最优化计算方法

邓乃扬　　诸梅芳

（北京工业大学）

中国人民解放军空军工程学院翻印

一九八三年七月

序　言

“最优化计算方法”是一个新兴的数学分支，它是选择“最优方案”的一种有力工具，能应用于多种工程技术和经济管理部门，并产生直接的经济效益。许多高等院校已陆续开设了该课程，在为我校工科各专业及有关单位开设该课的基础上，我们编写了本书。

本书着重介绍那些适用面较广、使用方便、具有实效的方法，并力求反映先进成果。按照本书所述算法，可以直接编制计算机程序，用来解决实际问题。考虑到本书的主要对象是工科学生、工程技术人员以及经济管理干部等非数学工作者，我们尽量利用平面或空间的图象直观地引进各种概念和方法。对于必要的理论分析，根据不同情况，分别采用几何解释、证明或说明等方式加以阐述。使之易于理解和掌握。

由于编者水平有限，书中错误与不妥之处在所难免，欢迎批评指正。

符　号

$f(x)$	目标函数
$g(x)$，$\nabla f(x)$	目标函数$f(x)$的梯度向量函数，$g(x)=\nabla f(x)=\left(\frac{\partial}{\partial x_1}f(x),\ \cdots,\ \frac{\partial}{\partial x_n}f(x)\right)^T$
$G(x)$，$\nabla^2 f(x)$	目标函数$f(x)$的Hess矩阵，其第i行第j列的元素为$\frac{\partial}{\partial x_i \partial x_j}f(x)$
$C_i(x)$	第i个约束函数
$C(x)$	由约束函数构成的向量函数，例如 $c(x)=(C_1(x),\ \cdots,\ C_l(x))^T$
$\nabla C_i(x)$	第i个约束函数$C_i(x)$的梯度向量函数
x^*	无约束问题或约束问题的解
$x^{(k)}$	对解x^*的第k次近似
$f^{(k)}$	$f^{(k)}=f(x^{(k)})$
$g^{(k)}$	$g^{(k)}=g(x^{(k)})=\nabla f(x^{(k)})$
$G^{(k)}$	$G^{(k)}=G(x^{(k)})=\nabla^2 f(x^{(k)})$
λ	拉格朗日乘子向量
R^n	n维欧氏空间
$\langle\cdot,\cdot\rangle$	向量的数量积
$\|\cdot\|$	向量的范数或矩阵的范数
a^T	向量 a 的转置
A^T	矩阵A的转置
$\mathrm{sign}(\cdot)$	符号函数：当$x>0$时，$\mathrm{sign}(x)=1$；当$x<0$时，$\mathrm{sign}(x)=-1$；$\mathrm{sign}(0)=0$
$\min\{f(x) \mid x\in D\}$	函数$f(x)$在集合D上的最小值
$A\setminus B$	集合A和集合B的差集，即属于集合A但不属于集合B的所有元素组成的集合
$A\cup B$	集合A和集合B的并集，即集合A与集合B中所有元素组成的集合
$A\cap B$	集合A和集合B的交集，即集合A与集合B中所有公共元素组成的集合

目　　录

第一章　概　　论

本章首先介绍最优化问题的数学形式，然后简单地阐明求解最优化问题的算法的收敛性概念。

§1　最优化问题实例

什么是最优化？用数学语言来说，就是找出多变量函数的极小点，我们通过两个实例予以说明。

例1.1.1　曲线拟合问题

设有两个物理量ξ和η，根据某一物理定律知道它们满足下列函数关系

$$\eta = a + b\xi^c \tag{1.1.1}$$

其中a,b,c是三个常数，在不同情况下它们取不同的数值。假定现在由实验测得了(ξ,η)的m组数据

$$(\xi_1,\eta_1),\quad(\xi_2,\eta_2),\quad\cdots\cdots\quad(\xi_m,\eta_m) \tag{1.1.2}$$

试选择a,b,c的值，使得曲线$\eta=a+b\xi^c$尽可能靠近所有的实验点(ξ_i,η_i)，$i=1,\cdots,m$。参看图1.1.1。这个问题可用最小二乘法求解，即选择a,b,c的一组值，使偏差的平方和

