Robust Control of An Inverted Pendulum System Based on Policy Iteration in Reinforcement Learning

Xu Dengguo *, Ma Yan , Huang Jiashun , Li Yahui

Posted Date: 17 October 2023

doi: 10.20944/preprints202310.1100.v1

Keywords: robust control; optimal control; inverted pendulum system; reinforcement learning

Article

Robust Control of An Inverted Pendulum System Based on Policy Iteration in Reinforcement Learning

Yan Ma [1], Dengguo Xu [1,2,*], Jiashun Huang [1] and Yahui Li [1]

[1] School of Automation, Guangxi University of Science and Technology, Liuzhou, 545000, China.
[2] School of Physics and Electrical Engineering, Liupanshui Normal University, Liupanshui, 553000, China.
* Correspondence: dengguoxu@163.com (Dengguo Xu)

Abstract: This paper is primarily focused on the robust control of an inverted pendulum system based on the policy iteration in reinforcement learning. First, a mathematical model of the single inverted pendulum system is established through a force analysis of the pendulum and trolley. Second, based on the theory of robust optimal control, the robust control of the uncertain linear inverted pendulum system is transformed into an optimal control problem with an appropriate performance index. Moreover, for the uncertain linear and nonlinear systems, two reinforcement-learning control algorithms are proposed using the policy iteration method. Finally, two numerical examples are provided to validate the reinforcement learning algorithms for the robust control of the inverted pendulum systems.

Keywords: robust control; optimal control; inverted pendulum system; reinforcement learning

1. Introduction

In the last decade, there has been an increasing interest in the robust control of the inverted pendulum system (IPS) owing to its high potential in testing a variety of advanced control algorithms. Robust control is widely used in power electronic [1], flight control [2], motion control [3], network control [4], and IPSs [5], in addition to other fields. The research on the robust control of an IPS has provided advantageous results in recent years. An inverted pendulum is an experimental device having insufficient drive, absolute instability, and uncertainty. It has become an excellent benchmark in the field of automatic control over the last few decades as it provides better explanations for model-based nonlinear control techniques and is a typical experimental platform for verifying classical and modern control theories[6].

Although the earliest research on the IPS can be traced back to 1908[7], there was almost no literature on this subject between 1908 and 1960. Until 1960, a number of tall, slender structures survived the Chilean earthquake, while structures that appeared more stable were severely damaged. Therefore, some scholars conducted more in-depth research to obtain a suitable explanation [8]. The pendulum structure under the effect of earthquakes is modeled as a base and rigid block system, and block overturning is studied by applying a horizontal acceleration, sinusoidal pulses, and seismic-type excitations to the system. It was observed that there is an unexpected scaling effect that makes the large block more stable than the small block among two geometrically similar blocks. Furthermore, tall blocks exhibit greater stability during earthquakes when exposed to horizontal forces. Since then, with the development of modern control theory, various control methods have been applied to different types of IPSs, such as the proportional-integral–derivative control [9,10], cloud model control [11,12], fuzzy control [13–16], sliding mode control [17,18], and neural network control methods [19–21]. These methods provide different ideas for the control of IPSs.

As is known, the IPS is an uncertain system, and the uncertainty of its model is naturally within the scope of consideration. The aim of the robust control of an IPS is to find a controller capable of addressing system uncertainties. When the system is disturbed by uncertainty, robust control laws can stabilize the system. Because it is difficult to directly solve the robust control problem, Lin et al. transformed the robust control problem into an optimal control problem[22–24].

 doi:10.20944/preprints202310.1100.v1

However, the pioneering methods for solving optimal control problems mainly include the dynamic programming [25] and maximum principle [26] methods. In the dynamic programming method, solving the Hamilton–Jacobi–Belman (HJB) equation yields the optimal control of the system. However, the HJB equation is a nonlinear partial differential equation, and obtaining its solution has proven to be more difficult than solving the optimal control problem. As for the optimal control problem of a linear system with a quadratic performance index, irrespective of whether it is a continuous system or a discrete system, it finally comes down to solving an algebraic Riccati equation (ARE). However, when the dimension of the state vector or control input vector in the dynamic system is large, the so-called "curse of dimensionality" appears when the dynamic programming method is used to solve the optimal control problem [27]. To overcome this weakness, some scholars have proposed the use of the reinforcement learning (RL) policy to solve the optimal control problem [28,29].

When RL was initially used for system control, it was primarily focused on discrete-time systems or discretized continuous-time systems in research on problems such as the billiard game problem [30], scheduling problem [31–33], and robot navigation problem [34]. Furthermore, the application of RL algorithms to continuous-time and continuous-state systems was initially extended by Doya et al. [35]. They used the known system model to learn the optimal control policy. In the context of control engineering, the RL and adaptive dynamic programming link traditional optimal control methods with adaptive control methods [36–38]. Vrabie et al.[39] used the RL algorithm to solve the optimal control problem of the continuous time system. In the case of the linear system, the system data is collected, and the solution of the HJB equation is obtained via online policy iteration (PI) using the least squares method. Xu et al. [40,41] proposed an RL algorithm based on linear and nonlinear continuous-time systems to solve the robust tracking problems through online PI. The algorithm takes into consideration the uncertainty in the system's state and input matrices and improves the method for solving robust tracking.

The use of the RL method to iteratively solve the optimal control problem has attracted extensive attention from scholars and resulted in relatively good results. The IPS demonstrates a positive impact in the validation of advanced control algorithms. Although there is a significant amount of literature on the IPS [9,11,15], to the best of our knowledge, there are almost no research results comprising the use of the RL for solving the control of IPS. In this study, an attempt is made to solve the robust control problem of an uncertain continuous-time IPS using the RL algorithm. The dynamic equations of the nominal system need not be known when using the RL control algorithm, and this study lays a theoretical foundation for the wide application of the RL control algorithm in engineering systems. The main contributions of this study are as follows.

1) The state space model of the IPS is established and a robust optimal control design method for uncertain systems is proposed. By constructing appropriate performance indicators, the optimal control method is used for the first time to design a robust control law for an uncertain IPS, which is an original approach.

2) A PI algorithm in the RL has been designed for realizing the robust optimal control of an IPS. The use of RL algorithms to solve the robust control problem of the IPS does not require that the nominal system matrix be known, and it can also overcome the challenges resulting from the "curse of dimensionality." The first application of the RL for solving the control problem of an IPS has significance for its potential application in practical engineering.

The organization of this paper is as follows. In Section 2, the state-space equation of the IPS and a linearization model are established. The robust control and RL algorithm for linearized the IPS are presented in Sections 3 and 4. In Section 5, we establish the nonlinear state-space model of the IPS and propose the corresponding RL algorithm. The RL algorithm is then verified via a simulation in Section 6. Finally, we summarize the work of this paper and potential future research directions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

2. Preliminaries

In this section, we established a physical model of a first-order linear IPS according to Newton's second law. By selecting appropriate state variables, the state space model with uncertainty is derived.

2.1. Modeling of Inverted Pendulum System

The inverted pendulum experimental device comprises a pendulum and a trolley. Its structure is presented in Figure 1. Moreover, its simplified physical model is presented in Figure 2, which mainly includes the pendulum and trolley. In Figure 2, owing to the interaction between the trolley and pendulum, the trolley is subjected to a force F_3 from the pendulum, which acts in the lower left direction. Furthermore, the pendulum is subjected to a force F_4 from the trolley, which acts in the upper right direction. In addition, the pendulum and trolley are also subjected to other forces, as shown in Figures 3 and 4, respectively. The trolley is driven by a motor to perform horizontal movements on the guide rail. In Figure 3, the trolley is subjected to the force F_1 from the motor and gravity. F_2 represents the resistance between the trolley and guide rail. Furthermore, N_1 and P_1 are the two components of force F_3. In Figure 4, the pendulum is subjected to gravity $G = m_1 g$, and N_2 and P_2 are the two components of force F_4.

