Identification and Control of Nonlinear Singularly Perturbed Systems Using Multi-time-scale Neural Networks

Dongdong Zheng

A Thesis

In the Department

of

Mechanical & Industrial Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy (Mechanical Engineering) at

Concordia University

Montréal, Québec, Canada

January 2017

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Mr. Dongdong Zheng**

Entitled: **Identification and Control of Nonlinear Singularly Perturbed Systems Using Multi-time-scale Neural Networks**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Mechanical Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Rajagopalan Jayakumar

_____ External Examiner
Dr. Yajun Pan

_____ External to Program
Dr. Nizar Bouguila

_____ Examiner
Dr. Youmin Zhang

_____ Examiner
Dr. Henry Hong

_____ Supervisor
Dr. Wen-Fang Xie

Approved by _____
Martin D. Pugh, Chair
Department of Mechanical & Industrial Engineering

_____ 2017 _____
Amir Asif, Dean
Faculty of Engineering and Computer Science

Abstract

Identification and Control of Nonlinear Singularly Perturbed Systems Using Multi-time-scale Neural Networks

Dongdong Zheng, Ph.D.

Concordia University, 2017

Many industrial systems are nonlinear with "slow" and "fast" dynamics because of the presence of some "parasitic" parameters such as small time constants, resistances, inductances, capacitances, masses and moments of inertia. These systems are usually labeled as "singularly perturbed" or "multi-time-scale" systems. Singular perturbation theory has been proved to be a useful tool to control and analyze singularly perturbed systems if the full knowledge of the system model parameters is available. However, the accurate and faithful mathematical models of those systems are usually difficult to obtain due to the uncertainties and nonlinearities.

To obtain the accurate system models, in this research, a new identification scheme for the discrete time nonlinear singularly perturbed systems using multi-time-scale neural network and optimal bounded ellipsoid method is proposed firstly. Compared with other gradient descent based identification schemes, the new identification method proposed in this research can achieve faster convergence and higher accuracy due to the adaptively adjusted learning gain. Later, the optimal bounded ellipsoid based identification method for discrete time systems is extended to the identification of continuous singularly perturbed systems. Subsequently, by adding two additional terms in the weight's updating laws, a modified identification scheme is proposed to guarantee the effectiveness of the identification algorithm during the whole identification process. Lastly, through introducing some filtered variables, a robust neural network training algorithm is proposed for the system identification problem subjected to measurement noises.

Based on the identification results, the singular perturbation theory is introduced to decompose a high order multi-time-scale system into two low order subsystems – the reduced slow subsystem and the reduced fast subsystem. Then, two controllers are designed for the two subsystems separately. By using the singular perturbation theory, an adaptive controller for a regulation problem is designed in this research firstly. Because the system order is reduced, the adaptive controller proposed in this research has a simpler structure and requires much less computational resources, compared with other conventional controllers. Afterwards, an indirect adaptive controller is proposed for solving the trajectory tracking problem. The stability of both identification and control schemes are analyzed through the Lyapunov approach, and the effectiveness of the identification and control algorithms are demonstrated using simulations and experiments.

Acknowledgments

Firstly, I would like to express my greatest gratitude to my supervisor, Dr. Wen-Fang Xie, for providing financial support for my Ph.D study, and also for her insightful guidance, selfless help and enormous encouragement during my research. She spent plenty of time in solving the problems in my research and reviewing my papers, word by word. She always replied to me timely no matter it's early in the morning or late in the night. She showed great tolerance and patience when I made mistakes. Without her help, I couldn't accomplish what I have done so far.

Secondly, I would like to thank my parents, my sister, my grandparents and my uncles and aunts, for their love and care. Although I am thousands of miles far away from home, their kind words always make me feel warm in my heart. Also I would like to thank my girlfriend, for her accompaniment, which brings brightness and vividness to my life.

I am also grateful to my committee members, Dr. Henry Hong, Dr. Rolf Wuthrich, Dr. Nizar Bouguila, Dr. Youmin Zhang, Dr. Mingyuan Chen, Dr. Yajun Pan, for their comments as well as their time spent and efforts made in reviewing my reports and coming to my exams and defense.

My co-authors are also greatly acknowledged, Dr. Jing Na, Dr. Xuemei Ren, Dr. Zhijun Fu for their help in solving the theoretical problems and reviewing my papers. I also would like to thank my friends and colleagues, Tingting Shu, Juqi Hu, Baoguang Xu, Xiaoming Zhang, Shutong Li, Pengcheng Li, Rui Zeng, for their valuable suggestions and kind help.

Contents

List of Figures

List of Tables

List of Acronyms

SPS	Singularly perturbed system
SPT	Singular perturbation theory
NN	Neural network
FNN	Feedforward neural network
RNN	Recurrent neural network
HONN	High order neural network
RHONN	Recurrent high order neural network
BP	Backpropagation
RTRL	Real time recurrent learning
EKF	Extended Kalman filter
DEKF	Dual extended Kalman filter
OBE	Optimal bounded ellipsoid
PAS	Practically asymptotically stable
UUB	Uniformly ultimately bounded

List of Notations

ε	Small parameter related to system time scale
u	Control input
x, y	Slow and fast states of the nonlinear SPS
$\hat{x}, \hat{y}$	Estimation of the slow and fast states
ς_x, ς_y	Identification errors of x and y
A, B	Stable matrices which determine the linear parts of the NN
W_1, W_2, W_3, W_4	NN weight matrices (vectors)
$W_1^*, W_2^*, W_3^*, W_4^*$	Nominal NN weight matrices (vectors)
V_1, V_2, V_3, V_4	NN hidden layer weight matrices
V_1^*, V_2*, V_3*, V_4*	Nominal NN hidden layer weight matrices
$\bar{W}_1, \bar{W}_2, \bar{W}_3, \bar{W}_4$	Upper bounds of the Frobenius norm of W_1, W_2, W_3, W_4
$\bar{V}_1, \bar{V}_2, \bar{V}_3, \bar{V}_4$	Upper bounds of the Frobenius norm of V_1, V_2, V_3, V_4
$\Psi_1, \Psi_2, \Psi_3, \Psi_4$	NN activation function matrices (vectors)
ζ_x, ζ_y	Modeling errors of x and y in identification
$\bar{\zeta}_x, \bar{\zeta}_y$	Upper bounds of the errors ζ_x, ζ_y
x_i, y_j	The ith and jth element of x and y, respectively
θ_{xi}, θ_{yj}	NN weight vectors corresponding to x_i and y_j
$\theta_{xi}^*, \theta_{yj}^*$	Nominal NN weight vectors corresponding to x_i and y_j, respectively
H_{xi}, H_{yj}	NN regressors corresponding to x_i and y_j
ζ_{xi}, ζ_{yj}	The ith and jth elements of ζ_x and ζ_y

$\bar{\zeta}_{xi}, \bar{\zeta}_{yj}$	Upper bounds of ζ_{xi} and ζ_{yj}
τ_{xi}, τ_{yj}	Auxiliary system outputs
$\hat{\tau}_{xi}, \hat{\tau}_{yj}$	Estimated auxiliary system outputs
e_{xi}, e_{yj}	Auxiliary output errors
$\tilde{\theta}_{xi}, \tilde{\theta}_{yj}$	Weight error vectors
E_x, E_y	Tracking errors
$\hat{E}_x, \hat{E}_y$	Estimated tracking errors
$S_{x,k}, S_{y,k}$	Weight error ellipsoids in discrete system
$E_{x,k}, E_{y,k}$	Modeling error ellipsoids in discrete system
S_{xi}, S_{yj}	Weight error ellipsoids in continuous system
E_{xi}, E_{yj}	Modeling error ellipsoids in continuous system
$\lambda_{xi}, \lambda_{xi}$	Designed parameters which determines the learning gains
δ'_x, δ'_y	Modeling errors in control
δ_x, δ_y	Augmented modeling errors in control
x_d, y_d	Reference signals
P_{xie}, P_{yje}	Equilibrium points of P_{xi} and P_{yj}
p_{xr}	The rth element of P_{xi}
p_{xre}	The equilibrium point of p_{xr}
J	The moment of inertia
ω	The angular velocity
k_t	The torque force constant
k_b	The back electromotive force constant
τ	The "stretched" time variable
τ_x, τ_y	Auxiliary system outputs
υ_x, υ_y	The approximate solution errors
er_x, er_y	The state errors
τ_{xf}, τ_{yf}	Filtered auxiliary system outputs

$\hat{\tau}_x, \hat{\tau}_y$	The estimation of auxiliary system outputs
$\hat{\tau}_{xf}, \hat{\tau}_{yf}$	Filtered estimation of auxiliary system outputs
$\tilde{x}, \tilde{y}$	Filtered identification errors
e_{xf}, e_{yf}	Filtered auxiliary output errors

Chapter 1

Introduction

1.1 Motivation

Many industrial systems are nonlinear with "slow" and "fast" dynamic states because of the presence of some "parasitic" parameters such as small time constants, resistances, inductances, capacitances, masses and moments of inertia. These phenomenons can be found in harmonic drive systems [1], flight test trajectory control systems [2], flexible link manipulators [3], and power systems [4], etc. These systems are usually labeled as "singularly perturbed" or "multi-time-scale" systems. For example, in the DC motor system, the current is the fast dynamic while the angular velocity is the slow dynamic [5]. In the flexible joint robot, the joint angle is the slow dynamic while the difference between the joint angle and the motor angle is the fast dynamic [6]. In the aircraft, the heading is the slow dynamic while the altitude and flight path angle are the fast dynamics [7]. Typical SPS is presented in Fig. 1.1. Due to the existence of the fast dynamic states, the singularly perturbed systems usually have high system orders, which greatly increases the difficulties in system modeling, analysis and controller design. A simple way to reduce the system order is to neglect those fast dynamic system states. However, a design based on a simplified model may result in a system with much worse performance, or even in an unstable system [5]. An effective way to overcome this problem is to separate the original system states into those that

(a) DC motor. (b) Flexible joint robot [8]. (c) Aircraft [9].

Figure 1.1: Typical singularly perturbed systems.

change rapidly and those that vary slowly on the chosen time scale using singularly perturbation technique.

Singular perturbation theory (SPT) has been proved to be a useful tool to control and analyze the singularly perturbed systems (SPSs), because of its remedial features of both dimensional reduction and stiffness relief [10]. The extensive research in the field of singular perturbations and time-scales has resulted in the publication of numerous survey papers, reports and proceedings of conferences [5, 10–14]. So far, many well-established theories are focused on linear or nonlinear systems with full knowledge of system model parameters [3, 5, 11, 14, 15]. Nevertheless, the accurate and faithful mathematical models of those systems are usually difficult to obtain due to the uncertainties and nonlinearities. In this case, system identification becomes important and necessary before a singular perturbation theory based control scheme can be designed.

Recently, the identification of nonlinear singularly perturbed system using multi-time-scale neural networks (NNs) has been investigated by some researchers. In [16, 17], the single layer recurrent neural network (RNN) with different time scales were used to identify the nonlinear systems. System identification schemes via two-time-scale multilayer RNN were proposed in [18] later. In [19–21], the stability properties of RNN with different time scales were discussed. Robustness stability results for uncertain two-time scale RNN under parameter perturbations were established in [22]. However, it should be pointed out that the most popular training methods for the NNs were gradient-like learning laws, such as backpropagation (BP) method. The main drawback of these methods was that their convergence speed was relatively slow.

In [23–25], the direct adaptive control using singular perturbation theorem and neural network was discussed. Nevertheless, the indirect adaptive control using singular perturbation theorem and multi-time-scale NN was rarely studied. In [26,27], the indirect adaptive controllers were designed based on the multi-time-scale NN identification results. However, the authors did not take advantage of the identified model to design two controllers for the slow and fast subsystems respectively using singular perturbation theorem. Instead, the authors treated the system as a regular system, and designed a controller for the whole system. Thus, the order of the matrices in the controller could be very high if the slow and fast system states have high dimensions, and the matrices in the controller could be ill-conditioned because the inverse of the singular perturbation parameter was involved, and the singular perturbation parameter was usually very small.

Due to the slow learning speed of existing training methods for multi-time-scale neural networks, a new identification scheme is expected to achieve faster convergence. Meanwhile, since the combination of multi-time-scale neural network and singular perturbation theory is not well investigated, it is meaningful to conduct research on how to design an adaptive controller to achieve satisfactory closed-loop system properties based on the identified system model using the multi-time-scale neural networks. These are the main motivations of this research project.

1.2 Research Objectives

After reviewing the identification and control methods for nonlinear singularly perturbed systems that have been presented in literature so far, it is noted that there are still many problems remaining unsolved. For instance:

1) Although multi-time-scale NNs have been applied in identification for nonlinear SPSs, the most widely used training methods are based on gradient descent updating algorithm with fixed "learning gain", such as BP and RTRL algorithm. The main drawback of these training methods is that the convergence speed may be very slow [28]. To solve this problem, a weight estimation algorithm based on extended Kalman filter theory, or optimal bounded ellipsoid theory could be used.

