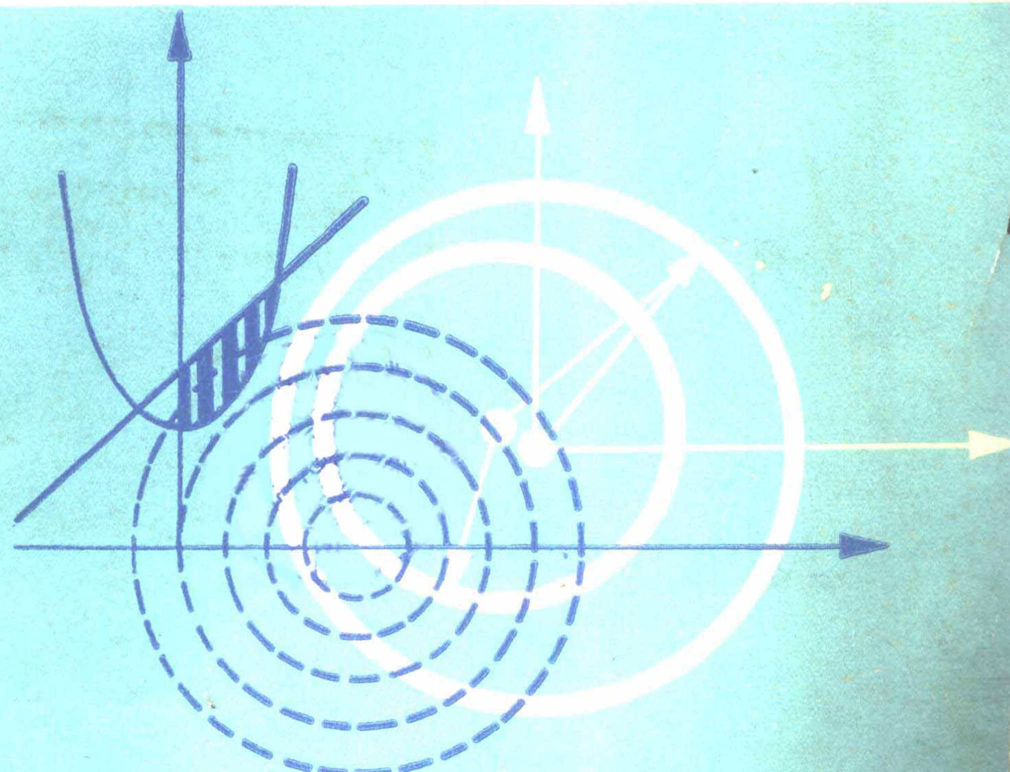
高等学校教学用书

非线性规划

朱儒楷　黄　皓　朱开永　编著



中国矿业大学出版社

高等学校教学用书

非 线 性 规 划

朱儒楷　黄　皓　朱开永

中国矿业大学出版社

内 容 简 介

本书较详细讨论了非线性规划的理论与方法。主要包括一维搜索的各种算法；无约束最优化方面的最速下降法、牛顿法、变尺度法、共轭梯度法、以及Powell方法等直接方法；约束最优化方面的可行方向法、梯度射影法、惩罚函数法、广义拉格朗日乘子法、线性逼近法等。各种方法都阐明其理论根据、计算步骤、算法框图，附有一些算法的计算机程序和较多的例题。本书还简介了学科的发展动向，列出了较多的参考文献。

本书可作为高等院校有关专业高年级学生、研究生教材，也可供有关教师和广大优化工作者和工程技术人员参考。

责任编辑：阎前辉
技术设计：杜锦芝
责任校对：关湘雯

高等学校教学用书
非线性规划
朱儒楷 黄 皓 朱开永 编著

中国矿业大学出版社出版
江苏省新华书店经销　　中国矿业大学印刷厂印刷
开本787×1092毫米1/16　印张16.5字数393千字
1990年5月第一版　　1990年5月第一次印刷
印数：1—1000册

ISBN 7-81021-225-7

F·15　　定价：3.30元

前　　言

非线性规划作为数学最优化理论与方法的一个重要分支，越来越受到有关数学工作者和广大科技人员的重视。随着电子计算机的普及，它已广泛应用于各工程技术部门和经济、管理等学科领域。作者之一在给工科系统工程等专业的研究生讲授非线性规划课时，曾选用中外多种参考书。各书都有其优点。有的以其理论的严谨著称，但比较忽视具体算法和例题；有的重视实际解法和对某个专业的应用，但理论又讲得不够；有的确有兼顾理论与方法的优点，但又要求偏高，内容偏多，不太符合我国工科研究生的教学要求，更不适宜广大工程技术人员参考。鉴于上述情况，作者在吸收各书优点的基础上，参照教学实际经验，编成这一本理论与方法并重、适应我国广大读者需要的非线性规划教材。

本书在理论方面，如凸集与凸函数的有关理论(第二章)、算法映射及其收敛性定理(第四章)、无约束极值和约束极值的充要条件、以及对偶理论与鞍点最优性准则(第一章、第八章)等，精选最必要内容，尽可能阐述清楚，重要定理都加以证明。在算法方面，选择常用的、实践证明较重要或行之有效的各种方法，有作为最优化算法基础的一维搜索诸算法(第三章)；无约束最优化方法(第五、六、七章)；约束最优化方法(第九、十、十一章)。每种算法均大致依次由算法导引、计算步骤、例题、框图、收敛性或算法的理论依据等几部分组成。附录中有部分算法的 BASIC 程序，供读者选用。绪论和第十二章，举了一些各学科中应用非线性规划的实例。在第七章和第十一章的后面，均专有一节扼要综述无约束最优化方法和约束最优化方法的学科发展动向。书后列有相应的参考文献，以满足进一步关心理论研究的读者需要。

