

最优化方法

ZUIYOUHUA FANGFA

编著 郝燕玲 徐耀群

罗跃生 沈继红

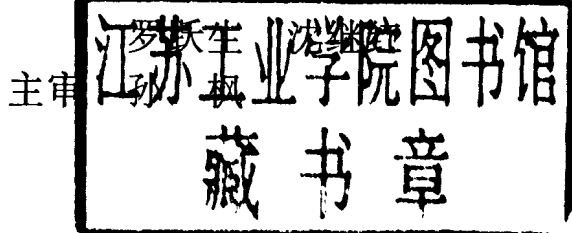
主审 孙 枫

哈尔滨工程大学出版社

最优化方法

编著 郝燕玲 徐耀群

主审



哈尔滨工程大学出版社

内 容 简 介

本书系统地介绍了最优化理论及其应用,内容包括无约束优化、线性规划、整数规划、目标规划、动态规划、非线性规划、遗传算法、神经网络优化、混沌优化。本书在编写过程中力求理论严谨,叙述概念清晰、凝练,并配有实例介绍最优化方法的应用。书中各章配有一定量习题,以供读者理解、巩固知识,锻炼解决问题的能力。

本书可供高等院校工程及管理类专业的硕士及博士研究生作教材使用,也可作为工程技术人员及各类管理人员学习最优化理论的参考读物。

最优化方法

编著 郝燕玲 徐耀群

罗跃生 沈继红

责任编辑 程小东

*

哈尔滨工程大学出版社出版发行

哈尔滨市南通大街145号哈工程大学11号楼

发行部电话(0451) 2519328 邮编:150001

新华书店经销

哈尔滨工业大学印刷厂印刷

*

开本 850mm×1168mm 1/32 印张 8.5 字数 220 千字

2001年3月第1版 2001年3月第1次印刷

印数:1~1 000 册

ISBN7-81073-127-0

TP·24 定价:11.00 元

前 言

追求最优目标是大自然最普遍的运行规则。最优化方法是在最优目标的前提下,从众多可能方案中选择最佳者的科学。最优化理论是一门比较新的应用数学分支,在工程技术、自然科学及经济管理等诸多领域中有其广泛的应用。在有限的资源条件下如何使产出最多(或利润最大),在市场需求一定的情况下如何节省人力、物力、财力,在满足系统设计指标的条件下如何使系统结构最简单,节省时间和能量,这些实际问题促使最优化理论产生、发展,其应用已广泛渗透到社会的各个领域。可以断言,随着信息社会的进一步发展,最优化理论和技术必将在社会的诸多方面发挥越来越大的作用。

本书是根据我校控制类博士生最优化理论课程的教学讲义整理而成的。考虑到工程院校的特点,本书适当增加了优化理论在工程技术应用方面的比重。全书共分九章,系统地介绍了最优化理论的基本内容。考虑到博士生的层次较高,本书在概念叙述上省去了许多详细的说明,在举例时也尽量做到凝练简洁,没有给出过多的冗繁例子;对于一些证明较冗长和复杂的定理,只给出了定理叙述,证明从略。总体上讲,本书在内容编排上尽量做到系统、全面,重点突出,选材针对性强,具有较好的可读性,既可使本科阶段学习过最优化课程的同学进一步学习,同时又可作为研究生的学习教材。

最优化理论和方法的内容极其广博。本书力图涵盖最优化理论的最新内容,将近年来出现的遗传算法、神经网络优化及混沌优化在最优化理论的框架中加以介绍,使读者可以了解最优化理论发展的现状。书中很多内容现在正处于研究发展阶段,也可为感

兴趣的读者做入门介绍,以便使更多的科研、工程技术及管理人员参与到最优化理论的研究与开发应用上来。

本书由郝燕玲教授主编,孙枫教授主审,沈继红、罗跃生、徐耀群三位同志参加了编写工作。在编写过程中得到了哈尔滨工程大学部分教师的帮助,在从讲义整理成教材的过程中,部分博士生也提出了一些宝贵建议,在此表示感谢。

