DOI: 10.7641/CTA.2017.70674

非线性零和微分对策的事件触发自适应动态规划算法

崔黎黎[1†], 张 勇[1], 张 欣[2]

(1. 沈阳师范大学 科信软件学院, 辽宁 沈阳 110034; 2. 中国石油大学(华东) 信息与控制工程学院, 山东 青岛 266580)

摘要: 针对一类非线性零和微分对策问题, 本文提出了一种事件触发自适应动态规划(event-triggered adaptive dynamic programming, ET–ADP)算法在线求解其鞍点. 首先, 提出一个新的自适应事件触发条件. 然后, 利用一个输入为采样数据的神经网络(评价网络)近似最优值函数, 并设计了新型的神经网络权值更新律使得值函数、控制策略及扰动策略仅在事件触发时刻同步更新. 进一步地, 利用Lyapunov稳定性理论证明了所提出的算法能够在线获得非线性零和微分对策的鞍点且不会引起Zeno行为. 所提出的ET–ADP算法仅在事件触发条件满足时才更新值函数、控制策略和扰动策略, 因而可有效减少计算量和降低网络负荷. 最后, 两个仿真例子验证了所提出的ET–ADP算法的有效性.

关键词: 自适应动态规划; 非线性零和微分对策; 事件触发; 神经网络; 最优控制

引用格式: 崔黎黎, 张勇, 张欣. 非线性零和微分对策的事件触发自适应动态规划算法. 控制理论与应用, 2018, 35(5): 610 – 618

中图分类号: TP273 **文献标识码**: A

Event-triggered adaptive dynamic programming algorithm for the nonlinear zero-sum differential games

CUI Li-li[1†], ZHANG Yong[1], ZHANG Xin[2]

(1. Software College, Shenyang Normal University, Shenyang Liaoning 110034, China;
2. College of information and control engineering, China University of Petroleum, Qingdao Shandong 266580, China)

Abstract: In this paper, an event-triggered adaptive dynamic programming algorithm (ET–ADP) is proposed to solve the saddle point of a class of nonlinear zero-sum differential games. Firstly, a new adaptive event-triggered condition is proposed. Then, a neural network (critic network) with the sampled state as its input is utilized to approximate the optimal value function. The new neural network weights updating law is designed to enable the value function, the control strategy and the disturbance strategy to be updated synchronously only at the event-triggered time. Further, the Lyapunov stability theory is used to prove that the proposed algorithm can obtain the saddle point of nonlinear zero-sum differential games online and avoid the occurrence of Zeno behavior. In the proposed ET–ADP algorithm, the value function, the control strategy and the disturbance strategy are updated only when the event-triggered condition is satisfied, as a result of which the computational burden is reduced and the network burden is eased effectively. Finally, two simulation examples validate the effectiveness of the proposed ET–ADP algorithm.

Key words: adaptive dynamic programming; nonlinear zero-sum differential games; event-triggered; optimal control

Citation: CUI LiLi, ZHANG Yong, ZHANG Xin. Event-triggered adaptive dynamic programming algorithm for the nonlinear zero-sum differential games. *Control Theory & Applications*, 2018, 35(5): 610 – 618

1 引言(Introduction)

在实际中, 有一大类非线性系统是由一个以上的控制器所控制, 这些控制器共同作用在一个系统上并相互影响, 在某个性能指标约束下合作或者对抗从而形成对策. 其中, 零和微分对策是最具代表性的一类, 近年来一直是控制领域的研究热点, 在导弹拦截、电力系统、复杂工业控制和多智能体协调控制[1–4]等领域具有广泛应用.

求解零和微分对策的本质就是找到对策的最优解, 也即是对策的鞍点. 对于非线性微分对策而言, 获得

收稿日期: 2017−09−15; 录用日期: 2017−12−30.

†通信作者. E-mail: cuilili8396@163.com; Tel.: +86 24-22921176.

本文责任编委: 梅生伟.

国家自然科学基金项目(61703289), 山东省自然科学基金项目(BX2015DX009), 辽宁省高等学校基本科研项目专项资金(LQN201720, LQN201702), 沈阳师范大学科技项目(L201510)资助.

Supported by the National Natural Science Foundation of China (61703289), the National Natural Science Foundation of Shandong Province (BX2015DX009), the Special Fund of Liaoning Province Universities' Fundamental Scientific Research Projects (LQN201720, LQN201702) and the Science and Technology Project of Shenyang Normal University (L201510).

其鞍点需要求解Hamilton-Jacobi-Issacs(HJI)方程, 而该方程是非线性偏微分方程, 具有本质非线性, 因而很难直接求解. 考虑到求解HJI方程非常困难, 一些求解方法被提出以获得HJI方程的近似解, 如泰勒展开法、有限元法等. 然而, 这些方法计算负担比较大. 近年来, 自适应动态规划(adaptive dynamic programming, ADP)作为一种有效的智能控制方法吸引了广大学者的注意. ADP方法的基本原理是利用函数近似结构(如神经网络等)来近似性能指标函数, 然后根据贝尔曼最优性原理更新函数近似结构的参数, 从而获得最优性能指标函数和最优控制. 由于具有自学习和优化能力, ADP方法在求解非线性系统最优控制方面具有强大优势, 目前已在一些文献中被用来解决非线性微分对策问题. 文献[5]提出了在线同步策略迭代算法得到了L_2增益最优控制问题中出现的非线性零和微分对策的近似最优反馈策略. 文献[6]提出了单网络ADP方法解决了一类非线性二人非零和对策问题. 文献[7]针对非线性多人微分对策问题提出了无需系统模型已知的同步策略迭代算法. 文献[8]提出了迭代ADP算法分别解决了鞍点存在时和鞍点不存在时的一类连续时间非线性二人零和微分对策问题.

上述大多数方法都是在时间触发机制下给出的, 其中控制器均采用的是连续更新的方式. 然而, 在很多实际情况下由于通讯带宽的限制控制器和被控对象间的持续通讯并不太容易实现, 因而限制了上述方法在实际当中的应用. 与时间触发控制不同, 事件触发控制不再是传统的连续控制或者周期采样控制, 而是当事件触发条件成立时控制器才进行更新, 因而在相同时间内可以大大降低控制器的更新频率和数据传输次数. 由此, 文献[9–10]提出了最优自适应事件触发控制算法, 解决了一类连续非线性系统的最优控制问题. 文献[11]基于ADP方法研究了一类不确定连续非线性系统的鲁棒最优控制问题. 文献[12]考虑了输入受限的局部动态未知的连续非线性系统的最优控制问题. 文献[13]提出了基于事件触发的观测器和最优控制器解决了一类内部状态未知的连续非线性系统的最优控制问题. 文献[14]将带有不确定性的非线性系统鲁棒控制问题转化为标称系统的最优控制问题, 提出了基于神经动态规划的事件触发自适应鲁棒控制策略. 上述结果均考虑的是仅有一个控制器的非线性系统的最优控制问题, 而两个控制器作用下的非线性系统零和微分对策问题的研究成果目前尚不多见. 文献[15]采用ADP算法研究了Buck型DC–DC变换器系统的输出跟踪问题. 文献[16]针对一类典型的带有控制约束的非线性离散时间系统, 提出了一种基于ADP算法的多设定值跟踪控制方法. 文献[17]针对离散非线性系统的二人零和微分对策问题提出了一个事件触发的无模型全局二次启发式规划算法. 文献[18–19]将连续非线性系统的H_∞控制问题转化为非线性零和微分对策问题, 然后利用自适应评价学习方法设计了基于事件触发的最优控制策略和基于时间触发的最优扰动策略. 值得指出的是, 现有的基于ADP的事件触发最优控制策略均是仅控制网络在事件触发时刻更新, 而评价网络仍然采取的是连续更新的方式.

受上述文献启发, 本文提出了一种事件触发近似动态规划(event-triggered adaptive dynamic programming algorithm, ET–ADP)算法在线求解一类非线性零和微分对策问题的鞍点. 首先, 提出了一个新的自适应事件触发条件. 然后, 在所提出的自适应触发条件下利用一个输入为采样数据的神经网络(评价网络)近似最优值函数, 接着设计了新型的评价网络权值更新律使其在线学习HJI方程的解,从而得到非线性零和微分对策的鞍点. 与文献[18–19]不同, 本文所提出的算法采用状态的采样数据作为评价网络的输入, 仅在事件触发时刻更新评价网络权值, 并且设计的最优控制策略和最优扰动策略均是仅在事件触发条件满足时与评价网络一起同步更新, 因而可有效减少计算量和降低网络传输负荷. 利用Lyapunov稳定性理论证明了所提出的的算法能够在线获得零和微分对策的鞍点和保证闭环系统的一致最终有界稳定性, 并且不会引起Zeno行为. 最后, 通过数值仿真验证所提出算法的有效性.

