Acta Automatica Sinica
Vol. 35, No. 6 June, 2009
Available online at www.sciencedirect.com
ScienceDirect

Data-based Optimal Control for Discrete-time Zero-sum Games of 2-D Systems Using Adaptive Critic Designs

WEI Qing-Lai[1] ZHANG Hua-Guang[2] CUI Li-Li[2]

Abstract In this paper, an iterative adaptive critic design (ACD) algorithm is proposed to solve a class of discrete-time two-person zero-sum games for Roesser type 2-D system. The idea is to use adaptive critic technique to obtain the optimal control pair iteratively to make the performance index function reach the saddle point of the zero-sum games. The proposed iterative ACD algorithm can be implemented based on the input and state data without the system model. Stability analysis of the 2-D system is presented and the convergence property of the performance index function is also proved. Neural networks are used to approximate the performance index function and compute the optimal control policies, respectively, for facilitating the implementation of the iterative ACD algorithm. The optimal control scheme of the air drying process is given to illustrate the performance of the proposed method.

Key words Adaptive critic designs (ACD), optimal control, zero-sum game, 2-D system, neural networks

A large class of complicated practical systems are controlled by more than one controller or decision maker with each using an individual strategy. These controllers often operate in a group with a general performance index function as a game[1]. Zero-sum game theory has been widely applied to decision making and control engineering problems[2−5]. In these situations, many control schemes are presented in order to reach some form of optimality[6−7]. In [8], zero-sum game was proposed to solve multiuser optimal flow control. In [9], the zero-sum game problem was discussed for noncooperative decision makers. Based on the zero-sum theory, the designs of controller in the worst case and the design of H_∞ controller were proposed in [10 − 12]. However, aforementioned results on zero-sum game are only for the one-dimensional systems. In the real world, many complicated control systems are described by 2-dimensional (2-D) structures[13−14]. The key feature of a 2-D system is that the information is propagated along two independent directions. Many physical processes, such as thermal processes, image processing, signal filtering, etc., have a clear 2-D structure. The 2-D system theory is frequently used as an analysis tool to solve some problems, e.g., iterative learning control[15] and repetitive process control[16]. So many control schemes are presented for 2-D system in order to obtain the optimal performance[17−18], while there are few results on the zero-sum games for 2-D systems. The great difficulty of the zero-sum games for 2-D systems is that the optimal recurrent equation, so called Hamilton-Jacobi-Isaacs (HJI) equation, is invalid in 2-D structure, which means that the optimal control pair cannot be obtained by the classical dynamic programming theory. Another difficulty lies in the fact that for many 2-D systems the model of the system cannot be obtained inherently. So it is important and necessary to give a new method to solve the zero-sum games for 2-D system without a system model. This motivates our research.

The adaptive critic designs (ACDs) are very useful tools in solving the optimal control problems and have received considerable attention for the past three decades[19−22]. ACDs were firstly proposed in [23 − 25] as a way to solve optimal control problems forward-in-time. ACDs combine reinforcement learning technique and dynamic programming theory with neural networks. In [13], the ACDs were classified into four main schemes: heuristic dynamic programming (HDP), dual heuristic dynamic programming (DHP), action dependent heuristic dynamic programming (ADHDP), also known as Q-learning[23], and action dependent dual heuristic dynamic programming (ADDHP). In [26], another two ACD schemes known as globalized-DHP (GDHP) and ADGDHP were developed. Though in recent years, ACDs have been further studied by many researchers such as [27 − 35], wherein most results focus on the optimal control problem with a single controller. Only in [36], based on HJI equation, zero-sum game was discussed for 1-D system. To the best of our knowledge, there are no results discussing how to solve the zero-sum game problem for 2-D systems.

In brief, it is the first time for the zero-sum game to solve for a 2-D system by ACD technique. The main contributions of this paper include:

1) Propose a new optimality principle for Roesser type 2-D system and obtain the optimal control formulation in theory.

2) Propose an iterative algorithm based on ACD technique (iterative ACD algorithm for brief) to obtain the optimal control pair iteratively with rigorous stability and convergence analysis.

3) Develop the iterative ACD algorithm into data-driven situation. What is needed to know is only the input and state data, and the model of the system is not required.

This paper is organized as follows. Section 1 presents the preliminaries and assumptions. In Section 2, the optimal control for zero-sum games for 2-D systems is proposed and the properties of the optimal control are also discussed. In Section 3, data-based iterative ACD algorithm is proposed with the convergence analysis. In Section 4, the neural network implementation for the control scheme is discussed. In Section 5, an example is given to demonstrate the effectiveness of the proposed control scheme. The conclusion is drawn in Section 6.

