工程优化方法

GONGCHENG YOUHUA FANGFA

陈卫东 蔡荫林 于诗源
吴限德 刘淼群 张文松 ▶ 编著

HEUP 哈爾濱工程大學出版社

工程优化方法

陈卫东　蔡荫林　于诗源
吴限德　刘淼群　张文松　编著

HEUP 哈爾濱工程大学出版社

内容简介

本书阐述工程优化方法的基本理论和算法，内容主要包括线性规划、非线性规划、整数规划、模糊规划和多目标规划，并对如何建立数学模型、如何选择优化方法和提高优化效率，以及若干新算法做了适当的介绍。书中从工程应用的角度出发，注重算法基本思想和方法的阐述，力求深入浅出，通俗易懂。

本书为工科研究生的教材或主要参考书，也可作为航空、航天、船舶、工程力学、机械、土木等有关专业高年级本科生的选修教材或主要参考书，亦可供工程技术人员自学和参考。

图书在版编目(CIP)数据

工程优化方法／陈卫东等编著．—哈尔滨：哈尔滨工程大学出版社，2017.11
ISBN 978-7-5661-1718-2

Ⅰ.①工… Ⅱ.①陈… Ⅲ.①工程-最优设计
Ⅳ.①TB21

中国版本图书馆 CIP 数据核字(2017)第 269119 号

责任编辑 王洪菲
封面设计 博鑫设计

出版发行 哈尔滨工程大学出版社
社　　址 哈尔滨市南岗区东大直街 124 号
邮政编码 150001
发行电话 0451-82519328
传　　真 0451-82519699
经　　销 新华书店
印　　刷 北京中石油彩色印刷有限责任公司
开　　本 787 mm×960 mm　1/16
印　　张 27
字　　数 592 千字
版　　次 2017 年 11 月第 1 版
印　　次 2017 年 11 月第 1 次印刷
定　　价 81.00 元
http://www.hrbeupress.com
E-mail:heupress@hrbeu.edu.cn

前　　言

工程优化设计是20世纪60年代以来出现的一种工程设计方法。它以最优化数学理论为基础,借助电子计算机来合理地选择设计方案。国内外大量的实践已经表明,当一个设计所希望达到的目标以及必须满足的限制条件都能用数学关系式表达时,这种方法既可以大大缩短设计周期,又可使设计质量显著提高。而且随着最优化理论、方法和电子计算技术的迅速发展,这种应用显得越来越迫切,而且越来越广泛。因此,越来越多的工程技术人员希望能了解和掌握工程优化设计的基本理论和方法。国内不少理工科大学已把工程优化方法作为研究生或高年级本科生的课程。

本书是在我们已有的"工程优化方法"研究生讲义的基础上编写的。该讲义经多年应用,得到学生的一致好评。本次编写中,一是增加了十几年来发展起来的新的理论和方法,如线性规划的内点算法及进化算法、信赖域方法、极大熵方法,等等;二是将我们20余年来,应用工程优化方法于结构优化设计与基于可靠性的结构优化设计方面的研究成果,给予了适当的介绍,用以作为若干算法在工程设计中应用的实例。

全书共分为10章。第1章给出了优化问题的数学表达和基本概念;第2章是无约束优化方法,它们可直接用来解决实际问题,又可以作为求解约束问题的工具;第3章介绍线性规划的解法;第4章讨论非线性规划问题的计算方法,首先简要地做一些理论讨论,然后把讨论的问题与第2章、第3章的内容相联系,并介绍若干求解非线性规划问题的有效算法;第5章介绍离散变量优化与整数规划;第6章介绍模糊规划;第7章介绍多目标规划;第8章集中介绍了优化方法的若干新进展;第9章提供了选择优化算法的一般原则和提高优化算法效率的途径,其中许多内容是作者20余年来应用优化方法的体会;第10章较详细介绍了作者应用优化算法于工程实践的实例。本书的最后,提供了所介绍的优化理论、方法或应用它们的原始文献,供读者参考。相信这些内容对读者,特别是对工程技术人员会有所裨益。

本书主要面向工科研究生、高年级本科生和工程技术人员。因此编写中着重介绍最优化理论和方法的基本原理和在实际中比较有效的计算方法,尽可能把所介绍的方法的基本思想、理论分析、计算的方法和步骤阐述清楚,并介绍应用算法的例题,而不是主要对所介绍的方法进行严密的数学推导和证明。在编写中尽量做到思路清晰,深入浅出,通俗易懂。凡具有微积分和线性代数知识的读者都能读懂本书大部分内容。顺便指出,若将本书作为研究生的教材或主要参考书,带"＊＊"号的章节为选学内容;若作为高年级本科生教材或主要参考书,可选

