actuators

MDPI

Article

Integral Reinforcement Learning-Based Stochastic Guaranteed Cost Control for Time-Varying Systems with Asymmetric Saturation Actuators

Yuling Liang [1,*], Mengjia Xie [1], Juan Zhang [2], Zhongyang Ming [2] and Zhiyun Gao [3]

[1] School of Artificial Intelligence, Shenyang University of Technology, Shenyang 110870, China; mengjiax@smail.sut.edu.cn
[2] School of Information Science and Engineering, Northeastern University, Shenyang 110819, China; zhangjuan@ise.neu.edu.cn (J.Z.); mingzhongyang@ise.neu.edu.cn (Z.M.)
[3] School of Transportation Science and Engineering, Civil Aviation University of China, Tianjin 300300, China; zygao@cauc.edu.cn
* Correspondence: liangwaxx@stu.edu.cn

Abstract

This study explores a stochastic guarantee cost control (GCC) for time-varying systems with random parameters and asymmetric saturation actuators by employing the integral reinforcement learning (IRL) method in the dynamic event-triggered (DET) mode. Firstly, a modified Hamilton–Jacobi–Isaac (HJI) equation is formulated, and then the worst-case disturbance policy and the asymmetric saturation optimal control signal can be obtained. Secondly, the multivariate probabilistic collocation method (MPCM) is used to evaluate the value function at designated sampling points. The purpose of introducing the MPCM is to simplify the computational complexity of stochastic dynamic programming (SDP) methods. Furthermore, the DET mode is utilized to solve the SDP problem to reduce the computational burden on communication resources. Finally, the Lyapunov stability theorem is applied to analyze the stability of time-varying systems, and the simulation shows the feasibility of the designed method.

Keywords: stochastic systems; asymmetric saturation actuators; multivariate probabilistic collocation method; adaptive dynamic programming

check for updates

Academic Editor: Paolo Mercorelli

Received: 3 September 2025
Revised: 5 October 2025
Accepted: 17 October 2025
Published: 19 October 2025

Citation: Liang, Y.; Xie, M.; Zhang, J.; Ming, Z.; Gao, Z. Integral Reinforcement Learning-Based Stochastic Guaranteed Cost Control for Time-Varying Systems with Asymmetric Saturation Actuators. *Actuators* **2025**, *14*, 506. https://doi.org/10.3390/act14100506

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Over the past few decades, stochastic control has been extensively studied across numerous practical domains, including robotic systems [1], microgrid setups [2], intelligent transportation networks [3], and other practical control fields. Compared with deterministic system control, stochastic system control has problems such as computational complexity and model mismatch, which makes its design challenging [4]. In [5], the stochastic control problem was studied in data-driven model predictive control by using neural networks (NNs) or establishing a stochastic sampling model. A stochastic control approach applicable to discrete-time nonlinear systems was put forward in [6]. The security issues of stochastic systems were also addressed in [7]. A novel stochastic control method was designed for a linear system with multiplicative noise and input constraints in [8]. Regarding stochastic discrete-time linear systems under bounded input constraints, a receding horizon control technique was formulated in [9]. However, none of the studies mentioned above considered optimal control under asymmetric saturation actuators (ASAs).

Actuator saturation is a common nonlinear problem that can affect system stability and performance and also pose significant challenges to controller design. In [10], a robust adaptive fuzzy fault-tolerant path planning control method was proposed to solve the actuator saturation problem. In [11], an adaptive compensation technique integrated into the backstepping framework was proposed for controlled systems with actuator saturation. In [12], an adaptive NN control method was proposed for systems with input saturation nonlinearities. These methods have symmetric saturation actuators, and in order to optimize control strategies, researchers have proposed ASA control strategies to ensure the safety and stability of the controlled system. In [13], a composite adaptive control method was proposed for tethered aircraft systems with ASAs. Ref. [14] proposed a sliding mode control method with ASAs for vehicle systems. In [15], an asymptotic stabilization method was proposed for unmanned surface vessels with actuator dead zones and bow constraints. However, the methods mentioned above do not consider the optimal performance of the controlled system. The adaptive dynamic programming (ADP) algorithm, which is a method that combines dynamic programming (DP) and integral reinforcement learning (RL) ideas, addresses the "curse of dimensionality" issue in conventional DP [16–19]. For the Itô-type stochastic systems, the authors employed the ADP method to address the tracking control issue in [20]. In [21], a self-learning optimal operation control method based on a compensator and RL was proposed for input-constrained dual-time scale systems. A Q-learning-based fault-tolerant control method was designed without using the knowledge of system function in [22]. In [23], an RL-based optimal control (OC) method was devised for linear stochastic systems, with the least squares approach adopted to derive approximate optimal solutions. Further, an integral RL-based model-free tracking approach was developed for linear stochastic systems in [24]. In [25], an ADP-based GCC method was presented for nonlinear systems. By introducing the GCC approach, one can ensure that the control performance index is below a certain bound. The above methods may cause unnecessary waste of communication resources because these approaches are designed based on time-triggered control mechanisms.

Event-triggered control (ETC) is a non-periodic control strategy based on triggering conditions; the core idea of ETC is to determine when to update control signals or communicate through preset event-triggered conditions, rather than traditional periodic sampling or time triggering [26,27]. In [28], for Itô-type time-delayed stochastic systems subject to uncertainties, an OC strategy under the ETC mechanism was designed by using the ADP method and constructing an actor–critic architecture. Moreover, an optimal consensus control (OCC) strategy was proposed by using fuzzy technology for stochastic MASs with time delays in [29]. To further decrease control input update frequency and increase communication resource utilization, the dynamic event-triggered control (DETC) mechanism serves as an upgraded version of the conventional ETC approach. It is a dynamically adjustable mechanism that further optimizes resource usage, system performance, and adaptability [30]. In [31], under the DETC mechanism, the H2/H∞ problem was explored for partially unknown nonlinear stochastic systems by employing the ADP algorithm. In [32], focusing on uncertain microgrids, control inputs and external disturbances were treated as two participants in zero-sum differential game scenarios. Under the DETC mechanism, the authors put forward a frequency recovery control approach based on integral reinforcement learning (IRL). In [33], a dynamically adjustable triggering condition was designed based on the OC method for stochastic nonlinear systems by using backstepping technology and the RL algorithm in the DETC mode.

In addition, it is difficult to obtain accurate mathematical models for actual control systems, and estimating system dynamics through extended simulation experiments is a conventional method [34,35]. If random time-varying parameters exist in actual nonlinear

controlled system dynamics, the expected control performance function is always estimated by using uncertainty evaluation methods [36], such as the Monte Carlo (MC) approach and its extensive version, which are implemented based on simulation experiments [37]. However, approximating the expected control performance index requires a large number of simulation experiments when using the MC method and its extensive version [38]. In [39], to simplify the computational complexity, the MPCM method, which is an effective technology for estimating uncertainties, was used to estimate the expected value of the cost function by taking values at certain specific sampling points. Stochastic GCC approaches have been put forward by using RL (or IRL) and MPCM methods in [36,40]. However, the issue of ASAs remains unaddressed. In our design, we put forward a GCC approach in DETC mode for stochastic systems subject to ASAs by employing the the MPCM integrated with IRL algorithms. The main contributions are as follows:

1. This study develops an innovative DETC-based GCC approach for stochastic systems by using IRL algorithms with the MPCM. This approach can ensure that the system performance index is less than a certain upper bound.
2. By solving an improved HJI equation by designing a modified long-term performance cost function, the control inputs under ASAs can be obtained via an actor–critic–disturbance NN structure.
3. Through the introduction of dynamic parameters into triggering conditions in the DETC mechanism, the update rule for control inputs can be adjusted dynamically, which can reduce the computational complexity of sampling data.

This paper's structure is arranged as follows. Section 2 presents the issue description. Section 3 develops stochastic systems' optimal GCC design via a modified HJI equation. A stochastic GCC method is designed by using the MPCM and IRL in Section 4. Section 5 presents the event-triggered structure for optimal GCC. A simulation example is shown to verify the feasibility of the strategy in Section 6. The conclusion is shown in Section 7.

2. Problem Statement

Consider the following stochastic system [36,41]:

$$\dot{\varsigma} = \chi(g)\varsigma + B\rho + Cd, \tag{1}$$

where $\chi(g) \in R^{n\times n}$ represents a time-varying matrix related to the stochastic vector $g(t)$. Matrices $B \in R^{n\times q}$ and $C \in R^{n\times p}$ are the system functions. The disturbance policy, control input, and system state are $d \in R^p$, $\rho \in R^q$ and $\varsigma(t) \in R^n$, respectively. The control signal $\rho = [\rho_1, \rho_2, \ldots, \rho_q]^T$ satisfies $\rho_{min} \le \rho_i \le \rho_{max}$, $|\rho_{\min}| \neq |\rho_{\max}|$, for $i = 1, 2, \ldots, q$. $\rho_{\min}$ and $\rho_{\max}$ stand for the smallest and largest constraint limits of control inputs ρ_i. $g_p(t)$ is the sample function of the pth element in the random vector $g(t)$, and the dynamics of the sample system (1) are ordinary differential equations; $g_p(t)$ exhibits good behavior.

Assumption 1. *The system (1) satisfies Lipschitz continuity over the compact set $\Omega_\varsigma \subset R^n$ [36,42,43]. In addition, $\chi(g) \le b_a$, where b_a is a positive constant.*

Remark 1. *Unlike general stochastic models with drift and diffusion terms, Ref. [41] focused on zero-sum game studies of time-varying stochastic systems (1) with environment-influenced random variables. The system model (1) is applicable to practical scenarios. For example, aircraft systems have been modeled as $\dot{\varsigma} = -K\varsigma(t) + M_u(t) + M_d(t)$, where K is a weather-dependent random variable, $M_u(t)$ refers to regulated thrust, and $M_d(t)$ is the interference force. Other fields also explore such system dynamics [44].*

Assumption 2. *Let β_1 denote a positive constant; the disturbance approach $d(\varsigma)$ is such that $d^T(\varsigma)d(\varsigma) \leq \beta_1 \varsigma^T \varsigma$, with $d(0) = 0$.*

The system (1) with nominal form becomes

$$\dot{\varsigma} = \chi(g)\varsigma + B\rho. \tag{2}$$

The positive definite matrix is represented as $Q \in R^{n\times n}$, and the long-term control performance function is

$$\bar{\Theta}(\varsigma(0), \rho) = E\left[\int_0^\infty \left(\varsigma^T Q \varsigma + \mathcal{W}(\rho)\right) dt\right]. \tag{3}$$

with

$$\mathcal{W}(\rho) = 2\kappa_2 \int_{\kappa_1}^{\rho(\varsigma)} \tilde{\rho}_a^{-T}\left(\kappa_2^{-1}(\vartheta - \kappa_1)\right) R \, \mathrm{d}\vartheta,$$
$$\kappa_1 = \frac{\rho_{\max} + \rho_{\min}}{2}, \quad \kappa_2 = \frac{\rho_{\max} - \rho_{\min}}{2},$$

where $\tilde{\rho}_a(\cdot) = \tanh(\cdot)$. For $\tilde{r}_i > 0, i = 1, \ldots, q$, $R = diag\left(\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_q\right)$.

By calculation, we have

$$\mathcal{W}(\rho) = 2\kappa_2 \tanh^{-T}\left(\frac{\rho(\varsigma) - \kappa_1}{\kappa_2}\right) R(\rho(\varsigma) - \kappa_1)$$
$$+ \kappa_2^2 \bar{R}^T \ln\left(\bar{\mathbf{1}} - \left(\frac{\rho(\varsigma) - \kappa_1}{\kappa_2}\right)^2\right),$$

where $\bar{R} = \left[\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_q\right]^T$ and $\bar{\mathbf{1}} = \underbrace{[1, \ldots, 1]^\top}_{q}$.

Remark 2. *Ref. [36] studied the GCC problem of stochastic systems under an event-triggered mechanism. However, some practical systems face the nonlinear control problem of ASAs [13–15], which limits the application of the method in [36]. Therefore, this paper proposes a GCC strategy for stochastic systems under the ASA condition with $|\rho_{\min}| \neq |\rho_{\max}|$. By introducing a non-quadratic function $\mathcal{W}(\rho)$ related to the control input ρ to relax the asymmetric constraints and incorporating dynamic parameters into the triggering condition, the number of communications can be effectively reduced, and the utilization rate of communication resources can be further improved.*

Within our design framework, the primary objective is to establish a GCC strategy for systems with ASAs by solving the HJI equation, which can ensure that the cost function is less than a certain upper bound and guarantee the stability and optimal control performance of stochastic systems under the DETC mechanism.

3. Stochastic Optimal GCC Design

In our design, an auxiliary policy is designed as $\eth$, and the auxiliary system is established as

$$\dot{\varsigma} = \chi(g)\varsigma + B\rho + C\eth. \tag{4}$$

Let $r(\varsigma, \rho, \eth) = \varsigma^T Q \varsigma + \mathcal{W}(\rho) - \gamma^2 \eth^T \eth$, where γ represents a pre-specified disturbance attenuation level.

The modified cost function is

$$\Theta(\varsigma(0),\rho,\eth)=E\left[\int_0^\infty\left(r(\varsigma,\rho,\eth)+\gamma^2\beta_1\varsigma^T\varsigma\right)dt\right]. \tag{5}$$

The value function is

$$\Gamma(\varsigma(t))=E\left[\int_t^\infty\left(r(\varsigma,\rho,\eth)+\gamma^2\beta_1\varsigma^T\varsigma\right)dt\right]. \tag{6}$$

The derivative value of $\Gamma(\varsigma(t))$ is

$$E\left[\dot{\Gamma}(\varsigma(t))\right]=-\gamma^2\beta_1\varsigma^T\varsigma-r(\varsigma,\rho,\eth). \tag{7}$$

Let $\nabla\Gamma(\varsigma)=\frac{\partial\Gamma(\varsigma)}{\partial\varsigma}$, inspired by [36,41], the Hamiltonian function is designed in the form of a mean, which is given by

$$\begin{aligned}\mathcal{H}(\varsigma,\nabla\Gamma,\rho,\eth)=&E\left[(\nabla\Gamma)^T(\chi(g)\varsigma+B\rho+C\eth)\right]\\&+\gamma^2\beta_1\varsigma^T\varsigma+r(\varsigma,\rho,\eth).\end{aligned} \tag{8}$$

Bellman's equation refers to the principle of optimization in DP problems. In light of Bellman's optimality principle, the following holds:

$$\min_{\rho}\max_{\eth}\mathcal{H}(\varsigma,\nabla\Gamma^*,\rho,\eth)=0, \tag{9}$$

with $\eth^*$, ρ^*, and $\Gamma^*(\varsigma)$ being the worst-case disturbance policy, optimal control signal, and optimal value function, respectively. By calculation, we have

$$\rho^*=-\kappa_2\tanh(N^*)+\kappa_1, \tag{10}$$

$$\eth^*=\frac{1}{2\gamma^2}C^T\nabla\Gamma^*(\varsigma), \tag{11}$$

where $N^*=\frac{1}{2\kappa_2}R^{-1}B^T\nabla\Gamma^*(\varsigma)$.