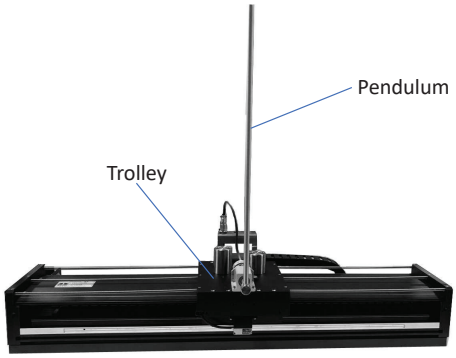

Figure 1. Inverted pendulum system diagram

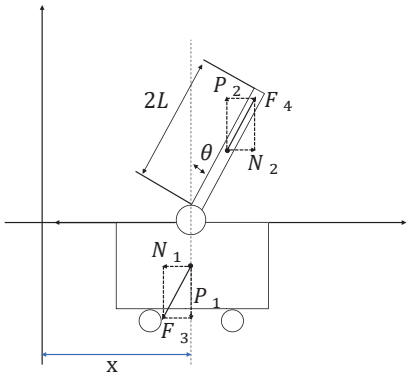

Figure 2. First-order inverted pendulum physical model

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

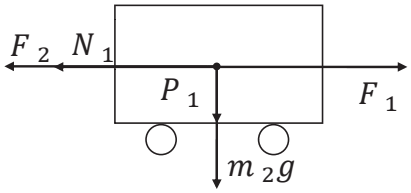

Figure 3. Force analysis of the trolley

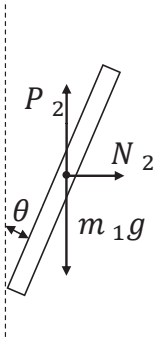

Figure 4. Force analysis of the pendulum

To facilitate subsequent calculations, we define the parameter of the first-order IPS, as shown in Table 1. The time parameter symbol (t) is omitted, which indicates that x represents $x(t)$. According to Newton's second Law, in the horizontal direction, the trolley satisfies the following equation

Table 1. IPS parameter symbols

Parameter	Unit	Significance
m_1	kg	Mass of the trolley
m_2	kg	Mass of the pendulum
L	m	Half the length of the pendulum
z	$N/m/s$	Friction coefficient between the trolley and guide rail
x	m	Displacement of the trolley
θ	rad	Angle from the upright position
I	$kg \cdot m^2$	Moment of inertia of pendulum

$$F_1 - N_1 - F_2 = m_2\ddot{x} \tag{1}$$

We assume that the resistance is proportional to the speed of the trolley. Therefore, $F_2 = z\dot{x}$, z is the proportional coefficient. Moreover, in the horizontal direction, the pendulum satisfies the following equation

$$N_2 = m_1\frac{d^2}{dt^2}(x - Lsin\theta) = m_1\ddot{x} + m_1L\dot{\theta}^2sin\theta - m_1L\ddot{\theta}cos\theta \tag{2}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

Considering that $N_1 = N_2$, and on substituting (2) into (1), we obtain

$$F_1 = (m_1 + m_2)\ddot{x} + z\dot{x} + m_1 L\dot{\theta}^2 sin\theta - m_1 L\ddot{\theta} cos\theta \tag{3}$$

Next, we analyze the resultant force in the vertical direction of the pendulum, and the following equation can be obtained.

$$P_2 - m_1 g = m_1 \frac{d^2}{dt^2}(Lcos\theta) = -m_1 L\dot{\theta}^2 cos\theta - m_1 L\ddot{\theta} sin\theta \tag{4}$$

The component force of N_2 in the direction perpendicular to the pendulum is

$$N_2 cos\theta = m_1 \frac{d^2}{dt^2}(x - Lsin\theta)cos\theta = m_1 \ddot{x} cos\theta + m_1 L\dot{\theta}^2 sin\theta \cos\theta - m_1 L\ddot{\theta} cos^2\theta \tag{5}$$

Based on the torque balance, we can obtain the following equation

$$I\ddot{\theta} = P_2 Lsin\theta + N_2 Lcos\theta \tag{6}$$

where I is the moment of inertia of the pendulum. On substituting equations (4) and (5) into equation (6),

$$(I + m_1 L^2)\ddot{\theta} - m_1 gLsin\theta = m_1 L\ddot{x} cos\theta \tag{7}$$

Thus far, equations (3) and (7) constitute the dynamic model of the IPS. Moreover, it can be assumed that the rotation angle of the pendulum is very small, that is, $\theta << 1(radian)$. Therefore, it can be approximated that

$$\sin\theta \approx \theta, \cos\theta \approx 1$$

Therefore, it follows from equations (3) and (7),

$$\begin{aligned} F_1 &= (m_1 + m_2)\ddot{x} + z\dot{x} + m_1 L\dot{\theta}^2\theta - m_1 L\ddot{\theta} \\ (I + m_1 L^2)\ddot{\theta} &- m_1 gL\theta = m_1 L\ddot{x} \end{aligned} \tag{8}$$

2.2. State Space Model with Uncertainty

In Section 2.1, we established the dynamic model of the IPS as shown in equation (8). Next, we will derive the state space model of the IPS.

As the rotation angle of the pendulum θ is very small, it can be approximated that $\dot{\theta} \approx 0$, $\dot{\theta}^2 \approx 0$. It follows from (8) that

$$\begin{aligned} F_1 &= (m_1 + m_2)\ddot{x} + z\dot{x} - m_1 L\ddot{\theta} \\ (I + m_1 L^2)\ddot{\theta} &- m_1 gL\theta = m_1 L\ddot{x} \end{aligned} \tag{9}$$

The state variables of the system can be defined as

$$x_1 = x, \quad x_2 = \dot{x}, \quad x_3 = \theta, \quad x_4 = \dot{\theta}.$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

Therefore, the following state space equation can be derived.

$$\begin{aligned}
\dot{x_1} =& x_2 \\
\dot{x_2} =& \frac{-(I+m_1L^2)z}{I(m_1+m_2)+m_1m_2L^2}x_2 + \frac{m_1^2gL^2}{I(m_1+m_2)+m_1m_2L^2}x_3 + \frac{I+m_1L^2}{I(m_1+m_2)+m_1m_2L^2}u \\
\dot{x_3} =& x_4 \\
\dot{x_4} =& \frac{-m_1Lz}{I(m_1+m_2)+m_1m_2L^2}x_2 + \frac{m_1gL(m_1+m_2)}{I(m_1+m_2)+m_1m_2L^2}x_3 + \frac{m_1L}{I(m_1+m_2)+m_1m_2L^2}u
\end{aligned} \tag{10}$$

where u represents the force F_3 from the motor. Using $W = I(m_1+m_2)+m_1m_2L^2$, equation (10) can be written as

$$\dot{x} = Ax(t) + Bu(t)$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & \frac{-(I+m_1L^2)z}{W} & \frac{m_1^2gL^2}{W} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{-m_1Lz}{W} & \frac{m_1gL(m_1+m_2)}{W} & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & \frac{I+m_1L^2}{W} & 0 & \frac{m_1L}{W} \end{bmatrix}^T$$

However, the accurate model of the IPS is difficult to obtain, and all its parameters have uncertainties. In this paper, the friction coefficient z between the trolley and guide rail is selected as an uncertain parameter. The numerical values of the other parameters in Table 1 are known, where $m_1 = 0.109, m_2 = 1.096, L = 0.25, and I = 0.0034$. Therefore, the state space model of the uncertain IPS can be abbreviated as

$$\dot{x} = A(z)x + Bu \tag{11}$$

where

$$A(z) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -0.8832z & 0.6293 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -2.3566z & 27.8285 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 0.8832 & 0 & 2.3566 \end{bmatrix}^T$$

Here we choose $z = 0.1$ as the nominal value and denote the nominal matrix of the system as $A(z_0)$. Therefore, the nominal system corresponding to the uncertain system (11) is

$$\dot{x} = A(z_0)x + Bu \tag{12}$$

where

$$A(z_0) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -0.0883 & 0.6293 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -0.2357 & 27.8285 & 0 \end{bmatrix}$$

3. Robust Control of Uncertain Linear System

This section mainly presents with robust optimal control methods for the uncertain IPS modeled in the previous section by selecting the appropriate performance index function and solving an ARE to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

construct the robust control law. When the uncertain parameters of the system change within a certain range, this robust control law can cause the system to become asymptotically stable.

