Research Article

A Direct Reinforcement Learning Approach for Nonautonomous Thermoacoustic Generator

Trong Tan Do,¹,² Duc Chuyen Tran,³, Duy Tung Le,¹ and Phuong Nam Dao¹

¹School of Electrical Electronic Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam
²Center for Micro Electronics and Information Technology, Hanoi, Vietnam
³University of Economics-Technology for Industries, Hanoi, Vietnam

Correspondence should be addressed to Duc Chuyen Tran; tdchuyen@uneti.edu.vn and Phuong Nam Dao; nam.daophuong@hust.edu.vn

Received 20 February 2022; Accepted 5 April 2022; Published 31 May 2022

Copyright © 2022 Trong Tan Do et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For nonautonomous nonlinear systems, the optimal control design is affected by the terms of partial derivative. If a reinforcement learning (RL) strategy is developed to approximate the optimal control scheme in nonautonomous nonlinear systems, then the closed control system might be unstabilizing. Therefore, in this article, the approach of direct RL law for a nonautonomous thermoacoustic generator (TAG) is investigated. We establish the mathematical model of TAG by partial differential equations (PDEs) and then transforming them into time varying nonlinear systems. The direct RL technique with Newton–Leibniz formula is implemented to consider the partial derivative term from classical policy iteration (PI) method by modifying the computation using data collection between the two sampling times. Finally, several simulation studies with some comparisons are conducted to validate the theoretical analyses.

1. Introduction

As a key problem of energy mission, the thermoacoustic generator (TAG) has been attracted by many scholars [1]. The consideration of transforming from high-temperature source to the high-efficient heat engine process as well as the process of converting heat to electricity were investigated in [1]. However, the main shortcoming is that the control design problem of TAG has not been much discussed. In the actual engineering system, optimal control is a remarkable technique, which gives their essential effectiveness for balancing the tracking problem and performance. Regarding the theoretical problem, for achieving optimal control design, one needs to solve the Hamilton-Jacobi-Bellman (HJB) equation in nonlinear systems or Riccati equation in linear systems subject to a user-defined performance index [2]. However, due to the challenge of analytically solving these equations as well as the existence of uncertainties, adaptive reinforcement learning (RL)-based techniques have been employed to approximate optimal control solution [2–4]. On the other side, it should be noted that some different methods can be developed without RL technique, such as the approach in [5] to solve robust LQR in the presence of constraint by Min-Max optimization.

As an important approach in the modern control technique, the RL-based control method aims to obtain the minimization of performance index while achieving the stability of closed systems. Up to now, many important researches have been implemented for linear systems using Kronecker product [6] to conveniently compute the quadratic form as well as for nonlinear systems by approximating with neural networks [7] to describe Bellman function. However, due to the time varying description of closed systems, it is challenging to investigate RL-based control strategy for practical systems such as robotics and generators. To overcome this challenge, the model transformation method [8] and direct RL technique for nonautonomous systems [9–11] are introduced. The technique of transforming closed systems is implemented by considering the desired trajectory as new state variables and changing the cost function by a different form [8]. After obtaining the autonomous systems, the online actor-critic strategy was discussed in
by generating the adaptation law of weights in actor/critic neural networks, which approximated the optimal control and Bellman function. Thanks to the property of Hamiltonian, it implies the method of training was achieved by minimizing square of Hamiltonian term. The classical actor-critic can also be extended for nonlinear continuous time-delayed dynamical systems [12] by adding the time delay-based integral term into optimal value functional. Moreover, it leads to the modification of the integral temporal difference error (ITDE) depending on time delay [12]. However, the implementation of traditional actor-critic strategy [8] usually requires certain model in the computation. For nonlinear systems containing unmodeled dynamics, Yang et al. in [13] proposed a robust online actor-critic strategy with the addition of robustifying term combining with fuzzy logic systems-based approximator. Likewise, RL method is developed for general HJB problem with the additional variable [3], HJI based approximator. RL controls solution for time varying systems [9–11] due to the existence of the term $\partial V(x, t)/\partial t$. Authors in [9] improved conventional policy iteration (PI) technique with the addition of partial derivative with respect to time in each step. Because of nonlinear property, two NNs are also utilized in both Actor NN and Critic NN with the weights to be trained using the addition of partial derivative in terms of states [9]. However, the direct RL techniques in [9–11] are only purely mathematical analyses. A different approach of direct RL control can be mentioned as off-policy technique [20, 21]. Due to the property of keeping the input control while computing RL algorithm, the off-policy technique is able to address to completely uncertain systems in linear systems [20] and nonlinear systems [21]. Therefore, it can be determined that direct RL-based controllers for time varying closed practical systems are challenging issues and these motivate us to study this problem in TAG systems.

