A Brief Lecture Notes on Optimization for Microeconomic Analysis[1]

冯　曲

复旦大学经济学院
中国经济研究中心

第一稿 2003 年 1 月

一切匠心都归功于 Dixit, Chiang, Takayama 等杰出的前辈；所有的错误、遗漏为编者所有。

他是个顽皮的孩子，不肯就安分的待在身旁；我们要做的是，不能让他的玩耍脱离我们的视线。

[1] 欢迎指出错误及评论。qufeng@fudan.edu.cn

目　　录

A.最优规划问题：

$$\max_x f(x)$$
$$s.t.$$
$$g(x)=c$$

这样一个规划问题可以用来表达一个在一定资源约束情况下的经济决策问题，其中 $f(x)$ 称为目标函数，x 为选择变量，$g(x)$ 为约束函数。如果用集合 S ＝$\{x|g(x)=c\}$表示约束，则此规划问题可以看作是在集合 S（可以看作是欧氏空间的一个子集，称为可行集）上选择一个点 x（或向量），使得目标函数 f(x)的值最大。如下图，在二维情形下，S 表示可行集，$f(x)=k$ 表示目标函数的等值线，则上述规划问题就变成了在 S 中找一点，使得目标函数的等值线达到一个最高的位置。

从这样的角度看规划问题，我们可以将研究的重点放在目标函数的性质和可行集的性质两个部分。

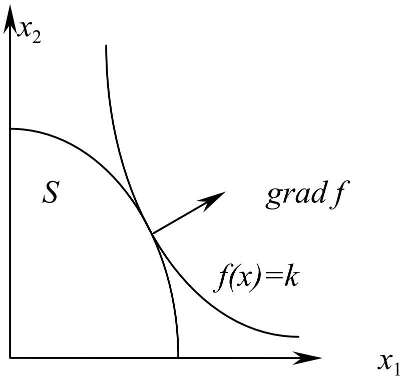


图 A.1

B.梯度向量

梯度 $grad\ f$ 的概念一般和目标函数的等值线（面）的变化有关。我们分两种情况来从梯度的角度看最优化问题。

B.1.无约束极值问题：

一维情况下：

$$df(x)=f^{'}(x)\cdot dx$$

此时，目标函数值的变化，是两部分的乘积，第一部分是导数，第二部分是 x 的微分，或者说是 x 的变化，可以取正也可以取负。

若 $f'(x)>0$，则可以取 $dx>0$，使得 $df(x)>0$，即通过点 x 的移动，可以使目标函数值增加；若 $f'(x)<0$，则可以取 $dx<0$，从而 $df(x)>0$，目标函数

值还可以继续增加。因此，当 $f(x)$取得极值，即目标函数值不再增加时，上述情形不可能发生，即有 $f'(x)=0$。

含义：一维无约束极值问题中，目标函数的梯度就是导数，表示目标函数值变化（增加）的方向。

多维情形下：

$$df(x)=f_x\cdot dx$$

此时，f_x即为$f(x)$在点x处的梯度向量，为横向量（$f_1,\cdots,f_n$），其中每一个分量为偏导数；dx为纵向量，表示 x 的变化。上式表示目标函数$f(x)$的变化可以用梯度向量和dx的<u>内积</u>来衡量：

$$df(x)=f_x\cdot dx=(f_1,\cdots,f_n)\begin{pmatrix}dx_1\\ \vdots\\ dx_n\end{pmatrix}=f_1dx_1+\cdots+f_ndx_n$$

B.1.1 几何意义：向量的内积

x,y 是两个向量，其内积定义为：

$$x\cdot y=(x_1,\cdots,x_n)\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}=x_1y_1+\cdots+x_ny_n$$

同时，也可以表示成：$x\cdot y=|x||y|\cos\alpha$，或$\cos\alpha=\dfrac{x\cdot y}{|x||y|}$，其中$\alpha$表示向量 x,y 之间的夹角，$|x|,|y|$分别表示向量 x,y 的模（原点到点 x、y 的距离）。

如果 $0\le\alpha<\dfrac{\pi}{2}$，则$x\cdot y>0$；

$\alpha=\dfrac{\pi}{2}$，则$x\cdot y=0$，即向量 x,y 正交（垂直）；

$\dfrac{\pi}{2}<\alpha\le\pi$，则$x\cdot y<0$

现在分析梯度向量f_x和dx的几何意义：

dx代表欧氏空间中一点 x 微小变化的向量，方向可以是前后左右上下，取决于每一个分量变化的大小。具体的，如下图，以二维为例，$dx=(dx_1,dx_2)$的分量分别表示从点 x 出发横轴和纵轴变化的方向，向量dx则表示从点 x 变化的整体方向，具体的，满足平行四边形法则。

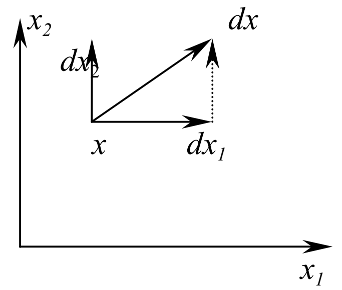


图 B.1

与无约束问题相比，存在约束时，现在可以选择的点（可行集）不再是整个欧氏空间，而是由具体的约束条件构成的欧式空间的一个子集。因此，此时从点 x 移动的方向不再是任意的，而只能在这个可行集内移动。可以想象一下：在一个没有围墙的校园内，可以向任何方向行走；如果有了围墙，则在围墙的附近就不能朝任意方向了，最多只能沿着围墙走。如在消费者选择的问题（两个商品的情形），其约束满足：

$$p_1x_1 + p_2x_2 = I,\quad x_1 \ge 0, x_2 \ge 0$$

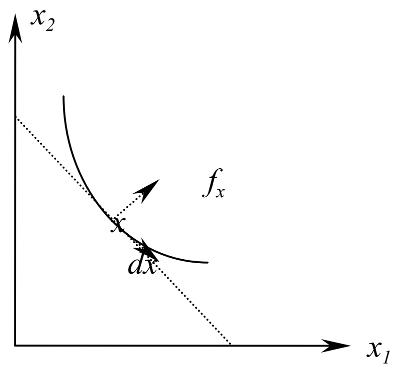


图 B.2

预算线如同围墙一般，在其附近就只能沿其方向移动。

从代数上看，我们也可以得到上述直观的理解，对上式两边全微分，并且 p_1, p_2, I 是参数，保持不变，得：

$$p_1dx_1 + p_2dx_2 = 0$$

与无约束相比，dx_1, dx_2 的大小不是任意的，而是相互联系的，即确定了 dx_1，则 dx_2 也就相应确定了（一般的，相互联系的机制有约束条件决定）。在这个例子中，dx_2 和 dx_1 的比例就等于预算线的斜率，作为平行四边形对角线的 dx 的方向也就确定了，就是沿预算线移动。

梯度向量 f_x 的几何意义是：等值面 $f(x) = k$ 在点 x 处指向值增加（变化）的法方向，也就是在点 x 处 $f(x)$ 值增加最快的方向，其变化率为 f_x 的模。如上

图，无差异曲线（二维情形下，等值面就是无差异曲线）上点 x 处作切线（多维情形下就是切平面），f_x 就是和切线垂直的、无差异曲线增加的向量。

由上述的讨论，对于 $df(x)=f_x \cdot dx$，如果梯度向量 f_x 和 dx 的方向是一致的（夹角小于 $\frac{\pi}{2}$），则 $df(x)>0$，即从点 x 移动，可以使 $f(x)$ 的值增加；如果梯度向量 f_x 和 dx 的方向不一致（夹角大于 $\frac{\pi}{2}$），则 $df(x)<0$，即从点 x 移动，可以使 $f(x)$ 的值减少；当且仅当梯度向量 f_x 和 dx 正交或其中有一个向量为零向量时（当其中有一个向量为零向量时，我们也可以将这两个向量看作时垂直的），$df(x)=0$，即随点 x 的微小移动，目标函数值不再增加，或者讲在点 x 处的一个领域内，找不到一个可移动的方向，使得目标函数值增加，即在最优解 $\bar{x}$ 处，一定有 $f_x \cdot dx = 0$。

在无约束极值问题中，dx 可以是任意向量，因此只要 f_x 不是零向量，总可以在点 x 处找到一个变化方向，使得向量 dx 和梯度向量 f_x 成锐角，从而有 $df(x)>0$。故而当目标函数在点 $\bar{x}$ 取得极值处，一定有梯度向量 $f_x(\bar{x})=0$，这就是无约束极值问题的一阶条件。

B.2.约束极值问题

当存在约束（不管是等式约束还是不等式约束）时，点 x 处的变化方向是有限制的，即向量 dx 不是任意的。比如在上面所举的消费者选择的例子中，在预算线上，向量 dx 的方向为沿预算线移动。此时，按我们上面对梯度向量 f_x 和向量 dx 的几何意义的讨论，当点 x^1（x^3）处梯度向量 f_x 和向量 dx 不正交时，所以向右（左）移动，可以使目标函数值增大；在点 x^2 处，梯度向量 f_x 和向量 dx 正交，$df(x)=0$，目标函数取得极值。由前面的讨论，向量 dx 的方向即为预算线的方向，而梯度向量 f_x 为无差异曲线的法方向，也就是与无差异曲线切线垂直的方向，由几何知识，我们知道，在点 x^2（最优解）处，无差异曲线的切线的斜率一定于预算线的斜率相同，如图示，无差异曲线在点 x^2 处一定于预算线相切。这个结论与我们在消费者选择定性讨论的结果是一致的。

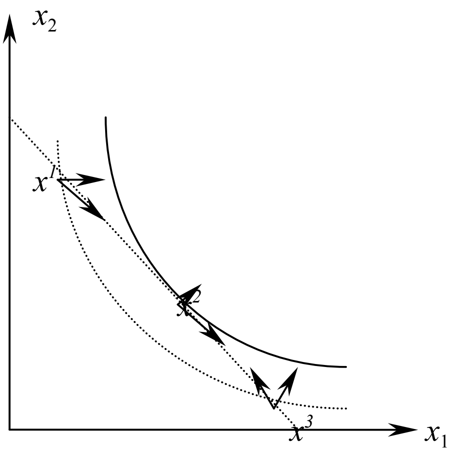


图 B.3

一般的，约束由等式$G(x)=c$表示，点 x 的变化向量dx是受限制的，具体的，将约束等式全微分，

$$G_1 dx_1+\cdots+G_n dx_n=0$$

由于c是常数，等式的右边等于零。写成向量的形式为：

$$(G_1,\cdots,G_n)\cdot\begin{pmatrix}dx_1\\ \vdots\\ dx_n\end{pmatrix}=0\text{，或}G_x\cdot dx=0\text{，其中}G_x=(G_1,\cdots,G_n)$$

因此，在点x处，向量dx的方向由上式确定，即向量dx与向量G_x垂直。

由在最优解$\bar{x}$处，满足$f_x\cdot dx=0$，即梯度向量f_x与向量dx垂直。由几何知识，我们知道，在最优解$\bar{x}$处，梯度向量f_x一定与向量G_x平行，即存在常数λ使得：

$$f_x=\lambda G_x\text{，或}\frac{f_x}{G_x}=\lambda$$

这就是我们熟悉的等式约束极值问题的一阶条件。

正如前面对梯度向量f_x的讨论，此处我们也可以将向量G_x看作时约束函数$G(x)$在点x处的梯度向量，即等值面$G(x)=c$在点x处的法向量。由$G_x\cdot dx=0$，向量dx与梯度向量G_x是垂直的，或者说向量dx可以看作是等值面$G(x)=c$的切平面中的一个向量。又由在点$\bar{x}$处，$f_x\cdot dx=0$，即梯度向量f_x和向量dx垂直，也就是说向量dx也在等值面$f(x)=k$的切平面中。因此，在最优解点$\bar{x}$处，两个等值面$f(x)=k$、$G(x)=c$的切平面重合，推得法向量f_x、G_x平行，即有

$f_x = \lambda G_x$。

梯度向量 $gradf$ 在一维的情形下就是导数 $f'(x)$，而在 n 维情形下就是偏导数向量 f_x。

B.2.1 雅克比矩阵

在上面的例子中，如果约束不是单个等式，而是由 m 个等式组成的，即 $G(x)$ 和 c 都是向量，具体的：

$$
\begin{aligned}
&G^1(x_1,\cdots,x_n) = c_1 \\
&\vdots \\
&G^m(x_1,\cdots,x_n) = c_m
\end{aligned}
$$

则此时，$G_x \equiv \dfrac{\partial(G^1,\cdots,G^m)}{\partial(x_1,\cdots,x_n)} \equiv \begin{pmatrix} \dfrac{\partial G^1}{\partial x_1},\cdots,\dfrac{\partial G^1}{\partial x_n} \\ \vdots \\ \dfrac{\partial G^m}{\partial x_1},\cdots,\dfrac{\partial G^m}{\partial x_n} \end{pmatrix}$，我们称之为 $G(x)$ 的雅克比矩阵。在隐函数定理的讨论中，会涉及到雅克比矩阵的概念。