$$\delta(a,b,c)=\sum_{i=1}^{m}(a+b\xi_i^c-\eta_i)^2 \tag{1.1.3}$$

取最小值。换句话说，欲求出三个变量的函数$\delta(a,b,c)$的极小点。以此作为问题的解。

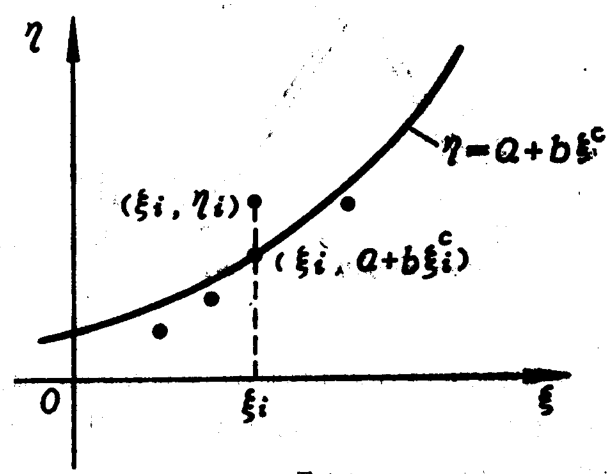

图 1.1.1

例1.1.2　生产安排问题

某工厂生产D,E两种产品，每种产品均经三道工序加工而成。假定每生产1米^{3}D种产品需用A种机器加工7小时，用B种机器加工3小时，用C种机器加工4.2小时。而每生产1米^{3}E种产品需用A种机器加工2.8小时，用B种机器加工9小时，用C种机器加工4小时。又已知每生产1米^{3}D种产品可盈利500元，每生产1米^{3}E种产品可盈利800元。现设一个月中A种机器工作时间不能超过560小时，B种机器不能超过450小时，C种机器不能超过336小时。问每月D，E两种产品各生产少可多盈利最多？

设每月生产D种产品x_1米^3，生产E种产品x_2米^3。这两种产品共盈利$500x_1+800x_2$元，我们要选择x_1和x_2，使它尽可能地大。然而由于受到机器加工时间的限制，x_1和x_2都不能太大。事实上，这两种产品共需用A种机器加工$7x_1+2.8x_2$小时，它不能超过560小时；需用B种机器加工$3x_1+9x_2$小时，它不能超过450小时；需用C种机器加工$4.2x_1+4x_2$小时，它不能超过336小时。总之我们应该在约束条件

$$7x_1+2.8x_2\leqslant 560$$
$$3x_1+9x_2\leqslant 450$$
$$4.2x_1+4x_2\leqslant 336$$
$$x_1\geqslant 0,\quad x_2\geqslant 0$$

下，寻找使函数

$$500x_1+800x_2$$

取最大值的x_1^*和x_2^*。

例1.1.3 人字架最优设计问题

考虑如图1.1.2所示钢管构成的人字架，设钢管壁厚$t=\bar{t}$和半跨度$s=\bar{s}$已经给定，试求能承受负荷2P的最轻设计。

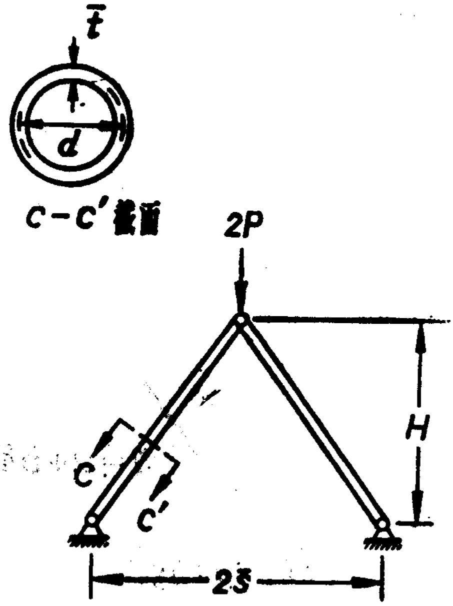

图 1.1.2

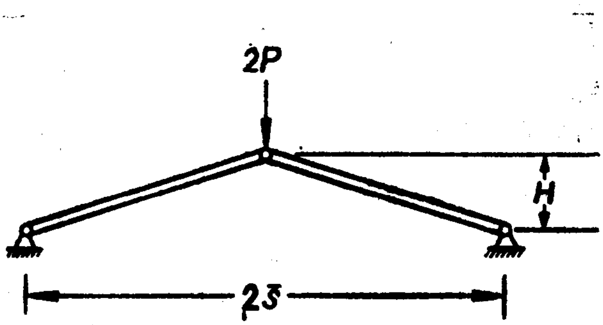

图 1.1.3

在这个问题中，需要选择调整的参数有两个：钢管的平均直径d（d等于内径D_1与外径D_2的平均值）和人字架的高H，我们先看看用定性分析的方法能否解决它。既然希望人字架轻，自然会想到钢管要尽可能短些，但是由于跨度$2\bar{s}$已经取定，这样做时人字架的张角将很大（参看图1.1.3）负荷2P就会在钢管上引起很大的分力，因而要求选用较粗的钢管。钢管变粗会增加人字架的重量，所以这未必是最优的方案。由此可见，进行以下的定量分析是必要的。

给定一组d和H的值后，立刻可以算出钢管的横截面积A和单根钢管的长L

$$A=\frac{1}{4}\pi(D_2^2-D_1^2)=\frac{\pi}{4}(D_2+D_1)(D_2-D_1)=\pi\cdot d\cdot\bar{t}$$

$$L=(\bar{s}^2+H^2)^{1/2}$$

因而人字架的重量可表为

$$W=W(d,\ H)=\rho\cdot 2\pi\cdot d\cdot \bar{t}(\bar{s}^2+H^2)^{1/2} \qquad (1.1.4)$$

