多学科学术讲座丛书

3

广义变分原理

钱伟长 著

多学科学术讲座丛书

3

广义变分原理

钱伟长著

知识出版社

1985・3・上海

责任编辑：陈荣乐
封面设计：张苏予

多学科学术讲座丛书
3
广义变分原理
钱 伟 长著
知 识 出 版 社 出 版
（上海古北路 650 号）
新华书店上海发行所发行　常熟梅李印刷厂印刷
开本 850×1156 毫米 1/32　印张 11.75　插页 4　字数 285,000
1985 年 3 月第 1 版　1985 年 3 月第 1 次印刷
印数：1—6,000

书号：13214·1022　定价：2.30 元

序　言

1982年冬，参加中国人民政治协商会议全国委员会五届五次会议的民盟小组委员，讨论如何开发盟内智力资源的问题，研究了由民盟中央发动盟内力量，筹备组织全国性的多学科学术讲座的倡议，深得与会委员的积极响应。随后中国民主同盟中央常务委员会决定从1983年暑期开始举办多学科学术讲座，并责成民盟中央文教委员会和科技委员会主持具体工作。经过半年的努力和筹备，讲座于1983年6月12日开学，开学典礼由民盟中央副主席楚图南同志主持，中共中央统战部杨静仁部长、李定副部长亲临指导。李定同志在讲话中肯定了民盟中央举办多学科学术讲座的重要意义，他指出，这不仅符合民盟盟员拥有大量学有专长的专家学者的智力集团的特点，而且也是民主党派工作中的一个创举，深得全国学术界的重视，是党的知识分子政策的胜利。知识分子通过讲座的形式，积极主动地把自己的智慧贡献给祖国的社会主义建设事业，反映了知识分子在党的教育下的又一次觉醒。

民盟中央1983年举办的讲座，分10个专题，每题十讲，每一专题，一般由主讲教授一人负责，也有少数专题由两位或三位主讲教授共同负责。听讲者一千余人。他们来自全国各地，有不少

人本身就是专家教授，有的已年逾花甲仍孜孜不倦为四化远涉千里来到北京，和中青年共同学习和学术交流，这也是前所少见的现象，具体地反映了党中央自三中全会以来的正确领导，动员了千百万知识分子，初步实现了团结奋斗的新局面。

1983 年的讲座内容，涉及人文科学、社会科学和自然科学的一些方面。人文科学方面有朱光潜、黄药眠、常任侠三位教授的《美学和中国美术史》，吴组缃和张毕来教授的《谈红楼梦和红学四论》，商承祚、陆宗达教授的《中国文字学和训诂学》。在社会科学方面，有关梦觉教授的《陈云同志的经济思想》，千家驹教授的《中国经济问题》及徐铸成教授的《新闻艺术》。在自然科学方面，有马大猷教授的《语言通信》，叶培大教授的《光纤理论》，钱伟长教授的《广义变分原理》等。准备在 1984 年参加讲座的主讲教授有唐敖庆、余瑞璜、张文佑、费孝通、陶大镛等 20 余位。

1983 年主讲教授 14 人，平均年龄 76岁，最高年龄 86 岁。他们以年逾古稀的高龄，冒着酷暑，一丝不苟地为学员认真讲解，亲切座谈，深受广大学员的欢迎和爱戴。

这次讲座努力贯彻了百家争鸣的学术方针，提倡严肃的学术民主。主讲教授都能在尊重不同意见的同时，深入透彻地讲解自己的学术观点，有些主讲教授对那些学术上的不正之风，进行了认真、严肃而又满腔热忱的批评和教育。这反映了老一辈学者对当前学术界不正之风的否定而又负责的态度。殷切期望我们的中青年学术接班人，发扬良好学风。有的主讲教授就在同一讲座上，以友好的态度各自讲解分析了双方不同学术观点的矛盾，而不以自己的观点强加给对方，更不以自己的观点来打倒别人的不同观点。这样就能以人之长补己之短，就能达到不同观点的相互融化，逐步走上更高水平的学术境地，从而更有利于为我国社会主义建设服务。

