DEEP REINFORCEMENT LEARNING EMPOWERED ACTIVE VIBRATION CONTROL STRATEGIES

ZHANG YIANG

PhD

The Hong Kong Polytechnic University

2026

The Hong Kong Polytechnic University

Department of Civil and Environmental Engineering

Deep Reinforcement Learning Empowered Active Vibration Control Strategies

ZHANG YIANG

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

August 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

ZHANG YIANG_____ (Name of student)

To the darkest times

ABSTRACT

Although classical active control techniques demonstrate superior performance in controlling structural vibrations, deriving the optimal control strategy often faces challenges due to the difficulty or complexity in accurately modeling dynamic systems in practical scenarios. In contrast, machine learning based (data driven) control methods eliminate such a requirement by directly learning control strategies from structural behavior. However, past studies on learning-based control algorithms primarily focused on their applicability without guaranteeing optimal control performance, resulting in a notable performance gap between learning-based model-free vibration control and model-based optimal control. This raises a challenging question:

How can optimal control performance be achieved without accurate knowledge of structural dynamics?

To answer this question, this thesis established and developed a novel model-free approach step-by-step, from linear systems to nonlinear systems, and validated it through numerical and experimental examples. The initially proposed framework was based on neural networks (NNs) and deep reinforcement learning (DRL) to control both linear and nonlinear systems effectively. The proposed learning algorithm could determine the control policy through interaction with the environment (i.e., the actual structures) without knowing structural dynamic models. First, the proposed method was demonstrated on linear systems, including a single-degree-of-freedom (SDOF) model, and then extended to multi-degree-of-freedom (MDOF) systems (such as a shear building model). The DRL-based strategy achieved control performance comparable to that of classical model-based controllers, such as the linear quadratic regulator (LQR),

even under conditions of partial observation and measurement noise. Therefore, the DRL-based strategy could provide optimal control without the need for detailed structural models, which has never been reported in the literature. Following that, the research extended to a nonlinear system, i.e., a shallow, simply-supported buckled beam. The findings illustrate the adaptability and efficacy of the DRL-based approach in preventing buckling under high loading levels without additional energy requirements. This work advances the current understanding of the DRL applications in structural control.

Though the exploration of pure DRL methods in structural control reveals their potential, the inherent limitations of DRL, such as extensive training data requirements and instability in real-world applications, limit their practical applicability. A pre-training stage was added by utilizing Imitation learning (IL). This IL-DRL framework initially involved training an NN controller through behavior cloning of an extremely simple but stable control strategy, followed by fine-tuning using model-free DRL. The framework's feasibility and effectiveness were systematically examined in simulation by using a verified numerical model of a full-scale stay cable system with an active damper. The control performance of the proposed model-free method closely approached that of the model-based full-state feedback controller, even under conditions of a reduced number of observed states or measurement noises. Moreover, it demonstrated superior sample efficiency compared with other state-of-the-art model-free DRL algorithms. The combined IL-DRL framework was validated through a shaking table experiment using an active mass damper to control an SDOF structure. It was demonstrated that the proposed approach could successfully train an optimal NN controller in the experiment without using any structural information. The response of the controlled structure under different excitations was compared to a classical optimal linear quadratic Gaussian (LQG) controller. The proposed controller achieved more displacement and acceleration reduction with a similar control power requirement. The results show that combining IL and DRL can significantly improve the sampling efficiency and performance of NN controllers and is suitable for real-world structural vibration control applications.

Furthermore, this thesis introduced a multi-objective control strategy based on the IL-DRL framework to enhance the control performance, enabling the controller to

optimize multiple objectives concurrently. This approach was validated experimentally on a three-story frame and through numerical simulations of the Canton Tower under wind-induced vibrations. The results demonstrate that the IL-DRL method achieves rapid convergence and surpasses the performance of classical model-based controllers (like LQR and LQG) with direct optimization to the target control performance indices (like the inter-story drift). These results affirm the method's efficacy as a model-free, multi-objective control strategy suitable for real-world applications.

In summary, a series of comprehensive investigations was conducted in this thesis to evaluate the efficacy and applicability of the model-free learning-based control strategies for structural vibration mitigation. For the first time in structural control literature, DRL was rigorously demonstrated to achieve optimal control performance comparable to conventional model-based methods. Building upon this foundation, a novel IL-DRL framework was proposed, representing the first successful integration of IL with DRL for vibration control applications. The developed strategy significantly enhanced performance, efficiency, and adaptability, showing considerable promise for future real-world applications, particularly in tackling complex, high-degree-of-freedom structures that have not been reported in the current study. Furthermore, the framework's inherent multi-objective control capability was rigorously examined and substantially strengthened within the IL-DRL architecture, achieving performance that surpasses model-based methods. This research contributes substantively to the advancement of intelligent structural control technologies while establishing a robust methodological foundation for future investigations in adaptive vibration mitigation.

PUBLICATIONS ARISING FROM THE THESIS

Journal Articles:

Zhang, Y. and Zhu, S.* Novel model-free optimal active vibration control strategy based on deep reinforcement learning. *Structural Control and Health Monitoring 2023;2023,* DOI: 10.1155/2023/6770137 (Top cited article in *Structural control and Health Monitoring* among work published between January 1, 2023, and December 31, 2023)

Zhang, Y. and Zhu, S.* Active nonlinear vibration control of a buckled beam based on deep reinforcement learning. *Journal of Vibration and Control 2024;31(1-2):70-81,* DOI: 10.1177/10775463241264112

Zhang, Y. and Zhu, S.* Active vibration control paradigm for stay cable vibration: A hybrid imitation and deep reinforcement learning approach. *Structural Control and Health Monitoring (under review)*

Zhang, Y., Ke, S. and Zhu, S. * Model-free active vibration control based on imitation learning and deep reinforcement learning: experimental implementation in an SDOF structure. *Smart Materials and Structures (under review)*

Zhang, Y. and Zhu, S.* Multi-objective active vibration control strategy combining imitation learning and deep reinforcement learning: experimental implementation on a multi-story frame. *Engineering Structures (under review)*

Zhang, Y. and Zhu, S.* Model-free active control of wind-induced vibration of

benchmark TV tower based on imitation learning and deep reinforcement learning. *Journal of Structural Engineering (under review)*

Conference Papers:

Zhang, Y. and Zhu, S.* Novel active structural vibration control strategy based on deep reinforcement learning. 10th ECCOMAS Thematic Conference on Smart Structures and Materials (SMART 2023), Patras, Greece, 03 - 05 July 2023. DOI: 10.7712/150123.9806.444204

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Songye Zhu, for his valuable guidance, continuous support, and insightful advice throughout my PhD journey at the Hong Kong Polytechnic University. His profound knowledge, rigorous academic attitude, and patience have been instrumental in shaping my research and professional growth. Without his encouragement and mentorship, this thesis would not have been possible.

I would like to extend my appreciation to the Hong Kong Polytechnic University for offering me the financial support and providing an excellent academic environment and resources, which enabled me to focus on my research.

I am also sincerely grateful to other group members for their constructive feedback and encouragement during my research. Their expertise and suggestions have greatly enriched my work. The stimulating discussions and collaborations with my colleagues and fellow researchers have been a source of inspiration and motivation.

Special thanks go to my family for their unconditional love, understanding, and unwavering belief in me. Their sacrifices and encouragement have been my pillar of strength during the challenging moments of this journey.

Lastly, I am indebted to my friends for their companionship. This thesis is dedicated to all those who have contributed, directly or indirectly, to my academic and personal growth.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

The main abbreviations used in this thesis are listed below:

AI	Artificial intelligence
AMD	Active mass damper
ANN	Artificial neural network
ARE	Algebra Riccati equation
ATMD	Active tuned mass damper
BC	Behavior cloning
DDPG	Deep deterministic policy gradient
DL	Deep learning
DNN	Deep neural network
DP	Dynamic programming
DQN	Deep Q-network
DRL	Deep reinforcement learning
IL	Imitation learning
LQR	Linear quadratic regulator
LQG	Linear quadratic gaussian
MDOF	Multi-degree-of-freedom
MDP	Markov decision process
MFRL	Model-free reinforcement learning
MSE	Mean squared error
NN	Neural network
PI	Policy iteration

PPO	Proximal policy optimization
PSD	Power spectral density
RL	Reinforcement learning
RMS	Root mean square
SAC	Soft actor critic
SDOF	Single-degree-of-freedom
SHM	Structural health monitoring
SOTA	State-of-the-art
TD	Temporal difference
TD3	Twin delayed deep deterministic
TMD	Tuned mass damper
TRPO	Trust region policy optimization
VI	Value iteration

CHAPTER 1

Introduction

1.1 Research Motivations

Structural vibration control is an essential aspect in various engineering disciplines, encompassing mechanical, aerospace, and civil engineering. Structural vibration control involves the application of various techniques and methods to reduce undesirable vibrations that may adversely affect the serviceability, safety, and longevity of structures and machinery.

In civil engineering, structural vibration control is vital for protecting buildings and infrastructures from the destructive dynamic loads (e.g., earthquakes and wind). In recent decades, an ongoing trend toward constructing supertall (300 meters and taller) buildings that are highly flexible has been observed (CTBUH, 2023). While increasing the stiffness of a structure can directly reduce vibration response, it inevitably increases structural weight and the associated costs. Modern sustainability-oriented design prefers less material usage in construction. Balancing the need to minimize risk and avoid excessive material use requires the development of smart control systems, which are designed to reduce structure vibrations under severe dynamic loadings, thereby preventing excessive damage and ensuring occupant comfort.

Over the past five decades, extensive research has aimed at mitigating structural responses to dynamic effects by employing various structural vibration control methods. These systems can be generally categorized into three types based on operational mechanisms: passive, active, and semi-active control systems (Spencer and Nagarajaiah, 2003). While passive and semi-active methods are noted for reliability and cost effectiveness, their adaptability to different types of excitations is restricted due to the internal limitations of passive components (Cheng et al., 2008). By contrast, active control techniques, such as active mass dampers (AMD) and active brace systems, can provide high-performance structural response reduction by calculating and generating the desired actuator control forces according to real-time observations (Housner et al., 1997).

A typical active control system includes actuators for force generation, sensors for response measurement, and control algorithms for force determination. The control algorithm is the core element in active control because it allows a structure to adapt to different excitations. Various approaches have been proposed over the past three decades. When the system information is fully known, model-based control algorithms, such as linear quadratic regulator (LQR), are often the choice (Yang et al., 1991; Spencer et al., 1993; Dyke et al., 1996). However, their performance is inevitably affected by the accuracy of the system dynamics and parameters. By contrast, developing intelligent learning-based algorithms, such as neural network (NN) control (Zhang and Jing, 2021), requires less system information. NN control solves the parameter uncertainty problem by reducing the structural response through learning; thus, it avoids the requirement of precise system information (Wen et al., 1995). However, most previous studies emphasized its simplicity and versatility in practical applications rather than control performance (Datta, 2003). This aspect leads to the first gap: the control performance gap between model-based and model-free vibration control methods.

Over the past decade, substantial advancements in artificial intelligence (AI) technology have benefited many industries. In particular, the development of deep reinforcement learning (DRL) has greatly enhanced NN controllers' capability across many fields, including gaming, autonomous driving, and robotic control (Silver et al., 2016; Hameed et al., 2025). DRL offers a powerful tool to bridge the gap between

learning-based control and optimal control (Li, 2017). Nevertheless, research on implementing DRL in the structural vibration control field is still lacking (Xie et al., 2020). How DRL can be applied to structural control is worth exploring. The advantages of DRL are apparent, but some disadvantages need to be addressed before implementation in real structures. The massive data requirement and unstable training prevent its direct use in actual structures due to hardware capability and safety concerns. Therefore, the second gap is a technological gap when advancing from simulation to real-world implementation.

Pre-training can substantially enhance the performance of DRL by providing an acceptable starting point, which accelerates training processes and improves training stability (Cruz Jr et al., 2017; Hester et al., 2018). By using techniques like imitation learning (IL), pre-training equips the DRL agent with a basic strategy, reduces the time needed to explore suboptimal strategies, and avoids the instability of initial random exploration (Ross et al., 2011). This approach also lessens the need for extensive interactions with the real environment, and saves time and resources while addressing hardware and safety concerns. Incorporating IL with DRL could make DRL more efficient and effective in complex real-world applications, which is still an underdeveloped area presenting the third research gap. The possibilities of applying it to structural vibration control are worth exploring.

The primary motivation behind this thesis is to address three research questions:

(1) Is a model-free DRL-based active controller whose performance is comparable to that of a model-based optimal controller feasible?

(2) How can this DRL strategy be made practicable for real-world structures?

(3) Can a model-free controller be improved to outperform a model-based optimal controller?

This thesis presents systematic simulations and experimental investigations for various scenarios to answer these questions step by step.

1.2 Research Objectives

The overarching objective of this research is to develop and enhance model-free active control of structural vibrations using DRL. Findings from six chapters, each contributing to the development and validation of the model-free control strategies, are integrated. The detailed objectives are summarized as follows:

(1) This work aims to propose and validate a novel active vibration control strategy based on DRL that does not require prior knowledge of structural dynamics. This strategy aims to find optimal solutions that traditional NN control methods could not achieve, with performance comparable to that of model-based optimal active controllers in linear systems with different observation conditions.

(2) This work aims to develop an advanced model-free control methodology for nonlinear vibration control that outperforms conventional model-based polynomial controllers. This methodology should be highly adaptable and easily implementable with customizable training parameters suitable for various application scenarios.

(3) This work aims to design and implement a model-free, learning-based control framework that employs IL for rapid initial training and DRL for subsequent optimization. This framework aims to enhance training stability and efficiency considerably. The efficacy of this combined approach is compared against traditional model-based optimal controllers using simulations on a full-scale stay cable under diverse excitations.

(4) This work aims to validate the IL-DRL framework experimentally on an AMD-controlled single-degree-of-freedom (SDOF) structure and demonstrates its transition from virtual environments to real-world applications. The proposed method's performance is compared with that of classical analytic optimal controllers to showcase its superiority under different excitation scenarios.

(5) This work aims to develop a multi-objective vibration control strategy based on the proposed IL-DRL framework. The reward criteria of the DRL consider multiple selective objective functions and enable the controller to optimize multiple objectives

simultaneously. This work addresses a limitation of classical model-based control algorithms, in which the form of objective functions is often fixed and cannot be adjusted to fit different situations. By embedding constraints and design features during training, the robustness and practical applicability of the IL-DRL framework are enhanced in the shaking table experiment of a three-degree-of-freedom structure with AMD.

(6) This work aims to investigate the application of the proposed multi-objective IL-DRL strategy in mitigating wind-induced vibrations of a full-scale structure. Canton Tower, equipped with an AMD controller, serves as the benchmark problem. The proposed control strategy is assessed under various wind speed levels to evaluate its effectiveness.

The research presented in this thesis aims to bridge the gap between model-free and model-based control methods and to ensure the practicability and robustness of DRL-enhanced active vibration control strategies in real-world structural applications.

1.3 Thesis Outlines

This thesis is organized into nine chapters, as illustrated in Figure 1.1.

Chapter 1 (i.e., this chapter) is an introduction. First, the research background and main objectives are stated, and the outline of this thesis is given.

Chapter 2 provides a detailed literature review, including classical active control theories and a brief history of classical NN methods in structural control. Then a general review of reinforcement learning (RL) methods, especially the state-of-the-art (SOTA) DRL techniques, follows. The limitations and challenges of existing studies are summarized.

Chapter 3 proposes a novel active vibration control strategy based on DRL. The proposed learning algorithm can determine the control policy through interaction with the environment (i.e., the controlled structures) without knowing the structural dynamic models. The proposed model-free strategy can provide optimal control performance comparable with that of traditional model-based optimal controllers for various systems and excitations. The proposed control strategy is first verified on an SDOF model and subsequently extended to a multiple-degree-of-freedom (MDOF) shear-building model.

Chapter 4 extends the framework in Chapter 3 to the nonlinear control area, which employs DRL to train an NN controller for nonlinear systems. The effectiveness and practicability of the proposed method are validated on a shallow, simply supported buckled beam. The proposed control strategy demonstrates excellent adaptability and ease of implementation.

Chapter 5 further combines IL and DRL to solve the unaddressed challenges of the direct implementation of DRL to vibration control of real structures. The proposed IL-DRL framework enhances adaptability, efficiency, and practical applicability compared with the previously proposed method in Chapter 3. The performance of the proposed control algorithm is examined in the simulation of a full-scale stay cable under various excitations.

Chapter 6 validates the proposed IL-DRL framework in the experiment. The proposed IL-DRL controller is trained and validated through a shaking table experiment of an SDOF structure controlled by an AMD. The proposed approach can successfully train an optimal NN controller in the experiment without using any structural information.

Chapter 7 further improves the IL-DRL framework and enables the controller to optimize multiple objectives simultaneously. This approach addresses the issue associated with classical model-based optimal control algorithms whose objective functions are often in a rigid form and cannot be adjusted to fit different situations. This framework is experimentally validated on a three-story structure equipped with an AMD.

Chapter 8 applies the method proposed in Chapter 7 to suppress the wind-induced vibration of the Canton Tower equipped with an AMD. A comprehensive numerical study based on the benchmark model of the Canton Tower is conducted to validate the efficacy of the control strategy.

Chapter 9 summarizes the conclusions of this thesis and provides discussions on future research directions.

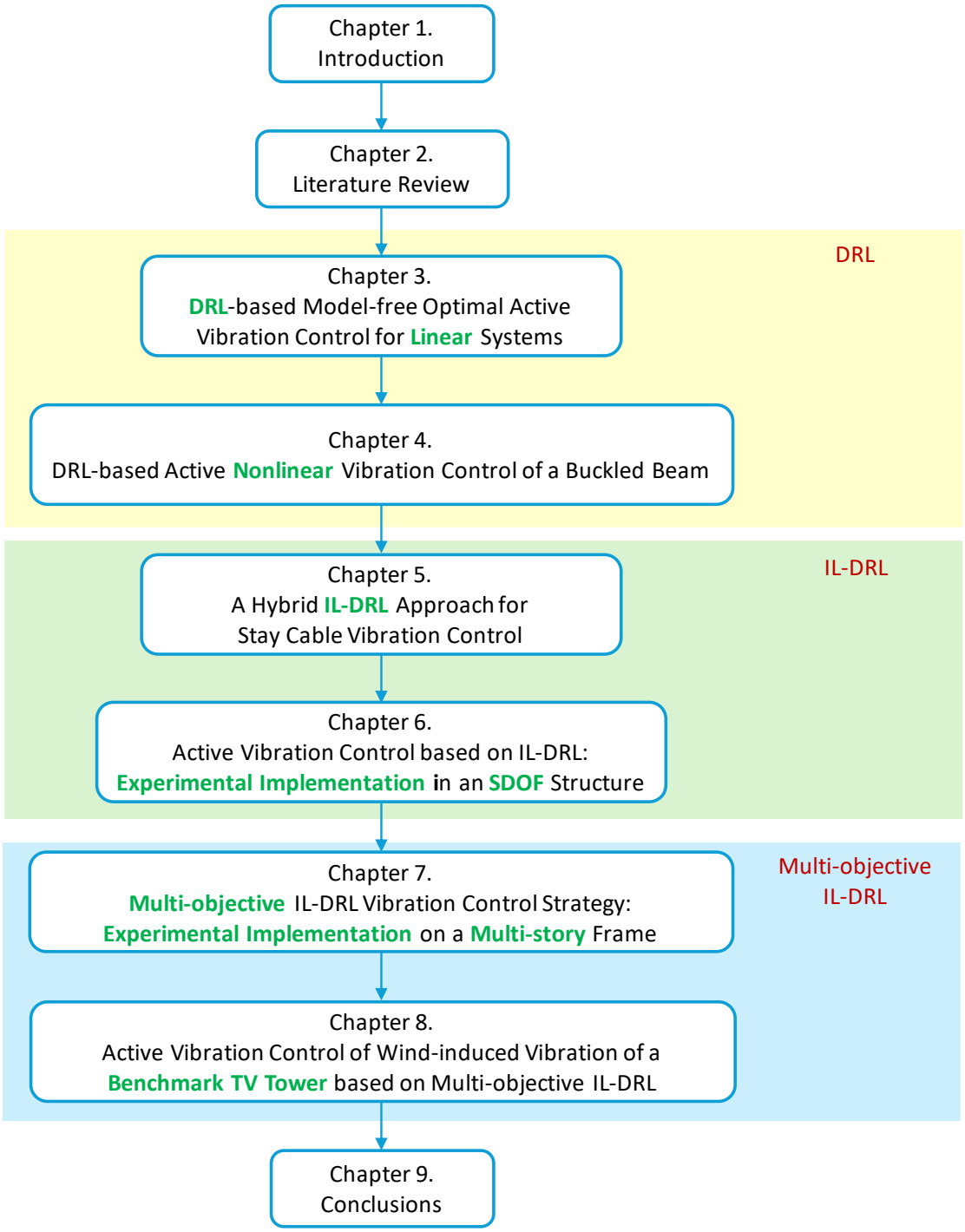


Figure 1.1 Organization of thesis

CHAPTER 2

Literature Review

With the main objective of this thesis presented in Chapter 1.2, this chapter focuses on two parts: the previous research on active structural control and advanced deep learning (DL) methods that will be adopted in the succeeding chapters for the development of model-free controllers. The scope of the reviewed works is presented by a flowchart in Figure 2.1. First, the classical active structural control is reviewed in two aspects: model-based control and learning-based (data-driven) control. Next, DRL, including its key elements and the corresponding algorithms, is reviewed. In addition, some concepts about pre-training are introduced.

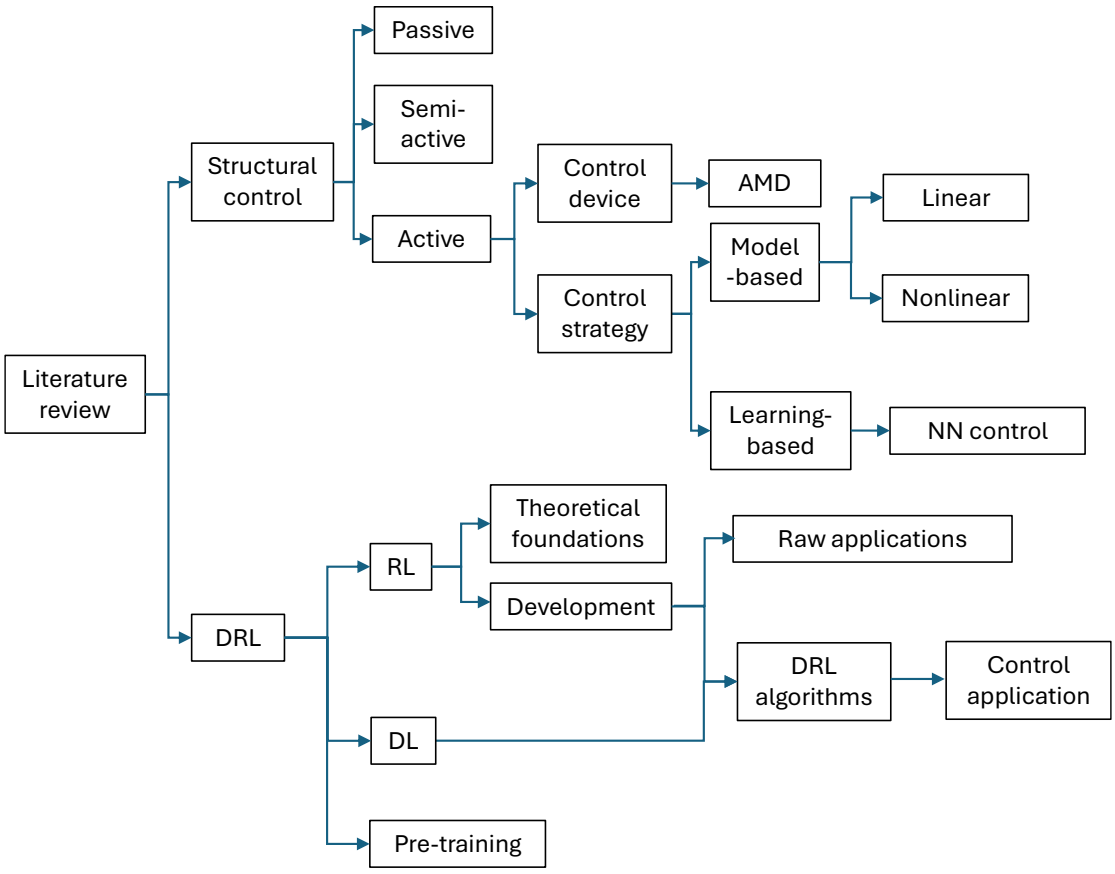
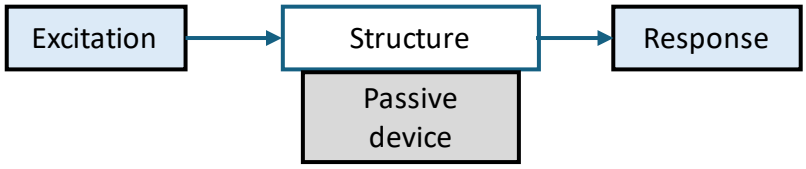


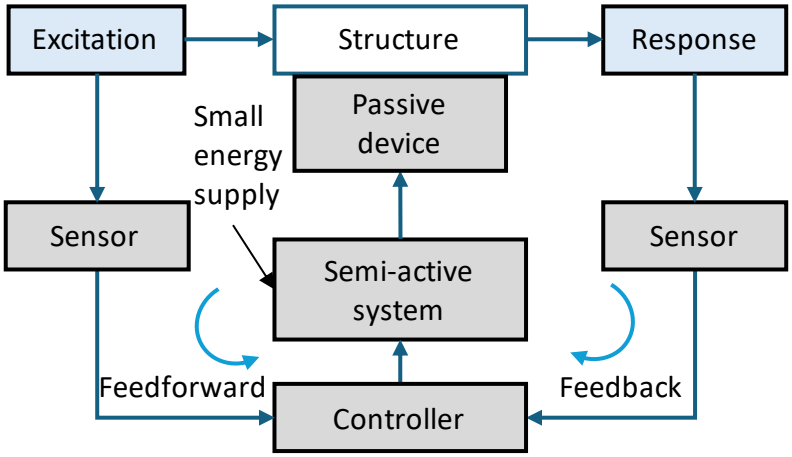
Figure 2.1 Directory of literature review

2.1 Overview of Structural Control

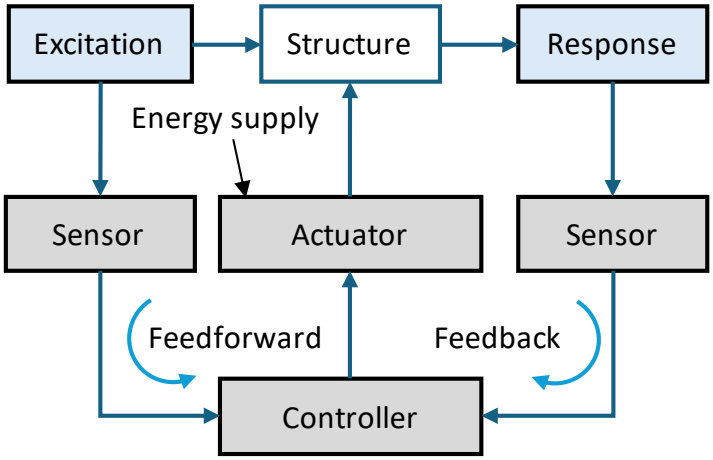
Structural vibration control involves the use of devices and algorithms to improve the performance of structures under dynamic loadings. The goal is to mitigate the negative effects of vibrations caused by wind, earthquakes, waves, traffic, and other dynamic loads (Housner et al., 1997). Because strength or stiffness design alone is often not adequate to prevent excessive vibration, numerous structural control systems have been established. Generally, these systems can be classified into three main groups based on their different operational mechanisms: passive, semi-active, and active types (Spencer and Nagarajaiah, 2003). The essential difference is the need for an external energy supply and a feedback/feedforward system, as illustrated in Figure 2.2.



(a)



(b)



(c)

Figure 2.2 Block diagram of structural control systems (Symans and Constantinou, 1999): (a) passive control system, (b) semi-active control system, and (c) active control system

2.1.1 Passive systems

Passive control systems suppress vibration by dissipating a portion of the input energy without requiring external power supply (Saaed et al., 2015). Two common types of these systems include isolation systems (Kelly, 1993; Skinner, et al., 1993; Soong and Constantinou, 2014) and energy dissipation devices (Constantinou et al., 1998). Unlike active control systems, passive devices operate without feedback mechanisms. Instead, they convert the kinetic energy of the host structure into heat through relative motion, thereby damping vibrations directly, as shown in Figure 2.2a.

Various passive control systems developed encompass diverse technologies with distinct working principles and applications (Symans et al., 2008). For example, metallic yield dampers utilize inelastic deformation of components (such as torsional beams or V-strips) to absorb energy and demonstrate reliable hysteretic performance and good fatigue durability, though their nonlinear behavior requires advanced analytical modeling (Li and Li, 2007; Moreschi and Singh, 2003). Friction dampers, adapted from vehicle braking mechanisms, provide effective energy dissipation through regulated sliding surfaces, which are particularly effective for seismic applications (Cameron et al., 1990; Mualla and Belev, 2015). Viscoelastic damping devices operate through shearing deformation of the polymer layer, and their performance is usually dependent on operational frequency, deformation magnitude, and environmental temperature (Lin et al., 1991). Viscous fluid dampers (VFD) achieve energy absorption through piston movement in specialized high-viscosity fluids (Lee and Taylor, 2001; Samali and Kwok, 1995).

Tuned mass dampers (TMDs), consisting of auxiliary mass–spring–damper assemblies precisely calibrated to structural frequencies, represent a widely implemented solution for wind-induced vibrations of high-rise structures (Kaynia et al., 1981; Tsai and Lin, 1993). A landmark implementation occurred in Boston's 244-m John Hancock Tower (Soto and Adeli, 2013). Tuned liquid dampers (TLDs) constitute a TMD variant that harnesses liquid sloshing effects for vibration control, predominantly in wind-sensitive buildings (Fujino et al., 1992; Kim and Adeli, 2005; Sun et al., 1992). Enhanced versions include multiple TMD configurations arranged in series or parallel to broaden effective frequency ranges (Igusa and Xu, 1994; Kareem

and Kline, 1995; Ying et al., 2003), and pendulum TMD systems whose frequencies are adjusted through cable length variation, as notably implemented in Taipei 101 (Tuan and Shang, 2014).

Passive control devices are usually tuned to address specific dynamic loading scenarios and offer inherent stability and simple design implementation (Spencer and Nagarajaiah, 2003). However, these systems exhibit three fundamental limitations: (1) Their energy dissipation capacity depends entirely on structural response and lacks adaptability to changing loading conditions or structural modifications. (2) Despite demonstrated reliability and robustness, their effectiveness remains confined to a narrow frequency range. (3) They are inadequate in reducing responses in local regions (Connor and Klink, 1996). These limitations are particularly evident in devices such as TMDs, whose performance is highly sensitive to tuning ratios and may suffer from higher-mode amplification during seismic pulses (Den Hartog, 1985). Such constraints underscore the necessity for adaptive control strategies in applications, requiring wider operational bandwidths or tunable control characteristics.

2.1.2 Semi-active systems

Semi-active systems originate from passive control systems. While maintaining the reliability of the passive components, the mechanical properties are adjustable through the controllable parts (Symans and Constantinou, 1999). Semi-active control typically requires fewer external power sources than active control. The history of research on semi-active dampers in structural engineering dates back to 1983 (Hrovat et al., 1983) and has been receiving growing attention over the past four decades.

Semi-active systems primarily function through several distinct mechanisms: stiffness-modulating devices that actively regulate bracing system rigidity (Kobori et al., 1993; Yang et al., 1996), variable fluid dampers that employ controllable valves to adjust output forces (Kawashima et al., 1992), and variable friction dampers capable of modifying normal forces to control energy dissipation (Xu et al., 2001). Smart materials adopted include electro-rheological dampers that utilize electric fields to control shear properties (McMahon and Makris, 1997; Tang et al., 2011; Zuo et al., 2013), and magneto-rheological (MR) dampers governed by magnetic field variations (Spencer et

al., 1997; Jansen and Dyke, 2000).

The semi-active principle has been successfully extended to tuned mass systems, including TMDs with adjustable dynamic parameters (Hrovat et al., 1983; Abe, 1996) and TLDs where tank geometry modifications enable frequency tuning (Chang et al., 1998). A notable application demonstrating the effectiveness of this technology is the hybrid system developed by Kim and Adeli (2005), which combines passive VFDs with semi-active TLDs for the mitigation of the wind-induced vibration of a 76-story high-rise building; this system is robust against structural stiffness modeling uncertainties.

As illustrated in Figure 2.2b, semi-active control systems dynamically adjust their operational parameters through continuous feedback from either excitation inputs or structural response measurements. This adaptive capability is achieved through a controller unit that processes real-time sensor data and generates appropriate command signals to modulate the semi-active component's properties based on pre-defined control algorithms. The power requirements of semi-active control systems are modest, typically in the range of tens of watts (Symans and Constantinou, 1999). Semi-active systems outperform passive counterparts by enabling real-time adaptability to structural demands. Semi-active approach achieves enhanced control efficacy through energy-efficient adjustments rather than full-force generation.

A more detailed review of semi-active structural control can be found in Spencer and Nagarajaiah (2003). Semi-active control systems, while offering improved adaptability compared with passive systems, still exhibit the fundamental limitation. Although capable of parameter adjustment to some extent, their tuning range and response speed are constrained by the passive components, which makes ideal real-time control difficult to achieve. These technical constraints result in control performance that, while superior to passive systems, generally falls short of that achievable through fully active control implementations (Shi et al., 2021).

2.1.3 Active systems

Passive and semi-active control systems offer cost-effective, reliable vibration mitigation solutions, but they share a fundamental limitation in their inability to adapt

to varying excitation characteristics fully (Cheng et al., 2008). This constraint is overcome in active control systems, which employ external power sources to generate precise control forces through electrohydraulic or electromechanical actuators, as conceptually illustrated in Figure 2.2c. Unlike their passive and semi-active counterparts, these systems can produce counteracting forces that are independent of structural motion, which enables superior adaptability to diverse dynamic loading conditions (Fujino et al., 1996). The active approach enhances performance through continuous monitoring of structural responses and real-time computation of optimal control forces, though at the cost of substantially higher energy requirements and system complexity.

The development of active structural control systems has evolved considerably since Yao's pioneering proposal (Yao 1972) for storm protection of tall buildings. Building upon this foundation, Soong and Manolis (1987) established the modern framework for active control in civil structures and examined various actuation methodologies. These systems fundamentally differ from passive approaches through their ability to inject and dissipate energy via actuator forces (Korkmaz, 2011) and employ diverse implementation strategies. Structural shape control methods include active bracing and tendon systems, whereas nonshape-altering approaches encompass AMDs and active tuned-mass dampers (ATMDs).

Particularly in cable-stayed bridges, active tendon systems have demonstrated superior vibration mitigation compared with traditional TMDs (Abdel-Rohman and Leipholz, 1983) and effectively maintained serviceability limits under dynamic loads (Ziegler, 2005). Experimental validations have confirmed their efficacy in reducing vertical vibrations from harmonic excitations (Warnitchai et al., 1993), while advanced implementations incorporate nonlinear decentralized control for seismic response mitigation (Magana and Rodellar, 1998) and piezoelectric actuators for precise force modulation (Perumeont et al., 2000). The field has been further advanced through comprehensive reviews examining tendon system applications in cable structures (Juan and Tur, 2008; Tibert et al., 2008; Rhode-Barbarigos et al., 2010), and active control has been established as a versatile solution for various structural challenges.

AMDs have emerged as one of the most effective active control strategies for civil

engineering structures and demonstrated remarkable vibration-reduction capabilities across various applications. The technology was first successfully implemented in the Hyobashi Seiwa Building in Japan and achieved 50% vibration reduction compared with uncontrolled responses (Kobori et al., 1991). Samali et al. (1985) validated the effectiveness of AMDs in mitigating wind-induced accelerations in high-rise structures through their numerical study of a 40-story building subjected to strong wind excitations. The versatility of AMD systems is further evidenced by their successful application in offshore structures, where Ma et al. (2006) demonstrated their efficacy in controlling vibrations caused by irregular wave loads. Two AMDs were installed on the roof of the "Sendagaya INTES" in Tokyo, Japan, the wind and earthquake records for five years were investigated by Yamamoto et al. (1998), and the results demonstrated effective vibration control by virtual damping increment. Figure 2.3 shows the building and AMDs in the study as an illustration of the AMD application. An AMD system installed on the Osman Gazi Bridge tower successfully reduced the vortex-induced vibration during construction and in-service conditions (Inoue et al., 2025). The locations, layout, and AMD unit are shown in Figure 2.4. These diverse applications collectively highlight AMD's superior performance in structural vibration control across a range of engineering scenarios.

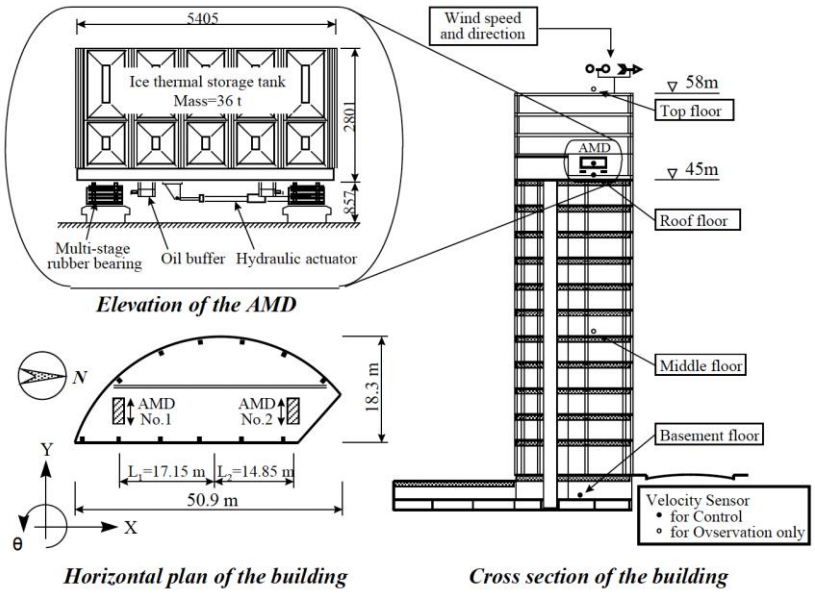


Figure 2.3 AMD system installed in the "Sendagaya INTES" (Tokyo, Japan) (Yamamoto et al., 1998)

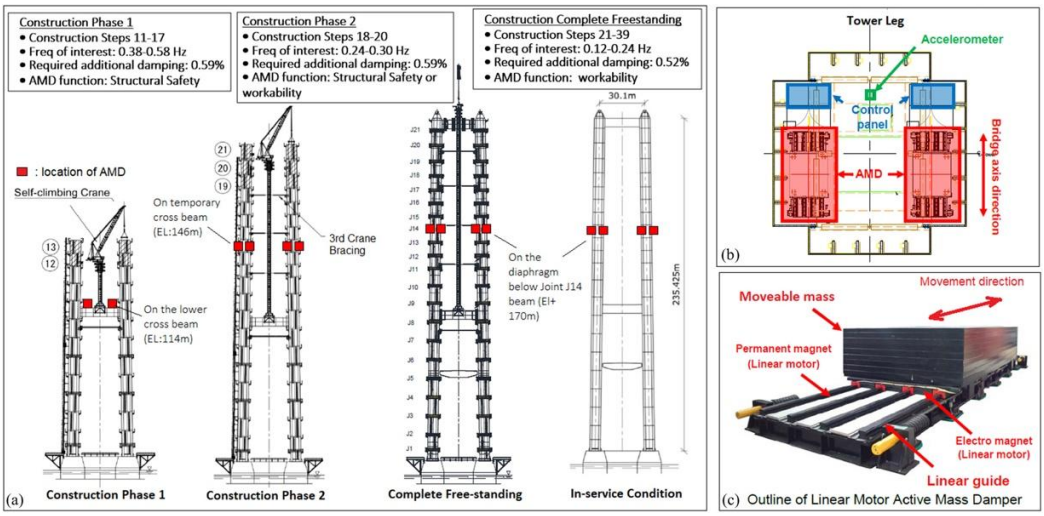


Figure 2.4 AMD system installed on the Osman Gazi Suspension Bridge Tower (Inoue et al., 2025)

ATMD is created through the connection of a standard TMD to an actuator, which was first proposed by Chang and Soong (1980) as an extension of a passive TMD. ATMD allowed smaller seismic masses and has served as the primary active control mechanism used for seismic protection with the development of sensing technology. More recently, ATMD technology has been implemented in supertall structures such as the Shanghai World Financial Center, which showcases its adaptability to different structural configurations and loading conditions (Xu et al., 2014). Modern super slender skyscrapers such as the Shenzhen's 600-meter Ping An Finance Center (PFAC) now employ ATMD that combine the frequency adaptability of active control with the fail-safe reliability of passive mass dampers (Zhang et al., 2019; Zhou et al., 2022); a schematic diagram and layout is shown in Figure 2.5.

A typical AMD/ATMD control system includes an actuator for force generation, sensors for vibration measurement, and a controller with a predetermined control algorithm. The control process involves three key phases: (1) measuring structural responses; (2) computing forces using sensor feedback; and (3) applying these forces to achieve desired performance bounds for accelerations (human comfort), displacements (structural integrity/safety), and stresses (material durability). The control algorithm is one of the core elements because it allows the structure to adapt to different excitations.

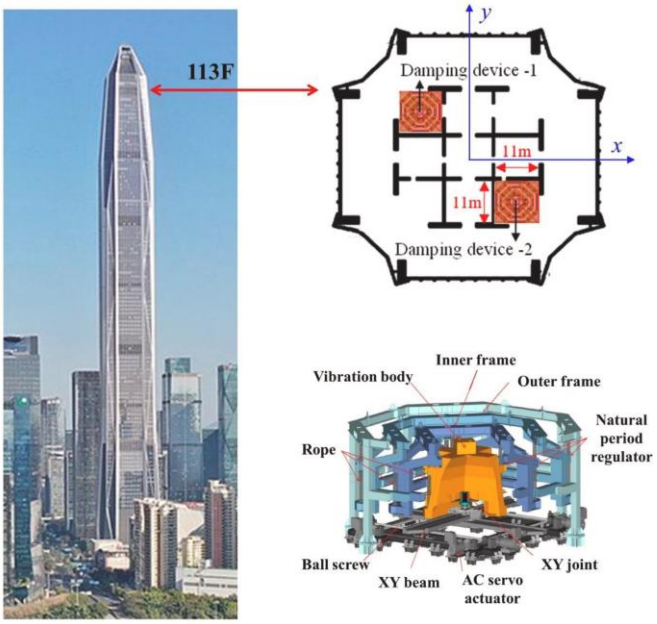


Figure 2.5 ATMD system installed on PFAC (Zhou et al., 2022)

Furthermore, combining the three major classes of control systems forms hybrid control systems.

2.2 Active Structural Control

As stated in subchapter 2.1, an active control system contains sensors, actuators, and controllers. The active control strategy, which dynamically adjusts actuator input signals to generate precise control forces, is the key to superior vibration suppression performance over broad frequency ranges. The effectiveness depends on controller design, system parameters, and performance metrics. The active control system requires substantial power supplies and real-time feedback from sensors monitoring structural responses and/or excitations.

A primary step in the control design for a smart adaptive structure is the selection of a control algorithm that can adjust the control forces to counteract various external excitations (Wani et al., 2022). An ideal control algorithm should exhibit robustness, versatility, design simplicity, and ease of implementation. Over the past three decades,

numerous control algorithms have been proposed to pursue superior control performance in different control problems (Datta, 2003; Fisco and Adeli, 2011). According to the information needed for calculating control forces, active control can be divided into feedforward and feedback control. This subchapter focuses on reviewing the classical feedback-based active control algorithms used to determine the control forces based on the response measurement from sensors.

2.2.1 Model-based control

2.2.1.1 Linear control

Based on mathematical modeling, linear control represents the type of active control in which all structural dynamic equations and operations are linear. The most popular linear control algorithm is linear optimal control theory.

Linear optimal control focuses on developing control strategies based on the linear models of structures, where the control forces are given as the gain matrix times the state vector. The gain matrix is derived through the minimization of a quadratic performance index. Techniques such as LQR are commonly used to minimize a cost function representing the trade-off between control effort and structural response. The control gain matrix is derived by solving the algebraic Riccati equation (ARE). A detailed description of this method, which is regarded as a standard classical control method compared with the proposed DRL controller, is discussed in Chapter 3. Many works have been done on linear optimal control for structural vibration, such as Soliman and Ray (1972), Yang (1975), and Sarbjeet and Datta (1998).

If the performance function is time dependent and external excitation is known, the minimization of the cost function over the time interval is classified as instantaneous optimal control (Abdel-Rohman and Leipholz, 1979; Yang et al., 1987). Incorporating the acceleration of the structure into the feedback, apart from displacement and velocity, forms the generalized feedback control, in which the modified cost function contains the acceleration term (Yang et al., 1991; Spencer et al., 1993; Dyke et al., 1996). An analytical solution for the LQR solution to protect structures from seismic motion was proposed by Aldemir et al. (2001). Combining observers such as Kalman filters with

LQR control strategy leads to linear quadratic Gaussian (LQG) control, often used for systems with Gaussian white noise (Fisco and Adeli, 2011).

For structures in which the dominant vibration modes are the first few modes, the gain matrix of the controller could be selected to modify the characteristics (e.g., the eigenvalues) of the considered modes. Although this pole assignment technique is not optimal, it could reduce the structural response (Abdel-Rohman and Nayfeh, 1987). Other widely conducted linear control strategies include the proportional-integral-derivative control (Guclu and Yazici, 2008; Nerves and Krishnan, 1995), H∞ control (Chen et al., 2007; Li et al., 2014; Park et al., 2008; Utkin, 1990), and skyhook control (Liu et al., 2019).

2.2.1.2 Nonlinear control

While active optimal control methods have achieved considerable success in linear systems, extending their applications to nonlinear systems still faces challenges. Nonlinear systems are inherently more complex and sensitive to initial conditions (Abdel-Rohman and Leipholz, 1978; Astolfi et al., 2008; Iqbal et al., 2017; Nayfeh et al., 1980; Xing et al., 2016). Direct implementation of active linear controllers to nonlinear systems relies on the linearization of nonlinear systems around specific operating points (Agrawal et al., 1998). However, this approach has limitations because errors are introduced whenever deviations from these predetermined points occur (Tomasula et al., 1996).

Historically, the methods for solving nonlinear feedback control problems can be classified into two main categories (Li and Todorov, 2004) based on Bellman's optimality principle and Pontryagin's maximum principle (Bryson and Ho, 1975; Lewis et al., 2012). Globally optimal solutions for the former existed, but the partial differential equation (Hamilton–Jacobi–Bellman [HJB] equation) is only solvable for low-dimensional systems. While groundbreaking in their time, these conventional methods necessitate a precise understanding of the system's mathematical model and parameters. These methods typically aim to minimize a polynomial performance index; by incorporating higher-order performance indices based on the LQR problem, a nonlinear control force can be deduced through cost function minimization, which

yields a control force that becomes a nonlinear function of the state vector (Shefer and Breakwell, 1987; Suhardjo et al., 1992; Wu et al., 1995; Agrawal et al., 1998). However, these strategies encounter limitations when applied to high-dimensional systems, as the HJB equation, central to their optimization, becomes increasingly difficult to solve.

Sliding mode control (SMC) (Mahmoudi et al., 2019) and its variant, fuzzy SMC, saturated PD SMC (Zhang et al., 2022; Yang et al., 2023) enrich the nonlinear control toolkit by offering robustness against external disturbances from their capacity to enforce a specific control law once the system's state attains the sliding surface, guaranteeing stability under diverse conditions. Nevertheless, they may cause chattering phenomena or increasing computational overhead (Allen et al., 2000; Yang et al., 1996; Zhao et al., 2000).

Additional nonlinear control techniques include pulse control (Pantelides and Nelson, 1995), perturbation method (Jansen, 2008), stochastic dynamic programming (Lü et al., 2020), and iterative LQR (Li and Todorov, 2004). The performance of these controllers degrades when the modeling assumptions cannot accurately capture the nonlinear dynamics.

In general, nonlinear physical systems inherently resist exact mathematical description, which makes an absolute performance guarantee unattainable. While engineers approximate dynamics through first-principle or data-driven models, these inevitably end in linear regimes where solutions remain tractable. Fundamental limitations persist: No theoretical framework exists for optimal model selection in control design, and critical states often elude direct measurement. These challenges, compounded by the high-power demands and sensor dependency of active systems, motivate data-driven strategies that bypass the requirements for precise system characterization.

2.2.2 Learning-based control

For all the algorithms mentioned in subchapter 2.2.1, system parameters are a prerequisite in controller design. Recently, research on intelligent learning-based control has emerged, eliminating the need for complete knowledge of system dynamics.

NN control leverages the learning capabilities of NNs (also known as artificial NNs) to develop control strategies. These machine learning (ML) methods can adapt to changing conditions and learn optimal control policies from data. Like most soft computing techniques that try to mimic human behavior, NNs are computational models designed to mimic the learning abilities of a human brain. A typical simple NN consists of three layers: input, hidden, and output. The processing units are called artificial neurons, which are inspired by biological neurons.

Individually, the neurons perform trivial functions, but collectively, in the form of a network, they can solve complicated problems. NNs rely on past knowledge, and when presented with input–output data pairs, they construct a functional relationship through a learning process. The learning ability of NNs is attributed to the adjustment in interneuron connections or the synaptic weight value. Adaptability to change input–output data, nonlinear function mapping, and the ability to capture unknown relationships make NNs a versatile tool for real-world modeling problems. Furthermore, by incorporating multiple hidden layers, the NN could be extended to a deep neural network (DNN), which enables superior model performance.

The NN controller has the advantages of avoiding the need to develop a control algorithm analytically (Datta, 2003) and the ability to be applied to nonlinear control. In addition, the NN controller can be used for structures with unknown dynamics. Various kinds of NN controllers have been suggested in the literature. In this part, the classical NN applications in active control are discussed. The term “classical” is used to distinguish the early NN strategies from the DL techniques developed in the last decade. Sidiique and Adeli (2013) discussed various ways of using NN for control. They classified the applications into two categories: system identification and controller design. This thesis mainly focuses on the latter one. More detailed reviews of other applications in civil engineering fields could be referred to Adeli (2001), Chandwani et al. (2013), and Waszczyszyn (2010).

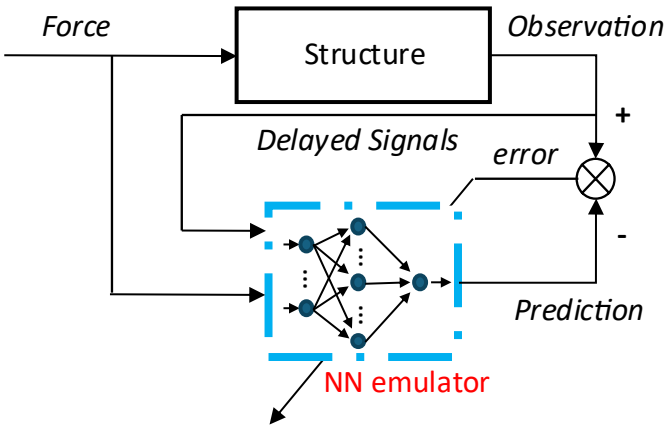


Figure 2.6 Training of emulator network (Ghaboussi and Joghatatie, 1995)

The majority of the past methods used in structural vibration control utilize NNs that contain an emulator network. The emulator network is trained to learn some relationship between the structure responses observed and the controller forces, as shown in the flowchart in Figure 2.6. The successfully trained network could either directly or indirectly obtain the control force or be used to train other networks. The early applications of NNs in structural control follow the strategy of using two NNs: one emulator network and one controller network. The learned emulator network is backpropagated to train the controller network to minimize the control criterion. After the training is finished, the controller network could control the structure responses independently (i.e., without the emulator network), like regular control algorithms.

Ghaboussi and Joghataie (1995) used a modified backpropagation (BP) algorithm for a three-story frame structure. The emulator network learns the relationships under different levels of the EI Centro earthquake. Moreover, the controller is trained for two criteria: one is that the average of an expected response for the future time steps is zero, and the other is a small value of the expected response. The control signals are calculated in the iterative loop to meet the criterion. The Jacobian is first derived to represent the sensitivity of the acceleration to the controller signals, and then the inverse is used to correct the signal through BP. However, a key limitation is that the instantaneous control objectives demand immediate suppression of vibrations at every moment, which can be overly stringent and energy inefficient that are difficult to achieve in practical applications.

Bani-Hani and Ghaboussi (1998) used a similar method for controlling a nonlinear three-story steel frame structure. They compared the capabilities of linearly and nonlinearly trained NN controllers. For each step, the actuator signal that satisfies the control criterion is searched for with the help of the emulator network.

Chen et al. (1995) also used BP to train the emulator and controller network using the historical data from an actual building subjected to recorded earthquake ground motions. The instantaneous error function adopted is the error between the estimated response of the next step and the desired zero response. The control error is backpropagated through the emulator to change the weights of the controller repeatedly until the actuator's capacity is reached. This training is shown in Figure 2.7.

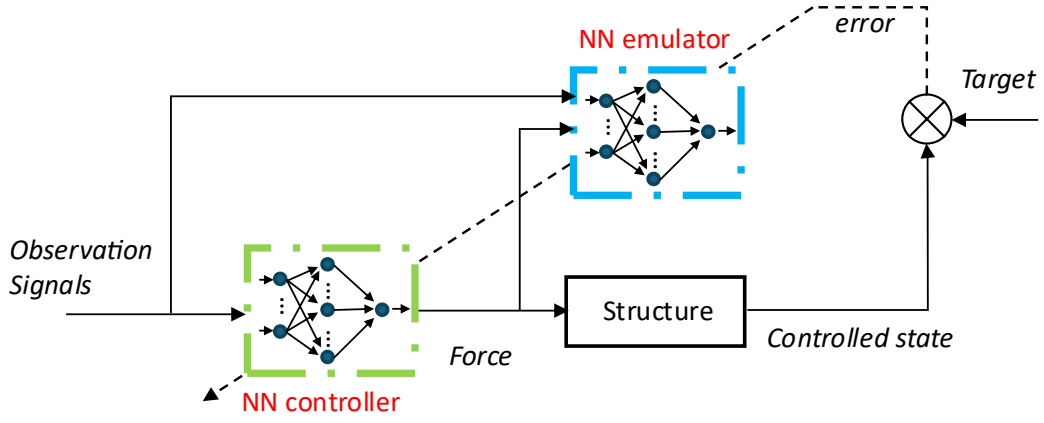


Figure 2.7 Training of NN controller using emulator network (Chen et al., 1995)

Tang (1996a) used a different strategy to control an SDOF system. The input of NN includes the displacement and velocity from the current and previous states, as well as the force input of the previous state, which includes the control force and excitations. The network output is the desired force of the current state. The NN is trained to predict the force input of every state. After training, it is first used to calculate the desired force of the current state, and by setting the input of the future state to zero, the NN is used again to determine the desired control force of the next state. Although only one network is used in this strategy, the implementation is more complicated compared with the emulator–controller strategy because for every time step, the NN must be implemented twice before the control force is generated. A simple nonlinear model was also used for testing, but the performance was poorer than that of the theoretical control.

Yen (1996) also applied the BP algorithm to control flexible multibody structures. Instead of using the standard feedforward networks, radial basis function networks were adopted. Nikzad et al. (1996) compared the performance of the NN controller with that of the conventional feedforward controller on a two-degree-of-freedom structure accounting for the adverse effects of the actuator dynamics and the computation phase delay. The training method was to use a modified BP algorithm to train an inverse emulator network and use it as the controller so that the process is completely supervised. The results showed that the NN achieves better performance compared with the conventional approaches, particularly in compensation for computational time delay and reduction of higher frequency noise. Because the NN is trained within the frequency range of interest, it ignores the higher frequency noise. Tang (1996b) used a multilayer feedforward perceptron network trained through the BP algorithm as the state estimator based on the partially observed states for active control. This method is straightforward to implement because it only involves matrix multiplication in the output computation.

Kim et al. (2000) followed the emulator–controller strategy but changed the error function to the quadratic cost function when training the NN controller. Because the strategy utilizes a pattern learning mode, the cost is minimized at each time step, which results in an instantaneously optimal solution. Thus, it can be regarded as an early implementation of a model-based reinforcement learning method. Linear and nonlinear SDOF systems were trained under the EI Centro earthquake using structural displacement, velocity, and ground acceleration as input. However, the performance has not been compared with that of the classical LQR algorithm that directly optimizes the global quadratic cost function.

Cho et al. (2005) applied a multilayer NN controller, in which the optimal number of hidden neurons is selected along with several iterative training cycles to a two-degree-of-freedom bridge system, and the robustness is shown by using randomly generated earthquake excitation for testing. Hung et al. (2000) applied the emulator–controller framework to train an active pulse controller using a modified BP algorithm, which learning ratio is determined through a search algorithm. Although the NN controller was compared with LQR, the results varied as the weighting matrix changed. Rao and Datta (2006) used two sets of NNs to control a 10-story building frame: One

for the generalization of actual acceleration, and the other outputs the control force by inputting the ground acceleration and the generalized acceleration. The time-delay effect is considered in this control scheme.