本书在写法上尽可能适应不同读者的需要。对理论要求较高的某些专业的大学生和研究生，授完全部内容约需60学时。若删去某些理论(如对偶理论和鞍点最优性准则)，某些定理可作直观理解(如凸集的分离定理，Farkas定理，约束最优化条件的某些引理等)，算法也可根据教学需要进一步精选，则可用 40 学时讲授。对于主要关心应用算法解决实际问题的读者，则可着重学习绪论，各种算法的导引、步骤与框图、例题，以及第十二章；对大多数理论推导和定理的证明，在阅读时都可略去，但要搞清楚一些定义，对某些重要结论也要有所了解，在掌握算法步骤的基础上，用程序上机进行习题演算，为解决实际问题作好准备。

本书是在朱儒楷1984年所编同名讲义的基础上，由朱儒楷、黄皓、朱开永修改、增订而成。使用过该讲义的教师和学生们曾提出过许多宝贵意见，雷阿庚同志花了许多时间热情帮助整理原稿，我们深表感谢。

编者　　1988年6月

目　录

无 约 束 问 题

有 约 束 问 题

绪　　论

随着生产的高速发展和计算机的广泛应用，最优化的方法在越来越多的领域得到了应用。反过来，最优化的方法也因此不断地丰富和完善。非线性规划是最优化的一个重要分支。其形式为

$$\begin{aligned} &\text{Min} \quad f(\mathbf{x}) \\ &\text{S.t.} \quad \mathbf{x} \in \Omega \subseteq \mathbf{E}_n \end{aligned} \tag{0-1}$$

即在 n 维空间 $\mathbf{E}_n$ 的一个子集Ω 中求函数 $f(\mathbf{x})$ 的极小值和极小值点。如果 $\Omega = \mathbf{E}_n$，则问题(0-1)被称为无约束非线性规划问题，如果 Ω 是 $\mathbf{E}_n$ 的真子集，则问题(0-1)被称为约束非线性规划问题。

非线性规划的问题出现在经济、管理、机械、化工、石油、交通、地质、勘探、自动化等工程技术领域。下面就几个简单的例子，说明非线性规划的应用背景。

例0-1　系统识别与最小二乘法问题。

假设有一个系统，如图 0-1 所示，输入变量 $\theta_1,\cdots,\theta_n$ 与输出变量 y 近似地呈线性关系

$$y = \theta_1 x_1 + \cdots + \theta_n x_n \tag{0-2}$$

所谓近似地呈线性关系，即任意给定参数 $(x_1,\cdots,x_n)$ 后，我们观察对应一组输入变量 $(\theta_1^k,\cdots,\theta_n^k)$，$k=1,\cdots,m$ 的输出变量 y_k，$k=1,\cdots,m$，一般与按式(0-2)计算出来的值 $\bar{y}_k = \theta_1^k x_1 + \cdots + \theta_n^k x_n$，$k=1,\cdots,m$ 不一致，即 $\sum_{k=1}^{m} [y_k - (\theta_1^k x_1 + \cdots + \theta_n^k x_n)]^2 \neq 0$。因此为了较好地用线性模型描述输入变量 $\theta_1,\cdots,\theta_n$ 与输出变量 y 之间的关系，通常选择参数 $x_1,\cdots,x_n$ 使实际的输出变量与计算所得的输出变量之间的偏差尽可能地小。即可抽象成如下的数学模型

$$\text{Min} \quad f(x_1,\cdots,x_n) = \sum_{k=1}^{m} [y_k - (\theta_1^k x_1 + \cdots + \theta_n^k x_n)]^2 \tag{0-3}$$

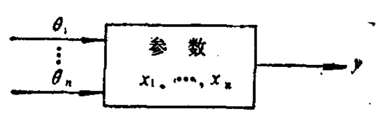


图0-1

例0-2　电厂发电机组间的经济负荷分配。

设有三台发电机组并联运行，功率分别为x_1,x_2,x_3（千瓦）。每台发电机的发电费用为

$$f_1(x_1) = 2x_1^2 + 3x_1 + 1$$
$$f_2(x_2) = x_2^2 + 4x_2 + 2$$
$$f_3(x_3) = x_3^2 + x_3 + 6$$

要求发电机组的总功率为 L，求发电机组间最优负荷分配，使发电总费用为最小。则

经济负荷分配的数学模型为

$$\begin{aligned} &\text{Min} \quad f_1(x_1)+f_2(x_2)+f_3(x_3) \\ &\text{S.t.} \quad x_1+x_2+x_3=L \end{aligned} \tag{0-4}$$

例0-3　大型零件加工余量的优化。

如图 0-2 所示，有一个大致圆型的毛坯，需要加工一个半径为 r 的圆。先在毛坯上找一个大致的中心 o，建立 xoy 坐标系，以一定的角度增量对毛坯的边缘进行测量，得到各点的极坐标值 (r_i,θ_i)，$i=1,\cdots,n$，换算为直角坐标为

