

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 (北京阜成门内百万庄南街一号)

(北京市书刊出版业营业登记证出字第 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 $850 \times 1168 \frac{1}{32}$ · 印张 $8 \frac{3}{4}$ · 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珍贵的。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。

IV

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	M	$H^{(-)}$ 的极大值
θ	飞行航向角、弹道角	m	质量
δu_i	u 的变分	μ	质流
ΔP	P 的规定改变量	μ_r	拉格朗日乘子
E	魏斯脱拉斯函数	ν_r	拉格朗日乘子 (常数)
E_j	终值	P	待极值化的终值
F	一般函数, 偶然为参照面	P^*	待极值化的代换值
$\tilde{F}$	积	p_A	大气压力
$\bar{F}$	F 的极值	p_j	最优的切线方向 (欧拉-拉格朗日方程)
F_E	喷嘴出口面积	$P_n(x)$	切比雪夫多项式
f	一般函数	ψ_j	最优的切线方向 (哈密顿-约可比方程)
G_j	一般方程	Q_j	一般终止条件
g	重力加速度	R	停止条件
g_0	地球表面的重力加速度	r	半径
$\bar{g}$	平均重力加速度	S	J 的极值
g_j	约束方程	$\tilde{S}$	推力
γ_j	一般约束	$\hat{S}$	不等式约束
H	经典理论的哈密顿算子	T	极小飞行时间
$H^{(-)}$	庞特里雅金理论的哈密顿算子	t	独立变量, 有时为时间
H^*	某些问题里 $H^{(-)}$ 的决定性部分	u_i	控制函数
h	高度	v	速度
h_i	伴随方程	W	阻力
J	待极值化的积分	$\hat{W}$	阻力项展开的特殊缩写
J^{P^0}	梯度理论中的特殊积分	$\tilde{W}$	加权函数
K_j	常数	w_A	排气速度
L	待极值化积分的被积函数	x_j	状态变量
λ_j	拉格朗日乘子 = 伴随变量	y_j	一般变量
λ_{u_i}	影响函数	z_j	一般变量

目 录

第一部分 基本概念

第一章 最优化方法介绍	1
§ 1.1 问题的说明	1
§ 1.2 各种最优化方法的特点	2
§ 1.3 各种方法的相互关系及历史演变	9
第二章 变分法概 概	10
§ 2.1 用卡勒茜奥多立方法解释基本概念	10
2.1.1 卡勒茜奥多立方法	10
2.1.2 欧拉和哈密顿-约可比微分方程	12
2.1.3 横截性	14
2.1.4 正规性	16
2.1.5 从等极值的曲线计算极值线及其逆问题	17
2.1.6 用欧拉和哈密顿-约可比微分方程处理极小值 问题的例子	18
2.1.7 欧特曼-魏斯脱拉斯角条件	22
§ 2.2 y' 的线性被积函数理论	22
2.2.1 问题的说明	22
2.2.2 不依赖于 y' 时极值线特征的建立	23
2.2.3 米勒问题的公式表示	24
2.2.4 探空火箭的最大爬升高度和米勒理论	27
§ 2.3 具有微分方程约束的变分问题	32
2.3.1 变分法基本概念的推广	32
2.3.2 作为约束的微分方程的特点	34
2.3.3 拉格朗日、迈耶和博尔扎问题	36

第二部分 间接方法

第一章 庞特里雅金极大值原理	39
§ 1.1 基本定理	39

VI

1.1.1 问题的说明	39
1.1.2 1.1.1的必要条件	40
1.1.3 附加评论	42
§ 1.2 庞特里雅金理论的定理	46
1.2.1 基本思想	46
1.2.2 庞特里雅金理论的主要定理	48
1.2.3 用庞特里雅金理论处理探空火箭的最大爬升	53
1.2.4 奇异弧	58
§ 1.3 线性时间最优系统	62
1.3.1 线性系统的特性	62
1.3.2 二个特例	65
1.3.3 一般关系	68
§ 1.4 综合问题	68
1.4.1 问题的系统叙述	68
1.4.2 综合问题的一般说明	71
第二章 对近代问题公式表示的变分法的调整	74
§ 2.1 以庞特里雅金方法区分状态变量和控制函数的迈耶和拉格朗日问题	74
2.1.1 迈耶和拉格朗日问题的重新公式表示	74
2.1.2 与庞特里雅金极大值原理的结果条件比较	75
2.1.3 用展开约束来简化问题的原则	77
2.1.4 例子:真空中的最优飞行	79
§ 2.2 由最优值存在性的约束引起的必要条件的简单推导法	83
2.2.1 欧拉方程的拉格朗日推导法	83
2.2.2 用形式上的推广对更一般的问题应用拉格朗日推导法	84
2.2.3 控制函数的简单约束	87
§ 2.3 不等式约束的一般处理	89
2.3.1 不等式约束的公式表示	89
2.3.2 不等式约束出现时的最优化条件	91
2.3.3 受约束部分欧拉方程的一个恒等解	95
2.3.4 带有不等式约束的最优化问题例子	96
§ 2.4 状态变量的跳跃	101