1 Preliminaries and assumptions

Basically, we consider the following discrete-time linear Roesser type 2-D system

$$\boldsymbol{x}^+(k,l) = A\boldsymbol{x}(k,l) + B\boldsymbol{u}(k,l) + C\boldsymbol{w}(k,l) \tag{1}$$

Received December 16, 2008; in revised form March 9, 2009
Supported by National High Technology Research and Development Program of China (863 Program) (2006AA04Z183), National Natural Science Foundation of China (60621001, 60534010, 60572070, 60774048, 60728307), Program for Changjiang Scholars and Innovative Research Groups of China (60728307, 4031002)
1. The Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P. R. China 2. The School of Information Science and Engineering, Northeastern University, Shenyang 110004, P. R. China
DOI: 10.1016/S1874-1029(08)60091-9

$$\boldsymbol{x}^{\rm h}(0,l)=\boldsymbol{f}(l),\quad \boldsymbol{x}^{\rm v}(k,0)=\boldsymbol{g}(k) \tag{2}$$

with

$$\boldsymbol{x}(k,l)=\begin{bmatrix}\boldsymbol{x}^{\rm h}(k,l)\\ \boldsymbol{x}^{\rm v}(k,l)\end{bmatrix},\quad \boldsymbol{x}^{+}(k,l)=\begin{bmatrix}\boldsymbol{x}^{\rm h}(k+1,l)\\ \boldsymbol{x}^{\rm v}(k,l+1)\end{bmatrix} \tag{3}$$

where $\boldsymbol{x}^{\rm h}(k,l)$ is the horizon state in $\mathbf{R}^{n_1}$, $\boldsymbol{x}^{\rm v}(k,l)$ is the vertical state in $\mathbf{R}^{n_2}$, $\boldsymbol{u}(k,l)$ and $\boldsymbol{w}(k,l)$ are the control inputs in $\mathbf{R}^{m_1}$ and $\mathbf{R}^{m_2}$. Let the system matrices $A\in\mathbf{R}^{(n_1+n_2)\times(n_1+n_2)}$, $B\in\mathbf{R}^{(n_1+n_2)\times n_1}$, and $C\in\mathbf{R}^{(n_1+n_2)\times m_2}$. Assume all the system matrices are nonsingular and the system matrices can be expressed by

$$A=\begin{bmatrix}A_1\ A_2\\ A_3\ A_4\end{bmatrix},\ B=\begin{bmatrix}B_1\\ B_2\end{bmatrix},\ C=\begin{bmatrix}C_1\\ C_2\end{bmatrix} \tag{4}$$

The function $\boldsymbol{f}(l)$ and $\boldsymbol{g}(k)$ are corresponding boundary conditions along two independent indirections.

We define the following denotements

$$\begin{aligned}&(k,l)\le(m,n) && \text{if and only if } k\le m \text{ and } l\le n\\ &(k,l)=(m,n) && \text{if and only if } k=m \text{ and } l=n\\ &(k,l)<(m,n) && \text{if and only if } (k,l)\le(m,n) \text{ and}\\ & && (k,l)\ne(m,n)\end{aligned} \tag{5}$$

Then, the infinite-time performance index function for 2-D systems can be given by

$$\begin{aligned}V(\boldsymbol{x}(0,0),\boldsymbol{u},\boldsymbol{w})=&\sum_{(0,0)\le(k,l)<(\infty,\infty)}\sum\Big(\boldsymbol{x}^{\rm T}(k,l)Q\boldsymbol{x}(k,l)+\\ &\boldsymbol{u}^{\rm T}(k,l)R\boldsymbol{u}(k,l)+\boldsymbol{w}^{\rm T}(k,l)S\boldsymbol{w}(k,l)\Big)\end{aligned} \tag{6}$$

where $Q\ge0$, $R>0$, and $S<0$ are with suitable dimensions and $L(\boldsymbol{x}(k,l),\boldsymbol{u}(k,l))=\boldsymbol{x}^{\rm T}(k,l)Q\boldsymbol{x}(k,l)+\boldsymbol{u}^{\rm T}(k,l)R\boldsymbol{u}(k,l)+\boldsymbol{w}^{\rm T}(k,l)S\boldsymbol{w}(k,l)$ is the utility function. For the above zero-sum game, the two control variables $\boldsymbol{u}$ and $\boldsymbol{w}$ are chosen, respectively, by player I and player II where player I tries to minimize the performance index function $V(\boldsymbol{x})$, while player II attempts to maximize it. The following assumptions are proposed that are in effect in the remaining sections.

Assumption 1. The 2-D system (1) is controllable under the control variables $\boldsymbol{u}$ and $\boldsymbol{w}$.

Assumption 2. For the boundary conditions for the 2-D system (1), the terms $\sum_{k=0}^{\infty}\boldsymbol{x}^{\rm vT}(k,0)\,\boldsymbol{x}^{\rm v}(k,0)$, $\sum_{l=0}^{\infty}\boldsymbol{x}^{\rm hT}(0,l)\,\boldsymbol{x}^{\rm h}(0,l)$, and $\sum_{(0,0)\le(k,l)<(\infty,\infty)}\sum\boldsymbol{x}^{\rm vT}(k,0)\times\boldsymbol{x}^{\rm h}(0,l)$ are all bounded.

