0$ 表示稳定

在 $x^*$ 附近做线性近似:

$$f(x) \approx f(x^*) + f'(x^*)(x - x^*).$$

因为 $f(x^*) = 0$, 所以

$$\dot{x} \approx f'(x^*)(x - x^*).$$

令偏离量 $y = x - x^*$, 则

$$\dot{y} \approx f'(x^*)y.$$

解为

$$y(t) = y(0)\mathrm{e}^{f'(x^*)t}.$$

若 $f'(x^*) < 0$, 指数项衰减, 偏离量回到 0; 若 $f'(x^*) > 0$, 指数项增长, 偏离量变大。

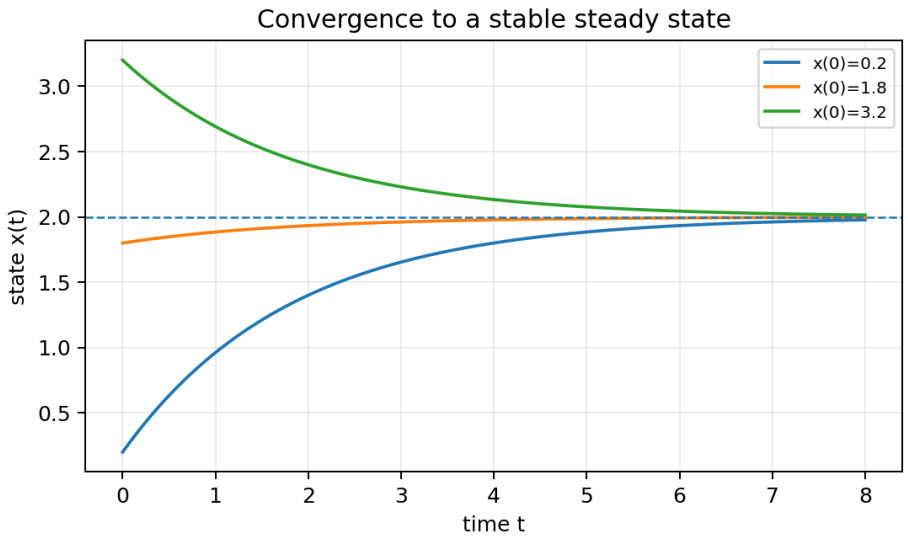



图 2: 稳定稳态附近, 不同初始点都会向同一个稳态收敛。

**方法模板**

**判断一维自治方程稳定性的步骤**

1. 找稳态: 解 $f(x)=0$。
2. 画相线: 判断各区间上 $f(x)$ 的正负。
3. 或者求导: 用 $f'(x^*)$ 判断局部稳定性。
4. 翻译经济含义: 稳态是吸引点、排斥点, 还是边界点。

# 7 一阶线性 ODE

## 7.1 标准形式

**定义 / 定理**

一阶线性 ODE 的标准形式是

$$\dot{x}+a(t)x=b(t).$$

若 $a(t)=a$, $b(t)=b$ 是常数, 则为常系数线性方程:

$$\dot{x}+ax=b.$$

经济学中大量动态方程都属于这个类型或可局部近似为这个类型。

## 7.2 常系数线性方程的解

考虑

$$\dot{x}+ax=b, \qquad a\neq 0.$$

稳态满足

$$0+ax^*=b,$$

即

$$x^*=\frac{b}{a}.$$

令 $y=x-x^*$, 则

$$\dot{y}=\dot{x}.$$

而

$$\dot{x}+ax=b.$$

把 $x=y+x^*$ 代入:

$$\dot{y}+a(y+x^*)=b.$$

由于 $ax^* = b$, 得

$$\dot{y} + ay = 0.$$

所以

$$y(t) = y(0)\mathrm{e}^{-at}.$$

因此

$$\boxed{x(t) = x^* + (x_0 - x^*)\mathrm{e}^{-at} = \frac{b}{a} + \left(x_0 - \frac{b}{a}\right)\mathrm{e}^{-at}.}$$

若 $a > 0$, 则 $\mathrm{e}^{-at} \to 0$, 路径收敛到 $b/a$。

## 7.3 积分因子法的详细推导

一般线性方程

$$\dot{x} + a(t)x = b(t)$$

可以用积分因子 integrating factor 求解。目标是把左边变成某个乘积的导数。设

$$\mu(t) = \exp\left(\int a(t)\,\mathrm{d}t\right).$$

则

$$\mu'(t) = a(t)\mu(t).$$

原方程两边乘以 $\mu(t)$:

$$\mu\dot{x} + \mu a x = \mu b.$$

注意

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mu(t)x(t)] = \mu(t)\dot{x}(t) + \mu'(t)x(t) = \mu\dot{x} + a\mu x.$$

因此

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mu(t)x(t)] = \mu(t)b(t).$$

两边积分:

$$\mu(t)x(t) = C + \int \mu(t)b(t)\,\mathrm{d}t.$$

所以

$$\boxed{x(t) = \frac{1}{\mu(t)}\left[C + \int \mu(t)b(t)\,\mathrm{d}t\right].}$$

**方法模板**

**积分因子法三句话**

1. 看到 $\dot{x} + a(t)x = b(t)$, 先找 $\mu(t) = \exp(\int a(t)\,\mathrm{d}t)$。
2. 乘以 $\mu(t)$, 左边变成 $\frac{\mathrm{d}}{\mathrm{d}t}[\mu(t)x(t)]$。

3. 积分再除以 $\mu(t)$, 最后代初始条件。

# 8 收敛速度与 half-life

如果路径满足

$$x(t) - x^* = (x_0 - x^*)\mathrm{e}^{-\lambda t}, \qquad \lambda > 0,$$

则偏离量以速度 $\lambda$ 衰减。$\lambda$ 越大, 收敛越快。

## 8.1 半衰期 half-life

半衰期是偏离量缩小一半所需时间。令

$$\mathrm{e}^{-\lambda T_{1/2}} = \frac{1}{2}.$$

取对数:

$$-\lambda T_{1/2} = -\ln 2.$$

因此

$$\boxed{T_{1/2} = \frac{\ln 2}{\lambda}.}$$

经济解释: 如果一个宏观变量与长期均衡的差距每隔 $T_{1/2}$ 年减半, 则 $T_{1/2}$ 是非常直观的调整速度指标。

# 9 经济应用 1: Solow 模型

## 9.1 模型方程

在最简 Solow 模型中, 人均有效资本 $k(t)$ 满足

$$\dot{k} = sf(k) - (n+\delta)k,$$

其中

- $s$ 是储蓄率;
- $f(k)$ 是人均有效产出函数;
- $n$ 是人口或有效劳动增长率;
- $\delta$ 是折旧率;
- $sf(k)$ 是实际投资;
- $(n+\delta)k$ 是维持资本不变所需的 break-even investment。

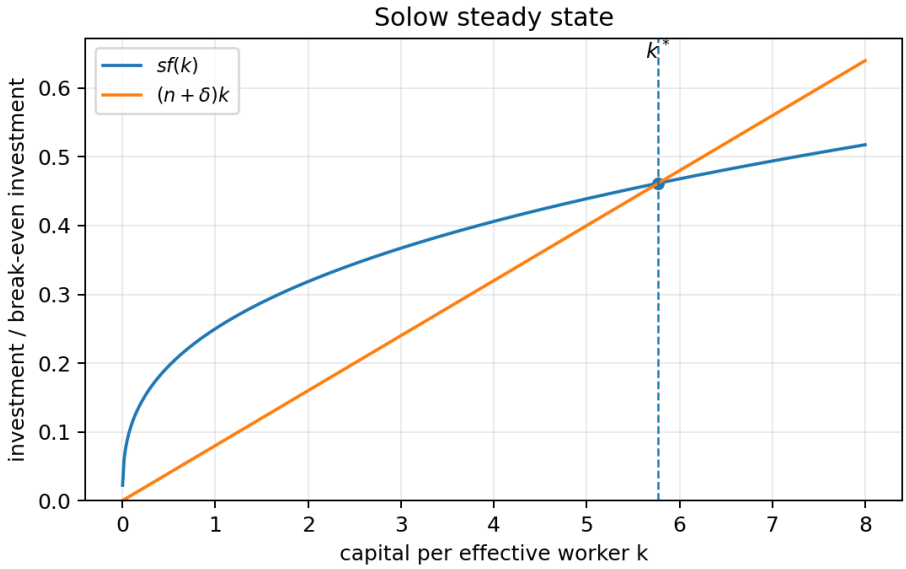



图 3: Solow 模型中, 稳态由实际投资与维持投资的交点给出。

## 9.2 稳态

稳态 $k^*$ 满足

$$0 = sf(k^*) - (n+\delta)k^*.$$

也就是

$$\boxed{sf(k^*) = (n+\delta)k^*.}$$

如果 $f(k) = k^\alpha$, $0 < \alpha < 1$, 则

$$s(k^*)^\alpha = (n+\delta)k^*.$$

若 $k^* > 0$, 两边除以 $(k^*)^\alpha$:

$$s = (n+\delta)(k^*)^{1-\alpha}.$$

所以

$$\boxed{k^* = \left(\frac{s}{n+\delta}\right)^{\frac{1}{1-\alpha}}.}$$

## 9.3 稳定性

令

$$F(k) = sk^\alpha - (n+\delta)k.$$

则

$$F'(k) = s\alpha k^{\alpha-1} - (n+\delta).$$

在稳态处，由

$$s(k^*)^{\alpha-1} = n+\delta$$

可得

$$F'(k^*) = \alpha(n+\delta) - (n+\delta) = -(1-\alpha)(n+\delta) < 0.$$

因此正的 Solow 稳态局部稳定。

### 9.4 收敛速度

局部线性化给出

$$\dot{k} \approx F'(k^*)(k - k^*).$$

由于

$$F'(k^*) = -(1-\alpha)(n+\delta),$$

所以偏离量大致满足

$$k(t) - k^* \approx (k_0 - k^*)\mathrm{e}^{-(1-\alpha)(n+\delta)t}.$$

收敛速度近似为

$$\lambda = (1-\alpha)(n+\delta).$$

**直觉盒子**

Solow 模型的 ODE 告诉你两件事。第一, 长期资本由“投资曲线”和“维持投资直线”的交点决定。第二, 如果边际资本回报递减, 低于稳态时资本上升, 高于稳态时资本下降, 因而稳态是吸引点。

## 10 经济应用 2: 债务/GDP 动态

### 10.1 基本方程

设 $b(t)$ 是政府债务占 GDP 的比率, $r$ 是实际利率, $g$ 是 GDP 增长率, $d$ 是 primary deficit 占 GDP 的比率。一个简化连续时间模型是

$$\dot{b} = (r-g)b + d.$$

### 10.2 解的形式

令

$$\rho = r - g.$$

则

$$\dot{b} = \rho b + d.$$

如果 $\rho \neq 0$, 稳态满足

$$0 = \rho b^* + d,$$

因此

$$b^* = -\frac{d}{\rho}.$$

解为

$$\boxed{b(t) = b^* + (b_0 - b^*)\mathrm{e}^{\rho t}.}$$

若 $r > g$, 即 $\rho > 0$, 偏离项 $\mathrm{e}^{\rho t}$ 增长, 债务动态通常不稳定。若 $r < g$, 即 $\rho < 0$, 偏离项衰减, 债务比率更容易稳定。

### 10.3 经济解释

- $r > g$: 利息负担增长快于经济规模, 债务/GDP 有爆炸压力。
- $r < g$: 经济规模增长快于利息累积, 债务/GDP 更容易被“稀释”。
- $d > 0$: 初级赤字推高债务比率。
- $d < 0$: 初级盈余降低债务比率。

**Worked Example**

[title=Worked Example 2: 债务可持续性] 设 $r = 0.04$, $g = 0.02$, $d = 0.01$, 初始债务比率 $b_0 = 0.6$。则 $\rho = 0.02$,

$$b^* = -\frac{0.01}{0.02} = -0.5.$$

解为

$$b(t) = -0.5 + (0.6 + 0.5)\mathrm{e}^{0.02t} = -0.5 + 1.1\mathrm{e}^{0.02t}.$$

因为指数项增长, 债务比率长期上升。这里的负稳态没有现实吸引意义, 它只是线性方程的数学稳态; 经济上更重要的是 $r - g > 0$ 带来的发散压力。

## 11 经济应用 3: 价格调整与市场稳定

### 11.1 模型

令 $D(p)$ 是需求, $S(p)$ 是供给。假设价格按 excess demand 调整:

$$\dot{p} = \lambda[D(p) - S(p)], \qquad \lambda > 0.$$

均衡价格 $p^*$ 满足

$$D(p^*) = S(p^*).$$

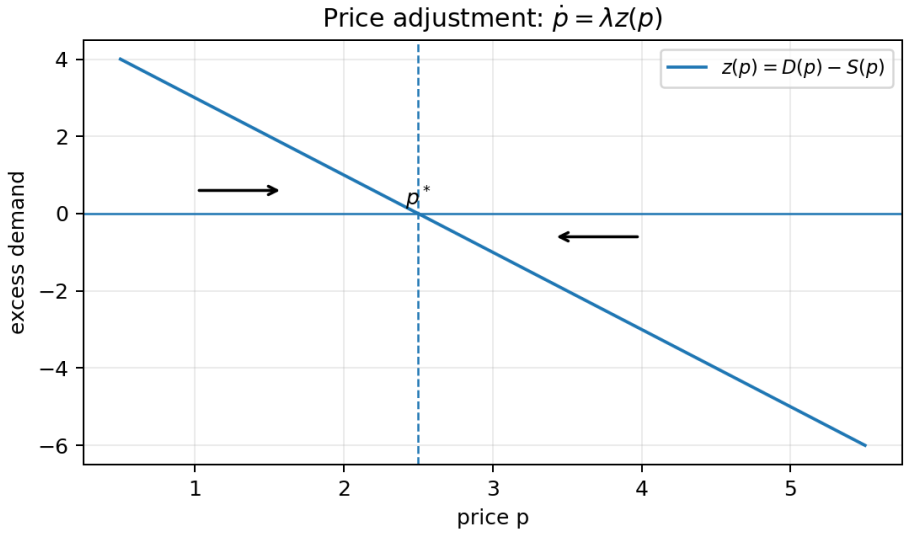



图 4: 若 excess demand 对价格递减, 价格向均衡调整。

## 11.2 局部稳定条件

令

$$z(p) = D(p) - S(p).$$

则

$$\dot{p} = \lambda z(p).$$

在均衡处 $z(p^*) = 0$。局部稳定要求

$$\lambda z'(p^*) < 0.$$

由于 $\lambda > 0$, 等价于

$$\boxed{z'(p^*) = D'(p^*) - S'(p^*) < 0.}$$

通常需求向下倾斜 $D' < 0$, 供给向上倾斜 $S' > 0$, 因而 $D' - S' < 0$, 均衡稳定。

## 11.3 经济解释

如果 $p < p^*$, excess demand 为正, 价格上升; 如果 $p > p^*$, excess demand 为负, 价格下降。因此价格被推回均衡。若反过来 $z'(p^*) > 0$, 则价格偏离后会被进一步推离均衡。

# 12 经济应用 4: 均值回复

金融中经常用均值回复 mean reversion 描述利率、利差、估值比率或某些风险因子。最简单模型是

$$\dot{x} = -a(x - \bar{x}), \qquad a > 0.$$

其中 $\bar{x}$ 是长期均值。

解法很直接。令 $y = x - \bar{x}$, 则

$$\dot{y} = -ay.$$

所以

$$y(t) = y(0)\mathrm{e}^{-at}.$$

因此

$$\boxed{x(t) = \bar{x} + (x_0 - \bar{x})\mathrm{e}^{-at}.}$$

经济金融解释: 当前值高于长期均值时, $x_0 - \bar{x} > 0$, 后续逐渐下降; 当前值低于长期均值时, 后续逐渐上升。

## 13 计算模板: 拿到一阶 ODE 怎么做

**方法模板**

**模板 1: 判断方程类型**

1. 如果是 $\dot{x} = ax$, 直接指数解。
2. 如果可以写成 $\dot{x} = g(t)h(x)$, 用可分离变量。
3. 如果是 $\dot{x} + a(t)x = b(t)$, 用积分因子。
4. 如果是 $\dot{x} = f(x)$ 且不好显式求解, 做稳态和相线图。

**方法模板**

**模板 2: 动态经济模型分析**

1. 找状态变量 $x(t)$。
2. 写出净变化率 $\dot{x}$: 流入减流出, 调整力量, 或 excess demand。
3. 找稳态: 令 $\dot{x} = 0$。
4. 判断稳定性: 看 $f'(x^*)$ 或画相线图。
5. 做比较静态: 改变参数, 看稳态和收敛速度怎么变。
6. 翻译经济含义: 增长、收敛、发散、调整速度、政策影响。

## 14 常见误区与纠正

**常见误区**

**误区 1：把 $\dot{x}$ 当成普通分数随意操作。**分离变量在很多情况下是合法的计算技巧，但背后依赖可微性和变量变换。初学时可以按形式操作，但要记得不是所有 ODE 都能分离变量。

**误区 2：只会求显式解，不会看相图。**经济模型常常没有简单显式解。稳态和稳定性比解析解更重要。

**误区 3：稳态等于“好结果”。**稳态只是 $\dot{x}=0$。它可能稳定，也可能不稳定；可能现实可行，也可能经济上无意义。

**误区 4：$r>g$ 与 $r<g$ 只是符号差别。**在债务动态中，$r-g$ 的符号决定指数项是增长还是衰减，经济含义完全不同。

**误区 5：认为 ODE 只是宏观。** ODE 同样出现在金融资产定价、利率模型、均值回复、学习动态、演化博弈和机制调整中。

## 15 本章小结图谱

| 概念 | 数学形式 | 经济金融含义 |
|---|---|---|
| 指数增长 | $\dot{x}=ax$ | 连续复利、人口增长、技术增长。 |
| 指数衰减 | $\dot{x}=-ax$ | 折旧、偏离量衰减、均值回复。 |
| 稳态 | $f(x^*)=0$ | 净变化为零的长期状态。 |
| 局部稳定 | $f'(x^*)<0$ | 偏离稳态后会被拉回。 |
| 可分离变量 | $\dot{x}=g(t)h(x)$ | 能通过两边积分求解的动态模型。 |
| 线性 ODE | $\dot{x}+ax=b$ | 债务、调整、均值回复的基本结构。 |
| 收敛速度 | $\mathrm{e}^{-\lambda t}$ | 偏离长期状态的消失速度。 |
| Solow 模型 | $\dot{k}=sf(k)-(n+\delta)k$ | 资本积累与长期稳态。 |
| 价格调整 | $\dot{p}=\lambda[D(p)-S(p)]$ | excess demand 推动价格变化。 |

## 16 存在唯一性与“路径不能交叉”的直觉

很多经济学生学习 ODE 时会马上开始求解，但动态模型背后还有一个重要问题：给定初始值以后，路径是否真的存在，是否唯一？如果不唯一，模型就会失去预测意义。

**定义 / 定理**

**存在唯一性的实用版本.** 对初始值问题

$$\dot{x} = f(t, x), \qquad x(t_0) = x_0,$$

如果 $f(t, x)$ 对 $t, x$ 连续, 并且对 $x$ 满足适当的 Lipschitz 条件, 那么在初始点附近存在唯一解。

这里不需要把 Lipschitz 条件学成纯分析难题。对多数经济模型, 如果 $f$ 连续可微, 就可以放心使用局部存在唯一性。它的经济含义是: 同样的经济状态和同样的制度环境, 不应该同时推出两条不同的未来路径。

### 16.1 为什么一维自治方程的路径不能交叉

考虑

$$\dot{x} = f(x).$$

如果从两个不同初始点出发的路径在某个时刻交于同一个状态 $\bar{x}$, 那么从这个时刻开始, 它们拥有相同当前状态。由于未来变化率只由当前状态决定, 后续路径必须相同。这会和之前两条路径不同相矛盾。因此在满足唯一性的情况下, 一维动态路径不会随意交叉。

**直觉盒子**

“路径不能交叉”是相图分析的隐含基础。它让我们可以只看箭头方向和稳态, 不必每次都求显式解。

## 17 斜率场 slope field: 不求解也能看路径

对非自治方程

$$\dot{x} = f(t, x),$$

每一个点 $(t, x)$ 都给出一条小斜率 $f(t, x)$。把这些小斜率画出来, 就得到 slope field。解曲线必须沿着这些斜率前进。

经济上, slope field 的价值是: 即使模型有时间变化政策、季节项、外生冲击, 我们仍然可以看出大致动态方向。例如

$$\dot{x} = t - x$$

表示目标水平随时间上升, 而 $x$ 不断向这个移动目标调整。

**Worked Example**

[title=Worked Example 3: 移动目标下的调整] 求解

$$\dot{x} + x = t, qquad x(0) = 0.$$

这是线性方程, 积分因子是 $\mu(t) = e^t$。两边乘以 $e^t$:

$$e^t\dot{x} + e^t x = te^t.$$

左边是

$$\frac{\mathrm{d}}{\mathrm{d}t}(e^t x).$$

因此

$$\frac{\mathrm{d}}{\mathrm{d}t}(e^t x) = te^t.$$

分部积分:

$$\int te^t \,\mathrm{d}t = te^t - e^t + C = e^t(t-1) + C.$$

所以

$$e^t x = e^t(t-1) + C,$$

即

$$x(t) = t - 1 + Ce^{-t}.$$

由 $x(0) = 0$ 得 $C = 1$, 所以

$$\boxed{x(t) = t - 1 + e^{-t}.}$$

这个结果说明 $x(t)$ 最终大约等于 $t-1$。它不是追上 $t$, 而是长期落后一个单位, 因为调整速度有限。

## 18 资本存量方程: 流量与存量的最基本关系

在宏观、公司金融和投资理论中, 常见形式是

$$\dot{K} = I(t) - \delta K.$$

其中 $K$ 是存量 stock, $I(t)$ 是投资流量 flow, $\delta K$ 是折旧流量。

**Worked Example**

[title=Worked Example 4: 常数投资下的资本积累] 设

$$\dot{K} = I - \delta K,$$

其中 $I > 0$, $\delta > 0$ 都是常数, 初始资本 $K(0) = K_0$。标准形式是

$$\dot{K} + \delta K = I.$$

稳态满足

$$\delta K^* = I,$$

所以

$$K^* = \frac{I}{\delta}.$$

用“稳态 + 偏离量”方法:

$$K(t) = K^* + (K_0 - K^*)e^{-\delta t} = \frac{I}{\delta} + \left(K_0 - \frac{I}{\delta}\right) e^{-\delta t}.$$

经济解释: 长期资本不是无限增长, 而是收敛到“投资流量 / 折旧率”。如果投资越高, 长期资本越高; 如果折旧率越高, 长期资本越低, 且调整速度越快。

# 19 比较静态: 稳态如何随参数变化

很多时候我们不只想知道路径如何走, 还想知道政策参数改变后, 长期稳态如何改变。设

$$\dot{x} = f(x, \theta),$$

稳态由

$$f(x^*(\theta), \theta) = 0$$

定义。对 $\theta$ 求导:

$$f_x(x^*, \theta)\frac{\mathrm{d}x^*}{\mathrm{d}\theta} + f_\theta(x^*, \theta) = 0.$$

如果 $f_x \neq 0$, 则

$$\boxed{\frac{\mathrm{d}x^*}{\mathrm{d}\theta} = -\frac{f_\theta}{f_x}.}$$

## 19.1 为什么这个公式重要

如果稳态稳定, 通常有 $f_x < 0$。此时若 $f_\theta > 0$, 即提高参数 $\theta$ 会在给定 $x$ 下提高 $\dot{x}$, 那么

$$\frac{\mathrm{d}x^*}{\mathrm{d}\theta} > 0.$$

经济语言是: 参数提高给系统更大的向上推动力, 稳定稳态会上移。

**Worked Example**

[title=Worked Example 5: Solow 稳态对储蓄率的比较静态] Solow 稳态满足

$$F(k, s) = sk^\alpha - (n + \delta)k = 0.$$

于是

$$F_s = k^{\alpha} > 0,$$

$$F_k = s\alpha k^{\alpha-1} - (n+\delta).$$

在稳定稳态处 $F_k < 0$。因此

$$\frac{\mathrm{d}k^*}{\mathrm{d}s} = -\frac{F_s}{F_k} > 0.$$

这说明储蓄率上升会提高长期人均有效资本。这个结论不需要先解出 $k^*$，只需要稳态方程和隐函数求导。

## 20 符号、术语与经济翻译

| 英文术语 | 中文含义 | 经济翻译 |
|---|---|---|
| state variable | 状态变量 | 描述系统当前位置的变量，如资本、债务、价格。 |
| flow variable | 流量变量 | 单位时间内的流入或流出，如投资、赤字、消费流。 |
| steady state | 稳态 | 净变化为零的长期状态。 |
| autonomous equation | 自治方程 | 变化率只依赖当前状态，不显式依赖时间。 |
| phase line | 相线图 | 一维动态中用箭头表示运动方向的图。 |
| local stability | 局部稳定 | 从稳态附近出发会回到稳态。 |
| linearization | 线性化 | 用一阶近似研究稳态附近动态。 |
| half-life | 半衰期 | 偏离量减半所需时间。 |
| integrating factor | 积分因子 | 把线性 ODE 左边变成乘积导数的工具。 |
| mean reversion | 均值回复 | 变量偏离长期均值后被拉回。 |

## 21 练习题

### A 组：基础计算

A1. 解 $\dot{x} = 3x$, $x(0) = 2$。

A2. 解 $\dot{x} = -0.4x$, $x(0) = 10$。求 $x(5)$。

A3. 解 $\dot{x} = 2tx$, $x(0) = 1$。

A4. 解 $\dot{x} + x = 4$, $x(0) = 0$。

A5. 解 $\dot{x} - 2x = 6$, $x(0) = 1$。

A6. 对 $\dot{x} = x(1 - x)$, 找稳态并判断每个稳态的稳定性。

A7. 对 $\dot{x} = (x - 1)(x - 3)$, 画相线图并判断稳定性。

A8. 若 $x(t) - x^* = 10\mathrm{e}^{-0.2t}$, 求 half-life。

### B 组: 概念判断

B1. 为什么 $\dot{x} = f(x)$ 的稳态由 $f(x) = 0$ 给出?

B2. 为什么 $f'(x^*) < 0$ 表示局部稳定?

B3. 解释 $\dot{x} = -a(x - \bar{x})$ 中 $a$ 的经济含义。

B4. 稳态存在是否意味着路径一定收敛到稳态? 说明原因。

B5. 在价格调整模型 $\dot{p} = \lambda[D(p) - S(p)]$ 中, 为什么 $D'(p^*) - S'(p^*) < 0$ 表示稳定?

### C 组: 混合题

C1. 解 logistic 方程 $\dot{x} = rx(1 - x)$, $x(0) = x_0 \in (0, 1)$。

C2. 给定 $\dot{x} = 5 - 0.5x$, 求稳态、显式解和 half-life。

C3. 给定 $\dot{x} = x^2 - 4x + 3$, 找稳态并判断稳定性。

C4. 给定 $\dot{x} = t - x$, $x(0) = 0$, 用积分因子法求解。

C5. 若 $\dot{x} = f(x, \theta)$ 且 $f(x^*(\theta), \theta) = 0$, 在 $f_x \neq 0$ 时推导 $\mathrm{d}x^*/\mathrm{d}\theta$。

### D 组: 经济金融建模题

D1. Solow 模型中 $\dot{k} = sk^{\alpha} - (n + \delta)k$。求稳态 $k^*$, 并判断稳定性。

D2. 债务动态 $\dot{b} = (r - g)b + d$。分别讨论 $r > g$ 与 $r < g$ 的长期含义。

D3. 价格调整模型中 $D(p) = 10 - p$, $S(p) = 2 + 3p$, $\lambda = 0.5$。求均衡价格并判断稳定性。

D4. 均值回复模型 $\dot{x} = -a(x - \bar{x})$。给定 $a = 0.25$, $\bar{x} = 4$, $x(0) = 10$, 求路径和 half-life。

D5. 一个企业的客户数量 $N(t)$ 满足 $\dot{N} = 0.3N(1 - N/1000)$。解释 0, 1000 两个稳态的含义和稳定性。

### E 组: 挑战 / 证明题

E1. 证明: 对一维自治方程 $\dot{x} = f(x)$, 若 $f'(x^*) < 0$, 则线性化系统在 $x^*$ 附近稳定。

E2. 设 $\dot{x} = f(x)$, 且存在 $x_1 < x_2$ 使 $f(x) > 0$ 对所有 $x \in (x_1, x_2)$ 成立。解释为什么路径会向右移动。

E3. 对常系数线性方程 $\dot{x} + ax = b$, 不使用积分因子, 通过“稳态 + 偏离量”方法推导通解。

E4. 在 Solow 模型中, 设 $f(k)$ 满足 $f' > 0$, $f'' < 0$, Inada 条件。说明为什么存在唯一正稳态。

E5. 对 $\dot{p} = \lambda z(p)$, 若 $z'(p^*) > 0$, 用线性化解释为什么均衡不稳定。

## 22 练习题详细解答

### A 组答案

A1. $\dot{x} = 3x$ 的解为 $x(t) = Ce^{3t}$。由 $x(0) = 2$ 得 $C = 2$, 所以

$$x(t) = 2e^{3t}.$$

A2. $x(t) = 10e^{-0.4t}$。因此

$$x(5) = 10e^{-2} \approx 1.353.$$

A3. 分离变量:

$$\frac{\dot{x}}{x} = 2t.$$

积分得 $\ln x = t^2 + C$。由 $x(0) = 1$ 得 $C = 0$, 所以

$$x(t) = e^{t^2}.$$

A4. 稳态为 $x^* = 4$。解为

$$x(t) = 4 + (0 - 4)e^{-t} = 4(1 - e^{-t}).$$

A5. 写成 $\dot{x} - 2x = 6$。常数特解满足 $-2x^* = 6$, 所以 $x^* = -3$。偏离量满足 $\dot{y} = 2y$, 故

$$x(t) = -3 + (1 + 3)e^{2t} = -3 + 4e^{2t}.$$

A6. 稳态为 $x = 0, 1$。$f(x) = x(1 - x)$, $f'(x) = 1 - 2x$。$f'(0) = 1 > 0$, 所以 $0$ 不稳定; $f'(1) = -1 < 0$, 所以 $1$ 稳定。

A7. $f(x) = (x-1)(x-3)$。稳态为 $1, 3$。区间 $(-\infty, 1)$ 上 $f > 0$, $(1, 3)$ 上 $f < 0$, $(3, \infty)$ 上 $f > 0$。因此 $x = 1$ 稳定, $x = 3$ 不稳定。

A8. $T_{1/2} = \ln 2 / 0.2 \approx 3.466$。

### B 组答案

B1. 稳态表示状态不再变化, 即 $\dot{x} = 0$。因为 $\dot{x} = f(x)$, 所以稳态必须满足 $f(x) = 0$。

B2. 在 $x^*$ 附近,

$$\dot{x} = f(x) \approx f'(x^*)(x - x^*).$$

若 $f'(x^*) < 0$, 偏离量满足指数衰减, 所以路径回到稳态。

B3. $a$ 是调整速度。$a$ 越大, $x(t)$ 越快回到长期均值 $\bar{x}$。half-life 为 $\ln 2/a$。

B4. 不一定。稳态可能不稳定。例如 $\dot{x} = x$ 有稳态 $x^* = 0$, 但任何非零初始值都会远离 0。

B5. 若 $p > p^*$ 时 excess demand 为负, 价格下降; 若 $p < p^*$ 时 excess demand 为正, 价格上升, 则价格回到均衡。条件 $D'(p^*) - S'(p^*) < 0$ 正是线性化意义上的这种回拉力量。

## C 组答案

C1. 对 $\dot{x} = rx(1 - x)$,

$$\frac{\mathrm{d}x}{x(1-x)} = r\,\mathrm{d}t.$$

因为

$$\frac{1}{x(1-x)} = \frac{1}{x} + \frac{1}{1-x},$$

所以

$$\ln \frac{x}{1-x} = rt + C.$$

令 $A = e^C$, 得

$$\frac{x}{1-x} = Ae^{rt}.$$

解出

$$x(t) = \frac{Ae^{rt}}{1 + Ae^{rt}}.$$

由 $x(0) = x_0$ 得 $A = x_0/(1 - x_0)$, 因此

$$\boxed{x(t) = \frac{x_0 e^{rt}}{1 - x_0 + x_0 e^{rt}}.}$$

C2. $\dot{x} = 5 - 0.5x$。稳态满足 $5 - 0.5x^* = 0$, 所以 $x^* = 10$。解为

$$x(t) = 10 + (x_0 - 10)e^{-0.5t}.$$

half-life 为 $\ln 2/0.5 = 2\ln 2 \approx 1.386$。

C3. $f(x) = x^2 - 4x + 3 = (x - 1)(x - 3)$, 稳态为 $1, 3$。由前面相线可知 1 稳定, 3 不稳定。

C4. $\dot{x} + x = t$。积分因子 $\mu(t) = e^t$。于是

$$\frac{\mathrm{d}}{\mathrm{d}t}(e^t x) = te^t.$$

积分得

$$e^t x = e^t(t - 1) + C.$$

因此

$$x = t - 1 + Ce^{-t}.$$

由 $x(0) = 0$ 得 $0 = -1 + C$, 所以 $C = 1$。最终

$$x(t) = t - 1 + e^{-t}.$$

C5. 对稳态条件

$$f(x^*(\theta), \theta) = 0$$

两边对 $\theta$ 求导:

$$f_x(x^*, \theta)\frac{\mathrm{d}x^*}{\mathrm{d}\theta} + f_\theta(x^*, \theta) = 0.$$

若 $f_x \neq 0$, 则

$$\boxed{\frac{\mathrm{d}x^*}{\mathrm{d}\theta} = -\frac{f_\theta}{f_x}.}$$

## D 组答案

D1. 稳态满足

$$s(k^*)^\alpha = (n + \delta)k^*.$$

正稳态为

$$k^* = \left(\frac{s}{n+\delta}\right)^{1/(1-\alpha)}.$$

令 $F(k) = sk^\alpha - (n + \delta)k$, 则

$$F'(k^*) = -(1 - \alpha)(n + \delta) < 0,$$

所以局部稳定。

D2. $\dot{b} = (r - g)b + d$。若 $r > g$, 指数项 $e^{(r-g)t}$ 增长, 债务比率有发散压力。若 $r < g$, 指数项衰减, 初始债务的影响会被经济增长逐渐稀释。

D3. 均衡满足 $10 - p = 2 + 3p$, 所以 $p^* = 2$。excess demand 为

$$z(p) = 8 - 4p.$$

$z'(p) = -4 < 0$, 且 $\lambda > 0$, 所以均衡稳定。

D4. 路径为

$$x(t) = 4 + (10 - 4)e^{-0.25t} = 4 + 6e^{-0.25t}.$$

half-life 为

$$T_{1/2} = \frac{\ln 2}{0.25} = 4\ln 2 \approx 2.773.$$

D5. 稳态为 $N = 0$ 和 $N = 1000$。当 $0 < N < 1000$ 时, $\dot{N} > 0$, 客户数量增长; 当 $N > 1000$ 时,

$\dot{N} < 0$, 客户数量下降。因此 1000 稳定, 0 在正半轴上不稳定。经济上 1000 表示市场容量上限。

### E 组答案

E1. 令 $y = x - x^*$。因为 $f(x^*) = 0$, 在 $x^*$ 附近有

$$\dot{y} = \dot{x} = f(x) \approx f'(x^*)y.$$

线性化方程的解为 $y(t) = y(0)e^{f'(x^*)t}$。若 $f'(x^*) < 0$, 则 $y(t) \to 0$, 所以线性化系统稳定。

E2. 因为 $\dot{x} = f(x) > 0$, 所以状态变量随时间增加。只要路径仍在区间 $(x_1, x_2)$ 内, 它就向右移动。

E3. 稳态 $x^* = b/a$。令 $y = x - x^*$, 则

$$\dot{y} = \dot{x} = b - ax = b - a(y + x^*) = b - ay - ax^* = -ay.$$

因此 $y(t) = y(0)e^{-at}$, 所以

$$x(t) = \frac{b}{a} + \left(x_0 - \frac{b}{a}\right) e^{-at}.$$

E4. 设 $F(k) = sf(k) - (n+\delta)k$。Inada 条件给出当 $k$ 很小时, $sf'(k)$ 很大, 因而 $F(k)$ 从正方向上升; 当 $k$ 很大时, 线性项 $(n+\delta)k$ 占主导, $F(k) < 0$。连续性保证存在交点。又因 $f'' < 0$, $sf(k)$ 是严格凹函数, 与直线 $(n+\delta)k$ 至多有一个正交点。因此正稳态唯一。

E5. 线性化:

$$\dot{p} \approx \lambda z'(p^*)(p - p^*).$$

若 $z'(p^*) > 0$, 且 $\lambda > 0$, 偏离量满足指数增长, 因此均衡不稳定。

## 23 本章到下一章的连接

本章处理连续时间动态:

$$\dot{x} = f(x, t).$$

下一章将处理离散时间动态:

$$x_{t+1} = F(x_t, t).$$

两者看起来相似, 但稳定条件会有重要区别。连续时间线性系统中

$$\dot{x} = -ax, \quad a > 0$$

稳定; 离散时间中

$$x_{t+1} = ax_t$$

稳定要求

$$|a| < 1.$$

这个区别会在蛛网模型、AR(1)、资产价格递推和理性预期模型中反复出现。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 11

# 差分方程与离散时间经济学

Chapter 10 用 ODE 描述连续时间动态。本章转向离散时间：变量只在 $t = 0, 1, 2, \ldots$ 这些时期被观察或决策。离散时间不是连续时间的低配版本，而是宏观、金融、计量和动态规划里最常用的表达方式。资产价格递推、蛛网模型、离散 Solow、AR(1) 时间序列、理性预期差分方程和 Bellman 方程，背后都依赖差分方程。本章目标是让你真正掌握三个问题：路径怎么解，稳态是否稳定，经济模型里的前瞻项和泡沫项为什么会出现。

math4econ personalized notes
2026

# 目录

**Chapter 11 总览** **2**

**1 本章为什么重要** **2**
1.1 连续时间与离散时间的区别 . . . . . 2
1.2 离散时间模型的五个关键词 . . . . . 3

**2 最小前置知识** **3**

**3 一阶线性差分方程** **3**
3.1 最核心模型 . . . . . 4
3.2 从递推展开到通解 . . . . . 4
3.3 稳定性分类 . . . . . 5

**4 计算方法模板** **6**

**5 非齐次项与时间变化外生项** **7**
5.1 政策乘数的动态解释 . . . . . 7

**6 蛛网模型 Cobweb Model** **8**
6.1 模型设定 . . . . . 8
6.2 稳态与稳定性 . . . . . 8

**7 离散时间 Solow 模型** **9**
7.1 基本方程 . . . . . 9
7.2 局部稳定性 . . . . . 10

**8 资产价格递推与泡沫项** **11**
8.1 向前迭代 . . . . . 11
8.2 常数股利与泡沫 . . . . . 12

**9 AR(1): 计量经济学中的差分方程** **12**
9.1 确定性版本 . . . . . 13
9.2 随机版本 . . . . . 13
9.3 半衰期 half-life . . . . . 14

**10 二阶差分方程与特征根** **14**
10.1 为什么猜 $x_t = \lambda^t$ . . . . . 14
10.2 路径形状 . . . . . 15

**11 系统差分方程与矩阵形式** **16**

**12 理性预期与前瞻差分方程入口** **17**

**13 Worked Examples** **18**

**14 常见错误与辨析** **20**

**15 本章小结: 概念 - 公式 - 应用** **20**

**16 练习题** **20**

A 组: 基础计算 . . . . . 20
B 组: 概念判断 . . . . . 21
C 组: 混合题 . . . . . 21
D 组: 经济金融建模题 . . . . . 21
E 组: 挑战 / 证明题 . . . . . 21

**17 练习题参考答案** **23**

A 组答案 . . . . . 23
B 组答案 . . . . . 23
C 组答案 . . . . . 24
D 组答案 . . . . . 24
E 组答案 . . . . . 25

**18 后续学习连接** **27**

# Chapter 11 总览

| | |
|---|---|
| **本章主题** | 差分方程 difference equations 与离散时间经济学。 |
| **核心问题** | 如果变量满足 $x_{t+1} = F(x_t, t)$ 或 $x_t = F(x_{t+1}, t)$，那么从初始状态出发路径如何演化，是否收敛到稳态，出现震荡或爆炸的条件是什么? |
| **数学工具** | 一阶线性差分方程、稳态、稳定性、特征根、二阶差分方程、系统差分方程、相图、前瞻方程、横截条件。 |
| **经济应用** | 离散时间 Solow、蛛网模型 cobweb model、资产定价递推、债务动态、AR(1)、理性预期基础、金融泡沫。 |
| **后续连接** | Chapter 12 线性系统与特征值，Chapter 13 非线性动态系统，Chapter 18 比较静态，Chapter 25 数值方法，以及高级宏观中的动态规划。 |

**直觉盒子**

ODE 问的是：每一个瞬间变量怎么动。差分方程问的是：每一期到下一期变量怎么变。它把时间切成 $0, 1, 2, \ldots$，于是动态规律变成

$$x_{t+1} = F(x_t).$$

你可以把它想成一台机器：输入今天的状态 $x_t$，输出明天的状态 $x_{t+1}$。反复把输出再放回输入，就得到整条路径。

## 1 本章为什么重要

经济金融中大量模型天然是离散时间的。原因很简单：数据按季度、年度、日度记录；家庭和企业按期决策；金融资产按期支付股利；计量模型按样本期排列。因此差分方程不是数学附属品，而是经济动态建模的基础语言。

### 1.1 连续时间与离散时间的区别

Chapter 10 的一阶 ODE 典型形式是

$$\dot{x}(t) = f(x(t)).$$

本章的差分方程典型形式是

$$x_{t+1} = F(x_t).$$

二者的共同点是都描述动态。区别在于：

- ODE 研究连续路径 $x(t)$；差分方程研究序列 $\{x_t\}_{t=0}^{\infty}$。
- ODE 的稳定性常看导数 $f'(x^*)$；一阶差分方程的稳定性常看斜率 $F'(x^*)$。
- ODE 的指数解是 $e^{at}$；差分方程的基本解是 $a^t$。
- 连续时间的均值回复是 $\dot{x} = -\lambda(x - \bar{x})$；离散时间的均值回复是 $x_{t+1} - \bar{x} = \rho(x_t - \bar{x})$。

**常见误区**

很多同学把 $x_{t+1} = ax_t + b$ 当成普通代数式看，这是错误的。它不是一个关于 $x$ 的方程，而是一个递推规则 recursion。你要解的是一整串数：$x_0, x_1, x_2, \ldots$。

### 1.2 离散时间模型的五个关键词

本章围绕五个关键词展开。

1. **recursion 递推:** 今天决定明天，明天再决定后天。
2. **steady state 稳态:** 如果进入某一点后不再变化，即 $x_{t+1} = x_t = x^*$。
3. **stability 稳定性:** 如果从附近出发能回到稳态，则稳定。
4. **root 特征根:** 线性动态的形状由根决定。
5. **forward-looking 前瞻性:** 有些经济方程不是由过去推未来，而是由未来约束现在。

## 2 最小前置知识

| 前置知识 | 本章如何使用 |
|---|---|
| 等比数列 | 解 $x_{t+1} = ax_t$ 和资产价格递推。 |
| 一元函数图像 | 用 $x_{t+1} = F(x_t)$ 与 $45^\circ$ 线画 cobweb。 |
| 二次方程 | 二阶差分方程的特征根。 |
| 线性代数初步 | 系统差分方程 $\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t + \boldsymbol{b}$。 |
| 概率与期望直觉 | AR(1)、随机冲击和条件期望。 |

## 3 一阶线性差分方程

## 3.1 最核心模型

**定义 / 定理**

**一阶线性差分方程.** 形式为

$$x_{t+1} = ax_t + b, \qquad t = 0, 1, 2, \dots$$

的递推关系称为一阶线性差分方程。这里 $a, b$ 是常数，$x_0$ 是初始条件。

它是全章最重要的方程。你在宏观、金融、计量里会不断见到它:

- $k_{t+1} = ak_t + b$: 资本积累的线性近似;
- $b_{t+1} = \frac{1+r}{1+g} b_t + d$: 债务/GDP 动态;
- $p_{t+1} = \alpha + \beta p_t$: 价格调整或预期递推;
- $y_t = \rho y_{t-1} + \varepsilon_t$: AR(1) 时间序列。

## 3.2 从递推展开到通解

先从初始值 $x_0$ 展开:

$$\begin{aligned} x_1 &= ax_0 + b, \\ x_2 &= ax_1 + b = a(ax_0 + b) + b = a^2 x_0 + ab + b, \\ x_3 &= ax_2 + b = a^3 x_0 + a^2 b + ab + b. \end{aligned}$$

所以一般地,

$$x_t = a^t x_0 + b(1 + a + \cdots + a^{t-1}).$$

若 $a \neq 1$，利用等比数列求和得到

$$1 + a + \cdots + a^{t-1} = \frac{1 - a^t}{1 - a},$$

因此

$$\boxed{x_t = a^t x_0 + \frac{1 - a^t}{1 - a} b.}$$

这个公式还可以写成更有经济含义的形式。稳态 $x^*$ 满足

$$x^* = ax^* + b,$$

所以

$$\boxed{x^* = \frac{b}{1 - a}}, \qquad a \neq 1.$$

把它代入通解：

$$\boxed{x_t - x^* = a^t(x_0 - x^*).}$$

**直觉盒子**

这就是一阶线性差分方程最重要的结论：**偏离稳态的距离每一期乘以** $a$。如果 $|a| < 1$，偏离会逐渐缩小；如果 $|a| > 1$，偏离会放大；如果 $a < 0$，路径会在稳态两边来回震荡。

## 3.3 稳定性分类

**定义 / 定理**

对一阶线性差分方程

$$x_{t+1} = ax_t + b,$$

若 $a \neq 1$，稳态为 $x^* = b/(1-a)$。其稳定性由 $a$ 决定：

- $|a| < 1$：稳态稳定，$x_t \to x^*$；
- $0 < a < 1$：单调收敛或不改变方向的收敛；
- $-1 < a < 0$：震荡收敛；
- $a > 1$：发散；
- $a < -1$：震荡发散；
- $a = 1$：无有限稳态，若 $b \neq 0$ 则线性漂移；
- $a = -1$：二周期震荡或不收敛。

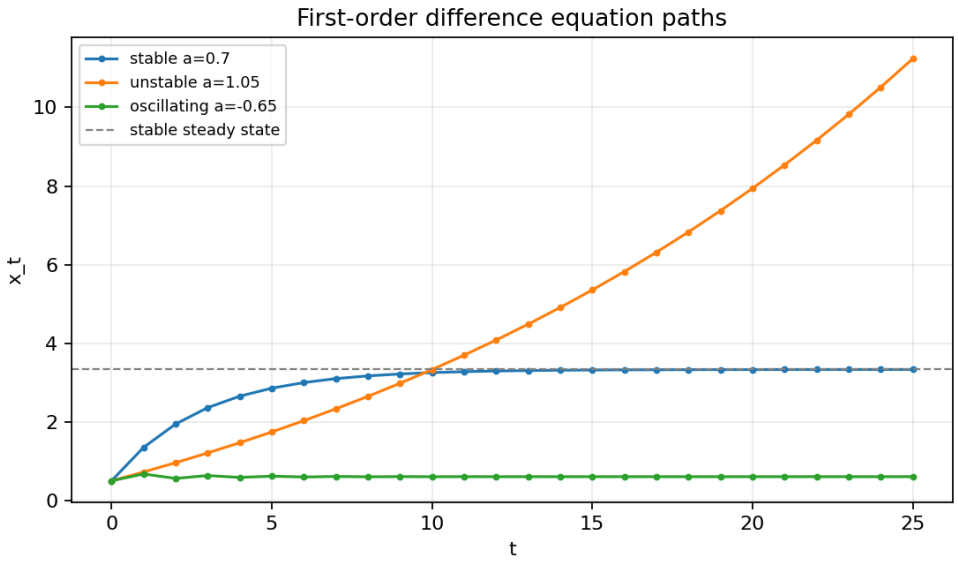



图 1：一阶差分方程的几类路径。稳定性本质上由 $a^t$ 是否趋于 0 决定。

# 4 计算方法模板

**方法模板**

**一阶线性差分方程解题模板**

面对

$$x_{t+1} = ax_t + b,$$

按以下步骤做：

1. 求稳态：令 $x_{t+1} = x_t = x^*$，得到 $x^* = ax^* + b$。
2. 解得 $x^* = b/(1-a)$，前提是 $a \neq 1$。
3. 写偏离形式：$x_{t+1} - x^* = a(x_t - x^*)$。
4. 递推得到：$x_t - x^* = a^t(x_0 - x^*)$。
5. 判断稳定性：看 $|a| < 1$ 是否成立。
6. 翻译经济含义：收敛、发散、震荡、均值回复或爆炸。

**Worked Example**

**例 1：债务/GDP 动态**

设政府债务占 GDP 比例为 $b_t$，实际利率为 $r$，GDP 增长率为 $g$，初级赤字占 GDP 比例为 $d$。离散时间债务动态可写成

$$b_{t+1} = \frac{1+r}{1+g} b_t + d.$$

令

$$a = \frac{1+r}{1+g}.$$

若 $a \neq 1$，稳态为

$$b^* = \frac{d}{1-a} = \frac{d}{1-\frac{1+r}{1+g}} = \frac{d(1+g)}{g-r}.$$

稳定条件是

$$\left|\frac{1+r}{1+g}\right| < 1.$$

在正常情形 $1+r>0, 1+g>0$ 下，这等价于

$$r < g.$$

经济解释：如果经济增长率高于实际利率，旧债务相对 GDP 会被增长稀释，债务比率更可能稳定；如果 $r > g$，债务滚动压力会放大，长期可持续性需要足够的初级盈余来抵消。

**常见误区**

债务动态里的 $r < g$ 不是说政府永远可以无限借债。它只是说明线性递推中旧债务项的系数小于 1。真实财政可持续性还取决于赤字路径、政治约束、市场利率内生变化和风险溢价。

## 5 非齐次项与时间变化外生项

经济模型里常见的是

$$x_{t+1} = ax_t + z_t,$$

其中 $z_t$ 是随时间变化的外生冲击、政策项或收入项。

递推展开：

$$\begin{aligned} x_1 &= ax_0 + z_0, \\ x_2 &= a^2x_0 + az_0 + z_1, \\ x_3 &= a^3x_0 + a^2z_0 + az_1 + z_2. \end{aligned}$$

一般地，

$$\boxed{x_t = a^t x_0 + \sum_{j=0}^{t-1} a^{t-1-j} z_j.}$$

**直觉盒子**

这个公式非常重要：今天的状态等于两部分。第一部分 $a^t x_0$ 是初始条件的遗留影响；第二部分是过去所有冲击 $z_j$ 的累积影响，但越久远的冲击乘上越高次的 $a$。如果 $|a| < 1$，远古冲击的影响会衰减。

### 5.1 政策乘数的动态解释

如果 $z_j$ 是某期政策冲击，那么它对 $x_t$ 的影响是

$$\frac{\partial x_t}{\partial z_j} = a^{t-1-j}, \qquad j < t.$$

这说明 $a$ 不只是数学系数，也是政策影响的持久性 persistence。

**Worked Example**

**例 2：暂时性刺激与持久性**

设产出缺口满足

$$y_{t+1} = 0.8y_t + g_t,$$

其中 $g_t$ 是财政刺激。若只有第 0 期有一次刺激 $g_0 = 1$，且 $y_0 = 0$，则

$$y_1 = 1, \quad y_2 = 0.8, \quad y_3 = 0.8^2, \quad \ldots, \quad y_t = 0.8^{t-1}.$$

一次性冲击不会永久影响水平，但如果 0.8 接近 1，影响会持续很久。

# 6 蛛网模型 Cobweb Model

蛛网模型是理解一阶差分方程经济含义的经典例子。它研究农业或生产有时滞的市场：供给由上一期价格决定，而需求由当前价格决定。

## 6.1 模型设定

设需求函数为

$$D(p_t) = \alpha - \beta p_t, \qquad \beta > 0,$$

供给函数由上一期价格决定：

$$S(p_{t-1}) = \gamma + \delta p_{t-1}, \qquad \delta > 0.$$

市场出清要求

$$D(p_t) = S(p_{t-1}).$$

于是

$$\alpha - \beta p_t = \gamma + \delta p_{t-1}.$$

整理得到价格递推：

$$\boxed{p_t = \frac{\alpha - \gamma}{\beta} - \frac{\delta}{\beta} p_{t-1}.}$$

这就是一阶线性差分方程，且斜率为负。

## 6.2 稳态与稳定性

稳态价格满足 $p_t = p_{t-1} = p^*$：

$$p^* = \frac{\alpha - \gamma}{\beta} - \frac{\delta}{\beta} p^*.$$

所以

$$p^* = \frac{\alpha - \gamma}{\beta + \delta}.$$

稳定条件为

$$\left| -\frac{\delta}{\beta} \right| < 1,$$

即

$$\boxed{\delta < \beta.}$$

经济解释：供给曲线对价格的反应不能太强。如果供给对上一期价格过度反应，价格会在均衡两侧越来越大幅度震荡。

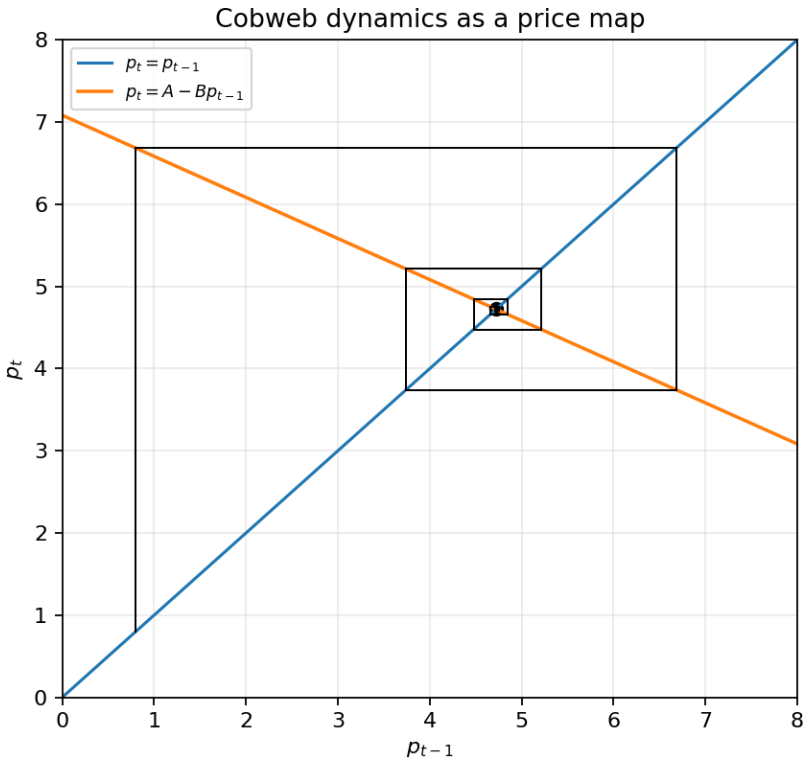



图 2: 蛛网模型可以看成价格映射 $p_t = A - Bp_{t-1}$ 与 $45^\circ$ 线之间的迭代。

**常见误区**

蛛网模型的负斜率意味着价格通常是震荡调整。稳定并不等于单调靠近均衡。只要 $-1 < -\delta/\beta < 0$，价格会左右震荡，但震荡幅度逐渐缩小。

# 7 离散时间 Solow 模型

Chapter 10 中我们看过连续时间 Solow:

$$\dot{k} = sf(k) - (n + \delta)k.$$

离散时间版本更加接近很多宏观教材和数值模拟。

## 7.1 基本方程

设人均资本满足

$$k_{t+1} = \frac{sf(k_t) + (1 - \delta)k_t}{1 + n}.$$

其中 $s$ 是储蓄率，$\delta$ 是折旧率，$n$ 是人口增长率。稳态满足

$$k^* = \frac{sf(k^*) + (1-\delta)k^*}{1+n}.$$

移项得

$$(n+\delta)k^* = sf(k^*).$$

这和连续时间稳态条件形式一致。

## 7.2 局部稳定性

定义映射

$$G(k) = \frac{sf(k) + (1-\delta)k}{1+n}.$$

模型为

$$k_{t+1} = G(k_t).$$

稳态附近线性化：

$$k_{t+1} - k^* \approx G'(k^*)(k_t - k^*).$$

因此局部稳定条件为

$$\boxed{|G'(k^*)| < 1.}$$

而

$$G'(k) = \frac{sf'(k) + 1 - \delta}{1+n}.$$

在标准凹生产函数下，通常稳态附近满足稳定条件。

**Advanced Core**

**从线性一阶方程到非线性映射**

即使 $k_{t+1} = G(k_t)$ 不是线性的，稳态附近仍然可以把它近似为

$$k_{t+1} - k^* \approx G'(k^*)(k_t - k^*).$$

这就是局部线性化。你可以把 $G'(k^*)$ 看成非线性动态在稳态附近的“有效 $a$”。所以一阶线性差分方程不是特例而已，它是理解一般动态系统的局部模板。

**Worked Example**

**例 3: Cobb-Douglas 离散 Solow 的稳态**

设 $f(k) = k^\alpha$，$0 < \alpha < 1$。稳态条件为

$$s(k^*)^\alpha = (n+\delta)k^*.$$

若 $k^* > 0$，两边除以 $(k^*)^\alpha$：

$$s = (n+\delta)(k^*)^{1-\alpha}.$$

因此

$$\boxed{k^* = \left(\frac{s}{n+\delta}\right)^{\frac{1}{1-\alpha}}.}$$

储蓄率 $s$ 上升会提高稳态资本；人口增长和折旧上升会降低稳态资本。

# 8 资产价格递推与泡沫项

差分方程在金融中最经典的应用是资产定价。考虑一个资产，下一期支付股利 $d_{t+1}$，折现因子为 $R = 1 + r$。无套利定价的简化形式是

$$p_t = \frac{d_{t+1} + p_{t+1}}{R}.$$

这不是普通的一阶后顾 backward equation，而是前瞻 forward-looking equation，因为今天的价格取决于明天的价格。

## 8.1 向前迭代

从

$$p_t = \frac{d_{t+1}}{R} + \frac{p_{t+1}}{R}$$

开始，将 $p_{t+1}$ 继续展开：

$$p_{t+1} = \frac{d_{t+2}}{R} + \frac{p_{t+2}}{R}.$$

代入得

$$p_t = \frac{d_{t+1}}{R} + \frac{d_{t+2}}{R^2} + \frac{p_{t+2}}{R^2}.$$

继续迭代 $T$ 期：

$$\boxed{p_t = \sum_{j=1}^{T} \frac{d_{t+j}}{R^j} + \frac{p_{t+T}}{R^T}.}$$

若满足横截条件 transversality condition

$$\lim_{T\to\infty} \frac{p_{t+T}}{R^T} = 0,$$

则基本面价格为

$$\boxed{p_t = \sum_{j=1}^{\infty} \frac{d_{t+j}}{R^j}.}$$

## 8.2 常数股利与泡沫

若 $d_t = d$ 为常数，则基本面价格为

$$p^F = \sum_{j=1}^{\infty} \frac{d}{R^j} = \frac{d}{R-1} = \frac{d}{r}.$$

但是差分方程还允许泡沫解。令

$$p_t = p^F + B_t.$$

代入定价方程可得泡沫项必须满足

$$B_t = \frac{B_{t+1}}{R},$$

也就是

$$B_{t+1} = RB_t.$$

所以

$$\boxed{B_t = R^t B_0.}$$

只要 $B_0 \neq 0$，泡沫项会按 $R$ 增长。

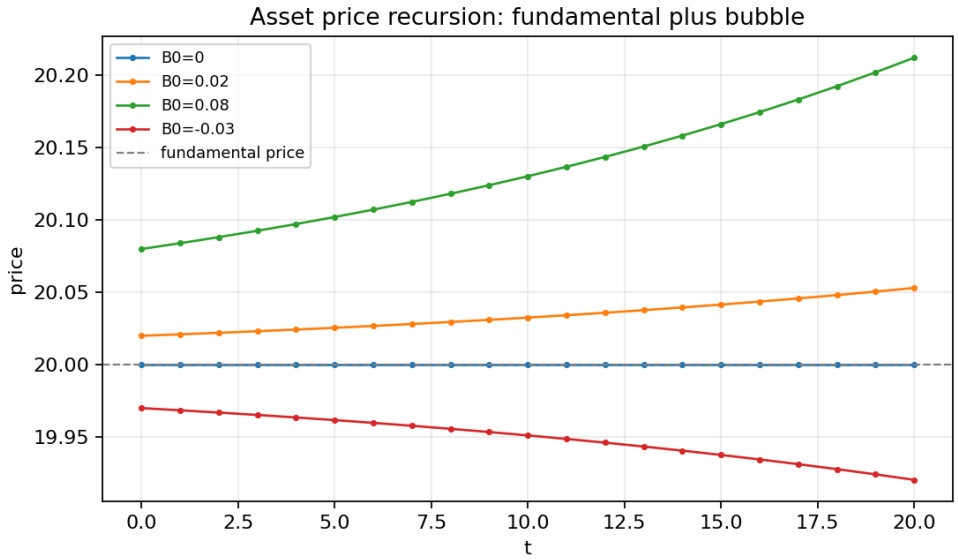



图 3: 资产价格递推中的基本面价格与泡沫项。横截条件用于排除爆炸性泡沫。

**直觉盒子**

前瞻差分方程最容易出错的地方是方向。后顾方程通常给定 $x_0$ 向前推；前瞻方程则常常需要一个终端条件或横截条件来排除不合理路径。金融资产定价中的“泡沫项”正是差分方程通解中未被基本面决定的齐次解。

# 9 AR(1)：计量经济学中的差分方程

## 9.1 确定性版本

AR(1) 的确定性骨架是

$$x_t = \rho x_{t-1}.$$

解为

$$x_t = \rho^t x_0.$$

稳定条件是

$$|\rho| < 1.$$

如果 $0 < \rho < 1$，变量逐渐回到 0；如果 $\rho$ 接近 1，冲击影响非常持久；如果 $\rho < 0$，变量在正负之间震荡。

## 9.2 随机版本

计量中更常见的是

$$x_t = \rho x_{t-1} + \varepsilon_t,$$

其中 $\varepsilon_t$ 是均值为 0 的冲击。向后展开：

$$\begin{aligned} x_t &= \rho x_{t-1} + \varepsilon_t \\ &= \rho(\rho x_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \rho^2 x_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t. \end{aligned}$$

继续展开得

$$\boxed{x_t = \rho^t x_0 + \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j}.}$$

若 $|\rho| < 1$，初始条件影响消失，过程有均值回复性质。

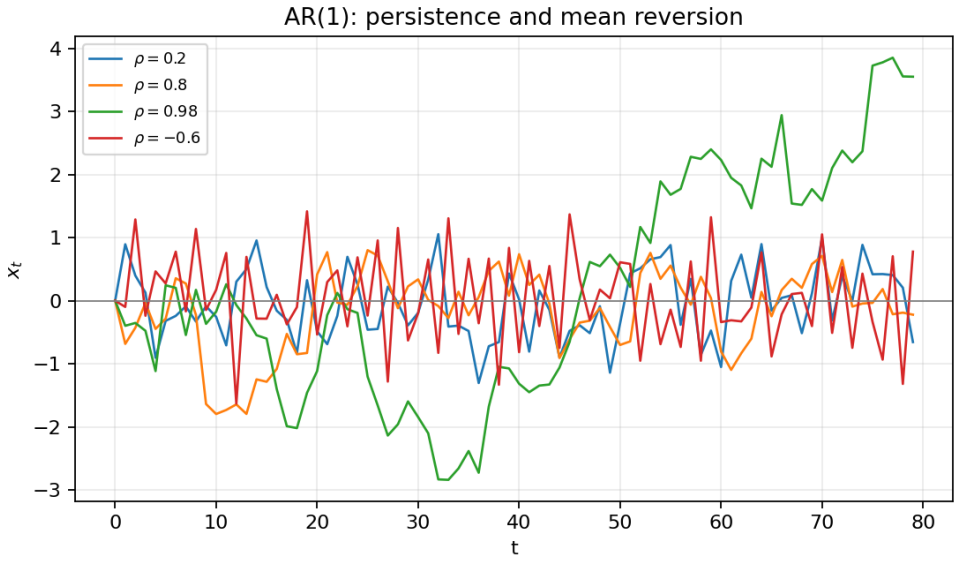



图 4: AR(1) 中 $\rho$ 控制持久性。$\rho$ 越接近 1，冲击消退越慢。

### 9.3 半衰期 half-life

若 $0 < \rho < 1$，一次冲击的影响在 $h$ 期后变成原来的一半：

$$\rho^h = \frac{1}{2}.$$

因此

$$\boxed{h = \frac{\log(1/2)}{\log \rho}.}$$

这在宏观和金融时间序列中非常常用：它把抽象的 $\rho$ 翻译成“冲击大约多久消退一半”。

## 10 二阶差分方程与特征根

一阶方程只依赖上一期。二阶差分方程依赖前两期：

$$x_{t+2} + a_1 x_{t+1} + a_0 x_t = 0.$$

它在经济学中常出现于动态优化、库存调整、蛛网模型扩展和理性预期模型。

### 10.1 为什么猜 $x_t = \lambda^t$

对于线性齐次方程，尝试指数型路径

$$x_t = \lambda^t.$$

代入

$$x_{t+2} + a_1 x_{t+1} + a_0 x_t = 0$$

得到

$$\lambda^{t+2} + a_1\lambda^{t+1} + a_0\lambda^t = 0.$$

如果 $\lambda \neq 0$，两边除以 $\lambda^t$：

$$\lambda^2 + a_1\lambda + a_0 = 0.$$

这叫特征方程 characteristic equation。

**定义 / 定理**

**二阶线性齐次差分方程的解.** 若特征方程

$$\lambda^2 + a_1\lambda + a_0 = 0$$

有两个不同实根 $\lambda_1, \lambda_2$，则通解为

$$\boxed{x_t = C_1\lambda_1^t + C_2\lambda_2^t.}$$

若 $|\lambda_1| < 1$ 且 $|\lambda_2| < 1$，则 $x_t \to 0$。

## 10.2 路径形状

根决定路径形状：

- 正实根：单调或非震荡路径；
- 负实根：符号交替，出现震荡；
- 复根：周期性震荡；
- 根的模小于 1：收敛；
- 根的模大于 1：爆炸。

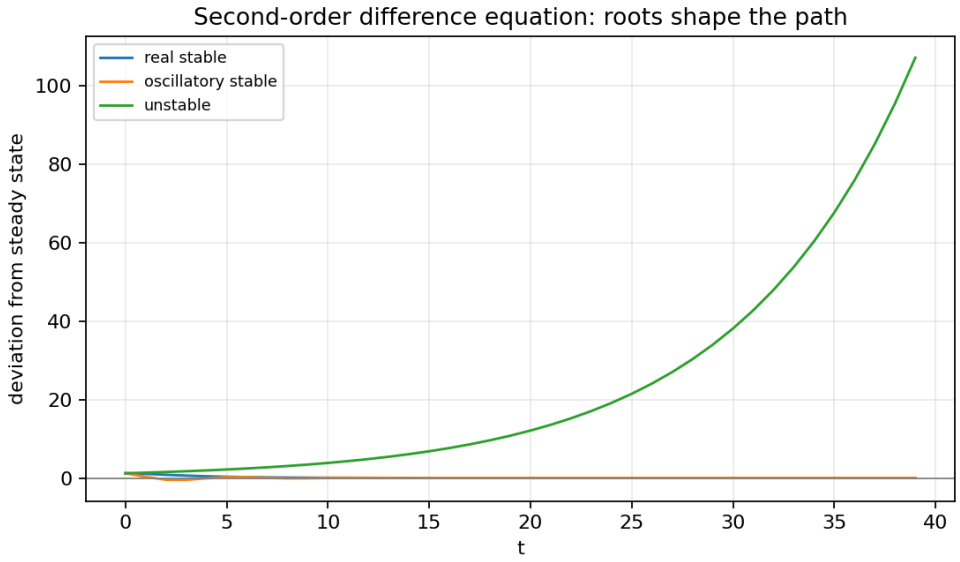



图 5: 二阶差分方程中，特征根决定路径是否震荡、收敛或发散。

**Worked Example**

**例 4: 二阶方程的完整求解**

求解

$$x_{t+2} - 1.2x_{t+1} + 0.35x_t = 0.$$

特征方程为

$$\lambda^2 - 1.2\lambda + 0.35 = 0.$$

因式分解:

$$(\lambda - 0.7)(\lambda - 0.5) = 0.$$

所以

$$\lambda_1 = 0.7, \qquad \lambda_2 = 0.5.$$

通解为

$$x_t = C_1(0.7)^t + C_2(0.5)^t.$$

两个根都在单位圆内，因此 $x_t \to 0$。若 $x_t$ 表示对稳态的偏离，则稳态局部稳定。

## 11 系统差分方程与矩阵形式

很多经济模型有多个状态变量。例如资本和消费、通胀和产出缺口、多个资产收益因子。此时写成向量形式更自然:

$$\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t + \boldsymbol{b}.$$

稳态满足

$$\boldsymbol{x}^* = A\boldsymbol{x}^* + \boldsymbol{b},$$

即

$$(I - A)\boldsymbol{x}^* = \boldsymbol{b}.$$

若 $I - A$ 可逆，则

$$\boxed{\boldsymbol{x}^* = (I - A)^{-1}\boldsymbol{b}.}$$

偏离稳态：

$$\boldsymbol{x}_{t+1} - \boldsymbol{x}^* = A(\boldsymbol{x}_t - \boldsymbol{x}^*).$$

所以

$$\boxed{\boldsymbol{x}_t - \boldsymbol{x}^* = A^t(\boldsymbol{x}_0 - \boldsymbol{x}^*).}$$

**Advanced Core**

**稳定性的矩阵版本**

一维时稳定条件是 $|a| < 1$。多维时，$a$ 被矩阵 $A$ 取代。稳定条件变成：$A$ 的所有特征值 $\lambda_i$ 都满足

$$|\lambda_i| < 1.$$

这就是 Chapter 12 会深入讲的内容。直觉上，系统会沿着不同特征向量方向以不同速度放大或缩小；只要每个方向都缩小，整体就稳定。

## 12 理性预期与前瞻差分方程入口

许多现代宏观金融模型不是简单的

$$x_{t+1} = ax_t + b,$$

而是含有预期项：

$$x_t = \alpha E_t x_{t+1} + z_t.$$

这类方程的逻辑是：今天的价格、消费或通胀取决于人们对未来的预期。

若暂时忽略不确定性，把 $E_t x_{t+1}$ 当作 $x_{t+1}$，得到

$$x_t = \alpha x_{t+1} + z_t.$$

如果 $|\alpha| < 1$，向前解可得

$$x_t = z_t + \alpha z_{t+1} + \alpha^2 z_{t+2} + \cdots,$$

即

$$\boxed{x_t = \sum_{j=0}^{\infty} \alpha^j z_{t+j}.}$$

**直觉盒子**

后顾模型 backward-looking model 说“过去影响现在”。前瞻模型 forward-looking model 说“未来预期影响现在”。这正是资产价格、通胀预期、消费储蓄选择和新凯恩斯宏观模型的核心结构。

## 13 Worked Examples

**Worked Example**

**例 5：一阶方程的完整路径、稳态与经济解释**

设

$$x_{t+1} = 0.6x_t + 4, \qquad x_0 = 2.$$

第一步，求稳态：

$$x^* = 0.6x^* + 4.$$

所以

$$0.4x^* = 4, \qquad x^* = 10.$$

第二步，写偏离形式：

$$x_{t+1} - 10 = 0.6(x_t - 10).$$

第三步，递推：

$$x_t - 10 = 0.6^t(x_0 - 10) = 0.6^t(2 - 10).$$

所以

$$\boxed{x_t = 10 - 8(0.6)^t.}$$

因为 $|0.6| < 1$，路径单调收敛到 10。经济解释：如果 $x_t$ 是某个经济变量的水平，当前水平低于稳态，因此之后逐步上升，但每期偏离只剩 60%。

**Worked Example**

**例 6：震荡收敛**

设

$$x_{t+1} = -0.5x_t + 6.$$

稳态为

$$x^* = \frac{6}{1 - (-0.5)} = 4.$$

偏离形式：

$$x_{t+1} - 4 = -0.5(x_t - 4).$$

所以

$$x_t - 4 = (-0.5)^t (x_0 - 4).$$

由于 $|-0.5| < 1$，路径收敛；由于系数为负，路径会在稳态两侧交替震荡。若 $x_0 > 4$，则 $x_1 < 4$，$x_2 > 4$，但偏离幅度每期减半。

**Worked Example**

**例 7：AR(1) 冲击的半衰期**

设产出缺口满足

$$y_t = 0.9 y_{t-1} + \varepsilon_t.$$

一次冲击的影响 $h$ 期后变成一半，满足

$$0.9^h = 0.5.$$

取对数：

$$h \log 0.9 = \log 0.5.$$

所以

$$h = \frac{\log 0.5}{\log 0.9} \approx 6.58.$$

解释：一次冲击大约需要 6 到 7 期才衰减一半。这说明 $\rho = 0.9$ 虽然小于 1，但持久性很强。

**Worked Example**

**例 8：资产价格的基本面价值**

设资产每期支付常数股利 $d = 2$，无风险毛收益率 $R = 1.05$。若排除泡沫，则价格为

$$p = \sum_{j=1}^{\infty} \frac{2}{1.05^j}.$$

这是等比级数：

$$p = 2 \cdot \frac{1/1.05}{1 - 1/1.05} = \frac{2}{1.05 - 1} = 40.$$

如果出现泡沫项 $B_t$，则总价格可以写成

$$p_t = 40 + B_t, \qquad B_{t+1} = 1.05 B_t.$$

排除泡沫等价于要求

$$\lim_{T \to \infty} \frac{B_{t+T}}{1.05^T} = 0.$$

若 $B_t = 1.05^t B_0$ 且 $B_0 \neq 0$，这个极限不为 0，因此被横截条件排除。

## 14 常见错误与辨析

**常见误区**

**误区 1：把稳态和稳定混为一谈。**稳态只是满足 $x_{t+1}=x_t$ 的点；稳定性问的是从附近出发是否会回到这个点。有稳态不代表稳定。

**误区 2：只看 $a<1$，忘记绝对值。**稳定条件是 $|a|<1$，不是 $a<1$。例如 $a=-2<1$，但路径震荡发散。

**误区 3：忘记初始条件。**差分方程的通解通常含有常数，必须用 $x_0$ 或其他初始条件确定路径。

**误区 4：前瞻方程方向搞反。**资产价格递推不能随便给 $p_0$ 向前推，否则可能保留泡沫项。要用横截条件或终端条件选择经济上合理的路径。

**误区 5：看到 $\rho$ 接近 1 就说不稳定。**只要 $|\rho|<1$，AR(1) 是稳定的；但 $\rho$ 接近 1 表示高度持久，有限样本中会看起来像随机游走。

## 15 本章小结：概念 - 公式 - 应用

| 概念 | 核心公式 | 经济金融含义 |
|---|---|---|
| 一阶线性差分 | $x_{t+1}=ax_t+b$ | 递推动态的基本模型。 |
| 稳态 | $x^*=b/(1-a)$ | 长期不再变化的状态。 |
| 偏离形式 | $x_t-x^*=a^t(x_0-x^*)$ | 稳定性由 $a^t$ 决定。 |
| 稳定条件 | $\|a\|<1$ | 偏离逐渐消失。 |
| 蛛网模型 | $p_t=A-Bp_{t-1}$ | 生产时滞导致价格震荡。 |
| AR(1) | $x_t=\rho x_{t-1}+\varepsilon_t$ | 持久性、均值回复、时间序列基础。 |
| 二阶差分 | $x_{t+2}+a_1x_{t+1}+a_0x_t=0$ | 特征根决定动态形状。 |
| 系统差分 | $\boldsymbol{x}_{t+1}=A\boldsymbol{x}_t+\boldsymbol{b}$ | 多变量动态系统。 |
| 资产价格 | $p_t=(d_{t+1}+p_{t+1})/R$ | 基本面价值与泡沫项。 |

## 16 练习题

### A 组：基础计算

A1. 求解 $x_{t+1}=0.5x_t+3$，$x_0=0$。求稳态并判断稳定性。

A2. 求解 $x_{t+1}=-0.8x_t+2$，$x_0=5$。判断路径是否震荡。

A3. 对 $x_{t+1} = 1.1x_t - 1$，求稳态并判断稳定性。

A4. 若 $x_{t+1} = ax_t + b$ 的稳态是 12，且 $a = 0.75$，求 $b$。

A5. 解二阶差分方程 $x_{t+2} - 0.9x_{t+1} + 0.2x_t = 0$。

A6. 对 AR(1) $y_t = 0.7y_{t-1} + \varepsilon_t$，写出 $y_t$ 关于 $y_0$ 和冲击序列的表达式。

### B 组：概念判断

B1. “只要 $a < 1$，$x_{t+1} = ax_t + b$ 就稳定。”这句话对吗？解释。

B2. 若 $a = -0.3$，路径为什么会震荡但仍然收敛?

B3. 稳态存在是否意味着路径一定会走向稳态？举例说明。

B4. 为什么前瞻资产价格方程需要横截条件?

B5. AR(1) 中 $\rho = 0.98$ 和 $\rho = 0.3$ 都稳定，它们的经济含义有什么不同?

### C 组：混合题

C1. 政府债务满足 $b_{t+1} = 1.02b_t - 0.01$。求稳态并判断稳定性。

C2. 蛛网模型中 $D(p) = 20 - 2p$，$S(p_{t-1}) = 4 + p_{t-1}$。写出价格递推、稳态和稳定条件。

C3. 离散 Solow 模型为 $k_{t+1} = 0.3k_t^{1/2} + 0.6k_t$。求正稳态，并用局部线性化判断稳定性。

C4. 若资产价格满足 $p_t = (1 + p_{t+1})/1.1$，求无泡沫基本面价格。

C5. 对 $x_{t+2} - 1.6x_{t+1} + 0.64x_t = 0$，求通解并说明是否稳定。

### D 组：经济金融建模题

D1. 设一个行业价格调整满足 $p_{t+1} = p_t + \lambda(D(p_t) - S(p_t))$，其中 $D(p) = 10 - p$，$S(p) = 2 + p$。写成一阶差分方程，并求稳定条件对 $\lambda$ 的限制。

D2. 某资产股利预计从下一期开始以增长率 $g$ 增长：$d_{t+j} = d_{t+1}(1+g)^{j-1}$，折现毛收益率为 $R$。在 $R > 1 + g$ 下求基本面价格。

D3. 产出缺口满足 $y_t = \rho y_{t-1} + u_t$。解释 $\rho$ 在财政政策冲击传播中的含义，并写出一次冲击 $u_0 = 1$ 对 $y_h$ 的影响。

D4. 离散债务动态 $b_{t+1} = \frac{1+r}{1+g}b_t + d$ 中，如果 $r > g$，政府怎样才能让债务比率不爆炸？用方程解释。

### E 组：挑战 / 证明题

E1. 证明：若 $|a| < 1$，则 $x_{t+1} = ax_t + b$ 对任意初始值都收敛到 $b/(1-a)$。

E2. 对非线性映射 $x_{t+1} = F(x_t)$，设 $F(x^*) = x^*$ 且 $|F'(x^*)| < 1$。说明为什么在稳态附近局部稳定。

E3. 设 $x_t = \rho x_{t-1} + \varepsilon_t$, $E\varepsilon_t = 0$, $\mathrm{Var}(\varepsilon_t) = \sigma^2$, 且 $|\rho| < 1$。推导平稳方差 $\mathrm{Var}(x_t) = \sigma^2/(1-\rho^2)$。

E4. 考虑前瞻方程 $x_t = \alpha x_{t+1} + z$，其中 $|\alpha| < 1$ 且 $z$ 常数。求满足无泡沫条件的解，并说明为什么还可能存在泡沫路径。

# 17 练习题参考答案

## A 组答案

A1. 稳态 $x^* = 3/(1-0.5) = 6$。偏离形式 $x_t - 6 = 0.5^t(x_0 - 6)$。因 $x_0 = 0$，

$$x_t = 6 - 6(0.5)^t.$$

$|0.5| < 1$，稳定。

A2. 稳态 $x^* = 2/(1+0.8) = 10/9$。解为

$$x_t - \frac{10}{9} = (-0.8)^t \left(5 - \frac{10}{9}\right).$$

因 $|-0.8| < 1$，收敛；因系数为负，震荡收敛。

A3. 稳态满足 $x^* = 1.1x^* - 1$，故 $-0.1x^* = -1$，$x^* = 10$。因 $|1.1| > 1$，不稳定。

A4. 稳态公式 $x^* = b/(1-a)$。因此 $12 = b/(1-0.75)$，所以 $b = 3$。

A5. 特征方程

$$\lambda^2 - 0.9\lambda + 0.2 = 0 = (\lambda - 0.5)(\lambda - 0.4).$$

通解

$$x_t = C_1(0.5)^t + C_2(0.4)^t.$$

A6. 递推展开：

$$y_t = 0.7^t y_0 + \sum_{j=0}^{t-1} 0.7^j \varepsilon_{t-j}.$$

## B 组答案

B1. 不对。稳定条件是 $|a| < 1$。例如 $a = -2 < 1$，但 $|-2| > 1$，路径震荡发散。

B2. 因为偏离稳态满足 $x_t - x^* = (-0.3)^t(x_0 - x^*)$。负号使偏离方向每期反转，$0.3^t \to 0$ 使幅度缩小。

B3. 不意味着。例如 $x_{t+1} = 2x_t - 1$ 有稳态 $x^* = 1$，但若 $x_0 \neq 1$，偏离满足 $x_t - 1 = 2^t(x_0 - 1)$，会发散。

B4. 因为前瞻方程通解中可能含有自我实现的泡沫项。横截条件用于排除现值不收敛或经济上不可接受的爆炸路径，从而选出基本面解。

B5. 二者都满足 $|\rho| < 1$，因此都稳定。但 $\rho = 0.98$ 的冲击非常持久，衰减很慢；$\rho = 0.3$ 的冲击很快消退。

## C 组答案

C1. 稳态 $b^* = -0.01/(1-1.02) = 0.5$。由于 $|1.02| > 1$，稳态不稳定。只有初始值恰好等于 0.5 才保持不变。

C2. 出清条件：

$$20 - 2p_t = 4 + p_{t-1}.$$

所以

$$p_t = 8 - 0.5p_{t-1}.$$

稳态 $p^* = 8/(1+0.5) = 16/3$。因 $|-0.5| < 1$，震荡收敛。

C3. 稳态满足

$$k = 0.3k^{1/2} + 0.6k.$$

正稳态中 $0.4k = 0.3k^{1/2}$，故 $k^{1/2} = 0.75$,

$$k^* = 0.5625.$$

映射 $G(k) = 0.3k^{1/2} + 0.6k$,

$$G'(k) = 0.15k^{-1/2} + 0.6.$$

在 $k^* = 0.5625$ 处，$\sqrt{k^*} = 0.75$，所以

$$G'(k^*) = 0.15/0.75 + 0.6 = 0.8.$$

因 $|0.8| < 1$，局部稳定。

C4. 无泡沫时常数价格 $p$ 满足

$$p = \frac{1+p}{1.1}.$$

所以 $1.1p = 1 + p$，$p = 10$。

C5. 特征方程

$$\lambda^2 - 1.6\lambda + 0.64 = 0 = (\lambda - 0.8)^2.$$

重根 $\lambda = 0.8$，通解为

$$x_t = (C_1 + C_2 t)(0.8)^t.$$

因 $t(0.8)^t \to 0$，稳定。

## D 组答案

D1. $D(p) - S(p) = (10-p) - (2+p) = 8 - 2p$。因此

$$p_{t+1} = p_t + \lambda(8 - 2p_t) = (1 - 2\lambda)p_t + 8\lambda.$$

稳定条件：

$$|1-2\lambda|<1.$$

即

$$-1<1-2\lambda<1 \quad\Longrightarrow\quad 0<\lambda<1.$$

D2. 基本面价格

$$p_t=\sum_{j=1}^{\infty}\frac{d_{t+j}}{R^j}=\sum_{j=1}^{\infty}\frac{d_{t+1}(1+g)^{j-1}}{R^j}.$$

提出第一项：

$$p_t=\frac{d_{t+1}}{R}\sum_{j=0}^{\infty}\left(\frac{1+g}{R}\right)^j.$$

当 $R>1+g$ 时，级数收敛：

$$p_t=\frac{d_{t+1}}{R}\cdot\frac{1}{1-(1+g)/R}=\frac{d_{t+1}}{R-1-g}.$$

若 $R=1+r$，则 $p_t=d_{t+1}/(r-g)$。

D3. 一次冲击 $u_0=1$ 对 $y_h$ 的影响为 $\rho^h$。$\rho$ 越大，冲击越持久；$\rho$ 越接近 0，政策影响越快消失。若 $\rho$ 为负，影响会正负交替。

D4. 令 $a=(1+r)/(1+g)$。若 $r>g$，则 $a>1$。债务动态为

$$b_{t+1}=ab_t+d.$$

若 $d$ 是常数赤字，稳态 $b^*=d/(1-a)$。因为 $a>1$，稳态不稳定。要避免爆炸，需要足够的初级盈余，即令 $d<0$，并且财政政策要随债务上升而提高盈余，例如 $d_t=d_0-\phi b_t$，使有效系数 $a-\phi$ 满足 $|a-\phi|<1$。

## E 组答案

E1. 稳态 $x^*=b/(1-a)$。偏离形式为

$$x_{t+1}-x^*=a(x_t-x^*).$$

递推得

$$x_t-x^*=a^t(x_0-x^*).$$

因 $|a|<1$，有 $a^t\to 0$。所以

$$\lim_{t\to\infty}(x_t-x^*)=0,$$

即 $x_t\to x^*$。

E2. 因 $F$ 在 $x^*$ 附近可微，Taylor 展开给出

$$F(x_t) - F(x^*) = F'(x^*)(x_t - x^*) + o(|x_t - x^*|).$$

又 $F(x^*) = x^*$，所以

$$x_{t+1} - x^* = F'(x^*)(x_t - x^*) + o(|x_t - x^*|).$$

在足够小邻域内，如果 $|F'(x^*)| < 1$，映射会压缩偏离，故局部稳定。严格证明可用压缩映射思想。

E3. 若过程平稳，令 $\mathrm{Var}(x_t) = \mathrm{Var}(x_{t-1}) = \sigma_x^2$。由于

$$x_t = \rho x_{t-1} + \varepsilon_t,$$

且 $\varepsilon_t$ 与过去不相关，

$$\mathrm{Var}(x_t) = \rho^2 \mathrm{Var}(x_{t-1}) + \mathrm{Var}(\varepsilon_t).$$

即

$$\sigma_x^2 = \rho^2 \sigma_x^2 + \sigma^2.$$

所以

$$\boxed{\sigma_x^2 = \frac{\sigma^2}{1 - \rho^2}.}$$

E4. 若 $z$ 为常数，猜测无泡沫常数解 $x^*$：

$$x^* = \alpha x^* + z,$$

得

$$x^* = \frac{z}{1 - \alpha}.$$

但齐次方程满足

$$x_t^H = \alpha x_{t+1}^H,$$

即

$$x_{t+1}^H = \frac{1}{\alpha} x_t^H.$$

因此

$$x_t^H = \left(\frac{1}{\alpha}\right)^t B_0.$$

当 $|\alpha| < 1$ 时，若 $B_0 \neq 0$，泡沫项爆炸。无泡沫条件排除它，留下

$$x_t = \frac{z}{1 - \alpha}.$$

## 18 后续学习连接

本章之后，Chapter 12 会把一维差分方程推广到二维和多维线性系统：

$$\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t.$$

这会引入特征值、特征向量、鞍路径 saddle path 和相图。你现在需要记住的是：一维的 $a$，在多维里变成矩阵 $A$ 的特征值；一维的 $|a| < 1$，在多维里变成所有特征值都在单位圆内。

**直觉盒子**

Chapter 10 的连续时间稳定条件常看特征值实部是否为负；Chapter 11 的离散时间稳定条件看特征值模是否小于 1。一个是“左半平面”，一个是“单位圆内”。这是宏观动态系统中最重要的对应关系之一。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 12
# 线性系统、特征值与相图

Chapter 10 用一阶 ODE 讲连续时间动态，Chapter 11 用差分方程讲离散时间递推。现实经济模型通常不只有一个变量：资本和消费一起动，通胀和产出缺口一起动，两个风险因子共同驱动资产收益，多个状态变量共同决定下一期价格。本章的核心目标是把单变量动态升级为系统动态：把模型写成 $\dot{x} = Ax$ 或 $x_{t+1} = Ax_t$，用特征值判断稳定性，用特征向量理解运动方向，用相图看经济路径。读完这一章，你应当不只是会算 eigenvalues，而是能看懂为什么 Ramsey 模型有 saddle path，为什么有些宏观动态需要跳变量一次性跳到稳定臂上，为什么金融中的多因子状态变量要看矩阵谱半径。

math4econ personalized notes
2026

# Contents

**Chapter 12 总览** **2**

**1 本章为什么重要** **2**
1.1 从一个变量到多个变量 . . . 2
1.2 为什么特征值会反复出现 . . . 3
1.3 本章服务哪些后续课程 . . . 3

**2 最小前置知识** **3**

**3 从标量动态到向量系统** **4**
3.1 连续时间线性系统 . . . 4
3.2 离散时间线性系统 . . . 4

**4 特征值与特征向量：动态系统的自然方向** **5**
4.1 特征向量为什么是“自然方向” . . . 5
4.2 二维矩阵的 trace 与 determinant . . . 5

**5 连续时间二维系统：分类与解法** **6**
5.1 基本解法：可对角化情形 . . . 6
5.2 稳定节点 Stable Node . . . 7
5.3 不稳定节点 Unstable Node . . . 7
5.4 鞍点 Saddle . . . 7
5.5 焦点与螺旋 Focus / Spiral . . . 8
5.6 二维连续时间系统的稳定性表 . . . 10

**6 离散时间系统：单位圆而不是左半平面** **10**

**7 相图 Phase Diagram 的读法** **11**
7.1 相图不是画图装饰，而是模型求解工具 . . . 11
7.2 相图中的四种信息 . . . 11

**8 Saddle Path：高级宏观里的核心图像** **12**
8.1 predetermined variable 与 jump variable . . . 12
8.2 为什么 saddle path 不是“玄学” . . . 12

**9 非线性系统的局部线性化** **13**
9.1 从 $\dot{x} = F(x)$ 到 $\dot{y} = Jy$ . . . 13
9.2 Jacobian 的经济含义 . . . 13

**10 经济应用一：Ramsey 模型的线性化直觉** **13**

**11 经济应用二：通胀与产出缺口的线性反馈** **14**

**12 金融应用：状态变量系统与风险因子** **15**

**13 Worked Examples** **15**

**14 常见错误与辨析** **17**

**15 本章小结：概念 - 公式 - 应用** **17**

**16 练习题** **18**

A 组：基础计算 . . . . . 18

B 组：概念判断 . . . . . 18

C 组：混合题 . . . . . 18

D 组：经济金融建模题 . . . . . 19

E 组：挑战 / 证明题 . . . . . 19

**17 练习题参考答案** **19**

A 组答案 . . . . . 19

B 组答案 . . . . . 20

C 组答案 . . . . . 20

D 组答案 . . . . . 21

E 组答案 . . . . . 21

**18 后续学习连接** **22**

# Chapter 12 总览

本章是动态系统部分的核心章节。Chapter 10 和 Chapter 11 已经分别给了你连续时间与离散时间的一维版本：

$$\dot{x} = f(x), \qquad x_{t+1} = g(x_t).$$

但经济金融模型里，真正有解释力的对象往往是向量：

$$x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix}.$$

例如宏观里 $x_1$ 可以是资本偏离稳态的程度，$x_2$ 可以是消费偏离稳态的程度；金融里 $x_1$ 可以是短端利率，$x_2$ 可以是风险溢价状态变量；计量里 $x_t$ 可以是一个 VAR 系统中的多个时间序列。于是模型就从一个方程变成一个系统：

$$\dot{x} = Ax, \qquad \text{or} \qquad x_{t+1} = Ax_t.$$

这里的矩阵 $A$ 不是普通系数表，而是局部动态结构。它告诉我们每个变量怎样影响每个变量的变化率或下一期取值。

**直觉框**

一维动态中，一个数 $a$ 决定稳定性；多维动态中，一个矩阵 $A$ 决定稳定性。矩阵本身看起来复杂，但它最重要的信息被压缩在特征值 eigenvalues 和特征向量 eigenvectors 里。特征值决定某个方向上是收敛、发散还是振荡；特征向量决定这些特殊方向在哪里。

本章结构如下：先把经济模型写成向量系统，再复习特征值的含义，然后分别讲连续时间和离散时间的稳定性判别，接着学习相图和 saddle path，最后把这些工具用到 Ramsey 模型、通胀-产出动态和金融状态变量系统中。

## 1 本章为什么重要

### 1.1 从一个变量到多个变量

一阶 ODE 讲的是“一个变量自己怎么动”。例如：

$$\dot{k} = sf(k) - (n+\delta)k.$$

这已经能描述 Solow 模型中资本的积累。但一旦进入更高级的经济理论，很多变量必须同时出现。

在 Ramsey 模型里，资本 $k$ 和消费 $c$ 共同演化：

$$\begin{cases} \dot{k} = f(k) - c - (n+\delta)k, \\ \dfrac{\dot{c}}{c} = \dfrac{f'(k) - \delta - \rho}{\theta}. \end{cases}$$

在 New Keynesian 风格的简化宏观动态里，通胀 $\pi$ 与产出缺口 $y$ 可能共同演化。在线性因子金融模型里，收益率、利率、风险溢价常常由一个状态向量驱动：

$$z_{t+1} = Az_t + \varepsilon_{t+1}.$$

这些都不是单变量动态，而是系统动态。

### 1.2 为什么特征值会反复出现

如果一个系统是线性的，或者在稳态附近可以近似为线性的，那么局部动态就由矩阵 $A$ 决定：

$$\dot{x} = Ax.$$

这个系统一般有很多变量交叉影响：

$$\dot{x}_1 = a_{11}x_1 + a_{12}x_2, \qquad \dot{x}_2 = a_{21}x_1 + a_{22}x_2.$$

直接盯着 $a_{ij}$ 很难看出路径。特征值方法做的事情是寻找“系统自己的自然方向”。如果某个方向 $v$ 满足

$$Av = \lambda v,$$

那么沿着这个方向运动时，矩阵的作用只是把向量放大、缩小或反向，而不改变方向。动态系统就可以拆解为若干个这样的方向。

**经济金融解释**

经济学中的“稳定均衡”不是一句直觉判断，而是特征值条件。连续时间里，特征值实部为负表示该方向收敛；离散时间里，特征值模长小于 1 表示该方向收敛。Ramsey 模型的 saddle path，本质上就是二维系统中一个稳定方向和一个不稳定方向同时存在。

### 1.3 本章服务哪些后续课程

本章会直接服务：

- 高级宏观：Ramsey 模型、RBC、New Keynesian 动态、saddle-path stability；
- 金融经济学：状态变量系统、均值回复、期限结构模型、动态资产定价；
- 计量经济学：VAR、状态空间模型、ARMA 的矩阵表达；
- 动态优化：Bellman 方程附近的局部线性化；
- 非线性动态：稳态附近的 Jacobian 与局部稳定性。

## 2 最小前置知识

本章需要你回忆四件事。

第一，矩阵乘法。若

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

则

$$Ax = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix}.$$

第二，特征值与特征向量。非零向量 $v$ 若满足

$$Av = \lambda v,$$

则 $\lambda$ 是特征值，$v$ 是对应的特征向量。求法是

$$\det(A - \lambda I) = 0.$$

第三，一维 ODE 与差分方程。连续时间中

$$\dot{x} = ax \quad \Longrightarrow \quad x(t) = x(0)e^{at}.$$

离散时间中

$$x_{t+1} = ax_t \quad \Longrightarrow \quad x_t = a^t x_0.$$

第四，局部线性化。对非线性系统 $\dot{x} = F(x)$，在稳态 $x^*$ 附近有

$$\dot{y} \approx J_F(x^*)y, \qquad y = x - x^*,$$

其中 $J_F(x^*)$ 是 Jacobian 矩阵。也就是说，非线性系统在稳态附近的第一近似就是线性系统。

# 3 从标量动态到向量系统

## 3.1 连续时间线性系统

连续时间线性系统的一般形式是

$$\dot{x}(t) = Ax(t) + b,$$

其中 $x(t) \in \mathbb{R}^n$，$A$ 是 $n \times n$ 矩阵，$b$ 是常数向量。若 $b = 0$，系统是齐次系统：

$$\dot{x}(t) = Ax(t).$$

若 $b \neq 0$，稳态 $x^*$ 满足

$$0 = Ax^* + b.$$

若 $A$ 可逆，则

$$x^* = -A^{-1}b.$$

令 $y = x - x^*$，则

$$\dot{y} = \dot{x} = A(x^* + y) + b = Ax^* + b + Ay = Ay.$$

所以非齐次线性系统可以通过平移稳态变成齐次系统。

**方法框**

连续时间线性系统的标准处理流程：

1. 先找稳态 $x^*$：解 $Ax^* + b = 0$；
2. 平移变量 $y = x - x^*$；
3. 得到 $\dot{y} = Ay$；
4. 求 $A$ 的特征值与特征向量；
5. 根据特征值判断局部稳定性，根据特征向量画相图。

## 3.2 离散时间线性系统

离散时间线性系统的一般形式是

$$x_{t+1} = Ax_t + b.$$

稳态满足

$$x^* = Ax^* + b, \qquad (I - A)x^* = b.$$

如果 $I-A$ 可逆，则

$$x^* = (I-A)^{-1}b.$$

令 $y_t = x_t - x^*$，则

$$y_{t+1} = Ay_t.$$

所以离散时间里也一样：先找稳态，再平移到齐次系统。

连续时间与离散时间的最大差别在稳定性标准：

$$\text{连续时间：}\quad \mathrm{Re}(\lambda_i) < 0; \qquad \text{离散时间：}\quad |\lambda_i| < 1.$$

前者看实部是否为负，后者看模长是否小于 1。

# 4 特征值与特征向量：动态系统的自然方向

## 4.1 特征向量为什么是“自然方向”

考虑二维线性系统

$$\dot{x} = Ax.$$

如果初始点刚好在某个特征向量方向上，即 $x(0) = cv$，且 $Av = \lambda v$，那么我们猜解

$$x(t) = z(t)v.$$

代入系统：

$$\dot{z}(t)v = A(z(t)v) = z(t)Av = z(t)\lambda v.$$

于是

$$\dot{z}(t) = \lambda z(t),$$

所以

$$z(t) = z(0)e^{\lambda t}.$$

因此

$$x(t) = ce^{\lambda t}v.$$

这说明沿着特征向量方向时，系统退化成一维指数运动。$\lambda < 0$ 时该方向收敛，$\lambda > 0$ 时该方向发散。

**直觉框**

特征向量是系统不会“拐弯”的方向。特征值是这个方向上的增长率。二维相图里，稳定臂、发散方向、saddle path，本质上都和特征向量有关。

## 4.2 二维矩阵的 trace 与 determinant

对二维矩阵

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

特征多项式是

$$\det(A - \lambda I) = 0.$$

计算得

$$\lambda^2 - \mathrm{tr}(A)\lambda + \det(A) = 0,$$

其中

$$\mathrm{tr}(A) = a + d, \qquad \det(A) = ad - bc.$$

所以二维系统特征值满足

$$\lambda_1 + \lambda_2 = \mathrm{tr}(A), \qquad \lambda_1 \lambda_2 = \det(A).$$

这两个量非常有用，因为很多时候不需要完整求出两个特征值，就能判断大致稳定性。

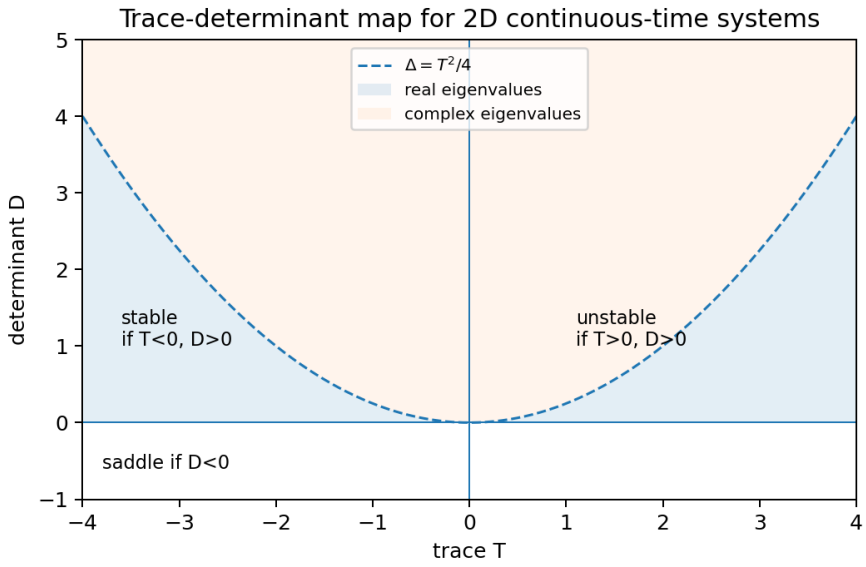



Figure 1: 二维连续时间系统的 trace-determinant 地图。横轴是 $T = \mathrm{tr}(A)$，纵轴是 $D = \det(A)$。$D < 0$ 表示 saddle；$D > 0, T < 0$ 是稳定区域；$D > 0, T > 0$ 是不稳定区域。

# 5 连续时间二维系统：分类与解法

## 5.1 基本解法：可对角化情形

设 $A$ 有两个线性无关的特征向量 $v_1, v_2$，对应特征值 $\lambda_1, \lambda_2$。那么通解为

$$x(t) = c_1 e^{\lambda_1 t} v_1 + c_2 e^{\lambda_2 t} v_2.$$

这和一维解法的关系很直接：每个特征方向上都有一个一维指数路径，整体路径是这些方向的线性组合。

如果写成矩阵形式，令

$$P = \begin{pmatrix} v_1 & v_2 \end{pmatrix}, \qquad D = \mathrm{diag}(\lambda_1, \lambda_2),$$

则

$$A = PDP^{-1},$$

并且

$$e^{At} = Pe^{Dt}P^{-1} = P \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix} P^{-1}.$$

所以

$$x(t) = e^{At} x(0).$$

### 5.2 稳定节点 Stable Node

如果两个特征值都是负实数：

$$\lambda_1 < 0, \qquad \lambda_2 < 0,$$

则任意方向都收敛到稳态。若一个特征值更接近 $0$，它对应的方向收敛更慢，长期路径会逐渐贴近慢方向。

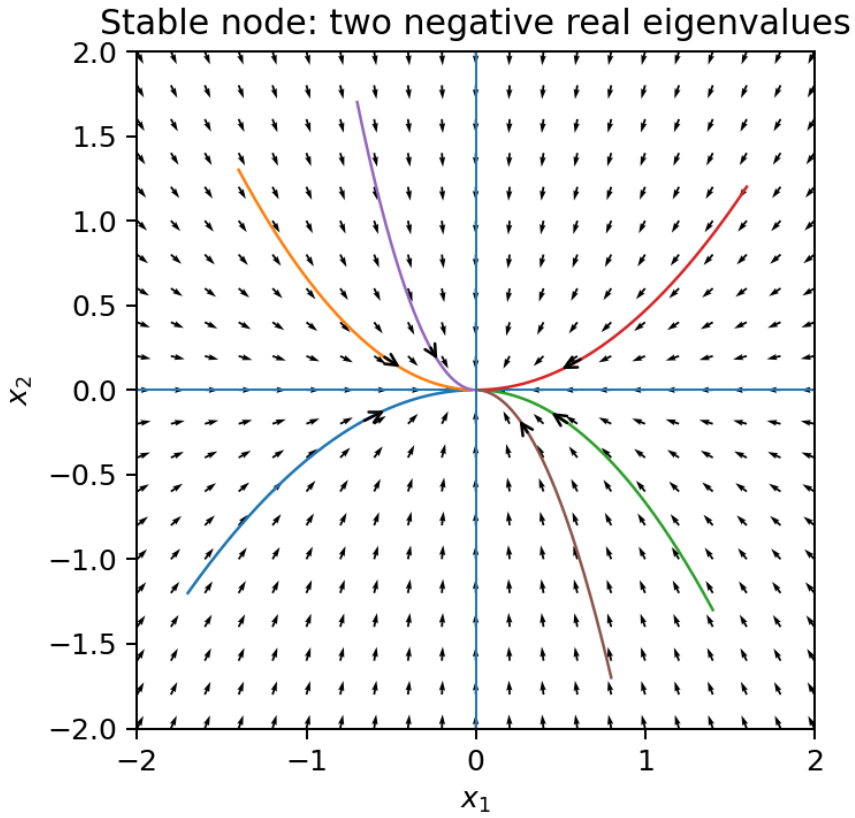



Figure 2: 稳定节点：两个负实特征值使所有路径收敛到稳态。越接近 $0$ 的特征值对应的方向收敛越慢。

经济含义：如果资本和价格都具有均值回复力量，并且没有不稳定的反馈回路，则系统可能表现为稳定节点。初始冲击会逐渐消失，但不同方向的调整速度不同。

### 5.3 不稳定节点 Unstable Node

如果两个特征值都是正实数：

$$\lambda_1 > 0, \qquad \lambda_2 > 0,$$

则除非初始点正好在稳态，否则路径会远离稳态。这在经济模型中通常代表强正反馈：偏离越大，增长率越让偏离进一步放大。

### 5.4 鞍点 Saddle

如果两个特征值一正一负：

$$\lambda_1 < 0 < \lambda_2,$$

则一个方向收敛，另一个方向发散。这样的稳态叫鞍点。鞍点不是“完全不稳定”，而是“只有一条或一个低维子空间稳定”。

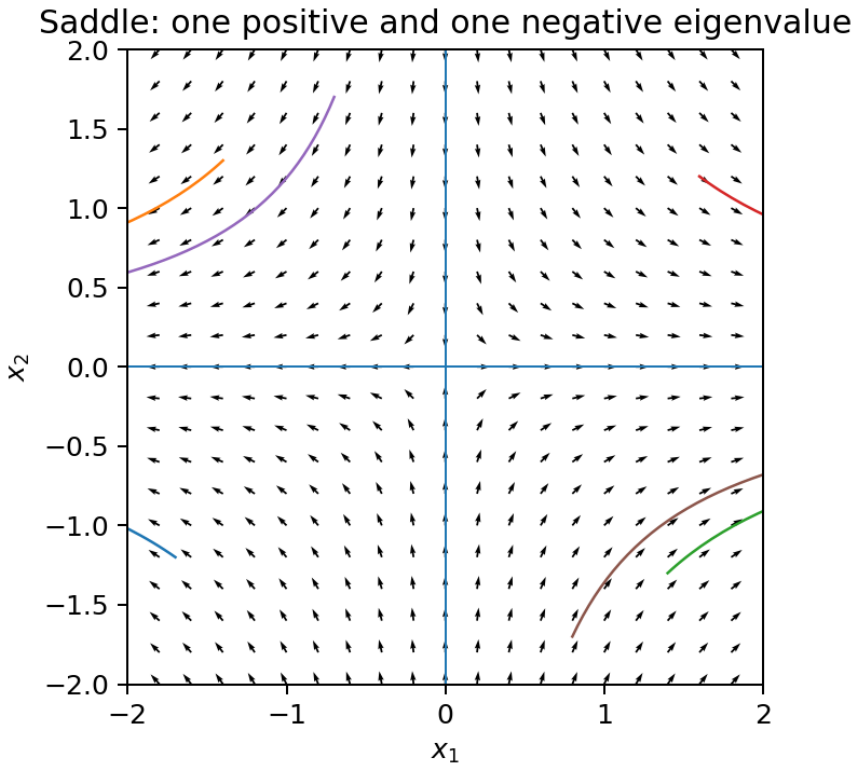



Figure 3: 鞍点：一个方向收敛，一个方向发散。经济学中的 saddle path stability 正是这个结构。

**经济金融解释**

Ramsey 模型中的 saddle path 就是鞍点结构。资本 $k$ 通常是 predetermined variable，不能瞬间跳；消费 $c$ 是 jump variable，可以在初始时刻调整。为了不让路径沿不稳定方向爆炸，消费必须选择到稳定臂上。这就是宏观里“唯一理性路径”的几何来源。

### 5.5 焦点与螺旋 Focus / Spiral

如果特征值是一对复数

$$\lambda_{1,2} = \alpha \pm i\beta,$$

则路径会旋转。$\alpha$ 决定半径变化，$\beta$ 决定旋转频率。

- $\alpha < 0$：稳定焦点，路径螺旋收敛；
- $\alpha > 0$：不稳定焦点，路径螺旋发散；
- $\alpha = 0$：中心 center，路径闭合或近似周期运动。

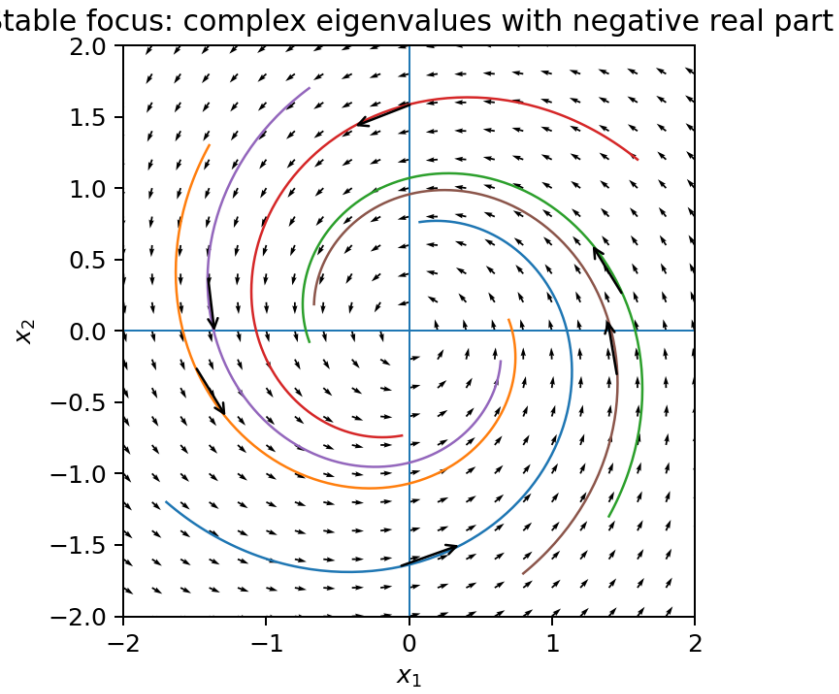



Figure 4: 稳定焦点：特征值有负实部和非零虚部，路径以螺旋方式收敛。

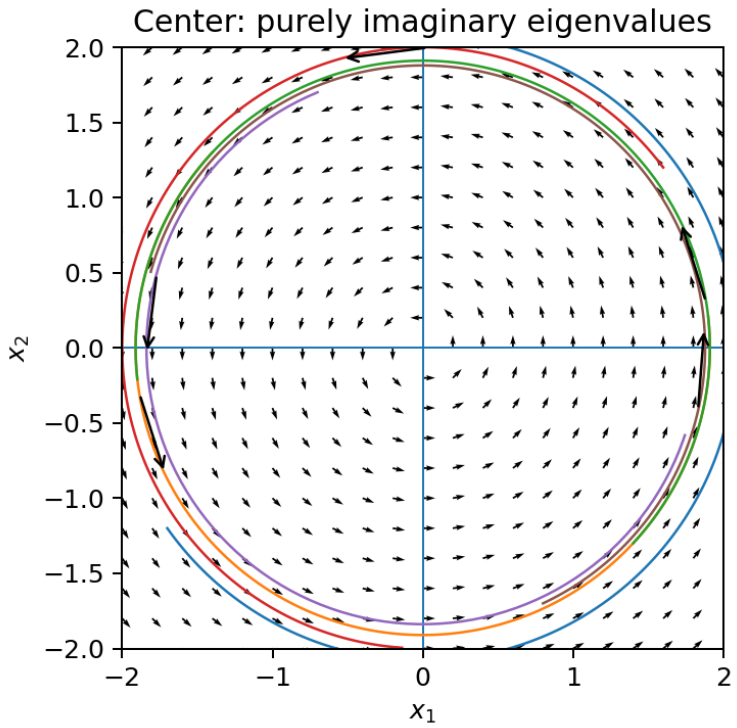



Figure 5: 中心：纯虚特征值使路径围绕稳态旋转。真实经济模型中，纯中心通常很脆弱，小扰动会变成稳定或不稳定焦点。

### 5.6 二维连续时间系统的稳定性表

| 特征值类型 | 条件 | 名称 | 路径形状 |
|---|---|---|---|
| 两个负实数 | $\lambda_1, \lambda_2 < 0$ | stable node | 所有路径收敛，无旋转 |
| 两个正实数 | $\lambda_1, \lambda_2 > 0$ | unstable node | 所有非稳态路径发散 |
| 一正一负 | $\lambda_1 \lambda_2 < 0$ | saddle | 一条稳定方向，一条发散方向 |
| 复数，负实部 | $\alpha < 0$ | stable focus | 螺旋收敛 |
| 复数，正实部 | $\alpha > 0$ | unstable focus | 螺旋发散 |
| 纯虚数 | $\alpha = 0$ | center | 周期或闭合运动 |

## 6 离散时间系统：单位圆而不是左半平面

离散时间系统

$$x_{t+1} = Ax_t$$

如果沿特征向量方向 $v$，且 $Av = \lambda v$，则

$$x_t = c\lambda^t v.$$

因此稳定性取决于 $\lambda^t$ 是否趋向 $0$：

$$|\lambda| < 1 \quad \Longleftrightarrow \quad \lambda^t \to 0.$$

这和连续时间的 $e^{\lambda t}$ 不同。

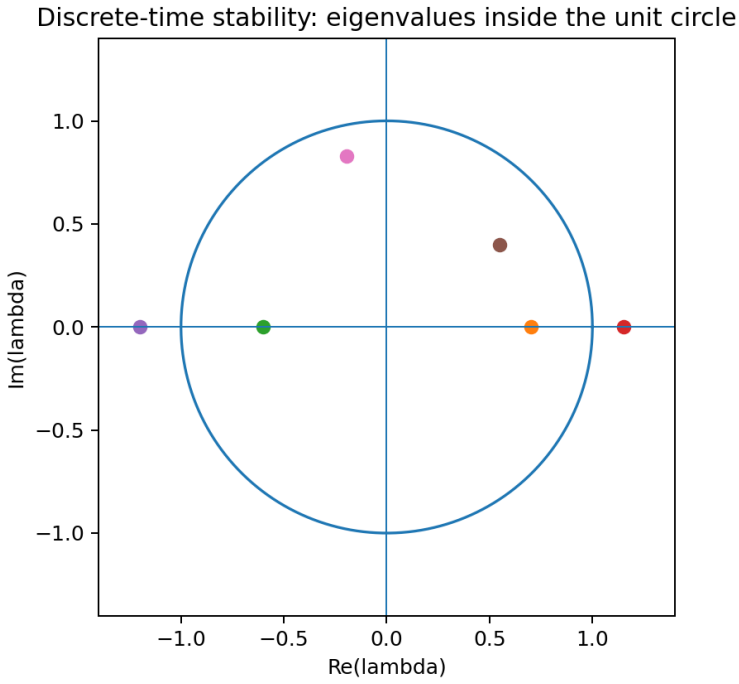



Figure 6: 离散时间稳定性：所有特征值必须位于复平面单位圆内，即 $|\lambda_i| < 1$。

**方法框**

离散时间稳定性判断：

1. 写出 $x_{t+1} = Ax_t + b$；

2. 求稳态 $x^* = (I - A)^{-1}b$；
3. 令 $y_t = x_t - x^*$，得到 $y_{t+1} = Ay_t$；
4. 求 $A$ 的特征值；
5. 若所有 $|\lambda_i| < 1$，稳态局部稳定；若某个 $|\lambda_i| > 1$，存在发散方向。

离散时间中负特征值有一个额外现象：符号交替。若 $\lambda = -0.8$，则 $(-0.8)^t$ 的绝对值下降，但正负交替，因此路径会在稳态两侧来回震荡收敛。若 $\lambda = -1.2$，则震荡发散。

# 7 相图 Phase Diagram 的读法

## 7.1 相图不是画图装饰，而是模型求解工具

对二维连续时间系统

$$\begin{cases} \dot{x} = f(x, y), \\ \dot{y} = g(x, y), \end{cases}$$

相图的核心是两条 nullclines：

$$\dot{x} = 0, \qquad \dot{y} = 0.$$

它们把平面分成若干区域。在每个区域里，$\dot{x}$ 的正负决定横向箭头，$\dot{y}$ 的正负决定纵向箭头。

**方法框**

画二维相图的标准步骤：

1. 写出两个动态方程；
2. 解 $\dot{x} = 0$ 与 $\dot{y} = 0$，画 nullclines；
3. 在各区域判断 $\dot{x}$ 与 $\dot{y}$ 的符号；
4. 找交点，即稳态；
5. 若需要判断局部稳定性，计算稳态处 Jacobian 的特征值；
6. 若是 saddle，标出稳定臂与不稳定臂。

## 7.2 相图中的四种信息

一张好的相图至少应该告诉你四件事。

第一，稳态在哪里。它是 nullclines 的交点。

第二，变量在每个区域的运动方向。例如 $\dot{k} > 0$ 表示资本增加，$\dot{c} < 0$ 表示消费下降。

第三，稳态附近的局部稳定类型。这个不能只靠肉眼，最好用 Jacobian 特征值验证。

第四，经济可行路径在哪里。很多宏观模型中，数学上存在很多路径，但只有满足预算约束、非负性、跨期最优和 transversality condition 的路径才是经济上可接受的。

# 8 Saddle Path：高级宏观里的核心图像

## 8.1 predetermined variable 与 jump variable

在许多动态经济模型中，不同变量有不同的时间性质。

资本 $k$ 是 predetermined variable。今天的资本来自过去投资，不能瞬间跳到任意值。消费 $c$、资产价格 $q$、通胀预期等则可能是 jump variable，可以因为预期变化而立刻调整。

如果系统有一个稳定方向和一个不稳定方向，即 saddle，那么一般初始点都会沿不稳定方向发散。为了让路径满足长期条件，jump variable 必须选择一个恰当初值，使整个系统刚好落在稳定臂上。

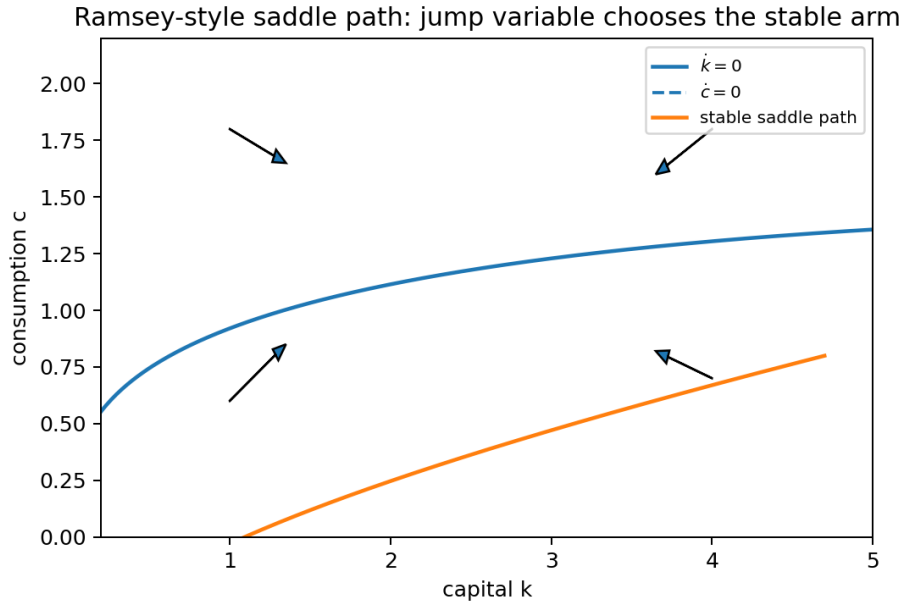



Figure 7: Ramsey 风格 saddle path：资本 $k$ 不能跳，消费 $c$ 可以跳。为了让经济沿稳定路径收敛到稳态，初始消费必须选在 saddle path 上。

## 8.2 为什么 saddle path 不是“玄学”

很多初学宏观的人会觉得 saddle path 很玄：为什么模型能自己选出一条路？其实逻辑是清楚的。

设线性化系统有两个特征值：

$$\lambda_s < 0 < \lambda_u.$$

通解为

$$y(t) = c_s e^{\lambda_s t} v_s + c_u e^{\lambda_u t} v_u.$$

由于 $e^{\lambda_u t} \to \infty$，只要 $c_u \neq 0$，路径最终爆炸。因此若模型要求路径有界、满足 transversality condition 或趋向稳态，就必须有

$$c_u = 0.$$

这相当于把初始状态限制在稳定特征向量张成的方向上。

**直觉框**

Saddle path 的本质：长期理性条件消除了不稳定成分。jump variable 的任务，就是在初始时刻调整到让不稳定成分系数为零的位置。

# 9 非线性系统的局部线性化

## 9.1 从 $\dot{x} = F(x)$ 到 $\dot{y} = Jy$

考虑非线性系统

$$\dot{x} = F(x), \qquad x \in \mathbb{R}^n.$$

稳态 $x^*$ 满足

$$F(x^*) = 0.$$

令

$$y = x - x^*.$$

对 $F$ 在 $x^*$ 附近做一阶 Taylor 展开：

$$F(x^* + y) = F(x^*) + J_F(x^*)y + o(\|y\|).$$

因为 $F(x^*) = 0$，所以

$$\dot{y} = J_F(x^*)y + o(\|y\|).$$

在足够接近稳态时，低阶项 $o(\|y\|)$ 比线性项小，于是局部动态由

$$\dot{y} \approx J_F(x^*)y$$

决定。

## 9.2 Jacobian 的经济含义

若

$$F(x, y) = \begin{pmatrix} f(x, y) \\ g(x, y) \end{pmatrix},$$

则 Jacobian 是

$$J_F(x, y) = \begin{pmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \\ \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{pmatrix}.$$

每个元素都是一个局部边际效应。例如 $\partial f/\partial y$ 表示变量 $y$ 对 $\dot{x}$ 的影响。宏观中这类交叉项很重要：消费变化影响资本积累，资本变化影响消费增长率，通胀影响实际利率，实际利率影响产出缺口。

**常见误区**

不要把 Jacobian 看成“多元求导表格”。在动态系统中，Jacobian 是稳态附近的反馈结构矩阵。它的特征值决定系统会不会回到稳态，它的特征向量决定沿什么方向回去或离开。

# 10 经济应用一：Ramsey 模型的线性化直觉

考虑简化 Ramsey 系统：

$$\begin{cases} \dot{k} = f(k) - c - (n + \delta)k, \\ \dfrac{\dot{c}}{c} = \dfrac{f'(k) - \delta - \rho}{\theta}. \end{cases}$$

定义

$$F_1(k, c) = f(k) - c - (n + \delta)k,$$

$$F_2(k,c) = c\frac{f'(k)-\delta-\rho}{\theta}.$$

稳态满足

$$f(k^*) - c^* - (n+\delta)k^* = 0,$$
$$f'(k^*) = \delta + \rho.$$

Jacobian 为

$$J(k,c) = \begin{pmatrix} f'(k)-(n+\delta) & -1 \\ \dfrac{cf''(k)}{\theta} & \dfrac{f'(k)-\delta-\rho}{\theta} \end{pmatrix}.$$

在稳态处，第二行第二列变成 0，因为 $f'(k^*)-\delta-\rho=0$。于是

$$J(k^*,c^*) = \begin{pmatrix} f'(k^*)-(n+\delta) & -1 \\ \dfrac{c^*f''(k^*)}{\theta} & 0 \end{pmatrix}.$$

在通常假设下，$f''(k^*)<0$。这个矩阵往往有一个正特征值和一个负特征值，因此稳态是 saddle。

**经济金融解释**

Ramsey 模型的 saddle 结构来自两个力量：资本方程中，消费越高，资本积累越低；Euler 方程中，资本越高，边际产出越低，从而消费增长率下降。这两个交叉反馈共同形成一个稳定方向和一个不稳定方向。

## 11 经济应用二：通胀与产出缺口的线性反馈

考虑一个非常简化的二维宏观动态：

$$\begin{cases} \dot{y} = -ay - b\pi, \\ \dot{\pi} = cy - d\pi, \end{cases} \qquad a,b,c,d>0.$$

写成矩阵形式：

$$\begin{pmatrix} \dot{y} \\ \dot{\pi} \end{pmatrix} = \begin{pmatrix} -a & -b \\ c & -d \end{pmatrix} \begin{pmatrix} y \\ \pi \end{pmatrix}.$$

trace 和 determinant 为

$$T = -(a+d) < 0,$$
$$D = ad + bc > 0.$$

因此稳态至少不会是 saddle。是否出现螺旋，取决于判别式

$$T^2 - 4D.$$

若 $T^2-4D<0$，则特征值为复数且实部为负，系统呈现稳定振荡：产出和通胀围绕稳态逐渐收敛。这给了商业周期模型一个最小数学图像：经济变量可能不是单调回稳态，而是振荡回稳态。

## 12 金融应用：状态变量系统与风险因子

在金融中，很多模型把价格、收益率或风险溢价写成状态变量 $z_t$ 的函数：

$$z_{t+1} = Az_t + \varepsilon_{t+1}.$$

若不考虑随机扰动，核心确定性部分为

$$z_{t+1} = Az_t.$$

如果所有特征值都在单位圆内，状态变量是均值回复的；如果有特征值接近 $1$，说明该因子非常持久；如果特征值大于 $1$，模型会产生爆炸路径，通常需要重新设定或加入均衡限制。

例如两个风险因子：

$$\begin{pmatrix} z_{1,t+1} \\ z_{2,t+1} \end{pmatrix} = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.6 \end{pmatrix} \begin{pmatrix} z_{1,t} \\ z_{2,t} \end{pmatrix} + \varepsilon_{t+1}.$$

这个矩阵的特征值决定两个独立动态成分的持久性。若一个特征值接近 $0.95$，另一个接近 $0.55$，那么一个因子消退慢，一个因子消退快。资产价格对长期因子的暴露通常更重要。

## 13 Worked Examples

**例 1：连续时间系统的通解**

求解

$$\dot{x} = Ax, \qquad A = \begin{pmatrix} -1 & 0 \\ 0 & -3 \end{pmatrix}, \qquad x(0) = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

**解：**因为 $A$ 已经是对角矩阵，两个特征值为

$$\lambda_1 = -1, \qquad \lambda_2 = -3.$$

两个特征向量分别是

$$v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

所以

$$x(t) = c_1 e^{-t} v_1 + c_2 e^{-3t} v_2.$$

由初始条件得 $c_1 = 2, c_2 = 1$，因此

$$x(t) = \begin{pmatrix} 2e^{-t} \\ e^{-3t} \end{pmatrix}.$$

经济解释：第二个变量收敛更快，因为 $-3$ 比 $-1$ 更负；长期接近稳态时，路径主要沿第一个变量方向缓慢靠近。

**例 2：判断 saddle**

考虑

$$A = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}.$$

判断 $\dot{x} = Ax$ 的稳定性。

**解：**特征值直接为 $2$ 与 $-1$。一个为正，一个为负，所以稳态是 saddle。通解为

$$x(t) = c_1 e^{2t}\begin{pmatrix}1\\0\end{pmatrix} + c_2 e^{-t}\begin{pmatrix}0\\1\end{pmatrix}.$$

如果 $c_1 \neq 0$，第一项爆炸；只有当 $c_1 = 0$ 时，路径才沿第二个方向收敛。这就是稳定臂的数学形式。

### 例 3：二维系统的 trace-determinant 判别

设

$$A = \begin{pmatrix}-2 & -1\\ 3 & -2\end{pmatrix}.$$

判断稳定类型。

**解：**先算

$$T = \operatorname{tr}(A) = -4,$$

$$D = \det(A) = (-2)(-2) - (-1)(3) = 7.$$

判别式为

$$T^2 - 4D = 16 - 28 = -12 < 0.$$

所以特征值是一对复数。由于 $T < 0$，实部为 $T/2 = -2 < 0$，因此是稳定焦点。路径会以螺旋形式收敛。

### 例 4：离散时间系统稳定性

考虑

$$x_{t+1} = Ax_t, \qquad A = \begin{pmatrix}0.5 & 0\\ 0 & -0.8\end{pmatrix}.$$

**解：**特征值为 $0.5$ 与 $-0.8$。由于

$$|0.5| < 1, \qquad |-0.8| < 1,$$

系统稳定。第一个方向单调收敛，第二个方向震荡收敛，因为负特征值导致符号交替。

### 例 5：非线性系统的局部线性化

考虑

$$\begin{cases}\dot{x} = x(1-x) - y,\\ \dot{y} = 0.5x - y.\end{cases}$$

求稳态并判断原点附近稳定性。

**解：**稳态满足

$$x(1-x) - y = 0, \qquad 0.5x - y = 0.$$

由第二式 $y = 0.5x$，代入第一式：

$$x(1-x) - 0.5x = x(0.5 - x) = 0.$$

所以稳态为

$$(0,0), \qquad (0.5, 0.25).$$

Jacobian 为

$$J(x,y)=\begin{pmatrix}1-2x & -1\\ 0.5 & -1\end{pmatrix}.$$

在原点处：

$$J(0,0)=\begin{pmatrix}1 & -1\\ 0.5 & -1\end{pmatrix}.$$

trace 与 determinant 为

$$T=0,\qquad D=-1+0.5=-0.5<0.$$

因此原点是 saddle。注意这是局部结论：它说明在原点附近存在一个稳定方向和一个不稳定方向，不等于全局路径都能用线性图像描述。

## 14 常见错误与辨析

**常见误区**

**错误一：把连续时间和离散时间稳定条件混用。**连续时间看 $\mathrm{Re}(\lambda)<0$；离散时间看 $|\lambda|<1$。$\lambda=-2$ 在连续时间中稳定，但在离散时间中不稳定，因为 $|-2|>1$。

**常见误区**

**错误二：只算 trace 和 determinant，却忘了判断判别式。** $D>0, T<0$ 说明稳定，但路径可能是 node，也可能是 focus。要区分是否旋转，需要看 $T^2-4D$。

**常见误区**

**错误三：认为 saddle 就是“不稳定，所以没意义”。**在经济学中，saddle 往往最有意义，因为 jump variable 可以选择稳定臂。Ramsey 模型、资产定价模型、理性预期模型里，saddle-path stability 是核心概念。

**常见误区**

**错误四：把相图当作严格证明。**相图帮助理解方向，但真正的局部稳定性判别应回到 Jacobian 和特征值。图像直觉和代数判别要相互校验。

## 15 本章小结：概念 - 公式 - 应用

| 概念 | 核心公式 | 经济金融含义 |
|---|---|---|
| 连续时间系统 | $\dot{x}=Ax+b$ | 多变量连续动态，如资本-消费系统 |
| 离散时间系统 | $x_{t+1}=Ax_t+b$ | VAR、资产价格递推、状态变量模型 |
| 稳态平移 | $y=x-x^*$ | 把非齐次系统转为齐次系统 |
| 特征方向 | $Av=\lambda v$ | 系统不改变方向的自然运动方向 |
| 连续稳定性 | $\mathrm{Re}(\lambda_i)<0$ | 所有方向都随时间衰减 |

| | | |
|---|---|---|
| 离散稳定性 | $\lvert\lambda_i\rvert < 1$ | 所有方向都随迭代衰减 |
| Saddle | 一正一负特征值 | 一条稳定臂，一条发散臂 |
| Jacobian | $J_F(x^*)$ | 非线性系统在稳态附近的反馈矩阵 |
| Ramsey saddle path | $c_u = 0$ | jump variable 消除不稳定成分 |

# 16 练习题

## A 组：基础计算

A1. 求矩阵 $A = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$ 的特征值与特征向量。

A2. 解连续时间系统 $\dot{x} = Ax$，其中 $A = \begin{pmatrix} -2 & 0 \\ 0 & -4 \end{pmatrix}$，$x(0) = (1, 3)'$。

A3. 判断 $A = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}$ 对应的连续时间系统稳定类型。

A4. 判断离散时间系统 $x_{t+1} = Ax_t$ 是否稳定，其中 $A = \begin{pmatrix} 0.3 & 0 \\ 0 & 1.1 \end{pmatrix}$。

A5. 对 $A = \begin{pmatrix} -1 & -2 \\ 2 & -1 \end{pmatrix}$，计算 trace、determinant 和判别式。

A6. 求系统 $\dot{x} = Ax + b$ 的稳态，其中 $A = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix}$，$b = (2, 4)'$。

## B 组：概念判断

B1. 连续时间中，$\lambda = -2$ 是否稳定？离散时间中，$\lambda = -2$ 是否稳定？解释。

B2. 如果二维连续时间系统有 $T < 0, D > 0$，是否一定是 stable node？

B3. 为什么 saddle path 在经济学中不是无用的发散路径？

B4. 为什么局部线性化只能保证稳态附近的性质，而不能自动推出全局稳定？

B5. 离散时间中负特征值的经济路径通常有什么可见特征？

## C 组：混合题

C1. 对 $A = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix}$，求特征值并判断 $\dot{x} = Ax$ 稳定性。

C2. 对 $A = \begin{pmatrix} 0 & 1 \\ -2 & 0 \end{pmatrix}$，判断连续时间系统的稳定类型。

C3. 对离散系统 $x_{t+1} = Ax_t$，$A = \begin{pmatrix} 0 & 1 \\ -0.2 & 0.7 \end{pmatrix}$，求特征值并判断稳定性。

C4. 非齐次系统 $x_{t+1} = Ax_t + b$ 中，$A = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.8 \end{pmatrix}$，$b = (1, 2)'$。求稳态并判断稳定性。

C5. 对非线性系统 $\dot{x} = x - y,\ \dot{y} = x^2 - y$，求原点处 Jacobian，并判断局部稳定类型。

### D 组：经济金融建模题

D1. 简化资本-消费系统为

$$\begin{pmatrix}\dot{k}\\ \dot{c}\end{pmatrix}=\begin{pmatrix}0.05 & -1\\ -0.02 & 0\end{pmatrix}\begin{pmatrix}k-k^*\\ c-c^*\end{pmatrix}.$$

判断稳态类型，并解释为什么它像 Ramsey saddle path。

D2. 金融状态变量系统为

$$z_{t+1}=\begin{pmatrix}0.95 & 0\\ 0 & 0.4\end{pmatrix}z_t+\varepsilon_{t+1}.$$

哪个因子更持久？若资产价格对第一个因子暴露很高，经济含义是什么？

D3. 一个通胀-产出系统为

$$\dot{y}=-y-2\pi,\qquad \dot{\pi}=3y-\pi.$$

判断路径是否振荡收敛。

D4. 解释为什么在有 predetermined variable 和 jump variable 的模型中，saddle path 可以给出唯一均衡路径。

### E 组：挑战 / 证明题

E1. 证明若 $A$ 可对角化，且所有连续时间特征值实部为负，则 $e^{At}x_0\to 0$。

E2. 证明若离散时间矩阵 $A$ 可对角化，且所有特征值模长小于 $1$，则 $A^tx_0\to 0$。

E3. 对二维连续系统，说明 $D<0$ 为什么意味着 saddle。

E4. 给出一个例子说明：非线性系统在稳态处 Jacobian 有零特征值时，仅靠线性化可能无法判断稳定性。

## 17 练习题参考答案

### A 组答案

**A1.** 特征值为 $1$ 和 $-2$。对应特征向量可取 $(1,0)'$ 与 $(0,1)'$。

**A2.** 解为

$$x(t)=\begin{pmatrix}e^{-2t}\\ 3e^{-4t}\end{pmatrix}.$$

两个分量都收敛到 $0$，第二个分量收敛更快。

**A3.** 特征值为 $-1$ 和 $2$，一负一正，所以是 saddle。

**A4.** 特征值为 $0.3$ 和 $1.1$。由于 $1.1>1$，离散系统不稳定。

**A5.** $T=-2$，$D=(-1)(-1)-(-2)(2)=5$，判别式

$$T^2-4D=4-20=-16<0.$$

所以特征值为复数且实部为负，是稳定焦点。

**A6.** 稳态满足 $Ax^*+b=0$，即

$$-x_1+2=0,\qquad -2x_2+4=0.$$

所以 $x^*=(2,2)'$。

## B 组答案

**B1.** 连续时间中稳定，因为 $e^{-2t} \to 0$。离散时间中不稳定，因为 $(-2)^t$ 绝对值发散。

**B2.** 不一定。$T < 0, D > 0$ 说明稳定，但可能是 stable node，也可能是 stable focus。要看判别式 $T^2 - 4D$。

**B3.** Saddle 有稳定方向。若模型中某些变量可以跳，且长期条件排除发散方向，那么经济路径可以被唯一选到稳定臂上。

**B4.** 线性化忽略了高阶项。只有在稳态附近，高阶项相对较小；远离稳态时，非线性项可能主导全局行为。

**B5.** 负特征值会导致路径在稳态两侧交替。若模长小于 1，震荡收敛；若模长大于 1，震荡发散。

## C 组答案

**C1.** 特征方程

$$\det \begin{pmatrix} -\lambda & 1 \\ -2 & -3-\lambda \end{pmatrix} = \lambda(3+\lambda) + 2 = \lambda^2 + 3\lambda + 2 = 0.$$