More advanced NNs, such as probabilistic NN (Jiang and Adeli, 2008), wavelet NN (Jaddu and Hiraishi, 2006; Khodabandehlou et al., 2018), and counter propagation network (CPN)(Madan，2004), have been adopted for structural control. Kim et al. (2008) proposed a method called lattice probabilistic NN to reduce the calculation time for the control forces by only considering the adjacent patterns (response of the structure). Madan (2005) used CPN, an unsupervised type of NN, to generate control forces without any target force and applied it to an eight-story building model. The input to the CPN is the sampled current and delayed structural response, excitation, and control forces. Each sample constitutes a training pattern, and the network outputs the control signal at the current step. Liu et al. (2009) proposed an improved BP-NN to predict the inverse model of the MR damper and time-delay elimination. Two controllers were trained: the first modifies the structural model offline, and the second amends the error through online feedback.

Other applications of ML in controlling civil engineering structures include genetic algorithms, fuzzy logic, and wavelet analysis. A more comprehensive review on the application of ML in structural control is available in Adeli (2001), Li and Adeli (2018), Thenozhi and Yu (2013), and Xie et al. (2020).

2.3 Deep Reinforcement Learning

In this subchapter, a general literature review of DRL is presented. The standard DRL setup is introduced along with reviews of DRL concepts and SOTA DRL algorithms. The algorithms adopted in the following chapters are briefly reviewed. The previous applications of DRL in the control field are also discussed.

2.3.1 Overview of DRL

DRL combines reinforcement learning (RL) with DNNs to handle high-dimensional state and action spaces. Originating from animal learning behavior, RL is a technique that enables the agents to find the optimal policy directly from their own experience within the environment and has been used by researchers to find the control policy in feedback control. RL is a general framework that can be used in various applications, such as robotics and computer games. A more thorough history and overview of RL can be found in Sutton and Barto (2018).

The development of RL algorithms bridges the gap between learning-based control and optimal control. Unlike supervised and unsupervised learning algorithms, which often focus on prediction or classification, RL agents learn by interacting with the environment to find a strategy that optimizes long-term reward. RL has been widely applied to computer games because the reward is clearly defined as the score. Whereas in structural control, the reward is implicit, and different reward functions are selected to meet different control performance requirements. The observation can be set as the acceleration, velocity, and displacement of the target structure in structural control.

This interactive learning renders RL an ideal method for realizing a model-free optimal controller by finding the extremum points of a given objective function (Ogata, 2010). While traditional NN-based control typically uses networks trained offline to approximate control laws or system dynamics, RL-based control leverages NNs as an adaptive approximation of key functions, such as policies, value functions, and action value functions. However, the classical RL methods are limited by the hardware abilities of the computing devices.

2.3.2 RL Setup

The RL problem includes five key components: environment, agent, state, action, and reward. In analogy to the feedback control system, the environment is equivalent to the plant model, and the agent acts as the controller. The state equals the complete observation of the environment, but the observation can also be part of the state. The action and observation are the same as the control scheme, and the agent interacts

directly with the environment by observing the state and performing the action (Sutton and Barto, 2018). The different kinds of actions are discrete, continuous, deterministic, and stochastic, and the set of all valid actions is called an action set. The distinction of different action spaces leads to different DRL methods, as discussed in a later part.

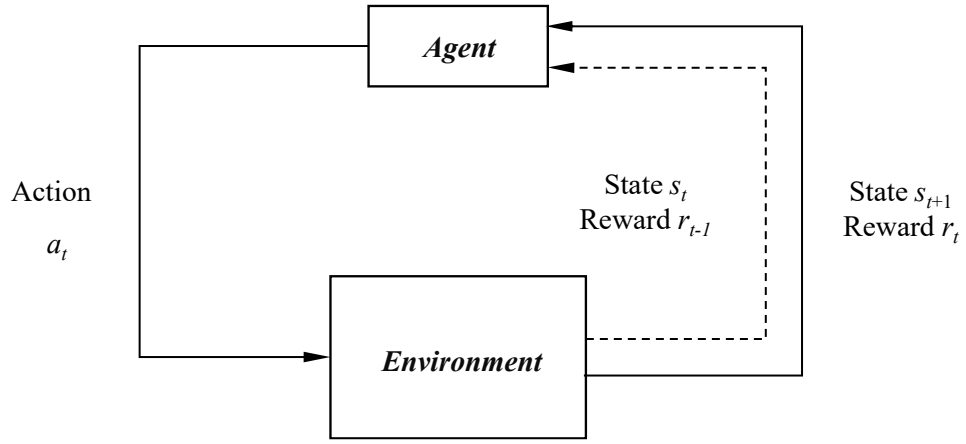


Figure 2.8 RL setup

The interaction between the agent and the environment in a standard RL setup is shown in Figure 2.8. To set up the RL environment, the dynamic system model is first discretized, and then the environment is formulated as a Markov decision process (MDP) model by the four-tuple (S, A, P, R), where the Markov property enforces that the future states depend solely on the current state and action.

S is the set of states, including the system state x in full state control, and observation state y for partial observable state control. A is the action space that serves as the control input. $P(s_{t+1}|s_t, a_t)$ is the transition probability distribution. R is the reward function.

For the agent-environment interaction, the agent at each time step t observes the state $s_t \in S$ and selects an action $a_t \in A$ based on policy π and receives the reward $r_t \in R\ (s_t, a_t)$, then ends in a new state s_{t+1} with the transition probability $P(s_{t+1}|s_t, a_t)$. The reward r_t quantifies the immediate efficacy of a_t and enables policy improvement through experience.

The RL agent aims to maximize the discounted future return H_t:

$$H_t = \sum_{t=t_0}^{\infty} \gamma^{t-t_0} r_t, \tag{2.1}$$

where $\gamma \in [0, 1]$ is the discount factor balancing immediate versus future rewards. The optimal policy π^* is solved as follows:

$$\pi^* = \arg \max_{\pi} \mathbb{E}[H_t], \tag{2.2}$$

where $\mathbb{E}[H_t]$ is the expectation of the discounted future rewards. For systems with limited sensing, the MDP generalizes to a partially observable MDP where partially observed states $o_t = f_{\text{obs}}(s_t)$ replace full states. The time horizon in Equation 2.1 can also be set to be finite to achieve optimal reward in a given finite time horizon T ($\sum_{t=t_0}^{T} \gamma^{t-t_0} r_t$).

When the RL framework is applied to the active control problem, the negative cost function can serve as the reward serves, and the aim is to control the vibration response due to the excitations to meet certain control criteria. For each state, the expected reward could represent future control performance. Thus, instead of obtaining an analytical solution or numerically searching for the optimal strategies, the control policy in RL is learned through interaction without the need for a full dynamic model of a structure.

The following subchapters present the literature review on different algorithms to determine the optimal policy π^* to maximize the return H_t.

2.3.3 Theoretical foundations of RL

The theoretical foundations of RL trace back to Bellman's dynamic programming (DP) (Bellman, 1957a) and the MDP framework (Puterman, 1994). Originally, RL algorithms generally fall into three categories (Bertsekas and Tsitsiklis, 1995): value iteration (VI) algorithms that iteratively estimate value functions to derive policies implicitly, policy iteration (PI) algorithms that directly optimize parametric policy representations, and temporal difference (TD) learning methods. Figure 2.9 illustrates the classical RL algorithms discussed in this subchapter.

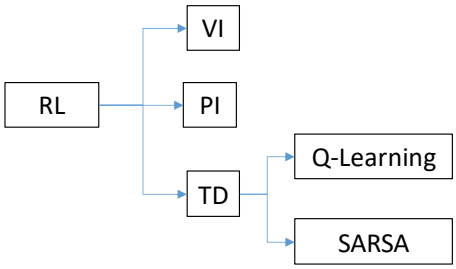


Figure 2.9 Classical RL algorithms

2.3.3.1 VI

The core idea of VI is to apply the Bellman optimality operator iteratively to derive implicit policy through greedy optimization (Bellman, 1958). The value of the state equals the next step reward plus the value of the next state, which is the foundation for modern value-based methods. $V(s_t)$, the value of state s_t, is initialized arbitrarily for all states and repeated:

$$V_{k+1}(s_t) = \max_{a_t} \mathbb{E}[r_t + \gamma V_k(s_{t+1})|s_t, a_t] \tag{2.3}$$

until convergence to compute the optimal value function $V^*(s)$:

$$V^*(s_t) = \max_{\pi} \mathbb{E}[H_t|s_t] \tag{2.4}$$

The optimized policy is extracted greedily:

$$\pi^*(s_t) = \arg\max_{a_t} \mathbb{E}[r_t + \gamma V^*(s_{t+1})|s_t, a_t] \tag{2.5}$$

The VI method guarantees convergence to V^* but requires knowledge of the transition dynamics $P(s_{t+1}|s_t, a_t)$, which makes it suitable for problems where accurate environment models are available. In case the accurate models are unattainable, it needs to learn the NN model of the environment to perform evaluation. Using the learned model for predictions is classified as model-based RL. In contrast to the model-based structural control that requires structural dynamic models, model-based RL uses the NN model to represent the environment (or structural dynamics).

2.3.3.2 PI

The core idea of PI is to evaluate and improve an explicit policy alternately until convergence to π^* (Howard, 1960). Unlike VI, PI directly optimizes the policy rather than the value function. PI is a two-phase process, and the first is policy evaluation. For a fixed policy π, the value $V^{\pi}(s_t)$ is computed by solving the Bellman expectation equation:

$$V^{\pi}(s_t) = \mathbb{E}[r_t + V^{\pi}\gamma(s_{t+1})|s_t], \tag{2.6}$$

through iterative implementation until $\|V_{k+1}^{\pi} - V_k^{\pi}\| < \epsilon$, ϵ is the loss defined. Then, the policy is updated greedily with respect to V^{π}:

$$\pi_{\text{new}}(s_t) = \arg\max_{a_t} \mathbb{E}[r_t + \gamma V^{\pi}(s_{t+1})|s_t, a_t] \tag{2.7}$$

VI focuses on learning a better representation of cost estimations based on given states or actions and uses it to determine actions with an implicit policy, such as a greedy policy; whereas PI is used to directly update the explicit policy function. PI typically converges faster than VI but shares the same model-based limitation.

The main advantage of DP is the guarantee of convergence and lower computation cost. Following the foundational work on VI and PI, the field of RL evolved to address scenarios where the state transition probability $P(s_{t+1}|s_t, a_t)$ is unavailable. This progression reflects the advance from model-based RL to model-free RL, which is the foundation of the algorithms used in the following chapters.

2.3.3.3 TD learning

TD learning is a commonly used model-free value-based algorithm (Sutton, 1988), which bridges the gap between DP and Monte Carlo methods. TD has the advantage of learning from incomplete episodes; it is the central idea of many DRL methods and has many variations. TD methods learn directly from sampled state reward trajectories (experience). The TD algorithm updates value estimates using a bootstrapping approach:

$$V(s_t) \leftarrow V(s_t) + \alpha[r_t + \gamma V(s_{t+1}) - V(s_t)], \tag{2.8}$$

where α is the learning rate hyperparameter. The TD error term $\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t)$ captures the discrepancy between the current and predicted values. While the TD algorithm in Equation 2.8 provides a value estimation method, it does not directly select actions. Using TD learning to learn an optimal policy is also known as TD control, and two ways are often used for update: on-policy and off-policy.

2.3.3.4 Q-learning (off-policy)

The development of Q-learning (Watkins and Dayan, 1992) marked a breakthrough in model-free RL control by introducing an off-policy approach to directly estimate action value Q. The update rule

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)], \qquad (2.9)$$

allows the agent to learn about the optimal policy while following an exploratory behavior policy. At each time step, the action is chosen greedily with respect to the Q values and does not require any pre-defined policy.

$$a_t = \arg\max_a Q(s_t, a), \qquad (2.10)$$

The Q-value function is updated afterward by Equation 2.10, where the next step action a_{t+1} is determined by estimating the maximum value of $Q(s_{t+1}, a_{t+1})$ independently of the previous policy; this approach is known as off-policy learning. The Q values could be estimated through the NN, which are elaborated in subchapter 2.3.6. As Q value estimation becomes more accurate through training, the chosen action can lead to more returns.

2.3.3.5 SARSA (on-policy)

Unlike Q-learning, SARSA (Rummery and Niranjan, 1994) is another well-known TD control algorithm that uses on-policy estimation update instead. The main difference is that the following action a_{t+1} is picked according to the current policy. The update rule of state-action value Q:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)], \quad (2.11)$$

The actions in SARSA are typically selected using an ϵ-greedy policy:

$$a_t = \arg\max_a Q(s_t, a), \text{ (with probability } 1 - \epsilon) \quad (2.12a)$$

$$a_t = \text{Random action, (with probability } \epsilon) \quad (2.12b)$$

Generally, on-policy provides more stable learning but is considered less sample efficient than off-policy in some tasks. Because no model is needed for action optimization, the class is called model-free RL.

Table 2.1 Comparison of classical RL algorithms

Algorithm	Type	Policy	Convergence	Sample Efficiency	Remarks
VI	Model-based RL	Implicit	Guaranteed	High	Linear systems
PI	Model-based RL	Explicit	Guaranteed	High	Interpretable policies
Q-learning	Model-free RL	Implicit (Off-policy)	Guaranteed (tabular)	Low	Offline optimization
SARSA	Model-free RL	Explicit (On-policy)	Probabilistic	Medium	Safe online adaptation

Table 2.1 compares the reviewed classical RL algorithms, progressing from model-based dynamic programming approaches (VI and PI) to model-free TD methods (Q-

learning and SARSA). The analysis reveals fundamental trade-offs between computational efficiency, convergence guarantees, and real-world applicability. No algorithm fits all cases, and each has its advantages. model-based RL methods (VI or PI) provide mathematically rigorous solutions for systems with well-characterized dynamics but become impractical for complex, high-dimensional structures. Model-free RL algorithms (Q-learning or SARSA) enable adaptive control without explicit dynamic knowledge but require careful consideration of exploration safety and sample efficiency.

While these methods are not directly incorporated in the following studies, the DRL algorithms adopted in all the chapters are developed based on these concepts.

2.3.4 RL applications in optimal structural control

RL, NN, and other AI methods are biologically inspired. Therefore, control through RL is closer to the computing and problem-solving strategies in nature. The control problem itself is a kind of inverse problem that human and animal brains deal with (Waszczyszyn, 2010). For example, when the brain controls the functioning organs, the forward problem sends a known signal to the organ and observes the response. The inverse is to determine the proper signal for the desired response.

RL and control systems share many things in common. The agent observes the environment and generates actions similar to the controller operation, where the actuator produces control forces to the plant based on the sensors' feedback. Thus, a control system can naturally be translated into an RL problem, where the effect of tuning the gain and parameters can be substantially reduced by constructing a proper reward representation in RL training. Moreover, for control problems that require complex, nonlinear control architectures, traditional model-based methods are often difficult to deal with. RL offers an end-to-end solution from data collected through system interactions.

Previous studies on finding model-free optimal control strategies for structural vibration control have been mostly conducted on solving ARE through PI. This data-driven technique directly finds the optimal policy from its own experience with the

environment and does not require knowledge of the full system dynamics. Vrabie et al. (2009) used integral RL to update the control policy for a full-state feedback system. The PI of integral RL is achieved by solving the least-square problem while integrating the target function along the state trajectory. The same approach has also been applied to output feedback problems (Zhu et al., 2015) and tracking control problems (Modares and Lewis, 2014). However, this method is unsuitable for high-dimensional states because it calculates the integrated quadratic cost for every step. Vibration problems in the civil engineering field are typically high-dimensional, considering the observation states, which adds difficulty to traditional policy iteration and value iteration algorithms, such as step-by-step least-square calculation. Hence, conventional RL algorithms are unsuitable for such problems. Consequently, an urgent need arises to develop a novel approach that can utilize the learning ability of NN controllers while attaining performance close to or even better than classical optimal control methods.

2.3.5 Deep learning

The development of DL in the past decade has dramatically enhanced the ability of RL. DL is a branch of ML that uses DNN to learn the representation of data. In contrast to traditional ML, no hand-crafted features are required. The early approach of DL was conceived in the 1970s (Werbos, 1974). DL scaled up the sensory input to be high-dimensional, and more complex control logic could be incorporated rather than using insufficient manually developed models. The earliest work of combining DL with RL could be traced back to the 1990s. Tesauro (1994) demonstrated a superior NN learned strategy for playing backgammon. However, the convergence of DL models in nonlinear RL scenarios is complex. Recently, DL has been applied to image classification, speech recognition, and many other applications and achieved remarkable results. In the civil engineering field, DL has been applied to system identification, damage detection, and fragility assessment (Salehi and Burgueno, 2018). A more comprehensive overview of DL can be found in LeCun et al. (2015) and Goodfellow et al. (2016). Using DNNs in DRL enables high-capacity function approximation that can solve human-level control problems.

2.3.6 Advanced DRL architectures

The recent success in DRL has proven the ability to find better solutions in complex environments. The achievements include the success of training agents in playing Atari games from raw pixels (Mnih et al., 2013; Mnih et al., 2015), the well-known Alpha Go that outperformed humans (Silver et al., 2016), and learning high-dimensional motion controllers. A more detailed survey on DRL algorithms and applications can be found in Li (2017). DRL has increased potential for dealing with structural control problems because vibration problems in the civil engineering field are usually high-dimensional, considering the observation states. DRL provides a new way of determining the control policy directly by interacting with the environment without solving the ARE. Moreover, the nonlinear techniques inherited from DL parts offer a new advantage to the nonlinearity problems in structural control.

2.3.6.1 Actor–critic architecture

Most of the DRL algorithms adopted in this thesis used the actor–critic architecture, which is briefly introduced in this part. Figure 2.10 shows a typical architecture of the actor–critic architecture. The two models could optionally share parameters, and the formulation is very flexible, which makes it capable of learning policies fast under complex action spaces.

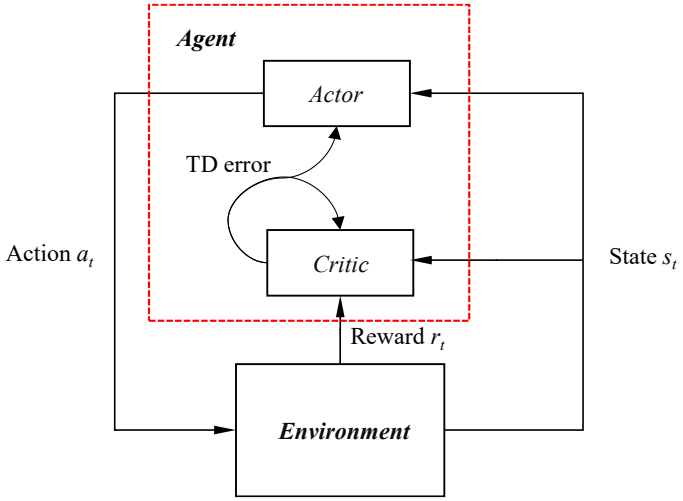


Figure 2.10 Actor–critic architecture

This framework combines two NNs: (1) a critic that evaluates action quality through value function $V(s)$ approximated from VI or through the action–value function $Q(s,a)$ approximated from Q-learning, and (2) an actor that directly parameterizes and optimizes the policy under the guidance of the critic. The optimal policy is the one that maximizes the value function $V_\phi(s)$ or the action–value function $Q_\phi(s,a)$. ϕ is the parameter of the critic network. When estimating the value function, the critic network's objective is to minimize the TD error through the following loss function:

$$Loss_{critic} = \mathbb{E}_{(s_t,r_t,s_{t+1})\sim\mathcal{D}}[(r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t))^2], \quad (2.13)$$

where $\mathcal{D}$ denotes the experience replay buffer. The mean squared error (MSE) formulation enables stable learning of value estimates by reducing variance through bootstrapping (using current value estimates) and by providing smooth gradient signals for policy improvement.

The actor parameter θ is updated using gradient ascent on the expected return $J(\pi_\theta)$:

$$\nabla_\theta J(\pi_\theta) = \mathbb{E}\left[\sum_{t=0}^{T} \nabla_\theta log\pi_\theta(a_t|s_t) \cdot (r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t))\right], \quad (2.14)$$

$$\theta_{k+1} = \theta_k + \alpha\nabla_\theta J(\pi_\theta)|\theta_k. \quad (2.15)$$

The critic does not directly modify the policy but provides a scaling factor that weights how much to reinforce/discourage specific actions.

2.3.6.2 Policy gradient

Policy gradient is another concept in DRL that models and optimizes the policy directly and works in many newly proposed DRL algorithms in recent years (Sliver et al., 2014). The general framework is described herein. Considering a stochastic, parameterized policy π_θ, the aim is to maximize the expected return $J(\pi_\theta) = \mathbb{E}[H(\tau)]$, where τ is the trajectory, and $H(\tau)$ is the expected return along the trajectory. If the policy is optimized by gradient ascent (Equation 2.15), the simplest form of the policy

gradient can be expressed as follows:

$$\nabla_\theta J(\pi_\theta) = \mathbb{E}[\sum_{t=0}^{T} \nabla_\theta log\pi_\theta(a_t|s_t)H(\tau)]. \quad (2.16)$$

The algorithms using this method for policy optimization are called the policy gradient algorithms, ranging from the simple Vanilla policy gradient to the more advanced algorithms (Schulman et al., 2015). This form is not fixed and can also be expressed by other forms, if $H(\tau)$ is replaced by other choices such as Q-value and advantage value. In-depth discussion can be found in Schulman et al. (2017). For example, using the advantage estimate $Adv(s_t, a_t)$

$$Adv(s_t, a_t) = Q(s_t, a_t) - V(s_t), \quad (2.17)$$

$$\nabla_\theta J(\pi_\theta) = \mathbb{E}[\sum_{t=0}^{T} \nabla_\theta log\pi_\theta(a_t|s_t)Adv(\tau)]. \quad (2.18)$$

Policy gradient is naturally more helpful in a continuous action space because the need to search for an infinite number of actions is avoided. The policy parameter θ is moved toward the gradient direction. Despite many policy gradient algorithms, in general, off-policy is more sample efficient compared with on-policy because they do not require whole trajectories for an update.

The variations of the actor-critic formulation lead to different algorithms (Konda and Borkar, 1999; Groundman et al., 2012). Comparing the actor–critic methods with the classical NN controllers reviewed in Subchapter 2.2.2, the critic and the emulator share something in common because they are both used in training the policy (i.e., actor) network for control. However, their basic concepts are totally different. The critic aims to learn a representation of the action–state value rather than the model, so it avoids the modeling error encountered in the emulator–controller structure. Moreover, in DRL, the critic and actor networks are typically trained together.

The development of the algorithms adopted in this thesis is reviewed in the following parts. The implementation details are provided in later chapters.

2.3.6.3 Development of DRL algorithms

Novel DRL algorithms demonstrate the possibilities of DRL strategies to tackle complex human-level problems. For example, deep Q-Network (DQN) (Mnih et al., 2013) establishes the base of DRL as a combination of Q-learning and DNN, in which the Q value estimation is done on a DNN. However, DQN uses discrete actions and is not suitable for continuous control problems. Newly developed algorithms, such as deterministic policy gradient (DPG) (Sliver et al., 2014), trust region policy optimization (TRPO) (Schulman et al., 2015), and proximal policy optimization (PPO) (Schulman et al., 2017), have demonstrated notable performance in robotic control tasks for an action–state domain. PPO is a policy gradient method that simplifies and improves upon previous algorithms by using a clipped objective function to constrain policy updates and ensure more stable and reliable learning.

From the point of view of policy function, DRL algorithms could generally be classified into stochastic and deterministic policies. The function $\pi(a|s)$ is modeled as a probability distribution over actions in the stochastic policy. Examples of stochastic algorithms include asynchronous advantage actor–critic (Mnih et al., 2016), TRPO, PPO, and soft actor–critic (SAC) (Haarnoja et al., 2018).

Deep deterministic policy gradient (DDPG), a combination of DPG and DQN, was introduced by Lillicrap et al. (2015). The advantages of both algorithms include that the Q-learning makes training stable, and the deterministic policy extends discrete action space to continuous space. Compared with batch algorithms, the DDPG algorithm continuously improves the policy as it explores the environment. Using a deterministic policy network as an actor, this algorithm can provide continuous control and be applied in robotics, quadrotors, and autonomous vehicle fields. Based on the DDPG, a couple of improved algorithms are presented by adding different techniques to enhance the DDPG. Distributed distributional DDPG (D4PG) runs in a distributional fashion (Barth-Maron et al., 2018). Multi-agent DDPG (Lowe et al., 2017) involves multiple agents for coordination in changing environments or the need for interaction between agents.

Twin delayed deep deterministic (TD3) (Fujimoto et al., 2018) adds three more

tricks to improve the DDPG. Double Q-learning is introduced in TD3 to prevent overestimating the Q network by adding a separate critic network and using the minimum estimation; the policy is updated at a lower frequency than the Q-function; third, a small amount of clipped random noise is averaged among the mini batches and added to the selected action when updating the value function to make the regularization smoother.

SAC (Haarnoja et al., 2018) extends the DDPG style to stochastic policy optimization. The key feature of SAC is entropy regularization, by which the trade-off between expected return and entropy is trained to be maximized. Entropy represents the randomness of the policy. Increasing entropy leads to more exploration and prevents the policy from converging to an undesired local optimum. SAC also uses the double-Q trick, but the target update is from the current policy.

Figure 2.11 presents a non-exhaustive taxonomy of standard DRL algorithms. The red ones are used later in different chapters of this thesis. Most of them have been introduced above. Model-based RL with model-free fine-tuning (MBMF) combines a learned environment model for efficient planning with model-free methods like TRPO to refine the policy (Nagabandi et al., 2018). Model-based value expansion (Feinberg et al., 2018) uses a dynamic model to simulate the short-term horizon and Q-learning for long-term estimation. Imagination-augmented agents (Weber et al., 2017) embed the planning procedure into the policy and train the agent to learn when and how to use the plans.

Despite numerous DRL algorithms, none work for all problems. Different types of problems lead to different state and action spaces. Because the aim of structural control is to determine specific control forces from sensor feedback, a policy that operates in continuous action space is chosen in this thesis.

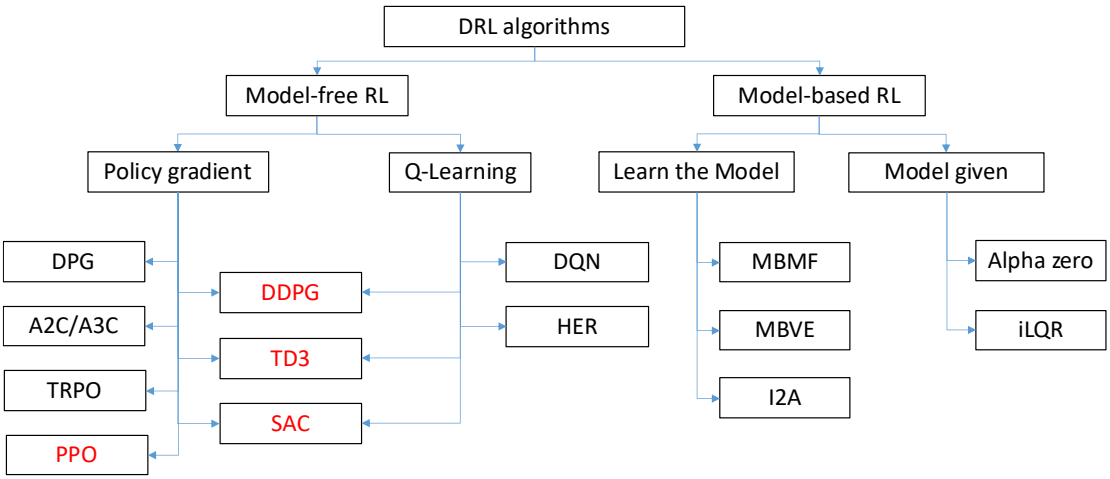


Figure 2.11 DRL algorithm taxonomy

2.3.7 DRL applications in control

Duan et al. (2016) reviewed a couple of benchmarks for DRL continuous control tasks, including cart-pole balancing, mountain car, double inverted pendulum balancing, and some locomotion tasks. However, most of the control-related DRL research is based on either OpenAI gyms or the robotic field (Nagabandi, 2020). The applications in structural vibration control are rare, and that is one of the motivations of this thesis. The feasibility of using RL in structural vibration control was validated on a full-scale tensegrity structure in the earlier days (Adam and Smith, 2008). DQN successfully generated discrete control signals in previous applications. A discrete controller was trained by DQN to suppress the vibration of an SDOF structure under earthquake excitations (Rahmani et al., 2019). Similar to the model-based controller that suppresses vibration responses based on some control criteria, for each state, the expected reward represents future control performance. Thus, instead of obtaining an analytical solution or numerically searching for the optimal strategies, the policy in DRL is learned through interaction without the need for a complete dynamic model.

A few more studies have explored the implementation of DRL under different conditions. By considering uncertainties in seismic inputs, the Q-learning RL algorithm was utilized to tune the gains of a fuzzy proportional derivative controller to achieve the stability of high-rise building vibration (Khalatbarisoltani et al., 2019). An actor-only DRL method was used for semi-active control of a simply supported aluminum

beam (Pisarski and Jankowski, 2023). This method showed low computational complexity in real time (Panda et al., 2024). In addition, PPO was used to mitigate the control–structure interaction effects of AMD systems in a wind-excited 76-story benchmark building (Yao et al., 2024). Moreover, gradient descent algorithms were used in some research for active control (Eshkevari et al., 2023; Nagendra et al., 2017).

2.3.8 Pre-training

Despite growing research in applying DRL to structural vibration control, to the author's best knowledge, real-world implementations of model-free structural vibration control using DRL have not been reported in the literature. One major obstacle is that DRL typically requires a long training time before achieving satisfactory performance (Schulman et al., 2015) and is highly unstable in training processes (Mysore et al., 2021). One potential solution is to incorporate pre-training into the NN controller (Hester et al., 2018), which can substantially reduce the required training time and make experimental or real-world implementations more feasible. In DRL, pre-training the agent network model improves sampling efficiency and increases training stability (Cruz Jr et al., 2017).

The pre-training methods can be categorized into two approaches: pre-training a state dynamic prediction model and pre-training an initial policy directly. The former improves the agent evaluation, whereas the latter reduces the training time and achieves reasonable performance more efficiently (Anderson et al., 2015; Subramanian et al., 2016). Policy pre-training provides less understanding of the target system's nature, but it is more straightforward and ensures the initial performance, which is suitable when the training is done on an actual structure. A common approach for policy initialization is IL, which is particularly useful when a demonstrator is available. The choice of demonstrator is flexible and can come from humans or other existing policies. It could provide an initialized policy for the DRL training because it can mimic a workable policy in the experiment and save the training time required to reach a reasonable performance (Ross et al., 2011). Although it cannot capture the inherent system nature, it is a straightforward way to ensure a guaranteed initial performance.

2.4 Remarks on Literature Review

Civil engineering structures present unique control challenges that distinguish them from conventional mechanical systems. Three fundamental limitations of conventional active control emerge as particularly consequential:

First, the massive scale of civil infrastructures necessitates correspondingly large actuation forces and imposes stringent power requirements on active systems. Second, inherent modeling uncertainties persist due to the imperfect characterization and modeling of structural properties and the stochastic nature of environmental excitations. Third, measurement constraints arise from practical implementations because of sparse, noise-contaminated sensor data (typically accelerometers or strain gauges) that cannot fully capture the system state. This partial observability is compounded by the reality that critical control variables often cannot be directly measured and require state estimation or indirect inference. These challenges require control algorithms capable of determining precise control forces, handling modeling uncertainties, and managing incomplete measurements to optimize power usage and ensure efficient control performance. They collectively motivate the model-free framework developed in subsequent chapters based on DRL.

Several past studies have been conducted in recent years to investigate the potential of DRL for structural vibration control. These studies have demonstrated the efficacy of DRL in this context. However, they have not yet developed a systematic approach. Further research is needed to explore how DRL can be applied to different systems in different working conditions to improve control results. Concurrently, the limitations of DRL must be acknowledged. Additional research is necessary to identify effective countermeasures and solutions for leveraging its advantages in actual structural vibration control applications. This research area deserves further investigation.

CHAPTER 3

DRL-based Model-free Optimal Active Vibration Control for Linear Systems

3.1 Introduction

As mentioned in Chapters 1 and 2, structural vibration control aims to suppress the vibrations of civil and mechanical structures induced by dynamic loads. A variety of structural control systems have been proposed and built to alleviate structure responses under various dynamic loads (such as seismic and wind loads), particularly when the original structural resistance is insufficient. Although passive and semi-active systems are cost-effective and reliable in operation, their performance is limited because they cannot adapt to different excitations. Active control systems can provide high-performance structural response reduction by calculating the desirable actuator control force according to real-time observations.

A general review of active control algorithms is presented in Chapter 2. Most optimal control algorithms necessitate complete system dynamics knowledge, and their control performance is primarily dependent on the accuracy of model parameters. In contrast to these model-based algorithms, an NN controller offers a model-free

control solution that is more versatile in design and more convenient in applications. NN controllers can eliminate the need for analytically developing control algorithms, which is often difficult for structures with unknown dynamics and strong nonlinearity. However, their control performance in early applications in the literature is mostly not comparable with model-based optimal control methods because the error functions used for training can only represent simple control schemes. Consequently, an urgent need arises to develop a novel approach that can utilize the learning ability of NN controllers while attaining performance close to or even better than classical optimal control methods.

Research on achieving optimal structural vibration control performance using a model-free method is still absent in the literature reviewed in Chapter 2. Existing methods exhibit at least one of the following limitations: (1) the control performance is nonoptimal, (2) the method cannot be applied to MDOF systems, or (3) the method cannot be extended to partially observed systems. This chapter proposes a novel active vibration control strategy based on DRL to fill these existing gaps. This control strategy can be applied to various linear vibration control problems because its model-free nature does not require any knowledge of structural dynamics. Compared with traditional active control based on NN, the current chapter has demonstrated the ability to find optimal strategies that traditional NN methods cannot achieve. The major contributions of this chapter are summarized as follows.

In fully observable SDOF and MDOF systems, the proposed model-free NN controller is compared with a model-based optimal LQR controller. Both the control performance under free and random vibrations and the robustness against measurement noise are examined.

The same strategy can be directly applied to a partially observed system. The corresponding control performance is compared with a full-state controller and an output feedback controller under different excitations.

The remainder of this chapter is organized as follows. The vibration problem is formulated in subchapter 3.2. Then, the settings of two traditional model-based methods, namely, LQR and output feedback controllers, are presented as the baseline

cases for comparison. The RL setup and the proposed DRL controller are presented in subchapter 3.3. The details of implementation and simulation results under various test conditions are shown in subchapter 3.4. Finally, subchapter 3.5 provides a summary of this chapter.

3.2 Control Problem Formulation

The vibration control of a linear continuous-time system is introduced in this subchapter. The state-space representation can be written as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{B}_g\ddot{x}_g, \tag{3.1}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x}, \tag{3.2}$$

where **x** is the system state vector, **y** is the system output, **u** is the control action, $\ddot{x}_g$ is the excitation, **A** is the system matrix, **B** is the input matrix, $\mathbf{B}_g$is the excitation input matrix, and **C** is the output matrix. We usually assume that pair (**A**, **B**) is controllable, and pair (**A**, **C**) is observable.

Two scenarios of the feedback control problem are considered. In the first scenario, the full knowledge of the system state vector x is measurable. In the second scenario, only a part of the state vector is observed. Two model-based controllers have been selected to compare with the proposed model-free controller. The block diagrams of the traditional optimal controllers in these two scenarios are illustrated in Figure 3.1, in which the control forces are determined based on the state vector x and the output vector **y**, respectively, in these two controllers.

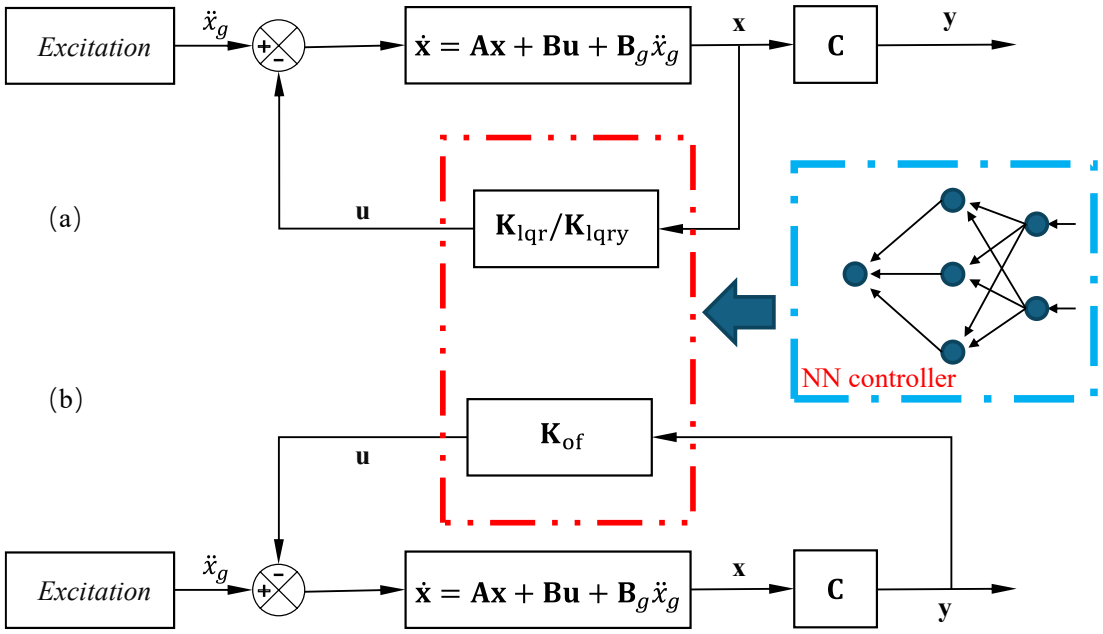


Figure 3.1 Block diagrams of classical control: (a) state-space model with LQR and LQRy controllers and (b) state-space model with output feedback controller

3.2.1 Full-state feedback control

For the full-state observable system, the classical active LQR controller is selected for comparison with the proposed novel DRL-based controller. The control action $\mathbf{u}$ is given as

$$\mathbf{u} = -\mathbf{K}_{\mathrm{lqr}}\mathbf{x}, \tag{3.3}$$

$\mathbf{K}_{\mathrm{lqr}}$ is the optimal feedback gain that minimizes the quadratic cost

$$J_{\mathrm{x}} = \int_0^{\infty} (\mathbf{x}^T\mathbf{Q}_{\mathrm{x}}\mathbf{x} + \mathbf{u}^T\mathbf{R}\mathbf{u})dt, \tag{3.4}$$

where $\mathbf{Q}_{\mathrm{x}}$ and $\mathbf{R}$ are the state and control force weight matrices, respectively; $\mathbf{x}^T\mathbf{Q}_{\mathrm{x}}\mathbf{x}$ stands for the state cost; and $\mathbf{u}^T\mathbf{R}\mathbf{u}$ is the force cost. The optimal feedback gain $\mathbf{K}_{\mathrm{lqr}}$ is determined by

$$\mathbf{K}_{\mathrm{lqr}} = \mathbf{R}^{-1}\mathbf{B}^T\mathbf{P}, \tag{3.5}$$

where **P** is the solution for ARE.

$$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} - \mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{P} + \mathbf{Q}_x = \mathbf{0}. \tag{3.6}$$

P can be determined by solving the ARE matrix. Using Equation 3.5 obtains the optimal gain matrix $\mathbf{K}_{lqr}$. The system response with the LQR controller can now be described by

$$\dot{\mathbf{x}} = (\mathbf{A} - \mathbf{B}\mathbf{K}_{lqr})\mathbf{x}. \tag{3.7}$$

If the quadratic cost function is calculated on the basis of the output vector **y**, instead of the vector **x**

$$J_y = \int_0^\infty (\mathbf{y}^T\mathbf{Q}_y\mathbf{y} + \mathbf{u}^T\mathbf{R}\mathbf{u})dt = \int_0^\infty (\mathbf{x}^T\mathbf{C}^T\mathbf{Q}_y\mathbf{C}\mathbf{x} + \mathbf{u}^T\mathbf{R}\mathbf{u})dt, \tag{3.8}$$

The corresponding algorithm is often termed LQRy. Note that the control action $\mathbf{u} = -\mathbf{K}_{lqry}\mathbf{x}$ is still based on the full-state vector, where $\mathbf{K}_{lqry}$ is the optimal feedback gain that minimizes the quadratic cost in Equation 3.8. Accordingly, the ARE is transformed into

$$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} - \mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{P} + \mathbf{C}^T\mathbf{Q}_y\mathbf{C} = 0. \tag{3.9}$$

The block diagram corresponding to the LQR/LQRy controller is presented in Figure 3.1a.

3.2.2 Output feedback control

It is usually impractical to observe a full-state vector in a large-scale structure. Therefore, the second scenario represents a practical condition when the state vector is only partially observable. The output feedback controller is adopted, in which the control action is determined directly based on the output measurement,

$$\mathbf{u} = -\mathbf{K}_{of}\mathbf{y}. \tag{3.10}$$

The gain matrix $\mathbf{K}_{\text{of}}$ can be calculated as

$$\mathbf{K}_{\text{of}} = \mathbf{R}^{-1}\mathbf{B}^T\mathbf{P}\mathbf{S}\mathbf{C}^T(\mathbf{C}\mathbf{S}\mathbf{C}^T)^{-1}, \tag{3.11}$$

where matrices $\mathbf{P}$ and $\mathbf{S}$ are the solutions for the following Lyapunov equations:

$$(\mathbf{A}-\mathbf{B}\mathbf{K}_{\text{of}}\mathbf{C})^T\mathbf{P}+\mathbf{P}(\mathbf{A}-\mathbf{B}\mathbf{K}_{\text{of}}\mathbf{C})+\mathbf{C}^T{\mathbf{K}_{\text{of}}}^T\mathbf{R}\mathbf{K}_{\text{of}}\mathbf{C}+\mathbf{C}^T\mathbf{Q}_{\text{y}}\mathbf{C}=0, \tag{3.12}$$

$$(\mathbf{A}-\mathbf{B}\mathbf{K}_{\text{of}}\mathbf{C})^T\mathbf{S}+\mathbf{S}(\mathbf{A}-\mathbf{B}\mathbf{K}_{\text{of}}\mathbf{C})+\mathbf{X}=0. \tag{3.13}$$

$\mathbf{X}$ is the expected value, i.e.,

$$\mathbf{X}=\mathbb{E}[\mathbf{x}(0)\mathbf{x}^T(0)]. \tag{3.14}$$

Considering that an initial gain matrix iterative solution algorithm can be used to search for the optimal gain matrix $\mathbf{K}_{\text{of}}$, the system response under output feedback control is written as

$$\dot{\mathbf{x}}=(\mathbf{A}-\mathbf{B}\mathbf{K}_{\text{of}}\mathbf{C})\mathbf{x}. \tag{3.15}$$

The block diagram of the output feedback control is presented in Figure 3.1b. Note that the LQR, LQRy, and output feedback controllers represent the classic model-based control algorithms that require the full knowledge of the state space models.

3.3 Vibration Control based on DRL

3.3.1 RL for vibration control

The basic setup and review of RL methods are presented in subchapter 2.3. The RL problem includes five key components: environment, agent, state, action, and reward. When applied to a feedback control system, the environment is equivalent to the plant model; the agent acts as the controller; the state is equivalent to the complete

or partial observation of the environment; the action is the same as that in the control scheme; and the agent interacts directly with the environment by observing the state and performing the action. In addition, a reward is generated in each step to measure the success or failure of the agent's action, and the agent can learn the action policy through experience.

The agent aims to find the optimal policy μ^* to maximize the total discounted reward. If a model-free RL algorithm is adopted, the optimal controllers can be determined without knowledge of system dynamics. The interaction between an agent and the environment in a standard RL setup is already illustrated in Figure 2.8.

To implement the RL for structural vibration control, the dynamic system model is first discretized when designing an RL environment for vibration control. Then, the environment is formulated as an MDP model, in which the future state only depends on the most current state and action (referred to as the Markov property), as discussed in subchapter 2.3.

At each time step t, the environment experiences the state s_t, receives the action a_t based on the policy μ, the corresponding reward is r_t. Then, it ends in a new state s_{t+1}. The total discounted reward from time t is defined as the return H_t, with the discount factor $\gamma \in [0, 1]$.

$$H_t = \sum_{t=t_0}^{\infty} \gamma^{t-t_0} r_t. \tag{3.16}$$

which is also presented as Equation 2.1 in Chapter 2.

If the state is partially observed, then the system is regarded as a partially observable MDP (POMDP). The time horizon can also be set to be finite to achieve the optimal reward in the given steps. The optimal policy μ^* shall maximize the discounted reward.

$$\mu^* = \arg \max_{\mu} \mathbb{E}[H_t]. \tag{3.17}$$

The objective is to suppress vibrations under excitations via the quadratic cost

functions. For each step, the expected reward H_t represents future control performance. Thus, instead of obtaining an analytical solution or numerically searching for optimal strategies, the policy in RL is learned through interaction without requiring a full dynamics model. The step reward for full-state feedback control is defined as the negative quadratic cost for the vibration control problem.

$$r_t = -\mathbf{x}_t^T \mathbf{Q}_{\mathrm{x}} \mathbf{x}_t - \mathbf{u}_t^T \mathbf{R} \mathbf{u}_t. \tag{3.18}$$

The step reward for partial observable state control is

$$r_t = -\mathbf{y}_t^T \mathbf{Q}_{\mathrm{y}} \mathbf{y}_t - \mathbf{u}_t^T \mathbf{R} \mathbf{u}_t. \tag{3.19}$$

Consequently, maximizing the total reward is equivalent to minimizing the quadratic cost function.

3.3.2 DRL-based vibration control strategy

This chapter chooses the DDPG algorithm to train the NN controller. DDPG is a model-free, off-policy, online RL algorithm, as introduced in subchapter 2.3.6. The DDPG algorithm uses a deterministic policy instead of the traditional stochastic policy. Thus, the DDPG agent outputs a deterministic action for each observation state, and the policy of DDPG is modeled as a DNN. Consequently, it can work with continuous action spaces, and this capability is critical for high-dimensional space vibration control.

The actor–critic structure is represented by DNNs in the DDPG agent. The actor outputs the corresponding action to the environment, and the critic approximates the long-term reward based on the observation and action. The expectation Q^{μ} of a state–action pair (s_t, a_t) is

$$Q^{\mu}(s_t, a_t) = \mathbb{E}[r(s_t, a_t) + \gamma Q^{\mu}(s_{t+1}, \mu(s_{t+1}))]. \tag{3.20}$$

The critic network Q^{μ} parameterized by θ^Q is optimized by minimizing loss L as follows:

$$L = \mathbb{E}[\left(Q^{\mu}(s_t, a_t|\theta^{Q^{\mu}}) - Y_t\right)^2], \tag{3.21}$$

where Y_t is the value function target at time t, i.e.,

$$Y_t = r(s_t, a_t) + \gamma Q^{\mu}(s_{t+1}, \mu(s_{t+1})|\theta^{Q^{\mu}}). \tag{3.22}$$

Two significant changes introduced into the Q-learning by DDPG are a replay buffer and two separate networks for critic and actor updates to increase the learning stability. The critic is learned using Bellman's equation, i.e., Equation 3.20 and the actor is updated through gradient ascent. As an off-policy algorithm, DDPG exhibits the advantage of independent exploration from learning. Temporally correlated exploration noise is added to the action during training for better exploration. The flowchart of using DDPG to train the NN controller is presented in Figure 3.2, and the training follows Algorithm 3.1.

Algorithm 3.1: DRL controller training

Initialization step: initial DDPG agent, initial state vector $\mathbf{x}(0)$
repeat
  Observe state s_t and get action $a_t = \mu(s_t)$
  Execute a_t in the environment
  Observe the next state
  Get the reward $r_t(s_t, a_t)$
  If s_{t+1} is terminal, then reset the initial states and start a new episode
  If it is time to update, then
    μ' =DDPG(μ, $[s_t, r_t(s_t, a_t), s_{t+1}]$)
  end if
until convergence

After the training process, the actor network can be used as the NN controller separately to generate control forces by accepting the states as input, similar to a regular controller, as shown in Figure 3.2.

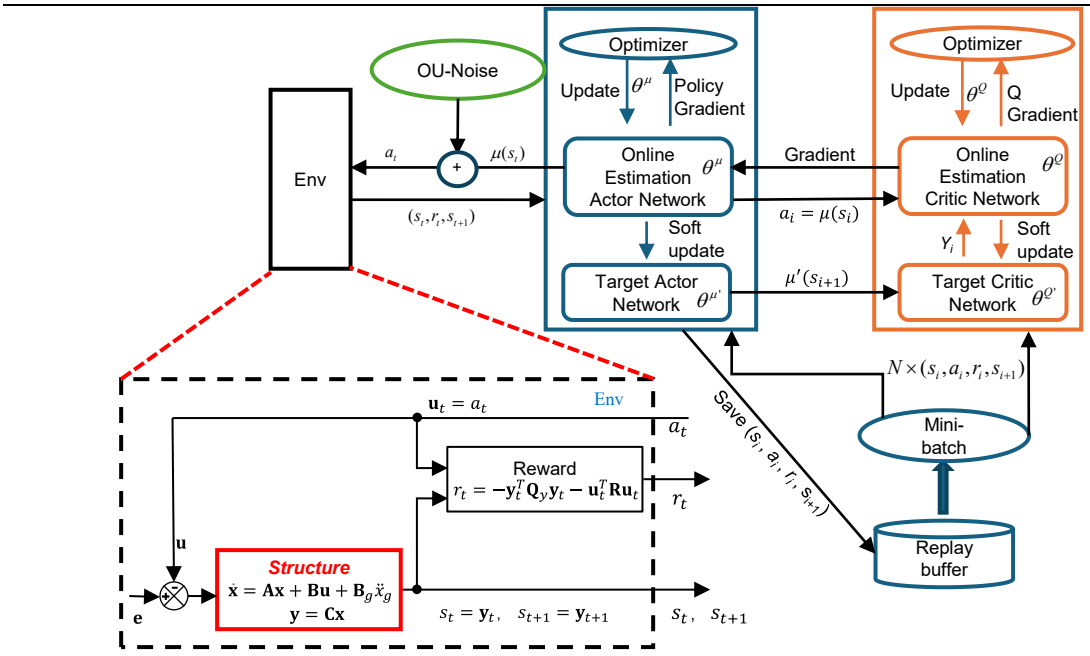


Figure 3.2 Flowchart of using the DDPG algorithm to train a vibration controller

3.4 Simulation Results

In this subchapter, various cases are presented to demonstrate the effectiveness of DRL agents in vibration control. Starting with a simple SDOF model (subchapter 3.4.1), the performance of the DRL-trained NN controller is compared with the LQR control under free and random vibrations. Subsequently, this strategy is extended to a six-story shear building model with an actuator installed in the first story. Full-state control (subchapter 3.4.2) and partially observed state control (subchapter 3.4.3) are examined. Moreover, the robustness of the DRL-based controller is tested by including measurement noise (subchapter 3.4.4).

3.4.1 SDOF system

Consider an SDOF system with the system matrices as follows:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{3.23}$$

The weighting matrices $\mathbf{Q}_x$ and $\mathbf{R}$ in the cost function are selected to be

$$\mathbf{Q}_x = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \mathbf{R} = 1. \tag{3.24}$$

Given an initial displacement and velocity, the SDOF system undergoes free vibration with the control action. 10-s free vibration records under the current agent are used for each training episode. The actor network has one hidden layer with a dimension of 64, and the critic network has two hidden layers, each with a dimension of 64. Both networks are composed of rectified linear unit (ReLU) activations. The selection of hyperparameters in DDPG is discussed in detail in Section 3.6 Appendix. The training progress is shown in Figure 3.3, and the agent reaches the total reward of the LQR controller in less than 500 episodes.

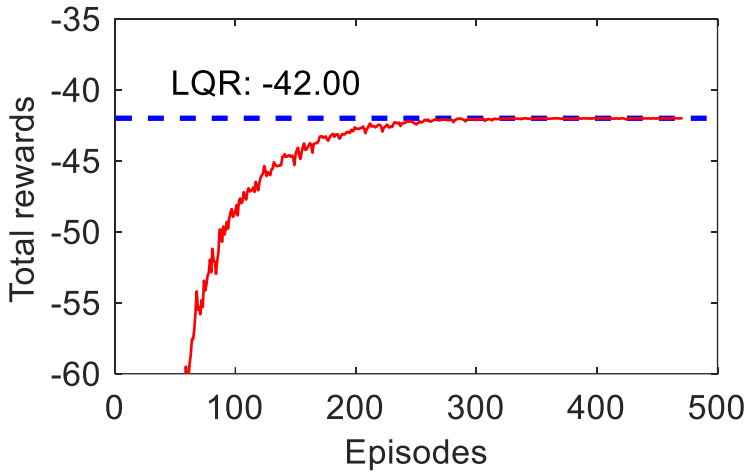


Figure 3.3 Episode reward of the DDPG agent during training for the SDOF system

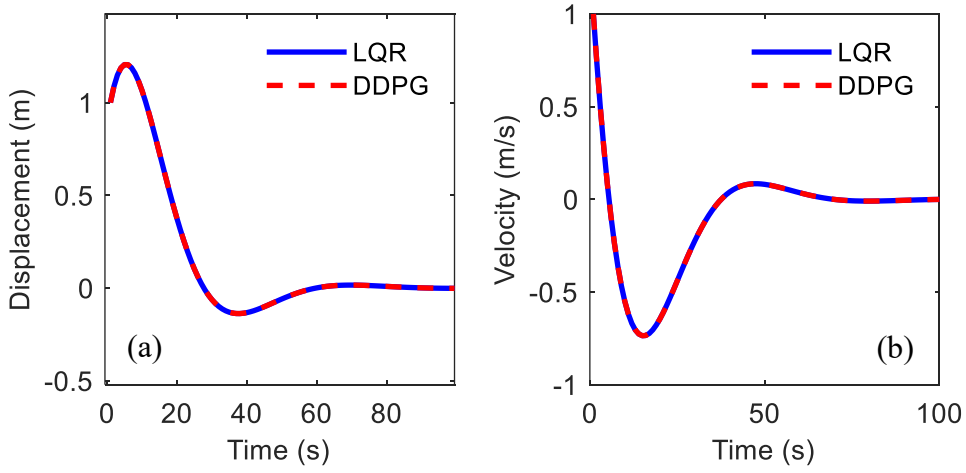


Figure 3.4 Comparison of SDOF free vibration: (a) displacement and (b) velocity

The performance of the LQR controller (model-based) and the DDPG agent (model-free) under free vibrations is compared in Figures 3.4a and 3.4b. Given the initial state [1, 1], the displacement and velocity trajectories of the SDOF system controlled by the DDPG agent and the LQR controller are nearly the same.

Table 3.1 Comparison of cumulative cost under white noise excitations

Duration (s)	Excitation level	DDPG	LQR
500	Low	720.96	720.94
	Middle	902.70	902.73
	High	7.18×10^4	7.18×10^4

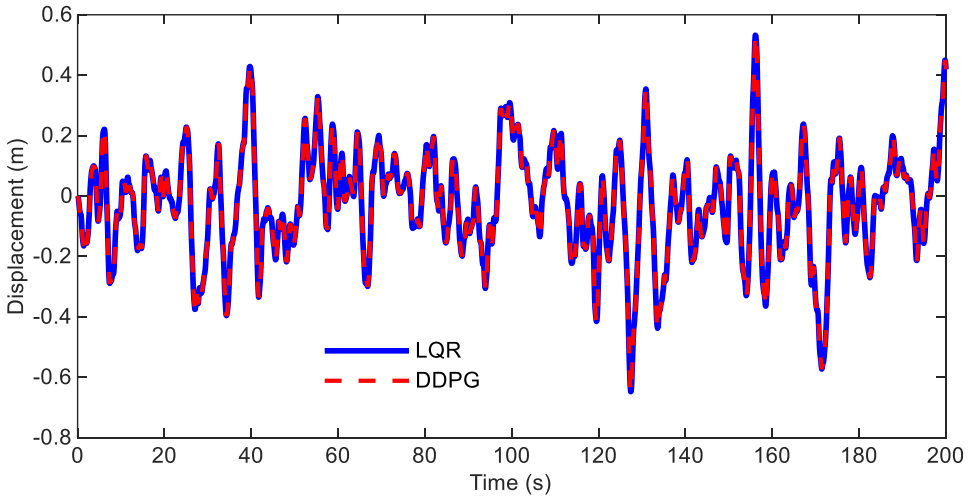


Figure 3.5 Displacement comparison of the SDOF system under random excitations

Although the DDPG agent is only trained on the basis of free vibration, white noise random excitations are also applied to test the performance. Random excitations of three different levels (low, middle and high) are applied to examine if the DDPG agent produces any nonlinear effect that would affect the control performance, where the low, middle and high levels correspond to the power spectral density (PSD) levels

of 0.04, 0.4, and 4.0, respectively, i.e., the PSD levels differ by two orders of magnitude. The cumulative cost for each simulation is provided in Table 3.1. The result of the DDPG controller is still very close to that of the LQR controller. Figure 3.5 presents the displacement comparison in one random excitation case. The performance of the two controllers is generally consistent. Although trained on free vibration only, the model-free DDPG controller can still offer an optimal control performance comparable to the classical LQR controller for the SDOF system under random situations.