讲座的讲解，都是各主讲教授长期或毕生从事的学术工作，还

有的是当前在四化建设第一线战斗岗位上总结提出的主要贡献。主讲人对讲稿都做了充分的准备，在讲座中又通过听讲学者的学习讨论，再次进行增删修改，才最后定稿。现蒙知识出版社编为丛书，按讲题分别出版。希望本丛书对于我国学术工作，产生有益的影响。

钱 伟 长

1983年7月26日于北京

目　录

第一章　变分原理和拉氏乘子法

§1.1　函数的极值和拉氏乘子法

在研究变分原理之前，让我们先研究某一函数的极值问题，然后再研究某一函数在某些约束条件下的极值问题.

如果 $f(x)$ 在 $x=x_0$ 处满足

$$f(x_0+\Delta x)-f(x_0)<0,\ \text{而且}\ 0<|\Delta x|<\varepsilon \tag{1.1}$$

则称 $f(x)$ 在 $x=x_0$ 处有一相对极大或局部极大，式(1.1)中的 ε 为一正的小量.

如果 $f(x)$ 在 $x=x_0$ 处满足

$$f(x)-f(x_0)\leqslant 0,\ \text{而且}\ a\leqslant x\leqslant b,\ a\leqslant x_0\leqslant b \tag{1.2}$$

则 $f(x)$ 在 $[a, b]$ 中的 $x=x_0$ 处有一绝对极大或全域极大.

如果(1.1)、(1.2)式中第一式的"小于"符号 $<$ 改成"大于"符号 $>$，则我们称 $f(x)$ 在 $x=x_0$ 分别有一相对极小和绝对极小．只要不引起误会，我们经常略去"相对"和"绝对"字样，简称极大或极小.

只要 $f''(x)$ 存在而且在 $[a, b]$ 中是连续的，则用泰勒级数把 $f(x)$ 在 $x=x_0$ 附近展开，得

$$f(x_0+\Delta x)-f(x_0)=f'(x_0)\Delta x+\frac{1}{2}f''(x_0)\Delta x^2+\Phi\Delta x^3 \tag{1.3}$$

其中 $\Phi\Delta x^3$ 为级数的余项，和

$$\Phi=\frac{1}{3!}f'''(x_0+\theta\Delta x),\quad 0<\theta<1 \tag{1.4}$$

设 $f'(x_0)>0$，则在足够小的 Δx 值下，$f(x_0+\Delta x)-f(x_0)$ 决

定于级数(1.3)的首项$f'(x_0)\Delta x$，而且当Δx是正时，$f'(x_0)\Delta x$为正，Δx是负时，$f'(x_0)\Delta x$为负．这不满足(1.1)式所规定的，不论Δx的正负，$f(x_0+\Delta x)-f(x_0)$都必须小于零，所以，这个x_0处，不是$f(x)$的相对极大．同样也不是相对极小．设$f'(x_0)<0$，则当Δx为正时，$f'(x)\Delta x$为负，而Δx为负时，$f'(x)\Delta x$为正，所以同样不满足相对极大和极小的定义．所以，$f'(x_0)<0$时，x_0处也不是极大或极小．

这就证明了，只有当$f'(x_0)=0$时，$x=x_0$处才能满足极大或极小的条件(1.1)，但这只是必要的，还不是充分的．

极大和极小的充分条件决定于$f''(x_0)\Delta x^2$的正负号，当$f'(x_0)=0$时，$f(x_0+\Delta x)-f(x_0)$的正负号决定于$f''(x_0)\Delta x^2$，不论Δx是正是负，这个正负号是由$f''(x_0)$决定的．$f''(x_0)$小于零时，$f(x_0+\Delta x)-f(x_0)$即小于零，根据(1.1)，$f(x)$在$x=x_0$处是相对极小．反之，当$f''(x_0)$大于零时，$f(x_0+\Delta x)-f(x_0)$即大于零，所以，$f(x)$在$x=x_0$处是相对极大．

总结起来说，相对极小的必要条件是$f'(x_0)=0$，而其充要条件是$f'(x_0)=0, f''(x_0)>0$．反之，相对极大的必要条件是$f'(x_0)=0$，而其充要条件则是$f'(x_0)=0$，$f''(x_0)<0$．