所以 $\lambda = -1, -2$，系统是 stable node。

**C2.** 特征方程为

$$\lambda^2 + 2 = 0,$$

所以 $\lambda = \pm i\sqrt{2}$。实部为 0，是 center。线性系统路径为周期运动。

**C3.** 特征方程

$$\lambda^2 - 0.7\lambda + 0.2 = 0.$$

判别式为 $0.49 - 0.8 = -0.31 < 0$，特征值为复数。乘积为 0.2，复特征值模长满足 $|\lambda|^2 = 0.2$，所以 $|\lambda| = \sqrt{0.2} < 1$，离散系统稳定，并呈现阻尼振荡。

**C4.** 稳态满足

$$(I - A)x^* = b,$$

即

$$\begin{pmatrix} 0.5 & 0 \\ 0 & 0.2 \end{pmatrix} x^* = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

所以

$$x^* = \begin{pmatrix} 2 \\ 10 \end{pmatrix}.$$

特征值为 $0.5, 0.8$，都在单位圆内，因此稳定。

**C5.** 系统为

$$F(x, y) = \begin{pmatrix} x - y \\ x^2 - y \end{pmatrix}.$$

Jacobian 为

$$J(x, y) = \begin{pmatrix} 1 & -1 \\ 2x & -1 \end{pmatrix}.$$

原点处

$$J(0, 0) = \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix}.$$

特征值为 1 和 $-1$，所以原点是 saddle。

## D 组答案

**D1.** 矩阵为

$$A = \begin{pmatrix} 0.05 & -1 \\ -0.02 & 0 \end{pmatrix}.$$

trace 为 $T = 0.05$，determinant 为

$$D = 0.05 \cdot 0 - (-1)(-0.02) = -0.02 < 0.$$

所以是 saddle。它像 Ramsey 模型，因为一个方向稳定，一个方向不稳定；若 $k$ 不能跳而 $c$ 可以跳，则 $c$ 必须选到稳定臂上。

**D2.** 第一个因子特征值 $0.95$，第二个因子 $0.4$。第一个更持久。资产价格若对第一个因子暴露高，说明它对长期风险或持久冲击更敏感，估值可能对折现率或长期增长预期非常敏感。

**D3.** 矩阵为

$$A = \begin{pmatrix} -1 & -2 \\ 3 & -1 \end{pmatrix}.$$

trace $T = -2$，determinant

$$D = 1 + 6 = 7.$$

判别式

$$T^2 - 4D = 4 - 28 = -24 < 0.$$

复特征值实部为 $T/2 = -1 < 0$，因此路径振荡收敛。

**D4.** predetermined variable 的初值由历史给定，不能任意选择；jump variable 可以瞬间调整。若系统是 saddle，稳定臂维度通常等于 predetermined variables 的维度。长期有界条件排除不稳定方向，于是 jump variable 的初值被唯一决定，使路径落在稳定臂上。

## E 组答案

**E1.** 若 $A = PDP^{-1}$，其中 $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$，则

$$e^{At} = Pe^{Dt}P^{-1}.$$

而

$$e^{Dt} = \mathrm{diag}(e^{\lambda_1 t}, \ldots, e^{\lambda_n t}).$$

若所有 $\mathrm{Re}(\lambda_i) < 0$，则每个 $e^{\lambda_i t} \to 0$。因此 $e^{Dt} \to 0$，从而 $e^{At}x_0 \to 0$。

**E2.** 若 $A = PDP^{-1}$，则

$$A^t = PD^tP^{-1}.$$

其中

$$D^t = \mathrm{diag}(\lambda_1^t, \ldots, \lambda_n^t).$$

如果所有 $|\lambda_i| < 1$，则 $\lambda_i^t \to 0$，所以 $D^t \to 0$，从而 $A^t x_0 \to 0$。

**E3.** 二维系统中 $D = \lambda_1\lambda_2$。若 $D < 0$，则两个特征值乘积为负，所以一个正一个负。连续时间里，一个方向发散，一个方向收敛，因此是 saddle。

**E4.** 例如一维系统

$$\dot{x} = -x^3.$$

原点是稳态，线性化导数为 $0$，线性化不能判断稳定性。但实际中若 $x > 0$，$\dot{x} < 0$；若 $x < 0$，$\dot{x} > 0$，所以原点稳定。另一个例子

$$\dot{x} = x^3$$

在原点也有零导数，但原点不稳定。这个对比说明零特征值时高阶项不可忽略。

# 18 后续学习连接

本章之后，你对动态系统的理解应当完成一次升级：从“解一条递推式”变成“分析一个多变量系统”。接下来 Chapter 13 会进入非线性动态系统初步，重点是多个稳态、局部稳定、分岔和 poverty trap 等现象。之后 Chapter 14 到 Chapter 20 会转向优化与均衡理论，但本章的特征值、Jacobian 和局部线性化仍会反复出现：Hessian 是静态优化里的二阶局部结构，Jacobian 是动态系统里的反馈结构，二者共同构成经济数学中“局部近似”的核心语言。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 13
# 非线性动态系统初步

Chapter 10 讲一维连续时间动态，Chapter 11 讲离散时间递推，Chapter 12 讲线性系统、特征值与相图。本章进入动态系统的下一层：现实经济金融模型通常不是线性的。资本积累有递减回报，金融杠杆有反馈放大，贫困陷阱有门槛效应，技术扩散可能有 S 型曲线，宏观政策规则可能在不同区间改变系统稳定性。本章的目标不是把你训练成动力系统专家，而是让你掌握经济金融中最常用的非线性动态语言：steady state、Jacobian 线性化、local stability、nullcline、多重均衡、threshold、bifurcation。读完这一章，你应当能把一个非线性动态模型先找稳态，再在稳态附近线性化，再用特征值判断局部稳定性，最后用图像和经济直觉解释为什么初始条件、参数变化和反馈机制会改变长期结果。

math4econ personalized notes
2026

# Contents

**Chapter 13 总览** **2**

**1 本章为什么重要** **2**

1.1 线性模型只是局部近似 . . . 2
1.2 非线性带来经济上更丰富的现象 . . . 3
1.3 本章服务哪些后续内容 . . . 3

**2 最小前置知识** **3**

2.1 一维动态的稳定性 . . . 3
2.2 矩阵特征值 . . . 3
2.3 Taylor 展开 . . . 4
2.4 Jacobian 矩阵 . . . 4

**3 非线性动态系统的基本对象** **4**

3.1 连续时间非线性系统 . . . 4
3.2 稳态 steady state . . . 5
3.3 离散时间非线性系统 . . . 5

**4 一维非线性动态：相线图与门槛** **5**

4.1 相线图的基本方法 . . . 5
4.2 导数判别的推导 . . . 6
4.3 Logistic 动态：非线性收敛的标准例子 . . . 6
4.4 门槛模型与贫困陷阱 . . . 7

**5 二维非线性系统：nullcline 与相图** **8**

5.1 二维系统的形式 . . . 8
5.2 Nullcline 的含义 . . . 8
5.3 一个非线性二维例子 . . . 9

**6 Jacobian 线性化：非线性系统的核心工具** **10**

6.1 局部线性化公式 . . . 10
6.2 二维 Jacobian . . . 11
6.3 连续时间二维系统的迹-行列式判别 . . . 11

**7 Worked Example 1：Solow 模型的非线性稳定性** **12**

**8 Worked Example 2：二维非线性系统的局部稳定性** **12**

**9 离散时间非线性系统** **13**

9.1 局部线性化 . . . 13

9.2 一维离散系统 . . . . . 14
9.3 Logistic map 与复杂动态 . . . . . 14

**10 多重均衡与路径依赖 15**
10.1 多重均衡的数学结构 . . . . . 15
10.2 经济解释：贫困陷阱 . . . . . 15
10.3 金融解释：杠杆和流动性陷阱 . . . . . 16

**11 分岔 bifurcation：参数改变结构 16**
11.1 什么是 bifurcation . . . . . 16
11.2 Saddle-node bifurcation . . . . . 16

**12 Worked Example 3：非线性债务动态 17**

**13 Worked Example 4：金融市场的非线性均值回复 18**

**14 局部稳定、全局稳定与 basin of attraction 19**
14.1 局部稳定不等于全局稳定 . . . . . 19
14.2 吸引域 basin of attraction . . . . . 19

**15 经济金融应用地图 19**
15.1 宏观：从 Solow 到 Ramsey 再到非线性宏观 . . . . . 19
15.2 金融：非线性反馈和危机状态 . . . . . 20
15.3 计量：非线性时间序列 . . . . . 20

**16 本章方法总表 20**

**17 分层练习 20**

**18 详细解答 22**

**19 本章总结图谱 24**

# Chapter 13 总览

前面三章已经建立了动态模型的基本工具。Chapter 10 中，一维连续时间模型的典型形式是

$$\dot{x} = f(x).$$

Chapter 11 中，一维离散时间模型是

$$x_{t+1} = g(x_t).$$

Chapter 12 中，多维线性系统是

$$\dot{x} = Ax, \qquad x_{t+1} = Ax_t.$$

本章把这些东西合起来，研究一般的非线性系统：

$$\dot{x} = F(x;\theta), \qquad x \in \mathbb{R}^n,$$

或

$$x_{t+1} = G(x_t;\theta), \qquad x_t \in \mathbb{R}^n.$$

这里 $\theta$ 是参数，可能代表储蓄率、贴现率、技术水平、政策强度、风险偏好、杠杆上限或货币政策规则。

**直觉框**

非线性动态系统的学习顺序很简单：先找不动的点，再看这些点附近怎么动，再看参数改变时这些点和路径会不会发生结构性变化。对应数学语言就是：steady states，local linearization，Jacian eigenvalues，phase diagram，bifurcation。

本章最重要的桥梁是：虽然模型本身非线性，但在稳态附近可以用线性系统近似。也就是如果 $x^*$ 满足 $F(x^*) = 0$，令 $y = x - x^*$，则

$$\dot{y} = F(x^* + y) \approx J_F(x^*)y.$$

于是 Chapter 12 的矩阵特征值工具仍然能用。非线性动态的第一步，不是立刻求精确解，而是把系统在稳态附近线性化。

## 1 本章为什么重要

### 1.1 线性模型只是局部近似

经济学中很多模型一开始看起来线性，是因为我们在稳态附近做了近似。真正的结构往往是非线性的。例如 Solow 模型：

$$\dot{k} = sf(k) - (n+\delta)k.$$

如果 $f(k) = k^\alpha$，那么

$$\dot{k} = sk^\alpha - (n+\delta)k,$$

这显然不是线性的。我们只有在稳态 $k^*$ 附近，才可以写成

$$\dot{k} \approx \left[s\alpha(k^*)^{\alpha-1} - (n+\delta)\right](k - k^*).$$

这个系数决定局部收敛速度，但整个路径仍然由非线性方程决定。

### 1.2 非线性带来经济上更丰富的现象

线性系统已经能描述收敛、发散、鞍点和振荡。但很多经济现象必须靠非线性表达：

- **多重均衡**：同一个经济体可能走向高收入均衡，也可能陷入低收入均衡；
- **门槛效应**：初始资本、教育、人力资本或金融信心低于某个水平时，系统向下走；超过门槛后才向上走；
- **金融反馈**：价格下跌导致抵押品价值下降，进一步导致去杠杆和抛售；
- **政策非线性**：小政策变化没有明显影响，但超过阈值后会改变均衡数量或稳定性；
- **周期与混沌**：离散时间非线性模型中，参数变大时可能从稳定收敛变成周期波动甚至复杂轨迹。

**经济金融解释**

如果你只会线性动态，你会倾向于认为经济冲击的影响是平滑、比例、可逆的。但非线性模型提醒我们：同样大小的冲击在不同状态下可能有完全不同的效果，政策参数的小变化可能导致均衡突然消失，初始条件可能决定长期命运。

### 1.3 本章服务哪些后续内容

本章服务五类后续课程：

1. 高级宏观：Ramsey、RBC、内生增长、贫困陷阱、金融加速器；
2. 金融经济学：杠杆循环、均值回复状态变量、波动率反馈、非线性风险溢价；
3. 计量与时间序列：非线性 AR、threshold models、state-dependent effects；
4. 动态优化：Bellman 方程的局部性质和策略函数近似；
5. 博弈与机制：多重均衡、学习动态、演化稳定性。

## 2 最小前置知识

本章需要四个前置工具。

### 2.1 一维动态的稳定性

连续时间中，如果

$$\dot{x} = f(x),$$

稳态 $x^*$ 满足 $f(x^*) = 0$。若在 $x^*$ 附近，左边 $f(x) > 0$，右边 $f(x) < 0$，则轨迹向 $x^*$ 靠近，稳态局部稳定。若左边 $f(x) < 0$，右边 $f(x) > 0$，则轨迹远离 $x^*$，稳态不稳定。

更简洁的导数判别是：若 $f'(x^*) < 0$，则连续时间一维系统局部稳定；若 $f'(x^*) > 0$，则局部不稳定。

### 2.2 矩阵特征值

对线性系统

$$\dot{y} = Ay,$$

如果所有特征值的实部都为负，则稳态局部渐近稳定；如果存在实部为正的特征值，则存在不稳定方向。

对离散系统

$$y_{t+1} = Ay_t,$$

如果所有特征值的模长都小于 1，则稳态局部渐近稳定；如果存在模长大于 1 的特征值，则存在不稳定方向。

### 2.3 Taylor 展开

一元函数在 $x^*$ 附近：

$$f(x) = f(x^*) + f'(x^*)(x - x^*) + \frac{1}{2}f''(\xi)(x - x^*)^2.$$

多元函数在 $x^*$ 附近：

$$F(x) = F(x^*) + J_F(x^*)(x - x^*) + \text{higher order terms.}$$

非线性系统的局部线性化正是这个展开的动态版本。

### 2.4 Jacobian 矩阵

若 $F : \mathbb{R}^n \to \mathbb{R}^n$，写作

$$F(x) = \begin{pmatrix} F_1(x) \\ \vdots \\ F_n(x) \end{pmatrix},$$

则 Jacobian 是

$$J_F(x) = \begin{pmatrix} \dfrac{\partial F_1}{\partial x_1} & \cdots & \dfrac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F_n}{\partial x_1} & \cdots & \dfrac{\partial F_n}{\partial x_n} \end{pmatrix}.$$

它是非线性系统在某一点附近的最佳线性近似。

## 3 非线性动态系统的基本对象

### 3.1 连续时间非线性系统

连续时间非线性系统的一般形式是

$$\dot{x} = F(x; \theta), \qquad x \in \mathbb{R}^n.$$

这里 $x$ 是状态向量，$F$ 是变化率函数。若 $n = 1$，就是一维非线性 ODE：

$$\dot{x} = f(x; \theta).$$

若 $n = 2$，就是二维相图中最常见的系统：

$$\begin{cases} \dot{x} = F_1(x, y; \theta), \\ \dot{y} = F_2(x, y; \theta). \end{cases}$$

### 3.2 稳态 steady state

连续时间系统的稳态是让所有状态变量都不再变化的点。形式上，$x^*$ 是稳态，如果

$$F(x^*;\theta)=0.$$

一维时是 $f(x^*)=0$；二维时是

$$F_1(x^*,y^*)=0,\qquad F_2(x^*,y^*)=0.$$

**直觉框**

稳态不是“系统一定会到达的点”，而只是“不动的点”。稳定性要另外判断。一个模型可以有多个稳态，其中有的稳定，有的不稳定，有的是鞍点。

### 3.3 离散时间非线性系统

离散时间系统的一般形式是

$$x_{t+1}=G(x_t;\theta).$$

稳态或固定点满足

$$x^*=G(x^*;\theta).$$

这和连续时间的 $F(x^*)=0$ 形式不同。连续时间看变化率为零，离散时间看下一期等于本期。

**常见误区**

连续时间和离散时间的稳定性条件不能混用。连续时间一维系统中，$f'(x^*)<0$ 通常表示稳定；离散时间一维系统中，稳定条件是 $|g'(x^*)|<1$。一个负的导数在离散时间中不一定稳定，如果 $g'(x^*)=-2$，轨迹会交替发散。

## 4 一维非线性动态：相线图与门槛

### 4.1 相线图的基本方法

考虑

$$\dot{x}=f(x).$$

相线图 phase line 的方法是：

1. 解 $f(x)=0$ 找所有稳态；
2. 在每两个稳态之间选测试点；
3. 判断 $f(x)$ 的正负；
4. 若 $f(x)>0$，则 $x$ 增加，箭头向右；若 $f(x)<0$，则 $x$ 减少，箭头向左；
5. 根据箭头是否指向稳态判断稳定性。

### 4.2 导数判别的推导

设 $x^*$ 是稳态，即 $f(x^*)=0$。令 $y=x-x^*$。则

$$\dot{y}=\dot{x}=f(x^*+y).$$

对 $f$ 在 $x^*$ 附近做一阶 Taylor 展开：

$$f(x^*+y)=f(x^*)+f'(x^*)y+o(y).$$

因为 $f(x^*)=0$，所以

$$\dot{y}=f'(x^*)y+o(y).$$

当 $y$ 很小时，主导项是

$$\dot{y}\approx f'(x^*)y.$$

其线性近似解为

$$y(t)\approx y(0)e^{f'(x^*)t}.$$

因此：

- 若 $f'(x^*)<0$，指数衰减，局部稳定；
- 若 $f'(x^*)>0$，指数增长，局部不稳定；
- 若 $f'(x^*)=0$，一阶线性化失效，需要看更高阶项。

**推导与证明**

这个推导的经济含义是：稳定性由稳态附近的边际反馈决定。若变量略高于稳态时系统把它往下拉，略低于稳态时系统把它往上推，就是负反馈；若偏离越多系统越把它推远，就是正反馈。

### 4.3 Logistic 动态：非线性收敛的标准例子

考虑

$$\dot{x}=rx\left(1-\frac{x}{K}\right),\qquad r>0,\quad K>0.$$

稳态满足

$$rx\left(1-\frac{x}{K}\right)=0,$$

所以

$$x^*=0,\qquad x^*=K.$$

令

$$f(x)=rx\left(1-\frac{x}{K}\right).$$

则

$$f'(x)=r-\frac{2r}{K}x.$$

在 $x^*=0$ 处：

$$f'(0)=r>0,$$

所以 $0$ 不稳定。在 $x^*=K$ 处：

$$f'(K)=r-2r=-r<0,$$

所以 $K$ 局部稳定。

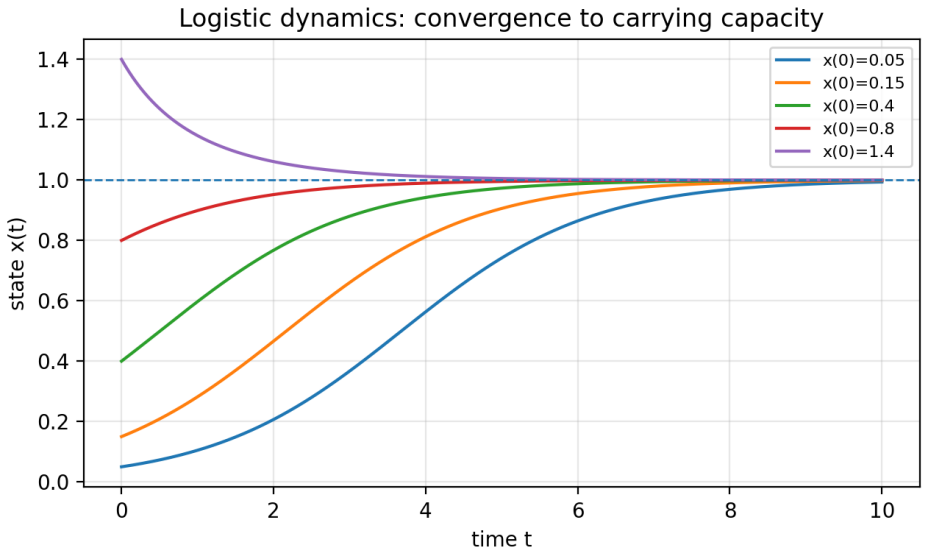



Figure 1: Logistic 动态：不同初始值都向承载水平 $K$ 收敛，但路径速度和方向不同。

**经济金融解释**

Logistic 形式可以解释技术扩散、平台用户增长、信息传播和市场渗透率。早期 $x$ 很小时，增长近似指数；接近上限 $K$ 时，增长变慢。这种 S 型过程不能用简单线性模型表达。

### 4.4 门槛模型与贫困陷阱

考虑

$$\dot{x} = x(1-x)(x-a), \qquad 0 < a < 1.$$

稳态为

$$x^* = 0, \qquad x^* = a, \qquad x^* = 1.$$

看不同区间的符号：

- 若 $x < 0$，符号不重要，因为经济变量通常取非负；
- 若 $0 < x < a$，则 $x > 0$，$1 - x > 0$，$x - a < 0$，所以 $\dot{x} < 0$；
- 若 $a < x < 1$，三个因子都为正，所以 $\dot{x} > 0$；
- 若 $x > 1$，$1 - x < 0$，所以 $\dot{x} < 0$。

因此 $0$ 和 $1$ 是稳定稳态，而 $a$ 是不稳定门槛。

**经济金融解释**

把 $x$ 理解为人力资本、基础设施、金融信任、产业集聚程度或技术采纳率，$a$ 就是临界门槛。低于门槛时，个体或地区缺乏足够互补条件，系统退回低水平均衡；高于门槛时，正反馈推动系统走向高水平均衡。这就是 poverty trap 和 big push 理论的动态骨架。

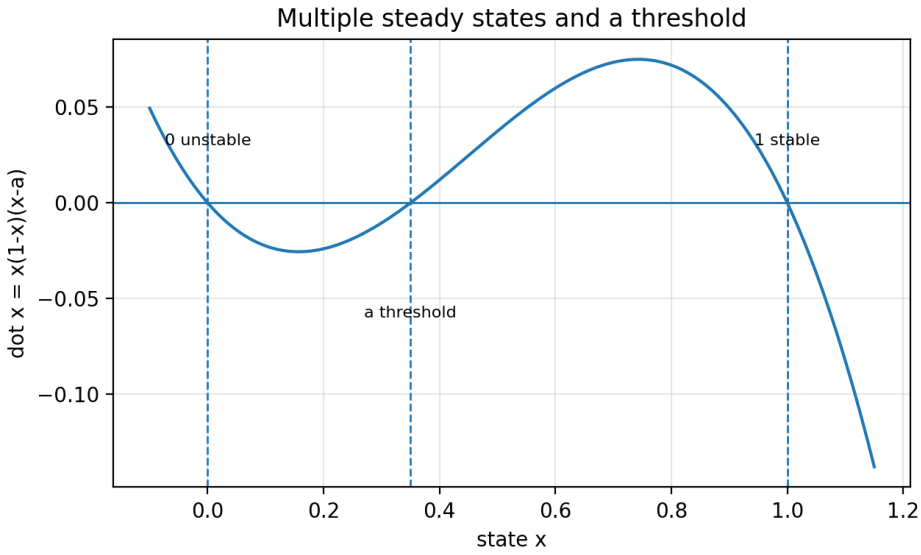



Figure 2: 多重稳态与门槛：$a$ 是分水岭，低于它向低稳态走，高于它向高稳态走。

# 5 二维非线性系统：nullcline 与相图

## 5.1 二维系统的形式

二维连续时间非线性系统写作

$$\begin{cases} \dot{x} = F(x, y), \\ \dot{y} = G(x, y). \end{cases}$$

稳态满足

$$F(x^*, y^*) = 0, \qquad G(x^*, y^*) = 0.$$

在二维里，仅靠相线不够，需要相图 phase diagram。

## 5.2 Nullcline 的含义

$x$-nullcline 是让 $x$ 不变的点集：

$$\dot{x} = F(x, y) = 0.$$

在这条曲线上，轨迹的水平变化为零，只可能向上或向下动。

$y$-nullcline 是让 $y$ 不变的点集：

$$\dot{y} = G(x, y) = 0.$$

在这条曲线上，轨迹的竖直变化为零，只可能向左或向右动。

两条 nullcline 的交点就是稳态。

**方法框**

二维相图画法：

1. 写出 $\dot{x} = 0$ 和 $\dot{y} = 0$ 两条 nullcline；
2. 求交点，得到稳态；
3. 在各区域判断 $\dot{x}$ 和 $\dot{y}$ 的正负；
4. 用箭头表示局部运动方向；

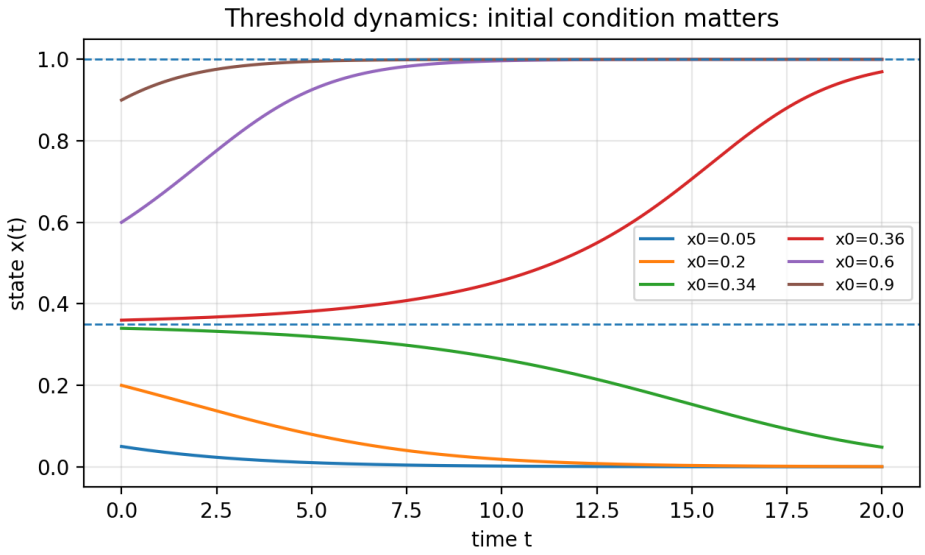



Figure 3: 门槛动态路径：初始条件决定长期命运。

5. 在稳态处计算 Jacobian，用特征值判断局部稳定类型。

### 5.3 一个非线性二维例子

考虑

$$\begin{cases} \dot{x} = x(1-x) - 0.45y, \\ \dot{y} = 0.65x - y. \end{cases}$$

第一条 nullcline 是

$$y = \frac{x(1-x)}{0.45}.$$

第二条 nullcline 是

$$y = 0.65x.$$

交点满足

$$\frac{x(1-x)}{0.45} = 0.65x.$$

一个交点是 $x = 0, y = 0$。若 $x \neq 0$，两边除以 $x$ 得

$$\frac{1-x}{0.45} = 0.65,$$

所以

$$1 - x = 0.2925, \qquad x = 0.7075, \qquad y = 0.65x \approx 0.4599.$$

**直觉框**

Nullcline 的意义是把二维运动分解成“哪里横向不动”和“哪里纵向不动”。经济学相图中经常出现 $\dot{k} = 0$ 曲线和 $\dot{c} = 0$ 曲线，正是这个思想。

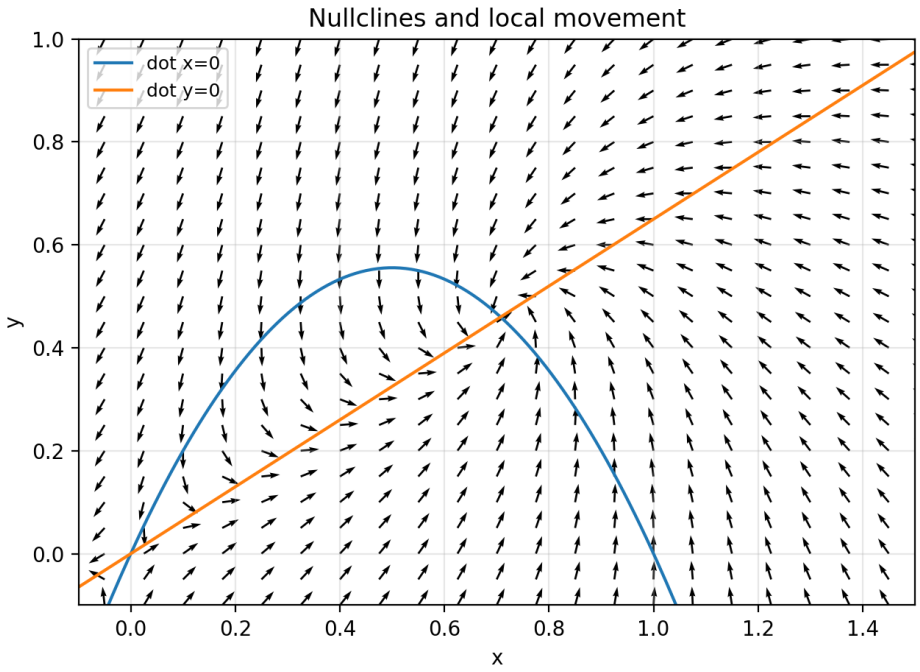



Figure 4: Nullclines 与向量场：交点是稳态，箭头显示局部运动方向。

# 6 Jacobian 线性化：非线性系统的核心工具

## 6.1 局部线性化公式

考虑

$$\dot{x} = F(x), \qquad x \in \mathbb{R}^n.$$

设 $x^*$ 是稳态，满足

$$F(x^*) = 0.$$

令

$$y = x - x^*.$$

则

$$\dot{y} = \dot{x} = F(x^* + y).$$

对 $F$ 在 $x^*$ 附近做一阶 Taylor 展开：

$$F(x^* + y) = F(x^*) + J_F(x^*)y + R(y),$$

其中余项满足

$$\frac{\|R(y)\|}{\|y\|} \to 0 \quad \text{as } y \to 0.$$

因为 $F(x^*) = 0$，得到

$$\dot{y} = J_F(x^*)y + R(y).$$

局部线性化就是忽略高阶小量：

$$\dot{y} \approx J_F(x^*)y.$$

推导与证明

这一步把非线性系统转化为 Chapter 12 的线性系统。也就是说，稳态附近的局部动态由 Jacobian 矩阵决定。Jacobian 的特征值告诉我们偏离稳态以后，系统会被拉回、推远、绕圈，还是沿着某些方向稳定而另一些方向不稳定。

## 6.2 二维 Jacobian

若

$$\begin{cases} \dot{x} = F(x, y), \\ \dot{y} = G(x, y), \end{cases}$$

则 Jacobian 是

$$J(x, y) = \begin{pmatrix} F_x(x, y) & F_y(x, y) \\ G_x(x, y) & G_y(x, y) \end{pmatrix}.$$

在稳态 $(x^*, y^*)$ 处的线性化为

$$\begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} = J(x^*, y^*) \begin{pmatrix} u \\ v \end{pmatrix},$$

其中

$$u = x - x^*, \qquad v = y - y^*.$$

## 6.3 连续时间二维系统的迹-行列式判别

对于二维线性系统

$$\dot{z} = Az, \qquad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

特征方程是

$$\lambda^2 - \operatorname{tr}(A)\lambda + \det(A) = 0.$$

记

$$T = \operatorname{tr}(A) = a + d, \qquad D = \det(A) = ad - bc.$$

判别规则：

- 若 $D < 0$，两个实特征值异号，稳态是 saddle；
- 若 $D > 0$ 且 $T < 0$，两个特征值实部为负，局部稳定；
- 若 $D > 0$ 且 $T > 0$，两个特征值实部为正，局部不稳定；
- 若 $T^2 - 4D < 0$，特征值为复数，路径带旋转；
- 若 $T^2 - 4D > 0$，特征值为实数，路径不旋转。

方法框

二维非线性系统局部稳定性流程：

1. 求稳态 $(x^*, y^*)$；
2. 写 Jacobian $J(x, y)$；
3. 代入稳态得到矩阵 $J^* = J(x^*, y^*)$；
4. 计算 $T = \operatorname{tr}(J^*)$ 与 $D = \det(J^*)$；
5. 用 $D, T, T^2 - 4D$ 判断局部类型。

## 7 Worked Example 1：Solow 模型的非线性稳定性

Worked Example

考虑 Solow 型资本动态：

$$\dot{k} = sk^{\alpha} - (n+\delta)k, \qquad 0 < \alpha < 1, \quad s, n+\delta > 0.$$

求稳态并判断稳定性。

**Step 1：找稳态。**稳态满足

$$sk^{\alpha} - (n+\delta)k = 0.$$

因式分解：

$$k\left(sk^{\alpha-1} - (n+\delta)\right) = 0.$$

一个稳态是 $k = 0$。正稳态满足

$$sk^{\alpha-1} = n+\delta.$$

因为 $\alpha - 1 < 0$，所以

$$k^{\alpha-1} = \frac{n+\delta}{s},$$

等价于

$$k^* = \left(\frac{s}{n+\delta}\right)^{\frac{1}{1-\alpha}}.$$

**Step 2：计算导数。**令

$$f(k) = sk^{\alpha} - (n+\delta)k.$$

则

$$f'(k) = s\alpha k^{\alpha-1} - (n+\delta).$$

在正稳态处，由稳态条件 $s(k^*)^{\alpha-1} = n+\delta$，所以

$$f'(k^*) = \alpha(n+\delta) - (n+\delta) = -(1-\alpha)(n+\delta) < 0.$$

因此正稳态局部稳定。

**Step 3：经济解释。**如果 $k$ 略低于 $k^*$，由于资本边际产出较高，储蓄带来的投资超过折旧和稀释，资本上升。如果 $k$ 略高于 $k^*$，递减回报导致新增投资不足以抵消折旧稀释，资本下降。这是负反馈。

## 8 Worked Example 2：二维非线性系统的局部稳定性

Worked Example

考虑系统

$$\begin{cases} \dot{x} = x(1-x) - 0.45y, \\ \dot{y} = 0.65x - y. \end{cases}$$

判断非零稳态的局部稳定性。

**Step 1：求稳态。**前面已经求得非零稳态为

$$x^* = 0.7075, \qquad y^* = 0.65x^* \approx 0.459875.$$

**Step 2：写 Jacobian。令**

$$F(x,y)=x(1-x)-0.45y,\qquad G(x,y)=0.65x-y.$$

则

$$F_x=1-2x,\qquad F_y=-0.45,\qquad G_x=0.65,\qquad G_y=-1.$$

所以

$$J(x,y)=\begin{pmatrix}1-2x & -0.45\\ 0.65 & -1\end{pmatrix}.$$

**Step 3：代入稳态。**

$$J^*=\begin{pmatrix}1-2(0.7075) & -0.45\\ 0.65 & -1\end{pmatrix}=\begin{pmatrix}-0.415 & -0.45\\ 0.65 & -1\end{pmatrix}.$$

迹为

$$T=-0.415-1=-1.415<0.$$

行列式为

$$D=(-0.415)(-1)-(-0.45)(0.65)=0.415+0.2925=0.7075>0.$$

因为 $D>0$ 且 $T<0$，所以该稳态局部稳定。
**Step 4：是否振荡？**看判别式：

$$T^2-4D=(-1.415)^2-4(0.7075)=2.002225-2.83<0.$$

所以特征值是复数，且实部为负。路径会以旋转方式收敛，即 stable focus。

# 9 离散时间非线性系统

## 9.1 局部线性化

考虑

$$x_{t+1}=G(x_t),$$

稳态 $x^*$ 满足

$$x^*=G(x^*).$$

令 $y_t=x_t-x^*$，则

$$y_{t+1}=x_{t+1}-x^*=G(x^*+y_t)-G(x^*).$$

Taylor 展开得到

$$y_{t+1}\approx J_G(x^*)y_t.$$

所以离散时间非线性系统的局部稳定性由 $J_G(x^*)$ 的特征值决定。

**方法框**

离散时间局部稳定性：若 $J_G(x^*)$ 的所有特征值 $\lambda_i$ 都满足

$$|\lambda_i|<1,$$

则稳态局部渐近稳定。若存在 $|\lambda_i|>1$，则存在不稳定方向。

## 9.2 一维离散系统

一维离散系统

$$x_{t+1} = g(x_t)$$

的稳态满足

$$x^* = g(x^*).$$

局部线性化为

$$y_{t+1} \approx g'(x^*)y_t.$$

所以稳定条件是

$$|g'(x^*)| < 1.$$

如果

$$0 < g'(x^*) < 1,$$

则轨迹单调收敛；如果

$$-1 < g'(x^*) < 0,$$

则轨迹交替收敛；如果

$$g'(x^*) > 1,$$

则单调发散；如果

$$g'(x^*) < -1,$$

则交替发散。

## 9.3 Logistic map 与复杂动态

离散 Logistic map 是

$$x_{t+1} = rx_t(1 - x_t), \qquad x_t \in [0, 1].$$

它看起来和连续时间 logistic ODE 类似，但动态性质完全不同。稳态满足

$$x = rx(1 - x).$$

所以

$$x^* = 0, \qquad x^* = 1 - \frac{1}{r} \quad (r > 1).$$

对

$$g(x) = rx(1 - x),$$

有

$$g'(x) = r(1 - 2x).$$

在非零稳态 $x^* = 1 - 1/r$ 处：

$$g'(x^*) = r\left[1 - 2\left(1 - \frac{1}{r}\right)\right] = r(-1 + 2/r) = 2 - r.$$

稳定条件是

$$|2 - r| < 1.$$

即

$$1 < r < 3.$$

当 $r$ 超过 3 后，非零稳态失去稳定性，系统进入周期，再进一步可能出现复杂动态。

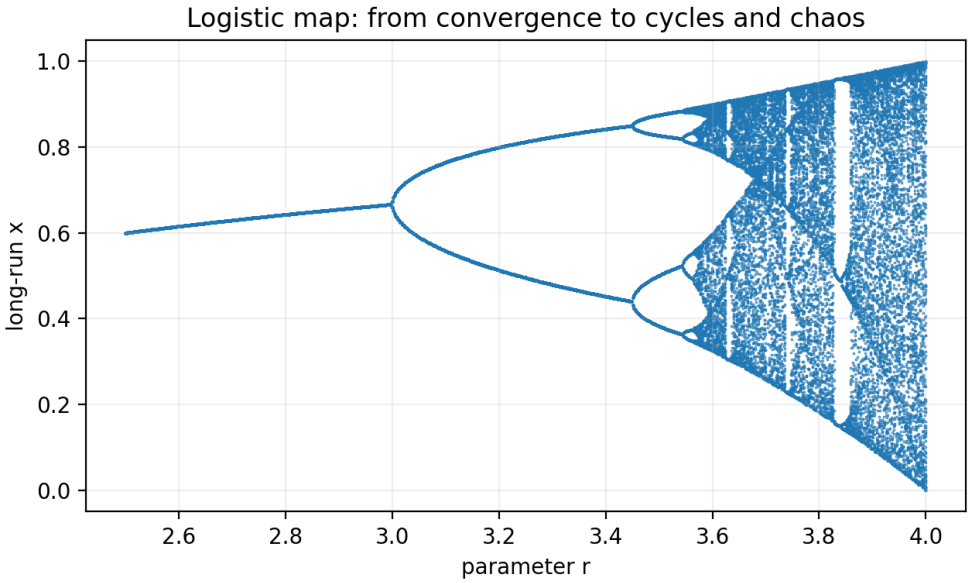



Figure 5: 离散 Logistic map 的参数变化：从收敛到周期，再到复杂动态。

**常见误区**

连续时间 logistic ODE 与离散时间 logistic map 不能混为一谈。前者通常平滑收敛到承载水平；后者在参数较大时可以出现周期和复杂动态。离散时间模型更容易因为“过度反应”而产生振荡。

# 10 多重均衡与路径依赖

## 10.1 多重均衡的数学结构

多重均衡通常来自非线性反馈。一个最简单的一维表达是

$$\dot{x} = f(x)$$

有多个根：

$$f(x_1^*) = f(x_2^*) = f(x_3^*) = 0.$$

这些根不一定都是稳定的。常见结构是：

$$\text{stable} \quad - \quad \text{unstable threshold} \quad - \quad \text{stable}.$$

这意味着初始条件决定最终均衡。

## 10.2 经济解释：贫困陷阱

设 $x$ 表示人均资本、人力资本或产业集聚程度。若

$$\dot{x} = x(1-x)(x-a),$$

那么低于 $a$ 时，系统向低稳态 $0$ 走；高于 $a$ 时，系统向高稳态 $1$ 走。政策含义是：小规模补贴可能不够，因为它不能把系统推过门槛；一次性大推动 big push 可能改变长期路径。

**经济金融解释**

贫困陷阱模型中，政策效果不是线性的。若政策只把 $x$ 从 0.2 提到 0.3，但门槛 $a = 0.35$，长期仍然回到低均衡。若政策把 $x$ 推到 0.36，哪怕只超过一点，系统也可能自己走向高均衡。非线性动态给了“大推动”理论数学基础。

### 10.3 金融解释：杠杆和流动性陷阱

在金融市场中，状态变量 $x$ 可以表示市场信心、抵押品价值或流动性。若价格下跌会降低抵押品价值，抵押品下降又会迫使机构卖出资产，卖出进一步压低价格，就形成正反馈。系统可能有两个稳定状态：正常流动性状态和危机状态，中间有一个不稳定门槛。

## 11 分岔 bifurcation：参数改变结构

### 11.1 什么是 bifurcation

Bifurcation 指的是：当参数 $\theta$ 连续变化时，系统的长期动态结构发生质变，例如稳态数量改变、稳定性改变、周期出现或消失。

这不是普通比较静态。普通比较静态问：

$$\theta \uparrow \quad \Longrightarrow \quad x^* \text{ 怎么连续变化?}$$

而 bifurcation 问：

$$\theta \text{ 变到某个临界值时，均衡结构是否突然变了?}$$

### 11.2 Saddle-node bifurcation

考虑

$$\dot{x} = r - x^2.$$

稳态满足

$$r - x^2 = 0.$$

若 $r > 0$，有两个稳态：

$$x^* = \sqrt{r}, \qquad x^* = -\sqrt{r}.$$

若 $r = 0$，两个稳态合并为一个。若 $r < 0$，没有实稳态。

稳定性由

$$f'(x) = -2x$$

判断。在 $x^* = \sqrt{r}$ 处，

$$f'(\sqrt{r}) = -2\sqrt{r} < 0,$$

所以稳定。在 $x^* = -\sqrt{r}$ 处，

$$f'(-\sqrt{r}) = 2\sqrt{r} > 0,$$

所以不稳定。

**经济金融解释**

Saddle-node bifurcation 可以帮助理解“政策参数或市场条件低于临界值时，某种稳定状态根本不存在”。例如金融市场中，如果流动性支持不足，正常均衡可能消失；经济只能滑向危机动态。政策不是让均衡稍微移动，而是让均衡重新出现。

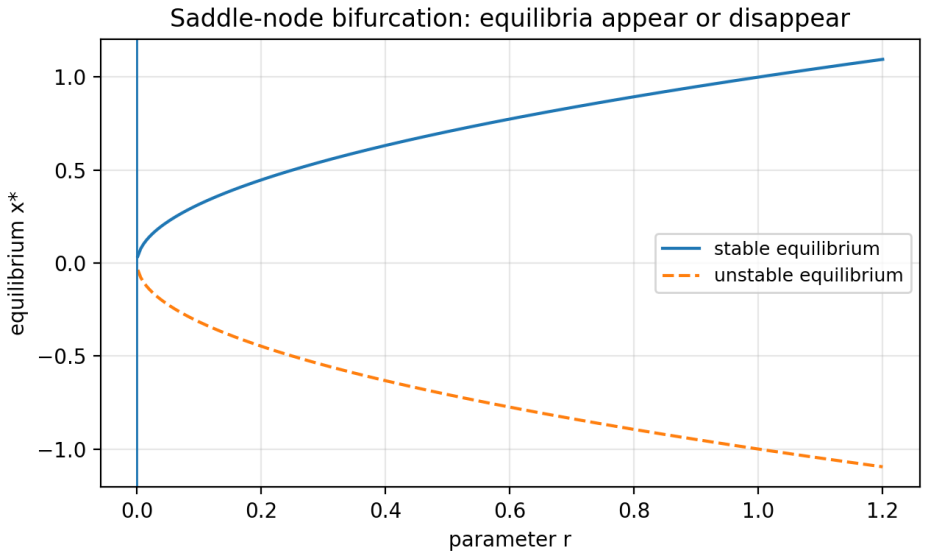



Figure 6: Saddle-node bifurcation：参数低于临界值时没有稳态，高于临界值时出现一稳一定两个稳态。

## 12 Worked Example 3：非线性债务动态

**Worked Example**

设政府债务比率 $b(t)$ 的动态为

$$\dot{b} = (r(b) - g)b - s,$$

其中 $s > 0$ 是 primary surplus，占 GDP 比例；$g$ 是增长率；利率随债务上升而上升：

$$r(b) = r_0 + \phi b, \qquad \phi > 0.$$

于是

$$\dot{b} = (r_0 - g)b + \phi b^2 - s.$$

分析稳态结构。

**Step 1：写成二次方程。**稳态满足

$$(r_0 - g)b + \phi b^2 - s = 0.$$

即

$$\phi b^2 + (r_0 - g)b - s = 0.$$

**Step 2：判断稳态数量。**判别式为

$$\Delta = (r_0 - g)^2 + 4\phi s > 0.$$

因此总有两个实根：

$$b^*_{1,2} = \frac{-(r_0 - g) \pm \sqrt{(r_0 - g)^2 + 4\phi s}}{2\phi}.$$

由于 $s > 0$，一个根通常为正，一个根为负；经济上只关注正根。

**Step 3：稳定性。**令

$$f(b) = (r_0 - g)b + \phi b^2 - s.$$

则

$$f'(b) = r_0 - g + 2\phi b.$$

在较大的正根处，$f'(b^*)$ 往往为正，因此局部不稳定。这意味着债务略高于该阈值时，利率上升反馈会使债务进一步增长。

**Step 4：经济解释。**如果利率不依赖债务，债务动态大致线性；如果利率会随债务比率上升，则系统出现非线性反馈。债务高时融资成本更高，融资成本更高又推高债务。这是债务危机的基本非线性机制。

## 13 Worked Example 4：金融市场的非线性均值回复

Worked Example

考虑一个状态变量 $x_t$ 表示风险溢价偏离长期均值的程度。假设离散动态为

$$x_{t+1} = \rho x_t - \gamma x_t^3, \qquad 0 < \rho < 1, \quad \gamma > 0.$$

分析 $x^* = 0$ 的稳定性。

**Step 1：求稳态。**稳态满足

$$x = \rho x - \gamma x^3.$$

移项：

$$(1-\rho)x + \gamma x^3 = 0.$$

即

$$x\left(1 - \rho + \gamma x^2\right) = 0.$$

因为 $1-\rho > 0$ 且 $\gamma x^2 \geq 0$，唯一实稳态是

$$x^* = 0.$$

**Step 2：局部线性化。**令

$$g(x) = \rho x - \gamma x^3.$$

则

$$g'(x) = \rho - 3\gamma x^2.$$

在 $0$ 处：

$$g'(0) = \rho.$$

因为 $0 < \rho < 1$，所以 $|g'(0)| < 1$，稳态局部稳定。

**Step 3：非线性项的含义。**在 $x$ 很小时，动态近似为

$$x_{t+1} \approx \rho x_t.$$

但当 $|x|$ 大时，$-\gamma x_t^3$ 会产生更强的拉回效应。也就是说，小偏离时近似 $\mathrm{AR}(1)$，大偏离时均值回复更强。

# 14 局部稳定、全局稳定与 basin of attraction

## 14.1 局部稳定不等于全局稳定

局部稳定 local stability 意味着：如果初始值足够接近稳态，轨迹会回到该稳态。但它不保证远处初始值也会回来。

全局稳定 global stability 意味着：从很大范围甚至整个状态空间出发，轨迹都会收敛到该稳态。

**常见误区**

经济学中常犯的错误是：在稳态附近线性化，得到局部稳定，然后口头说“这个经济会稳定”。严格地说，你只能说“该稳态局部稳定”。若模型存在多个稳态，局部稳定只说明其中一个吸引域里的轨迹会靠近它。

## 14.2 吸引域 basin of attraction

某个稳定稳态的吸引域是所有会收敛到它的初始条件集合。对于贫困陷阱模型

$$\dot{x} = x(1-x)(x-a),$$

低稳态 0 的吸引域是

$$(0, a),$$

高稳态 1 的吸引域是

$$(a, \infty)$$

在经济可行范围内则是 $(a, \infty)$ 或 $(a, 1]$。

**经济金融解释**

吸引域是政策分析中非常重要的概念。政策未必需要永久改变系统，只要一次性把状态变量推过分界线，系统后续可能自动走向更好的稳态。反过来，如果政策没有跨过分界线，长期效果可能消失。

# 15 经济金融应用地图

## 15.1 宏观：从 Solow 到 Ramsey 再到非线性宏观

Solow 模型是一维非线性动态。Ramsey 模型是二维非线性系统。RBC 和 New Keynesian 模型通常会在稳态附近 log-linearization，得到线性系统。但原始模型大多是非线性的。

典型步骤是：

1. 写出非线性均衡条件；
2. 求稳态；
3. 对稳态附近变量做 log-deviation；
4. 线性化得到矩阵系统；
5. 用特征值和 saddle-path 条件判断解的稳定性。

### 15.2 金融：非线性反馈和危机状态

金融危机常常不是线性冲击叠加，而是非线性反馈。比如：

$$\text{资产价格下跌} \to \text{抵押品价值下降} \to \text{融资约束收紧} \to \text{强制卖出} \to \text{价格进一步下跌}.$$

这种反馈可以产生多个状态：正常状态、脆弱状态、危机状态。线性模型通常只能描述小波动，非线性模型才能表达 regime shift。

### 15.3 计量：非线性时间序列

线性 AR(1) 是

$$x_{t+1} = \rho x_t + \varepsilon_{t+1}.$$

非线性版本可以是

$$x_{t+1} = g(x_t) + \varepsilon_{t+1},$$

或 threshold model：

$$x_{t+1} = \begin{cases} \rho_1 x_t + \varepsilon_{t+1}, & x_t \le c, \\ \rho_2 x_t + \varepsilon_{t+1}, & x_t > c. \end{cases}$$

这类模型表达的是状态依赖效应 state-dependent effects。比如经济衰退期财政乘数可能大于扩张期，金融压力高时货币政策传导可能不同。

## 16 本章方法总表

| 问题 | 数学步骤 | 经济解释 |
|---|---|---|
| 找长期状态 | 解 $F(x^*)=0$ 或 $x^*=G(x^*)$ | 找不再变化的状态，不等于一定会达到 |
| 判断一维连续稳定性 | 看 $f'(x^*)$ 符号 | 负反馈稳定，正反馈不稳定 |
| 判断一维离散稳定性 | 看 $\lvert g'(x^*)\rvert<1$ | 反馈不能过强，否则振荡或发散 |
| 二维相图 | 画 $\dot{x}=0$ 和 $\dot{y}=0$ nullclines | 分析两个状态变量如何共同运动 |
| 非线性线性化 | 计算 $J_F(x^*)$ | 稳态附近的局部结构 |
| 连续时间多维稳定 | 特征值实部 $<0$ | 所有方向都收敛 |
| 离散时间多维稳定 | 特征值模长 $<1$ | 下一期反馈不把偏离放大 |
| 多重均衡 | 多个稳态和吸引域 | 初始条件和政策门槛重要 |
| 分岔 | 参数改变导致稳态数量或稳定性改变 | 小参数变化可能造成结构性变化 |

## 17 分层练习

### A 组：基础计算

1. 对系统 $\dot{x}=x(2-x)$，求所有稳态并判断稳定性。
2. 对系统 $\dot{x}=x(x-1)(2-x)$，画相线图并判断三个稳态的稳定性。
3. 对离散系统 $x_{t+1}=0.6x_t+0.2x_t^2$，求 $x^*=0$ 的局部稳定性。
4. 对离散系统 $x_{t+1}=rx_t(1-x_t)$，求非零稳态并给出局部稳定条件。

5. 给定二维系统
$$\dot{x} = x - y, \qquad \dot{y} = x^2 - y,$$
写出两条 nullcline。

## B 组：概念判断

6. “稳态存在”是否意味着“系统会收敛到稳态”？解释。
7. 为什么连续时间稳定性看特征值实部，而离散时间稳定性看特征值模长？
8. 什么情况下一阶线性化无法判断稳定性？举一个一维例子。
9. 什么是 basin of attraction？为什么它对政策分析重要？
10. Nullcline 的交点一定是稳态吗？稳态一定是两条 nullcline 的交点吗？在二维系统中说明。

## C 组：混合题

11. 考虑
$$\dot{x} = x(1-x) - y, \qquad \dot{y} = x - 2y.$$
求稳态，并对每个稳态用 Jacobian 判断局部稳定性。
12. 考虑 Solow 动态
$$\dot{k} = sk^{\alpha} - (n+\delta)k.$$
证明正稳态的收敛速度为 $(1-\alpha)(n+\delta)$，其中“速度”指线性化系数绝对值。
13. 对系统
$$\dot{x} = rx - x^3,$$
分析 $r<0, r=0, r>0$ 时稳态数量与稳定性。
14. 对离散系统
$$x_{t+1} = ax_t - bx_t^3, \quad b>0,$$
判断 0 的局部稳定性如何依赖 $a$。

## D 组：经济金融建模题

15. 设技术采用率 $x$ 满足
$$\dot{x} = x(1-x)(x-a).$$
解释为什么这可以表达网络外部性或协调失败。
16. 设资产价格偏离基本价值的程度 $q$ 满足
$$\dot{q} = \mu q - \eta q^3.$$
分析 $\mu$ 的变化如何影响泡沫型稳态。
17. 设债务动态为
$$\dot{b} = (r_0 + \phi b - g)b - s.$$
讨论 $\phi$ 增大如何改变债务风险。
18. 设财政乘数在衰退期和扩张期不同，请用一个 threshold difference equation 的形式表达这种状态依赖。

### E 组：挑战/证明题

19. 证明：若连续时间一维系统 $\dot{x} = f(x)$ 中 $f'(x^*) < 0$，则 $x^*$ 局部渐近稳定。可以用 Taylor 展开的直觉证明，不要求完整 Lyapunov 证明。
20. 对二维连续系统，说明为什么 $D < 0$ 意味着 saddle。提示：二维矩阵特征值乘积等于 determinant。
21. 构造一个有三个稳态的一维连续系统，其中中间稳态稳定，两边稳态不稳定。
22. 思考题：为什么在宏观 log-linearization 中，我们通常只能得到局部结论，而不能直接宣称全局稳定？

## 18 详细解答

### A 组解答

**1.** $\dot{x} = x(2-x)$。稳态满足

$$x(2-x) = 0,$$

所以 $x^* = 0, 2$。令 $f(x) = x(2-x) = 2x - x^2$，则

$$f'(x) = 2 - 2x.$$

$f'(0) = 2 > 0$，所以 0 不稳定；$f'(2) = -2 < 0$，所以 2 稳定。

**2.** $f(x) = x(x-1)(2-x)$，稳态为 $0, 1, 2$。取区间测试：$(0,1)$ 中，如 $x = 0.5$，$f < 0$；$(1,2)$ 中，如 $x = 1.5$，$f > 0$；$x > 2$ 时 $f < 0$。因此 0 稳定，1 不稳定，2 稳定。若考虑 $x < 0$，可自行判断箭头。

**3.** $g(x) = 0.6x + 0.2x^2$，$g'(x) = 0.6 + 0.4x$。在 0 处，$g'(0) = 0.6$，满足 $|0.6| < 1$，所以 0 局部稳定。

**4.** 非零稳态满足 $x = rx(1-x)$。若 $x \neq 0$，则 $1 = r(1-x)$，所以 $x^* = 1 - 1/r$。导数 $g'(x) = r(1-2x)$，代入得 $g'(x^*) = 2 - r$。稳定条件为 $|2-r| < 1$，即 $1 < r < 3$。

**5.** $\dot{x} = 0$ 给 $x - y = 0$，所以 $y = x$。$\dot{y} = 0$ 给 $x^2 - y = 0$，所以 $y = x^2$。两条 nullcline 的交点满足 $x = x^2$，即 $x = 0$ 或 $x = 1$，对应稳态 $(0,0)$ 与 $(1,1)$。

### B 组解答

**6.** 不是。稳态只是让变化率为零的点，可能稳定也可能不稳定。例如 $\dot{x} = x$ 有稳态 0，但任何正的初始偏离都会发散。

**7.** 连续时间线性解含有 $e^{\lambda t}$，因此看 $\mathrm{Re}(\lambda)$ 是否为负。离散时间线性解含有 $\lambda^t$，因此看 $|\lambda|$ 是否小于 1。

**8.** 当 $f'(x^*) = 0$ 时，一阶线性化无法判断。例如 $\dot{x} = -x^3$ 在 0 处有 $f'(0) = 0$，但它稳定；$\dot{x} = x^3$ 在 0 处也有 $f'(0) = 0$，但它不稳定。

**9.** Basin of attraction 是所有会收敛到某个稳态的初始条件集合。政策若能把经济状态推出低均衡的吸引域，就可能改变长期路径；若没有跨过分界线，长期可能回到原来的低水平均衡。

**10.** 在二维系统中，nullcline 是 $\dot{x} = 0$ 与 $\dot{y} = 0$ 的点集。两条 nullcline 的交点同时满足两个变化率为零，因此是稳态。反过来，任何稳态都必须同时满足 $\dot{x} = 0$ 和 $\dot{y} = 0$，因此一定在两条 nullcline 的交点上。

## C 组解答

**11.** 系统为

$$\dot{x} = x(1-x) - y, \qquad \dot{y} = x - 2y.$$

稳态满足

$$y = x(1-x), \qquad y = \frac{x}{2}.$$

所以

$$x(1-x) = \frac{x}{2}.$$

即

$$x\left(1 - x - \frac{1}{2}\right) = 0,$$

得到 $x = 0$ 或 $x = 1/2$。对应稳态为

$$(0,0), \qquad \left(\frac{1}{2}, \frac{1}{4}\right).$$

Jacobian：

$$J(x,y) = \begin{pmatrix} 1-2x & -1 \\ 1 & -2 \end{pmatrix}.$$

在 $(0,0)$ 处：

$$J_0 = \begin{pmatrix} 1 & -1 \\ 1 & -2 \end{pmatrix}, \quad T = -1, \quad D = -1 < 0.$$

所以是 saddle。在 $(1/2, 1/4)$ 处：

$$J_1 = \begin{pmatrix} 0 & -1 \\ 1 & -2 \end{pmatrix}, \quad T = -2, \quad D = 1 > 0.$$

$T < 0, D > 0$，局部稳定。判别式 $T^2 - 4D = 4 - 4 = 0$，是临界情形，重复特征值 $-1$。

**12.** 正稳态满足 $s(k^*)^{\alpha-1} = n + \delta$。线性化系数为

$$f'(k^*) = s\alpha(k^*)^{\alpha-1} - (n+\delta) = \alpha(n+\delta) - (n+\delta) = -(1-\alpha)(n+\delta).$$

连续时间中局部收敛速度通常取该负系数的绝对值：

$$(1-\alpha)(n+\delta).$$

**13.** 对 $\dot{x} = rx - x^3 = x(r - x^2)$。若 $r < 0$，只有 $x = 0$ 一个稳态，$f'(0) = r < 0$，稳定。若 $r = 0$，$\dot{x} = -x^3$，0 稳定但一阶线性化失效。若 $r > 0$，稳态为

$$0, \quad \sqrt{r}, \quad -\sqrt{r}.$$

导数 $f'(x) = r - 3x^2$。$f'(0) = r > 0$，所以 0 不稳定；$f'(\pm\sqrt{r}) = r - 3r = -2r < 0$，所以 $\pm\sqrt{r}$ 稳定。这是 pitchfork bifurcation 的基本形式。

**14.** $g(x) = ax - bx^3$，$g'(x) = a - 3bx^2$。在 0 处，$g'(0) = a$。离散时间局部稳定条件为

$$|a| < 1.$$

参数 $b$ 不影响 0 的局部一阶稳定性，但影响远离零时的非线性动态。

### D 组解答

**15.** 当技术采用率低于门槛时，采用者太少，互补性不足，收益低，采用率下降；高于门槛时，网络外部性增强，采用收益上升，采用率继续上升。因此 $a$ 表示协调门槛。低稳态和高稳态分别对应失败协调和成功协调。

**16.** $\dot{q} = \mu q - \eta q^3 = q(\mu - \eta q^2)$。若 $\mu < 0$，只有 $q = 0$ 稳定。若 $\mu > 0$，出现 $q = \pm\sqrt{\mu/\eta}$ 两个非零稳态，$0$ 变为不稳定。这可以解释当投机反馈强度 $\mu$ 足够大时，市场可能出现正负方向的泡沫偏离。

**17.** $\phi$ 越大，债务对利率的反馈越强。数学上，$\dot{b}$ 中的二次项 $\phi b^2$ 更强，意味着高债务区间中债务增长速度更快，阈值附近更脆弱。经济上，这是债务风险溢价和可持续性之间的非线性反馈。

**18.** 可以写成

$$y_{t+1} = \begin{cases} \rho_R y_t + m_R G_t + \varepsilon_{t+1}, & y_t < c, \\ \rho_E y_t + m_E G_t + \varepsilon_{t+1}, & y_t \geq c, \end{cases}$$

其中 $y_t$ 是产出缺口，$G_t$ 是财政支出冲击，$m_R$ 是衰退期乘数，$m_E$ 是扩张期乘数。若 $m_R > m_E$，表示衰退期财政政策更有效。

### E 组解答

**19.** 设 $f(x^*) = 0$ 且 $f'(x^*) < 0$。由 Taylor 展开，

$$f(x) = f'(x^*)(x - x^*) + o(|x - x^*|).$$

当 $x$ 足够接近 $x^*$ 时，$f(x)$ 与 $(x - x^*)$ 符号相反：如果 $x > x^*$，则 $f(x) < 0$，轨迹向左；如果 $x < x^*$，则 $f(x) > 0$，轨迹向右。因此两侧箭头都指向 $x^*$，局部渐近稳定。

**20.** 二维矩阵的两个特征值乘积等于 determinant。若 $D < 0$，则两个特征值乘积为负，说明它们异号。连续时间中，一个正特征值方向发散，一个负特征值方向收敛，因此稳态是 saddle。

**21.** 一个例子是

$$\dot{x} = -(x + 1)x(x - 1).$$

稳态为 $-1, 0, 1$。检查符号可知 $x = 0$ 稳定，而 $x = -1$ 与 $x = 1$ 不稳定。也可以用导数法：$f(x) = -(x^3 - x) = -x^3 + x$，$f'(x) = -3x^2 + 1$。$f'(0) = 1 > 0$，这说明该构造其实让 $0$ 不稳定；所以需要调整为

$$\dot{x} = x(x + 1)(x - 1) = x^3 - x.$$

此时 $f'(0) = -1 < 0$，$0$ 稳定；$f'(\pm 1) = 2 > 0$，两边不稳定。

**22.** Log-linearization 是在稳态附近的一阶近似。它告诉我们足够小的偏离如何演化，但忽略了高阶项。若远离稳态，高阶项可能主导动态；若存在多个稳态，局部线性化只能描述一个稳态的吸引域附近；若参数变化导致分岔，原来的局部结构可能失效。因此宏观 log-linear 模型通常给出局部结论，而不是全局稳定证明。

## 19 本章总结图谱

| 概念 | 公式 | 应用 | 常见错误 |
|---|---|---|---|
| 连续稳态 | $F(x^*) = 0$ | 资本、债务、通胀稳态 | 以为存在就稳定 |
| 离散稳态 | $x^* = G(x^*)$ | AR、资产价格递推 | 和连续时间条件混用 |
| 一维连续稳定 | $f'(x^*) < 0$ | Solow 收敛 | 忘记 $f' = 0$ 时失效 |
| 一维离散稳定 | $\lvert g'(x^*)\rvert < 1$ | Logistic map、AR(1) | 只看正负不看模长 |

| | | | |
|---|---|---|---|
| Nullcline | $\dot{x}=0,\dot{y}=0$ | Ramsey 相图 | 把 nullcline 当轨迹 |
| Jacobian 线性化 | $\dot{y}\approx J_F(x^*)y$ | 局部稳定分析 | 从局部推出全局 |
| 多重均衡 | 多个 $f(x)=0$ 根 | 贫困陷阱、协调失败 | 忽视初始条件 |
| Bifurcation | 参数改变结构 | 危机阈值、政策临界点 | 当作普通比较静态 |

**直觉框**

本章最重要的一句话：非线性系统通常不能直接求完整解，但可以先找稳态，再在稳态附近线性化，再用特征值判断局部稳定，最后用相图、吸引域和分岔直觉理解全局结构。这套方法会在高级宏观、金融经济学、计量时间序列和动态优化中不断出现。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 14
# 多元优化与 Hessian

Chapter 14 是本讲义从动态系统转入经济理论核心工具的起点。经济学里几乎所有基础模型都可以被写成优化问题：消费者选择效用最大化，厂商选择利润最大化，成本函数来自成本最小化，投资者选择投资组合，计量中的 OLS 是平方损失最小化，机器学习里的估计也是目标函数优化。本章只讨论最基本但最关键的一类：**无约束多元优化**。你要掌握两件事：第一，梯度为零是什么意思；第二，Hessian 的正定、负定、不定为什么决定局部最小、局部最大或鞍点。读完本章，你不应只是会算偏导，而应该能把一个经济金融目标函数翻译成一阶边际条件和二阶曲率条件，并解释这些条件背后的经济含义。

math4econ personalized notes
2026

# Contents

**Chapter 14 总览** **3**

**1 本章为什么重要** **3**

1.1 优化是经济模型的基本语法 . . . 3

1.2 一阶条件不是最优条件的全部 . . . 4

1.3 Hessian 是经济学里的曲率语言 . . . 4

**2 最小前置知识** **4**

2.1 一元函数的最优化回顾 . . . 4

2.2 梯度与方向导数 . . . 5

2.3 二阶 Taylor 展开 . . . 5

**3 多元无约束优化的形式** **6**

3.1 最大化与最小化问题 . . . 6

3.2 驻点 critical point . . . 6

3.3 边界问题暂时不处理 . . . 6

**4 一阶必要条件：为什么梯度必须为零** **6**

**5 Hessian 与二阶 Taylor 展开** **7**

5.1 Hessian 的定义 . . . 7

5.2 Hessian 与二次型 . . . 7

**6 正定、负定、半定与不定** **8**

6.1 定义 . . . 8

6.2 二阶条件 . . . 8

6.3 半定时为什么麻烦 . . . 9

**7 如何判断正定与负定** **9**

7.1 方法一：看特征值 . . . 9

7.2 方法二：二维 Hessian 的行列式判别 . . . 9

7.3 方法三：Sylvester 判别法 . . . 10

**8 Worked Examples：从计算到解释** **11**

**9 经济金融应用模块** **13**

9.1 消费者效用最大化的无约束版本 . . . 13

9.2 生产函数的边际与曲率 . . . 14

9.3 估计理论中的曲率：为什么 Hessian 影响标准误 . . . 14

**10 常见错误与纠正** **14**

**11 本章小结图谱** **15**

**12 练习题** **15**

**13 练习题详细答案** **17**

**本章最后提醒** **21**

# Chapter 14 总览

从 Chapter 5 开始，我们已经把多元函数看作一个局部对象：梯度告诉你最陡上升方向，全微分告诉你局部线性近似，Hessian 则告诉你二阶弯曲。本章把这些工具放进优化问题。

一个无约束多元优化问题的形式是

$$\max_{x\in\mathbb{R}^n} f(x) \quad 或 \quad \min_{x\in\mathbb{R}^n} f(x).$$

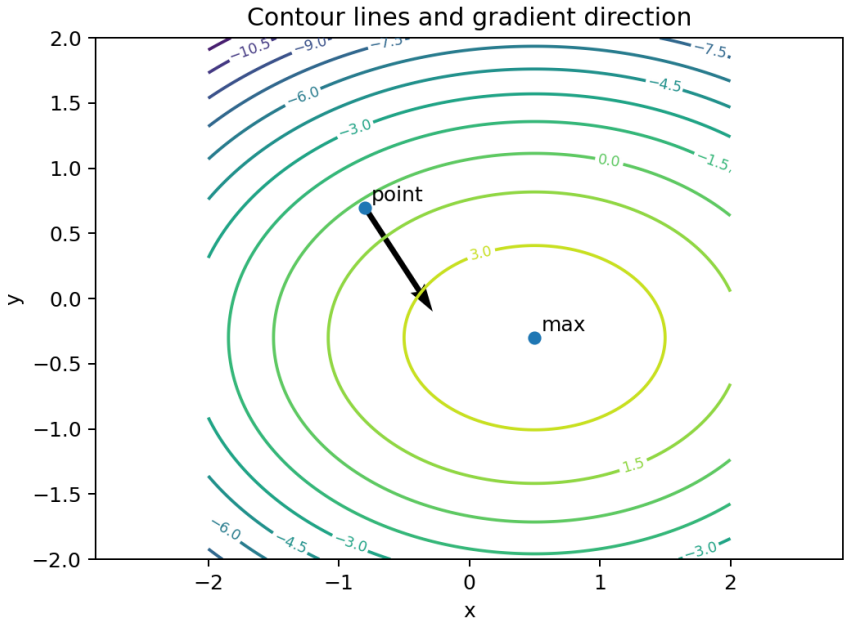
其中 $x=(x_1,\dots,x_n)'$ 是选择变量，$f$ 是目标函数。经济学上，$x$ 可能是消费束、投入向量、投资组合权重、政策变量或参数估计值。

本章主线如下：

$$\text{目标函数} \to \text{梯度} \to \text{一阶条件} \to \text{Hessian} \to \text{二阶条件} \to \text{经济解释}.$$

**直觉框**

一阶条件回答：在最优点附近，朝任何微小方向移动，目标函数的一阶变化是否已经消失？
二阶条件回答：如果一阶变化消失，目标函数在该点附近是像碗一样向上弯，像帽子一样向下弯，还是一个方向上升、另一个方向下降的马鞍？




图中等高线越靠近高处，目标函数越大。梯度垂直于等高线，指向局部最快上升方向。若某点是内部最优点，梯度必须消失；否则沿梯度方向移动就能继续提高目标函数。

## 1 本章为什么重要

### 1.1 优化是经济模型的基本语法

经济学之所以大量使用数学，不是为了装饰模型，而是因为经济行为通常被建模为某种选择：

- 消费者选择 $x$ 最大化 $u(x)$；

- 厂商选择投入 $(K, L)$ 最大化利润 $\pi(K, L)$；
- 政府选择税率 $\tau$ 最大化社会福利 $W(\tau)$；
- 投资者选择权重 $w$ 在收益和风险之间权衡；
- 计量经济学中选择参数 $\hat{\beta}$ 最小化样本平方误差。

只要模型里出现 “choice”，背后就有优化结构。

### 1.2 一阶条件不是最优条件的全部

很多初学者学优化时最容易犯的错误是：看到 $f_x = 0, f_y = 0$ 就宣布找到了最大值或最小值。但梯度为零只说明**局部一阶平坦**，并不说明是最大、最小还是鞍点。

例如

$$f(x, y) = x^2 - y^2.$$

在 $(0, 0)$，有

$$\nabla f(0, 0) = (0, 0),$$

但沿 $x$ 轴方向，$f(x, 0) = x^2 > 0$；沿 $y$ 轴方向，$f(0, y) = -y^2 < 0$。所以 $(0, 0)$ 既不是局部最大，也不是局部最小，而是鞍点。

**常见误区**

在经济学里，FOC 只是候选条件。真正判断最优，还要看二阶条件、边界条件、全局凹性/凸性，以及约束是否存在。本章先处理无约束内部点的二阶条件；等式约束和不等式约束会在 Chapter 15 与 Chapter 16 中处理。

### 1.3 Hessian 是经济学里的曲率语言

Hessian 矩阵记录所有二阶偏导：

$$H_f(x) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

在经济学中，它对应以下概念：

- 消费理论中的边际效用递减和交叉边际效应；
- 生产理论中的边际产出递减和要素互补/替代；
- 成本函数和利润函数的曲率；
- 投资组合方差 $w'\Sigma w$ 的风险曲率；
- 估计问题中目标函数的局部曲率与标准误。

## 2 最小前置知识

### 2.1 一元函数的最优化回顾

一元函数 $f : \mathbb{R} \to \mathbb{R}$ 的内部局部最优点 $x^*$ 通常满足

$$f'(x^*) = 0.$$

如果

$$f''(x^*) < 0,$$

则 $x^*$ 是局部最大点；如果

$$f''(x^*) > 0,$$

则 $x^*$ 是局部最小点；如果 $f''(x^*) = 0$，二阶判别失败，需要更高阶或其他方法。

多元优化是这个逻辑的推广：

$$f'(x^*) = 0 \quad \rightsquigarrow \quad \nabla f(x^*) = 0,$$

$$f''(x^*) \quad \rightsquigarrow \quad H_f(x^*).$$

区别是：一元函数的二阶导只是一个数，多元函数的 Hessian 是一个矩阵。矩阵没有简单的“正负”概念，因此我们需要正定、负定、不定。

## 2.2 梯度与方向导数

若 $f : \mathbb{R}^n \to \mathbb{R}$ 可微，则梯度是

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(x) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

沿单位方向 $v$ 的方向导数是

$$D_v f(x) = \nabla f(x)' v.$$

这句话非常重要：梯度不仅是偏导数组成的向量，它把任意方向的局部变化率都统一表示出来。

**直觉框**

如果 $\nabla f(x) \neq 0$，那么选择 $v = \nabla f(x)/\|\nabla f(x)\|$，就有

$$D_v f(x) = \|\nabla f(x)\| > 0.$$

这说明沿梯度方向移动可以提高 $f$。所以内部局部最大点不可能有非零梯度。

## 2.3 二阶 Taylor 展开

若 $f$ 二阶连续可微，则在 $x^*$ 附近，令 $h = x - x^*$，有

$$f(x^* + h) = f(x^*) + \nabla f(x^*)' h + \frac{1}{2} h' H_f(x^*) h + o(\|h\|^2).$$

如果 $x^*$ 是一阶驻点，即 $\nabla f(x^*) = 0$，则

$$f(x^* + h) - f(x^*) = \frac{1}{2} h' H_f(x^*) h + o(\|h\|^2).$$

因此在驻点附近，函数值上升还是下降，主要由二次型 $h' H_f(x^*) h$ 决定。

# 3 多元无约束优化的形式

## 3.1 最大化与最小化问题

无约束最大化问题写作

$$\max_{x\in\mathbb{R}^n} f(x).$$

若 $x^*$ 满足某个邻域 $B_\epsilon(x^*)$ 内

$$f(x^*)\geq f(x),\qquad \forall x\in B_\epsilon(x^*),$$

则 $x^*$ 是局部最大点。如果对所有 $x\in\mathbb{R}^n$ 都成立，则是全局最大点。

无约束最小化问题写作

$$\min_{x\in\mathbb{R}^n} f(x).$$

若在某个邻域内

$$f(x^*)\leq f(x),$$

则 $x^*$ 是局部最小点。

**常见误区**

局部最优和全局最优不是一回事。经济学里很多模型只通过 FOC 找局部最优，只有在目标函数凹/凸、可行集凸等额外条件下，局部最优才可以升级为全局最优。凸性与全局性会在 Chapter 17 系统讨论。

## 3.2 驻点 critical point

如果

$$\nabla f(x^*)=0,$$

则 $x^*$ 称为驻点或临界点。驻点只是候选点，包括局部最大、局部最小和鞍点。

## 3.3 边界问题暂时不处理

本章讨论无约束内部点。如果经济问题有预算约束、资源约束、非负约束，则不能直接套本章的无约束条件。例如消费者问题

$$\max_{x,y} u(x,y)\quad \text{s.t.}\quad p_x x+p_y y=w$$

不是本章问题，而是 Chapter 15 的 Lagrange 乘子问题。若还有 $x\geq 0, y\geq 0$，则进入 Chapter 16 的 KKT 问题。

# 4 一阶必要条件：为什么梯度必须为零

**命题 14.1：内部局部最优的一阶必要条件**

设 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $x^*$ 处可微，且 $x^*$ 是内部局部最大点或局部最小点。则

$$\nabla f(x^*)=0.$$

**证明。**任取方向 $v \in \mathbb{R}^n$，考虑一元函数

$$\phi(t) = f(x^* + tv).$$

因为 $x^*$ 是 $f$ 的局部最优点，所以 $t = 0$ 是 $\phi$ 的一元局部最优点。因此一元微积分给出

$$\phi'(0) = 0.$$

由链式法则，

$$\phi'(0) = \nabla f(x^*)'v.$$

于是对任意 $v$ 都有

$$\nabla f(x^*)'v = 0.$$

取 $v = \nabla f(x^*)$，得到

$$\|\nabla f(x^*)\|^2 = 0.$$

所以 $\nabla f(x^*) = 0$。证毕。

这个证明很值得记住，因为它说明了多元优化的本质：局部最优意味着沿任何一条穿过该点的直线看，都是一元局部最优；因此所有方向导数必须为零。

**方法框**

求无约束内部候选点的标准步骤：

1. 写出目标函数 $f(x_1, \dots, x_n)$；
2. 计算梯度 $\nabla f$；
3. 解方程组 $\nabla f = 0$；
4. 得到驻点集合；
5. 用 Hessian 和其他信息分类。

# 5 Hessian 与二阶 Taylor 展开

## 5.1 Hessian 的定义

若 $f : \mathbb{R}^n \to \mathbb{R}$ 二阶可微，则 Hessian 矩阵是

$$H_f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right]_{i,j=1}^{n}.$$

如果二阶偏导连续，则 Schwarz 定理给出

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i},$$

所以 Hessian 是对称矩阵。

## 5.2 Hessian 与二次型

在驻点 $x^*$ 附近，二阶 Taylor 展开给出

$$f(x^* + h) - f(x^*) \approx \frac{1}{2} h' H_f(x^*) h.$$

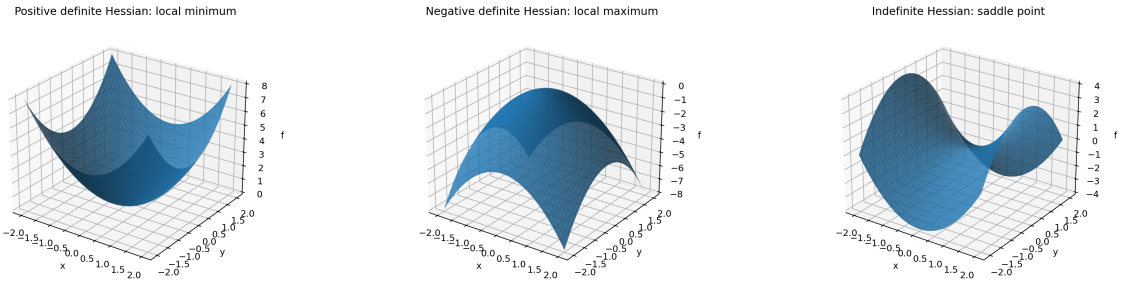
这里

$$h' H_f(x^*) h$$

称为由 Hessian 生成的二次型。

**直觉框**

Hessian 本身是一个矩阵，但真正决定函数在某个方向 $h$ 上弯曲的是二次型 $h'Hh$。如果所有非零方向 $h$ 上 $h'Hh > 0$，函数像一个碗；如果所有方向上 $h'Hh < 0$，函数像一顶帽子；如果有的方向正、有的方向负，函数像马鞍。




# 6 正定、负定、半定与不定

## 6.1 定义

设 $A$ 是 $n \times n$ 对称矩阵。

- 若对所有 $h \neq 0$，有 $h'Ah > 0$，则 $A$ 正定 positive definite；
- 若对所有 $h \neq 0$，有 $h'Ah < 0$，则 $A$ 负定 negative definite；
- 若对所有 $h$，有 $h'Ah \geq 0$，则 $A$ 半正定 positive semidefinite；
- 若对所有 $h$，有 $h'Ah \leq 0$，则 $A$ 半负定 negative semidefinite；
- 若存在 $h_1, h_2$，使 $h_1'Ah_1 > 0$ 且 $h_2'Ah_2 < 0$，则 $A$ 不定 indefinite。

## 6.2 二阶条件

**命题 14.2：二阶充分条件**

设 $f$ 在 $x^*$ 附近二阶连续可微，且 $\nabla f(x^*) = 0$。

1. 若 $H_f(x^*)$ 负定，则 $x^*$ 是严格局部最大点；
2. 若 $H_f(x^*)$ 正定，则 $x^*$ 是严格局部最小点；
3. 若 $H_f(x^*)$ 不定，则 $x^*$ 是鞍点，不是局部最优点。

**证明思路。**因为

$$f(x^* + h) - f(x^*) = \frac{1}{2} h' H_f(x^*) h + o(\|h\|^2).$$

如果 Hessian 负定，则存在 $c > 0$，使得足够小的 $h \neq 0$ 满足

$$h' H_f(x^*) h \leq -c\|h\|^2.$$

高阶项 $o(\|h\|^2)$ 相对于 $\|h\|^2$ 可以忽略，所以

$$f(x^* + h) - f(x^*) < 0$$

在足够小邻域内成立，故 $x^*$ 是严格局部最大点。正定情形同理给出严格局部最小点。不定时，存在两个方向使二次项一正一负，沿不同方向函数值分别上升和下降，所以是鞍点。证毕。

### 6.3 半定时为什么麻烦

如果 Hessian 半正定或半负定，二阶条件通常不能直接给出严格结论。例如

$$f(x,y) = x^4 + y^4$$

在 $(0,0)$ 的 Hessian 是零矩阵，既半正定也半负定，但 $(0,0)$ 是严格局部最小点。另一个函数

$$g(x,y) = x^4 - y^4$$

在 $(0,0)$ 的 Hessian 也是零矩阵，但 $(0,0)$ 是鞍点。因此 Hessian 半定或零矩阵时，需要高阶项或其他结构判断。

**常见误区**

考试和模型推导中，看到 Hessian 的行列式等于零，或者特征值有零，不要草率说“最大”或“最小”。这通常意味着二阶判别失败，而不是得到半个结论。

## 7 如何判断正定与负定

### 7.1 方法一：看特征值

对称矩阵 $A$ 可以正交对角化：

$$A = Q\Lambda Q',$$

其中 $Q'Q = I$，$\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$。令 $z = Q'h$，则

$$h'Ah = h'Q\Lambda Q'h = z'\Lambda z = \sum_{i=1}^{n} \lambda_i z_i^2.$$

因此：

- 所有特征值正 $\Rightarrow$ $A$ 正定；
- 所有特征值负 $\Rightarrow$ $A$ 负定；
- 有正有负 $\Rightarrow$ $A$ 不定；
- 所有特征值非负且至少一个为零 $\Rightarrow$ $A$ 半正定；
- 所有特征值非正且至少一个为零 $\Rightarrow$ $A$ 半负定。

### 7.2 方法二：二维 Hessian 的行列式判别

二维情形最常见。设

$$H = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

则

$$\det(H) = ac - b^2.$$

在驻点处：

- 若 $\det(H) > 0$ 且 $a < 0$，则 $H$ 负定，局部最大；
- 若 $\det(H) > 0$ 且 $a > 0$，则 $H$ 正定，局部最小；
- 若 $\det(H) < 0$，则 $H$ 不定，鞍点；
- 若 $\det(H) = 0$，二阶判别失败。

**方法框**

二维题最实用口诀：先看 $D = f_{xx}f_{yy} - f_{xy}^2$。若 $D < 0$，鞍点；若 $D > 0$，再看 $f_{xx}$：$f_{xx} < 0$ 是局部最大，$f_{xx} > 0$ 是局部最小；若 $D = 0$，不能判别。

### 7.3 方法三：Sylvester 判别法

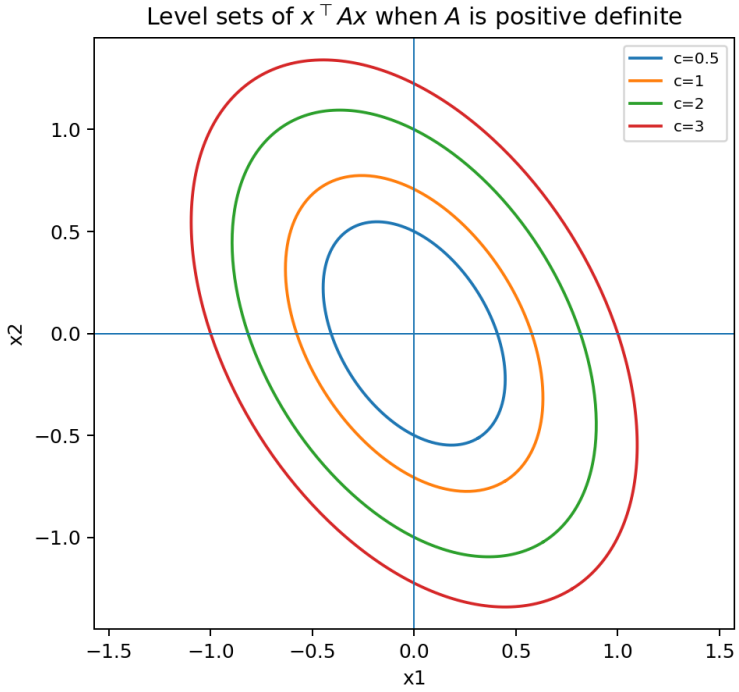
对于 $n \times n$ 对称矩阵 $A$，记前 $k$ 阶顺序主子式为

$$\Delta_k = \det(A_{1:k,1:k}).$$

则：

- $A$ 正定当且仅当 $\Delta_1 > 0, \Delta_2 > 0, \ldots, \Delta_n > 0$；
- $A$ 负定当且仅当 $\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \ldots, (-1)^k\Delta_k > 0$。

在经济学中，二维和三维问题经常用这个方法，尤其是 Hessian 或 bordered Hessian 判别。




正定二次型的等值线是椭圆。这就是为什么正定 Hessian 对应局部最小：离开中心时，二次型在任何方向都增加。

# 8 Worked Examples：从计算到解释

**例 1：分类三个驻点**

考虑

$$f(x,y)=x^3-3x+y^2.$$

求驻点并分类。

**解。**一阶偏导为

$$f_x=3x^2-3,\qquad f_y=2y.$$

令 $f_x=f_y=0$，得到

$$x^2=1,\qquad y=0.$$

所以驻点是

$$(1,0),\qquad (-1,0).$$

Hessian 为

$$H_f(x,y)=\begin{pmatrix}6x & 0\\ 0 & 2\end{pmatrix}.$$

在 $(1,0)$，

$$H_f(1,0)=\begin{pmatrix}6 & 0\\ 0 & 2\end{pmatrix},$$

正定，所以 $(1,0)$ 是严格局部最小点。

在 $(-1,0)$，

$$H_f(-1,0)=\begin{pmatrix}-6 & 0\\ 0 & 2\end{pmatrix},$$

一个方向负，一个方向正，不定，所以 $(-1,0)$ 是鞍点。

**经济解释。**驻点只是边际收益为零的点。只有当所有方向的二阶变化都支持同一种弯曲方向时，才能说是局部最大或最小；否则某些方向改善、某些方向恶化，经济上就不是稳定最优选择。

**例 2：利润最大化中的 Hessian**

设厂商选择资本 $K$ 和劳动 $L$，利润为

$$\pi(K,L)=pK^{\alpha}L^{\beta}-rK-wL,$$

其中 $0<\alpha,\beta<1$ 且 $\alpha+\beta<1$。求内部一阶条件，并解释二阶条件。

**解。**一阶条件是

$$\pi_K=p\alpha K^{\alpha-1}L^{\beta}-r=0,$$
$$\pi_L=p\beta K^{\alpha}L^{\beta-1}-w=0.$$

也就是

$$p\cdot MP_K=r,\qquad p\cdot MP_L=w.$$

经济含义是：每种投入的边际收益产品等于其要素价格。

Hessian 为

$$H_\pi=\begin{pmatrix}p\alpha(\alpha-1)K^{\alpha-2}L^{\beta} & p\alpha\beta K^{\alpha-1}L^{\beta-1}\\ p\alpha\beta K^{\alpha-1}L^{\beta-1} & p\beta(\beta-1)K^{\alpha}L^{\beta-2}\end{pmatrix}.$$

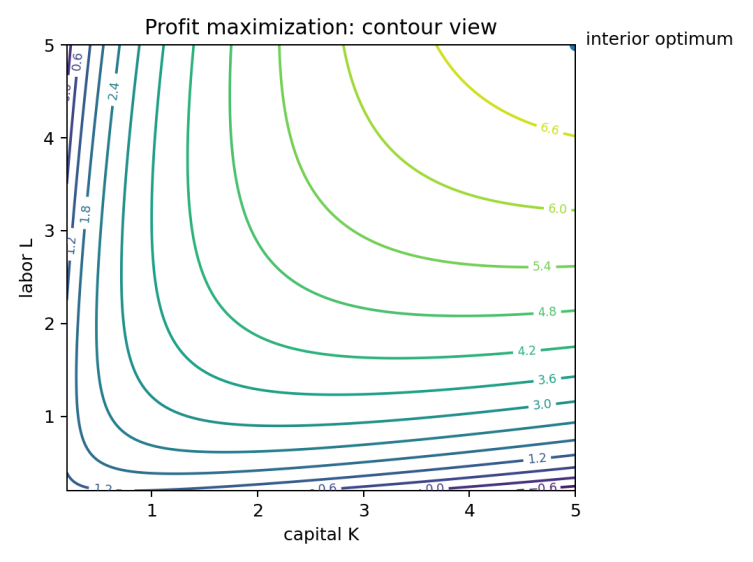
因为 $\alpha-1<0$ 且 $\beta-1<0$，对角线元素为负。再看行列式：

$$\det(H_\pi)=p^2K^{2\alpha-2}L^{2\beta-2}\alpha\beta[(\alpha-1)(\beta-1)-\alpha\beta].$$

括号内

$$(\alpha-1)(\beta-1)-\alpha\beta=\alpha\beta-\alpha-\beta+1-\alpha\beta=1-\alpha-\beta>0.$$

所以 $\det(H_\pi)>0$ 且第一个对角元素为负，Hessian 负定。利润函数局部严格凹，内部驻点是严格局部最大点。




**关键理解。**一阶条件说“边际收益产品等于投入价格”，二阶条件说“这个边际条件给出的点确实是利润峰顶，而不是谷底或鞍点”。

## 例 3：OLS 作为多元最小化

给定 $y\in\mathbb{R}^n$，$X\in\mathbb{R}^{n\times k}$，OLS 选择 $\beta\in\mathbb{R}^k$ 最小化

$$S(\beta)=(y-X\beta)'(y-X\beta).$$

求一阶条件与 Hessian。

**解。**展开：

$$S(\beta)=y'y-2\beta'X'y+\beta'X'X\beta.$$

梯度为

$$\nabla_\beta S(\beta)=-2X'y+2X'X\beta.$$

一阶条件：

$$X'X\hat\beta=X'y.$$

如果 $X'X$ 可逆，则

$$\hat\beta=(X'X)^{-1}X'y.$$

Hessian 是

$$H_S=2X'X.$$

若 $X$ 满列秩，则 $X'X$ 正定，$H_S$ 正定，所以 $S(\beta)$ 是严格凸的，OLS 解是唯一全局最小点。

**计量解释。** OLS 的一阶条件是残差与解释变量正交：

$$X'(y-X\hat\beta)=0.$$

Hessian 正定说明平方损失曲面像一个碗，所以这个正交投影点不是局部假象，而是唯一最小误差点。

**例 4：均值-方差投资组合优化的无约束形式**

设投资者选择权重 $w \in \mathbb{R}^n$，目标函数为

$$U(w) = \mu' w - \frac{\gamma}{2} w' \Sigma w,$$

其中 $\mu$ 是期望收益向量，$\Sigma$ 是协方差矩阵，$\gamma > 0$ 是风险厌恶系数。暂时忽略权重和为 1 的约束。求最优 $w$。

**解。**梯度为

$$\nabla_w U(w) = \mu - \gamma \Sigma w.$$

一阶条件：

$$\mu - \gamma \Sigma w = 0.$$

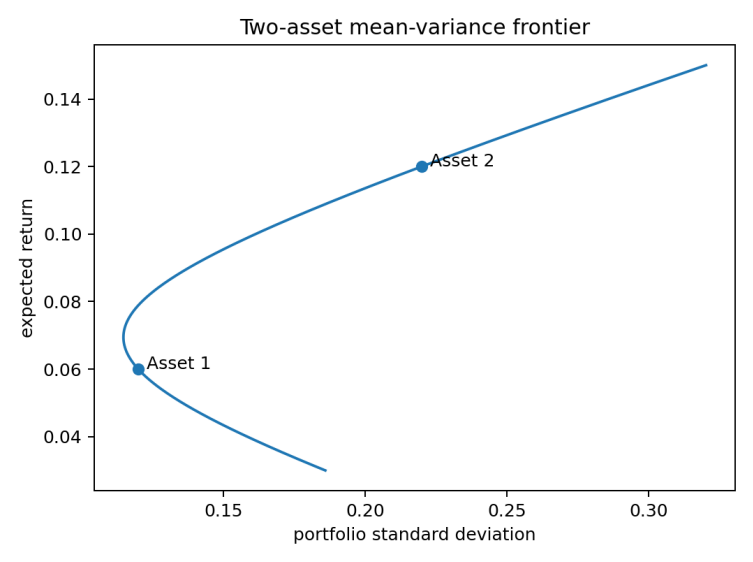
若 $\Sigma$ 正定，则

$$w^* = \frac{1}{\gamma} \Sigma^{-1} \mu.$$

Hessian 为

$$H_U = -\gamma \Sigma.$$

由于 $\Sigma$ 正定且 $\gamma > 0$，$H_U$ 负定，所以 $w^*$ 是严格局部最大点，事实上也是全局最大点。




**金融解释。** $\Sigma^{-1}\mu$ 是把收益信号按风险结构调整后的方向。风险厌恶 $\gamma$ 越大，仓位整体越小。Hessian 负定来自风险惩罚项 $-\frac{\gamma}{2} w' \Sigma w$ 的严格凹性。

# 9 经济金融应用模块

## 9.1 消费者效用最大化的无约束版本

严格地说，消费者问题通常有预算约束。但很多时候我们会先研究效用函数本身的曲率。例如

$$u(x, y) = \ln x + \ln y.$$

梯度为

$$\nabla u(x, y) = \begin{pmatrix} 1/x \\ 1/y \end{pmatrix},$$

Hessian 为

$$H_u(x,y)=\begin{pmatrix}-1/x^2 & 0\\ 0 & -1/y^2\end{pmatrix}.$$

在 $x,y>0$ 上 Hessian 负定，所以效用函数严格凹。这说明边际效用递减，并且在凸预算集上会带来唯一需求。真正的约束版本会在下一章处理。

### 9.2 生产函数的边际与曲率

对 Cobb-Douglas 生产函数

$$F(K,L)=AK^{\alpha}L^{\beta},$$

一阶偏导给出边际产出：

$$F_K=A\alpha K^{\alpha-1}L^{\beta},\qquad F_L=A\beta K^{\alpha}L^{\beta-1}.$$

二阶偏导给出曲率和交叉效应：

$$F_{KK}=A\alpha(\alpha-1)K^{\alpha-2}L^{\beta}<0,$$

$$F_{LL}=A\beta(\beta-1)K^{\alpha}L^{\beta-2}<0,$$

$$F_{KL}=A\alpha\beta K^{\alpha-1}L^{\beta-1}>0.$$

其中 $F_{KK}<0$ 和 $F_{LL}<0$ 表示单个要素边际产出递减；$F_{KL}>0$ 表示资本越多，劳动的边际产出越高，反之亦然。

### 9.3 估计理论中的曲率：为什么 Hessian 影响标准误

在最大似然估计中，我们最大化 log likelihood

$$\ell(\theta)=\sum_{i=1}^{n}\log f(y_i\mid\theta).$$

一阶条件是 score 为零：

$$\nabla_\theta\ell(\hat\theta)=0.$$

Hessian

$$H_\ell(\hat\theta)$$

描述 log likelihood 在估计点附近有多尖。如果峰很尖，小幅偏离就导致目标函数明显下降，估计更精确；如果峰很平，很多参数值都差不多，估计不精确。这就是为什么渐近方差与信息矩阵，即负 Hessian 的期望，有密切关系。

## 10 常见错误与纠正

**误区 1：把 FOC 当成最优解**

错误：解出 $\nabla f=0$ 就直接说最大值或最小值。
纠正：$\nabla f=0$ 只给候选点。还要用 Hessian 分类，或利用全局凹/凸性，或检查边界和约束。

**误区 2：只看 $f_{xx}$ 和 $f_{yy}$**

二维函数中，即使 $f_{xx}<0$ 且 $f_{yy}<0$，也不一定是局部最大，还要看交叉项 $f_{xy}$ 是否过大。正确判别需要

$$D=f_{xx}f_{yy}-f_{xy}^2.$$

如果 $D<0$，就是鞍点。

**误区 3：混淆最大化和最小化的 Hessian 符号**

最大化需要 Hessian 负定；最小化需要 Hessian 正定。因为最大点像帽子，最小点像碗。

**误区 4：忽视经济问题里的约束**

很多经济问题表面上是选择变量，其实有预算、资源、非负、借贷、权重和为 1 等约束。本章只适用于无约束内部点。后面两章会处理约束。

## 11 本章小结图谱

| 概念 | 数学表达 | 经济金融含义 |
|---|---|---|
| 目标函数 | $f(x)$ | 效用、利润、损失、福利、风险调整收益 |
| 梯度 | $\nabla f(x)$ | 各选择变量的边际效应向量 |
| 一阶条件 | $\nabla f(x^*)=0$ | 所有方向的一阶改善机会消失 |
| Hessian | $H_f(x^*)$ | 二阶曲率、边际效应如何变化 |
| 正定 | $h'Hh>0$ | 碗形，局部最小 |
| 负定 | $h'Hh<0$ | 帽形，局部最大 |
| 不定 | 正负方向并存 | 鞍点，不是局部最优 |
| 特征值判别 | $\lambda_i$ 全正/全负/混合 | 判断曲率方向 |
| OLS | $\min(y-X\beta)'(y-X\beta)$ | 投影和平方误差最小化 |
| 投资组合 | $\max \mu'w-\frac{\gamma}{2}w'\Sigma w$ | 收益-风险权衡 |

## 12 练习题

### A 组：基础计算

A1. 求 $f(x,y)=x^2+xy+y^2-2x-4y$ 的驻点，并用 Hessian 分类。

A2. 求 $f(x,y)=x^3+y^3-3xy$ 的所有驻点，并分类。

A3. 判断矩阵 $A=\begin{pmatrix}4&1\\1&3\end{pmatrix}$ 的正定性。

A4. 判断矩阵 $B=\begin{pmatrix}-2&1\\1&-2\end{pmatrix}$ 的负定性。

A5. 对 $f(x,y,z)=x^2+2y^2+3z^2+2xy-2xz$，写出 Hessian，并判断是否正定。

A6. 对 $f(x,y)=\ln x+\theta\ln y$，$x,y>0,\theta>0$，求梯度和 Hessian。

### B 组：概念判断

B1. 如果 $\nabla f(x^*)=0$，是否一定有局部最优？说明理由。

B2. 如果 $H_f(x^*)$ 负定，但 $\nabla f(x^*)\neq 0$，能否说 $x^*$ 是局部最大点？

B3. 为什么二维 Hessian 判别中不能只看 $f_{xx}$ 和 $f_{yy}$？

B4. 解释 $h'Hh$ 中方向 $h$ 的经济含义。

B5. 为什么 OLS 的 Hessian 是 $2X'X$？它什么时候正定？

### C 组：混合题

C1. 设 $f(x,y)=-(x-1)^2-2(y+1)^2+3$。求最大点，并用梯度与 Hessian 解释。

C2. 设 $f(x,y)=x^4+y^4-4xy$。求驻点，并尝试分类。

C3. 设 $f(x,y)=x^2+2axy+y^2$。讨论参数 $a$ 取什么值时 Hessian 正定。

C4. 设 $f(x,y)=e^x+e^y-e^{x+y}$。求驻点并分类。

C5. 设 $f(x,y)=x^2y+y^2x$。求驻点并说明二阶判别是否充分。

### D 组：经济金融建模题

D1. 厂商利润为 $\pi(K,L)=10K^{1/3}L^{1/3}-K-L$。求内部一阶条件，并判断利润函数在正象限是否凹。

D2. 投资者目标为 $U(w)=0.08w_1+0.12w_2-2(w_1^2+0.5w_2^2+0.3w_1w_2)$。求无约束最优权重，并判断 Hessian。

D3. OLS 目标函数 $S(\beta)=\sum_{i=1}^{n}(y_i-x_i'\beta)^2$。从矩阵形式推导 normal equations，并解释正定性。

D4. 社会福利函数 $W(g,t)=a\ln g+b\ln t-c(g+t)$，其中 $a,b,c>0$。求内部最优公共支出 $g$ 与转移支付 $t$。

D5. 设生产函数 $F(K,L)=K^{\alpha}L^{1-\alpha}$。计算 $F_{KL}$ 并解释资本与劳动的边际互补性。

### E 组：挑战/证明题

E1. 证明：若 $A$ 对称且所有特征值正，则 $A$ 正定。

E2. 构造一个函数，使得 $\nabla f(0,0)=0$，$H_f(0,0)=0$，但 $(0,0)$ 是严格局部最小点。

E3. 构造一个函数，使得 $\nabla f(0,0)=0$，$H_f(0,0)=0$，但 $(0,0)$ 是鞍点。

E4. 设 $f$ 二阶连续可微且 Hessian 在凸集 $C$ 上处处半负定。说明为什么 $f$ 是凹函数的直觉，并尝试用一维限制函数证明。

E5. 对均值-方差目标 $U(w)=\mu'w-\frac{\gamma}{2}w'\Sigma w$，证明当 $\Sigma$ 正定时目标函数严格凹。

# 13 练习题详细答案

## A 组答案

**A1.** 梯度：

$$f_x = 2x + y - 2, \qquad f_y = x + 2y - 4.$$

解

$$2x + y = 2, \qquad x + 2y = 4.$$

得 $x = 0, y = 2$。Hessian 为

$$H = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

$\det(H) = 3 > 0$ 且 $f_{xx} = 2 > 0$，所以正定，$(0, 2)$ 是严格局部最小点。

**A2.** 梯度为

$$f_x = 3x^2 - 3y, \qquad f_y = 3y^2 - 3x.$$

FOC 给出 $y = x^2$，$x = y^2$，所以 $x = x^4$，即 $x = 0$ 或 $x = 1$。驻点为 $(0, 0)$ 和 $(1, 1)$。Hessian：

$$H = \begin{pmatrix} 6x & -3 \\ -3 & 6y \end{pmatrix}.$$

在 $(0, 0)$，$\det(H) = -9 < 0$，鞍点。在 $(1, 1)$，$\det(H) = 36 - 9 = 27 > 0$ 且 $f_{xx} = 6 > 0$，局部最小点。

**A3.** $\Delta_1 = 4 > 0$，$\Delta_2 = 4 \cdot 3 - 1 = 11 > 0$，所以 $A$ 正定。

**A4.** $\Delta_1 = -2 < 0$，$\Delta_2 = 4 - 1 = 3 > 0$。符号满足负定的二维条件，所以 $B$ 负定。

**A5.** Hessian 为

$$H = \begin{pmatrix} 2 & 2 & -2 \\ 2 & 4 & 0 \\ -2 & 0 & 6 \end{pmatrix}.$$

顺序主子式：

$$\Delta_1 = 2 > 0, \quad \Delta_2 = 8 - 4 = 4 > 0,$$

$$\Delta_3 = \det(H) = 16 > 0.$$

所以 Hessian 正定，函数严格凸。

**A6.** 梯度：

$$\nabla f = \begin{pmatrix} 1/x \\ \theta/y \end{pmatrix}.$$

Hessian：

$$H = \begin{pmatrix} -1/x^2 & 0 \\ 0 & -\theta/y^2 \end{pmatrix}.$$

在 $x, y > 0$ 且 $\theta > 0$ 时负定，效用函数严格凹。

## B 组答案

**B1.** 不一定。例如 $f(x, y) = x^2 - y^2$ 在 $(0, 0)$ 梯度为零，但它是鞍点。

**B2.** 不能。局部最大点的必要条件是梯度为零。如果梯度非零，沿梯度方向可以提高目标函数，所以不是内部局部最大点。

**B3.** 因为交叉项 $f_{xy}$ 可能很大，导致某些混合方向上曲率符号改变。例如

$$H = \begin{pmatrix} -1 & 10 \\ 10 & -1 \end{pmatrix}$$

对角线都负，但行列式 $1 - 100 < 0$，矩阵不定。

**B4.** $h$ 是从当前选择点出发的一小步变化方向。$h'Hh$ 衡量沿这个联合调整方向，目标函数的二阶变化。例如企业同时改变资本和劳动时，不仅有各自的二阶效应，还有交叉效应。

**B5.** 因为

$$S(\beta) = y'y - 2\beta'X'y + \beta'X'X\beta.$$

对 $\beta$ 求梯度得 $-2X'y + 2X'X\beta$，再求导就是 $2X'X$。当 $X$ 满列秩时，$X'X$ 正定，所以 Hessian 正定。

## C 组答案

**C1.** 梯度：

$$f_x = -2(x-1), \qquad f_y = -4(y+1).$$

FOC 得 $x = 1, y = -1$。Hessian

$$H = \begin{pmatrix} -2 & 0 \\ 0 & -4 \end{pmatrix}$$

负定，所以 $(1, -1)$ 是严格局部最大点，也是全局最大点。

**C2.** 梯度：

$$f_x = 4x^3 - 4y, \qquad f_y = 4y^3 - 4x.$$

FOC 给出 $y = x^3$，$x = y^3$，所以 $x = x^9$，即 $x = 0, \pm 1$。驻点为 $(0,0), (1,1), (-1,-1)$。Hessian：

$$H = \begin{pmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{pmatrix}.$$

在 $(0,0)$，$\det(H) = -16 < 0$，鞍点。在 $(1,1)$ 与 $(-1,-1)$，$\det(H) = 144 - 16 = 128 > 0$ 且 $f_{xx} = 12 > 0$，局部最小点。

**C3.** Hessian 为

$$H = \begin{pmatrix} 2 & 2a \\ 2a & 2 \end{pmatrix}.$$

正定需要 $2 > 0$ 且

$$\det(H) = 4 - 4a^2 > 0.$$

所以 $|a| < 1$。

**C4.** 梯度：

$$f_x = e^x - e^{x+y}, \qquad f_y = e^y - e^{x+y}.$$

FOC 给出 $e^x = e^{x+y}$，所以 $y = 0$；又 $e^y = e^{x+y}$，所以 $x = 0$。驻点为 $(0,0)$。Hessian：

$$H = \begin{pmatrix} e^x - e^{x+y} & -e^{x+y} \\ -e^{x+y} & e^y - e^{x+y} \end{pmatrix}.$$

在 $(0,0)$，

$$H = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix},$$

行列式 $-1 < 0$，所以是鞍点。

**C5.** 梯度：

$$f_x = 2xy + y^2, \qquad f_y = x^2 + 2xy.$$

FOC 为 $y(2x+y)=0$，$x(x+2y)=0$。唯一共同解是 $(0,0)$。Hessian：

$$H = \begin{pmatrix} 2y & 2x+2y \\ 2x+2y & 2x \end{pmatrix}.$$

在 $(0,0)$ 是零矩阵，二阶判别失败。沿 $y=x$，$f=2x^3$，有正有负；所以 $(0,0)$ 是鞍点。

## D 组答案

**D1.** 一阶条件：

$$\pi_K = \frac{10}{3}K^{-2/3}L^{1/3} - 1 = 0,$$

$$\pi_L = \frac{10}{3}K^{1/3}L^{-2/3} - 1 = 0.$$

两式相除得到 $L/K = 1$，即 $L = K$。代入第一式：

$$\frac{10}{3}K^{-1/3} = 1 \Rightarrow K = \left(\frac{10}{3}\right)^3,$$

所以 $L = K = (10/3)^3$。由于 Cobb-Douglas 的指数和为 $2/3 < 1$，生产函数凹，减去线性成本后利润函数凹。更直接地可检验 Hessian 负半定；在内部最优附近满足最大化的曲率要求。

**D2.** 目标为

$$U = 0.08w_1 + 0.12w_2 - 2w_1^2 - w_2^2 - 0.6w_1w_2.$$

梯度：

$$U_{w_1} = 0.08 - 4w_1 - 0.6w_2,$$

$$U_{w_2} = 0.12 - 2w_2 - 0.6w_1.$$

解线性方程组：

$$4w_1 + 0.6w_2 = 0.08, \qquad 0.6w_1 + 2w_2 = 0.12.$$

得到近似

$$w_1 \approx 0.01175, \qquad w_2 \approx 0.05647.$$

Hessian 为

$$H = \begin{pmatrix} -4 & -0.6 \\ -0.6 & -2 \end{pmatrix}.$$

$\Delta_1 = -4 < 0$，$\Delta_2 = 8 - 0.36 = 7.64 > 0$，所以负定，解是严格最大点。

**D3.** 矩阵推导见例 3。Normal equations 是

$$X'X\hat{\beta} = X'y.$$

正定性来自：对任意 $a \neq 0$，

$$a'X'Xa = (Xa)'(Xa) = \|Xa\|^2.$$

若 $X$ 满列秩，则 $Xa \neq 0$，所以 $a'X'Xa > 0$，即 $X'X$ 正定。

**D4.** 梯度：

$$W_g = \frac{a}{g} - c, \qquad W_t = \frac{b}{t} - c.$$

FOC 给出

$$g^* = \frac{a}{c}, \qquad t^* = \frac{b}{c}.$$

Hessian：

$$H = \begin{pmatrix} -a/g^2 & 0 \\ 0 & -b/t^2 \end{pmatrix}$$

负定，所以是严格局部最大点。

**D5.** 对

$$F(K, L) = K^\alpha L^{1-\alpha},$$

有

$$F_K = \alpha K^{\alpha-1} L^{1-\alpha}.$$

再对 $L$ 求导：

$$F_{KL} = \alpha(1-\alpha) K^{\alpha-1} L^{-\alpha} > 0.$$

经济含义是：劳动越多，资本的边际产出越高；资本越多，劳动的边际产出越高。两种要素在边际上互补。

## E 组答案

**E1.** 因为 $A$ 对称，可正交对角化：

$$A = Q\Lambda Q'.$$

对任意 $h \neq 0$，令 $z = Q'h$。由于 $Q$ 可逆，$z \neq 0$。于是

$$h'Ah = z'\Lambda z = \sum_i \lambda_i z_i^2.$$

若所有 $\lambda_i > 0$，且 $z \neq 0$，则和式严格大于零。所以 $A$ 正定。

**E2.** 例如

$$f(x, y) = x^4 + y^4.$$

在 $(0,0)$，梯度为零，Hessian 为零矩阵。但只要 $(x, y) \neq (0,0)$，就有 $f(x, y) > 0 = f(0,0)$，所以 $(0,0)$ 是严格局部最小点。

**E3.** 例如

$$f(x, y) = x^4 - y^4.$$

在 $(0,0)$，梯度为零，Hessian 为零矩阵。沿 $x$ 轴，$f(x, 0) = x^4 > 0$；沿 $y$ 轴，$f(0, y) = -y^4 < 0$。因此 $(0,0)$ 是鞍点。

**E4.** 任取 $x, y \in C$，考虑一元函数

$$\phi(t) = f((1-t)x + ty), \qquad t \in [0, 1].$$

由链式法则，

$$\phi''(t) = (y-x)' H_f((1-t)x + ty)(y-x).$$

若 Hessian 处处半负定，则 $\phi''(t) \leq 0$，所以 $\phi$ 是一元凹函数。于是

$$\phi(t) \geq (1-t)\phi(0) + t\phi(1),$$

也就是

$$f((1-t)x + ty) \geq (1-t)f(x) + tf(y).$$

这正是凹函数定义。

**E5.** Hessian 为

$$H_U = -\gamma\Sigma.$$

对任意非零 $h$，由于 $\Sigma$ 正定，

$$h'\Sigma h > 0.$$

所以

$$h'H_U h = -\gamma h'\Sigma h < 0.$$

因此 Hessian 负定，目标函数严格凹。

## 本章最后提醒

本章最关键的一句话是：**FOC 找候选点，Hessian 判别局部形状。**但这句话还不是完整经济学优化理论。下一章加入等式约束后，一阶条件会变成 Lagrange 条件；再下一章加入不等式约束后，会出现 KKT 条件和互补松弛。你现在需要把本章打牢，因为 Lagrange 和 KKT 的二阶直觉仍然来自这里：最优点附近，允许方向上的二阶变化必须朝正确方向弯曲。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 15
# Lagrange 乘子与等式约束

Chapter 15 接在无约束多元优化之后，处理经济金融中更常见的情形：选择必须满足某个资源、预算、技术、会计或组合权重约束。消费者不能随便选消费束，因为有预算约束；厂商不能随便选产量和投入，因为有生产目标或技术约束；投资者不能随便选权重，因为权重通常必须加总为一，有时还要满足目标收益。Lagrange 乘子的核心作用，是把“在约束集合上最优”翻译成“目标函数边际变化与约束边际变化成比例”。这一章要学会三件事：第一，为什么最优点处目标函数等高线要和约束曲线相切；第二，为什么乘子可以解释为影子价格；第三，如何用一阶条件、切空间二阶条件和 bordered Hessian 判断候选解。

math4econ personalized notes
2026

# Contents

**Chapter 15 总览** **2**

**1 本章为什么重要** **3**
1.1 约束是经济学模型的基本结构 . . . 3
1.2 Lagrange 乘子不是纯计算技巧 . . . 3
1.3 等式约束是 KKT 和动态优化的入口 . . . 3

**2 最小前置知识** **3**
2.1 梯度、全微分与等高线 . . . 3
2.2 可行集合与可行方向 . . . 4
2.3 线性代数里的正交补 . . . 4

**3 一个等式约束下的一阶条件** **5**
3.1 问题形式 . . . 5
3.2 切空间推导 . . . 5
3.3 Lagrangian 写法 . . . 5

**4 二维消费者问题中的几何解释** **6**

**5 多个等式约束** **7**
5.1 问题形式 . . . 7
5.2 正则性：约束梯度线性无关 . . . 7
5.3 多个乘子的一阶条件 . . . 7

**6 影子价格与包络思想** **8**
6.1 Value function . . . 8
6.2 乘子等于最优值对约束的边际变化 . . . 8
6.3 为什么经济学特别喜欢乘子 . . . 9

**7 二阶条件：为什么 FOC 仍然不够** **9**
7.1 约束下的二阶方向 . . . 9
7.2 Bordered Hessian 的二维版本 . . . 10

**8 计算流程总结** **10**

**9 Worked Example 1：Cobb-Douglas 消费者问题** **11**

**10 Worked Example 2：成本最小化与边际成本** **12**

**11 Worked Example 3：均值-方差投资组合** **13**

**12 Worked Example 4：社会计划者的资源分配** **14**

**13 Worked Example 5：切空间二阶条件** **15**

**14 常见误区** **16**

**15 经济金融应用地图** **16**

**16 练习题** **16**

A 组：基础计算 . . . . . 16
B 组：概念判断 . . . . . 17
C 组：混合题 . . . . . 17
D 组：经济金融建模题 . . . . . 17
E 组：挑战/证明题 . . . . . 17

**17 详细解答** **18**

A 组解答 . . . . . 18
B 组解答 . . . . . 19
C 组解答 . . . . . 19
D 组解答 . . . . . 20
E 组解答提示 . . . . . 21

**18 本章总结图谱** **21**

**19 术语表** **22**

**20 下一章预告** **22**

# Chapter 15 总览

Chapter 14 讨论的问题是

$$\max_{x\in\mathbb{R}^n} f(x),$$

也就是选择变量可以在整个空间里自由移动。但经济学中的选择几乎总是有约束的。最标准的等式约束问题是

$$\max_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x)=b.$$

这里 $f(x)$ 是目标，$g(x)=b$ 是约束。比如：

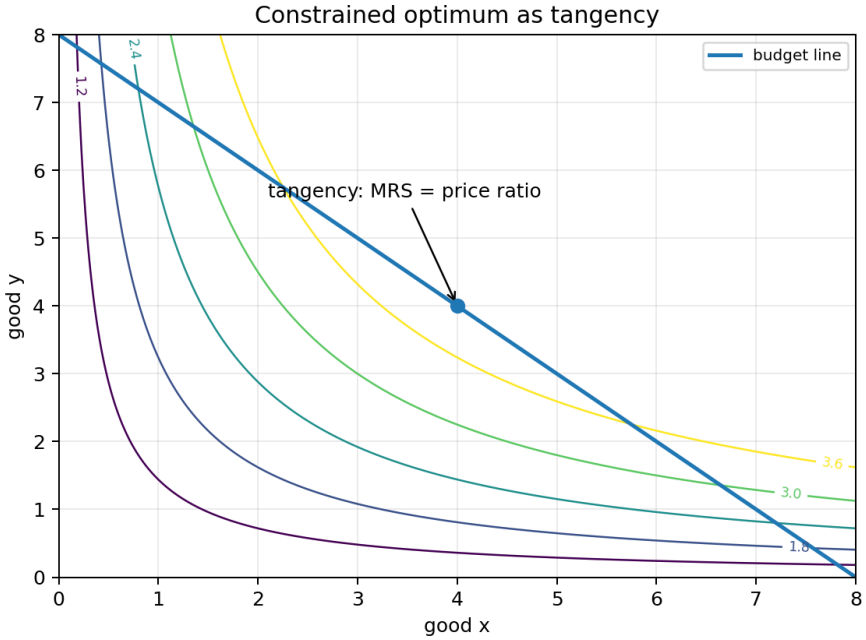
- 消费者问题：$g(x)=p\cdot x=w$；
- 成本最小化：$F(K,L)=q$；
- 投资组合：$\mathbf{1}'w=1$，有时还要求 $\mu'w=\bar{r}$；
- 社会计划者问题：$\sum_i x_i=\bar{X}$；
- 统计估计：参数满足 normalization 或 moment restrictions。

本章主线是

约束最优化 → 可行方向 → 切空间 → 梯度共线 → Lagrangian FOC → 乘子的影子价格解释.

**直觉框**

无约束最优要求“所有方向的一阶变化都为零”；等式约束最优只要求“所有**可行方向**的一阶变化都为零”。所谓可行方向，就是沿着它走一小步后仍然不会立刻违反约束的方向。Lagrange 方法的核心就是把“沿所有可行方向的变化为零”改写成一个容易计算的代数条件。




图中预算线上的最优点不是效用函数在整个平面上的无约束最大值，而是在预算线这个一维可行集合上的最大值。最优点处，无差异曲线与预算线相切；否则沿预算线还能移动到更高的无差异曲线。

# 1 本章为什么重要

## 1.1 约束是经济学模型的基本结构

经济学不是研究“想要什么”，而是研究“在约束下选择什么”。因此约束优化比无约束优化更接近经济模型的真实语言。典型例子包括：

- 预算约束：收入有限，消费束必须满足 $p \cdot x \le w$；
- 资源约束：一个经济体的总资源有限，配置必须满足 $\sum_i x_i = \bar{X}$；
- 技术约束：生产某个产量 $q$ 需要投入组合满足 $F(K, L) = q$；
- 会计约束：投资组合权重必须满足 $\sum_i w_i = 1$；
- 政策约束：政府预算平衡、排放总量限制、债务动态限制等。

## 1.2 Lagrange 乘子不是纯计算技巧

很多教材把 Lagrange 方法讲成“写出 $\mathcal{L} = f + \lambda g$，求偏导等于零”。这会让人误以为乘子只是为了求解方便引入的辅助变量。实际上，乘子在经济学里有非常明确的含义：**约束放松一点时，最优目标值改变多少**。

如果问题写作

$$V(b) = \max_x f(x) \quad \text{s.t.} \quad g(x) = b,$$

并采用

$$\mathcal{L}(x, \lambda; b) = f(x) + \lambda[b - g(x)],$$

那么在正则条件下有

$$V'(b) = \lambda^*.$$

这就是影子价格。它告诉你约束右端资源 $b$ 增加一单位，最大化后的目标值大约增加多少。

**经济金融解释**

在消费者问题中，预算约束的乘子可以解释为**财富的边际效用**。在成本最小化问题中，产量约束的乘子可以解释为**多生产一单位产量导致的最小成本增加**，也就是边际成本。影子价格是高微、宏观、金融、公共经济学和机制设计里反复出现的语言。

## 1.3 等式约束是 KKT 和动态优化的入口

Chapter 16 会处理不等式约束与 KKT 条件，而 KKT 本质上是 Lagrange 方法的扩展。动态优化中的 Euler equation、Hamiltonian、Bellman 方程也都与“在约束下选择路径”有关。因此本章不是孤立技巧，而是后面很多理论的入口。

# 2 最小前置知识

## 2.1 梯度、全微分与等高线

若 $f : \mathbb{R}^n \to \mathbb{R}$ 可微，则

$$\mathrm{d}f = \nabla f(x)' \mathrm{d}x.$$

梯度 $\nabla f(x)$ 指向局部最快上升方向，并且垂直于等高面 $f(x) = c$。在二维中，等高线是曲线；在三维中，等高面是曲面；在 $n$ 维中，等高集合是抽象的超曲面。

## 2.2 可行集合与可行方向

考虑一个约束

$$g(x) = b.$$

可行集合是

$$M = \{x \in \mathbb{R}^n : g(x) = b\}.$$

在可行点 $x^*$ 附近，若沿方向 $h$ 做微小移动 $x^* + th$，约束的一阶变化为

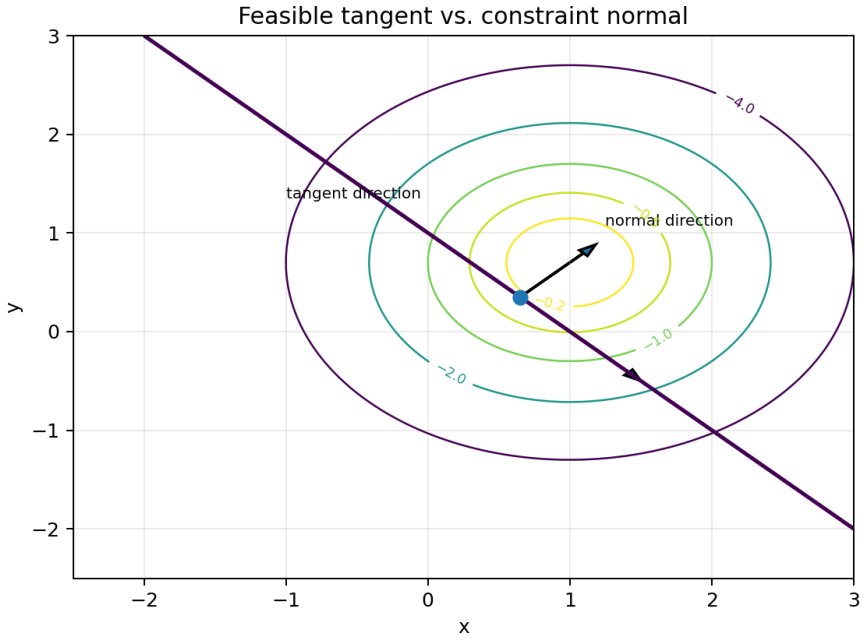
$$g(x^* + th) - g(x^*) \approx t\nabla g(x^*)'h.$$

为了保持在约束集合上一阶可行，必须有

$$\nabla g(x^*)'h = 0.$$

所以约束集合在 $x^*$ 处的切空间是

$$T_{x^*}M = \{h \in \mathbb{R}^n : \nabla g(x^*)'h = 0\}.$$




**直觉框**

约束梯度 $\nabla g(x^*)$ 垂直于约束集合，切方向 $h$ 与 $\nabla g(x^*)$ 正交。约束最优要求沿所有切方向都不能一阶提高目标函数，也就是 $\nabla f(x^*)'h = 0$ 对所有可行切方向成立。

## 2.3 线性代数里的正交补

若一个向量 $a \neq 0$，则集合

$$\{h : a'h = 0\}$$

是一个超平面。若另一个向量 $b$ 对所有满足 $a'h = 0$ 的 $h$ 都有 $b'h = 0$，那么 $b$ 必须与 $a$ 共线，即存在 $\lambda$ 使得

$$b = \lambda a.$$

这个线性代数事实正是 Lagrange 一阶条件背后的核心。

# 3 一个等式约束下的一阶条件

## 3.1 问题形式

先考虑最简单的情况：

$$\max_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x)=b.$$

假设 $x^*$ 是内部可行局部最优点，且约束是正则的：

$$\nabla g(x^*)\neq 0.$$

## 3.2 切空间推导

由于 $x^*$ 是约束集合上的局部最大点，所以沿任何可行切方向 $h\in T_{x^*}M$，目标函数的一阶变化必须为零：

$$\nabla f(x^*)'h=0 \quad \text{for all } h \text{ such that } \nabla g(x^*)'h=0.$$

根据上一节的线性代数事实，这意味着

$$\nabla f(x^*)=\lambda^*\nabla g(x^*)$$

对某个标量 $\lambda^*$ 成立。

这就是一个约束下的 Lagrange 条件。

推导与证明

**为什么梯度共线?**
令 $a=\nabla g(x^*)$，$b=\nabla f(x^*)$。正则性给出 $a\neq 0$。可行切方向是所有满足 $a'h=0$ 的方向。最优性要求对所有这样的 $h$，都有 $b'h=0$。这说明 $b$ 与 $a$ 的正交补中所有向量都正交，因此 $b$ 只能落在 $a$ 张成的一维空间里，即 $b=\lambda a$。
直观地说，如果 $\nabla f$ 不是 $\nabla g$ 的倍数，那么 $\nabla f$ 必然有某个沿约束切方向的分量。沿这个切方向移动，会使目标函数一阶上升或下降，从而不可能是约束最优点。

## 3.3 Lagrangian 写法

定义 Lagrangian：

$$\mathcal{L}(x,\lambda)=f(x)+\lambda[b-g(x)].$$

一阶条件是

$$\frac{\partial\mathcal{L}}{\partial x_i}=0, \quad i=1,\dots,n,$$

以及

$$\frac{\partial\mathcal{L}}{\partial\lambda}=b-g(x)=0.$$

合起来就是

$$\nabla f(x^*)-\lambda^*\nabla g(x^*)=0, \quad g(x^*)=b.$$

等价地，

$$\nabla f(x^*)=\lambda^*\nabla g(x^*).$$

**常见误区**

有些教材写 $\mathcal{L} = f + \lambda[g(x) - b]$，这时 FOC 会变成 $\nabla f + \lambda \nabla g = 0$，乘子的符号也会相反。两种写法都可以，但必须在解释影子价格时保持一致。本讲义采用

$$\mathcal{L} = f + \lambda[b - g(x)]$$

使得 $V'(b) = \lambda^*$。

## 4 二维消费者问题中的几何解释

考虑

$$\max_{x,y} u(x,y) \quad \text{s.t.} \quad p_x x + p_y y = w.$$

写成

$$g(x,y) = p_x x + p_y y = w.$$

Lagrangian 为

$$\mathcal{L}(x,y,\lambda) = u(x,y) + \lambda[w - p_x x - p_y y].$$

FOC 是

$$u_x(x,y) = \lambda p_x,$$
$$u_y(x,y) = \lambda p_y,$$
$$p_x x + p_y y = w.$$

前两式相除得到

$$\frac{u_x}{u_y} = \frac{p_x}{p_y}.$$

左边是边际替代率的绝对值，右边是价格比。因此最优条件就是

$$MRS = \frac{p_x}{p_y}.$$

**经济金融解释**

这个条件的经济含义是：消费者主观上愿意用多少 $y$ 换一单位 $x$，必须等于市场上实际需要放弃多少 $y$ 才能多买一单位 $x$。如果两者不相等，消费者就可以沿预算线调整消费束，提高效用。

乘子 $\lambda$ 的含义来自

$$u_x = \lambda p_x.$$

若花 $p_x$ 单位财富多买一点 $x$，获得边际效用 $u_x$，那么每一单位财富对应的边际效用是

$$\frac{u_x}{p_x} = \lambda.$$

同理，

$$\frac{u_y}{p_y} = \lambda.$$

最优时，每一元钱花在任何商品上得到的边际效用相同。否则就应该把钱从“单位货币边际效用低”的商品转到“单位货币边际效用高”的商品。

# 5 多个等式约束

## 5.1 问题形式

更一般地，考虑

$$\max_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad g_1(x)=b_1,\dots,g_m(x)=b_m.$$

令

$$g(x)=\begin{pmatrix} g_1(x)\\ \vdots\\ g_m(x)\end{pmatrix},\quad b=\begin{pmatrix} b_1\\ \vdots\\ b_m\end{pmatrix}.$$

问题写成

$$\max_{x} f(x) \quad \text{s.t.} \quad g(x)=b.$$

## 5.2 正则性：约束梯度线性无关

设 Jacobian 为

$$Dg(x)=\begin{pmatrix} \nabla g_1(x)'\\ \vdots\\ \nabla g_m(x)'\end{pmatrix}.$$

正则条件是

$$\operatorname{rank} Dg(x^*)=m.$$

也就是说，$m$ 个约束的梯度线性无关。若约束梯度线性相关，可能说明某些约束重复，或者约束集合在该点有尖点/奇异性，这会破坏标准 Lagrange 条件的必要性。

## 5.3 多个乘子的一阶条件

采用

$$\mathcal{L}(x,\lambda;b)=f(x)+\lambda'[b-g(x)],$$

其中 $\lambda=(\lambda_1,\dots,\lambda_m)'$。FOC 是

$$\nabla_x\mathcal{L}(x^*,\lambda^*;b)=0,$$

即

$$\nabla f(x^*)-Dg(x^*)'\lambda^*=0,$$

以及

$$g(x^*)=b.$$

写成展开形式：

$$\nabla f(x^*)=\lambda_1^*\nabla g_1(x^*)+cdots+\lambda_m^*\nabla g_m(x^*).$$

**直觉框**

一个约束时，目标函数梯度必须落在一个约束梯度张成的直线上；多个约束时，目标函数梯度必须落在所有约束梯度张成的空间中。换句话说，目标函数不能有任何沿可行切空间的分量。

# 6 影子价格与包络思想

## 6.1 Value function

定义最优值函数

$$V(b) = \max_x f(x) \quad \text{s.t.} \quad g(x) = b.$$

如果 $b$ 表示资源、财富、产量目标、目标收益等外生约束水平，那么 $V(b)$ 告诉我们：约束水平为 $b$ 时，最优目标值是多少。

## 6.2 乘子等于最优值对约束的边际变化

在本讲义的符号约定下

$$\mathcal{L}(x, \lambda; b) = f(x) + \lambda'[b - g(x)],$$

并在可微和正则条件下，有

$$\frac{\partial V}{\partial b_j} = \lambda_j^*.$$

推导与证明

**一维约束下的推导。**
设 $x^*(b)$ 是最优解，$\lambda^*(b)$ 是对应乘子，则

$$V(b) = f(x^*(b)).$$

对 $b$ 求导：

$$V'(b) = \nabla f(x^*(b))' \frac{\mathrm{d}x^*(b)}{\mathrm{d}b}.$$

由 FOC，

$$\nabla f(x^*) = \lambda^* \nabla g(x^*).$$

所以

$$V'(b) = \lambda^* \nabla g(x^*)' \frac{\mathrm{d}x^*}{\mathrm{d}b}.$$

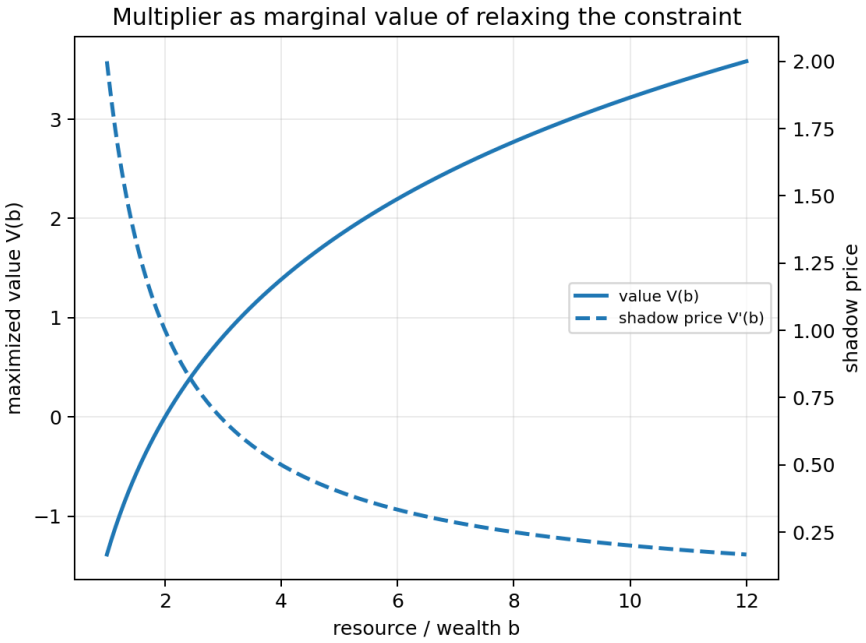
又因为约束 $g(x^*(b)) = b$，对 $b$ 求导得

$$\nabla g(x^*)' \frac{\mathrm{d}x^*}{\mathrm{d}b} = 1.$$

代回得到

$$V'(b) = \lambda^*.$$

这就是影子价格公式。




### 6.3 为什么经济学特别喜欢乘子

乘子把“最优解怎么变”与“最优值怎么变”区分开来。很多时候，我们不关心最优选择本身如何复杂调整，而只关心约束放松后目标值增加多少。乘子直接给出这个边际影响。

**经济金融解释**

比如社会计划者分配稀缺资源，资源约束乘子衡量一单位额外资源的社会边际价值；企业成本最小化中，产量约束乘子衡量一单位额外产量的最小成本增加；投资组合中，目标收益约束的乘子衡量要求更高目标收益会让最小风险增加多少。

## 7 二阶条件：为什么 FOC 仍然不够

### 7.1 约束下的二阶方向

在无约束问题中，驻点处的二阶条件看 Hessian：

$$h' H_f(x^*) h.$$

在约束问题中，不是所有方向 $h$ 都允许，只能看可行切方向：

$$T_{x^*} M = \{h : Dg(x^*)h = 0\}.$$

定义 Lagrangian 对 $x$ 的 Hessian：

$$H_{\mathcal{L}}(x^*, \lambda^*) = \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) = \nabla^2 f(x^*) - \sum_{j=1}^{m} \lambda_j^* \nabla^2 g_j(x^*).$$

约束局部最大点的充分条件通常是：

$$h' H_{\mathcal{L}}(x^*, \lambda^*) h < 0 \quad \text{for all } h \neq 0 \text{ with } Dg(x^*)h = 0.$$

约束局部最小点的充分条件通常是：

$$h' H_{\mathcal{L}}(x^*, \lambda^*) h > 0 \quad \text{for all } h \neq 0 \text{ with } Dg(x^*)h = 0.$$

**直觉框**

约束下的二阶条件不是问函数在整个空间里像碗还是像帽子，而是问它**沿着可行集合**像碗还是像帽子。目标函数在不可行方向上的弯曲，与约束最优无关。

### 7.2 Bordered Hessian 的二维版本

对于二维一个等式约束

$$\max_{x,y} f(x,y) \quad \text{s.t.} \quad g(x,y) = b,$$

bordered Hessian 写为

$$B = \begin{pmatrix} 0 & g_x & g_y \\ g_x & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ g_y & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{pmatrix}.$$

在常见正则情形下：

- 若 $\det B > 0$，候选点通常对应约束局部最大；
- 若 $\det B < 0$，候选点通常对应约束局部最小。

不过本讲义建议你优先掌握**切空间二阶条件**，因为它更不容易记错，也更适合高维问题。Bordered Hessian 可以作为计算捷径，但不要把它当成唯一理解方式。

**常见误区**

二阶条件有必要条件和充分条件之分，而且 bordered Hessian 的符号规则会随“最大/最小、约束个数、变量个数、Lagrangian 符号约定”而变化。考试中若只记符号规则，很容易错。最稳健的办法是：先求 FOC，再找切方向 $h$，最后检查 $h' H_{\mathcal{L}} h$ 在切空间上的正负。

## 8 计算流程总结

方法框

**等式约束优化的标准步骤。**
给定

$$\max_x f(x) \quad \text{s.t.} \quad g(x) = b.$$

1. 写出 Lagrangian：
$$\mathcal{L}(x, \lambda) = f(x) + \lambda'[b - g(x)].$$
2. 求一阶条件：
$$\nabla_x \mathcal{L} = 0, \quad b - g(x) = 0.$$
3. 解出候选点 $(x^*, \lambda^*)$。
4. 检查正则性：约束梯度是否满秩。
5. 若需要判断最大/最小，检查二阶条件：
$$h' H_{\mathcal{L}} h \quad \text{on } \{h : Dg(x^*)h = 0\}.$$
6. 解释乘子：
$$\lambda_j^* = \frac{\partial V}{\partial b_j}.$$

## 9 Worked Example 1：Cobb-Douglas 消费者问题

Worked Example

考虑消费者

$$\max_{x,y>0} u(x, y) = \alpha \ln x + (1 - \alpha) \ln y \quad \text{s.t.} \quad p_x x + p_y y = w,$$

其中 $0 < \alpha < 1$。
**Step 1: 写 Lagrangian.**

$$\mathcal{L} = \alpha \ln x + (1 - \alpha) \ln y + \lambda(w - p_x x - p_y y).$$

**Step 2: FOC.**

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\alpha}{x} - \lambda p_x = 0,$$
$$\frac{\partial \mathcal{L}}{\partial y} = \frac{1 - \alpha}{y} - \lambda p_y = 0,$$
$$w - p_x x - p_y y = 0.$$

由前两式：

$$x = \frac{\alpha}{\lambda p_x}, \quad y = \frac{1 - \alpha}{\lambda p_y}.$$

代入预算约束：

$$p_x \frac{\alpha}{\lambda p_x} + p_y \frac{1 - \alpha}{\lambda p_y} = w.$$

即

$$\frac{\alpha}{\lambda} + \frac{1 - \alpha}{\lambda} = w,$$

所以

$$\lambda^* = \frac{1}{w}.$$

因此 Marshallian demand 为

$$x^* = \frac{\alpha w}{p_x}, \quad y^* = \frac{(1-\alpha)w}{p_y}.$$

**Step 3: 解释。**消费者把财富的比例 $\alpha$ 花在商品 $x$ 上，把比例 $1-\alpha$ 花在商品 $y$ 上。乘子

$$\lambda^* = \frac{1}{w}$$

是财富的边际效用。财富越多，额外一元钱带来的边际效用越低。

推导与证明

**影子价格验证。**
把最优解代回效用函数：

$$V(w, p_x, p_y) = \alpha \ln\left(\frac{\alpha w}{p_x}\right) + (1-\alpha)\ln\left(\frac{(1-\alpha)w}{p_y}\right).$$

对 $w$ 求导：

$$\frac{\partial V}{\partial w} = \alpha\frac{1}{w} + (1-\alpha)\frac{1}{w} = \frac{1}{w}.$$

这正好等于 $\lambda^*$。

## 10 Worked Example 2：成本最小化与边际成本

Worked Example

厂商生产函数为

$$q = F(K, L) = K^a L^{1-a}, \quad 0 < a < 1.$$

资本价格为 $r$，劳动价格为 $w$。给定产量目标 $\bar{q}$，求成本最小化问题：

$$\min_{K,L>0} rK + wL \quad \text{s.t.} \quad K^a L^{1-a} = \bar{q}.$$

为了让乘子直接表示成本对产量目标的边际变化，写

$$\mathcal{L} = rK + wL + \mu[\bar{q} - K^a L^{1-a}].$$

FOC 为

$$r - \mu a K^{a-1} L^{1-a} = 0,$$
$$w - \mu(1-a) K^a L^{-a} = 0,$$
$$K^a L^{1-a} = \bar{q}.$$

前两式相除：

$$\frac{r}{w} = \frac{aK^{a-1}L^{1-a}}{(1-a)K^a L^{-a}} = \frac{a}{1-a}\frac{L}{K}.$$

整理得

$$\frac{K}{L} = \frac{a}{1-a}\frac{w}{r}.$$

这说明最优投入比例由技术参数和要素价格共同决定。
进一步可解得条件要素需求：

$$K^*(\bar{q}, r, w) = \bar{q}\left(\frac{a}{1-a}\frac{w}{r}\right)^{1-a},$$

$$L^*(\bar{q}, r, w) = \bar{q}\left(\frac{1-a}{a}\frac{r}{w}\right)^{a}.$$

代回成本函数：

$$C(\bar{q}, r, w) = \bar{q}\, r^a w^{1-a}\, a^{-a}(1-a)^{-(1-a)}.$$

所以边际成本为

$$\frac{\partial C}{\partial \bar{q}} = r^a w^{1-a}\, a^{-a}(1-a)^{-(1-a)}.$$

而由影子价格解释，

$$\mu^* = \frac{\partial C}{\partial \bar{q}}.$$

**经济金融解释**

成本最小化中的乘子是“多要求一单位产量，最小成本增加多少”。这就是边际成本。你以后在生产者理论中看到 Shephard's lemma、Hotelling's lemma、对偶理论时，这种乘子解释会反复出现。

## 11 Worked Example 3：均值-方差投资组合

**Worked Example**

投资者有 $n$ 个资产，权重向量为 $w \in \mathbb{R}^n$，协方差矩阵为 $\Sigma$，期望收益向量为 $\mu$。考虑在给定目标收益 $\bar{r}$ 下最小化组合方差：

$$\min_{w} \frac{1}{2} w'\Sigma w \quad \text{s.t.} \quad \mathbf{1}'w = 1, \quad \mu' w = \bar{r}.$$

写 Lagrangian：

$$\mathcal{L} = \frac{1}{2} w'\Sigma w + u(1 - \mathbf{1}'w) + \eta(\bar{r} - \mu' w).$$

FOC：

$$\Sigma w - \nu \mathbf{1} - \eta \mu = 0.$$

因此

$$w = \Sigma^{-1}(\nu \mathbf{1} + \eta \mu).$$

代入两个约束。定义

$$A = \mathbf{1}'\Sigma^{-1}\mathbf{1}, \quad B = \mathbf{1}'\Sigma^{-1}\mu, \quad C = \mu'\Sigma^{-1}\mu.$$

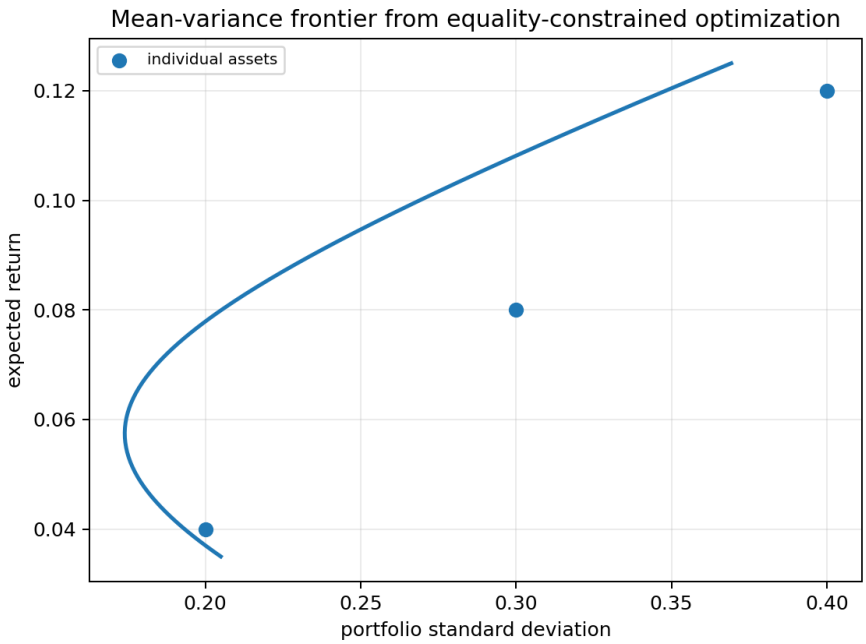
则

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix} \begin{pmatrix} \nu \\ \eta \end{pmatrix} = \begin{pmatrix} 1 \\ \bar{r} \end{pmatrix}.$$

解出 $\nu, \eta$ 后得到最小方差组合。组合的最小方差为

$$\sigma_p^2(\bar{r}) = \frac{A\bar{r}^2 - 2B\bar{r} + C}{AC - B^2}.$$

这就是均值-方差前沿的数学来源。




**经济金融解释**

投资组合问题中的两个等式约束分别是“钱要全部投出去”和“达到目标收益”。其中目标收益约束的乘子衡量：如果投资者提高一点目标收益，最小风险目标值会如何变化。注意这里目标函数是风险而不是效用，所以乘子的经济解释要结合具体 Lagrangian 符号和目标方向来读。

## 12 Worked Example 4：社会计划者的资源分配

**Worked Example**

社会中有两个个体，总资源为 $\bar{X}$。计划者最大化加权社会福利：

$$\max_{x_1, x_2 > 0} \theta \ln x_1 + (1-\theta) \ln x_2 \quad \text{s.t.} \quad x_1 + x_2 = \bar{X},$$

其中 $0 < \theta < 1$。

Lagrangian：

$$\mathcal{L} = \theta \ln x_1 + (1-\theta) \ln x_2 + \lambda(\bar{X} - x_1 - x_2).$$

FOC：

$$\frac{\theta}{x_1} = \lambda, \quad \frac{1-\theta}{x_2} = \lambda, \quad x_1 + x_2 = \bar{X}.$$

由前两式：

$$x_1 = \frac{\theta}{\lambda}, \quad x_2 = \frac{1-\theta}{\lambda}.$$

代入资源约束：

$$\frac{1}{\lambda} = \bar{X},$$

所以

$$\lambda^* = \frac{1}{\bar{X}}, \quad x_1^* = \theta\bar{X}, \quad x_2^* = (1-\theta)\bar{X}.$$

计划者按照社会福利权重分配资源。

**经济金融解释**

资源约束乘子 $\lambda^*$ 是额外一单位总资源的社会边际价值。在更一般的社会计划者问题中，乘子常常对应均衡价格、资源稀缺程度或贴现后的影子价值。

## 13 Worked Example 5：切空间二阶条件

**Worked Example**

考虑

$$\max_{x,y} f(x,y) = xy \quad \text{s.t.} \quad x + y = 10.$$

约束为 $g(x,y) = x + y = 10$，采用

$$\mathcal{L} = xy + lambda(10 - x - y).$$

FOC：

$$y - \lambda = 0,$$
$$x - \lambda = 0,$$
$$x + y = 10.$$

因此

$$x^* = y^* = 5, \quad \lambda^* = 5.$$

现在检查二阶条件。Lagrangian 对 $(x,y)$ 的 Hessian 为

$$H_{\mathcal{L}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

切方向满足

$$\nabla g' h = (1,1)\begin{pmatrix} h_x \\ h_y \end{pmatrix} = 0,$$

即

$$h_y = -h_x.$$

令 $h = (t, -t)'$。则

$$h' H_{\mathcal{L}} h = \begin{pmatrix} t & -t \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} t \\ -t \end{pmatrix} = -2t^2 < 0$$

对所有 $t \neq 0$ 成立。因此 $(5,5)$ 是约束局部最大点。

这个例子也可以直接化为一元问题：由 $y = 10 - x$，

$$f(x, 10 - x) = x(10 - x) = 10x - x^2,$$

显然在 $x = 5$ 取最大。切空间二阶条件只是把这个一元替代法推广到高维情形。

## 14 常见误区

**常见误区**

**误区 1：忘记约束方程本身。**
很多人只写 $\partial \mathcal{L}/\partial x_i = 0$，却忘了 $\partial \mathcal{L}/\partial \lambda = 0$。乘子的 FOC 正是原约束。少了它，方程数量不够，也没有可行性。

**常见误区**

**误区 2：把乘子符号看成绝对含义。**
乘子的符号依赖 Lagrangian 写法。若你写 $\mathcal{L} = f + \lambda(b - g)$，则 $V'(b) = \lambda$；若写 $\mathcal{L} = f + \lambda(g - b)$，则 $V'(b) = -\lambda$。经济解释前必须检查符号约定。

**常见误区**

**误区 3：把 FOC 当成充分条件。**
Lagrange FOC 只是正则局部最优的必要条件。若目标函数不凹、约束集合不凸、存在多个候选点或边界问题，只靠 FOC 不足以保证全局最优。

**常见误区**

**误区 4：混淆等式约束和不等式约束。**
等式约束的乘子没有非负要求。不等式约束的乘子才会出现 $\lambda \geq 0$ 和 complementary slackness。这个内容放在 Chapter 16。

## 15 经济金融应用地图

| 领域 | 约束优化形式 | 乘子解释 |
|---|---|---|
| 消费者理论 | $\max u(x)$ s.t. $p \cdot x = w$ | 财富的边际效用 |
| 生产者理论 | $\min wL + rK$ s.t. $F(K, L) = q$ | 多生产一单位产量的边际成本 |
| 社会计划者 | $\max \sum_i \theta_i u_i(x_i)$ s.t. $\sum_i x_i = \bar{X}$ | 资源的社会边际价值 |
| 投资组合 | $\min w'\Sigma w$ s.t. $\mathbf{1}'w = 1, \mu'w = \bar{r}$ | 预算和目标收益约束的影子成本 |
| 计量/统计 | $\min Q(\theta)$ s.t. $r(\theta) = 0$ | restriction 的边际损失 |
| 动态优化 | 选择路径满足状态方程 | 当前资源或状态变量的影子价值 |

## 16 练习题

### A 组：基础计算

A1. 求解 $\max_{x,y} xy$ s.t. $2x + y = 12$，并解释乘子含义。

A2. 求解 $\max_{x,y>0} \ln x + 2\ln y$ s.t. $x + 4y = 20$。

A3. 求解 $\min_{x,y} x^2 + y^2$ s.t. $x + y = 10$。

A4. 求解 $\max_{x,y,z>0} \ln x + \ln y + \ln z$ s.t. $x + y + z = 30$。

A5. 对 $f(x, y) = x^2 + y^2$，约束 $x - y = 2$，用切空间二阶条件判断候选点是最大还是最小。

## B 组：概念判断

B1. 为什么约束最优点不一定满足 $\nabla f = 0$?

