计算数学讲义（五）

最优化方法

南京大学数学系计算数学专业 编

科学出版社

计算数学讲义(五)

最优化方法

南京大学数学系计算数学专业 编

科学出版社

内 容 简 介

近年来，最优化方法引起普遍重视，已形成一门学科．目前，这方面的新算法不断出现，应用于科学技术各个领域．本书介绍现有算法中简单易行、效果较好者，大约分为三类：

1）最小二乘法（包括阻尼最小二乘法）．这类方法适用于平方和形式的目标函数；

2）斜量方法（包括共轭斜量法和变尺度方法）．这类方法适用于一般的目标函数；

3）直接方法（包括单纯形法和共轭方向法）．这类方法适用于一般目标函数，且不需要计算导数．

计算数学讲义（五）

最 优 化 方 法

南京大学数学系计算数学专业 编

*

科 学 出 版 社 出版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行　各地新华书店经售

*

1978 年 8 月第 一 版　开本：787×1092　1/32

1978 年 8 月第一次印刷　印张：6

印数：0001—128,400　字数：135,000

统一书号：13031·758

本社书号：1084·13—1

定价：0.62 元

说　　明

1. 这一套《计算数学讲义》是在我专业过去所编教材的基础上修改补充而成的.

2. 这套讲义共分下列九册:

(一)数值逼近方法,

(二)线性代数计算方法,

(三)常微分方程数值解法,

(四)偏微分方程数值解法,

(五)最优化方法,

(六)概率统计基础和概率统计方法,

数学基础之一: 线性代数,

数学基础之二: 常微分方程,

数学基础之三: 偏微分方程.

由于篇幅的原因,我们把《概率统计基础》和《概率统计方法》合并为一册.

3. 这套讲义可作为综合性大学理科计算数学专业教材,也可供利用电子计算机从事科学计算的科技人员参考.

4. 这套《计算数学讲义》的主编是何旭初同志.

讲义各册由我专业有关同志分工负责.

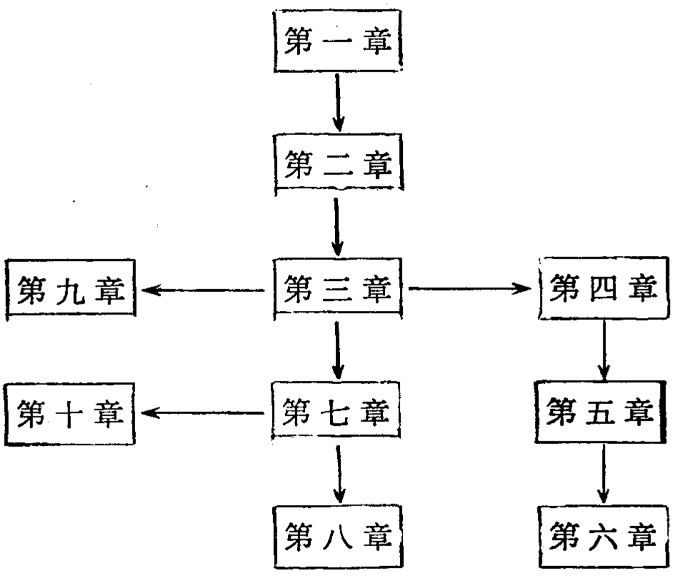
这册《最优化方法》的编写者为何旭初同志,林应举、颜起居、盛松柏、李明霞等同志也参加了部分工作.

5. 由于理论水平和实践经验所限,讲义中的缺点和错误在所难免,我们衷心盼望读者提出宝贵意见,以便进一步修改.

南京大学数学系计算数学专业

1978 年 1 月

各章阅读次序框图



目　　录

第一章　最优化问题的例

§1. 引　　言

最优化方法(或技术)是在高速电子计算机出现以后由于大量社会实践的需要而迅速发展起来的一门新兴学科. 在许多部门,如建筑结构、船体设计、化学工程、空间科学、电子科学、核动力以及许多工程设计、生产控制等方面,最优化方法得到了日益广泛的应用. 为了使读者了解最优化方法的作用,这一章列举一些实例来加以说明.