如果$f''(x_0)=0$，则相对极大或极小的充分条件还要根据更高次的级数项决定．例如，当

$$f'(x_0)=f''(x_0)=0,$$

而

$$f'''(x_0)\neq 0$$

时，$x=x_0$是$f(x)$的一个拐点．

在习惯上，我们把极大和极小统称极点（extremum），把极大、极小和拐点合在一起，统称驻点(stationary points)．极点上的函数值统称极值，驻点上的函数值统称驻值．

设f为两个变量x_1，x_2的函数．它可以在$x_1^{(0)}$，$x_2^{(0)}$附近展开为泰勒级数

$$
\begin{aligned}
&f(x_1^{(0)}+\Delta x_1,\ x_2^{(0)}+\Delta x_2)-f(x_1^{(0)},\ x_2^{(0)})\\
&\quad=\frac{\partial f}{\partial x_1}(x_1^{(0)},\ x_2^{(0)})\Delta x_1+\frac{\partial f}{\partial x_2}(x_1^{(0)},\ x_2^{(0)})\Delta x_2\\
&\qquad+\frac{1}{2}\left\{\frac{\partial^2 f}{\partial x_1^2}(x_1^{(0)},\ x_2^{(0)})(\Delta x_1)^2+2\frac{\partial^2 f}{\partial x_1\partial x_2}(x_1^{(0)},\ x_2^{(0)})\Delta x_1\Delta x_2\right.\\
&\qquad\left.+\frac{\partial^2 f}{\partial x_2^2}(x_1^{(0)},\ x_2^{(0)})(\Delta x_2)^2\right\}+\sum_{i,j,k=1}^{2}K_{ijk}\Delta x_i\Delta x_j\Delta x_k
\end{aligned}
\tag{1.5}
$$

其 K_{ijk} 为和三阶导数有关的余项. 和一元函数相类似, $f(x_1^{(0)}, x_2^{(0)})$ 为极大或极小的必要条件为

$$
\frac{\partial f}{\partial x_1}(x_1^{(0)},\ x_2^{(0)})=0,\qquad \frac{\partial f}{\partial x_2}(x_1^{(0)},\ x_2^{(0)})=0 \tag{1.6}
$$

为了列出其充分条件, 让我们定义 $n\times n$ 矩阵 s_{ij} 为正定和负定如下:

设 $\xi_1,\ \xi_2,\ \cdots,\ \xi_n$ 为一组不全同时等于零的实数, 则

(1) $\sum\limits_{i,j=1}^{n}s_{ij}\xi_i\xi_j>0$ 时, s_{ij} 为正定;

(2) $\sum\limits_{i,j=1}^{n}s_{ij}\xi_i\xi_j<0$ 时, s_{ij} 为负定.

根据这个定义, $f(x_1,\ x_2)$ 在 $(x_1^{(0)},\ x_2^{(0)})$ 点上是相对极大的充要条件为

$$
\left.\begin{array}{l}
\dfrac{\partial f}{\partial x_1}(x_1^{(0)},\ x_2^{(0)})=0,\ \ \dfrac{\partial f}{\partial x_2}(x_1^{(0)},\ x_2^{(0)})=0\\
\dfrac{\partial^2 f}{\partial x_i\partial x_j}(x_1^{(0)},\ x_2^{(0)})\text{为负定},\ \ (i,\ j=1,\ 2)
\end{array}\right\} \tag{1.7}
$$

同样, $f(x_1,\ x_2)$ 在 $(x_1^{(0)},\ x_2^{(0)})$ 点上是相对极小的充要条件为

$$
\left.\begin{array}{l}
\dfrac{\partial f}{\partial x_1}(x_1^{(0)},\ x_2^{(0)})=0,\ \ \dfrac{\partial f}{\partial x_2}(x_1^{(0)},\ x_2^{(0)})=0\\
\dfrac{\partial^2 f}{\partial x_i\partial x_j}(x_1^{(0)},\ x_2^{(0)})\text{为正定},\ \ (i,\ j=1,\ 2)
\end{array}\right\} \tag{1.8}
$$