According to (10) and (11), we get

$$\begin{aligned}\mathcal{W}(\rho^*)&=2\kappa_2\int_{\kappa_1}^{\rho^*(\varsigma)}\tilde{\rho}_a^{-T}\left(\kappa_2^{-1}(\vartheta-\kappa_1)\right)R\mathrm{d}\vartheta\\&=2\kappa_2\int_{\kappa_1}^{-\kappa_2\tanh(N^*)+\kappa_1}\tilde{\rho}_a^{-T}\left(\kappa_2^{-1}(\vartheta-\kappa_1)\right)R\mathrm{d}\vartheta\\&=2\kappa_2^2N^{*T}R\tanh(N^*)+k_2^2\bar{R}^T\ln\left(\bar{\mathbf{1}}-\tanh^2(N^*)\right),\end{aligned} \tag{12}$$

$$U_2^*=\gamma^2P^{*T}P^*, \tag{13}$$

with $P^*=\frac{C^T\nabla\Gamma^*}{2\gamma^2}$ and $r(\varsigma,\rho^*,\eth^*)=\varsigma^TQ\varsigma+\mathcal{W}(\rho^*)-U_2^*$.

Based on (8), (12) and (13), the HJI equation is restructured as

$$\begin{aligned}&\mathcal{H}(\varsigma,\nabla\Gamma^*,\rho^*,\eth^*)\\=&E\left[(\nabla\Gamma^*)^T(\chi(g)\varsigma+B\rho^*+C\eth^*)\right]+\gamma^2\beta_1\varsigma^T\varsigma+\varsigma^TQ\varsigma+\mathcal{W}(\rho^*)-\gamma^2\eth^{*T}\eth^*\\=&E\Big[\varsigma^TQ\varsigma+\nabla\Gamma^{*T}\chi(g)\varsigma+2\kappa_2^2N^{*T}R\tanh(N^*)+\kappa_2^2\bar{R}^T\ln\left(\bar{\mathbf{1}}-\tanh^2(N^*)\right)\\&+\gamma^2P^{*T}P^*+\gamma^2\beta_1\varsigma^T\varsigma+2\kappa_2N^{*T}R(-\kappa_2\tanh(N^*)+\kappa_1)\Big]\\=&E\left[\varsigma^TQ_1\varsigma+\nabla\Gamma^{*T}\chi(g)\varsigma+\gamma^2P^{*T}P^*+\kappa_2^2\bar{R}^T\ln\left(\bar{\mathbf{1}}-\tanh^2(N^*)\right)+2\kappa_2N^{*T}R\kappa_1\right]=0\end{aligned} \tag{14}$$

where $I_{n\times n}$ is an $n \times n$ identity matrix, which satisfies $Q_1 = Q + \gamma^2\beta_1 I_{n\times n}$ and $\Gamma(0) = 0$.

Assumption 3. *λ_1 and λ_2 are denoted as positive numbers. $\Gamma(\varsigma)$ is continuously differentiable, which satisfies $\lambda_1\|\varsigma\|^2 \leq \Gamma(\varsigma) \leq \lambda_2\|\varsigma\|^2$.*

Remark 3. *Due to the existence of uncertainty and random parameters, the OC problem of system (1) is difficult to solve. The GCC problem can obtain the upper bound of the cost function of system (1), which is the basic idea of the GCC problem. This article constructs an auxiliary system, and by solving the optimal value function of the auxiliary system, the guaranteed cost (GC) function of the original system (1) can be obtained; that is, the value function of the auxiliary system is the optimal GC function of the original system. A related theorem is given as follows.*

Theorem 1. *Considering system (1), define the cost function of auxiliary system (4) as (5); the asymptotic stability characteristic in the mean of stochastic system (1) can be guaranteed if a differentiable value function Γ^* satisfies the HJI equation in (14) and the asymmetric cost control signal in (10) and the auxiliary policy in (11) are utilized.*

Proof. Choose the candidate Lyapunov function as

$$\Gamma^*(\varsigma(t)) = E\left[\int_t^\infty \left(\varsigma^T Q_1 \varsigma + \mathcal{W}(\rho^*) - \gamma^2 \eth^{*T}\eth^*\right) d\varrho\right].$$

When ρ^* is substituted into (1), system (1) becomes

$$\dot{\varsigma} = \chi(g)\varsigma + B\rho^* + Cd. \tag{15}$$

Through the addition and substitution of $\nabla\Gamma^{*T}(\varsigma)C\eth^*$, we derive

$$\begin{aligned} E\left[\dot{\Gamma}^*(\varsigma)\right] &= E\left[(\nabla\Gamma^*(\varsigma))^T(\chi(g)\varsigma + B\rho^* + Cd)\right] \\ &= E\left[(\nabla\Gamma^*(\varsigma))^T(\chi(g)\varsigma + B\rho^* + C\eth^*)\right] \\ &\quad + E\left[(\nabla\Gamma^*(\varsigma))^T C(d - \eth^*)\right]. \end{aligned} \tag{16}$$

Based on (8), (9) and (11), one gets

$$\begin{aligned} &E\left[(\nabla\Gamma^*(\varsigma))^T(\chi(g)\varsigma + B\rho^* + C\eth^*)\right] + E\left[(\nabla\Gamma^*(\varsigma))^T C(d - \eth^*)\right] \\ =&E\left[-\gamma^2\beta_1\varsigma^T\varsigma - \varsigma^T Q\varsigma - \mathcal{W}(\rho^*) + \gamma^2\eth^{*T}\eth^*\right] \\ &+ E\left[2\gamma^2\eth^{*T}(d - \eth^*)\right]. \end{aligned} \tag{17}$$

Using Assumption 2 and the inequality $2\gamma^2\eth^{*T}d \leq \gamma^2\eth^{*T}\eth^* + \gamma^2 d^T d$, one obtains

$$E\left[\dot{\Gamma}^*(\varsigma)\right] \leq E\left[-\varsigma^T Q\varsigma - \mathcal{W}(\rho^*)\right] \leq 0. \tag{18}$$

Integrating Equation (18) over the interval $[0, \infty)$, we have

$$E\left[\int_0^\infty \left(\varsigma^T Q\varsigma + \mathcal{W}(\rho^*\right) d\varrho\right] \leq \Gamma^*(\varsigma(0)). \tag{19}$$

From (18), [36] (Lemma 1) and Assumption 3, the asymptotic stability characteristic in the mean of stochastic system (1) can be guaranteed under the application of the control pairs $(\rho^*, \eth^*)$ given in (10) and (11). Additionally, based on (18), we have $\bar{\Theta}(\varsigma(0), \rho^*) \leq \Gamma^*(\varsigma(0))$. □

4. Stochastic GCC Method Design via MPCM and IRL Algorithm

In this section, a novel stochastic GCC method is proposed for the controlled system with asymmetric constrained inputs by using the MPCM and IRL algorithms.

Using (7), we derive the integral Bellman equation based on the optimality principle as $\Gamma^*(\varsigma(t)) = E\Big[\int_t^{t+T}(\gamma^2\beta_1\varsigma^T\varsigma + r(\varsigma,\rho^*,\eth^*))d\varrho + \Gamma^*(\varsigma(t+T))\Big]$ with $T \in (0,t)$. For any admissible control policies $(\rho(\varsigma),\eth(\varsigma))$, the value function satisfies the condition that

$$\Gamma(\varsigma(t)) = E\bigg[\int_t^{t+T}\Big(\gamma^2\beta_1\varsigma^T\varsigma + r(\varsigma,\rho,\eth)\Big)d\varrho + \Gamma(\varsigma(t+T))\bigg]. \tag{20}$$

Accordingly, one obtains

$$\begin{aligned}\check{e}_B =& E\bigg[\int_t^{t+T}\Big(r(\varsigma,\rho,\eth) + \gamma^2\beta_1\varsigma^T\varsigma\Big)d\varrho \\ &+ \Gamma(\varsigma(t+T))\bigg] - \Gamma(\varsigma(t))\end{aligned} \tag{21}$$

where $\check{e}_B$ is the Bellman error.

Remark 4. *Equation (21) will be employed in the subsequent on-policy IRL design procedure. In our framework, the IRL algorithm serves to drive the Bellman error $\check{e}_B$ in (21) toward zero.*

As $t \to \infty$, $(\rho(\varsigma), d(\varsigma))$ converges to $(\rho^*(\varsigma), d^*(\varsigma))$ after several iterations in the learning process, and concurrently, $\Gamma(\varsigma)$ approaches $\Gamma^*(\varsigma)$.

4.1. On-Policy GCC Design

By designing an NN that includes the activation function, the expected weight vector and the approximation error are denoted as $\varphi_c(\varsigma)$, W_c, and $\varepsilon_c(\varsigma)$, respectively. We approximate the value function $\Gamma^*(\varsigma)$ as

$$\Gamma^*(\varsigma) = W_c^T\varphi_c(\varsigma) + \varepsilon_c(\varsigma). \tag{22}$$

Let $\hat{W}_c$ denote the estimated value for W_c. At time step $s+1$, the estimate value of $\Gamma(\varsigma)$ is

$$\Gamma^{(s+1)}(\varsigma) = \hat{W}_c^{(s+1)T}\varphi_c(\varsigma). \tag{23}$$

Based on [40] (Th.2), we adopt the MPCM to reduce the number of simulations from $2^m\prod_{q=1}^m N_q$ to $\prod_{q=1}^m N_q$, which enables the prediction of the output mean of the system mapping $G(\eta_1,\eta_2,\ldots,\eta_m)$. In this part, the GCC algorithm for random systems with stochastic uncertainty is presented in Algorithm 1.

The specific steps of Algorithm 1 include the following.

(1) Select a set of sampling points for the uncertain variable $g(t)$ based on the MPCM and calculate the cumulative cost function G_Γ for the future time period at each sampling point. This function is composed of the output estimate $\hat{W}_c^T\varphi_c$ of the next evaluation critic NN and the integral term containing control energy consumption and state cost in the current time period. (2) Calculate the mathematical expectation of the G_Γ function values at all sampling points to obtain the updated value function estimate for this iteration, denoted as $\Gamma^{(s+1)}$. With this estimated value as the objective, update the weight vector $\hat{W}_c$ of the evaluation NN by solving the HJI equation so that its output accurately matches the value function estimate. (3) Using the gradient information of the updated value function, design control strategy $\rho^{(s+1)}$ and disturbance strategy $\eth^{(s+1)}$, respectively.

Algorithm 1 Model-based GCC algorithm for stochastic uncertain system (1).

Initialization Set initial admissible policies $\rho(0)$ and $\eth(0)$.

Step 1: A set of sampling points for the uncertain variable $g(t)$ is selected on the basis of the MPCM ([40], Section II). Compute the value of G_Γ at each sampling point:

$$\begin{aligned} & G_{\Gamma^{(s+1)}}\left(\varsigma, \eth^{(s)}, \rho^{(s)}, g\right) \\ =& \hat{W}_c^{(s+1)T} \varphi_c\left(\varsigma(t+T)\right) \\ & + \int_t^{t+T} \Big(r\left(\varsigma(\varrho), \rho^{(s)}(\varrho), \eth^{(s)}(\varrho)\right) \\ & + \gamma^2 \beta_1 \varsigma^T \varsigma \Big) d\varrho. \end{aligned} \tag{24}$$

Step 2: Compute $\Gamma^{(s+1)}(\varsigma(t))$ by computing the mean value of $G_{\Gamma^{(s+1)}}(\cdot)$.

$$\Gamma^{(s+1)}\left(\varsigma(t)\right) = E\Big[G_{\Gamma^{(s+1)}}\left(\varsigma, \rho^{(s)}, \eth^{(s)}, g\right)\Big]. \tag{25}$$

Step 3: Update $W_c^{(s+1)}$ on the basis of solving the HJI equation:

$$\hat{W}_c^{(s+1)T} \varphi_c(\varsigma(t)) = \Gamma^{(s+1)}\left(\varsigma(t)\right). \tag{26}$$

Step 4: Design the control pairs via

$$\rho^{(s+1)} = -\kappa_2 \tanh\left(\frac{1}{2\kappa_2} B^T \nabla \Gamma^{(s+1)}(\varsigma)\right) + \kappa_1, \tag{27a}$$

$$\eth^{(s+1)} = \frac{1}{2\gamma^2} C^T \nabla \Gamma^{(s+1)}(\varsigma). \tag{27b}$$

Let s be updated as $s+1$. If $\|\Gamma^{(s+1)} - \Gamma^{(s)}\| \leq \varsigma$ where ς is a small positive number and stops at step 4; Otherwise, please return to step 1.

Here, we define a new system function involving the uncertain parameter $g(t)$ as $G_{\Gamma(t)}(\varsigma, \rho, \eth, g) = \int_t^{t+T} (\gamma^2 \beta_1 \varsigma^T \varsigma + \gamma(\varsigma, \rho, \eth)) d\varrho + \Gamma(\varsigma(t+T))$. From (20), it follows that $\Gamma(\varsigma) = E\big[G_\Gamma(\varsigma, \rho, \eth, g)\big]$. Specifically, select a set of samples according to the probability density functions (pdfs) $F_{\chi_q}(g_q)$ of the uncertain parameters $g(t)$, and then compute the value of $E\big[G_{\Gamma(t)}(\varsigma, \rho, \eth, g)\big]$ via simulating at these chosen samples. Suppose the degree of each uncertain variable g_q is updated to $2N_q - 1$; $G_{\Gamma(t)}(\varsigma, \rho, \eth, g)$ can be expressed as

$$\begin{aligned} & G_{\Gamma(t)}(\varsigma, \rho, \eth, g) \\ =& \sum_{p_1=0}^{2N_1-1} \sum_{p_2=0}^{2N_2-1} \cdots \sum_{p_m=0}^{2N_m-1} \Psi_{p_1, p_2, \ldots, p_m}(\varsigma, \rho, \eth) \prod_{q=1}^{m} g_q^{p_q}. \end{aligned} \tag{28}$$

Using the MPCM, a low-order mapping is used to approximate $G_{\Gamma(t)}(\varsigma, \rho, \eth, g)$ as follows:

$$\begin{aligned} & G'_{\Gamma(t)}(\varsigma, \rho, \eth, g) \\ =& \sum_{p_1=0}^{N_1-1} \sum_{p_2=0}^{N_2-1} \cdots \sum_{p_m=0}^{N_m-1} \Omega_{p_1, p_2, \ldots, p_m}(\varsigma, \rho, \eth) \prod_{q=1}^{m} g_q^{p_q}. \end{aligned} \tag{29}$$

On the basis of [40] (Th.2), one obtains $\Gamma(\varsigma(t)) = E\big[G_{\Gamma(t)}(\varsigma, \rho, \eth, g)\big] = E\big[G'_{\Gamma(t)}(\varsigma, \rho, \eth, g)\big]$.

Lemma 1. *The auxiliary system of controlled system* (1) *is given as* (4)*, and the value function is obtained by* (23)*. In each iteration, the MPCM is employed to sample the uncertainty space, and the policies* ρ *and* $\eth$ *obtained from Algorithm 1 remain optimal if Algorithm 1 converges.*

Proof. Using [40] (Th.2) and the relation $\Gamma(\varsigma) = E\left[G_\Gamma(\varsigma, \rho, \eth, g)\right]$, the control pairs $(\rho^*, \eth^*)$ can be derived from (10) and (11). To establish this theorem, we need to show that the optimal solution obtained by evaluating $G_{\Gamma(t)}(\varsigma, \rho, \eth, g)$ is identical to that derived from computing the reduced-order mapping $G'_{\Gamma(t)}(\varsigma, \rho, \eth, g)$. The equivalence of these two optimal solutions are proven via a contradiction method, as detailed in [39] (Th.1). □

4.2. IRL-based GCC Design with Asymmetric Constrained Inputs

Unlike Algorithm 1, an IRL-based GCC method is developed for the system (1) without relying on the knowledge of function B and C. Based on [45], two exploration signals e_1 and e_2 are, respectively, added to the control policies $\rho^{(s)}$ and $\eth^{(s)}$. Consequently, the system (4) becomes

$$\dot{\varsigma} = \chi(g)\varsigma + B(\rho + e_1) + C(\eth + e_2). \tag{30}$$

The specific steps of Algorithm 2 include the following.

