

分类号	
学校代码	10700
学 号	1170311018

西昌理工大學

# 博士学位论文

不确定系统的自适应对偶控制方法研究

马雪卉

学科门类: 工 学

一级学科: 控制科学与工程

二级学科: 控制理论与控制工程

指导教师: 钱富才 教授

申请日期: 2022年12月

## 独 创 性 声 明

本人所呈交的学位论文是在导师指导下进行的研究工作及取得的成果。尽我所知，除特别加以标注的地方外，论文中不包含其他人的研究成果。与我一同工作的同志对本文的研究工作和成果的任何贡献均已在论文中作了明确的说明并已致谢。

本论文及其相关资料若有不实之处，由本人承担一切相关责任

论文作者签名: 弓雪才 2022年 12月 11日

## 学 位 论 文 使 用 授 权

本人作为学位论文作者了解并愿意遵守学校有关保留、使用学位论文的规定，即：在导师的指导下创作完成的学位论文的知识产权归西安理工大学所有，本人今后在使用或发表该论文涉及的研究内容时，会注明西安理工大学。西安理工大学拥有学位论文的如下使用权，包括：学校可以保存学位论文；可以采用影印、缩印或其他复制手段保存论文；可以查阅或借阅。本人授权西安理工大学对学位论文全部内容编入公开的数据库进行检索。本学位论文全部或部分内容的公布（包括刊登）授权西安理工大学研究生院办理。

涉密的学位论文按照《西安理工大学研究生学位论文涉密认定和管理办法》要求进行密级认定，学校按照密级对学位论文进行分类管理。

保密的学位论文在解密后，适用本授权。

论文作者签名: 弓雪才 导师签名: 钱宣才 2022年 12月 11日

**论文题目：不确定系统的自适应对偶控制方法研究**

**学科名称：控制理论与控制工程**

**研究生：马雪卉**

**签 名： 马雪卉**

**指导老师：钱富才 教授**

**签 名： 钱富才**

## 摘要

实际系统中往往存在着不确定因素，例如系统中的未知参数、非线性环节、结构变化、元器件故障等，传统的基于确定性等价原理的自适应控制设计思想，是将未知参数的估计值视为真值，以此为基础直接求解控制律。这种设计思想并没有考虑到参数估计误差，因而设计出的控制器使得闭环系统远未接近最优。而本文研究的自适应对偶控制方法，不仅要求系统能够谨慎地跟踪目标轨迹，而且还要充分地激励系统去主动学习更多的未知信息，从而使得该控制律更接近最优。尽管已经有一些自适应对偶控制方法可以解决实际系统中存在不确定因素的影响，如参数未知、非线性环节、结构变化等，但除此之外，其他不确定因素，如孤立点噪声、不同分布类型的随机噪声、随机变化的扰动等，目前在自适应对偶控制框架下鲜有研究。

本文围绕影响系统控制性能的上述不确定性因素，开展了具有主动学习特点的自适应对偶控制方法研究，主要研究内容如下：

(1) 针对参数未知的线性系统的控制问题，尽管现有基于强化 Q-学习自适应控制方法，可以通过试错学习的方式获得最优控制律，然而在试错学习阶段，探索信号需要人为设定，一旦探索信号过大引起超调过大，因而难以用于实际系统。针对该问题，提出了一种在试错学习阶段可以主动调节探索信号的方法，一方面能够让系统状态谨慎跟踪目标，另一方面还能够激励系统获取更为丰富的动态信息，以便控制器对未知参数进行主动学习与探索。该方法是通过同时优化系统状态跟踪性能指标和 Q 函数核矩阵估计性能指标来实现，由此计算出的主动调整的探索信号，减少了试错过程中超调过大或者探索不够充分的情况，即对 Q-学习中探索与利用的冲突问题进行了最优平衡，进而提升了控制性能和实用价值。

(2) 在实际不确定系统中，由于传感器失灵、网络数据传输错误、设备损坏、外界入侵等原因，会给系统观测数据以及系统控制过程本身带来孤立点噪声，使得过去仅针对具有高斯白噪声的未知系统的自适应对偶控制方法不再有效。对此问题，本文一方面仅利用实时观测数据设计了孤立点的距离和方向两个判断准则，在数据被用于求解自适应对偶控制律之前，就将其中的孤立点数据实时检测出来并剔除，为自适应对偶控制赋

予了新的功能。另一方面在系统模型中加入不可控的系统激励成分，并在这个新的模型的基础上，提出了具有主动学习特性的对偶控制方法，减少了系统参数的不确定性。

(3) 由于一些实际的系统，例如经济系统、社会决策系统、生态系统中的随机噪声，通常具有尖峰、厚尾、非对称特征，而非普通的高斯白噪声，因而过去研究中基于卡尔曼滤波器的自适应对偶控制方法无法应对这种随机噪声。针对此问题，提出了自适应分位数控制方法，该方法设计了一种基于贝叶斯方法的分位数求和估计器，能够综合不同分位数下未知参数的估计值，从而使非高斯白噪声下的系统获得更好的参数学习效果，从而提升系统的控制性能。

(4) 针对完全未知的非线性系统的控制问题，经典的基于神经网络的自适应控制使用验前历史输入-输出数据进行建模，然后利用该网络模型计算控制律，这种方法并没有考虑网络模型与真实系统之间的误差，导致控制性能不佳，甚至远离最优控制。因此设计了一种具有主动学习特征的自适应对偶控制方法，一方面使用自动分配资源的神经网络模型对系统进行在线学习，并使用信息熵描述网络模型的学习效果，另一方面在求取自适应控制律时，不仅考虑了系统最优跟踪控制性能，同时还考虑到了如何优化网络模型的学习性能，使得控制律具有了主动学习的特性，进一步提升整体控制性能指标。

(5) 针对系统中包含不可测量的动态干扰问题，现有的理论是在最坏情况下设计鲁棒控制，尽管控制器能够应对允许范围内的动态干扰，但是过于保守。因此本文设计了一种具有主动学习特性的抗扰动控制器，提出了同时包含加性扰动和乘性扰动的神经网络建模方法，然后基于乘性和加性扰动的验前知识，生成一组有限候选集，利用贝叶斯理论确定离真实干扰最接近的候选值，最终由神经网络计算出具有抗扰动特点的控制律。该方法在候选值的学习过程中充分考虑到了学习的误差，赋予控制律对未知扰动的主动学习性质，从而进一步减少系统的不确定性并提高了控制性能。

**关键词：**不确定性；对偶控制；Q-学习；自适应控制；孤立点

**Title: RESEARCH ON ADAPTIVE DUAL CONTROL METHOD FOR UNCERTAIN SYSTEMS**

**Major: Control Theory and Control Engineering**

**Name: Xuehui Ma**

**Signature:** Xuehui Ma

**Supervisor: Prof. Fucai Qian**

**Signature:** Fucai Qian

## Abstract

Most practical systems are negatively affected by uncertainties, such as unknown system parameters, nonlinearity, structural variation, unpredicted component failures, etc. The traditional certainty-equivalence-based adaptive control treats the estimated parameters as the truth values and drives the control law directly upon this estimation. However, this method does not consider the parameter estimation error, leading to the designed controller making the closed-loop system far from optimal. In contrast, the adaptive dual control method studied in this paper can drive systems not only to track the target trajectory cautiously, but also fully motivate the system to actively learn more information about the system dynamics, so that the control law can be closer to the optimal one. Although previous adaptive dual controllers have handled the mentioned uncertainties, other types of uncertainties, such as random noise with non-Gaussian distributions, outliers, random disturbances, gained insufficient attention in the adaptive dual control.

This paper conducts research on adaptive dual control methods with active learning for systems suffering from different types of uncertainties. Our research efforts are as follows:

(1) Although the traditional reinforcement Q-learning based adaptive control method for linear systems with unknown parameters can derive the optimal control law by trial-and-error learning, however, their exploration signal needs manually setting, and the large exploration signal will bring large overshoots. We present an active learning exploration signal algorithm, resulting that the system state can cautiously track the target trajectory, and the system is stimulated toward enriched information, where the controller can actively explore unknown parameters. The actively adjusted exploration signal is calculated by optimizing the state tracking performance and the Q-function kernel matrix estimation performance simultaneously, so as to avoid overshoot and insufficient exploration in exploration period and optimally balance the conflict between exploration and exploitation in Q-learning.

(2) Outliers, resulting from sensor malfunctions, network data transmission errors, device malfunctions, adversary attacks, etc., exist in most practical systems. The existing adaptive dual

control methods for uncertain systems are aimed at systems polluted by white Gaussian noise, yet they are no longer valid for outliers. We use the real-time observation data to design two criteria for distance and direction, and detect outliers in the data stream before deriving the adaptive dual control law. On the other hand, we devise uncontrollable system excitation components in our method, and propose a dual adaptive control with active learning to reduce the uncertainty of system parameters.

(3) Random noises in practical systems, such as economic systems, social decision-making systems, and ecosystems, are usually characterized by peak, thick-tail, and asymmetric, rather than Gaussian white noises. As such, Kalman Filter based adaptive dual control cannot cope with this kind of random noise. We proposed an adaptive quantile control method with a Bayesian quantile sum estimator, which can synthesize the unknown parameters estimations under different quantiles, so as to improve the parameter estimation accuracy quality and the control performance of the system.

(4) The traditional neural network based adaptive control for completely unknown nonlinear systems uses historical input-output data for modeling, and then calculates the control law without considering the approximation error, which will degrade the control performance. We design an adaptive dual control method with active learning, which uses resource allocated neural network model to learn the system online, and leverage the information entropy to quantify the model learning performance. The active learning featured control law is derived by considering the optimal tracking control performance and the model learning performance, which can further improve the control performance.

(5) The traditionally robust control for the system with unmeasurable disturbances is designed under the worst-case assumption, which is too conservative for desirable control performance. This thesis designs an anti-disturbance controller with active learning, where we propose a neural network model that integrates both additive and multiplicative disturbances, and then design a finite candidate set according to the prior knowledge of multiplicative and additive disturbances. We use the Bayesian theory to learn which candidate value is closest to the ground truth value, and drive the anti-disturbance control law by the learned disturbed neural network. In the candidate values learning, the learning error is fully considered, and the control law is derived based on the active learning property for unknown disturbance, which aims to reduce the system uncertainties and improving the control performance.

**Key words:** uncertainties; dual control; Q-learning; adaptive control; outlier

# 目录

1 绪论 .....	1
1.1 研究背景与意义 .....	1
1.2 国内外研究现状 .....	2
1.3 研究内容与组织结构 .....	8
2 几种自适应控制方法的回顾 .....	11
2.1 引言 .....	11
2.2 基于确定性等价原理的自校正控制 .....	11
2.3 有谨慎特点的自校正控制 .....	13
2.4 具有对偶特性的自校正控制 .....	14
2.5 基于强化 Q-学习的自适应控制 .....	15
2.6 仿真结果及分析 .....	18
2.7 本章小结 .....	23
3 具有探索-利用平衡特性的 Q-学习自适应控制 .....	25
3.1 引言 .....	25
3.2 未知系统的线性二次调节问题 .....	26
3.3 基于 Q-学习的具有探索和利用平衡特性的自适应控制策略 .....	27
3.4 仿真实验 .....	32
3.4.1 一阶线性系统 .....	33
3.4.2 二阶线性系统 .....	36
3.5 本章小结 .....	39
4 具有孤立点噪声的随机系统自适应对偶控制 .....	41
4.1 引言 .....	41
4.2 问题描述 .....	42
4.3 控制器设计 .....	44
4.3.1 在线孤立点检测 .....	44
4.3.2 具有不可控激励的未知系统的双准则自适应对偶控制 .....	46
4.4 仿真实验 .....	49
4.4.1 在线孤立点检测仿真实验 .....	50
4.4.2 观测噪声中存在孤立点时的控制仿真实验 .....	52
4.4.3 过程噪声中存在孤立点时的控制仿真实验 .....	54
4.4.4 生物发酵连续灭菌过程的控制仿真 .....	57
4.5 本章小结 .....	59
5 具有非对称拉布拉斯噪声的随机系统自适应对偶控制 .....	61

5.1 引言.....	61
5.2 问题描述.....	62
5.3 控制器设计.....	67
5.3.1 差分模型 .....	68
5.3.2 迭代分位数估计器 .....	69
5.3.3 贝叶斯分位数求和估计器 .....	71
5.3.4 自适应对偶控制器 .....	72
5.4 仿真实验.....	73
5.4.1 贝叶斯分位数求和估计器 .....	74
5.4.2 最小相位系统仿真实验 .....	76
5.4.3 非最小相位系统仿真实验 .....	78
5.5 本章小结.....	81
6 基于可自动分配资源神经网络的自适应对偶控制.....	83
6.1 引言.....	83
6.2 问题描述.....	84
6.3 可自动分配资源的神经网络.....	85
6.4 网络模型参数估计.....	86
6.5 基于信息熵的自适应对偶控制.....	88
6.6 仿真实验.....	91
6.6.1 可自动分配资源的神经网络建模分析 .....	91
6.6.2 基于信息熵的自适应对偶控制实验分析 .....	94
6.7 本章小结.....	97
7 具有主动学习特性的抗扰动自适应对偶控制.....	99
7.1 引言.....	99
7.2 问题描述.....	100
7.3 控制器设计.....	101
7.3.1 针对具有扰动的非线性系统的特殊神经网络 .....	102
7.3.2 设计有限扰动候选集对扰动进行近似 .....	103
7.3.3 抗扰动自适应对偶控制器 .....	105
7.4 仿真实验.....	110
7.4.1 非线性系统受乘性扰动的影响 .....	111
7.4.2 非线性系统受加性扰动的影响 .....	113
7.4.3 在不同的估计误差容许值下的蒙特卡洛仿真实验 .....	116
7.4.4 高速列车的速度控制 .....	118
7.5 本章小结.....	121

8 总结与展望 .....	123
8.1 本文工作总结 .....	123
8.2 未来工作展望 .....	124
致谢 .....	125
参考文献 .....	127
攻读博士学位期间完成的主要工作 .....	139



# 1 绪论

## 1.1 研究背景与意义

不确定性在实际系统控制过程中普遍存在，而被控系统的不确定性按来源分可以分成两类，一类来源于被控系统的外部环境，另一类来源于系统内部变化<sup>[1-3]</sup>。例如，在高速列车的速度控制系统中，不确定性可来自于外部的环境变化，包括天气变化如刮风、降雨、降雪、轨道突发状况等；也可能来自于列车系统本身的变化，如零部件的正常磨损与老化、机械部件和电子部件的故障、车辆结构、摩擦因数的变化等<sup>[4,5]</sup>。在生物发酵连续灭菌系统的控制过程中，不确定性可源于外部环境的温度、湿度、气压等变化；也可能来源于系统内部，如物料的投放流量和温度的变化、加热蒸汽温度的变化、热交换设备效率的波动、传感器的故障和测量误差等等<sup>[6,7]</sup>。在卫星姿态控制系统中不可避免地存在转动惯量和干扰力矩不确定性的问题，来自于外界的不确定性包括重力梯度力矩、太阳辐射力矩、气动力矩以及地磁力矩等，内部干扰力矩包括执行机构的内部摩擦、活动部件的转动、执行机构的安装误差等，以及卫星有效载荷运动、太阳帆板的展开及转动、喷气的消耗等原因均会导致转动惯量的不确定性<sup>[8,9]</sup>。

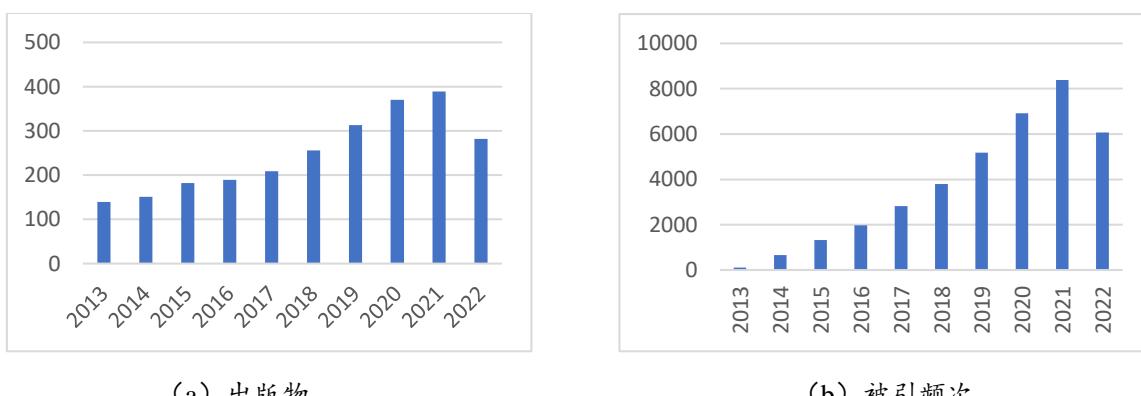
不确定性在实际的车辆系统、工业控制过程、航空航天等领域普遍存在，而这些不确定性往往会对系统的控制性能带来负面影响，甚至对系统造成永久损坏。因此在控制系统的应用设计中，如何减少系统的不确定性对系统造成的负面影响以及提高系统的鲁棒性就成为一个极其重要的问题。如果系统的不确定性部分可以直接测量，那么前馈控制是一个高效的鲁棒控制方案，但是在实际的系统中，不确定成分往往难以直接测量，还具有随机性，这对系统控制带来巨大的挑战，从而推动了大量的鲁棒控制策略和自适应控制策略的发展<sup>[10]</sup>。传统鲁棒控制的思想通常是考虑不确定性影响最差情况下的最优控制律，然而基于这个思想的控制方案往往过于保守，以牺牲了部分系统控制性能为代价，来确保系统的鲁棒性<sup>[11]</sup>。本文着眼于不确定性条件下的控制性能最优，因此这里不对该种方法进行详细研究。

研究表明，不确定性可以分为两类，一类是系统外部环境产生的不确定性，这类不确定性是一种客观存在，不可减少，只能用适当的滤波算法对被污染的信号进行估计，另一类是系统内部参数变化而产生的不确定性，由于未知参数的信息寓于输出信号中，可以通过一段时间的测量值对其在某种指标意义下进行最优估计，显然这类不确定性通过学习可以减少。自适应控制策略是针对系统中可减少的不确定性，一边实时学习以减少系统的不确定性，一边使用学习到的知识进行系统控制<sup>[12]</sup>。然而自适应控制中存在一个棘手的问题，即不确定成分的学习过程往往与系统最优控制目标相互冲突<sup>[13]</sup>，简单来说就是，在系统的控制器学习方面，需要对系统施加较大的探索激励信号来激发系统尽可能多的模态信息，促使控制器学习到更多的不确定性成分的信息，但是较大的激励信

号会导致系统输出产生超调和波动，降低系统的控制性能，甚至会损坏系统；而另一方面，若使用较小的学习探索激励信号，则可以减少系统的超调，保持系统的稳定，却减少了系统对未知信息的探索学习，那么不充分的学习结果必然导致系统控制无法进一步寻找到更高性能的控制律<sup>[14]</sup>。该问题于 1961 年由前苏联学者 Feldbaum 首次提出<sup>[15]</sup>，并命名为对偶控制问题，在 2000 年的 IEEE control system society 大会上被列为上世纪对控制领域最具影响力的问题之一。尽管目前针对现实系统中存在未知参数、非线性环节、结构变化等不确定因素的控制问题，已经有一些基于自适应对偶控制的解决方法，但不确定因素远不止这些。其他不确定因素如孤立点噪声，不同分布类型的随机噪声，随机变化的扰动等，目前极少有相对应的自适应对偶控制方法的研究，实际系统的控制中也缺乏应对多种不确定因素的有效解决方案。因此，针对具有不确定性系统的自适应对偶控制做进一步研究具有重要意义和价值。

## 1.2 国内外研究现状

对于不确定性系统的控制问题，自适应对偶控制方法因具有巨大潜力与进一步研究发展的空间而引起国内外学者的广泛关注。近些年来，学术界和工业界对自适应对偶控制的理论与应用进行了大量的研究，根据 web of science 数据库的检索结果，分别给出了自适应对偶控制相关文章发表情况以及引用情况。通过图 1-1 中的子图（a）可以看到，有关自适应对偶控制的 SCI 索引文章从 2013 年至 2022 年 9 月份都在呈现逐年增长的趋势。子图（b）展示了自适应对偶控制相关文献被引用的数量，也呈现逐年增长的趋势，且自 2013 年以来的总计被引用频次为 71180，平均引用次数为 21.59 次。这充分说明自适应对偶控制的研究是目前研究的热门课题。本文接下来将详细描述自适应对偶控制的研究现状。



(a) 出版物

(b) 被引频次

图 1-1 2013-2022 年自适应对偶控制的出版物数量以及出版物被引频次

Fig.1-1 The publications and citations of adaptive dual control from 2013 to 2022

自适应控制的思想是，针对数学模型并不明确已知，或者系统参数发生变化，或者结构发生变化，或者系统噪声不满足平稳性假设的不确定系统，采取一边进行在线辨识，

一边进行系统最优控制计算的方法<sup>[1,2,12]</sup>。因此自适应控制一般包含两个方面的功能：（1）对于被控对象的学习功能，即在控制的过程中积累有关被控对象特性的知识；（2）控制功能，即在积累知识的同时对系统施加控制，使之达到某种期望的效果。1958 年 Kalman 提出了自校正控制的思想，1973 年后鉴于计算机技术的迅速发展，由 Astrom 等人的研究，自校正控制有了突破性发展<sup>[18]</sup>。自校正控制的结构如图 1-2 所示，其中参数估计与控制器的设计是分离的。控制器的设计是基于确定性等价控制的思想，即在控制过程中不考虑参数的不确定性，认为参数的估计值是未知参数的真值。这个方法的执行步骤简单，操作方便，在发展初期至今都有极为广泛的应用。

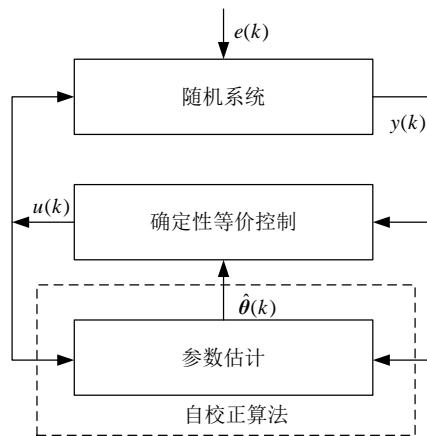


图 1-2 自校正控制结构框图  
Fig.1-2 The block diagram of self-tunning control

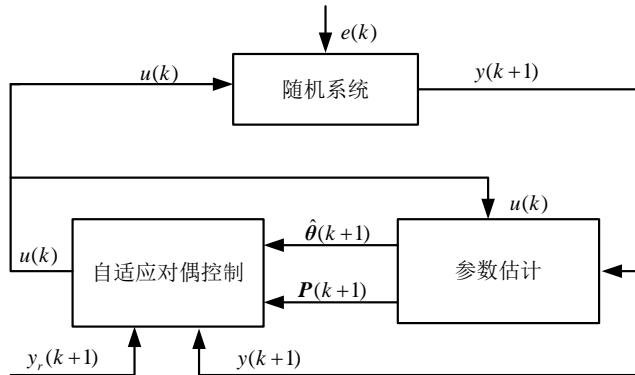


图 1-3 自适应对偶控制结构框图  
Fig.1-3 The block diagram of adaptive dual control

在 Feldbaum 早期关于最优控制的问题的工作中指出（1960-61, 1965），基于确定性等价原理的自适应控制并不是最优的<sup>[15,19]</sup>。他认为最优自适应控制系统应该具有两个主要特性：（1）系统输出能谨慎跟踪期望信号；（2）控制信号能充分激励系统主动学习更多的未知信息，从而在未来时刻得到更好的控制效果。这两个特性就是对偶特性（Dual Properties），能表现出这两个特性的自适应控制系统称为自适应对偶控制

(Adaptive Dual Control)。图 1-3 是自适应对偶控制的结构框图, 与基于确定性等价原理的自调节控制相比, 自适应对偶控制律设计的过程中还考虑到了参数估计的质量问题, 在图中表示为参数估计误差协方差矩阵  $P(k+1)$ 。

Feldbaum 在文献[19]中使用动态规划求最优自适应对偶控制的标准解, 但是由于维数灾难等问题, 即使是一个简单的被控对象, 由于底层空间维数增加, 得到解析解或数值解都比较困难<sup>[20,21]</sup>。Sternby 求得了一个简单对偶控制问题的解析解, 但是这个问题只考虑到少数几个可能的状态<sup>[22]</sup>。大部分的自适应控制系统在系统达到平衡且参考信号没有变化时, 就无法继续正常运行, 参数估计也因此停止, 这就是关断效应<sup>[23]</sup>。在系统状态没转移的情况下, 参数估计也一直都在运行, 关断效应会导致参数估计算法的信息矩阵的行列式的值接近零, 当求取该矩阵的逆时会有较大的计算错误出现。这个结果随后会导致维数爆炸, 参数估计值会非常大且不切实际, 并且系统的输出出现绝对值超大的情况。关断效应问题和最优对偶控制难以求解的问题, 推动了自适应对偶控制次优解的方法研究, 随之出现了大量的关于自适应对偶控制次优解的方法研究和应用的成果报道<sup>[13,24,25]</sup>。这些方法大致有六类: (1) 在谨慎控制器中加入摄动信号<sup>[26-28]</sup>; (2) 限制参数估计的方差<sup>[29]</sup>; (3) 损失函数的级数展开<sup>[30-32]</sup>; (4) 损失函数变形<sup>[33,34]</sup>; (5) 有限参数集<sup>[35,36]</sup>; (6) 从鲁棒控制思想设计控制器<sup>[37,38]</sup>。

以上是关于次优自适应对偶控制设计方法的总结, 近些年来自适应对偶控制有了新领域的发展, 有(1) 基于神经网络的自适应对偶控制; (2) 基于模型预测控制的自适应对偶控制; (3) 对偶控制在强化学习的探索与利用间实现平衡的应用和发展。下面对这些新发展进行逐一说明。

### (1) 基于神经网络的自适应对偶控制

Kadirkamanathan 首次利用高斯径向基函数 (Radial Basis Function, RBF) 神经网络模型在线学习系统非线性函数, 其控制器的计算使用了基于后验概率的多模型对偶控制概念<sup>[39]</sup>。随后 Fabri 和 Kadirkamanathan 提出基于新息的神经网络模型自适应对偶控制, 其中网络模型是基于 RBF 和 MLP (Multilayer Perceptron, MLP) 的非线性网络模型, 网络模型参数估计使用了扩展卡尔曼滤波<sup>[40]</sup>。Simandl 等人将双准则对偶控制方法应用到基于 MLP 神经网络模型的非线性系统中, 使用高斯和方法来估计网络模型参数<sup>[41]</sup>。Kral 研究了当系统中观测噪声有孤立点的情况下, 使用混合高斯噪声对观测噪声建模, 并使用混合高斯噪声估计器对网络模型参数进行学习, 控制器也是基于双准则对偶控制方法<sup>[42]</sup>。Fabri 等人还研究了 Hammerstein 模型的对偶控制方法<sup>[43]</sup>。Bugeja 和 Fabri 等成功的将基于神经网络的自适应对偶控制应用到移动机器人的控制中<sup>[44-46]</sup>。Kral 和 Simandl 提出了基于神经网络模型的非线性随机系统的预测对偶控制<sup>[47]</sup>。Fabri 和 Bugeja 提出了非线性多输入多输出系统的神经网络模型自适应对偶控制方法<sup>[48]</sup>。

### (2) 基于模型预测控制的自适应对偶控制

近十年来自适应对偶控制的思想被引入模型预测控制 (Model Predictive Control,

MPC) 的研究中<sup>[49]</sup>。Kim 等人在前人鲁棒模型预测自适应控制的基础上, 对参数估计误差建立了一个单调递减的边界数值函数, 并给出了一个减小参数不确定性的非传统形式的自适应鲁棒控制方法<sup>[50]</sup>。Marafioti 等人通过持续输入激励信号, 确保控制系统的对偶特性<sup>[51,52]</sup>。Žáčeková 提出了一种基于信息矩阵最小特征值最大化的新算法, 以获得满足控制要求并能获取充分信息的激励信号, 实现模型预测控制的对偶特性<sup>[53]</sup>。王超等人基于状态反馈 Tube 不变集鲁棒 MPC 算法, 利用自适应集员滤波在线估计系统过程噪声边界及状态可达集, 给出了确保系统状态鲁棒渐近稳定, 并收敛于终端干扰不变集的鲁棒自适应 MPC 算法<sup>[54]</sup>。Houska 等人针对含状态估计器的 MPC 控制, 提出一种名为 self-reflective 的非线性 MPC 控制策略, 通过在目标函数中加入未来状态估计方差的函数实现对偶控制<sup>[55]</sup>。Feng 等人延续了 self-reflective 的非线性 MPC 控制策略, 为 self-reflective MPC 提供了目标函数权重自动确定的方法<sup>[56]</sup>。Heirung 等人针对参数估计不确定性, 在其目标函数中加入未来参数估计的信息矩阵或协方差矩阵的函数项, 优化性能指标的同时减小自适应 MPC 中的参数不确定性<sup>[57]</sup>。随后针对随机系统中存在的不确定性问题, 基于正交基模型, 通过添加对未来跟踪误差的统计描述项, 建立了一个以最小化期望性能损失为目标的随机最优控制问题, 并将其近似为二次约束二次规划问题求解<sup>[58]</sup>。Kumar 等人用将对偶控制的思想用于基于维纳模型的非线性模型预测控制系统, 其中使用基准正交滤波器来进行参数估计<sup>[59]</sup>。曹文祺等人将自适应控制与鲁棒控制相结合, 将系统不确定性参数化, 并设定边界约束并作为优化问题的约束, 在优化控制目标同时减少不确定性对控制的影响<sup>[60]</sup>。Anilkumar 等人首次提出使用自适应对偶模型预测控制方法, 解决模型参数不匹配以及同时具有外界扰动的系统的输出跟踪控制问题<sup>[61]</sup>。Lin 等人提出在自适应对偶模型预测控制中, 使用在线集员辨识方法和递归最小二乘方法来减少参数的不确定性和得到参数估计值<sup>[62]</sup>。

### (3) 强化学习中探索与利用的平衡

近几年来, 强化学习结合深度学习在游戏策略设计中的成功应用<sup>[63,64]</sup>, 激发了学者们对强化学习研究的兴趣, 其中的根本任务是在动态不确定性环境中学习做出决策, 这个问题和对偶控制问题有很强的关联。强化学习中有两个目标: (i) 目标被量化为奖励函数, 求取奖励函数最大; (ii) 由于不确定性, 通过探索学习环境的信息。这两个目标相互冲突, 称为强化学习中的“探索-利用”平衡问题, 在控制中称为“对偶特性”。强化学习与控制之间的关系在最近的文献<sup>[65-67]</sup>中有详细的描述。Klenske 等人在环境模型是神经网模型的情况下, 首次提出将对偶控制中的近似方法用在贝叶斯强化学习中, 得到了不同于标准强化学习方法的结构化探索策略<sup>[68]</sup>。Andream 等人研究了有限时间长度下的瞬态特征, 提出了半定程序的方法解决具有未知参数系统的线性二次型问题, 得到探索与利用平衡的优化方法<sup>[69]</sup>。Jack 等人认为探索应该具有针对性, 需要在收集对目标有用的信息的同时, 不损害系统且确保系统的安全和可靠运行。他们提出了一种新的参数估计误差谱性质的高概率界, 对最坏情况的奖励函数进行凸近似求解, 称该方法为鲁棒强

化学习<sup>[70]</sup>。Chen 等人借用对偶控制的思想，解决了在未知环境中使用强化学习自主搜寻的过程中的探索与利用的平衡问题，具体就是在性能指标函数中增加了系统不确定性的估计性能<sup>[71,72]</sup>。

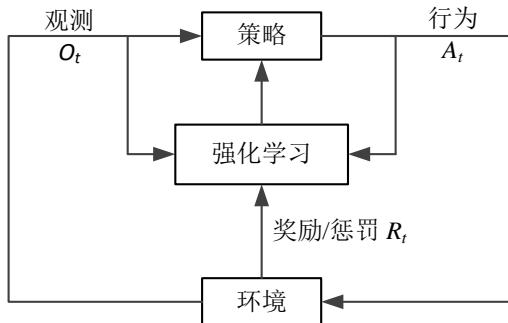


图 1-4 强化学习结构框图  
Fig.1-4 The block diagram of reinforcement learning

自 2017 年 AlphaGo 击败李世石以来，强化学习的研究成为了一个热点。强化学习是智能体通过与未知不确定的环境交互来改变自己的行为的方法，其与未知环境的交互就相当于一个试错的过程，从与环境的交互过程中所接收到的刺激或反馈中学习到未知的信息，然后利用得到的信息去调节行为<sup>[16]</sup>。图 1-4 是强化学习方法的结构框图，图中为智能体通过行为与环境进行交互，行为作用于环境得到奖励或者惩罚，然后智能体从奖励或者惩罚中修正自己的行为以获得更高的奖励。从理论的角度来看，强化学习与直接自适应控制有着很强的联系，都是一边利用反馈的信息对系统未知部分进行学习，一边利用学习到的新的知识做出最优的行为<sup>[17]</sup>。Sutton 于 1992 年发表在 IEEE Control Systems Magazine 的一篇文章就提到强化学习就是直接自适应控制，也就是说最优自适应控制可以通过强化学习来实现。

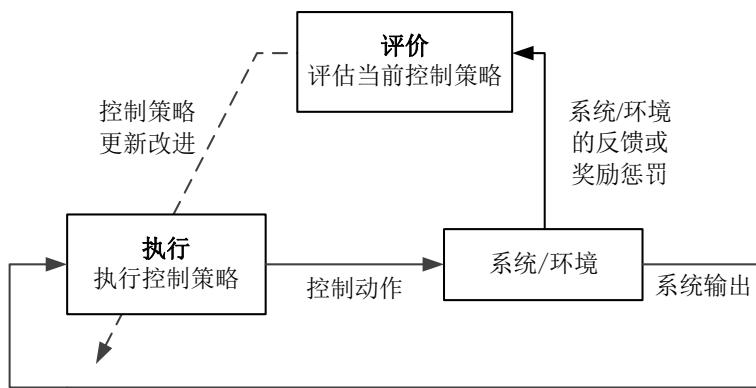


图 1-5 行为-评价结构  
Fig.1-5 The Actor-Critic structure

因而在基于强化学习的自适应控制中存在着与传统自适应对偶控制类似的问题，即探索与利用的平衡问题，系统在试错也就是探索的过程中，如果采用较为激进的行为进

行探索，那么系统就会受到严重惩罚，从而造成经济成本等各个方面的损失，但是如果探索行为过于保守，更多的利用已知信息来做出最优的行为策略，那么系统会更为稳定，但是牺牲了进一步探索更优的控制策略的机会。同样的在强化学习中，探索与利用的冲突问题实际上没有彻底解决，所以对该问题进行进一步探索极为重要。下面给出了基于强化学习思想的自适应控制的研究现状。

图 1-5 所示的是行为-评价结构下的强化学习算法，该算法是一种在时间上是从前往后推进的计算最优决策的方法，执行器对环境施加一个动作或者控制策略，然后评价器就对刚刚施加的动作或者控制策略进行价值评估，这些过程也都是实时进行的。图中所示的基于行为-评价结构的强化学习有两个迭代步骤，分别为评价网络的策略评估和行为网络的策略更新。通过观察当前控制动作作用于环境后得到的结果反馈来计算策略评估值，而评估值通常就是与最优控制目标的差距，其对应的物理意义可以表达为最少燃料、最少能量、最小风险、最大收益、最小跟踪偏差等等。根据对系统性能的评估，可以通过一些方法来提高控制策略，使得新的控制策略能够比之前的策略拿到更好的评估值。直接自适应控制是利用观测值直接实时辨识控制器的参数来调整控制器，而基于行为-评价结构的强化学习类似，即利用观测数据来辨识当前策略下的性能评估值，再根据性能评估来更新控制策略。

Werbos 在 1974 年的研究中首次提出将强化学习的思想应用于最优控制中，随后该方法获得学术界的关注和进一步发展，该方法目的是利用学习的方法使闭环系统的性能指标达到最优，解决了最优调节问题、最优跟踪问题、以及鲁棒控制问题

### (1) 最优调节问题

文献[73]给出了离散线性系统的基于强化学习的线性二次调节器。文献[74]提供了基于行为-评价结构的线性二次调节器的全局收敛性的证明。文献[75]给出了一种时间刻度下行为-评价结构算法的线性二次调节器。文献[76]提出了一种利用动态输出反馈来学习最优控制参数的强化学习方法，解决了连续时间系统的线性二次型问题。文献[77]利用神经网络来逼近方法，解决了基于强化学习非线性系统无限时间最优控制问题。文献[78]提出了同步强化学习算法，行为-评价网络的权系数能够同步更新，并设计了闭环系统稳定并且满足性能指标函数最优的控制器。另外基于强化学习的先行二次调节器已经被应用于许多实际的控制系统中，例如无人驾驶直升机<sup>[79]</sup>，辅助人机交互系统<sup>[80]</sup>，两轮自平衡机器人<sup>[81]</sup>，微小型四旋翼飞行器<sup>[82]</sup>等等。

### (2) 最优跟踪问题

文献[83]使用行为-评价结构的强化学习方法解决了离散线性系统的无限时间线性二次跟踪问题。文献[84]利用神经网络拟合行为-评价网络，提出了基于强化学习思想的非线性系统的跟踪控制方法。文献[85]给出了多输入多输出的非线性仿射系统的强化学习自适应跟踪控制。文献[86]利用积分强化学习实现了部分未知系统的自适应最优跟踪控制。文献[87]提出了基于强化 Q-学习的自适应最优控制方案，对完全未知的离散时间线

性系统进行最优跟踪控制，并且还证明了算法的收敛性。文献[88]针对完全未知的连续时间非线性系统，给出了事件触发的强化学习自适应跟踪控制方法。基于强化学习的最优跟踪器也得到了广泛的应用，例如，机械臂轨迹控制<sup>[89]</sup>，轮式机器人航行控制<sup>[90]</sup>，水下机器人的控制<sup>[91]</sup>，风能转换系统<sup>[92]</sup>等。

### (3) 鲁棒控制问题

文献[93]提出了基于强化学习思想的  $H_\infty$  状态反馈控制器，分别设计了基于模型的离线同步策略更新算法和无模型的在线同步策略更新算法对黎卡提方程进行求解。文献[94]利用积分强化学习的方法在线得到两层零和博弈的纳什平衡的解，用自适应动态规划的方法求解零和博弈问题中的代数黎卡提方程。文献[95]提供了一种使用强化学习方法的去解决非线性系统的  $H_\infty$  控制问题，其中利用基于神经网络的 off-line 强化学习去学习 Hamilton-Jacobi-Isaacs 方程。文献[96]通过 off-line 强化学习方法对完全未知的非线性连续时间系统设计了一种  $H_\infty$  跟踪控制方法。

## 1.3 研究内容与组织结构

本文主要考虑具有不同种类的不确定因素的系统的控制问题，有针对性的提出不同特点的解决方案，进一步完善不确定系统的自适应对偶控制方法研究。本文主要研究内容如下：

(1) 对结构已知、参数未知的系统，在基于强化 Q-学习的自适应控制的框架下，借鉴对偶控制的思想，对控制器设计中关于控制信号的探索-利用的冲突问题给出了一种平衡方法。基于强化 Q-学习的自适应控制方法在试错学习阶段，如果探索信号过大会引起较大的超调，探索信号过小会引起系统的关断效应，也就是所谓的探索-利用的冲突问题。对此问题设计了可以主动调整的探索信号，该信号通过同时优化两个性能指标函数求得，即系统状态跟踪误差最小和系统动态估计误差最大。通过对一阶和二阶线性系统的数值仿真，验证了所提出的控制策略对系统探索阶段性能的改善。

(2) 实际系统往往由于传感器失灵、数据传输错误、零部件损坏、黑客入侵等原因会被孤立点噪声污染。因此在参数未知的线性被控制系统中，分别针对传感器测量数据和系统过程本身被孤立点污染这两种情况，提出了一种能够在线检测孤立点的自适应对偶控制方法。其中在线孤立点检测算法由两个判定准则组成：当前被检测数据与上一时刻的数据之间的期望距离边界和期望方向边界。如果当前被检测数据既不在期望距离内也不在期望方向内，那么该数据就被判定为孤立点。同时，针对系统中不可控激励信号进行建模，并将所设计的在线孤立点检测算法嵌入到具有不可控激励信号模型的自适应对偶控制的框架中，确保系统或者传感器测量数据即使受到孤立点的干扰，也能平稳运行且具有良好的控制跟踪效果。最后使用数学模型生成的数据进行仿真实验去验证算法的有效性，并进一步使用实际生物发酵连续灭菌过程中测量到的数据进行实验，说明

了该方法的实用性。

(3) 在一些经济系统、社会决策系统、生态系统中的噪声往往不是理想的高斯白噪声，而是一些具有尖峰、厚尾、非对称特性的随机噪声，针对包含这一类特点的随机噪声的线性未知系统，提出了贝叶斯分位数求和估计器进行在线估计未知参数，并将其嵌入到自适应对偶控制的框架中。与理想的高斯白噪声不同的是，具有尖峰、厚尾、非对称特征的随机噪声可以使用非对称拉布拉斯分布进行描述，基于非对称拉布拉斯随机噪声的特点设计了贝叶斯分位数求和估计器。该参数估计器分别用不同分位数进行参数估计，然后以实时更新的贝叶斯后验概率作为相应的权值，对不同分位数下的参数估计值进行加权求和得到最优估计值。最后将参数的最优估计值用于自适应对偶控制律的求解中，保证了系统在受到具有尖峰、厚尾、非对称特性的随机噪声干扰的情况下，仍然能够平稳运行，且具有良好的跟踪控制性能。

(4) 针对完全未知的随机非线性系统，经典的基于神经网络模型的自适应控制使用神经网络进行建模，然后根据确定性等价原理，将建好的模型直接用于计算控制律，并没有考虑建模误差，因此导致系统控制性能不佳。对此问题，提出了一种基于自动分配资源的神经网络和信息熵的非线性自适应对偶控制方法。该方法一方面利用测量的系统输入-输出数据，实时更新可自动分配资源的神经网络模型的参数，另一方面使用信息熵的概念来描述系统信息增量，在代价函数中增加系统学习的信息增量来控制对未知信息的学习效率，从而得到具有主动学习特性的自适应对偶控制律。仿真实验表明该方法对未知非线性系统具有良好控制性能。

(5) 针对具有不可测动态加性扰动和乘性扰动的未知非线性系统，提出了一种具有主动学习特点的抗扰动对偶控制方法。该方法首先设计了一个特殊神经网络模型来学习未知系统，其中包含了未知干扰，由加性和乘性扰动的方式来表示。然后将特殊神经网络嵌入到对偶控制结构中，对其中的未知干扰进行学习，并对该系统进行输出跟踪控制。主动学习未知扰动的策略是给未知扰动参数分配一组候选值，根据系统输入-输出数据，并行的实时更新这些候选值的贝叶斯后验概率。每个候选扰动值都对应一个控制律，最终的控制律为所有控制律的后验概率加权的总和。数值仿真实验以及高铁速度控制模型下的实验表明该方法对具有未知动态扰动的非线性系统具有良好的控制性能。

论文的具体章节安排如下：

第 1 章：阐述了具有多种不确定性因素的系统的自适应对偶控制的研究背景及意义，以及自适应对偶控制和强化学习自适应控制的研究现状，最后给出了论文的研究内容与组织结构。

第 2 章：分别回顾了四种基本的参数未知的随机系统的自适应控制方法：基于确定性等价原理的自校正控制、具有谨慎特点的自校正控制、具有对偶特性的自校正控制、基于确定性等价原理的强化 Q-学习自适应控制，并且对这四种方法进行探讨和实验分析，为本文的研究和创新工作提供基础。

第 3 章：研究了基于强化 Q-学习的自适应控制中探索与利用之间平衡的问题，提供了一种主动调整探索信号的方法，以寻求探索与利用的最优平衡。

第 4 章：针对被孤立点干扰的系统，提出了具有在线检测孤立点功能的未知系统自适应对偶控制方法。

第 5 章：研究了被尖峰、厚尾、非对称特性的随机噪声污染的未知系统的自适应对偶控制方法。

第 6 章：针对完全未知的随机非线性系统，提出了基于可自动分配资源的神经网络与信息熵的非线性自适应对偶控制方法。

第 7 章：针对具有不可测量的乘性和加性干扰的未知系统，提出了具有主动学习特点的自适应对偶控制方法。

第 8 章：对全文工作进行总结，并对未来的研究进行展望。

## 2 几种自适应控制方法的回顾

自适应控制是一种有效解决不确定系统控制问题的方法，该方法能够一边利用输入-输出数据对系统不确定部分进行学习，一边利用学习到的信息计算控制信号，使系统朝着期望目标运行。本章节首先回顾了几种经典的自适应控制方法，包括基于确定性等价原理的自校正控制、具有谨慎特点的自校正控制、具有对偶特性的自校正控制以及基于强化 Q-学习的自适应控制，并对这些控制方法进行探讨和实验结果分析，在对比实验中分析自适应对偶控制的优点，为本文后续的研究和创新工作提供基础。

### 2.1 引言

目前大部分自适应控制都基于确定性等价原理，其参数估计与控制器设计是相互分离的，即在计算控制律时，不考虑参数估计的误差，而是将参数估计值直接作为真值使用。自校正控制就是基于确定性等价原理的控制方法<sup>[18]</sup>，该方法计算量小且执行步骤简单，从提出后至今在实际系统中都有极为广泛的应用。

Feldbaum 在六十年代初期就指出基于确定性等价原理的自适应控制并不是最优控制，甚至远未达到最优控制<sup>[15]</sup>。他认为最优自适应控制因该具有两个特性，也被称为对偶特性，即（1）系统输出能够谨慎地跟踪期望轨迹；（2）能充分激励系统主动学习未知参数，从而获得更优的控制效果，具有这两个特性的控制系统被称为自适应对偶控制系统。

Watkins 于九十年代提出基于强化 Q-学习的自适应控制方法<sup>[97]</sup>，也被称为动作依赖的启发式动态规划。对于未知的不确定系统，可以使用该方法进行试错探索学习，得到最优控制律。

下面就对基于确定性等价原理的自校正控制、具有谨慎特点的自校正控制、具有对偶特性的自校正控制以及基于强化 Q-学习的自适应控制这几种典型的控制方法进行简单回顾，并基于简单的数学模型进行理论分析和实验比较，以说明每个方法各自的特点，为本文后续研究工作做基础。

### 2.2 基于确定性等价原理的自校正控制

考虑如下离散时间单输入单输出系统：

$$\begin{aligned} y(k+1) = & a_1(k)y(k) + a_2(k)y(k-1) + \cdots + a_m(k)y(k-m+1) \\ & + b_1(k)u(k) + b_2(k)u(k-1) + \cdots + b_n(k)u(k-n+1) + e(k+1) \end{aligned} \quad (2-1)$$

其中系统参数  $\theta(k) = [b_1(k), b_2(k), \dots, b_n(k), a_1(k), a_2(k), \dots, a_m(k)]$  是未知参数，且服从均值为  $\hat{\theta}(0)$ ，方差为矩阵  $P(0)$  的高斯分布；  $y(k)$  是系统输出，  $u(k)$  是系统控制输入，  $e(k)$  是均值为零，方差为  $r$  的高斯白噪声，且与参数  $\theta(k)$  相互独立。假设初始状态

$\mathfrak{I}_0 = \{u(-1), \dots, u(-n+1), y(0), \dots, y(0), \dots, y(-m+1)\}$  已知, 且系统为有限步长  $k = 0, 1, \dots, N$ 。

控制目标是最小化系统输出  $y(k)$  与给定轨迹  $y_r(k+1)$  的偏差, 性能指标函数为

$$J[u(k)] = E \{ [y(k+1) - y_r(k+1)]^2 | \mathfrak{I}_k \}, \quad (2-2)$$

其中  $\mathfrak{I}_k$  是  $k$  时刻的信息状态, 包含了  $k$  时刻前的历史输入数据和输出数据, 定义为  $\mathfrak{I}_k = \{x(k), x(k-1), \dots, x(1), u(k-1), \dots, u(1)\}$ , 通过最小化性能指标函数 (2-2) 得到控制律  $u(k)$ 。

假设未知参数  $\theta(k)$  是个常量, 则其状态方程为  $\theta(k+1) = \theta(k)$ 。根据式 (2-1), 以  $\theta(k)$  为状态的观测方程为

$$y(k+1) = \theta^T(k) \varphi(k) + e(k+1) \quad (2-3)$$

其中  $\varphi(k) = [u(k), u(k-1), \dots, u(k-n+1), y(k), y(k-1), \dots, y(k-m+1)]$  为截止  $k$  时刻的数据向量, 假设给定条件  $\mathfrak{I}_k$ , 系统未知参数  $\theta(k)$  服从正态分布, 则参数的条件均值和协方差矩阵就可以用卡尔曼滤波迭代公式进行估计更新

$$\hat{\theta}(k+1) = \hat{\theta}(k) + K(k+1)[y(k+1) - \hat{\theta}^T(k) \varphi(k)] \quad (2-4)$$

$$K(k+1) = P(k) \varphi(k) [\varphi^T(k) P(k) \varphi(k) + r]^{-1} \quad (2-5)$$

$$P(k+1) = P(k) - K(k+1) \varphi^T(k) P(k) \quad (2-6)$$

确定性等价原理值指的是在自适应控制的过程中, 由参数估计器实时得到的参数估计值被认为是实际系统参数的真值来计算控制律。换句话说, 就是在求解控制律时, 假设系统参数  $\theta$  等于参数估计的均值  $\hat{\theta}$ , 则确定性等价控制的代价函数可以表达为

$$\begin{aligned} J_{CE}[u(k)] &= E \{ [y(k+1) - y_r(k+1)]^2 | \theta = \hat{\theta}(k) \} \\ &= E \{ [\hat{\theta}^T(k) \varphi(k) + e(k+1) - y_r(k+1)]^2 \} \\ &= E \{ [\hat{b}_1(k) u(k) + \hat{\theta}_0^T(k) \varphi_0(k) + e(k+1) - y_r(k+1)]^2 \} \end{aligned} \quad (2-7)$$

确定性等价控制律  $u_{CE}(k)$  通过最小化代价函数  $J_{CE}[u(k)]$  得到

$$u_{CE}(k) = \left\{ \arg \min J_{CE}[u(k)] \right\}_{\theta=\hat{\theta}(k)} \quad (2-8)$$

令  $\partial J_{CE}[u(k)] / \partial u(k) = 0$ , 可以求得确定性等价控制的解析解为

$$u_{CE}(k) = \frac{y_r(k+1) - \hat{\theta}_0^T(k) \varphi_0(k)}{\hat{b}_1(k)}, \quad (2-9)$$

其中估计的参数向量  $\hat{\theta}(k)$  可分割为  $\hat{\theta}(k) = [\hat{b}_1(k) \ : \ \hat{\theta}_0(k)]$ , 观测向量  $\varphi(k)$  可分割为  $\varphi(k) = [u(k) \ : \ \varphi_0(k)]$ 。这里得到的确定性等价控制律  $u_{CE}(k)$  就是自校正控制律。

这里将参数  $\theta$  未知情况下确定性等价控制器的性能指标与参数  $\theta$  已知情况下的最优控制器的性能指标对比进行系统控制性能分析。假设参数  $\theta$  已知, 则可以得到最优控制的解析解

$$u_{OPT}(k) = \frac{y_r(k+1) - \theta_0^T(k)\varphi_0(k)}{b_1(k)}, \quad (2-10)$$

以及对应的最优性能指标

$$J_{OPT}(k) = J_{OPT}(k) | u_{OPT}(k) = r \quad (2-11)$$

为了后续便于计算和控制性能的比较，假设  $\theta_0^T(k) = 1$ ，则  $u_{CE}(k)$  对应的性能指标为

$$\begin{aligned} J_{CE}(k) &= J_{CE}(k) | u_{CE}(k) \\ &= E \left\{ [b_1(k)u_{CE}(k) + \varphi_0(k) + e(k+1) - y_r(k+1)]^2 | \mathfrak{I}_k \right\} \\ &= E \left\{ \left[ \frac{b_1(k) - \hat{b}_1(k)}{\hat{b}_1(k)} y_r(k+1) + \frac{\hat{b}_1(k) - b_1(k)}{\hat{b}_1(k)} \varphi_0(k) + e(k+1) \right]^2 \middle| \mathfrak{I}_k \right\} \\ &= \frac{P_{uu}(k)}{\hat{b}_1^2(k)} y_r^2(k+1) + \frac{P_{uu}(k)}{\hat{b}_1^2(k)} \varphi_0^2(k) + r \end{aligned} \quad (2-12)$$

通过比较可得

$$J_{CE}(k) \geq J_{OPT}(k) \quad (2-13)$$

即确定性等价控制所得到的系统性能不是最优的。由上式的计算结果可知，导致代价函数值增大来源于式 (2-12) 的前两项，这是由参数的估计误差引起的，也就是参数的不确定性引起的，如果参数  $\theta$  已知就无需估计，此时也就没有估计误差，即  $P(k) = 0$ ，同时  $P_{uu}(k)$ ,  $P_{\varphi\varphi}(k)$ ,  $P_{\varphi u}(k)$  也均为 0，那么  $J_{CE}(k) = J_{OPT}(k)$ 。

### 2.3 有谨慎特点的自校正控制

上一小节中将估计的参数  $\hat{\theta}$  作为真值得出了确定性等价控制器，却没有考虑参数的不确定性，即未知参数的估计误差。这里可以用协方差矩阵  $P(k)$  来表示参数的不确定性。则完全考虑参数不确定性的谨慎控制律的代价函数计算过程为：

$$\begin{aligned} J_{CAU}[u(k)] &= E \left\{ [y(k+1) - y_r(k+1)]^2 | \mathfrak{I}_k \right\} \\ &= E \left\{ [\tilde{b}_1(k)u(k) + \tilde{\theta}_0^T(k)\varphi_0(k) + e(k+1) + \hat{b}_1(k)u(k) + \hat{\theta}_0^T(k)\varphi_0(k) - y_r(k+1)]^2 | \mathfrak{I}_k \right\} \\ &= [P_{uu}(k) + \hat{b}_1^2(k)]u^2(k) + 2[P_{\varphi u}^T(k)\varphi_0(k) + \hat{b}_1(k)\hat{\theta}_0^T(k)\varphi_0(k) - \hat{b}_1(k)y_r(k+1)]u(k) + C \end{aligned} \quad (2-14)$$

令  $\partial J_{CAU}[u(k)] / \partial u(k) = 0$  得到谨慎控制律  $u_{CAU}(k)$

$$u_{CAU}(k) = -\frac{P_{\varphi u}^T(k)\varphi_0(k) + \hat{b}_1(k)\hat{\theta}_0^T(k)\varphi_0(k) - \hat{b}_1(k)y_r(k+1)}{P_{uu}(k) + \hat{b}_1^2(k)} \quad (2-15)$$

不同于确定性等价控制，谨慎控制考虑到了参数的不确定性，则谨慎控制  $u_{CAU}(k)$  对应的性能指标值为

$$\begin{aligned}
J_{CAU}(k) &= J_{CAU}(k) | u_{CAU}(k) = E \left\{ [y(k+1) - y_r(k+1)]^2 | \mathfrak{I}_k \right\} \\
&= E \left\{ \left[ \frac{\tilde{b}_1(k)\hat{b}_1(k) - P_{uu}(k)}{\alpha} y_r(k+1) + \frac{\tilde{\theta}_0^T(k)\alpha - \tilde{b}_1(k)\beta + \gamma}{\alpha} \varphi_0(k) + e(k+1) \right]^2 \middle| \mathfrak{I}_k \right\} \quad (2-16) \\
&= \frac{P_{uu}(k)}{\alpha} y_r^2(k+1) + \frac{\alpha^2 P_{\varphi\varphi}(k) + P_{uu}(k)\beta^T\beta - 2\alpha P_{\varphi u}(k)\beta + \gamma^2}{\alpha^2} \varphi_0^T(k)\varphi_0(k) + r
\end{aligned}$$

为便于比较，假设  $\theta_0^T(k)=1$ ，此时其控制律为

$$u_{CAU}(k) = -\frac{\hat{b}_1(k)\varphi_0(k) - \hat{b}_1(k)y_r(k+1)}{P(k) + \hat{b}_1^2(k)} \quad (2-17)$$

当未知参数  $b_1(k)$  的不确定性增大时，即  $P(k)$  变大，根据上式，谨慎控制器的绝对值变小，当  $P(k)$  趋向于无穷大时，谨慎控制器趋向于零。换句话说，谨慎控制就是对不熟悉的系统，即不确定性较大的系统，在初始阶段不会使用较大的控制量，这就是所谓的谨慎特点。

此时谨慎控制的性能指标值为

$$J_{CAU}(k) = \frac{P(k)}{P(k) + \hat{b}_1^2(k)} y_r^2(k+1) + \frac{P(k)}{P(k) + \hat{b}_1^2(k)} \varphi_0^2(k) + r \quad (2-18)$$

比较确定性等价控制和谨慎控制的性能指标值有

$$J_{CE}(k) \geq J_{CAU}(k) \quad (2-19)$$

由此可以得出，确定性等价控制控制效果没有谨慎控制好。

## 2.4 具有对偶特性的自校正控制

上一小节中的谨慎控制器过于保守，不具备主动探测学习的功能。这里介绍了基于新息的对偶控制方法<sup>[98]</sup>，将系统的新息加入到代价函数中，以增加控制信号的探测作用，来减少参数的不确定性，提高参数的估计精度。包含新息的代价函数为

$$J_{DUAL}[u(k)] = E \left\{ [y(k+1) - y_r(k+1)]^2 - \lambda v^2(k+1) | \mathfrak{I}_k \right\} \quad (2-20)$$

其中  $v(k+1)$  就是包含了新的参数信息的新息，这是由新的观测输出  $y(k+1)$  带来了新的信息，新息  $v(k+1)$  的表达式为

$$v(k+1) = y(k+1) - \hat{\theta}^T(k)\varphi(k) \quad (2-21)$$

这里可以通过最大化新息来获取最多的探测信息，以提高参数估计的精度，减少参数的不确定性。在代价函数中可以用系数  $\lambda(k+1)$  调节探测信号的大小，该系数的取值范围为  $0 \leq \lambda \leq 1$ 。基于新息的自适应对偶控制的代价函数计算过程为

$$\begin{aligned}
J_{DUAL}[u(k)] &= E \left\{ [y(k+1) - y_r(k+1)]^2 - \lambda v^2(k+1) \mid \mathfrak{I}_k \right\} \\
&= E \left\{ [\theta^T(k)\varphi(k) + e(k+1) - y_r(k+1)]^2 - \lambda [\theta^T(k)\varphi(k) + e(k+1) - \hat{\theta}^T(k)\varphi(k)]^2 \mid \mathfrak{I}_k \right\} \\
&= E \left\{ \left. \begin{aligned} &[\tilde{b}_1(k)u(k) + \tilde{\theta}_0^T(k)\varphi_0(k) + e(k+1) + \hat{b}_1(k)u(k) + \hat{\theta}_0^T(k)\varphi_0(k) - y_r(k+1)]^2 \\ &- \lambda[\tilde{b}_1(k)u(k) + \tilde{\theta}_0^T(k)\varphi_0(k) + e(k+1)]^2 \end{aligned} \right| \mathfrak{I}_k \right\} \quad (2-22) \\
&= 2[(1-\lambda)P_{\varphi u}^T(k)\varphi_0(k) + \hat{b}_1(k)\hat{\theta}_0^T(k)\varphi_0(k) - \hat{b}_1(k)y_r(k+1)]u(k) \\
&\quad + [(1-\lambda)P_{uu}^T(k) + \hat{b}_1^2(k)]u^2(k) + C
\end{aligned}$$