如果 f 是 $x_1,\ x_2,\ \cdots,\ x_n$ 的函数, 则我们可以将该函数简写为 $f(x_i)$, 其中 (x_i) 为 $(x_1,\ x_2,\ \cdots,\ x_n)$ 的简写, 于是 $f(x_i)$ 在 $x_i=x_i^{(0)}$ 点的相对极大或相对极小的必要条件为

$$\frac{\partial f}{\partial x_k}(x_i^{(0)})=0, \quad k=1, 2, \cdots, n \tag{1.9}$$

同样，相对极大的充要条件为

$$\frac{\partial f}{\partial x_k}(x_i^{(0)})=0, \quad \frac{\partial^2 f}{\partial x_k \partial x_j}(x_i^{(0)}) \text{ 为负定} \tag{1.10}$$

而相对极小的充要条件为

$$\frac{\partial f}{\partial x_k}(x_i^{(0)})=0, \quad \frac{\partial^2 f}{\partial x_k \partial x_j}(x_i^{(0)}) \text{ 为正定} \tag{1.11}$$

现在让我们研究 $f(x_1, x_2)$ 在 x_1, x_2 受约束条件 $g(x_1, x_2)=0$ 的限制下的极值问题．几何上讲，这个约束条件限制了 x_1, x_2 值，使它们都在 $g(x_1, x_2)=0$ 这条曲线上．例如，$f(x_1, x_2)$ 代表地面上各点的海拔高度，$g(x_1, x_2)=0$ 代表一条公路的所经各点，我们的问题就相当于求在公路各点上，哪几点是海拔高度的相对极值．

由于 $g(x_1, x_2)=0$，变量 x_1, x_2 并不是独立的．用一般方法，即根据(1.6)式求 $f(x_1, x_2)$ 的极值并不一定正确，只有当(1.6)式求得的 $x_1^{(0)}$, $x_2^{(0)}$ 点恰好在 $g(x_1, x_2)=0$ 这条曲线上时，才碰巧满足要求．一般求 $f(x_1, x_2)$ 在满足 $g(x_1, x_2)=0$ 这个条件时的极值的步骤如下：先求解 $g(x_1, x_2)=0$，使一个变量 x_1 用一个 x_2 的函数 $h(x_2)$ 表示，即

$$x_1=h(x_2) \tag{1.12}$$

把它代入 $f(x_1, x_2)$，得

$$f=f(h(x_2), x_2) \tag{1.13}$$

(1.13)式是单一变量 x_2 的函数，其极值条件为

$$\frac{df}{dx_2}=\frac{\partial f}{\partial h}\frac{dh}{dx_2}+\frac{\partial f}{\partial x_2}=0 \tag{1.14}$$

根据(1.12)式，上式也可以写作

$$\frac{\partial f}{\partial x_1}\frac{dx_1}{dx_2}+\frac{\partial f}{\partial x_2}=0 \tag{1.15}$$

式中的 $\frac{dx_1}{dx_2}$ 即为 $\frac{dh}{dx_2}$，它可以从(1.12)式求得，也可以从

$g(x_1, x_2)=0$ 求得．微分 $g(x_1, x_2)=0$，得

$$\frac{\partial g}{\partial x_1}\frac{dx_1}{dx_2}+\frac{\partial g}{\partial x_2}=0 \tag{1.16}$$

从(1.15)，(1.16)式中消去 $\frac{dx_1}{dx_2}$，得

$$-\frac{\partial f}{\partial x_1}\left(\frac{\frac{\partial g}{\partial x_2}}{\frac{\partial g}{\partial x_1}}\right)+\frac{\partial f}{\partial x_2}=0 \quad \left(\frac{\partial g}{\partial x_1}\neq 0\right) \tag{1.17}$$

把(1.17)式和约束条件

$$g(x_1, x_2)=0 \tag{1.18}$$

合在一起，就是求解极值坐标 $x_1^{(0)}$，$x_2^{(0)}$的两个方程．(1.17)式也可以写成

$$\frac{\partial f}{\partial x_1}-\left(\frac{\frac{\partial g}{\partial x_1}}{\frac{\partial g}{\partial x_2}}\right)\frac{\partial f}{\partial x_2}=0 \quad \left(\frac{\partial g}{\partial x_2}\neq 0\right) \tag{1.19}$$

