On Stability of Perturbed Nonlinear Switched Systems with Adaptive Reinforcement Learning

Phuong Nam Dao 1, Hong Quang Nguyen 2,*, Minh-Duc Ngo 2,* and Seon-Ju Ahn 3,*

1 Department of Automatic Control, Hanoi University of Science and Technology, 1 Dai Co Viet Road, Hanoi 100000, Vietnam; nam.daophuong@hust.edu.vn
2 Department of Automation, Thai Nguyen University of Technology, 666, 3/2 Street, Tich Luong Ward, Thai Nguyen City 251750, Vietnam
3 Department of Electrical Engineering, Chonnam National University, Gwangju 61186, Korea

* Correspondence: quang.nguyenhong@tnut.edu.vn (H.Q.N.); ngoduc198-tdh@tnut.edu.vn (M.-D.N.); sjahn@jnu.ac.kr (S.-J.A.); Tel.: +84-982-092-013 (H.Q.N.); +84-392-681-318 (M.-D.N.); +82-62-530-1738 (S.-J.A.)

Received: 19 August 2020; Accepted: 24 September 2020; Published: 27 September 2020

Abstract: In this paper, a tracking control approach is developed based on an adaptive reinforcement learning algorithm with a bounded cost function for perturbed nonlinear switched systems, which represent a useful framework for modelling these converters, such as DC–DC converter, multi-level converter, etc. An optimal control method is derived for nominal systems to solve the tracking control problem, which results in solving a Hamilton–Jacobi–Bellman (HJB) equation. It is shown that the optimal controller obtained by solving the HJB equation can stabilize the perturbed nonlinear switched systems. To develop a solution to the translated HJB equation, the proposed neural networks consider the training technique obtaining the minimization of square of Bellman residual error in critic term due to the description of Hamilton function. Theoretical analysis shows that all the closed-loop system signals are uniformly ultimately bounded (UUB) and the proposed controller converges to optimal control law. The simulation results of two situations demonstrate the effectiveness of the proposed controller.

Keywords: adaptive dynamic programming (ADP); adaptive reinforcement learning (ARL); switched systems; HJB equation; uniformly ultimately bounded (UUB); Lyapunov stability theory

1. Introduction

Many power electronic converters play a remarkable role in industrial applications, such as electrical drives, renewable energy systems, etc. [1–5]. The modelling technique in power electronic converters is usually implemented by considering the average small-signal analysis. This method enables us to design the voltage, current controller, as well as implement the space vector modulation technique. However, the average small-signal model method is disadvantageous for complicated applications, with the requirement of tracking control systems. Switched systems have been known as a special class of hybrid systems describing an active application area in the field of power electronics, such as DC–DC converter, power source, etc. [6–15]. The approach of employing switched systems representation in control design has been the efficient method to model many practical systems based on multi-subsystem. Basic issues in the control of switched systems are to find the control input under arbitrary switching signal and/or to find the appropriate switching signal obtaining the tracking, stability of closed-loop systems. Additionally, the identification of the active mode based on estimating the delay to analyze the switched systems was implemented in [6]. The authors in [7] proposed a novel Lyapunov function combining linear matrix inequalities (LMIs) for switched linear systems in the presence of exogenous disturbance. The control design problem with stochastic
stability based on a complete probability space under the general random switching signal was investigated in [8]. The stabilizing switching signal was designed using average dwell time with proposed existence conditions [9]. Moreover, several conditions to characterize the mode-dependent dwell-times have been given in [10]. The switching event triggered and back stepping technique-based tracking controller achieved the bound of all signals and tracking error in the closed-loop system [11]. However, the optimal control with effectiveness in dealing with constraint has not been mentioned in switched systems yet [6–11]. For the optimal control design, it is necessary to find the result of a Ricatti equation for linear systems or a Hamilton–Jacobi–Bellman (HJB) equation for nonlinear systems [16]. However, in the general case, it is hard to solve the HJB (Hamilton–Jacobi–Bellman) equation for finding the optimal controller. In [16–19], thanks to the approximation description of Neural Networks (NNs), the weight of the actor and critic were updated simultaneously under the optimization problem for continuous/discrete time systems. The convergence effectiveness of weights in actor/critic policy is guaranteed by the persistence excitation (PE) condition. It is worth emphasizing that to deal with the uncertainties of nonlinear continuous systems, the identifier was inserted into the control structure [19] and the off-policy technique under the separation between actor/critic policy and control input in the first stage is presented [20]. Moreover, the constraint in input/output was mentioned in the adaptive/approximate dynamic programming (ADP) technique [17,18,21] by using the appropriate cost function and the dynamic programming principle. The framework of ADP was utilized in control scheme for surface vessel systems [22], a spring-Mass-Damper [23], two mass-spring systems [20], Wheeled mobile robotic systems [24]. Based on approximating the modified critic term in each control loop, the stability of whole cascade systems was guaranteed in surface vessel systems due to the consideration of the derivative of Lyapunov candidate function along the closed system [22,25]. However, all the ARL-based controllers presented in [20,22–24] are implemented under the consideration of nonlinear systems, which are not mentioned in the switching signals. The model-free problem, as well as the connection between continuous time systems and appropriate discrete time systems, ensure the development of off-policy algorithm [26,27]. To deal with the complete dynamic uncertainties, the modified cost function and critic enable us to obtain the approximately optimal control law [28–31]. Additionally, the general policy iteration (GPI) in linear discrete time systems [32] achieves the simultaneous computation and appropriate data collection.

To the best knowledge of the authors, adaptive reinforcement learning (ARL) for uncertain switched systems is still an open problem. Therefore, this paper attempts to implement an ARL algorithm for uncertain continuous-time nonlinear switched systems under arbitrary switching signals based on an optimization problem in NNs training. Inspired by the above works, this paper studies the ARL-based optimal control problem for a class of perturbed switched nonlinear continuous systems. First, the optimal control algorithm is established for switched systems. Then, the neural network is employed to approximate the Critic part in policy. Based on optimization principle, the training law is proposed to develop the optimal control strategy. The main contributions of this paper are as follows:

1. In comparison with the previous papers [12–15,18,22,29,30,33], an optimal control obtained from the nominal system is proposed for perturbed switched nonlinear continuous systems based on dynamic programming principle.