B2. 为什么一个等式约束下的 FOC 可以写成 $\nabla f = \lambda \nabla g$?

B3. 等式约束乘子是否必须非负？为什么?

B4. 若 $\nabla g(x^*) = 0$，标准 Lagrange 必要条件为什么可能失效?

B5. 为什么乘子可以解释为 shadow price？这个解释依赖什么符号约定?

## C 组：混合题

C1. 设 $u(x, y) = x^a y^{1-a}$，$p_x x + p_y y = w$。推导 Marshallian demand，并说明它与 log utility 情形的关系。

C2. 设 $F(K, L) = K^{1/2} L^{1/2}$，要生产 $q$，要素价格为 $r, w$。求成本函数和边际成本。

C3. 设 $f(x, y) = x + y$，约束 $x^2 + y^2 = 1$。求最大值、最小值和对应乘子。

C4. 设 $f(x, y) = x^2 - y^2$，约束 $x + y = 0$。求候选点并判断性质。

C5. 用一元替代法验证 A1 或 A3 的结果，并与 Lagrange 方法比较。

## D 组：经济金融建模题

D1. 消费者有三种商品，效用为 $u = x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3}$，其中 $\alpha_1 + \alpha_2 + \alpha_3 = 1$。推导需求函数。

D2. 社会计划者最大化 $\theta \ln x_1 + (1 - \theta) \ln x_2$，资源约束为 $x_1 + x_2 = \bar{X}$。解释 $\theta$ 变化如何影响配置。

D3. 两资产组合期望收益为 $\mu_1, \mu_2$，方差协方差矩阵为 $\Sigma$。在 $w_1 + w_2 = 1$ 下最小化方差，求最小方差组合。

D4. 设政府选择公共品 $G$ 和私人消费 $C$，社会福利 $W = \ln C + \beta \ln G$，资源约束 $C + G = Y$。求最优公共品支出份额。

D5. 解释为什么成本最小化问题中的产量约束乘子可被看成边际成本。

## E 组：挑战/证明题

E1. 证明若 $x^*$ 是正则约束局部最优点，则 $\nabla f(x^*)$ 位于约束梯度张成的空间。

E2. 对多个等式约束，写出切空间二阶充分条件，并解释它为什么只检查可行方向。

E3. 设 $V(b) = \max_x f(x)$ s.t. $g(x) = b$。在一阶条件和可微最优选择下推导 $V'(b) = \lambda^*$。

E4. 对均值-方差问题，推导 $\sigma_p^2(\bar{r}) = \frac{A\bar{r}^2 - 2B\bar{r} + C}{AC - B^2}$。

E5. 给出一个例子说明 Lagrange FOC 成立但不是最大或最小。

# 17 详细解答

## A 组解答

**A1.** 写

$$\mathcal{L} = xy + \lambda(12 - 2x - y).$$

FOC：

$$y - 2\lambda = 0, \quad x - \lambda = 0, \quad 2x + y = 12.$$

由 $x = \lambda$，$y = 2\lambda$，代入约束：

$$2\lambda + 2\lambda = 12,$$

所以 $\lambda = 3$，$x = 3, y = 6$。乘子表示约束右端 12 增加一单位时，最大化后的 $xy$ 大约增加 3。

**A2.** 写

$$\mathcal{L} = \ln x + 2\ln y + \lambda(20 - x - 4y).$$

FOC：

$$\frac{1}{x} = \lambda, \quad \frac{2}{y} = 4\lambda, \quad x + 4y = 20.$$

第二式给 $y = \frac{1}{2\lambda}$，第一式给 $x = \frac{1}{\lambda}$，所以 $x = 2y$。代入约束：

$$2y + 4y = 20,$$

得 $y = 10/3$，$x = 20/3$，$\lambda = 3/20$。

**A3.** 写

$$\mathcal{L} = x^2 + y^2 + \lambda(10 - x - y).$$

FOC：

$$2x - \lambda = 0, \quad 2y - \lambda = 0, \quad x + y = 10.$$

得 $x = y = 5$。由于目标函数严格凸，约束为线性集合，所以这是全局最小点。

**A4.** 写

$$\mathcal{L} = \ln x + \ln y + \ln z + \lambda(30 - x - y - z).$$

FOC：

$$\frac{1}{x} = \lambda, \quad \frac{1}{y} = \lambda, \quad \frac{1}{z} = \lambda, \quad x + y + z = 30.$$

所以 $x = y = z = 10$，$\lambda = 1/10$。

**A5.** FOC：

$$\mathcal{L} = x^2 + y^2 + \lambda(2 - x + y).$$

$$2x - \lambda = 0, \quad 2y + \lambda = 0, \quad x - y = 2.$$

由 $\lambda = 2x = -2y$ 得 $x = -y$，再由 $x - y = 2$ 得 $x = 1, y = -1$。Hessian 是 $2I$。切方向满足 $h_x - h_y = 0$，即 $h = (t, t)$。于是

$$h' 2Ih = 4t^2 > 0,$$

所以是约束局部最小点。

## B 组解答

**B1.** 因为选择不能朝任意方向移动，只能沿约束集合移动。约束最优只要求沿可行方向不能提高目标，而不要求所有方向的一阶变化都为零。

**B2.** 可行切方向满足 $\nabla g' h = 0$。约束最优要求 $\nabla f' h = 0$ 对所有这样的 $h$ 成立。因此 $\nabla f$ 必须与 $\nabla g$ 共线。

**B3.** 不必须。等式约束是必须严格满足的条件，乘子可以正、负或零。不等式约束中的乘子才通常有非负限制。

**B4.** 若 $\nabla g(x^*) = 0$，约束集合在该点可能没有正常的切空间结构，或者该点是奇异点。标准 Lagrange 定理需要正则条件，否则梯度共线结论可能不能推出。

**B5.** 乘子等于最优值函数对约束右端的边际导数。例如采用 $\mathcal{L} = f + \lambda(b - g)$ 时，$V'(b) = \lambda$。若采用相反符号，解释会变号。

## C 组解答

**C1.** 对 $u = x^a y^{1-a}$，最大化 $u$ 等价于最大化

$$a \ln x + (1-a) \ln y.$$

所以由 Cobb-Douglas log utility 的结果：

$$x^* = \frac{aw}{p_x}, \quad y^* = \frac{(1-a)w}{p_y}.$$

Cobb-Douglas 的幂效用和 log 形式在偏好表示上等价，给出同样需求。

**C2.** 由 Worked Example 2 取 $a = 1/2$，得

$$K^* = q \left(\frac{w}{r}\right)^{1/2}, \quad L^* = q \left(\frac{r}{w}\right)^{1/2}.$$

成本为

$$C(q, r, w) = 2q\sqrt{rw},$$

边际成本为

$$MC = 2\sqrt{rw}.$$

**C3.** 写

$$\mathcal{L} = x + y + \lambda(1 - x^2 - y^2).$$

FOC：

$$1 - 2\lambda x = 0, \quad 1 - 2\lambda y = 0, \quad x^2 + y^2 = 1.$$

得 $x = y$，所以 $2x^2 = 1$，即 $x = y = 1/\sqrt{2}$ 或 $x = y = -1/\sqrt{2}$。前者最大值为 $\sqrt{2}$，后者最小值为 $-\sqrt{2}$。

**C4.** 约束 $x + y = 0$，即 $y = -x$。目标函数

$$f(x, -x) = x^2 - x^2 = 0.$$

整个约束线上目标值都相同，所以每个可行点都是最大也是最小，但不是严格最大/最小。这说明有时 FOC 会给出一族解。

**C5.** A3 中由 $y = 10 - x$ 得

$$f(x) = x^2 + (10 - x)^2 = 2x^2 - 20x + 100.$$

一阶条件 $4x - 20 = 0$ 给出 $x = 5$，$y = 5$。二阶导 $4 > 0$，所以是最小点。Lagrange 方法保留了对称性和乘子解释，而替代法在低维时更直观。

## D 组解答

**D1.** 最大化 $\sum_i \alpha_i \ln x_i$ s.t. $\sum_i p_i x_i = w$。FOC：

$$\frac{\alpha_i}{x_i} = \lambda p_i.$$

所以

$$x_i = \frac{\alpha_i}{\lambda p_i}.$$

代入预算约束：

$$\sum_i p_i \frac{\alpha_i}{\lambda p_i} = \frac{\sum_i \alpha_i}{\lambda} = w.$$

因 $\sum_i \alpha_i = 1$，得 $\lambda = 1/w$，从而

$$x_i^* = \frac{\alpha_i w}{p_i}.$$

**D2.** 由 Worked Example 4，

$$x_1^* = \theta \bar{X}, \quad x_2^* = (1-\theta)\bar{X}.$$

$\theta$ 越大，计划者越重视个体 1，分给个体 1 的资源比例越高。

**D3.** 两资产权重 $w_1, w_2$，约束 $w_1 + w_2 = 1$，方差

$$\sigma_p^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1 w_2 \sigma_{12}.$$

令 $w_2 = 1 - w_1$，对 $w_1$ 求导：

$$2w_1\sigma_1^2 - 2(1-w_1)\sigma_2^2 + 2(1-2w_1)\sigma_{12} = 0.$$

整理得

$$w_1^* = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}, \quad w_2^* = 1 - w_1^*.$$

**D4.** 问题是

$$\max_{C,G} \ln C + \beta \ln G \quad \text{s.t.} \quad C + G = Y.$$

FOC：

$$\frac{1}{C} = \lambda, \quad \frac{\beta}{G} = \lambda.$$

所以 $G = \beta C$。代入 $C + G = Y$ 得

$$C^* = \frac{Y}{1+\beta}, \quad G^* = \frac{\beta Y}{1+\beta}.$$

公共品支出份额为

$$\frac{G^*}{Y} = \frac{\beta}{1+\beta}.$$

**D5.** 成本函数是

$$C(q) = \min_{K,L} rK + wL \quad \text{s.t.} \quad F(K,L) = q.$$

产量目标 $q$ 增加一点，最小成本的变化率就是 $C'(q)$。按 envelope/影子价格公式，若写 $\mathcal{L} = rK + wL + \mu(q - F(K,L))$，则

$$C'(q) = \mu^*.$$

因此产量约束乘子就是边际成本。

### E 组解答提示

**E1.** 设约束为 $g(x)=b$，正则性给出 $Dg(x^*)$ 满行秩。可行切空间为 $\{h: Dg(x^*)h=0\}$。最优性给出 $\nabla f(x^*)'h=0$ 对所有切方向成立。因此 $\nabla f(x^*)$ 属于 $\mathrm{Null}(Dg(x^*))^\perp$。线性代数基本定理说明

$$\mathrm{Null}(Dg)^\perp = \mathrm{Range}(Dg'),$$

所以存在 $\lambda$ 使得

$$\nabla f(x^*) = Dg(x^*)'\lambda.$$

**E2.** 切空间为 $T=\{h: Dg(x^*)h=0\}$。最大化的二阶充分条件是

$$h'\nabla^2_{xx}\mathcal{L}(x^*,\lambda^*)h<0$$

对所有非零 $h\in T$ 成立。只检查可行方向，是因为约束最优只允许沿约束集合移动，不可行方向上的上升或下降不影响约束局部最优。

**E3.** 见本章影子价格推导。核心步骤是

$$V'(b)=\nabla f(x^*)'x_b^*,$$

FOC 给 $\nabla f=\lambda\nabla g$，约束求导给 $\nabla g' x_b^*=1$，因此 $V'(b)=\lambda$。

**E4.** 由 FOC 得 $w=\Sigma^{-1}(\nu\mathbf{1}+\eta\mu)$。代入两个约束得到

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix}\begin{pmatrix}\nu\\ \eta\end{pmatrix}=\begin{pmatrix}1\\ \bar r\end{pmatrix}.$$

解出

$$\nu=\frac{C-B\bar r}{AC-B^2},\quad \eta=\frac{A\bar r-B}{AC-B^2}.$$

目标值 $w'\Sigma w$ 代入后可化简为

$$\sigma_p^2(\bar r)=\frac{A\bar r^2-2B\bar r+C}{AC-B^2}.$$

**E5.** 例子：

$$f(x,y)=x^2-y^2,\quad g(x,y)=x+y=0.$$

约束上 $y=-x$，所以 $f(x,-x)=0$，所有点目标值相同。也可以构造 $f(x,y)=x^2$ s.t. $y=0$，在 $(0,0)$ FOC 成立且为约束最小，但不是最大；或 $f(x,y)=x^3$ s.t. $y=0$，在 $(0,0)$ FOC 成立但不是局部极值。

## 18 本章总结图谱

| 概念 | 数学表达 | 经济含义 |
|---|---|---|
| 等式约束 | $g(x)=b$ | 资源、预算、技术、会计限制 |
| 可行方向 | $Dg(x^*)h=0$ | 沿着约束集合可以移动的方向 |
| Lagrange FOC | $\nabla f=Dg'\lambda$ | 目标边际变化由约束边际变化线性组合解释 |
| 乘子 | $\lambda_j=\partial V/\partial b_j$ | 第 $j$ 个约束的影子价格 |
| 切空间二阶条件 | $h'H_{\mathcal{L}}h<0$ 或 $>0$ | 沿可行方向判断最大/最小 |
| Bordered Hessian | 加边界的 Hessian 矩阵 | 低维问题中的二阶判别工具 |

# 19 术语表

| 英文术语 | 中文解释 |
|---|---|
| constrained optimization | 约束优化 |
| equality constraint | 等式约束 |
| feasible set | 可行集合 |
| feasible direction | 可行方向 |
| tangent space | 切空间 |
| normal vector | 法向量 |
| Lagrangian | 拉格朗日函数 |
| Lagrange multiplier | 拉格朗日乘子 |
| shadow price | 影子价格 |
| constraint qualification | 约束资格条件/正则性条件 |
| first-order condition | 一阶条件 |
| second-order condition | 二阶条件 |
| bordered Hessian | 加边 Hessian / bordered Hessian |
| value function | 最优值函数 |
| envelope theorem | 包络定理 |
| Marshallian demand | Marshallian 需求 / Walrasian 需求 |
| conditional factor demand | 条件要素需求 |
| mean-variance frontier | 均值-方差前沿 |

# 20 下一章预告

Chapter 15 处理的是必须等号成立的约束。Chapter 16 将进入更真实也更容易出错的情形：不等式约束。

例如消费者问题严格写作应该是

$$p \cdot x \le w, \quad x_i \ge 0,$$

投资组合问题可能有

$$w_i \ge 0$$

表示不能卖空，企业融资问题可能有 collateral constraint，合约理论可能有 limited liability。此时约束可能绑定，也可能不绑定；乘子也不再是任意符号，而会出现

$$\lambda_i \ge 0, \quad \lambda_i g_i(x) = 0.$$

这就是 Kuhn-Tucker 条件和 complementary slackness。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 16
# KKT 条件与不等式约束

Chapter 15 处理了等式约束：预算必须花完，成本约束必须满足，权重必须加总为 1。本章处理更真实的情形：消费不能为负，借款不能超过上限，投资组合不能卖空，企业利润不能低于有限责任边界，抵押约束可能绑定也可能不绑定。这类问题的核心不是“再多一个公式”，而是学会判断约束何时 active，何时 inactive，以及乘子何时为正、何时为零。KKT 条件就是不等式约束最优化的统一语言。

math4econ personalized notes
2026

# 目录

**Chapter 16 总览** **2**

**1 为什么需要 KKT: Lagrange 还不够** **2**

**2 一个一维例子: 为什么边界点不需要导数为零** **3**

**3 KKT 的标准问题与符号约定** **4**

**4 KKT 条件的四块** **5**

4.1 互补松弛的含义 . . . . . 5

**5 Active set: 先猜哪些约束绑定** **6**

**6 Worked Example 1: 非负约束下的一维最大化** **7**

**7 Worked Example 2: 消费者问题中的角点解** **8**

**8 几何解释: 梯度锥与可行方向** **9**

**9 约束资格条件 Constraint Qualification** **10**

9.1 为什么需要正则条件 . . . . . 10

9.2 Slater 条件 . . . . . 10

**10 KKT 的充分性: 什么时候解出来就是全局最优** **10**

**11 Worked Example 3: 借贷约束** **11**

**12 Worked Example 4: 卖空约束下的投资组合** **12**

**13 Worked Example 5: 有限责任与风险选择** **13**

**14 KKT 与包络定理的连接** **14**

**15 多个不等式约束: 如何系统分情况** **14**

**16 KKT 与线性规划的关系** **15**

**17 常见误区与纠正** **15**

**18 本章小结: 一张表** **15**

**19 练习题** **16**

A 组: 基础计算 . . . . . 16

B 组: 概念判断 . . . . . 16

C 组: 混合题 . . . . . 17
D 组: 经济金融建模题 . . . . . 17
E 组: 挑战 / 证明题 . . . . . 17

**20 练习题参考答案** **18**
A 组答案 . . . . . 18
B 组答案 . . . . . 18
C 组答案 . . . . . 18
D 组答案 . . . . . 19
E 组答案 . . . . . 20

**21 后续学习连接** **20**

# Chapter 16 总览

| | |
|---|---|
| **本章主题** | Kuhn–Tucker / KKT 条件, 不等式约束, 角点解, 互补松弛。 |
| **核心问题** | 当约束可能绑定也可能不绑定时, 如何同时刻画内点解和角点解? 为什么有些约束乘子为正, 有些为零? |
| **数学工具** | primal feasibility, dual feasibility, complementary slackness, stationarity, active set, constraint qualification, convex sufficiency。 |
| **经济应用** | 非负消费、借贷约束、卖空约束、抵押约束、有限责任、投资组合权重约束、产能上限、配给。 |
| **后续连接** | Chapter 17 凸性与全局最优, Chapter 18 隐函数与比较静态, Chapter 19 包络定理与对偶理论, 高级微观中的消费者/生产者理论。 |

**直觉盒子**

KKT 的一句话直觉是: **每个不等式约束只有两种状态**。

要么约束绑定, 乘子可能为正; 要么约束不绑定, 乘子必须为零。

这就是互补松弛 complementary slackness。它是理解角点解、借贷约束、卖空约束和有限责任的核心。

## 1 为什么需要 KKT: Lagrange 还不够

Chapter 15 中的等式约束问题通常长这样:

$$\max_x f(x) \quad \text{s.t.} \quad h(x) = 0.$$

等式约束没有“松不松”的问题。它必须成立, 所以 Lagrange 乘子可以直接进入一阶条件。

但现实经济问题常常是:

$$\max_x f(x) \quad \text{s.t.} \quad g(x) \le 0.$$

这时约束可能绑定, 也可能不绑定。比如:

- 消费者有非负消费约束 $x_i \ge 0$, 但最优时可能买正数量, 也可能完全不买;
- 家庭有借款约束 $b \ge -\bar{b}$, 但高收入家庭可能根本不受约束;
- 投资组合有卖空限制 $w_i \ge 0$, 但某个资产可能权重为正, 也可能被排除在组合之外;
- 企业有有限责任约束 payoff $\ge 0$, 坏状态下它可能绑定。

KKT 条件的任务是把这些情况放进同一个系统。

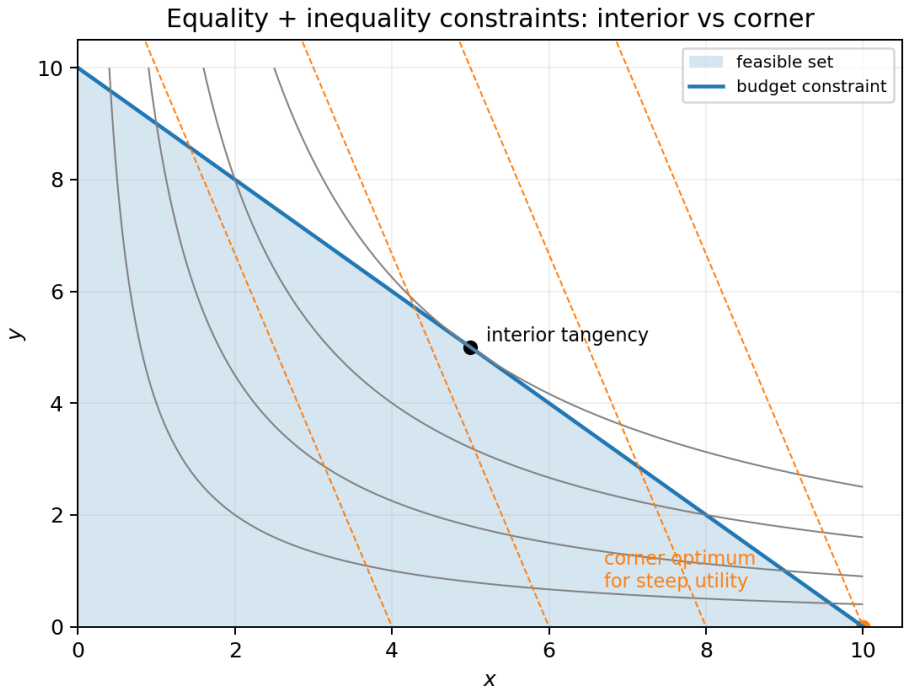



图 1：等式约束给出预算线，不等式约束给出可行区域。最优点可能是内点，也可能是角点。

**常见误区**

初学者最常见错误是：看到最优化就机械写“导数等于零”。这是内点解的条件，不是角点解的条件。若最优点在边界上，目标函数沿被禁止方向可能仍然想继续改进，只是约束不允许它继续走。

## 2 一个一维例子：为什么边界点不需要导数为零

考虑

$$\max_{x\geq 0} h(x) = -(x+1)^2 + 4.$$

如果没有 $x \geq 0$ 约束，一阶条件为

$$h'(x) = -2(x+1) = 0,$$

得到无约束最优点 $x = -1$。但 $x = -1$ 不可行。可行集合是 $[0, \infty)$，所以最优点是边界

$$x^* = 0.$$

此时

$$h'(0) = -2 \neq 0.$$

这不是错误，而是角点解的本质。

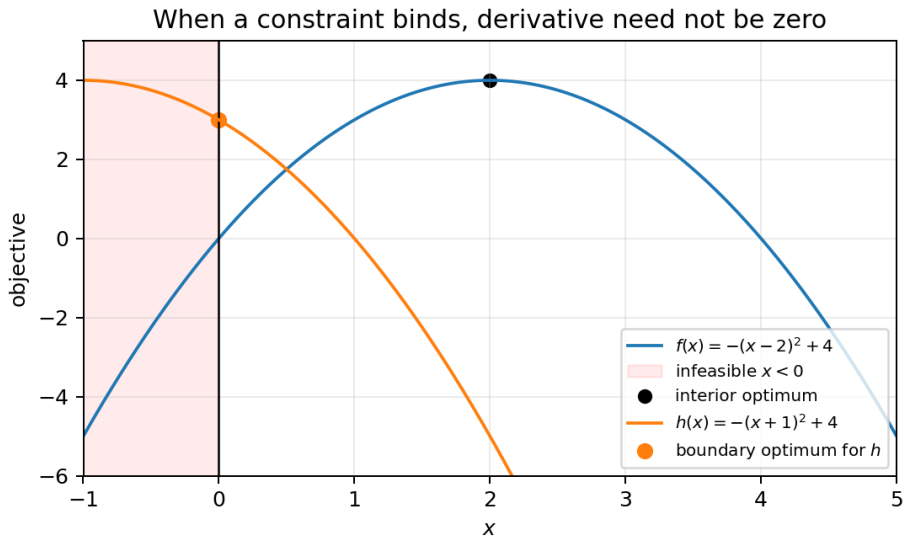



图 2: 内点最优需要导数为零; 边界最优只要求可行方向上不能改进。

**直觉盒子**

边界最优的经济含义: 如果没有约束, 你还想往不可行方向走; 但可行集合挡住了你。因此边界约束有价值, 它的乘子正是这种“挡住的价值”。

## 3 KKT 的标准问题与符号约定

本章采用一个固定符号约定, 避免乘子正负混乱。

**定义 / 定理**

**最大化问题的 KKT 约定.** 考虑

$$\max_{x\in\mathbb{R}^n} f(x)$$

subject to

$$g_i(x) \le 0, \quad i = 1, \dots, m,$$

$$h_j(x) = 0, \quad j = 1, \dots, r.$$

本章使用 Lagrangian

$$\boxed{\mathcal{L}(x, \lambda, \mu) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{r} \mu_j h_j(x),}$$

其中

$$\lambda_i \ge 0.$$

为什么最大化时写成 $-\lambda_i g_i$? 因为约束是 $g_i(x) \le 0$。如果约束放松, 允许 $g_i(x)$ 更大, 可行集合扩大, 最大值不应下降。用这个符号时, $\lambda_i$ 可以自然解释为放松约束的影子价值。

**常见误区**

不同教材有不同符号约定。有些教材写最小化问题并使用 $\mathcal{L} = f + \lambda g$, 有些把约束写成 $g(x) \geq 0$。不要死背乘子正负, 要先确认: 问题是最大化还是最小化? 约束写成 $\leq 0$ 还是 $\geq 0$? Lagrangian 前面是加还是减?

# 4 KKT 条件的四块

**定义 / 定理**

**KKT 条件, 最大化版本.** 在合适的正则条件下, 若 $x^*$ 是局部最优解, 则存在乘子 $\lambda_i \geq 0$ 和 $\mu_j \in \mathbb{R}$, 使得:

**1. Primal feasibility 原始可行性**

$$g_i(x^*) \leq 0, \qquad h_j(x^*) = 0.$$

**2. Dual feasibility 对偶可行性**

$$\lambda_i \geq 0.$$

**3. Complementary slackness 互补松弛**

$$\lambda_i g_i(x^*) = 0, \qquad i = 1, \ldots, m.$$

**4. Stationarity 驻点条件**

$$\nabla_x \mathcal{L}(x^*, \lambda, \mu) = 0.$$

这四块缺一不可。

## 4.1 互补松弛的含义

互补松弛条件

$$\lambda_i g_i(x^*) = 0$$

看起来很短, 但信息量最大。因为 $\lambda_i \geq 0$ 且 $g_i(x^*) \leq 0$, 它意味着:

- 若 $g_i(x^*) < 0$, 约束不绑定, 则必须 $\lambda_i = 0$;
- 若 $\lambda_i > 0$, 则必须 $g_i(x^*) = 0$, 约束绑定;
- 若 $g_i(x^*) = 0$, 乘子可以为正, 也可以为零, 取决于这个约束是否真的有边际价值。

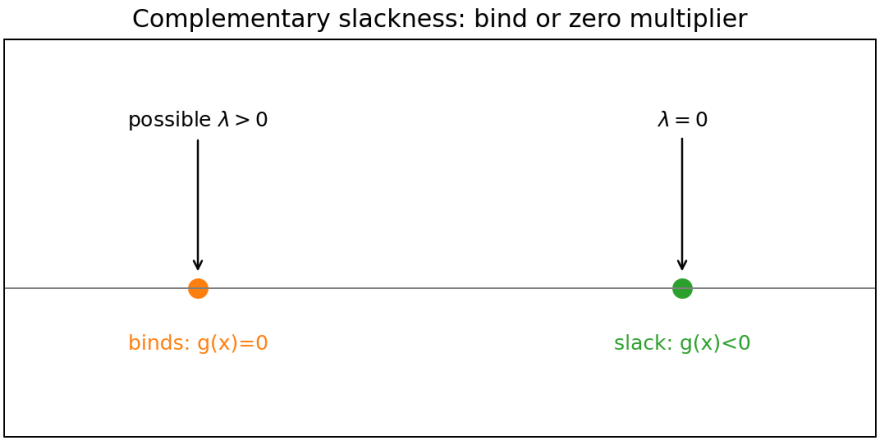



图 3: 互补松弛: slack constraint has zero multiplier; positive multiplier implies binding constraint.

**直觉盒子**

经济解释非常直接: 如果一个约束没有用尽, 放松它一点没有价值, 所以乘子为零。如果一个约束正在限制选择, 放松它可能提高目标值, 所以乘子可能为正。

## 5 Active set: 先猜哪些约束绑定

KKT 问题通常不能盲目一次解完。实操中常用 active set 思路。

**定义 / 定理**

**Active constraint.** 对约束 $g_i(x) \le 0$, 若

$$g_i(x^*) = 0,$$

则称它在最优点 active 或 binding。若

$$g_i(x^*) < 0,$$

则称它 inactive 或 slack。

Active set 方法就是:

1. 先猜哪些不等式约束绑定;
2. 对绑定约束当成等式处理;
3. 对不绑定约束设置对应乘子为 0;
4. 解 FOC;
5. 回头检查可行性、乘子符号和互补松弛。

**方法模板**

**KKT 解题模板**

1. 统一写成最大化且 $g_i(x) \le 0$。
2. 写 Lagrangian: $\mathcal{L} = f - \sum_i \lambda_i g_i + \sum_j \mu_j h_j$。
3. 写四类 KKT 条件: feasibility, dual feasibility, complementary slackness, stationarity。
4. 猜 active set, 分情况求解。
5. 检查所有约束和乘子符号。
6. 若问题是 concave maximization over convex feasible set, KKT 解就是全局最优。

## 6 Worked Example 1: 非负约束下的一维最大化

考虑

$$\max_{x \ge 0} f(x) = -(x+1)^2 + 4.$$

把约束写成

$$g(x) = -x \le 0.$$

最大化版本 Lagrangian 为

$$\mathcal{L}(x, \lambda) = f(x) - \lambda g(x) = -(x+1)^2 + 4 + \lambda x.$$

KKT 条件为:

$$\text{primal feasibility: } -x \le 0 \iff x \ge 0,$$

$$\text{dual feasibility: } \lambda \ge 0,$$

$$\text{complementary slackness: } \lambda(-x) = 0 \iff \lambda x = 0,$$

$$\text{stationarity: } \frac{\partial \mathcal{L}}{\partial x} = -2(x+1) + \lambda = 0.$$

互补松弛给出两种情况。

**情况 1: 内点 $x > 0$.** 则 $\lambda = 0$。stationarity 给出

$$-2(x+1) = 0 \implies x = -1,$$

不可行, 所以内点不可能。

**情况 2: 边界 $x = 0$.** stationarity 给出

$$-2(0+1) + \lambda = 0,$$

所以

$$\lambda = 2 > 0.$$

所有 KKT 条件满足, 所以

$$\boxed{x^* = 0, \quad \lambda^* = 2.}$$

**直觉盒子**

乘子 $\lambda = 2$ 的意思是: 非负约束正在限制你。如果把约束从 $x \geq 0$ 放松为 $x \geq -\varepsilon$, 最大值的一阶增量大约是 $2\varepsilon$。

## 7 Worked Example 2: 消费者问题中的角点解

考虑消费者问题:

$$\max_{x,y\geq 0} \; u(x,y) = \ln(x+1) + 2\ln(y+1)$$

subject to

$$p_x x + p_y y \leq w.$$

为简单起见, 令 $p_x = p_y = 1$, $w = 1$。问题为

$$\max_{x,y} \; \ln(x+1) + 2\ln(y+1)$$

subject to

$$x + y - 1 \leq 0, \qquad -x \leq 0, \qquad -y \leq 0.$$

Lagrangian:

$$\mathcal{L} = \ln(x+1) + 2\ln(y+1) - \lambda(x+y-1) + \mu_x x + \mu_y y.$$

其中 $\lambda, \mu_x, \mu_y \geq 0$。

Stationarity:

$$\frac{1}{x+1} - \lambda + \mu_x = 0,$$

$$\frac{2}{y+1} - \lambda + \mu_y = 0.$$

Complementary slackness:

$$\lambda(x+y-1) = 0, \qquad \mu_x x = 0, \qquad \mu_y y = 0.$$

由于效用对两种商品均严格递增, 预算必然绑定:

$$x + y = 1, \qquad \lambda > 0.$$

先猜内点 $x>0, y>0$。则 $\mu_x=\mu_y=0$, FOC 给出

$$\lambda=\frac{1}{x+1}=\frac{2}{y+1}.$$

因此

$$y+1=2(x+1)=2x+2,$$

即

$$y=2x+1.$$

但预算 $x+y=1$ 给出

$$x+2x+1=1 \quad\Longrightarrow\quad x=0,$$

这不是内点。说明真实解在 $x=0$ 的角点。

令 $x=0$, 则预算给出 $y=1$。Stationarity 第二式中 $y>0$, 所以 $\mu_y=0$:

$$\frac{2}{2}-\lambda=0 \quad\Longrightarrow\quad \lambda=1.$$

第一式给出

$$1-1+\mu_x=0 \quad\Longrightarrow\quad \mu_x=0.$$

所以一个 KKT 解是

$$\boxed{x^*=0,\quad y^*=1,\quad \lambda=1,\quad \mu_x=0,\quad \mu_y=0.}$$

这个例子有一个细节: $x=0$ 绑定, 但 $\mu_x=0$。这说明绑定约束不一定有正乘子。消费者刚好在边界处无差异地不买 $x$。

**常见误区**

互补松弛只说 $\lambda_i>0\Rightarrow g_i=0$ 和 $g_i<0\Rightarrow\lambda_i=0$。它不说 $g_i=0\Rightarrow\lambda_i>0$。约束绑定但乘子为零是可能的, 这种情况叫 weakly binding。

## 8 几何解释: 梯度锥与可行方向

等式约束中, 最优点处目标函数梯度必须落在约束梯度张成的线性空间中。KKT 中, 不等式约束带来的不是线性空间, 而是一个由 active constraints 形成的锥 cone。

若最大化 $f(x)$, active 约束为 $g_i(x)\le 0$, KKT stationarity 写成

$$\nabla f(x^*)=\sum_i \lambda_i\nabla g_i(x^*)-\sum_j \mu_j\nabla h_j(x^*),\qquad \lambda_i\ge 0.$$

这表示目标梯度不能指向任何可行改进方向。如果它指向某个不可行方向, active 约束的法向量会“挡住”它。

**直觉盒子**

KKT 的几何语言是：在最优点，没有任何可行方向能让目标函数继续上升。若目标函数还想往外走，那个方向正好被绑定约束挡住。乘子衡量每个挡板的重要性。

# 9 约束资格条件 Constraint Qualification

KKT 条件不是无条件成立。若约束形状太病态，最优点处可能不存在有意义的乘子。为了保证 KKT 必要性，需要 constraint qualification。

## 9.1 为什么需要正则条件

想象一个约束集合在最优点形成尖点、交叉、重叠或梯度消失。此时 active 约束的梯度可能不能正确描述可行方向，乘子系统就可能失效。

**定义 / 定理**

**LICQ, linear independence constraint qualification.** 在候选点 $x^*$，所有 active inequality constraints 的梯度和所有 equality constraints 的梯度线性无关。即

$$\{\nabla g_i(x^*) : g_i(x^*) = 0\} \cup \{\nabla h_j(x^*)\}$$

线性无关。

LICQ 是最常见、最好理解的 CQ。它说 active constraints 不能重复、不能退化，否则乘子可能不唯一或不存在。

## 9.2 Slater 条件

在凸优化中，更常见的是 Slater 条件。

**定义 / 定理**

**Slater condition, 直观版.** 对凸约束 $g_i(x) \le 0$ 和仿射等式约束，如果存在一个严格可行点 $\bar{x}$，使得

$$g_i(\bar{x}) < 0 \quad \text{for all inequality constraints,}$$

并满足等式约束，则 Slater 条件成立。

Slater 条件的经济含义是：可行集合内部不是空的，约束不是把选择空间压成奇怪的边界。

# 10 KKT 的充分性：什么时候解出来就是全局最优

KKT 条件一般是局部必要条件，不是所有问题的充分条件。但在凸优化结构下，它非常强。

**定义 / 定理**

**Concave maximization 的 KKT 充分性.** 若

- $f$ 是 concave objective;
- inequality constraints 写为 $g_i(x) \le 0$, 且 $g_i$ 是 convex, 因而可行集合凸;
- equality constraints 是 affine;

则任何满足 KKT 条件的点都是全局最大值。若 $f$ 严格凹, 全局最大值唯一。

**常见误区**

最大化问题中, 目标函数应该是凹函数 concave 才是好问题。最小化问题中, 目标函数应该是凸函数 convex 才是好问题。不要把 convex/concave 的方向搞反。

## 11 Worked Example 3: 借贷约束

考虑两期消费问题。消费者最大化

$$u(c_1) + \beta u(c_2)$$

subject to present-value budget

$$c_1 + \frac{c_2}{1+r} \le y_1 + \frac{y_2}{1+r},$$

以及借贷约束

$$c_1 \le y_1 + \bar{b}.$$

这里 $\bar{b} \ge 0$ 表示最多可借多少。

把约束写成 $g \le 0$:

$$g_1(c) = c_1 + \frac{c_2}{1+r} - y_1 - \frac{y_2}{1+r} \le 0,$$

$$g_2(c) = c_1 - y_1 - \bar{b} \le 0.$$

Lagrangian:

$$\mathcal{L} = u(c_1) + \beta u(c_2) - \lambda g_1(c) - \eta g_2(c),$$

其中 $\lambda, \eta \ge 0$。

FOC:

$$u'(c_1) - \lambda - \eta = 0,$$

$$\beta u'(c_2) - \frac{\lambda}{1+r} = 0.$$

即

$$u'(c_1) = \lambda + \eta,$$

$$\beta(1+r)u'(c_2) = \lambda.$$

消去 $\lambda$:

$$\boxed{u'(c_1) = \beta(1+r)u'(c_2) + \eta.}$$

互补松弛:

$$\eta(c_1 - y_1 - \bar{b}) = 0.$$

若借贷约束不绑定, $c_1 < y_1 + \bar{b}$, 则 $\eta = 0$, 得到标准 Euler equation:

$$u'(c_1) = \beta(1+r)u'(c_2).$$

若借贷约束绑定, $\eta > 0$, 则

$$u'(c_1) > \beta(1+r)u'(c_2).$$

经济含义: 消费者想把更多未来资源搬到今天, 但借贷约束阻止了他。因此今天消费的边际效用相对更高。

**直觉盒子**

借贷约束乘子 $\eta$ 是 Euler equation wedge。它不是噪音, 而是约束造成的边际条件扭曲。高级宏观里大量模型都在研究这种 wedge。

## 12 Worked Example 4: 卖空约束下的投资组合

设两个风险资产权重为 $w$ 和 $1-w$, 且不允许卖空第一只资产:

$$w \geq 0.$$

假设投资者的目标可简化为

$$\max_{w} U(w) = -(w + 0.25)^2 + 1$$

subject to

$$w \geq 0.$$

无约束最优为

$$U'(w) = -2(w + 0.25) = 0 \quad \Longrightarrow \quad w = -0.25.$$

但 $w = -0.25$ 是卖空, 不可行。所以约束下最优为

$$w^* = 0.$$

KKT 写法: 约束 $g(w) = -w \le 0$, Lagrangian

$$\mathcal{L} = -(w + 0.25)^2 + 1 - \lambda(-w) = -(w + 0.25)^2 + 1 + \lambda w.$$

Stationarity:

$$-2(w + 0.25) + \lambda = 0.$$

在 $w = 0$ 时,

$$-0.5 + \lambda = 0,$$

所以

$$\lambda = 0.5 > 0.$$

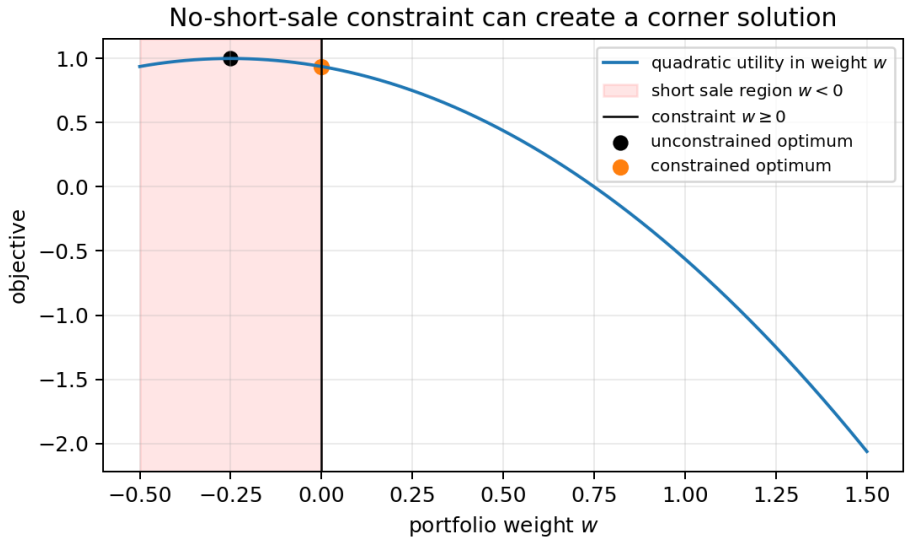



图 4: 卖空约束可能把无约束最优点推到边界, 从而产生角点组合。

## 13 Worked Example 5: 有限责任与风险选择

有限责任 limited liability 常写成 payoff 不能低于 0。考虑企业选择风险强度 $e \ge 0$, 项目成功收益为 $R(e)$, 失败时债权人承担部分损失, 股东 payoff 具有下界。一个极简数学表达是:

$$\max_{e \ge 0} \; V(e)$$

subject to

$$V(e) \ge 0.$$

若把约束写成

$$g(e) = -V(e) \le 0,$$

KKT Lagrangian 为

$$\mathcal{L} = V(e) - \lambda[-V(e)] = (1+\lambda)V(e), \quad \lambda \geq 0.$$

如果有限责任约束不绑定, $V(e) > 0$, 则 $g(e) < 0$, 所以 $\lambda = 0$。若坏状态下 $V(e) = 0$, 则 $\lambda$ 可能为正。

经济上, 有限责任使得企业或股东对下行损失不完全负责, 可能扭曲风险选择。KKT 乘子可以解释为下界约束的边际价值或扭曲强度。

**常见误区**

真实有限责任模型通常需要状态 contingent payoff, 债务面值, 违约边界和激励约束。本例只展示 KKT 语言如何处理 payoff 下界, 不代表完整公司金融模型。

## 14 KKT 与包络定理的连接

考虑带参数的约束问题:

$$V(b) = \max_x f(x) \quad \text{s.t.} \quad g(x) \leq b.$$

写成

$$g(x) - b \leq 0.$$

最大化 Lagrangian:

$$\mathcal{L} = f(x) - \lambda[g(x) - b] = f(x) - \lambda g(x) + \lambda b.$$

在正则条件下, 包络定理给出

$$\boxed{\frac{\mathrm{d}V}{\mathrm{d}b} = \lambda^*.}$$

这就是乘子的影子价格解释: 放松约束右端 $b$ 一点, 最优值增加约 $\lambda^*$。

如果约束不绑定, $\lambda^* = 0$, 放松约束没有价值。如果约束绑定且有边际价值, $\lambda^* > 0$。

## 15 多个不等式约束: 如何系统分情况

当有多个不等式约束时, active set 数量可能很多。如果有 $m$ 个约束, 理论上有 $2^m$ 种 active/inactive 组合。但经济结构通常可以帮助减少情况。

**方法模板**

**多约束 KKT 的实用策略**

1. 先用经济直觉判断哪些约束可能绑定。例如效用递增通常意味着预算约束绑定。

2. 对显然不可能绑定的约束, 先设乘子为零。
3. 对可能绑定的核心约束分情况。
4. 每个候选解都必须回查: 原始可行性、乘子非负、互补松弛。
5. 若目标凹且可行集凸, 找到的 KKT 解就是全局解。

## 16 KKT 与线性规划的关系

线性规划 linear programming 是 KKT 的重要特例:

$$\max_x c^T x \quad \text{s.t.} \quad Ax \le b, \quad x \ge 0.$$

在线性规划中, 目标和约束都是线性的, 可行集合是多面体。最优解通常出现在顶点, 所以角点解不是异常, 而是常态。

KKT 条件在这里变成 primal feasibility, dual feasibility 和 complementary slackness 的代数系统。这也是对偶理论和 shadow price 的来源之一。

**直觉盒子**

线性规划告诉你: 经济学里的角点不是麻烦, 而是一种结构。配给、产能、非负生产、不可卖空、预算上限, 都会自然产生顶点解。

## 17 常见误区与纠正

**常见误区**

**误区 1: 绑定约束一定有正乘子**。错误。绑定约束可能 weakly binding, 乘子为零。

**误区 2: 不绑定约束可以有正乘子**。错误。互补松弛要求 slack constraint 的乘子必须为零。

**误区 3: KKT 条件永远充分**。错误。一般非凸问题中, KKT 可能只是局部必要条件。凹最大化/凸可行集下才有强充分性。

**误区 4: 乘子符号可以随便写**。错误。乘子符号取决于最大化/最小化和约束方向。

**误区 5: 一看到角点就不会做**。角点正是 KKT 的优势。用 active set 分情况, 再回查即可。

## 18 本章小结: 一张表

| 概念 | 公式 | 经济含义 |
|---|---|---|
| Primal feasibility | $g_i(x^*) \le 0, h_j(x^*) = 0$ | 选择必须可行。 |
| Dual feasibility | $\lambda_i \ge 0$ | 放松约束的价值非负。 |
| Complementary slackness | $\lambda_i g_i(x^*) = 0$ | 不绑定约束无影子价值。 |
| Stationarity | $\nabla_x \mathcal{L} = 0$ | 没有可行的一阶改进方向。 |
| Active constraint | $g_i(x^*) = 0$ | 约束刚好绑定。 |
| Inactive constraint | $g_i(x^*) < 0$ | 约束有余量, 乘子为零。 |
| Constraint qualification | LICQ, Slater | 保证 KKT 乘子存在的正则条件。 |
| Convex sufficiency | concave max + convex feasible | KKT 解是全局最优。 |
| Shadow price | $\mathrm{d}V/\mathrm{d}b = \lambda$ | 放松约束的边际价值。 |

## 19 练习题

### A 组: 基础计算

A1. 把约束 $x \ge 0$ 写成 $g(x) \le 0$ 的形式。

A2. 对问题 $\max_{x \ge 0} -(x-2)^2$, 写 KKT 条件并求解。

A3. 对问题 $\max_{x \ge 0} -(x+3)^2$, 写 KKT 条件并求解。

A4. 对约束 $x + y \le 10$, $x \ge 0$, $y \ge 0$, 全部写成 $g_i(x, y) \le 0$。

A5. 解释互补松弛 $\lambda_i g_i(x^*) = 0$ 的两种情况。

A6. 判断: 若 $g_i(x^*) = 0$, 是否必有 $\lambda_i > 0$?

### B 组: 概念判断

B1. KKT 条件和 Lagrange 乘子法有什么关系?

B2. 为什么角点解处目标函数导数可能不为零?

B3. 为什么不绑定的借贷约束乘子必须为零?

B4. 什么是 constraint qualification? 为什么需要它?

B5. 在 concave maximization 中, KKT 为什么可以给出全局最优?

### C 组：混合题

C1. 求解 $\max_{x\geq 1} -(x-3)^2+5$。

C2. 求解 $\max_{x\geq 1} -(x-0)^2$。

C3. 求解 $\max_{x,y\geq 0} \ln x + \ln y$ subject to $x+y\leq 10$。

C4. 求解 $\max_{x,y\geq 0} x+2y$ subject to $x+y\leq 1$。

C5. 对 $\max_x f(x)$ subject to $g(x)\leq b$, 推导 $\mathrm{d}V/\mathrm{d}b=\lambda$ 的直觉。

### D 组：经济金融建模题

D1. 两期消费模型中, 借贷约束绑定时 Euler equation 如何改变? 解释 wedge。

D2. 投资组合问题中, 若无约束最优权重 $w_i<0$, 但有 $w_i\geq 0$ 约束, KKT 如何解释角点 $w_i=0$?

D3. 企业有产能约束 $q\leq \bar{q}$。写出利润最大化问题的 KKT 条件, 并解释产能约束乘子。

D4. 消费者有最低消费需求 $x\geq \underline{x}$。若该约束绑定, 乘子表示什么?

D5. 有限责任如何通过不等式约束改变最优风险选择?

### E 组：挑战 / 证明题

E1. 证明: 在 concave maximization with convex feasible set 中, KKT 条件充分。

E2. 给出一个绑定约束但乘子为零的例子。

E3. 对 $\max f(x)$ subject to $g(x)\leq 0$, $h(x)=0$, 推导 KKT stationarity 的几何含义。

E4. 解释 LICQ 失败时乘子为什么可能不唯一。

E5. 比较最大化约束 $g(x)\leq 0$ 与最小化约束 $g(x)\leq 0$ 的 Lagrangian 符号。

# 20 练习题参考答案

## A 组答案

A1. $x \geq 0$ 等价于 $-x \leq 0$, 所以可设 $g(x) = -x$。

A2. 问题为 $\max_{x \geq 0} f(x) = -(x-2)^2$。约束 $g = -x \leq 0$。Lagrangian:

$$\mathcal{L} = -(x-2)^2 + \lambda x.$$

KKT:

$$x \geq 0, \quad \lambda \geq 0, \quad \lambda x = 0,$$

$$-2(x-2) + \lambda = 0.$$

内点 $x > 0$ 时 $\lambda = 0$, 得 $x = 2$, 可行。所以 $x^* = 2, \lambda = 0$。

A3. $f(x) = -(x+3)^2$, Lagrangian $\mathcal{L} = -(x+3)^2 + \lambda x$。KKT stationarity:

$$-2(x+3) + \lambda = 0.$$

内点给 $x = -3$ 不可行。边界 $x = 0$ 时 $\lambda = 6 > 0$。所以 $x^* = 0$。

A4. 可写为

$$g_1 = x + y - 10 \leq 0, \quad g_2 = -x \leq 0, \quad g_3 = -y \leq 0.$$

A5. 若 $g_i(x^*) < 0$, 约束不绑定, 则 $\lambda_i = 0$。若 $\lambda_i > 0$, 则必须 $g_i(x^*) = 0$, 约束绑定。

A6. 不一定。绑定约束可以有零乘子, 称为 weakly binding。

## B 组答案

B1. Lagrange 乘子法处理等式约束。KKT 是它对不等式约束的推广, 增加了 dual feasibility 和 complementary slackness。

B2. 因为边界点处某些方向不可行。目标函数可能仍然想往不可行方向增加, 但约束挡住了它, 所以不需要普通导数为零。

B3. 如果借贷约束不绑定, 放松它一点不会改变最优选择或最优值, 所以影子价格为零。

B4. Constraint qualification 是保证约束梯度正确描述可行方向的正则条件。没有它, KKT 乘子可能不存在或不唯一。

B5. 因为凹目标加凸可行集使任何局部最优都是全局最优。KKT 排除所有可行一阶改进方向, 凹性保证没有隐藏的更好点。

## C 组答案

C1. 无约束最优 $x = 3$, 满足 $x \geq 1$, 所以 $x^* = 3$, 约束不绑定, 乘子为 0。

C2. 无约束最优 $x=0$, 不满足 $x\geq 1$, 所以 $x^*=1$。写约束 $1-x\leq 0$, 最大化 Lagrangian

$$\mathcal{L}=-x^2-\lambda(1-x)=-x^2-\lambda+\lambda x.$$

stationarity:

$$-2x+\lambda=0.$$

在 $x=1$ 时 $\lambda=2>0$。

C3. 效用递增, 预算绑定。内点 FOC:

$$\frac{1}{x}=\lambda,\quad \frac{1}{y}=\lambda.$$

所以 $x=y$。由 $x+y=10$, 得 $x^*=y^*=5$。非负约束不绑定。

C4. 线性目标 $x+2y$ 在预算线上最大化时应把全部资源给边际收益更高的 $y$。所以 $x^*=0, y^*=1$。这是角点解。

C5. Lagrangian 为 $\mathcal{L}=f(x)-\lambda[g(x)-b]=f(x)-\lambda g(x)+\lambda b$。包络定理给出 $V_b=\partial\mathcal{L}/\partial b=\lambda$。

## D 组答案

D1. 无约束 Euler equation 为 $u'(c_1)=\beta(1+r)u'(c_2)$。借贷约束绑定时出现乘子 $\eta>0$:

$$u'(c_1)=\beta(1+r)u'(c_2)+\eta.$$

$\eta$ 是 borrowing wedge, 表示消费者想把更多未来资源转移到当期但受限。

D2. 约束 $w_i\geq 0$ 可写成 $-w_i\leq 0$。若无约束最优 $w_i<0$, 受卖空约束后最优可能是 $w_i=0$。对应乘子为正, 表示允许卖空会提高目标值。

D3. 问题可写为

$$\max_q \pi(q)=pq-C(q)\quad \text{s.t.}\quad q-\bar{q}\leq 0.$$

Lagrangian:

$$\mathcal{L}=pq-C(q)-\lambda(q-\bar{q}).$$

FOC:

$$p-C'(q)-\lambda=0.$$

若产能绑定, $q=\bar{q}$ 且 $\lambda=p-C'(q)>0$, 表示提高产能上限的边际利润。

D4. 最低消费约束绑定说明消费者被迫消费至少 $\underline{x}$。乘子表示放松最低要求、允许减少该消费的边际价值; 符号需根据约束写法解释。

D5. 有限责任给 payoff 加下界, 使决策者不完全承担下行风险。KKT 乘子刻画下界约束的边际价值, 它可能导致风险选择偏离社会最优。

### E 组答案

E1. 设 $x^*$ 满足 KKT。对任意可行 $x$, concavity 给出

$$f(x) \le f(x^*) + \nabla f(x^*)^T (x - x^*).$$

Stationarity 表示 $\nabla f(x^*)$ 可由 active constraints 的法向量表示。结合 convex constraints 和 complementary slackness 可推出右侧不超过 $f(x^*)$。因此 $f(x) \le f(x^*)$, 故全局最优。

E2. 例: $\max_{x \ge 0} -x^2$。最优 $x^* = 0$, 约束绑定。KKT stationarity: $-2x + \lambda = 0$, 在 $x = 0$ 得 $\lambda = 0$。所以绑定但乘子为零。

E3. Stationarity 表示目标函数梯度可由 active inequality constraints 和 equality constraints 的梯度线性组合表示, 且 inequality multipliers 非负。这说明目标梯度指向的改进方向被可行集边界挡住。

E4. 若两个 active constraints 梯度线性相关, 同一个法向量可以用多组乘子表示, 因而乘子可能不唯一。例如约束 $x \le 0$ 和 $2x \le 0$ 在 $x = 0$ 同时 active, 梯度共线。

E5. 对最大化 $\max f$ with $g \le 0$, 本章用 $\mathcal{L} = f - \lambda g$, $\lambda \ge 0$。对最小化 $\min f$ with $g \le 0$, 常用 $\mathcal{L} = f + \lambda g$, $\lambda \ge 0$。二者符号相反, 因为优化方向相反。

## 21 后续学习连接

Chapter 17 将进入凸性、凹性与经济最优化。KKT 在非凸问题中只是必要条件, 但在凸/凹结构下会变成强大的全局工具。因此下一章的重点是回答:

什么时候局部条件足以保证全局最优?

这对高级微观、金融优化、机制设计和计量估计都非常关键。

**直觉盒子**

如果 Chapter 15 的核心是“等式约束的影子价格”, Chapter 16 的核心就是“不等式约束的开关机制”。KKT 把每个约束变成一个开关: 绑定时可能有价格, 不绑定时价格为零。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 17
# 凸性、凹性与经济最优化

前两章我们已经学会了 Lagrange 与 KKT: 如何写一阶条件, 如何处理等式约束与不等式约束。本章回答更深的问题: **什么时候这些一阶条件足以保证全局最优?** 答案主要来自凸性与凹性。凸集保证可行域没有“断裂”和“凹陷”; 凹目标保证最大化不会陷入坏的局部峰; 凸目标保证最小化不会被局部谷误导。对经济学而言, 凸性是消费者理论、生产理论、一般均衡、博弈论、金融组合优化和对偶理论的共同地基。

math4econ personalized notes
2026

# 目录

**Chapter 17 总览** **2**

**1 为什么经济学如此重视凸性** **2**

**2 凸集: 可行域的几何结构** **3**
2.1 定义 . . . 3
2.2 经济解释 . . . 3
2.3 凸集的交仍是凸集 . . . 4

**3 凸函数与凹函数** **4**
3.1 定义与图像 . . . 5
3.2 经济直觉 . . . 5
3.3 严格凸与严格凹 . . . 5

**4 Hessian 判别: 与 Chapter 14 的连接** **6**

**5 局部最优与全局最优** **6**
5.1 证明: 为什么凹性让一阶条件变充分 . . . 7

**6 凸优化与 KKT 充分性** **8**

**7 Quasi-concavity 与偏好凸性** **8**
7.1 凹性与 quasi-concavity 的关系 . . . 9

**8 Jensen 不等式与风险厌恶** **10**
8.1 Jensen 不等式 . . . 10
8.2 风险厌恶解释 . . . 10

**9 分离超平面: 价格支持的几何入口** **11**

**10 金融应用: 投资组合分散化与凸性** **12**
10.1 组合方差是凸函数 . . . 12

**11 Worked Examples** **13**

**12 常见错误与辨析** **15**

**13 本章小结: 概念 - 公式 - 应用** **15**

**14 练习题** **16**
A 组: 基础计算 . . . 16
B 组: 概念判断 . . . 16

C 组: 混合题 . . . . . 16
D 组: 经济金融建模题 . . . . . 16
E 组: 挑战 / 证明题 . . . . . 17

**15 练习题参考答案** **18**
A 组答案 . . . . . 18
B 组答案 . . . . . 18
C 组答案 . . . . . 19
D 组答案 . . . . . 19
E 组答案 . . . . . 19

**16 后续学习连接** **20**

# Chapter 17 总览

| | |
|---|---|
| **本章主题** | convexity, concavity, quasi-concavity, global optimum, Jensen inequality, separating hyperplane intuition。 |
| **核心问题** | 为什么有些最优化问题一阶条件就足够, 而有些问题会出现多个局部最优? 为什么经济学偏好凸性和生产集凸性如此重要? 为什么风险厌恶和 Jensen 不等式有关? |
| **数学工具** | convex set, convex/concave function, strict convexity, Hessian tests, quasi-concavity, upper contour set, Jensen inequality, first-order concavity inequality, separating hyperplane。 |
| **经济应用** | 消费者偏好凸性, 风险厌恶, 成本函数凸性, 利润最大化, 投资组合分散化, social welfare aggregation, duality 预备。 |
| **后续连接** | Chapter 18 隐函数定理与比较静态, Chapter 19 包络定理与对偶理论, Chapter 20 固定点与均衡存在性, 高级微观中的凸分析。 |

**直觉盒子**

本章的核心直觉是: **凸性让局部信息变成全局信息**。在非凸世界里, 一阶条件只告诉你“这里斜率为零”, 但可能只是局部最优、局部最差或鞍点。在凸/凹世界里, 曲率结构足够好, 所以一阶条件往往可以直接推出全局最优。

## 1 为什么经济学如此重视凸性

在经济学里, “凸性”不是纯数学审美, 而是模型可解性、均衡存在性和比较静态的基础。

- **消费者理论:** 偏好凸性意味着消费者喜欢多样化, upper contour set 是凸的。这样需求问题通常有良好解。
- **生产理论:** 生产集凸性意味着可以混合不同生产计划, 规模平均不会变得更差。
- **金融理论:** 投资组合权重集合通常是凸集, 分散化就是在资产组合之间取 convex combination。
- **优化理论:** 最大化凹函数或最小化凸函数时, 局部最优就是全局最优。
- **一般均衡:** 固定点定理、分离超平面、支持价格都依赖凸性直觉。

**常见误区**

本章不要学成“会背定义”即可。你要建立一个判断习惯: 每当看到一个优化问题, 先问三件事: 可行集是否凸? 目标函数是否凹/凸? 一阶条件能否推出全局最优? 这比单纯算

导数更像研究生经济学的数学思维。

# 2 凸集：可行域的几何结构

## 2.1 定义

**定义 / 定理**

集合 $C \subseteq \mathbb{R}^n$ 称为凸集 convex set, 如果对任意 $x, y \in C$ 和任意 $\lambda \in [0,1]$, 都有

$$\lambda x + (1-\lambda)y \in C.$$

也就是说, 连接集合中任意两点的整条线段仍然留在集合内部。

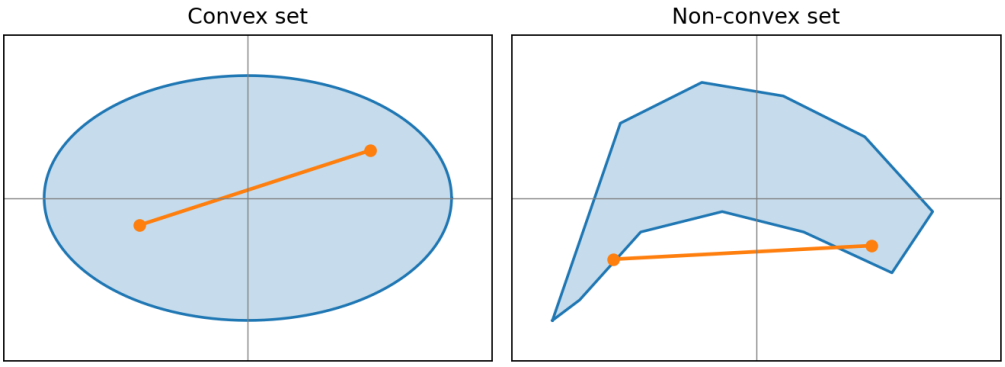



图 1: 凸集与非凸集。凸集中的任意两点连线仍在集合中; 非凸集可能出现连线穿出集合。

## 2.2 经济解释

如果 $x$ 和 $y$ 是两个可行消费束, 那么

$$\lambda x + (1-\lambda)y$$

是二者的混合消费束。可行集凸表示: 如果两个选择都可行, 那么它们的任意平均也可行。

典型凸集包括:

- 预算集 $\{x \geq 0 : p \cdot x \leq w\}$;
- 投资组合权重集 $\{w : \mathbf{1}'w = 1, w_i \geq 0\}$;
- 线性不等式定义的多面体;
- 球 $\{x : \|x\| \leq r\}$;
- 半空间 $\{x : a'x \leq b\}$。

非凸集常出现在固定成本、不可分商品、规模经济、门槛技术、最低投资规模等情形。非凸性会带来多重均衡、跳跃选择和局部最优问题。

**Worked Example**

**例 1：预算集是凸集**

考虑预算集

$$B = \{x \in \mathbb{R}^n_+ : p \cdot x \le w\},$$

其中 $p_i > 0$。若 $x, y \in B$，则

$$p \cdot x \le w, \qquad p \cdot y \le w.$$

对任意 $\lambda \in [0,1]$，

$$p \cdot [\lambda x + (1-\lambda)y] = \lambda p \cdot x + (1-\lambda) p \cdot y \le \lambda w + (1-\lambda) w = w.$$

同时 $x, y \ge 0$ 推出 $\lambda x + (1-\lambda)y \ge 0$。所以混合束仍在预算集内，$B$ 是凸集。

### 2.3 凸集的交仍是凸集

**定义 / 定理**

如果 $C_1, \dots, C_m$ 都是凸集，则

$$C = \bigcap_{i=1}^{m} C_i$$

也是凸集。

证明很简单。若 $x, y \in C$，则 $x, y$ 同时属于每个 $C_i$。由于每个 $C_i$ 凸，对任意 $\lambda \in [0,1]$，

$$\lambda x + (1-\lambda)y \in C_i, \quad i = 1, \dots, m.$$

所以它属于所有 $C_i$ 的交集。

**直觉盒子**

这解释了为什么许多线性约束叠加后仍给出凸可行集。预算约束、非负约束、产能约束、卖空约束、风险约束，只要各自是凸的，交出来的可行集仍然凸。

## 3 凸函数与凹函数

## 3.1 定义与图像

**定义 / 定理**

函数 $f : C \to \mathbb{R}$ 定义在凸集 $C$ 上。

- $f$ 是凸函数 convex function, 如果对任意 $x, y \in C$ 和 $\lambda \in [0, 1]$,

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y).$$

- $f$ 是凹函数 concave function, 如果

$$f(\lambda x + (1-\lambda)y) \ge \lambda f(x) + (1-\lambda) f(y).$$

凸函数的图像在弦线下方, 凹函数的图像在弦线上方。

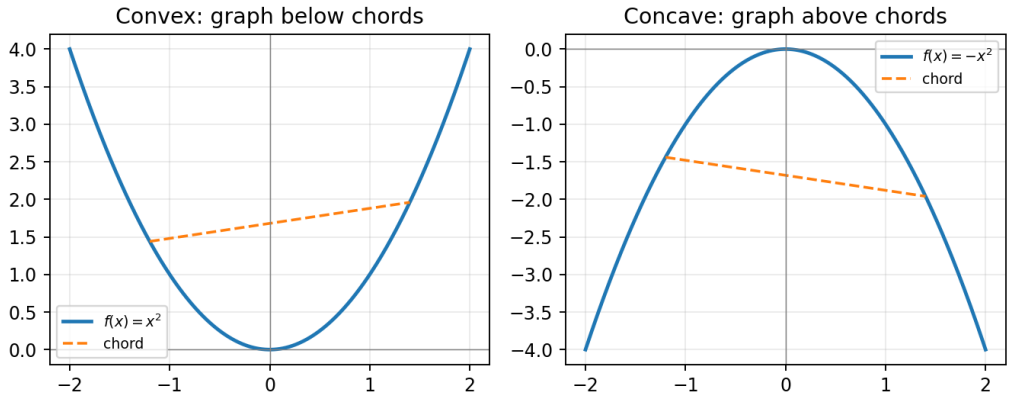



图 2: 凸函数与凹函数。左图 $x^2$ 是凸的; 右图 $-x^2$ 是凹的。

## 3.2 经济直觉

- 成本函数常常是凸的: 产量越高, 边际成本可能上升。
- 效用函数常常是凹的: 财富越多, 边际效用越低。
- 负利润函数或损失函数常常被设为凸的, 方便最小化。
- 均值方差组合中的方差 $w'\Sigma w$ 是凸函数, 因为协方差矩阵半正定。

## 3.3 严格凸与严格凹

若 $x \neq y$ 且 $\lambda \in (0, 1)$ 时不等号严格成立, 则称严格凸或严格凹。
严格凹最大化有一个重要好处: 如果最优解存在, 通常唯一。严格凸最小化同理。

**常见误区**

凹函数用于最大化, 凸函数用于最小化。不要混淆方向。最大化凸函数通常是困难的非凸优化; 最小化凹函数也通常不是好问题。

## 4 Hessian 判别: 与 Chapter 14 的连接

如果 $f$ 二阶连续可微, 凸性和 Hessian 直接相关。

**定义 / 定理**

设 $C$ 是凸开集, $f : C \to \mathbb{R}$ 二阶连续可微。

- $f$ 凸当且仅当对所有 $x \in C$, Hessian $\nabla^2 f(x)$ 半正定。
- $f$ 凹当且仅当对所有 $x \in C$, Hessian $\nabla^2 f(x)$ 半负定。
- 若 Hessian 正定, 则 $f$ 严格凸; 若 Hessian 负定, 则 $f$ 严格凹。

**Worked Example**

**例 2: Cobb-Douglas 效用的凹性**

设

$$u(x, y) = x^{\alpha} y^{1-\alpha}, \qquad x, y > 0, \quad 0 < \alpha < 1.$$

它是凹函数, 但直接用 Hessian 验证会比较繁琐。更常用的方法是取对数:

$$\ln u = \alpha \ln x + (1 - \alpha) \ln y.$$

$\ln x$ 与 $\ln y$ 是凹函数, 正权重和仍然凹。因此 $u$ 是 log-concave, 在很多最优化问题中具有良好性质。注意 log-concavity 比 quasi-concavity 强, 足以保证偏好凸性。

**Worked Example**

**例 3: 二次型的凸性**

设

$$f(x) = \frac{1}{2} x' A x + b' x + c,$$

其中 $A$ 对称。则

$$\nabla f(x) = Ax + b, \qquad \nabla^2 f(x) = A.$$

所以 $f$ 凸当且仅当 $A$ 半正定; 严格凸当 $A$ 正定。OLS 的平方损失和 portfolio variance 都是这个结构。

## 5 局部最优与全局最优

这是本章最核心的经济含义。

**定义 / 定理**

- $x^*$ 是局部最大点, 如果存在邻域 $N$ 使得对所有 $x \in N \cap C$, $f(x^*) \geq f(x)$。
- $x^*$ 是全局最大点, 如果对所有 $x \in C$, $f(x^*) \geq f(x)$。

最小点定义类似。

在非凸问题中, 局部最优和全局最优可能不同。

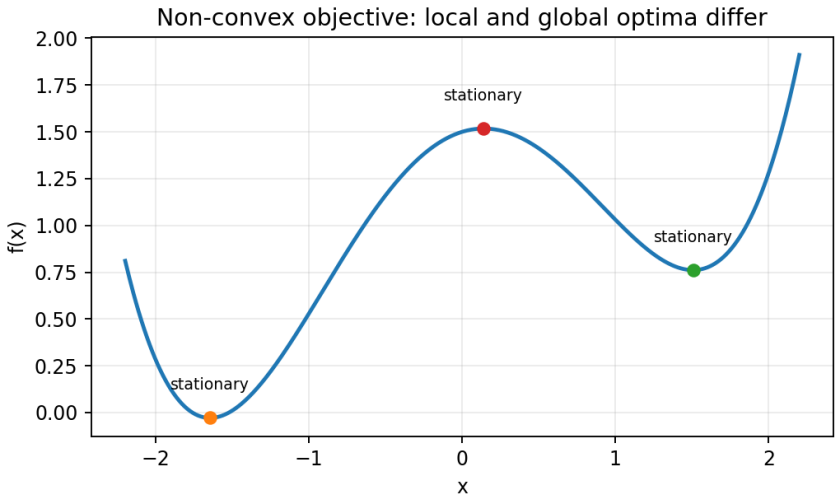



图 3: 非凸目标函数可能有多个 stationary points, 局部最优不一定是全局最优。

**定义 / 定理**

**凹最大化的一阶条件充分性.** 若 $C$ 是凸集, $f$ 是凹函数, 且 $x^* \in C$ 满足

$$\nabla f(x^*)'(x - x^*) \leq 0 \quad \text{for all } x \in C,$$

则 $x^*$ 是全局最大点。
若 $f$ 可微且 $x^*$ 是内点, 该条件简化为

$$\nabla f(x^*) = 0.$$

## 5.1 证明: 为什么凹性让一阶条件变充分

凹函数满足一阶上支撑性质:

$$f(x) \leq f(x^*) + \nabla f(x^*)'(x - x^*)$$

对所有 $x \in C$ 成立。若

$$\nabla f(x^*)'(x - x^*) \leq 0,$$

则

$$f(x) \le f(x^*)$$

对所有 $x \in C$ 成立。因此 $x^*$ 是全局最大点。

**直觉盒子**

凹函数像一座单峰山。只要你在某点发现所有可行方向都不能上升，那它就是全局山顶。凸函数最小化也是同样逻辑，只是山换成碗。

## 6 凸优化与 KKT 充分性

Chapter 16 学过 KKT 条件，但当时主要把它作为必要条件。本章给出更强版本。

**定义 / 定理**

**KKT 充分性，最大化版本.** 考虑

$$\max_x f(x)$$

subject to

$$g_i(x) \le 0, \qquad h_j(x) = 0.$$

若 $f$ 是凹函数，每个不等式约束函数 $g_i$ 是凸函数，等式约束 $h_j$ 是 affine 函数，并且某个可行点 $x^*$ 与乘子满足 KKT 条件，则 $x^*$ 是全局最优解。

为什么要求 $g_i$ 凸？因为可行集

$$\{x : g_i(x) \le 0\}$$

是 convex sublevel set。凸可行集 + 凹目标最大化，就是良好的 convex optimization 结构。

**常见误区**

“凸优化”这个词容易让人困惑。标准说法中，convex optimization 通常是“最小化凸函数”或等价地“最大化凹函数”。经济学里常做最大化，所以你经常看到的是 concave maximization over convex set。

## 7 Quasi-concavity 与偏好凸性

经济学不总是要求效用函数凹，很多时候只需要 quasi-concavity。

**定义 / 定理**

函数 $u : C \to \mathbb{R}$ 是 quasi-concave, 如果对任意 $a \in \mathbb{R}$, upper contour set

$$\{x \in C : u(x) \geq a\}$$

是凸集。

对消费者理论, upper contour set 是“至少和某个效用水平一样好”的消费束集合。它凸表示: 若两个消费束都至少一样好, 那么它们的混合也至少一样好。这就是偏好凸性。

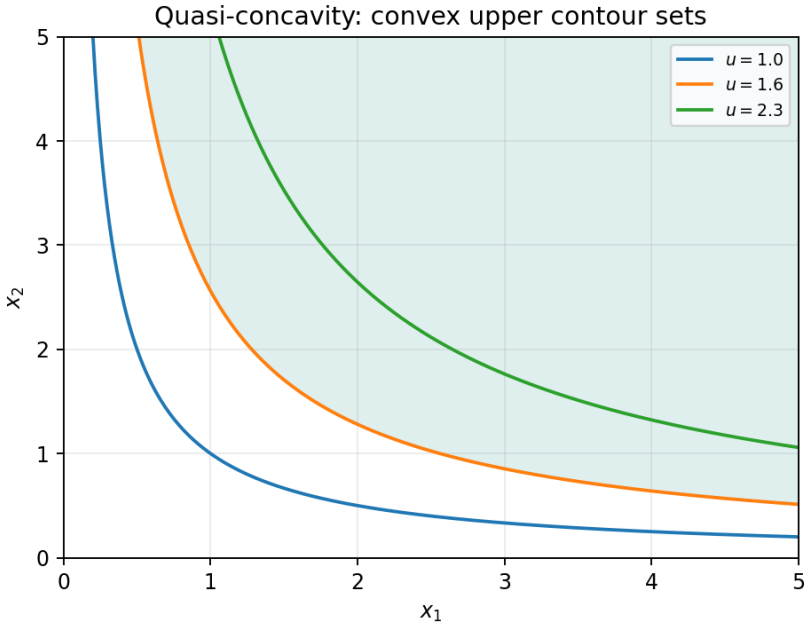



图 4: Quasi-concavity: upper contour sets 是凸集。Cobb-Douglas 型无差异曲线给出凸偏好。

## 7.1 凹性与 quasi-concavity 的关系

若 $u$ 是凹函数, 则 $u$ 一定 quasi-concave。证明如下。

若 $u(x) \geq a$ 且 $u(y) \geq a$, 对任意 $\lambda \in [0, 1]$, 凹性给出

$$u(\lambda x + (1 - \lambda)y) \geq \lambda u(x) + (1 - \lambda)u(y) \geq \lambda a + (1 - \lambda)a = a.$$

所以 upper contour set 凸。

但反过来不一定: quasi-concave 比 concave 弱。很多效用函数经过单调变换后仍表示同一偏好, 凹性可能改变, 但 quasi-concavity 作为偏好凸性的性质保持得更自然。

**Worked Example**

**例 4:** $u(x,y)=xy$ **是 quasi-concave**

在 $x,y>0$ 上, upper contour set 为

$$\{(x,y):xy\geq a\}.$$

当 $a>0$ 时边界是 $y=a/x$, 上方区域是凸的。更一般地, Cobb-Douglas 效用 $x^{\alpha}y^{1-\alpha}$ 是 quasi-concave, 因此表示凸偏好。

# 8 Jensen 不等式与风险厌恶

## 8.1 Jensen 不等式

**定义 / 定理**

若 $f$ 是凸函数, 则

$$f(\mathbb{E}[X])\leq \mathbb{E}[f(X)].$$

若 $u$ 是凹函数, 则

$$u(\mathbb{E}[X])\geq \mathbb{E}[u(X)].$$

这就是 Jensen inequality。

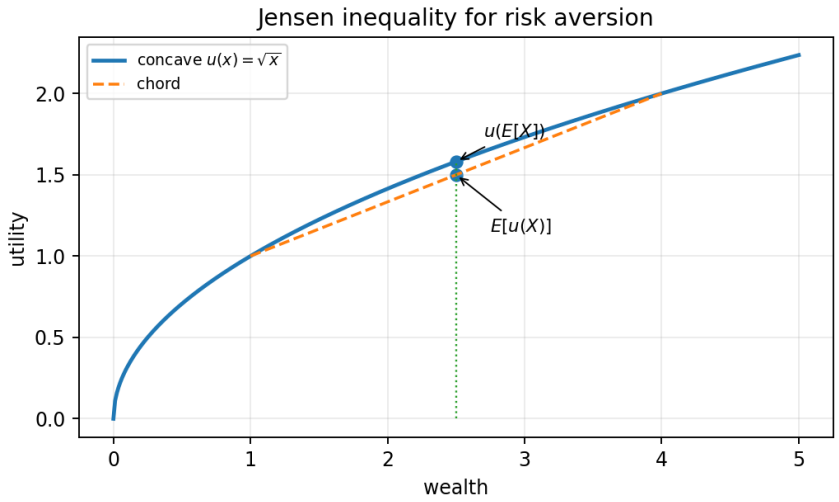



图 5: 凹效用下 $u(E[W])\geq E[u(W)]$。这就是风险厌恶的几何来源。

## 8.2 风险厌恶解释

设财富 $W$ 是随机变量, 效用函数 $u$ 凹。Jensen 给出

$$u(\mathbb{E}[W])\geq \mathbb{E}[u(W)].$$

左边是拿到确定财富 $\mathbb{E}[W]$ 的效用；右边是面对随机财富的期望效用。凹性表示边际效用递减，所以风险本身降低期望效用。

**直觉盒子**

风险厌恶不是“讨厌波动”的心理口号，而是凹效用的数学结果。由于坏状态损失带来的效用下降大于好状态收益带来的效用上升，同均值风险会降低期望效用。

## 9 分离超平面：价格支持的几何入口

凸分析中一个核心思想是：凸集和外部点可以被一个超平面分开。二维里就是一条直线。

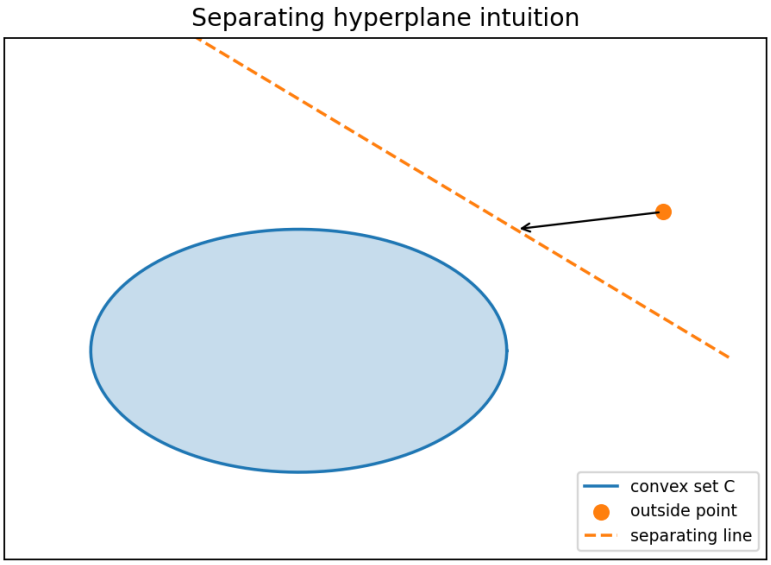



图 6：分离超平面直觉：凸集和外部点之间可以放置一条支持线。

**定义 / 定理**

**分离超平面直觉版。** 若 $C$ 是闭凸集，点 $y \notin C$，在适当条件下存在向量 $p \neq 0$ 和常数 $\alpha$，使得

$$p \cdot x \leq \alpha < p \cdot y$$

对所有 $x \in C$ 成立。

经济学里 $p$ 常常可以解释为价格向量。一般均衡中的支持价格、生产者利润最大化、消费者预算支持，都和这种几何思想有关。

**Advanced Core**

高级微观中常见一句话：“凸集可以被价格超平面支持。”这句话背后就是 separating hyperplane theorem。它解释了为什么在凸技术集合下，边界上的有效生产计划可以由某个价格向量支持为利润最大化选择。

# 10 金融应用：投资组合分散化与凸性

投资组合集合通常是凸的。若 $w^{(1)}$ 和 $w^{(2)}$ 是两个可行组合，则

$$\lambda w^{(1)} + (1-\lambda) w^{(2)}$$

也是可行组合。这正是基金组合、资产配置和风险分散的数学基础。

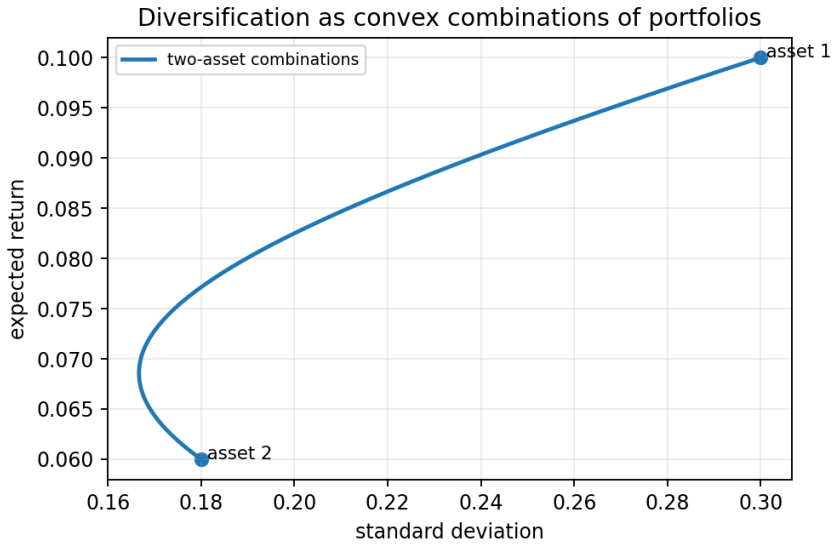



图 7: 两个资产组合之间的凸组合给出可行投资组合集合。相关系数小于 1 时，分散化可以降低风险。

## 10.1 组合方差是凸函数

设组合权重为 $w$，协方差矩阵为 $\Sigma$。组合方差是

$$\sigma_p^2 = w' \Sigma w.$$

由于协方差矩阵半正定，

$$z' \Sigma z \geq 0$$

对所有 $z$ 成立，所以 $w' \Sigma w$ 是凸函数。均值方差优化

$$\min_{w} w' \Sigma w \quad \text{s.t.} \quad \mu' w \geq \bar{r}, \quad \mathbf{1}' w = 1$$

就是凸优化问题。局部最优就是全局最优。

# 11 Worked Examples

**Worked Example**

**例 5：判断一个可行集是否凸**

判断

$$C = \{(x, y) : x \geq 0, y \geq 0, x + y \leq 1\}$$

是否凸。

若 $(x_1, y_1), (x_2, y_2) \in C$，对任意 $\lambda \in [0, 1]$，令

$$(x_\lambda, y_\lambda) = \lambda(x_1, y_1) + (1 - \lambda)(x_2, y_2).$$

由于 $x_1, x_2, y_1, y_2 \geq 0$，得 $x_\lambda, y_\lambda \geq 0$。又

$$x_\lambda + y_\lambda = \lambda(x_1 + y_1) + (1 - \lambda)(x_2 + y_2) \leq \lambda + (1 - \lambda) = 1.$$

所以 $(x_\lambda, y_\lambda) \in C$，$C$ 是凸集。

**Worked Example**

**例 6：用 Hessian 判断凹性**

设

$$f(x, y) = \ln x + \ln y, \qquad x, y > 0.$$

梯度为

$$\nabla f = \left( \frac{1}{x}, \frac{1}{y} \right).$$

Hessian 为

$$\nabla^2 f = \begin{pmatrix} -1/x^2 & 0 \\ 0 & -1/y^2 \end{pmatrix}.$$

对任意非零向量 $z = (z_1, z_2)'$，

$$z' \nabla^2 f z = -\frac{z_1^2}{x^2} - \frac{z_2^2}{y^2} < 0.$$

所以 Hessian 负定，$f$ 严格凹。最大化 $\ln x + \ln y$ over convex budget set 时，一阶条件给出唯一全局最优。

**Worked Example**

**例 7：凹最大化中的一阶条件足够**

考虑

$$\max_{x>0,y>0} \ln x + \ln y$$

subject to

$$p_x x + p_y y = w.$$

预算集是凸的, 目标函数严格凹, 所以任何满足 Lagrange FOC 的点都是唯一全局最优。
Lagrangian:

$$\mathcal{L} = \ln x + \ln y + \lambda(w - p_x x - p_y y).$$

FOC:

$$\frac{1}{x} = \lambda p_x, \qquad \frac{1}{y} = \lambda p_y, \qquad p_x x + p_y y = w.$$

由前两式得

$$\frac{y}{x} = \frac{p_x}{p_y}.$$

于是

$$p_x x = p_y y = \frac{w}{2}.$$

所以

$$x^* = \frac{w}{2p_x}, \qquad y^* = \frac{w}{2p_y}.$$

由于目标严格凹, 这个解不是局部候选点而已, 它就是唯一全局最优。

**Worked Example**

**例 8: Jensen 与风险溢价**

设 $u(w) = \sqrt{w}$, 财富随机变量为

$$W = \begin{cases} 100, & \text{概率} 1/2, \\ 0, & \text{概率} 1/2. \end{cases}$$

则

$$\mathbb{E}[W] = 50, \qquad u(\mathbb{E}[W]) = \sqrt{50} \approx 7.071.$$

期望效用为

$$\mathbb{E}[u(W)] = \frac{1}{2}\sqrt{100} + \frac{1}{2}\sqrt{0} = 5.$$

确定性等价 $CE$ 满足

$$\sqrt{CE} = 5,$$

所以

$$CE = 25.$$

风险溢价为

$$\mathbb{E}[W] - CE = 50 - 25 = 25.$$

凹效用使得同均值风险的确定性等价低于均值。

## 12 常见错误与辨析

**常见误区**

**误区 1：把 convex 和 concave 的优化方向搞反**。最大化要偏好凹目标；最小化要偏好凸目标。

**误区 2：以为 Hessian 在某一点负定就能推出全局凹**。全局凹性要求整个定义域上 Hessian 半负定。某一点的二阶条件只给局部信息。

**误区 3：以为 quasi-concavity 等于 concavity**。凹性更强，quasi-concavity 只要求 upper contour sets 凸。消费者理论中常常只需要后者。

**误区 4：忽视可行集凸性**。即使目标函数凹，如果可行集非凸，一阶条件仍可能不足以排除其他可行区域中的更好点。

**误区 5：把 Jensen 当作背诵公式**。 Jensen 的经济含义是：曲率决定平均之后再函数变换与函数变换之后再平均的大小关系。

## 13 本章小结：概念 - 公式 - 应用

| 概念 | 核心公式 / 条件 | 经济含义 |
|---|---|---|
| 凸集 | $\lambda x + (1-\lambda)y \in C$ | 混合可行选择仍可行。 |
| 凸函数 | $f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y)$ | 适合最小化；成本、风险常凸。 |
| 凹函数 | 反向不等式 | 适合最大化；效用、生产常凹。 |
| Hessian 判别 | $\nabla^2 f \succeq 0$ 或 $\preceq 0$ | 用曲率判断凸/凹。 |
| 凹最大化 | convex feasible set + concave objective | 局部最优等于全局最优。 |
| Quasi-concavity | upper contour sets convex | 偏好凸性，多样化偏好。 |
| Jensen | $u(EW) \ge E[u(W)]$ for concave $u$ | 风险厌恶与边际效用递减。 |
| 分离超平面 | $p \cdot x \le \alpha < p \cdot y$ | 支持价格、对偶理论入口。 |

| | | |
|---|---|---|
| 组合方差 | $w'\Sigma w$ convex | 投资组合优化是凸优化。 |

## 14 练习题

### A 组：基础计算

A1. 判断集合 $C = \{(x, y) : x + y \leq 1\}$ 是否凸。

A2. 判断集合 $C = \{(x, y) : xy \geq 1, x > 0, y > 0\}$ 是否凸。

A3. 判断 $f(x) = x^4$ 是否凸。

A4. 判断 $f(x) = \ln x$ 在 $x > 0$ 上是否凹。

A5. 对 $f(x, y) = x^2 + 2y^2 + xy$, 用 Hessian 判断凸性。

A6. 对 $f(x, y) = -(x^2 + y^2)$, 判断凹性和严格凹性。

### B 组：概念判断

B1. 为什么最大化凹函数比最大化一般函数更容易?

B2. 凹函数一定 quasi-concave 吗? 反过来呢?

B3. 为什么消费者理论更关心偏好凸性而不一定要求效用函数凹?

B4. Jensen 不等式如何解释风险厌恶?

B5. 为什么非凸生产集可能导致多重均衡或不连续选择?

### C 组：混合题

C1. 证明任意半空间 $\{x : a'x \leq b\}$ 是凸集。

C2. 证明凸函数的非负加权和仍然凸。

C3. 设 $f$ 凹且可微, 证明 $f(y) \leq f(x) + \nabla f(x)'(y - x)$。

C4. 设 $u(x, y) = \min\{x, y\}$, 判断其是否 concave, 是否 quasi-concave。

C5. 设 $\Sigma$ 半正定, 证明 $w'\Sigma w$ 是凸函数。

### D 组：经济金融建模题

D1. 解释为什么预算集 $\{x \geq 0 : p \cdot x \leq w\}$ 是凸集, 以及这对消费者问题有什么意义。

D2. 若效用函数严格凹且预算集凸, 为什么消费者最优解唯一?

D3. 用 Jensen 不等式解释为什么风险厌恶者愿意支付保险费。

D4. 均值方差问题 $\min_w w'\Sigma w$ s.t. $\mu'w \geq r_0, \mathbf{1}'w = 1$ 为什么是凸优化?

D5. 固定成本会怎样破坏生产集凸性? 给出经济解释。

### E 组：挑战 / 证明题

E1. 证明：若 $f$ 是凹函数，$C$ 是凸集，且 $x^*$ 满足 $\nabla f(x^*)'(x - x^*) \le 0$ 对所有 $x \in C$，则 $x^*$ 是全局最大点。

E2. 证明：若 $f$ 严格凹，则最大化问题最多有一个全局最优解。

E3. 设 $u$ 是严格递增的凹效用函数。证明对非退化风险 $W$，若 $E[W]$ 有限，通常有 $u(EW) > E[u(W)]$。

E4. 解释 separating hyperplane theorem 如何为“价格支持有效配置”提供几何直觉。

# 15 练习题参考答案

## A 组答案

A1. 是凸集。若 $x_1+y_1\le 1$, $x_2+y_2\le 1$, 则混合点满足

$$\lambda(x_1+y_1)+(1-\lambda)(x_2+y_2)\le 1.$$

A2. 是凸集。因为 $xy\ge 1$ 在 $x,y>0$ 上等价于 $\ln y\ge -\ln x$ 或 $y\ge 1/x$。函数 $1/x$ 在 $x>0$ 上凸, 其 epigraph 是凸集。

A3. $f''(x)=12x^2\ge 0$, 所以 $x^4$ 凸。但它不是处处严格凸的 Hessian 正定形式, 因为 $f''(0)=0$; 仍可证明它严格凸。

A4. $f''(x)=-1/x^2<0$, 所以 $\ln x$ 在 $x>0$ 上严格凹。

A5. Hessian 为

$$H=\begin{pmatrix}2 & 1\\ 1 & 4\end{pmatrix}.$$

主子式 $2>0$, determinant $8-1=7>0$, 所以正定, $f$ 严格凸。

A6. Hessian 为

$$H=\begin{pmatrix}-2 & 0\\ 0 & -2\end{pmatrix},$$

负定, 所以 $f$ 严格凹。

## B 组答案

B1. 因为凹函数的局部最大点就是全局最大点。一阶条件不只是必要条件, 在凸可行集上还可以成为充分条件。

B2. 凹函数一定 quasi-concave, 因为 upper contour sets 凸。反过来不一定, quasi-concavity 更弱。

B3. 效用函数只代表偏好, 可经过严格递增变换而不改变偏好。凹性可能因变换改变, 但 upper contour sets 的凸性更直接刻画“喜欢混合”的偏好结构。

B4. 凹效用下 $u(EW)\ge E[u(W)]$, 说明确定拿到期望财富比面对同均值风险更好, 因此风险厌恶者愿意支付保险费以减少风险。

B5. 固定成本和规模经济会让平均生产计划不可行或低效, 使两个可行生产计划的混合不一定可行或不一定保持同样技术性质, 从而造成非凸性和多重局部选择。

## C 组答案

C1. 若 $x, y$ 都满足 $a'x \le b$, $a'y \le b$, 则

$$a'[\lambda x + (1-\lambda)y] = \lambda a'x + (1-\lambda)a'y \le b.$$

所以半空间凸。

C2. 若 $f_i$ 凸且 $\alpha_i \ge 0$, 则

$$\sum_i \alpha_i f_i(\lambda x + (1-\lambda)y) \le \sum_i \alpha_i[\lambda f_i(x) + (1-\lambda) f_i(y)].$$

整理得非负加权和凸。

C3. 对凹函数, 图像位于任一点切平面下方。可由一维函数 $g(t) = f(x + t(y-x))$ 的凹性证明:

$$g(1) \le g(0) + g'(0),$$

即

$$f(y) \le f(x) + \nabla f(x)'(y-x).$$

C4. $u(x,y) = \min\{x, y\}$ 是凹函数, 因为两个 affine 函数的 pointwise minimum 是凹函数。既然凹, 它也 quasi-concave。其 upper contour set 是 $\{x \ge a, y \ge a\}$, 是凸集。

C5. Hessian 为 $2\Sigma$。若 $\Sigma$ 半正定, 则 $2\Sigma$ 半正定, 因此 $w'\Sigma w$ 是凸函数。

## D 组答案

D1. 预算集是线性不等式 $p \cdot x \le w$ 与非负半空间的交集, 所以凸。它保证两个可行消费束的混合仍可行, 让消费者问题具有良好几何结构。

D2. 严格凹函数在凸集上的最大化最多有一个最优解。若存在两个不同最优点, 它们的中点可行, 严格凹性会使中点效用严格高于两端, 矛盾。

D3. 保险把随机财富变得更确定。对凹效用, 减少同均值风险会提高期望效用。只要保险费不超过风险溢价, 风险厌恶者愿意购买。

D4. 目标函数 $w'\Sigma w$ 因 $\Sigma \succeq 0$ 而凸; 约束 $\mu' w \ge r_0$ 和 $\mathbf{1}'w = 1$ 是 affine/半空间约束, 可行集凸。因此这是凸优化问题。

D5. 固定成本意味着生产少量和不生产之间可能有跳跃。例如某技术只有产量超过阈值才可行。两个大规模生产计划和零生产计划的平均不一定对应可行或高效技术, 生产集可能非凸。

## E 组答案

E1. 凹性给出

$$f(x) \le f(x^*) + \nabla f(x^*)'(x - x^*).$$

由条件 $\nabla f(x^*)'(x-x^*)\le 0$, 得 $f(x)\le f(x^*)$ 对所有 $x\in C$ 成立。因此 $x^*$ 是全局最大点。

E2. 若存在两个不同全局最优点 $x_1,x_2$, 且最大值为 $M$, 则对 $\lambda\in(0,1)$, 严格凹性给出

$$f(\lambda x_1+(1-\lambda)x_2)>\lambda M+(1-\lambda)M=M.$$

这与 $M$ 是全局最大值矛盾。

E3. Jensen 给出 $u(EW)\ge E[u(W)]$。若 $u$ 严格凹且 $W$ 非退化, 即不是几乎处处常数, 则严格 Jensen 给出

$$u(EW)>E[u(W)].$$

这说明确定性均值严格优于同均值风险。

E4. 若某个有效配置位于可行集边界, 凸性使得存在支持超平面把该点和不可行改进方向分开。这个超平面的法向量可解释为价格向量。价格向量支持该配置, 意味着在这些价格下没有可行选择能带来更高价值或利润。

## 16 后续学习连接

下一章 Chapter 18 会进入隐函数定理、反函数定理与比较静态。凸性告诉我们什么时候最优解是全局且稳定的; 隐函数定理进一步告诉我们, 当参数变化时, 这个最优解或均衡点如何平滑变化。Chapter 19 的包络定理与对偶理论会大量使用本章思想: 凸可行集、凹目标、支持超平面和影子价格。

**直觉盒子**

到这里, 优化主线已经形成闭环: Chapter 14 学无约束和 Hessian, Chapter 15 学等式约束, Chapter 16 学不等式约束和 KKT, Chapter 17 学凸性保证全局最优。后面几章会把这些工具用于比较静态、对偶理论和均衡存在性。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 18
# 隐函数定理、包络定理与比较静态

前几章已经解决“如何求最优”的问题: Hessian, Lagrange, KKT, convexity。经济学真正关心的下一步通常是: **如果价格、收入、税率、利率、技术、风险厌恶这些参数改变, 最优选择和价值函数会怎样改变?** 本章围绕三个工具展开: 隐函数定理把一阶条件转化成最优选择函数; 比较静态给出参数变化的方向和大小; 包络定理说明价值函数的导数不需要追踪最优选择如何变化。它们是高级微观、宏观、金融、计量和动态优化中最常用的数学语言。

math4econ personalized notes
2026

# 目录

# Chapter 18 总览

| | |
|---|---|
| **本章主题** | implicit function theorem, comparative statics, envelope theorem, value function derivatives。 |
| **核心问题** | 已知均衡或最优条件 $F(x,\theta)=0$, 如何求 $x^*(\theta)$ 对参数 $\theta$ 的反应? 已知价值函数 $V(\theta)=\max_x f(x,\theta)$, 如何求 $V'(\theta)$ 而不显式求出 $x^*(\theta)$? |
| **数学工具** | scalar IFT, multivariate IFT, total differentiation, matrix inverse formula, FOC differentiation, envelope theorem, constrained envelope theorem, comparative statics sign analysis。 |
| **经济应用** | 市场均衡比较静态、垄断产量、税收影响、消费者需求、成本最小化、利润函数、价值函数、投资组合权重、债务稳态。 |
| **后续连接** | Chapter 19 对偶理论与支出/成本/利润函数, Chapter 20 固定点与均衡存在性, Chapter 21 动态规划, 高级微观的 Slutsky equation 和 Shephard/Roy/Hotelling 引理。 |

**直觉盒子**

本章的核心直觉是: **经济模型里很多对象不是显式写出来的函数, 而是由方程间接定义的函数**。例如均衡价格由供需相等定义, 最优消费由一阶条件和预算约束定义, 稳态资本由净投资为零定义。隐函数定理告诉你这些对象什么时候可以看成参数的函数, 以及怎么求导。

## 1 为什么比较静态是经济学的核心语言

经济学通常不满足于问“均衡是多少”。更常见的问题是:

- 税率上升会使劳动供给吗上升还是下降?
- 利率上升会使资产价格下降多少?
- 工资上升会使企业雇佣量下降多少?
- 风险厌恶增强会使最优风险资产权重下降多少?
- 储蓄率上升会怎样改变 Solow 稳态资本?
- 价格上涨会怎样改变消费者需求和间接效用?

这些问题都属于 comparative statics, 即比较不同参数值下的均衡或最优状态。

### 1.1 从“求解”到“求导”

很多模型不能轻易写出闭式解。比如一个均衡由

$$F(x,\theta)=0$$

定义，你可能无法直接解出

$$x=x^*(\theta).$$

但只要能求偏导，就能得到

$$\frac{\mathrm{d}x^*}{\mathrm{d}\theta}=-\frac{F_\theta}{F_x}.$$

这就是本章最重要的思想：**即使解不出来，也可以对解求导**。

**常见误区**

比较静态不是把两个数字代进去比大小。真正的比较静态是把模型内部的均衡条件或最优条件作为方程系统，对参数求导，解释导数的符号和大小。

## 2 标量隐函数定理：一条方程定义一个函数

### 2.1 从一个简单例子开始

考虑方程

$$F(x,\theta)=x^2+\theta^2-4=0.$$

它定义了一个圆。严格来说，这个方程不在全局上定义唯一的 $x(\theta)$，因为对同一个 $\theta$ 可能有两个 $x$。但在上半支附近，可以写成

$$x(\theta)=\sqrt{4-\theta^2}.$$

这个例子说明：隐函数通常是**局部**存在，而不是全局存在。

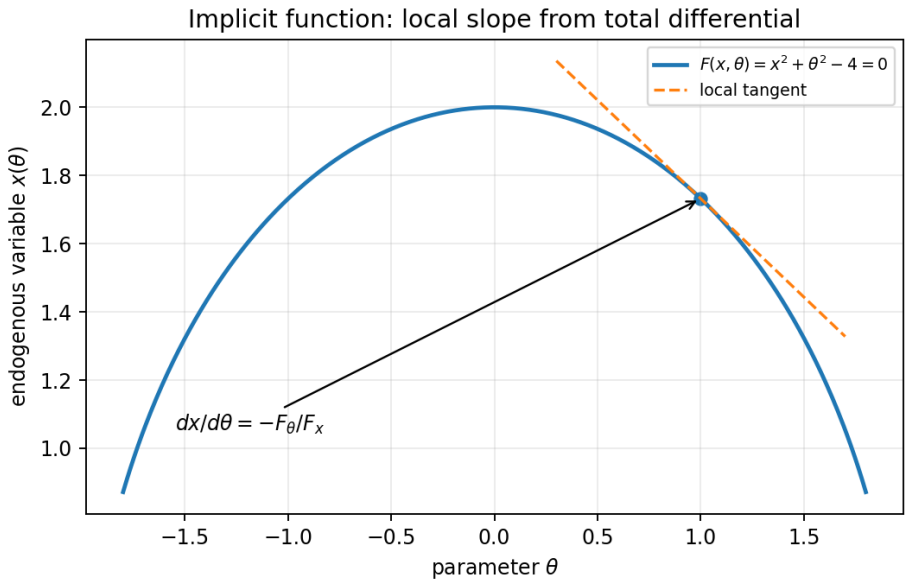



图 1: 隐函数定理: 方程 $F(x,\theta)=0$ 在局部定义一条函数曲线, 斜率来自全微分。

## 2.2 标量隐函数定理

**定义 / 定理**

**标量隐函数定理, 实用版本.** 设 $F(x,\theta)$ 连续可微, 且在某点 $(x_0,\theta_0)$ 满足

$$F(x_0,\theta_0)=0.$$

如果

$$F_x(x_0,\theta_0)\neq 0,$$

那么在 $(x_0,\theta_0)$ 附近, 可以把 $x$ 看成 $\theta$ 的函数 $x=x(\theta)$, 且满足

$$F(x(\theta),\theta)=0.$$

并且

$$\boxed{\frac{\mathrm{d}x}{\mathrm{d}\theta}=-\frac{F_\theta}{F_x}.}$$

## 2.3 公式推导: 全微分法

因为

$$F(x(\theta),\theta)=0,$$

对 $\theta$ 求导:

$$\frac{\mathrm{d}}{\mathrm{d}\theta}F(x(\theta),\theta)=0.$$

链式法则给出

$$F_x \frac{\mathrm{d}x}{\mathrm{d}\theta} + F_\theta = 0.$$

如果 $F_x \neq 0$, 则

$$\boxed{\frac{\mathrm{d}x}{\mathrm{d}\theta} = -\frac{F_\theta}{F_x}.}$$

**直觉盒子**

$F_x \neq 0$ 的直觉是: 方程对 $x$ 的变化有反应。如果 $x$ 稍微变动不会改变方程左边, 那么你很难用 $x$ 来抵消参数 $\theta$ 的变化。数学上, $F_x \neq 0$ 保证方程能在局部解出 $x$。

**Worked Example**

**例 1: 市场均衡价格对需求冲击的反应**

设需求和供给为

$$D(p, a) = a - bp, \qquad S(p) = c + dp,$$

其中 $a$ 是需求水平参数。均衡由

$$F(p, a) = D(p, a) - S(p) = a - bp - c - dp = 0$$

定义。求 $\mathrm{d}p^*/\mathrm{d}a$。

由隐函数公式:

$$F_a = 1, \qquad F_p = -(b + d).$$

所以

$$\frac{\mathrm{d}p^*}{\mathrm{d}a} = -\frac{F_a}{F_p} =$$

rac1b+d>0.

经济解释: 需求水平上升使均衡价格上升, 但价格反应越小当需求和供给对价格越敏感, 即 $b + d$ 越大。

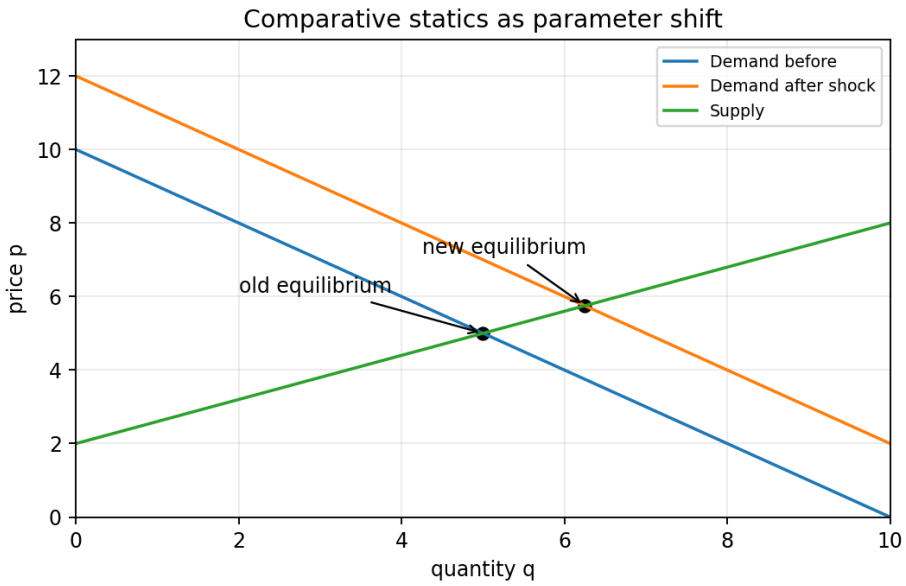



图 2: 比较静态: 参数变化移动需求曲线, 均衡价格和数量随之变化。

# 3 从均衡条件到比较静态

## 3.1 通用模板

很多一维经济模型可以写成

$$F(x,\theta)=0.$$

这时比较静态的模板是:

$$\frac{\mathrm{d}x^*}{\mathrm{d}\theta}=-\frac{F_\theta}{F_x}.$$

要判断符号, 只需要判断 $F_\theta$ 和 $F_x$ 的符号。

**方法模板**

**一维比较静态四步法**

1. 写均衡或一阶条件: $F(x,\theta)=0$。
2. 检查 $F_x\neq 0$。
3. 用隐函数公式: $\mathrm{d}x^*/\mathrm{d}\theta=-F_\theta/F_x$。
4. 用经济含义解释 $F_x$ 与 $F_\theta$ 的符号。

**Worked Example**

**例 2: 垄断产量对成本参数的反应**

逆需求为

$$P(Q)=a-bQ,$$

成本为

$$C(Q) = cQ.$$

利润为

$$\pi(Q; c) = (a - bQ)Q - cQ.$$

一阶条件为

$$F(Q, c) = a - 2bQ - c = 0.$$

求 $\mathrm{d}Q^*/\mathrm{d}c$。
由隐函数公式:

$$F_c = -1, \qquad F_Q = -2b.$$

所以

$$\frac{\mathrm{d}Q^*}{\mathrm{d}c} = -\frac{-1}{-2b} = -\frac{1}{2b} < 0.$$

经济解释: 边际成本上升降低垄断产量。需求越平坦, 即 $b$ 越小, 产量反应越大。

**Worked Example**

**例 3: Solow 稳态对储蓄率的反应**

Solow 稳态满足

$$F(k, s) = sf(k) - (n + \delta)k = 0.$$

隐函数公式给出

$$\frac{\mathrm{d}k^*}{\mathrm{d}s} = -\frac{F_s}{F_k}.$$

其中

$$F_s = f(k) > 0,$$

$$F_k = sf'(k) - (n + \delta).$$

在稳定稳态处, 通常有 $F_k < 0$。因此

$$\frac{\mathrm{d}k^*}{\mathrm{d}s} > 0.$$

经济解释: 储蓄率上升提高净投资, 稳态资本上升。

## 4 多变量隐函数定理

经济模型常常不是一个方程一个未知数, 而是一组方程决定一组内生变量。例如消费者最优包含多个一阶条件和预算约束, 一般均衡包含多个市场出清条件。

## 4.1 矩阵形式

设

$$\boldsymbol{F}(\boldsymbol{x},\boldsymbol{\theta})=\mathbf{0},$$

其中

$$\boldsymbol{x}\in\mathbb{R}^n,\qquad \boldsymbol{\theta}\in\mathbb{R}^m,\qquad \boldsymbol{F}:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^n.$$

如果在点 $(\boldsymbol{x}_0,\boldsymbol{\theta}_0)$ 满足

$$\boldsymbol{F}(\boldsymbol{x}_0,\boldsymbol{\theta}_0)=\mathbf{0},$$

并且 Jacobian

$$F_x=D_{\boldsymbol{x}}\boldsymbol{F}(\boldsymbol{x}_0,\boldsymbol{\theta}_0)$$

可逆, 则局部可以把 $\boldsymbol{x}$ 写成 $\boldsymbol{\theta}$ 的函数:

$$\boldsymbol{x}=\boldsymbol{x}(\boldsymbol{\theta}).$$

对

$$\boldsymbol{F}(\boldsymbol{x}(\boldsymbol{\theta}),\boldsymbol{\theta})=0$$

求导得到

$$F_x\frac{\partial\boldsymbol{x}}{\partial\boldsymbol{\theta}}+F_\theta=0.$$

因此

$$\boxed{\frac{\partial\boldsymbol{x}}{\partial\boldsymbol{\theta}}=-F_x^{-1}F_\theta.}$$

**定义 / 定理**

**多变量隐函数定理, 实用版本.** 如果方程数量等于内生变量数量, 且对内生变量的 Jacobian 矩阵可逆, 那么局部解函数存在, 并且其导数由

$$D\boldsymbol{x}(\boldsymbol{\theta})=-F_x^{-1}F_\theta$$

给出。

**常见误区**

多变量 IFT 中最常见的错误是维度混乱。若 $\boldsymbol{F}\in\mathbb{R}^n$, $\boldsymbol{x}\in\mathbb{R}^n$, $\boldsymbol{\theta}\in\mathbb{R}^m$, 则 $F_x$ 是 $n\times n$, $F_\theta$ 是 $n\times m$, 所以 $-F_x^{-1}F_\theta$ 是 $n\times m$, 正好是 $\partial\boldsymbol{x}/\partial\boldsymbol{\theta}$。

### 4.2 二元市场均衡例子

**Worked Example**

**例 4：两个市场同时出清**

设两个内生价格 $p_1, p_2$ 由两条市场出清条件决定：

$$F_1(p_1, p_2, a) = a - p_1 - 0.5p_2 = 0,$$

$$F_2(p_1, p_2, a) = 2 + p_1 - 2p_2 = 0.$$

求需求冲击 $a$ 对价格向量的影响。

写成矩阵：

$$\boldsymbol{F}(\boldsymbol{p}, a) = \begin{pmatrix} a - p_1 - 0.5p_2 \\ 2 + p_1 - 2p_2 \end{pmatrix} = 0.$$

对价格的 Jacobian 为

$$F_p = \begin{pmatrix} -1 & -0.5 \\ 1 & -2 \end{pmatrix}.$$

对 $a$ 的导数为

$$F_a = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

因此

$$\frac{\partial \boldsymbol{p}}{\partial a} = -F_p^{-1} F_a.$$

计算逆矩阵。行列式为

$$(-1)(-2) - (-0.5)(1) = 2.5.$$

所以

$$F_p^{-1} = \frac{1}{2.5} \begin{pmatrix} -2 & 0.5 \\ -1 & -1 \end{pmatrix}.$$

于是

$$\frac{\partial \boldsymbol{p}}{\partial a} = -\frac{1}{2.5} \begin{pmatrix} -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.4 \end{pmatrix}.$$

经济解释：第一个市场的需求冲击不仅提高 $p_1$，也通过联立市场结构提高 $p_2$。

## 5 一阶条件定义的最优选择函数

## 5.1 无约束最优化比较静态

考虑

$$x^*(\theta) = \arg\max_x f(x, \theta).$$

一阶条件为

$$f_x(x^*(\theta), \theta) = 0.$$

令

$$F(x, \theta) = f_x(x, \theta).$$

由隐函数公式:

$$\frac{\mathrm{d}x^*}{\mathrm{d}\theta} = -\frac{F_\theta}{F_x} = -\frac{f_{x\theta}}{f_{xx}}.$$

若是最大化且二阶条件满足 $f_{xx} < 0$, 则 $\mathrm{d}x^*/\mathrm{d}\theta$ 的符号与 $f_{x\theta}$ 相同:

$$\boxed{\frac{\mathrm{d}x^*}{\mathrm{d}\theta} = -\frac{f_{x\theta}}{f_{xx}}.}$$

因为 $f_{xx} < 0$, 若 $f_{x\theta} > 0$, 则 $x^*$ 随 $\theta$ 上升。

**直觉盒子**

$f_{x\theta}$ 衡量参数 $\theta$ 是否提高选择变量 $x$ 的边际收益。如果提高 $\theta$ 使边际收益曲线上移, 那么最优 $x$ 往往上升。分母 $f_{xx}$ 是曲率, 它越接近 0, 目标越平, 最优选择对参数越敏感。

**Worked Example**

**例 5: 最优努力对奖金强度的反应**

设代理人选择努力 $e$, 收益为

$$f(e, b) = be - \frac{1}{2}ce^2,$$

其中 $b$ 是奖金强度, $c > 0$。一阶条件:

$$f_e = b - ce = 0.$$

最优努力为 $e^* = b/c$。用公式也可得

$$\frac{\mathrm{d}e^*}{\mathrm{d}b} = -\frac{f_{eb}}{f_{ee}} = -\frac{1}{-c} = \frac{1}{c} > 0.$$

经济解释: 奖金强度越大, 努力越高; 努力成本曲率 $c$ 越大, 反应越弱。

## 5.2 多变量无约束优化

如果

$$\boldsymbol{x}^*(\theta) = \arg\max_{\boldsymbol{x}} f(\boldsymbol{x}, \theta),$$

一阶条件为

$$\nabla_x f(\boldsymbol{x}^*(\theta), \theta) = 0.$$

对 $\theta$ 求导:

$$f_{xx}\frac{\partial \boldsymbol{x}^*}{\partial \theta} + f_{x\theta} = 0.$$

因此

$$\boxed{\frac{\partial \boldsymbol{x}^*}{\partial \theta} = -(f_{xx})^{-1} f_{x\theta}.}$$

这里 $f_{xx}$ 是 Hessian, $f_{x\theta}$ 是边际条件对参数的导数向量。

## 6 约束优化的一阶条件与比较静态

经济学中更常见的是约束最优化:

$$\max_x f(x, \theta) \quad \text{s.t.} \quad g(x, \theta) = 0.$$

令 Lagrangian 为

$$\mathcal{L}(x, \lambda, \theta) = f(x, \theta) + \lambda g(x, \theta).$$

一阶条件是

$$\mathcal{L}_x(x, \lambda, \theta) = 0,$$

$$g(x, \theta) = 0.$$

把未知数写成

$$z = (x, \lambda).$$

方程系统为

$$\boldsymbol{F}(z, \theta) = \begin{pmatrix} \mathcal{L}_x \\ g \end{pmatrix} = 0.$$

由 IFT:

$$\frac{\mathrm{d}z^*}{\mathrm{d}\theta} = -F_z^{-1} F_\theta.$$

**Advanced Core**

**约束优化的比较静态矩阵.** 若 $x \in \mathbb{R}^n$, 有一个等式约束 $g(x, \theta) = 0$, 则对 $(x, \lambda)$ 的 Jacobian 为

$$F_z = \begin{pmatrix} \mathcal{L}_{xx} & g_x' \\ g_x & 0 \end{pmatrix}.$$

这个矩阵和 bordered Hessian 关系很近。它描述了最优性方程系统对选择变量和乘子的局部反应。

## 6.1 消费者问题的比较静态

考虑二商品消费者问题:

$$\max_{x_1,x_2} u(x_1,x_2) \quad \text{s.t.} \quad p_1x_1+p_2x_2=m.$$

Lagrangian:

$$\mathcal{L}=u(x_1,x_2)+\lambda(m-p_1x_1-p_2x_2).$$

FOC:

$$u_1-\lambda p_1=0, \qquad u_2-\lambda p_2=0,$$

$$m-p_1x_1-p_2x_2=0.$$

这些方程在合适条件下由 IFT 定义 Marshallian demand:

$$x_i=x_i(p_1,p_2,m).$$

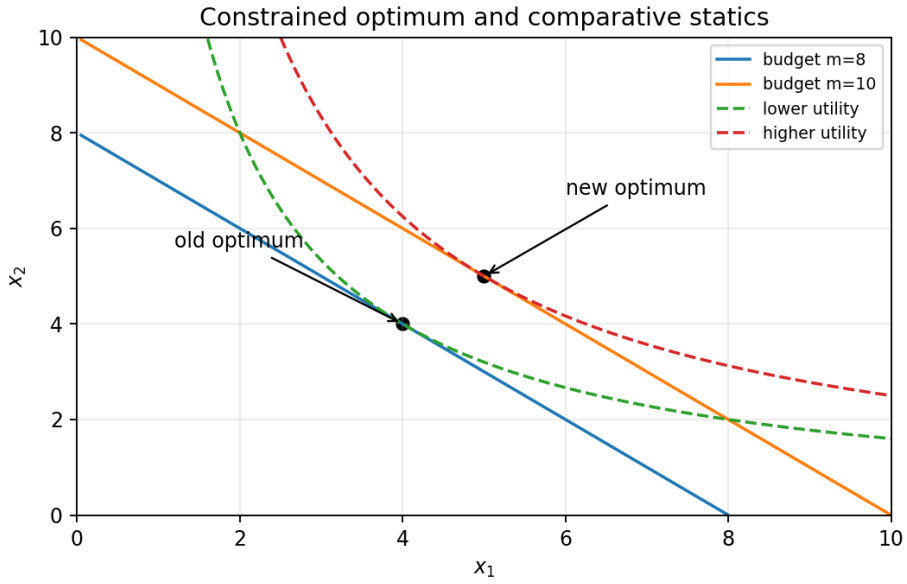



图 3: 约束最优的比较静态: 收入上升移动预算线, 最优点随之移动。

**常见误区**

需求比较静态的完整分析会通向 Slutsky equation: 价格效应可以分解为替代效应和收入效应。本章先建立 IFT 语言, 后续对偶理论中再系统处理。

# 7 包络定理: 价值函数导数为什么不用管 $x^*$ 的变化

## 7.1 无约束包络定理

设

$$V(\theta) = \max_{x} f(x, \theta),$$

最优选择为 $x^*(\theta)$。则

$$V(\theta) = f(x^*(\theta), \theta).$$

对 $\theta$ 求导:

$$V'(\theta) = f_x(x^*, \theta)\frac{\mathrm{d}x^*}{\mathrm{d}\theta} + f_\theta(x^*, \theta).$$

由于最优点满足一阶条件

$$f_x(x^*, \theta) = 0,$$

所以第一项消失:

$$\boxed{V'(\theta) = f_\theta(x^*(\theta), \theta).}$$

**定义 / 定理**

**包络定理，无约束版本.** 当 $x^*(\theta)$ 是内部最优且满足一阶条件时，价值函数对参数的导数等于目标函数对参数的直接偏导，在最优点处评价:

$$\frac{\mathrm{d}V}{\mathrm{d}\theta} = f_\theta(x^*(\theta), \theta).$$

不需要计算 $\mathrm{d}x^*/\mathrm{d}\theta$。

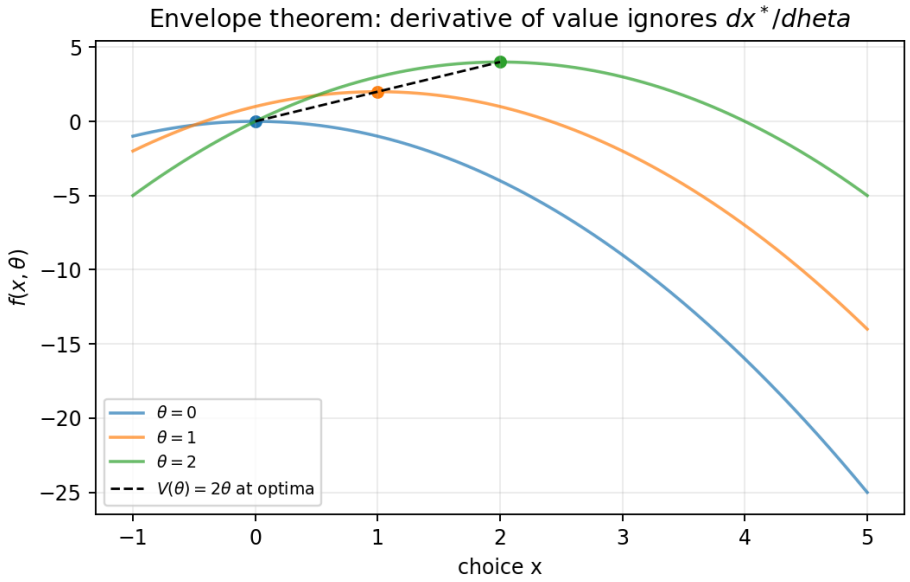



图 4: 包络定理: 价值函数沿着最优点包络一族目标函数; 求导时只保留直接效应。

## 7.2 为什么 $dx^*/d\theta$ 不出现

包络定理最容易让初学者困惑。因为 $V$ 明明等于

$$f(x^*(\theta),\theta),$$

为什么 $x^*$ 的变化不用管？答案是：在内部最优点，$x$ 方向的一阶变化对目标没有一阶影响。也就是说，你已经站在峰顶，沿 $x$ 小幅移动，一阶效应为零。参数变化对价值的一阶影响只来自目标函数本身的直接变化。

**Worked Example**

**例 6：包络定理验证**

设

$$f(x,\theta)=\theta x-\frac{1}{2}x^2.$$

最优条件：

$$f_x=\theta-x=0,$$

所以

$$x^*(\theta)=\theta.$$

价值函数：

$$V(\theta)=\theta\cdot\theta-\frac{1}{2}\theta^2=\frac{1}{2}\theta^2.$$

因此

$$V'(\theta)=\theta.$$

包络定理给出

$$f_\theta(x,\theta)=x.$$

在最优点处评价：

$$f_\theta(x^*,\theta)=x^*(\theta)=\theta.$$

两者一致。

## 7.3 约束包络定理

考虑

$$V(\theta)=\max_x f(x,\theta)\quad\text{s.t.}\quad g(x,\theta)=0.$$

使用

$$\mathcal{L}(x,\lambda,\theta)=f(x,\theta)+\lambda g(x,\theta).$$

若最优满足 FOC，则

$$\boxed{V'(\theta)=\mathcal{L}_\theta(x^*,\lambda^*,\theta)=f_\theta(x^*,\theta)+\lambda^* g_\theta(x^*,\theta).}$$

**直觉盒子**

约束包络定理说明: 参数改变价值的直接影响有两部分。第一部分是目标函数本身直接改变; 第二部分是约束集合改变, 其价值由 Lagrange multiplier 衡量。因此乘子自然解释为 shadow price。

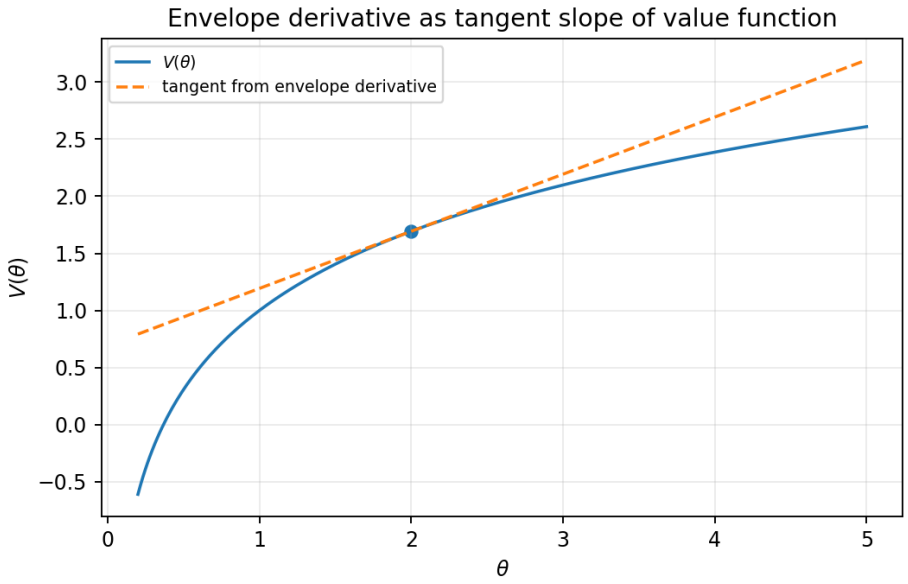



图 5: 价值函数的斜率可以由包络定理直接得到, 不需要显式求最优选择导数。

# 8 乘子的影子价格与包络定理

## 8.1 资源约束的影子价值

考虑

$$V(m) = \max_x u(x) \quad \text{s.t.} \quad m - px = 0.$$

Lagrangian:

$$\mathcal{L} = u(x) + \lambda(m - px).$$

约束包络定理:

$$V'(m) = \mathcal{L}_m = \lambda.$$

所以 Lagrange multiplier $\lambda$ 就是收入或资源增加一单位带来的最大效用增加, 即 shadow value of wealth。

## 8.2 成本最小化中的条件要素需求

成本函数定义为

$$C(w, q) = \min_x w \cdot x \quad \text{s.t.} \quad f(x) \geq q.$$

在标准条件下, Shephard's lemma 表示

$$\boxed{\frac{\partial C(w,q)}{\partial w_i} = x_i^*(w,q).}$$

它本质上是包络定理: 工资 $w_i$ 上升一点, 成本的直接增加等于最优使用的第 $i$ 个要素数量。

### 8.3 利润函数中的供给和要素需求

利润函数

$$\pi(p,w) = \max_{y,x} py - wx \quad \text{s.t.} \quad y \le f(x).$$

包络定理给出 Hotelling's lemma 直觉:

$$\frac{\partial \pi}{\partial p} = y^*(p,w),$$

$$\frac{\partial \pi}{\partial w} = -x^*(p,w).$$

价格提高一单位对最大利润的影响等于最优产量; 要素价格提高一单位对利润的影响等于负的最优要素需求。

**Advanced Core**

Shephard, Roy, Hotelling 这些经典引理都不是孤立公式, 它们都是包络定理在消费者/生产者对偶问题中的具体版本。下一章若进入对偶理论, 会系统整理这些对象之间的关系。

## 9 应用一: 税收对劳动供给的比较静态

考虑消费者选择劳动 $l$ 和消费 $c$。令总时间为 $T$, 闲暇为 $T-l$, 税后工资为 $(1-\tau)w$。预算约束:

$$c = (1-\tau)wl + R.$$

效用:

$$u(c, T-l).$$

劳动选择问题:

$$\max_{l} u((1-\tau)wl + R, T-l).$$

一阶条件可写为

$$F(l,\tau) = u_c(\cdot)(1-\tau)w - u_\ell(\cdot) = 0,$$

其中 $\ell = T-l$ 是闲暇。

由隐函数定理:

$$\frac{\mathrm{d}l^*}{\mathrm{d}\tau} = -\frac{F_\tau}{F_l}.$$

这里 $F_l$ 反映二阶条件，通常在最大化问题中为负。$F_\tau$ 则包含工资下降的替代效应与收入效应，因而总符号不一定确定。这解释了为什么劳动供给对税率的反应在理论上可能不明确。

**常见误区**

比较静态符号不确定不是失败，而是经济机制冲突的体现。税率上升降低税后工资，替代效应使劳动减少，但收入效应可能使劳动增加。数学上表现为 $F_\tau$ 的符号无法仅凭一般假设确定。

## 10 应用二：投资组合权重对风险厌恶的反应

考虑均值方差投资者：

$$\max_w \; w\mu - \frac{\gamma}{2}\sigma^2 w^2,$$

其中 $w$ 是风险资产权重，$\mu$ 是风险溢价，$\sigma^2$ 是方差，$\gamma > 0$ 是风险厌恶。

FOC：

$$F(w,\gamma) = \mu - \gamma\sigma^2 w = 0.$$

因此

$$w^*(\gamma) = \frac{\mu}{\gamma\sigma^2}.$$

用 IFT：

$$\frac{\mathrm{d}w^*}{\mathrm{d}\gamma} = -\frac{F_\gamma}{F_w}.$$

其中

$$F_\gamma = -\sigma^2 w, \qquad F_w = -\gamma\sigma^2.$$

所以

$$\frac{\mathrm{d}w^*}{\mathrm{d}\gamma} = -\frac{-\sigma^2 w}{-\gamma\sigma^2} = -\frac{w}{\gamma} < 0$$

若 $w > 0$。风险厌恶越强，风险资产配置越低。

多资产版本中

$$\max_w \; w'(\mu - r\mathbf{1}) - \frac{\gamma}{2} w'\Sigma w$$

给出

$$w^* = \frac{1}{\gamma}\Sigma^{-1}(\mu - r\mathbf{1}),$$

因此

$$\frac{\partial w^*}{\partial \gamma} = -\frac{1}{\gamma^2}\Sigma^{-1}(\mu - r\mathbf{1}) = -\frac{1}{\gamma} w^*.$$

## 11 应用三：资产价格对利率的反应

无增长永续股利资产价格为

$$P(r) = \frac{D}{r}.$$

直接求导:

$$P'(r) = -\frac{D}{r^2} < 0.$$

利率越高, 贴现越强, 价格越低。

如果股利以增长率 $g$ 增长, Gordon formula 为

$$P(r,g) = \frac{D_1}{r-g}, \quad r > g.$$

比较静态:

$$\frac{\partial P}{\partial r} = -\frac{D_1}{(r-g)^2} < 0,$$

$$\frac{\partial P}{\partial g} = \frac{D_1}{(r-g)^2} > 0.$$

而且当 $r-g$ 很小时, 导数绝对值非常大。这是久期和估值敏感性的一个基础直觉。

## 12 应用四: 间接效用与 Roy's identity 预告

消费者间接效用函数定义为

$$V(p,m) = \max_x u(x) \quad \text{s.t.} \quad m - p \cdot x = 0.$$

Lagrangian:

$$\mathcal{L} = u(x) + \lambda(m - p \cdot x).$$

包络定理给出

$$V_m = \lambda,$$

$$V_{p_i} = -\lambda x_i^*.$$

因此

$$\boxed{x_i^*(p,m) = -\frac{V_{p_i}(p,m)}{V_m(p,m)}.}$$

这就是 Roy's identity 的形式。它说明需求可以从间接效用函数的价格和收入导数中恢复出来。

**直觉盒子**

Roy's identity 的直觉是: 价格上升使效用下降, 下降幅度等于边际财富价值 $\lambda$ 乘以消费量 $x_i$。所以把价格导数除以收入导数, 就能把货币价值单位消掉, 恢复需求量。

## 13 比较静态与二阶条件

### 13.1 为什么分母经常是二阶导数

无约束一维最大化给出

$$\frac{\mathrm{d}x^*}{\mathrm{d}\theta} = -\frac{f_{x\theta}}{f_{xx}}.$$

分母 $f_{xx}$ 是曲率。若 $f_{xx}$ 很负，目标函数峰很尖，最优选择对参数不敏感。若 $f_{xx}$ 接近 0，目标函数很平，一点参数变化可能导致最优选择大幅移动。

### 13.2 比较静态的经济解释模板

**方法模板**

**比较静态解释模板**

1. 分子：参数如何改变边际收益或均衡压力。
2. 分母：系统自身的回拉力量或二阶曲率。
3. 符号：方向由二者共同决定。
4. 大小：分母越小，系统越敏感；分母越大，系统越稳定。

**Worked Example**

**例 7：线性供需比较静态的一般解释**

市场出清：

$$a - bp = c + dp.$$

均衡压力函数：

$$F(p, a) = a - c - (b + d)p.$$

比较静态：

$$\frac{\mathrm{d}p^*}{\mathrm{d}a} = \frac{1}{b + d}.$$

分子 1 表示需求冲击直接增加 excess demand；分母 $b+d$ 表示价格上升后需求减少和供给增加共同吸收冲击。市场越有弹性，价格反应越小。

## 14 不可微、角点和多重均衡时要小心

隐函数定理和包络定理都有条件。以下情况需要特别谨慎：

- **角点解：** 一阶条件可能不是等号，而是 KKT 不等式。
- **约束切换：** active constraint 变化时，最优选择函数可能不可微。
- **多重均衡：** 同一个参数可能对应多个解，隐函数只在某个分支上局部成立。
- **Jacobian 奇异：** 若 $F_x = 0$ 或 $F_x$ 不可逆，比较静态可能爆炸或发生 bifurcation。
- **非凹问题：** 一阶条件分支可能不是全局最优分支。

**常见误区**

看见 $F(x,\theta)=0$ 不要机械套公式。先检查：解是否局部唯一？$F_x$ 是否非零？当前解是否仍是最优或均衡？是否有约束切换？这些问题在金融危机、多重均衡、贫困陷阱和政策门槛模型中尤其重要。

## 15 本章小结图谱

| 概念 | 核心公式 | 经济含义 |
| --- | --- | --- |
| 隐函数定理 | $x_\theta = -F_\theta / F_x$ | 方程定义的均衡/最优选择如何随参数变化。 |
| 多变量 IFT | $D_\theta x = -F_x^{-1} F_\theta$ | 联立方程系统的比较静态。 |
| 无约束最优选择 | $x_\theta^* = -f_{x\theta}/f_{xx}$ | 参数改变边际收益，曲率决定反应强度。 |
| 约束优化比较静态 | $z_\theta^* = -F_z^{-1} F_\theta$ | FOC 和约束共同定义选择与乘子。 |
| 包络定理 | $V_\theta = f_\theta(x^*, \theta)$ | 价值函数导数只看直接效应。 |
| 约束包络 | $V_\theta = \mathcal{L}_\theta$ | 参数改变约束时，乘子衡量影子价值。 |
| Shephard | $C_{w_i} = x_i^*$ | 成本函数对要素价格的导数等于条件要素需求。 |
| Roy | $x_i = -V_{p_i}/V_m$ | 需求可由间接效用导数恢复。 |
| Hotelling | $\pi_p = y^*$, $\pi_w = -x^*$ | 利润函数导数给出供给和要素需求。 |

# 16 练习题

## A 组：基础计算

A1. 设 $F(x,\theta)=x^2+\theta x-1=0$。求 $\mathrm{d}x/\mathrm{d}\theta$ 的一般表达式。

A2. 设市场均衡由 $a-bp=c+dp$ 定义。求 $\mathrm{d}p^*/\mathrm{d}a$ 和 $\mathrm{d}p^*/\mathrm{d}c$。

A3. 设 $f(x,\theta)=\theta\ln x-x$, $x>0$。求 $x^*(\theta)$ 和 $\mathrm{d}x^*/\mathrm{d}\theta$。

A4. 设 $V(\theta)=\max_x\{\theta x-x^2\}$。用直接法和包络定理分别求 $V'(\theta)$。

A5. 设约束问题 $V(m)=\max_{x\geq 0}\ln x$ s.t. $x=m$。求 $V'(m)$ 并解释乘子。

## B 组：概念判断

B1. 隐函数定理为什么要求 $F_x\neq 0$?

