IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 16, 2019, accepted December 16, 2019, date of publication December 26, 2019, date of current version January 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962583

H_∞ Control for a Class of Discrete-Time Systems via Data-Based Policy Iteration With Application to Wind Turbine Control

YANG LIU, (Member, IEEE), ZUOXIA XING, LEI CHEN, JIAN XU, YUNLU LI, AND HAIXIN WANG, (Member, IEEE)

School of Electrical Engineering, Shenyang University of Technology, Shenyang 110870, China

Corresponding author: Zuoxia Xing (xingzuox@163.com)

This work was supported in part by the China Postdoctoral Science Foundation under Grant 2019M651144.

ABSTRACT This paper concerns a novel data-based policy iteration control method for discrete-time systems with disturbances. To avoid the need of system dynamic and state, a new system equation is constructed by using system operation data. Based on this, we derive a new data-based Bellman equation that can be applied to the design of model-free policy iteration algorithm. For discrete-time system with disturbances, we transform H_∞ control into a zero-sum game problem, and then use the proposed new data-based PI method to solve it. Wind turbine control is a typical control problem with time delay and disturbance. Mathematical simulation results are given to demonstrate the effectiveness of the proposed data-based PI control method.

INDEX TERMS Policy iteration, data-driven, discrete-time system, H_∞ control.

I. INTRODUCTION

Discrete system control problem [1]–[3] involves many key fields such as logistics system, event-driven system and impulse control. The corresponding control methods usually include optimal control [4], tracking control [5], [6] and optimization control, and so on. For the linear discrete-time (DT) system, we usually assumed that the performance index function of the controlled system is a quadratic form of system states, and then the optimal controller is designed using a linear quadratic regulator method. However, in practical applications, there are usually disturbances in the controlled system, such as external disturbances in signal transmission and wind turbulence in wind power generation. Therefore, the design of a H_∞ controller suitable for DT system with disturbances is of great research significance and application value.

Adaptive dynamic programming (ADP) [7]–[13] is an adaptive learning algorithm evolved from dynamic programming. Its purpose is to solve the curse of dimensionality problem encountered by dynamic programming by means of iterative learning. As an important branch of ADP structure, policy iteration (PI) algorithm has attracted extensive attention and achieved rapid development. The PI-based control method has been applied to many fields and proved to be an efficient control method [14]–[16]. By combining neural network and behrman equation, an PI-based optimal control method is proposed in [14], which is an iterative algorithm capable of dealing with general affine pseudo-linearized nonlinear systems. Luo *et al.* [15] proposed a QoS prediction method based on ADP, in which ADP method was used to learn the parameters of fuzzy rules and effectively improve the prediction performance. For the continuous system with uncertainties, in [16], an observer-based control strategy was designed by using HJB equation to solve the optimal control policies. In this method, the state observer was used to replace the state of the system, and NNs were adopted to determine the approximate control strategy.

For the H_∞ control [17]–[21] of DT system, PI algorithms also has some preliminary research work [22]–[24]. In [22], an event-driven PI algorithm was studied to deal with the H_∞ control strategy. In [24], by transforming the H_∞control into a zero sum game problem, the optimal control strategy with perturbation was obtained in [24], and the proof of convergence has also been given.

However, these existing PI-based control methods usually require the system dynamic information to be known. In practical applications, it is unrealistic to obtain all the dynamic

The associate editor coordinating the review of this manuscript and approving it for publication was Bohui Wang.

information, which leads to a decrease in the performance of the control algorithm. Therefore, a class of model-free H_∞ control methods based on PI has been studied [25]–[28]. A PI-based H_∞ optimal control method was designed by using generalized fuzzy hyperbolic model in [25]. For H_∞ control problem of discrete nonlinear systems, [26] presented a tracking control method by using PI algorithm combined with neural network with asymptotic stability of estimated error. In [27], the PI structure was used to design a H_∞ control strategy of dynamic unknown discrete system in which the algorithm structure is composed of three NNs. In [28], an ADP algorithm with three neural networks was presented to find the zero-sum game solution of DT systems online, the explicit update rules of the three neural networks were updated by using the system operation data.

Moreover, these methods are usually based on state feedback control methods, so all the states of the system need to be used. However, in many controlled systems, all system states are often unmeasurable, which is a challenge for designing controllers. Therefore, data-based control methods [29]–[35] have become a hot issue. Using the data generated during the operation of the system to assist the controller design is a realistic development direction in engineering applications. Therefore, the ADP algorithm combined with data can effectively improve the control performance of the ADP-based controller in practical applications [36]–[39]. In [36], a H_∞ control was designed for the continuous time system with disturbances, and the system operation data was adopted to find the zero-sum optimal solution. In [37], a PI-based structure was designed by using data-driven method to solve the multivariable tracking control policy. For optimal control of DT system, a policy gradient adaptive dynamic programming control method was presented in [38] using system operation data. In [39], a data-based PI structure was designed for the linear systems. In this method, a traditional value iterative algorithm was designed by using stochastic approximation theory.

Although the control design based on the ADP algorithm has the above research work. However, how to use the data-based PI structure to design H_∞ control of DT system with disturbances has not been studied. Therefore, in this paper, we design the H_∞ controller with the aid of data-based PI, which provides a discrete system H_∞ control scheme and extends the application of the PI-based control method. The innovations of this paper are as follows:

(1) Based on the system measured data, a novel PI-based H_∞ control of DT system is designed, which is an online model-free algorithm without the need of accurate model of system dynamics, which is of great significance in practical application.

(2) A new data-based system equation consists of system input and output sequences is constructed that effectively avoids the requirement of system state, which extends the application scope of the PI-based control method.

(3) For the H_∞ control of DT system, the corresponding new data-based Bellman equation is proposed to facilitate the design of PI algorithm.

This paper is organized as follows. In Section II, the problem formulation is presented. The data-based H_∞ control is derived in Section III. The PI-based H_∞ control method is proposed in Section IV. Simulation results of DT system with disturbances and wind turbine control are presented in Section V. Conclusion is given in Section VI.

