2025 年数理经济学笔记

授课: 杨佳楠老师

作者：徐靖

组织：PKU

时间：Febuary 28, 2025

Pkuhub.cn

个人笔记, 如有谬误, 欢迎指正! 联系方式: 2200012917@stu.pku.edu.cn

目录

第 1 章 Linear Algebra[1]

内容提要

- Leading Principal Minor : 顺序主子式
- Orthogonal matrix : 正交矩阵
- Symmetric matrix : 对称矩阵
- Positive definite matrix : 正定矩阵
- Positive semi-definite matrix : 半正定矩阵
- Determinant : 行列式

定义 1.1

For $N \times N$ matrix $A = (a_{ij})$, using any row or column:

$$\det A = \sum_{i=1}^{N} (-1)^{i+j} a_{ij} \det A_{ij}$$

where A_{ij} is the $(N-1) \times (N-1)$ matrix obtained by deleting the i-th row and j-th column of A. ♣

定理 1.1

$$A^{-1} = \frac{1}{\det A} \tilde{A}$$

where $\tilde{a_{mn}} = (-1)^{m+n} \det A_{nm}$. ♡

定义 1.2

Orthogonal matrix : $P^T P = I$.

Symmetric matrix : $A^T = A$.

Positive definite matrix : $x^T Ax > 0$ for all $x \neq 0$.

Positive semi-definite matrix : $x^T Ax \geq 0$ for all x. ♣

定义 1.3

Leading Principal Minor : determinant of the first $k \times k$ submatrix of A. For real symmetric matrix A, A is positive definite if and only if all its leading principal minors are positive. ♣

定义 1.4

$$Av = \lambda v$$

where v is **eigenvector**, λ is **eigenvalue**. λ is a root of the **characteristic polynomial** $\det(A - \lambda I) = 0$. ♣

定义 1.5

Complex inner product :

$$\langle x, y \rangle = x^* y = \sum_{i=1}^{n} \bar{x}_i y_i$$

where $\bar{x}$ is the **complex conjugate** and x^* is the **conjugate transpose** (adjoint). ♣

---

[1] 只记一些矩阵分解吧, 以防忘了

定义 1.6

Hermitian matrix : $A^* = A$.

For real matrices, Hermitian matrix is symmetric.

For Hermitian matrix, all eigenvalues are real.

定理 1.2 (Diagonalization of Symmetric Matrices)

$$P^T AP = \text{diag}\{\lambda_1, \dots, \lambda_n\}$$

where P is orthogonal matrix, λ_i are eigenvalues of A.

第 2 章 Topology of $\mathbb{R}^N$[1]

Keywords

- Topology 拓扑
- Metric Space 度量空间
- Convergence 收敛
- interior 内部
- closure 闭包
- boundary 边界
- compact set 紧集
- cluster point 聚点
- Lipschitz continuity 利普希茨连续
- semicontinuity 半连续
- Bolzano-Weierstrass Theorem 博尔扎诺-魏尔斯特拉斯定理
- Heine-Borel Theorem 海涅-波雷尔定理
- Contraction Mapping Theorem 压缩映射定理
- Intermediate Value Theorem 中值定理

2.1 Metric Spaces

2.1.1 Definition of Metric Spaces

定义 2.1

Let X be a set. A function $d : X \times X \to \mathbb{R}$ is called a **metric** (or **distance**) on X if :

1. (positivity) $d(x, y) \geq 0$ for all $x, y \in X$ and $d(x, y) = 0$ if and only if $x = y$.
2. (symmetry) $d(x, y) = d(y, x)$ for all $x, y \in X$.
3. (triangle inequality) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$.

♣

A set X together with a metric d is called a **metric space**, denoted by (X, d).

2.1.2 Examples of metrics in $\mathbb{R}^N$

- **Euclidean metric**: $d(x, y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}$.
- L^p **metric** (for $p \geq 1$): $d(x, y) = (\sum_{i=1}^N |x_i - y_i|^p)^{1/p}$.
- **Sup norm** (when $p = \infty$): $d(x, y) = \max_{i=1}^N |x_i - y_i|$.

2.2 Convergence of sequences

2.2.1 Definition of Convergence

定义 2.2

Let (X, d) be a metric space. A sequence $\{x_n\}$ in X is said to **converge** to a point $x \in X$ if for every $\epsilon > 0$, there exists an integer N such that $d(x_n, x) < \epsilon$ for all $n \geq N$. In this case, we write $\lim_{n\to\infty} x_n = x$. A sequence that converges is called **convergent**, otherwise it is called **divergent**.

♣

[1]点集拓扑对应数分高代这一级别的数学基础课, 只有 sms 和图班的同学学过, 所以我简单记一下, 无需关注证明, 数理经济学只用到结论

定义 2.3

When metric space is $\mathbb{R}^N$, we say that $\{x_n\}$ is **bounded** if there exists a real number M such that $\|x_k\| \leq M$ for all n. ♣

2.2.2 Cauchy Sequences and Complete Metric Spaces

- **Cauchy sequence**: A sequence $\{x_n\}$ in a metric space (X, d) is called a **Cauchy sequence** if for every $\epsilon > 0$, there exists an integer N such that $d(x_n, x_m) < \epsilon$ for all $n, m \geq N$.
- **Complete metric space**: A metric space (X, d) is called **complete** if every Cauchy sequence in X converges to a point in X.

定理 2.1

Any convergent sequence in a metric space is a Cauchy sequence. ♡

2.2.3 Example: Cauchy Sequence Not Convergent in $\mathbb{Q}$

Consider the metric space $(\mathbb{Q}, d)$, where $d(x, y) = |x - y|$.

Fibonacci sequence : Let $\{F_k\}$ be the Fibonacci sequence, defined by

$$F_1 = F_2 = 1, F_{k+1} = F_k + F_{k-1}, k \geq 2$$

A Special Sequence: Define $a_k = \frac{F_{k+1}}{F_k}$. Then $\{a_k\}$ is a Cauchy sequence in $\mathbb{Q}$ but does not converge in $\mathbb{Q}$.