B.2.2 隐函数定理

i)在隐函数 $f(x,y)=0$ 中（x,y 都是一维向量），如果在点 (x^0,y^0) 的一个领域内有 $f_y \neq 0$，则 y 可以表示成 x 的函数，即 $y=\phi(x)$，并且 $\dfrac{dy}{dx} = -\dfrac{f_x}{f_y}$。

ii)当 x 是 n 维向量时，如果在点 (x^0,y^0) 的一个领域内有 $f_y \neq 0$，则 y 可以表示成 n 维向量 x 的函数，即 $y=\phi(x)$，且 $\dfrac{\partial y}{\partial x_i} = -\dfrac{f_i}{f_y}, i=1,\cdots,n$。

这可以通过对隐函数 $f(x,y)=0$ 求全微分，$f_1 dx_1 + \cdots + f_n dx_n + f_y dy = 0$ 求得。

iii)当 y 是一个 m 维向量时，根据方程组的知识，y 的值要能够被确定，$f(x,y)$ 也一定是一个 m 维的向量，即：

$$
\begin{aligned}
&f^1(x_1,\cdots,x_n;y_1,\cdots,y_m) = 0 \\
&\vdots \\
&f^m(x_1,\cdots,x_n;y_1,\cdots,y_m) = 0
\end{aligned}
$$

此时，$f_y = \dfrac{\partial(f^1,\cdots,f^m)}{\partial(y_1,\cdots,y_m)} = \begin{pmatrix} \dfrac{\partial f^1}{\partial y_1},\cdots,\dfrac{\partial f^1}{\partial y_m} \\ \vdots \\ \dfrac{\partial f^m}{\partial y_1},\cdots,\dfrac{\partial f^m}{\partial y_m} \end{pmatrix}$，我们称之为 $f^1,\cdots,f^m$ 对于 $y_1,\cdots,y_m$ 的雅克比矩阵（$m\times m$）。

如果在点 (x^0,y^0) 的一个领域，有雅克比矩阵 f_y 是非奇异的，或 $\det f_y \neq 0$，即雅克比行列式不等于零，则向量 y 可以写出向量 x 的函数，即 $y=\phi(x)$，其中 $\phi(x)$ 是一个向量，并且有：

$$\begin{pmatrix} \dfrac{\partial y_1}{\partial x_i} \\ \vdots \\ \dfrac{\partial y_m}{\partial x_i} \end{pmatrix} = -f_y^{-1}\cdot f_i = -\begin{pmatrix} \dfrac{\partial f^1}{\partial y_1},\cdots,\dfrac{\partial f^1}{\partial y_m} \\ \vdots \\ \dfrac{\partial f^m}{\partial y_1},\cdots,\dfrac{\partial f^m}{\partial y_m} \end{pmatrix}^{-1} \cdot \begin{pmatrix} \dfrac{\partial f_1}{\partial x_i} \\ \vdots \\ \dfrac{\partial f^m}{\partial x_i} \end{pmatrix}$$

此结果可以通过对下列式子全微分，

$$\begin{aligned} &f^1(x_1,\cdots,x_n;y_1,\cdots,y_m)=0 \\ &\vdots \\ &f^m(x_1,\cdots,x_n;y_1,\cdots,y_m)=0 \end{aligned}$$

并利用克莱姆法则求得。

B.2.3 超平面

在上面消费者选择的例子中，预算约束等式为 $p_1x_1+p_2x_2=I$，即 $p\cdot x=I$ 在平面 x_1-x_2 中表现为一条直线；当 x 是三维向量时，预算约束等式为 $p_1x_1+p_2x_2+p_3x_3=I$，在三维空间 (x_1,x_2,x_3) 中表示一个平面，法方向为 (p_1,p_2,p_3)；当 n>3 时，预算约束 $p\cdot x=I$ 在 n 维欧氏空间 $(x_1,\cdots,x_n)$ 中表现为超平面。

C.等式约束极值的拉格朗日求解法（Lagrange’s Method）

前面我们从梯度的角度得出了约束等式极值问题

$$\max_x f(x)$$
$$s.t.$$
$$G(x)=c$$

在最优解 $\bar{x}$ 处的一阶条件，即 $f_x=\lambda G_x$，其中 λ 为常数。

此处，我们给出一个一般的求解方法——拉格朗日方法：通过引入一个参数，将一个约束极值问题转化成无约束极值问题。具体的，通过构造拉格朗日函数：

$$L(x,\lambda)=f(x)+\lambda(c-G(x))$$

其中 x 是 n 维向量，λ 参数，我们称之为拉格朗日乘子。上述约束极值在最优解 $\bar{x}$ 处的一阶条件可以用拉格朗日函数的一阶导数表示：

$$\frac{\partial L(x,\lambda)}{\partial x_i}=f_i-\lambda G_i=0, i=1,\cdots,n$$
$$\frac{\partial L(x,\lambda)}{\partial \lambda}=c-G(x)=0$$

第一个式子用向量表示，即为 $f_x=\lambda G_x$，其中 f_x, G_x 分别表示目标函数和约束函数在最优解点 $\bar{x}$ 处梯度向量的取值；第二个式子是约束等式的重新表述。共有 n+1 个变量 $(x_1,\cdots,x_n,\lambda)$，和 n+1 个等式，一般能直接求出最优解 $(\bar{x},\lambda)$，拉格朗日乘子 λ 是作为最优解的一部分求出来的。

当约束 $G(x)=c$ 表示 m 个等式时，求解过程与上面一样，此时拉格朗日乘子 λ 是一个 m 维的向量 $(\lambda_1,\cdots,\lambda_m)$，一阶条件由 $f_x=\lambda\cdot G_x$ 和约束等式 $G(x)=c$ 组成，n＋m 个变量 $(x_1,\cdots,x_n;\lambda_1,\cdots,\lambda_m)$，共 n+m 个等式，在一定的条件下（在最优解 $\bar{x}$ 处，雅克比矩阵 G_x 是非奇异的），可以求解出最优解 $(\bar{x},\lambda)$。

D.非线性规划问题的求解：库恩－塔克条件

前面我们讨论了等式约束的极值问题，但在经济学中，很多经济问题的约束都是以不等式形式出现的，如前面讨论的消费者选择问题中，我们给出的预算约束是所有的收入必须花完，但更为现实的是只要满足支出不超过收入就可以了。此外，我们还要加上选择变量非负的约束，比如商品的消费量不能小于零。

不等式约束及选择变量非负的最优化问题称之为非线性规划问题。一般的，可以用如下数学模型表示：

$$\max_x f(x)$$

$$s.t.$$

$$G(x) \le c$$

$$x \ge 0$$

对于这个问题的求解，我们只要在等式约束极值问题解法的基础上作一点修正：

第一步：作拉格朗日函数

$$L(x,\lambda) = f(x) + \lambda(c - G(x))$$

这个函数与等式约束极值问题的拉格朗日函数是一样的。

第二步：求一阶条件。显然，因为选择变量加上了非负的约束，所以原来在最优解 $\bar{x}$ 处拉格朗日函数对选择变量的一阶偏导等于零的一阶条件就要变为：

$$\frac{\partial L}{\partial x_i} = f_i - \lambda G_i \le 0, x_i \ge 0, x_i \frac{\partial L}{\partial x_i} = 0, i = 1, \cdots, n$$

在最优解 $\bar{x}$ 处，选择变量和相应的一阶偏导数乘积为零，表明当选择变量为正时（我们称之为内点解），一阶偏导为零，这是和等式约束一样的；而当选择变量取零时（我们称之为角点解），一阶偏导小于或等于零。这样的关系，我们称之为互补松弛条件。

在等式约束极值问题的一阶条件中拉格朗日函数对乘子 λ 的一阶导数等于零就是约束等式的重新表述；现在约束是以不等式形式出现的，对 λ 的一阶条件相应的修正为：

$$\frac{\partial L}{\partial \lambda} = c - G(x) \ge 0, \lambda \ge 0, \lambda \frac{\partial L}{\partial \lambda} = 0$$

和选择变量非负约束一样，不等式约束时的一阶条件需要添加一个互补松弛条件，即当 λ 为正时，约束取等号，此时我们称之为约束是紧的（binding）；λ 为零时，约束取不等号，则称此时约束为松的（non-binding）。这个互补松弛条件告诉我们，拉格朗日乘子的符号与约束的松紧情况是相关的：正的拉格朗日乘子对应于紧的约束；拉格朗日乘子为零则对应于松的约束。

把上面两个式子结合起来，就是非线性规划问题在最优解 $\bar{x}$ 处的一阶必要条

件，我们称之为库恩－塔克条件（Kuhn-Tucker Condition）。

当约束个数是 m (m>1)个时，$G(x), c, \lambda$ 都表示 m 维的向量，相应的乘积变为内积即可。

与等式约束极值问题的一阶必要条件相比，库恩－塔克条件是用不等式形式出现的，因此给求解带来了很大的麻烦。具体的，在前面的两种商品的消费者选择例子中，等式约束极值问题的一阶条件表现为 (x_1, x_2, λ) 三个未知量，三个方程；而当预算约束以不等式形式出现，并且加上 $x_1 \geq 0, x_2 \geq 0$ 条件后，库恩－塔克条件表现为三个未知量，三个不等式（及互补松弛条件），具体的求解需要讨论选择变量 $x_1 > 0, = 0; x_2 > 0, = 0; \lambda > 0, = 0$ 等各种情况后才能将不等式转化成等式，再进一步求解。三个变量每一个都有为正、零两种情况，所以组合起来一般的求解过程需要讨论 8 种情形，并在每一种情形下求出最优解。下面我们用一个具体的例子来看非线性规划库恩－塔克条件的求解过程。

例子：拟线性偏好（Quasi-Linear Preference）

$$\max_{x_1, x_2} \quad u(x_1, x_2) = x_2 + a \ln x_1$$

$$s.t. \qquad \begin{aligned} & p_1 x_1 + p_2 x_2 \leq I \\ & x_1, x_2 \geq 0 \end{aligned}$$

作拉格朗日函数 $L(x_1, x_2, \lambda) = x_2 + a \ln x_1 + \lambda(I - p_1 x_1 - p_2 x_2)$，一阶必要条件（库恩－塔克条件）：

$$\frac{\partial L}{\partial x_1} = \frac{a}{x_1} - \lambda p_1 \leq 0, x_1 \geq 0, x_1 \frac{\partial L}{\partial x_1} = 0$$

$$\frac{\partial L}{\partial x_2} = 1 - \lambda p_2 \leq 0, x_2 \geq 0, x_2 \frac{\partial L}{\partial x_2} = 0$$

$$\frac{\partial L}{\partial \lambda} = I - p_1 x_1 - p_2 x_2 \geq 0, \lambda \geq 0, \lambda \frac{\partial L}{\partial \lambda} = 0$$

如何求解？一般情况下，我们要讨论 $x_1 > 0, = 0; x_2 > 0, = 0; \lambda > 0, = 0$ 组合成的 8 种情况，不过，在这一具体的问题中，我们可以通过分析题目中隐含的条件（经济含义）初步确定未知量的范围，以减少讨论的可能情形。

比如，在此问题中，边际效用 u_1, u_2 均为正，推得不会有收入剩余，否则可以通过继续增加消费，使得效用增加，因此预算约束是紧的，即 $\lambda > 0$；另外，由效用函数的形式，$x_1 > 0$。因此，需要讨论的情形只有两种了：

第一种情形：$x_1 > 0, x_2 = 0, \lambda > 0$。此时库恩－塔克条件变为：

$$\frac{\partial L}{\partial x_1} = \frac{a}{x_1} - \lambda p_1 = 0$$
$$\frac{\partial L}{\partial x_2} = 1 - \lambda p_2 \le 0$$
$$\frac{\partial L}{\partial \lambda} = I - p_1 x_1 = 0$$

求得$x_1 = \frac{I}{p_1}, x_2 = 0, \lambda = \frac{a}{I}$,并且满足参数条件$I \le ap_2$。

第二种情形：$x_1 > 0, x_2 > 0, \lambda > 0$

此时，库恩－塔克条件变为：

$$\frac{\partial L}{\partial x_1} = \frac{a}{x_1} - \lambda p_1 = 0$$
$$\frac{\partial L}{\partial x_2} = 1 - \lambda p_2 = 0$$
$$\frac{\partial L}{\partial \lambda} = I - p_1 x_1 - p_2 x_2 = 0$$

求得$x_1 = \frac{ap_2}{p_1}, x_2 = \frac{I - ap_2}{p_2}, \lambda = \frac{1}{p_2}$，并且有$I > ap_2$。