2. The Neural Network training law with optimization principle is developed to achieve the ARL-based optimal control strategy.

3. The strict proof concerning UUB stability of the closed system and the convergence of the controller to optimal control input are given based on Lyapunov stability theory and reinforcement learning scheme.

The rest of this article is organized as follows. The preliminaries and problem statements are presented in Section 2. The main results are given in Section 3. Two simulation cases are presented in Section 4 to illustrate the effectiveness of the proposed solution. Finally, the conclusions are exhibited in Section 5.
where \( x(t) \in \Omega_u \subset \mathbb{R}^n \) and \( u(t) \in \Omega_u \subset \mathbb{R}^m \) are the vector of state variables and the control inputs, respectively. The function \( \sigma : [0, +\infty) \to \Omega = \{1, 2, ..., n\} \) is an unknown switching signal and \( n \) is the number of subsystems. \( f_i(x) \) (\( i \in \Omega \)) are unknown smooth vector functions satisfying \( f_i(0) = 0 \) and \( g_i(x) \) (\( \forall i \in \Omega \)) are known smooth vector functions such that \( G_{\text{min}} \leq \|g_i(x)\| \leq G_{\text{max}} \).

**Assumption 1.** There exists a known function \( \rho(x) \) satisfying that: \( \|\Delta(x,t)\| \leq \rho(x) \).

Regarding the perturbed continuous-time nonlinear switched system (1), we introduce the cost function formulated as:

\[
J(x(t), u(t)) = \int_{t}^{\infty} r(x(\tau), u(\tau)) d\tau
\]

where \( r(x(\tau), u(\tau)) = x^T Q x + u^T R u \) and \( Q = Q^T \geq 0; R = R^T > 0 \)

Control Objective: This article aims at designing the optimal guaranteed cost control scheme \( u^* = \arg\min_{u \in \Omega_u} K(x, u) \) despite arbitrary switching law, in which the feedback control law \( u \) and a finite upper bound function \( K(x, u) \) are satisfied, not only is the closed-loop system (1) robustly stable but the cost function (2) also satisfies the condition: \( J(x, u) \leq K(x, u) \leq M \).

**Definition 1.** The function \( K(x, u) \) can be known as the guaranteed cost function. Therefore, the control law \( u^* \) with \( u^* = \arg\min_{u \in \Omega_u} K(x, u) \) is known as the optimal guaranteed cost control law.

**Remark 1.** It is worth emphasizing that the main objective of this work is to find the optimal control for the equivalent nominal system with the performance index being modified. Compared with the control objective in [33], this work investigates the tracking control for switched systems, for which it is hard to develop the optimal control algorithm. Additionally, it is obviously different from the existing work of robust ADP in [31], since the proposed optimal control design is implemented based on the upper bound function \( K(x, u) \).

### 3. Adaptive Reinforcement Learning-Based Control Design

In this section, we investigate the ARL-based optimal control for perturbed nonlinear switched systems. Due to the difficulties in implementing directly the optimal control scheme for perturbed nonlinear systems, the strategy is proposed by three steps as follows. First, based on Dynamic Programming, we obtain the optimal control design for the corresponding nominal switched systems obtained by eliminating the uncertainties. Then, the ARL algorithm is developed by Neural Network technique for this nominal system. Finally, we carry out a stability analysis of the closed-loop system involving the perturbed switched system and the proposed ARL controller.

According to perturbed system (1), the nominal system is obtained by eliminating the term of uncertainties as:

\[
\dot{x} = f_\sigma(x) + g_\sigma(x) u
\]

Because the purpose is to consider the tracking problem of optimal control law and the influence of uncertainties, the cost function is modified to satisfy that \( J(x(t), u(t)) \leq J_1(x(t), u(t)) \). Therefore, the cost function associated with (3) can be represented as:
\[ f_1(x(t), u(t)) = \int_{t}^{\infty} \left[ r(x(\tau), u(\tau)) + \lambda \left( \rho(x) \right)^2 \right] d\tau \] (4)

It should be noted that \( f_1(x(t), u(t)) \) with \( \lambda \geq \| R \| \) is one of the guaranteed cost function associated with system (1). According to (2) and (4), it can be seen that:

\[ J(x(t), u(t)) \leq J_1(x(t), u(t)) \] (5)

Based on the Dynamic programming principle, establishing the Bellman function as: \( V^*(t) = \min_{u \in \Omega_u} f_1(x(t), u(t)), \) we have the cost function formulated as:

\[ V^*(t) = \min_{u \in \Omega_u} \int_{t}^{\infty} \left[ r(x(\tau), u(\tau)) + \lambda \left( \rho(x) \right)^2 \right] d\tau \] (6)

\[ V^*(t) = \min_{u \in \Omega_u} \frac{t + \Delta t}{t} \int_{t}^{t + \Delta t} \left[ r(x(\tau), u(\tau)) + \lambda \left( \rho(x) \right)^2 \right] d\tau + \min_{u \in \Omega_u} \int_{t + \Delta t}^{\infty} \left[ r(x(\tau), u(\tau)) + \lambda \left( \rho(x) \right)^2 \right] d\tau \] (7)

\[ V^*(x) = \min_{u \in \Omega_u} \frac{t + \Delta t}{t} \int_{t}^{t + \Delta t} \left[ r(x(\tau), u(\tau)) + \lambda \left( \rho(x) \right)^2 \right] d\tau + V^*(t + \Delta t) - V^*(t) \] (8)

\[ \min_{u \in \Omega_u} \left\{ \frac{1}{\Delta t} \int_{t}^{t + \Delta t} \left[ r(x(\tau), u(\tau)) + \lambda \left( \rho(x) \right)^2 \right] d\tau + \frac{V^*(t + \Delta t) - V^*(t)}{\Delta t} \right\} = 0 \] (9)

As the convergence of \( \Delta t \rightarrow 0^+ \), we can derive that:

\[ \min_{u \in \Omega_u} \left[ r(x(t), u(t)) + \lambda \rho^2(x) + (\nabla V^*)^T (f_r(x) + g_r(x) u) \right] = 0 \] (10)

where \( \nabla V^* = \frac{\partial V^*}{\partial x} \).

