

## 具有未知参数的 LQG 对偶控制 算法研究

尚婷<sup>1</sup> 钱富才<sup>1,2</sup> 张晓艳<sup>1</sup> 谢国<sup>1</sup>

**摘要** 对于具有未知参数的 LQG (Linear quadratic Gaussian) 问题, 提出了一种次优对偶控制方法, 用 Kalman 滤波处理过程噪声和测量噪声, 用前一时刻的后验概率对 Cost-to-go 进行线性近似, 然后, 用动态规划获得了次优控制律. 最后, 用一个例子说明了本文设计的控制器的实施过程. 结果表明, 该控制律具有良好的对偶性质, 并能在学习和控制之间实现较好平衡.

**关键词** LQG (Linear quadratic Gaussian), 动态规划, 对偶控制, 最优控制

**引用格式** 尚婷, 钱富才, 张晓艳, 谢国. 具有未知参数的 LQG 对偶控制算法研究. 自动化学报, 2017, 43(8): 1478–1484

**DOI** 10.16383/j.aas.2017.c160401

### Research on Dual Control Algorithm for LQG with Unknown Parameters

SHANG Ting<sup>1</sup> QIAN Fu-Cai<sup>1,2</sup> ZHANG Xiao-Yan<sup>1</sup>  
XIE Guo<sup>1</sup>

**Abstract** For the LQG problem with unknown parameters, a novel suboptimal dual control approach is proposed in this paper. First, Kalman filter is used to deal with the noises of process and measurement and posterior probabilities at the previous moment are used to linearly approximate the cost-to-go at the present moment. Then dynamic programming is adopted to obtain a suboptimal control law. Finally, an example is presented to illustrate the implementation process of the developed controller. The result shows that this control law has good dual property and achieves a better balance between learning and control.

**Key words** LQG (Linear quadratic Gaussian), dynamic programming, dual control, optimal control

**Citation** Shang Ting, Qian Fu-Cai, Zhang Xiao-Yan, Xie Guo. Research on dual control algorithm for LQG with unknown parameters. *Acta Automatica Sinica*, 2017, 43(8): 1478–1484

LQG (Linear quadratic Gaussian) 控制问题, 由于滤波器与控制器之间具有分离性质, 受到了理论界和工程界的高度重视. 目前, LQG 理论已经成功用于航空、航天和许多工业控制中<sup>[1-4]</sup>. 然而, 当系统方程中存在未知参数时, 分离性质不再成立, LQG 控制方法失去优势. 早在上世纪 60 年代

初, 前苏联学者 Feldbaum 就注意到这类问题, 并在他的系列论文中指出, 仅有极少数几个简单问题能够求出解析解, 而对于绝大多数问题即便是想获得数值解也极为困难, 与此同时解决这类问题的对偶控制 (Dual control) 方法<sup>[2-3]</sup> 应运而生. 2000 年 IEEE Control Systems Society 将对偶控制列为上世纪对控制领域最具影响的 25 篇论文之一, 事实上这一难题至今尚未彻底解决.

对于这类问题进行如此艰难的研究, 到底有没有实际需求, 以下事例给出了肯定的回答. 用现有控制理论设计出的控制器一般只能保证系统以最优的方式运行, 但是, 近几十年来, 空难、海难、核爆炸等事故的发生使人民生命和国家财产遭受了重大损失, 不得不将很多系统的可靠性和安全性放在首位. 的确, 当系统内部元件出现故障时, 控制器如果能够以较快的速度作出反应, 也许一些严重后果就可以避免. 元器件的故障, 如执行器、传感器的卡死反映在控制模型上就是存在一些未知参数<sup>[5]</sup>; 高速列车在行驶过程中受到空气的阻力  $w$  与车速  $v$  的关系为:  $w = c_0 + c_1v + c_2v^2$ , 在跨区域、大范围的运行中对控制器来说,  $c_0$ 、 $c_1$  和  $c_2$  都是未知的<sup>[6]</sup>; 在大型建筑结构、桥梁、机床、轮船、飞行器、装甲车和急救车中都会装有由质量块-弹簧-阻尼器组成的减振系统. 系统中弹簧的刚性系数  $K$  和阻尼器的阻尼系数  $C$  可视为常数, 但往往是未知的, 只能预先知道它们在标称值上下 20% 或 50% 内变化<sup>[1,7]</sup>, 体现在控制问题中就是模型中存在未知参数; 对于一般的非线性系统, 内部零部件的参数很难或者无法测量, 即使能够精确知道, 但由于建模简化或者环境恶劣使得模型的等效参数与元部件的实际物理参数不存在一对一对应关系, 必须通过在线辨识来确定, 这些情况也可视为模型中有未知参数. 因此, 我们有充分的理由认为, 模型中存在未知参数的控制问题普遍存在.

对于存在未知参数且为常数的随机对偶控制问题, 已经进行了大量研究, 2004 年, Filatov 等总结了对偶控制自提出到当时的研究成果<sup>[4]</sup>, 目前主要的方法有对偶自适应控制<sup>[8-9]</sup>、新息对偶控制<sup>[10]</sup>、交换算法<sup>[11]</sup>、方差最小化<sup>[12-15]</sup>、标称对偶控制<sup>[16-18]</sup> 等. 所有这些方法的缺点是要求未知参数有明确的统计特性, 要么在为数不多的有限集内取某个固定值, 对系统不确定性因素的先验信息要求太高, 而且控制器设计时往往将一个多阶段最优控制问题强行用多个单阶段控制问题来近似, 导致控制器具有“近视”行为, 只能保证单步最优, 而不是全局最优. 本文拓宽了这些要求, 仅要求未知参数在某个区间, 也就是说未知参数最大不会超过的上限和最小不能小于的下限是已知的, 借助于对偶控制思想, 导出了一个次优控制律, 通过对一个例子的仿真结果可以看出, 该控制律有良好的学习和控制效果. 就学习而言, 经过有限时间我们可以学习出包含未知参数的区间, 这个区间长度可以事先任意给定, 区间长度越小, 表明对未知参数产生的不确定性消除得越彻底, 但以模型数目增大为代价. 就控制而言, 本文导出的控制律和最优控制基本重合, 两者之间形成的面积很小. 当未知参数属于有界集合时, 也可以用鲁棒理论设计出所谓的鲁棒控制器, 实践与理论均已表明, 这类控制器往往比较保守. 而本文设计出的控制器, 就未知参数为常数时, 可以以概率 1 学习到包含未知参数的最小区间.