§2. 化学反应的平衡组成

复杂的有机化合物的合成以及复杂燃料(例如火箭燃料)的配方涉及如下问题: 若参与化学反应的物质中含有m种不同的元素(其比例为已知),反应的生成物中含有N种化合物(其比例待定),问如何确定这N种生成物的比例关系?这个问题对调整参与反应物质的比例、改进配方有重要关系. 在一定的温度和压力下,每一种化合物都具有一定的自由能. 根据化学热力学原理,化学反应达到平衡状态时,系统的总自由能为最小. 据此,我们可以建立如下的数学模型.

先设参与反应的元素为 $E^{(1)}, E^{(2)}, \cdots, E^{(m)}$,可能的生成物为 $C^{(1)}, C^{(2)}, \cdots, C^{(N)}$. 第 i 个生成物 $C^{(i)}$ 中含有 a_{ij} 个原子 $E^{(j)}$, 即生成物的分子式为

$$C^{(1)} = E^{(1)}_{a_{11}}\ E^{(2)}_{a_{12}} \cdots E^{(m)}_{a_{1m}},$$

$$C^{(2)} = E_{a_{21}}^{(1)} E_{a_{22}}^{(2)} \cdots E_{a_{2m}}^{(m)},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$C^{(N)} = E_{a_{N1}}^{(1)} E_{a_{N2}}^{(2)} \cdots E_{a_{Nm}}^{(m)}.$$

次设参与反应的第 j 种元素有 $b_j (j = 1, \cdots, m)$ 个原子量，生成 $x_i (i = 1, 2, \cdots, N)$ 个分子 $C^{(i)}$:

$$b_1 E^{(1)} + b_2 E^{(2)} + \cdots + b_m E^{(m)} \to$$
$$x_1 C^{(1)} + x_2 C^{(2)} + \cdots + x_N C^{(N)}.$$

在这个系统中，第 i 种化合物的自由能为

$$f_i(x_1, x_2, \cdots, x_N) = x_i(c_i + \ln x_i - \ln \bar{x}),$$

其中(在生成物为气体时)

$$c_i = \left(\frac{F}{RT}\right)_i + \ln P,$$

$$\bar{x} = \sum_{i=1}^{N} x_i,$$

P 为压力，T 为绝对温度，R 为克分子气体常数，F 为与相应化合物有关的常数. 系统的总自由能为

$$\Phi(x_1, x_2, \cdots, x_N) = f_1(x_1, x_2, \cdots, x_N) + \cdots + f_N(x_1, \cdots, x_N). \qquad (1)$$

在反应达到平衡状态时，$x_1, x_2, \cdots, x_N$ 应使总自由能 Φ 为最小.

应当指出，诸 x_i 不能是任意的. 首先，它们都不能是负数(生成物的克分子数为负是没有意义的)，即

$$x_i \geqslant 0, \quad i = 1, 2, \cdots, N. \qquad (2)$$

其次，它们应当满足质量守恒原理，即，设 b_j 为参与反应的第 j 种元素的原子量总数，a_{ij} 为第 i 种生成物中第 j 种元素的原子数，b_j 与 a_{ij} 均为已知. 根据质量守恒原理，反应前后物质的总量不变，因此，x_i, a_{ij}, b_j 应满足下列关系:

$$\sum_{i=1}^{N} a_{ij} x_i = b_j, \ j = 1, 2, \cdots, m. \tag{3}$$

于是,问题就归结如下:

求 $x_1, x_2, \cdots, x_N$,使函数 Φ 在满足条件(2)和(3)的情况下为最小.

§3. 光学系统的自动设计

一个光学系统,如照像机镜头,显微镜的物镜和目镜,望远镜以及电影放映机的镜头等,一般是由若干个球面透镜组成的,这些透镜有一条共同的光轴(即对称轴),而制造透镜的材料是具有不同折射率的光学玻璃. 各镜面都有一定大小的半径和厚度,各镜面之间保持一定的间隔. 一个光学系统中各透镜镜面的半径、厚度、镜面间隔以及各透镜所用光学玻璃的折射率,便是这个光学系统的*结构参数*. 这些结构参数的大小比例,直接影响系统的质量. 所谓光学系统的设计问题,就是根据对一个光学系统的要求,设法确定各结构参数的大小,使该系统成像的质量尽可能地好.