B2. 包络定理为什么可以忽略 $\mathrm{d}x^*/\mathrm{d}\theta$?

B3. 多重均衡时，为什么比较静态可能只在局部分支上成立?

B4. 若目标函数非常平，为什么最优选择对参数更敏感?

B5. 约束包络定理中 Lagrange multiplier 为什么可以解释为影子价格?

## C 组：混合题

C1. 垄断者需求 $P(Q)=a-bQ$，成本 $C(Q)=cQ+F$。求 $Q^*$ 对 $a,b,c$ 的比较静态。

C2. Solow 稳态 $sk^\alpha=(n+\delta)k$。求 $\partial k^*/\partial s$, $\partial k^*/\partial n$, $\partial k^*/\partial\delta$ 的符号。

C3. 设 $F_1(x,y,\theta)=x+y-\theta=0$, $F_2(x,y,\theta)=x-y^2=0$。写出矩阵公式求 $\partial(x,y)/\partial\theta$。

C4. 投资组合问题 $\max_w w\mu-\gamma\sigma^2w^2/2$。求 $\partial w^*/\partial\mu$, $\partial w^*/\partial\sigma^2$, $\partial w^*/\partial\gamma$。

C5. 设 $V(p,m)=\max_{x_1,x_2}\{\ln x_1+\ln x_2:p_1x_1+p_2x_2=m\}$。用包络定理求 $V_m$ 和 $V_{p_i}$，并验证 Roy's identity。

## D 组：经济金融应用题

D1. 解释为什么利率上升会降低永续债价格，并用导数说明利率越低时价格越敏感。

D2. 在劳动供给模型中，为什么税率对劳动供给的总效应可能符号不确定?

D3. 企业利润函数 $\pi(p,w)$ 对产出价格和工资的导数分别有什么经济含义?

D4. 成本函数 $C(w,q)$ 对工资的导数为什么等于条件劳动需求?

D5. 在多重均衡模型中，为什么政策参数的小变化可能导致均衡跳跃? 这和 IFT 的条件有什么关系?

### E 组：挑战 / 证明题

E1. 从链式法则严格推导标量隐函数公式 $x_\theta = -F_\theta / F_x$。

E2. 从 $V(\theta) = f(x^*(\theta), \theta)$ 出发推导无约束包络定理。

E3. 对约束问题 $\max_x f(x, \theta)$ s.t. $g(x, \theta) = 0$, 推导 $V_\theta = \mathcal{L}_\theta$。

E4. 设 $x^*(\theta)$ 是最大化问题的一阶条件分支。说明为什么若二阶条件失效, 该分支的比较静态未必有经济意义。

E5. 用矩阵微分推导多变量比较静态公式 $D_\theta x = -F_x^{-1} F_\theta$。

# 17 练习题参考答案

## A 组答案

A1. 对 $F(x,\theta)=0$ 用隐函数公式。这里

$$F_x = 2x+\theta, \qquad F_\theta = x.$$

因此

$$\frac{\mathrm{d}x}{\mathrm{d}\theta} = -\frac{x}{2x+\theta}.$$

A2. 均衡可写为

$$F(p,a,c) = a - bp - c - dp = 0.$$

$$F_p = -(b+d), \quad F_a = 1, \quad F_c = -1.$$

所以

$$\frac{\partial p^*}{\partial a} = \frac{1}{b+d}, \qquad \frac{\partial p^*}{\partial c} = -\frac{1}{b+d}.$$

需求上升推高价格，供给截距上升压低价格。

A3. 一阶条件：

$$f_x = \frac{\theta}{x} - 1 = 0.$$

所以

$$x^*(\theta) = \theta.$$

因此 $\mathrm{d}x^*/\mathrm{d}\theta = 1$。用公式也可得

$$f_{x\theta} = 1/x, \qquad f_{xx} = -\theta/x^2.$$

在 $x^*=\theta$ 处，

$$-\frac{f_{x\theta}}{f_{xx}} = -\frac{1/\theta}{-1/\theta} = 1.$$

A4. 一阶条件：

$$\theta - 2x = 0, \qquad x^* = \theta/2.$$

价值函数：

$$V(\theta) = \theta(\theta/2) - (\theta/2)^2 = \theta^2/4.$$

所以

$$V'(\theta) = \theta/2.$$

包络定理：$f_\theta = x$，在最优点处为 $x^*=\theta/2$。一致。

A5. 约束 $x = m$ 使 $V(m) = \ln m$。所以

$$V'(m) = 1/m.$$

Lagrangian 可写为 $\mathcal{L} = \ln x + \lambda(m - x)$。FOC 给出 $1/x - \lambda = 0$, 故 $\lambda = 1/x = 1/m$。乘子等于资源 $m$ 增加一单位带来的最大价值增加。

## B 组答案

B1. $F_x \neq 0$ 表示方程左边对内生变量有局部反应, 因而可以用内生变量的变化抵消参数变化。如果 $F_x = 0$, 曲线可能垂直、交叉或出现折点, 局部不一定能解出唯一函数。

B2. 因为在内部最优点 $f_x = 0$。对 $V(\theta) = f(x^*(\theta), \theta)$ 求导时, $f_x dx^*/d\theta$ 项为零, 只剩直接效应 $f_\theta$。

B3. 多重均衡意味着同一参数可能对应多个解。IFT 只能在某个满足非奇异条件的解附近给出局部分支, 不能说明全局会选择哪个分支。

B4. 目标函数很平意味着二阶导数绝对值小。公式 $x^*_\theta = -f_{x\theta}/f_{xx}$ 中, 分母小会使反应大。

B5. 约束放松一单位对最优价值的边际影响由乘子衡量。包络定理给出 $V_m = \lambda$, 所以乘子是资源或约束的影子价格。

## C 组答案

C1. 利润:

$$\pi = (a - bQ)Q - cQ - F.$$

FOC:

$$a - 2bQ - c = 0.$$

因此

$$Q^* = \frac{a - c}{2b}.$$

比较静态:

$$Q^*_a = \frac{1}{2b} > 0, \qquad Q^*_c = -\frac{1}{2b} < 0, \qquad Q^*_b = -\frac{a - c}{2b^2} < 0$$

若 $a > c$。

C2. 写

$$F(k, s, n, \delta) = sk^\alpha - (n + \delta)k = 0.$$

在稳定稳态处 $F_k < 0$。有

$$F_s = k^\alpha > 0, \quad F_n = -k < 0, \quad F_\delta = -k < 0.$$

因此

$$k^*_s = -F_s/F_k > 0,$$

$$k_n^* = -F_n/F_k < 0, \qquad k_\delta^* = -F_\delta/F_k < 0.$$

C3. 方程系统

$$F = \begin{pmatrix} x + y - \theta \\ x - y^2 \end{pmatrix} = 0.$$

$$F_{(x,y)} = \begin{pmatrix} 1 & 1 \\ 1 & -2y \end{pmatrix}, \qquad F_\theta = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

所以

$$\begin{pmatrix} x_\theta \\ y_\theta \end{pmatrix} = -\begin{pmatrix} 1 & 1 \\ 1 & -2y \end{pmatrix}^{-1} \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

C4. 最优权重

$$w^* = \frac{\mu}{\gamma\sigma^2}.$$

因此

$$\frac{\partial w^*}{\partial \mu} = \frac{1}{\gamma\sigma^2} > 0,$$

$$\frac{\partial w^*}{\partial \sigma^2} = -\frac{\mu}{\gamma(\sigma^2)^2} < 0$$

若 $\mu > 0$,

$$\frac{\partial w^*}{\partial \gamma} = -\frac{\mu}{\gamma^2\sigma^2} < 0.$$

C5. Cobb-Douglas 需求为

$$x_1^* = \frac{m}{2p_1}, \qquad x_2^* = \frac{m}{2p_2}.$$

间接效用

$$V(p,m) = \ln\frac{m}{2p_1} + \ln\frac{m}{2p_2} = 2\ln m - \ln p_1 - \ln p_2 - 2\ln 2.$$

因此

$$V_m = 2/m, \qquad V_{p_1} = -1/p_1, \qquad V_{p_2} = -1/p_2.$$

Roy:

$$-\frac{V_{p_1}}{V_m} = \frac{1/p_1}{2/m} = \frac{m}{2p_1} = x_1^*.$$

同理得到 $x_2^*$。

## D 组答案

D1. 永续债价格 $P = D/r$。导数 $P_r = -D/r^2 < 0$, 所以利率上升降低价格。且 $|P_r| = D/r^2$, 当 $r$ 越低时, 导数绝对值越大, 价格对利率越敏感。

D2. 税率上升降低税后工资。替代效应使闲暇变便宜, 劳动供给下降; 但收入效应可能使消费者

为了维持消费而增加劳动。两者方向相反, 因此总效应不确定。

D3. 包络定理给出 $\pi_p = y^*$, 产出价格上升一单位提高利润的幅度等于最优产量; $\pi_w = -x^*$, 工资上升一单位降低利润的幅度等于最优要素使用量。

D4. 成本函数是以最优要素组合达到产量 $q$ 的最小成本。工资上升一点时, 直接增加的成本等于原最优劳动使用量; 最优劳动调整的一阶影响由包络定理消掉。因此 $C_w = L^*$。

D5. 多重均衡通常意味着方程有多个解分支。当参数变化导致某个分支消失或失去稳定性时, 均衡可能跳到另一个分支。数学上常发生在 Jacobian 奇异或 IFT 条件失效附近。

## E 组答案

E1. 若 $F(x(\theta), \theta) = 0$, 对 $\theta$ 求导:

$$F_x(x(\theta), \theta)x_\theta + F_\theta(x(\theta), \theta) = 0.$$

若 $F_x \neq 0$, 则

$$x_\theta = -F_\theta / F_x.$$

E2. $V(\theta) = f(x^*(\theta), \theta)$。求导:

$$V_\theta = f_x(x^*, \theta)x^*_\theta + f_\theta(x^*, \theta).$$

内部最优给出 $f_x(x^*, \theta) = 0$, 因此

$$V_\theta = f_\theta(x^*, \theta).$$

E3. $V(\theta) = f(x^*(\theta), \theta)$, 约束 $g(x^*(\theta), \theta) = 0$。对价值求导:

$$V_\theta = f_x x_\theta + f_\theta.$$

Lagrangian FOC 为 $f_x + \lambda g_x = 0$, 所以 $f_x = -\lambda g_x$。约束求导:

$$g_x x_\theta + g_\theta = 0.$$

因此

$$V_\theta = -\lambda g_x x_\theta + f_\theta = \lambda g_\theta + f_\theta = \mathcal{L}_\theta.$$

E4. 一阶条件分支只说明 $f_x = 0$。若二阶条件失效, 该点可能是局部最小、鞍点, 或只是局部最优而非全局最优。此时沿该分支求得的比较静态可能描述的是一个非最优点, 经济上没有选择含义。

E5. 若 $F(x(\theta), \theta) = 0$, 对向量参数求导:

$$F_x D_\theta x + F_\theta = 0.$$

若 $F_x$ 可逆, 左乘 $F_x^{-1}$:

$$D_\theta x = -F_x^{-1} F_\theta.$$

## 18 本章到下一章的连接

本章已经建立了两个核心工具: 隐函数定理和包络定理。下一章将把它们用于经济学中最重要的对偶结构: 支出函数、间接效用函数、成本函数、利润函数和需求/供给函数之间的关系。你会看到很多看似神秘的经典公式其实都来自同一条线:

$$\text{优化问题} \Rightarrow \text{价值函数} \Rightarrow \text{包络定理} \Rightarrow \text{需求/供给/要素需求}.$$

**直觉盒子**

如果说 Chapter 14-17 让你会“求最优”, Chapter 18 则让你会“研究最优如何变化”。这是从做题计算进入经济理论分析的重要一步。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 19
# 对偶性、经济函数与包络引理

Chapter 18 已经讲了隐函数定理和包络定理。本章把这些工具放进经济学最常用的一组“经济函数”中: indirect utility, expenditure function, cost function, profit function, Marshallian demand, Hicksian demand, conditional factor demand, supply function。你会看到, 很多看似不同的结论, 例如 Roy’s identity, Shephard’s lemma, Hotelling’s lemma, Slutsky equation 的入口, 本质上都是同一个思想: **最优化问题的价值函数对参数求导时, 可以直接对目标函数或约束里的参数求偏导**。

math4econ personalized notes
2026

# 目录

# Chapter 19 总览

| | |
|---|---|
| **本章主题** | duality, economic functions, value functions, envelope lemmas, demand and supply derivatives。 |
| **核心问题** | 一个经济选择问题可以从两个方向看：给定预算最大化效用，或给定效用最小化支出；给定产量最小化成本，或给定价格最大化利润。这些 primal 与 dual 问题之间怎样相互转化？价值函数的导数怎样给出需求、供给、成本和福利变化？ |
| **数学工具** | Lagrange/KKT, convexity, envelope theorem, homogeneity, duality identities, Roy's identity, Shephard's lemma, Hotelling's lemma, Slutsky decomposition。 |
| **经济应用** | 消费者需求、福利测度、价格指数、成本函数、条件要素需求、利润函数、供给函数、投入需求、均值方差投资组合的对偶解释。 |
| **后续连接** | Chapter 20 固定点与均衡存在性, Chapter 21 变分法与 Euler equation, Chapter 22 动态规划, 高级微观中的一般均衡、福利经济学和机制设计。 |

**直觉盒子**

对偶性 duality 的直觉是：**同一个经济对象可以从“最大化收益”或“最小化成本”两个角度刻画**。例如消费者问题可以问：给定收入 $w$，最多能达到多少效用？也可以问：给定目标效用 $u$，最少需要多少支出？两个问题看似方向相反，但在良好条件下描述的是同一条无差异曲线和同一个切点。

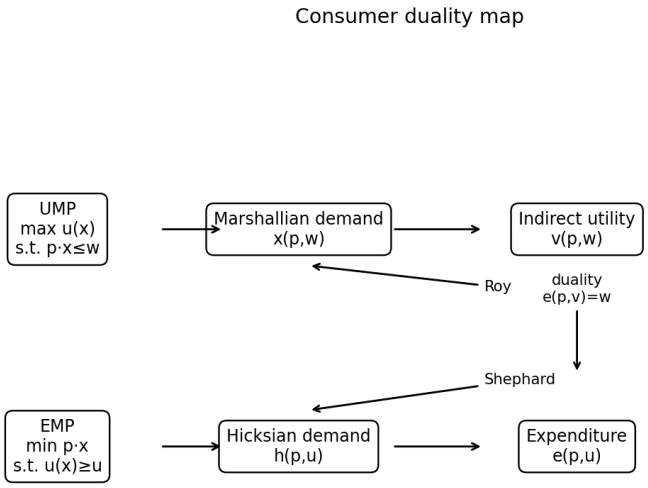



图 1: 消费者对偶性地图: UMP, EMP, Marshallian demand, Hicksian demand, indirect utility, expenditure function。

# 1 为什么经济学需要对偶性

在很多课程中, 消费者问题被写成

$$\max_{\boldsymbol{x}\geq 0} u(\boldsymbol{x}) \quad \text{s.t.} \quad \boldsymbol{p}\cdot\boldsymbol{x}\leq w.$$

这当然重要, 但如果只停留在这个问题, 很多高级微观和福利分析会变得笨重。对偶性告诉我们, 还可以换一个角度:

$$\min_{\boldsymbol{x}\geq 0} \boldsymbol{p}\cdot\boldsymbol{x} \quad \text{s.t.} \quad u(\boldsymbol{x})\geq \bar{u}.$$

第一个问题给出“在收入约束下能达到的最高效用”; 第二个问题给出“达到某个效用至少需要多少钱”。两者之间的互相转化, 是需求理论、价格指数、福利测度和 Slutsky 方程的基础。

企业理论中也完全类似。给定产量最小化成本:

$$c(\boldsymbol{w},q)=\min_{\boldsymbol{x}\geq 0}\ \boldsymbol{w}\cdot\boldsymbol{x} \quad \text{s.t.} \quad f(\boldsymbol{x})\geq q.$$

给定价格最大化利润:

$$\pi(p,\boldsymbol{w})=\max_{q,\boldsymbol{x}}\ pq-\boldsymbol{w}\cdot\boldsymbol{x} \quad \text{s.t.} \quad q\leq f(\boldsymbol{x}).$$

这些经济函数不只是“结果”, 它们本身就是模型的核心对象。

## 1.1 从 choice function 到 value function

一个优化问题一般有两类输出:

- **choice function**: 最优选择本身, 例如 $x(p,w)$、$h(p,u)$、$z(w,q)$;

- **value function**: 最优目标值, 例如 $v(p,w)$、$e(p,u)$、$c(w,q)$、$\pi(p,w)$。

很多时候, 价值函数比选择函数更稳健、更容易比较静态。原因是包络定理: 价值函数对参数的导数可以直接读出最优选择或影子价格。

**常见误区**

对偶性不是“把最大化换成最小化”的形式游戏。它依赖偏好或技术的凸性、单调性、连续性等条件。条件不好时, primal 和 dual 之间可能有 gap, 或需求不唯一, 或价值函数不可微。

# 2 消费者 primal: Utility Maximization Problem

## 2.1 问题设定

消费者的效用最大化问题 UMP 是

$$v(\boldsymbol{p},w)=\max_{\boldsymbol{x}\geq 0}\ u(\boldsymbol{x})\quad \text{s.t.}\quad \boldsymbol{p}\cdot\boldsymbol{x}\leq w.$$

其中 $\boldsymbol{p}\gg 0$ 是价格向量, $w>0$ 是财富或收入。最优解称为 Marshallian demand:

$$\boldsymbol{x}(\boldsymbol{p},w)\in\arg\max_{\boldsymbol{x}\geq 0}\{u(\boldsymbol{x}):\boldsymbol{p}\cdot\boldsymbol{x}\leq w\}.$$

价值函数 $v(\boldsymbol{p},w)$ 称为 indirect utility function。

**定义 / 定理**

**Marshallian demand 与 indirect utility.**

$$\boldsymbol{x}(\boldsymbol{p},w)=\arg\max_{\boldsymbol{x}\geq 0}\{u(\boldsymbol{x}):\boldsymbol{p}\cdot\boldsymbol{x}\leq w\},$$

$$v(\boldsymbol{p},w)=u(\boldsymbol{x}(\boldsymbol{p},w)).$$

Marshallian demand 描述在给定价格和收入下实际购买多少; indirect utility 描述由价格和收入间接决定的最高效用。

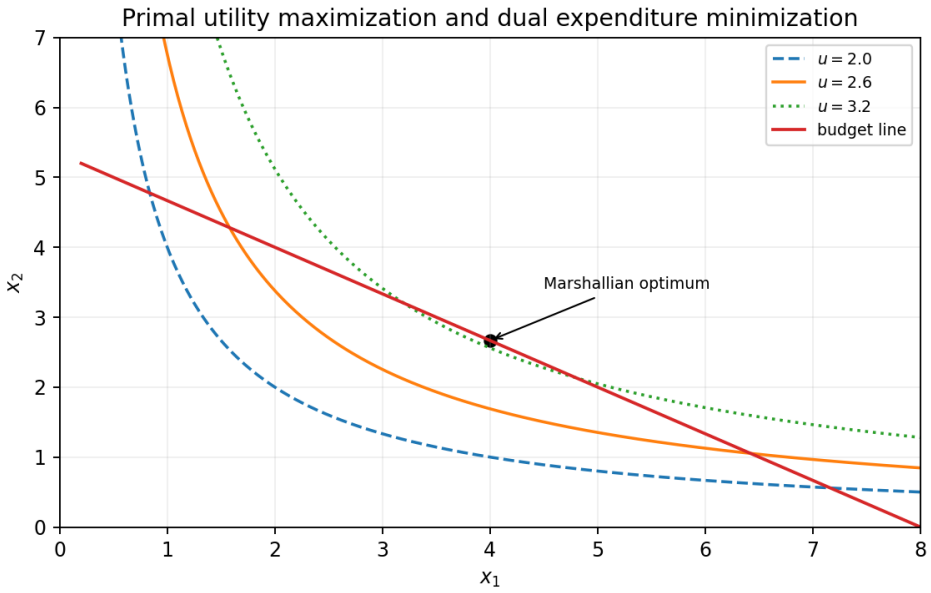



图 2: 效用最大化与支出最小化在同一个切点相遇。

## 2.2 Lagrange 条件与边际替代率

若内点解存在, Lagrangian 为

$$\mathcal{L} = u(\boldsymbol{x}) + \lambda(w - \boldsymbol{p} \cdot \boldsymbol{x}).$$

FOC 为

$$u_i(\boldsymbol{x}) = \lambda p_i, \qquad i = 1, \dots, n.$$

所以任意两种商品满足

$$\frac{u_i}{u_j} = \frac{p_i}{p_j}.$$

这就是 MRS equals price ratio。它说明最优点处, 无差异曲线斜率与预算线斜率相同。

## 2.3 间接效用函数的性质

在标准偏好下, $v(\boldsymbol{p}, w)$ 有以下性质:

1. 对 $w$ 非递减: 收入越高, 可行集越大;
2. 对价格 $p_i$ 非递增: 某商品变贵, 可行集收缩;
3. 零次齐次: 对任意 $t > 0$,
$$v(t\boldsymbol{p}, tw) = v(\boldsymbol{p}, w).$$
因为价格和收入同比例放大, 预算集不变;
4. 在价格上通常 quasi-convex, 但不一定 concave;
5. 在良好条件下连续, 且可用包络定理求导。

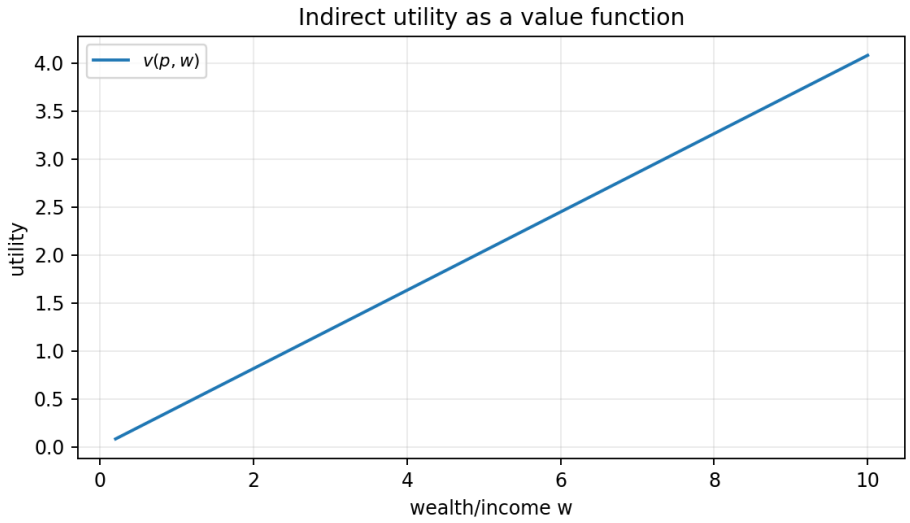



图 3: 固定价格下, 间接效用函数随收入上升。

# 3 消费者 dual: Expenditure Minimization Problem

## 3.1 问题设定

支出最小化问题 EMP 是

$$e(\boldsymbol{p}, u) = \min_{\boldsymbol{x} \geq 0} \ \boldsymbol{p} \cdot \boldsymbol{x} \quad \text{s.t.} \quad u(\boldsymbol{x}) \geq u.$$

最优解称为 Hicksian demand 或 compensated demand:

$$\boldsymbol{h}(\boldsymbol{p}, u) \in \arg\min_{\boldsymbol{x} \geq 0} \{\boldsymbol{p} \cdot \boldsymbol{x} : u(\boldsymbol{x}) \geq u\}.$$

价值函数 $e(\boldsymbol{p}, u)$ 称为 expenditure function。

**定义 / 定理**

**Hicksian demand 与 expenditure function.**

$$\boldsymbol{h}(\boldsymbol{p}, u) = \arg\min_{\boldsymbol{x} \geq 0} \{\boldsymbol{p} \cdot \boldsymbol{x} : u(\boldsymbol{x}) \geq u\},$$

$$e(\boldsymbol{p}, u) = \boldsymbol{p} \cdot \boldsymbol{h}(\boldsymbol{p}, u).$$

Hicksian demand 表示“为了达到效用 $u$, 在价格 $\boldsymbol{p}$ 下成本最低的消费束”。它剔除了价格变化带来的收入效应, 因此又叫 compensated demand。

## 3.2 支出函数的性质

在标准条件下, $e(\boldsymbol{p}, u)$ 满足:

1. 对价格 $p_i$ 非递减;
2. 对目标效用 $u$ 非递减;
3. 对价格一阶齐次:
$$e(t\boldsymbol{p}, u) = te(\boldsymbol{p}, u), \qquad t > 0;$$
4. 对价格 concave? 注意这里容易错。支出函数作为价格的函数是**凹?** 实际上它是价格的**凹或凸?** 让我们谨慎: $e(p, u) = \min_x p \cdot x$ 是线性函数族的下包络, 因此对价格是 concave? 线性函数的点态下确界是凹函数。常见教材表述为 expenditure function is concave in prices;
5. 若 Hicksian demand 唯一, 则可微, 且由 Shephard's lemma 给出导数。

**常见误区**

要区分两个“下包络/上包络”: 利润函数是线性收益的上包络, 所以对价格是 convex; 支出函数和成本函数是线性成本的下包络, 所以对价格是 concave。这一点在凸分析里很重要, 但很多入门教材不会强调。

## 4 对偶恒等式: v 与 e 如何互相反解

消费者对偶性的核心恒等式是

$$\boxed{e(\boldsymbol{p}, v(\boldsymbol{p}, w)) = w}$$

以及

$$\boxed{v(\boldsymbol{p}, e(\boldsymbol{p}, u)) = u.}$$

第一条意思是: 在价格 $\boldsymbol{p}$ 下, 达到“收入 $w$ 能买到的最高效用”所需的最小支出正好是 $w$。第二条意思是: 花“达到效用 $u$ 的最低支出”所能得到的最高效用正好是 $u$。

这些恒等式成立需要局部非饱和等条件, 使预算约束在最优时用尽。

**直觉盒子**

UMP 问: 给我 $w$ 元, 我最多多快乐? EMP 问: 给我一个快乐水平 $u$, 我最少要花多少钱? 如果偏好良好, 二者像温度和热量一样互相换算: $v$ 把钱转成效用, $e$ 把效用转成钱。

## 5 Shephard's Lemma: 支出函数导数就是 Hicksian demand

### 5.1 定理表述

**定义 / 定理**

**Shephard's lemma, 消费者版本.** 若 $e(\boldsymbol{p}, u)$ 在 $\boldsymbol{p}$ 处可微, 且 Hicksian demand 唯一, 则

$$\boxed{\frac{\partial e(\boldsymbol{p}, u)}{\partial p_i} = h_i(\boldsymbol{p}, u).}$$

向量形式为

$$\boxed{\nabla_{\boldsymbol{p}} e(\boldsymbol{p}, u) = \boldsymbol{h}(\boldsymbol{p}, u).}$$

### 5.2 包络推导

EMP 的目标函数是

$$\boldsymbol{p} \cdot \boldsymbol{x} = \sum_i p_i x_i.$$

在最优消费束 $\boldsymbol{h}(\boldsymbol{p}, u)$ 处, 价值函数为

$$e(\boldsymbol{p}, u) = \boldsymbol{p} \cdot \boldsymbol{h}(\boldsymbol{p}, u).$$

如果直接对 $p_i$ 求导, 似乎要追踪 $h_j$ 如何随 $p_i$ 变化。但包络定理告诉我们: 最优选择变化项因为 FOC 被消掉, 所以只需对目标函数中显式的 $p_i$ 求偏导:

$$\frac{\partial}{\partial p_i}(\boldsymbol{p} \cdot \boldsymbol{x}) = x_i.$$

在最优点代入 $x_i = h_i(\boldsymbol{p}, u)$, 得

$$\frac{\partial e}{\partial p_i} = h_i(\boldsymbol{p}, u).$$

**直觉盒子**

价格 $p_i$ 上升一小点, 达到同样效用所需最低支出会上升多少? 如果最优补偿需求中购买 $h_i$ 单位商品 $i$, 那么一单位价格上升造成的支出增加就是 $h_i$。这就是 Shephard 引理的经济含义。

## 6 Roy's Identity: 间接效用导数给出 Marshallian demand

### 6.1 定理表述

**定义 / 定理**

**Roy's identity.** 若 $v(\boldsymbol{p}, w)$ 可微且 $v_w(\boldsymbol{p}, w) > 0$, 则 Marshallian demand 满足

$$\boxed{x_i(\boldsymbol{p}, w) = -\frac{\partial v(\boldsymbol{p}, w)/\partial p_i}{\partial v(\boldsymbol{p}, w)/\partial w}.}$$

向量形式为

$$\boxed{\boldsymbol{x}(\boldsymbol{p}, w) = -\frac{\nabla_{\boldsymbol{p}} v(\boldsymbol{p}, w)}{v_w(\boldsymbol{p}, w)}.}$$

### 6.2 从包络定理推导

UMP 的 Lagrangian 是

$$\mathcal{L} = u(\boldsymbol{x}) + \lambda(w - \boldsymbol{p} \cdot \boldsymbol{x}).$$

由包络定理,

$$\frac{\partial v}{\partial w} = \lambda,$$

而

$$\frac{\partial v}{\partial p_i} = -\lambda x_i.$$

因此

$$x_i = -\frac{v_{p_i}}{v_w}.$$

**直觉盒子**

$p_i$ 上升会降低间接效用, 降低多少取决于消费者买多少商品 $i$ 以及收入的边际效用 $\lambda$。除以 $v_w$ 是把“效用损失”换算回“收入单位”, 所以得到需求量。

## 7 Cobb-Douglas 消费者: 完整双重例子

**Worked Example**

**例 1: Cobb-Douglas UMP, EMP, Roy, Shephard**

设

$$u(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha}, \qquad 0 < \alpha < 1.$$

## 1. Utility maximization

UMP 为

$$\max x_1^\alpha x_2^{1-\alpha} \quad \text{s.t.} \quad p_1x_1 + p_2x_2 = w.$$

FOC 给出

$$\frac{\alpha}{1-\alpha}\frac{x_2}{x_1} = \frac{p_1}{p_2}.$$

结合预算约束, 得 Marshallian demand:

$$\boxed{x_1(\boldsymbol{p}, w) = \frac{\alpha w}{p_1}, \qquad x_2(\boldsymbol{p}, w) = \frac{(1-\alpha)w}{p_2}.}$$

间接效用为

$$\boxed{v(\boldsymbol{p}, w) = \left(\frac{\alpha w}{p_1}\right)^\alpha \left(\frac{(1-\alpha)w}{p_2}\right)^{1-\alpha} = \alpha^\alpha(1-\alpha)^{1-\alpha}\frac{w}{p_1^\alpha p_2^{1-\alpha}}.}$$

## 2. Expenditure minimization

EMP 为

$$\min p_1x_1 + p_2x_2 \quad \text{s.t.} \quad x_1^\alpha x_2^{1-\alpha} \geq u.$$

Hicksian demand 为

$$\boxed{h_1(\boldsymbol{p}, u) = \alpha\frac{e(\boldsymbol{p}, u)}{p_1}, \qquad h_2(\boldsymbol{p}, u) = (1-\alpha)\frac{e(\boldsymbol{p}, u)}{p_2}.}$$

支出函数是

$$\boxed{e(\boldsymbol{p}, u) = \frac{u\,p_1^\alpha p_2^{1-\alpha}}{\alpha^\alpha(1-\alpha)^{1-\alpha}}.}$$

## 3. Shephard 检查

对 $p_1$ 求偏导:

$$\frac{\partial e}{\partial p_1} = \alpha\frac{up_1^{\alpha-1}p_2^{1-\alpha}}{\alpha^\alpha(1-\alpha)^{1-\alpha}} = \alpha\frac{e}{p_1} = h_1(\boldsymbol{p}, u).$$

同理 $\partial e/\partial p_2 = h_2$。

## 4. Roy 检查

由

$$v = C\frac{w}{p_1^\alpha p_2^{1-\alpha}},$$

有

$$v_w = \frac{v}{w}, \qquad v_{p_1} = -\alpha \frac{v}{p_1}.$$

所以

$$-\frac{v_{p_1}}{v_w} = \frac{\alpha v/p_1}{v/w} = \frac{\alpha w}{p_1} = x_1(\boldsymbol{p}, w).$$

## 8 Slutsky 分解：需求变化的两种效应

价格变化对 Marshallian demand 的影响可分解为 substitution effect 和 income effect。核心关系是 Hicksian demand 与 Marshallian demand 的连接：

$$\boldsymbol{h}(\boldsymbol{p}, u) = \boldsymbol{x}(\boldsymbol{p}, e(\boldsymbol{p}, u)).$$

特别地，若 $u = v(\boldsymbol{p}, w)$，则

$$\boldsymbol{h}(\boldsymbol{p}, v(\boldsymbol{p}, w)) = \boldsymbol{x}(\boldsymbol{p}, w).$$

对价格求导可以得到 Slutsky equation：

$$\boxed{\frac{\partial x_i}{\partial p_j} = \frac{\partial h_i}{\partial p_j} - x_j \frac{\partial x_i}{\partial w}.}$$

**定义 / 定理**

**Slutsky equation，直观版。**

$$\text{总价格效应} = \text{替代效应} + \text{收入效应}.$$

更具体地：

$$\frac{\partial x_i}{\partial p_j} = \underbrace{\frac{\partial h_i}{\partial p_j}}_{\text{substitution effect}} - \underbrace{x_j \frac{\partial x_i}{\partial w}}_{\text{income effect}}.$$

**直觉盒子**

Hicksian demand 固定效用，所以只反映相对价格变化导致的替代行为。Marshallian demand 固定收入，所以价格上升还会使消费者实际购买力下降。Slutsky 方程把这两种效应拆开。

## 9 福利测度：EV, CV 与价格指数

对偶理论的一个重要用途是福利测度。假设价格从 $\boldsymbol{p}^0$ 变为 $\boldsymbol{p}^1$，初始收入为 $w$。

## 9.1 Equivalent Variation

Equivalent variation, EV, 问的是: 在旧价格 $\boldsymbol{p}^0$ 下, 要拿走多少钱, 才会使消费者像价格变化后一样差? 定义为

$$EV = e(\boldsymbol{p}^0, u^1) - e(\boldsymbol{p}^0, u^0),$$

其中

$$u^0 = v(\boldsymbol{p}^0, w), \qquad u^1 = v(\boldsymbol{p}^1, w).$$

若价格上升导致效用下降, EV 通常为负或表示愿意支付避免变化的金额。

## 9.2 Compensating Variation

Compensating variation, CV, 问的是: 在新价格 $\boldsymbol{p}^1$ 下, 要补偿多少钱, 才能让消费者回到旧效用? 定义为

$$CV = e(\boldsymbol{p}^1, u^1) - e(\boldsymbol{p}^1, u^0)$$

或根据符号约定写成补偿金额

$$e(\boldsymbol{p}^1, u^0) - e(\boldsymbol{p}^1, u^1).$$

不同教材符号略有差别, 关键是要说清楚“给钱”还是“拿钱”。

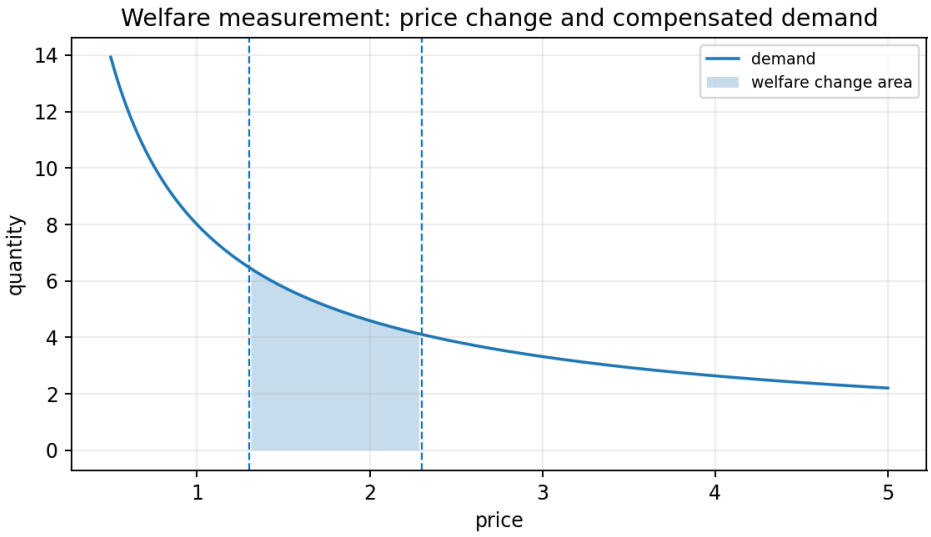



图 4: 价格变化的福利测度可用补偿需求曲线下的面积来表达。

**常见误区**

EV/CV 的符号很容易混乱。考试或写作中, 不要只写公式, 要用一句话说明: 我定义的是“消费者愿意支付多少避免价格变化”, 还是“需要补偿多少接受价格变化”。

# 10 生产者 dual: Cost Minimization

## 10.1 成本最小化问题

企业要生产产量 $q$, 要素价格为 $\boldsymbol{w} \gg 0$, 生产函数为 $f(\boldsymbol{x})$。成本最小化问题是

$$c(\boldsymbol{w}, q) = \min_{\boldsymbol{x} \geq 0} \ \boldsymbol{w} \cdot \boldsymbol{x} \quad \text{s.t.} \quad f(\boldsymbol{x}) \geq q.$$

最优投入称为 conditional factor demand:

$$\boldsymbol{x}(\boldsymbol{w}, q) \in \arg\min\{\boldsymbol{w} \cdot \boldsymbol{x} : f(\boldsymbol{x}) \geq q\}.$$

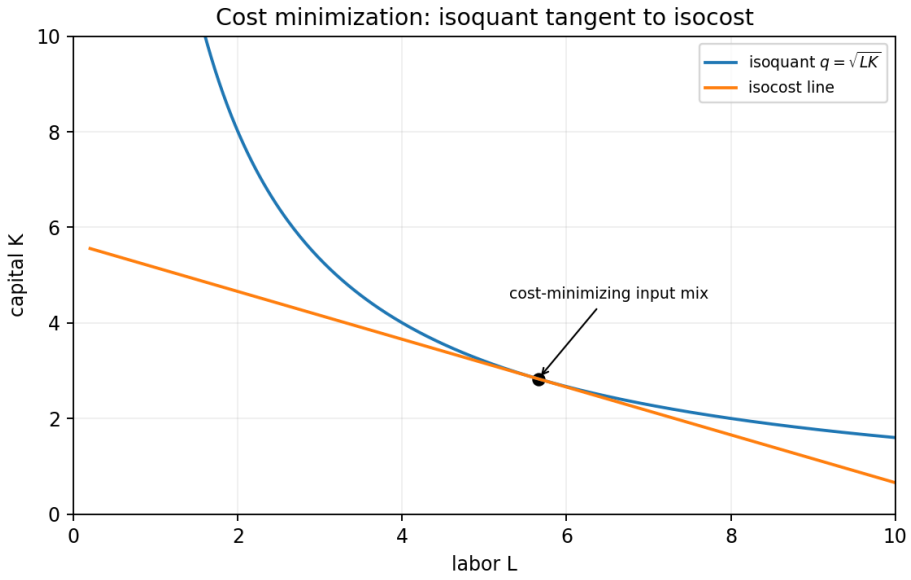



图 5: 成本最小化: 等产量线与等成本线相切。

## 10.2 成本函数性质

标准条件下, 成本函数 $c(\boldsymbol{w}, q)$ 满足:

1. 对要素价格非递减;
2. 对产量 $q$ 非递减;
3. 对要素价格一阶齐次:
$$c(t\boldsymbol{w}, q) = tc(\boldsymbol{w}, q), \quad t > 0;$$
4. 对要素价格 concave, 因为它是线性成本函数的下包络;
5. 若条件要素需求唯一, 则可微。

# 11 Shephard's Lemma: 成本函数版本

**定义 / 定理**

**Shephard's lemma, producer version.** 若成本函数 $c(\boldsymbol{w}, q)$ 对 $w_i$ 可微, 则

$$\boxed{\frac{\partial c(\boldsymbol{w}, q)}{\partial w_i} = x_i(\boldsymbol{w}, q),}$$

其中 $x_i(\boldsymbol{w}, q)$ 是条件要素需求。

经济解释非常直接: 工资 $w_i$ 上升一小点, 为生产同样产量 $q$ 所需最低成本增加多少? 如果最优方案中使用 $x_i$ 单位要素 $i$, 成本增加就是 $x_i$。

**Worked Example**

**例 2: Cobb-Douglas 技术的成本函数**

设生产函数

$$q = L^{\alpha} K^{1-\alpha}, \qquad 0 < \alpha < 1.$$

要素价格为 $w$ 和 $r$。成本最小化:

$$\min_{L,K} \; wL + rK \quad \text{s.t.} \quad L^{\alpha} K^{1-\alpha} = q.$$

FOC 给出

$$\frac{\alpha}{1-\alpha}\frac{K}{L} = \frac{w}{r}.$$

整理得

$$K = \frac{1-\alpha}{\alpha}\frac{w}{r} L.$$

代入生产约束可得条件要素需求:

$$L(w, r, q) = q \left( \frac{\alpha}{1-\alpha}\frac{r}{w} \right)^{1-\alpha},$$

$$K(w, r, q) = q \left( \frac{1-\alpha}{\alpha}\frac{w}{r} \right)^{\alpha}.$$

成本函数为

$$\boxed{c(w, r, q) = \frac{q\, w^{\alpha} r^{1-\alpha}}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}.}$$

由 Shephard 引理,

$$\frac{\partial c}{\partial w} = L(w, r, q), \qquad \frac{\partial c}{\partial r} = K(w, r, q).$$

# 12 利润函数与 Hotelling's Lemma

## 12.1 利润最大化问题

给定产出价格 $p$ 和投入价格 $\boldsymbol{w}$, 企业利润最大化:

$$\pi(p,\boldsymbol{w}) = \max_{q,\boldsymbol{x}} \; pq - \boldsymbol{w}\cdot\boldsymbol{x} \quad \text{s.t.} \quad q \le f(\boldsymbol{x}).$$

如果已经有成本函数, 则可以写成

$$\boxed{\pi(p,\boldsymbol{w}) = \max_{q\ge 0} \; pq - c(\boldsymbol{w},q).}$$

FOC 为

$$p = c_q(\boldsymbol{w},q),$$

即价格等于边际成本。

**定义 / 定理**

**Hotelling's lemma.** 若利润函数 $\pi(p,\boldsymbol{w})$ 可微, 则

$$\boxed{\frac{\partial \pi(p,\boldsymbol{w})}{\partial p} = q(p,\boldsymbol{w}),}$$

$$\boxed{\frac{\partial \pi(p,\boldsymbol{w})}{\partial w_i} = -x_i(p,\boldsymbol{w}).}$$

利润函数对产出价格的导数是供给量, 对投入价格的导数是负的投入需求。

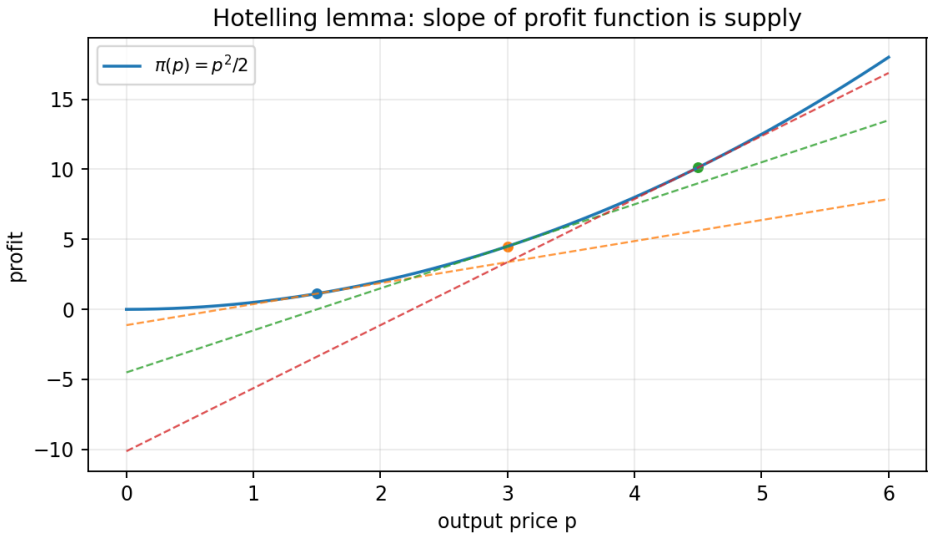



图 6: 利润函数作为价格的函数, 其斜率等于供给量。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 20
# 不动点定理与均衡存在

前面几章解决的是“给定一个优化问题, 最优选择是什么”。经济学的下一步常常是: 每个人都在最优化, 市场也要清算, 这些个体选择能否拼成一个整体均衡? 本章讲 fixed point theorem, fixed point iteration, equilibrium existence。你会看到 Brouwer, Kakutani, contraction mapping, Tarski 这些名字并不是抽象数学装饰, 而是一般均衡、Nash 均衡、动态规划、搜索匹配、金融市场清算和学习动态背后的存在性语言。

math4econ personalized notes
2026

# 目录

**Chapter 20 总览** **2**

**1 为什么经济学需要不动点** **2**
1.1 市场均衡 . . . 2
1.2 博弈均衡 . . . 3
1.3 动态规划 . . . 3

**2 从方程根到不动点** **3**
2.1 根和不动点的互相转化 . . . 3
2.2 一维图像 . . . 4

**3 Brouwer fixed point theorem** **5**
3.1 定理表述 . . . 5
3.2 为什么凸性和紧性重要 . . . 5
3.3 Brouwer 的经济解读 . . . 6

**4 Contraction mapping theorem** **6**
4.1 压缩映射 . . . 6
4.2 Banach fixed point theorem . . . 7
4.3 为什么 contraction 会唯一 . . . 7
4.4 误差界 . . . 8

**5 动态规划中的 contraction** **8**
5.1 Blackwell 条件的直觉 . . . 9

**6 Kakutani fixed point theorem** **10**
6.1 Correspondence . . . 10
6.2 定理表述 . . . 11
6.3 Upper hemicontinuity 的直觉 . . . 11

**7 Nash equilibrium as fixed point** **12**
7.1 有限玩家博弈 . . . 12
7.2 Nash 存在性定理的结构 . . . 13

**8 一般均衡存在性入口** **13**
8.1 市场清算和价格 simplex . . . 14
8.2 Walras law . . . 14
8.3 固定点证明的一般想法 . . . 15

**9 Tarski fixed point theorem 的直觉** **15**
9.1 单调映射 . . . . . 15
9.2 战略互补的例子 . . . . . 15

**10 算法：如何找不动点** **16**
10.1 固定点迭代 . . . . . 16
10.2 Damping . . . . . 16
10.3 经济解释 . . . . . 16

**11 Worked Examples** **17**

**12 常见误区与纠正** **19**

**13 本章小结** **19**

**14 练习题** **20**
A 组：基础计算 . . . . . 20
B 组：概念判断 . . . . . 20
C 组：混合题 . . . . . 20
D 组：经济金融应用 . . . . . 21
E 组：挑战 / 证明题 . . . . . 21

**15 练习题参考答案** **22**
A 组答案 . . . . . 22
B 组答案 . . . . . 22
C 组答案 . . . . . 23
D 组答案 . . . . . 24
E 组答案 . . . . . 24

**16 后续学习连接** **25**

# Chapter 20 总览

| | |
|---|---|
| **本章主题** | fixed point theorems, equilibrium existence, contraction mapping, correspondence, best response, market clearing。 |
| **核心问题** | 一个经济系统里, 个体的最优反应、价格调整、市场清算或动态规划算子是否存在自洽解? 数学上就是寻找 $x^*$ 使得 $x^* = F(x^*)$ 或 $x^* \in F(x^*)$。 |
| **数学工具** | Brouwer fixed point theorem, Kakutani fixed point theorem, contraction mapping theorem, Banach fixed point theorem, upper hemicontinuity, compactness, convexity, correspondence, Tarski monotone fixed point theorem 的直觉。 |
| **经济应用** | 一般均衡存在性、Nash 均衡存在性、动态规划 Bellman operator、市场清算价格、最优反应映射、价格调整算法、理性预期均衡。 |
| **后续连接** | Chapter 21 变分法与 Euler equation, Chapter 22 动态规划, Chapter 23 最优控制, Chapter 24 随机过程与马尔可夫决策, 以及高级微观中的一般均衡和博弈论。 |

**直觉盒子**

不动点 fixed point 的意思很简单: 一个点被函数或映射送回自己。

$$x^* = F(x^*).$$

经济学里很多“均衡”就是不动点: 给定价格, 个体最优选择产生总需求和总供给; 如果这些选择反过来支持原来的价格, 这个价格就是均衡价格。给定别人策略, 我的最优反应; 如果所有人的策略都同时是彼此的最优反应, 这个策略组合就是 Nash equilibrium。

## 1 为什么经济学需要不动点

经济学中“均衡”通常不是直接求一个最大值或最小值, 而是求一组互相兼容的对象。举几个典型例子。

### 1.1 市场均衡

价格 $p$ 影响消费者需求和企业供给, 因而决定 excess demand:

$$z(p) = D(p) - S(p).$$

市场清算要求

$$z(p^*) = 0.$$

如果我们把价格调整写成

$$p_{t+1} = F(p_t),$$

那么均衡价格就是

$$p^* = F(p^*).$$

### 1.2 博弈均衡

在一个博弈里, 每个玩家的最优策略依赖于其他玩家策略。令 $BR_i(s_{-i})$ 表示玩家 $i$ 的 best response。Nash 均衡满足

$$s_i^* \in BR_i(s_{-i}^*)$$

对所有 $i$ 成立。把所有玩家的 best response 合起来, 得到一个 correspondence:

$$BR(s) = BR_1(s_{-1}) \times \cdots \times BR_n(s_{-n}).$$

Nash 均衡就是

$$s^* \in BR(s^*).$$

### 1.3 动态规划

Bellman equation 写成

$$V = TV.$$

这里 $T$ 是 Bellman operator。价值函数 $V^*$ 是算子 $T$ 的不动点:

$$V^* = TV^*.$$

如果 $T$ 是 contraction, Banach fixed point theorem 不仅保证存在唯一 $V^*$, 还告诉我们 value iteration 会收敛。

**常见误区**

在经济学中, “求均衡”不一定等于“解一阶条件”。一阶条件通常解决个体优化; 均衡存在性还要证明这些个体最优选择能够互相匹配。这个匹配问题通常需要 fixed point theorem。

## 2 从方程根到不动点

### 2.1 根和不动点的互相转化

很多初学者觉得 $z(p) = 0$ 和 $p = F(p)$ 是两类问题。其实它们可以互相转化。

若要解

$$g(x) = 0,$$

可以构造

$$F(x) = x - \alpha g(x),$$

则

$$g(x^*) = 0 \quad \Longleftrightarrow \quad x^* = F(x^*).$$

反过来, 若要解

$$x = F(x),$$

可以定义

$$g(x) = F(x) - x,$$

则不动点就是 $g(x) = 0$ 的根。

**直觉盒子**

root finding 和 fixed point finding 本质上是同一类问题的两种语言。经济学更偏爱 fixed point language, 因为均衡通常表达为“个体最优反应与总体状态互相一致”。

## 2.2 一维图像

一维不动点最容易看。画出 $y = F(x)$ 和 $y = x$。交点就是不动点。

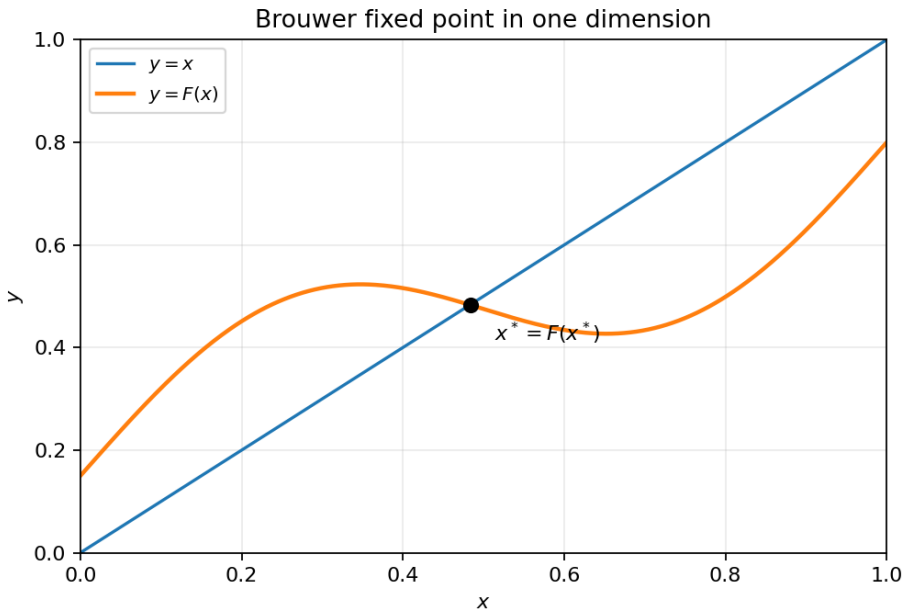



图 1: 一维不动点: $F(x)$ 与 $45°$ 线的交点。

如果 $F:[0,1]\to[0,1]$ 连续, 那么至少有一个不动点。直觉是: 定义

$$h(x)=F(x)-x.$$

因为 $F(0)\geq 0$, 所以

$$h(0)=F(0)-0\geq 0.$$

因为 $F(1)\leq 1$, 所以

$$h(1)=F(1)-1\leq 0.$$

连续函数从非负到非正, 中间必然穿过 0。于是存在 $x^*$ 使 $h(x^*)=0$, 即 $F(x^*)=x^*$。

# 3 Brouwer fixed point theorem

## 3.1 定理表述

**定义 / 定理**

**Brouwer fixed point theorem, 实用版.** 设 $X\subset\mathbb{R}^n$ 是非空、紧、凸集合。如果

$$F:X\to X$$

连续, 那么存在 $x^*\in X$ 使得

$$x^*=F(x^*).$$

这句话中每个条件都重要。

- 非空: 至少要有可选点。
- 紧 compact: 集合闭且有界, 防止点逃到无穷远或边界外。
- 凸 convex: 两点之间的线段还在集合内, 排除奇怪洞洞。
- 连续: 小变化不能导致跳跃式大变化。
- 自映射 $F:X\to X$: 映射后仍留在同一个集合里。

## 3.2 为什么凸性和紧性重要

经济学中的策略集合、消费集合、价格 simplex 经常被构造为 compact convex set。例如价格向量可以归一化为

$$\Delta=\left\{p\in\mathbb{R}^L_+:\sum_{\ell=1}^{L}p_\ell=1\right\}.$$

这个 simplex 是紧凸集合。只要我们能构造一个连续价格调整函数

$$F:\Delta\to\Delta,$$

Brouwer 就给出不动点。

**Advanced Core**

Brouwer 定理本身是拓扑学结论。经济学使用它时，通常不证明 Brouwer，而是把经济问题转化为满足 Brouwer 条件的自映射。真正困难的部分往往是：如何把“市场清算”或“最优反应”写成连续自映射。

### 3.3 Brouwer 的经济解读

Brouwer 的经济语言是：如果一个经济系统从一个可行状态出发，经过一套连续调整规则后仍然回到可行状态集合，那么至少存在一个状态在调整后不变。这就是 equilibrium existence 的最朴素版本。

**Worked Example**

**例 1：单区间上的不动点**

设

$$F(x) = 0.2 + 0.6x, \qquad F : [0,1] \to [0,1].$$

求不动点。

解：不动点满足

$$x = 0.2 + 0.6x.$$

所以

$$0.4x = 0.2, \qquad x^* = 0.5.$$

因为 $F$ 连续并把 $[0,1]$ 映射到自身，Brouwer 保证至少存在不动点。这里还可以直接算出唯一不动点。

## 4 Contraction mapping theorem

Brouwer 保证存在，但不一定保证唯一，也不一定告诉你如何找到不动点。Contraction mapping theorem 更强：它给出存在、唯一和迭代收敛。

### 4.1 压缩映射

**定义 / 定理**

**Contraction.** 在一个度量空间 $(X,d)$ 上，映射 $T : X \to X$ 称为 contraction，如果存在常数 $\beta \in [0,1)$ 使得对所有 $x, y \in X$，

$$d(Tx, Ty) \le \beta d(x,y).$$

意思是：映射 $T$ 会把任意两点之间的距离至少按比例 $\beta$ 缩短。只要不断迭代，点之间的距离会越来越小。

## 4.2 Banach fixed point theorem

**定义 / 定理**

**Banach fixed point theorem.** 若 $(X, d)$ 是 complete metric space，且 $T : X \to X$ 是 contraction，则：

1. 存在唯一不动点 $x^* \in X$，满足 $Tx^* = x^*$；
2. 对任意初始点 $x_0 \in X$，迭代

$$x_{n+1} = Tx_n$$

都收敛到 $x^*$。

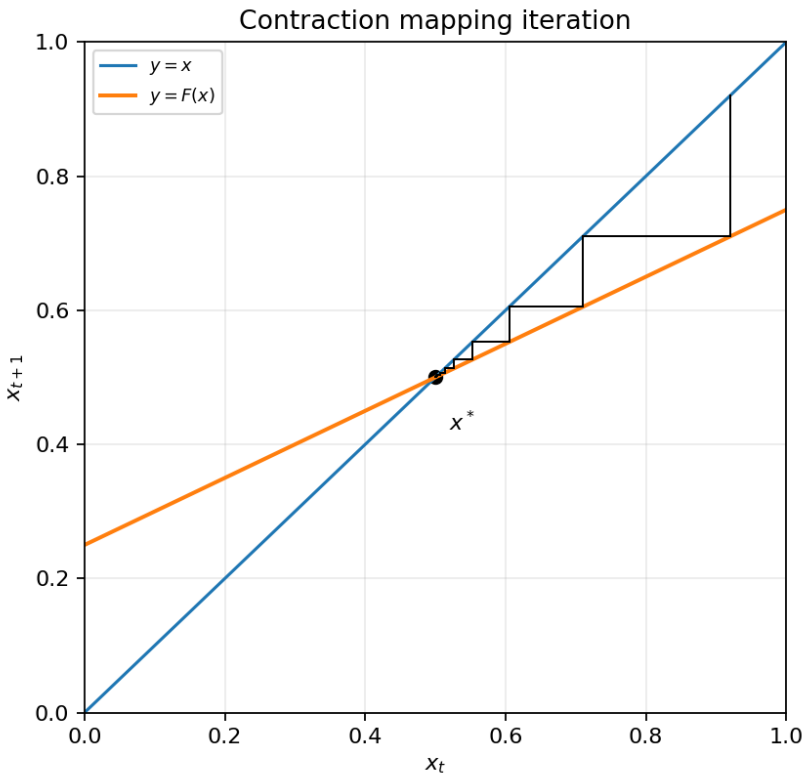



图 2：压缩映射迭代：不断把点送入 $F$，最终收敛到不动点。

## 4.3 为什么 contraction 会唯一

假设 $x^*$ 和 $y^*$ 都是不动点，即

$$Tx^* = x^*, \qquad Ty^* = y^*.$$

由 contraction,

$$d(x^*, y^*) = d(Tx^*, Ty^*) \le \beta d(x^*, y^*).$$

因为 $0 \le \beta < 1$, 只有可能

$$d(x^*, y^*) = 0.$$

所以

$$x^* = y^*.$$

### 4.4 误差界

设 $x^*$ 是不动点, 且 $x_{n+1} = Tx_n$。则

$$d(x_n, x^*) = d(Tx_{n-1}, Tx^*) \le \beta d(x_{n-1}, x^*).$$

反复递推:

$$\boxed{d(x_n, x^*) \le \beta^n d(x_0, x^*).}$$

这不仅说明收敛, 还说明收敛速度。

**方法模板**

**证明一个算子是 contraction 的常见步骤**

1. 选定空间 $X$ 和距离 $d$ 或范数 $\|\cdot\|$。
2. 证明 $T$ 是自映射: $T : X \to X$。
3. 对任意 $x, y \in X$, 估计 $\|Tx - Ty\|$。
4. 找到 $\beta < 1$ 使得 $\|Tx - Ty\| \le \beta \|x - y\|$。
5. 应用 Banach theorem: 存在唯一不动点, 迭代收敛。

## 5 动态规划中的 contraction

动态规划中最重要的不动点是 Bellman equation:

$$V = TV.$$

Bellman operator 通常形如

$$(TV)(x) = \max_{a \in A(x)} \{u(x, a) + \beta V(f(x, a))\},$$

其中 $0 < \beta < 1$ 是折现因子。

## 5.1 Blackwell 条件的直觉

在很多模型中, Bellman operator 是 contraction。直觉来自折现因子。若两个候选价值函数 $V, W$ 相差最多 $\|V-W\|_\infty$, 则未来价值项的差异在 Bellman operator 中被乘以 $\beta$:

$$|TV(x) - TW(x)| \le \beta \|V - W\|_\infty .$$

因此

$$\|TV - TW\|_\infty \le \beta \|V - W\|_\infty .$$

因为 $\beta < 1$, $T$ 是 contraction。

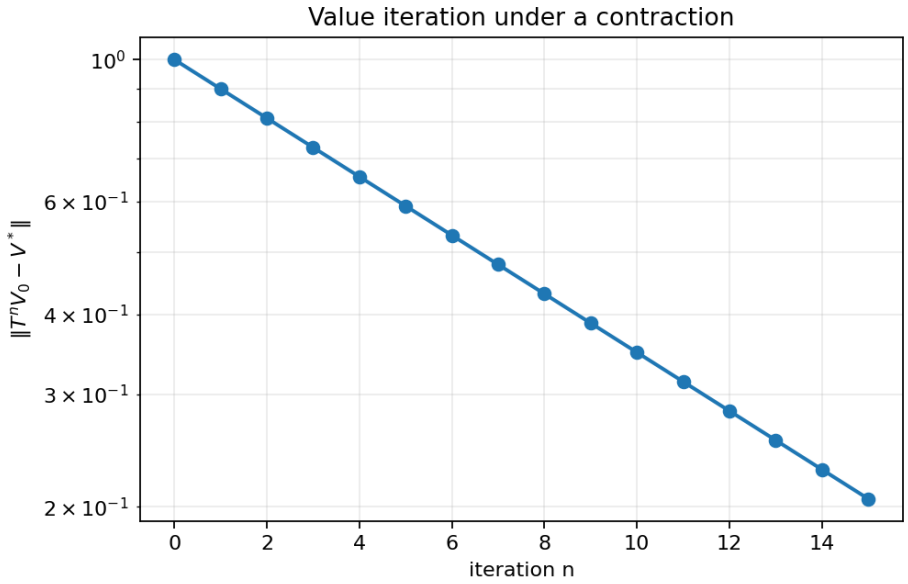



图 3: Bellman value iteration 在 contraction 下以几何速度收敛。

**直觉盒子**

折现因子 $\beta$ 不只是偏好参数, 它也是数学上的收缩系数。未来越不重要, Bellman operator 越强烈地压缩差异, value iteration 越快收敛。

**Worked Example**

**例 2: 一个最简单的 Bellman operator**

设状态集合有限, Bellman operator 为

$$(TV)(x) = r(x) + \beta \max_{y \in \Gamma(x)} V(y), \qquad 0 < \beta < 1.$$

证明它是 contraction。

解: 对任意 $V, W$,

$$|TV(x) - TW(x)| = \beta \left| \max_{y \in \Gamma(x)} V(y) - \max_{y \in \Gamma(x)} W(y) \right|.$$

利用最大值的 Lipschitz 性质:

$$\left| \max_y V(y) - \max_y W(y) \right| \le \max_y |V(y) - W(y)| \le \|V - W\|_\infty .$$

所以

$$|TV(x) - TW(x)| \le \beta \|V - W\|_\infty .$$

对所有 $x$ 取上确界:

$$\|TV - TW\|_\infty \le \beta \|V - W\|_\infty .$$

因此 $T$ 是 contraction。

# 6 Kakutani fixed point theorem

Brouwer 处理函数 $F : X \to X$。但经济学中很多最优反应不是单值函数, 而是集合值 correspondence。例如某个价格下消费者有多个最优消费束, 某个策略下玩家有多个 best response。

## 6.1 Correspondence

**定义 / 定理**

**Correspondence.** 若对每个 $x \in X$, $F(x)$ 不是一个点, 而是 $Y$ 的一个子集, 则称

$$F : X \rightrightarrows Y$$

为 correspondence 或 set-valued map。

不动点条件相应变成

$$x^* \in F(x^*).$$

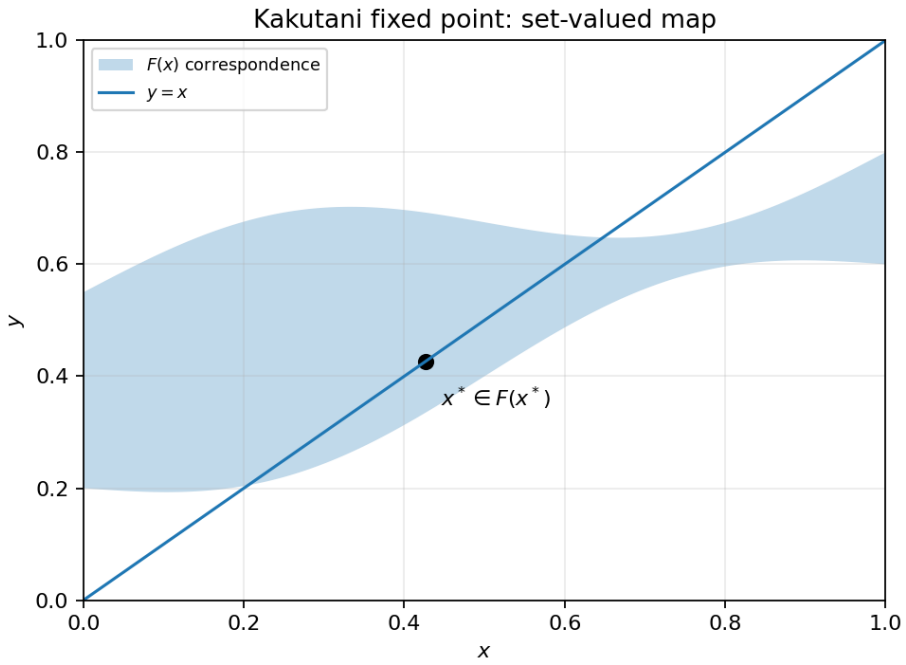



图 4: Kakutani 不动点: 图中阴影区域代表集合值映射 $F(x)$, 只要 $x^*$ 落在自己的对应集合中即可。

## 6.2 定理表述

**定义 / 定理**

**Kakutani fixed point theorem, 实用版.** 设 $X \subset \mathbb{R}^n$ 非空、紧、凸。若 correspondence

$$F : X \rightrightarrows X$$

满足:

1. 对每个 $x \in X$, $F(x)$ 非空、凸、紧;
2. $F$ upper hemicontinuous;

则存在 $x^* \in X$ 使得

$$x^* \in F(x^*).$$

## 6.3 Upper hemicontinuity 的直觉

Upper hemicontinuity, 简称 u.h.c., 是 correspondence 的连续性概念之一。直觉是: 当 $x_n \to x$ 且 $y_n \in F(x_n)$, 如果 $y_n \to y$, 那么 $y \in F(x)$。也就是说, 最优集合不会在极限处突然冒出不相干的点。

经济学中, 最大值定理 maximum theorem 经常用来证明最优反应 correspondence 是非空、紧值、u.h.c. 的。若目标函数 concave, 可行集 convex, 则最优集合还会是 convex valued。这样就能套 Kakutani。

**Advanced Core**

Kakutani 是 Nash equilibrium existence 和 Arrow-Debreu general equilibrium existence 的核心工具之一。它比 Brouwer 更适合经济学, 因为经济选择常常不是唯一的。

# 7 Nash equilibrium as fixed point

## 7.1 有限玩家博弈

考虑 $n$ 个玩家。玩家 $i$ 的策略集合为 $S_i$, 总策略空间为

$$S = S_1 \times \cdots \times S_n.$$

给定其他玩家策略 $s_{-i}$, 玩家 $i$ 的最优反应集合是

$$BR_i(s_{-i}) = \arg\max_{s_i \in S_i} u_i(s_i, s_{-i}).$$

定义整体 best response correspondence:

$$BR(s) = BR_1(s_{-1}) \times \cdots \times BR_n(s_{-n}).$$

Nash equilibrium 满足

$$\boxed{s^* \in BR(s^*).}$$

因此 Nash equilibrium 就是 best response correspondence 的不动点。

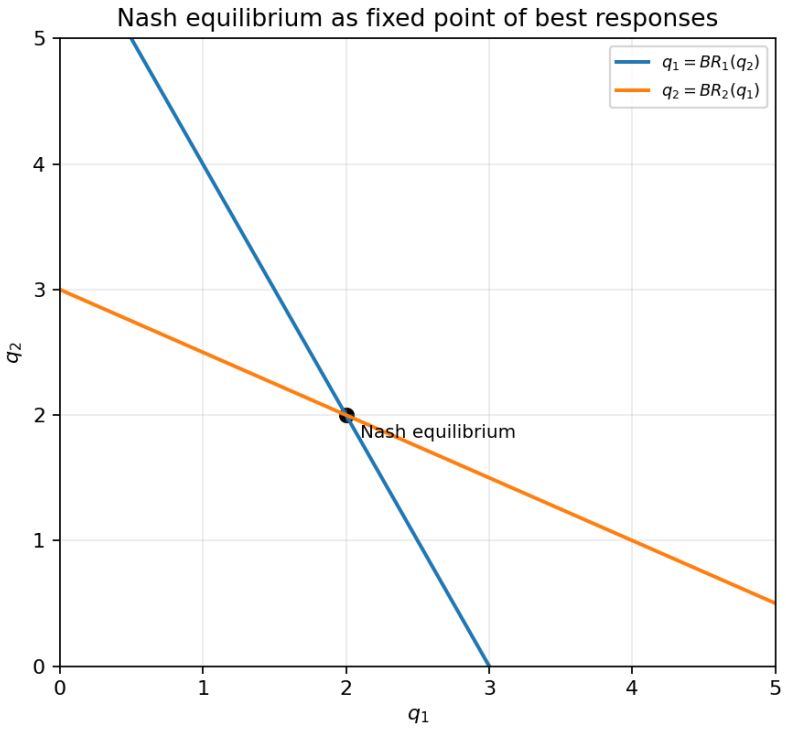



图 5: Nash equilibrium 是两个玩家 best response 的交点, 也就是整体 best response correspondence 的不动点。

### 7.2 Nash 存在性定理的结构

一个常见版本是: 若每个 $S_i$ 非空、紧、凸, 每个 $u_i$ 连续, 且对自己的策略 $s_i$ quasi-concave, 则 Nash equilibrium 存在。

证明结构如下:

1. $S$ 是非空、紧、凸集合。
2. 最大值定理保证 $BR_i$ 非空紧值且 u.h.c.。
3. quasi-concavity 保证 $BR_i$ convex valued。
4. 因此 $BR$ 满足 Kakutani 条件。
5. 存在 $s^* \in BR(s^*)$, 即 Nash equilibrium。

**常见误区**

Nash 存在性定理证明的是“至少存在一个均衡”, 不保证唯一, 也不保证均衡容易算, 更不保证均衡好、有效或公平。

## 8 一般均衡存在性入口

## 8.1 市场清算和价格 simplex

在一般均衡中, 有 $L$ 种商品, 价格向量 $p \in \mathbb{R}^L_+$。由于价格只在相对意义上重要, 可以归一化为 simplex:

$$\Delta = \left\{ p \in \mathbb{R}^L_+ : \sum_{\ell=1}^{L} p_\ell = 1 \right\}.$$

总 excess demand 为

$$z(p) = \sum_i x_i(p, w_i) - \sum_j y_j(p) - \omega.$$

均衡价格 $p^*$ 满足市场清算:

$$z(p^*) \le 0,$$

并且若 $p^*_\ell > 0$, 则第 $\ell$ 个市场清算为等号。常见简化写法是

$$z(p^*) = 0.$$

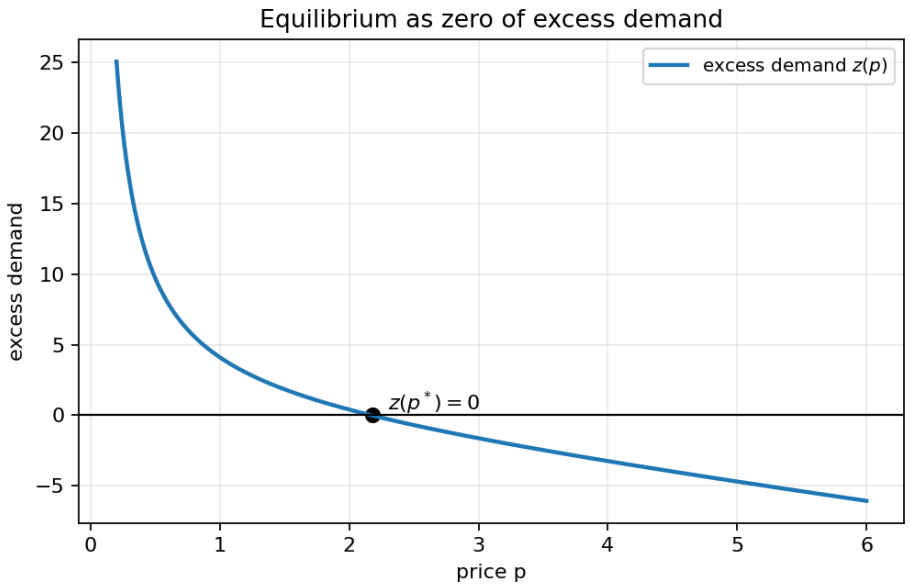



图 6: 一维价格下, 市场均衡可写成 excess demand 的零点。多维情况下, 需要固定点定理。

## 8.2 Walras law

Walras law 表示在每个价格向量下, 总预算约束导致

$$p \cdot z(p) = 0.$$

这意味着总 excess demand 的价值为零。若所有商品价格为正, 且除一个市场外其他市场都清算, 则最后一个市场也自动清算。

### 8.3 固定点证明的一般想法

一般均衡存在性证明通常不直接解 $z(p)=0$, 而是构造一个价格调整映射 $F:\Delta\to\Delta$。例如让 excess demand 为正的商品价格上升, excess demand 为负的商品价格下降, 然后再归一化回 simplex。

Brouwer 保证存在 $p^*=F(p^*)$。再利用 $F$ 的构造和 Walras law, 证明这个 fixed point 实际上是 market clearing price。

**Advanced Core**

Arrow-Debreu 一般均衡存在性证明在形式上更复杂, 因为消费者需求、企业供给可能是 correspondence, 价格也在 simplex 上, 还要处理边界价格和非唯一选择。但核心逻辑仍然是: 把均衡问题变成一个满足 Brouwer 或 Kakutani 条件的不动点问题。

## 9 Tarski fixed point theorem 的直觉

除了 Brouwer 和 Kakutani, 经济学还常见 Tarski fixed point theorem, 尤其在 supermodular games, monotone comparative statics, network games, dynamic complementarity 中。

### 9.1 单调映射

若集合有偏序 $\le$, 且映射 $F$ 满足

$$x\le y \quad\Longrightarrow\quad F(x)\le F(y),$$

则称 $F$ 是 monotone 或 order-preserving。

Tarski 的思想是: 在完整格 complete lattice 上, 单调映射存在不动点, 而且不动点集合本身也构成完整格。经济含义是: 若策略之间存在战略互补 strategic complementarity, 那么通常会有最大均衡和最小均衡。

**直觉盒子**

Brouwer/Kakutani 强调连续性、紧性、凸性。Tarski 强调序结构和单调性。它非常适合分析“别人做得越多, 我越想做得多”的经济系统, 例如投资互补、技术采用、银行挤兑、网络外部性。

### 9.2 战略互补的例子

设两个企业的投资选择满足

$$x_1=F_1(x_2),\qquad x_2=F_2(x_1),$$

并且 $F_1, F_2$ 都递增。若企业 2 投资更多会提高企业 1 的投资回报，企业 1 的最优投资也上升。这种情况下可能存在低投资均衡和高投资均衡。

# 10 算法：如何找不动点

存在性定理告诉你“有解”，但模型应用还需要“怎么找”。

## 10.1 固定点迭代

最直接方法是

$$x_{n+1} = F(x_n).$$

若 $F$ 是 contraction，该方法保证收敛。若不是 contraction，可能震荡或发散。

## 10.2 Damping

当直接迭代太激进时，可以使用 damping：

$$x_{n+1} = (1-\lambda)x_n + \lambda F(x_n), \qquad 0 < \lambda < 1.$$

这相当于只向 $F(x_n)$ 走一小步。

## 10.3 经济解释

价格调整、学习动态、best response dynamics 都可以看成固定点迭代。但请注意：存在均衡不代表某个具体调整过程一定收敛到均衡。存在性和稳定性是不同问题。

**方法模板**

**从经济问题到固定点定理的标准流程**

1. 明确均衡对象：价格、策略、价值函数、政策函数或分配。
2. 定义可行集合 $X$。
3. 证明 $X$ 非空、紧、凸，或具有完整格结构。
4. 构造映射或 correspondence $F$。
5. 证明 $F : X \to X$，并验证连续性 / u.h.c. / convex valued / contraction / monotonicity。
6. 套用 Brouwer, Kakutani, Banach 或 Tarski。
7. 解释 fixed point 与经济均衡条件等价。

# 11 Worked Examples

**Worked Example**

**例 3: 线性映射的不动点和 contraction**

设

$$F(x) = ax + b$$

在 $\mathbb{R}$ 上。若 $|a| < 1$, 求不动点并说明迭代收敛。

解: 不动点满足

$$x = ax + b.$$

所以

$$x^* = \frac{b}{1-a}.$$

对任意 $x, y$,

$$|F(x) - F(y)| = |a||x - y|.$$

若 $|a| < 1$, $F$ 是 contraction。因此从任意 $x_0$ 出发,

$$x_{n+1} = F(x_n)$$

都收敛到 $x^*$。这和 Chapter 11 的一阶差分方程稳定性完全一致。

**Worked Example**

**例 4: Cournot duopoly 的 Nash 均衡是不动点**

两个企业产量为 $q_1, q_2$, 反需求为

$$P = A - B(q_1 + q_2),$$

边际成本为 $c$, 其中 $A > c$, $B > 0$。企业 1 的利润为

$$\pi_1(q_1, q_2) = [A - B(q_1 + q_2) - c]q_1.$$

对 $q_1$ 求一阶条件:

$$A - c - 2Bq_1 - Bq_2 = 0.$$

所以 best response 为

$$BR_1(q_2) = \frac{A-c}{2B} - \frac{1}{2}q_2.$$

同理

$$BR_2(q_1) = \frac{A-c}{2B} - \frac{1}{2}q_1.$$

Nash equilibrium 满足

$$q_1 = BR_1(q_2), \qquad q_2 = BR_2(q_1).$$

对称均衡 $q_1 = q_2 = q$ 给出

$$q = \frac{A-c}{2B} - \frac{1}{2}q,$$

所以

$$\frac{3}{2}q = \frac{A-c}{2B}, \qquad \boxed{q^* = \frac{A-c}{3B}.}$$

这就是 best response 映射的不动点。

**Worked Example**

**例 5: 价格调整映射和市场均衡**

设 excess demand 为

$$z(p) = 10 - 2p.$$

市场均衡满足 $z(p^*) = 0$, 所以 $p^* = 5$。现在构造价格调整映射

$$F(p) = p + \alpha z(p) = p + \alpha(10 - 2p).$$

不动点满足

$$p = F(p) = p + \alpha(10 - 2p).$$

若 $\alpha \neq 0$, 得

$$10 - 2p = 0,$$

所以 $p = 5$。因此 fixed point 等价于 market clearing。
若研究迭代稳定性, 有

$$F'(p) = 1 - 2\alpha.$$

局部收敛需要

$$|1 - 2\alpha| < 1,$$

即

$$0 < \alpha < 1.$$

**Worked Example**

**例 6: 一个 correspondence 的不动点**

设 $X=[0,1]$, correspondence 为

$$F(x)=\begin{cases}[0,0.6], & 0\le x<0.4,\\ [0.3,0.8], & 0.4\le x\le 0.7,\\ [0.5,1], & 0.7<x\le 1.\end{cases}$$

找一个不动点。
解: 只需要找 $x\in F(x)$。例如 $x=0.5$ 属于区间 $[0.4,0.7]$, 此时

$$F(0.5)=[0.3,0.8].$$

因为

$$0.5\in[0.3,0.8],$$

所以 $0.5$ 是 fixed point。这个例子说明 correspondence 的不动点条件是 $x\in F(x)$, 不是 $x=F(x)$。

## 12 常见误区与纠正

**常见误区**

**误区 1: 有不动点就一定唯一。** Brouwer 和 Kakutani 通常只保证存在, 不保证唯一。唯一性需要额外条件, 例如 contraction。

**误区 2: 有均衡就一定能通过简单迭代找到。**存在性和稳定性不同。一个均衡可能存在但不稳定; best response dynamics 也可能循环。

**误区 3: 把 correspondence 当函数。**若最优反应不唯一, 必须写 $x\in F(x)$, 不能写 $x=F(x)$。

**误区 4: 忽略 compactness。**没有紧性时, 序列可能逃向无穷远, 不动点可能不存在。例如 $F(x)=x+1$ 在 $\mathbb{R}$ 上没有不动点。

**误区 5: 忽略凸性。** Kakutani 需要 convex valued。如果最优集合不凸, 可能需要混合策略、随机化或更高级工具。

## 13 本章小结

| 概念 | 数学形式 | 经济含义 |
|---|---|---|
| Fixed point | $x^*=F(x^*)$ | 自洽状态, 调整后不变。 |

| | | |
|---|---|---|
| Correspondence fixed point | $x^* \in F(x^*)$ | 最优反应可能不唯一时的均衡。 |
| Brouwer | continuous $F : X \to X$ | 连续自映射保证存在均衡。 |
| Kakutani | u.h.c. convex-valued correspondence | 最优反应集合存在不动点。 |
| Contraction | $d(Tx, Ty) \le \beta d(x, y)$ | 存在唯一不动点, 迭代收敛。 |
| Tarski | monotone map on lattice | 战略互补下有最大和最小均衡。 |
| Nash equilibrium | $s^* \in BR(s^*)$ | 策略互为最优反应。 |
| General equilibrium | market clearing price | 个体最优选择和市场清算同时成立。 |
| Bellman equation | $V = TV$ | 价值函数是 Bellman operator 的不动点。 |

## 14 练习题

### A 组: 基础计算

A1. 求 $F(x) = 0.3x + 2$ 的不动点, 并判断 $F$ 是否为 contraction。

A2. 求 $F(x) = x^2$ 在 $[0, 1]$ 上的不动点。

A3. 设 $F(x) = 1/x$ 定义在 $[1/2, 2]$ 上。求不动点, 并判断能否直接用 Brouwer。

A4. 设 $F(x) = 0.25 + 0.5x$ 在 $[0, 1]$ 上。若 $x_0 = 1$, 写出前 4 次迭代。

A5. 若 $T$ 是 contraction, 收缩系数为 $0.8$, 且 $d(x_0, x^*) = 10$, 给出 $d(x_5, x^*)$ 的上界。

A6. 把方程 $g(x) = x^2 - 3x + 2 = 0$ 写成一个 fixed point problem。

### B 组: 概念判断

B1. Brouwer fixed point theorem 保证唯一性吗?

B2. Contraction mapping theorem 比 Brouwer 强在哪里?

B3. 为什么 Nash equilibrium 可以写成 best response correspondence 的不动点?

B4. 为什么动态规划里的 Bellman equation 是固定点问题?

B5. 存在均衡是否意味着经济调整过程一定收敛到均衡? 解释。

### C 组: 混合题

C1. 证明 $F(x) = ax + b$ 在 $\mathbb{R}$ 上是 contraction 当且仅当 $|a| < 1$。

C2. 设 $F : [0, 1] \to [0, 1]$ 连续。用介值定理证明存在 $x^*$ 使 $F(x^*) = x^*$。

C3. 设价格调整为 $p_{t+1} = p_t + \alpha(6 - 3p_t)$。求均衡价格和局部收敛条件。

C4. 两企业 Cournot 模型中 $P = 12 - Q$, 边际成本 $c = 3$。求对称 Nash 均衡产量。

C5. 设 Bellman operator 满足 $\|TV - TW\|_\infty \le 0.95 \|V - W\|_\infty$。若初始误差为 100, 至多多少次迭代后误差上界小于 1?

### D 组: 经济金融应用

D1. 一般均衡中为什么要把价格归一化到 simplex?

D2. 解释 Walras law 如何帮助从价格不动点推出市场清算。

D3. 给出一个金融市场中 fixed point 的例子, 并说明固定点对象是什么。

D4. 解释战略互补为什么可能导致多个均衡。

D5. 动态规划中, 折现因子 $\beta$ 越接近 1, value iteration 为什么通常越慢?

### E 组: 挑战 / 证明题

E1. 证明 contraction 不动点唯一。

E2. 证明 contraction iteration 的误差界 $d(x_n, x^*) \le \beta^n d(x_0, x^*)$。

E3. 用 Kakutani 的条件解释 Nash 存在性证明的逻辑链条。

E4. 设 $T$ 是 Bellman operator, 证明在有界函数空间的 sup norm 下, 折现因子 $\beta < 1$ 会带来 contraction 的核心不等式。

E5. 构造一个连续函数 $F : (0, 1) \to (0, 1)$ 没有不动点, 并解释为什么这不违反 Brouwer。

# 15 练习题参考答案

## A 组答案

A1. 不动点满足

$$x = 0.3x + 2.$$

所以

$$0.7x = 2, \qquad x^* = \frac{20}{7}.$$

又

$$|F(x) - F(y)| = 0.3|x - y|,$$

因此是 contraction, 收缩系数 $0.3$。

A2. $x = x^2$ 给出

$$x(x-1) = 0.$$

所以不动点是 $0$ 和 $1$。

A3. 不动点满足

$$x = \frac{1}{x},$$

即 $x^2 = 1$。在 $[1/2, 2]$ 上不动点为 $x = 1$。函数 $F(x) = 1/x$ 连续, 且把 $[1/2, 2]$ 映射到 $[1/2, 2]$, 所以可以用一维 Brouwer。

A4. 迭代为

$$x_1 = 0.25 + 0.5(1) = 0.75,$$

$$x_2 = 0.25 + 0.5(0.75) = 0.625,$$

$$x_3 = 0.25 + 0.5(0.625) = 0.5625,$$

$$x_4 = 0.25 + 0.5(0.5625) = 0.53125.$$

不动点为 $0.5$。

A5. 误差界为

$$d(x_5, x^*) \le 0.8^5 \cdot 10 = 3.2768.$$

A6. 例如构造

$$F(x) = x - \alpha(x^2 - 3x + 2).$$

则 $F(x) = x$ 等价于 $x^2 - 3x + 2 = 0$。也可以取 $\alpha = 1$。

## B 组答案

B1. 不保证。Brouwer 只保证至少存在一个不动点。例如 $F(x) = x$ 在 $[0, 1]$ 上每个点都是不动点。

B2. Contraction theorem 不仅保证存在, 还保证唯一, 并且告诉我们从任意初始点做迭代都会收敛到唯一不动点。Brouwer 通常不提供算法收敛性。

B3. Nash equilibrium 要求每个玩家的策略都是对其他玩家策略的最优反应。因此整体策略向量 $s^*$ 必须属于整体 best response correspondence $BR(s^*)$, 即 $s^* \in BR(s^*)$。

B4. Bellman equation 是 $V = TV$, 其中 $T$ 把一个候选价值函数映射为一个更新后的价值函数。真正的价值函数在这个更新操作下不变, 所以是 $T$ 的 fixed point。

B5. 不意味着。存在性只是说明有均衡点; 收敛性取决于具体动态过程。价格调整或 best response dynamics 可能震荡、循环或发散。

## C 组答案

C1. 对任意 $x, y$,

$$|F(x) - F(y)| = |ax + b - ay - b| = |a||x - y|.$$

若 $|a| < 1$, 取 $\beta = |a|$, 是 contraction。若 $|a| \geq 1$, 不存在 $\beta < 1$ 使不等式对所有 $x, y$ 成立, 因而不是 contraction。

C2. 定义

$$h(x) = F(x) - x.$$

因为 $F(0) \in [0, 1]$, 所以 $h(0) = F(0) \geq 0$。因为 $F(1) \in [0, 1]$, 所以 $h(1) = F(1) - 1 \leq 0$。若某端点已经为 0, 端点就是不动点; 否则由介值定理, 存在 $x^* \in (0, 1)$ 使 $h(x^*) = 0$, 即 $F(x^*) = x^*$。

C3. 均衡满足

$$p = p + \alpha(6 - 3p).$$

若 $\alpha \neq 0$, 则 $6 - 3p = 0$, 得

$$p^* = 2.$$

映射为

$$F(p) = (1 - 3\alpha)p + 6\alpha.$$

局部收敛条件为

$$|1 - 3\alpha| < 1,$$

即

$$0 < \alpha < \frac{2}{3}.$$

C4. Cournot 中 $A = 12$, $c = 3$, $B = 1$。对称均衡产量

$$q^* = \frac{A - c}{3B} = \frac{9}{3} = 3.$$

所以两企业均衡产量均为 3。

C5. 需要

$$100(0.95)^n < 1.$$

等价于

$$(0.95)^n < 0.01.$$

取对数:

$$n\log(0.95) < \log(0.01).$$

因为 $\log(0.95) < 0$, 得

$$n > \frac{\log(0.01)}{\log(0.95)} \approx 89.78.$$

因此至少 90 次迭代。

## D 组答案

D1. 因为一般均衡中只有相对价格重要, 把所有价格同时乘以正数不会改变预算集和需求供给。因此可以用归一化 $\sum p_\ell = 1$ 把价格放入 compact convex simplex, 以便应用 fixed point theorem。

D2. Walras law 给出 $p \cdot z(p) = 0$。若构造出来的不动点价格使得所有正价格商品 excess demand 非正, 那么总价值为零迫使这些 excess demand 在价值上不能严格为负; 结合非浪费或 free disposal 条件, 可以推出市场清算。

D3. 例子: 资产定价中价格函数满足

$$p = \frac{d + E[p']}{R}.$$

均衡价格是定价算子的不动点。若价格影响需求和风险溢价, 则 fixed point 对象可能是整个价格向量或收益率函数。

D4. 战略互补表示一个人的行动越高, 其他人的最优行动也越高。这会使系统可能同时支持低行动自洽状态和高行动自洽状态, 例如低采用率技术均衡和高采用率技术均衡。

D5. Bellman operator 的 contraction modulus 通常是 $\beta$。当 $\beta$ 越接近 1, 差异每次迭代被缩小的比例越小, 所以误差下降越慢。

## E 组答案

E1. 假设 $x^*, y^*$ 都是不动点。则

$$d(x^*, y^*) = d(Tx^*, Ty^*) \le \beta d(x^*, y^*).$$

移项得

$$(1-\beta)d(x^*, y^*) \le 0.$$

因为 $1-\beta > 0$ 且距离非负, 只能有 $d(x^*, y^*) = 0$, 所以 $x^* = y^*$。

E2. 因为 $x_n = T^n x_0$, 且 $x^* = Tx^*$,

$$d(x_n, x^*) = d(Tx_{n-1}, Tx^*) \le \beta d(x_{n-1}, x^*).$$

反复应用得到

$$d(x_n, x^*) \le \beta^n d(x_0, x^*).$$

E3. Nash 证明逻辑: 策略空间 $S_i$ 非空紧凸, 所以乘积空间 $S$ 非空紧凸。收益函数连续保证最大值存在, 所以 best response 非空紧值且 u.h.c.; 对自己策略 quasi-concave 保证 best response convex valued。于是整体 $BR : S \rightrightarrows S$ 满足 Kakutani 条件, 存在 $s^* \in BR(s^*)$, 即 Nash equilibrium。

E4. 若

$$(TV)(x) = \max_a \{u(x,a) + \beta V(f(x,a))\},$$

则

$$|TV(x) - TW(x)| \le \beta \sup_{x'} |V(x') - W(x')|.$$

因此

$$\|TV - TW\|_\infty \le \beta \|V - W\|_\infty .$$

只要 $\beta < 1$, $T$ 就是 contraction。

E5. 例如

$$F(x) = \frac{x+1}{2}$$

定义在 $(0,1)$ 上。它把 $(0,1)$ 映射到 $(1/2,1)$, 仍在 $(0,1)$ 中。若有不动点, 则

$$x = \frac{x+1}{2}$$

得 $x = 1$, 但 $1 \notin (0,1)$。所以没有不动点。这不违反 Brouwer, 因为 $(0,1)$ 不是 compact, 它不是闭集。

## 16 后续学习连接

本章的 fixed point language 会在后面多次出现。动态规划中, 价值函数是 Bellman operator 的不动点。最优控制中, 稳态和 saddle path 可以看成动态系统的固定结构。随机动态模型中, invariant distribution 是 Markov operator 的不动点。一般均衡和博弈论中, equilibrium 是个体最优反应和市场清算条件的自洽点。

下一章将转向变分法和 Euler equation。那里我们会从“静态最优”进入“路径最优”: 不再只选一个 $x$, 而是选择一整条路径 $x(t)$ 或序列 $\{x_t\}$。不过背后的思想仍然相同: 最优性条件、稳定性和均衡自洽性。

**直觉盒子**

如果说 Chapter 14–19 解决的是“单个优化问题如何求解”，那么 Chapter 20 解决的是“很多优化问题拼在一起是否存在自洽均衡”。这是从数学工具走向高级经济理论的关键一步。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 21
# 动态优化基础与 Euler 方程

前面的章节已经建立了静态优化、对偶性、比较静态、固定点和动态系统的基础。本章开始进入动态优化 dynamic optimization。动态优化的核心问题不是“今天最优选择是什么”，而是“一整条跨期路径怎样最优”。本章重点讲清楚三件事：跨期资源约束如何写，Euler 方程为什么是跨期边际条件，终端条件和横截条件为什么用来排除不合理路径。消费储蓄、资产定价、资本积累、投资调整和 Ramsey 模型都从这里开始。

math4econ personalized notes
2026

# 目录

**Chapter 21 总览** **2**

**1 本章为什么重要** **2**
1.1 动态优化与前面章节的关系 . . . . . 3
1.2 本章要形成的核心直觉 . . . . . 3

**2 两期消费储蓄模型** **3**
2.1 Lagrange 推导 . . . . . 4
2.2 log utility 的显式解 . . . . . 5

**3 多期有限期问题与 Euler 方程** **5**
3.1 问题设定 . . . . . 6
3.2 用扰动法推导 Euler 方程 . . . . . 6
3.3 Lagrange 乘子推导 . . . . . 6

**4 无限期问题与横截条件** **7**
4.1 无限期目标 . . . . . 7
4.2 为什么 Euler 方程还不够 . . . . . 8
4.3 横截条件的直觉 . . . . . 8

**5 Euler 方程如何决定消费增长** **9**
5.1 经济解释 . . . . . 9

**6 资产定价 Euler 方程** **10**
6.1 无风险资产 . . . . . 10
6.2 风险资产 . . . . . 10

**7 资本积累与 Ramsey 模型入口** **11**
7.1 社会规划者问题 . . . . . 11
7.2 Lagrangian 推导 . . . . . 11
7.3 稳态条件 . . . . . 12

**8 现值乘子与当前值乘子** **13**
8.1 现值乘子 . . . . . 13
8.2 当前值乘子 . . . . . 13

**9 企业投资与调整成本** **13**

**10 连续时间版本预告：Euler-Lagrange 与 Hamiltonian** **14**

**11 动态优化的计算流程** **15**

**12 常见错误与纠正** **16**

**13 本章小结表** **16**

**14 练习题** **16**

A 组: 基础计算 . . . . . 17
B 组: 概念判断 . . . . . 17
C 组: 混合题 . . . . . 17
D 组: 经济金融建模题 . . . . . 17
E 组: 挑战 / 证明题 . . . . . 17

**15 练习题参考答案** **19**

A 组答案 . . . . . 19
B 组答案 . . . . . 20
C 组答案 . . . . . 20
D 组答案 . . . . . 21
E 组答案 . . . . . 22

**16 后续学习连接** **23**

# Chapter 21 总览

| | |
|---|---|
| **本章主题** | 动态优化 dynamic optimization、跨期选择 intertemporal choice、Euler equation、终端条件 terminal condition、横截条件 transversality condition。 |
| **核心问题** | 个体、企业或社会规划者如何在多个时期之间分配资源？什么时候应该多消费，什么时候应该多储蓄或投资？怎样用一阶条件刻画整条最优路径？ |
| **数学工具** | 有限期 Lagrange 方法、无限期贴现目标、Euler 方程、现值乘子和当前值乘子、终端条件、横截条件、相图直觉。 |
| **经济应用** | 两期消费储蓄、永久收入假说入口、资产定价 Euler 方程、资本积累、企业投资调整成本、Ramsey 增长模型直觉。 |
| **后续连接** | Chapter 22 将系统讲动态规划与 Bellman 方程；后面会继续进入 Hamiltonian、最优控制、随机动态优化和资产定价。 |

**直觉盒子**

静态优化问：在一个约束下，怎样选 $x$ 最大化 $u(x)$？动态优化问：怎样选一整串变量

$$c_0, c_1, c_2, \dots$$

最大化

$$\sum_{t=0}^{\infty} \beta^t u(c_t),$$

同时满足跨期资源约束。Euler 方程的直觉就是：最优路径上，把一单位资源从今天挪到明天，不能带来一阶收益。否则你还可以重新安排消费或投资，从而不是最优。

## 1 本章为什么重要

经济学最核心的模型往往不是静态的。家庭要决定今天消费多少、储蓄多少；企业要决定今天投资多少、明天产能如何；政府要决定今天举债还是征税；投资者要决定持有哪种资产。这些问题都含有一个共同结构：**今天的选择会改变明天的状态**。

在静态优化中，我们常写

$$\max_x f(x).$$

在动态优化中，选择对象是一条路径：

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t)$$

subject to

$$k_{t+1} = F(k_t) - c_t.$$

这时最优条件不再只是 $u'(c) = \lambda p$，而会出现跨期边际条件：

$$u'(c_t) = \beta u'(c_{t+1}) F'(k_{t+1}).$$

这就是 Euler equation 的典型形式。

### 1.1 动态优化与前面章节的关系

本章不是突然换主题，而是把前面知识合在一起：

- Chapter 10–13 的动态系统告诉我们路径如何演化；
- Chapter 14–17 的优化告诉我们一阶条件、二阶条件和凸性；
- Chapter 18–19 的包络与对偶告诉我们价值函数和影子价格；
- Chapter 20 的固定点与压缩映射是动态规划存在性的重要基础；
- Chapter 21 现在把这些都用于“跨期最优路径”。

### 1.2 本章要形成的核心直觉

动态优化中最重要的三句话是：

1. **状态 state** 连接过去和未来，例如资本 $k_t$、财富 $a_t$、债务 $b_t$。
2. **控制 control** 是当前选择，例如消费 $c_t$、储蓄 $s_t$、投资 $i_t$。
3. **Euler 方程** 是相邻两期之间的边际权衡条件。

**常见误区**

很多同学学动态优化时只记 Euler 方程公式，这是不够的。真正要理解的是：Euler 方程是一种**无套利型的跨期边际平衡**。如果今天少消费一点、把资源带到明天，明天多消费带来的贴现边际效用必须正好等于今天放弃消费的边际效用。

## 2 两期消费储蓄模型

最简单的动态优化是两期模型。消费者在第 1 期和第 2 期有收入 $y_1, y_2$，选择消费 $c_1, c_2$，可以以利率 $r$ 借贷。令

$$R = 1 + r.$$

终身预算约束为

$$c_1 + \frac{c_2}{R} = y_1 + \frac{y_2}{R}.$$

右边是 lifetime wealth：

$$W = y_1 + \frac{y_2}{R}.$$

问题是

$$\max_{c_1, c_2} u(c_1) + \beta u(c_2)$$

subject to

$$c_1 + \frac{c_2}{R} = W.$$

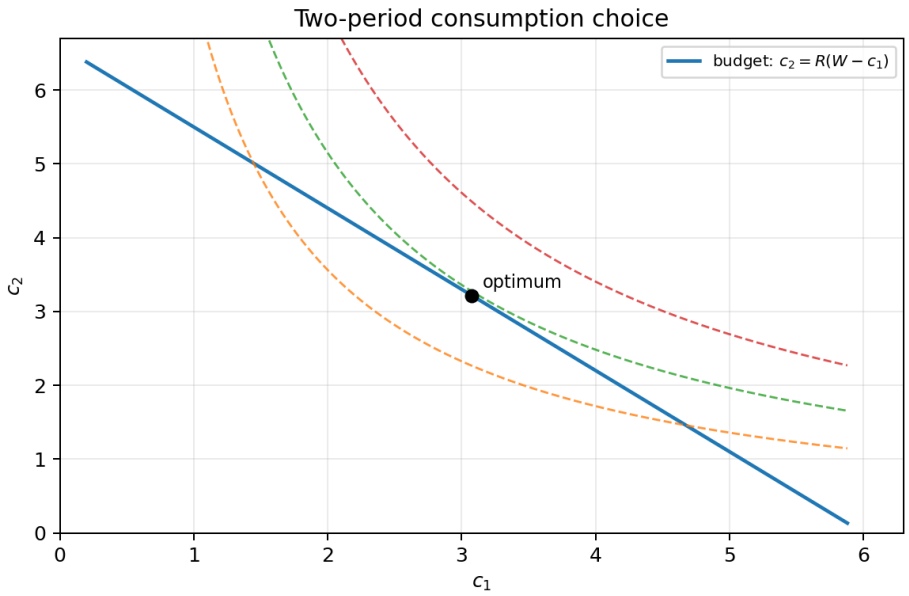



图 1：两期消费选择：最优点处无差异曲线与跨期预算线相切。

## 2.1 Lagrange 推导

写 Lagrangian：

$$\mathcal{L} = u(c_1) + \beta u(c_2) + \lambda \left( W - c_1 - \frac{c_2}{R} \right).$$

一阶条件为

$$\frac{\partial \mathcal{L}}{\partial c_1} = u'(c_1) - \lambda = 0,$$

$$\frac{\partial \mathcal{L}}{\partial c_2} = \beta u'(c_2) - \frac{\lambda}{R} = 0.$$

因此

$$\lambda = u'(c_1),$$

且

$$\lambda = \beta R u'(c_2).$$

合并得

$$\boxed{u'(c_1) = \beta R u'(c_2).}$$

这就是两期 Euler 方程。

**直觉盒子**

左边 $u'(c_1)$ 是今天多消费一单位的边际效用。右边 $\beta R u'(c_2)$ 是今天少消费一单位、存起来后明天得到 $R$ 单位消费，其贴现后的边际效用。最优时二者相等。

## 2.2 log utility 的显式解

若

$$u(c) = \log c,$$

则 Euler 方程为

$$\frac{1}{c_1} = \beta R \frac{1}{c_2}.$$

所以

$$c_2 = \beta R c_1.$$

代入预算约束：

$$c_1 + \frac{\beta R c_1}{R} = c_1 + \beta c_1 = (1+\beta)c_1 = W.$$

因此

$$\boxed{c_1^* = \frac{W}{1+\beta}}, \qquad \boxed{c_2^* = \frac{\beta R W}{1+\beta}}.$$

**Worked Example**

**例 1：两期消费储蓄的完整计算**

设 $u(c) = \log c$，$\beta = 0.95$，$r = 0.05$，$y_1 = 10$，$y_2 = 12$。则

$$R = 1.05, \qquad W = 10 + \frac{12}{1.05} \approx 21.4286.$$

所以

$$c_1^* = \frac{21.4286}{1.95} \approx 10.989,$$

$$c_2^* = \beta R c_1^* = 0.95 \times 1.05 \times 10.989 \approx 10.961.$$

第 1 期储蓄为

$$s_1 = y_1 - c_1^* \approx -0.989.$$

所以消费者第 1 期借款，第 2 期偿还。经济解释是：第 2 期收入更高，消费者为了平滑消费，会把未来收入的一部分提前到今天消费。

# 3 多期有限期问题与 Euler 方程

## 3.1 问题设定

考虑 $T+1$ 期消费储蓄问题。个体最大化

$$\sum_{t=0}^{T} \beta^t u(c_t),$$

subject to

$$a_{t+1} = R(a_t + y_t - c_t), \qquad t = 0, 1, \dots, T,$$

并给定初始资产 $a_0$。为简化，先假设终端资产 $a_{T+1} \geq 0$，且最优时通常有 $a_{T+1} = 0$。

这里：

- $a_t$ 是第 $t$ 期初资产；
- $y_t$ 是第 $t$ 期收入；
- $c_t$ 是消费；
- $R = 1 + r$ 是 gross return。

## 3.2 用扰动法推导 Euler 方程

不急着写大 Lagrangian，我们先用经济直觉推导。考虑最优路径上，把第 $t$ 期消费减少一个很小的量 $\varepsilon$，并把节省下来的资源存到第 $t+1$ 期。第 $t+1$ 期可多消费 $R\varepsilon$。

效用变化近似为

$$-\beta^t u'(c_t)\varepsilon + \beta^{t+1} u'(c_{t+1}) R\varepsilon.$$

最优时，这种局部挪动不能带来一阶收益，所以

$$-\beta^t u'(c_t) + \beta^{t+1} R u'(c_{t+1}) = 0.$$

两边除以 $\beta^t$，得到

$$\boxed{u'(c_t) = \beta R u'(c_{t+1}).}$$

**直觉盒子**

Euler 方程不是神秘公式。它来自一个很小的跨期扰动：今天少消费一点，明天多消费一点。最优路径上，这种挪动的边际收益必须为零。

## 3.3 Lagrange 乘子推导

写现值 Lagrangian：

$$\mathcal{L} = \sum_{t=0}^{T} \beta^t u(c_t) + \sum_{t=0}^{T} \mu_t \left[ R(a_t + y_t - c_t) - a_{t+1} \right].$$

对 $c_t$ 求导：

$$\beta^t u'(c_t) - R\mu_t = 0.$$

所以

$$\mu_t = \frac{\beta^t u'(c_t)}{R}.$$

对 $a_{t+1}$ 求导。$a_{t+1}$ 在第 $t$ 个约束中以 $-a_{t+1}$ 出现，在第 $t+1$ 个约束中以 $Ra_{t+1}$ 出现，所以

$$-\mu_t + R\mu_{t+1} = 0.$$

因此

$$\mu_t = R\mu_{t+1}.$$

代入消费一阶条件：

$$\frac{\beta^t u'(c_t)}{R} = R\frac{\beta^{t+1} u'(c_{t+1})}{R}.$$

整理得

$$u'(c_t) = \beta R u'(c_{t+1}).$$

**常见误区**

不同教材的资产约束写法可能不同，例如

$$a_{t+1} = Ra_t + y_t - c_t$$

或

$$a_{t+1} = R(a_t + y_t - c_t).$$

Euler 方程中的 $R$ 位置会随 timing convention 改变。考试或建模时，先看清楚资产是在消费前计息还是消费后计息。

# 4 无限期问题与横截条件

## 4.1 无限期目标

无限期动态优化的典型形式是

$$\max_{\{c_t, a_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t)$$

subject to

$$a_{t+1} = R(a_t + y_t - c_t).$$

这里 $0 < \beta < 1$。贴现因子保证未来效用权重下降，也常用于让目标函数有良好定义。

## 4.2 为什么 Euler 方程还不够

Euler 方程给出相邻两期之间的最优关系：

$$u'(c_t) = \beta R u'(c_{t+1}).$$

但它通常不能单独确定唯一最优路径。原因是：差分方程需要边界条件。初始资产 $a_0$ 给了一个起点，但无限期问题还需要一个远期条件来排除“不合理地把资源留到无限远”的路径。

这个远期条件就是横截条件 transversality condition。

**定义 / 定理**

**横截条件的常见形式.** 在许多消费储蓄问题中，横截条件可写成

$$\boxed{\lim_{T\to\infty} \beta^T u'(c_T) a_{T+1} = 0.}$$

它表示：在无限远处剩下的资产，其贴现边际价值必须趋于零。

## 4.3 横截条件的直觉

如果最优路径留下大量资产到无限远，并且这些资产的贴现边际价值不为零，那么消费者其实可以在某个有限时期多消费一点而提高效用。那就说明原路径不是最优。

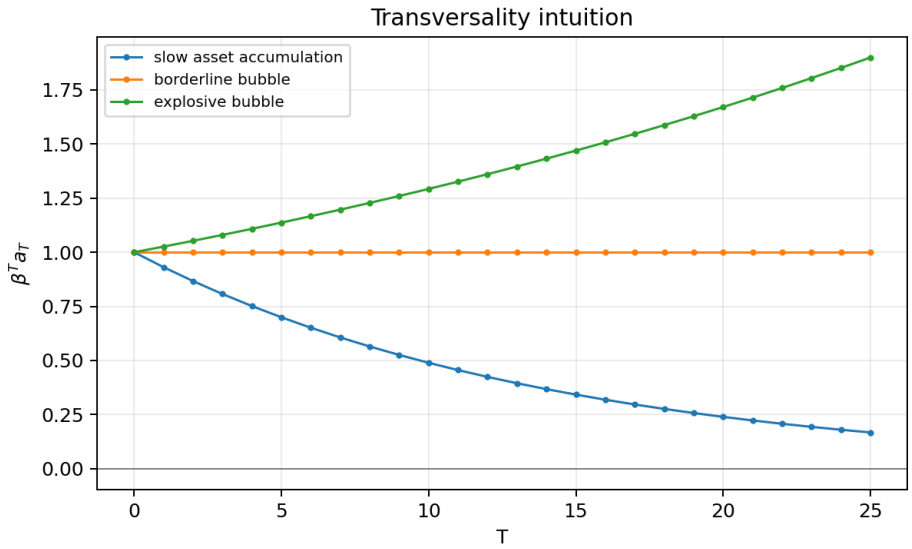



图 2: 横截条件直觉：贴现后的终端资产价值必须消失，否则存在未使用的有价值资源。

**常见误区**

横截条件不是“资产必须等于 0”。在无限期模型里，资产可以增长，但不能增长得太快，以至于贴现后的边际价值仍然为正。它排除的是 Ponzi path、bubble path 或非最优的无限囤积。

# 5 Euler 方程如何决定消费增长

假设效用函数为 CRRA：

$$u(c) = \frac{c^{1-\theta}}{1-\theta}, \qquad \theta > 0, \quad \theta \neq 1.$$

则

$$u'(c) = c^{-\theta}.$$

Euler 方程

$$u'(c_t) = \beta R u'(c_{t+1})$$

变成

$$c_t^{-\theta} = \beta R c_{t+1}^{-\theta}.$$

因此

$$\left(\frac{c_{t+1}}{c_t}\right)^{\theta} = \beta R.$$

所以

$$\boxed{\frac{c_{t+1}}{c_t} = (\beta R)^{1/\theta}.}$$

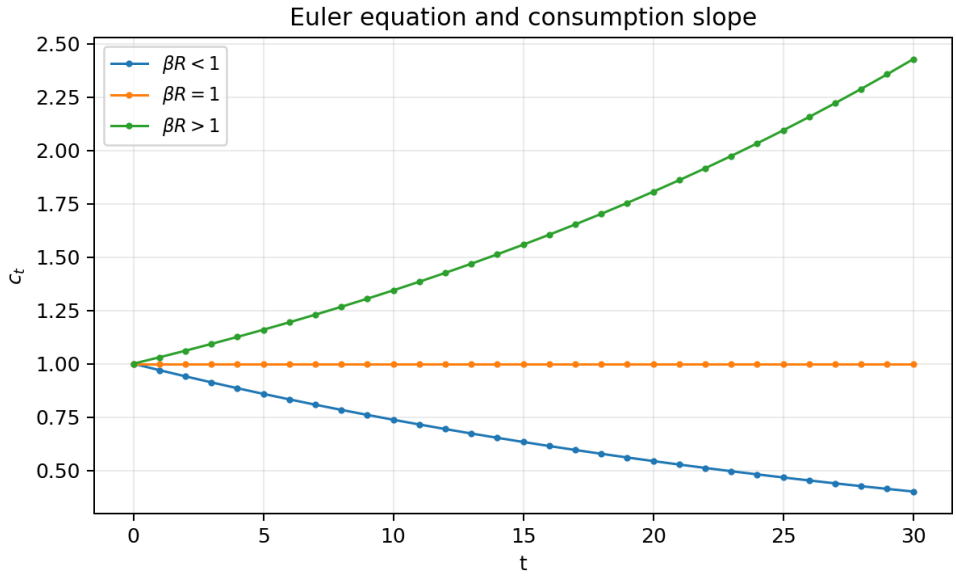



图 3: 在 log utility 情况下，$c_{t+1}/c_t = \beta R$。若 $\beta R > 1$，消费上升；若 $\beta R < 1$，消费下降。

## 5.1 经济解释

- 若 $\beta R > 1$，储蓄回报相对于耐心程度较高，消费者愿意推迟消费，消费路径上升。
- 若 $\beta R < 1$，未来贴现较重或回报较低，消费者倾向于更早消费，消费路径下降。
- $\theta$ 越大，消费者越不愿意让消费在时间上波动，消费增长对 $\beta R$ 的反应越小。

# 6 资产定价 Euler 方程

动态优化和金融资产定价的关系非常紧密。考虑消费者可以持有一项资产，价格为 $p_t$，下一期支付 payoff $x_{t+1}$。如果第 $t$ 期多买一点资产，需要牺牲 $p_t$ 单位消费；第 $t+1$ 期获得 $x_{t+1}$ 单位消费。

最优时：

$$u'(c_t)p_t = \beta \mathbb{E}_t[u'(c_{t+1})x_{t+1}].$$

因此

$$\boxed{p_t = \mathbb{E}_t\left[\beta \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1}\right].}$$

定义 stochastic discount factor：

$$m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}.$$

则资产定价公式为

$$\boxed{p_t = \mathbb{E}_t[m_{t+1}x_{t+1}].}$$

**直觉盒子**

这说明资产价格不是单纯由 payoff 的平均值决定，而是由 payoff 与边际效用状态的关系决定。在坏状态中支付更多的资产更有价值，因为那时 $u'(c_{t+1})$ 更高。

## 6.1 无风险资产

若无风险资产 gross return 为 $R_f$，支付满足 $x_{t+1} = R_f p_t$。代入 Euler 方程：

$$p_t = \mathbb{E}_t[m_{t+1}R_f p_t].$$

若 $p_t > 0$，则

$$\boxed{1 = R_f \mathbb{E}_t[m_{t+1}].}$$

所以

$$\boxed{R_f = \frac{1}{\mathbb{E}_t[m_{t+1}]}.}$$

## 6.2 风险资产

若资产 gross return 为 $R_{t+1}$，Euler 方程写成

$$\boxed{1 = \mathbb{E}_t[m_{t+1}R_{t+1}].}$$

因此风险溢价来自

$$\mathbb{E}_t[R_{t+1}] - R_f$$

与边际效用状态的协方差。粗略地说，如果资产在坏时候收益低，它必须提供更高平均回报来补偿投资者。

# 7 资本积累与 Ramsey 模型入口

## 7.1 社会规划者问题

考虑一个最简单的确定性增长模型。社会规划者选择消费和资本路径，最大化

$$\sum_{t=0}^{\infty} \beta^t u(c_t)$$

subject to

$$k_{t+1} = f(k_t) + (1-\delta)k_t - c_t,$$

给定 $k_0$。

这里 $k_t$ 是状态变量，$c_t$ 是控制变量。资本转移方程说明：明天资本等于今天产出、未折旧资本减去消费。

## 7.2 Lagrangian 推导

写 Lagrangian：

$$\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \left[ u(c_t) + \lambda_t \big( f(k_t) + (1-\delta)k_t - c_t - k_{t+1} \big) \right].$$

对 $c_t$ 求导：

$$u'(c_t) - \lambda_t = 0,$$

即

$$\lambda_t = u'(c_t).$$

对 $k_{t+1}$ 求导。$k_{t+1}$ 在第 $t$ 期约束中以 $-k_{t+1}$ 出现，在第 $t+1$ 期约束中通过 $f(k_{t+1})+(1-\delta)k_{t+1}$ 出现：

$$-\beta^t \lambda_t + \beta^{t+1} \lambda_{t+1} [f'(k_{t+1}) + 1 - \delta] = 0.$$

整理：

$$\lambda_t = \beta \lambda_{t+1} [f'(k_{t+1}) + 1 - \delta].$$

代入 $\lambda_t = u'(c_t)$：

$$\boxed{u'(c_t) = \beta u'(c_{t+1}) [f'(k_{t+1}) + 1 - \delta].}$$

这就是 Ramsey 模型的 Euler 方程。

**直觉盒子**

今天少消费一单位，会多留一单位资本给明天。明天这一单位资本产生的 gross return 是

$$f'(k_{t+1})+1-\delta.$$

最优时，今天消费的边际效用等于明天增加消费可能带来的贴现边际效用。

## 7.3 稳态条件

在稳态，$c_{t+1}=c_t=c^*$，$k_{t+1}=k_t=k^*$。Euler 方程变成

$$1=\beta[f'(k^*)+1-\delta].$$

所以

$$\boxed{f'(k^*)=\frac{1}{\beta}-1+\delta.}$$

这说明最优稳态资本由资本边际产品、折旧和贴现因子共同决定。

资源约束在稳态给出

$$k^*=f(k^*)+(1-\delta)k^*-c^*,$$

即

$$\boxed{c^*=f(k^*)-\delta k^*.}$$

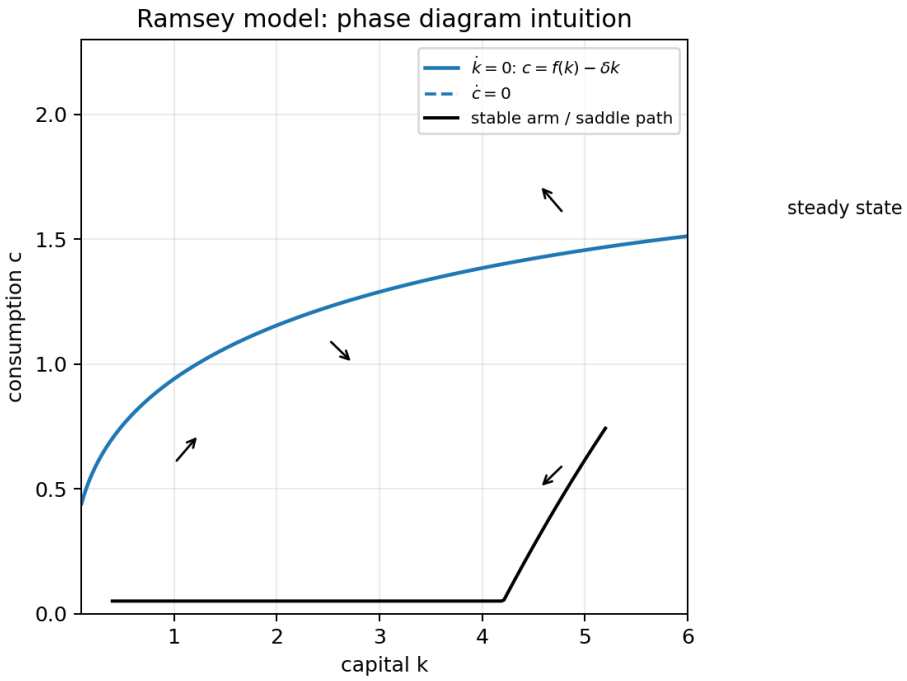



图 4: Ramsey 模型的相图直觉：最优路径必须落在 saddle path 上，同时满足资源约束、Euler 方程和横截条件。

# 8 现值乘子与当前值乘子

动态优化中，乘子有两种常见写法：present-value multiplier 和 current-value multiplier。

## 8.1 现值乘子

如果目标函数是

$$\sum_{t=0}^{\infty} \beta^t u(c_t),$$

并把约束直接乘以 $\mu_t$，则 $\mu_t$ 通常是现值乘子，它已经反映第 $t$ 期约束价值在第 0 期的贴现价值。

## 8.2 当前值乘子

也可以写成

$$\sum_{t=0}^{\infty} \beta^t \left[u(c_t) + \lambda_t (extconstraint)\right].$$

此时 $\lambda_t$ 是当前值乘子 current-value multiplier，表示第 $t$ 期约束在第 $t$ 期的边际价值。

二者关系通常是

$$\mu_t = \beta^t \lambda_t.$$

**常见误区**

现值乘子和当前值乘子的差别经常导致符号和贴现因子错误。判断方法是：如果乘子在求和符号外没有乘 $\beta^t$，它多半是现值乘子；如果整个括号都乘 $\beta^t$，括号内的乘子通常是当前值乘子。

# 9 企业投资与调整成本

动态优化不仅用于消费者，也用于企业投资。设企业资本为 $K_t$，投资为 $I_t$，资本积累为

$$K_{t+1} = (1-\delta)K_t + I_t.$$

若投资有调整成本，例如

$$\frac{\phi}{2}\left(\frac{I_t}{K_t}\right)^2 K_t,$$

企业不会无限快调整资本。一个典型结果是投资率与 marginal q 相关：

$$\boxed{\frac{I_t}{K_t} = \frac{q_t - 1}{\phi}.}$$

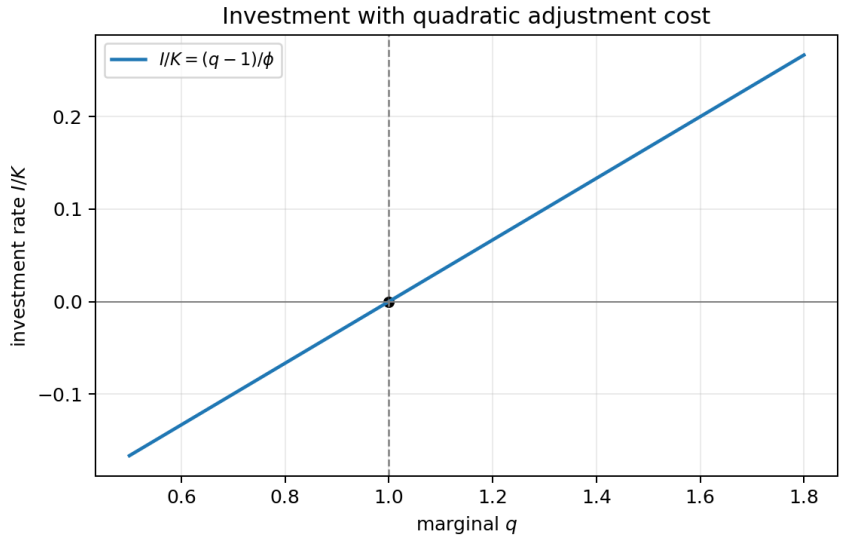



图 5: 有二次调整成本时，marginal $q$ 越高，企业越愿意投资；$q = 1$ 时净投资率为 0。

**直觉盒子**

静态投资理论可能说“只要边际收益高于成本就投资到相等为止”。动态投资理论强调：资本调整有成本，今天投资不仅影响今天利润，还影响未来资本状态。因此最优投资由资本的影子价值 $q$ 决定。

## 10 连续时间版本预告：Euler-Lagrange 与 Hamiltonian

本章主要讲离散时间动态优化，但后面还会进入连续时间最优控制。连续时间问题常写成

$$\max_{c(t)} \int_0^{\infty} e^{-\rho t} u(c(t))\,\mathrm{d}t$$

subject to

$$\dot{k}(t) = f(k(t)) - c(t) - \delta k(t).$$

其当前值 Hamiltonian 是

$$\mathcal{H} = u(c) + \lambda[f(k) - c - \delta k].$$

一阶条件为

$$u'(c) = \lambda,$$

协态方程为

$$\dot{\lambda} = \rho\lambda - \lambda[f'(k) - \delta].$$

合并后可得连续时间 Ramsey Euler 方程：

$$\boxed{\frac{\dot{c}}{c} = \frac{1}{\theta}[f'(k) - \delta - \rho]}$$

在 CRRA 效用下成立。

**Advanced Core**

离散时间 Euler 方程和连续时间 Euler 方程是一套思想的两个版本。离散时间中比较的是 $t$ 和 $t+1$；连续时间中比较的是相邻瞬间。离散式

$$u'(c_t) = \beta u'(c_{t+1}) R_{t+1}$$

在时间间隔趋近于零时，会变成关于 $\dot{c}/c$ 的微分方程。

## 11 动态优化的计算流程

**方法模板**

**离散时间动态优化的六步法**

1. 明确状态变量 state，例如 $a_t, k_t, b_t$。
2. 明确控制变量 control，例如 $c_t, i_t, s_t$。
3. 写目标函数：通常是 $\sum \beta^t u(\cdot)$。
4. 写状态转移方程：例如 $a_{t+1} = R(a_t + y_t - c_t)$。
5. 写 Lagrangian 或做边际扰动，推出 Euler 方程。
6. 加上初始条件和终端/横截条件，确定最优路径。

**方法模板**

**检查 Euler 方程经济含义的模板**

1. 左边通常是今天控制变量的边际收益或边际效用。
2. 右边通常是明天收益乘以贴现因子和回报率。
3. 如果右边大于左边，说明应该把资源从今天挪到明天。
4. 如果左边大于右边，说明应该把资源从明天挪回今天。
5. 最优时二者相等。