In this Ph.D. research project, the neural network training algorithm based on optimal bounded ellipsoid (OBE) theory will be investigated, because the theoretical analysis of EKF based training algorithm requires the modeling uncertainty of the NN to be Gaussian process, which may not be true in the real applications [29].

2) Most research works on controller design for singularly perturbed systems presented in literature are focused on linear or nonlinear SPSs with known system model and parameters. Very rare research results on indirect adaptive control for nonlinear SPSs with unknown system model and parameters are found in the literature. The other objective of this Ph.D. research project is to study the indirect adaptive controller design scheme for nonlinear singularly perturbed systems using multi-time-scale neural networks and singular perturbation theory.

A roadmap of this research is presented in Fig. 1.2.

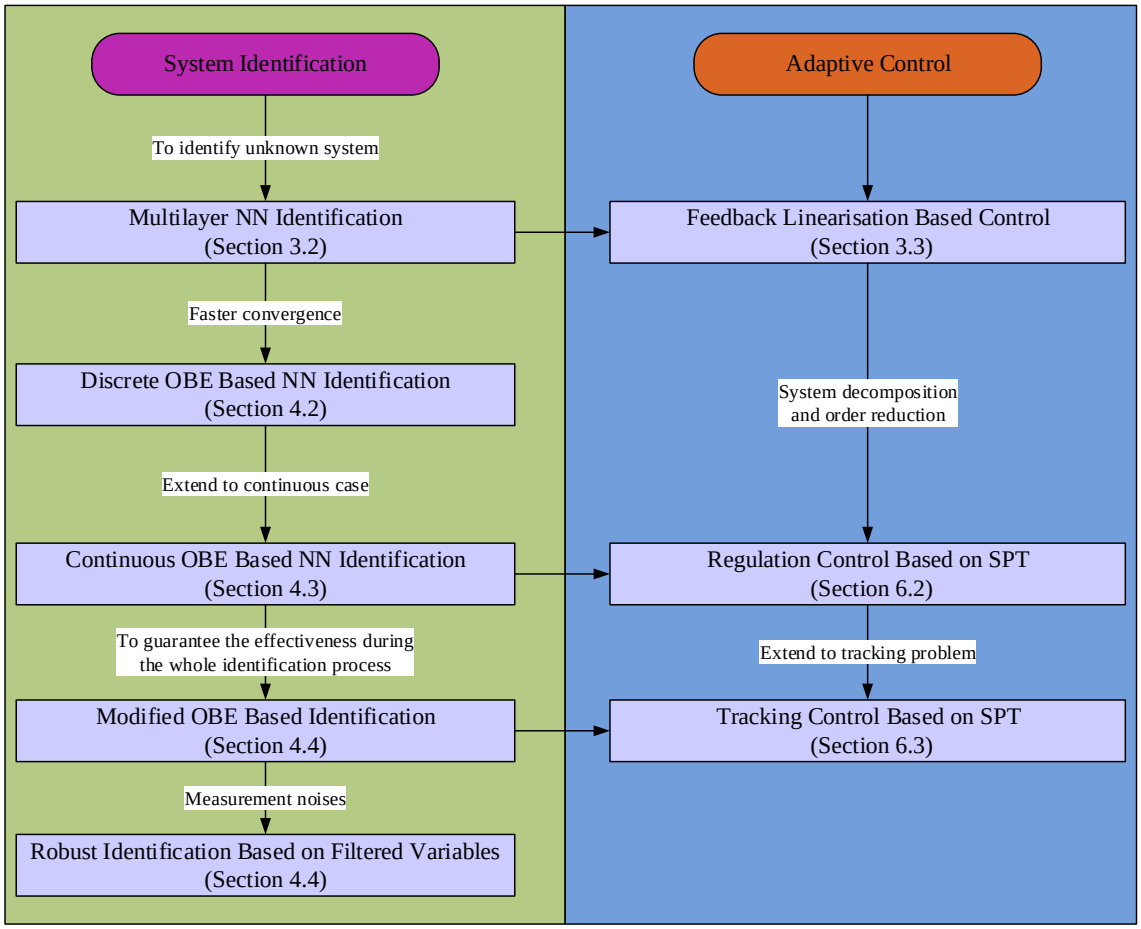


Figure 1.2: Roadmap of the research.

1.3 Contributions

In this Ph.D research, the identification and control of the nonlinear singularly perturbed systems using multi-time-scale neural networks are investigated. The contributions of this Ph.D research are listed as follows:

1) An identification scheme using the multilayer neural network is proposed which can achieve more accurate identification results due to the extra hidden layer.

2) The optimal bounded ellipsoid algorithm based identification scheme is proposed for a discrete time system. This new scheme can achieve faster convergence because the learning gain can be adjusted adaptively during the identification process.

3) The discrete time OBE based identification scheme is extended to a continuous case.

4) A modified identification scheme is proposed where two extra terms are added into the weight's updating laws. As a result, the gain matrix will converge to a user-defined equilibrium instead of $\mathbf{0}$, thus the identification scheme will remain effective during the whole identification process.

5) In order to avoid using the derivatives of the identification errors, which usually magnifies the noises, a robust identification scheme using filtered variables is proposed. Thus this new identification scheme is more robust to measurement noises.

6) An indirect adaptive controller using feedback linearisation and sliding mode technique is designed based on the identified models.

7) To solve the regulation problem, an indirect adaptive controller is designed based on the PAS theory. By using the singular perturbation theory, the original high order multi-time-scale system can be decomposed into two reduced order subsystems. The indirect controllers are designed for the reduced order subsystems. Thus the controller structure is simplified and the required computational resources are reduced.

8) To solve the trajectory tracking problem, an adaptive controller is designed without using the PAS theory. Through Lyapunov approach, the upper bound of ε is found, and the closed-loop stability is guaranteed for any $0 < \varepsilon < \varepsilon^*$.

9) A harmonic drive DC motor system is set up for experimental purpose. The identification and control schemes proposed in this research are tested on this DC motor system, and the effectiveness of the proposed schemes are verified.

1.4 Publications

The presented research work is documented in a number of journals and conference proceedings. The following is the list of author's publications.

Journal Publications:

1. D.-D. Zheng and W.-F. Xie, "Identification and trajectory tracking control of nonlinear singularly perturbed system," *IEEE Transactions on Industrial Electronics*, 2016, accepted.

2. D.-D. Zheng, W.-F. Xie, X. Ren, and J. Na, "Identification and control for singularly perturbed systems using multitime-scale neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, no. 99, pp. 1–13, 2016.

3. D.-D. Zheng, Z. Fu, W.-F. Xie, and W. Luo, "Indirect adaptive control of nonlinear system via dynamic multilayer neural networks with multi-time scales," *International Journal of Adaptive Control and Signal Processing*, vol. 29, no. 4, pp. 505–523, 2015.

Conference Publications:

1. D.-D. Zheng and W.-F. Xie, "Indirect adaptive control of flexible joint robotic manipulator," in *Proceedings of the World Congress of the International Federation of Automatic Control*, Submitted for publication.

2. ——, "Robust identification for nonlinear singularly perturbed systems using multi-time-scale recurrent neural network," in *Proceedings of the American Control Conference*, Submitted for publication.

3. D.-D. Zheng, W.-F. Xie, and X. Ren, "Identification and control for singularly perturbed systems using multi-time-scale neural networks," in *Proceedings of the 2015 IEEE International Conference on Information and Automation*, Lijiang, China, Aug. 2015, pp. 1233–1239.

4. D.-D. Zheng and W.-F. Xie, "Identification for nonlinear singularly perturbed system using recurrent high-order multi-time scales neural network," in *Proceedings of the American Control Conference*, Chicago, USA, Jul. 2015, pp. 1824–1829.

5. D.-D. Zheng, W.-F. Xie, and S. Dai, "Identification of singularly perturbed nonlinear system using recurrent high-order neural network," in *Proceedings of the 11th World Congress on Intelligent Control and Automation*, Shenyang, China, Jul. 2014, pp. 5770–5775.

1.5 Thesis Organization

The rest of the thesis is organized as follows. A comprehensive literature review is presented in Chapter 2. In Chapter 3, a system identification scheme for a class of nonlinear SPSs using multilayer multi-time-scale neural network is proposed, where the NN weights are updated by a gradient-like training algorithm. Based on the identified system model, a feedback controller is designed for a trajectory tracking problem. In Chapter 3, the identification for both discrete time and continuous time nonlinear SPSs based on the OBE algorithm are discussed. Also, a modified OBE based training algorithms are proposed to guarantee the effectiveness of the identification scheme during the whole identification scheme. In Chapter 5, a new identification scheme using filtered variables is proposed, which is more robust to measurement noises. In Chapter 6, an indirect adaptive controller using the SPT based on the identified NN models is proposed firstly for a regulation problem. Subsequently, the SPT based indirect adaptive controller for a trajectory tracking problem is also investigated. The Chapter 7 presents the conclusions of the thesis and the future work based on current research.

Chapter 2

Literature Review

2.1 Neural Networks

In 1943, a computational model for neural networks was created by W. McCulloch and W. Pitts based on mathematics and algorithms [38]. This model is called threshold logic. Since then, the investigation on application of neural networks to artificial intelligence has been conducted by more and more researchers, especially after the backpropagation algorithm, which effectively solved the exclusive-or problem, was created by P. Werbos in 1974 [39].

From a control engineer's point of view, the most significant ability of neural networks is its universal approximation to any nonlinear function, which can be used to deal with nonlinear systems [40]. In order to control these nonlinear systems with different time scales, one may need to construct the system model. Although many control methods are not dependent on system model, such as PID controller or Bang-bang controller, some other controllers can only be designed when a precise system model is built. However, in real systems, it is a nontrivial task to know the exact system model and its parameters. Sometimes, the system plants are even considered to be a black box.

To solve this problem, neural networks have been extensively studied in the past decades, since they have many excellent properties such as parallel distributed processing, learning and adaption,

data fusion, multi-input multi-output processing. The research results demonstrate that neural network is very effective in identification and control of nonlinear systems with unknown dynamics due to its powerful nonlinear approximation property. For instance, in [41–43], feedforward neural networks (FNNs) were utilized to identify the system models. In [44], an adaptive feedforward neural network tracking controller was proposed for a robotic manipulator with input deadzone and output constraint. In [45], an adaptive predictor incorporated with a high-order neural network (HONN) observer was proposed to obtain the future system states predictions, which were used in the control design to circumvent the input delay and nonlinearities. An adaptive neural network control method was investigated in [46] to stabilize a class of uncertain nonlinear strict-feedback systems with full-state constraints.

Recent results show that recurrent neural network (RNN) is more effective than feedforward neural network in system identification, since RNN incorporates feedback, and has dynamic memory, which provides more powerful representation capability [47–49]. The identification of nonlinear singularly perturbed systems using multi-time-scale RNN was further established in [16–18]. In [16, 17], single layer RNN with two time scales were used in system identification. Later, a multilayer RNN with different time scales was used in [18] for system identification purpose.

2.1.1 Feedforward and Recurrent Neural Networks

Until now, the most popular neural networks used by the researchers are feedforward neural networks and recurrent (dynamic) neural networks. Feedforward neural networks are usually used as the representation of a nonlinear function in the right-hand side of the equations of the dynamic system models [40, 48]. The structure of a typical FNN is shown in Fig. 2.1. It is noticed from Fig. 2.1 that in a feedforward neural network, the information always moves in one direction and never comes back. Feedforward neural networks have been widely implemented in identification and controller design for many systems. For instance, F. L. Lewis et al. used a multilayer neural network to design a controller for a general serial-link rigid robot arm [50]. In [51], W. Chen et al. used multilayer neural network to design an adaptive controller for a class of strict-feedback

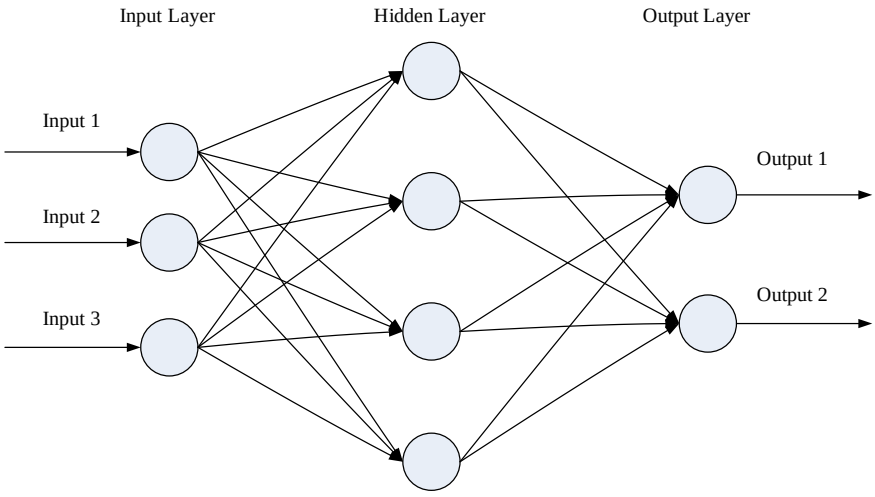


Figure 2.1: Structure of the feedforward neural network.

systems with unknown time-varying disturbance of known periods which appear in unknown nonlinear functions. FNN was also used to study the problem of identification for nonlinear systems in the presence of unknown driving noise in [41]. The methods for identification and control of dynamic systems by using different types of FNNs and generalized weight adaptation algorithms were discussed in [42].

In spite of its immense popularity in nonlinear system approximation and control, feedforward neural networks have some drawbacks, such as long computation time, sensitivity to external noise and difficulty in obtaining an independent system simulator [52]. Also the information on the local data structure is not used in weights updating, and the function approximation is sensitive to the training data [47]. Unlike the feedforward neural networks, the recurrent neural networks incorporate feedback, and have dynamic memory, which provides more powerful representation capabilities. Hence dynamic neural network is more suitable for representing dynamic systems [47, 52]. The structure of a common recurrent neural network is shown in Figure 2.2.

2.1.2 Single Layer, Mulilayer and High Order Neural Networks

Neural networks can be generally classified as single layer NN and multi-layer NN. The structure of a single layer NN is shown in Fig. 2.3, and the FNN and RNN shown in Fig. 2.1 and 2.2

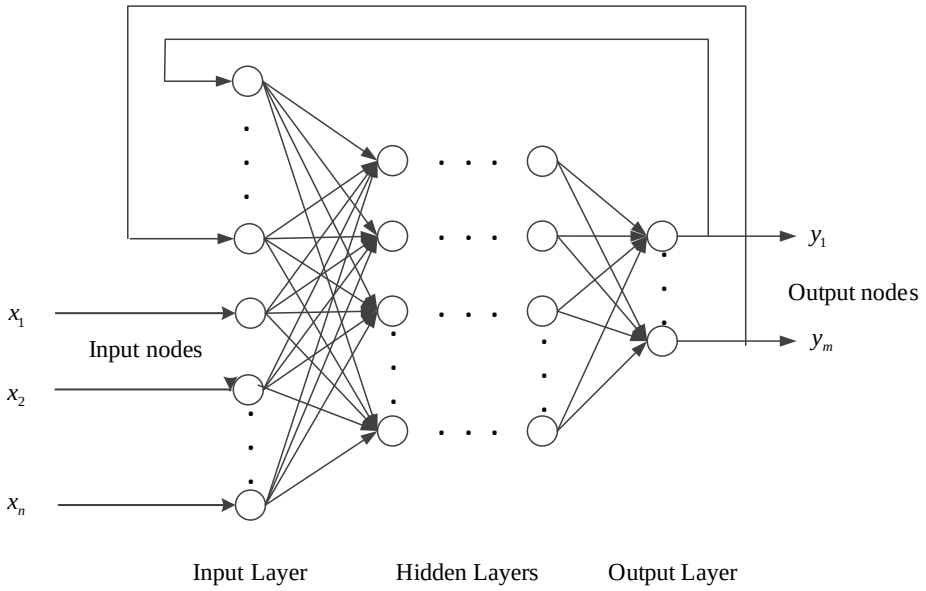


Figure 2.2: Structure of the recurrent neural network.

are multi-layer NN. Both the single layer NN and multi-layer NN have been widely used [53–55]. Single layer NN has a simple structure, but the approximation capability is poor. Multilayer NN is more powerful in nonlinearity approximation, but it requires a great deal of computational time due to its complex structure, which hinders its application in real-time identification and control. Unlike the single layer NN or the multilayer NN, high order neural network can achieve superior approximation capability with simple structure, because it allows higher-order interconnections between the neurons. It is also demonstrated that if enough high-order connections are allowed, this HONN can be used to approximate arbitrary dynamic system [56]. The feasibility and advantages of HONN in applications for system identification and control are demonstrated by many researchers [45, 57–59].

2.1.3 Multi-time-scale Neural Networks

In 1995, Meyer-Base et al. studied the short-term and long-time memory of competitive neural networks with different time-scale using quadratic-type Lyapunov functions in [60]. New methods of analyzing the dynamics of a multi-time-scale competitive neural system was proposed in [61].

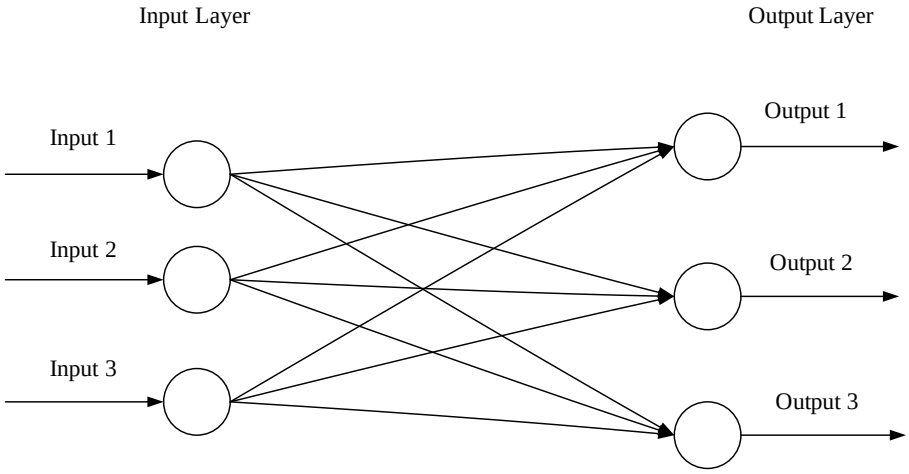


Figure 2.3: Structure of the single layer neural network.

Later, a global stability method and a modality of detecting the local stability behavior around individual eqilibrium points was presented in [62]. In [63], a new method of analyzing the dynamics of a neural network with different time-scale based on the theory of flow invariance was proposed. Recent results on stability property analysis of multi-time-scale neural networks could be found in [20, 21].

The identification of nonlinear singularly perturbed systems using multi-time-scale RNN was investigated in [16–18]. In [16, 17], single layer RNN with two time-scale were used in system identification, where the identification of the linear part matrices of the RNN was proposed and dead-zone indicators were introduced to prevent the weights of neural network from drifting. Although single layer RNN has a simple structure, which does not require much computation resources, it has limited approximation capability. To overcome this drawback, a multilayer RNN with different time-scale was used in [18], because the extra hidden layer offers the possibility of more complex nonlinear mapping between the NN inputs and outputs, which led to better approximation performance. But the computational load of multi-time-scale multilayer RNN is also very high, which limits its application in real time identification and control.

2.1.4 Training Methods for Neural Networks

In 1974, the backpropagation algorithm was proposed by P. Werbos, which effectively solved the exclusive-or problem [39]. Since then, backpropagation has become a common method in training different kinds of neural networks. BP algorithm is used in conjunction with an optimization method such as gradient descent, which calculates the gradient of a loss function with respects to all the weights in the neural network. The limitation for BP and other gradient descent training methods are that the convergence speed may be very slow, and the training process of the NN is sensitive to measurement noise, etc. [28].

To overcome these drawbacks, some researchers treated the learning algorithm as parameter estimation problem for a nonlinear system. By introducing the estimation methods, such as extended Kalman filter, which was originally developed for parameter estimation of the general nonlinear systems, the minimum variance estimation of the link weights of the neural networks could be obtained. For instance, Y. Iiguni et al. proposed a real-time training algorithm for a multilayer neural network based on the EKF in [28], which could converge in fewer iterations than BP method. Meanwhile, the tuning parameters that crucially govern the convergence properties were not included which could make its application easier. In [64], the authors developed an effective EKF based RNN training approach with a controllable training convergence, and the training convergence problem was proposed and studied by updating two artificial training noise parameters. In [65], the usage of dual extended Kalman filter (DEKF) in estimation of both hidden layer states and RNN weights, and the removal of some unimportant weights from a trained RNN were discussed. The DEKF algorithm was also used in the training of RNN with special emphasis on its application to control system design in [66]. The drawback of Kalman Filter based training methods is that the modeling uncertainty of NN needs to be Gaussian process in their theoretical analysis [29].

In 1979, L. G. Khachiyan first indicated how an ellipsoid method for linear programming could be implemented in polynomial time [67]. This result has caused great excitement and led to massive research on this topic. Recent results show that bounded ellipsoid is effective in improving the

learning speed of NN. For instance, a modified optimal bounded ellipsoid algorithm was proposed to train the weights of a RNN in the identification process of a nonlinear system in [68]. In [69], the OBE algorithm was applied to feedforward NN weights training. Wen Yu and Jose de Jesus Rubio proposed an ellipsoid propagation algorithm to train the weights of both hidden layer and output layer of a RNN, and analyzed the stability property of the identification process in [29]. It should be pointed out that most of previous research were conducted for discrete time single-time-scale systems. The identification for continuous multi-time-scale systems using OBE algorithm had not been developed.

2.2 Singularly Perturbed Systems and Singular Perturbation Theory

Many industrial systems are nonlinear with"slow" and"fast" dynamics because of the presence of some parasitic parameters such as small time constants, resistances, inductances, capacitances, masses and moments of inertia. These parasitic parameters are often the cause of the increased order and "stiffness" of a real physical system. These systems in which the suppression of a small parameter will result in the degeneration of dimension, are called "singularly perturbed" systems, or more generally, multi-time-scale systems [11, 12]. The "curse of dimensionality" of singularly perturbed systems often poses formidable amount of computation in analysis and controller design. Although many control theories are valid for any system order, their actual use is often limited to lower order models. The interaction of "slow" and "fast" dynamics in high order systems results in "stiff" numerical problems which require expensive integration routines [11].

The singular perturbation theory, a traditional tool of fluid dynamics, is recognized to be effective in reducing the dimensions and relieving stiffness, since it first became a means for simplified computation of optimal trajectories in 1960s [12]. This methodology has an impressive record of applications in wide spectrum of fields, and a lot of research surveys, journal papers and book chapters have been published so far [5, 10–14].

A fundamental feature of control methods based on singular perturbation theory is a decomposition of the feedback control design problem into two design subproblems for the slow and fast dynamics respectively. The two sub-designs are then combined to give the design for the full system. It should be pointed out that most publications are focused on linear or nonlinear singularly perturbed systems with known system model and parameters. For example, in [3, 70], the integral manifold concept was used to design nonlinear controllers for tip position trajectory tracking of the flexible link manipulators. In [71], Narang addressed the control problem for a general class of non-affine, non-standard singularly perturbed continuous-time systems. Other control methods, such as composite fast-slow model predictive control (MPC) [72], bond graph approach [73], composite observer-based feedback control based on the Lyapunov stability theorem and linear matrix inequality (LMI) [74], feedback linearizion control [75], linear-quadratic regulator [76], sliding mode control [15, 77], and switched output feedback control [78] were also reported in literature. More examples can be found in [10, 12, 13] and references therein.

To solve the control problem for SPSs with uncertainties, different robust control methods have been proposed in literature. In [79], a robust controller was designed for a nonlinear SPS with vanishing uncertainties. A robust output feedback control scheme was developed for nonlinear SPSs with time-varying uncertain variables in [80]. In [81], passivity-based integral sliding-mode control of uncertain SPSs was discussed. Wang et al. investigated the robust asymptotic stablization of a class of nonlinear uncertain SPSs by using the nonlinear PI control techniques in [82]. An adaptive control based on on-line estimated system parameters was also adopted to handle the SPSs with uncertainties. In [83–87], the unknown system parameters of robotic manipulators were estimated during the control process, and different controllers were designed based on the estimated system parameters, such as integral manifold concept controller, sliding mode controller, feedback linearisation controller and H_∞ controller. Zheng et al. [88] investigated the nonlinear adaptive sliding mode control problem of induction motors, in which the estimated fluxes and the rotor resistance were guaranteed to converge to their true values.

Adaptive neural networks control is classified into two kinds of structure: indirect and direct

adaptive control. In direct adaptive control, the parameters of the controller are changed directly without determining the characteristics of the process and its disturbance first [89–93]. In indirect adaptive control, the controller is designed based on the process model and possibly the disturbance model which are determined first [48, 49, 94, 95]. Although there are limited adaptive control methods developed for few specific class of systems with unknown parameters, they are not applicable when the system model is not known, or when the system plant is considered to be a "black box" [86]. The research on indirect adaptive control for nonlinear SPSs based on neural networks is rare, because the structure of the neural networks is usually very complicated, and it is very difficult to decompose the original system into reduced slow subsystem and reduced fast subsystems such that the SPT can be applied, if the original system are represented by NN. Although the identification algorithms for nonlinear SPSs using multi-time-scale RNN have been well developed, the controller design based on the identified model is still a very challenging problem. In [26, 27], a multi-time-scale RNN was employed for the identification purpose, but the authors did not make the best use of the identified model, i.e., they did not use singular perturbation theory to reduce the system order in the controller design process. When the system order is high, the controller could be very complex and require huge computational resource.

Chapter 3

Identification and Control Using Gradient Descent and Feedback Linearisation

3.1 Introduction

In literature, the adaptive identification and control for nonlinear singularly perturbed systems via multi-time-scale neural networks have been established. However, the dynamic neural networks used in [16, 26, 27] only contain single output layer. In order to improve the system performance, it is reasonable to implement multilayer dynamic neural networks instead of single layer one for system identification and control, since the extra hidden layer gives the possibility of more complex nonlinear mapping between the inputs and the outputs [40, 47, 50] which can improve the approximation performance [40].

