复杂非线性系统的自适应优化控制

罗艳红　张化光　崔黎黎◎著

科学出版社

复杂非线性系统的自适应优化控制

罗艳红　张化光　崔黎黎　著

科学出版社
北京

内 容 简 介

本书系统地研究了基于自适应动态规划方法的复杂非线性系统的自适应迭代最优控制理论和相关的应用问题. 全书共分为8章, 分别针对离散时间非线性系统和连续时间非线性系统的最优镇定控制和最优跟踪控制问题进行了深入的探讨. 第1章介绍了自适应动态规划的基本理论和发展现状; 第2章研究了执行器死区约束下的非线性系统自适应跟踪控制问题; 第3~6章详细研究了几类离散非线性系统的自适应最优镇定和最优跟踪控制问题; 第7、8章分别基于神经网络和模糊模型深入研究了两类连续时间非仿射非线性系统的鲁棒自适应评价设计问题.

本书可作为高等学校自动化、电气工程及其自动化、测控技术等相关专业高年级本科生、研究生的教材, 也可供相关学科的工程技术人员参考.

图书在版编目(CIP)数据

复杂非线性系统的自适应优化控制/罗艳红, 张化光, 崔黎黎著.
—北京: 科学出版社, 2013

ISBN 978-7-03-037810-1

Ⅰ. ①复… Ⅱ. ①罗… ②张… ③崔… Ⅲ. ①非线性系统(自动化)-自适应控制 Ⅳ. ①TP273

中国版本图书馆 CIP 数据核字(2013) 第 126241 号

责任编辑: 张海娜 / 责任校对: 桂伟利
责任印制: 张 倩 / 封面设计: 蓝正设计

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

新科印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2013 年 6 月第 一 版 开本: B5(720 × 1000)
2013 年 6 月第一次印刷 印张: 10 3/4
字数: 214 000

定价: 55.00 元

(如有印装质量问题, 我社负责调换)

前　言

自 20 世纪末以来, 自适应动态规划进入了蓬勃发展的阶段. 当前, 自适应动态规划已被公认是解决复杂非线性系统最优控制的一种行之有效的方法.

传统最优控制理论在工业生产、军事科学、空间技术等领域取得了许多成功的应用. 但随着科学技术和生产力的不断发展, 实际的被控对象越来越复杂, 因而对最优控制方法在变量存储和计算速度方面提出了更高的要求. 遗憾的是, 传统的变分法、最大值原理和动态规划方法虽然理论上能够得到问题的解析解, 但由于各自存在的一些局限性, 在实际应用时仅能处理低维系统的最优控制问题. 因此, 如何找到一种方法能够在保证系统性能最优的前提下快速求得问题的可行解, 在很长一段时间内成为最优控制领域的热点问题. 于是在 1977 年, 自适应动态规划方法应运而生. 自适应动态规划的主要思想是首先利用一个函数近似结构 (如神经网络、模糊模型、多项式等) 来逼近代价函数, 进而基于贝尔曼最优性原理给出最优控制策略的表达式, 通过代价函数和控制策略的顺序或交替迭代, 逐步求解最优的控制策略和最优值函数, 从而按时间正向求解动态规划问题. 近 10 年来, 由于自适应动态规划方法融合了模糊控制、神经网络及增强学习技术, 为解决非线性系统的最优控制问题提供了新的思路, 从而成为了国内外控制领域研究的热点.

本书从迭代优化控制角度出发, 采用自适应动态规划方法进行复杂非线性系统的分析和综合问题的研究. 主要基于神经网络和模糊逼近原理, 采用 Lyapunov 稳定性理论和自适应双环-多环迭代等技术, 对各种复杂非线性系统的镇定控制和跟踪控制进行研究, 给出一些新的研究成果.

本书首先针对目前的自适应动态规划算法多数未考虑执行器的饱和或死区特性, 令所求得控制器的性能大打折扣的问题, 提出了一种能够处理执行器饱和问题的新型双启发式动态规划算法, 通过在性能指标中引入一个非二次型泛函, 并建立协状态迭代方程来直接求解最优控制律, 解决了复杂非线性系统在近似最优控制应用上的一些实际问题.