（其中ρ是钢管材料的重度）。展在要调整d，H使W取最小值。由于整个结构必须能够承受负荷2P，所以d和H不能随意取值，它们应该满足如下两个条件：

(i) 压应力不能超过材料的许用应力σ_y。容易算出，在负荷2P作用下杆件上的压应力为

$$\sigma(d,\ H)=\frac{P}{\pi\bar{t}}\cdot\frac{(\bar{s}^2+H^2)^{1/2}}{Hd}$$

故应有

$$\frac{P}{\pi\bar{t}}\cdot\frac{(s^2+H^2)^{1/2}}{Hd}\leqslant\sigma_y \qquad (1.1.5)$$

这个限制称为屈服约束。

(ii) 压应力不能超过杆件的临界应力σ_e。临界应力σ_e可由Euler公式算得

$$\sigma_e=\sigma_e(d,H)=\frac{\pi^2E}{8}\ \frac{d^2+\bar{t}^2}{s^2+H^2}$$

（其中E是钢管材料的杨氏模量），故应有

$$\frac{P}{\pi\bar{t}}\ \frac{(\bar{s}^2+H^2)^{1/2}}{Hd}\leqslant\frac{\pi^2E}{8}\ \frac{d^2+\bar{t}^2}{\bar{s}^2+H^2} \qquad (1.1.6)$$

这个限制称为屈曲约束。

总之，我们是要求出d，H的一组值d*，H*，使得在满足屈服约束（1.1.5）和屈曲约束（1.1.6）的条件下，使重量（1.1.4）取最小值。

§2 最优化问题的数学形式

2.1 最优化问题的数学形式

前节考虑的三个例题中都包含着若干个需要选择或调整的量，而调整的目的都是使这些量的某个函数达到其最小值或最大值。这类问题称为最优化问题。然而三个例题也有不同之处：第一个例题中要选择的量是不受限制的，而后两个例题中要选择的量只能在某个限定的范围内取值。它们各自代表着一类最优化问题，前者称为无约束问题，后者称为约束问题。下面分别研究其数学表达形式。

1. 无约束问题

考虑例1.1.1所代表的问题，为了用统一的数学形式表达它们，先把例1.1.1写成标准形式。分别记需要选择的量a,b,c为x_1，x_2，x_3，并用一个三维欧氏空间R^3中的向量$\mathbf{x}=(x_1,\ x_2,\ x_3)^T$表示它们。再把欲求其最小值的函数$\delta(a,b,c)=\delta(x_1,\ x_2,\ x_3)$记作$f(\mathbf{x})=f(x_1,$

x_2, x_3)。最后把例1.1.1所述的问题表示为

$$\min f(x), \quad x=(x_1, x_2, x_3)^T \in R^3$$

其含义是求三维欧氏空间R^3上的函数 $f(x)$的最小点。可以想像，如果有n个需要选择或调整的量，我们可以用n维欧氏空间的向量$x=(x_1, \cdots, x_n)^T$表示它们，而把欲求其最小点的函数记为$f(x)$，从而问题可表示为

$$\text{(UP)} \quad \min f(x), \quad x \in R^n \tag{1.2.1}$$

这就是对自变量没有限制的最优化问题的一般形式，即无约束问题的一般形式，其中的$f(x)$称为目标函数。因此无约束问题就是寻求一个 n 维欧氏空间R^n上的函数$f(x)$的最小点。

定义1.2.1 若存在着$x^* \in R^n$，使得对任意的$x \in R^n$，都有

$$f(x) \geqslant f(x^*) \tag{1.2.2}$$

则称x^*是$f(x)$的最小点，或无约束问题(UP)的整体解。若对任意的$x \in R^n$，$x \neq x^*$，都有

$$f(x) > f(x^*) \tag{1.2.3}$$

则称x^*是$f(x)$的严格最小点，或无约束问题(UP)的严格整体解。

2. 约束问题

考虑例1.1.2 若记$x=(x_1, x_2)^T$，并令

$$f(x) = -(500x_1 + 800x_2)$$

则使$500x_1+800x_2$取最大值等价于使$f(x)$取最小值。另外若再令

$$C_1(x) = 7x_1 + 2.8x_2 - 560$$

$$C_2(x) = 3x_1 + 9x_2 - 450$$

$$C_3(x) = 4.2x_1 + 4x_2 - 336$$

$$C_4(x) = -x_1$$

$$C_5(x) = -x_2$$

我们可把问题简单地表为

$$\min f(x), \quad x=(x_1, x_2)^T \in R^2$$

$$\text{s.t.} C_i(x) \leqslant 0, \quad i=1,2,3,4,5$$

类似地若在例1.1.3中分别用x_1和x_2代替d和H，令$x=(x_1, x_2)^T$，再把函数W记作 f

$$f(x) = 2\pi\rho\, t x_1(\bar{s}^2 + x_2^2)^{1/2}$$

并令

$$C_1(x) = \frac{P}{\pi \bar{t}} \quad \frac{(\bar{s}^2 + x_2^2)^{1/2}}{x_1 x_2} - \sigma_y$$