2　问题描述(Problem statement)

考虑如下的连续非线性系统:

$$\dot{x} = f(x) + g(x)u + k(x)w, \tag{1}$$

其中: $x(t) \in \mathbb{R}^n$为系统状态, $u(t) \in \mathbb{R}^m$为控制输入, $w(t) \in \mathbb{R}^q$为扰动输入. 假设$f(x)$, $g(x)$, $k(x)$局部Lipschitz连续, 且$f(0) = 0$. 定义值函数为

$$V(x, u, w) = \int_t^\infty r(x(\tau), u(\tau), w(\tau))\mathrm{d}\tau, \tag{2}$$

其中$r(x, u, w) = Q(x) + u^{\mathrm{T}}Ru - \gamma^2 w^{\mathrm{T}}Pw$. 令$Q(x) \geqslant 0$, $R \in \mathbb{R}^{m\times m}$和$P \in \mathbb{R}^{q\times q}$为正定矩阵, $\gamma > 0$为正常数. 控制u和扰动w可看作是非线性零和微分对策的局中人, 其中: 控制u的作用是最小化值函数(2), 扰动w的作用是最大化值函数(2). 本文的控制目标是: 针对系统(1)设计基于事件触发ADP设计最优的控制策略u和扰动策略w从而得到非线性零和微分对策的鞍点(u,w). 若存在最优值函数$V^*(x)$满足下式:

$$V^* = \min_u \max_w \int_t^\infty r(x(\tau), u(\tau), w(\tau))\mathrm{d}\tau = \max_w \min_u \int_t^\infty r(x(\tau), u(\tau), w(\tau))\mathrm{d}\tau, \tag{3}$$

则上述非线性零和微分对策问题具有唯一解. 根据式(2), 定义如下的Hamilton函数:

$$H(x,u,w)=\nabla V(f(x)+g(x)u+k(x)w)+r(x,u,w), \tag{4}$$

其中$\nabla V=\dfrac{\partial V}{\partial x}$. 根据Bellman最优性原理可得最优性条件为

$$\min_u\max_w H(x,\nabla V^*,u,w)=0. \tag{5}$$

相应的最优控制策略u^*和扰动策略w^*如下:

$$\begin{cases} u^*(x)=-\dfrac{1}{2}R^{-1}g^{\mathrm{T}}(x)\nabla V^*(x),\\ w^*(x)=\dfrac{1}{2\gamma^2}P^{-1}k^{\mathrm{T}}(\,x)\nabla V^*(x). \end{cases} \tag{6}$$

将式(6)代入式(5)可得如下的Hamilton-Jacobi-Isaacs (HJI)方程:

$$Q(x)+\nabla V^{\mathrm{T}}f(x)-\frac{1}{4}\nabla V^{\mathrm{T}}g(x)R^{-1}g^{\mathrm{T}}(x)\nabla V+\frac{1}{4\gamma^2}\nabla V^{\mathrm{T}}k(x)P^{-1}k^{\mathrm{T}}(x)\nabla V=0. \tag{7}$$

若能够获得HJI方程的解V^*, 则可得相应的最优控制策略u^*和最优扰动策略w^*. 然而, HJI方程本质上为非线性偏微分方程, 因而一般情况下很难求得其解析解. 本文将提出一种新的ET–ADP算法, 该算法不仅能够在线学习HJI方程(7)的解从而得到非线性零和微分对策的鞍点, 还能减少计算量和降低网络负荷.

3 ET–ADP 算法设计 (ET–ADP algorithm design)

为了提出ET–ADP算法, 首先引入事件触发采样系统, 即在事件触发时刻对系统状态采样(事件触发时刻记为$t_1,t_2,\cdots,t_j,t_{j+1},\cdots$). 当事件触发时刻为$t_j$时, 定义状态采样为

$$\bar{x}_j=x(t_j),\ t=t_j. \tag{8}$$

定义测量误差为

$$e_j(t)=\bar{x}_j-x(t),\ t_j<t\leqslant t_{j+1}, \tag{9}$$

则当$t=t_j$时, 有$e_j(t_j)=0$.

本文提出如下的自适应事件触发条件:

$$\mathscr{D}(\|e_j\|)>\xi(x,\|\hat{W}_{\mathrm{c}}\|), \tag{10}$$

其中: $\xi(x,\|\hat{W}_{\mathrm{c}}\|)=\min\{CQ(x)/\beta_1,\sqrt{CQ(x)/\beta_2}\}$, $0<C<1$为常数, β_1和β_2见式(42)和式(43), $\hat{W}_{\mathrm{c}}$为评价网络理想权值W_{c}的估计. 当$\|x\|>\mathcal{B}_x$时死区算子$\mathscr{D}(\cdot)=\cdot$, $\mathcal{B}_x$如式(46)所示; 否则, $\mathscr{D}(\cdot)=0$, 该事件触发条件和下面设计的事件触发最优控制器可保证闭环系统的稳定性.

与时间触发ADP算法不同, 本文提出的ET–ADP算法采用状态采样$\bar{x}_j$作为反馈信号, 将其代入式(6)可得事件触发最优控制策略$u^*(\bar{x}_j)$和最优扰动策略$w^*(\bar{x}_j)$如下:

$$\begin{cases} u^*(\bar{x}_j)=-\dfrac{1}{2}R^{-1}g^{\mathrm{T}}(\bar{x}_j)\nabla V^*(\bar{x}_j),\\ w^*(\bar{x}_j)=\dfrac{1}{2\gamma^2}P^{-1}k^{\mathrm{T}}(\bar{x}_j)\nabla V^*(\bar{x}_j), \end{cases} \tag{11}$$

其中$t_j<t\leqslant t_{j+1}$. 由式(11)可以看出, 若V^*已知, 则可获得$u^*(\bar{x}_j)$和$w^*(\bar{x}_j)$. 控制策略和扰动策略仅在事件触发时刻更新, 而在相邻的两个事件间则保持不变, 直到下一个事件触发时刻到来. 然而, 由于V^*未知, 无法由式(11)得到最优控制策略$u^*(\bar{x}_j)$和最优扰动策略$w^*(\bar{x}_j)$. 接下来, 本文提出采用一个以状态采样$\bar{x}_j$为输入的神经网络近似最优值函数. 根据Weierstrass定理, 最优值函数可由如下的神经网络(评价网络)近似:

$$V^*(x)=W_{\mathrm{c}}^{\mathrm{T}}\phi(x)+\epsilon_{\mathrm{c}}=W_{\mathrm{c}}^{\mathrm{T}}\phi(\bar{x}_j)+\bar{\epsilon}(\bar{x}_j,e_j), \tag{12}$$

其中: W_{c}为评价网络理想权值, $\phi(x)$为评价网络激活函数, $\bar{\epsilon}(\bar{x}_j,e_j)=W_{\mathrm{c}}^{\mathrm{T}}(\phi(x)-\phi(\bar{x}_j))+\epsilon_{\mathrm{c}}(\bar{x}_j-e_j)$ 为评价网络的近似误差. 将式(12)代入式(7)可得以如下形式的基于状态采样$\bar{x}_j$和测量误差$e_j(t)$的HJI方程:

$$Q(x)+(\nabla_x\phi^{\mathrm{T}}(\bar{x}_j)\hat{W}_{\mathrm{c}}+\nabla_x\bar{\epsilon}(\bar{x}_j,e_j))^{\mathrm{T}}f(x)-\frac{1}{4}(\nabla_x\phi^{\mathrm{T}}(\bar{x}_j)\hat{W}_{\mathrm{c}}+\nabla_x\bar{\epsilon}(\bar{x}_j,e_j))(D_1-\frac{D_2}{\gamma^2})\times(\nabla_x\phi^{\mathrm{T}}(\bar{x}_j)\hat{W}_{\mathrm{c}}+\nabla_x\bar{\epsilon}(\bar{x}_j,e_j))=0, \tag{13}$$

其中:

$$D_1=g(\bar{x}_j)R^{-1}g^{\mathrm{T}}(\bar{x}_j),$$
$$D_2=k(\bar{x}_j)P^{-1}k^{\mathrm{T}}(\bar{x}_j),$$
$$\nabla_x\bar{\epsilon}(\bar{x}_j,e_j)=\frac{\partial\bar{\epsilon}(\bar{x}_j,e_j)}{\partial x},$$

则最优值函数的估计为

$$\hat{V}(\bar{x}_j)=\hat{W}_{\mathrm{c}}^{\mathrm{T}}\phi(\bar{x}_j),\ t_j<t\leqslant t_{j+1}. \tag{14}$$

将式(14)代入式(11)可得事件触发的控制策略和扰动策略如下:

$$\begin{cases} u(\bar{x}_j)=-\dfrac{1}{2}R^{-1}g^{\mathrm{T}}(\bar{x}_j)\nabla_x\phi^{\mathrm{T}}(\bar{x}_j)\hat{W}_{\mathrm{c}},\\ w(\bar{x}_j)=\dfrac{1}{2\gamma^2}P^{-1}k^{\mathrm{T}}(\bar{x}_j)\nabla_x\phi^{\mathrm{T}}(\bar{x}_j)\hat{W}_{\mathrm{c}}, \end{cases} \tag{15}$$

其中$t_j<t\leqslant t_{j+1}$.