Inspired by the above analysis, we investigate the application of direct RL procedure in control system of time varying TAG systems, which are described by PDE. We use some appropriate modifications to transform PDE of TAG systems into a time varying dynamic equation. The major contributions of this article are given as follows:

(1) First, it is obviously different from [2, 22] realizing RL procedure for robotic systems using differential equations; the structure of time varying RL is presented to obtain the optimal control for time varying TAG systems to be expressed by PDE.

(2) Second, differential from researches of TAG systems [1], optimal control strategy is investigated with the related comparison to be discussed through theoretical analysis and simulation studies.

The remainder of this article is organized as follows. Section 2 gives the problem statement of this article. Section 3 focuses on analysis of mathematical modelling of TAG systems. The direct RL procedure-based control scheme for TAGs is discussed in Section 4. Simulation studies in a TAG control system are presented in Section 5 and the conclusion remarks are presented in Section 6.

2. Preliminaries and Problem Statement

The thermoacoustic generator (TAG) is a device that can generate thermal energy or consume acoustic energy to transfer heat from low-temperature to high-temperature sources; from there, electricity can be obtained through electromechanical converters. In this paper, we investigate a fundamental structure of thermoacoustic generator (TAG), as shown in Figure 1, which includes 5 parts: regenerator (REG), heat exchanger (HHX, CHX), alternator (ALT), stub (STUB), and feedback pipe (FBP). According to thermoacoustic theory and the partial element equivalent circuit (PEEC) method [23], we construct a theoretical model for thermoacoustic generator from the equivalent parts model. Then, we apply the adaptive dynamic programming method to TAG from the obtained mathematical model. Moreover, based on physical phenomena of TAGs, some assumptions are required to represent TAGs by several partial derivative equations (PDEs) as well as equivalent circuits in next sections.
3. Mathematical Modelling of Thermoacoustic Generator Systems

In this section, we proceed to model the thermoacoustic generator based on the operating principles combined with the partial element equivalent circuit (PEEC) method [23] to give the analogous circuit structure of the object. Then, we provide the mathematical model based on assumptions about the thermoacoustic generator’s operating conditions and mathematical transformations.

3.1. Regenerator. According to the linear thermoacoustic theory [1], the interaction between the acoustic and temperature fields can be given by the following equation (1) as

\[
[1 - (\gamma - 1)f_k]P_1 + \frac{y_p}{\omega} \frac{d}{dx} \left( 1 - \frac{f_k}{\rho_M} \frac{dp_1}{dx} \right)
\]

\[
- \frac{a}{\omega} \frac{f_k - f_v}{1 - \sigma} \frac{dT_M}{T_M} \frac{dp_1}{dx} = 0,
\]

where \( \rho, U, T \) are the pressure (Pa), flow (m\(^3\)/s), and temperature of the gas (K), respectively; \( \rho, \gamma \) are specific weight (kg/m\(^3\)) and the adiabatic ratio of gases; \( \omega, a \) are the angular frequency (s\(^{-1}\)) and the speed of sound (m/s); \( f_k, f_v \) are viscous function and spatially averaged thermal function, respectively. For the convenience of analysis, we can separate equation (1) into the two following dynamic equations [24]:

\[
\frac{dp_1}{dx} = \frac{-i\omega p_M}{A(1 - f_v)} U_1,
\]

\[
\frac{dU_1}{dx} = \frac{i\omega A}{\gamma p_M} [1 + (\gamma - 1)f_k]P_1 + \frac{(f_k - f_v)}{(1 - \sigma)(1 - \sigma)} \frac{dT_M}{T_M} U_1.
\]

