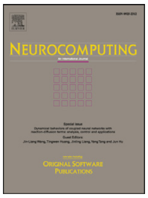
Neurocomputing 285 (2018) 51–59

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



Value iteration based integral reinforcement learning approach for H_∞ controller design of continuous-time nonlinear systems

Geyang Xiao[a,b], Huaguang Zhang[a,b,*], Kun Zhang[a,b], Yinlei Wen[a,b]

[a] College of Information Science and Engineering, Northeastern University, Box 134, 110819 Shenyang, PR China
[b] The Key Laboratory of Integrated Automation of Process Industry, National Education Ministry, Northeastern University, 110004 Shenyang, PR China

ARTICLE INFO

Article history:
Received 9 March 2017
Revised 2 January 2018
Accepted 11 January 2018

Communicated by Dr Bin xu

Keywords:
Value iteration
H_∞ control
Reinforcement learning
Continuous-time systems

ABSTRACT

In this paper, a novel integral reinforcement learning approach is developed based on value iteration (VI) for designing the H_∞ controller of continuous-time (CT) nonlinear systems. First, the VI learning mechanism is introduced to solve the zero-sum game problems, which is equivalent to the Hamilton–Jacobi–Isaacs (HJI) equation arising in H_∞ control problems. Since the proposed method is based on VI learning mechanism, it does not require the admissible control for the implementation, and thus satisfies a more general initial condition than the works based on policy iteration (PI). The iterative property of the value function is analysed with an arbitrary initial positive function, and the H_∞ controller can be derived as the iteration converges. For the implementation of the proposed method, three neural networks are introduced to approximate the iterative value function, the iterative control policy and the iterative disturbance policy, respectively. To verify the effectiveness of the VI based method, a linear case and a nonlinear case are presented, respectively.

1. Introduction

In various industrial applications, disturbance exists in many situations and always influences the controlled systems negatively. To handle this control problem, H_∞ control has been widely investigated and becomes an essential part of robust control. The goal of H_∞ control is to find a feedback controller for a given system while considering the robustness and control performance. In the early years, the H_∞ control problem was studied for the linear systems [1,2]. Later, some researchers [3–8] well developed the H_∞ control theory arising in the nonlinear systems. The work of [6] indicated that the H_∞ control problem could be equivalent to a two-player zero-sum differential game. The Nash equilibrium solution of the game could be solved by a equation called Hamilton–Jacobi–Isaacs (HJI), which is a nonlinear partial differential equation (PDE). For the linear case, the HJI equation reduces to a Riccati equation which can be efficiently solved. However, for the nonlinear case, there is still no approach to solve the HJI equation analytically. This has inspired researchers to study approaches for solving the HJI equation approximately, and some direct approaches have been proposed in early period [4,9]. Unfortunately, the proposed direct approaches were restricted by computational load. In recent years, some researchers developed an indirect approach to approximate the solution of HJI equation by introducing reinforcement learning (RL) technique.

Over the last several decades, RL has been widely studied [10–13], which attempts to imitate the natural law of learning in mammals. The concept of RL is learning how to map situations to actions, so as to maximize a numerical reward signal [12]. Unlike most forms in machine learning, the learner is not told which actions to take, but instead discover which actions can result in the most wanted reward by trying them. Actually, according to the RL technique, actions may affect not only the immediate reward but also the next situation and, through that, all subsequent rewards. Because of this important distinguishing feature, some researchers [14–16] introduced the idea of RL into solving the optimal problem arising in nonlinear control, and proposed an actor-critic structure to solve a nonlinear PDE called Hamilton–Jacobi–Bellman (HJB) equation approximately to derive the solution. This RL-based technique is named as approximate dynamic programming, or adaptive dynamic programming (ADP). Since the HJI equation is also a nonlinear PDE, much attention have been attracted to introduce this RL-based technique to seek for the solution of HJI [17–32]. Generally, there are two typical way in the ADP framework to solve for the PDE, the policy iteration (PI) and the value iteration (VI) [14].

[*] Corresponding author at: The Key Laboratory of Integrated Automation of Process Industry of the National Education Ministry, Northeastern University, 110004 Shenyang, PR China.

E-mail addresses: xgyalan@outlook.com (G. Xiao), hgzhang@ieee.org, zhg516516@gmail.com (H. Zhang), nukgnahz@163.com (K. Zhang), wenyl.d.h@hotmail.com (Y. Wen).

https://doi.org/10.1016/j.neucom.2018.01.029
0925-2312/© 2018 Elsevier B.V. All rights reserved.

For the H_∞ control problem arising in the continuous-time (CT) nonlinear systems, various works have been studied based on PI method. One feature of the PI is that it requires to solve a value function associated with an admissible control policy in the policy evaluation step [33,34]. In [17,20], the authors proved that the HJI equation can be solved by using PI, and the iterative convergence to the available storage function associated with a given L_2-gain was proposed. In [18,19], the H_∞ control problem with finite-horizon was studied by using PI. In [21,26], a developed PI based method was proposed which can deal with the systems with unknown drift dynamics and be implemented in an on-line manner. In [22], a PI based method was proposed to shown that the mixed optimum of the zero-sum game can be derived even the saddle point solution does not exist. In [23,24], the authors attempted to design PI based algorithms to seek for the solution of the HJI equation by using only one neural network. In [25], the authors developed a PI based integral reinforcement learning algorithm [34] for the H_∞ control of unknown CT linear systems. In [28], a novel PI based technique called off-policy was introduced to solve the HJI equation and arbitrary policies can be applied to generate the system data to tune the algorithm rather than the evaluating policy. The authors of [29,31] developed the off-policy technique to design the H_∞ controller for unknown CT nonlinear systems. Although various well developed methods were proposed for the H_∞ controller design of CT nonlinear systems, all of them were based on PI, and thus the initial admissible control is assumed [33]. From a mathematical point of view, an admissible control can be regarded as a suboptimal control which requires to solve the nonlinear partial differential equations analytically. Thus, to ensure the admissibility may be a serious restrictive condition actually. To the best of our knowledge, there is still no approach to obtain such a control, especially for the nonlinear systems with the existence of disturbance.