The following lemmas are proposed to prove the main results of this paper.

Lemma 1. *The nominal system (12) corresponding to system (11) is stabilizeable.*

Proof. For the four-dimensional continuous time-invariant system presented in system (12), the controllability matrix is constructed as

$$G = [B \quad A(z_0)B \quad A(z_0)^2B \quad A(z_0)^3B]$$

Therefore, we have

$$rank(G) = rank \begin{bmatrix} 0 & 0.8832 & -0.0780 & 1.4899 \\ 0.8832 & -0.0780 & 1.4899 & -0.2626 \\ 0 & 2.3566 & -0.2082 & 65.5990 \\ 2.3566 & -0.2082 & 65.5990 & -6.1442 \end{bmatrix} = 4$$

Therefore, system (12) is completely controllable, which means that the system can be stabilized. This concludes the proof. □

Lemma 2. *There is an $m \times n$ matrix $\Delta(z)$, such that the system matrices $A(z)$ and $A(z_0)$ satisfy the following matched condition.*

$$A(z) - A(z_0) = B\Delta(z) \tag{13}$$

Proof.

$$A(z) - A(z_0) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.8832(0.1-z) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 2.3566(0.1-z) & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.8832 \\ 0 \\ 2.3566 \end{bmatrix} \Delta(z) = B\Delta(z) \tag{14}$$

where

$$\Delta(z) = \begin{bmatrix} 0 & 0.1-z & 0 & 0 \end{bmatrix} \tag{15}$$

This concludes the proof. □

Lemma 3. *For any $z \in [0,1]$, there exists a positive semidefinite matrix F, such that $\Delta(z)$ satisfies*

$$\Delta(z)^T\Delta(z) \leq F \tag{16}$$

where $F \geq 0$.

Proof. According to lemma 2, we can obtain

$$\Delta(z)^T\Delta(z) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & (0.1-z)^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \leq \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.81 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = F \tag{17}$$

This concludes the proof. □

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

For nominal system (12), we construct the following ARE.

$$SA(z_0) + A(z_0)^T S + Q - SBB^T S = 0 \tag{18}$$

where $Q = F + I$. According to the above three lemmas and ARE (18), we propose the following theorem.

Theorem 1. Let us suppose that S is a symmetric positive definite solution to ARE (18). Then, for all uncertainties $z \in [0, 1]$, the feedback control $u = Kx$, $K = -B^T S$ can make the system (11) asymptotically stable.

Proof of Theorem 1. We define the following Lyapunov function.

$$V(x) = x^T S x \tag{19}$$

We set $u = Kx$ and take the time derivative of the Lyapunov function (19) along the system (11). We can then obtain

$$\dot{V}(x) = x^T \left[(A(z))^T + K^T B^T \right] Sx + x^T S \left[(A(z)) + BK \right] x$$

According to Lemma 2, we can obtain

$$\begin{aligned} \dot{V}(x) =& x^T \left[(A(z_0) + B\Delta(z))^T + K^T B^T \right] Sx + x^T S \left[(A(z_0) + B\Delta(z)) + BK \right] x \\ =& x^T [A(z_0)^T S + SA(z_0) + \Delta(z)^T B^T S] x + x^T SB\Delta(z)x + 2x^T SBKx \end{aligned}$$

On substituting ARE (18) into the above equation, we obtain

$$\dot{V}(x) = - x^T Q x + x^T SB^T Sx + x^T \Delta(z)^T B^T Sx + x^T SB\Delta(z)x + 2x^T SBKx$$

because $K = -B^T S$,

$$\dot{V}(x) = - x^T Q x + x^T K^T Kx - x^T \Delta(z)^T Kx + x^T SB\Delta(z)x + 2x^T SBKx$$

As

$$- x^T K^T Kx - 2x^T K^T \Delta(z)x = -x^T \left(K + \Delta(z)\right)^T \left(K + \Delta(z)\right)x + x^T \Delta(z)^T \Delta(z)x \leq x^T \Delta(z)^T \Delta(z)x$$

we can obtain

$$\begin{aligned} \dot{V}(x) =& - x^T Q x + x^T K^T Kx - x^T \Delta(z)^T Kx + x^T SB\Delta(z)x + 2x^T SBKx \\ =& - x^T Q x - x^T K^T Kx - 2x^T \Delta(z)^T Kx \\ =& - x^T Q x - x^T \left(K + \Delta(z)\right)^T \left(K + \Delta(z)\right)x + x^T \Delta(z)^T \Delta(z)x \\ \leq& - x^T Q x + x^T \Delta(z)^T \Delta(z)x \\ \leq& - x^T (Q - F)x \\ \leq& - x^T x \end{aligned}$$

Therefore,

$$\begin{aligned} \dot{V}(x) &= 0 \quad x = 0 \\ \dot{V}(x) &\leq 0 \quad x \neq 0 \end{aligned}$$

According to the Lyapunov stability theorem, the uncertain system (11) is asymptotically stable. Theorem 1 has thus been proved. □

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

4. RL Algorithm for Robust Optimal Control

In this section, we propose an RL algorithm for solving the robust control problem of an IPS through online PI. According to ARE (18), the following optimal control problem is constructed. For the nominal system,

$$\dot{x} = A(z_0)x + Bu$$

we find a control u, such that the following performance index reaches a minimum.

$$J = \int_t^{\infty} [x^T Q x + u^T u] dt \tag{20}$$

where $Q = F + I > 0$. For any initial time t, the optimal cost can be written as

$$\begin{aligned} V[x(t)] &= \int_t^{\infty} \left(x^T Q x + u^T u \right) dt \\ &= \int_t^{t+\Delta t} \left(x^T Q x + u^T u \right) dt + \int_{t+\Delta t}^{\infty} \left(x^T Q x + u^T u \right) dt \\ &= \int_t^{t+\Delta t} \left(x^T Q x + u^T u \right) dt + V[x(t+\Delta t)] \end{aligned}$$

From Lyapunov function (19), we obtain

$$x(t)^T S x(t) = \int_t^{t+\Delta t} \left(x^T Q x + u^T u \right) dt + x(t+\Delta t)^T S x(t+\Delta t) \tag{21}$$

where S is the solution to ARE (18). We propose the following RL algorithm for solving a robust controller.

Algorithm 1. *RL algorithm for robust control of uncertain linear IPS*

(1) $Q = F + I$ is computed.
(2) An initial stabilization control gain K_0 is selected.
(3) Policy evaluation: S_i is solved using the equation $x^T(t) S_i x(t) = \int_t^{t+\Delta t} x^T (Q + K_i^T K_i) x dt + x^T (t + \Delta t) S_i x(t + \Delta t)$.
(4) Policy improvement: $K_{i+1} = -B^T S_i$.
(5) We set $i = i + 1$, and steps 3 and 4 are repeated until $\|S_{i+1} - S_i\| \leq \epsilon$, where $\epsilon > 0$ is a small constant.

In Algorithm 1, by providing an initial stabilizing control law, repeated iterations are performed between steps 3 and 4 until convergence. We can then obtain the robust control gain K of system (11).

Remark 1. *Step 3 in Algorithm 1 is the policy evaluation, and step 4 is the policy improvement. Equivalently, the solving of the equation in step 3 is actually solving a least squares problem. In the integral interval, if sufficient data are obtained in the system, the least square method can be used to solve S_i. Although it is also possible to directly solve the ARE (18) to obtain S_i, the state matrix of the system (12) is required to be known. The implementation of Algorithm 1 does not necessitate that the state matrix of the system be known.*

Next, we prove the convergence of Algorithm 1. However, it is necessary to prove the following Lemma first.

Lemma 4. *On assuming that the matrix $A(z_0) + BK_i$ is stable, solving the matrix S_i from step 3 of Algorithm 1 becomes equivalent to solving the following equation.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

$$S(A(z_0)+BK_i)+(A(z_0)+BK_i)^TS+Q+K_i^TK_i=0 \tag{22}$$

Proof. We rewrite the equation of step 3 in Algorithm 1 as follows

$$\lim_{\Delta t\to 0}\frac{x^T(t+\Delta t)S_ix(t+\Delta t)-x^T(t)S_ix(t)}{\Delta t}+\lim_{\Delta t\to 0}\frac{\int_t^{t+\Delta t}x^T(Q+K_i^TK_i)xdt}{\Delta t}=0 \tag{23}$$

According to the definition of the derivative, it can be observed that the first term of equation (23) is the derivative of $x^T(t)S_ix(t)$ with respect to time t. We thus obtain

$$\frac{d(x^T(t)S_ix(t))}{dt}+\lim_{\Delta t\to 0}\frac{d}{d\Delta t}\int_t^{t+\Delta t}x^T(Q+K_i^TK_i)xdt=0 \tag{24}$$

Further re-arranging equation (24) yields

$$x^T[S(A(z_0)+BK_i)+(A(z_0)+BK_i)^TS+Q+K_i^TK_i]x=0 \tag{25}$$

which means that (22) is established. Next, we reverse the process.

Along the stable system $\dot{x}=(A+BK_i)x$, the time derivative of the Lyapunov function $V_i(x)=x^TS_ix$ is calculated. We can then obtain

$$\dot{V}_i(x)=\frac{dx^T(t)S_ix(t)}{dt}=x^T(A(z_0)+BK_i)^TS_ix+x^TS_i(A(z_0)+BK_i)x$$

On integrating both sides of the above equation in the interval $(t,\Delta t)$, we obtain

$$x^T(t+\Delta t)S_ix(t+\Delta t)-x^T(t)S_ix(t)=\int_t^{t+\Delta t}x^T(Q+K_i^TK_i)xdt$$

This concludes the proof.

According to the existing conclusions [42], iterative relations (22) and step 3 of Algorithm 1 converge to the solution of ARE (18).

5. Robust Control of Nonlinear IPS

In this section, we establish a nonlinear state-space model of the IPS and construct an appropriate auxiliary system and a performance index. The problem of the robust control of the IPS is then transformed into the optimal control problem of the auxiliary system. We finally propose the corresponding RL algorithm.

5.1. Nonlinear State-Space Representation of IPS

Based on the uncertain linear inverted pendulum model (11) established in Section 2.1, we consider the following uncertain nonlinear system.

$$\dot{x}=A(z)x(t)+Bu(t)+F(z,x) \tag{26}$$

where $F(z,x)$ represents the nonlinear perturbation of the system and can be used to represent various nonlinearity factors in the system. Based on the modeling process in Section 2 and literature [23], it is assumed that

$$F(z,x)=\begin{bmatrix}0\\ \frac{-(I+m_1L^2)z}{W}(cos(x_1x_2+x_3x_4)+\frac{0.5x_1+2x_3-4x_4}{x_2}-1)x_2\\ 0\\ \frac{-m_1Lz}{W}(cos(x_1x_2+x_3x_4)+\frac{0.5x_1+2x_3-4x_4}{x_2}-1)x_2\end{bmatrix}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

where the parameters I, m_1, L, and W are the same as those in (10). On rewriting system (26), we obtain

$$\dot{x} = \bar{A}x + Bu + \bar{F}(z,x) \tag{27}$$

where

$$\bar{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{m_1^2 g L^2}{W} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{m_1 g L(m_1+m_2)}{W} & 0 \end{bmatrix}, \bar{F}(z,x) = \begin{bmatrix} 0 \\ \frac{-(I+m_1L^2)z}{W}(cos(x_1x_2+x_3x_4) + \frac{0.5x_1+2x_3-4x_4}{x_2})x_2 \\ 0 \\ \frac{-m_1Lz}{W}(cos(x_1x_2+x_3x_4) + \frac{0.5x_1+2x_3-4x_4}{x_2})x_2 \end{bmatrix}$$

On substituting the parameter values into system (27), we can obtain

$$\bar{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0.6293 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 27.8285 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0.8832 \\ 0 \\ 2.3566 \end{bmatrix}, \bar{F}(z,x) = \begin{bmatrix} 0 \\ -0.8832zx_2 \\ 0 \\ -2.3566zx_2 \end{bmatrix} [cos(x_1x_2+x_3x_4) + \frac{0.5x_1+2x_3-4x_4}{x_2}]$$

5.2. Robust Control of Nonlinear IPS

To obtain the robust control law for an uncertain nonlinear IPS, we propose the following two lemmas.

Lemma 5. *There exists an uncertain function $G(z,x)$ such that $\bar{F}(z,x)$ can be decomposed into the following form.*

$$\bar{F}(z,x) = BG(z,x)$$

Proof.

$$\begin{aligned} \bar{F}(z,x) &= \begin{bmatrix} 0 \\ \frac{-(I+m_1L^2)z}{N}x_2 \\ 0 \\ \frac{-m_1Lz}{N}x_2 \end{bmatrix} [cos(x_1x_2+x_3x_4) + \frac{0.5x_1+2x_3-4x_4}{x_2}] \\ &= \begin{bmatrix} 0 \\ \frac{I+m_1L^2}{N} \\ 0 \\ \frac{m_1L}{N} \end{bmatrix} (-zx_2[cos(x_1x_2+x_3x_4) + \frac{0.5x_1+2x_3-4x_4}{x_2}]) = BG(z,x) \end{aligned}$$

where $G(z,x) = -zx_2[cos(x_1x_2+x_3x_4) + \frac{0.5x_1+2x_3-4x_4}{x_2}]$. Proof completed. □

Lemma 6. *There exists an upper bound function $f_{max}(x)$ such that $G(z,x)$ satisfies*

$$\mid G(z,x \mid \leq f_{max}(x) \tag{28}$$

Proof.

$$\begin{aligned} \mid G(z,x) \mid &= \mid -zx_2[cos(x_1x_2+x_3x_4) + \frac{0.5x_1+2x_3-4x_4}{x_2}] \mid \\ &= \mid zx_2cos(x_1x_2+x_3x_4) + z(0.5x_1+2x_3-4x_4) \mid \\ &\leq \mid x_2cos(x_1x_2+x_3x_4) + z(0.5x_1+2x_3-4x_4) \mid \\ &\leq \mid x_2+0.5x_1+2x_3-4x_4 \mid \\ &= f_{max}(x) \end{aligned}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

This concludes the proof. □

We constructing the optimal control problems for a nominal system.

$$\dot{x} = \bar{A}x(t) + Bu(t) \tag{29}$$

We determine a controller u such that $u = u(x)$ to minimize the following performance index.

$$J(x_0, u) = \int_0^\infty [f_{max}^2(x) + x^T x + u^T u]dt \tag{30}$$

According to the performance index function (30), the cost function corresponding to the admissible control policy $u(x)$ is

$$V(x) = \int_t^\infty [f_{max}^2(x) + x^T x + u^T u]dt \tag{31}$$

Taking the time derivative on both sides of function (31) yields the following Bellman equation.

$$f_{max}^2(x) + x^T x + u^T u + \nabla V^T[\bar{A}x + Bu] = 0 \tag{32}$$

where ∇V is the gradient of the cost function $V(x)$ with respect to x.

We define the following Hamiltonian function.

$$H(x, u, \nabla V) = f_{max}^2(x) + x^T x + u^T u + \nabla V^T[\bar{A}x + Bu] \tag{33}$$

On assuming that the minimum exists and is unique, the optimal control function for the given problem is then obtained as

$$u^* = -\frac{1}{2}B^T \nabla V^* \tag{34}$$

On substituting equation (34) into equation (32), the HJB equation that satisfies the optimal function $V^*(x)$ can be obtained as

$$f_{max}^2(x) + x^T x + \nabla V^{*T}\bar{A}x - \frac{1}{4}\nabla V^{*T} B B^T \nabla V^* = 0 \tag{35}$$

with the initial condition $V^*(0) = 0$.

On solving the optimal function $V^*(x)$ from equation (35), the solution of the optimal control problem can be obtained. The solution of the robust control problem can then be obtained. The following theorem shows that the optimal control $u^* = -\frac{1}{2}B^T \nabla V^*$ is a robust controller for a nonlinear IPS.

