applied sciences

MDPI

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 545616, China; mayan_yyjj@163.com (Y.M.); 18707519318@163.com (J.H.); lyh7ooo@163.com (Y.L.)
[2] School of Physics and Electrical Engineering, Liupanshui Normal University, Liupanshui 553004, China
* Correspondence: dengguoxu@163.com

Abstract: This paper is primarily focused on the robust control of an inverted pendulum system based on 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

check for updates

Citation: Ma, Y.; Xu, D.; Huang, J.; Li, Y. Robust Control of an Inverted Pendulum System Based on Policy Iteration in Reinforcement Learning. *Appl. Sci.* **2023**, *13*, 13181. https://doi.org/10.3390/app132413181

Academic Editor: Alessandro Gasparetto

Received: 12 October 2023
Revised: 26 November 2023
Accepted: 29 November 2023
Published: 12 December 2023

1. Introduction

In the last decade, there has been increased 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 electronics, flight control, motion control, network control, and IPSs, in addition to other fields [1,2]. Research on the robust control of an IPS has provided advantageous results in recent years. An inverted pendulum is an experimental device that has 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.

Although the earliest research on IPSs can be traced back to 1908 [3], there is almost no literature on this subject between 1908 and 1960. In 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 [4]. A pendulum structure under the effect of an earthquake was modeled as a base and rigid block system, and block overturning was 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 proportional–integral–derivative control, cloud model control, fuzzy control, sliding mode control, and neural network control methods [5–7]. 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, some scholars transformed the robust control problem into an optimal control problem. In [8], the authors proposed a robust optimal control method for linear systems with matching uncertainty. However, the situation where the uncertainty does not meet matching conditions has not been considered. Lin et al. [9,10] conducted research on the robust optimal control of uncertain systems by adjusting the value of the weighting matrix and solving an algebraic Riccati equation (ARE) to obtain robust control laws. Zhang et al. [11] presented a unified framework for studying robust optimal control problems with adjustable uncertainty sets. Wang et al. [12] developed a novel adaptive critical learning approach for robust optimal control of a class of uncertain affine nonlinear systems with matching uncertainties. And the data-based adaptive critical designs were developed to solve the Hamilton–Jacobi–Bellman (HJB) equation corresponding to the transformed optimal control problem.

In fact, the pioneering methods for solving optimal control problems mainly include dynamic programming [13] and maximum principles [14]. With the dynamic programming method, solving the HJB equation yields optimal control of the system. 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 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 [15]. To overcome this weakness, some scholars have used the reinforcement learning (RL) policy to solve the optimal control problem [16,17].

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 [18], scheduling problem [19], and robot navigation problem [20]. Furthermore, the application of RL algorithms to continuous-time and continuous-state systems was initially extended by Doya et al. [21]. They used the known system model to learn the optimal control policy. In the context of control engineering, RL and adaptive dynamic programming link traditional optimal control methods to adaptive control methods [22–24]. Vrabie et al. [25] used the RL algorithm to solve the optimal control problem of the continuous time system. In the case of the linear system, system data are collected, and the solution of the HJB equation is obtained via online policy iteration (PI) using the least squares method. Xu et al. [26,27] proposed an RL algorithm based on linear continuous-time systems to solve the robust control and 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 control.

The IPS demonstrates a positive impact in the validation of RL algorithms. There are some literatures on RL to solve the control problem of inverted pendulum systems. Bates [28] harnessed GPUs to quickly train a simulation of an inverted pendulum to balance itself. Israilov et al. [29] used two model-free RL algorithms to control targets and proposed a general framework to reproduce successful experiments and simulations based on the inverted pendulum. In addition, there are still many studies of this kind, for example [30–32]. However, the results of these studies focused more on reducing the time to reach equilibrium, without fully considering uncertainty in the system. We attempt to solve the robust control problem of an uncertain IPS using RL algorithms. Only the input and output data need to be collected when using the RL control algorithm and the state matrix of the nominal system need not be known. This study lays out 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 with uncertainty is established and a robust optimal control method is applied to the IPS model. By constructing an appropriate performance index, the optimal control method is used to design a robust control law for an uncertain IPS.

(2) A PI algorithm in the RL has been designed to realize the robust optimal control of an IPS. The use of the RL algorithm to solve the robust control problem of IPS does not require knowing the state matrix, only collecting input and output data. The 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 linearizing 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.

2. Model Formulation

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 [33]. Its structure is presented in Figure 1. The encoder is a photoelectric rotary one, and the motor is a direct-current servo motor. For a detailed information on the experimental platform, see [34]. 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.