Algorithm 2 IRL-based GCC algorithm for system (1) with asymmetric constrained control.

Initialization Set the initial admissible policies $\rho(0)$ and $\eth(0)$.
Step 1: Choose a set of sampling points for the uncertain variable $g(t)$ based on MPCM ([40], Section II). For each sampling point, compute the value of

$$\begin{aligned} &G_{\Gamma^{(s+1)}}\left(\varsigma, \rho^{(s)}, d^{(s)}, g\right) \\ &= \int_t^{t+T} \left(\gamma^2 \beta_1 \varsigma^T \varsigma + r\left(\varsigma(\varrho), \rho^{(s)}(\varrho), \eth^{(s)}(\varrho)\right) d\varrho \right. \\ &\quad \left. + \hat{W}_c^{(s+1)T} \varphi_c(\varsigma(t+T)\right). \end{aligned} \tag{31}$$

Step 2: Compute $\Gamma^{(s+1)}(\varsigma(t))$ by calculating the mean of $G_{\Gamma^{(s+1)}}(\cdot)$.

$$\Gamma^{(s+1)}(\varsigma(t)) = E\left[G_{\Gamma^{(s+1)}}\left(\varsigma, \rho^{(s)}, \eth^{(s)}, g\right)\right]. \tag{32}$$

Step 3: Update $\Gamma^{(s+1)}$, $\rho^{(s+1)}$ and $\eth^{(s+1)}$ through solving the HJI equation:

$$\begin{aligned} &\Gamma^{(s+1)}(\varsigma(t+T)) - \Gamma^{(s+1)}(\varsigma(t)) \\ =&E\left[\int_t^{t+T} \left(-2\kappa_2 \tanh^{-T}\left(\kappa_2^{-1}(\rho^{(s+1)} - \kappa_1)\right) R e_1 + 2\gamma^2 \eth^{(s+1)T} e_2\right) d\varrho\right] \\ &+ \int_t^{t+T} \left(-\gamma^2 \beta_1 \varsigma^T \varsigma - r^{(s)}(\varsigma, \rho, \eth)\right) d\varrho. \end{aligned} \tag{33}$$

Set s to $s+1$. If $\|\Gamma^{(s+1)} - \Gamma^{(s)}\| \leq \varsigma$, where ς is a chosen positive number, end at step 3; otherwise, return to step 1.

(1) Based on the MPCM, select a set of sampling points for the uncertain variable $g(t)$ and calculate the cumulative cost function G_Γ for the future time period at each sampling point. This function consists of the integral term containing the state penalty and control cost in the current time period, as well as the output estimate $\hat{W}_c^{(s+1)T} \varphi_c$ of the evaluation NN at the next time. (2) By calculating the mathematical expectation of the G_Γ function values at all sampling points, the updated value function estimate for this iteration, $\Gamma^{(s+1)}$,

is obtained. (3) Solve an HJI equation embedded with exploration signals e_1, e_2, while updating the value function $\Gamma^{(s+1)}$, control strategy $\rho^{(s+1)}$, and disturbance strategy $\eth^{(s+1)}$.

Taking (7) into account and utilizing (10) and (11), we have

$$\begin{aligned}
&E\left[\dot{\Gamma}^{(s+1)}(\varsigma(t))\right] \\
=&E\left[\left(\nabla\Gamma^{(s+1)\mathrm{T}}(\varsigma)\right)\left(\chi(g)\varsigma + B\rho^{(s)} + C\eth^{(s)}\right)\right. \\
&\left.+\left(\nabla\Gamma^{(s+1)\mathrm{T}}(\varsigma)\right)(Be_1 + Ce_2)\right] \\
=&E\left[-2\kappa_2\tanh^{-T}\left(\kappa_2^{-1}\left(\rho^{(s+1)} - \kappa_1\right)\right)Re_1\right. \\
&\left.+2\gamma^2\eth^{(s+1)\mathrm{T}}e_2\right] - \gamma^2\beta_1\varsigma^{\mathrm{T}}\varsigma - r^{(s)}(\varsigma,\rho,\eth).
\end{aligned} \tag{34}$$

Based on (34), one obtains

$$\begin{aligned}
&E\left[\Gamma^{(s+1)}(\varsigma(t+T))\right] - \Gamma^{(s+1)}(\varsigma(t)) \\
=&\int_t^{t+T}\left(-\gamma^2\beta_1\varsigma^T\varsigma - r^{(s)}(\varsigma,\rho,\eth)\right)d\varrho \\
&+E\left[\int_t^{t+T}(-2\kappa_2\tanh^{-T}\left(\kappa_2^{-1}\left(\rho^{(s+1)} - \kappa_1\right)\right)Re_1 + 2\gamma^2 d^{(s+1)\mathrm{T}}e_2)d\varrho\right].
\end{aligned} \tag{35}$$

Algorithm 2 presents the GCC approach via the MPCM and IRL algorithm, where the target control policies can be approximated through

$$\rho^*(\varsigma) = -\kappa_2\tanh\left(W_\rho^T\varphi_\rho(\varsigma) + \varepsilon_\rho(\varsigma)\right) + \kappa_1, \tag{36}$$

$$\eth^*(\varsigma) = W_\eth^T\varphi_\eth(\varsigma) + \varepsilon_\eth(\varsigma), \tag{37}$$

where the activation functions are denoted as $\varphi_\rho(\varsigma)$ and $\varphi_\eth(\varsigma)$, the expected weight vectors are denoted as W_ρ and $W_\eth$, and approximation errors are denoted as $\varepsilon_\rho(\varsigma)$ and $\varepsilon_\eth(\varsigma)$. The subscripts ρ and $\eth$ represent the actor NN and disturbance NN, respectively.

The control pairs in actual situations are

$$\rho^{(s+1)}(\varsigma) = -\kappa_2\tanh\left(\hat{W}_\rho^{(s+1)T}\varphi_\rho(\varsigma)\right) + \kappa_1, \tag{38}$$

$$\eth^{(s+1)}(\varsigma) = \hat{W}_\eth^{(s+1)T}\varphi_\eth(\varsigma). \tag{39}$$

Theorem 2. *The auxiliary system of controlled system* (1) *is given as* (4)*. The value function and control pairs are approximated via* (23)*,* (38) *and* (39)*, respectively. Assume Algorithm 2 converges. At each iteration, the MPCM is used to sample from the uncertain parameters;* $\hat{\Gamma}^{(s)}(\varsigma)$*,* $\hat{\rho}^{(s)}(\varsigma)$ *and* $\hat{\eth}^{(s)}(\varsigma)$ *given by* (33) *will ultimately converge to the optimal values* $\Gamma^{(*)}(\varsigma)$*,* $\rho^{(*)}(\varsigma)$*, and* $\eth^*(\varsigma)$*.*

Proof. The solution obtained using the GCC algorithm based on IRL has been proven to be the same as the solution obtained using the GCC algorithm with asymmetric constraint control based on IRL in [46] (Th. 3). The prior conclusion derived via the MPCM for each sample point holds valid as well. Therefore, the two approaches achieve identical optimal value functions. □

5. Event-Triggered Construction of Optimal GCC

Time-triggered control strategies involve substantial data transmission, which causes computational overload. In contrast to time-triggered methods, this section presents a GCC approach in the DETC mode for the stochastic system (1) based on Theorems 1–2 and Lemma 1, aiming to decrease the unnecessary consumption of communication resources.

5.1. Event-Triggered GCC Design

The control pairs will update when the triggering condition is violated with the DETC strategy. At each triggering instant, $\varsigma(s_j) = \bar{\varsigma}_j$, where s_j stands for the j th sampling moment. The event-triggered error is

$$e_j(t) = \varsigma(t) - \bar{\varsigma}_j, \quad \forall t \in [s_j, s_{j+1}) \tag{40}$$

For $j \in N$, $e_j(t) = 0$ holds when $t = s_j$. Based on (40), at triggering instants, the ETC policies become

$$\rho(\bar{\varsigma}_j) = \rho(\varsigma(t) - e_j), \tag{41}$$
$$\eth(\bar{\varsigma}_j) = \eth(\varsigma(t) - e_j). \tag{42}$$

By applying (41) and (42), and considering system (4) for all $\forall t \in [s_j, s_j + 1)$, the Hamiltonian function in (8) becomes

$$\begin{aligned}&\mathcal{H}(\varsigma, \bar{\rho}, \bar{\eth}, \nabla\Gamma(\varsigma))\\ &= E\Big[(\nabla\Gamma(\varsigma))^T(\chi(g)\varsigma + B\bar{\rho} + C\bar{\eth})\Big]\\ &\quad + \gamma^2\beta_1\varsigma^T\varsigma + \varsigma^T Q\varsigma + \mathcal{W}(\bar{\rho}) - \gamma^2\bar{\eth}^T\bar{\eth}.\end{aligned} \tag{43}$$

where $\bar{\rho} = \rho(\bar{\varsigma}_j)$ and $\bar{\eth} = \eth(\bar{\varsigma}_j)$.

Using (10) and (11), the ETC policies can be constructed as

$$\bar{\rho}^* = -\kappa_2 \tanh\left(\frac{1}{2\kappa_2} R^{-1} B^T \nabla\Gamma^*(\bar{\varsigma}_j)\right) + \kappa_1, \tag{44}$$
$$\bar{\eth}^* = \frac{1}{2\gamma^2} C^T \nabla\Gamma^*(\bar{\varsigma}_j). \tag{45}$$

with $\bar{\rho}^* = \rho^*(\bar{\varsigma}_j)$ and $\bar{\eth}^* = \eth^*(\bar{\varsigma}_j)$.

Substituting (44) and (45) into (43), the result is

$$\begin{aligned}&\mathcal{H}(\varsigma, \bar{\rho}^*, \bar{\eth}^*, \nabla\Gamma^*(\varsigma))\\ =&\gamma^2\beta_1\varsigma^T\varsigma + \varsigma^T Q\varsigma + \mathcal{W}(\bar{\rho}^*) - \gamma^2\bar{\eth}^{*T}\bar{\eth}^*\\ &+ E\Big[(\nabla\Gamma^*(\varsigma))^T(\chi(g)\varsigma + B\bar{\rho}^* + C\bar{\eth}^*)\Big].\\ =&E\Big[(\nabla\Gamma^*(\varsigma))^T\chi(g)\varsigma - \kappa_2(\nabla\Gamma^*(\varsigma))^T B \tanh\left(\frac{1}{2\kappa_2} R^{-1} B^T \nabla\Gamma^*(\bar{\varsigma}_j)\right)\\ &+ (\nabla\Gamma^*(\varsigma))^T B\kappa_1 + \frac{1}{2\gamma^2}(\nabla\Gamma^*(\varsigma))^T CC^T\nabla\Gamma^*(\bar{\varsigma}_j)\\ &- \frac{1}{4\gamma^2}(\nabla\Gamma^*(\bar{\varsigma}_j))^T CC^T\nabla\Gamma^*(\bar{\varsigma}_j)\Big] + \gamma^2\beta_1\varsigma^T\varsigma + \varsigma^T Q\varsigma + \mathcal{W}(\bar{\rho}^*).\end{aligned} \tag{46}$$

Assumption 4. *The control pairs are Lipschitz-continuous, satisfying*

$$\|\rho^* - \bar{\rho}^*\| \le l_\rho \|E[e_j(t)]\|, \tag{47}$$
$$\|\eth^* - \bar{\eth}^*\| \le l_\eth \|E[e_j(t)]\|, \tag{48}$$

where l_ρ and $l_\eth$ are positive constants, $\rho^* = \rho^*(\varsigma)$ and $\eth^* = \eth^*(\varsigma)$.

Lemma 2. *Given that Assumption 4 is valid, the following inequality holds:*

$$\begin{aligned}&\left\|\mathcal{H}(\varsigma,\bar{\rho}^*,\bar{\eth}^*,\nabla\Gamma^*(\varsigma))-\mathcal{H}(\varsigma,\rho^*,\eth^*,\nabla\Gamma^*(\varsigma))\right\|\\&\leq m_1+(l_\rho^2+\gamma^2 l_\eth^2)\left\|\left[E(e_j(t))\right]\right\|^2,\end{aligned}\tag{49}$$

where m_1 is given in (52).

Proof. Subtracting (14) from (43), we have

$$\begin{aligned}&\mathcal{H}(\varsigma,\bar{\rho}^*,\bar{\eth}^*,\nabla\Gamma^*(\varsigma))-\mathcal{H}(\varsigma,\rho^*,\eth^*,\nabla\Gamma^*(\varsigma))\\=&E\Big[\left(\nabla\Gamma^*(\varsigma)\right)^T\left(\chi(g)\varsigma+B\bar{\rho}^*+C\bar{\eth}^*\right)\Big]+\gamma^2\beta_1\varsigma^T\varsigma+r(\varsigma,\bar{\rho}^*,\bar{\eth}^*)\\&-E\Big[\left(\nabla\Gamma^*(\varsigma)\right)^T\left(\chi(g)\varsigma+B\rho^*+C\eth^*\right)\Big]-\gamma^2\beta_1\varsigma^T\varsigma-r(\varsigma,\rho^*,\eth^*)\\=&W(\bar{\rho}^*)-W(\rho^*)-\gamma^2\bar{\eth}^{*T}\bar{\eth}^*+\gamma^2\eth^{*T}\eth^*+2\gamma^2\eth^{*T}\bar{\eth}^*-2\gamma^2\eth^{*T}\eth^*\\&+E\Big[\left(\nabla\Gamma^*(\varsigma)\right)^T B(\bar{\rho}^*-\rho^*)\Big]\\=&W(\bar{\rho}^*)-W(\rho^*)-\gamma^2\left(\eth^*-\bar{\eth}^*\right)^T\left(\eth^*-\bar{\eth}^*\right)\\&+E\Big[\left(\nabla\Gamma^*(\varsigma)\right)^T B(\bar{\rho}^*-\rho^*)\Big].\end{aligned}\tag{50}$$

Based on (47), we can get

$$\|\mathcal{W}(\bar{\rho}^*)-\mathcal{W}(\rho^*)\|=M_{\mathcal{W}}\|\bar{\rho}^*-\rho^*\|\leq M_{\mathcal{W}}l_\rho\|E[e_j(t)]\|\tag{51}$$

where $M_{\mathcal{W}}=\left|2\kappa_2\tanh^{-T}\left(\frac{\rho_0-\kappa_1}{\kappa_2}\right)R\right|$ with $\rho_0=\left|\frac{\bar{\rho}^*+\rho^*}{2}\right|$. Using Young's inequality, we get $\|\mathcal{W}(\bar{\rho}^*)-\mathcal{W}(\rho^*)\|\leq\frac{M_{\mathcal{W}}^2}{2}+\frac{l_\rho^2\|E[e_j(t)]\|^2}{2}$.