2.2.4 Properties of Convergent Sequences in $\mathbb{R}^N$

Consider $\mathbb{R}^N$ with the Euclidean metric. Let $\{x_n\}$ and $\{y_n\}$ be two sequences.

- **Preservation of Addition/Subtraction**: If $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} y_n = y$, then $\lim_{n\to\infty}(x_n \pm y_n) = x \pm y$.
- **Preservation of Multiplication**: If $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} y_n = y$, then $\lim_{n\to\infty}(x_n \cdot y_n) = x \cdot y$.
- **Preservation of Division**: If $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} y_n = y \neq 0$, then $\lim_{n\to\infty} \frac{x_n}{y_n} = \frac{x}{y}$.
- **Preservation of Inequality**: If $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} y_n = y$, then $x_n \leq y_n$ for all n implies $x \leq y$.

2.2.5 Properties of Sequences in $\mathbb{R}^N$

性质 A convergent sequence in $\mathbb{R}^N$ is bounded.

A sequence $\{x_{n_k}\}$ is called a **subsequence** of $\{x_n\}$ if $n_1 < n_2 < n_3 < \cdots$.

性质 subsequences of a convergent sequence in $\mathbb{R}^N$ also converge to the same limit.

2.2.6 Limit Superior and Limit Inferior

定义 2.4

Let $\{x_n\}$ be a sequence in $\mathbb{R}^N$. The **limit superior** of $\{x_n\}$ is defined by

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty}\left(\sup_{k\geq n} x_k\right)$$

The **limit inferior** of $\{x_n\}$ is defined by

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty}\left(\inf_{k\geq n} x_k\right)$$

♣

2.3 Topological properties

2.3.1 Open and Closed Sets

定义 2.5

In a metric space (X, d), a set $U \subset X$ is called **open** if for every $x \in U$, there exists an $\epsilon > 0$ such that $B(x, \epsilon) \subset U$. A set $F \subset X$ is called **closed** if its complement $F^c \stackrel{\text{def}}{=} X \backslash F$ is open.

性质 For open sets:

1. The union of any collection of open sets is open.
2. The intersection of finitely many open sets is open.

For closed sets:

1. The intersection of any collection of closed sets is closed.
2. The union of finitely many closed sets is closed.

2.3.2 Interior, Closure, and Boundary of Sets

定义 2.6

The **interior** of a set $A \subset X$ is defined as:

$$\text{int}(A) = \bigcup \{U \subset A : U \text{ is open}\}$$

The **closure** of a set $A \subset X$ is defined as:

$$\overline{A} = \bigcap \{F \supset A : F \text{ is closed}\}$$

The **boundary** of a set $A \subset X$ is defined as:

$$\partial A = \overline{A} \backslash \text{int}(A)$$

命题 2.1

- $A \subset X$ is open if and only if $\partial A \subset A$.
- $A \subset X$ is closed if and only if $\partial A \subset A$.

2.3.3 Bounded Sets and Compact Sets in $\mathbb{R}^N$

定义 2.7

A set $A \subset \mathbb{R}^N$ is called **bounded** if there exists a real number M such that $\|x\| \leq M$ for all $x \in A$.
A set $A \subset \mathbb{R}^N$ is called **compact** if for any sequence $\{x_n\}$ in A, there exists a subsequence $\{x_{n_k}\}$ that converges to a point in A.

定理 2.2 (Heine-Borel Theorem)

In $\mathbb{R}^N$, a set A is compact if and only if it is closed and bounded.

2.4 Continuous functions

2.4.1 Cluster Points in Metric Spaces

定义 2.8

Let (X, d) be a metric space and $A \subset X$. A point $x \in X$ is called a **cluster point** of A if for every $\epsilon > 0$, there exists a point $y \in A$ such that $d(x, y) < \epsilon$ and $x \neq y$.

Equivalently, x is a cluster point of A if there exists a sequence $\{x_n\}$ in A such that $\lim_{n\to\infty} x_n = x$ and $x_n \neq x$ for all n.

♣

2.4.2 Limits of Functions at Cluster Points

定义 2.9

Let (X, d) and (Y, ρ) be metric spaces, $A \subset X$, $f : A \to Y$, and x be a cluster point of A. We say that f has a **limit** $y \in Y$ at x if for every $\epsilon > 0$, there exists a $\delta > 0$ such that $\rho(f(x_0), y) < \epsilon$ for all $x_0 \in A$ such that $0 < d(x_0, x) < \delta$.

Equivalently, using neighborhoods: f has a limit y at x if for every neighborhood V of y, there exists a neighborhood U of x such that $f(U \cap A) \subset V$.

♣

性质

1. $\lim_{x\to\bar{x}} f(x) = f(\bar{x})$ if and only if for every sequence $\{x_n\}$ in A such that $\lim_{n\to\infty} x_n = \bar{x}$, we have $\lim_{n\to\infty} f(x_n) = f(\bar{x})$.
2. If f has a limit at x, then the limit is unique.

2.4.3 Continuity of Functions

定义 2.10

Let (X, d) and (Y, ρ) be metric spaces, and $f : X \to Y$.

- f is **continuous at** $\bar{x} \in X$ if:
$$\forall \varepsilon > 0, \exists \delta > 0 : \forall x \in X, d(x, \bar{x}) < \delta \Rightarrow \rho(f(x), f(\bar{x})) < \varepsilon$$
Equivalently:
$$\forall \varepsilon > 0, \exists \delta > 0 : f(B_\delta(\bar{x})) \subseteq B_\varepsilon(f(\bar{x}))$$
- f is **continuous on** X (or simply **continuous**) if:
$$\forall \bar{x} \in X, f \text{ is continuous at } \bar{x}$$

♣

命题 2.2

Let (X, d) and (Y, ρ) be metric spaces, $f : X \to Y$, and $x \in X$. The following are equivalent:

1. f is continuous at x.
2. For every sequence $\{x_n\}$ in X such that $\lim_{n\to\infty} x_n = x$, we have $\lim_{n\to\infty} f(x_n) = f(x)$.
3. For every open set $V \subset Y$, $f^{-1}(V)$ is open in X.