此外, 现存的大多数针对非线性迭代最优控制的算法都只能处理带有仿射控制输入的非线性系统. 而实际上, 很多工业应用中导出的系统都是带有强非线性结构的系统, 即不仅相对于状态是非线性的, 对于控制输入也是非线性的, 从而导出的是非仿射非线性系统. 由于非仿射系统的控制极其复杂, 由仿射系统导出的结果通常不适用于非仿射的对象. 本书针对非仿射系统的最优镇定和最优跟踪控制问题, 进行了较为深入的分析和研究, 是仿射非线性系统最优控制研究的一个延伸, 也是

该领域研究的深化.

最后, 考虑到实际非仿射控制系统中的控制增益符号可能未知的问题, 即当控制方向未知时, 我们不能确定控制沿着什么方向作用在被控对象上, 这使得系统的自适应迭代优化控制设计变得更为复杂. 本书通过引入 Nussbaum 函数和中值定理来处理非线性函数隐含控制输入及控制方向未知的问题, 给出一些新的研究成果.

本书提到的复杂非线性系统的自适应迭代优化控制是非线性控制理论的一个有益的补充, 是我们多年来在近似最优控制领域内取得的一些创新性的研究成果, 属于当前所属研究领域的前沿问题, 具有重要的理论与实际应用价值.

本书的出版得到了国家自然科学基金(61104010, 61034005, 61273029, 61273027)、教育部博士点新教师基金 (20110042120032)、中国博士后科学基金 (2012M510825)、国家重点基础研究发展计划 (2009CB320601)、东北大学优秀博士后基金以及东北大学研究生教育科研计划的资助, 在此表示感谢.

由于作者的水平和研究范围有限, 书中缺点和不足在所难免, 恳请读者批评指正.

罗艳红　张化光　崔黎黎

2013 年 4 月

目　　录

第1章 绪 论

1.1 引 言

众所周知, 针对线性定常系统的最优控制的理论和方法已经非常成熟[1-4]. 而针对非线性系统而言, 虽然人们也对其最优控制问题进行了相应的研究, 但由于其高度的复杂性, 一般很难得到其解析的最优控制解. 然而, 非线性是自然界和工程技术领域中最普遍的现象, 许多实际工程系统都具有本质上的非线性, 必须用非线性系统描述才合理. 因此, 研究非线性系统的最优控制问题, 无论在理论上还是在实践上, 都具有重大的意义.

动态规划是贝尔曼于 20 世纪 50 年代提出的求解多阶段决策过程最优化的一种数学方法, 现已在最优控制领域获得广泛应用. 然而, 随着系统维数和时间段的增加, 该方法显示出了一个致命的缺点, 即其计算量和存储量呈现惊人的增长, 出现了所谓的 "维数灾" 问题. 为了克服这些缺点, Werbos 于 1977 年首先提出了自适应动态规划 (ADP) 方法的框架[5], 其主要思想是利用一个函数近似结构 (如神经网络、模糊模型、多项式等) 来估计代价函数, 用于按时间正向求解动态规划问题. 近些年来, ADP 方法获得了控制界广泛的关注, 由于其融合了模糊控制、神经网络及增强学习技术, 为解决非线性系统的最优控制问题提供了新的思路, 从而成为了近年来国内外控制领域研究的热点.

另一方面, 由于神经网络具有强大的非线性逼近能力和学习能力, 为复杂非线性系统的控制开辟了一条崭新的途径. 神经网络自适应控制就是基于自适应的基本原理, 结合神经网络的特点和理论产生的新方法, 它简化了单纯自适应控制系统设计的复杂性, 发挥了自适应与神经网络各自的长处, 因而在智能控制研究领域中受到了广泛的关注, 并取得了大量的研究成果[6-10].

模糊控制自诞生以来, 经历了近 40 年的完善和发展, 逐步被认为是解决复杂非线性系统建模和控制的一种行之有效的方法. 通过模糊逻辑, 把语言信息构造到控制系统上, 从而对难以建立精确数学模型的对象, 提供了新颖的系统分析与设计的理念. 可以预见, 如果将模糊控制和最优控制技术相结合将是解决复杂非线性系统控制问题的新途径.