本文的组织如下: 第 1 节详细描述了本文要解决的控制问题, 在第 2 节中我们将本文具有未知参数的 LQG 问题转化为对偶控制问题, 第 3 节利用线性近似的方法设计出了具有对偶特点的次优控制器, 第 4 节通过一个例子说明了控制器的性质, 结论放在第 5 节中.

收稿日期 2016-05-17 录用日期 2017-03-30  
Manuscript received May 17, 2016; accepted March 30, 2017  
国家自然科学基金 (61273127, U1534208), 陕西省科技攻关项目 (2016GY-108), 航天器在轨故障诊断与维修实验室开放课题 (SDML-OF2015004)  
Supported by National Natural Science Foundation of China (61273127, U1534208), Science and Technology Project of Shaanxi Province (2016GY-108), and the Key Laboratory for Fault Diagnosis and Maintenance of Spacecraft in Orbit (SDML-OF2015004)  
本文责任编辑 方海涛  
Recommended by Associate Editor FANG Hai-Tao  
1. 西安理工大学自动化与信息工程学院 西安 710048 2. 西安工业大学陕西省自主系统与智能控制国际联合研究中心 西安 710021  
1. School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048 2. The International Joint Research Center of Autonomous Systems and Intelligent Control, Xi'an Technological University, Xi'an 710021

## 1 问题描述

考虑如下具有未知参数的离散时间动态系统:

$$x(t+1) = \Phi(\alpha)x(t) + G(\alpha)u(t) + \xi(t), \quad t = 0, 1, \dots, N-1 \quad (1)$$

$$z(t) = H(\alpha)x(t) + \eta(t), \quad t = 1, 2, \dots, N \quad (2)$$

其中,  $x(t) \in \mathbf{R}^n$  是状态向量,  $u(t) \in \mathbf{R}^p$  是控制向量,  $z(t) \in \mathbf{R}^q$  是输出向量,  $\xi(t), \eta(t)$  和  $x(0)$  为相互独立的高斯白噪声, 其分布分别为:

$$\xi(t) \sim N(0, \Sigma_\xi)$$

$$\eta(t) \sim N(0, \Sigma_\eta)$$

$$x(0) \sim N(0, P_0)$$

$\Phi(\alpha), G(\alpha), H(\alpha)$  为适当维数的矩阵, 可以是时变的, 也可以是定常, 对后面控制器的设计没有影响, 为书写方便, 本文假设它们都是定常的. 另外, 假定未知参数  $\alpha \in \mathbf{R}$  以线性形式进入矩阵  $\Phi(\cdot), G(\cdot), H(\cdot)$ , 且  $\alpha$  在整个控制过程中是未知常数, 包含它的有限区间  $[\alpha_{\min}, \alpha_{\max}]$  是验前已知的.

动态系统的性能指标取为状态与控制的二次型形式:

$$J = E \left\{ \sum_{t=0}^{N-1} [x(t)^T A x(t) + u(t)^T B u(t)] + x(N)^T A x(N) \right\} \quad (3)$$

其中  $A, B$  为适当维数的半正定和正定对称阵.

设  $t$  为当前时刻, 则  $t$  时刻的实时信息为

$$I^t = \{u(0), u(1), \dots, u(t-1); z(1), z(2), \dots, z(t)\}$$

本文解决的控制问题为: 对于系统 (1) 和 (2), 求出一个允许的控制律使性能指标 (3) 中的  $J$  最小. 允许控制律是指其具有形式  $u(t) = \mu_t(I^t)$ , 其中  $\mu_t(\cdot)$  是待定的非线性函数.

对于上述控制问题, 当  $\alpha$  已知时, 本文研究的就是标准的 LQG 问题, 已经有成熟的解法; 当  $\alpha$  未知时系统中就有两种不确定性, 一种是环境噪声  $\xi(t)$  和量测噪声  $\eta(t)$ , 这种不确定性是客观存在, 无法减少, 只能用滤波技术对状态进行估计; 另外一种是由参数  $\alpha$  的未知性引起的, 由于  $\alpha \in [\alpha_{\min}, \alpha_{\max}]$ , 因此这种不确定性是一种有界不确定性.

目前, 对于随机不确定性和有界不确定性的研究已经相对成熟, 取得了丰富成果并用于实际. 然而, 长期以来处理这两种不确定性的理论被作为两个独立领域平行发展, 高度分化, 非此即彼. 一方面导致在 Bayesian 框架下的随机滤波由于非高斯噪声的存在而过于乐观, 甚至不能收敛, 另一方面由于高斯噪声的存在使得鲁棒滤波的噪声边界在选择时过于保守, 精度下降. 本文将两种不确定性视为一个共同体, 在同一个框架下统一处理.

最近二十多年也出现了一些随机系统鲁棒分析与控制器设计方法<sup>[19]</sup>, 其基本思想为首先在参数变化空间内选取  $M$  个设计点, 其次, 在每个设计点为被控对象设计出相应的控制律, 最后用随机鲁棒代价函数对  $M$  个控制律进行融合, 在飞行导弹控制中, 该方法鲁棒性极强, 即使参数有 20% 的波动也能保证满意的控制性能. 然而, 依据 Bernoulli 大数定律, 要保证融合出来的控制律具有一定的可信度,  $M \geq 1/(4\epsilon^2\beta)$ ,

其中  $\epsilon$  为控制精度,  $\beta$  为置信系数. 当  $\epsilon = \beta = 0.001$  时,  $M \geq 2.5 \times 10^8$ , 最小设计点的选取数目巨大.

从机理上说, 尽管模型中存在未知参数  $\alpha$ , 但系统的输出  $z(t)$  中始终包含了  $\alpha$  的信息, 因此, 通过不断地从系统获得测量信息, 可以减少这类不确定性. 还有, 当两种不确定性共存时, 控制器一方面要使性能指标达到最优, 控制信号不宜太大, 另一方面还要充分激励系统, 以获得状态和未知参数的更好估计, 控制信号要尽可能大, 显然, 两种作用互相冲突.

本文将依据对偶控制原理, 建立同时具有控制和学习 (估计) 特点的控制器的设计理论, 另外, 依据本文提出的控制器设计方法, 不确定参数设计点的个数仅为  $1/\epsilon$ , 在  $\epsilon = 0.001$  时,  $M = 100$ .

## 2 对偶控制问题

为了解决上节提出的控制问题, 本节将对偶控制问题来逼近真实问题.