## 12 常见错误与纠正

**常见误区**

**误区 1：把 Euler 方程当成预算约束。** Euler 方程是边际最优条件，不是资源约束。完整解必须同时满足 Euler 方程、预算/状态转移方程、初始条件和终端/横截条件。

**误区 2：忘记 timing convention。** 资产何时计息、资本何时生产、投资何时转化为资本，会影响 Euler 方程里的 $R$、$f'(k)$ 和折旧项位置。

**误区 3：以为横截条件可有可无。** 无限期问题中，Euler 方程通常允许很多路径。横截条件负责排除泡沫、Ponzi 路径或无限囤积资源的非最优路径。

**误区 4：混淆现值乘子和当前值乘子。** 现值乘子与当前值乘子差一个 $\beta^t$。如果不分清楚，协态方程和 Euler 方程很容易多一个或少一个贴现因子。

**误区 5：只会写公式，不会解释经济含义。** 动态优化最重要的是跨期边际替代：今天和未来之间的边际收益必须平衡。

## 13 本章小结表

| 概念 | 核心公式 | 经济含义 |
|---|---|---|
| 两期预算 | $c_1 + c_2/R = W$ | 未来资源折现到今天。 |
| 两期 Euler | $u'(c_1) = \beta R u'(c_2)$ | 今天消费与明天消费的边际平衡。 |
| CRRA 消费增长 | $c_{t+1}/c_t = (\beta R)^{1/\theta}$ | 利率、耐心和跨期替代弹性决定消费斜率。 |
| 横截条件 | $\lim \beta^T u'(c_T) a_{T+1} = 0$ | 排除无限远处留下有价值资源。 |
| 资产定价 | $p_t = \mathbb{E}_t[m_{t+1} x_{t+1}]$ | payoff 在坏状态越高，资产越有价值。 |
| Ramsey Euler | $u'(c_t) = \beta u'(c_{t+1})(f'(k_{t+1}) + 1 - \delta)$ | 今天消费与投资成未来资本之间的权衡。 |
| 稳态资本 | $f'(k^*) = 1/\beta - 1 + \delta$ | 最优长期资本由耐心、折旧和边际产品决定。 |
| 当前值乘子 | $\mu_t = \beta^t \lambda_t$ | 区分第 0 期价值和第 $t$ 期价值。 |

## 14 练习题

### A 组：基础计算

A1. 两期模型中，$u(c) = \log c$，$\beta = 1$，$R = 1.1$，$y_1 = 10, y_2 = 11$。求 $c_1^*, c_2^*$。

A2. 证明两期模型的一阶条件为 $u'(c_1) = \beta R u'(c_2)$。

A3. 对 CRRA 效用 $u(c) = c^{1-\theta}/(1-\theta)$，从 Euler 方程推出 $c_{t+1}/c_t = (\beta R)^{1/\theta}$。

A4. 若 $\beta = 0.96$，$R = 1.04$，$\theta = 2$，求消费增长率 $c_{t+1}/c_t - 1$。

A5. 若资产 payoff 为无风险 $R_f$，从 $1 = \mathbb{E}_t[m_{t+1} R_f]$ 推出 $R_f = 1/\mathbb{E}_t[m_{t+1}]$。

### B 组：概念判断

B1. Euler 方程是否足以唯一决定无限期最优路径？为什么？

B2. 横截条件的经济含义是什么？

B3. 为什么在 Ramsey 模型中，资本的回报是 $f'(k_{t+1}) + 1 - \delta$？

B4. $\theta$ 在 CRRA 效用中如何影响消费平滑？

B5. 为什么资产在坏状态支付更多会更有价值？

### C 组：混合题

C1. 两期模型中，$u(c) = \sqrt{c}$。推导 $c_2/c_1$ 与 $\beta R$ 的关系。

C2. 多期消费储蓄问题中，用边际扰动法推导 Euler 方程。

C3. Ramsey 模型中，$f(k) = Ak^\alpha$。求稳态 $k^*$ 的表达式。

C4. 设 $u(c) = \log c$, Ramsey Euler 方程为 $1/c_t = \beta(1/c_{t+1})(f'(k_{t+1}) + 1 - \delta)$。整理出 $c_{t+1}/c_t$。

C5. 若 $m_{t+1}$ 与资产收益 $R_{t+1}$ 负相关，解释为什么该资产需要更高平均收益。

### D 组：经济金融建模题

D1. 写出一个三期消费储蓄问题，包括目标函数、预算约束和终端资产条件。

D2. 用 Euler 方程解释永久收入假说的基本直觉。

D3. 写出 Lucas asset pricing 的基本 Euler 方程，并解释 stochastic discount factor。

D4. 企业有调整成本时，为什么投资取决于 marginal $q$ 而不只是当前利润？

D5. Ramsey 模型中，若 $\beta$ 上升，稳态资本会怎样变化？用公式解释。

### E 组：挑战 / 证明题

E1. 对有限期消费储蓄问题，写出完整 Lagrangian，并推导 Euler 方程和终端条件。

E2. 在两期模型中，证明利率上升对 $c_1$ 的影响包含替代效应和财富效应。只需给出直觉和公式方向。

E3. 对 Ramsey 模型，推导稳态条件 $f'(k^*) = 1/\beta - 1 + \delta$ 与 $c^* = f(k^*) - \delta k^*$。

E4. 设 $u$ 严格凹、约束集凸。解释为什么满足 Euler 方程和横截条件的路径是全局最优的候选，并说明还需要哪些条件。

# 15 练习题参考答案

## A 组答案

A1. lifetime wealth 为

$$W = 10 + \frac{11}{1.1} = 20.$$

对 log utility，有

$$c_1^* = \frac{W}{1+\beta} = 10, \qquad c_2^* = \beta R c_1^* = 11.$$

A2. Lagrangian 为

$$\mathcal{L} = u(c_1) + \beta u(c_2) + \lambda\left(W - c_1 - \frac{c_2}{R}\right).$$

一阶条件：

$$u'(c_1) = \lambda, \qquad \beta u'(c_2) = \lambda / R.$$

第二式乘以 $R$ 得 $\lambda = \beta R u'(c_2)$。与第一式合并：

$$u'(c_1) = \beta R u'(c_2).$$

A3. CRRA 下 $u'(c) = c^{-\theta}$。Euler 方程给出

$$c_t^{-\theta} = \beta R c_{t+1}^{-\theta}.$$

整理为

$$\left(\frac{c_{t+1}}{c_t}\right)^{\theta} = \beta R.$$

因此

$$\frac{c_{t+1}}{c_t} = (\beta R)^{1/\theta}.$$

A4.

$$\beta R = 0.96 \times 1.04 = 0.9984.$$

当 $\theta = 2$,

$$\frac{c_{t+1}}{c_t} = 0.9984^{1/2} \approx 0.9992.$$

所以消费增长率约为 $-0.08\%$。

A5. 若 $R_f$ 在 $t+1$ 已知，则

$$1 = \mathbb{E}_t[m_{t+1} R_f] = R_f \mathbb{E}_t[m_{t+1}].$$

因此

$$R_f = \frac{1}{\mathbb{E}_t[m_{t+1}]}.$$

## B 组答案

B1. 不足以。Euler 方程只给出相邻两期的边际平衡，通常还需要预算约束、初始条件和终端/横截条件来排除不合理路径。

B2. 横截条件表示无限远处剩余资产的贴现边际价值必须为零。否则说明经济主体留下了有价值但未使用的资源，可以通过提高有限期消费来改善效用。

B3. 今天多留一单位资本到明天，明天既能产生边际产出 $f'(k_{t+1})$，又有未折旧资本 $1-\delta$。所以 gross return 是 $f'(k_{t+1})+1-\delta$。

B4. $\theta$ 是相对风险厌恶系数，也是跨期替代弹性的倒数。$\theta$ 越大，消费者越不愿意让消费在不同期之间波动，消费增长对回报变化越不敏感。

B5. 坏状态通常消费低，边际效用高。如果资产在坏状态支付更多，它提供保险，因此更有价值，所需平均回报可以更低。

## C 组答案

C1. $u(c)=\sqrt{c}$，所以

$$u'(c)=\frac{1}{2\sqrt{c}}.$$

Euler 方程为

$$\frac{1}{2\sqrt{c_1}}=\beta R\frac{1}{2\sqrt{c_2}}.$$

整理得

$$\sqrt{\frac{c_2}{c_1}}=\beta R,$$

因而

$$\frac{c_2}{c_1}=(\beta R)^2.$$

C2. 在最优路径上，令第 $t$ 期消费减少 $\varepsilon$，第 $t+1$ 期消费增加 $R\varepsilon$。效用一阶变化为

$$-\beta^t u'(c_t)\varepsilon+\beta^{t+1}u'(c_{t+1})R\varepsilon.$$

最优时等于 0，得到

$$u'(c_t)=\beta R u'(c_{t+1}).$$

C3. Ramsey 稳态满足

$$1=\beta[A\alpha(k^*)^{\alpha-1}+1-\delta].$$

所以

$$A\alpha(k^*)^{\alpha-1}=\frac{1}{\beta}-1+\delta.$$

因此

$$k^*=\left(\frac{A\alpha}{1/\beta-1+\delta}\right)^{\frac{1}{1-\alpha}}.$$

C4. log utility 下，Euler 方程为

$$\frac{1}{c_t} = \beta \frac{1}{c_{t+1}} [f'(k_{t+1}) + 1 - \delta].$$

因此

$$\frac{c_{t+1}}{c_t} = \beta [f'(k_{t+1}) + 1 - \delta].$$

C5. 若 $m_{t+1}$ 与 $R_{t+1}$ 负相关，说明资产在边际效用高的坏状态收益低，在好状态收益高。它不能提供保险，反而加剧消费风险。因此投资者要求更高平均收益作为补偿。

## D 组答案

D1. 一个三期问题可写为

$$\max_{c_0, c_1, c_2, a_1, a_2, a_3} u(c_0) + \beta u(c_1) + \beta^2 u(c_2)$$

subject to

$$a_{t+1} = R(a_t + y_t - c_t), \qquad t = 0, 1, 2,$$

给定 $a_0$，并可设终端条件 $a_3 \geq 0$。若无遗产动机且效用递增，通常最优时 $a_3 = 0$。

D2. 永久收入假说的核心是消费者希望平滑消费。Euler 方程说明，在没有不确定性且 $\beta R \approx 1$ 时，边际效用近似恒定；如果效用严格凹，则消费本身也近似平滑。消费者不是根据当期收入消费，而是根据终身资源消费。

D3. Lucas asset pricing 的基本 Euler 方程为

$$p_t = \mathbb{E}_t \left[ \beta \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1} \right].$$

其中

$$m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}$$

是 stochastic discount factor，表示未来 payoff 转换为今天价值的状态价格权重。

D4. 因为投资改变未来资本存量，且资本调整有成本。当前利润只反映今天收益，而 marginal $q$ 反映多一单位资本对未来利润现值的边际贡献。因此有调整成本时，投资由 $q$ 决定。

D5. 稳态条件

$$f'(k^*) = \frac{1}{\beta} - 1 + \delta.$$

若 $\beta$ 上升，则 $1/\beta$ 下降，右边下降。若 $f'' < 0$，要让 $f'(k^*)$ 下降，必须提高 $k^*$。所以更有耐心的经济体选择更高稳态资本。

## E 组答案

E1. 有限期 Lagrangian 可写为

$$\mathcal{L} = \sum_{t=0}^{T} \beta^t u(c_t) + \sum_{t=0}^{T} \mu_t [R(a_t + y_t - c_t) - a_{t+1}].$$

对 $c_t$：

$$\beta^t u'(c_t) = R\mu_t.$$

对 $a_{t+1}$：

$$\mu_t = R\mu_{t+1}, \quad t = 0, \dots, T-1.$$

合并得

$$u'(c_t) = \beta R u'(c_{t+1}).$$

若终端资产可自由且无遗产动机，递增效用意味着不浪费资源，故终端资产约束通常绑定，例如 $a_{T+1} = 0$ 或对应互补条件成立。

E2. 利率上升有两个方向的影响。替代效应：未来消费相对今天消费变便宜，因为储蓄回报更高，所以倾向于减少 $c_1$、增加 $c_2$。财富效应：如果个体是净储蓄者，利率上升提高终身财富，可能增加当前消费；如果是净借款者，利率上升降低终身财富，减少当前消费。因此总效应取决于借贷身份和效用函数。

E3. Ramsey Euler 稳态中 $c_{t+1} = c_t$，所以

$$1 = \beta[f'(k^*) + 1 - \delta].$$

整理得

$$f'(k^*) = \frac{1}{\beta} - 1 + \delta.$$

资源约束稳态中 $k_{t+1} = k_t = k^*$，所以

$$k^* = f(k^*) + (1-\delta)k^* - c^*.$$

整理得

$$c^* = f(k^*) - \delta k^*.$$

E4. 严格凹效用和凸约束集保证目标函数是凹的，可行集是凸的，因此满足一阶必要条件的内点路径通常也是全局最优候选。还需要可行性、适当的边界条件、横截条件以及避免非内点约束遗漏。如果存在借贷约束或非负消费约束，还需要 KKT 条件和互补松弛。

## 16 后续学习连接

本章用 Lagrange 方法推导了 Euler 方程。下一章将转向动态规划 dynamic programming。动态规划不直接选择整条路径，而是把问题写成递归形式：

$$V(k) = \max_{c} \{u(c) + \beta V(k')\}.$$

它的核心思想是 Bellman principle of optimality：从今天最优路径的明天开始，剩余路径也必须是明天状态下的最优路径。届时我们会看到，Euler 方程也可以从 Bellman 方程的一阶条件和 envelope condition 推出。

**直觉盒子**

本章的 Euler 方程回答“相邻两期边际条件是什么”。下一章的 Bellman 方程回答“每一个状态下的最优价值是什么”。二者是动态优化的两种语言：Euler equation 更像路径条件，Bellman equation 更像递归价值条件。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 22
# 动态规划与 Bellman 方程

Chapter 21 用 Euler 方程刻画了跨期最优路径。本章进入动态规划 dynamic programming。它的核心思想是：一个复杂的多期优化问题可以被拆成“今天的选择”和“明天开始的同类问题”。Bellman 方程把无限期问题转化为一个关于价值函数的递归方程。对经济金融来说，它是现代宏观、资产定价、搜索匹配、储蓄投资、企业动态、风险控制和数值求解的基础语言。

math4econ personalized notes
2026

# 目录

# Chapter 22 总览

| | |
|---|---|
| **本章主题** | 动态规划 dynamic programming、Bellman equation、value function、policy function、contraction mapping、value iteration。 |
| **核心问题** | 如何把一个多期甚至无限期优化问题拆成递归形式？怎样从 Bellman 方程得到最优策略？为什么价值迭代会收敛？Bellman 与 Euler 方程有什么关系？ |
| **数学工具** | 有限期 backward induction、无限期 Bellman 方程、Bellman operator、压缩映射、价值函数、策略函数、包络条件、Euler 方程、马尔可夫状态、数值网格。 |
| **经济应用** | cake-eating problem、消费储蓄、确定性增长、随机收入储蓄、资产价格、工作搜索、企业进入退出、动态规划数值方法。 |
| **后续连接** | Chapter 23 会继续进入最优控制和 Hamiltonian；后续章节会把 Bellman 方程连接到随机动态优化、资产定价和数值计算。 |

**直觉盒子**

动态规划的核心不是一个公式，而是一句话：**从明天开始的最优问题，和今天面对的问题在结构上是同一个问题**。因此可以把“未来全部效用”压缩成一个函数 $V(\text{state})$。今天只需要在“当前收益”和“明天价值”之间做权衡：

$$V(k) = \max_{c} \{u(c) + \beta V(k')\}.$$

这就是 Bellman 方程的直觉。

## 1 本章为什么重要

Chapter 21 的 Euler 方程方法很直接：写出跨期 Lagrangian，对每一期选择变量求一阶条件，然后整理得到相邻两期的边际条件。它适合推导理论，也能帮助理解跨期替代。但是当问题变得更复杂时，Euler 方程方法会很笨重：

- 状态变量很多，例如财富、资本、生产率、债务、资产持仓；
- 不确定性存在，例如收入冲击、资产收益、失业风险；
- 约束可能绑定，例如借贷约束、不可逆投资、有限责任；
- 问题可能无限期，无法直接写出所有期的一阶条件；
- 我们希望数值求解，而不是只做解析推导。

动态规划的优势在于，它把整个未来压缩成一个 value function，并用递归结构求解。

## 1.1 从 Euler 到 Bellman 的转变

Euler 方法关注路径:

$$\{c_t, k_{t+1}\}_{t=0}^{\infty}.$$

Bellman 方法关注函数:

$$V(k), \qquad c = g(k), \qquad k' = h(k).$$

也就是说，Euler 方法问“最优路径满足什么条件”；Bellman 方法问“给定任何状态，最优决策规则是什么”。

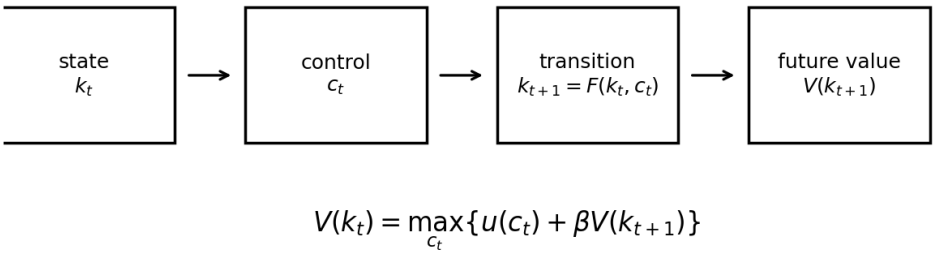



图 1: Bellman 方程把当前状态、当前控制、状态转移和未来价值连接起来。

## 1.2 动态规划的三个对象

| 对象 | 符号 | 经济含义 |
|---|---|---|
| 状态变量 | $x$ or $k$ | 历史带到今天的东西,例如财富、资本、生产率、债务。 |
| 控制变量 | $c$ or $a'$ | 今天可以选择的东西,例如消费、储蓄、投资、求职强度。 |
| 价值函数 | $V(x)$ | 从状态 $x$ 出发，未来能获得的最大贴现效用或利润。 |
| 策略函数 | $g(x)$ | 给定状态 $x$ 时的最优选择规则。 |
| 转移方程 | $x' = F(x, c, z')$ | 今天的状态和选择如何决定下一期状态。 |

# 2 有限期动态规划：Backward Induction

先从有限期开始。有限期问题最清楚，因为它有明确终点。设问题从 $t=0$ 到 $T$，状态为 $x_t$，控制为 $a_t$，即时收益为 $u(x_t,a_t)$，转移方程为

$$x_{t+1}=F(x_t,a_t).$$

目标是

$$\max_{\{a_t\}_{t=0}^{T}} \sum_{t=0}^{T} \beta^t u(x_t,a_t)$$

subject to

$$x_{t+1}=F(x_t,a_t), \qquad x_0 \text{ given}.$$

## 2.1 最后一期先解

在最后一期 $T$，没有未来，因此价值函数为

$$V_T(x)=\max_{a\in A(x)} u(x,a).$$

第 $T-1$ 期的价值是

$$V_{T-1}(x)=\max_{a\in A(x)} \{u(x,a)+\beta V_T(F(x,a))\}.$$

一般地，递归为

$$\boxed{V_t(x)=\max_{a\in A(x)} \{u(x,a)+\beta V_{t+1}(F(x,a))\}, \qquad t=T-1,\dots,0.}$$

这叫 backward induction。

**方法模板**

**有限期动态规划四步法**

1. 明确状态 $x_t$、控制 $a_t$、约束集 $A(x_t)$ 和转移方程。
2. 写终端价值 $V_T(x)$。
3. 从 $T-1$ 往前递推 Bellman 方程。
4. 每一期得到一个策略函数 $a_t=g_t(x_t)$。

**Worked Example**

**例 1：两期 cake-eating problem**

消费者有一块蛋糕，初始大小为 $W_0$。第 0 期消费 $c_0$，剩余 $W_1 = W_0 - c_0$ 留到第 1 期。没有生产，也没有利息。效用为

$$u(c_0) + \beta u(c_1), \qquad c_0 + c_1 = W_0.$$

动态规划写法如下。最后一期价值为

$$V_1(W) = \max_{0 \le c \le W} u(c) = u(W).$$

第 0 期价值为

$$V_0(W) = \max_{0 \le c \le W} \{u(c) + \beta V_1(W - c)\}.$$

因为 $V_1(W - c) = u(W - c)$，所以

$$V_0(W) = \max_{0 \le c \le W} \{u(c) + \beta u(W - c)\}.$$

一阶条件为

$$u'(c_0) = \beta u'(W - c_0) = \beta u'(c_1).$$

这正是两期 Euler 方程。

## 3 无限期 Bellman 方程

有限期问题中价值函数带有时间下标 $V_t$。无限期 stationary problem 的重要简化是：如果环境不随时间变化，那么从每一期开始面对的问题都一样。因此价值函数不再依赖日历时间 $t$，而只依赖状态 $x$。

**定义 / 定理**

**无限期 Bellman 方程.** 对问题

$$\max_{\{a_t\}} \sum_{t=0}^{\infty} \beta^t u(x_t, a_t)$$

subject to

$$x_{t+1} = F(x_t, a_t),$$

stationary Bellman equation 为

$$\boxed{V(x) = \max_{a \in A(x)} \{u(x, a) + \beta V(F(x, a))\}.}$$

若 $g(x)$ 是达到最大值的选择，则称 $g$ 为 policy function。

### 3.1 Bellman 方程的经济解释

Bellman 方程可以拆成三部分：

$$V(x) = \max_a \{ \underbrace{u(x,a)}_{\text{current payoff}} + \beta \underbrace{V(F(x,a))}_{\text{continuation value}} \}.$$

它说：今天的最优价值等于当前收益加上贴现后的明天最优价值。注意这里的 $V(F(x,a))$ 已经假定从明天开始继续最优。

**常见误区**

Bellman 方程不是把未来“忘掉”，而是把未来全部压缩到 $V(x')$ 中。最常见误解是以为动态规划只看眼前收益。事实上，Bellman 方程通过 continuation value 精确地保留了全部未来。

### 3.2 Bellman operator

为了讨论存在性和迭代计算，我们把右边定义为一个算子。给定任意函数 $V$，定义

$$(\mathcal{T}V)(x) = \max_{a \in A(x)} \{u(x,a) + \beta V(F(x,a))\}.$$

Bellman 方程就是 fixed point equation：

$$\boxed{V = \mathcal{T}V.}$$

这与 Chapter 20 的不动点定理直接相连。

## 4 压缩映射与价值迭代

动态规划之所以强大，是因为 Bellman operator 在常见条件下是 contraction。也就是说，对两个候选价值函数 $V, W$，它会把二者之间的距离缩小。

**定义 / 定理**

**压缩映射直觉.** 若存在 $0 < \beta < 1$，使得

$$\|\mathcal{T}V - \mathcal{T}W\|_\infty \le \beta \|V - W\|_\infty,$$

则 $\mathcal{T}$ 是压缩映射。Banach fixed point theorem 保证：

1. 存在唯一不动点 $V^*$；
2. 从任意初始函数 $V_0$ 出发，迭代 $V_{n+1} = \mathcal{T}V_n$ 会收敛到 $V^*$。

### 4.1 为什么贴现因子产生压缩

Bellman operator 中未来价值前面有 $\beta$：

$$u(x,a)+\beta V(x').$$

如果你把候选价值函数从 $V$ 改成 $W$，当前收益不变，只有 continuation value 改变，而且这个改变被乘以 $\beta$。因此两个价值函数之间的差距每次迭代大约缩小为原来的 $\beta$。

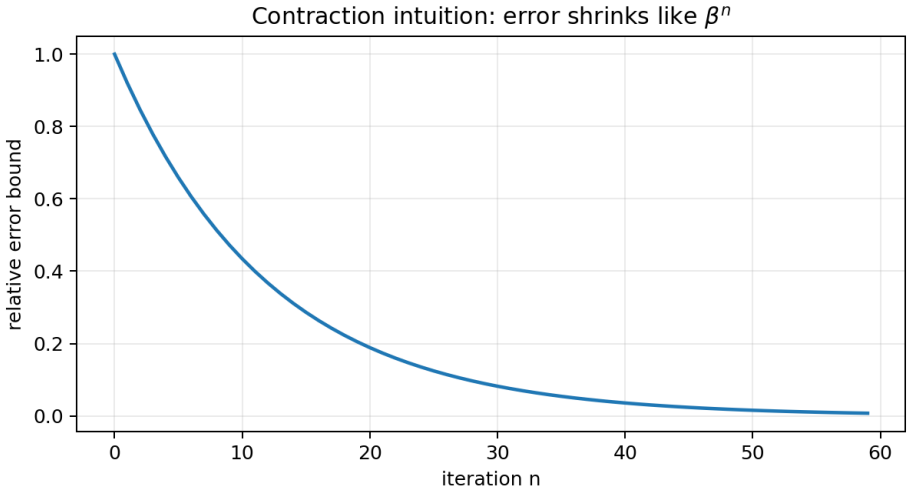



图 2：价值迭代的误差上界通常按 $\beta^n$ 衰减。$\beta$ 越接近 1，收敛越慢。

**方法模板**

**Value iteration 算法**

1. 选择初始猜测 $V_0(x)$，例如 $V_0(x)=0$。
2. 对每个状态 $x$，计算

$$V_{n+1}(x)=\max_{a\in A(x)}\{u(x,a)+\beta V_n(F(x,a))\}.$$

3. 重复直到 $\|V_{n+1}-V_n\|$ 足够小。
4. 由最终价值函数求 policy function：

$$g(x)=\arg\max_{a\in A(x)}\{u(x,a)+\beta V(F(x,a))\}.$$

## 5 价值函数与策略函数

## 5.1 价值函数 value function

价值函数 $V(x)$ 表示：从状态 $x$ 出发，未来所有最优选择带来的最大贴现价值。它不是当前效用，而是整个未来的总结。

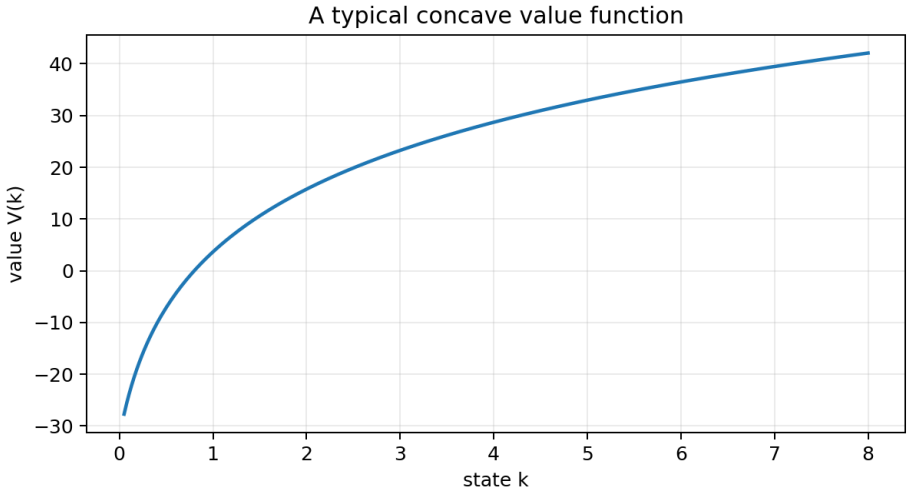



图 3: 很多经济问题中，价值函数是递增且凹的：状态越多越好，但边际价值递减。

如果状态是财富 $a$，则 $V(a)$ 是从财富 $a$ 开始可达到的最大终身效用。如果状态是资本 $k$，则 $V(k)$ 是从资本 $k$ 开始的最大社会福利或企业价值。

## 5.2 策略函数 policy function

策略函数 $g(x)$ 告诉我们：给定状态 $x$，最优控制变量是多少。例如消费储蓄问题中，策略函数可能是

$$c = g(a),$$

表示财富为 $a$ 时最优消费多少。资本积累问题中可能是

$$k' = h(k),$$

表示当前资本为 $k$ 时下期资本选择。

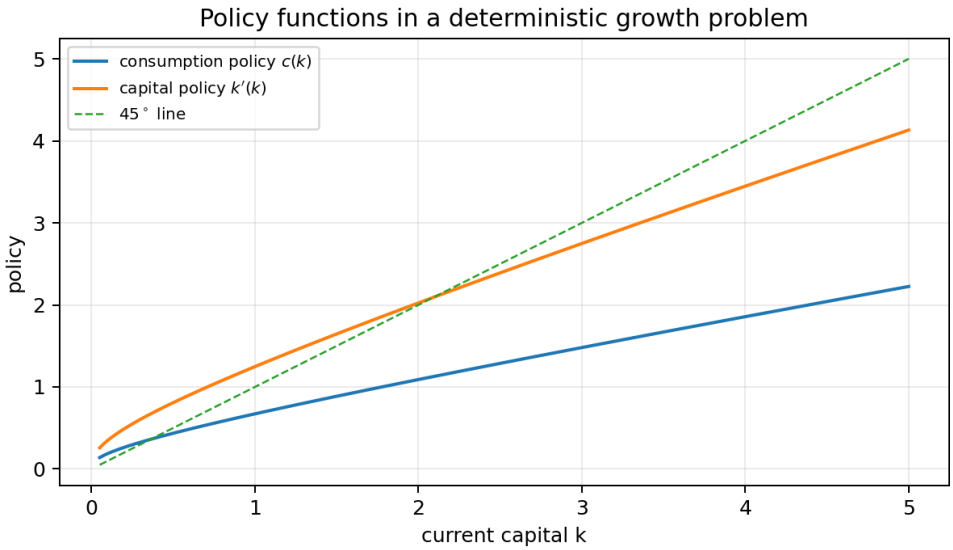



图 4：确定性增长模型中，策略函数给出每个资本状态下的最优消费和下期资本。

# 6 Bellman 与 Euler 方程的关系

动态规划不是替代 Euler 方程，而是生成 Euler 方程。用一个确定性增长问题说明。

## 6.1 确定性增长模型

社会规划者问题：

$$\max_{\{c_t,k_{t+1}\}} \sum_{t=0}^{\infty} \beta^t u(c_t)$$

subject to

$$c_t + k_{t+1} = f(k_t) + (1-\delta)k_t,$$

$$c_t \geq 0, \qquad k_{t+1} \geq 0.$$

Bellman 方程为

$$\boxed{V(k) = \max_{0 \leq k' \leq f(k)+(1-\delta)k} \left\{ u(f(k) + (1-\delta)k - k') + \beta V(k') \right\}.}$$

这里控制变量可以选 $k'$，消费由资源约束决定：

$$c = f(k) + (1-\delta)k - k'.$$

## 6.2 一阶条件

对 $k'$ 求导。目标函数为

$$u(c) + \beta V(k'), \qquad c = f(k) + (1-\delta)k - k'.$$

所以

$$\frac{\partial c}{\partial k'} = -1.$$

一阶条件为

$$-u'(c) + \beta V'(k') = 0,$$

即

$$\boxed{u'(c_t) = \beta V'(k_{t+1}).}$$

### 6.3 包络条件

对状态 $k$ 使用 envelope theorem：

$$V'(k) = u'(c)\left[f'(k) + 1 - \delta\right].$$

把 $k$ 换成下一期 $k_{t+1}$：

$$V'(k_{t+1}) = u'(c_{t+1})[f'(k_{t+1}) + 1 - \delta].$$

代入一阶条件，得到 Euler 方程：

$$\boxed{u'(c_t) = \beta u'(c_{t+1})[f'(k_{t+1}) + 1 - \delta].}$$

**直觉盒子**

Bellman 的 FOC 给出“今天少存一单位资本的边际效用损失等于明天资本价值”。Envelope condition 给出“明天一单位资本值多少钱”。两者合在一起，就是 Euler equation。

## 7 解析例子：无限期 Cake-Eating Problem

考虑没有生产、没有利息的蛋糕问题。初始资源为 $W_0$，每期消费 $c_t$，剩余资源转到下一期：

$$W_{t+1} = W_t - c_t, \qquad 0 \le c_t \le W_t.$$

目标为

$$\max_{\{c_t\}} \sum_{t=0}^{\infty} \beta^t \log c_t, \qquad 0 < \beta < 1.$$

Bellman 方程为

$$V(W) = \max_{0 \le c \le W} \{\log c + \beta V(W - c)\}.$$

### 7.1 猜测并验证

猜测价值函数形式为

$$V(W) = A + B \log W.$$

则

$$V(W-c) = A + B\log(W-c).$$

Bellman 方程右边为

$$\log c + \beta A + \beta B \log(W-c).$$

对 $c$ 求一阶条件：

$$\frac{1}{c} - \frac{\beta B}{W-c} = 0.$$

因此

$$W - c = \beta B c,$$

即

$$c = \frac{W}{1+etaB}.$$

如果最优消费是固定比例 $c = (1-eta)W$，则需要

$$\frac{1}{1+etaB} = 1 - eta.$$

解得

$$1 + etaB = \frac{1}{1-eta},$$

所以

$$B = \frac{1}{1-eta}.$$

于是最优策略为

$$\boxed{c_t = (1-eta)W_t, \qquad W_{t+1} = \beta W_t.}$$

路径为

$$W_t = \beta^t W_0, \qquad c_t = (1-eta)\beta^t W_0.$$

### 7.2 经济解释

贴现因子 $\beta$ 越大，消费者越有耐心，消费比例 $1-\beta$ 越小，留下的资源比例 $\beta$ 越大。这个例子非常简单，但它展示了动态规划的核心力量：我们不直接求整条无穷维路径，而是求一个函数 $V(W)$ 和一个规则 $c(W)$。

## 8 随机动态规划

现实经济中状态转移往往含有不确定性。例如收入冲击、技术冲击、资产收益冲击。设随机冲击为 $z'$，当前状态为 $x$，控制为 $a$，下一期状态为

$$x' = F(x, a, z').$$

Bellman 方程变为

$$\boxed{V(x) = \max_{a \in A(x)} \left\{ u(x,a) + \beta \mathbb{E}[V(F(x,a,z')) \mid x,a] \right\}.}$$

其中期望是对下一期随机冲击取的。

## 8.1 Markov 状态

动态规划通常要求状态是 Markov state：给定当前状态和当前选择后，未来分布不再依赖更早历史。形式上，

$$\Pr(x_{t+1} \mid x_t, a_t, x_{t-1}, a_{t-1}, \dots) = \Pr(x_{t+1} \mid x_t, a_t).$$

经济含义是：状态变量必须足够完整，能概括所有对未来有用的历史信息。

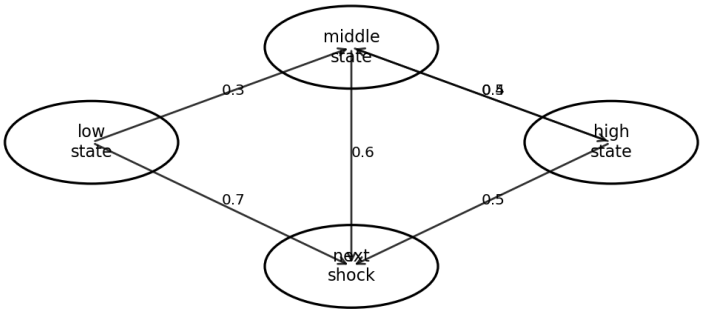



图 5：随机动态规划中，当前状态和决策决定下一期状态的概率分布。

**Worked Example**

**例 2：随机收入储蓄问题**

家庭拥有资产 $a$，收入状态为 $y$。选择消费 $c$ 和下一期资产 $a'$：

$$c + a' = (1+r)a + y,$$

$$a' \geq \underline{a}.$$

收入按 Markov chain 转移：

$$\Pr(y' = y_j \mid y = y_i) = \pi_{ij}.$$

Bellman 方程为

$$V(a,y)=\max_{a'\ge \underline{a}}\left\{u((1+r)a+y-a')+\beta\sum_{y'}\pi(y'\mid y)V(a',y')\right\}.$$

如果借贷约束绑定，Euler 方程会变成不等式；这正是动态规划比简单 Euler 方法更方便的地方。

## 9 借贷约束与不等式：为什么 Bellman 更方便

考虑消费储蓄问题：

$$V(a)=\max_{a'\ge \underline{a}}\{u((1+r)a+y-a')+\beta V(a')\}.$$

如果内点，FOC 为

$$u'(c)=\beta V'(a').$$

结合 envelope condition，得到 Euler equation：

$$u'(c_t)=\beta(1+r)u'(c_{t+1}).$$

但如果借贷约束绑定，$a'=\underline{a}$，最优点是角点，FOC 不再等号成立，而是

$$u'(c_t)\ge\beta(1+r)\mathbb{E}_t u'(c_{t+1}).$$

直觉是：如果家庭想借更多但不能借，那么今天消费的边际效用会高于把资源转移到未来的贴现边际收益。

**常见误区**

动态规划处理不等式约束时特别自然，因为每个状态下直接在可行集合 $A(x)$ 上最大化。Euler 方程只对内点最优最干净；一旦出现 borrowing constraint、short-sale constraint、irreversible investment，Bellman 写法往往更稳健。

## 10 数值动态规划：从函数到网格

很多 Bellman 方程没有解析解，需要数值求解。最常见的初学方法是 value function iteration on a grid。

## 10.1 状态网格

设状态 $k$ 位于区间 $[\underline{k}, \bar{k}]$。选择网格：

$$\underline{k} = k_1 < k_2 < \cdots < k_N = \bar{k}.$$

在每个网格点上近似价值函数：

$$V(k_i) \approx V_i.$$

然后对每个 $k_i$，枚举可能的 $k'_j$，计算

$$u(f(k_i) + (1-\delta)k_i - k'_j) + \beta V_n(k'_j),$$

取最大值作为 $V_{n+1}(k_i)$。

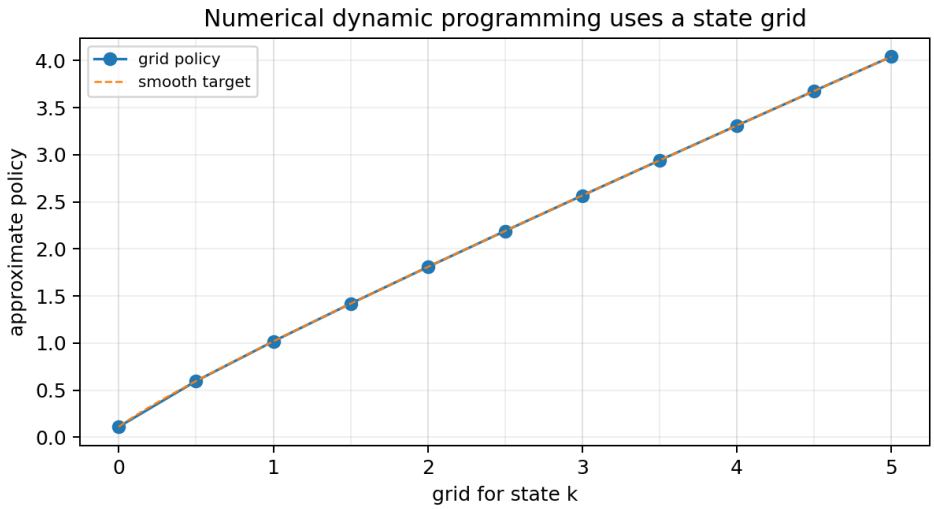



图 6: 数值动态规划用有限网格近似连续状态空间。网格越密，计算越精确但成本越高。

**方法模板**

**确定性增长模型的 value iteration 伪代码**

1. 建立资本网格 $k_1, \ldots, k_N$。
2. 初始化 $V_0(k_i) = 0$。
3. 对每个 $k_i$，枚举 $k'_j$，要求消费

$$c = f(k_i) + (1-\delta)k_i - k'_j$$

为正。

4. 计算

$$u(c) + \beta V_n(k'_j).$$

5. 取最大值，得到 $V_{n+1}(k_i)$ 和对应 policy $k' = h(k_i)$。

6. 重复直到价值函数变化足够小。

### 10.2 价值迭代、策略迭代和投影

数值动态规划有很多方法。最基础的是 value iteration，最直观但可能慢。policy iteration 先固定策略、求其价值，再改进策略，通常更快。projection/collocation 方法用基函数近似价值函数或策略函数，适合更高维问题。

**常见误区**

高维动态规划会遇到 curse of dimensionality。状态变量维度每增加一个，网格点数量可能指数级上升。例如每个维度取 100 个点，5 维状态就有 $100^5 = 10^{10}$ 个点，直接网格几乎不可行。这也是现代计算经济学、机器学习和近似动态规划的重要原因。

## 11 Bellman 方程的性质

在标准条件下，价值函数常有良好性质。

### 11.1 单调性

如果状态更多总是不会更差，例如财富更高、资本更多，则价值函数通常递增：

$$x_1 \geq x_0 \quad \Longrightarrow \quad V(x_1) \geq V(x_0).$$

经济直觉是：拥有更多资源时，至少可以模仿较少资源时的选择，还可能做得更好。

### 11.2 凹性

如果即时效用是凹的，技术和约束是凸的，价值函数通常也是凹的。凹性意味着边际价值递减：

$$V''(x) < 0.$$

这非常重要，因为它保证最优选择较为平滑，也有助于 Euler 方程的唯一性和数值求解。

### 11.3 包络性质

Bellman 方程中的 envelope condition 连接价值函数导数和影子价格。例如消费储蓄问题中：

$$V(a) = \max_{a'} \{u((1+r)a + y - a') + \beta V(a')\}.$$

在内点且可微时，

$$V'(a) = u'(c)(1+r).$$

这说明财富的边际价值等于当前消费边际效用乘以资产回报。

# 12 应用一：工作搜索模型

工作搜索是动态规划的经典应用。失业者每期收到工资 offer $w$，可以接受或拒绝。接受后获得永久工资 $w$；拒绝则得到失业补助 $b$，下期重新抽工资。

## 12.1 Bellman 方程

设 $V(w)$ 是看到工资 offer $w$ 时的价值，$U$ 是拒绝后的失业价值。接受价值为

$$A(w) = \frac{w}{1-\beta}.$$

拒绝价值为

$$R = b + eta\mathbb{E}V(w').$$

所以

$$V(w) = \max\left\{\frac{w}{1-\beta},\ b + eta\mathbb{E}V(w')\right\}.$$

## 12.2 保留工资 reservation wage

存在一个保留工资 $w^*$，使得当 $w \geq w^*$ 时接受，$w < w^*$ 时拒绝。临界点满足

$$\frac{w^*}{1-\beta} = b + eta\mathbb{E}V(w').$$

经济解释：如果工资足够高，接受工作的永久收益超过继续搜索的 option value。

**直觉盒子**

工作搜索模型展示了动态规划的一个重要概念：option value。拒绝当前 offer 的价值不是 $b$ 本身，而是 $b$ 加上未来可能遇到更好 offer 的价值。

# 13 应用二：企业进入退出

企业每期有生产率 $z$，选择继续经营或退出。继续经营得到利润 $\pi(z)$，下期生产率 $z'$ 按 Markov 过程转移；退出得到 scrap value $S$。

Bellman 方程为

$$V(z) = \max\left\{S,\ \pi(z) + \beta\mathbb{E}[V(z') \mid z]\right\}.$$

如果生产率低到一定程度，继续经营的 continuation value 低于退出价值，则企业退出。退出门槛 $z^*$ 满足

$$S = \pi(z^*) + \beta\mathbb{E}[V(z') \mid z^*].$$

### 13.1 经济含义

这个模型说明：企业是否退出不仅取决于当前利润，还取决于未来恢复的可能性。如果当前亏损但未来生产率可能改善，继续经营仍可能最优。

## 14 应用三：资产定价的 Bellman 视角

资产价格也可以看成动态规划。投资者持有资产，今天价格为 $p_t$，明天股利和价格为 $d_{t+1}+p_{t+1}$。在无套利和风险中性简化下：

$$p_t = \frac{1}{R}\mathbb{E}_t[d_{t+1}+p_{t+1}].$$

这和 Bellman 结构相似：今天价值等于明天 payoff 的贴现期望。更一般地，随机贴现因子为 $m_{t+1}$ 时，资产定价 Euler 方程是

$$\boxed{p_t = \mathbb{E}_t[m_{t+1}(d_{t+1}+p_{t+1})].}$$

这可以看成金融中的递归价值关系。

## 15 常见误区与纠正

**常见误区**

**误区 1：Bellman 方程只是另一个 Euler 方程。**
纠正：Euler 方程是最优路径的一阶条件；Bellman 方程是价值函数的不动点方程。Bellman 可以推出 Euler，但也能处理角点、随机性、离散选择和数值求解。

**误区 2：状态变量可以随便选。**
纠正：状态变量必须总结所有影响未来的信息。如果漏掉状态，问题就不满足 Markov 性，Bellman 方程会写错。

**误区 3：价值函数就是效用函数。**
纠正：效用函数 $u(c)$ 是当期收益；价值函数 $V(x)$ 是从状态 $x$ 出发的最大未来贴现收益。

**误区 4：$\beta<1$ 只是偏好假设。**
纠正：贴现因子也提供数学上的收敛和压缩性质。$\beta$ 越接近 1，经济主体越有耐心，同时数值迭代通常越慢。

**误区 5：数值动态规划只是编程。**
纠正：数值方法的背后仍是经济结构。网格、插值、约束、状态变量选择、收敛标准都会影响经济结论。

## 16 本章小结图谱

| 概念 | 公式 | 经济含义 |
|---|---|---|
| Bellman 方程 | $V(x)=\max_a\{u(x,a)+\beta V(x')\}$ | 当前收益加未来价值。 |
| 价值函数 | $V(x)$ | 从状态 $x$ 出发的最大未来贴现价值。 |
| 策略函数 | $g(x)=\arg\max\{\cdots\}$ | 给定状态时的最优决策规则。 |
| Bellman operator | $\mathcal{T}V$ | 把一个价值函数猜测更新成新的价值函数。 |
| 压缩映射 | $\|\mathcal{T}V-\mathcal{T}W\|\le\beta\|V-W\|$ | 保证唯一价值函数和迭代收敛。 |
| Euler 方程 | $u'(c_t)=\beta u'(c_{t+1})R_{t+1}$ | 相邻两期边际效用平衡。 |
| Markov 状态 | $P(x'\mid x,a)$ | 当前状态足以预测未来分布。 |
| Value iteration | $V_{n+1}=\mathcal{T}V_n$ | 动态规划的基础数值算法。 |

## 17 练习题

### A 组：基础概念

A1. 解释状态变量、控制变量、价值函数和策略函数的区别。

A2. 为什么无限期 stationary problem 的价值函数通常不带时间下标?

A3. 写出有限期 Bellman 递推式，并说明 backward induction 的含义。

A4. 解释为什么 Bellman 方程可以看成不动点方程。

A5. 什么是 Markov state? 为什么状态变量必须足够完整?

### B 组：计算题

B1. 两期 cake-eating：最大化 $\log c_0+\beta\log c_1$，约束 $c_0+c_1=W$。求 $c_0,c_1$。

B2. 无限期 cake-eating：已知最优规则 $c_t=(1-\beta)W_t$，求 $W_t,c_t$ 的路径。

B3. 对 Bellman 方程

$$V(a)=\max_{a'}\{\log((1+r)a+y-a')+\beta V(a')\},$$

写出内点一阶条件和 envelope condition。

B4. 确定性增长模型 Bellman 方程为

$$V(k)=\max_{k'}\{u(f(k)+(1-\delta)k-k')+\beta V(k')\}.$$

推导 FOC 和 envelope condition。

B5. 若 Bellman operator 是 $\beta = 0.9$ 的压缩映射，初始误差上界为 10，迭代 20 次后的误差上界是多少?

### C 组：综合推导

C1. 从确定性增长模型 Bellman 方程推导 Euler 方程。

C2. 证明无限期 cake-eating with log utility 的价值函数形如 $A + B \log W$，并求 $B$。

C3. 说明 value iteration 为什么会收敛到唯一价值函数。

C4. 在随机收入储蓄问题中，写出含 Markov 收入状态的 Bellman 方程，并解释期望项。

C5. 对工作搜索模型，写出接受价值、拒绝价值和保留工资条件。

### D 组：经济金融应用

D1. 借贷约束为什么会使 Euler 方程从等号变成不等式? 用经济语言解释。

D2. 在企业退出模型中，为什么当前亏损企业仍可能选择继续经营?

D3. 资产价格递归 $p_t = E_t[m_{t+1}(d_{t+1} + p_{t+1})]$ 和 Bellman 方程有什么共同结构?

D4. 为什么 $\beta$ 越接近 1，value iteration 通常越慢?

D5. 高维动态规划为什么会出现 curse of dimensionality? 给一个数字例子。

### E 组：挑战题

E1. 证明 Bellman operator 的 Blackwell sufficient conditions 直觉：monotonicity 和 discounting 如何导向 contraction? 只要求经济数学解释，不要求完全严格证明。

E2. 设价值函数凹且约束集凸，说明为什么 policy function 往往更容易唯一。

E3. 在随机储蓄模型中，如果效用满足 Inada 条件但存在借贷下限，解释为什么最优消费函数可能在约束附近出现 kink。

E4. 对有限状态 Markov 决策问题，说明 Bellman 方程如何变成有限个非线性方程。

# 18 练习题参考答案

## A 组答案

A1. 状态变量是从过去带到现在并影响未来的变量；控制变量是当前可选择的变量；价值函数是从某状态出发可获得的最大未来贴现收益；策略函数是把状态映射到最优控制的规则。

A2. 因为 stationary infinite-horizon problem 的环境不随日历时间变化。从第 0 期看未来和从第 10 期看未来的结构相同，只要状态一样，价值就一样，因此写 $V(x)$ 而不是 $V_t(x)$。

A3. 有限期递推为

$$V_t(x) = \max_{a \in A(x)} \{u(x,a) + \beta V_{t+1}(F(x,a))\}.$$

backward induction 是从最后一期先求，再一步步往前求。

A4. 定义 Bellman operator:

$$(\mathcal{T}V)(x) = \max_a \{u(x,a) + \beta V(F(x,a))\}.$$

Bellman 方程是 $V = \mathcal{T}V$，所以是一个函数空间上的不动点问题。

A5. Markov state 是能概括所有影响未来分布的信息的状态变量。若状态漏掉重要历史信息，则仅凭当前状态无法决定下一期分布，Bellman 方程就不成立。

## B 组答案

B1. 问题为

$$\max_{c_0, c_1} \log c_0 + \beta \log c_1$$

subject to $c_0 + c_1 = W$。令 $c_1 = W - c_0$，FOC:

$$\frac{1}{c_0} = \frac{\beta}{W - c_0}.$$

因此

$$W - c_0 = \beta c_0,$$

得

$$c_0 = \frac{W}{1+\beta}, \qquad c_1 = \frac{\beta W}{1+\beta}.$$

B2. 若 $c_t = (1-\beta)W_t$，则

$$W_{t+1} = W_t - c_t = \beta W_t.$$

因此

$$W_t = \beta^t W_0,$$

$$c_t = (1-\beta)\beta^t W_0.$$

B3. 消费为

$$c = (1+r)a + y - a'.$$

内点 FOC：

$$-u'(c) + \beta V'(a') = 0.$$

在 log 情况下 $u'(c) = 1/c$。Envelope condition：

$$V'(a) = u'(c)(1+r).$$

B4. FOC：

$$-u'(c) + \beta V'(k') = 0,$$

即

$$u'(c) = \beta V'(k').$$

Envelope condition：

$$V'(k) = u'(c)[f'(k) + 1 - \delta].$$

B5. 误差上界为

$$10 \cdot 0.9^{20} \approx 1.216.$$

## C 组答案

C1. 由 FOC：

$$u'(c_t) = \beta V'(k_{t+1}).$$

由下一期 envelope：

$$V'(k_{t+1}) = u'(c_{t+1})[f'(k_{t+1}) + 1 - \delta].$$

合并得

$$u'(c_t) = \beta u'(c_{t+1})[f'(k_{t+1}) + 1 - \delta].$$

C2. 猜测 $V(W) = A + B\log W$。Bellman 方程右边为

$$\log c + \beta A + \beta B \log(W - c).$$

FOC：

$$\frac{1}{c} = \frac{\beta B}{W - c}.$$

最优规则意味着 $c = (1-\beta)W$，所以

$$\frac{1}{1+\beta B} = 1 - \beta.$$

解得

$$B = \frac{1}{1-\beta}.$$

C3. 若 Bellman operator 是压缩映射，则 Banach fixed point theorem 保证唯一不动点，并且从任意初始函数开始迭代都会收敛到该不动点。贴现因子 $\beta < 1$ 是误差缩小的核心原因。

C4. 状态为 $(a, y)$，控制为 $a'$，消费为 $c = (1+r)a + y - a'$。Bellman 方程为

$$V(a,y) = \max_{a' \geq \underline{a}} \left\{ u((1+r)a + y - a') + \beta \sum_{y'} \pi(y' \mid y) V(a', y') \right\}.$$

期望项表示：下期收入有多个可能状态，未来价值要按转移概率加权平均。

C5. 接受价值为

$$A(w) = \frac{w}{1-\beta}.$$

拒绝价值为

$$R = b + \beta \mathbb{E} V(w').$$

Bellman 方程为

$$V(w) = \max\{A(w), R\}.$$

保留工资 $w^*$ 满足

$$\frac{w^*}{1-\beta} = b + \beta \mathbb{E} V(w').$$

## D 组答案

D1. 如果借贷约束绑定，家庭想把更多未来资源转移到今天但做不到。因此今天消费的边际效用可能高于储蓄带来的贴现未来边际效用，Euler 条件变成不等式。

D2. 因为继续经营有 option value。当前利润为负不代表未来利润一直为负；若生产率可能恢复，继续经营的期望 continuation value 可能超过退出价值。

D3. 二者都具有递归结构：今天价值等于当前 payoff 加上贴现未来价值。资产定价中 payoff 是股利和未来价格；Bellman 中 payoff 是当前效用和未来价值。

D4. value iteration 的误差大致按 $\beta^n$ 收缩。$\beta$ 越接近 1，$\beta^n$ 下降越慢，因此迭代收敛越慢。

D5. 若每个状态维度取 100 个网格点，1 维是 100 点，3 维是 $100^3 = 1,000,000$ 点，5 维是 $100^5 = 10,000,000,000$ 点。维度增加导致计算量爆炸，这就是 curse of dimensionality。

### E 组答案

E1. Monotonicity 表示如果 $V \geq W$，则 $\mathcal{T}V \geq \mathcal{T}W$，因为未来价值更高不会降低最大化结果。Discounting 表示给价值函数整体加常数 $c$，Bellman 更新只增加 $\beta c$，即

$$\mathcal{T}(V + c) = \mathcal{T}V + \beta c.$$

二者合起来说明 Bellman operator 会把函数之间的最大差距按 $\beta$ 缩小，从而形成 contraction。

E2. 如果目标函数对控制变量严格凹、可行集合凸，则最大化问题通常有唯一解。价值函数凹和即时收益凹会让 Bellman 右边保持凹性，因此 policy function 更容易唯一。

E3. 借贷约束附近，如果最优 $a'$ 被卡在下限 $\underline{a}$，策略函数从“内点 Euler 决定”切换为“约束边界决定”。这种规则切换会导致消费函数斜率发生变化，出现 kink。

E4. 如果状态只有有限个 $x_1, \ldots, x_N$，价值函数就是有限个未知数 $V_1, \ldots, V_N$。Bellman 方程对每个状态给出一个方程：

$$V_i = \max_{a \in A_i} \left\{ u(x_i, a) + \beta \sum_j P_{ij}(a) V_j \right\}.$$

因此原来的函数方程变成有限维非线性方程组。

## 19 本章到下一章的连接

Bellman 方程是离散时间动态优化的核心工具。下一章会进入连续时间最优控制 optimal control 与 Hamiltonian 方法。两者思想相通：

$$\text{Bellman: } V(x) = \max_a \{u(x, a) + \beta V(x')\},$$

$$\text{Hamiltonian: } \mathcal{H} = u(x, a) + \lambda f(x, a).$$

Bellman 的价值函数导数 $V'(x)$ 和 Hamiltonian 中的 costate variable $\lambda$ 本质上都表示状态变量的影子价值。掌握这一点，后面 Ramsey 模型、资产定价、HJB 方程和连续时间金融都会更容易理解。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 23
# 连续时间最优控制与 Hamiltonian

Chapter 21 讲了离散和有限期动态优化的 Euler 方程，Chapter 22 讲了 Bellman 方程和动态规划。本章进入连续时间最优控制 optimal control。它研究的问题是：当状态变量按微分方程运动，决策者可以选择控制变量时，怎样选择一整条控制路径，使折现效用、利润或福利最大？本章的核心工具是 Hamiltonian、costate variable、Pontryagin 最大值原理、当前值/现值写法和横截条件。你可以把它看成动态优化的“微分方程版一阶条件”。

math4econ personalized notes
2026

# 目录

**Chapter 23 总览** **2**

**1 本章为什么重要** **2**
1.1 从 Euler 方程到 Hamiltonian . . . . . 2
1.2 从 Bellman 方程到最优控制 . . . . . 3

**2 标准最优控制问题** **3**
2.1 有限期问题 . . . . . 3
2.2 为什么不能只做逐点最大化 . . . . . 4

**3 现值 Hamiltonian 与 Pontryagin 条件** **4**
3.1 现值 Hamiltonian . . . . . 4
3.2 三个条件的经济含义 . . . . . 5

**4 当前值 Hamiltonian** **5**

**5 从变分直觉推导 Hamiltonian 条件** **6**

**6 横截条件 Transversality Condition** **7**
6.1 典型形式 . . . . . 7
6.2 TVC 的经济解释 . . . . . 8

**7 Ramsey-Cass-Koopmans 模型** **8**
7.1 社会规划者问题 . . . . . 8
7.2 一阶条件 . . . . . 9
7.3 Euler 方程 . . . . . 9
7.4 相图与鞍路径 . . . . . 10

**8 不可再生资源与 Hotelling 规则** **10**
8.1 模型 . . . . . 11

**9 投资调整成本与 Tobin's q** **12**
9.1 模型 . . . . . 12
9.2 一阶条件 . . . . . 13

**10 当前值与现值写法对照** **14**

**11 一般求解模板** **14**

**12 Worked Examples** **15**

**13 常见误区与纠正** **17**

**14 术语表** **18**

**15 本章总结图谱** **18**

**16 练习题** **19**

A 组：基础计算 . . . . . 19
B 组：概念判断 . . . . . 19
C 组：混合推导 . . . . . 19
D 组：经济金融应用 . . . . . 20
E 组：挑战 / 证明题 . . . . . 20

**17 练习题参考答案** **21**

A 组答案 . . . . . 21
B 组答案 . . . . . 21
C 组答案 . . . . . 22
D 组答案 . . . . . 23
E 组答案 . . . . . 23

**18 后续学习连接** **24**

# Chapter 23 总览

| | |
|---|---|
| **本章主题** | 连续时间最优控制 optimal control, Hamiltonian, Pontryagin maximum principle, costate variable。 |
| **核心问题** | 状态变量 $x(t)$ 按 $\dot{x}=f(x,u,t)$ 运动，控制变量 $u(t)$ 由决策者选择。如何选择整条路径 $u(t)$，最大化 $\int_0^T e^{-\rho t}F(x,u,t)\,\mathrm{d}t+e^{-\rho T}S(x(T))$? |
| **数学工具** | Hamiltonian、状态方程、伴随方程、控制一阶条件、当前值 Hamiltonian、现值 Hamiltonian、横截条件、相图、saddle path。 |
| **经济应用** | Ramsey 增长模型、不可再生资源开采、投资调整成本、Tobin's $q$、最优储蓄、财政政策、资产定价 Euler 方程。 |
| **后续连接** | Hamilton-Jacobi-Bellman 方程、连续时间资产定价、随机控制、宏观 saddle-path stability、动态一般均衡。 |

**直觉盒子**

最优控制的思想可以用一句话概括：**今天的选择既带来今天的收益，也改变明天的状态。**例如多消费一点会提高今天效用，但减少明天资本；多开采一点资源会增加今天利润，但减少未来库存；多投资一点会增加未来资本，但今天要付出调整成本。Hamiltonian 就是把“今天的收益”和“通过状态变量传递到未来的价值”放在同一个式子里。

## 1 本章为什么重要

很多经济金融模型都是连续时间动态优化。与 Chapter 10 的 ODE 不同，这里不是给定 $\dot{x}=f(x)$ 后看路径怎样走，而是决策者主动选择控制变量 $u(t)$，让状态变量沿最优路径走。典型问题包括：

- 家庭选择消费路径 $c(t)$，资本或财富 $k(t)$ 随储蓄变化；
- 社会规划者选择消费和投资，最大化代际福利；
- 资源所有者选择开采路径，使资源租金现值最大；
- 企业选择投资路径，在资本积累和调整成本之间权衡；
- 政府选择税收、债务或公共资本路径，最大化社会福利。

### 1.1 从 Euler 方程到 Hamiltonian

有限期离散时间问题中，我们经常通过变分或 Lagrange 方法得到 Euler 方程。例如

$$u'(c_t)=\beta(1+r)u'(c_{t+1}).$$

连续时间中，对应的关系往往变成微分方程。例如 Ramsey 模型中消费增长满足

$$\frac{\dot{c}}{c} = \frac{1}{\theta}(f'(k) - \delta - \rho).$$

Hamiltonian 方法给出这些方程的系统推导方式。

### 1.2 从 Bellman 方程到最优控制

Chapter 22 的 Bellman 方程把动态问题写为

$$V(x) = \max_{u}\{r(x,u) + \beta V(x')\}.$$

连续时间版本的 Bellman 思想最终会导向 HJB 方程。最优控制中的 costate variable $\lambda(t)$ 与动态规划中的价值函数导数 $V_x$ 有深刻关系：

$$\lambda(t) \approx V_x(x(t), t).$$

也就是说，伴随变量不是神秘符号，而是状态变量的边际价值 shadow value。

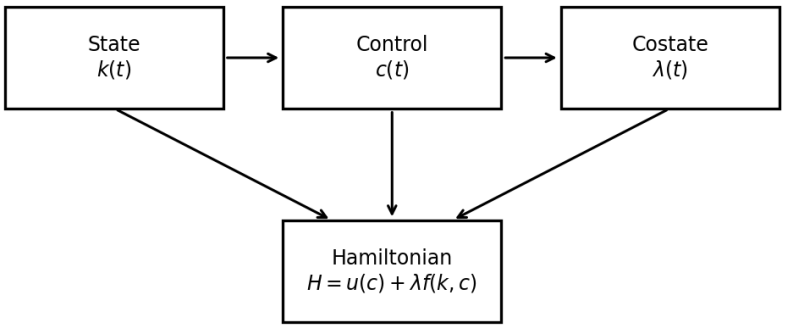



图 1：Hamiltonian 把状态、控制和伴随变量放在同一个一阶条件系统中。

## 2 标准最优控制问题

### 2.1 有限期问题

先考虑一个有限期问题：

$$\max_{u(t)} \int_0^T e^{-\rho t} F(x(t), u(t), t)\,\mathrm{d}t + e^{-\rho T} S(x(T))$$

subject to

$$\dot{x}(t) = f(x(t), u(t), t), \qquad x(0) = x_0.$$

这里：

- $x(t)$ 是状态变量 state variable；
- $u(t)$ 是控制变量 control variable；
- $F$ 是即时收益 flow payoff；
- $S(x(T))$ 是终端价值 terminal value；
- $\rho$ 是折现率 discount rate；
- $f$ 是状态转移方程 state equation。

**定义 / 定理**

**状态变量与控制变量.** 状态变量是由过去积累而来的“存量”，通常不能瞬间自由选择，例如资本、财富、资源存量、债务存量。控制变量是当期可以选择的“流量”或决策，例如消费、投资、开采量、税率、努力水平。

### 2.2 为什么不能只做逐点最大化

如果没有状态方程，问题会变成每个时刻单独最大化 $F(x, u, t)$。但状态方程使今天的 $u(t)$ 改变未来的 $x(s)$, $s > t$。因此不能只看今天收益，必须把未来状态价值纳入当期决策。

例如 Ramsey 模型中，消费 $c$ 增加会提高当前效用 $u(c)$，但状态方程

$$\dot{k} = f(k) - c - \delta k$$

说明更多消费会减少资本积累，降低未来产出。因此最优条件必然包含一个“资本的边际价值”。这就是 costate variable。

## 3 现值 Hamiltonian 与 Pontryagin 条件

### 3.1 现值 Hamiltonian

对有限期折现问题，定义现值 Hamiltonian present-value Hamiltonian：

$$\boxed{H^{PV}(x, u, \mu, t) = e^{-\rho t} F(x, u, t) + \mu f(x, u, t).}$$

其中 $\mu(t)$ 是现值 costate variable。

**定义 / 定理**

**Pontryagin 最大值原理，现值版本.** 若 $u^*(t), x^*(t)$ 是内点最优路径，则存在伴随变量 $\mu(t)$，使得:

$$H_u^{PV}(x^*, u^*, \mu, t) = 0,$$

$$\dot{x}^*(t) = H_\mu^{PV} = f(x^*, u^*, t),$$

$$\dot{\mu}(t) = -H_x^{PV}.$$

若终端状态自由且终端价值为 $e^{-\rho T}S(x(T))$，则

$$\mu(T) = e^{-\rho T}S'(x(T)).$$

若没有终端价值且 $x(T)$ 自由，则 $\mu(T) = 0$。

### 3.2 三个条件的经济含义

- $H_u = 0$: 当期控制变量的边际收益与它通过状态变量造成的边际未来价值损失相等。
- $\dot{x} = H_\mu$: 状态方程本身。
- $\dot{\mu} = -H_x$: 状态变量的影子价值如何随时间变化。

**直觉盒子**

如果 $x$ 是资本，$\mu$ 就是多一单位资本在现值意义下的价值。如果资本当前更有生产力，则 $H_x$ 大，资本的影子价值会受到影响。伴随方程不是机械公式，而是在说“状态变量的影子价格必须按无套利方式演化”。

## 4 当前值 Hamiltonian

经济学里更常用当前值 current-value Hamiltonian, 因为它把折现因子从即时收益中拿掉，表达更干净。

定义当前值 costate variable

$$\lambda(t) = e^{\rho t}\mu(t),$$

即

$$\mu(t) = e^{-\rho t}\lambda(t).$$

当前值 Hamiltonian 定义为

$$\boxed{H(x, u, \lambda, t) = F(x, u, t) + \lambda f(x, u, t).}$$

**定义 / 定理**

**Pontryagin 条件，当前值版本.** 对自治型折现问题

$$\max_{u(t)} \int_0^T e^{-\rho t} F(x,u)\,\mathrm{d}t + e^{-\rho T} S(x(T))$$

subject to $\dot{x} = f(x,u)$，当前值 Hamiltonian 为

$$H = F(x,u) + \lambda f(x,u).$$

内点最优条件为

$$H_u = 0,$$

$$\dot{x} = H_\lambda = f(x,u),$$

$$\dot{\lambda} = \rho\lambda - H_x.$$

终端条件为

$$\lambda(T) = S'(x(T))$$

若终端状态自由且无终端价值，则 $\lambda(T) = 0$。

**常见误区**

现值和当前值最容易混淆。现值伴随变量 $\mu$ 满足 $\dot{\mu} = -H_x^{PV}$。当前值伴随变量 $\lambda$ 满足 $\dot{\lambda} = \rho\lambda - H_x$。多出来的 $\rho\lambda$ 来自把折现因子吸收到 costate variable 里。

## 5 从变分直觉推导 Hamiltonian 条件

这一节不做完全严格证明，但给出经济学生应掌握的推导直觉。

考虑

$$\max_{u(t)} \int_0^T e^{-\rho t} F(x,u)\,\mathrm{d}t$$

subject to

$$\dot{x} = f(x,u), \qquad x(0) = x_0.$$

引入时间路径上的 Lagrange multiplier $\mu(t)$，把约束写进目标：

$$\int_0^T \left[ e^{-\rho t} F(x,u) + \mu(t)\left(f(x,u) - \dot{x}\right) \right] \mathrm{d}t.$$

对 $u$ 的变动给出

$$e^{-\rho t} F_u + \mu f_u = 0,$$

这就是 $H_u^{PV} = 0$。

对 $x$ 的变动，需要处理 $-\mu\dot{x}$。分部积分：

$$\int_0^T -\mu(t)\dot{x}(t)\,\mathrm{d}t = [-\mu(t)x(t)]_0^T + \int_0^T \dot{\mu}(t)x(t)\,\mathrm{d}t.$$

因此关于 $x$ 的内点项给出

$$e^{-\rho t}F_x + \mu f_x + \dot{\mu} = 0,$$

即

$$\dot{\mu} = -(e^{-\rho t}F_x + \mu f_x) = -H_x^{PV}.$$

边界项则给出终端条件。

**直觉盒子**

这段推导说明：Hamiltonian 条件其实就是“无限多个时点上的 Lagrange 条件”。区别是状态约束含有导数 $\dot{x}$，所以分部积分会产生 costate 的动态方程和终端条件。

# 6 横截条件 Transversality Condition

有限期问题通常有明确终点 $T$。无限期问题更常见于宏观和金融：

$$\max_{u(t)} \int_0^\infty e^{-\rho t}F(x,u)\,\mathrm{d}t.$$

这时没有固定终端期，但仍然需要排除不合理路径。横截条件 TVC 就是做这件事。

## 6.1 典型形式

若 $x(t)$ 是非负资产或资本，当前值 costate 是 $\lambda(t)$，一个常见 TVC 是

$$\boxed{\lim_{t\to\infty} e^{-\rho t}\lambda(t)x(t) = 0.}$$

含义是：到无限远未来时，不能留下有正现值的资产而不使用。否则说明你本可以稍微减少储蓄、增加当前消费，从而提高福利。

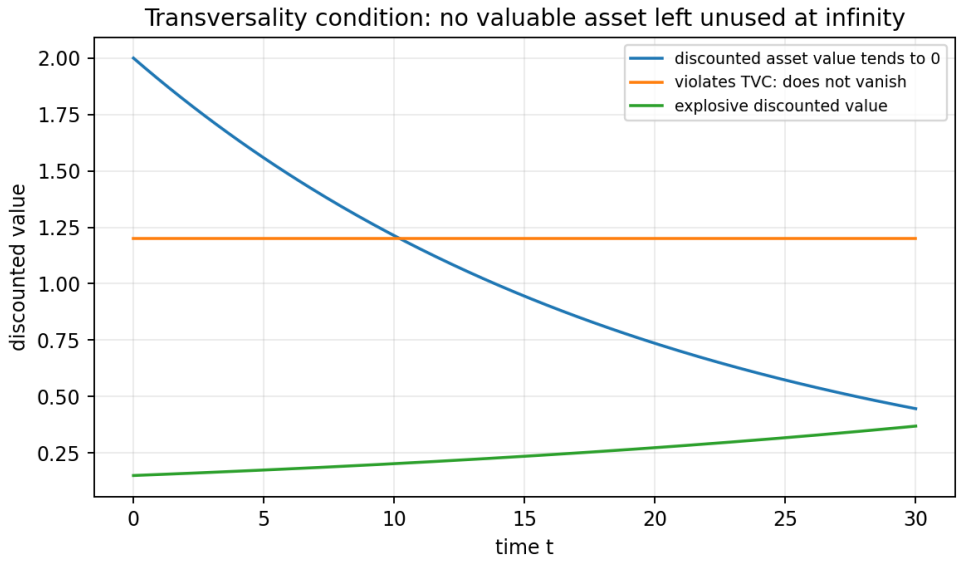



图 2: 横截条件排除“无限远未来仍留下正现值资产”的路径。

### 6.2 TVC 的经济解释

TVC 经常被误解成一个纯数学尾项条件。其实它是动态优化中的“无浪费条件”和“无泡沫条件”。

- 对家庭储蓄问题，TVC 排除永远积累财富却不消费的路径。
- 对资产定价，TVC 排除爆炸性泡沫项。
- 对 Ramsey 模型，TVC 选择 saddle path，而不是任意满足局部方程的路径。

## 7 Ramsey-Cass-Koopmans 模型

### 7.1 社会规划者问题

考虑最经典的连续时间增长模型：

$$\max_{c(t)} \int_0^{\infty} e^{-\rho t} u(c(t))\,\mathrm{d}t$$

subject to

$$\dot{k} = f(k) - c - \delta k,$$

$$k(0) = k_0,$$

其中 $u'(c) > 0, u''(c) < 0,\ f'(k) > 0, f''(k) < 0$。

当前值 Hamiltonian 为

$$H = u(c) + \lambda[f(k) - c - \delta k].$$

## 7.2 一阶条件

对 $c$ 求一阶条件：

$$H_c = u'(c) - \lambda = 0.$$

所以

$$\boxed{\lambda = u'(c).}$$

资本的状态方程：

$$\boxed{\dot{k} = f(k) - c - \delta k.}$$

伴随方程：

$$\dot{\lambda} = \rho\lambda - H_k.$$

而

$$H_k = \lambda[f'(k) - \delta].$$

所以

$$\dot{\lambda} = \lambda[\rho - f'(k) + \delta].$$

## 7.3 Euler 方程

由 $\lambda = u'(c)$，对时间求导：

$$\dot{\lambda} = u''(c)\dot{c}.$$

因此

$$u''(c)\dot{c} = u'(c)[\rho - f'(k) + \delta].$$

两边除以 $u'(c)$：

$$\frac{u''(c)c}{u'(c)}\frac{\dot{c}}{c} = \rho - f'(k) + \delta.$$

定义相对风险厌恶或跨期替代弹性倒数

$$\theta = -\frac{u''(c)c}{u'(c)} > 0.$$

则

$$\boxed{\frac{\dot{c}}{c} = \frac{1}{\theta}[f'(k) - \delta - \rho].}$$

这就是 Ramsey 模型的消费 Euler 方程。

**直觉盒子**

如果资本净回报 $f'(k) - \delta$ 高于主观折现率 $\rho$，未来回报高，社会愿意推迟消费，消费增长率为正。如果资本回报低于折现率，则当前消费相对更有吸引力，消费增长率为负。

### 7.4 相图与鞍路径

Ramsey 系统由两条微分方程组成：

$$\dot{k} = f(k) - c - \delta k,$$

$$\frac{\dot{c}}{c} = \frac{1}{\theta}[f'(k) - \delta - \rho].$$

$\dot{k} = 0$ locus 为

$$c = f(k) - \delta k.$$

$\dot{c} = 0$ locus 为

$$f'(k) = \rho + \delta.$$

由于 $f''(k) < 0$，这个方程通常确定唯一 $k^*$。

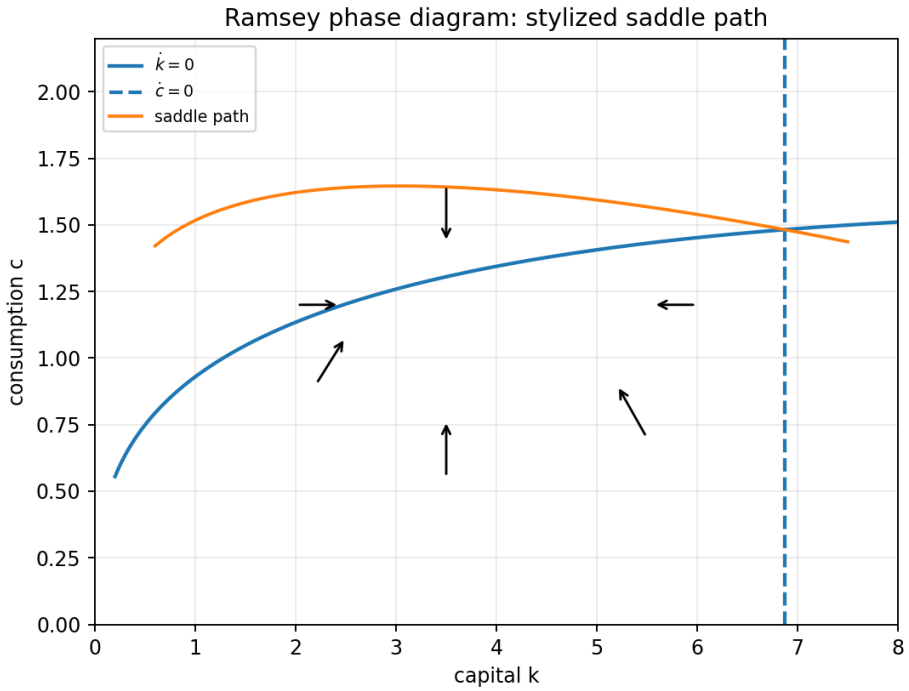



图 3: Ramsey 模型的典型相图。只有在 saddle path 上的初始消费选择才能收敛到稳态。

**常见误区**

Ramsey 模型中 $k(0)$ 给定，但 $c(0)$ 是可选择的跳变量 jump variable。不是所有 $c(0)$ 都可行或最优。过高的初始消费会耗尽资本，过低的初始消费会过度积累资本并违反 TVC。最优路径是穿过给定 $k(0)$ 的 saddle path。

## 8 不可再生资源与 Hotelling 规则

## 8.1 模型

资源所有者拥有初始资源库存 $S_0$，选择开采量 $q(t)$，资源库存满足

$$\dot{S}(t) = -q(t), \qquad S(0) = S_0.$$

假设价格 $p(t)$ 外生给定，单位开采成本为常数 $mc$，目标是最大化利润现值：

$$\max_{q(t)} \int_0^T e^{-rt}[p(t) - mc]q(t)\,\mathrm{d}t$$

subject to $\dot{S} = -q$, $S(t) \geq 0$。

现值 Hamiltonian：

$$H^{PV} = e^{-rt}[p(t) - mc]q + \mu(-q).$$

内点 FOC：

$$e^{-rt}[p(t) - mc] - \mu = 0.$$

所以

$$\mu = e^{-rt}[p(t) - mc].$$

伴随方程：

$$\dot{\mu} = -H_S = 0.$$

因此 $\mu$ 为常数。于是

$$e^{-rt}[p(t) - mc] = \text{constant}.$$

这意味着

$$\boxed{p(t) - mc \text{ grows at rate } r.}$$

即

$$\boxed{\frac{\dot{p}(t)}{p(t) - mc} = r}$$

在成本不变时成立。这就是 Hotelling rule。

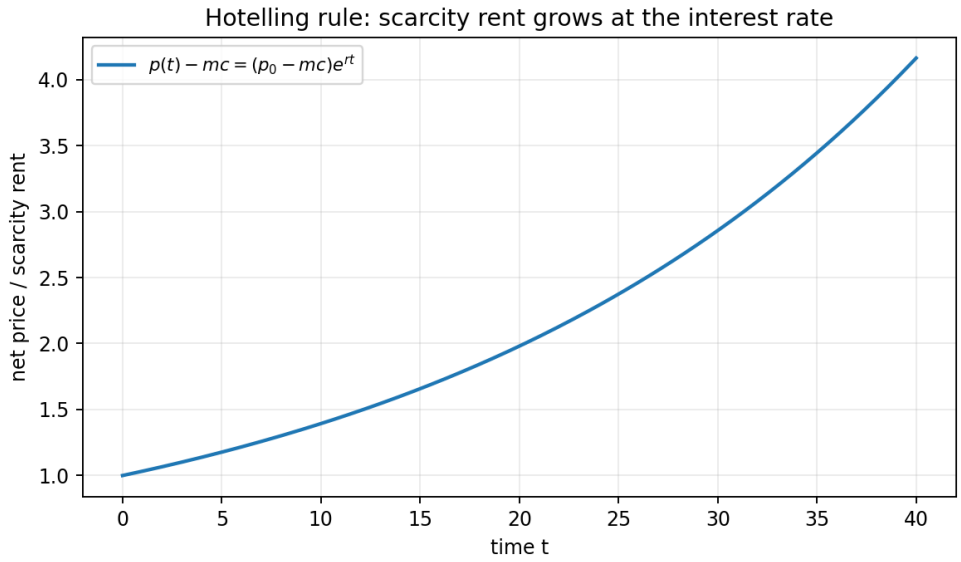



图 4: Hotelling 规则：资源稀缺租金按利率增长。

**直觉盒子**

资源留在地下也是一种资产。如果今天不开采，未来还能开采。因此资源的净价格或 scarcity rent 必须像金融资产一样以利率增长。若增长慢于利率，所有者更愿意今天开采并投资金融资产；若增长快于利率，则更愿意推迟开采。

# 9 投资调整成本与 Tobin's q

## 9.1 模型

企业选择投资 $I(t)$，资本积累为

$$\dot{K} = I - \delta K.$$

利润流为

$$\Pi(K) - I - \frac{\phi}{2} I^2,$$

其中 $\phi > 0$ 是调整成本参数。企业最大化

$$\max_{I(t)} \int_0^\infty e^{-rt} \left[ \Pi(K) - I - \frac{\phi}{2} I^2 \right] \mathrm{d}t.$$

当前值 Hamiltonian：

$$H = \Pi(K) - I - \frac{\phi}{2} I^2 + q(I - \delta K).$$

这里 $q$ 是资本的当前值影子价格，也可以理解为 marginal Tobin's q。

## 9.2 一阶条件

对 $I$ 求导:

$$H_I = -1 - \phi I + q = 0.$$

所以

$$\boxed{I = \frac{q-1}{\phi}.}$$

资本积累:

$$\boxed{\dot{K} = \frac{q-1}{\phi} - \delta K.}$$

伴随方程:

$$\dot{q} = rq - H_K.$$

而

$$H_K = \Pi'(K) - q\delta.$$

所以

$$\boxed{\dot{q} = (r+\delta)q - \Pi'(K).}$$

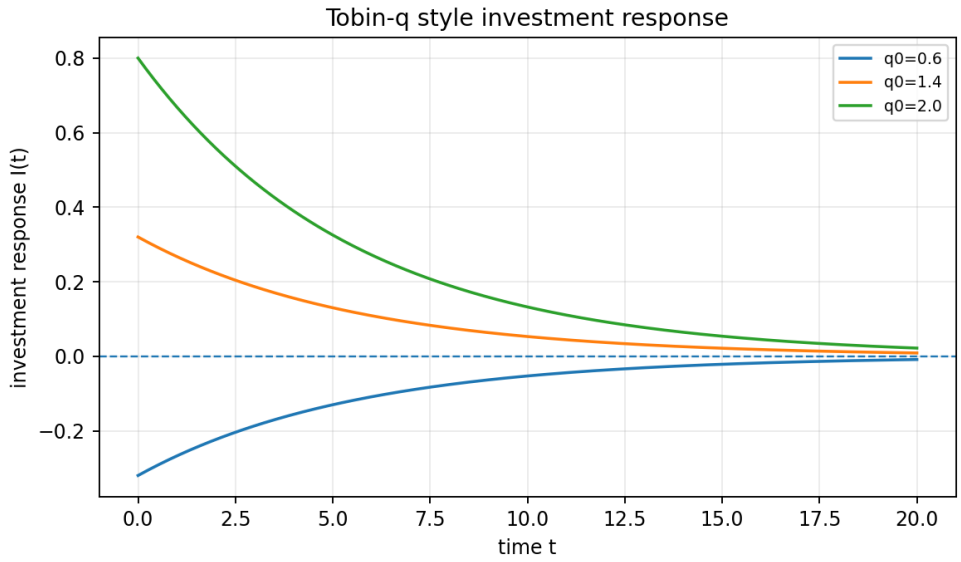



图 5: 调整成本模型中，$q > 1$ 鼓励投资，$q < 1$ 抑制投资。

**直觉盒子**

如果 $q > 1$，多装一单位资本的价值超过购买成本，企业投资为正。如果 $q < 1$，资本边际价值低于成本，企业减少投资或不投资。调整成本使投资不会瞬间跳到无限大，而是随 $q - 1$ 平滑变化。

## 10 当前值与现值写法对照

| 项目 | 现值写法 | 当前值写法 |
|---|---|---|
| Hamiltonian | $H^{PV} = e^{-\rho t}F + \mu f$ | $H = F + \lambda f$ |
| Costate variable | $\mu$ | $\lambda = e^{\rho t}\mu$ |
| 控制 FOC | $H_u^{PV} = 0$ | $H_u = 0$ |
| 状态方程 | $\dot{x} = H_\mu^{PV}$ | $\dot{x} = H_\lambda$ |
| 伴随方程 | $\dot{\mu} = -H_x^{PV}$ | $\dot{\lambda} = \rho\lambda - H_x$ |
| 常见用途 | 数学推导更直接 | 经济解释更清楚 |

**方法模板**

**实用建议：**经济学中优先用当前值 Hamiltonian。因为 flow payoff $F$ 不带折现项，$\lambda$ 直接解释为“当前价值单位下的状态变量边际价值”。但做严谨推导时，要记得当前值伴随方程多出 $\rho\lambda$。

## 11 一般求解模板

**方法模板**

**连续时间最优控制五步法**

给定

$$\max_{u(t)} \int_0^\infty e^{-\rho t}F(x,u)\,\mathrm{d}t$$

subject to

$$\dot{x} = f(x,u), \qquad x(0) = x_0.$$

步骤如下：

1. 写当前值 Hamiltonian:
$$H = F(x,u) + \lambda f(x,u).$$
2. 控制变量 FOC:
$$H_u = 0.$$
3. 状态方程:
$$\dot{x} = f(x,u).$$
4. 伴随方程:
$$\dot{\lambda} = \rho\lambda - H_x.$$

5. 加上初始条件和横截条件:

$$x(0)=x_0,\qquad \lim_{t\to\infty} e^{-\rho t}\lambda(t)x(t)=0.$$

## 12 Worked Examples

**Worked Example**

**例 1：最简单的消费储蓄模型**

设家庭最大化

$$\int_0^{\infty} e^{-\rho t}\ln c(t)\,\mathrm{d}t$$

subject to

$$\dot a = ra + y - c,$$

其中 $a$ 是资产，$y$ 是常数收入，$r$ 是利率。当前值 Hamiltonian 为

$$H=\ln c+\lambda(ra+y-c).$$

控制 FOC:

$$H_c=\frac{1}{c}-\lambda=0.$$

所以

$$\lambda=\frac{1}{c}.$$

伴随方程:

$$\dot\lambda=\rho\lambda-H_a.$$

而

$$H_a=\lambda r.$$

所以

$$\dot\lambda=(\rho-r)\lambda.$$

由 $\lambda=1/c$，有

$$\frac{\dot\lambda}{\lambda}=-\frac{\dot c}{c}.$$

因此

$$-\frac{\dot c}{c}=\rho-r.$$

即

$$\boxed{\frac{\dot c}{c}=r-\rho.}$$

经济解释：若市场回报 $r$ 高于主观折现率 $\rho$，家庭愿意推迟消费，消费增长；若 $r<\rho$，家庭更偏好当前消费，消费下降。

**Worked Example**

**例 2：线性二次调节问题**

考虑

$$\min_{u(t)} \int_0^{\infty} e^{-\rho t}\left[\frac{1}{2}x(t)^2+\frac{\gamma}{2}u(t)^2\right]\mathrm{d}t$$

subject to

$$\dot{x}=u.$$

为了用最大化形式，写作最大化负损失：

$$H=-\frac{1}{2}x^2-\frac{\gamma}{2}u^2+\lambda u.$$

FOC：

$$H_u=-\gamma u+\lambda=0.$$

所以

$$u=\frac{\lambda}{\gamma}.$$

伴随方程：

$$\dot{\lambda}=\rho\lambda-H_x=\rho\lambda+x.$$

状态方程：

$$\dot{x}=\frac{\lambda}{\gamma}.$$

这给出一个二维线性系统：

$$\begin{pmatrix}\dot{x}\\ \dot{\lambda}\end{pmatrix}=\begin{pmatrix}0 & 1/\gamma\\ 1 & \rho\end{pmatrix}\begin{pmatrix}x\\ \lambda\end{pmatrix}.$$

它通常有一个稳定方向和一个不稳定方向，最优路径由 TVC 选择稳定鞍路径。

**Worked Example**

**例 3：有限期终端条件**

考虑

$$\max_{u(t)} \int_0^{T}[x(t)-\frac{1}{2}u(t)^2]\,\mathrm{d}t$$

subject to

$$\dot{x}=u,\qquad x(0)=x_0,$$

且终端状态 $x(T)$ 自由，无终端价值。Hamiltonian 为

$$H = x - \frac{1}{2}u^2 + \lambda u.$$

FOC：

$$-u + \lambda = 0,$$

所以 $u = \lambda$。伴随方程：

$$\dot{\lambda} = -H_x = -1.$$

由于终端状态自由且无终端价值，横截条件为

$$\lambda(T) = 0.$$

因此

$$\lambda(t) = T - t.$$

所以

$$u(t) = T - t.$$

状态方程给出

$$\dot{x} = T - t.$$

积分：

$$x(t) = x_0 + Tt - \frac{1}{2}t^2.$$

这个例子说明终端条件可以从未来反向确定 costate，再决定控制路径。

## 13 常见误区与纠正

**常见误区**

**误区 1：把 $\lambda$ 当作普通 Lagrange multiplier，不理解其动态含义。**
静态乘子是约束放松一单位的边际价值；动态 costate 是状态变量在每个时点的边际价值，并且它自己必须满足一个动态方程。

**误区 2：忘记状态变量不能自由跳。**
资本、资产、资源库存通常是 predetermined state。控制变量可以跳，状态变量一般不能瞬间跳。

**误区 3：现值和当前值 Hamiltonian 混用。**
如果用当前值 Hamiltonian，伴随方程是 $\dot{\lambda} = \rho\lambda - H_x$；如果用现值 Hamiltonian，伴随方程是 $\dot{\mu} = -H_x^{PV}$。

**误区 4：只写 FOC，不写 TVC。**

无限期问题中，FOC 通常只能给出一族路径。TVC 用来排除非最优、泡沫或浪费路径。

**误区 5：看到 saddle path 就以为不稳定。**
Saddle path 的含义是：系统整体有稳定和不稳定方向。由于某些变量是跳变量，可以选择初始跳点进入稳定路径，因此模型可以有确定的最优收敛路径。

## 14 术语表

| 英文术语 | 中文译法 | 本章含义 |
|---|---|---|
| optimal control | 最优控制 | 在动态约束下选择控制路径。 |
| state variable | 状态变量 | 由过去累积决定, 通常不能瞬间跳。 |
| control variable | 控制变量 | 决策者当期可以选择的变量。 |
| Hamiltonian | 哈密顿量 | 当前收益加上状态变化的影子价值。 |
| costate variable | 伴随变量 | 状态变量的边际价值或影子价格。 |
| Pontryagin maximum principle | 庞特里亚金最大值原理 | 最优控制的一阶必要条件系统。 |
| transversality condition | 横截条件 | 无限期或自由终端问题的边界条件。 |
| current-value Hamiltonian | 当前值 Hamiltonian | 把折现因素吸收到 costate 中的写法。 |
| present-value Hamiltonian | 现值 Hamiltonian | 直接保留折现因子的写法。 |
| saddle path | 鞍路径 | 收敛到稳态的特殊路径。 |

## 15 本章总结图谱

| 对象 | 数学形式 | 经济含义 |
|---|---|---|
| 控制问题 | $\max \int e^{-\rho t} F(x,u)\,\mathrm{d}t$ | 选择整条动态路径。 |
| 状态方程 | $\dot{x} = f(x,u)$ | 当前决策如何改变未来状态。 |
| Hamiltonian | $H = F + \lambda f$ | 当期收益加未来状态价值。 |
| 控制 FOC | $H_u = 0$ | 控制变量的边际收益与边际未来成本相等。 |

| | | |
|---|---|---|
| 伴随方程 | $\dot{\lambda} = \rho\lambda - H_x$ | 状态变量影子价值的动态。 |
| TVC | $\lim e^{-\rho t}\lambda x = 0$ | 排除浪费、泡沫或爆炸路径。 |
| Ramsey Euler | $\dot{c}/c = (f'(k) - \delta - \rho)/\theta$ | 消费增长由资本净回报和耐心决定。 |
| Hotelling rule | $(\dot{p - mc})/(p - mc) = r$ | 稀缺租金按利率增长。 |
| Tobin's q | $I = (q-1)/\phi$ | 资本影子价值驱动投资。 |

## 16 练习题

### A 组：基础计算

A1. 给定 $H = \ln c + \lambda(ra + y - c)$，写出关于 $c$、$a$、$\lambda$ 的最优控制条件。

A2. 给定 $H = u(c) + \lambda[f(k) - c - \delta k]$，推导 Ramsey 模型的 $H_c = 0$ 和 $\dot{\lambda}$ 方程。

A3. 现值 Hamiltonian 为 $H^{PV} = e^{-\rho t}F(x,u) + \mu f(x,u)$。写出现值形式的三个条件。

A4. 当前值 Hamiltonian 为 $H = F(x,u) + \lambda f(x,u)$。写出当前值形式的三个条件。

A5. 若 $\lambda = u'(c)$ 且 $u(c) = c^{1-\theta}/(1-\theta)$，求 $\dot{\lambda}/\lambda$ 与 $\dot{c}/c$ 的关系。

A6. 若 $\lim_{t\to\infty} e^{-\rho t}\lambda(t)x(t) > 0$，解释为什么可能违反横截条件。

### B 组：概念判断

B1. 为什么 costate variable 可以解释为 shadow value?

B2. 当前值 Hamiltonian 和现值 Hamiltonian 的区别是什么?

B3. 为什么无限期问题不能只写一阶条件，还要写 TVC?

B4. Ramsey 模型中，为什么 $c$ 可以跳而 $k$ 通常不能跳?

B5. Hotelling 规则为什么可以理解为无套利条件?

### C 组：混合推导

C1. 对消费储蓄问题 $\max \int_0^\infty e^{-\rho t}\ln c\,\mathrm{d}t$ subject to $\dot{a} = ra + y - c$，推导消费增长率。

C2. 对 Ramsey 模型，设 $u(c) = \ln c$，推导 $\dot{c}/c = f'(k) - \delta - \rho$。

C3. 对资源开采问题，推导 $p(t) - mc$ 按利率增长。

C4. 对投资调整成本模型，推导 $I = (q-1)/\phi$。

C5. 给定 $\dot{x} = ax + bu$，收益 $F = -(x^2 + u^2)/2$，写出 Hamiltonian 和一阶条件。

### D 组：经济金融应用

D1. 用最优控制语言解释“多储蓄一单位资本”的收益和成本。

D2. 解释 Ramsey 模型中 $f'(k)-\delta>\rho$ 时消费为什么增长。

D3. 在资源模型中，如果预计资源净价格增长率低于利率，资源所有者应如何调整开采?

D4. 在 Tobin’s q 模型中，为什么 $q>1$ 会带来正投资?

D5. 解释 TVC 在资产定价中排除泡沫项的作用。

### E 组：挑战 / 证明题

E1. 从现值 Hamiltonian 出发，推导当前值伴随方程 $\dot{\lambda}=\rho\lambda-H_x$。

E2. 对 Ramsey 模型，说明稳态条件为什么是 $f'(k^*)=\rho+\delta$ 和 $c^*=f(k^*)-\delta k^*$。

E3. 对有限期问题，解释终端状态自由时为什么 costate 终端值与终端价值函数导数有关。

E4. 比较 Bellman 方程中的 $V_x$ 与 Hamiltonian 中的 $\lambda$。

# 17 练习题参考答案

## A 组答案

A1. 控制 FOC：

$$H_c = \frac{1}{c} - \lambda = 0.$$

状态方程：

$$\dot{a} = ra + y - c.$$

伴随方程：

$$\dot{\lambda} = \rho\lambda - H_a = \rho\lambda - r\lambda = (\rho - r)\lambda.$$

A2. $H_c = u'(c) - \lambda = 0$，所以 $\lambda = u'(c)$。又

$$H_k = \lambda[f'(k) - \delta].$$

当前值伴随方程为

$$\dot{\lambda} = \rho\lambda - \lambda[f'(k) - \delta] = \lambda[\rho - f'(k) + \delta].$$

A3. 现值形式：

$$H_u^{PV} = 0, \qquad \dot{x} = H_\mu^{PV} = f(x,u), \qquad \dot{\mu} = -H_x^{PV}.$$

A4. 当前值形式：

$$H_u = 0, \qquad \dot{x} = H_\lambda = f(x,u), \qquad \dot{\lambda} = \rho\lambda - H_x.$$

A5. 若

$$u(c) = \frac{c^{1-\theta}}{1-\theta},$$

则

$$u'(c) = c^{-\theta}.$$

因此 $\lambda = c^{-\theta}$。取对数：

$$\ln\lambda = -\theta \ln c.$$

对时间求导：

$$\frac{\dot{\lambda}}{\lambda} = -\theta\frac{\dot{c}}{c}.$$

A6. 若极限为正，说明无限远未来仍有正现值的状态资产未被使用。这通常意味着可以减少一点未来资产积累、增加当前消费，从而改善目标函数，因此不是最优路径。

## B 组答案

B1. 因为 costate 衡量状态变量增加一单位对目标函数的边际影响。例如资本多一单位能增加未来产出和消费，因此 $\lambda$ 是资本的边际价值。

B2. 现值 Hamiltonian 保留折现因子，costate 是现值单位；当前值 Hamiltonian 去掉折现因子，costate 是当前值单位。当前值伴随方程多出 $\rho\lambda$ 项。

B3. FOC 只给出局部动态关系，可能允许许多路径，包括爆炸路径或泡沫路径。TVC 用来排除不满足全局最优的路径。

B4. $k$ 是资本存量，由过去投资积累而来，短期不能瞬间跳变。$c$ 是控制变量，决策者可以在初始时刻选择不同消费水平。

B5. 资源留在地下相当于持有一项资产。若净价格增长率低于利率，应今天开采并投资金融资产；若高于利率，应推迟开采。均衡中净价格增长率等于利率。

## C 组答案

C1. Hamiltonian:

$$H = \ln c + \lambda(ra + y - c).$$

FOC: $1/c = \lambda$。伴随方程: $\dot{\lambda} = (\rho - r)\lambda$。由于 $\lambda = 1/c$，所以

$$-\frac{\dot{c}}{c} = \rho - r,$$

即

$$\frac{\dot{c}}{c} = r - \rho.$$

C2. 若 $u(c) = \ln c$，则 $u'(c) = 1/c$，相对风险厌恶系数 $\theta = 1$。Ramsey Euler 方程

$$\frac{\dot{c}}{c} = \frac{1}{\theta}[f'(k) - \delta - \rho]$$

变为

$$\frac{\dot{c}}{c} = f'(k) - \delta - \rho.$$

C3. 现值 Hamiltonian:

$$H^{PV} = e^{-rt}[p(t) - mc]q + \mu(-q).$$

FOC:

$$e^{-rt}[p(t) - mc] - \mu = 0.$$

因 $H_S = 0$，$\dot{\mu} = 0$。所以 $e^{-rt}[p(t) - mc]$ 为常数，即 $p(t) - mc$ 按 $r$ 增长。

C4. Hamiltonian:

$$H = \Pi(K) - I - \frac{\phi}{2}I^2 + q(I - \delta K).$$

FOC:

$$H_I = -1 - \phi I + q = 0.$$

因此

$$I = \frac{q - 1}{\phi}.$$

C5. Hamiltonian:

$$H = -\frac{1}{2}(x^2 + u^2) + \lambda(ax + bu).$$

FOC:

$$H_u = -u + b\lambda = 0,$$

所以

$$u = b\lambda.$$

状态方程：$\dot{x} = ax + bu$。伴随方程:

$$\dot{\lambda} = \rho\lambda - H_x = \rho\lambda - (-x + a\lambda) = x + (\rho - a)\lambda.$$

## D 组答案

D1. 多储蓄一单位资本的成本是少消费一单位，当前效用下降；收益是未来资本更高，产出更高，未来消费可能增加。最优时，二者按折现后的边际价值相等。

D2. 若 $f'(k) - \delta > \rho$，资本净回报高于主观折现率，推迟消费的回报较高，因此最优路径让消费随时间增长。

D3. 如果净价格增长率低于利率，地下资源作为资产的回报低于金融资产。所有者会增加当前开采，把收入投资到金融资产，直到净价格增长率被推高到利率。

D4. $q$ 是一单位资本的边际价值。若 $q > 1$，资本边际价值高于购买成本 1，投资有利可图，所以最优投资为正。

D5. 在资产定价中，差分方程可能允许基本面价格加泡沫项。TVC 或无泡沫条件要求远期价格的贴现值趋于零，从而排除自我实现的爆炸价格路径。

## E 组答案

E1. 现值 Hamiltonian:

$$H^{PV} = e^{-\rho t}F(x, u) + \mu f(x, u).$$

现值伴随方程为

$$\dot{\mu} = -H_x^{PV} = -e^{-\rho t}F_x - \mu f_x.$$

定义 $\lambda = e^{\rho t}\mu$，即 $\mu = e^{-\rho t}\lambda$。对 $\lambda$ 求导:

$$\dot{\lambda} = \rho e^{\rho t}\mu + e^{\rho t}\dot{\mu}.$$

代入 $\dot{\mu}$:

$$\dot{\lambda} = \rho\lambda + e^{\rho t}[-e^{-\rho t}F_x - e^{-\rho t}\lambda f_x] = \rho\lambda - F_x - \lambda f_x.$$

当前值 Hamiltonian 为 $H = F + \lambda f$，所以 $H_x = F_x + \lambda f_x$。因此

$$\boxed{\dot{\lambda} = \rho\lambda - H_x.}$$

E2. Ramsey 稳态要求 $\dot{c} = 0$ 和 $\dot{k} = 0$。由 Euler 方程

$$\frac{\dot{c}}{c} = \frac{1}{\theta}[f'(k) - \delta - \rho],$$

得 $\dot{c} = 0$ 时

$$f'(k^*) = \rho + \delta.$$

由资本积累方程

$$\dot{k} = f(k) - c - \delta k,$$

得 $\dot{k} = 0$ 时

$$c^* = f(k^*) - \delta k^*.$$

E3. 若终端目标含 $S(x(T))$，则终端状态多一单位带来的价值是 $S'(x(T))$。由于 costate 是状态变量边际价值，终端处必须满足 $\lambda(T) = S'(x(T))$。若无终端价值且终端状态自由，则 $S' = 0$，所以 $\lambda(T) = 0$。

E4. Bellman 方程中 $V_x$ 是状态变量对价值函数的边际贡献。Hamiltonian 方法中的 $\lambda$ 是状态变量的影子价值。在光滑条件下，二者本质相同：$\lambda(t) = V_x(x(t), t)$ 或当前值版本的相应转换。动态规划用函数方程表达，最优控制用微分方程的一阶条件表达。

## 18 后续学习连接

本章给出确定性连续时间最优控制。下一步自然是两条线:

- **HJB 方程:** 把 Hamiltonian 条件和动态规划统一起来，得到连续时间 Bellman 方程。
- **随机控制:** 在状态方程中加入 Brownian motion 或随机冲击, 进入连续时间金融、Merton portfolio problem 和 stochastic dynamic programming。

到这里，你已经拥有从静态优化到动态优化的基本工具链:

Lagrange/KKT $\to$ Envelope/Duality $\to$ Euler/Bellman $\to$ Hamiltonian/Optimal Control.

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 24
# HJB 方程与连续时间动态规划

Chapter 23 用 Hamiltonian 和 Pontryagin 条件处理连续时间最优控制。本章从另一个角度看同一类问题：动态规划。核心对象从控制路径变成价值函数 $V(x,t)$, 核心方程从 Hamiltonian 一阶条件变成 Hamilton-Jacobi-Bellman equation, HJB 方程。HJB 的思想是：最优价值等于“当前瞬间收益 + 状态变化带来的价值变化 + 未来价值”。它是连续时间宏观、金融、搜索、投资、资产定价、随机控制和强化学习式经济模型的共同语言。

math4econ personalized notes
2026

# 目录

**Chapter 24 总览** **2**

**1 本章为什么重要** **3**
1.1 Pontryagin 与 HJB 的差异 . . . . . 3
1.2 为什么经济金融离不开 HJB . . . . . 3

**2 从离散 Bellman 到连续时间 HJB** **3**
2.1 离散时间 Bellman 回顾 . . . . . 4
2.2 Taylor 展开推导 HJB . . . . . 4

**3 非平稳 HJB: 有终端时间的问题** **5**

**4 HJB 与 Hamiltonian 的对应关系** **6**
4.1 从 HJB 推出 costate 方程 . . . . . 6

**5 Verification theorem: 为什么解了 HJB 还要验证** **7**
5.1 证明直觉 . . . . . 8

**6 Worked Example 1: 线性状态、二次成本的 HJB** **8**
6.1 写 HJB . . . . . 8
6.2 猜二次价值函数 . . . . . 9

**7 Worked Example 2: 确定性消费储蓄问题** **10**
7.1 HJB 方程 . . . . . 10
7.2 经济含义 . . . . . 10
7.3 从 HJB 推 Euler 方程 . . . . . 11

**8 Worked Example 3: 企业投资与 Tobin's q** **11**
8.1 HJB . . . . . 12
8.2 Tobin's q 的含义 . . . . . 12

**9 HJB 的边界条件与状态约束** **13**
9.1 常见边界条件 . . . . . 13
9.2 状态约束 . . . . . 13

**10 HJB、PDE 与资产定价入口** **14**
10.1 确定性资产价格的 HJB/PDE 结构 . . . . . 14
10.2 随机版本的预告 . . . . . 14

**11 HJB 的数值直觉** **15**
11.1 有限差分方法的基本结构 . . . . . 15
11.2 与 Bellman value iteration 的关系 . . . . . 15

**12 价值函数的形状：单调性与凹性** **16**
12.1 为什么凹性重要 . . . . . 16

**13 常见误区与纠正** **16**

**14 本章总结图谱** **17**

**15 练习题** **17**
A 组：基础概念与公式 . . . . . 17
B 组：推导与计算 . . . . . 17
C 组：经济应用 . . . . . 18
D 组：挑战题 . . . . . 18

**16 练习题详细答案** **19**
A 组答案 . . . . . 19
B 组答案 . . . . . 19
C 组答案 . . . . . 20
D 组答案 . . . . . 21

**17 后续学习连接** **22**

# Chapter 24 总览

| | |
|---|---|
| **本章主题** | Hamilton-Jacobi-Bellman 方程，连续时间动态规划 continuous-time dynamic programming, verification theorem, stationary HJB。 |
| **核心问题** | 如果状态变量按 $\dot{x} = f(x,u,t)$ 运动，控制变量 $u$ 可选，价值函数 $V(x,t)$ 应满足什么微分方程？如何从 HJB 推出最优策略？如何证明猜出来的 $V$ 真的是价值函数？ |
| **数学工具** | 动态规划原理、短时间展开、Taylor 近似、HJB 方程、策略函数、verification theorem、平稳 HJB、边界条件、与 Hamiltonian/Pontryagin 的对应。 |
| **经济应用** | 连续时间消费储蓄、资产定价 PDE、企业投资与 Tobin's $q$、工作搜索、资源开采、随机控制入口、Merton 投资组合模型预告。 |
| **后续连接** | 随机过程、Ito 引理、Black-Scholes PDE、连续时间资产定价、宏观金融动态规划、强化学习与数值 HJB。 |

**直觉盒子**

HJB 的一句话直觉是：**站在状态 $x$，最优者在一个很短的时间 $\Delta t$ 内先拿到当前收益，再把状态推进一点点，然后继续拥有未来价值**。令 $\Delta t \to 0$，离散 Bellman 方程就变成一个关于 $V$ 的微分方程。

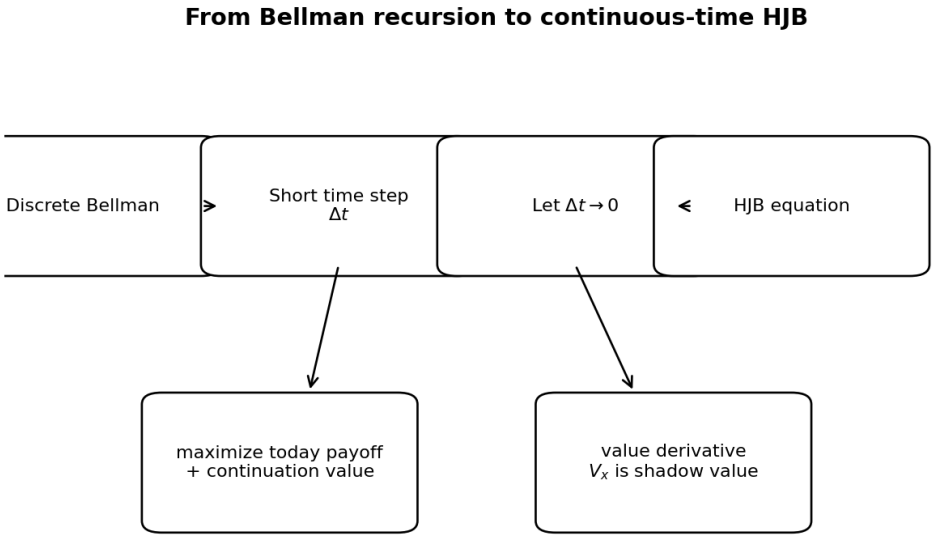



图 1：HJB 是 Bellman recursion 的连续时间极限。

# 1 本章为什么重要

Chapter 23 的 Pontryagin maximum principle 给出的是路径上的必要条件：状态方程、伴随方程、控制一阶条件、横截条件。它很像把 Lagrange multiplier 方法搬到无限维路径空间中。HJB 则从价值函数出发，强调每个状态下的最优选择。两种方法经常给出同样的经济结论，但视角不同。

## 1.1 Pontryagin 与 HJB 的差异

| | **Pontryagin / Hamiltonian** | **HJB / 动态规划** |
|---|---|---|
| 核心未知量 | 最优路径 $x(t), u(t)$ 与 costate $\lambda(t)$ | 价值函数 $V(x,t)$ 与策略函数 $u^*(x,t)$ |
| 局部对象 | Hamiltonian $\mathcal{H}$ | HJB 方程 |
| 经济直觉 | 伴随变量是状态的影子价格 | 价值函数导数是状态的边际价值 |
| 条件性质 | 常给出必要条件，也可在凹性下充分 | verification 可给充分性证明 |
| 适合场景 | 相图、解析路径、saddle path | 数值动态规划、随机控制、策略函数 |

## 1.2 为什么经济金融离不开 HJB

HJB 方程在以下问题中反复出现：

- 家庭连续时间消费储蓄：财富是状态，消费是控制；
- 企业投资：资本是状态，投资是控制，$V_K$ 对应 Tobin's $q$；
- 资产定价：价格或价值函数满足某种 PDE；
- 搜索与匹配：就业状态、工资报价、保留工资可以由 HJB 表示；
- 金融投资组合：财富动态带随机项，HJB 加上 Ito 引理得到 Merton 模型；
- 动态产业组织：企业状态、进入退出、投资、价格策略经常用 HJB/离散 Bellman 表达。

**常见误区**

HJB 不是“另一个公式”。它是一种建模语言。你应该学会从经济问题写出状态、控制、流收益、状态方程，然后把它们放进 HJB。比起机械解 PDE，更重要的是知道每一项的经济含义。

# 2 从离散 Bellman 到连续时间 HJB

## 2.1 离散时间 Bellman 回顾

离散时间动态规划通常写为

$$V(x) = \max_u \{r(x,u) + \beta V(x')\},$$

其中

$$x' = G(x,u).$$

如果时间间隔是 $\Delta t$，折现因子约为

$$\beta = e^{-\rho \Delta t} \approx 1 - \rho \Delta t.$$

在 $\Delta t$ 内获得的收益约为

$$r(x,u)\Delta t.$$

状态变化为

$$x(t+\Delta t) \approx x + f(x,u)\Delta t.$$

因此短时间 Bellman 方程可以写成

$$V(x) = \max_u \left\{ r(x,u)\Delta t + e^{-\rho \Delta t} V(x + f(x,u)\Delta t) \right\}.$$

## 2.2 Taylor 展开推导 HJB

对价值函数做一阶 Taylor 展开：

$$V(x + f\Delta t) = V(x) + V_x(x) f(x,u)\Delta t + o(\Delta t).$$

折现项

$$e^{-\rho \Delta t} = 1 - \rho \Delta t + o(\Delta t).$$

代入 Bellman 方程：

$$V(x) = \max_u \left\{ r\Delta t + [1 - \rho \Delta t] \left[ V + V_x f \Delta t + o(\Delta t) \right] \right\}.$$

忽略 $o(\Delta t)$ 和二阶小量，得到

$$V(x) = \max_u \{ V + r\Delta t + V_x f \Delta t - \rho V \Delta t \}.$$

两边减去 $V$，再除以 $\Delta t$：

$$0 = \max_u \{ r(x,u) + V_x(x) f(x,u) - \rho V(x) \}.$$

等价地,

$$\boxed{\rho V(x) = \max_u \{r(x,u) + V_x(x) f(x,u)\}.}$$

这就是最简单的平稳 HJB 方程。

**定义 / 定理**

**平稳 HJB 方程.** 对无限期、时间不显式出现的问题

$$\max_{u(t)} \int_0^\infty e^{-\rho t} r(x(t), u(t))\, \mathrm{d}t, \qquad \dot{x} = f(x,u),$$

价值函数若足够光滑，通常满足

$$\rho V(x) = \max_u \{r(x,u) + V_x(x) f(x,u)\}.$$

**直觉盒子**

HJB 左边 $\rho V$ 可以理解成“持有这个价值函数要求的机会成本”。右边由两部分组成：当前流收益 $r(x,u)$，以及状态变化带来的资本利得 $V_x f(x,u)$。最优控制就是在当前收益与未来价值变化之间权衡。

## 3 非平稳 HJB: 有终端时间的问题

如果问题有有限终点 $T$，价值函数依赖时间：

$$V(x,t) = \max_{u(s), s\in[t,T]} \int_t^T e^{-\rho(s-t)} r(x(s), u(s), s)\, \mathrm{d}s + e^{-\rho(T-t)} S(x(T)).$$

短时间展开会得到

$$\boxed{\rho V(x,t) = V_t(x,t) + \max_u \{r(x,u,t) + V_x(x,t) f(x,u,t)\}.}$$

等价地也常写为

$$\boxed{0 = \max_u \{r(x,u,t) + V_x f + V_t - \rho V\}.}$$

终端条件是

$$\boxed{V(x,T) = S(x).}$$

**方法模板**

**有限期 HJB 解题模板**

1. 确定状态 $x$、控制 $u$、流收益 $r(x,u,t)$、状态方程 $\dot{x} = f(x,u,t)$。
2. 写价值函数 $V(x,t)$。

3. 写 HJB：$\rho V = V_t + \max_u\{r + V_x f\}$。
4. 对 $u$ 求一阶条件，得到策略函数 $u^*(x,t)$。
5. 代回 HJB，得到关于 $V$ 的 PDE/ODE。
6. 使用终端条件 $V(x,T) = S(x)$ 或边界条件解出 $V$。

# 4 HJB 与 Hamiltonian 的对应关系

平稳 HJB：

$$\rho V(x) = \max_u\{r(x,u) + V_x(x)f(x,u)\}.$$

这和当前值 Hamiltonian 非常像。Chapter 23 中当前值 Hamiltonian 是

$$H(x,u,\mu) = r(x,u) + \mu f(x,u),$$

其中 $\mu$ 是当前值 costate。HJB 告诉我们：

$$\boxed{\mu = V_x(x).}$$

也就是说，costate variable 是价值函数对状态的边际导数。

## 4.1 从 HJB 推出 costate 方程

设最优策略为 $u^*(x)$，HJB 在最优处为

$$\rho V(x) = r(x,u^*) + V_x(x)f(x,u^*).$$

对 $x$ 求导，并利用 envelope 思想忽略 $u^*$ 的间接影响：

$$\rho V_x = r_x + V_{xx}f + V_x f_x.$$

沿最优路径，令

$$\mu(t) = V_x(x(t)).$$

则

$$\dot{\mu} = V_{xx}\dot{x} = V_{xx}f.$$

所以

$$\rho\mu = r_x + \dot{\mu} + \mu f_x.$$

整理得到

$$\boxed{\dot{\mu} = \rho\mu - r_x - \mu f_x.}$$

这正是当前值 costate 方程。

**直觉盒子**

Pontryagin 方法从“路径约束的乘子”得到 $\mu$；HJB 方法从“价值函数的边际价值”得到 $V_x$。两者在最优路径上对应：

$$\boxed{\mu(t) = V_x(x(t), t).}$$

这也是为什么 HJB 特别适合解释 shadow value。

## 5 Verification theorem: 为什么解了 HJB 还要验证

HJB 看起来给出了最优条件，但数学上仍需要验证：一个光滑函数 $W(x)$ 如果满足 HJB，是否一定就是价值函数 $V(x)$？某个由 HJB 推出的策略是否真的最优？Verification theorem 做的就是这件事。

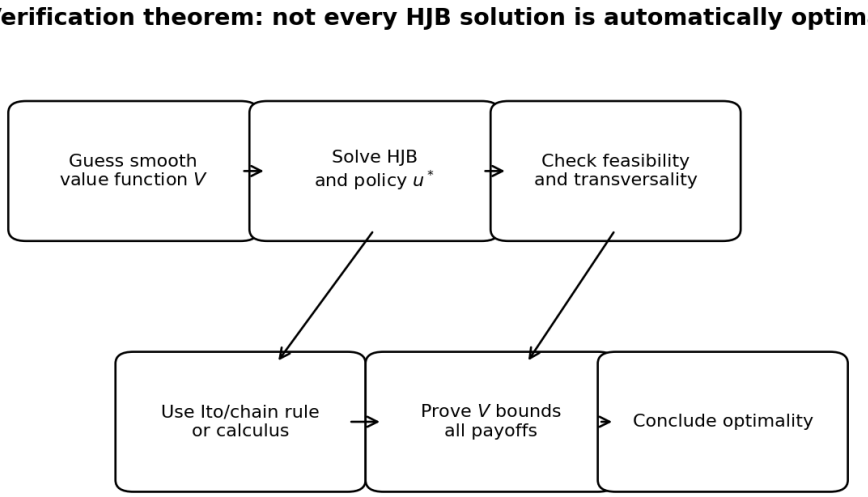



图 2: Verification theorem 的逻辑：猜测、求解、检查、证明。

**定义 / 定理**

**Verification theorem 的直观版本.** 假设存在一个足够光滑的函数 $W(x)$，满足 HJB

$$\rho W(x) \geq r(x,u) + W_x(x) f(x,u)$$

对所有可行 $u$ 成立，并且某个策略 $u^*(x)$ 使等号成立。同时满足适当边界条件和横截条件。那么 $W$ 是价值函数，$u^*$ 是最优策略。

### 5.1 证明直觉

对任意可行路径 $x(t)$，由 HJB 不等式：

$$r(x,u) \le \rho W(x) - W_x(x)\dot{x}.$$

两边乘以 $e^{-\rho t}$：

$$e^{-\rho t} r(x,u) \le e^{-\rho t}\rho W(x) - e^{-\rho t} W_x(x)\dot{x}.$$

注意

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-\rho t}W(x(t))\right) = e^{-\rho t}W_x\dot{x} - \rho e^{-\rho t}W.$$

所以

$$e^{-\rho t} r(x,u) \le -\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-\rho t}W(x(t))\right).$$

从 0 积分到 $T$：

$$\int_0^T e^{-\rho t} r(x,u)\,\mathrm{d}t \le W(x_0) - e^{-\rho T}W(x(T)).$$

如果横截项满足

$$\liminf_{T\to\infty} e^{-\rho T}W(x(T)) \ge 0,$$

则任意策略的收益不超过 $W(x_0)$。如果策略 $u^*$ 处处使 HJB 等号成立，并且横截项正好消失，则收益等于 $W(x_0)$，所以 $u^*$ 最优。

**常见误区**

很多教材直接写 HJB 并求解，但真正严谨时必须检查可行性、边界条件、横截条件和 concavity。否则可能得到满足一阶条件但不是最优的候选解，尤其在非凹问题、多重均衡和状态约束问题中。

## 6 Worked Example 1：线性状态、二次成本的 HJB

考虑一个简单调节问题：

$$\max_{u(t)} \int_0^\infty e^{-\rho t}\left[-\frac{1}{2}x(t)^2 - \frac{1}{2}u(t)^2\right]\mathrm{d}t$$

subject to

$$\dot{x} = u.$$

这里决策者想让状态 $x$ 靠近 0，但调整速度 $u$ 有成本。

### 6.1 写 HJB

平稳 HJB 是

$$\rho V(x) = \max_u \left\{-\frac{1}{2}x^2 - \frac{1}{2}u^2 + V_x(x)u\right\}.$$

对 $u$ 求一阶条件：

$$-u + V_x = 0.$$

所以

$$\boxed{u^*(x) = V_x(x).}$$

因为目标是最大化负成本，价值函数通常为负且 $V_x$ 的符号会推动 $x$ 回到 0。

## 6.2 猜二次价值函数

猜

$$V(x) = -\frac{1}{2}Px^2,$$

其中 $P > 0$。则

$$V_x = -Px.$$

最优控制为

$$u^*(x) = -Px.$$

这说明若 $x > 0$，选择负的 $u$ 把状态拉回；若 $x < 0$，选择正的 $u$。

代入 HJB：

$$\rho\left(-\frac{1}{2}Px^2\right) = -\frac{1}{2}x^2 - \frac{1}{2}(P^2x^2) + (-Px)(-Px).$$

右边为

$$-\frac{1}{2}x^2 - \frac{1}{2}P^2x^2 + P^2x^2 = -\frac{1}{2}x^2 + \frac{1}{2}P^2x^2.$$

所以

$$-\frac{1}{2}\rho Px^2 = \frac{1}{2}(P^2 - 1)x^2.$$

化简：

$$P^2 + \rho P - 1 = 0.$$

正根为

$$\boxed{P = \frac{-\rho + \sqrt{\rho^2 + 4}}{2}.}$$

因此

$$\boxed{V(x) = -\frac{1}{2}Px^2, \qquad u^*(x) = -Px.}$$

**直觉盒子**

这个例子展示了 HJB 的典型操作：猜价值函数形式，求一阶条件，代回 HJB，解出未知参数。线性-二次问题 linear-quadratic regulator, LQR 是宏观、金融和控制理论里最常见的可解析 HJB 模型之一。

# 7 Worked Example 2：确定性消费储蓄问题

考虑无限期消费储蓄问题：

$$\max_{c(t)} \int_0^{\infty} e^{-\rho t} u(c(t))\,\mathrm{d}t$$

subject to

$$\dot{a} = ra + y - c,$$

其中 $a$ 是财富，$r$ 是利率，$y$ 是劳动收入，$c$ 是消费。

## 7.1 HJB 方程

平稳 HJB：

$$\boxed{\rho V(a) = \max_c \{u(c) + V_a(a)(ra + y - c)\}.}$$

对 $c$ 的一阶条件：

$$u'(c) - V_a(a) = 0.$$

因此

$$\boxed{u'(c^*(a)) = V_a(a).}$$

## 7.2 经济含义

$V_a(a)$ 是多一单位财富带来的边际价值。最优消费使当前边际效用等于财富的边际价值：

$$\text{marginal utility of consuming now} = \text{marginal value of saving one more unit.}$$

如果 $V_a$ 很高，说明财富稀缺，应该少消费；如果 $V_a$ 较低，说明财富充裕，可以多消费。

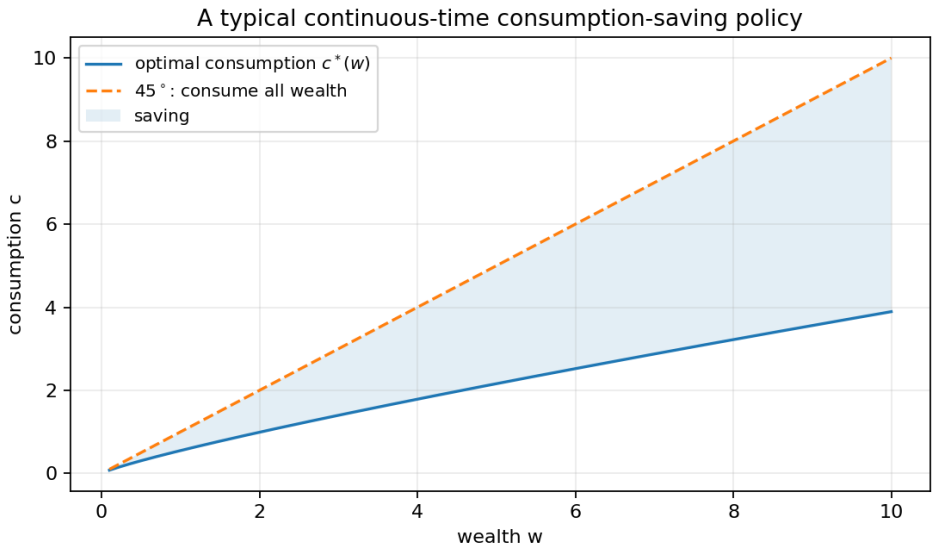



图 3：典型消费策略函数：财富越高，消费越高，但可能低于 45 度线，差额为储蓄。

### 7.3 从 HJB 推 Euler 方程

由 HJB 的 envelope 条件:

$$\rho V_a = u'(c)\cdot 0 + V_{aa}(ra + y - c) + V_a r,$$

更直接写为

$$\rho V_a = V_{aa}\dot{a} + rV_a.$$

利用一阶条件 $V_a = u'(c)$，沿路径求导:

$$\dot{V}_a = u''(c)\dot{c}.$$

又

$$\dot{V}_a = V_{aa}\dot{a}.$$

所以

$$\rho u'(c) = u''(c)\dot{c} + ru'(c).$$

整理:

$$u''(c)\dot{c} = (\rho - r)u'(c).$$

若效用为 CRRA:

$$u(c) = \frac{c^{1-\theta}}{1-\theta},$$

则

$$\frac{u'(c)}{-u''(c)c} = \frac{1}{\theta}.$$

因此

$$\boxed{\frac{\dot{c}}{c} = \frac{r-\rho}{\theta}.}$$

这就是连续时间消费 Euler 方程。

## 8 Worked Example 3: 企业投资与 Tobin's q

企业最大化利润现值:

$$\max_{I(t)} \int_0^\infty e^{-rt}\left[\pi(K(t)) - I(t) - \frac{\phi}{2}I(t)^2\right]\,\mathrm{d}t$$

subject to

$$\dot{K} = I - \delta K.$$

## 8.1 HJB

平稳 HJB:

$$rV(K) = \max_I \left\{ \pi(K) - I - \frac{\phi}{2} I^2 + V_K(K)(I - \delta K) \right\}.$$

对 $I$ 求一阶条件:

$$-1 - \phi I + V_K = 0.$$

因此

$$\boxed{I^*(K) = \frac{V_K(K) - 1}{\phi}.}$$

若存在不可逆投资约束 $I \geq 0$，则

$$\boxed{I^*(K) = \max\left\{0, \frac{V_K(K) - 1}{\phi}\right\}.}$$

## 8.2 Tobin's q 的含义

定义

$$q = V_K(K).$$

它是资本的边际价值。最优投资条件变成

$$q = 1 + phiI.$$

若 $q > 1$，资本边际价值高于购买一单位资本的成本，企业愿意投资；若 $q < 1$ 且投资不可逆，则企业可能选择 $I = 0$。

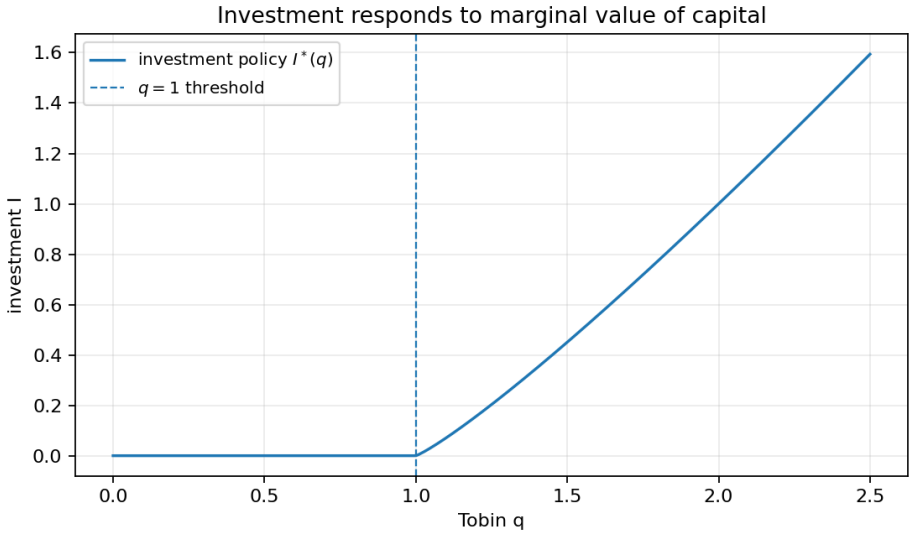



图 4: Tobin's $q$ 与投资策略：资本边际价值越高，投资越强。

**直觉盒子**

HJB 视角下，Tobin's $q$ 不是外生指标，而是价值函数的导数 $V_K$。它衡量多一单位资本给企业带来的价值。投资的一阶条件就是“边际安装成本 = 资本边际价值 - 资本购买成本”。

# 9 HJB 的边界条件与状态约束

HJB 是微分方程。微分方程只有方程本身通常不够，还需要边界条件 boundary condition 或终端条件。

## 9.1 常见边界条件

| 场景 | 条件形式 | 经济含义 |
|---|---|---|
| 有限期问题 | $V(x,T)=S(x)$ | 到终点后获得终端价值。 |
| 破产边界 | $V(0)=\bar{V}$ | 财富为 0 时退出或破产。 |
| 资源耗尽 | $V(0)=0$ 或 salvage value | 资源用尽后无继续开采价值。 |
| 自由边界 | value matching / smooth pasting | 退出、违约、行权、搜索接受工资。 |
| 横截条件 | $\lim e^{-\rho t}V_x x=0$ | 排除不合理爆炸路径。 |

## 9.2 状态约束

如果状态必须满足 $x\geq 0$，那么在边界 $x=0$，可行控制必须保证状态不越界。例如财富不能为负时，

$$\dot{a}=ra+y-c\geq 0 \quad \text{at } a=0.$$

这会限制消费：

$$c\leq y.$$

因此 HJB 不只是内点最大化问题，还要把可行控制集写清楚：

$$\rho V(a)=\max_{c\in\Gamma(a)}\{u(c)+V_a(ra+y-c)\}.$$

其中 $\Gamma(a)$ 是状态 $a$ 下的可行控制集合。

**常见误区**

在有借贷约束、不可逆投资、违约边界、退出选择时，HJB 的一阶条件经常不能直接使用。要先判断约束是否绑定，再写对应的 KKT/HJB 条件。

# 10 HJB、PDE 与资产定价入口

当状态是时间和资产价格，价值函数可能满足偏微分方程 PDE。即使暂时不学随机过程，也可以先理解确定性版本。

## 10.1 确定性资产价格的 HJB/PDE 结构

假设某资产或项目的状态 $x$ 按

$$\dot{x} = g(x)$$

演化，持有者每期获得现金流 $\pi(x)$，折现率为 $r$。价值函数满足

$$\boxed{rV(x) = \pi(x) + V_x(x)g(x).}$$

这其实是没有控制变量的 HJB。左边是资本化要求回报，右边是现金流加资本利得。

## 10.2 随机版本的预告

如果状态满足随机过程

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t,$$

那么 HJB 会多出二阶项：

$$\boxed{\rho V(x) = \max_u \left\{ r(x,u) + V_x \mu(x,u) + \frac{1}{2} V_{xx} \sigma(x,u)^2 \right\}.}$$

这个二阶项来自 Ito 引理。下一部分如果进入金融数学或随机过程，就会从这里走向 Black-Scholes PDE、Merton portfolio problem 和连续时间资产定价。

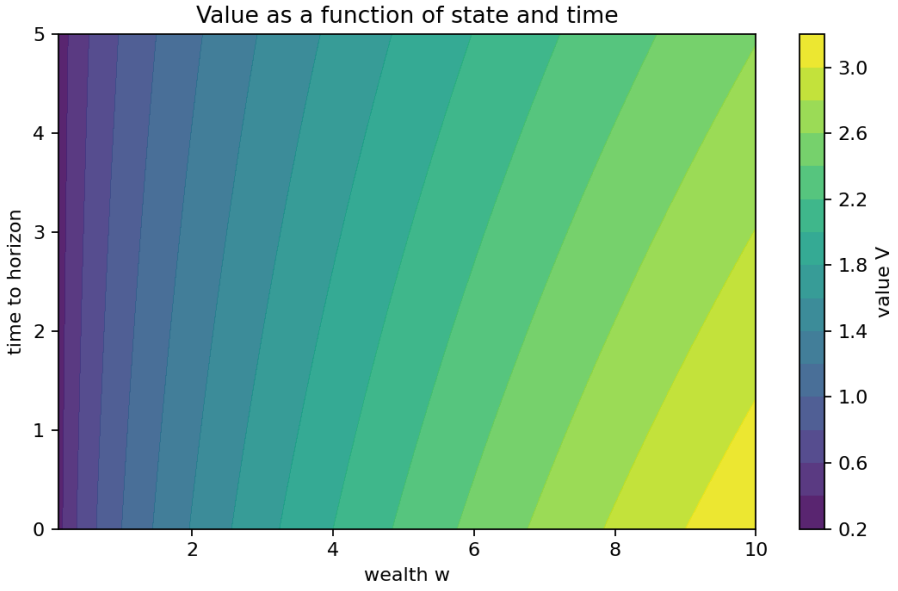



图 5: 有限期问题中，价值函数通常同时依赖状态和时间。HJB 是关于 $V(x,t)$ 的 PDE。

# 11 HJB 的数值直觉

很多 HJB 方程没有解析解，需要数值方法。最基本思路是把状态空间离散化，然后近似导数。

## 11.1 有限差分方法的基本结构

以平稳 HJB 为例：

$$\rho V(x) = \max_u \{r(x,u) + V_x(x) f(x,u)\}.$$

在网格点 $x_i$ 上，用差分近似导数：

$$V_x(x_i) \approx \frac{V(x_{i+1}) - V(x_i)}{\Delta x}$$

或者

$$V_x(x_i) \approx \frac{V(x_i) - V(x_{i-1})}{\Delta x}.$$

选择前向还是后向差分，通常要根据状态漂移方向 $f(x,u)$ 决定，这叫 upwind scheme 的直觉。

## 11.2 与 Bellman value iteration 的关系

连续时间 HJB 可以通过离散时间近似来解：

$$V^{new}(x_i) = \max_u \left\{ r(x_i,u)\Delta t + e^{-\rho \Delta t} V^{old}(x_i + f(x_i,u)\Delta t) \right\}.$$

这和 Chapter 22 的 value iteration 完全一致。差别只是连续时间需要更小的时间步和导数近似。

**方法模板**

**数值 HJB 的最小流程**

1. 设定状态网格 $x_1, \dots, x_N$。
2. 给定初始猜测 $V^0$。
3. 在每个网格点求控制最大化，得到 $u^n(x_i)$。
4. 用差分近似 $V_x$，更新 HJB。
5. 迭代直到 $V$ 和策略函数收敛。
6. 检查边界条件和策略是否可行。

# 12 价值函数的形状：单调性与凹性

很多经济问题中，价值函数具有良好形状。若状态越多越好，则 $V_x \geq 0$；若边际价值递减，则 $V_{xx} \leq 0$。

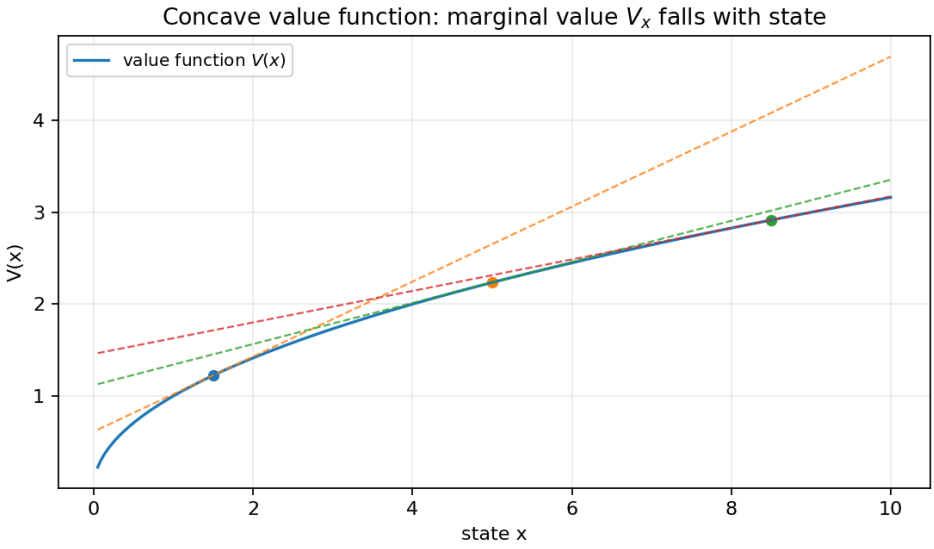



图 6: 凹价值函数表示状态变量的边际价值递减。

## 12.1 为什么凹性重要

凹性带来三个好处：

1. 一阶条件更可能是充分条件；
2. 策略函数通常更稳定、更容易解释；
3. 数值方法更不容易跑出奇怪的局部解。

例如消费储蓄问题中，财富越高，价值越高，但额外一单位财富带来的边际价值下降。于是富人消费更多，但不一定按比例消费更多；借贷约束附近，$V_a$ 很高，消费受到压抑。

# 13 常见误区与纠正

**常见误区**

**误区 1：把 HJB 当成普通最大化。** HJB 中的最大化不仅影响当前收益，还通过 $V_x f$ 影响未来价值。

**误区 2：忘记价值函数的状态变量。**写 HJB 前先问：状态是什么？如果状态选错，HJB 就写不对。

**误区 3：把 $V_x$ 当成普通导数，不解释经济含义。** $V_x$ 是状态变量的边际价值，也就是 shadow value。它通常对应 Hamiltonian 方法中的 costate variable。

**误区 4：求出一阶条件就结束**。一阶条件只给候选策略。还要检查约束、边界、凹性、横截条件和 verification。

**误区 5：忽略状态约束**。借贷约束、不可逆投资、资源非负约束都可能使内点 FOC 失效。HJB 必须写成对可行集 $\Gamma(x)$ 的最大化。

## 14 本章总结图谱

| 概念 | 公式 | 经济含义 |
|---|---|---|
| 平稳 HJB | $\rho V = \max_u \{r + V_x f\}$ | 价值的机会成本等于当前收益加资本利得。 |
| 非平稳 HJB | $\rho V = V_t + \max_u \{r + V_x f\}$ | 有终端时间时价值依赖时间。 |
| 策略函数 | $u^*(x) = \arg\max \{r + V_x f\}$ | 每个状态下的最优决策。 |
| 边际价值 | $V_x$ | 状态变量的 shadow value。 |
| Hamiltonian 对应 | $\mu = V_x$ | costate 等于价值函数导数。 |
| Verification | HJB 不等式 + 等号策略 | 证明候选函数和策略真正最优。 |
| 状态约束 | $u \in \Gamma(x)$ | 可行控制取决于当前状态。 |
| 随机 HJB | $\rho V = \max\{r + V_x \mu + \frac{1}{2} V_{xx} \sigma^2\}$ | Ito 项带来二阶导数。 |

## 15 练习题

### A 组：基础概念与公式

A1. 写出无限期确定性问题 $\max \int_0^\infty e^{-\rho t} r(x,u)\,\mathrm{d}t$ subject to $\dot{x} = f(x,u)$ 的平稳 HJB。

A2. 解释 HJB 中 $V_x(x) f(x,u)$ 的经济含义。

A3. 对有限期问题写出非平稳 HJB 和终端条件。

A4. 为什么 HJB 中的 $V_x$ 对应 Hamiltonian 方法中的 costate variable?

