优选法基础

D. J. 华尔德
C. S. 皮特勒 著

科 学 出 版 社

优选法基础

D. J. 华尔德
C. S. 皮特勒 著

尤云程 译

科学出版社

1978

内 容 简 介

本书对优选法和最优化的基础知识与实用算法作了较全面的介绍和统一处理．全书包括：微分方法，有约束最优化，几何规划，直接消去法，爬山法，动态规划和大系统的分部最优化，决策改进和最优原理，有试验误差的随机逼近法等内容．书中以推理和实例演示相结合的形式列举了许多实用性的成果．本书可供应用和研究优选法的科技人员以及高等学校有关专业的师生参考．

D. J. Wilde, C. S. Beightler

FOUNDATIONS OF OPTIMIZATION

1967

优 选 法 基 础

D. J. 华尔德
C. S. 皮特勒 著

尤云程 译

*

科学出版社 出版

北京朝阳门内大街137号

中国青年出版社印刷厂印刷

新华书店北京发行所发行　各地新华书店经售

*

1978年11月第一版　开本：850×1168 1/32

1978年11月第一次印刷　印张：12 3/4

印数：0001—90,100　字数：334,000

统一书号：13031·797

本社书号：1144·13—1

定价：1.55 元

代　　序

在工程设计、生产技术、科学试验和计划管理等许多方面，人们总想采取种种措施以获得最好的结果．这一类问题，长期以来经过数学的归纳与提炼，就是范围很广的最优化问题．

五十年代之前，用于解决最优化问题的数学方法仅限于古典的微分法和变分法．近二十几年来，由于实际需要和电子计算机的应用，促进了最优化理论的探索．

优选法，简单地说，就是寻求最好的方式来解决最优化问题．它不仅要找出问题的最优解，而且要提高求解的效率．这当然具有重要的现实意义，特别是对于高度复杂性和紧迫性的问题，效率上的差别也将对整个事情的成功与否发生影响．

本书在基础的数学水平上介绍了优选法和最优化的主要内容，收集了国外的部分研究成果，通过实际例子的数值计算讲述了基本的算法．

最优并不是一劳永逸的概念，优选法和最优化本身尚有不少课题有待进一步研究．随着计算技术的发展，今后在工业控制机上直接应用优选法和最优化数学模型寻找最优工况，实现自动化的最优控制，也许是较为理想的途径．

在本书的翻译中，进行了必要的内容调整和删改．译书中存在的缺点和错误恳望得到同志们的批评和指正．

翻译工作中得到湖北省科委推广优选法办公室的支持与指导，在此致谢．

尤云程

一九七五年五月

目　　录

第一章　间接的微分法

许多具有简单而规整的数学形式的最优化问题可以通过微分方法来解决. 这种方法把寻求目标函数最优值的问题间接地归结为解它的一阶导数为零所形成的方程组. 论述最优化经常以考察这一熟悉的方法开始，在本章中还将提出一些用于处理最优化的新概念.

1-01. 基本概念

在讨论最优化之前，我们必须明确定义所要用到的一些基本概念.

一般的目标函数 y 是依赖于 n 个为实数的独立变量 x_1, x_2, $\cdots$, x_n 的函数，这 n 个变量可简记为具有 n 个分量的列向量或点 $\mathbf{x}$:

$$\mathbf{x}=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}=(x_1, x_2, \cdots, x_n)^T. \tag{1-1}$$

上式中 T 表示向量或矩阵的转置. 当 $\mathbf{x}$ 的分量被定为一组特定数时，所得的向量有时也称为一项策略或设计. 用 $y(\mathbf{x})$ 表示 y 是 $\mathbf{x}$ 的函数.

实践中，许多人为地主观想象的策略在技术上是不现实的，不合理的，不安全的，或者事先已知是不经济的. 因此，独立变量 $\mathbf{x}$ 的取值范围通常要加以一定的数学上的限制，满足这些限制亦即约束条件的点的全体就构成了问题的可行域，记为 F. 如果可行域包括其边界上的所有点，则为闭域. 如果 F 的边界有一部分不

属于 F，则为开域.