**定理 1.** 对于任意给定的正数  $\epsilon$ , 存在正整数  $s$  及  $[\alpha_{\min}, \alpha_{\max}]$  上的分点

$$\alpha_{\min} = \alpha_0 < \alpha_1 < \dots < \alpha_s = \alpha_{\max}$$

使得  $\bigcup_{i=1}^s [\alpha_{i-1}, \alpha_i] = [\alpha_{\min}, \alpha_{\max}]$  且  $|\alpha_i - \alpha_{i-1}| < \epsilon$ .

**证明.** 对于任意给定的正数  $\epsilon$ , 取正整数  $s$  满足

$$s - 1 = \left\lceil \frac{\alpha_{\max} - \alpha_{\min}}{\epsilon} \right\rceil \quad (4)$$

其中,  $[x]$  为不超过  $x$  的最大整数, 则

$$\frac{\alpha_{\max} - \alpha_{\min}}{\epsilon} - \left\lfloor \frac{\alpha_{\max} - \alpha_{\min}}{\epsilon} \right\rfloor < 1$$

即,

$$\frac{\alpha_{\max} - \alpha_{\min}}{\epsilon} < s$$

上式等价于

$$\frac{\alpha_{\max} - \alpha_{\min}}{s} < \epsilon \quad (5)$$

将区间  $[\alpha_{\min}, \alpha_{\max}]$  进行  $s$  等分, 则每个子区间的长度为  $(\alpha_{\max} - \alpha_{\min})/s$ , 取

$$\alpha_0 = \alpha_{\min} \\ \alpha_i = \alpha_0 + i \frac{\alpha_{\max} - \alpha_{\min}}{s}, \quad i = 1, 2, \dots, s$$

则定理得证.  $\square$

该定理的意义是系统模型中的未知常数一定包含在一系列互不相交的子区间内, 且仅属于某个区间.  $\epsilon$  越小, 包含真值的区间越短, 但分点会增加, 从而覆盖先验区间  $[\alpha_{\min}, \alpha_{\max}]$  的子区间越多.

取每个子区间  $[\alpha_{i-1}, \alpha_i]$  的中点为  $\theta_i$ , 即

$$\theta_i = \frac{\alpha_{i-1} + \alpha_i}{2} \quad (6)$$

由于本文假定未知参数  $\alpha$  为常数, 因此,  $\alpha$  的真值  $\alpha^*$  一定在某个子区间内, 假定在第  $i$  个区间, 那么,  $|\alpha^* - \theta_i| < \epsilon$ . 对偶控制除了具有控制作用, 还能学习到未知参数, 一旦以概率 1 学到了  $\theta_i$ , 则可认为学到了包含真值  $\alpha^*$  的最小区间. 具有控制和学习特点的对偶控制是控制器设计的一种理念, 不同的问题, 设计方法不同, 下面我们先给出对偶控制的数

学描述. 本文解决的对偶控制问题 (Dual control problem, DCP) 如下:

$$\begin{aligned} \min J &= E\{x(N)^T A(N)x(N) + \\ &\sum_{t=0}^{N-1} [x(t)^T A(t)x(t) + u(t)^T B(t)u(t)]\} \\ \text{s.t. } x(t+1) &= \Phi(\alpha)x(t) + G(\alpha)u(t) + \xi(t), \\ &t = 0, 1, \dots, N-1 \\ z(t) &= H(\alpha)x(t) + \eta(t), t = 1, 2, \dots, N \end{aligned}$$

其中, 未知参数  $\alpha$  属于有限集合, 即  $\alpha \in \Omega = \{\theta_1, \theta_2, \dots, \theta_s\}$ , 真值是这  $s$  个值的其中之一, 不确定性参数  $\alpha$  可用验前概率描述,  $q_i(0) = 1/s, i = 1, 2, \dots, s$ . 未知参数以等概率取值于  $\Omega$  集合中的每个数, 这表明验前未知参数的不确定性最大.

为了后面书写方便, 本文做如下标记: 当  $\alpha = \theta_i$  时, 令  $\Phi_i = \Phi(\theta_i), G_i = G(\theta_i), H_i = H(\theta_i)$ .

注意, 如果在对偶控制问题 (DCP) 中, 参数是已知的, 没有不确定性, 仅有噪声不确定性, 那么, DCP 就是一个 LQG 问题, 有成熟解法; 如果未知参数有不确定性, 那么设计出的控制器应该具备双重作用, 一方面要使性能指标达到最优, 另一方面又要能从  $s$  个可能的参数中学出真实参数.

### 3 控制器设计

定义  $\theta_i$  对应的第  $i$  个模型在  $t$  时间关于信息集  $I^t$  的状态估计为  $\hat{x}_i(t|t)$ , 即:

$$\hat{x}_i(t+1|t+1) = E\{x(t+1)|\alpha = \theta_i, I^{t+1}\}$$

那么, 依据 Kalman 滤波, 我们有:

$$\hat{x}_i(t+1|t+1) = \Phi_i \hat{x}_i(t|t) + G_i u_i(t) + F_i(t+1) \nu_i(t+1) \tag{7}$$

$$F_i(t+1) = P_i(t+1|t) H_i^T \times [H_i P_i(t+1|t) + H_i^T \Sigma_\eta]^{-1} \tag{8}$$

$$P_i(t+1|t) = \Phi_i P_i(t|t) \Phi_i^T + \Sigma_\xi \tag{9}$$

$$P_i(t+1|t+1) = [I - F_i(t+1) H_i] \times P_i(t+1|t) \tag{10}$$

$$\nu_i(t) = z(t) - H_i \hat{x}_i(t|t) \tag{11}$$

其中  $z(t)$  为来自于系统的测量.

对偶控制的学习性质在于控制律中含有后验概率, 依据文献 [20], 第  $i$  个模型在  $t$  时刻的后验概率定义如下:

$$\tau_i(t) = P(\alpha = \theta_i | I^t) \tag{12}$$

当获得当前系统的测量输出  $z(t)$  后, 根据文献 [19], 用 Bayes 公式, 后验概率  $\tau_i(t)$  可以如下计算:

$$\tau_i(t) = \frac{M_i(t) \tau_i(t-1)}{\sum_{j=1}^s \tau_j(t-1) M_j(t)} \tag{13}$$

其中, 在式 (12) 中的初值  $\tau_i(0)$  验前给定, 且

$$\begin{aligned} M_i(t) &= |P_{z_i}(t|t-1, \theta_i)|^{-1/2} \times \\ &\exp \left[ -\frac{1}{2} \nu_i(t)^T P_{z_i}^{-1}(t|t-1, \theta_i) \nu_i(t) \right] \times \\ P_{z_i}(t|t-1, \theta_i) &= H_i P_i(t|t-1, \theta_i) H_i^T + \Sigma_\eta \end{aligned}$$