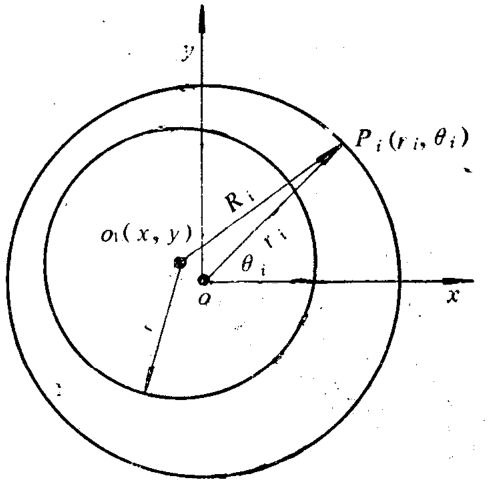


图0-2

$$x_i=r_i\cos\theta_i,\quad y_i=r_i\sin\theta_i,\quad i=1,\cdots,n$$

加工时，若把圆心取 $o_1(x,y)$ 点，则 o_1 到点 $P_i=(r_i,\theta_i)$ 的距离为

$$R_i=\sqrt{(x_i-x)^2+(y_i-y)^2}$$

于是，点 P_i 的加工余量为

$$\Delta_i=R_i-r$$

比较各点的加工余量，可找出最小的加工余量 $\Delta_{\min}$，它是圆心 o_1 的坐标 (x,y) 的函数

$$\Delta_{\min}=f(x,y)=\min_{1\leqslant i\leqslant n}\left[\sqrt{(x_i-x)^2+(y_i-y)^2}-r\right]$$

加工时，我们取的圆心 o_1 的坐标应该能使 $\Delta_{\min}$ 尽可能地大。因此这个问题的数学模型为

$$\text{Max}\quad f(x,y)=\text{Max}\min_{1\leqslant i\leqslant n}\left[\sqrt{(x_i-x)^2+(y_i-y)^2}-r\right] \tag{0-5}$$

例0-4　化学反应平衡问题。

19世纪末，人们发现化学反应平衡时，生成物的自由能达到最小。利用这一原理可以设计复杂的有机化合物的合成以及复燃料(例如火箭燃料)的配方。若参与化学反应的物质中有 m 种不同的元素，设为 $E^{(1)},\cdots,E^{(m)}$，反应的生成物有 n 种化合物 $C^{(1)},\cdots,C^{(n)}$。第 i 种生成物 $C^{(i)}$ 中含有 a_{ij} 个原子 $E^{(j)}$，即 $C^{(i)}$ 的分子式为

$$C^{(i)}=E^{(1)}_{a_{i1}}\cdots E^{(m)}_{a_{im}}$$

并设参与反应的原子 $E^{(j)}$ 有 b_j 个，生成物 $C^{(i)}$ 有 x_i 个分子。由物质不灭定律我们有

$$\sum_{j=1}^{m}b_j=\sum_{j=1}^{m}\sum_{i=1}^{n}x_i a_{ij}$$

这个系统的第 i 种生成物的自由能的计算公式为

$$g_i(x_1,\cdots,x_n)=x_i\left(c_i+\ln\frac{x_i}{\sum\limits_{i=1}^{n}x_i}\right)$$

其中 c_i 是与系统的压力、温度、化合物的参数有关的参数。这样，系统的化学反应平衡时，反应物的总自由能为

$$f(x_1,\cdots,x_n)=\sum_{i=1}^{n}g_i(x_1,\cdots,x_n)$$

因此，生成物 $C^{(i)}$ 的数量 x_i 是下列问题的解

$$\begin{aligned}&\operatorname{Min} f(x_1,\cdots,x_n)\\&\text{S.t.}\sum_{i=1}^{n}a_{ij}x_i=b_j\quad j=1,\cdots,m\end{aligned}\tag{0-6}$$

例0-5　光学系统的设计问题。

一个光学系统，例如照像机、望远镜、显微镜、电影放映机等，一般是由若干个透镜组成的。各透镜的半径、厚度、透镜之间的距离以及材料的折射率是这个系统的结构参数。这些参数直接影响着系统的质量。因此，在设计一个光学系统时，就要确定这些参数的大小，使这个系统的质量尽可能地好。

最理想的光学系统是物体在像平面上所成的像的形状与物平面上的物体的形状是几何相似的。即光线从物平面上的一点 (u,v) 出发，沿着方向 (α,β,γ)（其中 $\alpha^2+\beta^2+\gamma^2=1$）经过系统后，在像平面上的点为 $(u_p,v_p)=(Mu,Mv)$，M为常数，称为放大系数。但实际上这是不可能的。像点的坐标 (u_p,v_p) 取决于点 (u,v)、光线的发射角度 (α,β)〔因为 $\gamma=(1-\alpha^2-\beta^2)$ 取决于 α、β〕及结构参数 $x_1,\cdots,x_n$，它与理想点(Mu,Mv)的偏差记为

$$G(u,v,\alpha,\beta,x_1,\cdots,x_n)=u_p-Mu$$
$$H(u,v,\alpha,\beta,x_1,\cdots,x_n)=v_p-Mv$$

设计时在物平面上选择若干个点 (u_i,v_i)，$i=1,\cdots,m$ 从每一点出发选择若干条光线的方向(α_{ij},β_{ij})，$j=1,\cdots,l$，记

$$G_{ij}(x_1,\cdots,x_n)=G(u_i,v_i,\alpha_{ij},\beta_{ij},x_1,\cdots,x_n)$$
$$H_{ij}(x_1,\cdots,x_n)=H(u_i,v_i,\alpha_{ij},\beta_{ij},x_1,\cdots,x_n)$$