II. PROBLEM FORMULATION

Consider the following DT system

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + D\omega_k \\ y_k &= Cx_k, \end{aligned} \tag{1}$$

with the system state $x_k \in \mathbb{R}^n$, system output $y_k \in \mathbb{R}^p$, control input $u_k \in \mathbb{R}^m$ and external disturbance $\omega_k \in \mathbb{R}^q$. And assumed that the controlled system is controllable and observable.

The performance index function of the corresponding H_∞ control problem is selected as

$$\begin{aligned} J_k(x_k, u_k, \omega_k) &= \min_{u_i} \max_{\omega_i} \sum_{i=k}^{\infty} U_i \\ &= \min_{u_i} \max_{\omega_i} \sum_{i=k}^{\infty} (x_i^T Qx_i + u_i^T Ru_i - \delta^2 \omega_k^T \omega_k), \end{aligned} \tag{2}$$

with Q and R are both predetermined symmetric positive definite matrices, and δ is the upper bound of the L_2-gain of the controlled system (1), which satisfies

$$\sum_{k=0}^{\infty} [x_k^T Qx_k + u_k^T Ru_k] \leq \sum_{k=0}^{\infty} \delta^2 \omega_k^T \omega_k \tag{3}$$

where $\omega_k \in L_2[0, \infty)$ and is usually a given level of disturbance attenuation.

The ideal H_∞ control u_k^* ensures that the system is asymptotically stable and can minimize the performance index function in the presence of disturbances.

Based on the Bellman optimality principle, we rewritten (2) as

$$\begin{aligned} V(x_k) &= U_k + V(x_{k+1}) \\ &= x_k^T Qx_k + u_k^T Ru_k - \delta^2 \omega_k^T \omega_k + V(x_{k+1}). \end{aligned} \tag{4}$$

The performance index function (2) is assumed to have the following form:

$$V(x_k) = x_k^T Px_k. \tag{5}$$

And the Bellman equation (4) is given by

$$x_k^T Px_k = x_k^T Qx_k + u_k^T Ru_k - \delta^2 \omega_k^T \omega_k + x_{k+1}^T Px_{k+1} \tag{6}$$

The H_∞ control problem solved in this paper can be regarded as the minimax problem of the performance index function (4), that is

$$V^*(x_k) = \min_{u_k}\max_{\omega_k}(U_k + V^*(x_{k+1})). \quad (7)$$

Then, the Bellman temporal difference error satifies

$$\lim_{k\to\infty} e_k = \lim_{k\to\infty} (x_k^T Q x_k + u_k^T R u_k - \delta^2 \omega_k^T \omega_k + x_{k+1}^T P x_{k+1} - x_k^T P x_k = 0. \quad (8)$$

Define the Hamiltonian function as

$$H_k = x_k^T Q x_k + u_k^T R u_k - \delta^2 \omega_k^T \omega_k + x_{k+1}^T P x_{k+1} - x_k^T P x_k. \quad (9)$$

Based on (7) and (9), the control policy u_k^* can be obtained by solving for $\partial H_k/\partial u_k = 0$, when ω_k^* is the solution to $\partial H_k/\partial \omega_k = 0$.

Therefore, the optimal control can be obtained as follows:

$$u_k^* = -(R_1 - R_2 R_4^{-1} R_3)^{-1}(B_1 - R_2 R_4^{-1} B_2)x_k. \quad (10)$$

The worst case disturbance can be obtained from the following equation

$$\omega_k^* = (R_4 - R_3 R_1^{-1} R_2)^{-1}(B_2 - R_3 R_1^{-1} B_1)x_k. \quad (11)$$

with

$$P = A^T P A + Q - [B_1^T \; B_2^T]\begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix}\begin{bmatrix} B_1 \\ B_2 \end{bmatrix}. \quad (12)$$

where $B_1 = B^T PA$, $B_2 = D^T PA$, $R_1 = R + B^T PB$, $R_2 = B^T PD$, $R_3 = R_2^T$, $R_4 = D^T PD - \delta^2 I$.

Based on this, the following algorithm can be designed to solve the H_∞ control strategy.

Algorithm 1 (*PI for H_∞ control*)

Perform policy evaluation and policy improvement iterative calculation process from i to n:

Policy evaluation: Solve for P^{i+1} by the following equation

$$x_k^T P^{i+1} x_k = x_k^T Q x_k + (u_k^i)^T R u_k^i - \delta^2 (\omega_k^i)^T \omega_k^i + x_{k+1}^T P^{i+1} x_{k+1} \quad (13)$$

Policy improvement: Solve for the control policy u_k^{i+1} by the following equation

$$u_k^{i+1} = -(R_1^{i+1} - R_2^{i+1}(R_4^{i+1})^{-1}R_3^{i+1})^{-1} \times (B_1^{i+1} - R_2^{i+1}(R_4^{i+1})^{-1}B_2^{i+1})x_k. \quad (14)$$

The updated disturbance ω_k^{i+1} is

$$\omega_k^{i+1} = -(R_4^{i+1} - R_3^{i+1}(R_1^{i+1})^{-1}R_2^{i+1})^{-1} \times (B_2^{i+1} - R_3^{i+1}(R_1^{i+1})^{-1}B_1^{i+1})x_k. \quad (15)$$

with $B_1^{i+1} = B^T P^{i+1} A$, $B_2^{i+1} = D^T P^{i+1} A$, $R_1^{i+1} = R + B^T P^{i+1} B$, $R_2^{i+1} = B^T P^{i+1} D$, $R_3^{i+1} = (R_2^{i+1})^T$, $R_4 = D^T P^{i+1} D - \delta^2 I$.

It can be seen from Algorithm 1 that system dynamics and state information are needed, which is very demanding in practical situations. Therefore, the next step the data-based H_∞ control, which utilizes the measured data of the system.