动态规划是求解 LQG 的有力工具, 除了能获得递推形式的控制增益外, 还能得到反馈形式的控制律, 这是工程上所期望的. 为能够应用动态规划, 用  $W^*(t)$  表示  $t$  时刻到终端时刻  $N$  二次性能指标的最优损失泛函, 即

$$W^*(t) = \min_{u(t), \dots, u(N-1)} E \left\{ \sum_{i=t}^N J(i) | I^t \right\} \tag{14}$$

其中,

$$\begin{aligned} J(i) &= x^T(i) A x(i) + u^T(i) B u(i) \\ J(N) &= x^T(N) A x(N) \end{aligned}$$

那么根据式 (14),

$$\begin{aligned} W^*(t) &= \min_{u(t)} E \{ J(t) + \\ &\min_{u(t+1), \dots, u(N-1)} E [ \sum_{i=t+1}^N J(i) | I^{t+1} ] | I^t \} = \\ &\min_{u(t)} E \{ [ J(t) + W^*(t+1) ] | I^t \} \end{aligned} \tag{15}$$

边界条件为  $W^*(N) = \hat{x}^T(N) A \hat{x}(N)$ , 上式的第二个等号用到了期望算子的关于信息集的光滑性, 可以参看文献 [14]. 一般情况下, 我们称式 (15) 中的  $W^*(t)$  为 Cost-to-go.

方程 (15) 是求解 DCP 的最优性原理, 正如 LQG 求解过程那样, 在每一个时刻由后向前, 就可以得到具有反馈形式的控制律. 然而, 整个对偶控制问题的全部复杂性都体现在方程 (15) 的求解上, 由于性能指标与  $s$  个模型有关, 相关程度用后验概率  $\tau_i(t)$  来度量, 显然, 后验概率由系统的实时测量所决定, 导致了 LQG 问题中滤波与增益的计算相互耦合, 分离性质不再成立, 动态规划中的 Cost-to-go 不再具有统一的递归形式. 文献 [16] 的研究表明, 每阶段的 Cost-to-go 在后向递推过程中, 后验概率以非线性形式严重阻碍动态规划的进行, 一个能使控制律既有学习又有控制功能的次优算法, 就是在每个阶段 Cost-to-go 关于后验概率线性化. 下面基于这一思想推导一个新的对偶控制律.

为书写方便, 对于参数  $\theta_i, i = 1, 2, \dots, s$ , 定义:

$$\begin{aligned} \bar{\Phi} &= \text{diag}\{\Phi_1, \Phi_2, \dots, \Phi_s\} \\ \bar{G} &= [G_1^T, G_2^T, \dots, G_s^T]^T \\ \bar{Q}(t, A) &= \text{diag}\{\tau_1(t)A, \tau_2(t)A, \dots, \tau_s(t)A\} \\ \bar{P}(t) &= \text{diag}\{P_1(t|t), P_2(t|t), \dots, P_s(t|t)\} \\ \hat{X}(t) &= [\hat{x}_1^T(t|t), \hat{x}_2^T(t|t), \dots, \hat{x}_s^T(t|t)]^T \end{aligned}$$

**定理 2.** 对于问题 DCP,  $t = N-1, N-2, \dots, 0$ , 其次优解和对应的 Cost-to-go 分别为:

$$u^*(t) = -L(t) \hat{X}(t|t) \tag{16}$$

$$W^*(t) = \hat{X}^T(t) \tilde{S} \hat{X}(t) + T(t) \tag{17}$$

其中,

$$L(t) = D^{-1}(t)\bar{G}^T\bar{Q}(t, \tilde{S})\bar{\Phi} \quad (18)$$

$$D(t) = \bar{G}^T\bar{Q}(t, \tilde{S})\bar{G} + B \quad (19)$$

$$S(t) = \bar{\Phi}^T\bar{Q}(t, \tilde{S})\bar{\Phi} + \bar{Q}(t, \tilde{S}) - L^T(t)D(t)L(t) \quad (20)$$

$$T(t) = \text{tr}(\bar{\Phi}^T\bar{Q}(t, \tilde{S})\bar{\Phi}\bar{P}(t)) + \text{tr}(\tilde{S}\Sigma_\xi) \quad (21)$$

$$\tilde{S} = \tau_1(t)R_1(t-1) + \tau_2(t)R_2(t-1) + \dots + \tau_s(t)R_s(t-1) \quad (22)$$

边界条件为  $\tilde{S}(N-1) = A$ .

在求解问题 DCP 的过程中, 用动态规划可以得到  $S(t)$ , 它是后验概率  $\tau_i(t), i = 1, 2, \dots, s$  的非线性函数, 导致动态规划不能继续进行. 本文将  $S(t)$  在前一时刻  $t-1$  的后验概率  $\tau_i(t-1), i = 1, 2, \dots, s$  处线性展开, 线性部分记为  $\tilde{S}$ , 展开后  $\tau_i(t-1), i = 1, 2, \dots, s$  的系数记为  $R_i(t-1)$ . 这样近似有 3 个优点: 1) 在  $t$  时刻, 前一时刻的后验概率  $\tau_i(t-1), i = 1, 2, \dots, s$  是已知的, 因此, 在该点线性化是可行的; 2) 线性化后, 用后验概率的线性部分  $\tilde{S}$  代替非线性函数  $S(t)$ , 动态规划能够顺利进行, 保证了次优解偏离最优解不会太远; 3) 由于线性化后  $\tilde{S}$  中依然保留了后验概率, 因此导出的控制增益具有学习性质, 从而保证了本文导出的次优控制律具有对偶性质.