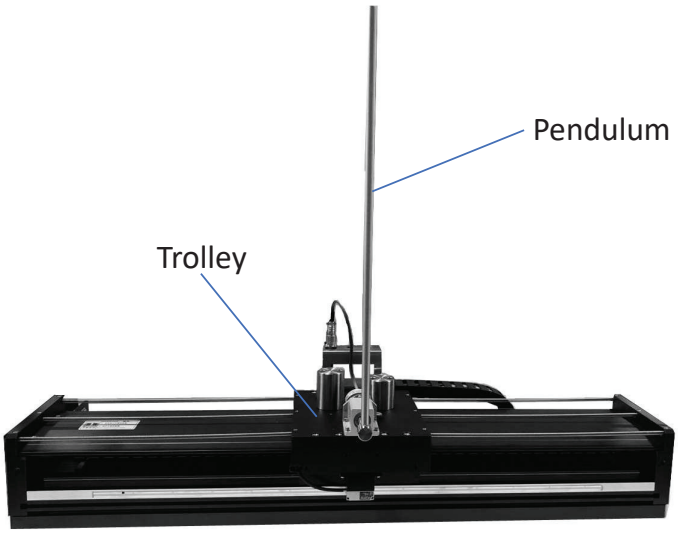

Figure 1. Inverted pendulum system diagram.

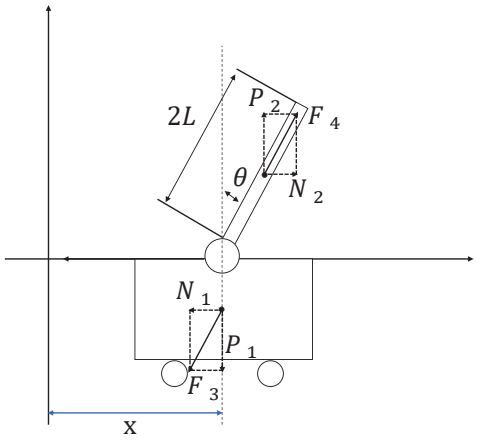

Figure 2. First-order inverted pendulum physical model.

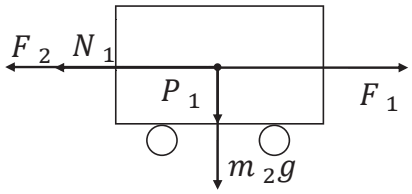

Figure 3. Force analysis of the trolley.

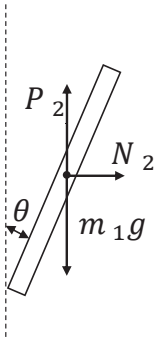

Figure 4. Force analysis of the pendulum.

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 $x(t)$. In Figure 3, according to Newton's second Law, the trolley satisfies the following equation in the horizontal direction:

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

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·m^2	Moment of inertia of pendulum

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 Figure 4, the pendulum satisfies the following equation in the horizontal direction:

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

Considering that $N_1 = N_2$ in Figure 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, in Figure 4, we analyze the resultant force in the vertical direction of the pendulum, and the following equation can be obtained.

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

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

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

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

$$I\ddot{\theta} = P_2 L\sin\theta + N_2 L\cos\theta \tag{6}$$

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

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

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),

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

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

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

Equation (9) is the linearized dynamic model of the system. The first equation comes from (1)–(3), which is the equilibrium force equation of the system in the horizontal direction. The second equation comes from (4)–(7), which is the equilibrium force equation of the system in the vertical direction. The state variables of the system can be defined as

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

Therefore, the following state-space equation can be derived.

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

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

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

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

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

Here we choose $z_0 = 0.1$ 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 the 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 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 completes 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 completes 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 completes the proof. □

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

$$SA(z_0) + A(z_0)^TS + M - SBB^TS = 0 \tag{18}$$

where $M = F + I$. 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 $u = Kx$, $K = -B^TS$ can make system (11) asymptotically stable.*

Proof. We define the following Lyapunov function.

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

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

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

According to Lemma 2, we can obtain

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

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

$$\dot{V}(x) = - x^T M x + x^T S B B^T S x + x^T \delta(z)^T B^T S x + x^T S B \delta(z) x + 2 x^T S B K x$$

because $K = -B^T S$,

$$\dot{V}(x) = - x^T M x + x^T K^T K x - x^T \delta(z)^T K x + x^T S B \delta(z) x + 2 x^T S B K x$$

As

$$- x^T K^T K x - 2 x^T K^T \delta(z) x = -x^T \big(K + \delta(z)\big)^T \big(K + \delta(z)\big) x + x^T \delta(z)^T \delta(z) x \leq x^T \delta(z)^T \delta(z) x$$

we can obtain

$$\begin{aligned} \dot{V}(x) =& - x^T M x + x^T K^T K x - x^T \delta(z)^T K x + x^T S B \delta(z) x + 2 x^T S B K x \\ =& - x^T M x - x^T K^T K x - 2 x^T \delta(z)^T K x \\ =& - x^T M x - x^T \big(K + \delta(z)\big)^T \big(K + \delta(z)\big) x + x^T \delta(z)^T \delta(z) x \\ \leq& - x^T M x + x^T \delta(z)^T \delta(z) x \\ \leq& - x^T (M - F) x \\ \leq& - x^T x \end{aligned}$$

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 [35], the uncertain system (11) is asymptotically stable. Theorem 1 has thus been proved. □

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,

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

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

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

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

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

From Lyapunov Function (19), we obtain

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

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

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).