择未带“ * ”和“ * * ”号的章节。为了帮助学习和复习,在每章后面附有习题和复习思考题,习题中带“ * ”的为选做题。

由于编者的水平有限,书中如有不妥之处,恳请批评指正。

编著者

2017年8月于哈尔滨工程大学

主要符号的说明

$\boldsymbol{X}=[x_1,x_2,\cdots,x_n]$——$n$ 维行向量

$\boldsymbol{X}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$——$n$ 维列向量,其中的 T 表示转置

$\boldsymbol{C}=\{X \mid X$ 所满足的性质$\}$—— 满足某种性质的 X 的全体(集合)

$\boldsymbol{R}^n$——n 维实数空间

min—— 极小化

max—— 极大化

s. t. —— 在 …… 约束条件下满足于,是“Subject to” 的缩写

$f(\boldsymbol{X})$—— 目标函数

$\nabla f(\boldsymbol{X})$——$f(\boldsymbol{X})$ 的梯度

$\nabla^2 f(\boldsymbol{X})$——$f(\boldsymbol{X})$ 的二阶导数矩阵

$g_i(\boldsymbol{X})$—— 第 i 个不等式约束

$h_j(\boldsymbol{X})$—— 第 j 个等式约束

$\boldsymbol{X}^*$—— 最优点

$f(\boldsymbol{X}^*)$—— 最优目标函数值

$\|\cdot\|$—— · 的模

$|\cdot|$—— · 的绝对值

$\subset$—— 被包含

$\supset$—— 包含

$\in$—— 属于

$\notin$—— 不属于

$\forall$—— 对于任意的

$\approx$—— 近似等于

$\equiv$—— 恒等

$\vee$—— 集合的并集

$\wedge$—— 集合的交集

$\varnothing$—— 空集

【—— 证明的开始

】—— 证明的结束

目　　录

第1章　极值理论与最优化问题的数学表达

所谓**最优化**(Optimization)就是追求最好的结果或最优的目标。因此它是在所有可能方案中选择最合理的一种方案以达到最优目标的一门学科。而寻求最优方案的方法就是最优化方法。从数学上讲,凡是追求最优目标的数学问题都属于最优化问题。也就是说,最优化问题是在一定约束条件下寻求函数极值的问题。可见,作为最优化问题,第一,有多个可供选择的方案;第二,有追求的目标,而且追求的目标是方案的函数。

最简单的最优化问题在微积分中已经遇到过,就是函数极值问题。例如:对边长为 a 的正方形铁板,在四个角处剪去相等的正方形以制成方形无盖水槽,问如何剪法可使水槽的容积最大?如图1-1所示,设剪去的正方形边长为 x,则水槽的容积是

$$f(x) = (a-2x)^2 x$$

令 $f'(x)=0$ 可解得,$x_1=\frac{a}{2}$ 和 $x_2=\frac{a}{6}$,x_1 不符题意,故 x_2 为所求。由于 $f''(x)\big|_{x=\frac{a}{6}}=-4a<0$,说明所求 $x_2=\frac{a}{6}$ 的确是使容积 $f(x)$ 最大的极值点。

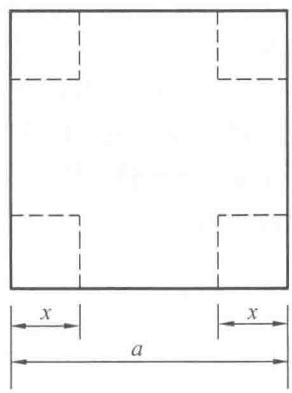

图1-1

又例如:把半径为1的实心金属球熔化后铸成一个实心圆柱体,问圆柱体取什么样的尺寸才能使它的表面积最小?