例如，由某化工厂的技术经济分析简化而得的目标函数为

$$y = 1000x_1 + 4 \times 10^9 x_1^{-1} x_2^{-1} + 2.5 \times 10^5 x_2. \tag{1-2}$$

在对此函数选取最小点的问题中，如按技术要求，x_1 和 x_2 若满足条件

$$\begin{aligned} 0 \leqslant x_1 \leqslant 2200, \\ 0 \leqslant x_2 \leqslant 8, \end{aligned} \tag{1-3}$$

则这两个不等式在坐标平面上围划的一块长方形区域就是该问题的可行域.

考察 F 中的一个特殊点 $\mathbf{x}^{**}$，该点上的目标函数值比可行域中任何其他点上的函数值小，即对 F 中一切 $x \neq x^{**}$ 有

$$y(\mathbf{x}^{**}) < y(\mathbf{x}), \tag{1-4}$$

则 $y(\mathbf{x}^{**})$ 或 y^{**} 是 y 的最小值，而 $\mathbf{x}^{**}$ 叫做最小点，它是唯一的. 如果对 F 中一切 $\mathbf{x} \neq \mathbf{x}^{**}$，(1-4) 式由下式取代:

$$y(\mathbf{x}^{**}) \leqslant y(\mathbf{x}), \tag{1-5}$$

则存在不止一个最小点，但目标的最小值仍是唯一的. 上式可写为

$$y^{**} = \min_{\mathbf{x} \in F} [y(\mathbf{x})]. \tag{1-6}$$

如果 $\mathbf{x}^{**}$ 在 F 的边界上，则 y^{**} 为边界最小值；否则便是内部最小值.

在相当常见的情况下，最小值并不一定存在，特别当可行域是开集时，往往是如此. 对于这种情形，考虑不一定在 F 中的点 $\mathbf{x}'$，这点上的 y 值比 F 中任何一点上的值都小. 对 $\mathbf{x} \in F$ 有

$$y(\mathbf{x}') < y(\mathbf{x}), \tag{1-7}$$

这样的点 $\mathbf{x}'$ 叫做 $y(\mathbf{x})$ 的一个下界点. 令 F' 为所有下界点的全体，函数 $y(\mathbf{x})$ 的下确界（即最大下界）是满足下式的点 $\bar{\mathbf{x}}$，即对 $\mathbf{x}' \in F'$ 有

$$y(\bar{\mathbf{x}}) \geqslant y(\mathbf{x}'). \tag{1-8}$$

y 的下确界记为

$$\bar{y} = y(\bar{\mathbf{x}}) = \inf_{\mathbf{x} \in F} [y(\mathbf{x})]. \tag{1-9}$$

为了避免这种繁琐，采用闭区域．数学分析中已经证明，闭区域上的连续函数必达到最小值．如果 F 是闭域，

$$y^{**} = \min_{\mathbf{x}} [y(\mathbf{x})] = \inf_{\mathbf{x}} [y(\mathbf{x})] = \bar{y}. \qquad (1\text{-}10)$$

以后在清楚的场合，$\mathbf{x}$ 总是取在可行域中，可以省去 $\mathbf{x} \in F$ 的详细说明．

y 的最大值和对应的最大点是以同样的方式来定义和表记的，只要把上面一些不等式中的不等号反向即可．当然，在具体问题中，到底要求最小还是最大应是清楚的．

$$y^{**} = y(\mathbf{x}^{**}) = \max_{\mathbf{x}} [y(\mathbf{x})] \qquad (1\text{-}11)$$

为了一般地统一成最优化的概念，我们称 y^{**} 为最优值，而 $\mathbf{x}^{**}$ 为最优点，记为

$$y^{**} = y(\mathbf{x}^{**}) = \operatorname*{opt}_{\mathbf{x}} [y(\mathbf{x})]. \qquad (1\text{-}12)$$