$$C_2(x) = \frac{P}{\pi \bar{t}} \quad \frac{(\bar{s}^2 + x_2^2)^{1/2}}{x_1 x_2} - \frac{\pi^2 E}{8} \quad \frac{\bar{t}^2 + x_1^2}{\bar{s}^2 + x_2^2}$$

则例1.1.3可表为

$$\min f(x), \quad x=(x_1, x_2)^T \in R^2$$

$$\text{s.t.} C_i(x) \leqslant 0, \quad i=1,2$$

从以上两个例题可以想见，在有些问题中会有更多的要选择的量和更多的约束条件。当有n个需要选择的量、m个不等式约束时，问题可以表为

$$\text{(IP)} \quad \begin{array}{ll} \min f(x), & x=(x_1, \cdots, x_n)^T \in R^n \\ \text{s.t.} C_i(x) \leqslant 0, & i \in I = \{1, \cdots, m\} \end{array} \tag{1.2.4}$$

这类问题称为不等式约束问题。

有时我们会遇到由若干个等式限制自变量取值范围的问题，其一般形式为

$$
(\mathrm{EP})\quad \begin{array}{ll} \min f(\mathbf{x}), & \mathbf{x}=(x_1,\ \cdots,\ x_n)^T\in R^n \\ \text{s.t.}\, C_i(\mathbf{x})=0, & i\in E=\{1,\ \cdots,\ l\} \end{array} \tag{1.2.5}
$$

这类问题称为等式约束问题。

更加一般的约束问题可以既包含等式约束，又包含不等式约束，其一般形式为

$$
(\mathrm{GP})\quad \begin{array}{ll} \min f(\mathbf{x}), & \mathbf{x}=(x_1,\ \cdots,\ x_n)^T\in R^n \\ \text{s.t.}\, C_i(\mathbf{x})=0, & i\in E=\{1,\ \cdots,\ l\} \\ C_i(\mathbf{x})\leqslant 0, & i\in I=\{l+1,\ \cdots,\ l+m\} \end{array} \tag{1.2.6}
$$

这类问题称为一般约束问题。在问题（GP）中，$f(\mathbf{x})$称为目标函数，$C_i(\mathbf{x})=0(i\in E)$和$C_i(\mathbf{x})\leqslant 0(i\in I)$称为约束条件。满足约束条件的自变量$\mathbf{x}$组成的集合称为可行域。与定义1.2.1相对应，对问题（GP）可给出如下定义：

定义1.2.2 设D是问题（GP）的可行域

$$
D=\{\mathbf{x}\,|\,C_i(\mathbf{x})=0,\ i\in E;\ C_i(\mathbf{x})\leqslant 0,\ i\in I,\ \mathbf{x}\in R^n\} \tag{1.2.7}
$$

若存在着$\mathbf{x}^*\in D$，使得对任意的$\mathbf{x}\in D$，都有

$$
f(\mathbf{x})\geqslant f(\mathbf{x}^*) \tag{1.2.8}
$$

则称$\mathbf{x}^*$是目标函数$f(\mathbf{x})$在可行域D上的最小点，或约束问题（GP）的整体解，若对任意$\mathbf{x}\in D$，$\mathbf{x}\neq\mathbf{x}^*$，都有

$$
f(\mathbf{x})>f(\mathbf{x}^*) \tag{1.2.9}
$$

则称$\mathbf{x}^*$是目标函数$f(\mathbf{x})$在可行域D上的严格最小点，或约束问题（GP）的严格整体解。

这样，所谓求解约束问题（GP）的含义也就清楚了。

2.2 最优化问题的几何解释

两个自变量（$n=2$）的最优化问题，有着十分清晰的几何意义，我们通过几个例子予以说明。这些直观的图象对于研究最优化计算方法有着极大的启发性，应当引起我们的注意。