Theorem 2. On considering the nominal system (29) with the performance index (30) and assuming that solution $V^*(x)$ of the HJB equation (35) exists, the optimal control law (34) can then globally stabilize the IPS (27).

Proof of Theorem 2. We select $V^*(x)$ as the Lyapunov function. On considering the performance index function (30), $V^*(x)$ is obviously positive definite, and $V^*(0) = 0$. Taking the time derivative of $V^*(x)$ along system (27) yields

$$\frac{dV^*}{dt} = \nabla V^{*T}[\bar{A}x + F(z, x)] - \frac{1}{2}\nabla V^{*T} B B^T \nabla V$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

According to Lemma 5, it follows that

$$\frac{dV^*}{dt} = \nabla V^{*T}\bar{A}x + \nabla V^{*T}BG(z,x) - \frac{1}{2}\nabla V^{*T}BB^T\nabla V \tag{36}$$

According to HJB equation (35), we can obtain

$$\nabla V^{*T}\bar{A}x = -f_{max}^2(x) - x^Tx + \frac{1}{4}\nabla V^{*T}BB^T\nabla V^* \tag{37}$$

On substituting equation (37) into equation (36), we obtain

$$\begin{aligned}\frac{dV^*}{dt} &= -f_{max}^2(x) - x^Tx + \frac{1}{4}\nabla V^{*T}BB^T\nabla V^* + \nabla V^{*T}BG(z,x) - \frac{1}{2}\nabla V^{*T}BB^T\nabla V \\ &= -f_{max}^2(x) - x^Tx - \frac{1}{4}\nabla V^{*T}BB^T\nabla V^* + \nabla V^{*T}BG(z,x)\end{aligned} \tag{38}$$

From equation (38), we can obtain

$$\begin{aligned}\frac{dV^*}{dt} &= -f_{max}^2(x) - x^Tx - \frac{1}{4}[\nabla V^{*T}BB^T\nabla V^* - 4\nabla V^{*T}BG(z,x) + 4G^T(z,x)G(z,x)] + G^T(z,x)G(z,x) \\ &= -x^Tx + G^T(z,x)G(z,x) - f_{max}^2(x) - \frac{1}{4}H^T(z,x)H(z,x) \\ &\leq -x^Tx\end{aligned} \tag{39}$$

where $H(z,x) = B^T\nabla V^* - 2G(z,x)$. According to the Lyapunov stability criterion, the optimal control (34) can make the uncertain nonlinear IPS (27) asymptotically stable for all the allowable uncertainties. Therefore, for a constant $p > 0$, there is a neighborhood $\mathcal{N} = \{x : \|x\| < p\}$ of the origin, such that, if the state $x(t)$ enters the neighborhood $\mathcal{N}$, then $x \to 0$ when $t \to \infty$. However, $x(t)$ cannot remain outside the domain $\mathcal{N}$ forever; else, $\|x(t)\| \geq p$ for all $t > 0$, which implies that

$$\begin{aligned}V^*[x(t)] - V^*[x(0)] &= \int_0^t \dot{V}^*(x(\tau))d\tau \\ &\leq \int_0^t -x^Txd\tau \\ &\leq \int_0^\tau -p^2d\tau \\ &= -p^2t\end{aligned}$$

Let $t \to \infty$, then

$$V^*[x(t)] \leq V^*[x(0)] - p^2t \to -\infty$$

5.3. RL Algorithm for Nonlinear IPS

For a nonlinear IPS, we consider the optimal control problems (29) and (30). For any admissible control, the cost function corresponding to the optimal control problem can be expressed as

$$\begin{aligned}V[x(t)] &= \int_t^\infty [f_{max}^2(x) + x^Tx + u^Tu]dt \\ &= \int_t^{t+T} [f_{max}^2(x) + x^Tx + u^Tu]dt + \int_{t+T}^\infty [f_{max}^2(x) + x^Tx + u^Tu]dt\end{aligned}$$

where $T > 0$ is an arbitrarily selected constant. We can then obtain the integral reinforcement relation satisfied by the cost function

$$V[x(t)] = \int_t^{t+T} [f_{max}^2(x) + x^Tx + u^Tu]dt + V[x(t+T)] \tag{40}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

Based on the integral-based reinforcement relations (40) and the optimal control (34), the RL algorithm for the robust control of the nonlinear IPS is given below.

Algorithm 2. *RL algorithm for robust control of uncertain nonlinear IPS*

(1) A non-negative function $f_{max}(x)$ is selected.
(2) An initial stabilization control law $u_0(x)$ is selected.
(3) Policy evaluation: the $V_i(x)$ from $V_i[x(t)] = \int_t^{t+T}[f_{max}^2(x) + x^T x + u_i^T(x)u_i(x)]dt + V_i[x(t+T)]$ is solved.
(4) Policy improvement: $u_{i+1}(x) = -\frac{1}{2}B^T\nabla V_i$.
(5) $i = i+1$ is set and steps 3 and 4 are repeated until $\|V_{i+1} - V_i\| \leq \epsilon$, where $\epsilon > 0$ is a small constant.

In Algorithm 2, by providing an initial stabilizing control law, the algorithm iterates repeatedly between steps 3 and 4 until convergence. We can then obtain the robust control gain u of system (27).

Remark 2. *Although it is possible to directly solve the ARE to obtain S_i, the state matrix of the system (12) is required to be known. The implementation of Algorithm 2 does not necessitate that the state matrix of the system be known.*

Next, we prove the convergence of Algorithm 2. The following conclusion provides an equivalent form of the integral strengthening relation in step 3.

Lemma 7. *On assuming that $u_i(x)$ is the stabilization control function of the nominal system (29), solving the value function $V_i(x)$ from the equation in step 3 in Algorithm 2 is equivalent to solving the following equation.*

$$f_{max}^2(x) + x^T x + u_i^T(x)u_i(x) + \nabla V_i[\bar{A}x + Bu_i] = 0 \tag{41}$$

Proof. On dividing both sides of the equation in step 3 by T and taking the limit, we obtain

$$\lim_{T\to 0}\frac{V_i x(t+T) - V_i x(t)}{T} + \lim_{T\to 0}\frac{\int_t^{t+T}[f_{max}^2(x) + x^T x + u_i^T(x)u_i(x)]dt}{T} = 0$$

Based on the definition of the function limit and L' Hopital's rule, we obtain

$$\frac{dV_i x(t)}{dt} + \lim_{T\to 0}\frac{d}{dT}\int_t^{t+T}[f_{max}^2(x) + x^T x + u_i^T(x)u_i(x)]dt = 0$$

Therefore,

$$f_{max}^2(x) + x^T x + u_i^T(x)u_i(x) + \nabla V_i[\bar{A}x + Bu_i] = 0$$

However, along the stable system $\dot{x} = \bar{A}x + Bu_i$, taking the time derivative of $V_i(x)$ yields

$$\frac{d}{dt}(V_i(x)) = \nabla V_i(\bar{A}x + Bu_i)$$

On integrating both sides of the above equation from t to $t+T$, we obtain

$$V_i[x(t+T)] - V_i[x(t)] = \int_t^{t+T}\nabla V_i(\bar{A}x + Bu_i)d\tau$$

Then, from (41), we obtain

$$V_i[x(t)] = \int_t^{t+T} f_{max}^2(x) + x^T x + u_i^T(x)u_i(x)d\tau + V_i[x(t+T)]$$

The above equation is the equation in the third step of Algorithm 2. This concludes the proof. □

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

According to the conclusion of [39,43], if the stabilizing initial control policy $u_0(x)$ is given, the follow-up control policy calculated using the iterative relation (34) and equation (41) is also a stabilizing control policy. Furthermore, the iteratively calculated cost function sequence converges to the optimal value function. From Lemma 7, we know that equation (41) and the equation of step 3 are equivalent. Therefore, the iterative relationship between steps 3 and 4 in Algorithm 2 converges on the optimal control and optimal cost functions. This contradicts the positive definiteness of $V^*[x(t)]$. Therefore, the system (27) is globally asymptotically stable. This concludes the proof. □

6. Numerical Simulation Results

In this section, two simulation examples are provided to demonstrate the feasibility of the theoretical results for the robust control of the uncertain IPS.

6.1. Example 1

Considering system (11), our objective is to obtain a robust control u such that it is stable. Based on Lemmas 1, 2, and 3, the weighting matrix Q is selected as

$$Q = F + I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1.81 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We present the initial stability control law

$$u_0 = [1.0900 \quad 4.1230 \quad -24.8908 \quad -6.7726]x$$

The initial state of the nominal system is selected as $x_0 = [0 \quad 1 \quad 1 \quad 1]^T$. The time step size for the collecting system status and input information is set as 0.01 s. Algorithm 1 converges after six iterations, and the S_d matrix and control gain K_d converge to the following optimal solutions:

$$S_d = \begin{bmatrix} 2.4465 & 2.0822 & -6.2489 & -1.2066 \\ 2.0822 & 4.3346 & -14.0702 & -2.7082 \\ -6.2489 & -14.0702 & 100.8262 & 18.9646 \\ -1.2066 & -2.7082 & 18.9646 & 3.6646 \end{bmatrix} \tag{42}$$

and

$$K_d = [1.0044 \quad 2.5538 \quad -32.2652 \quad -6.2440] \tag{43}$$

There are 10 independent numerical samples in the matrix S_d. These 10 numerical samples are collected in each iteration to address with the least squares problem. The evolution of the control signal u is presented in Figure 5. Figure 6 illustrates the iterative convergence process of the S matrix, where $S(i, j)$ represents the element lying at the intersection of the i-th row and the j-th column in the symmetric matrix S, where $i = 1, 2, 3, 4$, $j = 1, 2, 3, 4$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

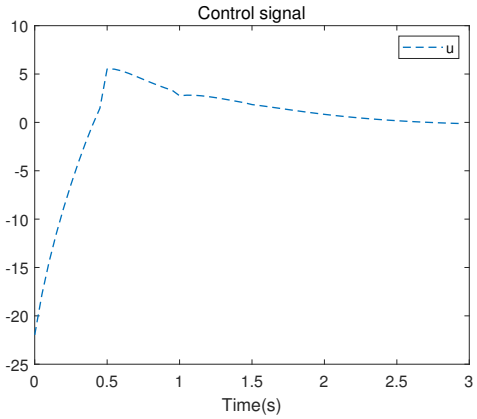

Figure 5. Control signal u of the linearized system

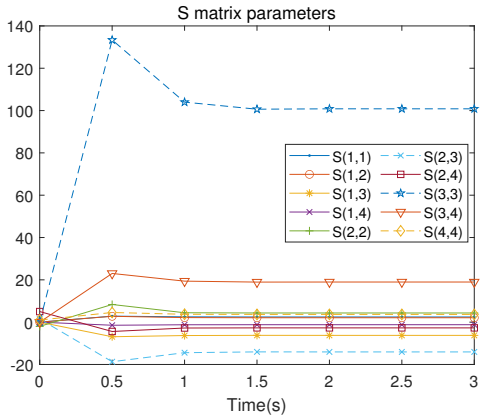

Figure 6. S-matrix iterative process of the linearized system

Using MATLAB to directly solve the ARE (18), we obtain the following optimal feedback gain and the S matrix. To avoid confusion, we use the following notations.

$$S = \begin{bmatrix} 2.4455 & 2.0802 & -6.2326 & -1.2039 \\ 2.0802 & 4.3307 & -14.0381 & -2.7032 \\ -6.2326 & -14.0381 & 100.5677 & 18.9228 \\ -1.2039 & -2.7032 & 18.9228 & 3.6579 \end{bmatrix} \tag{44}$$

$$K = \begin{bmatrix} 1.000 & 2.5455 & -32.1952 & -6.2327 \end{bmatrix} \tag{45}$$

As is apparent, the results obtained using the RL method are only marginally different from those obtained via the direct solution of the ARE. Figure 7 presents the closed-loop trajectory of the system when $z = 0.1, 0.4, 0.7, 1.0$. It is easy to observe that the system is stable, which means that the controller is valid.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

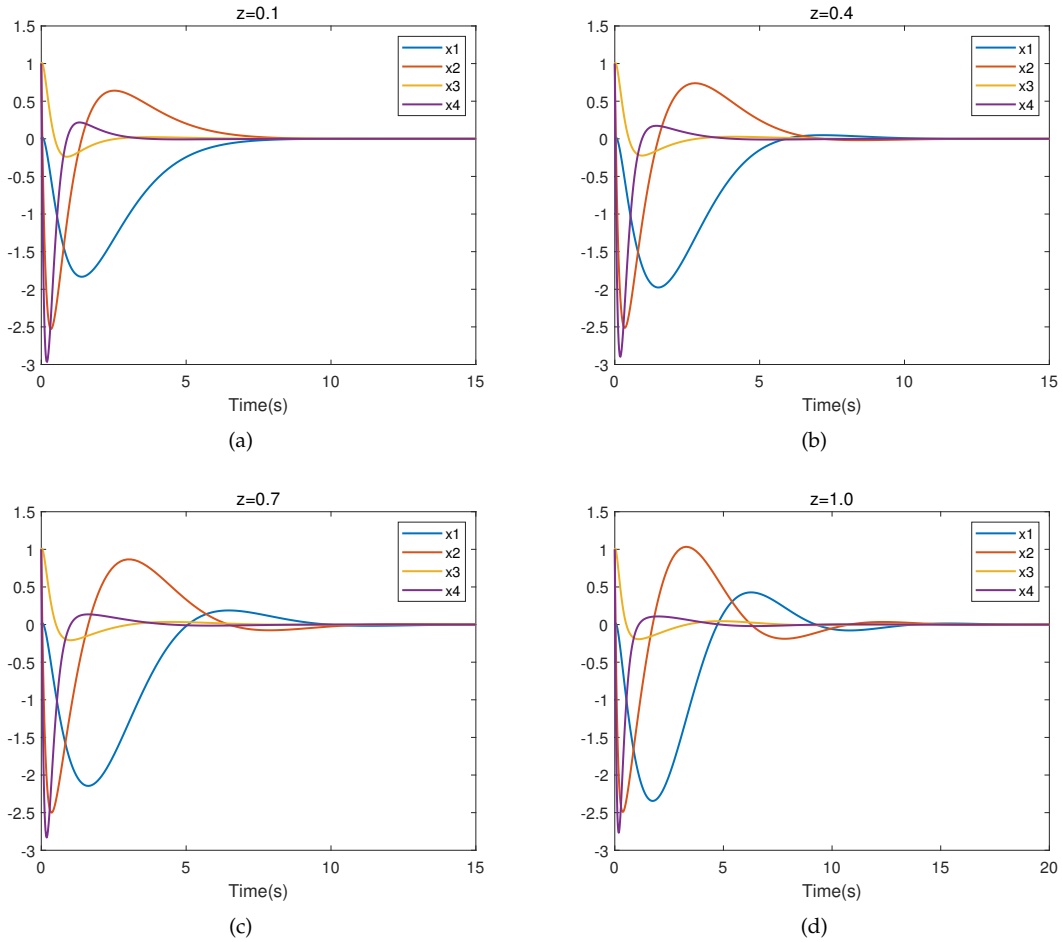

Figure 7. Trajectory of closed-loop linearized system

The corresponding partial eigenvalues of the closed-loop uncertain linear system with $u = Kx$ for different z are listed in Table 2. From Table 2, we can observe that the eigenvalues of the closed-loop system all have negative real parts. Thus, the uncertain linear system (11) with robust control $u = Kx$ is asymptotically stable for all $0 \leq z \leq 1$.