本章首先介绍了非线性系统最优控制的基本理论, 其次介绍了非线性系统自适应控制的发展历程, 接着重点介绍了基于神经网络的自适应动态规划算法的基本理

论和研究现状, 最后简述了作者在本领域所取得的主要研究成果.

1.2 非线性系统最优控制理论概述

最优控制理论是现代控制理论的重要组成部分, 其形成与发展奠定了整个现代控制理论的基础. 早在 20 世纪 50 年代初, Bushaw 就研究了伺服系统的时间最优控制问题, 他用几何方法证明了继电式的控制可以用最短的时间将伺服系统的误差调节到零. 之后, LaSalle 发展了时间最优控制的理论, 即所谓的 Bang-Bang 控制理论. 50 年代空间技术开始获得迅猛的发展. 导弹、卫星等都是复杂的多输入–多输出非线性系统, 而且在性能上有严格的要求. 这种工程上的要求刺激了最优控制理论的发展. 人们发现, 最优控制问题就其本质来说, 是一个变分学的问题. 然而, 经典变分学只能解决控制作用不受限制的情况. 实际上常常碰到控制作用受到限制的情况, 这就要求人们开辟求解最优控制的新途径. 1953 年至 1957 年间, 美国学者 Bellman 创立了 "动态规划" 理论[11], 发展了变分学中的 Hamilton-Jacobi 理论. 1956 年至 1958 年间苏联学者 Pontryagin 等创立了 "极大值原理"[12]. 这两种方法成为了目前最优控制理论的两个柱石. 时至今日, 最优控制理论的研究无论在深度上和广度上都有了很大的发展, 例如发展了对分布参数系统、随机系统、广义系统的最优控制理论的研究等. 毫不夸张地说, 最优控制理论仍是一个活跃的科学研究领域, 它在国民经济和国防事业中将继续发挥重要的作用.

众所周知, 动态规划从本质上讲, 是一种非线性规划方法, 其核心是贝尔曼最优性原理[13-16]: "一个多级决策问题的最优决策具有这样的性质: 当把其中任何一级及其状态作为初始级和初始状态时, 则不管初始状态是什么, 达到这个初始状态的决策是什么, 余下的决策对此初始状态必定构成最优策略." 根据这个原理, 动态规划解决多级决策问题时, 是从末端开始, 由最后一级逆向递推到始端的.

针对离散系统和连续系统, 动态规划方法各有发展.

1. 离散系统的动态规划

离散系统的动态过程是一个多阶段控制过程, 最优性原理是离散系统动态规划方法的理论基础.

假设一个系统的动态方程为

$$x(k+1) = F\left(x(k), u(k), k\right), \quad k = 0, 1, \cdots \tag{1.1}$$

其中, $x \in \mathbb{R}^n$ 是系统的状态向量; $u \in \mathbb{R}^m$ 为控制输入向量. 系统相应的代价函数 (或性能指标函数) 形式为

$$J\left(x(i),i\right)=\sum_{k=i}^{\infty}\gamma^{k-i}l\left(x(k),u(k),k\right) \tag{1.2}$$

其中, 初始状态 $x(k)=x_k$ 给定; $l\left(x(k),u(k),k\right)$ 是效用函数; γ 为折扣因子且满足 $0<\gamma\leqslant 1$. 控制目标就是求解容许决策 (或控制) 序列 $u(k)(k=i,i+1,\cdots)$, 使得代价函数 (1.2) 最小.