III. DATA-BASED H_∞ CONTROL

First, we construct a system equation consisting of input and output sequences, then a data-based Bellman equation is established. Finally, an PI-based H_∞ control method is presented. From (1), we have

$$x_k = Ax_{k-1} + Bu_{k-1} + D\omega_{k-1}. \quad (16)$$

Then substituting the above equation from k to $k-N$ into the equation of state, we can get:

$$x_k = A^N x_{k-N} + \begin{bmatrix} B & AB & \cdots & A^{N-1}B \end{bmatrix}\begin{bmatrix} u_{k-1} \\ \vdots \\ u_{k-N} \end{bmatrix} + \begin{bmatrix} D & AD & \cdots & A^{N-1}D \end{bmatrix}\begin{bmatrix} \omega_{k-1} \\ \vdots \\ \omega_{k-N} \end{bmatrix}$$
$$\doteq A^N x_{k-N} + E\bar{u}_{(k-1,k-N)} + F\bar{\omega}_{(k-1,k-N)}, \quad (17)$$

with $A^N \in \mathbb{R}^{n\times n}$, $E \in \mathbb{R}^{n\times Nm}$ and $F \in \mathbb{R}^{n\times Nq}$.

In the same way, the output equation can be written as

$$\begin{bmatrix} y_{k-1} \\ \vdots \\ y_{k-N} \end{bmatrix} = \begin{bmatrix} CA^{N-1} \\ CA^{N-2} \\ \vdots \\ C \end{bmatrix} x_{k-N} + \begin{bmatrix} \mathbf{0} & CD & \cdots & CA^{N-3}D & CA^{N-2}D \\ & \mathbf{0} & \ddots & CA^{N-4}D & CA^{N-3}D \\ & & \ddots & \vdots & \vdots \\ & \mathbf{0} & & \mathbf{0} & CD \\ & & & & \mathbf{0} \end{bmatrix}\begin{bmatrix} \omega_{k-1} \\ \omega_{k-2} \\ \vdots \\ \omega_{k-N} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & CB & \cdots & CA^{N-3}B & CA^{N-2}B \\ & \mathbf{0} & \ddots & CA^{N-4}B & CA^{N-3}B \\ & & \ddots & \vdots & \vdots \\ & \mathbf{0} & & \mathbf{0} & CB \\ & & & & \mathbf{0} \end{bmatrix}\begin{bmatrix} u_{k-1} \\ u_{k-2} \\ \vdots \\ u_{k-N} \end{bmatrix}$$
$$\doteq Gx_{k-N} + H\bar{u}_{(k-1,k-N)} + L\bar{\omega}_{(k-1,k-N)}, \quad (18)$$

where $G \in \mathbb{R}^{Np\times n}$, $H \in \mathbb{R}^{Np\times Nm}$ and $L \in \mathbb{R}^{Np\times Nq}$.

Assume that there is an appropriate observability index Q, so that

$$A^N = QG, \quad (19)$$

with G is a full column rank matrix.

Theorem 1: By using the system input and output sequence, the state equation (1) can be expressed as follows

$$x_k = Wz_k \quad (20)$$

where $W = [W_1\ W_2\ W_3]$, $W_1 = E - Q_1H$, $W_2 = F - Q_1L$, $W_3 = Q_1$, $Q_1 = A^N G^-$ and the data-based state vector $z_k = [\bar{u}_{(k-1,k-N)}\ \bar{\omega}_{(k-1,k-N)}\ \bar{y}_{(k-1,k-N)}]^T$.

Proof: According to (19), by substituting (18) into (17), we have

$$A^N x_{k-N} = Q\bar{y}_{(k-1,k-N)} - QH\bar{u}_{(k-1,k-N)} - QL\bar{\omega}_{(k-1,k-N)} \tag{21}$$

And the left inverse of G is given by

$$G^- = (G^T G)^{-1} G^T. \tag{22}$$

The general solution of (19) is

$$Q = A^N G^- + Z(I - GG^-) \doteq Q_1 + Q_2. \tag{23}$$

Note that $Q_2 G = Z(I - GG^-)G = 0$, then (21) can be rewritten as

$$x_k = Q_1\bar{y}_{(k-1,k-N)} + (E - Q_1H)\bar{u}_{(k-1,k-N)} + (F - Q_1L)\bar{\omega}_{(k-1,k-N)}, \tag{24}$$

The above equation can be defined as

$$\begin{aligned} x_k &= W_1\bar{u}_{(k-1,k-N)} + W_2\bar{\omega}_{(k-1,k-N)} \\ &\quad \times W_3\bar{y}_{(k-1,k-N)} \\ &= \begin{bmatrix} W_1 & W_2 & W_3 \end{bmatrix} \begin{bmatrix} \bar{u}_{(k-1,k-N)} \\ \bar{\omega}_{(k-1,k-N)} \\ \bar{y}_{(k-1,k-N)} \end{bmatrix}, \end{aligned} \tag{25}$$

which is the state equation (20) represented by the system measured data. ■

Remark 1: The construction principle of [29] can be referred to to ensure the observability and controllability of the data-based system.

The feedback control policy considered in this paper is assumed to have the following form

$$u_k = -Kz_k. \tag{26}$$

According to theorem 1, the performance index function (5) is given by

$$V(x_k) = z_k^T W^T P W z_k \doteq z_k^T \widehat{P} z_k. \tag{27}$$

Then, based on the Bellman equation (6), we have

$$z_k^T \widehat{P} z_k = z_k^T \widehat{Q} z_k + u_k^T R u_k - \delta^2 \omega_k^T \omega_k + z_{k+1}^T \widehat{P} z_{k+1} \tag{28}$$

with $\widehat{Q} = W^T Q W$, which is the data-based Bellman equation represented by measured data.