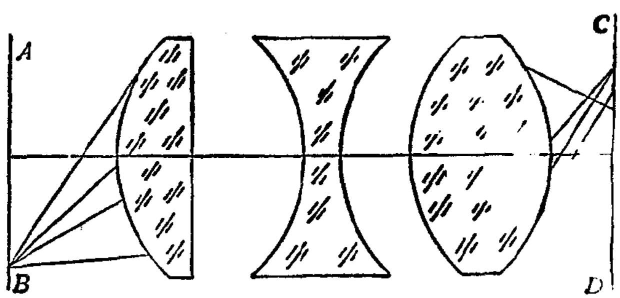


图 1-1

图 1-1 是一个简单的光学系统,AB 为物平面,CD 为像平面. 对于一个理想的光学系统,发自物平面上某一点的光

束，通过光学系统后应会聚在像平面上的一个点处，而且像的形状还保持与物的形状的几何相似性. 但是，在实际上，由于镜面为球面，且各色光在同一介质中的传播速度不同，发自物面一点的光束，并不会聚在像面上的一个点处，而是交像平面于无穷多个点处，分散在一定的范围内；同时，像并不保持与物的几何相似性. 这样就产生了所谓像差. 像差的存在使像的清晰度受到影响，并且失掉了真实形状. 设计者的任务就是设法调整光学系统的结构参数，使各种像差不超出规定的范围，尽可能使成像质量达到完善的程度.

在设计光学系统的过程中，需要利用光路计算公式算出发自物平面上一点处的一条光线通过该系统后达到像平面上的位置. 物平面和像平面垂直于光轴，取它们和光轴的交点为原点，分别建立座标系. 物面上点的座标记作 (u, v)，像平面上相应点的座标记作 (u_F, v_F). 发自一点的一条光线由它的方向余弦 α, β, γ 所确定，因为这三个方向余弦满足等式

$$\alpha^2+\beta^2+\gamma^2=1.$$

所以，实际上只有两个独立量，取其为 α, β. 从而可以用 (u, v, α, β) 来表示发自物面上一点的一条光线.

设光学系统的结构参数为 $x_1, x_2, \cdots, x_N$，则像点座标 (u_F, v_F) 既依赖于相应光线的座标 (u, v, α, β)，同时也依赖于系统的结构参数 $x_1, x_2, \cdots, x_N$，即 u_F, v_F 都是 $N+4$ 个变量 $u, v, \alpha, \beta, x_1, x_2, \cdots, x_N$ 的函数. 为了成像清晰并保持原形，应使它们尽可能满足关系：

$$\begin{aligned} u_F &= Mu, \\ v_F &= Mv, \end{aligned} \tag{4}$$

其中 M 为光学系统的放大率. 差数 $u_F - Mu, v_F - Mv$ 实际上并不为零，它们是 $u, v, \alpha, \beta, x_1, \cdots, x_N$ 的函数. 设

$$
\begin{aligned}
G(u, v, \alpha, \beta, x_1, \cdots, x_N) &\equiv u_F - Mu, \\
H(u, v, \alpha, \beta, x_1, \cdots, x_N) &\equiv v_F - Mv.
\end{aligned} \tag{5}
$$

在物面上选取若干个适当的点$(u_i, v_i)(i = 1, 2, \cdots, m)$，从每一点$(u_i, v_i)$再选若干条光线，它们的相应方向余弦为$\alpha_{ij}$, $\beta_{ij}(j = 1, 2, \cdots, n)$. 对于每一条光线

$$(u_i, v_i, \alpha_{ij}, \beta_{ij}),$$

记函数G和H为G_{ij}和H_{ij}.