由于编写者水平有限,编写时间匆忙,书中不妥之处,敬请广大读者批评指正。

编 者

2001年1月

目 录

1	无约束规划	1
1.1	基本概念.....	1
1.2	一维搜索.....	5
1.3	无约束极值问题的解法.....	8
	习题 1	30
2	线性规划	32
2.1	线性规划问题的数学模型.....	32
2.2	二维线性规划问题的图解法.....	36
2.3	单纯形法.....	38
2.4	初始基可行解的获得.....	52
2.5	改进单纯形法.....	56
2.6	线性规划的对偶问题.....	60
2.7	系数对解的影响和含参数线性规划.....	67
	习题 2	70
3	整数规划	73
3.1	整数规划的定义.....	73
3.2	割平面法.....	75
3.3	分枝定界法.....	78
3.4	分配问题与匈牙利法.....	80
3.5	解 0-1 规划问题的隐枚举法	86
	习题 3	90
4	目标规划	93
4.1	多目标线性规划.....	93
4.2	目标规划问题	101

4.3	目标规划的求解方法	105
	习题 4	118
5	动态规划	120
5.1	多阶段的决策问题	120
5.2	动态规划的基本方程	128
5.3	阶段数不定的动态规划	133
5.4	随机性动态规划	142
	习题 5	147
6	非线性规划	152
6.1	最优性条件	153
6.2	二次规划	158
6.3	可行方向法	162
6.4	制约函数法	164
	习题 6	172
7	遗传算法	175
7.1	遗传算法的由来及优点	175
7.2	遗传算法的计算步骤	176
7.3	数值例子	178
7.4	0、1 二进制编码的遗传算法	181
7.5	可行解本身作编码的遗传算法	185
7.6	遗传算法的数学理论	188
	习题 7	199
8	神经网络优化	200
8.1	感知器及其应用	201
8.2	Hopfield 神经网络	205
8.3	模拟退火算法	211
8.4	LSSM 系统	215
8.5	Karmarkar 算法的神经网络实现	221
8.6	二次离散神经优化计算	223

8.7 凸性神经优化网络设计	226
习题 8	231
9 混沌优化方法	235
9.1 符号动力系统	236
9.2 基本定义	240
9.3 测度熵和拓扑熵	241
9.4 混沌	243
9.5 李雅普诺夫指数	244
9.6 轨道点的密度分布	246
9.7 Logistic 映射	250
9.8 混沌优化方法	252
9.9 混沌神经网络	253
习题 9	261
参考文献	262

1 无约束规划

无约束规划是指没有任何约束条件的优化问题,这是优化问题的一种理想状态。一般来讲,经济及工程技术领域中的优化问题往往有控制品质、资源、资金、利润等一些限制。但是,当变量的限制不是很强(如自变量的上限很大,可按无穷大处理)时,也可把有约束规划近似看成无约束规划问题处理。无约束规划问题广泛应用于各类工程技术问题中。

设 $f(X)$ 为定义在 n 维欧氏空间 R^n 上的 n 元实函数,我们寻找 $f(X)$ 的极小点的问题称为一个无约束最优化问题。这个问题可表示为

$$\min f(X), \quad X = (x_1, x_2, \dots, x_n)^T \in R^n \quad (1.1)$$

其中 $f(X)$ 称为目标函数。

在本章中,第一节给出了一些基本概念;第二节给出两种一维搜索的方法;第三节讲述了三种求解无约束极值问题的方法:梯度法、共轭梯度法、变尺度法。

1.1 基本概念

函数的极值分为局部极值和全局极值,下面给出它们的定义。

定义 1 (1) 设 $f(X)$ 为定义在 n 维欧氏空间 R^n 的某一区域 Ω 上的 n 元实函数,其中 $X = (x_1, x_2, \dots, x_n)^T$ 。对于 $X^* \in R^n$, 如果存在某个 $\epsilon > 0$, 使所有与 X^* 的距离小于 ϵ 的 $X \in \Omega$ (即 $X \in \Omega$ 且 $\|X - X^*\| < \epsilon$) 均满足不等式 $f(X) \geq f(X^*)$, 则称 X^* 为 $f(x)$ 在 R^n 上的局部极小点, $f(X^*)$ 为局部极小值。若对所有 $X \neq X^*$ 且与 X^* 的距离小于 ϵ 的 $X \in \Omega$, 有 $f(X) >$

$f(X^*)$, 则称 X^* 为 $f(X)$ 在 Ω 上的严格局部极小点, $f(X^*)$ 为严格局部极小值。

(2) 对于点 $X^* \in \Omega$, 若对所有 $X \in \Omega$ 都有 $f(X) \geq f(X^*)$, 则称 X^* 为 $f(X)$ 在 Ω 上的全局极小点, $f(X^*)$ 为全局极小值。若对于所有 $X \in \Omega$ 且 $X \neq X^*$, 都有 $f(X) > f(X^*)$, 则称 X^* 为 $f(X)$ 在 Ω 上的严格全局极小点, $f(X^*)$ 为严格全局极小值。