♠

2.4.4 Bolzano-Weierstrass Theorem

定理 2.3 (Bolzano-Weierstrass Theorem)

If $K \subset \mathbb{R}^N$ is compact and nonempty, and $f : K \to \mathbb{R}^M$ is continuous, then :

1. $f(K)$ is compact.
2. f attains its maximum and minimum on K.

♡

2.4.5 Semicontinuity

定义 2.11

For $f : \mathbb{R}^N \to \mathbb{R}^M$:

- f is **upper semicontinuous** at x if :

$$f(x) \leq \limsup_{y \to x} f(y) \text{ for all } x \in \mathbb{R}^N$$

- f is **lower semicontinuous** at x if :

$$f(x) \geq \liminf_{y \to x} f(y) \text{ for all } x \in \mathbb{R}^N$$

♣

注 f is upper semicontinuous $\Leftrightarrow -f$ is lower semicontinuous.

定理 2.4 (Extrema of semicontinuous Functions)

Let $K \subset \mathbb{R}^N$ be compact and $f : K \to \mathbb{R}$ be upper semicontinuous. Then f attains its maximum on K. If f is lower semicontinuous, then f attains its minimum on K.

♡

2.4.6 Lipschitz Continuity

定义 2.12

A function $f : \mathbb{R}^N \to \mathbb{R}^M$ is called **Lipschitz continuous** if there exists a constant $K > 0$ such that:

$$\|f(x) - f(y)\| \leq K\|x - y\| \text{ for all } x, y \in \mathbb{R}^N$$

where K is called the **Lipschitz constant** of f. If $K < 1$, then f is called a **contraction mapping**.

♣

注 Lipschitz continuity implies uniform continuity, but the converse is not true. For example, $f(x) = x^2$.

定理 2.5 (Contraction Mapping Theorem)

Let (X, d) be a complete metric space and $f : X \to X$ be a contraction mapping. Then f has a unique fixed point $x^* \in X$, i.e., $f(x^*) = x^*$.

♡

定理 2.6 (Intermediate Value Theorem)

Let $f : D \to \mathbb{R}$ be a continuous function and $D \subset \mathbb{R}$. If :

- $[a, b] \subset D$ (closed interval)
- y is between $f(a)$ and $f(b)$

then there exists a point $c \in [a, b]$ such that $f(c) = y$.

♡

第 3 章 Multi-Variable Calculus[1]

Keywords

- ❏ gradient 梯度

3.1 introduction

3.1.1 Motivation and Insight

- Many practical problems involve optimization with multple variables.
- Real-world applications often require optimizing several variables simultaneoutly.
- Linear functions are easy to understand and manipulate.
    - Not all interesting functions are linear, but many can be approximated by linear functions.
    - The gradient is a generalization of the derivative to functions of multiple variables.

[1] 多元微分还能有不会的吗, 这真不用记了吧

第 4 章 Multi-Variable Unconstrained Optimization

Keywords

- ❑ First Order Condition 一阶条件
- ❑ Bisection Method 二分法
- ❑ Secant Method 割线法
- ❑ False Position Method 假位法
- ❑ Newton's Method 牛顿法

4.1 First Order Condition

An Unconstrained Optimization Problem is :

$$\min_{x \in \mathbb{R}^n} f(x)$$

定义 4.1

First Order Condition (FOC): $\nabla f(x^*) = 0$.

- x^* is a **stationary point** (驻点) of f.
- It is necessary but not sufficient.

global minimum: $f(x^*) \le f(x)$ for all $x \in \mathbb{R}^n$.

local minimum: $f(x^*) \le f(x)$ for all $x \in B(x^*, \epsilon)$ for some $\epsilon > 0$.

♣

命题 4.1

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function, and $\nabla f(x^*) = 0$. If $\nabla^2 f(x^*)$ is:

- positive definite, then x^* is a local minimum.
- negative definite, then x^* is a local maximum.
- indefinite, then x^* is a **saddle point**. (鞍点)

♠

4.2 Convex Optimization

定义 4.2 (Convex Function)

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if for all $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$:

$$f(\lambda x + (1 - \lambda)y) \le \lambda f(x) + (1 - \lambda)f(y)$$

♣

定理 4.1

A twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its Hessian matrix $\nabla^2 f(x)$ is positive semidefinite for all $x \in \mathbb{R}^n$.

♡

命题 4.2

Let f be differentiable. Then f is (strictly) convex if and only if:

$$f(y) - f(x)(>) \ge \nabla f(x) \cdot (y - x)$$

for all $x, y \in \mathbb{R}^n$.

♠

定理 4.2 (Minimum/maximum Characterization)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex (concave) function. Then x^* is a local minimum (maximum) if and only if:

$$\nabla f(x^*) = 0$$

- If f is strictly **convex**, then x^* is a global minimum.
- If f is strictly **concave**, then x^* is a global maximum.

♡

4.3 Numerical Optimization

4.3.1 Bisection Method

定义 4.3 (Bisection Method)

A simple, robust method for finding roots of continuous functions on bounded intervals

- Start with an interval $[a, b]$ such that $f(a)f(b) < 0$.
- Compute the midpoint $c = \frac{a+b}{2}$, and evaluate $f(c)$.
- Replace a or b with c based on the sign of $f(c)$.
- Iterate until desired precision.

♣

定义 4.4 (Convergence Rate and Order 收敛速度和阶)

For iteration x_n approaching the root r, the convergence rate C and order ρ are defined as:

$$\lim_{n\to\infty} \frac{|x_{n+1} - r|}{|x_n - r|^\rho} = C$$

- Linear convergence: $\rho = 1$, $C < 1$.
- Quadratic convergence: $\rho = 2$, $C < 1$.
- Superlinear convergence: $\rho > 1$, $C < 1$.

