21 数学天元基金

现代应用数学方法丛书 ①

数值优化中的二次逼近法

赵凤治 著

科学出版社
Science Press

现代应用数学方法丛书 1

数值优化中的二次逼近法

赵凤治 著

科学出版社

2000

内 容 简 介

数值优化在实践中有着广泛的应用.本书侧重于介绍作者近年来在大规模数值优化方法研究中取得的若干结果，着重介绍方法的思想及计算方案.主要内容有无约束极值的解法和有约束极值的解法.最后一章介绍了一些应用实例.

本书可供从事工程设计、管理科学、军事运筹等工作的人员阅读.

现代应用数学方法丛书 1

数值优化中的二次逼近法

赵风治 著

责任编辑 毕 颖

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

新蕾印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1994年4月第 一 版 开本:850×1168 1/32

2000年4月第二次印刷 印张:4 3/4

印数:1 501—3 500 字数:119 000

ISBN 7-03-003931-9/O·685

定 价:12.00元

（如有印装质量问题，我社负责调换〈新欣〉）

总　序

应用数学的发展与自然科学和社会科学广泛的交叉和渗透密切相关．一方面，它为形形色色的物理、化学、生物、社会等现象提供描述和分析的数学工具．另一方面，这些实际问题的解决又为数学学科的发展提供了动力和永不枯竭的源泉．许多成功的应用数学方法，如解非线性方程的牛顿-高斯法、曲线拟合的最小二乘法、线性规划的单纯形法等，成了当今应用数学工作者手中不可缺少的工具．它们之所以有如此强大的生命力，原因在于方法本身有坚实的理论基础，同时又有鲜明的应用背景，能用于不同的领域．因此，成功的应用数学方法是理论联系实际的桥梁和纽带．

我国的数学要达到世界先进水平，要对人类有所贡献，一个重要的方面是要有一批独创的应用数学方法．《现代应用数学方法》丛书的出版，希望能为鼓励和促进我国的数学工作者创造或介绍更多的现代应用数学方法增加一个舞台．

这套丛书的宗旨是介绍现代应用数学方法．这些方法应该是目前世界上最先进的，或者是我国独创的，或者是国外已经普遍使用但国内知之甚少的方法．丛书着重阐明所介绍方法的应用背景和思想，避开深奥的数学论证，力求深入浅出、图文并茂，有数值及应用性的例子，使读者易于理解和使用．丛书要求短小精炼，突出新的方法，不求齐全，一般每种篇幅在 10 万字左右．书末所附的文献指出方法的理论背景以及最近的进展，以便读者进一步深入研究．

本丛书的出版得到国家“天元”项目的资助，得到科学出版社的大力支持，是全体编委努力的成果．我们要特别感谢许多作者在百忙中为丛书撰写文稿，付出了辛勤的劳动．我们希望这套丛

书的出版对我国应用数学的发展起到促进作用，衷心地希望丛书成为广大读者的良师益友.

胡国定 （南开大学）

方开泰 （中国科学院应用数学所）

序

数值优化已在我国国民经济建设各领域中得到了广泛的应用,而且其应用效果都是很好的. 但是,也应该看到数值优化在我国四化建设中的作用并没有得到充分发挥,其潜力仍是巨大的.挖掘这一潜力的一支巨大力量存在于工程设计人员及其他各个领域的技术人员之中. 事实上,他们之中有很多人,不但有渊博的专业知识和丰富的实践经验,能应用数值优化技术于实际,而且有很好的数学造诣,对于数值优化方法的最新进展也颇为熟悉. 但是,也有一些工程技术人员及数值优化工作者对于数值优化方法不很清楚,他们迫切希望了解这方面知识.

作者基于这种估计不惭孤陋寡闻把近年来应用中认为有效的方法以及本人和学生的点滴工作奉献出来, 希望它们对读者能有一点助益,更希望得到指教.

这一小册子以介绍方法为主. 为了帮助读者解决在实用中可能会碰上的问题,以及实用中对方法进行改造,也用了一些篇幅介绍方法的基本思想.

在最后一章介绍了一些实例,除了本人所接触的题目外,也摘引了陈开周教授[4]、刘夏石教授[10]、丘昌涛教授、王德人教授、段虞荣教授、张可村教授[2]等许多专家卓越的工作. 对于他们的支持作者十分感谢!

作者非常感谢王长钰教授、赖炎连教授对本书提出的很好的修改意见及付出的劳动. 并感谢方开泰教授、程侃教授的鼓励与支持.