From the PEEC method and the mathematical transformations, we can obtain the formula for the equivalent current source and the resistance of regenerator as follows:

\[
U_1 = \frac{T_h - T_c}{T_c} A \rho_M \frac{-1}{\rho_M} \frac{2\mu}{\rho_M} \left( \frac{1 - 1}{r_0} \frac{2\mu}{\rho_M} \right) \frac{dp_1}{dx},
\]

\[
R_{REG} = \frac{\omega p_M l_{stack}}{A} \left[ \frac{1}{(2 - r_0/\sqrt{\omega p_M/2\mu}) - 2(r_0/\sqrt{2\mu\omega p_M})} \right]
\]

where \( T_h, T_c \) are the temperatures of the hot heat source and cold heat source (K); \( r_0, l_{stack} \) are the diameter (m\(^2\)) and length of the stack (m); and \( \mu \) is the absolute viscosity coefficient of the gas (Pa · s).

3.2. Alternator. A simple linear model [25] characterizing the loudspeaker as a linear alternator is shown in Figure 2. The acoustic wave imposes an oscillating pressure on the diaphragm, which has an effective area \( S \), as shown in Figures 2(a) and 2(b) and shows the equivalent circuit of the physical model, as shown in Figure 2(a). The diaphragm and the coil, with a total mass, \( M_m \), are subjected to oscillatory motion. The loudspeaker has a mechanical stiffness, \( K_m \), and a mechanical resistance, \( R_m \). The coil has an electrical inductance, \( L_e \), and an electrical resistance, \( R_e \). The force factor is \( Bl \). A pure electrical resistance, \( R_L \), is connected as a load to extract electrical power in this model. The voltage on the load resistance is \( V_L \), and the current is \( I_L \).

Assume that all parameters are linear and frequency-independent. Ignoring hysteresis losses, the alternator’s impedance in Figure 2 can be written as [22].

Assumption 1. The necessary assumptions to apply the linear thermoacoustic theory into modelling the thermoacoustic generator are as follows:

1. The material’s surface is smooth, the heat radiation is negligible, and the plates are rigid and stationary.
2. The acoustic pressure is \( x \) direction only and the viscosity is independent of temperature.
3. The length of the plates is small compared to the size of the resonator.

The control objective is to establish the control input \( u(x) \) of TAG systems for minimizing the performance index under the consideration of time varying systems. The control design is implemented by approximating with RL technique after obtaining the model of TAG.

Remark 1. It is worth emphasizing that unlike the work in [2,22] studying RL technique for time varying control system of robotics by indirect methods, the proposed method in this article develops the direct RL method for time varying TAG systems.
Notice that the pressure loss is much less than in traditional systems since the reactive component of the impedance is absolutely small even when the alternator is off resonance by a few Hz [26]:

$$\omega M_m - \frac{K_m}{\omega} \ll \left(\frac{B^2}{R_c + R_L} + R_m\right).$$

(5)

Therefore, we can rewrite alternator’s equivalent impedance as

$$Z_{ALT} = \frac{1}{A_{ALT}^2} \left[ \frac{(B^2)}{R_c + R_L} + R_m \right].$$

(6)

3.3. Stub. The stub is a piston placed perpendicular to the central conduit and shown in Figure 1. The principal objective of stubs is to accommodate for the pressure and flow phase shift after passing through alternator. Considering a straight acoustic duct, the relationship between the input $Z_1$ and the output acoustic impedance $Z_2$ can be expressed as [26]

$$Z_1 = Z_0 Z_2 + j Z_0 \tan kl.$$  

(7)

