ISSN: 2476-6909; Modares. Mech. Eng. *2026;26(04):241-251.*

Modares Mechanical Engineering

DOI: 10.48311/mme.2025.117163.82870
journal homepage: mme.modares.ac.ir

# Reinforcement Learning and Sliding Mode Hybrid Controller for an Enhanced Inverted Pendulum on Cart System

Ali Jalali, Navid Mohammadi*, Morteza Tayefi

Institute of Intelligent Control Systems, K. N. Toosi University of Technology, Tehran, Iran

**ARTICLE INFO**

**Article Type**
Original Research

**Article History**
Received: October 20, 2025
Revised: November 26, 2025
Accepted: December 06, 2025
ePublished: February 16, 2026

**ABSTRACT**

This paper presents a comprehensive control study of an inverted pendulum on cart system enhanced with torsional spring-damper dynamics and cart damping. A Linear Quadratic Regulator (LQR) was designed using a linearized model, while Model Predictive Control (MPC) and Reinforcement Learning (RL) controllers were augmented with Monte Carlo simulations to evaluate robustness and sensitivity. Results demonstrated comparable performance among all methods in linear regimes, with stabilization times under 20 seconds and overshoot variation of 3%. To address nonlinear dynamics, a hybrid SMC-RL strategy was developed, reducing settling time and improving capability of maintaining stability under nonlinear behavior and large initial angles to 120-150°. The proposed SMC-RL framework achieved a success rate in stabilizing the system from diverse initial conditions, significantly outperforming standalone controllers in transient response and adaptability. System stability was formally verified through Lyapunov analysis and empirically confirmed by Monte Carlo simulations, which demonstrated consistent performance with minimal standard deviation across 80 randomized trials.

**Keywords:** Hybrid SMC-RL Controller, Reinforcement Learning, Model Predictive Control, Monte Carlo Simulation, Inverted Pendulum on cart

**How to cite this article**
Jalali A, Mohammadi N, Tayefi M, Reinforcement Learning and Sliding Mode Hybrid Controller for an Enhanced Inverted Pendulum on Cart System. Modares Mechanical Engineering; *2026;26(04):241-251.*

* *Corresponding author's email: navid.mohammadi@email.kntu.ac.ir*
**Corresponding ORCID ID:0009-0008-4207-3781*

شاپا: ۷۶-۶۹۰۹ ۲۴۷۶؛ مهندسی مکانیک مدرس. ۲۵۱-۲۴۱:(۰۴)۲۶;۱۴۰۵

مهندسی مکانیک مدرس

DOI: 10.48311/mme.2025.117163.82870

صفحه اصلی مجله: mme.modares.ac.ir

# کنترل‌کننده هیبریدی یادگیری تقویتی و مد لغزشی برای سیستم بهبودیافته آونگ معکوس بروی گاری

علی جلالی، نوید محمدی*، مرتضی طایفی

پژوهشکده سیستم‌های کنترل هوشمند، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

**اطلاعات مقاله**

**نوع مقاله**

مقاله پژوهشی

**تاریخچه مقاله**

دریافت: ۱۴۰۴/۰۷/۲۸

بازنگری: ۱۴۰۴/۰۹/۰۵

پذیرش: ۱۴۰۴/۰۹/۱۵

ارائه آنلاین: ۱۴۰۴/۱۱/۲۷

**چکیده**

این مقاله به ارائه یک مطالعه جامع در زمینه کنترل آونگ معکوس روی گاری می‌پردازد که با افزودن دینامیک فنر-دمپر پیچشی و دمپر گاری بهبود یافته است. برای این منظور، یک تنظیم‌کننده بهینه مرتبه دوم (LQR) با استفاده از مدل خطی‌سازی‌شده طراحی شد. همچنین کنترل‌کننده‌های پیش‌بین مدل (MPC) و یادگیری تقویتی (RL) با شبیه‌سازی‌های مونت‌کارلویی ترکیب شدند تا مقاومت و حساسیت سیستم ارزیابی گردد. نتایج در ناحیه خطی عملکرد قابل مقایسه‌ای را بین تمام روش‌ها نشان دادند که در آن زمان نشست زیر ۲۰ ثانیه و تغییرات فراجهش در حدود ۳٪ ثبت شد. به منظور مقابله با دینامیک‌های غیرخطی، یک راهبرد کنترلی هیبریدی مد لغزشی-یادگیری تقویتی (SMC-RL) توسعه یافت که منجر به کاهش زمان نشست و بهبود چشمگیر قابلیت حفظ پایداری در شرایط رفتار غیرخطی و زوایای اولیه بزرگ تا ۱۲۰-۱۵۰ درجه گردید. چارچوب پیشنهادی SMC-RL نرخ موفقیت بالایی در پایدارسازی سیستم از شرایط اولیه گوناگون به دست آورد و به طور معناداری در پاسخ گذرا و پایداری، از کنترلگرهای منفرد پیشی گرفت. پایداری سیستم از طریق معیار لیاپانوف تأیید و به صورت تجربی توسط شبیه‌سازی‌های مونت‌کارلو اثبات شد؛ به‌گونه‌ای که عملکردی پایدار با حداقل انحراف معیار در ۸۰ آزمون تصادفی را نشان داد.

**کلیدواژه‌ها:** کنترل هیبریدی SMC–RL، یادگیری تقویتی، کنترل پیش‌بین مدل، شبیه‌سازی مونت‌کارلو، آونگ معکوس روی گاری

**نحوه ارجاع به این مقاله**

جلالی علی، محمدی نوید، طایفی مرتضی، کنترل‌کننده هیبریدی یادگیری تقویتی و مد لغزشی برای سیستم بهبودیافته آونگ معکوس بروی گاری ، مهندسی مکانیک مدرس. ۲۵۱-۲۴۱:(۰۴)۲۶;۱۴۰۵

*پست الکترونیکی نویسنده عهده‌دار مکاتبات: navid.mohammadi@email.kntu.ac.ir

*شناسه ارکید نویسنده عهده‌دار مکاتبات: ۰۰۰۹-۰۰۰۸-۴۲۰۷-۳۷۸۱

## 1- Introduction

The inverted pendulum on a cart is one of the most traditional and classical control system problems, widely studied as a model for testing and validating various dynamic systems, including hoverboards, Maglev Systems (Magnetic Levitation Systems), balancing Robots (e.g., two-wheeled robots) [1], and aerospace systems. Initially, this famous system consists of a pendulum attached to a cart that can move horizontally, allowing the pendulum to balance at its unstable equilibrium point by controlling the cart's motion. Due to its unstable nature and non-linearity, the inverted pendulum on a cart has become a popular test for evaluating the performance of advanced control algorithms. An accurate mathematical model for the mentioned system is crucial, as it provides the foundation of our designed controller, which is set to stabilize the pendulum and reach the desired reference. The significance and history of this system lie beyond academic research, with applications in robotics, aerospace, and industrial automation, where similar dynamics are encountered.

As stated earlier, numerous studies and articles have explored various control strategies and algorithms to stabilize the inverted pendulum on a cart system. In terms of traditional methods, the PID (Proportional-Integral-Derivative) controller is the first candidate that comes to mind. Alongside PID, the LQR controller is frequently used due to its simplicity and effectiveness in linearized systems [2-4]. For instance, [5] demonstrated the robustness of LQR in stabilizing the system under small disturbances. Meanwhile, the MPC controller has consistently proven helpful due to its ability to handle constraints and optimize control actions, as mentioned in [6-10]. As intelligent controllers have developed over recent years, the RL has shown remarkable assistance in controlling many complex nonlinear systems, making it the center of attention [11-14]. However, many RL-based studies have primarily focused on the Rotary Inverted Pendulum system [15-17] or Double Inverted Pendulum on a Cart [18], leaving a gap in the application of RL to the cart-mounted inverted pendulum system.

In this paper, we intend to fill this gap by designing controllers and comparing each output for the inverted pendulum on a cart while having inclusive linear and rotational dampers on both the cart and the pendulum itself, which helps make the system more realistic, using three distinctive control approaches: LQR, MPC, and RL. The inclusion of translational damping on the cart and torsional damping at the pivot in the inverted pendulum system is fundamentally motivated by the necessity to emulate real-world energy dissipation mechanisms, thereby enhancing model fidelity. Translational damping accounts for frictional losses between the cart and track, while torsional damping captures energy dissipation at the pivot joint due to material deformation and air resistance. From a dynamic's perspective, these damping terms stabilize high-frequency oscillations, improve control realism and enhance disturbance rejection by providing energy dissipation. Considering that the dynamic model remains the same, we expect the output for each of these controllers to be sufficiently similar. To further enhances robustness in nonlinear operating conditions, we extend our analysis by introducing SMC and implementing a hybrid switching strategy between SMC and RL control. This emphasizes transient performance and disturbance rejection in highly nonlinear conditions. Also, by incorporating stochastic sampling techniques, Monte Carlo evaluates a wide range of possible future scenarios, making it particularly effective in handling data sensitivity and variability. Furthermore, we provide a comprehensive comparison of the three methods, evaluating their performance in terms of stability, convergence, and robustness to disturbances.

The LQR controller has been extensively studied as a robust control strategy for inverted pendulum systems, offering optimal performance under well-defined dynamics. Rani and Kamlu [19] demonstrated the superiority of LQG controllers combined with Unscented Kalman Filters (UKF) over conventional PID and MPC methods, particularly in disturbance rejection. The integration of IoT for real-time LQR control, as proposed by the Mechatronics Polytechnic of Sanata Dharma [20], highlights the method's adaptability to modern technological frameworks. For hardware-aligned applications, an unnamed study designed an LQR controller accounting for mechanical transmission and DC motor dynamics, ensuring feasibility for physical implementations. Further, Erkol [21] advanced LQR tuning via the Grey Wolf Optimizer, automating weight matrix selection to enhance stabilization. Cheng et al. [22] reinforced these findings by systematizing LQR design principles for pendulum control. Collectively, these works validate LQR's versatility in balancing precision and robustness, while underscoring ongoing innovations in optimization and real-world deployment.

Recent advances in RL have demonstrated significant potential for controlling inverted pendulum systems, addressing key challenges such as model uncertainty, disturbance rejection, and adaptive stabilization. Yildiran [23] proposed an innovative hybrid approach that combines RL with LQR to achieve adaptive control without relying on precise system modeling, thereby showcasing robustness to parametric variations. Safeea and Neto [24] explored discrete-action Q-learning for continuous pendulum control, emphasizing the critical role of accurate simulation-to-reality transfer in RL training. For complex nonlinear dynamics, Hill [25] employed Deep Deterministic Policy Gradient (DDPG) to reject stochastic disturbances in a cart-pole system with uncertain parameters, outperforming classical methods. Liu et al. [26] extended RL applications to both swing-up and balance control using DDPG, validating their framework through simulations and hardware experiments. Complementing these technical studies, an educational framework illustrated RL-based pendulum stabilization using simplified state representations and discrete actions, highlighting pedagogical value. Collectively, these works highlight RL's versatility in handling the inverted pendulum's nonlinearity and under actuation, while also revealing open challenges in real-world deployment. Similarly, Bajelani [27] proposed an inner-outer loop control architecture for unstable systems, implementing it on both multicopters and inverted pendulums.

Furthermore, RL has recently been applied in flight control tasks to handle actuator faults and system uncertainties without relying on precise system models. For instance, the [28] employs a Q-learning-based RL approach to design a fault-tolerant flight controller for a passenger aircraft experiencing elevator actuator degradation. By modeling the system with variable actuator gains and time delays, the authors demonstrate that the Q-learning algorithm can adaptively learn optimal control policies through interaction, achieving stable performance under both nominal and faulty conditions. This model-free strategy, similar in spirit to the RL-based control we investigate for the inverted pendulum-cart system, shows strong resilience to actuator anomalies and highlights the potential of reinforcement learning in real-world control applications. These developments align with broader trends in robust and adaptive control for underactuated systems. For instance, Asrari et al. [29] developed a robust state-feedback controller for Linear Parameter-Varying (LPV) systems, validating their method on a cart-pendulum system with parametric variations. In a more complex application, Soltani and Kamari [30] investigated a hybrid position and force control strategy for a spherical inverted pendulum mounted on a quadrotor undergoing constrained motion, demonstrating advanced stabilization techniques for coupled aerial-manipulator systems.

The key contributions of this work are threefold. First, this study introduces a novel dynamic modeling framework for the inverted pendulum-cart system by incorporating both torsional spring-damper elements and cart translational damping, representing the first comprehensive derivation of dynamic equations under these combined realistic physical constraints. Second, we systematically design and compare multiple control algorithms (LQR, MPC, and RL) under small-angle initial conditions, subsequently applying them to

the full nonlinear dynamics of the system. Third, building on the performance analysis of these controllers, we develop an innovative hybrid SMC-RL control architecture that automatically switches between control strategies, by employing SMC during large angle deviations and severe nonlinear regimes, while transitioning to RL control for stabilization under small-angle conditions. This hybrid approach represents a significant advancement in adaptive control for underactuated systems. Through extensive simulations, we demonstrate that the proposed hybrid controller achieves superior performance compared to individual methods, particularly in handling uncertainties and strong nonlinearities. This study not only advances the theoretical understanding of hybrid control strategies but also provides a practical framework for implementing adaptive control in real-world applications with varying operational conditions.

Following this introduction, the remainder of this paper is organized to present a comprehensive study of inverted pendulum control. Section 2 develops the complete mathematical modeling of the inverted pendulum system, deriving both nonlinear equations of motion and their linearized approximations around the unstable equilibrium point. Section 3 systematically details the design and implementation of three distinct control approaches: the classic LQR, MPC, and RL algorithms, discussing their theoretical foundations and practical considerations. Moreover, a hybrid SMC strategy with RL switching for enhanced nonlinear robustness is being added. Section 4 presents extensive simulation results comparing the performance of these controllers in terms of stabilization capability, disturbance rejection, and robustness to parameter variations. Section 5 presents a critical discussion of the findings, highlighting the relative advantages and limitations of each control strategy, as well as their computational requirements. Finally, the paper concludes by identifying promising directions for future research, including hardware implementation challenges and potential hybrid control architectures.

## 2- Mathematical Modeling

This section establishes the dynamic model of the enhanced inverted pendulum on cart system using the Euler-Lagrange formulation. The derived equations of motion form the foundational basis for the subsequent design and analysis of the proposed control strategies.

### 2-1- Euler-Lagrange Equations

Consider a cart mass $M$ that moves along a horizontal axis $x$. A pendulum of mass $m$ and length $l$ is pivoted to the cart. Both the cart and the pendulum have a damper.

Assuming that the cart moves without slipping on the surface, the pendulum is a rigid rod with its mass concentrated at the center of mass; A damper with a damping coefficient $b_c$ providing a resistive force proportional to the cart's velocity $\dot{x}$ and a rotational damper with a damping coefficient $b_p$ providing a resistive torque proportional to the pendulum's angular velocity $\dot{\theta}$. Using the Euler-Lagrange equation:

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{x}_i}\right)-\frac{\partial T}{\partial x_i}+\frac{\partial P_E}{\partial x_i}=\sum F \qquad (1)$$

where $T$ and $P_E$ are corresponding kinetic and potential energy of the system. Also $x_i$ represents the generalized coordinates ($x$ and $\theta$ in this case). Equations for both kinetic and potential energies are derived as follows:

$$T=\frac{1}{2}M\dot{x}^2+m(2\dot{x}^2+\frac{7}{8}l^2\dot{\theta}^2-\frac{5}{2}\dot{x}l\dot{\theta}\cos\theta) \qquad (2)$$

$$\mathrm{P_E}=\frac{1}{2}k\theta^2+\frac{5}{2}mgl\cos\theta \qquad (3)$$

By applying the Euler-Lagrange equation, two coupled nonlinear differential equations that describe the dynamics of the system can be obtained:

$$(M+4m)\ddot{x}-\frac{5}{2}ml\ddot{\theta}\cos\theta+\frac{5}{2}ml\dot{\theta}sin\theta=u-b_c\dot{x} \qquad (4)$$

$$\frac{7}{4}ml^2\ddot{\theta}-\frac{5}{2}ml\ddot{x}\cos\theta+k\theta-\frac{5}{2}mgl\sin\theta=-b_p\dot{\theta} \qquad (5)$$

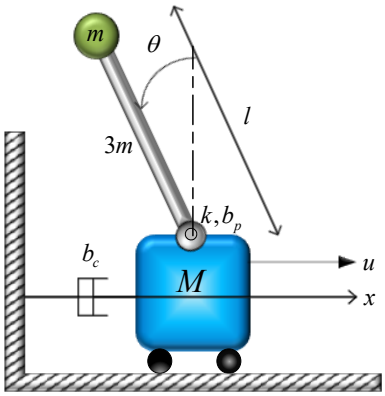


**Fig.1)** Schematics of Inverted Pendulum on cart

### 2-2- Linearization and State-Space Representation

As mentioned, Eq. (4), (5) These are nonlinear differential equations of the model. Since the inverted pendulum is inherently unstable around the upright position ($\theta=0$), we linearize the equations of motion around this equilibrium point. Assuming small angles ($\theta\sim0$), we approximate $sin\,\theta\sim\theta$ and $cos\,\theta\sim1$. Therefore, the linearized equations will be the following:

$$(M+4m)\ddot{x}-\frac{5}{2}ml\ddot{\theta}=u-b_c\dot{x} \qquad (6)$$

$$\frac{7}{4}ml^2\ddot{\theta}-\frac{5}{2}ml\ddot{x}+k\theta-\frac{5}{2}mgl\theta=-b_p\dot{\theta} \qquad (7)$$

The linear equations can be expressed in the form of a matrix system of equations as follows:

$$\dot{X}=AX+BU \qquad Y=CX+DU \qquad (8)$$

where $X=[x\quad \dot{x}\quad \theta\quad \dot{\theta}]^T$ is the state vector, U is the control input vector, also A, B, C and D are matrices derived from the linearized dynamics. Substituting the values from the table into the equation, the corresponding matrices could be shown as:

**Table 1.** Identified parameters of the Inverted Pendulum on Cart System

| Parameter | Value | Unit |
|---|---|---|
| $M$ | 5 | kg |
| $m$ | 1 | kg |
| $l$ | 1 | m |
| $k$ | 10 | N/m |
| $b_c$ | 5 | N.s/m |
| $b_p$ | 5 | N.s/rad |

$$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{\theta}\\ \ddot{\theta}\end{bmatrix}=\begin{bmatrix}0 & 1 & 0 & 0\\ 0 & 0.192 & -0.412 & 0.274\\ 0 & 0 & 0 & 1\\ 0 & 2.691 & 1.482 & 0.988\end{bmatrix}\begin{bmatrix}x\\ \dot{x}\\ \theta\\ \dot{\theta}\end{bmatrix}+\begin{bmatrix}0\\ -0.04\\ 0\\ -0.54\end{bmatrix}u \qquad (9)$$

$$y=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x\\ \dot{x}\\ \theta\\ \dot{\theta}\end{bmatrix} \qquad (10)$$

Finally in order to enhance the physical realism of our dynamic system modeling, an accurate electromechanical actuator model is incorporated. This implementation provides crucial fidelity in simulating the torque generation and transient behavior of the cart's actuation system. This approach bridges the gap between idealized control signals and realistic actuator behavior in our inverted pendulum simulations.

## 3- Control Algorithms

After the introductory explanations and system modeling, we proceed to its control. In this paper, LQR, MPC, and RL algorithms followed

by hybrid SMC-RL controller for nonlinear conditions are employed to design a controller for the system.

**3-1- LQR**

LQR is a commonly used optimal control algorithm for linear systems. Its primary advantage is to minimize the quadratic cost function that balances the trade-off between state deviations and control effort. Note that this cost function consists of two integral terms: the first term accounts for the system output error, and the other is associated with the energy of the control signals. The mathematical formulation of LQR control is expressed as follows:

$$J = \int_0^\infty ((X - X_d)^T Q (X - X_d) + U^T R U) dt \tag{11}$$

Where Q and R are the weight matrices for state and control errors, respectively, the LQR control system is optimized by solving the Riccati equations as follows:

$$A^T P + PA - PBR^{-1}B^T P + Q = 0 \tag{12}$$

Where $P$ is the symmetric matrix of the optimizer, $R^{-1}$ represents the inverse matrix of measurements (control weight). The control matrix $K$ is as follows:

$$K = R^{-1}B^T P \tag{13}$$

Using $K$, the optimal control can be calculated as follows:

$$U = -K(X - X_d) \tag{14}$$

These optimal control inputs can provide appropriate state feedback for the dynamic system states.

**3-2- MPC**

Model Predictive Control is one of the advanced methods of process control that has gained widespread popularity for its ability to handle complex, multivariable systems with constraints. Despite its computational complexity, advancements in optimization algorithms and computing power have made MPC a practical and powerful tool for modern control challenges. Unlike traditional control methods, MPC calculates the optimal signal corresponding to each input by considering the constraints. It predicts the future by making use of the input and output values of the preceding state over a specified time span. At each control interval, an optimization problem is solved to determine the sequence of control actions that minimizes a predefined cost function while satisfying system constraints. The discrete-time state-space system is presented below: [31]

$$\begin{aligned} X(k+1) &= AX(k) + BU(k) \\ Y(k) &= CX(k) + DU(k) \end{aligned} \tag{15}$$

In which $X \in R^n$ is our system state vector, $U \in R^m$ is the calculated control input vector, $Y \in R^p$ is the output vector, $A \in R^{n\times n}$ is the state matrix, $B \in R^{n\times m}$ is the input matrix, $C \in R^{p\times n}$ is the output matrix, and $k$ denotes the sampling number.

$$\begin{aligned} &X_{n\times 1}(k+1|k) = A_{n\times n}X_{n\times 1}(k|k) + B_{n\times m}U_{m\times 1}(k|k) \\ &X(k+2|k) = A^2X(k|k) + ABU(k|k) + BU(k+1|k) \\ &\vdots \\ &X(k+N_p|k) = A^{N_p}X(k|k) + A^{N_p-1}BU(k|k) \\ &\qquad + \cdots A^{N_p-N_c}BU(k+N_c-1|k) \\ &X_{nN_p\times 1} = F_{nN_p\times n}X_{n\times 1}(k|k) + \Phi_{nN_p\times mN_c}U_{mN_c\times 1} \end{aligned} \tag{16}$$

Where $X$ and $U$ are predicted state and predicted input vectors:

$$X = \begin{bmatrix} X(k+1|k) \\ X(k+2|k) \\ \vdots \\ X(k+N_p|k) \end{bmatrix}, U = \begin{bmatrix} U(k+1|k) \\ U(k+2|k) \\ \vdots \\ U(k+N_p|k) \end{bmatrix} \tag{17}$$

And F and $\Phi$ are defined as follows:

$$F = \begin{bmatrix} A \\ A^2 \\ \vdots \\ A^{N_p-1} \\ A^{N_p} \end{bmatrix}_{nN_p\times n}, \tag{18}$$

$$\Phi = \begin{bmatrix} B & \bar{0} & \dots & \bar{0} \\ AB & B & \dots & \vdots \\ \vdots & \vdots & \ddots & \bar{0} \\ A^{N_p-1}B & A^{N_p-2}B & \dots & A^{N_p-N_c}B \end{bmatrix}_{nN_p\times mN_c}$$

Prediction input-output is written as:

$$\begin{aligned} Y(k+1|k) &= CX(k+1|k) \\ &\vdots \\ Y(k+N_p|k) &= CA^{N_p}.X(k|k) + CA^{N_p-1}BU(k|k) \\ &\quad + \cdots + CA^{N_p-N_c}BU(k+N_c-1|k) \\ Y_{pN_p\times 1} &= \bar{F}_{pN_p\times n}X(k|k)_{n\times 1} + \bar{\Phi}_{pN_p\times mN_c}U \end{aligned} \tag{19}$$

where $\bar{F}$ and $\bar{\Phi}$ are defined as follows:

$$\begin{aligned} \bar{F} &= \begin{bmatrix} CA \\ CA^2 \\ \vdots \\ CA^{N_p} \end{bmatrix}_{pN_p\times n}, \\ \bar{\Phi} &= \begin{bmatrix} CB & \bar{0} & \dots & \bar{0} \\ CAB & B & \dots & \vdots \\ \vdots & \vdots & \ddots & \bar{0} \\ CA^{N_p-1}B & CA^{N_p-2}B & \dots & CA^{N_p-N_c}B \end{bmatrix}_{pN_p\times mN_c} \end{aligned} \tag{20}$$

The quadratic cost function is defined as:

$$\begin{aligned} J \triangleq\ & (X - X_d)^T(k+N_p|k)P(X - X_d)(k+N_p|k) \\ &+ \sum_{i=1}^{N_p-1} (X - X_d)^T(k+i|k)Q(X - X_d)(k+i|k) \\ &+ \sum_{j=1}^{N_c} U^T(k+j-1|k)RU(k+j-1|k) \end{aligned} \tag{21}$$

Another representation of the above function is:

$$\begin{aligned} J = \Big[ & \frac{1}{2}U^T HU + (X - X_d)^T(k|k)MU \\ & + \frac{1}{2}(X - X_d)^T(k|k)N(X \\ & - X_d)(k|k) \Big] \end{aligned} \tag{22}$$

where $H = 2(\Phi^T\bar{Q}\Phi + \bar{R})$, $M = 2F^T\bar{Q}F$ and $N = 2F^T\bar{Q}\Phi$. The cost function is minimized in the control process.

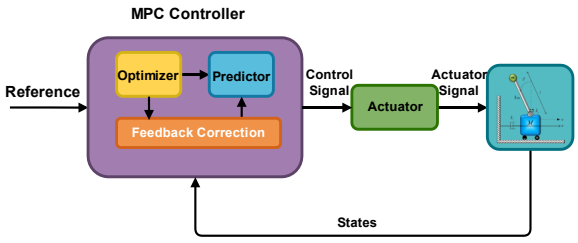


**Fig.2)** Schematics of MPC control algorithm

**3-3- RL**

RL is a cutting-edge approach in control engineering, where an agent interacts with a dynamic system and periodically receives rewards. In contrast to most classical control methods, RL does not require a precise mathematical model of the system. Instead, it relies on trial and error, where a software agent makes observations and takes actions within an environment, and in return, it receives rewards. The main objective in reinforcement learning is to maximize its expected long-term rewards.

In controlling dynamic systems, RL offers significant advantages, such as adaptability to changing environments and the ability to handle high-dimensional state and action spaces.

The dynamics of the system are characterized by a 4-state linear system articulated as follows: [28]

$$X_{k+1} = AX_k + BU_k \tag{23}$$

where $X \in \mathbb{R}^4$ is the state vector that includes linear position/velocities, angular position/velocities. Also $U \in \mathbb{R}^1$ is the control input vector. $A \in \mathbb{R}^{4\times 4}$ is the system matrix and $B \in \mathbb{R}^{4\times 1}$ is

the control input matrix. The optimal control policy in a model-free reinforcement learning approach is based on observed samples of outcomes to maximize expected rewards, which are then employed in Q-learning. The main goal of Q-learning is achieving a policy, denoted as π(s), which aims to establish a mapping from states $s$ to optimal actions a with targeting the maximum expected cumulative reward in the future.

$$J(\pi) = E(R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \cdots) \quad (24)$$

Here, $R_t$ represents the reward at time $t$, and $\gamma$ is the discount factor accounting for future rewards. The optimal Q function $Q^*(s,a)$ adheres to the Bellman equation:

$$Q^*(s,a) = \mathbb{E}[R_t + \gamma max_{a'} Q^*(s',a')] \quad (25)$$

where s′ denotes the next state after action a is taken in state s. The Q-learning algorithm initializes $Q(s_t, a_t)$values arbitrarily and refines these estimates through iterative updates:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[R_t + \gamma max_a Q(s_{t+1},a_t) - Q(s_t,a_t)] \quad (26)$$

Here, $a$ is the learning rate, and experiences in the form of ($s$, $a$, $R$, $s'$) are sampled to update $Q$.

### 3-4- SMC

SMC is a robust nonlinear control strategy designed to cope with system uncertainties and external disturbances. The fundamental principle relies on the definition of a sliding manifold $s(x,t) = 0$, designed such that the system exhibits reduced-order stable dynamics once the states reach this manifold. For a control-affine system of the form:

$$\dot{x}(t) = f(x,t) + g(x,t)u(t) + d(t) \quad (27)$$

with state $x \in \mathbb{R}^4$, input $u \in \mathbb{R}^1$, and bounded disturbance $d(t)$, the sliding variable is typically expressed as a linear combination of the states, for instance:

$$s(x,t) = c^T x(t); \quad c \in \mathbb{R}^4 \quad (28)$$

The derivative of the sliding variable is:

$$\dot{s} = \frac{\partial s}{\partial x}(f(x,t) + g(x,t)u(t) + d(t)) + \frac{\partial s}{\partial t} \quad (29)$$

To maintain trajectories on the manifold, an equivalent control is defined as:

$$u_{eq} = -\left(\frac{\partial s}{\partial x} g(x,t)\right)^{-1} \left[\frac{\partial s}{\partial x} f(x,t) + \frac{\partial s}{\partial t}\right] \quad (30)$$

However, in the presence of disturbances, the complete law becomes:

$$u = u_{eq} - K\, sign(s) \quad (31)$$

where $K > 0$ is selected such that $K > \Delta$, with Δ donating the upper bound on the disturbance term projected along the sliding surface. The stability of the sliding motion can be verified using the Lyapunov function candidate:

$$V(s) = \frac{1}{2}s^2, \quad \dot{V} = s\dot{s} \quad (32)$$

Substituting the control law yields $\dot{V} \leq -(K-\Delta)|s| < 0$, ensuring finite-time convergence of $s$ to be zero. Once on the manifold, the reduced-order dynamics are governed exclusively by the equivalent control, guaranteeing robustness and trajectory invariance to uncertainties.

In practice, the discontinuous nature of the $sign$(s) function induces high-frequency oscillations, known as chattering, which can excite unmodeled dynamics and degrade performance. To mitigate this effect, boundary-layer methods employ a continuous saturation function, while advanced strategies such as Higher-Order SMC (e.g., the super-twisting algorithm) generate continuous control signals that preserve robustness. These improvements extend the applicability of SMC to domains such as robotics, power converters, process control, and aerospace systems, where both precision and robustness are critical. The Fig.3 shows the control schematics as below:

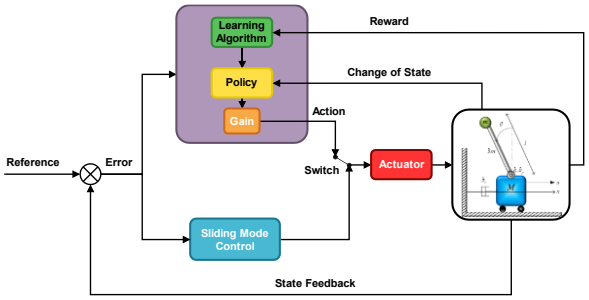


**Fig.3)** Schematics of SMC-RL control algorithm

## 4- Results and Simulations

In this section, we present the results obtained by implementing our designed control algorithms on the inverted pendulum system mounted on a cart, as reported through simulations using MATLAB/Simulink. The effectiveness of the proposed control strategies, as mentioned in the previous section, will be thoroughly assessed in terms of stabilizing the system, specifically LQR, MPC, and RL.

The simulations were carried out under specific initial conditions. For example, the pendulum was set at small angles close to the vertical position to evaluate stability in the near-equilibrium point. More to mention, different sorts of initial conditions, such as the cart's initial position and velocity, besides the pendulum's initial angle and angular velocity. The diversity in initial conditions, aside from assuring the effectiveness of our designed controllers, helped to analyze the system's behavior and response in multiple variations of initial conditions.

Applying the mentioned control algorithm to the plant gave us fascinating outputs. Despite having three controllers, each having completely different insights, the results seemed relatively too close to differ. Although LQR achieved a great deal in terms of stability and rapid convergence in regulating the output, while providing optimum control gains, it faced difficulties in terms of external disturbances and meeting the constraints. On the one hand, MPC demonstrated excellent performance in constrained environments, proving its effectiveness in predicting future states and related control actions. However, on the other hand, its complex mathematical computations require higher processor memory limits, making it less suitable for real-time applications. Lastly, it is worth mentioning that RL requires a considerable amount of time and training to achieve an adequate performance level.

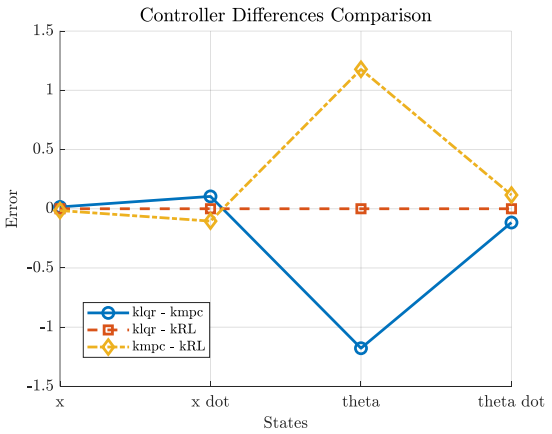


**Fig.4)** Comparison of different algorithms outputs

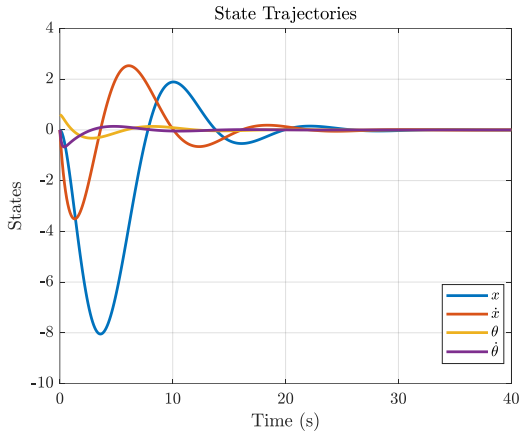


**Fig.5)** Linear state results in sample of time

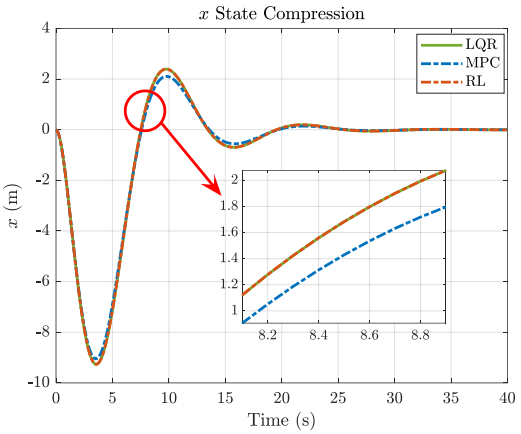


**Fig.7)** $x$ State Compression in sample of time

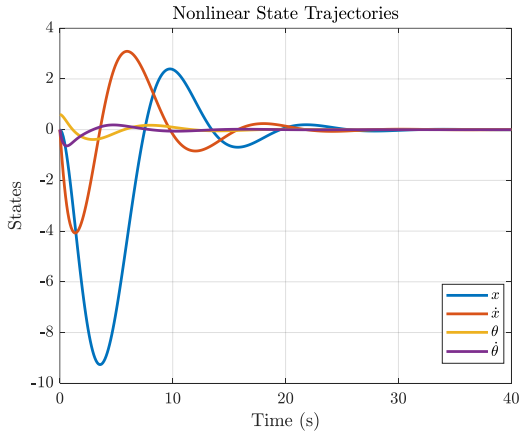


**Fig.6)** Nonlinear state results in sample of time

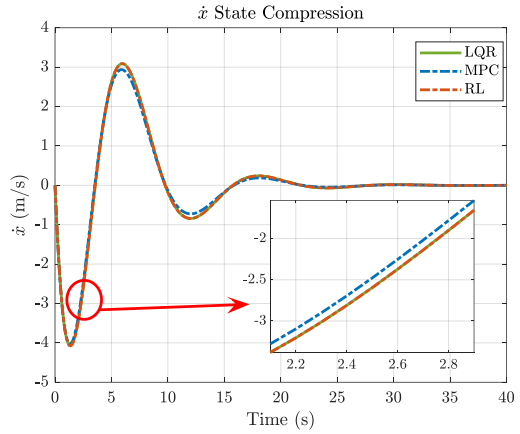


**Fig.8)** $\dot{x}$ State Compression in sample of time

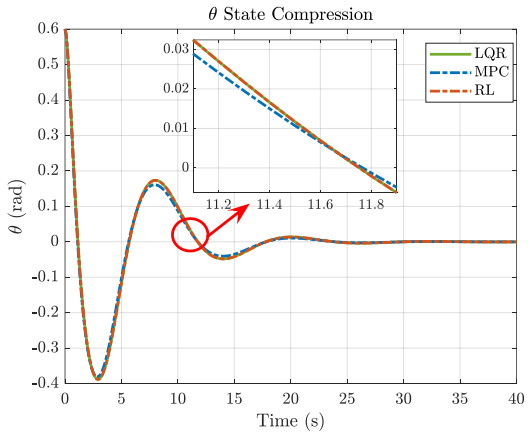


**Fig.9)** $\theta$ State Compression in sample of time

As illustrated in Fig.4, the slight difference between the control gain algorithms shows that both MPC and RL are close to optimal control gains. Also Fig.5 and Fig.6 clearly show that all three control strategies demonstrated satisfactory performance in stabilizing both linear and nonlinear dynamic equations of the inverted pendulum on a cart system. Upon closer examination, one can detect the differences in the response's behavior. The MPC illustrated a slight overshoot in its responses, which can be attributed to its defined weight matrices and reliance on the input control signals. While this overshoot does not significantly affect the plant's performance, it helps to detect the system's limitation in achieving stability, especially in the presence of external disturbances or dynamic uncertainties.