On the other hand, the learning mechanism of VI ensures more free in the initial condition than PI, where the admissible control assumption is not required [35–41]. In [35], the convergence of VI method was proved with an initial zero value function for the optimal control arising in the discrete-time (DT) nonlinear systems. In [39], the authors discussed the convergence of VI in a more general way for the optimal control problem of DT nonlinear systems, where the algorithm can be initialized with an arbitrary positive value function. Since the benefits of initial condition, some researchers introduced VI to solve the H_∞ control problem arising in DT systems [42,43]. In [42], the authors introduced the VI learning mechanism into the Q-learning method for the H_∞ control problem of DT linear systems. In [43], the authors developed a VI based algorithm to seek for the solution of the zero-sum game for DT nonlinear systems, which is equal to the solution of the HJI equation associated with H_∞ control problems. However, the above works were proposed for the DT nonlinear systems. The discussions on solving the H_∞ control problem by VI method for CT nonlinear systems are scarce, which motives our research.

In this paper, a novel VI based integral reinforcement learning method is proposed to design the H_∞ controller for CT nonlinear systems. First, the algorithm is proposed by introducing the VI learning mechanism into the integral reinforcement learning to solve the HJI equation arising in H_∞ control problems for CT nonlinear systems. Since the proposed method is based on VI learning mechanism, it satisfies a more general initial condition than the works based on PI which requires an initial admissible control for implementation. The iterative property of the value function is analysed with an arbitrary initial positive function, and the H_∞ controller can be derived as the iteration converges. For the implementation of the proposed method, three neural networks are introduced to approximate the iterative value function, the iterative control policy and the iterative disturbance policy, respectively. At last, two simulation cases are presented to illustrate the effectiveness of the proposed method.

2. Problem statement

Consider the CT nonlinear system described as

$$\begin{aligned} \dot{x} &= f(x) + g(x)u + p(x)d \\ y &= z(x), \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$ is the system state vector, $u \in \mathbb{R}^m$ is the control input, $d \in \mathbb{R}^p$ is the external disturbance and $y \in \mathbb{R}^q$ is the output. The dynamics of the system $f(x) \in \mathbb{R}^n$, $g(x) \in \mathbb{R}^{n\times m}$ and $p(x) \in \mathbb{R}^{n\times p}$ are Lipschitz continuous on a set $\Omega \subseteq \mathbb{R}^n$ and satisfy $f(0) = 0$. The output dynamic satisfies the zero-state observability.

The control objective of H_∞ controller design is to seek for a control policy $u(x)$ to ensure the asymptotically stability of system (1), and satisfying the following L_2-gain condition with a prescribed level γ

$$\int_0^\infty (y^{\mathrm{T}}y + \|u\|_R^2)\mathrm{d}\tau \le \gamma^2 \int_0^\infty \|d\|^2 \mathrm{d}\tau, \tag{2}$$

where $\|u\|_R^2 = u^{\mathrm{T}}Ru$ with $R > 0$.

The performance index is defined as

$$\begin{aligned} J(x_0, u, d) &= \int_0^\infty (y^{\mathrm{T}}y + \|u\|_R^2 - \gamma^2\|d\|^2)\mathrm{d}\tau \\ &= \int_0^\infty U(x, u, d)\mathrm{d}\tau \le 0 \end{aligned} \tag{3}$$

for all $d \in L_2[0, \infty)$ and $x(0) = 0$, where $U(x, u, d) = y^{\mathrm{T}}(x)y(x) + \|u\|_R^2 - \gamma^2\|d\|^2$ denotes the utility function.

For fixed control and disturbance policies $u(x)$ and $d(x)$, define the value function as

$$V(x(t), u, d) = \int_t^\infty U(x, u, d)\mathrm{d}\tau. \tag{4}$$

Differentiating the above value function, we can derive

$$U(x, u, d) + \nabla V^{\mathrm{T}}(x)(f(x) + g(x)u + p(x)d) = 0, V(0) = 0, \tag{5}$$

where $\nabla V = \partial V/\partial x$.

Define the Hamiltonian function for the value function (4) with associated control policy u and disturbance policy d as

$$H(x, u, d, \nabla V) \triangleq U(x, u, d) + \nabla V^{\mathrm{T}}(x)(f(x) + g(x)u + p(x)d). \tag{6}$$

According to the game theory [6], the control problem can be referred to a two-player zero-sum differential game:

$$V^*(x_0) = \min_u \max_d J(x_0, u, d), \tag{7}$$

which exists the unique solution of the saddle point if the Nash condition holds

$$V^*(x_0) = \min_u \max_d J(x_0, u, d) = \max_d \min_u J(x_0, u, d). \tag{8}$$

Suppose the $V^*(x)$ is continuous differentiable, based on the Bellman principle, we have

$$\min_u \max_d [H(x, u, d, \nabla V^*)] = 0, \tag{9}$$

and then

$$u^* = -\frac{1}{2}R^{-1}g^{\mathrm{T}}(x)\nabla V^*(x), \tag{10}$$

$$d^* = \frac{1}{2\gamma^2}p^{\mathrm{T}}(x)\nabla V^*(x). \tag{11}$$

Substituting the control policy (10) and disturbance policy (11) into (5), we can derive the following HJI equation: