第七讲 二阶条件

范翻

中央财经大学 (CCFD)

二阶性质

- 一阶导数：函数值在某点附近的变化情况；
- 二阶导数：函数一阶导数的变化情况，即函数值在某点附近的曲率。

考察目标函数和约束函数在最优值点附近的曲率，如果曲率满足一定条件，则最大值会是充分的。

1 无约束最大化问题

2 有约束最大化问题

无约束最大化问题

在某一点 $\bar{x}$ 附近对函数 $F(x)$ 进行泰勒展开:

$$F(x) = F(\bar{x}) + F^{'}(\bar{x})(x-\bar{x}) + \frac{1}{2}F^{''}(\bar{x})(x-\bar{x})^2 + \cdots$$

最优化的一阶必要条件是 $F^{'}(\bar{x}) = 0$，因此

$$F(x) - F(\bar{x}) = \frac{1}{2}F^{''}(\bar{x})(x-\bar{x})^2 + \cdots$$

对于足够接近于 $\bar{x}$ 的 x 而言，二阶项就支配了泰勒展开中的高阶项。因此，如果 $F^{''}(\bar{x})$ 为正，我们就可以找到一个足够接近于 $\bar{x}$) 的 x，使得 $F(x) > F(\bar{x})$。换句话说，$\bar{x}$ 是 $F(x)$ 的局部或全局最大值的二阶必要条件是

$$F^{''}(\bar{x}) \leq 0.$$

无约束最大化问题

如果 $F''(\bar{x})$ 严格为负，那么在 $\bar{x}$ 附近足够小的区间内，不管高阶项的符号如何，我们都有 $F(x) < F(\bar{x})$。因此

$$F''(\bar{x}) < 0.$$

是 $\bar{x}$ 产生 $F(x)$ 的一个局部最大值的二阶充分条件。

二阶必要条件和二阶充分条件之间存在两个不同之处：

- 前者是一个弱的不等式，而后者是相应的严格不等式；
- 前者是局部或全局最大值的一个必要条件，而后者仅是局部最大值的一个充分条件。

含参数的无约束最大化问题

假定最大化问题包含一个参数 θ。则一阶条件为：

$$F_x(\bar{x}, \theta) = 0.$$

我们希望知道最优选择如何随着 θ 的变动而变化，将上式全微分可得

$$F_{xx}(\bar{x}, \theta)d\bar{x} + F_{x\theta}(\bar{x}, \theta)d\theta = 0.$$

或

$$d\bar{x}/d\theta = -F_{x\theta}(\bar{x}, \theta)/F_{xx}(\bar{x}, \theta).$$

在最优解处，等式右边分母为负。因此 $d\bar{x}/d\theta$ 的符号与 $F_{x\theta}$ 在最优解处的符号相同。

比较静态分析

考虑选择变量为向量的最优化情形，泰勒展开可以写作：

$$F(x) = F(\bar{x}) + F_x(\bar{x})(x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T F_{xx}(\bar{x})(x - \bar{x}) + \cdots$$

此时 F_{xx} 是由二阶偏导数 $F_{jk} \equiv \partial^2 F / \partial x_j \partial x_k$ 组成的对称方阵。上标 T 代表矩阵的转置。二阶项此时是二次型：

$$(x - \bar{x})^T F_{xx}(\bar{x})(x - \bar{x}) = \sum_{j=1}^{n} \sum_{k=1}^{n} F_{jk}(\bar{x})(x_j - \bar{x}_j)(x_k - \bar{x}_k).$$

二次型与凸性

R^n 上的一个二次型是一个定义在 R^n 上的函数，表达式为 $y^T My$，其中 M 是一个对称矩阵，矩阵 M 称为关于二次型的矩阵（二次型矩阵）：

- 如果对于所有的 $y \neq 0$，二次型 $y^T My$ 的值均为负的，那么该二次型被称为负定的；如果二次型 $y^T My$ 的值均为非正的，那么该二次型被称为半负定的。
- 如果对称矩阵 M 的 k 阶主子式 M_k，即由任意的 k 行和 k 列元素组成的子矩阵而言，都有 $(-1)^k |M_k| \geq 0$ 。
- 如果 $F_{xx}(\bar{x})$ 是半负定的二次型矩阵，那么 F 在 $\bar{x}$ 处是凹的。
- 最大化问题的二阶充分条件等价于 $(x-\bar{x})^T F_{xx}(\bar{x})(x-\bar{x})$ 是负定的；二阶必要条件等价于 $(x-\bar{x})^T F_{xx}(\bar{x})(x-\bar{x})$ 是半负定的。