如将上述不等式反向, 即可得到相应的极大点和极大值的定义。

下面说明极值点存在的必要条件和充分条件。

定理 1(必要条件) 设 Ω 是 n 维欧氏空间 R^n 上的某一开集, $f(X)$ 在 Ω 上有一阶连续偏导数, 且在点 $X^* \in \Omega$ 处取得局部极值, 则必有

$$\frac{\partial f(X^*)}{\partial x_1} = \frac{\partial f(X^*)}{\partial x_2} = \dots = \frac{\partial f(X^*)}{\partial x_n} = 0 \quad (1.2)$$

或

$$\nabla f(X^*) = 0 \quad (1.3)$$

上式中

$$\nabla f(X^*) = \left(\frac{\partial f(X^*)}{\partial x_1}, \frac{\partial f(X^*)}{\partial x_2}, \dots, \frac{\partial f(X^*)}{\partial x_n} \right)^T \quad (1.4)$$

为函数 $f(X)$ 在点 X^* 处的梯度。

由数学分析知道, $\nabla f(X)$ 的方向为 $f(X)$ 的等值面的法线(在点 X 处)方向, 沿这个方向函数值增加最快。

满足式(1.2)或式(1.3)的点称为平稳点或驻点。在区域内部, 极值点必为平稳点, 但平稳点不一定是极值点。

定理 2(充分条件) 设 Ω 是 n 维欧氏空间 R^n 上的某一开集, $f(X)$ 在 Ω 上具有二阶连续偏导数, $X^* \in \Omega$, 若 $\nabla f(X^*) = 0$, 且对任何非零向量 $Z \in R^n$ 有

$$Z^T H(X^*) Z > 0 \quad (1.5)$$

则 X^* 为 $f(X)$ 的严格局部极小点。此处 $H(X^*)$ 为 $f(X)$ 在点

X^* 处的海赛(Hesse) 矩阵

$$H(X^*) = \begin{bmatrix} \frac{\partial^2 f(X^*)}{\partial x_1^2} & \frac{\partial^2 f(X^*)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(X^*)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(X^*)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(X^*)}{\partial x_2^2} & \dots & \frac{\partial^2 f(X^*)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(X^*)}{\partial x_n \partial x_1} & \frac{\partial^2 f(X^*)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(X^*)}{\partial x_n^2} \end{bmatrix} \quad (1.6)$$

为了求解某可微函数的最优解,可令该函数的梯度等于零,由此求得平稳点;然后用充分条件进行判别,求出所要的解。对某些较简单的函数,这样做有时是可行的;但对一般 n 元函数来说,由条件 $\nabla f(X) = 0$ 得到的常是一个非线性方程组,解它相当困难。对于不可微函数,当然谈不上使用这样的方法。为此,常直接使用迭代法。

迭代法的基本思想是:为了求函数 $f(X)$ 的最优解,首先给定一个初始估计 $X^{(0)}$,然后按某种规则(即算法)找出比 $X^{(0)}$ 更好的解 $X^{(1)}$,再按此规则找出比 $X^{(1)}$ 更好的解 $X^{(2)}$, $\dots$,如此即可得到一个解的序列 $\{X^{(k)}\}$ 。若这个序列有极限 X^* ,即

$$\lim_{k \rightarrow \infty} \|X^{(k)} - X^*\| = 0 \quad (1.7)$$

则称它收敛于 X^* 。

若这算法是有效的,那么它所产生的解的序列将收敛于该问题的最优解。不过,由于计算机只能进行有限次迭代。一般说很难得到准确解,只能得到近似解。当这一近似解满足所要求的精度时,即可停止迭代。

若由某算法所产生的解的序列 $\{X^{(k)}\}$ 使目标函数值 $f(X^{(k)})$ 逐步减小,则称这种算法为下降算法。

现假定已经迭代到点 $X^{(k)}$,若从 $X^{(k)}$ 出发沿任何方向移动都不能使目标函数值下降,则 $X^{(k)}$ 是一局部极小点,迭代停止。若从 $X^{(k)}$ 出发至少存在一个方向可使目标函数值有所下降,则可选定能使目标函数值下降的某方向 $P^{(k)}$,沿此方向迈进适当的一步,得到下一