假定结构参数 $(x_1,\cdots,x_n)$ 的取值范围是 Ω，则设计者必须求解下列问题

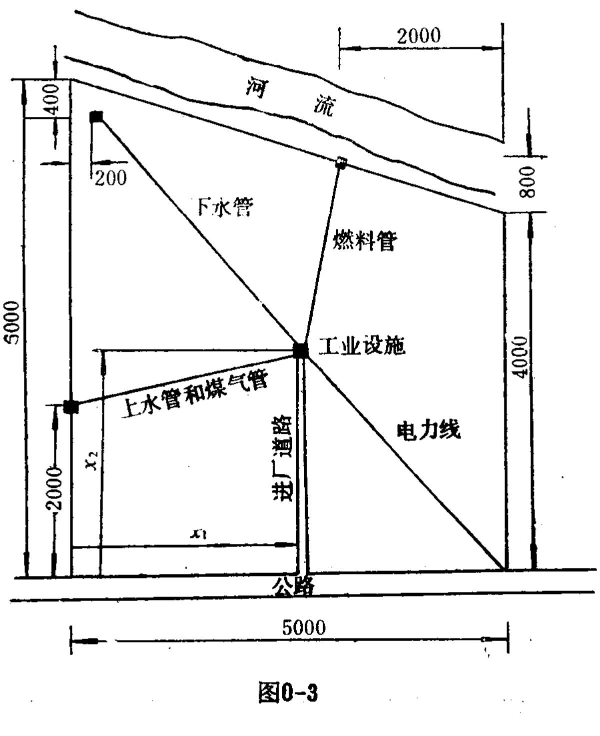


图0-3

$$\mathrm{Min} f(x_1,\cdots,x_n)=\sum_{i=1}^{m}\sum_{j=1}^{l}[G_{ij}^2(x_1,\cdots,x_n)+H_{ij}^2(x_1,\cdots,x_n)] \tag{0-7}$$

$$\mathrm{S.t.}\quad (x_1,\cdots,x_n)\in\Omega$$

例0-6 建造工厂的位置问题。

在图 0-3 所示的场地上欲建一工业设施，所用燃料管、下水管、电力线、上水管和煤气管均由现有汇接点引入，此外还要修建一条垂直于公路的进厂道路。单位长度管线、桩和道路修建费示于表 0-1 中。地质勘探说明有一层硬粘土层，从公路向河流方向倾斜，其上被一层软积土所覆盖(见图 0-4)。由于土质条件决定使用 150 根支承桩。现在设计者要确定使建造费最小的工业设施位置。

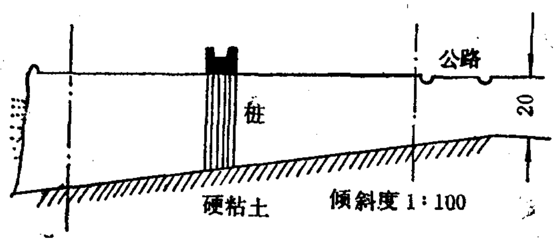


图0-4

表 0-1

项　　目	进厂道路	电 力 线	上水管和煤 气 管	下 水 管	燃 料 管	桩
费用(元/尺)	15	8	3	4	12	1.5

如图 0-3 所示的变量 x_1, x_2，我们可确定目标函数和约束条件如下

目标函数

$$f(x_1,x_2)=15x_2+3[(5000-x_1)^2+x_2^2]^{\frac{1}{2}}+5[x_1^2+(x_2-2000)^2]^{\frac{1}{2}}$$
$$+4[(x_1-200)^2+(5600-x_2)^2]^{\frac{1}{2}}+12[(3000-x_1)^2+(4800-x_2)^2]^{\frac{1}{2}}$$
$$+225\left(\frac{x_2}{100}\right)$$

约束条件

$$0\leqslant x_1\leqslant 5000$$

$$0\leqslant x_2\leqslant -\frac{2x_1}{5}+6000$$

总结起来，非线性规划的一般模型具有如下的型式

$$\begin{aligned}&\mathrm{Min} f(\mathrm{x})\\ &\mathrm{S.t.}\ g_i(\mathrm{x})\leqslant 0\quad i=1,\cdots,m\\ &\qquad h_j(\mathrm{x})=0\quad j=1,\cdots,l\end{aligned} \tag{0-8}$$

即在 n 维空间 E_n 的满足条件 $g_i(\mathrm{x})\leqslant 0\ \ i=1,\cdots,m, h_j(\mathrm{x})=0,\ j=1,\cdots,l$ 的点 x 中寻找一个使函数 $f(\mathrm{x})$ 取值最小的点，其中 $f(\mathrm{x}), g_i(\mathrm{x}), i=1,\cdots,m,\ h_j(\mathrm{x})=0,\ j=1,\cdots,l$ 是 n 元函数。当 $m=l=0$ 时，问题(0-8)称为无约束问题，当 $m^2+l^2>0$ 时，问题(0-8)称为约束问题。求解非线性规划的主要方法是寻找一个序列 $\{\mathrm{x}_k\}$，使得这个序列收敛于

问题(0-8)的最优点 x^*。具体的做法是从已得到的近似点 x_k 起，寻找一个方向 d_k，然后确定一个步长 λ_k，从而按照公式

$$x_{k+1}=x_k+\lambda_k d_k \tag{0-9}$$

计算下一个近似点 x_{k+1}。对于无约束问题，选择 d_k 的准则是，目标函数 $f(x)$ 在方向 d_k 上是下降的，如果取 d_k 为 x_k 点的负梯度方向 $-\nabla f(x_k)$，则构成最速下降法，我们在 §5.1 中介绍了这个方法；牛顿法、变尺度法、共轭梯度法等选择 d_k 时是将负梯度方向 $-\nabla f(x_k)$ 偏转一定的角度(≤90，保证 d_k 是 $f(x)$ 在 x_k 的下降方向)；或者轮流地取坐标方向为 d_k，以及根据已得到的近似点的函数值来作一个调整，第七章将介绍这样的方法。对于约束问题选择 d_k 主要有两种方法，一个方法是：所选择的 d_k 是目标函数 $f(x)$ 的下降方向，并且 d_k 是可行方向，即只要 λ 充分小，则点 $x_k+\lambda d_k$ 是满足问题(0-8)的约束条件的点，第九章较详细地介绍了这种方法；另一个方法是：所选择的 d_k 是目标函数的下降方向，或者从 x_k 沿 d_k 出发离约束集合更近，或者两者都是。其办法是，在目标函数上加上一个惩罚函数，构成一个新的目标函数，d_k 是新的目标函数的下降方向，在第十章比较详细地介绍了这种方法——制约函数法。除此之外，在第十一章简单介绍了割平面方法、用简单函数(例如线性函数)逼近目标函数和约束函数的近似型算法。