根据贝尔曼最优性原理, 始自第 k 时刻任意状态的最小代价包括两部分, 其中一部分是第 k 时刻内所需最小代价, 另一部分是从第 $k+1$ 时刻开始到无穷的最小代价累加和, 即

$$J^*\left(x(k)\right)=\min_{u(k)}\left\{l\left(x(k),u(k)\right)+\gamma J^*\left(x(k+1)\right)\right\} \tag{1.3}$$

相应的 k 时刻的控制策略 $u(k)$ 也达到最优, 表示为

$$u^*(k)=\arg\min_{u(k)}\left\{l\left(x(k),u(k)\right)+\gamma J^*\left(x(k+1)\right)\right\} \tag{1.4}$$

2. 连续系统的动态规划

对于一个连续系统, 最优性原理也同样成立, 具体可以陈述如下: 对于初始时刻 t_0 和初始状态 $x(t_0)$, 若 $u^*(t),t\in[t_0,t_{\mathrm{f}}]$ 和 $x^*(t)$ 是系统的最优控制和最优状态轨迹, 则对于时刻 $t_1(t_1>t_0)$ 和相应状态 $x(t_1)$ 来说, $u^*(t)(t\in[t_1,t_{\mathrm{f}}])$ 和 $x^*(t)$ 仍是系统在时刻 t_1 之后的最优控制和最优状态轨迹.

如下的连续时间系统:

$$\dot{x}(t)=F\left(x(t),u(t),t\right),\quad t\geqslant t_0 \tag{1.5}$$

其中, $F(x,u,t)$ 为任意连续函数. 求一容许控制策略 $u(t)$ 使得代价函数 (或性能指标函数) 最小:

$$J\left(x(t),t\right)=\int_t^{\infty}l\left(x(\tau),u(\tau)\right)\mathrm{d}\tau \tag{1.6}$$

我们可以通过离散化的方法将连续问题转换为离散问题, 然后通过离散动态规划方法求出最优控制, 当离散化时间间隔趋于零时, 两者必趋于一致. 通过应用贝尔曼最优性原理, 我们可以得到 DP 的连续形式为

$$\begin{aligned}-\frac{\partial J^*}{\partial t}=&\min_{u\in U}\left\{l\left(x(t),u(t),t\right)+\left(\frac{\partial J^*}{\partial x(t)}\right)^{\mathrm{T}}F(x(t),u(t),t)\right\}\\=&l\left(x(t),u^*(t),t\right)+\left(\frac{\partial J^*}{\partial x(t)}\right)^{\mathrm{T}}F\left(x(t),u^*(t),t\right)\end{aligned} \tag{1.7}$$

我们可以看出上式是 $J^*\left(x(t),t\right)$ 以 $x(t)$、t 为自变量的一阶非线性偏微分方程, 在数学上称其为哈密顿–雅可比–贝尔曼 (HJB) 方程.

1.3　非线性系统自适应控制理论概述

20 世纪 50 年代末期, 由于飞行控制的需要, 美国麻省理工学院的 Whitaker 教授首先提出了模型参考自适应控制方法 (后来被称为 MIT 方法), 企图用它来解决飞行器的自动驾驶问题. 该方法提出采用局部参数优化理论设计自适应控制律, 但由于其设计控制器时没有考虑系统的稳定性, 因而限制了这一方法的实际应用. 1966 年, 德国学者 Parks 提出了利用 Lyapunov 第二方法来推导自适应控制律的自适应系统设计方法. 该方法可以通过 Lyapunov 的稳定性定理保证自适应系统的全局渐近稳定性, 但需要用到被控对象输入和输出的各阶导数来构成自适应律, 这就减低了自适应系统对干扰的抑制能力. 为了克服上述缺点, 包括Åström 和 Narendra 在内的众多学者提出了很多不同的改进方案[17, 18]. 在 1973 年瑞典学者Åström 和 Wittenmark 首次提出了自校正调节器的概念, 接着 1975 年 Clark 等提出了自校正控制器设计方法, 1979 年 Wellstead 和Åström 提出了极点配置自校正调节器的设计方案. 另外, 法国学者 Landau 把罗马尼亚学者 Popov 在 1963 年提出的超稳定性理论应用于模型参考自适应控制方案中, 证明了所设计的模型参考自适应系统是全局渐近稳定的.