The corresponding Hamiltonian function is given by

$$H_k = z_k^T \widehat{Q} z_k + u_k^T R u_k - \delta^2 \omega_k^T \omega_k + z_{k+1}^T \widehat{P} z_{k+1} - z_k^T \widehat{P} z_k \tag{29}$$

IV. PI-BASED H_∞ CONTROL

Since $\widehat{P}$ is a symmetric matrix, we can write $\widehat{P}$ as a block matrix of the form

$$\widehat{P} = \begin{bmatrix} p_{uu} & p_{u\bar{u}} & p_{u\omega} & p_{u\bar{\omega}} & p_{u\bar{y}} \\ p_{\bar{u}u} & p_{\bar{u}\bar{u}} & p_{\bar{u}\omega} & p_{\bar{u}\bar{\omega}} & p_{\bar{u}\bar{y}} \\ p_{\omega u} & p_{\omega\bar{u}} & p_{\omega\omega} & p_{\omega\bar{\omega}} & p_{\omega\bar{y}} \\ p_{\bar{\omega}u} & p_{\bar{\omega}\bar{u}} & p_{\bar{\omega}\omega} & p_{\bar{\omega}\bar{\omega}} & p_{\bar{\omega}\bar{y}} \\ p_{\bar{y}u} & p_{\bar{y}\bar{u}} & p_{\bar{y}\omega} & p_{\bar{y}\bar{\omega}} & p_{\bar{y}\bar{y}} \end{bmatrix} \tag{30}$$

The goal of H_∞ control is to minimize the performance index function with disturbance. According to the Hamiltonian function (9), the optimal control is solved by $\partial H_k/\partial u_k = 0$ and the disturbance is the solution of $\partial H_k/\partial \omega_k = 0$.

Then the optimal control is

$$\begin{aligned} u_k = &-(R + p_{uu})^{-1}(p_{u\bar{u}}\bar{u}_{(k-1,k-N+1)} + p_{u\omega}\omega_k \\ &+ p_{u\bar{\omega}}\bar{\omega}_{(k-1,k-N+1)} + p_{u\bar{y}}\bar{y}_{(k,k-N+1)}). \end{aligned} \tag{31}$$

And the corresponding disturbance is given by

$$\begin{aligned} \omega_k = &-(I\delta^2 - p_{\omega\omega})^{-1}(p_{\omega u}u_k + p_{\omega\bar{u}}\bar{u}_{(k-1,k-N+1)} \\ &+ p_{\omega\bar{\omega}}\bar{\omega}_{(k-1,k-N+1)} + p_{\omega\bar{y}}\bar{y}_{(k,k-N+1)}). \end{aligned} \tag{32}$$

Theorem 2: If the data-based state equation (20) holds and O_1 and O_2 exist, then the H_∞ control policy can be represented by the system measured data as follows

$$u_k = -\Phi_1\bar{u}_{(k-1,k-N+1)} - \Phi_2\bar{\omega}_{(k-1,k-N+1)} - \Phi_3\bar{y}_{(k,k-N+1)}. \tag{33}$$

with $O_1 = (I - \Phi_u p_{u\omega}\Phi_w p_{\omega u})^{-1}$, $\Phi_1 = O_1\Phi_u(p_{u\bar{u}} - p_{u\omega}\Phi_w p_{\omega\bar{u}})$, $\Phi_2 = O_1\Phi_u(p_{u\bar{\omega}} - p_{u\omega}\Phi_w p_{\omega\bar{\omega}})$ and $\Phi_3 = O_1\Phi_u(p_{u\bar{y}} - p_{u\omega}\Phi_w p_{\omega\bar{y}})$.

$$\omega_k = -\Psi_1\bar{u}_{(k-1,k-N+1)} - \Psi_2\bar{\omega}_{(k-1,k-N+1)} - \Psi_3\bar{y}_{(k,k-N+1)}. \tag{34}$$

with $O_2 = (I - \Phi_w p_{\omega u}\Phi_u p_{u\omega})^{-1}$, $\Psi_1 = O_2\Phi_w(p_{\omega\bar{u}} - p_{\omega u}\Phi_u p_{u\bar{u}})$, $\Psi_2 = O_2\Phi_w(p_{u\bar{\omega}} - p_{u\omega}\Phi_w p_{\omega\bar{\omega}})$ and $\Psi_3 = O_2\Phi_w(p_{u\bar{y}} - p_{u\omega}\Phi_w p_{\omega\bar{y}})$.

Proof: According to (32), (31) has the following form

$$\begin{aligned} u_k = &-\Phi_u[p_{u\bar{u}}\bar{u}_{(k-1,k-N+1)} - p_{u\omega}\Phi_w(p_{\omega u}u_k \\ &+ p_{\omega\bar{u}}\bar{u}_{(k-1,k-N+1)} + p_{\omega\bar{\omega}}\bar{\omega}_{(k-1,k-N+1)} \\ &+ p_{\omega\bar{y}}\bar{y}_{(k,k-N+1)}) + p_{u\bar{\omega}}\bar{\omega}_{(k-1,k-N+1)} \\ &+ p_{u\bar{y}}\bar{y}_{(k,k-N+1)}]. \end{aligned} \tag{35}$$

with $\Phi_u = (p_{uu} + R)^{-1}$ and $\Phi_w = (I\delta^2 - p_{\omega\omega})^{-1}$.

Further, we can get

$$\begin{aligned} u_k(I - \Phi_u p_{u\omega}\Phi_w p_{\omega u}) = &-\Phi_u \\ &\cdot [(p_{u\bar{u}} - p_{u\omega}\Phi_w p_{\omega\bar{u}})\bar{u}_{(k-1,k-N+1)} \\ &+ (p_{u\bar{\omega}} - p_{u\omega}\Phi_w p_{\omega\bar{\omega}})\bar{\omega}_{(k-1,k-N+1)} \\ &+ (p_{u\bar{y}} - p_{u\omega}\Phi_w p_{\omega\bar{y}})\bar{y}_{(k,k-N+1)}]. \end{aligned} \tag{36}$$

Then, if $O_1 = (I - \Phi_u p_{u\omega}\Phi_w p_{\omega u})^{-1}$ exists, the control policy can be solved by

$$u_k = -\Phi_1\bar{u}_{(k-1,k-N+1)} - \Phi_2\bar{\omega}_{(k-1,k-N+1)} - \Phi_3\bar{y}_{(k,k-N+1)}. \tag{37}$$

with $\Phi_1 = O_1\Phi_u(p_{u\bar{u}} - p_{u\omega}\Phi_w p_{\omega\bar{u}})$, $\Phi_2 = O_1\Phi_u(p_{u\bar{\omega}} - p_{u\omega}\Phi_w p_{\omega\bar{\omega}})$ and $\Phi_3 = O_1\Phi_u(p_{u\bar{y}} - p_{u\omega}\Phi_w p_{\omega\bar{y}})$.