**证明.** 从  $N-1$  阶段开始, 反向递推到初始阶段 0, 对时间  $t$  用数学归纳法, 具体推导如下:

假定  $t = N-1$ , 则根据最优性原理 (15),

$$W(N-1) = \text{E}\{[J(N-1) + W^*(N)]|I^{N-1}\} = \text{E}\{[J(N-1) + J(N)]|I^{N-1}\} \quad (23)$$

式 (23) 中的条件期望具有如下性质:

$$\begin{aligned} \text{E}\{(\cdot)|I^{N-1}\} &= \sum_{i=1}^s P(\alpha = \theta_i|I^{N-1})\text{E}\{(\cdot)|I^{N-1}, \alpha = \theta_i\} = \\ &= \sum_{i=1}^s \tau_i(N-1)\text{E}\{(\cdot)|I^{N-1}, \alpha = \theta_i\} \end{aligned}$$

将 Kalman 滤波方程 (7)~(11) 以及动态方程 (1) 和 (2) 代入上式, 经过简单地整理后, 则有以下结果:

$$\begin{aligned} W(N-1) &= \hat{X}^T(N-1|N-1)[\bar{Q}(N-1, A) + \\ &\quad \bar{\Phi}^T\bar{Q}(N-1, A)\bar{\Phi}]\hat{X}(N-1|N-1) + \\ &\quad \text{tr}(\bar{\Phi}^T\bar{Q}(N-1, A)\bar{\Phi}\bar{P}(N-1)) + \\ &\quad 2u^T(N-1)\bar{G}^T\bar{Q}(N-1, A) \times \\ &\quad \bar{\Phi}\hat{X}(N-1|N-1) + \\ &\quad u^T(N-1)[\bar{G}^T\bar{Q}(N-1, A)\bar{G} + B] \times \\ &\quad u(N-1) + \text{tr}(A\Sigma_\xi) \end{aligned}$$

上式是控制  $u(N-1)$  的二次形式, 直接关于控制求导, 并令其为零, 可以得到使上式最小的控制  $u^*(N-1)$ , 即

$$u^*(N-1) = -L(N-1)\hat{X}(N-1|N-1) \quad (24)$$

其中

$$L(N-1) = D^{-1}(N-1)\bar{G}^T\bar{Q}(N-1, A)\bar{\Phi}$$

$$D(N-1) = \bar{G}^T\bar{Q}(N-1, A)\bar{G} + B$$

把最优控制 (24) 代入  $W(N-1)$ , 对应的最优 Cost-to-go 为

$$W^*(N-1) = \hat{X}^T(N-1)S(N-1)\hat{X}(N-1) + T(N-1)$$

其中

$$S(N-1) = \bar{\Phi}^T\bar{Q}(N-1, A)\bar{\Phi} + \bar{Q}(N-1, A) - L^T(N-1)D(N-1)L(N-1)$$

$$T(N-1) = \text{tr}(\bar{\Phi}^T\bar{Q}(N-1, A)\bar{\Phi}\bar{P}(N-1)) + \text{tr}(A\Sigma_\xi)$$

显然,  $L(N-1)$  是后验概率  $\tau_i(N-1)$  的非线性函数, 导致  $S(N-1)$  也是  $\tau_i(N-1)$  的非线性函数. 由于在  $N-2$  时刻的后验概率  $\tau_i(N-2), i = 1, 2, \dots, s$  是已知的, 因此, 在  $\tau_i(N-2)$  处, 对  $S(N-1)$  进行线性近似, 即把  $S(N-1)$  在  $\tau_i(N-2)$  处一阶展开, 忽略高次项, 这样就可得到  $S(N-1)$  的如下近似表达式:

$$\begin{aligned} \tilde{S} &= \tau_1(N-1)R_1(N-2) + \\ &\quad \tau_2(N-1)R_2(N-2) + \dots + \\ &\quad \tau_s(N-1)R_s(N-2) \end{aligned}$$

其中,  $R_i(N-2)$  与  $\tau_i(N-1)$  无关, 仅与  $\tau_i(N-2)$  有关, 而  $\tau_i(N-2)$  在  $N-1$  时刻是已知. 这样  $\tilde{S}$  关于每个  $\tau_i(N-1)$  是线性的.

用上式的  $\tilde{S}$  代替  $W^*(N-1)$  中的  $S(N-1)$ , 则

$$W^*(N-1) = \hat{X}^T(N-1)\tilde{S}\hat{X}(N-1) + T(N-1)$$

比较在  $N-1$  时刻得到的控制与近似的 Cost-to-go, 这些结果与定理 2 的结论完全一致. 这表明在  $t = N-1$  时, 定理结论成立.

假设在  $t+1$  时刻, 定理 2 的结论成立, 则根据最优性原理 (15),

$$W^*(t) = \min_{u(t)} \text{E}\{J(t) + W^*(t+1)|I^t\} \quad (25)$$

在式 (25) 中, 用在  $t$  时刻线性化的  $\tilde{S}$  近似  $W^*(t+1)$  中的  $S(t+1)$ , 那么下式成立,

$$W(t) = \text{E}\{x^T(t)Ax(t) + u^T(t)Bu(t) + T(t+1) + \hat{X}^T(t+1)\tilde{S}\hat{X}(t+1)|I^t\}$$

根据式 (22),  $\tilde{S}$  为  $q_i(t+1)$  的线性函数, 这样, 根据期望算子的光滑性, 上式中的最后一项可以用与  $N-1$  阶段相同的方法求出解析表达式, 即

$$\begin{aligned} W(t) &= \hat{X}^T(t)[\bar{Q}(t, \tilde{S}) + \\ &\quad \bar{\Phi}^T\bar{Q}(t, \tilde{S})\bar{\Phi}]\hat{X}(t) + \text{tr}(\bar{\Phi}^T\bar{Q}(t, \tilde{S})\bar{\Phi}\bar{P}(t)) + \\ &\quad 2u^T(t)\bar{G}^T\bar{Q}(t, \tilde{S})\bar{\Phi}\hat{X}(t) + \\ &\quad u^T(t)[\bar{G}^T\bar{Q}(t, \tilde{S})\bar{G} + B]u(t) + \text{tr}(\tilde{S}\Sigma_\xi) \end{aligned}$$



对  $W(t)$  关于  $u(t)$  求极小, 即令  $\frac{\partial W(t)}{\partial u(t)} = 0$ , 可以推导出  $t$  时刻的控制为

$$u^*(t) = -L(t)\hat{X}(t)$$

其中,  $L(t), D(t), S(t), \tilde{S}$  和  $T(t)$  与定理 2 中的表达式完全相同. 把  $u^*(t)$  代入  $W(t)$  中, 并用  $\tilde{S}$  代替  $S(t+1)$ , 对应于  $t$  时刻的 Cost-to-go 的似解析表达式仍然为二次型, 即

$$W^*(t) = \hat{X}^T(t)\tilde{S}\hat{X}(t) + T(t) \quad (26)$$

定理得证. □

在  $N-1$  时刻, 定理 2 给出的控制律是最优的. 因为动态规划首次运用时,  $W^*(N)$  中不包含后验概率, 因此无需近似. 而在后续的反向递推中,  $W^*(t+1)$  都与后验概率相关, 通过线性近似以保证动态规划的顺利进行, 导致对应的控制律为次优控制.