最近二三十年来，非线性规划取得许多进展，尤其近几年来，许多新的更有效的方法不断涌现。本书在选择所介绍的方法时，尽量选择在实践中证明了是比较好的方法。限于篇幅，还有许多方法没有介绍，但是读者可以从参考文献中找到。

第一章　多元函数极值的条件

本章对数学分析中有关多元函数极值的理论作扼要回顾。在复习一些必要的基本概念及符号的基础上，首先证明重要的多元台劳公式，然后推导多元函数极值的一些必要条件和充分条件，最后介绍为解决具有等式约束的多元函数极值问题的拉格朗日乘子法。

§1.1　预 备 知 识

一、几个常用的记号

1. 若 $f(\mathbf{x})$ 在 $\mathbf{E}_n$ 的某个开区域上连续，则记为 $f(\mathbf{x})\in C$。若 $f(\mathbf{x})$ 在 $\mathbf{E}_n$ 的某个开区域上有直至 p 阶连续偏导数，则记为 $f(\mathbf{x})\in C^p$。若无特别说明，以下开区域用符号 $\mathbf{D}$ 表示，简称区域。

2. 若 $\mathbf{x}\in\mathbf{D}\subset\mathbf{E}_n$，$f(\mathbf{x})\in C^1$，定义 $f(\mathbf{x})$ 在点 $\mathbf{x}$ 的梯度向量为

$$\nabla f(\mathbf{x})=\left(\frac{\partial f(\mathbf{x})}{\partial x_1},\cdots,\frac{\partial f(\mathbf{x})}{\partial x_n}\right) \tag{1-1}$$

3. 若 $\mathbf{x}\in\mathbf{D}\subset\mathbf{E}_n$，$f(\mathbf{x})\in C^2$，则 $f(\mathbf{x})$ 在点 $\mathbf{x}$ 的海赛(Hesse)矩阵为一个 $n\times n$ 对称矩阵，记为 $\nabla^2 f(\mathbf{x})$ 或 $\mathbf{H}(\mathbf{x})$，定义如下

$$\nabla^2 f(\mathbf{x})=\left[\frac{\partial^2 f(\mathbf{x})}{\partial x_i\partial x_j}\right],\quad i,j,=1,\cdots,n \tag{1-2}$$

或

$$\nabla^2 f(\mathbf{x})=\begin{bmatrix}\frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1\partial x_n}\\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2\partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & & \frac{\partial^2 f(\mathbf{x})}{\partial x_2\partial x_n}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n\partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n\partial x_2} & & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2}\end{bmatrix}$$

4. 对于 $n\times n$ 对称矩阵 $\mathbf{A}=[a_{ij}]$，其二次型 $\mathbf{x}^T\mathbf{A}\mathbf{x}$ 为正定的充要条件是

$$a_{11}>0;\quad \begin{vmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{vmatrix}>0;\cdots;\quad \begin{vmatrix}a_{11} & \cdots & a_{1n}\\ \cdots & \cdots & \cdots\\ a_{n1} & \cdots & a_{nn}\end{vmatrix}>0$$

其二次型 $\mathbf{x}^T\mathbf{A}\mathbf{x}$ 为负定的充要条件是

$$a_{11}<0;\quad \begin{vmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{vmatrix}>0;\cdots;\quad (-1)^n\begin{vmatrix}a_{11} & \cdots & a_{1n}\\ \cdots & \cdots & \cdots\\ a_{n1} & \cdots & a_{nn}\end{vmatrix}>0$$

二、极值定义

$f(\mathbf{x})$ 是定义在区域 $\mathbf{D}\subset\mathbf{E}_n$ 上的实函数，点 $\mathbf{x}\in\mathbf{D}$。

1. 若 $\exists\delta>0$，对一切满足 $\|\mathbf{x}-\mathbf{x}^*\|<\delta$ 和 $\mathbf{x}\in\mathbf{D}$ 的点 $\mathbf{x}$，都有

(1) $f(\mathbf{x})\geqslant f(\mathbf{x}^*)$，则称点 $\mathbf{x}^*$ 为 $f(\mathbf{x})$ 在区域 $\mathbf{D}$ 上的局部极小点；$f(\mathbf{x}^*)$ 称为局部

极小值；

(2) $f(\mathbf{x}) > f(\mathbf{x}^*)$， $\mathbf{x} \neq \mathbf{x}^*$， 则称点 $\mathbf{x}^*$ 为 $f(\mathbf{x})$ 在区域 D 上的严格局部极小点；$f(\mathbf{x}^*)$ 为严格局部极小值。

2．若对一切 $\mathbf{x} \in D$，都有

(1) $f(\mathbf{x}) \geqslant f(\mathbf{x}^*)$，则称点 $\mathbf{x}^*$ 为 $f(\mathbf{x})$ 在区域 D 上的全局极小点；$f(\mathbf{x}^*)$ 为全局极小值；