Consider the Hamilton function obtaining from the nominal system and performance index (4):

\[ H(x, u, V^*) = r(x(t), u(t)) + \lambda \rho^2(x) + (\nabla V^*)^T (f_r(x) + g_r(x) u) \] (11)

The control input is computed by minimizing this function with \( u \in \Omega_u \) as

\[ H(x, u^*, V^*) = \min_{u \in \Omega_u} H(x, u, V^*) = 0 \] (12)

Therefore, we find that (13) yields

\[ \left. \frac{\partial H(x, u, V^*)}{\partial u} \right|_{u = u^*} = 0 \Rightarrow u^* = -\frac{1}{2} R^{-1} (g_r(x))^T \nabla V^* \] (13)

We implement to develop this control algorithm (13) for nonlinear switched system (1) and achieve the following result:

**Theorem 1.** Take system (1) with the feedback control law \( u^*(x) = -\frac{1}{2} R^{-1} (g_r(x))^T \nabla V^* \) into consideration. Then, the cost function \( V^*(t) = \int_{t}^{\infty} \left[ r(x^*(\tau), u^*(\tau)) + \lambda \left( \rho(x^*) \right)^2 \right] d\tau \) is the Lyapunov function candidate with \( \lambda \geq \| R \| \) can guarantee that the system (1) is stable.
Proof. Consider the derivative of $V^*(t)$ along the solution of Equation (1), we achieve the result formulated as:

$$V^*(t) = (\nabla V^* T (f_i(x^*) + g_i(x^*) (u^* + \Delta(x^*, t))))$$

(14)

It follows from $u^*(x) = -\frac{1}{2} R^{-1} (g_i(x))^T \nabla V^*$ and according to (14), (11) and (12), we have:

$$8V^*(t) = -r(x^*(t), u^*(t)) - \lambda \rho^2 (x) + (g_i(x^*))^T \nabla V^* \Delta(x^*, t)$$
$$= -x^T Q x^* - u^* T R u^* - \lambda \rho^2 (x^*) - 2u^* T R \Delta(x^*, t)$$

(15)

$$= -x^T Q x^* - \lambda \rho^2 (x^*) - \left( u^* T + \Delta(x^*, t)^T \right) R (u^* + \Delta(x, t)) + \Delta(x^*, t)^T R \Delta(x^*, t)$$

(16)

$$= -x^T Q x^* - \left( \lambda \rho^2 (x^*) - \Delta(x^*, t)^T R \Delta(x^*, t) \right) - (u^* + \Delta(x, t))^T R (u^* + \Delta(x, t))$$

(17)

According to Assumption 1 and $\lambda \geq ||R||$, it can be seen that:

$$\dot{V}^*(t) \leq -x^T Q x^* - \left( \lambda \rho^2 (x^*) - ||R|| \Delta(x^*, t) ||^2 \right) \leq -x^T Q x^* - (\lambda - ||R||) \rho^2 (x^*)$$

(18)

$$\dot{V}(t) \leq -x^T Q x^*$$

(19)

Therefore, the system (1) is stable under the optimal control for equivalent nominal system. □

It is noteworthy that Theorem 1 has a new extension in switched systems compared with nonlinear systems in the existing result [33].

However, it is impossible to find the Bellman function $V^*(x)$ based on solving analytically the HJB nonlinear Equation (12). Hence, to solve it, we construct a critic network under the framework of adaptive critic learning. Using the approximation property of neural networks described in [28], the critic associated with system (3) can be described as:

$$V^*(x) = w^T \sigma(x) + \varepsilon(x)$$

(20)

where $\sigma(x) : \mathbb{R}^n \rightarrow \mathbb{R}^N; \sigma(0) = 0$ is the vector activation function including $N$ linearly independent elements, $N$ is the number of neurons in the hidden layer of Radial Basis Function (RBF) network [28] and it is also the number of dimensions of $\sigma(x) : \mathbb{R}^n \rightarrow \mathbb{R}^N$, and $\varepsilon(x)$ is the function reconstruction error playing the role in finding the training law the in next steps, $w \in \mathbb{R}^N$ is the idea weight vector and is generally unavailable. As $N \rightarrow \infty$ converges to infinity, we obtain that $\varepsilon(x) \rightarrow 0$ and $\nabla \varepsilon(x) \rightarrow 0$. The following assumption is considered for each fixed $N$.

Assumption 2. The estimation of terms in NNs can be described as: $||\varepsilon(x)|| \leq \varepsilon_{\text{max}}; ||\nabla \varepsilon(x)|| \leq \nabla \varepsilon_{\text{max}}; \nabla \sigma_{\text{min}} \leq ||\nabla \sigma(x)|| \leq \nabla \sigma_{\text{max}}; ||w|| \leq w_{\text{max}}$

According two Equations (12) and (13), it holds that:

$$H(x, u^*, V^*) = r(x(t), u^*(t)) + \lambda \rho^2 (x) + (\nabla V^*)^T f_i(x) + g_i(x) u^* = 0$$

(21)

$$= x^T Q x + \lambda \rho^2 (x) + (\nabla V^*)^T f_i(x) - \frac{1}{4} (\nabla V^*)^T g_i(x) R^{-1} g_i(x)^T (\nabla V^*) = 0$$

(22)

It follows from (20) to:

$$\nabla V^* = (\nabla \sigma(x))^T w + \nabla \varepsilon(x)$$

(23)

The NN-based HJB equation can be represented as follows:

$$e_{\text{NN}} = x^T Q x + \lambda \rho^2 (x) + (\nabla \sigma(x)^T w) f_i(x) - \frac{1}{4} (\nabla \sigma(x)^T w)^T g_i(x) R^{-1} g_i(x)^T (\nabla \sigma(x)^T w)$$

(24)

$$e_{\text{NN}} = x^T Q x + \lambda \rho^2 (x) + w^T \nabla \sigma(x) f_i(x) - \frac{1}{4} w^T \nabla \sigma(x) g_i(x) R^{-1} g_i(x)^T \nabla \sigma(x)^T w$$

(25)
Thus, the residual error is formulated by the function approximation error:

$$e_{NN} = -\nabla \epsilon (x)^T (f_i (x) + g_i (x) u^*) + \frac{1}{4} \nabla \epsilon (x)^T g_i (x) R^{-1} g_i (x)^T \nabla \epsilon (x)$$

