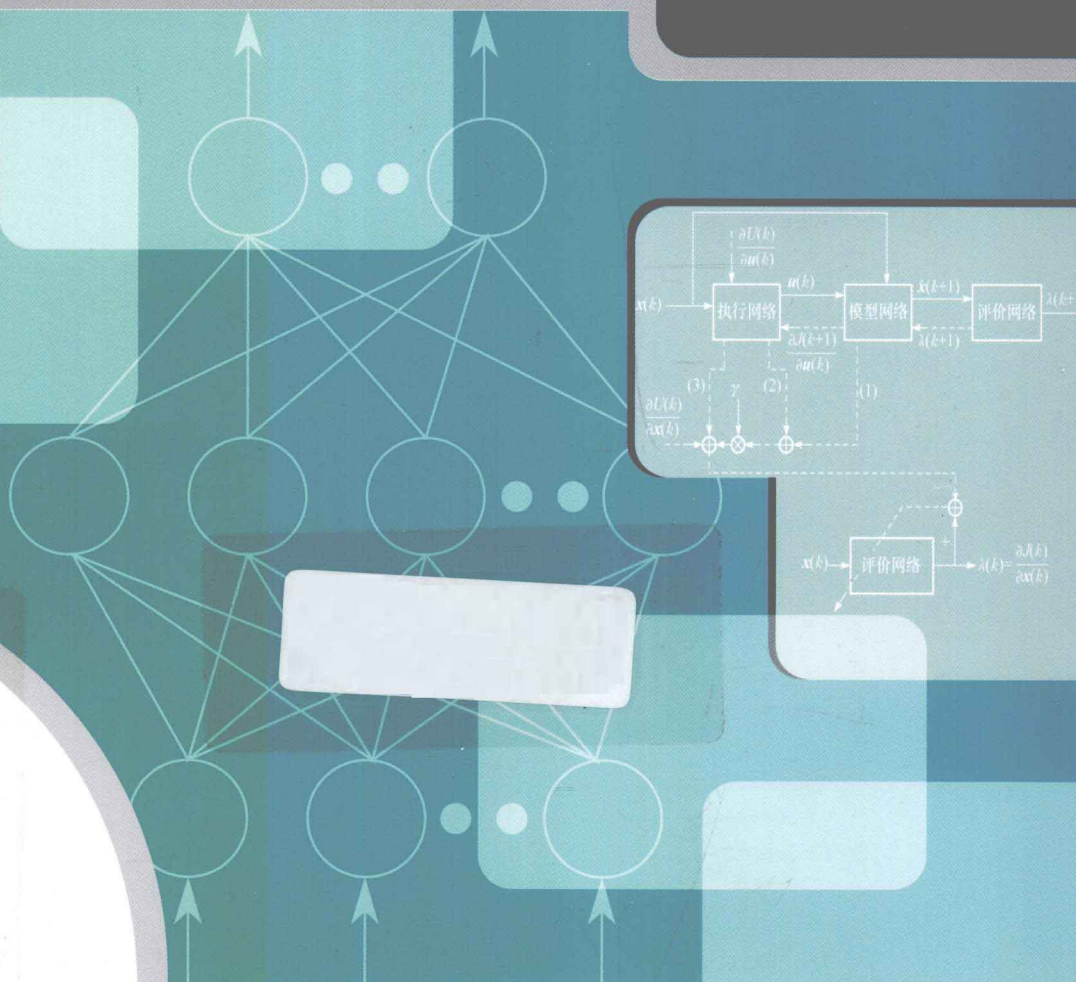


基于自适应动态规划的智能优化控制

林小峰 宋绍剑 宋春宁 著



基于自适应动态规划的 智能优化控制

林小峰 宋绍剑 宋春宁 著

科学出版社

北 京

内 容 简 介

自适应动态规划以传统的最优控制为理论基础,融合人工智能的先进方法,基于强化学习原理,模拟人通过环境反馈进行学习的思路,采用非线性函数拟合方法逼近动态规划的性能指标,有效地解决了动态规划“维数灾”的难题,为高维复杂非线性系统的最优控制提供了一种切实可行的理论和方法。自适应动态规划体现了人工智能和控制领域的一个重要发展方向。

本书对基于自适应动态规划的智能优化控制进行了较为详尽的分析和论述,主要内容包括:动态规划、强化学习理论、自适应动态规划的基本原理、自适应动态规划的基本结构、自适应动态规划的主要算法和实现技术、自适应动态规划的应用实例,以及自适应动态规划的发展趋势。