个迭代点 $X^{(k+1)}$, 并使 $f(X^{(k+1)}) < f(X^{(k)})$, 这相当于在射线

$$X = X^{(k)} + \lambda P^{(k)} \quad (1.8)$$

上选定新点

$$X^{(k+1)} = X^{(k)} + \lambda_k P^{(k)} \quad (1.9)$$

式中, $P^{(k)}$ 称为搜索方向; λ_k 称为步长或步长因子。

下降迭代算法的步骤可总结如下:

- (1) 选定某一初始点 $X^{(0)}$, 并令 $k = 0$ 。
- (2) 确定搜索方向 $P^{(k)}$ 。
- (3) 从 $X^{(k)}$ 出发, 沿方向 $P^{(k)}$ 求步长 λ_k , 以产生下一个迭代点 $X^{(k+1)}$ 。

(4) 检查得到的新点 $X^{(k+1)}$ 是否为极小点或近似极小点。若是, 则停止迭代。否则, 令 $k = k + 1$, 转回(2) 继续进行迭代。

在以上步骤中, 选取搜索方向 $P^{(k)}$ 是最关键的一步。各种算法的区分, 主要在于确定搜索方向的方法不同。

确定步长 λ_k 可选用不同的方法。最简单的一种方法是令它等于某一常数(例如令 $\lambda_k = 1$), 这样做计算简便, 但不能保证使目标函数值下降。第二种方法称为可接受点算法, 只要能使目标函数值下降, 可任意选取步长 λ_k 。第三种方法是沿搜索方向使目标函数值下降最多, 即沿射线 $X = X^{(k)} + \lambda P^{(k)}$ 求目标函数 $f(X)$ 的极小

$$\lambda_k : \min f(X^{(k)} + \lambda P^{(k)}) \quad (1.10)$$

值。由于这项工作是求以 λ 为变量的一元函数 $f(X^{(k)} + \lambda P^{(k)})$ 的极小点 λ_k , 故常称这一过程为(最优) 一维搜索或线搜索, 这样确定的步长为最佳步长。

一维搜索有个十分重要的性质: 在搜索方向上所得最优点处的梯度和该搜索方向正交。

定理3 设目标函数 $f(X)$ 具有一阶连续偏导数, $X^{(k+1)}$ 按下述规则产生

$$\left. \begin{aligned} \lambda_k : \min f(X^{(k)} + \lambda P^{(k)}) \\ X^{(k+1)} = X^{(k)} + \lambda_k P^{(k)} \end{aligned} \right\} \quad (1.11)$$

则有

$$\nabla f(X^{(k+1)})^T P^{(k)} = 0 \quad (1.12)$$

证明:构造函数 $\phi(\lambda) = f(X^{(k)} + \lambda P^{(k)})$, 则得

$$\begin{cases} \phi(\lambda_k) = \min_{\lambda} \phi(\lambda) \\ X^{(k+1)} = X^{(k)} + \lambda_k P^{(k)} \end{cases}$$

即 λ_k 为 $\phi(\lambda)$ 的极小点。此外

$$\phi'(\lambda) = \nabla f(X^{(k)} + \lambda P^{(k)})^T P$$

由 $\phi'(\lambda) \Big|_{\lambda=\lambda_k} = 0$ 可得

$$\nabla f(X^{(k)} + \lambda_k P^{(k)})^T P^{(k)} = \nabla f(X^{(k+1)})^T P^{(k)} = 0$$

1.2 一维搜索

当用迭代法求函数的极小点时,常常要用到一维搜索,即沿某一已知方向求目标函数的极小点。一维搜索的方法很多,常用的有:(1) 试探法(斐波那契法,0.618法等);(2) 插值法(抛物线插值法,三次插值法等);(3) 微积分中的求根法(牛顿法,二分法等)。

1.2.1 牛顿法

由于一维搜索就是求以 λ 为变量的一元函数的极小点,因此我们的问题就是求解

$$\min f(x), \quad -\infty < x < +\infty \quad (1.13)$$

或记为

$$\min f(x), \quad x \in R \quad (1.14)$$

假定 $f(x)$ 连续可微,则由极值的必要条件,极小点必为稳定点,这时问题可化为求解方程 $f'(x) = 0$,由方程求根的牛顿迭代法,迭代公式为

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})} \quad (1.15)$$

这就是求一维无约束极值的牛顿迭代公式。

牛顿迭代法的计算方法如下：

设已知初始点 x_0 ，并令 $k = 0$ 。

1. 计算 $f'(x_k), f''(x_k)$ 。

2. 求 $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$ 。

3. 若 $|x_{k+1} - x_k| \leq \epsilon$ ，则求出最优解的近似解 $x^* = x_{k+1}$ ，停止计算；否则转 4。

4. 令 $k = k + 1$ ，转 1。

注意：二次函数 $\phi(x)$ 仅当 $f''(x_0) > 0$ 时才有极小点存在。因此为了保证牛顿法收敛到极小点，应要求 $f''(x_k) > 0$ ，至少对充分大的 k 应如此。