例1.2.1 考虑无约束问题

$$
\min f(\mathbf{x})=(x_1-2)^2+(x_2-2)^2+4 \tag{1.2.10}
$$

我们知道，目标函数$f(\mathbf{x})$的等高线是以$(2,2)^T$为圆心的一族圆

$$
(x_1-2)^2+(x_2-2)^2+4=k \tag{1.2.11}
$$

而且圆的半径越小，其相应的函数值也越小，参看图1.2.1和1.2.2。由图易见问题(1.2.10)的解为$x^*=(2,2)^T$。

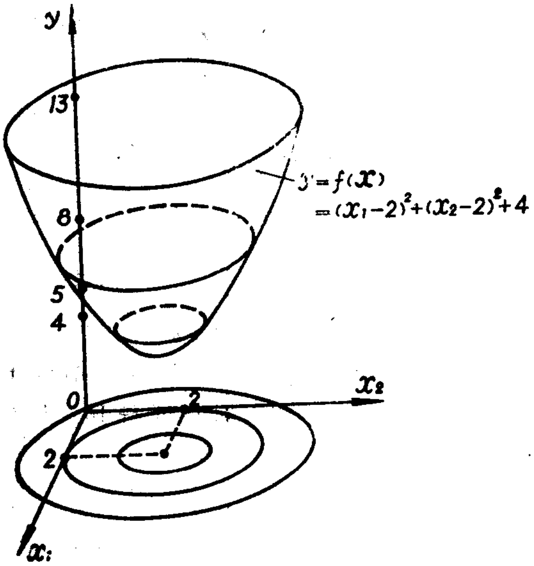

图 1.2.1

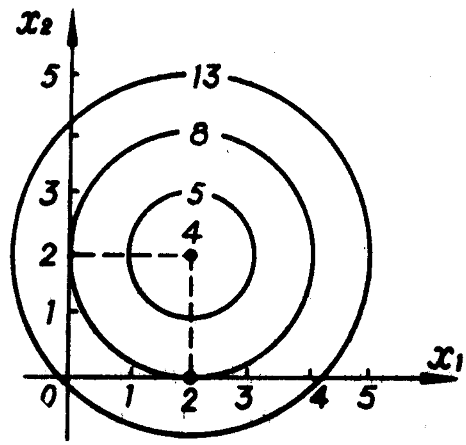

图 1.2.2

例1.2.2 考虑例1.1.2中问题的几何意义

该问题的目标函数的等高线为

$$f(x)=-(500x_1+800x_2)=k$$

它们是一族平行线（图1.2.3中的虚线），其可行域是五边形ABCDO的内部（包括边界）。

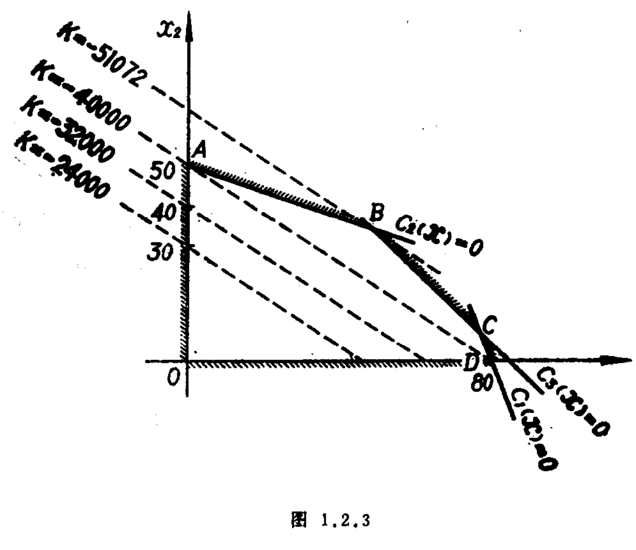

图 1.2.3

由此易见问题的解x^*位于顶点B处，它是直线$C_2(x)=0$和$C_3(x)=0$的交点

$$x^*=(x_1^*,\ x_2^*)^T=(47.44,\ 34.19)^T$$

该点处的目标函数值为$f(x^*)=-51072$。

例1.2.3 设例1.1.3中的已知参数取下例各值

$$\bar{s}=76\text{厘米},\quad \bar{t}=0.25\text{厘米},\quad P=15000\text{公斤}$$

$$\sigma_y=4200\text{公斤},\quad E=2.1\times10^6\text{公斤/厘米}^2,\quad \rho=8.30\times10^{-3}\text{公斤/厘米}^3$$

试讨论该问题的解的几何意义。

首先考虑在$(x_1,\ x_2)=(d,\ H)$平面上达到屈服极限的点的轨迹

$$C_1(x_1,\ x_2)=\frac{P}{\pi\bar{t}}\ \frac{(\bar{s}^2+x_2^2)^{1/2}}{x_1x_2}-\sigma_y=0$$