Algorithm 1 RL Algorithm for Uncertain Linear IPS

(1) $M = F + I$ is computed.
(2) An initial stabilization control gain K_0 is selected.
(3) Policy evaluation: S_i is solved using the equation $x^T(t)S_i x(t) = \int_t^{t+\delta t} x^T(M + K_i^T K_i)xdt + x^T(t+\delta t)S_i x(t+\delta t)$.
(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.

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.*

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.*

$$S(A(z_0) + BK_i) + (A(z_0) + BK_i)^T S + M + K_i^T K_i = 0 \tag{22}$$

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

$$\lim_{\delta t \to 0} \frac{x^T(t+\delta t)S_i x(t+\delta t) - x^T(t)S_i x(t)}{\delta t} + \lim_{\delta t \to 0} \frac{\int_t^{t+\delta t} x^T(M + K_i^T K_i)xdt}{\delta t} = 0 \tag{23}$$

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

$$\frac{d(x^T(t)S_i x(t))}{dt} + \lim_{\delta t \to 0} \frac{d}{d\delta t} \int_t^{t+\delta t} x^T(M + K_i^T K_i)xdt = 0 \tag{24}$$

Further re-arranging Equation (24) yields

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

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

$$\dot{V}_i(x)=\frac{dx^T(t)S_ix(t)}{dt}=x^T(A(z_0)+BK_i)^TS_ix+x^TS_i(A(z_0)+BK_i)x \tag{26}$$

On integrating both sides of the equation (26) in the interval $(t,\delta t)$, we obtain

$$x^T(t+\delta t)S_ix(t+\delta t)-x^T(t)S_ix(t)=\int_t^{t+\delta t}x^T(M+K_t^TK_i)xdt$$

This completes the proof.

According to the existing conclusions [36], iterative relations (22) and step 3 of Algorithm 1 converge to form the solution of ARE (18). □

Remark 2. *The behavior of the control is evaluated using a cost function, which is similar to the reward in RL. The agent corresponds to the controller in optimal control, and the control process corresponds to the environmental model in RL. In control engineering, maximizing rewards is equivalent to minimizing the cost function, so the ultimate goal of the controller is to develop an optimal control policy by learning.*

5. Robust Control of Nonlinear IPS

In this section, a nonlinear state-space model of the IPS is established. Moreover, we construct a suitable auxiliary system and corresponding 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.

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

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 [8], it is assumed that

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

where the parameters I, m_1, L, and W are the same as those in (10). On rewriting system (27), we obtain

$$\dot{x}=\bar{A}x+Bu+\bar{F}(z,x) \tag{28}$$

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 (28), 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 $G(z,x) = -zx_2[cos(x_1x_2+x_3x_4) + \frac{0.5x_1+2x_3-4x_4}{x_2}]$. This completes the proof. □

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

$$\mid G(z,x \mid \leq f_{max}(x) \tag{29}$$

Proof.

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

This completes the proof. □

We construct the optimal control problems for a nominal system.

$$\dot{x} = \bar{A}x(t) + Bu(t) \tag{30}$$

We determine a controller u, which minimizes the following performance index

$$J(x_0, u) = \int_0^\infty [f_{max}^2(x) + x^T x + u^T u]dt \tag{31}$$

Based on the performance index Function (31), the cost function to the admissible control policy $u(x)$ is

$$V(x) = \int_t^\infty [f_{max}^2(x) + x^T x + u^T u]dt \tag{32}$$

We define ∇V as the gradient of $V(x)$ with respect to x. Finding differentiation on both sides of (32) with respect to t yields the following Bellman equation.

$$f_{max}^2(x) + x^T x + u^T u + \nabla V^T[\bar{A}x + Bu] = 0 \tag{33}$$

Then, we define the following Hamiltonian function.

$$H(x, u, \nabla V) = f_{max}^2(x) + x^T x + u^T u + \nabla V^T[\bar{A}x + Bu] \tag{34}$$

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

$$u_{opt} = -\frac{1}{2}B^T \nabla V_{opt} \tag{35}$$

On substituting Equation (35) into Equation (33), the HJB equation that satisfies the optimal function $V_{opt}(x)$ can be obtained as

$$f_{max}^2(x) + x^T x + \nabla V_{opt}^T \bar{A}x - \frac{1}{4}\nabla V_{opt}^T BB^T \nabla V_{opt} = 0 \tag{36}$$

with the initial condition $V_{opt}(0) = 0$.

On solving the optimal function $V_{opt}(x)$ from Equation (36), 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 $u_{opt} = -\frac{1}{2}B^T \nabla V_{opt}$ is a robust controller for a nonlinear IPS.