A5. 什么是 verification theorem? 它解决什么问题?

### B 组：推导与计算

B1. 从短时间 Bellman 方程推导平稳 HJB。

B2. 对问题 $\rho V(x) = \max_u \{-\frac{1}{2}u^2 + V_x u\}$ 求最优 $u^*$。

B3. 对消费储蓄 HJB $\rho V(a) = \max_c \{\log c + V_a(ra + y - c)\}$ 写出消费一阶条件。

B4. 对企业投资 HJB $rV(K) = \max_I\{\pi(K) - I - \frac{\phi}{2}I^2 + V_K(I - \delta K)\}$ 求最优投资。

B5. 若 $V(x) = A\log x + B$，计算 $V_x$ 和 $V_{xx}$，并解释边际价值递减。

### C 组：经济应用

C1. 写出确定性资产价值方程 $rV(x) = \pi(x) + V_x g(x)$，并解释每一项。

C2. 在消费储蓄问题中，如果存在借贷约束 $a \geq 0$，HJB 的可行控制集合应如何改变?

C3. 在投资问题中，为什么 $V_K$ 可以解释为 Tobin's $q$?

C4. 资源开采问题中，资源存量 $S$ 满足 $\dot{S} = -q$。写出 HJB 并解释 $V_S$ 的含义。

C5. 工作搜索模型中，失业者选择是否接受工资 $w$。为什么这类问题通常会出现 value matching 和 reservation wage?

### D 组：挑战题

D1. 对 CRRA 消费储蓄问题，从 HJB 推导 Euler 方程 $\dot{c}/c = (r - \rho)/\theta$。

D2. 设 $\rho V(x) = \max_u\{r(x,u) + V_x f(x,u)\}$。推导当前值 costate 方程。

D3. 对线性二次调节问题，验证 $V(x) = -\frac{1}{2}Px^2$ 时 $P$ 满足 $P^2 + \rho P - 1 = 0$。

D4. 写出随机 HJB 的一般形式，并说明为什么会出现 $V_{xx}$ 项。

D5. 说明 HJB 与 Pontryagin 最大值原理在凹问题中的等价性直觉。

# 16 练习题详细答案

## A 组答案

A1. 平稳 HJB 为

$$\rho V(x) = \max_u \{r(x,u) + V_x(x) f(x,u)\}.$$

它表示价值的机会成本等于当前收益加状态变化造成的价值变化。

A2. $V_x f$ 是 capital gain 或 continuation value effect。$f$ 表示状态每单位时间变化多少，$V_x$ 表示状态多一单位的价值，所以乘积表示状态变化给价值带来的瞬时变化。

A3. 有限期价值函数 $V(x,t)$ 满足

$$\rho V(x,t) = V_t(x,t) + \max_u \{r(x,u,t) + V_x(x,t) f(x,u,t)\}.$$

若终端价值为 $S(x)$，终端条件是

$$V(x,T) = S(x).$$

A4. Hamiltonian 方法中的当前值 costate $\mu$ 是状态变量的影子价格；HJB 中 $V_x$ 是价值函数对状态的边际导数，也表示状态多一单位带来的价值增加。因此在最优路径上 $\mu = V_x$。

A5. Verification theorem 用来证明某个满足 HJB 的候选函数确实是价值函数，某个由 HJB 最大化得到的策略确实最优。它防止我们把只满足一阶条件的候选解误认为真正最优解。

## B 组答案

B1. 短时间 Bellman 方程为

$$V(x) = \max_u \{r(x,u)\Delta t + e^{-\rho \Delta t} V(x + f(x,u)\Delta t)\}.$$

Taylor 展开：

$$e^{-\rho \Delta t} \approx 1 - \rho \Delta t,$$

$$V(x + f\Delta t) \approx V(x) + V_x f \Delta t.$$

代入、减去 $V(x)$、除以 $\Delta t$ 并令 $\Delta t \to 0$，得到

$$\rho V(x) = \max_u \{r(x,u) + V_x f(x,u)\}.$$

B2. 最大化对象为

$$-\frac{1}{2}u^2 + V_x u.$$

一阶条件：

$$-u + V_x = 0.$$

所以

$$u^* = V_x.$$

二阶导数为 $-1 < 0$，故为最大值。

B3. 对

$$\rho V(a) = \max_c \{\log c + V_a(ra + y - c)\}$$

求一阶条件：

$$\frac{1}{c} - V_a = 0.$$

因此

$$c^*(a) = \frac{1}{V_a(a)}.$$

B4. 对 $I$ 求一阶条件：

$$-1 - \phi I + V_K = 0.$$

因此

$$I^*(K) = \frac{V_K(K) - 1}{\phi}.$$

若约束 $I \geq 0$，则

$$I^*(K) = \max\left\{0, \frac{V_K(K) - 1}{\phi}\right\}.$$

B5. 若

$$V(x) = A \log x + B,$$

则

$$V_x = \frac{A}{x}, \qquad V_{xx} = -\frac{A}{x^2}.$$

如果 $A > 0$，则 $V_x > 0$ 且 $V_{xx} < 0$，说明状态越多价值越高，但边际价值递减。

## C 组答案

C1. 资产价值方程为

$$rV(x) = \pi(x) + V_x(x)g(x).$$

左边是持有价值 $V$ 要求的回报；右边第一项是当前现金流，第二项是状态变化导致的资本利得。

C2. 借贷约束 $a \geq 0$ 意味着在边界 $a = 0$ 处不能选择让 $\dot{a} < 0$ 的消费。因为 $\dot{a} = ra + y - c$，在 $a = 0$ 时必须有

$$y - c \geq 0,$$

即 $c \leq y$。所以 HJB 应写为

$$\rho V(a) = \max_{c \in \Gamma(a)} \{u(c) + V_a(ra + y - c)\},$$

其中 $\Gamma(a)$ 反映借贷约束。

C3. $V_K$ 表示多一单位资本给企业带来的边际价值。Tobin's $q$ 正是资本的边际价值相对于购买成本的度量。在有调整成本时，一阶条件 $V_K = 1 + \phi I$ 表示资本边际价值越高，投资越强。

C4. 若资源存量 $S$、开采量 $q$，收益为 $\pi(q, S)$，状态方程为 $\dot{S} = -q$，则 HJB 为

$$\rho V(S) = \max_q \{\pi(q, S) - V_S(S)q\}.$$

因为开采减少资源存量，$V_S$ 是保留一单位资源在地下的边际价值，也就是资源的影子价格。

C5. 接受工资意味着从失业价值切换到就业价值。临界工资处，接受和拒绝的价值相等，这叫 value matching。使两种选择价值相等的工资就是 reservation wage。若工资高于保留工资，接受；低于则继续搜索。

## D 组答案

D1. 见正文推导。HJB 为

$$\rho V(a) = \max_c \{u(c) + V_a(ra + y - c)\}.$$

FOC:

$$u'(c) = V_a.$$

Envelope:

$$\rho V_a = V_{aa}\dot{a} + rV_a.$$

沿路径 $\dot{V}_a = u''(c)\dot{c} = V_{aa}\dot{a}$，所以

$$\rho u'(c) = u''(c)\dot{c} + ru'(c).$$

整理:

$$u''(c)\dot{c} = (\rho - r)u'(c).$$

CRRA 下得到

$$\frac{\dot{c}}{c} = \frac{r - \rho}{\theta}.$$

D2. HJB:

$$\rho V = r(x, u^*) + V_x f(x, u^*).$$

对 $x$ 求导并用 envelope 定理:

$$\rho V_x = r_x + V_{xx} f + V_x f_x.$$

令 $\mu = V_x$，则 $\dot{\mu} = V_{xx}\dot{x} = V_{xx}f$。所以

$$\rho\mu = r_x + \dot{\mu} + \mu f_x.$$

整理得到

$$\dot{\mu} = \rho\mu - r_x - \mu f_x.$$

D3. 线性二次问题的 HJB：

$$\rho V = \max_u \{-\frac{1}{2}x^2 - \frac{1}{2}u^2 + V_x u\}.$$

猜 $V = -\frac{1}{2}Px^2$，则 $V_x = -Px$，FOC 给出 $u^* = -Px$。代入：

$$-\frac{1}{2}\rho P x^2 = -\frac{1}{2}x^2 - \frac{1}{2}P^2x^2 + P^2x^2.$$

所以

$$-\rho P = P^2 - 1,$$

即

$$P^2 + \rho P - 1 = 0.$$

D4. 随机 HJB 通常为

$$\rho V(x) = \max_u \left\{ r(x,u) + V_x \mu(x,u) + \frac{1}{2} V_{xx} \sigma(x,u)^2 \right\}.$$

$V_{xx}$ 项来自 Ito 引理：随机扩散项会影响价值函数的期望变化，其大小由价值函数曲率和波动率平方决定。

D5. 在凹问题中，Hamiltonian 对控制的 FOC 给出全局最优控制；HJB 中的最大化也给出同样控制。HJB 的 $V_x$ 等于 Hamiltonian 的 costate，HJB 对状态求导得到 costate 方程。因此两者描述同一最优路径，只是一个强调路径乘子，一个强调价值函数。

## 17 后续学习连接

本章完成了确定性 HJB 的基础。下一步自然是随机过程与 Ito calculus。加入 Brownian motion 后，状态不再按普通 ODE 运动，而是按 stochastic differential equation 运动：

$$\mathrm{d}X_t = \mu(X_t, u_t)\,\mathrm{d}t + \sigma(X_t, u_t)\,\mathrm{d}W_t.$$

此时 HJB 会变成二阶 PDE：

$$\rho V = \max_u \left\{ r + V_x \mu + \frac{1}{2} V_{xx} \sigma^2 \right\}.$$

金融中的 Black-Scholes PDE、Merton portfolio problem、real options、违约边界、公司投资和宏观金融模型都从这里展开。

**直觉盒子**

如果 Chapter 23 的关键词是 Hamiltonian 和 costate，那么 Chapter 24 的关键词就是 value function 和 HJB。两者的桥梁是

$$\boxed{\text{costate} = \text{marginal value} = V_x.}$$

理解这句话，你就理解了连续时间动态优化的核心结构。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 25
# 数值方法与计算经济学入门

前面几章讲了动态规划、最优控制和 HJB。理论上,我们可以写出 Bellman 方程、Euler 方程或 HJB 方程；但在真实经济金融模型中，解析解往往不存在。本章的目标是把数学模型变成可计算对象：如何离散化状态空间，如何求根，如何做数值优化，如何迭代 Bellman operator，如何检查误差和收敛。你不需要立刻成为数值分析专家，但需要知道计算经济学的基本语言和风险。

math4econ personalized notes
2026

# 目录

**Chapter 25 总览** **2**

**1 为什么经济学需要数值方法** **2**
1.1 解析解与数值解 . . . . . 2
1.2 计算经济学的一般流程 . . . . . 3

**2 离散化：把连续问题变成有限问题** **3**
2.1 状态空间网格 . . . . . 3
2.2 网格密度的权衡 . . . . . 4

**3 插值：网格之间怎么办** **4**
3.1 线性插值 . . . . . 5
3.2 插值误差 . . . . . 5

**4 求根：从均衡条件到数值解** **5**
4.1 二分法 bisection method . . . . . 6
4.2 Newton 法 . . . . . 6

**5 数值优化：最大化和最小化** **8**
5.1 网格搜索 . . . . . 8
5.2 梯度法 . . . . . 8
5.3 Newton 优化 . . . . . 8

**6 固定点迭代与压缩映射** **9**
6.1 收敛条件 . . . . . 9
6.2 停止准则 . . . . . 9

**7 价值迭代 Value Function Iteration** **10**
7.1 离散状态和离散控制的版本 . . . . . 10

**8 策略迭代与 Howard Improvement** **12**
8.1 两步结构 . . . . . 12

**9 Euler Equation Iteration** **13**
9.1 Euler error . . . . . 13

**10 有限差分与 HJB 数值入口** **13**
10.1 有限差分公式 . . . . . 13
10.2 Upwind scheme 的经济直觉 . . . . . 14

**11 随机模型与离散化冲击** **14**

**12 模拟 Simulation** **15**
12.1 burn-in . . . . . 15
12.2 Monte Carlo 误差 . . . . . 15

**13 误差、收敛与验证** **15**

**14 Worked Example: 离散消费储蓄问题** **16**
14.1 离散网格算法 . . . . . 16
14.2 经济解释 . . . . . 16

**15 Worked Example: 用二分法求市场清算价格** **17**

**16 常见错误与调试清单** **17**

**17 本章小结** **18**

**18 练习题** **18**
A 组: 基础计算 . . . . . 18
B 组: 概念判断 . . . . . 18
C 组: 混合题 . . . . . 19
D 组: 经济金融建模题 . . . . . 19
E 组: 挑战 / 证明题 . . . . . 19

**19 练习题参考答案** **20**
A 组答案 . . . . . 20
B 组答案 . . . . . 20
C 组答案 . . . . . 21
D 组答案 . . . . . 22
E 组答案 . . . . . 22

**20 后续学习连接** **23**

# Chapter 25 总览

| | |
|---|---|
| **本章主题** | 数值方法 numerical methods 与计算经济学 computational economics 的基本入口。 |
| **核心问题** | 当模型没有解析解时，怎样用计算机求近似解？怎样知道近似解可信？怎样避免把数值误差误认为经济机制？ |
| **数学工具** | 离散化、网格、插值、二分法、Newton 法、固定点迭代、梯度法、Newton 优化、价值迭代、策略迭代、Euler equation error、有限差分。 |
| **经济应用** | 动态规划、消费储蓄、工作搜索、企业退出、资产定价、HJB、数值比较静态、模拟矩。 |
| **后续连接** | Chapter 26 概率论与随机过程，Chapter 27 随机动态模型，Chapter 28 数值模拟与 Monte Carlo，后续计量和 AI 方法中的数值计算。 |

**直觉盒子**

数值方法不是“理论不行时的临时替代”。在现代宏观、金融、产业组织和计量经济学中，很多模型从一开始就被设计为需要计算。理论告诉你问题长什么样；数值方法告诉你如何真正求出政策函数、价值函数、均衡价格和比较静态。

## 1 为什么经济学需要数值方法

如果所有经济模型都有解析解，学习动态规划、最优控制和 HJB 就会简单很多。但现实不是这样。稍微加入一点异质性、借贷约束、非线性效用、随机收入、资本调整成本、不可逆投资、违约风险、离散选择，解析解通常就消失了。

### 1.1 解析解与数值解

解析解 analytic solution 是用有限个公式把答案写出来。例如

$$x_{t+1} = ax_t + b$$

的解是

$$x_t = x^* + a^t(x_0 - x^*).$$

但数值解 numerical solution 通常是一个近似对象，例如一张表、一组网格点上的函数值、一条策略函数曲线、一个迭代后的数组。

**常见误区**

数值解不等于随便算一个答案。好的数值解必须回答三个问题：

1. 近似对象是什么？例如网格上的价值函数还是参数化策略函数。
2. 误差从哪里来？例如离散化误差、插值误差、迭代误差、优化误差。
3. 如何验证？例如 Euler equation error、Bellman residual、网格加密、与已知解析解比较。

### 1.2 计算经济学的一般流程

**方法模板**

**从经济模型到数值解的六步流程**

1. 写出理论模型：目标函数、约束、状态变量、控制变量和冲击。
2. 选择递归表示：Bellman 方程、Euler 方程、HJB 或均衡条件。
3. 选择近似空间：网格、基函数、插值、参数化函数或模拟样本。
4. 选择算法：求根、优化、价值迭代、策略迭代、Newton 法、时间迭代。
5. 检查收敛和误差：停止准则、残差、网格敏感性、政策函数形状。
6. 做经济解释：比较静态、政策实验、福利、分布效应。

## 2 离散化：把连续问题变成有限问题

大多数经济状态变量是连续的。例如资本 $k \in [0, \bar{k}]$，资产 $a \in [\underline{a}, \bar{a}]$，生产率 $z$ 可能服从连续过程。但计算机无法直接处理无限多个状态点。第一步通常是离散化 discretization。

### 2.1 状态空间网格

给定区间 $[\underline{k}, \bar{k}]$，选择 $N$ 个网格点：

$$k_1, k_2, \dots, k_N.$$

最简单的是均匀网格：

$$k_i = \underline{k} + (i-1)\Delta k, \qquad \Delta k = \frac{\bar{k} - \underline{k}}{N-1}.$$

也可以使用非均匀网格，例如在借贷约束附近放更多点，因为政策函数在那里可能弯曲更剧烈。

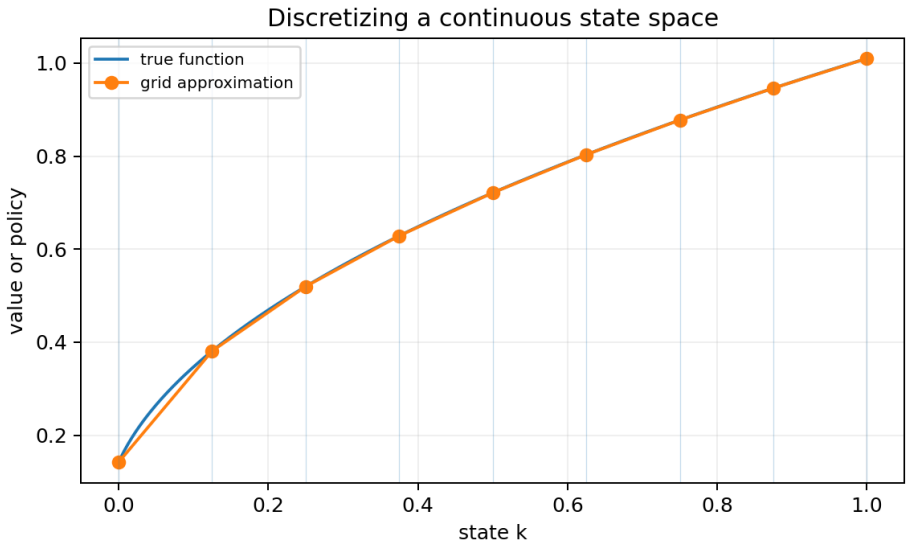



图 1：把连续状态空间离散成有限网格。网格越密，近似通常越好，但计算成本越高。

**直觉盒子**

离散化的本质是用有限维对象近似无限维对象。价值函数本来是一个函数 $V(k)$，离散化后变成一个向量

$$(V(k_1), V(k_2), \ldots, V(k_N)).$$

这就是为什么线性代数和数值方法会进入动态经济学。

### 2.2 网格密度的权衡

网格太粗，会错过函数曲率、约束点和阈值；网格太密，计算成本快速上升。若状态变量有 $d$ 个维度，每个维度有 $N$ 个点，总点数是

$$N^d.$$

这就是 curse of dimensionality 维数灾难。

**常见误区**

维数灾难是计算经济学的核心困难之一。一个状态变量 500 个点不算大；两个状态变量就是 250000 个组合；五个状态变量就几乎不可直接用简单网格暴力计算。

## 3 插值：网格之间怎么办

如果只在网格点上知道函数值，但最优选择落在两个网格点之间，就需要 interpolation 插值。

### 3.1 线性插值

若 $x \in [x_i, x_{i+1}]$，且已知 $f(x_i)$、$f(x_{i+1})$，线性插值为

$$f(x) \approx \frac{x_{i+1} - x}{x_{i+1} - x_i} f(x_i) + \frac{x - x_i}{x_{i+1} - x_i} f(x_{i+1}).$$

它的含义是：把两个网格点之间的函数看成一条直线。

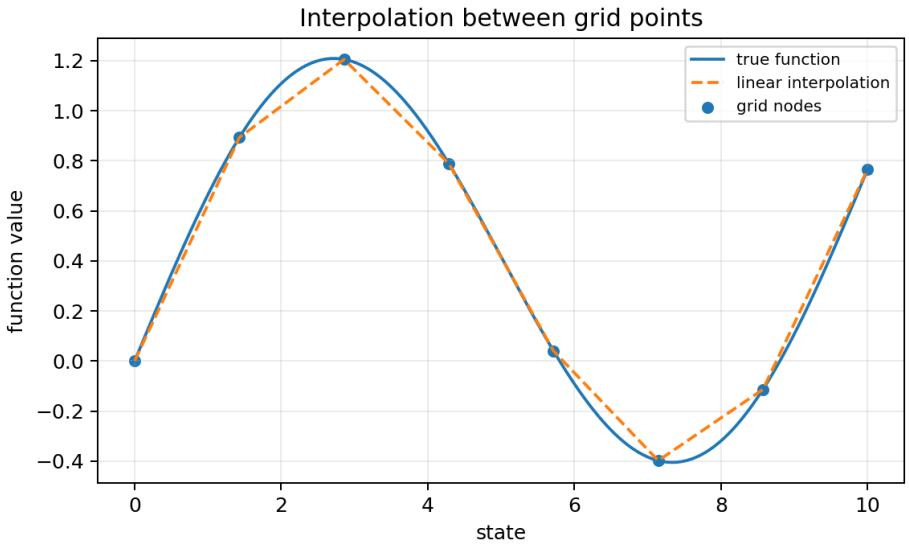



图 2：线性插值用相邻网格点之间的直线近似真实函数。

### 3.2 插值误差

如果函数很平滑，且网格足够密，线性插值效果通常不错。但如果函数在某处有 kink，例如借贷约束或不可逆投资约束导致政策函数不可微，简单插值可能产生明显误差。

**方法模板**

**插值检查清单**

1. 函数是否单调？插值后是否破坏单调性？
2. 函数是否凹？插值后是否出现不该有的凸起？
3. 约束点附近是否加密网格？
4. 是否比较过不同插值方法和不同网格密度？

## 4 求根：从均衡条件到数值解

很多经济问题最终可以写成求根：

$$f(x) = 0.$$

例如市场清算、Euler 方程残差、稳态条件、FOC 系统。

## 4.1 二分法 bisection method

如果 $f$ 连续，且

$$f(a)f(b)<0,$$

则在 $[a,b]$ 中至少有一个根。二分法不断取中点

$$m=\frac{a+b}{2},$$

然后保留有符号变化的一半区间。

**方法模板**

**二分法算法**

1. 找到 $[a,b]$，使 $f(a)f(b)<0$。
2. 取中点 $m=(a+b)/2$。
3. 若 $f(a)f(m)<0$，令 $b=m$；否则令 $a=m$。
4. 重复直到区间长度足够小。

二分法慢，但很稳健。它不需要导数，只需要符号变化。

## 4.2 Newton 法

Newton 法用切线近似函数。给定当前猜测 $x_n$，更新为

$$\boxed{x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}.}$$

如果初始点好、函数足够光滑、导数不接近 0，Newton 法非常快。

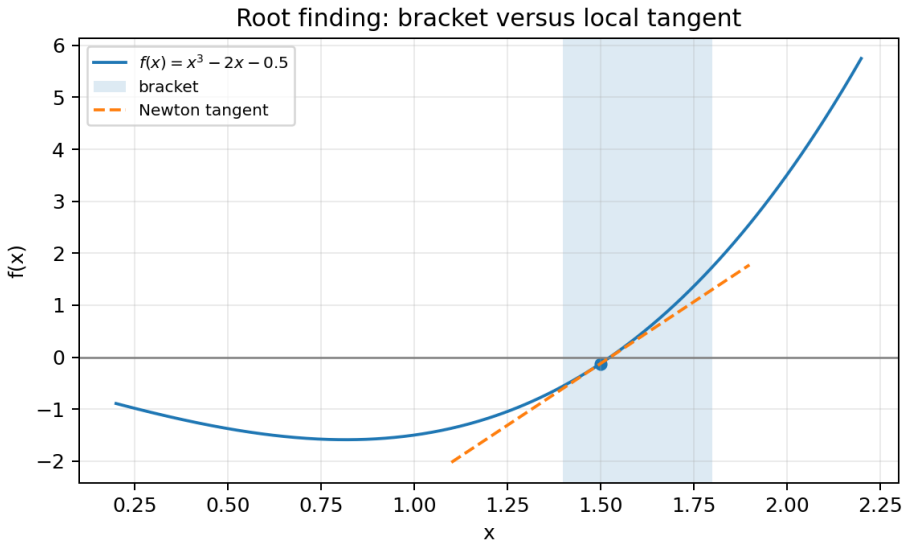



图 3: 二分法依赖 bracket，Newton 法依赖局部切线。前者稳健，后者快但可能不稳。

**常见误区**

Newton 法常见失败原因：

- 初始点离根太远；
- 导数太小，导致跳跃过大；
- 函数有多个根，算法收敛到错误根；
- 迭代跑出经济可行域，例如消费变成负数。

实际计算中常用 damped Newton，即

$$x_{n+1} = x_n - \lambda \frac{f(x_n)}{f'(x_n)}, \quad 0 < \lambda \le 1.$$

**Worked Example**

**例 1: 用 Newton 法求 Euler 方程根**

考虑简单消费储蓄问题的一阶条件写成

$$f(c) = u'(c) - \beta R u'(w - c) = 0.$$

若 $u(c) = \log c$，则

$$f(c) = \frac{1}{c} - \beta R \frac{1}{w - c}.$$

导数为

$$f'(c) = -\frac{1}{c^2} - \beta R \frac{1}{(w - c)^2}.$$

Newton 更新为

$$c_{n+1} = c_n - \frac{\frac{1}{c_n} - \beta R \frac{1}{w-c_n}}{-\frac{1}{c_n^2} - \beta R \frac{1}{(w-c_n)^2}}.$$

注意必须保持 $0 < c_n < w$，否则效用或边际效用没有经济意义。

# 5 数值优化：最大化和最小化

优化问题是经济学的核心。理论里我们写

$$\max_x f(x),$$

数值计算里则需要算法寻找最优 $x$。

## 5.1 网格搜索

最直接的方法是在网格上计算目标函数：

$$x^* \approx \arg\max_{x_i \in \mathcal{G}} f(x_i).$$

优点是稳健、不需要导数；缺点是慢且精度受网格限制。

## 5.2 梯度法

若要最大化 $f(x)$，梯度上升法为

$$x_{n+1} = x_n + \eta_n \nabla f(x_n),$$

其中 $\eta_n$ 是 step size。若要最小化则反向：

$$x_{n+1} = x_n - \eta_n \nabla f(x_n).$$

## 5.3 Newton 优化

若 $f$ 二阶可微，最大化时 Newton 更新为

$$x_{n+1} = x_n - [\nabla^2 f(x_n)]^{-1} \nabla f(x_n).$$

这里 Hessian 决定曲率。若 Hessian 接近奇异或不是负定，Newton 步可能不可靠。

**方法模板**

**优化算法选择的经验规则**

1. 低维、目标函数可能不光滑：先用网格搜索。

2. 光滑且一维：二分法/黄金分割/Brent 法通常稳健。
3. 光滑且导数可得：梯度法或 Newton 法。
4. 约束复杂：使用 constrained optimization，并检查 KKT residual。
5. 动态规划内部优化：常常先粗网格，再局部 refine。

# 6 固定点迭代与压缩映射

很多经济问题可以写成固定点：

$$x = T(x).$$

例如均衡价格、策略函数、价值函数。最自然的算法是 fixed point iteration：

$$x_{n+1} = T(x_n).$$

## 6.1 收敛条件

如果 $T$ 是 contraction mapping，即存在 $0 < q < 1$ 使

$$\|T(x) - T(y)\| \le q \|x - y\|,$$

则迭代收敛到唯一固定点。

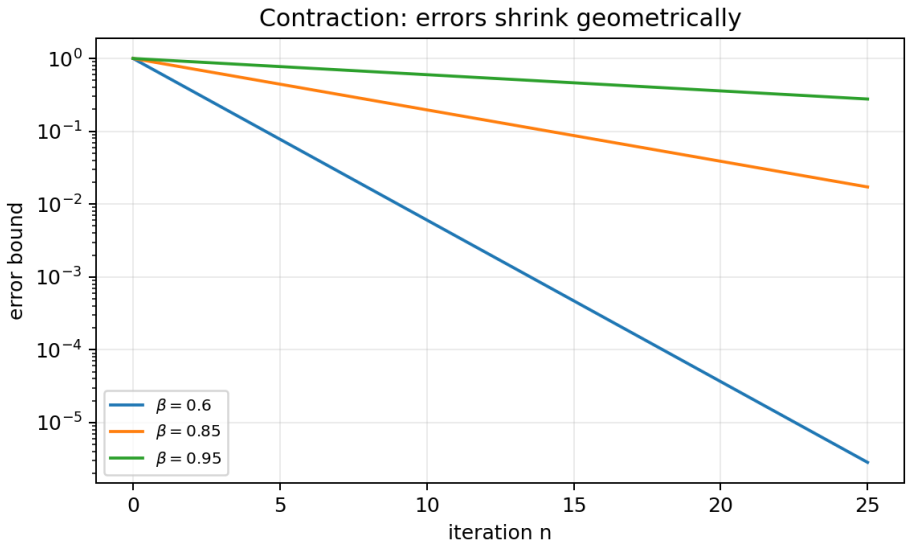



图 4: 压缩映射下，误差按几何速度下降。折现因子越接近 1，价值迭代越慢。

## 6.2 停止准则

常见停止准则是

$$\|x_{n+1} - x_n\| < \varepsilon.$$

但要注意，迭代差小不一定代表真实误差小。若 $T$ 是 contraction，且压缩系数为 $q$，则有界估计

$$\|x_n - x^*\| \le \frac{q}{1-q}\|x_n - x_{n-1}\|.$$

当 $q$ 接近 1 时，即使相邻迭代差不大，真实误差仍可能不小。

# 7 价值迭代 Value Function Iteration

Chapter 22 中的 Bellman 方程是

$$V(s) = \max_{a\in\Gamma(s)}\{u(s,a) + \beta V(s')\}.$$

数值方法里，我们定义 Bellman operator：

$$(TV)(s) = \max_{a\in\Gamma(s)}\{u(s,a) + \beta V(s')\}.$$

然后从一个初始猜测 $V_0$ 开始：

$$V_{n+1} = TV_n.$$

**定义 / 定理**

若 $0 < \beta < 1$，在通常条件下 Bellman operator 是压缩映射，因此价值迭代收敛到唯一价值函数 $V^*$。

## 7.1 离散状态和离散控制的版本

假设状态网格为

$$s_1, \dots, s_N,$$

控制网格为

$$a_1, \dots, a_M.$$

对每个状态点 $s_i$，计算所有可行控制 $a_j$ 的 RHS：

$$RHS_{ij} = u(s_i, a_j) + \beta V_n(s'(s_i, a_j)).$$

选择最大值：

$$V_{n+1}(s_i) = \max_j RHS_{ij}.$$

同时记录最优控制：

$$g_{n+1}(s_i) = \arg\max_j RHS_{ij}.$$

**方法模板**

**价值迭代算法**

1. 建立状态网格和控制网格。
2. 给定初始价值函数 $V_0$。
3. 对每个状态点，枚举或优化控制变量。
4. 得到新的 $V_{n+1} = TV_n$ 和策略函数 $g_{n+1}$。
5. 若 $\|V_{n+1} - V_n\|$ 足够小，停止；否则继续。
6. 用 Euler error、policy convergence 和网格加密检查结果。

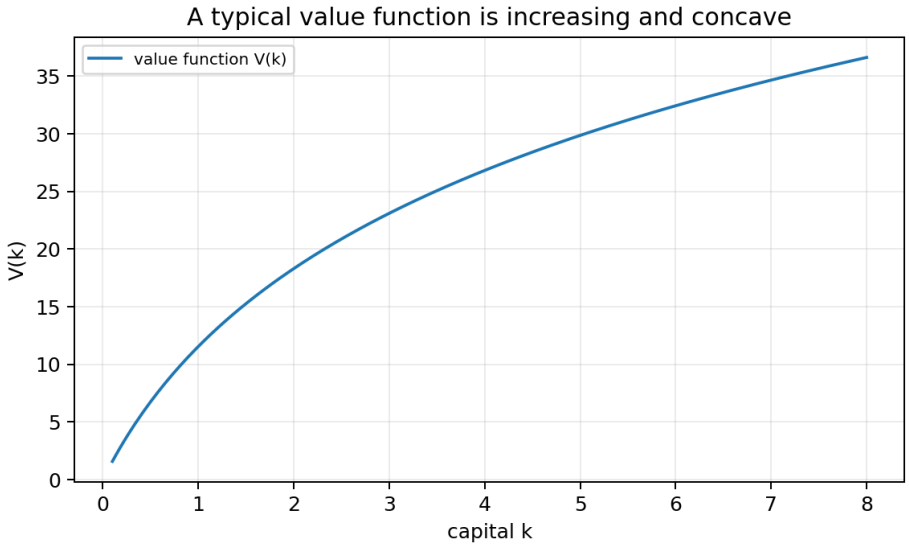



图 5: 典型消费储蓄问题中，价值函数通常递增且凹。数值解应符合这些理论性质。

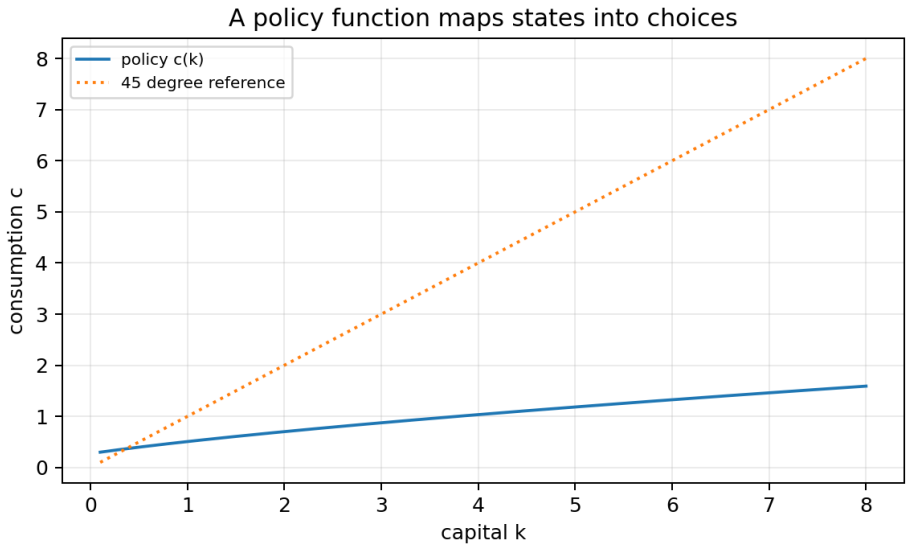



图 6: 策略函数 policy function 把状态映射到最优选择，是动态模型数值解最重要的输出之一。

# 8 策略迭代与 Howard Improvement

价值迭代稳健，但当 $\beta$ 接近 1 时会很慢。策略迭代 policy iteration 的思想是：先给定一个策略，计算它对应的价值，再改进策略。

## 8.1 两步结构

1. **Policy evaluation:** 给定策略 $g$，求解

$$V_g(s) = u(s, g(s)) + \beta V_g(s'(s, g(s))).$$

2. **Policy improvement:** 用 $V_g$ 更新策略：

$$g_{new}(s) = \arg\max_a \{u(s, a) + \beta V_g(s'(s, a))\}.$$

若策略不再改变，则得到最优策略。

**直觉盒子**

价值迭代每次只把价值函数往正确方向推一点；策略迭代则是“先彻底算清楚当前策略的价值，再重新优化”。在很多离散动态规划问题中，策略迭代比单纯价值迭代快很多。

# 9 Euler Equation Iteration

动态规划可以用 Bellman 方程求解，也可以用 Euler 方程求解。以消费储蓄问题为例：

$$u'(c_t) = \beta R E_t u'(c_{t+1}).$$

如果猜测一个消费政策函数 $c = g(a, z)$，就可以根据 Euler 方程更新它。

## 9.1 Euler error

Euler equation error 衡量一组数值策略是否满足一阶条件。例如定义残差

$$\mathcal{E}(s) = 1 - \frac{\beta R E[u'(c(s'))]}{u'(c(s))}.$$

若 $\mathcal{E}(s)$ 接近 0，说明 Euler 方程近似满足。

也常用消费单位误差：假设需要把当前消费调整为 $\tilde{c}$ 才能精确满足 Euler 方程，那么

$$\log_{10}\left|\frac{\tilde{c} - c}{c}\right|$$

越小越好。例如 $-3$ 表示相对误差约为 $0.1\%$。

**常见误区**

只看 Bellman 迭代差可能误导。一个策略函数看起来平滑、价值函数看起来收敛，并不一定代表 Euler 方程精确满足。尤其在借贷约束、非均匀网格和高持久冲击下，应检查 Euler error。

# 10 有限差分与 HJB 数值入口

Chapter 24 的 HJB 典型形式是

$$\rho V(x) = \max_a \{u(x, a) + V_x(x) f(x, a)\}.$$

数值上，需要近似导数 $V_x$。这就是 finite difference 有限差分。

## 10.1 有限差分公式

设网格间距为 $\Delta x$。

前向差分：

$$V_x(x_i) \approx \frac{V_{i+1} - V_i}{\Delta x}.$$

后向差分：

$$V_x(x_i) \approx \frac{V_i - V_{i-1}}{\Delta x}.$$

中心差分：

$$V_x(x_i) \approx \frac{V_{i+1} - V_{i-1}}{2\Delta x}.$$

二阶导数：

$$V_{xx}(x_i) \approx \frac{V_{i+1} - 2V_i + V_{i-1}}{(\Delta x)^2}.$$

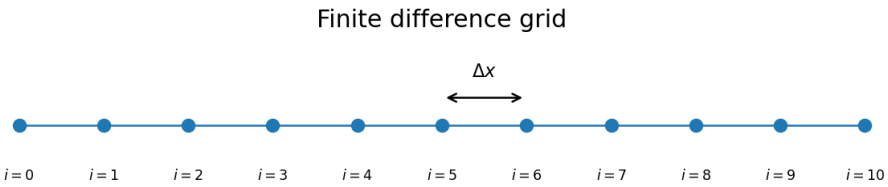



图 7：有限差分用相邻网格点的函数值近似导数。

### 10.2 Upwind scheme 的经济直觉

如果状态变量按照

$$\dot{x} = f(x, a)$$

向右移动，则导数近似应使用来自左侧的信息；如果向左移动，则使用右侧信息。这类方法叫 upwind scheme。

经济直觉是：HJB 是沿着状态运动方向传播信息。若你用错方向，可能产生数值振荡或不稳定结果。

## 11 随机模型与离散化冲击

很多动态模型包含随机冲击 $z_t$。若 $z_t$ 连续，例如 AR(1)：

$$z_{t+1} = \rho z_t + \varepsilon_{t+1},$$

数值计算通常把它离散成有限状态 Markov chain：

$$z \in \{z_1, \dots, z_M\},$$

并用转移矩阵

$$P_{ij} = \Pr(z_{t+1} = z_j \mid z_t = z_i).$$

此时 Bellman 方程变为

$$V(a_i, z_m) = \max_{a'} \left\{ u(c) + \beta \sum_{j=1}^{M} P_{mj} V(a', z_j) \right\}.$$

**方法模板**

**随机动态规划计算对象**

- 状态网格：资产 $a_i$ 与冲击 $z_m$ 的组合。
- 控制变量：下一期资产 $a'$ 或消费 $c$。
- 转移矩阵：$P_{mj}$。
- 输出：价值函数 $V(a,z)$、政策函数 $a'=g(a,z)$、模拟路径、稳态分布。

# 12 模拟 Simulation

求得政策函数后，常常需要模拟模型经济。给定初始状态 $s_0$，反复使用

$$a_t = g(s_t), \qquad s_{t+1} = F(s_t, a_t, \varepsilon_{t+1}).$$

得到一条样本路径。模拟可以用来计算均值、方差、分布、转移矩阵和政策实验。

## 12.1 burn-in

如果初始状态不代表长期分布，模拟前面若干期常被丢弃，称为 burn-in。例如模拟 10000 期，丢弃前 1000 期，只用后 9000 期计算统计量。

## 12.2 Monte Carlo 误差

模拟样本有限，所以统计量有随机误差。样本长度越长，Monte Carlo error 通常越小。但如果过程高度持久，独立信息增长很慢，需要更长模拟。

# 13 误差、收敛与验证

数值结果必须被验证。常见验证方法包括：

| 检查方法 | 含义 |
|---|---|
| Bellman residual | 检查 $V$ 是否接近满足 Bellman 方程。 |
| Euler equation error | 检查策略是否满足一阶条件。 |
| Grid refinement | 加密网格后结果是否稳定。 |
| Known solution test | 用有解析解的特殊情形测试代码。 |
| Shape restrictions | 价值函数是否递增、凹；策略函数是否单调。 |
| Budget feasibility | 消费、资产、资本是否满足约束。 |
| Policy convergence | 策略函数是否稳定，而不只是价值函数稳定。 |

| | |
|---|---|
| Simulation moments | 模拟矩是否与理论或数据目标一致。 |

**常见误区**

一个看起来漂亮的图不等于正确的数值解。最危险的情况是：代码运行、图像平滑、结果符合直觉，但预算约束或 Euler 方程其实有系统误差。因此计算经济学必须同时看经济直觉和数值残差。

# 14 Worked Example: 离散消费储蓄问题

考虑一个确定性消费储蓄问题：

$$V(a) = \max_{a' \in [\underline{a}, \bar{a}]} \left\{ \log c + \beta V(a') \right\},$$

约束为

$$c = Ra + y - a', \qquad c > 0.$$

## 14.1 离散网格算法

1. 选择资产网格 $a_1, \dots, a_N$。
2. 给定初始价值函数 $V_0(a_i) = 0$。
3. 对每个当前资产 $a_i$，枚举下一期资产 $a_j$。
4. 计算消费

$$c_{ij} = Ra_i + y - a_j.$$

若 $c_{ij} \le 0$，设 RHS 为非常小的惩罚值。

5. 计算

$$RHS_{ij} = \log c_{ij} + \beta V_n(a_j).$$

6. 对 $j$ 取最大：

$$V_{n+1}(a_i) = \max_j RHS_{ij}.$$

记录最优 $a_j$ 为策略函数。

## 14.2 经济解释

资产越高，消费可行集合越大，价值函数应更高。若效用是凹的，价值函数通常也是凹的。若收入和回报固定，策略函数通常随资产上升而上升。若存在借贷约束 $a' \ge \underline{a}$，低资产处约束可能 binding，Euler 方程以不等式形式成立。

## 15 Worked Example: 用二分法求市场清算价格

设需求和供给分别为

$$D(p) = 10 - 2p, \qquad S(p) = 1 + p^2.$$

市场清算要求

$$f(p) = D(p) - S(p) = 0.$$

即

$$f(p) = 9 - 2p - p^2.$$

在 $p = 1$ 时，

$$f(1) = 9 - 2 - 1 = 6 > 0.$$

在 $p = 3$ 时，

$$f(3) = 9 - 6 - 9 = -6 < 0.$$

因此根在 $[1, 3]$。二分法第一步取

$$m = 2.$$

此时

$$f(2) = 9 - 4 - 4 = 1 > 0.$$

所以根在 $[2, 3]$。继续取 2.5：

$$f(2.5) = 9 - 5 - 6.25 = -2.25 < 0.$$

所以根在 $[2, 2.5]$。不断重复即可。

解析解是

$$p^* = -1 + \sqrt{10} \approx 2.1623.$$

这个例子说明：二分法虽然慢，但只要 bracket 正确，就非常稳。

## 16 常见错误与调试清单

**常见误区**

**错误 1：忽略可行域**。例如消费储蓄问题中出现 $c \le 0$，但代码仍然计算 $\log c$。

**错误 2：网格边界太窄**。最优策略总是贴着上界或下界，可能说明边界设置不合理。

**错误 3：只检查 value convergence**。价值函数变化很小，但策略函数仍可能跳动。

**错误 4：把插值外推当成插值**。如果选择跑出网格，需要外推 extrapolation，误差可能很大。

**错误 5：没有与简单解析例子对照**。写复杂模型前，应先用能手算的模型验证代码。

## 17 本章小结

| 概念 | 公式/方法 | 经济含义 |
|---|---|---|
| 离散化 | $k_i = \underline{k} + (i-1)\Delta k$ | 把连续状态变成有限数组。 |
| 插值 | 线性组合相邻网格点 | 处理网格之间的状态或选择。 |
| 二分法 | 保留符号变化区间 | 稳健求均衡价格或 FOC 根。 |
| Newton 法 | $x_{n+1} = x_n - f/f'$ | 快速但依赖初始点和导数。 |
| 价值迭代 | $V_{n+1} = TV_n$ | 用 Bellman operator 求价值函数。 |
| 策略迭代 | evaluation + improvement | 常比价值迭代更快。 |
| Euler error | FOC residual | 检查政策函数是否准确。 |
| 有限差分 | $(V_{i+1} - V_i)/\Delta x$ | HJB 数值求解入口。 |
| 模拟 | $s_{t+1} = F(s_t, g(s_t), \varepsilon)$ | 生成模型数据和矩。 |

## 18 练习题

### A 组：基础计算

A1. 在区间 $[0, 10]$ 上建立 6 个均匀网格点，写出所有点和 $\Delta x$。

A2. 已知 $f(2) = 3$，$f(4) = 7$，用线性插值估计 $f(3.5)$。

A3. 对 $f(x) = x^2 - 2$，从区间 $[1, 2]$ 开始做二分法两步，写出新区间。

A4. 对 $f(x) = x^2 - 2$，写出 Newton 更新公式，并从 $x_0 = 1$ 算出 $x_1$。

A5. 若 Bellman operator 的压缩系数为 $\beta = 0.9$，相邻两次价值函数 sup norm 差为 $10^{-4}$，用误差界估计真实误差上界。

A6. 用中心差分近似 $V_x(x_i)$ 的公式是什么?

### B 组：概念判断

B1. 网格越密是否一定越好? 说明计算成本和误差之间的权衡。

B2. Newton 法为什么可能失败? 列出至少两个原因。

B3. 为什么动态规划里只看价值函数收敛可能不够?

B4. Euler equation error 的经济含义是什么?

B5. HJB 数值解中为什么要关心 upwind scheme?

### C 组：混合题

C1. 对方程 $x = \cos x$，写成固定点迭代形式，并从 $x_0 = 1$ 计算两步。

C2. 设市场清算条件为 $f(p) = 20 - 3p - p^2 = 0$。找一个包含正根的 bracket，并做一次二分法。

C3. 给定一阶线性插值公式，证明权重之和等于 1。

C4. 对消费储蓄问题 $V(a) = \max_{a'}\{\log(Ra + y - a') + \beta V(a')\}$，写出离散网格上的 Bellman 更新式。

C5. 如果 AR(1) 冲击被离散成 3 个状态，Bellman 方程中的期望如何写成矩阵加权求和?

### D 组：经济金融建模题

D1. 解释为什么借贷约束附近通常需要更密的资产网格。

D2. 在企业退出模型中，价值函数满足 $V(z) = \max\{0, \pi(z) + \beta E[V(z')]\}$。说明如何用价值迭代求解。

D3. 在资产定价模型中，价格满足 $p = \beta E[m'(d + p')]$。为什么这可以看作固定点问题?

D4. 用数值方法做政策实验时，为什么必须同时报告算法误差或稳健性检查?

### E 组：挑战 / 证明题

E1. 证明二分法每一步把 bracket 长度减半。因此经过 $n$ 步后区间长度是多少?

E2. 若 $T$ 是 contraction，证明固定点迭代误差满足 $\|x_{n+1} - x^*\| \le q\,\|x_n - x^*\|$。

E3. 对一维 Newton 法，说明为什么在根附近可能具有二阶收敛。只要求直觉推导。

E4. 设 $V_{n+1} = TV_n$，且 $T$ 是折现因子 $\beta$ 的压缩映射。推导误差界 $\|V_n - V^*\| \le \frac{\beta}{1-\beta}\,\|V_n - V_{n-1}\|$。

# 19 练习题参考答案

## A 组答案

A1. 6 个点意味着 $N=6$，

$$\Delta x=\frac{10-0}{6-1}=2.$$

网格点为

$$0,2,4,6,8,10.$$

A2. 线性插值权重为

$$\frac{4-3.5}{4-2}=0.25,\qquad \frac{3.5-2}{4-2}=0.75.$$

所以

$$f(3.5)\approx 0.25\cdot 3+0.75\cdot 7=6.$$

A3. 初始区间 $[1,2]$。中点 $1.5$，

$$f(1.5)=2.25-2=0.25>0.$$

因 $f(1)<0$，新区间为 $[1,1.5]$。第二步中点 $1.25$，

$$f(1.25)=1.5625-2<0.$$

新区间为 $[1.25,1.5]$。

A4. Newton 公式为

$$x_{n+1}=x_n-\frac{x_n^2-2}{2x_n}.$$

从 $x_0=1$ 出发，

$$x_1=1-\frac{-1}{2}=1.5.$$

A5. 误差界为

$$\|V_n-V^*\|\le\frac{\beta}{1-\beta}\|V_n-V_{n-1}\|=\frac{0.9}{0.1}10^{-4}=9\times10^{-4}.$$

A6. 中心差分为

$$V_x(x_i)\approx\frac{V_{i+1}-V_{i-1}}{2\Delta x}.$$

## B 组答案

B1. 不一定。网格更密通常降低离散化误差，但会提高计算成本，尤其多维状态下点数按 $N^d$ 增长。更好的做法是在曲率大或约束附近加密网格，而不是盲目全局加密。

B2. Newton 法可能失败，因为初始点太远、导数接近 0、函数有多个根、迭代跑出可行域、目标函数不光滑等。

B3. 因为价值函数变化小不一定意味着策略函数准确。政策函数可能在约束附近跳动, Euler 方程也可能有残差。

B4. Euler error 衡量数值策略是否满足跨期边际条件。若误差大，说明消费/储蓄策略在边际替代上不准确。

B5. HJB 中信息沿状态运动方向传播。upwind scheme 根据漂移方向选择差分方向，可以避免错误方向导致的数值震荡。

## C 组答案

C1. 固定点形式为

$$x_{n+1} = \cos x_n.$$

从 $x_0 = 1$ 出发:

$$x_1 = \cos 1 \approx 0.5403,$$

$$x_2 = \cos 0.5403 \approx 0.8576.$$

C2. $f(2) = 20 - 6 - 4 = 10 > 0$，$f(4) = 20 - 12 - 16 = -8 < 0$，所以 bracket 可取 $[2, 4]$。中点为 3,

$$f(3) = 20 - 9 - 9 = 2 > 0.$$

新区间为 $[3, 4]$。

C3. 权重为

$$w_i = \frac{x_{i+1} - x}{x_{i+1} - x_i}, \qquad w_{i+1} = \frac{x - x_i}{x_{i+1} - x_i}.$$

因此

$$w_i + w_{i+1} = \frac{x_{i+1} - x + x - x_i}{x_{i+1} - x_i} = 1.$$

C4. 若资产网格为 $a_1, \dots, a_N$，则

$$V_{n+1}(a_i) = \max_{j : c_{ij} > 0} \left\{ \log c_{ij} + \beta V_n(a_j) \right\},$$

其中

$$c_{ij} = R a_i + y - a_j.$$

C5. 若冲击状态为 $z_1, z_2, z_3$，转移概率为 $P_{mj}$，则期望项写为

$$E[V(a', z') \mid z_m] = \sum_{j=1}^{3} P_{mj} V(a', z_j).$$

## D 组答案

D1. 借贷约束附近政策函数经常有 kink，Euler 方程从等式变成不等式。若网格太粗，会错过约束开始 binding 的位置，导致消费和储蓄误差较大。

D2. 定义 Bellman operator:

$$(TV)(z_i) = \max\left\{0, \pi(z_i) + \beta \sum_j P_{ij} V(z_j)\right\}.$$

从 $V_0$ 开始重复 $V_{n+1} = TV_n$，直到收敛。退出阈值由继续价值和 0 的比较决定。

D3. 因为价格函数 $p$ 出现在等式两边：当前价格等于由未来价格和 payoff 计算出的贴现价值。定义 operator

$$(Tp)(s) = \beta E[m'(d(s') + p(s')) \mid s],$$

均衡价格满足 $p = Tp$。

D4. 政策实验结果可能受网格、插值、迭代误差影响。如果不报告误差检查，读者无法判断结果来自经济机制还是数值误差。

## E 组答案

E1. 初始区间长度为 $L_0 = b - a$。二分法每次保留一半，所以第 $n$ 步后长度为

$$L_n = \frac{b-a}{2^n}.$$

E2. 若 $x^*$ 是固定点，即 $T(x^*) = x^*$，且 $T$ 是 contraction，则

$$\|x_{n+1} - x^*\| = \|T(x_n) - T(x^*)\| \le q \|x_n - x^*\|.$$

因 $q < 1$，误差几何下降。

E3. 在根 $x^*$ 附近，Taylor 展开有

$$f(x_n) = f'(x^*)(x_n - x^*) + \frac{1}{2} f''(x^*)(x_n - x^*)^2 + \cdots.$$

Newton 法用线性项消除主要误差，因此下一步误差大致与当前误差的平方成比例。这就是二阶收敛的直觉。

E4. 因 $V^* = TV^*$,

$$\|V_n - V^*\| \le \|V_n - V_{n+1}\| + \|V_{n+1} - V^*\|.$$

又

$$\|V_{n+1} - V^*\| = \|TV_n - TV^*\| \le \beta \|V_n - V^*\|.$$

所以

$$(1-\beta)\,\|V_n - V^*\| \le \|V_n - V_{n+1}\|\,.$$

一个常用后验界也可写为

$$\|V_n - V^*\| \le \frac{\beta}{1-\beta}\,\|V_n - V_{n-1}\|\,,$$

它来自把 $\|V_n - V_{n+1}\|$ 用压缩性界定为不超过 $\beta\,\|V_n - V_{n-1}\|$。

## 20 后续学习连接

本章完成后，你已经拥有把理论模型变成数值模型的基本工具。下一章将进入概率论与随机过程基础。原因是：真实经济金融模型不仅非线性、动态，而且充满不确定性。消费储蓄模型需要收入风险，资产定价需要随机折现因子，宏观模型需要生产率冲击，金融模型需要波动率和随机过程。因此，数值方法和随机过程会在后续章节不断结合。

**直觉盒子**

经济金融数学的路线到这里已经很清晰：优化告诉你个体如何选择，动态系统告诉你变量如何演化，动态规划告诉你如何递归求最优，数值方法告诉你如何真正求解，而概率论将告诉你如何把风险和不确定性放进这些模型。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 26

# 概率论基础、随机变量与风险

前面几章完成了优化、动态规划、最优控制和数值方法。接下来进入随机世界。经济金融模型里的收益、收入、冲击、状态转移、样本误差、资产风险、违约、失业、政策不确定性，本质上都需要概率语言。本章不是为了把概率论讲成抽象公理课, 而是为了建立经济金融中真正高频使用的工具：随机变量、分布、期望、方差、协方差、条件期望、大数定律和中心极限定理。

math4econ personalized notes
2026

# 目录

**Chapter 26 总览** **2**

**1 为什么经济金融需要概率论** **2**
1.1 从确定变量到随机变量 . . . . . 2
1.2 概率论在经济金融中的四个角色 . . . . . 3

**2 概率空间：样本空间、事件与概率** **3**
2.1 三个组成部分 . . . . . 3
2.2 概率的基本性质 . . . . . 3

**3 随机变量与分布** **4**
3.1 随机变量 . . . . . 4
3.2 CDF: 累积分布函数 . . . . . 5
3.3 PMF 与 PDF . . . . . 5

**4 期望：平均结果与决策基准** **6**
4.1 离散型期望 . . . . . 6
4.2 连续型期望 . . . . . 7
4.3 期望的线性性 . . . . . 7
4.4 Jensen 不等式入口 . . . . . 8

**5 方差、标准差与风险** **8**
5.1 定义 . . . . . 8
5.2 尺度变换 . . . . . 9

**6 协方差、相关系数与组合风险** **9**
6.1 协方差 . . . . . 9
6.2 相关系数 . . . . . 10
6.3 投资组合方差 . . . . . 10

**7 条件概率与 Bayes 规则** **11**
7.1 条件概率 . . . . . 11
7.2 乘法公式与全概率公式 . . . . . 11
7.3 Bayes 规则 . . . . . 12

**8 独立性** **12**
8.1 事件独立 . . . . . 12
8.2 随机变量独立 . . . . . 13

**9 条件期望：最优预测与回归入口** **13**
9.1 定义与直觉 . . . . . 13
9.2 最优预测性质 . . . . . 14
9.3 迭代期望法则 . . . . . 15
9.4 全方差公式 . . . . . 15

**10 常用分布** **15**
10.1 Bernoulli 分布 . . . . . 15
10.2 Binomial 分布 . . . . . 15
10.3 Uniform 分布 . . . . . 16
10.4 Normal 分布 . . . . . 16

**11 大数定律：样本均值为什么可靠** **16**
11.1 样本均值 . . . . . 16
11.2 为什么 LLN 对计量重要 . . . . . 17

**12 中心极限定理：为什么正态分布无处不在** **17**
12.1 经济金融解释 . . . . . 18

**13 经济金融应用 1：风险溢价与期望收益** **18**

**14 经济金融应用 2：回归为什么是条件期望** **19**

**15 经济金融应用 3：保险、违约与条件概率** **19**

**16 常见错误与辨析** **20**

**17 本章小结：概念 - 公式 - 经济含义** **20**

**18 练习题** **21**
A 组：基础计算 . . . . . 21
B 组：概念判断 . . . . . 21
C 组：混合题 . . . . . 21
D 组：经济金融建模题 . . . . . 21
E 组：挑战 / 证明题 . . . . . 22

**19 练习题参考答案** **23**
A 组答案 . . . . . 23
B 组答案 . . . . . 23
C 组答案 . . . . . 24
D 组答案 . . . . . 24
E 组答案 . . . . . 25

**20 后续学习连接** **26**

# Chapter 26 总览

| | |
|---|---|
| **本章主题** | 概率论基础、随机变量、分布、矩、条件期望和极限定理。 |
| **核心问题** | 当经济结果不是确定数，而是带不确定性的随机变量时，如何描述风险、计算平均结果、衡量波动、理解变量之间的共同变化，并用样本推断总体? |
| **数学工具** | 概率空间、随机变量、CDF/PDF/PMF、期望、方差、协方差、相关系数、条件概率、条件期望、全期望公式、全方差公式、LLN、CLT。 |
| **经济金融应用** | 风险和收益、投资组合均值方差、保险定价、资产收益协方差、回归中的条件均值、计量估计量的样本分布、Monte Carlo 模拟。 |
| **后续连接** | Chapter 27 随机过程与 Markov 链, Chapter 28 Monte Carlo 与模拟，后续计量经济学中的估计、检验和因果推断。 |

**直觉盒子**

概率论的核心不是“算骰子题”，而是用数学方式描述不确定性。经济学中，价格、收益、收入、失业、消费、违约、政策变化都不是完全确定的。概率论给你三种语言：第一，什么事情可能发生；第二，各种结果发生的可能性多大；第三，在这些可能性下，平均结果和风险是多少。

## 1 为什么经济金融需要概率论

如果经济世界完全确定，那么动态优化和最优控制已经足够。但真实经济中的选择几乎总是面对不确定性：明年的收入是多少，股票回报是多少，贷款人是否违约，消费者是否购买，政策会不会改变，估计量离真实参数有多远。

### 1.1 从确定变量到随机变量

确定模型中，我们写

$$y = f(x).$$

随机模型中，我们写

$$Y = f(X, \varepsilon),$$

其中 $X$ 可能是可观测状态，$\varepsilon$ 是冲击。现在结果 $Y$ 不再是一个确定数，而是随机变量 random variable。

**常见误区**

随机变量不是“随机的变量名”。它是一个函数：把每个可能状态 $\omega$ 映射成一个数 $X(\omega)$。经济中我们通常不直接看到 $\omega$，而是看到由它产生的收入、收益、价格和选择。

### 1.2 概率论在经济金融中的四个角色

1. **风险描述：** 资产收益不是一个数，而是分布。均值描述平均回报，方差描述风险。
2. **决策准则：** 期望效用 $\mathbb{E}[u(c)]$ 是不确定性下选择的基础。
3. **信息更新：** 条件概率和 Bayes rule 描述获得信息后如何更新信念。
4. **统计推断：** 大数定律和中心极限定理解释为什么样本均值能估计总体均值，为什么估计量常近似正态。

## 2 概率空间：样本空间、事件与概率

### 2.1 三个组成部分

**定义 / 定理**

一个概率空间 probability space 通常写作

$$(\Omega, \mathcal{F}, \mathbb{P}).$$

其中：

- $\Omega$ 是样本空间 sample space，包含所有可能状态；
- $\mathcal{F}$ 是事件集合 event sigma-algebra，包含可以讨论概率的事件；
- $\mathbb{P}$ 是概率测度 probability measure，给每个事件分配概率。

在本科经济金融应用里，你不需要立刻进入测度论细节，但要理解下面的语言：事件是样本空间的子集，概率是事件发生的大小。

### 2.2 概率的基本性质

对任意事件 $A, B \in \mathcal{F}$：

$$0 \le \mathbb{P}(A) \le 1,$$

$$\mathbb{P}(\Omega) = 1,$$

若 $A \cap B = \varnothing$，则

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

一般地，

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).}$$

**Worked Example**

**例 1：违约事件与经济解释**

设 $D$ 表示贷款违约，$R$ 表示经济衰退。如果

$$\mathbb{P}(D) = 0.04, \qquad \mathbb{P}(R) = 0.10, \qquad \mathbb{P}(D \cap R) = 0.025,$$

则“违约或衰退至少发生一个”的概率为

$$\mathbb{P}(D \cup R) = 0.04 + 0.10 - 0.025 = 0.115.$$

这说明不能直接把 $0.04$ 和 $0.10$ 相加，因为两个事件可能同时发生。

# 3 随机变量与分布

## 3.1 随机变量

**定义 / 定理**

随机变量 random variable 是从样本空间到实数的映射：

$$X : \Omega \to \mathbb{R}.$$

如果 $X$ 只能取有限或可数多个值，称为离散随机变量。若 $X$ 在区间上连续取值，称为连续随机变量。

经济例子：

- 股票收益率 $R$；
- 家庭收入 $Y$；
- 违约指示变量 $D \in \{0, 1\}$；
- 明年生产率冲击 $Z$；
- 回归误差项 $u$。

## 3.2 CDF: 累积分布函数

**定义 / 定理**

随机变量 $X$ 的累积分布函数 cumulative distribution function, CDF 定义为

$$F_X(x) = \mathbb{P}(X \le x).$$

CDF 的优点是离散、连续都能用。它告诉我们 $X$ 落在某个阈值以下的概率。

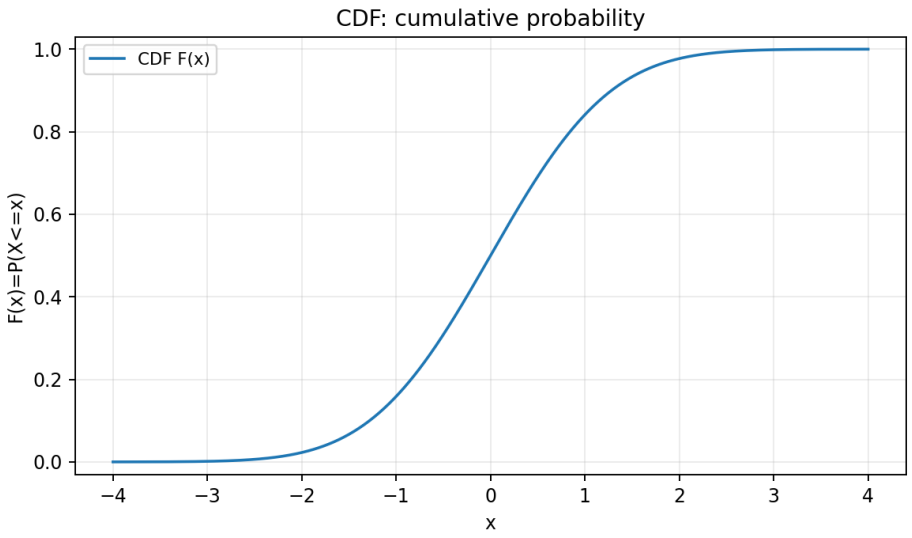



图 1: CDF 表示累积概率 $F(x) = \mathbb{P}(X \le x)$。

## 3.3 PMF 与 PDF

如果 $X$ 是离散变量，概率质量函数 probability mass function, PMF 为

$$p_X(x) = \mathbb{P}(X = x).$$

并满足

$$\sum_x p_X(x) = 1.$$

如果 $X$ 是连续变量，概率密度函数 probability density function, PDF 为 $f_X(x)$，满足

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t,$$

且

$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1.$$

区间概率由面积给出：

$$\boxed{\mathbb{P}(a \le X \le b) = \int_a^b f_X(x)\,\mathrm{d}x.}$$

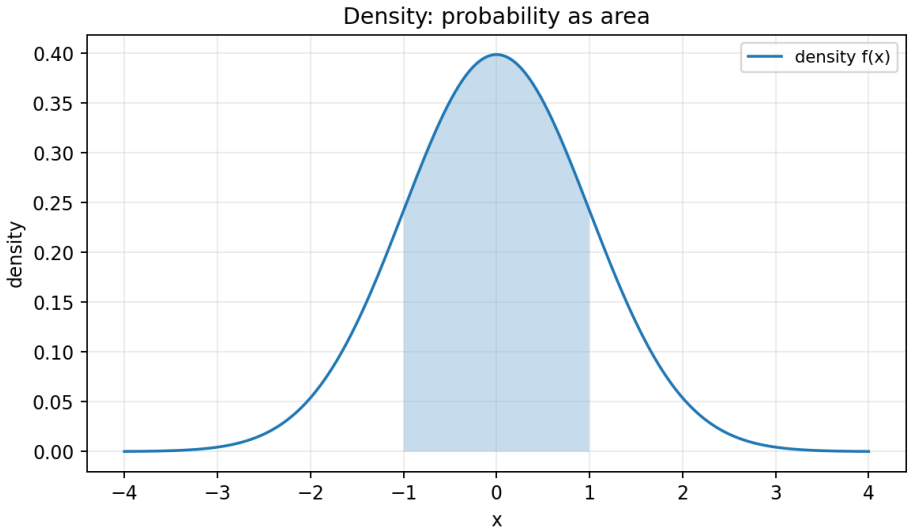



图 2: 连续变量中，概率不是点的高度，而是密度曲线下的面积。

**常见误区**

连续随机变量中，$f_X(x)$ 不是概率本身。通常

$$\mathbb{P}(X = x) = 0.$$

真正有意义的是区间概率 $\mathbb{P}(a \le X \le b)$。密度可以大于 1，只要总面积为 1 即可。

# 4 期望：平均结果与决策基准

## 4.1 离散型期望

若 $X$ 取值 $x_1, \dots, x_n$，概率为 $p_1, \dots, p_n$，则

$$\boxed{\mathbb{E}[X] = \sum_i x_i p_i.}$$

这是概率加权平均。

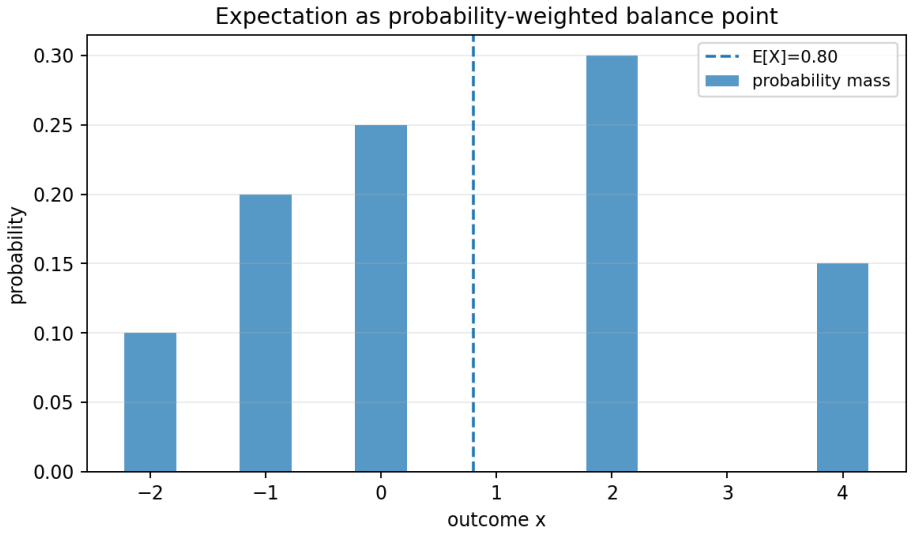



图 3: 期望可以理解为概率加权的平衡点。

## 4.2 连续型期望

若 $X$ 有密度 $f_X$，则

$$\boxed{\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x.}$$

更一般地，对函数 $g(X)$：

$$\boxed{\mathbb{E}[g(X)] = \int g(x) f_X(x)\,\mathrm{d}x}$$

或离散情形下

$$\mathbb{E}[g(X)] = \sum_{x} g(x) p_X(x).$$

## 4.3 期望的线性性

无论 $X, Y$ 是否独立，只要期望存在，都有

$$\boxed{\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c.}$$

这是概率论中最常用的性质之一。

**Worked Example**

**例 2：投资组合收益的期望**

设两个资产收益为 $R_1, R_2$，投资权重为 $w$ 和 $1-w$。组合收益为

$$R_p = wR_1 + (1-w)R_2.$$

则

$$\mathbb{E}[R_p] = w\mathbb{E}[R_1] + (1-w)\mathbb{E}[R_2].$$

注意这里不需要 $R_1, R_2$ 独立。期望的线性性永远成立。

### 4.4 Jensen 不等式入口

若 $u$ 是凹函数 concave function，则

$$\boxed{\mathbb{E}[u(X)] \le u(\mathbb{E}[X]).}$$

这就是风险厌恶的数学核心。若 $u$ 凹，随机消费的期望效用小于确定消费 $\mathbb{E}[X]$ 带来的效用。

**直觉盒子**

期望不等于“最可能结果”。均值、众数、中位数是不同概念。在偏态分布中，均值可能被极端值显著拉动。金融收益、收入和财富分布尤其需要注意这一点。

## 5 方差、标准差与风险

### 5.1 定义

**定义 / 定理**

随机变量 $X$ 的方差 variance 定义为

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

标准差 standard deviation 为

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

常用计算公式：

$$\boxed{\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.}$$

推导：

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[(X-\mu)^2] \\ &= \mathbb{E}[X^2 - 2\mu X + \mu^2] \\ &= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \\ &= \mathbb{E}[X^2] - \mu^2.\end{aligned}$$

其中 $\mu = \mathbb{E}[X]$。

### 5.2 尺度变换

若 $Y = aX + b$，则

$$\mathbb{E}[Y] = a\mathbb{E}[X] + b,$$

$$\boxed{\mathrm{Var}(Y) = a^2\,\mathrm{Var}(X).}$$

平移 $b$ 不改变风险，放大倍数 $a$ 会按平方放大方差。

**Worked Example**

**例 3：离散收益的均值和方差**

某资产有三种收益：

$$R = \begin{cases} -0.10, & p = 0.2, \\ 0.05, & p = 0.5, \\ 0.20, & p = 0.3. \end{cases}$$

期望收益为

$$\mathbb{E}[R] = (-0.10)(0.2) + (0.05)(0.5) + (0.20)(0.3) = 0.065.$$

再算二阶矩：

$$\mathbb{E}[R^2] = 0.01(0.2) + 0.0025(0.5) + 0.04(0.3) = 0.01525.$$

所以

$$\mathrm{Var}(R) = 0.01525 - 0.065^2 = 0.011025.$$

标准差为

$$\sigma_R = \sqrt{0.011025} \approx 0.105.$$

## 6 协方差、相关系数与组合风险

### 6.1 协方差

**定义 / 定理**

两个随机变量 $X, Y$ 的协方差 covariance 定义为

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

等价公式为

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

协方差衡量两个变量是否一起偏离各自均值。若 $X$ 高于均值时 $Y$ 也常高于均值，协方差为正；若一个高时另一个低，协方差为负。

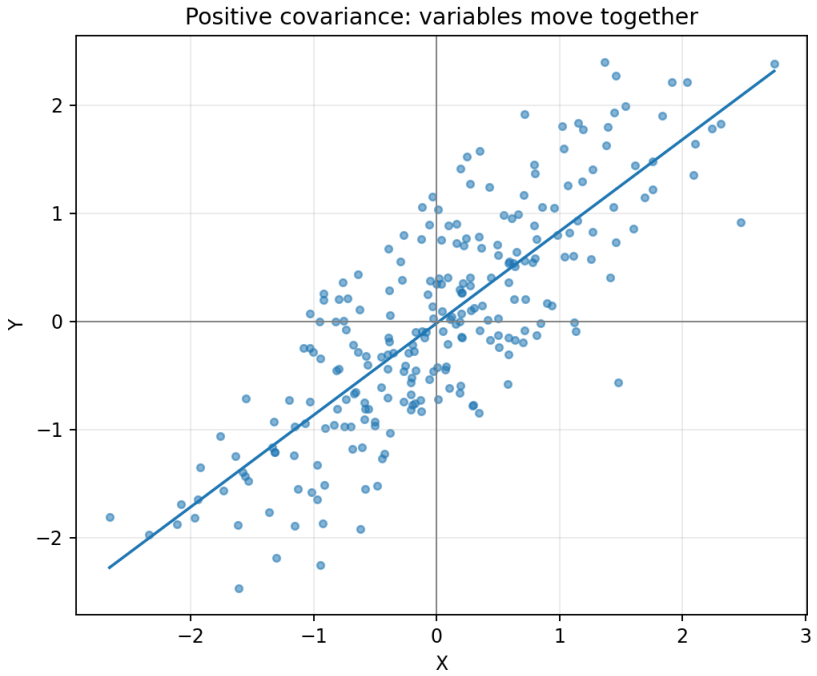



图 4: 正协方差表示两个变量倾向于同向变化。

## 6.2 相关系数

协方差受单位影响，所以常用相关系数 correlation：

$$\boxed{\mathrm{Corr}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}.}$$

它满足

$$-1\le \mathrm{Corr}(X,Y)\le 1.$$

## 6.3 投资组合方差

若

$$R_p=wR_1+(1-w)R_2,$$

则

$$\boxed{\mathrm{Var}(R_p)=w^2\sigma_1^2+(1-w)^2\sigma_2^2+2w(1-w)\,\mathrm{Cov}(R_1,R_2).}$$

若用相关系数 $\rho$ 表示：

$$\mathrm{Var}(R_p)=w^2\sigma_1^2+(1-w)^2\sigma_2^2+2w(1-w)\rho\sigma_1\sigma_2.$$

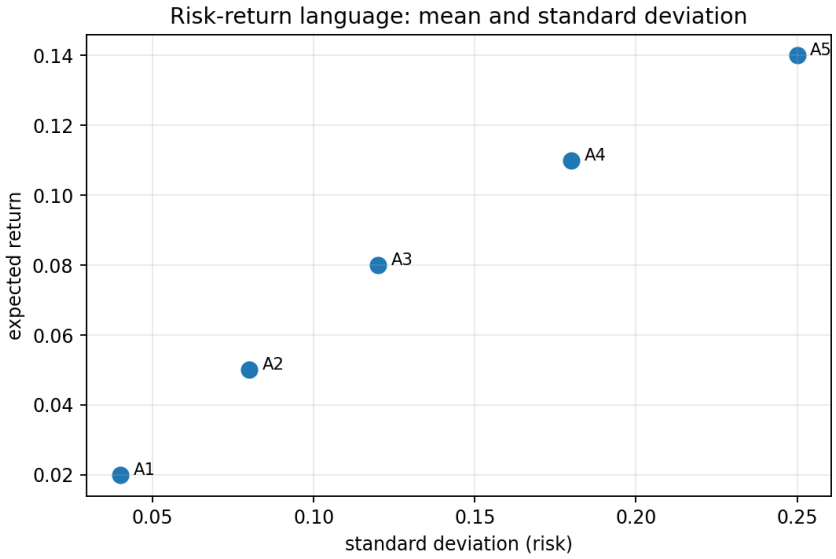



图 5：金融中的风险-收益语言通常用均值和标准差描述。

**直觉盒子**

分散化 diversification 的核心不是资产数量，而是协方差。如果两个资产相关性很高，放在一起并不能显著降低风险；如果相关性低甚至负相关，组合风险会明显下降。

# 7 条件概率与 Bayes 规则

## 7.1 条件概率

**定义 / 定理**

若 $\mathbb{P}(B) > 0$，事件 $A$ 在 $B$ 已发生条件下的概率为

$$\boxed{\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.}$$

这表示：当我们知道 $B$ 已经发生时，样本空间缩小到 $B$，然后看其中有多少也属于 $A$。

## 7.2 乘法公式与全概率公式

由条件概率定义可得

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A).$$

若 $A_1, \dots, A_n$ 构成互斥且完备的划分，则

$$\boxed{\mathbb{P}(B) = \sum_i \mathbb{P}(B \mid A_i)\mathbb{P}(A_i).}$$

这叫全概率公式。

### 7.3 Bayes 规则

**定义 / 定理**

Bayes 规则为

$$\boxed{\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}.}$$

若 $A_i$ 构成划分，则

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}.$$

**Worked Example**

**例 4：违约预警信号**

设某企业真实违约状态为 $D$，预警模型发出信号 $S$。已知

$$\mathbb{P}(D) = 0.02, \quad \mathbb{P}(S \mid D) = 0.80, \quad \mathbb{P}(S \mid D^c) = 0.10.$$

求看到预警信号后企业真的违约的概率。

由 Bayes 规则：

$$\mathbb{P}(D \mid S) = \frac{\mathbb{P}(S \mid D)\mathbb{P}(D)}{\mathbb{P}(S \mid D)\mathbb{P}(D) + \mathbb{P}(S \mid D^c)\mathbb{P}(D^c)}.$$

代入：

$$\mathbb{P}(D \mid S) = \frac{0.80 \cdot 0.02}{0.80 \cdot 0.02 + 0.10 \cdot 0.98} = \frac{0.016}{0.114} \approx 0.140.$$

虽然模型对违约企业有 80% 的识别率，但由于违约本身很少见，看到预警后的真实违约概率只有约 14%。

**常见误区**

这是金融风险和医疗检测中都非常常见的 base-rate problem。不要只看 $\mathbb{P}(S \mid D)$，真正决策常需要 $\mathbb{P}(D \mid S)$。

## 8 独立性

### 8.1 事件独立

事件 $A, B$ 独立表示

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).}$$

若 $\mathbb{P}(B) > 0$，等价于

$$\mathbb{P}(A \mid B) = \mathbb{P}(A).$$

也就是说，知道 $B$ 是否发生不会改变 $A$ 的概率。

### 8.2 随机变量独立

随机变量 $X, Y$ 独立意味着二者的联合分布可以分解为边际分布的乘积。离散情形下：

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

连续情形下：

$$f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

若 $X, Y$ 独立，则

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y],$$

因此

$$\mathrm{Cov}(X, Y) = 0.$$

但反过来一般不成立：协方差为 0 不必然独立。

**常见误区**

“无相关”不等于“独立”。如果 $Y = X^2$ 且 $X$ 对称分布，可能有 $\mathrm{Cov}(X, Y) = 0$，但 $Y$ 完全由 $X$ 决定，显然不独立。

## 9 条件期望：最优预测与回归入口

### 9.1 定义与直觉

条件期望 $\mathbb{E}[Y \mid X]$ 是给定 $X$ 信息后对 $Y$ 的平均预测。若 $X = x$，常写作

$$m(x) = \mathbb{E}[Y \mid X = x].$$

它是 $X$ 的函数。

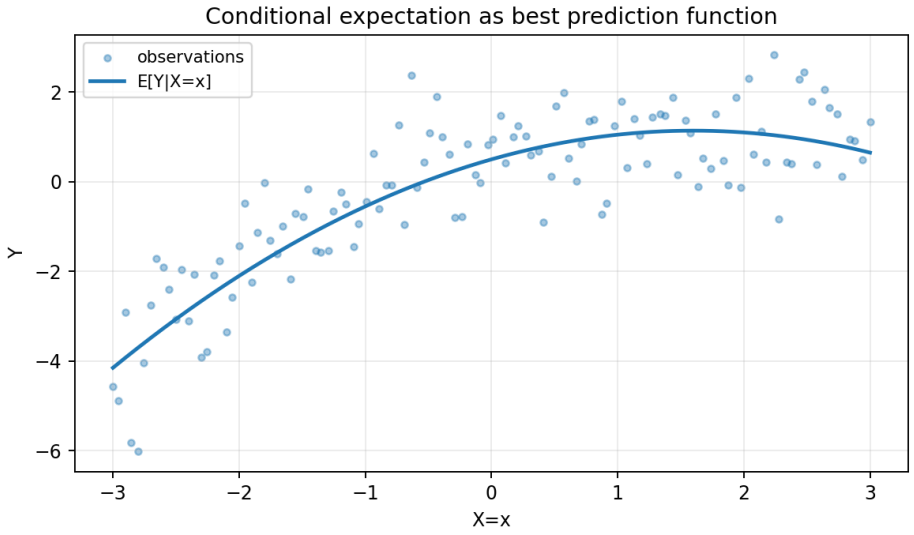



图 6: 条件期望 $\mathbb{E}[Y \mid X = x]$ 是给定 $X = x$ 后对 $Y$ 的最佳均方预测。

## 9.2 最优预测性质

**Advanced Core**

在所有只依赖 $X$ 的函数 $g(X)$ 中，条件期望 $\mathbb{E}[Y \mid X]$ 最小化均方预测误差：

$$\mathbb{E}[(Y - g(X))^2].$$

也就是说，

$$\boxed{\mathbb{E}[Y \mid X] = \arg\min_{g(X)} \mathbb{E}[(Y - g(X))^2].}$$

这是回归的概率论核心。

若定义误差

$$u = Y - \mathbb{E}[Y \mid X],$$

则

$$\boxed{\mathbb{E}[u \mid X] = 0.}$$

进一步，对任何合适函数 $h(X)$，有

$$\mathbb{E}[uh(X)] = 0.$$

这就是回归正交条件的来源。

### 9.3 迭代期望法则

**定义 / 定理**

全期望公式 law of iterated expectations:

$$\boxed{\mathbb{E}[Y] = \mathbb{E}\left[\mathbb{E}[Y \mid X]\right].}$$

更一般地，若信息集 $\mathcal{G} \subset \mathcal{H}$，则

$$\mathbb{E}[\mathbb{E}[Y \mid \mathcal{H}] \mid \mathcal{G}] = \mathbb{E}[Y \mid \mathcal{G}].$$

### 9.4 全方差公式

$$\boxed{\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y \mid X)] + \mathrm{Var}(\mathbb{E}[Y \mid X]).}$$

解释：总不确定性 = 给定 $X$ 后仍然存在的平均不确定性 + $X$ 本身解释掉的均值差异。

**直觉盒子**

这条公式对经济学很有用。例如收入不平等可以分解为组内不平等和组间不平等；资产收益风险可以分解为条件风险和预期收益变化带来的风险；预测误差可以分解为不可预测噪声和模型解释部分。

## 10 常用分布

### 10.1 Bernoulli 分布

若 $X \in \{0, 1\}$，且

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p,$$

则 $X \sim \mathrm{Bernoulli}(p)$。

$$\mathbb{E}[X] = p, \qquad \mathrm{Var}(X) = p(1 - p).$$

经济例子：是否违约、是否就业、是否购买、政策是否通过。

### 10.2 Binomial 分布

若 $X$ 是 $n$ 次独立 Bernoulli 试验中成功次数，则

$$X \sim \mathrm{Binomial}(n, p).$$

概率为

$$\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}.$$

矩为

$$\mathbb{E}[X]=np,\qquad \mathrm{Var}(X)=np(1-p).$$

### 10.3 Uniform 分布

若 $X\sim U[a,b]$，则密度为

$$f(x)=\frac{1}{b-a},\qquad a\le x\le b.$$

期望和方差为

$$\mathbb{E}[X]=\frac{a+b}{2},\qquad \mathrm{Var}(X)=\frac{(b-a)^2}{12}.$$

### 10.4 Normal 分布

若 $X\sim\mathcal{N}(\mu,\sigma^2)$，则密度为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

正态分布重要不是因为所有经济变量都正态，而是因为中心极限定理使大量估计量和样本均值近似正态。

标准化：

$$Z=\frac{X-\mu}{\sigma}\sim\mathcal{N}(0,1).$$

## 11 大数定律：样本均值为什么可靠

### 11.1 样本均值

设 $X_1,\dots,X_n$ 独立同分布，均值为 $\mu$，样本均值为

$$\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i.$$

**定义 / 定理**

**大数定律 law of large numbers, LLN.** 在适当条件下，

$$\boxed{\bar{X}_n\to\mu}$$

当 $n\to\infty$。更精确地说，样本均值依概率收敛到总体均值。

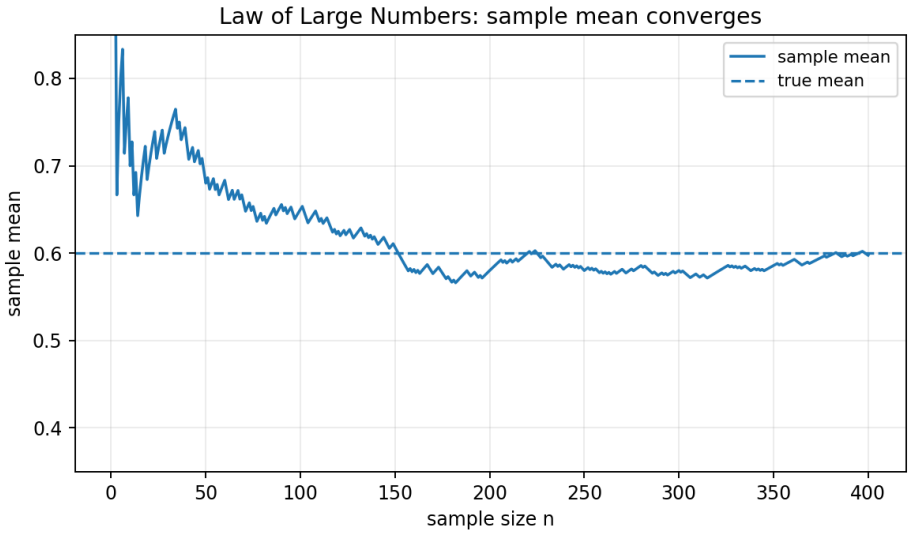



图 7: 大数定律说明样本均值会随着样本量增加而靠近真实均值。

### 11.2 为什么 LLN 对计量重要

样本均值估计总体均值:

$$\bar{X}_n \approx \mathbb{E}[X].$$

回归、矩估计、GMM、最大似然等方法都依赖类似逻辑：样本矩会逼近总体矩。

**常见误区**

大数定律不表示样本路径会单调接近真实均值，也不表示小样本一定准确。它说的是样本量趋于无穷时的长期性质。

## 12 中心极限定理：为什么正态分布无处不在

**定义 / 定理**

**中心极限定理 central limit theorem, CLT.** 设 $X_1, \ldots, X_n$ 独立同分布，均值为 $\mu$，方差为 $\sigma^2 < \infty$。则

$$\boxed{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \Rightarrow \mathcal{N}(0,1).}$$

这里 $\Rightarrow$ 表示分布收敛。

等价地，当 $n$ 大时，

$$\bar{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

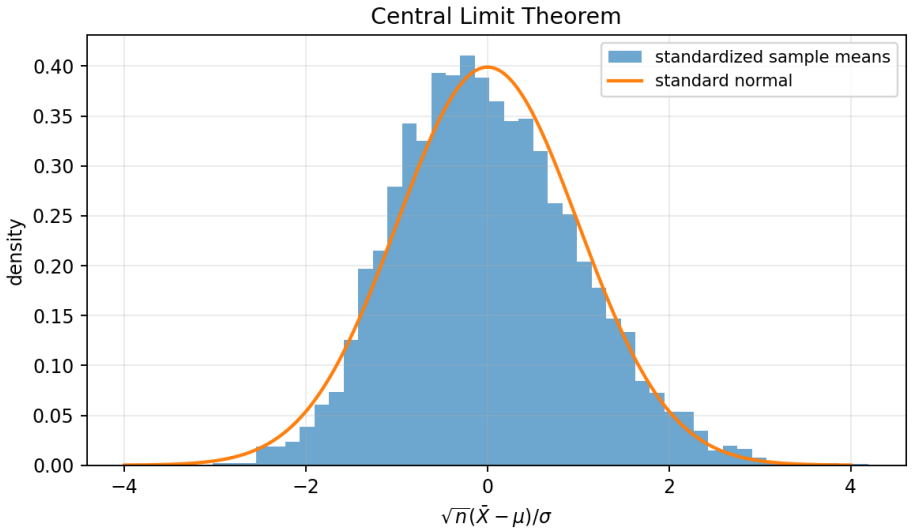



图 8: 即使原始变量不是正态，标准化样本均值也常近似正态。

### 12.1 经济金融解释

CLT 的意义不是“所有变量都正态”，而是“很多平均型统计量近似正态”。例如：

- 样本平均收益；
- 回归系数估计量；
- 样本矩；
- Monte Carlo 模拟均值；
- 大样本下的最大似然估计量。

## 13 经济金融应用 1：风险溢价与期望收益

资产定价中，收益 $R$ 是随机变量。投资者关心的不只是 $\mathbb{E}[R]$，还关心风险。一个简单风险溢价可写成

$$\mathbb{E}[R_i] - R_f,$$

其中 $R_f$ 是无风险收益。

若使用均值-方差偏好，投资者可能最大化

$$\mathbb{E}[R_p] - \frac{\gamma}{2}\mathrm{Var}(R_p),$$

其中 $\gamma > 0$ 表示风险厌恶程度。这里的概率工具直接进入优化目标。

**Worked Example**

**例 5：两资产组合的方差**

设 $\sigma_1 = 0.10$，$\sigma_2 = 0.20$，相关系数 $\rho = 0.25$，权重 $w = 0.6$。则

$$\mathrm{Var}(R_p) = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\rho\sigma_1\sigma_2.$$

代入：

$$\mathrm{Var}(R_p) = 0.36(0.01) + 0.16(0.04) + 2(0.6)(0.4)(0.25)(0.10)(0.20).$$

$$\mathrm{Var}(R_p) = 0.0036 + 0.0064 + 0.0024 = 0.0124.$$

标准差为

$$\sigma_p = \sqrt{0.0124} \approx 0.111.$$

## 14 经济金融应用 2：回归为什么是条件期望

计量经济学里常写

$$Y = m(X) + u, \qquad \mathbb{E}[u \mid X] = 0.$$

这里

$$m(X) = \mathbb{E}[Y \mid X].$$

若进一步假设 $m(X)$ 是线性的：

$$\mathbb{E}[Y \mid X] = X'\beta,$$

就得到线性回归模型。

OLS 的一阶条件

$$\mathbb{E}[X(Y - X'\beta)] = 0$$

正是条件期望误差正交性质的线性版本。

**直觉盒子**

回归不是凭空给一条线。它是在用 $X$ 所包含的信息预测 $Y$。最一般的最佳预测函数是 $\mathbb{E}[Y \mid X]$，线性回归只是把这个函数限制为线性形式。

## 15 经济金融应用 3：保险、违约与条件概率

保险定价中，若损失为随机变量 $L$，精算公平保费为

$$\pi = \mathbb{E}[L].$$

若保险公司观察到风险类型 $X$，更合理的保费是

$$\pi(X) = \mathbb{E}[L \mid X].$$

这就是条件期望的应用。

违约风险中，贷款人的关键问题也是条件概率：

$$\mathbb{P}(D \mid \text{borrower characteristics}).$$

信用评分、机器学习违约预测、logit/probit 模型，本质上都在估计条件违约概率。

## 16 常见错误与辨析

**常见误区**

**误区 1：把密度当概率。**连续变量中 $f(x)$ 是密度，不是 $\mathbb{P}(X = x)$。

**误区 2：认为期望就是最可能结果。**期望是加权平均，不一定是众数或中位数。

**误区 3：独立和不相关混淆。**独立推出协方差为零，但协方差为零不一定推出独立。

**误区 4：大数定律被理解成小样本保证。** LLN 是渐近结论，小样本仍然可能偏离很大。

**误区 5：中心极限定理被理解成原始变量正态。** CLT 说的是标准化样本均值近似正态，不是说原始变量一定正态。

**误区 6：条件概率方向搞反。** $\mathbb{P}(A \mid B)$ 和 $\mathbb{P}(B \mid A)$ 通常完全不同。

## 17 本章小结：概念 - 公式 - 经济含义

| 概念 | 核心公式 | 经济金融含义 |
|---|---|---|
| 概率空间 | $(\Omega, \mathcal{F}, \mathbb{P})$ | 描述不确定性的数学环境。 |
| 随机变量 | $X : \Omega \to \mathbb{R}$ | 收益、收入、冲击等不确定结果。 |
| CDF | $F(x) = \mathbb{P}(X \le x)$ | 阈值以下的累计概率。 |
| PDF | $\mathbb{P}(a \le X \le b) = \int_a^b f(x)\,\mathrm{d}x$ | 概率是面积。 |
| 期望 | $\mathbb{E}[X]$ | 平均结果、精算公平价格。 |
| 方差 | $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2]$ | 风险、波动、离散程度。 |
| 协方差 | $\mathrm{Cov}(X, Y)$ | 两变量共同变化，组合风险核心。 |
| 条件概率 | $\mathbb{P}(A \mid B)$ | 信息更新、违约预测。 |
| 条件期望 | $\mathbb{E}[Y \mid X]$ | 最优预测、回归函数。 |

| | | |
|---|---|---|
| LLN | $\bar{X}_n \to \mu$ | 样本均值估计总体均值。 |
| CLT | $\sqrt{n}(\bar{X}-\mu)/\sigma \Rightarrow N(0,1)$ | 大样本推断和标准误基础。 |

## 18 练习题

### A 组：基础计算

A1. 设事件 $A, B$ 满足 $\mathbb{P}(A)=0.4$，$\mathbb{P}(B)=0.5$，$\mathbb{P}(A\cap B)=0.2$。求 $\mathbb{P}(A\cup B)$ 和 $\mathbb{P}(A\mid B)$。

A2. 随机变量 $X$ 取值 $0,1,3$，概率分别为 $0.2, 0.5, 0.3$。求 $\mathbb{E}[X]$ 和 $\mathrm{Var}(X)$。

A3. 若 $Y=2X+5$，且 $\mathbb{E}[X]=3$，$\mathrm{Var}(X)=4$，求 $\mathbb{E}[Y]$ 和 $\mathrm{Var}(Y)$。

A4. 若 $X\sim \mathrm{Bernoulli}(p)$，推导 $\mathbb{E}[X]$ 和 $\mathrm{Var}(X)$。

A5. 若 $X, Y$ 独立，$\mathbb{E}[X]=2$，$\mathbb{E}[Y]=4$，求 $\mathbb{E}[XY]$ 和 $\mathrm{Cov}(X,Y)$。

A6. 若 $X\sim\mathcal{N}(10,4)$，求 $\mathbb{P}(X\le 10)$ 与标准化变量 $Z$。

### B 组：概念判断

B1. 为什么连续随机变量满足 $\mathbb{P}(X=x)=0$，但密度 $f(x)$ 可以大于 0?

B2. 独立和不相关有什么区别?

B3. 为什么期望的线性性不要求随机变量独立?

B4. 解释 $\mathbb{E}[Y\mid X]$ 为什么可以看成最佳预测函数。

B5. 大数定律和中心极限定理分别解决什么问题?

### C 组：混合题

C1. 某贷款组合中，违约概率为 $0.03$，违约时损失为 $100$，不违约时损失为 $0$。求期望损失和损失方差。

C2. 两资产收益标准差分别为 $0.15$ 和 $0.25$，相关系数为 $0.4$，组合权重为 $0.5, 0.5$。求组合方差。

C3. 设 $\mathbb{P}(D)=0.01$，$\mathbb{P}(S\mid D)=0.9$，$\mathbb{P}(S\mid D^c)=0.05$。求 $\mathbb{P}(D\mid S)$。

C4. 设 $Y=X+u$，且 $\mathbb{E}[u\mid X]=0$。证明 $\mathbb{E}[Y\mid X]=X$。

C5. 若 $X_1,\dots,X_n$ iid，均值为 $\mu$，方差为 $\sigma^2$，求 $\mathbb{E}[\bar{X}_n]$ 和 $\mathrm{Var}(\bar{X}_n)$。

### D 组：经济金融建模题

D1. 用期望和方差解释为什么两个资产均值相同但风险不同，会影响风险厌恶投资者选择。

D2. 在保险定价中，为什么公平保费应等于 $\mathbb{E}[L]$? 如果能观察到类型 $X$，为什么保费应改为 $\mathbb{E}[L\mid X]$?

D3. 在线性回归中，为什么假设 $\mathbb{E}[u \mid X] = 0$ 比 $\mathbb{E}[u] = 0$ 更强?

D4. 解释为什么金融收益低相关性会带来分散化收益。

D5. Monte Carlo 模拟中，为什么重复模拟次数增加可以降低模拟均值误差? 用 LLN/CLT 解释。

### E 组: 挑战 / 证明题

E1. 证明 $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$。

E2. 证明 $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$。

E3. 证明若 $X, Y$ 独立，则 $\mathrm{Cov}(X, Y) = 0$。

E4. 推导全方差公式 $\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y \mid X)] + \mathrm{Var}(\mathbb{E}[Y \mid X])$。

E5. 设 $\bar{X}_n$ 是 iid 样本均值。用方差计算说明为什么样本均值随着 $n$ 增大更集中。

# 19 练习题参考答案

## A 组答案

A1.

$$\mathbb{P}(A \cup B) = 0.4 + 0.5 - 0.2 = 0.7.$$

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{0.2}{0.5} = 0.4.$$

A2.

$$\mathbb{E}[X] = 0(0.2) + 1(0.5) + 3(0.3) = 1.4.$$

$$\mathbb{E}[X^2] = 0^2(0.2) + 1^2(0.5) + 3^2(0.3) = 3.2.$$

$$\mathrm{Var}(X) = 3.2 - 1.4^2 = 1.24.$$

A3.

$$\mathbb{E}[Y] = 2\mathbb{E}[X] + 5 = 11.$$

$$\mathrm{Var}(Y) = 2^2\,\mathrm{Var}(X) = 16.$$

A4. Bernoulli 变量满足 $X^2 = X$，所以

$$\mathbb{E}[X] = p, \qquad \mathbb{E}[X^2] = p.$$

因此

$$\mathrm{Var}(X) = p - p^2 = p(1-p).$$

A5. 独立给出

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = 8.$$

所以

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0.$$

A6. 因正态分布关于均值对称，

$$\mathbb{P}(X \le 10) = 0.5.$$

标准差为 2，所以

$$Z = \frac{X - 10}{2} \sim \mathcal{N}(0, 1).$$

## B 组答案

B1. 连续变量的概率由区间面积给出，单点没有长度，因此单点概率为 0。密度是单位长度附近概率的强度，不是点概率本身。

B2. 独立表示联合分布可分解为边际分布乘积；不相关只表示协方差为 0。独立通常推出不相关，但不相关不一定推出独立。

B3. 因为期望本质上是加权求和或积分，求和与积分本身是线性的，所以不要求随机变量独立。

B4. 在所有只依赖 $X$ 的预测函数中，$\mathbb{E}[Y \mid X]$ 最小化均方误差 $\mathbb{E}[(Y - g(X))^2]$，因此是最佳均方预测。

B5. LLN 说明样本均值会靠近总体均值，解决一致性问题；CLT 说明标准化样本均值近似正态，解决大样本分布和推断问题。

## C 组答案

C1. 损失 $L = 100D$，其中 $D \sim Bernoulli(0.03)$。期望损失：

$$\mathbb{E}[L] = 100(0.03) = 3.$$

方差：

$$\mathrm{Var}(L) = 100^2 \,\mathrm{Var}(D) = 10000(0.03)(0.97) = 291.$$

C2.

$$\mathrm{Var}(R_p) = 0.5^2(0.15^2) + 0.5^2(0.25^2) + 2(0.5)(0.5)(0.4)(0.15)(0.25).$$

$$= 0.005625 + 0.015625 + 0.0075 = 0.02875.$$

C3.

$$\mathbb{P}(D \mid S) = \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.05 \cdot 0.99} = \frac{0.009}{0.0585} \approx 0.1538.$$

C4.

$$\mathbb{E}[Y \mid X] = \mathbb{E}[X + u \mid X] = X + \mathbb{E}[u \mid X] = X.$$

C5.

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n} \sum_i \mathbb{E}[X_i] = \mu.$$

独立同分布给出

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_i \mathrm{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

## D 组答案

D1. 若两个资产均值相同，风险厌恶者会偏好方差更低的资产，因为凹效用下波动会降低期望效用。均值只描述平均收益，方差描述围绕均值的不确定性。

D2. 公平保费等于期望损失，因为长期平均赔付为 $\mathbb{E}[L]$。若能观察类型 $X$，不同类型风险不同，保费应按条件期望 $\mathbb{E}[L \mid X]$ 定价，否则会产生逆向选择。

D3. $\mathbb{E}[u]=0$ 只说明误差总体平均为零；$\mathbb{E}[u\mid X]=0$ 说明在每个 $X$ 条件下误差平均都为零。后者排除了误差与解释变量的系统关系，是回归因果/预测解释中更强的条件。

D4. 组合方差包含协方差项。相关性低时，某个资产高于均值时另一个资产不一定也高于均值，因此波动会相互抵消，降低组合总风险。

D5. 模拟均值是 iid 模拟结果的样本均值。LLN 保证模拟次数增加时样本均值靠近真实期望；CLT 给出误差大约为 $O(1/\sqrt{n})$，所以模拟次数越多，误差越小。

## E 组答案

E1. 令 $\mu=\mathbb{E}[X]$。则

$$\mathrm{Var}(X)=\mathbb{E}[(X-\mu)^2]=\mathbb{E}[X^2-2\mu X+\mu^2]=\mathbb{E}[X^2]-2\mu^2+\mu^2=\mathbb{E}[X^2]-\mu^2.$$

E2.

$$\mathrm{Var}(aX+b)=\mathbb{E}[(aX+b-\mathbb{E}(aX+b))^2].$$

因 $\mathbb{E}(aX+b)=a\mathbb{E}[X]+b$，所以

$$aX+b-a\mathbb{E}[X]-b=a(X-\mathbb{E}[X]).$$

因此

$$\mathrm{Var}(aX+b)=\mathbb{E}[a^2(X-\mathbb{E}X)^2]=a^2\,\mathrm{Var}(X).$$

E3. 独立意味着 $\mathbb{E}[XY]=\mathbb{E}[X]\mathbb{E}[Y]$，所以

$$\mathrm{Cov}(X,Y)=\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]=0.$$

E4. 记 $m(X)=\mathbb{E}[Y\mid X]$。有

$$Y-\mathbb{E}[Y]=(Y-m(X))+(m(X)-\mathbb{E}[Y]).$$

两边平方取期望:

$$\mathrm{Var}(Y)=\mathbb{E}[(Y-m)^2]+\mathbb{E}[(m-\mathbb{E}Y)^2]+2\mathbb{E}[(Y-m)(m-\mathbb{E}Y)].$$

交叉项为 0，因为

$$\mathbb{E}[Y-m\mid X]=0.$$

第一项是 $\mathbb{E}[\mathrm{Var}(Y\mid X)]$，第二项是 $\mathrm{Var}(\mathbb{E}[Y\mid X])$。故

$$\mathrm{Var}(Y)=\mathbb{E}[\mathrm{Var}(Y\mid X)]+\mathrm{Var}(\mathbb{E}[Y\mid X]).$$

E5. 由 C5,

$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

随着 $n$ 增大，样本均值的方差趋于 0，说明样本均值越来越集中在 $\mu$ 附近。这是大数定律的方差直觉。

## 20 后续学习连接

本章建立了随机变量和分布的基础。下一章将把单个随机变量推进到随时间演化的随机对象：随机过程、Markov 链、随机游走、AR(1)、白噪声和平稳性。你需要记住本章三条主线：

distribution describes uncertainty,

expectation and variance summarize risk,

conditional expectation is prediction and regression’s foundation.

后续的计量经济学、金融资产定价和随机动态规划都会反复调用这些概念。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 27
# 条件期望与回归的数学基础

Chapter 26 建立了概率、随机变量、分布、期望与方差。本章进入计量经济学真正的数学入口：条件期望 conditional expectation 与回归 regression。你会看到，OLS 不是一套机械公式，而是一个投影问题；$E[Y \mid X]$ 不是一句口号，而是给定信息时对 $Y$ 的最优均方预测；$E[\varepsilon \mid X] = 0$ 与 $E[X\varepsilon] = 0$ 则是识别、因果推断和回归理论的核心语言。

math4econ personalized notes
2026

# 目录

# Chapter 27 总览

| | |
|---|---|
| **本章主题** | 条件期望、投影、population regression、OLS 正交条件与计量经济学基础。 |
| **核心问题** | 给定信息 $X$ 后，如何用最小均方误差预测 $Y$？线性回归为什么等价于投影？$E[\varepsilon \mid X]=0$ 和 $E[X\varepsilon]=0$ 有何区别？什么时候回归系数能解释因果效应？ |
| **数学工具** | 条件分布、条件期望、迭代期望法则、最小均方预测、Hilbert space 投影、正交条件、best linear predictor、population OLS、样本 OLS。 |
| **经济应用** | 收入函数、教育回报、消费函数、资产收益因子模型、风险溢价预测、因果识别中的 omitted variable bias 与 exogeneity。 |
| **后续连接** | Chapter 28 的收敛概念与极限定理，Chapter 29 的多元正态与金融风险，计量经济学中的一致性、渐近正态、IV、DID、GMM。 |

**直觉盒子**

本章最重要的一句话是：

$$\boxed{E[Y \mid X] \text{ 是在知道} X \text{ 后对} Y \text{ 的最佳均方预测。}}$$

如果你只允许使用线性函数 $a+bX$ 来预测 $Y$，得到的是 best linear predictor；如果你允许使用所有关于 $X$ 的函数来预测 $Y$，得到的就是 conditional expectation。OLS 是这个思想的样本版本。

## 1 为什么这一章重要

你以后学计量经济学，会反复看到这些式子：

$$Y_i = X_i'\beta + u_i,$$

$$E[u_i \mid X_i] = 0,$$

$$E[X_i u_i] = 0,$$

$$\hat{\beta} = (X'X)^{-1}X'Y.$$

很多教材会很快把它们变成估计和检验公式，但如果你没有真正理解条件期望和投影，这些式子会变成一堆机械符号。

本章要解决的不是“OLS 怎么按计算器算”，而是以下问题：

- 为什么条件期望是最优预测?
- 为什么回归残差和解释变量正交?
- $E[u \mid X] = 0$ 为什么比 $E[Xu] = 0$ 更强?
- 为什么 omitted variable 会造成偏误?
- 线性投影系数和因果效应有什么关系，又什么时候没有关系?
- 金融因子模型里的 beta 为什么也是投影系数?

## 1.1 从概率论到计量经济学

Chapter 26 中，期望是一个总体平均:

$$E[Y] = \int y f_Y(y)\,\mathrm{d}y.$$

可是经济学中我们常常不是无条件地问 $Y$ 的平均值，而是问：在知道某些信息后，$Y$ 的平均值是多少？例如:

- 已知教育年限，工资的平均值是多少?
- 已知资产市场 beta，预期收益是多少?
- 已知家庭收入，消费的平均值是多少?
- 已知政策处理状态，潜在结果的条件均值是多少?

这些问题都需要条件期望。

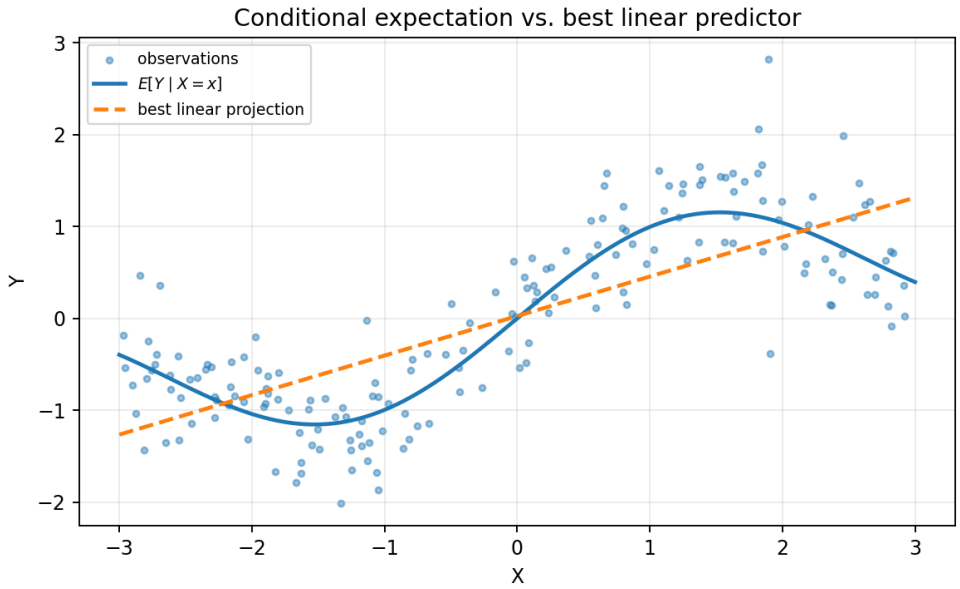



图 1: 条件期望 $E[Y \mid X = x]$ 可以是非线性的；线性回归只是用直线逼近它。

# 2 条件分布与条件期望

## 2.1 离散情形

若 $X, Y$ 是离散随机变量，条件概率为

$$P(Y = y \mid X = x) = \frac{P(Y = y, X = x)}{P(X = x)}, \qquad P(X = x) > 0.$$

于是条件期望为

$$\boxed{E[Y \mid X = x] = \sum_{y} y\, P(Y = y \mid X = x).}$$

这只是把无条件平均换成“在 $X = x$ 这个子群体中的平均”。

## 2.2 连续情形

若 $(X, Y)$ 有联合密度 $f_{X,Y}(x, y)$，边际密度为

$$f_X(x) = \int f_{X,Y}(x, y)\, \mathrm{d}y,$$

则条件密度为

$$\boxed{f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}, \qquad f_X(x) > 0.}$$

条件期望为

$$\boxed{E[Y \mid X = x] = \int y f_{Y|X}(y \mid x)\, \mathrm{d}y.}$$

**直觉盒子**

条件期望可以理解为“局部平均”。把所有 $X$ 接近 $x$ 的样本放在一起，对它们的 $Y$ 取平均，就得到对 $E[Y \mid X = x]$ 的经验近似。非参数回归、kernel regression、local polynomial regression 都是这个直觉的精细版本。

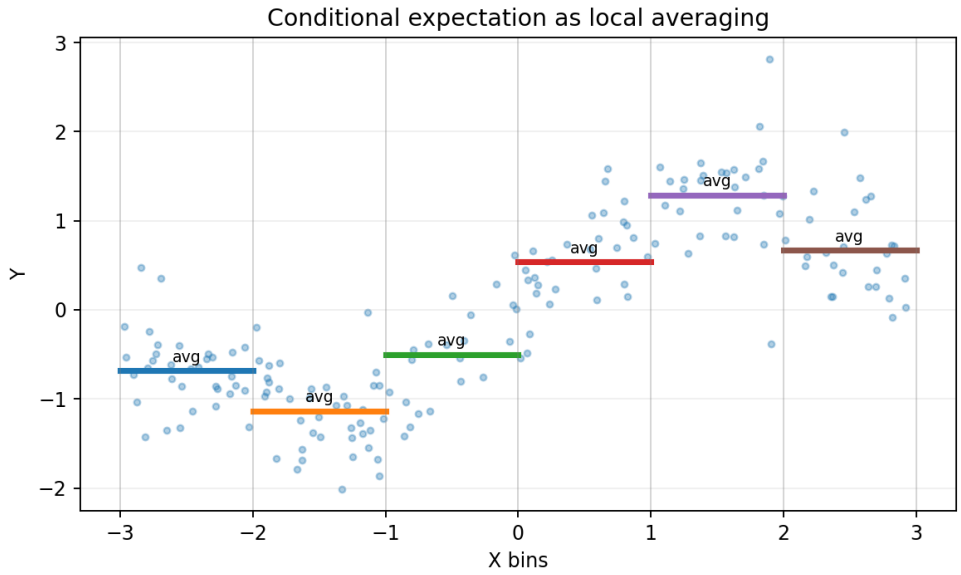



图 2：条件期望的经验直觉：按照 $X$ 分箱，对每个箱中的 $Y$ 求平均。

### 2.3 条件期望是随机变量

严格地说，$E[Y \mid X = x]$ 是一个关于 $x$ 的函数。记作

$$m(x) = E[Y \mid X = x].$$

而

$$E[Y \mid X]$$

是把随机变量 $X$ 代入这个函数后得到的新随机变量：

$$E[Y \mid X] = m(X).$$

这点非常重要。比如如果

$$E[Y \mid X = x] = 2 + 3x,$$

那么

$$E[Y \mid X] = 2 + 3X.$$

**常见误区**

常见错误：把 $E[Y \mid X]$ 当成一个普通常数。它通常不是常数，而是一个随机变量，因为它随 $X$ 的取值而变化。只有当 $Y$ 与 $X$ 均值独立时，$E[Y \mid X] = E[Y]$ 才不随 $X$ 变化。

## 3 条件期望的基本性质

## 3.1 可测函数可以拿出来

若 $g(X)$ 是只依赖 $X$ 的函数，则

$$\boxed{E[g(X)Y \mid X] = g(X)E[Y \mid X].}$$

直觉：在已经知道 $X$ 的情况下，$g(X)$ 已经是已知数，所以可以像常数一样提出。

## 3.2 迭代期望法则 Law of Iterated Expectations

**定义 / 定理**

**迭代期望法则.** 若期望存在，则

$$\boxed{E[E[Y \mid X]] = E[Y].}$$

更一般地，若信息集满足 $X$ 是 $(X, Z)$ 的一部分，则

$$\boxed{E[E[Y \mid X, Z] \mid X] = E[Y \mid X].}$$

这个定理是计量经济学里最常用的概率工具之一。它说明：先在细分组里取平均，再对细分组平均，等于在粗分组里取平均。

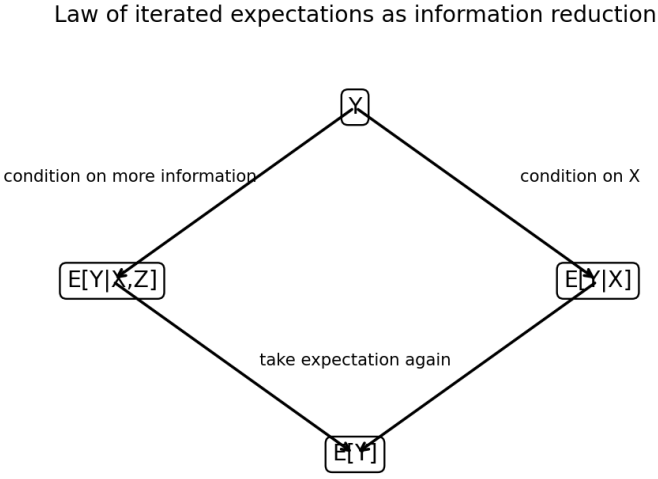



图 3: 迭代期望法则：从更细信息到更粗信息，再到无条件平均。

## 3.3 条件期望的正交性质

令

$$\eta = Y - E[Y \mid X].$$

则

$$E[\eta \mid X] = 0.$$

因为

$$E[Y - E[Y \mid X] \mid X] = E[Y \mid X] - E[Y \mid X] = 0.$$

进一步，对任何只依赖于 $X$ 的函数 $g(X)$，有

$$\boxed{E[g(X)\eta] = 0.}$$

证明：

$$E[g(X)\eta] = E[E[g(X)\eta \mid X]] = E[g(X)E[\eta \mid X]] = 0.$$

**直觉盒子**

条件期望留下的预测误差 $Y - E[Y \mid X]$ 和所有关于 $X$ 的信息都不相关。换句话说，如果你已经用 $X$ 做了最佳预测，那么残差里不应该再有任何可由 $X$ 预测的系统成分。

# 4 条件期望为什么是最佳预测

## 4.1 最小均方误差问题

给定随机变量 $X$，我们想用某个函数 $g(X)$ 预测 $Y$。目标是最小化均方误差：

$$\min_{g} E[(Y - g(X))^2].$$

结论是：

$$\boxed{g^*(X) = E[Y \mid X].}$$

## 4.2 证明：平方损失分解

令

$$m(X) = E[Y \mid X].$$

对任何候选函数 $g(X)$，有

$$Y - g(X) = Y - m(X) + m(X) - g(X).$$

平方并取期望：

$$\begin{aligned} E[(Y - g(X))^2] = {} & E[(Y - m(X))^2] \\ & + E[(m(X) - g(X))^2] \\ & + 2E[(Y - m(X))(m(X) - g(X))]. \end{aligned}$$

最后一项为 0，因为 $m(X)-g(X)$ 是 $X$ 的函数，而 $Y-m(X)$ 与任何 $X$ 的函数正交。因此

$$\boxed{E[(Y-g(X))^2]=E[(Y-m(X))^2]+E[(m(X)-g(X))^2].}$$

第二项非负，且只有当 $g(X)=m(X)$ 时为 0。所以最佳预测是

$$g^*(X)=m(X)=E[Y\mid X].$$

**Advanced Core**

这就是 Hilbert space 投影定理的概率版本。随机变量可以看成一个带内积的空间，内积为

$$\langle A,B\rangle=E[AB].$$

所有关于 $X$ 的函数构成一个子空间。$E[Y\mid X]$ 是 $Y$ 在这个子空间上的投影。

# 5 从条件期望到回归

## 5.1 真正的条件均值函数

在总体中，最根本的对象是

$$\boxed{m(x)=E[Y\mid X=x].}$$

如果 $m(x)$ 是线性的，例如

$$m(x)=\alpha+\beta x,$$

那么线性回归模型就是正确指定的条件均值模型。

但一般情况下，$m(x)$ 可以是非线性的。此时线性回归并不是“真实函数”，而是对 $m(x)$ 的线性逼近。

## 5.2 Best Linear Predictor

如果我们只允许用线性函数预测：

$$g(X)=\alpha+\beta X,$$

那么最佳线性预测 best linear predictor 由

$$(\alpha^*,\beta^*)=\arg\min_{\alpha,\beta}E[(Y-\alpha-\beta X)^2]$$

定义。

一阶条件为

$$E[Y-\alpha^*-\beta^*X]=0,$$

$$E[X(Y-\alpha^*-\beta^*X)]=0.$$

令

$$u=Y-\alpha^*-\beta^*X.$$

则

$$\boxed{E[u]=0,\qquad E[Xu]=0.}$$

这就是线性投影的正交条件。

### 5.3 单变量回归系数公式

由一阶条件可得

$$\alpha^*=E[Y]-\beta^*E[X].$$

代入第二个条件：

$$E[X(Y-\alpha^*-\beta^*X)]=0.$$

整理得

$$\boxed{\beta^*=\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}.}$$

截距为

$$\boxed{\alpha^*=E[Y]-\beta^*E[X].}$$

**Worked Example**

**例 1：教育回报的线性投影**

设 $Y$ 是 log wage，$X$ 是 schooling years。Best linear predictor 是

$$Y\approx\alpha+\beta X.$$

总体中的 $\beta$ 为

$$\beta=\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}.$$

如果 $E[Y\mid X]$ 真的是线性的，$\beta$ 是条件均值函数的斜率。如果 $E[Y\mid X]$ 非线性，$\beta$ 是用一条直线拟合整个条件均值函数的最佳线性斜率。

## 6 回归的投影几何

### 6.1 随机变量空间的内积

定义随机变量 $A,B$ 的内积：

$$\langle A,B\rangle=E[AB].$$

如果

$$E[AB] = 0,$$

就说 $A$ 和 $B$ 正交。Best linear predictor 的条件

$$E[Xu] = 0$$

就是残差 $u$ 与解释变量 $X$ 正交。

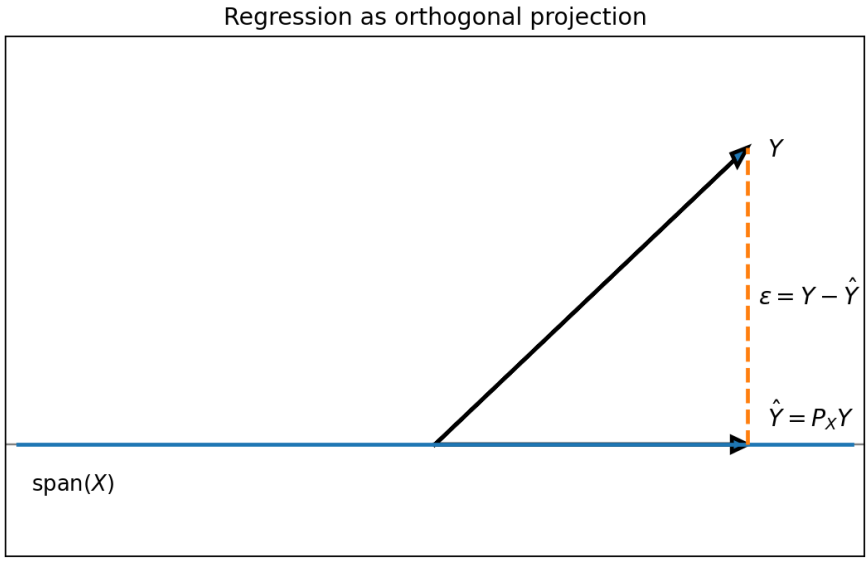



图 4: 回归是投影：拟合值是 $Y$ 在解释变量张成空间上的投影，残差与该空间正交。

## 6.2 样本 OLS 的几何版本

样本中，令

$$\boldsymbol{y} = (y_1, \dots, y_n)', \qquad X = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}.$$

OLS 解为

$$\boxed{\hat{\beta} = (X'X)^{-1}X'\boldsymbol{y}.}$$

拟合值为

$$\hat{\boldsymbol{y}} = X\hat{\beta}.$$

残差为

$$\hat{\boldsymbol{u}} = \boldsymbol{y} - X\hat{\beta}.$$

一阶条件为

$$\boxed{X'\hat{\boldsymbol{u}} = 0.}$$

这说明样本残差与每一列解释变量正交。

### 6.3 总体 OLS 的矩阵形式

如果解释变量是向量 $X \in \mathbb{R}^k$，定义 best linear predictor:

$$\beta^* = \arg\min_b E[(Y - X'b)^2].$$

一阶条件为

$$E[X(Y - X'\beta^*)] = 0.$$

整理得

$$E[XX']\beta^* = E[XY].$$

若 $E[XX']$ 可逆，则

$$\boxed{\beta^* = (E[XX'])^{-1}E[XY].}$$

这就是 population OLS coefficient。

**方法模板**

**记忆方式**

$$\text{sample OLS:} \quad \hat{\beta} = (X'X)^{-1}X'Y.$$

$$\text{population OLS:} \quad \beta = (E[XX'])^{-1}E[XY].$$

样本平均 $\frac{1}{n}X'X$ 对应总体期望 $E[XX']$；样本平均 $\frac{1}{n}X'Y$ 对应总体期望 $E[XY]$。

## 7 条件均值零与正交条件

### 7.1 $E[u \mid X] = 0$

如果模型写成

$$Y = X'\beta + u,$$

并且满足

$$\boxed{E[u \mid X] = 0,}$$

则说明给定 $X$ 后，误差项没有系统性均值。

这时

$$E[Y \mid X] = E[X'\beta + u \mid X] = X'\beta + E[u \mid X] = X'\beta.$$

所以 $X'\beta$ 就是条件均值函数。

### 7.2 $E[Xu] = 0$

由 $E[u \mid X] = 0$ 可以推出

$$E[Xu] = 0.$$

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 28
# 收敛概念与极限定理

Chapter 27 把条件期望和回归解释为投影。本章进一步回答一个计量经济学最核心的问题：为什么样本中的估计量可以代表总体中的参数？当样本量越来越大时，样本均值、OLS、MLE、GMM 估计量会怎样靠近真实值，又为什么可以用正态分布近似做置信区间和假设检验？本章的关键词是 convergence, LLN, CLT, Slutsky, continuous mapping, delta method, consistency, asymptotic normality。

math4econ personalized notes
2026

# 目录

**Chapter 28 总览** **2**

**1 为什么这一章重要** **2**
1.1 从“算出一个数”到“知道这个数可靠吗” . . . 3
1.2 经济金融中的典型场景 . . . 3

**2 随机变量序列与估计量序列** **3**
2.1 为什么要研究序列 . . . 3

**3 四种收敛概念** **4**
3.1 依概率收敛 convergence in probability . . . 4
3.2 几乎处处收敛 almost sure convergence . . . 4
3.3 依分布收敛 convergence in distribution . . . 5
3.4 均方收敛 mean-square convergence . . . 5

**4 大数定律 LLN** **6**
4.1 弱大数定律 . . . 6
4.2 证明骨架 . . . 6
4.3 LLN 在计量中的作用 . . . 7

**5 中心极限定理 CLT** **7**
5.1 标准形式 . . . 8
5.2 为什么是 $\sqrt{n}$ . . . 9

**6 一致性与渐近正态** **9**
6.1 一致性 consistency . . . 9
6.2 渐近正态 asymptotic normality . . . 10

**7 连续映射定理 CMT** **10**
7.1 为什么重要 . . . 10

**8 Slutsky 定理** **11**
8.1 标准误为什么可以被估计值替代 . . . 11

**9 Delta Method** **12**
9.1 一维 Delta method . . . 12
9.2 证明直觉 . . . 12
9.3 多维 Delta method . . . 13

**10 OLS 的一致性与渐近正态** **13**
10.1 模型与估计量 . . . . . 14
10.2 一致性 . . . . . 14
10.3 渐近正态 . . . . . 14

**11 MLE 与 GMM 的大样本逻辑入口** **15**
11.1 M-estimation 的统一形式 . . . . . 16
11.2 GMM 的矩条件 . . . . . 16

**12 置信区间与假设检验** **16**
12.1 单个参数的置信区间 . . . . . 16
12.2 t statistic . . . . . 17

**13 金融应用：均值、波动率与 Sharpe ratio** **17**
13.1 估计期望收益 . . . . . 17
13.2 估计 Sharpe ratio . . . . . 17

**14 常见误区** **18**

**15 本章小结图谱** **18**

**16 练习题** **19**
A 组：基础计算 . . . . . 19
B 组：概念判断 . . . . . 19
C 组：混合题 . . . . . 19
D 组：经济金融应用题 . . . . . 19
E 组：挑战 / 证明题 . . . . . 20

**17 练习题参考答案** **21**
A 组答案 . . . . . 21
B 组答案 . . . . . 21
C 组答案 . . . . . 22
D 组答案 . . . . . 22
E 组答案 . . . . . 23

**18 下一章连接** **23**

# Chapter 28 总览

| | |
|---|---|
| **本章主题** | 收敛概念、LLN、CLT、Slutsky 定理、连续映射定理、Delta method，以及它们在计量推断中的作用。 |
| **核心问题** | 样本均值为什么接近总体均值？估计量为什么会一致？为什么标准化后的估计误差常常近似正态？对估计量做非线性变换后，渐近分布如何改变？ |
| **数学工具** | convergence in probability, convergence in distribution, almost sure convergence, LLN, CLT, continuous mapping theorem, Slutsky theorem, delta method。 |
| **经济应用** | OLS 一致性与渐近正态、MLE/GMM 推断、标准误、置信区间、假设检验、风险溢价估计、样本均值和样本方差的极限性质。 |
| **后续连接** | Chapter 29 多元正态与二次型、Chapter 30 Monte Carlo 与计算统计、计量经济学中的大样本理论、金融风险估计。 |

**直觉盒子**

本章最重要的主线是：

$$\boxed{\text{样本对象} \longrightarrow \text{总体对象} \longrightarrow \text{渐近分布} \longrightarrow \text{推断}}$$

LLN 告诉你“估计量会靠近哪里”；CLT 告诉你“靠近的误差大概长什么样”；Slutsky 和 Delta method 告诉你“把估计量放进更复杂公式后，极限结果如何变化”。

## 1 为什么这一章重要

计量经济学和金融实证中，你几乎每天都会看到这些句子：

$$\hat{\beta} \xrightarrow{p} \beta,$$

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, V),$$

$$\frac{\hat{\beta} - \beta_0}{\widehat{se}(\hat{\beta})} \xrightarrow{d} \mathcal{N}(0, 1).$$

如果没有本章的收敛语言，这些式子会像一堆神秘符号。它们真正表达的是：样本量大时，估计量靠近真值，并且误差经过合适缩放后可以用正态分布近似。

### 1.1 从“算出一个数”到“知道这个数可靠吗”

假设你用 500 个家庭数据估计教育回报，得到

$$\hat{\beta} = 0.08.$$

单独这个数字不够。你还需要知道：

- 如果换一批样本，估计值会差多少？
- 样本量变大时，它是否会靠近真实教育回报？
- 标准误怎么来？
- 什么时候可以说结果显著？

这些问题都不是有限样本代数能解决的，而是大样本理论 asymptotic theory 的问题。

### 1.2 经济金融中的典型场景

- **宏观：**用样本平均增长率估计长期增长率。
- **金融：**用历史平均收益估计期望收益，用样本协方差估计风险矩阵。
- **计量：**用 OLS 估计处理效应或需求弹性。
- **结构模型：**用 MLE 或 GMM 估计偏好参数、技术参数、状态转移概率。
- **因果推断：**用大样本正态近似构造置信区间和 p-value。

**常见误区**

大样本理论不是说“样本一大就万事大吉”。LLN 和 CLT 需要条件，比如独立性、弱相关、有限方差、识别条件、矩条件连续性等。经济金融数据常有时间相关、异方差、厚尾和结构变化，所以你要知道定理在说什么，也要知道它的边界在哪里。

## 2 随机变量序列与估计量序列

### 2.1 为什么要研究序列

估计量通常依赖样本量。样本量为 $n$ 时的估计量记为

$$\hat{\theta}_n.$$

当 $n$ 改变时，估计量本身也变，所以我们研究的是随机变量序列

$$\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n, \ldots$$

收敛概念就是用来描述这个序列的长期行为。

例如样本均值

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

随着 $n$ 增大通常会靠近 $\mu = E[X]$。但因为 $\bar{X}_n$ 是随机变量，所以“靠近”必须用概率语言定义。

# 3 四种收敛概念

## 3.1 依概率收敛 convergence in probability

**定义 / 定理**

若对任意 $\varepsilon > 0$，都有

$$P(|X_n - X| > \varepsilon) \to 0,$$

则称 $X_n$ 依概率收敛到 $X$，记作

$$X_n \xrightarrow{p} X.$$

直觉：当 $n$ 很大时，$X_n$ 偏离 $X$ 超过任意固定容忍度的概率趋于零。

在计量经济学中，一致性 consistency 通常就是依概率收敛：

$$\hat{\theta}_n \xrightarrow{p} \theta_0.$$

也就是说，估计量会以概率意义靠近真实参数。

## 3.2 几乎处处收敛 almost sure convergence

**定义 / 定理**

如果

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1,$$

则称 $X_n$ 几乎处处收敛到 $X$，记作

$$X_n \xrightarrow{a.s.} X.$$

几乎处处收敛比依概率收敛更强。它说的是：除了概率为零的一小部分样本路径外，$X_n(\omega)$ 作为一个普通数字序列会收敛到 $X(\omega)$。

## 3.3 依分布收敛 convergence in distribution

**定义 / 定理**

若 $X_n$ 的分布函数 $F_n(x)$ 在 $X$ 的分布函数 $F(x)$ 的所有连续点处满足

$$F_n(x) \to F(x),$$

则称 $X_n$ 依分布收敛到 $X$，记作

$$X_n \xrightarrow{d} X.$$

依分布收敛关心的是分布形状，而不是随机变量本身是否靠近。CLT 就是一个依分布收敛定理:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1).$$

## 3.4 均方收敛 mean-square convergence

若

$$E[(X_n - X)^2] \to 0,$$

则称 $X_n$ 均方收敛到 $X$。均方收敛通常推出依概率收敛。证明用 Chebyshev 不等式:

$$P(|X_n - X| > \varepsilon) \le \frac{E[(X_n - X)^2]}{\varepsilon^2} \to 0.$$

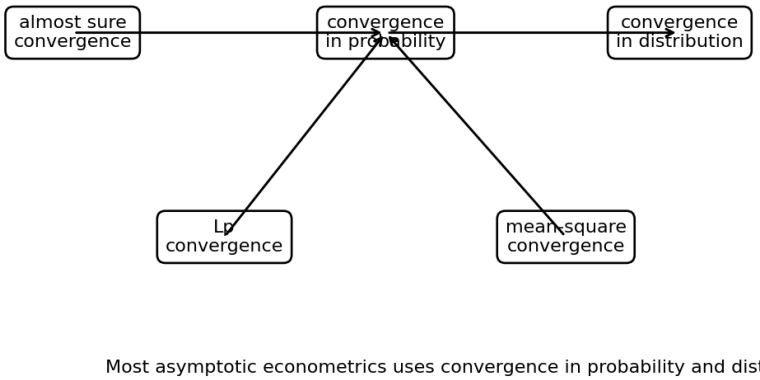



图 1: 几种收敛概念之间的常见关系。计量渐近理论最常使用依概率收敛和依分布收敛。

**方法模板**

**记忆方式**

- 一致性：通常看 $\hat{\theta}_n \xrightarrow{p} \theta_0$。
- 渐近分布：通常看 $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} Z$。
- 大样本推断：先证明一致性，再证明渐近正态，再估计方差。

# 4 大数定律 LLN

## 4.1 弱大数定律

**定义 / 定理**

设 $X_1, X_2, \dots$ 独立同分布，且 $E[X_i] = \mu$，$\mathrm{Var}(X_i) = \sigma^2 < \infty$。则样本均值

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

满足

$$\boxed{\bar{X}_n \xrightarrow{p} \mu.}$$

这称为弱大数定律 weak law of large numbers。

## 4.2 证明骨架

样本均值的期望为

$$E[\bar{X}_n] = \mu.$$

若独立同分布，则

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\sigma^2}{n}.$$

由 Chebyshev 不等式，任意 $\varepsilon > 0$，

$$P(|\bar{X}_n - \mu| > \varepsilon) \le \frac{\mathrm{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0.$$

所以

$$\bar{X}_n \xrightarrow{p} \mu.$$

**直觉盒子**

LLN 的核心不是“样本均值等于总体均值”，而是“样本越大，样本均值偏离总体均值的概率越小”。随机性没有消失，但平均化让误差变小。

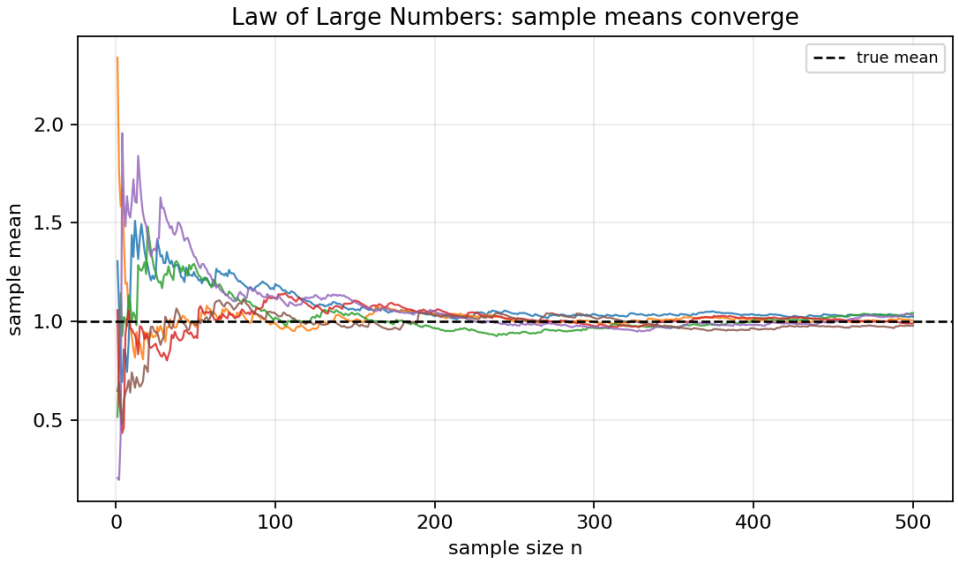



图 2: 大数定律模拟：不同样本路径的样本均值都逐渐靠近总体均值。

### 4.3 LLN 在计量中的作用

以 population OLS 为例:

$$\beta_0 = (E[X_i X_i'])^{-1} E[X_i Y_i].$$

样本 OLS 是

$$\hat{\beta} = (\frac{1}{n}\sum_i X_i X_i')^{-1} \left(\frac{1}{n}\sum_i X_i Y_i\right).$$