令  $\partial J_{DUAL}[u(k)] / \partial u(k) = 0$ ，得到自适应对偶控制律

$$u_{DUAL}(k) = -\frac{(1-\lambda)P_{\varphi u}^T(k)\varphi_0(k) + \hat{b}_1(k)\hat{\theta}_0^T(k)\varphi_0(k) - \hat{b}_1(k)y_r(k+1)}{(1-\lambda)P_{uu}^T(k) + \hat{b}_1^2(k)} \quad (2-23)$$

从公式形式来看，自适应对偶控制律与谨慎控制律以及确定性等价控制律有直观的联系。当  $\lambda=1$  时，自适应对偶控制律就等同于基于确定性等价原理的控制律，即  $u_{DUAL}(k)|_{\lambda=1}=u_{CE}(k)$ ；当  $\lambda=0$  时，自适应对偶控制律就等同于具有谨慎特点的控制律，即  $u_{DUAL}(k)|_{\lambda=0}=u_{CAU}(k)$ 。这表明，自适应对偶控制就是通过  $\lambda$  来调节控制律，使其值在确定性等价控制律和谨慎控制律之间。确定性等价控制过于激进忽略了参数的不确定性；而谨慎控制考虑参数的不确定性，但过于谨慎，在参数不确定性过于大的情况下会导致系统关断。对偶控制可以通过调节参数  $\lambda$  来平衡控制量，使其既不像确定性等价控制一样过于激进，也不像谨慎控制一样过于保守。上述对偶控制律可以改写成如下包含谨慎控制律和确定性等价控制律的形式

$$u_{DUAL}(k) = u_{CE}(k) \frac{\hat{b}_1^2(k)}{(1-\lambda)P_{uu}^T(k) + \hat{b}_1^2(k)} - (1-\lambda) \frac{P_{\varphi u}^T(k)\varphi_0(k)}{(1-\lambda)P_{uu}^T(k) + \hat{b}_1^2(k)} \quad (2-24)$$

$$u_{DUAL}(k) = u_{CAU}(k) \frac{P_{uu}^T(k) + \hat{b}_1^2(k)}{(1-\lambda)P_{uu}^T(k) + \hat{b}_1^2(k)} + \lambda \frac{P_{\varphi u}^T(k)\varphi_0(k)}{(1-\lambda)P_{uu}^T(k) + \hat{b}_1^2(k)} \quad (2-25)$$

从式 (2-24) 可以看出自适应对偶控制律具有两个部分，一部分是控制部分，另一部分是探测部分。由于  $0 \leq \lambda \leq 1$ ，则  $\frac{\hat{b}_1^2(k)}{(1-\lambda)P_{uu}^T(k) + \hat{b}_1^2(k)} < 1$ ，那么对偶控制的控制部分是由确定性等价控制律乘以一个小于 1 的系数得到，由此表明对偶控制比确定性等价控制更谨慎。在式 (2-25) 中， $\frac{P_{uu}^T(k)\hat{b}_1^2(k)}{(1-\lambda)P_{uu}^T(k) + \hat{b}_1^2(k)} > 1$ ，那么对偶控制的控制部分是由谨慎控制律乘以一个大于 1 的系数得到，由此表明对偶控制比谨慎控制更激进大胆。

## 2.5 基于强化 Q-学习的自适应控制

本小节介绍一种基于强化学习 Q-学习算法的自适应控制方法，这个方法可以在线求解系统参数未知的离散系统的线性二次调节问题 (linear quadratic regulator, LQR)。简

而言之，该方法可以在系统参数未知的情况下，仅仅使用实时观测的状态值，在线求解代数黎卡提方程，得到系统的最优控制的解。

考虑如下离散时间线性状态空间方程

$$x(k+1) = Ax(k) + Bu(k) \quad (2-26)$$

其中  $x(k)$  是  $m$  维状态向量， $u(k)$  是  $l$  维控制向量， $k$  是时间步数，模型参数  $A$  是  $m \times m$  状态转移矩阵， $B$  是  $m \times n$  控制矩阵，且  $A$  和  $B$  是未知且定常的。定义系统性能指标函数为

$$J(k) = \frac{1}{2} \sum_{i=0}^{\infty} [x^T(i)Qx(i) + u^T(k)Ru(k)] \quad (2-27)$$

假设控制策略为  $u(k) = \pi[x(k)]$ ，并且定义相应的值函数为

$$V[x(k)] = \frac{1}{2} \sum_{i=k}^{\infty} [x^T(i)Qx(i) + u^T(k)Ru(k)] \quad (2-28)$$

为了使用动态规划求解下述问题，假设值函数是关于状态  $x(k)$  的二次型，即  $V[x(k)] = \frac{1}{2} x^T(k)Px(k)$ 。在 Q-学习算法中，定义 Q 函数为

$$Q[x(k), u(k)] = \frac{1}{2} [x^T(k)Qx(k) + u^T(k)Ru(k)] + V[x(k+1)] \quad (2-29)$$

将状态方程 (2-26) 代入式 (2-29) 可得

$$\begin{aligned} Q[x(k), u(k)] &= \frac{1}{2} [x^T(k)Qx(k) + u^T(k)Ru(k)] + \frac{1}{2} [Ax(k) + Bu(k)]^T P [Ax(k) + Bu(k)] \\ &= \frac{1}{2} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}^T \begin{bmatrix} A^T PA + Q & A^T PB \\ B^T PA & B^T PB + R \end{bmatrix} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \end{aligned} \quad (2-30)$$

定义核矩阵

$$S = \begin{bmatrix} A^T PA + Q & B^T PA \\ A^T PB & B^T PB + R \end{bmatrix} \quad (2-31)$$

将核矩阵  $S$  代入 Q 函数可得

$$Q[x(k), u(k)] = \frac{1}{2} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}^T S \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}^T \begin{bmatrix} S_{xx} & S_{xu} \\ S_{ux} & S_{uu} \end{bmatrix} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \quad (2-32)$$

令  $\partial Q[x(k), u(k)] / \partial u(k) = 0$ ，可得最优控制律

$$u(k) = -S_{uu}^{-1} S_{ux} x(k) \quad (2-33)$$

由上式可知，求最优控制律  $u(k)$  只需要 Q 函数的核矩阵  $S$ ，而不需要确切的系统参数  $A$  和  $B$  的值。下面介绍如何使用时序差分方法，用观测值来在线学习 Q 函数的核矩阵  $S$ 。这个过程也就是所谓的 Q-学习算法，可以在参数未知的情况下，在线求解贝尔曼方程，从而得到最优控制律。

由式 (2-32) 可得,  $Q$  函数是系统状态变量  $x(k)$  和控制变量  $u(k)$  的二次型, 令  $z(k) = [x^T(k) \ u^T(k)]^T$ , 则  $Q$  函数可以写成关于  $z(k)$  的二次型

$$Q[x(k), u(k)] = Q[z(k)] = \frac{1}{2} z^T(k) S z(k) \quad (2-34)$$

其中矩阵  $S$  可以用系统辨识的方法进行在线估计。将以上  $Q$  函数写成下面的形式便于进行参数辨识

$$Q[x(k), u(k)] = Q[Z(k)] = W^T(k) \phi[z(k)] \quad (2-35)$$

其中将矩阵  $S$  变化为参数向量  $W(k)$ ,  $W(k)$  包含了参数矩阵  $S$  中的元素, 基向量  $\phi[z(k)]$  式包含了向量  $z(k)$  的二次型。例如, 当  $z(k) = [x_1(k) \ x_2(k) \ u(k)]^T$  时, 参数向量  $W(k)$  为  $\begin{bmatrix} \frac{1}{2} S_{11} & S_{12} & S_{13} & \frac{1}{2} S_{22} & S_{23} & \frac{1}{2} S_{33} \end{bmatrix}^T$ , 基向量  $\phi[z(k)]$  为  $[x_1^2(k) \ x_1(k)x_2(k) \ x_1(k)u(k) \ x_2^2(k) \ x_2(k)u(k) \ u^2(k)]^T$ 。

根据  $Q$  函数的定义, 可知  $Q$  函数的值等于价值函数  $V(k)$ , 即  $V[x(k)] = Q[x(k), u(k)]$ 。相应的关于  $Q$  函数的贝尔曼方程为

$$Q[x(k), u(k)] = \frac{1}{2} [x^T(k) Q x(k) + u^T(k) R u(k)] + Q[x(k+1), u(k+1)] \quad (2-36)$$

将式 (2-35) 代入到上式 (2-36) 中, 那么此时  $Q$  函数的贝尔曼方程可以写成如下形式

$$W^T(k) \{\phi[z(k)] - \phi[z(k+1)]\} = \frac{1}{2} [x^T(k) Q x(k) + u^T(k) R u(k)] \quad (2-37)$$

这里可以使用迭代最小二乘法或者梯度下降法来在线辨识参数  $W(k)$ , 然后将辨识后的  $W(k)$  转换成参数矩阵  $S$ , 从而得到新的反馈控制增益  $K = S_{uu}^{-1} S_{ux}$ 。在第  $k$  时刻, 将当前的控制量  $u(k) = -Kx(k)$  应用到被控系统后, 可以实时观测到  $k+1$  时刻的状态  $x(k+1)$ , 并且应用当前的反馈控制增益得到  $k+1$  时刻控制量  $u(k+1) = -Kx(k+1)$ , 从而就能计算出基向量  $\phi[z(k)]$  和  $\phi[z(k+1)]$ 。在已知  $\{x(k), u(k), \phi[z(k)], \phi[z(k+1)]\}$  情况下, 可以使用迭代最小二乘法或梯度下降法更新参数向量  $W(k)$ 。该方法下的最优控制策略  $u^*(k) = \pi^*[x(k)]$  能够使最优  $Q$  函数最小

$$u^*(k) = \pi^*[x(k)] = \arg \min \left\{ \frac{1}{2} [x^T(k) Q x(k) + u^T(k) R u(k)] + Q^*[x(k+1), u(k+1)] \right\} \quad (2-38)$$

在使用该方程进行求解最优控制律时, 假设当前的得到的控制策略  $\hat{\pi}^*[x(k)]$  就是实际系统的最优策略, 即  $\hat{\pi}^*[x(k)] = \pi^*[x(k)]$ 。而用当前控制策略所估计的最优  $Q$  函数值  $\hat{Q}^*[x(k), \hat{\pi}^*[x(k)]]$ , 也被视为实际最优  $Q$  函数值, 即  $\hat{Q}^*[x(k), \hat{\pi}^*[x(k)]] = Q^*[x(k), \pi^*[x(k)]]$ 。因此, 值函数的估计误差并没有被考虑, 所以该方法具有确定性等价控制器的特性。

对于 LQR 问题, 在使用参数辨识技术估计  $Q$  函数的参数时, 得到并不是  $Q$  函数的真实参数  $W(k)$ , 而是估计值  $\hat{W}(k)$ , 存在一定的估计误差,  $\tilde{W}(k) = W(k) - \hat{W}(k)$ 。所以

用估计参数  $\hat{W}(k)$  来更新的控制策略  $\hat{\pi}^*[x(k)]$  并不是最优控制策略。

## 2.6 仿真结果及分析

本小节对以上介绍的基于确定性等价原理的自校正控制，具有谨慎特点的自校正控制，具有对偶特性的自校正控制，以及基于强化 Q-学习的自适应控制进行数值仿真实验及结果分析。

考虑式 (2-1) 差分方程形式的离散线性系统，参数真值设置为

$$a_1 = -1.41, \quad a_2 = 0.9, \quad b = 0.5, \quad m = 2, \quad n = 1 \quad (2-39)$$

其中系统噪声是高斯白噪声  $e(k) \sim N(0, 0.0025)$ 。系统的目标跟踪轨迹  $y_r$  设置为由传递函数为  $1/(s+1)$  的滤波器将 0.1Hz 的方波信号滤波后的数据。设置卡尔曼滤波的参数初值为  $\theta^T(1) = [0.1 \ 0.1 \ 0.1]$ ，参数协方差矩阵的初值为  $P(1) = 1000I_{3 \times 3}$ 。

实验对比了确定性等价控制，谨慎控制和  $\lambda = 0.85$  时的对偶控制的输出跟踪效果，输出跟踪误差和累计性能指标。在该实验中，需要确保系统的初始状态，系统噪声和跟踪目标是一样的，只有控制律是采用了不同的三种方法。

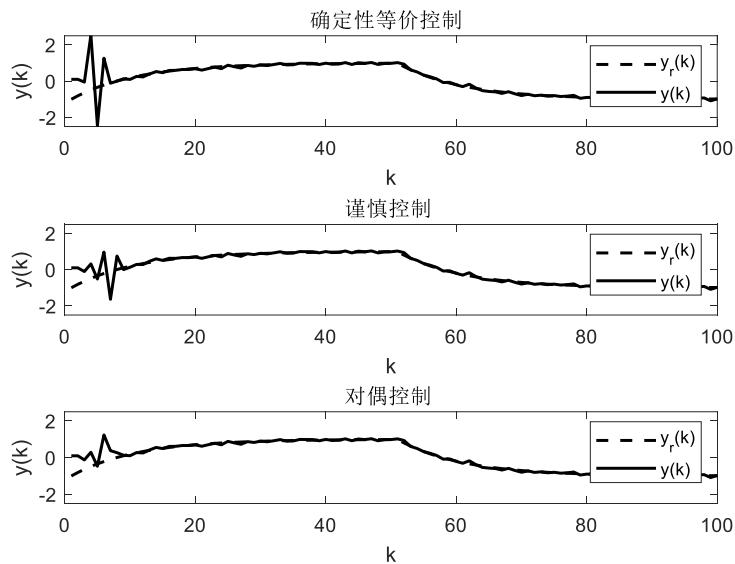


图 2-1 确定性等价控制，谨慎控制，自适应对偶控制的输出跟踪效果图

Fig.2-1 The system output tracking performance for certainty equivalence based control, cautious control and adaptive dual control

图 2-1 是在确定性等价控制，谨慎控制和对偶控制三种控制方法下，系统输出跟踪目标轨迹的效果图。可以看到在系统启动阶段，系统输出状态和理论分析一致，确定性等价控制的启动阶段的响应比较剧烈，具有较大的超调，因为其在设计控制律的时候没有将参数估计的误差考虑进去。经过启动阶段，大概运行 7 步之后参数收敛，系统的才显示出较为良好的输出跟踪效果。相比较而言，谨慎控制在启动阶段的响应比较慢，因

为在设计控制律的时候考虑到了参数的估计值是由误差的，谨慎控制大概在运行 11 步之后参数收敛。对偶控制即是在确定性等价控制和谨慎控制这两种极端情况之间取了一个折中的方案，在图中可以清楚地看到，对偶控制在启动阶段和确定性等价控制相比没有很大的超调，和谨慎控制相比能够更快的进行参数辨识，大概在运行 9 步之后参数就能够收敛，跟踪控制进入稳定状态。图 2-2 是在确定性等价控制，谨慎控制和对偶控制三种控制方法下，系统输出跟踪目标轨迹的误差图。

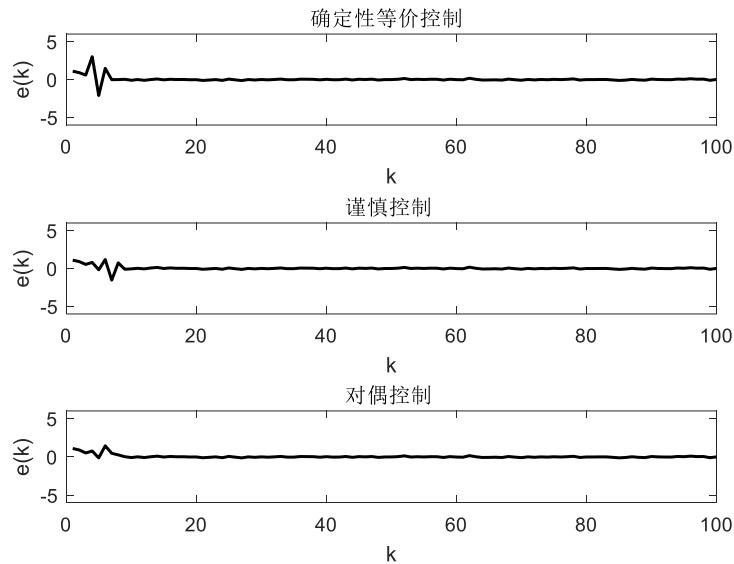


图 2-2 确定性等价控制，谨慎控制，自适应对偶控制的输出跟踪误差效果图  
Fig.2-2 The system output tracking error for certainty equivalence based control, cautious control and adaptive dual control

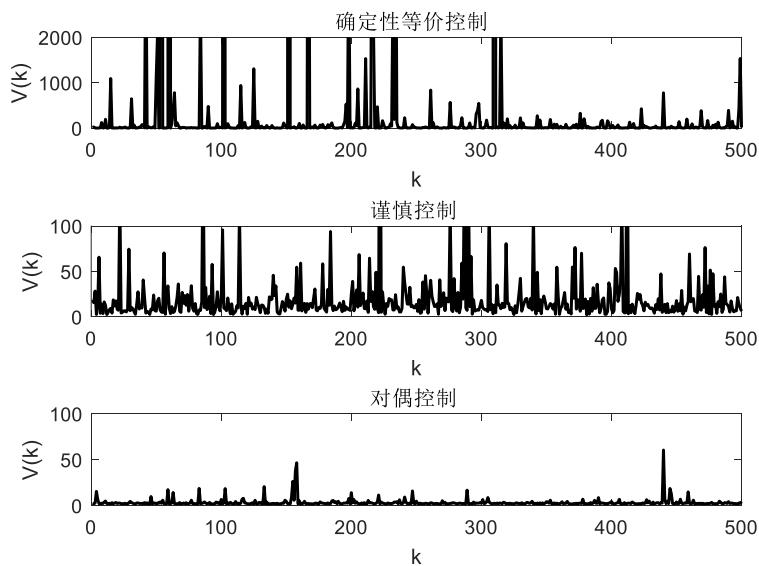


图 2-3 确定性等价控制，谨慎控制，自适应对偶控制的累计代价  
Fig.2-3 The accumulated costs for certainty equivalence based control, cautious control and adaptive dual control

为了量化三种控制方案下系统的性能指标，引入了累积代价

$$V(T) = \sum_{k=0}^T [y(k) - y_r(k)]^2 \quad (2-40)$$

其中  $T$  为仿真步长。在量化分析随机系统的控制性能指标时，需要做蒙特卡洛实验进行性能的量化分析，而不只是做一次实验。本实验进行了 500 次蒙特卡洛实验，得到了每次实验的累积代价。在图 2-3 中分别显示了三种控制方法下每次实验的累积代价值。从图中明显可以看到确定性等价控制的累积代价值最大，而对偶控制的累积代价值最小。分别计算这三种控制方案下 500 次蒙特卡洛实验的平均累积代价，其中确定性等价控制的平均累积代价值为 724.7429，谨慎控制的平均累积代价值为 20.1329，对偶控制的平均累积代价值为 3.1813。由此可以得出结论，无论是系统在初始起步阶段的瞬时响应，还是 500 次蒙特卡洛统计实验的平均累积代价值，对偶控制相比确定性等价控制和谨慎控制都表现出更为优秀的控制性能。

下面对基于强化 Q-学习的自适应控制进行实验并对结果进行分析。考虑式 (2-26) 的状态空间方程形式的离散线性系统，其中参数分别设置为

$$A = \begin{bmatrix} 0 & 1 \\ -0.16 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad Q = 1, \quad R = 1 \quad (2-41)$$

首先假设系统模型参数  $A$  和  $B$  已知，那么这就是经典的 LQR 问题。可以通过反向迭代求解数黎卡提方程即可得到最优核矩阵  $P^*$ ，从而计算出最优反馈控制的反馈增益  $K^*$  以及最优控制律  $u^*(k) = -K^*x(k)$ 。在本实验中，求得的最优核矩阵值为

$$P^* = \begin{bmatrix} 1.0186 & 0.1117 \\ 0.1117 & 2.6831 \end{bmatrix} \quad (2-42)$$

最优反馈增益为

$$K^* = \begin{bmatrix} -0.1166 & -0.6982 \end{bmatrix} \quad (2-43)$$

可以求得 Q 函数的最优核矩阵  $S^*$  的值为

$$S^* = \begin{bmatrix} 1.0687 & 0.4114 & -0.4293 \\ 0.4114 & 4.4783 & -2.5714 \\ -0.4293 & -2.5714 & 3.6831 \end{bmatrix} \quad (2-44)$$

假设模型参数  $A$  和  $B$  未知，那么本实验采用 Q-学习算法，在线学习 Q 函数中的核矩阵  $S$ ，从而计算出状态反馈控制增益  $K = S_{uu}^{-1}S_{ux}$ ，进而计算出基于确定性等价控制原理下的最优控制律。在进行 Q 函数参数辨识时，使用了最小二乘法，Q 函数核矩阵初值设置为

$$\hat{S}(1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2-45)$$

系统初值设置为

$$x(1) = \begin{bmatrix} x_1(1) \\ x_2(1) \end{bmatrix} = \begin{bmatrix} 10 \\ -10 \end{bmatrix} \quad (2-46)$$

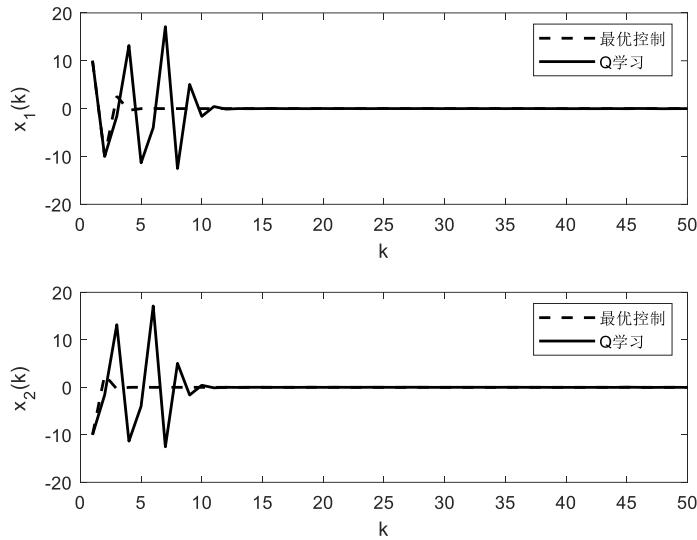


图 2-4 在最优控制和 Q-学习算法下系统状态  $x_1(k)$  和  $x_2(k)$  的调节过程  
Fig.2-4 The system state  $x_1(k)$  and  $x_2(k)$  for optimal control and Q-learning

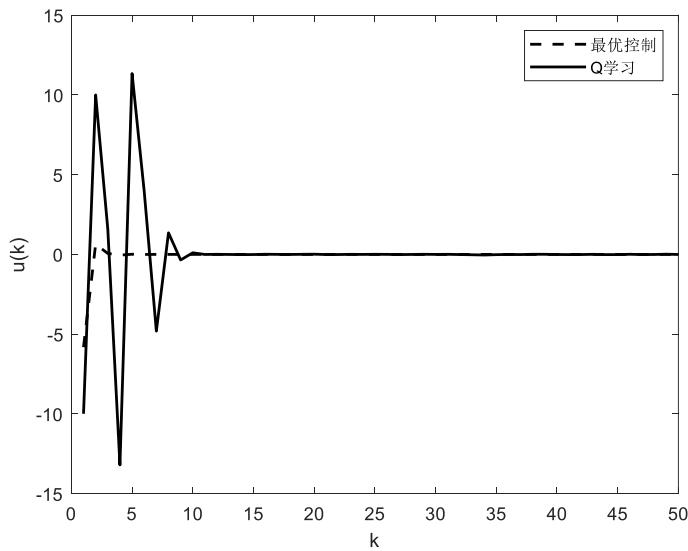


图 2-5 在最优控制和 Q-学习算法下的控制信号  $u(k)$   
Fig.2-5 The control signal  $u(k)$  for optimal control and Q-learning

图 2-4 在最优状态反馈控制和基于 Q-学习算法的自适应控制两种方法下系统状态  $x_1(k)$  和  $x_2(k)$  调节过程的效果图。其中最优控制是指系统模型参数完全已知的情况下，

通过线下求解代数黎卡提方程，得到最优状态反馈控制增益，用对应的最优控制律对系统进行状态调节。基于 Q-学习的自适应方法对应的系统模型参数完全未知，通过观测得到的反馈值直接学习 Q 函数的核矩阵  $\hat{S}(k)$ ，从而计算出相应的最优控制律对系统状态进行调节。从图 2-4 中可以看出，参数已知情况下的最优控制使得系统状态  $x_1(k)$  和  $x_2(k)$  在第 4 步就调节到 0，参数未知情况下的 Q-学习自适应控制需要学习时间，系统状态  $x_1(k)$  和  $x_2(k)$  在第 11 步调节到了 0，且状态在调节过程中超调比较大。

图 2-5 显示了最优控制和基于 Q-学习算法的自适应控制两种方法下系统的控制信号  $u(k)$ ，明显看出在第 1 步到第 10 步之间，Q-学习的控制信号幅值比较大。

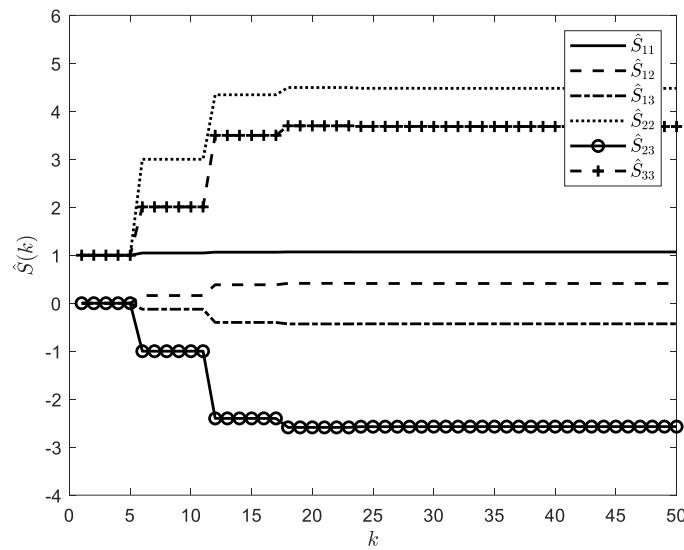


图 2-6 Q-学习算法下 Q 函数的核矩阵  $\hat{S}(k)$  的收敛过程

Fig.2-6 The kernel matrix  $\hat{S}(k)$  for Q-learning based adaptive control

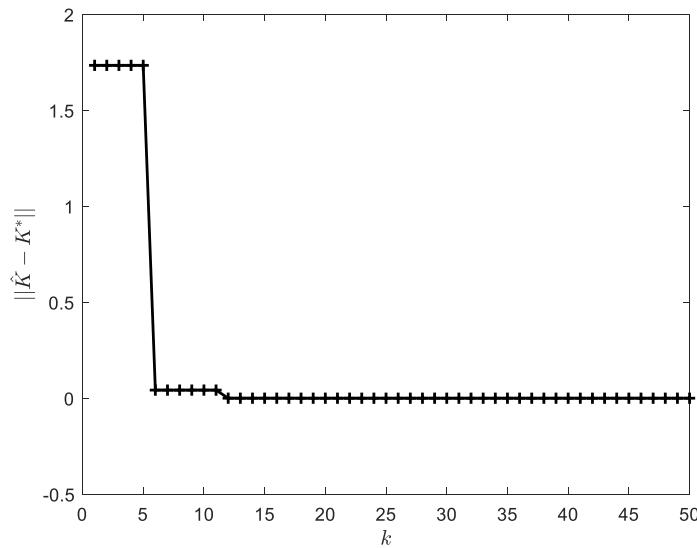


图 2-7 在 Q-学习算法下状态反馈控制的增益  $\hat{K}$  收敛到  $K^*$  的过程

Fig.2-7 The convergence of state feedback gain  $\hat{K}$  for Q-learning based adaptive control

图 2-6 是参数未知的系统在 Q-学习算法下, Q 函数的核矩阵  $\hat{S}(k)$  的学习收敛过程。从图中可以看出, 核矩阵  $\hat{S}(k)$  在第 18 步收敛到最优核矩阵  $S^*$ , 对应每个元素为  $\hat{S}_{11}(18)=1.0691$  ,  $\hat{S}_{12}(18)=0.4140$  ,  $\hat{S}_{13}(18)=-0.4316$  ,  $\hat{S}_{22}(18)=4.4956$  ,  $\hat{S}_{23}(18)=-2.5874$  ,  $\hat{S}_{33}(18)=3.6975$  。

图 2-7 显示了在 Q-学习算法下状态反馈控制的增益  $\hat{K}$  收敛到  $K^*$  的过程, 其中纵轴显示的是在线学习的反馈增益  $\hat{K}$  与最优反馈增益  $K^*$  差值的范数。从图中明显可以看出  $\|\hat{K}-K^*\|$  第 11 步就趋向于 0, 也就是反馈增益  $\hat{K}$  在第 11 步就收敛到最优反馈增益值  $K^*$ , 表明系统在第 11 步学习到了最优控制律  $u^*$  。

## 2.7 本章小结

本章通过数学模型简单介绍了基于确定性等价原理的自校正控制, 具有谨慎特点的自校正控制, 具有对偶特性的自校正控制, 以及基于强化 Q-学习的自适应控制的原理, 并分别进行仿真实验和比较分析。通过理论分析和仿真实验结果表明, 具有对偶特性的自校正控制的性能指标明显优于确定性等价原理下的自校正控制和谨慎控制。另外对目前传统的基于强化 Q-学习的自适应控制进行分析, 分析结果表明该方法是基于确定性等价原理的。本章通过仿真实验将基于强化 Q-学习的自适应控制与最优控制对比, 发现 Q-学习的自适应控制在学习阶段有较大的超调。接下来的第三章将对这一问题进行进一步研究和探索。



## 3 具有探索-利用平衡特性的 Q-学习自适应控制

### 3.1 引言

本章节研究了未知系统的线性二次调节（Linear Quadratic Regulation, LQR）问题。该问题的控制目标是通过优化控制信号  $u(k)$  来驱动系统到达期望的系统状态  $x(k)$ <sup>[99]</sup>。在确定性系统中，可以利用系统状态方程直接计算出最优控制律  $u^*(k) = \phi(k, x(k))$ ；而对于未知系统，只有包含控制序列  $\{u(k), u(k-1), \dots, u(1)\}$  和系统状态  $\{x(k), x(k-1), \dots, x(1)\}$  的实时状态信息  $z(k)$  可以使用<sup>[100]</sup>，因而无法直接求解最优控制律。绝大部分实际系统，如生态系统和经济系统，通常都是未知的，因此有必要对未知系统的 LQR 问题进行研究。

Watkins 首次提出了 Q-学习方法，也被称为动作依赖启发式动态规划(Action dependent heuristic dynamic programming, AD-HDP)，该方法旨在解决未知系统的 LQR 问题<sup>[97]</sup>。随后 Landelius 为 Q-学习的方法可以求解参数未知 LQR 问题中的黎卡提方程提供了有力证明<sup>[101]</sup>。该方法在一些实际系统中已得到应用，例如 Masoud 等人设计了一种基于 Q-学习算法的主频率控制器，用于调节智能电网中存在机械参数、负载、扰动等各种未知特征的现代化微电网的频率<sup>[102]</sup>。Shi 等人开发了多个伪 Q-学习策略来解决参数未知的自主水下航行器的轨迹跟踪控制问题<sup>[103]</sup>。插电式混合动力汽车最优能量控制的研究中由于系统参数未知，也采用了 Q-学习的方法来求解控制律<sup>[104]</sup>。

然而传统的基于 Q-学习的控制方法都存在忽视参数不确定性的方法，即在 Q-学习的策略评估过程中，将学习到的 Q 函数的核矩阵的估计值作为真值直接用于控制律推导，该方法遵循确定性等效原理，但忽略了 Q 函数核矩阵的估计误差<sup>[101]</sup>，因此，在确定性等价原理下的 Q-学习控制方法并不是最优的。在 Q 函数核矩阵的试错学习中，传统的基于 Q-学习的控制策略都是在控制律中添加一个探索信号，直接对 Q 函数核矩阵进行探索学习，也就是对系统未知信息的探索，这就是强化学习中的探索。通常所取的探索信号值比较大，因此会对系统进行过度探索，会导致系统在试错学习阶段具有较大的超调。在许多实际的系统中，较大的超调会直接损坏系统，因而导致该方法无法得到实际应用<sup>[105]</sup>。与基于确定性等价原理的控制方法相比，谨慎控制是一种更为保守的控制策略，它在利用当前采集的信息来推导控制策略时过度考虑了估计误差，这就是强化学习中的利用。在探测信息较少的情况下，谨慎控制将具有较长的调节时间，甚至系统可能会产生关断效应，特别是在系统参数出现比较大的不确定性时<sup>[106]</sup>。因此，如何在基于强化 Q-学习的 LQR 问题中平衡探索与利用的冲突关系，以提高未知系统在试错学习阶段的控制性能极具挑战性。

本章提出一种新的基于 Q-学习的线性二次型调节器，可以平衡未知系统的探索和与利用之间的冲突。该方法通过构造两个性能指标函数来解决 LQR 问题，一个是系统状态跟踪性能指标，即通过最小化状态调节误差得到目标状态，衡量了系统利用的程度，另

一个是  $Q$  函数核矩阵估计性能指标，即通过最大化信息增益来促进系统对未知信息的学习，代表系统探索的程度。同时对上述两个性能指标函数进行双目标优化问题的求解，可以得到探索于利用平衡的控制策略。其中采用双准则方法来求解这个双目标优化问题，所得到的控制策略包含两部分：（1）能够让系统状态进行调节的谨慎控制信号；（2）增加学习信息的探索信号。通过对参数未知的一阶线性系统和二阶线性系统的数值仿真实验，说明所提出的控制策略相比于传统方法可以改善和提高系统的控制性能。

### 3.2 未知系统的线性二次调节问题

考虑如下离散时间线性随机系统

$$x(k+1) = A(k)x(k) + B(k)u(k) + \omega(k+1) \quad (3-1)$$

其中  $x(k) \in \mathbb{R}^n$  是系统状态， $u(k) \in \mathbb{R}^r$  是控制信号， $\omega(k)$  是高斯白噪声

$$\omega(k) \sim \mathcal{N}(0, R_w) \quad (3-2)$$

系统参数由矩阵  $A(k)$  和  $B(k)$  表示。假设这两个参数未知，可以用随机变量  $\theta(k)$  来表示，写成  $A(\theta(k))$  和  $B(\theta(k))$  的形式。假设随机变量  $\theta(k)$  也服从高斯分布

$$\theta(k) \sim \mathcal{N}(\bar{\theta}, R_\theta) \quad (3-3)$$

并且定义参数  $\theta(k)$  的状态方程为

$$\theta(k+1) = \theta(k) + \delta(k+1) \quad (3-4)$$

其中  $\delta(k)$  为  $n$  维高斯白噪声

$$\delta_i(k) \sim \mathcal{N}(0, R_{\delta_i}), i \in \{1, 2, \dots, n\} \quad (3-5)$$

其中假设  $\delta(k)$  与  $\omega(k)$  相互独立。针对随机系统的 LQR 问题，需要使用期望计算，因此值函数定义为

$$V(k) = E \left\{ \sum_{i=k}^N \gamma^{i-k} r[x(i), u(i)] \middle| \mathfrak{I}_k \right\} \quad (3-6)$$

其中  $r[x(i), u(i)]$  是一步代价函数，定义为

$$r[x(i), u(i)] = x^T(i)Qx(i) + u^T(i)Ru(i) \quad (3-7)$$

信息状态  $\mathfrak{I}_k$  定义为

$$\mathfrak{I}_k = \{x(k), x(k-1), \dots, x(1), u(k-1), \dots, u(1)\} \quad (3-8)$$

这里采用了折扣因子  $0 < \gamma < 1$ ，该折扣因子可以对未来时刻的代价赋予较小的权重，距离当前时刻较近的代价赋予较高的权重，从而减少未来时刻对当前决策的影响，让策略更注重近期的收益。因此 LQR 问题的目标是找到一个控制策略  $u^*(k)$  能够最小化式 (3-6) 的价值函数，该价值函数可以写成如下形式

$$V^*(k) = \min_{u^*(k)} E \left\{ \sum_{i=k}^N \gamma^{i-k} r[x(i), u(i)] | \mathfrak{I}_k \right\} \quad (3-9)$$

将值函数 (3-6) 重写为

$$V(k) = E \left\{ r[x(k), u(k)] + \sum_{i=k+1}^N \gamma^{i-(k+1)} r[x(i), u(i)] | \mathfrak{I}_k \right\} \quad (3-10)$$

那么可以得到

$$V(k) = E \{ r[x(k), u(k)] + \gamma V(k+1) | \mathfrak{I}_k \} \quad (3-11)$$

这就是贝尔曼方程，因此式 (3-9) 的最优值函数可以写成贝尔曼最优方程

$$V^*(k) = \min_{u(k)} E \{ r[x(k), u(k)] + \gamma V^*(k+1) | \mathfrak{I}_k \} \quad (3-12)$$

那么最优控制策略由下式可以求得

$$u^*(k) = \arg \min_{u(k)} E \{ r[x(k), u(k)] + \gamma V^*(k+1) | \mathfrak{I}_k \} \quad (3-13)$$

### 3.3 基于 Q-学习的具有探索和利用平衡特性的自适应控制策略

本节根据上一小节中提出的控制问题，提出了一种探索和利用相平衡的基于 Q-学习的未知系统控制方案。该方案通过优化两个代价函数来实现：一是最小化系统状态调节以实现最优利用；另一个最大化信息增益以实现最优探索。控制策略的详细推导如下。

首先根据上一小节的内容定义 Q 函数

$$Q(x(k), u(k)) = E \{ r[x(k), u(k)] + \gamma V(k+1) | \mathfrak{I}_k \} \quad (3-14)$$

其中  $Q[x(k), u(k)] = V(k)$ ，因此 Q 函数也是二次型

$$V(k) = x^T(k) S x(k) \quad (3-15)$$

将式 (3-1)，(3-7)，以及 (3-15) 代入式 (3-14)，可以得到如下形式的 Q 函数

$$\begin{aligned} Q[x(k), u(k)] &= E \{ x^T(k) Q x(k) + u^T(k) R u(k) \\ &\quad + \gamma [A(k)x(k) + B(k)u(k) + w(k+1)]^T \\ &\quad S [A(k)x(k) + B(k)u(k) + w(k+1)] \} \end{aligned} \quad (3-16)$$

为便于式 (3-16) 中的期望计算，对式中的参数做如下定义

$$\begin{aligned} A(k) &= \hat{A}(k) + \tilde{A}(k) \\ B(k) &= \hat{B}(k) + \tilde{B}(k) \end{aligned} \quad (3-17)$$

其中参数  $\hat{A}(k)$  是  $A(k)$  的估计值，是在信息状态  $\mathfrak{I}_k$  下的估计  $\hat{A}(k) = E\{A(k) | \mathfrak{I}_k\}$ ，参数  $\tilde{A}(k)$  是估计误差  $\tilde{A}(k) = E\{[A(k) - \hat{A}(k)] | \mathfrak{I}_k\}$ ，同样的参数  $\hat{B}(k)$  是  $B(k)$  的估计值， $\tilde{B}(k)$  是参数估计误差  $\tilde{B}(k) = E\{[B(k) - \hat{B}(k)] | \mathfrak{I}_k\}$ 。将 (3-17) 代入到式 (3-16) 中，通过计算

可以得到 Q 函数

$$Q[x(k), u(k)] = \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}^T \left\{ \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} + \gamma \begin{bmatrix} P_{xx} & P_{xu} \\ P_{ux} & P_{uu} \end{bmatrix} + \gamma \begin{bmatrix} \hat{T}_{xx} & \hat{T}_{xu} \\ \hat{T}_{ux} & \hat{T}_{uu} \end{bmatrix} \right\} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} + \text{tr}\{SR_w\} \quad (3-18)$$

该式的具体计算过程如下所示。将式 (3-17) 代入到 Q 函数 (3-16) 中可得

$$\begin{aligned} Q[x(k), u(k)] &= E\{[x^T(k)Qx(k) + u^T(k)Ru(k)] \\ &+ \gamma[\tilde{A}(k)x(k) + \tilde{B}(k)u(k)]^T S[\tilde{A}(k)x(k) + \tilde{B}(k)u(k)] \\ &+ \gamma[\hat{A}(k)x(k) + \hat{B}(k)u(k) + w(k+1)]^T S[\hat{A}(k)x(k) + \hat{B}(k)u(k) + w(k+1)] | \mathfrak{I}_k\} \end{aligned} \quad (3-19)$$

首先计算上式的第二个部分

$$\begin{aligned} &E\{[\tilde{A}(k)x(k) + \tilde{B}(k)u(k)]^T S[\tilde{A}(k)x(k) + \tilde{B}(k)u(k)]\} \\ &= E\left\{\begin{bmatrix} x(k) \\ u(k) \end{bmatrix}^T \begin{bmatrix} \tilde{A}^T(k)S\tilde{A}(k) & \tilde{A}^T(k)S\tilde{B}(k) \\ \tilde{B}^T(k)S\tilde{A}(k) & \tilde{B}^T(k)S\tilde{B}(k) \end{bmatrix} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}\right\} \end{aligned} \quad (3-20)$$

定义参数矩阵  $\tilde{T}$

$$\tilde{T} = \begin{bmatrix} \tilde{A}^T(k)S\tilde{A}(k) & \tilde{A}^T(k)S\tilde{B}(k) \\ \tilde{B}^T(k)S\tilde{A}(k) & \tilde{B}^T(k)S\tilde{B}(k) \end{bmatrix} = \begin{bmatrix} \tilde{T}_{xx} & \tilde{T}_{xu} \\ \tilde{T}_{ux} & \tilde{T}_{uu} \end{bmatrix} \quad (3-21)$$

则对应的期望为

$$E\{\tilde{T}\} = P_T(k) = \begin{bmatrix} P_{xx}(k) & P_{xu}(k) \\ P_{ux}(k) & P_{uu}(k) \end{bmatrix} = E\left\{\begin{bmatrix} \tilde{A}^T(k)S\tilde{A}(k) & \tilde{A}^T(k)S\tilde{B}(k) \\ \tilde{B}^T(k)S\tilde{A}(k) & \tilde{B}^T(k)S\tilde{B}(k) \end{bmatrix}\right\} \quad (3-22)$$

将上式代入式 (3-20) 可得

$$\begin{aligned} &E\{[\tilde{A}(k)x(k) + \tilde{B}(k)u(k)]^T S[\tilde{A}(k)x(k) + \tilde{B}(k)u(k)]\} \\ &= \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}^T \begin{bmatrix} P_{xx}(k) & P_{xu}(k) \\ P_{ux}(k) & P_{uu}(k) \end{bmatrix} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \end{aligned} \quad (3-23)$$

然后对式 (3-19) 的第三个部分进行计算

$$\begin{aligned} &E\{[\hat{A}(k)x(k) + \hat{B}(k)u(k) + w(k+1)]^T S[\hat{A}(k)x(k) + \hat{B}(k)u(k) + w(k+1)]\} \\ &= \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}^T \begin{bmatrix} \hat{A}^T(k)S\hat{A}(k) & \hat{A}^T(k)S\hat{B}(k) \\ \hat{B}^T(k)S\hat{A}(k) & \hat{B}^T(k)S\hat{B}(k) \end{bmatrix} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} + \text{tr}\{SR_w\} \end{aligned} \quad (3-24)$$

定义矩阵

$$T = \begin{bmatrix} \hat{A}^T(k)S\hat{A}(k) & \hat{A}^T(k)S\hat{B}(k) \\ \hat{B}^T(k)S\hat{A}(k) & \hat{B}^T(k)S\hat{B}(k) \end{bmatrix} = \begin{bmatrix} \hat{T}_{xx}(k) & \hat{T}_{xu}(k) \\ \hat{T}_{ux}(k) & \hat{T}_{uu}(k) \end{bmatrix} \quad (3-25)$$

将该矩阵代入到式 (3-24) 得

$$\begin{aligned}
& E \left\{ [\hat{A}(k)x(k) + \hat{B}(k)u(k) + w(k+1)]^T S [\hat{A}(k)x(k) + \hat{B}(k)u(k) + w(k+1)] \right\} \\
&= \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}^T \begin{bmatrix} \hat{T}_{xx}(k) & \hat{T}_{xu}(k) \\ \hat{T}_{ux}(k) & \hat{T}_{uu}(k) \end{bmatrix} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} + \text{tr} \{ SR_w \}
\end{aligned} \quad (3-26)$$

将式 (3-24) 和 (3-26) 代入到 Q 函数中即可得到结果 (3-18)，推导结束。

定义矩阵  $H$  为

$$H = \gamma \begin{bmatrix} P_{xx} & P_{xu} \\ P_{ux} & P_{uu} \end{bmatrix} + \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} + \gamma \begin{bmatrix} \hat{T}_{xx} & \hat{T}_{xu} \\ \hat{T}_{ux} & \hat{T}_{uu} \end{bmatrix} = \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \quad (3-27)$$

以及向量  $z(k) = [x(k) \ u(k)]^T$ 。因此 Q 函数可以化简为

$$Q[x(k), u(k)] = E \{ z^T(k) H z(k) + \text{tr} \{ SR_w \} \} \quad (3-28)$$

将  $H(k) = \hat{H}(k) + \tilde{H}(k)$  代入到式 (3-28) 可得

$$\begin{aligned}
Q[x(k), u(k)] &= E \{ z^T(k) [\hat{H}(k) + \tilde{H}(k)] z(k) \} + \text{tr} \{ SR_w \} \\
&= z^T(k) E \{ \tilde{H}(k) \} z(k) + z^T(k) \hat{H}(k) z(k) + \text{tr} \{ SR_w \}
\end{aligned} \quad (3-29)$$

定义核矩阵  $H(k)$  的估计误差协方差为

$$E \{ \tilde{H}(k) \} = P^H = \begin{bmatrix} P_{xx}^H & P_{xu}^H \\ P_{ux}^H & P_{uu}^H \end{bmatrix} \quad (3-30)$$

将上式协方差矩阵代入到式 (3-29) 中可得到

$$Q[x(k), u(k)] = u^T(k) (P_{uu}^H + \hat{H}_{uu}) u(k) + 2u^T(k) (P_{ux}^H + \hat{H}_{ux}) x(k) + C \quad (3-31)$$

其中  $C$  是常数。式 (3-31) 关于  $u(k)$  求偏导，并且令  $\partial Q[x(k), u(k)] / \partial u(k) = 0$ ，可以求出具有谨慎特性的控制律

$$u_{caut}(k) = -[P_{uu}^H + \hat{H}_{uu}]^{-1} [P_{ux}^H + \hat{H}_{ux}] x(k) \quad (3-32)$$

接下来求解式 (3-32) 中的核矩阵  $H(k)$  的估计值  $\hat{H}$ ，以及估计误差协方差矩阵  $P^H$ 。为了便于后续对核矩阵的实时学习，将二次型  $z^T(k) H z(k)$  转换成 Kronecker 乘积的形式  $z^T(k) H z(k) = \text{vec}(H) z(k) \otimes z(k)$ ，其中  $\text{vec}(H)$  是把核矩阵  $H$  的列向量排成单个列向量，并且令  $H_v = \text{vec}(H)$ ， $\otimes$  是 Kronecker 乘积，并且令  $z_v(k) = z(k) \otimes z(k)$ ，因此二次型可以简写为  $z^T(k) H z(k) = H_v z_v(k)$ ，将其代入式 (3-14) 定义的 Q 函数，可以得到如下形式的贝尔曼方程

$$H_v z_v(k) = x^T(k) Q x(k) + u^T(k) R u(k) + \gamma H_v z_v(k+1) \quad (3-33)$$

根据式 (3-33)，可以利用卡尔曼滤波对未知向量  $H_v$  进行估计，迭代过程如下所示

$$\begin{aligned}
\hat{H}_v(k+1) &= \hat{H}_v(k) + K(k+1) \left[ Q_v(k+1) - \hat{H}_v(k) z_v(k) \right] \\
K(k+1) &= P^{H_v}(k) z_v(k) \left[ R_v + z_v^T(k) P^{H_v}(k) z_v(k) \right]^{-1} \\
P^{H_v}(k+1) &= [I - K(k+1) z_v(k)] P^{H_v}(k)
\end{aligned} \tag{3-34}$$

通过使用式(3-34)的迭代计算实时更新矩阵向量  $\hat{H}_v$  和估计误差协方差矩阵  $P^{H_v}$ , 可以得到基于 Q-学习的谨慎控制律, 该控制律考虑到了估计误差对系统的影响。然而, 谨慎控制是一种相对保守的控制方法, 该方法限制了系统的探索功能, 这将会降低系统对未知信息的学习性能, 甚至会产生关断效应。因此有必要对系统施加一个持续的激励信号去刺激系统探索未知信息。因此, 在控制律中增加了一个探索激励信号, 则控制信号可以写成如下形式

$$u(k) = u_{caut}(k) + u_{prob}(k) \tag{3-35}$$

其中  $u_{prob}(k)$  为探索信号。在大多数 Q-学习自适应控制方案中,  $u_{prob}(k)$  通常是人为设定的随机数据或者是一个函数来生成探索信号, 并且该信号在系统学习过程收敛后会终止使用。也就是说  $u_{prob}(k)$  不是自动调节的而是固定选取的, 过去的文献通常会选取一个较大的探索信号以探索到更多的未知信息, 而由此产生的代价是系统的控制在探索阶段会有很大的超调, 这在实际应用中不切实际, 往往会给系统带来永久损害。同样, 如果选取了较小的探索信号, 那么对系统未知部分的学习也不够充分, 也会降低系统的控制性能。

接下来设计能主动调节探测信号的方法, 为保证探测信号能够增加在系统试错学习阶段进行探索时所需信息的丰富程度, 并且还能够保持系统控制在探索和利用之间的平衡, 设计了两个代价函数并构成双目标优化问题, 并且使用了双准则的方法求解该双目标优化问题, 得到的探索信号可以调节系统探索与利用之间的冲突。两个代价函数分别为(1)最小化系统的调节误差, 也就是利用; (2)最大化信息的丰富程度, 也就是探索。这里的前一个代价函数就是式(3-18)所表示的 Q 函数, 第二个代价函数就是量化系统不确定性的函数, 如下所示

$$Q_a(k) = E \left\{ z^T(k) \tilde{H}(k) z(k) \right\} = x^T(k) P_{xx}^H x(k) + 2x^T(k) P_{xu}^H u(k) + u^T(k) P_{uu}^H u(k) \tag{3-36}$$

通过最大化该代价函数  $Q_a(k)$  来增加系统探索未知部分所需信息的丰富程度, 从而得对应的最优解  $u_a^*(k)$

$$Q_a^*(k) = \max_{u_a^*(k)} \left\{ x^T(k) P_{xx}^H x(k) + 2x^T(k) P_{xu}^H u(k) + u^T(k) P_{uu}^H u(k) \right\} \tag{3-37}$$

接下来就是找到能够同时满足这两个优化问题的最优控制律  $u^*(k)$ , 则控制律求解问题可以转换成双目标优化问题

$$\begin{aligned} & \min_{u(k)} \{Q(k), -Q_a(k)\} \\ s.t. \quad & x(k+1) = A(k)x(k) + B(k)u(k) + w(k+1) \\ & \theta(k+1) = \theta(k) + \delta(k+1) \end{aligned} \quad (3-38)$$

这里使用双准则方法来求解最优控制律  $u^*(k)$ ，可以同时最小化  $Q(k)$  和  $-Q_a(k)$ 。解的形式如下

$$u^*(k) = u_{caut}(k) + u_{prob}(k) \quad (3-39)$$

其中

$$u_{prob}(k) = \sigma(k) \operatorname{sign}\{\beta(k)\} \quad (3-40)$$

$$\Omega(k) = [u_{caut}(k) - \sigma(k), u_{caut}(k) + \sigma(k)] \quad (3-41)$$

$$\sigma(k) = \alpha \operatorname{tr}\{P^H(k)\}, \alpha \geq 0 \quad (3-42)$$

$$\beta(k) = x^T(k)P_{xu}^H + P_{uu}^H u(k) \quad (3-43)$$

下面详细给出了控制律  $u^*(k)$  的推导过程。根据双准则方法， $u^*(k)$  是通过求解在以谨慎控制律为中心的一个区间内的代价函数  $Q_a(k)$  的最大值得到，该区间为  $\Omega(k) = [u_{caut}(k) - \sigma(k), u_{caut}(k) + \sigma(k)]$ ，其中  $\sigma(k)$  为误差估计协方差矩阵的迹。通过比较在区间边界对应的值，并选择最大的值所对应的点

$$\begin{aligned} & Q_a[u_{caut}(k) - \sigma(k)] - Q_a[u_{caut}(k) + \sigma(k)] \\ &= 2x^T(k)P_{xu}^H[u_{caut}(k) - \sigma(k)] + [u_{caut}(k) - \sigma(k)]^T P_{uu}^H [u_{caut}(k) - \sigma(k)] \\ &\quad - 2x^T(k)P_{xu}^H[u_{caut}(k) + \sigma(k)] - [u_{caut}(k) + \sigma(k)]^T P_{uu}^H [u_{caut}(k) + \sigma(k)] \\ &= -4\sigma(k)[x^T(k)P_{xu}^H + P_{uu}^H u_{caut}(k)] \end{aligned} \quad (3-44)$$

为简化表达，定义

$$\beta(k) = x^T(k)P_{xu}^H + P_{uu}^H u_{caut}(k) \quad (3-45)$$

如果  $\beta(k)$  小于 0，这表明  $Q_a[u_{caut}(k) - \sigma(k)]$  比  $Q_a[u_{caut}(k) + \sigma(k)]$  大，那么控制律为  $u_{caut}(k) - \sigma(k)$ 。如果  $\beta(k)$  大于 0，这表明  $Q_a[u_{caut}(k) - \sigma(k)]$  比  $Q_a[u_{caut}(k) + \sigma(k)]$  小，那么控制律为  $u_{caut}(k) + \sigma(k)$ 。在式 (3-39) 中用函数  $\operatorname{sign}\{x\}$  来表达。

下面详细给出估计误差写方程矩阵  $P^H$  的计算过程，核矩阵  $H$  可以写成如下形式

$$H = \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1n} \\ H_{21} & H_{22} & \dots & H_{2n} \\ \vdots & \vdots & & \vdots \\ H_{n1} & H_{n2} & \dots & H_{nn} \end{bmatrix} \quad (3-46)$$

将其改写成向量形式为

$$H_v = [H_{11}, \dots, H_{n1}, H_{12}, \dots, H_{n2}, \dots, H_{nn}]^T \quad (3-47)$$

则对应的向量形式的核矩阵的估计误差协方差矩阵为

$$\begin{aligned}
 P^{H_v} &= E\left\{[H_v - \hat{H}_v]^T [H_v - \hat{H}_v]\right\} = E\left\{\tilde{H}_v^T \tilde{H}_v\right\} \\
 &= E\left\{\begin{bmatrix} \tilde{H}_{11}\tilde{H}_{11} & \tilde{H}_{11}\tilde{H}_{21} & \dots & \tilde{H}_{11}\tilde{H}_{nn} \\ \tilde{H}_{21}\tilde{H}_{11} & \tilde{H}_{21}\tilde{H}_{21} & \dots & \tilde{H}_{21}\tilde{H}_{nn} \\ \vdots & \vdots & & \vdots \\ \tilde{H}_{n1}\tilde{H}_{11} & \tilde{H}_{n1}\tilde{H}_{21} & \dots & \tilde{H}_{n1}\tilde{H}_{nn} \\ \tilde{H}_{12}\tilde{H}_{11} & \tilde{H}_{12}\tilde{H}_{21} & \dots & \tilde{H}_{12}\tilde{H}_{nn} \\ \vdots & \vdots & & \vdots \\ \tilde{H}_{n2}\tilde{H}_{11} & \tilde{H}_{n2}\tilde{H}_{21} & \dots & \tilde{H}_{n2}\tilde{H}_{nn} \\ \vdots & \vdots & & \vdots \\ \tilde{H}_{nn}\tilde{H}_{11} & \tilde{H}_{nn}\tilde{H}_{21} & \dots & \tilde{H}_{nn}\tilde{H}_{nn} \end{bmatrix}\right\} \\
 &= \begin{bmatrix} E\{\tilde{H}_{11}\tilde{H}_{11}\} & E\{\tilde{H}_{11}\tilde{H}_{21}\} & \dots & E\{\tilde{H}_{11}\tilde{H}_{nn}\} \\ E\{\tilde{H}_{21}\tilde{H}_{11}\} & E\{\tilde{H}_{21}\tilde{H}_{21}\} & \dots & E\{\tilde{H}_{21}\tilde{H}_{nn}\} \\ \vdots & \vdots & & \vdots \\ E\{\tilde{H}_{nn}\tilde{H}_{11}\} & E\{\tilde{H}_{nn}\tilde{H}_{21}\} & \dots & E\{\tilde{H}_{nn}\tilde{H}_{nn}\} \end{bmatrix} \tag{3-48}
 \end{aligned}$$

该矩阵的主对角线元素的值为

$$E\{\tilde{H}_{ij}\tilde{H}_{ij}\} = P_{i+(j-1)n, i+(j-1)n}^{H_v} \tag{3-49}$$

为便于表达将  $P_{i+(j-1)n, i+(j-1)n}^{H_v}$  写为  $P_{i+(j-1)n}^{H_v}$ , 由式 (3-48) 得  $\tilde{H}_{ij}$  的期望值为

$$E\{\tilde{H}_{ij}\} = \sqrt{\frac{P_{i+(j-1)n}^{H_v}}{2}} \tag{3-50}$$

根据式 (3-46) 的形式,  $P^H$  估计误差协方差矩阵可以写成如下形式

$$P^H = E\{\tilde{H}\} = \sqrt{\frac{1}{2}} \begin{bmatrix} \sqrt{P_1^{H_v}} & \sqrt{P_2^{H_v}} & \dots & \sqrt{P_n^{H_v}} \\ \sqrt{P_{1+n}^{H_v}} & \sqrt{P_{2+n}^{H_v}} & \dots & \sqrt{P_{n+n}^{H_v}} \\ \vdots & \vdots & & \vdots \\ \sqrt{P_{1+(n-1)n}^{H_v}} & \sqrt{P_{2+(n-1)n}^{H_v}} & \dots & \sqrt{P_{n+(n-1)n}^{H_v}} \end{bmatrix} \tag{3-51}$$

### 3.4 仿真实验

本节通过数值仿真实验验证了所提出的基于 Q-学习的具有探索与利用平衡特性的自适应控制方法的有效性, 并且将其与基于确定性等价原理的 Q-学习方法和参数已知的最优控制方法进行了比较。实验分别用一阶线性系统和二阶线性系统测试了所提出的 Q-学习方法, 特别是在探索阶段也就是试错学习阶段, 对探索信号如何平衡系统的状态调节和系统未知部分的探索进行了深入的分析。

以下为本章所提出的探索与利用相平衡的 Q-学习自适应控制方法的执行步骤：

#### 初始化：

初始化参数向量  $H_v(:,1)$ ，信息状态  $z(1) = [x(1); u(1)]$ ，估计误差协方差矩阵  $P^{H_v}(:,1)$ ，参数  $\alpha > 0$ 。

#### 迭代过程：

- (1) 利用迭代公式 (3-34) 更新  $k+1$  时刻估计的核矩阵向量  $\hat{H}_v(:,k+1)$ 。
- (2) 根据  $\hat{H}_v(:,k+1)$  计算核矩阵  $\hat{H}(::,k+1)$ 。
- (3) 根据  $P^{H_v}(:,:,k+1)$  计算核矩阵估计误差协方差  $P^H(:,:,k+1)$ 。
- (4) 由式 (3-32) 计算谨慎控制律  $u_{caut}(k+1)$ 。
- (5) 由式 (3-43) 计算判据  $\beta(k+1)$ 。
- (6) 如果判据  $\beta(k+1)$  大于 0，探索信号就设置为  $\delta(k+1)$ ；如果判据  $\beta(k+1)$  小于 0，探索信号就设置为  $-\delta(k+1)$ ；如果判据  $\beta(k+1)$  等于 0，探索信号设置为 0。
- (7) 根据式 (3-39) 和 (3-40) 计算控制律  $u^*(k+1)$ 。
- (8) 将控制律  $u^*(k+1)$  施加到系统中，然后重复执行步骤 (1) 到步骤 (7)。

#### 3.4.1 一阶线性系统

本实验的被控对象为一个参数未知的一阶线性系统

$$x(k+1) = ax(k) + bu(k) + \omega(k), \quad k = 1, 2, \dots, N \quad (3-52)$$

其中  $a \sim N(0.8, 0.0001)$ ， $b \sim N(0.5, 0.0001)$ ， $N = 35$ 。代价函数中的参数设置为  $Q = 2.5$  和  $R = 1$ 。系统噪声为高斯白噪声  $\omega(k) \sim N(0, 0.01)$ 。

系统状态初值设置为  $x(1) = 5$ ，核矩阵向量初值设置为  $H_v(:,1) = [0.1; 0.1; 0.1]$ ，参数估计误差协方差初值为  $P^{H_v}(:,1) = I_{3 \times 3}$ 。基于确定性等价原理的 Q-学习控制方法的激励信号设置为方差为 0.0001 的随机噪声。这里所提出的主动调节探索信号的参数设置为  $\alpha = 0.15$ 。

基于以上参数设置，本实验对比了三种控制方法，即所提出的探索与利用相平衡的 Q-学习自适应控制 (QLEE)，基于确定性等价原理的 Q-学习自适应控制 (CE)，以及可作为控制性能比较基准的参数假定已知的最优控制 (OPT)。

如图 3-1 所示，首先分析系统控制的起步阶段的瞬时性能，由图可知本章所提出的探索与利用的相平衡的控制方法在系统起步阶段所带来的超调远远小于确定性等价原理下的控制方案，并且状态  $x(k)$  随后调节收敛到 0，这表明本章所提出的具有主动调节性质的探索信号更有利于提高系统在过渡时刻的瞬态控制性能。

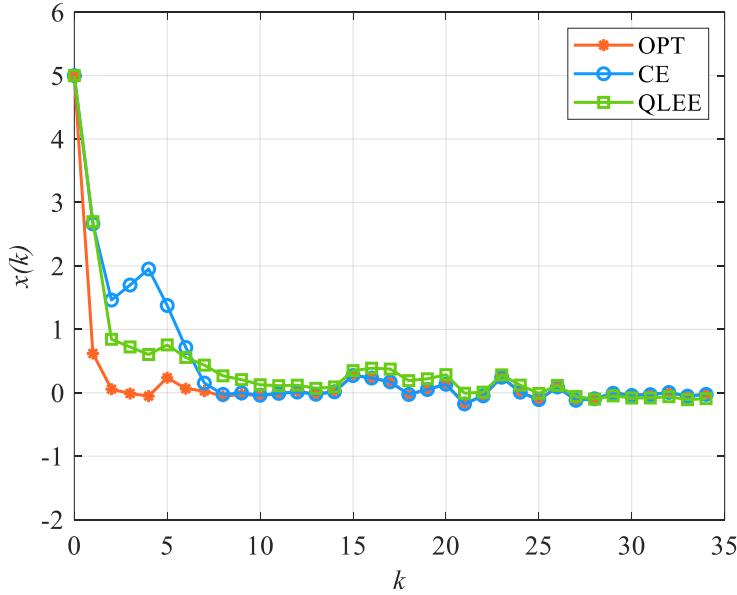


图 3-1 理想最优控制 (OPT)，基于确定性等价原理的 Q-学习 (CE)，以及探索与利用平衡的 Q-学习 (QLEE) 下的系统状态  $x(k)$  轨迹

Fig.3-1 The system state  $x(k)$  for the ideal benchmark optimal control, the CE-based Q-learning, and the exploration and exploitation balanced Q-learning.