利用式(4)(14)–(15)可得近似Hamilton函数如下:

$$\hat{H}(\bar{x}_j,\hat{W}_{\mathrm{c}})=Q(\bar{x}_j)+\hat{W}_{\mathrm{c}}^{\mathrm{T}}\nabla_x\phi(\bar{x}_j)f(\bar{x}_j)-\frac{1}{4}\hat{W}_{\mathrm{c}}^{\mathrm{T}}\times\nabla_x\phi(\bar{x}_j)(D_1-\frac{D_2}{\gamma^2})\nabla_x\phi^{\mathrm{T}}(\bar{x}_j)\hat{W}_{\mathrm{c}},\quad t_j<t\leqslant t_{j+1}. \tag{16}$$

由式(8)可得事件触发时刻t_j时的近似Hamilton函数如下:

$$\hat{H}^+(x,\hat{W}_{\mathrm{c}})=$$

$$Q(x)+\hat{W}_{\rm c}^{\rm T}\nabla_x\phi(x)f(x)-\frac{1}{4}\hat{W}_{\rm c}^{\rm T}\nabla_x\phi(x)\times$$
$$(D_1-\frac{D_2}{\gamma^2})\nabla_x\phi^{\rm T}(x)\hat{W}_{\rm c},\ t=t_j. \tag{17}$$

然后, 利用改进的梯度下降法设计评价网权值调节律如下:

$$\begin{cases}\hat{W}_{\rm c}^{+}=\hat{W}_{\rm c}-\dfrac{\alpha_{\rm c}\hat{\sigma}\hat{H}^{+}(x,\hat{W}_{\rm c})}{(1+\hat{\sigma}^{\rm T}\hat{\sigma})^2},\ t=t_j,\\ \dot{\hat{W}}_{\rm c}=0, \qquad\qquad t_j<t\leqslant t_{j+1},\end{cases} \tag{18}$$

其中: $\alpha_{\rm c}>0$,

$$\hat{\sigma}=\nabla_x\phi(x)(f(x)-\frac{1}{2}(D_1-\frac{D_2}{\gamma^2})\nabla_x\phi^{\rm T}(x)\hat{W}_{\rm c}).$$

定义评价网权值估计误差为

$$\tilde{W}_{\rm c}=W_{\rm c}-\hat{W}_{\rm c}, \tag{19}$$

则评价网权值估计误差动态为

$$\begin{cases}\tilde{W}_{\rm c}^{+}=\tilde{W}_{\rm c}+\dfrac{\alpha_{\rm c}\hat{\sigma}\hat{H}^{+}(x,\hat{W}_{\rm c})}{(1+\hat{\sigma}^{\rm T}\hat{\sigma})^2},\ t=t_j,\\ \dot{\tilde{W}}_{\rm c}=0, \qquad\qquad t_j<t\leqslant t_{j+1}.\end{cases} \tag{20}$$

利用式(7)可将式(17)重写为如下形式:

$$\hat{H}^{+}(x,\tilde{W}_{\rm c})=-\tilde{W}_{\rm c}^{\rm T}\hat{\sigma}+\frac{1}{4}\tilde{W}_{\rm c}^{\rm T}\nabla_x\phi(x)(D_1-\frac{D_2}{\gamma^2})\times$$
$$\nabla_x\phi^{\rm T}(x)\tilde{W}_{\rm c}-\epsilon_{\rm h},\ t=t_j, \tag{21}$$

其中

$$\epsilon_{\rm h}=\nabla_x^{\rm T}\epsilon_{\rm c}(f(x)+g(x)u^*+k(x)w^*)-$$
$$\frac{1}{4}\nabla_x^{\rm T}\epsilon_{\rm c}(D_1-\frac{D_2}{\gamma^2})\nabla_x\epsilon_{\rm c}.$$

由式(1)(6)(11)可得$t=t_j$时闭环系统动态方程如下:

$$\dot{x}=$$
$$f(x)+g(x)u^*(x)+k(x)w^*(x)+$$
$$\frac{1}{2}g(x)R^{-1}(g^{\rm T}(x)-g^{\rm T}(\bar{x}_j))\nabla_x\phi^{\rm T}(x)W_{\rm c}+$$
$$\frac{1}{2}g(x)R^{-1}g^{\rm T}(\bar{x}_j)\nabla_x\phi^{\rm T}(x)\tilde{W}_{\rm c}+$$
$$\frac{1}{2}g(x)R^{-1}g^{\rm T}(\bar{x}_j)(\nabla_x\phi^{\rm T}(x)-\nabla_x\phi^{\rm T}(\bar{x}_j))\hat{W}_{\rm c}+$$
$$\frac{1}{2}g(x)R^{-1}g^{\rm T}(x)\nabla_x\epsilon_{\rm c}-$$
$$\frac{1}{2\gamma^2}k(x)P^{-1}(k^{\rm T}(x)-k^{\rm T}(\bar{x}_j))\nabla_x\phi^{\rm T}(x)W_{\rm c}-$$
$$\frac{1}{2\gamma^2}k(x)P^{-1}k^{\rm T}(\bar{x}_j)\nabla_x\phi^{\rm T}(x))\tilde{W}_{\rm c}-$$
$$\frac{1}{2\gamma^2}k(x)P^{-1}k^{\rm T}(\bar{x}_j)(\nabla_x\phi^{\rm T}(x)-\nabla_x\phi^{\rm T}(\bar{x}_j)\hat{W}_{\rm c}-$$
$$\frac{1}{2\gamma^2}k(x)P^{-1}k^{\rm T}(x)\nabla_x\epsilon_{\rm c},\ t_j<t\leqslant t_{j+1}. \tag{22}$$

本文提出的ET-ADP算法总结如下:

步骤 1　初始化系统状态$x(0)$和评价网权值$\hat{W}_{\rm c}(0)$及初始稳定控制律u_0和w_0, 并给出计算精度$\zeta>0$. 然后, 将u_0和w_0作用到系统(1)上;

步骤 2　检查事件触发条件(10)是否满足. 若在t_j时刻事件触发条件满足, 则对系统状态采样得到$\bar{x}_j$, 进一步得到$e_j=0$, 更新评价网权值:

$$\hat{W}_{\rm c}^{+}=\hat{W}_{\rm c}+\frac{\alpha_{\rm c}\hat{\sigma}\hat{H}^{+}(x,\hat{W}_{\rm c})}{(1+\hat{\sigma}^{\rm T}\hat{\sigma})^2},\ t=t_j, \tag{23}$$

并根据式(15)更新控制策略u和扰动策略w; 若不满足, 则评价网权值$\hat{W}_{\rm c}$、控制策略u和扰动策略w保持不变.

步骤 3　如果$\|\hat{W}_{\rm c}^{+}-\hat{W}_{\rm c}\|<\zeta$, 则停止, 由式(14)得到近似最优值函数; 否则, 将步骤2中得到的控制策略u和扰动策略w作用到系统(1)上并转到步骤2.

注 1　文献[18–19]中评价网络的权值是连续更新的, 而本文所提出的的ET-ADP算法仅在事件触发时刻同步更新评价网络权值$\hat{W}_{\rm c}$、控制策略$u(\bar{x}_j)$和扰动策略$w(\bar{x}_j)$. 当事件触发条件不满足时, 评价网络权值$\hat{W}_{\rm c}$、控制策略$u(\bar{x}_j)$和扰动策略$w(\bar{x}_j)$保持不变, 由零阶保持器使得控制输入和扰动输入保持连续. 与文献[18–19]相比, 本文提出的方法所需计算量大大降低.