Assumption 3. There exists an unique saddle point of the zero-sum game for the 2-D system (1).

There are some important characters that must be pointed out. Firstly, for the 1-D control system, the boundary condition is just an initial point of state, while the boundary conditions of 2-D system are two given state curves along two different directions. Secondly, for the zero-sum games of 2-D system under the infinite time horizon, the boundary state trajectories are uncontrollable and so the terms $\sum_{l=0}^{\infty}\boldsymbol{x}^{\rm T}(k,0)Q\boldsymbol{x}(k,0)$, $\sum_{l=0}^{\infty}\boldsymbol{x}^{\rm T}(0,l)Q\boldsymbol{x}(0,l)$, and $\sum_{(0,0)\le(k,l)<(\infty,\infty)}\sum\boldsymbol{x}^{\rm T}(k,0)Q\boldsymbol{x}(0,l)$ may be infinite, which means the performance index function (6) is infinite. Therefore, Assumption 2 is necessary. Thirdly, the boundary conditions $\boldsymbol{f}(l)$ and $\boldsymbol{g}(k)$ in (2) should be boundary, but not necessary smooth or continuous functions. For example, let

$$\boldsymbol{f}(l)=\begin{cases}\boldsymbol{c}, & l\le T\\ \boldsymbol{0}, & l>T\end{cases} \tag{7}$$

where c is any real constant number and T is a given real number. So, Assumption 2 is not very strong.

According to Assumption 3, the optimal performance index function can be expressed as

$$\begin{aligned}V^*(\boldsymbol{x}(k,l))=&\min_{\boldsymbol{u}}\max_{\boldsymbol{w}}\sum_{(k,l)\le(i,j)<(\infty,\infty)}\sum\Big(\boldsymbol{x}^{\rm T}(i,j)Q\boldsymbol{x}(i,j)+\\ &\boldsymbol{u}^{\rm T}(i,j)R\boldsymbol{u}(i,j)+\boldsymbol{w}^{\rm T}(i,j)S\boldsymbol{w}(i,j)\Big)=\\ &\max_{\boldsymbol{w}}\min_{\boldsymbol{u}}\sum_{(k,l)\le(i,j)<(\infty,\infty)}\sum\Big(\boldsymbol{x}^{\rm T}(i,j)Q\boldsymbol{x}(i,j)+\\ &\boldsymbol{u}^{\rm T}(i,j)R\boldsymbol{u}(i,j)+\boldsymbol{w}^{\rm T}(i,j)S\boldsymbol{w}(i,j)\Big)\end{aligned} \tag{8}$$

2 The optimal control for the zero-sum games for 2-D systems

For zero-sum games for 1-D systems, the optimal performance index function can be written by a recurrent formulation according to the dynamic programming principle[36]. However, for zero-sum games for 2-D systems, the dynamic programming principle may not be true. The main difficulty lies in the state of the 2-D system in the next stage coupling with the states of two different directions in the current stage and then the dynamic programming equation of the zero-sum games for 2-D systems does not exist. So in this paper, we propose an optimality principle for 2-D system and obtain the expressions of optimal control pair for the zero-sum game.

2.1 The optimality principle for zero-sum games for 2-D systems

In this subsection, we will propose the optimality principle for the zero-sum games for 2-D systems, and discuss the properties of the optimal control pair derived by the principle.

Theorem 1. Given the performance index function defined as (6), if $\boldsymbol{u}(k,l)$ minimizes and $\boldsymbol{w}(k,l)$ maximizes the performance index function (6), respectively, subject to the system equation (1), and then there are $(n+m)$-dimensional vector sequences $\boldsymbol{\lambda}(k,l)$ and $\boldsymbol{\lambda}^{+}(k,l)$ defined as

$$\boldsymbol{\lambda}^{+}(k,l)=\begin{bmatrix}\boldsymbol{\lambda}^{\rm h}(k+1,l)\\ \boldsymbol{\lambda}^{\rm v}(k,l+1)\end{bmatrix},\quad \boldsymbol{\lambda}(k,l)=\begin{bmatrix}\boldsymbol{\lambda}^{\rm h}(k,l)\\ \boldsymbol{\lambda}^{\rm v}(k,l)\end{bmatrix} \tag{9}$$

where $\boldsymbol{\lambda}^{\rm h}(k,l)\in\mathbf{R}^{n_1}$ and $\boldsymbol{\lambda}^{\rm v}(k,l)\in\mathbf{R}^{n_2}$, such that for all $(0,0)\le(k,l)<(\infty,\infty)$

1) State equation:

$$\boldsymbol{x}^{+}(k,l)=\frac{\partial H(k,l)}{\partial\boldsymbol{\lambda}^{+}(k,l)} \tag{10}$$

2) Costate equation:

$$\boldsymbol{\lambda}(k,l)=\frac{\partial H(k,l)}{\partial\boldsymbol{x}(k,l)} \tag{11}$$