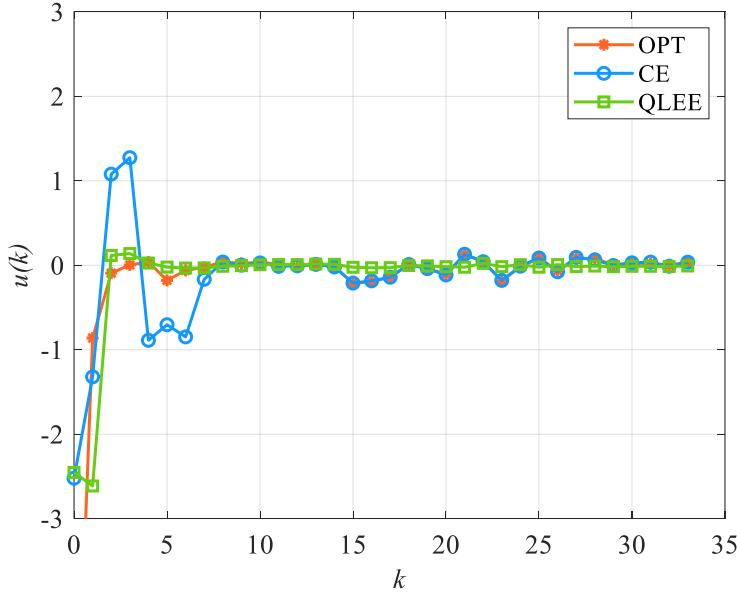


图 3-2 理想最优控制 (OPT)，基于确定性等价原理的 Q-学习 (CE)，以及探索与利用平衡的 Q-学习 (QLEE) 下的系统控制信号  $u(k)$

Fig.3-2 The control signal  $u(k)$  for the ideal benchmark optimal control, the CE-based Q-learning, and the exploration and exploitation balanced Q-learning

图 3-2 是在不同方法下生成的控制信号，从图中可以观察到在本章所提出的探索与利用的相平衡的控制方法下生成的控制信号是相对比较稳定的，没有大的波动。相比之下，确定性等价原理下的 Q-学习方法在起步阶段产生具有较大波动的控制信号，较大的控制信号会给系统带来较大的超调，表明系统做出过大的探索，也就是探索与利用之间

没有做好平衡关系。

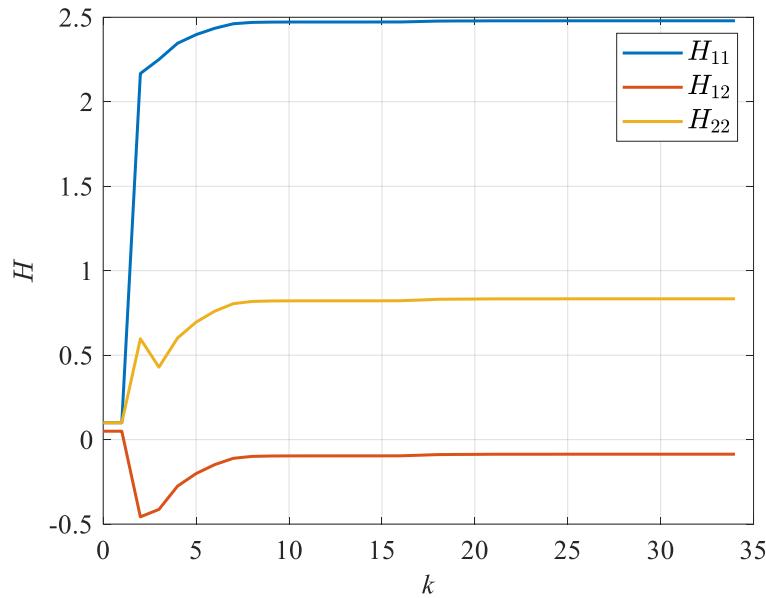


图 3-3 核矩阵  $H$  的学习收敛过程  
Fig.3-3 The convergence of the parameters of kernel matrix  $H$

图 3-3 展示了提出的 Q-学习算法的核心矩阵  $H$  的收敛过程，其中  $H_{11}$ 、 $H_{12}$  和  $H_{22}$  分别为核矩阵  $H$  中的元素。该方法在系统启动后的 6 步内收敛，表明该算法学习参数的快速性和有效性。

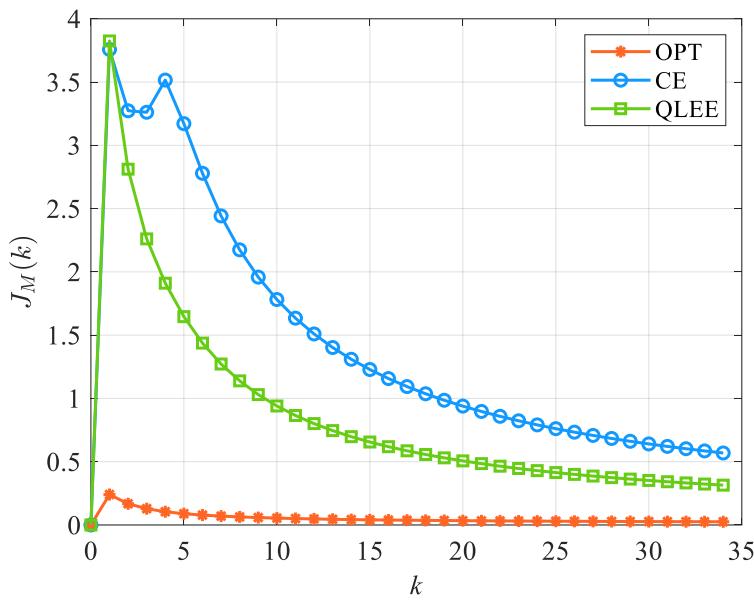


图 3-4 100 次蒙特卡洛仿真实验的平均性能指标  $J_M(k)$   
Fig.3-4 The average performance index  $J_M(k)$  for 100 Monte Carlo simulation results

以上的实验是单次仿真实验的结果，下面通过统计实验来对所提出的方法进行分析。统计实验使用蒙特卡洛仿真进行控制性能的分析，首先定义单次蒙特卡洛仿真的控制性

能指标为

$$J(k) = \frac{1}{k} \sum_{i=1}^k x(i)^T x(i) \quad (3-53)$$

其中  $k$  是仿真步数。通过  $M$  次蒙特卡洛仿真实验，其平均性能指标定义为

$$J_M(k) = \frac{1}{M} \sum_{i=1}^M J_k(i) \quad (3-54)$$

其中  $J_k(i)$  是第  $i$  次蒙特卡洛仿真是  $k$  步控制性能。

图 3-4 是 100 次蒙特卡洛仿真实验的结果，图中曲线记录了在不同控制策略下每一步的平均控制性能指标。由图可知，与基于确定性等价原理的 Q-学习方法相比，本章所提出的具有主动调整特性的探索信号对系统控制性能有了显著的改善，并且能够更快地收敛到理想状态。

表 3-1 理想最优控制，基于确定性等价原理的 Q-学习，以及探索与利用平衡的 Q-学习下的蒙特卡洛仿真实验的平均性能指标

Tab.3-1 Comparison of the average performance index  $J_M$  (35) for the ideal benchmark optimal control, the CE-based Q-learning, and the exploration and exploitation balanced Q-learning

最优控制	确定性等价原理 Q-学习	探索与利用平 衡的Q-学习
平均性能指标	0.0248	0.5575

表 3-1 总结了蒙特卡洛仿真实验的控制性能指标。由该表中的结果可知，本章所提出的控制策略比确定性等价原理的控制策略更接近最优控制的性能。

### 3.4.2 二阶线性系统

本实验的被控对象为一个参数未知的二阶线性系统

$$x(k+1) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} x(k) + \begin{bmatrix} b_{11} \\ b_{12} \end{bmatrix} u(k) + \begin{bmatrix} c_{11} \\ c_{12} \end{bmatrix} \omega(k), \quad k = 1, 2, \dots, N \quad (3-55)$$

其中  $a_{11} \sim N(0, 0.0001)$ ,  $a_{12} \sim N(1, 0.0001)$ ,  $a_{21} \sim N(-0.16, 0.0001)$ ,  $a_{22} \sim N(-1, 0.0001)$ ,  $b_{11} \sim N(0, 0.0001)$ ,  $b_{12} \sim N(1, 0.0001)$ ,  $c_{11} = 0$ ,  $c_{12} = 1$ , 系统运行步数为  $N = 35$ , 代价函数中的参数设置为  $Q = 1$ ,  $R = 1$ 。系统噪声为  $\omega(k) \sim N(0, 0.0025)$ 。

系统初值设置为  $x(:,1) = [10; -10]$ , 核心矩阵向量的初值设置为  $H_v(:,1) = [0.1; 0.1; 0.1; 0.1; 0.1]$ , 估计误差协方差矩阵的初值设置为  $P^{H_v} = 100I_{6 \times 6}$ 。基于确定性等价原理的 Q-学习控制方法的探索信号设置为方差为 0.01 的随机噪声。所提出的主动调整的探索信号的参数设置为  $\alpha = 0.15$ 。本实验对式 (3-55) 中的系统模型进行数值仿真，将本章所提出的探索与利用相平衡的 Q-学习自适应控制，基于确定性等价原理的 Q-学习自适应控

制，以及参数假定已知的最优控制进行对比。

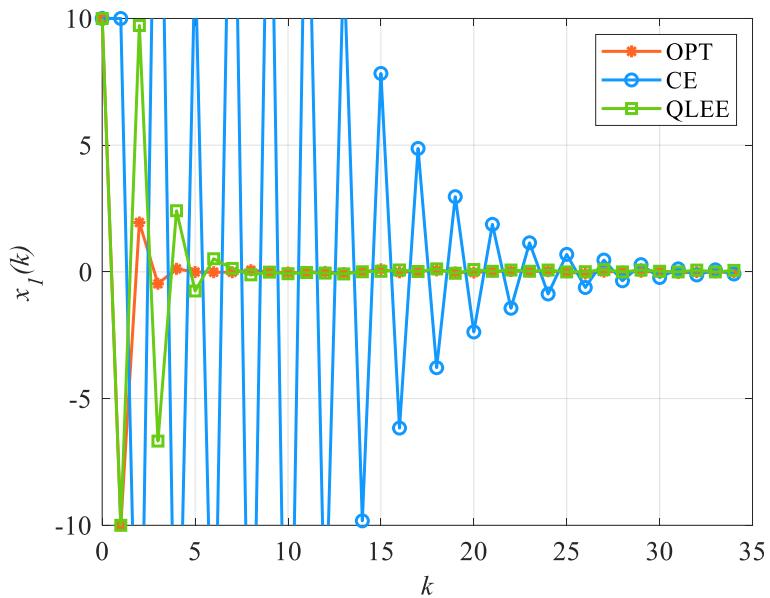


图 3-5 理想最优控制 (OPT)，基于确定性等价原理的 Q-学习 (CE)，以及探索与利用平衡的 Q-学习 (QLEE) 下的系统状态  $x_1(k)$  轨迹

Fig.3-5 The system state  $x_1(k)$  for the ideal benchmark optimal control, the CE-based Q-learning, and the exploration and exploitation balanced Q-learning

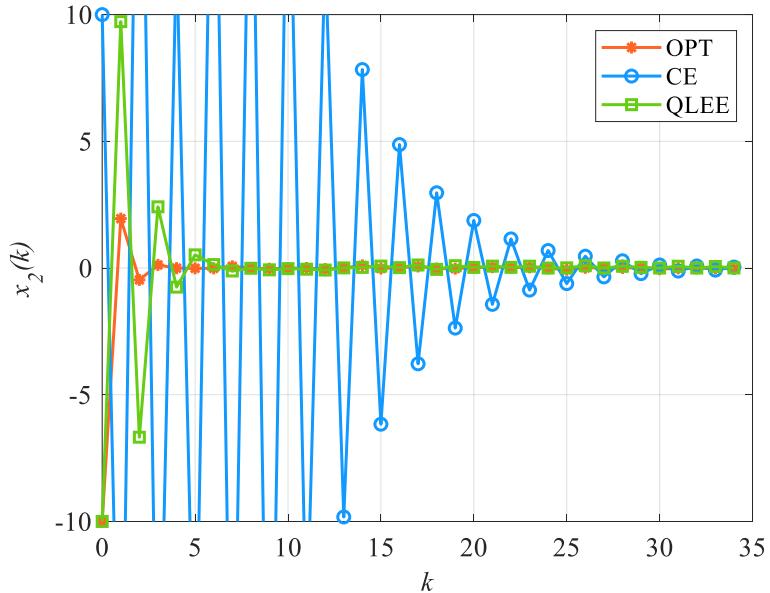


图 3-6 理想最优控制 (OPT)，基于确定性等价原理的 Q-学习 (CE)，以及探索与利用平衡的 Q-学习 (QLEE) 下的系统状态  $x_2(k)$  轨迹

Fig.3-6 The system state  $x_2(k)$  for the ideal benchmark optimal control, the CE-based Q-learning, and the exploration and exploitation balanced Q-learning

图 3-5 和图 3-6 分别给出了状态  $x_1(k)$  和  $x_2(k)$  的调节轨迹。与基于确定性等价原理的 Q-学习相比，本章所提出的探索与利用相平衡的 Q-学习方法下，系统状态  $x_1(k)$  和  $x_2(k)$  轨迹在起步阶段都显示出更少的超调，并且收敛的时间更短。随着时间的推移，状态  $x_1(k)$  和  $x_2(k)$  均调节到 0。

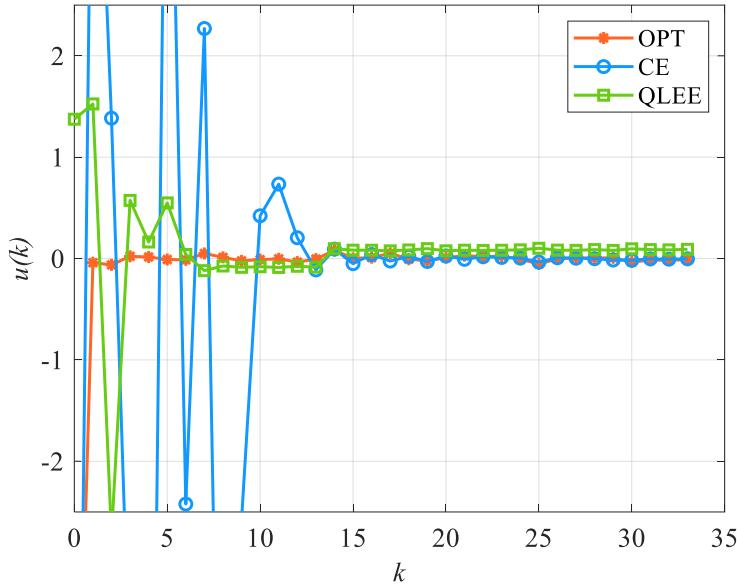


图 3-7 理想最优控制 (OPT)，基于确定性等价原理的 Q-学习 (CE)，以及探索与利用平衡的 Q-学习 (QLEE) 下的系统控制信号  $u(k)$

Fig.3-7 The control signal  $u(k)$  for the ideal benchmark optimal control, the CE-based Q-learning, and the exploration and exploitation balanced Q-learning

图 3-7 是不同控制策略下的控制信号曲线。从图中可以观察到，与基于确定性等价原理的 Q-学习自适应控制相比，本章提出的具有主动调整功能的探索信号提供了更为平稳的控制信号。实验结果表明随着时间的推移，该探索信号能在系统的探索和利用之间实现了有效的平衡。

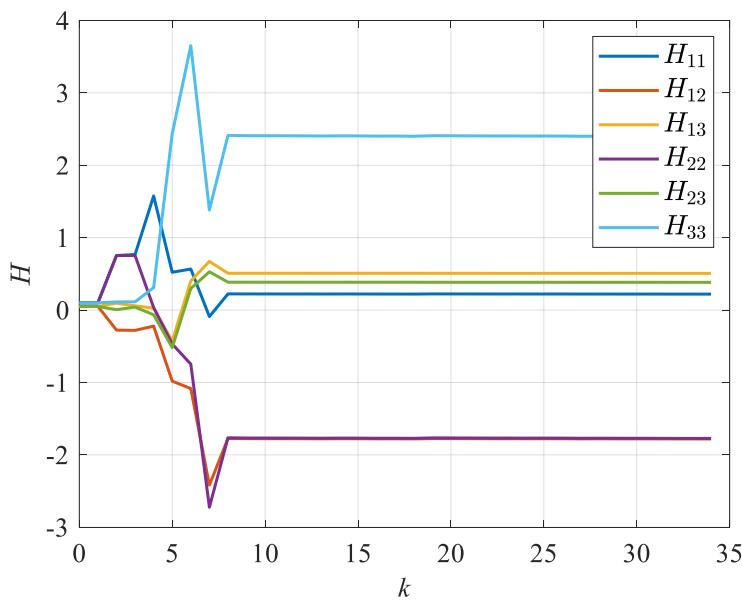


图 3-8 核矩阵参数  $H$  的学习收敛过程  
Fig.3-8 The convergence of the parameters of kernel matrix  $H$

图 3-8 是核心矩阵  $H$  的收敛过程，矩阵  $H$  在 7 步之内就快速收敛。下面对该系统进行 100 次蒙特卡洛仿真实验以分析其统计结果。

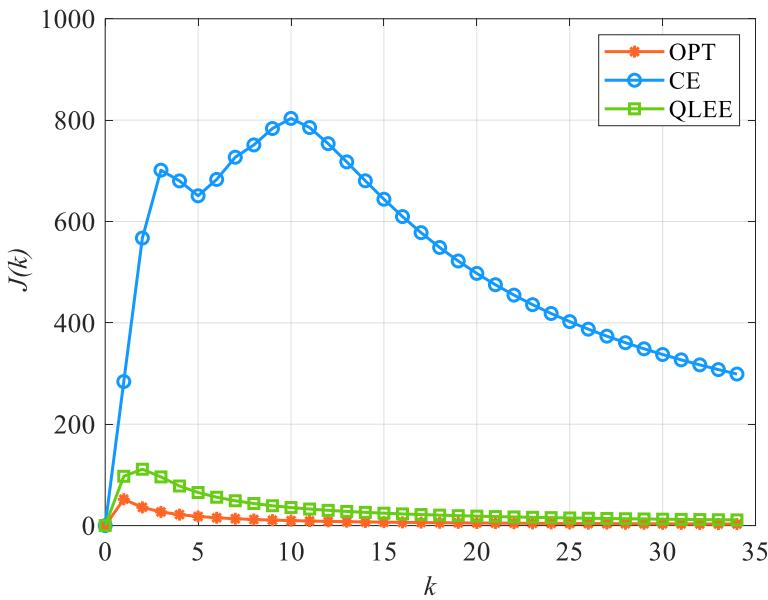
图 3-9 100 次蒙特卡洛仿真实验的平均性能指标  $J_M(k)$ Fig.3-9 The average performance index  $J_M(k)$  for 100 Monte Carlo simulation results

图 3-9 是 100 次蒙特卡洛实验的平均性能指标图。与基于确定性等价原理的 Q-学习自适应控制相比，本章所提出的方法在性能指标上有显著的改进，随着时间的推移，平均控制性能指标  $J_M(k)$  更接近最优控制。

表 3-2 理想最优控制，基于确定性等价原理的 Q-学习，以及探索与利用平衡的 Q-学习下的蒙特卡洛仿真实验的平均性能指标

Tab.3-2 Comparison of the average performance index  $J_M(35)$  for the ideal benchmark optimal control, the CE-based Q-learning, and our proposed algorithm

	最优控制	确定性等价原理 Q-学习	探索与利用平衡 的Q-学习
平均性能指标	3.1019	300.6339	11.3303

表 3-2 是 35 步内的平均控制性能指标，由表可得，本章提出的控制策略比确定性等价原理的控制策略更接近最优控制的性能。

### 3.5 本章小结

本章针对参数未知的线性系统，提出了一种基于 Q-学习的探索与利用相平衡的线性二次型调节器。与现有的基于 Q-学习的确定性等价原理的控制方法不同，该方法设计控制律时考虑了 Q 函数估计误差，并通过同时优化系统状态跟踪性能指标和 Q 函数核矩阵

估计性能指标得以实现。该方法设计出的控制律由两部分组成，一部分是谨慎控制律，能够让系统状态谨慎跟踪目标，另一部分是可主动调整的探索信号，能够激励系统获得更为丰富的信息来进行未知部分的探索，减少了探索学习过程中超调过大或者探索不够充分的情况，即对 Q-学习中探索与利用的冲突问题进行最优平衡，进而提升了控制性能。最后将所提出的具有探索与利用平衡特性的 Q-学习控制方法分别对一阶线性系统和二阶线性系统进行仿真实验，并在同一实验条件下将该方法与确定性等价原理的 Q-学习方法以及参数已知的最优控制方法进行对比。仿真结果表明，本章所提出的具有探索与利用平衡特性的 Q-学习控制方法与确定性等价原理的 Q-学习控制方法相比，减少了系统在探索阶段的超调，其探索阶段的瞬时性能指标以及统计意义上的性能指标更接近最优控制，这意味着该方法在系统状态调节和系统不确定性探索之间取得了良好的平衡，从而取得了更好的控制性能。

## 4 具有孤立点噪声的随机系统自适应对偶控制

### 4.1 引言

在实际随机系统的控制过程中，由传感器实时测量的系统观测值通常会受到大量随机噪声的污染，而这些噪声通常不是现有研究中经常假设的高斯白噪声。这些随机噪声中对控制系统最具伤害性的是孤立点。在系统观测值中，孤立点表现为某个观测值偏离正常观测值非常大，并且这种情况在整个观测过程中出现的频率比较低<sup>[107,108]</sup>。在实际的控制系统中，观测值中的孤立点可能源于传感器失灵、数据传输错误、恶意攻击、或者幅值较大的观测噪声等<sup>[109,110]</sup>。这些存在于观测值中的孤立点和正常观测值比幅值过于大，会对系统的实时参数估计带来恶劣影响，导致系统控制性能的严重下降。另外，孤立点也可能存在于系统本身的过程噪声中，这可能源于系统设备的失灵，恶意进攻等等<sup>[111]</sup>。本章针对被孤立点污染的参数未知随机系统进行自适应对偶控制方法研究。

目前已经存在多种孤立点检测的方法，例如：基于分布的方法<sup>[112]</sup>，基于深度的方法<sup>[113]</sup>，基于距离的方法<sup>[114]</sup>，基于密度的方法<sup>[115]</sup>，基于聚类的方法<sup>[116,117]</sup>等。然而这些方法需要整个样本集，且只能应用于静态数据，并不适用于在自适应控制中需要对观测数据进行实时处理的情况。也有一些方法能够实时检测孤立点，例如迭代聚类的方法可以自适应的实时检测出孤立点，但是该方法只适用于输入输出空间数据，不适用于具有时序特征的数据<sup>[118-120]</sup>；高斯求和估计器可以使自适应控制不受观测值孤立点的影响，但是该方法假设系统的观测噪声服从混合高斯分布，然而实际的场景中的噪声往往不一定具有混合高斯分布的特点<sup>[121,122]</sup>；基于贝叶斯的检测方法可以实时检测出时序数据中的孤立点，但是该方法必须提前知道数据的具体分布<sup>[123,124]</sup>。Ma 等人提出了一种基于概率包络约束的滤波器，可以实时检测时变系统的测量值中的孤立点，但是这个方法假设系统模型参数是已知的，并不适用于实际情况中参数未知的系统<sup>[125]</sup>。鲁棒估计器也常用于具有孤立点的系统的参数估计中，但是该方法目前仅适用于目标模型为高斯马尔可夫模型的情况<sup>[126,127]</sup>。Angelo 等人提出一种移动窗口估计器可以实时处理观测值中的孤立点并对系统未知参数进行估计，但是这种方法仅限于在滑动窗口中只有一个孤立点的情况下，当窗口中包含多个孤立点时并不适用<sup>[107,108]</sup>。Seckin 等人提出了一种移动均值的方法去减小孤立点对模型参数估计质量的负面影响，该方法不需要在进行参数估计时剔除观测值中的孤立点<sup>[128]</sup>。

实际系统往往都是未知系统，自适应控制方法能够一边学习系统未知参数，一边进行系统输出跟踪控制。自适应对偶控制由 Feldbaum 首次提出，该方法能够主动探索系统未知信息，在提高系统参数的辨识精度的同时，还能够让系统输出谨慎的跟踪期望轨迹，以减少系统不确定性。本章着眼于带有孤立点噪声的未知系统控制，目标是降低孤立点对未知系统的学习和控制性能的影响。

本章依托于自适应对偶控制思想，针对参数未知的线性系统，提出一种新的自适应对偶控制方法，该方法能够在随机系统在遭受孤立点污染下仍然能够保证系统的控制性能。为了达到这个目标，设计了一个具有在线检测孤立点的自适应对偶控制方法。该方法提出了一个实时孤立点检测机制，可以在控制过程中实时检测出数据中的孤立点，并且不需要任何输入输出数据的先验知识。该检测机制包含两个检测准则：期望的距离边界和期望的方向边界。如果被检测的数据既不在期望的距离界内，也不再期望的方向界内，那么这个数据就会被标记为孤立点。另外在实际的过程控制中，系统中通常存在一些不可控的激励信号，同时这些不可控的变量可以用传感器进行实时测量，这些实时测量数据也会受到孤立点噪声的污染。该方法将这些不可控的激励信号也加入到经典的自适应对偶控制的框架之中。为了验证方法的有效性，本章实验部分不仅对数学模型生成的数值进行仿真实验，还用实际生物发酵连续灭菌过程中测量到的数据进行仿真实验，并且将该方法和具有移动平均估计器的控制方法<sup>[107,108]</sup>、具有鲁棒移动窗口估计器的控制<sup>[110]</sup>方法做了对比实验，进一步说明该方法的有效性。

## 4.2 问题描述

传统的自适应对偶控制是对具有未知参数和随机噪声的不确定系统进行控制。考虑如下离散时间、单输入单输出的参数未知线性系统

$$\begin{aligned} y(k+1) = & b_1 u(k) + b_2 u(k-1) + \cdots + b_n u(k-m+1) + a_1 y(k) + a_2 y(k-1) + \cdots \\ & + a_m y(k-n+1) + \omega(k), \quad k = 0, 1, \dots, M-1 \end{aligned} \quad (4-1)$$

其中  $u(k)$  是控制输入，  $y(k)$  是系统输出，  $\{b_1, \dots, b_m, a_1, \dots, a_n\}$  是未知定常参数，假设阶数  $m$  和  $n$  已知，  $\omega(k)$  是高斯白噪声。

以上差分方程模型是传统的自适应对偶控制的研究对象，该模型中仅包含系统的输入信号  $u(k)$  和系统输出信号  $y(k)$ ，而在实际控制系统中往往还存在着一类不可控制的激励信号，例如在生物发酵连续灭菌系统中，原始培养基物料的投入流动速度、未灭菌时原始物料的温度、以及蒸汽温度的变化都无法精准的控制，但是这些变量都会影响灭菌的效果，对此本章设计了一种新的差分方程模型，该模型中包含了不可控的激励信号

$$\begin{aligned} y(k+1) = & b_1 u(k) + b_2 u(k-1) + \cdots + b_n u(k-m+1) + a_1 y(k) + a_2 y(k-1) + \cdots \\ & + a_m y(k-n+1) + c_1 x_1(k) + c_2 x_2(k) \cdots + c_l x_l(k) + \omega(k), \quad k = 1, 2, \dots, M \end{aligned} \quad (4-2)$$

其中  $\{x_1(k), x_2(k), \dots, x_l(k)\}$  是不可控的激励信号，  $\omega(k)$  为过程噪声。考虑到系统具有观测噪声，则观测值可表达为

$$Z(k) = [y(k), x_1(k), x_2(k), \dots, x_l(k)] + v(k) \quad (4-3)$$

其中  $Z(k)$  为观测向量，向量  $v(k)$  为观测噪声。

在以往的研究中，系统的过程噪声  $\omega(k)$  和观测噪声  $v(k)$  均被假设为高斯白噪声，然

而在实际系统中，由于传感器失灵，设备失灵，数据传输失误，恶意数据进攻等原因，会出现的一种频率低幅值高的噪声，即孤立点。因此将孤立点考虑到噪声建模当中，相应的表达式如下

$$\omega(k) = G_\omega(k) + O_\omega(k) \quad (4-4)$$

$$v(k) = G_v(k) + O_v(k) \quad (4-5)$$

其中  $G_\omega(k)$  和  $G_v(k)$  是高斯白噪声，分别服从分布  $G_\omega(k) \sim N(0, R_w)$  和  $G_v(k) \sim N(0, R_v)$ ；  
 $O_\omega(k)$  和  $O_v(k)$  是孤立点且其具体的分布未知。

针对以上问题，这里将设计一个实时检测观测值  $Z(k)$  中孤立点的算法。一旦检测出  $Z(k)$  中的孤立点，就用滤波后的数据  $Z^f(k) = \{f^f(k), x_1^f(k), x_2^f(k), \dots, x_l^f(k)\}$  去替代孤立点，进而用于自适应对偶控制中参数估计和对偶控制律的更新。为便于后续基于双准则对偶控制（Bi-criterial Dual Control, BDC）的求解，这里将未知系统（3-2）简写为如下形式

$$y(k+1) = \phi^T(k)\theta(k) + \omega(k) \quad (4-6)$$

式中  $\phi(k)$  为截止  $k$  时刻的数据向量

$$\phi^T(k) = [u(k), \dots, u(k-m+1), y(k), \dots, y(k-n+1), x_1(k), \dots, x_l(k)] \quad (4-7)$$

参数向量  $\theta(k)$  包含了截止  $k$  时刻的参数

$$\theta^T(k) = [b_1(k), \dots, b_m(k), a_1(k), \dots, a_n(k), c_1(k), \dots, c_l(k)] \quad (4-8)$$

双准则对偶控制是一种次优的对偶控制策略，该方法通过同时优化两个相互冲突的性能指标函数得到：一个是参数估计质量指标函数，另一个是输出跟踪控制性能指标函数。

衡量控制性能的代价函数可以表达为

$$J_c = E \left\{ [y(k+1) - y_r(k+1)]^2 \mid \mathfrak{I}_k \right\} \quad (4-9)$$

其中  $y_r(k+1)$  是期望系统输出， $\mathfrak{I}_k$  是信息状态，在所提出的控制方法中，信息状态  $\mathfrak{I}_k$  中包含的观测值是已经被滤除孤立点之后的数据  $\mathfrak{I}_k = \{u(1), \dots, u(k-1), Z^f(1), \dots, Z^f(k)\}$ 。该代价函数描述了系统的跟踪误差。

衡量参数估计质量的代价函数可以写成如下形式

$$J_a = E \left\{ [y(k+1) - \hat{y}(k+1)]^2 \mid \mathfrak{I}_k \right\} \quad (4-10)$$

其中  $\hat{y}(k+1)$  是基于当前估计参数的系统输出的一步超前预测。该代价函数描述了系统的估计误差。能够同时优化这两个代价函数的控制律就是对偶控制律，下一小节给出了详细的控制器设计过程。

### 4.3 控制器设计

本小节详细描述了针对受孤立点噪声污染的随机系统的自适应对偶控制方法的设计。图 4-1 是所提出的具有在线孤立点检测的自适应对偶控制方法的整体结构。

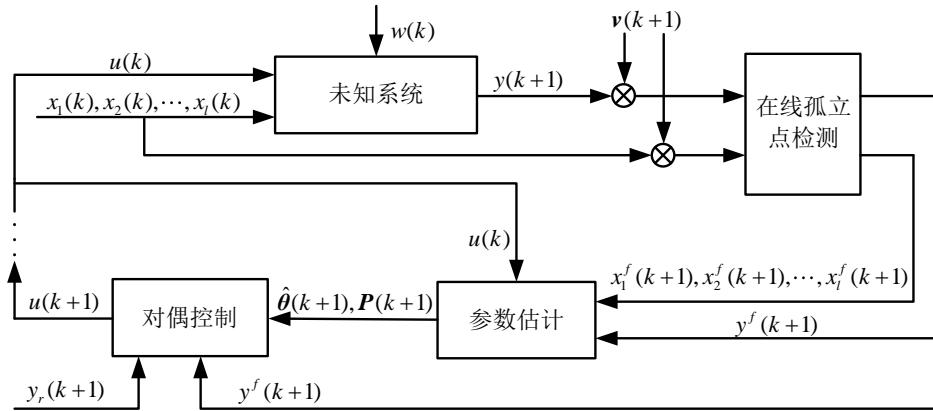


图 4-1 具有在线孤立点检测的自适应对偶控制整体结构框图  
Fig4-1 Overview of adaptive dual control with online outlier detection

在图 4-1 中，未知系统采用式 (4-2) 进行描述，并假设系统观测噪声和过程噪声均会被孤立点污染。在线孤立点检测算法用来实时检测观测值中的孤立点，并将其用期望的值来替换，这样滤波后的观测值不再有孤立点，滤波后的观测值为  $Z^f(k+1) = \{y^f(k+1), x_1(k+1), x_2(k+1), \dots, x_l(k+1)\}$ ，该观测值被用于后续的参数估计。基于 BDC 的自适应对偶控制  $u(k+1)$  是通过同时最优化式 (4-9) 和 (4-10) 两个代价函数求得。接下来分别对在线孤立点检测和基于 BDC 的自适应对偶控制做详细介绍。

#### 4.3.1 在线孤立点检测

本小节详细介绍了所提出的在线孤立点检测算法，该算法能够在参数估计前对观测值进行实时孤立点检测。这就需要设计的孤立点检测算法执行步骤简单且定位孤立点速度快，能够在采样到观测值的瞬间判断出其是否为孤立点。



图 4-2 在线孤立点检测的结构框图  
Fig4-2 The block diagram of online outlier detection

根据式 (4-3)，在第  $i$  步的观测值可以表达成如下向量形式

$$Z(i) = [y(i), x_1(i), x_2(i), \dots, x_l(i)] + v(i) \quad (4-11)$$

定义一组数据，其中包含孤立点数据，孤立点数据随机分布在这组数据中，数据如下

$$\{Z(1), Z(2), \dots, Z(i-1), Z(i), \dots, Z(k), Z(k+1)\} \quad (4-12)$$

如图 4-2 所示，通过设计距离判据和方向判据对  $Z(k+1)$  进行检测，判断当前需要检测的  $Z(k+1)$  是否为孤立点。这两个判据分别是当前观测值  $Z(k+1)$  的期望距离和方向范围，如果  $Z(k+1)$  同时超出这两个范围，那么它就被认为是孤立点。这两个判据的具体描述如下。

距离判据：如果  $\|Z(k+1) - Z(k)\|$  不在区间  $\Omega_s(k+1)$  内，则观测向量  $Z(k+1)$  被判为疑似孤立点，区间  $\Omega_s(k+1)$  就是  $Z(k+1)$  和  $Z(k)$  之间的期望距离区间

$$\Omega_s(k+1) = [E_s(k+1) - \eta_s \sigma_s(k+1), E_s(k+1) + \eta_s \sigma_s(k+1)], \quad (4-13)$$

其中  $\eta_s$  是自定义参数， $\sigma_s(k+1)$  是从第  $(k-N+1)$  步到第  $k$  步近邻数据的欧氏距离的标准差

$$\sigma_s(k+1) = \sqrt{\frac{1}{N} \sum_{i=k-N+1}^k \|Z(i) - Z(i-1)\|^2}, \quad (4-14)$$

$E_s(k+1)$  是  $Z(k+1)$  和  $Z(k)$  之间的期望距离

$$E_s(k+1) = \sum_{i=k-N+1}^k \frac{2i}{N(2k-N+1)} \|Z(i) - Z(i-1)\|, \quad (4-15)$$

其中  $\|Z(i) - Z(i-1)\|$  是  $Z(i)$  和  $Z(i-1)$  之间的欧式距离； $N$  是窗口长度。

方向判据：如果  $\frac{Z(k+1) - Z(k)}{\|Z(k+1) - Z(k)\|}$  在区间  $\Omega_r(k+1)$  之外，则观测向量  $Z(k+1)$  被判为疑似孤立点，区间  $\Omega_r(k+1)$  就是  $Z(k+1)$  和  $Z(k)$  之间的期望方向区间

$$\Omega_r(k+1) = [E_r(k+1) - \eta_r \sigma_r(k+1), E_r(k+1) + \eta_r \sigma_r(k+1)] \quad (4-16)$$

其中  $\eta_r$  是自定义参数， $\sigma_r(k+1)$  是从第  $(k-N+1)$  步到第  $k$  步近邻数据的方向的标准差

$$\sigma_r(k+1) = \sqrt{\frac{1}{N} \sum_{i=k-N+1}^k \frac{Z(i) - Z(i-1)}{\|Z(i) - Z(i-1)\|}} \quad (4-17)$$

$E_r(k+1)$  是在第  $k+1$  步的期望方向

$$E_r(k+1) = \sum_{i=k-N+1}^k \frac{2i}{N(2k-N+1)} \frac{Z(i) - Z(i-1)}{\|Z(i) - Z(i-1)\|} \quad (4-18)$$

注释 1：式 (4-13) 的期望距离区间  $\Omega_s(k+1)$  和式 (4-16) 的期望方向区间  $\Omega_r(k+1)$  都由最新的  $N$  个历史近邻数据  $\{Z(k-N+1), \dots, Z(k-1), Z(k)\}$  进行更新。也就是说期望距离和方向区间是根据最新的  $N$  个历史近邻数据来自动调整的，因而孤立点检测准则是随实时数据自动调整的。

注释 2：在式 (4-15) 和 (4-18) 里的系数  $\frac{2i}{N(2k-N+1)}$  具有如下特征

$$\sum_{i=k-N+1}^k \frac{2i}{N(2k-N+1)} = 1 \quad (4-19)$$

该系数可以在计算期望距离和方向选择数据时，使靠近被检测值的数据的权重更大一些，而距离被检测数据越远的数据的权重更小一些。这就使得期望距离  $E_s(k+1)$  和期望方向  $E_r(k+1)$  更多的依赖于靠近被检测值的数据而不是即将过期的数据，从而使孤立点检测准则中的期望区间更为合理。

注释 3：如果  $Z(k+1)$  的近邻距离  $\|Z(k+1)-Z(k)\|$  和近邻方向  $\frac{Z(k+1)-Z(k)}{\|Z(k+1)-Z(k)\|}$  均不在

期望距离区间  $\Omega_s(k+1)$  和期望方向区间  $\Omega_r(k+1)$  之内，那么  $Z(k+1)$  就被认为是孤立点。孤立点  $Z(k+1)$  将被期望观测值  $Z_E(k+1)$  替换，期望观测值将被用于后续的参数估计和控制律求解。定义期望观测值  $Z_E(k+1)$  为

$$Z_E(k+1) = Z(k) + E_s(k+1)E_r(k+1) \quad (4-20)$$

其中  $E_s(k+1)$  和  $E_r(k+1)$  分别根据式 (4-15) 和式 (4-18) 获得。

注释 4：在式 (4-13) 和式 (4-16) 中的系数  $\eta_s > 0$  和  $\eta_r > 0$  是根据实验经验数据进行选取的。如果选取的  $\eta_s$  和  $\eta_r$  比较大，那么相应的期望区间  $\Omega_s(k+1)$  和  $\Omega_r(k+1)$  也就比较大，近邻距离  $\|Z(k+1)-Z(k)\|$  和近邻方向  $\frac{Z(k+1)-Z(k)}{\|Z(k+1)-Z(k)\|}$  就会有更高的可能性落入期望区间，使孤立点判据不会过于严苛。同理，较小的  $\eta_s$  和  $\eta_r$  会导致较为严苛的孤立点判据，容易引发数据的误判。这里利用历史数据通过网格搜寻的方法找到最优的  $\eta_s$  和  $\eta_r$  数值。

### 4.3.2 具有不可控激励的未知系统的双准则自适应对偶控制

本小节详细介绍了针对具有不可控激励的未知系统的自适应对偶控制。双准则对偶控制方法是由 Filatov 和 Unbehauen 首次提出<sup>[27]</sup>，并且是基于极点配置的控制器。这里将双准则自适应对偶控制方法用于式 (4-6) 所示具有不可控激励的参数未知系统。

如图 4-1 所示，观测值经过孤立点检测和滤波之后的值为  $Z^f(k+1) = \{y^f(k+1), x_1(k+1), x_2(k+1), \dots, x_l(k+1)\}$ ，然后将  $Z^f(k+1)$  用于参数估计，这里参数估计使用了卡尔曼滤波器，具体描述如下。式 (4-6) 中的向量  $\phi^T(k)$  由滤波后的观测值替代，可以写成

$$\phi^T(k) = [u(k), \dots, u(k-m+1), y^f(k), \dots, y^f(k-n+1), x_1^f(k+1), \dots, x_l^f(k+1)] \quad (4-21)$$

基于卡尔曼滤波的参数估计的迭代公式如下所示

$$\hat{\theta}(k+1) = \hat{\theta}(k) + K(k+1)[y(k+1) - \hat{\theta}(k)\phi(k)] \quad (4-22)$$

$$K(k+1) = P(k)\phi(k)[\phi^T(k)P(k)\phi(k) + R_\omega]^{-1} \quad (4-23)$$

$$P(k+1) = P(k) - K(k+1)\phi^T(k)P(k) \quad (4-24)$$

其中估计参数  $\hat{\theta}(k+1)$  和误差协方差矩阵  $P(k+1)$  将用于生成第  $k+1$  步的对偶控制律  $u(k+1)$ 。对偶控制律就是同时使代价函数  $J_c$  和  $-J_a$  取最小值的  $u(k+1)$ 。首先通过最小化代价函数  $J_c$ ，可以得到谨慎控制  $u_c(k) = \arg \min_{u(k)} J_c(k)$

$$u_c(k) = -\frac{[P_{\alpha b_1}^T(k) + \hat{b}_1(k)\alpha^T(k)]\varphi(k) - \hat{b}_1(k)y_r(k+1)}{P_{b_1}(k) + \hat{b}_1^2(k)} \quad (4-25)$$

具体推导过程如下，首先计算代价函数  $J_c(k)$

$$\begin{aligned} J_c &= E\left\{ [y(k+1) - y_r(k+1)]^2 \middle| \mathfrak{I}_k \right\} \\ &= E\left\{ [\phi^T(k)\theta(k) + \omega(k) - y_r(k+1)]^2 \middle| \mathfrak{I}_k \right\} \end{aligned} \quad (4-26)$$

参数估计值  $\hat{\theta}(k)$  是使用状态信息  $\mathfrak{I}_k$  进行参数估计得到，可以表达为  $\hat{\theta}(k) = E\{\theta(k) | \mathfrak{I}_k\}$ ， $\tilde{\theta}(k)$  为参数估计误差，可以表达为  $\tilde{\theta}(k) = E\{\theta(k) - \hat{\theta}(k) | \mathfrak{I}_k\}$ ，那么可以得到关于参数  $\theta(k)$  的等式  $\theta(k) = \tilde{\theta}(k) + \hat{\theta}(k)$ ，并将其代入到式 (4-26) 可得

$$\begin{aligned} J_c(k) &= E\left\{ [\phi^T(k)\tilde{\theta}(k) + \omega(k) + \phi^T(k)\hat{\theta}(k) - y_r(k+1)]^2 \middle| \mathfrak{I}_k \right\} \\ &= \phi^T(k)P(k)\phi(k) + R_\omega + [\phi^T(k)\hat{\theta}(k) - y_r(k+1)]^2 \end{aligned} \quad (4-27)$$

其中误差协方差矩阵为  $P(k) = E\{\tilde{\theta}(k)\tilde{\theta}^T(k) | \mathfrak{I}_k\}$ 。

下面将参数估计值  $\hat{\theta}(k)$ 、向量  $\phi(k)$  和估计误差协方差矩阵  $P(k)$  进行分块处理

$$\hat{\theta}^T(k) = [\hat{b}_1(k) \quad | \quad \hat{\alpha}^T(k)] \quad (4-28)$$

$$\phi^T(k) = [u(k) \quad | \quad \varphi^T(k)] \quad (4-29)$$

$$P(k) = \begin{bmatrix} P_{b_1}(k) & | & P_{b_1\alpha}^T(k) \\ -- & -|- & -- \\ P_{b_1\alpha}(k) & | & P_\alpha(k) \end{bmatrix} \quad (4-30)$$

将式 (4-28)，(4-29) 和 (4-30) 代入式 (4-27)，代价函数  $J_c(k)$  可以写成含有  $u(k)$  的表达式

$$\begin{aligned} J_c(k) &= P_{b_1}(k)u_c^2(k) + 2P_{b_1\alpha}^T(k)\varphi(k)u_c(k) + \varphi^T(k)P_\alpha(k)\varphi(k) + R_\omega \\ &\quad + [\hat{b}_1(k)u_c(k) + \varphi(k)\hat{\alpha}(k) - y_r(k+1)]^2 \\ &= [P_{b_1}(k) + \hat{b}_1^2(k)]u_c^2(k) + 2\{P_{b_1\alpha}^T(k)\varphi(k) + \hat{b}_1(k)[\varphi(k)\hat{\alpha}(k) - y_r(k+1)]\}u_c(k) \\ &\quad + \varphi^T(k)P_\alpha(k)\varphi(k) + R_\omega + [\varphi(k)\hat{\alpha}(k) - y_r(k+1)]^2 \end{aligned} \quad (4-31)$$

$J_c(k)$  关于  $u(k)$  的偏导为

$$\frac{\partial J_c(k)}{\partial u(k)} = 2[P_{b_1}(k) + \hat{b}_1^2(k)]u_c(k) + 2\{P_{b_1\alpha}^T(k)\varphi(k) + \hat{b}_1(k)[\varphi(k)\hat{\alpha}(k) - y_r(k+1)]\} \quad (4-32)$$

令  $\frac{\partial J_c(k)}{\partial u(k)} = 0$  可求得谨慎控制律  $u_c(k)$ , 即得式 (4-25)。

在以谨慎控制律  $u_c(k)$  为区间的区间  $I_k = [u_c(k) - \sigma(k), u_c(k) + \sigma(k)]$  内, 求取使第二个代价函数  $-J_a$  最小的  $u(k)$  就是最终所需的自适应对偶控制律

$$u(k) = u_c(k) + \sigma(k) \text{sign}(\rho) \quad (4-33)$$

其中  $\sigma(k) = \beta \text{trace}\{P(k)\}$ ,  $\beta > 0$ ,  $\rho = P_{b_1}(k)u_c(k) + P_{b_1\alpha}^T(k)\varphi(k)$ , 详细求解过程如下所示。

代价函数  $-J_a$  的计算过程为

$$\begin{aligned} -J_a(k) &= -E\{[y(k+1) - \hat{y}(k+1)]^2 | \mathfrak{I}_k\} \\ &= -E\{[\phi^T(k)\theta(k) + \omega(k) - \hat{\phi}^T(k)\hat{\theta}(k)]^2 | \mathfrak{I}_k\} \\ &= -E\{[\phi^T(k)\tilde{\theta}(k) + \omega(k)]^2 | \mathfrak{I}_k\} \\ &= -[\phi^T(k)P(k)\phi(k) + R_\omega] \end{aligned} \quad (4-34)$$

将式 (4-28), (4-29) 和 (4-30) 代入式 (4-34) 可得

$$-J_a(k) = -\left[ P_{b_1}(k)u^2(k) + 2P_{b_1\alpha}^T(k)\varphi(k)u(k) + \varphi^T(k)P_\alpha(k)\varphi(k) + R_w \right] \quad (4-35)$$

在区间  $I_k = [u_c(k) - \sigma(k), u_c(k) + \sigma(k)]$  内找到能够使  $-J_a(k)$  最小的  $u(k)$ 。由于  $P_{b_1}(k) > 0$ ,  $-J_a(k)$  是一个开口朝下的关于  $u(k)$  的二次函数, 所以在区间  $I_k$  的两端  $u_c(k) - \sigma(k)$  或者  $u_c(k) + \sigma(k)$  取得  $-J_a(k)$  的最小。下面比较取两个端点下的  $-J_a(k)$  的大小

$$\begin{aligned} -J_a[u_c(k) - \sigma(k)] &+ J_a[u_c(k) + \sigma(k)] \\ &= -P_{b_1}(k)[u_c(k) - \sigma(k)]^2 - 2P_{b_1\alpha}^T(k)\varphi(k)[u_c(k) - \sigma(k)] \\ &\quad + P_{b_1}(k)[u_c(k) + \sigma(k)]^2 + 2P_{b_1\alpha}^T(k)\varphi(k)[u_c(k) + \sigma(k)] \\ &= 4P_{b_1}(k)\sigma(k)u_c(k) + 4P_{b_1\alpha}^T(k)\varphi(k)\sigma(k) \\ &= 4\sigma(k)[P_{b_1}(k)u_c(k) + P_{b_1\alpha}^T(k)\varphi(k)] \end{aligned} \quad (4-36)$$

为简化表达, 引入定义

$$\rho = P_{b_1}(k)u_c(k) + P_{b_1\alpha}^T(k)\varphi(k) \quad (4-37)$$

如果  $\rho < 0$ , 表明  $-J_a[u_c(k) - \sigma(k)]$  比  $-J_a[u_c(k) + \sigma(k)]$  小, 那么对偶控制律为  $u_c(k) - \sigma(k)$ 。

如果  $\rho > 0$ , 表明  $-J_a[u_c(k) - \sigma(k)]$  比  $-J_a[u_c(k) + \sigma(k)]$  大, 那么对偶控制律为  $u_c(k) + \sigma(k)$ 。

对偶控制律可以用符号函数  $\text{sign}(x)$  简化表达, 即可得到如式 (4-33) 的简化形式。

**注释 5:** 式 (4-33) 是自适应对偶控制律表达式, 其结构为谨慎控制律  $u_c(k)$  加上一个探测信号  $\sigma(k)$ 。探测信号  $\sigma(k)$  是参数估计误差协方差矩阵的迹, 它可以使得用于参数估计的信息更为丰富, 促进系统主动进行未知参数的学习, 同时也平衡了参数估计与跟踪控制之间的冲突, 使控制性能达到最佳状态。

**注释 6:** 在设计控制律时加入在线孤立点检测算法, 可以得到更为精确的参数估计值, 从而使得控制律更为精确, 最终提高系统控制性能。

注释 7：探测信号  $\sigma(k)$  中的系数  $\beta > 0$ 。当  $\beta = 0$  时，对偶控制律就等于谨慎控制律。当  $\beta$  取值过于大时，探测信号  $\sigma(k)$  就会很大，会给系统在启动阶段带来较大的超调，使系统的瞬态性能指标变差，这里可以用历史数据搜寻最优的  $\beta$  值。

## 4.4 仿真实验

本小节分别用数学模型生成的数据和实际生物发酵连续灭菌过程数据对提出的具有孤立点检测的自适应对偶控制算法的有效性进行验证。在 4.4.1 小节中，用数学模型生成的数据对在线孤立点检测算法进行仿真实验。第 4.4.2 和 4.4.3 小节仍然使用数学模型生成的数据对整个具有孤立点检测算法的自适应对偶控制进行仿真实验。第 4.4.2 小节实验假设孤立点出现在观测噪声之中，而第 4.4.3 小节的实验中假设系统的系统噪声中存在孤立点。4.4.4 小节用实际工厂中生物发酵连续灭菌过程中记录的数据进行试验，并且将所提出的方法与其他方法做了对比实验。

以下为本章所提出的具有在线孤立点检测的自适应对偶控制的具体执行步骤：

步骤 1：初始化参数向量  $\theta(1)$ ，信息状态  $\mathfrak{I}_1$ ，协方差矩阵  $P(1)$ 。设置系数  $\eta_s > 0$ ，  
 $\eta_r > 0$ ， $N \geq 5$  和  $\beta > 0$ ；

步骤 2：当传感器观测到数据  $Z(k+1) = \{y(k+1), x_1(k+1), \dots, x_l(k+1)\}$  时，根据式 (4-13) 计算期望距离区间  $\Omega_s(k+1)$ ；

步骤 3：根据式 (4-16) 计算期望方向区间  $\Omega_r(k+1)$ ；

步骤 4：计算在第  $k+1$  步的近邻欧式距离  $\|Z(k+1) - Z(k)\|$ ，假如第  $k$  步的数据是孤立点，那么应该使用  $\|Z(k+1) - Z_E(k)\|$  计算第  $k+1$  步的近邻距离，其中  $Z_E(k)$  是第  $k$  步的期望数据；

步骤 5：计算在第  $k+1$  步的近邻方向向量  $\frac{Z(k+1) - Z(k)}{\|Z(k+1) - Z(k)\|}$ ，假如第  $k$  步的数据是孤立点，那么在第  $k+1$  步的近邻方向应该使用  $\frac{Z(k+1) - Z_E(k)}{\|Z(k+1) - Z_E(k)\|}$  计算，其中  $Z_E(k)$  是第  $k$  步的期望数据；

步骤 6：如果  $\|Z(k+1) - Z(k)\|$  和  $\frac{Z(k+1) - Z(k)}{\|Z(k+1) - Z(k)\|}$  都分别超出区间  $\Omega_s(k+1)$  和  $\Omega_r(k+1)$ ，那么数据  $Z(k+1)$  就被判定为孤立点，由期望的数据  $Z_E(k+1)$  替代；否则数据  $Z(k+1)$  就被视为正常值，可直接使用；

