

基于自适应动态规划的导弹制导律研究综述

孙景亮¹ 刘春生¹

摘要 自适应动态规划 (Adaptive dynamic programming, ADP) 作为最优控制领域的近似优化方法, 是求解复杂非线性系统最优控制问题的有力工具. 近年来, 已成为控制理论与计算智能领域的研究热点. 本文着重介绍 ADP 算法的理论研究进展及其在航空航天领域的应用. 分析了几种典型的制导律优化设计方法, 以及 ADP 方法在导弹制导律设计中的应用现状和前景.

关键词 自适应动态规划, 最优控制, 导弹, 制导律

引用格式 孙景亮, 刘春生. 基于自适应动态规划的导弹制导律研究综述. 自动化学报, 2017, 43(7): 1101–1113

DOI 10.16383/j.aas.2017.c160735

An Overview on the Adaptive Dynamic Programming Based Missile Guidance Law

SUN Jing-Liang¹ LIU Chun-Sheng¹

Abstract Adaptive dynamic programming (ADP) is a powerful tool for optimal control of complicated nonlinear system, which is a novel approximate optimal control method. Recently, it has become a hot topic in the field of control theory and computational intelligence. This paper focuses on giving a review of ADP on the development of ADP algorithms and its aerospace applications. The design methods of classic missile guidance law are introduced, as well as the present and potential applications of ADP in the guidance law design of missiles.

Key words Adaptive dynamic programming (ADP), optimal control, missile, guidance law

Citation Sun Jing-Liang, Liu Chun-Sheng. An overview on the adaptive dynamic programming based missile guidance law. *Acta Automatica Sinica*, 2017, 43(7): 1101–1113

自适应动态规划 (Adaptive dynamic programming, ADP) 技术是最优控制领域新近兴起的一种近似最优控制方法, 是当前国际最优化领域的研究热点^[1], 其本质是基于强化学习 (Reinforcement learning, RL) 原理, 模拟人类通过环境进行反馈学习, 是一种非常接近人脑智能的控制方法. ADP 技术的基本结构是执行器–评价器结构, 执行器模块通过和环境交互产生决策或控制, 评价器模块通过评价函数判断系统性能的优劣来调整控制策略. ADP 理论融合了强化学习、动态规划以及函数近似等方法, 利用函数近似结构估计代价函数, 采用离线或在线更新方法, 逼近系统的最优解, 从而有效地解

决非线性系统的最优控制问题. 从学科的角度来看, ADP 依托现代控制理论和计算智能理论, 是一种新的智能控制理论与方法^[2].

上世纪 80 年代, Werbos^[3] 首次提出自适应动态规划的思想, 命名为 ACD (Adaptive critic design). 其基本思想是利用一个函数近似结构估计代价函数, 用于按时间正向求解动态规划问题, 避免了传统动态规划计算维数灾问题, 为高维复杂系统的最优控制提供了一个便捷、有效的解决方法. 随后, ADP 方法受到广大学者的关注, 产生了一系列同义名词, 例如自适应评价设计 (Adaptive critic design)^[4–5]、近似动态规划 (Approximate dynamic programming)^[6]、神经元动态规划 (Neuro-dynamic programming)^[7]、强化学习 (Reinforcement learning)^[8–9] 等. 近十年来, ADP 方法已经成为控制理论与计算智能领域的研究热点. 2006 年, 在美国科学基金会组织的 “2006 NSF Workshop and Outreach Tutorials on Approximate Dynamic Programming” 研讨会上, 建议将该方法统一为 Adaptive/Approximate dynamic programming. 从 2007 年起, IEEE 计算智能协会每两年召开一次 ADP 与 RL 国际专题研讨会, 且于 2008 年专门成立了 ADP 与 RL 技术委员会, 标志着 ADP 成为一个重要的研究领域. 迄今为止, 有关

收稿日期 2016-10-25 录用日期 2017-02-06
Manuscript received October 25, 2016; accepted February 6, 2017

国家自然科学基金 (61473147), 江苏省普通高校学术学位研究生科研创新计划项目 (KYLX16_0376), 南京航空航天大学博士学位论文创新与创优基金 (BCXJ16-02) 资助

Supported by National Natural Science Foundation of China (61473147), Funding of Jiangsu Innovation Program for Graduated Education (KYLX16_0376), and Funding for Outstanding Doctoral Dissertation in Nanjing University of Aeronautics and Astronautics (NUAA) (BCXJ16-02)

本文责任编辑 魏庆来
Recommended by Associate Editor WEI Qing-Lai

1. 南京航空航天大学自动化学院 南京 211106
1. College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106

ADP 的研究已经取得丰硕成果,在国际上已经出版数本专著^[10-12]以及多篇综述性文章^[1-2,13-15]。

随着航天技术的飞速发展,各国之间的空间竞争范围不断扩大且日益激烈。精确制导武器的发展已经在各军事强国中引起高度重视。导弹作为一种强威慑与大威力打击武器,是军事强国军事变革中武器装备战略转型的优选目标,在当今信息化战争中可以“首当其冲,先发制人”,这必然推动传统作战方式的重大变革,对现代战争观念产生巨大而深远的影响^[16]。因此,导弹制导技术的发展至关重要且备受关注。在导弹制导与控制过程中,末端制导律的设计直接影响打击精度。为了实现战场中对目标的精确打击,制导律在攻击末端对脱靶量要求尤为严格。传统的导弹制导技术(例如比例导引及其变形)主要是打击固定目标或者有人驾驶飞行器等速度较慢的目标,由于结构简单、需要的制导参数少而得到广泛应用^[17]。然而,随着现代军事科技的进步,战术弹道导弹、高速智能巡航导弹、无人驾驶飞行器等高速飞行器的威胁越来越突出。由于此类新目标具有较快的速度和较强的机动能力,要求导弹以最小脱靶量命中目标还不够,还希望能够在满足某些约束条件(例如终端角度、终端速度等)时命中目标,这就对制导系统提出了新的挑战。再者,现代和未来战争的战场环境越来越复杂,机动目标有可能会使用各种干扰手段(例如伪装、隐蔽、欺骗等)试图躲避打击^[18],从而达到突防的目的,这就对导弹制导系统的稳定性与鲁棒性提出了更高的要求。近年来,基于现代控制理论和智能控制理论发展的最优控制、滑模变结构控制、智能控制、自适应控制、鲁棒控制等诸多控制方法开始用于导弹制导律的设计。基于 ADP 技术的制导方式能够使导弹根据目标的机动实时在线调整和更新末端制导律,达到独立自主、智能作战的要求,能够根据外界环境的变化以及系统自身的特性,自适应地、实时地应对突发情况,是一种全新的智能优化控制方法,有望成为未来智能制导系统的理论依据。

为了进一步提高导弹制导系统的稳定性,拓宽 ADP 方法的应用范围,使之能够在导弹制导过程中发挥主导作用,本文对 ADP 方法的理论研究进展进行全面梳理,详细介绍其在航空航天领域的应用现状,简单介绍几种典型的制导律方法的研究进展,并从导弹制导律设计的方法和特点出发,对 ADP 方法在导弹制导律设计中的应用进行了探讨。

1 ADP 算法的发展

ADP 算法可以按照不同的标准进行分类。按照算法迭代方式主要分为策略迭代法(Policy iteration)和值迭代法(Value iteration)(统称迭代 ADP

算法);按照算法执行方式可以分为离线迭代算法和在线自适应算法。其理论研究主要涉及稳定性分析和收敛性证明以及在各类控制问题中的应用。本文主要按照以上两种分类标准,分别对 ADP 算法进行比较全面的梳理。

1.1 迭代 ADP 算法

迭代 ADP 算法主要分为策略迭代法和值迭代法。策略迭代算法是从一个初始稳定的控制策略开始,通过求解一系列 Lyapunov 等式,逐步改进控制策略从而达到最优。该算法主要包括策略评价和策略提高两个过程;值迭代算法建立在值函数更新和策略提高两个过程上,不需要给定一个初始的稳定控制策略,而是从任意一个正定的初始值函数开始。

多年来,研究者针对策略迭代与值迭代进行了大量研究,取得了丰硕成果。其中,文献[19]首次分析了离散非线性系统的策略迭代算法特性,并严格证明了在初始稳定控制器下,其迭代值函数是单调非增的且收敛于最优值。文献[20]针对离散系统设计了一种局部策略迭代算法,避免了传统策略迭代算法要求全局状态信息的苛刻要求。文献[21]考虑离散系统的时滞问题,提出了基于策略迭代的最优跟踪控制算法,并给出了稳定性证明。文献[15]首次针对连续非线性系统提出了基于策略迭代算法的最优控制策略。文献[22]考虑很多实际系统中状态与控制输入为复数形式,讨论了非线性系统的复值策略迭代算法。文献[23]和文献[24]融合策略迭代与值迭代算法,提出了广义策略迭代 ADP 算法。文献[25]针对离散系统提出了迭代零和 ADP 算法,该算法可以使用任意半正定函数初始化迭代的上下界。随后基于 Q-学习的策略迭代算法也被应用于控制器的设计^[26-28]。