本书可供从事智能控制、优化控制、计算智能、人工智能与智能信息处理的科技人员使用,也可供高等院校有关专业师生参考。

图书在版编目(CIP)数据

基于自适应动态规划的智能优化控制/ 林小峰,宋绍剑,宋春宁著. —北京:科学出版社,2013

ISBN 978-7-03-036433-3

I. ①基… II. ①林… ②宋… ③宋… III. ①自适应-控制-智能控制
IV. ①TP273

中国版本图书馆 CIP 数据核字(2013)第 008083 号

责任编辑:李岚峰 张丽花/ 责任校对:刘小梅
责任印制:闫 磊 / 封面设计:迷底书装

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

新科印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2013 年 2 月第 一 版 开本:720×1000 B5

2013 年 2 月第一次印刷 印张:14

字数:273 000

定价:56.00 元

(如有印装质量问题,我社负责调换)

前 言

自适应动态规划(adaptive/approximate dynamic programming, ADP)是人工智能和控制技术发展交汇形成的。一方面,在控制领域,Bellman 提出了动态规划(dynamic programming, DP),实现了有效求解非线性系统的最优控制问题。工程技术以及经济领域中的许多问题都可以表述为目标函数的极小化或极大化问题,如误差极小化问题、能量极小化问题和效益极大化问题等。理论上讲,动态规划方法可以用来解决这些问题。动态规划最优控制理论积累了大量成果,并已应用于实际工业生产过程。但动态规划在实际应用中的缺点也非常明显,其时间上向后走的算法会产生巨大的计算量,使得该算法只适用于小规模、非线性系统的最优控制问题。另一方面,在人工智能领域,强化学习、人工神经网络、模糊系统、演化计算等方法的发展和丰富,带来了许多求解非线性系统优化问题的解决思路和技术方法。而自适应动态规划则以传统的最优控制为理论基础,融合人工智能的先进方法,提出了大规模复杂非线性系统优化控制问题的解决方法。自适应动态规划由美国学者 Paul J. Werbos 首先提出,其本质是基于强化学习原理,模拟人通过环境反馈进行学习的思路,是一种非常接近人脑智能的方法。自适应动态规划采用非线性函数拟合方法逼近动态规划的性能指标,有效地解决了动态规划“维数灾”的难题,为高维复杂非线性系统的最优控制提供了一种切实可行的解决方案。自适应动态规划体现了人工智能和控制领域的一个重要发展方向。

我们自 2006 年开始了基于自适应动态规划的智能优化控制的研究工作。2008 年 2 月至 2009 年 2 月,林小峰教授在美国伊利诺伊大学芝加哥分校(The University of Illinois at Chicago)进行访问研究,开展基于自适应动态规划的智能优化控制的研究工作,并得到国际知名的自适应动态规划学者 Derong Liu 的指导和帮助。在美国期间,林小峰教授专程到亚利桑那州立大学(Arizona State University),与国际知名的自适应动态规划学者 Jennie Si 进行学术讨论,拓宽了研究思路。我们数次出席与自适应动态规划相关的国际会议,多次聆听自适应动态规划的创始人(美国国家自然科学基金委员会 Paul J. Werbos 博士)的报告,这给我们很大的启迪,使我们进一步明确了自适应动态规划的研究方向。在研究过程中,我们得到了国内许多专家和学者的热情指导和具体帮助,他们是:中国科学院自动化研究所刘德荣、易建强、赵冬斌研究员,魏庆来副研究员;东北大学张化光教授,罗艳红副教授;清华大学陆超副教授。在此,作者谨向他们表示衷心的感谢。

本书是作者在承担的国家自然科学基金项目(60964002)、国家自然科学基金重点项目(61034002)、广西自然科学基金项目(2011GXNSFC018017)、中国科学院复杂系统与智能科学重点实验室开放基金项目(20070101)、广西科学研究与技术开发计划项目(桂科攻 11107004-2)、广西理工科学实验中心重点项目(LGZX201106)研究成果的基础上撰写的,也是作者近年来潜心学习和研究基于自适应动态规划的智能优化控制理论、方法、应用成果的一个总结。

本书结合我们近年来在自适应动态规划领域的研究成果,对自适应动态规划的各个方面进行较系统的阐述和讨论。全书共5章,第1章介绍动态规划的基本概念、基本原理以及基于动态规划的最优控制理论;第2章介绍强化学习的基本概念、学科基础、形成与发展及其主要算法,第1、2章是自适应动态规划的来源和基础;第3章介绍自适应动态规划的基本原理、基本结构、主要算法、研究概况及其发展趋势;第4章介绍自适应动态规划的实现技术以及ADP主要算法的推导和实现步骤;第5章介绍自适应动态规划在制糖、水泥、电力等行业的应用实例。

本书的第1~4章、第5章第1~4节由林小峰教授撰写和统稿,第5章第5节由宋春宁副教授撰写和统稿,第5章第6节由宋绍剑教授撰写和统稿。在本书的撰写、编辑、修改及参考文献整理的过程中,研究生张志刚、段栓峰、张衡、滕千里、谢树平、黄元君、孙欣、周显俊、杨继云、黄际乐、曹怒云、余乾仲、彭奎、周晓华等做了大量的具体工作,作者对他们的辛勤工作表示由衷的感谢。同时感谢科学出版社的张濮、李岚峰、张丽花编辑对本书出版所做的认真、细致的工作。