Based on (50), we have

$$\begin{aligned}&\left\|\mathcal{H}(\varsigma,\bar{\rho}^*,\bar{\eth}^*,\nabla\Gamma^*(\varsigma))-\mathcal{H}(\varsigma,\rho^*,\eth^*,\nabla\Gamma^*(\varsigma)\right\|\\\leq&\frac{\left\|E\left[\nabla\Gamma^*(\varsigma)B\right]\right\|^2+l_\rho^2\left\|\left[E\left[e_j(t)\right]\right\|^2}{2}\\&+\frac{M_{\mathcal{W}}^2}{2}+\frac{l_\rho^2\|E[e_j(t)]\|^2}{2}+\gamma^2 l_\eth^2\left\|\left[E(e_j(t))\right]\right\|^2\\=&m_1+(l_\rho^2+\gamma^2 l_\eth^2)\left\|\left[E(e_j(t))\right]\right\|^2.\end{aligned}\tag{52}$$

where $m_1=\frac{\left\|E\left[\nabla\Gamma^*(\varsigma)B\right]\right\|^2}{2}+\frac{M_{\mathcal{W}}^2}{2}$. □

Theorem 3. *Assuming that $\Gamma^*(\varsigma(t))>0$ is a continuous function satisfying the improved HJI equation (14) and the OC pairs are given as (44) and (45), then system (1) is mean asymptotically stable if the following condition holds:*

$$\|E[e_j(t)]\|^2\leq\frac{\beta\tilde{\sigma}_a-\alpha^2\lambda_{\min}(Q)\|\varsigma\|^2-m_1}{l_\rho^2+2\gamma^2 l_\eth^2}=e_T,\tag{53}$$

where α is the designed parameter, $0<\alpha<1$, $\beta>0$, and e_T is the trigger threshold. m_1 is given in (52). The internal dynamic signal is

$$\dot{\tilde{\sigma}}_a=-\beta\tilde{\sigma}_a+\lambda_{\min}(Q)\|\varsigma\|^2.\tag{54}$$

Proof. The Lyapunov function is $L=l_1+l_2=\Gamma(\varsigma)+\tilde{\sigma}_a$, where $l_1=\Gamma(\varsigma)$ and $l_2=\tilde{\sigma}_a$.

In light of (43), the following holds:

$$\begin{aligned} & E\Big[\big(\nabla\Gamma^*(\varsigma)\big)^T\big(\chi(g)\varsigma + B\bar{\rho}^* + C\bar{\eth}^*\big)\Big] \\ =& -\varsigma^T Q_1\varsigma - \mathcal{W}(\bar{\rho}^*) + \gamma^2\bar{\eth}^{*T}\bar{\eth}^* \\ & + \mathcal{H}\big(\varsigma, \bar{\rho}^*, \bar{\eth}^*, \nabla\Gamma^*(\varsigma)\big). \end{aligned} \tag{55}$$

By substituting $\bar{\rho}^*$ from (44) and $\bar{\eth}^*$ from (45) into (1), and relying on (55) and Lemma 2, one obtains

$$\begin{aligned} & E[\dot{\Gamma}^*(\varsigma(t))] \\ =& E\Big[\big(\nabla\Gamma^*(\varsigma)\big)^T\big(\chi(g)\varsigma + B\bar{\rho}^* + C\bar{\eth}^*\big) + \big(\nabla\Gamma^*(\varsigma)\big)^T C\big(d(\varsigma) - \bar{\eth}^*\big)\Big] \\ =& -\varsigma^T Q_1\varsigma - \mathcal{W}(\bar{\rho}^*) + \gamma^2\bar{\eth}^{*T}\bar{\eth}^* + E\Big[2\gamma^2\bar{\eth}^{*T}\big(d(\varsigma) - \bar{\eth}^*\big)\Big] \\ & + \mathcal{H}\big(\varsigma, \bar{\rho}^*, \bar{\eth}^*, \nabla\Gamma^*(\varsigma)\big) \\ \leq& E\Big[\gamma^2\big(\bar{\eth}^* - \bar{\eth}^*\big)^T\big(\bar{\eth}^*(\varsigma) - \bar{\eth}^*\big) - \varsigma^T Q\varsigma + m_1 + (l_\rho^2 + \gamma^2 l_{\bar{\eth}}^2)\big\|E[e_j(t)]\big\|^2\Big] \\ \leq& E\Big[-\alpha^2\lambda_{\min}(Q)\|\varsigma\|^2 + \gamma^2 l_{\bar{\eth}}^2\big\|E[e_j(t)]\big\|^2 \\ & + m_1 + (l_\rho^2 + \gamma^2 l_{\bar{\eth}}^2)\big\|E[e_j(t)]\big\|^2 - (1-\alpha^2)\lambda_{\min}(Q)\|\varsigma\|^2\Big], \end{aligned} \tag{56}$$

where $\lambda_{\min}(Q)$ denotes the smallest eigenvalue of Q.

According to (54), we obtain

$$\dot{l}_2 = -\beta\tilde{\sigma}_a + \lambda_{\min}(Q)\|\varsigma\|^2, \tag{57}$$

Taking (56) and (57) into account together, if the condition (53) holds, we derive $E[L] \leq -\alpha^2\lambda_{\min}(Q)\|\varsigma\|^2 < 0$, for any $\varsigma = 0$. According to [36] (Lemma 1) and Assumption 3, the system (1) achieves mean asymptotic stability. □

Theorem 4. *Consider system (2); when adopting the optimal ETC policy $\rho^*(\bar{\varsigma}_j)$ from (44) and employing the OC signal $\rho^* = \bar{\rho}^*$, the following result is obtained:*

$$\bar{\Theta}(\varsigma(0)), \rho^*) \leq \Theta^*\big(\varsigma(0), \rho^*, \eth^*\big). \tag{58}$$

Proof. With respect to any admissible control function ρ, by referring to (4), (6), and (7) and calculating the system dynamics in (1), one obtains

$$\begin{aligned} \bar{\Theta}(\varsigma(0), \rho) &= E\Big[\int_0^\infty \Big(\varsigma^T Q\varsigma + \mathcal{W}(\rho) + \dot{\Gamma}^*(\varsigma)\Big)d\varrho\Big] + \Theta^*(\varsigma(0)) \\ &= E\Big[\int_0^\infty \Big(\big(\nabla\Gamma^*(\varsigma)\big)^T(\chi(g)\varsigma + B\rho + Cd) + \varsigma^T Q\varsigma \\ &\quad + \mathcal{W}(\rho)\Big)d\varrho\Big] + \Theta^*(\varsigma(0)). \end{aligned} \tag{59}$$

On the basis of (8), one obtains

$$\begin{aligned} & E\Big[\big(\nabla\Gamma^*(\varsigma)\big)^T(\chi(g)\varsigma)\Big] \\ =& E\Big[-\big(\nabla\Gamma^*(\varsigma)\big)^T B\rho^*(\varsigma) - \gamma^2\eth^{*T}\eth^* - \gamma^2\beta_1\varsigma^T\varsigma \\ & -\varsigma^T Q\varsigma - \mathcal{W}(\rho^*)\Big]. \end{aligned} \tag{60}$$

Based on (60), we have

$$\begin{aligned} & E\Bigg[\int_0^\infty \Big((\nabla\Gamma^*(\varsigma))^T(\chi(g)\varsigma + B\rho + Cd) \\ & \quad +\varsigma^T Q\varsigma + \mathcal{W}(\rho)\Big)d\varrho\Bigg] \\ =& E\Big[(\nabla\Gamma^*(\varsigma))^T B(\rho-\rho^*) - \gamma^2\eth^{*T}\eth^* \\ & \quad -\gamma^2\beta_1\varsigma^T\varsigma + \mathcal{W}(\rho) - \mathcal{W}(\rho^*) + 2\gamma^2\eth^{*T}d\Big]. \end{aligned} \tag{61}$$

According to (61), Assumption 2, the Cauchy–Schwarz inequality, and $2\gamma^2\eth^{*T}d \leq \gamma^2\eth^{*T}\eth^* + \gamma^2 d^T d$, we obtain

$$\begin{aligned} & E\Bigg[\int_0^\infty \Big(\nabla\Gamma^{*T}(\varsigma)(\chi(g)\varsigma + B\rho + Cd) + \varsigma^T Q\varsigma + \mathcal{W}(\rho)\Big)d\varrho\Bigg] \\ \leq& E\Big[(\nabla\Gamma^*(\varsigma))^T B(\rho-\rho^*) + \mathcal{W}(\rho) - \mathcal{W}(\rho^*)\Big]. \end{aligned} \tag{62}$$

Applying $\rho = \bar{\rho}^*$ to (62), one obtains the result (58). □

5.2. NN-Based Control Design

Based on Theorem 4, we have $\bar{\Theta}(\varsigma(0), \rho^*) \leq \Theta^*\big(\varsigma(0), \rho^*(\varsigma), \eth^*(\varsigma)\big)$, and the GCC algorithm using IRL is used to achieve an approximation of $\Theta^*\big(\varsigma(0), \rho^*(\varsigma), \eth^*(\varsigma)\big)$. In the proposed method, $\Gamma^*(\varsigma)$ is approximated by (22). Based on the work in [47], the optimal GCC pairs are given as

$$\rho^*(\varsigma) = -\kappa_2 \tanh\Big(W_\rho^T\varphi_\rho(\bar{\varsigma}_j) + \varepsilon_{e,\rho}(\bar{\varsigma}_j, e_j)\Big) + \kappa_1 \quad \forall t \in [s_j, s_{j+1}), \tag{63}$$

$$\eth^*(\varsigma) = W_\eth^T\varphi_\eth(\bar{\varsigma}_j) + \varepsilon_{e,\eth}(\bar{\varsigma}_j, e_j) \quad \forall t \in [s_j, s_{j+1}), \tag{64}$$

where $\varepsilon_{e,\rho} = W_\rho^T\big[\varphi_\rho(\bar{\varsigma}_j + e_j) - \varphi_\rho(\bar{\varsigma}_j)\big] + \varepsilon_\rho(\bar{\varsigma}_j + e_j)$ and $\varepsilon_{e,\eth} = W_\eth^T[\varphi_\eth(\bar{\varsigma}_j + e_j) - \varphi_\eth(\bar{\varsigma}_j)] + \varepsilon_\eth(\bar{\varsigma}_j + e_j)$ denote the event-sampled approximation errors.

Based on (14), (63) and (64), the HJI equation takes the form

$$\begin{aligned} & E\Big[(W_c^T\nabla\varphi_c(\varsigma))\chi(g)\varsigma + W_c^T\nabla\varphi_c(\varsigma)B\rho^*\Big] + \gamma^2\beta_1\varsigma^T\varsigma \\ & + \varsigma^T Q\varsigma + \mathcal{W}(\rho^*) + \gamma^2\Big(W_\eth^T\varphi_\eth(\bar{\varsigma}_j)\Big)^T\Big(W_\eth^T\varphi_\eth(\bar{\varsigma}_j)\Big) \\ & + \varepsilon_{HJI} \\ =& \varsigma^T Q_1\varsigma + (W_c^T\nabla\varphi_c(\varsigma))\chi(g)\varsigma + \kappa_2^2\bar{R}^T\ln\Big(\bar{\mathbf{1}} - \tanh^2(N_1)\Big) \\ & + \gamma^2 P_1^T P_1 + 2\kappa_2 N_1^T R\kappa_1 \\ & + \varepsilon_{HJI1} \\ =& 0, \end{aligned} \tag{65}$$

where $\varepsilon_{HJI} = \nabla\varepsilon_c^T\chi(g)\varsigma + \nabla\varepsilon_c^T B\rho^* + 2\gamma^2 W_\eth^T\varphi_\eth(\bar{\varsigma}_j)\varepsilon_{e,\eth} + \gamma^2\varepsilon_{e,\eth}^T\varepsilon_{e,\eth}$. $P_1 = W_\eth^T\varphi_\eth(\bar{\varsigma}_j)$, $\varepsilon_{HJI1} = \nabla\varepsilon_c^T\chi(g)\varsigma + \kappa_2^2\bar{R}^T\ln\Big(\bar{\mathbf{1}} - \tanh^2(N_1 + \varepsilon_{e,\rho})\Big) - \kappa_2^2\bar{R}^T\ln\Big(\bar{\mathbf{1}} - \tanh^2(N_1)\Big) + \gamma^2\varepsilon_{e,\eth}^T\varepsilon_{e,\eth} + 2\gamma^2 W_\eth^T\varphi_\eth\varepsilon_{e,\eth} + 2\kappa_2\varepsilon_{e,\rho}R\kappa_1$ and $N_1 = W_\rho^T\varphi_\rho(\bar{\varsigma}_j)$.

Assumption 5. *The actor, critic, and disturbance NNs, the NN weight vectors, activation function, reconstruction errors, and gradient of relevant parameters are bounded, which satisfy* $\|W_c\| \leq c_1$, $\|W_\rho\| \leq c_2$, $\|W_\eth\| \leq c_3$, $\|W_a\|_F \leq c_2'$, $\|W_\eth\|_F \leq c_3'$; $\|\varphi_\rho\| \leq \varphi_{\rho M}$, $\|\varphi_\eth\| \leq \varphi_{\eth M}$, $\|\nabla\varphi_c\| \leq \varphi_{cM}$, $\varepsilon_c(\varsigma) \leq \varepsilon_{cM}$, $\varepsilon_\rho(\varsigma) \leq \varepsilon_{\rho M}$, $\varepsilon_\eth(\varsigma) \leq \varepsilon_{\eth M}$, $\|\varepsilon_{ca}(\bar{\varsigma}_j)\| \leq b_{ca}$ *and* $\nabla\varepsilon_c \leq c_\rho$, *respectively, where* c_1, c_2, c_3, c_2', c_3', $\varphi_{\rho M}$, $\varphi_{\eth M}$, φ_{cM}, ε_{cM}, $\varepsilon_{\rho M}$, $\varepsilon_{\eth M}$, b_{ca} *and* c_ρ *are positive constants.*

The estimates for Γ, ρ and $\eth$ are expressed as

$$\hat{\Gamma}(\varsigma) = \hat{W}_c^T \varphi_c(\varsigma), \tag{66a}$$

$$\hat{\rho}(\bar{\varsigma}_j) = -\kappa_2 \tanh\left(\hat{W}_\rho^T \varphi_\rho(\bar{\varsigma}_j)\right) + \kappa_1, \tag{66b}$$

$$\hat{\eth}(\bar{\varsigma}_j) = \hat{W}_\eth^T \varphi_\eth(\bar{\varsigma}_j), \tag{66c}$$

where $\hat{W}_c$, $\hat{W}_\rho$ and $\hat{W}_\eth$ stand for the estimates of W_c, W_ρ and $W_\eth$, respectively.