Table 2. Characteristic root of system (11) when z takes different values.

z	$\lambda 1$	$\lambda 2$	$\lambda 3$	$\lambda 4$
0.1	-6.60	-4.23	-0.85+0.32i	-0.85-0.32i
0.2	-6.73	-4.33	-0.78+0.43i	-0.78-0.43i
0.3	-6.86	-4.41	-0.71+0.50i	-0.71-0.50i
0.4	-7.00	-4.48	-0.65+0.55i	-0.65-0.55i
0.5	-7.14	-4.54	-0.60+0.59i	-0.60-0.59i
0.6	-7.28	-4.59	-0.55+0.62i	-0.55-0.62i
0.7	-7.42	-4.63	-0.50+0.65i	-0.50-0.65i
0.8	-7.56	-4.67	-0.46+0.67i	-0.46-0.67i
0.9	-7.70	-4.70	-0.42+0.68i	-0.42-0.68i
1.0	-7.84	-4.73	-0.38+0.69i	-0.38-0.69i

6.2. Example 2

Let us consider the nonlinear IPS (27). According to Lemma 5, the system (27) can be rewritten as

$$\dot{x} = \bar{A}x + Bu + BG(z, x) \tag{46}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

The optimal control problem for the IPS is as follows: for nominal system (29), we find the control function u such that the performance index (30) achieves a minimum.

According to lemma 6, we obtain

$$\mid G(z,x) \mid = \mid -zx_2[cos(x_1x_2 + x_3x_4) + \frac{0.5x_1 + 2x_3 - 4x_4}{x_2}] \mid \leq |x_2 + 0.5x_1 + 2x_3 - 4x_4| = f_{max}(x)$$

then

$$f_{max}^2(x) = (x_2 + 0.5x_1 + 2x_3 - 4x_4)^2 = x^T \begin{bmatrix} 0.25 & 0.5 & 1 & -2 \\ 0.5 & 1 & 2 & -4 \\ 1 & 2 & 4 & -8 \\ -2 & -4 & -8 & 16 \end{bmatrix} x$$

According to performance index (30), the weight matrix Q is selected as

$$Q = \begin{bmatrix} 1.25 & 0.5 & 1 & -2 \\ 0.5 & 2 & 2 & -4 \\ 1 & 2 & 5 & -8 \\ -2 & -4 & -8 & 17 \end{bmatrix}$$

Based on Algorithm 2, present give the initial control policy

$$u_0 = [1.0900 \quad 4.1230 \quad -24.8908 \quad -6.7726]x$$

The initial state of the system is selected as $x_0 = [0 \quad 1 \quad 1 \quad 1]^T$. Algorithm 2 converges after six iterations, and the S_d matrix and control gain K_d converge to the following optimal solutions.

$$S_d = \begin{bmatrix} 2.4325 & 2.4398 & -6.2874 & -1.3888 \\ 2.4398 & 5.0469 & -14.0539 & -3.0045 \\ -6.2874 & -14.0539 & 130.4535 & 18.9735 \\ -1.3888 & -3.0045 & 18.9735 & 4.2715 \end{bmatrix} \tag{47}$$

and

$$K_d = [1.1180 \quad 2.6229 \quad -32.3006 \quad -7.4126] \tag{48}$$

The evolution of the control signal u is presented in Figure 8. Figure 9 presents the convergence process of the S_d matrix.

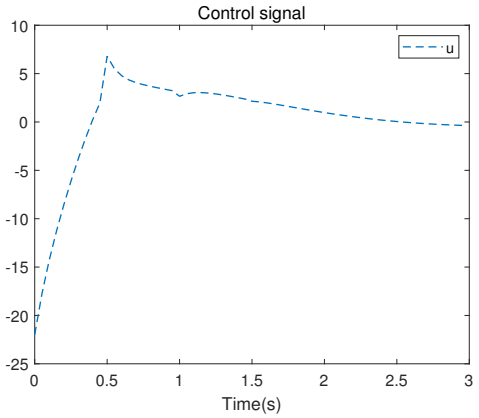

Figure 8. Control signal u of the nonlinear system

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2023 doi:10.20944/preprints202310.1100.v1

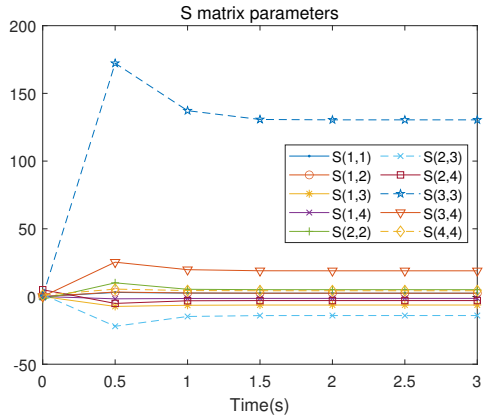

Figure 9. S-matrix iterative process of the nonlinear system

We also selected $z = 0.1, 0.4, 0.7, 1.0$. Figure 10 presents the trajectory of the closed-loop system for different values of z.

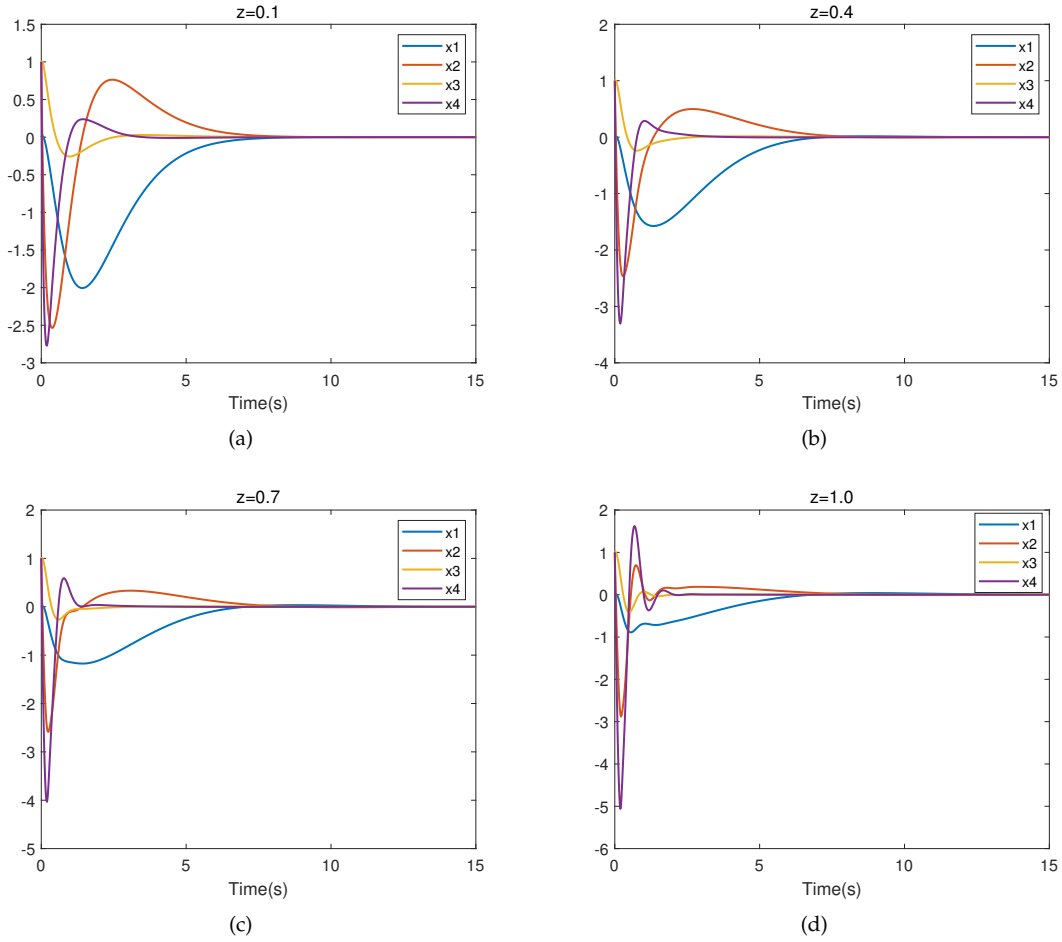

Figure 10. Trajectory of closed-loop nonlinear system

7. Conclusions

In this paper, the robust control problem of the first-order IPS is studied. The linearization and nonlinear state-space representation are established, and an RL algorithm for the robust control of the IPS is proposed. The controller of the uncertain system is obtained using the method of online PI. The results thus obtained show that the error between the controller obtained using the RL algorithm

and by directly solving ARE is very small. Moreover, the use of the RL algorithm can effectively circumvent the "curse of dimensionality." Moreover, the algorithm can provide a controller that meets the requirements without the nominal matrix of the system being known. This improves the current state at which the robust control of the IPS relies excessively on the nominal matrix. In future research, we intend to take into consideration that the input matrix of the system also has uncertainty and extend the RL algorithm to more general systems.