Here, $l$ is the length of duct ($n$), $k = 2\pi/\lambda$ is the acoustic propagation coefficient, and $Z_0 = \rho M a/A$ is the characteristic impedance of duct. The stub is basically a closed-end portion of an acoustic duct, $Z_2 = \infty$. Therefore, the input acoustic impedance of the stub can be approximately written as

$$Z_s = j \omega L_s + \frac{1}{j \omega C_s}.$$  

(8)

where

$$L_s = \frac{I \rho M}{A} (H),$$  

(9)

$$C_s = \frac{V}{\rho M a^2} (F).$$

3.4. Heat Exchanger and Feedback Pipe. The heat exchangers are utilized as heat recovery equipment and are suitable for intake and return air systems. Two heat exchangers are employed in the thermoacoustic generator construction to recover and maintain the temperature of the hot and cold heat sources. The heat exchanger has a very low porosity while having a relatively long length. Compared to the thermoacoustic core’s enormous cross-sectional area, the heat exchanger effectively adds a long but short cross section channel into the loop locally. As a result, from an acoustic standpoint, the heat exchanger exhibits a considerable inertance effect and the average acoustic resistance (Figure 3)

$$Z_{CHX} = R_{CHX} + i \omega L_{CHX}.$$  

(10)

The remainder of the feedback pipe is just a lossy acoustic waveguide. Each unit section can be modelled as a mixture of resistance $R_{fb}$, inductance $L_{fb}$, and capacitance $C_{fb}$, all of which can be determined using equations (8) and (9).

After modelling the five elements of thermoacoustic generator, we merge the results into the closed-circuit shown in Figure 4.

Assumption 2. Attempts to simplify the thermoacoustic generator’s equivalent circuit:

1. The gas movement is optimal, neglecting all of the friction with the duct. As a result, the energy of the gas flow is conserved, allowing us to ignore the influence of the feedback pipe (FBP) component throughout the circuit survey.
(2) In the heat exchanger, the equivalent inductance is very small compared to the impedance; at the same time, the impedance value can be adjusted through the solvent flow in the device. Therefore, the heat exchanger can be equivalent to a rheostat in the analogous circuit for modelling convenience.
Denote \( u = T_h - T_c / T_o \), \( x \) is the pressure of gas (Pa), \( y \) is the output power on alternator (W), and the thermoacoustic generator’s mathematical model is as follows:

\[
\begin{align*}
\dot{x} &= \frac{y_{ax}}{u} \sin (\omega t + \phi), \\
y &= \frac{A^2 y_{x}}{\rho_M} \left[ \frac{(B l)^2}{R_c + R_L} + R_m \right] \cdot Z_{H}^\gamma \sin (\omega t + \theta).
\end{align*}
\]

The phase difference angles are defined as follows:

\[
\begin{align*}
\varphi &= \arctan \left( r_0 \sqrt{\frac{\omega \rho M}{2 \mu}} - 1 \right), \\
\theta &= \arctan \left( \frac{-2PR/Q^2 + R^2 \left( S + Z_{H}^\gamma + P/Q^2 + R^2 \right)}{(S + Z_{H}^\gamma + P/Q^2 + R^2)^2 - (PC/Q^2 + R^2)^2} \right).
\end{align*}
\]

(12)

Remark 2. It is worth emphasizing that the thermoacoustic generator system can be enabled when the temperature differential between the hot and cold sources is positive. It means that the input control signal is always positive, and it is the system’s input constraint. Furthermore, the pressure of gas (state) is always positive when the system is performed.