现在考虑一个可行点 $\mathbf{x}^*$ 和在 $\mathbf{x}^*$ 的一个可行邻域中的点集．这些点不仅满足规定 F 的限制条件，而且满足附加条件

$$0 < |\mathbf{x} - \mathbf{x}^*| < \varepsilon, \qquad (1\text{-}13)$$

其中

$$|\mathbf{x} - \mathbf{x}^*| = \left[\sum_{j=1}^{n} (x_j - x_j^*)^2\right]^{1/2}. \qquad (1\text{-}14)$$

这可行邻域记为 N．如果存在一个包含 $\mathbf{x}^*$ 的可行邻域且对 $\mathbf{x} \in N$ 有

$$y(\mathbf{x}^*) \leqslant y(\mathbf{x}) \qquad (1\text{-}15)$$

则 $y(\mathbf{x}^*)$ 称为一个局部最小值，而 $\mathbf{x}^*$ 为局部最小点．

$$y^* = y(\mathbf{x}^*) = \operatorname*{\mathit{l}min}_{\mathbf{x}} [y(\mathbf{x})]. \qquad (1\text{-}16)$$

y^* 与 y^{**} 之间有区别，后者是最希望的结果，而前者却是间接微分法通常能够达到的． y^{**} 有时也叫全域最小值．可类似说明可行邻域，以建立局部下确界、上确界、最大值和最优值等概念．只有唯一的局部亦即全域最优值的目标函数称为单峰函数．

最小和最大之间存在着一定的基本关系． 若 b 为正数，a 为

任意实数，如果 $\mathbf{x}^{**}$ 是 $y(\mathbf{x})$ 的最小点，它便是 $a+by(\mathbf{x})$ 的最小点：

$$a+by(\mathbf{x}^{**})=\min_{\mathbf{x}}[a+by(\mathbf{x})]. \tag{1-17}$$

进而它又是 $a-by(\mathbf{x})$ 的最大点：

$$a-by(\mathbf{x}^{**})=\max_{\mathbf{x}}[a-by(\mathbf{x})]. \tag{1-18}$$

这里可以看出两点．其一，加上常数与乘以正数并不影响最优点的位置；其二，只要改变目标函数的符号，最小化方法可用来解决求最大点的问题，反之亦真．

1-02. 微分方法

最优化方法可以分为两类：直接法和间接法．

直接法是从任取一点出发，逐步改进而趋向最优点．这种方法留待以后讲．古典的微分法之所以是间接的，因为它最后涉及到解一个方程或方程组，而不是直接搜寻最优点，解出方程的根就容易知道最优点的位置．这种微分方法不需要测试非最优点即能选出最优点．因此，只要能用上，它是很有效的．

单变量函数 $y(x)$ 关于 x 的一阶导数定义为

$$\frac{dy}{dx}=\lim_{\Delta x\to 0}\frac{y(x+\Delta x)-y(x)}{\Delta x}=y', \tag{1-19}$$

而多变量函数 $y(\mathbf{x})$ 关于 n 个变量中的某一变量 x_i 的一阶偏导数是

$$\frac{\partial y}{\partial x_i}=\lim_{\Delta x_i\to 0}\frac{y(x_1,\cdots,x_{i-1},x_i+\Delta x_i,x_{i+1},\cdots,x_n)-y(\mathbf{x})}{\Delta x_i}. \tag{1-20}$$

对某一变量求偏导数时，其他变量视为常数不变．

如用 $\partial\mathbf{x}$ 表示独立变量的微小摄动组成的列向量

$$\partial\mathbf{x}=\begin{pmatrix}\partial x_1\\ \partial x_2\\ \vdots\\ \partial x_n\end{pmatrix}, \tag{1-21}$$