2.4.1 问题的说明	101
2.4.2 状态变量跳跃的条件	101
第三章 非线性常微分方程系统边值问题的数值解	104
§ 3.1 基本概念	104
3.1.1 问题的公式化	104
3.1.2 基本方程	105
§ 3.2 满足微分方程时边界条件的叠代实现	106
3.2.1 系统地变化初始值	107
3.2.2 例子: 真空中的最优飞行	109
3.2.3 关于自由始值的偏导数的精确计算	115
3.2.4 进一步的方法	117
§ 3.3 微分方程的叠代实现	120
3.3.1 基本原理	120
3.3.2 上述方法与确定函数 $f(\bar{z}) = 0$ 之根的牛顿-拉福森方法的类似性	123
3.3.3 附加评论	124

第三部分 直接方法

第一章 一阶梯度法	127
§ 1.1 用梯度法解一般极值问题	127
1.1.1 基本方程	127
1.1.2 火箭分级的最优化——梯度法的一个例子	130
1.1.3 带有约束的一般极值问题	134
1.1.4 梯度法的极限	136
§ 1.2 用梯度法解简单变分问题	137
§ 1.3 用梯度法解带有微分方程约束的最优化问题	138
1.3.1 解带有微分方程约束的最优化问题时的关键方程	138
1.3.2 对各种简单问题的讨论	142
1.3.3 控制函数和状态变量的约束	146
1.3.4 用梯度法处理探空火箭的最大爬升问题	148
1.3.5 对一般问题公式表示的讨论	154
1.3.6 最初轨线的逼近度	159

第二章 一阶梯度法的推广及其有关方法	162
§ 2.1 二阶梯度法	162
2.1.1 二阶梯度法的定义	162
2.1.2 公式的推导	163
§ 2.2 到二阶为止的部分展开法	170
2.2.1 部分展开法的基本公式	170
2.2.2 用部分展开法解一般问题	172
2.2.3 极值-H法(极值-哈密顿法)	174
2.2.4 极值H法的推广	179
§ 2.3 最优值必要条件的数值解与梯度法之间的关系	181
2.3.1 系统的交叉连接	181
2.3.2 不同方法的质量比较	182
第三章 贝尔曼动态规划法	185
§ 3.1 用贝尔曼法解一般极值问题	185
3.1.1 序论	185
3.1.2 贝尔曼法的最简单形式	187
3.1.3 一个初级例题的解	190
3.1.4 贝尔曼法的一般优点及其与系统的网格搜索法的比较	193
§ 3.2 用贝尔曼法解简单的变分问题	195
3.2.1 贝尔曼步骤	195
3.2.2 作为贝尔曼法极限情况的哈密顿-约可比微分方程	198
§ 3.3 用贝尔曼法解具有微分方程约束的最优化问题	200
3.3.1 庞特里雅金问题的处理	200
3.3.2 对一个例题的基本讨论	201
3.3.3 在过渡至极限的例题中, 贝尔曼法与庞特里雅金极大值原理的分析比较	206
3.3.4 结论	209
3.3.5 贝尔曼法情况下的变分可能性	210
§ 3.4 贝尔曼法的数值方面的问题	211
3.4.1 计算工作量的估计	211
3.4.2 用多项式近似来减少维数的问题	211
3.4.3 用贝尔曼法考虑探空火箭的最大爬升高度	216

§ 3.5 有二次性能准则的线性过程.....	224
3.5.1 基本概念.....	224
3.5.2 实例.....	225
3.5.3 用贝尔曼法时的分析解.....	226
附录 例题	230
参考文献	253
I 书籍	253
II 论文	255
III 述评	260
IV 评论	260
中英译名索引	261

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

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

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

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

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 $u_i(t)$ 的话，那么基本的问题是：确定控制函数 $u_i(t)$ ，使得

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

并在下述约束条件和边界条件下成立。

约束条件：

$$\dot{x}_j = g_j[t, x_i(t), u_i(t)]$$

边界条件:

$$x_l(t_1) = A_l; \quad x_l(t_2) = E_l$$

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

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

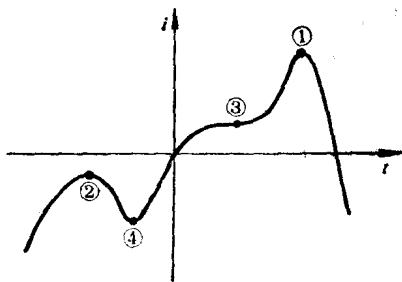


图1-1 $\frac{di}{dt} = 0$ 的点

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 $\frac{di}{dt} = 0$ 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

$$S = \int ds = \int \left(\frac{\partial S}{\partial t} dt + \frac{\partial S}{\partial y} dy \right)$$

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坐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 $x_j(t)$ 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。

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

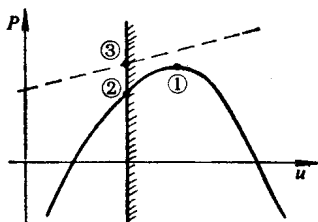


图1-2 边界对极值的影响

$P(u)$ 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

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

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 $u_k(t)$, 一般能表示为位置坐标 $x_j(t)$ 和某些附加函数 $\lambda_s(t)$, $s = 1, 2, \dots, m$ 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 $x_j(t_0) = A_j$ 和 m 个终值 $x_j(t_1) = E_j$ 。因此, 如果我们想要得到某个问题的具体极值 $x_j(t)$, 那么就求得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, 是否能够通过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。

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。

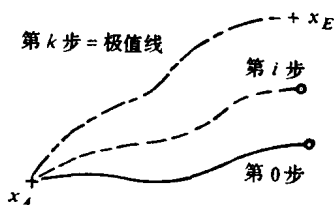


图 1-3 在受梯度法影响的一个近似解中的变化

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 $x_j(t_E)$ 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)。