令

$$\phi(x_1, x_2, \cdots, x_N) = \sum_{j=1}^{n} \sum_{i=1}^{m} (G_{ij}^2 + H_{ij}^2). \tag{6}$$

设计者的任务就是设法求得使

$$\phi(x_1, x_2, \cdots, x_N) = \text{极小} \tag{7}$$

的结构参数.

还要指出，结构参数$x_1, \cdots, x_N$不是完全独立的，它们应满足某些约束条件.

§4. 参数估计和电路设计

同轴电缆传输多道载波系统中各类均衡网络的电路设计具有下述共同的数学模型：

电路的特性函数$\varphi(\omega)$是给定的，要求用某种形式的电路结构使之具有这种特性. 设电路参数为$x_1, \cdots, x_N$，对于所选用的电路结构，特性函数的表达式为$f(x_1, \cdots, x_N; \omega)$. 问题是设法确定这$N$个电路参数$x_1, \cdots, x_N$，使实际上的特性函数$f$在$\omega$的某一区间$[\alpha, \beta]$上逼近规定的特性函数$\varphi(\omega)$并达到指定的精确度，即，使

$$\max_{\alpha \leqslant \omega \leqslant \beta} |f(x_1, \cdots, x_N; \omega) - \varphi(\omega)| \leqslant \varepsilon. \tag{8}$$

这本来是一个连续区间上的逼近问题，但为了易于求解，常把

它离散化：在区间 $[\alpha, \beta]$ 上取 M 个 ω 值

$$\omega_1, \omega_2, \cdots, \omega_M,$$

令

$$\varphi_i = \varphi(\omega_i),$$

则问题(8)便可以近似地归纳如下：求 $x_1, \cdots, x_N$ 使

$$|f(x_1, \cdots, x_N; \omega_i) - \varphi_i| \leqslant \varepsilon, \quad i = 1, \cdots, M. \tag{9}$$

这时，问题(8)的解决就成为解不等式(9)了. 因为 $f(x_1, \cdots, x_N; \omega)$一般都是 $x_1, \cdots, x_N$ 的非线性函数，而解非线性不等式组比较困难，因此，又往往把解非线性不等式组(9)的问题转化为求 $x_1, \cdots, x_N$ 使

$$F(x_1, \cdots, x_N) = \sum_{i=1}^{M} f_i^2(x_1, \cdots, x_N) = \text{极小}, \tag{10}$$

其中

$$f_i(x_1, \cdots, x_N) = f(x_1, \cdots, x_N; \omega_i) - \varphi_i,$$
$$i = 1, \cdots, M.$$

问题中的变元 $x_1, \cdots, x_N$ 一般不能是随意的，而要受到某些限制(即约束).

我们称前述确定参数 $x_1, \cdots, x_N$ 的问题为非线性参数估计问题.

§5. 非线性方程组的求解

解非线性方程组是相当困难的一类问题，由于最优化方法的发展，对解非线性方程组提供了一种有力的手段. 设求解的方程组如下：

$$\begin{aligned} &f_1(x_1, x_2, \cdots, x_N) = 0, \\ &f_2(x_1, x_2, \cdots, x_N) = 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &f_N(x_1, x_2, \cdots, x_N) = 0, \end{aligned} \tag{11}$$

可令

$$\phi(x_1, x_2, \cdots, x_N) = \sum_{i=1}^{N} f_i^2(x_1, x_2, \cdots, x_N). \quad (12)$$

若方程组(11)有解，则此解必为函数 $\phi(x_1, x_2, \cdots, x_N)$的极小点. 因为 ϕ 是一个平方和，其值最小为零，因此，解非线性方程组(11)的问题便化为求 $x_1, x_2, \cdots, x_N$ 使

$$\phi(x_1, x_2, \cdots, x_N) = \text{极小} \quad (13)$$

的问题，从而最优化方法可用于非线性方程组的求解.