♣

Method	Definition	Rate	Order
Bisection	Iteratively bisects an interval and selects a subinterval	Linear ($C = 0.5$)	1
Secant	Root approximation via secant line through two points	Superlinear ($C \approx 1.618$)	1.618
False Position	Bisection variant with linear interpolation updates	Linear	1
Newton-Raphson	Derivative-based iterative root-finding	Quadratic ($C \propto f''$)	2
Gradient method	Function minimization via negative gradient direction	Linear ($C \propto \kappa$)	1

表 4.1: Compact Comparison of Numerical Methods

定义 4.5 (Methods)

Secant Method:

- Compute the secant line through $(x_0, f(x_0))$ and $(x_1, f(x_1))$.
- Find the intersection with the x-axis to get the next approximation x_2.
- Iterate until convergence.

$$x_{n+1} = x_n - \frac{f(x_n)(x_n - x_{n-1})}{f(x_n) - f(x_{n-1})} \tag{4.1}$$

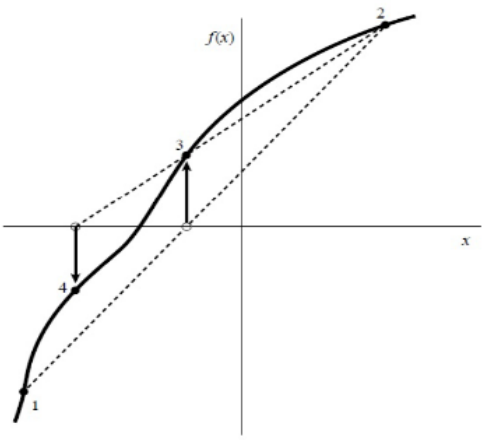

图 4.1: Secant Method

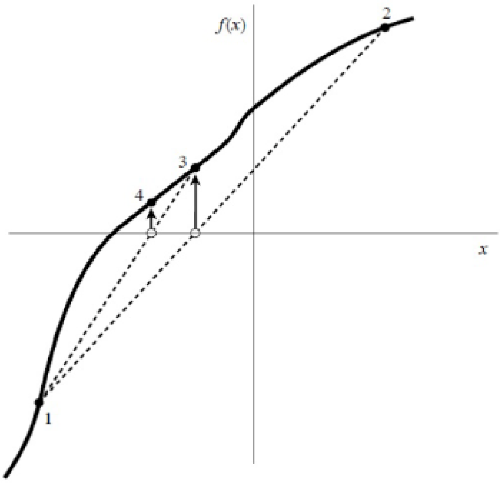

图 4.2: False Position Method

False Position Method:

- Similar to the secant method, but always keeps the interval $[a, b]$ such that $f(a)f(b) < 0$.
- Update a or b based on the sign of $f(c)$.
- Iterate until convergence.

$$c = \frac{af(b) - bf(a)}{f(b) - f(a)},$$
$$[a, b] \leftarrow [a, c] \quad \text{if} \quad f(a)f(c) < 0,$$
$$[a, b] \leftarrow [c, b] \quad \text{if} \quad f(b)f(c) < 0.$$

♣

笔记 若初始值足够接近根且函数光滑, 则 Secant Method 收敛速度优于 False Position Method, 但可能因迭代点跳出根的邻域而发散. False Position Method 保证收敛, 但多一个异号的初始条件且速度较慢.

定义 4.6 (Newton-Raphson Method)

- Start with an initial guess x_0.
- Compute the next approximation using the formula:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

- Iterate until convergence.

♣

问题 4.1 为什么牛顿法是二阶收敛的?

解 对 $f(x)$ 在 x_n 处做泰勒展开, 对于 $f(r) = 0$:

$$f(r) = f(x_n) + f'(x_n)(r - x_n) + \frac{f''(x_n)}{2}(r - x_n)^2 + O((x - x_n)^3)$$

带入牛顿法迭代公式 $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$:

$$x_{n+1} - r = (x_n - r)^2 \cdot \frac{f''(x_n)}{2f'(x_n)}$$

笔记 牛顿法初期可能出问题, 如果不满足足够接近根的假设.

牛顿法可以很好地应用到多变量上, 但过程中 Hessian 矩阵的逆矩阵计算量较大, 并且他是一个 local method.

例题 4.1 将牛顿法应用到求解二次可微函数的极值问题, 可以求解 first order condition:

$$x_{n+1} = x_n - [\nabla^2 f(x_n)]^{-1} \nabla f(x_n)$$

定义 4.7 (Gradient Method)

- Start with an initial guess x_0 and error tolerance ϵ.
- Iterate until $\|x_{n+1} - x_n\| < \epsilon$:
  - Compute the gradient $\nabla f(x_n)$.
  - Define $\phi(t) = f(x_n - t\nabla f(x_n))$.
  - Find the minimum of $\phi(t)$ using a one-variable optimization method (e.g., bisection, secant, or Newton's method).
  - Defint $x_{n+1} = x_n - t^*\nabla f(x_n)$.

♣

第 5 章 Multi-Variable Optimization with Equality Constraints

Keywords

- Equality Constraints 等式约束
- Lagrange Multiplier 拉格朗日乘数法
- Nondegenerate Constraint Qualification, NDCQ 非退化约束条件
- Cobb-Douglas Utility Function 柯布-道格拉斯效用函数

笔记 现在我们考虑有约束条件的优化问题,这一关键是将约束视为函数方程并引入拉格朗日乘数法.

定义 5.1 (Optimization with Equality Constraints)

设 $f(x)$ 是可微函数, $g(x)=0$ 是可微约束条件组, 那么我们要优化的问题可以表示为:

$$\max f(x) \quad s.t. \quad g(x)=0 \tag{5.1}$$

设 f 是 n 维向量, g 是 m 维向量, x 是 k 维向量.

定义 5.2 (Lagrange Multiplier)

对于上述问题, 我们可以构造拉格朗日函数:

$$L(x,\lambda)=f(x)+\lambda^T g(x) \tag{5.2}$$

其中 λ 是拉格朗日乘数. 通过对 L 求导数, 我们可以得到一组方程:

$$\frac{\partial L}{\partial x}=0, \quad \frac{\partial L}{\partial \lambda}=g(x)=0 \tag{5.3}$$

其解 (x^*,λ^*) 就是我们要找的最优解.