然而,策略迭代算法要求系统初始稳定控制器这一要求,对于很多非线性系统不易实现,而值迭代算法却对此没有要求,因此受到很多研究者的青睐。自从 Bertsekas 等首次提出了值迭代算法概念以来^[7],该算法得到了很大的发展。2002 年, Murray 等首次提出了针对连续系统的值迭代 ADP 算法,并给出了系统的稳定性和迭代的收敛性证明^[29]。文献[30]考虑一般非线性形式,提出了基于值迭代的启发式动态规划方法,并给出了该算法的收敛性证明。Wang 等考虑有限域 ADP 算法,提出了基于 ϵ -误差界值算法。该算法保证了迭代性能指标函数能够保持在一个 ϵ 界值内^[31]。考虑有限域输入受限情况,文献[32]舍弃了评价网与控制网并存的双网络结构,构建了有限域单网络评价结构,并给出了算法的稳定性证明。近年来,魏庆来等针对值迭代算法做了大量的研究^[33-39]。其中,文献[40]针对离散系统,

提出局部值迭代 ADP 算法, 重点分析了迭代 ADP 算法的可容许性能, 并建立了新的终止准则. 文献 [34] 提出了一种新的值迭代 ADP 算法, 该算法不再局限于零初始性能指标的条件, 可以实现以任意半正定函数作为初始性能指标函数的稳定控制. 文献 [39] 针对离散非线性系统, 建立了数字迭代 ADP 算法结构. 此外, 基于 Q-学习的值迭代算法也得到关注^[35-36].

策略迭代与值迭代算法是迭代 ADP 算法的两种主要迭代方式, 其实现形式可分为离线迭代方式和在线实现方式两种. 下面就离线迭代与在线实现两种情况分别阐述.

1.2 离线迭代算法

相对于连续系统而言, 针对离散系统设计的离线 ADP 算法成果比较丰富. 其中, Al-Tamimi 等针对离散系统的最优控制问题, 提出一种不要求初始稳定控制的贪婪迭代 ADP 算法^[30], 并证明了算法的收敛性. Wei 等通过设计 ADP 算法, 解决了在考虑神经网络有限逼近误差的情况下, 离散系统的最优跟踪问题^[41]. 此外, 在针对带有饱和约束条件下的近似最优控制问题^[21, 41]、带有时滞的离散非线性系统控制问题^[43-44] 和有限时间最优控制问题^[32, 45] 等方面, ADP 算法均取得了丰硕的研究成果.

2002 年, Murray 等提出针对连续非线性系统的策略迭代 ADP 算法^[29], 该算法要求系统具有初始稳定控制策略, 并详细给出了系统稳定性和算法收敛性的证明. 这是第一次从数学上证明了具有初始稳定控制策略的迭代算法能够保证系统的稳定性和迭代算法的收敛性, 对 ADP 理论的发展具有里程碑的意义. 随后, Abu-Khalaf 等使用基于神经网络的值迭代 ADP 算法^[46], 求解具有执行器饱和的非线性系统的最优控制问题, 该算法不要求系统具有初始稳定控制器. Tassa 等利用神经网络的非线性逼近能力, 逼近值函数, 而后利用最小二乘法求解连续 HJB 方法^[47]. Song 等考虑模型未知且存在干扰情况下, 构建执行网-评价网结构, 利用积分强化学习算法, 提出离线迭代控制方法, 通过设计干扰补偿项来保证迭代控制器收敛到最优值^[48]. 文献 [49] 针对连续非线性系统非零和微分对策问题, 采用积分强化学习方法, 设计了离线策略积分强化学习算法, 并证明了闭环系统的渐近稳定性.

然而, 上述方法均采用离线迭代的方式, 一旦系统发生变化, 就需要重新离线计算, 不仅给控制器造成不必要的计算损耗, 而且无法实时保证闭环系统的稳定性. 因此, 发展在线 ADP 算法具有十分重要的理论意义和工程价值.

1.3 在线自适应算法

为了能够满足控制系统对自身与外界环境变化的自适应性和鲁棒性等要求, 在线 ADP 算法越来越凸显出其巨大的潜力. 近十年来, 在线 ADP 算法的理论研究很受欢迎, 并取得了丰硕的研究成果.

Lewis 及其团队多年来一直致力于 ADP 算法的研究, 并形成了一套较完整的理论体系^[10, 12]. Vrabie 等首先提出了具有执行器-评价器结构的在线积分强化学习算法解决线性连续系统的最优控制问题^[50], 该方法不需要已知系统全部动态特性, 只要求系统部分特性(输出特性)已知. 随后, 将其推广到非线性连续系统^[51]. 但该控制器是通过策略评价与策略提高两步骤交替进行, 即在策略评价过程中, 策略提高部分需要保持不变, 反之亦然. 这就不可避免地表现出其内部的离散特性, 而这一特性使得闭环系统稳定性的分析具有一定的困难. 因此, Vamvoudakis 等提出了同时调整评判网络和执行网络的在线同步策略迭代算法^[52-53], 通过 Lyapunov 函数法可以证明闭环系统的一致最终有界稳定性, 但该方法要求系统的全部动态特性.

张化光及其团队深入开展了 ADP 的理论研究, 针对仿射系统、非仿射系统、切换系统以及时滞系统等进行了 ADP 算法的研究. 其中, 文献 [54] 首次设计了一个新型的鲁棒项, 保证了误差系统的渐近稳定性. 刘德荣及其团队关于在线 ADP 算法的研究也取得了丰硕的研究成果^[55-58]. 王鼎等通过适当定义性能指标函数, 解决了非线性系统存在不确定项的最优控制问题^[57-59].

ADP 算法在稳定性方面主要是通过 Lyapunov 函数法证明闭环系统的稳定性, 但到目前为止, 大部分 ADP 算法只能证明一致最终有界稳定, 离希望的渐近稳定的要求还有一定的距离. 在线 ADP 算法的发展已经拓展到镇定调节控制、输入约束控制、跟踪控制、鲁棒控制、微分对策等问题. 下面就这几类控制问题, 详述在线 ADP 算法的设计方法.

在线 ADP 算法关于镇定调节控制问题的研究比较广泛. 文献 [52-53, 60] 为了实现策略迭代算法的在线执行过程, 通过构建评价网络和执行网络, 并设计权值更新律, 利用 Lyapunov 函数方法证明了权值收敛特性. 文献 [61] 针对模型部分未知情况, 通过构建辨识器-评价器结构, 实现了辨识网络和评价网络同步更新, 保证闭环系统的稳定性. 文献 [62] 将 Backstepping 技术、平方和技术与 ADP 技术相结合, 基于控制 Lyapunov 函数思想, 设计了一种全新的在线 ADP 控制方法, 该方法得益于 Backstepping 技术逐步反推的优势, 可以有效减轻在线 ADP 算法的计算负荷. 文献 [63] 通过重复利

用存储的数据执行评价网络在线运算,有效避免了自适应控制要求的持续激励条件,通过增加鲁棒控制项,保证了闭环系统的渐近稳定性。

跟踪控制问题相对于调节问题更加复杂,对控制器的要求更高,但在工程应用中却更加普遍。因此,基于 ADP 算法的最优跟踪控制问题得到了广泛的研究。文献 [64] 研究了不确定非线性系统的保性能 ADP 跟踪控制问题。Dierks 等针对严格反馈形式的非线性连续系统,首先通过 Backstepping 技术将其转换为仿射形式,然后,利用单网络在线逼近器求解近似最优控制器^[65-66],实现非线性系统的 ADP 跟踪控制,该方法不需要系统的初始稳定控制器。文献 [67] 和文献 [68] 通过将误差动态系统和期望跟踪系统增广为一个新系统,重新定义性能指标函数,实现了模型部分未知的在线 ADP 跟踪控制。文献 [69] 考虑了多输入多输出非线性系统的在线 ADP 跟踪控制问题。Zhang 等通过数据驱动方法,实现了模型未知情况下的在线 ADP 跟踪控制,通过设计一个新型的鲁棒控制项,保证了跟踪误差渐近收敛到零^[54]。

输入受限问题是实际工程应用必须考虑的重要因素之一,它直接影响控制器的控制性能,甚至会导致闭环系统不稳定。因此,对于 ADP 算法的输入受限问题,在近几年的研究中得到了广泛的关注。Lewis 等比较早地研究了 ADP 算法输入受限问题,通过定义一个非二次型性能指标函数,将其表示为输入自由的非线性形式,然后利用 ADP 方法实现在线自适应求解^[46-47, 70-72]。类似地,刘德荣、张化光等关于 ADP 输入受限问题的研究也取得了很多成果^[55-56, 73]。该方法能够有效地处理输入受限控制问题,得到大家一致认同,并被广泛应用,但其稳定性证明比较复杂且仅限于处理常值受限问题。