3.4.2 Full-state observable MDOF system

Table 3.2 Six-story shear building model properties

Story	Mass ($\times 10^3$ kg)	Stiffness ($\times 10^6$ N/m)
1	25	6.82
2–5	20	6.06
6	15	6.06

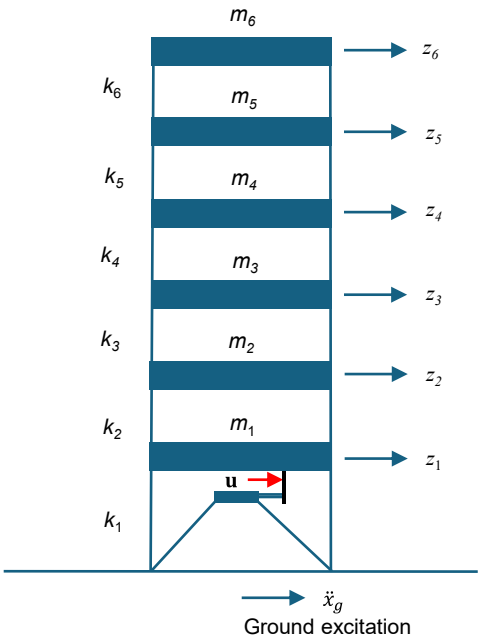


Figure 3.6 Six-story shear building configuration

The proposed control strategy is extended from SDOF to a higher-order MDOF system, namely, a six-story shear building with an actuator installed in the first story, as shown in Figure 3.6. A couple of assumptions are made for this simplified model. The floor is assumed to be rigid without rotation, and the floor mass is modeled as a lumped mass. Table 3.2 provides each floor's structural properties. The damping ratio is set to 5% for all modes. Notably, although the actuator is assumed to be installed in the first story of the building in this simulation case, the presented control strategy can be extended to other actuator locations or even other structures.

Only the in-plane vibration is considered. The corresponding equation of motion can be written as

$$\mathbf{m}\ddot{\mathbf{z}} + \mathbf{c}\dot{\mathbf{z}} + \mathbf{k}\mathbf{z} = -\mathbf{m}\mathbf{b}_g\ddot{x}_g + \mathbf{b}\mathbf{u}, \tag{3.25}$$

where **m**, **c**, and **k** are the mass, damping, and stiffness matrices of the shear building, respectively; $\mathbf{z} = [z_1 \dots z_6]^T$ includes the displacement of each floor; $\ddot{x}_g$ is the ground excitation; **u** is the control force; $\mathbf{b} = [1\ 0\ 0\ 0\ 0\ 0]^T$; and $\mathbf{b}_g = [1\ 1\ 1\ 1\ 1\ 1]^T$.

This equation can transform into a state-space representation as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{B}_g\ddot{x}_g, \tag{3.26}$$

where **x** is the full-state vector of velocity and displacements. **A** is the state matrix, i.e.,

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{m}^{-1}\mathbf{k} & -\mathbf{m}^{-1}\mathbf{c} \end{bmatrix}. \tag{3.27}$$

$\mathbf{B}_g$ is the input matrix for the excitation force, i.e.,

$$\mathbf{B}_g = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}. \tag{3.28}$$

B is the input matrix for the control force, i.e.,

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} \\ \mathbf{m}^{-1}\mathbf{b} \end{bmatrix}. \tag{3.29}$$

The continuous state-space model is further discretized with a sampling interval of 0.01 s. The performance indices $\mathbf{Q}_x$ and $\mathbf{R}$ in the cost function are set as $\mathbf{Q}_x = 10 \times \mathbf{I}_{12\times12}$ and $\mathbf{R} = 10^{-4}$. The agent critic has three hidden layers, each with a dimension of 64 and ReLU activations. The actor is represented by an NN composed of tanh nonlinearities and two hidden layers, each with a dimension of 64. The training process is the same as that for the SDOF system. A step excitation is applied at the starting point, and the building is allowed to perform under free vibration, and 5-s free vibration records are used in the training process.

The total cost of the LQR controller and the trained DDPG agent is 292.23 and 295.45. The total cost of the DDPG agent is slightly higher, but its performance is still very close to that of the LQR controller. Figures 3.7a and 3.7b show the displacement and velocity time histories of two representative floors. The vibration decay rates of the two controllers are extremely close to each other and are in the same phase.

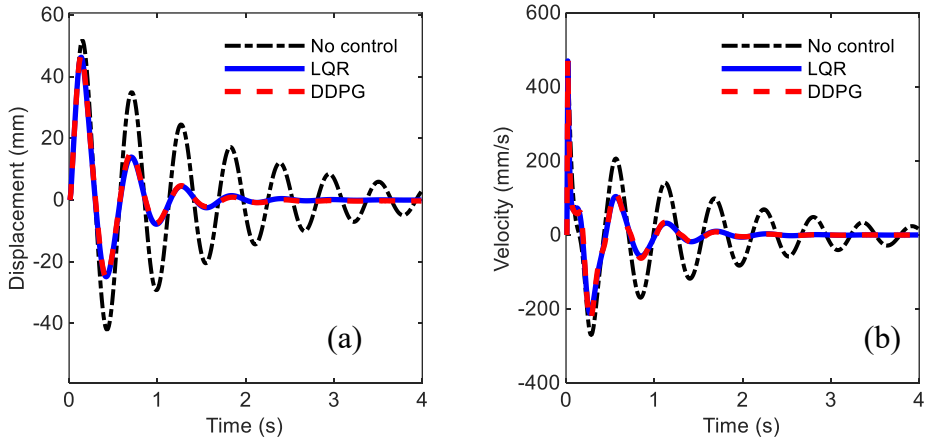


Figure 3.7 Free vibration of full-state observable MDOF: (a) top floor displacement and (b) first floor velocity

The NN controller trained on free vibration is further tested under random excitations and sine sweep excitations (sweeping from 0.5 Hz to 10 Hz in 20 s). Figure 3.8a depicts the displacement response of the first floor under random excitations.

Figure 3.8b shows the root-mean-square (RMS) displacement responses at different floors under random excitations. The RMS displacement of the DDPG agent is smaller than that of the LQR controller, indicating a slightly better control performance of the DDPG agent.

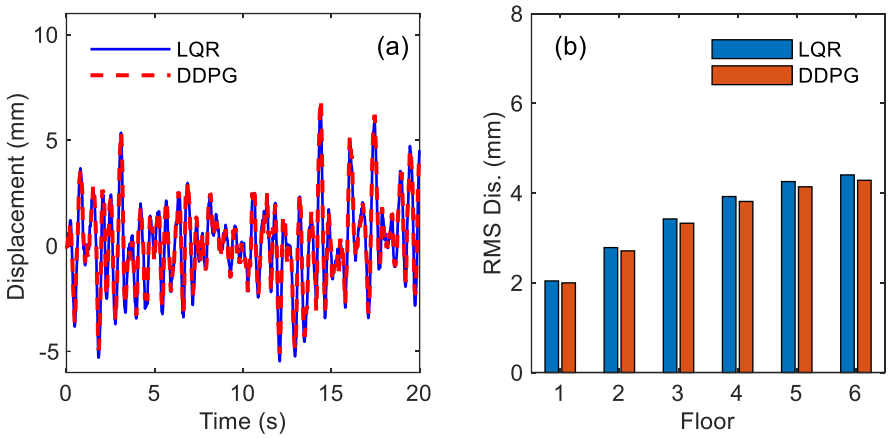


Figure 3.8 Full-state observable MDOF under random excitations: (a) top floor displacement and (b) RMS displacement of different floors

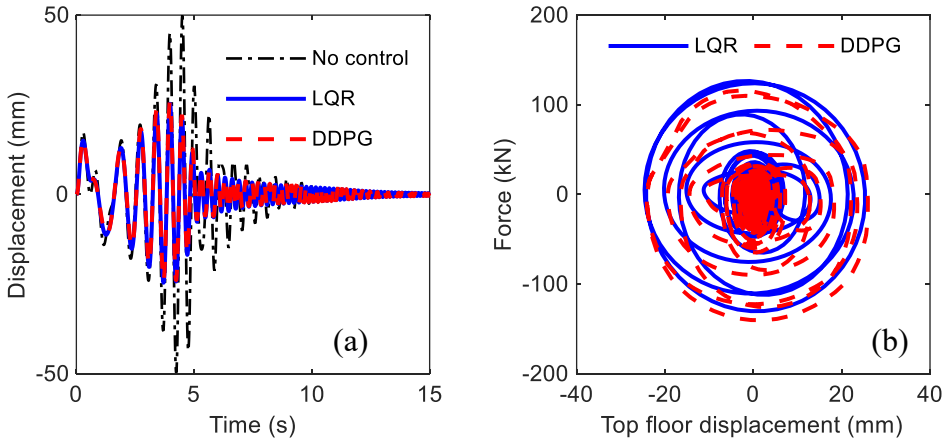


Figure 3.9 Full-state observable MDOF under sine sweep excitation: (a) top floor displacement response and (b) control force vs. first-floor displacement relationship under sine sweep excitation

Figure 3.9a presents the displacement response of the sixth floor under sine sweep excitations. The response time histories are nearly identical for the two controllers

under all frequencies. Figure 3.9b illustrates the relationship between floor displacement and control force. The two controllers demonstrate very close features. Vibration comparison under different excitations indicates that the model-free NN controller trained through the DRL method without any knowledge of system dynamics under various occasions can compete with the performance of the model-based LQR controller in full-state feedback control

3.4.3 Partially observed MDOF system

Another case with a partially observed state is further investigated using the same six-story shear building model. Only six states are observed: the displacements and velocities of the first-, third-, and fifth-floors. The agent structure follows that of the full-state controller except for reducing the nodes of the input layers. The only difference in training is the reward calculation. The weighting matrix $\mathbf{Q}_y$ is selected to be $10 \times \mathbf{I}_{6\times6}$, while $\mathbf{R}$ is the same. The system performs training under free vibration for 6 s during each training episode.

Table 3.3 Total cost for free vibration of partially observed MDOF system

Controller	No. of observable states	Total cost
LQR*y*	12	164.56
Output feedback	6	203.56
DDPG	6	168.80

The free vibration results are provided in Table 3.3, where the cost function is defined in Equation 3.19. The performance of LQRy is the best among the three because it uses the fully observable state to calculate the control force. The trained DDPG agent using partially observable states can achieve performance similar to that of LQRy, and the DDPG agent can outperform the output feedback controller with the same number of observable states. From the training progress shown in Figure 3.10, the episode reward of the DDPG agent surpasses that of the output feedback controller after a dozen training steps. When training stops, the total reward is extremely close

to the result of the full-state controller LQR*y*. The control performance of the output feedback controller is the worst among the three. Figures 3.11a and 11b present the displacement responses of the first floor and top floor, respectively.

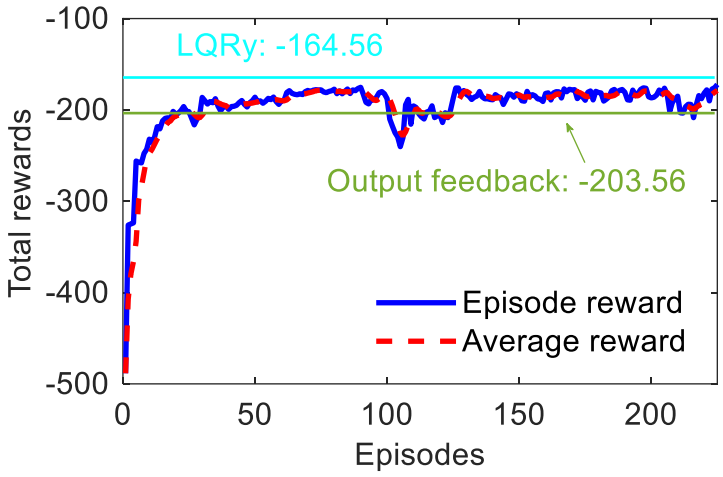


Figure 3.10 Episode and average reward of DDPG during training of the partially observed MDOF system

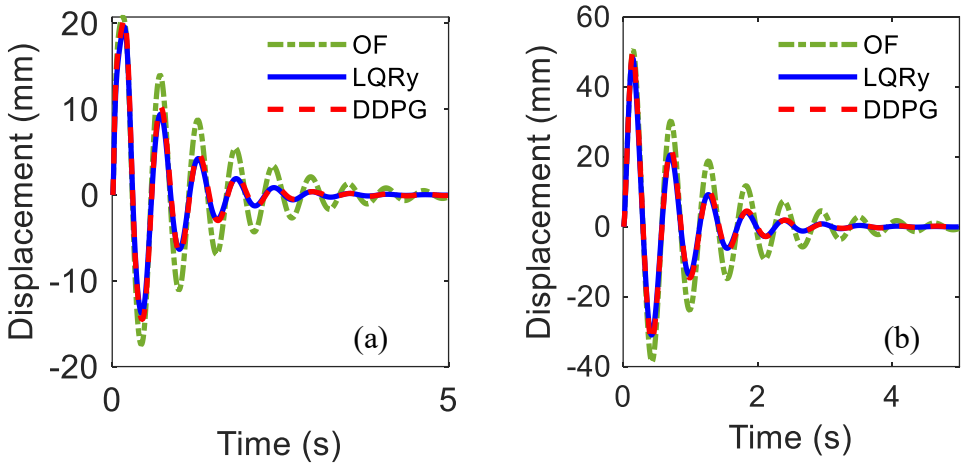


Figure 3.11 Displacement response of the partially observed MDOF system under free vibration: (a) first floor and (b) top floor

Random and sine sweep excitations are also applied. Figures 3.12a and 3.12b show the displacement of the first floor under two different excitations. Similarly, the control performance of the DDPG agent is close to that of the full-state LQRy

controller and exceeds that of the output feedback controller. Figure 3.12c illustrates the force-displacement relationship of the three controllers, in which the negative stiffness features of the LQRy and DDPG agents are close to each other and more significant than that of the output feedback controller when selecting the same weighting matrix. The comparison of the RMS displacement of all the floors is depicted in Figure 3.12d.

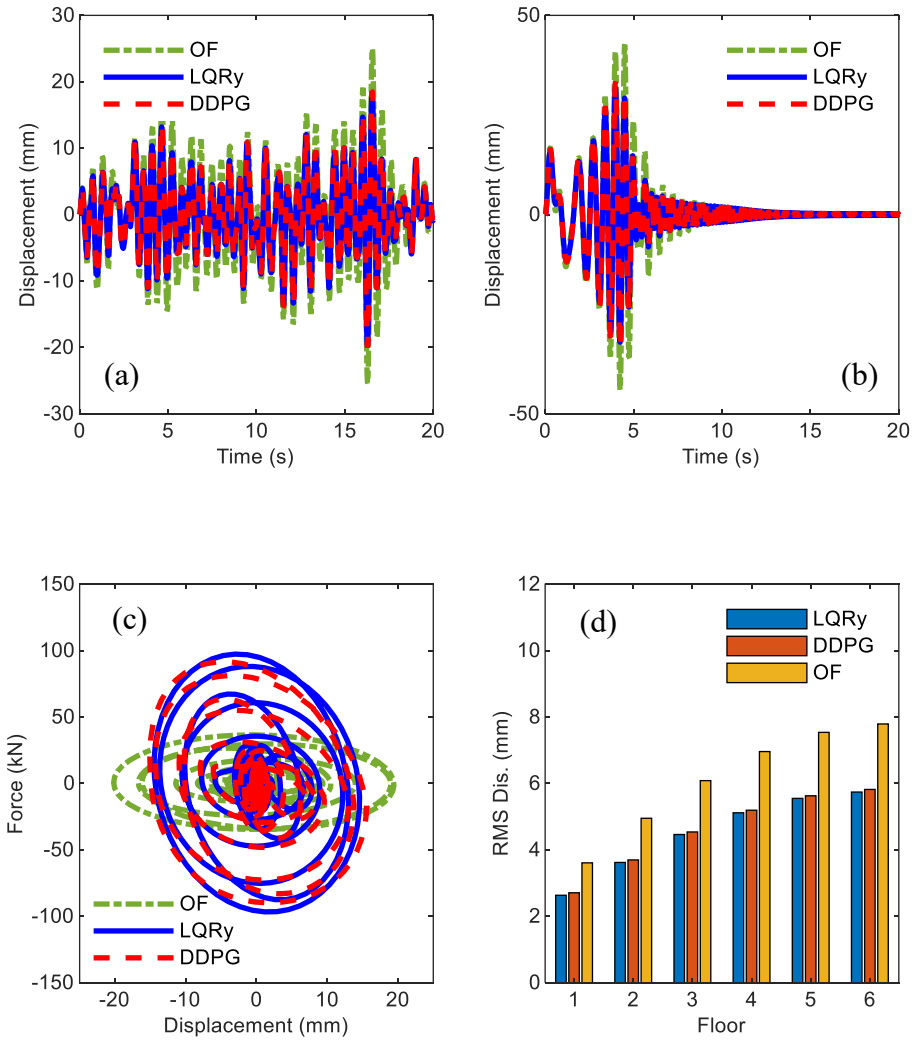


Figure 3.12 Partially observed MDOF under different excitations: (a) top floor displacement under random excitations, (b) top floor displacement under sine sweep excitations, (c) control force vs. first-floor displacement relationship under sine sweep excitation, and (d) RMS displacement at different floors under random excitations

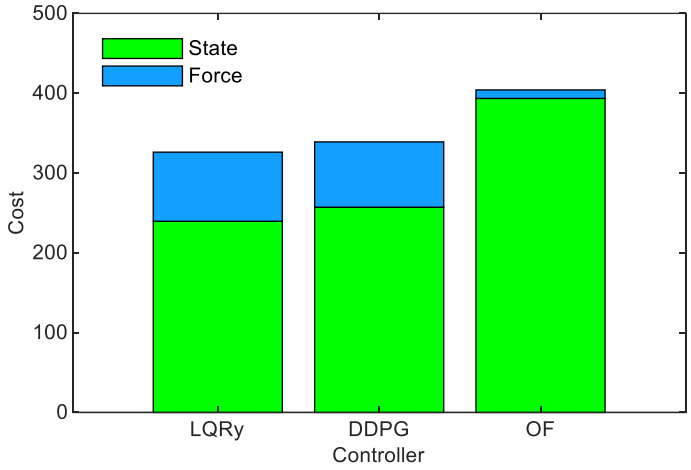


Figure 3.13 Components of total cost under sweep signal excitations

Figure 3.13 shows the components within the total cost (defined in Equation 3.19) under sine sweep excitations. The output feedback controller has the highest cost. Although the force component cost of the output feedback controller is considerably lower than the other two, its state cost is enormous. The DDPG controller has a slightly higher state cost than LQRy, leading to a higher total cost.

3.4.4 Full-state observable MDOF system with measurement noise

Measurement noise in the observed state vector is considered to demonstrate the robustness of the DRL strategy. Thus, the full-state observation becomes

$$\mathbf{y}_n = \mathbf{y} + \mathrm{w}_n. \tag{3.30}$$

Table 3.4 Total cost of MDOF under free vibration with and without noise

Noise	LQR	DDPG	Difference (%)
Without noise	292.23	296.96	1.59
With noise	293.42	298.26	1.62

The training progress stays the same as that in subchapter 3.4.2, except adding the measurement noise, and the trained controller is compared with the LQR controller. The total cost after the training is summarized in Table 3.4. Similar to the case without noise, the total cost of LQR is still slightly lower than that of DDPG when the measurement noise is considered. In both cases without and with noise, the differences in the total cost are less than 2%. The displacement time histories of the first and top floors are shown in Figures 3.14a and 3.14b, respectively. The vibration decay rates are nearly the same when considering the measurement noise in the observation function. The control force–displacement relationship is presented in Figure 3.15. The maximum force provided by the DDPG agent is lower than that of LQR, but the overall trend is highly similar.

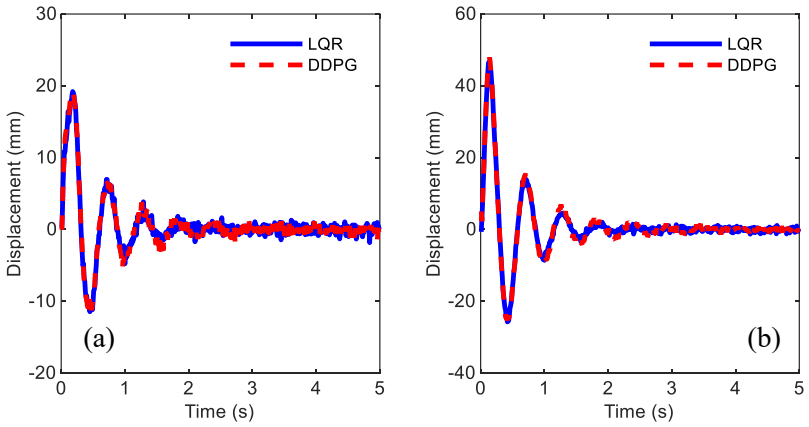


Figure 3.14 Displacement response of MDOF with measurement noise under free vibration: (a) first floor and (b) top floor

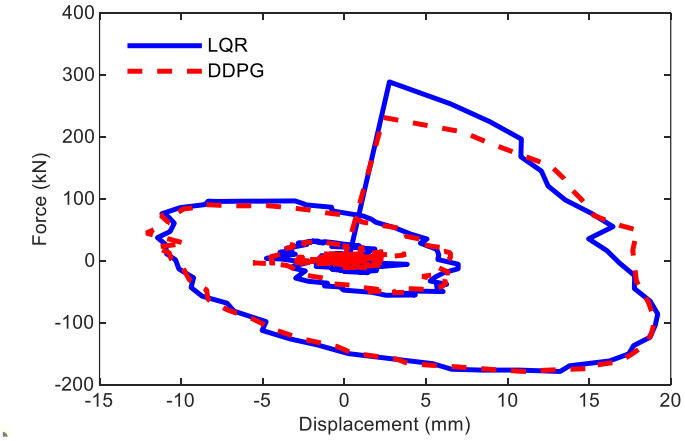


Figure 3.15 Control force vs. first-floor displacement relationship of MDOF with measurement

3.5 Summary

A novel DRL-based vibration control strategy is presented in this chapter, which, to the best of the author's knowledge, presents the first study that introduces DRL into the optimal structure control domain. An off-policy model-free algorithm, i.e., DDPG, is used to train the NN controller, and its effectiveness is tested under various conditions.

The proposed DRL-based controller does not require any information regarding structural dynamic models but can still achieve control performance comparable to that of the traditional model-based LQR controllers. The proposed method is numerically tested in an SDOF system and a six-story shear building MDOF system. The agents are only trained under free vibrations and then tested under different levels of random excitations. For all the cases, the total costs and displacement levels are nearly the same as those of LQR. In addition, the control performance is relatively consistent under varying measurement noise. When feedback is not in a full state, the DRL controller is considerably better than the output feedback controller with the same number of observed states. The control performance of the DRL controller does not degrade too much compared with the full-state feedback controller LQR*y*, although the former uses fewer measured states in the feedback than the latter.

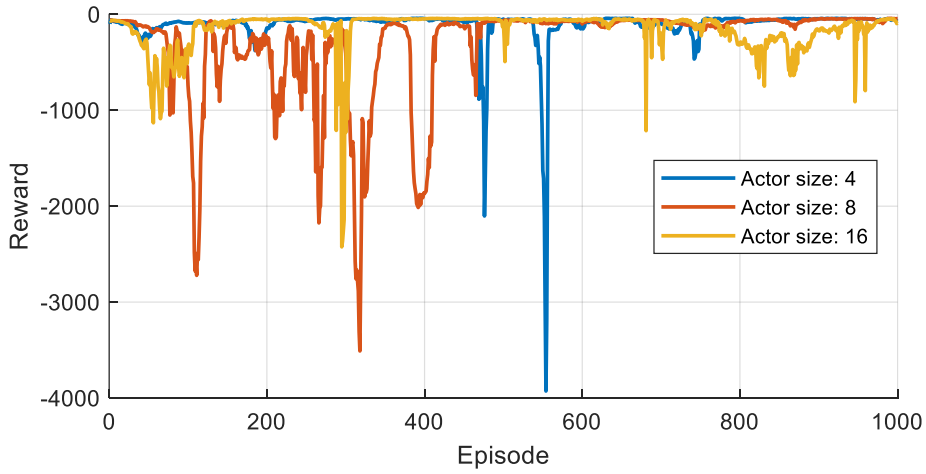
This chapter only focuses on linear structural systems. Implementing the proposed strategy in nonlinear systems will be investigated in Chapter 4, in which the strength of DRL is expected to be better utilized.

3.6 Appendix

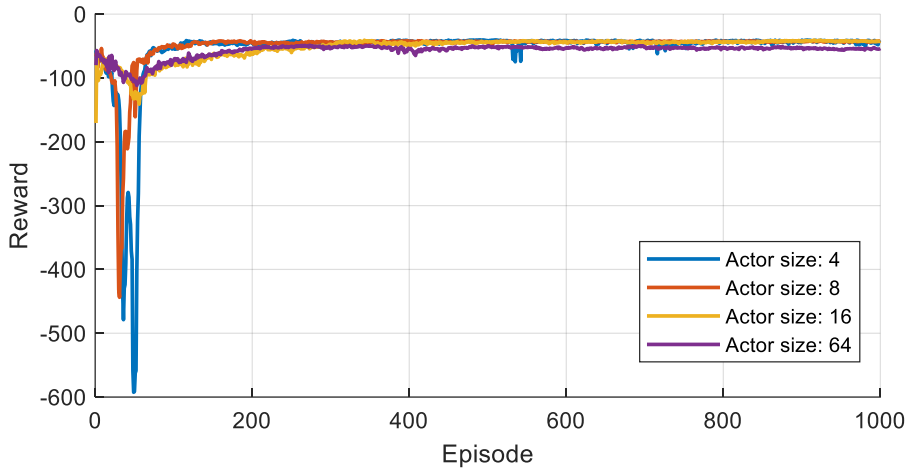
3.6.1 Selection of hyperparameter in DDPG

Hyperparameters have a critical influence on the performance and stability of DRL algorithms. In the DDPG setup used in this chapter, more than ten hyperparameters need to be determined, and there is no universal guideline for their tuning. The numerical model of the SDOF-AMD system described in Chapter 6 is used as an example to discuss the sensitivity of the DDPG training to the hyperparameters.

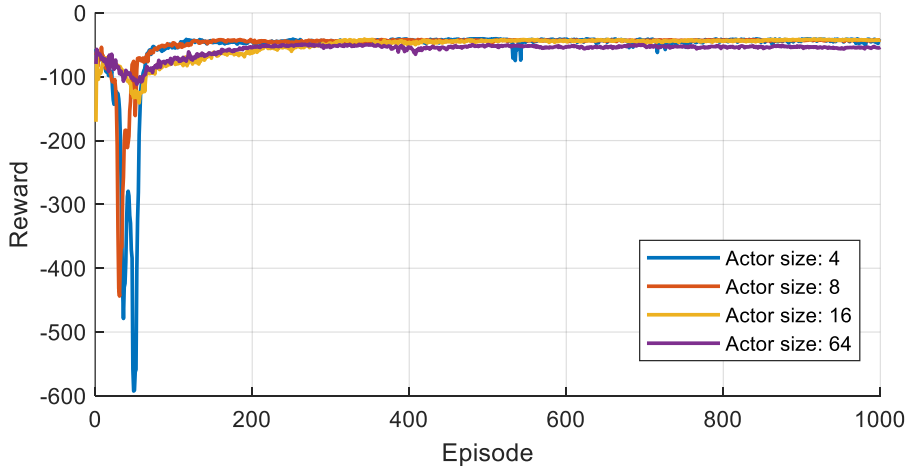
The DDPG agent adopts an actor–critic DRL structure. The critic network is designed with sufficient capacity to approximate the action–value function, while the actor network is generally smaller. For both networks, the choice of architecture (number of layers and units) and learning rate are key, as the latter directly controls the magnitude of the weight updates at the end of each training step. The influence of the network sizes is illustrated in Figure 3.16, where different actor sizes are tested with critic hidden layer sizes of 16, 64, and 256 neurons. As the critic size increases, the training curves become much smoother and converge faster, and overly small critic networks (e.g., 16 neurons) lead to unstable learning, whereas moderate actor sizes (8–16 neurons) already provide satisfactory performance.



(a)



(b)



(c)

Figure 3.16 DDPG training rewards for varying critic hidden layer sizes and actor sizes: (a) critic 16 nodes; (b) critic 64 nodes; (c) critic 256 nodes

nodes.

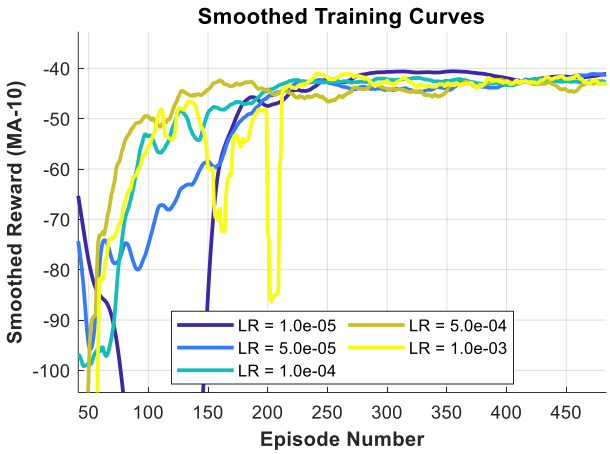


Figure 3.17 Reward of DDPG training with different actor learning rates

The learning rate requires meticulous tuning. Figure 3.17 compares smoothed training curves for actor learning rates between 1×10^{-5} and 1×10^{-3}. Very small learning rates lead to slow convergence, while excessively large values result in oscillatory or even divergent behaviour in the early training episodes. An intermediate value of 5×10^{-4} offers the best trade-off between convergence speed and stability for the SDOF–AMD benchmark and is therefore used as the default setting in this chapter.

In addition to the network architecture, several training-related hyperparameters are specific to each DRL algorithm. The representative examples in DDPG include the mini-batch size, the length of the experience replay buffer, the smoothing factor used for updating the target actor and critic networks, and the policy update frequency. Because DRL training is highly stochastic, there is no guarantee that a particular hyperparameter setting is globally optimal. Among all tested configurations, the network architecture and the learning rates were found to have the strongest impact on performance, while the mini-batch size mainly affects training efficiency.

The effect of the mini-batch size is shown in Figure 3.18, where batch sizes ranging from 64 to 1024 are compared. Smaller batches converge faster initially but exhibit larger fluctuations in the episode reward, whereas larger batches yield smoother learning curves at the expense of higher computational effort. A batch size in the range of 128–256 samples provides a good trade-off between convergence speed and stability and is therefore adopted in the main numerical studies. All batch sizes eventually reach comparable performance while differing mainly in the smoothness of the learning process.

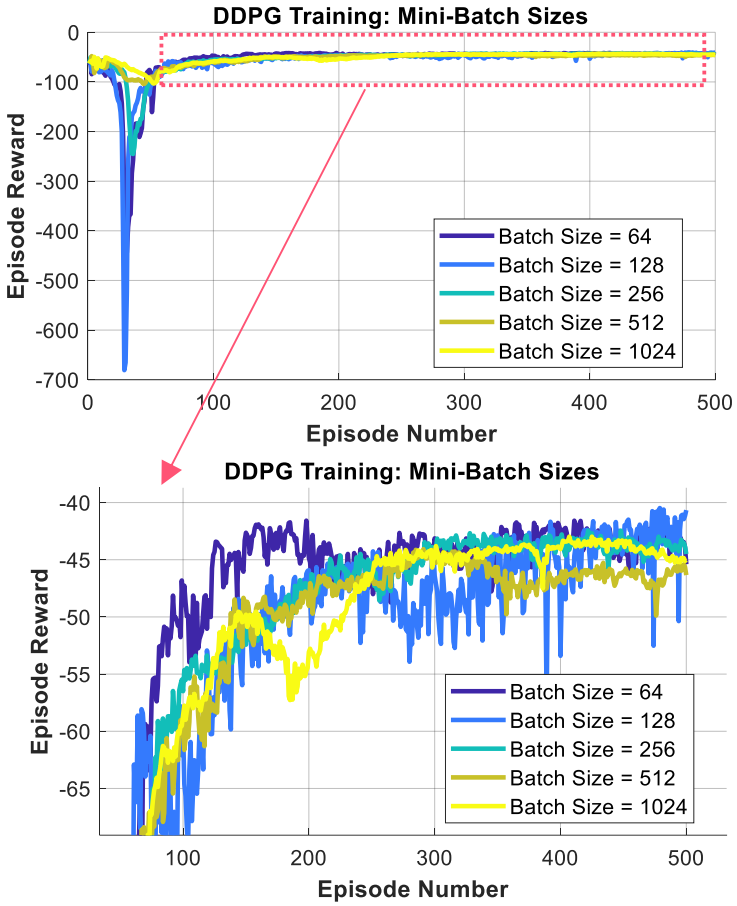


Figure 3.18 Reward of DDPG training with different mini-batch sizes

3.6.2 Practical recommendations

In practice, it is recommended to start from moderate hyperparameter values suggested in literature or from previous related studies, and then perform systematic

variations around this baseline using a simplified numerical model of the target system. By monitoring convergence speed, stability, and final control performance for different network sizes, batch sizes, and learning rates, suitable hyperparameter ranges can be identified before applying the DRL controller to more complex structural systems.

CHAPTER 4

DRL-based Active Nonlinear Vibration Control of a Buckled Beam

4.1 Introduction

As architectural trends lean towards more slender structural elements, nonlinear effects, particularly geometric nonlinearities, become increasingly important (Ali and Moon, 2018). While increasing the stiffness of a structure can directly reduce displacements and mitigate nonlinear effects, it also inevitably increases structural weight and the associated costs. Achieving a balance between minimizing risk and avoiding excessive material use necessitates the development of advanced, adaptive, and scalable control strategies to manage nonlinear vibration challenges effectively. Following the work in Chapter 3, this chapter aims to extend the DRL-based vibration control strategy to the nonlinear vibration control area.

As previously mentioned in subchapter 2.2.1.2, while active control methods have achieved considerable success in linear systems, extending their applications to nonlinear systems faces new challenges. The performance of model-based controllers degrades when the modeling assumptions cannot accurately capture the nonlinear dynamics.

This chapter developed and validated a model-free active control strategy for nonlinear vibration control. An NN controller is trained by DRL to eliminate uncertainties and complexities of the target nonlinear system. The proposed strategy was numerically tested on a shallow, simply-supported buckled beam with nonlinearities up to the 3rd order. Under certain loading conditions, the beam may lose stability at load levels below the inherent strength of the material and suffer snap-through buckling (Pinto and Gonçalves, 2000). The DRL-based methodology demonstrated superior performance through comparison with prevailing model-based polynomial controllers. The strategy not only attenuated vibrational responses but also significantly enhanced the safety margins of the structural element, effectively reducing the risk of snap-through buckling while maintaining a conservative energy consumption.

The key contributions of this chapter can be summarized as follows:

(a) Development of an advanced active control methodology for nonlinear vibration control, which outperforms conventional model-based polynomial controllers.

(b) Introduction of a highly adaptable and easily implementable active nonlinear control strategy with customizable training parameters to suit various application scenarios.

The structure of this chapter is organized in the following manner: Subchapter 4.2 elaborates on the DRL controller that we propose. Subchapter 4.3 provides a detailed description of the nonlinear system utilized for testing. Subchapter 4.4 discloses the implementation details and the simulation results under various test conditions. Lastly, subchapter 4.5 offers a summary and conclusion of this chapter.

4.2 Nonlinear Vibration Control

4.2.1 Classical model-based optimal control

For a fully observable and controllable dynamical system, the state-space representation can be expressed as:

$$\dot{\mathbf{x}} = f[\mathbf{x}(t), \mathbf{u}(t), t]. \tag{4.1}$$

where $\mathbf{x}(t)$ is the state vector, $\mathbf{u}(t)$ is the control vector, t states the time, and f is the system's dynamic function. In the case of a linear system, optimal control approaches such as the LQR controller are directly applicable. The optimal feedback gain $\mathbf{K}_{\mathrm{lqr}}$ is determined through the resolution of the ARE. Consequently, the control action $\mathbf{u}$ can be determined according to Equation 3.3, where the optimal feedback gain $\mathbf{K}_{\mathrm{lqr}}$ is to minimize the quadratic cost function defined in Equation 3.4.

The efficacy of linear control gains, however, diminishes when addressing nonlinear systems unless these systems are linearized around specific operational points. This limitation necessitates the exploration of nonlinear control strategies. As detailed by Suhardjo et al. (1992), a polynomial controller offers one such alternative. Recursive equations can be formulated and solved by expressing the state equation and control force as polynomial functions. Although these equations can extend to any polynomial degree, it is essential to note that as the order increases, so does computational complexity; at the same time, the incremental contribution to control performance becomes marginal. This research compares the polynomial controller up to the 3$^{\mathrm{rd}}$ order with the DRL-based control methodology proposed in the following subchapter.

4.2.2 DRL-based nonlinear control

To address the abovementioned problems, the current study takes advantage of the DRL method introduced in Chapter 3 to train an optimal NN controller for nonlinear vibration control without directly solving the ARE. The control problem is systematically formalized into a DRL paradigm, as discussed in subchapter 3.3.1. At every time step t, the agent experiences the state $\mathbf{s}_t$ and takes the action a_t according to policy μ. The agent receives the reward r_t and ends up in a new state $\mathbf{s}_{t+1}$ with the probability P. The agent aims to find the optimal policy μ^* to maximize the total discounted reward H_t with the discount factor $\gamma \in [0, 1]$, as defined in Equation 3.16.

For the specific problem of vibration suppression under variable excitation conditions, the step reward is formulated as a negative quadratic cost:

$$r_t = -\mathbf{s}_t^T \mathbf{Q}_s \mathbf{s}_t - a_t^T \mathbf{R} a_t. \tag{4.2}$$

Thus, as the expected reward represents the future control performance, maximizing the total return in Equation 3.16 is equivalent to minimizing the quadratic cost function in Equation 3.4. In this way, the controller's performance is optimized. By utilizing a model-free DRL algorithm, the study leverages the benefit of not requiring any precise knowledge of system dynamics, thereby enhancing the controller's robustness to adapt to uncertainties prevalent in real-world control applications. The illustration of training a control policy for a target nonlinear plant by DRL is shown in Figure 4.1.

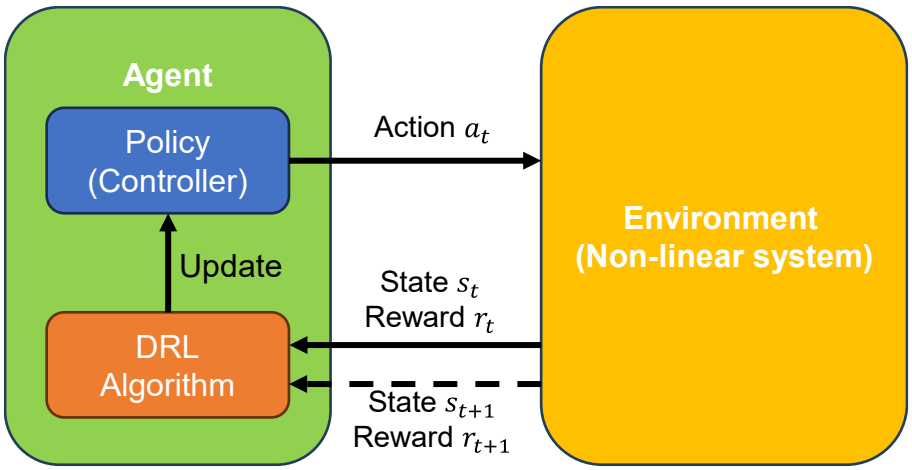


Figure 4.1 DRL structure for nonlinear vibration control

In the present investigation, the system can oscillate freely or under external excitation for a predetermined duration. The control policy is subsequently updated based on the data recorded. This data-driven update mechanism utilizes the TD3 algorithm, obviating the need for solving complex nonlinear equations traditionally associated with model-based controllers. TD3 is a model-free, online, and off-policy algorithm that employs a deterministic policy for action selection. The TD3 agent comprises two key components, namely the 'actor' and the 'critic', both implemented using DNN. The actor is responsible for generating control actions based on the current system state or observations. In comparison, the critic estimates the Q-value of state-action pairs, thereby guiding the policy update mechanism.

TD3 represents an advanced variation of the DDPG algorithm and incorporates

three key improvements to mitigate the overestimation bias in the Q-value function, as discussed in subchapter 2.3.6. The algorithm employs double Q-learning by instantiating two separate critic networks and adopting the lower of the two Q-value estimates for policy evaluation. The policy is updated at a reduced frequency compared with the Q-function updates, contributing to more stable learning dynamics. A small amount of clipped random noise is averaged among the mini-batches and added to the selected action when updating the value function to make the regularization smoother. In TD3, exploration is facilitated by incorporating noise into the actions determined by the policy network throughout the training phase, thereby enriching the agent's experience by promoting the discovery of a broad range of strategies within the action space.

Figure 4.2 compares the training curves of TD3 and DDPG on the six-story shear building problem described in subchapter 3.4.2. It can be observed that the training process of TD3 exhibits slightly greater stability. Given the inherent uncertainties associated with nonlinear systems, a more stable training process is usually preferable; therefore, TD3, instead of DDPG, is adopted in this chapter.

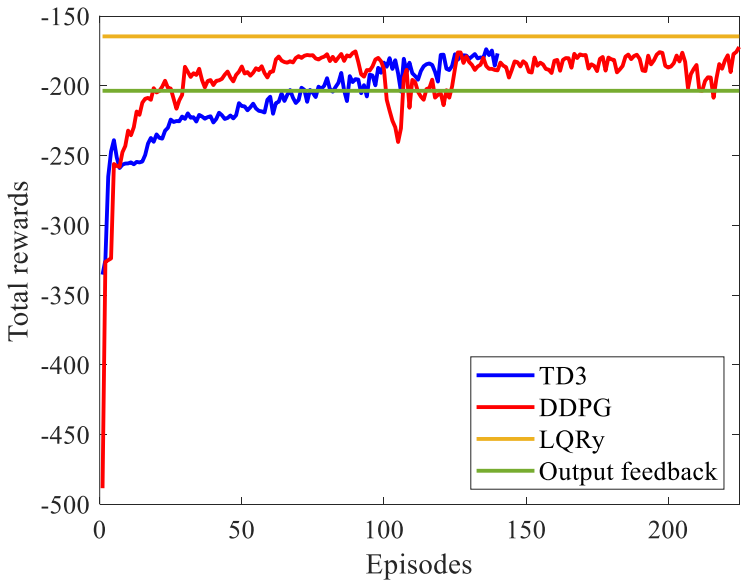


Figure 4.2 Comparison of TD3 and DDPG training curves

Upon completing the training process outlined in Algorithm 4.1, the trained actor alone will serve as an offline controller. The main difference from Algorithm 3.1 is the

replacement of DDPG by TD3. It can generate control forces in response to observed system states, like a traditional model-based controller, but with the additional benefits of enhanced adaptability and robustness.

Algorithm 4.1: Nonlinear controller training based on DRL

Initialization step: initial agent, initial state vector
repeat
  Observe state s_t and get action $a_t = \mu(s_t)$
  Execute a_t in the environment
  Observe the next state
  Get the reward $r_t(s_t, a_t)$
  If s_{t+1} is a terminal, then reset the initial states and start a new episode
  If it is time to update, then
    μ' =TD3(μ, $[s_t, r_t(s_t, a_t), s_{t+1}]$)
  end if
until convergence

4.3 Nonlinear Control Problem Formulation

4.3.1 Buckled beam

To evaluate the efficacy of the proposed DRL-based control methodology, this subchapter investigates the nonlinear oscillations of a shallow, simply-supported buckled beam subjected to transversal dynamic loads. The SDOF model derivation is aligned with the work by Pinto and Gonçalves (2002), who employed polynomial controllers up to the third order for the same problem.

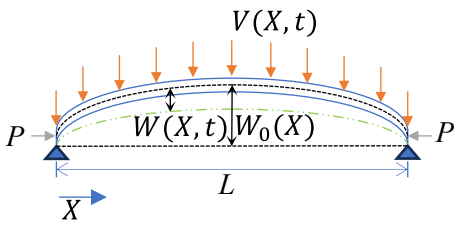


Figure 4.3 Buckled beam setup

Consider a homogeneous elastic simply-supported beam, as shown in Figure 4.3. The section properties are defined by length L, Young's modulus E, and moment of inertia I. Assuming the response W and the lateral load V are symmetric. For a large initial deflection, non-symmetric responses involving higher vibration modes may occur. In this chapter, the initial deflection is assumed to be very shallow, and only the first mode is considered. The equation of motion of the undamped SDOF model can be written as follows in a dimensionless form:

$$\ddot{\alpha} + (1 - p + 2\beta^2)\alpha + 3\beta\alpha^2 + \alpha^3 = -v, \tag{4.3}$$

$$w(x,t) = \alpha(t)\sin(\pi x), \tag{4.4}$$

$$p = \frac{P}{P_{cr}} = \frac{PL^2}{\pi^2 EI}, v = \frac{VL^4}{2\pi^4 EIr}, w = \frac{W}{2r}, x = \frac{X}{L}, \tag{4.5}$$

$$\beta = \pm\frac{L}{6\pi r}\sqrt{-24 + 18p \pm 6\sqrt{9p^2 - 8}}. \tag{4.6}$$

where α represents the amplitude of the first mode. p, v, w, x are dimensionless quantities. The initial axial load of the beam is P, P_{cr} is the static Euler critical load of the beam subjected to axial compressive load, and r is the radius of gyration. Amplitude β is the solution to the static buckling problem, assuming the initial static deflection form is the first buckling mode.

The control mechanism of the buckled beam consists of applying moment couples at specific points along the beam axis. The dimensionless system equation of motion of the controlled beam is:

$$\ddot{\alpha} + (1 - p + 2\beta^2)\alpha + 3\beta\alpha^2 + \alpha^3 - \sum_{i=1}^{k} 2\cos(\pi d_i)\, u_i = -v, \quad (4.7)$$

where u_i is the dimensionless quantity of the ith control moment located at a distance d_i from the origin, and k is the number of the applied control moment.

4.3.2 Static behavior

The behavior of the buckled beam under static loading is initially analyzed. Considering a time-independent transversal load v, the static equilibrium equation could be obtained from Equation 4.3 as:

$$(1 - p + 2\beta^2)\alpha + 3\beta\alpha^2 + \alpha^3 = -v. \quad (4.8)$$

For p = 1.005 and β = 1, the static response of the buckled beam with a very shallow arch is shown in Figure 4.4, indicating two stable equilibrium positions (α = -2.005 and α = 0) and one unstable equilibrium in the middle (the black cross markers). The snap-through buckling path (red dashed line) from point P1 to P2 occurs upon a slight incremental load from the local peak at P1, and the deflection significantly increases. Similar behavior is observed from P3 to P4 if the load is reduced from P3. This dynamic behavior leads to oscillations around point P2 or P4 in the absence of damping.

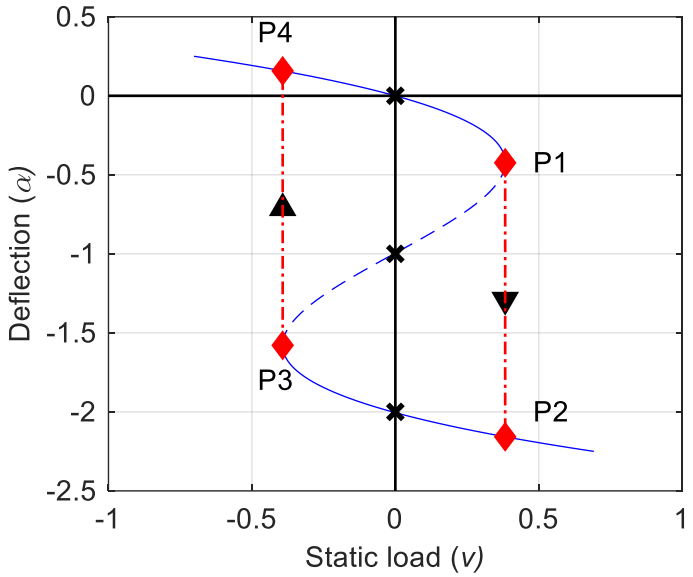


Figure 4.4 Snap-through behavior of the bulked beam

4.3.3 Dynamic behavior

If a linear viscous damping equivalent to 3.54% of critical damping of the linearized system is assumed, and two actuators are located at L/3 and 2L/3, the equation of motion for the uncontrolled system becomes:

$$\ddot{\alpha} + 0.1\dot{\alpha} + 1.995\alpha + 3\alpha^2 + \alpha^3 - 2u = -v. \tag{4.9}$$

In the uncontrolled scenario, the system's final state under free vibration is sensitive to its initial condition, as two stable equilibrium points exist in the system. Figure 4.5 illustrates the displacement responses of the beam under free vibration for two distinct initial states, where the escape point is α = –1.00. The MATLAB function ode45 was employed to perform the integration for the numerical solution of these nonlinear equations.

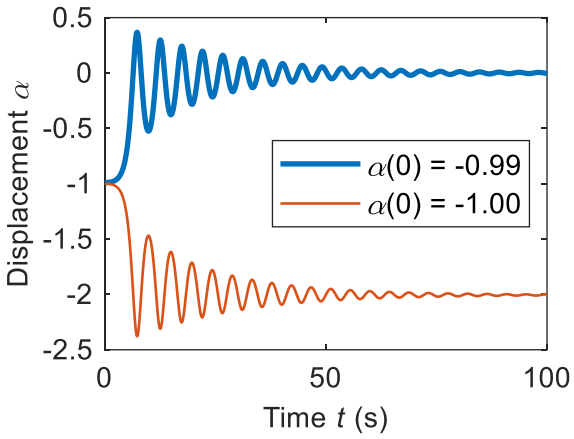
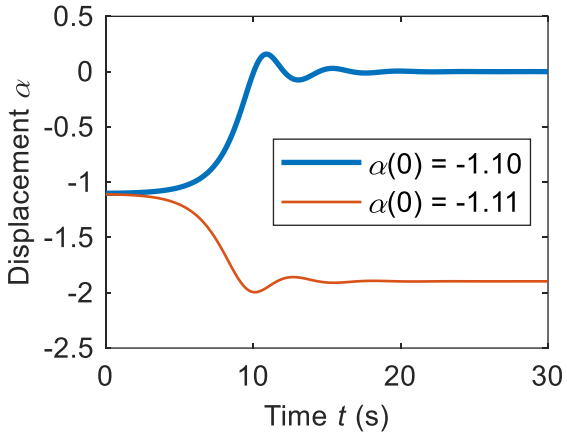
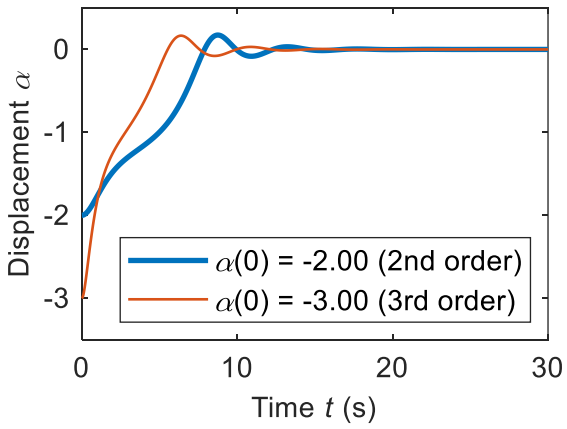


Figure 4.5 Free vibration response of the uncontrolled system



(a)



(b)

Figure 4.6 Free vibration response of the system with (a) the first order controller; (b) the second and third order controllers

With the weighting matrix in Equation 4.2 chosen to be $\mathbf{Q}(1,1) = 0.1, \mathbf{Q}(2,2) = 0.1$ and $R = 1$ (other coefficients are set to zero), the 1st order LQR controller was implemented, and its impact on the free vibration behavior near the escape point is depicted in Figure 4.6a. The escape point shifted to $\alpha = -1.11$.

By employing higher-order controllers based on the polynomial strategies proposed by Suhardjo et al. (1992), the following observations were made. For the 2nd order control, the system invariably settles to the first equilibrium point $\alpha = 0$, regardless of where the initial position is, even at the second equilibrium point $\alpha = -2.00$. A similar trend is seen with the 3rd order control. Figure 4.6b shows the displacement response from an initial state $\alpha = -2.00$ with the 2nd order control and changing the initial state to $\alpha = -3.00$ with the 3rd order control. The weighting coefficients for the higher-order terms of the polynomial controller were set to zero for this analysis.

When subjected to a sudden step load, the system has the possibility to cross the energy barrier, resulting in dynamic buckling even at loads lower than the static critical load. The load needed for this dynamic transition increases with the order of the control; however, the increase becomes marginal beyond third-order control. In the case of an uncontrolled system, the escape load stands at 0.31, which is significantly lower (by 18.4%) than the static critical load of 0.38.

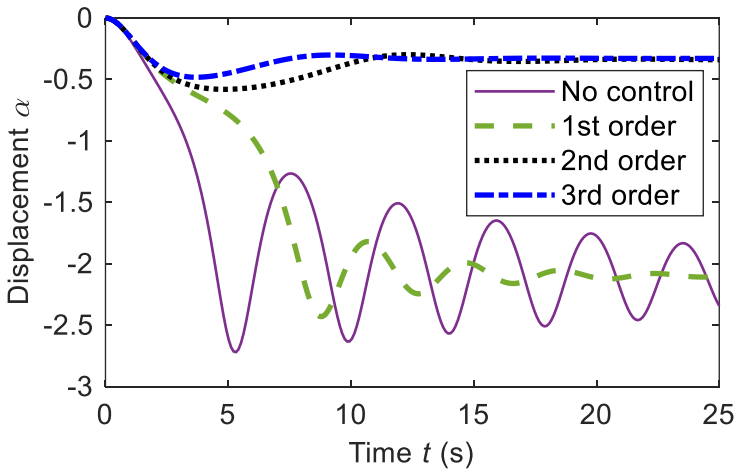


Figure 4.7 Time response of different orders of control under a step load of infinite duration ($v = 0.42$)

Figure 4.7 shows the response time history of the beam subject to a step load of infinite duration (v = 0.42) under the regulation of the active controllers of varying orders. At this load level, the 1st order controller proves insufficient in averting snap-through buckling. The beam transitions to a post-buckling state, indicative of the limitations of linear control schemes for systems exhibiting nonlinear behaviors. In contrast, the 2nd and 3rd order controllers are competent in retaining the system within the first potential well, thereby preventing buckling. In the following subchapter, the performance of the DRL-trained controller is presented and compared to the model-based polynomial controllers.

4.4 Results and Discussions

4.4.1 DRL training

The subsequent subchapters will delve into the performance and characteristics of the controller trained using DRL. Based on the behavior of the system mentioned above, the DRL training is performed by applying a step load of infinite duration (0.48), slightly higher than the escape load of the 3rd order (0.47) controller. The training allows the agents to explore during interaction with the environment. Each training episode lasts for 25 seconds. The training methodology employs the TD3 algorithm, featuring a training duration of 800 episodes. The training was conducted using *MATLAB* on an Intel i7 9700 CPU. From the onset, exploration is initiated by implementing a noise-added action policy, with the exploration rate starting at an initial action noise variance of 0.001 and a decay rate of 1×10^{-5}. The batch size for training is set at 128. The architecture for the actor-network includes two hidden layers with 16 nodes for each layer, and employs a hyperbolic tangent (Tanh) activation function. The actor learning rate is set at 1×10^{-4}. Both critic networks feature three hidden layers with 64 nodes in each layer, with rectified linear units (ReLU) as the activation function and a learning rate of 0.001. The discount factor is γ=0.99.

The observation includes the system state $[\alpha, \dot{\alpha}]^T$. The reward is the negative quadratic cost (Equation 4.10) by setting the weighting matrix the same as the 1st order

controller, i.e., $\mathbf{Q}(1,1) = 0.1, \mathbf{Q}(2,2) = 0.1$ and R = 1 (other coefficients are set to zero). The performance cost J is the combination of state response and control force cost.

$$J = \sum_{i=1}^{i=max\ step}([\alpha_i, \dot{\alpha}_i]\mathbf{Q}[\alpha_i, \dot{\alpha}_i]^T + u_i R u_i). \tag{4.10}$$

During the training process, the DRL-trained controller's performance was analyzed with an initial system state at $\alpha = -1.50$. As depicted in Figure 4.8, the TD3 agent rapidly outperforms all three traditional controllers, highlighting its advanced adaptive capabilities. The standard deviation of over 5 trials shows a smooth training process. Critically, the optimal quadratic performance cost J of the TD3 agent is found to be 3.51 under the step load of 0.48. This marks a dramatic reduction of approximately 90%, compared with the 3rd order controller, which registered a J value of 35.26. Notably, the 3rd order controller was unable to avert snap-through buckling at this elevated load level. The substantial disparity in performance J between the DRL-trained and the 3rd order controllers underscores the former's superior capability in managing complex nonlinear dynamics, particularly under challenging loading conditions. The TD3 agent avoids snap-through buckling and accomplishes this with significantly less energy consumption, demonstrating its robustness and efficiency.

Notably, although the DRL training is done under a single step load (v = 0.48), the same trained DRL-based controller will be tested under different loading conditions in this subchapter, including step loads with different amplitudes and random excitations.