比较静态分析

对一阶条件 $F_x(\bar{x}, \theta) = 0$ 全微分

$$F_{xx}(\bar{x}, \theta) d\bar{x} + F_{x\theta}(\bar{x}, \theta) d\theta = 0.$$

此时，$d\bar{x}$ 和 $d\theta$ 均是向量，F_{xx} 和 $F_{x\theta}$ 均是矩阵。$d\bar{x}$ 的解是

$$d\bar{x} = -F_{xx}(\bar{x}, \theta)^{-1} F_{x\theta}(\bar{x}, \theta) d\theta$$

二阶条件的使用

考虑一个企业在 w 的价格下购买了投入向量 x，生产了产出 $y = f(x)$，最后将其出售获得收入 $R(y)$。
它的利润可以表示为选择变量 x 和投入价格（参数）w 的函数

$$F(x, w) = R(f(x)) - wx$$

试找出 w 的变动对最优选择 x 的影响?

二阶条件的使用

假设最优选择 x 是参数 w 的函数

$$x = x(w)$$

目标函数可以改写为 $F(x(w), w)$,

1 无约束最大化问题

2 有约束最大化问题

约束最优化

考虑两个选择变量和一个等式约束的最优化问题，即在约束 $G(x_1, x_2) = c$ 下最大化 $F(x_1, x_2)$，其中 F 和 G 都是自变量的增函数。把 x_2 视作沿着每一条 F 的等值线上关于 x_1 的函数：

$$dx_2/dx_1 = -F_1(x_1, x_2)/F_2(x_1, x_2).$$

x_2 作为 x_1 的函数，将上式再次微分有

$$\begin{aligned}\frac{d^2x_2}{dx_1^2} &= \frac{d[-F_1/F_2]}{dx_1}\\ &= -\frac{F_2(F_{11} + F_{12}dx_2/dx_1) - F_1(F_{21} + F_{22}dx_2/dx_1)}{F_2^2}\\ &= -\frac{F_2^2F_{11} - 2F_1F_2F_{12} + F_1^2F_{22}}{F_2^3}\end{aligned}$$

约束最优化

类似的表达式也可以沿着约束曲线求二阶导数得到，$\bar{x}$ 为局部最优解的二阶充分条件为 d_2^x/dx_1^2 沿着 F 的等值线的值应该比它沿着 G 的等值线的值更大。利用一阶必要条件

$$F_j(\bar{x}) = \lambda G_j(\bar{x}), j = 1, 2$$

化简可得

$$G_2^2(F_{11} - \lambda G_{11}) - 2G_1G_2(F_{12} - \lambda G_{12}) + G_1^2(F_{22} - \lambda G_{22}) < 0.$$

其矩阵形式为:

$$det \begin{bmatrix} F_{11} - \lambda G_{11} & F_{12} - \lambda G_{12} & -G_1 \\ F_{21} - \lambda G_{21} & F_{22} - \lambda G_{22} & -G_2 \\ -G_1 & -G_2 & 0 \end{bmatrix} > 0$$

约束最优化

在上述问题中的函数 F 和 G 中加入一个 s 维的参数向量 θ，那么一阶条件为

$$F_x(\bar{x},\theta)-\lambda G_x(\bar{x},\theta)=0,\ G(\bar{x},\theta)=0$$

对一阶条件全微分有

$$\begin{aligned}&\sum_{k=1}^{n}(\partial^2 F/\partial x_j\partial x_k)d\bar{x}_k+\sum_{r=1}^{s}(\partial^2 F/\partial x_j\partial\theta_r)d\theta_r\\&-\sum_{i=1}^{m}\lambda_i\{\sum_{k=1}^{n}(\partial^2 G/\partial x_j\partial x_k)d\bar{x}_k+\sum_{r=1}^{s}(\partial^2 G/\partial x_j\partial\theta_r)d\theta_r\}\\&-\sum_{i=1}^{m}d\lambda_i\partial G^i\partial x_j=0\end{aligned}$$

约束最优化

上式可以用矩阵表示为:

$$\begin{bmatrix} F_{xx}-\lambda G_{xx} & -G_x^T \\ -G_x & 0 \end{bmatrix}\begin{bmatrix} d\bar{x} \\ d\lambda^T \end{bmatrix} = -\begin{bmatrix} F_{x\theta}-\lambda G_{x\theta} \\ -G_\theta \end{bmatrix}$$