鲁棒控制是处理系统内部摄动和外界干扰等不确定因素行之有效的控制方法。ADP 控制器的设计大部分是基于系统精确数学模型。但在实际控制系统中,外界干扰和内部扰动等不确定性因素不可避免,因此,发展鲁棒自适应动态规划方法必不可少。Jiang 等结合现代控制理论,较早地提出鲁棒自适应动态规划的概念^[74],随后,将其发展到非线性系统^[75-77],并保证了闭环系统的全局渐近稳定性,系统动态特性和系统阶数不需要精确已知。王鼎等通过适当定义性能指标函数,使之既能够体现最优控制要求,又能反映不确定特性,从而实现非线性系统的鲁棒 ADP 控制^[57-59, 78]。Zhang 等通过在 ADP 控制器上添加鲁棒控制项的方法,实现非线性系统的鲁棒控制^[79]。文献 [80] 通过将 ADP 理论和现代控制理论相结合,提出基于 ADP 技术的滑模控制方法。

为了放松对系统模型完全或者部分已知的要求,Zhang 等提出了一个数据驱动的鲁棒近似最优跟踪策略^[54]。通过可获得的数据建立数据驱动模型,用来重构系统。然后在建立的数据驱动模型上采用在线 ADP 方法求解近似最优解,并且设计了一个新型的鲁棒项,保证跟踪误差渐近收敛。文献 [81] 和文献 [82] 提出了基于神经网络观测器的在线 ADP 算法,通过对未知模型的估计,建立系统的估计模型,并在此基础上设计在线 ADP 控制器。Zhu 等利用神经网络逼近器分别对控制输入、干扰和值函数进行估计,实现基于数据驱动的 ADP 零和微分对策控制^[83]。文献 [84] 采用输入输出数据,提出了一种多重评价器-执行器在线 ADP 算法,用于解决模型未知非线性系统的最优控制问题。

同理,作为一种在线优化方法,ADP 理论被认为是求解微分对策纳什均衡解的一种有效、便捷的方法,并被广泛应用于微分博弈问题中。Vrabie 等采用积分强化学习方法,提出一种在线 ADP 方法求解纳什均衡解,该方法仅要求系统的部分模型。文献 [85] 提出策略同步更新算法求解鞍点,采用一个评价器和两个执行器同步在线调整,即控制律部分与干扰部分同时更新参数。考虑系统存在外界干扰,文献 [86] 采用数据驱动方法,建立微分对策数据模型,通过执行网络和评价网络在线求解鞍点值。Lewis 等设计同步策略迭代算法求解非线性系统的多人非零和博弈问题^[87],并将其推广到多人合作博弈问题,设计了分布式自适应学习算法^[88-89]。Zhang 等利用模糊自适应动态规划技术,结合多人微分对策理论,研究了多智能体一致性最优控制问题^[90]。利用 ADP 方法研究多智能控制问题已成为研究热点^[91-94]。文献 [95] 采用数据驱动强化学习方法,研究了未知动态特性的多智能体最优一致性控制问题。

2 ADP 的航空航天应用

由于 ADP 方法的独特算法和结构,使其在解决复杂非线性系统最优控制问题时表现出巨大潜力。因此,基于 ADP 方法的最优控制已经在电力系统^[96-99]、智能交通控制^[100-102]和工业控制^[103-105]等领域取得了大量成功的应用。

一直以来,人们不断尝试将 ADP 方法应用于具有强耦合、复杂非线性以及高复杂特性的航空航天飞行器中。然而,遗憾的是,有关 ADP 理论在航空航天方面的应用成果很少,在导弹制导律的设计中的应用更是至今鲜见。本节主要对现有相关文献作具体详细的分析与概述。Nodland 等利用 ADP 思想设计了自适应优化的输出反馈控制器^[106],实现了无人直升机的轨迹跟踪控制。该方法没有采用策略迭代或值迭代思想,而结合 Backstepping 技术和

神经网络观测器, 通过设计单网络结构在线逼近最优控制器, 并通过 Lyapunov 方法保证闭环跟踪系统的稳定性. Lewis 等考虑存在多种干扰和模型不确定条件下的复杂非线性四旋翼直升机的控制问题, 通过构建单一评价网络和多重执行网络结构在线学习控制参数, 将 ADP 方法应用于直升机控制问题中, 从而有效避免维数灾问题^[107]. Xie 等采用基于核函数的双重启发式动态规划结构^[108], 实现了飞行器的自动降落控制问题. 文献 [109] 结合滑模控制和 ADP 方法, 设计了数据驱动形式的补充控制器, 解决了高超声速飞行器的跟踪控制问题. 该控制器中的 ADP 控制部分主要针对滑模控制下的跟踪误差进行补充控制, 同时在线调节控制器参数, 使之能够满足不同条件下的参数不确定性和干扰, 从而保证闭环系统的稳定性, 值得注意的是该控制器不依赖于系统精确模型, 而是采用数据驱动方式.

由于 ADP 技术的发展, 出现了很多同义词. 早期的有关研究大都不是直接以 ADP 方法命名, 但其基本思想一致. 其中, 文献 [110] 针对飞行器最优控制问题, 提出了基于自适应评价结构的神经网络控制方法. 该控制器包含一个评价网络和一个执行网络, 并且不需要系统精确的数学模型. 文献 [111] 和文献 [112] 以 AH-64 阿帕奇武装直升机为研究背景, 考虑其复杂非线性数学模型难以精确获得, 且存在多种未知干扰, 设计了一种神经元动态规划控制器. 通过在线学习方式, 实现对该直升机的精确控制, 并表现出比较好的鲁棒性能. 随后, Ferrari 等针对六自由度飞行器, 提出了基于神经网络的非线性控制. 该控制器采用一种新型的离线训练技术来训练控制器, 同时采用自适应评价网络在线优化性能指标函数. 通过对该飞行器的数值仿真表现出很强的鲁棒性^[113]. 文献 [114] 研究了变形翼无人飞行器的改进型自适应强化学习控制算法. 该方法将强化学习变形翼控制函数与自适应动态逆控制轨迹跟踪函数相结合, 利用基于 Q-学习方法的无监督学习算法取代经典的评价器-执行器网络结构. 最后利用基于伽辽金法 (Galerkin method) 的分散式数据逼近技术实现目标函数的逼近. 文献 [115] 采用 H_∞ 技术实现高超声速飞行器的连续非线性控制. 为了求解 H_∞ 控制技术中的耦合 Hamilton-Jacobi-Isaacs (HJI) 方程, 文中采用在线同步策略迭代技术实现其在线求解. 文献 [116] 针对高超声速飞行器的纵向控制, 提出了直接启发式动态规划方法. 文中采用模糊神经网络技术来强化该算法的自学习能力和鲁棒性, 且不需要已知系统数学模型, 而是利用在线学习方式解决动态问题. 文献 [117] 采用终端多重滑模控制算法实现星际机器人精准着陆导航问题, 然而, 经研究发现, 该控制器性能的好坏很大程度上取决于控

制器中制导参数的选择. 因此采用强化学习的思想优化与调节制导参数, 并取得了良好的效果. Zhou 等考虑复杂的飞行控制系统一般难以获得其数学模型, 通过权衡 ADP 控制算法的求解复杂度和系统性设计思想, 提出了增量式 ADP 控制方法. 该方法通过对原飞行器非线性模型变换, 得到其线性化的增量式模型, 然后基于模型自由的线性 ADP 技术进行控制^[118-119].

导弹作为一种特殊的飞行器, 表现出更加复杂的动态特性, 对其控制性能要求更严格, 制导精度要求更高. 下面将从导弹制导律的设计方法出发, 阐述制导律的优化设计思想.

3 导弹制导律优化设计方法

导弹制导规律即导引律是实现精确打击或拦截导引的火控系统的关键技术之一. 因此, 导引律的选择至关重要, 它直接影响着导弹是否可以精确打击目标或者成功拦截. 导引律是根据导弹-目标的相对位置、速度和加速度等基本信息, 引导导弹接近目标, 实施攻击. 目前比较典型的制导方式主要有: 比例导引律^[120]、最优制导律^[121]以及滑模变结构制导律^[122]等. 其中, 比例导引律在工程应用中最为常见, 也易于实现, 是目前应用最广泛的导引方法. 但应用比例导引法时, 导弹命中目标时所需的法线过载与导弹速度和攻击角有直接的关系. 当导弹发射角较大或者目标具有强机动性时, 它的性能将急剧下降. 随着现代科学技术的发展, 导弹正在迈向更精确化、更灵巧化、更智能化、更远程化的新型发展方向. 因此, 比例导引律似乎已经力不从心. 基于现代控制理论设计的最优制导律、滑模制导律等得到广大学者的青睐. 由于滑模控制对参数的不确定性以及外部干扰具有良好的自适应性和鲁棒性, 促使其在导弹制导控制中被广泛应用^[122-124]. 但滑模制导律在设计过程中, 其制导参数的选取比较困难, 且在制导末端, 容易发生抖振现象, 亦没有考虑性能指标的最优特性来保证导弹能够在中远程距离执行任务.

