最优化方法

〔西德〕海宁·吐尔著

Optimization Methods

机械工业出版社

最优化方法

〔西德〕海宁·吐尔 著
欧天垣 曹叔维 译
马国选 校

机械工业出版社

最优化方法

〔西德〕海宁·吐尔　著
欧天垣　曹叔维　译
马国选　校

*

机械工业出版社出版（北京阜成门外百万庄南街一号）
（北京市书刊出版业营业许可证出字第 117 号）
机械工业出版社印刷厂印刷
新华书店北京发行所发行·新华书店经售

*

开本 850×1168 1/32·印张 8 3/4·字数 224 千字
1982 年 11 月北京第一版·1982 年 11 月北京第一次印刷
印数　0,001—8,000·定价 1.10 元

*

统一书号：15033·5115

前　言

建立在变分法基础上的最优化方法，从本世纪三十年代卡勒茜奥多立开始，已经历了半个世纪。但这类问题的求解只有在个别情况下才有可能，并且大多数在数值处理过程中还会碰到计算上的困难。从五十年代以来，数字计算机的出现，逐步使我们能实际地应用变分法所提供的公式，并能更多地从数值分析上加以考虑。这一点，特别在控制系统工程和空间飞行技术领域中被证明是非常成功的。因为，有时从最优化得到的，甚至是相对很小的增益也是很可贵的。如果，我们在一个工程问题中，有可能选取一个关于时间的函数，则它的绝对最优过程总会显得十分有用。因为：

1．在任何情况下，我们必须对“时间”函数选择某个过程，这样来获得一个最优过程是非常实际的。

2．同时，当选择某个在技术上合适的过程时，我们总想知道这个系统的工作情况，还有哪些可改进的地方。特别在当前世界上消耗性资源日趋减少时，人类更加重视获得最好的实际性能指标。

本书是针对各种工程技术人员和对这些现代方法感兴趣的数学工作者而写的。重点是介绍五十年代后所出现的最优化方法。给出了最优化方法的全貌，而省略了大部分的数学证明。第一部分介绍变分法的基本概念；第二部分讨论庞特里雅金极大值原理及利用间接法来处理具有非线性微分方程的问题。第三部分讨论梯度法及贝尔曼动态规划。在最后的附录中介绍了九个例题，加深读者对这些方法的理解。由于作者早期曾在空间工业方面担任工作，所以这些例子几乎全部来自空间飞行技术，但并不意味着所述过程仅对空间飞行技术适用。

本书保持了基本文献中的常用记号，也使用了某些附加符号，例如在庞特里雅金极大值原理中，用$H^{(-)}$来代替H，以避免对不同的物理量使用相同的符号。在三个主要部分中，每一部分的公式都是连贯编号的。在同一部分中，直接标出了编号，比如(75)。在其它部分中，前面加标该部分的数码，比如（1-75)。全书编写中采用了分章的办法，原则上可以逐章独立地阅读。

本书由同济大学欧天垣及曹叔维两位同志翻译。华东师范大学马国选同志校订。

本书根据1975年英文版翻译，订正了若干印刷上和计算上的错误。

缩写符号表

A_i	初始值、初值
ϑ	飞行航向角、弹道角
δu_l	u 的变分
ΔP	P 的规定改变量
E	魏斯脱拉斯函数
E_j	终值
F	一般函数，偶然为参照面积
$\widetilde{F}$	F 的极值
F_E	喷嘴出口面积
f	一般函数
G_j	一般方程
g	重力加速度
g_0	地球表面的重力加速度
$\tilde{g}$	平均重力加速度
g_j	约束方程
γ_j	一般约束
H	经典理论的哈密顿算子
$H^{(-)}$	庞特里雅金理论的哈密顿算子
H^*	某些问题里 $H^{(-)}$ 的决定性部分
h	高度
h_j	伴随方程
J	待极值化的积分
J^{PQ}	梯度理论中的特殊积分
K_j	常数
L	待极值化积分的被积函数
λ_j	拉格朗日乘子＝伴随变量
λ_{ul}	影响函数
M	$H^{(-)}$ 的极大值
m	质量
μ	质流
μ_p	拉格朗日乘子
ν_r	拉格朗日乘子（常数）
P	待极值化的终值
P^*	待极值化的代换值
p_A	大气压力
p_j	最优的切线方向（欧拉-拉格朗日方程）
$P_n(x)$	切比雪夫多项式
ψ_j	最优的切线方向（哈密顿-约可比方程）
Q_j	一般终止条件
R	停止条件
r	半径
S	J 的极值
$\widetilde{S}$	推力
$\hat{S}$	不等式约束
T	极小飞行时间
t	独立变量，有时为时间
u_l	控制函数
v	速度
W	阻力
$\hat{W}$	阻力项展开的特殊缩写
$\widehat{W}$	加权函数
w_A	排气速度
x_j	状态变量
y_j	一般变量
z_j	一般变量