控制律是在基本遵循最优性原理 (15) 的基础上导出来的, 因此具有次优性质, 控制律中含有各阶段的后验概率, 而后验概率具有学习性质, 保证了控制律具有学习特点. 既有控制 (优化) 功能又有学习特点的控制就是对偶控制<sup>[14]</sup>.

#### 4 对偶控制算法

依据定理 1 和定理 2, 总结以上结果, 对于本文研究的控制问题, 可用以下算法实施:

步骤 1. 对于给定的  $\varepsilon$ , 用式 (4) 求出模型个数  $s$ ;

步骤 2. 用式 (6) 求出  $\theta_i$ ;

步骤 3. 令  $t = 0$ ;

步骤 4. 用递推方程 (18)~(22) 和式 (16) 求出  $t$  时刻的最优控制;

步骤 5. 测量系统输出  $z(t)$ , 用 Kalman 滤波 (7)~(11) 求出  $\hat{X}(t)$ ;

步骤 6. 用式 (13) 求出后验概率;

步骤 7. 如果  $t$  等于  $N-1$ , 结束; 否则, 返回步骤 4.

下面用一个简单的例子, 来说明本文提出的对偶控制算法的实施过程, 并验证控制器所具有的对偶性质.

$$x(t+1) = \Phi(\alpha)x(t) + G(\alpha)u(t) + \xi(t)$$

$$z(t) = H(\alpha)x(t) + \eta(t)$$

其中,  $\Phi = 0.7 + \alpha, G = 0.9 - \alpha, H = 2\alpha + 0.1$ .

假设初始状态  $x(0)$ , 过程噪声  $\xi(t)$  与测量噪声  $\eta(t)$  均服从均值为 0、方差为 1 的高斯分布  $N(0, 1)$ , 未知参数  $\alpha \in [0, 1.2]$ , 给定  $\varepsilon = 0.6$ , 则,  $s = 3$ ,

$$[\alpha_{\min}, \alpha_{\max}] = [\alpha_0, \alpha_1] \cup [\alpha_1, \alpha_2] \cup [\alpha_2, \alpha_3] = [0, 0.4] \cup [0.4, 0.8] \cup [0.8, 1.2]$$

$\theta$  取每个子区间的中点, 依次为  $\theta_1 = 0.2, \theta_2 = 0.6, \theta_3 = 1$ , 假定  $\alpha$  的真值为 0.95, 则真值落在第三个子区间  $[0.8, 1.2]$  内,  $\alpha$  取三个  $\theta$  值的先验概率假定为  $\tau_1(0) = 1/3, \tau_2(0) = 1/3, \tau_3(0) = 1/3$ , 也就是说最开始参数  $\alpha$  以等概率属于每个子区间, 此时, 参数的不确定性最大.

采用不同的控制策略对系统施行控制, 当未知参数已知时, 对应的控制问题为 LQG, 获得的控制律为最优控制, 对应的性能指标是其他控制的下界. 由于参数未知, 学习过程不可缺少, 因此, 这个下界永远不能达到.

比较各种控制策略优劣的性能指标为:

$$J = E\{Ax^2(25) + \sum_{t=0}^{24} [Ax^2(t) + Bu^2(t)]\}$$

其中,  $A = 0.4, B = 0.3$ .

图 1 是未知参数  $\alpha$  取  $\theta_1, \theta_2$  和  $\theta_3$  分别对应的后验概率演化图. 下面的两条“点实线”为  $\theta_1$  和  $\theta_2$  对应的后验概率, 上面的“圈线”为  $\theta_3$  对应的后验概率, 它们均从  $1/3$  (此时不确定性最大) 出发, 经过 25 个采样时刻,  $\theta_3$  对应的后验概率接近于 1, 而其余两个的后验概率接近于 0, 这表明  $\alpha$  从初始的最大不确定性开始, 通过从系统不断获得测量后, 其不确定性逐渐减小, 直至完全消除, 这是控制器的学习特点.  $\theta_3$  对应的区间为  $[0.8, 1.2]$ , 因此, 利用本文的算法可以获得包含未知参数的最小区间. 依据定理 1,  $\varepsilon$  越小, 包含真值的区间越小, 控制器的学习精度越高. 然而, 区间的个数  $s$  会越大, 对应的模型越多, 自然会消耗更多的时间. 图 1 说明了本文给出的对偶控制具有学习特点.

图 2 中的“实线”代表最优控制, 也就是未知参数  $\alpha$  是真值 0.95 时, 通过求解 LQG 问题得到的最优控制. 图 2 中的“圈线”表示本文推导出的对偶控制, 即假定参数  $\alpha$  未知, 一边控制、一边学习, 可以看出, 两条折线很快几乎重合, 夹在它们之间的面积很小, 这表明新的对偶控制和最优控制几乎有相同的控制效果. 然而, 两者永远不会完全重合, 毕竟控制器对未知参数的真值需要一个学习过程, 这段时间两个控制律不会重合. 其实, 即使学习结束, 理论上两个控制律也会存在差异, 因为,  $\theta_3$  仅是包含真值区间的标识, 而不是真正的真值. 两个控制律的充分接近, 表明用本文方法设计的控制律有控制或者优化功能, 能够迫使系统朝着期望的目标运行, 保证二次性能指标最优.

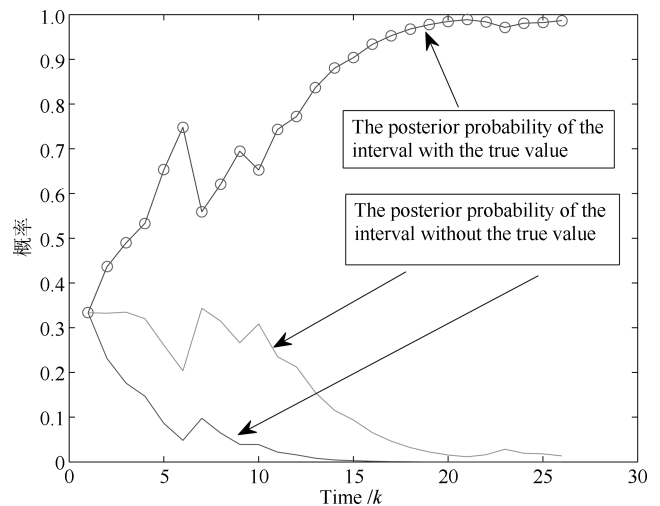


图 1 后验概率演化

Fig. 1 The posterior probabilities evolving

对偶控制在试图控制系统的过程中必须对未知参数进行辨识, 辨识的精度决定了控制的效果, 因而辨识与控制之间有着本质的联系. 从图 2 可以看出, 在开始阶段对偶控制比最优控制能量要大, 其作用展示在后验概率的演化图 1 中, 明显可以看出, 包含真值的区间对应的后验概率趋向 1, 而其他两个区间对应的后验概率趋于 0. 这个事实表明在控制的初始阶段控制器投入了更大的能量去学习未知参数, 一旦学

出了未知参数, 对偶控制的模型与最优控制是相同的, 两者也如图 2 末端表示的那样, 几乎重合. 控制器不再分配能量去学习, 所有能量用来实现控制目标, 充分体现了对偶控制在学习与控制之间的平衡.