In the same way, the disturbance is given by

$$\begin{aligned}\omega_k &= -\Phi_w[-p_{\omega u}\Phi_u(p_{u\bar{u}}\bar{u}_{(k-1,k-N+1)} + p_{u\omega}\omega_k \\ &\quad + p_{u\bar{\omega}}\bar{\omega}_{(k-1,k-N+1)} + p_{u\bar{y}}\bar{y}_{(k,k-N+1)}) \\ &\quad + p_{\omega\bar{u}}\bar{u}_{(k-1,k-N+1)} + p_{\omega\bar{\omega}}\bar{\omega}_{(k-1,k-N+1)} \\ &\quad + p_{\omega\bar{y}}\bar{y}_{(k,k-N+1)}]. \end{aligned} \tag{38}$$

If $O_2 = (I - \Phi_u p_{u\omega}\Phi_w p_{\omega u})^{-1}$ exists, the disturbance can be expressed as

$$\omega_k = -\Psi_1\bar{u}_{(k-1,k-N+1)} - \Psi_2\bar{\omega}_{(k-1,k-N+1)} - \Psi_3\bar{y}_{(k,k-N+1)}. \tag{39}$$

with $\Psi_1 = O_2\Phi_w(p_{\omega\bar{u}} - p_{\omega u}\Phi_u p_{u\bar{u}})$, $\Psi_2 = O_2\Phi_w(p_{u\bar{\omega}} - p_{u\omega}\Phi_w p_{\omega\bar{\omega}})$ and $\Psi_3 = O_2\Phi_w(p_{u\bar{y}} - p_{u\omega}\Phi_w p_{\omega\bar{y}})$. This completes the proof. ■

Theorem 2 gives the form of the system inputs represented by the system measurement data. Based on this, we propose the following PI based PI-based H_∞ control algorithm.

Algorithm 2 (*Data-based PI for H_∞ control*)

Start the iterative algorithm with an initial admissible u_k^0. Perform policy evaluation and policy improvement iterative calculation process from i to n:

Policy evaluation: Calculate $\widehat{P}^{i+1}$ with a certain precision

$$\begin{aligned} z_k^T\widehat{P}^{i+1}z_k &= z_k^T\widehat{Q}z_k + (u_k^{i+1})^T R u_k^{i+1} - \delta^2(\omega_k^{i+1})^T\omega_k^{i+1} \\ &\quad + z_{k+1}^T\widehat{P}^{i+1}z_{k+1} \end{aligned} \tag{40}$$

Policy improvement: Calculate the update control

$$\begin{aligned} u_k^{i+1} &= -\Phi_1^{i+1}\bar{u}_{(k-1,k-N+1)} - \Phi_2^{i+1}\bar{\omega}_{(k-1,k-N+1)} \\ &\quad - \Phi_3^{i+1}\bar{y}_{(k,k-N+1)}. \end{aligned} \tag{41}$$

Calculate the update disturbance

$$\begin{aligned} \omega_k^{i+1} &= -\Psi_1^{i+1}\bar{u}_{(k-1,k-N+1)} - \Psi_2^{i+1}\bar{\omega}_{(k-1,k-N+1)} \\ &\quad - \Psi_3^{i+1}\bar{y}_{(k,k-N+1)}. \end{aligned} \tag{42}$$

The iteration is stopped when the PI algorithm converges. ■

The policy evaluation equation (40) is derived from (28), which is equivalent to the Bellman equation in [40]. And the goal of each policy improvement equation (41) is to minimize the respective (4) in the case of worst disturbances. Therefore, the convergence of the two algorithms is equivalent, and the detailed convergence proof can refer to [40]–[42].

Since this paper uses only input/output data to construct the PI-based structure to directly solve the H_∞ control policy, it requires a sufficient training process.

V. SIMULATION

A. THE DT SYSTEM WITH DISTURBANCES

Consider the DT system with disturbances:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + D\omega_k \\ y_k &= Cx_k, \end{aligned} \tag{43}$$

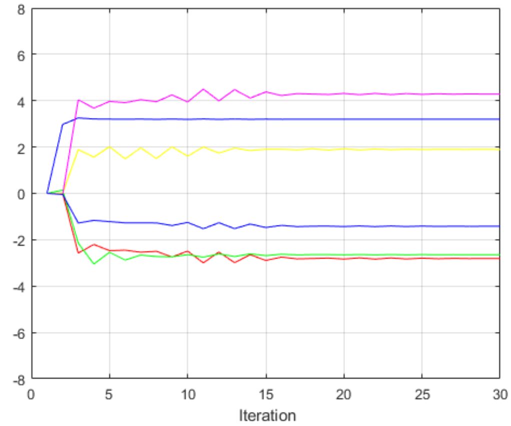


FIGURE 1. **Convergence process of $\widetilde{P}_{1i}$.**

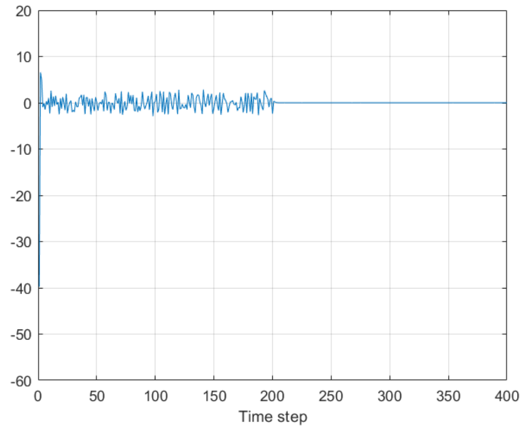


FIGURE 2. **Control policy during the training process.**

with system dynamics given by $A = [1.3000 - 0.4200; 1\ 0]$, $B = [1; 0]$, $C = [1\ \ -0.8]$ and $D = [0.01; 0.05]$. The parameter of the cost function were set to $R = 1$, $Q = 3 \times I_{3\times3}$. A probing signal was injected into the learning process to satisfy the PE condition.