设所铸成的圆柱体的底面半径为 r,高为 h,表面积为 s,则这个问题可以表示为求 r 与 h,条件是

$$r^2h-\frac{4}{3}=0$$

使 $s=2\pi rh+2\pi r^2$ 最小。

这个问题可以采用拉格朗日(Lagrange)乘子法求解这个有等式约束的函数极值问题。拉格朗日函数是

$$L(r,h,\lambda)=2\pi rh+2\pi r^2+\lambda\left(r^2h-\frac{4}{3}\right)$$

将其分别对 r,h,λ 求偏导并令其等于零,得到联立代数方程

$$\left.\begin{aligned}\frac{\partial L}{\partial r} &= 2\pi h + 4\pi r + 2rh\lambda = 0\\ \frac{\partial L}{\partial h} &= 2\pi r + \lambda r^2 = 0\\ \frac{\partial L}{\partial \lambda} &= r^2 h - \frac{4}{3} = 0\end{aligned}\right\}$$

由第一、二个方程，得 $h = 2r$，再与第三个方程联立可解出

$$r = \left(\frac{2}{3}\right)^{\frac{1}{3}} \text{和} h = 2\left(\frac{2}{3}\right)^{\frac{1}{3}}$$

此时圆柱体的表面积是 $s = 6\pi\left(\frac{2}{3}\right)^{\frac{2}{3}}$。

还有，在变分学中，把求泛函的极值问题化为求解相应的微分方程等。上面所述求解极值问题的方法都归结为非线性方程组的求解，如果不借助于计算机，它们只有在极特殊的情况下才能求解出来。因此，通常微积分的极值问题大都限于一元和二元的问题，我们称它为经典的最优化方法。

近几十年来，特别是第二次世界大战以后，随着运筹学的形成和发展，以及电子计算机的飞速发展，使最优化方法获得迅速的发展，创立了许多新的理论和方法来求解各种大型问题，从而形成了近代的最优化理论和方法。同时，由于它是一门新的学科，还不很成熟，许多问题还有待解决。

1.1　极值理论简介

为了介绍近代最优化方法，本节对微分学应用中的极值理论作一简要的介绍。其中，有关的概念，例如梯度、方向导数等，是十分重要的，是学习以后各章的重要基础。

1.1.1　一元函数的极值

图 1－2 所示为定义在区间$[a,b]$上的函数$f(x)$的图形。我们分别称x_1和x_2为函数$f(x)$的极大点和极小点，统称为$f(x)$的极值点。而称$f(x_1)$和$f(x_2)$为函数$f(x)$的极大值和极小值，统称为函数$f(x)$的极值。

函数的极值点是函数的一阶导数为零的点。即对于在$[a,b]$内处处有一阶导数的函数$f(x)$，极值点x存在的必要条件为

$$f'(x) = 0 \tag{1-1}$$

在一般情况下，函数的一阶导数$f'(x) = 0$的点并不一定都是极值点。我们称使$f(x)$的一

阶导数$f'(x)=0$的点为函数$f(x)$的驻点。可见极值点必为驻点，而驻点不一定是极值点。要确定驻点是否是极值点还要考虑函数$f(x)$在该点附近的变化情况。即：若函数$f(x)$在该点的左边递增，而在右边递减，则可以断定该点为$f(x)$的一个极大点，如图1－2中的x_1点；如果$f(x)$在该点左边递减，而在右边递增，则可以断定该点为$f(x)$的一个极小点，如图1－2中的x_2点。上述即是判定驻点是否为极值点的充分条件。因为函数$f(x)$值的递增与递减又可由函数二阶导数表示，故极值点的充分必要条件可归结为：若函数在驻点附近有连续的二阶导数且小于零($f''(x)<0$)时则该点为极大点；若函数在驻点附近有连续的二阶导数且大于零($f''(x)>0$)时则该点为极小点。

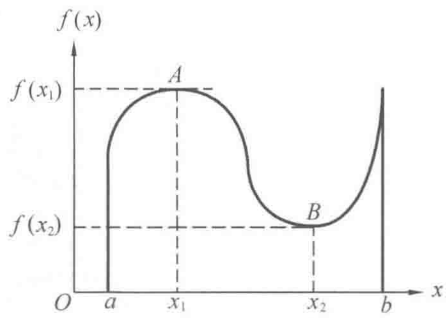

图1－2

例1－1 求$y=\sin x$在$[0,\pi]$内的极值点。

解 求导数$y'=\cos x$。

解方程$y'=\cos x=0$，得$x=\dfrac{\pi}{2}$为函数$\sin x$的驻点。

求二阶导数$y''=-\sin x$。在$x=\dfrac{\pi}{2}$处，$y''\left(\dfrac{\pi}{2}\right)=-\sin\left(\dfrac{\pi}{2}\right)=-1<0$，故$x=\dfrac{\pi}{2}$为函数$y=\sin x$的极大点，见图1－3。