§6. 光学多层膜系的设计

图 1-2 是光在N层介质膜上反射和透射的示意图. 从薄膜的光学理论知道，当一束光(波长为 λ)以入射角为 φ_0 的方

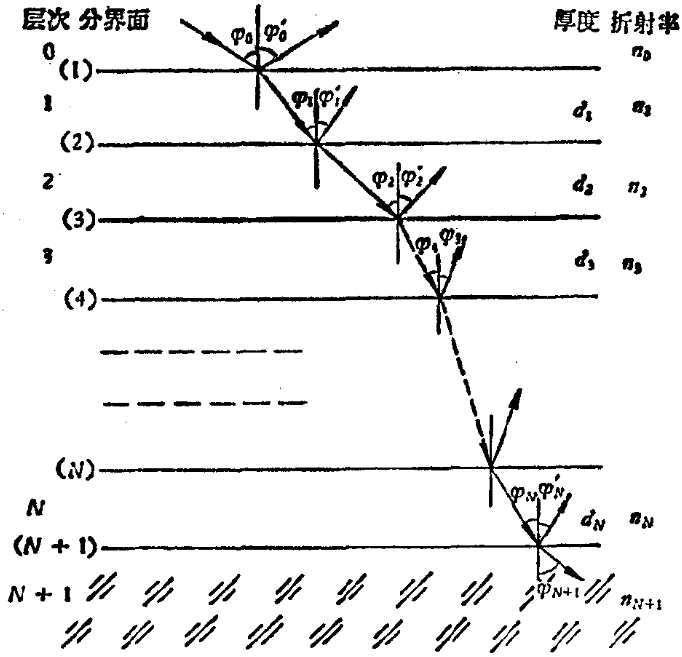


图 1-2　光在N层介质膜上反射和透射示意图

向通过折射率呈不连续变化的层状介质时，则发生多次的反射和透射. 若以 R 表示反射部分或反射率（即反射能量和入射能量的比值）；以 T 表示透射部分或透射率（即透射能量和入射能量的比值）；以 A 表示被介质吸收的部分或吸收率（即被吸收的能量和入射能量的比值），则应该有

$$R + T + A = 1.$$

但是，当把光束限制在很小的范围内时，吸收部分便可以忽略不计，即 $A \doteq 0$. 因此，一般有

$$R + T = 1, \tag{14}$$

即入射光束的光能被多层膜系分成反射和透射两个部分，而且这两部分的能量总和等于入射光束的能量，这就是光学多层膜系的分光特性. 利用这个特性，我们可以设计各种膜系，应用于各种光学仪器，如激光技术中使用的特定光谱波段就具有高反射率的介质膜，高透射膜；彩色电视中使用的分色膜；还有光学仪器中使用的宽带增透膜等等.

从薄膜的光学理论知道，一组膜系在入射光线的入射角 φ_0，入射介质的折射率 n_0 和衬底（通常是用玻璃做的）的折射率 n_{N+1} 给定的条件下，一组膜系的反射率 R 取决于所用介质（膜料）的折射率 $n_1, n_2, \cdots, n_N$ 和各膜层的光学厚度（即几何厚度 d 和相应的膜料折射率 n 的乘积）$x_1, x_2, \cdots, x_N$ 以及所考虑的波段的光波的波长 λ，即

$$R = R(n_1, n_2, \cdots, n_N; x_1, \cdots, x_N; \lambda). \tag{15}$$

具体地讲，一个膜系的结构参数有 $n_1, n_2, \cdots, n_N$; $x_1, x_2, \cdots, x_N$; λ. 但是，由于目前可供镀膜的材料不多，一般设计都是选好了膜料，只把光学厚度作为结构参数，故 R 可简单地表示如下：

$$R = R(x_1, x_2, \cdots, x_N; \lambda). \tag{16}$$

具体计算可采用光学导纳矩阵法，它是一个递推的方

法：

$$R(\lambda)=\left(\frac{\eta_0-Y}{\eta_0+Y}\right)\overline{\left(\frac{\eta_0-Y}{\eta_0+Y}\right)}=\gamma_1^2+\gamma_2^2, \tag{17}$$