Theorem 2. *On considering the nominal system (30) with the performance index (31) and assuming that solution $V_{opt}(x)$ of the HJB Equation (36) exists, the optimal control law (35) can then globally stabilize the IPS (28).*

Proof. We select $V_{opt}(x)$ as the Lyapunov function. On considering the performance index function (31), $V_{opt}(x) \geq 0$ is in evidence, and $V_{opt}(0) = 0$. Solving the derivative of $V_{opt}(x)$ with respect to t along system (28) yields

$$\frac{dV_{opt}}{dt} = \nabla V_{opt}^T[\bar{A}x + F(z, x)] - \frac{1}{2}\nabla V_{opt}^T BB^T \nabla V$$

According to Lemma 5, it follows that

$$\frac{dV_{opt}}{dt} = \nabla V_{opt}^T \bar{A}x + \nabla V_{opt}^T BG(z, x) - \frac{1}{2}\nabla V_{opt}^T BB^T \nabla V \tag{37}$$

According to HJB Equation (36), we can obtain

$$\nabla V_{opt}^T \bar{A}x = -f_{max}^2(x) - x^T x + \frac{1}{4}\nabla V_{opt}^T BB^T \nabla V_{opt} \tag{38}$$

On substituting Equation (38) into Equation (37), we obtain

$$\begin{aligned}\frac{dV_{opt}}{dt} &= -f_{max}^2(x) - x^T x + \frac{1}{4}\nabla V_{opt}^T BB^T \nabla V_{opt} + \nabla V_{opt}^T BG(z,x) - \frac{1}{2}\nabla V_{opt}^T BB^T \nabla V \\ &= -f_{max}^2(x) - x^T x - \frac{1}{4}\nabla V_{opt}^T BB^T \nabla V_{opt} + \nabla V_{opt}^T BG(z,x)\end{aligned} \tag{39}$$

From Equation (39), we can obtain

$$\begin{aligned}\frac{dV_{opt}}{dt} &= -f_{max}^2(x) - x^T x - \frac{1}{4}[\nabla V_{opt}^T BB^T \nabla V_{opt} - 4\nabla V_{opt}^T BG(z,x) + 4G^T(z,x)G(z,x)] + G^T(z,x)G(z,x) \\ &= -x^T x + G^T(z,x)G(z,x) - f_{max}^2(x) - \frac{1}{4}H^T(z,x)H(z,x) \\ &\leq -x^T x\end{aligned} \tag{40}$$

where $H(z,x) = B^T \nabla V_{opt} - 2G(z,x)$. According to the Lyapunov stability criterion, the optimal controller (35) can asymptotically stabilize the uncertain nonlinear IPS (28) for all the allowable uncertainties. Therefore, for a constant $p > 0$, there exists a neighborhood $\mathcal{N} = \{x : \|x\| < p\}$ near the origin, so that, if $x(t) \in \mathcal{N}$, then $x \to 0$ when $t \to \infty$. However, $x(t)$ cannot remain outside the domain $\mathcal{N}$ forever, or else $\|x(t)\| \geq p$ for all $t > 0$, which implies that

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

Let $t \to \infty$, then

$$V_{opt}[x(t)] \leq V_{opt}[x(0)] - p^2 t \to -\infty$$

This completes the proof. □

5.3. RL Algorithm for Nonlinear IPS

For a nonlinear IPS, we consider the optimal control problems (30) and (31). For any admissible control, the cost function corresponding to the optimal control problem can be expressed as

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

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

$$V[x(t)] = \int_t^{t+\phi} [f_{max}^2(x) + x^T x + u^T u]dt + V[x(t+\phi)] \tag{41}$$

According to the integral-based reinforcement relations (41) and the optimal controller (35), the RL algorithm for the robust control of the nonlinear IPS is as follows.

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 (28).

Algorithm 2 RL Algorithm 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 $V_i[x(t)] = \int_t^{t+\phi}[f_{max}^2(x) + x^T x + u_i^T(x)u_i(x)]dt + V_i[x(t+\phi)]$ 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.

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 (30), solving the cost function $V_i(x)$ from the equation in step 3 in Algorithm 2 can be equivalent to solving Equation (42).*

$$f_{max}^2(x) + x^T x + u_i^T(x)u_i(x) + \nabla V_i[\bar{A}x + Bu_i] = 0 \tag{42}$$

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

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

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

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

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 $\dot{x} = \bar{A}x + Bu_i$, finding the derivative of $V_i(x)$ with respect to t 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 $t+\phi$, we obtain

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

Then, from (42), we obtain

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

The above equation is consistent with the third step of Algorithm 2. This completes the proof. □

According to the conclusions of [25,37], if the initial control policy $u_0(x)$ can stabilize the system, the control policy taken using the optimal control Function (35) and Equation (42) also can stabilize the system. Furthermore, the iteratively calculated cost function sequence converges to the optimal cost function. From Lemma 7, we know that Equation (42) 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.

6. Numerical Simulation Results

This section includes two simulation examples to illustrate the practical applicability of the theoretical results in the robust control of the uncertain IPS.

6.1. Example 1

Considering system (11), whose state-space model can be referenced in [34], our objective is to obtain a robust control u such that it is stable. Based on Lemmas 1–3, the weighting matrix M is selected as

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

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:

$$S_d = \begin{bmatrix} 2.4465 & 2.0822 & -6.2489 & -1.2066 \\ 2.0822 & 4.3346 & -14.0702 & -2.7082 \\ -6.2489 & -14.0702 & 100.8262 & 18.9646 \\ -1.2066 & -2.7082 & 18.9646 & 3.6646 \end{bmatrix} \tag{43}$$

and

$$K_d = [1.0044 \quad 2.5538 \quad -32.2652 \quad -6.2440] \tag{44}$$

There are 10 independent numerical samples in the matrix S_d. These 10 numerical samples are collected in each iteration to address 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 $i = 1, 2, 3, 4$, $j = 1, 2, 3, 4$.