作　者

1991.12

《现代应用数学方法丛书》编委会

目 录

第一章　绪　　论

§1.1　最优化数值方法的功用

目的性是人类活动的显著特点，尽管有的人很明确，有的人不够明确；有的人很自觉，有的人不够自觉；有的人直言不讳，有的人闪烁其词，但是毫无目的的人类活动是没有的．万千年来人类的历史就是这样过来了，直到近一个世纪以来，人们才逐渐认识到用科学的方法来认识这一事实，研究这一事实．伴随着这些研究，也形成了一套数学手段——最优化．

人们用数学语言把某一活动的目的描述出来，如生产活动中如何获得最大产值，军事活动中如何最多地杀伤敌人及最有效地保存自己，等等．这一数学形式被称为目标函数．自然，客观上一定会有一些因素使我们为获得最佳效果的行为受到各种限制．人们用数学语言把这些限制刻画出来，称为约束条件．至此便形成了数学上表示为

$$\max f(\boldsymbol{x}) \tag{1}$$

满足于约束条件

$$\begin{aligned} &g_i(\boldsymbol{x}) \geqslant 0,\ \ i = 1,2,\cdots,m_1, \\ &h_j(\boldsymbol{x}) = 0,\ \ j = 1,2,\cdots,m_2, \end{aligned} \tag{2}$$

的最优化问题．

这种数学问题是那样的广泛，只要你略加思索，就会发现就在你的身边会有大量的最优化问题存在着．

也许你不愿意接受这种陈述问题的方式．那么你每天总要干点什么事吧！如果你要干点什么事，你只要略加思索，你一定会面

临干什么，怎么干等问题．把这些问题用数学的工具科学地描绘出来，就得出了最优化的数学模型．

总之，最优化是实用价值很大的一个数学工具．人们在各自工作的领域中用上最优化这一科学手段后得到了许多好处，社会的、技术的、经济的．于是最优化从实际中产生，经科学家的手又回到实际中，受到了普遍的欢迎．

读者自然明白，最优化的应用建立模型是首要的任务．自然，这种模型可以是显式的数学公式，也可以是无明显的数学公式．前者有时称为有模型的；后者有时称为无模型的．实质上无模型的也有一定的数学关系存在，只不过没有把这种关系的数学式子列出来罢了．

由实际问题抽象出数学模型是一件复杂的工作，它要求人们对实际问题的深入理解，对数学工具的熟悉及对计算数学的了解．从数学描述的形成方式，大致可以分为从客观实际的理论结构出发建立的机理模型；从人们的工作积累出发建立的经验模型；从模拟人工、模拟仪表、模拟实际过程出发建立的仿真模型．也可以把3种手段混合使用建立模型． 至于得到的模型可以是确定性的、随机性的、模糊性的．

模型是实际与理论之间的桥梁，因此建模是一件必须认真对待的工作．无论数学加工及求解，都离不开这项基础工作．并且，其他工作从模型得到的启发往往有很大的价值．

形成数学模型之后必要的数学讨论与数学加工是不能省掉的．如解的性状的讨论，为求解进行的简化，数学模型的有效范围等等．

数学模型的求解自然是极为重要的一环．这也正是写此小册子的目的．对此，我们将在后面展开讨论．姑且假定我们从数学模型中解得了 x 值，不妨称它们为‘策略’．所求策略自然要投入实际中去使用．使用中又会发现新的问题．于是形成反馈到模型中，循环下去．每循环一次，便会使我们的认识前进一步，直到得出满意的策略为止．

目前最优化数值方法有很多分支、各分支又硕果累累．但是客观实际问题更是变化万千的，形成的数学模型也自然是千姿百态五花八门的．面对这一客观现实，最优化数值方法确实又显得不够用了．许多实际问题的求解时时来困扰我们．为了摆脱这些困扰，近几年来我们做了一些研究工作，得出一些成果．这些工作得到了一些专家及同行的鼓励．现在我们把其中一部分介绍给读者，希望使读者在解决实际问题时得到一点助益．

§1.2 数值最优化方法简顾

数值优化是运筹学的一个分支，与其他分支不同，它又是运筹学问题数值求解的一个侧面．因为工程设计、国民经济、管理科学中也要用到数值优化，所以它又以某种意义超出了运筹学的范畴．数值优化包括的分支很多，如线性规划、二次规划、非线性规划、无约束极值、几何规划、动态规划、多目标规划、离散规划、组合规划、互补问题、不动点问题、不可微优化问题等等．因为我们这本小册子只限于讨论无约束极值与非线性规划，所以就把后面的议论也限制在这样一个较小的范围内．但是，应该看到，即使这样一个较小的范围在实用中也有很广泛的天地．甚至可以说，工程优化设计中所用到的方法很多都属于这一范围．