Using (23) and (31)–(33), the residual error can be given as

$$\begin{aligned}
\check{e}_B &= E\left[\bar{G}_{\Gamma(s+1)}(\varsigma, \hat{\rho}, \hat{\eth}, g)\right] - \hat{W}_c^T \varphi_c(\varsigma) \\
&= E\left[\int_t^{t+T} \left(\varsigma^T Q_1 \varsigma + \mathcal{W}(\hat{\rho}) - \gamma^2 \hat{\eth}^T \hat{\eth}\right) d\varrho\right] \\
&\quad - \kappa_2 \int_t^{t+T} \left((2e_1^T R) \bigotimes \varphi_\rho(\bar{\varsigma}_j)\right) d\varrho \cdot \mathrm{vec}(\hat{W}_\rho) \\
&\quad - \int_t^{t+T} \left((2\gamma^2 e_2^T) \bigotimes \varphi_\eth(\bar{\varsigma}_j)\right) d\varrho \cdot \mathrm{vec}(\hat{W}_\eth) \\
&\quad + E\left[\hat{W}_c^T \varphi_c(\varsigma(t+T))\right] - \hat{W}_c^T \varphi_c(\varsigma).
\end{aligned} \tag{67}$$

where $\hat{\rho} = \hat{\rho}(\bar{\varsigma}_j)$ and $\hat{\eth} = \hat{\eth}(\bar{\varsigma}_j)$. Design a new augmented weight vector as $\hat{W} = \left[\hat{W}_c, \mathrm{vec}(\hat{W}_\rho), \mathrm{vec}(\hat{W}_\eth)\right]^T$ and the target weight vector $W = \left[W_c, \mathrm{vec}(W_\rho), \mathrm{vec}(W_\eth)\right]^T$. Denote

$$\begin{aligned}
S_1 &= E\left[\int_t^{t+T} \left(\varsigma^T Q_1 \varsigma + \mathcal{W}(\hat{\rho}) - \gamma^2 \hat{\eth}^T \hat{\eth}\right) d\varrho\right], \\
S_2 &= E[\varphi_c(\varsigma(t+T)) - \varphi_c(\varsigma(t))], \\
S_3 &= -\kappa_2 \int_t^{t+T} \left((2e_1^T R) \bigotimes \varphi_\rho(\bar{\varsigma}_j)\right) d\varrho, \\
S_4 &= -\int_t^{t+T} \left((2\gamma^2 e_2^T) \bigotimes \varphi_\eth(\bar{\varsigma}_j)\right) d\varrho.
\end{aligned} \tag{68}$$

Then, we have

$$\check{e}_B = E\left[\bar{S}\hat{W} + S_1\right], \tag{69}$$

with $\bar{S} = [S_2, S_3, S_4]$.

Let $\xi_1 = \frac{1}{2}\check{e}_B^T \check{e}_B$, α_1 denote an adaptive gain parameter; the weight adjustment rule is developed as

$$\dot{\hat{W}} = -E\left[\alpha_1 \frac{\bar{S}^T}{(\bar{S}\bar{S}^T + 1)^2} \check{e}_B\right], \tag{70}$$

Assumption 6. *$\tilde{\eta}_1$ is continuously excited within the interval of $[t, t+T]$, and one obtains*

$$\gamma_1 I \leq \int_t^{t+T} \tilde{\eta}_1 \tilde{\eta}_1^T dt \leq \gamma_2 I, \tag{71}$$

where $\tilde{\eta}_1 = \left(\bar{S}^T / \left(\bar{S}\bar{S}^T + 1\right)\right)$, $T > 0$, $\gamma_1 \leq \gamma_2$, and γ_1 and γ_2 are constants.

Considering $\dot{\varsigma} = \chi(g)\varsigma + B\hat{\rho} + C\hat{\eth}$ and $\tilde{W} = W - \hat{W}$, we define the augmented system state as $\mho = \left[\varsigma^T, \bar{\varsigma}_j^T, \tilde{W}^T\right]^T$, we have

$$\dot{\mho}(t) = \begin{bmatrix} \chi(g)\varsigma + B\hat{\rho} + C\hat{\eth} \\ 0 \\ E\left[\alpha_1 \frac{\bar{S}^T}{(\bar{S}\bar{S}^T+1)^2}\check{e}_B\right] \end{bmatrix}, \quad t \in [s_j, s_{j+1}) \tag{72}$$

Furthermore, one obtains

$$\mho(t) = \mho(t^-) + \begin{bmatrix} 0 \\ \varsigma - \bar{\varsigma}_j \\ 0 \end{bmatrix}, \quad t = s_{j+1} \tag{73}$$

Here, $\mho(t^-) = \lim_{\lambda \to 0} \mho(t - \lambda)$, where $\lambda \in (0, s_{j+1} - s_j)$.

Remark 5. *The residual error $\check{e}_B$, which requires minimization in (67), is derived from the expectation of the function $E\left[\bar{G}_{\Gamma(s+1)}(\varsigma, \hat{\rho}, \hat{\eth}, g)\right]$. Detailed procedures for calculating this expectation are provided in [40].*

5.3. Stability Analysis

Theorem 5. *The ETC policies are given as (66b) and (66c) for the dynamical system (4), and the weight adjustment law $\hat{W}$ is designed as (70). The system states and NN weight estimation errors are uniformly bounded (UB) in the mean. The triggering condition is designed as*

$$\|E[e_j(t)]\|^2 \leq \frac{\beta\tilde{\sigma}_b - \varpi^2\lambda_{\min}(Q_1)\|\varsigma\|^2 - \Psi(\bar{\varsigma}_j)}{l_{\rho\eth}} - \frac{C_0 + \gamma^2\left\|\hat{\eth}\right\|^2}{l_{\rho\eth}} = e_T, \tag{74}$$

where C_0 and $l_{\rho\eth}$ are defined in (81), $\varpi \in (0,1)$. The internal dynamic signal is given by

$$\dot{\tilde{\sigma}}_b = -\beta\tilde{\sigma}_b + \lambda_{\min}(Q_1)\|\varsigma\|^2. \tag{75}$$

Proof. A Lyapunov function is chosen as

$$\Lambda_\Gamma(t) = \Lambda_1(t) + \Lambda_2(t) + \Lambda_3(t) + \Lambda_4(t), \tag{76}$$

with $\Lambda_1(t) = E[\Gamma^*(\varsigma)]$, $\Lambda_2(t) = E[\Gamma^*(\bar{\varsigma}_j)]$, $\Lambda_3(t) = E\left[\frac{1}{2\alpha_1}\tilde{W}^T\tilde{W}\right]$ and $\Lambda_4(t) = \tilde{\sigma}_b$.

Case 1: When $\forall t \in [s_j, s_{j+1})$, $j \in N$, based on (4), (66b) and (66c), one obtains

$$\begin{aligned} \dot{\Lambda}_1(t) &= E\left[(\nabla\Gamma^*(\varsigma))^T\left(\chi(g)\varsigma + B\hat{\rho} + C\hat{\eth}\right)\right], \\ \dot{\Lambda}_2(t) &= 0, \\ \dot{\Lambda}_3(t) &= E\left[\frac{1}{\alpha_1}\tilde{W}^T\dot{\tilde{W}}\right], \\ \dot{\Lambda}_4(t) &= \dot{\tilde{\sigma}}_b. \end{aligned}$$

Based on (22), we can obtain

$$\nabla\Gamma^*(\bar{\varsigma}_j) = \left(\nabla\varphi_c(\bar{\varsigma}_j)\right)^T W_c + \nabla\varepsilon_c(\bar{\varsigma}_j). \tag{77}$$

According to $2a^Tb \le \|a\|^2 + \|b\|^2$ and Assumption 5, we get

$$\begin{aligned}
\dot{\Lambda}_1(t) &= E\Big[(\nabla\Gamma^*(\varsigma))^T(\chi(g)\varsigma + B\rho^* + C\eth^*)\Big] \\
&\quad + (\nabla\Gamma^*(\varsigma))^T B(\hat{\rho} - \rho^*) \\
&\quad + (\nabla\Gamma^*(\varsigma))^T C(\hat{\eth} - \eth^*) \\
&= E\Big[\gamma^2\eth^{*T}\eth^* - \gamma^2\beta_1\varsigma^T\varsigma \\
&\quad -\varsigma^T Q\varsigma - \mathcal{W}(\rho^*) + (\nabla\Gamma^*(\varsigma))^T B(\hat{\rho} - \rho^*) \\
&\quad +2\gamma^2\eth^{*T}(\hat{\eth} - \eth^*)\Big] \\
&\le E\Big[\gamma^2\hat{\eth}^T\hat{\eth} - \varsigma^T Q_1\varsigma - \mathcal{W}(\rho^*) \\
&\quad +\frac{1}{2}\|B(\hat{\rho} - \rho^*)\|^2 - \gamma^2\left\|\eth^* - \hat{\eth}\right\|^2 + \left(c_\rho^2 + \varphi_{cM}^2 c_1^2\right)\Big] \\
&\le E\Big[\gamma^2\hat{\eth}^T\hat{\eth} - \varsigma^T Q_1\varsigma - \mathcal{W}(\rho^*) \\
&\quad +\frac{1}{2}\|B(\hat{\rho} - \rho^*)\|^2 + \left(c_\rho^2 + \varphi_{cM}^2 c_1^2\right)\Big]
\end{aligned} \tag{78}$$

Substituting (77) into (10), one obtains

$$\rho^*(\bar{\varsigma}_j) = -\kappa_2\tanh(\mathcal{N}^*) + \varepsilon_{ca}(\bar{\varsigma}_j) + \kappa_1,\ t \in [s_l, s_{l+1})) \tag{79a}$$

$$\hat{\rho}(\bar{\varsigma}_j) = -\kappa_2\tanh\left(\frac{1}{2\kappa_2}B^T\left(\nabla\varphi_c(\bar{\varsigma}_j)\right)^T\hat{W}_c\right) + \kappa_1, \tag{79b}$$

where $\mathcal{N}^* = (1/(2\kappa_2))R^{-1}B^T\left(\nabla\varphi_c(\bar{\varsigma}_j)\right)^T W_c$ and $\varepsilon_{ca}(\bar{\varsigma}_j) = -(1/2)(I_q - \mathfrak{D}(\epsilon(\bar{\varsigma}_j)))R^{-1}$ $B^T(\bar{\varsigma}_j)\nabla\varepsilon_c(\bar{\varsigma}_j)$ with $\mathfrak{D}(\epsilon(\bar{\varsigma}_j)) = \mathrm{diag}\{\tanh^2(\epsilon_i(\bar{\varsigma}_j))\}, i = 1, 2, \ldots, q.\ \epsilon(\bar{\varsigma}_j) = [\epsilon_1(\bar{\varsigma}_j), \epsilon_2(\bar{\varsigma}_j),$ $\ldots, \epsilon_q(\bar{\varsigma}_j)]^T \in R^q$, $\epsilon(\bar{\varsigma}_j)$ is selected between $\mathcal{N}^*$ and N^* and $\hat{\rho}(\bar{\varsigma}_j)$ is an approximate value of $\rho^*(\bar{\varsigma}_j)$.

According to $\|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2$, (79a) and (79b), we get

$$\begin{aligned}
&\|\hat{\rho} - \bar{\rho}^* + \bar{\rho}^* - \rho^*\|^2 \\
\le& 2\|\hat{\rho} - \bar{\rho}^*\|^2 + 2\|\bar{\rho}^* - \rho^*\|^2 \\
\le& 2\|\kappa_2\tanh\left(\frac{1}{2\kappa_2}R^{-1}B^T\left(\nabla\varphi_c(\bar{\varsigma}_j)\right)^T W_c\right) \\
&- \kappa_2\tanh\left(\frac{1}{2\kappa_2}R^{-1}B^T\left(\nabla\varphi_c(\bar{\varsigma}_j)\right)^T\hat{W}_c\right) - \varepsilon_{ca}(\bar{\varsigma}_j)\|^2 + 2\|l_\rho E[e_j(t)]\|^2 \\
\le& 4\|\kappa_2\tanh\left(\frac{1}{2\kappa_2}R^{-1}B^T\left(\varphi_c(\bar{\varsigma}_j)\right)^T W_c\right) \\
&- \kappa_2\tanh\left(\frac{1}{2\kappa_2}R^{-1}B^T\left(\varphi_c(\bar{\varsigma}_j)\right)^T\hat{W}_c\right)\|^2 + 4\|\varepsilon_{ca}(\bar{\varsigma}_j)\|^2 \\
&+ 2\|l_\rho E[e_j(t)]\|^2 \\
\le& 8\|\kappa_2\tanh\left(\frac{1}{2\kappa_2}R^{-1}B^T\left(\varphi_c(\bar{\varsigma}_j)\right)^T W_c\right)\|^2 \\
&+ 8\|\kappa_2\tanh\left(\frac{1}{2\kappa_2}R^{-1}B^T\left(\varphi_c(\bar{\varsigma}_j)\right)^T\hat{W}_c\right)\|^2 + 4\|\varepsilon_{ca}(\bar{\varsigma}_j)\|^2 \\
&+ 2\|l_\rho E[e_j(t)]\|^2 \\
\le& 16\kappa_2^2 + 4b_{ca}^2 + 2l_\rho^2\left\|E[e_j(t)]\right\|^2.
\end{aligned} \tag{80}$$

As proved in [48], $\mathcal{W}(\rho^*) > 0$, and one has

$$\begin{aligned}\dot{\Lambda}_1(t) &\le E\Big[\gamma^2\hat{\eth}^T\hat{\eth} - \varsigma^T Q_1\varsigma + \Big(c_\rho^2 + \varphi_{cM}^2 c_1^2\Big) \\ &\quad + \frac{1}{2}\|B\|^2\Big(16\kappa_2^2 + 4b_{ca}^2 + 2l_\rho^2 E[e_j(t)]^2\Big)\Big] \\ &\le E\Big[\gamma^2\hat{\eth}^T\hat{\eth} - \varsigma^T Q_1\varsigma + l_{\rho\eth}\|E[e_j(t)]\|^2 + C_0\Big] \end{aligned} \tag{81}$$

where $C_0 = \Big(c_\rho^2 + \varphi_{cM}^2 c_1^2\Big) + \frac{1}{2}\|B\|^2\big(16\kappa_2^2 + 4b_{ca}^2\big)$ and $l_{\rho\eth} = \|B\|^2 l_\rho^2$.

According to (12), we get

$$\begin{aligned}\mathcal{W}(\hat{\rho}) &= 2\kappa_2\int_{\kappa_1}^{-\kappa_2\tanh\left(\hat{W}_\rho^T\varphi_\rho(\bar{\varsigma}_j)\right)+\kappa_1}\tilde{\rho}_a^{-T}\Big(\kappa_2^{-1}(\vartheta-\kappa_1)\Big)R\mathrm{d}\vartheta \\ &= 2\kappa_2\int_{\kappa_1}^{-\kappa_2\tanh(\hat{W}_\rho^T\varphi_\rho(\bar{\varsigma}_j))+\kappa_1}\tilde{\rho}_a^{-T}\Big(\kappa_2^{-1}(\vartheta-\kappa_1)\Big)R\mathrm{d}\vartheta \\ &= 2\kappa_2^2\int_0^{-\tanh(\hat{W}_\rho^T\varphi_\rho(\bar{\varsigma}_j))}\tilde{\rho}_a^{-T}\Big(\kappa_2^{-1}(\vartheta-\kappa_1)\Big)R\mathrm{d}\Big(\kappa_2^{-1}(\vartheta-\kappa_1)\Big) \\ &= 2\kappa_2^2\varphi_\rho(\bar{\varsigma}_j)^T\hat{W}_\rho R\tanh(\hat{W}_\rho^T\varphi_\rho(\bar{\varsigma}_j)) + \kappa_2^2\bar{R}^T\ln\Big(\bar{\mathbf{1}} - \tanh^2(\hat{W}_\rho^T\varphi_\rho(\bar{\varsigma}_j))\Big). \end{aligned} \tag{82}$$

By calculating formula (69), the residual error $\check{e}_B$ is re-expressed as follows:

$$\begin{aligned}\check{e}_B &= E\Big[\bar{S}(W-\tilde{W}) + S_1\Big] \\ &= E\Big[-\bar{S}\tilde{W} + \bar{S}W + \int_t^{t+T}\Big(\varsigma^T Q_1\varsigma + \mathcal{W}(\hat{\rho}) - \hat{U}_2\Big)d\varrho\Big], \end{aligned} \tag{83}$$

$$\mathcal{W}(\hat{\rho}) = 2\kappa_2^2\hat{N}^T R\tanh(\hat{N}) + \kappa_2^2\bar{R}^T\ln\Big(\bar{\mathbf{1}} - \tanh^2(\hat{N})\Big), \tag{84}$$

$$\hat{N} = \hat{W}_\rho^T\varphi_\rho(\bar{\varsigma}_j), \tag{85}$$

$$\hat{U}_2 = \gamma^2\hat{P}^T\hat{P}, \tag{86}$$

$$\hat{P} = \hat{W}_\eth^T\varphi_\eth(\bar{\varsigma}_j). \tag{87}$$

Based on (65), one obtains

$$\check{e}_B = E\Big[\int_t^{t+T}\Big(-\mathcal{W}(\bar{\rho}) + \mathcal{W}(\hat{\rho}) + \bar{U}_2 - \hat{U}_2 - (W_c^T\nabla\varphi_c(\varsigma))\chi(g)\varsigma - \varepsilon_{HJI1} + Z_0\Big)d\varrho - \bar{S}\tilde{W} + \bar{S}W\Big], \tag{88}$$

where

$$\begin{aligned} Z_0 &= -2\kappa_2 N_1^T R\kappa_1 - 2\gamma^2 P_1^T P_1 + 2\kappa_2^2 N_1^T R\tanh(N_1), \\ \mathcal{W}(\bar{\rho}) &= 2\kappa_2^2 N_1^T R\tanh(N_1) + \kappa_2^2\bar{R}^T\ln\Big(\bar{\mathbf{1}} - \tanh^2(N_1)\Big), \\ \bar{U}_2 &= \gamma^2 P_1^T P_1. \end{aligned} \tag{89}$$

where $\varepsilon_{HJI1} \le \varepsilon_h$, with ε_h being a positive number.