Figure 1 shows the convergence process of P solved by the data-based PI method as the iterative learning continues. And we can see that the convergence process of the PI-based control method was slowed down by the influence of disturbance, and more iterative learning steps are needed to achieve convergence.

From figure 2 we can see the control input with the probing noise signal. The probing noise signal added to the training process of the PI algorithm is shown in figure 3 and removed at 200 time steps. Since the probing signal can fully simulate the situation of the system, the ADP algorithm can obtain the optimal information by obtaining the information of the system through the data.

B. WIND TURBINE CONTROL

For large wind turbines, when the wind speed exceeds the rated wind speed, PI variable pitch control is usually adopted to adjust the blade pitch angle to achieve constant power

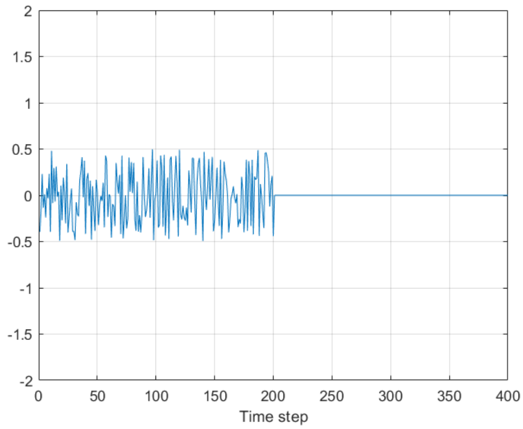


FIGURE 3. Probing noise signal during the training process.

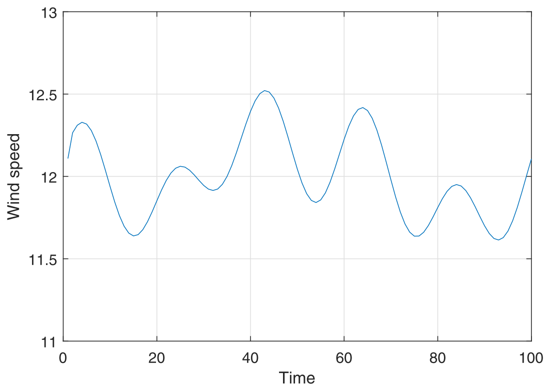


FIGURE 4. Wind speed.

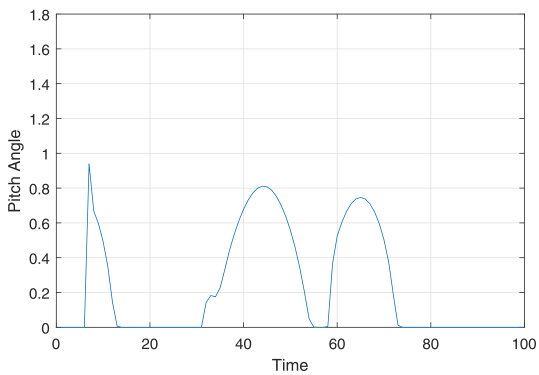


FIGURE 5. Pitch angle of the wind turbine.

control. Due to the characteristics of time delay and disturbance in large wind turbines, if the parameters of the variable pitch controller are improperly selected, the output power will fluctuate greatly, which will cause adverse effects on the power grid. Therefore, it is particularly important to study the parameter optimization of the variable pitch controller. Considering the control problem of SUT1000 doubly-fed wind generator, the rated power is 1MW and the rated wind speed is 12m/s. The parameter of the cost function were set to $R = 2$, $Q = 2 \times I_{3\times3}$.

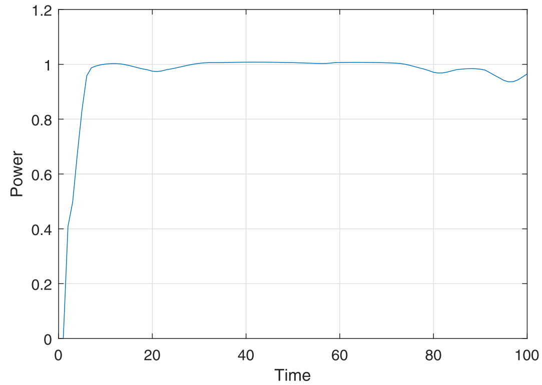


FIGURE 6. Output power of the wind turbine.

The wind speed during simulation is shown in figure 4. Through iterative learning, the variable pitch control of the variable pitch system is shown in figure 5. The output power of the wind turbine is presented in figure 6. It can be seen that the control method can ensure the stability of the output power under the condition that the wind speed fluctuates greatly.

VI. CONCLUSION

For the optimal control of discrete system with disturbances, a novel data-based PI algorithm was presented. By using the input and output data in a finite horizon, the new data-based system equations were constructed, thus effectively avoiding the need for system dynamics and state. Based on the new system equation, the data-based bellman equation was obtained. For the DT systems with disturbance, the optimal control was transformed into the optimal solution for the zero-sum game. Then the PI-based control method was designed to deal with the H_∞ control problem using only system data. As can be seen from the simulation, due to the disturbance, the convergence process of the data-based PI control method becomes longer, and more iteration steps are needed to identify the internal information of the system. Future work will attempt to deal with the H_∞ control problem for time-varying delay systems.