把它画在图1.2.4中。就这个约束而言，该曲线右上方的点是可以接受的，另一侧是不可接

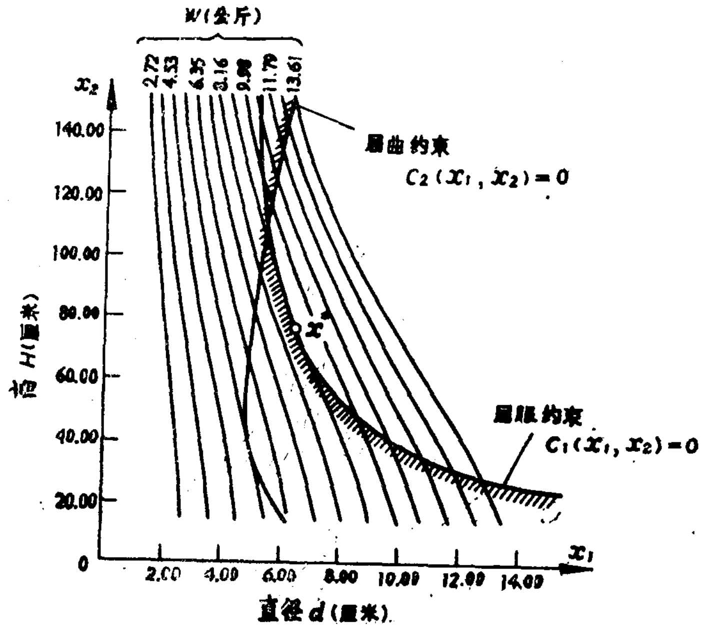

图 1.2.4

受的。同样可画出达到屈曲极限的点的轨迹

$$C_2(x_1,\ x_2)=\frac{P(\bar{s}^2+x_2^2)^{1/2}}{\pi\bar{t}x_1x_2}-\frac{\pi^2E(\bar{t}^2+x_1^2)}{8(\bar{s}^2+x_2^2)}=0$$

就这个约束而言，该曲线右方的点是可接受的，另一侧是不可接受的。由此可见，图1.2.4中的区域D是可行域。结合目标函数的等高线可以看出，最轻重量设计在点x^*处达到，这里x^*是目标函数等高线与约束曲线$C_1(x_1,\ x_2)=0$的切点。

2.3 算法和算法的收敛性

本书介绍的求解无约束问题（1.2.1）或约束问题（1.2.6）的方法是可以在数字电子计算机上实现的，这些方法称为算法，每个算法都是在试图按一定规则构造出逐渐接近于问题解x^*的一个点列$\{x^{(k)}\}$。如果某算法构造出的点列$\{x^{(k)}\}$能够在有限步内达到问题的解x^*，或者收敛到x^*，那么就称该算法对所论问题是收敛的。显然，只是具有收敛性的算法才是有意义的，我们建立的算法应该对于相当广泛的一类问题具有这种收敛性质。

第二章　无约束问题的局部解

及下降算法概述

在第一章中，我们已经说过，所谓求解一个无约束最优化问题，就是寻找目标函数$f(\mathbf{x})$在整个R^n上的最小点$\mathbf{x}^*$，然而一般来说这是一个很困难的问题，至今没有很好地解决，现有的方法大多只能保证近似地找出目标函数$f(\mathbf{x})$在某个局部范围内的最小点，即所谓问题的局部解（其确切含义将在下面介绍）。要想求得整体的最小点（即问题的整体解），可以设法多找出几个局部解，然后把其中使目标函数$f(\mathbf{x})$取值最小的那个局部解近似地做为欲求的解点。

本章主要解决以下两个问题：

(i) 讨论无约束问题局部解的性质，导出它的必要条件和充分条件。

(ii) 讨论求解无约束问题的一般下降算法的基本思想，以及我们希望算法具有的某些性质。

§1　无约束问题的局部解

1.1 无约束问题的局部解

我们从一个例子讲起。

例2.1.1　考虑两个变量的无约束问题

$$\min f(\mathbf{x}) = 4 + \frac{9}{2}x_1 - 4x_2 + x_1^2 + 2x_2^2 - 2x_1x_2 + x_1^4 - 2x_1^2x_2$$

该问题的目标函数的等高线如图2.1.1所示。由于$f(\mathbf{x}^*)=0.5134$，$f(\widetilde{\mathbf{x}}^*)=0.9855$，所以$\mathbf{x}^*$是整体最小点，但是就局部范围来说，$\mathbf{x}^*$和$\widetilde{\mathbf{x}}^*$都是最小点，这种点的特点是，与其附近各点的目标函数值相比较，该点处函数值最小，我们称这种点为问题的局部解，其严格定义如下：

定义2.1.1　考虑无约束问题（UP）。若对$\mathbf{x}^*=(x_1^*, \cdots, x_n^*)^T \in R^n$，存在着$\varepsilon>0$，使得当$\mathbf{x}=(x_1, \cdots, x_n)^T$与$\mathbf{x}^*$的距离小于$\varepsilon$，即当$\|\mathbf{x}-\mathbf{x}^*\|<\varepsilon$时，总有

$$f(\mathbf{x}) \geqslant f(\mathbf{x}^*) \tag{2.1.1}$$

则称$\mathbf{x}^*$为问题的局部解（简称问题的解）或目标函数$f(\mathbf{x})$的局部极小点。若当$\|\mathbf{x}-\mathbf{x}^*\|<\varepsilon$，$\mathbf{x} \neq \mathbf{x}^*$时，总有