(26)

It should be noted that, as $N$ goes to infinity, $e_{NN}$ converges uniformly to zero. Hence, the residual error $e_{NN}$ is bounded for each fixed $N$. Under the framework of ADP-based approximate optimal control design, a critic neural network is established by an estimated weight vector $\hat{w}$ given as:

$$V = \hat{w}^T \sigma (x) = \sigma (x)^T \hat{w}; \dot{u} = -\frac{1}{2} R^{-1} (g_i (x))^T \nabla V$$

(27)

The approximate error of the critic network can be formulated as:

$$e_{HJB} = r (x (t), \dot{u} (t)) + \lambda^2 (x) + \left( \nabla \sigma (x)^T \hat{w} \right)^T (f_i (x) + g_i (x) \dot{u})$$

(28)

$$e_{HJB} = x^T Q x + \lambda^2 (x) + \hat{w}^T \nabla \sigma (x) f_i (x) - \frac{1}{4} \hat{w}^T \nabla \sigma (x) g_i (x) R^{-1} g_i (x)^T \nabla \sigma (x)^T \hat{w}$$

(29)

The weight vector is determined by a steepest descent algorithm:

$$\frac{d}{dt} \hat{w} = -\alpha \frac{\partial E}{\partial \hat{w}}$$

(30)

with $E = \frac{1}{2} e_{HJB}^T e_{HJB}$

**Remark 2.** The weight vector $\hat{w}$ is designed as in (30) to minimize the quadratic function $E = \frac{1}{2} e_{HJB}^T e_{HJB}$. The fact is that the following inequality:

$$\dot{E} = \frac{\partial E}{\partial t} = \frac{\partial E}{\partial \hat{w}} \frac{\partial \hat{w}}{\partial t} = \frac{\partial E}{\partial \hat{w}} \left( \frac{d}{dt} \hat{w} \right) = -\alpha \left( \frac{\partial E}{\partial \hat{w}} \right)^2 \leq 0$$

(31)

**Theorem 2.** The proposed optimal control (27) with the weight vector of the critic network (30) can force the system (1) to be uniform ultimate bounded (UUB) stability.

**Proof.** Define: $\tilde{w} = w - \hat{w} \Rightarrow \frac{d}{dt} \tilde{w} = -\frac{d}{dt} \hat{w}$. Consider the Lyapunov function:

$$V (t) = V_1 (t) + V_2 (t), \text{ where: } V_1 (t) = \frac{1}{2\alpha} \tilde{w} (t)^T \hat{w} (t), V_2 (t) = V^*$$

For deriving the term $V_1 (t)$, we obtain that:

$$\frac{d}{dt} V_1 (t) = \frac{1}{\alpha} \tilde{w} (t)^T \left( \frac{d}{dt} \hat{w} (t) \right) = -\frac{1}{\alpha} \tilde{w} (t)^T \left( \frac{d}{dt} \hat{w} (t) \right) = \hat{w} (t)^T \frac{\partial E}{\partial \hat{w}}$$

$$V_1 = \tilde{w}^T e_{HJB} \nabla \sigma (x) (f_i (x) + g_i (x) \dot{u})$$

According to (27) and (13), it follows that:

$$\dot{u} - u^* = -\frac{1}{2} R^{-1} (g_i (x))^T (\nabla V - \nabla V^*)$$

(32)

$$= -\frac{1}{2} R^{-1} (g_i (x))^T \left( \nabla \sigma (x)^T \hat{w} - (\nabla \sigma (x))^T \hat{w} - \nabla \epsilon (x) \right)$$

(33)

$$= \frac{1}{2} R^{-1} (g_i (x))^T \left( (\nabla \sigma (x))^T \hat{w} + \nabla \epsilon (x) \right)$$

(34)
Moreover, we have:

\[
\nabla \sigma(x) \left( f_1(x) + g_i(x) \tilde{u} \right) = \nabla \sigma(x) \left( f_1(x) + g_i(x) u^* \right) + \nabla \sigma(x) g_i(x) \left( \tilde{u} - u^* \right) \tag{35}
\]

\[
\nabla \sigma(x) \left( f_1(x) + g_i(x) \tilde{u} \right) = \nabla \sigma(x) \left( f_1(x) + g_i(x) u^* + \frac{1}{2} g_i(x) R^{-1} g_i(x)^T \left( \nabla \sigma(x)^T \tilde{\omega} + \nabla \epsilon(x) \right) \right) \tag{36}
\]

From (26) and (29), it follows:

\[
e_{HJB} - e_{NN} = \left( \left( \nabla \sigma(x)^T \tilde{\omega} \right)^T f_1(x) - \left( \nabla \sigma(x)^T w \right)^T f_1(x) \right) + \frac{1}{4} \left( w^T \nabla \sigma(x) g_i(x) R^{-1} g_i(x)^T \nabla \sigma(x)^T w - \tilde{\omega}^T \nabla \sigma(x) g_i(x) R^{-1} g_i(x)^T \nabla \sigma(x)^T \tilde{\omega} \right) \tag{37}
\]

\[
e_{HJB} - e_{NN} = \left( \nabla \sigma(x)^T \tilde{\omega} \right)^T f_1(x) + \frac{1}{4} \left( \nabla \sigma(x)^T w \right)^T f_1(x) \tag{38}
\]

\[
e_{HJB} - e_{NN} = -\left( \nabla \sigma(x)^T \tilde{\omega} \right)^T f_1(x) + \frac{1}{4} \left( \nabla \sigma(x)^T w \right)^T f_1(x) \tag{39}
\]

\[
e_{HJB} - e_{NN} = \left( \nabla \sigma(x)^T \tilde{\omega} \right)^T f_1(x) + \frac{1}{4} \left( \nabla \sigma(x)^T w \right)^T f_1(x) \tag{40}
\]

\[
e_{HJB} - e_{NN} = -\frac{1}{2} \tilde{\omega}^T \nabla \sigma(x) g_i(x) R^{-1} g_i(x)^T \nabla \sigma(x)^T \tilde{\omega} \tag{41}
\]