最优制导律是通过最优控制理论推导出的导引律, 它不仅能够考虑导弹-目标的动力学问题, 而且可以考虑制导过程起点或终点的约束条件, 例如终端脱靶量最小、最短时间和最小控制能量等要求, 根据给定的不同性能指标, 推导出最优制导规律^[125]. 它主要分为两大类: 线性最优制导律和非线性最优制导律. 其中线性最优制导律是在导弹与目标非线性拦截几何关系线性化基础上, 以二次型性能指标推导出来的, 它以状态反馈的形式给出. 20 世纪 70 年代, Kishi 等首次将最优控制理论应用于导弹制导律的设计中, 发表“次优与最优比例导引”一文, 揭开了最优导引律的序幕. 此后, 就如何利用

最优控制理论研究性能最优的导引律得到广泛的研究^[121, 126-128]。随着现代作战环境的恶化与战斗节奏的加快, 要求导弹以最小脱靶量命中目标还不够, 还希望能够以一定的角度攻击目标, 从而使战斗部发挥最大效能, 取得最大的毁伤效果。随着 Kim 于 1973 年首次推导出带落角约束形式的末制导律以来, 国内外学者纷纷开始对具有约束条件下的制导律进行研究, 并取得了丰硕的成果。Taub 等应用线性二次型最优控制理论, 针对任意时变加速度约束的导弹, 研究了满足落角约束的次优制导律^[129]。陈克俊^[130]、赵汉元^[131]等在针对高超声速再入飞行器末端攻击地面目标时, 针对落角约束问题, 将三维空间的制导问题转化为俯冲平面和转弯平面上的最优控制问题, 在两个平面上分别设计最优制导律。Lee 等设计了同时满足终端角度和加速度约束的最优制导律^[132]。文献 [133] 推导出同时满足终端角度和飞行时间约束的最优制导律。由最优控制理论推导出的最优导引律在理想情况下具有良好的性能。当系统存在不确定参数摄动或外界干扰时, 最优导引律将产生比较大的偏差。为了提高导引律对外界干扰的自适应性和鲁棒性, 胡正东等结合最优制导方法和变结构制导方法, 设计了一种带落角约束的再入机动弹头的复合导引律^[134]。该导引律在最优导引律的基础上引入滑模变结构控制, 增强导引律的鲁棒性。

最优导引律一般是在假设目标运动已知的情况下得到的, 而实际上拦截目标都会采取一定的规避动作, 避开导弹的攻击, 这类问题就变成了研究双方最优策略问题, 因此产生了以微分对策理论为基础的导引律。与最优导引律相比, 微分对策导引律是一种真正的双方动态控制。文献 [135] 以拦截机动目标为背景, 设计了基于 SDR 方法的微分对策制导律, 并与线性二次型微分对策 (LQDG) 制导律作了比较。李登峰等对微分对策导引律进行了大量深入的研究, 并作了一定的总结^[136]。孙传鹏在其论文中应用博弈论思想, 研究了拦截制导问题, 以及多飞行器协同制导方法^[18]。

然而, 上述最优制导方法大部分是基于线性模型发展而来。在很多情况下, 这种线性模型与实际模型存在较大差异, 造成制导精度显著下降。随着高性能导弹的不断问世, 导弹模型表现出的复杂非线性、强耦合、不确定性干扰等特点越来越突出, 其精确数学模型越来越难以获得。Yang 等通过求解 Hamilton-Jacobi (HJ) 偏微分不等式得到制导段非线性模型的 H_∞ 鲁棒制导律^[137]。然而, 基于 H_∞ 鲁棒理论的制导律需要求解偏微分不等式, 而此偏微分不等式的解析解很难用解析方法得到。总之, 不论是基于非线性最优控制理论还是基于非线性微分对策理论

设计制导律, 都不可避免地需要求解相应 Hamilton-Jacobi-Bellman (HJB) 或者 HJI 方程。而这两个方程均属于非线性偏微分方程, 要得到它们的解析解非常困难, 给工程应用造成了很大麻烦, 极大地限制了优化制导律的发展。因此, 如何快速、有效地求解 HJB 或 HJI 方程的解, 甚至可以不依赖于系统模型, 实现智能化学习制导, 对于非线性优化制导律的发展至关重要。

ADP 技术是结合最优控制、自适应控制、强化学习、智能控制等先进控制理论发展而来的一种新型近似最优控制方法。它能够利用函数近似结构来近似 HJB 或 HJI 方程的解, 采用离线或在线更新的方法, 获得系统的近似最优控制策略, 从而能够有效解决非线性系统的优化控制问题。由于 ADP 技术独特的优势, 广大学者一直致力于将这种更加智能化的优化控制技术应用于导弹制导律的设计中。经过多年的努力, 已经取得了一定的成绩。文献 [138] 提出了基于自适应评价结构的最优制导律设计方法, 并与传统最优制导律进行对比, 表现出较强的鲁棒性。Han 等基于导弹输入受约束情况, 采用自适应评价网络设计制导律, 从而保证敏捷导弹能够在最小的时间内由初始给定的马赫数增加到终端确定的马赫数, 同时保证该导弹航迹角的反向^[139-140]。随后, 利用自适应评价网络结构, 研究了中段制导律的设计^[141]。在文献 [142] 中, Han 等考虑敏捷导弹最小飞行马赫数要求, 即状态受约束, 设计了基于自适应评价网络的制导律, 实现导弹在最小时间内由变化的初始马赫数增加到终端马赫数, 同时保证其航迹角的反向。Bertsekas 等将包含动态资源分配的导弹防御问题描述为一个动态规划或马尔科夫决策过程, 通过利用神经元动态规划技术实现该问题的求解^[143]。文献 [144] 研究了针对来袭导弹, 如何选取防御武器发射数量的问题, 并将其描述为一个马尔科夫决策过程, 利用 ADP 技术求解最优策略。Lin 研究倾斜转弯导弹的自适应评价自动驾驶仪设计方法。在自适应评价结构的设计过程中, 采用基于模糊-神经网络的关联搜索单元来逼近导弹的复杂非线性函数, 而自适应评价单元产生强化学习信号来调节关联搜索单元。文中设计的制导律能够消除逼近误差以及干扰影响, 并通过在线学习制导参数从而有效缩短调节时间^[145]。文献 [146] 介绍了强化学习在空空导弹精确制导律设计中的应用。文献 [147] 研究了无人飞行器的一对一空战问题, 利用 ADP 技术逼近最优控制策略。通过一种拥有快速性和有效性的策略提取方法在线完成该制导任务。文献 [148] 基于强化学习方法, 设计了寻的制导律, 并保证其最优性, 论文表明该方法明显优于传统比例导引法。Lee 等研究了导弹-目标追逃问题, 其中导弹采用纯比

例引导方法, 而目标则利用强化学习算法实现躲避策略^[149]. Sun 等考虑导弹-目标质点模型的相对运动关系, 结合微分对策思想, 描述了非线性系统存在不确定项时的鲁棒最优控制问题, 利用在线 ADP 技术, 实现了拦截制导问题基于微分对策的在线 ADP 制导控制^[150].

从以上分析可以看到, 不论是基于强化学习的制导律设计, 亦或是基于自适应评价结构的制导律设计还是神经元动态规划制导律设计, 均属于 ADP 制导问题. 因此, 将 ADP 技术应用于制导律的设计中, 不仅可以解决最优制导律设计中的技术难题, 而且对发展更加智能化、灵巧化、精确化的新型制导系统具有重要的理论价值和深远的实际意义, 为制导系统的开发提供一条新的研究思路.

4 可能的研究方向及展望

ADP 技术自提出以来就受到工作和科研领域人员的关注和重视. 在理论研究和实际应用中都取得了重要的成果, 并且展现出巨大的发展潜力. 然而在航空航天领域, 特别是在线 ADP 技术在导弹制导方面却鲜有涉及. 作为一种近似最优控制方法, ADP 技术在导弹制导律设计中可以发挥不可限量的作用. 下面简单介绍 ADP 技术在导弹制导律设计中可能发挥的作用以及发展趋势:

1) 多约束条件下的 ADP 制导. 随着战场环境的日益复杂, 为了有效提高导弹的战斗能力, 在研究制导控制问题时应该考虑一些约束条件. 从控制的角度来说, 约束主要分为两类: 一是控制约束, 即对输入的约束. 例如过载的饱和特性、姿态控制中气动舵的饱和特性等; 二是状态约束, 即在制导控制的某些时刻或者全过程中, 对某些状态变量的约束. 例如终端角度约束、终端速度约束、弹着时间约束等. 约束条件下的制导控制问题实质上是状态变量或控制输入受多种不同形式约束的一类非线性系统的控制问题^[151], 而非线性最优控制是解决约束控制问题的有力工具. ADP 制导不仅能够解决非线性最优制导求解困难的问题, 而且能够有效地处理多种约束条件下的制导问题. 目前, ADP 方法有关输入受约束的理论研究比较多, 但均属于简单的定常输入约束, 而无法处理复杂形式的约束 (例如时变的区间型或函数型约束), 对状态受约束问题更是研究甚少. 因此发展多约束条件下的 ADP 制导是十分有必要的.