本文所提出的的ET-ADP控制系统结构如图1所示.

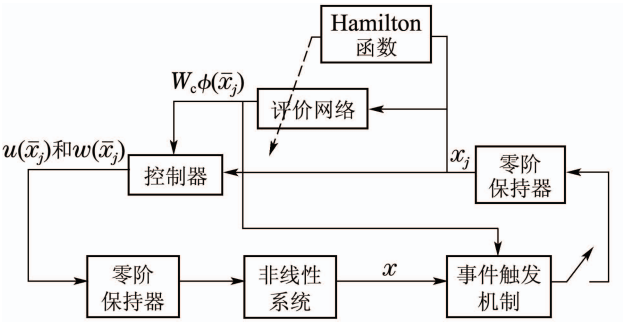

图 1　ET-ADP控制系统结构图

Fig. 1　The diagram of the ET-ADP control system

4　稳定性分析(Stability analysis)

为了便于分析, 首先给出下列假设.

假设 1　激活函数是Lipschitz连续的, 即

$$\|\nabla_x\phi^{\rm T}(x)-\nabla_x\phi^{\rm T}(\bar{x}_j)\|\leqslant l\|e_j\|,$$

其中l为正数.

假设 2　$g(\cdot)$和$k(\cdot)$有界, 即

$$g_{\rm m}\leqslant\|g(\cdot)\|\leqslant g_{\rm M},\ k_{\rm m}\leqslant\|k(\cdot)\|\leqslant k_{\rm M},$$

其中$g_{\rm m}$, $g_{\rm M}$, $k_{\rm m}$, $k_{\rm M}$均为正数.

假设 3　评价网络的理想权值、激活函数及其导数、近似误差及其导数均有界, 即

$\|W_{\rm c}\| \leqslant W_{\rm cM},\ \phi_{\rm m} \leqslant \|\phi(\cdot)\| \leqslant \phi_{\rm M},$

$\|\epsilon_{\rm c}\| \leqslant \epsilon_{\rm cM},\ \|\nabla_x\phi(\cdot)\| \leqslant \phi'_{\rm M},\ \|\nabla_x\epsilon_{\rm c}(\cdot)\| \leqslant \epsilon'_{\rm cM},$

其中$W_{\rm cM}, \phi_{\rm m}, \phi_{\rm M}, \epsilon_{\rm cM}, \phi'_{\rm cM}, \epsilon'_{\rm cM}$均为正数.

定理 1 考虑系统(1)和最优控制策略和扰动策略(6), 假设系统状态满足持续激励条件, 则下列不等式成立:

$$\|f(x) + gu^* + kw^*\| \geqslant \psi, \tag{24}$$

其中$\psi = \dfrac{\lambda_{\min}(Q)\eta_x^2}{\phi'_{\rm M}W_{\rm cM} + \epsilon_{\rm cM}}$.

证 根据最优值函数定义可得

$$\|\nabla V^*(f(x) + g(x)u^* + k(x)w^*)\| \leqslant -x^{\rm T}Qx. \tag{25}$$

由系统状态满足持续激励条件可得存在一个小的正整数$\eta_x > 0$使得$\|x\| \geqslant \eta_x$. 对上式取范数, 并利用式(12)及不等式$\|a^{\rm T}b\| \leqslant \|a\|\|b\|$, 可得式(22). 证毕.

定理 2 对任意正整数$\kappa > 0$, 下列不等式满足:

$$\begin{aligned} -\frac{\tilde{W}_{\rm c}^{\rm T}\hat{\sigma}^{\rm T}\hat{\sigma}\tilde{W}_{\rm c}}{(1+\hat{\sigma}^{\rm T}\hat{\sigma})^2} \leqslant\ & -\frac{1}{4(\kappa+1)}\phi_{\rm m}'^4\|D_1 - \frac{D_2}{\gamma^2}\|_{\rm m}^2\|\tilde{W}_{\rm c}\|^4 + \\ & \frac{1}{\kappa}\phi_{\rm m}'^2\psi^2\|\tilde{W}_{\rm c}\|^2. \end{aligned} \tag{26}$$

证 利用$\hat{\sigma}$的定义, Young不等式及式(6)可得

$$\begin{aligned} & -\tilde{W}_{\rm c}^{\rm T}\hat{\sigma}^{\rm T}\hat{\sigma}\tilde{W}_{\rm c} \leqslant \\ & \frac{1}{\kappa}\tilde{W}_{\rm c}^{\rm T}(-\nabla_x\phi(x)(f(x)+g(x)u^* + \\ & k(x)w^*))^{\rm T}(-\nabla_x\phi(x)(f(x)+g(x)u^* + \\ & k(x)w^*))\tilde{W}_{\rm c} - \frac{1}{4(\kappa+1)}\tilde{W}_{\rm c}^{\rm T}(\nabla_x\phi(x)(D_1 - \frac{D_2}{\gamma^2}))\cdot \\ & \nabla_x\phi^{\rm T}(x)\tilde{W}_{\rm c})^{\rm T} \times (\nabla_x\phi(x)\cdot \\ & (D_1 - \frac{D_2}{\gamma^2})\nabla_x\phi^{\rm T}(x)\tilde{W}_{\rm c})\tilde{W}_{\rm c}. \end{aligned} \tag{27}$$

利用假设1及定理1即可证明式(27)成立. 证毕.

定理 3 考虑非线性系统(1), 事件触发控制策略和扰动策略为(16), 评价网权值调节律为(18), 自适应事件触发条件为(10). 假设系统状态满足持续激励条件, 则评价网络权值估计误差$\tilde{W}_{\rm c}$一致最终有界.

证 首先, 选取如下的Lyapunov函数:

$$L_{\rm W} = {\rm tr}\{\tilde{W}_{\rm c}^{\rm T}\tilde{W}_{\rm c}\}, \tag{28}$$

然后, 分两种情况证明$\tilde{W}_{\rm c}$的有界性.

情况 1 当$t_j < t \leqslant t_{j+1}$时: 对Lyapunov函数(28)求导可得

$$\dot{L}_{\rm W} = 2{\rm tr}\{\tilde{W}_{\rm c}^{\rm T}\dot{\tilde{W}}_{\rm c}\} = 0. \tag{29}$$

情况 2 当$t = t_j$时: 对Lyapunov函数(28)求差分可得

$$\Delta L_{\rm W} = {\rm tr}\{\tilde{W}_{\rm c}^{+{\rm T}}\tilde{W}_{\rm c}^{+}\} - {\rm tr}\{\tilde{W}_{\rm c}^{\rm T}\tilde{W}_{\rm c}\}. \tag{30}$$

将式(19)代入式(30)可得

$$\begin{aligned} \Delta L_{\rm W} = & [\frac{\alpha_{\rm c}\hat{\sigma}\hat{H}^+(x,\hat{W}_{\rm c})}{(1+\hat{\sigma}^{\rm T}\hat{\sigma})^2}]^{\rm T}[\frac{\alpha_{\rm c}\hat{\sigma}\hat{H}^+(x,\hat{W}_{\rm c})}{(1+\hat{\sigma}^{\rm T}\hat{\sigma})^2}] + \\ & \frac{2\alpha_{\rm c}\tilde{W}_{\rm c}^{\rm T}\hat{\sigma}\hat{H}^{+{\rm T}}(x,\hat{W}_{\rm c})}{(1+\hat{\sigma}^{\rm T}\hat{\sigma})^2}. \end{aligned} \tag{31}$$

利用假设1、定理1、定理2及Young不等式和式(21)可得

$$\begin{aligned} & \Delta L_{\rm W} = \\ & -\alpha_{\rm c}(\frac{1}{4} - 3\alpha_{\rm c})\frac{\tilde{W}_{\rm c}^{\rm T}\hat{\sigma}\hat{\sigma}^{\rm T}\tilde{W}_{\rm c}}{(1+\hat{\sigma}\hat{\sigma}^{\rm T})^2} + \\ & \frac{\alpha_{\rm c}}{2(1+\hat{\sigma}\hat{\sigma}^{\rm T})^2}(\frac{1}{\kappa}\phi_{\rm m}'^2\psi^2\|\tilde{W}_{\rm c}\|^2 - \\ & \frac{1}{4(\kappa+1)}\phi_{\rm m}'^4\|D_1 - \frac{D_2}{\gamma^2}\|_{\rm m}^2\|\tilde{W}_{\rm c}\|^4) + \\ & \frac{4\alpha_{\rm c}+3\alpha_{\rm c}^2}{(1+\hat{\sigma}\hat{\sigma}^{\rm T})^2}\|\epsilon_{\rm h}\|^2 + \frac{\alpha_{\rm c}+3\alpha_{\rm c}^2}{16(1+\hat{\sigma}\hat{\sigma}^{\rm T})^2} \\ & \phi_{\rm M}'^2\|D_1 - \frac{D_2}{\gamma^2}\|_{\rm M}^2\|\tilde{W}_{\rm c}\|^4. \end{aligned} \tag{32}$$