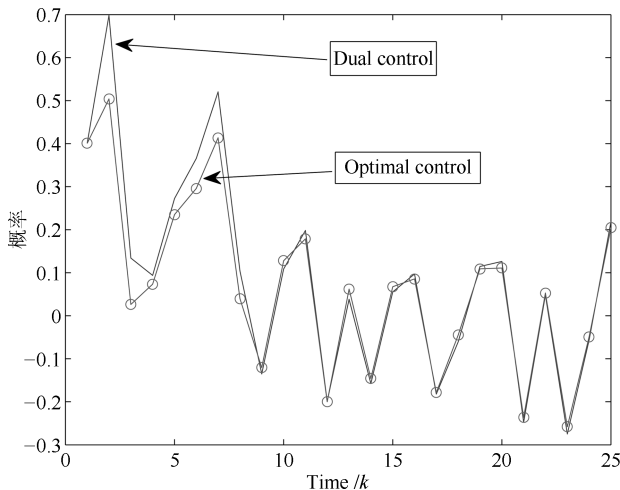


图 2 对偶控制与最优控制比较  
Fig. 2 Comparison of the dual control with the optimal control

系统中未知参数  $\alpha \in [\alpha_{\min}, \alpha_{\max}]$ . 我们作以下规定: 当  $\alpha$  等于真值 0.95 时, 求解相应的 LQG 问题, 所得的控制律自然就是最优控制; 当  $\alpha$  取值于区间的左端点, 即  $\alpha = \alpha_{\min} = 0$  时, 求解相应的 LQG 问题, 所得的控制律称为非对偶控制 1;  $\alpha$  取右端点, 即  $\alpha = \alpha_{\max} = 1.2$  时, 解相应的 LQG 问题, 所得的控制律称为非对偶控制 2; 用本文提出的方法, 导出的控制律称为对偶控制. 对上述系统, 用 4 个不同的控制律进行控制, 会得到它们各自的性能指标. 由于是随机系统, 每次性能指标是不同的, 为了获得可靠的结论, 我们进行 1000 次 Monte Carlo 仿真, 仿真结果如表 1. 可以看出, 对偶控制最接近于最优控制, 其他两个控制对应的性能指标和最优控制有着数量级的差别. 原因在于系统的输出含有未知参数  $\alpha$  的信息, 用输出不断校正后验概率, 使参数的不确定性持续减小, 最终未知参数真值所在区间的中点  $\theta_3$  在控制量的计算中以概率 1 发挥作用, 而其他两个区间参数的代表值几乎不参与控制量的计算. 如果简单地用未知参数的上界或者下界代替真实参数, 正如表中性能指标体现的那样控制效果不堪设想, 因此, 三个控制律中对偶控制最好, 最优控制对应的性能指标是它们永远不能获得的下界.

表 1 不同不确定参数下 Monte Carlo 仿真性能指标比较  
Table 1 Performance index of Monte Carlo runs comparison for different uncertainty parameters

最优控制	非对偶控制 1	非对偶控制 2	对偶控制
30.1702	$5.1102 \times 10^5$	379.9141	38.6375

DUL 算法是自适应对偶控制的一个经典方法<sup>[15]</sup>, 它的基本思想为, 在 DCP 问题中当固定  $\alpha$  为  $\theta_i, i = 1, 2, \dots, s$  时, 其退化为 LQG 问题, 在  $k$  时刻依据动态规划与 Kalman 滤波可以求出最优控制  $u_i^*(k), i = 1, 2, \dots, s$ , 以后验概率  $\tau_i(k)$  为权系数对  $s$  个最优控制  $u_i^*(k)$  进行加权和, 得到  $u^*(k) = \sum_{i=1}^s \tau_i(k)u_i^*(k)$ , 视  $u^*(k)$  为当前时刻施加于系统

的实际控制, 就得到了次优的 DUL 算法. DUL 算法控制器设计简单, 计算量小, 后验概率的学习性质赋予了控制器的学习特点, 是一个优秀的次优学习算法. 然而, DUL 仅用当前的系统信息校正前一时刻的后验概率, 而完全忽略了未来系统的实际输出, 是一种被动学习算法. 本文提出的控制器设计方法除了用到了当前的后验概率, 还在每一个阶段用前一时刻后验概率对 Cost-to-go 进行线性近似, 即控制律的计算用到了未来对系统输出的预测信息, 因此, 本文的方法是一种主动学习算法. 由于是随机系统单次控制效果不能对控制器进行有效评估, 为此, 我们对本文导出的控制律和 DUL 控制律在同样噪声及参数不确定性条件下进行 100 次 Monte Carlo 仿真测试, 从目标函数的统计平均值、目标函数的最大值、目标函数的标准差来比较两个控制律的性能. 表 2 是较结果.

表 2 两个控制律性能比较  
Table 2 Performance comparison for two control laws in example

控制律	DUL 控制	对偶控制
统计平均值	56.4350	39.2685
最大值	148.6953	126.3292
标准差	27.9879	21.2748

在表 2 中目标函数的 100 次统计平均值、最大值和标准差表明对偶控制要优于 DUL 控制, 这表明对偶控制比 DUL 控制在最优性和可靠性都会更好, 原因在于它充分利用了来自系统的信息.

## 5 结论

本文研究了具有未知参数的 LQG 控制问题, 给出了控制问题的数学描述. 由于问题中包含两种不确定性, 采用 Kalman 滤波处理过程噪声和测量噪声产生的不确定性; 用后验概率学习到了包含未知参数真值的给定区间, 该区间的精度可以事先任意设定, 正如仿真例子揭示的那样, 后验概率从最大的不确定性开始, 到控制结束时几乎没有不确定性. 尽管动态规划是求解具有反馈形式控制律的有力工具, 但由于每个阶段的 Cost-to-go 中含有后验概率, 导致动态规划不能顺利进行, 本文用线性近似的方法得到了一个次优控制律. 本文的研究表明, 具有未知参数的 LQG 问题的难点在于 Cost-to-go 与后验概率的非线性依赖关系, 正确和有效处理这一关系是获得控制增益的关键. 除了本文提出的线性近似外, 我们期待其他更好的方法. 另外, 将这种具有学习性质的控制方法用于传感器、执行器中的卡死故障及减振系统中的在线参数辨识, 以提高系统的可靠性, 也是未来进一步要开展的工作.