2) 输入不理想条件下的 ADP 制导. 在实际制导控制过程中, 经常会出现执行机构卡死、损坏等故障, 而不能提供理想控制输入的情况. 目前, 处理执行器故障问题主要采用容错控制方法, 而有关 ADP 容错控制的理论研究成果目前很少. 因此, 开展基于

容错控制的 ADP 制导问题研究不仅能够丰富 ADP 理论的应用范围, 而且对输入不理想条件下的 ADP 制导具有重要的意义.

3) 有限时间 ADP 制导. 由于在实际制导过程中, 通常需要导弹能够在有限时间内击中目标, 从而保证打击效果. 而目前 ADP 绝大部分理论都是基于无限时间最优控制. 有限时间 ADP 控制问题的理论研究仍然是一个难点.

4) ADP 协同制导. 随着战场环境的复杂以及未来战争的需要, 发展多导弹协同制导系统已经势不可挡. 目前最优控制和微分对策在多弹协同制导过程中被广泛应用, 但就目前的研究现状而言, 大多以末制导初始时刻的视线为基准进行线性化, 在惯性直角坐标系下采用线性模型研究协同制导问题. 当面对具有大机动能力的高速飞行目标时, 线性假设已不再成立, 由此导致脱靶量大大增加. 而 ADP 方法能够有效解决复杂非线性系统的最优控制问题, 因此, 考虑更贴合实际的非线性模型, 研究基于 ADP 技术的多导弹协同制导问题将是一个十分有益的探索.

5 结论

自适应动态规划技术是解决非线性系统最优控制问题的有效、便捷工具, 受到越来越多研究者的关注与重视. 本文梳理了自适应动态规划技术的理论研究进展, 重点介绍了其在处理不同控制问题时的应用, 分析和总结了现有导引律设计的优缺点, 并着重介绍了最优导引律和微分对策导引律在导弹制导中的应用. 自适应动态规划技术本质上属于最优控制领域, 因此, 研究基于自适应动态规划技术的导弹制导律不仅能够有效处理非线性最优制导律求解困难等问题, 而且为实现导弹的智能化、精确化、灵巧化提供了坚实的理论基础, 具有较高的应用价值.

References

- 1 Zhang H G, Zhang X, Luo Y H, Yang J. An overview of research on adaptive dynamic programming. *Acta Automatica Sinica*, 2013, **39**(4): 303-311
- 2 Liu D R, Li H L, Wang D. Data-based self-learning optimal control: research progress and prospects. *Acta Automatica Sinica*, 2013, **39**(11): 1858-1870
- 3 Werbos P. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences [Ph.D. dissertation], Harvard University, USA, 1974.
- 4 Prokhorov D V, Wunsch D C. Adaptive critic designs. *IEEE Transactions on Neural Networks*, 1997, **8**(5): 997-1007
- 5 Padhi R, Unnikrishnan N, Wang X H, Balakrishnan S N. A single network adaptive critic (SNAC) architecture for

- optimal control synthesis for a class of nonlinear systems. *Neural Networks*, 2006, **19**(10): 1648–1660
- 6 Wang Y, O'Donoghue B, Boyd S. Approximate dynamic programming via iterated Bellman inequalities. *International Journal of Robust and Nonlinear Control*, 2015, **25**(10): 1472–1496
 - 7 Bertsekas D P, Tsitsiklis J N. Neuro-dynamic programming: an overview. In: *Proceedings of the 34th IEEE Conference on Decision and Control*. New Orleans, LA, USA: IEEE, 1995. 560–564
 - 8 Zhu L M, Modares H, Peen G O, Lewis F L, Yue B Z. Adaptive suboptimal output-feedback control for linear systems using integral reinforcement learning. *IEEE Transactions on Control Systems Technology*, 2015, **23**(1): 264–273
 - 9 Bhasin S. Reinforcement Learning and Optimal Control Methods for Uncertain Nonlinear Systems [Ph.D. dissertation], University of Florida, USA, 2011.
 - 10 Vrabie D, Vamvoudakis K G, Lewis F L. *Optimal Adaptive Control and Differential Games by Reinforcement Learning Principles*. London: IET, 2012.
 - 11 Zhang H, Liu D, Luo Y, Wang D. *Adaptive Dynamic Programming for Control: Algorithms and Stability*. London: Springer-Verlag, 2013.
 - 12 Lewis F L, Liu D R. *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*. New Jersey: IEEE Press, 2013.
 - 13 Jiang Z P, Jiang Y. Robust adaptive dynamic programming for linear and nonlinear systems: an overview. *European Journal of Control*, 2013, **19**(5): 417–425
 - 14 Khan S G, Herrmann G, Lewis F L, Pipe T, Melhuish C. Reinforcement learning and optimal adaptive control: an overview and implementation examples. *Annual Reviews in Control*, 2012, **36**(1): 42–59
 - 15 Buşoniu L, Ernst D, De Schutter B, Babuška R. Approximate reinforcement learning: an overview. In: *Proceedings of the 2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning*. Paris, France: IEEE, 2011. 1–8
 - 16 Diao Zhao-Shi. Research on High-precision and High-efficiency Terminal Guidance and Control Key Technologies for Missiles [Ph.D. dissertation], Beijing Institute of Technology, China, 2015.
(刁兆师. 导弹精确高效末制导与控制若干关键技术研究 [博士学位论文], 北京理工大学, 中国, 2015.)
 - 17 Li Yun-Qian. Integrated Guidance and Control for Endo-Atmospheric Interceptors [Ph.D. dissertation], Harbin Institute of Technology, China, 2011.
(李运迁. 大气层内拦截弹制导控制及一体化研究 [博士学位论文], 哈尔滨工业大学, 中国, 2011.)
 - 18 Sun Chuan-Peng. Research on Interception Guidance Based on Game Theory [Ph.D. dissertation], Harbin Institute of Technology, China, 2014.
(孙传鹏. 基于博弈论的拦截制导问题研究 [博士学位论文], 哈尔滨工业大学, 中国, 2014.)
 - 19 Liu D R, Wei Q L. Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, **25**(3): 621–634
 - 20 Wei Q L, Liu D R, Lin Q, Song R Z. Discrete-time optimal control via local policy iteration adaptive dynamic programming. *IEEE Transactions on Cybernetics*, 2016, DOI: 10.1109/TCYB.2016.2586082
 - 21 Zhang H G, Song R Z, Wei Q L, Zhang T Y. Optimal tracking control for a class of nonlinear discrete-time systems with time delays based on heuristic dynamic programming. *IEEE Transactions on Neural Networks*, 2011, **22**(12): 1851–1862
 - 22 Song R Z, Xiao W D, Zhang H G, Sun C Y. Adaptive dynamic programming for a class of complex-valued nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, **25**(9): 1733–1739
 - 23 Wei Q L, Liu D R, Yang X. Infinite horizon self-learning optimal control of nonaffine discrete-time nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26**(4): 866–879
 - 24 Wei Q L, Liu D R, Lewis F L, Liu Y, Zhang J. Mixed iterative adaptive dynamic programming for optimal battery energy control in smart residential microgrids. *IEEE Transactions on Industrial Electronics*, 2017, DOI: 10.1109/TIE.2017.265087
 - 25 Wei Q L, Liu D R, Lin Q, Song R Z. Adaptive dynamic programming for discrete-time zero-sum games. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, DOI: 10.1109/TNNLS.2016.2638863
 - 26 Wei Q L, Liu D R. A novel policy iteration based deterministic Q-learning for discrete-time nonlinear systems. *Science China Information Sciences*, 2015, **58**(12): 1–15
 - 27 Kiumarsi B, Lewis F L, Modares H, Karimpour A, Naghibi-Sistani M B. Reinforcement Q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics. *Automatica*, 2014, **50**(4): 1167–1175
 - 28 Vamvoudakis K G. Non-zero sum Nash Q-learning for unknown deterministic continuous-time linear systems. *Automatica*, 2015, **61**: 274–281
 - 29 Murray J J, Cox C J, Lendaris G G, Saeks R. Adaptive dynamic programming. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 2002, **32**(2): 140–153
 - 30 Al-Tamimi A, Lewis F L, Abu-Khalaf M. Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2008, **38**(4): 943–949
 - 31 Wang F Y, Jin N, Liu D R, Wei Q L. Adaptive dynamic programming for finite-horizon optimal control of discrete-time nonlinear systems with ϵ -error bound. *IEEE Transactions on Neural Networks*, 2011, **22**(1): 24–36
 - 32 Heydari A, Balakrishnan S N. Finite-horizon control-constrained nonlinear optimal control using single network adaptive critics. *IEEE Transactions on Neural Networks and Learning Systems*, 2013, **24**(1): 145–157