$$f(\mathbf{x}) > f(\mathbf{x}^*) \tag{2.1.2}$$

则称$\mathbf{x}^*$为问题的严格局部解（简称严格解）或目标函数$f(\mathbf{x})$的严格局部极小点。

不难看出，按照上述定义来判断，x^*和$\tilde{x}^*$都是例2.1.1所述问题的局部解。

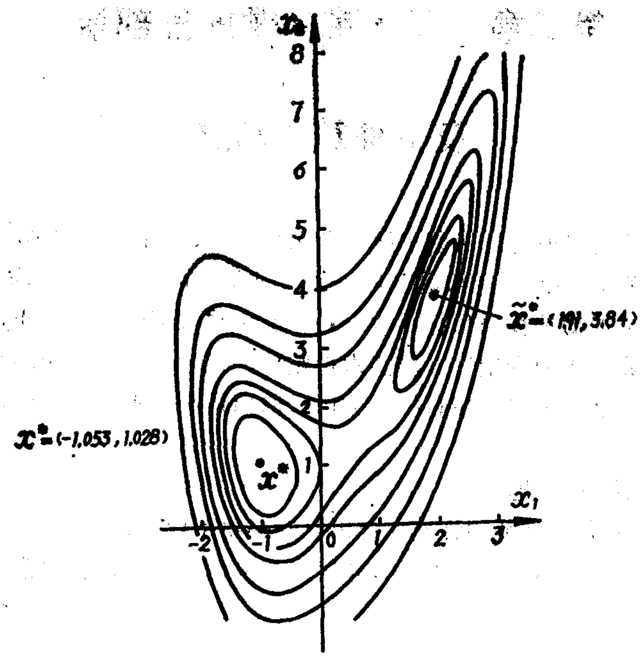

图 2.1.1

1.2 无约束问题局部解的必要条件和充分条件

一般来说，按照定义2.1.1检验局部解是困难的，因为这牵涉到附近所有点上函数值的大小。下面我们讨论局部解的必要条件和充分条件。

定理2.1.1 （无约束问题解的一阶必要条件）若函数f(x)连续可微且x^*是问题（UP）的局部解，则f(x)在x^*处沿任意方向的方向导数都是零，或者等价地说，f(x)在x^*处的梯度向量是零向量即

$$\mathbf{g}(x^*)=\nabla f(x^*)=0 \qquad (2.1.3)$$

证明 任取单位向量$\mathbf{d}\in R^n$，令

$$\varphi(\alpha)=f(x^*+\alpha\mathbf{d})$$

因为x^*是函数f(x)的局部极小点，所以$\alpha=0$是函数$\varphi(\alpha)$的局部极小点（图2.1.2画出了$n=2$的情形）从而

$$0=\varphi'(0)=\langle \mathbf{g}(x^*),\ \mathbf{d}\rangle \qquad (2.1.4)$$

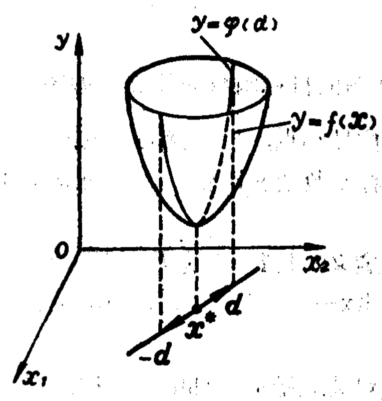

图 2.1.2

这表明f(x)在x^*处沿方向$\mathbf{d}$的方向导数为零。因为$\mathbf{d}$是任意的单位向量，故有$\mathbf{g}(x^*)=0$。定理证毕。

定理2.1.2 （无约束问题解的二阶必要条件）若函数 $f(x)$ 二次连续可微且 x^* 是问题（UP）的局部解，则

(i) $g(x^*)=0$；

(ii) $f(x)$ 在 x^* 处沿任意方向的二阶方向导数都是非负数，或者等价地说，$f(x)$ 的Hess矩阵 $G(x^*)=\nabla^2 f(x^*)$ 半正定。

证明 由定理2.1.1知 $g(x^*)=0$，现只需证明结论（ii）成立。任取单位向量 $d\in R^n$，令

$$\varphi(\alpha)=f(x^*+\alpha d)$$

它的二阶Taylor展式为

$$\varphi(\alpha)-\varphi(0)=\varphi'(0)\alpha+\frac{1}{2}\varphi''(0)\alpha^2+o(\alpha^2)$$

因为 $\alpha=0$ 是 $\varphi(\alpha)$ 的局部极小点，并注意到 $\varphi'(0)=0$，便可推知当 α 充分小时有

$$0\leqslant\varphi(\alpha)-\varphi(0)=\frac{1}{2}\varphi''(0)\alpha^2+o(\alpha^2)$$