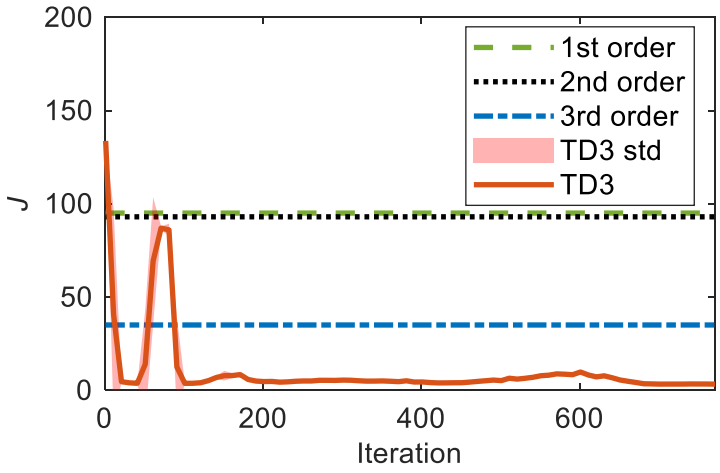


Figure 4.8 The evolution of the cost function J with the iteration of DRL training

4.4.2 Escape point

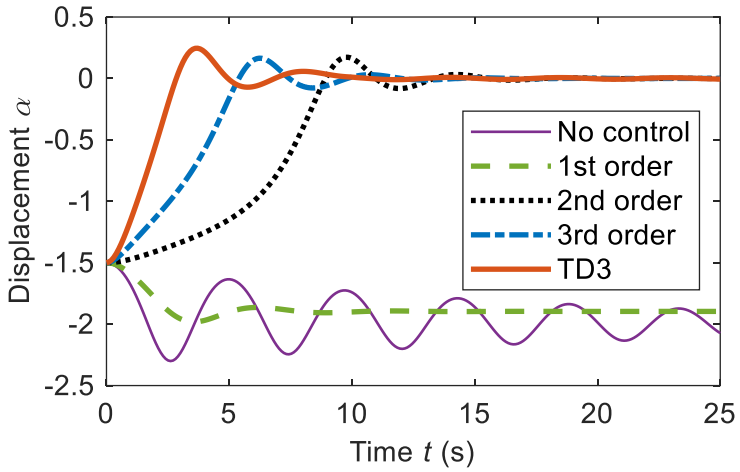


Figure 4.9 Time response of TD3 agent controlled free vibration of the buckled beam

The DRL-trained controller's performance is further assessed by monitoring the system's free vibration response. Results obtained using the same weighting matrix are visualized in Figure 4.9. The initial displacement is set at $\alpha = -1.50$. The displacement response of the DRL-trained controller declines more swiftly than that of the 3rd order polynomial controller. In addition, the DRL-trained agent exhibits the lowest performance cost J, registering a value of 6.96, compared with a cost of 7.67 for the 3rd order polynomial controller. Like high-order polynomial controllers, the DRL-trained agent returns to the zero-equilibrium state irrespective of the initial conditions.

4.4.3 Safety region

The system's displacement responses under the effect of different controllers are computed and compared under the step load v = 0.45 in Figure 4.10. After about 3 s, the 1st and 2nd order controllers could not keep the beam in the pre-buckling position, while the displacement of the TD3 agent controller remains in the safe region and converges quickly. Increasing the step load greater than the escape load will lead the system to experience dynamic buckling. The variation of escape load v_e is detailed in Table 4.1, where Δv_e is the improvement in comparison with the uncontrolled case. The trained agent increased the escape load of the system considerably by 22.58%,

compared to the 3rd order polynomial controller.

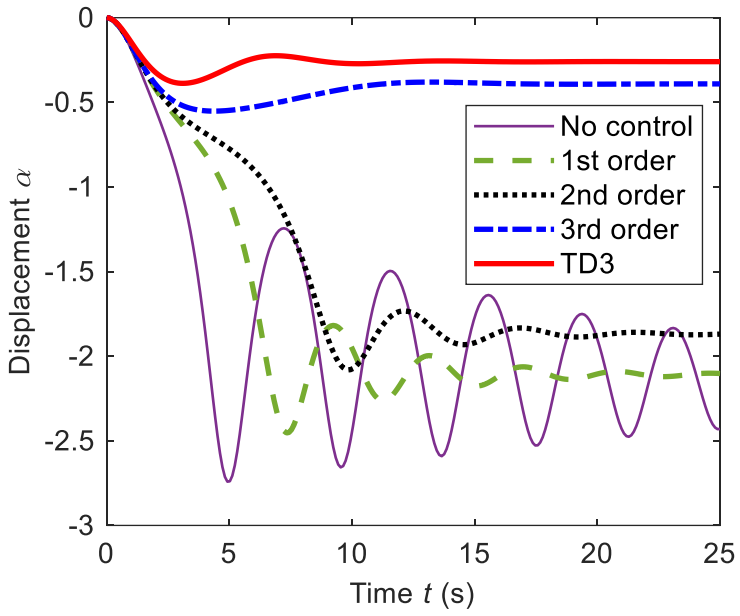


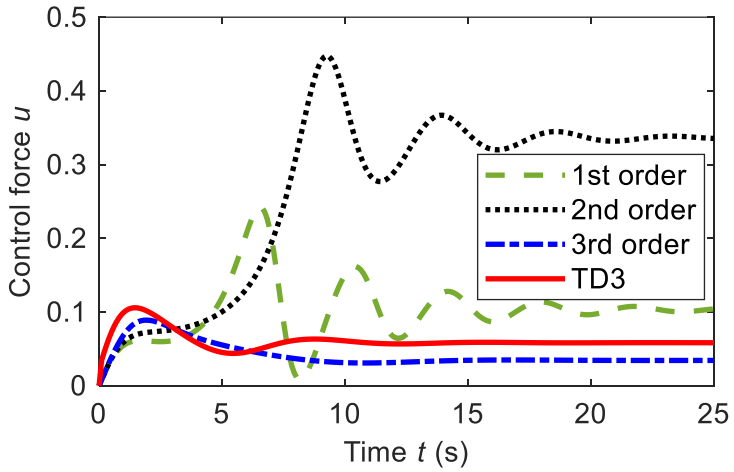
Figure 4.10 Response time histories of the controlled buckled beam system under a step load

Table 4.1 Variation of the escape load

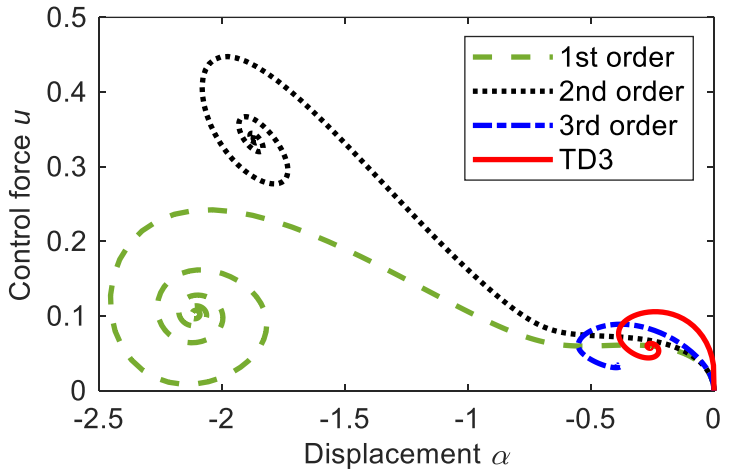
Controller		v_e	Δv_e (%)
No control		0.31	
Traditional	1st order	0.39	25.81
	2nd order	0.43	38.71
	3rd order	0.47	51.61
DRL	TD3	0.54	74.19

Figures 4.11a and 4.11b display the control force history and the force-displacement relationship under the chosen step load. For the 1st and 2nd order controllers, the control force rapidly escalates after buckling, while the DRL-based controller requires slightly more peak force than the 3rd order controller but achieves

significantly better performance in terms of the cost function J. The DRL-trained agent tends to generate a larger control force at lower displacement levels, resulting in a smaller resting displacement. This aspect is readily observable from the force-displacement plot in Figure 4.11b. The peak and RMS values of displacement, velocity, and force for different controllers are compared in Table 4.2. In terms of system safety and robustness, the DRL-trained controller significantly outperforms the traditional controllers. Its capability to maintain the system within a safe region under varied loading conditions, combined with its efficiency in control force and energy use, highlights its effectiveness in handling complex nonlinear systems.



(a)



(b)

Figure 4.11 Response of the controlled buckled beam under step load $v = 0.45$ (a) control force history; (b) control force-displacement relationship

Table 4.2 Responses and control forces of the buckled beam system under step load v = 0.45 with different controllers

Controller		Displacement		Velocity		Control Force		Cost J
		Peak	RMS	Peak	RMS	Peak	RMS	
Traditional	1st order	2.45	1.87	0.82	0.25	0.24	0.11	92.53
	2nd order	2.08	1.60	0.47	0.16	0.45	0.30	84.81
	3rd order	0.55	0.42	0.26	0.07	0.09	0.05	4.85
DRL	TD3	0.39	0.27	0.22	0.05	0.11	0.06	2.76

4.4.4 Performance cost

An important measure of the controller's performance is the performance cost J under different levels of step loads, as a weighted combination of control effort and system state response reduction. The performance function J of the system under varying levels of step loads was recorded and plotted in Figure 4.12. For loads below the escape load ($v < 0.24$), the TD3 agent exhibits slightly higher J compared to traditional polynomial controllers. As the system surpasses its escape load, a sharp uptick in J value is observed for all controllers. However, the rate of increase for the TD3 agent is more moderate, which could be attributed to its higher escape load. The behavior of the TD3 agent at lower loads might be an artifact of its training regimen. The observed phenomenon may result from the training conditions, specifically the choice of step load levels, which were geared towards preventing post-buckling behavior rather than optimizing performance cost at lower load levels. The training was focused on preventing buckling under loads greater than the escape load of the 3rd order controller. Consequently, the TD3 agent may not have been optimized for minimizing the performance function under conditions where the system is naturally stable.

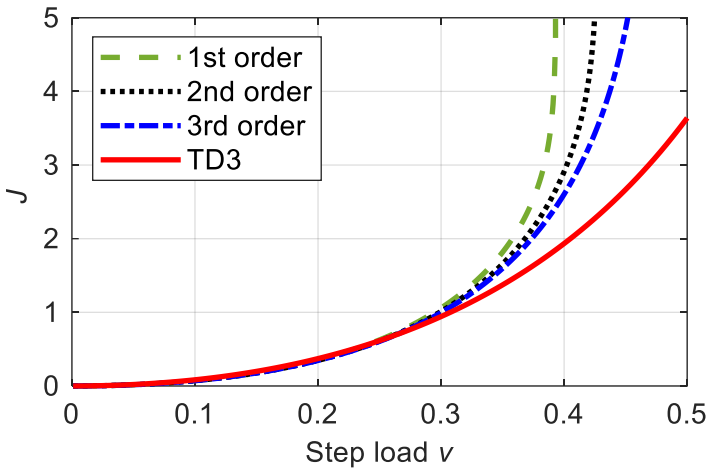


Figure 4.12 Performance cost J vs step load v for different controllers

4.4.5 Random excitation

To evaluate the efficacy of the controller trained through DRL under uncertain conditions, an imposed random excitation lasting 1,000 s was applied to the system. The displacement response under random white-noise excitation, as illustrated in Figure 4.13, makes it evident that neither the linear controller nor the 2nd order controller could prevent the beam from experiencing snap-through buckling. In contrast, both the 3rd order polynomial controller and the TD3 agent effectively mitigated the risk of buckling under the applied level of random excitations. This performance attests to a notable robustness in both controllers when dealing with unpredictable external influences.

Evaluation metrics, nondimensionalized with the responses of an uncontrolled scenario, were selected to analyze the maximum and RMS states of the controlled system. As indicated in Table 4.3, the values of these criteria decrease as the control order increases. This decreasing trend suggests that increasing nonlinearity in the controllers corresponds to enhanced control performance. Interestingly, the TD3 agent registered the lowest values among all the tested controllers, suggesting that the agent may have learned some intrinsic features of the nonlinear system. One salient observation pertains to the TD3 agent's superior capability to minimize maximum

displacement, even when compared to the 3rd order controller. Furthermore, this was achieved while necessitating a substantially lower maximum control force, 0.45 for the TD3 agent versus 0.69 for the 3rd order controller. This underscores the DRL-trained controller's heightened efficiency and effectiveness.

Increasing the order of a polynomial controller can indeed enhance control performance, but this improvement is associated with significantly escalated computational demands as the controller's complexity rises. The proposed DRL control strategy outperforms traditional model-based control methods in terms of adaptability and efficiency. Since DRL training can balance the trade-off between control performance and operation costs, the proposed control strategy shows excellent flexibility. For instance, a higher escape load and broader safety margin can be achieved by setting a larger step load in the training process. In addition, the DRL method learns from the data directly without requiring any model information. This model-free approach can avoid the problems associated with system uncertainties and model inaccuracies, leading to robust control performance and great adaptability in broad vibration scenarios.

Table 4.3 Evaluation Criteria for different controllers under random excitation

Controller		Displacement		Velocity	
		Max	RMS	Max	RMS
Traditional	1st order	0.75	0.47	0.39	0.36
	2nd order	0.56	0.22	0.38	0.36
	3rd order	0.32	0.17	0.36	0.35
DRL	TD3	0.31	0.12	0.35	0.31

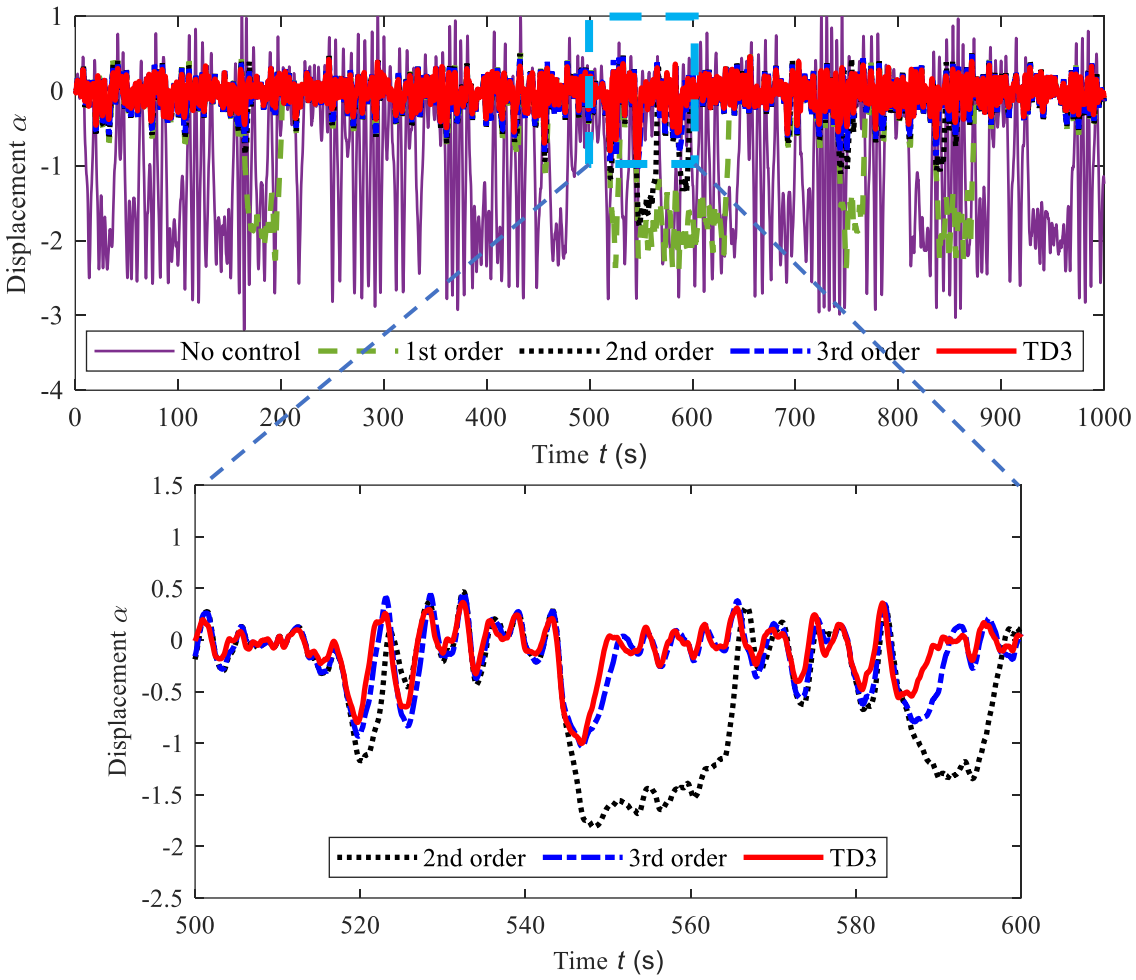


Figure 4.13 Time history of random excitation

4.5 Summary

This chapter extends the DRL control framework to realize model-free nonlinear active vibration control, which was proven to be more flexible and adaptable than the traditional model-based polynomial control strategies.

An NN controller was trained by the TD3 algorithm to realize optimal control performance. The numerical tests were conducted on a simply-supported buckled beam in various vibration scenarios. Superior performances of the TD3-trained NN controller were demonstrated under both static loadings and random excitations. Compared with the system controlled by the 3rd order polynomial controller, the DRL-controlled system improved the safety margin, and the escaped load increased by nearly 20%. Moreover,

such performance improvement did not increase the control force correspondingly, making it more energy efficient than the conventional controllers. Although being trained under step loads, the NN controller could reduce displacement and control force more effectively than the 3rd order polynomial controller under random excitations. This reveals the robustness and adaptability of the proposed method.

Even though DRL-based methods can offer robust and scalable solutions for nonlinear vibration control, real-world applications involve more complex characteristics, including uncertainty, noise, and operational variability of the control target. How to train agents to adapt to practical vibration scenarios effectively remains a question. In the next chapter, how to increase the practicality of DRL is investigated.

4.6 Appendix

4.6.1 Buckled beam with higher arch

Regarding the buckled beam shown in Figure 4.3, the non-symmetric effects should be taken into account when the initial deflection is higher. For a higher arch (p = 1.01 and β = 3), the second mode should also be considered. The dimensionless system equation of motion:

$$\ddot{\alpha}_1 + (1 - p + 2\beta^2)\alpha_1 + 3\beta{\alpha_1}^2 + 4\beta{\alpha_2}^2 + {\alpha_1}^3 + 4\alpha_1{\alpha_2}^2 = -v_1 + u. \quad (4.10)$$

$$\ddot{\alpha}_2 + (16 - p)\alpha_2 + 8\beta\alpha_1\alpha_2 + 16{\alpha_2}^3 + 4{\alpha_1}^2\alpha_2 = -v_2 + 2u. \quad (4.11)$$

where α_1 and α_2 represent the amplitude of the first and second modes, respectively. A small asymmetric load component v_2 is assumed as 0.5% of the symmetric load v_1. This Appendix presents the control results associated with a buckled beam with a high arch.

4.6.2 Results

The same procedure as that in subchapter 4.4.1 is used to train the NN controller for vibration control in this case. The step load adopted for training is 6, which is higher

than the escape load of the 3rd order controller ($v_1 = 5.15$). The trained TD3 agent is evaluated against the 1st order and 3rd order controller under the step load of $v_1 = 7.77$, different from the value used in the training process. Figure 4.14 compares the time response of the symmetric mode under different controllers, while Figure 4.15 presents the corresponding response of the asymmetric mode. Figure 4.16 illustrates the time history of the control force generated by each controller. Under this loading level, only the TD3 agent is able to prevent the dynamic buckling of the system while requiring a smaller control effort, highlighting the potential of DRL for controlling nonlinear vibrations involving multiple modes.

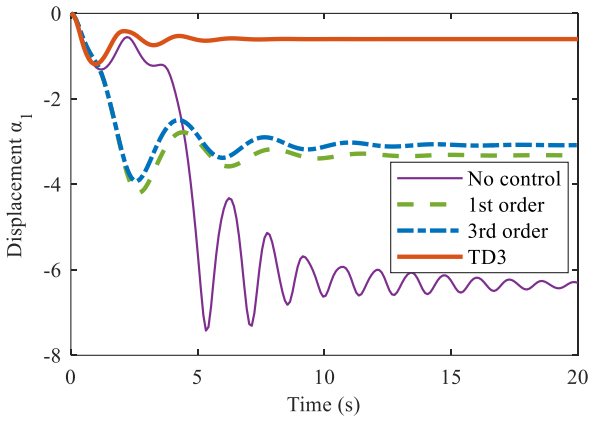


Figure 4.14 Response time histories of the symmetric mode α_1

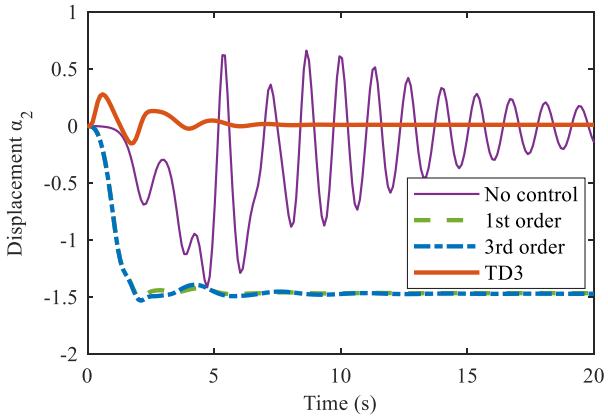


Figure 4.15 Response time histories of the asymmetric mode α_2

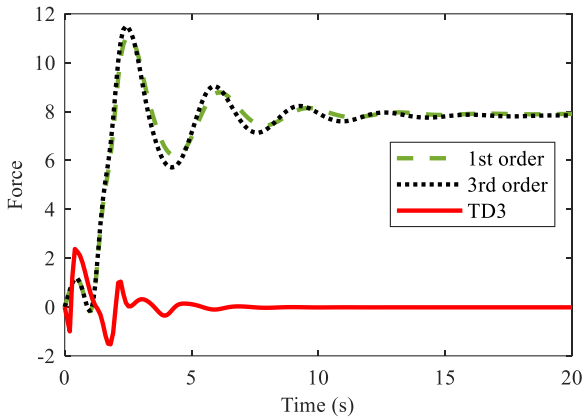


Figure 4.16 Response time histories of the control force

CHAPTER 5

A Hybrid IL-DRL Approach for Stay Cable Vibration Control

5.1 Introduction

While Chapters 3 and 4 demonstrated the direct applications of DRL to optimal structural control, several critical challenges remain unaddressed. Major among these are the substantial sample requirements and instability issues of the initial agent, which pose significant obstacles to directly applying DRL to vibration control of real-world structures. For example, the SDOF training shown in Figure 3.3 took more than 300 episodes to converge, where the agent could hardly control the system during the first 100 episodes. These findings emphasize the necessity for improved methodologies and techniques that mitigate these limitations while preserving the advantages of DRL in structural control applications. Moreover, there remains a lack of studies regarding the use of learning-based control methods to achieve optimal control performance in real structures with high degrees of freedom.

This chapter introduces a novel learning-based control framework that leverages the synergy between IL and DRL to enhance adaptability, efficiency, and practical applicability in active structural vibration control, built upon advancements in sample efficiency achieved by pre-training NN models. The introduced data-driven control algorithm is distinctively designed to function autonomously without necessitating precise knowledge of structural and actuator dynamics. The performance of the proposed control algorithm is numerically validated on a full-scale stay cable under various excitations.

Stay cables are essential load-carrying elements used in cable-stayed bridges with long spans. Stay cables are susceptible to large amplitude vibrations caused by wind or wind-and-rain actions due to their slenderness, considerable flexibility, low inherent damping, and low-frequency characteristics (Hikami and Shiraishi 1988; Matsumoto et al. 1998; Yamaguchi 1990). Various vibration control devices and methods have been studied and implemented in recent decades. For example, passive viscous dampers (Krenk 2000; Pacheco et al. 1993) have been applied, and the optimal additional damping for a specific target mode of a cable was estimated through eigenvalue analysis. Their capacity, however, is limited to roughly one-half of the distance from the cable anchorage to the damper over the full cable length (Main and Jones 2002; Xu and Yu 1998). Semi-active control approaches are considered more efficient methods for vibration mitigation and provide more remarkable damping enhancement compared with passive devices (Chen and Ni 2017; Johnson et al. 2007; Weber and Boston 2011). Active control techniques provide the best performance and can significantly reduce cable vibration (Iemura and Pradono 2005; Li et al. 2008; Shi et al. 2017). The active control forces produced by an actuator are calculated based on real-time feedback from the sensors and require advanced control algorithms to determine optimal control forces. The design of active control algorithms significantly affects control performance. Therefore, the IL-DRL controller is applied to active mitigation of the stay cable vibration in this chapter.

The primary objective of this chapter is listed as follows:

a) Design and validate a model-free, learning-based control framework that employs IL for rapid initial training and DRL for subsequent optimization. This approach aims to explore the potential of IL in providing a stable basis

for DRL. Compared with the pure DRL controller presented in Chapters 3 and 4, which requires hundreds of episodes or even more to converge to a workable but non-optimal control policy, the IL-DRL controller is expected to be able to enhance training stability and efficiency considerably.

b) Examine the framework's capability to negate the need for intricate knowledge about structural dynamics, while achieving notable sample efficiency in model-free learning. The efficacy of this combined approach will be benchmarked against traditional model-based optimal controllers using simulations on a full-scale stay cable under diverse excitations.

c) Validate the methodology's effectiveness and adaptability across scenarios with varying observation points in the presence of measurement noise and actuation force uncertainties, ensuring its robustness for practical applications.

Notably, the employed cable model is only for data generation in this study. The proposed model-free method does not utilize any model information in the training and testing, making it immune to any uncertainties in structural and actuator models that represent a great challenge in the traditional model-based control methods. By eliminating dependencies on specific models, this approach emphasizes innovation in developing robust control strategies that maintain strong performance in unpredictable real environments.

More importantly, adding the pre-training stage saves training time and avoids instability issues in the initial random exploration stage of pure DRL training, making the IL-DRL control method more feasible for practical applications than the pure DRL controller introduced in Chapters 3 and 4.

5.2 Model-free IL-DRL Controller

This chapter proposes a model-free IL-DRL controller designed to control structures without requiring precise knowledge of system dynamics. Meanwhile, this approach tackles the challenges associated with pure DRL, specifically demanding a large amount of data. Notably, without any structural dynamic models, DRL-based control methods aim to learn control strategies from real structural responses or structural tests.

However, the need for large numbers of training samples requires many structural tests to be conducted, making pure DRL methods impractical for real-world implementations. For example, as presented in Chapters 3 and 4, pure DRL methods need hundreds of iterations to converge even for SDOF systems. To address this issue, the learning-based control framework introduced in this section comprises two main components: IL for agent pre-training using a simple control strategy and DRL for subsequent fine-tuning. By employing IL-based pre-training, this hybrid method can improve sample efficiency and reduce training time, ultimately enabling practical applications in the field of structural control. Figure 5.1 provides an overview of the proposed IL-DRL controller.

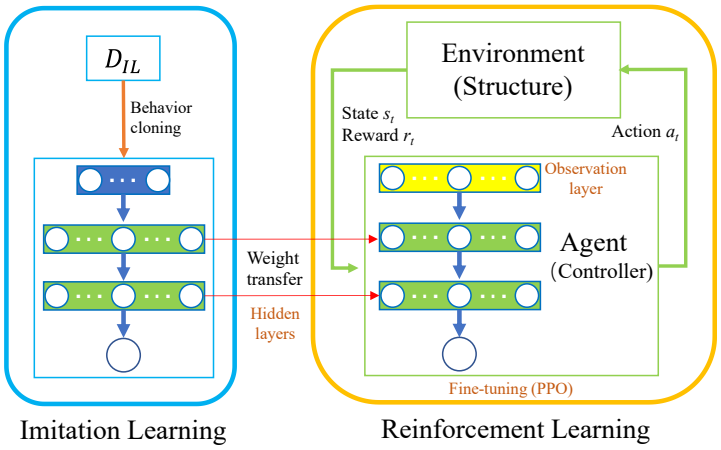


Figure 5.1 Illustration of the model-free IL-DRL controller

5.2.1 IL for pre-training

In this chapter, the IL is primarily adopted to pre-train a controller before applying DRL, rather than random initialization of the agent. This pre-training can increase sample efficiency and reduce total training time. Specifically, behavioral cloning (BC), the most basic form of IL, is adopted herein to replicate a straightforward control strategy without relying on any prior knowledge of structural dynamics.

To collect training data for controlling the target structure, a simple demonstration control strategy is initially implemented to mitigate vibrations under random excitation. Given the unknown system model, there is no specific criterion for selecting the control strategy. The demonstration strategy is deemed acceptable if it

increases the structure's damping ratio while keeping the control force within the actuator's range. A negative stiffness damper (NSD) that combines negative stiffness and viscous damping in parallel is adopted as a reference strategy because a passive NSD has been proven effective in vibration control of stay cables by Shi et al. (2016). More details about this reference strategy will be provided in subchapter 5.3.2. The demonstration dataset, denoted by D_{IL}, is compiled by recording feedback observations $\mathbf{s}_t$ and applied control forces $\mathbf{a}_t$ at each time step and then constructing state-action pairs ($\mathbf{s}_t$, $\mathbf{a}_t$) to be used in training. No structural dynamics is required at this stage.

The policy $\mu_\theta(\mathbf{s}_t)$ was formulated as a DNN that employs supervised learning to minimize the MSE error $\varepsilon(\theta)$ between the demonstration control strategy and the network outputs.

$$\varepsilon(\theta) = \frac{1}{|D_{IL}|}\sum_{(\mathbf{s}_t,\mathbf{a}_t)\in D_{IL}} \frac{1}{2}[\mu_\theta(\mathbf{s}_t) - \mathbf{a}_t]^2, \tag{5.1}$$

where $\mathbf{a}_t$ is the recorded control force at time t in the demonstration strategy, $\mu_\theta(\mathbf{s}_t)$ is the NN's output control force given the observation state $\mathbf{s}_t$ at time t, and the term $1/|D_{IL}|$ computes the average loss across all samples. Therefore, Equation 5.1 is essential to minimize the difference between the recorded control force $\mathbf{a}_t$ in the demonstration control strategy and the predicted control force $\mu_\theta(\mathbf{s}_t)$ by the NN controller, so that the NN controller can replace the demonstration controller in the control loop. However, in real-world scenarios where encountered states fall outside the training dataset, the IL-trained NN may fail due to accumulated errors. Dataset aggregation algorithms resolve this problem by incorporating data for previously unvisited states. Although the pure BC method is adopted here, it is noted that the subsequent DRL fine-tuning can rectify any resultant errors.

The pre-trained control network's weight was transferred to a new DRL agent for fine-tuning, as shown in Figure 5.1. The connections between the hidden layers and the current observation nodes in the top layer were retained, while the number of nodes in the input layer was adjustable to accommodate more observation points or previous measurements. The connections between the newly added nodes and the first hidden layer were initialized to zero, ensuring that the IL policy would not degrade.

5.2.2 DRL-based fine-tuning

Unlike the initial IL, DRL aims to learn the optimal control strategy by interacting with the environment, instead of attempting to match a presumed strategy. The control problem in DRL is modeled as a Markov decision process, wherein the next state depends only on the current state and action. At each time step t, the agent observes state $\mathbf{s}_t$, selects an action $\mathbf{a}_t$, receives a reward $r_t = r(\mathbf{s}_t, \mathbf{a}_t)$, and the environment then transits to the next state $\mathbf{s}_{t+1}$. The objective of DRL is to learn a policy that maximizes the expected sum of discounted future rewards (Equation 3.16).

Numerous DRL algorithms have been developed, showcasing remarkable performance in complex control problems. This study follows the model-free DRL framework, whose objective is to optimize the policy network directly without learning the environment's dynamics model. Considering the potential instability in active control of stay cables when higher vibration modes are excited, this study uses PPO to fine-tune the pre-trained policy, as it offers a balance between performance and stability during training. As an improved version of the on-policy policy gradient algorithm, the TRPO maintains the new policy close to the old while progressing. This choice is preferred over other SOTA off-policy algorithms, such as SAC and TD3 policy gradient, which are known for their high sample efficiency. In contrast to off-policy algorithms (such as DDPG used in Chapter 3 and TD3 used in Chapter 4) that necessitate critic or value function initialization and use experience buffers, PPO does not require these components. This feature makes it more convenient to preserve the pre-trained agent's functionality in precarious active control problems (e.g., stay cables), where off-policy algorithms may encounter difficulties. Therefore, PPO is chosen in this chapter. The pre-trained control network serves as the initial policy for PPO, but the framework is flexible enough to incorporate other model-free algorithms, if necessary.

The negative quadratic cost function was selected as the reward, as

$$r(\mathbf{s}_t, \mathbf{a}_t) = -(\mathbf{s}_t^T \mathbf{Q} \mathbf{s}_t + \mathbf{a}_t^T \mathbf{R} \mathbf{a}_t) \tag{5.2}$$

which is also presented as Equation 3.19 in Chapter 3.

By maximizing the long-term reward, the optimal solution for the linear quadratic control problem can be obtained using the negative quadratic reward function. The collected state-action pairs, together with the cable's vibration, are used to update the policy through the PPO algorithm. The final fine-tuned policy network operates independently without the critic network as an NN controller, thus replacing the classical controller

5.3 Numerical Examples

Using a numerical example of a full-scale stay cable, this subchapter assesses the performance of the suggested intelligent model-free learning-based controller, wherein the simulations were designed to demonstrate the proposed controller's ability to effectively mitigate stay cable vibrations while minimizing the control force required. The results were then compared to those obtained with the classical model-based optimal controllers to determine the effectiveness and efficiency of the intelligent controller in achieving the desired control objectives. The training was done on an Intel i7-9700 CPU.

5.3.1 Cable configuration

The simulation is based on a full-scale (135 m long) stay cable depicted in Figure 5.2, which was manufactured by OVM Co. Ltd. (based in Hubei, China) and served as the cable model. The cable's dynamic behavior was tested using various passive and semi-active control strategies in experiments (Li et al., 2019; Li et al., 2020); in their studies, the modeling accuracy was validated. The cable was discretized into 100 uniformly distributed segments, with a total of 99 internal nodes. The actuator was located at the 3% position from the anchorage, which was connected to the third node in the numerical model. The feedback signal included 8 observations, i.e., the displacements and velocities at four locations: the damper position and the 1/4, 1/2, and 3/4 span positions. Table 5.1 shows the cable parameters used in the simulations. These specifications were utilized to assess the proposed learning-based controller's performance and compare it to that of the classical optimal controller, as previously described in subchapter 3.2.

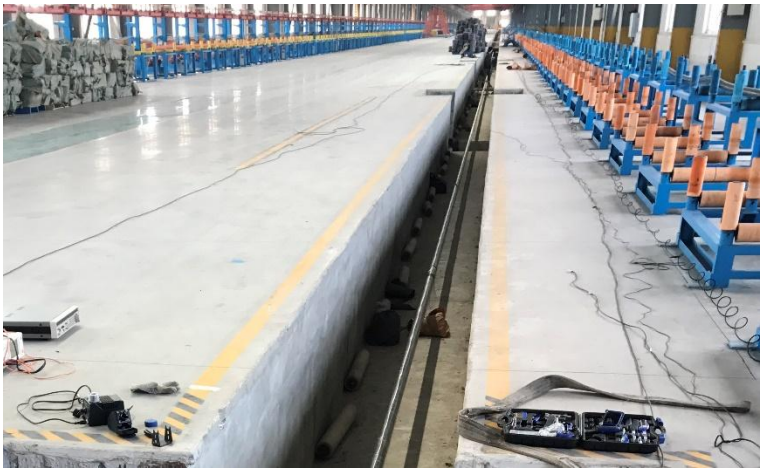


Figure 5.2 Photograph of the cable

Table 5.1 Cable parameters

Item	Value
Length (m)	135
Diameter (cm)	8.5
Cross-section area (cm^2)	57
Moment of inertia ($N \times m^2$)	2.56×10^6
Young's modulus (N/m^2)	1.95×10^{11}
Mass per unit length (kg/m)	16.65
Tension (kN)	1,200

Figure 5.3 illustrates the configuration of the stay cable system. The following partial differential equation of motion describes the transverse vibration of the system, when subjected to external excitation:

$$T\frac{\partial^2 v(x,t)}{\partial x^2} - m\frac{\partial^2 v(x,t)}{\partial t^2} - c\frac{\partial v(x,t)}{\partial t} - \frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 v(x,t)}{\partial x^2}\right) = w(x,t) + f(t)\delta(x - L_d) \quad (5.3)$$

where $v(x,t)$ is the transverse displacement at location x and time t, m is the mass per unit length, T is the tension force, c is the damping per unit length, EI is the flexural rigidity, L is the total length, L_d is the distance of the damper to the left anchorage, δ is the Dirac delta function, f denotes the damper force, and w represents the excitation.

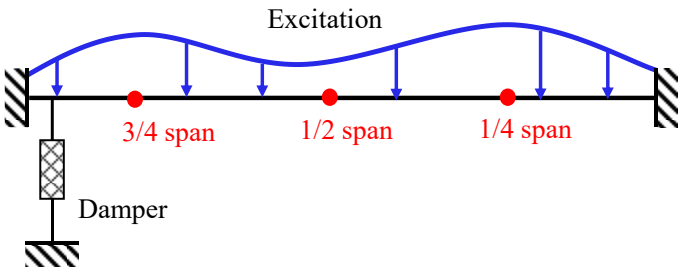


Figure 5.3 Configurations of cable-damper system

This study uses the finite-difference approach (Mehrabi and Tabatabai, 1998) to discretize Equation 5.3 into n+1 elements with element length l. Various studies provided further descriptions and validation of this method (Shi et al., 2017; Shen and Zhu, 2015). The stiffness matrix of an uncontrolled stay cable with a uniform cross-section can be written as

$$\mathbf{K} = \begin{bmatrix} H & W & V & & & 0 \\ W & S & W & V & & \\ V & W & S & W & \ddots & \\ & V & \ddots & \ddots & \ddots & V \\ & & \ddots & \ddots & S & W \\ 0 & & & V & W & H \end{bmatrix}. \tag{5.4}$$

where,

$$S = \frac{6EI}{l^3} + \frac{2T}{l}, W = -\frac{4EI}{l^3} - \frac{T}{l}, V = \frac{EI}{l^3} \tag{5.5}$$

For fixed ends

$$H = \frac{7EI}{l^3} + \frac{2T}{l} \tag{5.6}$$

The mass matrix can thus be expressed as the lumped mass model:

$$\mathbf{M} = \mathbf{I}_N \cdot ml \tag{5.7}$$

where $\mathbf{I}_N$ is an $n \times n$ identity matrix. The damping matrix $\mathbf{C}$ is constructed following Rayleigh damping. The equation of motion in matrix form is:

$$\mathbf{M}\ddot{\mathbf{v}} + \mathbf{C}\dot{\mathbf{v}} + \mathbf{K}\mathbf{v} = \mathbf{w} + \mathbf{f} \tag{5.8}$$

The corresponding state matrix $\mathbf{A}$ in Equation 3.1 is:

$$\mathbf{A} = \begin{bmatrix} 0 & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix} \tag{5.9}$$

$\mathbf{B_f}$ is

$$\mathbf{B_f} = \begin{bmatrix} 0 \\ \mathbf{M}^{-1}\boldsymbol{\gamma} \end{bmatrix} \tag{5.10}$$

where

$$\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \cdots, \gamma_n]^T$$

$$\gamma_i = \begin{cases} 0, i \neq j \\ 1, i = j \end{cases} \tag{5.11}$$

The active damper is located at the *j*th node ($j = 3$ in this study). $\mathbf{B_w}$, expressed as:

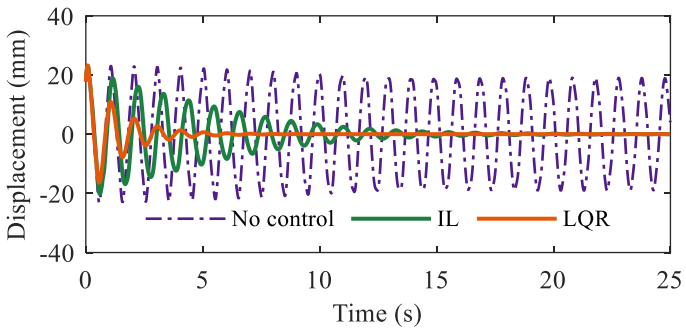
$$\mathbf{B_w} = \begin{bmatrix} 0 \\ \mathbf{M}^{-1} \end{bmatrix} \tag{5.12}$$

It is noteworthy that the above-described cable model is intended to generate vibration response data of a cable only. Such a dynamic model of a cable is no longer needed, if the cable's response data can be collected by sensors in real-world scenarios. The controller can autonomously compute control forces based on sensor feedback signals. Therefore, a major advantage of the proposed model-free method is its immunity to the inaccuracy of modeling in structural dynamics (or actuator dynamics).

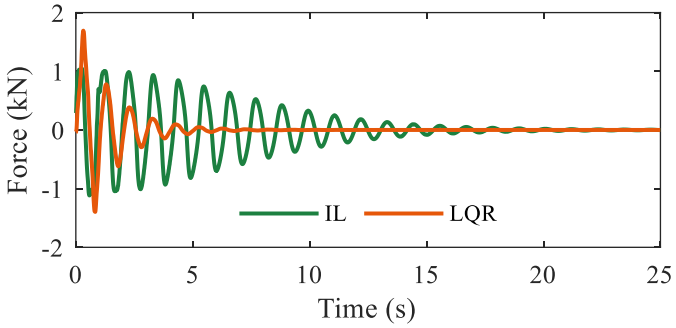
5.3.2 Pre-training stage

The dataset used for IL was collected by sampling the observations and actuator force of an NSD, which does not require the cable model. The actuator force $f_0 = -\mathbf{G}_0[\mathbf{v}_d, \dot{\mathbf{v}}_d]^{\boldsymbol{T}}$ was calculated based on inputting a series of random displacements and velocities to the damper with the control gain $\mathbf{G}_0 = [-220, 2]$. These coefficients of the NSD could increase the stay cable's damping ratio to some extent. However, because the cable model was assumed unknown, these coefficients were not deliberately optimized and thus could not provide optimal damping forces to this cable. The training dataset contained 2.5×10^5 pairs of observations (damper displacements and velocities) and control forces.

The training of the NN model followed the previously outlined procedure, wherein the NN model contained 2 hidden layers, each with 64 dimensions and Tanh nonlinearities. The policy was trained using the Adam optimizer with a batch size of 1028 and a learning rate of 0.001.



(a)



(b)

Figure 5.4 Control effect of pre-trained NN controller on free vibration of stay cable: (a) displacement at mid-span (b) control force.

Subsequently, the pre-trained NN controller can be applied to control the cable vibration. Figure 5.4 compares the control performance of the IL-trained NN controller and the LQR controller under free vibrations. Figure 5.4a presents the displacement response time histories at the middle span of the stay cable, and Figure 5.4b shows the time histories of the control forces. The weighting matrix of LQR was designed to be $\mathbf{Q} = \mathbf{I}$ (8×8) and $\mathbf{R} = 10^{-6}$. Although the IL-trained NN controller successfully lowers vibration levels compared with an uncontrolled cable, it underperforms in displacement reduction after the first cycle when measured against an optimal LQR controller. Meanwhile, the IL-trained NN controller used large control

forces most of the time during the free vibrations, implying more energy consumption. This performance gap arises because the IL-trained NN controller has not been optimally tuned based on system characteristics.

5.3.3 Fine-tuning stage

Given the poor performance of the IL-trained initial NN controller, fine-tuning in the second stage is inevitable. To implement DRL for stay cable control, the cable was subjected to random white noise excitation and then suddenly released to undergo free vibrations. In real applications, such free vibration tests can be excited and repeated on real cables, and the corresponding responses can be measured. Because the structural dynamic equation is unknown, the total time steps of each episode were extended to be sufficiently long to guarantee the complete attenuation of cable vibrations and ensure control stability. Exclusive reliance on training under random vibration could not ensure stability.

It is essential to emphasize that only the simulated responses served as the input in training without utilizing any model information. This approach is independent of the underlying system dynamics knowledge, so it can be regarded as a model-free method.

The policy acquired through IL was used as the initial agent to implement DRL. The weights and biases from IL were transferred directly to the hidden layers, while modifications were made to the observation layer by incorporating more observations from the current and previous steps. The connections of the current-step observation and the first hidden layer remain unchanged, while the connections with the previous-step observation were unavailable in IL and initially set to zero. Consequently, the control force's initial contribution predominantly originates from the current-step observation.

Table 5.2 compares the performance between the LQR controller, the initial IL agent, and various observation combinations. A total of 15 tests were conducted to account for the inherent stochasticity in DRL methods. The training is considered successful if it achieves at least 90% of the LQR controller's performance (i.e, around 110% of the LQR's cost). After comparing the results of these 15 random trials, the

final observation adopted the observation of the current step combined with the previous step ($\mathbf{y}_t$, $\mathbf{y}_{t-1}$) due to its relatively lower cost and highest success rate.

The concept of 'number of successful cases' should be interpreted in the context of DRL's intrinsic variability and trial-and-error nature. A low number of successful cases is not necessarily indicative of a flawed model but rather reflects the explorative and stochastic characteristics of DRL. Given the limitations of a small state observation space in this study, a broader set of input states was chosen. This decision aimed at allowing the agent to capture more of the system's intrinsic dynamics. It was hypothesized that additional input could improve the agent's learning effectiveness. However, it's important to note that an increase in input states does not guarantee better performance. The optimal selection of observation states remains an unanswered question and offers a direction for future research.

Table 5.2 Performance comparison between different observations input

Controller	No. of time steps	Input state variables	Cost J	No. of successful cases
LQR	1	$\mathbf{y}_t$	271	
IL (initial agent)	1	$\mathbf{y}_t$	638	
IL-DRL (fine-tuned)	1	$\mathbf{y}_t$	294	1
	2	$\mathbf{y}_t$, $\mathbf{y}_{t-1}$	282	7
	3	$\mathbf{y}_t$, $\mathbf{y}_{t-1}$, $\mathbf{y}_{t-2}$	286	4

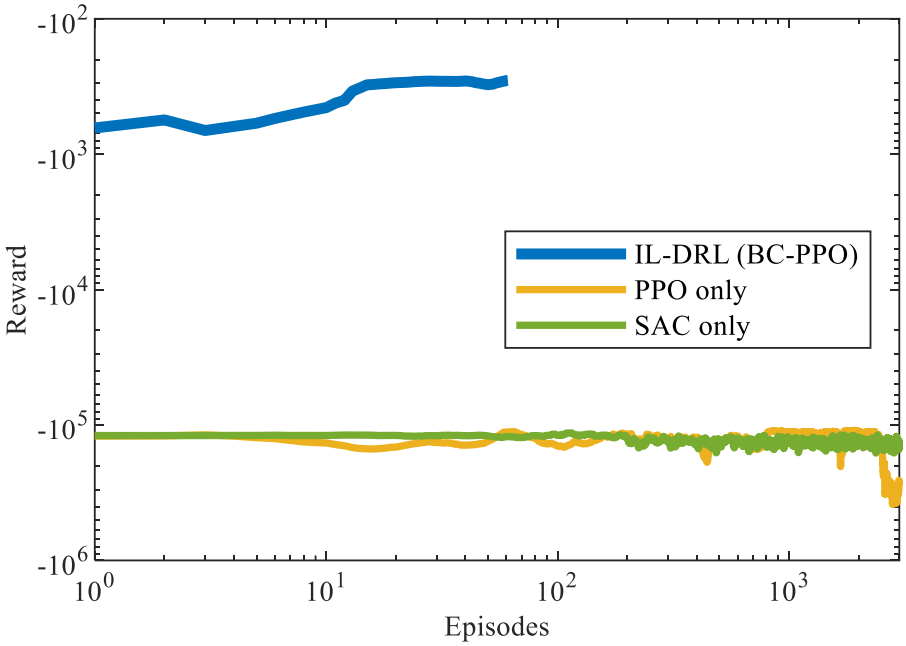


Figure 5.5 Average reward comparison

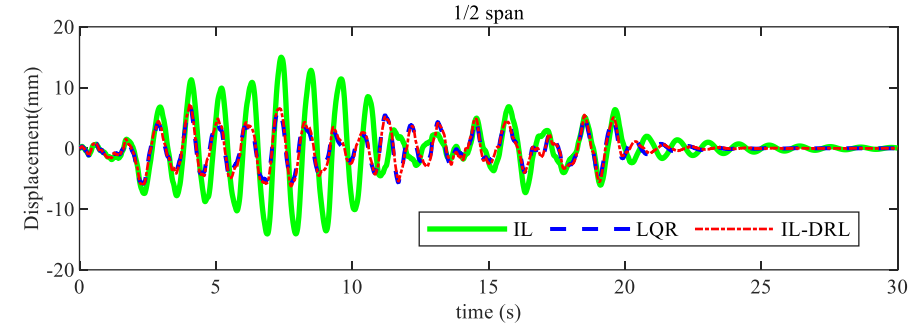
Figure 5.5 compares the average reward (moving average of 5 episodes) of the proposed IL-DRL approach (with pre-training) with that of pure model-free DRL algorithms PPO and SAC (without pre-training). By partially transferring the weights of the initial agent from the imitation learned NN controller, the proposed agent achieves a reward very close to the value of the model-based LQR controller in less than 100 episodes (5×10^6 sets of data). Given a sampling frequency of 100 Hz, a 50-s time history of cable response was used in each episode. The average training time is half an hour. In contrast, the pure PPO and SAC agents still struggle to find a viable strategy even after more than 3000 episodes, which is unaffordable in real cable tests. This result demonstrates the proposed approach's effectiveness in improving the sample efficiency of DRL training for active control of stay cables.

5.4 Random Excitation Performance

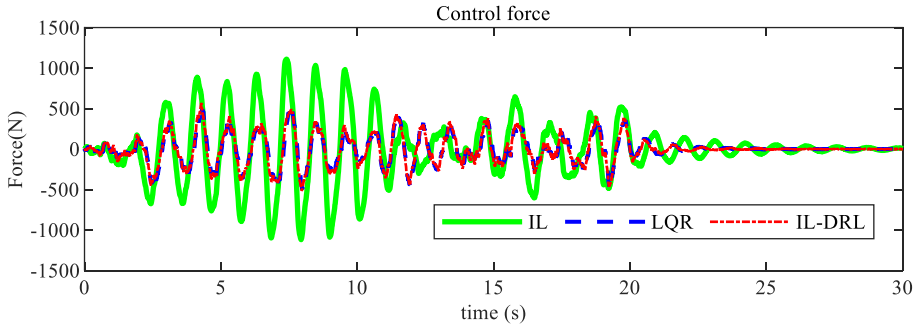
After training under free vibrations, the IL-DRL controller was directly tested under random excitations. Figure 5.6a and 5.6b compare the mid-span displacement response and control force under random excitations, respectively. A white noise excitation force was applied to all nodes for the first 20 seconds, and the cable performed free vibrations after 20 seconds to evaluate the stability of the controllers.

As shown in Figure 5.6a, the performance of the IL-DRL controller was extremely close to the optimal LQR controller, indicating the proposed approach's effectiveness. The improvement was significant when comparing the IL-DRL controller to the IL controller without fine-tuning, which shows that fine-tuning the pre-trained policy network through DRL could improve its performance considerably.

Figure 5.7 presents the displacement-force relationships of the IL-DRL and LQR controllers at the damper position, both of which show similar negative stiffness features. The RMS displacements of the IL-DRL controller are 1.69, 2.40, 1.80, and 0.73 mm at the 1/4, 1/2, 3/4 span and the damper position, respectively. Meanwhile, the corresponding values of the LQR controller are 1.69, 2.37, 1.72, and 0.70 mm, respectively, slightly lower than those of the IL-DRL controller, as shown in Figure 5.8. But the fine-tuned IL-DRL controller outperformed the initial pre-trained IL controller considerably.



(a)



(b)

Figure 5.6 Cable response time histories under random excitations: (a) displacement at mid-span (b) control force

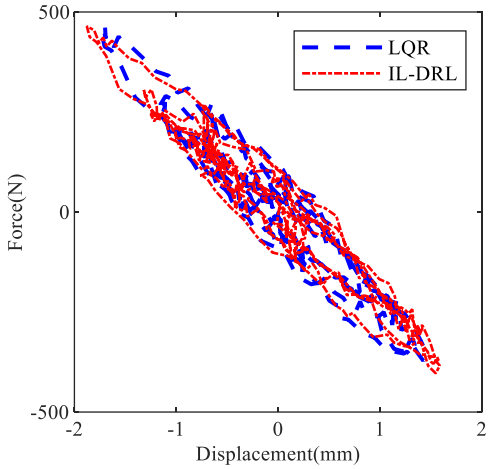


Figure 5.7 Control force-displacement relationships under random excitations

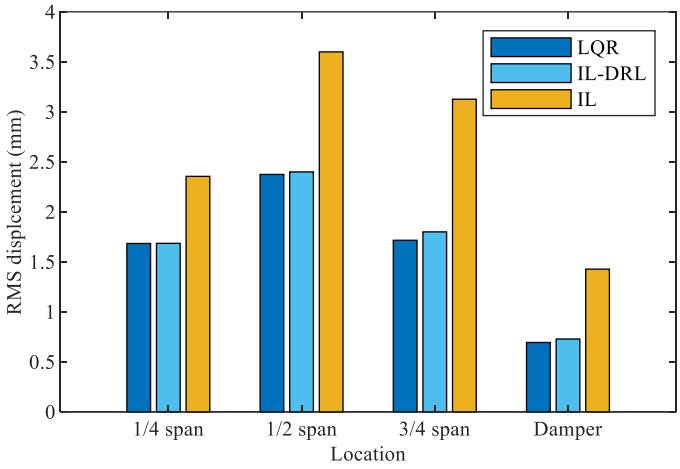


Figure 5.8 RMS displacement at various locations

Figure 5.9 shows the effect of control weight **R** on the control performance of the proposed IL-DRL controller under the same white noise excitation. The parameters of PPO were kept the same for all cases. The blue dots represent the RMS displacements at mid-span, and the red triangles indicate the RMS control force. As **R** decreases from 10^{-4} to 10^{-6}, the displacement decreases while the control force increases, which parallels the trend observed for the LQR controller. Figure 5.10 compares the RMS displacement reduction by our IL-DRL controller and the classical LQR controller at mid-span with different **R** values. The IL-DRL controller's displacement reduction tends to be more when **R** is larger than 10^{-5}. This may be because the maximum control force used in the IL pre-training is close to the LQR controller when **R** equals 10^{-5}. Notably, the hyperparameters are set the same for all cases in this chapter. Case-

specific hyperparameter tuning could enhance the performance for different **R** weights.

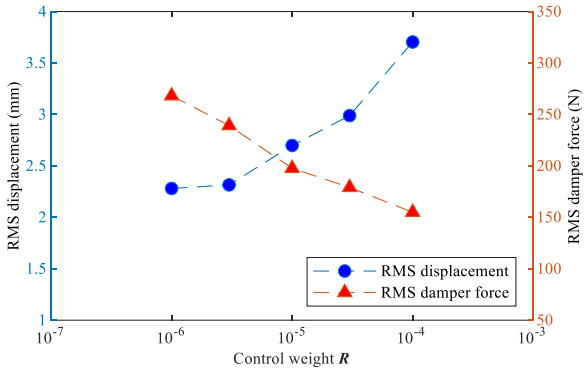


Figure 5.9 RMS cable response and RMS damper force of the IL-DRL controller regarding different control weights

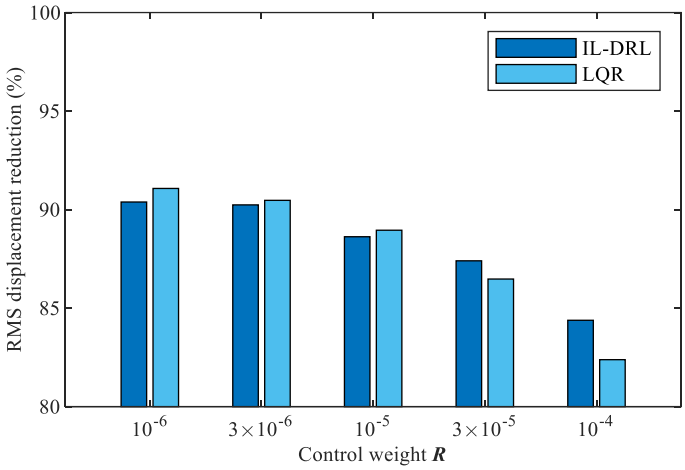
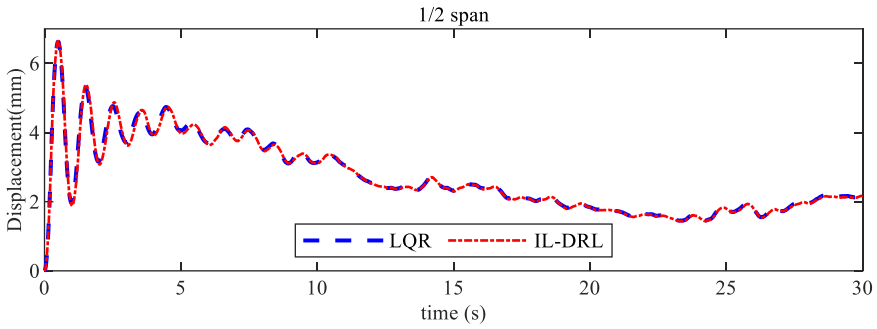


Figure 5.10 RMS displacement reduction with different control weights

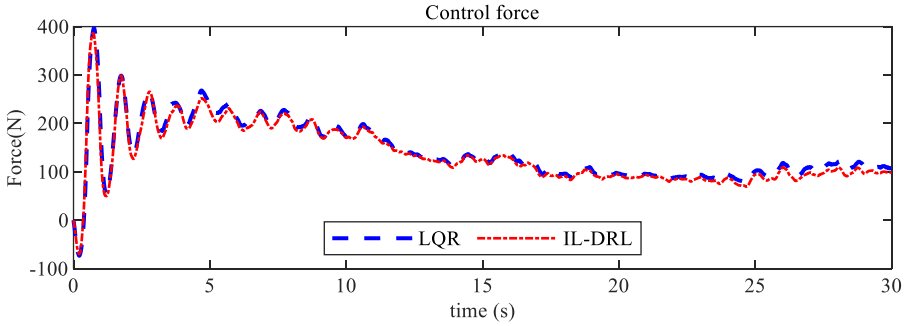
Figure 5.11 displays the mid-span displacement responses and control forces under wind force excitations, where the wind load used herein included the mean and dynamic wind components with a mean velocity of 10 m/s. Due to the quasi-static mean component, the response of the cable is not centered around zero, as seen in Figure 5.11a. As the IL-DRL controller is fine-tuned to achieve optimal performance, its control effect, shown in Figure 5.11b, is almost indistinguishable from that of the analytically derived LQR controller that was designed with the same control weights. Therefore, the proposed IL-DRL controller can achieve optimal mean and dynamic control performance.