- 33 Wei Q L, Liu D R, Xu Y C. Neuro-optimal tracking control for a class of discrete-time nonlinear systems via generalized value iteration adaptive dynamic programming approach. *Soft Computing*, 2016, **20**(2): 697–706
- 34 Wei Q L, Liu D R, Lin H Q. Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems. *IEEE Transactions on Cybernetics*, 2016, **46**(3): 840–853
- 35 Wei Q L, Lewis F L, Sun Q Y, Yan P F, Song R Z. Discrete-time deterministic Q-learning: a novel convergence analysis. *IEEE Transactions on Cybernetics*, 2017, **47**(5): 1024–1037
- 36 Wei Q L, Song R Z, Sun Q Y. Nonlinear neuro-optimal tracking control via stable iterative Q-learning algorithm. *Neurocomputing*, 2015, **168**: 520–528
- 37 Wei Q L, Liu D R. Stable iterative adaptive dynamic programming algorithm with approximation errors for discrete-time nonlinear systems. *Neural Computing and Applications*, 2014, **24**(6): 1355–1367
- 38 Wei Q L, Liu D R. Adaptive dynamic programming for optimal tracking control of unknown nonlinear systems with application to coal gasification. *IEEE Transactions on Automation Science and Engineering*, 2014, **11**(4): 1020–1036
- 39 Wei Q L, Liu D R. Numerical adaptive learning control scheme for discrete-time non-linear systems. *IET Control Theory and Applications*, 2013, **7**(11): 1472–1486
- 40 Wei Q L, Liu D R, Lin Q. Discrete-time local value iteration adaptive dynamic programming: admissibility and termination analysis. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, DOI: 10.1109/TNNLS.2016.2593743
- 41 Wei Q L, Wang F Y, Liu D R, Yang X. Finite-approximation-error-based discrete-time iterative adaptive dynamic programming. *IEEE Transactions on Cybernetics*, 2014, **44**(12): 2820–2833
- 42 Zhang H G, Luo Y H, Liu D R. Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints. *IEEE Transactions on Neural Networks*, 2009, **20**(9): 1490–1503
- 43 Wei Q L, Zhang H G, Liu D R, Zhao Y. An optimal control scheme for a class of discrete-time nonlinear systems with time delays using adaptive dynamic programming. *Acta Automatica Sinica*, 2010, **36**(1): 121–129
- 44 Song R Z, Wei Q L, Sun Q Y. Nearly finite-horizon optimal control for a class of nonaffine time-delay nonlinear systems based on adaptive dynamic programming. *Neurocomputing*, 2015, **156**: 166–175
- 45 Wang D, Liu D R, Wei Q L. Finite-horizon neuro-optimal tracking control for a class of discrete-time nonlinear systems using adaptive dynamic programming approach. *Neurocomputing*, 2012, **78**(1): 14–22
- 46 Abu-Khalaf M, Lewis F L. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. *Automatica*, 2005, **41**(5): 779–791
- 47 Tassa Y, Erez T. Least squares solutions of the HJB equation with neural network value-function approximators. *IEEE Transactions on Neural Networks*, 2007, **18**(4): 1031–1041
- 48 Song R Z, Lewis F L, Wei Q L, Zhang H G. Off-policy actor-critic structure for optimal control of unknown systems with disturbances. *IEEE Transactions on Cybernetics*, 2016, **46**(5): 1041–1050
- 49 Song R Z, Lewis F L, Wei Q L. Off-policy integral reinforcement learning method to solve nonlinear continuous-time multiplayer nonzero-sum games. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, **28**(3): 704–713
- 50 Vrabie D, Pastravanu O, Abu-Khalaf M, Lewis F L. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica*, 2009, **45**(2): 477–484
- 51 Vrabie D, Lewis F. Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems. *Neural Networks*, 2009, **22**(3): 237–246
- 52 Vamvoudakis K G, Lewis F L. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica*, 2010, **46**(5): 878–888
- 53 Vamvoudakis K G, Vrabie D, Lewis F L. Online adaptive learning of optimal control solutions using integral reinforcement learning. In: *Proceedings of the 2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning*. Paris, France: IEEE, 2011. 250–257
- 54 Zhang H G, Cui L L, Zhang X, Luo Y H. Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method. *IEEE Transactions on Neural Networks*, 2011, **22**(12): 2226–2236
- 55 Liu D R, Yang X, Wang D, Wei Q L. Reinforcement-learning-based robust controller design for continuous-time uncertain nonlinear systems subject to input constraints. *IEEE Transactions on Cybernetics*, 2015, **45**(7): 1372–1385
- 56 Yang X, Liu D R, Huang Y Z. Neural-network-based online optimal control for uncertain non-linear continuous-time systems with control constraints. *IET Control Theory and Applications*, 2013, **7**(17): 2037–2047
- 57 Wang D, Liu D R, Zhang Q C, Zhao D B. Data-based adaptive critic designs for nonlinear robust optimal control with uncertain dynamics. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2016, **46**(11): 1544–1555
- 58 Wang D, Liu D R, Li H L. Policy iteration algorithm for online design of robust control for a class of continuous-time nonlinear systems. *IEEE Transactions on Automation Science and Engineering*, 2014, **11**(2): 627–632
- 59 Wang D, Liu D R, Li H L, Ma H W. Neural-network-based robust optimal control design for a class of uncertain nonlinear systems via adaptive dynamic programming. *Information Sciences*, 2014, **282**: 167–179

- 60 Vamvoudakis K, Vrabie D, Lewis F. Online policy iteration based algorithms to solve the continuous-time infinite horizon optimal control problem. In: Proceedings of the 2009 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning. Nashville, TN, USA: IEEE, 2009.
- 61 Yang X, Liu D R, Wei Q L. Online approximate optimal control for affine non-linear systems with unknown internal dynamics using adaptive dynamic programming. *IET Control Theory and Applications*, 2014, **8**(16): 1676–1688
- 62 Wang Z, Liu X P, Liu K F, Li S, Wang H Q. Backstepping-based Lyapunov function construction using approximate dynamic programming and sum of square techniques. *IEEE Transactions on Cybernetics*, 2016, DOI: 10.1109/TCYB.2016.2574747
- 63 Vamvoudakis K G, Miranda M F, Hespanha J P. Asymptotically stable adaptive-optimal control algorithm with saturating actuators and relaxed persistence of excitation. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, **27**(11): 2386–2398
- 64 Yang X, Liu D R, Wei Q L, Wang D. Guaranteed cost neural tracking control for a class of uncertain nonlinear systems using adaptive dynamic programming. *Neurocomputing*, 2016, **198**: 80–90
- 65 Zargarzadeh H, Dierks T, Jagannathan S. State and output feedback-based adaptive optimal control of nonlinear continuous-time systems in strict feedback form. In: Proceedings of the 2012 American Control Conference. Montréal, Canada: IEEE, 2012. 6412–6417
- 66 Zargarzadeh H, Dierks T, Jagannathan S. Optimal control of nonlinear continuous-time systems in strict-feedback form. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26**(10): 2535–2549
- 67 Modares H, Lewis F L. Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning. *Automatica*, 2014, **50**(7): 1780–1792
- 68 Kamalapurkar R, Dinh H, Bhasin S, Dixon W E. Approximate optimal trajectory tracking for continuous-time nonlinear systems. *Automatica*, 2015, **51**: 40–48
- 69 Zhou Q, Shi P, Tian Y, Wang M Y. Approximation-based adaptive tracking control for mimo nonlinear systems with input saturation. *IEEE Transactions on Cybernetics*, 2015, **45**(10): 2119–2128
- 70 Modares H, Lewis F L, Sistani M B N. Online solution of nonquadratic two-player zero-sum games arising in the H_∞ control of constrained input systems. *International Journal of Adaptive Control and Signal Processing*, 2014, **28**(3–5): 232–254
- 71 Modares H, Sistani M B N, Lewis F L. A policy iteration approach to online optimal control of continuous-time constrained-input systems. *ISA Transactions*, 2013, **52**(5): 611–621
- 72 Abu-Khalaf M, Lewis F L, Huang J. Neurodynamic programming and zero-sum games for constrained control systems. *IEEE Transactions on Neural Networks*, 2008, **19**(7): 1243–1252
- 73 Yang X, Liu D R, Wang D. Reinforcement learning for adaptive optimal control of unknown continuous-time nonlinear systems with input constraints. *International Journal of Control*, 2014, **87**(3): 553–566
- 74 Jiang Y, Jiang Z P. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. *Automatica*, 2012, **48**(10): 2699–2704
- 75 Jiang Y, Jiang Z P. Robust approximate dynamic programming and global stabilization with nonlinear dynamic uncertainties. In: Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference. Orlando, FL, USA: IEEE, 2011. 115–120
- 76 Jiang Y, Jiang Z P. Robust adaptive dynamic programming and feedback stabilization of nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, **25**(5): 882–893
- 77 Jiang Y, Jiang Z P. Global adaptive dynamic programming for continuous-time nonlinear systems. *IEEE Transactions on Automatic Control*, 2015, **60**(11): 2917–2929
- 78 Wang D, Liu D R, Li H L, Ma H W. Adaptive dynamic programming for infinite horizon optimal robust guaranteed cost control of a class of uncertain nonlinear systems. In: Proceedings of the 2015 American Control Conference. Chicago, IL, USA: IEEE, 2015. 2900–2905
- 79 Luo Y H, Sun Q Y, Zhang H G, Cui L L. Adaptive critic design-based robust neural network control for nonlinear distributed parameter systems with unknown dynamics. *Neurocomputing*, 2015, **148**: 200–208
- 80 Fan Q Y, Yang G H. Adaptive actor-critic design-based integral sliding-mode control for partially unknown nonlinear systems with input disturbances. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, **27**(1): 165–177
- 81 Liu D R, Huang Y Z, Wang D, Wei Q L. Neural-network-observer-based optimal control for unknown nonlinear systems using adaptive dynamic programming. *International Journal of Control*, 2013, **86**(9): 1554–1566
- 82 Lv Y F, Na J, Yang Q M, Wu X, Guo Y. Online adaptive optimal control for continuous-time nonlinear systems with completely unknown dynamics. *International Journal of Control*, 2016, **89**(1): 99–112
- 83 Zhu Y H, Zhao D B, Li X J. Iterative adaptive dynamic programming for solving unknown nonlinear zero-sum game based on online data. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, **28**(3): 714–724
- 84 Song R Z, Lewis F L, Wei Q L, Zhang H G, Jiang Z P, Levine D. Multiple actor-critic structures for continuous-time optimal control using input-output data. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26**(4): 851–865
- 85 Vamvoudakis K, Vrabie D, Lewis F L. Adaptive optimal control algorithm for zero-sum Nash games with integral reinforcement learning. In: Proceedings of the 2012 AIAA Guidance, Navigation, and Control Conference. Minneapolis, Minnesota, USA: AIAA, 2012.