(2) $f(\mathbf{x}) > f(\mathbf{x}^*)$， $\mathbf{x} \neq \mathbf{x}^*$， 则称点 $\mathbf{x}^*$ 为 $f(\mathbf{x})$ 在区域 D 上的严格全局极小点；$f(\mathbf{x}^*)$ 为严格全局极小值。

关于局部、全局极大点，极大值定义，读者显然容易由上述定义对比而得到，从略。

三、多元台劳定理

假设联结 E_n 上的两点 $\mathbf{x}_1$ 和 $\mathbf{x}_2$ 的线段被包含在区域 $D \subset E_n$ 中，并且在区域 D，$f(\mathbf{x}) \in C^{k+1}$，则在联结两点 $\mathbf{x}_1$ 和 $\mathbf{x}_2$ 的线段上存在一点 ξ，使

$$\begin{aligned} f(\mathbf{x}_2) = f(\mathbf{x}_1) &+ \sum_{i=1}^{n} \frac{\partial f(\mathbf{x}_1)}{\partial x_i}(x_i^2 - x_i^1) + \\ &+ \frac{1}{2!}\sum_{i,j=1}^{n} \frac{\partial^2 f(\mathbf{x}_1)}{\partial x_i \partial x_j}(x_i^2 - x_i^1)(x_j^2 - x_j^1) + \cdots + \\ &+ \frac{1}{k!}\sum_{i_1,\cdots,i_k=1}^{n} \frac{\partial^k f(\mathbf{x}_1)}{\partial x_{i_1}\cdots\partial x_{i_k}}(x_{i_1}^2 - x_{i_1}^1)\cdots(x_{i_k}^2 - x_{i_k}^1) + \\ &+ \frac{1}{(k+1)!}\sum_{i_1,\cdots,i_{k+1}=1}^{n} \frac{\partial^{k+1} f(\xi)}{\partial x_{i_1}\cdots\partial x_{i_{k+1}}}(x_{i_1}^2 - x_{i_1}^1)\cdots(x_{i_{k+1}}^2 - x_{i_{k+1}}^1) \end{aligned} \tag{1-3}$$

证：在联结点 $\mathbf{x}_1$ 和 $\mathbf{x}_2$ 的线段上，定义一元函数 $g(t)$ 如下

$$g(t) = f[\mathbf{x}_1 + t(\mathbf{x}_2 - \mathbf{x}_1)] = f[x_1^1 + t(x_1^2 - x_1^1), \cdots, x_n^1 + t(x_n^2 - x_n^1)]$$

根据多元复合函数的导数计算公式，容易求出 $g(t)$ 的各阶导数

$$g'(t) = \sum_{i=1}^{n} \frac{\partial f[\mathbf{x}_1 + t(\mathbf{x}_2 - \mathbf{x}_1)]}{\partial x_i}(x_i^2 - x_i^1)$$

$$g''(t) = \sum_{i,j=1}^{n} \frac{\partial^2 f[\mathbf{x}_1 + t(\mathbf{x}_2 - \mathbf{x}_1)]}{\partial x_i \partial x_j}(x_i^2 - x_i^1)(x_j^2 - x_j^1)$$

……………………………………………………

$$g^{k}(t) = \sum_{i_1,\cdots,i_k=1}^{n} \frac{\partial^k f[\mathbf{x}_1 + t(\mathbf{x}_2 - \mathbf{x}_1)]}{\partial x_{i_1}\cdots\partial x_{i_k}}(x_{i_1}^2 - x_{i_1}^1)\cdots(x_{i_k}^2 - x_{i_k}^1)$$

$$g^{k+1}(t) = \sum_{i_1,\cdots,i_{k+1}=1}^{n} \frac{\partial^{k+1} f[\mathbf{x}_1 + t(\mathbf{x}_2 - \mathbf{x}_1)]}{\partial x_{i_1}\cdots\partial x_{i_{k+1}}}(x_{i_1}^2 - x_{i_1}^1)\cdots(x_{i_{k+1}}^2 - x_{i_{k+1}}^1)$$

由一元台劳定理

$$g(1) = g(0) + g'(0) + \frac{1}{2!}g''(0) + \cdots + \frac{1}{k!}g^{k}(0) + \frac{1}{(k+1)!}g^{k+1}(\tau) \tag{1-4}$$

其中 $0<\tau<1$。

根据 $g(t)$ 的式子求出 $g(1)$ 和 $g(0)$，根据$g'(t),\cdots,g^k(t)$的式子分别求出$g'(0)$，$\cdots,g^k(0)$，又根据 $g^{k+1}(t)$ 的式子求出 $g^{k+1}(\tau)$，又注意到 $x_2+\tau(x_2-x_1)$ 正是联结x_1和x_2的线段上一点，记为ξ。将以上所得代入(1-4)式，即得所欲证的(1-3)式。(1-3)式称为n元台劳公式，在数学分析中是很有用的一个公式。

§1.2 极值的必要条件

在本章的定理中，一律假设 $f(x)$ 定义在区域 $D\subset E_n$ 上，所论及的点 x 和 x^* 等是区域 D 的内点，又点 x 的 ε 邻域是指集合 $N_\varepsilon(x)=\{y;\ \|y-x\|<\varepsilon\}$。

