"十一五"重点规划教材

高等学校自动化系列教材

最优化方法与最优控制

王晓陵 陆 军/主编

重点规划教材

哈尔滨工程大学出版社
Harbin Engineering University Press

最优化方法与最优控制

王晓陵　陆　军　主编

哈尔滨工程大学出版社

内容简介

本书是为研究生课程最优化理论或最优控制系统编写的教材，书中深入浅出地阐述了最优化方法和最优控制系统的基础理论、基本方法，并配有丰富的例题和习题，帮助读者理解书中所阐述的内容。

本书的内容分为两大部分，第一部分包括第 1 章、第 2 章和第 3 章，阐述了最优化方法的一般概念和静态最优化方法（线性规划和非线性规划）的一些基本理论和计算方法；第二部分包括第 4 章至第 7 章，阐述了动态最优化方法的基本内容，包括变分极值问题、最小值原理、线性二次型最优控制系统和动态规划的各种基本算法。

本书各章节注重基本原理和基本概念的阐述，容易理解。本书可作为高等工业院校自动化、测控技术与仪器、电气工程及其自动化等相关专业的高年级本科生或研究生的教材；也可供从事控制工程的科研工作者和工程技术人员自学和参考。

图书在版编目(CIP)数据

最优化方法与最优控制/王晓陵，陆军主编．—哈尔滨：哈尔滨工程大学出版社，2008.4
ISBN 978-7-81133-234-6

Ⅰ.最…　Ⅱ.①王…②陆…　Ⅲ.①最优化②最优控制　Ⅳ.O224　O232

中国版本图书馆 CIP 数据核字(2008)第 038544 号

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　肇东粮食印刷厂
开　　本　787mm×1 092mm　1/16
印　　张　11
字　　数　268 千字
版　　次　2008 年 4 月第 1 版
印　　次　2008 年 4 月第 1 次印刷
定　　价　22.00 元
http://press.hrbeu.edu.cn
E-mail:heupress@hrbeu.edu.cn

前　言

人们在处理日常生活、生产过程、经营管理、社会发展等实际问题时，都希望获得最佳的处理结果。在有多种方案及各种具体措施可供选择时，处理结果与所选取方案和具体措施密切相关。如何获取最佳处理结果的问题称为最优化问题。针对最优化问题，如何选取满足要求的方案和具体措施，使所得结果最佳的方法称为最优化方法。

实际问题中所提出的最优化问题大体分为两类，一类是求函数的极值；另一类是求泛函（函数的自变量是函数）的极值。求函数极值的数值方法或试验最优化方法称为数学规划，包括线性规划和非线性规划。数学规划所处理的问题一般是静态问题，因此求函数极值问题又被称为静态最优化问题。求泛函极值问题需要应用变分法、最小（大）值原理或动态规划来处理，所处理的问题一般是动态问题，这一类问题就称为动态最优化问题，通常称为最优控制问题。

静态最优化和动态最优化问题并无截然的界限，它们都有度量处理结果优劣的目标函数、描述问题的数学模型，处理方法中有分析算法和直接算法。两类问题的直接算法基本相同，只是动态最优化处理的问题要复杂一些。但是，在数学基础上两类问题分属两个不同范畴，静态最优化问题属于运筹学范畴，而动态最优化问题属于变分学范畴，其基础理论框架是不一样的。为便于适应各种不同的实际最优化问题的需要，本书包含有静态最优化和动态最优化的内容，可使读者在阅读本书后，能够获得比较宽广的理论基础和解决问题的基本技能。

本书是作者在总结多年从事最优控制系统教学经验及控制工程实践的基础上，参考了许多优秀教材编写的。本书较侧重基本原理和应用，对于基本定理，避开严格的数学证明，而给出原理上和概念上的简洁阐述，使读者易于理解并能牢固地掌握基本概念和基本理论。

本书由王晓陵教授、陆军副教授主编，史震教授主审。

限于作者水平，书中的内容难免存在错误、遗漏和不妥，希望读者指正。

编　者

2008年2月

目　　录

第 1 章　最优化方法的一般概念 …… 1
1.1　目标函数、约束条件和求解方法 …… 1
1.2　静态最优化问题与动态最优化问题 …… 2
1.3　线性规划和非线性规划问题 …… 2
1.4　最优化方法在控制领域中的应用 …… 3
习　题 …… 4
第 2 章　非线性规划 …… 6
2.1　一元函数的极小化 …… 7
2.2　多元函数无约束的极小化 …… 18
2.3　求解多元函数无约束极值的直接法 …… 29
2.4　多元函数带约束极小化 …… 32
2.5　非线性规划应用举例 …… 40
习　题 …… 41
第 3 章　线性规划 …… 43
3.1　线性规划的数学模型 …… 43
3.2　图解法 …… 45
3.3　线性规划的数学基础 …… 45
3.4　线性规划的单纯形法 …… 48
3.5　线性规划的对偶问题 …… 64
3.6　对偶单纯形法 …… 68
3.7　线性规划应用举例 …… 71
习　题 …… 77
第 4 章　最优控制与变分法 …… 80
4.1　最优控制问题的数学描述 …… 80
4.2　无约束条件的动态最优化问题 …… 84
4.3　带等式约束的动态最优化问题 …… 93
4.4　用哈密顿函数求解最优控制问题 …… 95
习　题 …… 98
第 5 章　最小值原理 …… 102
5.1　最小值原理 …… 102
5.2　快速最优控制 …… 109
5.3　奇异最优控制 …… 116
5.4　一些典型性能指标下的最优控制 …… 120
习　题 …… 125