目　录

第一部分　基本概念

第二部分　间接方法

第三部分 直接方法

第一部分　基 本 概 念

第一章　最优化方法介绍

§1.1　问题的说明

最简单的最优化问题是确定一些独立变量的值，它们能使这些变量的一个给定函数极大化或极小化。这就是说，在最简单的情况下，我们需要确定 t 的值，在这个 t 值时，使

$$i = I(t) = \begin{cases} \text{极大} \\ \text{或极小} \end{cases} \qquad (1)$$

比较复杂一些的最优化问题是要找出变量函数的导数，变量函数及其导数结合成一个函数式，使这个函数的积分成为极值。此时最简单的情况是找出 $y'(t)$ 的值，使

$$J = \int_{t_A}^{t_E} L[t, y(t), y'(t)]dt = \begin{cases} \text{极大} \\ \text{或极小} \end{cases} \qquad (2)$$

且有 $$y(t_A) = A; \quad y(t_E) = E$$

在物理问题中，一般会出现独立变量的几个函数，其中的一些函数必须附带地满足一组常微分方程，如牛顿的运动方程 $\vec{F} = m\vec{a}$。如果把由微分方程所确定的函数记为状态变量 $x_i(t)$，而把自由函数记为控制函数 $u_l(t)$ 的话，那么基本的问题是：确定控制函数 $u_l(t)$，使得

$$P = \int_{t_A}^{t_B} L[t, x_i(t), u_l(t)]dt = \begin{cases} \text{极大，} \\ \text{或极小，} \end{cases} \qquad (3)$$

并在下述约束条件和边界条件下成立。

约束条件：

$$\dot{x}_i = g_i[t, x_i(t), u_l(t)]$$

边界条件：

$$x_j(t_A) = A_j;\quad x_j(t_E) = E_j$$

$$j、\hat{j} = 1, 2, \cdots, m;\quad l = 1, 2, \cdots, k$$

解决象方程（1）这样的一般极值问题，经典方法是通过微分计算；欧拉（Euler）和拉格朗日（Lagrange）对方程（2）这样的最优化问题提出了采用变分计算的方法（见[11]）。

我们将主要介绍方程（3）这样的最优化问题，并讨论以下方法：

（i）庞特里雅金（Pontryagin）极大值原理

（ii）相应的变分法的发展

（iii）梯度法

（iv）贝尔曼（Bellman）“动态规划”。

式（2）这样的问题将是我们讨论最优化方法的出发点，因为在绝大部分最优化问题的较新著作中都采用了变分法这个基本概念。

§1.2　各种最优化方法的特点

现在我们简略地比较一下本书讨论的几种最优化方法。

解一个问题的经典数学方法是经过分析讨论。因而，用变分法来导出方程（2）的解答所满足的分析条件，我们必须分以下四个基本问题：

（i）必要条件的公式表示，

（ii）充分条件的公式表示，

（iii）解的存在性问题，

（iv）解的唯一性问题。

但我们只讨论必要条

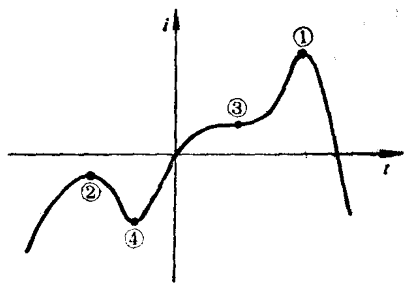

图1-1　$\frac{di}{dt} = 0$ 的点

件，因为在许多情况下，其他问题的解答能够从问题的物理说明中直接得到。

可以从一般的极值计算中导出那些可望从必要条件唯一地得到的信息。必要条件$\frac{\mathrm{d}i}{\mathrm{d}t}=0$给出了图1-1中的点（1）到点（4）的特征。举例来说，我们要求得到的是绝对极大值（1），但我们可能得到局部极大值（2）、稳态点（3），甚至极小值（4）。这里稳态点和极小值可能比局部极大值更容易找出来。