Because \( u^* = -\frac{1}{2} R^{-1} (g_i(x))^T \left( (\nabla \sigma(x))^T w + \nabla \epsilon(x) \right) \), we have:

\[
e_{HJB} - e_{NN} = -\frac{1}{2} \tilde{\omega}^T \nabla \sigma(x) f_i(x) + \frac{1}{2} \tilde{\omega}^T \nabla \sigma(x) g_i(x) \left( -u^* - \frac{1}{2} R^{-1} g_i(x)^T \nabla \epsilon(x) \right) \tag{42}
\]

\[
e_{HJB} - e_{NN} = -\frac{1}{2} \tilde{\omega}^T \nabla \sigma(x) f_i(x) + \tilde{\omega}^T \nabla \sigma(x) g_i(x) \left( -u^* - \frac{1}{2} R^{-1} g_i(x)^T \nabla \epsilon(x) \right) \tag{43}
\]

Assumption 3. \( ||f_i(x) + g_i(x) u^*|| \leq \mu_{max} \)

Define: \( \mu_i = f_i(x) + g_i(x) u^* \); \( G_i = g_i(x) R^{-1} g_i(x)^T \); \( \nabla \sigma = \nabla \sigma(x) ; \nabla \epsilon = \nabla \epsilon(x) \)

From (35) we obtain:

\[
V_1(t) = -\tilde{\omega}^T \left( -e_{NN} + \tilde{\omega}^T \nabla \sigma \mu_i + \frac{1}{2} \tilde{\omega}^T \nabla \sigma G_i \nabla \epsilon + \frac{1}{4} \tilde{\omega}^T \nabla \sigma G_i \nabla \sigma^T \tilde{\omega} \right) \nabla \sigma(x) \left( \mu_i + \frac{1}{2} G_i \left( \nabla \sigma^T \tilde{\omega} + \nabla \epsilon \right) \right) \tag{44}
\]

\[
V_1(t) = -\frac{1}{8} \tilde{\omega}^T \left( \tilde{\omega}^T \nabla \sigma G_i \nabla \sigma^T \tilde{\omega} \right) \nabla \sigma \left( G_i \nabla \sigma^T \tilde{\omega} \right) - \left( \frac{1}{2} \tilde{\omega}^T \left( \tilde{\omega}^T \nabla \sigma \mu_i \right) \nabla \sigma \left( G_i \nabla \sigma^T \tilde{\omega} \right) + \frac{1}{4} \tilde{\omega}^T \left( \tilde{\omega}^T \nabla \sigma G_i \nabla \sigma^T \tilde{\omega} \right) \nabla \sigma \mu_i \right) \tag{45}
\]

\[
-\tilde{\omega}^T \left( \tilde{\omega}^T \nabla \sigma \mu_i + \frac{1}{2} \tilde{\omega}^T \nabla \sigma G_i \nabla \epsilon \right) \nabla \sigma \mu_i + \frac{1}{2} \tilde{\omega}^T e_{NN} \nabla \sigma \left( G_i \nabla \sigma^T \tilde{\omega} \right) + \tilde{\omega}^T e_{NN} \nabla \sigma \mu_i \tag{46}
\]

}\]
According to (49) and (50), the inequality can be obtained

Define:

\[ A = \mathbf{m}^T \nabla g \nabla \sigma \mathbf{w}; B = \frac{3}{4} \left( \mathbf{w}^T \nabla \sigma \mathbf{u} \right) + \frac{1}{4} \left( \mathbf{w}^T \nabla g \nabla \sigma \right) + \frac{1}{2} e_{NN}; \]

It is obvious that:

\[ V_1 (t) = -\frac{1}{8} A^2 - BA - C^2 - DC = -\frac{1}{8} (A + 4B)^2 + 2B^2 - \left( C + \frac{D}{2} \right)^2 + \frac{D^2}{4} \]

We have:

\[ A + 4B \geq \| \mathbf{w} \| (G_{\min})^2 \lambda_{\min} \left( R^{-1} \right) \left( \nabla \sigma_{\max} \right)^2 - \| \mathbf{w} \| (3 \mathbf{v} \sigma_{\max} \sigma_{\min} + \nabla \sigma_{\max} (G_{\max})^2 \lambda_{\max} \left( R^{-1} \right) \nabla \epsilon_{\max} ) - 2 \epsilon_{\max} \]

\[ 16B^2 + 2D^2 \leq \left( \frac{\| \mathbf{w} \| (3 \mathbf{v} \sigma_{\max} \sigma_{\min} + \nabla \sigma_{\max} (G_{\max})^2 \lambda_{\max} \left( R^{-1} \right) \nabla \epsilon_{\max} )}{2} + 2 \epsilon_{\max} \right)^2 \]

According to (49) and (50), the inequality can be obtained

\[ (A + 4B)^2 - (16B^2 + 2D^2) \geq \pi_1 \]

It can be obtained that \( \pi_1 > 0 \) with a big enough variable \( \| \mathbf{w} \| \) and the highest order coefficient \( \left( (G_{\min})^2 \lambda_{\min} \left( R^{-1} \right) \left( \nabla \sigma_{\min} \right)^2 \right)^2 > 0 \). Therefore, we can determine the positive number \( \theta_1 \) such that:

\[ \forall \| \mathbf{w} \| > \theta_1, \text{ we imply that} \ (A + 4B)^2 - (16B^2 + 2D^2) \geq \pi_1, \text{ and from } (48) \text{ we obtain: } V_1 (t) \leq -\pi_1. \]