其中记号“$\bar{x}$”表示取共轭复数，而

$$Y=\frac{C}{B}, \tag{18}$$

$$\begin{bmatrix} B \\ C \end{bmatrix}=\prod_{k=1}^{N}\begin{bmatrix} \cos\delta_k & i\sin\delta_k/\eta_k \\ i\eta_k\sin\delta_k & \cos\delta_k \end{bmatrix}\begin{bmatrix} 1 \\ \eta_{N+1} \end{bmatrix}, \tag{19}$$

其中

$$\delta_k=\frac{2\pi x_k\cos\varphi_k}{\lambda},\ k=1,2,\cdots,N; \tag{20}$$

$$\eta_k=\begin{cases} n_k\cos\varphi_k, & S\text{ 分量}, \\ n_k/\cos\varphi_k, & P\text{ 分量}, \end{cases} k=0,1,2,\cdots,N+1, \tag{21}$$

这里的 φ_k 由下式求出：

$$n_0\sin\varphi_0=n_k\sin\varphi_k,\ k=1,2,\cdots,N. \tag{22}$$

上面的 Y 称为等效导纳，δ_k 称为位相厚度，η_k 称为光学导纳.

设计一个膜系，就是在给定的波段 $\lambda_a\leqslant\lambda\leqslant\lambda_b$ 上用某种方法确定结构参数 $x_1, x_2,\cdots, x_N$，以便获得所要求的光谱特性(能量反射率或能量透射率). 例如在给定的波段上选取 $M(M>N)$ 个点 $\lambda_j(\lambda_{a_j}\leqslant\lambda_j\leqslant\lambda_{b_j})$，用某种方法确定结构参数 $x_1, x_2,\cdots,x_N$，使在这 M 个点上，其反射率达到预先给定的值 $R_1^0, R_2^0,\cdots,R_M^0$，即

$$|R_j-R_j^0|\leqslant\varepsilon_j,\ j=1,2,\cdots,M, \tag{23}$$

其中 R_j 是计算出来的值，ε_j 是预先给定的精确度. 这样确定出来的 $x_1, x_2,\cdots, x_N$ 就是一种最优的设计方案.

若令

$$f_j=R_j-R_j^0,\ j=1,2,\cdots,M, \tag{24}$$

$$\boldsymbol{x}=[x_1, x_2,\cdots, x_N]^T,$$

则称 f_j 为对于波长为 λ_j 的光束对于膜系的反射率残差，$\boldsymbol{x}$ 称为膜系的一种设计方案.

传统的设计方法，就是根据光学多层膜的光学性质对给定的条件进行分析，考虑到工艺上实现的可能性，加上设计者在这方面的经验，预先选定膜系的层数 N，选取膜料和衬底(即给出 $n_0, n_1, \cdots, n_N, n_{N+1}$)，确定入射角 φ_0 以及初始的设计方案 $\boldsymbol{x}^0=(x_1^0, x_2^0, \cdots, x_N^0)^T$，然后计算反射率 $R_j(j=1,2,\cdots,M)$，看它们是否达到预先给定的值 $R_j^0(j=1,2,\cdots,M)$. 若不满足，再修改初始结构参数，重复以上过程，一直进行到满足要求为止. 由于计算诸 R_j 的工作量大，初始结构参数的选取和结构参数的修改往往带有盲目性，因此，就出现了利用最优化方法在电子数字计算机上进行最优设计.

膜系设计中的评价函数为

$$\Phi(\boldsymbol{x})=\sum_{k=1}^{M}\mu_k f_k^2, \tag{25}$$

其中 μ_k 称为权重因子，其相对大小取决于光源的能量分布，受光器的灵敏度以及诸残差对膜系的相对重要性.

设计者的任务就是设法求得使

$$\Phi(\boldsymbol{x})=\sum_{k=1}^{M}\mu_k f_k^2=\text{极小} \tag{26}$$

的 $\boldsymbol{x}^*=[x_1^*, x_2^*, \cdots, x_N^*]^T$，这就是所谓膜系的最优设计问题.

最后还必须指出，由于工艺上实现的可能性以及其他问题，对结构参数是有一定的限制的，即有约束条件，这在解题时是必须加以考虑的.

§7. 小　　结

前面列举的一些问题，有一个共同的特点，就是均为求变