REFERENCES

[1] Q. Wei, R. Song, Z. Liao, B. Li, and F. L. Lewis, "Discrete-time impulsive adaptive dynamic programming," *IEEE Trans. Cybern.*, to be published. [Online]. Available: https://ieeexplore.ieee.org/document/8685683, doi: 10.1109/TCYB.2019.2906694.

[2] M. Bi, A. Wang, J. Xu, and F. Zhou, "Anomaly behavior detection of database user based on discrete-time Markov chain," *J. Shenyang Univ. Technol.*, vol. 40, no. 1, pp. 70–76, 2018.

[3] Y. Ma and G. Wang, "Detection for bridge surface crack based on 2D complex discrete wavelet packet transform," *J. Shenyang Univ. Technol.*, vol. 40, no. 6, pp. 659–663, 2018.

[4] Q. Wei, B. Li, and R. Song, "Discrete–time stable generalized self–learning optimal control with approximation errors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1226–1238, Apr. 2018.

[5] Y. Liu, H. Zhang, Y. Luo, and J. Han, "ADP based optimal tracking control for a class of linear discrete-time system with multiple delays," *J. Franklin Inst.*, vol. 353, no. 9, pp. 2117–2136, Jun. 2016.

[6] Y. Liu and R. Yu, "Model-free optimal tracking control for discrete-time system with delays using reinforcement Q-learning," *Electron. Lett.*, vol. 54, no. 12, pp. 750–752, Jun. 2018.

[7] H. Zhang, Y. Liang, H. Su, and C. Liu, "Event-driven guaranteed cost control design for nonlinear systems with actuator faults via reinforcement learning algorithm," *IEEE Trans. Syst. Man Cybern, Syst.*, to be published, doi: 10.1109/tsmc.2019.2946857.

[8] H. Zhang, K. Zhang, Y. Cai, and J. Han, "Adaptive fuzzy fault–tolerant tracking control for partially unknown systems with actuator faults via integral reinforcement learning method," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 10, pp. 1986–1998, Oct. 2019.

[9] H. Zhang, D. Liu, Y. Luo, and D. Wang, *Adaptive Dynamic Programming for Control: Algorithms and Stability*. London, U.K.: Springer, 2012.

[10] D. Liu, Q. Wei, D. Wang, X. Yang, and H. Li, *Adaptive Dynamic Programming With Applications in Optimal Control*. London, U.K.: Springer 2017.

[11] G. Zhu, M. Lei, and X. Zhao, "Adaptive iterative learning control for permanent magnet synchronous motor servo system," *J. Shenyang Univ. Technol.*, vol. 40, no. 1, pp. 6–11, 2018.

[12] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control*. Hoboken, NJ, USA: Wiley, 2012.

[13] D. Liu, Y. Xu, Q. Wei, and X. Liu, "Residential energy scheduling for variable weather solar energy based on adaptive dynamic programming," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 1, pp. 36–46, Jan. 2018.

[14] L. Zhang, Y. Zhu, H. Shi, and K.-S. Hwang, "Adaptive dynamic programming approach on optimal control for affinely pseudo-linearized nonlinear system," *IEEE Access*, vol. 7, pp. 75132–75142, 2019.

[15] X. Luo, Y. Lv, R. Li, and Y. Chen, "Web service QoS prediction based on adaptive dynamic programming using fuzzy neural networks for cloud services," *IEEE Access*, vol. 3, pp. 2260–2269, 2015.

[16] C. Mu, Y. Zhang, and K. Wang, "Observer–based adaptive control of uncertain nonlinear systems via neural networks," *IEEE Access*, vol. 6, pp. 42675–42686, 2018.

[17] S. H. Mousavi, M. Davari, and H. J. Marquez, "An innovative event–based filtering scheme using $\boldsymbol{H}_\infty$ performance for stochastic LTI systems considering a practical application in smart modernized microgrids," *IEEE Access*, vol. 7, pp. 48138–48150, 2019.

[18] P.-J. Ko and M.-C. Tsai, "H_∞ control design of PID–like controller for speed drive systems," *IEEE Access*, vol. 6, pp. 36711–36722, 2018.

[19] S. Lu, S. Fu, and W. Tong, "Robust control for variable pitch of wind turbines," *J. Shenyang Univ. Technol.*, vol. 40, no. 6, pp. 627–631, 2018.

[20] T. Kerdphol, F. S. Rahman, M. Watanabe, and Y. Mitani, "Robust virtual inertia control of a low inertia microgrid considering frequency measurement effects," *IEEE Access*, vol. 7, pp. 57550–57560, 2019.

[21] R. Hayat, M. Leibold, and M. Buss, "Robust-adaptive controller design for robot manipulators using the $\mathcal{H}_\infty$ approach," *IEEE Access*, vol. 6, pp. 51626–51639, 2018.

[22] D. Wang, C. Mu, D. Liu, and H. Ma, "On mixed data and event driven design for adaptive-critic-based nonlinear H_∞ control," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 993–1005, Apr. 2018, doi: 10.1109/cdc.2016.7799163.

[23] Y. Zhu, D. Zhao, X. Yang, and Q. Zhang, "Policy iteration for H_∞ optimal control of polynomial nonlinear systems via sum of squares programming," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 500–509, Feb. 2018.

[24] A. Al-Tamimi, M. Abu-Khalaf, and F. L. Lewis, "Adaptive critic designs for discrete-time zero-sum games with application to H_∞ control," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 1, pp. 240–247, Feb. 2007.

[25] H. Su, H. Zhang, D. W. Gao, and Y. Luo, "Adaptive dynamics programming for H_∞ control of continuous-time unknown nonlinear systems via generalized fuzzy hyperbolic models," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2019.2900750.

[26] J. Hou, D. Wang, D. Liu, and Y. Zhang, "Model-free H_∞ optimal tracking control of constrained nonlinear systems via an iterative adaptive learning algorithm," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2018.2863708.