So far, some techniques have been developed to solve the potential singularity problem and ensure bounded control signal. In [96, 97], the control law was set to zero when the system parameter went into a ball near the singularity point. Lewis proposed a control structure consists of a robustifying portion which keeps the control signal bounded in [98]. In [48], an identification and control scheme based on projection algorithm was designed to solve the singularity problem. However, none of the above mentioned method was designed for a multi-time scale system with

multilayer neural networks.

In this chapter, a new system identification scheme based on multi-time-scale multilayer neural network is presented. Then, based on the identified system model, an indirect adaptive controller is designed for a trajectory tracking problem. As a main contribution of this chapter, the on-line updating laws for both the hidden layers and output layers of the recurrent neural networks are proposed. In addition, the e-modification [99] and a novel correction term are introduced in the on-line updating laws to guarantee bounded parameter estimations. Also, the potential singularity problem is solved by designing the identification algorithm so that the determinant of the control gain matrix will stay away from zero all the time during the identification process. The stability property of the identification and control scheme are discussed via Lyapunov approach.

3.2 Preliminaries

In this section, some basic definitions and terminologies will be introduced.

1. **Smoothness**.

The smoothness of a function is a property described by the number of its continuous derivatives. A function f is said to be differentiable of class C^k (or f is C^k) if the derivatives f', f'', $\cdots$, f^k exist and are continuous. The function f is C^∞, or smooth if it has derivatives of all orders [100].

2. **Norm**.

(1) Frobenius Norm. The Frobenius norm of a $m \times n$ matrix A is defined as:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}, \tag{3.1}$$

where a_{ij} is the element in ith row and jth column of A.

(2) L_1-Norm. The L_1-norm of a vector $x \in \Re^n$ is defined as:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|, \tag{3.2}$$

where x_i is the ith element of x.

(3) L_2-Norm. The L_2-norm of a vector $x \in \Re^n$ is defined as:

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}, \tag{3.3}$$

where x_i is the ith element of x.

3. **Semi-globally Uniformly Boundedness (SGUUB)**.

The solution of $\dot{x} = f(x,t)$ is SGUUB if for any compact set Ω_0, there exists an $S > 0$ and $T(S, x(t_0))$ such that $\|x(t)\| \leq S$ for all $x(t_0) \in \Omega_0$ and $t \geq t_0 + T$.

4. **Affine-in-control system**.

A nonlinear system is called a control-affine system or affine-in-control system if it can be expressed as

$$\dot{x} = f(x) + \sum_{i=1}^{m} g_i(x) u_i, \tag{3.4}$$

where $x \in \Re^n$ is the system state, $u \in \Re^m$ is the control input, f and g_i are smooth functions with appropriate dimensions [101].

3.3 System Identification Using Multilayer Neural Network

In this section, a new identification scheme will be proposed for a class of singular perturbed nonlinear systems with two different time scales described by [5]:

$$\begin{aligned} \dot{x} &= f_1(x, y, u, t), \\ \varepsilon \dot{y} &= f_2(x, y, u, t), \end{aligned} \tag{3.5}$$

where $x \in \Re^n$ and $y \in \Re^m$ are slow and fast state variables, $u \in \Re^{m+n}$ is the control input vector and ε is a small parameter. $f_1(\cdot)$, $f_2(\cdot)$ are unknown smooth functions.

In order to identify the nonlinear dynamic system (3.5), the following recurrent neural network with two-time-scale is employed:

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + W_1\Psi_1(V_1[x; y]) + W_2\Psi_2(V_2[x; y])u, \\ \varepsilon\dot{\hat{y}} &= B\hat{y} + W_3\Psi_3(V_3[x; y]) + W_4\Psi_4(V_4[x; y])u, \end{aligned} \tag{3.6}$$

where $\hat{x} \in \Re^n$, $\hat{y} \in \Re^m$ are the estimation of the slow and fast state variables using neural networks, $A \in \Re^{n\times n}$ and $B \in \Re^{m\times m}$ are stable matrices, $W_1 \in \Re^{n\times(n+m)}$, $W_2 = [\text{diag}(w_{21}, \cdots, w_{2n}), \mathbf{0}] \in \Re^{n\times(n+m)}$, $W_3 \in \Re^{m\times(n+m)}$, $W_4 = [\mathbf{0}, \text{diag}(w_{41}, \cdots, w_{4m})] \in \Re^{m\times(n+m)}$ are the weights in the output layers, $V_i \in \Re^{(n+m)\times(n+m)}$, $i = 1, \cdots, 4$ are the weights in the hidden layers, $\Psi_i(V_i[x; y]) = \big[\psi_i\big((V_i[x; y])_1\big), \cdots, \psi_i\big((V_i[x; y])_{n+m}\big)\big] \in \Re^{(n+m)}$, $i = 1, 3$, $\Psi_j(V_j[x; y]) = \text{diag}\big[\psi_j\big((V_j[x; y])_1\big), \cdots, \psi_j\big((V_j[x; y])_{n+m}\big)\big] \in \Re^{(n+m)\times(n+m)}$, $j = 2, 4$. $(V_i[x; y])_q$, $q = 1, \cdots, n + m$ and $(V_j[x; y])_q$, $q = 1, \cdots, n + m$ are q^{th} elements of $(V_i[x; y])$ and $(V_j[x; y])$ respectively. Typical presentation of $\psi_i(\cdot)$ is the sigmoid function of the form

$$\psi_i(z) = \frac{\alpha_{i,1}}{1 + e^{-\alpha_{i,2}z}} + \alpha_{i,3},\ i = 1, \cdots, 4. \tag{3.7}$$

When ε is equal to 1, the recurrent neural network(3.6) becomes a normal one [102].

Remark 3.1. *In a general case, the parameters $\alpha_{i,1}, \alpha_{i,2}, i = 1, \cdots, 4$ in* (3.7) *are positive real*

numbers and $\alpha_{i,3}, i = 1, \cdots, 4$ *are real numbers. These parameters can be chosen a priori based on trial and error. The most commonly selected values are* $\alpha_{i,1} = \alpha_{i,2} = 1, \alpha_{i,3} = 0$*, where the logistic function can be obtained. The other common selection is* $\alpha_{i,1} = \alpha_{i,2} = 2, \alpha_{i,3} = -1$*, where a hyperbolic tangent function is obtained [103]. In this section, the parameters* $\alpha_{i,j},\ i = 1, 3,\ j = 1, 2, 3$ *can be chosen arbitrarily, and* $\alpha_{i,j},\ i = 2, 4,\ j = 1, 2, 3$ *are chosen to guarantee the existence of* $[\Psi_i(V_i[x; y])]^{-1},\ i = 2, 4$.

The structure of the NN identifier is shown in Fig. 3.1.

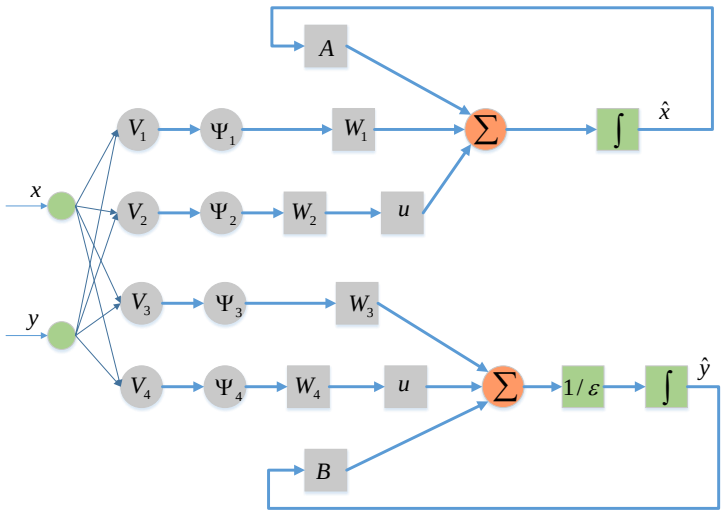


Figure 3.1: Structure of the NN identifier.

Assume that a nominal neural network model of the nonlinear system (1) with modeling error is described by the following equations

$$\begin{aligned} \dot{x} &= A^* x + W_1^* \Psi_1(V_1^*[x; y]) + W_2^* \Psi_2(V_2^*[x; y])u + \zeta_x, \\ \varepsilon \dot{y} &= B^* y + W_3^* \Psi_3(V_3^*[x; y]) + W_4^* \Psi_4(V_4^*[x; y])u + \zeta_y, \end{aligned} \tag{3.8}$$

where $W_i^*,\ i = 1, \cdots, 4$, $V_j^*,\ j = 1, \cdots, 4$ are unknown nominal constant matrices, A^*, B^* are unknown nominal constant Hurwitz matrices, the vectors ζ_x, ζ_y are modeling errors.

Assumption 3.1. *The nominal weight values and the modeling errors are bounded as*

$$\|W_1^*\| \leq \bar{W}_1,\ \|W_2^*\| \leq \bar{W}_2,\ \|W_3^*\| \leq \bar{W}_3,\ \|W_4^*\| \leq \bar{W}_4,\ \|\zeta_x\| \leq \bar{\zeta}_x,$$
$$\|V_1^*\| \leq \bar{V}_1,\ \|V_2^*\| \leq \bar{V}_2,\ \|V_3^*\| \leq \bar{V}_3,\ \|V_4^*\| \leq \bar{V}_4,\ \|\zeta_y\| \leq \bar{\zeta}_y,$$

where $\bar{W}_i,\ \bar{V}_i,\ \bar{Z}_i,\ i = 1, \cdots, 4,\ \bar{\zeta}_x,\ \bar{\zeta}_y$ *are prior known boundaries,* $\|\cdot\|$ *is the Frobenius norm.*

Remark 3.2. *It is reasonable to assume that the modeling errors are bounded by upper bounds. Similar assumptions can be found in [16, 29, 65, 69] and many other references. In fact, arbitrary small modeling errors can be obtained by increasing the number of neurons used in the NN [104].*

For notational convenience, define

$$\check{W}_i^* = \begin{bmatrix} W_i^* \\ \mathbf{0} \end{bmatrix} \in \Re^{(n+m)\times(n+m)}, \check{W}_i = \begin{bmatrix} W_i \\ \mathbf{0} \end{bmatrix} \in \Re^{(n+m)\times(n+m)},$$
$$Z_i^* = \begin{bmatrix} \check{W}_i^* & \mathbf{0} \\ \mathbf{0} & V_i^* \end{bmatrix} \in \Re^{2(n+m)\times 2(n+m)}, Z_i = \begin{bmatrix} \check{W}_i & \mathbf{0} \\ \mathbf{0} & V_i \end{bmatrix} \in \Re^{2(n+m)\times 2(n+m)}.$$

Thus the nominal weights can be further bounded as

$$\|Z_i^*\| \leq \bar{Z}_i.$$

It is assumed that the system states are measurable. The identification errors are defined by

$$\begin{aligned} \varsigma_x &= x - \hat{x}, \\ \varsigma_y &= y - \hat{y}. \end{aligned} \tag{3.9}$$

From (3.6) and (3.8), one can obtain the error dynamic equations

$$\begin{aligned} \dot{\varsigma}_x &= A^*\varsigma_x + \tilde{A}\varsigma_x + W_1^*\tilde{\Psi}_1 + \tilde{W}_1\Psi_1(V_1[x;y]) + W_2^*\tilde{\Psi}_2 u + \tilde{W}_2\Psi_2(V_2[x;y])u + \zeta_x, \\ \varepsilon\dot{\varsigma}_y &= B^*\varsigma_y + \tilde{B}\varsigma_y + W_3^*\tilde{\Psi}_3 + \tilde{W}_3\Psi_3(V_3[x;y]) + W_4^*\tilde{\Psi}_4 u + \tilde{W}_4\Psi_4(V_4[x;y])u + \zeta_y, \end{aligned} \tag{3.10}$$

where $\tilde{A} = A^* - A$, $\tilde{B} = B^* - B$, $\tilde{W}_i = W_i^* - W_i$, $\tilde{\Psi}_i = \Psi_i(V_i^*[x;y]) - \Psi_i(V_i[x;y])$, $i = 1, \cdots, 4$.

With Taylor's series expansion, it can be obtained that:

$$\begin{aligned}
\tilde{\Psi}_i &= D_{\Psi_i}\tilde{V}_i[x;y] + O(\tilde{V}_i[x;y])^2,\ i = 1, 3,\\
\tilde{\Psi}_j u &= D_{\Psi_j}\tilde{V}_j[x;y] + O(\tilde{V}_j[x;y])^2,\ j = 2, 4,\\
D_{\Psi_i} &= \frac{\partial \Psi_i(V_i[x;y])}{\partial(V_i[x;y])},\ D_{\Psi_j} = \frac{\partial \Psi_j(V_{j,q}[x;y])u}{\partial(V_j[x;y])},
\end{aligned} \tag{3.11}$$

where $\tilde{V}_i = V_i^* - V_i$, $\tilde{V}_j = V_j^* - V_j$.

The error system is further represented as:

$$\begin{aligned}
\dot{\varsigma}_x =& A^*\varsigma_x + \tilde{A}\varsigma_x + W_1 D_{\Psi_1}\tilde{V}_1[x;y] - \tilde{W}_1 D_{\Psi_1} V_1[x;y] + \tilde{W}_1\Psi_1(V_1[x;y]) + W_2 D_{\Psi_2}\tilde{V}_2[x;y]\\
&- \tilde{W}_2 D_{\Psi_2} V_2[x;y] + \tilde{W}_2\Psi_2(V_2[x;y])u + \xi_x,\\
\varepsilon\dot{\varsigma}_y =& B^*\varsigma_y + \tilde{B}\varsigma_y + W_3 D_{\Psi_3}\tilde{V}_3[x;y] - \tilde{W}_3 D_{\Psi_3} V_3[x;y] + \tilde{W}_3\Psi_3(V_3[x;y]) + W_4 D_{\Psi_4}\tilde{V}_4[x;y]\\
&- \tilde{W}_4 D_{\Psi_4} V_4[x;y] + \tilde{W}_4\Psi_4(V_4[x;y])u + \xi_y,
\end{aligned} \tag{3.12}$$

and the disturbance terms are

$$\begin{aligned}
\xi_x &= \tilde{W}_1 D_{\Psi_1} V_1^*[x;y] + \tilde{W}_2 D_{\Psi_2} V_2^*[x;y] + W_1^* O(\tilde{V}_1[x;y])^2 + W_2^* O(\tilde{V}_2[x;y])^2 u + \zeta_x,\\
\xi_y &= \tilde{W}_3 D_{\Psi_3} V_3^*[x;y] + \tilde{W}_4 D_{\Psi_4} V_4^*[x;y] + W_3^* O(\tilde{V}_3[x;y])^2 + W_4^* O(\tilde{V}_4[x;y])^2 u + \zeta_y.
\end{aligned} \tag{3.13}$$

For sigmoid, Radial Basis Function (RBF) and tanh activation functions, the higher-order terms in the Taylor series are bounded by [89]

$$\begin{aligned}
\|O(\tilde{V}_i[x;y])^2\| &\le C_{i1} + C_{i2}\|\tilde{V}_i\|\|[x;y]\|,\ i = 1, 3,\\
\|O(\tilde{V}_j[x;y])^2 u\| &\le C_{j1}\|u\| + C_{j2}\|\tilde{V}_j\|\|[x;y]\|\|u\|,\ j = 2, 4,
\end{aligned} \tag{3.14}$$

where $C_{i,1}$, $C_{i,2}$, $C_{j,1}$, $C_{j,2}$ are positive constants.

Based on Assumption 3.1 and using (3.14), the disturbance terms (3.13) are bounded as

$$\begin{aligned} \|\xi_x\| &\leq C_{x1} + C_{x2}\|\tilde{Z}_1\| + C_{x3}\|\tilde{Z}_2\|, \\ \|\xi_y\| &\leq C_{y1} + C_{y2}\|\tilde{Z}_3\| + C_{y3}\|\tilde{Z}_4\|, \end{aligned} \tag{3.15}$$

where C_{x1}, C_{x2}, C_{x3}, C_{y1}, C_{y2}, C_{y3} are positive constants, $\tilde{Z}_i = Z_i^* - Z_i,\ i = 1, \cdots, 4$.

For stable nominal matrices A^*, B^* and any positive definite matrices Q_x, Q_y, there exist positive definite matrices P_x, P_y satisfying the following equations [105]:

$$\begin{aligned} A^{*T}P_x + P_xA^* &= -Q_x, \\ B^{*T}P_y + P_yB^* &= -Q_y. \end{aligned} \tag{3.16}$$

Rewrite (3.6) as

$$\begin{bmatrix} \dot{\hat{x}} \\ \dot{\hat{y}} \end{bmatrix} = \begin{bmatrix} A\hat{x} \\ (1/\varepsilon)B\hat{y} \end{bmatrix} + \begin{bmatrix} W_1\Psi_1(V_1[x;y]) \\ (1/\varepsilon)W_3\Psi_3(V_3[x;y]) \end{bmatrix} + \begin{bmatrix} W_2\Psi_2(V_2[x;y]) \\ (1/\varepsilon)W_4\Psi_4(V_4[x;y]) \end{bmatrix} u, \tag{3.17}$$

and define

$$M = \begin{bmatrix} W_2\Psi_2(V_2[x;y]) \\ (1/\varepsilon)W_4\Psi_4(V_4[x;y]) \end{bmatrix}. \tag{3.18}$$

M can be viewed as a control gain matrix for system. In order to avoid the potential singularity problem in controller design, $\det(M) \neq 0$ should be guaranteed. It can be proved that if only $w_{2i} > \varpi,\ i = 1, \cdots, n$ and $w_{4j} > \varpi,\ j = 1, \cdots, m$, where ϖ is a sufficiently small constant, then the eigenvalues of M will not equal zero, therefore $\det(M) \neq 0$. Inspired by [48], the authors propose the following NN weight update rules for $w_{2i},\ i = 1, \cdots, n$ in W_2 and $w_{4j},\ j = 1, \cdots, m$ in W_4:

1) Whenever any $|w_{2i}| = \varpi,\ i = 1, \cdots, n$ or $|w_{4j}| = \varpi,\ j = 1, \cdots, m$, use

$$\dot{w}_{2i} = \begin{cases} \dot{W}_{2i}, & \text{if } \dot{W}_{2i}\text{sign}(w_{2i}) > 0 \\ 0, & \text{if } \dot{W}_{2i}\text{sign}(w_{2i}) \leq 0 \end{cases}, \dot{w}_{4j} = \begin{cases} \dot{W}_{4j}, & \text{if } \dot{W}_{4j}\text{sign}(w_{4j}) > 0 \\ 0, & \text{if } \dot{W}_{4j}\text{sign}(w_{4j}) \leq 0 \end{cases}. \tag{3.19}$$

2) Otherwise, the updating law is given as:

$$\dot{w}_{2i} = \dot{W}_{2i},\ \dot{w}_{4j} = \dot{W}_{4j}, \tag{3.20}$$

where $\dot{W}_{2i}$ is the i^{th} element of i^{th} row in $\dot{W}_2$, and $\dot{W}_{4j}$ is the $(n+j)^{\text{th}}$ element of j^{th} row in $\dot{W}_4$. $\dot{W}_2$ and $\dot{W}_4$ are defined in (3.21).

The updating laws for NN weights are given as follows:

$$\begin{aligned}
\dot{A} &= s_x(k_A \varsigma_x \hat{x}^T), \\
\dot{W}_1 &= s_x\{K_1 P_x \varsigma_x \Psi_1^T(V_1[x;y]) - K_1 P_x \varsigma_x (V_1[x;y])^T D_{\Psi_1} - k_x K_1 \|\varsigma_x\| W_1\}, \\
\dot{W}_2 &= s_x\{K_2 P_x [u^T \Psi_2(V_2[x;y])\varsigma_x] - K_2 P_x \varsigma_x (V_2[x;y])^T D_{\Psi_2} - k_x K_2 \|\varsigma_x\| W_2\}, \\
\dot{V}_1 &= s_x\{L_1 (W_1 D_{\Psi_1})^T P_x \varsigma_x [x;y]^T - k_x L_1 \|\varsigma_x\| V_1\}, \\
\dot{V}_2 &= s_x\{L_2 (W_2 D_{\Psi_2})^T P_x \varsigma_x [x;y]^T - k_x L_2 \|\varsigma_x\| V_2\}, \\
\dot{B} &= s_y(\varepsilon^{-1} k_B \varsigma_y \hat{y}^T), \\
\dot{W}_3 &= s_y\{\varepsilon^{-1} K_3 P_y \varsigma_y \Psi_3^T(V_3[x;y]) - \varepsilon^{-1} K_3 P_y \varsigma_y (V_3[x;y])^T D_{\Psi_3} - k_y K_3 \|\varsigma_y\| W_3\}, \\
\dot{W}_4 &= s_y\{\varepsilon^{-1} K_4 P_y \varsigma_y [\Psi_4(V_4[x;y]u)]^T - \varepsilon^{-1} K_4 P_y \varsigma_y (V_4[x;y])^T D_{\Psi_4} - k_y K_4 \|\varsigma_y\| W_4\}, \\
\dot{V}_3 &= s_y\{\varepsilon^{-1} L_3 (W_3 D_{\Psi_3})^T P_y \varsigma_y [x;y]^T - k_y L_3 \|\varsigma_y\| V_3\}, \\
\dot{V}_4 &= s_x\{\varepsilon^{-1} L_4 (W_4 D_{\Psi_4})^T P_y \varsigma_y [x;y]^T - k_y L_4 \|\varsigma_y\| V_4\},
\end{aligned} \tag{3.21}$$

with

$$s_x = \begin{cases} 1, & \|\varsigma_x\| \geq b_x \\ 0, & \|\varsigma_x\| < b_x \end{cases}, s_y = \begin{cases} 1, & \|\varsigma_y\| \geq b_y \\ 0, & \|\varsigma_y\| < b_y \end{cases}, \tag{3.22}$$

where $L_i,\ i = 1, \cdots, 4$, $K_i,\ i = 1, 3$ are positive definite matrices, $K_j,\ j = 2, 4$ are diagonal positive definite matrices. k_A, k_B, k_x, k_y are positive constants. b_x, b_y are specified as:

$$b_x = \frac{2^{-1}(k_x C_{x4}^2 + k_x C_{x5}^2) + 2\|P_x\| C_{x1}}{\lambda_{min}(Q_x)},\ b_y = \frac{2^{-1}\varepsilon(k_y C_{y4}^2 + k_y C_{y5}^2) + 2\|P_y\| C_{y1}}{\lambda_{min}(Q_y)}, \tag{3.23}$$

where C_{x4}, C_{x5}, C_{y4}, C_{y5} are positive constants to be defined later.

Theorem 3.1. *Consider the nonlinear system* (3.5) *and identification model* (3.6)*. With the updating laws proposed in* (3.19)*,* (3.20)*,* (3.21)*, the identification process can guarantee the following stability properties:*

$$\varsigma_x,\ \varsigma_y,\ A,\ B,\ W_i,\ V_i \in \mathcal{L}_\infty.$$

Proof. Case I: $s_x = 1$ and $s_y = 1$. Consider the following Lyapunov function candidate:

$$\begin{aligned} V_I =& V_x + V_y, \\ V_x =& \varsigma_x^T P_x \varsigma_x + \text{tr}\{\tilde{W}_1^T K_1^{-1} \tilde{W}_1\} + \text{tr}\{\tilde{W}_2^T K_2^{-1} \tilde{W}_2\} + \text{tr}\{\tilde{V}_1^T L_1^{-1} \tilde{V}_1\} + \text{tr}\{\tilde{V}_2^T L_2^{-1} \tilde{V}_2\} \\ & + k_A^{-1} \text{tr}\{\tilde{A}^T P_x \tilde{A}\}, \\ V_y =& \varsigma_y^T P_y \varsigma_y + \text{tr}\{\tilde{W}_3^T K_3^{-1} \tilde{W}_3\} + \text{tr}\{\tilde{W}_4^T K_4^{-1} \tilde{W}_4\} + \text{tr}\{\tilde{V}_3^T L_3^{-1} \tilde{V}_3\} + \text{tr}\{\tilde{V}_4^T L_4^{-1} \tilde{V}_4\} \\ & + k_B^{-1} \text{tr}\{\tilde{B}^T P_y \tilde{B}\}. \end{aligned} \tag{3.24}$$