第 6 章　线性二次型最优控制系统 …… 129
6.1　线性二次型最优控制系统 …… 129
6.2　状态调节问题 …… 130
6.3　$t_f \to \infty$ 时的状态调节问题 …… 135
6.4　能够保证衰减速度的最优控制 …… 138
6.5　在阶跃干扰作用下的状态调节器 …… 140
6.6　输出调节问题 …… 143
6.7　最优跟踪问题 …… 145
习　题 …… 149
第 7 章　动态规划 …… 152
7.1　多级决策过程 …… 152
7.2　最优性原理 …… 154
7.3　离散系统的线性调节问题 …… 157
7.4　动态规划的连续形式 …… 161
7.5　用动态规划求解连续线性二次型最优调节问题 …… 164
7.6　动态规划的应用示例 …… 166
习　题 …… 167
参考文献 …… 170

第1章　最优化方法的一般概念

人们在处理日常生活、生产过程、经营管理、社会发展等实际问题时，都希望获得最佳的处理结果。在有多种方案及各种具体措施可供选择时，处理结果与所选取方案、具体措施密切相关。获取最佳处理结果的问题称为最优化问题。针对最优化问题，如何选取满足要求的方案和具体措施，使所得结果最佳的方法称为最优化方法。

1.1　目标函数、约束条件和求解方法

目标函数就是用数学方法描述处理问题所能够达到结果的函数，该函数的自变量是表示可供选择的方案及具体措施的一些参数或函数，最佳结果表现为目标函数取极值。在处理实际问题时，通常会受到经济效率、物理条件、政策界限等许多方面的限制，这些限制的数学描述称为最优化问题的约束条件。求解方法是获得最佳结果的必要手段，该方法使目标函数取极值，所得结果称为最优解。针对各种类型的最优化问题，找出可靠、快捷的处理方法是最优化方法（理论）的研究范畴。

目标函数、约束条件和求解方法是最优化问题的三个基本要素。无约束条件的最优化问题称为理想最优化问题，所得结果称为理想最优解。下面用三个简单的例子，说明最优化问题的目标函数和约束条件。

例1-1　有一块薄的塑料板，宽为 a，对称地把两边折起，做成槽（如图1-1）。欲使槽的横截面积 S 最大，x_1，x_2 和 θ 的最优值是多少？

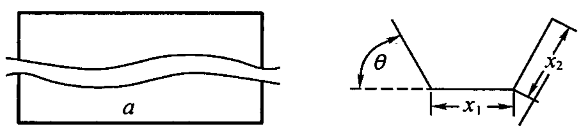

图1-1　横截面积与参数关系图

该问题要找出最优参数 x_1，x_2 和 θ，使槽的横截面积 S 最大，所以，目标函数为

$$\max S=(x_1+x_2\cos\theta)\cdot x_2\sin\theta \tag{1-1}$$

由于底边与两个斜边的总长度应等于塑料板宽度 a，因此约束条件为

$$x_1+2x_2=a \tag{1-2}$$

有许多最优化问题可以方便地将等式约束条件代入目标函数中，使原问题转换为无约束条件的最优化问题，便于求解。例1-1为无约束条件的最优化问题时，目标函数为

$$\max S=(a-2x_2+x_2\cos\theta)\cdot x_2\sin\theta \tag{1-3}$$

例1-2　仓库里存有20 m长的钢管，现场施工需要100根6 m长和80根8 m长的钢管，问最少需要领取多少根20 m长的钢管？

用一根20 m长的钢管，截出8 m管或6 m管的方法只有三种，设 x_1 为一根长管截成两

根 8 m 管的根数；x_2 为一根长管截成一根 8 m 管和两根 6 m 管的根数；x_3 为一根长管截成三根6 m管的根数。该问题的目标函数为

$$\min\ n = x_1 + x_2 + x_3 \tag{1-4}$$

现场施工需要 80 根 8 m 长和 100 根 6 m 长的钢管，即约束条件为

$$\begin{cases} 2x_1 + x_2 \geqslant 80, \\ 2x_2 + 3x_3 \geqslant 100, \end{cases} x_i \geqslant 0,\ i = 1,2,3 \tag{1-5}$$