Given their close control performance, the major differences between the IL-DRL

and LQR controllers should be noted: (1) the former is a model-free method, while the latter is model-based; (2) the former uses only eight state variables as feedback, while the latter uses the full state (198 state variables); and (3) the former uses the feedback in two steps, while the latter uses one step only.



(a)



(b)

Figure 5.11 Cable response under wind excitation: (a) displacement at mid-span (b) control force

5.4.1 Effect of the number of feedback state variables

To further evaluate the proposed controller's performance, the number of observer points was reduced to two (mid-span and damper locations), and the feedback signal included the displacement and velocity at the two measurement points. The IL training setup was the same as before. For the DRL fine-tuning, the only change made to the NN structure was the reduction of the input layer's node numbers. The **Q** in the reward function was adjusted to $\mathbf{Q} = \mathbf{I}$ (4×4) to match the number of observations. The training plot of the episode reward and the average score of 10 episodes are shown in Figure 5.12. It took 140 episodes to reach the optimal reward, slightly longer than the previous

cases using four observation points. However, compared with the performance of pure DRL training by using the SOTA algorithms shown in Figure 5.5, this increase is nonetheless negligible.

Figure 5.13 shows the mid-span displacement time history and the control forces of different controllers under 30-second random excitations. While the force of the LQR controller at peaks is slightly larger than that of the IL-DRL controller, the overall control effects of the two controllers are in good agreement. This emphasizes the effectiveness of our proposed IL-DRL method, which can achieve optimal control performance using only four state variables as feedback, compared to the full-state feedback (198 state variables) required by LQR.

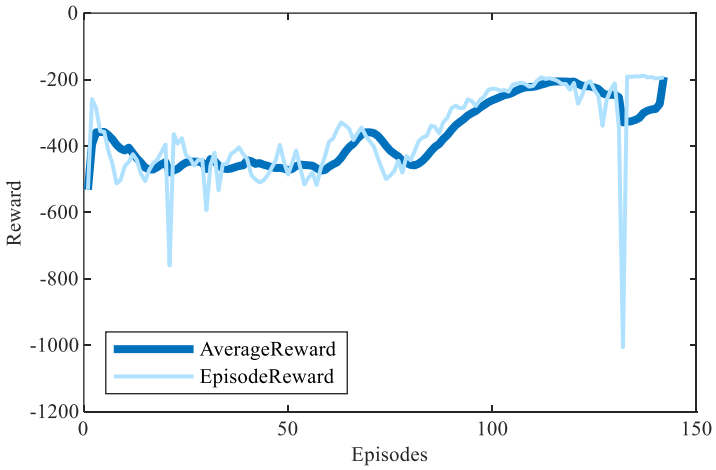
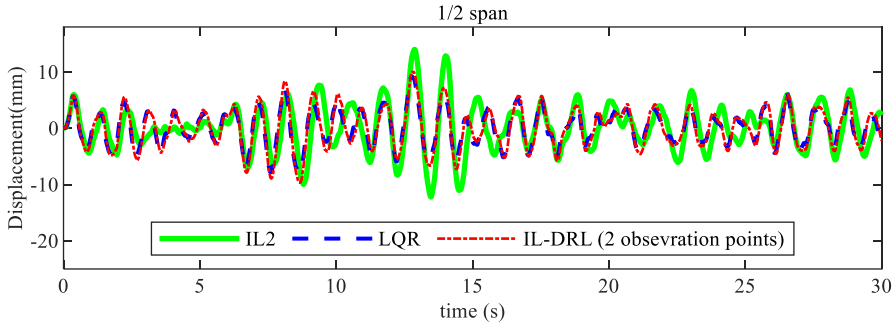
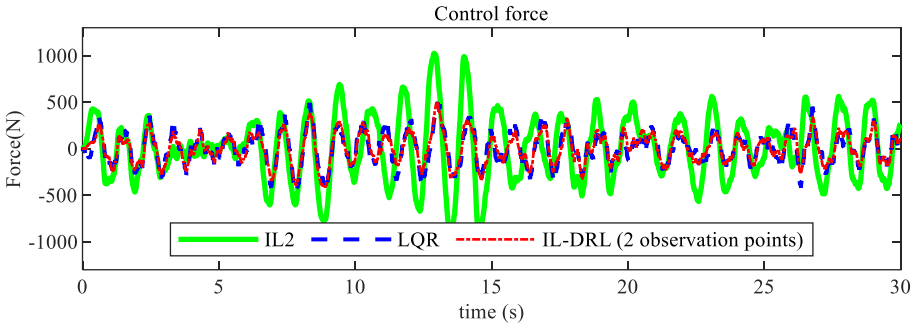


Figure 5.12 Fine-tune training plot of the IL-DRL controller with 2 observation points



(a)



(b)

Figure 5.13 Cable response under random vibration (2 observation points): (a) displacement at mid span (b) control force

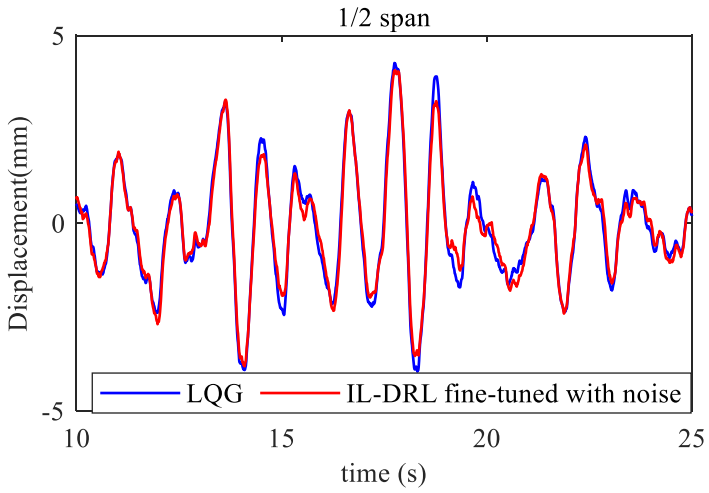
5.4.2 Effect of measurement and actuation noise

To assess the robustness of the proposed IL-DRL method, Gaussian white noise is added to all eight observed states during the DRL training as measurement noise, while maintaining the pre-trained network. The noise-to-signal ratio is set as 5%. Meanwhile, considering the potential uncertainty in actuator dynamics, 10% noise is added to the actual control force applied to the cable at each step in the performance testing. The improved agent was then compared with the optimal LQG feedback controller using a consistent control weight matrix, aiming to examine the effect of measurement noise and unpredictable actuator dynamics on the control performance.

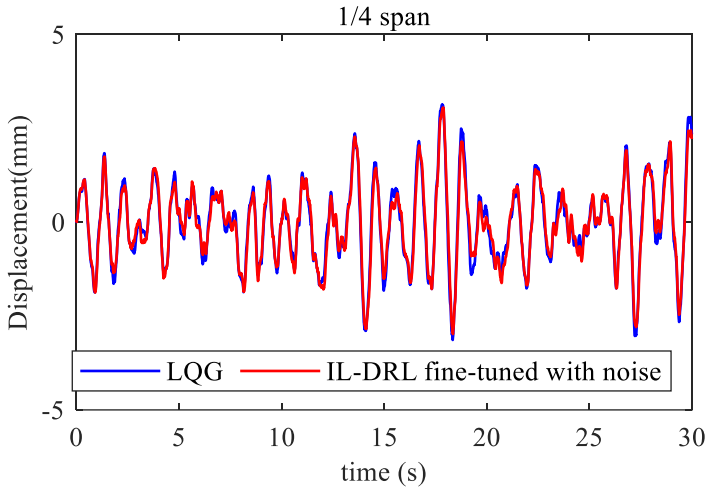
Figure 5.14a and 5.14b illustrate the displacement responses at mid-span and the 1/4 span when the cable is randomly excited. Here, the IL-DRL control outperforms the LQG controller in displacement reduction. Because the fine-tuning process involves measurement noise and the PPO algorithm inherently incorporates a level of randomness in force generation, the IL-DRL controller is more resistant to

measurement noise and actuation uncertainty. In contrast, the control force of the LQG controller relies on the full-state estimation through the Kalman filter, and the performance is slightly more sensitive to the noise estimations and force discrepancies.

Figure 5.15 compares the RMS displacement at four observation points. The IL-DRL controller fine-tuned with noisy measurements outperforms the LQG controller on every observation point except the damper position, illustrating the robust performance acquired through DRL training.



(a)



(b)

Figure 5.14 Cable response under random vibration with measurement noise and force discrepancy: (a) displacement at mid span, (b) displacement at 1/4 span

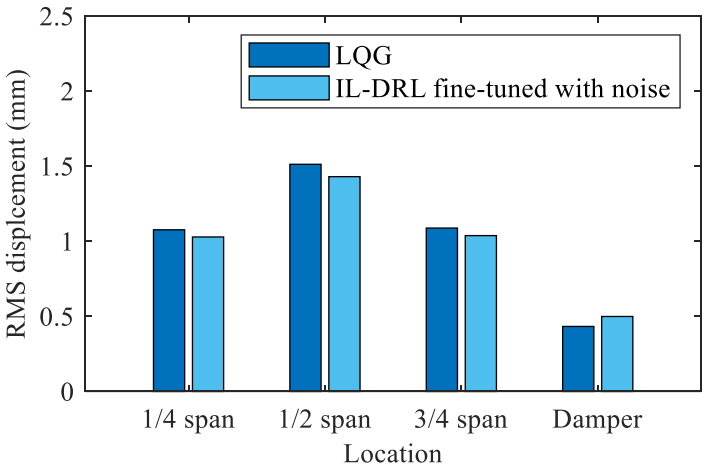


Figure 5.15 RMS displacement at various locations under random vibration with measurement noise and force discrepancy

The above analyses reveal the advantages of the model-free attribute of the proposed IL-DRL approach in addressing various uncertainties in vibration control. In contrast to conventional model-based methods, the proposed method sidesteps the need for accurate knowledge of system dynamics; instead, it learns optimal strategies directly from the data. The DRL fine-tuning phase conducted in an actual application environment allows the proposed method to adapt to unforeseen discrepancies and variances in real-world scenarios. The adaptability and robustness ensure that the IL-DRL controller remains effective, even in the presence of unanticipated uncertainties in system and actuator dynamics.

5.5 Summary

Based on the pure DRL controller developed in Chapters 3 and 4, this chapter presents an improved methodology for controlling structural vibrations, eliminating the need for dynamic system models by integrating IL and DRL technologies. The methodology is notably time-efficient in training. Compared with the pure DRL methods, the IL-DRL method can reduce training time to less than 1/30, rendering it more applicable for real-world applications. Its efficacy and practicality are validated through rigorous testing on a full-scale stay cable equipped with an active damper.

Initially, an NN controller is trained through BC to mimic a simple control strategy, followed by fine-tuning through DRL. In scenarios with random excitations, the

control performance of the proposed model-free strategy closely approximates that of the full-state model-based feedback controller, LQR, despite operating with limited state variables. Additionally, the IL-DRL controller exhibits higher resistance to measurement noise and actuation force uncertainty than the optimal LQG controller. This indicates its robustness and adaptability, primarily attributed to the DRL fine-tuning process, enabling the controller to adapt itself to the characteristics of the structure in a real-world environment. In summary, this research successfully establishes an innovative, robust, and efficient IL-DRL controller that excels in both performance and practicality.

The framework offers a new alternative for active vibration control, making it a compelling candidate for practical applications. In the following chapters, this IL-DRL approach will be further tested in a series of experiments.

CHAPTER 6

Active Vibration Control based on IL-DRL:

Experimental Implementation in an SDOF Structure

6.1 Introduction

In Chapter 5, the IL-DRL control framework is proposed to solve the problem that pure DRL typically requires a large amount of training data.

Building on the development in Chapter 5, this chapter extends the proposed IL-DRL framework to experimental validation, aiming to demonstrate its potential applicability in real-world structural vibration control. This approach is experimentally validated on an AMD-controlled SDOF structure to bring IL-DRL-based vibration control from virtual environments to real-world applications. In the IL-DRL approach, an NN controller is first pre-trained via IL by mimicking a simple reference strategy, ensuring that the initial policy is reliable to be applied in physical experiments directly. Subsequently, the controller is fine-tuned through DRL by interacting with the actual structure directly. The entire training process is conducted without requiring any prior knowledge of structural dynamics. To evaluate the performance, the proposed controller is compared with a classical analytic optimal controller, LQG, under different excitation cases. The results demonstrate the potential of the proposed method to be

further extended to MDOF structures and a wide range of vibration control applications.

The remainder of this chapter is organized as follows. Subchapter 6.2 describes the system used in this study, followed by a brief description of the adopted IL-DRL controller in subchapter 6.3. The experimental validation results and analysis are presented in subchapter 6.4.

6.2 SDOF-AMD System

Figure 6.1 shows the schematic representation of an SDOF frame structure equipped with an AMD system to be investigated in this study, where x_c is the relative displacement between the AMD mass to the floor, and x_f is the relative displacement of the floor to the ground, F_c is the AMD control force, and $\ddot{x}_g$ is the ground acceleration.

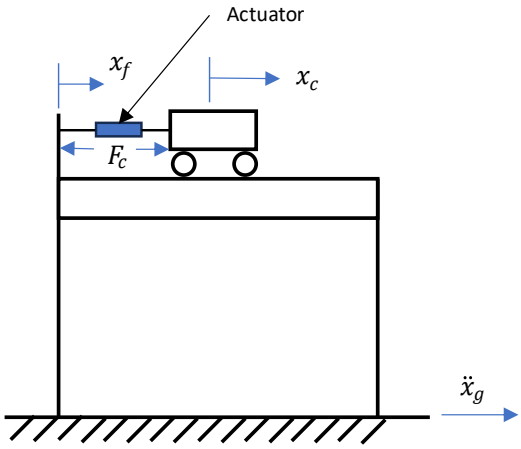


Figure 6.1 The SDOF-AMD system

The state-space representation of this system can be written as:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u + \mathbf{B}_g\ddot{x}_g + \mathbf{w}, \tag{6.1}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}u + \mathbf{D}_g\ddot{x}_g + \mathbf{v}, \tag{6.2}$$

where $\mathbf{x} = [x_c, x_f, \dot{x}_c, \dot{x}_f]$ is the state vector, $\mathbf{y}$ is the observed state depending on the sensor installation, $\mathbf{A}$ is the system matrix, $\mathbf{B}$ is the control input matrix, $\mathbf{B}_g$ is the ground excitation input matrix, $\mathbf{C}$ is the system output matrix, $\mathbf{D}$ is the control input direct feedthrough matrix, $\mathbf{D}_g$ is the ground excitation feedthrough matrix, u is the

control signal input to the AMD. In this case, the control force F_c is assumed to be a linear superposition of two terms

$$F_c = -p_1\dot{x}_c + p_2 u, \tag{6.3}$$

where the coefficients p_1 and p_2 are related to the AMD parameters. The first term represents the equivalent viscous damping force due to the AMD motion, and the second term represents the force component proportional to the control signal. If the above state-space model is known, classical model-based control strategies, like the LQG, can be employed. The LQG controller combines a Kalman filter for state estimation and an LQR for optimal feedback gain as

$$u = -\mathbf{K}_{\text{lqg}}\hat{\mathbf{x}}, \tag{6.4}$$

where $\hat{\mathbf{x}}$ represents the estimated system state provided by the Kalman filter, $\mathbf{K}_{\text{lqg}}$ is the optimal feedback gain that minimizes the quadratic cost function

$$J_{\text{lqg}} = \int (\mathbf{y}^T \mathbf{Q}_{\text{lqg}} \mathbf{y} + u^T R_{\text{lqg}} u) dt, \tag{6.5}$$

where $\mathbf{Q}_{\text{lqg}}$ and R_{lqg} are weighting matrices for output states and control energy, respectively. The feedback gain $\mathbf{K}_{\text{lqg}}$ is derived by solving the ARE. Figure 6.2 illustrates the feedback control diagrams for LQG.

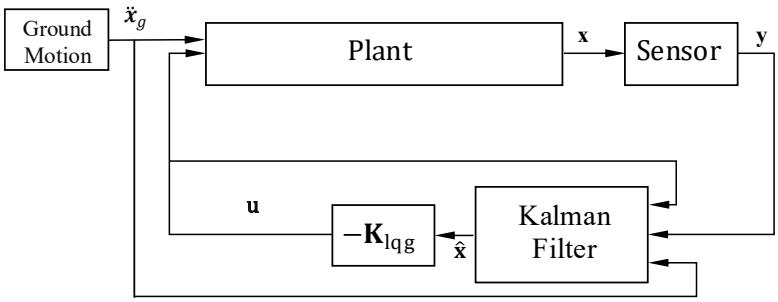


Figure 6.2 Block diagram of LQG controller

However, if the state-space model is unknown, the classical model-based control strategies are no longer applicable. Model-free methods are necessary to find the optimal control policy $f(\mathbf{y})$ that determines the control input signal $u = f(\mathbf{y})$. The following subchapters detail the proposed IL-DRL framework for this purpose.

6.3 IL-DRL Controller for SDOF-AMD

Chapters 3 and 4 have demonstrated the ability of pure DRL to achieve optimal model-free control of structural vibrations. Despite its superior performance without requiring explicit structural dynamic models, DRL remains challenging when applied in real-world structural control problems due to its dependency on massive trial-and-error interactions in the training process. Chapter 5 proposed the IL-DRL method to address these challenges and improve sample efficiency and safety in the practical experimental setting. This subchapter applies such a hybrid IL-DRL controller to an SDOF-AMD experiment. Subchapter 6.3.1 describes how the controller network is initialized by IL, and subchapter 6.3.2 illustrates how the controller is further improved by DRL.

6.3.1 Behavior cloning (BC)

Following Chapter 5, IL is also realized by using BC in this chapter, where BC is to mimic a simple reference control strategy for policy initialization purposes. The reference control strategy adopted is not designed to optimize any objective functions; instead, it can be any stable strategy that can increase the structural damping ratio while satisfying the constraints of the control devices.

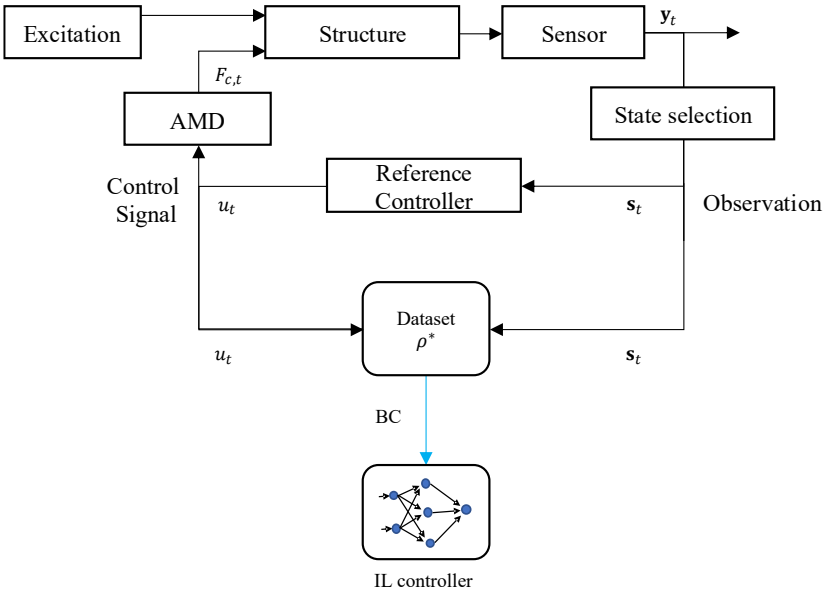


Figure 6.3 Block diagram of imitation learning

Figure 6.3 demonstrates the framework of IL adopted in this chapter, in which the reference control strategy is applied to the tested structure under random excitations to collect training data. The observation $\mathbf{s}_t$ is part of the sensor measurement $\mathbf{y}_t$ ($\mathbf{s}_t \subseteq \mathbf{y}_t$). The control signal u_t is determined and recorded according to the reference controller. In this case, the reference strategy is simply chosen as a negative control gain matrix $\boldsymbol{G_r}$ based on the sensor feedback $\mathbf{s}_t$ to prevent excessive motions of the AMD mass and structure.

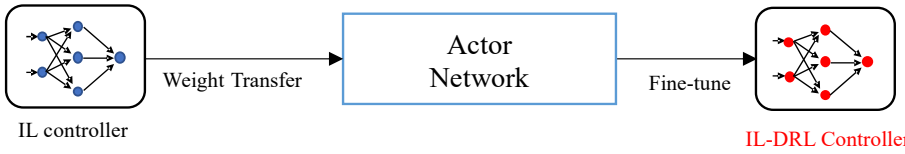
$$u_t = \boldsymbol{G_r}\mathbf{s}_t \tag{6.6}$$

where the control gain matrix $\boldsymbol{G_r}$ is a diagonal matrix with negative diagonal elements, u_t is the control signal sent to the AMD, and the sensor feedback $\mathbf{s}_t$ contains the relative position of the AMD cart x_c and the acceleration of the floor $\ddot{x}_f$ (i.e., $\mathbf{s}_t = [x_c \quad \ddot{x}_f]^T$). More details about the parameter setting will be provided in the experimental section.

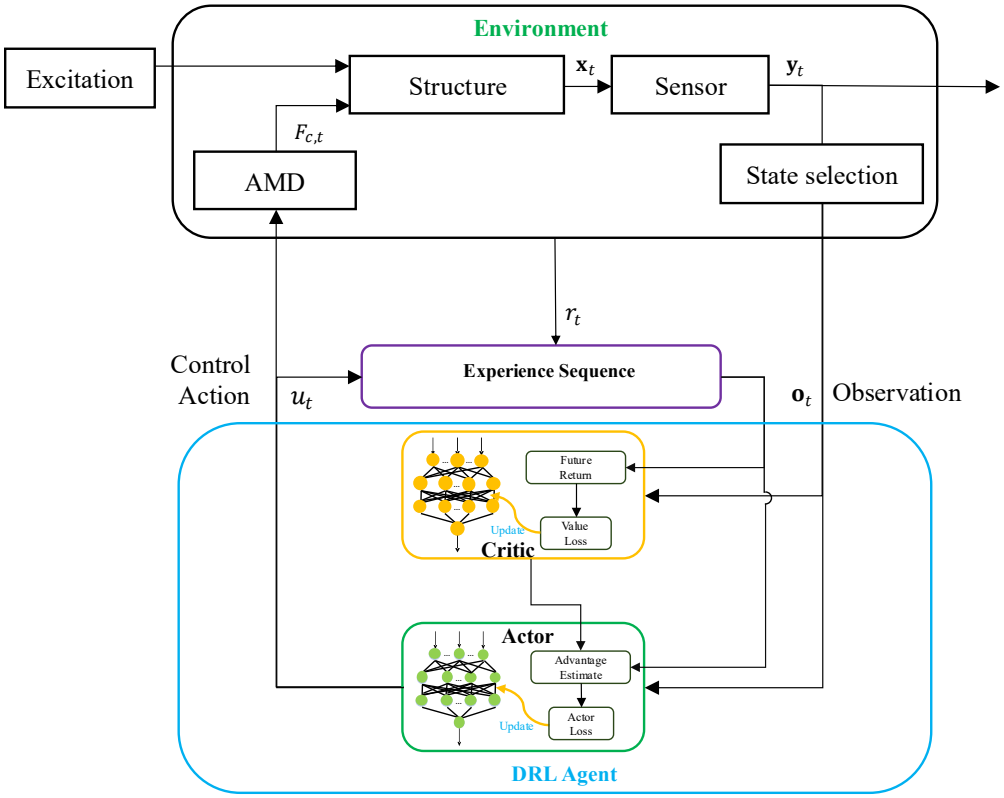
The sensor measurements and control signal at each time step form the training data set ρ^* consisting of observation-action pairs $(\mathbf{s}_t, u_t)$, where observation $\mathbf{s}_t$ and action u_t are the input and output of the reference controller. Notably, the reference strategy in Equation 6.6 can be easily designed and implemented on the physical structure without knowing the dynamic model of the structure. With the collected training dataset ρ^*, supervised learning is then applied to train the NN controller $\mu_\theta(\mathbf{s}_t)$ to mimic the reference control policy. The network parameter θ is optimized by minimizing the MSE $\varepsilon(\theta)$ (Equation 5.1) using gradient ascent. After the successful training, the IL-controller can completely replace the reference controller in the feedback system.

The weights of the successfully trained IL-controller are then transferred to the actor network of the DRL agent for fine-tuning described in the following subchapter. Although the BC is done on a limited dataset, it could enable a workable initial control policy for the DRL fine-tuning. As the parameters in the reference strategy are chosen quite arbitrarily, the reference strategy and the pre-trained initial policy by no means represent optimal control strategies. The pre-training with IL significantly reduces the number of interactions required by DRL, as the policy begins from a workable solution rather than random weights.

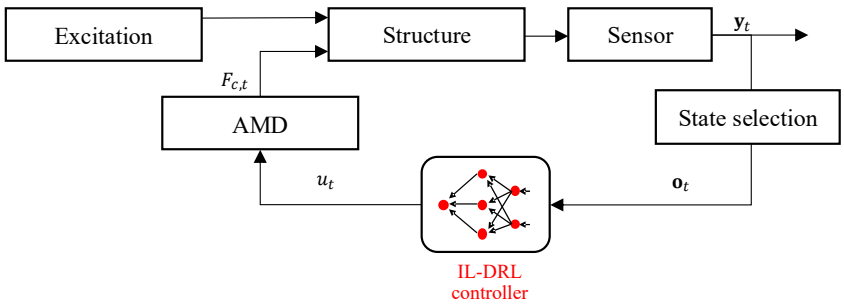
6.3.2 DRL fine-tuning



(a)



(b)



(c)

Figure 6.4 Block diagram of DRL (a) fine-tuning; (b) network evolution; (c) IL-DRL control

Since the previous IL-trained NN network provides a workable initial policy μ_θ, a model-free DRL algorithm can be employed to improve the control performance further, to minimize the number of physical tests to reduce the training cost and time.

First, the pre-trained control network's weight was transferred to the actor network of the DRL agent Figure 6.4a. The connections between the hidden layers and the current observation nodes in the top layer were retained, while the number of nodes in the input layer was adjustable to accommodate more observation points or previous measurements. The connections between the newly added nodes and the first hidden layer were initialized to zero, ensuring that the IL policy would not degrade. Then, instead of imitating the reference policy, the goal of DRL is to adapt by interacting with the environment to achieve the optimal strategy. As shown in Figure 6.4b, the structural vibration control problem is first regarded as an MDP model, where the next state at $t+1$ is decided only by the current state and the action at time t. At each time t, the DRL agent observes the state $\mathbf{o}_t$, executes the action $u_t = \mu(\mathbf{o}_t)$ according to the current controller (i.e., the current policy μ), receives the reward r_t, and the environment (i.e., the target structure) transitions to the next state at $t+1$. In this work, the state $\mathbf{o}_t$ is the measured structural responses (including other sensor measurements in addition to the observation $\mathbf{s}_t$, i.e., $\mathbf{s}_t \subseteq \mathbf{o}_t \subseteq \mathbf{y}_t$), action u_t is the control signal, and the reward function in this study is chosen in a negative quadratic form:

$$r_t = -\mathbf{o}_t^T \mathbf{Q}_{\text{DRL}} \mathbf{o}_t - u_t^T R_{\text{DRL}} u_t \tag{6.8}$$

where $\mathbf{Q}_{\text{DRL}}$ and R_{DRL} are the weighting matrices, which is also presented as Equation 3.19 in Chapter 3. The first part serves as a penalty for the structural responses, while the second term is related to control energy consumption. By maximizing the total reward, it is possible to achieve an equilibrium between minimizing vibration responses and avoiding excessive energy use. This adopted reward function is similar to the cost function (i.e., Equation 6.5) in the LQG controller to facilitate the performance comparison of different controllers with similar objective functions. However, unlike the traditional LQG controller that requires a standard cost function in Equation 6.4, the reward function of DRL can be flexibly designed in various forms other than Equation 6.8.

The chapter still employs PPO, the same as that in Chapter 5, to fine-tune the pre-

trained NN controller based on several considerations. The most significant advantage is that the update can be conducted separately after data collection, thus reducing the computational and data transmission requirements between the computer and the controller. Furthermore, the value function of PPO does not require any initialization, and a randomized value network can be combined with the pre-trained actor network. The process of model-free fine-tuning by PPO is outlined in Algorithm 6.1.

Algorithm 6.1: DRL fine-tune based on PPO

Initialization

Initialize agent with NN weights pre-trained by IL

For each episode:

Reset the structure to the initial state

For $t = 0$ to t_{end}

Get current state $\mathbf{o}_t$ from sensors

Select action $u_t = \mu(\mathbf{o}_t)$

Send u_t to AMD

Receive the reward $r_t(\mathbf{o}_t, u_t)$

If $t = t_{\text{end}}$, update:

$$\mu' = \text{PPO}(\mu^k, [\mathbf{o}_t, r_t(\mathbf{o}_t, u_t)])$$

End current episode

until convergence

Each update aims to find the policy μ that optimizes the objective function:

$$J_{PPO}^{\mu^k}(\mu) = \sum_{(\mathbf{o}_t, u_t)} \frac{p_\mu(u_t|\mathbf{o}_t)}{p_{\mu^k}(u_t|\mathbf{o}_t)} \alpha^{\mu^k}(\mathbf{o}_t, u_t) - \beta KL(\mu, \mu^k) \tag{6.10}$$

$$\alpha^{\mu^k}(\mathbf{o}_t, u_t) = H_t - V_\emptyset(\mathbf{o}_t) \tag{6.11}$$

where μ^k is the policy obtained from the last iteration k. $p_\mu(u_t|\mathbf{o}_t)$ is the probability of selecting control action u_t based on the observation $\mathbf{o}_t$. The advantage function $\alpha^{\mu^k}(\mathbf{o}_t, u_t)$ directly estimates the improvement of the discounted future reward H_t to the value estimation of the critic network value $V_\emptyset(\mathbf{o}_t)$ by the current action u_t. $\beta KL(\mu, \mu^k)$ is the adaptive KL penalty, a regularization term that constrains policy

updates by limiting the divergence between the new and old policies, ensuring stable training while preventing drastic changes that could degrade performance.

The subsequent subchapter demonstrates the efficiency of the proposed IL-DRL approach through experiments.

6.4 Experiments and Results Analysis

6.4.1 Experiment setup

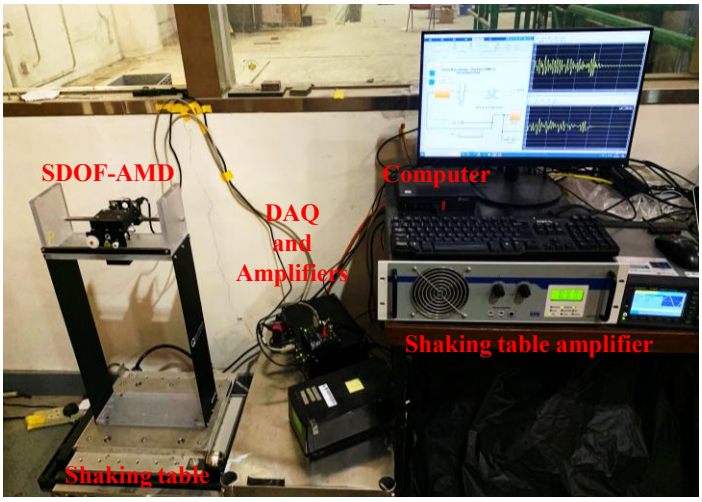


Figure 6.5 Experiment setup of the SDOF structure equipped with AMD

This subchapter presents the experimental evaluation of the proposed model-free IL-DRL controller on an SDOF frame structure equipped with an AMD system on top. The frame is made of aluminum. The experiment setup is shown in Figure 6.5. A shaking table (model No.: APS 420) generates ground excitations. Two accelerometers are installed on the top floor and at the base to measure the absolute floor acceleration $\ddot{x}_f$ and the base acceleration $\ddot{x}_g$, respectively. An encoder is installed on the AMD cart to measure the cart motion x_c. The AMD (model No.: Quanser AMD-1) consists of a linear cart driven by a direct current (DC) motor. The cart serves as the moving mass and slides along a stainless-steel shaft. The parameters of the system are given in Table 6.1.

Table 6.1 The parameters of the SDOF structure

Item	Parameter	Value
Structure	Floor total mass m_f (kg)	1.38
	Rack mass	0.7
	Floor linear stiffness K_f (N/m)	195.66
	Flexible structure height (m)	0.50
	Flexible structure depth (m)	0.11
AMD	Cart mass m_c (kg)	0.39
	Cart weight mass (kg)	0.13
	Cart motor torque constant K_t (N.m/A)	0.00767
	Cart planetary gearbox gear ratio K_g	3.71
	Cart motor armature resistance R_m (U)	2.6
	Cart back-electro motive-force constant K_m (V.s/rad)	0.00767
	Cart motor pinion radius r_{mp} (m)	0.00635
	Motor efficiency η_m (%)	100
	Gearbox efficiency η_g (%)	100

When the motor turns, the generated torque is transferred to the control force F_c through the rack and pinion mechanism. According to the supplier, the control force F_c in Equation 6.3 can be expressed as:

$$F_c = -\frac{\eta_g K_g^2 \eta_m K_t K_m \dot{x}_{c,t}}{R_m r_{mp}^2} + \frac{\eta_g K_g \eta_m K_t V_m}{R_m r_{mp}} \tag{6.12}$$

where all the parameters are defined in Table 6.1. In this case, the control signal u_t is taken as the motor voltage V_m

$$u_t = V_m \tag{6.13}$$

A data acquisition device (DAQ) (model No.: Quanser Q2-USB) collects and

transmits sensor measurements to the computer with the Intel i7-9700 CPU. Then, the IL-DRL agent receives the observations $\mathbf{o}_t = [x_{c,t}, \dot{x}_{c,t}, \ddot{x}_{f,t}, \ddot{x}_{g,t}, u_{t-1}]^T$, computes the control signal in the SIMULINK, and outputs the control voltage u_t to the cart motor through the amplifier (model No.: Quanser VoltPAQ-X1). The maximum input voltage for AMD cart operation is limited to 5 V, and the stroke is capped at 15 cm. To mitigate the structural vibration, an active control strategy is required to determine the control voltage in real time. The following part describes the implementation of the proposed model-free IL-DRL method.

6.4.2 Training on shaking table

A simple reference control strategy based on the feedback of the AMD cart displacement x_c and absolute floor acceleration $\ddot{x}_f$ was applied to the SDOF-AMD system for the IL purpose, where the control signal u_t was determined as:

$$u_t = [-10, -0.5] \times \begin{bmatrix} x_{c,t} \\ \ddot{x}_{f,t} \end{bmatrix} \tag{6.14}$$

The SDOF-AMD system was subjected to 400-s random vibrations to collect training data with a sampling interval of 0.002 s. The demonstration data set ρ^* contained 2×10^5 state-action pairs $(\mathbf{o}_t, u_t)$.

BC was used to pre-train an NN to imitate this reference control strategy. The network architecture had two hidden layers, and each layer had a dimension of four with *tanh* activations. The policy training used the Adam optimizer with a learning rate of 0.001 and a batch size of 512.

The trained IL policy was subsequently transferred to the actor network of PPO as the initial policy and then fine-tuned in shaking table tests. The reward function was constructed as a negative quadratic cost of the control voltage, cart relative displacements, and absolute and relative floor accelerations based on the sensor installation.

$$r_t = -\left[x_{c,t}, \ddot{x}_{f,t}, \left(\ddot{x}_{f,t} - \ddot{x}_{g,t}\right)\right]_t^T \mathbf{Q}_{\mathrm{IL-DRL}} [x_{c,t}, \ddot{x}_{f,t}, (\ddot{x}_{f,t} - \ddot{x}_{g,t})] - u_t^T R_{\mathrm{IL-DRL}} u_t \tag{6.15}$$

The weighting matrix $\mathbf{Q}_{\mathrm{IL-DRL}}$ was a diagonal matrix, in which all the non-diagonal elements are zero and the diagonal elements are $\mathbf{Q}_{\mathrm{IL-DRL}}(1,1) = 10000$, $\mathbf{Q}_{\mathrm{IL-DRL}}(2,2) = \mathbf{Q}_{\mathrm{IL-DRL}}(3,3) = 0.5$; and $\mathbf{R}_{\mathrm{IL-DRL}} = 1$.

The DRL training was done via shaking table tests. In each training episode, the SDOF-AMD system was subjected to 30-s random excitation, which was kept the same in all the training episodes, and 10-s free vibration, which allowed the vibration to decay. The time step was set as 0.002 s. The PPO algorithm was implemented in MATLAB in the experiment, with the following parameters: batch size 64, discount factor 0.99, actor network learning rate 0.0001, and critic network learning rate 0.001. If the actor network displayed instability, its learning rate was halved to stabilize the training process, and a penalty of -10^5 is added to the reward of the step that violates the constraint. The critic network comprises three hidden layers, each with a dimension of 256 and *ReLU* activations.

Figure 6.6 shows the high sampling efficiency of the proposed IL-DRL approach, compared with the pure DRL method. Notably, the training of the pure DRL with PPO was too long to be implemented in the experiments, so its training was done through simulations using the state-space model of the SDOF-AMD system in a SIMULINK environment. In contrast, the IL-DRL training was conducted through real shaking table experiments. The training parameters and excitation input were the same for the pure DRL and IL-DRL agents. The IL-DRL agent could achieve optimal control performance about ten times faster than the pure DRL method. Moreover, the training process of the pre-trained IL-DRL agent was apparently more stable than that of the randomly initialized DRL agent. The performance drops in the pure DRL training (episodes 50 – 100) arise from constraint violations in random exploration, i.e., the DRL agent temporarily exceeded motion or voltage limits, triggering penalty terms in the reward function. In contrast, the IL-DRL agent avoided such instability because its pre-trained policy, initialized via IL, inherently respects operational constraints. This not only accelerates convergence but also ensures safer real-world training by eliminating catastrophic failure in the exploration phases.

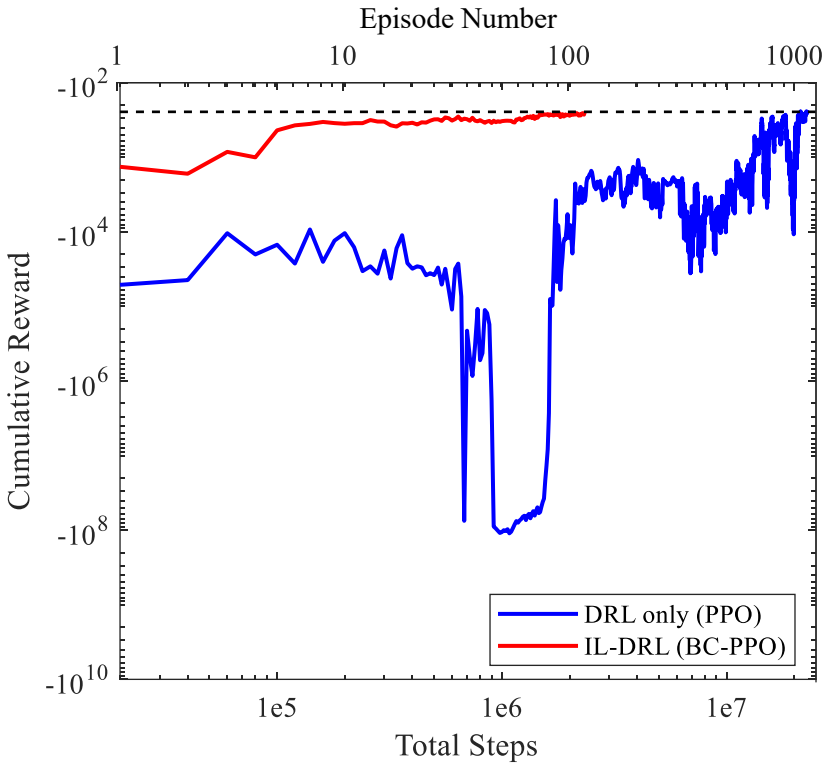


Figure 6.6 Cumulative reward plot of IL-DRL approach and DRL-only approach

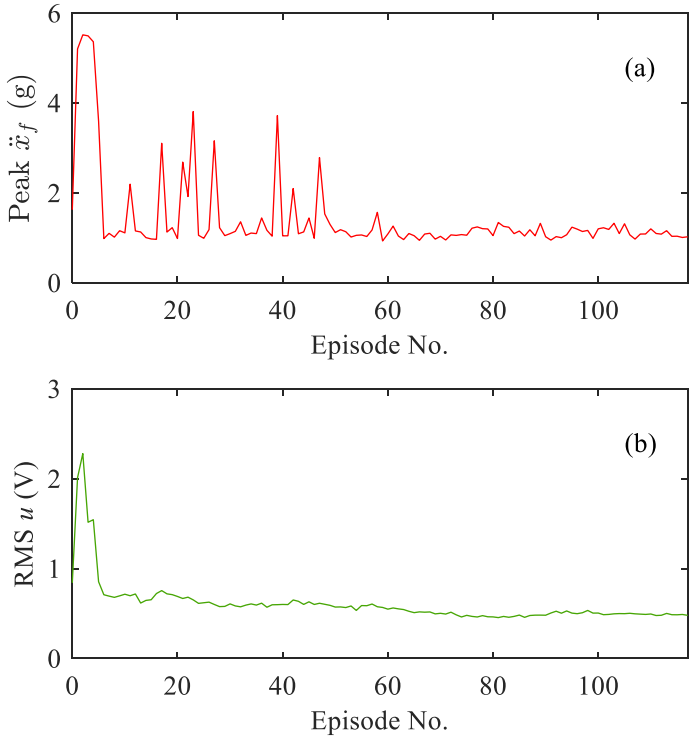


Figure 6.7 IL-DRL agent performance during the experiments: (a) peak floor acceleration; (b) RMS control voltage

Figures 6.7a and 6.7b indicate the peak absolute floor acceleration $\ddot{x}_f$ and the RMS control voltage u during the training, respectively. Both curves exhibit an upward trend at the initial stage, and they begin to decline steadily after approximately 60 episodes. At episode 70, both values are already comparably low. The training goes to 110 episodes to see stabilization. The final values were approximately half of the initial values (episode 0) before fine-tuning. The reduction in acceleration indicates better control performance, while the decreasing RMS control voltage stands for a lower energy consumption. In addition, no reasonable controller is obtained through pure DRL until approximately 800 episodes in simulation, so the pure DRL control performance is not included in Figure 6.7.

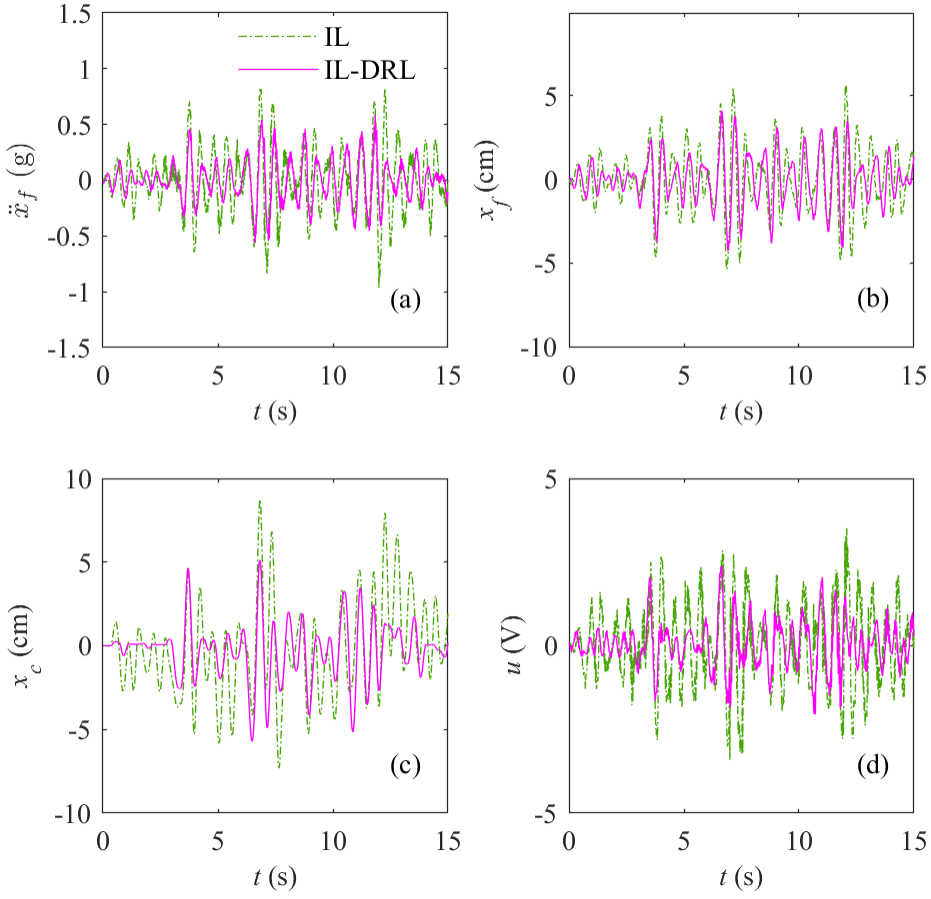


Figure 6.8 Random vibration responses of the SDOF-AMD system before and after DRL fine-tuning: (a) absolute floor acceleration; (b) relative floor displacement; (c) relative cart displacement; (d) control voltage

Figure 6.8 shows the structural responses and control voltage (only the first 15 s)

under the random excitation used in the fine-tuning process, when controlled by the fine-tuned IL-DRL agent and the initial IL controller without fine-tuning. The IL-DRL agent could provide better control performance while requiring less control voltage. Figure 6.8a shows the absolute floor acceleration response, where the peak values are significantly reduced from 0.96 g to 0.59 g (corresponding to a 39% reduction) after the DRL fine-tuning, and the RMS value is decreased by 38%. Figure 6.8b shows that the relative displacement control is also enhanced, where the peak value is reduced by 25% and the RMS value is reduced by 24%. The extent of the relative displacement reduction is not as significant as the acceleration, which may be due to the selection of the weighting matrix. Figure 6.8c shows the AMD cart response, where the maximum and RMS cart displacements are reduced by over 35% in comparison with the IL-trained controller. Figure 6.8d shows the time history of the control voltage, where the peak control voltage is reduced by 32% (from 3.5 V to 2.4 V) and the RMS control voltage is reduced by 42% (from 1.08 V to 0.63 V). In addition to its lower control signal level, the IL-DRL controller's control signal is smoother than the initial IL controller.

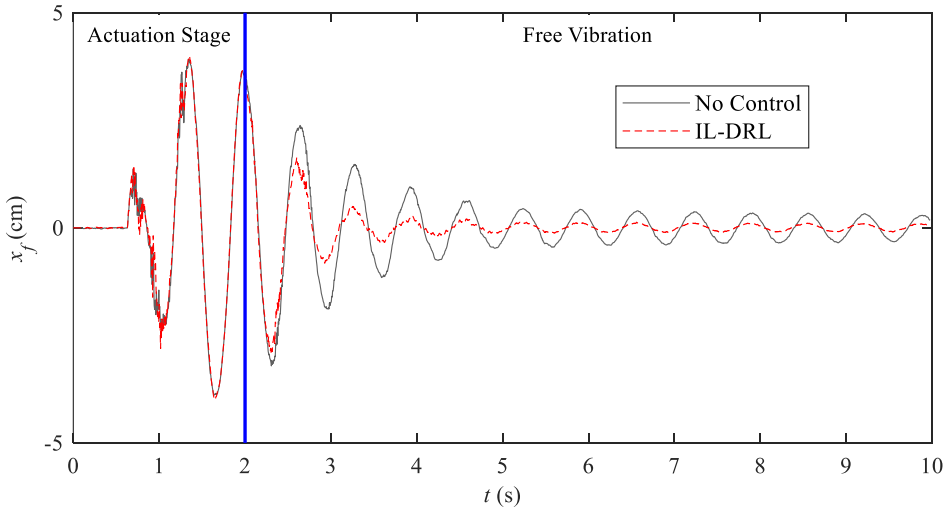


Figure 6.9 Relative floor displacement responses under free vibration

The free vibration behavior of the trained controller was further investigated before testing other types of excitations. The AMD cart was first used as the exciter to input sinusoidal excitation into the structure, and after 2 s, the AMD control mode was switched to either the IL-DRL controller or no control. Figure 6.9 shows the

displacement responses of the floor under free vibration. The IL-DRL controller quickly damped the floor vibration after the first cycle, compared to the uncontrolled case. In the first free vibration cycle, the displacement amplitude with the IL-DRL agent decreased from 3.7 cm to 1.6 cm, a 32% reduction compared to the uncontrolled case (from 3.7 cm to 2.4 cm). The damping ratio of the IL-DRL controlled response was calculated as 0.13, compared to 0.07 of the uncontrolled case, indicating a significant improvement of 86%. These results show the ability of the IL-DRL-trained controllers to learn how to suppress the vibrations of the target structure without knowing any structural model information.

6.4.3 Comparison with the model-based LQG controller

The model-free IL-DRL controller was further compared to a classic optimal model-based LQG controller under different vibration scenarios to examine the controllers' performance. As previously described in subchapter 6.2, the feedback gain $\mathbf{K}_{\text{lqg}}$ was calculated based on the ARE using the precise state space matrix knowledge of the system. The observation vector in Equation 6.2 was chosen as $\mathbf{y} = [x_c, \dot{x}_c, (\ddot{x}_{f,t} - \ddot{x}_g)\,]^T$ including the relative displacement and velocity of the AMD cart, and the relative acceleration of the floor to the shaking table. The state weight matrix $\mathbf{Q}_{\text{lqg}}$ in Equation 6.5 was chosen as

$$J_{\text{lqg}} = \int (\mathbf{y}^T \mathbf{Q}_{\text{lqg}} \mathbf{y} + u^T R_{\text{lqg}} u) dt, \tag{6.16}$$

$$\mathbf{Q}_{\text{lqg}} = \begin{bmatrix} 10000 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{6.17}$$

The weighting matrix $\mathbf{Q}_{\text{lqg}}$ was determined according to the reward function of DRL. The control weighting matrix R_{lqg} was 0.5 to ensure the RMS control voltage was at the same level as the IL-DRL controller under random excitations. The difference in the weighting matrices is because the state-space model used for the LQG controller in this study is constructed by using the relative motion of the structure to the ground, and the system output cannot contain the direct accelerometer measurement term (e.g., the absolute floor acceleration $\ddot{x}_{f,t}$). In this experiment, the ground displacement measurement was absent, and the state-space model used for the LQG was developed

relative to the ground.

6.4.3.1 Random vibration

A 250-s random excitation, which was different from the random excitation used in the training, was inputted to the SDOF-AMD system through the shaking table to compare the performance of different controllers. Figure 6.9a shows the time history of the floor displacement relative to the ground (i.e., the shaking table) under white-noise excitation. The IL-DRL controller suppressed the vibration more effectively (with lower peaks of floor displacements) than the LQG controller. The IL-DRL controller reduced peak floor displacements to 2.46 cm, outperforming the LQG controller (with 3.53 cm), achieving 61.9 % and 45.2 % reductions, respectively, compared with the uncontrolled case. Figure 6.9b illustrates the result of the absolute floor accelerations in the frequency domain after applying a fast Fourier transform (FFT). This figure reveals the difference between the two controllers. The uncontrolled structure exhibited a natural frequency of 1.5 Hz. The IL-DRL controller tended to shift the natural frequency by 6.7 % to 1.6 Hz, while the LQG controller shifted it by –18.8 % to 1.3Hz. This indicates the IL-DRL controller modifies the structural dynamic frequency less than the LQG controller. In terms of the integrated energy (i.e., the area under the FFT curve across the frequency range), the IL-DRL controller showed better attenuation than the LQG controller, with reductions of 67.7 % and 64.4 %, respectively, compared with the uncontrolled case.

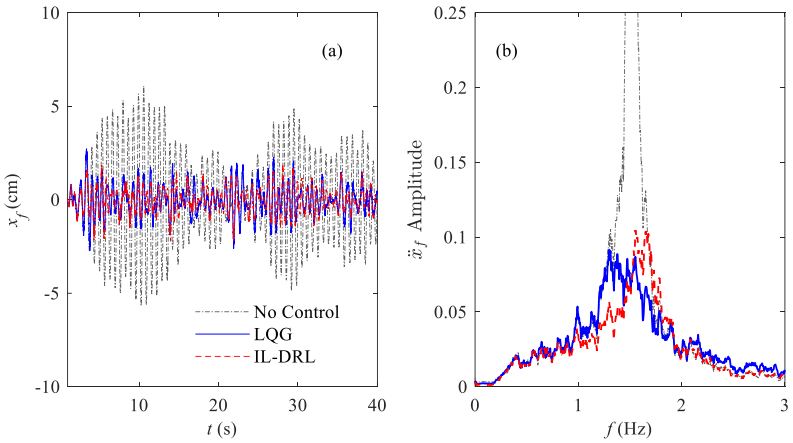


Figure 6.10 Random vibration of the SDOF-AMD system without control, with LQG and IL-DRL control (a) time history of the floor displacement relative to the ground; (b) FFT of the floor acceleration

Figure 6.10 compares the peak and RMS responses of the IL-DRL controller, the LQG controller, and the uncontrolled case, demonstrating the superior control performance of the IL-DRL controller. Figures 6.10a and 6.10b plot the absolute floor acceleration and relative floor displacement, respectively. Compared with the uncontrolled case, the relative peak displacement was reduced by 20% more by the IL-DRL controller than the LQG controller, and the peak absolute acceleration was further reduced by 10%. Figure 6.10c reveals the close control effort between the two controllers, where the RMS control voltage of IL-DRL is slightly higher, while the LQG controller shows a higher max voltage requirement. These results clearly illustrate that with similar control effort, the IL-DRL controller learned to reduce the floor responses better than the traditional LQG controller under white-noise excitations.

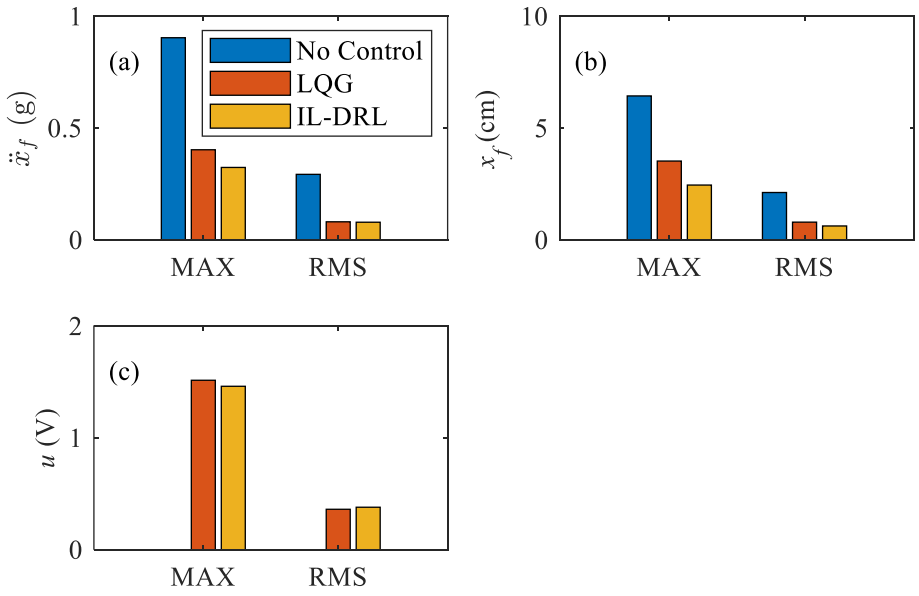


Figure 6.11 Peak and RMS responses to random excitation: (a) absolute floor acceleration; (b) floor displacement relative to the ground; (c) control voltage.