则由此引起函数值的变化 ∂y，即 $y(\mathbf{x}+\partial\mathbf{x})$ 和 $y(\mathbf{x})$ 之差可以用泰勒级数来表示；

$$\begin{aligned}\partial y &= y(\mathbf{x}+\partial\mathbf{x})-y(\mathbf{x})\\ &=\left(\frac{\partial y}{\partial x_1}\right)\partial x_1+\left(\frac{\partial y}{\partial x_2}\right)\partial x_2+\cdots+\left(\frac{\partial y}{\partial x_n}\right)\partial x_n+O(\partial\mathbf{x}^2)\\ &=\sum_{i=1}^{n}\left(\frac{\partial y}{\partial x_i}\right)\partial x_i+O(\partial\mathbf{x}^2),\end{aligned}\tag{1-22}$$

其中，$O(\partial\mathbf{x}^2)$ 表示至少含有两个变量微分之乘积的无穷项之总和．如当 ∂x_i 趋于零时，比值 $O(\partial\mathbf{x}^2)/\partial x_i$ 极限地趋于零，则 $O(\partial\mathbf{x}^2)$ 项可以近似地忽略．在这样的假设下，考虑将有限变分写成

$$\Delta y=\nabla y\Delta\mathbf{x},\tag{1-23}$$

对于微分写为

$$\partial y=\nabla y\partial\mathbf{x},\tag{1-24}$$

这里 ∇y 是一阶偏导数组成的行向量，

$$\nabla y=\left(\frac{\partial y}{\partial x_1},\ \frac{\partial y}{\partial x_2},\ \cdots,\ \frac{\partial y}{\partial x_n}\right)\tag{1-25}$$

称为 y 的*梯度*．

在一个局部内最优点 $\mathbf{x}^*$ 处必须成立的条件是由它的定义和在任一包含 $\mathbf{x}^*$ 的可行开邻域 N 中的泰勒级数展开式推导出来的．设 $\mathbf{x}^*$ 是一个局部最小点，对所有可能的摄动 ∂x_i，极限地成立

$$\sum_{i=1}^{n}\left(\frac{\partial y}{\partial x_i}\right)^*\partial x_i=\partial y=y(\mathbf{x}^*+\partial\mathbf{x})-y(\mathbf{x}^*)\geqslant 0.\tag{1-26}$$

由 ∂x_i 符号的任意性推知，每一个偏导数都必须为零；

$$\left(\frac{\partial y}{\partial x_i}\right)^*=0,\quad i=1,2,\cdots,n.\tag{1-27}$$

如果不是这样，设有一个 $\left(\frac{\partial y}{\partial x_i}\right)^*\neq 0$，则取与此 $\left(\frac{\partial y}{\partial x_i}\right)^*$ 符号相反的单变量摄动 ∂x_i，而令其他变量的摄动均为零，则产生在 y^* 的每一可行邻域中的函数值更小的点，这与 y^* 是局部最小值相矛盾．同样可以证明，(1-27) 式对局部内最大点也成立，因此，$\mathbf{x}^*$ 是内

部最优点的必要条件是那里的梯度为零，

$$\nabla y(\mathbf{x}^*) = \nabla y^* = \mathbf{0}, \tag{1-28}$$

即梯度为 n 维零向量. 几何上讲，最优点处的切线(切平面)是水平的.

例如，对于 (1-2) 式所示的目标函数，要求在 (1-3) 式约束下的最小值及最小点. 按上面所述的方法，令函数关于 x_1 和 x_2 的两个一阶偏导数都为零，这样得到两个联立方程

$$\frac{\partial y}{\partial x_1} = 1000 - 4 \times 10^9 x_1^{-2} x_2^{-1} = 0, \tag{1-29}$$

$$\frac{\partial y}{\partial x_2} = 2.5 \times 10^5 - 4 \times 10^9 x_1^{-1} x_2^{-2} = 0. \tag{1-30}$$

解此联立方程，求得唯一的极值点

$$x_1^* = 1000, \quad x_2^* = 4.$$

这里的局部最小点也就是目标函数在整个可行域上的最小值点，所得 $y^* = 3 \times 10^6$. 图 1-1 画出了这一目标函数在可行域中的等高线.