例 1-3 物体在液体中作直线运动时，它所受到的阻力与运动速度的平方成正比。现假设该物体要在规定的时间$[0, t_f]$内，从起点 $x(0) = 0$ 到达终点 $x(t_f) = S$，且终点速度不受限制。问该物体采用什么运动方式 $x(t)$，它所消耗的能量最少？

消耗的能量等于克服阻力所作的功，为运动速度的平方乘以一个比例常数，由于该常数在求极值过程中不起作用，因此目标函数为

$$\min\ J = \int_0^{t_f} \dot{x}^2(t)\mathrm{d}t \tag{1-6}$$

约束（边界）条件为

$$x(0) = 0, \quad x(t_f) = S \tag{1-7}$$

上述三个简单的最优化问题中，例 1-1 和例 1-2 的目标函数的自变量都是参数，而例 1-3的目标函数的自变量是表示物体运动方式的时间函数。不同类型的最优化问题的合适求解方法是不相同的，有针对性的最优化方法将在后续章节分别讨论。

1.2 静态最优化问题与动态最优化问题

静态是指无时间变量的系统或处于平衡工作状态的动态系统，系统的数学模型是代数方程，而不是微分方程或差分方程。静态最优化问题，就是选择系统的最优参数使目标函数取极值。对于动态系统而言，就是选择最优的平衡（工作）点参数。例 1-1 和例 1-2 都是静态最优化问题的示例。

动态最优化问题的目标函数的自变量中含有动态系统的状态变量，状态变量一般是时间的函数。动态最优化问题习惯上又称为最优控制问题，即选择系统最优的运动轨线，使目标函数取极值。解动态最优化（求泛函的极值）问题通常采用变分法、最大（小）值原理和动态规划等方法。例 1-3 是动态最优化问题的示例。

1.3 线性规划和非线性规划问题

非线性规划和线性规划是静态最优化问题的两个分支。非线性规划问题的范围很宽，针对不同类型的最优化问题都有各自适用的求解方法。静态最优化问题实质上是目标函数求极值的问题。主要求解方法分别在第 2 章和第 3 章介绍。

线性规划问题：该类问题的目标函数和约束条件都是变量的线性函数。例 1-2 就是一个线性规划问题的例子。线性规划问题是最简单的最优化问题，同时也是很重要的、具有普遍实际意义的最优化问题。该类问题的求解方法相对简单，对于特殊类型的线性规划问题

有更简便的求解方法。

非线性规划问题：该类问题的目标函数和约束条件中，含有变量的非线性函数。例1-1就是一个非线性规划问题的例子。非线性规划问题也是生产过程、经营管理、社会发展等实践中常常遇到的实际问题。求解有约束的非线性规划问题，通常要将问题转化为无约束的非线性规划问题求解。求解方法通常有两类：解析法和直接法。

1.4 最优化方法在控制领域中的应用

最优化方法应用范围很广，在不同领域应用时，有不同的相应名称。下面给出一些控制领域中的应用实例。

例1-4 参数估计 系统建模和自适应控制都是现代控制理论的重要分支，这两个分支都使用参数估计，即在已知系统结构条件下，经过对系统输入输出的观测，估计出系统参数的最优值，使数据拟合的残差平方和最小。单输入-单输出系统式(1-8)的参数估计问题的数学描述为

$$y(k)=\boldsymbol{\phi}(k)\boldsymbol{\theta}+\varepsilon(k) \tag{1-8}$$

$$\boldsymbol{\theta}=[-a_1 \quad \cdots \quad -a_n \quad b_1 \quad \cdots \quad b_m \quad c_1 \quad \cdots \quad c_r]^{\mathrm{T}} \tag{1-9}$$

$$\boldsymbol{\phi}(k)=[y(k-1) \quad \cdots \quad y(k-n) \quad u(k-1) \quad \cdots \quad u(k-m) \quad \varepsilon(k-1) \quad \cdots \quad \varepsilon(k-r)] \tag{1-10}$$

式中 $u(k)$，$y(k)$和 $\varepsilon(k)$——系统输入、系统输出和拟合残差；

$\boldsymbol{\theta}$——待估计的系统参数向量；

$\boldsymbol{\phi}(k)$——观测数据向量。

式(1-8)是该问题的约束条件，目标函数为

$$\min J(\hat{\boldsymbol{\theta}})=\frac{1}{2}\sum_{k=1}^{N}\varepsilon^2(k) \tag{1-11}$$

目标函数的自变量，即所要优化的变量是系统参数的估计值 $\hat{\boldsymbol{\theta}}$。

该问题使用的最优化方法称为增广最小二乘法(Extended Least Squares)，具体计算方法参见系统辨识的有关内容。

例1-5 最小方差控制 在随机控制理论中，使被控系统的输出方差最小的控制策略称为最小方差控制。目标函数是使系统输出方差最小；约束条件是系统的数学模型；目标函数的自变量是调节器的参数。例如，线性时不变系统为

$$\begin{aligned}y(k)+a_1y(k-1)+\cdots+a_ny(k-n)&=b_0u(k-d)+b_1(k-d-1)+\cdots+b_mu(k-d-m)\\&\quad+e(k)+c_1e(k-1)+\cdots+c_re(k-r)\end{aligned} \tag{1-12}$$