6.4.3.2 Sweep excitation

Sinusoidal sweeping ground motions from 0.5 to 3 Hz were applied to the SDOF-AMD system, with a slow incremental rate of 0.01 Hz/s and a constant signal input gain. Figure 6.12 compares the experimental results of different control cases. Figure 6.12a plots the floor acceleration response against excitation frequency, and Figure 6.12b shows the FFT result of the relative floor displacement response. The peak floor

acceleration of the first resonance mode reached 1.05 g for the uncontrolled system, while the corresponding values were decreased to 0.16 g by both the IL-DRL and the LQG controllers. Figure 6.12b shows the same pattern as Figure 6.10b in that the IL-DRL controller tends to reduce more responses than the LQG controller below 1.55 Hz. In contrast, the LQG controller performs better as frequency increases above the fundamental frequency. The pre-trained IL controller is also included in Figure 6.12b to show the learning ability of the DRL process. Although the initial IL control strategy (without fine-tuning) could reduce the displacement response below 1.7 Hz, the displacement was even higher than the uncontrolled case at a higher frequency, leading to poor vibration control performance in a high frequency range. In addition, the RMS control voltage by the initial IL controller is almost 2 times higher than the fine-tuned IL-DRL controller. After fine-tuning through DRL, the resonant peak shifted from 1.83 Hz to 1.64 Hz, and no vibration amplification was observed in a high frequency range.

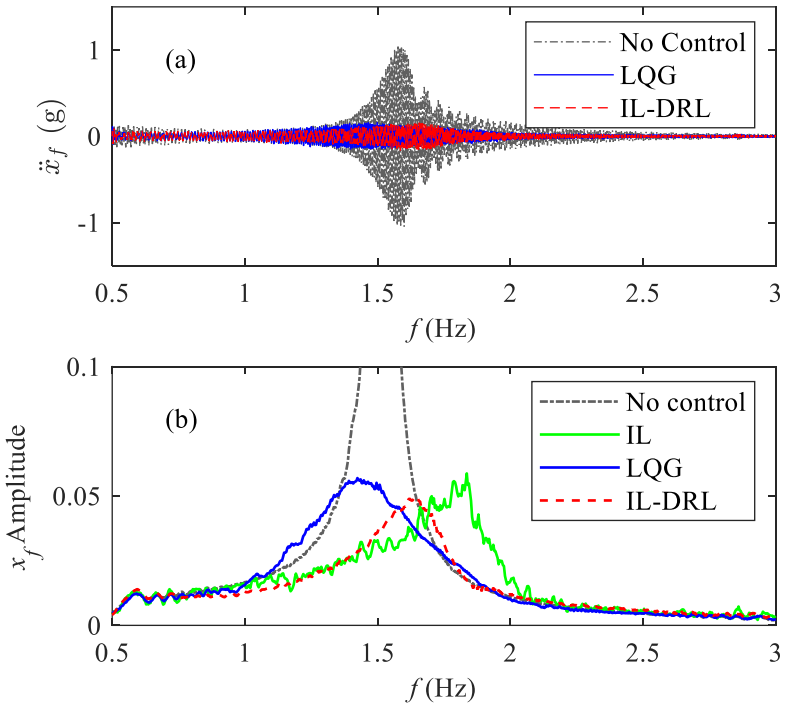


Figure 6.12 Responses to the sweep excitation: (a) absolute floor acceleration; (b) FFT of relative floor displacement response

6.4.3.3 Earthquake excitation

The performance of the proposed IL-DRL controller is further evaluated under

earthquake excitations. Figures 6.13a and 6.13b plot the time histories of structural displacement and acceleration responses with different controllers, when subjected to the scaled El Centro earthquake record. The IL-DRL controller showed a better peak value control performance. The peak relative floor displacement response of the uncontrolled case was 7.9 cm. Implementing the IL-DRL and the LQG controllers reduced the peak by 32 % and 23 %, respectively. The comparison of peak and RMS response values is plotted in Figure 6.14. With similar RMS control voltage, the proposed IL-DRL controller outperformed the LQG controller concerning the floor acceleration and displacement. Notably, although the IL-DRL controller was trained under random excitations, its performance under unseen earthquake excitation was even better than that of the model-based LQG controller, showing the robustness of the proposed model-free control strategy.

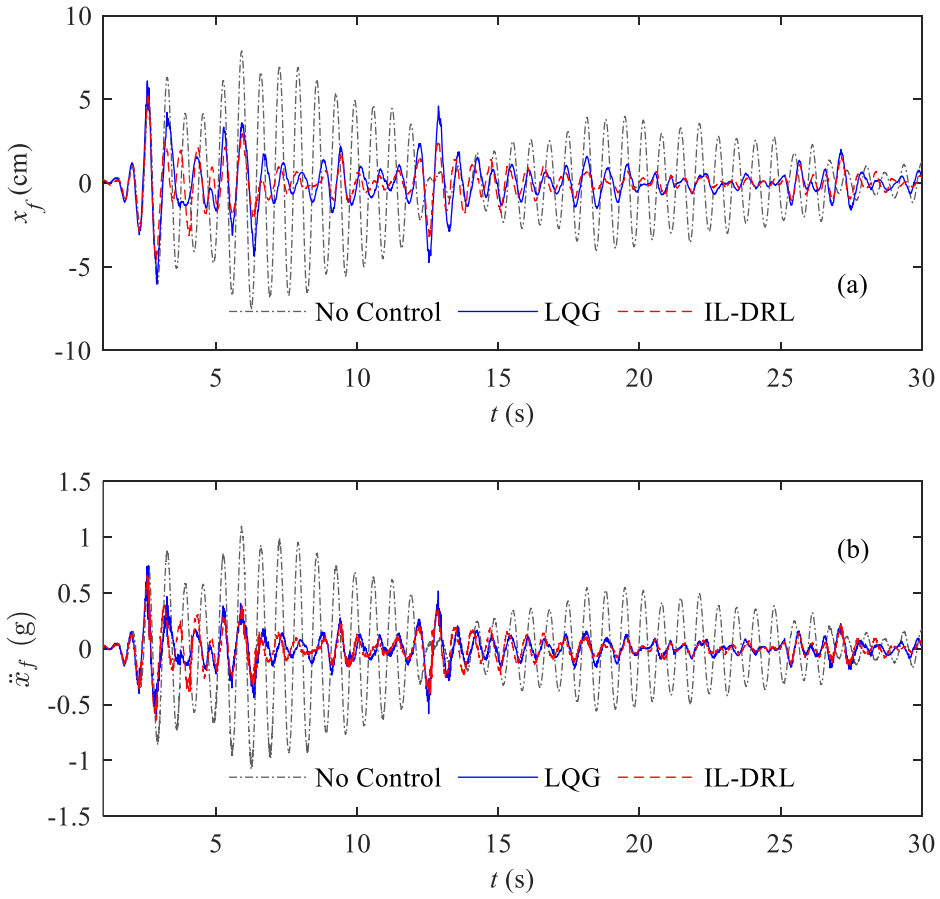


Figure 6.13 Comparison of responses to El Centro earthquake: (a) relative floor displacement; (b) absolute floor acceleration

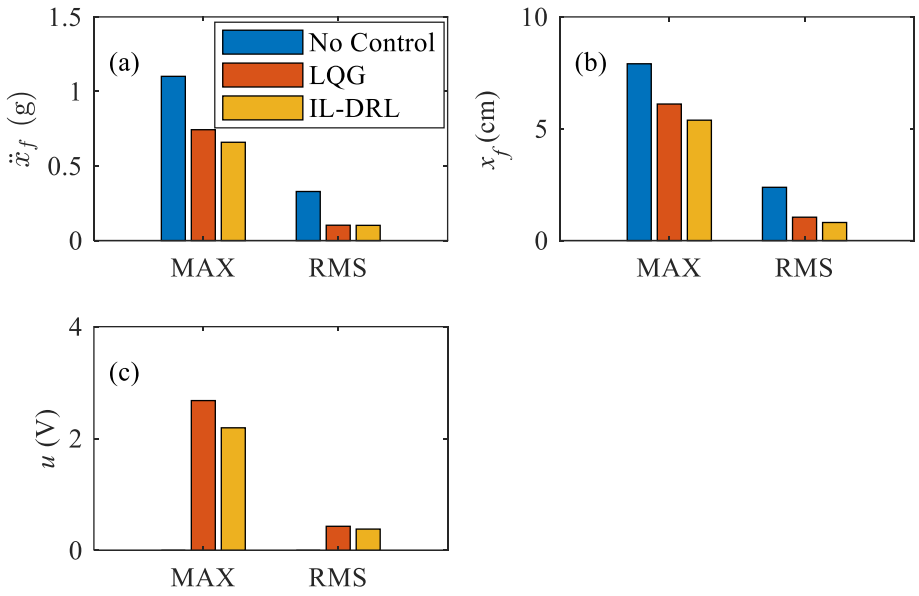


Figure 6.14 Peak and RMS responses to El Centro earthquake: (a) absolute floor acceleration; (b) floor displacement relative to the ground; (c) control voltage.

6.5 Summary

In this chapter, the novel model-free active vibration control method proposed in Chapter 5 that combines IL and DRL is successfully implemented in a series of laboratory experiments on an SDOF structure equipped with an AMD system. To address the challenge associated with a large number of training epochs required by pure DRL, the NN controller was initialized by BC with a simple reference control strategy and fine-tuned by PPO to obtain optimal performance under random excitation.

The proposed IL-DRL controller shows the following advantages in the experiments, which are consistent with the numerical results and observations reported in Chapter 5:

(1) Training efficiency and stability: Incorporating IL in the pre-training process, even with a casually chosen reference policy, can significantly reduce the training time of DRL while enhancing stability. This makes experimental and field implementations more practical in comparison with pure DRL, as the unstable exploration phases have safety risks.

(2) Model-free superiority: Unlike classical control methods that require precise structural dynamic models, the IL-DRL controller does not rely on structural models. Moreover, the IL-DRL controller outperforms the classical model-based LQG controller. With a similar control voltage level, the fine-tuned IL-DRL controller can further reduce the floor acceleration and relative displacement in comparison to the LQG controller, when subjected to various excitations.

Despite its advantages in training efficiency, model-free characteristics, and optimal control performance, the proposed IL-DRL also exhibits certain limitations. For example, the training process is sensitive to the choice of learning rate. An overly large initial learning rate may lead the agent to explore unstable policies. To prevent training instability in such cases, a failure-proof mechanism was implemented in this study: when the cart reached its stroke limit, the controller was automatically switched back to the last known stable policy.

The next chapter will extend this method's capability to higher degrees of freedom structure with more complicated dynamics, while trying to enhance the control performance by leveraging the flexibility in the design of cost functions. Unlike classical control methods that usually optimize rigidly defined objective functions in a specific form, the model-free nature of IL-DRL allows the exploration of control strategies without any restrictions in predefined cost functions.

CHAPTER 7

Multi-objective IL-DRL Vibration Control Strategy: Experimental Implementation on a Multi-story Frame

7.1 Introduction

In the previous chapters, the proposed IL-DRL framework was first applied to stay cable vibration control in Chapter 5 and subsequently validated through experiments on an SDOF structure with AMD in Chapter 6. In this chapter, the primary objective is to develop a multi-objective IL-DRL vibration control strategy by leveraging the flexibility of the reward design of DRL. As discussed in Chapters 5 and 6, the hybrid IL-DRL approach is designed to address the crucial challenges of DRL, such as extensive training requirements and potential instabilities when interacting with real-world systems. By intricately embedding constraints and design features during training, this chapter aims to elevate the IL-DRL framework's theoretical robustness and practical applicability in structural vibration control. Another objective is to conduct a comprehensive comparison of the proposed IL-DRL approach with classical model-based controllers like the LQR and the LQG, to validate the effectiveness under different observers. A series of experiments was conducted on a three-degree-of-

freedom (3DOF) structure with an AMD equipped and exposed to diverse excitations.

The key contributions of this chapter can be summarized as follows:

a) Based on the SDOF experiment in Chapter 6, this chapter extends the IL-DRL framework to control the vibration of a 3DOF structure controlled by an AMD in experiments.

b) The integration of multiple selective objectives into the DRL's reward function facilitates fine-tuning and allows for the direct optimization of a diverse array of control objectives, improving the cost function in a rigid form (Equation 6.8) used in Chapters 5 and 6.

c) The multi-objective IL-DRL strategy is used to train two controllers with full state and partial state feedback and compared with LQR and LQG under various excitations in shaking table experiments to demonstrate the superior control performance and the advantages of multi-objective functions.

This chapter is structured as follows: Subchapter 7.2 briefly describes the formulation control and classical optimal control methods used for later comparison. Subchapter 7.3 elaborates on the proposed multi-objective IL-DRL methodology. In Subchapter 7.4, the experimental setup used for performance validation is described. Subchapter 7.5 discloses the implementation specifics and presents the results under various testing conditions. Finally, subchapter 7.6 provides a summary of this chapter.

7.2 Classical Control Formulation

The continuous-time state-space representation of a linear structure subjected to a ground motion can be expressed as Equations 6.1 and 6.2.

When the system matrices are precisely known, traditional model-based control strategies can be effectively employed to determine the optimal active control force. In Chapters 3 and 6, the LQR and LQG controllers have been discussed. Although these traditional approaches are effective under ideal conditions, they heavily rely on the accuracy of the system parameters. Moreover, the form of the cost function J_0 is fixed.

$$J_0 = \int_0^\infty (\mathbf{x}^T \mathbf{Q}_x \mathbf{x} + \mathbf{u}^T \mathbf{R} \mathbf{u}) dt, \tag{7.1}$$

where $\mathbf{Q}_x$ and $\mathbf{R}$ are the state and control force weight matrices, respectively. The term $\mathbf{x}^T\mathbf{Q}_x\mathbf{x}$ represents the state cost, and $\mathbf{u}^T\mathbf{R}\mathbf{u}$ accounts for the control energy cost.

These two limitations lead to the development of the proposed IL-DRL methodology. Notably, the state-space model presented herein is only used for the classical model-based controllers. The proposed IL-DRL controller is essentially model-free and does not need any state-space model information.

7.3 Multi-objective IL-DRL Controller

7.3.1 Overview of the multi-objective IL-DRL approach

This chapter extends the previously proposed IL-DRL controller to a flexible model-free multi-objective control strategy to overcome the shortcomings of the classical model-based controllers. The primary advantage of IL-DRL is that the entire process is data-driven, limiting the reliance on precise system matrices. The IL-DRL approach circumvents the initial instability issue of DRL by utilizing IL to efficiently pre-train an initial control policy. The subsequent fine-tuning is conducted using DRL, allowing for fast convergence and improved adaptability to unseen scenarios.

Furthermore, although the classical model-based controllers offer solutions for systems with prior known dynamics, the objective function in the controller design may not be directly related to the performance evaluation criteria. For example, the LQR and LQG controllers optimize a trade-off between structural kinetic energy and control effort, but their predefined cost function may not fully capture critical performance metrics of particular interest. When controlling vibrations in the MDOF frame structure illustrated in Figure. 7.1, the key performance metrics, such as the maximum inter-story drift and maximum floor acceleration, cannot be explicitly optimized within the conventional LQR framework. By contrast, DRL can provide superior design flexibility, allowing for the simultaneous optimization of multiple competing objectives through carefully designed reward functions during the training process.

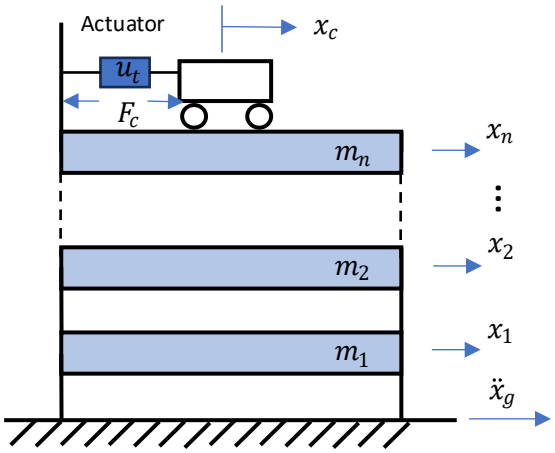


Figure 7.1 The schematic of an MDOF frame equipped with an AMD at the top

This chapter uses the MDOF frame controlled by an AMD as a demonstration structure to illustrate the multi-objective IL–DRL method. x_c signifies the relative displacement of the AMD to the top floor, x_n is the relative displacement of the nth floor to the ground, u_t stands for the control signal sent to the AMD at time t, and F_c is the control force of AMD.

7.3.2 IL-DRL design for MDOF-AMD

In this study, IL serves as a preliminary step to pre-train the controller for subsequent DRL. The pre-training improves sampling efficiency and reduces overall training time. The concept and procedure have been illustrated in the previous chapters. The training flow can be found in subchapter 6.3

The data set ρ_{IL}, which comprises time-stamped state-action pairs $(\mathbf{s}_t, u_t)$, is collected from a real structure when the reference controller is executed on the target structure under random excitations. The observation vector $\mathbf{s}_t$, selected from the sensor measurements $\mathbf{y}_t$ ($\mathbf{s}_t \subseteq \mathbf{y}_t$), depending on the input requirement of the reference controller, and u_t is the control signal sent by the reference controller to the AMD. In this MDOF-AMD system, the reference controller is chosen as a simple position-velocity proportional controller of the AMD mass, which only relies on the sensor measurement of the relative displacement x_c and velocity $\dot{x}_c$.

$$u_t = \mathbf{K}_{\mathrm{pv}}\mathbf{s}_t, \tag{7.2}$$

where $\mathbf{K}_{\text{pv}}$ is the control gain matrix, and $\mathbf{s}_t = [x_c\ \dot{x}_c]^T$. The choice of reference controller here is not designed to optimize any control objectives but to provide a stable initial controller for the later fine-tuning stage. BC is applied after the data collection process. After the training, the IL controller can replace the reference controller, and the network weights are transferred to the actor network of a DRL agent to ensure that compatibility is maintained by retaining connections between the hidden layers and the current observation nodes, while also allowing for flexibility in the input layer to accommodate additional observation points.

The DRL fine-tuning design is based on the previous concept discussed in Chapter 6. In this MDOF-AMD system, $\mathbf{o}_t$ is the combination of $\mathbf{s}_t$ and other measured floor responses. The pre-trained IL control network's weight is initially transferred to the actor network of the DRL agent, keeping the connections between the hidden layers and the observation nodes $\mathbf{s}_t$ in the input layer. Additional observation nodes are added in the input layer to include other state measurements in $\mathbf{o}_t$. The weights of the new connections are zero at first to prevent instability.

For each time step *t*, the DRL agents obtain the observation state $\mathbf{o}_t$ from the sensors, output the action u_t to the control device (i.e., AMD), and receive the reward r_t; afterward, the environment (i.e., the target controlled structure) transitions to the next state at step $t + 1$, and $[\mathbf{o}_t, u_t, r_t]$ is recorded in the experience buffer. This chapter also adopted PPS as the training algorithm. Upon successfully training, only the actor network is deployed as the final controller on its own (without the critic network).

7.3.3 Multi-objective reward function

The classical model-based LQR controller faces an inherent limitation in the control objective function design. The LQR cost function adopts a rigid form that combines quadratic costs to balance the structural response and the control effort through energy-based metrics, as shown in Equation 3.4.

By contrast, the proposed IL–DRL framework enables flexible multi-objective control by designing the reward function r_t as a weighted combination of different structural performance metrics. The weighted sum approach is widely adopted in multi-objective optimization to balance competing targets. In the vibration control of the

MDOF-AMD system shown in Figure 7.1, the control performance is typically evaluated based on the reduction in inter-story drift d_n and floor acceleration $\ddot{x}_n$. The inter-story drift quantifies the relative displacement between adjacent floors, reflecting structural deformation and damage. Meanwhile, floor acceleration directly affects human comfort, equipment safety, and non-structural damage. Accordingly, the maximum inter-story drift and maximum floor acceleration recorded at each time step are included directly in the reward function. In addition, the AMD's mass position $x_{c,t}$ and velocity $\dot{x}_{c,t}$ are included to regulate the movement of the AMD.

$$r_t = -[\beta_1 x_{c,t}^2 + \beta_2 \dot{x}_{c,t}^2 + \beta_3 \max(d_{i,t}^2) + \beta_4 \max(a_{i,t}^2)], \tag{7.3}$$

where $\beta_1, \beta_2, \beta_3,$ and β_4 are the weighting coefficients for each performance metric; $\max(d_{i,t}^2)$ is the squared maximum inter-story drift among all the stories at time *t*; and $\max(a_{i,t}^2)$ is the squared maximum floor acceleration among all the floors at time *t*. The fine-tuning process of the DRL aims to find the optimal balance between these competing control objectives to ensure simultaneous suppression of structural lateral motion, minimization of AMD movement, and preservation of human comfort and structural integrity. Notably, this framework is flexible, and the rewards can be easily redefined by incorporating new objectives.

Further details on the adjustment of the reward functions to different observation scenarios are provided in subchapter 7.5, which introduces the experimental implementation of the proposed IL–DRL. Safety remains a priority during the training phase, as the data are collected on a real structure during the experiment. Accordingly, constraints are activated when the damper exceeds its permissible displacement limits; consequently, the controller is switched to the last workable one in the fine-tuning. In such cases, the most recently effective control policy is reverted to, thus averting any severe structural damage. Meanwhile, the learning rate of the actor component in the PPO algorithm is halved, slowing down the policy updates to exercise additional caution.

7.4 Experimental Study

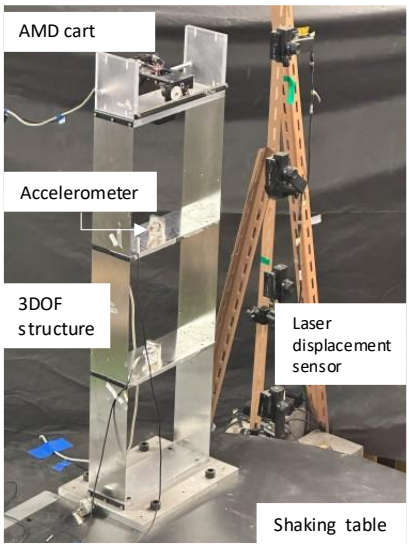
7.4.1 Experimental setup

This subchapter describes the experimental setup used to evaluate the control strategies discussed earlier. The experiment features a 3DOF frame structure made of 304 stainless steel, equipped with an AMD (model No.: Quanser AMD-1) on top. The experimental setup and schematic design of the test system are shown in Figures 7.2a and 7.2b, respectively. Three accelerometers are positioned at each floor level to measure the absolute floor acceleration (a_1, $a_{2,}$, a_3). The ground excitation $\ddot{x}_g$ is measured by another accelerometer installed on the shaking table (model No.: MTC22009). In addition, four laser displacement sensors (model No.: KEYENCE LK-500) are installed to measure the absolute floor and ground displacements, which can be used to compute the relative floor displacement against the shaking table (x_1, $x_{2,}$, x_3). Sensor feedback signals are collected with a Quanser Q8-USB DAQ (Figure 7.2c). This DAQ device forwards the collected data to the computer for real-time or post-processing analysis. The control voltage u_t is sent to the cart motor through the amplifier (model No.: Quanser VoltPAQ-X1), with a maximum input voltage of $u_{\max}$ = 5V. This constraint is crucial for designing and implementing control algorithms operating within safe voltage limits. The CPU is an Intel i7-9700, and the DRL process is computed in SIMULINK.

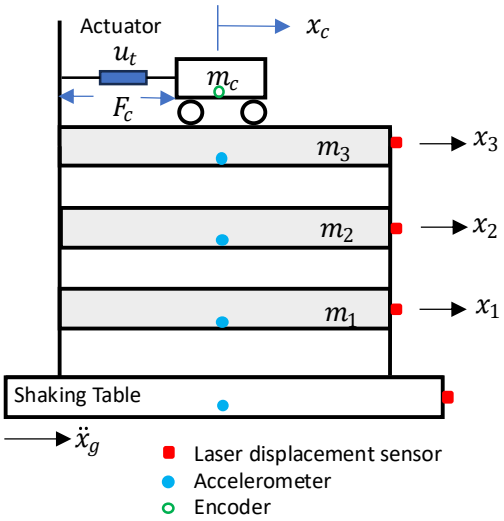
The key parameters of the 3DOF frame structure and the AMD cart, including the material properties and dimensions, are detailed in Table 7.1. The natural frequencies of the uncontrolled system are measured at 1.63, 4.59, and 6.31 Hz for the three vibration modes. The AMD cart used is manufactured by *Quanser Consulting Inc.*, equipped with an encoder to measure the cart's relative position to the top floor x_c and velocity $\dot{x}_c$. The AMD system consists of a linear cart that is driven by a DC motor. The cart serves as the moving mass and slides along a stainless-steel shaft. The cart serves as the moving mass and slides along a stainless-steel shaft. When the motor turns, the torque created is transferred through the rack and pinion mechanism to the control force F_c. The control signal u_t is taken as the motor voltage V_m.

Table 7.1 The parameters of 3DOF frame and AMD

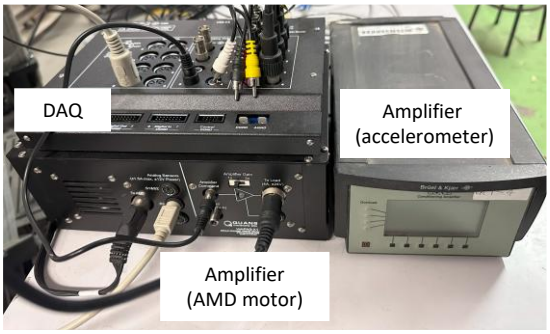
Item	Parameter	Value
Structure	Third floor mass m_3 (kg)	2.41
	Second floor mass m_2 (kg)	2.66
	First floor mass m_1 (kg)	2.66
	Floor linear stiffness K_f (N/m)	1323.35
	Structure floor height (m)	0.32
	Structure floor depth (m)	0.01
	Structure floor width (m)	0.11
AMD	Moving mass of the AMD cart m_c (kg)	0.41
	Cart motor torque constant K_t (N.m/A)	0.00767
	Cart planetary gearbox gear ratio K_g	3.71
	Cart motor armature resistance R_m (U)	2.6
	Cart back-electromotive-force constant K_m (V.s/rad)	0.00767
	Cart motor pinion radius r_{mp} (m)	0.00635
	Motor efficiency η_m (%)	100
	Gearbox efficiency η_g (%)	100



(a)



(b)



(c)

Figure 7.2 Experiment setup (a) 3DOF frame with AMD (b) schematic drawing; (c) DAQ system and amplifiers

7.4.2 Test scenarios

The 3DOF-AMD system is tested in two feedback scenarios, namely, full and partial state feedback. The full-state observation vector $\mathbf{o}_{t,\mathrm{FS}} = [x_{c,t}, x_{1,t}, x_{2,t}, x_{3,t}, \dot{x}_{c,t}, \dot{x}_{1,t}, \dot{x}_{2,t}, \dot{x}_{3,t}]^T$ includes the relative displacement and velocity of the AMD cart to the top floor and the relative displacement and velocity of each floor to the shaking table. By contrast, the partial-state observation vector $\mathbf{o}_{t,\mathrm{PS}} = [x_{c,t}, x_{3,t}, \dot{x}_{c,t}, \dot{x}_{3,t}]^T$ only includes the displacement and velocity of the cart and the top floor. The former and latter are compared with the classical LQR and LQG controllers, respectively.

In both scenarios, the IL pre-training is the same, while the DRL fine-tuning is slightly adjusted to fit the different dimensions of the observation vectors. The NN for the IL process is constructed with three hidden layers, each containing 64 nodes, with Tanh activation function. The position-velocity control gain used for BC is $\mathbf{K}_{\mathrm{pv}} = [-10,5]$. A single random ground excitation of 500 s is applied to the structure to collect the dataset ρ_{IL} for the BC training.

The trained NN is subsequently transferred to the actor network of the PPO agent. The critic network in the PPO agent consists of three hidden layers with 256 nodes each, using a *ReLU* activation function. The critic and actor learning rates are initialized at 1e−3. When the actor network displays potential instability, its learning rate is halved to stabilize the training process, thereby ensuring a smooth update.

An additional penalty term is added to discourage excessive actuation power beyond 90% of the maximum limit ($u_{\max} = 5V$). The reward function in Equation 7.3 becomes

If $u < 0.9u_{\max}$:

$$r_t = -[\beta_1 x_{c,t}^2 + \beta_2 \dot{x}_{c,t}^2 + \beta_3 \max(d_{i,t}^2) + \beta_4 \max(a_{i,t}^2)], \tag{7.4}$$

If $u > 0.9u_{\max}$:

$$r_t = -[\beta_1 x_{c,t}^2 + \beta_2 \dot{x}_{c,t}^2 + \beta_3 \max(d_{i,t}^2) + \beta_4 \max(a_{i,t}^2)] e^{\frac{u_t}{0.9u_{\max}}}. \tag{7.5}$$

where the weight coefficient values are $\beta_1 = 1, \beta_2 = 0.1,\ \beta_3 = 1000,\ \beta_4 = 0.1$. The second condition is only triggered when the control voltage approaches the maximum limit to discourage maximum power usage.

7.5 Experimental Results and Discussions

7.5.1 Fine-tuning of full-state feedback controller

The first experiment is to fine-tune the controller with full-state feedback $\mathbf{o}_{t,\mathrm{FS}}$. Figure 7.3 shows the evolution of training rewards over episodes, capturing the learning trajectory of the agent. Each training episode consisted of a 30-s test under white noise ground excitation, followed by 10 s of zero excitation to observe free vibration decay. The ground excitation record was maintained the same throughout the training. After each iteration, the agent's policy is updated in around 30 s. Therefore, the entire training phase is accomplished in approximately an hour.

Figure 7.4 displays the agent's incremental performance gains as the training advances. Specifically, both the maximum inter-story drift and maximum floor acceleration among the three floors decline rapidly, with the latter showing a highly pronounced rate of improvement. The structure achieves its optimal controlled state as the training approaches completion, as indicated by the minimized inter-story drift. The fine-tuned IL-DRL controller achieved significant performance improvement, decreasing peak inter-story drift by 56.6% and peak floor acceleration by 40.8%, compared with the initial performance levels of the IL controller without any fine-tuning.

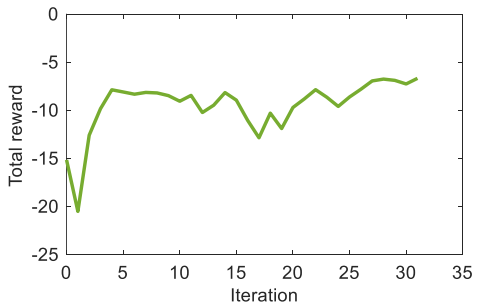
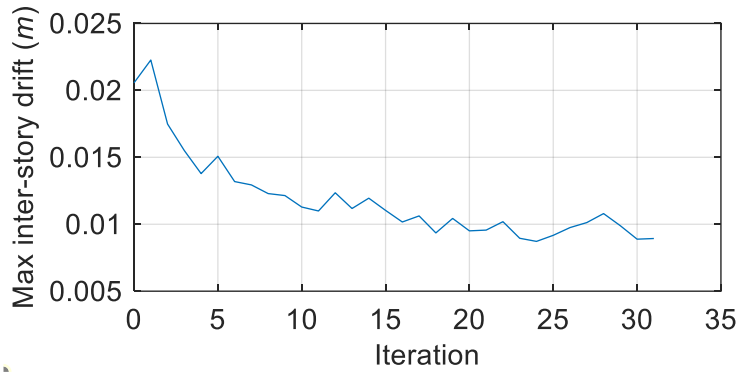
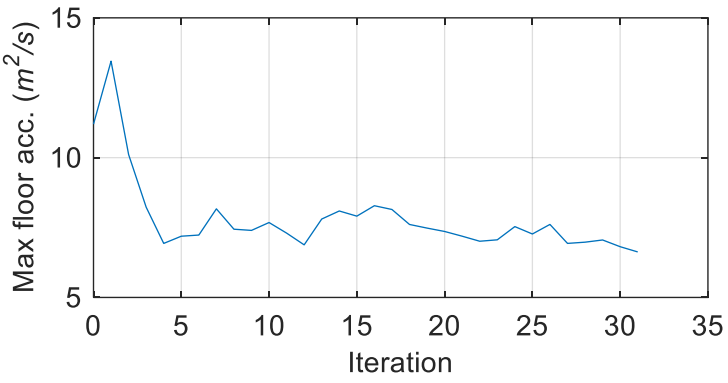


Figure 7.3 Training reward against episode during DRL fine-tuning (full state feedback)



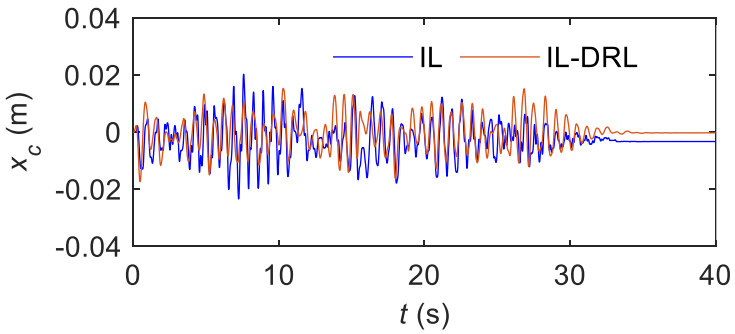
(a)



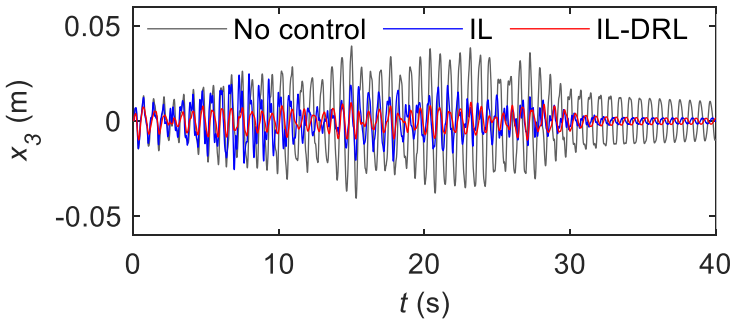
(b)

Figure 7.4 Peak level during DRL fine-tuning (with full state feedback): (a) peak inter-story drift (b) peak floor acceleration

Figure 7.5a and 7.5b further compare the displacement time histories of the cart and top floor between the initial IL controller and the fine-tuned IL-DRL controllers under the same excitation used in the training process. The fine-tuned IL-DRL controller significantly outperforms the IL controller, reducing the peak displacement of the third floor by 40% and the RMS displacement of the cart by 14%. While the IL controller should theoretically return the AMD cart to its neutral position (due to its position feedback term), its performance degrades in the experiment when encountering unobserved states outside the training data distribution. The IL-DRL controller compensates for this weakness through the DRL fine-tune, ensuring consistent restoration to the neutral position and robust performance across new states.



(a)



(b)

Figure 7.5 The responses of the 3DOF frame-AMD system before/after fine-tuning by DRL: (a) cart displacement (b) third floor displacement

7.5.2 Fine-tuning of partial state feedback controller

This experimental study is expanded to include the training of a controller with partial-state feedback $\mathbf{o}_{t,\mathrm{PS}}$ to further explore the flexibility and adaptability of the proposed approach. The NN structure remains consistent with the full-state feedback model, with the only difference being a halving of the nodes in the input layer to accommodate the reduced dimension of the observation vector.

Given the limited observation in the partial-state feedback, the reward function requires modification accordingly. The inter-story drift value cannot be calculated from the top floor displacement. Therefore, the square value of peak inter-story drift $\max(d_{i,t}^2)$ in the reward function is replaced by the square value of the top floor displacement $x_{3,t}^2$

$$r_t = -[\beta_1 x_{c,t}^2 + \beta_2 \dot{x}_{c,t}^2 + \beta_3 \max(x_{3,t}^2) + \beta_4 \max(a_{i,t}^2)] \quad (7.6)$$

where $\beta_3 = 10000$, and the values of β_1, β_2, and β_4 are the same as those in subchapter 7.4.2

Figure 7.6 shows the DRL fine-tuning process for the partial-state feedback controller, with the total reward progression over 35 training iterations. The overall duration of the training phase is still approximately an hour. The learning curve exhibits a consistent upward trend throughout the training course with no significant regressions, indicating stable policy updates.

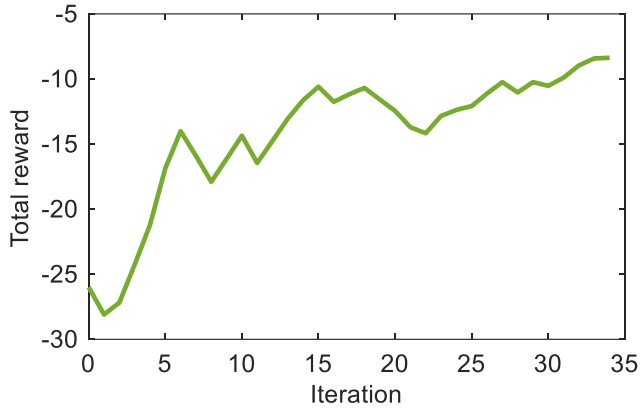
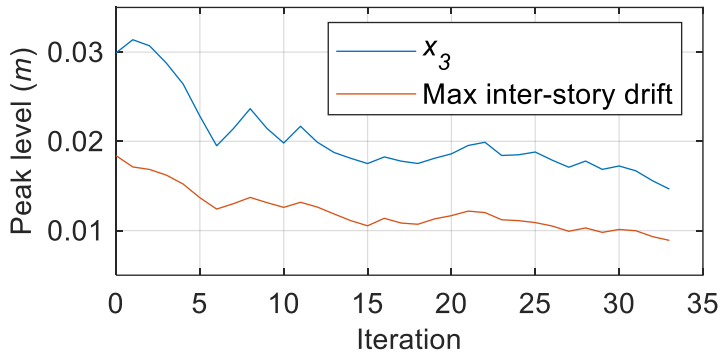
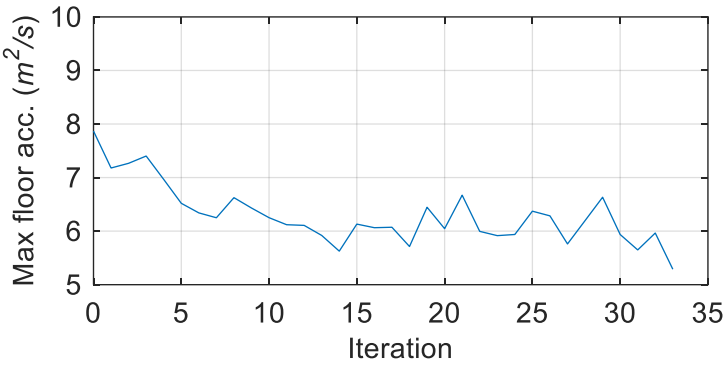


Figure 7.6 Training reward against episode during DRL fine-tuning (with partial-state feedback)

Figure 7.7 illustrates the peak values recorded during the training phase. Compared with the initial IL controller without fine-tuning, the fine-tuned partial state IL-DRL controller reduces the peak inter-story drift by 51.6%, the top floor displacement by 51 %, and the top floor acceleration by 32.8 %. The top floor displacement shows a pattern close to that of the inter-story drift, despite the two occasionally being at odds. Specifically, the peak inter-story drift doesn't always occur between the top two floors. Nevertheless, the diagram underscores the effectiveness of the chosen reward function. Even in the absence of direct inter-story drift metrics within the reward function, the controller successfully manages to reduce the inter-story drift throughout the training course. This outcome is an encouraging validation of the adaptability and robustness of the modified reward function when transitioning from full-state to partial-state feedback.



(a)



(b)

Figure 7.7 Peak response during the DRL fine-tuning of partial-state feedback controller: (a) peak inter-story drift and top floor displacement; (b) peak floor acceleration

7.5.3 Random excitation

Another white-noise ground excitation of 250 s, different from the one used in the training, is applied to the 3DOF-AMD system by the shaking table to evaluate the performance of the two trained controllers under a ground excitation unseen in the training process. The IL–DRL controllers are compared against two classical model-based controllers, namely, the LQR controller with full-state feedback and the LQG controller with partial-state feedback. The weighting matrices for the LQR and LQG controllers are chosen to provide control performance similar to that of the IL-DRL controller. Given that the performance of the LQG controller is highly sensitive to assumed system noise and measurement noise, multiple variations are tested to identify

the best-performing version. The final selection of the system noise covariance is $\mathbf{V}_d = 1 \times 10^{-4}\mathbf{I}_{8\times8}$, and the measurement noise covariance is $\mathbf{W}_n = [2.5 \times 10^{-4}\mathbf{I}_{4\times4}, 0; 0, 0.003\mathbf{I}_{4\times4}]$.

To quantitatively evaluate the performance of the IL–DRL controllers and the classical controllers, six nondimensionalized performance metrics are adopted, which are inspired by the evaluation of another active control benchmark problem (Spencer et al., 1998).

$$J_1 = \max\left[\frac{\sigma_{d_1}}{\sigma_{x_{3\mathrm{uc}}}}, \frac{\sigma_{d_2}}{\sigma_{x_{3\mathrm{uc}}}}, \frac{\sigma_{d_3}}{\sigma_{x_{3\mathrm{uc}}}}\right] \tag{7.7}$$

$$J_2 = \max\left[\frac{\sigma_{a_1}}{\sigma_{a_{3\mathrm{uc}}}}, \frac{\sigma_{a_2}}{\sigma_{a_{3\mathrm{uc}}}}, \frac{\sigma_{a_3}}{\sigma_{a_{3\mathrm{uc}}}}\right] \tag{7.8}$$

$$J_3 = \frac{\sigma_u}{u_{\max}} \tag{7.9}$$

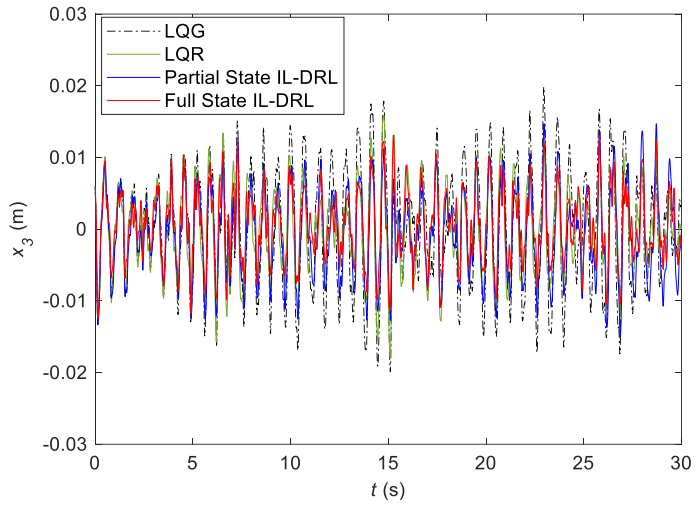
$$J_4 = \max\left[\left|\frac{d_1}{x_{3\mathrm{uc}}}\right|, \left|\frac{d_2}{x_{3\mathrm{uc}}}\right|, \left|\frac{d_3}{x_{3\mathrm{uc}}}\right|\right] \tag{7.10}$$

$$J_5 = \max\left[\left|\frac{a_1}{a_{3\mathrm{uc}}}\right|, \left|\frac{a_2}{a_{3\mathrm{uc}}}\right|, \left|\frac{a_3}{a_{3\mathrm{uc}}}\right|\right] \tag{7.11}$$

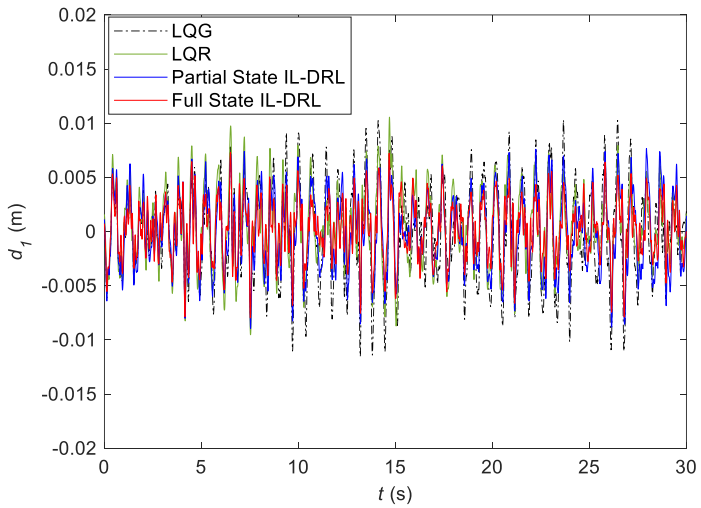
$$J_6 = \max\left[\frac{|u|}{u_{\max}}\right] \tag{7.12}$$

where the subscript uc denotes the uncontrolled measurements, and σ represents the RMS value. $\sigma_{x_{3\mathrm{uc}}}$ is the RMS displacement of the roof of the uncontrolled structure, and σ_{d_i} is the RMS inter-story drift of the *i*th story. σ_{a_i} is the RMS acceleration of the *i*th floor, and $\sigma_{a_{3\mathrm{uc}}}$ is the RMS roof acceleration of the uncontrolled structure. J_1 evaluates the maximum RMS of inter-story drift relative to the uncontrolled state, providing insights into the controller's capability to minimize building sway. J_2 captures the peak RMS of the absolute floor acceleration across different floors, normalized by the uncontrolled scenario. This variable indicates the occupants' comfort levels during controlled and uncontrolled vibrations. J_3 provides a measure of the control power required, measured in terms of the RMS of the control voltage applied to the AMD, scaled by the maximum permissible voltage $u_{\max}$. J_4 quantifies the peak inter-story drift normalized by the uncontrolled state, offering a measure of the structural integrity under control. J_5 measures peak acceleration levels, similar to J_2,

but focuses on the absolute maximums. J_6 captures the peak control voltage, providing another dimension to evaluate the required control resources.



(a)



(b)

Figure 7.8 Controlled response under random excitation comparison (a) roof displacement (b) first floor drift

Figure 7.8 shows the responses of the floors under white-noise shaking table excitation, where only 30-s time histories are shown for clarity. Figure 7.8a shows the displacement of the top floor, while Figure 7.8b focuses on the inter-story drift experienced by the first story. Both the IL–DRL controllers demonstrate superior peak

control, although the full-state feedback controller exhibits marginally improved performance. The LQR controller's results closely resemble those of the IL–DRL controllers, whereas the LQG exhibits significant degradation.

Table 7.2 compares these performance metrics across the different controllers during 250-s random excitation. The full-state IL–DRL controller achieved the lowest normalized RMS inter-story drift ($J_1 = 0.25$) and peak drift ($J_4 = 0.26$), which outperformed classical LQR ($J_1 = 0.28$) and LQG ($J_1 = 0.31$) by 10.7% and 19.4%, respectively. The fine-tuning process further reduced the peak drift by 30.6% compared with the pre-trained IL controller (before fine-tuning) ($J_4 = 0.37$). In terms of acceleration control, the full-state IL–DRL controller ($J_2 = 0.46$) matched the LQR's performance ($J_2 = 0.45$) and surpassed LQG's ($J_2 = 0.5$) by 8%, effectively balancing occupant comfort and structural safety.

The full-state IL–DRL controller also demonstrated better energy efficiency than the LQR and LQG in terms of peak control voltage ($J_6 = 0.48$), consuming 4% and 20% less than the LQR ($J_6 = 0.50$) and LQG ($J_6 = 0.60$) controllers, respectively. Although the RMS control voltage ($J_3 = 0.13$) of the full-state IL-DRL was marginally higher than LQR ($J_3 = 0.12$, 8.3% increase), this minor trade-off is significantly outweighed by its substantial improvements in vibration suppression.

The partial-state IL-DRL controller achieved comparable performance to the full-state version in acceleration control (J_2), despite using only limited sensor data. Therefore, the partial-state IL-DRL proves robust to incomplete measurements, offering a cost-effective, fault-tolerant solution for real-world deployments where full sensor coverage is impractical. Moreover, compared with the LQG controller, the partial-state IL-DRL controller performed comparably in drift reduction (J_1 and J_4), showed a 9% improvement in peak acceleration (J_5), but required slightly higher peak control effort (J_6).

Table 7.2 Performance metrics of different controllers under random excitation

Controller	J_1	J_2	J_3	J_4	J_5	J_6
IL (before fine-tuning)	0.36	0.67	0.06	0.37	0.78	0.21
Full-state IL–DRL	0.25	0.46	0.13	0.26	0.63	0.48
Partial-state IL–DRL	0.31	0.46	0.16	0.30	0.61	0.63
LQR	0.28	0.45	0.12	0.28	0.63	0.50
LQG	0.31	0.50	0.15	0.31	0.67	0.60

From the comparative analysis of the four controllers, it is easy to observe that the full-state IL-DRL demonstrates superior drift control performance (J_1 and J_4), attributable to its reward function, which explicitly penalizes maximum inter-story drift. In contrast, the other controllers exhibit weaker drift mitigation. While both IL-DRL controllers include an acceleration penalty in their reward design, their acceleration control performance differs only marginally (J_2 and J_5). These results demonstrate a critical advantage of the multi-objective IL-DRL framework: unlike conventional LQR/LQG controllers with fixed cost functions, IL-DRL enables flexible and customizable control objective formulation, particularly for challenging tasks like drift suppression.

Figure 7.9 plots the 10-s control voltage under the random excitation to show the control energy usage more directly. The full-state IL-DRL controller demonstrates superior voltage regulation. While the partial-state IL-DRL shows greater voltage fluctuation than its full-state counterpart, its peak magnitudes remain comparable to LQG. The LQR controller demonstrates intermediate performance, consistent with the quantitative indices J_3 and J_6 discussed above.

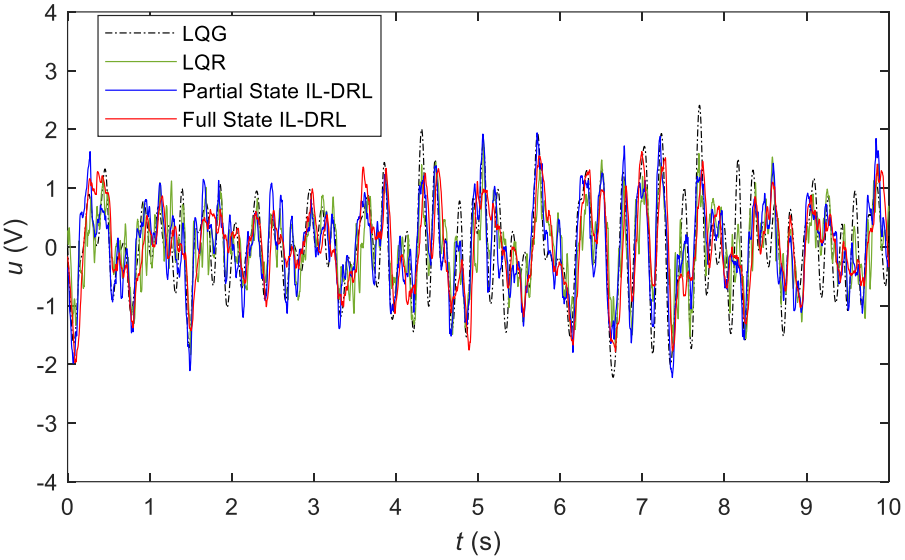


Figure 7.9 Comparison of control voltage u under random excitation

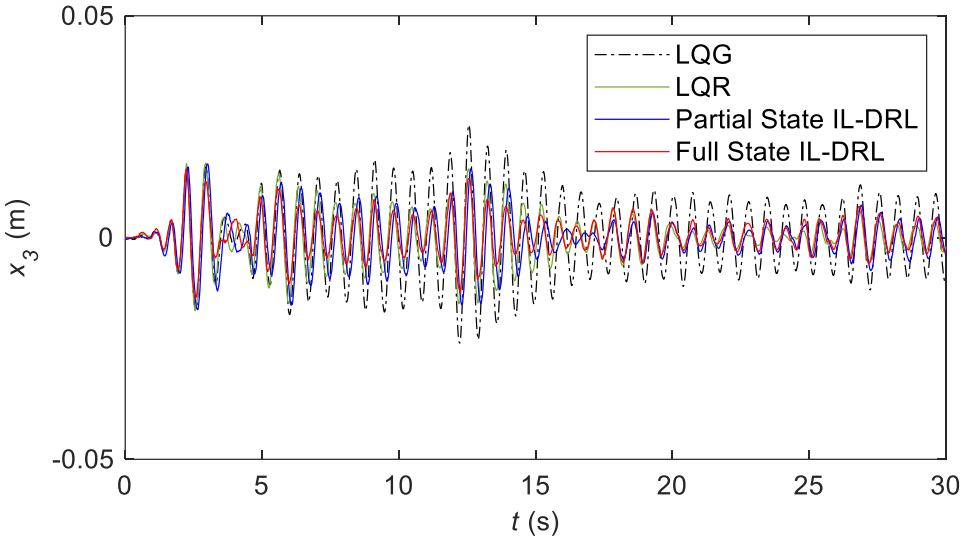
These results collectively underscore IL-DRL's robust control capability, with the full-state version delivering improvements in structural response metrics while maintaining competitive energy usage and model-free nature, substantiating its advantages over classical model-based approaches for random excitation scenarios. The results further confirm that the slight energy increase is substantially outweighed by the significant gains in vibration suppression performance.

7.5.4 Earthquake excitation

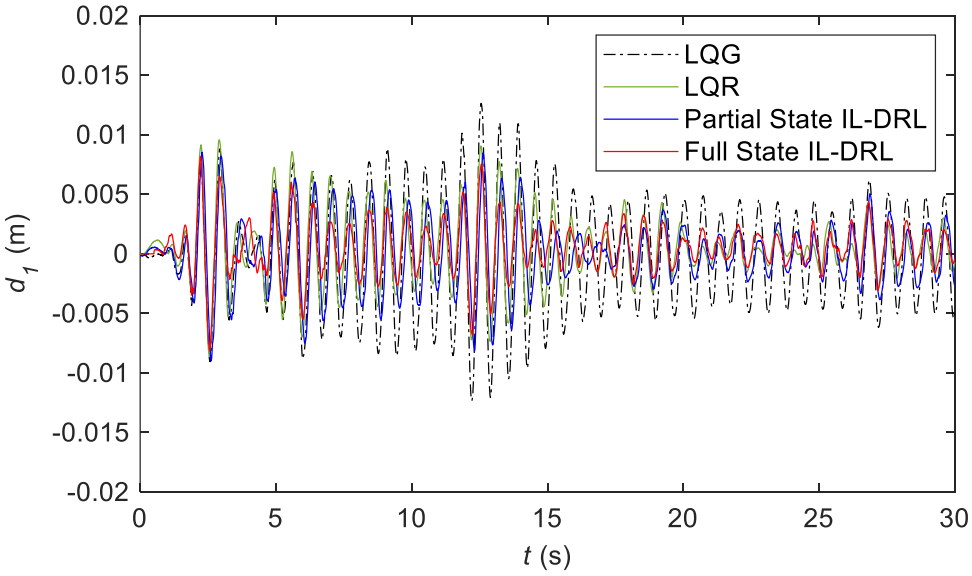
The structure was further tested under an earthquake excitation (i.e., a scaled El Centro earthquake) to validate the efficacy of the proposed IL-DRL controller. Notably, only the white-noise excitation was applied in the training stage. Figure 7.10 compares the response time histories of four different controllers. In an uncontrolled scenario, the peak relative displacement of the third floor reached 4.3 cm. The full-state IL-DRL controller showcases superior reduction capabilities, curtailing displacements by 64%, while the LQG manages only a 40% reduction.

Given the comparable RMS control voltage, the proposed IL-DRL controllers surpass the model-based controllers in controlling peak floor accelerations and displacements. Both IL-DRL controllers outperform the LQR and LQG. The performance enhancement can be attributed to the DRL fine-tuning process, enabling

the controllers to adapt to unforeseen excitations autonomously. This inherent flexibility of the IL-DRL controllers bestows an advantage in designing control systems, especially compared with the LQG controller that is primarily based on the assumption that the system and measurement noise is in the form of Gaussian white noise.



(a)



(b)

Figure 7.10 Controlled response under El Centro earthquake comparison (a) roof displacement (b) first floor drift

Table 7.3 compares these performance metrics across the different controllers under the earthquake excitation. The full-state IL-DRL controller achieved the lowest RMS inter-story drift ($J_1 = 0.12$), outperforming LQR and LQG by 14.3% and 42.9%, respectively. The acceleration control performance ($J_2 = 0.26$) of the full-state IL-DRL controller marked a 74.3% reduction from the pre-trained IL controller and a 13.3% improvement over LQR, while maintaining reasonable control energy demands (J_3 and J_6).

The partial-state IL-DRL controller closely matched the full-state IL-DRL controller in RMS drift (J_1) and acceleration (J_2) reduction, despite limited sensor inputs used in the feedback. Moreover, the partial-state IL-DRL controller outperformed LQG by 38.1% in drift control (J_1) and 36.6% in acceleration mitigation (J_2). Although this IL-DRL controller requires 56% more peak energy (J_6) than LQG, this cost is offset by the performance enhancement in structural response reduction. LQG's energy-efficient performance came at the expense of considerably poorer response control.

These results conclusively demonstrate that the IL-DRL strategy provides a better balance between exceptional seismic response mitigation and reasonable energy demands, even though it is trained under random excitations only.