定义 5.3 (NDCQ)

如果 $g(x)$ 在 x^* 处可微, 且 $Dg(x^*)$ 的秩为 m, 那么我们称 $g(x)$ 满足非退化约束条件 **(NDCQ)**. 其中,

$$Dg(x^*)=\begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_k} \end{pmatrix}$$

定理 5.1 (Lagrange Multiplier Theorem)

设 $f(x)$ 和 $g(x)$ 都是可微函数, 且 $g(x)$ 满足非退化约束条件. 那么 (x^*,λ^*) 是上述优化问题的最优解.

笔记 我们需要进一步判断最大值还是最小值.

定义 5.4 (Borderde Hessian Matrix)

$$H=\begin{pmatrix} 0 & Dg(x^*) \\ Dg(x^*)^T & D^2f(x^*) \end{pmatrix} \tag{5.4}$$

其中 $D^2f(x^*)$ 是 $f(x)$ 在 x^* 处的 Hessian 矩阵, $Dg(x^*)$ 是 $g(x)$ 在 x^* 处的 Jacobian 矩阵.
本质是求 Lagrange 函数的 Hessian 矩阵, 它是 $k+m$ 维的.

定理 5.2 (Sufficient Condition for Maximum)

设 H 是上述的 Borderde Hessian 矩阵, 那么如果顺序主子式交替符号, 且 $\det H$ 的符号与 $(-1)^m$ 相同, 则 (x^*, λ^*) 是局部极大值, 若顺序主子式恒为负, 则 (x^*, λ^*) 是局部极小值. ♡

笔记 这里我们不再用正定性来判断, 因为加边海色矩阵的左上 k 阶主子式是为 0.

第 6 章 Comparative Statics and Envelope Theorem

Keywords

- Generalized Comparative Statics 广义比较静态
- Cramer's Rule 克拉默法则
- Envelope Theorem 包络定理
- Shephard’ s Lemma 谢泼德引理

6.1 Comparative Statics

笔记 经济学中的比较静态分析是指在给定一个经济模型的情况下, 研究当 **exogenous variables** 发生变化时, **endogenous variables** 的值如何变化. 比如当给定消费者收入去讨论市场供需模型中的均衡价格, 给定税率去讨论对 Monopoly 的影响, 前者作为外生变量 (经济模型的输入), 后者作为内生变量 (经济模型的输出).

定义 6.1 (Generalized Comparative Statics)

We have an economic model, the equilibrium solution of which is given by the form:

$$F(x^*, \alpha) = 0$$

where x^* is the equilibrium solution of endogenous variables x, and α is a vector of exogenous variables. The key objective is to find the derivative $\frac{\partial x_i^*}{\partial \alpha_j}$ and identify its sign.

笔记 这里向量函数 f 的个数应当与内生变量的个数相同, 设为 n.

定理 6.1 (Cramer's Rule)

Let $F(x^*(\alpha), \alpha) = 0$ be a system of n equations in n unknowns. The Jacobian matrix of the system is given by:

$$\det J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

We have:

$$J\frac{\partial x^*}{\partial \alpha_j} + \frac{\partial F}{\partial \alpha_j} = 0, \forall j$$

The Cramer's Rule states that the derivative of the equilibrium solution with respect to the exogenous variable α_j is given by:

$$\frac{\partial x_i^*}{\partial \alpha_j} = -\frac{\det J_{ij}}{\det J}$$

where J_{ij} is the matrix obtained by replacing the i-th column of J with the vector $\frac{\partial F}{\partial \alpha_j}$.

6.1.1 Comparative Statics for Unconstrained Optimization

这时我们考虑一个最优化问题, 其形式为:

$$\max_x f(x; a)$$

其中 $f : \mathbb{R}^n \times R^m \to \mathbb{R}$ 是一个可微函数[1]. m 以下视为 1 (不考虑外生变量之间的影响).

[1] 请注意 $\nabla f = F$: 这里 $\nabla f = 0$ 是 1.1.1 小节中函数的 FOC, F 是广义比较静态的模型函数, 我们不是第一次用类似的记号

FOC $\nabla f = 0$ 可以这样写:

$$\frac{\partial f}{\partial x_i}(x_1^*, \ldots, x_n^*; a) = 0, \forall i$$

$f(\cdot)$ 的 FOC 的 Jacobian 矩阵也就是 $f(\cdot)$ 的 Hessian 矩阵:

$$\det J(x^*; a) = \frac{\partial^2 f}{\partial x^2}(x^*; a)$$

命题 6.1 (Implict Function Theorem)

If $\det J(x^*; a) \neq 0$, then the system implicitly defines differentiable functions:

$$x_i^* : a \to x_i^*(a), \forall i$$

And the derivatives of these functions are given by[a]:

$$\frac{\partial x_i^*}{\partial a} = -\frac{\det J_i(x^*; a)}{\det J(x^*; a)}$$

where J_i is the matrix obtained by replacing the i-th column of J with the vector $\frac{\partial f}{\partial a}$.

[a] 这个式子和上面的 Cramer's Rule 的结论是一样的, 区别是在优化问题中, 我们将向量函数指定为被优化函数的梯度, 那么 Jacobian 矩阵成为了一个特例, 也就是 Hessian 矩阵.

笔记 隐函数定理揭示了二阶条件与比较静态的存在性的关联

6.1.2 Comparative Statics for Equality Constrained Optimization

在有约束优化中, 我们先做拉格朗日再用同样的隐函数定理方法来进行比较分析, 实际上还是一样的, 因为要对 Lagrangian 函数求一阶条件.

$$\begin{aligned}
&L = f + \lambda h \\
&\frac{\partial L}{\partial \lambda}(x^*, \lambda^*; a) = h(x^*; a) = 0 \\
&\frac{\partial L}{\partial x_i}(x^*, \lambda^*; a) = \frac{\partial f}{\partial x_i}(x^*; a) + \lambda^* \frac{\partial h}{\partial x_i}(x^*; a) = 0 \quad \text{for } i = 1, 2, ..., n
\end{aligned}$$

Jacobian 是 Bordered Hessian 矩阵:

$$\det J_L(x^*; a) = \begin{bmatrix} 0 & \frac{\partial h}{\partial x}(x^*; a) \\ \frac{\partial h}{\partial x}(x^*; a) & \frac{\partial^2 f}{\partial x^2}(x^*; a) \end{bmatrix}$$

利用隐函数定理:

$$\frac{\partial x_i^*}{\partial a} = -\frac{\det J_i(x^*; a)}{\det J(x^*; a)}$$

其中 J_i 是通过将 J_L 的第 i 列替换为 $\frac{\partial L}{\partial a}$ 得到的矩阵.