式中，$d\geqslant1$ 是系统滞后的采样周期数；$e(k)$表示白噪声，且

$$w(k)=e(k)+c_1e(k-1)+\cdots+c_re(k-r)$$

$w(k)$是系统中的随机过程干扰。为便于叙述，采用时间移动算子 q，则式(1-12)改写为

$$A(q^{-1})y(k)=q^{-d}B(q^{-1})u(k)+C(q^{-1})e(k) \tag{1-13}$$

控制误差为
$$\varepsilon(k+d)=y(k+d)-\hat{y}(k+d) \tag{1-14}$$

最小方差控制目标函数为
$$\min J(\hat{\boldsymbol{\theta}})=\frac{1}{2}\sum_{k=1}^{N}\varepsilon^2(k) \tag{1-15}$$

最小方差控制规律为

$$u(k)=-\frac{F(q^{-1})}{B(q^{-1})G(q^{-1})}y(k) \tag{1-16}$$

系统预报误差(控制误差)理论值分别为

$$y(k+d)=G(q^{-1})e(k+d) \tag{1-17}$$

式(1-16)中的多项式 $F(q^{-1})$和 $G(q^{-1})$是下列多项式方程的解

$$A(q^{-1})G(q^{-1})+q^{-d}F(q^{-1})=C(q^{-1})$$

式中,$G(q^{-1})$是 $d-1$ 次的首一多项式,$F(q^{-1})$的次数满足多项式方程需要。由多项式方程可以得到多项式 $F(q^{-1})$和 $G(q^{-1})$的系数唯一解,即得到最小方差控制器的参数。

更多的应用例子将在后续章节里讨论。

习　题

1.1　某工厂生产 a_1,a_2 和 a_3 三种紧俏产品。每生产 1 万件 a_1,a_2,a_3 产品消耗的原料为 b_1,b_2,b_3,如表 1-1 所示。已知该厂现有原料 b_1,b_2 和 b_3 分别为 100 t,200 t 和 500 t。每万件 a_1,a_2 和 a_3 产品分别可获利润 4 万元、3 万元和 2 万元。现有订货合同中 a_1,a_2 和 a_3 产品分别为 10 万件、5 万件和 2 万件。如何安排生产才能获得最大利润? 试给出该问题的目标函数和约束条件(不必求解)。

1.2　某工厂生产的 A,B 和 C 三种产品,都需要经过机加工和组装车间加工。产品加工、组装所需工时和车间每周生产能力如表 1-2 所示。A,B 和 C 三种产品的利润分别为 150 元/只、100 元/只和 80 元/只。问:如何安排生产才能获得最大利润? 试给出该问题的目标函数和约束条件(不必求解)。

表 1-1　题 1.1 产品消耗原料表

		原料		
		b_1	b_2	b_3
产	a_1	5	10	10
	a_2	5	8	5
品	a_3	10	5	0

表 1-2　题 1.2 产品所需工时及车间生产能力表

	产品			每周生产能力
	A	B	C	
机加工	10	6	5	1 000
组装	6	4	3	600
利润/(元/只)	150	100	80	

1.3　已知一线性时不变系统的运动方程、初始状态、允许采用的控制量及终端状态为

$$\ddot{y}(t)=u(t);\quad y(0)=1,\dot{y}(0)=1,\ -1\leqslant u(t)\leqslant 1;\quad y(t_f)=\dot{y}(t_f)=0$$

要求选取最优控制策略,使得系统在最短时间内达到终端。试给出该问题的目标函数和约束条件(不必求解)。

1.4　已知一线性时不变系统的运动方程、初始状态及终端状态为

$$\dot{x}_1(t)=x_2(t),\dot{x}_2(t)=u;\quad x_1(0)=x_2(0)=1;\quad x_1(t)=x_2(5)=0$$

要求选取最优控制策略,使得系统在指定的时间段[0,5]内达到终端,且消耗的控制能量最少。试给出该问题的目标函数和约束条件(不必求解)。

1.5 已知一线性时不变系统的运动方程、初始状态及控制作用为

$$\ddot{y}(t)+2\zeta\dot{y}(t)+y(t)=u(t);\quad y(0)=\dot{y}(0)=0;\quad u(t)=1(t)$$

要求优化系统参数 ζ,使系统状态偏差平方和最小。试给出该问题的目标函数和约束条件(不必求解)。

第2章 非线性规划

在实践中,最优化问题的目标函数和(或)约束条件常常是非线性的,如例1-1和例2-1。这类问题称为非线性规划问题。

求解无约束条件的非线性规划问题通常采用逐步逼近法(俗称试探法),若在求解过程中使用了目标函数的导数,则称之为解析法;若不使用导数而直接利用目标函数进行比较、搜索,则称之为直接法。

对于有约束条件的非线性规划问题,可以将其转化为无约束条件的非线性规划问题后再进行求解。

本章介绍求解非线性规划的一些方法,对解的收敛性不作讨论。

例2-1 求解例1-1,即有一块薄的塑料板,宽为 a,对称地把两边折起,做成槽(如图2-1)。欲使槽的横截面积 S 最大,x_1,x_2 和 θ 的最优值是多少?