[27] H. Zhang, C. Qin, B. Jiang, and Y. Luo, "Online adaptive policy learning algorithm for H_∞ state feedback control of unknown affine nonlinear discrete–time systems," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2706–2718, Dec. 2014.

[28] X. Zhong, H. He, D. Wang, and Z. Ni, "Model–free adaptive control for unknown nonlinear zero–sum differential game," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1633–1646, May 2018.

[29] F. L. Lewis and K. G. Vamvoudakis, "Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data," *IEEE Trans. Syst., Man, Cybern. B*, vol. 41, no. 1, pp. 14–25, Feb. 2011.

[30] K. Xu, D. Feng, F. Wang, and X. Peng, "Layout of sensors in sunlight greenhouse based on data fusion," *J. Shenyang Univ. Technol.*, vol. 41, no. 1, pp. 91–97, 2019.

[31] Y. Liu, H. Zhang, R. Yu, and Q. Qu, "Data-driven optimal tracking control for discrete-time systems with delays using adaptive dynamic programming," *J. Franklin Inst.*, vol. 355, no. 13, pp. 5649–5666, Sep. 2018.

[32] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control," *Automatica*, vol. 43, no. 3, pp. 473–481, Mar. 2007.

[33] H. Zhang, Y. Liu, G. Xiao, and H. Jiang, "Data-based adaptive dynamic programming for a class of discrete-time systems with multiple delays," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2017.2758849.

[34] W. Han and C. Yu, "Distributed data storage architecture for data storage management of power transmission and transformation engineering," *J. Shenyang Univ. Technol.*, vol. 41, no. 4, pp. 366–371, 2019.

[35] B. Kiumarsi, F. L. Lewis, M.-B. Naghibi-Sistani, and A. Karimpour, "Optimal tracking control of unknown discrete–time linear systems using input–output measured data," *IEEE Trans. Cybern.*, vol. 45, no. 12, pp. 2770–2779, Dec. 2015.

[36] Q. Wei, R. Song, and P. Yan, "Data–driven zero–sum neuro–optimal control for a class of continuous–time unknown nonlinear systems with disturbance using ADP," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 2, pp. 444–458, Feb. 2016.

[37] C. Mu, Z. Ni, C. Sun, and H. He, "Data–driven tracking control with adaptive dynamic programming for a class of continuous–time nonlinear systems," *IEEE Trans. Cybern.*, vol. 47, no. 6, pp. 1460–1470, Jun. 2017.

[38] B. Luo, D. Liu, H.-N. Wu, D. Wang, and F. L. Lewis, "Policy gradient adaptive dynamic programming for data–based optimal control," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3341–3354, Oct. 2017.

[39] T. Bian and Z.-P. Jiang, "Value iteration and adaptive dynamic programming for data-driven adaptive optimal control design," *Automatica*, vol. 71, pp. 348–360, Sep. 2016.

[40] B. Kiumarsi, F. L. Lewis, and Z.-P. Jiang, "H_∞ control of linear discrete-time systems: Off-policy reinforcement learning," *Automatica*, vol. 78, pp. 144–152, Apr. 2017.

[41] Q. Wei, F. L. Lewis, D. Liu, R. Song, and H. Lin, "Discrete–time local value iteration adaptive dynamic programming: Convergence analysis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 6, pp. 875–891, Jun. 2018.

[42] Q. Wei, D. Liu, Q. Lin, and R. Song, "Adaptive dynamic programming for discrete–time zero–sum games," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 957–969, Apr. 2018.

YANG LIU (Member, IEEE) received the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2018.

He is currently a Postdoctoral Fellow with the School of Electrical Engineering, Shenyang University of Technology, Shenyang. His current research interests include adaptive dynamic programming, wind power control, reinforcement learning, optimal control, adaptive control, and their industrial applications.

ZUOXIA XING received the Ph.D. degree in automation of electric power systems from Beijing Jiaotong University, Beijing, China, in 2008.

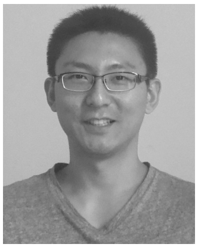
She has been a Professor with the Shenyang University of Technology, since 2018. Her current research interests include new energy control, multienergy complementary, energy storage, and consumption.

LEI CHEN received the B.S. degree in aerodynamics from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1997.

He is currently an Engineer with the New Energy Research Institute, Shenyang University of Technology, Shenyang, China. His current research interests include wind energy development and utilization, electric heat storage, and fluid dynamics.

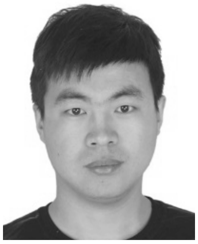
JIAN XU received the M.A.Eng. degree in electric machines and electric apparatus from the Shenyang University of Technology, China, in 2011, where he is currently pursuing the Ph.D. degree in electric engineering.

His current research mainly focuses on the Internet of energy and wind turbine control strategy.



YUNLU LI received the B.S. degree in electronic information engineering from the Shenyang University of Technology, Shenyang, China, in 2009, the M.S. degree in control engineering, in 2011, and the Ph.D. degree in power electronics and drives from Northeastern University, in 2017, Shenyang, China.

He was a Postdoctoral Researcher with the Shenyang University of Technology, from 2017 to 2019, where he is currently a Lecturer with the School of Electrical Engineering. His research interests include phase-locked loop and nonlinear filtering techniques for distributed systems and power quality and data-driven modeling of micro-grids.



HAIXIN WANG (Member, IEEE) received the B.Eng. degree in electrical engineering from Inner Mongolia Agricultural University, Hohhot, China, in 2011, and the M.Sc. and Ph.D. degrees in electrical engineering from the Shenyang University of Technology, Shenyang, China, in 2014 and 2017, respectively.

He is currently a Postdoctoral Researcher with the Shenyang University of Technology. He is also doing his postdoctoral research as a Visiting Scholar with the School of Engineering, Cardiff University, Cardiff, U.K. His research interests include cyber-physical systems, wind power generation systems, and integrated energy systems.