4. Direct Reinforcement Learning for Nonautonomous TAG Systems

In this section, we consider the extension from classical RL algorithm [2, 22, 27] to time varying RL for the general class of time varying nonlinear systems as

\[
\dot{x}(t) = f(x(t), u(t), t), \quad (\forall t \geq t_0).
\]

(13)

With the initial value, \( x(t_0) = x_0 \), and the associated cost function is defined as

\[
J(x_0, t_0, u(t)) = \int_{t_0}^{\infty} L(x(t), u(t), t) dt,
\]

(14)

where \( x(t) \in \mathbb{R}^n \) is the arbitrary states vector and \( u(t) \in \mathbb{R}^m \) is the vector of input signals. Additionally, \( f : \mathbb{R}^n \times \mathbb{R}^m \times [t_0, \infty) \rightarrow \mathbb{R}^n \) is a continuously differentiable and positive definite map and the partial derivatives in terms of \( x, u \), and \( t \) of \( L : \mathbb{R}^n \times \mathbb{R}^m \times [t_0, \infty) \rightarrow \mathbb{R} \) are continuous in \([t_0, \infty)\). The optimal control objective is to minimize the cost function \( J(x_0, t_0, u(t)) \), which is bounded to the class of admissible signals to be defined as follows.

Definition 1. (see [4]) The control signal \( u : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is considered as an admissible control signal if and only if

1. \( u(x(t)) \) is a piecewise continuous function with respect to time.

2. The solution \( x(t) \) of the time varying nonlinear system \((13)\) under the input signal \( u(x(t)) \) is converged to zero as time comes to infinity \((t \rightarrow \infty)\).

3. The integral \((14)\) is finite: \( \int_{0}^{\infty} L(x(t), u(x(t)), t) dt < \infty \) for all \( x_0 \).

According to dynamic programming (DP) principle [22, 27, 28], implementing the derivative with respect to time of optimal value \( f^*(x(t), t) \) along the system trajectory \((13)\), one can obtain the following equation:

\[
\frac{\partial f^*(x(t), t)}{\partial t} = \min_{u \in \mathbb{R}^m} \left\{ L(x(t), u(t), t) + (\partial f^*(x(t), t)/\partial x(t))^T f(x(t), u(t), t) \right\},
\]

(15)

where \( f \) is a continuously differentiable function with \( f \) being positive definite. Based on the view of the relations \( f^*(x(t), t) \rightarrow \nu^*(x(t)) \) (15) and \( \nu^*(x(t)) \rightarrow V^*(x(t), t) \) (16), the following iterative algorithm can be applied to the optimization problem of thermoacoustic generator systems.

In the light of [10], the improvement of traditional Algorithm 1 for the case of uncertain nonlinear systems can be completed after computing the deviation of two time varying value functions \( V^*(x(t), t) \) at the two different sampling times \( t_k \) and \( t_{k+1} \) by approximating with Newton–Leibniz formula and combining with function approximation theory. Therefore, Step 3 with Algorithm 1 in classical Algorithm 1 is transformed into the computation using data collection by two following steps. First, this deviation of two time varying value functions is calculated using Newton–Leibniz formula under the control input \( u_0 \) in time intervals \([t_k, t_{k+1}]\):

\[
\begin{align*}
V^*(x(t_k), t_k) - V^*(x(t_{k-1}), t_{k-1}) &= \\
&= \int_{t_{k-1}}^{t_k} \left( V^*_t(x(t), t) - V^*_t(x(t_{k-1}), t_{k-1}) \right) dt \\
&= \int_{t_{k-1}}^{t_k} \left( \frac{d}{dt} V^*_t(x(t), t) \right) dt \\
&= \int_{t_{k-1}}^{t_k} \left( \frac{d}{dt} V^*_t(x(t), t) \right) dt + \int_{t_{k-1}}^{t_k} L(x(t), \nu^{i-1}(x(t)), t) dt,
\end{align*}
\]

(17)

Second, since \( f (\cdot), \{ \nu^0 (\cdot) \}, \{ \nu^1 (\cdot) \}, \{ \nu^2 (\cdot) \} \) and \( \nu^{i-1} (\cdot) \) for \( i = 1, 2, \ldots \) are unknown functions, by using the basis function method [29], Step 3 in traditional Algorithm 1 can be easily realized with updation law of weights as follows [10]:

\[
\begin{align*}
\nu^i (x(t), t) &= \nu^{i-1} (x(t), t) + \alpha \left( \frac{d}{dt} V^*_t(x(t), t) - L(x(t), \nu^{i-1}(x(t)), t) - \frac{d}{dt} V^*_t(x(t), t) \right) dt,
\end{align*}
\]
where \( e_i^{-1}(x(t)), e_i^j(x(t), t), e_i^j(\cdot), e_i^j(\cdot) \) are approximate errors, which converge to zero as the iteration step comes to infinity. Based on the above analysis and transformation, a numerical RL algorithm can be proposed.