6.2 Envelope Theorem

还是之前的最优化问题:

$$V(a) = \max_x f(x; a)$$

- $F : \mathbb{R}^n \times R^m \to \mathbb{R}$ is differentiable.
- a is an exogenous variable.
- $x^*(a)$ is a local solution with differentiable components $x_i^*(a) : \mathbb{R} \to \mathbb{R}$.

定理 6.2 (Envelope Theorem)

The derivative of the value function with respect to the exogenous variable a is given by:

$$\begin{aligned}\frac{\partial V(a)}{\partial a} &= \frac{\partial f}{\partial a}(x^*(a);a) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^*(a);a)\frac{\partial x_i^*(a)}{\partial a} \\ &= \underbrace{\frac{\partial f}{\partial a}}_{\text{直接影响}} + \underbrace{\nabla_x f \cdot \frac{\partial x^*(a)}{\partial a}}_{\text{间接影响}}\end{aligned}$$

where $\nabla_x f$ is the gradient of the objective function with respect to the endogenous variables, and $\frac{\partial x^*(a)}{\partial a}$ is the derivative of the equilibrium solution with respect to the exogenous variable a. ♡

笔记 包络定理揭示了: 当外生参数变化时, 只需考虑该参数的直接影响, 而无需额外计算内生变量调整[2]带来的间接影响. 因为由一阶条件, $\nabla_x f$ 项在最优解处趋于零.

对于有约束优化也一样:

$$\frac{\partial V(a)}{\partial a} = \frac{\partial f}{\partial a} + \sum_{i=1}^{n} \lambda_i \frac{\partial g}{\partial a}$$

包络定理这一名称源于它在成本曲线上的应用:

- 长期总成本曲线是短期总成本曲线的包络线
- 短期总成本曲线是通过改变固定投入水平而产生的
- 包络定理为这种关系提供了严格的数学基础

命题 6.2 (Shephard’s Lemma)

当要素的价格上涨 1 单位时, 最小成本的边际增加量正好等于厂商对该要素的使用量.

$$\frac{\partial C(w_1,w_2,y)}{\partial w_i} = x_i(w_1,w_2,y)$$

♠

[2]如果是有约束优化, λ 的选取也包含在内

第 7 章 Multi-Variable Optimization with Inequality Constraints

Keywords

- Complementary slackness 互补松弛条件
- Karush-Kuhn-Tucker Theorem KKT 条件
- Constraints Qualification 约束资格条件
- Convex Optimization 凸优化
- Nondegenerate Constraint Qualification NDCQ 非退化约束资格条件
- Slater Condition SCQ 斯莱特条件

7.1 Optimization with linear inequality constraints

7.1.1 Geometric intuition for FOC

考虑一个优化问题:

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & Ax \leq b \\ & x \geq 0 \end{aligned}$$

我们的目标是去找到 x^* 作为局部最小的必要条件.

定义 7.1 (Constraint Set)

$$\Omega = \{x \in \mathbb{R}^n | Ax \leq b\}$$

其中 A 是 $m \times n$ 的矩阵, b 是 m 维向量. ♣

先考虑一条直线作为约束, 即 $a^T x \leq c$.

- 如果 $a^T x^* < c$, 那么满足 $\nabla f(x) = 0$ 的 x^* 是局部最优
- 如果 $a^T x^* = c$, 那么 x^* 是局部最优的充分必要条件是 $\nabla f(x^*)$ 和 a 线性相关.

这是因为对任意可行的方向 d, 有:

$$0 \leq \lim_{t\to 0} \frac{f(x^* + td) - f(x^*)}{t} = \langle \nabla f(x^*), d\rangle$$

这意味着 $\nabla f(x^*)$ 和 d 线性相关, 实际上

$$\nabla f(x^*) + \lambda a = 0, \lambda \geq 0$$

定义 7.2 (feasible direction)

d 是可行的方向, 如果 $x^* + td \in \Omega$ 对任意小的 $t > 0$ 都成立. ♣

考虑两条直线的约束, 对于其交点 x^*, 其成为局部最优的直观的必要条件是:

$$\nabla f(x^*) + \lambda_1 a_1 + \lambda_2 a_2 = 0, \lambda_1, \lambda_2 \geq 0$$

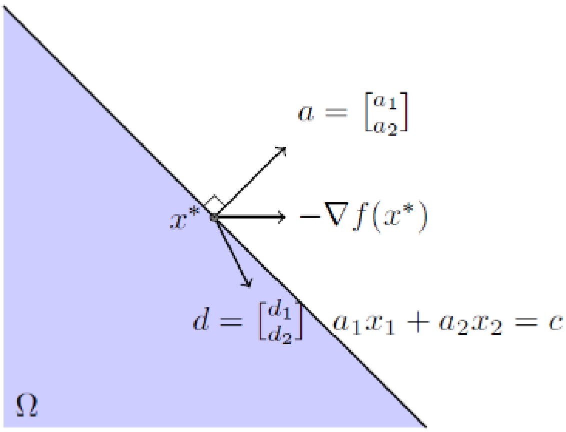

图 7.1: $-\nabla f$ 只能延 a 的方向

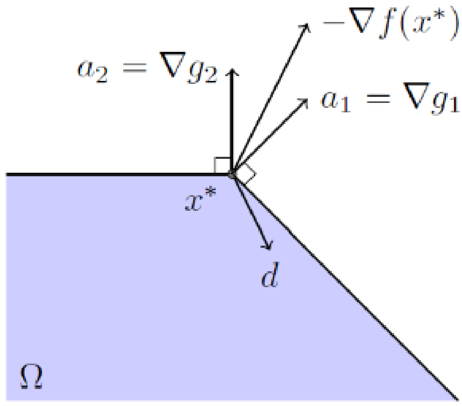

图 7.2: $-\nabla f$ 可以延 a_1, a_2 所夹任意方向

7.1.2 KKT

因此, 对于线性约束的优化问题, 我们给出一个相对统一严谨的必要条件:

定理 7.1 (Modified Karush-Kuhn-Tucker Theorem)

对于优化问题:

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0, i \in [I] \\ & h_j(x) = 0, j \in [J] \end{aligned}$$

其中, f 可微, $g_i = \langle a_i, x\rangle - c_i, h_j = \langle b_j, x\rangle - d_j$ 均为线性函数, $a_i, b_j \neq 0$. 如果 x^* 是局部最优解, 那么存在 λ_i 和 μ_j 使得:

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^{I} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{J} \mu_j \nabla h_j(x^*) = 0 \\ \lambda_i \ge 0, g_i(x^*) \le 0, \lambda_i g_i(x^*) = 0 \\ h_j(x^*) = 0 \end{aligned}$$

♡

笔记 KTT 条件可以由拉格朗日乘子法得到, 其中三个条件分别对应于: 一阶条件, 互补松弛, 等式约束.

7.2 Optimization with nonlinear inequality constraints

7.2.1 General KKT

线性约束的优化问题 (线性规划) 是简单的, 非线性是坏的性质. 一个直观的想法是用 Taylor 展开近似线性约束.

$$g_i(x) \approx g_i(x^*) + \langle \nabla g_i(x^*), x - x^* \rangle$$

因此我们引入一类条件, 统称约束资格条件 constraint qualigication (CQ). 使之代替原本的 g,h 线性条件. (本质是作近似) 然后重写 KTT [1]:

定理 7.2 (General Karush-Kuhn-Tucker Theorem)

对于优化问题:

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0, i \in [I] \\ & h_j(x) = 0, j \in [J] \end{aligned}$$

其中, f 可微, x^* 是局部最优解, 且约束资格条件 CQ 成立, 那么存在 λ_i 和 μ_j 使得:

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^{I} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{J} \mu_j \nabla h_j(x^*) = 0 \\ \lambda_i \ge 0, g_i(x^*) \le 0, \lambda_i g_i(x^*) = 0 \\ h_j(x^*) = 0 \end{aligned}$$

♡

7.2.2 Constraint Qualification

课堂上给出了两种:

定义 7.3 (Nondegenerate CQ (NDCQ))

令 $I(x^*)$ 为 x^* 积极约束的指标集 (set of active constraint indices), 即

$$I(x^*) = \{i | g_i(x^*) = 0\}$$

NDCQ : 活跃约束的梯度线性无关, 即:

$$\text{rank}\left(\nabla g_i(x^*), i \in I(x^*)\right) = |I(x^*)|$$

♣

笔记 线性无关的意义在于, 保证了 λ 的唯一性. 且确保 ∇f 一定在这些梯度张成的空间. 几何上可以理解为, 可行方向 (f 的梯度方向) 和起作用的/活跃的约束梯度方向正交

定义 7.4 (Slater Condition(SCQ))

- g_i' 是凸的
- 存在 x_0 使得 $g_i(x_0) < 0, i \in [I]$, 即 x_0 是严格可行的 (strictly feasible).

♣

7.3 Convex Optimization

应用到凸优化问题上, 我们可以得到更强的结论.

定义 7.5 (Convex Optimization (无等式约束版))

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0, i \in [I] \end{aligned}$$

[1] 实际上, 课堂所展示的 General KKT 忽略了等式约束的函数组 h, CQ 条件也忽略. (从 g 扩充到 h 是显然的, 例如 NDCQ 中 h 自然在 $I(x^*)$ 中, 但 g 不一定在 $I(x^*)$ 中.)

其中 f, g_i' 可微且凸.

命题 7.1

Necessity：如果 x^* 是局部最优解, 且存在 x_0 使得 $g_i(x_0) < 0, i \in [I]$, 那么存在拉格朗日乘子 λ_i 使得 KKT 条件成立.

Sufficiency：如果存在拉格朗日乘子 λ_i 使得 KKT 条件成立. 则 x^* 是全局最优解.

笔记 必要性即以 SCQ 为假设的 KTT. 充分性证明不难: x^* 最小化 $L(x, \lambda)$, 从而对任意可行的 x, 有 $f(x^*) = L(x^*, \lambda) \leq L(x, \lambda) \leq f(x)$.

理论成立, 做题可以三步走:

1. Verify Condition
   - f, g_i' 可微且凸吗
   - SCQ 成立吗
2. Form Lagrangian $L(x, \lambda)$
   - First-order condition
   - Complementary slackness
3. Solve System
   - Solve for x^*, λ^*
   - x^* 是原问题解, λ^* 提供灵敏度信息

第 8 章 From Thick to Thin

笔者的一些 insights, 便于理解和记忆数理经济学的体系, 也是考前的复习提纲

期中部分

前三章 介绍了前置数学工具, 包括线性代数, 欧式空间 (的拓扑性质), 以及多元微分. 其中相对核心的有:

- 正定矩阵及其判定, 对角化的一系列结论
- Bolzano-Weierstrass Theorem
- Hessian, Jacobian, lagrange

1. 极值与零点

- 无约束优化: $f:\mathbb{R}^n\to\mathbb{R}$, 求 x^* 使 $f(x^*)$ 最小
- 函数组零点: $F:\mathbb{R}^n\to\mathbb{R}^n$, 求 x^* 使 $F(x^*)=0$

以上两个问题的交集在于 FOC:

$$F=\nabla f=0$$

笔记 换句话说, 优化问题求解的第一步是零点问题, 零点问题的一些特例可以还原成优化问题 (显然不是所有的方程组都是某个函数的 FOC), 假如 $F=\nabla f$, 有:

- f 的 Hessian 矩阵是 F 的 Jacobian 矩阵, 他们都是对称矩阵

$$\begin{pmatrix}\frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1\partial x_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial^2 f}{\partial x_n\partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2}\end{pmatrix}=\begin{pmatrix}\frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial F_n}{\partial x_1} & \cdots & \frac{\partial F_n}{\partial x_n}\end{pmatrix}$$