LLN 告诉我们，在合适条件下，

$$\frac{1}{n}\sum_i X_i X_i' \xrightarrow{p} E[X_i X_i'],$$

$$\frac{1}{n}\sum_i X_i Y_i \xrightarrow{p} E[X_i Y_i].$$

再用连续映射定理，就得到

$$\hat{\beta} \xrightarrow{p} \beta_0.$$

## 5 中心极限定理 CLT

## 5.1 标准形式

**定义 / 定理**

设 $X_1, X_2, \ldots$ 独立同分布，且

$$E[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2 < \infty.$$

则

$$\boxed{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1).}$$

这称为中心极限定理 central limit theorem。

CLT 的形式也可以写成：

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

或者近似写为：

$$\bar{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

**直觉盒子**

LLN 告诉你 $\bar{X}_n$ 会靠近 $\mu$。CLT 进一步告诉你：靠近的误差大约是 $1/\sqrt{n}$ 量级，而且形状近似正态。标准误 standard error 正是这个误差尺度。

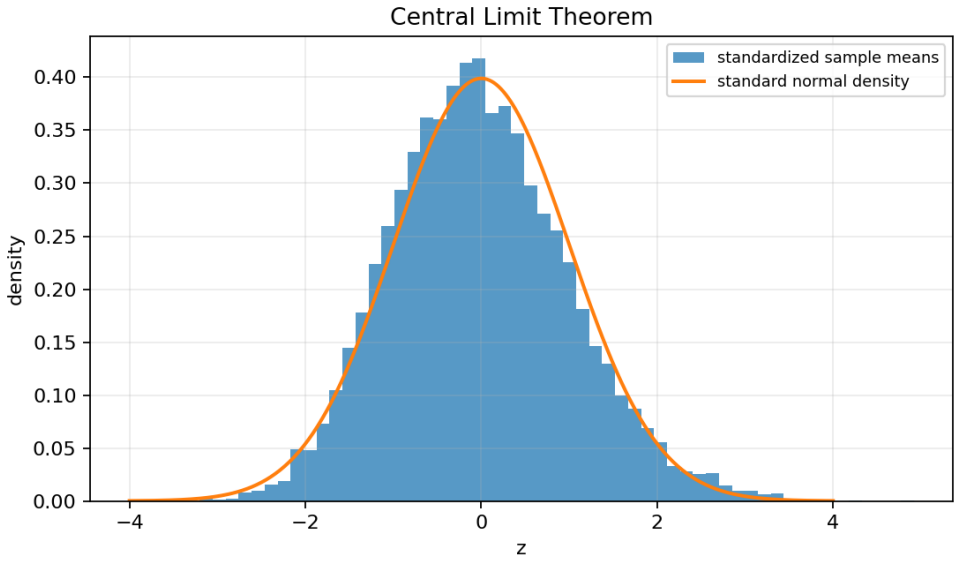



图 3: CLT 模拟：即使原始变量来自指数分布，标准化样本均值也逐渐接近标准正态。

### 5.2 为什么是 $\sqrt{n}$

样本均值的方差是

$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

所以 $\bar{X}_n - \mu$ 的标准差是

$$\frac{\sigma}{\sqrt{n}}.$$

如果不乘 $\sqrt{n}$，误差会收缩到 0，分布极限退化成一个点；乘以 $\sqrt{n}$ 后，误差被放大回非退化尺度：

$$\sqrt{n}(\bar{X}_n - \mu).$$

这就是渐近正态里的 $\sqrt{n}$ rate。

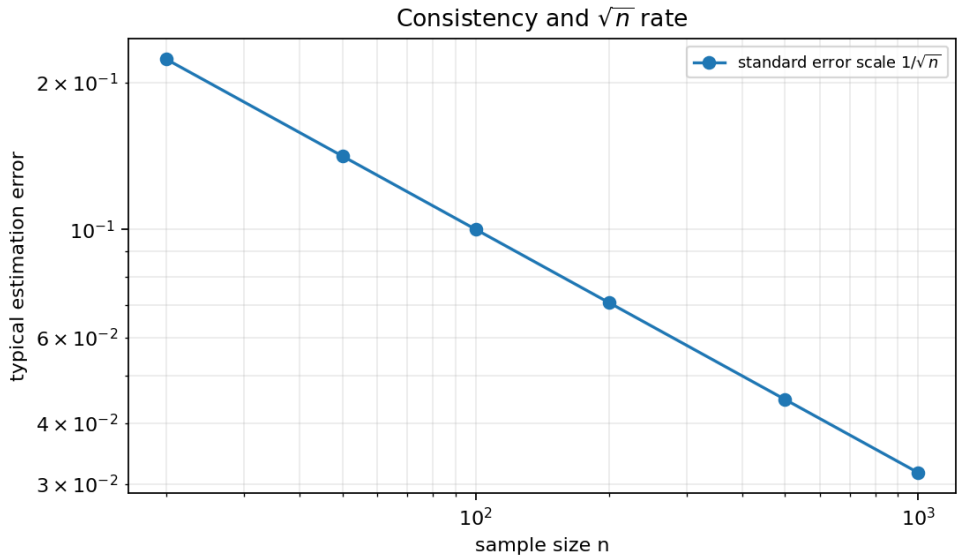



图 4: 估计误差的典型量级通常按 $1/\sqrt{n}$ 收缩。样本量扩大 4 倍，标准误大约减半。

## 6 一致性与渐近正态

### 6.1 一致性 consistency

**定义 / 定理**

若估计量 $\hat{\theta}_n$ 满足

$$\hat{\theta}_n \xrightarrow{p} \theta_0,$$

则称 $\hat{\theta}_n$ 是 $\theta_0$ 的一致估计量 consistent estimator。

一致性的含义是：样本量趋于无穷时，估计量会靠近真实参数。它不要求有限样本无偏。一个估计量可以有小偏误，但只要偏误随样本量消失，仍然可能一致。

### 6.2 渐近正态 asymptotic normality

**定义 / 定理**

若

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, V),$$

则称 $\hat{\theta}_n$ 渐近正态。此时近似有

$$\hat{\theta}_n \approx \mathcal{N}\left(\theta_0, \frac{V}{n}\right).$$

渐近正态是构造标准误、置信区间和假设检验的基础。例如若 $\widehat{V} \xrightarrow{p} V$，则

$$\frac{\hat{\theta}_n - \theta_0}{\sqrt{\widehat{V}/n}} \xrightarrow{d} \mathcal{N}(0,1).$$

## 7 连续映射定理 CMT

**定义 / 定理**

若 $X_n \xrightarrow{p} X$，且 $g$ 在 $X$ 的取值点连续，则

$$g(X_n) \xrightarrow{p} g(X).$$

若 $X_n \xrightarrow{d} X$，且 $g$ 连续，则

$$g(X_n) \xrightarrow{d} g(X).$$

这称为连续映射定理 continuous mapping theorem。

### 7.1 为什么重要

我们经常不是直接使用估计量本身，而是使用估计量的函数。例如：

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_i (X_i - \bar{X}_n)^2,$$

$$\widehat{Sharpe} = \frac{\bar{R} - r_f}{\hat{\sigma}_R},$$

$$\hat{\beta} = (\hat{Q}_{XX})^{-1}\hat{Q}_{XY}.$$

CMT 告诉我们：只要里面的样本矩收敛，且变换函数连续，组合出来的估计量也收敛。

**Worked Example**

**例 1：样本方差的一致性**

设 $E[X^2] < \infty$。样本方差可以写为

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i X_i^2 - \bar{X}_n^2.$$

由 LLN,

$$\frac{1}{n}\sum_i X_i^2 \xrightarrow{p} E[X^2], \qquad \bar{X}_n \xrightarrow{p} E[X] = \mu.$$

由连续映射定理,

$$\bar{X}_n^2 \xrightarrow{p} \mu^2.$$

所以

$$\hat{\sigma}^2 \xrightarrow{p} E[X^2] - \mu^2 = \mathrm{Var}(X).$$

# 8 Slutsky 定理

**定义 / 定理**

若

$$X_n \xrightarrow{d} X, \qquad Y_n \xrightarrow{p} c,$$

其中 $c$ 是常数，则

$$X_n + Y_n \xrightarrow{d} X + c,$$

$$X_n Y_n \xrightarrow{d} cX,$$

如果 $c \neq 0$，则

$$\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}.$$

这称为 Slutsky 定理。

## 8.1 标准误为什么可以被估计值替代

CLT 给出

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1).$$

但真实的 $\sigma$ 通常未知。我们用样本标准差 $\hat{\sigma}$ 代替。若

$$\hat{\sigma} \xrightarrow{p} \sigma,$$

则由 Slutsky 定理,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}} \xrightarrow{d} \mathcal{N}(0,1).$$

这就是 t-statistic 大样本近似正态的核心理由。

**直觉盒子**

Slutsky 定理的直觉：如果一个随机对象有非退化极限分布，另一个对象只是稳定地收敛到常数，那么把后者替进去不会改变一阶渐近分布。也就是说，在大样本里，“估计出来的标准误”可以代替“真实标准误”。

# 9 Delta Method

## 9.1 一维 Delta method

**定义 / 定理**

若

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, V),$$

且 $g$ 在 $\theta_0$ 处可微，则

$$\boxed{\sqrt{n}\left(g(\hat{\theta}_n) - g(\theta_0)\right) \xrightarrow{d} \mathcal{N}\left(0, [g'(\theta_0)]^2 V\right).}$$

这称为 Delta method。

## 9.2 证明直觉

用一阶 Taylor 展开：

$$g(\hat{\theta}_n) = g(\theta_0) + g'(\theta_0)(\hat{\theta}_n - \theta_0) + \text{higher order term}.$$

所以

$$\sqrt{n}[g(\hat{\theta}_n) - g(\theta_0)] \approx g'(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0).$$

若右边渐近正态，则左边也渐近正态，方差乘以 $[g'(\theta_0)]^2$。

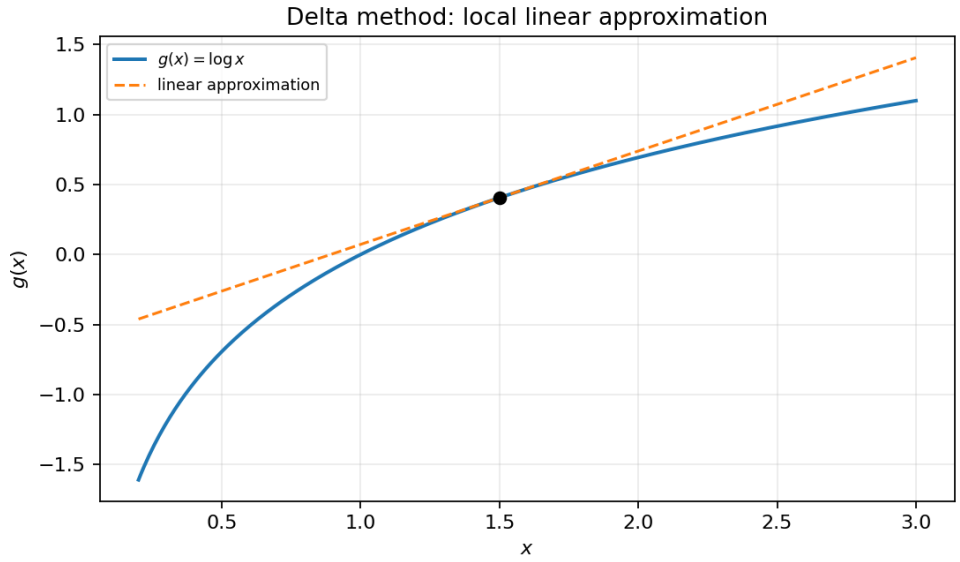



图 5: Delta method 的本质是对非线性函数在真值附近做局部线性化。

## 9.3 多维 Delta method

若

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, V),$$

且 $g : \mathbb{R}^k \to \mathbb{R}^m$ 在 $\boldsymbol{\theta}_0$ 可微，Jacobian 为

$$G = \frac{\partial g}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0),$$

则

$$\boxed{\sqrt{n}[g(\hat{\boldsymbol{\theta}}_n) - g(\boldsymbol{\theta}_0)] \xrightarrow{d} \mathcal{N}(\mathbf{0}, GVG').}$$

**Worked Example**

**例 2：log 变换的渐近分布**

若

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, V),$$

且 $\theta > 0$。令 $g(\theta) = \log \theta$，则

$$g'(\theta) = \frac{1}{\theta}.$$

由 Delta method，

$$\sqrt{n}(\log \hat{\theta} - \log \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{V}{\theta^2}\right).$$

这在估计弹性、增长率、比率和对数风险指标时很常见。

# 10 OLS 的一致性与渐近正态

## 10.1 模型与估计量

考虑线性模型

$$Y_i = X_i'\beta_0 + u_i,$$

其中 $X_i$ 是 $k \times 1$ 向量。OLS 为

$$\hat{\beta} = \left( \sum_i X_i X_i' \right)^{-1} \sum_i X_i Y_i.$$

代入 $Y_i = X_i'\beta_0 + u_i$：

$$\hat{\beta} - \beta_0 = \left( \sum_i X_i X_i' \right)^{-1} \sum_i X_i u_i.$$

等价地，

$$\hat{\beta} - \beta_0 = \left( \frac{1}{n} \sum_i X_i X_i' \right)^{-1} \left( \frac{1}{n} \sum_i X_i u_i \right).$$

## 10.2 一致性

若

$$\frac{1}{n} \sum_i X_i X_i' \xrightarrow{p} Q,$$

且 $Q$ 可逆，并且

$$\frac{1}{n} \sum_i X_i u_i \xrightarrow{p} 0,$$

则

$$\hat{\beta} - \beta_0 \xrightarrow{p} Q^{-1} 0 = 0.$$

所以

$$\boxed{\hat{\beta} \xrightarrow{p} \beta_0.}$$

这里的关键是矩条件：

$$E[X_i u_i] = 0.$$

LLN 把样本矩带到总体矩。

## 10.3 渐近正态

写成 $\sqrt{n}$ 形式：

$$\sqrt{n}(\hat{\beta} - \beta_0) = \left( \frac{1}{n} \sum_i X_i X_i' \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_i X_i u_i \right).$$

若

$$\frac{1}{n}\sum_i X_i X_i' \xrightarrow{p} Q,$$

且由 CLT,

$$\frac{1}{\sqrt{n}}\sum_i X_i u_i \xrightarrow{d} \mathcal{N}(0,\Omega),$$

则由 Slutsky,

$$\boxed{\sqrt{n}(\hat{\beta}-\beta_0) \xrightarrow{d} \mathcal{N}(0, Q^{-1}\Omega Q^{-1}).}$$

其中

$$\Omega = E[X_i X_i' u_i^2]$$

在异方差情形下成立。

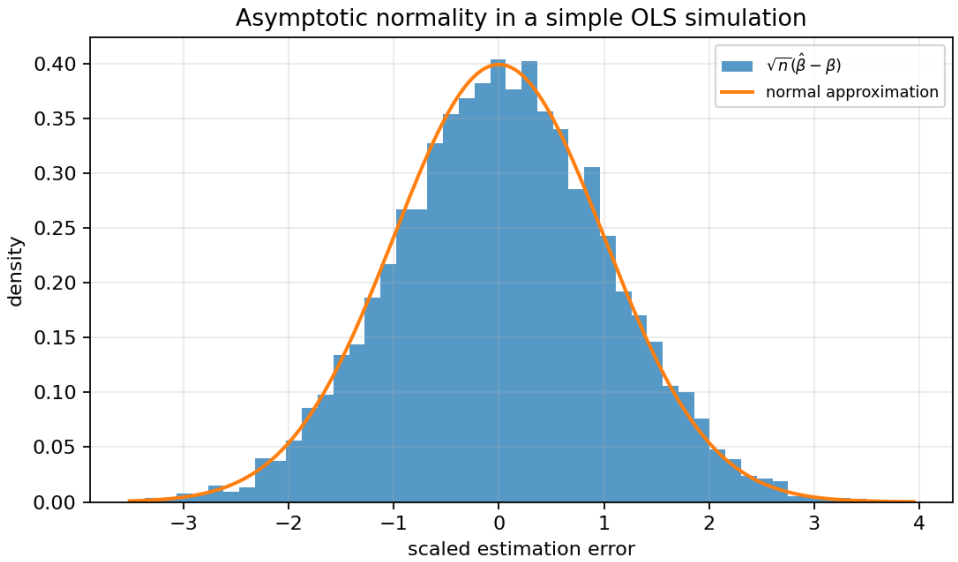



图 6: 简单 OLS 模拟：标准化后的估计误差逐渐呈现正态形状。

**Advanced Core**

**这就是 robust standard error 的来源**。异方差稳健方差估计对应

$$\widehat{V} = \hat{Q}^{-1}\hat{\Omega}\hat{Q}^{-1},$$

其中

$$\hat{Q} = \frac{1}{n}\sum_i X_i X_i', \qquad \hat{\Omega} = \frac{1}{n}\sum_i X_i X_i' \hat{u}_i^2.$$

它不是拍脑袋修正，而是来自 OLS 渐近方差的 sandwich 结构。

## 11 MLE 与 GMM 的大样本逻辑入口

### 11.1 M-estimation 的统一形式

很多估计量可以写成

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} q(W_i, \theta).$$

例如 MLE 中，$q$ 是 log-likelihood：

$$q(W_i, \theta) = \log f(W_i; \theta).$$

大样本逻辑是：由 LLN，

$$\frac{1}{n} \sum_{i} q(W_i, \theta) \xrightarrow{p} E[q(W_i, \theta)]$$

在每个 $\theta$ 附近成立。如果总体目标函数在 $\theta_0$ 处唯一最大，则样本最大点 $\hat{\theta}$ 会靠近总体最大点 $\theta_0$。

### 11.2 GMM 的矩条件

GMM 从矩条件开始：

$$E[g(W_i, \theta_0)] = 0.$$

样本矩为

$$\hat{g}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} g(W_i, \theta).$$

由 LLN，

$$\hat{g}_n(\theta_0) \xrightarrow{p} 0.$$

由 CLT，

$$\sqrt{n}\hat{g}_n(\theta_0) \xrightarrow{d} \mathcal{N}(0, S).$$

GMM 估计量通过让样本矩尽可能接近 0 来估计参数：

$$\hat{\theta} = \arg\min_{\theta} \hat{g}_n(\theta)' W_n \hat{g}_n(\theta).$$

后续计量理论会把 Taylor 展开、Slutsky 和 Delta method 结合起来，推出 GMM 的渐近正态分布。

## 12 置信区间与假设检验

### 12.1 单个参数的置信区间

若

$$\frac{\hat{\theta} - \theta_0}{\widehat{se}(\hat{\theta})} \xrightarrow{d} \mathcal{N}(0, 1),$$

则大样本 95% 置信区间为

$$\boxed{\hat{\theta} \pm 1.96\, \widehat{se}(\hat{\theta}).}$$

这里的 1.96 来自标准正态分布的 97.5% 分位数。

### 12.2 t statistic

检验

$$H_0 : \theta = \theta_0$$

时，使用

$$t = \frac{\hat{\theta} - \theta_0}{\widehat{se}(\hat{\theta})}.$$

在大样本下，若 $H_0$ 为真，

$$t \xrightarrow{d} \mathcal{N}(0,1).$$

因此若 $|t| > 1.96$，在近似 5% 水平上拒绝双侧原假设。

**常见误区**

“渐近正态”不等于“小样本精确正态”。样本很小、分布厚尾、数据强相关或存在聚类时，普通正态近似可能很差。此时需要有限样本方法、bootstrap、cluster robust standard errors 或更合适的模型。

## 13 金融应用：均值、波动率与 Sharpe ratio

### 13.1 估计期望收益

令 $R_t$ 是资产超额收益，$E[R_t] = \mu$。样本均值为

$$\bar{R}_T = \frac{1}{T}\sum_{t=1}^{T} R_t.$$

若收益独立同分布且方差有限，则

$$\bar{R}_T \xrightarrow{p} \mu,$$

$$\sqrt{T}(\bar{R}_T - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

这说明估计期望收益很难：标准误按 $1/\sqrt{T}$ 下降，若收益波动很大，需要很长样本才能精确估计均值。

### 13.2 估计 Sharpe ratio

Sharpe ratio 为

$$SR = \frac{\mu}{\sigma}.$$

样本估计为

$$\widehat{SR} = \frac{\bar{R}}{\hat{\sigma}}.$$

这是样本均值和样本标准差的非线性函数。它的一致性来自 CMT:

$$\bar{R} \xrightarrow{p} \mu, \qquad \hat{\sigma} \xrightarrow{p} \sigma \quad \Longrightarrow \quad \widehat{SR} \xrightarrow{p} \frac{\mu}{\sigma}.$$

它的渐近分布则需要多维 CLT 和 Delta method。

## 14 常见误区

**常见误区**

**误区 1：依概率收敛和依分布收敛是一回事**。不是。依概率收敛表示随机变量本身靠近；依分布收敛表示分布形状靠近。$X_n \xrightarrow{p} c$ 会推出 $X_n \xrightarrow{d} c$，但一般依分布收敛不推出依概率收敛。

**误区 2：CLT 意味着原始数据正态**。不是。CLT 说的是标准化样本均值近似正态，不是说 $X_i$ 本身服从正态。

**误区 3：一致性说明估计量无偏**。不是。无偏是有限样本性质，一致性是大样本性质。

**误区 4：标准误越小一定越好**。标准误小表示抽样不确定性小，但如果模型有内生性或识别错误，估计量可以非常精确地收敛到错误对象。

**误区 5：样本很大就可以忽略相关性**。时间序列和面板数据常有自相关、聚类和异方差。若标准误没有正确处理相关结构，推断会失真。

## 15 本章小结图谱

| 概念 | 核心公式 | 经济计量含义 |
|---|---|---|
| 依概率收敛 | $X_n \xrightarrow{p} X$ | 一致性的语言。 |
| 依分布收敛 | $X_n \xrightarrow{d} X$ | 渐近分布的语言。 |
| LLN | $\bar{X}_n \xrightarrow{p} \mu$ | 样本矩接近总体矩。 |
| CLT | $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$ | 标准误和正态近似。 |
| CMT | $X_n \to X \Rightarrow g(X_n) \to g(X)$ | 估计量函数的一致性。 |
| Slutsky | 随机极限加常数极限 | 用估计标准误替代真实标准误。 |
| Delta method | $\sqrt{n}(g(\hat{\theta}) - g(\theta))$ | 非线性变换的渐近分布。 |
| OLS 一致性 | $\frac{1}{n}\sum X_i u_i \xrightarrow{p} 0$ | 外生性矩条件。 |
| OLS 渐近正态 | sandwich variance | robust standard errors。 |

# 16 练习题

## A 组：基础计算

A1. 设 $X_i$ i.i.d., $E[X_i]=5$, $\mathrm{Var}(X_i)=9$。求 $E[\bar{X}_n]$ 与 $\mathrm{Var}(\bar{X}_n)$。

A2. 在 A1 条件下，用 Chebyshev 不等式给出 $P(|\bar{X}_n-5|>0.5)$ 的上界。

A3. 若 $\sqrt{n}(\hat{\theta}-\theta)\xrightarrow{d}N(0,4)$，求 $\hat{\theta}$ 的近似标准误。

A4. 若 $\sqrt{n}(\hat{\theta}-\theta)\xrightarrow{d}N(0,9)$，且 $g(\theta)=\log\theta$, $\theta=3$，用 Delta method 求 $\sqrt{n}(\log\hat{\theta}-\log\theta)$ 的极限方差。

A5. 若 $\hat{\sigma}^2\xrightarrow{p}\sigma^2$ 且 $\sigma^2>0$，证明 $\hat{\sigma}\xrightarrow{p}\sigma$。

A6. 若 $X_n\xrightarrow{d}N(0,1)$ 且 $Y_n\xrightarrow{p}2$，求 $X_n/Y_n$ 的极限分布。

## B 组：概念判断

B1. LLN 和 CLT 分别回答什么问题?

B2. 为什么 CLT 中要乘以 $\sqrt{n}$?

B3. 一致估计量一定无偏吗？无偏估计量一定一致吗?

B4. 依分布收敛为什么通常不足以说明估计量“靠近真值”?

B5. 为什么 Slutsky 定理对标准误替换很重要?

B6. Delta method 为什么本质上是 Taylor 展开?

## C 组：混合题

C1. 设 $X_i$ i.i.d., $E[X_i]=\mu$, $\mathrm{Var}(X_i)=\sigma^2$。推导 $\bar{X}_n$ 的 CLT 近似置信区间。

C2. 若 $\hat{\theta}\xrightarrow{p}\theta$，且 $\theta\neq0$，证明 $1/\hat{\theta}\xrightarrow{p}1/\theta$。

C3. 设 $\sqrt{n}(\hat{\theta}-\theta)\xrightarrow{d}N(0,V)$。令 $g(\theta)=\theta^2$。求 $g(\hat{\theta})$ 的渐近分布。

C4. 推导样本方差 $\frac{1}{n}\sum_i(X_i-\bar{X})^2$ 的一致性。

C5. 对简单回归 $Y_i=\beta X_i+u_i$，无截距，写出 $\sqrt{n}(\hat{\beta}-\beta)$ 的渐近表达式。

## D 组：经济金融应用题

D1. 用历史月度收益估计资产期望收益时，为什么均值比波动率更难估计？请用 $1/\sqrt{n}$ rate 解释。

D2. 给定 OLS 模型 $Y=X'\beta+u$，解释为什么 $E[Xu]=0$ 是一致性的核心条件。

D3. 假设收益存在自相关，但你仍按 i.i.d. 公式计算标准误。可能会造成什么问题?

D4. Sharpe ratio 是 $\mu/\sigma$。说明为什么估计 Sharpe ratio 的渐近分布需要 Delta method。

D5. 在政策评估中，一个估计量非常精确但不一致意味着什么?

### E 组：挑战 / 证明题

E1. 用 Chebyshev 不等式证明弱大数定律的一个简单版本。

E2. 证明：若 $X_n \xrightarrow{p} c$，其中 $c$ 为常数，则 $X_n \xrightarrow{d} c$。

E3. 推导异方差情形下 OLS 的渐近方差 $Q^{-1}\Omega Q^{-1}$。

E4. 设 $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(0, V)$，$g : \mathbb{R}^k \to \mathbb{R}$ 可微。写出多维 Delta method 并解释每一项的维度。

E5. 解释为什么一致性、渐近正态和方差一致估计三者合在一起，才能形成可用的大样本推断。

# 17 练习题参考答案

## A 组答案

A1.

$$E[\bar{X}_n] = 5, \qquad \mathrm{Var}(\bar{X}_n) = \frac{9}{n}.$$

A2. Chebyshev 不等式给出

$$P(|\bar{X}_n - 5| > 0.5) \le \frac{9/n}{0.5^2} = \frac{36}{n}.$$

A3. 渐近方差为 $4/n$，所以近似标准误为

$$\frac{2}{\sqrt{n}}.$$

A4. $g'(\theta) = 1/\theta = 1/3$。极限方差为

$$(1/3)^2 \cdot 9 = 1.$$

A5. 因为平方根函数在 $\sigma^2 > 0$ 附近连续，由连续映射定理，

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} \xrightarrow{p} \sqrt{\sigma^2} = \sigma.$$

A6. 由 Slutsky，

$$\frac{X_n}{Y_n} \xrightarrow{d} \frac{Z}{2}, \qquad Z \sim N(0,1).$$

所以极限分布是

$$N(0, 1/4).$$

## B 组答案

B1. LLN 回答“样本平均或样本矩会靠近总体对象吗”；CLT 回答“靠近时的误差分布是什么”。

B2. 因为样本均值误差的标准差是 $\sigma/\sqrt{n}$。乘以 $\sqrt{n}$ 后，误差具有非退化的极限分布。

B3. 不一定。一致性是大样本性质，无偏是有限样本性质。无偏也不一定一致，如果方差不随 $n$ 消失，仍可能不一致。

B4. 依分布收敛只说明分布形状靠近，不保证随机变量本身与真值距离变小。一致性需要依概率收敛到常数。

B5. 因为真实标准误通常未知。Slutsky 允许用一致估计的标准误替代真实标准误，而不改变 t-statistic 的极限分布。

B6. 因为 Delta method 用

$$g(\hat{\theta}) \approx g(\theta) + g'(\theta)(\hat{\theta} - \theta)$$

把非线性变换局部线性化。

## C 组答案

C1. CLT 给出

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1).$$

若用 $\hat{\sigma}$ 估计 $\sigma$，则近似 95% 置信区间为

$$\bar{X}_n \pm 1.96\frac{\hat{\sigma}}{\sqrt{n}}.$$

C2. 因为 $g(x) = 1/x$ 在 $\theta \neq 0$ 附近连续，由连续映射定理，

$$1/\hat{\theta} \xrightarrow{p} 1/\theta.$$

C3. $g'(\theta) = 2\theta$。所以

$$\sqrt{n}(\hat{\theta}^2 - \theta^2) \xrightarrow{d} N(0, 4\theta^2 V).$$

C4.

$$\frac{1}{n}\sum_i (X_i - \bar{X})^2 = \frac{1}{n}\sum_i X_i^2 - \bar{X}^2.$$

LLN 给出第一项收敛到 $E[X^2]$,第二项由 CMT 收敛到 $\mu^2$。所以样本方差收敛到 $E[X^2]-\mu^2$。

C5. 无截距估计量为

$$\hat{\beta} = \frac{\sum_i X_i Y_i}{\sum_i X_i^2}.$$

因而

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n}\sum_i X_i^2\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_i X_i u_i\right).$$

## D 组答案

D1. 期望收益的标准误约为 $\sigma/\sqrt{n}$。金融收益的 $\sigma$ 很大，而 $\mu$ 通常较小，所以即使 $n$ 不小，均值估计仍可能很不精确。波动率是二阶矩，通常相对更容易在有限样本中估得稳定一些。

D2. OLS 误差可写为

$$\hat{\beta} - \beta = \hat{Q}^{-1}\frac{1}{n}\sum_i X_i u_i.$$

如果 $E[Xu] = 0$，LLN 使样本矩趋于 0，从而估计误差趋于 0。

D3. 自相关会改变样本均值或回归估计量的长期方差。若仍用 i.i.d. 标准误，可能低估或高估真实不确定性，导致显著性检验失真。

D4. 因为 $\widehat{SR} = \bar{R}/\hat{\sigma}$ 是 $(\bar{R}, \hat{\sigma})$ 的非线性函数。先需要二者的联合渐近分布，再用 Delta method 得到比率的渐近分布。

D5. 它意味着抽样误差小，但估计目标错了。比如内生性导致估计量稳定地收敛到偏误参数，此时标准误再小也不能解释为真实因果效应。

### E 组答案

E1. 若 $X_i$ i.i.d. 且方差有限，则

$$E[\bar{X}_n] = \mu, \qquad \mathrm{Var}(\bar{X}_n) = \sigma^2/n.$$

Chebyshev 给出

$$P(|\bar{X}_n - \mu| > \varepsilon) \le \frac{\sigma^2}{n\varepsilon^2} \to 0.$$

所以 $\bar{X}_n \xrightarrow{p} \mu$。

E2. 对任意连续有界函数或通过分布函数定义均可证明。直觉上，$X_n$ 以概率趋近常数 $c$，所以其分布质量集中到 $c$。因此分布函数在 $c$ 以外连续点收敛到退化分布的分布函数。

E3. 从

$$\sqrt{n}(\hat{\beta} - \beta) = \hat{Q}^{-1} \frac{1}{\sqrt{n}} \sum_i X_i u_i$$

出发。若 $\hat{Q} \xrightarrow{p} Q$，且

$$\frac{1}{\sqrt{n}} \sum_i X_i u_i \xrightarrow{d} N(0, \Omega),$$

则由 Slutsky,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, Q^{-1}\Omega Q^{-1}).$$

E4. 若 $g : \mathbb{R}^k \to \mathbb{R}$，则 gradient 为 $k \times 1$ 向量 $G = \nabla g(\theta)$。多维 Delta method 给出

$$\sqrt{n}[g(\hat{\theta}) - g(\theta)] \xrightarrow{d} N(0, G'VG).$$

其中 $V$ 是 $k \times k$ 渐近协方差矩阵，$G'VG$ 是标量。

E5. 一致性保证估计对象正确；渐近正态给出误差分布形状；方差一致估计允许我们把未知方差替换成可计算标准误。三者合起来才能形成置信区间、t 检验和 p-value。

## 18 下一章连接

本章建立了大样本推断的基础。下一章将进入多元正态、二次型与金融风险。那一章会把本章的渐近正态进一步矩阵化：

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V)$$

中的 $V$ 是协方差矩阵；投资组合风险

$$\sigma_p^2 = w'\Sigma w$$

也是二次型。理解多元正态与二次型，会让你同时更懂回归检验、F 检验、Wald 检验、mean-variance portfolio 和 PCA。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 29
# 多元正态、二次型与金融风险

Chapter 28 讲了收敛概念和极限定理。本章把这些极限结果转化为经济金融中最常用的多元分布语言：向量均值、协方差矩阵、多元正态、二次型、卡方/t/F 分布、投资组合方差、风险分解和 PCA。你可以把本章看成计量推断和金融风险之间的桥：计量里很多检验统计量是二次型，金融里很多风险度量也是二次型。

math4econ personalized notes
2026

# 目录

# Chapter 29 总览

| | |
|---|---|
| **本章主题** | 多元随机向量、协方差矩阵、多元正态分布、二次型、卡方/t/F 分布、投资组合风险与 PCA。 |
| **核心问题** | 多个随机变量一起变化时，如何描述它们的联合波动？为什么投资组合方差是 $\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}$？为什么回归检验统计量经常是卡方、t 或 F？协方差矩阵的特征值和特征向量如何解释风险来源？ |
| **数学工具** | covariance matrix, positive semidefinite matrix, multivariate normal, quadratic form, eigenvalue decomposition, chi-square distribution, t distribution, F distribution, PCA。 |
| **经济金融应用** | 投资组合方差、diversification、mean-variance frontier、factor model、PCA 风险因子、OLS 检验统计量、Wald test、VaR 正态近似。 |
| **后续连接** | Chapter 30 Monte Carlo 与计算统计、金融风险建模、计量推断、时间序列和因子模型。 |

**直觉盒子**

本章最重要的主线是：

$$\boxed{\text{随机向量} \to \text{协方差矩阵} \to \text{二次型} \to \text{风险/检验统计量}}$$

单变量里，风险是方差 $\sigma^2$。多变量里，风险不再只是每个变量自己的方差，还包括变量之间如何一起动，也就是协方差。把这些信息集中起来，就是协方差矩阵 $\boldsymbol{\Sigma}$。

## 1 为什么这一章重要

Chapter 26 讲了随机变量和分布，Chapter 27 讲了条件期望与回归，Chapter 28 讲了极限定理。现在我们需要处理最现实的一步：经济金融里的随机对象很少是单变量。

- 一个投资组合有多只资产收益 $R_1, \dots, R_n$；
- 一个宏观系统有产出、通胀、利率、汇率等多个变量；
- 一个回归模型有多个解释变量和多个系数；
- 一个检验可能同时检验多个限制；
- 一个因子模型用少数共同因子解释很多资产收益。

这些问题都要求我们用随机向量和矩阵语言，而不仅是一元分布。

### 1.1 两个领域共用同一个数学对象

金融中，若资产收益向量为

$$\boldsymbol{R} = (R_1, \dots, R_n)',$$

权重为

$$\boldsymbol{w} = (w_1, \dots, w_n)',$$

则投资组合收益为

$$R_p = \boldsymbol{w}'\boldsymbol{R}.$$

投资组合方差是

$$\mathrm{Var}(R_p) = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}.$$

计量中，若估计量满足

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \mathcal{N}(\mathbf{0}, V),$$

则很多检验统计量也长成二次型：

$$n(R\hat{\boldsymbol{\beta}} - r)'(R\hat{V}R')^{-1}(R\hat{\boldsymbol{\beta}} - r).$$

**直觉盒子**

同一个数学结构在金融里叫“风险”，在计量里叫“估计误差的不确定性”。它们都由协方差矩阵和二次型控制。

## 2 随机向量与均值向量

### 2.1 随机向量

**定义 / 定理**

若 $X_1, \dots, X_k$ 是随机变量，则

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$$

称为 $k$ 维随机向量 random vector。

随机向量的均值是向量：

$$\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X}] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_k] \end{pmatrix}.$$

如果 $\in \mathbb{R}^k$ 是常数向量，则线性组合

$$Y =' \boldsymbol{X}$$

的期望为

$$\mathbb{E}[Y] = \mathbb{E}[' \boldsymbol{X}] =' \mathbb{E}[\boldsymbol{X}] =' \boldsymbol{\mu}.$$

## 2.2 协方差矩阵

**定义 / 定理**

随机向量 $\boldsymbol{X}$ 的协方差矩阵定义为

$$\boldsymbol{\Sigma} = \mathrm{Var}(\boldsymbol{X}) = \mathbb{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'].$$

其第 $(i, j)$ 个元素为

$$\Sigma_{ij} = \mathrm{Cov}(X_i, X_j).$$

因此

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_k) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_k, X_1) & \mathrm{Cov}(X_k, X_2) & \cdots & \mathrm{Var}(X_k) \end{pmatrix}.$$

**方法模板**

协方差矩阵的三条基本性质：

1. 对称：$\boldsymbol{\Sigma}' = \boldsymbol{\Sigma}$；
2. 半正定：对任意 ，都有 $'\boldsymbol{\Sigma} \geq 0$；
3. 对角线是方差，非对角线是协方差。

## 2.3 为什么协方差矩阵半正定

对任意常数向量 ，考虑随机变量

$$Y =' \boldsymbol{X}.$$

它的方差不能为负。又

$$\mathrm{Var}(Y) = \mathrm{Var}('\boldsymbol{X}) =' \mathrm{Var}(\boldsymbol{X}) =' \boldsymbol{\Sigma}.$$

所以

$$'\boldsymbol{\Sigma} \geq 0.$$

这正是半正定 positive semidefinite 的定义。

**常见误区**

协方差矩阵不是任意对称矩阵。它必须半正定。若你估计出的协方差矩阵有负特征值，通常说明数据、估计或数值处理有问题，或者你对矩阵做了不恰当的修正。

## 3 线性变换下的均值与协方差

这是本章最常用的一组公式。

设

$$= A\boldsymbol{X}+,$$

其中 $A$ 是常数矩阵， 是常数向量。若

$$\mathbb{E}[\boldsymbol{X}] = \boldsymbol{\mu}, \qquad \mathrm{Var}(\boldsymbol{X}) = \boldsymbol{\Sigma},$$

则

$$\boxed{\mathbb{E}[] = A\boldsymbol{\mu}+,}$$

$$\boxed{\mathrm{Var}() = A\boldsymbol{\Sigma}A'.}$$

### 3.1 推导

均值：

$$\mathbb{E}[] = \mathbb{E}[A\boldsymbol{X}+] = A\mathbb{E}[\boldsymbol{X}]+ = A\boldsymbol{\mu} + .$$

协方差：

$$-\mathbb{E}[] = A(\boldsymbol{X} - \boldsymbol{\mu}).$$

所以

$$\begin{aligned}\mathrm{Var}() &= \mathbb{E}[(-\mathbb{E})(-\mathbb{E})'] \\ &= \mathbb{E}[A(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'A'] \\ &= A\mathbb{E}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})']A' \\ &= A\boldsymbol{\Sigma}A'.\end{aligned}$$

**Worked Example**

**例 1：投资组合均值与方差**

设资产收益向量 $\boldsymbol{R}$ 的均值和协方差为

$$\mathbb{E}[\boldsymbol{R}] = \boldsymbol{\mu}, \qquad \mathrm{Var}(\boldsymbol{R}) = \boldsymbol{\Sigma}.$$

投资组合收益为

$$R_p = \boldsymbol{w}'\boldsymbol{R}.$$

这里 $A = \boldsymbol{w}'$ 是 $1 \times n$ 矩阵。因此

$$\mathbb{E}[R_p] = \boldsymbol{w}'\boldsymbol{\mu},$$

$$\mathrm{Var}(R_p) = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}.$$

这就是现代投资组合理论最核心的二次型。

# 4 二次型与风险

## 4.1 二次型是什么

**定义 / 定理**

给定对称矩阵 $A$ 和向量 $\boldsymbol{x}$，表达式

$$Q(\boldsymbol{x}) = \boldsymbol{x}'A\boldsymbol{x}$$

称为二次型 quadratic form。

展开为

$$\boldsymbol{x}'A\boldsymbol{x} = \sum_i a_{ii}x_i^2 + 2\sum_{i<j} a_{ij}x_i x_j.$$

这里对角线项表示“自身平方贡献”，非对角线项表示“交互贡献”。

## 4.2 投资组合方差展开

对于两只资产，权重为 $w$ 和 $1-w$，标准差为 $\sigma_1, \sigma_2$，相关系数为 $\rho$。投资组合方差为

$$\boxed{\sigma_p^2 = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\rho\sigma_1\sigma_2.}$$

最后一项是分散化 diversification 的关键。如果 $\rho < 1$，组合风险可能低于简单加权平均风险。若 $\rho$ 越低，分散化收益越大。

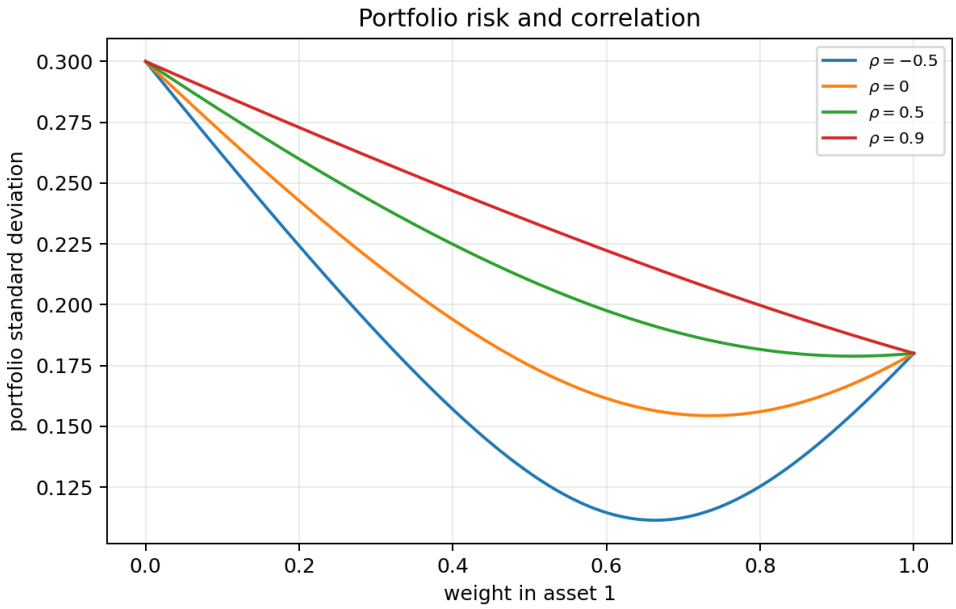



图 1：相关系数越低，投资组合标准差曲线越向下弯，分散化收益越明显。

**常见误区**

很多人以为分散化来自“资产数量多”。更准确地说，分散化来自“不完全正相关”。如果所有资产完全正相关，增加资产数量并不能真正降低系统性风险。

### 4.3 二次型的几何解释

如果 $A$ 是正定矩阵，则集合

$$\{\boldsymbol{x} : \boldsymbol{x}' A \boldsymbol{x} = c\}$$

通常是椭圆或椭球。矩阵的特征向量决定椭圆主轴方向，特征值决定主轴长度。

在金融中，协方差矩阵把收益空间变成风险椭圆：沿某些方向移动风险增长很快，沿另一些方向风险增长较慢。

## 5 特征值分解与 covariance ellipse

### 5.1 谱分解

若 $\boldsymbol{\Sigma}$ 是对称矩阵，则可以写成

$$\boxed{\boldsymbol{\Sigma} = Q \Lambda Q',}$$

其中

$$Q'Q = I,$$

且

$$\Lambda = \mathbf{diag}(\lambda_1, \ldots, \lambda_k).$$

这里 $Q$ 的列是特征向量，$\lambda_i$ 是特征值。

由于协方差矩阵半正定，所有特征值满足

$$\lambda_i \geq 0.$$

## 5.2 把风险旋转到主轴方向

令

$$= Q'(X - mu).$$

则

$$\mathrm{Var}() = Q'\mathbf{\Sigma}Q = \Lambda.$$

这说明通过正交旋转，可以把原来的相关随机变量变成不相关的主成分方向。每个方向的方差就是对应特征值。

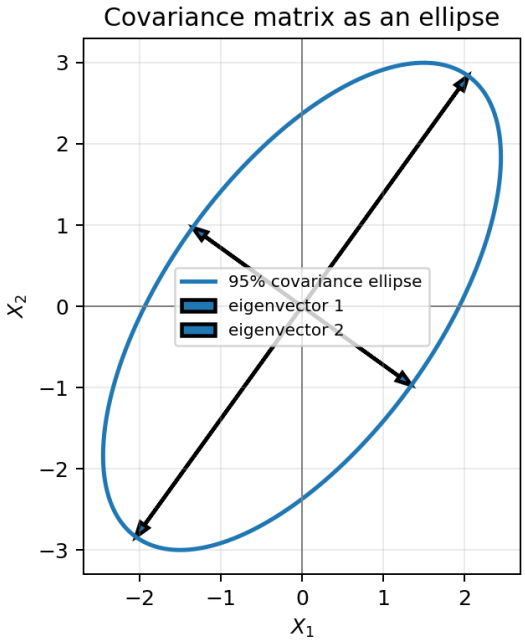



图 2: 协方差矩阵可以被理解为椭圆：特征向量给方向，特征值给该方向的方差大小。

**直觉盒子**

如果第一大特征值很大，说明数据中存在一个主要共同方向。金融中，这常被解释为市场因子或共同风险来源；宏观中，可能是共同周期；多资产收益中，可能是整体市场一起涨跌的方向。

# 6 多元正态分布

## 6.1 定义

**定义 / 定理**

若随机向量 $\boldsymbol{X} \in \mathbb{R}^k$ 的密度为

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right],$$

则称

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

其中

$$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

叫 Mahalanobis distance 的平方。它不是普通欧氏距离，而是按协方差结构调整后的距离。

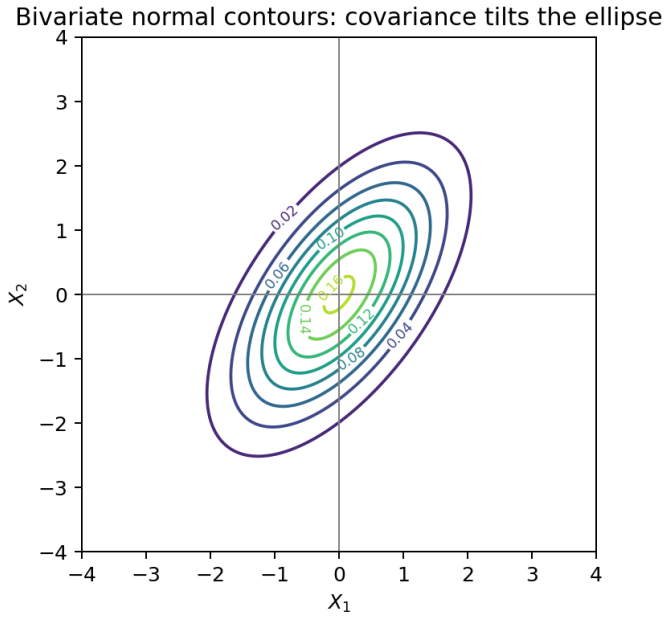



图 3: 二维正态的等高线是椭圆。协方差改变椭圆的倾斜方向和形状。

## 6.2 多元正态的核心性质

若

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

则有以下重要性质：

1. 任何线性组合仍然正态：

$$'\boldsymbol{X} \sim \mathcal{N}('\boldsymbol{\mu},' \boldsymbol{\Sigma}).$$

2. 任何线性变换仍然多元正态:

$$A\boldsymbol{X}+ \sim \mathcal{N}(A\boldsymbol{\mu}+, A\boldsymbol{\Sigma}A').$$

3. 在多元正态下，不相关等价于独立。一般分布中不相关不代表独立，但多元正态是特殊情况。

4. 条件分布仍然正态。这是高斯线性模型、Kalman filter 和 Bayesian normal model 的基础。

### 6.3 标准化

若 $\boldsymbol{\Sigma}$ 正定，可取矩阵 $\boldsymbol{\Sigma}^{-1/2}$，令

$$= \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}).$$

则

$$\sim \mathcal{N}(\mathbf{0}, I).$$

反过来，若

$$\sim \mathcal{N}(\mathbf{0}, I),$$

且 $AA' = \boldsymbol{\Sigma}$，则

$$\boldsymbol{X} = \boldsymbol{\mu} + A \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

这也是模拟多元正态的基本方法。

## 7 卡方分布与二次型

### 7.1 从标准正态到卡方

**定义 / 定理**

若 $Z_1, \dots, Z_m$ 独立且均服从标准正态，则

$$Q = Z_1^2 + \cdots + Z_m^2$$

服从自由度为 $m$ 的卡方分布，记作

$$Q \sim \chi^2_m.$$

用向量写，若

$$\sim \mathcal{N}(\mathbf{0}, I_m),$$

则

$$' \sim \chi^2_m.$$

### 7.2 一般正态二次型

若

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

且 $\boldsymbol{\Sigma}$ 正定，则

$$(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \sim \chi^2_k.$$

证明思路：令

$$= \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu}),$$

则 $\sim \mathcal{N}(\mathbf{0}, I)$，并且

$$' = (\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}).$$

**Advanced Core**

**Wald test 的根源**

如果

$$\sqrt{n}(-_0) \mathcal{N}(\mathbf{0}, V),$$

那么

$$n(-_0)'V^{-1}(-_0) \chi^2_k.$$

这就是多参数检验中 Wald statistic 的数学来源。换句话说，Wald 检验就是看估计量离原假设有多远，但这个“距离”不是普通欧氏距离，而是用协方差矩阵调整后的 Mahalanobis distance。

## 8 t 分布和 F 分布

### 8.1 t 分布

**定义 / 定理**

若

$$Z \sim \mathcal{N}(0,1), \qquad U \sim \chi^2_\nu,$$

且 $Z$ 与 $U$ 独立，则

$$T = \frac{Z}{\sqrt{U/\nu}}$$

服从自由度为 $\nu$ 的 t 分布，记作

$$T \sim t_\nu.$$

t 分布出现在样本方差替代真实方差的时候。直觉上，如果方差未知，我们用估计的标准误来标准化均值或系数，额外的不确定性会让尾部更厚。

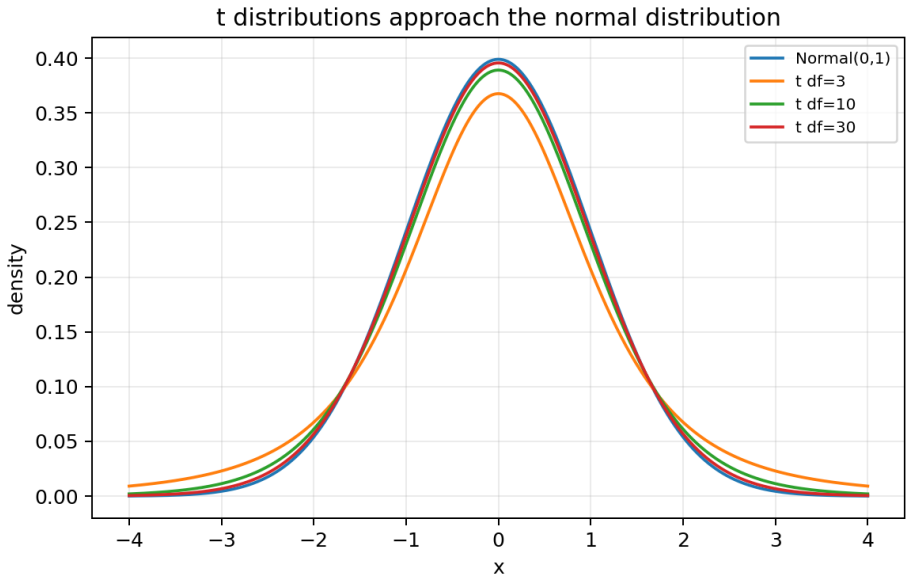



图 4：自由度越大，t 分布越接近标准正态。小样本下 t 分布尾部更厚。

## 8.2 F 分布

**定义 / 定理**

若

$$U_1 \sim \chi^2_{d_1}, \qquad U_2 \sim \chi^2_{d_2},$$

且相互独立，则

$$F = \frac{U_1/d_1}{U_2/d_2}$$

服从自由度为 $(d_1, d_2)$ 的 F 分布。

F 分布在回归中常用于联合显著性检验。例如同时检验多个系数是否为零：

$$H_0 : \beta_2 = \beta_3 = \cdots = \beta_k = 0.$$

# 9 回归中的二次型：OLS 推断入口

考虑线性模型

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I).$$

OLS 估计量为

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}.$$

### 9.1 OLS 估计量分布

代入模型：

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (X'X)^{-1}X'\boldsymbol{\varepsilon}.$$

由于 $\varepsilon$ 正态，线性变换仍然正态：

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2(X'X)^{-1}\right).$$

单个系数的 t 统计量为

$$t_j = \frac{\hat{\beta}_j - \beta_{j,0}}{\widehat{se}(\hat{\beta}_j)}.$$

在经典正态线性模型下，它服从 t 分布；在大样本下，它近似标准正态。

### 9.2 联合检验与 F/Wald

若要检验

$$H_0 : R\boldsymbol{\beta} = r,$$

其中 $R$ 是 $q \times k$ 矩阵，则 Wald 统计量为

$$W = (R\hat{\boldsymbol{\beta}} - r)' \left[R\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})R'\right]^{-1} (R\hat{\boldsymbol{\beta}} - r).$$

在大样本下，

$$W\chi^2_q.$$

在经典有限样本正态线性模型中，对应的 F 检验为

$$F = \frac{(SSR_R - SSR_U)/q}{SSR_U/(n-k)}.$$

**直觉盒子**

回归检验不是“拍脑袋看系数大不大”。它本质上是看估计结果与原假设之间的距离，而这个距离必须根据估计误差的协方差矩阵进行标准化。

## 10 投资组合风险：均值-方差框架

### 10.1 基本问题

设资产收益均值为 $\boldsymbol{\mu}$，协方差矩阵为 $\boldsymbol{\Sigma}$。投资组合权重为 $\boldsymbol{w}$，满足

$$\mathbf{1}'\boldsymbol{w} = 1.$$

则

$$\mathbb{E}[R_p] = \boldsymbol{w}'\boldsymbol{\mu},$$

$$\sigma_p^2 = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}.$$

均值-方差理论把投资组合选择写成二次规划。最小方差组合为

$$\min_{\boldsymbol{w}} \ \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \quad \text{s.t.} \quad \mathbf{1}'\boldsymbol{w} = 1.$$

## 10.2 全局最小方差组合

构造 Lagrangian：

$$\mathcal{L} = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} - \lambda(\mathbf{1}'\boldsymbol{w} - 1).$$

一阶条件：

$$2\boldsymbol{\Sigma}\boldsymbol{w} - \lambda\mathbf{1} = \mathbf{0}.$$

所以

$$\boldsymbol{w} = \frac{\lambda}{2}\boldsymbol{\Sigma}^{-1}\mathbf{1}.$$

代入约束：

$$\mathbf{1}'\boldsymbol{w} = \frac{\lambda}{2}\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1} = 1.$$

因此

$$\frac{\lambda}{2} = \frac{1}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}.$$

得到

$$\boxed{\boldsymbol{w}_{GMV} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}.}$$

**Worked Example**

**例 2：两资产组合方差最小化**

两资产组合方差为

$$\sigma_p^2 = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\rho\sigma_1\sigma_2.$$

对 $w$ 求导：

$$\frac{\mathrm{d}\sigma_p^2}{\mathrm{d}w} = 2w\sigma_1^2 - 2(1-w)\sigma_2^2 + 2(1-2w)\rho\sigma_1\sigma_2.$$

令其为 0，整理得

$$\boxed{w^* = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.}$$

如果相关系数越低，最小方差组合的风险下降越明显。

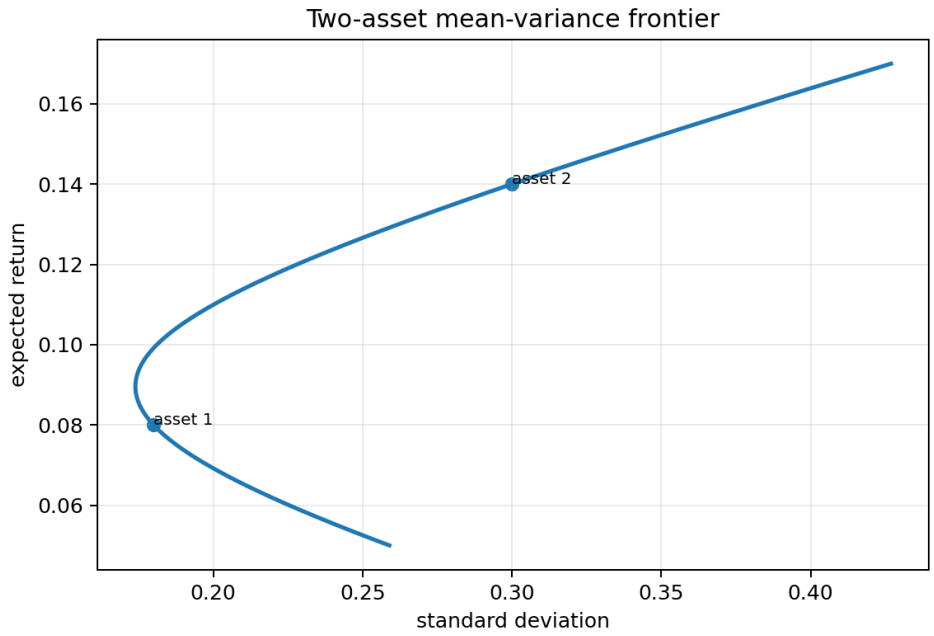



图 5: 两资产均值-方差 frontier。允许权重变化后，组合在风险-收益平面上形成曲线。

# 11 PCA 与因子风险

## 11.1 PCA 的核心思想

Principal Component Analysis, PCA, 的目标是找到一组正交方向，使得数据在第一方向上的方差最大，第二方向在与第一方向正交的条件下方差最大，依次类推。

数学上，如果协方差矩阵为 $\mathbf{\Sigma}$，PCA 就是特征值分解：

$$\mathbf{\Sigma} = Q\Lambda Q'.$$

第一主成分方向是最大特征值对应的特征向量。

## 11.2 方差解释比例

若特征值为

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0,$$

则总方差为

$$\mathrm{tr}(\mathbf{\Sigma}) = \lambda_1 + \cdots + \lambda_k.$$

第 $j$ 个主成分解释的方差比例为

$$\frac{\lambda_j}{\sum_{i=1}^{k} \lambda_i}.$$

前 $m$ 个主成分解释比例为

$$\frac{\lambda_1 + \cdots + \lambda_m}{\lambda_1 + \cdots + \lambda_k}.$$

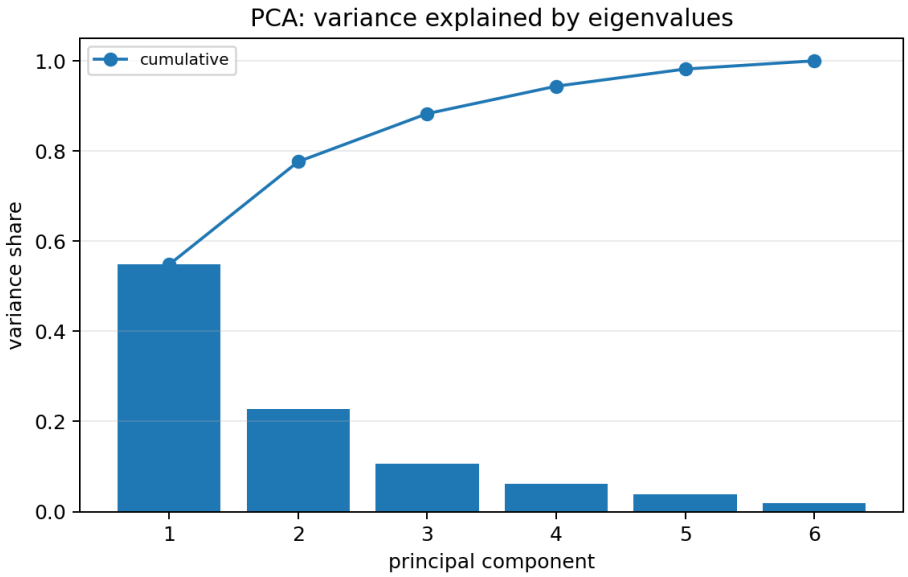



图 6: PCA 中，每个特征值代表一个主成分方向上的方差贡献。

### 11.3 金融解释

在资产收益中，如果第一主成分解释了很高比例的方差，通常说明很多资产共同受一个市场因子影响。第二、第三主成分可能对应行业、期限结构、信用风险等更细分的共同风险。

**常见误区**

PCA 因子是统计因子，不一定天然有经济含义。解释 PCA 结果时，需要结合资产类别、样本期、市场结构和经济理论。不能只看到第一主成分就机械说它一定是“市场因子”。

## 12 正态近似下的 VaR 与风险测度

### 12.1 单资产 VaR

若收益 $R$ 近似服从

$$R \sim \mathcal{N}(\mu, \sigma^2),$$

则损失可定义为 $L = -R$。若要计算置信水平 $\alpha$ 下的 VaR，令 $z_\alpha$ 是标准正态的 $\alpha$ 分位数，则

$$\mathrm{VaR}_\alpha(L) = -\mu + z_\alpha \sigma.$$

### 12.2 投资组合 VaR

若

$$\boldsymbol{R} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

组合收益为

$$R_p = \boldsymbol{w}' \boldsymbol{R}.$$

则

$$R_p \sim \mathcal{N}(\boldsymbol{w}'\boldsymbol{\mu}, \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}).$$

因此组合损失 $L_p = -R_p$ 的正态 VaR 为

$$\boxed{\mathrm{VaR}_\alpha(L_p) = -\boldsymbol{w}'\boldsymbol{\mu} + z_\alpha\sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}}.}$$

**常见误区**

正态 VaR 很方便，但金融收益常有厚尾、偏度、波动聚集和极端事件。正态近似可作为入口，但不能盲目当成真实风险分布。实际风险管理中常结合历史模拟、极值理论、GARCH、压力测试和情景分析。

## 13 常见分布之间的关系图谱

| 分布 | 来源 | 典型用途 |
|---|---|---|
| Normal | CLT 或正态假设 | 样本均值、OLS 渐近分布、资产收益近似。 |
| Multivariate Normal | 多个正态变量的联合分布 | 多资产收益、回归系数向量、状态空间模型。 |
| Chi-square | 标准正态平方和 | 方差推断、Wald 检验、似然比检验。 |
| t | 正态除以估计标准差 | 小样本均值检验、回归单系数检验。 |
| F | 两个独立卡方按自由度标准化后的比值 | 回归联合显著性检验、方差分析。 |

## 14 Worked Examples

**Worked Example**

**例 3：线性组合的分布**

设

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 9 \end{pmatrix} \right).$$

求

$$Y = 2X_1 - 3X_2$$

的分布。

令

$$= \begin{pmatrix} 2 \\ -3 \end{pmatrix}.$$

则

$$Y =' \boldsymbol{X}.$$

均值:

$$\mathbb{E}[Y] =' \boldsymbol{\mu} = 2 \cdot 1 - 3 \cdot 2 = -4.$$

方差:

$$\mathrm{Var}(Y) =' \boldsymbol{\Sigma}.$$

先算

$$\boldsymbol{\Sigma} = \begin{pmatrix} 4 & 1 \\ 1 & 9 \end{pmatrix} \begin{pmatrix} 2 \\ -3 \end{pmatrix} = \begin{pmatrix} 5 \\ -25 \end{pmatrix}.$$

再算

$$'\boldsymbol{\Sigma} = (2, -3) \begin{pmatrix} 5 \\ -25 \end{pmatrix} = 10 + 75 = 85.$$

所以

$$\boxed{Y \sim \mathcal{N}(-4, 85).}$$

**Worked Example**

**例 4: 二次型与卡方**

设

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

且 $\boldsymbol{\Sigma}$ 正定。证明

$$Q = (\boldsymbol{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2_k.$$

取 $\boldsymbol{\Sigma}^{-1/2}$，令

$$= \boldsymbol{\Sigma}^{-1/2} (\boldsymbol{X} - \boldsymbol{\mu}).$$

则

$$\sim \mathcal{N}(\mathbf{0}, I_k).$$

所以

$$' = \sum_{i=1}^{k} Z_i^2 \sim \chi^2_k.$$

又

$$' = (\boldsymbol{X}-\boldsymbol{\mu})'(\boldsymbol{\Sigma}^{-1/2})'\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu}).$$

若取对称平方根，则

$$(\boldsymbol{\Sigma}^{-1/2})'\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}.$$

因此

$$Q \sim \chi^2_k.$$

## Worked Example

**例 5：三资产组合方差**

设三资产协方差矩阵为

$$\boldsymbol{\Sigma} = \begin{pmatrix} 0.04 & 0.01 & 0.00 \\ 0.01 & 0.09 & 0.02 \\ 0.00 & 0.02 & 0.16 \end{pmatrix},$$

权重为

$$\boldsymbol{w} = \begin{pmatrix} 0.5 \\ 0.3 \\ 0.2 \end{pmatrix}.$$

求组合方差。

先算

$$\boldsymbol{\Sigma}\boldsymbol{w} = \begin{pmatrix} 0.04(0.5)+0.01(0.3)+0(0.2) \\ 0.01(0.5)+0.09(0.3)+0.02(0.2) \\ 0(0.5)+0.02(0.3)+0.16(0.2) \end{pmatrix} = \begin{pmatrix} 0.023 \\ 0.036 \\ 0.038 \end{pmatrix}.$$

因此

$$\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} = 0.5(0.023)+0.3(0.036)+0.2(0.038) = 0.0299.$$

组合标准差为

$$\sqrt{0.0299} \approx 0.1729.$$

## Worked Example

**例 6：Wald 检验的计算结构**

假设估计量满足

$$\hat{} = \begin{pmatrix} 1.2 \\ 0.7 \end{pmatrix}, \qquad \widehat{\mathrm{Var}}() = \begin{pmatrix} 0.04 & 0.01 \\ 0.01 & 0.09 \end{pmatrix}.$$

检验

$$H_0 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Wald 统计量为

$$W = (\hat{} - _0)' \widehat{\mathrm{Var}}(\hat{})^{-1} (\hat{} - _0).$$

偏离向量为

$$\hat{} - _0 = \begin{pmatrix} 0.2 \\ -0.3 \end{pmatrix}.$$

矩阵逆为

$$\begin{pmatrix} 0.04 & 0.01 \\ 0.01 & 0.09 \end{pmatrix}^{-1} = \frac{1}{0.0035} \begin{pmatrix} 0.09 & -0.01 \\ -0.01 & 0.04 \end{pmatrix}.$$

所以

$$W \approx 2.857.$$

在大样本下，$W$ 近似服从 $\chi^2_2$。若显著性水平为 5%，临界值约为 5.99，则不拒绝原假设。

## 15 常见误区与纠正

**常见误区**

**误区 1：协方差为零就一定独立。**一般情况下不对。只有在多元正态等特殊分布下，不相关才等价于独立。

**误区 2：只看资产各自波动率，不看相关性。**投资组合风险由 $\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}$ 决定。非对角线协方差项可能极大改变组合风险。

**误区 3：把正态 VaR 当成真实风险。**正态近似忽略厚尾和极端事件。金融风险管理中必须对尾部风险保持警惕。

**误区 4：PCA 因子一定有经济含义。** PCA 是统计分解。主成分是否对应经济因子，需要额外解释和验证。

**误区 5：Wald/t/F 检验只是公式。**这些检验本质上都是用协方差矩阵标准化估计误差，再判断距离是否过大。

## 16 本章总结图谱

| 概念 | 核心公式 | 经济金融含义 |
| --- | --- | --- |

| | | |
|---|---|---|
| 随机向量 | $\boldsymbol{X} = (X_1, \dots, X_k)'$ | 多个经济变量或资产收益一起变化。 |
| 均值向量 | $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X}]$ | 多变量平均水平。 |
| 协方差矩阵 | $\boldsymbol{\Sigma} = \mathbb{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})']$ | 联合波动和相关结构。 |
| 线性变换 | $\mathrm{Var}(A\boldsymbol{X}) = A\boldsymbol{\Sigma}A'$ | 投资组合、因子模型、回归估计量。 |
| 二次型 | $\boldsymbol{x}'A\boldsymbol{x}$ | 风险、距离、检验统计量。 |
| 多元正态 | $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ | 计量推断和金融建模的标准入口。 |
| 卡方 | $' \sim \chi^2_k$ | 方差推断、Wald 检验。 |
| t 分布 | $Z/\sqrt{U/\nu}$ | 小样本标准化统计量。 |
| F 分布 | $(U_1/d_1)/(U_2/d_2)$ | 联合限制检验。 |
| PCA | $\boldsymbol{\Sigma} = Q\Lambda Q'$ | 风险分解、统计因子。 |

## 17 练习题

### A 组：基础计算

A1. 设 $\boldsymbol{X} = (X_1, X_2)'$，$\mathbb{E}[X_1] = 1$，$\mathbb{E}[X_2] = 2$，$\mathrm{Var}(X_1) = 4$，$\mathrm{Var}(X_2) = 9$，$\mathrm{Cov}(X_1, X_2) = 3$。写出 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$。

A2. 在 A1 中，求 $Y = 2X_1 - X_2$ 的期望和方差。

A3. 证明 $\mathrm{Var}(A\boldsymbol{X}+) = A\,\mathrm{Var}(\boldsymbol{X})A'$。

A4. 设 $\sim \mathcal{N}(\boldsymbol{0}, I_3)$，求 $'$ 的分布。

A5. 若 $X \sim \mathcal{N}(\mu, \sigma^2)$，证明 $((X-\mu)/\sigma)^2 \sim \chi^2_1$。

A6. 设两资产波动率为 $0.2, 0.3$，相关系数为 $0.5$，权重分别为 $0.6, 0.4$。求组合方差。

### B 组：概念判断

B1. 为什么协方差矩阵必须半正定?

B2. 为什么相关系数越低，分散化收益通常越大?

B3. 多元正态下“不相关”和“独立”是什么关系？一般分布中仍然成立吗?

B4. Mahalanobis distance 与普通欧氏距离有什么区别?

B5. 为什么 t 分布比标准正态尾部更厚?

B6. PCA 第一主成分解释了最大方差，它是否一定是经济上的“市场因子”?

### C 组：混合题

C1. 设 $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$， 为常数向量。推导 $'\boldsymbol{X}$ 的分布。

C2. 设 $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，证明 $(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \sim \chi^2_k$。

C3. 两资产最小方差权重公式为

$$w^* = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

令 $\sigma_1 = 0.2$，$\sigma_2 = 0.3$，$\rho = 0.2$，求 $w^*$。

C4. 设协方差矩阵特征值为 $5, 2, 1, 1$。求第一主成分和前两个主成分解释的方差比例。

C5. 若 $T = Z/\sqrt{U/\nu}$，其中 $Z \sim N(0,1)$，$U \sim \chi^2_\nu$ 且独立，说明 $T$ 的分布以及在 $\nu \to \infty$ 时的极限直觉。

### D 组：经济金融建模题

D1. 设三资产收益均值为 $\boldsymbol{\mu} = (0.06, 0.08, 0.12)'$，权重为 $\boldsymbol{w} = (0.4, 0.4, 0.2)'$。求组合期望收益。

D2. 使用例 5 的协方差矩阵和 D1 的权重，计算组合方差。

D3. 解释为什么金融中的 factor model 可以被看成对协方差矩阵的低维近似。

D4. 若一个投资组合收益近似服从 $N(0.001, 0.02^2)$，求 95% 正态 VaR 的表达式，取 $z_{0.95} = 1.645$。

D5. 在回归中同时检验 3 个线性限制，Wald 统计量大样本下近似服从什么分布？如果统计量为 9.5，你会如何查表判断显著性?

### E 组：挑战 / 证明题

E1. 证明若 $\boldsymbol{\Sigma}$ 正定，则 $\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}$ 是严格凸函数。

E2. 推导全局最小方差组合

$$\boldsymbol{w}_{GMV} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}.$$

E3. 设 $\boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$，谱分解 $\boldsymbol{\Sigma} = Q\Lambda Q'$。证明 $Q'\boldsymbol{X}$ 的协方差矩阵为 $\Lambda$。

E4. 解释为什么 Wald statistic 本质上是估计量与原假设集合之间的标准化距离。

E5. 若收益分布有厚尾，说明正态 VaR 可能如何低估风险，并提出至少两种改进方向。

# 18 练习题参考答案

## A 组答案

A1.

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix}.$$

A2. 令 $= (2, -1)'$，则 $Y =' \boldsymbol{X}$。

$$\mathbb{E}[Y] = 2 \cdot 1 - 2 = 0.$$

$$\mathrm{Var}(Y) =' \boldsymbol{\Sigma}.$$

先算

$$\boldsymbol{\Sigma} = \begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}.$$

所以

$$\mathrm{Var}(Y) = (2, -1) \begin{pmatrix} 5 \\ -3 \end{pmatrix} = 13.$$

A3. 令 $= A\boldsymbol{X} +$。则 $\mathbb{E}[] = A\mathbb{E}[\boldsymbol{X}] +$，所以

$$-\mathbb{E}[] = A(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}]).$$

因此

$$\mathrm{Var}() = \mathbb{E}[A(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})'A'] = A\,\mathrm{Var}(\boldsymbol{X})A'.$$

A4.

$$' = Z_1^2 + Z_2^2 + Z_3^2 \sim \chi_3^2.$$

A5. 令 $Z = (X - \mu)/\sigma$，则 $Z \sim N(0, 1)$，所以 $Z^2 \sim \chi_1^2$。

A6.

$$\sigma_p^2 = 0.6^2(0.2)^2 + 0.4^2(0.3)^2 + 2(0.6)(0.4)(0.5)(0.2)(0.3).$$

计算得

$$0.0144 + 0.0144 + 0.0144 = 0.0432.$$

## B 组答案

B1. 对任意 ，$'\boldsymbol{\Sigma} = \mathrm{Var}('\boldsymbol{X}) \geq 0$，所以协方差矩阵半正定。

B2. 组合方差含有 $2w(1-w)\rho\sigma_1\sigma_2$。$\rho$ 越低，该项越小，组合风险越低，分散化收益越大。

B3. 多元正态下，不相关等价于独立。一般分布中不成立。

B4. 欧氏距离把所有方向同等看待；Mahalanobis distance 用 $\boldsymbol{\Sigma}^{-1}$ 调整不同方向的波动大小和相关性。高方差方向上的同样偏离没有低方差方向上那么“异常”。

B5. 因为 t 统计量的分母也随机，方差估计带来额外不确定性，小自由度时尾部更厚。

B6. 不一定。PCA 第一主成分是最大方差方向，是统计对象。它是否对应市场因子需要经济解释和实证验证。

## C 组答案

C1. 因多元正态的线性组合仍为正态，

$$'\boldsymbol{X} \sim \mathcal{N}('\boldsymbol{\mu},'\boldsymbol{\Sigma}).$$

C2. 令 $= \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu})$，则 $\sim \mathcal{N}(\mathbf{0}, I_k)$，所以 $' \sim \chi^2_k$。而 $' = (\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$。

C3.

$$w^* = \frac{0.3^2 - 0.2(0.2)(0.3)}{0.2^2 + 0.3^2 - 2(0.2)(0.2)(0.3)} = \frac{0.09 - 0.012}{0.04 + 0.09 - 0.024} = \frac{0.078}{0.106} \approx 0.736.$$

C4. 总方差为 $5+2+1+1=9$。第一主成分解释比例为 $5/9 \approx 55.6\%$。前两个主成分解释比例为 $7/9 \approx 77.8\%$。

C5. $T \sim t_\nu$。当 $\nu \to \infty$ 时，$U/\nu \to 1$，所以 $T$ 趋近于 $Z \sim N(0,1)$。

## D 组答案

D1.

$$\mathbb{E}[R_p] = 0.4(0.06) + 0.4(0.08) + 0.2(0.12) = 0.08.$$

D2. 使用例 5 的矩阵和 $\boldsymbol{w} = (0.4, 0.4, 0.2)'$。先算

$$\boldsymbol{\Sigma}\boldsymbol{w} = \begin{pmatrix} 0.04 & 0.01 & 0 \\ 0.01 & 0.09 & 0.02 \\ 0 & 0.02 & 0.16 \end{pmatrix} \begin{pmatrix} 0.4 \\ 0.4 \\ 0.2 \end{pmatrix} = \begin{pmatrix} 0.020 \\ 0.044 \\ 0.040 \end{pmatrix}.$$

所以

$$\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} = 0.4(0.020) + 0.4(0.044) + 0.2(0.040) = 0.0336.$$

D3. 因子模型把收益写成 $\boldsymbol{R} = B\boldsymbol{f}+$。如果因子维度远小于资产数量，则协方差矩阵近似为

$$\mathrm{Var}(\boldsymbol{R}) = B\,\mathrm{Var}(\boldsymbol{f})B' + \mathrm{Var}().$$

这用少数共同因子解释大部分协方差结构，因此是低维近似。

D4. 损失 $L=-R$。95% 正态 VaR 为

$$\mathrm{VaR}_{0.95}=-0.001+1.645(0.02)=0.0319.$$

D5. 近似服从 $\chi^2_3$。若统计量为 9.5，需要查 $\chi^2_3$ 的 95% 或 99% 临界值。95% 临界值约为 7.815，所以在 5% 水平拒绝；99% 临界值约为 11.345，所以在 1% 水平不拒绝。

## E 组答案

E1. 函数 $f(\boldsymbol{w})=\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}$ 的 Hessian 为 $2\boldsymbol{\Sigma}$。若 $\boldsymbol{\Sigma}$ 正定，则 $2\boldsymbol{\Sigma}$ 正定，因此 $f$ 严格凸。

E2. 见正文推导。设 Lagrangian 为

$$\mathcal{L}=\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}-\lambda(\mathbf{1}'\boldsymbol{w}-1).$$

FOC 为

$$2\boldsymbol{\Sigma}\boldsymbol{w}=\lambda\mathbf{1}.$$

得

$$\boldsymbol{w}=\frac{\lambda}{2}\boldsymbol{\Sigma}^{-1}\mathbf{1}.$$

用 $\mathbf{1}'\boldsymbol{w}=1$ 解出 $\lambda/2=1/(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})$，所以

$$\boldsymbol{w}_{GMV}=\frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}.$$

E3.

$$\mathrm{Var}(Q'\boldsymbol{X})=Q'\,\mathrm{Var}(\boldsymbol{X})Q=Q'\boldsymbol{\Sigma}Q.$$

因 $\boldsymbol{\Sigma}=Q\Lambda Q'$，所以

$$Q'\boldsymbol{\Sigma}Q=Q'Q\Lambda Q'Q=\Lambda.$$

E4. Wald statistic 形如

$$(\hat{\theta}-\theta_0)'\widehat{V}^{-1}(\hat{\theta}-\theta_0).$$

它衡量估计值与原假设之间的距离，但用 $\widehat{V}^{-1}$ 标准化。若某方向估计很不精确，协方差大，则同样偏离不那么显著；若某方向估计很精确，则同样偏离更显著。

E5. 厚尾意味着极端损失概率高于正态分布预测。正态 VaR 会低估尾部风险。改进方向包括：历史模拟 VaR、t 分布或 skew-t 分布、GARCH 波动模型、极值理论 EVT、压力测试和情景分析。

## 19 后续学习连接

本章完成了“多元概率 - 计量推断 - 金融风险”的桥梁。下一章可以自然进入计算统计和模拟方法:

Monte Carlo, bootstrap, simulation-based inference, risk simulation.

如果说 Chapter 29 给出了多元正态和二次型的解析公式，那么 Chapter 30 将回答：当解析公式不好用时，我们如何用计算实验逼近期望、分布、标准误、置信区间和风险指标。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 30

# Monte Carlo、Bootstrap 与模拟推断

前面三章完成了从概率到回归、从极限定理到多元正态与金融风险的过渡。本章把这些理论变成可计算的研究工具：当积分无法解析、分布太复杂、标准误难以手推、动态模型没有闭式解时，我们用随机数来近似期望、分布、风险和估计量的不确定性。Monte Carlo 不是“电脑帮忙算很多次”这么简单，它背后的数学是 LLN、CLT、条件期望、投影与极限定理；它的经济含义是用模拟实验理解模型、政策、风险和估计误差。

math4econ personalized notes
2026

# 目录

**Chapter 30 总览** **2**

**1 本章为什么重要** **2**
1.1 和前几章的关系 . . . . . 3

**2 Monte Carlo 估计期望** **3**
2.1 基本问题 . . . . . 3
2.2 无偏性与方差 . . . . . 4
2.3 Monte Carlo CLT . . . . . 5

**3 随机数生成与分布模拟** **6**
3.1 均匀分布是起点 . . . . . 6
3.2 反函数法 inverse transform . . . . . 6
3.3 多元正态模拟 . . . . . 7

**4 模拟积分与经济期望** **7**
4.1 从积分到样本平均 . . . . . 7
4.2 期望效用例子 . . . . . 8

**5 模拟误差、置信区间与方差缩减** **8**
5.1 Monte Carlo 置信区间 . . . . . 8
5.2 方差缩减 variance reduction . . . . . 9

**6 Bootstrap: 用样本近似抽样分布** **9**
6.1 问题: 标准误从哪里来 . . . . . 9
6.2 Bootstrap 标准误 . . . . . 10
6.3 Bootstrap 置信区间 . . . . . 10

**7 回归中的 Bootstrap** **11**
7.1 Pairs bootstrap . . . . . 11
7.2 Residual bootstrap . . . . . 11
7.3 Wild bootstrap . . . . . 11

**8 Permutation Test 与随机化推断** **11**
8.1 随机分配的思想 . . . . . 12
8.2 基本步骤 . . . . . 12
8.3 和 bootstrap 的区别 . . . . . 12

**9 金融风险模拟: VaR 与 Expected Shortfall** **13**
9.1 VaR . . . . . 13
9.2 Expected Shortfall . . . . . 13
9.3 参数模拟法 . . . . . 14

**10 路径模拟: 资产价格与动态模型** **14**

**11 模型模拟与政策实验** **15**
11.1 基本流程 . . . . . 15
11.2 Common random numbers 的政策含义 . . . . . 16

**12 MCMC 入口: 当不能直接抽样时** **16**
12.1 Markov Chain Monte Carlo . . . . . 16
12.2 Metropolis-Hastings 直觉 . . . . . 16

**13 模拟研究的可复现性与诊断** **17**
13.1 随机种子 . . . . . 17
13.2 收敛诊断 . . . . . 17

**14 Worked Examples** **17**

**15 常见错误与辨析** **19**

**16 本章小结图谱** **20**

**17 练习题** **20**
A 组: 基础计算 . . . . . 20
B 组: 概念判断 . . . . . 21
C 组: 混合题 . . . . . 21
D 组: 经济金融建模题 . . . . . 21
E 组: 挑战 / 证明题 . . . . . 21

**18 练习题参考答案** **23**
A 组答案 . . . . . 23
B 组答案 . . . . . 24
C 组答案 . . . . . 24
D 组答案 . . . . . 25
E 组答案 . . . . . 25

**19 后续学习连接** **26**

# Chapter 30 总览

| | |
|---|---|
| **本章主题** | Monte Carlo simulation, bootstrap, permutation test, simulation-based risk measurement, MCMC 入口。 |
| **核心问题** | 当数学模型给出一个期望、积分、估计量分布、风险测度或动态路径, 但解析公式很难写出时, 如何用随机模拟近似它, 并衡量模拟误差? |
| **数学基础** | 大数定律 LLN, 中心极限定理 CLT, 条件期望, 分布收敛, Slutsky theorem, quantile, covariance matrix。 |
| **经济金融应用** | 计算期望效用, 风险溢价, VaR/ES, 资产价格路径, 动态模型仿真, bootstrap 标准误, 随机化检验, 事件研究和政策实验。 |
| **后续连接** | 计量中的 bootstrap inference, simulated method of moments, Bayesian MCMC, 资产定价数值模拟, dynamic structural estimation。 |

**直觉盒子**

Monte Carlo 的核心思想只有一句话: 如果你想知道 $\mathbb{E}[g(X)]$, 但积分

$$\int g(x) f_X(x)\,\mathrm{d}x$$

不好算, 那就从分布中抽很多个样本 $X_1, \dots, X_N$, 用样本平均

$$\frac{1}{N}\sum_{i=1}^{N} g(X_i)$$

近似它。大数定律保证它会接近真值; 中心极限定理告诉你误差大约有多大。

## 1 本章为什么重要

经济金融中有很多量看起来只是一个期望, 但实际并不好解析计算。例如:

- 风险资产的 expected utility: $\mathbb{E}[u(W_0(1+R))]$;
- 期权价值: $\mathbb{E}[m_{t+1}\max(S_{t+1}-K,0)]$;
- 投资组合的 99% VaR 和 Expected Shortfall;
- 一个动态规划模型中未来状态的条件期望;
- 一个复杂估计量的标准误;
- 一个政策实验在有限样本下的随机化分布。

这些对象都可以写成积分或分布函数，但通常没有漂亮闭式解。Monte Carlo 与 bootstrap 提供了一个共同思路：**用可重复的随机实验来近似理论对象**。

### 1.1 和前几章的关系

| 前置章节 | 本章如何使用 |
|---|---|
| Ch26 概率论 | 随机变量、期望、方差、条件分布是模拟对象。 |
| Ch27 条件期望与回归 | 条件均值、投影、残差重抽样和 regression bootstrap 都依赖这些概念。 |
| Ch28 极限定理 | LLN 保证模拟均值收敛，CLT 给出 Monte Carlo 标准误。 |
| Ch29 多元正态与金融风险 | 多元正态模拟、投资组合风险、VaR/ES 和二次型分布是核心应用。 |
| Ch25 数值方法 | 求根、优化和动态规划经常需要和模拟结合。 |

**常见误区**

模拟不是“随便随机跑很多次”。严肃的 simulation 必须回答四个问题：

1. 你在近似哪个理论对象？
2. 随机数是否来自正确分布？
3. 模拟误差有多大？
4. 结果是否可复现 reproducible？

不回答这些问题，模拟图再漂亮也不可靠。

## 2 Monte Carlo 估计期望

### 2.1 基本问题

设 $X$ 是随机变量，我们想计算

$$\mu = \mathbb{E}[g(X)].$$

如果能从 $X$ 的分布中独立抽样

$$X_1, X_2, \dots, X_N,$$

则 Monte Carlo estimator 定义为

$$\boxed{\widehat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} g(X_i).}$$

**定义 / 定理**

**Monte Carlo estimator.** 对目标 $\mu = \mathbb{E}[g(X)]$, 若 $X_i$ 是来自 $X$ 分布的独立同分布样本, 则

$$\widehat{\mu}_N = N^{-1} \sum_{i=1}^{N} g(X_i)$$

是 $\mu$ 的 Monte Carlo 估计量。

## 2.2 无偏性与方差

若 $\mathbb{E}|g(X)| < \infty$, 则

$$\mathbb{E}[\widehat{\mu}_N] = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}[g(X_i)] = \mu.$$

所以 Monte Carlo 均值是无偏的。

若 $\mathrm{Var}(g(X)) = \sigma_g^2 < \infty$, 且样本独立, 则

$$\mathrm{Var}(\widehat{\mu}_N) = \mathrm{Var}\left( \frac{1}{N} \sum_{i=1}^{N} g(X_i) \right) = \frac{1}{N^2} \sum_{i=1}^{N} \mathrm{Var}(g(X_i)) = \frac{\sigma_g^2}{N}.$$

标准误为

$$\boxed{\mathrm{se}(\widehat{\mu}_N) = \frac{\sigma_g}{\sqrt{N}}.}$$

**直觉盒子**

Monte Carlo 的误差下降速度是 $1/\sqrt{N}$, 而不是 $1/N$。这很慢。想把误差缩小到原来的十分之一, 需要把模拟次数增加到原来的 100 倍。这是所有模拟研究都绕不开的基本成本。

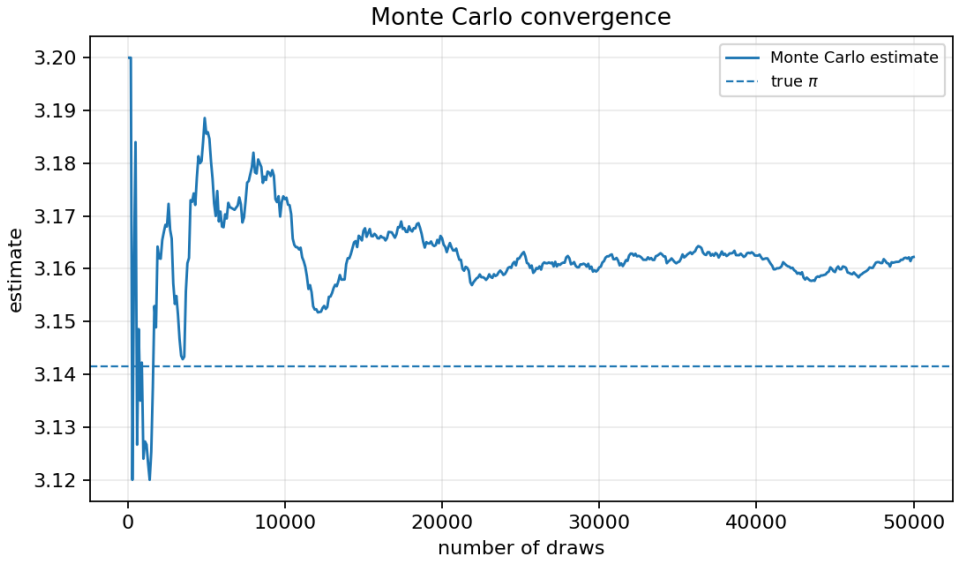



图 1: Monte Carlo 估计会随着抽样次数增加而收敛, 但路径存在随机波动。

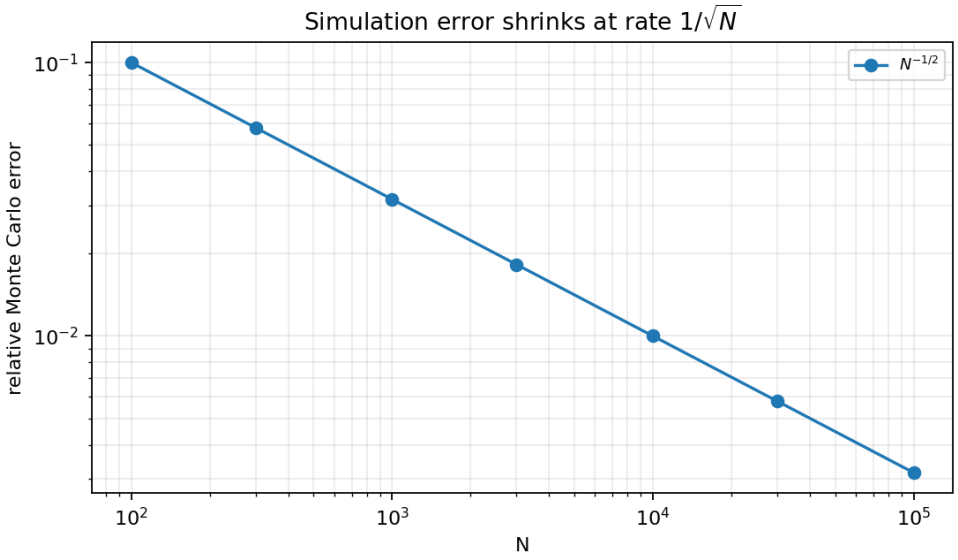



图 2: Monte Carlo 标准误的数量级是 $N^{-1/2}$。

## 2.3 Monte Carlo CLT

由中心极限定理, 若 $\mathrm{Var}(g(X)) = \sigma_g^2 < \infty$, 则

$$\sqrt{N}(\widehat{\mu}_N - \mu) \xrightarrow{d} N(0, \sigma_g^2).$$

等价地,

$$\widehat{\mu}_N \approx N\left(\mu, \frac{\sigma_g^2}{N}\right).$$

实践中 $\sigma_g^2$ 不知道, 用样本方差估计:

$$\widehat{\sigma}_g^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(g(X_i)-\widehat{\mu}_N\right)^2.$$

Monte Carlo 标准误为

$$\boxed{\widehat{\mathsf{se}}(\widehat{\mu}_N) = \frac{\widehat{\sigma}_g}{\sqrt{N}}.}$$

**方法模板**

**Monte Carlo 报告最小规范**

1. 说明目标量: 例如 $\mathbb{E}[g(X)]$, VaR, ES, policy value。
2. 说明抽样分布和随机数种子。
3. 报告模拟次数 $N$。
4. 报告估计值和 Monte Carlo standard error。
5. 做至少一个稳健性检查: 增加 $N$, 换 seed, 或比较解析基准。

# 3 随机数生成与分布模拟

## 3.1 均匀分布是起点

大多数计算机随机数生成器先生成近似均匀分布

$$U \sim U(0,1).$$

其他分布通常由 $U$ 转换而来。

## 3.2 反函数法 inverse transform

若随机变量 $X$ 的 CDF 为 $F$, 且 $F$ 可逆, 则

$$\boxed{X = F^{-1}(U), \qquad U \sim U(0,1)}$$

服从分布 $F$。

**Worked Example**

**例 1: 指数分布模拟**

指数分布 CDF 为

$$F(x) = 1 - e^{-\lambda x}, \qquad x \geq 0.$$

令 $U = F(X)$, 则

$$U = 1 - e^{-\lambda X}.$$

解出

$$e^{-\lambda X} = 1 - U,$$

$$X = -\frac{1}{\lambda}\log(1-U).$$

因为 $1-U$ 和 $U$ 同分布, 常写成

$$\boxed{X = -\frac{1}{\lambda}\log U.}$$

经济金融中, 指数分布可以用来模拟等待时间、违约到达时间、交易到达时间等。

### 3.3 多元正态模拟

若要模拟

$$X \sim N(\mu, \Sigma),$$

可以先模拟

$$Z \sim N(0, I),$$

再找矩阵 $L$ 使得

$$LL' = \Sigma.$$

例如 Cholesky decomposition。令

$$\boxed{X = \mu + LZ.}$$

则

$$\mathbb{E}[X] = \mu,$$

$$\mathrm{Var}(X) = L\,\mathrm{Var}(Z)L' = LL' = \Sigma.$$

**直觉盒子**

多元正态模拟本质上是把独立标准正态 $Z$ 通过线性变换 $L$ 拉伸、旋转、相关化。Ch29 的协方差椭圆与 PCA, 在这里变成了实际生成联合资产收益的算法。

## 4 模拟积分与经济期望

### 4.1 从积分到样本平均

许多经济对象可写成

$$I = \int h(x)\,\mathrm{d}x$$

或

$$\mathbb{E}[g(X)] = \int g(x) f(x)\, \mathrm{d}x.$$

如果维度很高, 传统网格积分会遇到 curse of dimensionality。比如每个维度取 100 个点, 10 维网格需要 $100^{10}$ 个点, 不可能计算。Monte Carlo 的优势是误差主要取决于样本数 $N$, 而不是直接以维度指数爆炸。

### 4.2 期望效用例子

**Worked Example**

**例 2: 用模拟估计期望效用**

设财富为

$$W = W_0(1+R),$$

效用为 CRRA:

$$u(W) = \frac{W^{1-\gamma}}{1-\gamma}, \qquad \gamma \neq 1.$$

若 $R$ 的分布复杂, 想算

$$\mathbb{E}[u(W_0(1+R))].$$

可模拟 $R_1, \dots, R_N$, 计算

$$\widehat{U} = \frac{1}{N}\sum_{i=1}^{N} u(W_0(1+R_i)).$$

若要比较两个投资组合 $A, B$, 可以分别计算

$$\widehat{U}_A = \frac{1}{N}\sum_{i=1}^{N} u(W_0(1+R_{A,i})),$$

$$\widehat{U}_B = \frac{1}{N}\sum_{i=1}^{N} u(W_0(1+R_{B,i})).$$

如果用相同随机数 common random numbers 生成两组情景, 比较差值的方差通常更小。

## 5 模拟误差、置信区间与方差缩减

### 5.1 Monte Carlo 置信区间

由 CLT,

$$\widehat{\mu}_N \pm 1.96 \frac{\widehat{\sigma}_g}{\sqrt{N}}$$

可作为目标 $\mu$ 的近似 95% Monte Carlo confidence interval。

这里的区间只反映 simulation error, 不一定包含模型误差、数据误差或估计误差。比如

你用错误的收益分布模拟，$N$ 再大也只是精确地估计了错误模型。

### 5.2 方差缩减 variance reduction

因为误差下降速度慢，很多技巧试图降低 $\sigma_g^2$。

| 方法 | 思想 | 经济金融例子 |
|---|---|---|
| Antithetic variates | 同时使用 $U$ 和 $1-U$，抵消部分随机波动 | 期权定价、风险测度模拟。 |
| Control variates | 找一个期望已知且与目标高度相关的变量 | 用 Black-Scholes 解析价格辅助复杂期权模拟。 |
| Common random numbers | 比较两个政策时使用同一组随机冲击 | 比较两种税制、投资组合或算法。 |
| Importance sampling | 更多抽取对目标贡献大的稀有区域 | 违约风险、极端损失、尾部概率。 |
| Stratified sampling | 把样本空间分层，每层都抽样 | 异质性家庭模型、收入分布模拟。 |

**常见误区**

方差缩减不是改变目标分布。特别是 importance sampling 需要重新加权，否则会得到有偏估计。模拟中最危险的错误之一就是“抽样分布变了，但权重没改”。

## 6 Bootstrap: 用样本近似抽样分布

### 6.1 问题: 标准误从哪里来

很多估计量 $\widehat{\theta} = T(X_1, \dots, X_n)$ 的标准误不好手推。Bootstrap 的想法是：如果总体分布 $F$ 不知道，就用经验分布 $\widehat{F}_n$ 代替它；从原样本中有放回抽样，反复计算估计量，得到估计量的近似抽样分布。

**定义 / 定理**

**Nonparametric bootstrap.** 给定样本 $X_1, \dots, X_n$，重复 $B$ 次：

1. 从原样本中有放回抽取 $n$ 个观测，得到 bootstrap sample $X_1^*, \dots, X_n^*$；
2. 计算 $\widehat{\theta}^* = T(X_1^*, \dots, X_n^*)$。

$\widehat{\theta}^{*(1)}, \dots, \widehat{\theta}^{*(B)}$ 的分布用于近似 $\widehat{\theta}$ 的抽样分布。

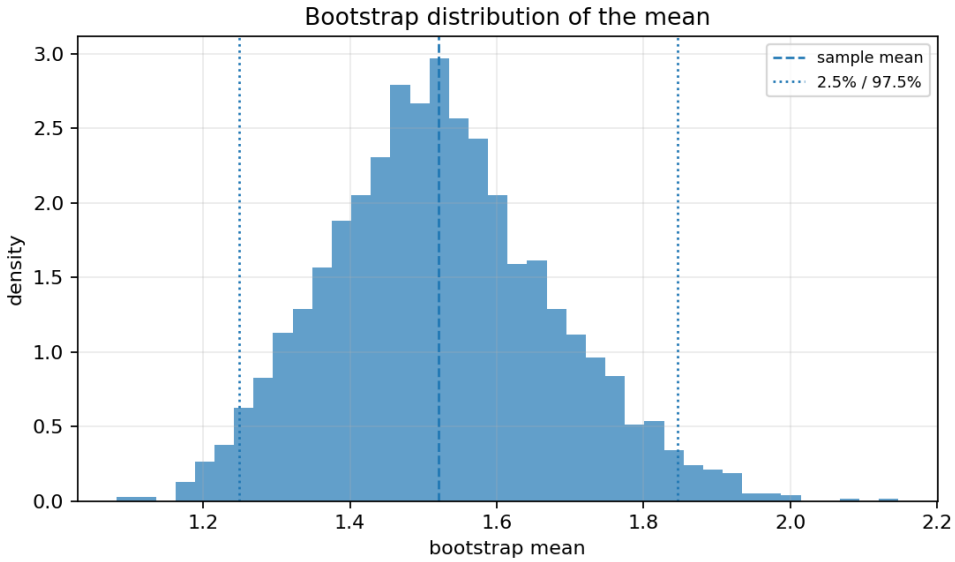



图 3: Bootstrap 用重抽样产生估计量分布, 进而得到标准误和置信区间。

## 6.2 Bootstrap 标准误

若有 $B$ 个 bootstrap 估计

$$\widehat{\theta}_1^*, \ldots, \widehat{\theta}_B^*,$$

则 bootstrap standard error 为

$$\boxed{\widehat{\mathrm{se}}_{boot}(\widehat{\theta}) = \left[ \frac{1}{B-1} \sum_{b=1}^{B} \left( \widehat{\theta}_b^* - \bar{\theta}^* \right)^2 \right]^{1/2},}$$

其中

$$\bar{\theta}^* = \frac{1}{B} \sum_{b=1}^{B} \widehat{\theta}_b^*.$$

## 6.3 Bootstrap 置信区间

最简单的 percentile interval 是取 bootstrap 分布的分位数:

$$\boxed{[q_{0.025}^*, q_{0.975}^*].}$$

它直观, 但不是所有情况下都最好。更高级的区间包括 basic bootstrap interval, studentized bootstrap interval, BCa interval。本章只要求理解 percentile interval 的逻辑。

**方法模板**

**Bootstrap 操作步骤**

1. 原样本估计 $\widehat{\theta}$。

2. 重复 $B$ 次有放回抽样。
3. 每次重新估计 $\widehat{\theta}_b^*$。
4. 用 $\{\widehat{\theta}_b^*\}$ 的标准差估计标准误。
5. 用分位数构造置信区间。

# 7 回归中的 Bootstrap

## 7.1 Pairs bootstrap

对回归数据 $(Y_i, X_i)$, pairs bootstrap 是从观测对

$$(Y_i, X_i)$$

中有放回抽样, 每次重新跑 OLS。它适合把整个观测视为 i.i.d. 抽样的情形。

## 7.2 Residual bootstrap

若固定设计 $X$ 更合理, 可使用 residual bootstrap:

1. 原回归得到 $\widehat{Y}_i = X_i'\widehat{\beta}$ 与残差 $\widehat{\varepsilon}_i$;
2. 对残差有放回抽样得到 $\widehat{\varepsilon}_i^*$;
3. 构造 $Y_i^* = X_i'\widehat{\beta} + \widehat{\varepsilon}_i^*$;
4. 用 $(Y_i^*, X_i)$ 重新估计。

## 7.3 Wild bootstrap

若存在异方差, residual bootstrap 可能不合适。Wild bootstrap 常写为

$$Y_i^* = X_i'\widehat{\beta} + \widehat{\varepsilon}_i v_i,$$

其中 $v_i$ 是均值 0、方差 1 的随机变量。它保留了残差大小随 $X_i$ 变化的结构。

**常见误区**

Bootstrap 不是万能的。时间序列有序列相关时不能随便打乱观测; 聚类数据不能把个体观测当成 i.i.d.; 重尾分布下某些统计量的 bootstrap 可能表现很差。选择 bootstrap 方法时必须尊重数据生成结构。

# 8 Permutation Test 与随机化推断

## 8.1 随机分配的思想

在随机实验中, treatment assignment 是随机的。若 sharp null hypothesis 成立, 即处理对每个个体都没有影响, 那么不同 treatment label 的分配只会改变标签, 不会改变潜在结果。因此可以通过重新随机打乱处理标签来构造检验统计量的随机化分布。

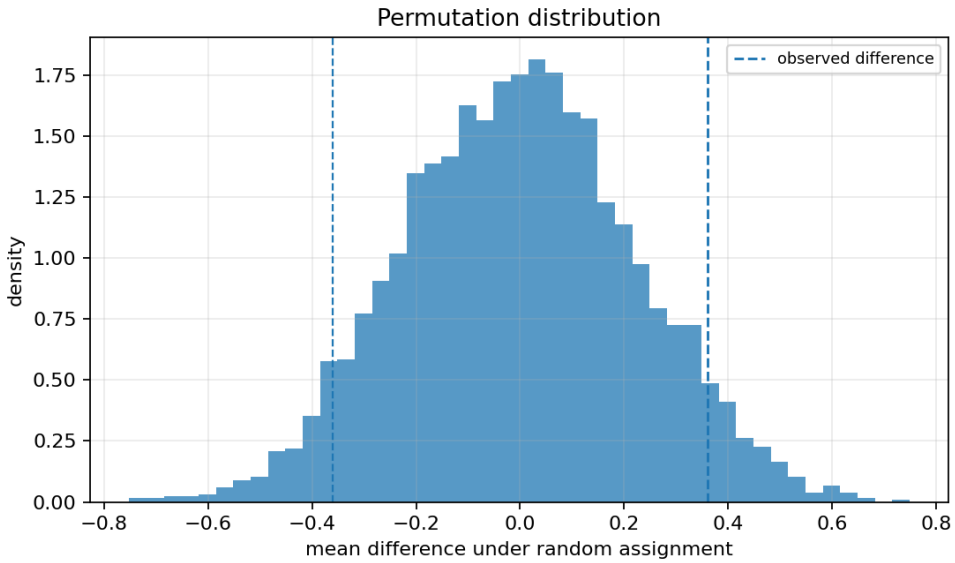



图 4: Permutation test 用重新随机分配标签产生零假设下的统计量分布。

## 8.2 基本步骤

设统计量为均值差

$$T = \bar{Y}_1 - \bar{Y}_0.$$

Permutation test 步骤:

1. 计算真实标签下的 $T_{obs}$;
2. 随机打乱 treatment label, 保持处理组人数不变;
3. 每次计算 $T^*$;
4. 比较 $|T_{obs}|$ 在 permutation distribution 中的位置;
5. 得到随机化 $p$-value。

## 8.3 和 bootstrap 的区别

| | **Bootstrap** | **Permutation test** |
|---|---|---|
| 核心问题 | 估计抽样分布和标准误 | 构造零假设下的随机化分布 |
| 抽样方式 | 有放回重抽样 | 打乱标签或重新分配 |

| | | |
|---|---|---|
| 典型用途 | 标准误、置信区间 | 实验检验、随机化推断 |
| 依赖逻辑 | 样本经验分布近似总体分布 | 随机分配机制已知 |

# 9 金融风险模拟: VaR 与 Expected Shortfall

## 9.1 VaR

设损失变量为 $L$. 置信水平 $\alpha$ 的 Value at Risk 定义为损失分布的 $\alpha$ 分位数:

$$\boxed{\mathrm{VaR}_\alpha(L) = \inf\{\ell : \mathbb{P}(L \le \ell) \ge \alpha\}.}$$

例如 95% VaR 是使得损失不超过该值的概率达到 95% 的阈值。

## 9.2 Expected Shortfall

Expected Shortfall, 也叫 Conditional VaR, 衡量尾部平均损失:

$$\boxed{\mathrm{ES}_\alpha(L) = \mathbb{E}[L \mid L \ge \mathrm{VaR}_\alpha(L)]}$$

在连续分布下这个写法很直观。ES 比 VaR 更关注尾部严重程度。

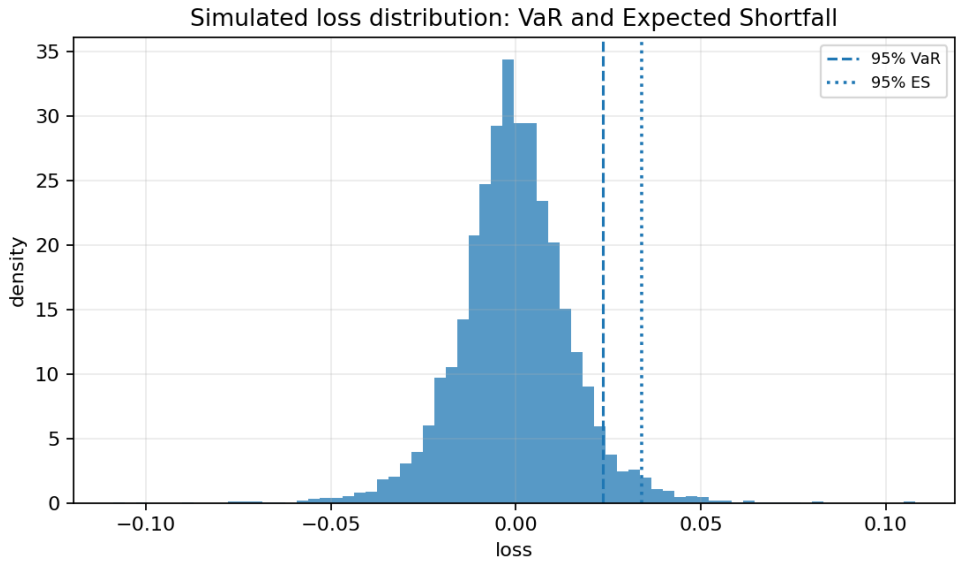



图 5: VaR 是尾部阈值, ES 是超过阈值后的平均损失。

**Worked Example**

**例 3: 用历史模拟法估计 VaR 和 ES**

设有投资组合收益样本 $R_1, \dots, R_N$, 损失定义为

$$L_i = -R_i.$$

把损失从小到大排序:

$$L_{(1)} \le \cdots \le L_{(N)}.$$

95% VaR 可估计为

$$\widehat{\mathrm{VaR}}_{0.95} = L_{(\lceil 0.95N \rceil)}.$$

ES 可估计为尾部平均:

$$\widehat{\mathrm{ES}}_{0.95} = \frac{1}{\#\{i : L_i \ge \widehat{\mathrm{VaR}}_{0.95}\}} \sum_{i: L_i \ge \widehat{\mathrm{VaR}}_{0.95}} L_i.$$

历史模拟法简单, 但假设过去收益分布能代表未来, 对结构变化和极端事件敏感。

### 9.3 参数模拟法

如果假设资产收益

$$R \sim N(\mu, \Sigma),$$

则可以模拟很多次多元正态收益, 计算组合损失

$$L = -w'R,$$

再取样本分位数和尾部平均。若收益存在厚尾, 可以改用 $t$ 分布或 bootstrap historical returns。

## 10 路径模拟: 资产价格与动态模型

许多金融对象依赖整条路径, 而不是只依赖终点。例如 barrier option, drawdown risk, dynamic portfolio, default time, consumption path。路径模拟就是反复生成

$$X_{t+1} = F(X_t, \varepsilon_{t+1}).$$

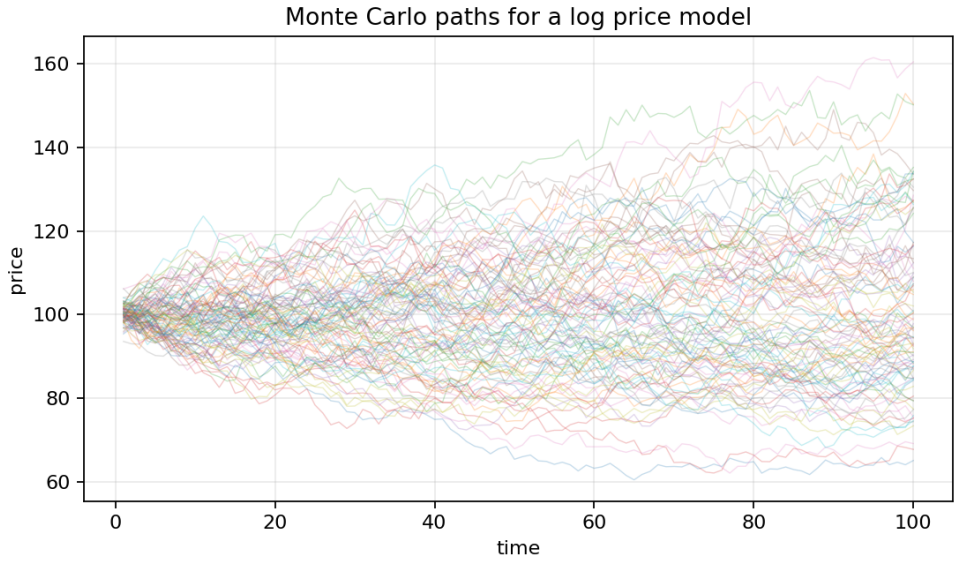



图 6: 随机路径模拟可以展示终点分布、路径风险和极端情景。

**Worked Example**

**例 4: 离散时间 log price 模型**

假设

$$\log S_{t+1} = \log S_t + \mu + \sigma \varepsilon_{t+1},$$

其中 $\varepsilon_{t+1} \sim N(0,1)$。给定 $S_0$, 生成一条路径:

1. 抽取 $\varepsilon_1, \dots, \varepsilon_T$;
2. 递推计算 $\log S_t$;
3. 转换为 $S_t = \exp(\log S_t)$。

重复很多次可得到终点价格分布、最大回撤分布、期权 payoff 分布等。

# 11 模型模拟与政策实验

在宏观和结构模型中, 模拟常用于回答 counterfactual question: 如果参数或政策改变, 模型生成的变量路径如何变化?

## 11.1 基本流程

**方法模板**

**经济模型仿真实验流程**

1. 写清状态变量、控制变量和冲击。
2. 校准或估计参数。

3. 指定冲击分布和初始状态。
4. 模拟 baseline economy。
5. 改变政策或参数, 模拟 counterfactual economy。
6. 用同一组随机数比较两种情景, 降低比较方差。
7. 报告均值、分位数、福利、风险、置信区间。

### 11.2 Common random numbers 的政策含义

如果比较两个政策 A 和 B, 最好在相同冲击路径下比较结果:

$$Y^A(\varepsilon_1, \dots, \varepsilon_T) - Y^B(\varepsilon_1, \dots, \varepsilon_T).$$

这样差值更能反映政策差异, 而不是被不同随机冲击掩盖。

## 12 MCMC 入口: 当不能直接抽样时

Monte Carlo 假设我们能从目标分布抽样。但有时目标分布只知道比例形式:

$$\pi(\theta \mid y) \propto L(y \mid \theta) p(\theta),$$

归一化常数很难计算。这在 Bayesian inference 中非常常见。

### 12.1 Markov Chain Monte Carlo

MCMC 的思想是构造一个 Markov chain

$$\theta^{(1)}, \theta^{(2)}, \dots$$

使其长期平稳分布正好是目标分布 $\pi(\theta \mid y)$。然后用链的后半段样本近似 posterior expectation:

$$\mathbb{E}[h(\theta) \mid y] \approx \frac{1}{M} \sum_{m=1}^{M} h(\theta^{(m)}).$$

### 12.2 Metropolis-Hastings 直觉

从当前点 $\theta$ 出发, 提议一个新点 $\theta'$, 若 posterior density 更高则容易接受, 若更低则以一定概率接受。这样链既会去高概率区域, 又不会完全困在局部峰值。

**常见误区**

MCMC 样本不是 i.i.d.。相邻样本相关, 所以有效样本量 effective sample size 可能远小于迭代次数。使用 MCMC 时必须检查 burn-in、mixing、autocorrelation 和多链收敛。

# 13 模拟研究的可复现性与诊断

## 13.1 随机种子

模拟研究必须设置 random seed。seed 不改变理论结果, 但保证别人能复现实验路径。报告中至少应说明：模拟次数、seed、分布设定、参数设定和代码版本。

## 13.2 收敛诊断

常见诊断包括:

- 把 $N$ 加倍, 看结果是否稳定;
- 报告 Monte Carlo standard error;
- 多个 seed 下重复;
- 与解析解或简单 benchmark 比较;
- 对极端尾部事件增加样本量;
- 画 running average 和 histogram。

**常见误区**

不要只报告“模拟结果为 0.1234”。如果 Monte Carlo standard error 是 0.02, 那么小数点后第三四位没有意义。模拟结果的有效数字必须和误差规模匹配。

# 14 Worked Examples

**Worked Example**

**例 5: 估计正态尾部概率**

目标是

$$p = \mathbb{P}(Z > 2), \qquad Z \sim N(0,1).$$

直接 Monte Carlo: 抽 $Z_1, \dots, Z_N$, 定义

$$I_i = \mathbf{1}\{Z_i > 2\}.$$

估计量为

$$\widehat{p} = \frac{1}{N}\sum_{i=1}^{N} I_i.$$

因为 $I_i$ 是 Bernoulli random variable,

$$\mathrm{Var}(I_i) = p(1-p).$$

所以

$$\mathrm{se}(\widehat{p}) = \sqrt{\frac{p(1-p)}{N}}.$$

实际中用

$$\widehat{\mathrm{se}}(\widehat{p}) = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{N}}.$$

这个例子说明尾部概率难估: 如果 $p$ 很小, 很多模拟抽样根本落不到尾部。

**Worked Example**

**例 6: Bootstrap 样本均值的标准误**

给定样本 $X_1, \ldots, X_n$, 原估计量为

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Bootstrap 做法: 每次从原样本中有放回抽 $n$ 个数, 得到 $\bar{X}^{*(b)}$。重复 $B$ 次。标准误为

$$\widehat{\mathrm{se}}_{boot}(\bar{X}) = \left[\frac{1}{B-1}\sum_{b=1}^{B}(\bar{X}^{*(b)} - \bar{X}^{*})^2\right]^{1/2}.$$

对均值而言, 解析标准误 $s/\sqrt{n}$ 很容易算, bootstrap 主要是帮助你理解重抽样逻辑。真正需要 bootstrap 的是中位数、分位数、复杂回归估计量、非线性函数等标准误不好写的对象。

**Worked Example**

**例 7: 投资组合 VaR 模拟**

设两资产收益

$$R \sim N(\mu, \Sigma),$$

权重为 $w$. 投资组合收益

$$R_p = w'R,$$

损失

$$L = -R_p.$$

模拟步骤：

1. 用 Cholesky 分解生成 $R^{(1)}, \dots, R^{(N)}$；
2. 计算 $L_i = -w' R^{(i)}$；
3. 排序 $L_{(1)} \le \cdots \le L_{(N)}$；
4. 取 $L_{(\lceil \alpha N \rceil)}$ 作为 VaR；
5. 对超过 VaR 的尾部损失取平均作为 ES。

如果 $R$ 真的是多元正态，VaR 也可解析计算；但模拟法更容易扩展到非正态、期权组合和路径依赖风险。

**Worked Example**

**例 8：模拟一个 AR(1) 过程**

设

$$y_t = \rho y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2).$$

模拟步骤：

1. 给定 $y_0$；
2. 抽取 $\varepsilon_1, \dots, \varepsilon_T$；
3. 递推 $y_t = \rho y_{t-1} + \varepsilon_t$；
4. 重复 $M$ 条路径，研究样本均值、方差、最大值或 hitting time。

如果 $|\rho| < 1$，理论平稳方差为

$$\mathrm{Var}(y_t) = \frac{\sigma^2}{1 - \rho^2}.$$

模拟方差可与该解析结果比较，作为代码检查。

## 15 常见错误与辨析

**常见误区**

**误区 1：模拟次数越大，模型就越真实。**增加 $N$ 只能降低 simulation error，不能修正错误模型。错误分布、错误参数、错误经济机制会导致精确但错误的答案。

**误区 2：忘记报告 Monte Carlo standard error。**没有误差报告的模拟结果不能判断精度。

**误区 3：Bootstrap 时破坏数据结构。**时间序列、面板聚类、空间相关数据不能随意 i.i.d. 重抽样。

**误区 4：把 bootstrap 分布当成总体分布。** Bootstrap 近似的是估计量的抽样分布，

不是原变量真实总体分布。

**误区 5：用 VaR 忽视尾部损失大小。** VaR 只给阈值，不告诉超过阈值后平均损失有多严重。ES 更关注尾部厚度。

**误区 6：MCMC 迭代次数等于有效样本量。** MCMC 样本相关，需要检查 effective sample size 和 mixing。

## 16 本章小结图谱

| 概念 | 核心公式 / 方法 | 经济金融含义 |
|---|---|---|
| Monte Carlo mean | $N^{-1}\sum g(X_i)$ | 用模拟近似期望、积分和价值函数。 |
| MC standard error | $\widehat{\sigma}_g/\sqrt{N}$ | 衡量模拟误差，决定有效数字。 |
| Inverse transform | $X = F^{-1}(U)$ | 从均匀随机数生成目标分布。 |
| Multivariate normal simulation | $X = \mu + LZ$, $LL' = \Sigma$ | 模拟相关资产收益和风险因子。 |
| Bootstrap | 从样本有放回重抽样 | 估计标准误和置信区间。 |
| Permutation test | 打乱处理标签 | 随机实验和 sharp null 检验。 |
| VaR | $\inf\{\ell : P(L \le \ell) \ge \alpha\}$ | 损失分布尾部阈值。 |
| ES | $E[L \mid L \ge VaR]$ | 超过阈值后的平均损失。 |
| MCMC | 构造目标平稳分布的 Markov chain | Bayesian posterior simulation。 |

## 17 练习题

### A 组：基础计算

A1. 设 $X \sim U(0,1)$，用 Monte Carlo 估计 $\mathbb{E}[X^2]$。写出估计量、真值和估计量方差。

A2. 设 $Z \sim N(0,1)$，用 indicator 估计 $P(Z > 1.96)$。写出估计量和标准误公式。

A3. 用反函数法推导如何模拟 CDF 为 $F(x) = x^3$，$0 \le x \le 1$ 的随机变量。

A4. 若 $\widehat{\mu} = 2.5$，Monte Carlo standard error 为 $0.04$，写出近似 95% simulation confidence interval。

A5. 设 bootstrap 估计值为 $1.1, 1.4, 1.2, 1.3, 1.0$，计算 bootstrap standard error。

A6. 给定损失样本 $1, 2, 3, 4, 10, 12, 15, 20, 25, 30$，求 90% historical VaR。

### B 组：概念判断

B1. 为什么 Monte Carlo 误差下降速度是 $1/\sqrt{N}$?

B2. 为什么模拟次数增加不能修正模型设定错误?

B3. Bootstrap 近似的是变量分布还是估计量分布?

B4. Permutation test 和 bootstrap 的核心区别是什么?

B5. 为什么 VaR 可能低估尾部风险?

B6. MCMC 为什么不能把迭代次数直接当作独立样本数?

### C 组：混合题

C1. 设 $X \sim Exp(\lambda)$, 用反函数法推导模拟公式, 并说明 $\lambda$ 对等待时间的影响。

C2. 目标为 $\mu = E[e^X]$, $X \sim N(0,1)$. 写出 Monte Carlo 估计量, 并给出其标准误估计。

C3. 设资产收益样本为 $-0.03, 0.01, 0.02, -0.05, 0.04, -0.01, 0.00, -0.02, 0.03, -0.08$. 以损失 $L = -R$ 计算 80% VaR 和 ES。

C4. 给定回归 $Y_i = X_i'\beta + \varepsilon_i$, 描述 pairs bootstrap 和 residual bootstrap 的区别。

C5. 设 $y_t = 0.8y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$. 写出模拟 $T$ 期路径的算法, 并给出理论平稳方差。

### D 组：经济金融建模题

D1. 你要比较两个税收政策对福利的影响。解释为什么使用 common random numbers 可以提高比较精度。

D2. 一个投资组合包含期权, payoff 非线性且路径依赖。解释为什么解析 VaR 难以计算, 以及如何用 simulation 估计。

D3. 在一个 RCT 中, 样本量很小。说明 permutation test 为什么可能比大样本正态近似更有说服力。

D4. 设模型需要估计 $E[V(k', z')]$, 其中 $z'$ 是离散 Markov shock。说明如何用模拟近似这个条件期望。

D5. 解释在结构模型中, simulation error 和 sampling error 的区别。

### E 组：挑战 / 证明题

E1. 证明 Monte Carlo estimator $\widehat{\mu}_N = N^{-1}\sum g(X_i)$ 的无偏性和方差公式。

E2. 利用 CLT 推导 Monte Carlo 置信区间。

E3. 设 $X \sim N(\mu, \Sigma)$, 证明若 $X = \mu + LZ$, $Z \sim N(0, I)$, $LL' = \Sigma$, 则 $X$ 的协方差为 $\Sigma$。

E4. 设 $x_t = \rho x_{t-1} + \varepsilon_t$, $|\rho| < 1$. 推导平稳方差并说明如何用模拟验证。

E5. 解释 importance sampling 为什么需要权重修正。可用 $E_f[h(X)]$ 和另一抽样密度 $q$ 写出公式。

# 18 练习题参考答案

## A 组答案

A1. 估计量为

$$\widehat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i^2.$$

真值

$$E[X^2] = \int_0^1 x^2 \,\mathrm{d}x = \frac{1}{3}.$$

方差需要 $E[X^4] = 1/5$, 所以

$$\mathrm{Var}(X^2) = \frac{1}{5} - \frac{1}{9} = \frac{4}{45}.$$

因此

$$\mathrm{Var}(\widehat{\mu}) = \frac{4}{45N}.$$

A2. 定义 $I_i = \mathbf{1}\{Z_i > 1.96\}$. 估计量

$$\widehat{p} = \frac{1}{N}\sum_{i=1}^{N} I_i.$$

标准误估计

$$\widehat{\mathrm{se}}(\widehat{p}) = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{N}}.$$

A3. 若 $U \sim U(0,1)$, 令 $U = F(X) = X^3$, 得

$$X = U^{1/3}.$$

A4. 区间为

$$2.5 \pm 1.96(0.04) = [2.4216, 2.5784].$$

A5. 均值为 $1.2$. 离差平方和为

$$(1.1-1.2)^2 + (1.4-1.2)^2 + (1.2-1.2)^2 + (1.3-1.2)^2 + (1.0-1.2)^2 = 0.10.$$

bootstrap standard error 为

$$\sqrt{0.10/(5-1)} = \sqrt{0.025} \approx 0.158.$$

A6. $N = 10$, $\lceil 0.9N \rceil = 9$. 排序后第 9 个损失是 25, 所以 90% VaR 为 25。

## B 组答案

B1. 因为样本均值方差为单个抽样方差除以 $N$, 标准差就是除以 $\sqrt{N}$。CLT 也给出 $\sqrt{N}(\widehat{\mu}-\mu)$ 有非退化极限分布。

B2. 模拟次数增加只会更精确地估计当前模型的目标。如果分布、参数或经济机制设定错了, 模拟会收敛到错误目标。

B3. Bootstrap 主要近似估计量的抽样分布, 例如均值、回归系数、中位数或分位数的分布。

B4. Bootstrap 是有放回重抽样, 用于估计抽样分布; permutation test 是在零假设下重新分配标签, 用于构造随机化检验分布。

B5. VaR 只告诉某个分位点, 不告诉超过该分位点后损失有多严重。两个组合可能 VaR 相同, 但一个组合的尾部损失远大于另一个。

B6. 因为 MCMC 样本来自 Markov chain, 相邻样本相关。相关性越强, 有效独立信息越少。

## C 组答案

C1. 指数分布 CDF 为 $F(x)=1-e^{-\lambda x}$. 令 $U=F(X)$ 得

$$X=-\frac{1}{\lambda}\log(1-U).$$

可写成 $X=-\lambda^{-1}\log U$. $\lambda$ 越大, 平均等待时间 $1/\lambda$ 越短。

C2. 估计量

$$\widehat{\mu}=\frac{1}{N}\sum_{i=1}^{N}e^{X_i}.$$

标准误估计

$$\widehat{\mathrm{se}}=\frac{1}{\sqrt{N}}\left[\frac{1}{N-1}\sum_{i=1}^{N}(e^{X_i}-\widehat{\mu})^2\right]^{1/2}.$$

C3. 收益排序不直接用, 先算损失:

$$0.03,-0.01,-0.02,0.05,-0.04,0.01,0,0.02,-0.03,0.08.$$

排序:

$$-0.04,-0.03,-0.02,-0.01,0,0.01,0.02,0.03,0.05,0.08.$$

80% VaR 取第 $\lceil 0.8\times 10\rceil=8$ 个, 即 $0.03$。ES 为不小于 $0.03$ 的损失均值:

$$(0.03+0.05+0.08)/3=0.0533.$$

C4. Pairs bootstrap 重抽 $(Y_i,X_i)$ 观测对, 适合 i.i.d. 抽样。Residual bootstrap 固定 $X_i$, 重抽残差并生成 $Y_i^*$, 适合固定设计和同方差误差的情形。

C5. 算法: 给定 $y_0$, 抽取 $\varepsilon_1, \ldots, \varepsilon_T$, 递推 $y_t = 0.8y_{t-1} + \varepsilon_t$. 平稳方差为

$$\frac{1}{1-0.8^2} = \frac{1}{0.36} = 2.7778.$$

## D 组答案

D1. 使用相同随机冲击路径时, 两种政策结果的差异更主要来自政策本身, 而不是不同随机抽样带来的噪声。这样差值估计的方差通常更小。

D2. 期权组合 payoff 非线性, 路径依赖意味着损失不只由终点决定, 还由路径最大值、最小值、触碰边界等决定。可以模拟大量资产价格路径, 每条路径计算组合损失, 再从损失样本中估计 VaR 和 ES。

D3. 小样本下 CLT 正态近似可能不准。若处理是随机分配的, permutation test 可以直接利用随机分配机制构造零假设分布, 因而更贴合实验设计。

D4. 如果 $z'$ 是离散 Markov shock, 且转移概率为 $P(z' \mid z)$, 可以精确求和

$$E[V(k', z') \mid z] = \sum_{z'} P(z' \mid z) V(k', z').$$

若状态太多或连续, 可以从条件分布模拟 $z'_1, \ldots, z'_N$, 用样本平均近似。

D5. Sampling error 来自实际数据样本有限, 即估计总体参数的不确定性。Simulation error 来自用有限模拟次数近似模型对象。即使数据无限, 模拟次数有限仍有 simulation error; 即使模拟次数无限, 样本有限仍有 sampling error。

## E 组答案

E1. 无偏性:

$$E[\widehat{\mu}_N] = N^{-1} \sum E[g(X_i)] = \mu.$$

方差: 独立性给出

$$Var(\widehat{\mu}_N) = N^{-2} \sum Var(g(X_i)) = \sigma_g^2 / N.$$

E2. CLT 给出

$$\frac{\widehat{\mu}_N - \mu}{\widehat{\sigma}_g / \sqrt{N}} \approx N(0, 1).$$

因此近似 95% 置信区间为

$$\widehat{\mu}_N \pm 1.96 \frac{\widehat{\sigma}_g}{\sqrt{N}}.$$

E3. 因 $E[Z] = 0$, $Var(Z) = I$, 有

$$E[X] = \mu + LE[Z] = \mu.$$

同时

$$Var(X) = Var(LZ) = LVar(Z)L' = LIL' = LL' = \Sigma.$$

E4. 平稳时 $Var(x_t) = Var(x_{t-1}) = \sigma_x^2$. 由

$$x_t = \rho x_{t-1} + \varepsilon_t$$

得

$$\sigma_x^2 = \rho^2 \sigma_x^2 + \sigma_\varepsilon^2.$$

所以

$$\sigma_x^2 = \frac{\sigma_\varepsilon^2}{1-\rho^2}.$$

可通过模拟长路径、丢弃 burn-in 后计算样本方差, 与理论值比较。

E5. 目标为

$$E_f[h(X)] = \int h(x) f(x)\,\mathrm{d}x.$$

若改从密度 $q$ 抽样, 则

$$E_f[h(X)] = \int h(x)\frac{f(x)}{q(x)} q(x)\,\mathrm{d}x = E_q\left[h(X)\frac{f(X)}{q(X)}\right].$$

因此必须乘权重 $f(X)/q(X)$。不加权就会估计 $E_q[h(X)]$, 不是原目标。

## 19 后续学习连接

Chapter 30 完成后, 你已经拥有计量、金融风险和计算实验所需的核心概率计算工具。下一章建议进入全书最后的地图章节:

高级微观　高级宏观　计量理论　金融数学　AI/机器学习经济学

分别需要哪些数学, 以及前 30 章如何连接这些方向。

**直觉盒子**

如果说 Ch26 到 Ch29 帮你建立了概率和统计理论底座, 那么 Ch30 的作用是让这些理论可以落地运行。以后你读 empirical paper、asset pricing paper、structural estimation 或 computational macro, 看到 simulation, bootstrap, Monte Carlo standard error, VaR, MCMC, 都应该能把它们还原成: 目标对象、抽样机制、近似误差和经济解释。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Chapter 31

# 通向高级课程的数学地图

## Advanced Micro, Macro, Econometrics, Finance, and AI Economics

这是正文的收束章。前 30 章已经把多元积分、向量分析、投影、动态系统、优化、对偶、固定点、动态规划、概率、回归、极限定理、金融风险和模拟推断串成了一套经济金融数学语言。本章不再引入一堆新技术，而是回答三个问题：这些工具以后各自通向哪些高级课？你该怎样复习和补强？整套讲义后面是否还需要附录、公式表和综合习题册？

math4econ personalized notes
2026

# 目录

# Chapter 31 总览

| | |
|---|---|
| **本章定位** | 全书路线图、复习方案、高级课程衔接表。 |
| **核心问题** | 学完这 31 章后，你怎样把数学工具迁移到高级微观、高级宏观、计量理论、金融工程、机器学习与 AI 经济学？哪些内容应该继续深挖，哪些只需够用？ |
| **本章输出** | 课程地图、技能清单、优先级、推荐教材、复习周期、论文阅读模板、综合练习、后续附录建议。 |
| **是否结束** | 正文部分到本章结束。后续最好不再开新“理论章节”，而是做全书公式大全、综合题册、索引、错题清单和考试/论文阅读版压缩讲义。 |

**直觉盒子**

如果说前 30 章是在造工具，本章就是在做工具箱标签。你真正需要的不是“我学过多少章”，而是看到一个模型时能马上判断：它是优化问题、动态系统、概率极限、投影问题、固定点问题，还是数值计算问题。

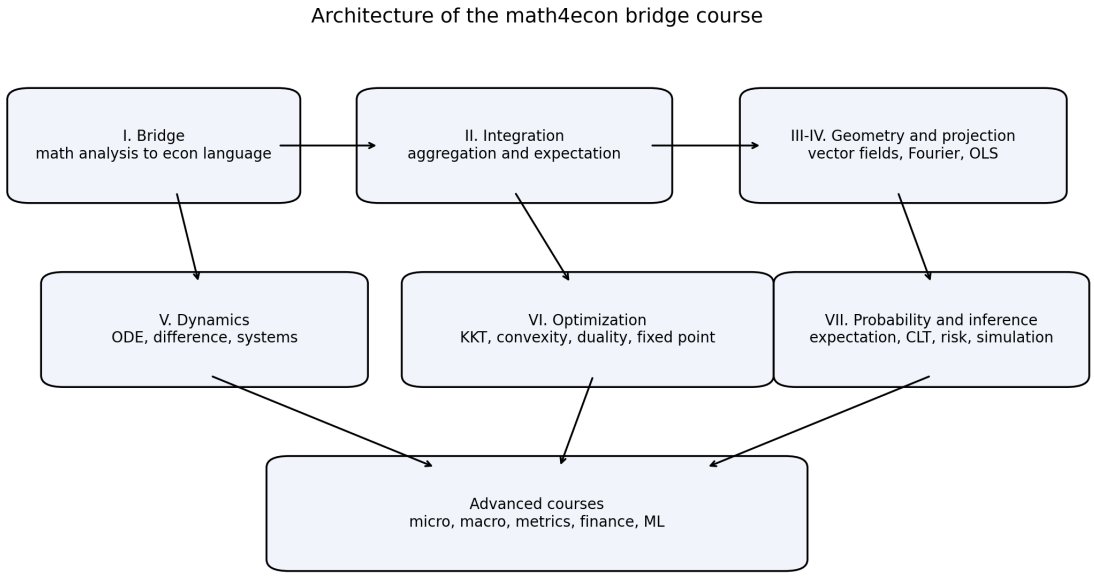



图 1：整套讲义的结构：从数学分析语言到高级经济金融课程。

## 1 这套讲义到这里是否结束

按正文结构，本章就是最后一章。原因有三点。

第一，原来的路线图最后一章就是“通向更高级课程的地图”。我们实际写作时把动态优化和数值方法扩展得更充分，所以正文编号从原计划的 26 章顺延到了 31 章，但功能上，本章仍

然承担最后的收束任务。

第二，核心数学闭环已经完成。你已经拥有以下七个模块：

1. 积分与期望：从二重积分到随机变量期望；
2. 几何与投影：从梯度、曲线积分到 OLS 与 Fourier；
3. 动态系统：ODE、差分方程、线性与非线性系统；
4. 静态优化：Hessian、Lagrange、KKT、凸性；
5. 均衡与对偶：隐函数、包络、对偶、固定点；
6. 动态优化：Euler、Bellman、最优控制、HJB；
7. 概率推断：条件期望、极限定理、多元正态、模拟推断。

第三，再往后继续写“新章”，很容易变成另一本教材：比如高级计量、随机过程、随机微积分、测度论、泛函分析、博弈论、机制设计、机器学习理论。这些当然都重要，但它们已经不是“经济金融数学分析补课讲义”的正文，而是后续专题课。

**方法模板**

所以接下来最合理的不是 Chapter 32，而是四个附录型交付：

1. **Formula Handbook**: 全书公式大全与条件说明；
2. **Mixed Review Problem Book**: 跨章节综合题册与详解；
3. **Concept Map and Index**: 概念索引、符号索引、模型索引；
4. **Paper Reading Toolkit**: 读经济金融论文时如何识别数学结构。

## 2 全书知识结构回顾

这套讲义的根本目标不是复制公共高数，也不是复制数学系分析，而是把数学变成经济金融建模语言。可以用下面的对应关系来总结。

| 数学对象 | 经济金融语言 | 典型模型 |
| --- | --- | --- |
| 导数与梯度 | 边际效应、局部变化 | 边际效用、边际成本、score function |
| 积分 | 聚合、总量、期望 | 总福利、期望效用、风险溢价 |
| Jacobian | 换变量、密度变换 | log return、联合分布变换、多元正态标准化 |
| 投影 | 最小二乘、最佳预测 | OLS、因子模型、条件期望近似 |
| ODE/差分方程 | 动态路径、收敛、稳定 | Solow、债务动态、AR(1)、蛛网模型 |

| | | |
|---|---|---|
| 矩阵与特征值 | 多变量动态方向 | Ramsey saddle path、VAR、状态空间模型 |
| Hessian | 曲率与二阶条件 | 利润最大化、均值方差优化 |
| Lagrange/KKT | 约束选择、影子价格 | 消费者、成本最小化、借贷约束 |
| 凸性 | 全局最优、唯一性 | 偏好凸性、生产集、风险分散 |
| 隐函数/包络 | 比较静态、价值变化 | 需求、供给、福利、成本函数 |
| 固定点 | 均衡存在 | 一般均衡、Nash 均衡、动态规划 |
| Bellman/HJB | 动态最优选择 | 消费储蓄、资产定价、投资 q 理论 |
| 极限定理 | 渐近推断 | OLS 一致性、CLT、GMM、t 检验 |
| 多元正态/二次型 | 风险与检验统计量 | portfolio variance、VaR、Wald/F test |
| Monte Carlo/Bootstrap | 模拟推断与计算实验 | 风险模拟、置信区间、政策仿真 |

**常见误区**

学完这些章节后，最危险的误区是“我都看过了，所以我会了”。数学工具真正内化的标志不是能复述定义，而是能在新模型里识别结构。例如看到

$$V(k) = \max_{c} \{u(c) + \beta V(f(k) - c)\}$$

你应当马上想到 Bellman 方程、压缩映射、策略函数、Euler 方程、数值迭代和 envelope condition。

## 3 通向高级微观经济学的地图

高级微观的数学底座主要是优化、凸分析、对偶、固定点、比较静态和博弈论。本讲义已经覆盖了最关键的前置部分。

### 3.1 消费者理论

消费者理论需要你熟练掌握：

- 偏好、效用函数、quasi-concavity；
- 预算集的凸性与紧性；
- Lagrange 与 KKT；

- indirect utility 和 expenditure function；
- Roy identity 与 Shephard lemma；
- Slutsky decomposition；
- welfare measures: CV, EV, consumer surplus。

  最核心的数学结构是

$$\max_x u(x) \quad \text{s.t.} \quad p \cdot x \le w,$$

以及它的对偶问题

$$\min_x p \cdot x \quad \text{s.t.} \quad u(x) \ge \bar{u}.$$

**方法模板**

读高级微观消费者理论时，每看到一个新命题，先问三件事：

1. 这是 primal 问题还是 dual 问题?
2. 它用的是 FOC、KKT、凸性，还是包络定理?
3. 它是在比较选择量，还是比较价值函数?