Denote $\bar{\mathcal{F}}_1 = \int_t^{t+T}(\bar{U}_2 - \mathcal{W}(\bar{\rho}) - \varepsilon_{HJI1} + Z_0)d\varrho$, $\bar{\mathcal{F}}_2 = \int_t^{t+T}(\mathcal{W}(\hat{\rho}) - \hat{U}_2)d\varrho$ and $\bar{\mathcal{F}}_3 = \int_t^{t+T}(-W_c^T\nabla\varphi_c(\varsigma)\chi(g)\varsigma)d\varrho$, where T is a small positive number. Then, we have $\bar{\mathcal{F}}_1 = T(\bar{U}_2 - \mathcal{W}(\bar{\rho}) - \varepsilon_{HJI1} - Z_0)$, $\bar{\mathcal{F}}_2 = T(\mathcal{W}(\hat{\rho}) - \hat{U}_2)$ and $\bar{\mathcal{F}}_3 = T(-W_c^T\nabla\varphi_c\chi(g)\varsigma)$. Based on Assumption 5, we have $\|\bar{\mathcal{F}}_1\| \le d_1$ with $d_1 > 0$.

Then, we have

$$\check{e}_B = E\left[\check{\mathcal{F}} - \bar{S}\tilde{W} + \bar{S}W\right]. \tag{90}$$

where $\check{\mathcal{F}} = \bar{\mathcal{F}}_1 + \bar{\mathcal{F}}_2 + \bar{\mathcal{F}}_3$.

Based on (70) and (83), we have

$$\begin{aligned} E\left[\frac{1}{\alpha_1}\tilde{W}^T\dot{\tilde{W}}\right] &= E\left[\frac{\tilde{W}^T\bar{S}^T}{(\bar{S}\bar{S}^T+1)^2}\left(\check{\mathcal{F}} - \bar{S}\tilde{W} + \bar{S}W\right)\right] \\ &\le E\Bigg[-\tilde{W}^T\tilde{\eta}_1\tilde{\eta}_1^T\tilde{W} + \tilde{W}^T\tilde{\eta}_1\tilde{\eta}_1^T W + \frac{1}{2}d_1\|\tilde{W}\| + \frac{1}{4}\|\tilde{W}\|^2 \\ &\quad + \Psi(\bar{\varsigma}_j) + \frac{1}{4}\|\tilde{W}\|^2 + \frac{1}{4}\lambda_m^2\|\varsigma\|^2\Bigg] \\ &\le E\Bigg[-\left(\frac{\gamma_1}{T} - \frac{1}{2}\right)\|\tilde{W}\|^2 + \Psi(\bar{\varsigma}_j) + \frac{1}{4}\lambda_m^2\|\varsigma\|^2 \\ &\quad + \left(\frac{\gamma_2}{T}\sqrt{c_1^2 + c_2'^2 + c_3'^2} + \frac{d_1}{2}\right)\|\tilde{W}\|\Bigg] \end{aligned} \tag{91}$$

where $\tilde{\eta}_1 = \left(\bar{S}^T/\left(\bar{S}\bar{S}^T+1\right)\right)$, $\Psi(\bar{\varsigma}_j) = \frac{T^2}{4}\left(\mathcal{W}(\hat{\rho}) - -\gamma^2\hat{\eth}^T(\bar{\varsigma}_j)\hat{\eth}(\bar{\varsigma}_j)\right)^2$ and $\lambda_m = Tc_1\varphi_{cM}b_a$.

Using Equations (81) and (91), $\dot{\Lambda}_\Gamma(t)$ becomes

$$\begin{aligned} \dot{\Lambda}_\Gamma(t) \le E\Big[&\gamma^2\hat{\eth}^T\hat{\eth} - \varsigma^T Q_1\varsigma + l_{\rho\eth}\|E[e_j(t)]\|^2 + C_0 \\ &-\left(\frac{\gamma_1}{T} - \frac{1}{2}\right)\|\tilde{W}\|^2 + \Psi(\bar{\varsigma}_j) + \frac{1}{4}\lambda_m^2\|\varsigma\|^2 \\ &+\left(\frac{\gamma_2}{T}\sqrt{c_1^2 + c_2'^2 + c_3'^2} + \frac{d_1}{2}\right)\|\tilde{W}\| + \dot{\check{\sigma}}_b\Big], \end{aligned} \tag{92}$$

If the triggering conditions (74) and (75) hold, one obtains

$$\begin{aligned} \dot{\Lambda}_\Gamma(t) &\le E\Big[\gamma^2\hat{\eth}^T\hat{\eth} - \varsigma^T Q_1\varsigma + l_{\rho\eth}\|E[e_j(t)]\|^2 + C_0 \\ &\quad -\left(\frac{\gamma_1}{T} - \frac{1}{2}\right)\|\tilde{W}\|^2 + \Psi(\bar{\varsigma}_j) + \frac{1}{4}\lambda_m^2\|\varsigma\|^2 \\ &\quad +\left(\frac{\gamma_2}{T}\sqrt{c_1^2 + c_2'^2 + c_3'^2} + \frac{d_1}{2}\right)\|\tilde{W}\| + \dot{\check{\sigma}}_b\Big] \\ &\le E\Bigg[-\left(\varpi^2\lambda_{\min}(Q_1) - \frac{1}{4}\lambda_m^2\right)\|\varsigma\|^2 - \left(\frac{\gamma_1}{T} - \frac{1}{2}\right) \\ &\quad \times\left(\|\tilde{W}\| - \frac{2\gamma_2\sqrt{c_1^2 + c_2'^2 + c_3'^2} + Td_1}{2(2\gamma_1 - T)}\right)^2 + m\Bigg], \end{aligned} \tag{93}$$

where $m = \frac{\left(2\gamma_2\sqrt{c_1^2+c_2'^2+c_3'^2}+Td_1\right)^2}{8T(2\gamma_1-T)}$.

We select the interval T and the parameter γ_1, ϖ under the condition that $\varpi^2\lambda_{\min}(Q_1) - 1/4\lambda_m^2 > 0$. Let parameter T be set such that $\gamma_1 > T$; then $\dot{\Lambda}_\Gamma(t) < 0$ given that

$$\|\varsigma\| > \sqrt{\frac{m}{\varpi^2\lambda_{\min}(Q_1) - \frac{1}{4}\lambda_m^2}},$$

$$\|\tilde{W}\| > \sqrt{\frac{2Tm}{2\gamma_1 - T}} + \frac{2\gamma_2\sqrt{c_1^2 + c_2'^2 + c_3'^2} + Td_1}{2(2\gamma_1 - T)}.$$

Case 2: For any $\forall t = s_{j+1}$, the derivation of $\Lambda_\Gamma(t)$ is given by

$$\begin{aligned}&\Delta\Lambda_\Gamma(t)\\ =&E\left[\frac{1}{2\alpha_1}\tilde{W}^T\left(\bar{\varsigma}(s_j^+)\right)\tilde{W}\left(\bar{\varsigma}(s_j^+)\right)\right]-E\left[\frac{1}{2\alpha_1}\tilde{W}^T\left(\bar{\varsigma}(s_j)\right)\tilde{W}\left(\bar{\varsigma}(s_j)\right)\right]\\ &+\Gamma^*\left(\bar{\varsigma}_{j+1}\right)-\Gamma^*\left(\bar{\varsigma}_j\right)++\tilde{\sigma}_b\left(\bar{\varsigma}(s_j^+)\right)-\tilde{\sigma}_b\left(\bar{\varsigma}(s_j)\right)+\Gamma^*\left(\bar{\varsigma}(s_j^+)\right)-\Gamma^*\left(\bar{\varsigma}(s_j)\right).\end{aligned}\tag{94}$$

Considering Case 1, $\Gamma(t)$ decreases strictly monotonically over the interval $[s_j, s_{j+1})$. This implies $\Gamma(s_j + t_1) < \Gamma(s_j)$ for all $\forall t_1 \in (0, s_{j+1} - s_j)$. Then, we have $\Lambda(s_j^+) \leq \Lambda(s_j)$ when calculating the limits on both sides of this inequality. Given this result, it follows that

$$\begin{aligned}&\sigma\left(\bar{\varsigma}(s_j^+)\right)+E\left[\frac{1}{2\alpha_1}\tilde{W}^T\left(\bar{\varsigma}(s_j^+)\right)\tilde{W}\left(\bar{\varsigma}(s_j^+)\right)\right]+\Gamma^*\left(\bar{\varsigma}(s_j^+)\right)\\ \leq&\sigma\left(\bar{\varsigma}(s_j)\right)+E\left[\frac{1}{2\alpha_1}\tilde{W}^T\left(\bar{\varsigma}(s_j)\right)\tilde{W}\left(\bar{\varsigma}(s_j)\right)\right]+\Gamma^*\left(\bar{\varsigma}(s_j)\right).\end{aligned}\tag{95}$$

Furthermore, for s_j, $j \in N$, according to Case 1, $\Gamma^*(\varsigma)$ constitutes a continuous difference function, so it follows that $\Gamma^*(\bar{\varsigma}_{j+1}) \leq \Gamma^*(\bar{\varsigma}_j)$. Based on (94) and (95), we derive $\Delta\Lambda_\Gamma(t) \leq 0$.

Based on the above analysis, the proof is completed. □

6. Simulation

Consider the following stochastic system [36]:

$$\dot{\varsigma} = \chi(g)\varsigma + B\rho + Cd,\tag{96}$$

with the system states and control policy being denoted as $\varsigma = [\varsigma_1, \varsigma_2]^T \in R^2$ and $\rho \in R$, respectively. d is the disturbance policy. The control policy ρ satisfies $-0.8 \leq \rho \leq 1$. The uncertain term is $d(\varsigma) = 0.5p_1 \sin(2p_2\varsigma_2)\cos(p_3\varsigma_1)$ with $p_1 \in (-1, 1)$, $p_2 \in (-0.5, 0.5)$ and $p_3 \in (0, 0.5)$. $\xi_i(t), i = 1, \ldots, 4$, is a random variable that satisfies a uniform distribution, and its pdfs are the same as example 1 in the simulation in [36].

Through calculation, one obtains $\|\eth(\varsigma)\|^2 \leq \sin(0.5\varsigma_2^2) \leq 0.5\|\varsigma\|^2$. Set the following auxiliary system as

$$\dot{\varsigma} = \chi(g)\varsigma + B\rho + C\eth,\tag{97}$$

where $\eth(\varsigma)$ denotes the auxiliary disturbance policy. The original state is $\varsigma_0 = [1, 2]^T$ and $Q = \mathrm{diag}\{1, 1\}$. The activation function is defined as follows: $\varphi_c = \left[\varsigma_1^2, \varsigma_1\varsigma_2, \varsigma_2^2\right]^T$, $\varphi_\rho = \varphi_\eth = [\varsigma_1, \varsigma_2]^T$; the weight vectors are $W_c = [W_{c1}, W_{c2}, W_{c3}]^T$, $W_\rho = [W_{\rho 1}, W_{\rho 2}]^T$, and $W_\eth = [W_{\eth 1}, W_{\eth 2}]^T$. The sample size used by the MPCM is 16. Table 1 shows the simulation parameters.

Table 1. Parameter settings.

Parameter	R	T	γ	β	$\rho_{\min}$	$\rho_{\max}$	ϖ	l_ρ
value	1	0.005	5	0.5	−0.8	1	0.5	10

The weight curves of NNs are shown in Figures 1 and 2. The weight updates of the critic, actor, and disturbance NNs gradually approach $[3.413, 2.916, 2.147]$, $[0.868, 0.648]$, and $[0.576, 0.208]$, respectively. Based on this weight convergence value of the critic NN, we can obtain that the GC of the original system is 17.86.

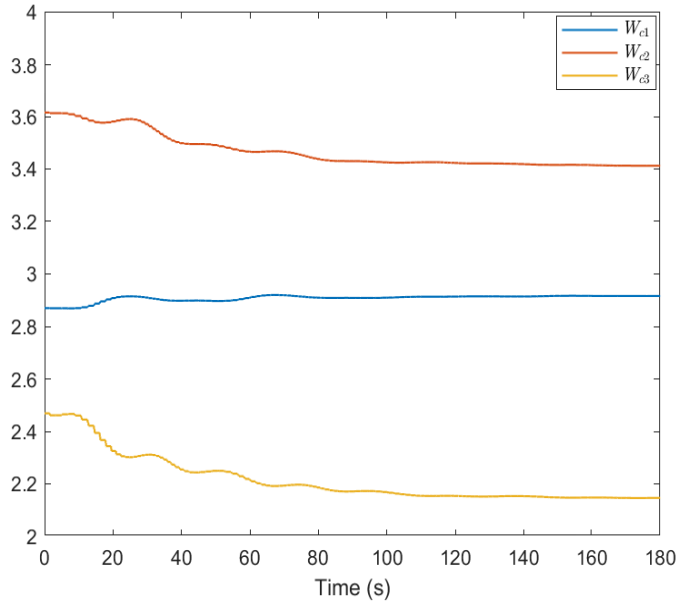


Figure 1. Weight trajectories of the critic NN in our design.