什么是数值优化？数值优化就是求极值！关于极值的工作，几乎与微积分的诞生同时问世，距今约有 300 年的历史了．而 Lagrange 关于条件极值方面的工作也有 200 多年的历史了．但是，人们习惯把这些称之为古典极值问题．

1948 年 John 首先讨论了具有不等式约束的极值问题，人们把这一工作视为最优化方法的开始．事实上，无约束极值的工作也有了新的进展．

最优化数值方法常被分为分析法与直接法两大类．分析法是指计算方案中利用到被极值化函数的一阶导数、二阶导数．这一类方法有最速下降法、牛顿法、共轭梯度法、拟牛顿法、容许方向

法、投影梯度法、简约梯度法、约束变尺度法等等。其中前面4类是求无约束极值的方法，后4类是求约束极值的方法。约束极值中具有奠基性质的工作是属于 Kuhn-Tucker 的。他们把约束极值的讨论转化成鞍点问题，得出了鞍点定理。事实上很多算法求出的点都是鞍点。而判别鞍点的所谓 K-T 条件成为优化工作者人人皆知的了。

众所周知，于 x_k 点附近将 $f(x)$ 做 Taylor 展开有

$$f(x)=f(x_k)+(x-x_k)^T\nabla f(x_k)+\frac{1}{2}(x-x_k)^T H(x_k)\cdot(x-x_k)+o(\|x-x_k\|^2). \tag{1.1}$$

其中

$$\nabla f(x_k)=\left(\frac{\partial f(x_k)}{\partial x_1},\ \frac{\partial f(x_k)}{\partial x_2},\cdots\ \frac{\partial f(x_k)}{\partial x_n}\right)^T \tag{1.2}$$

称为 $f(x)$ 于 x_k 点的梯度；

$$H(x_k)=\begin{pmatrix} \frac{\partial^2 f(x_k)}{\partial x_1^2} & \frac{\partial^2 f(x_k)}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 f(x_k)}{\partial x_1\partial x_n} \\ \frac{\partial^2 f(x_k)}{\partial x_2\partial x_1} & \frac{\partial^2 f(x_k)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x_k)}{\partial x_2\partial x_n} \\ \cdots\cdots\cdots & & & \\ \frac{\partial^2 f(x_k)}{\partial x_n\partial x_1} & \frac{\partial^2 f(x_k)}{\partial x_n\partial x_2} & \cdots & \frac{\partial^2 f(x_k)}{\partial x_n^2} \end{pmatrix} \tag{1.3}$$

称为 $f(x)$ 于 x_k 点的 Hesse 矩阵(也可简称为 Hessian)。$H(x_k)$ 一般总是对称的。

如果我们记

$$x-x_k=\lambda p_k,$$
$$x_{k+1}-x_k=\lambda_k p_k,$$

则有

$$f(x)=f(x_k)+\lambda p_k^T\nabla f(x_k)+\frac{1}{2}\lambda^2 p_k^T H(x_k)p_k+o(\lambda^2\|p_k\|^2).$$