Regarding the term of \( V_2 (t) \), from (20) we compute the derivative of them:

\[ V_2 = (\nabla v^*)^T (f_i + g_i (\theta + \Delta)) \]

\[ = -\left( x^T Q x + \lambda \sigma^2 (x) \right) + \frac{1}{4} (\nabla v^*)^T g_i (x) (R^{-1} g_i (x))^T (\nabla v^*) + (\nabla v^*)^T g_i \left( -\frac{1}{2} R^{-1} (g_i (x))^T \nabla \sigma (x) \right) + \Delta \]

\[ + \frac{1}{2} (\nabla v^*)^T g_i (x) \left( (\nabla v^*)^T \nabla \sigma - \nabla \varepsilon (x) \right) + (\nabla v^*)^T g_i \Delta \]

\[ = -\left( x^T Q x + \lambda \sigma^2 (x) \right) - \frac{1}{4} (\nabla v^*)^T g_i (x) (R^{-1} g_i (x))^T (\nabla v^*) \]

\[ + \frac{1}{2} (\nabla v^*)^T g_i (x) \left( (\nabla v^*)^T \nabla \sigma (x) \right) + (\nabla v^*)^T g_i \Delta \]

\[ = -\left( x^T Q x + \lambda \sigma^2 (x) \right) - \frac{1}{4} (\nabla v^*)^T g_i (x) (R^{-1} g_i (x))^T (\nabla v^*) \]

\[ + \frac{1}{2} (\nabla v^*)^T g_i (x) \left( (\nabla v^*)^T \nabla \sigma (x) \right) + (\nabla v^*)^T g_i \Delta \]
It can be considered that $\rho (x) = \omega \|x\|$. According to (51), we obtain:

$$\dot{V}_2 \leq - (\lambda_{\min} (Q) + \lambda \omega) \|x\|^2 + \theta^2$$  \hspace{1cm} (52)

with

$$\theta^2 = - \frac{1}{4} (\nabla V^*)^T g_i R^{-1} g_i^T (\nabla V^*) + \frac{1}{2} (\nabla V^*)^T g_i R^{-1} g_i^T (\nabla \sigma (x)^T \dot{w} + \nabla \varepsilon (x)) + (\nabla V^*)^T g_i \Delta$$  \hspace{1cm} (53)

According to the assumption 1 and assumption 2, it follows that:

$$\theta^2 \leq \frac{1}{4} (w_{\max} \nabla \sigma_{\max} + \nabla \varepsilon_{\max})^2 \delta_{\max} \lambda_{\max} \left( R^{-1} \right) + \frac{1}{2} (\theta \nabla \sigma_{\max} + \nabla \varepsilon_{\max})^2 \delta_{\max} \lambda_{\max} \left( R^{-1} \right)$$

$$+ (w_{\max} \nabla \sigma_{\max} + \nabla \varepsilon_{\max}) g_{\max} \omega \|x\|$$

(54)

It can be seen that $(\lambda_{\min} (Q) + \lambda \omega) \|x\|^2 - \theta^2 \geq \pi_2$ with $\pi_2 > 0$ also leads to other inequalities, including the polynomial quadratic form with the variable $\|x\|$ and the highest order coefficient $(\lambda_{\min} (Q) + \lambda \omega) > 0$. Hence, we can find the positive number $\vartheta_2$ such that: $\forall \|x\| > \vartheta_2$ we get $(\lambda_{\min} (Q) + \lambda \omega) \|x\|^2 - \theta^2 \geq \pi_2$, and from (52) we obtain

$$\dot{V}_2 (t) \leq - \pi_2.$$

\[\square\]

**Remark 3.** The numbers $\vartheta_1; \vartheta_2$ can be changed by establishing the neural network of the optimal cost function. Additionally, for any switching index, the variable $\|x\|$ and $\|\dot{x}\|$ converge to the specified domains. The approximate optimal control law $\hat{u}$ is considered in the equation as (27), which converges to the neighbourhood of the optimal control as $u^*$. Different from the controller in [33], the switched system with unknown switching signal is mentioned in this work by proposed adaptive optimal control.

**Proof.** It should be noted that according to (34), we imply that:

$$\|\hat{u} - u^*\| = \frac{1}{2} \|R^{-1} (g_i (x))^T \left( (\nabla \sigma (x))^T \dot{w} + \nabla \varepsilon (x) \right)\| \leq \frac{1}{2} \lambda_{\max} \left( R^{-1} \right) \|g_{\max} (\nabla \sigma_{\max} \cdot c_1 + \nabla \varepsilon_{\max}) = \vartheta_3$$  \hspace{1cm} (55)

\[\square\]

4. Simulation Results

In this section, we verify the effectiveness and performance of the proposed controller. The ARL algorithm is developed for solving the optimal and tracking problem by RBF network-based ADP, where a single critic Neural Network is used to approximate the Bellman function. The ARL control law and the weight of the Critic part are established as in (27) and (30) with the appropriate learning rate $\alpha$ and the coefficient $\lambda$. Moreover, to carry out the ARL algorithm (27) and (30), we need to employ the term of model $g_i (x)$, the term of cost function $Q, R$ and the appropriate function $\sigma (x)$ in Neural Network. To verify the proposed method of our algorithm, two different situations are implemented to do simulations with all the parameters and functions being listed in each case as follows:

4.1. The Second-Order Switched Nonlinear Systems

In this simulation experiment, we consider the case of switched systems including $N = 2$ subsystems to be described as Equations (56) and (57):

$$\begin{cases} x_1 = -x_1^3 - 2x_2 + (u + \Delta_1 (x, t)) \\ x_2 = x_1 + 0.5 \cos (x_1^2) \sin (x_2^3) - (u + \Delta_1 (x, t)) \end{cases}$$  \hspace{1cm} (56)
\begin{align}
\dot{x}_1 &= -x_1^5 \sin(x_2) + (u + \Delta_2(x,t)) \\
\dot{x}_2 &= \frac{1}{2} x_1 - \cos(x_1) \cos(x_3^2) - (u + \Delta_2(x,t))
\end{align}
\tag{57}

The initial value of state vector is selected as

\[ x(0) = \begin{bmatrix} 5 \\ -5 \end{bmatrix}^T \]

and choosing the parameters and matrices in cost function as follows: 

\[ R = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}; Q = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}; \alpha = 0.1; \lambda = 5. \]

Additionally, the function \( \sigma(x) \) in RBF Neural Network can be chosen as

\[ \sigma(x) = [x_1^2, x_1 x_2, x_2^2]^T. \]

Under the arbitrary switching law (Figure 1), the results are shown in Figures 2–6 by using proposed ARL law, in which the responses of state variables are described in Figures 2 and 3 under the control input (Figure 5) and the weights are trained as in Figure 6. It is obvious that the closed system is stable due to the convergence of state variables to zero (Figures 2 and 3). On the other hand, the convergence is also shown in the weights of Critic part (Figure 6).