选取$\alpha_{\rm c} = \min\{3/4, (1-\kappa)/(3(\kappa+1))\},\ 0 < \kappa < 1,$ 可得

$$\Delta L_{\rm W} \leqslant -\theta_2(\|\tilde{W}_{\rm c}\|^2 - \frac{\theta_1}{2\theta_2})^2 + \frac{\theta_1^2}{4\theta_2} + \theta_3\|\epsilon_{\rm h}\|^2, \tag{33}$$

其中:

$$\begin{aligned} \theta_1 &= \frac{\alpha_{\rm c}\phi_{\rm m}'^2\psi^2}{2\kappa}, \\ \theta_2 &= \frac{1}{4(\kappa+1)}\phi_{\rm m}'^4\|D_1 - \frac{D_2}{\gamma^2}\|_{\rm m}^2 - \\ &\quad \frac{\alpha_{\rm c}+3\alpha_{\rm c}^2}{16}\phi_{\rm M}'^2\|D_1 - \frac{D_2}{\gamma^2}\|_{\rm M}^2, \end{aligned} \tag{34}$$

$$\theta_3 = 4\alpha_{\rm c} + 3\alpha_{\rm c}^2. \tag{35}$$

则若不等式

$$\tilde{W}_{\rm c} > \sqrt{\frac{\theta_1}{2\theta_2} + \sqrt{\frac{\theta_1^2}{4\theta_2} + \theta_3\|\epsilon_{\rm h}\|^2}} \tag{36}$$

成立, 可保证$\Delta L_{\rm W} < 0$. 综合考虑情况1和情况2, 评价网络权值估计误差$\tilde{W}_{\rm c}$在$t_j < t \leqslant t_{j+1}$时保持不变, 在$t = t_j$时一致有界. 因此, 利用Lypunov理论可得评价网络权值估计误差$\tilde{W}_{\rm c}$一致最终有界.

定理 4 考虑系统(1), 事件触发控制策略和扰动策略为式(15), 评价网络权值调节律为式(18), 自适应事件触发条件为式(10). 假设系统状态满足持续激励条件, 则闭环系统一致最终有界, 并且所获得的控制

策略$u(\bar{x}_j)$和扰动策略$w(\bar{x}_j)$近似收敛到最优控制控制策略$u^*(\bar{x}_j)$和最优扰动策略$w^*(\bar{x}_j)$, 即$\|u-u^*\|\leqslant\varepsilon_u$ $\|w-w^*\|\leqslant\varepsilon_w$.

证　选取如下的Lyapunov函数:

$$L=L_{\rm W}+L_x+L_{\bar{x}_j},\tag{37}$$

其中: $L_{\rm W}=\mathrm{tr}\{\tilde{W}_{\rm c}^{\rm T}\tilde{W}_{\rm c}\}$, $L_x=V(x)$, $L_{\bar{x}_j}=\bar{x}_j^{\rm T}\bar{x}_j$与定理3类似, 分两种情况证明闭环系统的稳定性.

情况 1　当$t_j<t\leqslant t_{j+1}$时: 利用式(8)(18)(22)可得

$$\dot{L}_{\rm W}=2\mathrm{tr}\{\tilde{W}_{\rm c}^{\rm T}\dot{\tilde{W}}_{\rm c}\}=0,\tag{38}$$

$$\dot{L}_{\bar{x}_j}=2\dot{\bar{x}}_j^{\rm T}\bar{x}_j=0,\tag{39}$$

$$\begin{aligned}&\dot{L}_x\leqslant\\&-Q(x)+\nabla_x V(x)(\frac{1}{2}g(x)R^{-1}(g^{\rm T}(x)-g^{\rm T}(\bar{x}_j))\times\\&\nabla_x\phi^{\rm T}(x)W_{\rm c}+\frac{1}{2}g(x)R^{-1}g^{\rm T}(\bar{x}_j)\nabla_x\phi^{\rm T}(x)\tilde{W}_{\rm c}+\\&\frac{1}{2}g(x)R^{-1}g^{\rm T}(\bar{x}_j)(\nabla_x\phi^{\rm T}(x)-\nabla_x\phi^{\rm T}(\bar{x}_j))\hat{W}_{\rm c}-\\&\frac{1}{2}g(x)R^{-1}g^{\rm T}(x)\nabla_x\epsilon_{\rm c}-\\&\frac{1}{2\gamma^2}k(x)P^{-1}(k^{\rm T}(x)-k^{\rm T}(\bar{x}_j))\nabla_x\phi^{\rm T}(x)W_{\rm c}-\\&\frac{1}{2\gamma^2}k(x)P^{-1}k^{\rm T}(\bar{x}_j)\nabla_x\phi^{\rm T}(x))\tilde{W}_{\rm c}-\\&\frac{1}{2\gamma^2}k(x)P^{-1}k^{\rm T}(\bar{x}_j)(\nabla_x\phi^{\rm T}(x)-\nabla_x\phi^{\rm T}(\bar{x}_j)\hat{W}_{\rm c}-\\&\frac{1}{2\gamma^2}k(x)P^{-1}k^{\rm T}(x)\nabla_x\epsilon_{\rm c}).\end{aligned}\tag{40}$$

利用假设1和Young不等式, 可得

$$\dot{V}(x)\leqslant-Q(x)-\beta_1\|e_j\|+\beta_2\|e_j\|^2+\mathcal{B}_1,\tag{41}$$

其中:

$$\begin{aligned}&\beta_1=\\&\frac{1}{2}g_{\rm M}^2\lambda_{\max}(R^{-1})l\phi'_{\rm M}\|\hat{W}_{\rm c}\|^2+\\&\frac{1}{2\gamma^2}k_{\rm M}^2\lambda_{\max}(P^{-1})\phi'_{\rm M}\|\hat{W}_{\rm c}\|^2,\end{aligned}\tag{42}$$

$$\begin{aligned}&\beta_2=\\&\frac{1}{4}g_{\rm M}^2\lambda_{\max}(R^{-1})l\phi'_{\rm M}\|\hat{W}_{\rm c}\|+\\&\frac{1}{4\gamma^2}k_{\rm M}^2\lambda_{\max}(P^{-1})\|\hat{W}_{\rm c}\|,\end{aligned}\tag{43}$$

$$\begin{aligned}&\mathcal{B}_1=\\&\frac{1}{4}g_{\rm M}^2\lambda_{\max}(R^{-1})\phi'_{\rm M}(\|\tilde{W}_{\rm c}\|+W_{\rm cM})\times\\&(\phi'_{\rm M}\|\tilde{W}_{\rm c}\|+\epsilon'_{\rm cM})^2+\\&\frac{1}{4\gamma^2}k_{\rm M}^2\lambda_{\max}(P^{-1})\phi'_{\rm M}(\|\tilde{W}_{\rm c}\|+W_{\rm cM})\times\\&(\phi'_{\rm M}\|\tilde{W}_{\rm c}\|+\epsilon'_{\rm cM})^2+\\&g_{\rm M}^2\lambda_{\max}(R^{-1})\phi'_{\rm M}W_{\rm cM}(\phi'_{\rm cM}W_{\rm 1M}+\epsilon'_{\rm cM})+\\&\frac{1}{2}g_{\rm M}^2\lambda_{\max}(R^{-1})(\phi'_{\rm M}W_{\rm cM}+\epsilon'_{\rm cM})\|\tilde{W}_{\rm c}\|+\\&k_{\rm M}^2\lambda_{\max}(P^{-1})\phi'_{\rm M}W_{\rm cM}(\phi'_{\rm M}W_{\rm cM}+\epsilon'_{\rm cM})+\\&\frac{1}{2}k_{\rm M}^2\lambda_{\max}(P^{-1})(\phi'_{\rm M}W_{\rm cM}+\epsilon'_{\rm cM})\phi'_{\rm M}\|\tilde{W}_{\rm c}\|+\\&\frac{1}{2}k_{\rm M}^2\lambda_{\max}(P^{-1})(\phi'_{\rm M}W_{\rm cM}+\epsilon'_{\rm cM})\epsilon'_{\rm cM}.\end{aligned}\tag{44}$$

由定理3可知$\|\tilde{W}_{\rm c}\|$有界, 则可得$\mathcal{B}_1$有界, 记为$\mathcal{B}_{1\rm M}$. 将事件触发条件(10)代入上式, 可得

$$\dot{V}(x)\leqslant-(1-C)Q(x)+\mathcal{B}_{1\rm M},\tag{45}$$

其中$0<C<1$. 则若不等式

$$x>Q^{-1}(\frac{\mathcal{B}_{1\rm M}}{(1-C)})\triangleq\mathcal{B}_x\tag{46}$$

成立, 可保证$\dot{V}(x)<0$.

情况 2　当$t=t_j$时: 对Lyapunov函数(37)求差分, 令$x^+=\lim\limits_{\varrho\to0}x(t_j+\varrho)$. 注意到当$t=t_j$时$x^+=x$, 则可得

$$\Delta L_x=V(x^+)-V(x)=0,\tag{47}$$

$$\begin{aligned}\Delta L_{\bar{x}_j}&=V(\bar{x}_j^+)-V(\bar{x}_j)=\\&x^{\rm T}x-\bar{x}_j^{\rm T}\bar{x}_j=-\|\bar{x}_j\|^2+\mathcal{B}_x^2.\end{aligned}\tag{48}$$

由式(37)(46)–(48)可得

$$\Delta L=-\|\bar{x}_j\|^2-\theta_2(\|\tilde{W}_{\rm c}\|^2-\frac{\theta_1}{2\theta_2})^2+\mathcal{B}_{2\rm M},\tag{49}$$

其中$\mathcal{B}_{2\rm M}=\dfrac{\theta_1^2}{4\theta_2}+\theta_3\|\epsilon_{\rm h}\|^2+\mathcal{B}_x^2$, 则若不等式

$$\|\bar{x}_j\|>\sqrt{\mathcal{B}_{2\rm M}}\tag{50}$$

和

$$\|\tilde{W}_{\rm c}\|>\sqrt{\frac{\theta_1}{2\theta_2}+\sqrt{\mathcal{B}_{2\rm M}}}\triangleq\mathcal{B}_{\tilde{W}_{\rm c}}\tag{51}$$

成立, 可保证$\Delta L<0$. 综合考虑情况1和情况2, 当$t_j<t<t_{j+1}$时, 评价网权络值估计误差$\tilde{W}_{\rm c}$和状态采样$\bar{x}_j$保持不变, 系统状态x一致最终有界; 当$t=t_j$时, 评价网络权值估计误差$\tilde{W}_{\rm c}$、状态采样$\bar{x}_j$和系统状态x一致最终有界. 因此, 利用Lypunov理论可得系统状态x、状态采样$\bar{x}_j$、评价网络权值误差$\tilde{W}_{\rm c}$一致最终有界. 接下来证明$\|u-u^*\|=\varepsilon_u$和$\|w-w^*\|=\varepsilon_w$. 由式(11)(15), 假设3以及$\tilde{W}_{\rm c}$的有界性可得

$$\begin{aligned}\|u-u^*\|&=\|-\frac{1}{2}R^{-1}g^{\rm T}(\bar{x}_j)\nabla_x\phi^{\rm T}(\bar{x}_j)\tilde{W}_{\rm c}\|\leqslant\\&\frac{1}{2}\lambda_{\max}(R^{-1})g_{\rm M}\phi'_{\rm cM}\mathcal{B}_{\tilde{W}_{\rm c}}=\varepsilon_u.\end{aligned}\tag{52}$$

类似地, 可得

$$\|w-w^*\|=\|\frac{1}{2\gamma^2}P^{-1}k^{\rm T}(\bar{x}_j)\nabla_x\phi^{\rm T}(\bar{x}_j)\tilde{W}_{\rm c}\|\leqslant$$

$$\frac{1}{2\gamma^2}\lambda_{\max}(P^{-1})k_{\mathrm{M}}\phi'_{\mathrm{cM}}\mathcal{B}_{\tilde{W}_{\mathrm{c}}}=\varepsilon_w. \tag{53}$$

证毕.

在事件触发控制中, 发生Zeno行为会极大的消耗系统能量甚至影响系统的控制目标. 接下来, 定理5将证明本文提出的ET–ADP算法能够避免Zeno行为的产生.

定理 5 考虑系统(1), 事件触发控制策略和扰动策略为(15), 评价网络权值调解律为(18), 自适应事件触发条件为(10), 则最小采样间隔时间有下界, 即

$$\tau_{\min}\geqslant\frac{1}{k}\ln(1+\frac{K\xi_{\mathrm{m}}}{\mu})>0, \tag{54}$$

其中:

$$\begin{aligned}&K>0,\\&\xi_{\mathrm{m}}=\min_{j\in N}(\xi(x,\|\hat{W}_{\mathrm{c}}\|))=\\&\qquad\min_{j\in N}\{\frac{CQ_{\min}}{\beta_{1\mathrm{M}}},\sqrt{\frac{2CQ_{\min}}{\beta_{2\mathrm{M}}}}\}>0,\\&\mu=(g_{\mathrm{M}}^2\lambda_{\max}(R^{-1})\phi'_{\mathrm{M}})(W_{\mathrm{cM}}+\\&\qquad\frac{1}{2}\|\tilde{W}_{\mathrm{c}}\|+\|\hat{W}_{\mathrm{c}}\|+\frac{1}{2}\epsilon'_{\mathrm{cM}})+\\&\qquad(\frac{k_{\mathrm{M}}^2}{\gamma^2}\lambda_{\max}(P^{-1})\phi'_{\mathrm{M}})(W_{\mathrm{cM}}+\\&\qquad\frac{1}{2}\|\tilde{W}_{\mathrm{c}}\|+\|\hat{W}_{\mathrm{c}}\|+\frac{1}{2}\epsilon'_{\mathrm{cM}}).\end{aligned}$$

证 当$t_j<t<t_{j+1}$时, 闭环系统动态为式(22). 由于最优控制策略和扰动策略可镇定系统, 因此存在一个常数K满足不等式$\|f(x)+gu^*+kw^*\|\leqslant K\|x\|$. 利用式(22)、假设1及定理4可得, 当$t_j<t<t_{j+1}$时,

$$\|\dot{x}\|\leqslant K\|x\|+\mu. \tag{55}$$

由定理4可知, 评价网络权值估计误差$\tilde{W}_{\mathrm{c}}$一致最终有界, 则有μ有界. 对测量误差$e_j(t)$求导, 可得

$$\|\dot{e}_j\|\leqslant\|\dot{x}\|\leqslant K\|x\|+\mu. \tag{56}$$

利用比较原理求解式(56), 初始条件为$e_j(t)=\bar{x}_j-x(t)=0,\ t=t_j$, 则可得

$$\|e_j\|\leqslant\frac{\mu(\mathrm{e}^{K(t-t_j)}-1)}{K},\ t_j<t\leqslant t_{j+1}. \tag{57}$$

为了得到最小采样间隔时间的下界, 先利用式(10)得到$\xi(x,\|\hat{W}_{\mathrm{c}})\|$的下界$\xi_{\mathrm{m}}$. 当$t_j<t<t_{j+1}$时,

$$\begin{aligned}&\xi_{\mathrm{m}}=\min_{j\in N}(\xi(x,\|\hat{W}_{\mathrm{c}}\|))=\\&\qquad\min_{j\in N}\{\frac{CQ_{\min}}{\beta_{1\mathrm{M}}},\sqrt{\frac{2CQ_{\min}}{\beta_{2\mathrm{M}}}}\}>0,\end{aligned}$$

其中:

$$\begin{aligned}&0<Q_{\min}<Q(x),\\&\beta_{1\mathrm{M}}=\frac{1}{2}g_{\mathrm{M}}^2\lambda_{\max}(R^{-1})l\phi'_{\mathrm{M}}\|\hat{W}_{\mathrm{c,M}}\|^2)+\\&\qquad\frac{1}{2\gamma^2}k_{\mathrm{M}}^2\lambda_{\max}(P^{-1})\phi'_{\mathrm{M}}\|\hat{W}_{\mathrm{c,M}}\|^2,\\&\beta_{2\mathrm{M}}=\frac{1}{4}g_{\mathrm{M}}^2\lambda_{\max}(R^{-1})l\|\hat{W}_{\mathrm{c,M}}\|^2)+\\&\qquad\frac{1}{4\gamma^2}k_{\mathrm{M}}^2\lambda_{\max}(P^{-1})l\|\hat{W}_{\mathrm{c,M}}\|^2,\\&\hat{W}_{\mathrm{c,M}}=\max_{j\in N}(\|\hat{W}_{\mathrm{c}}\|).\end{aligned}$$

当采样间隔最小时, $\|e_{j+1}\|=\xi_{\mathrm{m}}$, 与式(57)比较, 则有$\xi_{\mathrm{m}}\leqslant\frac{\mu\mathrm{e}^{K(t_{j+1}-t_j)-1}}{K}$. 进一步求解可得$t_{j+1}-t_j=\tau_k\geqslant\frac{1}{\kappa}\ln(1+\frac{K\xi_{\mathrm{m}}}{\mu})>0$, 则最小采样间隔满足式(54).