In contrast, the RL achieved results that were notably more satisfactory. Considering that RL can adapt the system's dynamics through learning, rather than relying on fixed models or assumptions, its superiority in control and performance becomes readily apparent. These findings emphasize the advantages of RL in handling the plant's dynamic systems, while also highlighting the trade-offs associated with the other control methods mentioned, LQR and MPC.

Fig.7 through Fig.10 demonstrate that all four states are simulated with a minor difference, which is extremely small. This makes it clear that the three control algorithms have achieved the same results using three different methods. To demonstrate how minor the state differences are, Fig.11 is illustrated, which explains that all states are close enough to be considered identical. Furthermore, it is notable that although the differences can be regarded as negligible, at the extremum points, the error reaches its peak. Control signal output is shown in Fig.12 as well.

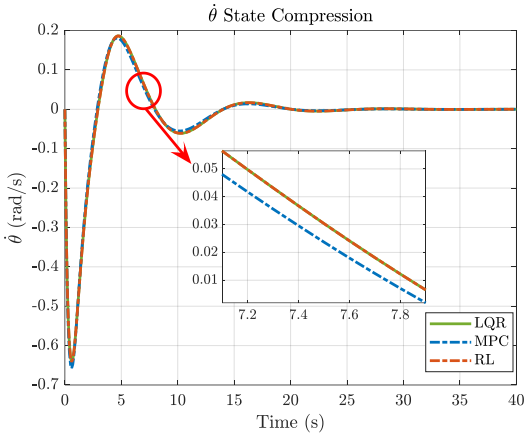


**Fig.10)** $\dot{\theta}$ State Compression in sample of time

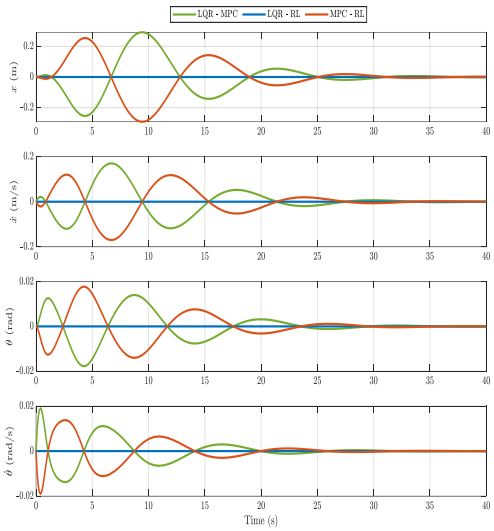


**Fig.11)** State Error in sample of time

**Table 2.** States comparison under different control strategies

| Parameter | Strategy | Rise time (s) | Settling time (s) | Max | Min |
|---|---|---|---|---|---|
| Cart Position | LQR | 7.969 | 19.427 | 2.3940 | -9.2649 |
| | MPC | 8.096 | 19.558 | 2.1076 | -9.0489 |
| | RL | 7.969 | 19.427 | 2.3949 | -9.2666 |
| Cart Velocity | LQR | 3.951 | 19.427 | 3.0895 | -4.0769 |
| | MPC | 3.928 | 19.427 | 2.9391 | -4.0315 |
| | RL | 3.951 | 19.427 | 3.0895 | -4.0769 |
| Pendulum Angle | LQR | 0.733 | 19.427 | 0.6000 | -0.3882 |
| | MPC | 0.716 | 19.427 | 0.6000 | -0.3814 |
| | RL | 0.733 | 19.427 | 0.6000 | -0.3882 |
| Pendulum Angular Velocity | LQR | 0.343 | 19.427 | 0.1864 | -0.6414 |
| | MPC | 0.353 | 19.427 | 0.1812 | -0.6571 |
| | RL | 0.343 | 19.427 | 0.1864 | -0.6414 |

Fig.13 depicts the temporal activation of RL and SMC within the hybrid scheme. A value of 0 corresponds to RL dominance, while 1 indicates SMC operation. After disturbing the system from its equilibrium point, initially the controller relies entirely on SMC to rapidly stabilize the system, and after approximately 6 seconds, the blending factor transits to RL, reflecting a handover to a more robust mode. Fig.14 and Fig.15 demonstrate states and control signal outputs under the hybrid controller. Primary transitions induced by SMC generate bounded oscillations in the states, especially in the pendulum angle and angular velocity. Following the mode switch to RL, the states converge smoothly toward zero with noticeably reduced overshoot, confirming the effectiveness of the hybrid strategy. The hybrid controller efforts to stabilize the system under input saturation constraints ( $|u| \leq 50\ N$). At first an initial high-amplitude input is applied to counteract large deviations, followed by a markedly smoother control signal, indicating successful chattering mitigation and stable closed-loop behavior.