步骤 7：用式 (4-22) - (4-24) 卡尔曼滤波估计系统未知参数  $\hat{\theta}(k+1)$  以及估计误差协方差矩阵  $P(k+1)$ ；

步骤 8：根据式 (4-33) 计算自适应对偶控制律  $u(k+1)$ ；

步骤 9：将  $u(k+1)$  作为系统的控制输入，并且返回到步骤 2。

#### 4.4.1 在线孤立点检测仿真实验

本小节用数学模型生成的数据来测试在线孤立点检测算法的有效性。设置数据长度为  $M = 100$ 。不可控激励信号  $x_1(k)$  和  $x_2(k)$  由下式生成

$$\begin{aligned}x_1(k) &= 0.1 \sin(0.02\pi k) + 0.2 \\x_2(k) &= 0.25 \cos(0.02\pi k) + 0.35\end{aligned}\quad (4-38)$$

系统输出  $y(k)$  由下式生成

$$y(k+1) = 0.5u(k) - 1.41y(k) + 0.9y(k-1) \quad (4-39)$$

其中控制信号  $u(k) = [y_r(k+1) + 1.41y(k) - 0.9y(k-1)]/0.5$ ,  $y_r(k)$  是 0.1Hz 的方波信号经过传递函数  $1/(s+1)$  滤波后的数据。生成的数据序列为  $Z(k) = \{y(k), x_1(k), x_2(k)\}$ , 分别在第 15, 20, 45, 65 和 78 个数据上加上孤立点噪声, 相应的孤立点设置为

$$O_v(15) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, O_v(20) = \begin{bmatrix} 0 \\ 0 \\ -1.2 \end{bmatrix}, O_v(45) = \begin{bmatrix} 0 \\ -0.6 \\ 0 \end{bmatrix}, O_v(65) = \begin{bmatrix} 0.85 \\ 0 \\ 0 \end{bmatrix}, O_v(78) = \begin{bmatrix} 0.3 \\ 0 \\ 0.4 \end{bmatrix} \quad (4-40)$$

数据中也存在分布为  $G_v(k) \sim N(0, 0.0005)$  的高斯白噪声。

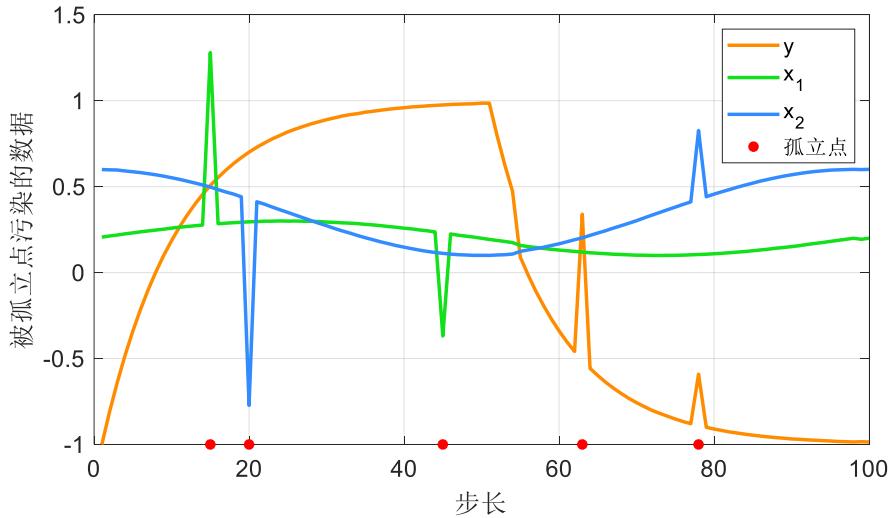


图 4-3 被孤立点污染的一组数据序列  
Fig.4-3 A sequence of data corrupted by outliers

图 4-3 是根据以上描述生成的被孤立点污染的数据序列, 本小节用这组数据验证所提出的在线孤立点检测方法的有效性, 在算法中参数设置为  $\eta_s = 3.5$ ,  $\eta_r = 0.9$ ,  $N = 5$ 。

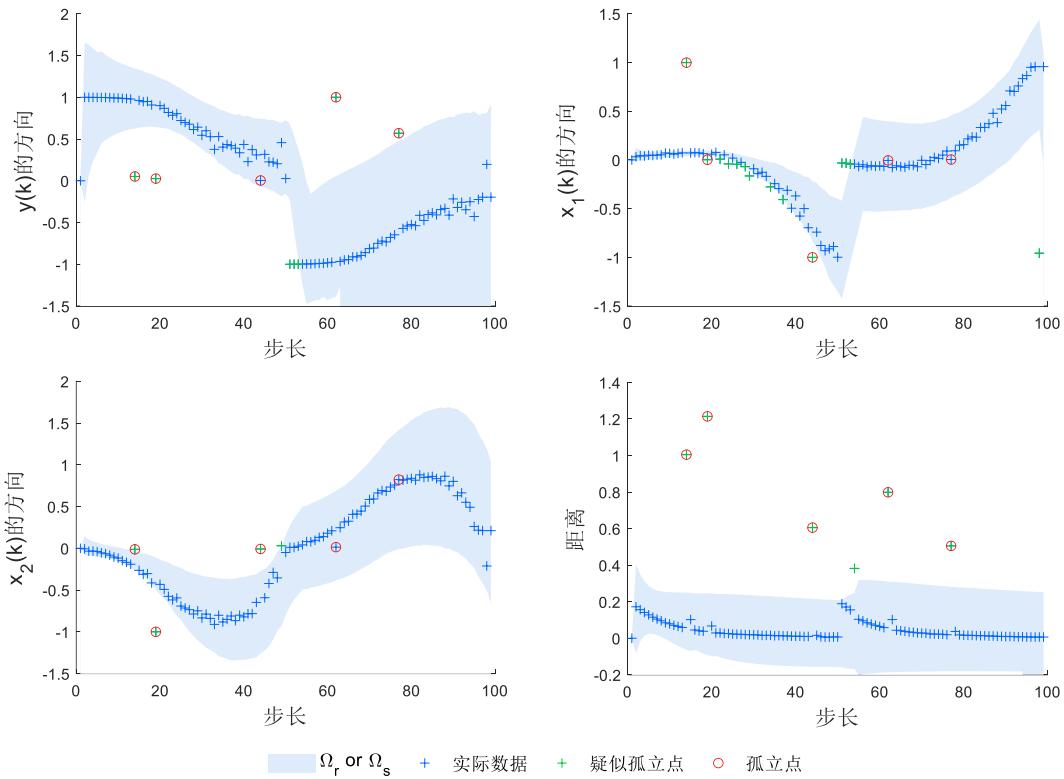


图 4-4 在线孤立点检测过程  
Fig.4-4 The process of online outlier detection

图 4-4 中浅蓝色区域分别代表期望方向区间  $\Omega_r$  和期望距离区间  $\Omega_s$ ，蓝色加号代表实际被孤立点污染过的数据序列所计算的在第  $k$  步的近邻距离  $\|Z(k+1)-Z(k)\|$  和近邻方向  $[Z(k+1)-Z(k)]/\|Z(k+1)-Z(k)\|$ ，绿色加号为疑似孤立点，表明在该点的近邻方向超出了期望方向边界或者近邻距离超出了期望距离。如果绿色加号不仅超出了期望距离边界，也超出了期望方向边界，该点就被判定为孤立点，并用红色圈标出。其中淡蓝色区域代表的期望方向边界和期望距离边界是根据最新的  $N$  个相邻的历史数据实时更新的。可以观察到图中第 55 步的数据  $Z(55)$  与第 54 步的数据  $Z(54)$  之间的近邻距离很大，已经超出了期望的距离边界，然而这两个数据之间的近邻方向很小，而且处于期望的方向边界之中，所以数据  $Z(55)$  不会被判定为孤立点。在第 52 步数据的方向变化很大，实际近邻方向已经超出了期望近邻边界，但是该点的近邻距离很小，并没有超出期望近邻距离边界，那么数据  $Z(52)$  也不会被判定为孤立点。实验中，实际的孤立点  $Z(15)$ ,  $Z(20)$ ,  $Z(45)$ ,  $Z(65)$  和  $Z(78)$ ，均被正确的检测了出来，表明使用所设计的两个判据可以实时检测出数据流中的孤立点。

#### 4.4.2 观测噪声中存在孤立点时的控制仿真实验

本小节验证了当观测噪声  $v(k)$  中存在孤立点时，所提出的控制方案的有效性。实验进行了单次仿真分析和蒙特卡洛统计实验分析，将具有孤立点检测的自适应对偶控制与没有孤立点检测的自适应对偶控制的控制性能进行比较。考虑如式（4-2）所描述的离散时间线性系统，参数设置如下

$$\begin{aligned} a_1 &= -1.41, \quad a_2 = 0.9, \quad n = 2, \\ b_1 &= 0.5, \quad m = 1, \\ c_1 &= 0.2, \quad c_2 = -0.5, \quad l = 2 \end{aligned} \quad (4-40)$$

自适应对偶控制和孤立点检测的系数分别设置为  $\beta = 0.1$ ,  $\eta_s = 4$ ,  $\eta_r = 1.5$ ,  $N = 10$ 。系统参数初值设置为  $\theta(:,1) = \{0.1, 0.1, 0.1, 0.1, 0.1\}$ , 协方差矩阵初值设置为  $P(:, :, 1) = I$ 。

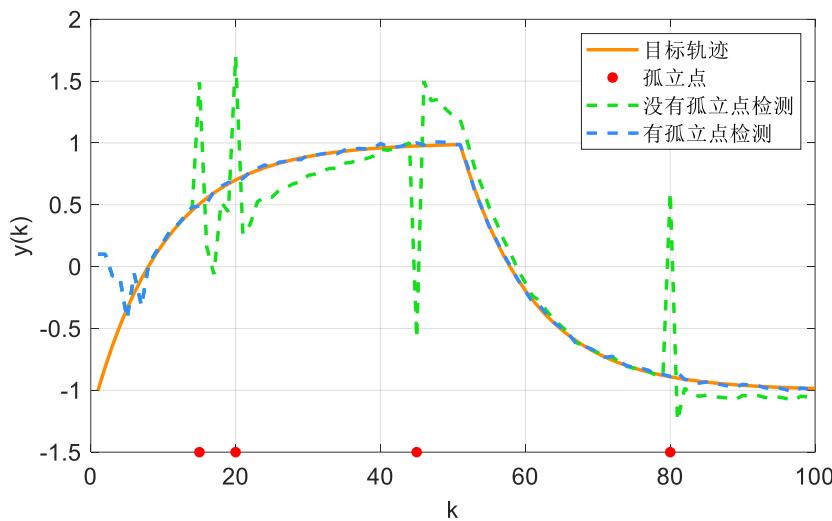


图 4-5 有孤立点检测和没有孤立点检测的自适应对偶控制下系统的输出跟踪  
Fig.4-5 The output of the system under adaptive dual control with or without outlier detection

图 4-5 描述了在单次仿真实验下，系统输出跟踪目标轨迹的效果图。橙色实线代表的是目标轨迹  $y_r(k)$ 。绿色虚线代表的是没有孤立点检测的自适应对偶控制的系统输出。蓝色虚线代表的是有孤立点检测的自适应对偶控制系统输出。在横轴上的红点代表的是孤立点出现的位置。从图中可以看出，和没有孤立点检测的控制输出相比，具有孤立点检测的自适应对偶控制在跟踪控制的过程中不再受孤立点的影响，而没有孤立点检测的自适应对偶控制下的系统输出在孤立点出现的时候出现了较大的尖峰和波动。因此在有孤立点检测的控制策略下的控制性能更好。

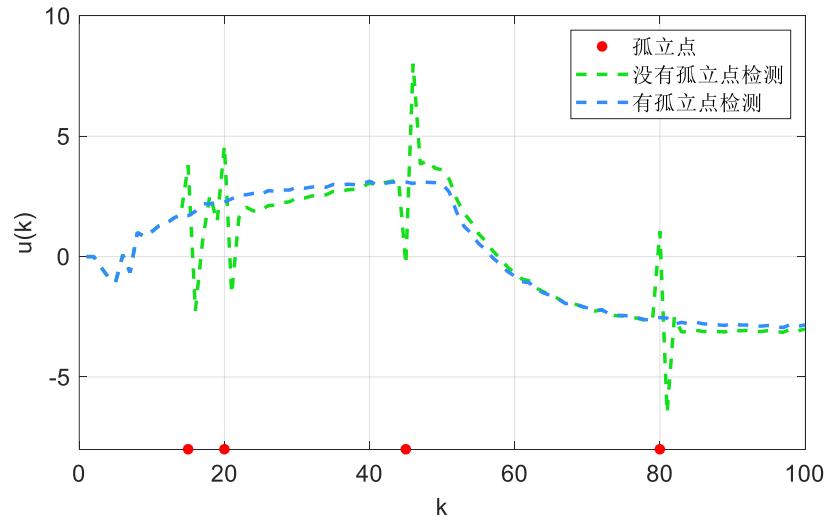


图 4-6 有孤立点检测和没有孤立点检测的自适应对偶控制下的控制信号  
Fig.4-6 The control signal for adaptive dual control with or without outlier detection

图 4-6 显示了具有孤立点检测和不具有孤立点检测的自适应对偶控制的控制信号图，明显可以看到具有孤立点检测的控制信号受孤立点的影响较小甚至几乎没有受到影响，而在没有孤立点检测的控制策略下得到的控制信号出现了较大的波动。

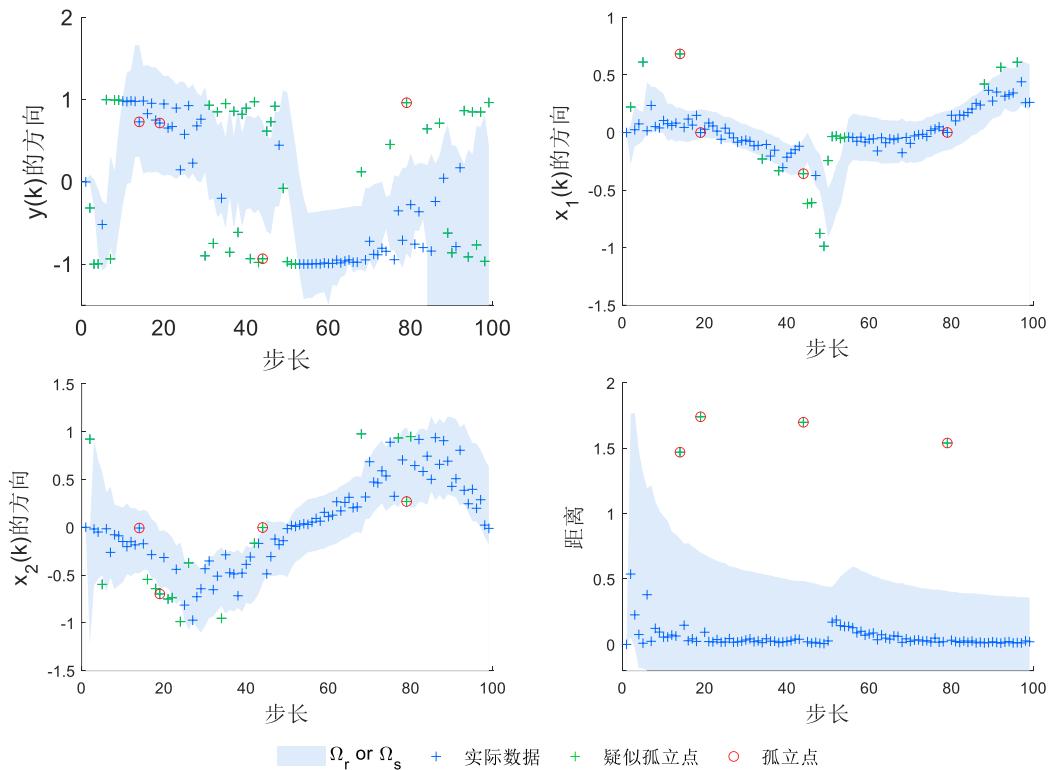


图 4-7 控制过程中的孤立点检测  
Fig.4-7 Online outlier detection during the control process

图 4-7 是在控制中的孤立点检测的过程，其中四个孤立点均能被实时检测出来。

为了量化控制性能指标，定义如下性能指标

$$J = \frac{1}{M} \sum_{k=1}^M [y(k) - y_r(k)]^2 \quad (4-41)$$

其中  $y(k)$  是系统输出， $y_r(k)$  是目标轨迹，执行  $N$  次蒙特卡洛实验，求  $N$  次实验的性能指标的均值

$$\bar{J} = \frac{1}{N} \sum_{i=1}^N J(i) \quad (4-42)$$

表 4-1 是执行 100 次蒙特卡洛实验后有孤立点检测和没有孤立点检测的系统控制的平均性能指标以及运行时间。

表 4-1 有孤立点检测和没有孤立点检测的控制性能和运行时间对比  
Tab.4-1 Comparison between the adaptive dual control with and without online outlier detection

孤立点检测	性能指标均值	运行时间
无	0.0704	0.0088
有	0.0047	0.0254

从表 4-1 中可以看到，具有孤立点检测的自适应对偶控制的性能指标值比没有孤立点检测的控制的性能指标小，说明对应的系统的输出跟踪性能更好。表中还显示了执行 100 次蒙特卡洛实验平均所需的时间，也就是从第 1 步运行到第 100 步的算法执行时间。从表中可以得出，尽管具有孤立点检测的算法相比而言执行时间长，但其执行时间仍然在较低级数范围内，完全满足系统控制的实时性要求。总结来说，这个实验表明当观测值被孤立点污染后，所提出的具有孤立点检测的自适应对偶控制方法能够提供更为稳定和精确的控制效果。

#### 4.4.3 过程噪声中存在孤立点时的控制仿真实验

本小节验证了当系统过程噪声  $\omega(k)$  中存在孤立点时，所提出的控制方案的有效性。实验进行了单次仿真分析和蒙特卡洛统计实验分析，将具有孤立点检测的自适应对偶控制与没有孤立点检测的自适应对偶控制的控制性能进行比较。

同 4.4.2 小节，考虑如式 (4-2) 所描述的离散时间线性系统，参数设置如式 (4-40) 所示，其中不可控激励信号由式 (4-38) 所示的函数生成。观测噪声  $v(k)$  中仅含有高斯白噪声  $G_v(k) \sim N(0, 0.0005)$ 。系统噪声  $\omega(k)$  中不仅含有高斯白噪声  $N(0, 0.01)$ ，且含有孤立点，孤立点出现在第 15, 20, 45 和 80 步，对应的幅值为 {1, 1.2, -1.6, 1.5}。对偶控制和孤立点检测的系数设置为  $\beta = 0.1$ ,  $\eta_s = 4$ ,  $\eta_r = 1.5$ ,  $N = 10$ 。系统参数初值设置为  $\theta(:,1) = \{0.1, 0.1, 0.1, 0.1, 0.1\}$ ，协方差矩阵初值设置为  $P(:, :, 1) = I$ 。

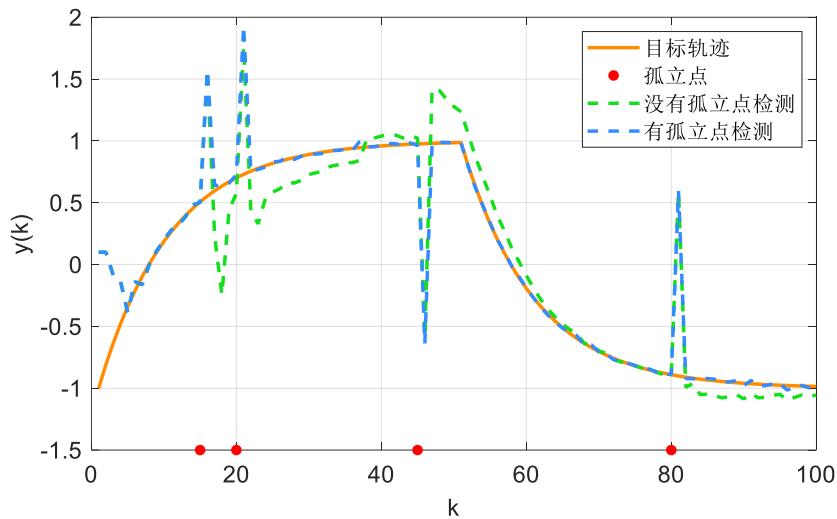


图 4-8 有孤立点检测和没有孤立点检测的自适应对偶控制下系统的输出跟踪  
Fig.4-8 The output of the system under adaptive dual control with or without outlier detection

图 4-8 描述了在单次实验下系统输出跟踪目标轨迹的效果图，其中橙色实线代表的是目标轨迹  $y_r(k)$ ，绿色虚线代表的是没有孤立点检测的自适应对偶控制的系统输出，蓝色虚线代表的是有孤立点检测的自适应对偶控制系统输出，在横轴上的红点代表的是孤立点出现的位置。从图中结果可以看出，和没有孤立点检测的控制输出相比，具有孤立点检测的自适应对偶控制的目标跟踪偏离程度比没有孤立点检测的控制方法小，控制相对更为稳定。

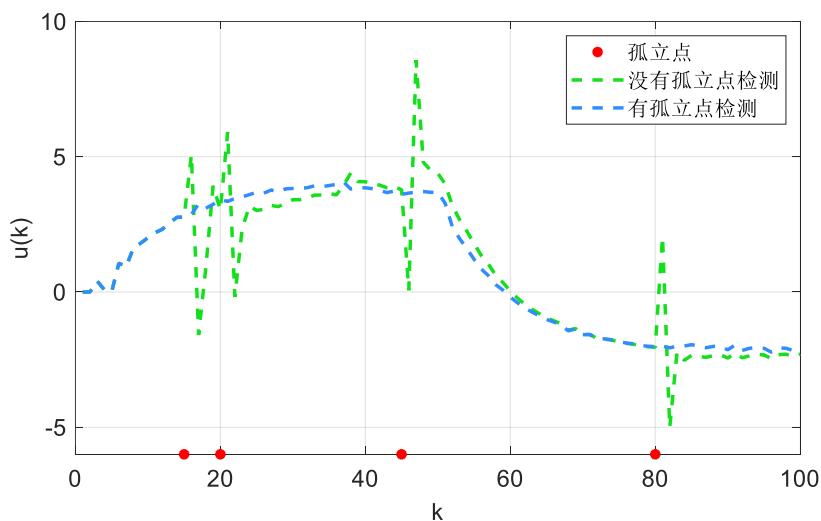


图 4-9 有孤立点检测和没有孤立点检测的自适应对偶控制下的控制信号  
Fig.4-9 The control signal for adaptive dual control with or without outlier detection

图 4-9 是具有孤立点检测和不具有孤立点检测的自适应对偶控制的控制信号图，可以看到具有孤立点检测的控制信号受孤立点的影响较小，几乎没有大的波动。

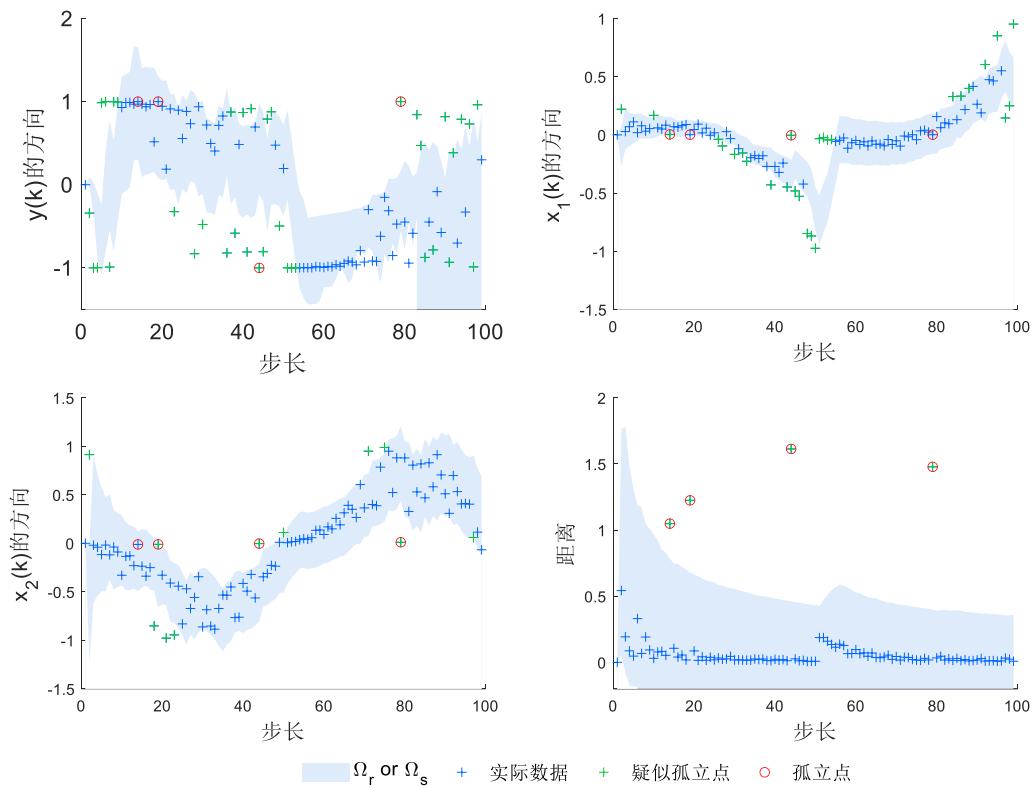


图 4-10 控制过程中的孤立点检测  
Fig.4-10 Online outlier detection during the control process

图 4-10 显示了在系统控制过程中实时孤立点检测的过程，其中四个孤立点均被识别出来。

表 4-2 有孤立点检测和没有孤立点检测的控制性能和运行时间对比  
Tab.4-2 Comparison between the adaptive dual control with and without online outlier

孤立点检测	性能指标均值	运行时间
无	0.0918	0.0071
有	0.0804	0.0240

表 4-2 是执行 100 次蒙特卡洛实验后得到的平均控制性能指标。具有孤立点检测的自适应对偶控制的性能指标值比没有孤立点检测的控制的性能指标小，说明对应的系统的输出跟踪性能更好。从表中可以得出，尽管具有孤立点检测的算法相比而言时间长，仍然完全满足系统控制的实时性要求。总结来说，所提出的具有孤立点检测的自适应对偶控制遭过程噪声中存在的孤立点影响相对较小，结合图 4-9 可以看出该方法生成的控制律对孤立点的出现并不敏感，这就会使系统参数估计更为精确，且得到更好的控制性能。

#### 4.4.4 生物发酵连续灭菌过程的控制仿真

本实验对实际生物发酵培养基连续灭菌过程中采集的数据进行仿真实验，并将所提出的具有孤立点检测的自适应对偶控制方法与其他估计器下的自适应对偶控制的控制性能进行对比，包括具有移动均值滤波的参数估计器<sup>[107,108]</sup>与鲁棒移动窗口参数估计器<sup>[128]</sup>。

在实际生产中，生物发酵培养基中的原始物料里面会存在一些发酵过程所不希望存在的微生物，这些微生物可以通过将培养基加热到一定温度来灭除，一般采用蒸汽喷射使原始物料升高到目标温度从而达到灭菌效果。图 4-11 是生物发酵培养基连续灭菌控制过程框图。

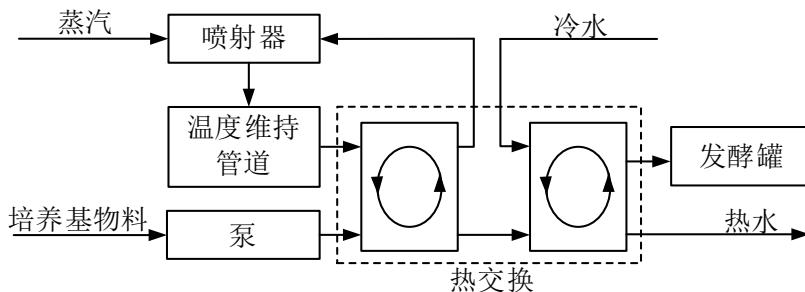


图 4-11 生物发酵培养基连续灭菌控制过程框图  
Fig4-11 The procedure of the fermentation sterilization control process

通过控制蒸汽喷射阀的开度  $O_v$ ，可以调节喷射器喷出的蒸汽的量，使原始物料在热交换中保持在一个期望的灭菌温度中。整个生物发酵培养基连续灭菌控制过程可以用式 (4-2) 来描述，其中不可控的激励信号有原始培养基物料的投入流动速度  $R_m$ ，未灭菌时原始物料的温度  $T_m$ ，蒸汽温度  $T_s$ 。这里的目标物料灭菌温度  $T_{out}$  为 124 摄氏度。物料流速  $R_m$ ，物料投入温度  $T_m$  和蒸汽温度  $T_s$  的取值范围分别为 0.1-26 吨/小时，0-70 摄氏度，150-210 摄氏度。整个生物发酵培养基连续灭菌控制过程表述为如下公式

$$T_{out}(k+1) = b_1 O_v(k) + a_1 T_{out}(k) + a_2 T_{out}(k-1) + c_1 R_m(k) + c_2 T_m(k) + c_3 T_s(k) + \omega(k) \quad (4-43)$$

$$Z(k+1) = [T_{out}(k+1), R_m(k), T_m(k), T_s(k)]^T + v(k+1) \quad (4-44)$$

本实验通过采集实际连续灭菌过程中传感器测量的数据，来估计式 (4-43) 描述的系统的参数值，并将其作为真值，用于自适应控制的仿真实验。生物发酵培养基连续灭菌控制过程中的参数为  $b_1 = 0.0049$ ， $a_1 = 0.9840$ ， $a_2 = 0.0091$ ， $c_1 = 0.0035$ ， $c_2 = 0.0042$ ， $c_3 = 0.0016$ 。图 4-12 是收集到的一组数据，用于测试所提出的控制方案的控制性能。

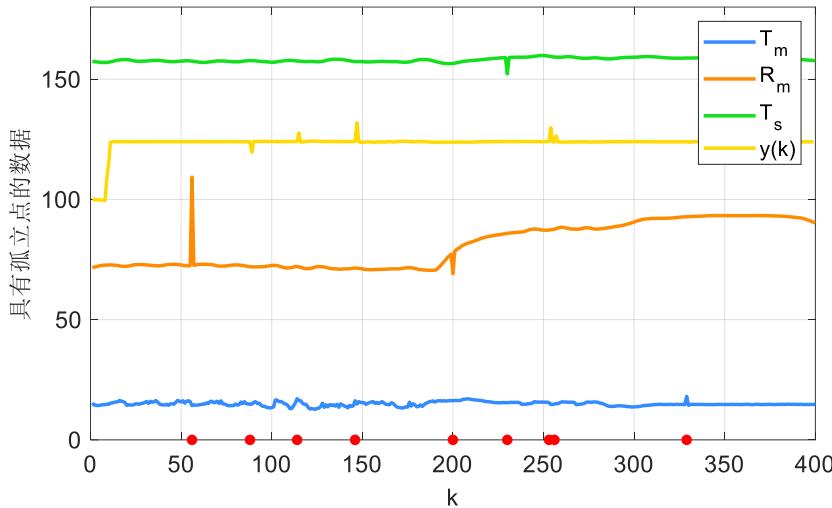


图 4-12 被孤立点污染的生物发酵连续灭菌过程数据

Fig.4-12 The observations of the fermentation sterilization control process corrupted by outliers

图 4-12 中观测数据在第 56, 88, 114, 146, 200, 230, 253, 256, 329 步出现了异常值，这里就被认为是孤立点。孤立点的幅值分别为

$$\begin{aligned}
 O_v(56) &= \begin{bmatrix} 0 \\ 0 \\ 36.8 \\ 0 \end{bmatrix}, & O_v(88) &= \begin{bmatrix} -4.4 \\ 0 \\ 0 \\ 0 \end{bmatrix}, & O_v(114) &= \begin{bmatrix} 3.6 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \\
 O_v(146) &= \begin{bmatrix} 8 \\ 0 \\ 0 \\ 0 \end{bmatrix}, & O_v(200) &= \begin{bmatrix} 0 \\ 0 \\ -8.8 \\ 0 \end{bmatrix}, & O_v(230) &= \begin{bmatrix} 0 \\ 0 \\ 0 \\ -6.8 \end{bmatrix}, \\
 O_v(253) &= \begin{bmatrix} 5.8 \\ 0 \\ 0 \\ 0 \end{bmatrix}, & O_v(256) &= \begin{bmatrix} 2.4 \\ 0 \\ 0 \\ 0 \end{bmatrix}, & O_v(329) &= \begin{bmatrix} 0 \\ 15.8 \\ 0 \\ 0 \end{bmatrix}.
 \end{aligned} \tag{4-45}$$

图 4-13 是在不同的鲁棒参数估计器下的自适应对偶控制的系统输出对比图，其中蓝色虚线是没有孤立点检测的自适应对偶控制的输出，绿色实线是本章新提出的具有在线孤立点检测的自适应对偶控制系统输出，黄色虚线是具有移动均值滤波的参数估计器的自适应对偶控制系统输出，紫红色虚线是具有鲁棒移动窗口参数估计器的自适应对偶控制系统输出。

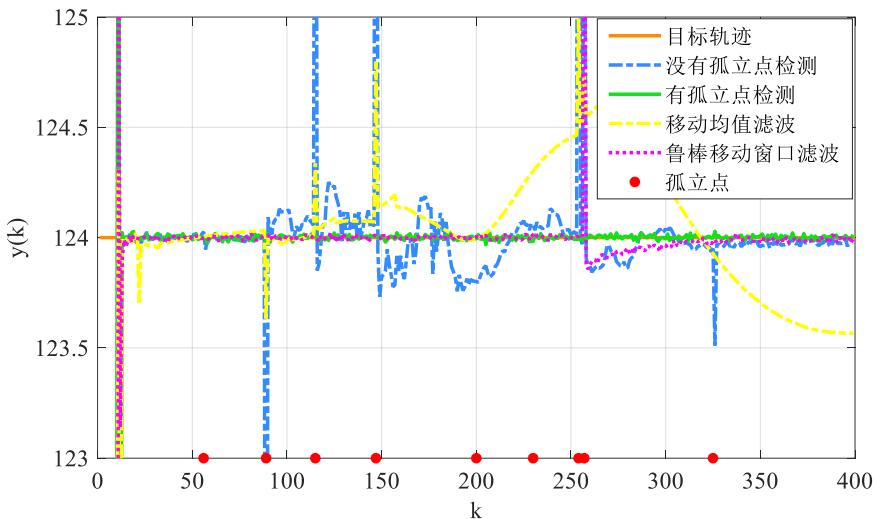


图 4-13 具有不同鲁棒参数估计器的自适应对偶控制下生物发酵连续灭菌过程中温度输出  
Fig.4-13 The output of the fermentation sterilization control process with different robust parameter estimators

从图 4-13 中明显可以看出，与没有孤立点处理的普通估计器相比，移动平均估计器明显减少了孤立点带来的输出波动。鲁棒移动窗口估计器也显示出在有孤立点污染时仍具有较好的鲁棒性，但是在 256 步时效果并不好，原因是在第 253 步和 256 步都出现了孤立点，然而该方法仅限于移动窗口中只存在一个孤立点的情况才有效。因为在第 256 步出现的孤立点和 253 步出现的孤立点在同一窗口，所以并没有得到检测和滤除。相比较而言，本章所提出的孤立点检测方法在这种情况下的检测效果更好，并且在整个控制过程中，仅仅在系统起步阶段有 12 步的调节时间，之后在孤立点噪声的污染下跟踪控制都相对比较平稳。

## 4.5 本章小结

本章提出了一种具有孤立点检测的自适应对偶控制方法，该方法对不确定系统中出现孤立点噪声时具有较强的鲁棒性。该方法在参数估计之前，通过对参与参数估计的数据进行有效的孤立点检测并用期望值替换孤立点，从而提高了参数估计的鲁棒性，进而提高系统控制性能。该孤立点检测算法利用实时观测数据设计了孤立点的距离和方向两个判断准则进行在线孤立点检测，计算量小，实时性强，适用于实际的工业控制过程。此外，该方法在设计控制器时还考虑到了不可控的激励信号，并将其嵌入到自适应对偶控制的结构中，以减少系统的不确定性，使其能更广泛的应用于实际场景。基于数学模型的数值实验表明，无论在观测噪声或系统过程噪声中出现孤立点，该方法都具有较高的鲁棒性。在实际的生物发酵连续灭菌过程的仿真对比实验中，与具有移动均值滤波的参数估计器和鲁棒移动窗口参数估计器下的自适应对偶控制方法相比，该方法具有更优的控制性能。



## 5 具有非对称拉布拉斯噪声的随机系统自适应对偶控制

### 5.1 引言

在一些存在随机噪声的实际系统，如经济系统、社会决策系统和生物生态系统中，随机噪声通常具有尖峰、厚尾、非对称的特征，而不是理想的高斯白噪声<sup>[129,130]</sup>。这种噪声会降低参数估计的精度，从而降低随机系统的控制性能，这一问题阻碍了随机控制在具有该特殊噪声的实际场景中的应用。例如，金融投资者希望设计一个调节器来更好地控制金融风险<sup>[131]</sup>；网络或服务行业的管理者希望做出尽可能满足大多数客户的决策，而不是平均水平下的决策<sup>[132,133]</sup>；政策制定者根据城市和农村之间的财富不平等来分配资源<sup>[134]</sup>等等。本章节研究了包含未知参数和具有尖峰、厚尾、非对称的特征的随机噪声的不确定随机系统的控制问题。

自适应对偶控制是解决不确定性随机系统控制问题的一种有效方法。该方法旨在设计一种具有对偶特性的控制律，可以一边通过学习来减少系统的不确定性，一边进行系统谨慎跟踪控制。例如，Milito 等人提出的基于新息的自适应对偶控制，是在代价函数中增加了系统新息，来减少系统的参数估计的不确定性<sup>[98]</sup>。Filatov 等人提出了双准则自适应对偶控制方法，即设计了两个代价函数，一个是系统的输出最优跟踪代价，一个是系统未知参数最优估计代价，对偶控制律可以通过同时优化这两个代价函数求得<sup>[106]</sup>。基于该方法的控制策略由于易于在微型机和微控制器单元上实现，目前在工业控制系统、航空航天系统、医疗系统以及其他系统中得到了广泛的应用<sup>[135-139]</sup>。然而上述自适应对偶控制方法仅考虑了具有理想高斯白噪声的参数未知系统的控制问题，通常使用卡尔曼滤波或者递归最小二乘作为未知参数的估计器。

由于本章研究对象为受到具有尖峰、厚尾、非对称随机噪声的影响的随机系统，所以上述假设随机噪声为高斯噪声的自适应对偶控制策略不再适用<sup>[129,130]</sup>。鲁棒控制是一类解决不确定性系统的控制问题的方案。例如，利用 Q-学习近似动态规划、启发式动态规划和双启发式动态规划，求解参数未知的离散时间线性系统的零和博弈相关的  $H_\infty$  最优控制问题<sup>[140-143]</sup>。然而鲁棒控制方法是在最坏干扰情况下的控制策略，因此控制策略比较保守，不考虑采用这种方法。

针对具有尖峰、厚尾、非对称特征的随机噪声的系统的控制，本章需要提出新的参数估计器，以提高系统参数估计的精度，从而提高自适应对偶控制性能。由于具有尖峰、厚尾、非对称特征的随机噪声可以用非对称拉布拉斯分布(Asymmetric Laplace Distribution, ALD)来描述，结合单分位点分位数回归可以用来估计包含 ALD 噪声的系统的模型参数<sup>[144-146]</sup>。然而当该噪声的分布未知时，在分位数回归中很难选择分位点<sup>[147]</sup>。因此就有学者在模型参数估计中结合了多个分位点的分位数回归的复合分位数回归方法。Zou 等人引入了一种结合多分位数回归模型的等加权复合分位数回归模型<sup>[148]</sup>。Zhao 等人

通过最小化渐近方差来计算不同分位数的不同权值<sup>[149]</sup>。Huang 等人提出了贝叶斯复合分位数回归，其中每个分位数的权重作为一个开放参数，通过马尔科夫链蒙特卡罗（Markov Chain Monte Carlo, MCMC）抽样程序进行估计<sup>[150]</sup>。关于分位数回归的工作大多是基于 MCMC 方法。然而 MCMC 方法并不适合用于自适应控制中的参数估计，因为它是离线进行的，而不是实时更新模型参数。

本章针对包含非对称拉布拉斯随机噪声的参数未知系统，提出了自适应分位数对偶控制方法。该方法包括两部分：（1）贝叶斯分位数求和估计器（Bayesian Quantile Sum Estimation, BQSE）；（2）具有对偶特性的输出跟踪控制器。所提出的 BQSE 参数估计器是一种实时递归参数估计器，用于估计存在 ALD 噪声的系统的参数。BQSE 综合了不同分位数下的参数估计值，通过对不同分位数下的参数估计值进行加权求和得到，权重为实时更新的贝叶斯后验概率。然后使用 BQSE 得到的参数估计值和估值误差协方差矩阵，计算基于新息的自适应对偶控制方法的控制律。最后在数值仿真实验中，将所提出的具有 BQSE 的自适应对偶控制方法与参数已知的最小方差控制，和基于 RLS 的自适应对偶控制进行比较，对比实验结果表明了该方法的有效性。

## 5.2 问题描述

考虑如下线性离散时间单输入单输出的随机系统

$$A(z^{-1})y(k) = B(z^{-1})u(k-d) + C(z^{-1})e(k) \quad k=0,1,\dots,N-1 \quad (5-1)$$

其中  $y(k)$  是系统输出， $u(k)$  是控制输入。 $A(z^{-1}) = 1 + \sum_{i=1}^n a_i z^{-i}$ ， $B(z^{-1}) = b_0 + \sum_{j=1}^m b_j z^{-j}$ ， $b_0 \neq 0$ ， $C(z^{-1}) = 1 + \sum_{i=1}^n c_i z^{-i}$ ，在  $A(z^{-1})$ ， $B(z^{-1})$  和  $C(z^{-1})$  里的参数未知。假设时间延迟系数  $d$ ，阶数  $n$  和  $m$  均已知。随机噪声  $e(k)$  具有尖峰、厚尾、非对称特征，且服从非对称拉布拉斯分布。

假设随机变量  $x$  服从非对称拉布拉斯分布，其相应的概率密度函数为

$$f_{pdf}(x) = \frac{\tau(1-\tau)}{\sigma} \begin{cases} e^{-\frac{(1-\tau)|x-\mu|}{\sigma}}, & x < \mu \\ e^{-\frac{\tau|x-\mu|}{\sigma}}, & x \geq \mu \end{cases} \quad (5-2)$$

其中  $\tau \in (0,1)$ ， $\mu \in (-\infty, \infty)$  和  $\sigma > 0$ 。

图 5-1 中高斯分布的均值设置为 0，方差为  $\sigma$ ，对称拉布拉斯分布的位置参数设置为  $\mu=0$ ，尺度参数为  $\sigma_1=1$  和  $\sigma_2=2$ ，不对称参数为  $\tau=0.5$ 。从图中可以看到与高斯分布相比，拉布拉斯分布具有尖峰和厚尾的特点。图 5-2 中不对称参数分别是  $\tau_2=0.9$ ， $\tau_3=0.2$ ， $\tau_4=0.9$ 。通常经济系统或生态系统等，其中的随机噪声服从如图 5-2 所示的具有尖峰、厚尾且非对称的特点的概率分布。

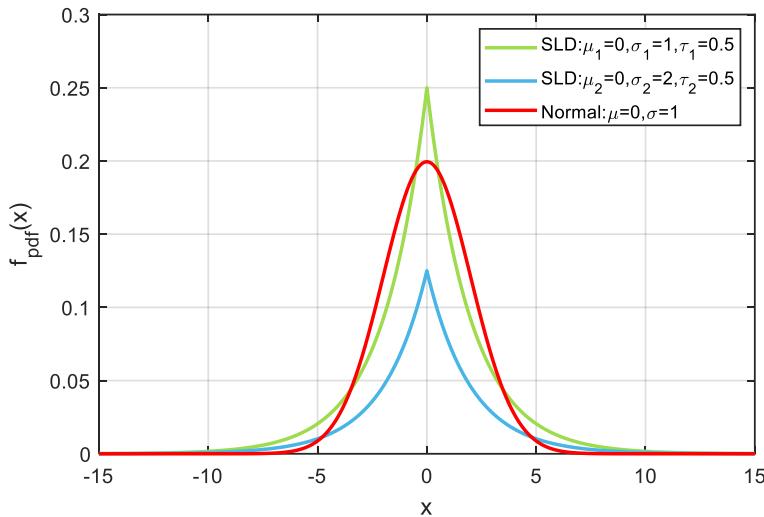


图 5-1 对称拉布拉斯分布和高斯分布的概率密度函数

Fig.5-1 The probability density function of Symmetric Laplace Distribution and Gaussian Distribution

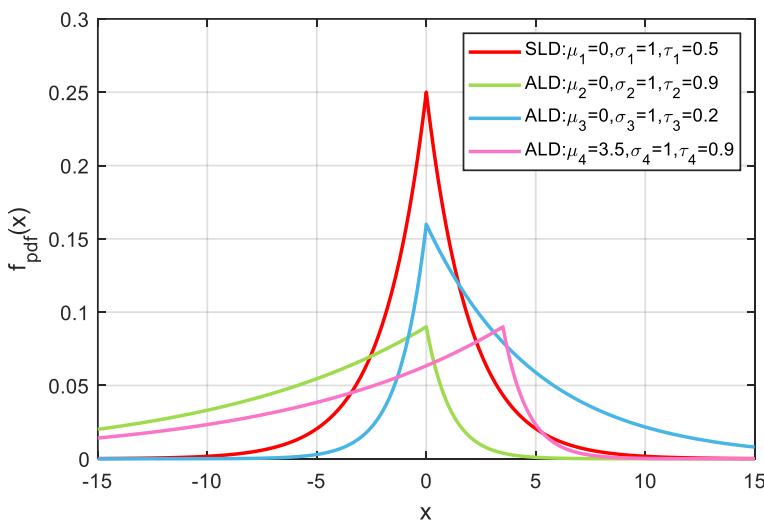


图 5-2 非对称拉布拉斯分布的概率密度函数

Fig.5-2 The probability density function of Asymmetric Laplace Distribution

根据式 (5-2)，非对称拉布拉斯分布的累积分布函数为

$$F(x) = \int_{-\infty}^x f(t)dt = \begin{cases} \tau e^{\frac{(1-\tau)}{\sigma}(x-\mu)}, & x < \mu \\ 1 - (1-\tau)e^{-\frac{\tau}{\sigma}(x-\mu)}, & x \geq \mu \end{cases} \quad (5-3)$$

具体推导过程如下所示：

如果  $x < \mu$ ，有

$$\begin{aligned} \int_{-\infty}^x f(t)dt &= \int_{-\infty}^x \frac{\tau(1-\tau)}{\sigma} e^{\frac{1-\tau}{\sigma}(t-\mu)} dt \\ &= \tau e^{\frac{1-\tau}{\sigma}(t-\mu)} \Big|_{-\infty}^x = \tau e^{\frac{1-\tau}{\sigma}(x-\mu)} \end{aligned} \quad (5-4)$$

如果  $x < \mu$ , 有

$$\begin{aligned}
\int_{-\infty}^x f(t)dt &= \int_{-\infty}^{\mu} f(t)dt + \int_{\mu}^x f(t)dt \\
&= \tau e^{\frac{1-\tau}{\sigma}(t-\mu)} \Big|_{-\infty}^{\mu} + \int_{\mu}^x \frac{\tau(1-\tau)}{\sigma} e^{\frac{-\tau}{\sigma}(t-\mu)} dt \\
&= \tau e^{\frac{1-\tau}{\sigma}(t-\mu)} \Big|_{-\infty}^{\mu} - (1-\tau)e^{\frac{-\tau}{\sigma}(t-\mu)} \Big|_{\mu}^x \\
&= 1 - (1-\tau)e^{\frac{-\tau}{\sigma}(x-\mu)}
\end{aligned} \tag{5-5}$$

非对称拉布拉斯噪声可以从相应的累积分布函数的反函数得到, 反函数为

$$F^{-1}(x) = \begin{cases} \mu + \frac{\sigma}{1-\tau} \ln \frac{1}{\tau} x, & 0 < x < \tau \\ \mu - \frac{\sigma}{\tau} \ln \frac{1}{1-\tau} (1-x), & \tau \leq x < 1 \end{cases} \tag{5-6}$$

具体的推导过程为:

如果  $x < \mu$ , 令  $y = \tau e^{\frac{1-\tau}{\sigma}(x-\mu)}$ , 则有  $y \in (0, \tau)$

$$\frac{y}{\tau} = e^{\frac{1-\tau}{\sigma}(x-\mu)} \Rightarrow \ln \frac{y}{\tau} = \frac{1-\tau}{\sigma}(x-\mu) \Rightarrow x = \mu + \frac{\sigma}{1-\tau} \ln \frac{y}{\tau} \tag{5-7}$$

如果  $x \geq \mu$ , 令  $y = 1 - (1-\tau)e^{\frac{-\tau}{\sigma}(x-\mu)}$ , 则有  $y \in [\tau, 1)$

$$\frac{1-y}{1-\tau} = e^{\frac{-\tau}{\sigma}(x-\mu)} \Rightarrow \ln \frac{1-y}{1-\tau} = \frac{-\tau}{\sigma}(x-\mu) \Rightarrow x = \mu - \frac{\sigma}{\tau} \ln \frac{1}{1-\tau} (1-y) \tag{5-8}$$

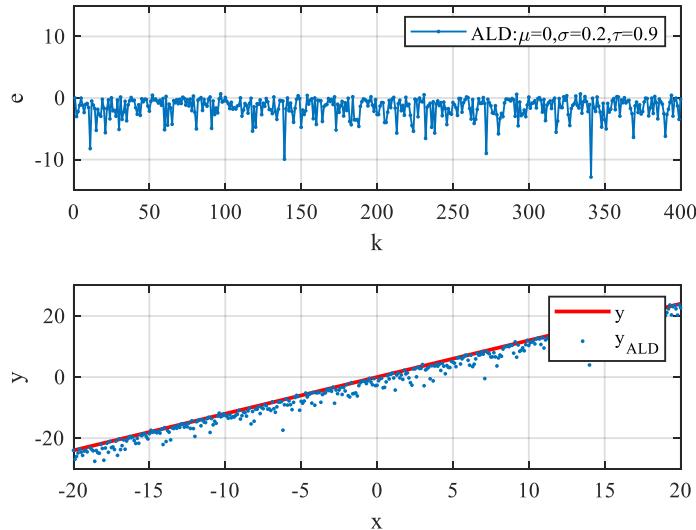


图 5-3 非对称拉布拉斯噪声  $e$ (非对称参数为  $\tau=0.9$ )和函数  $y=1.2x+e$

Fig.5-3 Asymmetric Laplace Noise  $e$  with the asymmetry parameter  $\tau=0.9$  and the output of  $y=1.2x+e$

图 5-3 的上半部分显示的是非对称拉布拉斯噪声  $e$ , 非对称参数  $\tau$  为 0.9, 位置参数

$\mu$  为 0, 尺度参数  $\sigma$  为 0.2。从图中可以看出, 大部分的噪声值都位于横轴  $x=0$  下方, 只有少部分点位于横轴上方。图 5-3 的下半部分显示的是含有噪声  $e$  的线性系统  $y=1.2x+e$  的输出。

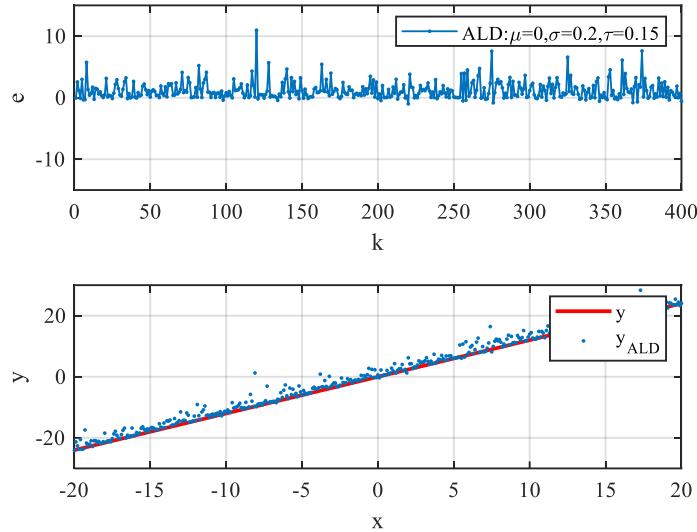


图 5-4 非对称拉布拉斯噪声  $e$ (非对称参数为  $\tau=0.15$ )和函数  $y=1.2x+e$

Fig.5-4 Asymmetric Laplace Noise  $e$  with the asymmetry parameter  $\tau=0.15$  and the output of  $y=1.2x+e$

图 5-4 显示了非对称参数  $\tau$  为 0.15 的非对称拉布拉斯噪声, 其值大部分位于横轴下方, 相应的  $y=1.2x+e$  如图下部分所示。

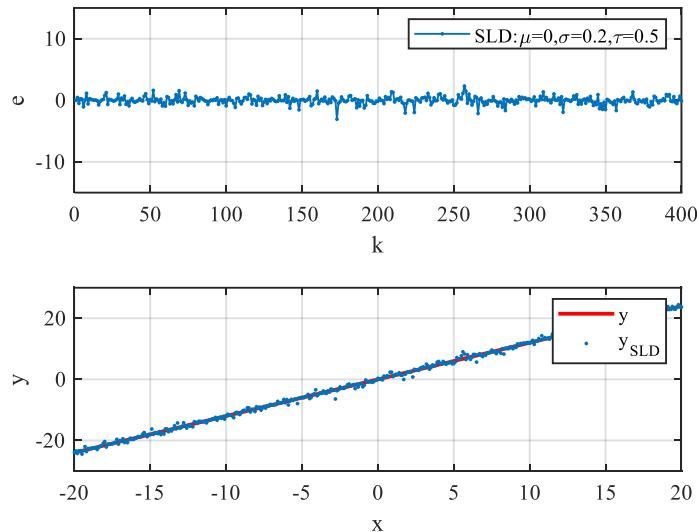


图 5-5 对称拉布拉斯噪声  $e$ (非对称参数为  $\tau=0.5$ )和函数  $y=1.2x+e$

Fig.5-5 Asymmetric Laplace Noise  $e$  with the asymmetry parameter  $\tau=0.5$  and the output of  $y=1.2x+e$

图 5-5 显示的是非对称参数  $\tau$  为 0.5 时的拉布拉斯噪声, 也就是对称拉布拉斯噪声, 其对称的分布在横轴上下, 相应的  $y=1.2x+e$  如图下部分所示。

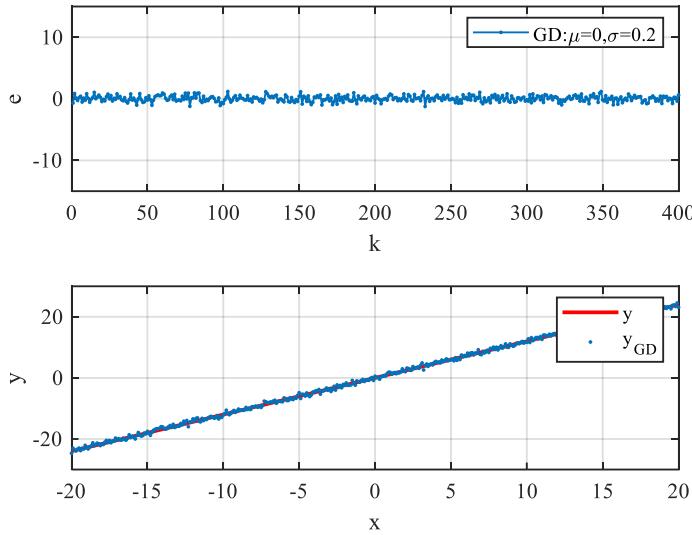
图 5-6 高斯白噪声  $e$  和函数  $y = 1.2x + e$ Fig.5-6 White Gaussian Noise  $e$  and the output of  $y = 1.2x + e$ 

图 5-6 显示的是均值  $\mu$  为 0, 方差  $\sigma$  为 0.2 的高斯白噪声, 相应的  $y = 1.2x + e$  如图下部分所示。

对于非对称拉布拉斯噪声, 定义其期望为  $\mu_{lap}$ , 方差为  $\sigma_{lap}$ , 计算如下

$$\mu_{lap} = \frac{\sigma(1-2\tau)}{\tau(1-\tau)} \quad (5-9)$$

$$\sigma_{lap}^2 = \frac{\sigma^2(1-2\tau+2\tau^2)}{\tau^2(1-\tau)^2} \quad (5-10)$$

其中假设位置参数  $\mu$  是 0。具体计算过程如下所示:

$$\begin{aligned} E(x) &= \int_{-\infty}^{+\infty} xf(x)dx \\ &= \frac{\tau(1-\tau)}{\sigma} \left[ \int_{-\infty}^{\mu} xe^{\frac{1-\tau}{\sigma}(x-\mu)} dx + \int_{\mu}^{+\infty} xe^{\frac{-\tau}{\sigma}(x-\mu)} dx \right] \\ &= \frac{\mu\tau(1-\tau) + \sigma - 2\tau\sigma}{\tau(1-\tau)} \end{aligned} \quad (5-11)$$

当  $\mu = 0$ , 则  $\mu_{lap} = E(x) = \frac{\sigma(1-2\tau)}{\tau(1-\tau)}$ 。

$$\begin{aligned} E(x^2) &= \int_{-\infty}^{+\infty} x^2 f(x)dx \\ &= \frac{\tau(1-\tau)}{\sigma} \left[ \int_{-\infty}^{\mu} x^2 e^{\frac{1-\tau}{\sigma}(x-\mu)} dx + \int_{\mu}^{+\infty} x^2 e^{\frac{-\tau}{\sigma}(x-\mu)} dx \right] \\ &= \tau\mu^2 - \frac{2\sigma\tau}{1-\tau} \left( \mu - \frac{\sigma}{1-\tau} \right) + (1-\tau)\mu^2 + \frac{2\sigma(1-\tau)}{\tau} \left( \mu + \frac{\sigma}{\tau} \right) \end{aligned} \quad (5-12)$$

当  $\mu=0$  时, 有  $E(x^2)=\frac{2\tau\sigma^2}{(1-\tau)^2}+\frac{2(1-\tau)\sigma^2}{\tau^2}$ , 则可得方差为

$$\begin{aligned} D(x) &= E(x^2)-E^2(x) \\ &= \frac{2\tau\sigma^2}{(1-\tau)^2}+\frac{2(1-\tau)\sigma^2}{\tau^2}-\frac{\sigma^2(1-2\tau)^2}{\tau^2(1-\tau)^2} \\ &= \frac{\sigma^2(1-2\tau+2\tau^2)}{\tau^2(1-\tau)^2} \end{aligned} \quad (5-13)$$