从这个例子的讨论中，我们知道，虽然非线性规化的库恩－塔克条件的求解过程一般需要讨论选择变量和拉格朗日乘子为零和为正所组合成的每一种情况，但可以通过对变量的范围的初步分析，以减少求解情形的可能性并简化求解的过程。

另外，从求解的具体过程看，虽然题目没有直接告诉我们，但其实问题中的参数是有条件的，参数空间可以划分成各个不同的部分，每一部分对应于求解的一个具体情形。如本题中，当$I \le ap_2$时，$x_1 > 0, x_2 = 0, \lambda > 0$；而当$I > ap_2$时，$x_1 > 0, x_2 > 0, \lambda > 0$。

这个例子告诉我们：在拟线性偏好的效用函数中，收入较低时，只消费商品 1；当收入超过一定水平时，两种商品都消费，但是第一种商品消费的量保持不变，而所有增加的收入都用来消费商品 2。从这个角度看，我们可以将此例中的商品 1 看作是生活中的必需品（如简单的衣食住行），而商品 2 可以看作是高档消费品，低收入时只消费必需品，而当收入上升到一定水平时，必需品的消费不再增加，转而消费高档品。

E.二阶条件

前面我们讨论了两类最优化问题：约束极值问题及非线性规化问题，并且给出了求解两类问题的拉格朗日方法和库恩－塔克条件。

不过，从求解的过程看，无论是前面梯度向量的角度，还是后面等式约束问题的拉格朗日方法、非线性规划的库恩－塔克条件，我们都只是分析了在最优解处应满足的性质，而并没有保证满足此性质的点一定是最优解，比如最大值问题和最小值问题的一阶条件是相同的，或者说由拉格朗日方法推导出的一阶条件和库恩－塔克条件只是解的必要条件。为此，我们需要找到新的条件，以保证由一阶必要条件所求得的解就是此最优规划问题的最优解，这就是二阶条件的讨论。

E.1 无约束极值问题

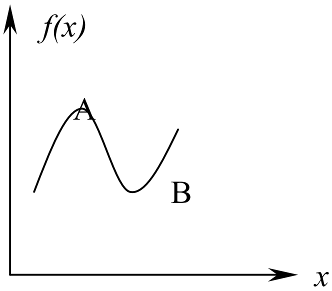


图 E.1

在选择变量 x 是一维情况下，一阶条件为 $f'(x)=0$，体现在平面 $x-f(x)$ 中，曲线 $f(x)$ 的斜率为零。从上图看，点 A 和点 B 都满足这样的性质。显然，这两个点的性质是不同的，函数在点 A 的一个领域内取得了极大值，而在点 B 取得了极小值。因此，从此例看，由一阶必要条件求得的解不能保证就是最大化极值问题的解。为此，我们需要寻找出新的条件，以保证满足此条件的点就是最优解。

E.1.1 泰勒展开

在进一步讨论之前，我们先给出函数 $f(x)$ 在点 x_0 附近的一个近似估计：

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{1}{2}f''(x_0)(x-x_0)^2+\cdots$$

或 $$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{1}{2}f''(x_0)(x-x_0)^2+o((x-x_0)^2)$$

等式右边的最后一项是一个与二阶项相比非常小、可忽略的余项（无穷小量）。因此，我们可以写出近似的形式：

$$f(x)\approx f(x_0)+f'(x_0)(x-x_0)+\frac{1}{2}f''(x_0)(x-x_0)^2$$

上述三个式子都可称为函数 $f(x)$ 在点 x_0 附近的泰勒展开。

现在，将$f(x)$在最优解$\bar{x}$处泰勒展开：

$$f(x) \approx f(\bar{x}) + f'(\bar{x})(x-\bar{x}) + \frac{1}{2}f''(\bar{x})(x-\bar{x})^2$$

由$f(x)$在最优解$\bar{x}$处的一阶导数为零，我们得：

$$f(x) - f(\bar{x}) \approx \frac{1}{2}f''(\bar{x})(x-\bar{x})^2$$

因此，等式左边的正负取决于函数$f(x)$在点$\bar{x}$处二阶导数的符号。显然，当$f''(\bar{x}) < 0$时，一定有 $f(x) - f(\bar{x}) < 0$，或 $f(\bar{x}) > f(x)$，即函数$f(x)$在点$\bar{x}$处附近取得了极大值。我们将$f''(\bar{x}) < 0$称为函数$f(x)$在点$\bar{x}$处附近取得了极大值的二阶充分条件。同样，极小值的二阶（充分）条件为$f''(\bar{x}) > 0$。

另外，当$f(x)$在点$\bar{x}$处附件取得了极大值时，在点$\bar{x}$处附近一定有$f(\bar{x}) \geq f(x)$。由泰勒展开及$f(x)$在点$\bar{x}$处附件的一阶条件为零，一定有$f''(\bar{x}) \leq 0$，我们称之为极大值问题的二阶必要条件。同样的，极小值问题的二阶必要条件为$f''(\bar{x}) \geq 0$。

当选择变量x是n维向量时，在最优解$\bar{x}$处相应的泰勒展开变为：

$$f(x) = f(\bar{x}) + gradf(\bar{x}) \cdot (x-\bar{x}) + \frac{1}{2}(x-\bar{x})^T \cdot H \cdot (x-\bar{x}) + o(\|x-\bar{x}\|^2)$$

此时，一阶导数由函数在点$\bar{x}$处的梯度向量表示，一阶项为梯度与向量$x-\bar{x}$的内积；二阶项中的二阶导数在n维情形下扩展为矩阵：

$H = \begin{pmatrix} f_{11},\cdots,f_{1n} \\ \vdots \\ f_{n1},\cdots,f_{nn} \end{pmatrix}$，其中$f_{ij}, i, j = 1,\cdots,n$表示函数$f(x)$在点$\bar{x}$处的二阶偏导，我们称矩阵 H 为海赛矩阵，向量$x-\bar{x}$的上标 T 表示转置。

余项为可以忽略不计的向量$x-\bar{x}$模的平方的无穷小量。

同样，在点$\bar{x}$处，梯度向量$gradf$为零，因此$f(x)-f(\bar{x})$的正负取决于二阶项的符号。而此形式是一个标准的二次型，在进一步讨论之前，我们先来看看有关二次型的知识：

E.1.2 二次型

我们称形如$Q(x)=x^T Ax$为二次型，其中$x=\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$，$x$的上标 T 表示转置，

即 $x^T=(x_1,\cdots,x_n)$，其中矩阵A为$n\times n$的对称阵，$A=\begin{pmatrix} a_{11},\cdots,a_{1n} \\ \vdots \\ a_{n1},\cdots,a_{nn} \end{pmatrix}$，其中

$a_{ij}=a_{ji},i,j=1,\cdots,n$，向量之间为内积。

如果除$x=0$以外，对任意的x，有$Q(x)>0(\geq 0)$，则我们称$Q(x)$正定（半正定）；

如果除$x=0$以外，对任意的x，有$Q(x)<0(\leq 0)$，则我们称$Q(x)$负定（半负定）。

x是任意的，因此二次型$Q(x)$的性质取决于矩阵A的性质，我们称矩阵A是（半）正定（或（半）负定）的，当二次型$Q(x)$为（半）正定（或（半）负定）。

而矩阵A是正定的等价于其顺序主子式为正，即：

$D_1>0,D_2>0,\cdots,D_n>0$

矩阵A是负定的等价于其顺序主子式符号负正依次相间，即：

$D_1<0,D_2>0,\cdots,(-1)^n D_n>0$

而矩阵A是半正定的，等价于k阶主子式非负，即：

$\overline{D_1}\geq 0,\overline{D_2}\geq 0,\cdots,\overline{D_n}\geq 0$

矩阵A是半负定的等价于其k阶主子式符号非正非负依次相间，即：

$\overline{D_1}\leq 0,\overline{D_2}\geq 0,\cdots,(-1)^n\overline{D_n}\geq 0$

顺序主子式$D_k=\begin{vmatrix} a_{11}a_{12}\cdots a_{1k} \\ \vdots \\ a_{k1}a_{k2}\cdots a_{kk} \end{vmatrix},k=1,2,\cdots,n$，

k 阶主子式$\overline{D_k}=\begin{vmatrix} a_{i_1i_1}a_{i_1i_2}\cdots a_{i_1i_k} \\ \vdots \\ a_{i_ki_1}a_{i_ki_2}\cdots a_{i_ki_k} \end{vmatrix},1\leq i_1\leq\cdots\leq i_k\leq n,k\leq n$

根据以上二次型的知识，我们知道二次项的符号就取决于海赛矩阵 H 的正

负定性，如果 H 为负定，则 $f(x)-f(\bar{x})<0$，即函数 $f(x)$ 在点 $\bar{x}$ 附近取得极大值；同样的，当海赛矩阵 H 为正定时，函数 $f(x)$ 在点 $\bar{x}$ 附近取得极小值。

因此，在选择变量为 n 维的无约束极值问题中，讨论二阶充分条件时只需要将原来二阶导数的概念扩展为矩阵——海赛矩阵。极大值的二阶充分条件由原来的二阶导数为负修正为海赛矩阵为负定；极小值的二阶充分条件变为海赛矩阵为正定。

而极大值问题的二阶必要条件为海赛矩阵半负定，极小值问题的二阶必要条件为海赛矩阵半正定。

海赛矩阵 $H=\begin{pmatrix} f_{11},\cdots,f_{1n} \\ \vdots \\ f_{n1},\cdots,f_{nn} \end{pmatrix}$ 为负定时，等价于

$$|f_{11}|<0,\begin{vmatrix} f_{11}f_{12} \\ f_{21}f_{22} \end{vmatrix}>0,\cdots,(-1)^n\begin{vmatrix} f_{11}\cdots f_{1n} \\ \vdots \\ f_{n1}\cdots f_{nn} \end{vmatrix}>0\text{；}$$

海赛矩阵 H 为正定时，等价于

$$|f_{11}|>0,\begin{vmatrix} f_{11}f_{12} \\ f_{21}f_{22} \end{vmatrix}>0,\cdots,\begin{vmatrix} f_{11}\cdots f_{1n} \\ \vdots \\ f_{n1}\cdots f_{nn} \end{vmatrix}>0$$

当 n＝1 时，我们可以检验一下上述结论是否与前面相一致。

E.2 等式约束极值问题：

以选择变量是二个、等式约束为一个的情形为例，即：

$$\max_{x_1,x_2} f(x_1,x_2)$$

$$s.t.$$

$$g(x_1,x_2)=c$$

拉格朗日方法的思想，就是通过引入拉格朗日乘子 λ，构造拉格朗日函数：

$$L(x_1,x_2,\lambda)=f(x_1,x_2)+\lambda(c-g(x_1,x_2))$$

将上述等式约束问题变成一个无约束问题。只要满足一定的约束规格，等式约束的最优解 $\bar{x}$ 就是无约束问题 $L(x_1,x_2,\lambda)$ 的最优解。而只要满足 $g(x_1,x_2)=c$（也就是拉格朗日函数 $L(x_1,x_2,\lambda)$ 对 λ 的一阶条件），函数 $f(x_1,x_2)$ 的值和 $L(x_1,x_2,\lambda)$ 是相等的。在最优解 $\bar{x}$ 处满足一阶条件：

$$\frac{\partial L}{\partial x_1} = f_1 - \lambda g_1 = 0$$
$$\frac{\partial L}{\partial x_2} = f_2 - \lambda g_2 = 0$$
$$\frac{\partial L}{\partial \lambda} = c - g(x_1, x_2) = 0$$

我们对 $f(x_1, x_2)$ 求全微分，并在点 $\bar{x}$ 处取值：

$df = f_1 dx_1 + f_2 dx_2$，其中 f_1, f_2 为函数 $f(x_1, x_2)$ 在点 $\bar{x}$ 处的偏导数。

在无约束问题中，微分 dx_1, dx_2 是可以自由变动的；而在约束极值问题中，始终满足条件 $g(x_1, x_2) = c$，等式两边全微分，也即始终满足条件：

$g_1 dx_1 + g_2 dx_2 = 0$，即 $dx_2 = -\frac{g_1}{g_2} dx_1$，其中 g_1, g_2 为约束函数的偏导数。从这个条件看，约束问题与无约束问题的区别在于，dx 不再是任意的，dx_1、dx_2 是互相联系的，dx_2 可以看作是变量 x_1, x_2, dx_1 的函数（dx_1 可以看作是自由变动的变量）。对函数 $f(x_1, x_2)$ 求两阶全微分：

$$d^2 f = d(df) = d(f_1 dx_1 + f_2 dx_2) = \frac{\partial (f_1 dx_1 + f_2 dx_2)}{\partial x_1} dx_1 + \frac{\partial (f_1 dx_1 + f_2 dx_2)}{\partial x_2} dx_2$$
$$= (f_{11} dx_1 + f_{21} dx_2 + f_2 \frac{\partial dx_2}{\partial x_1}) dx_1 + (f_{12} dx_1 + f_{22} dx_2 + f_2 \frac{\partial dx_2}{\partial x_2}) dx_2$$
$$= f_{11} dx_1^{\ 2} + f_{21} dx_2 dx_1 + f_2 \frac{\partial dx_2}{\partial x_1} dx_1 + f_{12} dx_1 dx_2 + f_{22} dx_2^{\ 2} + f_2 \frac{\partial dx_2}{\partial x_2} dx_2$$
$$= f_{11} dx_1^{\ 2} + 2 f_{12} dx_1 dx_2 + f_{22} dx_2^{\ 2} + f_2 d(dx_2)$$