Author Contributions: All the authors contributed equally to the development of the research.

Funding: This research was funded by the National Natural Science Foundation of China under Grant No. 61463002.

Acknowledgments: The authors thank to the Journal editors and the reviewers for their helpful suggestions and comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dai L. Strong decoupling in singular systems. Mathematical Systems Theory. 1989,22(1):275–2
2. Pal B, Chaudhuri B. Robust Control in Power Systems. Springer Science & Business Media
3. Liu L, Liu Y J, Tong S, et al. Relative threshold-based event-triggered control for nonlinear constrained systems with application to aircraft wing rock motion. IEEE Transactions on Industrial Informatics. 2022,18(2):911–921.
4. Wang Z, She J, Wang F, et al. Further Result on Reducing Disturbance-Compensation Error of Equivalent-Input-Disturbance Approach. IEEE/ASME Transactions on Mechatronics. IEEE/ASME Transactions on Mechatronics (Early Access), doi: 10.1109/TMECH.2023.3303029.
5. Yue D, Han QL, Lam J. Network-based robust H control of systems with uncertainty. Automatica. 2005,41(6):999–1007.
6. Grasser F, D'arrigo A, Colombi S, Rufer AC. JOE: a mobile, inverted pendulum. IEEE Transactions on industrial electronics. 2002,49(1):107–114.
7. Stephenson A. Memoirs and Proceedings of the Manchester Literary and Philosophical Society 52, 1–10. On a new type of dynamic stability. 1908.
8. Housner GW. The behavior of inverted pendulum structures during earthquakes. Bulletin of the seismological society of America. 1963,53(2):403–417.
9. Wang JJ. Simulation studies of inverted pendulum based on PID controllers. Simulation Modelling Practice and Theory. 2011,19(1):440–449.
10. Prasad LB, Tyagi B, Gupta HO. Optimal Control of Nonlinear Inverted Pendulum System Using PID Controller and LQR: Performance Analysis Without and With Disturbance Input. International Journal of Automation and Computing. 2014,11:661–670.
11. Li D, Chen H, Fan J, Shen C. A novel qualitative control method to inverted pendulum systems. IFAC Proceedings Volumes. 1999,32(2):1495–1500.
12. Kwon T, Hodgins JK. Momentum-Mapped Inverted Pendulum Models for Controlling Dynamic Human Motions. ACM Transactions on Graphics (TOG). 2017,36(1):1–14.
13. Yamakawa T. Stabilization of an inverted pendulum by a high-speed fuzzy logic controller hardware system. Fuzzy sets and Systems. 1989,32(2):161–180.
14. Huang CH, Wang WJ, Chiu CH. Design and implementation of fuzzy control on a two-wheel inverted pendulum. IEEE Transactions on Industrial Electronics. 2010,58(7):2988–3001.
15. Su X, Xia F, Liu J, Wu L. Event-triggered fuzzy control of nonlinear systems with its application to inverted pendulum systems. Automatica. 2018,94:236–248.
16. Nasir ANK, Razak AAA. Opposition-based spiral dynamic algorithm with an application to optimize type-2 fuzzy control for an inverted pendulum system. Expert Systems with Applications. 2022,195:116661.
17. Wai RJ, Chang LJ. Adaptive stabilizing and tracking control for a nonlinear inverted-pendulum system via sliding-mode technique. IEEE Transactions on Industrial Electronics. 2006,53(2):674–692.
18. Huang J, Guan ZH, Matsuno T, Fukuda T, Sekiyama K. Sliding-mode velocity control of mobile-wheeled inverted-pendulum systems. IEEE Transactions on robotics. 2010,26(4):750–758.

19. Jung S, Kim SS. Control experiment of a wheel-driven mobile inverted pendulum using neural network. IEEE Transactions on Control Systems Technology. 2008,16(2):297–303.
20. Yang C, Li Z, Li J. Trajectory planning and optimized adaptive control for a class of wheeled inverted pendulum vehicle models. IEEE Transactions on Cybernetics. 2012,43(1):24–36.
21. Yang C, Li Z, Cui R, Xu B. Neural network-based motion control of an underactuated wheeled inverted pendulum model. IEEE Transactions on Neural Networks and Learning Systems. 2014,25(11):2004–2016.
22. Feng Lin and A. W. Olbrot, "An LQR approach to robust control of linear systems with uncertain parameters," Proceedings of 35th IEEE Conference on Decision and Control, Kobe, Japan, 1996, pp. 4158-4163 vol.4, doi: 10.1109/CDC.1996.577433.
23. Lin F, Brandt RD. An optimal control approach to robust control of robot manipulators. IEEE Transactions on robotics and automation. 1998,14(1):69–77.
24. Lin F. An optimal control approach to robust control design. International journal of control. 2000,73(3):177–186.
25. Bellman R. Dynamic programming. Science. 1966,153(3731):34–37.
26. Neustadt LW, Pontrjagin LS, Trirogoff K. The mathematical theory of optimal processes. Interscience, 1962.
27. Powell WB. Approximate Dynamic Programming: Solving the curses of dimensionality. 703. John Wiley & Sons, 2007.
28. Li H, Liu D. Optimal control for discrete-time affine non-linear systems using general value iteration. IET Control Theory & Applications. 2012,6(18):2725–2736.
29. Wei Q, Liu D, Lin H. Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems. IEEE Transactions on cybernetics. 2015,46(3):840–853.
30. Tesauro G. TD-Gammon, a self-teaching backgammon program, achieves master-level play. Neural computation. 1994,6(2):215–219.
31. Crites R, Barto A. Improving elevator performance using reinforcement learning. Advances in neural information processing systems. 1995,8.
32. Zhang W, Dietterich T. High-performance job-shop scheduling with a time-delay TD (λ) network. Advances in neural information processing systems. 1995,8.
33. Singh S, Bertsekas D. Reinforcement learning for dynamic channel allocation in cellular telephone systems. Advances in neural information processing systems. 1996,9.
34. Maja J M. Reward Functions for Accelerated Learning. In: Cohen WW, Hirsh H. , eds. Machine learning proceedings 1994, 1 ed., Morgan Kaufmann, 1994,181–189.
35. Doya K. Reinforcement learning in continuous time and space. Neural computation. 2000,12(1):219–245.
36. Krstic M, Kokotovic PV, Kanellakopoulos I. Nonlinear and adaptive control design. John Wiley & Sons, Inc., 1995.
37. Ioannou P, Fidan B. Adaptive Control Tutorial, vol. 11 of Advances in Design and Control. SIAM, Philadelphia, Pa, USA,. 2006.
38. Åström KJ, Wittenmark B. Adaptive control. Courier Corporation, 2013.
39. Vrabie D, Pastravanu O, Abu-Khalaf M, Lewis FL. Adaptive optimal control for continuous-time linear systems based on policy iteration. Automatica. 2009,45(2):477–484.
40. Xu D, Wang Q, Li Y. Adaptive optimal control approach to robust tracking of uncertain linear systems based on policy iteration. Measurement and Control. 2021, 54(5-6): 668-680.
41. Xu D, Wang Q, Li Y. Optimal guaranteed cost tracking of uncertain nonlinear systems using adaptive dynamic programming with concurrent learning. International Journal of Control, Automation and Systems. 2020,18(5):1116–1127.
42. Kleinman D. On an iterative technique for Riccati equation computations. IEEE Transactions on Automatic Control. 1968,13(1):114–115.
43. Abu-Khalaf M, Lewis FL. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. Automatica. 2005,41(5):779–791.