控制目标为设计控制律  $u(k)$  使系统输出  $y(k)$  能够跟踪期望输出  $y_r(k)$ , 代价函数为

$$J = E \sum_{k=0}^{N-1} [y(k+1) - y_r(k+1)]^2 \quad (5-14)$$

自适应对偶控制问题为

$$\begin{aligned} \min_{u(k)} J \\ s.t. \quad A(z^{-1})y(k) = B(z^{-1})u(k-d) + C(z^{-1})e(k) \end{aligned} \quad (5-15)$$

其中参数  $A(z^{-1})$ ,  $B(z^{-1})$  和  $C(z^{-1})$  未知,  $e(k)$  为非对称拉布拉斯噪声。

### 5.3 控制器设计

本节详细介绍了具有 BQSE 的自适应对偶控制器的设计。图 5-7 是所提出的具有 BQSE 的自适应对偶控制的结构框图。

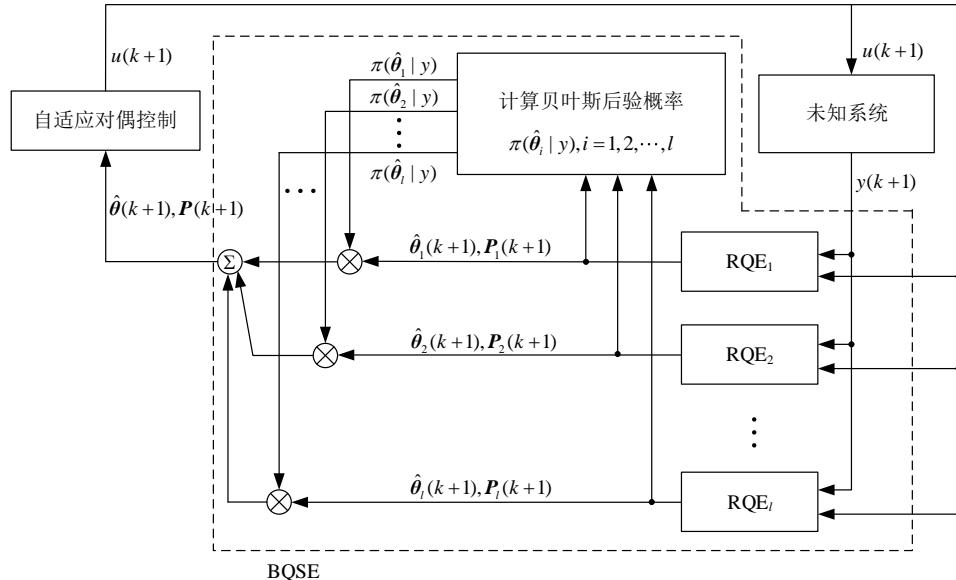


图 5-7 具有 BQSE 的自适应对偶控制的结构框图  
Fig 5-7 Block diagram illustrating the adaptive quantile control with BQSE

图 5-7 中未知系统由式 (5-1) 所示的具有非对称拉布拉斯噪声的差分模型表示。虚线框内显示了 BQSE 估计器的结构, 该估计器可以为后续的自适应对偶控制器实时提供

估计的参数  $\hat{\theta}(k)$  和参数估计误差协方差矩阵  $P(k)$ 。在 BQSE 中， $l$  个不同分位点下的单分位数估计器并行运行，每个估计器都生成对应的估计参数  $\hat{\theta}_i(k)$ 。最终估计的参数值为所有分位点估计的参数  $\hat{\theta}_i(k)$  的加权和，权重为每个分位点下的参数估计的贝叶斯后验概率  $\pi(\hat{\theta}_i | y)$ 。

### 5.3.1 差分模型

假设式 (5-1) 所示的系统的辅助输出表达式为

$$y_a(k) = P(z^{-1})y(k) + Q(z^{-1})u(k-d) - R(z^{-1})y_r(k) \quad (5-16)$$

其中  $y_r(k)$  是目标轨迹， $P(z^{-1}) = p_0 + \sum_{i=1}^{n_p} p_i z^{-i}$ ， $Q(z^{-1}) = 1 + \sum_{i=1}^{n_q} q_i z^{-i}$ ， $R(z^{-1}) = 1 + \sum_{i=1}^{n_r} r_i z^{-i}$ 。

引入式  $P(z^{-1})C(z^{-1}) = A(z^{-1})L(z^{-1}) + z^{-d}G(z^{-1})$ ，并带入式 (5-1) 得到

$$C(z^{-1})y_a(k+d) = F(z^{-1})u(k) + G(z^{-1})y(k) + H(z^{-1})y_r(k+d) + C(z^{-1})\bar{e}(k+d) \quad (5-17)$$

其中  $F(z^{-1}) = Q(z^{-1})C(z^{-1}) + B(z^{-1})L(z^{-1})$ ， $H(z^{-1}) = -C(z^{-1})R(z^{-1})$ ， $\bar{e}(k) = L(z^{-1})e(k)$ 。式 (5-17) 的具体推导过程如下所示：

将式 (5-16) 代入式 (5-1) 中得到

$$\begin{aligned} z^d y_a(k) &= z^d P(z^{-1})y(k) + Q(z^{-1})u(k) - z^d R(z^{-1})y_r(k) \\ \Rightarrow z^d y_a(k) &= z^d P(z^{-1}) \frac{B(z^{-1})u(k-d) + C(z^{-1})e(k)}{A(z^{-1})} + Q(z^{-1})u(k) - z^d R(z^{-1})y_r(k) \\ \Rightarrow z^d C(z^{-1})y_a(k) &= \left[ \frac{B(z^{-1})P(z^{-1})C(z^{-1})}{A(z^{-1})} + Q(z^{-1})C(z^{-1}) \right] u(k) \\ &\quad + z^d \frac{C(z^{-1})P(z^{-1})C(z^{-1})}{A(z^{-1})} e(k) - z^d R(z^{-1})C(z^{-1})y_r(k) \end{aligned} \quad (5-18)$$

将  $P(z^{-1})C(z^{-1}) = A(z^{-1})L(z^{-1}) + z^{-d}G(z^{-1})$  代入上式中得

$$\begin{aligned} z^d C(z^{-1})y_a(k) &= \left[ B(z^{-1})L(z^{-1}) + Q(z^{-1})C(z^{-1}) \right] u(k) + G(z^{-1})y(k) \\ &\quad + z^d C(z^{-1})L(z^{-1})e(k) - z^d R(z^{-1})C(z^{-1})y_r(k) \end{aligned} \quad (5-19)$$

将上式重新整理后可以得到式 (5-16) 的表达形式。

多项式  $D(z^{-1}) = P(z^{-1})B(z^{-1}) + Q(z^{-1})A(z^{-1})$  的特征根需位于  $z$  平面上的单位圆内以确保系统的稳定性，这可以通过选择合适的  $P(z^{-1})$  和  $Q(z^{-1})$  来满足。令系统未知参数向量为  $\theta^T = [f_0, f_1, \dots, f_{n_f}, g_0, g_1, \dots, g_{n_g}, h_0, h_1, \dots, h_{n_h}]$ ，那么式 (5-16) 所示的系统可以改写为如下形式

$$y_a(k+1) = \Phi^T(k)\theta(k) - y_r(k+1) + \bar{e}(k+1) \quad (5-20)$$

其中  $\Phi^T(k) = [u(k), \dots, u(k-n_f), y(k), \dots, y(k-n_g), y_r(k); \dots; y_r(k-n_h)]$ ， $\bar{e}(k)$  是非对称拉布拉

斯噪声，非对称系数为 $\bar{\tau}$ 。为了能简化接下来的方法推导，时延参数 $d$ 设置为1，系数 $C(z^{-1})$ 设置为1。

### 5.3.2 迭代分位数估计器

普通的基于迭代最小二乘的参数估计只适用于含有高斯白噪声的系统，而对于具有非高斯白噪声的系统并不适用。大量文献表明，在实际系统中，还有一种具有尖峰、厚尾和非对称特性的随机噪声。对此 Roger Koenker 提出了分位数回归，这在非高斯噪声的情况下被广泛应用。然而，该方法是基于 MCMC 采样方法，如 Metropolis-Hastings 和 Gibbs 采样等，这些方法并不适用于实时控制过程。为了能够实时估计模型参数，需要对传统的分位数回归进行改进。本节介绍了递归分位数估计方法。

给定随机变量 $\Phi$ ， $Y$ 的条件累积分布函数定义为 $F_{Y|\Phi}(y)$ ， $Y$ 的 $\tau$ -条件分位数为

$$Q_\tau(y|\Phi) = \inf_y \{y : F_{Y|\Phi}(y) \geq \tau\} \quad (5-21)$$

其中 $\inf\{\cdot\}$ 是下确界函数。考虑式(5-20)所描述的线性模型，生成 $n$ 个样本

$$y(k+1) = \Phi^T(k)\theta(k) + e(k+1), \quad k=1, 2, \dots, n \quad (5-22)$$

其中的一组观测样本表示为 $\{\Phi(k), y(k+1)\}$ ， $e(k+1)$ 为非对称拉布拉斯噪声， $y(k+1)$ 的 $\tau$ -条件分位数为 $Q_\tau\{y(k+1)|\Phi(k)\} = \Phi^T(k)\theta_\tau(k)$ 。通过最小化损失函数求得 $\theta$ 的 $\tau$ -分位数估计，损失函数为

$$\sum_{k=1}^n \rho_\tau\{y(k+1) - \Phi^T(k)\theta_\tau(k)\} \quad (5-23)$$

其中 $\rho_\tau(u) = \begin{cases} \tau u^2, & u \geq 0 \\ (\tau-1)u^2, & u < 0 \end{cases}$ 。最小化式(5-23)求 $\theta$ 的 $\tau$ -分位数估计，还可以写成如下形式

$$\begin{aligned} \theta_\tau = \arg \min \sum_{k=1}^n & \left[ (1-\tau) \sum_{y(k+1) < \Phi^T(k)\theta(k)} (y(k+1) - \Phi^T(k)\theta(k))^2 \right. \\ & \left. + \tau \sum_{y(k+1) \geq \Phi^T(k)\theta(k)} (y(k+1) - \Phi^T(k)\theta(k))^2 \right] \end{aligned} \quad (5-24)$$

式(5-24)是一个分段二次函数，通过求解该式就可以得到迭代分位数估计，具体的迭代计算过程如下

$$\begin{aligned} K(k+1) &= P(k)\Phi(k)[\lambda(k) + \Phi^T(k)P(k)\Phi(k)]^{-1} \\ \theta(k+1) &= \theta(k) + K(k+1)[y(k+1) - \theta^T(k)\Phi(k)] \\ P(k+1) &= [I - K(k+1)\Phi^T(k)]P(k) \end{aligned} \quad (5-25)$$

其中

$$\lambda(k) = \begin{cases} 1/(1-\tau), & y(k+1) < \theta^T(k)\Phi(k) \\ 1/\tau, & y(k+1) \geq \theta^T(k)\Phi(k) \end{cases} \quad (5-26)$$

详细的推导过程如下，根据最小二乘法，在第  $k$  步的参数估计值  $\theta$  为

$$\theta_k = (\Phi_k^T W_k \Phi_k)^{-1} \Phi_k^T W_k Y_k \quad (5-27)$$

其中  $W_{k-1} = \begin{bmatrix} W_{k-1} & \cdots \\ \cdots & \tau_k \end{bmatrix}$ ,  $\Phi_k = \begin{bmatrix} \Phi_{k-1} \\ \varphi(k) \end{bmatrix}$ ,  $Y_k = \begin{bmatrix} Y_{k-1} \\ y(k+1) \end{bmatrix}$ 。令

$$\begin{aligned} P(k) &= [\Phi_k^T W_k \Phi_k]^{-1} \\ &= [\Phi_{k-1}^T W_{k-1} \Phi_{k-1} + \tau \varphi(k) \varphi^T(k)]^{-1} \\ &= [P^{-1}(k-1) + \tau \varphi(k) \varphi^T(k)]^{-1} \end{aligned} \quad (5-28)$$

上式也可写成如下形式

$$P^{-1}(k) = P^{-1}(k-1) + \tau \varphi(k) \varphi^T(k) \quad (5-29)$$

将  $P(k) = [\Phi_k^T W_k \Phi_k]^{-1}$  代入到第  $k-1$  步的参数估计值  $\theta_{k-1}$  中得

$$\begin{aligned} \theta_{k-1} &= (\Phi_{k-1}^T W_{k-1} \Phi_{k-1})^{-1} \Phi_{k-1}^T W_{k-1} Y_{k-1} \\ &= P(k-1) \Phi_{k-1}^T W_{k-1} Y_{k-1} \end{aligned} \quad (5-30)$$

将  $P^{-1}(k) = P^{-1}(k-1) + \tau \varphi(k) \varphi^T(k)$  代入上式可得

$$\begin{aligned} \Phi_{k-1}^T W_{k-1} Y_{k-1} &= P^{-1}(k-1) \theta(k-1) \\ &= [P^{-1}(k) - \tau \varphi(k) \varphi^T(k)] \theta(k-1) \end{aligned} \quad (5-31)$$

那么在第  $k$  步的参数估计值  $\theta$  为

$$\begin{aligned} \theta(k) &= P(k) \Phi_k^T W_k Y_k \\ &= P(k) [\Phi_{k-1}^T W_{k-1} Y_{k-1} + \tau \varphi(k) y(k)] \\ &= P(k) \{ [P^{-1}(k) - \tau \varphi(k) \varphi^T(k)] \theta(k-1) + \tau \varphi(k) y(k) \} \\ &= \theta(k-1) + \tau P(k) \varphi(k) [y(k) - \varphi^T(k) \theta(k-1)] \\ &= \theta(k-1) + K(k) [y(k) - \varphi^T(k) \theta(k-1)] \end{aligned} \quad (5-32)$$

其中  $K(k) = \tau P(k) \varphi(k)$ 。那么  $P(k)$  可以改写为如下形式

$$P(k) = P(k-1) - \frac{\tau P(k-1) \varphi(k) \varphi^T(k) P(k-1)}{I + \tau \varphi^T(k) P(k-1) \varphi(k)} \quad (5-33)$$

将上式带入到  $K(k)$  中可得

$$K(k) = \frac{\tau P(k-1) \varphi(k)}{I + \tau \varphi^T(k) P(k-1) \varphi(k)} \quad (5-34)$$

将以上  $K(k)$  代入  $P(k)$  可得

$$P(k) = [I - K(k) \varphi^T(k)] P(k-1) \quad (5-35)$$

以上即为迭代分位数参数估计的详细推导过程。

### 5.3.3 贝叶斯分位数求和估计器

上一小节通过设置不对称参数  $\tau$  来得到不同分位点的估计参数。然而，由于噪声的分布未知，很难确定最优的  $\tau$ 。本小节提出了将多个递归分位数估计器组合成一个参数估计器，即贝叶斯分位数求和估计器，BQSE。BQSE 中通过将每个单独的递归分位数估计值以贝叶斯后验概率为权重来求和，得到最终的复合参数估计值。

通过求解以下代价函数来求取 BQSE 下的参数估计  $\theta$

$$\theta = \arg \min \sum_{i=1}^m \sum_{k=1}^n \left[ (1-\tau_i) \sum_{y(k+1) < \Phi^T(k)\theta(k)} (y(k+1) - \Phi^T(k)\theta(k))^2 + \tau_i \sum_{y(k+1) \geq \Phi^T(k)\theta(k)} (y(k+1) - \Phi^T(k)\theta(k))^2 \right] \quad (5-36)$$

其中分位点设置为  $0 < \tau_1 < \tau_2 < \dots < \tau_i < \dots < \tau_m < 1$ ，直接求解式 (5-36) 是很困难的。根据式 (5-25)，可以得到每个分位点  $\tau_i$  下的参数估计  $\theta_i$ ，然后用每个分位点下估计的参数的贝叶斯后验概率  $\pi_i$  作为相应的估计参数  $\theta_i$  的权重，进行加权求和运算得到最终的参数估计值  $\theta$ 。

给定观测值  $y = \{y_1, y_2, \dots, y_n\}$ ，参数  $\theta$  的条件概率定义为  $\pi(\theta | y)$ 。根据贝叶斯后验概率定义， $\pi(\theta | y)$  为

$$\pi(\theta | y) = \frac{f_{pdf}(y | \theta) \pi(\theta)}{\int_{\Theta} f_{pdf}(y | \theta) \pi(\theta) d\theta} \quad (5-37)$$

其中  $\Theta$  是参数空间， $\pi(\theta)$  是  $\theta$  的先验分布， $f_{pdf}(y | \theta)$  是模型的概率密度函数。相应的离散后验概率定义为

$$\pi(\theta_i | y) = \frac{f_{pdf}(y | \theta_i) \pi(\theta_i)}{\sum_i f_{pdf}(y | \theta_i) \pi(\theta_i)} \quad (5-38)$$

其中  $\pi(\theta_i | y)$  是先验概率，概率密度函数  $f_{pdf}(y | \theta_i)$  为

$$f_{pdf}(y | \theta_i) = \tau_i (1-\tau_i) \begin{cases} e^{-(1-\tau_i) \|y - \Phi^T \theta_i\|}, & (y - \Phi^T \theta_i) < 0 \\ e^{-\tau_i \|y - \Phi^T \theta_i\|}, & (y - \Phi^T \theta_i) \geq 0 \end{cases} \quad (5-39)$$

根据式 (5-25) 和式 (5-36)，在分位点  $\tau_i$  下的参数估计  $\theta_i$  的迭代过程为

$$\begin{aligned} K_i(k+1) &= P_i(k | k) \Phi(k) [\lambda_i(k) + \Phi^T(k) P_i(k) \Phi(k)]^{-1} \\ \theta_i(k+1) &= \theta_i(k) + K_i(k+1) [y(k+1) - \theta_i^T(k) \Phi(k)] \\ P_i(k+1) &= [I - K_i(k+1) \Phi^T(k)] P_i(k) \end{aligned} \quad (5-40)$$

其中

$$\lambda_i(k) = \begin{cases} 1/(1-\tau_i), & y(k+1) < \theta_i^T(k)\Phi(k) \\ 1/\tau_i, & y(k+1) \geq \theta_i^T(k)\Phi(k) \end{cases} \quad (5-41)$$

则贝叶斯分位数求和估计值为

$$\theta(k) = \sum_{i=1}^m \pi(\theta_i | y)\theta_i(k) \quad (5-42)$$

协方差矩阵为

$$P(k+1) = \sum_{i=1}^m \pi(\theta_i | y)P_i(k) \quad (5-43)$$

贝叶斯分位数求和估计的结构如图 5-7 中的虚线所示。

### 5.3.4 自适应对偶控制器

自适应对偶控制是基于随机系统的分析。在自适应对偶控制中，不仅考虑到了系统的输出跟踪效果，还考虑到系统未知参数的学习效果。该控制方案的优点是可以一边估计未知系统的模型参数，一边驱使系统做最优跟踪控制。由于式 (5-15) 中的对偶控制问题难以求解，根据基于新息的自适应对偶控制方法，可以通过最小化相应的包含系统新息的代价函数求得，具体过程如下所示

$$u^*(k) = \arg \min_{u(k)} E\{[y_a(k+1)]^2 - \beta(k)[v(k+1)]^2 | \theta_k, \mathfrak{I}_k, u(k)\} \quad (5-44)$$

其中  $y_a(k+1)$  是系统的辅助输出， $\mathfrak{I}_k = \{u(1), \dots, u(k-1), y(1), \dots, y(k)\}$  是信息状态， $v(k+1)$  就是预测误差，也就是系统新息

$$v(k+1) = \theta^T(k)\Phi(k) - \theta^T(k)\Phi(k) \quad (5-45)$$

$\beta(k)$  是学习系数，取值范围为  $0 \leq \beta(k) \leq 1$ 。学习系数  $\beta(k)$  由设计者选择，该系数衡量的是在控制方面的优化和学习方面的优化的折中程度。式 (5-44) 中明显表达出了控制器的对偶特性，即代价函数中  $-\beta(k)[v(k+1)]^2$  这一部分增加了估计误差，但代价是系统输出偏离目标轨迹变大。这就会增加参数估计时所需的信息丰富程度，从而提高参数估计的精度，进一步提高整体的控制性能。

下面将参数向量  $\theta(k)$ ，观测向量  $\Phi(k)$  和协方差矩阵  $P(k)$  进行分块处理

$$\theta^T(k) = [f_0(k) \ : \ \alpha^T(k)] \quad (5-46)$$

$$\Phi^T(k) = [u(k) \ : \ \varphi^T(k)] \quad (5-47)$$

$$P(k) = \begin{bmatrix} P_{f_0}(k) & \vdots & P_{\alpha f_0}^T(k) \\ \cdots & \cdots & \cdots \\ P_{\alpha f_0}(k) & \vdots & P_\alpha(k) \end{bmatrix} \quad (5-48)$$

将分块向量 (5-46) 和 (5-47) 代入辅助系统 (5-20) 可得

$$y_a(k+1) = f_0(k)u(k) + \alpha^T(k)\varphi(k) - y_r(k+1) + \bar{e}(k+1) \quad (5-49)$$

根据式 (5-46) 至 (5-49)，式 (5-44) 中的代价函数的计算过程如下

$$\begin{aligned} J(k) &= E\left\{[y_a(k+1)]^2 - \beta(k)[v(k+1)]^2 \mid \theta_k, \mathfrak{I}_k, u(k)\right\} \\ &= E\left\{[1-\beta(k)][\tilde{\theta}^T(k)\Phi(k)]^2 + [\tilde{\theta}^T(k)\Phi(k) - y_r(k+1) + \bar{e}(k+1)]^2\right\} \\ &= [1-\beta(k)][P_{f_0}(k)u^2(k) + 2P_{\alpha f_0}^T(k)\varphi(k)u(k) + P_\alpha(k)\varphi^T(k)\varphi(k)] \\ &\quad + [\hat{f}_0(k)u(k) + \hat{\alpha}^T(k)\varphi(k) - y_r(k+1)]^2 \\ &\quad + 2[\hat{f}_0(k)u(k) + \hat{\alpha}^T(k)\varphi(k) - y_r(k+1)]\mu_{Lap} + \sigma_{Lap} \end{aligned} \quad (5-50)$$

其中  $E\{\tilde{\theta}^T(k)\tilde{\theta}(k)\} = P(k)$ ， $E\{\bar{e}(k+1)\} = \mu_{lap}$ ， $E\{\bar{e}^2(k+1)\} = \sigma_{lap}$ 。令  $\frac{\partial J(k)}{\partial u(k)} = 0$  可求得

自适应对偶控制律

$$u^*(k) = -\frac{[1-\beta(k)]P_{\alpha f_0}^T(k)\varphi(k) + \hat{f}_0(k)[\hat{\alpha}^T(k)\varphi(k) - y_r(k+1)] + \hat{f}_0(k)\mu_{Lap}}{[1-\beta(k)]P_{f_0}(k) + \hat{f}_0^2(k)} \quad (5-51)$$

其中  $\mu_{Lap} = \{\sigma(1-2\tau)\}/\{\tau(1-\tau)\}$ ， $\tau = \sum_{i=1}^m \pi(\hat{\theta}_i \mid y)\hat{\tau}_i$ 。

## 5.4 仿真实验

本节对所提出的具有贝叶斯分位数求和估计器的自适应对偶控制方法进行了仿真实验，验证该方法的有效性。第 5.4.1 小节使用一组数学模型生成的数据对 BQSE 的参数估计性能进行了仿真实验和分析，并将其与递推最小二乘、递推分位数估计器进行比较，不仅比较了单次仿真实验结果，还比较了蒙特卡洛仿真实验下的统计结果；第 5.4.2 和 5.4.3 小节分别在最小相位和非最小相位系统中，对具有 BQSE 的自适应对偶控制器的性能进行了仿真实验和分析，并将其与参数已知的最优控制和基于递推最小二乘的自适应对偶控制进行比较。以下为本章所提出的具有贝叶斯分位数求和估计器的自适应对偶控制方法的具体执行步骤：

步骤 1：初始化参数向量  $\theta_i(1)$ ，信息状态  $\mathfrak{I}_1$ ，协方差矩阵  $P_i(1)$ ，非对称参数  $\tau_i$ ，贝叶斯后验概率  $\pi_i(1)$ ，其中贝叶斯后验概率的初值设置的时候需要满足在所有分位点下的概率和为 1，即  $\sum_{i=1}^l \tau_i = 1$ ，参数  $\tau_i$  的取值应该在区间 (0,1) 内；

步骤 2：根据式 (5-41) 在第  $k+1$  步计算分位点  $\tau_i$  下的估计参数  $\hat{\theta}_i(k+1)$ ；

步骤 3：根据式 (5-38) 计算第  $k+1$  步的贝叶斯后验概率  $\pi(\theta_i \mid y)$ ；

步骤 4：根据式 (5-42) 和式 (5-43) 分别计算  $\hat{\theta}(k+1)$  和  $P(k+1)$ ；

步骤 5：根据式 (5-51) 求自适应对偶控制律  $u^*(k+1)$ ，式 (5-51) 中非对称拉布拉

斯噪声的期望需要在每一步都进行更新，其中的参数就是由 $\tau = \sum_{i=1}^m \pi(\hat{\theta}_i | y) \hat{\tau}_i$ 进行更新；

步骤 6：将 $u^*(k+1)$ 作为系统的控制输入，并且返回到步骤 2。

### 5.4.1 贝叶斯分位数求和估计器

本小节通过仿真实验将所提出的 BQSE 参数估计方法和递推最小二乘、递推分位数估计方法进行对比。测试数据是由式（5-1）描述的系统生成，系统参数设置为

$$\begin{aligned} a_1 &= -1.41, \quad a_2 = 0.9, \quad n = 2, \\ b_0 &= 0.5, \quad m = 0, \quad d = 1, \\ \sigma &= 0.02, \quad \tau = 0.9, \quad \mu = 0. \end{aligned} \quad (5-52)$$

控制信号设置为 $u(k) = [y_r(k+1) - a_1 y(k) - a_2 y(k-1)] / b_0$ ，其中的 $y_r$ 设置为 0.1Hz 的方波信号经过传递函数 $1/(s+1)$ 滤波后的数据，非对称拉布拉斯噪声 $e(k)$ 设置为

$$e(k) = \begin{cases} \mu + \frac{\sigma}{1-\tau} \ln \frac{1}{\tau} x(k), & 0 < x(k) < \tau \\ \mu - \frac{\sigma}{\tau} \ln \frac{1}{1-\tau} (1-x(k)), & \tau \leq x(k) < 1 \end{cases} \quad (5-53)$$

其中 $x(k)$ 设置为均匀分布。根据以上设置收集了 100 组样本数据用于参数估计仿真。

初始参数设置为 $\theta_i^T(1) = [0.1, 0.1, 0.1]$ ，初始误差协方差矩阵设置为 $P_i(1) = 100I$ 。在不同的分位点 $\tau_i$ 下，初始贝叶斯后验概率 $\pi_i(1)$ 设置为 $\{\tau_1 = 0.9, \pi_1(1) = 1/9\}$ ， $\{\tau_2 = 0.8, \pi_2(1) = 1/9\}$ ， $\{\tau_3 = 0.7, \pi_3(1) = 1/9\}$ ， $\{\tau_4 = 0.6, \pi_4(1) = 1/9\}$ ， $\{\tau_5 = 0.5, \pi_5(1) = 1/9\}$ ， $\{\tau_6 = 0.4, \pi_6(1) = 1/9\}$ ， $\{\tau_7 = 0.3, \pi_7(1) = 1/9\}$ ， $\{\tau_8 = 0.2, \pi_8(1) = 1/9\}$ ， $\{\tau_9 = 0.1, \pi_9(1) = 1/9\}$ ，所有的贝叶斯后验概率总和是 1。

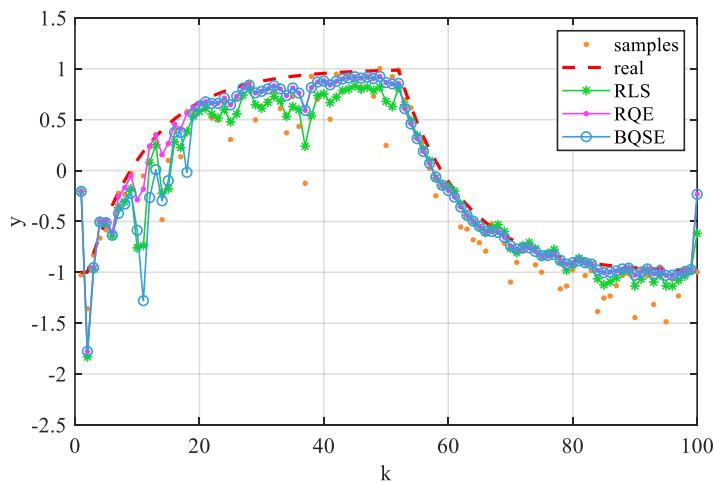


图 5-8 迭代最小二乘 (RLS)，递推分位数估计(RQE)，贝叶斯求和估计器 (BQSE) 对应的系统输出预测

Fig.5-8 The estimation of system output for RLS, RQE and BQSE

图 5-8 是使用了不同的参数估计器下系统的输出预测值，即根据估计的参数  $\hat{\theta}(k)$  实时预测的系统输出  $y(k)$ 。在图 5-8 中的“sample”指的就是用来做参数估计的样本数据，包含了非对称拉布拉斯噪声，“real”指没有噪声污染的实际模型生成的样本数据，“RLS”代表迭代最小二乘法，“RQE”代表递推分位数估计，在这个实验中分位点已知， $\tau = 0.9$ ，“BQSE”是贝叶斯分位数求和估计器，在这个实验中分位点是未知的。由于在 RQE 的实验中，分位点  $\tau$  是确切已知的，所以 RQE 的估计参数就是一个理想的估计结果，可以作为参考基准。从图中可以看出 RQE 和 BQSE 的估计值比 RLS 的估计系统输出更接近真实值。BQSE 的曲线在第 15 步迭代之后和 RQE 的曲线重合，表明贝叶斯后验概率在 15 步后开始收敛。

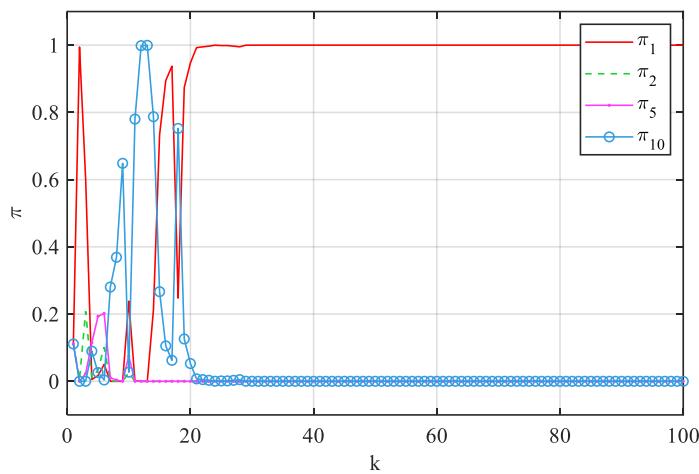
图 5-9 不同分位点  $\tau_i$  的贝叶斯后验概率  $\pi_i$ Fig.5-9 The Bayesian posterior probability  $\pi_i$  for different  $\tau_i$ 

图 5-9 显示了在 BQSE 参数估计器中，不同分位点对应的贝叶斯后验概率的收敛过程。分位点为  $\tau_1 = 0.9$  对应的后验概率  $\pi_1$  收敛到了 1，其他后验概率均收敛到 0，表明后验概率收敛到正确的分位点。

单次仿真实验的参数估计的性能可以用输出累积误差来衡量

$$V = \frac{1}{N} \sum_{k=1}^N [y_{est}(k) - y_{real}(k)]^2 \quad (5-54)$$

其中  $y_{real}(k)$  是不同方法下的估计输出， $y_{real}(k)$  是没有噪声污染的实际模型的输出， $N$  是仿真步长。 $M$  次蒙特卡洛实验的平均累积误差为

$$\bar{V} = \frac{1}{M} \sum_{i=1}^M V(i) \quad (5-55)$$

下面对该系统进行 1000 次蒙特卡洛仿真实验，并且计算在 RLS、RQE 和 BQSE 三种不同方法下的平均性能指标，也就是平均累积误差进行比较。

表 5-1 三种不同参数估计器的性能指标比较  
Tab.5-1 Comparison of performance for three estimation methods

估计方法	性能指标均值
RLS	0.0620
RQE	0.0242
BQSE	0.0380

表 5-1 是 1000 次蒙特卡洛仿真的平均估计性能，BQSE 的平均累积误差值比 RLS 低，且和 RQE 的值很接近，这意味着 BQSE 收敛到了真实的分位数下的参数估计值。

#### 5.4.2 最小相位系统仿真实验

本小节对最小相位系统进行仿真控制，将所提出的具有 BQSE 的自适应对偶控制与基于 RLS 的自适应对偶控制和最优控制进行比较。考虑如式 (5-1) 所描述的系统，相应的参数设置如下

$$\begin{aligned} a_1 &= -1.7, \quad a_2 = 0.7, \quad n = 2, \\ b_0 &= 1, \quad b_1 = 0.5, \quad m = 1, \quad d = 1, \\ \sigma &= 0.01, \quad \tau = 0.95, \quad \mu = 0. \end{aligned} \quad (5-56)$$

目标轨迹  $y_r(k)$  设置为 0.1Hz 的方波信号经过传递函数  $1/(s+1)$  滤波后的数据。初始参数估计设置为  $\theta_i^T(1) = [1, 1, 1, 1]$ ，初始协方差矩阵设置为  $P_i(1) = 100I$ 。初始控制信号设置为  $u(1) = 0.1$ 。初始贝叶斯后验概率设置为  $\{\tau_1 = 0.95, \pi_1(1) = 0.1\}$ ， $\{\tau_2 = 0.85, \pi_2(1) = 0.1\}$ ， $\{\tau_3 = 0.75, \pi_3(1) = 0.1\}$ ， $\{\tau_4 = 0.65, \pi_4(1) = 0.1\}$ ， $\{\tau_5 = 0.55, \pi_5(1) = 0.1\}$ ， $\{\tau_6 = 0.45, \pi_6(1) = 0.1\}$ ， $\{\tau_7 = 0.35, \pi_7(1) = 0.1\}$ ， $\{\tau_8 = 0.25, \pi_8(1) = 0.1\}$ ， $\{\tau_9 = 0.15, \pi_9(1) = 0.1\}$ ， $\{\tau_{10} = 0.05, \pi_{10}(1) = 0.1\}$ 。

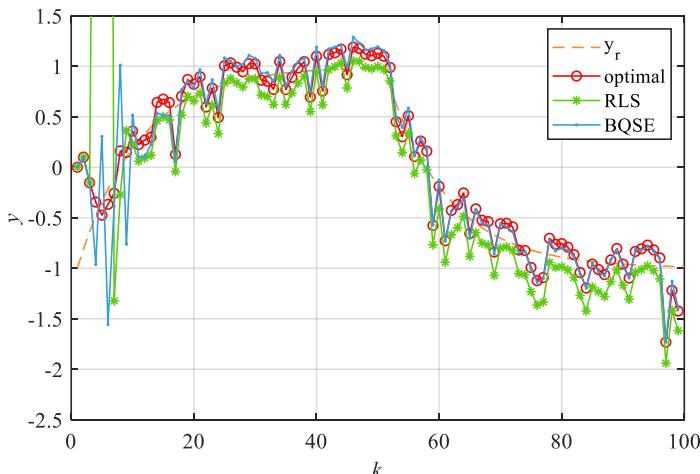


图 5-10 最优控制 (optimal)，迭代最小二乘 (RLS)，贝叶斯求和估计器 (BQSE) 对应的自适应控制下系统输出

Fig.5-10 The system output under different controllers

图 5-10 显示了系统参数假设已知的最优控制，基于 RLS 的自适应对偶控制和基于 BQSE 的自适应对偶控制下系统的输出跟踪对比。BQSE 对应的系统输出能够很好的跟踪上目标轨迹，并且能够在 20 步后收敛到最优控制下的输出轨迹。

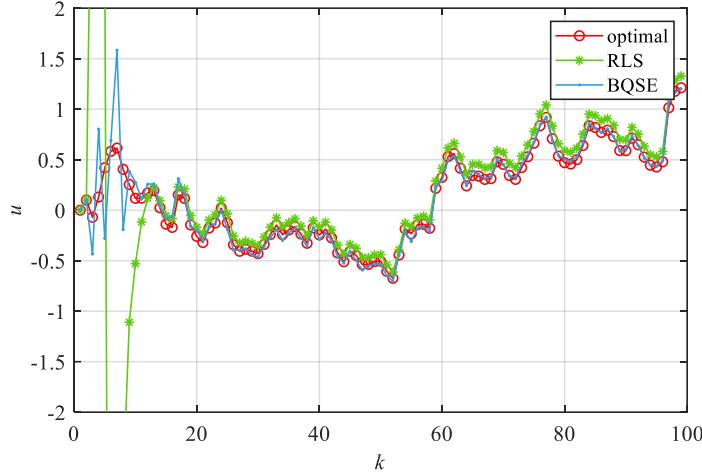


图 5-11 最优控制 (optimal)，迭代最小二乘 (RLS)，贝叶斯求和估计器 (BQSE) 对应的控制信号  
Fig.5-11 The control signal under different controllers

图 5-11 是不同控制方案下的控制信号  $u(k)$ ，从图中可以看到控制信号在 20 步以后收敛到最优控制信号。

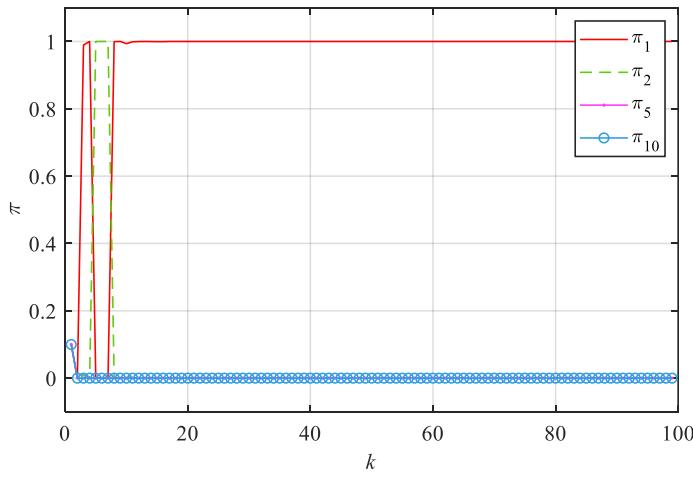


图 5-12 不同分位点  $\tau_i$  的贝叶斯后验概率  $\pi_i$   
Fig.5-12 The Bayesian posterior probability  $\pi_i$  for different  $\tau_i$

图 5-12 描绘了不同分位点下的贝叶斯后验概率  $\pi_i$  的收敛过程。分位点为  $\tau_1 = 0.95$  的后验概率  $\pi_1$  收敛到 1，其他的后验概率均收敛到 0，说明参数估计值收敛到正确的分位点下的估计值。

定义如下的性能指标来量化系统的跟踪控制性能

$$J = \frac{1}{N} \sum_{k=1}^N [y(k) - y_r(k)]^2 \quad (5-57)$$

其中  $y(k)$  和  $y_r(k)$  分别为  $k$  时刻系统的输出和目标轨迹,  $N$  是仿真长度。 $M$  次蒙特卡洛仿真实验的平均性能指标为

$$\bar{J} = \frac{1}{M} \sum_{i=1}^M J(i) \quad (5-58)$$

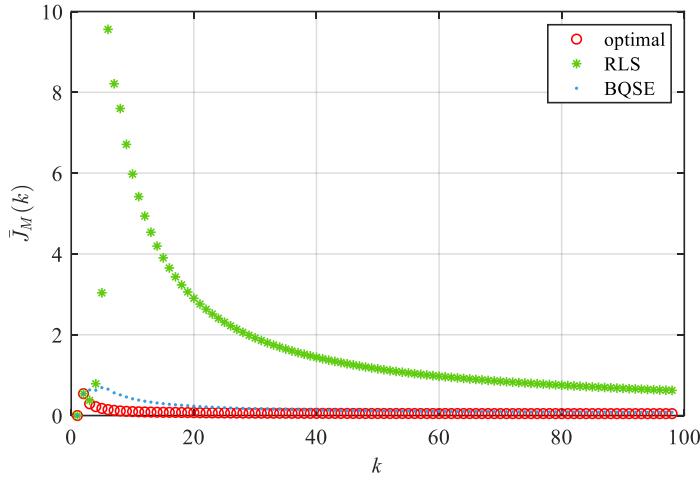


图 5-13 平均跟踪控制性能指标

Fig.5-13 The average control tracking performance index of 100 Monte Carlo simulations

图 5-13 是不同的控制方法下 100 次蒙特卡洛仿真实验的平均性能指标。从图中可得基于 BQSE 的自适应对偶控制的平均控制性能指标相比于基于 RLS 的方法更接近最优控制。

表 5-2 三种不同控制器的性能指标比较  
Tab.5-2 Comparison of three controllers

估计方法	性能指标均值
optimal	0.0460
RLS	0.9703
BQSE	0.3546

表 5-2 记录了在仿真的第 100 步, 不同的控制方案下的平均性能指标。基于 BQSE 的自适应对偶控制比 RLS 更接近最优控制的值。

### 5.4.3 非最小相位系统仿真实验

考虑非最小相位系统, 如式 (5-1) 所示, 相应的参数设置如下

$$\begin{aligned}
 a_1 &= -2, \quad a_2 = 0.7, \quad n = 2, \\
 b_0 &= 1, \quad b_1 = 2, \quad m = 1, \quad d = 1, \\
 \sigma &= 0.02, \quad \tau = 0.2, \quad \mu = 0.
 \end{aligned} \tag{5-59}$$

不同的分为点所对应的初始后验概率值设置为  $\{\tau_1 = 0.9, \pi_1(1) = 0.2\}$ ,  $\{\tau_2 = 0.7, \pi_2(1) = 0.2\}$ ,  $\{\tau_3 = 0.5, \pi_3(1) = 0.2\}$ ,  $\{\tau_4 = 0.3, \pi_4(1) = 0.2\}$ ,  $\{\tau_5 = 0.1, \pi_5(1) = 0.2\}$ 。

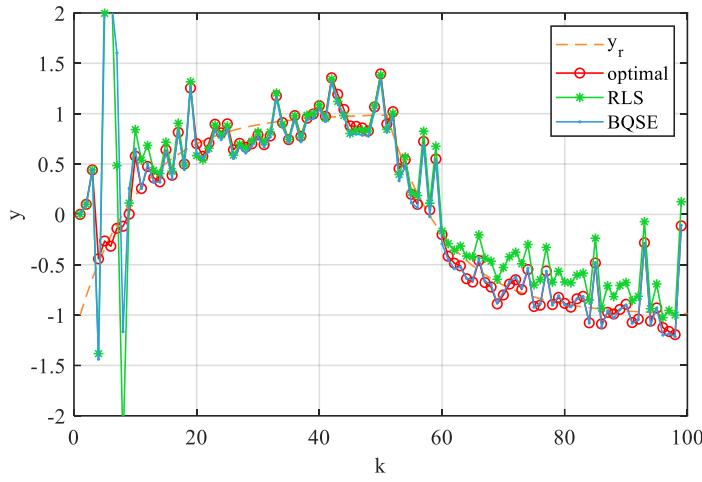


图 5-14 最优控制 (optimal) , 迭代最小二乘 (RLS) , 贝叶斯求和估计器 (BQSE) 对应的自适应控制下系统输出

Fig.5-14 The system output under different controllers

图 5-14 分别显示了在最优控制, 基于 RLS 的自适应对偶控制和基于 BQSE 的自适应对偶控制下系统的输出。基于 BQSE 的系统输出能够跟踪目标轨迹, 且在 15 步后收敛到最优控制的输出轨迹。

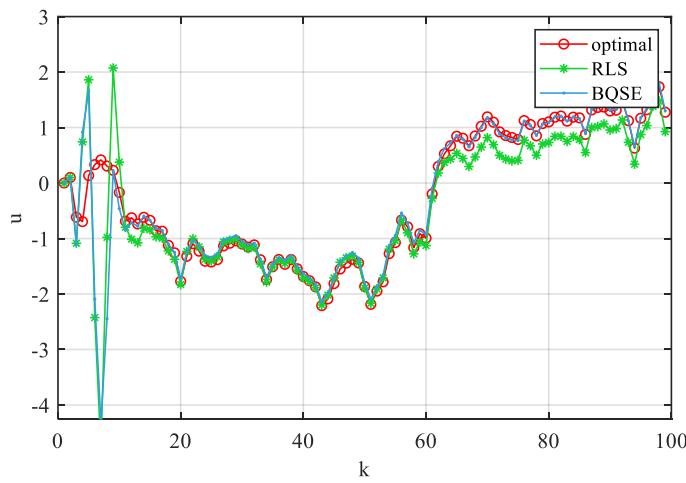


图 5-15 最优控制 (optimal) , 迭代最小二乘 (RLS) , 贝叶斯求和估计器 (BQSE) 对应的控制信号

Fig.5-15 The control signal under different controllers

图 5-15 是不同控制方案的控制信号。和 RLS 相比, 基于 BQSE 的控制信号更接近于最优控制信号。

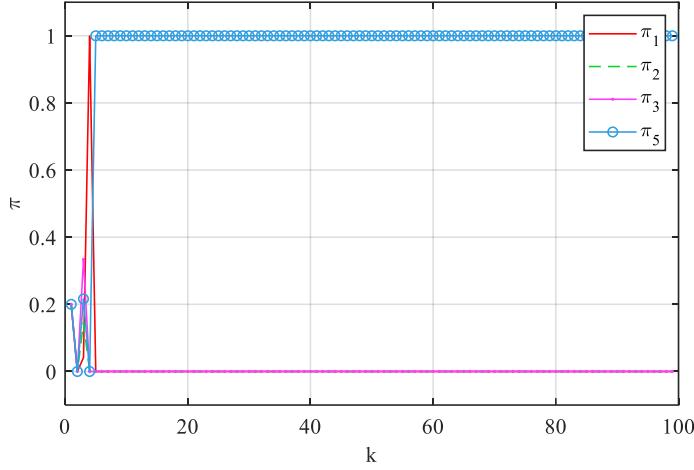
图 5-16 不同分位点  $\tau_i$  的贝叶斯后验概率  $\pi_i$ Fig.5-16 The Bayesian posterior probability  $\pi_i$  for different  $\tau_i$ 

图 5-16 显示了贝叶斯后验概率的收敛过程。分位点为  $\tau_5 = 0.1$  的后验概率  $\pi_5$  收敛到 1, 其他后验概率收敛到 0。

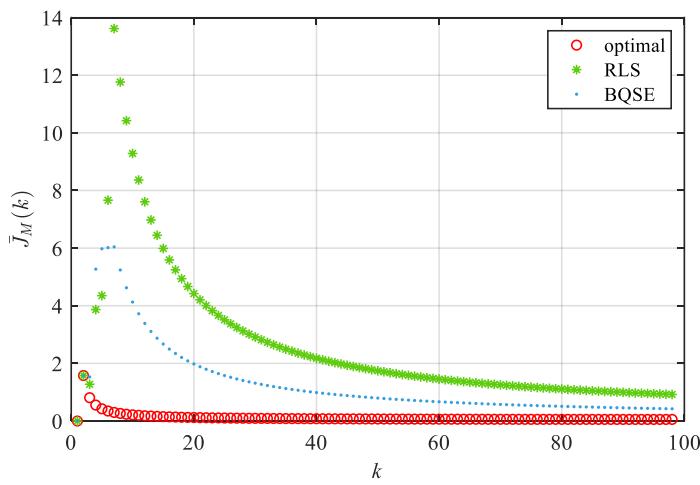


图 5-17 平均跟踪控制性能指标

Fig.5-17 The average control tracking performance index of 100 Monte Carlo simulations

图 5-17 是不同的控制方法下 100 次蒙特卡洛仿真实验的平均性能指标。从图中可得基于 BQSE 的自适应对偶控制的平均控制性能指标相比于基于 RLS 的方法更接近最优控制。

表 5-3 三种不同控制器的性能指标比较

Tab.5-3 Comparison of three controllers

估计方法	性能指标均值
optimal	0.0587
RLS	1.2979
BQSE	0.5032

表 5-3 记录了在仿真的第 100 步, 不同的控制方案下的平均性能指标。基于 BQSE 的

自适应对偶控制比基于 RLS 的自适应对偶控制更接近最优控制的值。

## 5.5 本章小结

本章设计了一种基于贝叶斯分位数求和估计器的自适应对偶控制方法，与传统的自适应对偶控制方法不同的是，该方法在控制器设计时首次考虑到被控系统具有尖峰、厚尾、非对称特征的随机噪声而不是理想高斯白噪声的情况。所设计的贝叶斯分位数求和估计器能够针对具有这种尖峰、厚尾、非对称特征的随机噪声的系统，提供更为精确的参数估计值，因为该参数估计器综合了不同分位数下的参数估计值。由贝叶斯分位数求和估计器实时估计出的系统参数，将被用于生成具有对偶特性的控制律，进一步提高了系统的控制性能。用数学模型分别对所提出的贝叶斯分位数求和估计器，以及在该估计器下的自适应对偶控制方法进行仿真实验，并将结果与最优控制和迭代最小二乘下的控制进行对比，说明了该方法的有效性。



## 6 基于可自动分配资源神经网络的自适应对偶控制

### 6.1 引言

对于完全未知的非线性系统，传统的神经网络自适应控制器的设计都是基于确定性等价原理的<sup>[151,152]</sup>，即在设计控制律时，把学习到的系统的神经网络模型视为真实模型，而没有考虑到网络建模的精度问题。这类方法在神经网络学习的初始阶段，网络模型的精度不够，也就是系统的不确定性比较大，此时使用确定性等价控制原理求解控制律，将会得到一个较大的控制信号，该控制信号会导致系统出现较大超调，从而损坏系统，因此无法在实际系统中得到应用<sup>[41]</sup>。针对这个问题，通常改进办法是利用系统历史输入输出数据，线下学习好系统的神经网络模型，以避免上述情况的发生，另外一种可行的做法是，利用自适应对偶控制原理去设计控制器<sup>[153]</sup>。

自适应对偶控制方法在设计控制器时，同时考虑到系统输出跟踪最优和系统估计最优，因此在系统中存在较大的不确定性时，控制信号不会过大，系统输出依旧能够谨慎的跟踪期望轨迹<sup>[13]</sup>。经典的自适应对偶控制方法有基于新息的自适应对偶控制<sup>[98]</sup>和双准则自适应对偶控制<sup>[106]</sup>，这两种方法都是针对的参数未知的线性系统。随后 Fabric 等人将新息对偶控制方法扩展到非线性系统中，分别使用了高斯径向基函数和多层感知器对非线性系统进行建模<sup>[40]</sup>。Simandl 等人将双准则自适应对偶控制方法用于非线性系统的控制，使用基于混合高斯分布的多层感知器对函数未知的非线性系统进行建模<sup>[41]</sup>。

然而上面提到的两种神经网络模型，在用于自适应对偶控制器之前，都需要人为设置一些网络参数。例如基于高斯径向基函数的神经网络模型，需要初始化网络神经元个数、径向基函数的中心点、宽度等，在控制过程中，仅学习网络的输出权系数。然而有些实际系统难以获取这些网络的初始参数，而不合理的参数设置会给系统带来较大的建模误差，从而降低系统的控制性能。本章设计了一种可以自动分配资源的神经网络模型，对非线性系统进行建模，该神经网络可以根据输入输出数据，自动分配网络节点，以提高网络学习的效率和精度。

信息熵是由信息论的创始人 Shannon 提出，用来衡量事件的信息量，在通信技术史的发展中有重要意义。信息熵在控制领域中有广泛的应用，例如有基于熵函数最小化的最优控制<sup>[154]</sup>、基于信息熵的自适应控制<sup>[155]</sup>、最小熵滤波等<sup>[156]</sup>。本章节将使用信息熵来描述神经网络建模过程中的信息增量，在控制律的设计中不仅最小化系统输出跟踪误差，同时还增加建模的信息增量，以在确保系统谨慎跟踪目标轨迹的同时，也能激励系统学习以减少系统的不确定性，也就是使控制律具有对偶特性。

本章针对未知非线性系统，利用可自动分配资源的神经网络模型对非线性系统进行建模，然后设计了基于信息熵的自适应对偶控制律，能够对网络模型参数进行学习的同时进行系统的跟踪控制。最后使用数学模型分别对可自动分配资源的神经网络模型对系

统建模，以及在此神经模型下基于信息熵的自适应对偶控制方法进行仿真实验，通过与最优控制与谨慎控制的结果对比，说明了该方法的有效性。

## 6.2 问题描述

本章节的控制对象是一个单输入单输出的随机非线性离散系统，该系统的非线性仿射模型表达式如下：

$$y(k) = f[x(k-1)] + g[x(k-1)]u(k-1) + e(k) \quad (6-1)$$

其中  $y(k)$  是系统输出， $u(k)$  是控制输入， $x(k)$  是系统状态向量， $x(k) = [y(k-n), \dots, y(k-1), u(k-1-m), \dots, u(t-2)]$ ，且  $m$  和  $n$  满足  $m \leq n$ ， $f[x(k-1)]$  和  $g[x(k-1)]$  是关于状态向量  $x(k-1)$  的非线性函数且未知，系统噪声为  $e(k)$ 。根据文献[41]给出如下假设：

假设 1：随机噪声  $e(k)$  是独立的，且服从均值为 0，方差为  $\sigma^2$  的高斯分布，其中方差  $\sigma^2$  是已知的。

假设 2：系统状态向量的维数  $m$  和  $n$  是已知的。

假设 3：系统为最小相位系统，且函数  $g[x(k-1)]$  不等于 0。

假设 4：参考输入  $y_r(k+1)$  是已知的，且提前一步预知。

对偶控制的控制律  $u(k)$  就是通过求解如下  $N$  步最小方差性能指标得到<sup>[40]</sup>

$$J_{dual} = E \left\{ \sum_{k=0}^{N-1} [y(k+1) - y_r(k+1)]^2 \mid I^k \right\} \quad (6-2)$$

其中  $y(k)$  是被控系统的输出， $y_r(k)$  是系统参考输入， $E\{\cdot\}$  为数学期望， $I^k$  是截止到  $k$  时刻的信息状态，定义为  $I^k = \{y(k), \dots, y(0), u(k-1), \dots, u(0)\}$ 。

理论上，以上对偶控制律可以通过动态规划对相应的贝尔曼方程进行求解获得解析解，但是在实际求解过程中，会伴随维数灾难，计算复杂，内存资源消耗大等问题。因此转向更容易求解的控制律，如确定性等价控制或者谨慎控制方案。然而这两种方案均有缺点：确定性等价控制的瞬态响应具有较大的超调；谨慎控制的响应时间太慢。Simon 和 Visakan 等人提出了针对非线性系统的次优对偶控制<sup>[40,41]</sup>，利用高斯径向基函数来拟合系统的非线性部分，且用基于新息的对偶控制方案或者双准则对偶控制方案，求解非线性系统的对偶控制的解析解。本章节的目标为，设计一种方法，一方面使用一种可自动分配资源的神经网络模型（Resource Allocating Network, RAN）来拟合系统的非线性部分，从而大大减少使用固定结构的神经网络的计算量，另一方面用交叉信息熵来衡量系统的不确定性，通过调控一步最小方差控制过程中的交叉信息熵来控制系统的丰富程度，从而平衡系统的最优跟踪控制和系统不确定性探索之间的关系，从而是控制系统具有对偶特性。

### 6.3 可自动分配资源的神经网络

可自动分配资源的神经网络是一种可以根据实时观测的数据自动分配神经元的网络模型。该网络通过分配新的资源，使学习可以在多项式级数时间内完成，解决了固定大小网络的 NP-完全问题<sup>[178]</sup>。在对非线性被控系统没有先验知识的情况下，使用固定大小的神经网络模型时是难以确定网络的大小的。如果网络规模过大，容易造成数据的过拟合；如果网络规模太小，就不能进行较好的插值。可自动分配资源的神经网络可以根据被控系统实时收集的系统状态和控制量，自动判断是否为当前的网络结构中增加神经元，从而解决网络结构的设置问题，得到更好的非线性函数的网络学习效果。

可自动分配资源的神经网络中包含一个隐含层，网络的输出是隐含层神经元的线性组合，表达式为

$$H[x(k)] = w_0 + \sum_{i=0}^K w_i \phi_i[x(k)] \quad (6-3)$$

其中  $\phi_i[x(k)]$  为隐含层第  $i$  个神经元，是关于网络输入层  $x(k)$  的响应。参数  $[w_1, \dots, w_i, \dots, w_K]$  是每个神经元的权重，代表了每个神经元的对系统模型输出的占比， $\alpha_0$  是网络的偏置项。这里神经元使用的是高斯径向基函数，表达式如下

$$\phi_i[x(k)] = \exp\left\{-\frac{1}{\sigma_i^2} \|x(k) - c_i\|^2\right\} \quad (6-4)$$

其中  $c_i$  是第  $i$  个高斯基函数的中心， $\sigma_i^2$  代表了高斯基函数的宽度。

可自动分配资源的神经网络的初始结构中没有隐含层节点，仅初始化网络偏置项参数为  $\alpha_0 = y(1)$ ，其中  $y(1)$  为系统状态为  $x(0)$  时首次观测到的系统输出。根据后续实时观测到的系统状态和输出数据对  $\{x(k), y(k+1)\}$ ，该网络选取其中的数据逐步增加隐含层的节点。数据对的选取标准有两条，分别为

$$\|x(k) - c_{\text{nearest}}\| > \delta_c(k) \quad (6-5)$$

$$\|y(k+1) - H[x(k)]\| > \delta_e(k) \quad (6-6)$$

其中  $c_{\text{nearest}}$  是距离当前状态  $x(k)$  最近的一个基函数中心点， $\delta_c(k)$  和  $\delta_e(k)$  为阈值。式 (6-5) 的条件表明，被选择的数据对中的  $x(k)$  需要离最近的基函数中心  $c_{\text{nearest}}$  足够远，即两者之间的欧氏距离要大于阈值  $\delta_c(k)$ ，其中  $\delta_c(k)$  就是中心点之间的最小间隔。阈值  $\delta_c(k)$  的取值并不是固定的，初始值为  $\delta_c(k) = \delta_{\max}$ ，随后乘以一个衰减系数，让阈值变小直至最小值  $\delta_{\min}$ ，具体表达为  $\delta_c(k) = \max\{\gamma^k \delta_{\max}, \delta_{\min}\}$ ，其中衰减系数  $0 < \gamma < 1$ ，阈值  $\delta_c(k)$  最终衰减到最小值  $\delta_{\min}$ 。式 (6-6) 表明，被选择的数据对中的  $y(k+1)$  与该神经网络预测的系统输出  $f[x(k)]$  之间的误差需要超过阈值  $\delta_e(k)$ ，其中  $\delta_e(k)$  就是期望的神经网络输出的最小误差。