Following that, we proceed to apply Algorithm 2 to the mathematical model of thermoacoustic generator (11) as analyzed:

\[
\begin{align*}
\dot{x} &= \frac{\gamma u x}{u} \sin(\omega t + \varphi), \\
y &= \frac{A^2 \gamma x}{\rho M} \left[ \frac{(B)^2}{R_c + R_L} \right] Z_{\text{HX}}^3 \sin(\omega t + \theta).
\end{align*}
\]

(19)

With the associated cost function,

\[
J(x_0, y_0, u(t)) = \int_{t_0}^{t_f} \left[ (x - x_0)^2 + (y - y_0)^2 + (u - u_0)^2 \right] dt,
\]

(20)

where \( x_0, u_0, y_0 \) are the initial state, input, and output signal, respectively. From the HJB equation, we employ the basis function method to approximate the unknown functions as (18) with \( N_1 = N_2 = N_3 = N_4 = 1 \) which are the number of basis function in each function vector. The optimal weights are determined according to Algorithm 2, where

\[
\Theta_i = [P_{k,1}, Q_{k,1}, R_{k,1}].
\]

(21)

Furthermore, the terms \( P_{k,1}, Q_{k,1}, R_{k,1} \) are obtained from the collected data as follows:

\[
\begin{align*}
P_{k,1} &= \phi_1(x(t_k), t_k) - \phi_1(x(t_{k-1}), t_{k-1}), \\
Q_{k,1} &= -\int_{t_{k-1}}^{t_k} \left( \phi_1(x, u_0, s) + \phi_1(x, t_1^{-1} \psi_1(x), s) \right) ds, \\
R_{k,1} &= -2 \int_{t_{k-1}}^{t_k} \alpha_1(x, s) ds.
\end{align*}
\]

(22)

Afterward, the optimal weight parameter \( \tilde{u} \) can be obtained by solving the problem as Algorithm 2 and calculating the control signal \( \tilde{v} \) via (18).

**Remark 3.** The requirement of convergences of control policy \( \tilde{v}(x(t)) \) as well as value function \( \tilde{V}(x(t), t) \) in Algorithm 2 is proved as described in [10] with the following steps. First, it is necessary to point out the property of admissible control input [4] \( v_i(x(t)) \) and positive definite optimal function \( V_i(x(t), t) \) which are kept after each iteration of Algorithm 1. Second, the decrement of optimal function \( V_i(x(t), t) \) is determined and the convergences of control policy \( \tilde{v}(x(t)) \) and value function \( \tilde{V}(x(t), t) \) in Algorithm 2 are guaranteed after obtaining the estimation of error between the control policies \( v_i(x(t)) \) and \( \tilde{v}_i(x(t)) \) as...
well as optimal functions $V_i(x(t), t)$ and $\tilde{V}_i(x(t), t)$ to be obtained in Algorithms 1 and 2. However, unlike the work in [10], the time varying RL is completely developed for a practical TAG system (19). Furthermore, it is obviously different from the method of time varying RL in [9] developing the actor/critic technique in [22]; RL procedure is implemented under the consideration of data collection in Algorithm 2. Additionally, to fulfill the term of partial derivative in HJB equation of nonautonomous systems, it is obviously different from indirect RL method studying the equivalent systems by adding more state variables [8, 22], the direct RL control for TAGs is able to keep the dynamic model without transforming systems.