## 3.2 生产者理论

生产者理论需要：

- 凸生产集与生产函数；
- 成本最小化和利润最大化；
- conditional factor demand；
- cost function 的性质；
- profit function 的性质；
- Shephard lemma 与 Hotelling lemma。

  你可以把生产者理论压缩成两个优化问题：

$$C(w,y) = \min_x w \cdot x \quad \text{s.t.} \quad f(x) \ge y,$$

$$\pi(p,w) = \max_{y,x} py - w \cdot x \quad \text{s.t.} \quad y \le f(x).$$

高级微观大量结果都来自这两个 value functions 的包络性质。

## 3.3 一般均衡与博弈论

一般均衡和博弈论的核心数学是固定点。

- Walrasian equilibrium: excess demand 或 price correspondence 的 fixed point；

- Nash equilibrium: best response correspondence 的 fixed point;
- mixed strategy equilibrium: compact simplex 上的 Kakutani fixed point;
- matching and mechanisms: correspondence、upper hemicontinuity、convex values。

你暂时不需要把拓扑证明学到数学系水平, 但要知道定理条件为什么像这样出现: compact 负责“不跑到无穷远”，convex 负责“平均仍可行”，upper hemicontinuity 负责“极限不跳出图像”。

# 4 通向高级宏观经济学的地图

高级宏观的核心不是静态均衡, 而是动态最优选择和均衡路径。因此它需要动态系统、差分方程、最优控制、Bellman 方程、HJB、线性化和数值方法。

## 4.1 增长模型

Solow 模型是动态系统入门:

$$\dot{k} = sf(k) - (n + \delta)k.$$

Ramsey 模型则是动态优化:

$$\max_{\{c_t\}} \sum_{t=0}^{\infty} \beta^t u(c_t) \quad \text{s.t.} \quad k_{t+1} = f(k_t) + (1 - \delta)k_t - c_t.$$

它对应的数学工具包括:

- Euler equation;
- Bellman equation;
- saddle path stability;
- linearization around steady state;
- value function iteration。

## 4.2 新凯恩斯模型与线性系统

新凯恩斯基本结构通常可以抽象成线性预期系统:

$$AE_t z_{t+1} = Bz_t + C\varepsilon_t.$$

你需要的数学能力是:

- 矩阵代数;
- 特征值与稳定根;

- forward-looking variables 与 predetermined variables 的区分；
- saddle path 与唯一理性预期均衡；
- 数值求解线性模型。

### 4.3 异质性主体宏观

Heterogeneous agent macro 需要：概率、Markov chain、动态规划、数值方法和分布演化。最常见对象是：

$$V(a,z)=\max_{c,a'}\{u(c)+\beta E[V(a',z')\mid z]\},$$

以及分布 law of motion

$$\mu_{t+1}=T(\mu_t,g_t).$$

这里 $V$ 是个体问题，$\mu$ 是总体分布，$g_t$ 是策略函数。你需要同时理解微观最优化和宏观聚合。

**Advanced Core**

如果你以后读 Aiyagari、Huggett、HANK、search and matching 模型，最重要的数学基础不是多算积分，而是：state space、policy function、stationary distribution、fixed point、simulation、Euler error。

## 5 通向计量经济学理论的地图

计量理论的底层语言是概率极限、条件期望、投影、矩条件和渐近正态。

### 5.1 OLS 与投影

Population OLS 的本质是投影：

$$\beta=\arg\min_b E[(Y-X'b)^2].$$

FOC 是正交条件：

$$E[X(Y-X'\beta)]=0.$$

所以

$$\beta=(E[XX'])^{-1}E[XY].$$

这说明回归不是魔法，而是 Hilbert space 中的最小平方近似。

### 5.2 一致性与渐近正态

计量推断一般有三步：

1. 用 LLN 证明样本目标函数或矩条件收敛到总体对象；

2. 用识别条件说明总体对象唯一；
3. 用 CLT、Taylor expansion、Slutsky 得到渐近正态。

   典型结构是：

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, V).$$

然后用估计的 $\hat{V}$ 构造 standard errors、confidence intervals 和 tests。

### 5.3 MLE, GMM, causal inference

| 方向 | 核心数学对象 | 你应复习的章节 |
|---|---|---|
| MLE | likelihood、score、Hessian、信息矩阵 | Ch4, Ch14, Ch28, Ch29 |
| GMM | moment condition、weighting matrix、quadratic form | Ch8, Ch27, Ch28, Ch29 |
| Causal inference | conditional expectation、selection bias、projection | Ch27, Ch28, Ch30 |
| Panel data | projection、fixed effects、LLN/CLT | Ch8, Ch27, Ch28 |
| Time series | difference equation、stationarity、AR(1)、CLT | Ch11, Ch27, Ch28, Ch29 |

**常见误区**

计量里最常见的学习错误是只背命令和公式。真正应该问的是：估计量在估计哪个 population object？识别条件是什么？样本对象为什么收敛？标准误来自哪个渐近分布？

## 6 通向金融数学与资产定价的地图

金融数学有三条主线：组合选择、资产定价和风险管理。

### 6.1 组合选择

组合选择从均值方差开始：

$$\max_{w} \; w'\mu - \frac{\gamma}{2} w'\Sigma w \quad \text{s.t.} \quad \mathbf{1}'w = 1.$$

它需要：

- quadratic form；
- positive definite covariance matrix；

- Lagrange 与 KKT;
- eigenvalue decomposition;
- PCA and factor models。

## 6.2 资产定价

资产定价最核心的 Euler equation 是

$$1 = E_t[M_{t+1}R_{t+1}],$$

其中 $M_{t+1}$ 是 stochastic discount factor。它背后需要:

- 条件期望;
- 动态优化;
- 资产价格递推;
- 随机过程;
- GMM estimation。

## 6.3 风险管理

风险管理常见对象包括:

$$\sigma_p^2 = w'\Sigma w,$$

$$\mathrm{VaR}_\alpha, \qquad ES_\alpha = E[L \mid L \geq \mathrm{VaR}_\alpha].$$

你需要理解它们不是孤立公式，而是概率分布、分位数、尾部期望和模拟方法的组合。

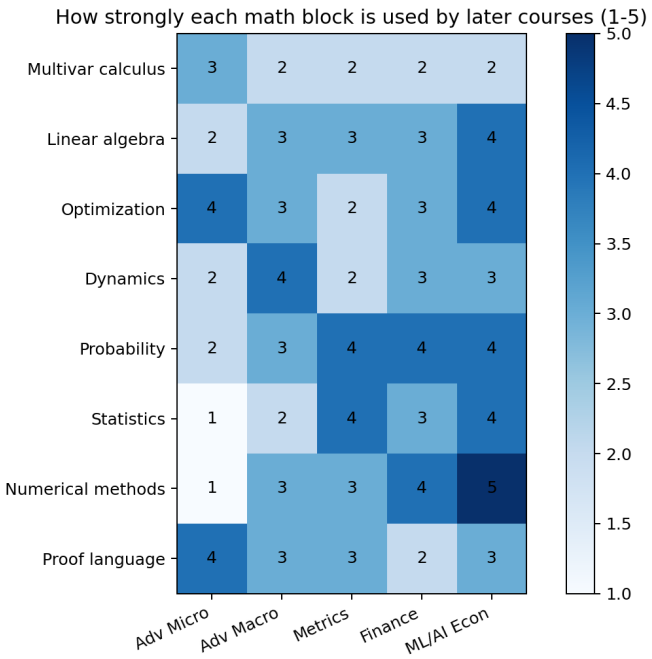



图 2: 不同高级课程对各数学模块的依赖强度。数字越大表示越重要。

# 7 通向机器学习与 AI 经济学的地图

AI economics、机器学习因果推断、算法经济学和 computational social science 需要的数学其实并不完全等于传统经济数学。你要特别补强四类能力。

## 7.1 优化与数值计算

机器学习训练的基本形式是经验风险最小化：

$$\hat{f} = \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \ell(y_i, f(x_i)) + \lambda \Omega(f).$$

这需要你理解：

- gradient descent；
- convex vs nonconvex optimization；
- regularization；
- cross-validation；
- numerical stability。

### 7.2 概率与泛化

泛化问题的底层语言是：训练误差和总体误差之间的差距。

$$R(f) = E[\ell(Y, f(X))], \qquad \hat{R}_n(f) = \frac{1}{n}\sum_i \ell(Y_i, f(X_i)).$$

这和 Ch28 的 LLN、CLT 以及 uniform convergence 有直接关系。

### 7.3 因果机器学习

因果机器学习常见目标是 conditional average treatment effect：

$$\tau(x) = E[Y(1) - Y(0) \mid X = x].$$

它需要：

- 条件期望；
- orthogonal moments；
- sample splitting；
- nuisance functions；
- asymptotic normality。

这些并不是单纯的 coding 技巧，而是概率、投影和极限定理的组合。

## 8 推荐教材与使用方式

下面的书不是让你全读完，而是按用途查阅。

| 用途 | 推荐材料 | 怎么用 |
|---|---|---|
| 经济数学总复习 | Simon and Blume; Sydsaeter and Hammond | 查公式、补标准经济例题，不必按页全读。 |
| 研究生数学经济 | de la Fuente; Sundaram; Ok | 用来补凸性、优化、动态系统、固定点和证明风格。 |
| 微积分/向量分析 | Stewart; Apostol Calculus; Marsden and Tromba | 只查多重积分、Jacobian、向量分析难点。 |
| 实分析严谨性 | Tao Analysis; Rudin; Abbott | 用来补极限、紧性、连续性、测度前置，不用一开始硬读 Rudin。 |

| | | |
|---|---|---|
| 计量理论 | Hayashi; Wooldridge; Newey and McFadden 章节 | 重点配合 Ch27-Ch29 看一致性、渐近正态、GMM。 |
| 高级宏观 | Ljungqvist and Sargent; Stokey Lucas Prescott; Adda and Cooper | 重点配合 Ch21-Ch25 看动态规划、递归均衡、数值方法。 |
| 金融数学 | Cochrane; Campbell Lo MacKinlay; Shreve 入门部分 | 先读离散时间资产定价和风险，再进入连续时间。 |
| 机器学习 | Hastie Tibshirani Friedman; Murphy; Chernozhukov DML papers | 从经验风险、正则化、orthogonal moments 读起。 |

**方法模板**

使用教材的原则：不要“从第一页开始虔诚阅读”。正确方式是带着模型找工具。例如你读到 Ramsey 模型不会 saddle path，就回 Ch12、Ch21、Ch23；读到 GMM 看不懂 weighting matrix，就回 Ch8、Ch27、Ch29；读到 VaR 或 factor model，就回 Ch29、Ch30。

## 9 六个月复习与进阶计划

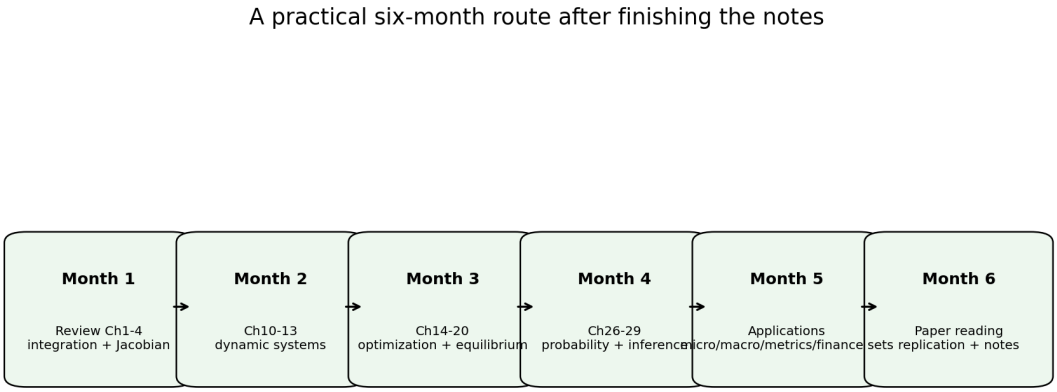



图 3: 完成本讲义后的六个月路线。

## 9.1 第一个月：把基础补成肌肉记忆

重点复习 Ch1-Ch4、Ch8-Ch9、Ch26-Ch27。

- 每天做 2 道多重积分或变量变换；
- 每天做 1 道条件期望或投影题；
- 每周总结一次“积分 = aggregation / expectation”的模型对应。

## 9.2 第二个月：动态系统

重点复习 Ch10-Ch13、Ch21-Ch24。

- 每天画一个 phase diagram；
- 每周解一个 Bellman 或 Euler 方程例子；
- 用 Python 做一次 value iteration。

## 9.3 第三个月：优化与均衡

重点复习 Ch14-Ch20。

- 每天做一个 KKT 或 Lagrange 问题；
- 每周做一个对偶性/包络定理例子；
- 阅读一个高级微观命题，把它拆成“优化 + 定理条件”。

## 9.4 第四个月：概率、计量和金融风险

重点复习 Ch26-Ch30。

- 每天做一个 LLN/CLT/Delta method 推导；
- 每周做一个 OLS 渐近推断题；
- 每周做一个 portfolio variance 或 VaR simulation。

## 9.5 第五个月：模型综合应用

选择四个综合模型：

1. 消费者问题与对偶；
2. Ramsey 模型；
3. OLS/GMM 推断；
4. 均值方差投资组合。

每个模型写一份 3-5 页推导笔记。

### 9.6 第六个月：论文阅读与复现

选择 6 篇论文，每周一篇，要求输出：

- 模型结构；
- 数学工具清单；
- 关键方程推导；
- 识别或均衡条件；
- 你自己的问题和延伸。

## 10 读论文时如何识别数学结构

读论文时不要被符号吓住。很多公式其实属于有限几类结构。

| 你看到的东西 | 本质结构 | 回去复习 |
|---|---|---|
| $\max_x u(x)$ with constraints | 静态优化 | Ch14-Ch17 |
| $F(x,\theta)=0$ | 隐函数/比较静态 | Ch18 |
| $V(\theta)=\max_x f(x,\theta)$ | 包络定理 | Ch18-Ch19 |
| $T(x)=x$ | 固定点/均衡存在 | Ch20 |
| $x_{t+1}=F(x_t)$ | 差分动态 | Ch11-Ch13 |
| $V(s)=\max_a\{u+\beta EV(s')\}$ | Bellman 方程 | Ch22 |
| $\mathcal{H}=u+\lambda f$ | 最优控制 | Ch23 |
| $E[m(Z,\theta)]=0$ | 矩条件/GMM | Ch27-Ch29 |
| $\sqrt{n}(\hat{\theta}-\theta_0)$ | 渐近正态 | Ch28 |
| $w'\Sigma w$ | 二次型/风险 | Ch29 |
| Simulation / bootstrap | 模拟推断 | Ch30 |

**Worked Example**

**例 1：看到一个高级微观命题**

命题说：若效用函数严格拟凹、预算集凸且紧，则消费者需求唯一，并且 indirect utility 对收入的导数等于 Lagrange multiplier。

拆解：

1. 严格拟凹 $\Rightarrow$ 最优选择唯一；
2. 预算集凸且紧 $\Rightarrow$ 最大值存在；

3. Lagrange FOC $\Rightarrow$ 边际替代率等于价格比；
4. Envelope theorem $\Rightarrow \partial v/\partial w = \lambda$。

所以这个命题不是神秘结论，而是 Ch15、Ch17、Ch18、Ch19 的组合。

**Worked Example**

**例 2：看到一个计量定理**

定理说：若 $E[X\varepsilon] = 0$ 且 $E[XX']$ 非奇异，则 OLS 一致；在同方差条件下，

$$\sqrt{n}(\hat{\beta} - \beta) \Rightarrow N(0, \sigma^2 Q^{-1}).$$

拆解：

1. $E[X\varepsilon] = 0$ 是 population orthogonality；
2. $E[XX']$ 非奇异是 identification；
3. LLN 给出样本矩阵收敛；
4. CLT 给出 $n^{-1/2}\sum X_i\varepsilon_i$ 的正态极限；
5. Slutsky 把二者合并。

所以这个定理是 Ch27 和 Ch28 的直接应用。

## 11 个人技能目标与检查表

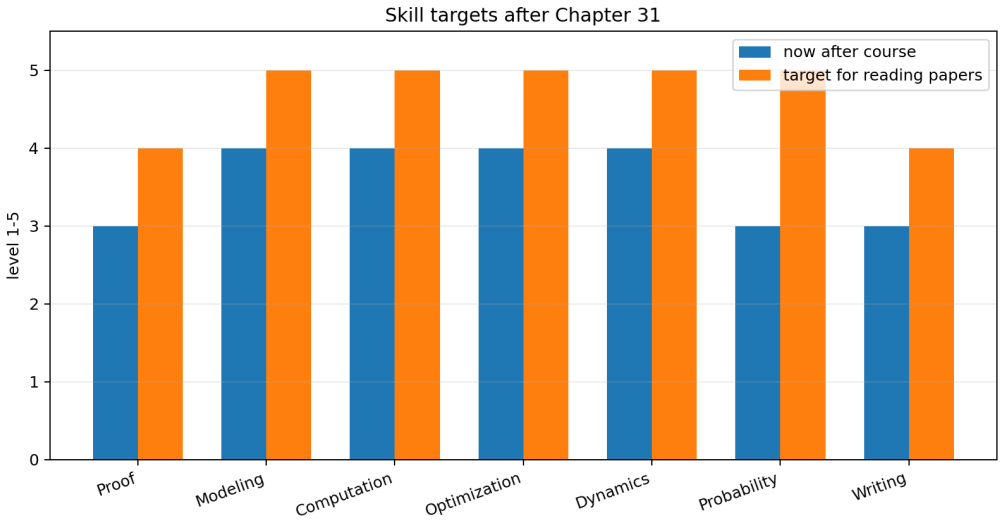



图 4：完成本讲义后的技能目标。

### 11.1 最低合格线

你应该能做到：

1. 看懂一个多元优化问题并写出 FOC/KKT；
2. 看懂一个动态方程并判断稳态稳定性；
3. 看懂一个 Bellman 方程并解释状态、控制、价值函数；
4. 看懂一个 OLS/GMM 推导中的正交条件；
5. 看懂 LLN/CLT/Slutsky 在推断中的作用；
6. 看懂 portfolio variance 与 covariance matrix；
7. 能用 Python 做基本模拟和数值迭代；
8. 能把论文中的公式映射回本讲义章节。

### 11.2 优秀线

如果要进一步冲击研究型学习，你应当做到：

1. 对每个定理知道条件为什么必要；
2. 能把一个模型从文字设定推到核心方程；
3. 能自己检查二阶条件、稳定性和识别条件；
4. 能把理论模型转成数值算法；
5. 能复现一篇论文的关键图表或主回归；
6. 能写出一页“模型数学结构说明”。

## 12 综合练习题

本章练习不再考单一章节技巧，而是训练你“看到问题选工具”。

### A 组：识别数学结构

对下列对象，判断它属于哪类数学结构，并指出应复习哪些章节。

A1. $\max_x u(x)$ subject to $p \cdot x \le w$。

A2. $x_{t+1} = 0.8x_t + \varepsilon_t$。

A3. $\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, V)$。

A4. $p_t = \sum_{j=1}^{\infty} \beta^j E_t d_{t+j}$。

A5. $E[Z_i(Y_i - X_i'\beta)] = 0$。

A6. $w'\Sigma w$。

A7. $V(k) = \max_c \{u(c) + \beta V(f(k) - c)\}$。

A8. $F(x, \theta) = 0$ and $\mathrm{d}x/\mathrm{d}\theta = -F_x^{-1} F_\theta$。

### B 组：课程路线设计

B1. 如果你的目标是高级微观，列出最优先复习的 8 章并说明原因。

B2. 如果你的目标是高级宏观，列出最优先复习的 8 章并说明原因。

B3. 如果你的目标是计量理论，列出最优先复习的 8 章并说明原因。

B4. 如果你的目标是量化金融，列出最优先复习的 8 章并说明原因。

B5. 设计一个 4 周复习计划，每周至少覆盖两个数学模块。

### C 组：综合推导题

C1. 用 Lagrange 方法推导 Cobb-Douglas 消费者需求，并说明乘子的经济含义。

C2. 从 population OLS 最小化问题推导正交条件和闭式解。

C3. 对 $x_{t+1} = ax_t + b$，推导稳态和稳定条件，并解释它如何用于债务动态。

C4. 对 $\dot{x} = f(x)$，说明如何用线性化判断局部稳定性。

C5. 对 $w'\Sigma w$，说明为什么 $\Sigma$ 必须半正定，并解释 diversification。

### D 组：论文阅读应用题

D1. 读到一个模型使用 Bellman 方程，写出你应检查的 5 个要素。

D2. 读到一个论文说“we estimate by GMM”，写出你应检查的 5 个要素。

D3. 读到一个论文说“the equilibrium exists by Kakutani”，写出你应检查的定理条件。

D4. 读到一个金融论文使用 PCA，说明它在数学上做了什么。

D5. 读到一个政策评估论文使用 bootstrap standard errors，说明 bootstrap 试图近似什么。

### E 组：自我诊断

E1. 列出你最不熟的 5 个章节，并给出具体补救题型。

E2. 选择一个经济模型，写出它使用了本讲义的哪些数学工具。

E3. 选择一篇论文，做一页“数学结构拆解”。

E4. 解释为什么本讲义不是公共高数，也不是数学系分析。

# 13 综合练习题参考答案

## A 组答案

A1. 约束优化、Lagrange/KKT、消费者理论。复习 Ch15-Ch19。

A2. 一阶差分方程、AR(1)、随机过程入口。复习 Ch11、Ch27-Ch29。

A3. 渐近正态。复习 Ch28，若涉及 Wald/t/F 统计量再复习 Ch29。

A4. 前瞻资产价格递推、条件期望、折现现值。复习 Ch11、Ch21、Ch27。

A5. 矩条件和 GMM。复习 Ch8、Ch27、Ch28、Ch29。

A6. 二次型和投资组合风险。复习 Ch14、Ch29。

A7. Bellman 方程和动态规划。复习 Ch22，也连接 Ch21、Ch25。

A8. 隐函数定理与比较静态。复习 Ch18。

## B 组答案

B1. 高级微观优先：Ch14 Hessian, Ch15 Lagrange, Ch16 KKT, Ch17 convexity, Ch18 IFT/envelope, Ch19 duality, Ch20 fixed point, Ch27 conditional expectation。理由：高级微观核心是选择、对偶、均衡和福利。

B2. 高级宏观优先：Ch10 ODE, Ch11 difference equations, Ch12 linear systems, Ch13 nonlinear dynamics, Ch21 Euler, Ch22 Bellman, Ch23 optimal control, Ch25 numerical methods。理由：宏观核心是动态路径、均衡稳定和计算。

B3. 计量理论优先：Ch8 projection, Ch14 Hessian, Ch26 probability, Ch27 conditional expectation/OLS, Ch28 asymptotics, Ch29 quadratic forms, Ch30 bootstrap/Monte Carlo, Ch18 identification/comparative statics。理由：计量依赖 population objects、识别、极限定理和检验统计量。

B4. 量化金融优先：Ch3 Jacobian, Ch11 asset recursion, Ch14 optimization, Ch16 constraints, Ch21 Euler, Ch26 probability, Ch29 multivariate normal/risk, Ch30 simulation。理由：金融依赖收益变换、组合优化、资产定价、风险和模拟。

B5. 一个合理 4 周计划：第一周 Ch1-Ch4 + Ch26，补积分和期望；第二周 Ch10-Ch13，补动态；第三周 Ch14-Ch20，补优化和均衡；第四周 Ch27-Ch30，补计量金融推断。每天必须做题，而不是只读。

## C 组答案

C1. 设 $u(x,y)=x^{\alpha}y^{1-\alpha}$，预算 $p_x x+p_y y=w$。Lagrangian:

$$\mathcal{L}=x^{\alpha}y^{1-\alpha}+\lambda(w-p_x x-p_y y).$$

FOC:

$$\alpha x^{\alpha-1}y^{1-\alpha}=\lambda p_x,\qquad (1-\alpha)x^{\alpha}y^{-\alpha}=\lambda p_y.$$

两式相除:

$$\frac{\alpha}{1-\alpha}\frac{y}{x}=\frac{p_x}{p_y}.$$

结合预算可得

$$x^*=\frac{\alpha w}{p_x},\qquad y^*=\frac{(1-\alpha)w}{p_y}.$$

乘子 $\lambda$ 是收入的边际效用，即 budget 放松一单位带来的 value 增加。

C2. Population OLS:

$$\min_b E[(Y-X'b)^2].$$

对 $b$ 求导:

$$-2E[X(Y-X'b)]=0.$$

因此

$$E[XY]=E[XX']b.$$

若 $E[XX']$ 可逆,

$$\beta=(E[XX'])^{-1}E[XY].$$

正交条件是 $E[X\varepsilon]=0$。

C3. 稳态满足 $x^*=ax^*+b$，所以 $x^*=b/(1-a)$。偏离满足

$$x_t-x^*=a^t(x_0-x^*).$$

稳定条件是 $|a|<1$。债务动态中 $a=(1+r)/(1+g)$，所以正常情形下稳定要求 $r<g$，或者财政反应使有效系数低于 1。

C4. 若 $\dot{x}=f(x)$ 且 $f(x^*)=0$，在稳态附近

$$\dot{x}\approx f'(x^*)(x-x^*).$$

若 $f'(x^*)<0$，偏离量指数衰减，稳态局部稳定；若 $f'(x^*)>0$，偏离量增长，稳态不稳定。

C5. 对任意权重 $w$，portfolio variance 是

$$\mathrm{Var}(w'R)=w'\Sigma w.$$

方差不能为负，所以必须有 $w'\Sigma w\geq 0$，即 $\Sigma$ 半正定。Diversification 来自协方差项：若资产不完全正相关，组合方差可低于单个资产风险的加权平均。

### D 组答案

D1. Bellman 方程检查：state variable 是什么？control variable 是什么？transition law 是什么？discount factor 是否小于 1？value function 是否满足 contraction 或 concavity？

D2. GMM 检查：moment condition 是什么？为什么等于零？参数是否识别？weighting matrix 如何选择？standard errors 是否稳健？

D3. Kakutani 检查：定义域是否 compact convex？correspondence 是否非空？取值是否凸？graph 是否 closed 或 upper hemicontinuous？fixed point 是否对应均衡？

D4. PCA 是对协方差矩阵做特征值分解，找到解释方差最大的正交方向。金融里它常用于提取共同风险因子。

D5. Bootstrap 试图用样本分布近似总体分布，从而近似估计量的抽样分布，进而构造标准误、置信区间或检验。

### E 组参考

E 组没有唯一答案，但合格答案应满足：

- 不说空话，要具体到章节、公式、题型；
- 模型拆解要包括目标函数、约束、状态变量、随机变量、均衡条件；
- 论文拆解要能指出用了哪些数学结构，而不是只概括主题；
- 对本讲义定位的解释应抓住“经济金融导向的数学语言”这一点。

## 14 最后的收束：真正的下一步

正文到这里结束，但学习不应该停止。最值得继续做的不是开新章，而是把这 31 章压缩成三种可用版本。

### 14.1 版本一：公式大全

按模块整理：

1. 积分与变量替换公式；
2. ODE/差分方程解法与稳定条件；
3. 优化 FOC/SOC/KKT/duality/envelope；
4. Bellman/Hamiltonian/HJB；
5. 概率、条件期望、LLN/CLT、Delta method；
6. 多元正态、二次型、金融风险；
7. Monte Carlo/Bootstrap。

公式大全不能只写公式，还要写“使用条件”和“经济含义”。

## 14.2 版本二：综合题册

题册应该每题跨越至少两个模块。例如：

- Ramsey 模型：动态优化 + 相图 + 稳定性；
- 消费者理论：KKT + 对偶 + 包络；
- OLS 推断：投影 + LLN + CLT + 二次型；
- 投资组合：Hessian + KKT + 多元正态 + 模拟风险。

## 14.3 版本三：论文阅读工具包

每篇论文用同一模板：

1. 研究问题；
2. 模型对象；
3. 数学结构；
4. 关键方程；
5. 识别或均衡条件；
6. 推导中最难的一步；
7. 可复现部分；
8. 你自己的批注。

**直觉盒子**

这套课真正完成的标志，不是“PDF 都生成了”，而是你拿到一篇经济金融论文时，不再被公式吓住，而是能说：这里是优化，这里是投影，这里是固定点，这里是极限定理，这里是数值算法。

**Advanced Core**

**最终结论**

本讲义正文可以在 Chapter 31 结束。后续最值得做的是：

$$\text{Formula Handbook} \quad + \quad \text{Mixed Review Problem Book} \quad + \quad \text{Paper Reading Toolkit}.$$

如果之后再扩展，建议另起专题，而不是继续 Chapter 32：

- 专题 A：测度论与严谨概率；
- 专题 B：随机过程与随机微积分；

- 专题 C：高级计量理论；
- 专题 D：高级宏观数值方法；
- 专题 E：金融工程与资产定价；
- 专题 F：机器学习与因果推断。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Appendix A
# 全书公式大全与速查手册

Formula Compendium for Economics and Finance

这是 Ch1–Ch31 的公式收束版。它不是新章节，而是一个复习和查阅工具：把多元积分、向量分析、投影、动态系统、优化、对偶、固定点、动态优化、概率、回归、极限定理、金融风险与模拟方法的核心公式按主题重新整理。每个公式都尽量给出三件事：条件、解释、常见用途。

math4econ personalized notes
2026

# 目录

**如何使用这份公式大全** **2**

**1 多元积分与概率化思维** **2**
1.1 二重积分、Fubini 与区域变换 . . . 2
1.2 三重积分与多维分布 . . . 3
1.3 变量替换与 Jacobian . . . 4
1.4 Leibniz Rule 与积分下求导 . . . 4

**2 向量分析、微分与路径** **5**
2.1 梯度、方向导数与全微分 . . . 5
2.2 曲线积分、保守场与路径无关 . . . 6
2.3 Green、Gauss、Stokes 的局部-整体公式 . . . 6

**3 投影、OLS 与 Fourier** **6**
3.1 线性投影与 OLS . . . 7
3.2 Frisch-Waugh-Lovell 定理 . . . 7
3.3 Fourier 级数 . . . 8

**4 动态系统：ODE、差分方程与相图** **8**
4.1 一阶 ODE . . . 8
4.2 经济动态方程速查 . . . 9
4.3 一阶差分方程 . . . 9
4.4 线性系统与特征值 . . . 10
4.5 非线性系统局部线性化 . . . 10

**5 静态优化、KKT 与凸性** **11**
5.1 无约束优化与 Hessian . . . 11
5.2 等式约束与 Lagrange 乘子 . . . 12
5.3 KKT 条件 . . . 12
5.4 凸性与全局最优 . . . 13

**6 比较静态、包络、对偶与固定点** **13**
6.1 隐函数定理与比较静态 . . . 13
6.2 包络定理 . . . 14
6.3 经济对偶公式 . . . 14
6.4 固定点与均衡存在 . . . 15

**7 动态优化、Bellman、最优控制与 HJB** **15**
7.1 离散时间 Euler 方程 . . . 15
7.2 Bellman 方程 . . . 16

7.3 连续时间最优控制 . . . . . 16
7.4 HJB 方程 . . . . . 17

**8 概率、条件期望与回归 17**
8.1 概率基础 . . . . . 17
8.2 条件期望 . . . . . 18
8.3 OLS 与识别公式 . . . . . 18

**9 收敛、极限定理与渐近推断 19**
9.1 收敛概念 . . . . . 19
9.2 LLN、CLT、Slutsky、Delta . . . . . 19

**10 多元正态、二次型与金融风险 20**
10.1 多元正态与协方差矩阵 . . . . . 20
10.2 二次型检验统计量 . . . . . 20
10.3 投资组合风险 . . . . . 21

**11 Monte Carlo、Bootstrap 与模拟推断 21**
11.1 Monte Carlo 估计 . . . . . 21
11.2 Bootstrap . . . . . 22

**12 模型识别：看到公式该想到什么 22**

**13 常见公式误用清单 23**

**14 一页终极复习图谱 24**

# 如何使用这份公式大全

**速查直觉**

公式不是用来背孤立符号的。经济金融里的公式通常回答四类问题：

1. **累加或平均：**用积分和期望把异质性个体、状态、风险加总。
2. **边际和最优：**用导数、梯度、KKT、包络定理描述选择。
3. **动态和均衡：**用差分方程、ODE、Bellman、HJB 描述时间路径。
4. **不确定性和推断：**用条件期望、极限定理、协方差矩阵和模拟方法连接数据。

如果你看到一个经济模型，先判断它属于哪类问题，再找对应公式。

**使用模板**

**速查建议**

- 考试前：先看每节的 boxed formula，再看“常见误用”。
- 做题时：先识别模型类型，不要一上来套公式。
- 读论文时：把论文里的核心等式归类为优化 FOC、均衡条件、动态递推、投影矩条件或概率极限。
- 复习时：把这份公式大全和综合习题册配合使用。只看公式不会真正掌握，必须用题目测试公式的适用条件。

# 1 多元积分与概率化思维

## 1.1 二重积分、Fubini 与区域变换

**核心公式**

矩形区域上的二重积分：

$$\iint_{[a,b]\times[c,d]} f(x,y)\,\mathrm{d}A = \int_a^b \int_c^d f(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$

Type I 区域：

$$R = \{(x,y) : a \le x \le b,\ g_1(x) \le y \le g_2(x)\},$$

$$\iint_R f(x,y)\,\mathrm{d}A = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$

Type II 区域：

$$R = \{(x,y) : c \le y \le d,\ h_1(y) \le x \le h_2(y)\},$$

$$\iint_R f(x,y)\,\mathrm{d}A = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

| 公式 | 条件/解释 | 经济金融用途 |
|---|---|---|
| $\iint_R f\,\mathrm{d}A$ | 区域上的累加。 | 总福利、总需求、密度下的期望。 |
| $\bar{f} = \frac{1}{\mathrm{Area}(R)} \iint_R f\,\mathrm{d}A$ | 区域平均值。 | 平均税负、平均效用、平均风险暴露。 |
| $\int\int f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$ | 联合密度归一化。 | 联合风险因子、异质性分布。 |
| $f_X(x) = \int f_{X,Y}(x,y)\,\mathrm{d}y$ | 边际密度。 | 从联合分布中抽出单个变量分布。 |
| $\mathbb{E}[g(X,Y)] = \iint g(x,y) f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$ | 函数的期望。 | portfolio payoff、welfare、loss function。 |

**常见误用**

改积分次序时，不能只交换 $\mathrm{d}x\,\mathrm{d}y$。必须重新画区域并改上下限。经济应用中，如果你把积分区域改错，相当于改变了样本空间或人群范围。

## 1.2 三重积分与多维分布

**核心公式**

三维联合密度和边际化：

$$\iiint f_{X,Y,Z}(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = 1,$$

$$f_{X,Y}(x,y) = \int f_{X,Y,Z}(x,y,z)\,\mathrm{d}z,$$

$$\mathbb{E}[g(X,Y,Z)] = \iiint g(x,y,z) f_{X,Y,Z}(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z.$$

多维积分最常见的经济解释是 aggregation over heterogeneous states。比如家庭有收入 $y$、财富 $a$、生产率 $z$，社会平均消费可写成

$$\bar{c} = \iiint c(y,a,z) f(y,a,z)\,\mathrm{d}y\,\mathrm{d}a\,\mathrm{d}z.$$

## 1.3 变量替换与 Jacobian

**核心公式**

二维变量替换：若

$$(x, y) = T(u, v),$$

且变换局部一一、可微、Jacobian 非零，则

$$\iint_R f(x, y)\,\mathrm{d}x\,\mathrm{d}y = \iint_S f(x(u, v), y(u, v))\left|\frac{\partial(x, y)}{\partial(u, v)}\right|\,\mathrm{d}u\,\mathrm{d}v.$$

Jacobian：

$$\frac{\partial(x, y)}{\partial(u, v)} = \det\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}.$$

| 变换 | Jacobian | 用途 |
|---|---|---|
| 极坐标 $x = r\cos\theta, y = r\sin\theta$ | $\lvert J\rvert = r$ | 圆形、扇形、二维正态。 |
| 线性变换 $\boldsymbol{x} = A\boldsymbol{u}$ | $\lvert J\rvert = \lvert\det A\rvert$ | 多元正态、portfolio 因子变换。 |
| 一维随机变量 $Y = g(X)$ | $f_Y(y) = f_X(g^{-1}(y))\left\lvert\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right\rvert$ | lognormal、收益率变换。 |
| 多维随机变量 $\boldsymbol{Y} = T(\boldsymbol{X})$ | $f_Y(y) = f_X(T^{-1}(y))\lvert\det DT^{-1}(y)\rvert$ | 多因子风险、坐标标准化。 |

**常见误用**

随机变量变换中容易把 Jacobian 放反。若已知 $x = T^{-1}(y)$，密度公式用的是 inverse Jacobian。积分变量替换和密度变换本质一致，但方向不同。

## 1.4 Leibniz Rule 与积分下求导

**核心公式**

若上下限依赖参数：

$$I(\theta) = \int_{a(\theta)}^{b(\theta)} f(x, \theta)\,\mathrm{d}x,$$

则

$$\frac{\mathrm{d}I}{\mathrm{d}\theta} = f(b(\theta), \theta)b'(\theta) - f(a(\theta), \theta)a'(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial f}{\partial\theta}(x, \theta)\,\mathrm{d}x.$$

固定区间时：

$$\frac{\partial}{\partial\theta}\int f(x,\theta)\,\mathrm{d}x=\int\frac{\partial f}{\partial\theta}(x,\theta)\,\mathrm{d}x.$$

典型应用：

$$\frac{\partial}{\partial\theta}\mathbb{E}_\theta[m(X,\theta)]=\int\frac{\partial m}{\partial\theta}f_\theta(x)\,\mathrm{d}x+\int m(x,\theta)\frac{\partial f_\theta(x)}{\partial\theta}\,\mathrm{d}x.$$

若写成 score function $s_\theta(x)=\partial\log f_\theta(x)/\partial\theta$，则

$$\frac{\partial}{\partial\theta}\mathbb{E}_\theta[m(X)]=\mathbb{E}_\theta[m(X)s_\theta(X)].$$

# 2 向量分析、微分与路径

## 2.1 梯度、方向导数与全微分

**核心公式**

对 $f:\mathbb{R}^n\to\mathbb{R}$，梯度为

$$\nabla f(x)=\begin{pmatrix}\frac{\partial f}{\partial x_1}\\ \vdots\\ \frac{\partial f}{\partial x_n}\end{pmatrix}.$$

方向导数：

$$D_uf(x)=\nabla f(x)'u,\qquad \|u\|=1.$$

全微分：

$$\mathrm{d}f=\sum_{i=1}^n f_i(x)\,\mathrm{d}x_i=\nabla f(x)'\,\mathrm{d}x.$$

二阶近似：

$$f(x+h)\approx f(x)+\nabla f(x)'h+\frac{1}{2}h'H(x)h.$$

经济解释：

- $\nabla u(x)$：边际效用向量；
- $\nabla C(q)$：边际成本向量；
- $\nabla f(x)'h$：沿政策变化 $h$ 的一阶福利变化；
- Hessian：曲率、替代性、风险厌恶、二阶条件。

### 2.2 曲线积分、保守场与路径无关

**核心公式**

标量场沿曲线的积分：

$$\int_C f(x,y)\,\mathrm{d}s = \int_a^b f(r(t))\|r'(t)\|\,\mathrm{d}t.$$

向量场做功型曲线积分：

$$\int_C \boldsymbol{F}\cdot\,\mathrm{d}\boldsymbol{r} = \int_a^b \boldsymbol{F}(r(t))\cdot r'(t)\,\mathrm{d}t.$$

若 $\boldsymbol{F}=\nabla\phi$，则

$$\int_C \boldsymbol{F}\cdot\,\mathrm{d}\boldsymbol{r} = \phi(B)-\phi(A),$$

只与端点有关，和路径无关。

二维保守场判别的常见充分条件：在单连通区域内，若

$$\boldsymbol{F}=(P,Q),\qquad \frac{\partial P}{\partial y}=\frac{\partial Q}{\partial x},$$

则 $\boldsymbol{F}$ 有势函数。

### 2.3 Green、Gauss、Stokes 的局部-整体公式

| 定理 | 公式 | 直觉 |
|---|---|---|
| Green | $\oint_{\partial R} P\,\mathrm{d}x+Q\,\mathrm{d}y=\iint_R (Q_x-P_y)\,\mathrm{d}A$ | 边界循环等于内部旋度总和。 |
| 二维散度 | $\oint_{\partial R}\boldsymbol{F}\cdot n\,\mathrm{d}s=\iint_R \nabla\cdot\boldsymbol{F}\,\mathrm{d}A$ | 边界净流出等于内部源汇总和。 |
| Gauss | $\iint_{\partial V}\boldsymbol{F}\cdot n\,\mathrm{d}S=\iiint_V \nabla\cdot\boldsymbol{F}\,\mathrm{d}V$ | 三维通量守恒。 |
| Stokes | $\oint_{\partial S}\boldsymbol{F}\cdot\,\mathrm{d}\boldsymbol{r}=\iint_S(\nabla\times\boldsymbol{F})\cdot n\,\mathrm{d}S$ | 边界循环等于曲面旋度通量。 |

经济金融中用得不如优化和概率频繁，但它们提供一个重要语言：local condition 如何 aggregate 成 global condition。

## 3 投影、OLS 与 Fourier

## 3.1 线性投影与 OLS

**核心公式**

有限维投影：将 $y$ 投影到 $X$ 的列空间，最小化

$$\min_b \|y - Xb\|^2.$$

一阶条件：

$$X'(y - X\hat{b}) = 0.$$

解：

$$\hat{b} = (X'X)^{-1}X'y.$$

投影矩阵和残差矩阵：

$$P_X = X(X'X)^{-1}X', \qquad M_X = I - P_X.$$

Population OLS：

$$\beta^* = \arg\min_b \mathbb{E}[(Y - X'b)^2],$$

FOC：

$$\mathbb{E}[X(Y - X'\beta^*)] = 0,$$

所以

$$\boxed{\beta^* = (\mathbb{E}[XX'])^{-1}\mathbb{E}[XY].}$$

**常见误用**

样本 OLS 的 $X'\hat{\varepsilon} = 0$ 是代数恒等式。因果识别需要的是总体条件，例如 $\mathbb{E}[\varepsilon \mid X] = 0$ 或 $\mathbb{E}[X\varepsilon] = 0$。不要把“回归残差与样本解释变量正交”误读成“模型没有内生性”。

## 3.2 Frisch-Waugh-Lovell 定理

对于回归

$$y = X_1\beta_1 + X_2\beta_2 + u,$$

$\beta_1$ 可以通过 partialling-out 得到：

$$\hat{\beta}_1 = (X_1'M_{X_2}X_1)^{-1}X_1'M_{X_2}y.$$

解释：先从 $y$ 和 $X_1$ 中剔除 $X_2$ 的线性影响，再用剩余部分回归。

### 3.3 Fourier 级数

**核心公式**

在 $[-\pi,\pi]$ 上：

$$f(x)\sim\frac{a_0}{2}+\sum_{n=1}^{\infty}(a_n\cos nx+b_n\sin nx),$$

$$a_0=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\,\mathrm{d}x,$$

$$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos nx\,\mathrm{d}x,\quad b_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin nx\,\mathrm{d}x.$$

Parseval identity：

$$\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)^2\,\mathrm{d}x=\frac{a_0^2}{2}+\sum_{n=1}^{\infty}(a_n^2+b_n^2).$$

奇偶性：偶函数只含余弦项，奇函数只含正弦项。经济直觉：低频捕捉趋势和周期，高频捕捉快速波动或噪声。

## 4 动态系统：ODE、差分方程与相图

### 4.1 一阶 ODE

**核心公式**

指数增长/衰减：

$$\dot{x}=ax\quad\Rightarrow\quad x(t)=x_0e^{at}.$$

均值回复：

$$\dot{x}=-a(x-\bar{x}),\quad a>0\quad\Rightarrow\quad x(t)=\bar{x}+(x_0-\bar{x})e^{-at}.$$

一阶线性 ODE：

$$\dot{x}+a(t)x=b(t),$$

积分因子：

$$\mu(t)=\exp\left(\int a(t)\,\mathrm{d}t\right),$$

$$x(t)=\frac{1}{\mu(t)}\left[C+\int\mu(t)b(t)\,\mathrm{d}t\right].$$

自治方程：

$$\dot{x}=f(x).$$

稳态：

$$f(x^*)=0.$$

局部稳定：

$$f'(x^*) < 0.$$

半衰期：若偏离为 $e^{-\lambda t}$，则

$$T_{1/2} = \frac{\ln 2}{\lambda}.$$

## 4.2 经济动态方程速查

| 模型 | 方程 | 稳态/条件 |
|---|---|---|
| Solow 连续时间 | $\dot{k} = sf(k) - (n+\delta)k$ | $sf(k^*) = (n+\delta)k^*$。 |
| 债务/GDP 连续时间 | $\dot{b} = (r-g)b + d$ | $r < g$ 时旧债务影响衰减。 |
| 价格调整 | $\dot{p} = \lambda[D(p) - S(p)]$ | 稳定条件 $D'(p^*) - S'(p^*) < 0$。 |
| 资本存量 | $\dot{K} = I - \delta K$ | $K^* = I/\delta$。 |
| 均值回复 | $\dot{x} = -a(x - \bar{x})$ | 速度 $a$，half-life $\ln 2/a$。 |

## 4.3 一阶差分方程

**核心公式**

一阶线性差分：

$$x_{t+1} = ax_t + b.$$

稳态：

$$x^* = \frac{b}{1-a}, \quad a \neq 1.$$

偏离形式：

$$x_t - x^* = a^t(x_0 - x^*).$$

稳定条件：

$$|a| < 1.$$

时间变化冲击：

$$x_{t+1} = ax_t + z_t \quad \Rightarrow \quad x_t = a^t x_0 + \sum_{j=0}^{t-1} a^{t-1-j} z_j.$$

蛛网模型：

$$D(p_t) = \alpha - \beta p_t, \quad S(p_{t-1}) = \gamma + \delta p_{t-1},$$

$$p_t = \frac{\alpha - \gamma}{\beta} - \frac{\delta}{\beta} p_{t-1},$$

稳定条件：

$$\left|\frac{\delta}{\beta}\right| < 1.$$

## 4.4 线性系统与特征值

**核心公式**

连续时间线性系统：

$$\dot{\boldsymbol{x}} = A\boldsymbol{x}.$$

若 $A$ 可对角化，$A = V\Lambda V^{-1}$，则

$$\boldsymbol{x}(t) = Ve^{\Lambda t}V^{-1}\boldsymbol{x}(0).$$

稳定条件：所有特征值实部为负，

$$\mathrm{Re}(\lambda_i) < 0.$$

离散时间线性系统：

$$\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t.$$

解：

$$\boldsymbol{x}_t = A^t\boldsymbol{x}_0.$$

稳定条件：所有特征值在单位圆内，

$$|\lambda_i| < 1.$$

二维系统类型：

| 类型 | 特征值条件 | 经济直觉 |
| --- | --- | --- |
| Stable node | 两个实特征值均负。 | 所有方向都收敛。 |
| Unstable node | 两个实特征值均正。 | 所有方向都发散。 |
| Saddle | 一个正，一个负。 | 只有稳定臂收敛，宏观 saddle path。 |
| Stable focus | 复根实部为负。 | 震荡收敛。 |
| Unstable focus | 复根实部为正。 | 震荡发散。 |
| Center | 复根实部为零。 | 闭合轨道或中性震荡。 |

## 4.5 非线性系统局部线性化

非线性系统：

$$\dot{\boldsymbol{x}} = F(\boldsymbol{x}).$$

稳态：

$$F(\boldsymbol{x}^*) = 0.$$

Jacobian：

$$J(\boldsymbol{x}^*) = DF(\boldsymbol{x}^*).$$

局部近似：

$$\dot{\boldsymbol{y}} = J(\boldsymbol{x}^*)\boldsymbol{y}, \qquad \boldsymbol{y} = \boldsymbol{x} - \boldsymbol{x}^*.$$

稳定性由 $J(\boldsymbol{x}^*)$ 的特征值决定。

# 5 静态优化、KKT 与凸性

## 5.1 无约束优化与 Hessian

**核心公式**

无约束问题：

$$\max_x f(x).$$

一阶条件：

$$\nabla f(x^*) = 0.$$

二阶条件：

- 若 $H_f(x^*)$ 负定，则严格局部最大；
- 若 $H_f(x^*)$ 正定，则严格局部最小；
- 若 Hessian 不定，则 saddle point。

二次型：

$$q(h) = h'Hh.$$

二元 Hessian 判别：

$$H = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}, \qquad D = f_{xx}f_{yy} - f_{xy}^2.$$

- $D > 0, f_{xx} < 0$：局部最大；
- $D > 0, f_{xx} > 0$：局部最小；
- $D < 0$：鞍点。

## 5.2 等式约束与 Lagrange 乘子

**核心公式**

等式约束问题：

$$\max_x f(x) \quad \text{s.t.} \quad g(x) = c.$$

Lagrangian：

$$\mathcal{L}(x, \lambda) = f(x) + \lambda[c - g(x)].$$

FOC：

$$\nabla_x f(x^*) = \lambda^* \nabla_x g(x^*), \qquad g(x^*) = c.$$

影子价格：

$$\frac{\partial V(c)}{\partial c} = \lambda^*.$$

消费者问题：

$$\max_x u(x) \quad \text{s.t.} \quad p'x = w.$$

FOC：

$$\nabla u(x^*) = \lambda p,$$

两商品时：

$$\frac{MU_1}{MU_2} = \frac{p_1}{p_2}.$$

## 5.3 KKT 条件

**核心公式**

最大化问题的一种常用写法：

$$\max_x f(x) \quad \text{s.t.} \quad g_i(x) \geq 0,\ h_j(x) = 0.$$

Lagrangian：

$$\mathcal{L} = f(x) + \sum_i \lambda_i g_i(x) + \sum_j \mu_j h_j(x).$$

KKT：

$$\nabla_x \mathcal{L} = 0,$$

$$g_i(x^*) \geq 0, \quad h_j(x^*) = 0,$$

$$\lambda_i \geq 0,$$

$$\lambda_i g_i(x^*) = 0.$$

互补松弛：

$$\lambda_i g_i(x^*) = 0$$

表示：约束不绑定 $g_i(x^*) > 0$ 时乘子为 0；约束绑定时 $g_i(x^*) = 0$，乘子可以为正。

### 5.4 凸性与全局最优

**核心公式**

凸集：若 $x, y \in C$，$\alpha \in [0,1]$，则

$$\alpha x + (1-\alpha)y \in C.$$

凸函数：

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y).$$

凹函数：

$$f(\alpha x + (1-\alpha)y) \ge \alpha f(x) + (1-\alpha) f(y).$$

Jensen：若 $f$ 凸，

$$f(\mathbb{E}X) \le \mathbb{E}[f(X)].$$

若 $u$ 凹，

$$u(\mathbb{E}W) \ge \mathbb{E}[u(W)],$$

对应风险厌恶。

优化充分性：在凸可行集上最大化凹函数，任何满足 KKT 的点都是全局最优。

## 6 比较静态、包络、对偶与固定点

### 6.1 隐函数定理与比较静态

**核心公式**

若

$$F(x, \theta) = 0,$$

且 $F_x$ 非奇异，则局部存在 $x(\theta)$，并且

$$\boxed{\frac{\mathrm{d}x}{\mathrm{d}\theta} = -F_x^{-1} F_\theta.}$$

一维形式：

$$\frac{\mathrm{d}x}{\mathrm{d}\theta} = -\frac{F_\theta}{F_x}.$$

市场均衡比较静态：

$$D(p,\tau)=S(p),$$
$$\frac{\mathrm{d}p^*}{\mathrm{d}\tau}=-\frac{D_\tau}{D_p-S_p}.$$

## 6.2 包络定理

**核心公式**

无约束 value function：

$$V(\theta)=\max_x f(x,\theta).$$

若 $x^*(\theta)$ 是内点解，则

$$\boxed{V_\theta(\theta)=f_\theta(x^*(\theta),\theta).}$$

约束问题：

$$V(a)=\max_x f(x)\quad \text{s.t.}\quad g(x)\le a.$$

若 Lagrangian 写作 $\mathcal{L}=f(x)+\lambda[a-g(x)]$，则

$$\frac{\partial V}{\partial a}=\lambda^*.$$

核心直觉：最优选择随参数变化带来的间接效应在一阶上消失，只剩 objective 或 constraint 对参数的直接效应。

## 6.3 经济对偶公式

消费者理论：

$$v(p,w)=\max_x u(x)\quad \text{s.t.}\quad p'x\le w.$$

支出函数：

$$e(p,\bar u)=\min_x p'x\quad \text{s.t.}\quad u(x)\ge \bar u.$$

Roy's identity：

$$x_i(p,w)=-\frac{\partial v(p,w)/\partial p_i}{\partial v(p,w)/\partial w}.$$

Shephard's lemma：

$$h_i(p,\bar u)=\frac{\partial e(p,\bar u)}{\partial p_i}.$$

Slutsky decomposition：

$$\frac{\partial x_i}{\partial p_j}=\frac{\partial h_i}{\partial p_j}-x_j\frac{\partial x_i}{\partial w}.$$

生产者理论：

$$\pi(p,w)=\max_{y,x} py-w'x\quad \text{s.t. production technology}.$$

Hotelling's lemma:

$$y_i(p,w) = \frac{\partial \pi(p,w)}{\partial p_i}.$$

成本函数:

$$C(w,y) = \min_x w'x \quad \text{s.t.} \quad f(x) \geq y.$$

Shephard input demand:

$$x_i(w,y) = \frac{\partial C(w,y)}{\partial w_i}.$$

### 6.4 固定点与均衡存在

**核心公式**

固定点:

$$x^* = T(x^*).$$

压缩映射: 若

$$d(Tx,Ty) \leq \beta d(x,y), \quad 0 < \beta < 1,$$

则存在唯一固定点, 且迭代 $x_{n+1} = T(x_n)$ 收敛。
Brouwer: 连续函数 $T : C \to C$, $C$ 非空、紧、凸, 则存在固定点。
Kakutani: 对应 $\Phi : C \rightrightarrows C$ 非空、凸值、紧值、上半连续, 则存在 $x^* \in \Phi(x^*)$。

经济应用: 一般均衡、Nash 均衡、混合策略、动态规划 Bellman operator。

## 7 动态优化、Bellman、最优控制与 HJB

### 7.1 离散时间 Euler 方程

典型消费储蓄问题:

$$\max_{\{c_t, a_{t+1}\}} \sum_{t=0}^{\infty} \beta^t u(c_t)$$

$$a_{t+1} = R(a_t + y_t - c_t).$$

Euler equation:

$$\boxed{u'(c_t) = \beta R \mathbb{E}_t[u'(c_{t+1})].}$$

资产定价 Euler:

$$1 = \mathbb{E}_t[M_{t+1} R_{t+1}],$$

其中 stochastic discount factor

$$M_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}.$$

## 7.2 Bellman 方程

**核心公式**

无限期离散动态规划：

$$V(s) = \max_{a \in \Gamma(s)} \{u(s,a) + \beta \mathbb{E}[V(s') \mid s, a]\}.$$

Bellman operator：

$$(TV)(s) = \max_{a \in \Gamma(s)} \{u(s,a) + \beta \mathbb{E}[V(s') \mid s, a]\}.$$

若 $T$ 是 $\beta$-contraction，则

$$\|TV - TW\| \le \beta \|V - W\|.$$

价值迭代：

$$V_{n+1} = TV_n.$$

## 7.3 连续时间最优控制

问题：

$$\max_{u(t)} \int_0^\infty e^{-\rho t} F(x(t), u(t))\, \mathrm{d}t$$

$$\dot{x} = f(x, u).$$

当前值 Hamiltonian：

$$\mathcal{H} = F(x,u) + \mu f(x,u).$$

必要条件：

$$\mathcal{H}_u = 0,$$

$$\dot{x} = \mathcal{H}_\mu = f(x,u),$$

$$\dot{\mu} = \rho\mu - \mathcal{H}_x.$$

横截条件常见形式：

$$\lim_{t \to \infty} e^{-\rho t} \mu(t) x(t) = 0.$$

### 7.4 HJB 方程

**核心公式**

连续时间 HJB：

$$\rho V(x) = \max_u \{F(x,u) + V_x(x) f(x,u)\}.$$

FOC：

$$F_u(x,u^*) + V_x(x) f_u(x,u^*) = 0.$$

随机扩散版本：若

$$\mathrm{d}X_t = \mu(X_t,u_t)\,\mathrm{d}t + \sigma(X_t,u_t)\,\mathrm{d}W_t,$$

则

$$\rho V(x) = \max_u \left\{ F(x,u) + V_x \mu(x,u) + \frac{1}{2} V_{xx} \sigma(x,u)^2 \right\}.$$

## 8 概率、条件期望与回归

### 8.1 概率基础

**核心公式**

期望：

$$\mathbb{E}[X] = \int x f_X(x)\,\mathrm{d}x.$$

方差：

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mathbb{E}X)^2] = \mathbb{E}[X^2] - (\mathbb{E}X)^2.$$

协方差：

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)] = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y.$$

相关系数：

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$

条件概率与 Bayes：

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)},$$

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)}.$$

## 8.2 条件期望

**核心公式**

离散条件期望：

$$\mathbb{E}[Y \mid X = x] = \sum_y yP(Y = y \mid X = x).$$

连续条件期望：

$$\mathbb{E}[Y \mid X = x] = \int y f_{Y|X}(y \mid x)\,\mathrm{d}y.$$

LIE：

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}(Y \mid X)].$$

条件方差分解：

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y \mid X)] + \mathrm{Var}(\mathbb{E}[Y \mid X]).$$

回归函数：

$$m(x) = \mathbb{E}[Y \mid X = x].$$

残差：

$$u = Y - m(X).$$

核心性质：

$$\mathbb{E}[u \mid X] = 0, \qquad \mathbb{E}[ug(X)] = 0.$$

## 8.3 OLS 与识别公式

Population OLS：

$$\beta = (\mathbb{E}[XX'])^{-1}\mathbb{E}[XY].$$

如果结构模型是

$$Y = X'\beta + \varepsilon,$$

且

$$\mathbb{E}[X\varepsilon] = 0,$$

则 OLS target 等于结构参数。

遗漏变量偏误两变量公式：若真模型

$$Y = \beta X + \gamma Z + u,$$

但回归遗漏 $Z$，则

$$\operatorname{plim} \hat{\beta}_{omit} = \beta + \gamma \frac{\mathrm{Cov}(X, Z)}{\mathrm{Var}(X)}.$$

二元处理：

$$Y = DY(1) + (1 - D)Y(0),$$

$$\mathbb{E}[Y \mid D=1]-\mathbb{E}[Y \mid D=0]=\mathbb{E}[Y(1)-Y(0) \mid D=1]+\{\mathbb{E}[Y(0) \mid D=1]-\mathbb{E}[Y(0) \mid D=0]\}.$$

第二项是 selection bias。

# 9 收敛、极限定理与渐近推断

## 9.1 收敛概念

| 概念 | 记号/定义 | 用途 |
|---|---|---|
| 依概率收敛 | $X_n \xrightarrow{p} X$, $P(\lvert X_n-X\rvert>\epsilon)\to 0$ | 一致性。 |
| 分布收敛 | $X_n \xrightarrow{d} X$ | 渐近分布、CLT。 |
| 几乎处处收敛 | $P(\lim X_n = X)=1$ | 强大数定律。 |
| 均方收敛 | $\mathbb{E}[(X_n-X)^2]\to 0$ | MSE、预测误差。 |

## 9.2 LLN、CLT、Slutsky、Delta

**核心公式**

LLN：若 $X_i$ i.i.d. 且 $\mathbb{E}\lvert X_i\rvert<\infty$，则

$$\bar{X}_n \xrightarrow{p} \mu.$$

CLT：若 $\mathbb{E}X_i=\mu$，$\mathrm{Var}(X_i)=\sigma^2<\infty$，则

$$\sqrt{n}(\bar{X}_n-\mu) \xrightarrow{d} N(0,\sigma^2).$$

Slutsky：若 $X_n \xrightarrow{d} X$，$Y_n \xrightarrow{p} c$，则

$$X_n+Y_n \xrightarrow{d} X+c, \quad X_nY_n \xrightarrow{d} cX.$$

Delta method：若

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,V),$$

则

$$\sqrt{n}(g(\hat{\theta})-g(\theta_0)) \xrightarrow{d} N(0,g'(\theta_0)Vg'(\theta_0)').$$

OLS 渐近正态：

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N(0,Q^{-1}\Omega Q^{-1}),$$

其中

$$Q=\mathbb{E}[X_iX_i'], \qquad \Omega=\mathbb{E}[X_iX_i'\varepsilon_i^2].$$

同方差时：

$$\Omega = \sigma^2 Q, \qquad \mathrm{Avar}(\hat{\beta}) = \sigma^2 Q^{-1}.$$

# 10 多元正态、二次型与金融风险

## 10.1 多元正态与协方差矩阵

**核心公式**

若 $X \sim N(\mu, \Sigma)$，密度为

$$f(x) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right].$$

线性变换：

$$AX + b \sim N(A\mu + b, A\Sigma A').$$

标准化二次型：

$$(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2_k.$$

协方差矩阵性质：

$$\Sigma = \mathbb{E}[(X-\mu)(X-\mu)'],$$

$$\Sigma \text{ symmetric positive semidefinite},$$

$$a'\Sigma a = \mathrm{Var}(a'X) \geq 0.$$

## 10.2 二次型检验统计量

若 $Z \sim N(0, I_k)$，则

$$Z'Z \sim \chi^2_k.$$

若

$$T = \frac{Z}{\sqrt{U/q}}, \quad Z \sim N(0,1),\ U \sim \chi^2_q,$$

则

$$T \sim t_q.$$

若

$$F = \frac{(U_1/q_1)}{(U_2/q_2)},$$

则

$$F \sim F_{q_1,q_2}.$$

### 10.3 投资组合风险

**核心公式**

投资组合收益：

$$R_p = w'R.$$

期望收益：

$$\mu_p = w'\mu.$$

方差：

$$\sigma_p^2 = w'\Sigma w.$$

两资产方差：

$$\sigma_p^2 = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\rho\sigma_1\sigma_2.$$

最小方差组合：

$$w_{GMV} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}.$$

PCA：若

$$\Sigma = Q\Lambda Q',$$

则主成分为

$$Z = Q'(X-\mu),$$

其方差为 $\Lambda$。第 $j$ 个主成分解释比例：

$$\frac{\lambda_j}{\sum_i \lambda_i}.$$

VaR 和 ES：若 loss $L$，置信水平 $\alpha$：

$$VaR_\alpha = \inf\{l : P(L \le l) \ge \alpha\},$$

$$ES_\alpha = \mathbb{E}[L \mid L \ge VaR_\alpha].$$

正态情形：若 $L \sim N(\mu, \sigma^2)$，

$$VaR_\alpha = \mu + \sigma z_\alpha,$$

$$ES_\alpha = \mu + \sigma\frac{\phi(z_\alpha)}{1-\alpha}.$$

## 11 Monte Carlo、Bootstrap 与模拟推断

### 11.1 Monte Carlo 估计

目标：

$$\theta = \mathbb{E}[g(X)].$$

模拟 $X_1,\dots,X_M$，估计：

$$\hat{\theta}_M = \frac{1}{M}\sum_{m=1}^{M} g(X_m).$$

标准误：

$$SE(\hat{\theta}_M) = \sqrt{\frac{\mathrm{Var}(g(X))}{M}}.$$

样本估计：

$$\widehat{SE} = \frac{s_g}{\sqrt{M}}.$$

误差阶：

$$\hat{\theta}_M - \theta = O_p(M^{-1/2}).$$

### 11.2 Bootstrap

给定样本 $Z_1,\dots,Z_n$，从经验分布重复有放回抽样，得到 bootstrap 样本 $Z_1^*,\dots,Z_n^*$，计算统计量 $T^*$。重复 $B$ 次，用 $\{T_b^*\}_{b=1}^B$ 近似统计量抽样分布。

Bootstrap 标准误：

$$\widehat{SE}_{boot} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(T_b^* - \bar{T}^*)^2}.$$

Percentile interval：

$$[Q_{\alpha/2}(T^*), Q_{1-\alpha/2}(T^*)].$$

**常见误用**

Bootstrap 不是魔法。它依赖样本能够代表总体。时间序列、聚类数据、重尾分布、极端分位数、弱识别问题中，普通 bootstrap 可能失败，需要 block bootstrap、cluster bootstrap 或专门方法。

## 12 模型识别：看到公式该想到什么

| 你看到的形式 | 应该想到的工具 | 典型课程 |
|---|---|---|
| $\int g(x)f(x)\,\mathrm{d}x$ | 期望、aggregation、welfare。 | 概率、微观、金融。 |
| $\nabla f = 0$ | 无约束优化 FOC。 | 微观、金融优化。 |
| $\nabla f = \lambda \nabla g$ | Lagrange、影子价格。 | 消费者/生产者理论。 |
| $\lambda g(x) = 0$ | KKT、角点解。 | 高微、金融约束。 |
| $F(x,\theta) = 0$ | 隐函数、比较静态。 | 高微、宏观均衡。 |
| $V(\theta) = \max_x f(x,\theta)$ | 包络定理。 | 高微、动态优化。 |
| $x = T(x)$ | 固定点和均衡存在。 | 一般均衡、博弈论。 |

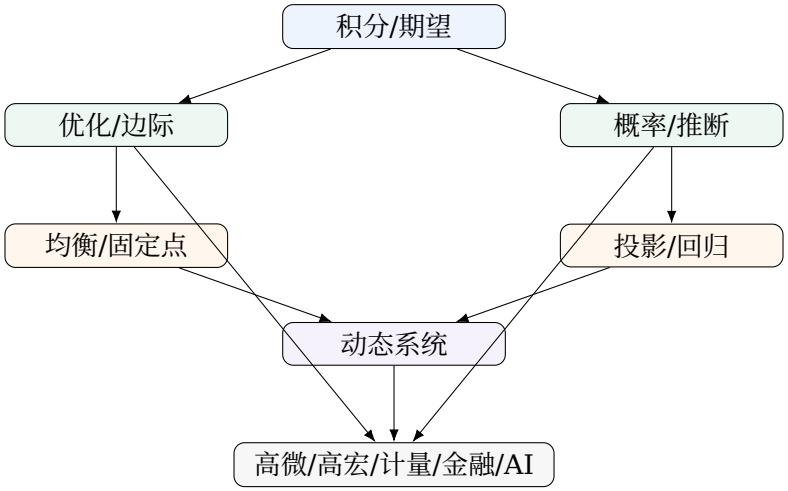
| | | |
|---|---|---|
| $\dot{x} = f(x)$ | ODE、相图、稳定性。 | 宏观动态。 |
| $x_{t+1} = F(x_t)$ | 差分方程、离散动态。 | 宏观、计量时间序列。 |
| $V = TV$ | Bellman、压缩映射。 | 动态规划。 |
| $\rho V = \max\{u + V_x f\}$ | HJB。 | 连续时间宏观/金融。 |
| $\mathbb{E}[\varepsilon \mid X] = 0$ | 外生性、条件均值独立。 | 计量。 |
| $\mathbb{E}[X\varepsilon] = 0$ | 矩条件、OLS/GMM。 | 计量理论。 |
| $w'\Sigma w$ | 二次型、组合风险。 | 金融。 |
| $\sqrt{n}(\hat{\theta} - \theta) \Rightarrow N(0, V)$ | 渐近正态、标准误。 | 计量、统计。 |

## 13 常见公式误用清单

**常见误用**

1. **把局部最优当全局最优**。只有凹性/凸性等条件成立时，FOC 或 KKT 才能保证全局最优。
2. **把样本正交当识别**。OLS 残差与样本解释变量正交是代数恒等式，不是外生性证据。
3. **忘记 Jacobian 绝对值**。变量替换中面积缩放必须取绝对值。
4. **把稳定条件写错**。连续时间看特征值实部小于 0；离散时间看特征值模小于 1。
5. **把影子价格符号弄反**。 Lagrangian 约束写法不同，乘子符号和解释会改变。
6. **对跳跃点乱用 Fourier 值**。 Fourier 级数在跳跃点通常收敛到左右极限平均。
7. **混淆条件期望和线性投影**。 $E[Y \mid X]$ 是最优非线性预测；OLS 是最佳线性预测。
8. **把 CLT 用在重尾无方差情形**。普通 CLT 需要有限方差或相应条件。
9. **忽略横截条件**。动态优化和资产定价中，不加 TVC 可能保留泡沫或 Ponzi 路径。
10. **把模拟样本数当真实样本数**。 Monte Carlo error 随模拟次数下降，但不能修复模型设定错误。

## 14 一页终极复习图谱




**进阶连接**

这份公式大全之后，最值得继续做的不是新理论章，而是两个训练型文件：

1. **Appendix B：综合习题册**。按模块混合出题，不提示用哪个公式。
2. **Appendix C：论文阅读工具包**。把论文中的模型、命题、证明、估计式拆成数学对象。

公式大全解决“我该查什么”；综合习题解决“我会不会用”；论文工具包解决“我能不能读研究”。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Appendix B
# 全书综合习题册与详细答案

本附录不是新章节，而是对 Ch1–Ch31 的综合训练。它故意不按单一章节机械排列，而是把多重积分、动态系统、优化、对偶、概率、计量和金融风险混合起来，让你练习真正的模型识别能力：看到问题先判断用什么工具，再写出公式、条件和经济解释。

math4econ personalized notes
2026

# 目录

# Appendix B 使用说明

**使用说明**

这份习题册建议按三轮使用。

1. **第一轮：看题目识别工具**。每题先写一句话：这题主要用积分、投影、动态、优化、概率还是计量？不要急着算。
2. **第二轮：独立做主干步骤**。对每题至少写出关键公式和一阶条件，哪怕细节算不完。
3. **第三轮：对照详细答案**。重点看答案中“为什么这样设”和“经济解释”部分，而不是只核对最终数值。

本册分为六组：多重积分与概率化思维、向量/投影/动态系统、优化与对偶、动态优化与计算、概率计量与金融风险、综合建模 Capstone。

**解题策略**

综合题的标准流程：

$$\text{识别对象} \to \text{写数学形式} \to \text{检查条件} \to \text{求解/证明} \to \text{经济金融解释} \to \text{误差与限制}$$

如果只会算而不会解释，你只能应付作业；如果能解释条件和限制，才算真正能读论文和建模型。

# 1 A 组：多重积分、Jacobian 与概率化思维

**A1. 改积分次序与异质性福利**

某经济体中个体由收入 $y$ 与教育年限 $s$ 描述。考虑区域

$$R = \{(y, s) : 0 \le s \le 2,\ s^2 \le y \le 4\}.$$

社会福利密度为 $w(y, s) = ys$。

(1) 写出总福利 $W = \iint_R ys \, \mathrm{d}y \, \mathrm{d}s$。

(2) 改写为先对 $s$ 积分、再对 $y$ 积分的形式。

(3) 计算 $W$。

(4) 解释“改积分次序”在异质性经济模型中的含义。

**A2. 随机变量变换与 log return**

资产价格 $P$ 服从 lognormal 分布，即 $X = \log P \sim N(\mu, \sigma^2)$。令收益变量 $R = P/P_0 - 1$，其中 $P_0 > 0$ 是常数。

(1) 写出 $P$ 的密度。

(2) 推导 $R$ 的密度。

(3) 说明为什么金融中常使用 $\log P$ 或 log return 而非直接价格。

**A3. 二维正态的线性组合**

设 $X, Y$ 的均值分别为 $\mu_X, \mu_Y$，方差为 $\sigma_X^2, \sigma_Y^2$，协方差为 $\sigma_{XY}$。投资组合收益为

$$R_p = wX + (1 - w)Y.$$

(1) 求 $\mathbb{E}[R_p]$ 与 $\mathrm{Var}(R_p)$。

(2) 对 $\mathrm{Var}(R_p)$ 关于 $w$ 求最小值点。

(3) 解释相关性如何影响分散化。

**A4. 含参积分与税收变化**

收入 $y$ 在 $[0, 1]$ 上密度为 $f(y) = 2y$。税率参数为 $\tau \in [0, 1)$，税后效用为

$$u(y, \tau) = \log(1 + (1 - \tau)y).$$

社会福利为

$$W(\tau)=\int_0^1 \log(1+(1-\tau)y)2y\,\mathrm{d}y.$$

(1) 在积分号下求 $W'(\tau)$。

(2) 判断 $W'(\tau)$ 的符号。

(3) 给出经济解释。

**A5. 联合密度与边际密度**

设 $(X,Y)$ 在三角区域 $0\le y\le x\le 1$ 上联合密度为 $f(x,y)=cx$。

(1) 求常数 $c$。

(2) 求边际密度 $f_X(x)$ 与 $f_Y(y)$。

(3) 求 $\mathbb{E}[Y\mid X=x]$。

## 2 B 组：投影、傅里叶、ODE 与差分方程

**B1. 函数投影与 OLS 类比**

在 $[-\pi,\pi]$ 上，用 $a\sin x$ 近似 $f(x)=x$，使

$$\int_{-\pi}^{\pi}(x-a\sin x)^2\,\mathrm{d}x$$

最小。

(1) 求最优 $a$。

(2) 说明该条件和 OLS 正交条件的关系。

**B2. Parseval 与波动分解**

已知

$$x=2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\sin(nx),\qquad -\pi<x<\pi.$$

(1) 使用 Parseval identity 推导 $\sum_{n=1}^{\infty}1/n^2$。

(2) 解释 Parseval 在金融收益频率分解中的含义。

**B3. 一阶 ODE 与半衰期**

宏观变量 $x(t)$ 满足

$$\dot{x}=-0.3(x-5),\qquad x(0)=11.$$

(1) 求路径 $x(t)$。

(2) 求 half-life。

(3) 解释 $0.3$ 的经济含义。

### B4. 离散债务动态

债务比率满足

$$b_{t+1} = 1.04 b_t - 0.02.$$

(1) 求稳态。

(2) 判断稳定性。

(3) 如果 $b_0 = 0.7$，写出 $b_t$。

(4) 给出财政含义。

### B5. 二维线性系统与 saddle

考虑连续时间系统

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

(1) 求通解。

(2) 判断原点稳定性。

(3) 解释 saddle path 的含义。

### B6. 非线性动态与局部线性化

设

$$\dot{x} = x(1-x) - y, \qquad \dot{y} = 0.5(x - y).$$

(1) 求稳态。

(2) 求 Jacobian。

(3) 在每个稳态处判断局部稳定性。

# 3 C 组：优化、KKT、对偶与固定点

## C1. Hessian 与利润最大化

企业利润为

$$\pi(x,y) = 10x + 8y - x^2 - y^2 - xy.$$

(1) 求无约束最优 $(x^*, y^*)$。

(2) 用 Hessian 判断是否为全局最大。

(3) 解释交叉项 $-xy$ 的经济含义。

## C2. 消费者 Lagrange 问题

消费者最大化

$$u(x,y) = \log x + 2\log y$$

约束为

$$p_x x + p_y y = m.$$

(1) 求 Marshallian demand。

(2) 求间接效用函数。

(3) 解释 Lagrange 乘子。

## C3. KKT 与角点解

求解

$$\max_{x \geq 0} \quad f(x) = -(x-2)^2$$

附加约束

$$x \leq 1.$$

(1) 用 KKT 条件求解。

(2) 哪个约束 binding?

(3) 解释互补松弛。

## C4. Jensen 与风险厌恶

设 $u(c) = \sqrt{c}$，且随机消费 $C$ 取 1 和 9，概率各为 $1/2$。

(1) 计算 $\mathbb{E}[u(C)]$ 与 $u(\mathbb{E}C)$。

(2) 用 Jensen inequality 解释差异。

(3) 求 certainty equivalent。

### C5. 包络定理与成本函数

厂商成本最小化：

$$C(w,r,q)=\min_{L,K} wL+rK \quad \text{s.t.} \quad L^{1/2}K^{1/2}\geq q.$$

(1) 求条件要素需求。

(2) 求成本函数 $C(w,r,q)$。

(3) 验证 Shephard's lemma。

### C6. Brouwer 不动点直觉

设 $F:[0,1]\to[0,1]$ 连续。

(1) 证明存在 $x^*$ 使 $F(x^*)=x^*$。

(2) 解释这和均衡存在有什么关系。

## 4 D 组：动态优化、Bellman、HJB 与数值方法

### D1. 两期消费储蓄 Euler 方程

消费者最大化

$$\log c_0+\beta\log c_1$$

约束为

$$c_0+s=y_0, \qquad c_1=(1+r)s+y_1.$$

(1) 推导 Euler 方程。

(2) 求 $c_0,c_1,s$。

(3) 解释利率上升的替代效应。

### D2. Bellman 方程与消费储蓄

无限期问题

$$V(a)=\max_{0\leq c\leq a}\{\log c+\beta V(R(a-c))\}.$$

猜测 $V(a)=A+B\log a$。

(1) 求最优消费比例。

(2) 求 $B$。

(3) 解释策略函数。

### D3. 最优控制与 Hotelling 规则

资源存量 $S$ 满足

$$\dot{S} = -q,$$

社会规划者最大化

$$\int_0^{\infty} e^{-\rho t} p q(t)\,\mathrm{d}t$$

且 $q(t) \geq 0$、总资源有限。若单位开采成本为 $c(S)$ 且边际净收益为 $p - c(S)$，说明 Hotelling 规则中影子价格的动态含义。

### D4. 数值固定点迭代

求解固定点

$$x = \cos x.$$

(1) 写出简单迭代算法。

(2) 给出收敛的直觉条件。

(3) 从 $x_0 = 1$ 手算前三次迭代到三位小数。

### D5. Monte Carlo 估计

设 $U \sim U[0,1]$，目标为

$$\theta = \mathbb{E}[e^{U^2}].$$

(1) 写出 Monte Carlo 估计量。

(2) 给出标准误估计。

(3) 说明样本量变为四倍时标准误如何变化。

## 5 E 组：概率、计量、金融风险与模拟推断

### E1. 条件期望与迭代期望

设

$$Y = 2X + \varepsilon, \qquad \mathbb{E}[\varepsilon \mid X] = 0.$$

(1) 求 $\mathbb{E}[Y \mid X]$。

(2) 求 $\mathbb{E}[Y]$。

(3) 证明 $\mathbb{E}[X\varepsilon] = 0$。

### E2. Population OLS

设 $Y = X'\beta + u$，且 $\mathbb{E}[Xu] = 0$，$\mathbb{E}[XX']$ 可逆。

(1) 从最小化 $\mathbb{E}[(Y - X'b)^2]$ 推导 $\beta$。

(2) 说明正交条件。

### E3. Omitted variable bias

真实模型为

$$Y = \beta X + \gamma Z + u, \qquad \mathbb{E}[u \mid X, Z] = 0.$$

如果回归时遗漏 $Z$，只回归 $Y$ 对 $X$，求 population slope。

### E4. LLN、CLT 与标准误

设 $X_1, \ldots, X_n$ i.i.d.，均值 $\mu$、方差 $\sigma^2$。

(1) 写出 LLN。

(2) 写出 CLT。

(3) 用它构造大样本置信区间。

### E5. Delta method

若

$$\sqrt{n}(\hat{\theta} - \theta) \Rightarrow N(0, V),$$

且 $g$ 可微。推导 $g(\hat{\theta})$ 的渐近分布。然后令 $g(\theta) = \log\theta$，写出具体结果。

### E6. 投资组合 VaR 与 ES

假设组合收益 $R_p \sim N(\mu, \sigma^2)$，损失定义为 $L = -R_p$。

(1) 写出置信水平 $\alpha$ 下的 VaR。

(2) 写出 Expected Shortfall。

(3) 解释正态假设的风险。

# 6 F 组：Capstone 综合建模题

**Capstone**

**F1. 从异质性收入到税收福利比较静态**收入 $Y$ 的密度为 $f(y;\theta)=\theta y^{\theta-1}$，$0<y<1$，$\theta>0$。政府征比例税 $\tau$，税收返还为均等转移 $T(\tau)=\tau\mathbb{E}[Y]$。个体消费为

$$c(y,\tau)=(1-\tau)y+T(\tau).$$

社会福利为

$$W(\tau,\theta)=\int_0^1 \log c(y,\tau)\,f(y;\theta)\,\mathrm{d}y.$$

(1) 求 $\mathbb{E}[Y]$。

(2) 写出 $W_\tau$ 的积分表达式。

(3) 解释为什么收入分布参数 $\theta$ 会影响最优税率。

(4) 如果要数值求最优税率，你会用什么算法和什么检查?

**Capstone**

**F2. 线性动态系统中的宏观稳定性**考虑

$$\begin{pmatrix} x_{t+1} \\ \pi_{t+1} \end{pmatrix} = \begin{pmatrix} 0.8 & -0.2 \\ 0.3 & 0.7 \end{pmatrix} \begin{pmatrix} x_t \\ \pi_t \end{pmatrix} + \begin{pmatrix} g_t \\ 0 \end{pmatrix}.$$

(1) 写出无冲击系统的稳定性判别方法。

(2) 计算特征值。

(3) 若 $g_0=1$ 且之后为 0，写出冲击响应的矩阵形式。

(4) 解释为什么特征值决定财政冲击的持久性。

**Capstone**

**F3. 消费储蓄、Euler 方程与回归检验**假设消费者 Euler 方程为

$$\mathbb{E}_t\left[\beta R_{t+1}\frac{u'(c_{t+1})}{u'(c_t)}-1\right]=0.$$

(1) 若 $u(c)=\log c$，写出 Euler moment。

(2) 说明如何构造样本矩条件。

(3) 解释为什么工具变量必须在 $t$ 期信息集中。

(4) 连接到 GMM 的正交条件。

# 7 详细答案

## 7.1 A 组答案

**A1 解答**

(1) 按区域定义，

$$W = \int_0^2 \int_{s^2}^4 ys \, \mathrm{d}y \, \mathrm{d}s.$$

(2) 改次序要先看区域。由 $s^2 \le y \le 4$ 和 $0 \le s \le 2$ 得 $0 \le y \le 4$，且 $0 \le s \le \sqrt{y}$。所以

$$W = \int_0^4 \int_0^{\sqrt{y}} ys \, \mathrm{d}s \, \mathrm{d}y.$$

(3) 计算：

$$W = \int_0^2 s \left[ \frac{y^2}{2} \right]_{s^2}^4 \mathrm{d}s = \int_0^2 s \left( 8 - \frac{s^4}{2} \right) \mathrm{d}s = \left[ 4s^2 - \frac{s^6}{12} \right]_0^2 = 16 - \frac{64}{12} = \frac{32}{3}.$$

(4) 改积分次序不是纯技巧，而是改变 aggregation 的视角：先按教育分组再在收入上加总，或先按收入分组再在教育上加总。异质性模型中，很多政策效应取决于你如何切分人群。

**A2 解答**

(1) $P = \mathrm{e}^X$，所以

$$f_P(p) = \frac{1}{p\sigma\sqrt{2\pi}} \exp\left[ -\frac{(\log p - \mu)^2}{2\sigma^2} \right], \qquad p > 0.$$

(2) $R = P/P_0 - 1$，所以 $P = P_0(1+R)$，要求 $R > -1$。Jacobian 为

$$\left| \frac{\mathrm{d}p}{\mathrm{d}r} \right| = P_0.$$

因此

$$f_R(r) = f_P(P_0(1+r))P_0 = \frac{1}{(1+r)\sigma\sqrt{2\pi}} \exp\left[ -\frac{(\log(P_0(1+r)) - \mu)^2}{2\sigma^2} \right], \quad r > -1.$$

(3) log price 把乘法增长转成加法增长，并且正价格自动满足。log return 可跨期相加，更适合连续复利和正态近似。

**A3 解答**

期望：

$$\mathbb{E}[R_p] = w\mu_X + (1-w)\mu_Y.$$

方差:

$$\mathrm{Var}(R_p) = w^2\sigma_X^2 + (1-w)^2\sigma_Y^2 + 2w(1-w)\sigma_{XY}.$$

对 $w$ 求导:

$$2w\sigma_X^2 - 2(1-w)\sigma_Y^2 + 2(1-2w)\sigma_{XY} = 0.$$

整理得

$$w^* = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}.$$

相关性越低，$\sigma_{XY}$ 越小，组合方差越能被分散化降低；若完全正相关，分散化效果最弱。

**A4 解答**

在常规支配条件下可积分下求导:

$$W'(\tau) = \int_0^1 \frac{\partial}{\partial\tau}\log(1+(1-\tau)y)2y\,\mathrm{d}y = \int_0^1 \frac{-y}{1+(1-\tau)y}2y\,\mathrm{d}y.$$

所以

$$W'(\tau) = -\int_0^1 \frac{2y^2}{1+(1-\tau)y}\,\mathrm{d}y < 0.$$

经济含义：在没有公共品或再分配收益进入效用的设定下，提高税率只降低税后消费，因此福利下降。

**A5 解答**

归一化:

$$1 = \int_0^1\int_0^x cx\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 cx^2\,\mathrm{d}x = \frac{c}{3},$$

所以 $c=3$。边际密度:

$$f_X(x) = \int_0^x 3x\,\mathrm{d}y = 3x^2, \quad 0\le x\le 1.$$

对 $Y$，给定 $y\le x\le 1$,

$$f_Y(y) = \int_y^1 3x\,\mathrm{d}x = \frac{3}{2}(1-y^2), \quad 0\le y\le 1.$$

条件密度

$$f_{Y|X}(y|x) = \frac{3x}{3x^2} = \frac{1}{x}, \quad 0\le y\le x.$$

所以 $Y|X=x \sim U[0,x]$,

$$\mathbb{E}[Y|X=x] = \frac{x}{2}.$$

## 7.2 B 组答案

**B1 解答**

最小化目标对 $a$ 求导:

$$\frac{\mathrm{d}}{\mathrm{d}a}\int_{-\pi}^{\pi}(x-a\sin x)^2\,\mathrm{d}x=-2\int_{-\pi}^{\pi}(x-a\sin x)\sin x\,\mathrm{d}x=0.$$

所以

$$a=\frac{\int_{-\pi}^{\pi}x\sin x\,\mathrm{d}x}{\int_{-\pi}^{\pi}\sin^2 x\,\mathrm{d}x}.$$

分子为 $2\int_0^{\pi}x\sin x\,\mathrm{d}x=2\pi$，分母为 $\pi$，所以 $a=2$。该条件就是残差 $x-a\sin x$ 与解释变量 $\sin x$ 正交，和 OLS 的 $E[X\varepsilon]=0$ 或样本中的 $X'\hat{\varepsilon}=0$ 完全同构。

**B2 解答**

Parseval 对该正弦展开给出

$$\frac{1}{\pi}\int_{-\pi}^{\pi}x^2\,\mathrm{d}x=\sum_{n=1}^{\infty}\left(\frac{2(-1)^{n+1}}{n}\right)^2.$$

左边为

$$\frac{1}{\pi}\cdot\frac{2\pi^3}{3}=\frac{2\pi^2}{3}.$$

右边为

$$4\sum_{n=1}^{\infty}\frac{1}{n^2}.$$

所以

$$\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{\pi^2}{6}.$$

金融含义：时域中的总平方波动可以分解到不同频率。低频可能对应长期趋势或商业周期，高频可能对应短期冲击或微观结构噪声。

**B3 解答**

令 $z=x-5$，则 $\dot{z}=-0.3z$，所以

$$z(t)=z(0)e^{-0.3t}=6e^{-0.3t}.$$

因此

$$x(t)=5+6e^{-0.3t}.$$

半衰期满足 $e^{-0.3T} = 1/2$,

$$T = \frac{\log 2}{0.3} \approx 2.31.$$

0.3 是调整速度，越大表示偏离长期均值消失越快。

**B4 解答**

稳态满足

$$b^* = 1.04b^* - 0.02,$$

故

$$-0.04b^* = -0.02, \quad b^* = 0.5.$$

偏离形式：

$$b_t - 0.5 = 1.04^t(b_0 - 0.5).$$

若 $b_0 = 0.7$,

$$b_t = 0.5 + 0.2(1.04)^t.$$

因 $|1.04| > 1$，稳态不稳定。财政含义：若有效滚债系数大于 1，偏离目标债务水平会被放大，除非财政反应随债务上升而增强。

**B5 解答**

两个方程分离：

$$\dot{x} = 2x, \qquad \dot{y} = -y.$$

通解为

$$x(t) = x_0 e^{2t}, \qquad y(t) = y_0 e^{-t}.$$

原点有一个正特征值和一个负特征值，因此是 saddle，不稳定。只有初始点位于稳定臂 $x_0 = 0$ 上时，路径才收敛到原点。宏观中的 saddle path 意味着某些跳跃变量必须立刻选到正确路径，否则系统发散。

**B6 解答**

稳态满足

$$y = x(1 - x), \qquad y = x.$$

代入得

$$x = x(1 - x) \Rightarrow x^2 = 0,$$

所以唯一稳态为 $(0,0)$。Jacobian 为

$$J(x,y)=\begin{pmatrix}1-2x & -1\\ 0.5 & -0.5\end{pmatrix}.$$

在 $(0,0)$ 处：

$$J=\begin{pmatrix}1 & -1\\ 0.5 & -0.5\end{pmatrix}.$$

迹为 0.5，行列式为 $1(-0.5)-(-1)(0.5)=0$。特征值为 0 和 0.5。由于存在正特征值，稳态非稳定。零特征值意味着线性化不能给出完全分类，但已经足以说明不是局部渐近稳定。

## 7.3 C 组答案

### C1 解答

一阶条件：

$$\pi_x=10-2x-y=0,\qquad \pi_y=8-2y-x=0.$$

解线性方程：从 $y=10-2x$，代入 $x+2y=8$：

$$x+20-4x=8\Rightarrow x=4,\quad y=2.$$

Hessian：

$$H=\begin{pmatrix}-2 & -1\\ -1 & -2\end{pmatrix}.$$

主子式：第一个为 $-2<0$，二阶行列式为 $4-1=3>0$，所以 $H$ 负定，利润严格凹，$(4,2)$ 是全局最大。交叉项 $-xy$ 表示两种投入或产品之间存在拥挤/替代式相互影响。

### C2 解答

Lagrangian：

$$\mathcal{L}=\log x+2\log y+\lambda(m-p_xx-p_yy).$$

FOC：

$$\frac{1}{x}=\lambda p_x,\qquad \frac{2}{y}=\lambda p_y.$$

相除得

$$\frac{y}{2x}=\frac{p_x}{p_y}\Rightarrow p_yy=2p_xx.$$

预算为 $p_x x + p_y y = m$，所以 $3p_x x = m$，

$$x^* = \frac{m}{3p_x}, \qquad y^* = \frac{2m}{3p_y}.$$

间接效用：

$$v(p_x, p_y, m) = \log \frac{m}{3p_x} + 2\log \frac{2m}{3p_y}.$$

乘子 $\lambda$ 是收入的边际效用，即放松预算一单位带来的最大效用增量。

**C3 解答**

约束写为 $g_1(x) = -x \le 0$，$g_2(x) = x - 1 \le 0$。最大化可转为最小化 $-f$，也可用最大化 KKT。直觉上无约束最大点 $x = 2$，但约束 $x \le 1$，所以最优为 $x^* = 1$。下界不 binding，上界 binding。互补松弛表示：若约束不紧，则其乘子为零；若乘子为正，则约束必须正好等号成立。这里 $x \le 1$ 限制了想选 2 的愿望，因此它有正影子价值。

**C4 解答**

$$\mathbb{E}[u(C)] = \frac{1}{2}\sqrt{1} + \frac{1}{2}\sqrt{9} = 2.$$

$$\mathbb{E}[C] = 5, \qquad u(\mathbb{E}C) = \sqrt{5} \approx 2.236.$$

由于 $u$ 凹，Jensen 给出

$$\mathbb{E}[u(C)] \le u(\mathbb{E}C).$$

certainty equivalent $CE$ 满足 $\sqrt{CE} = 2$，所以 $CE = 4$。风险溢价为 $\mathbb{E}C - CE = 1$。

**C5 解答**

约束在最优处 binding：$L^{1/2}K^{1/2} = q$，即 $LK = q^2$。Lagrangian：

$$\mathcal{L} = wL + rK + \lambda(q - L^{1/2}K^{1/2}).$$

FOC 给出 MRTS 条件：

$$\frac{MP_L}{MP_K} = \frac{w}{r}.$$

对 Cobb-Douglas $MP_L/MP_K = K/L$，所以

$$\frac{K}{L} = \frac{w}{r} \Rightarrow K = \frac{w}{r}L.$$

代入 $LK = q^2$：

$$L \cdot \frac{w}{r}L = q^2 \Rightarrow L^* = q\sqrt{\frac{r}{w}}, \quad K^* = q\sqrt{\frac{w}{r}}.$$

成本：

$$C = wL^* + rK^* = q\sqrt{wr} + q\sqrt{wr} = 2q\sqrt{wr}.$$

Shephard：

$$\frac{\partial C}{\partial w} = q\frac{r}{\sqrt{wr}} = q\sqrt{\frac{r}{w}} = L^*,$$

同理 $C_r = K^*$。

**C6 解答**

定义 $g(x) = F(x) - x$。因为 $F$ 连续，所以 $g$ 连续。由于 $F(0) \in [0,1]$,

$$g(0) = F(0) - 0 \geq 0.$$

由于 $F(1) \in [0,1]$,

$$g(1) = F(1) - 1 \leq 0.$$

若端点已有零点则结束；否则由介值定理，存在 $x^* \in (0,1)$ 使 $g(x^*) = 0$，即 $F(x^*) = x^*$。均衡存在证明通常也是构造一个从策略/价格集合到自身的连续映射或对应，然后用不动点定理找自我一致点。

## 7.4 D 组答案

**D1 解答**

用预算消去 $s$：$s = y_0 - c_0$,

$$c_1 = (1+r)(y_0 - c_0) + y_1.$$

FOC：

$$\frac{1}{c_0} = \beta(1+r)\frac{1}{c_1}.$$

这就是 Euler 方程：

$$\frac{u'(c_0)}{\beta u'(c_1)} = 1 + r.$$

由 $c_1 = \beta(1+r)c_0$ 与终身预算

$$c_0 + \frac{c_1}{1+r} = y_0 + \frac{y_1}{1+r}$$

得

$$c_0(1+eta) = y_0 + \frac{y_1}{1+r},$$

所以

$$c_0 = \frac{1}{1+\beta}\left(y_0 + \frac{y_1}{1+r}\right), \quad c_1 = \frac{\beta(1+r)}{1+\beta}\left(y_0 + \frac{y_1}{1+r}\right).$$

利率上升提高未来消费相对当前消费的回报，替代效应鼓励减少当前消费、增加储蓄。

**D2 解答**

猜测 $V(a) = A + B\log a$。令消费比例 $c = \alpha a$，下期资产 $a' = R(1-\alpha)a$。目标：

$$\log(\alpha a) + \beta[A + B\log(R(1-\alpha)a)].$$

与 $\alpha$ 有关的部分为

$$\log\alpha + \beta B\log(1-\alpha).$$

FOC：

$$\frac{1}{\alpha} - \frac{\beta B}{1-\alpha} = 0 \Rightarrow \alpha = \frac{1}{1+etaB}.$$

比较 $\log a$ 系数：

$$B = 1 + etaB \Rightarrow B = \frac{1}{1-\beta}.$$

所以

$$\alpha = \frac{1}{1+\beta/(1-\beta)} = 1-\beta.$$

策略函数为

$$c^*(a) = (1-\beta)a.$$

这表示固定消费财富的一部分，储蓄比例为 $\beta$。

**D3 解答**

Hotelling 规则的核心是：地下资源的影子价格必须按贴现率增长，否则规划者可以通过改变开采时点提高现值。若今天开采一单位资源，获得当前净收益；若保留到未来，获得未来净收益。最优路径要求边际上二者现值相等。影子价格是资源存量约束的乘子，表示多一单位地下资源对目标函数的边际价值。若资源越稀缺，影子价格越高；动态条件描述这个稀缺性租金如何随时间演化。

**D4 解答**

简单迭代：

$$x_{n+1} = \cos x_n.$$

若在固定点附近 $|g'(x)| = | -$

$sinx| < 1$，迭代局部收敛。手算：

$$x_0 = 1, \quad x_1 = \cos 1 \approx 0.540,$$

$$x_2 = \cos 0.540 \approx 0.858, \quad x_3 = \cos 0.858 \approx 0.654.$$

它会震荡但逐渐收敛到约 $0.739$。

**D5 解答**

生成 $U_1, \ldots, U_N \sim U[0,1]$，估计量为

$$\hat{\theta}_N = \frac{1}{N} \sum_{i=1}^{N} e^{U_i^2}.$$

样本标准误估计：

$$\widehat{se}(\hat{\theta}_N) = \frac{s_N}{\sqrt{N}},$$

其中 $s_N^2$ 是 $e^{U_i^2}$ 的样本方差。样本量变为四倍，标准误除以 2，这是 Monte Carlo 的 $1/\sqrt{N}$ 收敛率。

## 7.5 E 组答案

**E1 解答**

由条件均值零：

$$\mathbb{E}[Y|X] = \mathbb{E}[2X + \varepsilon|X] = 2X + \mathbb{E}[\varepsilon|X] = 2X.$$

由迭代期望：

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}(Y|X)] = 2\mathbb{E}[X].$$

并且

$$\mathbb{E}[X\varepsilon] = \mathbb{E}[\mathbb{E}(X\varepsilon|X)] = \mathbb{E}[X\mathbb{E}(\varepsilon|X)] = 0.$$

**E2 解答**

最小化

$$Q(b) = \mathbb{E}[(Y - X'b)^2].$$

FOC：

$$\frac{\partial Q}{\partial b} = -2\mathbb{E}[X(Y - X'b)] = 0.$$

所以

$$\mathbb{E}[XX']b = \mathbb{E}[XY].$$

若 $\mathbb{E}[XX']$ 可逆，

$$b = \mathbb{E}[XX']^{-1}\mathbb{E}[XY].$$

当 $Y = X'\beta + u$ 且 $\mathbb{E}[Xu] = 0$ 时，

$$\mathbb{E}[XY] = \mathbb{E}[X(X'\beta + u)] = \mathbb{E}[XX']\beta,$$

故 $b = \beta$。正交条件是解释变量与投影误差正交。

**E3 解答**

只回归 $Y$ 对 $X$ 的 population slope 为

$$\tilde{\beta} = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}.$$

代入真实模型：

$$\mathrm{Cov}(X, Y) = \beta\,\mathrm{Var}(X) + \gamma\,\mathrm{Cov}(X, Z) + \mathrm{Cov}(X, u).$$

由 $\mathbb{E}[u|X, Z] = 0$ 得 $\mathrm{Cov}(X, u) = 0$。所以

$$\tilde{\beta} = \beta + \gamma\frac{\mathrm{Cov}(X, Z)}{\mathrm{Var}(X)}.$$

遗漏变量偏误等于

$$\gamma\frac{\mathrm{Cov}(X, Z)}{\mathrm{Var}(X)}.$$

**E4 解答**

LLN：

$$\bar{X}_n \xrightarrow{p} \mu.$$

CLT：

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \Rightarrow N(0, 1).$$

若用样本标准差 $s$ 估计 $\sigma$，大样本置信区间为

$$\bar{X}_n \pm z_{1-\alpha/2}\frac{s}{\sqrt{n}}.$$

LLN 保证估计中心对，CLT 给出抽样误差形状。

**E5 解答**

Delta method：

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \Rightarrow N(0, [g'(\theta)]^2 V).$$

若 $g(\theta) = \log\theta$，则 $g'(\theta) = 1/\theta$，因此

$$\sqrt{n}(\log\hat{\theta} - \log\theta) \Rightarrow N\left(0, \frac{V}{\theta^2}\right).$$

**E6 解答**

损失 $L = -R_p \sim N(-\mu, \sigma^2)$。令 $z_\alpha = \Phi^{-1}(\alpha)$。VaR 定义为损失的 $\alpha$ 分位数：

$$VaR_\alpha = -\mu + \sigma z_\alpha.$$

正态下 ES：

$$ES_\alpha = -\mu + \sigma\frac{\phi(z_\alpha)}{1-\alpha}.$$

正态假设风险：金融收益常有厚尾、偏度、波动聚集，正态 VaR 可能低估极端损失。

## 7.6 F 组答案

**F1 解答**

(1) 对 $f(y;\theta) = \theta y^{\theta-1}$，

$$\mathbb{E}[Y] = \int_0^1 y\theta y^{\theta-1}\,\mathrm{d}y = \frac{\theta}{\theta+1}.$$

所以

$$T(\tau) = \tau\frac{\theta}{\theta+1}.$$

(2) 消费

$$c(y,\tau) = (1-\tau)y + \tau\frac{\theta}{\theta+1}.$$

因此

$$W_\tau = \int_0^1 \frac{-y + \theta/(\theta+1)}{(1-\tau)y + \tau\theta/(\theta+1)}\theta y^{\theta-1}\,\mathrm{d}y.$$

(3) $\theta$ 改变收入分布形状，也改变均等转移的平均收入基础。若低收入群体权重更高，再分配带来的边际效用收益更大，最优税率可能更高。(4) 数值求解可用网格搜索、二分法或 Newton method。需要检查：目标是否凹、边界点 $\tau = 0$ 或接近 1 是否最优、积分精度、步长敏感性和一阶条件符号变化。

**F2 解答**

无冲击系统为

$$\boldsymbol{z}_{t+1} = A\boldsymbol{z}_t, \quad A = \begin{pmatrix} 0.8 & -0.2 \\ 0.3 & 0.7 \end{pmatrix}.$$

稳定性看 $A$ 的所有特征值是否在单位圆内。迹为 1.5，行列式为

$$0.8 \cdot 0.7 - (-0.2)(0.3) = 0.56 + 0.06 = 0.62.$$

特征方程:

$$\lambda^2 - 1.5\lambda + 0.62 = 0.$$

判别式

$$1.5^2 - 4(0.62) = 2.25 - 2.48 = -0.23 < 0.$$

复根的模为 $\sqrt{\det A} = \sqrt{0.62} \approx 0.787 < 1$，所以系统稳定但有旋转/震荡成分。若 $g_0 = 1$ 后为 0,

$$\boldsymbol{z}_1 = A\boldsymbol{z}_0 + \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

之后

$$\boldsymbol{z}_t = A^{t-1}\boldsymbol{z}_1, \quad t \geq 1.$$

特征值模决定冲击衰减速度，特征值虚部导致震荡式响应。

**F3 解答**

若 $u(c) = \log c$，则 $u'(c) = 1/c$。Euler moment 为

$$\mathbb{E}_t\left[\beta R_{t+1}\frac{c_t}{c_{t+1}} - 1\right] = 0.$$

给定工具变量 $Z_t$ 属于 $t$ 期信息集，有无条件矩:

$$\mathbb{E}\left[Z_t\left(\beta R_{t+1}\frac{c_t}{c_{t+1}} - 1\right)\right] = 0.$$

样本矩条件:

$$\frac{1}{T}\sum_{t=1}^{T} Z_t\left(\beta R_{t+1}\frac{c_t}{c_{t+1}} - 1\right) = 0.$$

工具变量必须在 $t$ 期信息集中，因为 Euler 方程是条件于 $t$ 期信息的正交条件，未来信息不能作为当期决策依据。GMM 正是选择参数使样本矩尽量接近 0。

## 8 最后的复盘清单

**使用说明**

做完 Appendix B 后，你应该能完成以下动作：

1. 看到积分题能判断区域、变量替换和概率解释。
2. 看到动态方程能判断连续/离散、稳态、稳定性和经济含义。
3. 看到优化题能判断无约束、等式约束、不等式约束、KKT 和凸性充分性。
4. 看到 value function 能想到 envelope theorem、Euler equation 或 Bellman equation。
5. 看到回归公式能说出 population projection、条件期望和正交条件。
6. 看到渐近分布能知道 LLN/CLT/Slutsky/Delta method 分别负责哪一步。
7. 看到金融风险题能写出组合方差、正态 VaR/ES，并说明模型风险。

经济金融数学分析补课讲义

From Mathematical Analysis to Economic and Financial Modeling

# Appendix C
# 论文阅读工具包与建模速查表

正文 Ch1-Ch31 已经把工具讲完；Appendix A 整理公式，Appendix B 训练综合题。本附录回答最后一个问题：拿到一篇经济学、金融学、计量或政治经济学论文时，怎样把长篇文字、公式、表格和模型快速还原成可以学习、复述、批判和迁移的结构。本附录不是文献综述写作课，而是一套可执行的 paper reading protocol。

math4econ personalized notes
2026

# 目录

# Appendix C 总览

**使用直觉**

读论文不是把 PDF 从第一页看到最后一页。真正有效的读法是把论文拆成四层：

$$\text{question} \to \text{model/design} \to \text{math/estimation} \to \text{claim and limitation}.$$

你要训练的不是“看懂每一个句子”，而是能回答：作者想解释什么、用什么模型或识别策略、关键数学工具在哪里、结论靠什么条件成立、我能否把这个结构迁移到自己的项目。

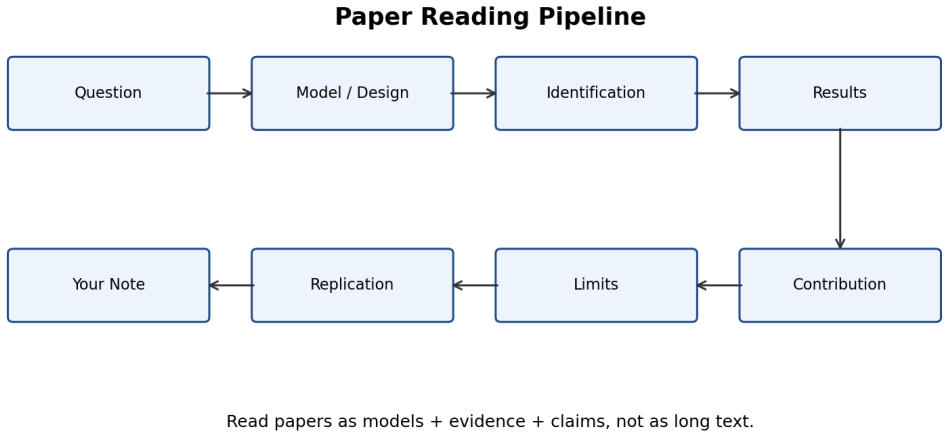



图 1：论文阅读的循环流程：问题、模型/设计、识别、结果、贡献、局限和复现。

| 本附录用途 | 你应该怎样使用 |
|---|---|
| 15 分钟扫读 | 判断一篇论文值不值得深读，定位 question、method、main result。 60 分钟精读 |
| 建立论文结构图，摘出模型、估计式、识别假设和结果表。半天深读 | 复推模型或关键证明，复刻核心回归/图表，写成可发布读书笔记。长期积累 |
| 把每篇论文转成标准化 note，形成自己的研究数据库。写作反向工程 | 学会从优秀论文中反推出 introduction、model、empirical design、robustness 的写法。 |

# 1 三层阅读法：15 分钟、60 分钟、半天

## 1.1 15 分钟扫读：只回答五个问题

第一遍不要试图完全理解。你只需要回答五个问题：

1. 作者研究的核心问题是什么？
2. 论文属于理论、实证、结构估计、实验、金融资产定价、还是方法论文？
3. 主要数学对象是什么：优化、均衡、回归、极限定理、动态规划、随机过程，还是数值模拟？
4. 最重要的结果是哪一个定理、命题、表格或图？
5. 论文的贡献是新问题、新模型、新数据、新识别、新机制，还是新测度？

**操作模板**

**15 分钟扫读顺序**

1. 读标题、摘要、introduction 最后两段。
2. 看所有 section 标题。
3. 找 main theorem、main regression table 或 main figure。
4. 看 conclusion 的第一段和 limitations。
5. 写 5 句话摘要，不允许超过 150 字。

## 1.2 60 分钟精读：建立论文骨架

第二遍开始建立结构。你要把论文压缩到一张“骨架表”。

| 模块 | 要提取的内容 | 对应数学工具 |
| --- | --- | --- |
| Research question | 因果问题、机制问题、定价问题、均衡问题。 | 变量定义、counterfactual、objective。Environment / Data |
| agents、states、sample、market、institution。 | 随机变量、状态空间、样本空间。Model / Design | 优化问题、均衡条件、回归式、实验设计。 |
| KKT、fixed point、projection、identification。Assumptions | 偏好、技术、外生性、单调性、正则条件。 | 凸性、独立性、连续性、rank condition。 Main result |

| | | |
|---|---|---|
| theorem、estimate、calibration result、welfare number。 | FOC、CLT、standard error、comparative statics。 Robustness | 替代样本、替代规格、placebo、sensitivity。 |
| overidentification、bounds、perturbation。Contribution | 相对于已有文献的新点。 | 方法迁移与理论边界。 |

### 1.3 半天深读：复推一个核心

真正学论文，必须复推一个核心。核心可以是：

- 一个命题的证明；
- 一个 Euler 方程；
- 一个 Bellman 方程；
- 一个 identification argument；
- 一个主要回归表；
- 一个 numerical experiment；
- 一个 welfare decomposition。

**常见误区**

最危险的读法是“我感觉懂了”。如果你没有复写出作者的关键式子，没有说清楚每个假设在哪里被用到，没有能独立解释一张表的每一列，你只是熟悉了论文的表面叙述，并没有真正吸收它。

## 2 理论论文阅读工具包

理论论文的难点是公式多、符号多、命题多。读理论论文时，不要从证明细节开始，而要先还原 model skeleton。

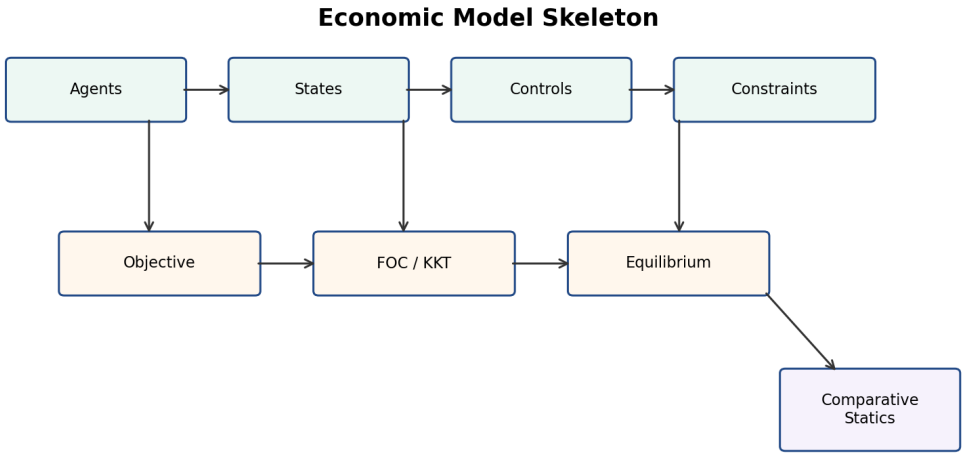



图 2: 理论模型的基本骨架: agents、states、controls、constraints、objective、FOC/KKT、equilibrium 和 comparative statics。

## 2.1 模型骨架表

| 问题 | 你要写下什么 | 常见数学表达 |
|---|---|---|
| Agents | 谁在决策? 家庭、企业、政府、投资者、平台、匹配双方。 | $i \in I$, continuum agents, representative agent。 State variables |
| 哪些变量无法瞬间选择，而是描述系统状态? | $k_t, a_t, z_t, b_t, \theta_t$。 Controls | 决策变量是什么? |
| $c_t, k_{t+1}, x_i, p, q, \pi$。 Objective | 最大化或最小化什么? | $\max \mathbb{E}\sum \beta^t u(c_t)$, $\max \pi(p)$。 Constraints |
| 预算、技术、资源、激励、参与、市场清算。 | $px \le w$, $g(x) \le 0$, $F(k,l)$。 Equilibrium | 谁的最优和哪些清算条件同时成立? |
| fixed point, market clearing, Nash equilibrium。 Comparative statics | 参数变化如何影响选择和均衡? | IFT, envelope, monotone comparative statics。 |

## 2.2 命题阅读五步

看到 Proposition/Theorem 时，按五步处理:

1. **翻译结论：**先用中文说出这个命题到底声称什么。

2. **标出条件：**哪些条件是技术条件，哪些条件有经济含义?
3. **找证明工具：**是凸性、固定点、包络、IFT、single crossing、martingale,还是 contraction?
4. **找最小例子：**写一个 2 变量或 2 状态的简单例子来验证命题直觉。
5. **问边界：**哪个假设去掉后结论可能失败?

**示范**

**示范：读一个一般均衡存在命题**

命题形式可能是：在偏好连续、局部非饱和、凸，消费集非空、闭、凸，并且总资源有限的经济中，竞争均衡存在。

阅读时不要一开始钻证明。先翻译：作者要说明“价格向量和配置可以同时满足消费者最优、企业最优和市场清算”。然后标条件：连续性用于需求对应的好性质；凸性用于 Kakutani 或 Brouwer；非饱和用于预算约束 binding；资源有限用于 compactness。最后问：如果偏好非凸，需求可能跳跃或对应值非凸，固定点定理可能无法直接使用。

## 2.3 常用理论工具速查

| 工具 | 论文中出现的信号 | 你应该问的问题 |
| --- | --- | --- |
| 凸性/凹性 | unique optimum, global optimum, concavity。 | 局部最优为什么是全局最优? 可行集是否凸? KKT |
| inequality constraints, corner solution。 | 哪些约束 active? 互补松弛如何解释? 隐函数定理 | local comparative statics, equilibrium response。 |
| Jacobian 对哪个变量非奇异? 包络定理 | value function derivative, welfare effect。 | 直接效应是什么? 间接选择效应为什么消失? 固定点 |
| equilibrium existence, best response correspondence。 | compact、convex、upper hemicontinuity 是否满足? 压缩映射 | Bellman operator, value iteration。 |
| 折现因子为何保证 contraction? Euler 方程 | intertemporal choice。 | 边际收益和边际成本在相邻期如何相等? HJB/Pontryagin |
| continuous-time dynamic optimization。 | state、control、costate/-value derivative 是什么? | |

# 3 实证论文阅读工具包

实证论文的核心不是“回归表显著吗”，而是：估计对象是什么，识别假设是什么，估计式是否真的对应问题。

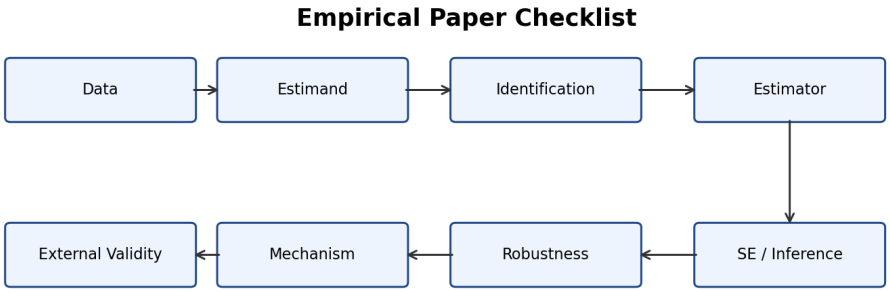



图 3: 实证论文阅读清单：从 data 到 estimand、identification、estimator、inference 和 robustness。

## 3.1 估计对象：先区分 estimand 与 estimator

**核心清单**

**Estimand** 是你想知道的总体对象，例如 $\mathbb{E}[Y(1)-Y(0)]$、$\beta$、ATE、ATT、elasticity、risk premium。
**Estimator** 是你用样本计算出来的东西，例如 $\hat{\beta}_{OLS}$、difference-in-differences estimator、IV estimator、MLE、GMM estimator。

很多论文读不懂，是因为把 estimand 和 estimator 混在一起。你应该先问：作者想估计什么总体对象？然后再问：这个样本方法在什么假设下能估计它？

## 3.2 识别假设表

| 设计 | 核心假设 | 阅读时要检查 |
|---|---|---|
| RCT | 随机分配，处理独立于潜在结果。 | attrition、noncompliance、spillover。 |
| OLS | 条件均值独立 $\mathbb{E}[\varepsilon \mid X]=0$。 | omitted variables、bad controls、functional form。 |
| DID | | parallel trends。 |

| | | |
|---|---|---|
| pre-trends、anticipation、composition change。 IV | relevance + exclusion + monotonicity。 | first stage、排除限制是否可信、LATE 人群。 RDD |
| cutoff 附近连续性。 | manipulation、bandwidth、polynomial、density test。 Event study | 无预趋势，事件时间正确。 |
| dynamic selection、overlapping events。 Panel FE | time-invariant unobservables 可被吸收。 | time-varying confounders、within variation。 |

**常见误区**

标准误显著不等于因果识别可信。实证论文最关键的是识别假设是否可信，而不是星号多少。星号回答的是 sampling uncertainty；识别假设回答的是 bias 是否存在。

### 3.3 回归表阅读模板

拿到一张 regression table，按以下顺序读：

1. 被解释变量是什么？单位是什么？是否标准化？
2. 核心解释变量是什么？处理变量、价格、政策、风险暴露还是交互项？
3. 系数的单位解释是什么？1 unit 变化对应 $Y$ 改变多少？
4. 样本大小如何变化？加入 controls 后样本是否改变？
5. fixed effects 是否改变识别来源？
6. standard errors 如何聚类？聚类层级是否合理？
7. 列与列之间的变化说明什么？
8. 系数大小是否有经济意义，而不只是统计显著？

**操作模板**

**一句话读表模板**

“在样本 sample 中，控制 controls/FE 后，X 增加 1 个单位，与 Y 增加/减少 coefficient 个单位相关；若识别假设 assumption 成立，这可以解释为 causal effect/structural parameter。”

## 4 公式到问题的翻译表

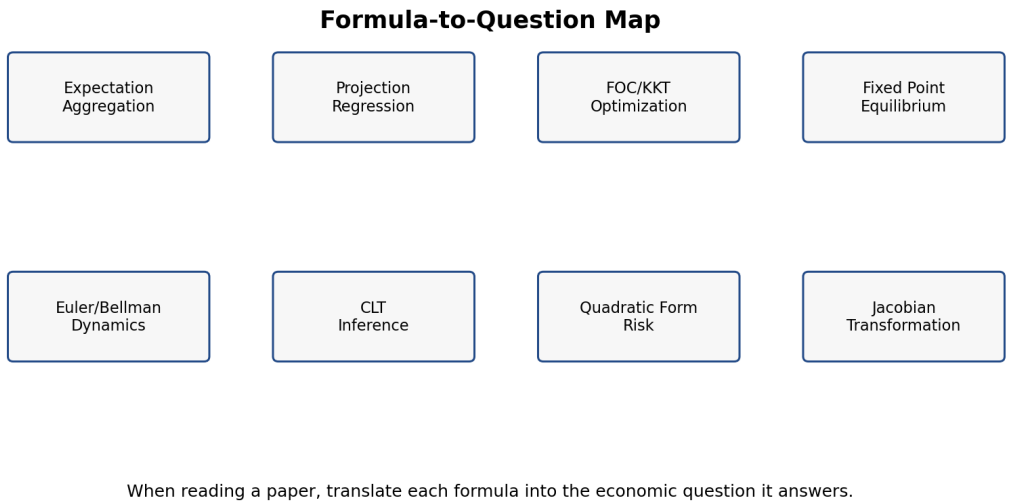



图 4: 看到公式时，不要只识别符号，要把公式翻译成它回答的经济问题。

| 你看到的公式 | 数学含义 | 经济/计量问题 |
|---|---|---|
| $\mathbb{E}[Y \mid X]$ | 条件期望。 | 给定特征时，平均结果是什么? $\mathbb{E}[\varepsilon \mid X] = 0$ |
| 条件外生性。 | 控制 $X$ 后是否还有系统性遗漏? $\mathbb{E}[X\varepsilon] = 0$ | 正交矩条件。 |
| 工具/解释变量是否与误差无关? $\min_b \mathbb{E}(Y - X'b)^2$ | population projection。 | 最佳线性预测是什么? $\nabla f(x^*) = 0$ |
| 一阶条件。 | 边际收益和边际成本在哪里相等? $\lambda_i g_i(x) = 0$ | 互补松弛。 |
| 哪些约束真的限制选择? $F(x, \theta) = 0$ | 隐式均衡。 | 参数变化如何移动均衡? $V(s) = \max_a \{u(s,a) + \beta \mathbb{E} V(s')\}$ |
| Bellman 方程。 | 当前选择如何权衡现在和未来? $p_t = \sum_j R^{-j} \mathbb{E}_t d_{t+j}$ | 资产基本面定价。 |
| 价格是未来现金流折现值吗? $\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, V)$ | 渐近正态。 | 标准误和置信区间为何有效? $w'\Sigma w$ |
| 二次型风险。 | 投资组合方差是多少? | |

# 5 理论模型速查表

## 5.1 消费者理论论文

| 模块 | 速查问题 |
| --- | --- |
| 偏好 | utility 是 cardinal 还是 ordinal？是否单调、凸、可微、可分？预算 |
| 预算约束是否 binding？价格和财富如何进入？需求 | 是 Marshallian 还是 Hicksian？是否使用 Roy 或 Shephard？福利 |
| 使用 EV、CV、consumer surplus 还是 value function change？异质性 | 偏好、收入、价格暴露、信息、约束如何异质？识别/估计 |
| demand system 是否满足 adding-up、homogeneity、symmetry？ | |

## 5.2 生产者与产业组织论文

| 模块 | 速查问题 |
| --- | --- |
| 技术 | production function、cost function、capacity constraint 是什么？市场结构 |
| 完全竞争、垄断、寡头、平台、拍卖、匹配？策略变量 | 价格、数量、质量、广告、进入退出、投资？均衡概念 |
| Nash、subgame perfect、Bayes Nash、competitive equilibrium？识别 | 成本、需求、markups 如何从数据中识别？反事实 |
| merger、tax、subsidy、entry regulation 的 welfare effect？ | |

### 5.3 宏观动态论文

| 模块 | 速查问题 |
| --- | --- |
| 状态变量 | capital、productivity、wealth distribution、debt、beliefs。冲击 |
| TFP、monetary、fiscal、preference、financial shock。家庭问题 | Euler 方程、预算约束、借贷约束、劳动供给。企业问题 |
| 投资、调整成本、融资约束、价格粘性。均衡 | 市场清算、政策规则、aggregate law of motion。数值方法 |
| value iteration、projection、perturbation、simulation、calibration。验证 | moments matched、impulse response、welfare decomposition。 |

## 6 因果识别论文速查

### 6.1 从回归式还原识别问题

看到回归式:

$$Y_{it} = \alpha + \beta D_{it} + X_{it}'\gamma + \mu_i + \lambda_t + \varepsilon_{it},$$

不要只说“这是双向固定效应”。要问:

1. $D_{it}$ 是什么处理? 什么时候、对谁变化?
2. $\beta$ 是 average treatment effect, ATT, 还是加权平均?
3. $\mu_i$ 吸收了什么? $\lambda_t$ 吸收了什么?
4. 剩下的识别变异来自哪里?
5. 误差项中还可能有哪些 time-varying confounders?
6. 标准误是否应该在处理层级聚类?

**示范**

**示范：DID 论文阅读**

若论文研究最低工资提高对就业的影响，DID 回归是

$$Employment_{st} = \alpha + \beta MW_{st} + \mu_s + \lambda_t + X_{st}'\gamma + \varepsilon_{st}.$$

你要把它翻译成：在控制州固定效应和时间固定效应后，最低工资变动州相对于未变动州的就业变化差异。核心识别假设是 parallel trends：如果没有最低工资政策，处理组和对照组就业趋势会相同。阅读重点不是先看 $\beta$ 是否显著，而是检查 pre-trend 图、政策实施是否内生、是否存在跨州 spillover，以及行业结构变化是否同时发生。

### 6.2 因果图式自检

每篇因果论文都可以用三句话自检：

1. 理想实验是什么?
2. 现实设计怎样逼近这个实验?
3. 还有哪条路径让处理变量和潜在结果相关?

## 7 金融论文阅读工具包

金融论文通常分成 asset pricing、corporate finance、market microstructure、household finance、financial econometrics、macro-finance。不同方向看法不同。

### 7.1 资产定价论文

| 模块 | 速查问题 |
| --- | --- |
| Test assets | 用哪些投资组合或单个资产？排序依据是什么？ Factors |
| 因子是 traded 还是 non-traded？经济含义是什么？ Pricing equation | $\mathbb{E}[mR]=1$、beta pricing、SDF、Euler equation？ Anomaly |
| 超额收益是否可被已有因子解释？是否 survives costs？ Inference | Fama-MacBeth、GMM、time-series alpha、clustered SE？ Risk or mispricing |
| 机制是风险补偿、限制套利、行为偏误还是数据挖掘？ | |

### 7.2 公司金融论文

| 模块 | 速查问题 |
| --- | --- |

| | |
|---|---|
| Decision | 投资、融资、分红、并购、治理、创新? Friction |
| 信息不对称、代理问题、税、破产成本、融资约束? Identification | shock 是否外生? firm FE 是否够? 是否有行业/地区时间冲击? Outcome |
| market value、investment、employment、productivity、default? Mechanism | cash flow channel、collateral channel、governance channel? |

### 7.3 金融计量论文

| 模块 | 速查问题 |
|---|---|
| Data frequency | daily、intraday、monthly、quarterly? microstructure noise 是否重要? Process |
| returns、volatility、jump、factor、state-space? Estimator | MLE、GMM、realized volatility、filtering、simulation? Asymptotics |
| $T \to \infty$、$N \to \infty$、high-frequency infill? Risk measure | VaR、ES、drawdown、tail beta、stress test? Validation |
| out-of-sample、backtesting、forecast comparison? | |

## 8 证明阅读与复写模板

### 8.1 证明不是逐字翻译，而是找依赖链

一个证明通常由以下零件组成:

- 定义展开;
- 构造对象;
- 使用引理;
- 验证定理条件;
- 得到存在性、唯一性或不等式;
- 用反证或极限收束。

**操作模板**

**证明复写四栏法**

| Step | 作者做了什么 | 用了哪个假设/定理 | 经济含义 |
|---|---|---|---|
| 1 | 展开定义 | definition | 这个对象是什么 2 |
| 构造函数/对应 | compactness/convexity | 为均衡存在准备 3 | 验证条件 |
| continuity/UHC | 最优反应好性质 4 | 应用定理 | Brouwer/Kakutani |
| 均衡存在 5 | 解释结论 | theorem statement | 模型有解 |

## 8.2 常见证明关键词

| 关键词 | 它通常意味着什么 |
|---|---|
| WLOG | without loss of generality, 作者在利用对称性或归一化。It suffices to show |
| 证明被化简为一个更小的目标。 By continuity | 要么用中间值定理, 要么用极限交换或闭图。By compactness |
| 要么取收敛子列，要么保证最大值存在。 By convexity | 局部到全局、一阶条件充分、混合策略/平均仍可行。 By envelope theorem |
| 价值函数导数只看直接效应。 By law of iterated expectations | 条件期望嵌套被消去。 By Slutsky |
| 极限分布中替换一致估计量。 | |

# 9 论文笔记模板

## 9.1 一页纸 note

**可复制模板**

**Paper One-Page Note**

1. Citation:
2. Research question in one sentence:
3. Why it matters:

4. Environment / data:
5. Main model or empirical design:
6. Key equation / theorem / table:
7. Identification or main assumptions:
8. Main result, with magnitude:
9. Mechanism:
10. Limitations:
11. What I learned mathematically:
12. One idea I can borrow:

## 9.2 深度 note

**可复制模板**

**Deep Reading Note**

A. Setup

- Agents:
- State variables:
- Controls:
- Constraints:
- Shocks/information:

B. Mathematical core

- Optimization/equilibrium problem:
- FOC/KKT/Euler/Bellman equation:
- Key theorem and assumptions:
- Comparative statics:

C. Empirical core

- Estimand:
- Estimator:
- Identification assumption:
- Standard errors:
- Robustness checks:

D. Replication plan

- Data needed:

- Variables to construct:
- Main table/figure to reproduce:
- Minimum code skeleton:

E. My evaluation

- Strongest part:
- Weakest assumption:
- Extension idea:

## 10 从论文到公众号/读书笔记

你曾经提到想每天读一篇 econ、社会学、政治学方向的 top paper 并写到公众号。这个附录给一个最稳妥的结构：既适合学习，也适合公开表达。

**操作模板**

**公众号读书笔记结构**

1. 开头 150 字：这篇论文问了什么，为什么重要。
2. 背景：用普通语言解释问题，不急着放公式。
3. 方法：讲模型或识别设计，用一张图/一个式子说明。
4. 结果：只讲 2-3 个最重要结果，必须带量级。
5. 数学工具：解释一个值得学习的公式或定理。
6. 评价：贡献、局限和可以继续做什么。
7. 你的 takeaway：这篇论文怎样改变你看问题的方式。

**常见误区**

公开写作时不要堆过多术语。读书笔记不是论文摘要，也不是机械翻译。最好的写法是用你的语言重构论文逻辑：问题为什么难，作者怎样绕开难点，数学工具在哪里起作用，结论有什么边界。

## 11 Agent/AI 辅助读论文的安全用法

AI 可以帮你加速，但不能替代你判断。推荐把 AI 用在以下任务：

- 提取论文结构；
- 解释符号和公式；

- 把证明拆成步骤；
- 生成复现计划；
- 检查你的读书笔记是否遗漏识别假设；
- 把表格结果翻译成自然语言。

不要把 AI 用在以下任务：

- 未核对原文就相信摘要；
- 让 AI 代替你判断识别是否可信；
- 不看数据和表格就写评价；
- 直接复制 AI 的文献综述。

**可复制模板**

**AI 读论文 Prompt 模板**

请按以下结构帮我读这篇论文，但不要替我下结论。请分别列出原文依据：

1. Research question;
2. Model or empirical design;
3. Key assumptions;
4. Main equations and what each symbol means;
5. Main result with magnitude;
6. Identification threats or theoretical weak points;
7. What math tools from my notes are used;
8. Three questions I should verify by reading the paper myself.

## 12 从数学工具到论文类型的映射

| 讲义章节 | 常见论文位置 | 阅读重点 |
| --- | --- | --- |
| 多重积分/Jacobian | 分布、期望、似然、变量变换。 | 积分区域、密度变换、期望对象。含参积分 |
| welfare、likelihood、score。 | 是否可以积分下求导，边界项是否存在。投影/Fourier | OLS、factor、basis expansion、signal processing。 |
| 投影空间、正交条件、频率解释。ODE/差分/系统 | 宏观动态、资产价格、时间序列。 | 稳态、稳定性、特征根。优化/KKT/凸性 |

| | | |
|---|---|---|
| 微观理论、机制设计、金融组合。 | FOC、角点、充分性、shadow price。 IFT/Envelope | 比较静态、福利分解、结构模型。 |
| 直接效应与间接效应, Jacobian 非奇异。固定点 | 一般均衡、Nash、动态规划。 | compact/convex/UHC 或 contraction。 Bellman/HJB/control |
| 宏观、金融、产业动态。 | state、control、value、policy、TVC。概率/条件期望 | 计量、因果推断、风险模型。 |
| estimand、exogeneity、projection。 极限定理 | 推断、standard error、GMM、MLE。 | 收敛类型、渐近方差、Slutsky。多元正态/二次型 |
| 金融风险、回归检验、PCA。 | covariance matrix、test statistic、risk decomposition。 Monte Carlo/bootstrap | 计算经济、simulation、inference。 |
| simulation error、resampling scheme、coverage。 | | |

# 13 综合训练任务

## Task A: 15 分钟扫读训练

任选一篇论文，只允许 15 分钟，完成以下卡片：

1. 研究问题：
2. 方法类型：理论/实证/结构/实验/金融/方法：
3. 核心数学工具：
4. 主结果：
5. 我是否值得继续深读，为什么：

## Task B: 模型骨架复原

对一篇理论论文, 填表：agents、states、controls、constraints、objective、equilibrium、comparative statics。然后选一个命题，写出它用到的三个数学工具。

### Task C: 回归表复述

任选一张主回归表，用自然语言复述每一列。要求说明：样本、controls、FE、standard errors、核心系数、经济量级、识别假设。

### Task D: 识别假设攻击

对一篇因果论文，写出三个可能威胁识别的机制。每个机制必须说明：它如何同时影响 treatment 和 outcome；作者是否处理；你会增加什么 robustness check。

### Task E: 公式翻译训练

从论文中抄下 5 个公式。每个公式必须写：

1. 符号含义；
2. 数学工具；
3. 经济含义；
4. 这个公式在论文论证中起什么作用。

## 14 本附录最终总结

**使用直觉**

整套讲义的终点不是记住公式，而是形成一种读论文的能力：

看到模型，知道它的数学骨架；看到回归，知道它的识别对象；看到定理，知道假设和结论的依赖链；看到

如果你能把每篇论文都压缩成 question、tool、identification、result、limitation、extension 六个部分，你就已经从“读者”开始接近“研究者”。

| 最后检查 | 你读完一篇论文后应该能回答 |
| --- | --- |
| 一句话问题 | 作者到底想解释什么？核心工具 |
| 用了哪一个数学/计量/金融工具？关键假设 | 结论最依赖哪个假设？主结果 |
| 结果方向、量级和显著性是什么？识别/机制 | 为什么这个结果可以被解释为作者声称的机制？局限 |
| 最可能失败的地方在哪里？迁移 | 这个模型、识别或写法能否用于你的项目？ |