其中最后一个等式利用了 Young 定理，$f_{12} = f_{21}$。

对约束等式 $g(x_1, x_2) = c$ 求二阶全微分，同样的，有：

$$g_{11} dx_1^{\ 2} + 2 g_{12} dx_1 dx_2 + g_{22} dx_2^{\ 2} + g_2 d(dx_2) = 0$$

代入上式，有：

$$d^2 f = (f_{11} - \frac{f_2}{g_2} g_{11}) dx_1^{\ 2} + 2(f_{12} - \frac{f_2}{g_2} g_{12}) dx_1 dx_2 + (f_{22} - \frac{f_2}{g_2} g_{22}) dx_2^{\ 2}$$

将有一阶条件 $f_2 - \lambda g_2 = 0$ 代入，得：

$$d^2 f = (f_{11} - \lambda g_{11}) dx_1^{\ 2} + 2(f_{12} - \lambda g_{12}) dx_1 dx_2 + (f_{22} - \lambda g_{22}) dx_2^{\ 2}$$

如果我们用拉格朗日函数$L(x_1, x_2, \lambda)$的偏导数

$$L_{11} = f_{11} - \lambda g_{11}$$
$$L_{12} = f_{12} - \lambda g_{12}$$
$$L_{22} = f_{22} - \lambda g_{22}$$

表示，则$d^2 f = L_{11} dx_1^{\ 2} + 2L_{12} dx_1 dx_2 + L_{22} dx_2^{\ 2}$，偏导数在点$\bar{x}$取值。写出矩阵的形式：

$$d^2 f = (dx)^T \cdot L_{xx} \cdot dx \text{，其中} dx = (dx_1, dx_2), L_{xx} = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix}$$

由微积分的知识，我们知道，如果在点$\bar{x}$处有$d^2 f < 0$，则函数$f(x_1, x_2)$取得极大值；如果在点$\bar{x}$处有$d^2 f > 0$，则函数$f(x_1, x_2)$取得极小值。

正如我们前面所讨论的，与无约束问题的区别在于，此处，dx_2是变量x_1, x_2, dx_1的函数，不能自由变动，因此上式不是一个二次型。将$dx_2 = -\dfrac{g_1}{g_2} dx_1$代入上式，得：

$$d^2 f = (L_{11} g_2^{\ 2} - 2L_{12} g_1 g_2 + L_{22} g_1^{\ 2}) \frac{1}{g_2^{\ 2}} dx_1^{\ 2}$$

显然，等式右边括号外边的部分为正，所以$d^2 f$的正负取决于括号部分的符号。

令：

$$\left|\overline{H}\right| = \begin{vmatrix} 0 & g_1 & g_2 \\ g_1 & L_{11} & L_{12} \\ g_2 & L_{21} & L_{22} \end{vmatrix} = -(L_{11} g_2^{\ 2} - 2L_{12} g_1 g_2 + L_{22} g_1^{\ 2})$$

$\left|\overline{H}\right|$是可以这么看，第一行和第一列是由约束函数的一阶偏导组成的，而余下的部分可以看作是无约束问题的拉格朗日函数的海赛矩阵，因此我们就把$\left|\overline{H}\right|$称为<u>加边的海赛行列式</u>，就像是由无约束极值问题的海赛行列式再加了两条边。

当$d^2 f < 0$时，函数$f(x)$在点$\bar{x}$处取得极大值，也即$-\left|\overline{H}\right| < 0$，这就是约束极大值问题二阶充分条件；同样，当$-\left|\overline{H}\right| > 0$，$d^2 f > 0$，函数函数$f(x)$在点$\bar{x}$

处取得极小值，此为约束极小值问题的二阶充分条件。

当选择变量是$n(n>2)$维，一个约束等式时，

$$|\overline{H}|=\begin{vmatrix} 0 & g_1 & \cdots & g_n \\ g_1 & L_{11} & \cdots & L_{1n} \\ \vdots & \vdots & & \\ g_n & L_{n1} & \cdots & L_{nn} \end{vmatrix}$$

函数$f(x_1,\cdots,x_n)$在$\overline{x}$处取得极大值的二阶充分条件也等价于：

$$|\overline{H}_2|>0,|\overline{H}_3|<0,\cdots$$

同样，函数$f(x_1,\cdots,x_n)$在$\overline{x}$处取得极小值的二阶充分条件等价于：

$$|\overline{H}_2|<0,|\overline{H}_3|<0,\cdots$$

其中：$|\overline{H}_2|=\begin{vmatrix} 0 & g_1 & g_2 \\ g_1 & L_{11} & L_{12} \\ g_2 & L_{21} & L_{22} \end{vmatrix},|\overline{H}_3|=\begin{vmatrix} 0 & g_1 & g_2 & g_3 \\ g_1 & L_{11} & L_{12} & L_{13} \\ g_2 & L_{21} & L_{22} & L_{23} \\ g_3 & L_{31} & L_{32} & L_{33} \end{vmatrix}$，依此类推。

当选择变量是n维，约束等式有m（$m<n$）个时，

$$L(x_1,\cdots,x_n;\lambda_1,\cdots,\lambda_m)=f(x_1,\cdots,x_n)+\sum_{j=1}^{m}\lambda_j[c_j-g^j(x_1,\cdots,x_n)]$$

$$|\overline{H}|=\left|\begin{array}{ccc:cccc} 0 & \cdots & 0 & g_1^1 & g_2^1 & \cdots & g_n^1 \\ \vdots & & \vdots & \vdots & & & \vdots \\ 0 & \cdots & 0 & g_1^m & g_2^m & \cdots & g_n^m \\ \hdashline g_1^1 & \cdots & g_1^m & L_{11} & L_{12} & \cdots & L_{1n} \\ g_2^1 & \cdots & g_2^m & L_{21} & L_{22} & \cdots & L_{2n} \\ \vdots & & \vdots & \vdots & & & \vdots \\ g_n^1 & \cdots & g_n^m & L_{n1} & L_{n2} & \cdots & L_{nn} \end{array}\right|$$

此时加边的海赛行列式可以分成四块：

$$|\overline{H}|=\left|\begin{array}{c:c} 0 & G \\ \hdashline G^T & L_{xx} \end{array}\right|$$

左上角是$m\times m$的零矩阵，右上角就是约束函数的雅克比矩阵G

$$G=\frac{\partial(g^1,\cdots,g^m)}{\partial(x_1,\cdots,x_n)},$$

左下角是雅克比矩阵G的转置，右下角为拉格朗日函数L的海赛矩阵。

在这种情形下，当$d^2f<0$时，函数$f(x)$在点$\overline{x}$处取得极大值，也即

$(-1)^m\left|\overline{H}\right|<0$；同样，当$(-1)^m\left|\overline{H}\right|>0$，$d^2f>0$，函数函数$f(x)$在点$\bar{x}$处取得极小值。

同样，此情形下的约束极值问题的二阶条件也可以用相应的加边海赛行列式的子式的符号表达：

极大值的二阶充分条件为$(-1)^m\left|\overline{H}_{m+k}\right|(k=1,\cdots,n-m)$ 的符号为$(-1)^k$，即$\left|\overline{H}_{m+k}\right|$的符号为$(-1)^{m+k}$；而极小值的二阶充分条件为$(-1)^m\left|\overline{H}_{m+k}\right|(k=1,\cdots,n-m)$的符号为正，即$\left|\overline{H}_{m+k}\right|$的符号为$(-1)^m$。

$$\left|\overline{H}_{m+k}\right|=\left|\begin{array}{ccc:cccc} 0 & \cdots & 0 & g_1^1 & g_2^1 & \cdots & g_{m+k}^1 \\ \vdots & & \vdots & \vdots & & & \vdots \\ 0 & \cdots & 0 & g_1^m & g_2^m & \cdots & g_{m+k}^m \\ \hdashline g_1^1 & \cdots & g_1^m & L_{11} & L_{12} & \cdots & L_{1,m+k} \\ g_2^1 & \cdots & g_2^m & L_{21} & L_{22} & \cdots & L_{2,m+k} \\ \vdots & & \vdots & \vdots & & & \vdots \\ g_{m+k}^1 & \cdots & g_{m+k}^m & L_{m+k,1} & L_{m+k,2} & \cdots & L_{m+k,m+k} \end{array}\right|$$

从前面的讨论看，无约束极值问题的二阶充分条件由海赛矩阵的正负定性来表达；而约束极值问题的二阶充分条件则与由相应的拉格朗日函数的海赛矩阵和约束函数的雅克比矩阵组成的加边海赛行列式的正负号相关。为了理解上的一致性，我们同样可以定义加边的海赛矩阵的正负定，以表达约束极值问题的二阶充分条件。我们称当加边海赛矩阵$\overline{H}$为负定时，函数$f(x)$在点$\bar{x}$处取得极大值。不过，与无约束极值问题相比，从我们的推导看，在用主子式的符号判断正负定性时，前面还要乘以$(-1)^m$（指数m是约束等式的个数），比如在选择变量为 2 个、约束等式为 1 个时，按海赛矩阵负定的含义，主子式的符号应该是负正交替，但由于在推导的过程中，二阶微分的符号由$-\left|\overline{H}\right|$决定，前面有个负号，所以最后以主子式的符号表达的二阶充分条件是正负交替。极小值的二阶充分条件为加边海赛矩阵为正定，不过在用主子式的符号表达二阶充分条件是，也要乘以$(-1)^m$。

关于二阶必要条件的问题，同无约束极值问题的讨论一样，只要将严格不等号改为小于等于（或大于等于）即可。

E.3 不等式约束问题

当规划问题的约束是以不等式形式出现的，则相应的二阶条件应如何表达？

正如非线性规划问题的库恩－塔克条件的求解，我们利用参数条件（不同参数空间的划分）将问题分成几种不同的情形，而在每一个情形下，规划问题是以等式约束形式出现的，相应的二阶条件就可以参照上面的分析进行。

F.凹规划(Concave Programming)

前面的讨论都是通过微分法，分析在最优解在一个小的领域内应满足的条件，所以我们说这样求得的解只是在这个小的领域内成立的，是局部的，因此我们称之为局部解(Local Optima)。在整个定义域内，可能存在很多个满足这样性质的点，因此我们需要比较各个局部解的目标函数值的大小，才能确定哪个是整个规划问题的最优解，我们称此解为全局解(Global Optima)。

如果通过前面我们介绍的拉格朗日方法、库恩－塔克条件求解出的局部解有很多个，通过相互比较确定全局解将是一个很繁琐的过程。另外，正如我们在上一小节所讨论的，要确定由一阶必要条件求出的解就是规划问题的极值需要通过二阶充分条件的检验，而这本身是很复杂的。通过我们对所求解经济问题的现实含义，对目标函数及约束函数进行凹凸性的假设，我们可以避免上述两个问题，保证一阶必要条件也是充分条件，同时所求的局部解就是规划问题的全局解。这是凹规划所讲的内容。

在进入凹规划的讨论之前，先来介绍几个与之相关的概念：

F.1.1 凸集（Convex Set）：

对集合 S，其中的任意两点 $x_1, x_2 \in S, \theta \in [0,1]$，如果有 $\theta x_1 + (1-\theta)x_2 \in S$，则称集合 S 为凸集。

也就是说，凸集有这样的性质：集合中的任何两点，其连线也在集合中。

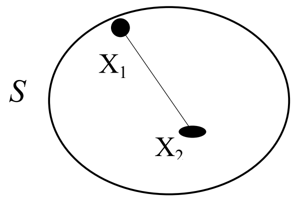


图 F.1

F.1.2 凸函数(Convex Function)：

对函数 $f(x)$，如果定义域 S（凸集）中的任何两点 $x_1, x_2 \in S, \theta \in [0,1]$，如果有 $f[\theta x_1 + (1-\theta)x_2] \le \theta f(x_1) + (1-\theta) f(x_2)$，则称函数 $f(x)$ 为凸函数。