自适应动态规划是一个多学科交叉的研究方向,其理论及应用均有许多问题尚待深入研究。由于作者水平有限,书中难免存在疏漏和不足之处,恳请同行、读者批评指正。

作 者

2012年9月

目 录

前言

第 1 章 动态规划	1
1.1 动态规划的基本方法	1
1.1.1 多级决策过程	1
1.1.2 最优性原理	4
1.1.3 动态规划的基本递推方程	5
1.2 离散系统的动态规划	10
1.2.1 离散最优问题的动态规划解	10
1.2.2 动态规划的数值算法	14
1.2.3 离散系统动态规划的特点	17
1.3 连续动态规划	17
1.3.1 连续时间系统的最优控制问题	18
1.3.2 哈密顿-雅可比-贝尔曼方程	18
1.3.3 连续动态规划的基本方程	20
参考文献	25
第 2 章 强化学习	26
2.1 强化学习的基本概念	26
2.2 强化学习的形成和发展	28
2.2.1 强化学习的学科基础	28
2.2.2 强化学习的形成阶段	30
2.2.3 强化学习的发展阶段	30
2.3 强化学习的主要算法	32
2.3.1 蒙特卡罗算法	32
2.3.2 时序差分算法	37
2.3.3 自适应启发评价算法	44
2.3.4 Q-学习算法	51
参考文献	56
第 3 章 自适应动态规划	60
3.1 自适应动态规划的研究概况	60
3.2 大脑智能与优化	64

3.2.1	神经系统	64
3.2.2	大脑的分区功能与统一协调	65
3.2.3	大脑是智能控制器	65
3.2.4	大脑的渐进学习	67
3.2.5	强化学习与最优化	69
3.2.6	大脑优化的两个问题	70
3.2.7	仿脑智能优化控制的实现	70
3.3	自适应动态规划的基本原理	71
3.3.1	自适应动态规划的基本思想	71
3.3.2	动态规划的前向、后向算法	72
3.3.3	自适应动态规划迭代算法	74
3.3.4	函数近似结构	76
3.3.5	自适应动态规划的优点	77
3.4	自适应动态规划的基本结构与分类	77
3.4.1	启发式动态规划	77
3.4.2	双启发式动态规划	79
3.4.3	执行依赖启发式动态规划	80
3.4.4	执行依赖双启发式动态规划	81
3.5	自适应动态规划的发展趋势	81
3.5.1	三代大脑智能 ADP 模型	81
3.5.2	自适应动态规划的研究动态	85
	参考文献	89
第4章	基于 BP 网络的自适应动态规划算法及实现	96
4.1	基于 BP 网络的 HDP 算法及实现	96
4.1.1	HDP 的结构和基本原理	96
4.1.2	模型网络	97
4.1.3	评价网络	99
4.1.4	执行网络	100
4.1.5	相关参数的选择	105
4.2	基于 BP 网络的 DHP 算法及实现	105
4.2.1	DHP 的结构和基本原理	105
4.2.2	模型网络	106
4.2.3	评价网络	107
4.2.4	执行网络	114
4.3	基于 BP 网络的 ADHDP 算法及实现	118

4.3.1	ADHDP 结构和基本原理	118
4.3.2	评价网络	120
4.3.3	执行网络	121
	参考文献	124
第 5 章	自适应动态规划的应用	126
5.1	水泥预分解窑系统的 HDP 控制	126
5.1.1	水泥预分解窑系统	126
5.1.2	水泥预分解窑的神经网络建模	130
5.1.3	水泥预分解窑 HDP 算法	134
5.1.4	HDP 训练策略及控制系统仿真	138
5.1.5	小结	141
5.2	糖厂澄清过程 pH 值的 DHP 控制	141
5.2.1	糖厂澄清过程工艺流程	141
5.2.2	制糖澄清过程系统建模	145
5.2.3	制糖澄清过程 DHP 算法	148
5.2.4	DHP 算法控制清汁 pH 值	151
5.2.5	小结	154
5.3	同步发电机励磁系统的 ADP 控制	155
5.3.1	同步发电机及其励磁系统	155
5.3.2	基于 ADHDP 的励磁控制	157
5.3.3	基于多目标 ADHDP 的励磁控制	164
5.4	基于 DHP 方法的发电商竞价策略模型	173
5.4.1	竞价策略研究	173
5.4.2	发电商市场竞价策略模型	175
5.4.3	仿真试验及分析	179
5.4.4	小结	182
5.5	ADHDP 在高压直流输电整流控制中的应用	182
5.5.1	高压直流输电系统	182
5.5.2	ADHDP 整流控制器设计	189
5.5.3	整流侧定电流、逆变侧定电压控制模型仿真	193
5.5.4	小结	201
5.6	三容液位系统的 ADHDP 控制	201
5.6.1	三容液位系统简介及其模型	201
5.6.2	三容液位系统 ADHDP 控制器设计	204
5.6.3	控制仿真及分析	206