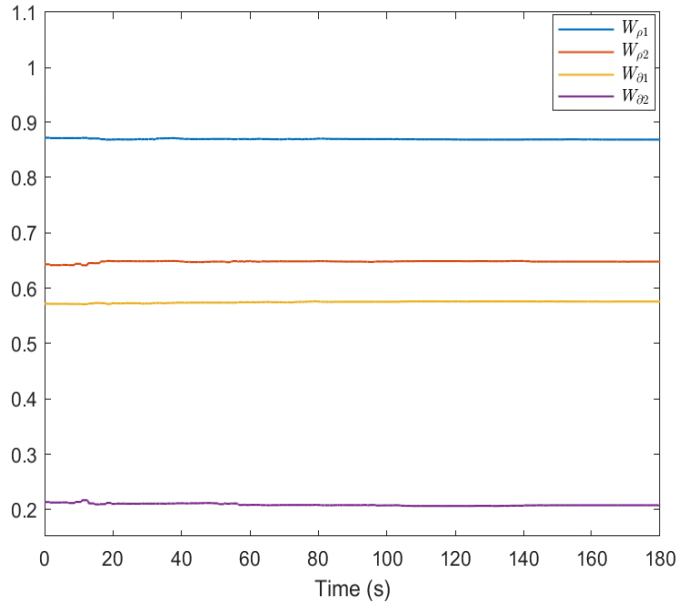


Figure 2. Weight curves of actor and disturbance NNs in our design.

Figure 3 shows the system state curves. Evidently, all state trajectories are capable of eventually converging to the equilibrium point.

Figure 4 gives the variation curves of control pairs according to the DETC mode. It is shown that, compared with the baseline, the curve of the control policy obtained by our designed method stays within the range of -0.8 to 1, which proves the effectiveness of the designed method with ASAs.

Figure 5 shows the trajectories of $||e_j||^2$ and $\bar{e}_T$. By using the DETC method, the number of controller updates can be greatly reduced. The sampling period of ETC is given in Figure 6. It displays the time interval between the previous and the current triggering moment. The triggering frequency is dynamically adjusted according to the triggering condition.

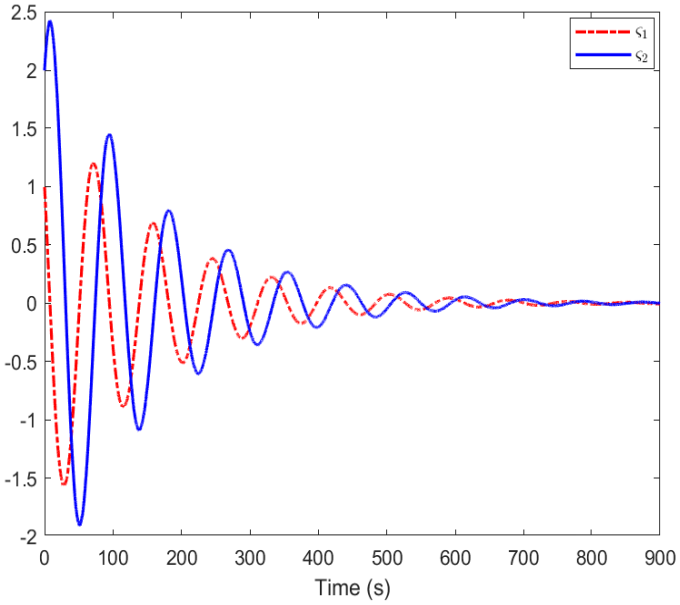


Figure 3. System state curves in our design.

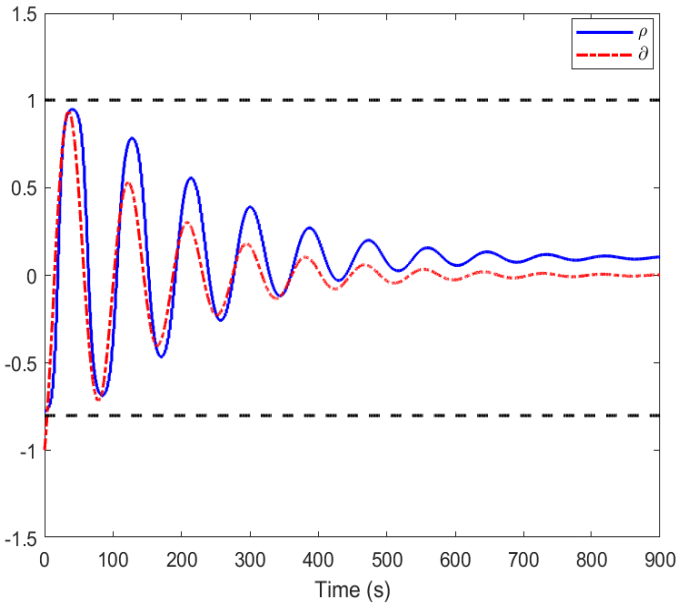


Figure 4. The change curves of the event-triggered optimization control strategy and worst-case disturbance strategy in our design.

Figure 7 illustrates the dynamic evolution process of internal signals within the DETC framework. It is clear that the trajectory of internal parameter changes is always greater than 0 and decreasing.

To demonstrate the control performance of the approach proposed in our design, a comparative simulation is given as follows by using the method proposed in [36].

Figures 8 and 9 show the convergence of the NN weight vectors. The weight updates of the critic, actor and disturbance NNs gradually approach $[3.121, 2.880, 2.295]$, $[0.900, 0.697]$ and $[0.599, 0.301]$, respectively.

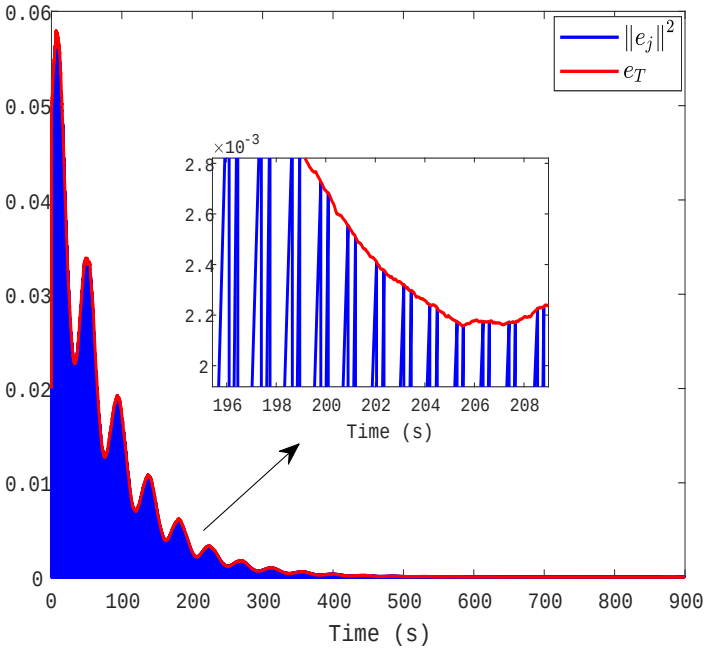


Figure 5. The triggering condition of $||e_j||^2$ and $\bar{e}_T$ in our design.

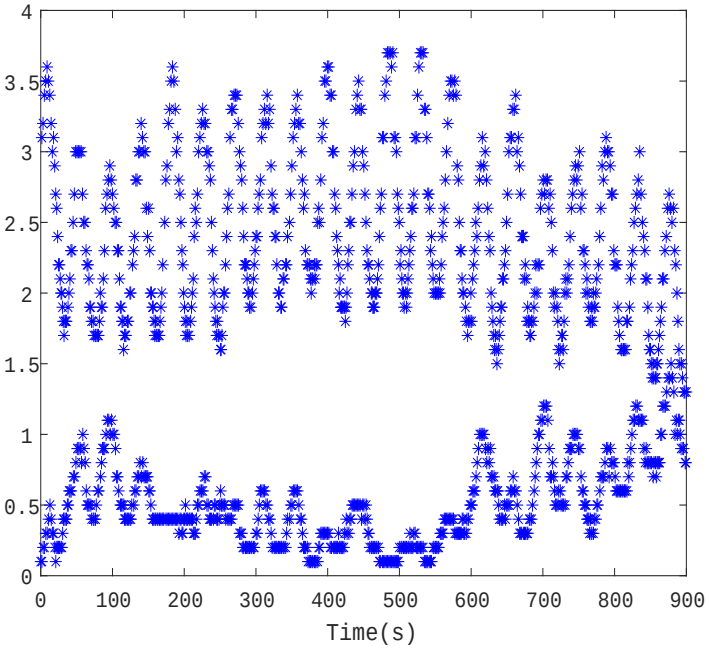


Figure 6. Sampling period of the learning stage in our design.

Figure 10 shows the system state curves obtained by using the method in [36].

Figure 11 shows the variation curves of control pairs based on the ETC mode obtained by using the method in [36]. The blue line denotes the variation curve of the control input, indicating that it exceeds the range of -0.8 to 1. Figure 4 is within the range of -0.8 to 1. By comparing Figures 4 and 11, it is clear that the control strategy in our design exhibits excellent performance.

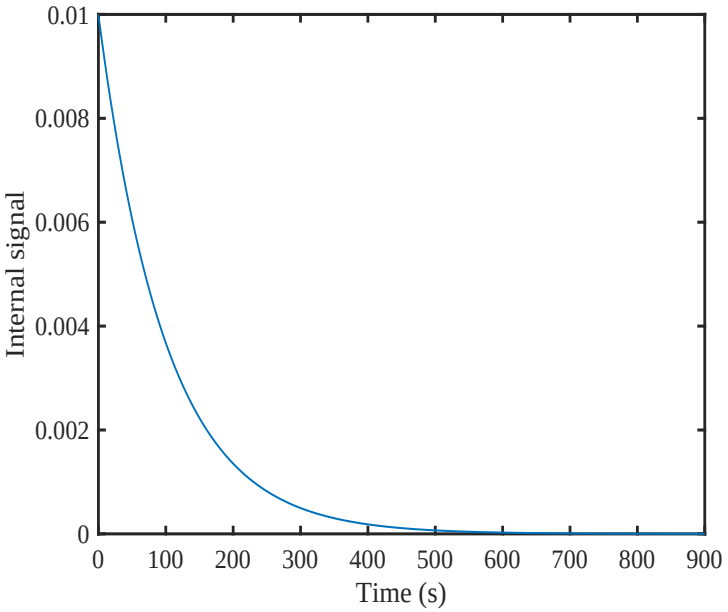


Figure 7. The change curve of the dynamic signal in the triggering condition in our design.

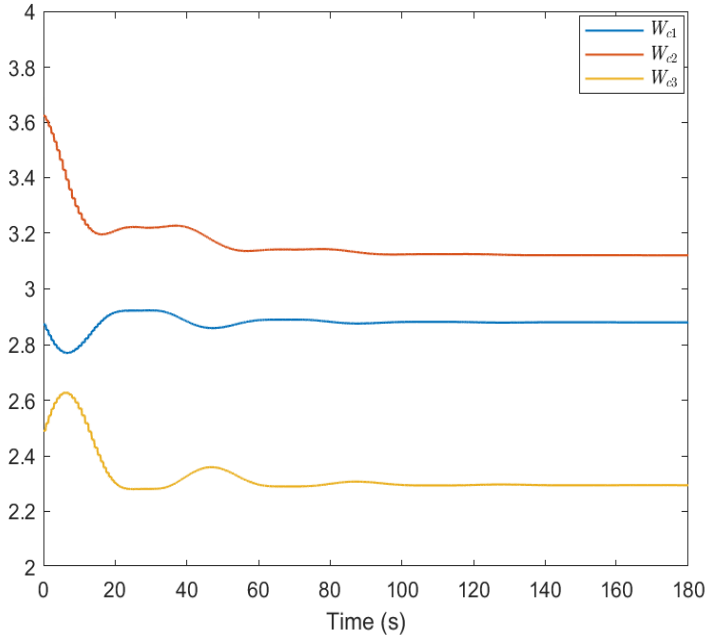


Figure 8. Weight trajectories of the critic NN by using the method in [36].

Figure 12 shows the triggering conditions of $||e_j||^2$ and $\bar{e}_T$ under event-triggering mode. It is obvious that the ETC signal requires 2574 state samples, while the DETC strategy proposed in this paper only uses 2028 state samples. It is evident that the DETC proposed in this article can improve controller updates by up to 21.2%.

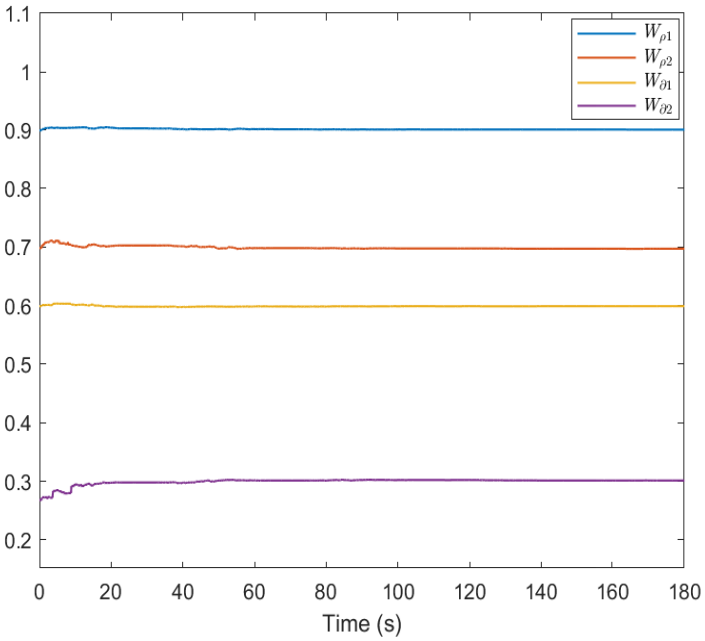


Figure 9. Weight curves of actor and disturbance NNs by using method in [36].

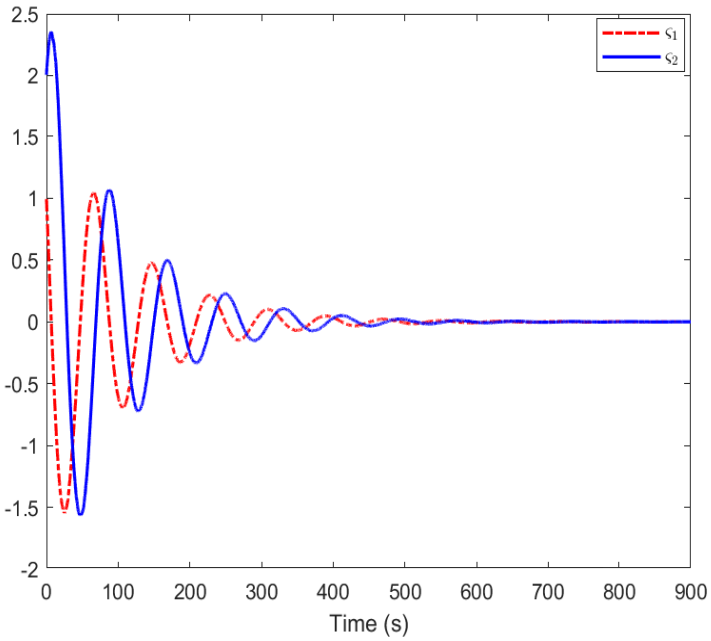


Figure 10. System state curves by using method in [36].