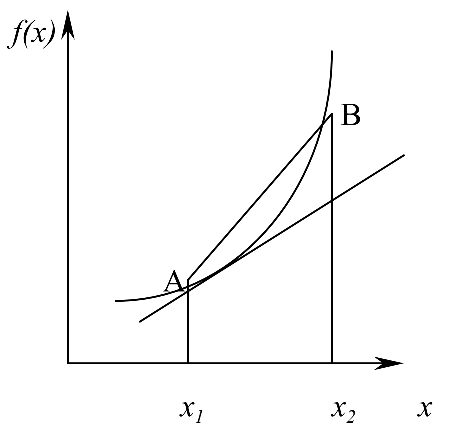


图 F.2

从图形上看，凸函数的定义指两点x_1, x_2连线的函数值（在曲线上）要比两点函数值 A、B 的连线（弦）低，或者说曲线上任何两点连成的弦 AB 在曲线上面。如果这两点非常近，（我们可以将点 B 逐步移向点 A，看此时弦 AB 的变化），则弦 AB 变成经过点 A 的切线。显然，此时经过点 A 的切线在曲线的下面。这个性质可以用如下的式子来表达：

$$f'(x_1)(x_2 - x_1) \le f(x_2) - f(x_1)$$

即在同一个点 x 用切线上的点估计的函数值小于曲线上的值。严格的证明利用微分中值定理很容易得到。

如果在点x_1处函数泰勒展开：

$$f(x_2) \approx f(x_1) + f'(x_1)(x_2 - x_1) + \frac{1}{2} f''(x_1)(x_2 - x_1)^2$$

则上述性质也等价于在点x_1的函数的二阶导数为非负。由于点x_1的一般性，我们可以得到凸函数的微分性质：

$$f''(x) \ge 0$$

总结一下，关于凸函数的性质如下：

i) 凸函数的几何性质表现为曲线上的任何两点连成的弦在曲线的上方。如果函数$f(x)$是可微的，则有$f'(x_1)(x_2 - x_1) \le f(x_2) - f(x_1)$。

ii) 如果函数$f(x)$是凸的，并且是二阶可微的，则有$f''(x) \ge 0$。在经济学中，常见的凸函数有成本函数，体现为边际成本递增；当选择变量为 n 为时，二阶条件变为相应的海赛矩阵为正定。

iii) 凸函数的非负线性组合是凸的，即如果 $f_i(x) (i = 1, \cdots, n)$ 是凸的，$k_i \ge 0, i = 1, \cdots, n$，则$\sum_i k_i f_i(x)$也是凸的。

iv) 如果前面的讨论中，所有的不等式是严格的，则我们称函数是严格凸的。

另外，在上述的讨论中，凸函数本身的定义并不要求函数是可微的。而自变量也不局限在一维的情形，多维的情形同样成立。

F.1.3.凹函数（Concave Function）：

对凹函数的一个简单定义是，如果函数$-f(x)$是凸函数，则$f(x)$是凹函数。因此，凹函数的性质只要将前面讨论中的不等式改向即可。正式的定义如下：

函数$f(x)$，对定义域 S（凸集）上任意两点$x_1, x_2 \in S, \theta \in [0,1]$，如果有

$f[\theta x_1 + (1-\theta) x_2] \ge \theta f(x_1) + (1-\theta) f(x_2)$，则我们称函数$f(x)$为凹函数。

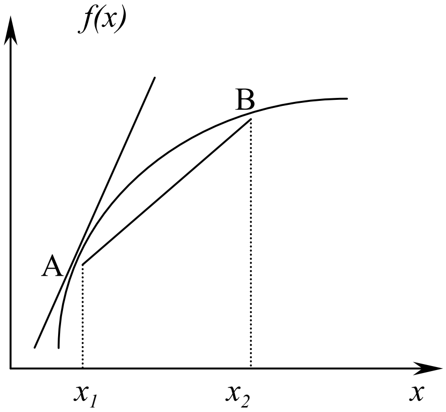


图 F.3

同凸函数一样，凹函数的性质可以总结如下：

i) 凹函数的几何性质表现为曲线上的任何两点连成的弦在曲线的下方。如果函数 $f(x)$ 是可微的，则有 $f'(x_1)(x_2-x_1)\ge f(x_2)-f(x_1)$。

ii) 如果函数 $f(x)$ 是凹的，并且是二阶可微的，则有 $f''(x)\le 0$。在经济学中，常见的凹函数有效用函数，体现为边际效用递减；而当选择变量为 n 维时，相应的海赛矩阵为负定。

iii) 凹函数的非负线性组合是凹的，即如果 $f_i(x)(i=1,\cdots,n)$ 是凹的，$k_i\ge 0, i=1,\cdots,n$，则 $\sum_i k_i f_i(x)$ 也是凹的。

iv)如果前面的讨论中，所有的不等式是严格的，则我们称函数是严格凹的。

F.2.凹规划

前面我们说过，只要对目标函数、约束函数的凹凸性作适当地假定，前面求解最优化问题过程中的一阶必要条件（库恩－塔克条件）所确定的解就是规划问题的解，即一阶必要条件就是充分条件；同时所求得的局部解就是全局解。我们用一个定理来总结上述性质：

对于非线性规划问题：

$$\max_x f(x)$$
$$s.t.$$
$$g(x)\le c$$
$$x\ge 0$$

如果目标函数 $f(x)$ 为凹函数，约束函数 $g(x)$ 为凸函数，且都可微，点 $\bar{x}$ 为由库恩－塔克条件所确定的解，则 $\bar{x}$ 就是规划问题的整体最优解。

这里，我们给出一个简单的证明：

首先，构造拉格朗日函数 $L(x,\lambda)=f(x)+\lambda[c-g(x)]$，因为 $f(x)$ 为凹函数，$g(x)$ 为凸函数，由前面我们对凹函数、凸函数性质的讨论知道，当把拉格朗日

乘子λ看作是常数时，拉格朗日函数$L(x,\lambda)$也是变量x的凹函数。因此，在点$\bar{x}$处$L(x,\lambda)$泰勒展开：

$$L(x,\lambda)\approx L(\bar{x},\lambda)+L_{\bar{x}}\cdot(x-\bar{x})+\frac{1}{2}(x-\bar{x})^T\cdot L_{\bar{x}\bar{x}}\cdot(x-\bar{x})$$

由凹函数的海赛矩阵负定的性质，我们有：

$$L_{\bar{x}}\cdot(x-\bar{x})\geq L(x,\lambda)-L(\bar{x},\lambda)$$

即　$L_{\bar{x}}\cdot x-L_{\bar{x}}\cdot\bar{x}\geq L(x,\lambda)-L(\bar{x},\lambda)$

进一步，$L(\bar{x},\lambda)\geq L(x,\lambda)-L_{\bar{x}}\cdot x+L_{\bar{x}}\cdot\bar{x}$

其中$L_{\bar{x}}$表示拉格朗日函数$L(x,\lambda)$在点$\bar{x}$处的导数（或梯度）。

由条件，点$\bar{x}$满足库恩－塔克条件，即

$$L_{\bar{x}}\leq 0,\bar{x}\geq 0,L_{\bar{x}}\cdot\bar{x}=0$$

另有，$x\geq 0$，得$L_{\bar{x}}\cdot x\leq 0,L_{\bar{x}}\cdot\bar{x}=0$。因此，有：

$$L(\bar{x},\lambda)\geq L(x,\lambda)$$

即　$f(\bar{x})+\lambda[c-g(\bar{x})]\geq f(x)+\lambda[c-g(x)]$

而点$\bar{x}$满足库恩－塔克条件，即

$$L_\lambda=c-g(\bar{x})\geq 0,\lambda\geq 0,\lambda\cdot L_\lambda=\lambda\cdot[c-g(\bar{x})]=0$$

另有，$c-g(x)\geq 0$，代入上述不等式，有：

$$f(\bar{x})\geq f(x)$$

即满足一阶必要条件的解就是规划问题的最优解，库恩－塔克条件同时也是充分条件，不需要进行二阶条件的检验。

另外，我们对目标函数、约束函数的凹凸性的假定是在整个定义域上的，上述的证明在整个定义域上成立，因此由库恩－塔克条件确定的局部解同时也是全局解。

F.3.拟凹函数（Quasi-concave Function）、拟凸函数（Quasi-convex Function）：

前面在凹函数的讨论中，我们说经济学中，一种常见的凹函数如效用函数，是用来描述通常所认为的人们在消费选择中的一些行为特征，如边际效用递减

等。不过，在微观经济学中对典型消费者的标准假设是，效用函数为拟凹的。

在理论上，更弱的假设可以增加理论的一般性和解释力。凹的效用函数已经能够反映对人们消费行为的很多直观认识，而拟凹函数的引入则可能放宽了这个假设，而足以满足理论上的要求。关于这一点，我们将在下一节关于最优化的解的理论中给出解释。

我们给出拟凹函数的定义：

函数 $f(x)$，对定义域 S（凸集）上任意两点 $x_1, x_2 \in S, \theta \in [0,1]$，如果有

$f[\theta x_1 + (1-\theta)x_2] \geq \min\{f(x_1), f(x_2)\}$，则我们称函数 $f(x)$ 为拟凹函数。

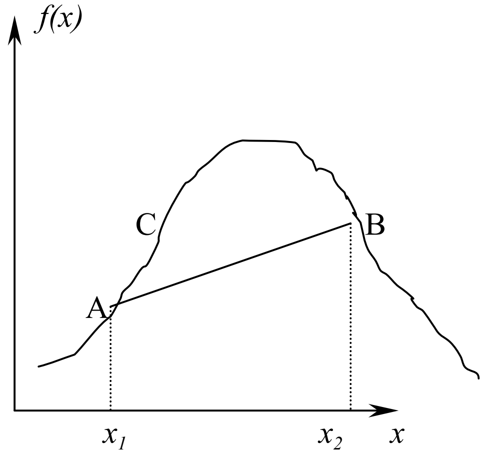


图 F.4

直观的，从图形上看，函数 $f(x)$ 为拟凹表示线段 x_1x_2 之间的点的函数值要高于点 A，或者说曲线 ACB 之间的点都高于点 A。显然，当函数 *f(x)* 是凹函数，曲线呈一个倒置的锅，则上述性质是满足的。从这一点看，凹函数一定是拟凹函数（代数的证明只要利用两者的定义即得）。但是，这不是必要的。如上图，在曲线 AC 段，函数是凹的；而在 CB 段，函数是凸的。这说明拟凹函数的概念要比凹函数更弱。如下图，凹函数、线性函数和凸函数三种情况都满足上述拟凹函数的定义，因此，简单起见，从比较的角度，我们可以将拟凹函数理解为凹函数、线性函数和凸函数的一个简单组合，当然不仅限于此。

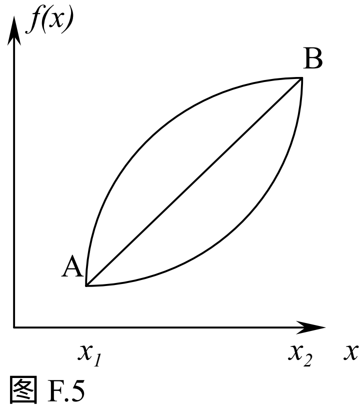


图 F.5

关于拟凹函数的一个重要性质是，如果函数 $f(x)$ 是拟凹的，则当且仅当集合 $\overline{S} = \{x \mid f(x) \geq k\}$ 是凸集，其中 k 是任意常数。我们称集合 $\overline{S}$ 为函数 $f(x)$ 的<u>上</u>

等值集（Upper Contour Set）。充分性利用定义马上得到，而必要性则可通过反证法得出，这里从略。拟凹函数与其上等值集为凸集是等价的，因此在一些教科书中，拟凹函数常常是通过其上等值集为凸集来定义的。

与拟凹函数相对的概念是拟凸函数，定义为：

函数 $f(x)$，对定义域 S（凸集）上任意两点 $x_1, x_2 \in S, \theta \in [0,1]$，如果有

$f[\theta x_1 + (1-\theta)x_2] \le \max\{f(x_1), f(x_2)\}$，则称函数 $f(x)$ 是拟凸的。

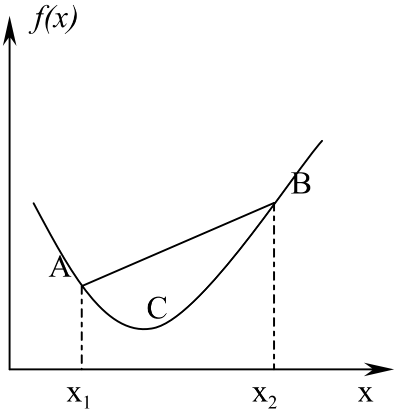


图 F.6

直观的看，函数 $f(x)$ 是拟凸的表示曲线 ACB 之间的点都低于 B 点。显然，如果函数 $f(x)$ 是凸的，则图形如一个正放的锅，弦在曲线上面，而弦上的点本身满足上述性质，因而一定是拟凸的。代数的证明只要利用两者的定义即得。但反向则不一定成立，如前面同是单调函数的凹函数、线性函数、凸函数的图形中，同样满足拟凸函数的定义，即拟凸函数可以是凹函数，也可以是凸函数。