证毕.

5 仿真(Simulation)

1) 考虑如下的F–16战斗机模型[20]:

$$\dot{x}=Ax+B_1u+B_2w, \tag{58}$$

其中:

$$A=\begin{bmatrix}-1.0087&0.90506&-0.00215\\0.82225&-1.07741&-0.17555\\0&0&-1\end{bmatrix},$$

$B_1=[0\ \ 0\ \ 1]^{\mathrm{T}},\ B_2=[1\ \ 0\ \ 0]^{\mathrm{T}}$.

系统状态$x=[a\ \ q\ \ \delta_{\mathrm{e}}]$, a表示攻角, q表示速度, δ_{e}表示升降舵偏角. 控制输入u为致动器电压, 扰动w为作用到攻角上的阵风. 值函数定义为式(2), 其中:

$$Q=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix},\ R=1,\ P=1,\ \gamma=5.$$

评价网络的激活函数设为$[x_1^2\ \ x_1x_2\ \ x_1x_3\ \ x_2^2\ \ x_2x_3\ \ x_3^2]$, 权值记为$\hat{W}_{\mathrm{c}}=[W_{\mathrm{c1}}\ \ W_{\mathrm{c2}}\ \ W_{\mathrm{c3}}\ \ W_{\mathrm{c4}}\ \ W_{\mathrm{c5}}\ \ W_{\mathrm{c6}}]$, 评价网络的自适应增益设为$\alpha_{\mathrm{c}}=0.3$. 事件触发条件中的参数设为$C=0.2,\ g_{\mathrm{M}}=k_{\mathrm{M}}=\phi'_{\mathrm{M}}=3,\ l=1$. 系统状态初值设为$x(0)=[20\ \ 20\ \ 20]^{\mathrm{T}}$. 学习时间设为$t=8000$ s. 评价网络的权值收敛轨迹如图2所示. 将所求得的事件触发最优控制作用到系统(58)上, 可得系统状态轨迹如图3所示, 事件触发条件e_j及$\xi(x,\|\hat{W}_{\mathrm{c}}\|)$的轨迹如图4所示. 累加的采样次数轨迹如图5所示. 由图5可以看出, 本文所提出的ET–ADP算法仅需对状态采样1749次, 而传统的时间触发ADP算法需要对状态采样4×10^4次. 本文提出的ET–ADP算法能够减少97.3%的计算量. 与文献[18–19]相比, 本文所提出的评价网络和控制器只需各更新1749次, 而文献[18–19]提出的方法仅评价网络就更新4×10^4次, 因此, 仿真结果表明本文提出的ET–ADP算法能够有效地减少计算量和降低网络负荷.

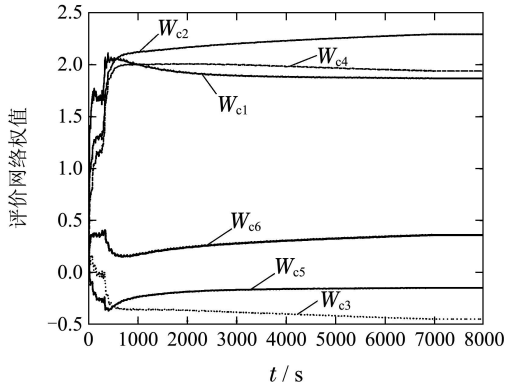

图 2　评价网络的权值收敛轨迹

Fig. 2　The convergent trajectories of critic NN weights

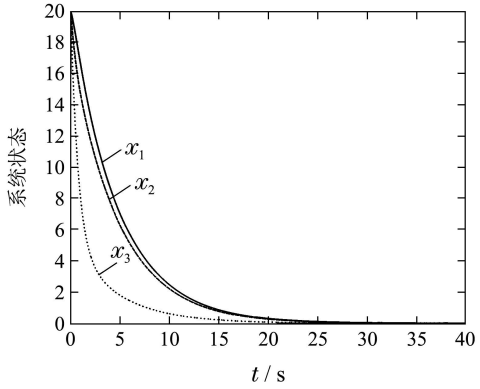

图 3　系统状态轨迹

Fig. 3　The trajectories of system states

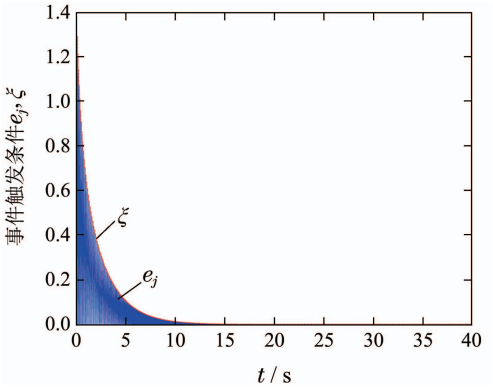

图 4　事件触发条件轨迹

Fig. 4　The trajectory of event-triggered condition

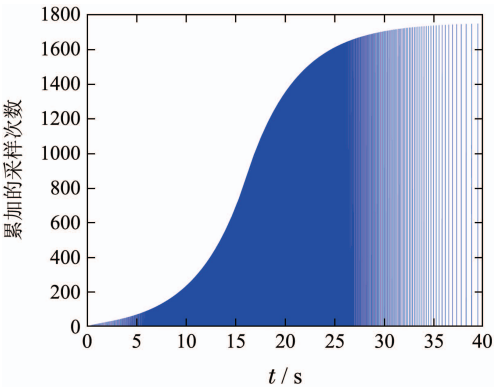

图 5　累加的采样次数

Fig. 5　The cumulative samples

2)　考虑如下的非线性系统:

$$\dot{x} = f(x) + g(x)u + k(x)w, \tag{59}$$

其中: $f(x) = [-x_1+x_2,\ -x_1-x_2+0.25x_2(\cos(2x_1)+2)^2+0.25x_2(\sin(4x_1)+2)^2]^{\mathrm{T}}$, $g(x)=[0\ \ \cos(2x_1)+2]^{\mathrm{T}}$, $k(x) = [0\ \ \sin(4x_1)+2]^{\mathrm{T}}$. 值函数定义为式(2), 其中:

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ R = 1,\ P = 1,\ \gamma = 8.$$

评价网络的激活函数设为$[x_1^2\ \ x_2^2\ \ x_1^4\ \ x_2^4]$, 权值记为$\hat{W}_{\mathrm{c}} = [W_{\mathrm{c1}}\ \ W_{\mathrm{c2}}\ \ W_{\mathrm{c3}}\ \ W_{\mathrm{c4}}]$, 评价网络的自适应增益设为$\alpha_{\mathrm{c}} = 0.3$. 事件触发条件中的参数设为$C = 0.4$, $g_{\mathrm{M}} = k_{\mathrm{M}} = l = \phi'_{\mathrm{M}} = 3$. 系统状态初值设为$x(0) = [4\ \ 5]^{\mathrm{T}}$. 学习时间设为$t = 2\times10^4$ s. 评价网络的权值收敛轨迹如图6所示. 将所求得的事件触发最优控制作用到系统(59)上, 可得系统状态轨迹如图7所示, 事件触发条件e_j及$\xi(x, \|\hat{W}_{\mathrm{c}}\|)$的轨迹如图8所示. 累加的采样次数轨迹如图9所示. 由图9可以看出, 本文所提出的ET–ADP算法仅需对状态采样508次, 而传统的时间触发ADP算法需要对状态采样5×10^3次, 本文提出的ET–ADP算法与时间触发ADP算法相比减少89.9%的计算量. 与文献[18–19]相比, 本文所提出的评价网络和控制器只需各更新508次, 而文献[18–19]提出的方法仅评价网络就更新5×10^3次. 因此, 仿真结果表明本文所提出的的ET–ADP算法的有效性.