**Remark 4.** The second-order switched nonlinear system was also mentioned in [14] by using nonlinear control law after the equivalent transform operation (Figures 1–6). However, it is noteworthy that the response of control input of proposed method has a better transient and steady performance compared with the nonlinear controller in [14] because of the effectiveness of handling the constraint in optimal control (Figures 4 and 5).

![Time Series Plot:](image)

**Figure 1.** The Switching Law of the second-order Switched System.
Figure 2. The response of state variable 1.

Figure 3. The response of state variable 2.
Figure 4. The response of Control Input.

Figure 5. The response of Control Input using Adaptive Reinforcement Learning.
4.2. The Third-Order Switched Nonlinear Systems

Here, we continue to investigate the case of switched systems including $N = 2$ subsystems as follows:

\[
\begin{align*}
\dot{x} &= \begin{bmatrix} -x_1 + x_2 \\ -0.2x_2 - \sin x_1 \cos x_3 \\ x_1x_2 - \sin x_3 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} u + \Delta_1(x, t) \\
\dot{x} &= \begin{bmatrix} -x_1 + x_2^2x_3 \\ -x_2 - \sin (x_1 \cos x_3) \\ \sin x_2 - x_3 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} u + \Delta_2(x, t)
\end{align*}
\]

where $x_0 = [5; -3; 2]^T$; $\Delta_i(x, t) = \frac{1}{10} \begin{bmatrix} d_1x_1 \\ d_2x_2 \\ d_3x_3 \end{bmatrix}$; $\|\Delta_i\| \leq \frac{1}{10} \|x\|$ with $d_1, d_2, d_3$ are the random functions.

The parameters of proposed control design can be chosen as $R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$; $\alpha = 0.01$; $Q = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$; $\lambda = 5$. Moreover, the function $\sigma(x)$ in RBF Neural Network can be chosen as $\sigma(x) = [x_1^2, x_2^2, x_3^2, x_1x_2, x_2x_3, x_1x_3]^T$. In this case, under the arbitrary switching law in Figure 7, the results are obtained in Figures 8–13 with the responses of state variables are shown in Figures 8–10 based on the control input (Figures 11 and 12), and the training weights are shown in Figure 13. It can be seen that the closed system is stable (Figures 8–10) and the convergence is also shown in the weights of Critic part (Figure 13).

**Remark 5.** It is worth noting that the third-order switched nonlinear system was also considered in [15] based on the adaptive backstepping output feedback nonlinear control scheme combined with the transform operation.

![Figure 6. The Training Parameters of adaptive dynamic programming (ADP).](image-url)
obtaining the response of state variables, control input in Figures 8–11 and adaptation law (Figure 14). However, thanks to the property of optimal control, the control input response of the proposed controller has a better transient compared with the adaptive backstepping nonlinear controller in [15] (Figures 11 and 12).

Figure 7. The Switching Law of the third-order Switched System.

Figure 8. The response of state variable 1.
Figure 9. The response of state variable 2.

Figure 10. The response of state variable 2.
Figure 11. The response of Control Input.

Figure 12. The response of Control Input using Adaptive Reinforcement Learning.
5. Conclusions

This paper investigated the optimal control design for perturbed switched nonlinear systems based on the adaptive dynamic programming technique. The optimal control is first designed under
the consideration of nominal systems. Then, the ADP technique is developed by using the Neural Networks. Due to the description of nonlinear systems, as well as unknown switch index, the neural network is considered to approximate the critic part of the iterative algorithm. Moreover, the UUB stability problem of the closed-loop system and the convergence of weight training are guaranteed under this solution. Finally, two simulation examples are given to verify the effectiveness of the presented ARL algorithm.

**Author Contributions:** P.N.D. was originally responsible for conceptualization, methodology, simulation and prepared the original draft of the article. H.Q.N., M.-D.N. verified the simulation data, data curation, investigation, and draft writing. S.-J.A. contribute a formal analysis of review, editing, and supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Korea Electric Power Corporation (Grant number: R18XA04).

**Acknowledgments:** The authors gratefully acknowledge Thai Nguyen University of Technology for supporting this work.

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

**Abbreviations**
The following abbreviations are used in this manuscript:

- ARL: Adaptive Reinforcement Learning
- ADP: Adaptive Dynamic Programming
- UUB: Uniformly Ultimately Bounded
- NNs: Neural Networks
- HJB: Hamilton–Jacobi–Bellman

**References**
1. Loschi, H.; Smolenski, R.; Lezynski, P.; Nascimento, D.; Demidova, G. Aggregated Conducted Electromagnetic Interference Generated by DC/DC Converters with Deterministic and Random Modulation. *Energies* 2020, 13, 3698. [CrossRef]
2. Bakeer, A.; Chub, A.; Vinnikov, D. Step-Up Series Resonant DC–DC Converter with Bidirectional-Switch-Based Boost Rectifier for Wide Input Voltage Range Photovoltaic Applications. *Energies* 2020, 13, 3747. [CrossRef]
3. Premkumar, M.; Subramaniam, U.; Haes Alhelou, H.; Siano, P. Design and Development of Non-Isolated Modified SEPIC DC-DC Converter Topology for High-Step-Up Applications: Investigation and Hardware Implementation. *Energies* 2020, 13, 3960. [CrossRef]
4. Korkh, O.; Blinov, A.; Vinnikov, D.; Chub, A. Review of Isolated Matrix Inverters: Topologies, Modulation Methods and Applications. *Energies* 2020, 13, 2394. [CrossRef]
5. Chen, B.-Y.; Shangguan, X.-C.; Jin, L.; Li, D.-Y. An Improved Stability Criterion for Load Frequency Control of Power Systems with Time-Varying Delays. *Energies* 2020, 13, 2101. [CrossRef]
6. Zhang, L.; Xiang, W. Mode-identifying time estimation and switching-delay tolerant control for switched systems: An elementary time unit approach. *Automatica* 2016, 64, 174–181. [CrossRef]
7. Yuan, S.; Zhang, L.; De Schutter, B.; Baldi, S. A novel Lyapunov function for a non-weighted L2 gain of asynchronously switched linear systems. *Automatica* 2018, 87, 310–317. [CrossRef]
8. Xiang, W.; Lam, J.; Li, P. On stability and $H_{\infty}$ control of switched systems with random switching signals. *Automatica* 2018, 95, 419–425. [CrossRef]
9. Lin, J.; Zhao, X.; Xiao, M.; Shen, J. Stabilization of discrete-time switched singular systems with state, output and switching delays. *J. Frankl. Inst.* 2019, 356, 2060–2089. [CrossRef]
10. Briat, C. Convex conditions for robust stabilization of uncertain switched systems with guaranteed minimum and mode-dependent dwell-time. *Syst. Control. Lett.* 2015, 78, 63–72. [CrossRef]
11. Lian, J.; Li, C. Event-triggered control for a class of switched uncertain nonlinear systems. *Syst. Control. Lett.* 2020, 135, 104592. [CrossRef]
12. Vu, T.A.; Nam, D.P.; Huong, P.T.V. Analysis and control design of transformerless high gain, high efficient buck-boost DC-DC converters. In Proceedings of the 2016 IEEE International Conference on Sustainable Energy Technologies (ICSET), Hanoi, Vietnam, 14–16 November 2016; pp. 72–77.