变分法中一个非常有趣的特点是，必要条件可以用二种不同的但是却是等价的方法来表示。为了说明此点，我们注意到，对一已知始点，一般（除奇异情况外）通过适当地选择方程（2）中的$y'(t)$，可以对任何一个终点(t_E, y_E)，确定一个极值。因此J的极值S，表示了一个纯粹的位置函数$S(t_E, y_E)$。于是，我们可以写

$$S=\int \mathrm{d}s=\int\left(\frac{\partial S}{\partial t}\mathrm{d}t+\frac{\partial S}{\partial y}\mathrm{d}y\right)$$

并在我们寻求表示极值的特征条件时，从积分号下S的偏导数和y的微分中推得偏微分方程或常微分方程——根据我们是对极值S的数值，还是对边界点的最佳连接$y(t)$感兴趣而定。二种方法要同时探索，以便得到最简单的方程。正如人们在投影几何里，对于互相对偶的点在平面座标中的位置所做的一样。

虽然，这里建议的二种方法，是通过常微分方程(欧拉方程)或一个偏微分方程（哈密顿-约可比方程）进行的，但在最优化问题（2）的整个过程中，同时都采用。对于带约束最优化问题的分析研究，以（作为最优值必要条件的）常微分方程为主，这对物理解释和数值计算都提供了一个最直接的途径。因此，庞特里雅金极大值原理（由此产生出最优化问题（3）的必要条件的分析推导）主要引入常微分方程作为给出最优值的曲线序列$x_i(t)$的特征，这一点就并不奇怪了。它不同于经典的变分法——其中带有约束的问题当然已作处理——有以下二个方面：

（i）仅对最优化问题（3）或其推广来说，人们第一次区分了控制函数与状态变量，因此只涉及了经典变分法的一部分；

（ii）控制函数可能属于一个闭区域，因此对因变量严格地假定在一个开区域上的经典变分法来说，它被相当地扩大了。

尤其因为大部分物理和技术的最优化问题能用这种方法来处理。对最优化问题（3）的限止，结合状态变量和控制函数的区分，已经导致许多新的研究。关于控制函数限止区域的意义，通过一个简化的例子就会弄清楚。

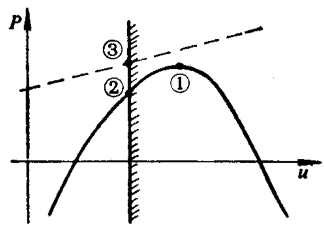

图1-2　边界对极值的影响

如果我们有一个对时间为常数的控制函数和函数 $P(u)$，如图 1-2，则在没有任何限止的情况下，（1）是极大值；当存在图1-2 中所示右侧边界时，（2）是极大值。若 $P(u)$ 如图中虚线这样一条直线，则对 u 不加限止时，不存在有限极大值。在大部分技术问题中，对控制函数的这种限止，是可以理解的。比如把 u 当作可控火箭发动机的质流，显然质流只能在零和由发动机所决定的一个极大值之间变动，因而

$$0 \leqslant u \leqslant u_{max}$$

成立。

在最优化问题（3）的处理中，庞特里雅金极大值原理和经典变分法有以下共同点，给出极值的控制函数 $u_k(t)$，一般能表示为位置座标 $x_j(t)$ 和某些附加函数 $\lambda_s(t)$，$s=1, 2, \cdots, m$ 的形式，后者又必须满足 m 个一阶常微分方程。它们与 x_j 的 m 个微分方程一起组成一个 $2m$ 个微分方程组，对此确定了 m 个初值 $x_j(t_A)=A_j$ 和 m 个终值 $x_j(t_E)=E_j$。因此，如果我们想要得到某个问题的具体极值 $x_j(t)$，那么就得求解一阶非线性常微分方程的边值问题。一般地说，这只有数值解才可能。