与拟凹函数相对，拟凸函数也有一个等价定义：如果函数 $f(x)$ 是拟凸的，当且仅当集合 $\underline{S} = \{x \mid f(x) \le c\}$ 是凸集，我们称集合 $\underline{S}$ 为函数 $f(x)$ 的下等值集（Lower Contour Set）。证明从略。

现将关于拟凹函数、拟凸函数的一些性质不加证明的归纳如下：

i) 如果函数 $f(x)$ 是凹（凸）的，则 $f(x)$ 也一定是拟凹（凸）的；反之则不成立；

ii) 如果函数 $f(x)$ 是拟凹（凸）的，则 $-f(x)$ 一定是拟凸（凹）的；

iii) 线性函数 $f(x)$ 既是拟凹的，也是拟凸的；

iv) 拟凹函数等价于凸的上等值集；拟凸函数等价于凸的下等值集。

另外，值得注意的是，与凹（凸）函数不同，拟凹（凸）函数的非负线性组合不是拟凹（凸）函数。

G.最优化问题的解

在前面的讨论中，我们假定最优规划问题的解是存在的。但事实上并非总是如此的，而只有当目标函数和约束函数满足一些基本的性质和条件时，最优化问题的解才存在。我们下面的工作就是要寻找这样的条件。另外，解的唯一性也是一个很重要的问题，我们也将讨论唯一性成立的条件。

在进一步讨论之前，我们先给出一些基本的概念：

G..1.基本概念

G.1.1 紧集：

如果度量空间 X 的任何一个序列都有收敛的子列，则称 X 是紧的。

紧集是一个很抽象的概念，与收敛的性质紧密相连。不过对一个具体的度量空间，欧式空间 R^n 中，我们有结论：有界闭集是紧集。

G.1.2.函数的连续：

$\forall \varepsilon > 0, \exists \delta > 0,$ 当 $|x - x_0| < \delta$ 时，如果有 $|f(x) - f(x_0)| < \varepsilon$，则称函数 $f(x)$ 在点 x_0 处连续。

换一种定义方式，如果有 $\lim\limits_{x \to x_0} f(x) = f(x_0)$，则称 $f(x)$ 在点 x_0 处连续。这两种定义方式都表达了函数连续的性质，即当自变量 x 逐渐靠近点 x_0 时，曲线上的点 $y = f(x)$ 也愈加接近点 $y_0 = f(x_0)$。

如果把这种性质从定义域 X 到值域 Y 的函数，推广到更加一般的空间 X 到空间 Y 的映射，我们有连续映射的定义：

$\forall \varepsilon > 0, \exists \delta > 0,$ 当 $d(x, x_0) < \delta$ 时，$x, x_0 \in X$，$d(x, x_0)$ 表示空间 X 中两点的一个度量，如果有 $d(f(x), f(x_0)) < \varepsilon$，其中 $f(x), f(x_0) \in Y$，则称映射 $f: X \mapsto Y$ 在点 x_0 处连续。

有一个很重要的定理将紧集和连续映射联在一起：

G.1.3Weierstrass Theorem （韦氏定理）：

如果 X 是紧集，映射 $f: X \mapsto Y$ 是连续的，则 $f(X)$ 也是紧集。

我们刚讲过，欧氏空间中的有界闭集是紧集，如果 X 是欧氏空间中的有界闭集，则其肯定是紧集；又映射 $f: X \mapsto Y$ 是连续的，则 $f(X)$ 将是有界闭集（也是紧集）。

而跟欧氏空间中的有界闭集相关的一个很重要的原理是确界原理：

有界闭集一定有确界。

比如讲区间[0,1]，显然是一个有界且闭的集合，因此有上确界 1 和下确界 0。

G.2.解的存在性和唯一性

G.2.1.存在性定理

将韦氏定理和确界原理结合起来，如果 X 是欧氏空间中的一个有界闭集，映射 $f:X\mapsto Y$ 是连续的，且 Y 为实数 R，即映射 $f:X\mapsto Y$ 实际上就是函数，则 $f(X)$ 是实数域 R 上的一个有界闭集；而按确界原理，则 $f(X)$ 一定有上确界 $\sup f(X)$ 和下确界 $\inf f(X)$，即函数 $f(x),x\in X$ 的最大值和最小值都是存在的。

介绍了这些概念和定理，到底和我们讨论的最优规划解的问题有何关系？正如在本文第一节中所讲的那样，我们可以将下面最优规划问题的解 $\bar{x}$ 看作在

$$
\begin{aligned}
&\max_x f(x)\\
&s.t.\\
&g(x)\geq c\\
&x\geq 0
\end{aligned}
$$

可行集 $S=\{x\mid g(x)\geq c,x\geq 0\}$ 上找一个点 $\bar{x}$ 使得函数 $f(x)$ 在这个点所达到的值最大。如此，如果将此处的集合 S（当然不能是空集）看作上面讨论中的 $X\subset R$，则只要函数 $f(x)$ 在 R 上是连续的，则这个点 $\bar{x}$ 一定能找到。这就是所谓的解的存在性定理：

当集合 S 是非空的有界闭集，且函数 $f(x)$ 是连续的，则最优规划问题的解一定存在。

在二维的情形下，我们可以用图示来说明这个问题。

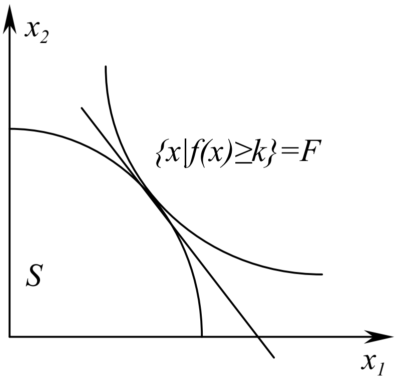


图 G.1

在平面 x_1—x_2 中，区域 S 表示规划问题的可行集 $\{x\mid g(x)\geq c,x\geq 0\}$，用点形图

表示，而斜纹图则表示集合 $\{x \mid f(x) \geq k\}$，我们用 F 来表示，其中 $f(x)=k$ 表示无差异曲线。因此最优规划问题就变成在区域 S 中寻找一点 $\bar{x}$，使得无差异曲线 $f(x)=k$ 可以达到最高的高度。存在性定理告诉我们，只要 S 是非空有界的闭集，且函数 $f(x)$ 是连续的，即无差异曲线不间断，则点 $\bar{x}$ 一定可以找到。

G.2.2 解的唯一性定理：

上面我们讨论了最优规划的解的存在性，从结论看，目标函数函数和可行集满足了一定的条件，规划问题的解就一定存在，所以我们在求解过程或者定性的分析中就无需考虑解是否存在的问题了。

与之相关的另一个重要问题是解的唯一性。我们知道，最优规划的解是一个关于参数的函数，而这些参数表达了外部的环境，如果最后的解只有一个，也就是说，人们的最优选择是与这些参数的一个函数，则对于经济学家而言，则可以通过研究参数的变化（代表外部环境变化）所引起的最优选择的变化来理解人们的行为方式（这种方法即比较静态分析，我们在后面将会讨论）。比如，在消费者行为选择中，商品的消费量是一个收入和价格的函数，经济学家通过研究收入或价格变化而引起的消费量的变动来理解消费者的行为。但如果最优规划的解不是唯一的，即最优选择不是参数的函数，而是与参数的一组对应关系，则与前者相比此处的分析要复杂许多，不同的解之间的差异很难确定是由于参数的变动还是由于不同最优解之间的跳跃。因此，对于经济学家而言，解的唯一性就显得十分重要。

显然，解的唯一性的条件肯定要比存在性条件更强。上图中的例子是解唯一的情形，直观的看，似乎与集合 S 的边际 $g(x)=c$ 和集合 F 的边际 $f(x)=k$ 的凹凸性有关。在进一步讨论之前，我们先给出定理，然后作简单的说明。

解的唯一性定理：

集合 S 是非空有界闭的凸集，且函数 $f(x)$ 是连续的，集合 F 也是凸集，如果凸集 S 和集合 F 两者至少有一个是严格凸的，则最优规划的解唯一。

只要利用简单的反证法，定理的证明很容易得到，这里从略。

当集合 S 和集合 F 都为凸集时，一定可以找到一条直线 l 将这两个集合分开，即集合 S、F 分处直线 l 的两边（只要 k 不断变化，集合 F 可以不断向上移动，直到存在直线 l 将两区域分离）。显然，可能的情形有四种，只有第二种情形下集合 S 和 F 都是非严格凸的（有部分是线性的），此时交点有很多个，即在重合的边界上，最优解有无数个。而在其余三种情形（集合 S、F 至少有一个是严格凸的），最优解唯一。

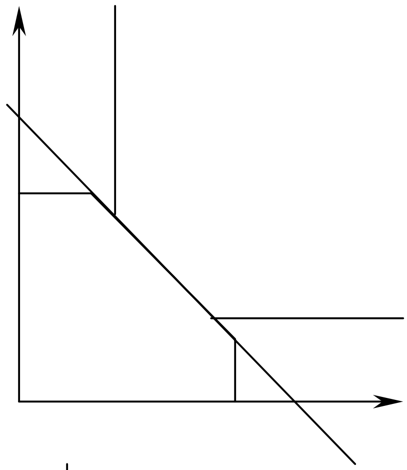
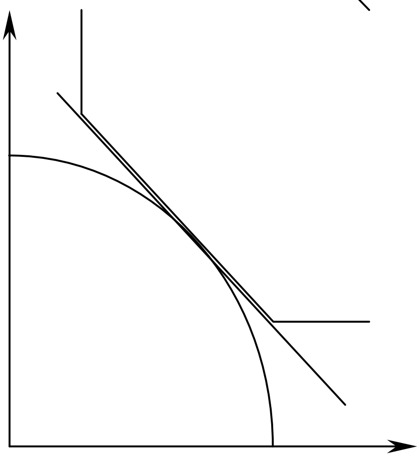


图 G.2



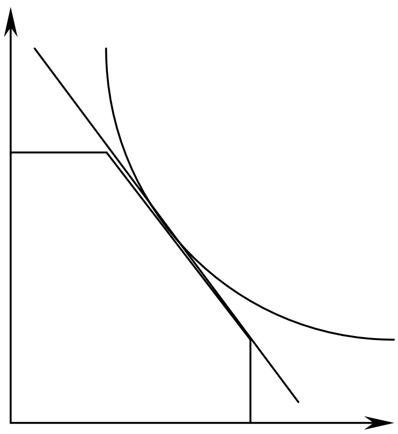


图 G.3　　　　图 G.4

G.3.分离

上述集合 S、F 可用一条直线分开的性质，我们用一个定理来归纳：
分离定理(Seperation Theorem)：

如果非空集合 S、F 是凸集，且没有共同的内点，则存在直线 l：$px = b$ 将集合 S、F 分开，且有：

i) $px \le b, \forall x \in S$

ii) $px \ge b, \forall x \in F$

当选择变量是三维的情形，则直线 l：$px = b$ 是一个平面，我们称之为分离平面；而当三维以上的情形，则相应的称之为分离超平面。

进一步的，我们来考察当目标函数、约束函数满足什么样的条件时，集合 S、F 是凸集。根据前面对拟凹函数、拟凸函数的讨论，我们知道：

函数 $f(x)$ 为拟凹 $\Leftrightarrow$ 其上等值集 $\{x \mid f(x) \ge k\}$ 是凸集，即集合 F 为凸集等价

于目标函数为拟凹函数；

函数 $g(x)$ 为拟凸 $\Leftrightarrow$ 其下等值集 $\{x \mid g(x) \le c\}$ 为凸集，即集合 S 为凸集等价于目标函数为拟凸函数。

由上述讨论，及分离定理，我们可以将一个最优规划问题分解成两个问题：

如果目标函数 $f(x)$ 为拟凹函数， $g(x)$ 是拟凸函数， $\bar{x}$ 为最优规划问题 $\max\limits_x f(x), x \in \{x \mid g(x) \le c, x \ge 0\}$ 的解，则当且仅当存在一个非零向量 p ，使得：

i) $\bar{x}$ 是最优规划问题 $\max\limits_x px$， $x \in \{x \mid g(x) \le c, x \ge 0\}$ 的解；

ii) $\bar{x}$ 是最优规划问题 $\min\limits_x px$， $x \in \{x \mid f(x) \ge k\}$ 的解。

另外，解的唯一性定理也可以重新阐述：

目标函数 $f(x)$ 是连续拟凹的，约束函数 $g(x)$ 是拟凸函数，如果函数 $f(x), g(x)$ 两者至少有一个是严格的，则最优规划的解唯一。

由前面关于函数凹凸性的讨论，凹函数一定是拟凹的，凸函数一定是拟凸的，则下列命题也成立：

目标函数 $f(x)$ 是连续凹函数，约束函数 $g(x)$ 是凸函数，如果函数 $f(x), g(x)$ 两者至少有一个是严格的，则最优规划的解唯一。

上述定理告诉我们，最优规划的解的性质取决于目标函数和约束函数的凹凸性。这不禁让我们联想起前面提到的凹规划：如果目标函数是凹函数，约束函数是凸函数，则由一阶必要条件确定的点就是规划问题的解，而且解是全局的。这一点就是集合 S、F 的相切之点，从图上可以很直观地得到。如果目标函数、约束函数在整个定义域上分别是凹的、凸的，则集合 S、F 都是凸集，因此由一阶条件确定的点就是规划问题的解，并且由于凸性是整个集合上的，因此解是全局的。