5.6.4 实时控制	209
5.6.5 小结	213
参考文献	213

第 1 章 动态规划

动态规划(Dynamic Programming, DP)是运筹学的一个分支,是求解决策过程最优化的数学方法。20 世纪 50 年代美国数学家贝尔曼等在研究多级决策过程的优化问题时,提出了著名的最优性原理,创立了解决这一类过程优化问题的新方法——动态规划。

动态规划的核心是贝尔曼最优性原理,它首先将一个多级(步)决策问题转化为一组单级决策问题,然后从最后一级状态开始逆向递推到初始级状态为止,即在求解多级决策问题时,要从末端开始到始端为止,逆向递推^[1]。动态规划在控制理论上的重要性表现为:对于离散控制系统可用得到某些理论结果建立起迭代计算程序;对于连续控制系统,除了可以得到某些新的理论结果外,还可建立起与变分法和极小值原理的联系^[2]。

尽管动态规划存在着没有构造模型的通用方法及用数值方法求解时存在“维数灾(Curse of Dimensionality)”等问题,但时至今日,由于计算技术和计算方法的迅速发展,动态规划仍在经济管理、生产调度、资源分配、设备更新、优化设计、最优控制、信息处理和模式识别等领域中获得广泛的应用^[3]。本章主要介绍动态规划在控制理论和控制工程中解决动态系统最优控制方面的理论。

1.1 动态规划的基本方法

1.1.1 多级决策过程

所谓多级决策过程(Multi-step Decision Process),是指将一个过程按时间或空间顺序分为若干级,即分为若干阶段,然后给每一级做出“决策”(当一个阶段的状态确定后,可以做出各种选择从而演变到下一阶段的某个状态,这种选择手段称为决策),以使整个过程取得最优的效果,即多次的决策最终要构成一个最优控制策略(决策组成的序列)。多级决策过程如图 1.1 所示。图中,对于中间的任意级,例如,第 $k+1$ 级做出相应的“决策”(或控制) u_k 后,才能确定该级输入状态与输出状态之间的关系,即从 x_k 变化到 x_{k+1} 的状态转移规律。显然,如果选择好每一级的“决策”(或控制) u_k ($k=1,2,\dots,N-1$),则整个过程的状态转移规律,即从 x_0 经 x_k 一直到 x_N ,也就完全被确定了。

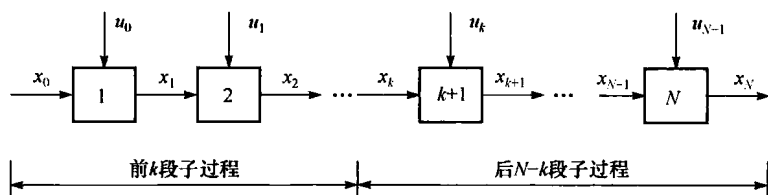


图 1.1 多级决策过程示意图

在多级决策过程中,当某阶段的状态给定时,这个阶段以后过程的演变与该阶段以前各阶段的状态无关,例如,在图 1.1 中,第 $k+1$ 级的输出 x_{k+1} 只由输入 x_k 和决策 u_k 决定,这种特有性质称为无后效性。

为了进一步阐明多级决策过程及动态规划的基本思想和方法,考察一个最短路线问题^[1]。

设由 A 地至 E 地的路线如图 1.2 所示,全程分为 4 级,且每个中间站又有若干个可供选择的路线,各站之间的距离已标注在图中。现要求从 A 地出发,经过 B、C、D 三地到达 E 地,且使所选择的路线最短。由图可见,由 A 地到终点 E 地有不同的路线,沿路线 A 地与 E 地之间的路程不同,为使 A 地与 E 地之间的路程最短,在路线的前 3 级要做出 3 次决策。也就是说,第 1 级由 A 地到 $B(B_1, B_2)$ 地,要求选择一条路线,使 A 地与 B 地之间路程最短,称为一级决策过程;第 2 级由 $B(B_1, B_2)$ 地到 $C(C_1, C_2)$ 地,要求选择一条路线,使 ABC 路程最短,称为二级决策过程;第 3 级由 $C(C_1, C_2)$ 地到 $D(D_1, D_2)$ 地,要求选择一条路线,使 ABCD 路程最短,称为三级决策过程;第 4 级由 $D(D_1, D_2)$ 地到 E 地,只有一条路线选择,所以本级无决策问题。