由上式可知，当 $\alpha\neq0$ 时有

$$\frac{1}{2}\varphi''(0)+\frac{o(\alpha^2)}{\alpha^2}\geqslant0$$

令 $\alpha\to0$，得到

$$\varphi''(0)=\langle d,\ G(x^*)d\rangle\geqslant0 \tag{2.1.5}$$

式（2.1.5）表明 $f(x)$ 在 x^* 处沿方向 d 的二阶方向导数是非负数，式（2.1.5）同时表明Hess矩阵 $G(x^*)$ 半正定。定理证毕。

定理2.1.3 （无约束问题解的二阶充分条件）若函数 $f(x)$ 二次连续可微，且在 x^* 处满足下列两个条件：

(i) $g(x^*)=0$

(ii) $f(x)$ 在 x^* 处沿任意方向的二阶方向导数都是正数，或者等价地说，$f(x)$ 的Hess矩阵 $G(x^*)=\nabla^2 f(x^*)$ 正定，

则 x^* 是问题（UP）的严格局部解。

说明 设 d 为单位向量，考虑从 x^* 出发沿方向 d 移动时目标函数值的变化情况。根据Taylor展开式有

$$f(x^*+\alpha d)-f(x^*)=\langle g(x^*),d\rangle\alpha+\frac{1}{2}\langle d,G(x^*)d\rangle\alpha^2+\cdots$$

但由定理条件知

$$\langle g(x^*),d\rangle=0,\ \langle d,G(x^*)d\rangle>0$$

所以从

$$f(x^*+\alpha d)-f(x^*)=\frac{1}{2}\langle d,G(x^*)d\rangle\alpha^2+\cdots$$

可以看出，只要步长 α 充分小（但 $\alpha\neq0$）就有

$$f(x^*+\alpha d)-f(x^*)>0\text{或}f(x^*+\alpha d)>f(x^*)$$

由于d是任意的单位向量，所以上式意味着从 x^* 出发沿任意方向作微小移动时，都会使函数值上升，因而 x^* 应该是问题（UP）的严格局部解。定理说明完毕。

例2.1.2 试证正定二次函数

$$f(x)=\frac{1}{2}x^TGx+r^Tx+\delta$$

（其中G是 $n\times n$ 阶正定对称矩阵，r 是某个 n 维向量，δ是某个数）有唯一的严格局部极小点，并说明此极小点的几何意义。

证明 因为 $g(x^*)=Gx+r$，所以由定理2.1.1知 $f(x)$ 的极小点应满足

$$Gx+r=0$$

注意到矩阵G正定，所以此方程有唯一解。

$$x^*=-G^{-1}r$$

这表明 $f(x)$ 最多只有一个极小点 $x^*=-G^{-1}r$。另一方面注意到 $f(x)$ 在 x^* 处的Hess矩阵 $G(x^*)=G$ 正定，由定理 2.1.3 即知 x^* 的确是 $f(x)$ 的严格局部极小点。

经过直接推算可以得到 $f(x)$ 的另一表达式

$$f(x)=\frac{1}{2}(x-x^*)^TG(x-x^*)+f(x^*)$$

由此易见 $f(x)$ 的等高线是一族超椭球面，极小点 x^* 恰好是其中心，参看图2.1.3。

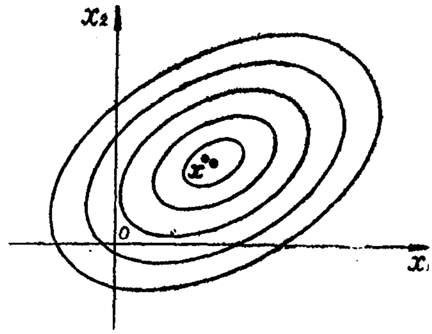

图 2.1.3

§2 从最速下降法到一般下降算法

2.1 最速下降法

1. 二元函数的最速下降法

考虑两个变量的无约束问题

$$\min f(x)=f(x_1,\ x_2) \tag{2.2.1}$$

不妨设想问题的目标函数的等高线如图2.2.1所示，其解为 x^*。由于我们总能得到 x^* 的一个估值 $x^{(1)}$（不管它们之间的接近程度如何），所以可把寻找 x^* 的问题归结为：如何从某点 $x^{(1)}$ 出发逐渐行进到 x^*。这样，首先要解决的问题是，在点 $x^{(1)}$ 处应该沿哪个方向前进。注意到我们的目的是要达到函数取极小值的点，自然会想到选取使函数值下降最快的方向。现在考查哪个方向具有这性一性质，记等高线 $f(x)=f(x^{(1)})=6$ 在 $x^{(1)}$ 处的切线为 t，显然与 t 垂直的方向 d 就是使函数值下降最快的方向，而我们知道，它恰好是函数 $f(x)$ 在 $x^{(1)}$ 处的负梯度方向

$$d=-g(x^{(1)})$$

这样说来，要从 $x^{(1)}$ 行进到 x^*，就可以考虑时时都沿着负梯度方向前进。鉴于梯度方向和等