从而，我们会问自己，是否能够不通过最优值存在性分析条件

的这种迂回方法，而直接用数值方法求解最优化问题。贝尔曼的动态规划和梯度法已做到了这一点。但是，我们必须认识到，在用数值方法时，我们得到的只是所研究例子的特解，而用庞特里雅金极大值原理和经典变分法来分析条件方程时，能作出关于最优化问题解的一般描述，而不必知道特解或确定数值解。

梯度法是从最优化问题（3）的约束微分方程中某个解开始的；在这里给定了一些我们期望的控制函数。把需要极值化的积分转变为一个需要极值化的终值。从而，问题就集中到所确定的控制函数的变分对终值有什么影响。因此我们推导的是控制函数的一个分步变分。比如说，它规定的终止条件是要从时间上尽快地达到（见图 1-3)，同时需要最优化的终值在所希望的方向上有尽可能大的变动。当规定的终止条件被足够精确地满足，而且需要的终值也已达到一个不能再用控制函数的变分来改进的极值时，步骤就算完成了。此时，实际上我们只是再一次获得了局部极值，且解曲线取决于任选的起始曲线，因为只有一部分对需要极值化的值产生强烈影响的起始曲线改变了形状。另一方面，每一步我们都改善了需要极值化的值，并且看到了解曲线的那个部分对极值化是有意义的。当对于可接受的解曲线有限止时，这一点将会特别重要。

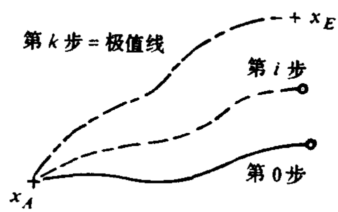

图1-3　在受梯度法影响的一个近似解中的变化

这些性质，和建立该方法时的某些自由度一起，使梯度法对于工程人员具有很大的吸引力。

通过计算一组带有线性化约束的微分方程，可以改进需要最优化的终值，这组微分方程是与一阶微分方程相一致的。后者，比如是从庞特里雅金极大值原理得出的一个最优情况存在的必要条件。

这就允许把梯度法与决定极值 $x_j(t)$ 之边值问题的数值解法联系在一起，后者是在最优化问题的间接分析中发生的。

在我们确定极值时，基本上有三个任务：

（i）满足边界条件 $x_j(t_A)=A_j$；$x_j(t_E)=E_j$，

（ii）满足 $x_j(t)$ 和 $\lambda_s(t)$ 的常微分方程组，

（iii）满足一个极值性条件。

正规解法（变分的经典方法）是计算 $x_j(t)$ 和 $\lambda_s(t)$ 的函数 $u_k(t)$，并求解仅依赖于 $x_j(t)$ 和 $\lambda_s(t)$ 的常微分方程组。在此，终值 $x_j(t_F)$ 对于自由初值 $\lambda_j(t_A)$ 的变分 $\delta\lambda_j$ 的依赖关系被展开为一个级数。

另一种方法，推广的牛顿-拉福森（Newton-Raphson）方法，从级数展开式和微分方程线性化着手，在边界条件必须严格满足的条件下，试图用叠代法来满足微分方程。

梯度法可以从级数展开和将需要极值化的 P 值线性化推出，如果最优化问题（3）（如已述的那样）已被变换，使 P 纯粹是终值的函数 $P\equiv P[t_E, x_j(t_E)]$。因此，梯度法能理解为最优化问题（3）的另两种数值解法的逻辑代用法。

P 值的级数展开式的梯度法推导不仅用于线性化这点上，而且还导致该方法的几种推广形式。如果我们不是截断到级数的线性项，而是截取到二次项的话，那么我们就得到二阶梯度法。如果我们并不考虑全部而只是部分二次项，那么我们得到“Extr.H”方法（Extremum-Hamilton Method；极值-哈密顿方法）。

贝尔曼的“动态规划”，初看起来显得与迄今为止用来解决最优化问题（3）的方法并无直接联系。它是建立在数值解法基础上的。每一种数值解法都是从把连续关系分解为离散过程着手的。让我们首先局限于找出在 x_1，x_2 平面上最优化问题（3）的解曲线，此曲线连接 A、E 点并使 P 成为极大。然后，最基本的数值解法，是在始点和终点之间嵌入可能的最好网格，并且寻找所有可能的连接路径。这样做时，我们获得一条曲线，作为最优曲线的一个近似，它逐段遵循各个格点的方向（图 1-4）。