定理1.1（一阶必要条件）

假设在 D 上 $f(x)\in C^1$，则 $f(x)$ 在点 x^* 有局部极小值的必要条件是

$$\nabla f(x^*)=0 \tag{1-5}$$

证明：用反证法。假设 $f(x)$ 在点 x^* 的梯度向量不是零向量，则存在 i 使得$\dfrac{\partial f(x^*)}{\partial x_i}\neq 0$，不妨设$\dfrac{\partial f(x^*)}{\partial x_i}>0$。根据 $f(x)\in C^1$，则对于任意给定的$\delta>0$，存在 h：$0<h<\delta$，当η是连接点 x^* 和点 $(x_1^*,\cdots x_{i-1}^*,x_i^*-h,x_{i+1}^*,\cdots,x_n^*)\triangleq x_2\in D$的线段上任一点时，有：$\dfrac{\partial f(\eta)}{\partial x_i}>0$。在一般 n 元台劳公式(1-3)中，取 $k=0$，$x_1=x^*$，以及 x_2，即可得

$$f(x_2)=f(x^*)+\frac{\partial f(\xi)}{\partial x_i}(-h)<f(x^*)$$

(这里的点 ξ 是上述的 η 中之一)。由以上得 x^* 的任一 δ 邻域中，必有 D 中的点 x_2，使

$$f(x_2)<f(x^*)$$

这与点 x^* 是 $f(x)$ 的局部极小点的假设相矛盾，故 $\nabla f(x^*)=0$ 得证。

显然(1-5)式也是 $f(x)$ 在点 x^* 有局部极大值的必要条件，读者可自行证明。以下定理都针对极小点叙述，相对应的极大点的有关定理，读者都可自己叙述并加以证明。

使 $\nabla f(x)=0$ 的点 x，称为 $f(x)$ 的稳定点或驻点，也可称为临界点。但可能 $f(x)$ 在某点的梯度向量为零向量，而该点既非局部极大点也非局部极小点。例如考虑二元函数 $f(x,y)=x^2-y^2$ 容易验证 $\nabla f(0,0)=0$，点(0，0)是驻点。显然有 $f(0,0)=0$，当$x\neq 0$时 $f(x,0)>0$；当 $y\neq 0$时 $f(0,y)<0$。可见驻点(0，0)既非局部极大点也非局部极小点。因此需要找出驻点为局部极小点或局部极大点的充分条件，在下节再讨论。

定理1.2 （二阶必要条件）

假设在 D 上 $f(x)\in C^2$，则 $f(x)$ 在点 x^* 有局部极小值的必要条件是，对一切 $z\in E_n$，总有 $z^T\nabla^2 f(x^*)z\geqslant 0$，此时也称点 x^* 的海赛矩阵 $\nabla^2 f(x^*)$ 是半正定的。

证明 由于已知点 x^* 是 $f(x)$ 的局部极小点，必存在 x^* 的某个邻域 $N_{\delta_1}(x^*)$，使对一切 $x\in N_{\delta_1}(x^*)$，必有 $f(x)\geqslant f(x^*)$。又已知 $f(x)\in C^2$，由定理 1.1 知，$\nabla f(x^*)=0$。

以下用反证法。假设存在某一 $z_0\in E_n$，使 $z_0^T\nabla^2 f(x^*)z_0<0$。由于 $f(x)\in C^2$，必存在x^*的某邻域$N_{\delta_2}(x^*)$，使对一切 $x\in N_{\delta_2}(x^*)$，有$z_0^T\nabla^2 f(x)z_0<0$。现在取 $x_1\in N_{\delta_2}(x^*)$

$\cap N_\delta(x^*)$及实数$\theta>0$，使满足 $x_1-x^*=\theta z_0$，即$z_0=\dfrac{x_1-x^*}{\theta}$，显然这是能够做到的。

对于点 x_1 和 x^*，利用 $k=1$ 时的台劳定理，有

$$f(x_1)=f(x^*)+\nabla f(x^*)^T(x_1-x^*)+\frac{1}{2}(x_1-x^*)^T\nabla^2 f(\xi)(x_1-x^*) \tag{1-6}$$

这里 ξ 是连接 x^* 和 x_1 线段上一点，所以 $\xi\in N_\delta(x^*)$，应有

$$z_0^T\nabla^2 f(\xi)z_0=\left(\frac{x_1-x^*}{\theta}\right)^T\nabla^2 f(\xi)\left(\frac{x_1-x^*}{\theta}\right)<0$$

可得：

$$\frac{1}{2}(x_1-x^*)^T\nabla^2 f(\xi)(x_1-x^*) < 0$$

并注意到 $\nabla f(x^*)=0$，这样，由 (1-6) 式得 $f(x_1)<f(x^*)$，与 x^* 是局部极小点的假设矛盾，定理得证。

必要条件从反面去应用，即不满足极值必要条件的点必然不是极值点。前面考虑的二元函数 $f(x,y)=x^2-y^2$，由定理1.1并不能判断驻点(0，0)是否为极值点。但在该点的海赛矩阵 $\nabla^2 f(0,0)=\begin{bmatrix}2 & 0\\ 0 & -2\end{bmatrix}$，是不定的。由定理1.2即知(0，0)不是极小点。应用类似于定理 1.2 的关于极大点的定理，也可知 (0，0) 也不是极大点。

§1.3　极值的充分条件

定理1.3　(充分条件之一)

假设在 D 上 $f(x)\in C^2$，又点 x^* 满足 $\nabla f(x^*)=0$，且对任何非零向量 z，有 $z^T\nabla^2 f(x^*)z>0$，亦称 $\nabla^2 f(x^*)$ 正定，则 $f(x)$ 在点 x^* 有严格局部极小值。

证明：因已知$\nabla^2 f(x^*)$正定，且$f(x)\in C^2$，能够证明必存在x^*的某个邻域$N_\delta(x^*)$，使对一切 $x\in N_\delta(x^*)$，$\nabla^2 f(x)$ 正定。以下的 1 、 2 两段先证明这个结论，为最后在第 3 段中完成此定理的证明作准备。