Table 7.3 Performance metrics of different controllers under the El Centro earthquake excitation

Controller	J_1	J_2	J_3	J_4	J_5	J_6
IL (before fine-tuning)	0.47	1.01	0.14	0.90	0.78	0.47
Full-state IL–DRL	0.12	0.26	0.09	0.20	0.46	0.33
Partial-state IL–DRL	0.13	0.26	0.10	0.21	0.45	0.36
LQR	0.14	0.30	0.08	0.22	0.45	0.37
LQG	0.21	0.41	0.07	0.29	0.56	0.23

The analysis demonstrates consistent advantages under both random and seismic excitation scenarios. The full-state IL-DRL controller delivers optimal drift control performance (J_1 and J_4). In contrast, the partial-state IL-DRL exhibits relatively weaker drift mitigation since its reward formulation lacks direct drift measurements, though it maintains comparable acceleration control performance (J_2 and J_5). These results

provide quantitative evidence that the multi-objective IL-DRL framework successfully addresses fundamental limitations of conventional model-based approaches, inflexible cost function formulations. This methodological advancement enables superior vibration suppression capabilities for structural control applications, particularly in scenarios requiring adaptive balance between competing control objectives.

Figure 7.11 shows the PSD of the second-floor acceleration a_2, revealing distinct control performance characteristics among the different controllers. Both the IL-DRL controllers demonstrate superior vibration suppression capability at the fundamental structural mode (1.63 Hz). The full-state IL-DRL controller achieves the minimum recorded PSD magnitude, and even the partial-state IL-DRL shows significant reductions of 30.6% and 57.6% compared to the LQR and LQG controllers, respectively.

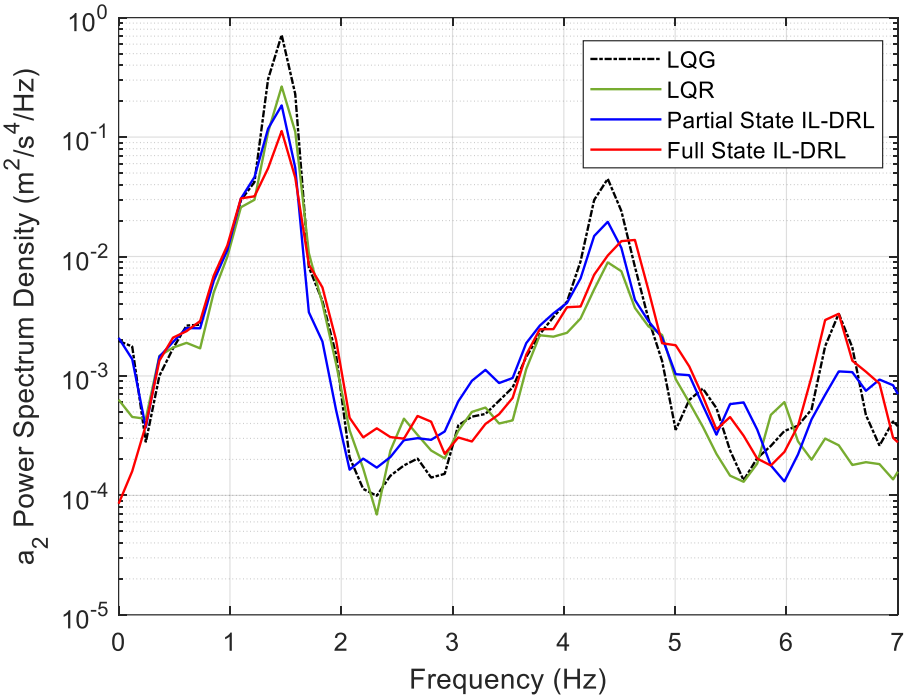


Figure 7.11. PSD of the second-floor acceleration

In contrast, while the LQR controller provides better attenuation at higher-order modes (4.59 and 6.31 Hz), the LQG controller demonstrates relatively poorer performance across the entire frequency spectrum. These results highlight the learning capability of the IL-DRL controllers, which emerges from their model-free adaptation to actual structural dynamics. The flexible control objective design enables the IL-DRL controllers to automatically identify and prioritize control of the structurally dominant

mode while optimally allocating control resources. This leads to more effective and energy-efficient vibration suppression compared to conventional LQR and LQG approaches.

7.6 Summary

This chapter extends the IL-DRL proposed in Chapter 5 to realize multi-objective control and tests it through laboratory experiments on an MDOF structure. Unlike Chapter 5 that still adopts the objective functions used by the classical model-based controllers, the multiple flexibly-selected objective functions in this chapter are integrated into the DRL's reward function, facilitating fine-tuning and allowing for the direct optimization of a diverse array of control objectives. By designing the reward function through different combinations of performance metrics, the proposed IL-DRL controller with full-state feedback achieved a 74% reduction in peak drift and 54% in RMS floor acceleration compared with the uncontrolled case under random excitation.

A series of shaking table experiments highlights the IL-DRL model's superior performance over classical model-based controllers (e.g., LQR and LQG) in diverse scenarios. Although the fine-tuning process is conducted only under random excitations, the full-state IL-DRL controller under the EI Centro earthquake excitation achieved 13% more reduction in RMS floor acceleration, while the peak voltage level is 11% lower, compared to LQR. The partial-state IL-DRL controller achieved similar performance with a 10% higher control voltage. Both IL-DRL controllers achieved 30% more peak drift reduction, and 36.6% RMS acceleration reduction compared with LQG. This innovative combination of DRL and IL offered enhanced capabilities in effectively attenuating vibrations and adapting to varied and unforeseen excitations.

By designing the reward function through different combinations of performance metrics, the proposed IL-DRL controller demonstrated remarkable performance in the experiment, providing insights into its potential for optimizing active control strategies in multiple control objectives. Despite the proposed strategy's promising results, the complicated nature of real-world structures necessitates further refinement and comprehensive validation. The uncertainties encountered in real structures pose persistent challenges, emphasizing the need for advancements in learning processes and

methodologies to affirm the reliability and robustness of the developed controllers in real-world conditions.

CHAPTER 8

Active Control of Wind-induced Vibration of a Benchmark TV Tower based on Multi-objective IL-DRL

8.1 Introduction

In this chapter, the real-world feasibility of the multi-objective IL-DRL-based active control strategy presented in Chapter 7 is further demonstrated by applying it to mitigate wind-induced vibration of a supertall structure, namely the Canton TV Tower. The Canton Tower with an AMD is selected as the benchmark case to evaluate the effectiveness under realistic conditions. The Hong Kong Polytechnic University established this numerical model as a benchmark problem for structural health monitoring (SHM) of high-rise structures (Ni, Xia et al. 2012). As a representative supertall and slender structure with low damping, the Canton Tower is inherently susceptible to wind (Ni, Wong et al. 2011). Wind-induced vibrations not only affect the comfort of the people inside but may also result in severe structural damage under extreme winds (Lin, Song et al. 2017).

Various levels of wind excitations are applied to evaluate the effectiveness and performance of the control strategy. To the best of the authors' knowledge, this study presents the first implementation of a model-free IL–DRL control strategy on a supertall structure, marking a novel contribution to the field of intelligent vibration control in real-world structures.

The remainder of this chapter is structured as follows: Subchapter 8.2 describes the benchmark problem, the Canton Tower. Subchapter 8.3 explains the AMD setup. Subchapter 8.4 shows how the IL-DRL is adapted, and subchapter 8.5 presents the results of the wind excitation analyses and a discussion thereof. Finally, subchapter 8.6 provides a summary of this study.

8.2 Dynamical Properties of the Canton Tower

8.2.1 General introduction

The Canton Tower (formally known as the Guangzhou New TV tower) is a 610 m-tall landmark structure in Guangzhou, China, primarily serving tourism and observation purposes. As shown in Figure 8.1(a), the tower comprises a 454 m-high main tower and a 156 m-high antenna mast made of steel (Ding and Zhu, 2017). The total weight of the tower is nearly 2×10^5 tons. The main tower structure is in a tube-in-tube form, composed of a concrete inner cube and a steel lattice outer tube that are connected on 37 different floors. Over 700 sensors, including accelerometers, anemometers, wind pressure meters, strain gauges, a seismograph, and a weather station, were installed, forming a comprehensive SHM system for in-service condition monitoring. This SHM system enables the integration of online health monitoring and real-time feedback for vibration control. The Canton Tower is equipped with hybrid mass dampers combining a two-stage passive TMD with an AMD, in which the AMD is driven by linear induction motors. The system defaults to the passive TMD operation if the active control fails (Tan et al., 2016). The design of the existing AMD control system is detailed in Xu et al. (2014).

8.2.2 Benchmark model

This subchapter employs the validated benchmark finite element model, originally developed for the Canton Tower's SHM system (Ni et al., 2012), for data generation. The reduced-order model strategically divides the structure into 37 beam elements, with 27 representing the main tower and 10 for the antenna mast. As shown in Figure 8.1b, this discretization creates 38 nodal points along the height, beginning at the ground (node 1) and terminating at the mast apex (node 38). Each intermediate node possesses five dynamic degrees of freedom (DOFs), namely, lateral translations in two orthogonal directions, and rotations about three axes, while excluding vertical translation due to its minimal influence on the overall structural lateral response.

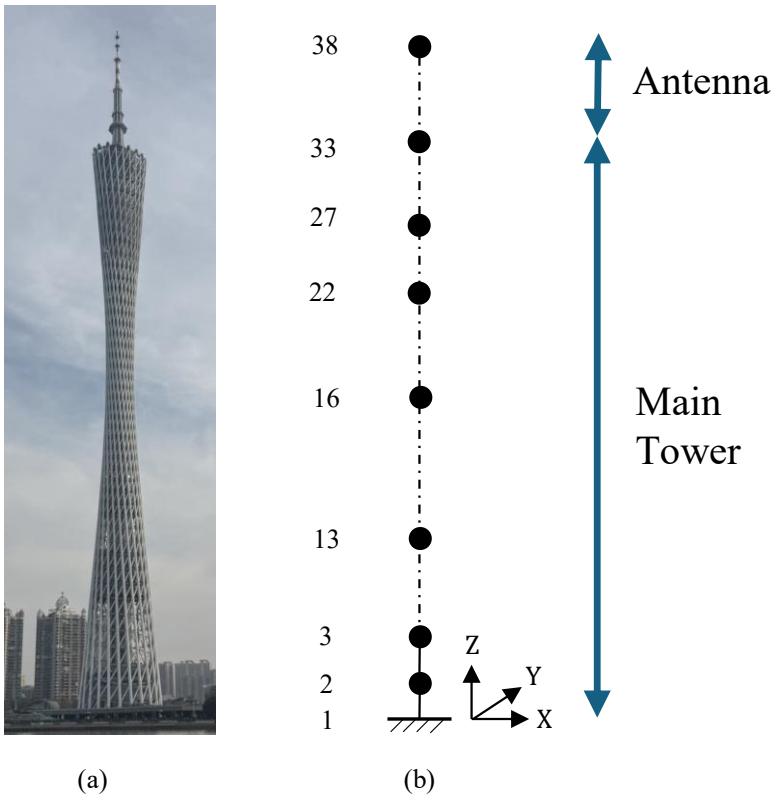


Figure 8.1 Canton Tower: (a) overall view; (b) reduced-order model

Table 8.1 lists the first eight natural frequencies identified from the reduced-order model and validated by the SHM data. Please refer to Chen et al. (2012) for further details on the dynamical properties of the Canton Tower. The structural short–axis of the inner tube has an 18° angle relative to the global X axis, as shown in Figure 8.2.

Table 8.1 First eight natural frequencies of the Canton Tower (Chen et al., 2012)

Mode	Value (Hz)	Description
1	0.110	Short-axis bending
2	0.159	Long-axis bending
3	0.347	Short-axis bending
4	0.369	Long-axis bending (mast)
5	0.400	Short-axis bending
6	0.462	Torsion
7	0.487	Long- and short-axis bending
8	0.738	Short-axis bending

The stochastic wind speed generation and the shape coefficient adopted in this study follow the previous wind tunnel test on an aeroelastic model conducted at Tongji University (Ding and Zhu, 2017).

The target profile of turbulent intensity adopted is

$$I_u = 0.1\left(\frac{Z}{H_G}\right)^{-\alpha-0.05}, \tag{8.1}$$

where the exponent α and the gradient height H_G are chosen as 0.22 and 400 m, respectively, for terrain type C in accordance with the Chinese Load Code for the Design of Building Structures (MCC MoCPRC). Z is the measurement height. The approximate windward area for each node is calculated on the basis of the benchmark model. The wind load is decomposed into two orthogonal directions (global X and Y) and applied to the nodes of the benchmark model.

8.3 AMD Control System

8.3.1 AMD parameters

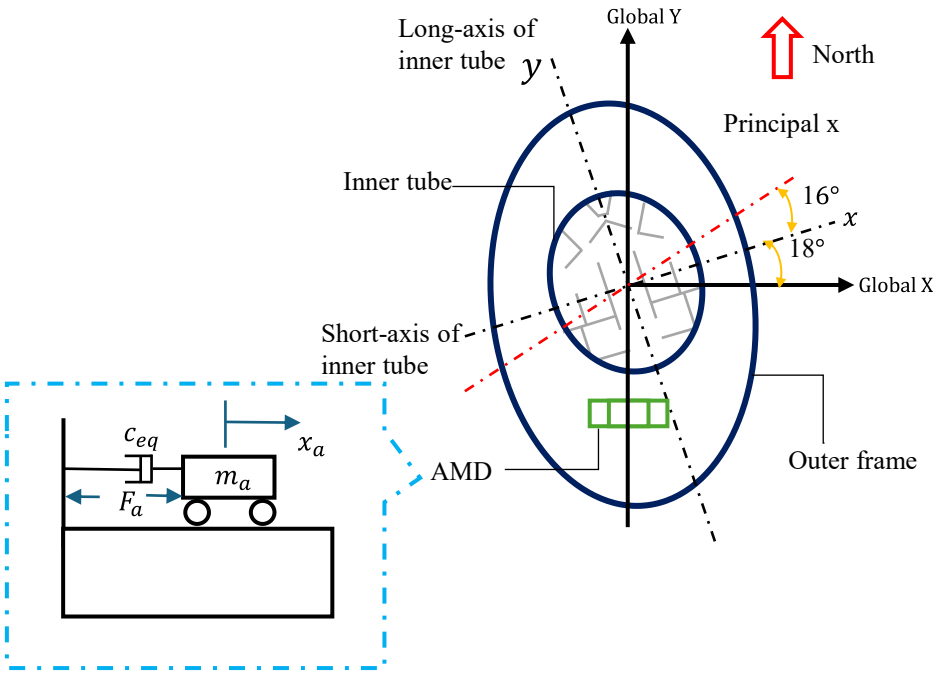


Figure 8.2 Tower section view and AMD at 438m

With reference to the modal analysis work (Xu et al., 2014), the principal x-axis direction of the bending modes is at a 16° angle relative to the short axis *x* of the inner tube, as shown in Figure 8.2. Accelerometer and velocity transducers are installed at heights of 168, 227, 386, 438, and 529 m, as shown in Figure 8.3. In the current Canton Tower, a set of two AMD control subsystems is installed on water tanks on the tower top (438 m) to suppress vibrations in the principal x-direction, where the water tank can also function as TMD inertial masses. But this study takes a different design. For the proposed data-driven methodology, no prior knowledge of the structural model and its corresponding principal bending directions is assumed. Using a fixed coordinate direction allows for the simple evaluation of the controller's effectiveness independent of modal principal direction assumptions. As shown in Figure 8.2, the AMD controller in this study is placed at the top of the main tower (node 27 at 438 m height), aligned with the global X-direction of the coordinate system, rather than following any presumed principal vibration direction or the short/long axes. All vibration responses

and control performance metrics discussed subsequently are reported in this global X-direction.

The dynamic model of the AMD system can be described as

$$F_{\mathrm{a}} = c_{\mathrm{eq}} v_{\mathrm{a}} + m_{\mathrm{a}} a_{\mathrm{a}}, \tag{8.2}$$

where F_{a} is the actual force on the tower generated by the AMD, as shown in Figure 8.2; $c_{\mathrm{eq}} v_{\mathrm{a}}$ is the resistance force mainly caused by the bearings; v_{a} is the relative velocity of the moving mass; m_{a} is the AMD mass; and a_{a} stands for the acceleration of the mass. The specifications of the AMD system adopted are listed in Table 8.2. The values for the AMD controller design are chosen with reference to Xu et al. (2014), close to the combination of the two AMD subsystems.

Table 8.2 Specification of the AMD system

Item	Value
Mass (t)	100
Maximum stroke (m)	$\pm$5
Maximum velocity (m/s)	1
Maximum thrust (kN)	300
Viscous coefficient c_{eq}(N·s/m)	11170

8.3.2 Feedback control system

The design of the sensor placement for the AMD controller also aligns with the feedback system described by Xu et al. (2014). The velocity and acceleration responses of the main tower at heights of 168, 227, 386, and 438 m are measured (i.e., nodes 13, 16, 22, 27). An additional accelerometer is placed on the antenna at 529 m (node 33). The relative displacement and velocity of the AMD to the top floor of the tower are also measured. The sensor locations are shown in Figure 8.3, corresponding to a total of 11 states in the feedback vector.

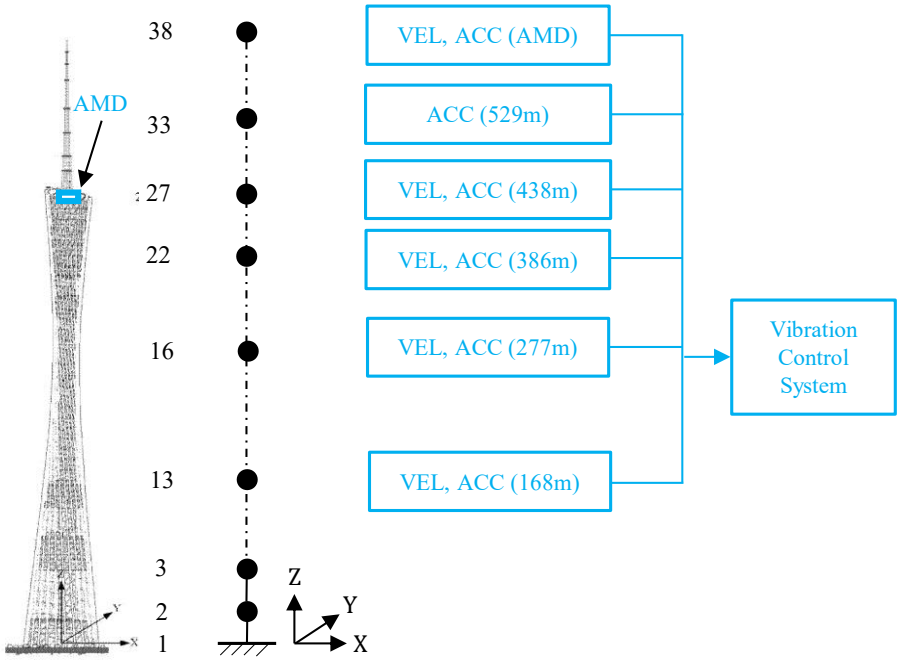


Figure 8.3 AMD control system

The state-space expression of the equation of motion of the tower with AMD control system is presented below:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}F_{\mathrm{a}} + \mathbf{B}_{\mathrm{w}}\mathbf{F}_{\mathrm{w}} + \mathbf{w}, \tag{8.3}$$

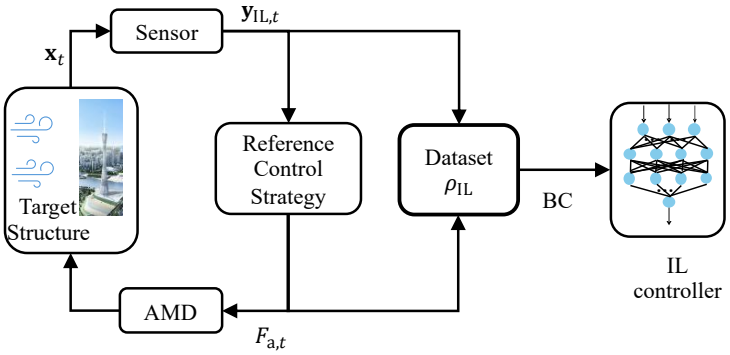
$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}F_{\mathrm{a}} + \mathbf{D}_{\mathrm{w}}\mathbf{F}_{\mathrm{w}} + \mathbf{v}, \tag{8.4}$$

where $\mathbf{x}$ is the state vector of the system; $\mathbf{y}$ is the observation vector; $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{B}_{\mathrm{w}}$, and $\mathbf{D}_{\mathrm{w}}$ are the corresponding system matrices. $\mathbf{F}_{\mathrm{w}}$ denotes the wind excitation force. $\mathbf{w}$ and $\mathbf{v}$ are the system error vector and measurement noise vector, respectively. The aim is to develop a control policy $F_{\mathrm{a}} = g(\mathbf{y})$ to determine the control force of the AMD based on the observations. When the exact knowledge of system matrices ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$) is known and partial states are observable, the LQG control can be applied. Two LQG controllers are later used for comparison. The methodology of LQG is introduced in subchapter 6.2.

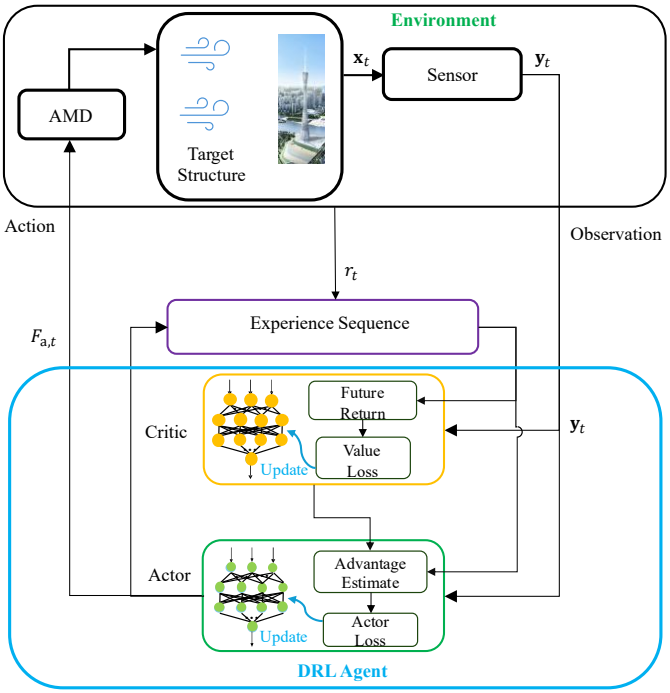
In the IL-DRL controller, the system matrices are assumed to be unknown.

8.4 Multi-objective IL-DRL Controller

The multi-objective IL–DRL approach proposed in Chapter 7 is used for controlling wind-induced vibration of the Canton Tower. As illustrated in Figure 8.4a, a simple reference control strategy is provided to the Canton Tower for pre-training an NN controller at first. The IL–trained NN is then transferred to the DRL agent and fine–tuned under free vibration, as shown in Figure 8.4b.



(a)



(b)

Figure 8.4 IL-DRL for Canton Tower control (a) IL pre-train; (b) DRL fine-tune

8.4.1 IL pre-train

The IL adopted is the BC method used in Chapter 7. Because the aim is to achieve a model–free controller and avoid the use of model dynamics, an all–state control gain derived from the optimal control methods is assumed unavailable. Instead, an extremely simple feedback controller is adopted here to be a reference strategy, which uses only three physically measurable states: the top floor velocity of the main tower (v_{27}) and the relative displacement (x_{a}) and velocity (v_{a}) of the AMD. The control force is computed as

$$F_{\mathrm{a}} = -\mathbf{K}_{\mathrm{r}}[v_{27};\ x_{\mathrm{a}}; v_{\mathrm{a}}], \tag{8.5}$$

where the gain matrix $\mathbf{K}_{\mathrm{r}} = [-10^5, 10^3, 1]$. The gain $\mathbf{K}_{\mathrm{r}}$ is selected to achieve two primary objectives. The negative gain -10^5 provides high damping for the tower's first bending mode through velocity feedback, and the smaller gains on AMD states $[10^3, 1]$ ensure the AMD stability by generating restoring forces and limiting excessive actuator motion (i.e., providing positive stiffness and damping). However, this over-simplified reference control strategy can by no means provide an optimal control performance.

The training data for IL are acquired through a one-time two-step procedure: an initial excitation phase in which the AMD applies a 55-s harmonic force with a 0.1-Hz frequency and a 9-kN force amplitude, followed by a free vibration phase in which the AMD switches to the reference control mode (i.e., Equation 8.5) to regulate the structural response. The resulting free vibration data are recorded at a 50-Hz sampling frequency for 500 s.

As shown in Figure 8.4a, the collected training data are used to train an NN controller. Specifically, the collected dataset ρ_{IL} is sliced into observation–action pairs $(\mathbf{y}_{\mathrm{IL},t}, F_{\mathrm{a},t})$. Then, BC is applied to train the NN controller $\mu_\theta(\mathbf{y}_{\mathrm{IL},t})$ to mimic the reference policy by minimizing the MSE error. The IL training utilizes only post-excitation free-vibration data, potentially restricting generalization to unseen excitation types.

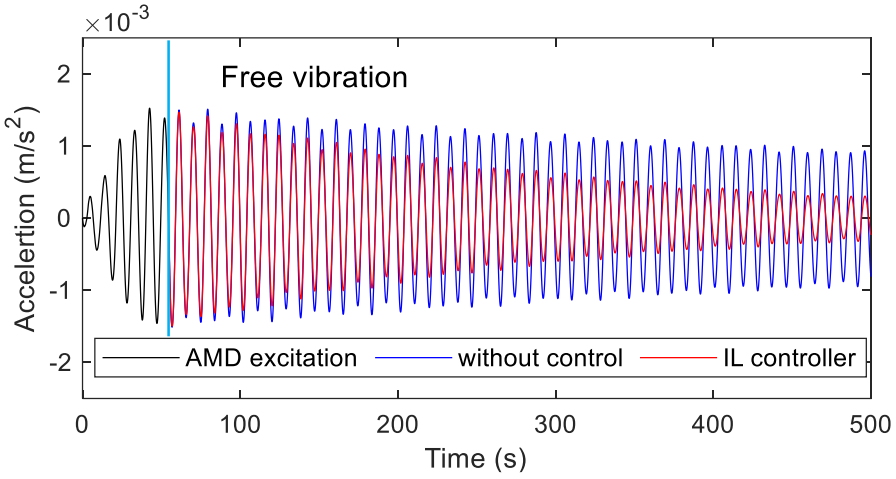


Figure 8.5 Acceleration of the structure at the tower top (at a height of 438 m) under harmonic excitation and free vibration

Figure 8.5 compares the free-vibration acceleration decay of the uncontrolled and IL-controlled cases after the initial harmonic excitation. The acceleration response in the uncontrolled case decays more slowly than the IL-controlled case. The former and latter correspond to damping ratios of 0.31 % and 0.70%, respectively.

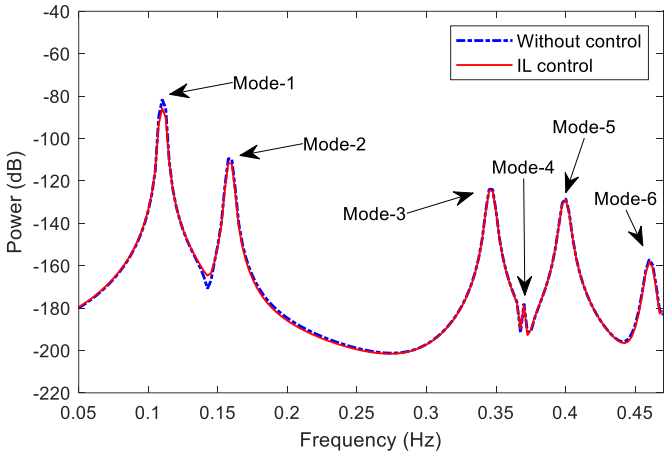


Figure 8.6 PSD of the free-vibration acceleration at the tower top (at a height of 438 m)

Figure 8.6 compares the PSD plot of the tower-top acceleration during free vibration between the uncontrolled and IL-controlled responses. A moderate reduction in the first six modal peaks can be observed. The IL controller reduced the amplitude of the first mode by -3.74 dB (i.e., 35%) compared with the uncontrolled one, and the rest modes achieved -1.13 to -2.39 dB (i.e., 12%–24%) reduction. As aforementioned, by

mimicking an oversimplified reference strategy, the IL controller's performance is far from optimal. However, the adopted reference control strategy ensures a stable initial control policy that contributes slight positive damping to the structure. The next step is to improve control performance through DRL

8.4.2 Multi-objective DRL fine-tune

Having the IL controller as the initial policy, the subsequent step is to fine-tune it through DRL for optimal control performance. The procedure follows the description in Chapter 7. In this case, the control environment is the Canton Tower, and the DRL agent acts like the controller in a classical active control system. As shown in Figure 8.4b, the agent receives the sensor measurement state $\mathbf{y}_t$ at time t, determines the control signal $F_{\mathrm{a},t}$ on the basis of the current control policy $\mu_{\theta,t}(\mathbf{y}_t)$, and applies the control force $F_{\mathrm{a},t}$ to the environment through AMD. The agent's performance is judged by reward $r_t = f(\mathbf{y}_t, F_{\mathrm{a},t})$, with the objective of maximizing the discounted sum of future rewards.

The observation state $\mathbf{y}$ consists of the velocity at nodes (13, 16, 22, 27), acceleration at nodes (13, 16, 22, 27, 33), and the relative displacement (x_{a}) and velocity (v_{a}) of the AMD, i.e.,

$$\mathbf{y} = [v_{13,}, v_{16}, v_{22}, v_{27}, a_{13}, a_{16}, a_{22}, a_{27}, a_{33}, x_{\mathrm{a}}, v_{\mathrm{a}}]^T. \tag{8.6}$$

Notably, the observation state $\mathbf{y}$ herein contains more state variables than that used in IL training (see Equation 8.5). As previously described, the classical optimal controllers derive the control strategy by minimizing a specific cost function shown in Equation 3.4. However, such a cost function in a rigid form may not be able to incorporate some primary control objectives of interest. In contrast, the DRL can address this limitation by defining its reward function in a flexible form. The following reward function is adopted in this study:

For $F_{\mathrm{a},t} \leq 90\% F_{\mathrm{a,max}}$,

$$r_t = -\mathbf{G}_{\mathrm{r}} \times [x_{\mathrm{a}}^2, v_{\mathrm{a}}^2, \max(|v_i|), \max(|a_i|), \mathrm{rms}(v_{i,t}), \mathrm{rms}(a_{i,t}), F_{\mathrm{a},t}^2]^T. \tag{8.8}$$

This function serves as a combination of the AMD motion and the structure

performance metrics.

For $F_{\mathrm{a},t} > 90\%F_{\mathrm{a,max}}$,

$$r_t = -\mathbf{G}_\mathrm{r} \times \left[x_\mathrm{a}^2, v_{\mathrm{a},t}^2, \max\left(|v_i|\right), \max(|a_i|), \mathrm{rms}\left(v_{i,t}\right), \mathrm{rms}\left(a_{i,t}\right), F_{\mathrm{a},t}^2\right]^T e^{\frac{F_{\mathrm{a},t}}{0.9F_{\mathrm{a,max}}}} \quad (8.9)$$

where i = 13, 16, 22, 27 stands for the four locations with sensors. As shown in Figure 8.3, all four nodes are on the tower, and no antenna nodes are selected in the reward function. This reward function combines the AMD motion and multiple structure performance metrics. The construction of the reward function is directly based on the control objectives. Given that the performance metrics for structure vibration control are often based on the reduction of the maximum displacement and acceleration responses, the maximum value of the sensor measurement, $\max\left(|v_i|\right)$, and $\max(|a_i|)$, along the tower height are included directly. The function rms( ) computes the root-mean-square (RMS) value of the sensor measurement at all locations (i = 13, 16, 22, 27) at time t. The RMS value of the acceleration response is added to consider human comfort. The relative displacement and velocity of the AMD are included to ensure that its movements are within the desired range. The control force is also involved to achieve a balance between control performance and power demand. The reward weighting matrix $\mathbf{G}_\mathrm{r}$ is chosen as $[4\times 10^{-5}, 4\times 10^{-6}, 0.5, 1, 1, 2, 1\times 10^{-12}]$ to normalize the measured response values of different orders of magnitude. The DRL algorithm adopted is PPO.

The training is conducted in *MATLAB* on an Intel i7 9700 CPU. The batch size is 256. The architecture for the actor network includes two hidden layers with 64 nodes for each and employs a hyperbolic tangent (Tanh) activation function. The actor's learning rate is set to 5×10^{-4}. The critic network is constructed using three hidden layers with 256 nodes in each layer, with rectified linear units as the activation function and a learning rate of 0.001. Before the start of each episode, the AMD is used as an exciter to provide a harmonic excitation at 0.1 Hz for 55 s. With the same setup as that in IL, the force amplitude is kept at 9 kN. Afterward, the AMD is switched to the controller mode and follows the policy of the PPO agent for 500 s with 50 Hz sampling frequency.

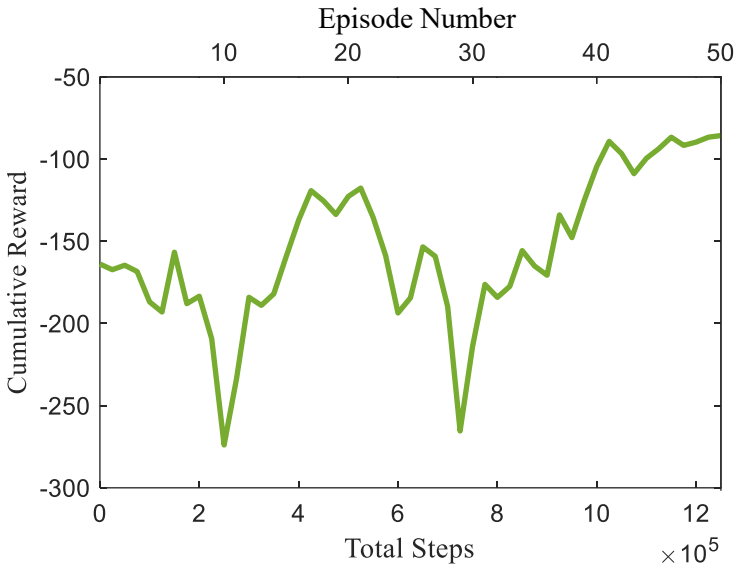


Figure 8.7 Cumulative reward plot of DRL training

Figure 8.7 shows the reward curve over the DRL training. The training takes 50 episodes, implying that the above excitation and free-vibration tests need to be conducted 50 times. Each episode contains 500-s free-vibration data (2.5×10^4 time steps), with a total of 1.25×10^6 s free vibration data. Figure 8.7 shows a rapid training process. At step 0, the training curve originates from the initial performance of the IL controller obtained in subchapter 8.4.1, which already demonstrates fair effectiveness with inherent stability and moderate vibration reduction capability. The two sudden performance drops, occurring around episodes 10 and 30, arise from peak force constraint violations in the agent's exploration, triggering the penalty terms in the reward function in Equation 8.9. In comparison, if a randomly initialized DRL agent is trained using the same procedure, no reasonable control strategy can be obtained even after training with 100 times more data. This indicates that a pure DRL controller without IL pre-training cannot be practically trained through real structural tests.

8.5 Results and Discussions

To comprehensively investigate the performance of the proposed IL–DRL controller, this subchapter conducts a series of simulations on the vibration control of the Canton Tower under strong wind. The wind speed design value is 36.7 m/s at 10 m above the ground for a 10-year return period. The time history of the wind speed at the tower top is generated from the Davenport spectrum, as shown in Figure 8.8, where the

duration of the turbulent wind is 720 s and the mean wind speed has been removed.

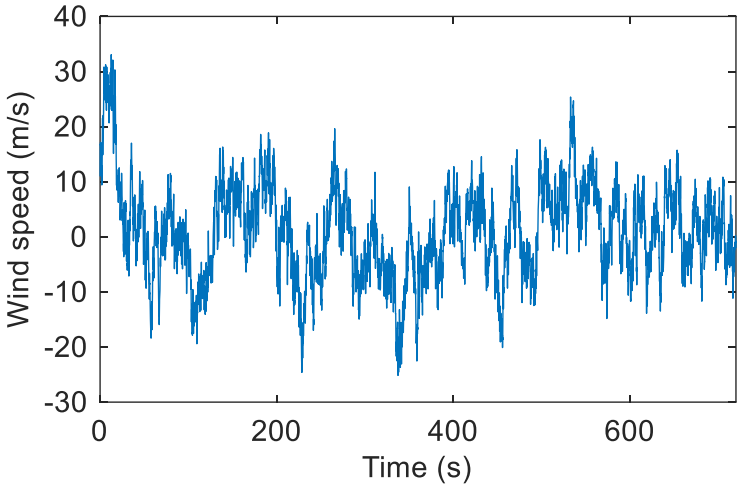
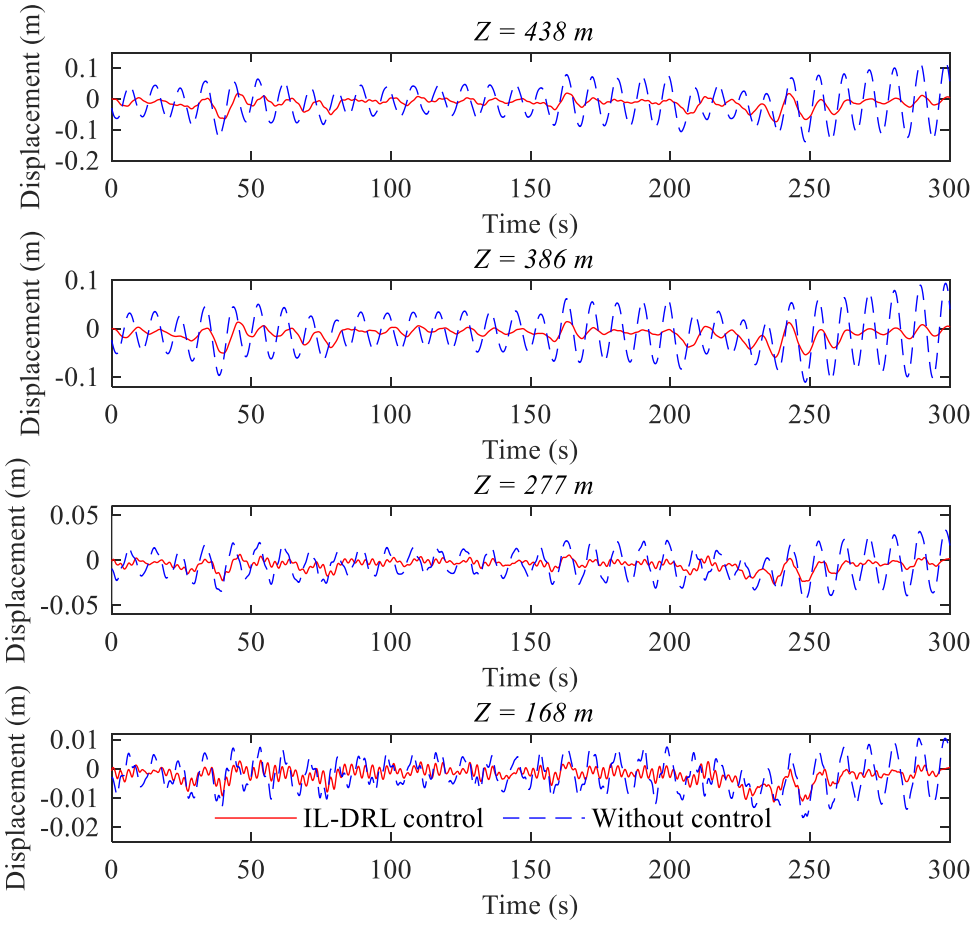


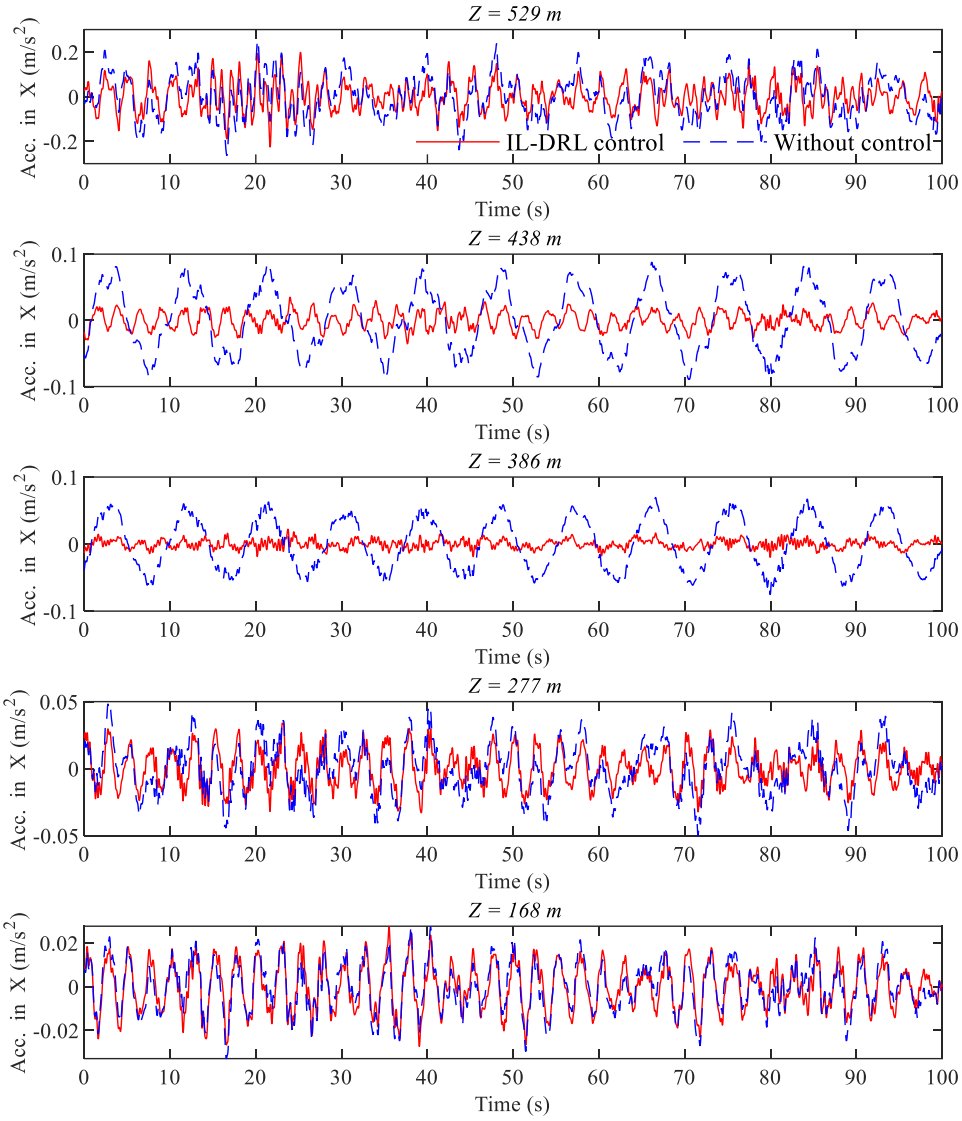
Figure 8.8 Wind speed generated on top of the main tower

Figure 8.9a and 8.9b compare the IL–DRL controller case with the uncontrolled case by plotting the dynamic time histories of the displacement and acceleration responses at different measurement points under the generated wind with an attack angle of 225°. The static component caused by the mean wind speed is removed. In both figures, the IL–DRL control scheme achieves significant reductions in structural responses, thereby improving the serviceability and safety of the structure. From the base to the top of the tower, an increase in the magnitude of uncontrolled vibration can be observed. The vibration near the tower top is significantly reduced as anticipated, given that the AMD directly counteracts vibrations at its point of installation. The vibration reduction is less pronounced at lower heights (e.g., 168 m). This pattern is particularly noticeable in the acceleration data, in which the vibration reduction effect intensifies with increasing height. This is because the design of the reward function in Equation 8.8 inherently emphasizes upper-floor responses strongly. It is noted that higher-order vibration modes (the 3rd and above), which predominantly influence acceleration responses, exert greater impact on the lower parts (i.e., at the height of 168m and 277m), whereas lower-order vibration modes show dominant influence on the upper part (e.g., 438m). However, the acceleration amplitude is relatively small at the lower part. In addition, the achieved vibration attenuation in the antenna is less substantial compared with that at the tower top (Figure 8.9b). It indicates that the AMD

installed at the tower top cannot effectively suppress the higher-order bending modes of the mast, likely because these modes are not prominent in the free vibration used during the training and the reward function in Equation 8.8 does not consider the acceleration of the mast in this study. These observations collectively highlight the adaptive, objective-oriented characteristics of the IL–DRL methodology, which strategically distributes control resources to areas of greatest significance rather than employing uniform control effects throughout the structure.



(a)



(b)

Figure 8.9 Dynamic response time histories in the X-direction at different heights: (a) displacement; (b) acceleration

The performance of the classical LQG controller can be tuned by modifying the weighting matrices. Therefore, the proposed IL–DRL controller is compared with two LQG controllers, namely LQG1 and LQG2, to examine the effectiveness of the proposed IL–DRL strategy over the classical model-based methods. The weighting matrices of the LQG1 controller are tuned to have a similar control force level to that of the IL–DRL controller, while the LQG2 controller is chosen to achieve a comparable vibration reduction level at the tower top to the proposed controller. The added measurement noise is Gaussian white noise in all cases. The noise-to-signal ratio in

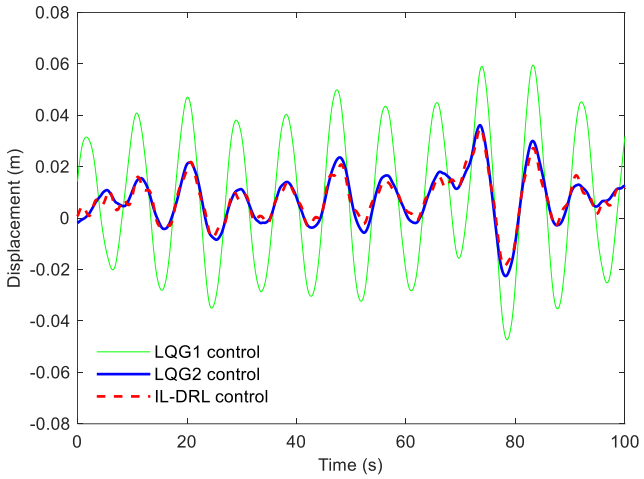
observation is set as 5%, while the noise in the control force is 10%.

Table 8.3 Comparison of peak and RMS responses under wind excitation

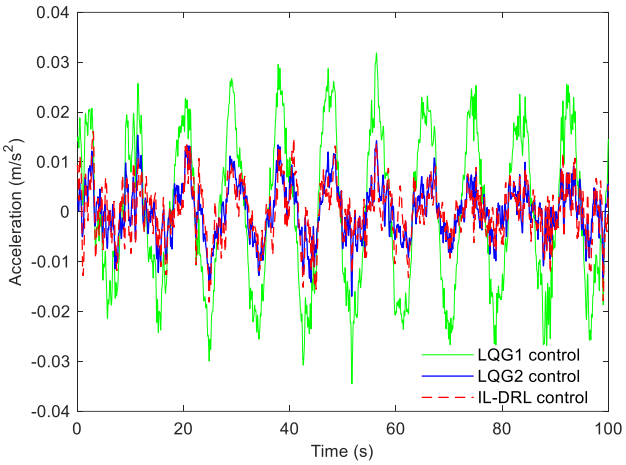
	UC	LQG1		LQG2		IL-DRL	
		Value	Red. %	Value	Red. %	Value	Red. %
Peak roof disp. (cm)	9.27	6.17	33	3.72	60	3.55	62
Peak roof acc. (cm/s^2)	8.96	3.78	58	1.97	78	2.33	74
RMS roof disp. (cm)	4.67	2.94	37	1.30	72	1.19	75
RMS roof acc. (cm/s^2)	4.74	1.63	66	0.59	88	0.64	86
Peak active force (kN)	-	146.78	-	249.28	-	152.60	-
RMS active force (kN)	-	52.48	-	79.12	-	47.96	-

Note: UC is the uncontrolled structure, LQG1 has control force levels comparable to IL-DRL, and LQG2 has response reduction comparable to IL-DRL.

A summary of the peak and RMS responses at the main tower top (438 m) is presented in Table 8.3, which also includes the reduction percentage compared to the value without control. The time histories of the displacement and acceleration responses on top of the main tower using three different controllers are also compared in Figures 8.11a and 8.11b, respectively. Figure 8.12a provides the time history plot of the AMD's control force, and Figure 8.12b shows the control force loops against the AMD's relative displacement. The variation trend and the control force loops of the AMD in IL–DRL are closer to those in LQG1. While LQG2 provides greater peak control forces, and consequently, increases the AMD relative displacement. IL–DRL successfully limits the AMD relative displacement within $\pm$2.5 m, half of the maximum stroke.



(a)



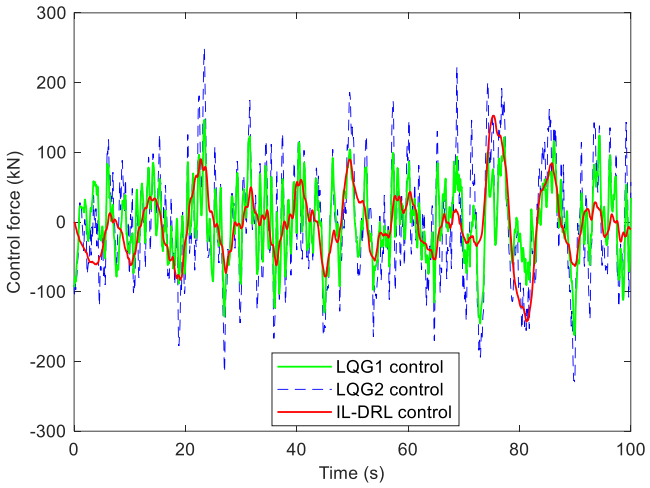
(b)

Figure 8.10 Dynamic time history response on top of the main tower under wind excitation: (a) displacement; (b) acceleration

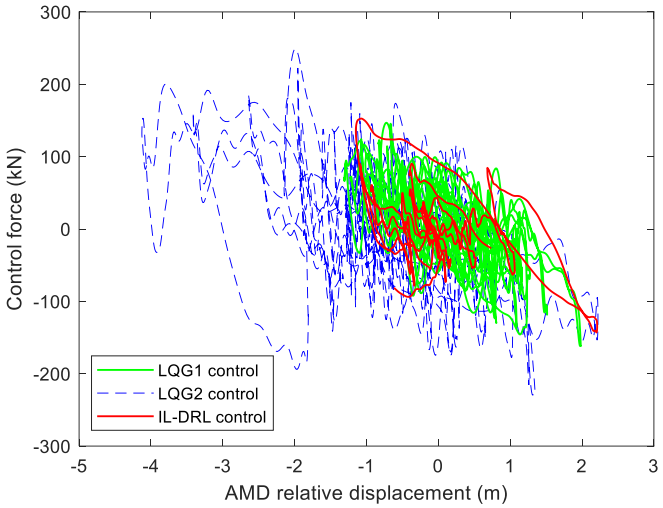
The comparison in Figures 8.11a and 8.11b indicates that the control performance of IL–DRL is close to that of LQG2, while LQG1 is obviously weaker than them. By contrast, the control force of IL–DRL is at the same level as that of LQG1 and significantly lower than that of LQG2. IL–DRL achieves the design objective of providing superior control performance with moderate energy consumption.

The IL–DRL controller reduces the peak displacement at the top of the tower by 62% in comparison with the uncontrolled case, almost twice as much as the LQG1 controller's reduction. The reduction in peak acceleration is 16% higher than that with

LQG1, with the exact values shown in Table 8.3. Hence, the proposed intelligent controller can protect the structure from excessive damage and provide better occupant comfort compared with the classical optimal controller. Although LQG2 achieves slightly better control performance in acceleration, the peak active force is much higher than that of the IL–DRL controller, and the RMS force is nearly 60% higher than that of IL–DRL. The control performance of IL–DRL in RMS acceleration also slightly falls behind that of LQG2, which may be due to the design of the DRL reward function that penalizes the velocity value more than the acceleration value.



(a)

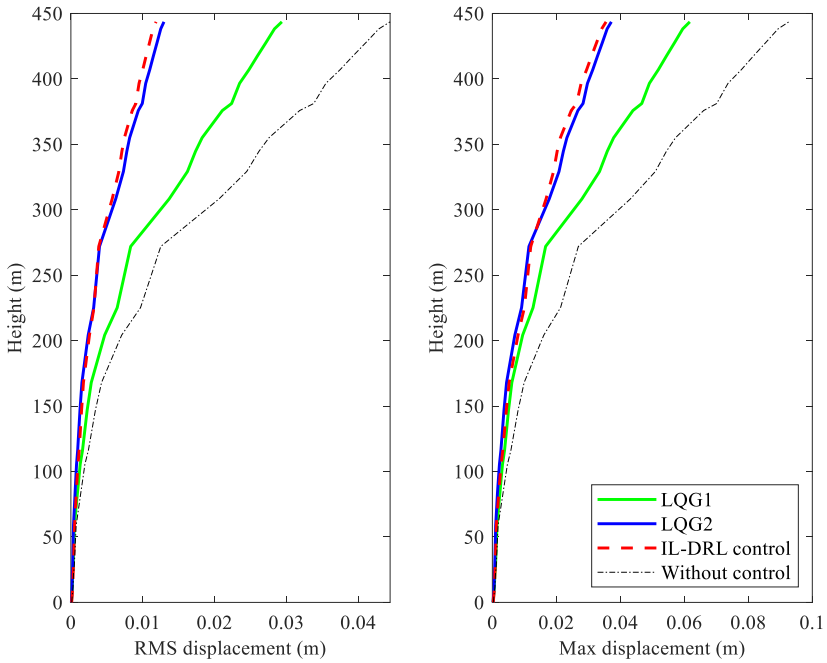


(b)

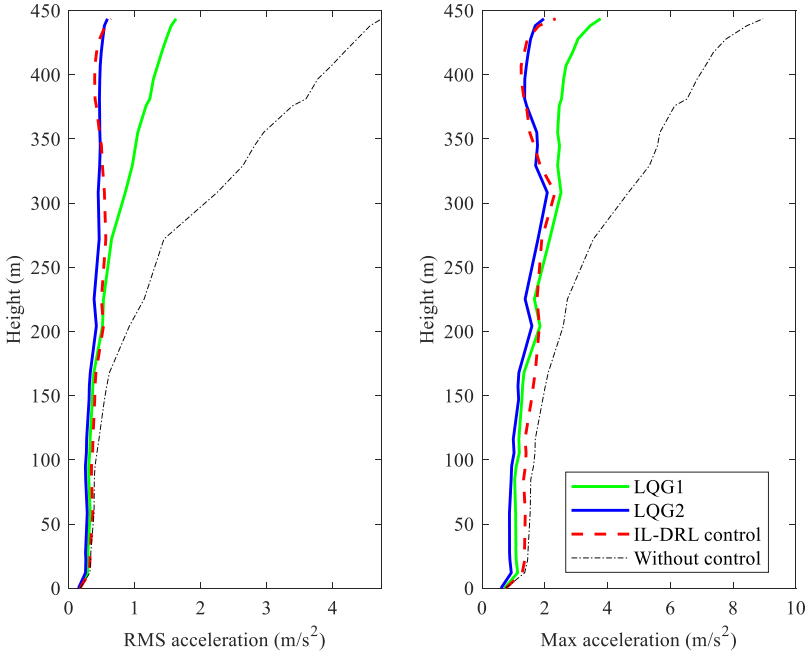
Figure 8.11 Control force under wind excitation: (a) time history; (b) control force against AMD relative displacement

The RMS and peak responses of the displacement and acceleration at different height levels are presented in Figures 8.12a and 8.12b, respectively, to comprehensively evaluate the performance of the proposed IL–DRL method. This evaluation provides a direct way to examine the control effect against heights by different strategies. The plots demonstrate that as heights increase, IL–DRL shows relatively better vibration control performance. This trend is more evident in the displacement plot. As illustrated in Figure 8.12a, IL–DRL demonstrates comparable control performance to that of the LQG2 controller, whereas the LQG1 controller exhibits a notable decline in its performance, particularly in the control of peak displacements. The gap between LQG1 and IL-DRL increases with height. Regarding the acceleration plot, the performance can be compared separately in two sectors. In the lower half (approximately below 225 m) of the tower, the acceleration reduction by IL–DRL is less than that by the two LQG controllers. While in the upper half of the tower, the performance of IL-DRL is comparable to that of LQG2 and considerably better than that of LQG1.

One possible explanation for this discrepancy is that the optimal controllers, LQG1 and LQG2, employ full-state control matrices, whereas IL–DRL is trained using only sensor measurements. The heights of the sensors placed on the tower are 168, 272, 381, and 438 m, of which three are in the upper half. The placement of a single sensor in the lower portion of the tower results in a small contribution to the total reward associated with the DRL fine-tuning process. This assessment also demonstrates how IL–DRL could achieve a better energy requirement. IL–DRL learns to emphasize more on the control of the parts where the vibration is more intense, which is the part close to the tower top; whereas the response reduction by IL-DRL is lower than that of the optimal controllers at lower heights. Notably, this outcome can be changed if the reward function is designed to give larger weights to the lower part responses.