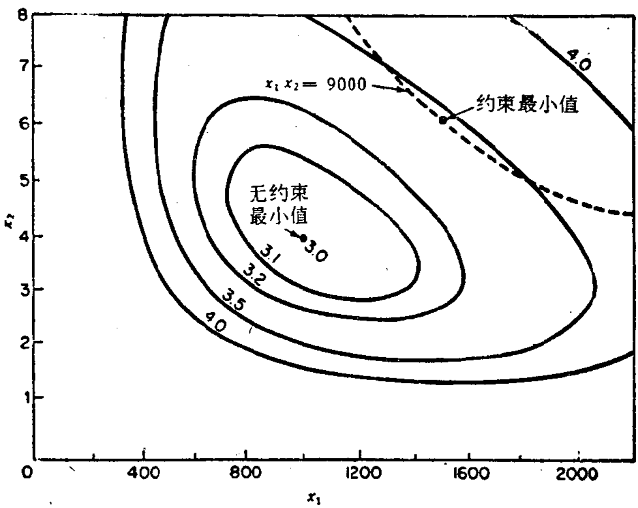

图 1-1 目标函数等高线

1-03. 解非线性方程组

因为微分方法最后归结为解非线性方程组，我们在此介绍解非线性方程(或方程组)的牛顿法.

把目标函数 $y(\mathbf{x})$ 的 n 个一阶偏导数记为

$$\frac{\partial y}{\partial x_i} = y_i'(\mathbf{x}),\quad i = 1, \cdots, n, \tag{1-31}$$

则 (1-27) 式即为

$$y_i'(\mathbf{x}^*) = 0,\quad i = 1, \cdots, n. \tag{1-32}$$

牛顿法的基本思路是从任一点开始，逐步以线性近似的形式迭代，最终使其收敛于方程组 (1-32) 式的根 $\mathbf{x}^*$. 设 $\mathbf{x}_k$ 为试验点，而 $\left(\frac{\partial y_i'}{\partial x_j}\right)_k$ 为在 $\mathbf{x}_k$ 点计算的 y_i' 的 n 个一阶偏导数，也就是原目标函数 y 的二阶偏导数，简记为

$$\left(\frac{\partial y_i'}{\partial x_j}\right)_k = \left(\frac{\partial^2 y}{\partial x_i \partial x_j}\right)_k = y_{ij}''(\mathbf{x}_k). \tag{1-33}$$

牛顿法利用这些二阶导数来估计在 $\mathbf{x}_k$ 邻域中的一阶导数的值，然后将这种估计的一阶导数全为零的新点 $\mathbf{x}_{k+1}$ 选为 $\mathbf{x}^*$ 的更好的逼近点，依此逐次迭代，直至所有 y_i' 变为充分小，令人满意为止. 将 y_i' 关于 $\mathbf{x}_k$ 展开就给出 $\mathbf{x}_{k+1}$ 为下式的解:

$$y_i'(\mathbf{x}_{k+1}) = y_i'(\mathbf{x}_k) + \sum_{j=1}^{n} y_{ij}''(\mathbf{x}_k)(x_{j,k+1} - x_{jk}) = 0. \tag{1-34}$$

既然 $y_i'(\mathbf{x}_k)$ 和 $y_{ij}''(\mathbf{x}_k)$ 均是已知的，上面这 n 个线性方程中只有 $\mathbf{x}_{k+1}$ 的 n 个坐标值 $x_{j,k+1}(j = 1, 2, \cdots, n)$为未知.

用矩阵记号来表述以上这些情形是方便的. 设将 n^2 个二阶偏导数排列成方阵如下:

$$\mathbf{H}(\mathbf{x}_k) = \mathbf{H}_k = \begin{pmatrix} y_{11}'' & y_{12}'' & \cdots & y_{1n}'' \\ y_{21}'' & y_{22}'' & \cdots & y_{2n}'' \\ \vdots & \vdots & & \vdots \\ y_{n1}'' & y_{n2}'' & \cdots & y_{nn}'' \end{pmatrix}. \tag{1-35}$$

这个矩阵可称为海辛 (Hessian) 阵 $\mathbf{H}$. 用矩阵乘法规则，将 (1-34)