牛顿法的几何解释如图 1-1，曲线 $y = f'(x)$ 在 x_k 处的切线与 x 轴的交点作为下一个点 x_{k+1} 。

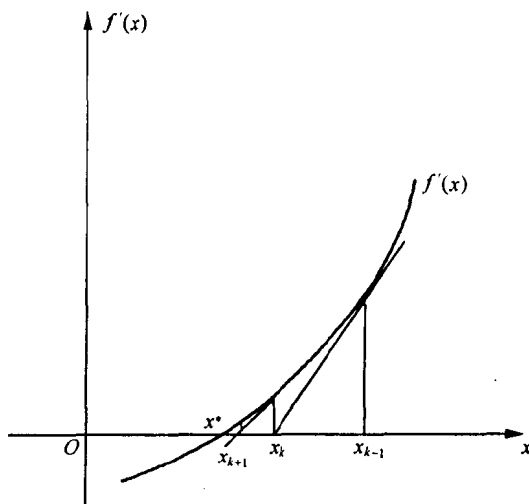


图 1-1

牛顿法的最大优点是收敛速度快。但是,这个方法也有缺点,要在每个迭代点处计算函数的二阶导数值,这样就增加了每次迭代的工作量。牛顿法的另一个缺点是要求初始点选得比较好,也就是说初始点不能离极小值点太远。

1.2.2 平分法

平分法的收敛速度虽然比牛顿法要慢得多,但每次迭代计算量较少,而且总能收敛到一个局部极小值点。所以它仍是一种受欢迎的方法。

我们知道,在极小点 x^* 处 $f'(x^*) = 0$, 并且当 $x < x^*$ 时, 函数是递减的, 即 $f'(x) < 0$; 而当 $x > x^*$ 时, 函数递增, 即 $f'(x) > 0$ 。如果我们找到了一个区间 $[a, b]$, 且 $f'(a) < 0, f'(b) > 0$, 则在 a, b 之间必有 $f(x)$ 的极小点 x^* , 并且 $f'(x^*) = 0$ 。为了找到 x^* , 我们取 $x_0 = \frac{a+b}{2}$, 若 $f'(x_0) > 0$, 则在区间 $[a, x_0]$ 中有极小点, 这时以 $[a, x_0]$ 作为新的区间 $[a, b]$; 若 $f'(x_0) < 0$, 则在 $[x_0, b]$ 中有极小点, 因此以 $[x_0, b]$ 作为新的区间 $[a, b]$ 。继续这个过程, 逐步将区间 $[a, b]$ 换小, 当区间 $[a, b]$ 的长度充分小时, 或者当 $f'(x_0)$ 充分小时, 即可将 $[a, b]$ 的中点取做极小点的近似点。这时有明显的估计 $\left| x^* - \frac{a+b}{2} \right| < \frac{b-a}{2}$ 。

至于区间 $[a, b]$ 的确定, 一般可采用下述方式。首先取一初始点 x_0 , 若 $f'(x_0) < 0$, 则在 x_0 右方取点 $x_1 = x_0 + \Delta x$ (Δx 也是事先给定的一个步长); 若 $f'(x_1) > 0$, 则令 $a = x_0, b = x_1$; 若仍有 $f'(x_1) < 0$, 则取 $x_2 = x_1 + \Delta x$ (或者先将 Δx 增大一倍, 再令 $x_2 = x_1 + \Delta x$), 求 $f'(x_2)$; 若 $f'(x_2) > 0$, 则以 $[x_1, x_2]$ 作为区间 $[a, b]$, 否则继续做下去。对 $f'(x_0) > 0$ 的情况, 也可作类似于 $f'(x_0) < 0$ 的情况去讨论, 这时应在 x_0 左方取点, 取 $\Delta x < 0$ 。