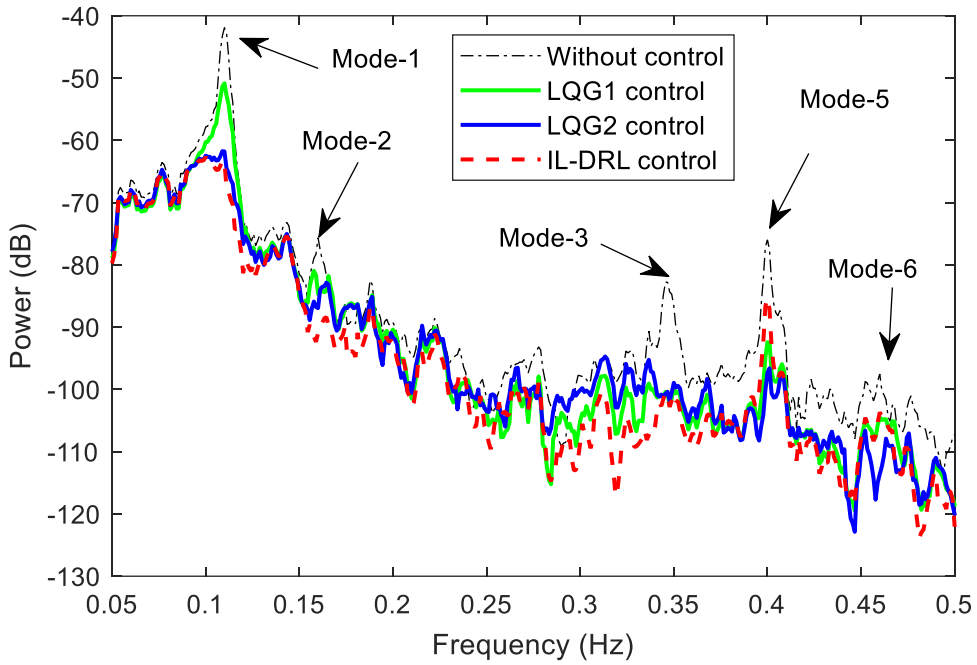


(a)

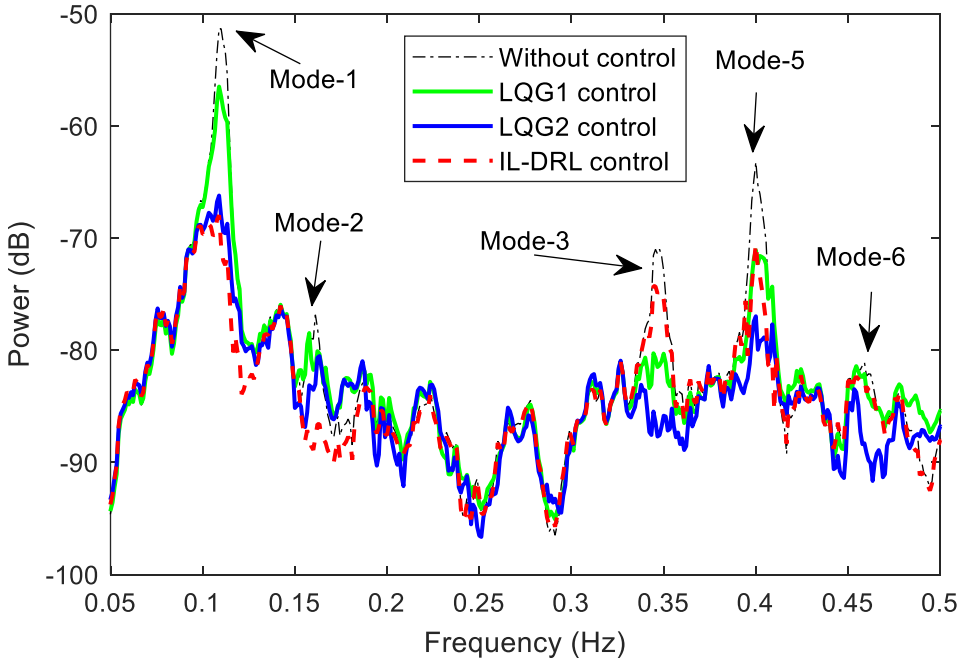


(b)

Figure 8.12 Max and RMS response level under wind excitation plot against height: (a) displacement; (b) acceleration



(a)



(b)

Figure 8.13 PSD of the main tower under wind excitation: (a) displacement at 438 m; (b) acceleration at 438 m

Figures 8.13a and 8.13b demonstrate the PSD of the displacement and acceleration responses at the top of the main tower under wind excitation, with different controllers in operation. The spectral analysis reveals that the higher modes contribute more to the acceleration responses than the displacements. For the prominent spectral peaks corresponding to fundamental bending modes (0.1–0.35 Hz), the IL–DRL controller achieves significant vibration suppression compared with the uncontrolled case. When benchmarked against model-based controllers, IL–DRL exhibits comparable performance to the optimal LQG2 controller, while outperforming LQG1 obviously, as proven in the previous analysis in the time domain.

The proposed IL-DRL controller demonstrates slightly reduced effectiveness in accelerations for higher modes (modes 3 and 5), though it maintains measurable vibration attenuation. Compared with the training condition (the free-vibration spectrum in Figure 8.6), the operational wind excitation amplifies higher mode participation, despite that the fundamental modes remain dominant. This finding suggests that the IL–DRL algorithm strategically prioritizes control effort for dominant vibration modes, while tolerating modest higher-mode responses to optimize energy efficiency, given that the reward function design emphasizes peak value control. By contrast, the model-based LQG controllers provide more uniform vibration reduction across all modes. A possible solution to enhance the control performance for higher modes is to use high-frequency excitation during the training process.

Notably, despite being trained exclusively on free vibration data, the IL–DRL controller maintains robust performance under wind-induced vibrations, demonstrating effective generalization capability. The comparative analysis underscores IL–DRL's balanced approach between performance and efficiency, particularly in managing the trade-off between fundamental mode control and higher-mode suppression.

8.6 Summary

This chapter utilizes the multi-objective vibration control to control the wind-induced vibration of the Canton Tower. The IL–DRL framework effectively addresses two major DRL implementation challenges by applying IL pre-training to significantly

reduce required training time and provide stable policy initialization that overcomes training instability. The key findings are as follows:

This model-free IL-DRL method demonstrates high feasibility. The DRL training is conducted using data from 50 episodes of free vibration tests. Compared to the training in the MDOF experiment presented in Chapter 7, the training in this chapter consists of 20 additional episodes. But this result still indicates the feasibility of transferring the proposed IL-DRL approach from laboratory tests to real-world control problems.

The proposed IL-DRL controller can enhance the performance of classical model-based controllers by formulating the reward function of DRL as a direct representation of multiple control objectives. The reward function design can consider both human comfort and structural integrity by directly using the peak and RMS response values at each time step in the reward function. The control force can also be added to constrain power demand.

The IL–DRL controller trained on free vibrations shows highly adaptive performance under unforeseen wind excitations. The IL–DRL controller achieves comparable peak vibration reduction to LQG controllers with 40% less control force. Compared with the LQG controller with a similar control energy demand, IL–DRL learns to distribute the control effort more on the upper half of the tower, where the vibration level is more intense, thus showing a better mitigation effect. Although only the vibration in the global X-direction is controlled in this chapter, the proposed control strategy can be easily extended to control vibrations in multiple directions.

While this chapter shows promising results of the proposed model-free control method on a supertall structure, the control performance may be influenced by the excitations used in the DRL training process. For simple structures, training based on random and free vibrations is generally sufficient to cover lower modes; whereas for more complex structures, they may not be adequate to excite all target vibration modes. To design excitations that can effectively and safely activate higher modes of interest in real structures will be an essential task in the training process.

In addition, in-service fine-tuning of the controller, which can continuously improve control performance during operation and adapt the controller to excitations

unseen during the training, will be highly preferable. A potential solution to this challenging task is to design a dimensionless reward, for example, a quadratic cost over a specified time window normalized by the excitation level, and then carry out continuous policy updates during service.

CHAPTER 9

Conclusions and Future Works

9.1 Summary

In the design of active vibration control strategies, the choice of control algorithm plays an important role. While the previous model-free methods were known for their simplicity, they often lacked the ability to achieve control performance as optimal as model-based controllers. The primary motivation of this thesis is to seek feasible and practical solutions to achieve optimal performance by using purely data-driven methods. The recent development of DRL and its success in the robotic control field make it a promising method for this task. This study numerically and experimentally investigated the possibility and effectiveness of the DRL applications in the active structural vibration control field. The following major tasks were completed:

(1) A DRL-based active control framework was proposed and applied to an SDOF structure and an MDOF shear building with different observation setups. The efficacy of the proposed method was demonstrated in comparison with the classical model-based optimal controllers under different vibrations.

(2) The proposed pure DRL framework was modified and extended to a nonlinear structure, namely, a buckled beam. The controlled dynamic behavior was carefully examined in comparison with analytical methods.

(3) To enhance the practical implementation ability of DRL, a pre-training stage was added to mimic simple control strategies and to avoid the uncertainty of random initiation of the DRL in real conditions, which may damage the structure.

(4) The proposed IL-DRL method was numerically validated on a full-scale stay cable, and its feasibility was demonstrated by showing optimal control performance and reduced training time.

(5) The proposed IL-DRL method was experimentally validated on an SDOF frame controlled by AMD. For the first time, data-driven methods could achieve performance comparable with that of model-based active controllers in an experimental setting.

(6) To leverage the flexibility of model-free methods, the design of a multi-objective reward was introduced. Performance indexes (e.g., inter-story drift) were directly added to the reward function structure. This improvement was demonstrated on an MDOF structure in a laboratory experiment.

(7) As an attempt towards real-world engineering deployment, the proposed multi-objective IL-DRL method was applied to a more complicated benchmark structure, namely, the Canton Tower, to control the wind-induced vibration.

9.2 Conclusions

In this thesis, the model-free active optimal control strategy based on DRL was first proposed and applied to linear and nonlinear systems. To address the inherent limitation of DRL and enhance its applicability in real-world scenarios, IL was combined to form the IL-DRL framework. This method was demonstrated through its application to vibration mitigation of a stay cable, and a shaking table experiment was conducted to prove the validity. By using the multi-objective reward function, the IL-DRL strategy successfully surpassed the performance of classical optimal controllers in the experiments on a 3DOF structure controlled by AMD and in the numerical investigation of the Canton Tower benchmark model. The major conclusions are summarized as follows:

(1) A model-free DRL approach is proposed for linear systems, which successfully trains the NN controllers using the DDPG algorithm under free vibrations. The performance is evaluated under various conditions using an SDOF model and an MDOF shear building model. Results indicate that the proposed model-free DRL controllers can achieve performance comparable with that of model-based LQR controllers. Notably, when only partial state feedback is available, the DRL controller outperforms the output feedback controllers with the same observation states and achieves performance close to full-state feedback LQR controllers despite using fewer states. These findings highlight the strong adaptability of the DRL approach in scenarios where accurate system models or complete state information are unavailable. By eliminating the dependency on detailed system model information, the DRL approach provides great potential as a practical alternative to model-based control for structural vibration.

(2) Building upon the model-free DRL for linear systems under free vibration, the TD3 algorithm is applied to active control of the nonlinear vibration of a simply-supported buckled beam with up to third-order nonlinearity. Comparative analyses show that the TD3-trained controller outperforms an analytical polynomial controller under static loadings and random excitations, achieving a nearly 20% increase in the escape load and improving the system safety margin. Notably, this performance gain is achieved without an increase in control force, which indicates higher energy efficiency and robustness of the DRL-based approach than conventional polynomial controllers. These findings not only demonstrate the generalizability of the proposed model-free DRL approach across both linear and nonlinear systems, but also confirm its robustness under a variety of loading conditions.

(3) The proposed model-free IL-DRL framework is applied to active vibration control of stay cables, reducing training time to less than 1/30 of that required by pure DRL methods, and enhancing its applicability in real-world structural control scenarios. Under random excitations, the IL-DRL controller achieves RMS displacement control comparable to that of a full-state model-based LQR feedback controller, despite operating with limited feedback state variables.

Additionally, the IL-DRL controller exhibits higher resistance against 5% measurement noise and 10% actuation force uncertainties compared with the classical LQG controller. This robustness is primarily attributed to the DRL fine-tuning process, which enables the controller to adapt to the actual structural dynamics encountered in practical applications.

(4) The IL-DRL method is further evaluated through shaking table experiments on an AMD-controlled SDOF structure. The incorporation of IL during the pre-training phase reduces the training time and stabilizes the learning process, which leads to efficient DRL training in the experiment without model knowledge. A series of comparative experiments validates the performance under different vibration scenarios. The IL-DRL controller achieves better peak floor displacement reduction than that of the model-based LQG controller by 16% under random excitation and by 9% under earthquake excitation, while maintaining similar control voltage levels. These experimental results validate the capability of the proposed IL-DRL approach to provide efficient and high-performance structural control in practical applications.

(5) Thanks to the flexibility of the IL-DRL approach, the framework can be further extended to consider multi-objective reward functions through different combinations of performance metrics. By directly incorporating peak inter-story drift and acceleration into the reward design, the IL-DRL controller can optimize multiple key performance indices of interest simultaneously, which enables the exploration of control strategies without restrictions on predefined cost functions, leading to more effective and flexible vibration control. In experiments on a 3DOF frame equipped with AMD under random excitation, the full-state IL-DRL controller reduces peak inter-story drift by 74% and RMS floor acceleration by 54% versus uncontrolled cases. Despite fine-tuning only under random excitations, it achieves 13% more RMS acceleration reduction and 11% lower peak control voltage than LQR under the EI Centro earthquake. The partial-state IL-DRL controller performs similarly with a 10% higher control voltage. Both controllers outperform LQG, reducing peak drift by 30% and RMS acceleration by 36.6%. These results validate the adaptability of multi-objective control functions and their capability to maintain superior

performance across different structural configurations, excitation types, and measurement availability conditions.

(6) When the multi-objective IL-DRL is applied to mitigate the wind-induced vibration of the Canton Tower, the results demonstrate that the IL-DRL controller achieves superior control performance compared to two LQG controllers with different control-energy balance setups. The reward function can be designed to directly consider the peak and RMS response values along with the control force at each time step, aiming to mitigate vibrations for structural integrity while maintaining human comfort and meeting power consumption requirements. Compared with the LQG controller with similar control performance, the peak control force of the IL-DRL is 40% less. Compared with that of the LQG controller with similar energy use, the IL-DRL adapts control efforts to focus more on the upper half of the tower, where vibrations are strongest, yielding an additional 16-38% improvement in vibration mitigation at the tower top. The method's flexibility allows adjusting reward functions to train the controller for varied and unforeseen excitations, confirming its robustness and practical applicability for large-scale structural vibration control.

(7) In conclusion, this thesis first proves the feasibility of achieving optimal control performance through model-free DRL methods in both linear and nonlinear systems. The practicability is improved by utilizing IL to pre-train an initial agent for DRL, which addresses the challenges of prolonged training time and instability issues commonly encountered in DRL methods. Furthermore, the control performance is improved by utilizing the flexibility of the DRL reward design to incorporate multi-objective control.

9.3 Future Works

This subchapter outlines the future directions arising from this thesis. The presented work has clearly demonstrated the potential of DRL in active structural vibration control. Specifically, the proposed IL-DRL framework lays the foundation to realize model-free optimal control for civil structures when dealing with active

vibration control problems. During this study, two major limitations inherent to current DRL methods were also identified, namely, the instability issue during training and the sample efficiency. Training instability poses significant safety concerns, especially when applied to real-world structures, because unstable control policies may lead to unsafe control actions. Meanwhile, sample inefficiency may restrict the method's ability to effectively generalize across complex structural dynamics, which typically require large volumes of training data.

Though these limitations have been partially addressed in this thesis by incorporating a pre-training stage and imposing constraints during training, the effectiveness and robustness of these strategies remain to be validated for more complex and uncertain structural systems.

Building on the current IL-DRL framework, future research shall include but not limited to:

(1) In Chapters 5–8, BC was employed to mimic simple controllers, such as negative stiffness dampers and position-velocity feedback controllers. Future research should investigate how the choice of reference control strategies impacts the DRL fine-tuning phase. It would be valuable to explore whether using more advanced or comprehensive reference controllers during pre-training can accelerate convergence and improve final policy quality. It is also worth examining transfer learning approaches, where knowledge from different controllers or structural configurations is adapted to new target structures to enhance efficiency and adaptability.

(2) The sensitivity of IL-DRL training to the initial learning rate has not been fully explored. Future studies could develop adaptive learning rate scheduling methods, such as performance-based decay or stability-driven adjustments, which can automatically tune the learning rate during training. This would help maintain stable policy updates, reduce the likelihood of divergence in early stages, and support faster, more robust convergence.

(3) To ensure safety during training, relying solely on designed hard constraints, as done in the current thesis, has certain limitations. Future research should explore safe RL methods (e.g., constrained policy optimization or Lyapunov-

based stability guarantees) that can enforce safety bounds throughout exploration without frequent manual intervention. Such methods will be critical for successful deployment on real-world structures where unsafe actions could cause devastating damage.

(4) The current fine-tuning stage relies entirely on online interaction with the target structure, meaning policy stability cannot be verified beforehand. A promising improvement would be to adopt a hybrid training framework that combines offline testing with subsequent online adaptation. Agents will be required to pass rigorous offline performance and safety evaluations before being deployed for online learning, thus reducing the risk of instability or failure during interaction with real structures.

(5) From a practical implementation perspective, the controller's ability to handle gradual or sudden changes in structural parameters during service has not yet been studied. A promising solution is to perform in-service tuning, in which the IL-DRL controller provides the baseline control performance, and in-service fine-tuning enables the agent to continuously learn and update the control policy online during operation, thus adapting to changes in structural dynamics over the service life. Future work will also consider sensor faults, multi-type sensor fusion (including vision-based sensing), sim-to-real transfer, and digital-twin integration.

(6) Finally, as structural models' complexity grows, real-time computation efficiency becomes a pressing challenge. Larger network architectures may improve control performance but also increase inference latency, making it harder to meet real-time requirements on embedded hardware. Future work should investigate network designs that balance complexity and efficiency, alongside techniques such as pruning, quantization, and lightweight model architectures, to ensure DRL controllers remain feasible for embedded, real-time, real-world applications. The effect of time delay should be systematically investigated.

The following promising directions are also worth further investigation:

(1) Generalizable and meta-RL for control

The development of a generalized control solution that can quickly optimize specific tasks is a brand-new, unexplored area in structural vibration control. This represents a step forward versus the current IL design, in which the pre-training structure is on the target structure. A more general control solution should be able to efficiently adapt to various control tasks and structural conditions, especially those that have never been encountered during training. To date, no relevant research has been conducted to address this direction within the field. The plan is to leverage the SOTA algorithms in meta-RL and unsupervised pre-training RL. However, because these methods are not designed for structural vibration control problems, many adjustments will be required. Therefore, additional experiments on various test structures will be necessary in future work. Moreover, large-scale experimental validation, such as hardware-in-the-loop tests and field deployments, will be essential to verify the reliability and scalability of DRL controllers in diverse and unpredictable vibration scenarios.

(2) Physics-informed machine learning (PIML) integration

Another promising direction is the integration of PIML. PIML aims to learn a physical model from data via optimization, which has potential applications in diverse areas, such as material design, fluid flow, robot dynamics, and digital twins. The PIML methods can embed physics in network design. Previous studies have shown that PIML can uncover coordinates and physical laws from data using methods like an autoencoder. An active vibration control problem can be written as follows:

$$m\ddot{x} + c\dot{x} + kx = Excitation + Control\ force \tag{9.1}$$

The PIML methods can be incorporated into two stages with DRL:

The first stage could use a physics-informed neural network (PINN) to determine m, c, k, given known excitation. PINN has been used for system identification tasks, such as equation discovery and parameter estimation. Moreover, it can be used to enhance learning methods to provide better estimation for sparse data. The trained PINN can be used in forward modeling as a model for simulation.

The second stage is to add physical law to the DRL training either by adding a physics-based loss term to the DRL loss function or by modifying the agent's NN architecture to ensure that its hidden states encode physically meaningful properties.

With these advancements, a unified and generalizable control strategy can be established and applied across different structures, which leads to a step forward in the field of structural vibration control.

(3) Hybrid control architectures

Combining DRL with classical model-based adaptive control methods can harness the advantages of both approaches. Hybrid frameworks offer improved reliability by providing fallback control options or composite signals, enhancing safety and performance in complex real-world applications.

(4) Long-term field trials and deployment

Before widespread real-world adoption, it is critical to assess the long-term performance of model-free DRL controllers through extensive experimental monitoring on actual civil structures. Such studies will reveal insights into stability, maintenance requirements, robustness under environmental variability, and the techno-economic viability of DRL-based vibration control under continuous operational conditions.

Collectively, these research frontiers, grounded in advances in meta-learning, physics-informed modeling, hybrid control strategies, and rigorous real-world testing, will pave the way for intelligent, scalable, robust, and broadly applicable active structural vibration control solutions and will accelerate the transition from promising laboratory research to practical field deployment in civil engineering.

REFERENCES

Abdel-Rohman, M., and Leipholz, H. H. (1978). "Active control of flexible structures." *Journal of the Structural Division*, 104(8), 1251-1266.

Abdel-Rohman, M., and Leipholz, H. H. (1979). "General approach to active structural control." *Journal of the Engineering Mechanics Division*, 105(6), 1007-1023.

Abdel-Rohman, M., and Leipholz, H. H. (1983). "Active control of tall buildings." *Journal of Structural Engineering*, 109(3), 628-645.

Abdel-Rohman, M., and Nayfeh, A. H. (1987). "Active control of nonlinear oscillations in bridges." *Journal of Engineering Mechanics*, 113(3), 335-348.

Abe, M. (1996). "Semi-active tuned mass dampers for seismic protection of civil structures." *Earthquake Engineering & Structural Dynamics*, 25(7), 743-749.

Adam, B., and Smith, I. F. (2008). "Reinforcement learning for structural control." *Journal of Computing in Civil Engineering*, 22(2), 133-139.

Adeli, H. (2001). "Neural networks in civil engineering: 1989–2000." *Computer-Aided Civil and Infrastructure Engineering*, 16(2), 126-142.

Agrawal, A., Yang, J., and Wu, J. (1998). "Non-linear control strategies for Duffing systems." *International Journal of Non-Linear Mechanics*, 33(5), 829-841.

Aldemir, U., Bakioglu, M., and Akhiev, S. (2001). "Optimal control of linear buildings under seismic excitations." *Earthquake Engineering & Structural Dynamics*, 30(6), 835-851.

Ali, M. M., and Moon, K. S. (2018). "Advances in structural systems for tall buildings: emerging developments for contemporary urban giants." *Buildings*, 8(8), 104.

Allen, M., Bernelli-Zazzera, F., and Scattolini, R. (2000). "Sliding mode control of a large flexible space structure." *Control Engineering Practice*, 8(8), 861-871.

Anderson, C. W., Lee, M., and Elliott, D. L. "Faster reinforcement learning after pretraining deep networks to predict state dynamics." *Proc., 2015 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 1-7.

Astolfi, A., Karagiannis, D., and Ortega, R. (2008). *Nonlinear and adaptive control with applications*, Springer.

Bani-Hani, K., and Ghaboussi, J. (1998). "Nonlinear structural control using neural networks." *J Eng Mech-Asce*, 124(3), 319-327.

Barth-Maron, G., Hoffman, M. W., Budden, D., Dabney, W., Horgan, D., Tb, D., Muldal, A., Heess, N., and Lillicrap, T. (2018). "Distributed distributional deterministic policy gradients." *arXiv preprint arXiv:1804.08617*.

Bellman, R. (1957). "A Markovian decision process." *Journal of Mathematics and Mechanics*, 679-684.

Bellman, R. (1958). " Dynamic programming and stochastic control processes." *Information and Control*(1(3)), 228-239.

Bertsekas, D. P., and Tsitsiklis, J. N. " Neuro-dynamic programming: an overview.
." *Proc., In Proceedings of 1995 34th IEEE conference on decision and control* IEEE.

Bryson, A. E., and Ho, Y.-C. (1975). *Applied optimal control: optimization, estimation, and control*, Routledge.

Cameron, T., Griffin, J., Kielb, R., and Hoosac, T. (1990). "An integrated approach for friction damper design." *Journal of Vibration and Acoustics*, 112(2), 175-182.

Chandwani, V., Agrawal, V., and Nagar, R. (2013). "Applications of soft computing in civil engineering: a review." *International Journal of Computer Applications*, 81(10).

Chang, J. C., and Soong, T. T. (1980). "Structural control using active tuned mass dampers." *Journal of the Engineering Mechanics Division*, 106(6), 1091-1098.

Chang, P.-M., Lou, J. Y., and Lutes, L. D. (1998). "Model identification and control of a tuned liquid damper." *Engineering Structures*, 20(3), 155-163.

Chen, C.-W., Yeh, K., Chiang, W.-L., Chen, C.-Y., and Wu, D.-J. (2007). "Modeling, H∞ control and stability analysis for structural systems using Takagi-Sugeno fuzzy model." *Journal of Vibration and Control*, 13(11), 1519-1534.

Chen, H. M., Tsai, K. H., Qi, G. Z., Yang, J. C. S., and Amini, F. (1995). "Neural-Network for Structure Control." *Journal of Computing in Civil Engineering*, 9(2), 168-176.

Chen, H.-P., Tee, K. F., and Ni, Y.-Q. (2012). "Mode shape expansion with consideration of analytical modelling errors and modal measurement uncertainty." *Smart Structures and Systems*, 10(4-5), 485-499.

Chen, Z., and Ni, Y. (2017). "Adaptive Semiactive Cable Vibration Control: A Frequency Domain Perspective." *Shock and Vibration*, 2017.

Cheng, F. Y., Jiang, H., and Lou, K. (2008). *Smart structures: innovative systems for seismic response control*, CRC press.

Cho, H. C., Fadali, M. S., Saiidi, M. S., and Lee, K. S. (2005). "Neural network active control of structures with earthquake excitation." *International Journal of Control, Automation, and Systems*, 3(2), 202-210.

Connor, J. J., and Klink, B. S. (1996). *Introduction to motion based design*, Computational Mechanics Publications.

Constantinou, M. C., Soong, T. T., and Dargush, G. F. (1998). *Passive energy dissipation systems for structural design and retrofit*, NCEER Monograph, National Center for Earthquake Engineering Research, Buffalo, NY.

Cruz Jr, G. V., Du, Y., and Taylor, M. E. (2017). "Pre-training neural networks with human demonstrations for deep reinforcement learning." *arXiv preprint arXiv:1709.04083*.

CTBUH (2023). "CTBUH Year in Review: Tall Trends of 2022." *CTBUH Journal*(1), 44-45.

Datta, T. (2003). "A state-of-the-art review on active control of structures." *ISET Journal of Earthquake Technology*, 40(1), 1-17.

Den Hartog, J. P. (1985). *Mechanical vibrations*, Courier Corporation.

Ding, Q., and Zhu, L. (2017). "Aeroelastic model test of a 610 m-high TV tower with complex shape and structure." *Wind & structures*, 25(4), 361-379.

Duan, Y., Chen, X., Houthooft, R., Schulman, J., and Abbeel, P. "Benchmarking deep reinforcement learning for continuous control." *Proc., International Conference on Machine Learning*, 1329-1338.

Dyke, S., Spencer Jr, B., Quast, P., Sain, M., Kaspari Jr, D., and Soong, T. (1996). "Acceleration feedback control of MDOF structures." *Journal of Engineering Mechanics*, 122(9), 907-918.

Eshkevari, S. S., Eshkevari, S. S., Sen, D., and Pakzad, S. N. (2023). "Active structural control framework using policy-gradient reinforcement learning." *Engineering Structures*, 274, 115122.

Feinberg, V., Wan, A., Stoica, I., Jordan, M. I., Gonzalez, J. E., and Levine, S. "Model-based value expansion for efficient model-free reinforcement learning." *Proc., Proceedings of the 35th International Conference on Machine Learning (ICML 2018)*.

Fisco, N., and Adeli, H. (2011). "Smart structures: part I—active and semi-active

control." *Scientia Iranica*, 18(3), 275-284.

Fujimoto, S., Hoof, H., and Meger, D. "Addressing function approximation error in actor-critic methods." *Proc., International Conference on Machine Learning*, PMLR, 1587-1596.

Fujino, Y., Soong, T., and Spencer, B. (1996). "Structural control: Basic concepts and applications."

Fujino, Y., Sun, L., Pacheco, B. M., and Chaiseri, P. (1992). "Tuned liquid damper (TLD) for suppressing horizontal motion of structures." *Journal of Engineering Mechanics*, 118(10), 2017-2030.

Ghaboussi, J., and Joghataie, A. (1995). "Active Control of Structures Using Neural Networks." *J Eng Mech-Asce*, 121(4), 555-567.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*, MIT press.

Grondman, I., Busoniu, L., Lopes, G. A., and Babuska, R. (2012). "A survey of actor-critic reinforcement learning: Standard and natural policy gradients." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6), 1291-1307.

Guclu, R., and Yazici, H. (2008). "Vibration control of a structure with ATMD against earthquake using fuzzy logic controllers." *Journal of Sound and Vibration*, 318(1-2), 36-49.

Gutierrez Soto, M., and Adeli, H. (2013). "Tuned mass dampers." *Archives of Computational Methods in Engineering*, 20, 419-431.

Hameed, I., Chao, X., Navarro-Alarcon, D., and Jing, X. (2025). "Deep reinforcement learning enabling a BCFbot to learn various undulatory patterns. " *Ocean Engineering*, 320, 120322.

Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S. "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor." *Proc., International Conference on Machine Learning*, PMLR, 1861-1870.

Hester, T., Vecerik, M., Pietquin, O., Lanctot, M., Schaul, T., Piot, B., Horgan, D., Quan, J., Sendonaris, A., and Osband, I. "Deep q-learning from demonstrations." *Proc., Thirty-second AAAI Conference on Artificial Intelligence*.

Hikami, Y., and Shiraishi, N. (1988). "Rain-wind induced vibrations of cables stayed bridges." *Journal of Wind Engineering and Industrial Aerodynamics*, 29(1-3), 409-418.

Housner, G., Bergman, L. A., Caughey, T. K., Chassiakos, A. G., Claus, R. O., Masri,

S. F., Skelton, R. E., Soong, T., Spencer, B., and Yao, J. T. (1997). "Structural control: past, present, and future." *Journal of Engineering Mechanics*, 123(9), 897-971.

Howard, R. A. (1960). *Dynamic programming and markov processes*, John Wiley.

Hrovat, D., Barak, P., and Rabins, M. (1983). "Semi-active versus passive or active tuned mass dampers for structural control." *Journal of Engineering Mechanics*, 109(3), 691-705.

Hung, S.-L., Kao, C., and Lee, J. (2000). "Active pulse structural control using artificial neural networks." *Journal of Engineering Mechanics*, 126(8), 839-849.

Iemura, H., and Pradono, M. H. (2005). "Simple algorithm for semi-active seismic response control of cable-stayed bridges." *Earthquake Engineering & Structural Dynamics*, 34(4-5), 409-423.

Igusa, T., and Xu, K. (1994). "Vibration control using multiple tuned mass dampers." *Journal of Sound and Vibration*, 175(4), 491-503.

Inoue, M., Siringoringo, D. M., Fujino, Y., and Koike, Y. (2025). "Identification of vortex-induced vibration on the osman gazi suspension bridge tower and mitigation by an active mass damper." *Journal of Bridge Engineering*, 30(2), 04024108.

Iqbal, J., Ullah, M., Khan, S. G., Khelifa, B., and Ćuković, S. (2017). "Nonlinear control systems-A brief overview of historical and recent advances." *Nonlinear Engineering*, 6(4), 301-312.

Jansen, E. (2008). "A perturbation method for nonlinear vibrations of imperfect structures: application to cylindrical shell vibrations." *International Journal of Solids and Structures*, 45(3-4), 1124-1145.

Jansen, L. M., and Dyke, S. J. (2000). "Semiactive control strategies for MR dampers: comparative study." *Journal of Engineering Mechanics*, 126(8), 795-803.

Johnson, E. A., Baker, G. A., Spencer Jr, B., and Fujino, Y. (2007). "Semiactive damping of stay cables." *Journal of Engineering Mechanics*, 133(1), 1-11.

Juan, S. H., and Tur, J. M. M. (2008). "Tensegrity frameworks: Static analysis review." *Mechanism and Machine Theory*, 43(7), 859-881.

Kareem, A., and Kline, S. (1995). "Performance of multiple mass dampers under random loading." *Journal of Structural Engineering*, 121(2), 348-361.

Kawashima, K., Unjoh, S., and Shimizu, H. "Seismic response control of highway bridges by variable damper." *Proc., Wind and Seismic Effects: Proceedings of the 24th Joint Meeting of the US-Japan Cooperative Program in Natural Resources Panel on Wind and Seismic Effects*, US Department of Commerce, National Institute

of Standards and Technology, 71.

Kaynia, A. M., Veneziano, D., and Biggs, J. M. (1981). "Seismic effectiveness of tuned mass dampers." *Journal of the Structural Division*, 107(8), 1465-1484.

Kelly, J. M. (1993). *Earthquake-resistant design with rubber*, Springer.

Khalatbarisoltani, A., Soleymani, M., and Khodadadi, M. (2019). "Online control of an active seismic system via reinforcement learning." *Structural Control and Health Monitoring*, 26(3), e2298.

Kim, D. H., Kim, D., Chang, S., and Jung, H.-Y. (2008). "Active control strategy of structures based on lattice type probabilistic neural network." *Probabilistic Engineering Mechanics*, 23(1), 45-50.

Kim, H., and Adeli, H. (2005). "Wind-induced motion control of 76-story benchmark building using the hybrid damper-TLCD system." *Journal of Structural Engineering*, 131(12), 1794-1802.

Kobori, T., Koshika, N., Yamada, K., and Ikeda, Y. (1991). "Seismic-response-controlled structure with active mass driver system. Part 1: Design." *Earthquake Engineering & Structural Dynamics*, 20(2), 133-149.

Kobori, T., Takahashi, M., Nasu, T., Niwa, N., and Ogasawara, K. (1993). "Seismic response controlled structure with active variable stiffness system." *Earthquake Engineering & Structural Dynamics*, 22(11), 925-941.

Konda, V. R., and Borkar, V. S. (1999). "Actor-critic--type learning algorithms for markov decision processes." *SIAM Journal on Control and Optimization*, 38(1), 94-123.

Korkmaz, S. (2011). "A review of active structural control: challenges for engineering informatics." *Computers & Structures*, 89(23-24), 2113-2132.

Krenk, S. (2000). "Vibrations of a taut cable with an external damper." *J. Appl. Mech.*, 67(4), 772-776.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). "Deep learning." *Nature*, 521(7553), 436-444.

Lee, D., and Taylor, D. P. (2001). "Viscous damper development and future trends." *The Structural Design of Tall Buildings*, 10(5), 311-320.

Lewis, F. L., Vrabie, D., and Syrmos, V. L. (2012). *Optimal control*, John Wiley & Sons.

Li, H., Liu, M., and Ou, J. (2008). "Negative stiffness characteristics of active and semi-active control systems for stay cables." *Structural Control and Health Monitoring: The Official Journal of the International Association for Structural Control and*

Monitoring and of the European Association for the Control of Structures, 15(2), 120-142.

Li, H.-N., and Li, G. (2007). "Experimental study of structure with “dual function” metallic dampers." *Engineering Structures*, 29(8), 1917-1928.

Li, H., Jing, X., and Karimi, H. R. (2013). "Output-feedback-based H∞ control for vehicle suspension systems with control delay. " *IEEE Transactions on Industrial Electronics*, 61(1), 436-446.

Li, J.-Y., Zhu, S., Shi, X., and Shen, W. (2020). "Electromagnetic shunt damper for bridge cable vibration mitigation: Full-scale experimental study." *Journal of Structural Engineering*, 146(1), 04019175.

Li, W., and Todorov, E. "Iterative linear quadratic regulator design for nonlinear biological movement systems." *Proc., ICINCO (1)*, Citeseer, 222-229.

Li, Y. (2017). "Deep reinforcement learning: An overview." *arXiv preprint arXiv:1701.07274*.

Li, Y., Shen, W., and Zhu, H. (2019). "Vibration mitigation of stay cables using electromagnetic inertial mass dampers: Full-scale experiment and analysis." *Engineering Structures*, 200, 109693.

Li, Z., and Adeli, H. (2018). "Control methodologies for vibration control of smart civil and mechanical structures." *Expert Systems*, 35(6), e12354.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. (2015). "Continuous control with deep reinforcement learning." *arXiv preprint arXiv:1509.02971*.

Lin, R., Liang, Z., Soong, T., Zhang, R., and Mahmoodi, P. (1991). "An experimental study on seismic behaviour of viscoelastically damped structures." *Engineering Structures*, 13(1), 75-84.

Lin, W., Song, G., and Chen, S. (2017). "PTMD control on a benchmark TV tower under earthquake and wind load excitations." *Applied Sciences*, 7(4), 425.

Liu, C., Chen, L., Yang, X., Zhang, X., and Yang, Y. (2019). "General theory of skyhook control and its application to semi-active suspension control strategy design." *IEEE Access*, 7, 101552-101560.

Liu, J., Xia, K., and Zhu, C. "Structural vibration intelligent control based on magnetorheological damper." *Proc., 2009 International Conference on Computational Intelligence and Software Engineering*, IEEE, 1-4.

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., and Mordatch, I. (2017). "Multi-agent

actor-critic for mixed cooperative-competitive environments." *arXiv preprint arXiv:1706.02275*.

Lü, Q., Wang, X., Lu, K., and Huan, R. (2020). "Nonlinear Stochastic Optimal Control Using Piezoelectric Stack Inertial Actuator." *Shock and Vibration*, 2020.

Ma, H., Tang, G.-Y., and Zhao, Y.-D. (2006). "Feedforward and feedback optimal control for offshore structures subjected to irregular wave forces." *Ocean Engineering*, 33(8-9), 1105-1117.

Madan, A. (2005). "Vibration control of building structures using self-organizing and self-learning neural networks." *Journal of Sound and Vibration*, 287(4-5), 759-784.

Magaña, M., and Rodellar, J. (1998). "Nonlinear decentralized active tendon control of cable-stayed bridges." *Journal of Structural Control*, 5(1), 45-62.

Mahmoudi, R., Ghaffarzadeh, H., Ahani, E., and Katebi, J. (2019). "Sliding mode control of linear structures and a Duffing system using active tendons." *Proceedings of the Institution of Civil Engineers-Engineering and Computational Mechanics*, 172(3), 106-117.

Main, J., and Jones, N. (2002). "Free vibrations of taut cable with attached damper. I: Linear viscous damper." *Journal of Engineering Mechanics*, 128(10), 1062-1071.

Matsumoto, M., Daito, Y., Kanamura, T., Shigemura, Y., Sakuma, S., and Ishizaki, H. (1998). "Wind-induced vibration of cables of cable-stayed bridges." *Journal of Wind Engineering and Industrial Aerodynamics*, 74, 1015-1027.

McMahon, S., and Makris, N. "Large-scale ER-damper for seismic protection." *Proc., Smart Structures and Materials 1997: Passive Damping and Isolation*, SPIE, 140-147.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., and Kavukcuoglu, K. "Asynchronous methods for deep reinforcement learning." *Proc., International Conference on Machine Learning*, PMLR, 1928-1937.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. (2013). "Playing atari with deep reinforcement learning." *arXiv preprint arXiv:1312.5602*.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., and Ostrovski, G. (2015). "Human-level control through deep reinforcement learning." *Nature*, 518(7540), 529-533.

Modares, H., and Lewis, F. L. (2014). "Linear Quadratic Tracking Control of Partially-Unknown Continuous-Time Systems Using Reinforcement Learning." *Ieee*

Transactions on Automatic Control, 59(11), 3051-3056.

Moreschi, L., and Singh, M. (2003). "Design of yielding metallic and friction dampers for optimal seismic performance." *Earthquake Engineering & Structural Dynamics*, 32(8), 1291-1311.

Mualla, I., and Belev, B. (2015). "Analysis, design and applications of rotational friction dampers for seismic protection." *Czasopismo Inżynierii Lądowej, Środowiska i Architektury*, 62(4), 335-346.

Mysore, S., Mabsout, B., Mancuso, R., and Saenko, K. (2021). "Honey, I Shrunk The Actor: A Case Study on Preserving Performance with Smaller Actors in Actor-Critic RL." *arXiv preprint arXiv:2102.11893*.

Nagabandi, A. (2020). "Model-based Deep Reinforcement Learning for Robotic Systems."

Nagabandi, A., Kahn, G., Fearing, R. S., and Levine, S. "Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning." *Proc., 2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 7559-7566.

Nagendra, S., Podila, N., Ugarakhod, R., and George, K. "Comparison of reinforcement learning algorithms applied to the cart-pole problem." *Proc., 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE, 26-32.

Nayfeh, A. H., Mook, D. T., and Holmes, P. (1980). "Nonlinear oscillations."

Nerves, A., and Krishnan, R. "Active control strategies for tall civil structures." *Proc., Proceedings of IECON'95-21st Annual Conference on IEEE Industrial Electronics*, IEEE, 962-967.

Ni, Y., Wong, K., and Xia, Y. (2011). "Health checks through landmark bridges to sky-high structures." *Advances in Structural Engineering*, 14(1), 103-119.

Ni, Y., Xia, Y., Lin, W., Chen, W., and Ko, J. (2012). "SHM benchmark for high-rise structures: a reduced-order finite element model and field measurement data." *Smart Structures and Systems*, 10(4-5), 411-426.

Nikzad, K., Ghaboussi, J., and Paul, S. L. (1996). "Actuator dynamics and delay compensation using neurocontrollers." *Journal of Engineering Mechanics*, 122(10), 966-975.

Ogata, K. (2010). *Modern control engineering*, Prentice hall Upper Saddle River, NJ.

Pacheco, B. M., Fujino, Y., and Sulekh, A. (1993). "Estimation curve for modal

damping in stay cables with viscous damper." *Journal of Structural Engineering*, 119(6), 1961-1979.

Panda, J., Chopra, M., Matsagar, V., and Chakraborty, S. (2024). "An iterative gradient descent-based reinforcement learning policy for active control of structural vibrations." *Computers & Structures*, 290, 107183.

Pantelides, C., and Nelson, P. (1995). "Continuous pulse control of nonlinear structures." *Computers & Structures*, 55(6), 997-1006.

Park, W., Park, K.-S., and Koh, H.-M. (2008). "Active control of large structures using a bilinear pole-shifting transform with H∞ control method." *Engineering Structures*, 30(11), 3336-3344.

Pinto, O., and Gonçalves, P. (2000). "Non-linear control of buckled beams under step loading." *Mechanical Systems and Signal Processing*, 14(6), 967-985.

Pinto, O., and Gonçalves, P. (2002). "Active non-linear control of buckling and vibrations of a flexible buckled beam." *Chaos, Solitons & Fractals*, 14(2), 227-239.

Pisarski, D., and Jankowski, Ł. (2023). "Reinforcement learning ‐ based control to suppress the transient vibration of semi ‐ active structures subjected to unknown harmonic excitation." *Computer-Aided Civil and Infrastructure Engineering*, 38(12), 1605-1621.

Preumont, A., Achkire, Y., and Bossens, F. (2000). "Active tendon control of large trusses." *AIAA Journal*, 38(3), 493-498.

Puterman, M. L. (1994). *Markov decision processes: discrete stochastic dynamic programming.*, John Wiley & Sons.

Rahmani, H. R., Chase, G., Wiering, M., and Konkea, C. (2019). "A framework for brain learning-based control of smart structures." *Advanced Engineering Informatics*, 42, 100986.

Rao, M., and Datta, T. (2006). "Modal seismic control of building frames by artificial neural network." *Journal of Computing in civil Engineering*, 20(1), 69-73.

Rhode-Barbarigos, L., Ali, N. B. H., Motro, R., and Smith, I. F. (2010). "Designing tensegrity modules for pedestrian bridges." *Engineering Structures*, 32(4), 1158-1167.

Ross, S., Gordon, G., and Bagnell, D. "A reduction of imitation learning and structured prediction to no-regret online learning." *Proc., Proceedings of the fourteenth international conference on artificial intelligence and statistics*, JMLR Workshop

and Conference Proceedings, 627-635.

Rummery, G. A., and Niranjan, M. (1994). *On-line Q-learning using connectionist systems*, University of Cambridge, Department of Engineering Cambridge, UK.

Saaed, T. E., Nikolakopoulos, G., Jonasson, J.-E., and Hedlund, H. (2015). "A state-of-the-art review of structural control systems." *Journal of Vibration and Control*, 21(5), 919-937.

Salehi, H., and Burgueño, R. (2018). "Emerging artificial intelligence methods in structural engineering." *Engineering Structures*, 171, 170-189.

Samali, B., and Kwok, K. (1995). "Use of viscoelastic dampers in reducing wind-and earthquake-induced motion of building structures." *Engineering Structures*, 17(9), 639-654.

Samali, B., Yang, J., and Yeh, C. (1985). "Control of lateral-torsional motion of wind-excited buildings." *Journal of Engineering Mechanics*, 111(6), 777-796.

Sarbjeet, S., and Datta, T. (1998). "Open-closed-loop linear control of building frames under seismic excitation." *Journal of Structural Engineering*, 124(1), 43-51.

Schulman, J., Levine, S., Abbeel, P., Jordan, M., and Moritz, P. "Trust region policy optimization." *Proc., International Conference on Machine Learning*, 1889-1897.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). "Proximal policy optimization algorithms." *arXiv preprint arXiv:1707.06347*.

Shefer, M., and Breakwell, J. (1987). "Estimation and control with cubic nonlinearities." *Journal of Optimization Theory and Applications*, 53(1), 1-7.

Shen, W., and Zhu, S. (2015). "Harvesting energy via electromagnetic damper: Application to bridge stay cables." *Journal of Intelligent Material Systems and Structures*, 26(1), 3-19.

Shi, X., Zhao, F., Yan, Z., Zhu, S., and Li, J. Y. (2021). "High-performance vibration isolation technique using passive negative stiffness and semiactive damping." *Computer-Aided Civil and Infrastructure Engineering*, 36(8), 1034-1055.

Shi, X., Zhu, S., Li, J. Y., and Spencer, B. F. (2016). "Dynamic behavior of stay cables with passive negative stiffness dampers." *Smart Materials and Structures*, 25(7), 075044.

Shi, X., Zhu, S., and Nagarajaiah, S. (2017). "Performance comparison between passive negative-stiffness dampers and active control in cable vibration mitigation." *Journal of Bridge Engineering*, 22(9), 04017054.

Siddique, N., and Adeli, H. (2013). *Computational intelligence: synergies of fuzzy logic,*

neural networks and evolutionary computing, John Wiley & Sons.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., and Lanctot, M. (2016). "Mastering the game of Go with deep neural networks and tree search." *Nature*, 529(7587), 484-489.

Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., and Riedmiller, M. "Deterministic policy gradient algorithms."

Skinner, R. I., Robinson, W. H., and McVerry, G. H. (1993). *An introduction to seismic isolation*, Wiley, Chichester.

Soliman, M., and Ray, W. (1972). "Optimal feedback control for linear-quadratic systems having time delays." *International Journal of Control*, 15(4), 609-627.

Soong, T. T., and Costantinou, M. C. (2014). *Passive and active structural vibration control in civil engineering*, Springer.

Soong, T. T., and Manolis, G. D. (1987). "Active structures." *Journal of Structural Engineering*, 113(11), 2290-2302.

Spencer, B., Dyke, S., Sain, M., and Quast, P. "Acceleration feedback control strategies for aseismic protection." *Proc., 1993 American Control Conference*, IEEE, 1317-1321.

Spencer Jr, B., Dyke, S., Sain, M., and Carlson, J. (1997). "Phenomenological model for magnetorheological dampers." *Journal of Engineering Mechanics*, 123(3), 230-238.

Spencer Jr, B., and Nagarajaiah, S. (2003). "State of the art of structural control." *Journal of Structural Engineering*, 129(7), 845-856.

Subramanian, K., Isbell Jr, C. L., and Thomaz, A. L. "Exploration from demonstration for interactive reinforcement learning." *Proc., Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, 447-456.

Suhardjo, J., Spencer Jr, B., and Sain, M. (1992). "Non-linear optimal control of a Duffing system." *International Journal of Non-linear Mechanics*, 27(2), 157-172.

Sun, L., Fujino, Y., Pacheco, B., and Chaiseri, P. (1992). "Modelling of tuned liquid damper (TLD)." *Journal of Wind Engineering and Industrial Aerodynamics*, 43(1-3), 1883-1894.

Sutton, R. S. (1988). " Learning to predict by the methods of temporal differences." *Machine Learning*, 3, 9-44.

Sutton, R. S., and Barto, A. G. (2018). *Reinforcement learning: An introduction*, MIT

press.

Symans, M., Charney, F., Whittaker, A., Constantinou, M., Kircher, C., Johnson, M., and McNamara, R. (2008). "Energy dissipation systems for seismic applications: current practice and recent developments." *Journal of Structural Engineering*, 134(1), 3-21.

Symans, M. D., and Constantinou, M. C. (1999). "Semi-active control systems for seismic protection of structures: a state-of-the-art review." *Engineering Structures*, 21(6), 469-487.

Tan, P., Liu, Y., Zhou, F., and Huan, S. "Application and Testing of Hybrid Mass Dampers for Vibration Control of Canton Tower." *Proc., EACS 2016 – 6th European Conference on Structural Control*.

Tang, X., Zuo, L., Lin, T., and Zhang, P. "Improved design of linear electromagnetic transducers for large-scale vibration energy harvesting." *Proc., Active and Passive Smart Structures and Integrated Systems 2011*, SPIE, 242-252.

Tang, Y. (1996). "Active control of SDF systems using artificial neural networks." *Computers & Structures*, 60(5), 695-703.

Tang, Y. (1996). "New algorithm for active structural control." *Journal of Structural Engineering*, 122(9), 1081-1088.

Tesauro, G. (1994). "TD-Gammon, a self-teaching backgammon program, achieves master-level play." *Neural Computation*, 6(2), 215-219.

Thenozhi, S., and Yu, W. (2013). "Advances in modeling and vibration control of building structures." *Annual Reviews in Control*, 37(2), 346-364.

Tibert, G., Skelton, R., de Oliveura, M., Micheletti, A., Pandia Raj, R., Guest, S. D., Shigematsu, M., Tanaka, M., Noguchi, H., and Maurin, B. (2008). *Advances in the Optimization and Form-finding of Tensegrity Structures*.

Tomasula, D., Spencer Jr, B., and Sain, M. (1996). "Nonlinear control strategies for limiting dynamic response extremes." *Journal of Engineering Mechanics*, 122(3), 218-229.

Tsai, H. C., and Lin, G. C. (1993). "Optimum tuned-mass dampers for minimizing steady-state response of support-excited and damped systems." *Earthquake Engineering & Structural Dynamics*, 22(11), 957-973.

Tuan, A. Y., and Shang, G. (2014). "Vibration control in a 101-storey building using a tuned mass damper." *Journal of Applied Science and Engineering*, 17(2), 141-156.

Utkin, V. I. (2013). *Sliding modes in control and optimization*, Springer Science &

Business Media.

Vrabie, D., Pastravanu, O., Abu-Khalaf, M., and Lewis, F. L. (2009). "Adaptive optimal control for continuous-time linear systems based on policy iteration." *Automatica*, 45(2), 477-484.

Wani, Z. R., Tantray, M., Farsangi, E. N., Nikitas, N., Noori, M., Samali, B., and Yang, T. (2022). "A critical review on control strategies for structural vibration control." *Annual Reviews in Control*, 54, 103-124.

Warnitchai, P., Fujino, Y., Pacheco, B. M., and Agret, R. (1993). "An experimental study on active tendon control of cable-stayed bridges." *Earthquake Engineering & Structural Dynamics*, 22(2), 93-111.

Waszczyszyn, Z. (2010). *Advances of soft computing in engineering*, Springer Science & Business Media.

Watkins, C. J., and Dayan, P. (1992). "Q-learning." *Machine Learning*, 8(3-4), 279-292.

Weber, F., and Boston, C. (2011). "Clipped viscous damping with negative stiffness for semi-active cable damping." *Smart Materials and Structures*, 20(4), 045007.

Weber, T., Racanière, S., Reichert, D. P., Buesing, L., Guez, A., Rezende, D. J., Badia, A. P., Vinyals, O., Heess, N., and Li, Y. (2017). "Imagination-augmented agents for deep reinforcement learning." *arXiv preprint arXiv:1707.06203*.

Wen, Y.-K., Ghaboussi, J., Venini, P., and Nikzad, K. (1995). "Control of structures using neural networks." *Smart Materials and Structures*, 4(1A), A149.

Werbos, P. (1974). "New tools for prediction and analysis in the behavioral sciences." *Ph. D. Dissertation, Harvard University*.

Wu, Z., Lin, R., and Soong, T. (1995). "Non-linear feedback control for improved peak response reduction." *Smart Materials and Structures*, 4(1A), A140.

Xie, Y. Z., Sichani, M. E., Padgett, J. E., and DesRoches, R. (2020). "The promise of implementing machine learning in earthquake engineering: A state-of-the-art review." *Earthquake Spectra*, 36(4), 1769-1801.

Xing, L., Wen, C., Liu, Z., Su, H., and Cai, J. (2016). "Event-triggered adaptive control for a class of uncertain nonlinear systems." *IEEE Transactions on Automatic Control*, 62(4), 2071-2076.

Xu, H.-b., Ou, J.-p., Zhang, C.-w., Li, H., Tan, P., and Zhou, F.-l. (2014). "Active mass driver control system for suppressing wind-induced vibration of the Canton Tower." *Smart Structures and Systems, An International Journal*, 13(2), 281-303.

Xu, Y. L., Qu, W., and Chen, Z. (2001). "Control of wind-excited truss tower using

semiactive friction damper." *Journal of Structural Engineering*, 127(8), 861-868.

Xu, Y. L., and Yu, Z. (1998). "Mitigation of three-dimensional vibration of inclined sag cable using disrete oil dampers—II. Application." *Journal of Sound and Vibration*, 214(4), 675-693.

Yamaguchi, H. (1990). "Analytical study on growth mechanism of rain vibration of cables." *Journal of Wind Engineering and Industrial Aerodynamics*, 33(1-2), 73-80.

Yamamoto, M., Higashino, M., Aizawa, S., and Toyama, K. "Application of active mass damper (AMD) system, and earthquake and wind observation results." *Proc., Proceedings of the 2nd World Conference on Structural Control*, 783-794.

Yang, J., Wu, J., and Li, Z. (1996). "Control of seismic-excited buildings using active variable stiffness systems." *Engineering Structures*, 18(8), 589-596.

Yang, J.-N. (1975). "Application of optimal control theory to civil engineering structures." *Journal of the Engineering Mechanics Division*, 101(6), 819-838.

Yang, J. N., Li, Z., and Liu, S. (1991). "Instantaneous optimal control with acceleration and velocity feedback." *Stochastic Structural Dynamics 2*, Springer, 287-306.

Yang, T., Zhou, S., Litak, G., and Jing, X. (2023). "Recent advances in correlation and integration between vibration control, energy harvesting and monitoring." *Nonlinear Dynamics*, 111(22), 20525-20562.

Yao, H., Tan, P., Yang, T., and Zhou, F. (2024). "Deep reinforcement learning-based active mass driver decoupled control framework considering control–structure interaction effects." *Computer-Aided Civil and Infrastructure Engineering*.

Yao, J. T. (1972). "Concept of structural control." *Journal of the Structural Division*, 98(7), 1567-1574.

Yen, G. (1996). "Distributive vibration control in flexible multibody dynamics." *Computers & Structures*, 61(5), 957-965.

Ying, Z., Ni, Y., and Ko, J. (2003). "Stochastic optimal coupling-control of adjacent building structures." *Computers & Structures*, 81(30-31), 2775-2787.

Zhang, J., and Li, Q. (2019). "Identification of modal parameters of a 600-m-high skyscraper from field vibration tests." *Earthquake Engineering & Structural Dynamics*, 48(15), 1678-1698.

Zhang, M., and Jing, X. (2021). "Adaptive neural network tracking control for double-pendulum tower crane systems with nonideal inputs." *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *52*(4), 2514-2530.

Zhang, M., Jing, X., Huang, W., and Li, P. (2022). "Saturated PD-SMC method for

suspension systems by exploiting beneficial nonlinearities for improved vibration reduction and energy-saving performance." *Mechanical Systems and Signal Processing*, 179, 109376.

Zhao, B., Lu, X., Wu, M., and Mei, Z. (2000). "Sliding mode control of buildings with base-isolation hybrid protective system." *Earthquake Engineering & Structural Dynamics*, 29(3), 315-326.

Zhou, K., Zhang, J.-W., and Li, Q.-S. (2022). "Control performance of active tuned mass damper for mitigating wind-induced vibrations of a 600-m-tall skyscraper." *Journal of Building Engineering*, 45, 103646.

Zhu, L. M. M., Modares, H., Peen, G. O., Lewis, F. L., and Yue, B. Z. (2015). "Adaptive Suboptimal Output-Feedback Control for Linear Systems Using Integral Reinforcement Learning." *Ieee Transactions on Control Systems Technology*, 23(1), 264-273.

Ziegler, F. (2005). "Computational aspects of structural shape control." *Computers & Structures*, 83(15-16), 1191-1204.

Zuo, L., and Tang, X. (2013). "Large-scale vibration energy harvesting." *Journal of Intelligent Material Systems and Structures*, 24(11), 1405-1430.