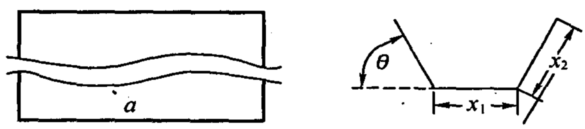

图2-1 横截面积与参数关系图

解 最优化问题的目标函数和约束条件分别为

$$\max\ S = (x_1 + x_2\cos\theta)\cdot x_2\sin\theta$$

$$x_1 + 2x_2 = a,\quad x_1 \geqslant 0,\quad x_2 \geqslant 0$$

分析:该问题的目标函数是非线性的,属非线性规划问题。目标函数的自变量中,只有两个是自由的,不妨以 x_2 和 $y=\cos\theta$ 为自变量,将其转化为无约束条件的非线性规划问题。

$$x_1 = a - 2x_2,\quad 0 < x_2 \leqslant 0.5a$$

目标函数改写为

$$\max\ J = (a - 2x_2 + x_2 y)x_2\sqrt{1-y^2}$$

原最优化问题简化成二元函数求极值问题,求解过程为

$$\frac{\partial J}{\partial x_2} = (y-2)x_2\sqrt{1-y^2} + (a - 2x_2 + x_2 y)\sqrt{1-y^2} = 0$$

$$\frac{\partial J}{\partial y} = x_2^2\sqrt{1-y^2} - \frac{(a-2x_2+x_2y)x_2y}{\sqrt{1-y^2}} = 0$$

$$\begin{cases} a - 4x_2 + 2x_2 y = 0 \\ ay - x_2 - 2x_2 y + 2x_2 y^2 = 0 \end{cases} \Rightarrow \begin{cases} x_2 = a/3 \\ y = 1/2 \end{cases},\quad \begin{cases} x_2 = a/3 \\ \theta = 60° \end{cases}$$

最优解为 $x_1 = a/3, x_2 = a/3, \theta = 60°, \max\ J = \sqrt{3}a^2/12$。

该例实质上是二元函数求极值问题,经解析方法计算使 J 的一阶导数等于零的参数值,且 J 的海森矩阵

$$H = \begin{bmatrix} -3\sqrt{3}/2 & \sqrt{3}a/3 \\ \sqrt{3}a/3 & -2\sqrt{3}a/9 \end{bmatrix}$$

是负定矩阵，因此目标函数 J 取极大值，即为最优解。

2.1 一元函数的极小化

1.极值（极大值或极小值）

若函数 $f(x)$ 在点 x_0 的双侧邻域中有定义，并且对于某邻域 $0<|x-x_0|<\delta$ 内的所有点，不等式

$$f(x)<f(x_0) \quad (\text{或 } f(x)>f(x_0))$$

成立，则称函数 $f(x)$ 在点 x_0 处有极大值（或极小值）。

2.极值存在的必要条件

假定函数 $f(x)$ 在区间 (a,b) 内存在有限导数，若函数在点 $x_0 \in (a,b)$ 处有极值，则必有

$$f'(x_0)=0$$

3.极值存在的充分条件

(1)若函数 $f(x)$ 满足条件：①在 x 的某邻域 $0<|x-x_0|<\delta$ 内有定义并且连续，且在点 x_0 处，$f'(x_0)=0$ 或不存在；②在 $0<|x-x_0|<\delta$ 范围内有有限的 $f'(x)$；③$f'(x)$ 在点 x_0 的两侧有固定的符号，则函数 $f(x)$ 在点 x_0 处的极值情况见表 2-1。

表 2-1 $f(x)$ 在点 x_0 处的极值情况

x	$x<x_0$	x_0	$x>x_0$	$f(x)$
	+		−	极大值
$f'(x)$	−	0 或不存在	+	极小值
	+		+	增函数
	−		−	减函数

(2)若函数 $f(x)$ 有二阶导数 $f''(x)$，并且在点 x_0 处下列条件成立：

$$f'(x_0)=0, \quad f''(x_0)\neq 0$$

则函数 $f(x)$ 在点 x_0 处有极值，当 $f''(x_0)<0$ 时，有极大值；当 $f''(x_0)>0$ 时，有极小值。

(3)设函数 $f(x)$ 在某邻域 $0<|x-x_0|<\delta$ 内有导数 $f'(x),\cdots,f^{(n)}(x)$，且

$$f^{(k)}(x_0)=0, k=1,\cdots,n-1; \quad f^{(n)}(x_0)\neq 0$$

若 n 为偶数，则函数 $f(x)$ 在点 x_0 处有极值，$f^{(n)}(x_0)<0$ 为极大值；$f^{(n)}(x_0)>0$ 为极小值；若 n 为奇数，则函数 $f(x)$ 在点 x_0 处无极值。