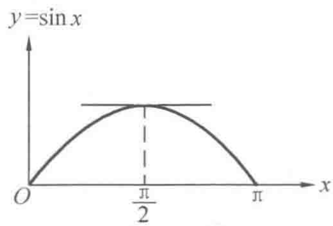

图1－3

1.1.2 二元函数的极值

二元函数$f(x,y)$的图形可以表示成空间的一个曲面。图1－4所示p点为函数$z=f(x,y)$在该点附近为极大点与极小点的情况。从图1－4可知，二元函数极值点存在的必要条件是过该点(x_0,y_0)函数$z=f(x,y)$的切面平行于xy平面。由于切面方程为

$$z-f(x_0,y_0)=f'_x(x_0,y_0)(x-x_0)+f'_y(x_0,y_0)(y-y_0)$$

式中$f'_x(x_0,y_0)$和$f'_y(x_0,y_0)$分别是函数$f(x,y)$在(x_0,y_0)点处对x和y的一阶偏导数，要它和xy平面平行，而x和y又有一定的任意性，故必须

$$\left.\begin{aligned}f'_x(x_0,y_0)&=0\\f'_y(x_0,y_0)&=0\end{aligned}\right\}\qquad(1-2)$$

式(1－2)即为点(x_0,y_0)为二元函数$f(x,y)$极值点的必要条件。我们称满足式(1－2)的点(x_0,y_0)为函数$f(x,y)$的驻点。一般，驻点不一定是极值点，但极值点必为驻点。

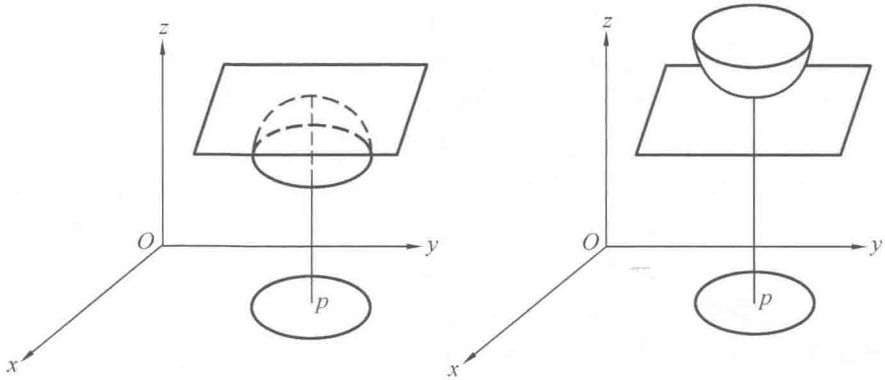

图 1－4

可以证明，点(x_0,y_0)为极值点的充分必要条件是

$$\left.\begin{aligned}&f'_x(x_0,y_0)=f'_y(x_0,y_0)=0\\&[f''_{xy}(x_0,y_0)]^2-f''_{xx}(x_0,y_0)f''_{yy}(x_0,y_0)<0\\&f''_{xx}(x_0,y_0)<0\ \text{时为极大点}\\&f''_{xx}(x_0,y_0)>0\ \text{时为极小点}\end{aligned}\right\}\tag{1-3}$$

式中$f''_{xx}(x_0,y_0)$，$f''_{xy}(x_0,y_0)$和$f''_{yy}(x_0,y_0)$是函数$f(x,y)$在点(x_0,y_0)处的二阶偏导数。

1.1.3　多元函数的极值

为表达方便，我们用向量（列阵）和矩阵表达n维空间点和函数的表达式。在n维欧氏空间$\boldsymbol{R}^n$内区域$\boldsymbol{D}$上的函数可表示为

$$f(\boldsymbol{X})\qquad(\boldsymbol{X}\in\boldsymbol{D}\subset\boldsymbol{R}^n)$$

其中向量$\boldsymbol{X}$的分量$x_1,x_2,\cdots,x_n$为函数$f(\boldsymbol{X})$的自变量。符号“$\in$”意为“属于”，“$\subset$”意为“被包含”。

1.1.3.1　函数$f(\boldsymbol{X})$的梯度　$f(\boldsymbol{X})$

若$f(\boldsymbol{X})$在$\boldsymbol{X}_0$处可微，则$f(\boldsymbol{X})$在该点必存在关于各变量的一阶偏导数。我们把这个$f(\boldsymbol{X})$在$\boldsymbol{X}_0$处的n个偏导数分量组成的向量称为$f(\boldsymbol{X})$在该点的**梯度**（Gradient），记作　$f(\boldsymbol{X}_0)$，即有

$$\nabla f(\boldsymbol{X}_0)=\left[\frac{\partial f(\boldsymbol{X}_0)}{\partial x_1},\frac{\partial f(\boldsymbol{X}_0)}{\partial x_2},\cdots,\frac{\partial f(\boldsymbol{X}_0)}{\partial x_n}\right]^{\mathrm{T}}\tag{1-4}$$