经过 50 多年的时间, 自适应控制无论在理论上还是在应用上都取得了长足的发展[19]. 特别是近 20 年来, 以神经网络、模糊逻辑和进化计算为代表的人工智能理论和方法开始应用于自适应控制理论中, 形成了神经网络自适应控制、模糊自适应控制等新的分支, 为非线性系统的自适应控制提供了新的方法和手段. 然而在应用自适应控制方法控制环境与过程高度不确定情况下的复杂系统时, 还存在一些问题, 如自适应控制器结构过于复杂, 模型参考自适应控制系统对确定性干扰不能确保零稳态误差等. 另一方面, 80 年代以来迅速发展起来的神经网络, 显示出它在解决高度非线性和严重不确定系统控制方面的巨大潜力[20], 其吸引力在于: ①能够充分逼近任意复杂的非线性映射关系; ②能够学习与适应严重不确定性系统的动态特性; ③有高度的鲁棒性和容错能力; ④采用并行分布处理, 使得快速进行复杂运算成为可能.

然而, 在 20 世纪 90 年代初, 大多数神经网络控制方法, 包括 Narendra 提出的神经网络控制方法[21], 都是通过静态和动态反传的梯度法实现的. 但是梯度法的缺点是不能保证整个闭环系统的稳定性、鲁棒性和动态性能, 特别是在控制器参数在线自适应调节的情况下. 因此, 为了克服这些问题, Sanner[22] 和 Polycarpou[23] 等提出了稳定的神经网络自适应控制. 虽然这些研究成果最初集中于线性参数化神经网络, 如 RBF 网络、B 样条网络、小波网络等. 但近 10 余年来, 基于前向神经网络的具有闭环稳定性的神经网络自适应镇定控制和跟踪控制方案受到了大家广

泛的关注, 涌现出了大量的研究成果 [6, 7, 9, 10].

此外, 模糊控制自诞生以来, 经历了近 50 年的完善和发展, 逐步被认为是解决复杂非线性系统建模和控制的一种行之有效的方法. 通过模糊逻辑, 把语言信息构造到控制系统上, 从而对难以建立精确数学模型的对象, 提供了新颖的系统分析与设计的理念. 随着模糊理论的发展, 人们将模糊控制与其他一些控制思想相结合, 衍生出许多不同而实用的控制方法, 例如, 模糊预测控制、模糊自适应控制、模糊鲁棒控制等. 文献 [24] 是一部介绍模糊自适应控制的重要专著, 详细地介绍了各种模糊自适应控制方法. 针对多输入多输出非线性系统, 文献 [25] 采用模糊模型参考自适应控制方法, 应用 Lyapunov 稳定性理论分析了整个系统的稳定性和鲁棒性. 文献 [26] 针对不确定系统设计了一致最终有界模糊自适应跟踪控制器. 文献 [27] 利用小增益原理设计了模糊自适应鲁棒跟踪控制器.

尽管神经网络自适应控制和模糊自适应控制已取得一些令人瞩目的成果, 但是在理论上都还存在一些亟待解决的问题, 仍需要进一步深入研究和探讨. 例如, 由于神经网络或模糊模型的引入, 其近似误差不可避免, 导致最终只能获得一致最终有界的结果; 稳定性只是对系统设计的一个基本要求, 如何将神经网络自适应控制或模糊自适应控制同最优控制相结合, 既能保证系统稳定性, 又能优化系统的性能也需要进一步的深入探讨.

1.4　自适应迭代最优控制的发展及研究现状

1.4.1　自适应动态规划算法的发展历程

动态规划法是求解离散和连续系统最优控制问题的一种计算方法[28, 29], 它适用的范围很广, 如多输入多输出的系统、线性系统和非线性系统. 但是由于该方法存在很致命的缺点, 主要就是其时间上从后往前递推的算法使得动态规划虽然比穷举法的计算量有所减少, 但是对于复杂问题, 例如状态变量和控制变量的数目多、级数多时, 应用动态规划方法产生的计算量和存储量仍旧非常大, 有时用一般计算机也解决不了, 出现了所谓的 “维数灾” 问题[11, 30], 故其一般只适用于小规模简单非线性系统的最优控制问题的求解.

正是为了解决动态规划的 “维数灾” 问题, Werbos 在 1977 年的一篇论文[5] 里提出了后来被称为 “自适应评价设计”(adaptive critic designs) 的自适应动态规划算法. 随后很多学者也开始探讨这方面的问题, 寻求动态规划问题的近似解法, 包括文献 [31]~[60]. 在现有文献中, “自适应动态规划” 有许多同义词, 如在文献 [33] 中被称为 “近似动态规划”(approximate dyanmic programming, ADP), 在文献 [39] 中被称为 “ 渐近动态规划”(asymptotic dynamic programming), 在文献 [32] 和 [40] 中被

称为 “启发式动态规划”(heuristic dynamic programming), 在文献 [37] 中被称为 “神经动态规划”(neuro-dynamic programming), 在文献 [47] 中被称为 “神经元动态规划”(neural dynamic programming), 在文献 [35]、[36]、[38] 和 [42] 中被称为 “自适应评价设计”(adaptive critic designs), 在文献 [61] 中被称为 “增强学习”(reinforcement learning).

虽然增强学习也是寻求动态规划问题的近似解, 但由于历史原因, 增强学习方面研究的发展有其自己的独立性[62, 63]. 增强学习方面的主要研究成果参见 Sutton 和 Barto 的书[61] 以及书中所列的参考文献. 这一领域里最著名的两个算法当属时间差分方法[64] 和 Q- 学习方法[65]. 虽然该领域的研究目前比较成熟并有大量的参考文献, 但其共同局限性就是状态变量只能取有限个离散点, 这样才使得目标函数可以用一个有限表格来表达.

另外, 文献 [66]~[68]给出了一种适用于有限基问题的自适应动态规划方法, 并称其为 “神经动态最优化”(neural dynamic optimization), 但这种方法的实现主要依赖于非线性最优化算法, 同时缺少算法的收敛性证明.

近些年自适应动态规划方面的理论研究工作已有一些成果, 目前在本领域里发表的较著名的文献是 [38]、[44]、[52]、[60]. 文献 [38] 对直到 1997 年自适应动态规划方面的研究做了很详尽的总结. 在这之前, 这方面的研究成果主要见 Werbos 的论文[5, 32, 33]. Werbos 曾经多次指出 “自适应动态规划是目前唯一有可能达到类人脑智能” 的研究方向[31, 34]. 文献 [39] 和 [44] 针对自适应动态规划方法的稳定性和最优性做了初步研究, 特别是文献 [44] 第一次从数学上严格证明了对于连续时间系统, 当从一个初始稳定的控制策略开始迭代, 通过给出的迭代算法能够求得系统的近似最优控制, 并基于 HJB 方程给出了算法的稳定性和收敛性分析. 而对于离散系统, 其最优迭代控制算法的研究一直进展不大, 虽然文献 [54] 也给出了从初始稳定控制策略开始迭代的离散系统最优控制迭代算法, 但其要求满足小摄动的假设. 直到 2007 年, 文献 [52] 和 [53] 提出了一种贪婪的 HDP 迭代方案以逐步求解非线性离散系统的最优控制问题, 其首次提出了求解非线性系统的近似最优控制可以从任意控制策略开始迭代的方案. 继而文献 [60] 对任意策略开始迭代的近似最优控制算法的稳定性和最优收敛性进行了较为全面的分析, 并给出了相应的数学证明.

1.4.2 自适应动态规划算法的基本理论

自适应动态规划是动态规划的一种近似实现形式, 故也被称为近似动态规划. 自适应动态规划的基本思想是首先通过某种函数近似工具实现对指标函数的近似表示, 然后再通过另外一个近似工具实现对最优控制序列的选择, 从而实现整个系统的近似最优控制.

在文献 [38] 中, 给出了三种基本的自适应动态规划算法: 启发式动态规划 (heuristic dynamic programming, HDP)、二次启发式规划 (dual heuristic programming, DHP) 和全局二次启发式规划 (globalized dual heuristic programming, GDHP), 并给出了上述三种算法的变形: 控制依赖启发式动态规划 (action dependent HDP, ADHDP)、控制依赖二次启发式动态规划 (ADDHP) 以及控制依赖全局二次启发式动态规划 (ADGDHP) 算法的构造原理和实现方法.