例 2-2 已知 $f(x)=x^4-2x^3-12.5x^2-10.5x$，求函数 $f(x)$，$x\in[-3,4]$ 的最大值和最小值。

解 先求出一阶导数等于零的点，即

$$f'(x)=4x^3-6x^2-25x+10.5=0$$

解得 $$x_1=3.5,\ x_2=-0.5,\ x_3=-1.5$$

有 $$f''(x)=12x^2-12x-25;\ f''(x_1)=80,\ f''(x_2)=-16,\ f''(x_3)=20$$

函数 $f(x)$ 在 x_1 处有极小值 $f(x_1)=-125.5625$,在 x_2 处有极大值 $f(x_2)=2.4375$,在 x_3 处有极小值 $f(x_3)=-0.5625$;$f(-3)=54$,$f(4)=-114$。

答案:函数 $f(x)$ 的最大值为 54,最小值为 -125.5625。

例 2-3 求函数 $f(x)$ 的极值点

$$f(x)=x^4-3x^3+3x^2-x+5 \qquad (2-1)$$

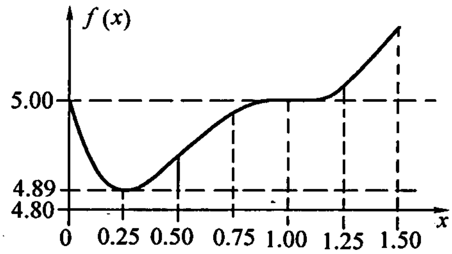

图 2-2 例 2-3 的 $f(x)$ 曲线

解 由已知有

$$f'(x)=4x^3-9x^2+6x-1=0$$

$$x_1=0.25,\ x_{2,3}=1$$

$$f''(0.25)=2.25,\ f''(1)=0;\quad f'''(1)=6$$

答案:函数 $f(x)$ 在极点 $x^*=0.25$ 处有极小值 $f(x^*)=4.89453125$。点 $x=1$ 是 $f(x)$ 的拐点,见图 2-2。

2.1.1 牛顿迭代公式

在求解函数 $f(x)$ 的局部极值时,如例 2-1、例 2-2 和例 2-3,必须解方程

$$g(x)=f'(x)=0 \qquad (2-2)$$

求解非线性方程通常是很困难的。这里介绍著名的牛顿迭代公式(Newton-Raphson)。该方法是一种快速逼近方法,简称牛顿法。

2.1.1.1 一般的牛顿法

假设函数 $g(x)$ 具有连续的一阶导数,且 x^* 是函数 $g(x)$ 的零点,x_0 是零点的初始估计值。函数 $g(x)$ 在其零点估计值 x_k 邻域的一阶泰勒展开式为

$$g(x^*)=g(x_k)+f'(x_k)(x^*-x_k),\ g(x^*)=0 \qquad (2-3)$$

式(2-3)是计算 x^* 的近似表达式。通常,每一次迭代只能得到更接近于真值 x^* 的估计值 x_{k+1}。迭代公式为

$$x_{k+1}=x_k-\alpha g(x_k)/g'(x_k),\ k=0,1,2,\cdots$$

式中,α 是线性搜索因子,用于保证算法收敛或调整收敛速度,常用值为 1。

牛顿法的计算步骤如下:

(1)选取区间 (a,b),使 $g(a)g(b)<0, g'(x)\neq 0$;

(2)选取初始零点估计值 $a<x_0<b$,允许误差 ε;

(3)计算 $g(x_k), g'(x_k)$;

(4)收敛性检查,若 $|g(x_k)|\leqslant\varepsilon$,则 $x^*=x_k$,终止计算;否则继续;

(5)计算 $x_{k+1}=x_k-\alpha g(x_k)/g'(x_k)$;

(6)返回到步骤(3),继续逼近零点。

例 2-4 求式(2-1)所示函数 $f(x)$ 的极值点。

解 记

$$g(x)=f'(x)=4x^3-9x^2+6x-1$$

选取区间[0,0.8], $g(0)=-1$, $g(0.8)=3.8$ 及 $g'(x)=12(x-0.75)^2-0.75$,满足步骤 1 中的条件;

取初始估计值 $x_0=0.4$,取允许误差 $\varepsilon=10^{-5}$。5 次迭代的结果见表 2-2。

表 2-2 例 2-4 牛顿法计算表

k	x_k	$g(x_k)$	$g'(x_k)$	k	x_k	$g(x_k)$	$g'(x_k)$
0	0.400 000 000	0.216 00	0.720 00	3	0.246 739 130	−0.007 40	2.289 26
1	0.100 000 000	−0.486 00	4.320 00	4	0.249 972 010	−0.000 06	2.250 34
2	0.212 500 000	−0.093 02	2.716 88	5	0.249 998 672	-4.0×10^{-9}	$\varepsilon=10^{-5}$