梯度也可以称为函数$f(\boldsymbol{X})$关于变量向量的一阶导数。其长度（模）为

$$\|\nabla f(\boldsymbol{X}_0)\|=[\nabla f(\boldsymbol{X}_0)^{\mathrm{T}}\nabla f(\boldsymbol{X}_0)]^{\frac{1}{2}}=\left[\sum_{i=1}^{n}\left(\frac{\partial f(\boldsymbol{X}_0)}{\partial x_i}\right)^2\right]^{\frac{1}{2}}\tag{1-5}$$

梯度有两个重要的性质。

第一,函数在某点的梯度若不为零,则必与过该点的函数等值面相“垂直”,如图 1 - 5 所示。图中 L 是等值面 $f(\boldsymbol{X}_0)$ 上的任一光滑曲线,S 是过点 $\boldsymbol{X}_0$ 的 L 的切线。即有

$$\nabla f(\boldsymbol{X}_0)^{\mathrm{T}} S = 0 \qquad (1-6)$$

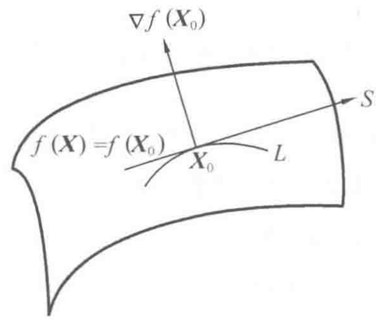

图 1 - 5

第二,梯度方向是函数在该点函数值具有最大变化率的方向。

以上两性质在最优化方法中非常重要,但要注意它只是函数在 $\boldsymbol{X}_0$ 点处附近的局部性质。

下面是一些特殊类型函数的梯度公式,我们常常要用到它们。

(1) 若 $f(\boldsymbol{X}) = c$(常数),则$\nabla f(\boldsymbol{X}) = 0$,即

$$c = 0 \qquad (1-7)$$

这是因为 c 对于各个变量的偏导数都为零。

(2)
$$(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}) = \boldsymbol{B} \qquad (1-8)$$

式中的 $\boldsymbol{B}$ 是分量全为常数的 n 维向量,$\boldsymbol{X}$ 也是 n 维向量。

(3)
$$(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}) = 2\boldsymbol{X} \qquad (1-9)$$

式中的 $\boldsymbol{X}$ 是 n 维向量。

(4) 若 $\boldsymbol{H}$ 是 $n \times n$ 阶对称方阵,则

$$(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{X}) = 2\boldsymbol{H}\boldsymbol{X} \qquad (1-10)$$

若函数为一向量值函数 $\boldsymbol{G}(x)$,它有 m 个分量,即有 $\boldsymbol{G}(\boldsymbol{X}) = [g_1(\boldsymbol{X}), g_2(\boldsymbol{X}), \cdots, g_m(\boldsymbol{X})]^{\mathrm{T}}$,则定义其在 $\boldsymbol{X}_0$ 点的梯度为

$$\boldsymbol{G}(\boldsymbol{X}_0) = \begin{bmatrix} \dfrac{\partial g_1(\boldsymbol{X}_0)}{\partial x_1} & \dfrac{\partial g_1(\boldsymbol{X}_0)}{\partial x_2} & \cdots & \dfrac{\partial g_1(\boldsymbol{X}_0)}{\partial x_n} \\ \dfrac{\partial g_2(\boldsymbol{X}_0)}{\partial x_1} & \dfrac{\partial g_2(\boldsymbol{X}_0)}{\partial x_2} & \cdots & \dfrac{\partial g_2(\boldsymbol{X}_0)}{\partial x_n} \\ \cdots & \cdots & & \cdots \\ \dfrac{\partial g_m(\boldsymbol{X}_0)}{\partial x_1} & \dfrac{\partial g_m(\boldsymbol{X}_0)}{\partial x_2} & \cdots & \dfrac{\partial g_m(\boldsymbol{X}_0)}{\partial x_n} \end{bmatrix} \qquad (1-11)$$

上式中的矩阵也称为向量函数 $\boldsymbol{G}(\boldsymbol{X})$ 在点 $\boldsymbol{X}_0$ 处的**雅可比**(Jacobi)矩阵。