由此可以看出，只要取 p_k 满足

$$p_k^T \nabla f(x_k) < 0, \tag{1.4}$$

λ_k 足够小，令

$$x_{k+1} = x_k + \lambda_k p_k, \tag{1.5}$$

就一定有

$$f(x_{k+1}) < f(x_k).$$

用公式（1.5）由 x_k 得出 x_{k+1} 建立的计算方案称为迭代法。p_k 称为搜索方向，λ_k 称为搜索步长；也称为迭代方向，迭代步长。

满足不同性质的迭代方向建立的方法有不同的特色，往往缀上不同的名字。但满足条件（1.4）的 p_k 所建立的方法统称为下降法（下山法）。

特别取

$$p_k = -\nabla f(x_k), \tag{1.6}$$

则有

$$p_k^T \nabla f(x_k) = -\|\nabla f(x_k)\|^2,$$

可知只要 $\|\nabla f(x_k)\| \neq 0$，用（1.6）定出的 p_k 构造的迭代法是下降算法。人们称这一类方法为梯度法。

满足

$$\|\nabla f(x_k)\| = 0 \tag{1.7}$$

的点称为稳定点。不难证明如果 x^* 是 $f(x)$ 的极小点就一定有

$$\|\nabla f(x^*)\| = 0.$$

有时可用 $f(x)$ 极小点的必要条件来求极小点。为了求出满足 $\|\nabla f(x^0)\| = 0$ 的 x^0，由（1.1）式得出

$$x^0 = x_k - H^{-1}(x_k)\nabla f(x_k).$$

其中 $H^{-1}(x_k)$ 为 Hesse 矩阵的逆矩阵。取迭代方向为

$$p_k = -H^{-1}(x_k)\nabla f(x_k).$$

所构成的算法称为 Newton 法。

因为 Newton 法中用到 $H^{-1}(x_k)$，自然 $H(x_k)$ 还必须具备一定性质，这就使 Newton 法的使用受到限制。为了克服 Newton 法的局限性，人们又做了许多工作，其中最著名的是取

$$p_k = -H_k \nabla f(x_k).$$

式中 H_k 利用公式

$$H_{k+1} = H_k + \Delta H_k$$

进行迭代。这便是著名的变尺度方法．变尺度方法中最著名的是取

$$\Delta H_k = \frac{\Delta x_k \Delta x_k^T}{\Delta x_k^T y_k} - \frac{H_k y_k y_k^T H_k^T}{y_k^T H_k y_k},$$

称为 DFP 方法;取

$$\Delta H_k = \left(1 + \frac{y_k^T H_k y_k}{y_k^T \Delta x_k}\right) \frac{\Delta x_k \Delta x_k^T}{y_k^T \Delta x_k} - \frac{y_k \Delta x_k^T H_k^T}{y_k^T \Delta x_k} - \frac{H_k \Delta x_k y_k^T}{y_k^T \Delta x_k},$$

称为 BFGS 方法．所有上式中

$$\Delta x_k = x_{k+1} - x_k,$$
$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

虽然,无约束极值中还有一些别的方法,但变尺度法、或其中的拟 Newton 法是最有效的一种解法．

应该看出,变尺度方法有如下 3 个方面还应改进．

i）它们没有脱离开 $f(x)$ 的局部展开性质，这与求极值这一具有某种意义的整体性质有一定思维上的抵触;

ii）计算中必须存贮矩阵 H_k，因而存贮量较大;

iii）由 x_k 到 x_{k+1} 的迭代过程，有了 p_k 之后还要用到一维搜索,调用 $f(x)$ 的次数较多．

我们正是基于上面看法，在无约束极值算法方便做了一些工作,这是本书的一部分内容．

目前约束极值计算方法中尽管有逐步线性化、罚函数、容许方向法、投影梯度法、简约梯度法等等许多类，但人们认为序列二次规划方法是目前较有效的一种方法．假定所讨论的问题是

$$\min f(x)$$

满足于约束条件

$$g_i(\boldsymbol{x})=0,\ i=1,2,\cdots,m.$$

于是在一点 $\boldsymbol{x}_k$ 我们可以把目标函数做二次展开得

$$f(\boldsymbol{x})\doteq f(\boldsymbol{x}_k)+(\boldsymbol{x}-\boldsymbol{x}_k)^T\nabla f(\boldsymbol{x}_k)+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_k)^T H(\boldsymbol{x}_k)(\boldsymbol{x}-\boldsymbol{x}_k),$$

把约束条件做线性展开得

$$g_i(\boldsymbol{x})\doteq g_i(\boldsymbol{x}_k)+(\boldsymbol{x}-\boldsymbol{x}_k)^T\nabla g_i(\boldsymbol{x}_k),\ i=1,2,\cdots,m.$$

如果用此二次规划解得 $\boldsymbol{d}^*=\boldsymbol{x}^*-\boldsymbol{x}_k$，再用公式 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\boldsymbol{d}^*$ 求出新点. 容易看出，这有许多缺点.

所以 Wilson, Han, Powell 等人建立了其主要思想如下的约束变尺度法.[36-33]

如果有一个迭代点 $\boldsymbol{x}_k$，我们便可求出

$$\boldsymbol{g}_k=\nabla f(\boldsymbol{x}_k),$$

用变尺度法的公式，可以用迭代的方式生成 H_k. 于是形成二次规划

$$\min\ Q(\boldsymbol{d})=\boldsymbol{g}_k^T\boldsymbol{d}+\frac{1}{2}\boldsymbol{d}^T H_k\boldsymbol{d}$$

满足于约束条件

$$\boldsymbol{d}^T\nabla g_i(\boldsymbol{x}_k)=-g_i(\boldsymbol{x}_k),\ i=1,2,\cdots,m.$$