1.1 零点求解方法

第四章 介绍了五种方法

- **Bisection**: 二分法, $x_{n+1}=\frac{x_n+x_{n-1}}{2}$
- **Secant**: 割线法, $x_{n+1}=x\text{axis}\cap\text{line}(x_n,x_{n-1})$
- **False Position**: 假位法, 是割线法的变种, 每次都在 x_n 和 x_{n-1} 之间取一个点与 x_{n+1} 构成新区间, 保证区间含根
- **Newton Raphson**: 牛顿法, $x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}$
- **Gradient Descent**: 梯度下降法, $x_{n+1}=x_n-t\nabla f(x_n)$, 其中 $t=\arg\min_{t>0}f(x_n-t\nabla f(x_n))$

1.2 极值点判定

规范方法: 判断 Hessian 矩阵的正定性, 本课程用顺序主子式法[1]来判断会比较快

H_f	H_i	x^*
正定	$H_i>0$	极小值
负定	$H_i<0$	极大值
不定	$H_i\neq 0$, 但不属于以上两种	鞍点

实用技巧:

[1] 正小负大

- 正定矩阵的特征值都是正数，负定矩阵的特征值都是负数
- 先代入临界点排除明显非极值情况
- 结合 f 的凹凸性判断，如果严格凸或者严格凹，则临界点一定是极小值或者极大值

2. 无约束优化与约束优化

约束优化是第五章的内容，即在可微函数组 $g(x)=0$ 上求 f 的极值，沟通二者的桥梁是拉格拉日乘数法与 **NDCQ**

$$L(x,\lambda)=f(x)+\lambda^\top g(x)$$

然后 (x^*,λ^*) 是一阶条件 $\nabla L(x^*,\lambda^*)=0$ 的解，我们似乎回到了无约束优化

2.1 非退化约束条件

NDCQ (非退化约束条件) 是让拉格朗日乘数法成立的条件

定理 8.1

假设 x 是 k 维向量，λ 是 m 维向量，则 L 是 $k+m$ 维向量，考虑

$$Dg(x^*)=\begin{pmatrix}\frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_k}\\ \vdots & \ddots & \vdots\\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_k}\end{pmatrix}$$

NDCQ: $\operatorname{rank}(Dg(x^*))=m$

- 假如 **NDCQ** 得到满足，则 (x^*,λ^*) 可能是极值点
- 假如不满足，即约束 g 是退化的，此时可能仍然存在极值点，但无法由拉格朗日乘数法得到

♡

2.2 加边海色矩阵

加边海色矩阵是判断有约束优化问题极值点的工具：

$$\bar{H}=\begin{pmatrix}0 & Dg(x^*)\\ Dg(x^*)^\top & H_f(x^*)\end{pmatrix}$$

加边后显然不正定了，判断方法[2]也有变化：

$\bar{H}$	$H_i, i\in[2m+1,n+m]$	x^*
	$\operatorname{sgn}H_i=(-1)^m$	极小值
	$\operatorname{sgn}H_i=(-1)^{i-m}$	极大值
	$\operatorname{sgn}H_i\neq 0$，但不属于以上两种	鞍点

笔记 对加边海色矩阵的一个直观理解[3]：假如 f 在 $g=0$ 这个 $n-m$ 维的嵌入 $\mathbb{R}^n$ 的流形上表现出了类似正定负定的性质，那么可以判断是极值还是鞍点，比如我们在 3 维空间中的球面上找极值.

做题时如果从几何解释出发，可以避免犯错

[2] 判断方法和完整证明参考Northwestern University 的笔记. 实际上，课堂上只讲了 $m=1$ 的情形，在此情形下极小值对应主子式恒负，极大值对应符号交替，这样就不对称，和无约束优化问题的海色矩阵对比起来显得不自然. 经济学中最简单的 $m=1,n=2$ 的情形，自然只需看 $|\bar{H}|$ 的符号，正大负小

[3] 这个直观理解来自于 CMU 的笔记

3. 外生变量

第六章 的内容探究外生变量的作用. 一是探究外生变量 (作为经济模型的输入) 对内生变量 (经济模型的输出) 的影响 (偏导数), 二是外生变量对最优值函数 $V = \max f$ 的影响 (包络定理).

打个比方, 企业的生产存在外部因素 (劳动力价格, 资本价格等), 决策得到的产量, 最终的成本/收益. 这三者分别对应外生变量, 内生变量, 最优值函数.

3.1 对内生变量的影响

无论无约束还是有约束, 对于多元函数 f (或 $L = f + \lambda g$), 和方程组 $F = \nabla f = 0$, 对外生变量 α 求导有:

$$\sum_{i=1}^{n} \frac{\partial F_k}{\partial x_i} \frac{\partial x_i}{\partial \alpha} + \frac{\partial F_k}{\partial \alpha} = 0$$

整理得:

$$J \frac{\partial x^*}{\partial \alpha} + \frac{\partial F}{\partial \alpha} = 0$$

其中 J 是 f 的 Hessian 矩阵 (Bordered Hessian 矩阵), 也是 F 的 Jacobian 矩阵. 规定 J_i 是用 F 取代原本第 i 列后的矩阵. 利用隐函数定理和克拉默法则, 我们可以得到:

$$\frac{\partial x_i^*}{\partial \alpha} = -\frac{\det J_i}{\det J}$$

3.1 对最优值函数的影响

外生变量对最优值函数的影响, 包含直接影响和通过内生变量作用的间接影响两部分. 而包络定理揭示了后者在最优解处趋于零 (因为 $\nabla f = 0$), 因此只需考虑前者.

$$\frac{\partial V(a)}{\partial a} = \underbrace{\frac{\partial f}{\partial a}}_{\text{直接影响}} + \underbrace{\nabla_x f \cdot \frac{\partial x^*(a)}{\partial a}}_{\text{间接影响}}$$

形象的经济学理解比如 Shephard's Lemma : 当要素的价格上涨 1 单位时, 最小成本的边际增加量正好等于厂商对该要素的使用量.

期末部分