常用的几个向量值函数的梯度公式如下。

(1)
$$\boldsymbol{C} = \boldsymbol{0} \qquad (1-12)$$

式中的 $\boldsymbol{C}$ 是分量全为常数的 n 维向量,$\boldsymbol{0}$ 是 $n \times n$ 阶零矩阵。这是因为常数关于变量 $x_1, x_2, \cdots, x_n$

的偏导数都为零。

(2)
$$X = I \tag{1-13}$$
式中 $\boldsymbol{X}$ 是 n 维向量,$\boldsymbol{I}$ 是 $n \times n$ 阶单位矩阵。

(3)
$$(\boldsymbol{HX}) = \boldsymbol{H} \tag{1-14}$$
式中 $\boldsymbol{H}$ 是 $n \times n$ 阶对称方阵,$\boldsymbol{X}$ 为 n 维向量。

1.1.3.2 方向导数

设在变量空间中一点 $\boldsymbol{X}_0$,对应的函数值为 $f(\boldsymbol{X}_0)$,若通过 $\boldsymbol{X}_0$ 点有某固定不变的向量 $\boldsymbol{P}$,$\boldsymbol{E}$ 是方向 $\boldsymbol{P}$ 上的单位向量,则当变量从 $\boldsymbol{X}_0$ 点沿方向 $\boldsymbol{P}$ 移动到距离为 t 时有
$$\boldsymbol{X} = \boldsymbol{X}_0 + t\boldsymbol{E} \tag{1-15}$$
则称极限
$$\lim_{t \to 0} \frac{f(\boldsymbol{X}_0 + t\boldsymbol{E}) - f(\boldsymbol{X}_0)}{t} = \frac{\partial f(\boldsymbol{X}_0)}{\partial \boldsymbol{P}} = f'_p(\boldsymbol{X}_0) \tag{1-16}$$
为函数 $f(\boldsymbol{X})$ 在点 $\boldsymbol{X}_0$ 处沿方向 $\boldsymbol{P}$ 的**方向导数**(Directional derivative)。

可见,若$\dfrac{\partial f(\boldsymbol{X}_0)}{\partial \boldsymbol{P}} < 0$,则 $f(\boldsymbol{X})$ 从 $\boldsymbol{X}_0$ 出发在 $\boldsymbol{X}_0$ 附近沿 $\boldsymbol{P}$ 方向是下降的;若$\dfrac{\partial f(\boldsymbol{X}_0)}{\partial P} > 0$,则 $f(\boldsymbol{X})$ 从 $\boldsymbol{X}_0$ 出发在 $\boldsymbol{X}_0$ 附近沿 $\boldsymbol{P}$ 方向是上升的。

可以证明,函数 $f(\boldsymbol{X})$ 在点 $\boldsymbol{X}_0$ 处沿 $\boldsymbol{P}$ 的方向导数,等于该点函数梯度与单位向量 $\boldsymbol{E}$ 的点乘积,即有
$$f'_P(\boldsymbol{X}_0) = \nabla f(\boldsymbol{X}_0)^{\mathrm{T}} \cdot \boldsymbol{E} \tag{1-17}$$
【证 $f(\boldsymbol{X})$ 在点 $\boldsymbol{X}_0$ 的泰勒展开式为
$$f(\boldsymbol{X}_0 + \boldsymbol{P}) = f(\boldsymbol{X}_0) + \nabla f(\boldsymbol{X}_0)^{\mathrm{T}}\boldsymbol{P} + 0(\|\boldsymbol{P}\|)$$
此时有 $f(\boldsymbol{X}_0 + t\boldsymbol{E}) = f(\boldsymbol{X}_0) + \nabla f(\boldsymbol{X}_0)^{\mathrm{T}}(t\boldsymbol{E}) + 0(\|t\boldsymbol{E}\|)$,由式(1-16)有
$$\begin{aligned} f'_P(\boldsymbol{X}_0) &= \lim_{t \to 0} \frac{f(\boldsymbol{X}_0) + \nabla f(\boldsymbol{X}_0)^{\mathrm{T}}(t\boldsymbol{E}) + 0(\|t\boldsymbol{E}\|) - f(\boldsymbol{X}_0)}{t} \\ &= \lim_{t \to 0}\left(\nabla f(\boldsymbol{X}_0)^{\mathrm{T}}\boldsymbol{E} + \frac{0(\|t\boldsymbol{E}\|)}{t}\right) = \nabla f(\boldsymbol{X}_0)^{\mathrm{T}}\boldsymbol{E} \end{aligned}$$
】