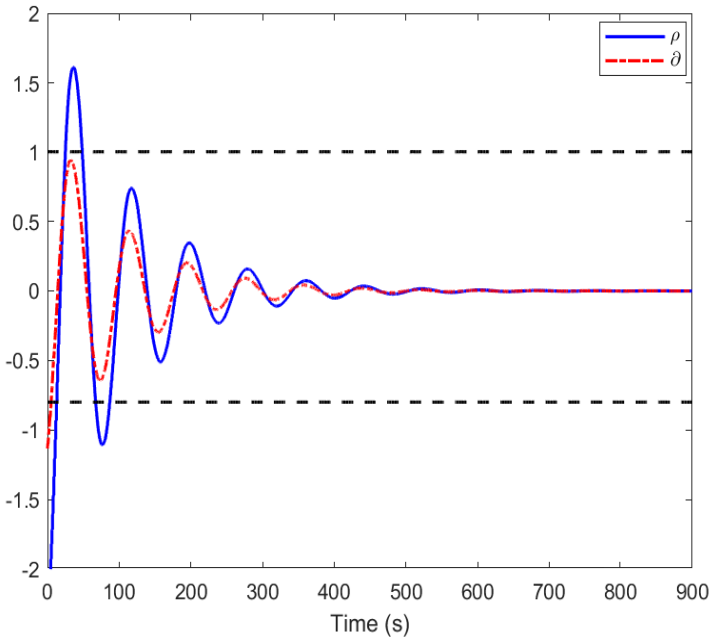


Figure 11. The change curves of the event-triggered optimization control signal and worst-case disturbance signal by using the method in [36].

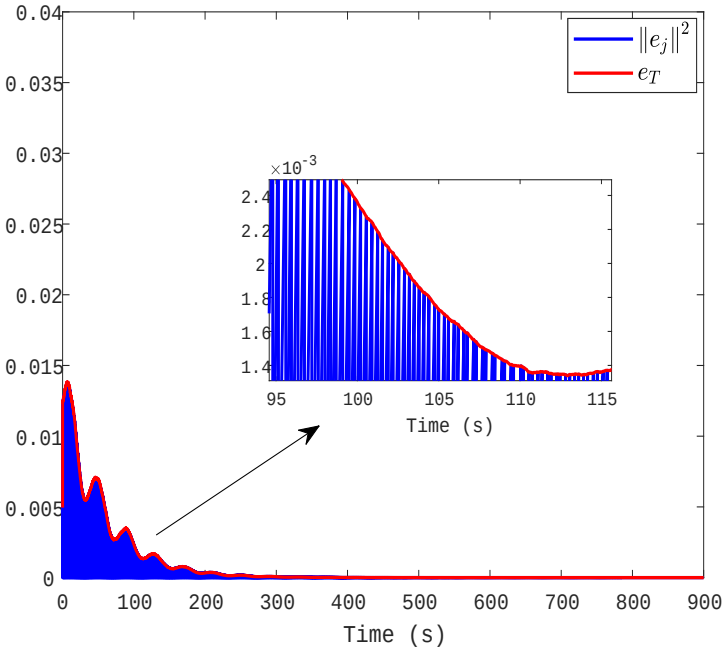


Figure 12. The triggering condition of $||e_j||^2$ and $\bar{e}_T$ by using the method in [36].

7. Conclusions

In our design, we have proposed dynamic event-triggered IRL-based control approach for a stochastic system with random parameters and ASAs. Firstly, a modified HJI equation has been formulated. Applying the actor–critic–disturbance NN architecture, both the OC signal under ASAs and the worst-case disturbance strategy have been derived by solving an improved HJI equation. Secondly, the MPCM has been employed to estimate the value function at specific sampling points. The utilization of the MPCM can reduce the computational complexity of sampling data in simulation experiments. Thirdly, through the introduction of dynamic parameters into the triggering conditions, the control input

has been adjusted dynamically. Moreover, when the triggering condition is met, the weight values of NNs can be tuned in a synchronous manner, which has significantly enhanced the efficiency of communication resource usage. Furthermore, the Lyapunov stability theorem has been employed to analyze and verify the time-varying system's stability. Finally, simulation results have confirmed the efficacy of the developed approach.

Author Contributions: Conceptualization, Y.L.; software, Y.L.; methodology, Y.L.; validation, M.X.; writing—original draft preparation, M.X. and Y.L.; rigorous analysis, J.Z.; supervision, Z.G.; data curation, J.Z. and Z.M.; writing—review and editing, M.X.; funding acquisition, Y.L.; visualization, J.Z. All authors have reviewed and agree to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (62403329).

Data Availability Statement: In terms of the data availability, if a researcher requires data from this article, the corresponding author can provide the simulation data.

Conflicts of Interest: The authors have declared that there are no conflicts of interest.

References

1. Xie, M.; Shakoor, A.; Wu, Z.; Jiang, B. Optical manipulation of biological cells with a robot-tweezers system: A stochastic control approach. *IEEE Trans. Circuits Syst. Ii: Express Briefs* **2020**, *67*, 3232–3236.
2. Bazmohammadi, N.; Tahsiri, A.; Anvari-Moghaddam, A.; Guerrero, J.M. Stochastic predictive control of multi-microgrid systems. *IEEE Trans. Ind. Appl.* **2019**, *55*, 5311–5319.
3. Dai, M.; Wu, C.; Wen, J. Vehicle longitudinal stochastic control for connected and automated vehicle platooning in highway systems. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 9563–9578. https://doi.org/10.1109/TITS.2025.3572555.
4. Liu, J.; Xu, J.; Zhang, H.; Fu, M. Stochastic LQ optimal control with initial and terminal constraints. *IEEE Trans. Autom. Control* **2024**, *69*, 6261–6268.
5. Sun, H.Y.; Mu, H.R.; Fu, S.J.; Han, H.G. Data-driven model predictive control for unknown nonlinear NCSs with stochastic sampling intervals and successive packet dropouts. *IEEE Trans. Cybern.* **2025**, *55*, 2899–2909. https://doi.org/10.1109/TCYB.2025.3555766.
6. Xu, J.; Xie, L.; Zhang, H. Solution to discrete-time linear FBSDEs with application to stochastic control problem. *IEEE Trans. Autom. Control* **2017**, *62*, 6602–6607.
7. Li, Y.; Voos, H.; Darouach, M.; Hua, C. An application of linear algebra theory in networked control systems: Stochastic cyber-attacks detection approach. *IMA J. Math. Control Inf.* **2016**, *33*, 1081–1102.
8. Cetinkaya, A.; Kishida, M. Instabilizability conditions for continuous-time stochastic systems under control input constraints. *IEEE Control Syst. Lett.* **2021**, *6*, 1430–1435.
9. Chatterjee, D.; Hokayem, P.; Lygeros, J. Stochastic receding horizon control with bounded control inputs: A vector space approach. *IEEE Trans. Autom. Control* **2011**, *56*, 2704–2710.
10. Nguyen, X.P.; Dang, X.K.; Do, V.D.; Corchado, J.M.; Truong, H.N. Robust adaptive fuzzy-free fault-tolerant path planning control for a semi-submersible platform dynamic positioning system with actuator constraints. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 12701–12715.
11. Wang, F.; Xie, X.; Zhou, C. Locally expanded constraint-boundary-based adaptive composite control of a constrained nonlinear system with time-varying actuator fault. *IEEE Trans. Fuzzy Syst.* **2023**, *31*, 4121–4136.
12. Sun, W.; Diao, S.; Su, S.F.; Sun, Z.Y. Fixed-time adaptive neural network control for nonlinear systems with input saturation. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *34*, 1911–1920.
13. Zhang, F.; Song, M.; Huang, B.; Huang, P. Adaptive tracking control for tethered aircraft systems with actuator nonlinearities and output constraints. *IEEE Trans. Aerosp. Electron. Syst.* **2024**, *60*, 3582–3597.
14. Gao, Z.; Zhang, Y.; Guo, G. Adaptive fixed-time sliding mode control of vehicular platoons with asymmetric actuator saturation. *IEEE Trans. Veh. Technol.* **2023**, *72*, 8409–8423.
15. Guo, G.; Zhang, P. Asymptotic stabilization of USVs with actuator dead-zones and yaw constraints based on fixed-time disturbance observer. *IEEE Trans. Veh. Technol.* **2019**, *69*, 302–316.
16. Wang, D.; Gao, N.; Liu, D.; Li, J.; Lewis, F.L. Recent progress in reinforcement learning and adaptive dynamic programming for advanced control applications. *IEEE/CAA J. Autom. Sin.* **2023**, *11*, 18–36.
17. Zhang, Y.; Wang, Y.; Cai, Y. Value iteration-based distributed adaptive dynamic programming for multi-player differential game with incomplete information. *IEEE/CAA J. Autom. Sin.* **2025**, *12*, 436–447.

18. Wei, Q.; Yang, Z.; Su, H.; Wang, L. Online adaptive dynamic programming for optimal self-learning control of VTOL aircraft systems with disturbances. *IEEE Trans. Autom. Sci. Eng.* **2022**, *21*, 343–352.
19. Wei, Q.; Chen, W.; Tan, X.; Xiao, J.; Dong, Q. Observer-based optimal Backstepping security control for nonlinear systems using reinforcement learning strategy. *IEEE Trans. Cybern.* **2024**, *54*, 7011–7023. https://doi.org/10.1109/TCYB.2024.3443522.
20. Ming, Z.; Zhang, H.; Li, W.; Luo, Y. Neurodynamic programming and tracking control for nonlinear stochastic systems by PI algorithm. *IEEE Trans. Circuits Syst. II Express Briefs* **2022**, *69*, 2892–2896.
21. Li, J.; Yang, M.; Lewis, F.L.; Zheng, M. Compensator-based self-learning: Optimal operational control for two-time-scale systems with input constraints. *IEEE Trans. Ind. Inform.* **2024**, *20*, 9465–9475.
22. Shi, H.; Gao, W.; Jiang, X.; Su, C.; Li, P. Two-dimensional model-free Q-learning-based output feedback fault-tolerant control for batch processes. *Comput. Chem. Eng.* **2024**, *182*, 108583.
23. Pang, B.; Jiang, Z.P. Reinforcement learning for adaptive optimal stationary control of linear stochastic systems. *IEEE Trans. Autom. Control* **2022**, *68*, 2383–2390.
24. Zhang, K.; Peng, Y. Model-free tracking control for linear stochastic systems via integral reinforcement learning. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 10835–10844. https://doi.org/10.1109/TASE.2025.3529895.
25. Zhang, H.; Qu, Q.; Xiao, G.; Cui, Y. Optimal guaranteed cost sliding mode control for constrained-input nonlinear systems with matched and unmatched disturbances. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 2112–2126.
26. Liu, T.; Jiang, Z.P. Event-based control of nonlinear systems with partial state and output feedback. *Automatica* **2015**, *53*, 10–22.
27. Lu, J.; Han, L.; Wei, Q.; Wang, X.; Dai, X.; Wang, F.Y. Event-triggered deep reinforcement learning using parallel control: A case study in autonomous driving. *IEEE Trans. Intell. Veh.* **2023**, *8*, 2821–2831.
28. Zhang, G.; Zhu, Q. Event-triggered optimized control for nonlinear delayed stochastic systems. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2021**, *68*, 3808–3821.
29. Zhang, G.; Liang, C.; Zhu, Q. Adaptive fuzzy event-triggered optimized consensus control for delayed unknown stochastic nonlinear multi-agent systems using simplified ADP. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 11780–11793. https://doi.org/10.1109/TASE.2025.3540468.
30. Xue, S.; Zhang, W.; Luo, B.; Liu, D. Integral reinforcement learning-based dynamic event-triggered nonzero-sum games of USVs. *IEEE Trans. Cybern.* **2025**, *55*, 1706–1716. https://doi.org/10.1109/TCYB.2025.3533139.
31. Ming, Z.g.; Zhang, H.; Tong, X.; Yan, Y. Mixed H2/H∞ control with dynamic event-triggered mechanism for partially unknown nonlinear stochastic systems. *IEEE Trans. Autom. Sci. Eng.* **2022**, *20*, 1934–1944.
32. Tong, X.; Ma, D.; Wang, R.; Xie, X.; Zhang, H. Dynamic event-triggered-based integral reinforcement learning algorithm for frequency control of microgrid with stochastic uncertainty. *IEEE Trans. Consum. Electron.* **2023**, *69*, 321–330.
33. Zhu, H.Y.; Li, Y.X.; Tong, S. Dynamic event-triggered reinforcement learning control of stochastic nonlinear systems. *IEEE Trans. Fuzzy Syst.* **2023**, *31*, 2917–2928.
34. Liu, M.; Wan, Y.; Lewis, F.L. Adaptive optimal decision in multi-agent random switching systems. *IEEE Control Syst. Lett.* **2019**, *4*, 265–270.
35. Liu, T.; Qin, Z.; Hong, Y.; Jiang, Z.P. Distributed optimization of nonlinear multiagent systems: A small-gain approach. *IEEE Trans. Autom. Control* **2021**, *67*, 676–691.
36. Liang, Y.; Zhang, H.; Zhang, J.; Ming, Z. Event-triggered guarantee cost control for partially unknown stochastic systems via explorized integral reinforcement learning strategy. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *35*, 7830–7844.
37. Yuan, R.; Ma, J.; Su, P.; Dong, Y.; Cheng, J. Monte-Carlo integration models for multiple scattering based optical wireless communication. *IEEE Trans. Commun.* **2019**, *68*, 334–348.
38. Wang, J.; Gao, X.; Cao, R.; Sun, Z. A multilevel Monte Carlo method for performing time-variant reliability analysis. *IEEE Access* **2021**, *9*, 31773–31781.
39. Xie, J.; Wan, Y.; Mills, K.; Filliben, J.J.; Lewis, F.L. A scalable sampling method to high-dimensional uncertainties for optimal and reinforcement learning-based controls. *IEEE Control Syst. Lett.* **2017**, *1*, 98–103.
40. Zhou, Y.; Wan, Y.; Roy, S.; Taylor, C.; Wanke, C.; Ramamurthy, D.; Xie, J. Multivariate probabilistic collocation method for effective uncertainty evaluation with application to air traffic flow management. *IEEE Trans. Syst. Man, Cybern. Syst.* **2014**, *44*, 1347–1363.
41. Liu, M.; Wan, Y.; Lewis, F.L.; Lopez, V.G. Adaptive optimal control for stochastic multiplayer differential games using on-policy and off-policy reinforcement learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 5522–5533.
42. Jin, Z. Global asymptotic stability analysis for autonomous optimization. *IEEE Trans. Autom. Control* **2025**, *70*, 6953–6960. https://doi.org/10.1109/TAC.2025.3567509.
43. Jin, Z.; Li, H.; Qin, Z.; Wang, Z. Gradient-free cooperative source-seeking of quadrotor under disturbances and communication constraints. *IEEE Trans. Ind. Electron.* **2024**, *72*, 1969–1979. https://doi.org/10.1109/TIE.2024.3423337.
44. Shi, K.; Tang, Y.; Zhong, S.; Yin, C.; Huang, X.; Wang, W. Nonfragile asynchronous control for uncertain chaotic lurie network systems with bernoulli stochastic process. *Int. J. Robust Nonlinear Control* **2018**, *28*, 1693–1714.

45. Cui, X.; Zhang, H.; Luo, Y.; Jiang, H. Adaptive dynamic programming for H∞ tracking design of uncertain nonlinear systems with disturbances and input constraints. *Int. J. Adapt. Control Signal Process.* **2017**, *31*, 1567–1583.
46. Zhang, H.; Cui, X.; Luo, Y.; Jiang, H. Finite-horizon H∞ tracking control for unknown nonlinear systems with saturating actuators. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *29*, 1200–1212.
47. Sahoo, A.; Jagannathan, S. Stochastic optimal regulation of nonlinear networked control systems by using event-driven adaptive dynamic programming. *IEEE Trans. Cybern.* **2016**, *47*, 425–438.
48. Yasini, S.; Naghibi Sitani, M.B.; Kirampor, A. Reinforcement learning and neural networks for multi-agent nonzero-sum games of nonlinear constrained-input systems. *Int. J. Mach. Learn. Cybern.* **2016**, *7*, 967–980.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.