Differentiating (3.24) and using (3.12) yield

$$\begin{aligned}\dot{V}_x &= L_A + L_{W1} + L_{W2} + L_{V1} + L_{V2} - \varsigma_x^T Q_x \varsigma_x + 2\varsigma_x^T P_x \xi_x,\\ \dot{V}_y &= L_B + L_{W3} + L_{W4} + L_{V3} + L_{V4} - \varepsilon^{-1}\varsigma_y^T Q_y \varsigma_y + 2\varsigma_y^T P_y \xi_y,\end{aligned} \tag{3.25}$$

where

$$\begin{aligned}L_A &= 2k_A^{-1}\text{tr}\{\dot{\tilde{A}}^T P_x \tilde{A}\} + 2\varsigma_x^T P_x \tilde{A}\hat{x},\\ L_{W1} &= 2\text{tr}\{\dot{\tilde{W}}_1^T K_1^{-1}\tilde{W}_1\} + 2\varsigma^T P_x \tilde{W}_1 \Psi_1(V_1[x;y]) - 2\varsigma_x^T P_x \tilde{W}_1 D_{\Psi_1} V_1[x;y],\\ L_{W2} &= 2\text{tr}\{\dot{\tilde{W}}_2^T K_2^{-1}\tilde{W}_2\} + 2\varsigma_x^T P_x \tilde{W}_2 \Psi_2(V_2[x;y])u - 2\varsigma_x^T P_x \tilde{W}_2 D_{\Psi_2} V_2[x;y],\\ L_{V1} &= 2\text{tr}\{\dot{\tilde{V}}_1^T L_1^{-1}\tilde{V}_1\} + 2\varsigma_x^T P_x \tilde{W}_1 D_{\Psi_1}\tilde{V}_1[x;y],\\ L_{V2} &= 2\text{tr}\{\dot{\tilde{V}}_2^T L_2^{-1}\tilde{V}_2\} + 2\varsigma_x^T P_x \tilde{W}_2 D_{\Psi_2}\tilde{V}_2[x;y],\\ L_B &= 2k_B^{-1}\text{tr}\{\dot{\tilde{B}}^T P_y \tilde{B}\} + 2\varepsilon^{-1}\varsigma_y^T P_y \tilde{B}\hat{y},\\ L_{W3} &= 2\text{tr}\{\dot{\tilde{W}}_3^T K_3^{-1}\tilde{W}_3\} + 2\varepsilon^{-1}\varsigma_y^T P_y \tilde{W}_3 \Psi_3(V_3[x;y]) - 2\varepsilon^{-1}\varsigma_y^T P_y \tilde{W}_3 D_{\Psi_3} V_3[x;y],\\ L_{W4} &= 2\text{tr}\{\dot{\tilde{W}}_4^T K_4^{-1}\tilde{W}_4\} + 2\varepsilon^{-1}\varsigma_y^T P_y \tilde{W}_4 \Psi_4(V_4[x;y])u - 2\varepsilon^{-1}\varsigma_y^T P_y \tilde{W}_4 D_{\Psi_4} V_4[x;y],\\ L_{V3} &= 2\text{tr}\{\dot{\tilde{V}}_3^T L_3^{-1}\tilde{V}_3\} + 2\varepsilon^{-1}\varsigma_x^T P_x \tilde{W}_3 D_{\Psi_3}\tilde{V}_3[x;y],\\ L_{V4} &= 2\text{tr}\{\dot{\tilde{V}}_4^T L_4^{-1}\tilde{V}_4\} + 2\varepsilon^{-1}\varsigma_x^T P_x \tilde{W}_4 D_{\Psi_4}\tilde{V}_4[x;y].\end{aligned}$$

Using updating rules (3.21), one has

$$\begin{aligned}\dot{V}_x =& -\varsigma_x^T Q_x \varsigma_x + 2k_x\|\varsigma_x\|\text{tr}\{(W_1^* - \tilde{W}_1)^T\tilde{W}_1\} + 2k_x\|\varsigma_x\|\text{tr}\{(V_1^* - \tilde{V}_1)^T\tilde{V}_1\}\\ &+ 2k_x\|\varsigma_x\|\text{tr}\{(W_2^* - \tilde{W}_2)^T\tilde{W}_2\} + 2k_x\|\varsigma_x\|\text{tr}\{(V_2^* - \tilde{V}_2)^T\tilde{V}_2\} + 2\varsigma_x^T P_x\xi_x\\ =& -\varsigma_x^T Q_x \varsigma_x + 2k_x\|\varsigma_x\|\text{tr}\{(Z_1^* - \tilde{Z}_1)^T\tilde{Z}_1\} + 2k_x\|\varsigma_x\|\text{tr}\{(Z_2^* - \tilde{Z}_2)^T\tilde{Z}_2\} + 2\varsigma_x^T P_x\xi_x\ \ \dot{V}_y = \quad -\varepsilon^{-1}\varsigma_y^T Q_y\varsigma_y + 2\\ &+ 2k_y\|\varsigma_y\|\text{tr}\{(W_4^* - \tilde{W}_4)^T\tilde{W}_4\} + 2k_y\|\varsigma_y\|\text{tr}\{(V_4^* - \tilde{V}_4)^T\tilde{V}_4\} + 2\varepsilon^{-1}\varsigma_y^T P_y\xi_y\\ =& -\varepsilon^{-1}\varsigma_y^T Q_y\varsigma_y + 2k_y\|\varsigma_y\|\text{tr}\{(Z_3^* - \tilde{Z}_3)^T\tilde{Z}_3\} + 2k_y\|\varsigma_y\|\text{tr}\{(Z_4^* - \tilde{Z}_4)^T\tilde{Z}_4\} + 2\varepsilon^{-1}\varsigma_y^T P_y\xi_y.\end{aligned}$$

Since $\text{tr}\{(Z^* - \tilde{Z})^T\tilde{Z}\} = \left\langle Z^*, \tilde{Z}\right\rangle_F - \|\tilde{Z}\|^2 \le \|\tilde{Z}\|\|Z^*\| - \|\tilde{Z}\|^2$ and using (3.13), the following

inequality holds

$$\begin{aligned}\dot{V}_x \leq & -\lambda_{min}(Q_x)\|\varsigma_x\|^2 + 2k_x\|\varsigma_x\|\|\tilde{Z}_1\|(\bar{Z}_1 - \|\tilde{Z}_1\|) + 2k_x\|\varsigma_x\|\|\tilde{Z}_2\|(\bar{Z}_2 - \|\tilde{Z}_2\|) + 2\|\varsigma_x\|\|P_x\|\|\xi_x\| \\ = & -\|\varsigma_x\|L_x, \\ \dot{V}_y \leq & -\frac{1}{\varepsilon}\lambda_{min}(Q_y)\|\varsigma_y\|^2 + 2k_y\|\varsigma_y\|\|\tilde{Z}_3\|(\bar{Z}_3 - \|\tilde{Z}_3\|) + 2k_y\|\varsigma_y\|\|\tilde{Z}_4\|(\bar{Z}_4 - \|\tilde{Z}_4\|) + \frac{2}{\varepsilon}\|\varsigma_y\|\|P_y\|\|\xi_y\| \\ = & -\frac{1}{\varepsilon}\|\varsigma_y\|L_y,\end{aligned}$$

where

$$\begin{aligned}L_x = & \lambda_{min}(Q_x)\|\varsigma_x\| + 2k_x\|\tilde{Z}_1\|(\|\tilde{Z}_1\| - \bar{Z}_1) - 2\|P_x\|(C_{x1} + C_{x2}\|\tilde{Z}_1\|) + 2k_x\|\tilde{Z}_3\|(\|\tilde{Z}_3\| - \bar{Z}_3) \\ & - 2C_{x3}\|P_x\|\|\tilde{Z}_3\|, \\ L_y = & \lambda_{min}(Q_y)\|\varsigma_y\| + 2\varepsilon k_y\|\tilde{Z}_1\|(\|\tilde{Z}_1\| - \bar{Z}_1) - 2\|P_y\|(C_{y1} + C_{y2}\|\tilde{Z}_1\|) + 2\varepsilon k_y\|\tilde{Z}_3\|(\|\tilde{Z}_3\| - \bar{Z}_3) \\ & - 2C_{y3}\|P_y\|\|\tilde{Z}_3\|,\end{aligned}$$

Thus, $\dot{V}_I$ is negative as long as L_x, L_y are positive.

Define $C_{x4} = \bar{Z}_1 + \frac{\|P_x\|C_{x2}}{k_x}$, $C_{x5} = \bar{Z}_2 + \frac{\|P_x\|C_{x2}}{k_x}$, $C_{y4} = \bar{Z}_3 + \frac{\|P_y\|C_{y3}}{k_y}$, $C_{y5} = \bar{Z}_4 + \frac{\|P_y\|C_{y4}}{k_y}$, then L_x and L_y become:

$$\begin{aligned}L_x = & \lambda_{min}(Q_x)\|\varsigma_x\| + 2k_x\|\tilde{Z}_1\|(\|\tilde{Z}_1\| - C_{x4}) - 2\|P_x\|C_{x1} + 2k_x\|\tilde{Z}_2\|(\|\tilde{Z}_2\| - C_{x5}) \\ = & 2k_x(\|\tilde{Z}_1\| - \frac{C_{x4}}{2})^2 - \frac{k_xC_{x4}^2}{2} + \lambda_{min}(Q_x)\|\varsigma_x\| - 2\|P_x\|C_{x1} + 2k_x(\|\tilde{Z}_2\| - \frac{C_{x5}}{2})^2 - \frac{k_xC_{x5}^2}{2}, \\ L_y = & \lambda_{min}(Q_y)\|\varsigma_y\| + 2\varepsilon k_y\|\tilde{Z}_3\|(\|\tilde{Z}_3\| - C_{y4}) - 2\|P_y\|C_{y1} + 2\varepsilon k_y\|\tilde{Z}_4\|(\|\tilde{Z}_4\| - C_{y5}) \\ = & 2\varepsilon k_y(\|\tilde{Z}_3\| - \frac{C_{y4}}{2})^2 - \frac{\varepsilon k_yC_{y4}^2}{2} + \lambda_{min}(Q_y)\|\varsigma_y\| - 2\|P_y\|C_{y1} + 2\varepsilon k_y(\|\tilde{Z}_4\| - \frac{C_{y5}}{2})^2 - \frac{\varepsilon k_yC_{y5}^2}{2}.\end{aligned}$$

The above terms are guaranteed positive as long as either

$$\|\varsigma_x\| > \frac{2^{-1}(k_xC_{x4}^2 + k_xC_{x5}^2) + 2\|P_x\|C_{x1}}{\lambda_{min}(Q_x)} = b_x,$$

$$\|\varsigma_y\| > \frac{2^{-1}\varepsilon(k_y C_{y4}^2 + k_y C_{y5}^2) + 2\|P_y\|C_{y1}}{\lambda_{min}(Q_y)} = b_y,$$

or

$$\|\tilde{Z}_1\| > C_{x4}/2 + \sqrt{(C_{x4}^2 + C_{x5}^2)/4 + \|P_x\|C_{x1}/k_x},$$
$$\|\tilde{Z}_2\| > C_{x5}/2 + \sqrt{(C_{x4}^2 + C_{x5}^2)/4 + \|P_x\|C_{x1}/k_x},$$
$$\|\tilde{Z}_3\| > C_{y4}/2 + \sqrt{(C_{y4}^2 + C_{y5}^2)/4 + \varepsilon^{-1}\|P_y\|C_{y1}/k_y},$$
$$\|\tilde{Z}_4\| > C_{y5}/2 + \sqrt{(C_{y4}^2 + C_{y5}^2)/4 + \varepsilon^{-1}\|P_y\|C_{y1}/k_y},$$

then $\dot{V}_I < 0$ is true. According to standard Lyapunov theorem extension [89,99], this demonstrates the uniformly ultimately boundedness (UUB) of $\|\varsigma_x\|$, $\|\varsigma_y\|$, $\|\tilde{Z}_1\|$, $\|\tilde{Z}_2\|$, $\|\tilde{Z}_3\|$, $\|\tilde{Z}_4\|$. This implies ς_x, ς_y, W_i, V_i, $i = 1, \cdots, 4$, A, $B \in \mathcal{L}_\infty$.

Case II: $s_x = 0$ and $s_y = 0$. In this case, the learning process is stopped (all right-hand sides of the differential equations in (3.21) are equal to 0) and NN weights remain constants, then the identification error and weight matrices remain bounded, i.e., ς_x, ς_y, W_i, V_i, $i = 1, \cdots, 4$, A, $B \in \mathcal{L}_\infty$.

Case III: $s_x = 0$, $s_y = 1$ or $s_x = 1$, $s_y = 0$. When $s_x = 0$, $s_y = 1$, the learning process for A, W_1, W_2, V_1, V_2 are stopped, and remain constants. Because $\|\varsigma_x\| < b_x$, it can be concluded that ς_x, W_1, V_1, W_2, V_2, $A \in \mathcal{L}_\infty$.

Following the same analysis procedure, it can be proved that $\dot{V}_y$ is negative definite as long as either

$$\|\varsigma_y\| > \frac{2^{-1}\varepsilon(k_y C_{y4}^2 + k_y C_{y5}^2) + 2\|P_y\|C_{y1}}{\lambda_{min}(Q_y)} = b_y,$$

or

$$\|\tilde{Z}_3\| > C_{y4}/2 + \sqrt{(C_{y4}^2 + C_{y5}^2)/4 + \varepsilon^{-1}\|P_y\|C_{y1}/k_y},$$

$$\|\tilde{Z}_4\| > C_{y5}/2 + \sqrt{(C_{y4}^2 + C_{y5}^2)/4 + \varepsilon^{-1}\|P_y\|C_{y1}/k_y}.$$

This implies ς_y, W_3, V_3, W_4, V_4, $B \in \mathcal{L}_\infty$. Hence it can be concluded that when $\|\varsigma_x\| < b_x$ and $\|\varsigma_y\| > b_y$, one has ς_x, ς_y, W_i, V_i, $i = 1, \cdots, 4$, A, $B \in \mathcal{L}_\infty$.