## References

- Asami T, Nishihara O.  $H_2$  optimization of the three-element type dynamic vibration absorbers. *Journal of Vibration and Acoustics*, 2002, **124**(5): 583–592
- Feldbaum A A. *Optimal Control Systems*. New York: Academic, 1965. 13–16
- Feldbaum A A. Dual control theory I-II. *Automatic Remote Control*, 1960, **21**(4): 1033–1039
- Filatov N M, Unbehauen H. *Adaptive Dual Control Theory and Applications*. Heidelberg: Springer Verlag, 2004. 75–86

- 5 Chen J, Patton R J [Author], Wu Jian-jun [Translator]. *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Beijing: National Defence Industry Press, 2009. 81–94 (Chen J, Patton R J [著], 吴建军 [译]. 动态系统基于模型的鲁棒故障诊断. 北京: 国防工业出版社, 2009. 81–94)
- 6 Zhang Dan. Parameter Identification of Dynamic Model of High Speed Train [Master dissertation], Xi'an University of Technology, China, 2016. (张丹. 高速列车动力学模型的参数辨识和状态估计 [硕士学位论文], 西安理工大学, 中国, 2016.)
- 7 Li Wei-Wei. Vibration Isolation Design of Vehicle Stretcher-supine Body System Based on Wire-rope Spring [Master dissertation], Tianjin University, China, 2013. (李维伟. 基于钢丝绳弹簧的车载担架-卧姿人体系统隔振设计研究 [硕士学位论文], 天津大学, 中国, 2013.)
- 8 Alster J, Bélanger P R. A technique for dual adaptive control. *Automatica*, 1974, **10**(6): 627–634
- 9 Bar-Shalom Y, Wall K D. Dual adaptive control and uncertainty effects in macroeconomic systems optimization. *Automatica*, 1990, **16**(2): 147–156
- 10 Milito R, Padilla C, Padilla R, Cadorn D. An innovations approach to dual control. *IEEE Transactions on Automatic Control*, 1982, **27**(1): 132–137
- 11 Deshpande J G, Upadhyay T N, Lainiotis D G. Adaptive control of linear stochastic systems. *Automatica*, 1973, **9**(1): 107–115
- 12 Fu P L, Li D, Qian F C. Active dual control for Linear-Quadratic Gaussian system with unknown parameters. In: Proceedings of the 15th IFAC World Congress. Barcelona, Spain: Pergamon, 2002. 337–342
- 13 Li D, Qian F C, Fu P L. Research on dual control. *Acta Automatica Sinica*, 2005, **31**(1): 32–42
- 14 Li D, Qian F C, Fu P L. Variance minimization in stochastic systems. *Stochastic Modeling and Optimization*. New York: Springer-Verlag, 2003. 2010–2020
- 15 Li D, Qian F C, Fu P L. Variance minimization approach for a class of dual control problems. *IEEE Transactions on Automatic Control*, 2002, **47**(12): 2010–2020
- 16 Li D, Qian F C, Fu P L. Optimal nominal dual control for discrete-time linear-quadratic Gaussian problems with unknown parameters. *Automatica*, 2008, **44**(1): 119–127
- 17 Qian Fu-Cai, Zhu Shao-Ping, Liu Ding. On LQG problems with unknown noises. *Control Theory & Applications*, 2010, **27**(8): 1017–1022 (钱富才, 朱少平, 刘丁. 噪声未知的 LQG 控制问题研究. 控制理论与应用, 2010, **27**(8): 1017–1022)
- 18 Chen Xiao-Ke. Research on Dual Control in LQG with Uncertainty Parameters Problem [Master dissertation], Xi'an University of Technology, China, 2004. (陈小可. 具有不确定参数的 LQG 对偶控制问题研究 [硕士学位论文], 西安理工大学, 中国, 2004.)
- 19 Wu Sen-Tang. *Stochastic Robustness Analysis and Design for Guidance and Control System of Winged Missile*. Beijing: National Defence Industry Press, 2010. (吴森堂. 飞航导弹制导控制系统随机鲁棒分析与设计. 北京: 国防工业出版社, 2010.)
- 20 Qian F C, Gao J J, Li D. Complete statistical characterization of discrete-time LQG and cumulant control. *IEEE Transactions on Automatic Control*, 2012, **57**(8): 2110–2115
- 尚 婷 西安理工大学自动化与信息工程学院博士研究生. 主要研究方向为最优控制, 随机控制, 系统辨识, 故障诊断. E-mail: tshang0722@126.com
- (SHANG Ting Ph.D. candidate at the School of Automation and Information Engineering, Xi'an University of Technology. Her research interest covers the optimal control, stochastic control, systems identification, fault diagnosis.)
- 钱富才 西安理工大学自动化与信息工程学院教授. 主要研究方向为随机控制, 系统辨识, 非线性控制, 最优控制, 故障诊断和全球定位系统. 本文通信作者. E-mail: qianfc@xaut.edu.cn (QIAN Fu-Cai Professor at the School of Automation and Information Engineering, Xi'an University of Technology. His research interest covers stochastic control, systems identification, nonlinear control, optimal control, fault diagnosis and global positioning system. Corresponding author of this paper.)
- 张晓艳 西安理工大学自动化与信息工程学院博士研究生. 主要研究方向为对偶控制, 最优控制, 故障诊断, 随机控制, 容错控制, 系统辨识. E-mail: xyzhang\_2016@163.com (ZHANG Xiao-Yan Ph.D. candidate at the School of Automation and Information Engineering, Xi'an University of Technology. Her research interest covers dual control, optimal control, fault diagnosis, stochastic control, tolerant-fault control and system identification.)
- 谢 国 西安理工大学自动化与信息工程学院副教授. 2013 年获得日本大学工学博士学位. 主要研究方向为轨道交通系统的安全性和可靠性, 最优控制, 和随机控制. E-mail: guoxie@xaut.edu.cn (XIE Guo Associate professor at the School of Automation and Information Engineering, Xi'an University of Technology. He received his Ph.D. degree from Vihon University, Japan in 2013. His research interest covers safety and reliability of railway system, optimal control and stochastic control.)