1. HDP

HDP 是自适应动态规划方法中最基本的一种, 其主要由三个模块组成: 评价模块 (critic)、模型模块 (model) 及控制模块 (action), 其中每个模块皆可通过函数近似工具如神经网络、模糊基函数等近似逼近, 因此它也被称为自适应评价设计 (ACD).

基于神经网络实现的 HDP 的结构如图 1.1 所示, 其中包含三个神经网络, 一个网络称为 "控制网络"(action network), 代表系统状态变量到控制变量之间的映射; 第二个网络称为 "模型网络"(model network), 用于对未知非线性系统进行建模; 第三个网络称为 "评价网络"(critic network), 以状态变量作为输入, 输出则是性能指标函数.

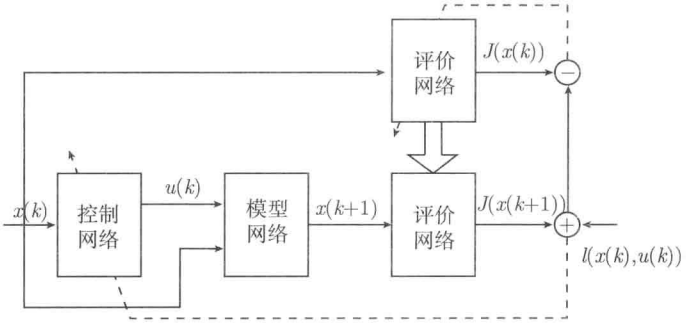

图 1.1　HDP 结构示意图

在 HDP 中, 性能指标函数可以写成如下表达式:

$$J(x(k)) = l(x(k), u(x(k))) + J(x(k+1)) \tag{1.8}$$

其中, $u(x(k))$ 是反馈控制变量; 性能指标函数 $J(x(k))$ 和 $J(x(k+1))$ 为评价神经网络的输出. 如果评价网络的权值设为 w, 我们可以令式 (1.8) 的等号右边为

$$d(x(k), w) = l(x(k), u(x(k))) + J(x(k+1), w) \tag{1.9}$$

同时式 (1.8) 的等号左边可以写为 $J(x(k),w)$. 因此可以通过调节评价神经网络权值 w 最小化如下均方误差函数:

$$w^* = \arg\min_w \left\{ |J(x(k),w) - d(x(k),w)|^2 \right\} \tag{1.10}$$

获得最优性能指标函数. 根据最优性原理, 最优控制应满足一阶微分必要条件, 即有

$$\begin{aligned} \frac{\partial J^*(x(k))}{\partial u(k)} =& \frac{\partial l\left(x(k),u(k)\right)}{\partial u(k)} + \frac{\partial J^*(x(k+1))}{\partial u(k)} \\ =& \frac{\partial l\left(x(k),u(k)\right)}{\partial u(k)} + \frac{\partial J^*(x(k+1))}{\partial x(k+1)} \frac{\partial f\left(x(k),u(k)\right)}{\partial u(k)} \end{aligned} \tag{1.11}$$

因此得到最优控制

$$u^* = \arg\min_u \left(\left| \frac{\partial J(x(k))}{\partial u(k)} - \frac{\partial l\left(x(k),u(k)\right)}{\partial u(k)} - \frac{\partial J^*(x(k+1))}{\partial x(k+1)} \frac{\partial f\left(x(k),u(k)\right)}{\partial u(k)} \right| \right) \tag{1.12}$$

其中, $\dfrac{\partial J^*(x(k+1))}{\partial x(k+1)}$ 可以通过评价网络权值 w 和输入输出关系式得出.

在上述 HDP 中, 如果控制网络的输出直接作为评价网络的部分输入, 则这种自适应评价设计被称为 ADHDP, 其工作原理与 HDP 基本相同, 其结构图如图 1.2 所示.