答案:函数 $f(x)$在 $x^*=0.249\ 998\ 672$ 处有极小值 $f(x_5)=4.894\ 531\ 25$。

2.1.1.2 近似牛顿法

如果 $g'(x)$不易算出,那么可用 $g(x)$的差分代替,得到近似牛顿迭代公式

$$x_{k+1}=x_k-g(x_k)\frac{2h}{g(x_k+h)-g(x_k-h)} \tag{2-4}$$

例 2-5 应用近似牛顿迭代公式(2-4),求式(2-1)所示函数 $f(x)$的极值点。

解 初始搜索区间同例 2-4,取 $x_0=0.4$, $\varepsilon=10^{-5}$, $h=10^{-5}$。5 次迭代的结果见表 2-3。

表 2-3 例 2-5 近似牛顿法计算表

k	x_k	$g(x_k)$	$g(x_k+h)$	$g(x_k-h)$
0	0.400 000 000	0.216 000 000	0.216 007 199	0.215 992 799
1	0.100 012 499	−0.485 946 003	−0.485 902 806	−0.485 989 202
2	0.212 504 882	−0.093 010 173	−0.092 983 005	−0.093 037 342
3	0.246 739 357	−0.007 400 376	−0.007 377 484	−0.007 423 269
4	0.249 971 994	−0.000 063 016	−0.000 040 513	−0.000 095 52
5	0.249 999 996	−0.000 000 007	$\varepsilon=10^{-5}$	$h=10^{-5}$

答案:函数 $f(x)$在 $x^*=0.249\ 999\ 996$ 处有极小值, $f(x_5)=4.894\ 531\ 25$。

如果初始极值点的估计值与极值点的值比较接近,则牛顿法能很快地得到满足精度要求的极值点估计值,即算法收敛速度很快,这是牛顿法的优点。但是,牛顿法需要计算函数 $f(x)$的一阶导数 $g(x)$和二阶导数 $g'(x)$,计算量较大。此外,若初始估计值选择不恰当,则会使$\alpha f'(x_0)/f''(x_0)$的绝对值过大,导致算法发散,计算失败。

若式(2-2)有重根,例如

$$g(x)=f'(x)=(x-a)^{2n+1}=0,\ n\geq1 \tag{2-5}$$

易知,函数 $f(x)$在 $x=a$ 处有极大值。但是,牛顿法迭代计算为

$$x_{k+1}=x_k-\alpha(x_k-a)/(2n+1)$$

在 $x=a$ 的邻域，因 x_k-a 近似等于零，所以每次迭代的修正量很小，使收敛速度逐渐降低。

相反，当函数 $f(x)$是自变量 x 的二次函数，即(式(2-5)中，$n=0$)

$$f(x)=ax^2+bx+c$$

时，无论初始估计值如何选取，只需一次迭代，就可得到极值点。理由如下：

$$f'(x_0)=2ax_0+b, f''(x_0)=2a;\quad \bar{x}=x_1=-b/(2a)$$

也就是说，如果函数 $f(x)$近似为自变量 x 的二次函数，那么采用牛顿法收敛速度很快。

2.1.2 求解一元方程的其他方法

应用牛顿法求解函数极值时，需要计算函数 $f(x)$的一阶导数和二阶导数。这里介绍不必计算二阶导数的两种迭代算法。它们的计算步骤有以下共同特点：

(1)选取区间(a_0,b_0)，使 $g(a_0)g(b_0)<0$；确定允许误差 ε；

(2)按照规则，选取新的估计值 x_k；

(3)收敛性检查，若$|g(x_k)|\leqslant\varepsilon$，则 $x^*=x_k$，终止计算；否则继续；

(4)选取新区间(a_{k+1},b_{k+1})，使 $g(a_{k+1})g(b_{k+1})<0$；

(5)返回步骤(2)，继续逼近零点。

2.1.2.1 平分法(二分法)

平分法是最简单的方法，它认为极值点可能在区间的中点，即新的估计值为

$$x_k=(a_k+b_k)/2$$

x_k 是新区间(a_{k+1},b_{k+1})的一个边界，而另一个边界是原区间边界中的一个，必须满足

$$g(a_k)g(x_k)<0 \quad 或 \quad g(x_k)g(b_k)<0$$

例 2-6 应用平分法，求式(2-1)所示函数 $f(x)$的极值点。

解 初始搜索区间与例 2-4 相同，取 $x_0=0.4$，$\varepsilon=10^{-5}$。4 次迭代的结果见表 2-4。

表 2-4 例 2-6 平分法计算表

k	x_k	$g(x_k)$	a_{k+1},b_{k+1}	k	x_k	$g(x_k)$	a_{k+1},b_{k+1}
0	0.400 000 00	0.216 00	0.0,0.4	2	0.300 000 00	0.098 00	0.2,0.3
1	0.200 000 00	-1.440 00	0.2,0.4	3	0.250 000 00	0.000 00	*