从而可见,若
$$\nabla f(\boldsymbol{X}_0)^{\mathrm{T}}\boldsymbol{P} < 0 \tag{1-18}$$
则 $\boldsymbol{P}$ 的方向是函数 $f(\boldsymbol{X})$ 在点 $\boldsymbol{X}_0$ 处的下降方向;若
$$\nabla f(\boldsymbol{X}_0)^{\mathrm{T}}\boldsymbol{P} > 0 \tag{1-19}$$
则 $\boldsymbol{P}$ 的方向是函数 $f(\boldsymbol{X})$ 在点 $\boldsymbol{X}_0$ 处的上升方向。也就是说,方向导数的正负决定了函数的升降,而升降的快慢由它的绝对值大小确定。

由式(1-17)可知
$$f'_P(\boldsymbol{X}_0) = \|\nabla f(\boldsymbol{X}_0)\| \cos\beta \tag{1-20}$$

式中的 β 是梯度 $\nabla f(\boldsymbol{X}_0)$ 与方向 $\boldsymbol{P}$ 之间的夹角。可见当 $\boldsymbol{P}=-\nabla f(\boldsymbol{X}_0)$ 时 $\beta=180°$，$f'_p(\boldsymbol{X}_0)$ 的值最小。因此函数的负梯度方向是它的最速下降方向(参见图 1-6)。同理,梯度方向是函数的最速上升方向。函数在与其梯度垂直(或称正交)的方向上变化率为零,即有(1-6)式。函数在与其梯度成锐角的方向上是上升的,而在成钝角的方向上是下降的。

例 1-2 求函数 $f(\boldsymbol{X})=x_1^2+x_2^2+1$ 在点 $\boldsymbol{X}_0=[0,3]^{\mathrm{T}}$ 处的最速下降方向,并求沿这个方向移动一个单位长后新点的函数值。

解 因为

$$\frac{\partial f(\boldsymbol{X})}{\partial x_1}=2x_1,\ \frac{\partial f(\boldsymbol{X})}{\partial x_2}=2x_2$$

所以最速下降方向是 $-\nabla f(\boldsymbol{X}_0)=\begin{bmatrix}-2x_1\\-2x_2\end{bmatrix}_{\boldsymbol{X}=\boldsymbol{X}_0}=\begin{bmatrix}0\\-6\end{bmatrix}$。这个方向上的单位向量是

$\boldsymbol{E}=\dfrac{-\nabla f(\boldsymbol{X}_0)}{\|-\nabla f(\boldsymbol{X}_0)\|}=\begin{bmatrix}0\\-1\end{bmatrix}$，新点是 $\boldsymbol{X}_1=\boldsymbol{X}_0+\boldsymbol{E}=\begin{bmatrix}0\\3\end{bmatrix}+\begin{bmatrix}0\\-1\end{bmatrix}=\begin{bmatrix}0\\2\end{bmatrix}$,对应的函数值为 $f(\boldsymbol{X}_1)=0^2+2^2+1=5$,参见图 1-7。

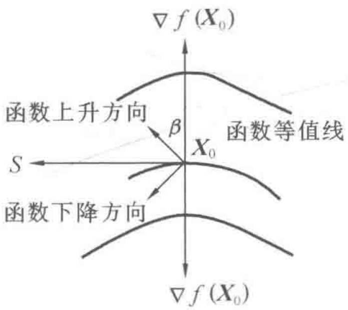

图 1-6

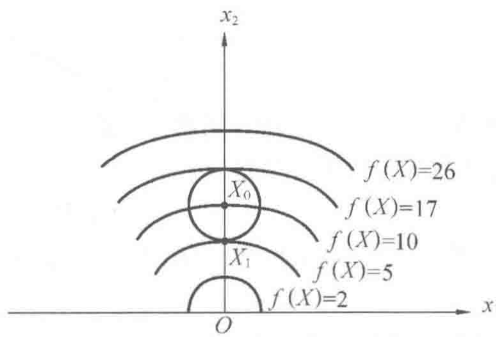

图 1-7

1.1.3.3 多元函数的泰勒(Taylor)展开式与近似表达式

由高等数学可知,若函数 $f(\boldsymbol{X})$ 在点 $\boldsymbol{X}_0$ 的邻域至少有连续的二阶微分,则在 $\boldsymbol{X}_0$ 处的泰勒展开式为

$$f(\boldsymbol{X})=f(\boldsymbol{X}_0)+\sum_{i=1}^{n}\frac{\partial f(\boldsymbol{X}_0)}{\partial x_i}\Delta x_i+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 f(\boldsymbol{X}_0)}{\partial x_i\partial x_j}\Delta x_i\Delta x_j+\text{高阶无穷小量} \tag{1-21}$$