5. Simulation Results

In this section, a TAG system with the parameters is to be given for control system as follows: $T_c = 300, t_0 = 0$, the sampling time $t_s = 0.05\, \text{s}$, the initial state and initial control signal are $x_0 = 1085000\, (\text{Pa}), u_0 = 0.2$, respectively, the approximation error $\epsilon = 0.01$, the designed value of state $x_m = 1.1 \cdot 10^5\, (\text{Pa})$, and the desired output $y_m = 65\, (\text{W})$. To
validate the effectiveness of the proposed Algorithm 2 with time varying RL, we implement two simulations using Algorithm 2 (Figures 4–6) and traditional PID control scheme (Figures 7–9). It is seen that the high performance of control input, state variable, and output signal is implemented by time varying RL control for TAG in comparison with traditional PID controller. The control signal of TAGs (Figure 7) is oscillated with high frequency in traditional PID controller. The disadvantage follows the oscillators in output and state signals of TAGs (Figures 8 and 9). In contrast to traditional PID control scheme, the control signal and state are stable under RL algorithm (Figures 4 and 5) as well as the average of output signal convergences to desired value (Figure 6).

6. Conclusion
This article proposes an application of direct time varying RL strategy to solve the optimal control for nonautonomous TAG systems subject to unknown system model parameters.
The PDE-based mathematical model of TAGs is transformed into time varying nonlinear dynamical systems. After that by collecting the data between two sampling times under the control signals and Newton–Leibniz approximation, the conventional RL technique is modified to handle for non-autonomous TAG systems. Numerical simulations and the comparison with traditional method of TAG control system verify the high performance of the proposed method.

**Data Availability**

This publication is supported by multiple datasets, which are available at locations cited in the reference section.

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**References**

[1] S. Backhaus, E. Tward, and M. Petach, “Traveling-wave thermoacoustic electric generator,” *Applied Physics Letters*, vol. 85, no. 6, pp. 1085–1087, 2004.

[2] P. N. Dao and Y. C. Liu, “Adaptive reinforcement learning strategy with sliding mode control for unknown and disturbed wheeled inverted pendulum,” *International Journal of Control, Automation and Systems*, vol. 19, pp. 1139–1150, 2021.

[3] T. Bian and Z. P. Jiang, “Reinforcement learning and adaptive optimal control for continuous-time nonlinear systems: a value iteration approach,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2021.

[4] H. Zhang, Q. Qu, G. Xiao, and Y. Cui, “Optimal guaranteed cost sliding mode control for constrained-input nonlinear systems with matched and unmatched disturbances,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2112–2126, 2018.

[5] A. Scampicchio, A. Aravkin, and G. Pillonetto, “Stable and robust lqr design via scenario approach,” *Automatica*, vol. 129, Article ID 109571, 2021.

[6] H. N. Wu and Z. Y. Liu, “Data-driven guaranteed cost control design via reinforcement learning for linear systems with parameter uncertainties,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 11, pp. 4151–4159, 2019.

[7] F. Adib Yaghaie and D. J. Braun, “Reinforcement learning for a class of continuous-time input constrained optimal control problems,” *Automatica*, vol. 99, pp. 221–227, 2019.

[8] P. N. Dao, P. T. Loc, T. Q. Huy, and V. T. Vu, “Sliding variable-based online adaptive reinforcement learning of uncertain/disturbed nonlinear mechanical systems,” *Journal of Control, Automation and Electrical Systems*, vol. 32, pp. 281–290, 2021.

[9] Q. Wei, Z. Liao, Z. Yang, B. Li, and D. Liu, “Continuous-time time-varying policy iteration,” *IEEE Transactions on Cybernetics*, vol. 50, no. 12, pp. 4958–4971, 2019.

[10] T. Sun and X. M. Sun, “An adaptive dynamic programming scheme for nonlinear optimal control with unknown dynamics and its application to turbofan engines,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 367–376, 2020.