现在让我们从几何角度来说明这个问题．设我们求 $f(x_1, x_2)$ 在条件 $g(x_1, x_2)=0$ 下的极小(或极大)问题．这里有两种可能：

(1) $f=C_i$ 这些等高线中，有 $f=C_0$ 是一个极小点 Q，而这个

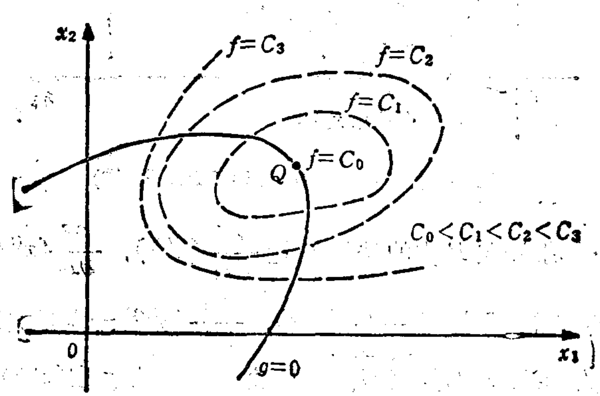

图1.1 $f=C_i$ 的极小点 Q 却好在 $g=0$ 曲线上．

极小点又恰好位在 $g(x_1, x_2)=0$ 的曲线上(图 1.1).

在这个 f 的极小点上，我们有

$$\frac{\partial f}{\partial x_1}(x_1^{(0)}, x_2^{(0)})=0, \qquad \frac{\partial f}{\partial x_2}(x_1^{(0)}, x_2^{(0)})=0 \qquad (1.20)$$

如果从(1.20)式中求解$(x_1^{(0)}, x_2^{(0)})$，把结果代入(1.18)式，恰好得

$$g(x_1^{(0)}, x_2^{(0)})=0 \qquad (1.21)$$

我们可以看到，在这种情况下，不论$\frac{\partial g}{\partial x_1}$, $\frac{\partial g}{\partial x_2}$在该点的值是多少，(1.17)或(1.18)式都是满足的.

(2) $g(x_1, x_2)=0$ 不通过 f 的极小点 Q，但和 $f=C_2$ 相切于 P 点. 则 P 点为曲线 $g(x_1, x_2)=0$ 上 f 值最小的点. 很易看到在曲线 $g(x_1, x_2)=0$ 上 P 点附近各点 P_1, P_2, P_3, P_4 上的 f 值都比 P 点的 f 值大(图 1.2).

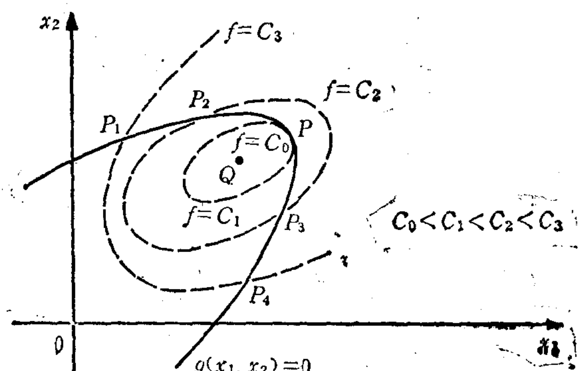

图 1.2 $g=0$ 曲线和 $f=C_1$ 切于 $P(x_1^{(P)}, x_2^{(P)})$

$f=C_1$ 曲线在 P 点的法线矢量 $\frac{\partial f}{\partial x_1}(x^{(P)})$, $\frac{\partial f}{\partial x_2}(x^{(P)})$ 和 $g=0$ 在 P 点的法线矢量 $\frac{\partial g}{\partial x_1}(x^{(P)})$, $\frac{\partial g}{\partial x_2}(x^{(P)})$ 一定是共线的，其中$(x^{(P)})$为$(x_1^{(P)}, x_2^{(P)})$的简写. 这里业已假定 f, g 在$(x^{(P)})$附近都是连续可导的. 这个共线条件可以用下列等式来表示，即