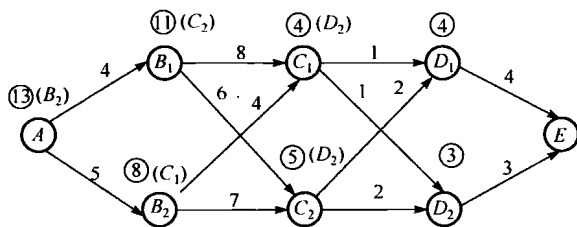


图 1.2 A 地至 E 地路线图

对于这样一个级数 $N=4$,每级有两个决策的简单问题,易知,共有 $2^{N-1}=8$ 种可能的路线。将每级路程加起来,就可得到每种路线的总路程。选择其中最小者,便可决定路程最短的路线。用这种“穷举法”确定最优路线,需要算出所有可能路线的路程。而每条路线要做 3 次加法,总共需相加 24 次。一般说来,如果全程有 N 级,就需相加 $(N-1)2^{N-1}$ 次。

另一种可确定最优路线的方法是动态规划法。动态规划法是从终点开始,以路程最短为目标,逐级向前逆推,依次算出各地到终点的路程最优值,并据此确定出最优路线。应用动态规划法求解此类问题可大大减少运算工作量,并明显提高运算速度。

设 N 为多级决策过程的级数; x 表示在任一级所处的位置,称为状态变量(State Variable); $S_N(x)$ 为决策变量(Decision Variable),表示状态 x 后所选取的下一点; $J_N(x)$ 表示从状态 x 到终点 E 的 N 级过程的最短距离,称为性能指标函数(Objective Function); $d(x, S_N)$ 表示从 x 点到 S_N 点之间的距离。对于图 1.2 路线问题,从最后一级开始计算。

1. 第 4 级(D 级)

由于本级从 D_1 到 E 及从 D_2 到 E 都只有一种可能,所以本级无决策问题, $D(D_1, D_2)$ 到 E 的距离可表示为

$$J_4(D_1) = d(D_1, E) = 4$$

$$J_4(D_2) = d(D_2, E) = 3$$

将 $D(D_1, D_2)$ 至 E 的距离数值标注于图 1.2 中,数字用圆圈圈起来,表示本级指标函数,即在 D_1 和 D_2 点处分别标注④和③。数字旁括号内填写相应的决策变量(本级无)。

2. 第 3 级(C 级)

本级决策有两种选择,每种选择有两条可能的路线。

若从 C_1 出发,可达 D_1 ,也可达 D_2 ,所以

$$J_3(C_1) = \min \begin{cases} d(C_1, D_1) + J_4(D_1) \\ d(C_1, D_2) + J_4(D_2) \end{cases} = \min \begin{cases} 1+4 \\ 1+3 \end{cases} = 4$$

由此可见, C_1 至 E 的最短距离为 4,路线为 $C_1 D_2 E$,决策变量为 $S_3(C_1) = D_2$ 。因而在图 1.2 中的 C_1 点处,标注④(D_2)。

若从 C_2 出发,则有

$$J_3(C_2) = \min \begin{cases} d(C_2, D_1) + J_4(D_1) \\ d(C_2, D_2) + J_4(D_2) \end{cases} = \min \begin{cases} 2+4 \\ 2+3 \end{cases} = 5$$

由此可见, C_2 至 E 的最短距离为 5,线路为 $C_2 D_2 E$,决策变量 $S_3(C_2) = D_2$ 。因而在图 1.2 中的 C_2 点处,标注⑤(D_2)。

3. 第 2 级(B 级)

本级有两种选择,每种选择有两种可能的路线

$$J_2(B_1) = \min \left\{ \begin{array}{l} d(B_1, C_1) + J_3(C_1) \\ d(B_1, C_2) + J_3(C_2) \end{array} \right\} = \min \left\{ \begin{array}{l} 8+4 \\ 6+5 \end{array} \right\} = 11$$

$$S_2(B_1) = C_2$$

$$J_2(B_2) = \min \left\{ \begin{array}{l} d(B_2, C_1) + J_3(C_1) \\ d(B_2, C_2) + J_3(C_2) \end{array} \right\} = \min \left\{ \begin{array}{l} 4+4 \\ 7+5 \end{array} \right\} = 8$$

$$S_2(B_2) = C_1$$

因而在图 1.2 中 B_1 和 B_2 点处,分别标注⑪(C_2)和⑧(C_1)。

4. 第 1 级(A 级)

本级决策是唯一的,它有两种可能的路线

$$J_1(A) = \min \left\{ \begin{array}{l} d(A, B_1) + J_2(B_1) \\ d(A, B_2) + J_2(B_2) \end{array} \right\} = \min \left\{ \begin{array}{l} 4+11 \\ 5+8 \end{array} \right\} = 13$$

$$S_1(A) = B_2$$

在图 1.2 中 A 点处,标注⑬(B_2)。

因此从 A 地开始,可顺序确定最短路线为 $AB_2C_1D_2E$,最短距离为

$$J^* = J_1(A) = 13$$

对于本题,采用的递推方程的一般形式为

$$J_N(x) = \min_{S_N(x)} \{ d[x, S_N(x)] + J_{N+1}[S_N(x)] \} \quad (N=3, 2, 1) \quad (1.1)$$