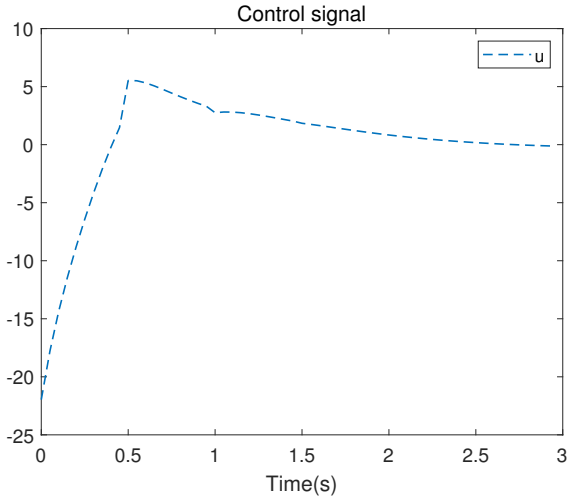

Figure 5. Control signal u of the linearized system.

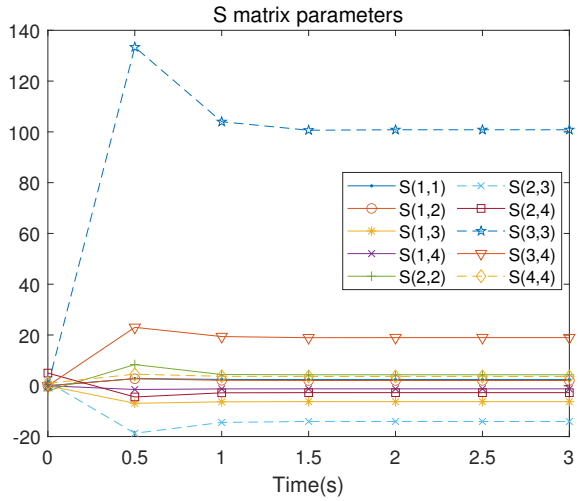

Figure 6. S-matrix iterative process of the linearized system.

The ARE (18) is solved directly by Matlab, the S matrix and optimal feedback K are obtained as follows.

$$S = \begin{bmatrix} 2.4455 & 2.0802 & -6.2326 & -1.2039 \\ 2.0802 & 4.3307 & -14.0381 & -2.7032 \\ -6.2326 & -14.0381 & 100.5677 & 18.9228 \\ -1.2039 & -2.7032 & 18.9228 & 3.6579 \end{bmatrix} \tag{45}$$

$$K = [1.000 \quad 2.5455 \quad -32.1952 \quad -6.2327] \tag{46}$$

As is apparent, the results from the two methods are very similar. Figure 7 presents the closed-loop trajectory of system (11). Figure 7a–d represent the closed-loop system trajectories for uncertain parameters $z = 0.1, 0.4, 0.7, 1.0$, respectively. It is easy to observe that the system is stable, which means that the controller is valid

Table 2 displays the respective partial eigenvalues of the system (11) with $u = Kx$ under varying values of z. 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