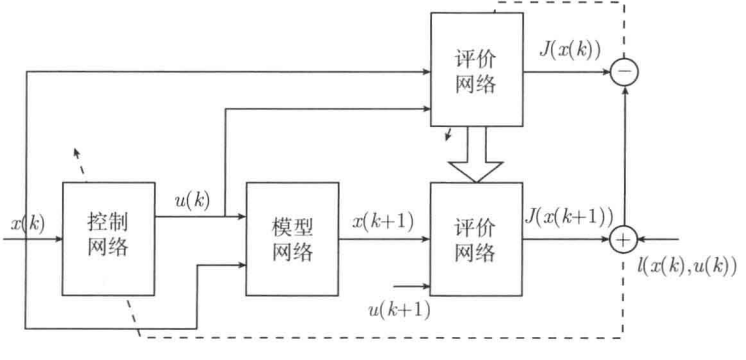

图 1.2　ADHDP 结构示意图

在实现上 HDP 与 ADHDP 的最大区别在于 ADHDP 的评价网络不但以系统状态作为输入, 同时也以控制变量作为输入, 评价网络的输出通常称为 Q-函数, 因此 ADHDP 也被称为 Q-学习[69]. 其具体算法可以参见 HDP 部分和相关文献[69, 38].

2. DHP

DHP 结构如图 1.3 所示, 同样包含三个神经网络, 分别是控制网络、模型网络和评价网络. 其中控制网络与模型网络的功能和作用与 HDP 相同. 但对于 DHP, 评价网络将逼近性能指标函数 J 对状态 x 的导数而不是性能指标函数 J 本身, 其中 $\dfrac{\partial J(x(k))}{\partial x(k)}$ 也叫做协状态 (costate). 为此, 我们需要知道效用函数对状态的导数 $\dfrac{\partial l(x(k),u(k))}{\partial x(k)}$ 以及系统函数对状态的导数 $\dfrac{\partial f(x(k),u(k))}{\partial x(k)}$.

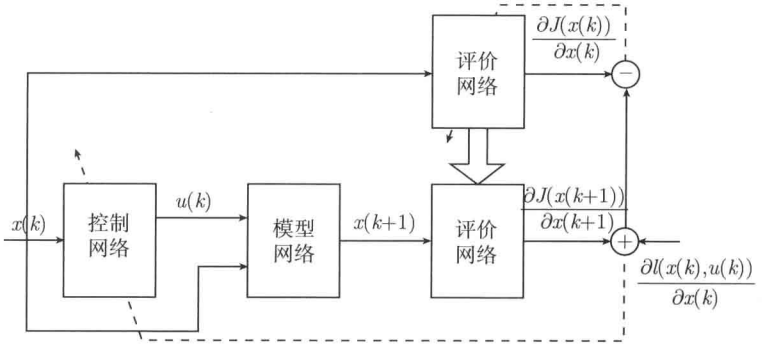

图 1.3　DHP 结构示意图

DHP 算法根据性能指标函数和效用函数对状态的导数进行迭代如下:

$$\frac{\partial J(x(k))}{\partial x(k)}=\frac{\partial l(x(k),u(x(k)))}{\partial x(k)}+\frac{\partial J(x(k+1))}{\partial x(k)} \tag{1.13}$$

其中, $u(x(k))$ 是反馈控制变量, 协状态 $\dfrac{\partial J(x(k))}{\partial x(k)}$ 和 $\dfrac{\partial J(x(k+1))}{\partial x(k)}$ 为评价网络的输出. 如果评价网络的权值设为 w, 则我们可以令式 (1.8) 的右式为

$$e(x(k),w)=\frac{\partial l(x(k),u(x(k)))}{\partial x(k)}+\frac{\partial J(x(k+1),w)}{\partial x(k)} \tag{1.14}$$

同时式 (1.8) 左式可以写为 $\dfrac{\partial J(x(k),w)}{\partial x(k)}$.

通过调节评价网络的权值 w 来最小化如下均方误差函数:

$$w^*=\arg\min_w\left\{\left|\frac{\partial J(x(k),w)}{\partial x(k)}-e(x(k),w)\right|^2\right\} \tag{1.15}$$

可以使得评价网络的输出为协状态.