及

$$J_1(x) = d(x, E) \quad (1.2)$$

与穷举法相比,动态规划法确实使计算量大为减少。对于 N 级决策问题,只需相加 $4(N-2)+2$ 次。

由本例的分析过程可知,一个 N 级最优过程,不论第一级决策如何,其余 $N-1$ 级决策过程至少必须依据第一级决策所形成的状态组成一个 $N-1$ 级最优过程。在此基础上选择第一级决策,必可使总的 N 级过程最优。在多级决策过程中,这种递推思想的核心就是贝尔曼提出的最优性原理(Principle of Optimality)。

1.1.2 最优性原理

贝尔曼 1957 年指出,多级决策过程的最优策略具有这样的性质:不论初始状态和初始决策如何,当把其中的任何一级和状态再作为初始级和初始状态时,其余的决策对此必定也是一个最优策略^[5]。

最优性原理的具体描述为若有一个初态为 $x(0)$ 的 N 级决策过程,其最优策略为 $\{u(0), u(1), \dots, u(N-1)\}$ 。那么,对以 $x(1)$ 为初态的 $N-1$ 级决策过程来说,决策集合 $\{u(1), u(2), \dots, u(N-1)\}$ 必定是最优策略。

证明 设策略或决策序列

$$u^* = \{u^*(0), u^*(1), \dots, u^*(N-1)\}$$

是使代价函数 J 最小的最优策略,相应的最小代价为

$$\begin{aligned} J^*[x(0)] &= J[x(0), u^*] \\ &= J[x(0), u^*(0), u^*(1), \dots, u^*(N-1)] \\ &= \min_{u(k)} \sum_{k=0}^{N-1} U[x(k), u(k), k] \end{aligned}$$

反设在 $k=r, r+1, \dots, N-1$ 区间内, $u^*(k)$ 不是最优决策序列,则必存在另一个决策序列 $\bar{u}(k), k=r, r+1, \dots, N-1$, 使得

$$J[x(r), \bar{u}(r), \bar{u}(r+1), \dots, \bar{u}(N-1)] < J[x(r), u^*(r), u^*(r+1), \dots, u^*(N-1)]$$

因而,有两个决策序列

$$\begin{aligned} &u^*(k), k=0, 1, \dots, N-1 \\ u^{**}(k) &= \begin{cases} u^*(k), & k \in [0, r-1] \\ \bar{u}(k), & k \in [r, N-1] \end{cases} \end{aligned}$$

使得

$$J^* > J^{**}$$

式中

$$\begin{aligned} J^* &= J[x(0), u^*(0), u^*(1), \dots, u^*(N-1)] \\ &= U[x(0), u^*(0), 0] + \dots + U[x(r-1), u^*(r-1), r-1] \\ &\quad + U[x(r), u^*(r), r] + \dots + U[x(N-1), u^*(N-1), N-1] \\ J^{**} &= J[x(0), u^*(0), \dots, u^*(r-1); \bar{u}(r), \bar{u}(r+1), \dots, \bar{u}(N-1)] \\ &= U[x(0), u^*(0), 0] + \dots + U[x(r-1), u^*(r-1), r-1] \\ &\quad + U[x(r), \bar{u}(r), r] + \dots + U[x(N-1), \bar{u}(N-1), N-1] \end{aligned}$$

式中, $U[x(k), u(k), k]$ 为由 $x(k)$ 出发采取的决策 $u(k)$ 使系统状态转移到下一个状态 $x(k+1)$ 的阶段指标值。

上述结果与 J^* 为最小代价的假设相矛盾。因此,反设不成立,最优性原理得证。

对于连续系统,最优性原理可以表述为:若对于初始时刻 t_0 和初始状态 $x(t_0)$, $u^*(t)$ 和 $x^*(t)$ 是某系统的最优控制和最优轨线,则对于时刻 t_1 ($t_1 > t_0$) 和相应状态 $x(t_1)$ 来说, $u^*(t), t \in [t_1, t_f]$ 和 $x^*(t), t \in [t_1, t_f]$ 仍是该系统往后的最优控制和最优轨线。即最优控制和最优轨线的一部分(从终端起的后一部分)必是当时状态为初态之下的最优控制和最优轨线。

最优性原理为解决多级决策过程的寻优问题提供了简便而有效的方法,该原理可将一个多级最优决策问题化为多个单级最优决策问题,并为导出动态规划的基本递推方程提供了理论基础。