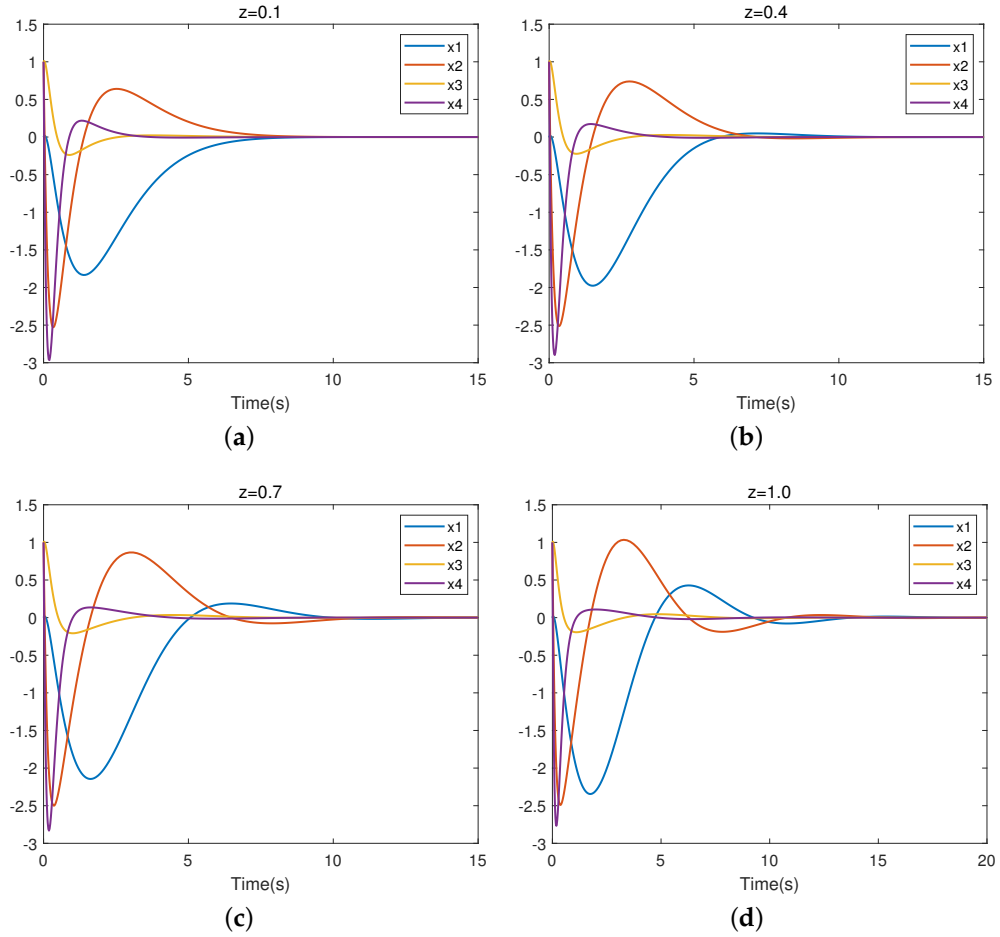

Figure 7. Trajectory of closed-loop linearized system.

6.2. Example 2

Let us consider the nonlinear IPS (28). According to Lemma 5, system (28) can be rewritten as

$$\dot{x} = \bar{A}x + Bu + BG(z, x) \tag{47}$$

The optimal control problem for the IPS is as follows: for nominal system (30), we find an optimal control u such that the performance index (31) achieves a minimum.

According to Lemma 6, we obtain

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

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 (31), the weight matrix M is selected as

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

Based on Algorithm 2, we 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.

$$S_d = \begin{bmatrix} 2.4325 & 2.4398 & -6.2874 & -1.3888 \\ 2.4398 & 5.0469 & -14.0539 & -3.0045 \\ -6.2874 & -14.0539 & 130.4535 & 18.9735 \\ -1.3888 & -3.0045 & 18.9735 & 4.2715 \end{bmatrix} \tag{48}$$

and

$$K_d = [1.1180 \quad 2.6229 \quad -32.3006 \quad -7.4126] \tag{49}$$

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

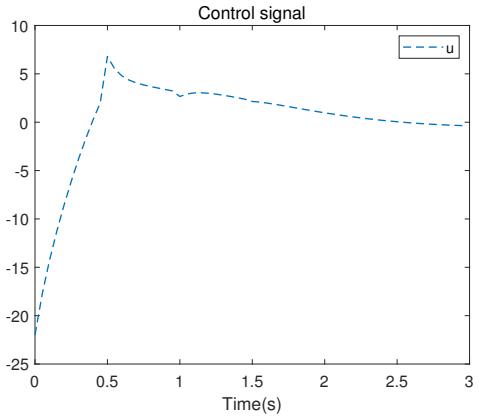

Figure 8. Control signal u of the nonlinear system.

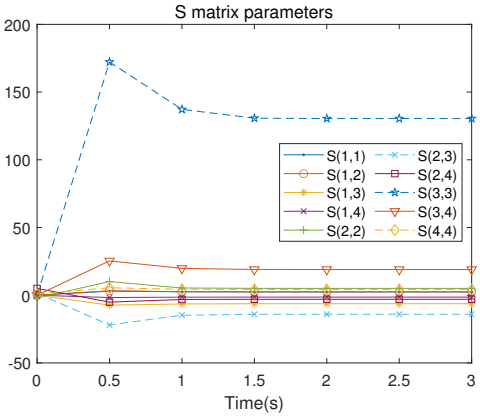

Figure 9. S-matrix iterative process of the nonlinear system.

We also selected z = 0.1, 0.4, 0.7, 1.0. Figure 10 presents the closed-loop trajectory of system (28). Figure 10a–d represent the closed-loop system trajectories for uncertain parameters z = 0.1, 0.4, 0.7, and 1.0, respectively. It is easy to observe that the system is stable, which means that the controller is valid.