1．设 H 是 $n\times n$ 对称正定矩阵，则有 $m>0$，使得当 $\|z\|=1$ 时，有 $z^THz>2m$。

证明　设 $u_1,u_2,\cdots,u_n$ 是 H 的标准化特征向量系，即

$$u_i^Tu_j=\begin{cases}0 & i\neq j\\ 1 & i=j\end{cases} \tag{1-7}$$

对于任意 $z\in E_n$，存在 $z_1,z_2,\cdots,z_n$，使得

$$z=\sum_{i=1}^n z_iu_i$$

再由已知 $\|z\|=1$，可得$\sum_{i=1}^n z_i^2=1$。计算

$$\begin{aligned}z^THz&=\left(\sum_{i=1}^n z_iu_i\right)^TH\left(\sum_{j=1}^n z_ju_j\right)\\&=\sum_{i=1}^n\sum_{j=1}^n z_iz_ju_i^THu_j\end{aligned}$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_j z_i z_j \mathbf{u}_i^T\mathbf{u}_j$$

$$=\sum_{i=1}^{n}\lambda_i z_i^2 \tag{1-8}$$

其中$\lambda_1,\lambda_2,\cdots,\lambda_n$是H的特征值。在上面推导中用到$\mathbf{H}\mathbf{u}_j=\lambda_j\mathbf{u}_j$与(1-7)式。

在(1-8)式中分别用最小特征值λ_1和最大特征值λ_n去替换每个λ_i并注意到$\sum_{i=1}^{n}z_i^2=1$，则有

$$\lambda_1\leqslant \mathbf{z}^T\mathbf{H}\mathbf{z}\leqslant\lambda_n$$

注意到$\lambda_1>0$，显然存在$m>0$，使$\mathbf{z}^T\mathbf{H}\mathbf{z}>2m$，得证。

2．由以上结论，因$\nabla^2 f(\mathbf{x}^*)$为正定，则有$m>0$，使得当$\|\mathbf{z}\|=1$时，有

$$\mathbf{z}^T\nabla^2 f(\mathbf{x}^*)\mathbf{z}=\sum_{i,j=1}^{n}\frac{\partial^2 f(\mathbf{x}^*)}{\partial x_i\partial x_j}z_i z_j>2m$$

设
$$M=\max\left\{\sum_{i,j=1}^{n}|z_i z_j|;\ \|\mathbf{z}\|=1\right\}$$

和
$$\varepsilon=m/M$$

由于$f(\mathbf{x})\in C^2$，存在$\delta>0$，当$\mathbf{x}\in N_\delta(\mathbf{x}^*)$时，

$$\left|\frac{\partial^2 f(\mathbf{x})}{\partial x_i\partial x_j}-\frac{\partial^2 f(\mathbf{x}^*)}{\partial x_i\partial x_j}\right|<\varepsilon$$

对一切$i,j=1,\cdots,n$成立

于是当$\mathbf{x}\in N_\delta(\mathbf{x}^*)$，$\|\mathbf{z}\|=1$时，有

$$\left|\sum_{i,j=1}^{n}\frac{\partial^2 f(\mathbf{x})}{\partial x_i\partial x_j}z_i z_j-\sum_{i,j=1}^{n}\frac{\partial^2 f(\mathbf{x}^*)}{\partial x_i\partial x_j}z_i z_j\right|<\varepsilon M=m$$

所以当$\mathbf{x}\in N_\delta(\mathbf{x}^*)$和$\|\mathbf{z}\|=1$时，有

$$\sum_{i,j=1}^{n}\frac{\partial^2 f(\mathbf{x})}{\partial x_i\partial x_j}z_i z_j=\mathbf{z}^T\nabla^2 f(\mathbf{x})\mathbf{z}>m$$

这就是说，当$\|\mathbf{z}\|>0$时，有

$$\mathbf{z}^T\nabla^2 f(\mathbf{x})\mathbf{z}>0$$

亦即存在$\mathbf{x}^*$的一个邻域$N_\delta(\mathbf{x}^*)$，当$\mathbf{x}\in N_\delta(\mathbf{x}^*)$时，$\nabla^2 f(\mathbf{x})$为正定，得证。

3．任取$\mathbf{x}\in N_\delta(\mathbf{x}^*)$，且$\mathbf{x}\neq\mathbf{x}^*$，由台劳中值定理，有

$$f(\mathbf{x})=f(\mathbf{x}^*)+\nabla f(\mathbf{x}^*)^T(\mathbf{x}-\mathbf{x}^*)+\frac{1}{2}(\mathbf{x}-\mathbf{x}^*)^T\nabla^2 f(\xi)(\mathbf{x}-\mathbf{x}^*) \tag{1-9}$$

其中ξ位于连接$\mathbf{x}^*$和$\mathbf{x}$的线段上，所以$\xi\in N_\delta(\mathbf{x}^*)$，$\nabla^2 f(\xi)$正定。又注意到定理的条件$\nabla f(\mathbf{x}^*)=0$，由(1-9)式即得：$f(\mathbf{x})>f(\mathbf{x}^*)$。即$f(\mathbf{x}^*)$是严格局部极小值，得证。

定理1.3 是根据函数$f(\mathbf{x})$在$\mathbf{x}^*$点的性态，给出$f(\mathbf{x})$在$\mathbf{x}^*$有严格局部极小值的充分条件。但容易找到不满足定理1.3中充分条件的极值的例子。例如一元函数$f(x)=$