当观测的数据对  $\{x(k), y(k+1)\}$  满足式 (6-5) 和 (6-6) 两个条件, 即可增加一个隐含层节点, 将状态  $x(k)$  作为相应的高斯基函数的中心  $c_i$ , 即

$$c_i = x(k) \quad (6-7)$$

高斯基函数的宽度的设置是基于当前状态  $x(k)$  与最靠近的中心点  $c_i$  之间的距离的, 设置为

$$\sigma_i = \lambda \|x(k) - c_{\text{nearest}}\| \quad (6-8)$$

该高斯基函数所对应的权重  $w_i$  初值设置为

$$w_i = \|y(k+1) - H[x(k)]\| \quad (6-9)$$

当实时观测的数据对  $\{x(k), y(k+1)\}$  不满足式 (6-5) 和 (6-6) 两个条件, 则该数据对不用于增加神经网络隐含层节点, 仅用于网络参数的调整更新。

## 6.4 网络模型参数估计

根据式 (6-1) 所描述的非线性仿射系统可以用高斯径向基函数进行拟合

$$\begin{aligned} H[x(k)] &= f[x(k)] + g[x(k)]u(k) \\ &= w_f^{*T}(k)\phi_f[x(k)] + w_g^{*T}(k)\phi_g[x(k)]u(k) \\ &= w^{*T}(k)\phi[x(k)] \end{aligned} \quad (6-10)$$

其中网络权重参数为

$$\begin{aligned} w^{*T}(k) &= [w_f^*(k); w_g^*(k)]^T \\ &= [w_{f_0}^*(k); w_{f_1}^*(k); \dots; w_{g_0}^*(k); w_{g_1}^*(k); \dots]^T \end{aligned} \quad (6-11)$$

高斯基函数  $\phi[x(k)]$  为

$$\phi[x(k)] = [\phi_{f_0}[x(k)]; \phi_{f_1}[x(k)]; \dots; \phi_{g_0}[x(k)]u(k); \phi_{g_1}[x(k)]u(k); \dots] \quad (6-12)$$

参照上一小节, 将自动分配资源的神经网络用于该非线性系统的在线学习。首先用观测的系统第一对输入输出数据  $\{u(0), y(1)\}$ , 初始化该网络的结构和参数

$$\hat{w}_{f_0}(1) = \hat{w}_{g_0}(1) = y(1)/[u(0)+1], \quad \phi_{f_0}[x(k)] = \phi_{g_0}[x(k)] = 1 \quad (6-13)$$

第二对数据  $\{u(1), y(2)\}$  用于构建网络隐含层的第一个节点  $\phi_{f_1}[x(k)]$  和  $\phi_{g_1}[x(k)]$ , 对应的高斯基函数的参数设置为

$$\begin{aligned} \hat{w}_{f_1}(2) &= \hat{w}_{g_1}(2) = \alpha \|y(2) - H[x(1)]\| \\ c_{f_1} &= c_{g_1} = x(1) \\ \sigma_{f_1} &= \sigma_{g_1} = \lambda \|x(1) - x(0)\| \end{aligned} \quad (6-14)$$

在后续实时观测的数据对中, 如果该数据  $\{x(k), y(k+1)\}$  满足如下两个条件式 (6-5) 以及式 (6-6) 这两个判据, 则该数据对可用于增加一个隐含层节点  $\phi_{f_i}[x(k)]$  和  $\phi_{g_i}[x(k)]$ ,

对应的高斯基函数的参数设置为

$$\begin{aligned}\hat{w}_{f_i}(k+1) &= \hat{w}_{g_i}(k+1) = \alpha \|y(k+1) - H[x(k)]\| \\ c_{f_i} &= c_{g_i} = x(k) \\ \sigma_{f_i} &= \sigma_{g_i} = \lambda \|x(k) - c_{nearest}\|\end{aligned}\quad (6-15)$$

如果该数据  $\{x(k), y(k+1)\}$  满足如下两个条件式 (6-5) 以及式 (6-6) 这两个判据，则该数据用于调整网络的权值参数  $w^T(k+1)$  以及中心位置  $c$ 。这里用扩展卡尔曼滤波 (Extended Kalman Filter, EKF) 来进行参数学习，相比于最小均方算法(Least Mean Square, LMS)，EKF 提高了网络的收敛率并且减少了网络的复杂程度。

定义网络参数为

$$\theta^T(k) = [w_{f_0}(k); w_{f_1}(k); \dots; c_{f_0}(k); c_{f_1}(k); \dots; w_{g_0}(k); w_{g_1}(k); \dots; c_{g_0}(k); c_{g_1}(k); \dots]^T \quad (6-16)$$

网络函数  $H[x(k)]$  对参数向量  $\theta(k)$  的梯度为  $\phi(k) = \nabla_\theta H[x(k)]$

$$\phi^T(k) = \left[ \begin{array}{l} 1; \phi_{f_1}[x(k)]; \dots; \phi_{f_1}[x(k)] \frac{2w_{f_1}}{\sigma_{f_1}^2} [x(k) - c_{f_1}(k)]^T; \dots; u(k); \phi_{g_1}[x(k)]u(k); \\ \dots; \phi_{g_1}[x(k)]u(k) \frac{2w_{g_1}}{\sigma_{g_1}^2} [x(k) - c_{g_1}(k)]^T; \dots \end{array} \right]^T \quad (6-17)$$

扩展卡尔曼滤波进行参数学习的迭代过程如下

$$\hat{\theta}(k+1) = \hat{\theta}(k) + K(k+1) \{y(k+1) - \hat{w}^T(k) \phi[x(k)]\} \quad (6-18)$$

$$K(k+1) = [R + \phi^T(k)P(k)\phi(k)]^{-1} P(k)\phi(k) \quad (6-19)$$

$$P(k+1) = [I - K(k+1)\phi^T(k)]P(k) + QI \quad (6-20)$$

式 (6-19) 中  $R$  是观测噪声的方差，由于卡尔曼滤波会使得参数快速收敛，以至于在后期抑制了参数的实时调整，式 (6-20) 协方差矩阵的更新中增加了随机游走模块  $QI$  以解决这个问题。新的误差协方差矩阵可以写成

$$P(k+1) = \begin{bmatrix} P_{w_f}(k) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & P_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & P_{c_f}(k) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & P_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & P_{w_g}(k) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & P_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & P_{c_g}(k) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & P_0 \end{bmatrix} \quad (6-21)$$

其中  $P_0$  为新增的参数误差协方差矩阵的元素，其值代表了参数估计的不确定性。由于在可自动分配资源的神经网络中，隐含层的神经元高斯基函数是实时自动分配的，所以网

络参数  $\theta$  也是变化的，相应的参数估计误差协方差矩阵  $P$  的维数也跟随调整。例如增加一个新的隐层节点，则增加四个新的参数，相应的协方差矩阵增的维数也增加四个。

## 6.5 基于信息熵的自适应对偶控制

本小节介绍了所提出的基于信息熵的自适应对偶控制。首先定义系统的条件熵，由于系统中包含未知的随机参数  $\theta$ ，所以系统的条件熵可以写成参数  $\theta$  的表达式

$$H[\theta | I_k] = -E[\log p(\theta | I_k)] \quad (6-22)$$

其中  $p(\theta | I_k)$  是给定信息状态  $I_k$  时， $\theta$  的条件概率密度，也可称为后验概率。 $k$  时刻的信息状态  $I_k$  定义为

$$I_k = \{u(1), \dots, u(k-1), y(1), \dots, y(k)\} \quad (6-23)$$

参数  $\theta$  的先验概率定义为  $p(\theta)$ ，则相应的熵为

$$H[\theta] = -E[\log p(\theta)] \quad (6-24)$$

那么从无信息状态  $I_k$  到有信息状态  $I_k$  下，给参数  $\theta$  所带来的信息增量(Expected Information Gain, EIG)，可以用其后验概率的熵减去先验概率的熵来衡量，表达式为

$$EIG[\theta | I_k] = E[\log p(\theta | I_k)] - E[\log p(\theta)] \quad (6-25)$$

该式同理描述的是系统参数  $\theta$  不确定性的减少量。因此，可以通过最大化  $EIG[\theta | I_k]$ ，使系统获取更多的未知信息，减少系统的不确定性，从而使估计参数  $\hat{\theta}$  更接近未知参数  $\theta$  的真值。

为了使系统得到一个较好的输出跟踪效果，可以使用一步超前最小方差控制，其代价函数为

$$J(k+1) = E\{[y(k+1) - y_r(k+1)]^2 | I_k\} \quad (6-26)$$

对偶控制律  $u^*(k)$  不仅有减少系统的不确定性，提高参数估计的质量的特点，同时还具有使系统输出跟踪目标轨迹的特点，因此，需要同时满足最大化信息增量  $EIG[\theta | I_{k+1}]$  和最小化跟踪误差  $J(k+1)$ ，即可变为一个双目标优化问题

$$\min_{u(k)} \{J(k+1), -EIG[\theta | I_{k+1}]\} \quad (6-27)$$

利用效用函数可以将其转化为单目标优化问题，再求解对偶控制律  $u^*(k)$ 。

$$\min_{u(k)} \{J_\mu(k+1)\} = \min_{u(k)} \{J(k+1) - \mu(k+1)EIG[\theta | I_{k+1}]\} \quad (6-28)$$

其中对偶控制的代价函数  $J_\mu(k+1)$  的计算结果为

$$\begin{aligned} J_\mu(k+1) &= \varphi^T(k)P(k)\varphi(k) + \sigma^2 + [\hat{w}^T(k)\phi(k) - y_r(k+1)]^2 \\ &\quad - \mu(k) \log(1 + \varphi^T(k)P(k)\varphi(k)) \end{aligned} \quad (6-29)$$

其计算过程如下所示，由于假设给定信息状态  $I_{k+1}$ ，参数  $\theta$  的条件分布为正态分布，均值为  $\hat{\theta}(k+1)$ ，方差为  $P(k+1)$ 。根据上一小节中的卡尔曼滤波，误差协方差矩阵定义为  $P(k+1) = E\{[\theta(k+1) - \hat{\theta}(k+1)]^T [\theta(k+1) - \hat{\theta}(k+1)]\}$ 。信息增量中的  $E[\log p(\theta | I_{k+1})]$  可以写成

$$\begin{aligned}
 E[\log p(\theta | I_{k+1})] &= E\left[\log \frac{1}{(2\pi)^{\frac{p}{2}} |P(k+1)|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} [\theta(k+1) - \hat{\theta}(k+1)]^T P^{-1}(k+1) [\theta(k+1) - \hat{\theta}(k+1)]\right)\right] \\
 &= -\frac{1}{2} \log[(2\pi)^p |P(k+1)|] - \frac{1}{2} E\{[\theta(k+1) - \hat{\theta}(k+1)]^T P^{-1}(k+1) \\
 &\quad [\theta(k+1) - \hat{\theta}(k+1)]\} (\log e) \\
 &= -\frac{1}{2} \log[(2\pi)^p |P(k+1)|] - \frac{1}{2} \log e \\
 &= -\frac{1}{2} \log[(2\pi)^p e |P(k+1)|]
 \end{aligned} \tag{6-30}$$

由于此时先验概率  $p(\theta)$  为上个时刻的后验概率  $\log p(\theta | I_k)$ ，因此同理可得  $E[\log p(\theta)] = -\frac{1}{2} \log[(2\pi)^p e |P(k)|]$ 。因此信息增量  $EIG[\theta | I_{k+1}]$  为

$$\begin{aligned}
 EIG[\theta | I_{k+1}] &= E[\log p(\theta | I_{k+1})] - E[\log p(\theta)] \\
 &= \frac{1}{2} \log[2\pi e |P(k+1)|] - \frac{1}{2} \log[2\pi e |P(k)|] \\
 &= \frac{1}{2} \log |P(k+1)| - \frac{1}{2} \log |P(k)|
 \end{aligned} \tag{6-31}$$

根据式 (6-20)，可将式  $P(k+1) = \frac{P(k)}{1 + \varphi^T(k)P(k)\varphi(k)}$  代入上式中得

$$\begin{aligned}
 EIG[\theta | I_k] &= \frac{1}{2} \log |P(k)| - \frac{1}{2} \log \left| \frac{P(k)}{1 + \varphi^T(k)P(k)\varphi(k)} \right| \\
 &= \frac{1}{2} \log (1 + \varphi^T(k)P(k)\varphi(k))
 \end{aligned} \tag{6-32}$$

根据式 (6-1) 对系统的描述，以及式 (6-10) 神经网络模型的定义，这里用一阶泰勒展开对系统输出  $y(k+1)$  关于估计参数  $\hat{w}(k)$  进行近似计算

$$y(k+1) \approx \hat{w}^T(k)\phi(k) + \varphi^T(k)[\theta(k) - \hat{\theta}(k)] + e(k+1) \tag{6-33}$$

将上式带入到代价函数  $J(k)$  中进行计算得

$$\begin{aligned}
 J(k) &= E\{[y(k+1) - y_r(k+1)]^2 | I_k\} \\
 &= E\{[\hat{w}^T(k)\phi(k) + \varphi^T(k)[\theta(k) - \hat{\theta}(k)] + e(k+1)]^2 | I_k\} \\
 &= \varphi^T(k)P(k)\varphi(k) + \sigma^2 + [\hat{w}^T(k)\phi(k) - y_r(k+1)]^2
 \end{aligned} \tag{6-34}$$

将式 (6-32) 以及式 (6-34) 代入到代价函数  $J_\mu(k+1)$  中，即可得式 (6-29)。由于

$u(k)$  隐藏在表达式 (6-29) 中, 因此无法求解出最值, 因此将观测向量  $\phi(k)$  和协方差矩阵  $P(k)$  进行分块处理, 使  $u(k)$  在表达式中显现出来便于求解

$$\begin{aligned}\phi^T(k) &= \begin{bmatrix} \phi_f^T[x(k)] & : & \nabla_{c_f} \phi_f^T[x(k)] & : & \phi_g^T[x(k)]u(k) & : & \nabla_{c_g} \phi_g^T[x(k)]u(k) \end{bmatrix} \\ &= \begin{bmatrix} \phi_f^T(k) & : & \phi_g^T(k)u(k) \end{bmatrix}\end{aligned}\quad (6-35)$$

$$P(k) = \begin{bmatrix} P_f(k) & : & P_{gf}^T(k) \\ \cdots & \cdots & \cdots \\ P_{gf}(k) & : & P_g(k) \end{bmatrix}\quad (6-36)$$

将上式 (6-35) 和 (6-36) 代入  $\phi^T(k)P(k)\phi(k)$  中得

$$\phi^T(k)P(k)\phi(k) = v_{ff} + 2v_{gf}u(k) + v_{gg}u^2(k)\quad (6-37)$$

其中  $v_{ff} = \phi_f^T(k)P_f(k)\phi_f(k)$ ,  $v_{gf} = \phi_g^T(k)P_{gf}(k)\phi_f(k)$ ,  $v_{gg} = \phi_g^T(k)P_g(k)\phi_g(k)$ 。将  $\hat{w}^T(k) = [\hat{w}_f(k); \hat{w}_g(k)]^T$  和  $\phi(k) = [\phi_f(k); \phi_g(k)u(k)]$  代入  $[\hat{w}^T(k)\phi(k) - y_r(k+1)]^2$  得

$$[\hat{w}^T(k)\phi(k) - y_r(k+1)]^2 = [\hat{h}_f - y_r(k+1)]^2 + 2[\hat{h}_f - y_r(k+1)]\hat{h}_g u(k) + \hat{h}_g^2 u^2(k)\quad (6-38)$$

其中  $\hat{h}_f = \hat{w}_f^T(k)\phi_f(k)$ ,  $\hat{h}_g = \hat{w}_g^T(k)\phi_g(k)$ 。

将式 (6-37) 和式 (6-38) 代入代价函数  $J_\mu(k)$  中,  $J_\mu(k)$  可以写成如下包含  $u(k)$  的表达式为

$$\begin{aligned}J_\mu(k) &= v_{ff} + 2v_{gf}u(k) + v_{gg}u^2(k) + \sigma^2 + [\hat{h}_f - y_r(k+1)]^2 \\ &\quad + 2[\hat{h}_f - y_r(k+1)]\hat{h}_g u(k) + \hat{h}_g^2 u^2(k) \\ &\quad - \mu(k) \log(1 + v_{ff} + 2v_{gf}u(k) + v_{gg}u^2(k))\end{aligned}\quad (6-39)$$

代价函数  $J_\mu(k)$  关于  $u(k)$  的一阶导数为

$$\begin{aligned}\frac{\partial J_\mu(k)}{\partial u(k)} &= 2v_{gf} + 2v_{gg}u(k) + 2[\hat{h}_f - y_r(k+1)]\hat{h}_g + 2\hat{h}_g^2 u(k) \\ &\quad - \mu(k) \frac{2v_{gf} + 2v_{gg}u(k)}{1 + v_{ff} + 2v_{gf}u(k) + v_{gg}u^2(k)}\end{aligned}\quad (6-40)$$

令  $\partial J_\mu(k)/\partial u(k) = 0$ , 通过化简可以得到如下关于  $u(k)$  的三次多项式方程

$$A(k)u^3(k) + B(k)u^2(k) + C(k)u(k) + D(k) = 0\quad (6-41)$$

其中参数  $A(k)$ ,  $B(k)$ ,  $C(k)$  和  $D(k)$  分别为

$$\begin{aligned}A(k) &= v_{gg}(v_{gg} + \hat{h}_g^2) \\ B(k) &= 2v_{gf}(v_{gg} + \hat{h}_g^2) + v_{gg}(v_{gf} + [\hat{h}_f - y_r(k+1)]\hat{h}_g) \\ C(k) &= (v_{gg} + \hat{h}_g^2)(1 + v_{ff}) + 2v_{gf}(v_{gf} + [\hat{h}_f - y_r(k+1)]\hat{h}_g) - \mu(k)v_{gg} \\ D(k) &= (v_{gf} + [\hat{h}_f - y_r(k+1)]\hat{h}_g)(1 + v_{ff}) - \mu(k)v_{gf}\end{aligned}\quad (6-42)$$

方程 (6-41) 的根即为对偶控制律  $u^*(k)$ ，为保证式 (6-41) 有唯一解，代价函数  $J_\mu(k)$  的关于  $u(k)$  的二阶偏导数需  $\partial^2 J_\mu(k)/\partial u(k)^2 \geq 0$ ，此时代价函数是关于  $u(k)$  的凸函数，其最小值存在且唯一，下面对该条件进行分析

$$\frac{\partial^2 J_\mu(k)}{\partial u(k)^2} = 2v_{gg} + 2\hat{h}_g^2 - \mu(k) \left( \frac{v_{gg}}{1 + \varphi^T(k)P(k)\varphi(k)} - \frac{2[v_{gf} + v_{gg}u(k)]^2}{[1 + \varphi^T(k)P(k)\varphi(k)]^2} \right) \quad (6-43)$$

将上式等号右侧乘以  $[1 + \varphi^T(k)P(k)\varphi(k)]^2$  得

$$2\hat{h}_g^2[1 + \varphi^T(k)P(k)\varphi(k)]^2 + 2v_{gg}[1 + \varphi^T(k)P(k)\varphi(k)][1 + \varphi^T(k)P(k)\varphi(k) - \mu(k)] + 2\mu(k)[v_{gf} + v_{gg}u(k)]^2 \quad (6-44)$$

式 (6-44) 的第一部分是正的。当  $1 + \varphi^T(k)P(k)\varphi(k) - \mu(k) \geq 0$ ，该式的第二部分也是正的。当  $\mu(k) \geq 0$  时，该式的第三部分也是正的。因此在选择参数  $\mu(k)$  时，只要符合条件  $0 \leq \mu(k) \leq 1 + \varphi^T(k)P(k)\varphi(k)$ ，即可保证对偶控制律  $u^*(k)$  有唯一解。

为便于对偶控制律的结构进行分析，定义  $\rho(k) = \frac{\mu(k)}{1 + \varphi^T(k)P(k)\varphi(k)}$ ，将其代入式 (6-40) 得

$$\begin{aligned} \frac{\partial J_\mu(k)}{\partial u(k)} &= 2v_{gf} + 2v_{gg}u(k) + 2[\hat{h}_f - y_r(k+1)]\hat{h}_g + 2\hat{h}_g^2u(k) \\ &\quad - \rho(k)(2v_{gf} + 2v_{gg}u(k)) \end{aligned} \quad (6-45)$$

令  $\partial J_\mu(k)/\partial u(k) = 0$ ，可得对偶控制律  $u^*(k)$  的表达式为

$$u^*(k) = -\frac{[1 - \rho(k)]v_{gf}(k) + [\hat{h}_f - y_r(k+1)]\hat{h}_g}{[1 - \rho(k)]v_g(k) + \hat{h}_g^2} \quad (6-46)$$

可以将式 (6-46) 所示的对偶控制律与经典的次优自适应对偶控制律进行对比分析。当  $\rho(k) = 1$  时，对偶控制就变成了熟知的自校正控制。当  $\rho(k) = 0$ ，对偶控制就变成了谨慎控制。根据定义， $\rho(k)$  中包含控制律  $u(k)$ ， $u(k)$  需要通过式 (6-46) 进行计算，因此无法得出  $\rho(k)$  的取值。根据条件  $0 \leq \mu(k) \leq 1 + \varphi^T(k)P(k)\varphi(k)$  以及  $\rho(k)$  的定义，可得该条件等价于条件  $0 \leq \rho(k) \leq 1$ ，因此  $\rho(k)$  可以在区间中任意取值。这就表明对偶控制律就是处于自校正控制和谨慎控制之间的控制律。Milito 的基于新息的对偶控制方法给出了类似的结论和类似形式的控制律。本章用基于信息熵的理论，给出了比基于新息的对偶控制更广义的形式。

## 6.6 仿真实验

### 6.6.1 可自动分配资源的神经网络建模分析

本实验用可自动分配资源的神经网络对 Hermite 多项式函数进行拟合，Hermite 多项

式的表达式为

$$f(x) = 1.1(1-x+2x^2)\exp\left\{-\frac{1}{2}x^2\right\} \quad (6-47)$$

其中  $x \in R$ 。  $x$  在区间  $[-4, +4]$  进行采样得到输出数据  $f(x)$ ，收集 200 组输入输出序列数据样本点，对序列样本数据进行自动可分配资源神经网络参数进行实时拟合。该实验的参数设置为  $\delta_c(k)=1$ ,  $\delta_e(k)=0.01$ ,  $\alpha=1$ ,  $\lambda=0.8$ ,  $P(1)=100$ ,  $Q=0.1$ 。

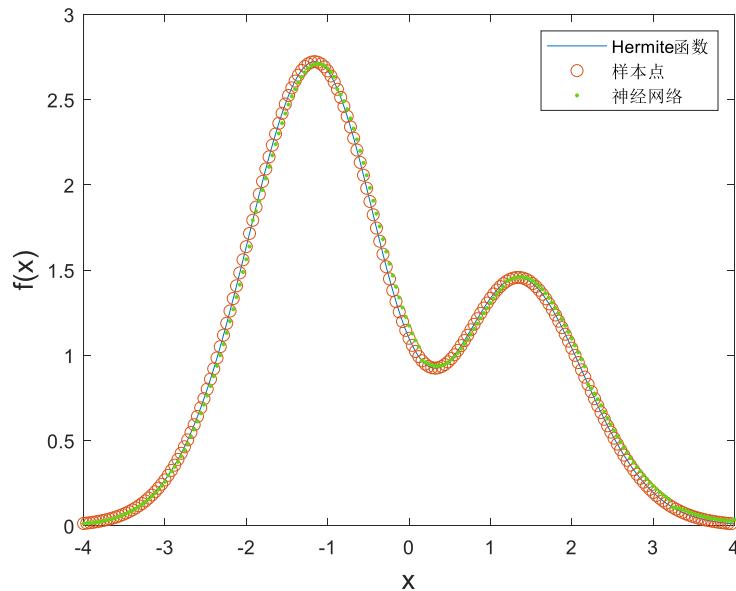


图 6-1 Hermite 函数和自动分配资源的神经网络的实时拟合输出  
Fig.6-1 Hermite function and the resource auto-allocating neural network approximation

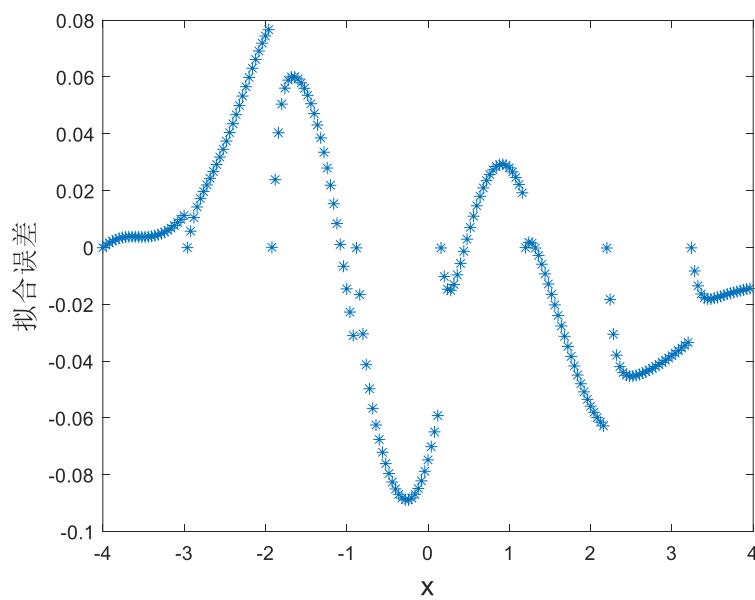


图 6-2 自动分配资源的神经网络对 Hermite 函数的拟合误差  
Fig.6-2 The approximation error of resource auto-allocating neural network approximation

图 6-1 描述的是可自动分配资源的神经网络对 Hermite 函数实时拟合的输出效果。其中蓝色实线是 Hermite 函数的输出曲线，橙色圈是用于神经网络学习的数据，绿色点是利用实时数据进行神经网络学习得到的网络输出数据。从图中明显可以看到神经网络的输出与实际 Hermite 函数输出数据很接近。

图 6-2 是实时学习的网络输出与真实 Hermite 函数的输出之间的误差，拟合误差范围在 -0.1 到 0.1 之间，完全符合拟合误差要求。

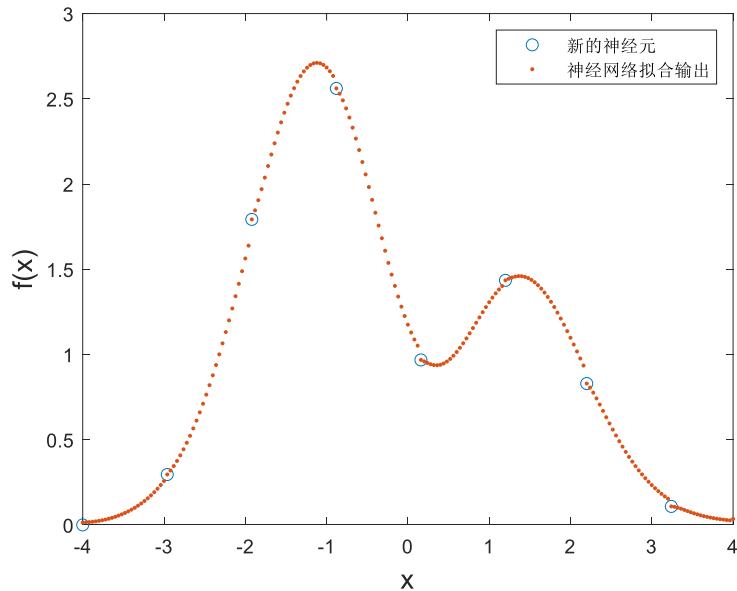


图 6-3 自动分配资源的神经网络拟合过程中的新增隐层节点  
Fig6-3 The new allocated neural unit in hidden layer

图 6-3 中橙色点是神经网络的输出，蓝色圈表明神经网络在此处自动增加了一个神经元。

表 6-1 参数  $\delta_c$  和  $\delta_e$  取不同值时自动分配资源的神经网络的隐层节点数  
Tab.6-1 The number of neural units with different  $\delta_c$  and  $\delta_e$

隐层节点数	$\delta_c$					
	0.20	0.30	0.50	0.75	1.00	
$\delta_e$	0.01	34	24	15	11	8
	0.03	29	21	14	10	8
	0.05	23	17	11	8	7
	0.07	18	13	9	7	6
	0.09	11	10	8	7	6

表 6-1 是当参数  $\delta_c$  和  $\delta_e$  分别取不同的值时，可自动分配资源的神经网络的隐层节点数。由表可以得出，随着参数  $\delta_c$  和  $\delta_e$  的增大，网络分配的隐含层的节点数减少，那么对

应的神经网络的拟合精度就会降低。

表 6-2 参数  $\delta_c$  和  $\delta_e$  取不同值时自动分配资源的神经网络的均方根误差Tab.6-2 The RMSE with different  $\delta_c$  and  $\delta_e$ 

RMSE	$\delta_c$				
	0.20	0.30	0.50	0.75	1.00
$\delta_e$	0.01	0.0176	0.0230	0.0298	0.0353
	0.03	0.0192	0.0242	0.0309	0.0360
	0.05	0.0222	0.0272	0.0341	0.0392
	0.07	0.0244	0.0293	0.0351	0.0395
	0.09	0.0303	0.0321	0.0361	0.0434

表 6-2 是当参数  $\delta_c$  和  $\delta_e$  分别取不同的值时，可自动分配资源的神经网络建模的均方根误差。从表中数据可得，与期望的情况一致，随着参数  $\delta_c$  和  $\delta_e$  的增大，自动分配资源的神经网络的拟合均方根误差也增大。

### 6.6.2 基于信息熵的自适应对偶控制实验分析

考虑如下非线性系统

$$y(k+1) = \sin(x(k)) + \cos(3x(k)) + (2 + \cos(x(t)))u(k) + e(k+1) \quad (6-48)$$

其中系统状态为  $x(t) = y(t)$ ，噪声为均值为 0，方差为  $\sigma^2 = 0.001$  的高斯白噪声。未知非线性函数为  $f(x) = \sin(x(k)) + \cos(3x(k))$ ， $g(x) = 2 + \cos(x(t))$ 。期望输出轨迹  $y_r(k)$  是 0.1Hz 的方波信号经过传递函数  $1/(s+1)$  滤波后的信号。自适应对偶控制和可自动分配资源的神经网络的系数分别设置为  $\delta_c(k) = 1$ ， $\delta_e(k) = 0.01$ ， $\alpha = 1$ ， $\lambda = 0.8$ ， $P(1) = 100$ ， $Q = 0.1$ ， $\rho(k) = 0.85$ 。

图 6-4 对比了基于神经网络模型的不同控制方法下，系统输出与参考信号的对比图。从图中明显可以看到，在系统起步阶段，基于信息熵的自适应对偶控制方法下的控制输出明显比基于确定性等价原理的系统输出的超调小，比谨慎控制下的系统输出的调节时间短。

图 6-5 是不同控制方案下的系统跟踪输出误差，其中自适应对偶控制方法下的跟踪误差最少。

通过 500 次蒙特卡洛实验，得到了每次实验的累积代价。在图 6-6 中分别显示了三种控制方法下每次实验的累积代价价值。从图中明显可以看到确定性等价控制的累积代价价值最大，而对偶控制的累积代价价值最小。

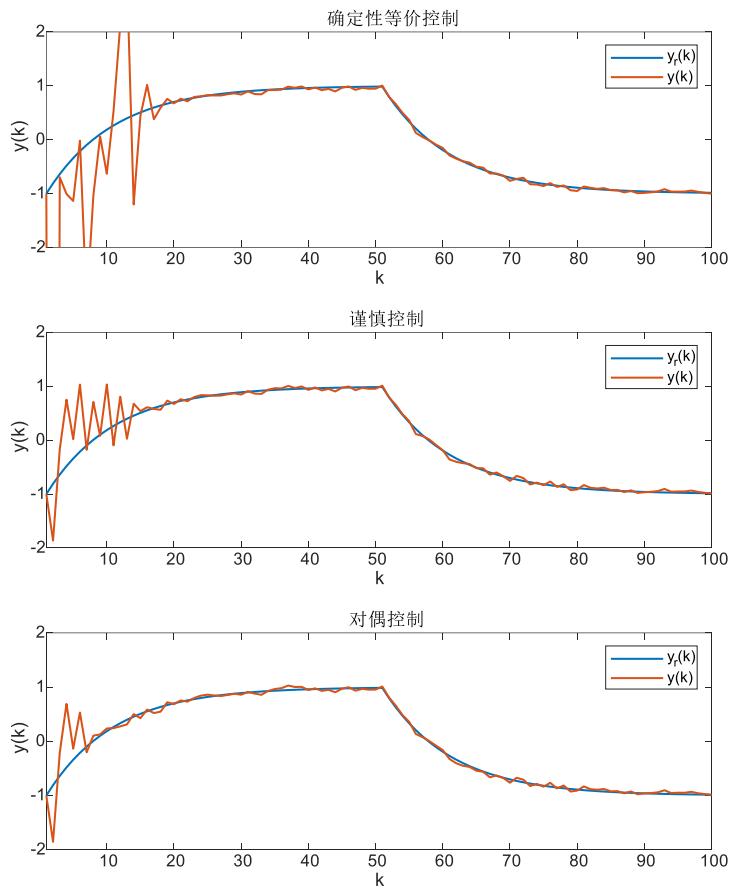


图 6-4 确定性等价控制，谨慎控制，自适应对偶控制的输出跟踪效果图

Fig.6-4 The system output tracking performance for certainty equivalence based control, cautious control and adaptive dual control

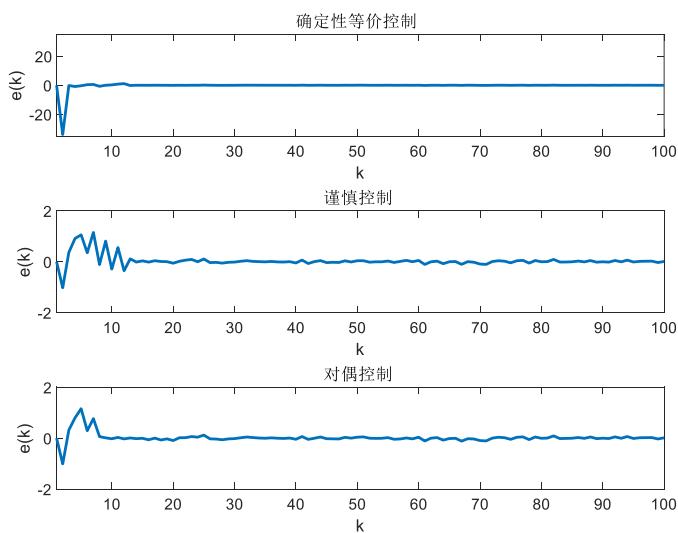


图 6-5 确定性等价控制，谨慎控制，自适应对偶控制的输出跟踪误差

Fig.6-5 The system output tracking error for certainty equivalence based control, cautious control and adaptive dual control

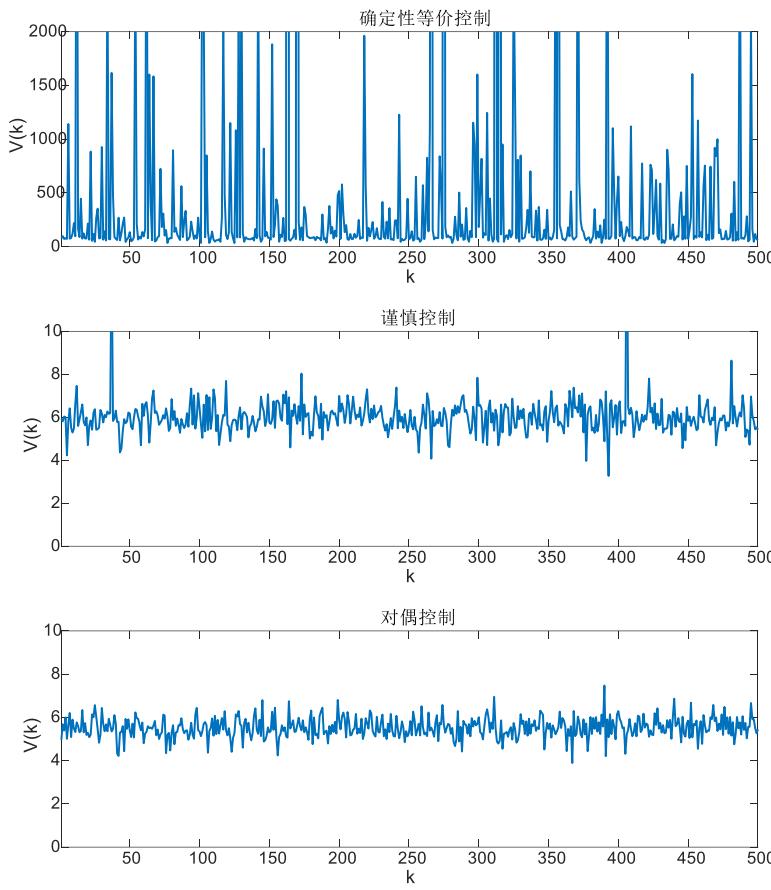


图 6-6 确定性等价控制, 谨慎控制, 自适应对偶控制的累计代价

Fig.6-6 The accumulated costs for certainty equivalence based control, cautious control and adaptive dual control

分别计算这三种控制方案下 500 次蒙特卡洛实验的平均累积代价，并列入表 6-3 中。

表 6-3 确定性等价控制, 谨慎控制, 自适应对偶控制的平均性能指标

Tab.6-4 The average performance index for certainty equivalence based control, cautious control and adaptive dual control

	确定性等价控制	谨慎控制	对偶控制
平均性能指标	862.0107	6.0160	5.5317

从表 6-3 中可以看到，确定性等价控制的平均累积代价值为 862.0107，谨慎控制的平均累积代价值为 6.0160，对偶控制的平均累积代价值为 5.5317。由此可以得出结论，无论是系统在初始起步阶段的瞬时响应，还是 500 次蒙特卡洛统计实验的平均累积代价值，自适应对偶控制相比确定性等价控制和谨慎控制都表现出更优的控制性能。

## 6.7 本章小结

本章针对未知非线性系统，提出一种基于信息熵的自适应对偶控制方法。首先设计了一种可自动分配资源的神经网络对未知系统进行建模，与其他网络不同的是，该网络可以根据实时输入输出数据调整网络节点和参数，无需提前设置神经网络的参数，并通过 Kalman 滤波对网络参数进行在线学习。不同于传统的自适应控制，该方法在设计控制律时认为学习到的神经网络模型存在建模误差。因此在代价函数中增加了由交叉信息熵来描述的系统信息增量，使控制器求取自适应控制律时，不仅考虑了系统最优跟踪控制性能，同时还考虑到了如何优化网络模型的学习性能，减少建模误差，使得控制律具有了主动学习的特性，也称为对偶特性，从而进一步提升整体控制性能指标。通过仿真实验表明，该方法对未知非线性系统的跟踪控制是有效和可行的。



## 7 具有主动学习特性的抗扰动自适应对偶控制

### 7.1 引言

绝大多数实际系统控制过程当中都存在扰动因素，这些扰动包括来自外部环境的扰动和来自系统自身的不确定性<sup>[157-159]</sup>。例如在高速列车的速度控制当中，扰动可能来自外部环境，如风、雨、雪等的干扰，也可能来自系统内部的干扰，如机械部件和电子元件劳损、老化、故障<sup>[160-162]</sup>。这些扰动会降低控制性能，甚至损坏系统，从而造成经济损失。因此对于实际系统，如何减少扰动带来的负面影响，提高控制系统的鲁棒性就极为关键。通常前馈控制策略可以消除可观测的扰动带来的负面影响，但在实际系统中，扰动在多数情况下是难以直接测量出来的。这个问题就推动了鲁棒控制和自适应控制在具有不可测量的动态扰动的系统控制中的研究<sup>[163]</sup>。

鲁棒控制可以有效解决具有扰动的系统控制问题，例如，Zhang 等人针对具有有界扰动的部分未知马尔可夫跳变系统提出了  $H_2/H_\infty$  模型预测控制<sup>[164]</sup>。Hooshmandi 等人提出了针对非线性样本数据系统的基于多项式线性参数变化模型的鲁棒  $H_\infty$  控制方法<sup>[165]</sup>。Liu 等人针对随机非线性系统设计了自适应神经网络指定性能指标的有界  $H_\infty$  控制器<sup>[166]</sup>。鲁棒控制策略一般假设扰动有界，并计算有界扰动下控制性能的最大性能指标，通过最小化最大扰动下的性能指标得到鲁棒控制律。然而，由于是系统处在最坏情况求得的控制律，所得到的控制律会过于保守。

与鲁棒控制不同的是，自适应控制策略通过学习系统未知参数，并利用学习到的信息来调整控制律，该方法相对于保守的鲁棒控制策略来说更为激进。然而自适应控制中的学习与控制是相互冲突的。也就是说，对系统未知部分的学习需要对系统输入较大的激励来获取用于减少系统不确定性的信息，而较大的激励会降低系统控制性能。反之，较少的激励可以使控制过程变得平稳，但这抑制了进一步探索系统以提高控制性能的机会。对偶控制由 Feldaum 首先提出，这是一种针对学习与控制之间的冲突问题的控制策略<sup>[15]</sup>。由于对偶控制的求解极其困难，因此后续学者针对参数未知的系统控制问题，提出了一系列次优解，这些次优自适应对偶控制方法也能够很好的平衡未知系统的学习和控制。近些年来对偶制理论得到了进一步的发展，并成功地应用于各个领域。例如 Tutsoy 等人开发了一种基于强化学习的无模型对偶控制方案，以平衡未知系统的学习与控制<sup>[167]</sup>。Chen 等人将对偶控制理论嵌入到一个并行学习框架，并将其用于未知环境下的自主源搜索<sup>[168]</sup>。Klenske 等人利用对偶控制理论解决贝叶斯强化学习中的探索-利用的权衡问题<sup>[169]</sup>。Liu 等人设计了一种针对失效随机系统的容错对偶控制方法<sup>[170]</sup>。自适应对偶控制在处理不确定系统的控制方面有很好的应用前景。因此本章基于对偶控制方法的思想，来解决具有未知不可测扰动的非线性系统的跟踪控制问题。

过往的研究中对系统未知扰动用不同的方法进行描述，并给出了一些处理方法以提

高系统控制性能。其中绝大部分用加性扰动对未知扰动进行描述。另外一些研究还提出在实际应用场景中干扰具有乘性特征，如石油地震勘探<sup>[171]</sup>、水下通信系统<sup>[172]</sup>、雷达目标跟踪<sup>[173]</sup>和信息物理系统<sup>[174]</sup>。近些年来，已有一些研究利用乘性扰动的概念用来分析系统从而提高控制的鲁棒性。例如，Zhao 等人利用线性矩阵不等式技术开发了具有乘性噪声的不确定马尔可夫跳变系统的  $H_\infty$  控制<sup>[175]</sup>。Mazouchi 等人提出了一种凸优化方法，利用系统级综合方法对带有乘性扰动的线性二次调节问题进行优化<sup>[176]</sup>。Wang 等人针对被乘性噪声干扰的严格反馈非线性系统建立了一种反步控制器<sup>[177]</sup>。

本章节将在未知系统控制过程中引入了加性扰动和乘性扰动，对此设计了一个特殊神经网络模型（Specialized Neural Network, SNN）来学习未知系统，其中的干扰是由加性和乘性方式来加入到网络模型中。然后将特殊神经网络嵌入到自适应对偶控制结构中，该对偶控制方法将学习和控制进行解耦，使系统一边对未知动态扰动进行主动学习，一边对系统输出进行跟踪控制。该方法首先为神经网络中的扰动参数分配一组候选值，然后并行实时学习这些候选值的贝叶斯后验概率，每个候选扰动值都对应一个控制律，最终的控制律为所有控制律的后验概率加权的总和。该方法的优点总结如下：

- (1) 针对具有动态扰动的未知非线性系统构造了一个特殊神经网络 SNN，结构为仿射非线性神经网络模型，并集成了加性扰动和乘性扰动来描述系统所受的扰动，所构造的 SNN 旨在精确地反映扰动对系统的影响。
- (2) 设计了实时主动学习 SNN 中扰动参数的方法。通过设计有限候选值集来逼近有界扰动，从而将对扰动参数的学习转化为更新候选值的后验概率。通过迭代计算每个候选扰动值的贝叶斯后验概率，当某个后候选值对应的验概率收敛为 1 时，该候选扰动值即为扰动估计值。
- (3) 提出了一种针对具有不可测量的扰动的非线性系统的自适应对偶控制方法。该方法打破了扰动学习与输出跟踪控制之间的耦合关系，将扰动的学习与系统输出跟踪控制分离开来，表现为扰动的学习与输出控制能够并行工作。
- (4) 使用单次实验和蒙特卡洛仿真实验验证所提出的方法。通过对具有动态扰动的非线性系统进行仿真实验，测试了所提出的控制方法的鲁棒性。为了获得更有说服力的结果，还将该方法在高速列车速度控制中进行仿真实验加以验证。

## 7.2 问题描述

考虑具有加性扰动和乘性扰动的离散时间仿射非线性系统

$$y(k+1) = \alpha(k)f[x(k)] + \beta(k)g[x(k)]u(k) + \gamma(k) + e(k) \quad (7-1)$$

其中  $x(k) = [y(k), y(k-1), \dots, y(k-n+1), u(k-1), u(k-2), \dots, u(k-m)]^T$  是系统状态向量， $y(k)$  是系统输出， $u(k)$  是系统控制信号， $f[x(k)]$  和  $g[x(k)]$  是关于系统状态向量  $x(k)$  的非线性函数， $\alpha(k)$  和  $\beta(k)$  是乘性扰动， $\gamma(k)$  是加性扰动， $e(k)$  是高斯白噪声，并做如

下假设：

假设 7.1：乘性扰动  $\alpha(k)$  和  $\beta(k)$ ，以及加性扰动  $\gamma(k)$  是未知且时变的，并且分别属于有界区间  $[\alpha_l, \alpha_u]$ ,  $[\beta_l, \beta_u]$  和  $[\gamma_l, \gamma_u]$ 。 $\alpha_l$  和  $\alpha_u$  分别是扰动  $\alpha(k)$  的下界和上界， $\beta_l$  和  $\beta_u$  分别是扰动  $\beta(k)$  的下界和上界， $\gamma_l$  和  $\gamma_u$  分别是扰动  $\gamma(k)$  的下界和上界。

假设 7.2：系统状态向量中的维度参数  $n$  和  $m$  已知。

假设 7.3：系统随机噪声  $e(k)$  服从均值为 0，方差为  $\sigma^2$  的高斯分布。

假设 7.4：系统是最小相位系统，且非线性函数  $f[x(k)]$  和  $g[x(k)]$  未知。

基于以上模型和假设，可以对受扰动的未知非线性系统进行控制。对于系统输出跟踪的控制目标，对偶控制性能指标为

$$J(k+1) = E \left\{ \sum_{k=1}^N [y(k+1) - y_r(k+1)]^2 \mid \mathfrak{I}^k \right\} \quad (7-2)$$

其中  $y_r(k)$  是系统输出参考轨迹， $\mathfrak{I}^k$  是在  $k$  时刻的状态信息

$$\mathfrak{I}^k = \{u(0), u(1), \dots, u(k-1), y(1), y(2), \dots, y(k)\} \quad (7-3)$$

控制目标是设计出一个反馈控制律

$$u(k) = \mu_k(\mathfrak{I}^k) \quad (7-4)$$

能够最小化性能指标  $J(k+1)$ ，其中控制策略  $\mu(\cdot)$  是非线性函数。

由于系统中存在未知扰动  $\alpha(k)$ ,  $\beta(k)$  和  $\gamma(k)$ ，以及非线性函数  $f[x(k)]$  和  $g[x(k)]$ ，因此求出控制策略  $\mu(\cdot)$  极具挑战性。传统的基于神经网络的自适应控制策略可以对未知扰动和参数进行一边学习一边控制。然而对扰动和参数的学习与输出跟踪控制本身是相互耦合的，也就是说， $k$  时刻的学习会受到  $k-1$  时刻的控制律的影响，同样  $k$  时刻的控制律也依赖于  $k-1$  时刻的扰动和参数的学习。这种耦合会在干扰或者系统参数发生变化甚至突变时，影响当前控制方案对未知系统的控制性能。本章节旨在解决这种耦合问题，设计一种对偶控制策略来打破耦合回路，并对式 (7-1) 中所示的非线性系统的动态扰动进行学习和系统输出跟踪。

### 7.3 控制器设计

这一节详细阐述了具有主动学习特点的抗扰动自适应对偶控制方法。首先设计包含乘性和加性扰动的特殊的神经网络对具有动态扰动的非线性系统进行建模，然后以该特殊神经网络作为基础来辨识系统参数和扰动，并将其用于后续未知系统的控制律推导，详见第 7.3.1 节。第 7.3.2 节描述了如何使用有限候选值集来逼近 SNN 中的有界扰动，这为主动学习未知扰动奠定了基础。第 7.3.3 节阐述了具有主动学习特性的抗扰动自适应对偶控制，该方法是通过学习扰动候选值的后验概率来辨识扰动参数，该方法通过对偶控制思想打破了控制与学习之间的耦合关系。

### 7.3.1 针对具有扰动的非线性系统的特殊神经网络

特殊神经网络 SNN 分别用两个高斯径向基函数  $\hat{f}$  和  $\hat{g}$  来逼近未知非线性函数  $f[x(k)]$  和  $g[x(k)]$ 。高斯径向基函数  $\hat{f}$  和  $\hat{g}$  分别为

$$\hat{f} = \hat{f}[\hat{w}_f, x(k)] = \hat{w}_f^T h_f[x(k)] \quad (7-5)$$

$$\hat{g} = \hat{g}[\hat{w}_g, x(k)] = \hat{w}_g^T h_g[x(k)] \quad (7-6)$$

其中  $\hat{w}_f$  和  $\hat{w}_g$  是神经网络的输出层参数， $h_f[x(k)]$  和  $h_g[x(k)]$  是高斯径向基函数向量，第  $i$  个元素为

$$h_{f_i}[x(k)] = \exp \left\{ \frac{-\|x(k) - \hat{c}_{f_i}\|^2}{2\hat{b}_{f_i}^2} \right\} \quad (7-7)$$

$$h_{g_i}[x(k)] = \exp \left\{ \frac{-\|x(k) - \hat{c}_{g_i}\|^2}{2\hat{b}_{g_i}^2} \right\} \quad (7-8)$$

其中  $\hat{c}_{f_i}$  和  $\hat{c}_{g_i}$  是第  $i$  个高斯径向基函数的坐标中心点， $\hat{b}_{f_i}^2$  和  $\hat{b}_{g_i}^2$  是对应的方差。

考虑到系统的加性和乘性扰动，非线性系统用特殊神经网络近似，可写成如下形式

$$y(k+1) \approx \alpha(k)\hat{w}_f^T h_f[x(k)] + \beta(k)\hat{w}_g^T h_g[x(k)]u(k) + \gamma(k) + e(k) \quad (7-9)$$

该网络结构如图 7-1 所示

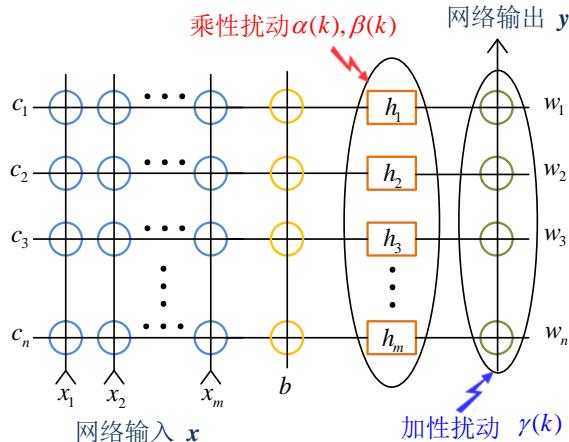


图 7-1 具有乘性扰动和加性扰动的特殊神经网络

Fig.7-1 The specialized neural network which contains multiplicative and additive disturbances

接下来考虑如何学习神经网络的参数。假设系统在平稳状态且扰动不变化的情况下，可以用历史输入输出数据线下学习网络的参数。首先假设系统的乘性扰动为恒值 1，加性扰动为 0，用输入输出数据对网络参数进行学习，得到的近似非线性系统表示为

$$\hat{y}(k+1) = \hat{w}_f^T h_f[x(k)] + \hat{w}_g^T h_g[x(k)]u(k) \quad (7-10)$$

**引理 7.1:** 假设在  $k$  时刻系统的乘性和加性扰动均已知，分别为  $\hat{\alpha}(k)$ ， $\hat{\beta}(k)$  和  $\hat{\gamma}(k)$ 。根据基于新息的自适应对偶控制原理，通过最小化以下代价函数

$$J(k+1) = E\{[y(k+1) - y_r(k+1)]^2 - \lambda[y(k+1) - \hat{y}(k+1)]^2 | \mathfrak{I}^k\} \quad (7-11)$$

可以得到自适应对偶控制律

$$u_t(k) = \frac{[y_r(k+1) - \hat{\alpha}(k)\hat{w}_f h_f - \hat{\gamma}(k)\hat{w}_g h_g]\hat{\beta}(k)\hat{w}_g h_g}{(1-\lambda)\hat{w}_g h_g P_\beta + [\hat{\beta}(k)\hat{w}_g h_g]^2} - \frac{(1-\lambda)(\hat{w}_f h_f P_{\alpha\beta} + P_{\gamma\beta})\hat{w}_g h_g}{(1-\lambda)\hat{w}_g h_g P_\beta + [\hat{\beta}(k)\hat{w}_g h_g]^2} \quad (7-12)$$

其中参数  $P_{\alpha\beta}$ ， $P_{\gamma\beta}$  和  $P_\beta$  分别是扰动估计误差协方差里面的元素， $\lambda$  为对偶控制系数。

### 7.3.2 设计有限扰动候选集对扰动进行近似

根据假设 7.1，扰动  $\alpha(k)$ ， $\beta(k)$  和  $\gamma(k)$  是未知但有界的，因此可以将相应的有界区间切分成多个小的子区间，然后可以通过实时的输入输出数据来学习出每个时刻最接近实际扰动的子区间，并认为该子区间的中值为当前时刻扰动的估计值。定理 7.1 详细给出了如何设置子区间。

**定理 7.1:** 给定任意值  $\varepsilon_\alpha > 0$ ， $\varepsilon_\beta > 0$  和  $\varepsilon_\gamma > 0$ ，存在正整数  $s_\alpha$ ， $s_\beta$  和  $s_\gamma$ ，以及三组扰动切分点

$$\begin{aligned} \alpha_l &= \alpha_1 < \alpha_2 < \cdots < \alpha_{s_\alpha} = \alpha_u \\ \beta_l &= \beta_1 < \beta_2 < \cdots < \beta_{s_\beta} = \beta_u \\ \gamma_l &= \gamma_1 < \gamma_2 < \cdots < \gamma_{s_\gamma} = \gamma_u \end{aligned} \quad (7-13)$$

使得

$$\begin{aligned} \bigcup_{i=1}^{s_\alpha} [\alpha_i, \alpha_{i+1}] &= [\alpha_l, \alpha_u], \quad |\alpha_i - \alpha_{i+1}| < \varepsilon_\alpha \\ \bigcup_{i=1}^{s_\beta} [\beta_i, \beta_{i+1}] &= [\beta_l, \beta_u], \quad |\beta_i - \beta_{i+1}| < \varepsilon_\beta \\ \bigcup_{i=1}^{s_\gamma} [\gamma_i, \gamma_{i+1}] &= [\gamma_l, \gamma_u], \quad |\gamma_i - \gamma_{i+1}| < \varepsilon_\gamma \end{aligned} \quad (7-14)$$

证明：对于任意正数  $\varepsilon_\alpha$ ，存在一个正整数  $s_\alpha$  满足

$$s_\alpha - 1 = \left\lceil \frac{\alpha_u - \alpha_l}{\varepsilon_\alpha} \right\rceil \quad (7-15)$$

其中  $[x]$  代表不超过  $x$  的最大整数，因此有

$$\frac{\alpha_u - \alpha_l}{\varepsilon_\alpha} - \left\lceil \frac{\alpha_u - \alpha_l}{\varepsilon_\alpha} \right\rceil < 1 \quad (7-16)$$

将式 (7-15) 代入到式 (7-16) 中得到

$$\frac{\alpha_u - \alpha_l}{\varepsilon_\alpha} < \left\lceil \frac{\alpha_u - \alpha_l}{\varepsilon_\alpha} \right\rceil + 1 = s_\alpha \quad (7-17)$$

式 (7-17) 也可以写成如下形式

$$\frac{\alpha_u - \alpha_l}{s_\alpha} < \varepsilon_\alpha \quad (7-18)$$

将扰动区间平均切分成  $s_\alpha$  个子区间，每个子区间的长度为

$$\frac{\alpha_u - \alpha_l}{s_\alpha} \quad (7-19)$$

区间  $[\alpha_l, \alpha_u]$  的分割点可以按如下公式设置

$$\alpha_1 = \alpha_l, \quad \alpha_i = \alpha_1 + \frac{\alpha_u - \alpha_l}{s_\alpha} i, \quad i = 1, 2, \dots, s_\alpha - 1, \quad \alpha_{s_\alpha+1} = \alpha_u \quad (7-20)$$

区间  $[\beta_l, \beta_u]$  和  $[\gamma_l, \gamma_u]$  的分割点同上。证毕。

根据定理 7.1，扰动  $\alpha(k)$ ， $\beta(k)$  和  $\gamma(k)$  会位于相对应的某个子区间内，这里使用子区间的中值

$$\theta_{\alpha_i} = \frac{\alpha_{i-1} - \alpha_i}{2}, \quad \theta_{\beta_i} = \frac{\beta_{i-1} - \beta_i}{2}, \quad \theta_{\gamma_i} = \frac{\gamma_{i-1} - \gamma_i}{2} \quad (7-21)$$

来分别逼近属于子区间  $[\alpha_{i-1}, \alpha_i]$ ， $[\beta_{i-1}, \beta_i]$  和  $[\gamma_{i-1}, \gamma_i]$  的实际扰动值。因此可以使用一系列的中点来逼近有界区间的扰动

$$\begin{aligned} [\alpha_l, \alpha_u] &= \Omega_\alpha = \{\theta_{\alpha_1}, \theta_{\alpha_2}, \dots, \theta_{\alpha_{s_\alpha}}\} \\ [\beta_l, \beta_u] &= \Omega_\beta = \{\theta_{\beta_1}, \theta_{\beta_2}, \dots, \theta_{\beta_{s_\beta}}\} \\ [\gamma_l, \gamma_u] &= \Omega_\gamma = \{\theta_{\gamma_1}, \theta_{\gamma_2}, \dots, \theta_{\gamma_{s_\gamma}}\} \end{aligned} \quad (7-22)$$

那么就可以通过更新当前实际扰动处于每个子区间的概率来得到扰动近似值。在每个子区间的初始概率可以设置为相等的概率，分别为  $\pi_{\alpha_i}(1) = 1/s_\alpha$ ， $\pi_{\beta_i}(1) = 1/s_\beta$  和  $\pi_{\gamma_i}(1) = 1/s_\gamma$ 。扰动估计的精度随着子区间的数量而增长，因为更多的子区间带来更高的分辨率。然而子区间数量的增加会增大有限集概率估计的计算量。因此，需要通过选择合适的子区间数来权衡扰动逼近精度与计算量之间的关系。