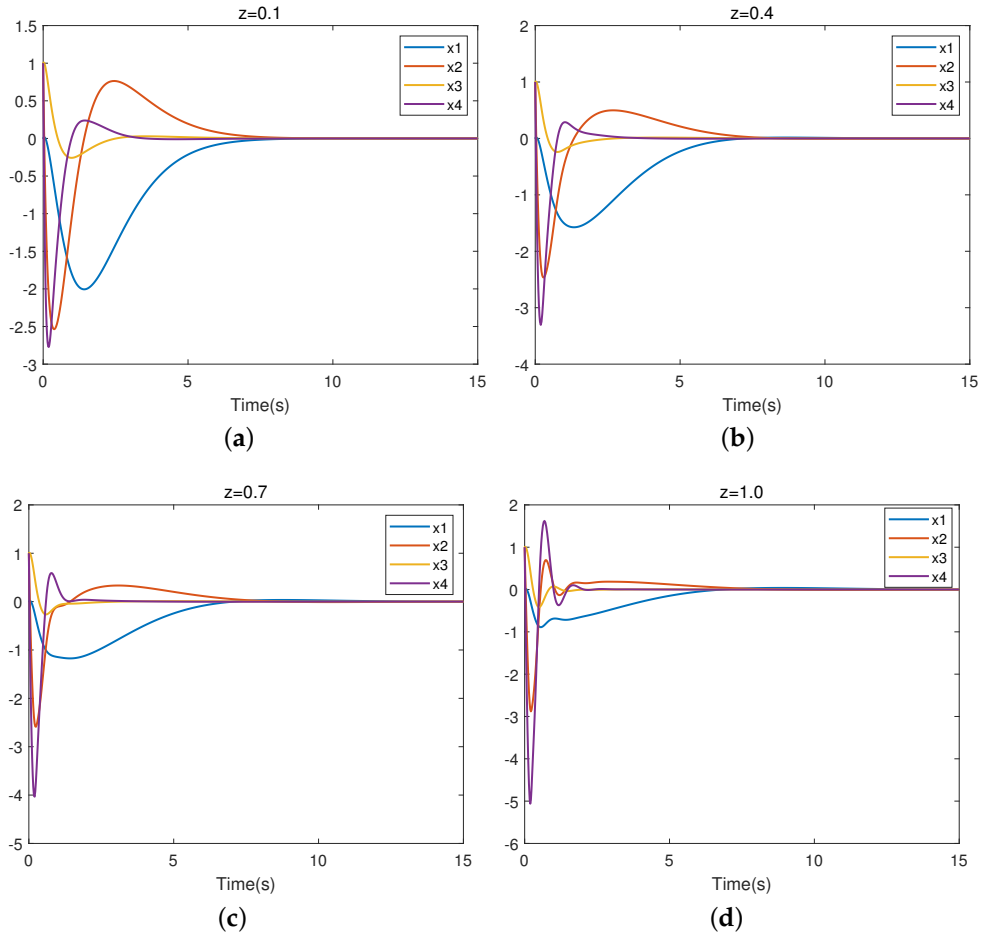

Figure 10. Trajectory of closed-loop nonlinear system.

7. Conclusions

In this paper, the robust control problem of a 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 algorithm can provide a controller that meets the requirements without the nominal matrix A of the system being known, only collecting input and output data. 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: Y.M.: investigation, methodology, software; D.X.: formal analysis; J.H.: investigation; Y.L.: writing—original draft preparation; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guizhou Province Natural Science Foundation of China under Grant No. Qiankehe Fundamentals−ZK[2021] General 322 and the Doctoral Foundation of Guangxi University of Science and Technology Grant No. Xiaokebo 22z04 .