式中 $\Delta x_i=x_i-x_{0i}(i=1,2,\cdots,n)$。

上式可写成矩阵形式

$$f(\boldsymbol{X}_0+\Delta X)=f(\boldsymbol{X}_0)+\nabla f(\boldsymbol{X}_0)^{\mathrm{T}}\Delta X+\frac{1}{2}\Delta \boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}(\boldsymbol{X}_0)\Delta X+\text{高阶无穷小量} \quad (1-22)$$

式中$\nabla f(\boldsymbol{X}_0)$为$f(\boldsymbol{X})$在点$\boldsymbol{X}_0$处的梯度，而

$$\boldsymbol{H}(\boldsymbol{X}_0)= {}^2 f(\boldsymbol{X}_0)= [\nabla f(\boldsymbol{X}_0)]=\begin{bmatrix}\frac{\partial^2 f(\boldsymbol{X}_0)}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{X}_0)}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{X}_0)}{\partial x_1\partial x_n}\\ \frac{\partial^2 f(\boldsymbol{X}_0)}{\partial x_2 x_1} & \frac{\partial^2 f(\boldsymbol{X}_0)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{X}_0)}{\partial x_2\partial x_n}\\ \vdots & \vdots & & \vdots\\ \frac{\partial^2 f(\boldsymbol{X}_0)}{\partial x_n\partial x_1} & \frac{\partial^2 f(\boldsymbol{X}_0)}{\partial x_n\partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{X}_0)}{\partial x_n^2}\end{bmatrix} \quad (1-23)$$

式(1-23)是$f(\boldsymbol{X})$在点$\boldsymbol{X}_0$处的二阶导数矩阵，称为**海辛(Hesse)矩阵**。它的每一元素$\frac{\partial^2 f(\boldsymbol{X}_0)}{\partial x_i\partial x_j}(i,j=1,2,\cdots,n)$是$f(\boldsymbol{X})$在点$\boldsymbol{X}_0$处的二阶偏导数。由于当$f(\boldsymbol{X})$的所有二阶偏导数连续时有

$$\frac{\partial^2 f(\boldsymbol{X}_0)}{\partial x_i\partial x_j}=\frac{\partial^2 f(\boldsymbol{X}_0)}{\partial x_j\partial x_i}(i,j=1,2,\cdots,n) \quad (1-24)$$

所以$\boldsymbol{H}(\boldsymbol{X}_0)$为$n\times n$阶的实对称矩阵。

若对于任何的ΔX，总有

$$\Delta \boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}(\boldsymbol{X}_0)\Delta X\geqslant 0 \quad (1-25)$$

则称$\boldsymbol{H}(\boldsymbol{X}_0)$是半正定的。特别的，若对于$\Delta X\neq 0$恒有

$$\Delta \boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}(\boldsymbol{X}_0)\Delta X>0 \quad (1-26)$$

则称$\boldsymbol{H}(\boldsymbol{X}_0)$是正定的。

当$\boldsymbol{H}(\boldsymbol{X}_0)$正定时，称$-\boldsymbol{H}(\boldsymbol{X}_0)$是负定的；当$\boldsymbol{H}(\boldsymbol{X}_0)$为半正定时，称$-\boldsymbol{H}(\boldsymbol{X}_0)$为半负定的。

附带说明一下，对称矩阵为正定的条件是矩阵的主对角线各元素以及其对应的行列式均为正，或者是矩阵的全部特征值为正。

在讨论函数的局部性质及研究算法时，经常需要用到多元函数的一阶近似（线性近似）和二阶近似（平方近似）表达式，这就是取泰勒展开式的一次和二次项来近似得到展开点附近的函数值。即有一阶近似公式为

$$f(\boldsymbol{X})\approx f(\boldsymbol{X}_0)+\nabla f(\boldsymbol{X}_0)^{\mathrm{T}}\Delta X \quad (1-27)$$

二阶近似公式为

$$f(\boldsymbol{X})\approx f(\boldsymbol{X}_0)+\nabla f(\boldsymbol{X}_0)^{\mathrm{T}}\Delta X+\frac{1}{2}\Delta \boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}(\boldsymbol{X}_0)\Delta X \quad (1-28)$$