H.比较静态分析

H.1.基本思想

以上我们讨论了最优规划问题的基本概念、解的求解方法（等式约束问题的拉格朗日方法、非线性规划的库恩－塔克条件）、二阶条件，和目标函数、约束函数满足一定凹凸性的规划问题的解的特性（凹规划）；以及规划问题的解的一般性问题。

现在我们假定这个代表某个经济问题的规划的解已经求出，下面的一个问题是：求出解有什么用？例如，在消费者的选择行为中，如两个商品的情形，效用函数为 $u(x_1,x_2)=\alpha\ln x_1+\beta\ln x_2$ ，且 $\alpha+\beta=1,\alpha,\beta>0$ ，预算约束为 $p_1x_1+p_2x_2=I$ ，其中 p_1,p_2,I 分别为两种商品的价格及收入。利用拉格朗日方法我们可以很容易的求得：

$$\overline{x_1}=\frac{\alpha I}{p_1},\overline{x_2}=\frac{\beta I}{p_2}$$

此解可以告诉我们，当价格、收入及偏好一定的情况下消费者选择的两种商品的数量。但是对于经济学家而言，仅仅这些结果是不够的。正如前面所提到的，他们更加关注当价格发生变化或收入发生变化了，消费者对两种商品的消费量会如何变动。如商品 1 的价格上升了一倍，变为 $2p_1$，则消费量为原来的一半，$\overline{x_1}'=\frac{1}{2}\overline{x_1}$ ，经济学家通过比较 $\overline{x_1}',\overline{x_1}$ 来理解消费者的行为。当其中一个参数发生变化，其余不变，比较前后不同均衡解的方法来理解这个参数所表示的外界条件对人们行为的影响就是比较静态均衡分析。

进一步的，如上面的例子中，如果均衡解 $\overline{x}$ 可以表达为一组参数的函数，则衡量一个因素（一个参数变动，其余保持不变）对均衡解的影响可以通过均衡解直接对此参数求偏导数即可，如在上例中，价格对商品消费量的影响，可以由 $\overline{x_1}$ 直接对 p_1 求偏导数 $\frac{\partial\overline{x_1}}{\partial p_1}=-\frac{\alpha I}{p_1{}^2}<0$ 得出，即价格越高，消费量越少。对表示收入的参数 I 和偏好的参数 α 对消费量的影响的分析类似。

总结一下，在这种情形下的比较静态分析的步骤：第一步，求解出均衡解，均衡解为表示外界条件的一组参数的函数；第二步，通过均衡解对参数求偏导数表示一个条件变化时对选择的影响，以此来表示人们的行为。

当目标函数不是以具体的函数形式出现的，如上例中，效用函数没有具体给出，而只有一些定性的性质，此时均衡解就不能表达成参数的显式函数的形式。此时，要研究一个参数的变动对选择变量的影响，就不能通过简单的求偏导数的方式得出。怎么办？

一般的，以等式约束为例，回想求解的过程，我们通过构造拉格朗日函数

将约束问题化为无约束问题，通过拉格朗日对选择变量和拉格朗日乘子的一阶导数为零来确定解，即一阶必要条件。如果能将解直接求出，即是上面所讲的情形；如果不能直接求出，则此一阶条件已经包含了最优解的性质，或者讲，用隐函数（即一阶条件）的方式定义了解与参数的关系。利用我们前面所讲的有关隐函数的知识，可以通过克莱姆法则（多维的情况下）求出均衡解对参数的偏导数。

H.2.一般方法

如效用函数是以一般形式$u(x_1,x_2)$出现的，则进行比较静态分析的第一步，构造拉格朗日函数，并求解一阶条件：

$$L(x_1,x_2,\lambda)=u(x_1,x_2)+\lambda(I-p_1x_1-p_2x_2)$$

$$\frac{\partial L}{\partial x_1}=u_1-\lambda p_1=0$$

$$\frac{\partial L}{\partial x_2}=u_2-\lambda p_2=0$$

$$\frac{\partial L}{\partial \lambda}=I-p_1x_1-p_2x_2=0$$

对上式进行全微分，得

$$u_{11}dx_1+u_{12}dx_2-p_1d\lambda=\lambda dp_1$$

$$u_{21}dx_1+u_{22}dx_2-p_2d\lambda=\lambda dp_2$$

$$-p_1dx_1-p_2dx_2=-dI+x_1dp_1+x_2dp_2$$

写成矩阵的形式，即

$$\begin{bmatrix} u_{11} & u_{12} & -p_1 \\ u_{21} & u_{22} & -p_2 \\ -p_1 & -p_2 & 0 \end{bmatrix}\begin{bmatrix} dx_1 \\ dx_2 \\ d\lambda \end{bmatrix}=\begin{bmatrix} \lambda dp_1 \\ \lambda dp_2 \\ x_1dp_1+x_2dp_2-dI \end{bmatrix}$$

其 x_1, x_2, λ取达到最优解时值，即满足一阶条件，$\overline{x}=(\overline{x_1},\overline{x_2}),\lambda=\overline{\lambda}$

以$\overline{x_1}$为例，比较静态结果为：

$$\frac{\partial \overline{x_1}}{\partial p_1}=\frac{-\lambda p_2^{\,2}+x_1(p_1u_{22}-p_2u_{12})}{D},\quad dp_1\neq 0,dp_2=0,dI=0$$

$$\frac{\partial \overline{x_1}}{\partial p_2}=\frac{-\lambda p_1^{\,2}+x_2(p_2u_{11}-p_1u_{21})}{D},\quad dp_2\neq 0,dp_1=0,dI=0 \qquad (2.30)$$

$$\frac{\partial \overline{x_1}}{\partial I}=\frac{-(p_1u_{22}-p_2u_{12})}{D},\quad dI\neq 0,dp_1=0,dp_2=0$$

所有的取值都在均衡解（$\overline{x}=(\overline{x_1},\overline{x_2}),\lambda=\overline{\lambda}$）取得，分析价格$p_1, p_2, I$对$x_1, x_2$的影响(比较静态意义)，看上式偏导的符号，若为正，则表明消费量的变动与价格(或收入)同方向变动；若为负，则表明两者是负相关的。故判断比较静态结

果，主要分析上述偏导的符号；而判断符号看分子与分母，而分母 D 与最优解的二阶条件有重要关系（从前面的例子看，D>0）。

更一般的，设一个系统有 n 个选择变量（内生变量）和 m 个参数（外生变量），满足一阶条件：

$$f^1(x_1,\cdots,x_n,\alpha_1,\cdots,\alpha_m)=0$$

$$\cdots\ \cdots$$

$$f^n(x_1,\cdots,x_n,\alpha_1,\cdots,\alpha_m)=0$$

我们知道，如果以一阶偏导数组成的雅可比矩阵的行列式

$$D=\begin{vmatrix} f_1^1\cdots f_n^1 \\ \cdots\ \cdots \\ f_1^n\cdots f_n^n \end{vmatrix}\neq 0$$

则由隐函数定理得，选择变量可以表示成一组参数的显式函数，即：

$$x_i=h^i(\alpha_1,\cdots,\alpha_m), i=1,\cdots,n$$

每个参数的变化对均衡解的影响可以用上面函数的偏导数 $\frac{\partial x_i}{\partial \alpha_j}=h_j^i$ 来表示，对一阶条件全微分，得：

$$\begin{aligned} &f_1^1dx_1+\cdots+f_n^1dx_n+f_{n+1}^1d\alpha_1+\cdots+f_{n+m}^1d\alpha_m=0 \\ &\cdots\cdots\qquad\cdots\cdots \\ &f_1^ndx_1+\cdots+f_n^ndx_n+f_{n+1}^ndx_1+\cdots f_{n+m}^nd\alpha_m=0 \end{aligned}\tag{2.26}$$

写成矩阵的形式为

$$\begin{bmatrix} f_1^1\cdots f_n^1 \\ \cdots\cdots \\ f_1^n\cdots f_n^n \end{bmatrix}\begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix}=\begin{bmatrix} -(f_{n+1}^1d\alpha_1+\cdots+f_{n+m}^1d\alpha_m) \\ \cdots\qquad\cdots \\ -(f_{n+1}^nd\alpha_1+\cdots+f_{n+m}^nd\alpha_m) \end{bmatrix}\tag{2.27}$$

由比较静态意义，即假定其它不变，$d\alpha_j\neq 0$, $d\alpha_k=0$，$k\neq j$，由克莱姆法则得 $\frac{dx_i}{d\alpha_j}=\frac{D_{ij}}{D}$，$D_{ij}$ 为雅可比矩阵行列式 D 的 i 列由 $(-f_{n+j}^1,\cdots,-f_{n+j}^n)^T$ 代替而成的行列式。

I.包络定理(Envelope Theorem)

与比较静态分析相关的一个重要工具是包络定理。比较静态分析的思想是在其它条件（参数）不变得前提下，研究单个参数的变化对均衡解的影响，以此来表达决策者的行为。而另一类重要的问题是，我们常常要考虑此参数的变化对目标函数（最大值）的影响，如一商品价格的变化对消费者的效用的影响，一投入要素价格的变化（或要素禀赋的变动）对厂商收入（或利润）的影响，都属此类情形。在进入正式的讨论之前，我们先介绍一个概念：最大值函数

I.1.最大值函数(Maximum Value Function)

在最优规划问题 $\max\limits_{x} f(x), x \in \{x \mid g(x)=c\}$ 中，我们知道最优解是与参数有关的函数，即 $\bar{x}=x(c)$，因此在最优解 $\bar{x}$ 处，目标函数的值为：

$$f(\bar{x}) = f(x(c)) \equiv V(c)$$

即目标函数在最优解处的解也是与参数 c 有关的函数，我们定义为 $V(c)$，称之为最大值函数。

更一般的，如果参数不仅出现在约束，而且也出现在目标函数，我们有：

$$V(\alpha) = \max_{x}\{f(x,\alpha) \mid g(x,\alpha)=0\} = f(x(\alpha),\alpha)$$

其中 $\bar{x}=x(\alpha)$ 为参数为 α 时的选择变量的最优解，或者称最优反应。因此，根据定义，如果对一任意 x'，则有 $f(x(\alpha),\alpha) \ge f(x',\alpha)$，等号当且仅当 $x'=\bar{x}=x(\alpha)$ 时取得。 因此，最大值函数 $V(\alpha)=f(x(\alpha),\alpha)$ 与函数 $f(x,\alpha)$ 是有区别的，一般而言，$V(\alpha) \ge f(x,\alpha)$，当且仅当 $x=x(\alpha)$ 时，$V(\alpha)=f(x,\alpha)$，即 x 为参数 α 的最优反应时取得等号。我们用一简单的图示来说明这一关系。

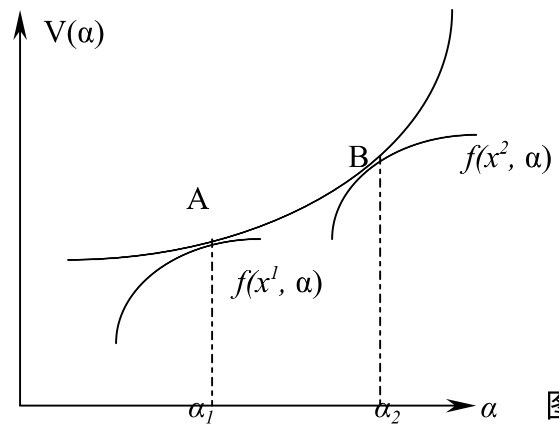


图 I.1

上图的横轴为 α，纵轴为 V(α)，图形上部的曲线为 V(α)，另两条曲线分别为 $f(x^1, \alpha)$、$f(x^2, \alpha)$，即当选择变量 x 分别为 x^1、x^2 时函数 $f(x, \alpha)$ 与 α 之间的关系。正如我们前面所讨论的：