假设  $\alpha^*$  是扰动  $\alpha(k)$  的真值，且位于子区间  $[\alpha_{i-1}, \alpha_i]$  中。根据式 (7-14) 有不等式  $|\alpha^* - \theta_{\alpha_i}| < \varepsilon_\alpha$ ，表明估计误差小于  $\varepsilon_\alpha$ ，因此可以用  $\varepsilon_\alpha$  来调整子区间的长度。根据式 (7-15)，子区间的个数设置为  $s_\alpha = \left\lceil \frac{\alpha_u - \alpha_l}{\varepsilon_\alpha} \right\rceil + 1$ 。 $\beta(k)$  和  $\gamma(k)$  子区间个数的选择同上。

基于以上对乘性扰动和加性扰动参数的逼近，通过求解下面的控制问题 ( $P$ ) 就可以实现 7.3.1 小节中提出的控制目标

$$\begin{aligned} (P) \quad \min J &= E\{[y(k+1) - y_r(k+1)]^2 - \lambda[y(k+1) - \hat{y}(k+1)]^2 | \mathfrak{F}^k\} \\ \text{s.t.} \quad y(k+1) &= \alpha(k)f[x(k)] + \beta(k)g[x(k)]u(k) + \gamma(k) + e(k), \end{aligned} \quad (7-23)$$

其中未知扰动属于有限集： $\alpha(k) \in \Omega_\alpha = \{\theta_{\alpha_1}, \theta_{\alpha_2}, \dots, \theta_{\alpha_{s_\alpha}}\}$ ， $\beta(k) \in \Omega_\beta = \{\theta_{\beta_1}, \theta_{\beta_2}, \dots, \theta_{\beta_{s_\beta}}\}$ ，

$$\gamma(k) \in \Omega_\gamma = \{\theta_{\gamma_1}, \theta_{\gamma_2}, \dots, \theta_{\gamma_{s_\gamma}}\}.$$

### 7.3.3 抗扰动自适应对偶控制器

基于上述特殊神经网络和未知扰动的近似，本小节设计了具有抗扰动功能的对偶控制器。针对式(7-23)描述的控制问题，在 $k$ 时刻的扰动 $\alpha(k)$ ， $\beta(k)$ 和 $\gamma(k)$ 分别为候选集 $\Omega_\alpha = \{\theta_{\alpha_i} : i=1, 2, \dots, s_\alpha\}$ ， $\Omega_\beta = \{\theta_{\beta_j} : j=1, 2, \dots, s_\beta\}$ ， $\Omega_\gamma = \{\theta_{\gamma_l} : l=1, 2, \dots, s_\gamma\}$ 中的一个元素。为简化表达，定义扰动向量为 $\theta_t = [\theta_{\alpha_i}, \theta_{\beta_j}, \theta_{\gamma_l}]$ ，扰动向量集为

$$\Omega = \{\theta_t : t=1, 2, \dots, s_\alpha s_\beta s_\gamma\} \quad (7-24)$$

这意味着扰动向量 $\theta_t$ 有 $s_\alpha s_\beta s_\gamma$ 种组合选项。定义扰动向量 $\theta_t$ 的先验概率为 $\pi(\theta_t)$ ，并且先验概率的和为

$$\sum_{t=1}^{s_\alpha s_\beta s_\gamma} \pi(\theta_t) = 1 \quad (7-25)$$

抗扰动对偶控制律为

$$u(k) = \sum_{t=1}^{s_\alpha s_\beta s_\gamma} \pi(\theta_t | \mathfrak{I}^k) u(k, \theta_t) \quad (7-26)$$

其中 $\pi(\theta_t | \mathfrak{I}^k)$ 是给定信息状态 $\mathfrak{I}^k$ 下扰动向量 $\theta_t$ 的贝叶斯后验概率。 $\pi(\theta_t | \mathfrak{I}^k)$ 根据下式随时间进行更新

$$\pi(\theta_t | \mathfrak{I}^k) = \frac{p(y(k) | \theta_t, \mathfrak{I}^k) \pi(\theta_t | \mathfrak{I}^{k-1})}{\sum_{t=1}^{s_\alpha s_\beta s_\gamma} p(y(k) | \theta_t, \mathfrak{I}^k) \pi(\theta_t | \mathfrak{I}^{k-1})} \quad (7-27)$$

其中

$$p(y(k) | \theta_t, \mathfrak{I}^k) = \frac{1}{\sqrt{2\pi\Sigma_y(k, \theta_t)}} \exp\left\{-\frac{\tilde{y}^2(k, \theta_t)}{2\Sigma_y(k, \theta_t)}\right\} \quad (7-28)$$

后验概率初值为

$$\pi(\theta_t | \mathfrak{I}^0) = \pi(\theta_t) = \frac{1}{s_\alpha s_\beta s_\gamma}, \quad t=1, 2, \dots, s_\alpha s_\beta s_\gamma \quad (7-29)$$

且

$$\sum_{t=1}^{s_\alpha s_\beta s_\gamma} \pi(\theta_t | \mathfrak{I}^k) = 1 \quad (7-30)$$

这里的控制律 $u(k, \theta_t)$ 是所有候选扰动值对应的控制律的加权和，权重为对应的后验概率。根据引理 7.1，扰动 $\theta_t$ 对应的反馈控制律为

$$u(k, \theta_t) = \frac{[y_r(k+1) - \theta_t(1)\hat{w}_f h_f - \theta_t(3)]\hat{\beta}(k)\hat{w}_g h_g}{(1-\lambda)\hat{w}_g h_g P_\beta + [\theta_t(2)\hat{w}_g h_g]^2} - \frac{(1-\lambda)(\hat{w}_f h_f P_{\alpha\beta} + P_{\gamma\beta})\hat{w}_g h_g}{(1-\lambda)\hat{w}_g h_g P_\beta + [\theta_t(2)\hat{w}_g h_g]^2} \quad (7-31)$$

该控制律也被称为  $\theta_t$ -候选控制律 ( $\theta_t$ -candidate)。其中  $\theta_t(1)$ ,  $\theta_t(2)$  和  $\theta_t(3)$  均为向量  $\theta_t$  的元素, 分别是扰动  $\alpha(k)$ ,  $\beta(k)$  和  $\gamma(k)$  的估计值。  $P_\beta$ ,  $P_{\alpha\beta}$  和  $P_{\gamma\beta}$  为扰动向量的估计误差协方差矩阵  $P(k) = E\{\tilde{\theta}_t \tilde{\theta}_t^T\}$  中的元素。扰动向量估计误差协方差矩阵的更新公式为

$$P(k+1) = \log_2[\eta / \pi(\theta_t | \mathfrak{I}^k) + 1]P(k) \quad (7-32)$$

其中系数  $\eta$  设置为  $1/s_\alpha s_\beta s_\gamma$ , 该迭代公式能够为后验概率较低的元素分配较大的方差值。

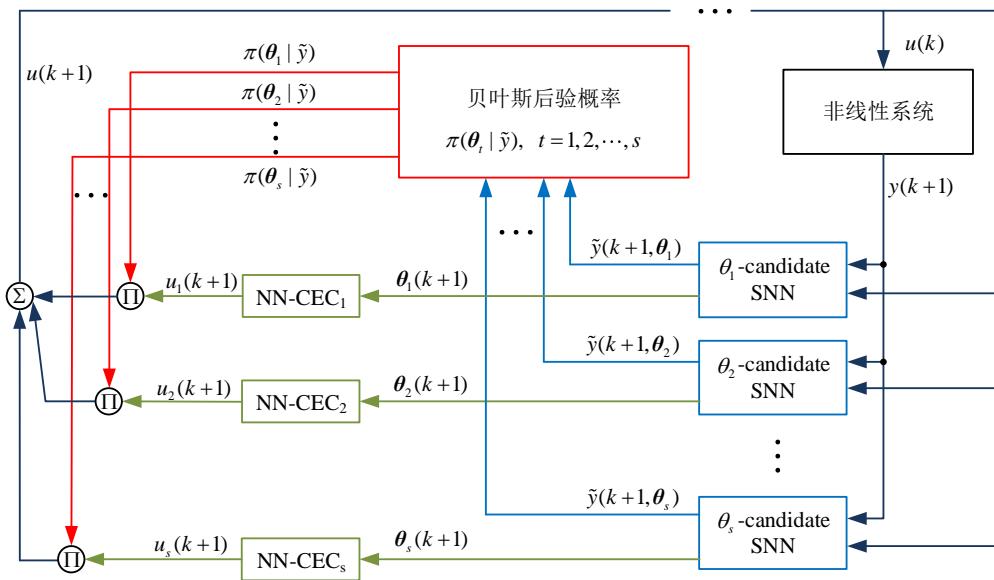


图 7-2 抗扰动对偶控制结构框图  
Fig. 7-2 The block diagram of the anti-disturbance dual control scheme

图 7-2 为抗扰动对偶控制策略的结构框图。如图所示, 所设计的控制律是通过对  $s = s_\alpha s_\beta s_\gamma$  个  $\theta_t$ -候选控制律求和得到, 其权重为每个候选值对应的贝叶斯后验概率  $\pi(\theta_t | \mathfrak{I}^k)$ 。该控制策略具有相互冲突的双重特性: (1) 驱动系统输出跟踪参考轨迹; (2) 主动学习系统动未知参扰动以减少控制的不确定性。如图所示, 该方法将跟踪控制和主动学习进行解耦。一方面所设计的特殊神经网络, 对应每一个候选扰动值  $\theta_t$  ( $t \in \{1, 2, 3, \dots, s_\alpha s_\beta s_\gamma\}$ ), 都有一个与之对应的  $\theta_t$ -候选神经网络 ( $\theta_t$ -candidate SNN), 每个候选神经网络都对应一个对偶控制律 (NN-DPC<sub>t</sub>)。另一方面, 基于  $\theta_t$ -candidate SNN 的一步超前输出预测误差为  $\tilde{y}(k+1, \theta_t)$ , 用输出预测误差作为新息来更新扰动候选值对应的贝叶斯后验概率, 实现对未知扰动  $\theta_t$  的主动学习。

因此这里所设计的抗干扰对偶控制方法具有以下特点。假设扰动的真实值更接近于候选扰动  $\theta_{t^*}$ , 那么通过主动学习扰动估计值可以收敛到  $\theta_{t^*}$ , 则  $\theta_{t^*}$ -candidate 对应的一步超前预测  $\tilde{y}(k+1, \theta_{t^*})$  是最优系统输出估计。接下来将在下面两个定理中阐明并证明了这一说法。定理 6.2 证明了  $\theta_{t^*}$ -candidate 一步预测误差  $\tilde{y}(k+1, \theta_{t^*})$  的方差与其他估计值

$\tilde{y}(k+1, \theta_t)$ ,  $t \neq t^*$  相比是最小的。定理 6.3 证明了不同扰动候选值对应的贝叶斯后验概率的收敛性。即当  $k \rightarrow \infty$ ,  $\theta_{t^*}$ -candidate 对应的后验概率  $\pi(\theta_{t^*} | \tilde{y}) \rightarrow 1$ , 并且对  $\forall t \neq t^*$ , 其他的后验概率有  $\pi(\theta_t | \tilde{y}) \rightarrow 0$ 。

**定理 7.2:** 假设扰动向量  $\theta_{t^*}$  最接近扰动  $\alpha(k)$ ,  $\beta(k)$  和  $\gamma(k)$  的真值, 则对其他的扰动向量  $\theta_t$ ,  $t \neq t^*$  有如下不等式

$$E\{\tilde{y}^2(k+1|k, \theta_t)\} > E\{\tilde{y}^2(k+1|k, \theta_{t^*})\} \quad (7-33)$$

其中  $E\{\tilde{y}^2(k+1|k, \theta_t)\}$  是在  $k$  时刻  $\theta_t$  候选值下的系统输出误差协方差。

证明: 定义  $\theta_{t^*}$  候选值下  $k$  时刻系统一步超前预测为

$$\hat{y}(k+1, \theta_{t^*}) = \theta_{t^*}(1)\hat{f}[x(k)] + \theta_{t^*}(2)\hat{g}[x(k)]u(k) + \theta_{t^*}(3) \quad (7-34)$$

定义一步超前预测误差为

$$\tilde{y}(k+1, \theta_{t^*}) = y(k+1) - \hat{y}(k+1, \theta_{t^*}) \quad (7-35)$$

将式 (7-1) 描述的系统代入到  $\tilde{y}(k+1, \theta_{t^*})$  中得到

$$\begin{aligned} \tilde{y}(k+1, \theta_{t^*}) &= \{\alpha(k)f[x(k)] - \theta_{t^*}(1)\hat{f}[x(k)]\} + \{\beta(k)g[x(k)] - \theta_{t^*}(2)\hat{g}[x(k)]\}u(k) \\ &\quad + \{\gamma(k) - \theta_{t^*}(3)\} + e(k) \end{aligned} \quad (7-36)$$

假设线下学习的非线性函数与真实函数非常接近并满足  $f[x(k)] = \hat{f}[x(k)]$  和  $g[x(k)] = \hat{g}[x(k)]$ 。那么在  $\theta_{t^*}$  下的一步超前预测误差可以写成

$$\begin{aligned} \tilde{y}(k+1, \theta_{t^*}) &= \tilde{\alpha}_{t^*}(k)\hat{f}[x(k)] + \tilde{\beta}_{t^*}(k)\hat{g}[x(k)]u(k) + \tilde{\gamma}_{t^*}(k) + e(k) \\ &= \tilde{\theta}_{t^*}(k)\phi(k) + e(k) \end{aligned} \quad (7-37)$$

其中  $\tilde{\alpha}_{t^*}(k) = \alpha(k) - \theta_{t^*}(1)$ ,  $\tilde{\beta}_{t^*}(k) = \beta(k) - \theta_{t^*}(2)$ ,  $\tilde{\gamma}_{t^*}(k) = \gamma(k) - \theta_{t^*}(3)$ ,  $\phi(k) = [\hat{f}[x(k)], \hat{g}[x(k)]u(k), 1]$ 。在  $\theta_{t^*}$  下的一步超前预测误差方差定义为

$$\begin{aligned} \Sigma_y(k+1, \theta_{t^*}) &= E\{\tilde{y}^2(k+1, \theta_{t^*})\} \\ &= \phi(k)^T E\{\tilde{\theta}_{t^*}(k)\tilde{\theta}_{t^*}^T(k)\}\phi(k) + E\{e^2(k)\} \\ &= \phi(k)^T P(k)\phi(k) + \sigma^2 \end{aligned} \quad (7-38)$$

根据假设 7.3, 过程噪声  $e(k)$  是均值为 0, 方差为  $\sigma^2$  的高斯噪声。由于  $\theta_{t^*}$  最接近扰动真值, 相应的后验概率  $\pi(\theta_{t^*} | \tilde{y})$  会收敛到 1, 那么式 (7-32) 中的  $\log_2[\eta / \pi(\theta_t | \mathfrak{I}^k) + 1]$  小于 1, 因此估计误差协方差  $P(k)$  收敛到 0。进而可以得到

$$\Sigma_y(k+1, \theta_{t^*}) = \sigma^2 \quad (7-39)$$

如果扰动向量  $\theta_t$  不是最接近真值, 即  $t \neq t^*$ , 则对应的后验概率  $\pi(\theta_t | \tilde{y})$  收敛到 0,  $\log_2[\eta / \pi(\theta_t | \mathfrak{I}^k) + 1]$  大于 1, 那么估计误差协方差  $P(k)$  就会远远大于 0。因此, 在  $\theta_t$  ( $t \neq t^*$ ) 下的一步超前预测误差方差定义为

$$\Sigma_y(k+1, \theta_t) = E\{\tilde{y}^2(k+1, \theta_t)\} = \phi(k)^T P(k) \phi(k) + \sigma^2 > \sigma^2 \quad (7-40)$$

对比式 (7-39) 和式 (7-40)，可以得到如下不等式

$$E\{\tilde{y}^2(k+1|k, \theta_t)\} > E\{\tilde{y}^2(k+1|k, \theta_{t^*})\} \quad (7-41)$$

证毕。

**定理 7.3:** 假设  $\theta_{t^*}$  是需要辨识的扰动的真值。那么当  $k \rightarrow \infty$ ，后验概率通过主动学习方法得到  $\pi(\theta_{t^*} | \mathfrak{I}^k) \rightarrow 1$ ，而对于任意  $t \neq t^*$ ，后验概率  $\pi(\theta_t | \mathfrak{I}^k) \rightarrow 0$ 。

证明：假设  $\tilde{y}(k|k, \theta_t)$  是弱渐近平稳序列，则平稳过程  $\tilde{y}(k|k, \theta_t)$  对  $\theta_t$  具有遍历性，且协方差矩阵为常矩阵并用  $\Sigma_t$  表示。定义

$$L_t(k) = \pi(\theta_t | \mathfrak{I}^k) \pi^{-1}(\theta_{t^*} | \mathfrak{I}^k), \quad t \in \{1, 2, \dots, s_\alpha s_\beta s_\gamma\} \quad (7-42)$$

将式 (7-27) 代入到  $L_t(k)$  的表达式中可得

$$L_t(k) = \frac{p(y(k) | \theta_t, \mathfrak{I}^k) \pi(\theta_t | \mathfrak{I}^{k-1})}{p(y(k) | \theta_{t^*}, \mathfrak{I}^k) \pi(\theta_{t^*} | \mathfrak{I}^{k-1})} = \frac{p(y(k) | \theta_t, \mathfrak{I}^k)}{p(y(k) | \theta_{t^*}, \mathfrak{I}^k)} L_t(k-1) \quad (7-43)$$

根据上式进行迭代可以得到

$$\frac{L_t(k+n-1)}{L_t(k-1)} = \prod_{\tau=k}^{k+n-1} \frac{p(y(\tau) | \theta_t, \mathfrak{I}^\tau)}{p(y(\tau) | \theta_{t^*}, \mathfrak{I}^\tau)} \quad (7-44)$$

将概率密度函数代入到上式中得

$$\frac{L_t(k+n-1)}{L_t(k-1)} = \left\{ \frac{\Sigma_t}{\Sigma_{t^*}} \right\}^{-\frac{n}{2}} \frac{\exp \left\{ \sum_{\tau=k}^{k+n-1} -\frac{\tilde{y}^2(\tau, \theta_t)}{2\Sigma_t} \right\}}{\exp \left\{ \sum_{\tau=k}^{k+n-1} -\frac{\tilde{y}^2(\tau, \theta_{t^*})}{2\Sigma_{t^*}} \right\}} \quad (7-45)$$

对  $\frac{L_t(k+n-1)}{L_t(k-1)}$  取自然对数可得

$$\ln \left\{ \frac{L_t(k+n-1)}{L_t(k-1)} \right\} = \frac{n}{2} \ln \left\{ \frac{\Sigma_{t^*}}{\Sigma_t} \right\} - \frac{1}{2} \sum_{\tau=k}^{k+n-1} \frac{\tilde{y}^2(\tau | \tau, \theta_t)}{\Sigma_t} + \frac{1}{2} \sum_{\tau=k}^{k+n-1} \frac{\tilde{y}^2(\tau | \tau, \theta_{t^*})}{\Sigma_{t^*}} \quad (7-46)$$

根据假设  $\theta_{t^*}$  是扰动的真值，那么可以得到

$$E\{\tilde{y}^2(k | k, \theta_{t^*})\} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\tau=k}^{k+n-1} \tilde{y}^2(\tau | \tau, \theta_{t^*}) = \Sigma_{t^*} \quad (7-47)$$

等式的右边还可以写成

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\tau=k}^{k+n-1} \frac{\tilde{y}^2(\tau | \tau, \theta_{t^*})}{\Sigma_{t^*}} = 1 \quad (7-48)$$

然而对其它的扰动向量  $\theta_t (t \neq t^*)$  有

$$E\{\tilde{y}^2(k|k, \theta_t)\} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\tau=k}^{k+n-1} \tilde{y}^2(\tau|\tau, \theta_t) \quad (7-49)$$

并且根据不等式 (7-41) 可以得

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\tau=k}^{k+n-1} \tilde{y}^2(\tau|\tau, \theta_t) > \Sigma_{t^*}, \quad \forall t \neq t^* \quad (7-50)$$

将式 (7-48) 和 (7-49) 代入式 (7-50) 可得

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{2}{n} \ln \left\{ \frac{L_t(k+n-1)}{L_t(k-1)} \right\} &= \ln \left\{ \frac{\Sigma_{t^*}}{\Sigma_t} \right\} - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\tau=k}^{k+n-1} \frac{\tilde{y}^2(\tau|\tau, \theta_t)}{\Sigma_t} + \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\tau=k}^{k+n-1} \frac{\tilde{y}^2(\tau|\tau, \theta_{t^*})}{\Sigma_{t^*}} \\ &= \ln \left\{ \frac{\Sigma_{t^*}}{\Sigma_t} \right\} - \frac{\Sigma_{t^*}}{\Sigma_t} - \frac{M_t}{\Sigma_t} + 1 \end{aligned} \quad (7-51)$$

其中

$$M_t = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\tau=k}^{k+n-1} \tilde{y}^2(\tau|\tau, \theta_t) - \Sigma_{t^*} \quad (7-52)$$

显然对  $\forall x > 0$  有  $\ln(x) - x + 1 \leq 0$ 。因此根据式 (7-51)，且  $\Sigma_{t^*} > 0$  和  $\Sigma_t > 0$ ，可以得到不等式

$$\ln \left\{ \frac{\Sigma_{t^*}}{\Sigma_t} \right\} - \frac{\Sigma_{t^*}}{\Sigma_t} + 1 \leq 0 \quad (7-53)$$

根据式 (7-50) 和 (7-52) 可得  $M_t > 0$ 。将式不等式 (7-53) 和  $M_t > 0$  代入到式 (7-51) 中得

$$\lim_{n \rightarrow \infty} \frac{2}{n} \ln \left\{ \frac{L_t(k+n-1)}{L_t(k-1)} \right\} = -c_1 \quad (7-54)$$

其中  $c_1 > 0$ 。根据式 (7-54) 可得

$$\lim_{n \rightarrow \infty} L_t(k+n-1) = \lim_{n \rightarrow \infty} c_2 L_t(k-1) \exp\{-nc_1/2\} = 0, \quad \forall t \neq t^* \quad (7-55)$$

其中  $c_2$  为常数。结合式 (7-42) 和 (7-55) 得

$$\lim_{n \rightarrow \infty} \pi(\theta_t | \mathfrak{I}^k) \pi^{-1}(\theta_{t^*} | \mathfrak{I}^k) = 0, \quad \forall t \neq t^* \quad (7-56)$$

该式等价于

$$\lim_{n \rightarrow \infty} \pi(\theta_t | \mathfrak{I}^k) = 0, \quad \forall t \neq t^* \quad (7-57)$$

由于所有候选扰动的概率之和为 1，因此可得

$$\lim_{n \rightarrow \infty} \pi(\theta_{t^*} | \mathfrak{I}^k) = 1 \quad (7-58)$$

表明通过主动学习，扰动估计值会收敛到真值。因此所设计的控制律学习到了正确的扰动向量，并得到最优控制律。证毕。

**注释 7.1:** 定理 6.3 的证明表明, 如果  $\theta_t^*$  是扰动的最优估计值, 那么其后验概率从初值  $1/s_\alpha s_\beta s_\gamma$  按指数收敛到 1。根据扰动向量值  $\theta_t = [\theta_{\alpha_i}, \theta_{\beta_j}, \theta_{\gamma_l}]$  的定义,  $\theta_t$  包含了所有可能的组合。后验概率的收敛也表明扰动实际值  $\alpha(k)$ ,  $\beta(k)$  和  $\gamma(k)$  分别位于子区间  $[\alpha_{i-1}, \alpha_i]$ ,  $[\beta_{j-1}, \beta_j]$  和  $[\gamma_{l-1}, \gamma_l]$ 。

**注释 7.2:** 式 (6-26) 表明所设计的控制器包含  $s = s_\alpha s_\beta s_\gamma$  个子控制器, 每个子控制器都对应一个  $\theta_t$  ( $t \in \{1, 2, 3, \dots, s_\alpha s_\beta s_\gamma\}$ ), 控制律是每个子控制器的加权和, 权重为每个控制器对应的  $\theta_t$  的后验概率  $\pi(\theta_t | \mathfrak{I}^k)$ 。在这里最优的估计值  $\theta_t^*$  会收敛到 1, 因此控制律最终等于  $\theta_t^*$  对应的子控制器, 且该控制律就是当前的最优控制律。

**注释 7.3:** 假设在  $k$  时刻, 后验概率  $\pi(\theta_t^* | \mathfrak{I}^k)$  已经收敛到 1, 其他扰动值的后验概率  $\pi(\theta_t | \mathfrak{I}^k)$  收敛到 0。根据式 (6-27), 后验概率  $\pi(\theta_t^* | \mathfrak{I}^k)$  和  $\pi(\theta_t | \mathfrak{I}^k)$  将不会变化, 也就是被锁死, 即使实际的扰动  $\alpha(k)$ ,  $\beta(k)$  和  $\gamma(k)$  在  $k$  时刻之后发生了较大的变化, 那么控制律也会被锁死在  $u(k) \equiv u(k, \theta_t^*)$ 。因此有必要设计一个能够检测到扰动发生变化的准则, 确保系统不会被锁死。因此设计了一个检测方法如下。如果系统的输出估计误差  $\tilde{y}(k+1, \theta_t^*) > \epsilon$ , 这就表明扰动发生了变化, 其中  $\epsilon$  就是系统跟踪可容许误差。当检测到系统扰动发生变化, 就会将所有的后验概率全部赋初值  $1/s_\alpha s_\beta s_\gamma$  以保证主动学习重新启动。

## 7.4 仿真实验

本小节给出了在不同工况下的具有主动学习特征的抗扰动自适应对偶控制的仿真实验。7.4.1 小节对一个非线性数学模型进行仿真控制, 系统中的乘性干扰随时间而变化, 输出参考轨迹为连续余弦波。7.4.2 小节是对相同的模型进行仿真实验, 但扰动是加性扰动, 输出参考轨迹为方波。7.4.3 小节通过蒙特卡洛仿真实验, 在统计层面分析了该方法的有效性。在 7.4.4 小节对存在乘性和加性干扰的高速列车的速度控制进行了仿真分析, 并将仿真结果与理想最优控制和无模型自适应控制方法进行了比较。本章所提出的方法的具体执行步骤如下所示:

**初始化:**

初始化系数  $\varepsilon_\alpha > 0$ ,  $\varepsilon_\beta > 0$ ,  $\varepsilon_\gamma > 0$  和  $\epsilon > 0$ , 扰动向量有限候选集  $\Omega = \{\theta_t : t = 1, 2, \dots, s_\alpha s_\beta s_\gamma\}$ , 状态信息  $\mathfrak{I}^0$ , 贝叶斯后验概率  $\pi(\theta_t | \mathfrak{I}^0)$ , 对偶特性系数  $0 < \lambda < 1$ , 扰动估计误差协方差矩阵  $P(0)$ , 控制信号  $u^*(1)$ , 后验概率阈值  $\phi$ 。

**迭代过程:**

- (1) 对系统加入控制信号  $u^*(k)$ , 并观测系统输出  $y(k+1)$ ;
- (2) 根据式 (7-27) 与 (7-28) 计算  $k+1$  时刻的贝叶斯后验概率  $\pi(\theta_t | \mathfrak{I}^k)$ ;
- (3) 根据式 (7-31) 计算出  $k+1$  时刻的  $\theta_t$ -候选控制律  $u(k+1, \theta_t)$ ;
- (4) 根据式 (7-26) 计算出所提出的控制律  $u^*(k+1)$ ;

- (5) 计算  $k+1$  时刻的系统输出预测  $\hat{y}(k+1, \theta_t)$  和输出预测误差  $\tilde{y}(k+1, \theta_t)$ ;
- (6) 如果  $\tilde{y}(k+1, \theta_t) > \epsilon$  且  $\pi(\theta_t | \mathfrak{I}^k) > \phi$ , 则后验概率重置为  $\pi(\theta_t | \mathfrak{I}^k) = 1 / s_\alpha s_\beta s_\gamma$ ;
- (7) 根据式 (7-32) 更新扰动估计误差协方差矩阵  $P(k+1)$ ;
- (8) 重复步骤 (1) 到步骤 (7)。

#### 7.4.1 非线性系统受乘性扰动的影响

考虑如下非线性仿射系统

$$y(k+1) = \alpha(k)[\sin(x(k)) + \cos(3x(k))] + \beta(k)(2 + \cos(x(t)))u(k) + \gamma(k) + e(k+1) \quad (7-59)$$

其中乘性扰动  $\alpha(k)$  和  $\beta(k)$  分别在有界区间  $[0.75, 1.25]$  和  $[0.75, 1.05]$  中变化, 加性噪声为 0。系统状态为  $x(k) = y(k)$ , 系统噪声服从高斯分布  $e(k) \sim N(0, 0.0004)$ 。该系统有两个非线性函数,  $f[x(k)] = \sin(x(k)) + \cos(3x(k))$  和  $g[x(k)] = 2 + \cos(x(t))$ 。系统输出跟踪的参考轨迹为余弦波  $y_r(k) = \cos(5\pi k / 600)$ 。扰动  $\alpha(k)$  和  $\beta(k)$  的实际值变化情况如下

$$\alpha(k) = \begin{cases} 1.0 & \text{for } 1 \leq k < 85 \\ 1.11 & \text{for } 85 \leq k < 180 \\ 0.78 & \text{for } 180 \leq k < 340 \\ 0.91 & \text{for } 340 \leq k < 520 \\ 1.18 & \text{for } 520 \leq k \leq 600 \end{cases} \quad (7-60)$$

$$\beta(k) = \begin{cases} 0.9 & \text{for } 1 \leq k < 180 \\ 0.82 & \text{for } 180 \leq k < 340 \\ 1.0 & \text{for } 340 \leq k \leq 600 \end{cases} \quad (7-61)$$

首先需要使用历史输入输出数据来学习特殊神经网络的参数, 对未知非线性系统进行建模。对非线性函数  $f[x(k)]$ , 神经网络  $\hat{f}[x(k)]$  中的高斯径向基函数的中心点选为  $\hat{c}_f = [-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2]$ , 宽度为  $\hat{b}_f^2 = 1$ , 学习到的网络的权系数为  $\hat{w}_f = [11.2856, -4.6174, -12.3754, 1.5622, 12.0864, 2.2351, -11.9197, -4.4981, 12.6531]$ 。同样的, 对非线性函数  $g[x(k)]$ , 神经网络  $\hat{g}[x(k)]$  中的高斯径向基函数的中心点选为  $\hat{c}_g = [-2, 0, 2]$ , 宽度为  $\hat{b}_g^2 = 3.6$ , 网络的权系数为  $\hat{w}_g = [-0.4449, 3.5097, -0.4449]$ 。

根据定理 7.3, 可容许的估计误差值设置为  $\varepsilon_\alpha = 0.1$  和  $\varepsilon_\beta = 0.1$ , 则对应的子区间的个数为  $s_\alpha = 5$  和  $s_\beta = 3$ 。乘性扰动的近似有限集为  $\Omega_\alpha = \{0.8, 0.9, 1.0, 1.1, 1.2\}$  和  $\Omega_\beta = \{0.8, 0.9, 1.0\}$ 。扰动向量的有限集中元素一共有 15 个, 对应的元素分别为  $\theta_1 = [0.8, 0.8, 0]$ ,  $\theta_2 = [0.8, 0.9, 0]$ ,  $\theta_3 = [0.8, 1.0, 0]$ ,  $\theta_4 = [0.9, 0.8, 0]$ ,  $\theta_5 = [0.9, 0.9, 0]$ ,  $\theta_6 = [0.9, 1.0, 0]$ ,  $\theta_7 = [1.0, 0.8, 0]$ ,  $\theta_8 = [1.0, 0.9, 0]$ ,  $\theta_9 = [1.0, 1.0, 0]$ ,  $\theta_{10} = [1.1, 0.8, 0]$ ,  $\theta_{11} = [1.1, 0.9, 0]$ ,  $\theta_{12} = [1.1, 1.0, 0]$ ,  $\theta_{13} = [1.2, 0.8, 0]$ ,  $\theta_{14} = [1.2, 0.9, 0]$ ,  $\theta_{15} = [1.2, 1.0, 0]$ 。每个扰动向量对应的贝叶斯后验概率的初值设置为  $\pi(\theta_t | \mathfrak{I}^0) = 1/15$ ,  $t = 1, 2, \dots, 15$ 。防锁死判据的参数设置为  $\epsilon = 0.08$ 。扰动估计误差方差的初值设置为  $P(1) = I_{3 \times 3}$ , 对偶特性参数

设置为  $\lambda = 0.9$ 。

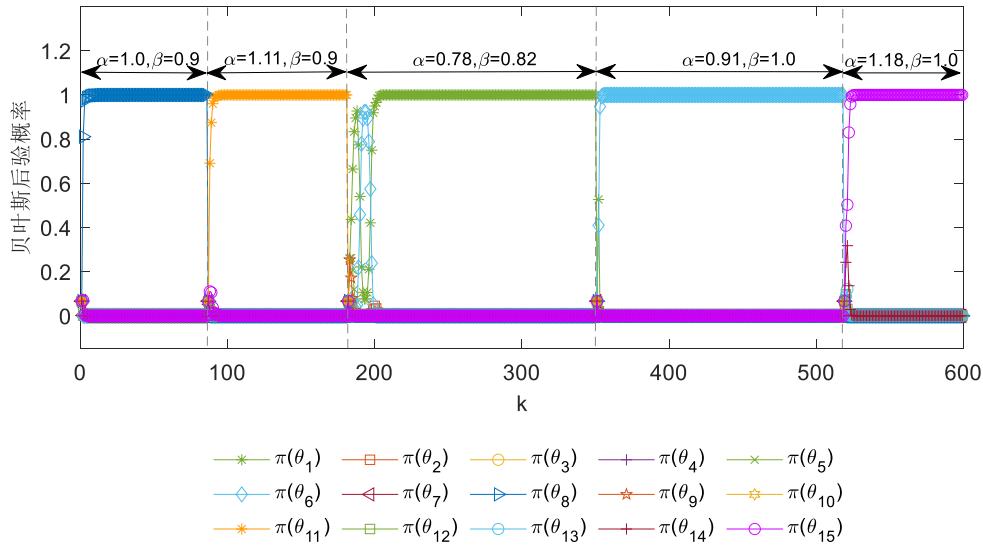


图 7-3 贝叶斯后验概率的收敛过程  
Fig.7-3 The convergence of Bayesian posterior probability

图 7-3 是所有贝叶斯后验概率的收敛过程。由图可得，当乘性扰动为  $\alpha=1.0$  和  $\beta=0.9$  时， $\pi(\theta_8)$  从  $1/15$  收敛到  $1$ ，其它的后验概率  $\pi(\theta_t)$  ( $t \neq 8$ ) 从  $1/15$  收敛到  $0$ 。当乘性扰动  $\alpha(k)$  在第 85 步从  $1.0$  变为  $1.11$  时，所有的贝叶斯后验概率重置为  $1/15$ ，表明此变化被成功检测到，然后主动学习重新开始运行，使得后验概率  $\pi(\theta_8)$  快速收敛到  $0$ ， $\pi(\theta_{11})$  收敛到  $1$ 。从第 180 步到第 340 步中，乘性扰动变为  $\alpha=0.78$  和  $\beta=0.82$ ，后验概率  $\pi(\theta_1)$  收敛到  $1$ ，这意味着乘性扰动  $\alpha(k)$  和  $\beta(k)$  都位于子区间  $[0.75, 0.85]$ ，因此扰动估计值为  $0.8$ 。在第 340 步扰动变为  $\alpha=0.91$  和  $\beta=1.0$  时， $\pi(\theta_6)$  收敛到  $1$ ，对应于  $\alpha(k)$  和  $\beta(k)$  位于子区间  $[0.85, 0.95]$  和  $[0.95, 1.05]$ ，则相应的扰动估计值分别为  $0.9$  和  $1.0$ 。在第 520 步，后验概率  $\pi(\theta_{15})$  收敛到  $1$ ，其他概率收敛到  $0$ ，相应的扰动估计值为  $1.2$  和  $1.0$ 。

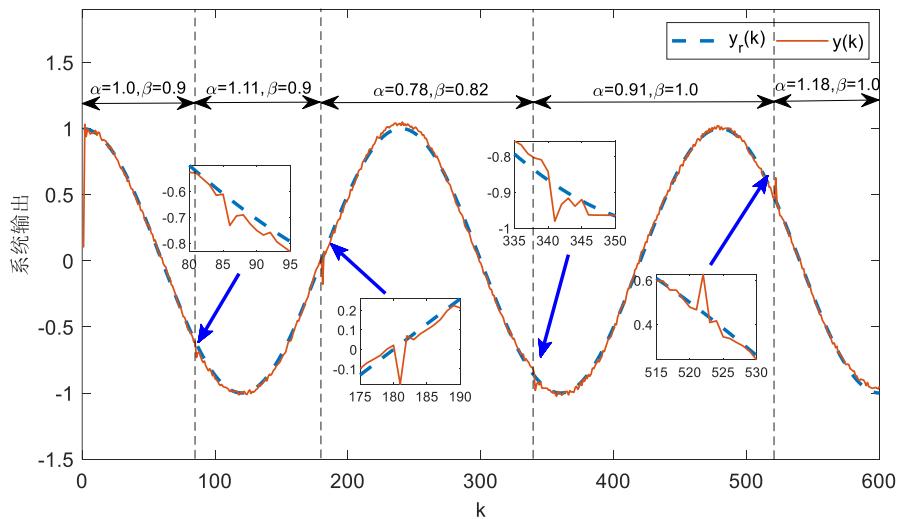


图 7-4 系统输出跟踪性能  
Fig.7-4 The system output tracking performance

图 7-4 是在抗扰动对偶控制方法下系统输出跟踪轨迹。图中的蓝色虚线表示参考轨迹，红线是使用所提出的控制策略的系统输出。尽管有未知的干扰存在，系统输出仍然能够精确地跟踪参考轨迹，除了在干扰变化后的 2 到 3 步内有尖峰，这个尖峰是不可避免的，因为当扰动发生变化时，后验概率需要一定的时间才能收敛到合适的值，那么系统相应的需要一定的时间才能重新跟踪上目标轨迹。

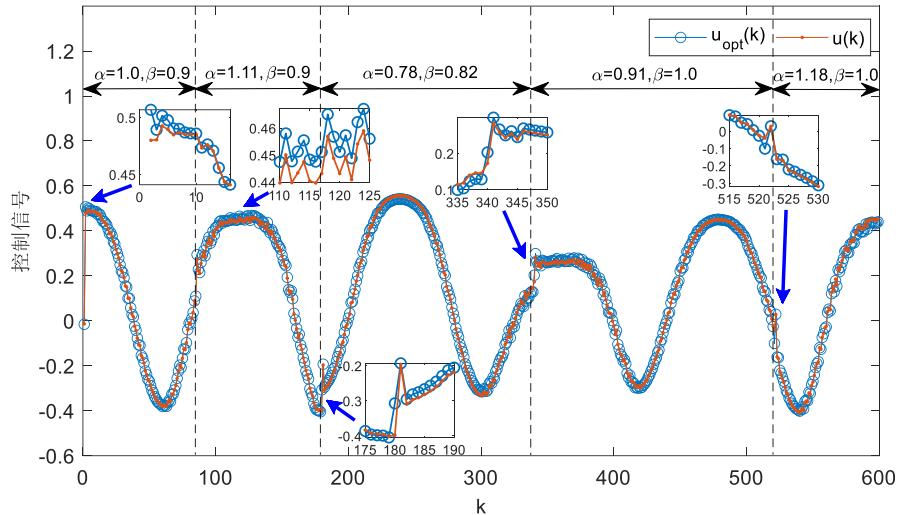


图 7-5 系统控制输入信号  
Fig.7-5 The system control input signal

图 7-5 是所提出的具有主动学习特性的对偶控制方法下产生的控制信号与最优控制方法产生的控制信号的对比。当系统的参数和扰动完全已知的情况下，就能够推导出最优控制律，作为用于对照的控制策略。与本章所提出的具有主动学习特性的对偶控制方法相比，由于最优控制方法没有扰动引起的不确定性，该最优控制可以作为理想参考基准策略。如图 7-5 所示，正如预期的那样，从第 1 步到第 85 步，控制信号快速收敛到最优控制信号，同时也表明学习到的扰动接近于真值。图 6-5 中的子图显示了所提出的控制律与最优控制律在第 1 步、第 85 步、第 180 步、第 340 步和第 520 步前后的偏差，这些偏差发生在扰动发生变化的地方。另外还需要注意的是，如果估计出的扰动不完全等于真值，那么控制信号的偏差会更大。例如，从第 180 步到第 340 步，扰动的真实值为  $\alpha = 0.78$  和  $\beta = 0.82$ ，而学习到的近似值分别为  $\hat{\alpha} = 0.8$  和  $\hat{\beta} = 0.8$ 。这里所提出的方法与最优控制的偏差可以通过将有界扰动划分为更多的子区间来减少，用来获得更精确的扰动估计值，这将在 7.4.3 小节中进一步讨论。

#### 7.4.2 非线性系统受加性扰动的影响

本实验的被控对象为式 (7-59) 描述的系统。假设加性扰动  $\gamma(k)$  处于有界区间  $[-1.45, 0.55]$ ，乘性扰动为  $\alpha(k)=1$  和  $\beta(k)=1$ ，系统噪声服从高斯分布  $e(k) \sim N(0, 0.0025)$ 。

系统输出参考轨迹为方波

$$y_r(k) = \begin{cases} 1 & \text{for } 1 \leq k < 150 \\ -1 & \text{for } 150 \leq k < 300 \\ 1 & \text{for } 300 \leq k < 450 \\ -1 & \text{for } 450 \leq k \leq 600 \end{cases} \quad (7-62)$$

加性扰动为

$$\gamma(k) = \begin{cases} -0.2 & \text{for } 1 \leq k < 200 \\ 0.32 & \text{for } 200 \leq k < 400 \\ -0.85 & \text{for } 400 \leq k < 500 \\ 0.5 & \text{for } 500 \leq k \leq 600 \end{cases} \quad (7-63)$$

可容许加性扰动近似误差设为  $\varepsilon_\gamma = 0.1$ ，则子区间个数为  $s_\gamma = 20$ 。相应的扰动候选值有限集是  $\Omega_\gamma = \{-1.4, -1.3, -1.2, -1.1, -1.0, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5\}$ ，扰动向量集  $\Omega$  中的元素个数为 20。贝叶斯后验概率初值设置为  $\pi(\theta_t | \mathfrak{I}^0) = 1/20$ ， $t = 1, 2, \dots, 20$ 。防锁死判据的参数设置为  $\epsilon = 0.5$ 。扰动估计误差协方差初值设为  $P(1) = I_{3 \times 3}$ ，对偶控制系数设为  $\lambda = 0.9$ 。

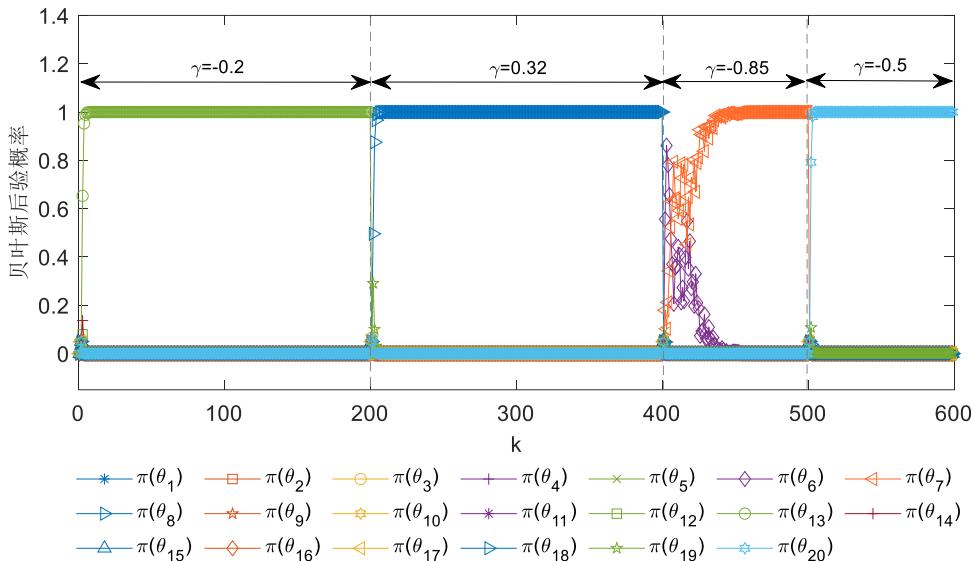


图 7-6 贝叶斯后验概率的收敛图  
Fig.7-6 The convergence of Bayesian posterior probability

图 7-6 描述了所提出的方法下每个候选扰动向量对应的贝叶斯后验概率收敛过程。图中，在  $k \in [1, 200]$  期间，加性扰动为  $\gamma = -0.2$ ，后验概率  $\pi(\theta_{13})$  从  $1/20$  收敛到 1，后验概率  $\pi(\theta_t)$  ( $t \neq 13$ ) 从  $1/20$  收敛到 0，这意味着扰动  $\gamma(k)$  接近于实际值 -0.2。在第 200 步扰动  $\gamma(k)$  从 -0.2 变化到 0.32，一旦系统检测到变化，所有后验概率都重置为  $\pi(\theta_i) = 1/20$ 。在第 200 步扰动变化后，后验概率  $\pi(\theta_{18})$  在 5 步之后收敛到 1。这表明加性扰动  $\gamma(k)$  位于子区间  $[-0.35, -0.25]$  中，那么  $\gamma(k) = 0.32$  近似等于 0.3。从第 400 步到第 500 步的扰动为-

0.85, 即在子区间  $[-0.95, -0.85]$  和  $[-0.85, -0.75]$  之间的边缘点。因此, 扰动可能收敛到子区间  $[-0.95, -0.85]$  或  $[-0.85, -0.75]$ 。在此期间的扰动可能近似等于-0.9 或-0.8。正如预期的那样, 在第 400 步后  $\pi(\theta_6)$  和  $\pi(\theta_7)$  在 0 和 1 之间波动, 最后  $\pi(\theta_6)$  收敛到 0,  $\pi(\theta_7)$  收敛到 1。扰动  $\gamma(k)$  在第 500 步变为-0.5, 随后  $\pi(\theta_{20})$  收敛到 1。

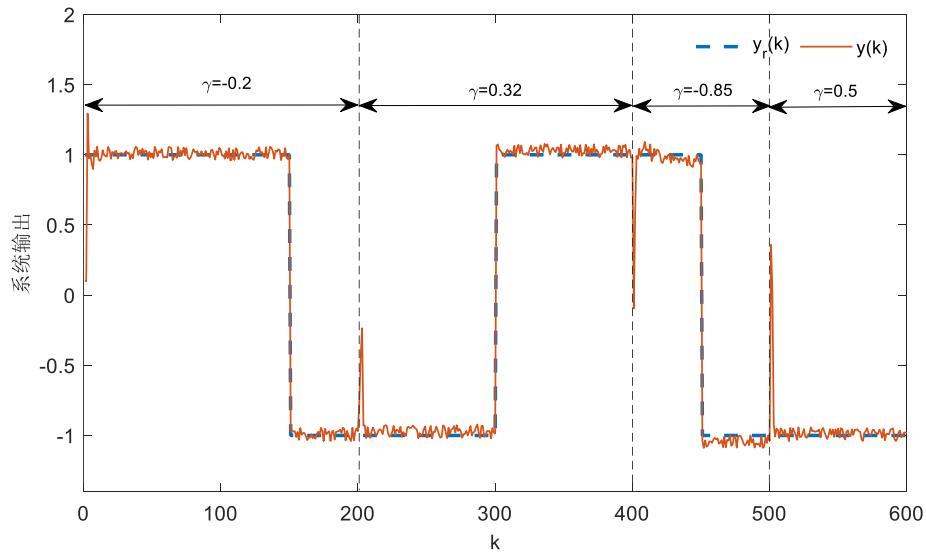


图 7-7 系统输出跟踪轨迹  
Fig.7-7 The system output tracking performance

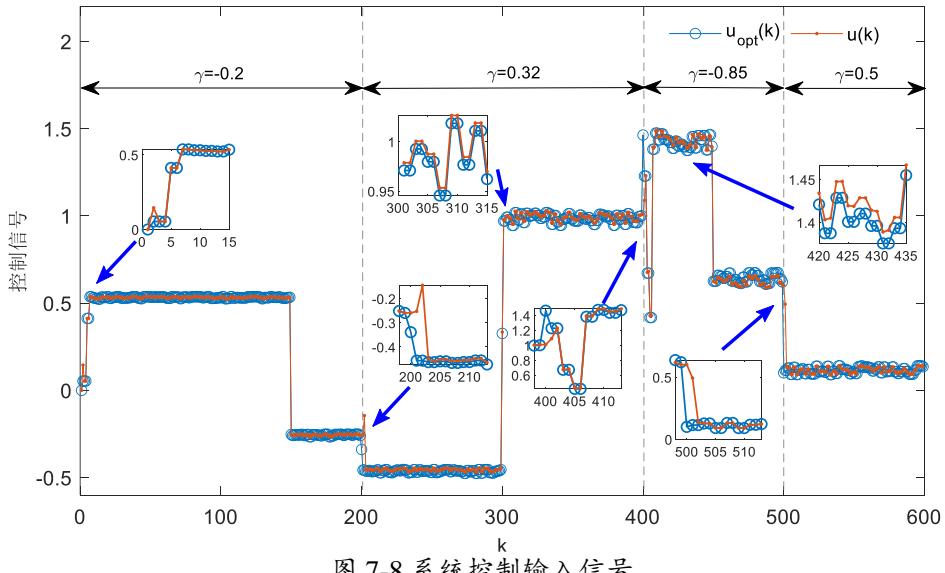


图 7-8 系统控制输入信号  
Fig.7-8 The system control input signal

图 7-7 是系统输出跟踪方波的轨迹。由图可知, 所提出的具有主动学习特性的对偶控制方法对阶跃参考信号的跟踪控制同样有良好的跟踪效果, 系统输出能准确地跟踪参考轨迹, 除了在每次在扰动发生变化之后的 5 步内出现了一个尖峰。这个尖峰也是不可避免地, 这是由于扰动的后验概率在扰动变化后需要利用系统的输入输出数据重新进行

扰动的学习，因此经过一段时间扰动估计值才能学习到正确的值。

图 7-8 将本章所提出的具有主动学习特性的对偶控制方法与理想基准最优控制信号进行了比较。由图可得，当扰动在第 1 步和第 500 步发生变化时，控制信号在 4 步内收敛到最优控制信号。扰动  $\gamma(k)$  在第 200 步从 -0.2 变化到 0.32，对偶控制方法的控制信号收敛到  $\hat{\gamma} = 0.3$  的控制律，且接近最优控制信号。在子图中详细体现了该方法与理想基准最优控制信号之间的偏差。当加性扰动在第 400 步变为  $\gamma(k) = -0.85$  时，控制信号在  $\theta_6$ -candidate 和  $\theta_7$ -candidate 控制信号之间波动，并逐渐收敛到  $\theta_6$ -candidate 控制信号，该控制信号也接近于基准控制信号。

### 7.4.3 在不同的估计误差容许值下的蒙特卡洛仿真实验

通过蒙特卡洛仿真对 7.4.1 和 7.4.2 中的实验进行统计研究，实验参数的设置与前两个实验一致，除了针对扰动  $\alpha$ 、 $\beta$  和  $\gamma$  的学习设置的容许估计误差  $\varepsilon_\alpha$ 、 $\varepsilon_\beta$  和  $\varepsilon_\gamma$ 。本节通过改变这些参数来分析不同的扰动估计分辨率是如何影响抗扰动对偶控制方法的控制性能。

对 7.4.1 的实验进行蒙特卡洛仿真，使用不同的参数  $\varepsilon_\alpha = \varepsilon_\beta = 0.05$ 、 $\varepsilon_\alpha = \varepsilon_\beta = 0.075$ 、 $\varepsilon_\alpha = \varepsilon_\beta = 0.1$  和  $\varepsilon_\alpha = \varepsilon_\beta = 0.2$  来测试所提出的方法。当参数  $\varepsilon_\alpha = \varepsilon_\beta = 0.05$  时， $\alpha(k)$  和  $\beta(k)$  对应的子区间数分别为 10 和 6。 $\alpha(k)$  和  $\beta(k)$  对应的扰动候选值有限集为  $\Omega_\alpha = \{0.775, 0.825, 0.875, 0.925, 0.975, 1.025, 1.075, 1.125, 1.175, 1.225\}$  和  $\Omega_\beta = \{0.775, 0.825, 0.875, 0.925, 0.975, 1.025\}$ 。当参数  $\varepsilon_\alpha = \varepsilon_\beta = 0.075$  时， $\alpha(k)$  和  $\beta(k)$  的子区间数分别为 7 和 4。 $\alpha(k)$  和  $\beta(k)$  对应的扰动候选值有限集为  $\Omega_\alpha = \{0.7857, 0.8571, 0.9285, 0.9999, 1.0713, 1.1427, 1.2142\}$  和  $\Omega_\beta = \{0.7875, 0.8625, 0.9375, 1.0125\}$ 。当参数  $\varepsilon_\alpha = \varepsilon_\beta = 0.1$  时， $\alpha(k)$  和  $\beta(k)$  的子区间数分别为 5 和 3。 $\alpha(k)$  和  $\beta(k)$  对应的扰动候选值有限集为  $\Omega_\alpha = \{0.8, 0.9, 1.0, 1.1, 1.2\}$  和  $\Omega_\beta = \{0.8, 0.9, 1.0\}$ 。当参数  $\varepsilon_\alpha = \varepsilon_\beta = 0.2$  时， $\alpha(k)$  和  $\beta(k)$  的子区间数分别为 3 和 2。 $\alpha(k)$  和  $\beta(k)$  对应的扰动候选值有限集为  $\Omega_\alpha = \{0.8334, 1.0001, 1.1667\}$  和  $\Omega_\beta = \{0.825, 0.975\}$ 。由于  $\varepsilon_\alpha$  和  $\varepsilon_\beta$  的值不同，导致子区间长度不同，扰动估计的分辨率不同。

这里用不同的容许估计误差  $\varepsilon_\alpha$  和  $\varepsilon_\beta$  对系统进行 100 次蒙特卡洛仿真实验，其中实际扰动值  $\alpha(k)$  和  $\beta(k)$  在有界区间  $[0.75, 1.25]$  和  $[0.75, 1.05]$  中随机取值。为了量化 100 次蒙特卡洛仿真下的该方法控制性能，定义如下平均性能指标

$$J_M = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \left( \frac{1}{N} \sum_{k=1}^N \sqrt{[y(k) - y_r(k)]^2} \right)_i \quad (7-64)$$

其中  $N_{MC}$  是蒙特卡洛仿真次数。

表 7-1 是参数  $\varepsilon_\alpha$  和  $\varepsilon_\beta$  取不同值时的蒙特卡洛仿真实验结果。

表 7-1 参数  $\varepsilon_\alpha$  和  $\varepsilon_\beta$  取不同值时的平均性能指标和运行时间Tab.7-1 The average performance index and running time with different  $\varepsilon_\alpha$  and  $\varepsilon_\beta$ 

$\varepsilon_\alpha$ , $\varepsilon_\beta$	$s_\alpha s_\beta$	$J_M$	$t_r(s)$
0.05	60	0.0022	0.4110
0.075	28	0.0027	0.1817
0.1	15	0.0029	0.0954
0.2	6	0.0042	0.0392
最优控制		0.0018	0.0036

由表 7-1 可知,  $\varepsilon_\alpha = \varepsilon_\beta = 0.05$  时的平均性能指标是最低的, 且平均性能指标随参数  $\varepsilon_\alpha$  和  $\varepsilon_\beta$  的增加而增加。由此可以得出  $\varepsilon_\alpha = \varepsilon_\beta = 0.05$  的控制性能最好, 且最接近理想基准最优控制。当  $\varepsilon_\alpha$  和  $\varepsilon_\beta$  值较小的情况下系统的控制更为精确, 但这是以更多的计算量为代价的。如表所示, 系统控制控制律计算时间  $t_r$  随候选扰动数量  $s_\alpha$  和  $s_\beta$  的增长而线性增长。

对 7.4.2 节的实验进行蒙特卡洛仿真, 使用不同的参数  $\varepsilon_\gamma = 0.05$ ,  $\varepsilon_\gamma = 0.1$ ,  $\varepsilon_\gamma = 0.2$ ,  $\varepsilon_\gamma = 0.4$ 。当  $\varepsilon_\gamma = 0.05$  时,  $\gamma(k)$  的子区间个数为 40, 对应的扰动候选值有限集为  $\Omega_\gamma = \{-1.425, -1.375, -1.275, -1.225, -1.175, -1.125, -1.075, -1.025, -0.975, -0.925, -0.875, -0.825, -0.775, -0.725, -0.675, -0.625, -0.575, -0.525, -0.475, -0.425, -0.375, -0.325, -0.275, -0.225, -0.175, -0.275, -0.075, 0.125, 0.175, 0.225, 0.275, 0.325, 0.375, 0.425, 0.475, 0.525\}$ 。当参数  $\varepsilon_\gamma = 0.1$  时,  $\gamma(k)$  的子区间个数为 20, 对应的扰动候选值有限集为  $\Omega_\gamma = \{-1.4, -1.3, -1.2, -1.1, -1.0, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5\}$ 。当参数  $\varepsilon_\gamma = 0.2$  时,  $\gamma(k)$  的子区间个数为 10, 对应的扰动候选值有限集为  $\Omega_\gamma = \{-1.35, -1.15, -0.95, -0.75, -0.55, -0.35, -0.15, 0.05, 0.25, 0.45\}$ 。当参数  $\varepsilon_\gamma = 0.4$  时,  $\gamma(k)$  的子区间个数为 5, 对应的扰动候选值有限集为  $\Omega_\gamma = \{-1.25, -0.85, -0.45, -0.05, 0.35\}$ 。

表 7-2 参数  $\varepsilon_\gamma$  取不同值时的平均性能指标和运行Tab.7-2 The average performance index and running time with different  $\varepsilon_\gamma$ 

$\varepsilon_\gamma$	$s_\gamma$	$J_M$	$t_r(s)$
0.05	40	0.0030	0.3021
0.1	20	0.0033	0.1472
0.2	10	0.0037	0.0732
0.4	5	0.0054	0.03
最优控制		0.0024	0.0036

对系统进行 100 次蒙特卡洛仿真实验, 加性扰动  $\gamma(k)$  在有界区间  $[-1.45, 0.55]$  内随机取值。表 7-2 是不同  $\varepsilon_\gamma$  下的蒙特卡洛仿真实验结果。正如预期的那样,  $\varepsilon_\gamma = 0.05$  时的平均性能指标最低的, 最接近最优控制。而系统的平均性能指标会随着参数  $\varepsilon_\gamma$  的增加而增