[11] J. Sun, H. Zhang, Y. Yan, S. Xu, and X. Fan, “Optimal regulation strategy for nonzero-sum games of the immune system using adaptive dynamic programming,” *IEEE Transactions on Cybernetics*, pp. 1–10, 2021.

[12] R. Moghadam and S. Jagannathan, “Optimal adaptive control of uncertain nonlinear continuous-time systems with input and state delays,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2021.

[13] Y. Yang, W. Gao, H. Modares, and C. Z. Xu, “Robust actor-critic learning for continuous-time nonlinear systems with unmodeled dynamics,” *IEEE Transactions on Fuzzy Systems*, 2021.

[14] S. A. Arogeti and F. L. Lewis, “Static output-feedback h control design procedures for continuous-time systems with different levels of model knowledge,” *IEEE Transactions on Cybernetics*, pp. 1–15, 2021.

[15] S. Li, P. Durdevic, and Z. Yang, “Model-free Hoo tracking control for de-oiling hydrocyclone systems via off-policy reinforcement learning,” *Automatica*, vol. 133, Article ID 109862, 2021.

[16] H. Zhang, Z. Ming, Y. Yan, and W. Wang, “Data-driven finite-horizon h tracking control with event-triggered mechanism for the continuous-time nonlinear systems,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2021.

[17] K. G. Vamvoudakis, “Q-learning for continuous-time linear systems: a model-free infinite horizon optimal control approach,” *Systems & Control Letters*, vol. 100, pp. 14–20, 2017.

[18] Q. Wei, J. Lu, T. Zhou, X. Cheng, and F. Wang, “Event-triggered near-optimal control of discrete-time constrained nonlinear systems with application to a boiler-turbine system,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 3926–3935, 2021.

[19] N. Chen, B. Li, B. Luo, W. Gui, and C. Yang, “Event-triggered optimal control for temperature field of roller kiln based on adaptive dynamic programming,” *IEEE Transactions on Cybernetics*, pp. 1–13, 2021.

[20] Y. Yang, B. Kiumarsi, H. Modares, and C. Xu, “Model-free λ-policy iteration for discrete-time linear quadratic regulation,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2021.

[21] Y. Yang, H. Modares, K. G. Vamvoudakis, W. He, C. Z. Xu, and D. C. Wunsch, “Hamiltonian-driven adaptive dynamic programming with approximation errors,” *IEEE Transactions on Cybernetics*, pp. 1–12, 2021.

[22] P. N. Dao, D. K. Do, and D. K. Nguyen, “Adaptive reinforcement learning-enhanced motion/force control strategy for multirobot systems,” *Mathematical Problems in Engineering*, vol. 2021, Article ID 5560277, 2021.

[23] L. Lombardi, R. Raimondo, and G. Antonini, “Electrothermal formulation of the partial element equivalent circuit method,” *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 31, no. 4, Article ID e2253, 2018.

[24] G. Swift, *A Unifying Perspective for Some Engines and Refrigerators*, Acoustical Society of America, Melville, 2002.

[25] Z. Yu, P. Saechan, and A. J. Jaworski, “A method of characterising performance of audio loudspeakers for linear alternator applications in low-cost thermoacoustic electricity generators,” *Applied Acoustics*, vol. 72, no. 5, pp. 260–267, 2011.
[26] Z. Yu, A. J. Jaworski, and S. Backhaus, “Travelling-wave thermoacoustic electricity generator using an ultra-compliant alternator for utilization of low-grade thermal energy,” Applied Energy, vol. 99, pp. 135–145, 2012.

[27] F. L. Lewis, D. Vrabie, and V. L. Syrmos, Optimal Control, John Wiley & Sons, Hoboken, NJ, USA, 2012.

[28] D. Bertsekas, Dynamic Programming and Optimal Control, I and II, Athena Scientific, Belmont, Massachusetts, 1995.

[29] H. Wendland, “Computational aspects of radial basis function approximation,” vol. 12, pp. 231–256, in Proceedings of the Studies in Computational Mathematics, vol. 12, Elsevier, Amsterdam, Netherlands, December 2006.