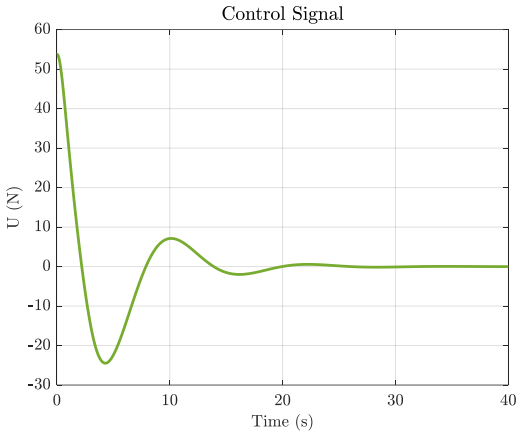


**Fig.12)** Control Signal output in sample of time

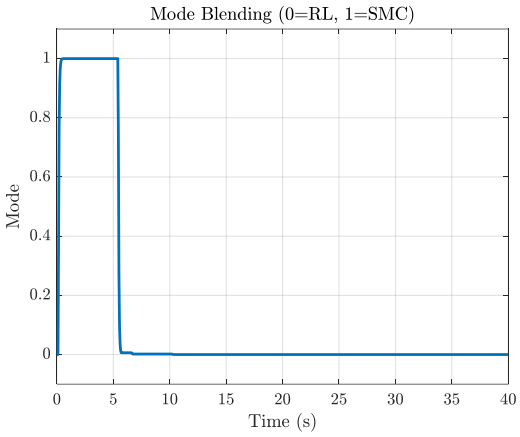


**Fig.13)** Mode blending between RL and SMC in sample of time

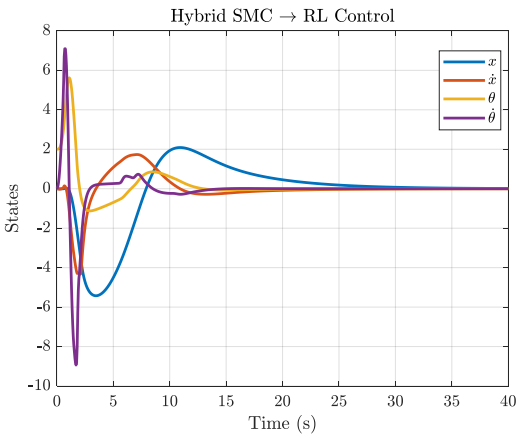


**Fig.14)** Hybrid states output in sample of time

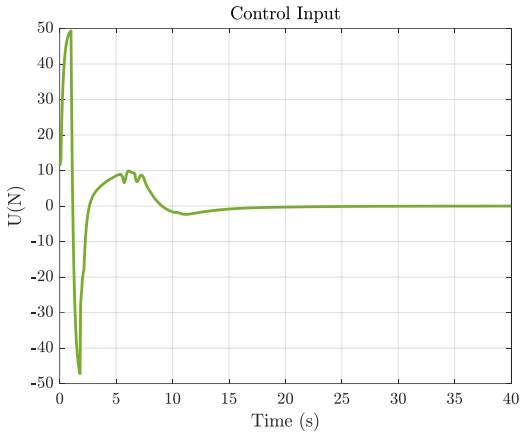


**Fig.15)** Hybrid control signal output in sample of time

**Table 3.** SMC-RL control strategy performance

| Parameter | Rise time (s) | Settling time (s) | Max | Min |
|---|---|---|---|---|
| Cart Position | 4.184 | 37.009 | 2.0875 | -5.4244 |
| Cart Velocity | 4.210 | 10.350 | 1.7330 | -4.3292 |
| Pendulum Angle | 0.650 | 12.700 | 5.6250 | -1.1244 |
| Pendulum Angular Velocity | 0.187 | 14.200 | 7.0959 | -8.9384 |

Using the Monte Carlo method, extensive training and evaluations were performed by 80 simulations, each with different initial conditions from one another. These simulations were conducted after the RL agent had been fully trained and had learned the control policy. The output data from the simulations were evaluated by their mean and standard deviation. As it is evident in the Fig.16 and Fig.17 Monte Carlo simulations demonstrate that the mean performance remained consistent throughout these 80 simulations, with minimal variation, illustrating that the system's behavior can be considered a stable and reliable response under a diverse range of initial conditions. Additionally, the low standard deviation further confirms that the control strategy was not sensitive to changes in the system's initial state. The effectiveness of this method in handling robustness and uncertainties makes this algorithm a suitable choice for implementing this controller on the plant, where the initial conditions have no specific constraints to be controlled.

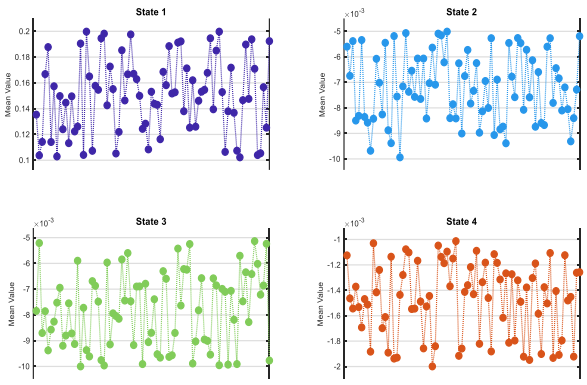


**Fig.16)** State's Mean Value after RL training

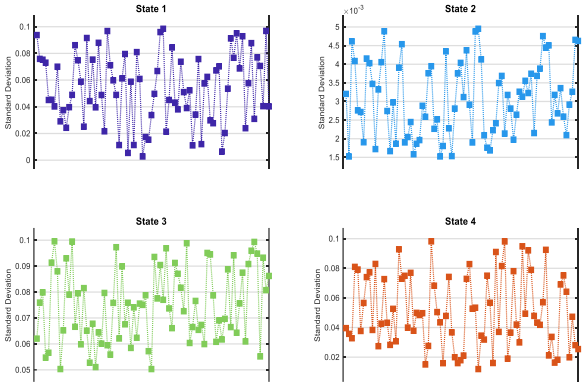


**Fig.17)** State's standard deviation after RL training

Eventually to ensure the validation of the stability properties of the controlled system, a Lyapunov-based stability analysis was conducted. The Lyapunov equations used in the stability proof are provided in Equation (32). The results confirm that the closed-loop system satisfies the necessary conditions for asymptotic stability under the proposed control scheme. This theoretical assurance is strongly supported by the empirical evidence from Monte Carlo simulations. For data-driven learning algorithms like RL, numerical verification through extensive simulations is the most practical approach to demonstrate stability. The conjunction of Lyapunov theory and statistical Monte Carlo evidence provides a comprehensive verification of system robustness and reliable performance. The conjunction of Lyapunov theory and statistical Monte Carlo evidence provides a comprehensive verification of system robustness and reliable performance.

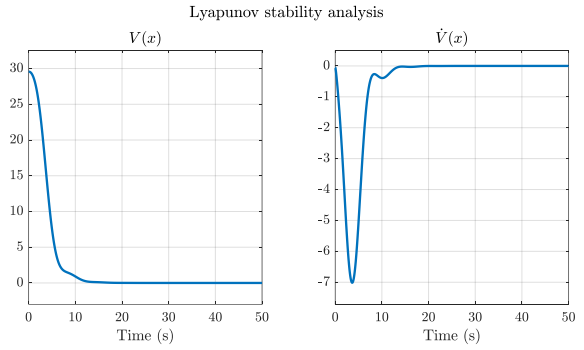


**Fig.18)** Lyapunov stability analysis

## 5- Conclusion

This work investigated LQR, MPC, and RL controllers on an enhanced inverted pendulum system with torsional spring–damper and cart damping. Simulations showed that although these controllers differ in formulation, their performance in the linear region was similar, with settling times difference of about 0.7%, maximum overshoot variation 1.6% for cart position and 2.3% pendulum angle, and steady-state angle errors under 0.5°. However, under nonlinear conditions and large initial angles, LQR, MPC and RL exhibited an inability to control and sustain accuracy while maintaining stability only up to initial angles of about 45-50°.

To overcome these limitations, a hybrid SMC-RL controller was developed, achieving significantly improved robustness across nonlinear regimes. The proposed approach successfully stabilized the pendulum from initial large angles up to 120°. Future work will focus on hardware-in-the-loop validation and extending the method to more complex underactuated systems.

**Ethics Approval**

The scientific content of this article is the result of the authors' research and has not been published in any Iranian or international journal.

**Conflict of Interest**

The authors declared that they have no conflicts of interests to this work.

## References


[1] M. Tayefi and Z. Geng, "Self-balancing controlled Lagrangian and geometric control of unmanned mobile robots," *Journal of Intelligent & Robotic Systems,* vol. 90, no. 1, pp. 253-265, 2018, doi: 10.1007/s10846-017-0666-7.

[2] M. Fauziyah, Z. Amalia, I. Siradjuddin, D. Dewatama, R. P. Wicaksono, and E. Yudaningtyas, "Linear quadratic regulator and pole placement for stabilizing a cart inverted pendulum system," *Bulletin of electrical engineering and informatics,* vol. 9, no. 3, pp. 914-923, 2020, doi: 10.11591/eei.v9i3.2017.

[3] I. Siradjuddin, B. Setiawan, A. Fahmi, Z. Amalia, and E. Rohadi, "State space control using LQR method for a cart-inverted pendulum linearised model," *International Journal of Mechanical and Mechatronics Engineering,* vol. 17, no. 1, pp. 119-126, 2017.

[4] S. J. Chacko and R. J. Abraham, "On LQR controller design for an inverted pendulum stabilization," *International Journal of Dynamics and Control,* vol.11,no.4,pp.1584-1592,2023, doi: 10.1007/s40435-022-01079-0.

[5] S. Y. Kim, C. H. Kang, and C. G. Park, "Multiple Frequency Tracking and Mitigation Based on RSPWVD and Adaptive Multiple Linear Kalman Notch Filter," *International Journal of Control, Automation and Systems,* vol. 18, no. 5, pp. 1139-1149, 2020/05/01 2020, doi: 10.1007/s12555-019-0123-4.