加, 表明  $\varepsilon_\gamma$  较低时的控制性能更好。类似于上一个蒙特卡洛仿真实验, 可以得出同样的结论, 较低的  $\varepsilon_\gamma$  能带来更多的扰动候选值, 但是需要更多的计算量从而增加了系统运行的时间。

#### 7.4.4 高速列车的速度控制

本实验用所提出的具有主动学习特点的抗扰动对偶控制策略对 CRH3 高速列车的速度控制进行仿真实验, 并且将该方法与另外两种控制方法进行对比, 即理想基准最优控制和近些年出现的无模型自适应控制方法。列车的速度控制模型如下

$$\begin{aligned} v(k+1) &= v(k) + \xi T \{ f(k) - W[v(k)] \} \\ W[v(k)] &= c_r + c_m v(k) + c_a v^2(k) \end{aligned} \quad (7-65)$$

其中  $v(k)$  是高速列车的速度,  $f(k)$  是牵引力/制动力,  $\xi$  是加速度系数,  $T$  是采样间隔,  $W[v(k)]$  是阻力, 包括滚动阻力  $c_r$ , 机械阻力  $c_m v(k)$  和气动阻力  $c_a v^2(k)$ 。高速列车的速度控制是通过调节牵引力或制动力使列车保持在一个目标的运行速度。然而, 列车的轨道状况异常、天气和气候的变化、机械磨损和老化等都可能对列车的速度控制产生不利影响。除此之外, 系数  $\xi$ 、 $c_r$ 、 $c_m$  和  $c_a$  都不是常数, 会受到复杂的外部环境和系统内部的干扰。

首先用特殊神经网络 SNN 来对该非线性系统进行建模

$$v(k+1) = \alpha(k) \{ v(k) - \xi T [c_r + c_m v(k) + c_a v^2(k)] \} + \beta(k) \xi T f(k) + \gamma(k) + e(k+1) \quad (7-66)$$

其中包括乘性扰动  $\alpha(k)$ 、 $\beta(k)$  和加性扰动  $\gamma(k)$ 。由于列车模型参数是未知的, SNN 使用两个非线性函数来对系统中的非线性部分进行建模, 相应的非线性函数的建模分别表示为  $\hat{f}[x(k)] \approx v(k) - \xi T [c_r + c_m v(k) + c_a v^2(k)]$  和  $\hat{g}[x(k)] \approx \xi T$ 。函数  $\hat{f}[x(k)]$  和  $\hat{g}[x(k)]$  中的参数可以利用历史观测数据进行离线学习得到。

本实验假设扰动分别位于有界区间  $[0.725, 1.075]$ ,  $[0.8, 1.2]$  和  $[-13, 2]$  内。随机噪声服从高斯分布  $e(k) \sim N(0, 1)$ 。在仿真实验中, 参数设置为  $\xi = 0.06$ ,  $c_r = 0.1$ ,  $c_m = 0.0064$ ,  $c_a = 0.000115$ , 采样间隔为  $T = 0.1$ 。其中变化的扰动设置为

$$\alpha(k) = \begin{cases} 1.0 & \text{for } 1 \leq k < 100 \\ 1.05 & \text{for } 100 \leq k < 250 \\ 0.95 & \text{for } 250 \leq k < 350 \\ 0.92 & \text{for } 350 \leq k < 500 \\ 0.95 & \text{for } 500 \leq k \leq 600 \end{cases} \quad (7-67)$$

$$\beta(k) = \begin{cases} 1.0 & \text{for } 1 \leq k < 100 \\ 0.9 & \text{for } 100 \leq k < 350 \\ 1.0 & \text{for } 350 \leq k < 500 \\ 1.15 & \text{for } 500 \leq k \leq 600 \end{cases} \quad (7-68)$$

$$\gamma(k) = \begin{cases} 0.0 & \text{for } 1 \leq k < 350 \\ -12.5 & \text{for } 350 \leq k < 500 \\ -2.25 & \text{for } 500 \leq k \leq 600 \end{cases} \quad (7-69)$$

参考速度轨迹设置为  $270 + 50(1 + \exp(-2k))$ 。

仿真参数设置为  $\varepsilon_\alpha = 0.05$ 、 $\varepsilon_\beta = 0.4$  和  $\varepsilon_\gamma = 1$ ，对应的扰动候选子区间个数分别为  $s_\alpha = 7$ 、 $s_\beta = 1$  和  $s_\gamma = 15$ 。扰动候选值有限集为  $\Omega_\alpha = \{0.75, 0.80, 0.85, 0.90, 0.95, 1.00, 1.05\}$ 、 $\Omega_\beta = \{1.00\}$  和  $\Omega_\gamma = \{-12.5, -11.5, -10.5, -9.5, -8.5, -7.5, -6.5, -5.5, -4.5, -3.5, -2.5, -1.5, -0.5, 0.5, 1.5\}$ 。候选扰动向量  $\theta_t$  的个数为 105，贝叶斯后置概率初值设置为  $\pi(\theta_t | \mathfrak{I}^0) = 1/105$ ， $t = 1, 2, \dots, 105$ 。防锁死判据参数设置为  $\epsilon = 1.5$ 。扰动估计误差协方差初值设为  $P(1) = I_{3 \times 3}$ ，对偶控制系数设为  $\lambda = 0.9$ 。

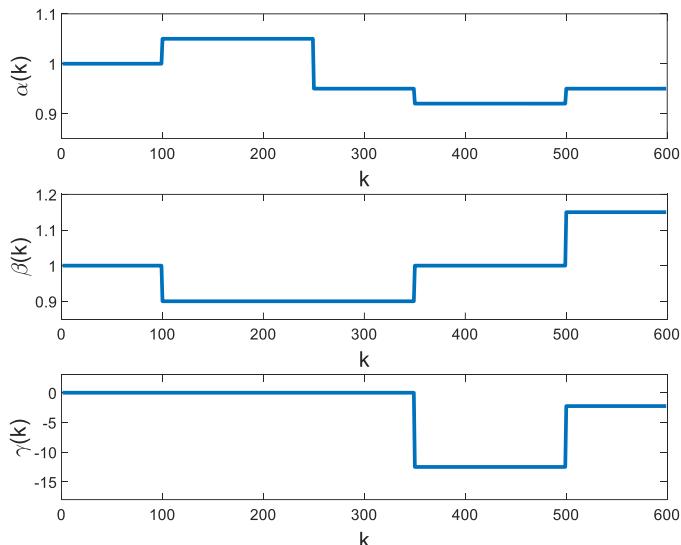


图 7-9 乘性扰动和加性扰动  
Fig. 7-9 The multiplicative and additive disturbances

图 7-9 是扰动  $\alpha(k)$ 、 $\beta(k)$  和  $\gamma(k)$  随时间变化图。

图 7-10 对比了所提出的方法下系统输出  $y(k)$  与理想基准最优控制和无模型自适应控制下的系统输出。由图可得，最优控制具有更好的输出跟踪性能，因为被控系统的干扰是完全已知的。本章所提出的控制方法下，除了扰动变化后的 3 到 4 步内出现尖峰外，其他时刻均表现出良好的跟踪控制性能。相比之下，无模型自适应控制具有较大的超调和较长的调节时间。

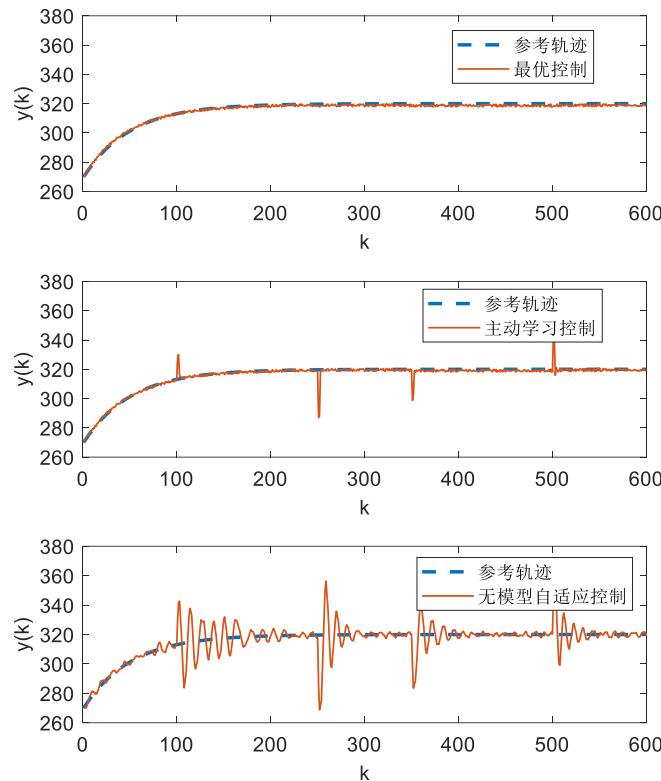


图 7-10 最优控制，主动学习控制和无模型自适应控制下的系统输出轨迹

Fig.7-10 The system output of the high-speed train by the optimal control, the proposed approach, and the model-free adaptive control

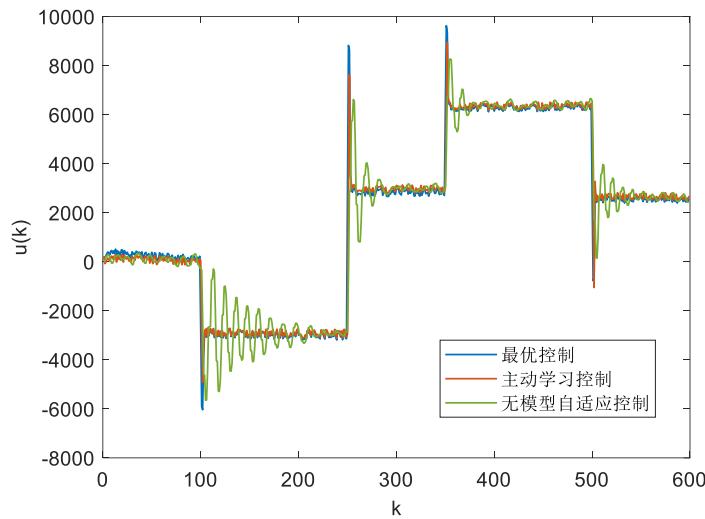


图 7-11 最优控制，主动学习控制和无模型自适应控制下的系统控制信号

Fig.7-11 The system control input signals by the optimal control, the proposed approach, and the model-free adaptive control

图 7-11 将本章所提出的控制方法的控制信号与理想基准最优控制和无模型自适应控制进行了比较，从图中可以看到，所提出方法提供了更接近最优控制的控制信号，并且比无模型自适应控制更能抵御干扰的影响。

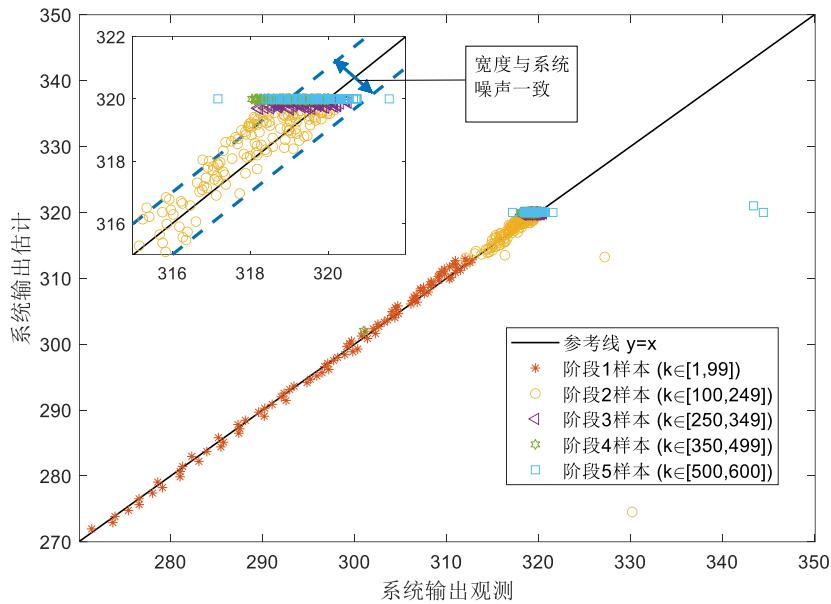


图 7-12 系统观测输出和估计输出之间的比较

Fig.7-12 The comparison of measured system output and the estimated system output

图 7-12 利用主动学习到的扰动估计系统输出  $\hat{y}(k)$  与真值进行比较, 可以看到系统的大部分估计输出与真实输出接近。估计的系统输出和真值之间的差距大多在  $[-1,1]$  内, 表明了主动学习的有效性, 所提出的方法对受扰动非线性系统进行了精确拟合。

## 7.5 本章小结

本章研究了具有不可测量的动态扰动的未知非线性系统的控制问题。针对这类系统, 本章提出了一种抗扰动对偶控制方法, 该方法通过主动实时学习扰动, 并将学习到的扰动用于生成对偶控制律, 减少了扰动对控制性能的影响。在该控制方法的设计中, 考虑了乘性和加性扰动的影响, 设计了特殊神经网络来对未知的系统和干扰进行建模。此外, 该方法在执行扰动学习和系统输出跟踪的同时, 设计了扰动主动学习与输出跟踪控制解耦的对偶控制方法, 从而增强了系统在出现突变扰动时控制的鲁棒性。最后通过对数学模型的仿真实验和对高速列车速度控制的仿真实验, 分析证明了本章所提出的具有主动学习特性的对偶控制方法的有效性。



## 8 总结与展望

### 8.1 本文工作总结

本文在自适应对偶控制方法的框架下，研究了具有不同类型的不确定因素的随机系统的控制问题，其中包括参数未知的线性系统，被孤立点干扰的系统，被具有尖峰、厚尾、非对称特点的随机噪声污染的系统，完全未知的具有高斯白噪声的非线性系统、具有不可测动态扰动的非线性系统。针对以上不同类型的系统，对自适应对偶控制方法设计做出了进一步研究。主要工作内容如下：

(1) 针对参数未知的线性系统，在基于强化 Q-学习的自适应控制的框架下，研究了控制器设计中探索-利用的平衡问题。与传统的方法不同的是，所提出的基于强化 Q-学习的自适应控制器设计中，借鉴了自适应对偶控制的思想，不仅考虑到了系统状态最优跟踪性能指标，还考虑到了估计误差问题，从而解决了强化学习自适应控制中的探索与利用的冲突问题，以达到最优的控制效果。通过对一阶和二阶线性系统的数值仿真，验证了所提出的控制策略对系统控制性能的改善。

(2) 针对受孤立点干扰的线性系统，提出了一种具有孤立点检测的自适应对偶控制方法，该方法对不确定系统的观测数据以及系统控制过程中出现孤立点噪声时具有较强的鲁棒性。该方法通过对参与参数估计的数据进行有效的孤立点检测，并用期望值替换孤立点来提高参数的估计精度，从而提高系统控制鲁棒性。该孤立点检测算法的计算量小，实时性强，适用于实际的工业控制过程。此外，在模型中还考虑到了不可控激励对控制性能的影响，并将其嵌入到自适应对偶控制的框架中，使其更适合实际应用场景。数值仿真的结果表明，无论在观测噪声还是在系统噪声中出现孤立点，所提出的方法都具有较高的鲁棒性。另外在生物发酵连续灭菌过程的仿真对比实验中，该方法与其他滤波方法下的控制结果相比具有更优的控制性能。

(3) 针对被具有尖峰、厚尾、非对称特性的随机噪声污染的系统，提出了具有贝叶斯分位数求和估计器的自适应对偶控制方法。所设计的贝叶斯分位数求和估计可以在控制过程中对模型参数进行实时估计，并且当系统受到此类噪声污染时具有更好的估计效果。在系统启动阶段，即使给模型参数设置任意初值，该参数估计器中不同分位数下的贝叶斯后验概率依旧能够快速收敛，即参数估计器能找到最优的分位点进行参数估计，最终能够为系统的控制带来更优的控制性能。数学模型的仿真结果表明，与其他估计器相比，贝叶斯分位数求和估计器对被具有尖峰、厚尾、非对称特性的随机噪声污染的系统的参数估计具有更好的估计性能，并且在控制过程中带来更好的输出跟踪控制。

(4) 针对一个完全未知的且具有高斯白噪声的随机非线性系统，提出了一种基于自动分配资源的神经网络自适应对偶控制方法。首先设计一个在非线性仿射模型框架下的可以自动分配资源的神经网络模型，然后基于该网络模型结构，设计对应的自适应对偶

控制器，能一方面利用测量的系统输入-输出数据，实时更新可自动分配资源的神经网络模型的参数。另一方面使用信息熵的概念来描述系统信息增量，在代价函数中增加系统学习的信息增量来控制未知信息的学习效率，从而得到具有主动学习特性的自适应对偶控制律，仿真实验表明所设计的自适应对偶控制器对未知非线性系统具有良好控制性能。

(5) 针对受乘性扰动和加性扰动干扰的非线性系统，提出了具有主动学习特性的抗扰动对自适应对偶控制方法。该方法设计了特殊神经网络结构来对未知的非线性系统和不可测动态扰动进行建模。然后基于该特殊神经网络，设计了能对未知扰动学习与输出跟踪控制解耦的对偶控制方法，增强了系统在出现突变扰动时系统控制的鲁棒性，最后通过对数学模型和对高速列车速度控制的仿真实验，说明了该控制方法的有效性。

## 8.2 未来工作展望

本文针对具有不同种类不确定性的系统的自适应对偶控制方法进行研究。但是，目前的工作仍然由一些不够完善的地方，成为未来需要进一步研究的内容。

(1) 在具有孤立点检测的自适应对偶控制方法研究和具有贝叶斯分位数求和估计器的自适应对偶控制方法研究中，都使用了离散线性差分方程来拟合工业控制过程，而通常工业控制过程为非线性高耦合的系统，因此存在一定的建模误差，在一定程度上影响系统的控制性能。在今后的研究中，可以尝试使用非线性模型或者神经网络模型对实际系统进行拟合，以提高控制性能指标。而且这两种方法研究的被控对象均为单输入-单输出的系统，然而在许多实际控制过程中，被控对象通常为多输入-多输出系统，因此可以将这两种方法推广成多输入-多输出系统的自适应对偶控制方法，使其在实际系统控制过程中得到更广泛的应用。

(2) 在文章中关于对偶控制问题求解时，对于最优估计和最优控制之间的平衡，所使用的系数  $\beta$  或者  $\rho$  是根据经验设置的值。因此有必要设计算法，能够在控制过程中实时选取最优的参数，进一步提高控制性能。而且在设计控制律时没有考虑实际控制问题中的输入-输出约束，一定程度上限制这些方法在实际系统中的应用。因此，在后续的研究中需要在控制器的设计中考虑实际系统的各种约束，以提高控制方法的实用性。

(3) 在基于强化 Q-学习的自适应控制的研究中，仅针对线性二次调节器中探索与利用平衡问题进行研究，未来可以扩展到针对非线性系统的强化学习控制方法中。

(4) 本文研究的自适应对偶控制是通过优化性能指标函数得到最优控制律，该方法注重改善系统在过渡时刻的瞬态控制性能，这不同于基于李雅普诺夫函数设计的控制器，虽然李雅普诺夫方法下的控制律不是最优的，但控制律、未知参数的自适应律能够确保系统控制的稳态性能。自适应对偶控制认为系统最终能够学习到真实参数，如果系统是线性的且能控能观，那么系统控制是稳定的，但对于非线性系统控制的稳定性问题，至今还是一个尚未解决的挑战性问题，因此今后可以对其稳定性做进一步研究。

## 致谢

写到这里就意味我的读博之旅已经接近尾声了，我的校园生活也即将结束，在西理工本硕博十二年的求学经历，此刻如电影一般在我脑海中回放，心中多少有点起伏。曾经有人问过我，你以后会怀念这个地方吗，当时只是戏谑的回答了一句，这有什么可怀念的。然而在此刻我觉得，那些在西理工肆意挥霍青春的日子，以后大概率还是会出现在我的梦境里。在离开西理工之前，我要向陪伴自己成长的老师、亲人、同学和朋友认真的道谢。

感谢我的博士导师钱富才教授，从 2017 年开始读博以来，我在科研中遇到的关键问题，都离不开钱老师的指导与帮助。钱老师学识渊博，对科学问题有自己独到的见解，每当我遇到科研上的问题时，钱老师都会从数学的角度帮我分析和理解问题，并且提供多种解决思路；钱老师治学严谨，处事细心，在科研过程中经常提醒我不要放过任何一个细节才能把科研工作做好；钱老师诲人不倦，一直以来悉心指导学生学业，对我的科研进程以及大小论文的撰写都提供了方方面面的帮助和指导；钱老师谦和平易，不给学生施加过多的压力，让我整个读博期间的科研生活比较宽松和自由。钱老师以身作则的治学、处事、为人的风格，都潜移默化的影响到了我，也将让我终生受益，我向钱老师致以崇高的敬意和诚挚的感谢！

感谢张晓晖教授，张老师是我的硕士导师，当时做完硕士课题，张老师说我有做科研的潜质，也因此一句话，点燃了心中的那团火，从此我便选择了读博，走上做科研的道路。我的读博道路一直都不算顺利，期间遇到各种困难，曾经想要放弃，张老师了解这些情况后会适时的鼓励我，让我重拾对科研的信心，让我能够继续走下去。在这里，我要向张老师表达衷心的谢意！

感谢我的先生张士良博士，我们于 2016 年夏天相遇，那时我也正处于迷茫期，不知自己该何去何从，当我决定要读博时，张博也是第一个站出来支持我的人。转眼间我们一起携手度过了七年的风雨与美好，回顾我整个博士期间的科研生活，张博一直以来都是我最坚强的后盾。当我因为论文一次次被拒而愤怒时，因为一直没有成果而焦虑失眠时，因为病痛而哭泣时，张博都会陪在我的身边。论文被拒后会跟我讨论问题所在，给我提供改进方案，一次次帮我逐字逐句修改语言表达，半夜焦虑睡不着的时候会陪我聊聊天，引导我放下思想压力，生病的时候耐心的照顾我的情绪和生活。在面对生活所带来的压力和困难时，张博常对我说，你要是心里想着算了吧，我不弄了，那么你就真的输给了生活，这肯定不是你想要的，所以心里的那股劲一定不要丢，我们就是能做成。感谢张博一直以来对我的包容、理解、帮助和鼓励。

感谢自动化与信息工程学院信控系的授课老师，刘丁教授、刘军教授、杨延西教授、刘涵教授、焦尚斌教授、胡绍林教授、张友民教授、朱虹教授、弋英民教授、谢国教授、辛菁教授，刘青教授，你们的授课让我对控制学科有了系统的认识。

感谢武莉师兄、刘磊师兄、尚婷师姐、晏琳师妹、沙忻煜师弟，我们同处一个教研室很多年，是相互陪伴时间最多的人，你们在科研、工作和生活中都给了我非常多的帮助，同时你们的优秀科研成果也激励了我。还要感谢武晗师妹、刘思宇师妹、刘浩林师弟、李薇师妹、刘康师弟、杨松楠师弟、王金宝师弟、郭燕师妹，我们一起讨论问题、带本科毕业设计、听讲座、出游、聚餐、看电影，这些美好的时光都会留在我的心里。感谢在西理工的同学，于雅洁、徐静、任俊超、杨洁、李晨晔，刘安东，感谢你们给予的帮助与陪伴。

感谢豹斐电科给我帮助和指导的人们，他们是王未韬师兄，王宝平师兄，王瑞斌师兄，李昂师兄。感谢瑞日电科帮助和指导过我的人们，他们是刘康总工，王浩师兄，王勇吉师兄，赵婉师姐。

感谢一起攀岩的小伙伴们，他们是媛媛、蟹老师、冷静姐、倩、剑、珏、豪哥、昆、厨子、黄司、兴林、朱司。我们曾经一起在岩馆磕线，一起去秦岭征服自然岩壁。攀岩更多的时候为我逃避科研提供了避难所，我在这里可以忘掉一切，只专注于去抓下一个岩点，享受完成一条线的喜悦，很好的缓解了科研压力以及强壮了身体。在攀岩中，我也认识到这么一个事，如果我怀疑够不到下一个岩点，那么我即使试一百次我都不可能够到那个点，但是如果我坚信自己可以做到，那么我才有可能真的把难点通过，也许这就是意念的力量。爬上峭壁俯瞰脚下的风景是无比美丽的，而攀爬峭壁的过程中用勇气和毅力克服恐惧和疼痛，完成一个个动作爬上去的过程更值得回味的。

感谢我的父亲和母亲，感谢你们的默默奉献和关爱。

感谢所有曾经给予我支持、帮助和关心的人们。

最后感谢国家自然科学基金的资助和各位审阅论文稿的评审老师们！

## 参考文献

- [1] 韩崇昭. 随机系统概论:分析、估计与控制[M]. 清华大学出版社, 2014.
- [2] 郭尚来. 随机控制[M]. 清华大学出版社, 1999.
- [3] Chen W H, Yang J, Guo L, et al. Disturbance-observer-based control and related methods-An overview[J]. IEEE Transactions on industrial electronics, 2015, 63(2): 1083-1095.
- [4] 杨辉, 刘盼, 李中奇. 基于 Elman 模型的高速列车速度跟踪控制[J]. 控制理论与应用, 2017, 34(1): 125-130.
- [5] Song Q, Song Y, Tang T, et al. Computationally inexpensive tracking control of high-speed trains with traction/braking saturation[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(4): 1116-1125.
- [6] Fermentation and biochemical engineering handbook[M]. William Andrew, 2014.
- [7] Zhang S, Cao H, Zhang Y, et al. Energy-efficient dynamic matrix control for biochemical continuous sterilization[C]//2016 Chinese Control and Decision Conference (CCDC). IEEE, 2016: 612-617.
- [8] 柯晓曼, 吴云华, 郑墨泓, 等. 基于改进迭代学习的参数不确定卫星姿态控制[J]. 系统工程与电子技术, 2021, 043(002): 508-518.
- [9] 管宇, 张迎春, 沈毅, 等. 基于迭代学习观测器的卫星姿态控制系统的鲁棒容错控制[J]. 宇航学报, 2012, 33(8): 1080-1086.
- [10] Guo L, Cao S. Anti-disturbance control theory for systems with multiple disturbances: A survey[J]. ISA transactions, 2014, 53(4): 846-849.
- [11] Murad Abu-Khalaf, Jie Huang, and Frank L Lewis. Nonlinear  $H_2 / H_\infty$  Constrained Feedback Control: A Practical Design Approach Using Neural Networks. Springer, 2006.
- [12] 庞中华. 系统辨识与自适应控制 MATLAB 仿真[M]. 北京航空航天大学出版社, 2009
- [13] Filatov N M, Unbehauen H. Adaptive dual control: Theory and applications[M]. Springer Science & Business Media, 2004.
- [14] Duan Li, Fucai Qian, and Peilin Fu. Optimal nominal dual control for discrete-time linear-quadratic gaussian problems with unknown parameters[J]. Automatica, 2008, 44(1): 119-127.
- [15] Feldbaum A A . Dual control theory I-IV[J]. Automation & Remote Control, 1960-61, 21: 874-880; 21: 1033-1039; 22: 1-12; 22: 109-121.
- [16] Reinforcement learning and approximate dynamic programming for feedback control[M]. John Wiley & Sons, 2013.

- [17] Sutton R S, Barto A G, Williams R J. Reinforcement learning is direct adaptive optimal control[J]. IEEE control systems magazine, 1992, 12(2): 19-22.
- [18] Kalman, R. E. Design of a self-optimizing control system. Transactions of the American Society of Mechanical Engineers, 1958, 80(2): 468-477.
- [19] Feldbaum A A . Optimal Control System[M]. New York: Academic Press, 1965.
- [20] Bohlin T. Optimal dual control of a simple process with unknown gain[J]. Technical Report, 1969.
- [21] Astrom K J , Wittenmark B . Problems of identification and control[J]. Journal of Mathematical Analysis & Applications, 1971, 34(1): 90-113.
- [22] Sternby J . A simple dual control problem with an analytical solution[J]. IEEE Transactions on Automatic Control, 1976, 21(6): 840-844.
- [23] Wittenmark B. An active suboptimal dual controller for systems with stochastic parameters[J]. Automatic Control Theory and Application, 1975, 3:13-19.
- [24] Filatov N M , Unbehauen H . Survey of adaptive dual control methods[J]. IEE Proceedings. Part D, 2000, 147(1): 118-128.
- [25] Mesbah A. Stochastic model predictive control with active uncertainty learning: A survey on dual control[J]. Annual Reviews in Control, 2018, 45: 107-117.
- [26] Wieslander J , Wittenmark B . An approach to adaptive control using real time identification[J]. Automatica, 1971, 7(2): 211-217.
- [27] Lozano R . Adaptive pole placement without excitation probing signals[J]. IEEE Transactions on Automatic Control, 2002, 39(1): 47-58.
- [28] Bar-Shalom Y. Stochastic dynamic programming: Caution and probing[J]. IEEE Transactions on Automatic Control, 1981, 26(5): 1184-1195.
- [29] Alster J , Pierre R. Bélanger. A technique for dual adaptive control.[J]. Automatica, 1974, 10(6): 627-634.
- [30] Mookerjee P , Bar-Shalom Y . An adaptive dual controller for a MIMO-ARMA system[J]. IEEE Transactions on Automatic Control, 1986, 34(7): 795-800.
- [31] Sternby J . A Regulator for Time-Varying Stochastic Systems[J]. IFAC Proceedings Volumes, 1978, 11(1): 2025-2032.
- [32] Birmiwal K , Bar-Shalom Y . Dual Control for Identification and Guidance[J]. Ifac Proceedings Volumes, 1984, 17(2): 841-846.
- [33] Wieslander J , Wittenmark B . An approach to adaptive control using real time identification[J]. Automatica, 1971, 7(2): 211-217.
- [34] Chan S S, Zarrop M B. A suboptimal dual controller for stochastic systems with unknown parameters[J]. International Journal of Control, 1985, 41(2): 507-524.

- [35] Casiello F, Loparo K A. Optimal control of unknown parameter systems[J]. IEEE transactions on automatic control, 1989, 34(10): 1092-1094.
- [36] Bernhardsson B. Dual control of a first-order system with two possible gains[J]. International Journal of Adaptive Control and Signal Processing, 1989, 3(1): 15-22.
- [37] Venkatasubramanian J, Köhler J, Berberich J, et al. Robust dual control based on gain scheduling[C]//2020 59th IEEE Conference on Decision and Control (CDC). IEEE, 2020: 2270-2277.
- [38] Iannelli A, Khosravi M, Smith R S. Structured exploration in the finite horizon linear quadratic dual control problem[J]. IFAC-PapersOnLine, 2020, 53(2): 959-964.
- [39] Kadirkamanathan V. A stochastic method for neural-adaptive control of multi-modal nonlinear systems[C]//IEE conference publication. 1998: 49-53.
- [40] Fabri S , Kadirkamanathan V . Dual adaptive control of nonlinear stochastic systems using neural networks[J]. Automatica, 1998, 34(2): 245-253.
- [41] Simandl, Miroslav, Král, Ladislav, Hering P . Neural network based bicriterial dual control of nonlinear systems[J]. IFAC Proceedings Volumes, 2005, 38(1): 58-63.
- [42] L. Kral, P. Hering, M. Simandl, Functional adaptive control for nonlinear stochastic systems in presence of outliers[J]. IFAC Proceedings Volumes, 2009, 42 (10): 1505-1510.
- [43] Fabri S G, Wittenmark B, Bugeja M K. Dual adaptive extremum control of Hammerstein systems[J]. International Journal of Control, 2015, 88(6): 1271-1286.
- [44] Bugeja M K, Fabri S G. Dual adaptive control for trajectory tracking of mobile robots[C]//Proceedings 2007 IEEE International Conference on Robotics and Automation. IEEE, 2007: 2215-2220.
- [45] Bugeja M K, Fabri S G, Camilleri L. Dual adaptive dynamic control of mobile robots using neural networks[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2008, 39(1): 129-141.
- [46] Bugeja M K, Fabri S G. Dual-adaptive computer control of a mobile robot based on the unscented transform[C]//2009 Third International Conference on Advanced Engineering Computing and Applications in Sciences. IEEE, 2009: 136-141.
- [47] Král L, Šimandl M. Predictive dual control for nonlinear stochastic systems modelled by neural networks[C]//2011 19th Mediterranean Conference on Control & Automation (MED). IEEE, 2011: 1277-1282.
- [48] Fabri S G, Bugeja M K. Unscented transform-based dual adaptive control of nonlinear MIMO systems[C]//2013 European Control Conference (ECC). IEEE, 2013: 392-397.
- [49] Mesbah A. Stochastic model predictive control with active uncertainty learning: A survey on dual control[J]. Annual Reviews in Control, 2018, 45: 107-117.

- [50] Kim T H, Sugie T. Adaptive receding horizon predictive control for constrained discrete-time linear systems with parameter uncertainties[J]. International Journal of Control, 2008, 81(1): 62-73.
- [51] Marafioti G , Stoican F , Bitmead R R , et al. Persistently exciting model predictive control[J]. International Journal of Adaptive Control & Signal Processing, 2012, 45(17): 448-453.
- [52] Marafioti G, Stoican F, Bitmead R R, et al. Persistently exciting model predictive control for siso systems[J]. IFAC Proceedings Volumes, 2012, 45(17): 448-453.
- [53] Žáčeková E, Prívara S, Pčolka M. Persistent excitation condition within the dual control framework[J]. Journal of Process Control, 2013, 23(9): 1270-1280.
- [54] 王超, 张胜修, 秦伟伟, 等. 具有自适应噪声边界的 Tube 可达集鲁棒预测控制[J]. 控制理论与应用, 2014, 31(1): 11-18.
- [55] Houska B, Telen D, Logist F, et al. Self-reflective model predictive control[J]. SIAM Journal on Control and Optimization, 2017, 55(5): 2959-2980.
- [56] Feng X, Houska B. Real-time algorithm for self-reflective model predictive control[J]. Journal of Process Control, 2018, 65: 68-77.
- [57] Heirung T A N, Foss B, Ydstie B E. MPC-based dual control with online experiment design[J]. Journal of Process Control, 2015, 32: 64-76.
- [58] Heirung T A N, Ydstie B E, Foss B. Dual adaptive model predictive control[J]. Automatica, 2017, 80: 340-348.
- [59] Kumar K, Patwardhan S C, Noronha S. Control of Systems Exhibiting Input Multiplicities using Dual Nonlinear MPC[J]. IFAC-PapersOnLine, 2018, 51(20): 110-115.
- [60] 曹文祺, 李少远. 具有可参数化不确定性系统的对偶自适应模型预测控制[J]. 控制理论与应用, 2019, 36(8): 1197-1206.
- [61] Parsi A, Iannelli A, Smith R S. An explicit dual control approach for constrained reference tracking of uncertain linear systems[J]. IEEE Transactions on Automatic Control, 2022.
- [62] Lin M, Ning Z, Li B, et al. Adaptive dual model predictive control for linear systems with parametric uncertainties[J]. IET Control Theory & Applications, 2022.
- [63] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. nature, 2015, 518(7540): 529-533.
- [64] Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. nature, 2016, 529(7587): 484-489.
- [65] Fazel M, Ge R, Kakade S M, et al. Global convergence of policy gradient methods for the linear quadratic regulator[J]. arXiv preprint arXiv:1801.05039, 2018.
- [66] Cohen A, Hassidim A, Koren T, et al. Online linear quadratic control[J]. arXiv preprint

- arXiv:1806.07104, 2018.
- [67] Dean S, Mania H, Matni N, et al. On the sample complexity of the linear quadratic regulator[J]. Foundations of Computational Mathematics, 2019: 1-47.
- [68] Klenske E D, Hennig P. Dual control for approximate Bayesian reinforcement learning[J]. The Journal of Machine Learning Research, 2016, 17(1): 4354-4383.
- [69] Larsson C A, Ebadat A, Rojas C R, et al. An application-oriented approach to dual control with excitation for closed-loop identification[J]. European Journal of Control, 2016, 29: 1-16.
- [70] Umenberger J, Ferizbegovic M, Schön T B, et al. Robust exploration in linear quadratic reinforcement learning[C]//Advances in Neural Information Processing Systems. 2019: 15336-15346.
- [71] Chen W H, Rhodes C, Liu C. Dual Control for Exploitation and Exploration (DCEE) in autonomous search[J]. Automatica, 2021, 133: 1-7.
- [72] Li Z, Chen W H, Yang J. Concurrent Learning Based Dual Control for Exploration and Exploitation in Autonomous Search[J]. arXiv preprint arXiv:2108.08062, 2021.
- [73] Bradtke S. Reinforcement learning applied to linear quadratic regulation[C]//Advances in Neural Information Processing Systems. 1992, 5: 294-302.
- [74] Yang Z, Chen Y, Hong M, et al. On the global convergence of actor-critic: A case for linear quadratic regulator with ergodic cost[J]. arXiv preprint arXiv:1907.06246, 2019.
- [75] Zhou M, Lu J. Single Time-scale Actor-critic Method to Solve the Linear Quadratic Regulator with Convergence Guarantees[J]. arXiv preprint arXiv:2202.00048, 2022.
- [76] Rizvi S A A, Lin Z. Reinforcement learning-based linear quadratic regulation of continuous-time systems using dynamic output feedback[J]. IEEE Transactions on Cybernetics, 2019, 50(11): 4670-4679.
- [77] Parisini T, Zoppoli R. Neural approximations for infinite-horizon optimal control of nonlinear stochastic systems[J]. IEEE transactions on neural networks, 1998, 9(6): 1388-1408.
- [78] Vamvoudakis K G, Lewis F L. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem[J]. Automatica, 2010, 46(5): 878-888.
- [79] Lee D J, Bang H. Model-free LQ control for unmanned helicopters using reinforcement learning[C]//2011 11th International Conference on Control, Automation and Systems. IEEE, 2011: 117-120.
- [80] Modares H, Ranatunga I, Lewis F L, et al. Optimized assistive human-robot interaction using reinforcement learning[J]. IEEE transactions on cybernetics, 2015, 46(3): 655-667.
- [81] Guo L, Rizvi S A A, Lin Z. Optimal control of a two-wheeled self-balancing robot by

- reinforcement learning[J]. International Journal of Robust and Nonlinear Control, 2021, 31(6): 1885-1904.
- [82] Yoo J, Jang D, Kim H J, et al. Hybrid reinforcement learning control for a micro quadrotor flight[J]. IEEE Control Systems Letters, 2020, 5(2): 505-510.
- [83] Kiumarsi-Khomartash B, Lewis F L, Naghibi-Sistani M B, et al. Optimal tracking control for linear discrete-time systems using reinforcement learning[C]//52nd IEEE Conference on Decision and Control. IEEE, 2013: 3845-3850.
- [84] Wen G, Chen C L P, Ge S S, et al. Optimized adaptive nonlinear tracking control using actor-critic reinforcement learning strategy[J]. IEEE transactions on industrial informatics, 2019, 15(9): 4969-4977.
- [85] Liu Y J, Tang L, Tong S, et al. Reinforcement learning design-based adaptive tracking control with less learning parameters for nonlinear discrete-time MIMO systems[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 26(1): 165-176.
- [86] Modares H, Lewis F L. Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning[J]. Automatica, 2014, 50(7): 1780-1792.
- [87] Kiumarsi B, Lewis F L, Modares H, et al. Reinforcement Q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics[J]. Automatica, 2014, 50(4): 1167-1175.
- [88] Guo X, Yan W, Cui R. Event-triggered reinforcement learning-based adaptive tracking control for completely unknown continuous-time nonlinear systems[J]. IEEE Transactions on Cybernetics, 2019, 50(7): 3231-3242.
- [89] Hu Y, Wang W, Liu H, et al. Reinforcement learning tracking control for robotic manipulator with kernel-based dynamic model[J]. IEEE transactions on neural networks and learning systems, 2019, 31(9): 3570-3578.
- [90] Ding L, Li S, Gao H, et al. Adaptive partial reinforcement learning neural network-based tracking control for wheeled mobile robotic systems[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018, 50(7): 2512-2523.
- [91] Yu R, Shi Z, Huang C, et al. Deep reinforcement learning based optimal trajectory tracking control of autonomous underwater vehicle[C]//2017 36th Chinese control conference (CCC). IEEE, 2017: 4958-4965.
- [92] Wei C, Zhang Z, Qiao W, et al. Reinforcement-learning-based intelligent maximum power point tracking control for wind energy conversion systems[J]. IEEE Transactions on Industrial Electronics, 2015, 62(10): 6360-6370.
- [93] Wu H N, Luo B. Simultaneous policy update algorithms for learning the solution of linear

- continuous-time  $H_\infty$  state feedback control[J]. Information Sciences, 2013, 222: 472-485.
- [94] Vrabie D, Lewis F. Adaptive dynamic programming for online solution of a zero-sum differential game[J]. Journal of Control Theory and Applications, 2011, 9(3): 353-360.
- [95] Luo B, Wu H N, Huang T. Off-policy reinforcement learning for  $H_\infty$  control design[J]. IEEE transactions on cybernetics, 2014, 45(1): 65-76.
- [96] Modares H, Lewis F L, Jiang Z P.  $H_\infty$  tracking control of completely unknown continuous-time systems via off-policy reinforcement learning[J]. IEEE transactions on neural networks and learning systems, 2015, 26(10): 2550-2562.
- [97] Watkins C J C H, Dayan P. Q-learning[J]. Machine learning, 1992, 8(3): 279-292.
- [98] Milito R , Padilla C , Padilla R , et al. An innovations approach to dual control[J]. IEEE Transactions on Automatic Control, 1982, 27(1):132-137.
- [99] Bertsekas D P. Dynamic programming and optimal control 3rd edition, volume ii[J]. Belmont, MA: Athena Scientific, 2011.
- [100] Bertsekas D. Reinforcement learning and optimal control[M]. Athena Scientific, 2019.
- [101] Landelius T, Knutsson H. Greedy adaptive critics for LQR problems: Convergence proofs (Tech. Rep. No. LiTH-ISY-R-1896)[J]. Linkoping, Sweden: Computer Vision Laboratory, 1996.
- [102] Davari M, Gao W, Jiang Z P, et al. An optimal primary frequency control based on adaptive dynamic programming for islanded modernized microgrids[J]. IEEE Transactions on Automation Science and Engineering, 2020, 18(3): 1109-1121.
- [103] Shi W, Song S, Wu C, et al. Multi pseudo Q-learning-based deterministic policy gradient for tracking control of autonomous underwater vehicles[J]. IEEE transactions on neural networks and learning systems, 2018, 30(12): 3534-3546.
- [104] Liu C, Murphrey Y L. Optimal power management based on Q-learning and neuro-dynamic programming for plug-in hybrid electric vehicles[J]. IEEE transactions on neural networks and learning systems, 2019, 31(6): 1942-1954.
- [105] Arcari E, Hewing L, Zeilinger M N. An approximate dynamic programming approach for dual stochastic model predictive control[J]. IFAC-PapersOnLine, 2020, 53(2): 8105-8111.
- [106] Filatov N M, Unbehauen H, Keuchel U. Dual pole-placement controller with direct adaptation[J]. Automatica, 1997, 33(1): 113-117.
- [107] Alessandri A, Awawdeh M. Moving-horizon estimation with guaranteed robustness for discrete-time linear systems and measurements subject to outliers[J]. Automatica, 2016, 67: 85-93.
- [108] Alessandri A, Awawdeh M. Moving-horizon estimation for discrete-time linear systems with measurements subject to outliers[C]//53rd IEEE Conference on Decision and Control.

- IEEE, 2014: 2591-2596.
- [109] Gustafsson F, Gustafsson F. Adaptive filtering and change detection[M]. New York: Wiley, 2000.
- [110] Akkaya A D, Tiku M L. Robust estimation in multiple linear regression model with non-Gaussian noise[J]. Automatica, 2008, 44(2): 407-417.
- [111] Ma X, Qian F, Zhang S. Dual control for stochastic systems with multiple uncertainties[C]//2020 39th Chinese Control Conference (CCC). IEEE, 2020: 1001-1006.
- [112] Vic Barnett, Toby Lewis. Outliers in statistical data[J]. Contemporary Sociology, 1980, 9(4): 560-561
- [113] Johnson T, Kwok I, Ng R T. Fast Computation of 2-Dimensional Depth Contours[C]//KDD. 1998: 224-228.
- [114] Knorr E M, Ng R T, Tucakov V. Distance-based outliers: algorithms and applications[J]. The VLDB Journal, 2000, 8(3): 237-253.
- [115] Breunig M M, Kriegel H P, Ng R T, et al. LOF: identifying density-based local outliers[C]//Proceedings of the 2000 ACM SIGMOD international conference on Management of data. 2000: 93-104.
- [116] Ali B, Azam N, Shah A, et al. A spatial filtering inspired three-way clustering approach with application to outlier detection[J]. International Journal of Approximate Reasoning, 2021, 130: 1-21.
- [117] Yu K, Chen H. Markov boundary-based outlier mining[J]. IEEE transactions on neural networks and learning systems, 2018, 30(4): 1259-1264.
- [118] Lin F, Cohen W W. Power iteration clustering[C]//ICML. 2010: 1-8.
- [119] Ismkhan H. Ik-means $\text{--}+$ : An iterative clustering algorithm based on an enhanced version of the k-means[J]. Pattern Recognition, 2018, 79: 402-413.
- [120] Prasad R K, Sarmah R, Chakraborty S. Incremental k-means method[C]//International Conference on Pattern Recognition and Machine Intelligence. Springer, Cham, 2019: 38-46.
- [121] Král L, Hering P, Šimandl M. Functional adaptive control for nonlinear stochastic systems in presence of outliers[J]. IFAC Proceedings Volumes, 2009, 42(10): 1505-1510.
- [122] Král L, Šimandl M. Neural network based bicriterial dual control with multiple linearization[J]. IFAC Proceedings Volumes, 2010, 43(10): 271-276.
- [123] Yuen K V, Mu H Q. A novel probabilistic method for robust parametric identification and outlier detection[J]. Probabilistic Engineering Mechanics, 2012, 30: 48-59.
- [124] Yoon J W. A simple sequential outlier detection with several residuals[C]//2015 23rd European Signal Processing Conference (EUSIPCO). IEEE, 2015: 2351-2355.

- [125] Ma L, Wang Z, Hu J, et al. Probability-guaranteed envelope-constrained filtering for nonlinear systems subject to measurement outliers[J]. IEEE Transactions on Automatic Control, 2020, 66(7): 3274-3281.
- [126] Wang L, Li Z, Yu F. Jackknife method for the location of gross errors in weighted total least squares[J]. Communications in Statistics-Simulation and Computation, 2019: 1-21.
- [127] Wang J, Zhao J, Liu Z, et al. Location and estimation of multiple outliers in weighted total least squares[J]. Measurement, 2021, 181: 1-17.
- [128] Karasu S, Altan A. Recognition model for solar radiation time series based on random forest with feature selection approach[C]//2019 11th international conference on electrical and electronics engineering (ELECO). IEEE, 2019: 8-11.
- [129] Xiao L, Wang Z, Wu Y. Composite Quantile Regression Estimation for Left Censored Response Longitudinal Data[J]. Acta Mathematicae Applicatae Sinica, English Series, 2018, 34(4): 730-741.
- [130] Altunbaş Y, Thornton J. The impact of financial development on income inequality: A quantile regression approach[J]. Economics Letters, 2019, 175: 51-56.
- [131] Jorion P. Value-at-Risk: The New Benchmark for Managing[J]. Financial Risk, 2006.
- [132] DeCandia G, Hastorun D, Jampani M, et al. Dynamo: Amazon's highly available key-value store[J]. ACM SIGOPS operating systems review, 2007, 41(6): 205-220.
- [133] Benoit D F, Van den Poel D. Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services[J]. Expert Systems with Applications, 2009, 36(7): 10475-10484.
- [134] Nguyen B T, Albrecht J W, Vroman S B, et al. A quantile regression decomposition of urban–rural inequality in Vietnam[J]. Journal of Development Economics, 2007, 83(2): 466-490.
- [135] Zhang N, Sun Q, Wang J, et al. Distributed adaptive dual control via consensus algorithm in the energy internet[J]. IEEE Transactions on Industrial Informatics, 2020, 17(7): 4848-4860.
- [136] Grado L L, Johnson M D, Netoff T I. Bayesian adaptive dual control of deep brain stimulation in a computational model of Parkinson’s disease[J]. PLoS computational biology, 2018, 14(12): 1-23.
- [137] Nechval K N, Nechval N A, Vasermanis E K. Adaptive dual control in one biomedical problem[J]. Kybernetes, 2003.
- [138] Pericic D, Vucetic B. Adaptive dual control for spacecrafts rendezvous[C]//Paris International Astronautical Federation Congress. 1982.
- [139] Bancroft J, Graham C A, Janssen E, et al. The dual control model: Current status and future

- directions[J]. Journal of sex research, 2009, 46(2-3): 121-142.
- [140] Al-Tamimi A, Lewis F L, Abu-Khalaf M. Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control[J]. Automatica, 2007, 43(3): 473-481.
- [141] Al-Tamimi A, Abu-Khalaf M, Lewis F L. Adaptive Critic Designs for Discrete-Time Zero-Sum Games With Application to  $H_\infty$  Control[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2007, 37(1): 240-247.
- [142] Kiumarsi B, Lewis F L, Jiang Z P.  $H_\infty$  control of linear discrete-time systems: Off-policy reinforcement learning[J]. Automatica, 2017, 78: 144-152.
- [143] Valadbeigi A P, Sedigh A K, Lewis F L.  $H_\infty$  Static Output-Feedback Control Design for Discrete-Time Systems Using Reinforcement Learning[J]. IEEE transactions on neural networks and learning systems, 2019, 31(2): 396-406.
- [144] Zhao K, Lian H. The Expectation-Maximization approach for Bayesian quantile regression[J]. Computational Statistics & Data Analysis, 2016, 96: 1-11.
- [145] Lancaster T, Jae Jun S. Bayesian quantile regression methods[J]. Journal of Applied Econometrics, 2010, 25(2): 287-307.
- [146] Ordiano J Á G, Gröll L, Mikut R, et al. Probabilistic energy forecasting using the nearest neighbors quantile filter and quantile regression[J]. International journal of forecasting, 2020, 36(2): 310-323.
- [147] Wang H, Li C. Distributed quantile regression over sensor networks[J]. IEEE Transactions on Signal and Information Processing over Networks, 2017, 4(2): 338-348.
- [148] Zou H, Yuan M. Composite quantile regression and the oracle model selection theory[J]. The Annals of Statistics, 2008, 36(3): 1108-1126.
- [149] Zhao Z, Xiao Z. Efficient regressions via optimally combining quantile information[J]. Econometric theory, 2014, 30(6): 1272-1314.
- [150] Huang H, Chen Z. Bayesian composite quantile regression[J]. Journal of Statistical Computation and Simulation, 2015, 85(18): 3744-3754.
- [151] Chen F C, Khalil H K. Adaptive control of a class of nonlinear discrete-time systems using neural networks[J]. IEEE Transactions on Automatic Control, 1995, 40(5): 791-801.
- [152] Nørgaard M, Ravn O, Poulsen N K, et al. Neural networks for modelling and control of dynamic systems[M]. Springer-Verlag London Limited. London. England, 2000.
- [153] Piche S, Sayyar-Rodsari B, Johnson D, et al. Nonlinear model predictive control using neural networks[J]. IEEE Control Systems Magazine, 2000, 20(3): 53-62.
- [154] Saridis G N. Entropy formulation of optimal and adaptive control[J]. IEEE Transactions on Automatic Control, 1988, 33(8): 713-721.

- [155] Tsai Y A, Casiello F A, Loparo K A. Discrete-time entropy formulation of optimal and adaptive control problems[J]. IEEE transactions on automatic control, 1992, 37(7): 1083-1088.
- [156] Wang H. Minimum entropy control of non-Gaussian dynamic stochastic systems[J]. IEEE Transactions on Automatic Control, 2002, 47(2): 398-403.
- [157] Chen W H, Yang J, Guo L, et al. Disturbance-observer-based control and related methods-An overview[J]. IEEE Transactions on industrial electronics, 2015, 63(2): 1083-1095.
- [158] Sariyildiz E, Oboe R, Ohnishi K. Disturbance observer-based robust control and its applications: 35th anniversary overview[J]. IEEE Transactions on Industrial Electronics, 2019, 67(3): 2042-2053.
- [159] He W, Sun Y, Yan Z, et al. Disturbance observer-based neural network control of cooperative multiple manipulators with input saturation[J]. IEEE transactions on neural networks and learning systems, 2019, 31(5): 1735-1746.
- [160] Song Q, Song Y, Tang T, et al. Computationally inexpensive tracking control of high-speed trains with traction/braking saturation[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(4): 1116-1125.
- [161] Song Q, Song Y D. Data-based fault-tolerant control of high-speed trains with traction/braking notch nonlinearities and actuator failures[J]. IEEE Transactions on Neural Networks, 2011, 22(12): 2250-2261.
- [162] Wang X, Zhu L, Wang H, et al. Robust distributed cruise control of multiple high-speed trains based on disturbance observer[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 22(1): 267-279.
- [163] Guo L, Cao S. Anti-disturbance control theory for systems with multiple disturbances: A survey[J]. ISA transactions, 2014, 53(4): 846-849.
- [164] Zhang Y, Lim C C, Liu F. Robust mixed  $H_2/H_\infty$  model predictive control for Markov jump systems with partially uncertain transition probabilities[J]. Journal of the Franklin Institute, 2018, 355(8): 3423-3437.
- [165] Hooshmandi K, Bayat F, Jahed-Motlagh M R, et al. Polynomial LPV approach to robust  $H_\infty$  control of nonlinear sampled-data systems[J]. International Journal of Control, 2020, 93(9): 2145-2160.
- [166] Liu H, Li X, Liu X, et al. Adaptive Neural Network Prescribed Performance Bounded- $H_\infty$  Tracking Control for a Class of Stochastic Nonlinear Systems[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 31(6): 2140-2152.
- [167] Tutsoy O, Barkana D E, Balikci K. A novel exploration-exploitation-based adaptive law for intelligent model-free control approaches[J]. IEEE Transactions on Cybernetics, 2021.

- [168] Li Z, Chen W H, Yang J. A Dual Control Perspective for Exploration and Exploitation in Autonomous Search[C]//2022 European Control Conference (ECC). IEEE, 2022: 1876-1881.
- [169] Klenske E D, Hennig P. Dual control for approximate Bayesian reinforcement learning[J]. The Journal of Machine Learning Research, 2016, 17(1): 4354-4383.
- [170] Liu L, Xie G, Qian F, et al. Dual fault-tolerant control for a class of stochastic systems with partial loss-of-control effectiveness[J]. International Journal of Robust and Nonlinear Control, 2022, 32(2): 947-959.
- [171] Liu W. Optimal filtering for discrete-time linear systems with time-correlated multiplicative measurement noises[J]. IEEE Transactions on Automatic Control, 2015, 61(7): 1972-1978.
- [172] Dai Y, Yu S, Yan Y. An adaptive EKF-FMPC for the trajectory tracking of UVMS[J]. IEEE Journal of Oceanic Engineering, 2019, 45(3): 699-713.
- [173] Frost V S, Stiles J A, Shanmugan K S, et al. A model for radar images and its application to adaptive digital filtering of multiplicative noise[J]. IEEE Transactions on pattern analysis and machine intelligence, 1982, 4(2): 157-166.
- [174] Song H, Ding D, Dong H, et al. Distributed maximum correntropy filtering for stochastic nonlinear systems under deception attacks[J]. IEEE Transactions on Cybernetics, 2020, 52(5): 3733-3744.
- [175] Zhao Y, Zhang W, Xia J, et al. Robust Stochastic Stability and Control for Uncertain Singular Markovian Jump Systems with Multiplicative Noise[J]. Asian Journal of Control, 2017, 19(6): 1891-1904.
- [176] Mazouchi M, Modares H. Data-driven Robust LQR with Multiplicative Noise via System Level Synthesis[J]. arXiv preprint arXiv:2204.02883, 2022.
- [177] Wang M, Wang Z, Dong H, et al. A novel framework for backstepping-based control of discrete-time strict-feedback nonlinear systems with multiplicative noises[J]. IEEE Transactions on Automatic Control, 2020, 66(4): 1484-1496.
- [178] Platt J. A resource-allocating network for function interpolation[J]. Neural computation, 1991, 3(2): 213-225.

## 攻读博士学位期间完成的主要工作

完成的论文：

1. Ma Xuehui, Qian Fucai, Zhang Shiliang, Wu Li. Adaptive dual control with online outlier detection for uncertain systems[J]. ISA Transactions, 2022, 129: 157-168. (SCI : 000875900500015, 中科院一区, IF: 5.911, 对应本文第 4 章)
2. Ma Xuehui, Qian Fucai, Zhang Shiliang, Wu Li, Liu Lei. Adaptive quantile control for stochastic system[J]. ISA Transactions, 2022, 123: 110-121. (SCI: 000793551800007, 中科院一区, IF: 5.911, 对应本文第 5 章)
3. Ma Xuehui, Qian Fucai, Zhang Shiliang. Dual control for stochastic systems with multiple uncertainties[C]//2020 39th Chinese Control Conference (CCC). IEEE, 2020: 1001-1006. (EI: 20203909241472, 对应本文第 4 章)
4. Ma Xuehui, Zhang Shiliang, Qian Fucai, Wang Jinbao, Yan lin. Q-learning based linear quadratic regulator with balanced exploration and exploitation for unknown systems[C]//2022 China Automation Congress (CAC). (EI: 已录用, 对应本文第 3 章)
5. Ma Xuehui, Zhang Shiliang, Qian Fucai. Active learning for anti-disturbance dual control of unknown nonlinear systems. (同行评审中, 对应本文第 7 章)
6. Wang Jinbao, Zhang Xiaohui, Ma Xuehui, Gao Yu-er, Li Ning. An Improved Model-free Adaptive Control Algorithm With Differential Element For Nonlinear Systems[C]//2022 China Automation Congress (CAC). (EI: 已录用)
7. Wu Li, Qian Fucai, Wang Lingzhi, Ma Xuehui. An improved type-reduction algorithm for general type-2 fuzzy sets[J]. Information Sciences, 2022, 593: 99-120. (SCI : 000770686400007, 中科院一区, IF: 8.233)
8. Liu Lei, Xie Guo, Qian Fucai, Wang Min, Ma Xuehui. Reliable control based on dual control for ARMAX system with abrupt faults[J]. Journal of the Franklin Institute, 2021, 358(11): 5694-5706. (SCI: 000702010000004, 中科院二区, IF: 4.504)
9. Liu Lei, Xie Guo, Qian Fucai, Guo Xiaohong, Ma Xuehui. Dual fault-tolerant control for a class of stochastic systems with partial loss-of-control effectiveness[J]. International Journal of Robust and Nonlinear Control, 2022, 32(2): 947-959. (SCI: 000711301600001, 中科院二区, IF: 4.406)

参与的项目：

1. 国家自然科学基金：混合不确定性系统的估计与控制问题研究。
2. 国家自然科学基金：不确定系统的概率鲁棒与对偶控制研究。