- 86 Wei Q L, Song R Z, Yan P F. Data-driven zero-sum neuro-optimal control for a class of continuous-time unknown nonlinear systems with disturbance using ADP. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, **27**(2): 444–458
- 87 Vamvoudakis K G, Lewis F L. Multi-player non-zero-sum games: online adaptive learning solution of coupled Hamilton-Jacobi equations. *Automatica*, 2011, **47**(8): 1556–1569
- 88 Vamvoudakis K G, Lewis F L, Hudas G R. Multi-agent differential graphical games: online adaptive learning solution for synchronization with optimality. *Automatica*, 2012, **48**(8): 1598–1611
- 89 Wei Q L, Liu D R, Lewis F L. Optimal distributed synchronization control for continuous-time heterogeneous multi-agent differential graphical games. *Information Sciences*, 2015, **317**: 96–113
- 90 Zhang H G, Zhang J L, Yang G H, Luo Y H. Leader-based optimal coordination control for the consensus problem of multiagent differential games via fuzzy adaptive dynamic programming. *IEEE Transactions on Fuzzy Systems*, 2015, **23**(1): 152–163
- 91 Nguyen T L. Adaptive dynamic programming-based design of integrated neural network structure for cooperative control of multiple MIMO nonlinear systems. *Neurocomputing*, 2017, **237**: 12–24
- 92 Jiao Q, Modares H, Xu S Y, Lewis F L, Vamvoudakis K G. Multi-agent zero-sum differential graphical games for disturbance rejection in distributed control. *Automatica*, 2016, **69**: 24–34
- 93 Jiao Q, Modares H, Lewis F L, Xu S Y, Xie L H. Distributed L_2 -gain output-feedback control of homogeneous and heterogeneous systems. *Automatica*, 2016, **71**: 361–368
- 94 Adib Yaghmaie F, Lewis F L, Su R. Output regulation of linear heterogeneous multi-agent systems via output and state feedback. *Automatica*, 2016, **67**: 157–164
- 95 Zhang H G, Jiang H, Luo Y H, Xiao G Y. Data-driven optimal consensus control for discrete-time multi-agent systems with unknown dynamics using reinforcement learning method. *IEEE Transactions on Industrial Electronics*, 2017, **64**(5): 4091–4100
- 96 Venayagamoorthy G K, Harley R G, Wunsch D C. Dual heuristic programming excitation neurocontrol for generators in a multimachine power system. *IEEE Transactions on Industry Applications*, 2003, **39**(2): 382–394
- 97 Park J W, Harley R G, Venayagamoorthy G K. Adaptive-critic-based optimal neurocontrol for synchronous generators in a power system using MLP/RBF neural networks. *IEEE Transactions on Industry Applications*, 2003, **39**(5): 1529–1540
- 98 Wei Q L, Liu D R, Shi G, Liu Y. Multibattery optimal coordination control for home energy management systems via distributed iterative adaptive dynamic programming. *IEEE Transactions on Industrial Electronics*, 2015, **62**(7): 4203–4214
- 99 Wei Q L, Liu D R, Shi G. A novel dual iterative Q-learning method for optimal battery management in smart residential environments. *IEEE Transactions on Industrial Electronics*, 2015, **62**(4): 2509–2518
- 100 Cai C, Wong C K, Heydecker B G. Adaptive traffic signal control using approximate dynamic programming. *Transportation Research Part C: Emerging Technologies*, 2009, **17**(5): 456–474
- 101 Zhao Dong-Bin, Liu De-Rong, Yi Jian-Qiang. An overview on the adaptive dynamic programming based urban city traffic signal optimal control. *Acta Automatica Sinica*, 2009, **35**(6): 676–681
(赵冬斌, 刘德荣, 易建强. 基于自适应动态规划的城市交通信号优化控制方法综述. *自动化学报*, 2009, **35**(6): 676–681)
- 102 Wang F Y. Agent-based control for networked traffic management systems. *IEEE Intelligent Systems*, 2005, **20**(5): 92–96
- 103 Lee J M, Lee J H. An approximate dynamic programming based approach to dual adaptive control. *Journal of Process Control*, 2009, **19**(5): 859–864
- 104 Wei Q L, Liu D R. Data-driven neuro-optimal temperature control of water-gas shift reaction using stable iterative adaptive dynamic programming. *IEEE Transactions on Industrial Electronics*, 2014, **61**(11): 6399–6408
- 105 Lin Xiao-Feng, Huang Yuan-Jun, Song Chun-Ning. Approximate optimal control with ε -error bound. *Control Theory and Applications*, 2012, **29**(1): 104–108
(林小峰, 黄元君, 宋春宁. 带 ε 误差限的近似最优控制. *控制理论与应用*, 2012, **29**(1): 104–108)
- 106 Nodland D, Zargarzadeh H, Jagannathan S. Neural network-based optimal adaptive output feedback control of a helicopter UAV. *IEEE Transactions on Neural Networks and Learning Systems*, 2013, **24**(7): 1061–1073
- 107 Stingu E, Lewis F L. An approximate dynamic programming based controller for an underactuated 6DOF quadrotor. In: *Proceedings of the 2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning*. Paris, France: IEEE, 2011.
- 108 Xie Q Q, Luo B, Tan F X, Guan X P. Optimal control for vertical take-off and landing aircraft non-linear system by online kernel-based dual heuristic programming learning. *IET Control Theory and Applications*, 2015, **9**(6): 981–987
- 109 Mu C X, Ni Z, Sun C Y, He H B. Air-breathing hypersonic vehicle tracking control based on adaptive dynamic programming. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, **28**(3): 584–598
- 110 Balakrishnan S N, Biega V. Adaptive-critic-based neural networks for aircraft optimal control. *Journal of Guidance, Control, and Dynamics*, 1996, **19**(4): 893–898
- 111 Enns R, Si J. Apache helicopter stabilization using neural dynamic programming. *Journal of Guidance, Control, and Dynamics*, 2002, **25**(1): 19–25
- 112 Enns R, Si J. Helicopter trimming and tracking control using direct neural dynamic programming. *IEEE Transactions on Neural Networks*, 2003, **14**(4): 929–939