13. Nam, D.P.; Thang, B.M.; Thanh, N.T. Adaptive Tracking Control for a Boost DC–DC Converter: A Switched Systems Approach. In Proceedings of the 2018 4th International Conference on Green Technology and Sustainable Development (GTSD), Ho Chi Minh City, Vietnam, 23–24 November 2018; pp. 702–705.

14. Thanh, N.T.; Sam, P.N.; Nam, D.P. An Adaptive Backstepping Control for Switched Systems in presence of Control Input Constraint. In Proceedings of the 2019 International Conference on System Science and Engineering (ICSSSE), Quang Binh, Vietnam, 19–21 July 2019; pp. 196–200.

15. Chiang, M.-L.; Fu, L.-C. Adaptive stabilization of a class of uncertain switched nonlinear systems with backstepping control. *Automatica* **2014**, *50*, 2128–2135. [CrossRef]

16. Vamvoudakis, K.G.; Vrabie, D.; Lewis, F.L. Online adaptive algorithm for optimal control with integral reinforcement learning. *Int. J. Robust Nonlinear Control* **2014**, *24*, 2686–2710. [CrossRef]

17. Bai, W.; Zhou, Q.; Li, T.; Li, H. Adaptive reinforcement learning neural network control for uncertain nonlinear system with input saturation. *IEEE Trans. Cybern.* **2019**, *50*, 3433–3443. [CrossRef] [PubMed]

18. Zhang, T.; Xu, H. Adaptive optimal dynamic surface control of strict-feedback nonlinear systems with output constraints. *J. Franklin. Inst.* **2020**, *30*, 2059–2078. [CrossRef]

19. Lv, Y.; Na, J.; Yang, Q.; Wu, X.; Guo, Y. Online adaptive optimal control for continuous-time nonlinear systems with completely unknown dynamics. *Int. J. Control* **2016**, *89*, 99–112. [CrossRef]

20. Chen, C.; Modares, H.; Xie, K.; Lewis, F.L.; Wan, Y.; Xie, S. Reinforcement learning-based adaptive optimal exponential tracking control of linear systems with unknown dynamics. *IEEE Trans. Autom. Control* **2019**, *64*, 4423–4438. [CrossRef]

21. Yang, X.; Wei, Q. Adaptive Critic Learning for Constrained Optimal Event-Triggered Control with Discounted Cost. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**. [CrossRef]

22. Wen, G.; Ge, S.S.; Chen, C.L.P.; Tu, F.; Wang, S. Adaptive tracking control of surface vessel using optimized backstepping technique. *IEEE Trans. Cybern.* **2018**, *49*, 3420–3431. [CrossRef]

23. Wang, D.; Mu, C. Adaptive-critic-based robust trajectory tracking of uncertain dynamics and its application to a spring–mass–damper system. *IEEE Trans. Ind. Electron.* **2017**, *65*, 654–663. [CrossRef]

24. Li, S.; Ding, L.; Gao, H.; Liu, Y.-J.; Huang, L.; Deng, Z. ADP-based online tracking control of partially uncertain time-delayed nonlinear system and application to wheeled mobile robots. *IEEE Trans. Cybern.* **2019**, *50*, 3182–3194. [CrossRef]

25. Wen, G.; Chen, C.L.P.; Ge, S.S.; Yang, H.; Liu, X. Optimized adaptive nonlinear tracking control using actor–critic reinforcement learning strategy. *IEEE Trans. Ind. Inform.* **2019**, *15*, 4969–4977. [CrossRef]

26. Vamvoudakis, K.G.; Ferraz, H. Model-free event-triggered control algorithm for continuous-time linear systems with optimal performance. *Automatica* **2018**, *87*, 412–420. [CrossRef]

27. Gao, W.; Jiang, Y.; Jiang, Z.-P.; Chai, T. Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming. *Automatica* **2016**, *72*, 37–45. [CrossRef]

28. Mu, C.; Zhang, Y.; Gao, Z.; Sun, C. ADP-based robust tracking control for a class of nonlinear systems with unmatched uncertainties. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**. [CrossRef]

29. Huang, Y. Optimal guaranteed cost control of uncertain non-linear systems using adaptive dynamic programming with concurrent learning. *IET Control Theory Appl.* **2018**, *12*, 1025–1035. [CrossRef]

30. Tang, D.; Chen, L.; Tian, Z.F.; Hu, E. Modified value-function-approximation for synchronous policy iteration with single-critic configuration for nonlinear optimal control. *Int. J. Control* **2019**, 1–13. [CrossRef]

31. Fan, B.; Yang, Q.; Tang, X.; Sun, Y. Robust ADP design for continuous-time nonlinear systems with output constraints. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 2127–2138. [CrossRef]

32. Chun, T.Y.; Lee, J.Y.; Park, J.B.; Choi, Y.H. Adaptive dynamic programming for discrete-time linear quadratic regulation based on multirate generalised policy iteration. *Int. J. Control* **2018**, *91*, 1223–1240. [CrossRef]

33. Mu, C.; Wang, D. Neural-network-based adaptive guaranteed cost control of nonlinear dynamical systems with matched uncertainties. *Neurocomputing* **2017**, *245*, 46–54. [CrossRef]

© 2020 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 (http://creativecommons.org/licenses/by/4.0/).