$V(\alpha) \ge f(x^1, \alpha)$，当且仅当在 α_1 处，$x^1 = x(\alpha_1)$，即 x^1 是参数 α_1 的最优反应，等号取得，或者说曲线 V(α)与 $f(x^1, \alpha)$ 相切于点 A，其余都在曲线 $f(x^1, \alpha)$ 上。曲线 V(α)与 $f(x^2, \alpha)$ 的关系也是一样，两者相切于点 B。可以想象一下，在上述平面中有一族曲线 $f(x, \alpha)$，不同的 x 就代表一条曲线，而每一条曲线都在曲线 V(α)的下方，而只与其相切于一点 $x=x(\alpha)$。我们将所有这样的点连起来，就得到曲线 V(α)，即为函数 $f(x, \alpha)$ 的包络线。

I.2.包络定理

包络曲线 V(α)与曲线 $f(x^1, \alpha)$ 相切于点 A，即两曲线在点 A 的斜率相等，用代数表达为：

$V_\alpha = f_\alpha(x^1, \alpha)$，在 $\alpha = \alpha_1$ 处取值

此等式即为包络定理。

更加一般的，对于最优规划问题：

$$\max_x f(x, \alpha)$$
$$s.t.$$
$$g(x, \alpha) = 0$$

其中选择变量 x 为 n 维向量，参数 α 为 m 维向量，包络定理可以为：

$$\frac{\partial V}{\partial \alpha_k} = L_{\alpha_k}(x, \alpha)\Big|_{\bar{x}}$$

即参数 α_k 对最大值函数（目标函数的最大值）的影响，就等于拉格朗日函数直接对参数 α_k 求偏导数，并在最优解 $\bar{x}$ 处取值。

下面，我们来严格地证明这个定理：

首先，构造拉格朗日函数 $L(x, \alpha; \lambda) = f(x, \alpha) - \lambda g(x, \alpha)$，一阶条件为：

$$\frac{\partial L}{\partial x_i} = f_i - \lambda g_i = 0, i = 1, \cdots, n$$
$$\frac{\partial L}{\partial \lambda} = g(x, \alpha) = 0$$

假设求得最优解 $\bar{x} = x(\alpha)$，则最大值函数 $V(\alpha) = f(x(\alpha), \alpha)$，两边对 α_k 求导，并在最优解 $\bar{x}$ 处取值：

$$\frac{\partial V}{\partial \alpha_k} = \sum f_i \frac{\partial x_i}{\partial \alpha_k} + f_{\alpha_k} = \sum \lambda g_i \frac{\partial x_i}{\partial \alpha_k} + f_{\alpha_k}$$

第二个等式利用了一阶条件。

由约束 $g(x,\alpha)=0$，等式两边对 α_k 求导，并在最优解 $\bar{x}$ 处取值：

$$\sum g_i \frac{\partial x_i}{\partial \alpha_k} + g_{\alpha_k} = 0$$

代入上式，得：

$$\frac{\partial V}{\partial \alpha_k} = f_{\alpha_k} - \lambda g_{\alpha_k} = L_{\alpha_k}(x,\alpha)\Big|_{\bar{x}}$$

从最大值函数的定义看，即 $V(\alpha) = f(x(\alpha),\alpha)$，参数对最大值的函数的影响，来自两部分，第一部分来自参数对目标函数的直接影响，用目标函数对参数的偏导数表示，即直接效应；第二部分来自通过影响选择变量 x 间接的影响目标函数，即间接效应。从证明的过程看，由于拉格朗日函数一阶条件为零，因此间接的效应为零，所以参数对目标函数最大值的影响，就等于拉格朗日函数对参数的偏导数，并在最优解处取值。这就是上面图示的含义。

上述讨论的是求最大值时的情形，在最小化问题中，只要将不等号方向改变就可以了。图示变为：

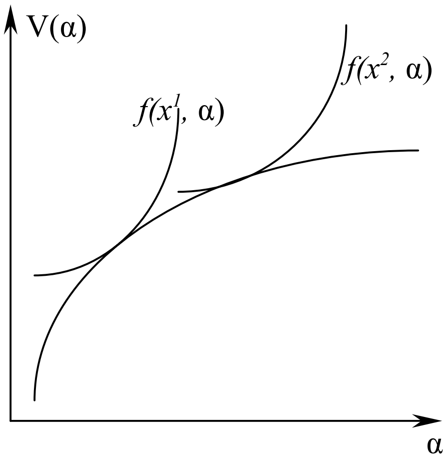


图 I.2

这样的图形让我们不禁联想起成本函数的性质：长期平均成本曲线是短期平均成本曲线（族）的下包络线。

I.3.拉格朗日乘子的含义

前面在等式约束极值问题的求解中，我们是通过引入拉格朗日乘子λ，将一个约束极值问题转化成无约束问题，而拉格朗日乘子也作为解的一部分求出。虽然在拉格朗日方法的求解中，拉格朗日乘子是作为操作性的工具引入的，但实际上它有着很丰富的经济含义。在这里，我们通过包络定理的一个应用，给出简单的说明。

在消费者行为选择中，消费者在预算约束下最大化其效用，即

$$\max_{x} u(x_1, x_2)$$

$$s.t.$$

$$p_1 x_1 + p_2 x_2 = I$$

构造拉格朗日函数 $L(x_1, x_2, \lambda) = u(x_1, x_2) + \lambda[I - p_1 x_1 - p_2 x_2]$，则利用包络定理，有：

$$\frac{\partial V}{\partial I} = \frac{\partial u(\bar{x})}{\partial I} = L_I\big|_{\bar{x}} = \lambda$$

即拉格朗日乘子就等于效用函数在最优解处对收入的偏导数，也就是在最优解处，增加一个单位的收入所带来的效用的增加；或者讲，拉格朗日乘子就是用在最优解处用效用来衡量的收入的价值。所以，我们可以称其为收入的边际效用。

更一般的，在规划问题：

$$\max_{x} f(x)$$

$$s.t.$$

$$g(x) = c$$

如果我们将选择变量 x 看作是产品，目标函数为企业的收入，约束表达投入的限制，比如对该企业而言，单一投入要素——劳动的总量是一定的，为 c，则根据包络定理，我们同样可以得到，此问题中的拉格朗日乘子λ为最优解处目标函数的对要素禀赋 c 的偏导数，即：

$$\frac{\partial V}{\partial c} = \frac{\partial f(\bar{x})}{\partial c} = L_c\big|_{\bar{x}} = \lambda$$

也就是说，拉格朗日乘子的值等于在最优解处增加一个单位的劳动投入所带来的收入。如果此时企业要增加一单位的劳动所需的成本小于λ，则企业将获利；相反，如果成本大于λ，则企业将亏损。因此，对于企业而言，拉格朗日乘子λ将是衡量劳动投入的一种价格，我们称之为影子价格。当然，上述规划问题中拉格朗日乘子的经济解释不仅限于企业生产的例子，一般的，将其看作是在最优解处以目标函数衡量的参数 c 的影子价格。

I.4.包络定理的应用——消费者选择理论

下面，我们通过消费者选择的例子来说明包络定理的应用。以二种商品为例：

$$\max_{x} u(x_1, x_2)$$

$$s.t.$$

$$p_1 x_1 + p_2 x_2 = I$$

最优解 $\bar{x}$ 将是与参数 p_1, p_2, I 有关的函数，即：

$$\overline{x_1} = x_1(p_1, p_2, I)$$
$$\overline{x_2} = x_2(p_1, p_2, I)$$

我们称之为马歇尔需求函数，分别记为 $D_1(p_1, p_2, I), D_2(p_1, p_2, I)$。

此时，效用函数的值为：

$$u(\overline{x_1}, \overline{x_2}) = u(x_1(p_1, p_2, I), x_2(p_1, p_2, I)) \equiv V(p_1, p_2, I)$$

即最大值函数，此处我们称之为间接效用函数，以区别于直接定义在商品的消费量上的效用函数 u。

构造拉格朗日函数 $L(x, \lambda, p, I) = u(x_1, x_2) + \lambda[I - p_1 x_1 - p_2 x_2]$，则利用包络定理：

$$V_I = \lambda$$
$$V_{p_1} = -\lambda x_1 = -\lambda D_1$$

第一个式子表明，拉格朗日乘子 λ 就等于收入的边际效用；将 λ 从第一式中代入第二式，得：

$$\overline{x_1} = x_1(p_1, p_2, I) = -V_{p_1} / V_I$$

我们称之为罗伊恒等式（Roy's Identity）。

考虑另一个对称的问题，在一定的效用水平下，消费者支出最小化：

$$\min_{x_1, x_2} p_1 x_1 + p_2 x_2$$
$$s.t.$$
$$u(x_1, x_2) = u$$

同样的，最优解是关于一组参数 p_1, p_2, u 的函数，即：

$$\overline{x_1} = x_1(p_1, p_2, u)$$
$$\overline{x_2} = x_2(p_1, p_2, u)$$

我们称之为希克斯需求函数（或补偿需求函数），分别记为 $H_1(p_1, p_2, u), H_2(p_1, p_2, u)$。

此时，支出的值为：

$$p_1 \overline{x_1} + p_2 \overline{x_2} = p_1 x_1(p_1, p_2, u) + p_1 x_2(p_1, p_2, u) \equiv E(p_1, p_2, u)$$

即最大值函数，此处我们称之为支出函数，类同于前面的间接效用函数 V。

构造拉格朗日函数 $L(x, \lambda, p, u) = p_1 x_1 + p_2 x_2 + \lambda[u - u(x_1, x_2)]$，则利用包络定理：

$$E_u = \lambda$$
$$E_{p_1} = x_1 = H_1(p_1, p_2, u)$$

第一个式子表明，拉格朗日乘子λ就等于支出函数对效用 u 的偏导数；第二式表明，希克斯需求函数就等于支出函数对价格的偏导数，我们将这个关系称之为谢菲德引理（Shepherd Lemma）。

我们将这两个问题放在一起，是因为它们是一个问题的两个方面，在许多地方蕴含着相关的特征。看下图：

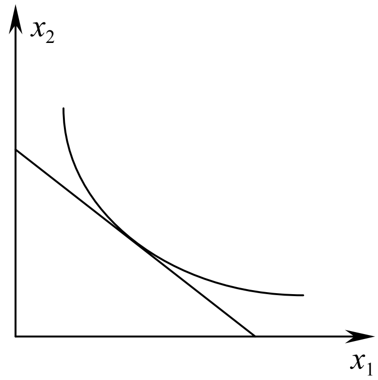


图 I.3

前面一个问题等价于，当预算线一定的情形下，在其上找一点使得无差异曲线达到最高，最优解在两者相切之点；而后一问题可以表达为，在无差异曲线一定的情况下，在其上寻找一点使得相应的预算线更加靠近原点，最优解也在两者相切之点。因此，当预算线不变的情况下，找到其上一点能够达到的一条的最高的无差异曲线，此时切点记为$\bar{x}$；若此时保持这条无差异曲线不变，选择在其上一点所能达到的最靠近的预算线，此时相切之点记为$\bar{x}'$。显然，此时的预算线就是前一问题中保持不变的那条预算线，切点也为同一个，即$\bar{x}'=\bar{x}$。反映在规划中，当后一问题中的参数u取 V 时，均衡解相同，即

$$H(p_1,p_2,u)=D(p_1,p_2,I),\text{当}\ u=V(p_1,p_2,I)\ \text{时。}$$

上述过程也可倒过来，当 $I=E$ 时，两个问题的均衡解也相同，即

$$H(p_1,p_2,u)=D(p_1,p_2,E(p_1,p_2,u))$$

上式两边对参数p_1求导，有：

$$\frac{\partial H_1}{\partial p_1}=\frac{\partial D_1}{\partial p_1}+\frac{\partial D_1}{\partial I}\frac{\partial E}{\partial p_1}=\frac{\partial D_1}{\partial p_1}+x_1\frac{\partial D_1}{\partial I}$$

第二个等式利用了谢菲德引理，将上式整理，得：

$$\frac{\partial D_1}{\partial p_1}=\frac{\partial H_1}{\partial p_1}-x_1\frac{\partial D_1}{\partial I}$$

所有的取值均在均衡解$\bar{x}$处。上式即为斯勒茨基方程（Slutsky Equation），价格对商品需求的影响有两个效应，第一部分为替代效应；第二部分为收入效应。

参考文献：

Chiang, Alpha C., *Fundamental Methods of Mathematical Economics* (3rd ed.), McGraw-Hill, 1984.（中译本《数理经济学的基本方法》，刘学译，商务印书馆，2001）

Dixit,A., *Optimization in Economic Theory*, Oxford University Press, 1990.

Gravelle, H. & Rees, R., *Microeconomic*s (2nd ed.) , Longman, 1992.

Sydsaeter, Knut et al., *Economists' Mathematical Manual* (3rd ed.), Springer, 1999. (中译本《经济学家数学手册》张涛译，复旦大学出版社，2001)。

高山晟，《经济学中的分析方法》，刘振亚译，中国人民大学出版社，2001。

张军（主编），《高级微观经济学》，复旦大学出版社，2002。

沈永欢等（编），《实用数学手册》，科学出版社，2002。