求出最优解后，不用

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\boldsymbol{d}_k$$

求 $\boldsymbol{x}_{k+1}$，而是从 $\boldsymbol{x}_k$ 出发沿 $\boldsymbol{d}_k$ 方向求目标函数及约束条件两种因素都考虑在内的一个效应函数

$$\varphi(\lambda)=f(\boldsymbol{x}_k+\lambda\boldsymbol{d}_k)+\sum_{i=1}^{m}M_i g_i(\boldsymbol{x}_k+\lambda\boldsymbol{d}_k)+\frac{1}{2}\sum_{i=1}^{m}\sigma_i g_i^2(\boldsymbol{x}_k+\lambda\boldsymbol{d}_k)$$

的极小得 $\boldsymbol{x}_{k+1}$.

读者不难看出，这一方法也有前面无约束极值变尺度法的缺

点．于此，我们也做了一些工作，现拣出其中一点介绍于此书中．

以上便是我们写此书时的取材动机．我们写此书的目的在于应用．所以，那些较单纯的数学讨论我们就不进行了，有兴趣的读者可以查后面所列参考书．但为了使读者把握住方法的基本思想，我们还是尽量做了一些较直观的交待．

§1.3 二次函数与二次近似

二次函数是本书的主要工具，本节对它略做陈述当做数学准备．

假定有矩阵

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots\cdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix},$$

向量

$$\boldsymbol{b}=(b_1,b_2,\cdots,b_n)^T,$$

及纯量 c，我们称

$$Q(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^TA\boldsymbol{x}+\boldsymbol{b}^T\boldsymbol{x}+c$$

为 $\boldsymbol{x}$ 的二次函数．并且一般常假定 A 为对称矩阵．

容易看出，$Q(\boldsymbol{x})$ 的梯度函数为

$$\nabla Q(\boldsymbol{x})=A\boldsymbol{x}+\boldsymbol{b},$$

$Q(\boldsymbol{x})$ 的 Hesse 矩阵为

$$H(\boldsymbol{x})=A.$$

在二次函数中特别称

$$Q(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^TA\boldsymbol{x}$$

为二次型．

如果二次型 $\frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x}$ 满足

$$\boldsymbol{x}^T A\boldsymbol{x} \geqslant 0, \ \forall \boldsymbol{x} \neq \boldsymbol{0},$$

则称其为半正定二次型,特别称满足

$$\boldsymbol{x}^T A\boldsymbol{x} > 0, \ \forall \boldsymbol{x} \neq \boldsymbol{0}$$

的二次型为正定二次型.

有时为了方便加一个系数 $\frac{1}{2}$.

对于一般 n 元非线性可微函数，于一点 $\boldsymbol{x}_k$ 处做 Taylor 展开有

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_k) + (\boldsymbol{x} - \boldsymbol{x}_k)^T \nabla f(\boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^T H(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k) + o(\|\boldsymbol{x} - \boldsymbol{x}_k\|^2).$$

我们取

$$Q(\boldsymbol{x}) = f(\boldsymbol{x}_k) + (\boldsymbol{x} - \boldsymbol{x}_k)^T \nabla f(\boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^T H(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k),$$

称 $Q(\boldsymbol{x})$ 为 $f(\boldsymbol{x})$ 于点 $\boldsymbol{x}_k$ 处的二次近似函数. 不难看出,若 $\boldsymbol{x}_k$ 满足

i) $\nabla f(\boldsymbol{x}_k) = 0$,

ii) $\boldsymbol{y}^T H(\boldsymbol{x}_k)\boldsymbol{y} \geqslant 0$, 对任何 $\boldsymbol{y} \neq \boldsymbol{0}$.

则 $\boldsymbol{x}_k$ 为 $f(\boldsymbol{x})$ 的一个极小值点. 由此可知,对于二次函数

$$\frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$$

若 $\boldsymbol{x}^*$ 满足 $A\boldsymbol{x}^* + \boldsymbol{b} = \boldsymbol{0}$, 而且 $\boldsymbol{x}^T A\boldsymbol{x} \geqslant 0$ 对任何 $\boldsymbol{x}$, 则 $\boldsymbol{x}^*$ 为所给定二次函数的极小值点. 有时也称半正定二次型对应的矩阵 A 为半正定矩阵. 对正定二次型也有相应的称呼. 此时可把上面的结论说成,对于函数 $f(\boldsymbol{x})$, 若点 $\boldsymbol{x}_k$ 使

i) $\nabla f(\boldsymbol{x}_k) = 0$,

ii) $H(\boldsymbol{x}_k)$ 半正定,