[6] E. S. Varghese, A. K. Vincent, and V. Bagyaveereswaran, "Optimal control of inverted pendulum system using PID controller, LQR and MPC," in *IOP Conference Series: Materials Science and Engineering*, 2017, vol. 263, no. 5: IOP Publishing,p.052007, doi: 10.1088/1757-899X/263/5/052007

[7] R. Balan, V. Maties, O. Hancu, and S. Stan, "A predictive control approach for the inverse pendulum on a cart problem," in *IEEE International Conference Mechatronics and Automation, 2005*, 2005, vol. 4: IEEE, pp. 2026-2031, doi: 10.1109/ICMA.2005.1626874.

[8] L. Messikh, E.-H. Guechi, and S. Blažič, "Stabilization of the cart–inverted-pendulum system using state-feedback pole-independent MPC controllers," *Sensors,* vol. 22, no. 1, p. 243, 2021, doi: https://doi.org/10.3390/s22010243.

[9] R. Balan, V. Maties, and S. Stan, "A solution of the inverse pendulum on a cart problem using predictive control," in *Proceedings of the IEEE International Symposium on Industrial Electronics, 2005. ISIE 2005.*, 2005, vol. 1: IEEE, pp. 63-68, doi: 10.1109/ISIE.2005.1528889.

[10] S. Nakatani and H. Date, "Swing up control of inverted pendulum on a cart with collision by monte carlo model predictive control," in *2019 58th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, 2019:IEEE,pp.1050-1055, doi: 10.23919/SICE.2019.8859912.

[11] D. S. Deighan, S. E. Field, C. D. Capano, and G. Khanna, "Genetic-algorithm-optimized neural networks for gravitational wave classification," *Neural Computing and Applications,* vol. 33, no. 20, pp. 13859-13883, 2021/10/01 2021, doi: 10.1007/s00521-021-06024-4.

[12] C. A. Manrique Escobar, C. M. Pappalardo, and D. Guida, "A parametric study of a deep reinforcement learning control system applied to the swing-up problem of the cart-pole," *Applied Sciences,* vol. 10, no. 24, p. 9013, 2020, doi: 10.3390/app10249013.

[13] A. Ataka, A. Sandiwan, H. Tnunay, D. R. Utomo, and A. I. Cahyadi, "Inverted pendulum control: A comparative study from conventional control to reinforcement learning," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi,* vol. 12, no. 3, pp. 197-204,2023, doi: 10.22146/jnteti.v12i3.7065.

[14] A. Dev, K. R. Chowdhury, and M. P. Schoen, "Q-learning based Control for Swing-up and Balancing of Inverted Pendulum," in *2024 Intermountain Engineering, Technology and Computing (IETC)*, 2024: IEEE, pp. 209-214, doi: 10.1109/IETC61393.2024.10564347.

[15] R. Hernandez, R. Garcia-Hernandez, and F. Jurado, "Modeling, Simulation, and Control of a Rotary Inverted Pendulum: A Reinforcement Learning-Based Control Approach. Modelling 2024, 5, 1824–1852," ed, 2024, doi: 10.3390/modelling5040095.

[16] R. S. Bhourji, S. Mozaffari, and S. Alirezaee, "Reinforcement learning DDPG–PPO agent-based control system for rotary inverted pendulum," *Arabian Journal for Science and Engineering,* vol. 49, no. 2, pp. 1683-1696, 2024, doi:10.1007/s13369-023-07934-2.

[17] P.-W.Cheng, "Reinforcement learning based controller design applied to rotary inverted pendulum system," M.S. thesis, National Yang Ming Chiao Tung Univ., Taiwan, 2021.

[18] T. M. Tijani and I. A. Jimoh, "Optimal control of the double inverted pendulum on a cart: A comparative study of explicit MPC and LQR," *Applications of Modelling and Simulation,* vol. 5, pp. 74-87, 2021.

[19] M. Rani and S. S. Kamlu, "Optimal LQG controller design for inverted pendulum systems using a comprehensive approach," *Scientific Reports,* vol. 15, no. 1, p. 4692, 2025, doi: 10.1038/s41598-025-85581-3.

[20] D. Maneetham and P. Sutyasadi, "System design for inverted pendulum using LQR control via IoT," *International Journal for Simulation and Multidisciplinary Design Optimization,* vol. 11, p. 12, 2020, doi: 10.1051/smdo/2020007.

[21] H. O. ERKOL, "Linear quadratic regulator design for position control of an inverted pendulum by grey wolf optimizer," *International Journal of Advanced Computer Science and Applications,* vol.

9, no. 4, 2018, doi: 10.14569/IJACSA.2018.090403.

[22] H. X. Cheng, J. X. Chen, J. Li, and L. Cheng, "Optimal control for single inverted pendulum based on linear quadratic regulator," in *MATEC Web of Conferences*, 2016, vol. 44: EDP Sciences, p. 02064, doi: 10.1051/matecconf/20164402064.

[23] U. Yıldıran, "Adaptive Control of an Inverted Pendulum by a Reinforcement Learningbased LQR Method," *Sakarya University Journal of Science,* vol. 27, no. 6, pp. 1311-1321, 2023, doi: 10.16984/saufenbilder.1286391.

[24] M. Safeea and P. Neto, "A Q-learning approach to the continuous control problem of robot inverted pendulum balancing," *Intelligent Systems with Applications,* vol. 21, p. 200313, 2024, doi: 10.1016/j.iswa.2023.200313.

[25] V. W. Hill, "Deep Reinforcement Learning Control for Disturbance Rejection in a Nonlinear Dynamic System with Parametric Uncertainty," *arXiv preprint arXiv:2404.04699,* 2024, doi: 10.48550/arXiv.2404.04699.

[26] J. Liu, X. Zhuan, and C. Lu, "Swing-up and balance control of cart-pole based on reinforcement learning DDPG," in *International Conference on Bio-Inspired Computing: Theories and Applications*, 2022: Springer, pp. 419-429, doi: 10.1007/978-981-99-1549-1_33.

[27] M. Bajelani and M. Tayefi, "Identification and control of unstable systems using internal and external loops implemented on a multirotor," *Mechanical Engineering of University of Tabriz,* vol. 51, no. 4, pp. 59-67, 2022, doi: 10.22034/jmeut.2022.12263.

[28] N. Mohammadi, M. Ebrahimi, M. Tayefi, and A. Nikkhah, "Reinforcement Q-learning based flight control for a passenger aircraft under actuator fault," *Discover Mechanical Engineering,* vol. 4, no. 1, pp. 1-22, 2025,doi: 10.1007/s44245-025-00090-x.

[29] H. Asrari, I. Mohammadzaman, and F. Allahverdizadeh, "Robust state-feedback controller for linear parameter-varying systems with time-invariant uncertainties," *Scientia Iranica,* vol. 30, no. 3, pp. 1148-1157, 2023,doi: 10.24200/sci.2021.56871.4953.

[30] A. Soltani and M. H. Kamari, "Hybrid Position and Force Control for a Spherical Inverted Pendulum Connected to a Quadrotor in a Constrained Motion," *Amirkabir Journal of Mechanical Engineering,* vol. 54, no. 10, pp. 2255-2276, 2022, doi: 10.22060/mej.2022.21308.7420.

[31] N. Mohammadi, M. Tayefi, and M. Zhu, "Vertical take-off and hover to cruise transition for a hybrid UAV using model predictive controller and MPC allocation," *Aircraft Engineering and Aerospace Technology,* vol. 95, no. 10, pp. 1642-1650, 2023, doi: 10.1108/AEAT-04-2023-0090.