下面给出平分法的计算公式, 首先给出寻找区间 $[a, b]$ 的计

算公式。

给定 $x_0, \Delta x$ 且 $\Delta x > 0$ 。

1. 求 $f'(x_0)$, 若 $f'(x_0) = 0$, 则求出 $x^* = x_0$, 停止整个运算; 若 $f'(x_0) < 0$, 做 2、3; 若 $f'(x_0) > 0$, 转向 4、5。

2. $x_1 = x_0 + \Delta x$ (或再做 $\Delta x = \Delta x + \Delta x$)。

3. 求 $f'(x_1)$, 若 $f'(x_1) = 0$, 则求出 $x^* = x_1$; 若 $f'(x_1) > 0$, 则求出区间 $[a, b] = [x_0, x_1]$; 若 $f'(x_1) < 0$, 则令 $x_0 = x_1$ 转向 2。

4. 令 $x_1 = x_0 - \Delta x$ (或再做 $\Delta x = \Delta x + \Delta x$)。

5. 求 $f'(x_1)$, 若 $f'(x_1) = 0$, 则求出 $x^* = x_1$; 若 $f'(x_1) < 0$, 则求出区间 $[a, b] = [x_1, x_0]$; 若 $f'(x_1) > 0$, 则令 $x_0 = x_1$ 转向 4。

找出了区间 $[a, b]$ 之后, 便可用平分法的计算过程。现在我们把其计算方法整理为:

给定 a, b, ϵ ,

1. 求 $x = \frac{a+b}{2}$, 若 $|b-a| \leq \epsilon$, 则求出稳定点, 即 $x^* = x$; 否则做 2。

2. 求 $f'(x)$, 若 $f'(x) < 0$, 则 $a = x$; 若 $f'(x) > 0$, 则 $b = x$, 这两种情形均转向 1。若 $f'(x) = 0$, 则求出稳定点 $x = x^*$ 。

1.3 无约束极值问题的解法

本节研究无约束极值问题的解法, 这种问题可表述为

$$\min f(X), \quad X \in R^n \quad (1.16)$$

在求解上述问题时常使用迭代法。迭代法可分为两大类: 一类要用到函数的一阶导数和二阶导数, 由于用到了函数的解析性质, 故称为解析法; 另一类在迭代过程中仅用到函数值, 而不要求函数的解

析性质,这类方法称为直接法。一般说来,直接法的收敛速度较慢,因此只是在变量较少时才适用。但是直接法的迭代步骤简单,特别是当目标函数的解析表达式十分复杂,甚至写不出具体表达式时,其导数很难求得,或者根本不存在,这时解析法就无能为力了,而应用直接法则能很好地解决问题。

1.3.1 梯度法(最速下降法)

在求解无约束极值问题的解析法中,梯度法是最为古老但又十分基本的一种求值方法。它的迭代过程简单,使用方便,而且又是理解某些其它最优化方法的基础,不过由于其收敛速度慢,因此人们已经不太使用它了。

1. 梯度法基本原理

假定无约束极值问题中的目标函数 $f(X)$ 有一阶连续偏导数,具有极小点 X^* 。以 $X^{(k)}$ 表示极小点的第 k 次近似,为了求其第 $k+1$ 次近似点 $X^{(k+1)}$,我们在 $X^{(k)}$ 点沿方向 $P^{(k)}$ 作射线

$$X = X^{(k)} + \lambda P^{(k)} \quad (\lambda \geq 0) \quad (1.17)$$

现将 $f(X)$ 在 $X^{(k)}$ 点处展成泰勒级数

$$\begin{aligned} f(X) &= f(X^{(k)} + \lambda P^{(k)}) = \\ &= f(X^{(k)}) + \lambda \nabla f(X^{(k)})^T P^{(k)} + o(\lambda) \end{aligned}$$

其中

$$\lim_{\lambda \rightarrow 0} \frac{o(\lambda)}{\lambda} = 0$$

对于充分小的 λ ,只要

$$\nabla f(X^{(k)})^T P^{(k)} < 0$$

即可保证 $f(X^{(k)} + \lambda P^{(k)}) < f(X^{(k)})$ 。这时若取

$$X^{(k+1)} = X^{(k)} + \lambda P^{(k)}$$

就能使目标函数值得到改善。

现考察不同的方向 $P^{(k)}$ 。假定 $P^{(k)}$ 的模一定且不为零,并设 $\nabla f(X^{(k)}) \neq 0$ (否则, $X^{(k)}$ 是平稳点),则使式(1.17)成立的 $P^{(k)}$