- 113 Ferrari S, Stengel R F. Online adaptive critic flight control. *Journal of Guidance, Control, and Dynamics*, 2004, **27**(5): 777–786
- 114 Valasek J, Doebbler J, Tandale M D, Meade A J. Improved adaptive-reinforcement learning control for morphing unmanned air vehicles. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2008, **38**(4): 1014–1020
- 115 Guo C, Wu H N, Luo B, Guo L. H_∞ control for air-breathing hypersonic vehicle based on online simultaneous policy update algorithm. *International Journal of Intelligent Computing and Cybernetics*, 2013, **6**(2): 126–143
- 116 Luo X, Chen Y, Si J, Feng L. Longitudinal control of hypersonic vehicles based on direct heuristic dynamic programming using ANFIS. In: Proceedings of the 2014 International Joint Conference on Neural Networks. Beijing, China: IEEE, 2014. 3685–3692
- 117 Furfaro R, Wibben D R, Gaudet B, Simo J. Terminal multiple surface sliding guidance for planetary landing: development, tuning and optimization via reinforcement learning. *The Journal of the Astronautical Sciences*, 2015, **62**(1): 73–99
- 118 Zhou Y, Van Kampen E J, Chu Q P. Nonlinear adaptive flight control using incremental approximate dynamic programming and output feedback. *Journal of Guidance, Control, and Dynamics*, 2017, **40**(S): 493–500
- 119 Zhou Y, Van Kampen E J, Chu Q P. An incremental approximate dynamic programming flight controller based on output feedback. In: Proceedings of the 2016 AIAA Guidance, Navigation, and Control Conference. San Diego, California, USA: AIAA, 2016.
- 120 Ghosh S, Ghose D, Raha S. Capturability of augmented pure proportional navigation guidance against time-varying target maneuvers. *Journal of Guidance, Control, and Dynamics*, 2014, **37**(5): 1446–1461
- 121 Shaferman V, Shima T. Linear quadratic guidance laws for imposing a terminal intercept angle. *Journal of Guidance, Control, and Dynamics*, 2008, **31**(5): 1400–1412
- 122 Lee Y, Kim Y, Moon G, Jun B E. Sliding-mode-based missile-integrated attitude control schemes considering velocity change. *Journal of Guidance, Control, and Dynamics*, 2016, **39**(3): 423–436
- 123 Kumar S R, Rao S, Ghose D. Nonsingular terminal sliding mode guidance with impact angle constraints. *Journal of Guidance, Control, and Dynamics*, 2014, **37**(4): 1114–1130
- 124 Zhou Hui-Bo. Study on Guidance Law and Cooperative Guidance for Multi-missiles Based on Finite-time and Sliding Mode Theory [Ph.D. dissertation], Harbin Institute of Technology, China, 2015.
(周慧波. 基于有限时间和滑模理论的导引律及多导弹协同制导研究 [博士学位论文], 哈尔滨工业大学, 中国, 2015.)
- 125 Zhang You-An, Huang Jie, Wang Li-Ying. Research progress of terminal guidance law with constraint. *Journal of Naval Aeronautical and Astronautical*, 2013, **28**(6): 581–586
(张友安, 黄洁, 王丽英. 约束条件下的末制导律研究进展. 海军航空工程学院学报, 2013, **28**(6): 581–586)
- 126 Imado F, Kuroda T, Miwa S. Optimal midcourse guidance for medium-range air-to-air missiles. *Journal of Guidance, Control, and Dynamics*, 1990, **13**(4): 603–608
- 127 Balakrishnan S N, Xin M. Robust state dependent Riccati equation based guidance laws. In: Proceedings of the 2001 American Control Conference. Arlington, VA, USA: IEEE, 2001. 3352–3357
- 128 Indig N, Ben-Asher J Z, Sigal E. Near-optimal minimum-time guidance under spatial angular constraint in atmospheric flight. *Journal of Guidance, Control, and Dynamics*, 2016, **39**(7): 1563–1577
- 129 Taub I, Shima T. Intercept angle missile guidance under time varying acceleration bounds. *Journal of Guidance, Control, and Dynamics*, 2013, **36**(3): 686–699
- 130 Chen Ke-Jun, Zhao Han-Yuan. An optimal reentry maneuver guidance law applying to attack the ground fixed target. *Journal of Astronautics*, 1994, **15**(1): 1–7, 94
(陈克俊, 赵汉元. 一种适用于攻击地面固定目标的最优再入机动制导律. 宇航学报, 1994, **15**(1): 1–7, 94)
- 131 Zhao Han-Yuan. *Reentry Vehicle Dynamics and Guidance*. Beijing: National University of Defense Technology, 1997.
(赵汉元. 飞行器再入动力学和制导. 北京: 国防科技大学出版社, 1997.)
- 132 Lee Y I, Ryoo C K, Kim E. Optimal guidance with constraints on impact angle and terminal acceleration. In: Proceedings of the 2003 AIAA Guidance, Navigation, and Control Conference and Exhibit. Austin, Texas, USA: AIAA, 2003.
- 133 Lee J I, Jeon I S, Tahk M J. Guidance law to control impact time and angle. *IEEE Transactions on Aerospace and Electronic Systems*, 2007, **43**(1): 301–310
- 134 Hu Zheng-Dong, Guo Cai-Fa, Cai Hong. Integrated guidance law of reentry maneuvering warhead with terminal angular constraint. *Journal of National University of Defense Technology*, 2008, **30**(3): 21–26
(胡正东, 郭才发, 蔡洪. 带落角约束的再入机动弹头的复合导引律. 国防科技大学学报, 2008, **30**(3): 21–26)
- 135 Bardhan R, Ghose D. Nonlinear differential games-based impact-angle-constrained guidance law. *Journal of Guidance, Control, and Dynamics*, 2015, **38**(3): 384–402
- 136 Fang Shao-Kun, Li Deng-Feng. Research advances on differential games and applications to military field. *Command Control and Simulation*, 2008, **30**(1): 114–117
(方绍琨, 李登峰. 微分对策及其在军事领域的研究进展. 指挥控制与仿真, 2008, **30**(1): 114–117)
- 137 Yang C D, Chen H Y. Nonlinear H_∞ robust guidance law for homing missiles. *Journal of Guidance, Control, and Dynamics*, 1998, **21**(6): 882–890
- 138 Dalton J, Balakrishnan S N. A neighboring optimal adaptive critic for missile guidance. *Mathematical and Computer Modelling*, 1996, **23**(1–2): 175–188
- 139 Han D C, Balakrishnan S N. Adaptive critic based neural networks for control-constrained agile missile control. In: Proceedings of the 1999 American Control Conference. San Diego, California, USA: IEEE, 1999. 2600–2604

- 140 Si J, Barto A, Powell W, Wunsch D. *Adaptive Critic Based Neural Network for Control-Constrained Agile Missile*. New Jersey: John Wiley and Sons, Inc., 2012.
- 141 Han D C, Balakrishnan S. Midcourse guidance law with neural networks. In: Proceedings of the 2000 AIAA Guidance, Navigation, and Control Conference and Exhibit. Denver, CO, USA: AIAA, 2000.
- 142 Han D C, Balakrishnan S N. State-constrained agile missile control with adaptive-critic-based neural networks. *IEEE Transactions on Control Systems Technology*, 2002, **10**(4): 481–489
- 143 Bertsekas D P, Homer M L, Logan D A, Patek S D, Sandell N R. Missile defense and interceptor allocation by neurodynamic programming. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 2000, **30**(1): 42–51
- 144 Davis M T, Robbins M J, Lunday B J. Approximate dynamic programming for missile defense interceptor fire control. *European Journal of Operational Research*, 2017, **259**(3): 873–886
- 145 Lin C K. Adaptive critic autopilot design of bank-to-turn missiles using fuzzy basis function networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2005, **35**(2): 197–207
- 146 Lu Chao-Qun, Jiang Jia-He, Ren Zhang. Research of precision guidance law based on Q-learning for air-to-air missile. *Control Technology of Tactical Missile*, 2006, (4): 19–22, 76
(卢超群, 江加和, 任章. 基于增强学习的空空导弹智能精确制导律研究. *战术导弹控制技术*, 2006, (4): 19–22, 76)
- 147 McGrew J S, How J P, Bush L, Williams B, Roy N. Air combat strategy using approximate dynamic programming. In: Proceedings of the 2008 AIAA Guidance, Navigation and Control Conference and Exhibit. Honolulu, Hawaii, USA: AIAA, 2008.
- 148 Gaudet B, Furfaro R. Missile homing-phase guidance law design using reinforcement learning. In: Proceedings of the 2012 AIAA Guidance, Navigation, and Control Conference. Minneapolis, Minnesota, USA: AIAA, 2012.
- 149 Lee D, Bang H. Planar evasive aircrafts maneuvers using reinforcement learning. *Intelligent Autonomous Systems 12: Advances in Intelligent Systems and Computing*. Berlin Heidelberg: Springer, 2013. 533–542
- 150 Sun J L, Liu C S, Ye Q. Robust differential game guidance laws design for uncertain interceptor-target engagement via adaptive dynamic programming. *International Journal of Control*, 2017, **64**(5): 4091–4100
- 151 Yao Yu, Zheng Tian-Yu, He Feng-Hua, Wang Long, Wang Yang, Zhang Xi, Zhu Bai-Yang, Yang Bao-Qing. Several hot issues and challenges in terminal guidance of flight vehicles. *Acta Aeronautica et Astronautica Sinica*, 2015, **36**(8): 2696–2716
(姚郁, 郑天宇, 贺风华, 王龙, 汪洋, 张曦, 朱柏羊, 杨宝庆. 飞行器末制导中的几个热点问题与挑战. *航空学报*, 2015, **36**(8): 2696–2716)



孙景亮 南京航空航天大学自动化学院博士研究生. 主要研究方向为最优控制, 微分对策, 自适应动态规划.

E-mail: sunjingliangac@163.com

(**SUN Jing-Liang** Ph. D. candidate at the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics. His research interest covers optimal control, differential game, and adaptive dynamic programming.)



刘春生 南京航空航天大学自动化学院教授. 主要研究方向为自适应控制, 最优控制, 故障诊断与容错控制及其在飞行器中的应用. 本文通信作者.

E-mail: liuchsh@nuaa.edu.cn

(**LIU Chun-Sheng** Professor at the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics. Her research interest covers adaptive control, optimal control, fault diagnosis and tolerant control with the application in aircraft. Corresponding author of this paper.)