1.1.3 动态规划的基本递推方程

问题 1-1^[1,2,4] 设 N 级决策过程的状态方程为

$$x(k+1) = f[x(k), u(k), k], \quad x(0) = x_0 \quad (1.3)$$

式中,状态约束 $x(k) \in X \subset \mathbf{R}^n$,控制(决策)约束 $u(k) \in \Omega \subset \mathbf{R}^m, k=0,1,\dots,N-1$ 。求容许控制(决策)序列 $u^*(k), k=0,1,\dots,N-1$,使代价函数(或性能指标)

$$J[x(0)] = \sum_{k=0}^{N-1} U[x(k), u(k), k] \quad (1.4)$$

最小。

在问题 1-1 中, k 表示 N 级决策过程中的阶段变量, $x(k)$ 表示第 $k+1$ 级的初始状态, $u(k)$ 表示第 $k+1$ 级所采用的控制(决策)向量, $U[x(k), u(k), k]$ 表示由 $x(k)$ 出发采取的决策 $u(k)$ 使系统状态转移到下一个状态 $x(k+1)$ 的阶段指标值。为了强调 J 是控制 u 的泛函,因此式(1.4)可表示为

$$J[x(0), u] = \sum_{k=0}^{N-1} U[x(k), u(k), k] \quad (1.5)$$

一般情况下,始于任何状态 $x(k)$ 的代价可记为 $J[x(k), u]$, 其中

$$u = \{u(k), u(k+1), \dots, u(N-1)\}$$

$J[x(k), u]$ 表示任意状态 $x(k)$ 在任意策略 u 作用下,变化到过程终点为止时所产生的代价。因此,始自 $x(k)$ 的最优代价为

$$J^*[x(k)] = J[x(k), u^*]$$

式中, u^* 表示最优策略。对于给定问题,当 $x(k)$ 固定时, u^* 是确定的,因此最优代价 $J^*[x(k)]$ 仅是初始状态 $x(k)$ 的函数,常记为 $J[x(k)]$ 。

为了求出问题 1-1 的最小代价 $J^*[x(0), 0]$, 根据动态规划的解题思想,可讨论求 $J^*[x(k), k]$ 最小的问题,其中将 $J^*[x(0)]$ 写为 $J^*[x(0), 0]$, 将 $J^*[x(k)]$ 写为 $J^*[x(k), k]$ 是为了强调 J 对初始时刻的依赖。研究如下问题

$$\min_{u \in \Omega} J[x(k), k] = \sum_{i=k}^{N-1} U[x(i), u(i), i] \quad (1.6)$$

式中, $x(k)$ 固定。状态方程约束为

$$x(i+1) = f[x(i), u(i), i], i=k, k+1, \dots, N-1 \quad (1.7)$$

式中

$$\begin{aligned} x(k) &\in X \subset \mathbf{R}^n \\ k &= 0, 1, \dots, N-1 \\ u(k) &\in \Omega \subset \mathbf{R}^m \end{aligned}$$

始自第 k 级任意容许状态 $x(k)$ 的最小代价为

$$\begin{aligned} J^*[x(k), k] &= \min_{\{u(k), u(k+1), \dots, u(N-1)\} \in \Omega} \left\{ \sum_{i=k}^{N-1} U[x(i), u(i), i] \right\} \\ &= \min_{\{u(k), u(k+1), \dots, u(N-1)\} \in \Omega} \left\{ U[x(k), u(k), k] + \sum_{i=k+1}^{N-1} U[x(i), u(i), i] \right\} \end{aligned} \quad (1.8)$$

式中,第一部分是第 k 级内所付出的代价,第二部分是从第 $k+1$ 级到第 N 级的代价和。将式(1.8)中求最小的运算也分解为两部分:在本级决策 $u(k)$ 下求最小,以及在剩余决策序列 $\{u(k+1), u(k+2), \dots, u(N-1)\}$ 下求最小。于是,式(1.8)可写为

$$J^*[x(k), k] = \min_{u(k) \in \Omega} \min_{\{u(k+1), \dots, u(N-1)\} \in \Omega} \left\{ U[x(k), u(k), k] + \sum_{j=k+1}^{N-1} U[x(i), u(i), i] \right\} \quad (1.9)$$

式中大括号内的第一项仅取决于 $u(k)$,而与 $u(i), i=k+1, k+2, \dots, N-1$ 无关,因此对 $u(i)$ 取极小没有意义。大括号内的第二项,当 $x(k+1)$ 固定时,其值仅取决于 $u(i), i=k+1, k+2, \dots, N-1$,而与 $u(k)$ 没有直接关系,但是 $u(k)$ 通过状态方程式(1.7)决定 $x(k+1)$,从而影响该项的值。于是,式(1.9)可写为

$$J^*[x(k), k] = \min_{u(k) \in \Omega} \left\{ U[x(k), u(k), k] + \min_{\{u(k+1), \dots, u(N-1)\} \in \Omega} \sum_{i=k+1}^{N-1} U[x(i), u(i), i] \right\} \quad (1.10)$$