Following similar analysis approach, it can be shown that ς_x, ς_y, W_i, V_i, $i = 1, \cdots, 4$, A, $B \in \mathcal{L}_\infty$. are also valid when $\|\varsigma_x\| > b_x$ and $\|\varsigma_y\| < b_y$. Theorem 3.1 is thus proved. □

Theorem 3.2. *In* (3.19) *and* (3.20)*, only the i^{th} element of i^{th} row in $\dot{W}_2$, and $(n+j)^{th}$ element of j^{th} row in $\dot{W}_4$ are needed to update W_2 and W_4. The stability property of the system will not be affected when other elements in $\dot{W}_2$ and $\dot{W}_4$ are neglected.*

Proof. Using the fact $\dot{\tilde{W}}_2 = -\dot{W}_2$, $\dot{\tilde{W}}_4 = -\dot{W}_4$, then the first terms of L_{W2}, L_{W4} in (3.25) equal

$$\begin{aligned}
2\text{tr}\{\dot{\tilde{W}}_2^T K_2^{-1}\tilde{W}_2\} &= -2\text{tr}\{\dot{W}_2^T K_2^{-1}\tilde{W}_2\},\\
2\text{tr}\{\dot{\tilde{W}}_4^T K_4^{-1}\tilde{W}_4\} &= -2\text{tr}\{\dot{W}_4^T K_4^{-1}\tilde{W}_4\}.
\end{aligned}$$

Because $\tilde{W}_2 = [\text{diag}(\tilde{w}_{21,\cdots,\tilde{w}_{2n}}), \mathbf{0}] \in \Re^{n\times(n+m)}$, $\tilde{W}_4 = [\mathbf{0}, \text{diag}(\tilde{w}_{41,\cdots,\tilde{w}_{4m}})] \in \Re^{m\times(n+m)}$, and K_2, K_4 are diagonal matrices, it can be obtained that

$$\begin{aligned}
\text{tr}\{\dot{\tilde{W}}_2^T K_2^{-1}\tilde{W}_2\} &= -\sum_{i=1}^{n} \dot{W}_{2i} k_{2i}^{-1}\tilde{w}_{2i},\\
\text{tr}\{\dot{\tilde{W}}_4^T K_4^{-1}\tilde{W}_4\} &= -\sum_{j=1}^{m} \dot{W}_{4j} k_{4j}^{-1}\tilde{w}_{4j},
\end{aligned}$$

where $\dot{W}_{2i}$, $\tilde{w}_{2i}$ are the i^{th} element of i^{th} row in $\dot{W}_2$ and $\tilde{W}_2$, respectively, $\dot{W}_{4j}$, $\tilde{w}_{4j}$ are the $(n+j)^{\text{th}}$ element of j^{th} row in $\dot{W}_4$ and $\tilde{W}_4$, respectively, k_{2i}, k_{4j} are the i^{th} and j^{th} diagonal elements in K_2 and K_4, respectively. This means only $\dot{W}_{2i}$, $\dot{W}_{4j}$ are involved in the Lyapunov analysis of the system. Hence, the stability property will not be affected if only $\dot{W}_{2i}$ and $\dot{W}_{4j}$ are used to update W_2 and W_4 in (3.19) and (3.20). Theorem 3.2 is thus proved. □

Remark 3.3. *Since the update gains in* (3.21) *can be chosen arbitrarily, the learning process of*

the recurrent neural network does not depend on the solution of Riccati equation (3.16)*. Hence* b_x*,* b_y *can be set as small as possible by choosing suitable* Q_x*,* Q_y*.*

Remark 3.4. *The first terms of* $\dot{W}_i$*,* $\dot{V}_i$*,* $i = 1, \cdots, 4$ *in the updating laws* (3.21) *are the back propagation of multilayer perceptrons and the last terms correspond to the e-modification [99] in standard use in adaptive control to guarantee bounded parameter estimates. The second terms of* $\dot{W}_i$ *are used to provide some corrections to the weight tuning for* W_i *and to assure the convergence properties of identification error.*

Remark 3.5. *When* ε *is very close to zero,* W_3*,* V_3*,* W_4*,* V_4 *present a high-gain behavior, causing the instability of identification algorithm. The Riccati equation* (3.16) *can be multiplied by any positive constant* ϱ*, i.e.,* $B^{*T}(\varrho P_y) + (\varrho P_y)B^* = -\varrho Q_y$*. It can be guaranteed that the learning gains of* W_3*,* V_3*,* W_4*,* V_4 *will not become too big if the* ϱ *is chosen as a very small number.*

A summary of the nonlinear identification via recurrent multilayer neural networks with two-time scales are listed as follows:

1) Construct the dynamic multilayer neural networks with two-time scales as (3.6) and choose suitable sigmoid functions ψ_i, $i = 1, \cdots, 4$. There is no preliminary off-line learning phase, so it is not necessary to provide the stable initial weights. A good choice is to select V_i, $i = 1, \cdots, 4$ arbitrarily, W_1, W_3 equal zero. W_2, W_4 are chosen to satisfy the condition $|w_{2i}| \geq \varpi$, $i = 1, \cdots, n$ and $|w_{4j}| \geq \varpi$, $j = 1, \cdots, m$.

2) Determine the constants in the update rules (3.21). b_x, b_y can be set to arbitrarily small by selecting suitable Q_x, Q_y. The learning gains K_1, K_3, L_i, $i = 1, \cdots, 4$ should be chosen as positive definite matrices, and K_2, K_4 should be chosen as diagonal positive definite matrices. k_x, k_y, k_A, k_B are positive constants.

3) On-line identification. The system states x, y can be obtained from the plant and the estimation of x, y can be obtained from (3.6). Then the weights of neural networks can be adjusted on-line using the identification errors.

3.4 Indirect Adaptive Control Based on Feedback Linearization

In this Section, the trajectory tracking problem based on the identification result from Section 3.3 will be considered. From Section 3.3 it is clear that a nonlinear singularly perturbed system can be modeled by the recurrent neural network as:

$$\begin{aligned} \dot{x} &= Ax + W_1\Psi_1(V_1[x;y]) + W_2\Psi_2(V_2[x;y])u + \delta'_x, \\ \varepsilon\dot{y} &= By + W_3\Psi_3(V_3[x;y]) + W_4\Psi_4(V_4[x;y])u + \delta'_y, \end{aligned} \tag{3.26}$$

where δ'_x and δ'_y are the modeling errors.

The control goal is to force the system states to track the desired signals, which are generated by a nonlinear reference model

$$\begin{aligned} \dot{x}_d &= f_r(x_d, y_d, t), \\ \varepsilon\dot{y}_d &= g_r(x_d, y_d, t). \end{aligned} \tag{3.27}$$

Define the tracking errors as:

$$\begin{aligned} E_x &= x - x_d, \\ E_y &= y - y_d. \end{aligned} \tag{3.28}$$

Then the error dynamic equations become:

$$\begin{aligned} \dot{E}_x &= Ax + W_1\Psi_1(V_1[x;y]) + W_2\Psi_2(V_2[x;y])u + \delta'_x - f_r, \\ \varepsilon\dot{E}_y &= By + W_3\Psi_3(V_3[x;y]) + W_4\Psi_4(V_4[x;y])u + \delta'_y - g_r. \end{aligned} \tag{3.29}$$

The control signal u consists of two parts:

$$u = u_L + u_\delta, \tag{3.30}$$

where u_L is a compensation for the known nonlinearity and u_δ is dedicated to deal with the model

errors, which can be left open if the model errors are zero or ignorable. Let u_L be

$$\begin{aligned} u_L &= \begin{bmatrix} W_2\Psi_2(V_2[x;y]) \\ \frac{1}{\varepsilon}W_4\Psi_4(V_4[x;y]) \end{bmatrix}^{-1} u'_L, \\ u'_L &= -\begin{bmatrix} Ax_d \\ \frac{1}{\varepsilon}By_d \end{bmatrix} - \begin{bmatrix} W_1\Psi_1(V_1[x;y]) \\ \frac{1}{\varepsilon}W_3\Psi_3(V_3[x;y]) \end{bmatrix} + \begin{bmatrix} f_r \\ \frac{1}{\varepsilon}g_r \end{bmatrix}. \end{aligned} \tag{3.31}$$

and rewrite (3.29) as

$$\begin{bmatrix} \dot{E}_x \\ \dot{E}_y \end{bmatrix} = \begin{bmatrix} Ax \\ \frac{1}{\varepsilon}By \end{bmatrix} - \begin{bmatrix} W_1\Psi_1(V_1[x;y]) \\ \frac{1}{\varepsilon}W_3\Psi_3(V_3[x;y]) \end{bmatrix} + \begin{bmatrix} W_2\Psi_2(V_2[x;y]) \\ \frac{1}{\varepsilon}W_4\Psi_4(V_4[x;y]) \end{bmatrix} u + \begin{bmatrix} \delta'_x \\ \frac{1}{\varepsilon}\delta'_y \end{bmatrix} - \begin{bmatrix} f_r \\ \frac{1}{\varepsilon}g_r \end{bmatrix}. \tag{3.32}$$

Then substituting (3.31) into (3.32) yields

$$\begin{bmatrix} \dot{E}_x \\ \dot{E}_y \end{bmatrix} = \begin{bmatrix} AE_x \\ \frac{1}{\varepsilon}BE_y \end{bmatrix} + \begin{bmatrix} W_2\Psi_2(V_2[x;y]) \\ \frac{1}{\varepsilon}W_4\Psi_4(V_4[x;y]) \end{bmatrix} u_\delta + \begin{bmatrix} \delta'_x \\ \frac{1}{\varepsilon}\delta'_y \end{bmatrix}. \tag{3.33}$$

The control signal u_δ is designed to compensate the unknown dynamic modeling errors. The sliding mode control method is applied to accomplish the task. Let u_δ be

$$u_\delta = \begin{bmatrix} W_2\Psi_2(V_2[x;y]) \\ \frac{1}{\varepsilon}W_4\Psi_4(V_4[x;y]) \end{bmatrix}^{-1} \begin{bmatrix} u'_{\delta x} \\ u'_{\delta y} \end{bmatrix}, \tag{3.34}$$

$$u'_{\delta x} = -AE_x - \eta_x[\text{sign}(E_{x1}), \cdots, \text{sign}(E_{xn})]^T, \tag{3.35}$$

$$u'_{\delta y} = -\frac{1}{\varepsilon}BE_y - \frac{1}{\varepsilon}\eta_y[\text{sign}(E_{y1}), \cdots, \text{sign}(E_{ym})]^T, \tag{3.36}$$

where E_{xi}, $i = 1, \cdots, n$, E_{yj}, $j = 1, \cdots, m$ are the i^{th} and j^{th} elements of E_x and E_y, respectively.

Then substituting (3.34) into (3.33) yields

$$\begin{aligned}\dot{E}_x &= -\eta_x[\text{sign}(E_{x1}), \cdots, \text{sign}(E_{xn})]^T + \delta'_x, \\ \dot{E}_y &= -\frac{1}{\varepsilon}\eta_y[\text{sign}(E_{y1}), \cdots, \text{sign}(E_{ym})]^T + \frac{1}{\varepsilon}\delta'_y.\end{aligned} \tag{3.37}$$

Theorem 3.3. *Consider the identification model given in* (3.26) *with reference model given in* (3.27)*. It can be proved that by using control strategies given in* (3.30)*,* (3.31) *and* (3.34)*, the following stability properties can be guaranteed:*

$$\lim_{t\to\infty} E_x = 0, \ \lim_{t\to\infty} E_y = 0.$$

Proof. Consider the following Lyapunov function candidate:

$$V_c = E_x^T E_x + E_y^T E_y. \tag{3.38}$$

Using (3.37), one can obtain the derivation of V_c as

$$\begin{aligned}\dot{V}_c =& 2E_x^T\dot{E}_x + 2E_y^T\dot{E}_y \\ =& 2E_x^T(-\eta_x[\text{sign}(E_{x1}), \cdots, \text{sign}(E_{xn})]^T + \delta'_x) + 2E_y^T(-\frac{\eta_y}{\varepsilon}[\text{sign}(E_{y1}), \cdots, \text{sign}(E_{ym})]^T + \frac{\delta'_y}{\varepsilon}) \\ =& -2\eta_x\|E_x\|_1 + 2E_x^T\delta'_x - \frac{2}{\varepsilon}\eta_y\|E_y\|_1 + \frac{2}{\varepsilon}E_y^T\delta'_y \\ \leq& -2\eta_x\|E_x\|_1 + 2\|E_x^T\|_1\|\delta'_x\|_1 - \frac{2}{\varepsilon}\eta_y\|E_y\|_1 + \frac{2}{\varepsilon}\|E_y^T\|_1\|\delta'_y\|_1 \\ =& -2(\eta_x - \|\delta'_x\|_1)\|E_x\|_1 - \frac{2}{\varepsilon}(\eta_y - \|\delta'_y\|_1)\|E_y\|_1,\end{aligned} \tag{3.39}$$

where $\|\cdot\|_1$ denotes the L_1-norm. If η_x, η_y are chosen to be $\eta_x > \bar{\delta}'_x$, $\eta_y > \bar{\delta}'_y$, where $\bar{\delta}'_x$, $\bar{\delta}'_y$ are upper boundaries of δ'_x and δ'_x. Then $\dot{V}_c < 0$ is true, which implies $\lim_{t\to\infty} E_x = 0$, $\lim_{t\to\infty} E_y = 0$. Theorem 3.3 is thus proved. □

Remark 3.6. *At the beginning of the identification process, the initial value of W_2, W_4 are chosen to satisfy the condition $|w_{2i}| > \varpi$, $|w_{4j}| > \varpi$. Then according to* (3.19) *and* (3.20)*, if $|w_{2i}| = \varpi$,*