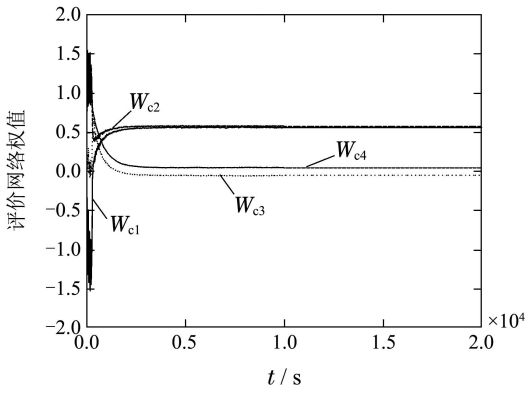

图 6　评价网络的权值收敛轨迹

Fig. 6　The convergent trajectories of critic NN weights

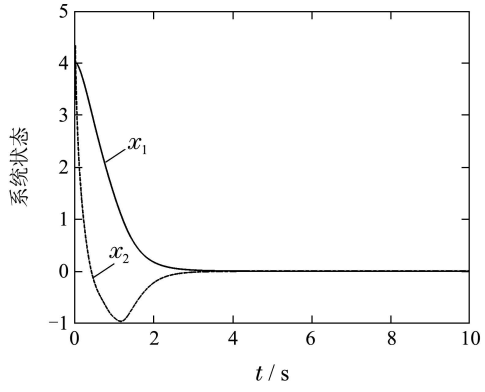

图 7　系统状态轨迹

Fig. 7　The trajectories of system states

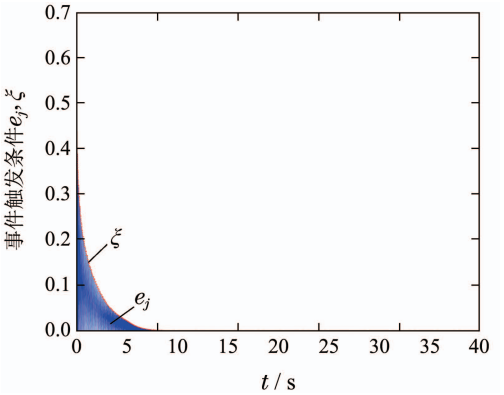

图 8 事件触发条件轨迹

Fig. 8 The trajectory of event-triggered condition

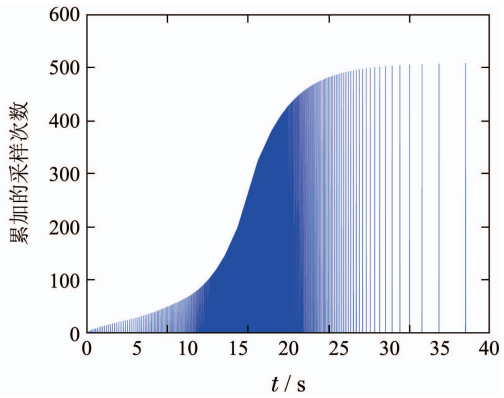

图 9 累加的采样次数

Fig. 9 The cumulative samples

6 结论(Conclusions)

针对一类非线性零和微分对策问题, 本文提出了一种事件触发近似动态规划(ET–ADP)算法. 该算法能够在线学习获得非线性零和微分对策的鞍点, 且仅在事件触发时刻同时更新值函数、控制策略和扰动策略. 通过Lyapunov稳定性理论证明了所提出的算法能够保证闭环系统的一致最终有界性并且不会引起Zeno行为. 最后, 仿真例子验证了所提出的ET–ADP算法的有效性.

参考文献(References):

[1] SHIMA T, GOLAN O M. Linear quadratic differential games guidance law for dual controlled missiles [J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2007, 43(3): 834 – 842.

[2] TAO L, ZHANG J. Asymptotically optimal decentralized control for large population stochastic multiagent systems [J]. *IEEE Transactions on Automatic Control*, 2008, 53(7): 1643 – 1660.

[3] GU D. A differential game approach to formation control [J]. *IEEE Transactions on Control Systems Technology*, 2008, 16(1): 85 – 93.

[4] YEUNG D W, PETROSYAN L A. A cooperative stochastic differential game of transboundary industrial pollution [J]. *Automatica*, 2008, 44(6): 1532 – 1544.

[5] ABU-KHALAF M, LEWIS F L, HUANG J. Neurodynamic programming and zero-sum games for constrained control systems [J]. *IEEE Transactions on Neural Networks*, 2015, 19(7): 1243 – 1252.

[6] ZHANG H G, CUI L L, LUO Y H. Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP [J]. *IEEE Transactions on Cybernetics*, 2013, 43(1): 206 – 216.

[7] LIU D, LI H, WANG D. Online synchronous approximate optimal learning algorithm for multi-player non-zero-sum games with unknown dynamics [J]. *IEEE Transactions on Systems Man & Cybernetics Systems*, 2014, 44(8): 1015 – 1027.

[8] ZHANG H G, WEI Q L, LIU D. An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games [J]. *Automatica*, 2011, 47(1): 207 – 214.

[9] VAMVOUDAKIS K G. Event-triggered optimal adaptive control algorithm for continuous-time nonlinear systems [J]. *IEEE/CAA Journal of Automatica Sinica*, 2015, 1(3): 282 – 293.

[10] SAHOO A, XU H, JAGANNATHAN S. Approximate optimal control of affine nonlinear continuous-time systems using event-sampled neurodynamic programming [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(3): 639 – 652.

[11] ZHANG Q C, ZHAO D B, WANG D. Event-based robust control for uncertain nonlinear systems using adaptive dynamic programming [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, Doi: 10.1109/TNNLS.2016.2614002.

[12] ZHU Y H, ZHAO D B, HE H B, et al. Event-triggered optimal control for partially-unknown constrained-input systems via adaptive dynamic programming [J]. *IEEE Transactions on Industrial Electronics*, 2017, 64(5): 4101 – 4109.

[13] ZHONG X N, HE H B. An event-triggered ADP control approach for continuous-time system with unknown internal states [J]. *IEEE Transactions on Systems, Man & Cybernetics: Systems*, 2017, 47(3): 683 – 694.

[14] WANG D, MU C, ZHANG Q C, et al. Event-driven adaptive robust control of nonlinear systems with uncertainties through NDP strategy [J]. *IEEE Transactions on Systems, Man & Cybernetics: Systems*, 2017, 47(7): 1358 – 1370.

[15] LI Jian, SHEN Yanjun, LIU Yungang. Adaptive dynamic programming algorithms for the output tracking of Buck converter systems [J]. *Control Theory & Applications*, 2017, 34(3): 393 – 400.
(李健, 沈艳军, 刘允刚. Buck型变换器输出跟踪的自适应动态规划算法 [J]. 控制理论与应用, 2017, 34(3): 393 – 400.)

[16] LI Xiaoli, LIU Dexin, JIA Chao, et al. Multiple set-points tracking control method based on adaptive dynamic programming [J]. *Control Theory & Applications*, 2013, 30(6): 709 – 716.
(李晓理, 刘德馨, 贾超, 等. 基于自适应动态规划的多设定值跟踪控制方法 [J]. 控制理论与应用, 2013, 30(6): 709 – 716.)

[17] ZHONG X N, HE H B, WANG D, et al. Model-free adaptive control for unknown nonlinear zero-sum differential game [J]. *IEEE Transactions on Cybernetics*, 2017, Doi: 10.1109/TCYB.2017.2712617.

[18] WANG D, MU C, ZHANG Q C, et al. Event-based input-constrained nonlinear H_∞ state feedback with adaptive critic and neural implementation [J]. *Neurocomputing*, 2016, 214(C): 848 – 856.

[19] ZHANG Q C, ZHAO D B, ZHU Y H. Event-triggered H_∞ control for continuous-time nonlinear system via concurrent learning [J]. *IEEE Transactions on Systems, Man & Cybernetics: Systems*, 2017, 47(7): 1071 – 1081.

[20] WU H N, LUO B, Neural Network Based online simultaneous policy update algorithm for solving the HJI equation in nonlinear H_∞ control [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, 23(13): 1884 – 1895.

作者简介:

崔黎黎 (1983–), 女, 博士, 讲师, 研究方向为自适应动态规划、非线性系统、微分对策、最优控制等, E-mail: cuilili8396@163.com;

张 勇 (1978–), 男, 硕士, 副教授, 研究方向为计算机控制、嵌入式系统等, E-mail: yzhang.sy@qq.com;

张 欣 (1982–), 女, 博士, 讲师, 研究方向为自适应动态规划、微分对策、神经网络自适应控制等, E-mail: zhangxin@upc.edu.cn.