答案：函数 $f(x)$在 $x^*=0.25$ 处有极小值 $f(x_3)=4.894\ 531\ 25$。

注：该例及其初始搜索区间，恰好使二分法的迭代次数最少。不能用本例证明二分法在收敛速度方面优于其他迭代算法。

2.1.2.2 弦截法(线性插值法)

弦截法认为极值点近似为满足端点条件的线性函数的零点，即新的估计值为

$$x_k=b_k-\frac{(b_k-a_k)g(b_k)}{g(b_k)-g(a_k)}$$

选取新区间的方法与平分法相同。

例 2-7 应用弦截法，求式(2-1)所示函数 $f(x)$的极值点。

解 初始搜索区间同例 2-4，取 $\varepsilon=10^{-5}$。17 次迭代的结果见表 2-5。

表 2-5 例 2-7 弦截法计算表

k	x_k	$g(x_k)$	a_{k+1},b_{k+1}	k	x_k	$g(x_k)$	a_{k+1},b_{k+1}
0	0.735 294 117	0.136 016 690	a_0,x_k	9	0.252 036 939	0.004 558 252	a_0,x_k
1	0.647 256 438	0.197 719 328	a_0,x_k	10	0.250 893 304	0.002 005 149	a_0,x_k
2	0.540 407 441	0.245 365 619	a_0,x_k	11	0.250 391 232	0.000 879 354	a_0,x_k
3	0.433 934 768	0.235 752 758	a_0,x_k	12	0.250 171 243	0.000 385 121	a_0,x_k
4	0.351 150 151	0.170 339 333	a_0,x_k	13	0.250 074 934	0.000 168 567	a_0,x_k
5	0.300 041 314	0.098 069 398	a_0,x_k	14	0.250 032 786	0.000 073 763	a_0,x_k
6	0.273 244 400	0.049 108 323	a_0,x_k	15	0.250 014 344	0.000 032 274	a_0,x_k
7	0.260 453 943	0.022 870 233	a_0,x_k	16	0.250 006 275	0.000 014 120	a_0,x_k
8	0.254 630 485	0.010 290 340	a_0,x_k	17	0.250 002 745	0.000 006 177	*

答案：函数 $f(x)$ 在 $x^*=0.250\,002\,745$ 处有极小值 $f(x_{17})=4.894\,531\,25$。

在函数 $f(x)$ 的导数近似为线性函数时，适合采用弦截法。显然，该例不是应用弦截法的恰当例子，因为其迭代次数较多。

2.1.3 求解单变量函数极值问题的直接法

上面介绍的各种方法都需要计算函数 $f(x)$ 的一阶导数 $g(x)$。有时，函数 $f(x)$ 很复杂或没有解析表达式，就不易或不能求导。下面介绍三种适用于单极值（峰或谷）函数求极值的直接方法，它们是通过直接比较函数值的大小来确定函数极值的。

2.1.3.1 0.618 法（黄金分割法）

假定函数 $f(x)$ 在区间 $[a_0,b_0]$ 上有极大值（或极小值）。为求取极值点，需要不断地缩小搜索区间。为此，要在区间 $[a_k,b_k]$ 上找出两点 x_{ka} 和 x_{kb}（$x_{ka}<x_{kb}$），比较 $f(x_{ka})$ 和 $f(x_{kb})$ 值的大小，单边缩小区间 $[a_k,b_k]$ 得到新搜索区间 $[a_{k+1},b_{k+1}]$。x_{ka} 和 x_{kb} 中有一个是新搜索区间的边界，而另一个则是新区间中的一点。只需再找出一点，就可以继续缩小搜索区间。

为了计算方便，每次更新的区间都应该是上次区间的 β 倍（$0<\beta<1$）。这就要求 x_{ka} 和 x_{kb} 在区间 $[a_k,b_k]$ 中的位置对称，即

$$x_{kb}-a_k=\beta(b_k-a_k) \quad 且 \quad b_k-x_{ka}=\beta(b_k-a_k) \tag{2-6}$$

因为，在 $k+1$ 次迭代时，x_{ka} 和 x_{kb} 及其函数值仍然有用。在区间 $[a_{k+1},b_{k+1}]$ 中的两点也要位置对称。例如，取新区间为 $[a_{k+1},b_{k+1}]=[x_{ka},b_k]$，要保证对称，则

$$b_k-x_{kb}=\beta(b_k-x_{ka}) \tag{2-7}$$

将式(2-6)代入式(2-7)中，得到

$$\beta^2+\beta-1=0$$

解得

$$\beta=(\sqrt{5}-1)/2\approx 0.618$$

该迭代算法使用的系数值等于 0.618，算法因此命名为 0.618 法。由于 x_{ka} 和 x_{kb} 在区间 $[a_k,b_k]$ 中的位置对称，因此 x_{ka} 和 x_{kb} 满足关系