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

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. Marrison, C.I.; Stengel, R.F. Design of Robust Control Systems for a Hypersonic Aircraft. *J. Guid. Control Dyn.* **1998**, *21*, 58–63. [CrossRef]
2. Yao, B.; Al-Majed, M.; Tomizuka, M. High-Performance Robust Motion Control of Machine Tools: An Adaptive Robust Control Approach and Comparative Experiments. *IEEE/ASME Trans. Mechatron.* **1997**, *2*, 63–76.
3. Stephenson, A. *A New Type of Dynamical Stability*; Manchester Philosophical Society: Manchester, UK, 1908; Volume 52, pp. 1–10.
4. Housner, G.W. The behavior of inverted pendulum structures during earthquakes. *Bull. Seismol. Soc. Am.* **1963**, *53*, 403–417. [CrossRef]
5. Wang, J.J. Simulation studies of inverted pendulum based on PID controllers. *Simul. Model. Pract. Theory* **2011**, *19*, 440–449. [CrossRef]
6. Li, D.; Chen, H.; Fan, J.; Shen, C. A novel qualitative control method to inverted pendulum systems. *IFAC Proc. Vol.* **1999**, *32*, 1495–1500. [CrossRef]
7. Nasir, A.N.K.; Razak, A.A.A. Opposition-based spiral dynamic algorithm with an application to optimize type-2 fuzzy control for an inverted pendulum system. *Expert Syst. Appl.* **2022**, *195*, 116661. [CrossRef]
8. Tsay, S.C.; Fong, I.K.; Kuo, T.S. Robust linear quadratic optimal control for systems with linear uncertainties. *Int. J. Control* **1991**, *53*, 81–96. [CrossRef]
9. Lin, F.; Brandt, R.D. An optimal control approach to robust control of robot manipulators. *IEEE Trans. Robot. Autom.* **1998**, *14*, 69–77.
10. Lin, F. An optimal control approach to robust control design. *Int. J. Control* **2000**, *73*, 177–186. [CrossRef]
11. Zhang, X.; Kamgarpour, M.; Georghiou, A.; Goulart, P.; Lygeros, J. Robust optimal control with adjustable uncertainty sets. *Automatica* **2017**, *75*, 249–259. [CrossRef]
12. Wang, D.; Liu, D.; Zhang, Q.; Zhao, D. Data-based adaptive critic designs for nonlinear robust optimal control with uncertain dynamics. *IEEE Trans. Syst. Man Cybern. Syst.* **2015**, *46*, 1544–1555. [CrossRef]
13. Bellman, R. Dynamic programming. *Science* **1966**, *153*, 34–37. [CrossRef]
14. Neustadt, L.W.; Pontrjagin, L.S.; Trirogoff, K. *The Mathematical Theory of Optimal Processes*; Interscience: London, UK, 1962.
15. Powell, W.B. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*; John Wiley & Sons: Hoboken, NJ, USA, 2007; Volume 703.
16. Li, H.; Liu, D. Optimal control for discrete-time affine non-linear systems using general value iteration. *IET Control Theory Appl.* **2012**, *6*, 2725–2736. [CrossRef]
17. Wei, Q.; Liu, D.; Lin, H. Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems. *IEEE Trans. Cybern.* **2015**, *46*, 840–853. [CrossRef] [PubMed]
18. Tesauro, G. TD-Gammon, a self-teaching backgammon program, achieves master-level play. *Neural Comput.* **1994**, *6*, 215–219. [CrossRef]
19. Singh, S.; Bertsekas, D. Reinforcement learning for dynamic channel allocation in cellular telephone systems. *Adv. Neural Inf. Process. Syst.* **1996**, *9*, 974–980.
20. Maja, J.M. Reward Functions for Accelerated Learning. In *Machine Learning Proceedings 1994*, 1st ed.; Cohen, W.W., Hirsh, H., Eds.; Morgan Kaufmann: Burlington, MA, USA, 1994; pp. 181–189.
21. Doya, K. Reinforcement learning in continuous time and space. *Neural Comput.* **2000**, *12*, 219–245. [CrossRef]
22. Krstic, M.; Kokotovic, P.V.; Kanellakopoulos, I. *Nonlinear and Adaptive Control Design*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1995.
23. Ioannou, P.; Fidan, B. *Adaptive Control Tutorial, Vol. 11 of Advances in Design and Control*; SIAM: Philadelphia, PA, USA, 2006.
24. Åström, K.J.; Wittenmark, B. *Adaptive Control*; Courier Corporation: North Chelmsford, MA, USA, 2013.
25. Vrabie, D.; Pastravanu, O.; Abu-Khalaf, M.; Lewis, F.L. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica* **2009**, *45*, 477–484. [CrossRef]
26. Xu, D.; Wang, Q.; Li, Y. Adaptive optimal control approach to robust tracking of uncertain linear systems based on policy iteration. *Meas. Control* **2021**, *54*, 668–680. [CrossRef]
27. Xu, D.; Wang, Q.; Li, Y. Optimal guaranteed cost tracking of uncertain nonlinear systems using adaptive dynamic programming with concurrent learning. *Int. J. Control Autom. Syst.* **2020**, *18*, 1116–1127. [CrossRef]
28. Bates, D. A hybrid approach for reinforcement learning using virtual policy gradient for balancing an inverted pendulum. *arXiv* **2021**, arXiv:2102.08362.
29. Israilov, S.; Fu, L.; Sánchez-Rodríguez, J.; Fusco, F.; Allibert, G.; Raufaste, C.; Argentina, M. Reinforcement learning approach to control an inverted pendulum: A general framework for educational purposes. *PLoS ONE* **2023**, *18*, e0280071. [CrossRef]
30. Lin, B.; Zhang, Q.; Fan, F.; Shen, S. A damped bipedal inverted pendulum for human–structure interaction analysis. *Appl. Math. Model.* **2020**, *87*, 606–624. [CrossRef]

31. Puriel-Gil, G.; Yu, W.; Sossa, H. Reinforcement learning compensation based PD control for inverted pendulum. In Proceedings of the 2018 15th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), Mexico City, Mexico, 5–7 September 2018; pp. 1–6.
32. Surriani, A.; Wahyunggoro, O.; Cahyadi, A.I. Reinforcement learning for cart pole inverted pendulum system. In Proceedings of the 2021 IEEE Industrial Electronics and Applications Conference (IEACon), Penang, Malaysia, 22–23 November 2021; pp. 297–301.
33. Landry, M.; Campbell, S.A.; Morris, K.; Aguilar, C.O. Dynamics of an inverted pendulum with delayed feedback control. *SIAM J. Appl. Dyn. Syst.* **2005**, *4*, 333–351. [CrossRef]
34. Muskinja, N.; Tovornik, B. Swinging up and stabilization of a real inverted pendulum. *IEEE Trans. Ind. Electron.* **2006**, *53*, 631–639. [CrossRef]
35. Bhatia, N.P.; Szegö, G.P. *Stability Theory of Dynamical Systems*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2002.
36. Kleinman, D. On an iterative technique for Riccati equation computations. *IEEE Trans. Autom. Control* **1968**, *13*, 114–115. [CrossRef]
37. Abu-Khalaf, M.; Lewis, F.L. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. *Automatica* **2005**, *41*, 779–791. [CrossRef]