根据最优性原理和状态方程式(1.3),如下关系式成立:

$$\begin{aligned} J^*[x(k+1), k+1] &= J^*[f(x(k), u(k), k), k+1] \\ &= \min_{\{u(k+1), \dots, u(N-1)\} \in \Omega} \sum_{i=k+1}^{N-1} U[x(i), u(i), i] \end{aligned} \quad (1.11)$$

将式(1.11)代入式(1.10),得动态规划的基本递推方程

$$J^*[x(k), k] = \min_{u(k) \in \Omega} \{ U[x(k), u(k), k] + J^*[x(k+1), k+1] \} \quad (k=0, 1, \dots, N-1) \quad (1.12)$$

或者

$$J^*[x(k), k] = \min_{u(k) \in \Omega} \{ U[x(k), u(k), k] + J^*[f(x(k), u(k), k), k+1] \} \quad (k=0, 1, \dots, N-1) \quad (1.13)$$

上述递推关系从过程的最后一级开始,逐级逆向递推。由式(1.12),令 $k=N-1$ 得

$$J^*[x(N-1), N-1] = \min_{u(N-1) \in \Omega} \{ U[x(N-1), u(N-1), N-1] + J^*[x(N), N] \} \quad (1.14)$$

式中, $J^*[x(N), N]$ 表示代价函数 J 中的末值项。对于问题 1-1 以及相应的嵌入式(1.6)代价函数中没有末值项,应取

$$J^*[x(N), N] = 0$$

于是,式(1.14)写为

$$J^*[x(N-1), N-1] = \min_{u(N-1) \in \Omega} U[x(N-1), u(N-1), N-1] \quad (1.15)$$

式(1.15)仅是函数 $U[x(N-1), u(N-1), N-1]$ 对控制 $u(N-1) \in \Omega$ 的单级最

优化问题,而不再是式(1.6)那样的多级优化问题,变得易于求解。求解时,可以对所有的 $x(N-1) \in X$, 求解式(1.15),以便得到 $J^*[x(N-1), N-1]$ 。然后由递推方程(1.12)逆向逐级递推,求出 $J^*[x(N-2), N-2], \dots, J^*[x(1), 1], J^*[x(0), 0]$ 。这最后一步的递推解 $J^*[x(0), 0]$ 以及得到的最优策略 $\{u^*(0), u^*(1), \dots, u^*(N-1)\}$ 正是问题 1-1 要求的解。

例 1-1 已知离散系统状态方程为

$$x(k+1) = x(k) + u(k), x(4) = 2$$

性能指标为

$$J = \sum_{k=0}^3 [x^3(k) + u(k)x(k) + u^2(k)x(k)]$$

式中,控制 $u(k)$ 限取 +1 或 -1,各级代价要求非负。试求最优控制序列 $u^*(k)$ 和最优曲线序列 $x^*(k), k=0, 1, 2, 3$ 。

解 本例为 $N=4$ 级最优决策问题。

(1) 令 $k=3$

$$\begin{aligned} x(4) &= x(3) + u(3) = 2, x(3) = 2 - u(3) \\ J^*[3] &= \min_{u(3) \in \{1, -1\}} \{x^3(3) + u(3)x(3) + u^2(3)x(3)\} \\ &= \min_{u(3) \in \{1, -1\}} \{8 - 10u(3) + 7u^2(3) - 2u^3(3)\} \\ &= \begin{cases} 3, u(3) = +1 \\ 27, u(3) = -1 \end{cases} \end{aligned}$$

故

$$u^*(3) = 1, x^*(3) = 2 - u^*(3) = 1, J^*(3) = 3$$

(2) 令 $k=2$

$$\begin{aligned} x(3) &= x(2) + u(2) = 1, x(2) = 1 - u(2) \\ J^*[2] &= \min_{u(2) \in \{1, -1\}} \{x^3(2) + u(2)x(2) + u^2(2)x(2)\} + J^*[3] \\ &= \min_{u(2) \in \{1, -1\}} \{1 - 2u(2) + 3u^2(2) - 2u^3(2)\} + 3 \\ &= \begin{cases} 3, u(2) = +1 \\ 11, u(2) = -1 \end{cases} \end{aligned}$$

故

$$u^*(2) = 1, x^*(2) = 1 - u^*(2) = 0, J^*(2) = 3$$

(3) 令 $k=1$

$$\begin{aligned} x(2) &= x(1) + u(1) = 0, x(1) = -u(1) \\ J^*[1] &= \min_{u(1) \in \{1, -1\}} \{x^3(1) + u(1)x(1) + u^2(1)x(1)\} + J^*[2] \\ &= \min_{u(1) \in \{1, -1\}} \{-2u^3(1) - u^2(1)\} + 3 \end{aligned}$$