$|w_{4j}| = \varpi$, *and* $\dot{w}_{2i}$*sign*$(w_{2i}) \geq 0$, $\dot{w}_{4j}$*sign*$(w_{4j}) \geq 0$ *are always true. Therefore,* $|w_{2i}| \geq \varpi$, $|w_{4j}| \geq \varpi$ *are always valid during the identification process [48]. Since* Ψ_2, Ψ_4 *are diagonal matrices and the existence of* Ψ_2^{-1}, Ψ_4^{-1} *are guaranteed by selecting suitable* $\alpha_{i,j}$ *in* (3.7)*,* $det(M)$ *defined in* (3.18) *is guaranteed to be non-zero and the potential singularity problem in the controller design is avoided, and the control signal in* (3.30) *is guaranteed bounded.*

3.5 Application

To demonstrate the effectiveness of the proposed identification and control scheme, the following nonlinear system is considered:

$$\begin{aligned} \dot{x} &= -5x + 3\text{sign}(y) + u_1, \\ \varepsilon\dot{y} &= -10y + 2\text{sign}(x) + u_2, \end{aligned} \tag{3.40}$$

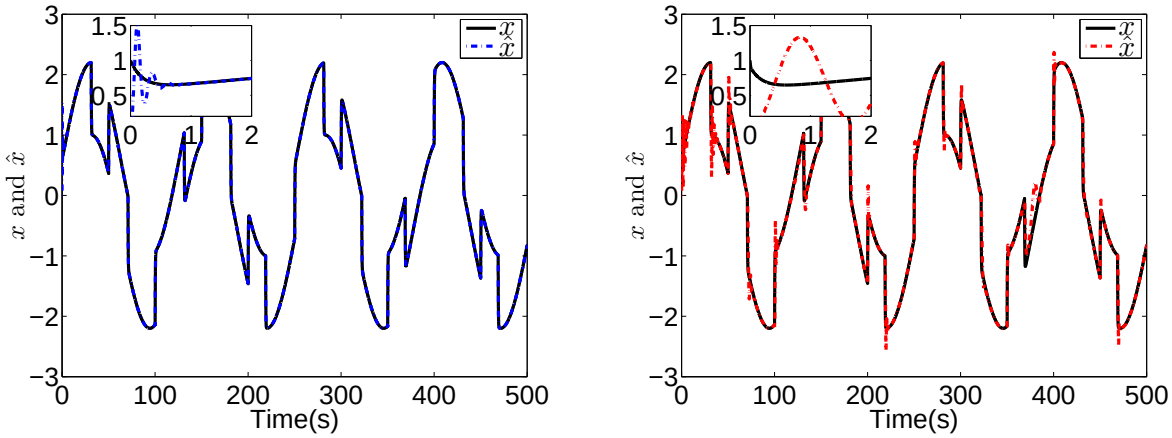
where $\varepsilon = 0.2$ and the initial states of the system are $x(0) = 1$, $y(0) = 0$. The nonlinear system given above, even simple, is interesting enough, since it has multiple isolated equilibriums. The simulation results using the identification and control scheme proposed in [26] is also provided for comparison purpose.

3.5.1 System Identification

To identify the nonlinear singularly perturbed system (3.40), the structure of the recurrent neural network is chosen as: $A \in \Re$, $B \in \Re$, $W_1,\ W_2,\ W_3,\ W_4 \in \Re^2$, $V_1,\ V_2,\ V_3,\ V_4 \in \Re^{2\times 2}$, $\psi_1,\ \psi_3 \in \Re^2$, $\psi_2,\ \psi_4 \in \Re^2$. The input signals are chosen the same as in [26], which are $u_1 = 8\text{sin}(0.05t)$ and u_2 is a saw-tooth function with the amplitude of 8 and the frequency of 0.02 Hz. The neural network parameters are chosen as: $k_A = -1000$, $k_B = -200$, $K_1 = -200\mathbf{I}$, $K_2 = -100\mathbf{I}$, $K_3 = K_4 = -20\mathbf{I}$ which are also the same as in [26]. The other parameters in the updating law (3.21) are chosen as $L_1 = L_2 = -0.05\mathbf{I}$, $L_3 = L_4 = -0.5\mathbf{I}$, $k_x = k_y = 0.05$, $\varpi = 0.05$. The sampling time in the simulation is 1 ms. The identification results are shown in the

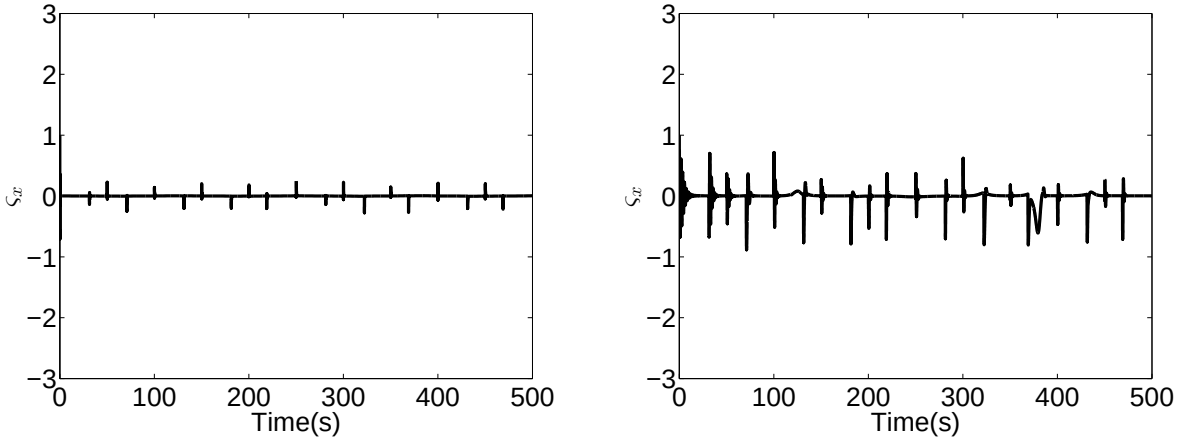
Figs. 3.2-3.5.

Remark 3.7. *The learning gains of the identification algorithm* (3.21) *are determined by* K_i,L_i, $i = 1, \cdots, 4$, k_x, k_y, k_A, k_B. *Generally speaking, larger learning gain will result in faster convergence, smaller identification errors with more oscillations. Hence, in practice, one can choose small learning gains at first, and increase them gradually until satisfactory identification results are achieved.*



(a) Proposed in Section 3.3. (b) Proposed in [26].

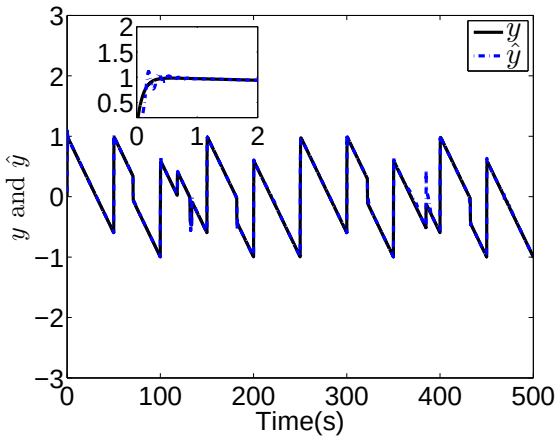
Figure 3.2: Identification results of x.



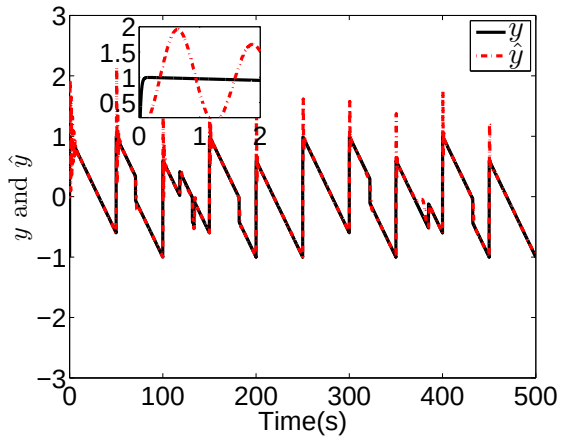
(a) Proposed in Section 3.3. (b) Proposed in [26].

Figure 3.3: Identification errors of x.

In Fig. 3.2(a), it can be seen that when the new identification algorithm proposed in Section 3.3 is used, the $\hat{x}$ will converge to x within 1s. However, when the identification proposed in [26]

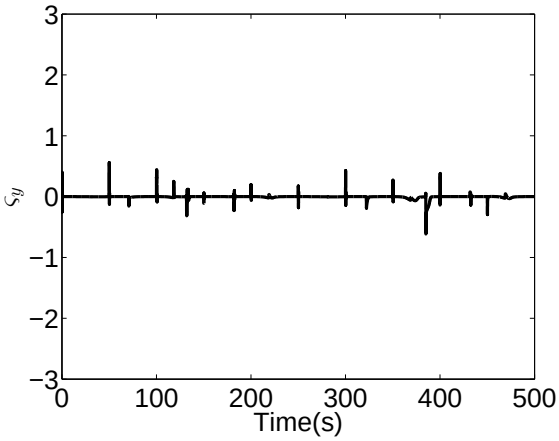


(a) Proposed in Section 3.3.

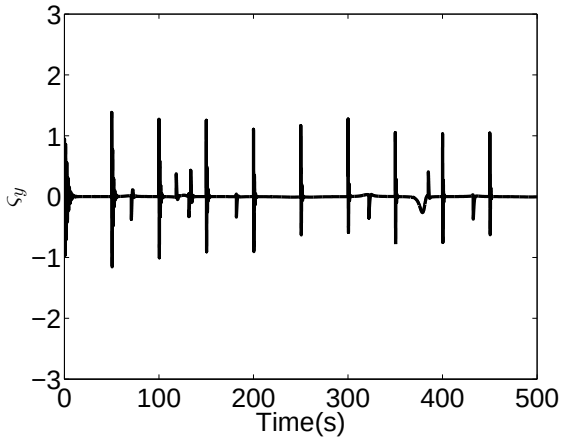


(b) Proposed in [26].

Figure 3.4: Identification results of y.



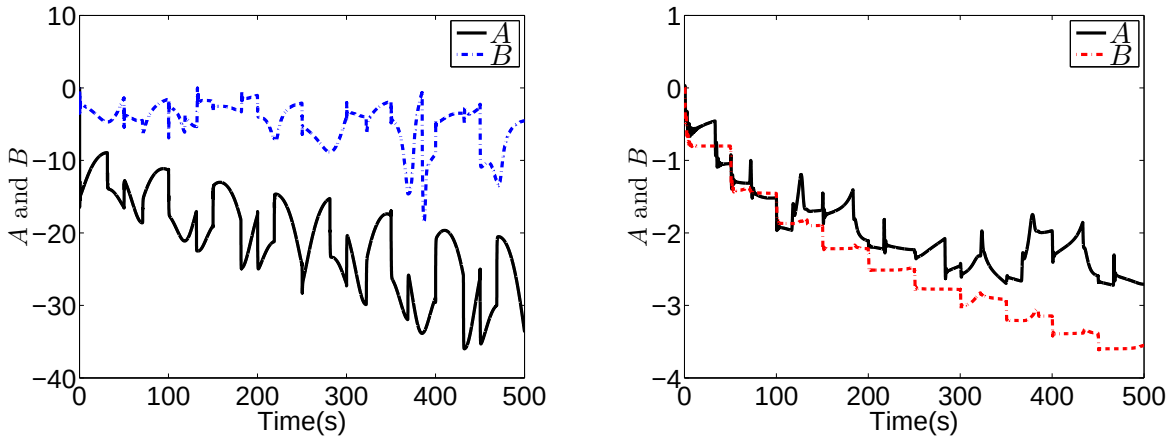
(a) Proposed in Section 3.3.



(b) Proposed in [26].

Figure 3.5: Identification errors of y.

is used, large difference between x and $\hat{x}$ can be observed, as shown in Fig. 3.2(b). In Fig. 3.3(a), it is obvious that when the new identification algorithm is used, the identification error ς_x is very small. While in Fig. 3.3(b), larger identification error is presented when the identification algorithm proposed in [26] is used. Similar phenomena can be noted in Fig. 3.4 and Fig. 3.5. Thus, it can be concluded that the state variables of the dynamic multi-time-scale NN follow those of the nonlinear system more accurately and quickly when the multilayer NN proposed in Section 3.3 is used compared to the results in [26]. The eigenvalues of the linear parameter matrix are shown in Fig. 3.6. The eigenvalues for both A and B are universally smaller than zero, which means they



(a) Proposed in Section 3.3.

(b) Proposed in [26].

Figure 3.6: Eigenvalues of A and B.

remain stable during the identification.

To show the performance of the proposed identification algorithm, the performance index-Root Mean Square (RMS) for the identification errors has been adopted, which is defined as:

$$RMS = \sqrt{\frac{\sum_{i=1}^{n} \varsigma^2(i)}{n}}, \tag{3.41}$$

where n is the number of the simulation steps, $\varsigma(i)$ is the identification error at i^{th} step. The RMS results are given in Table 3.1. From Table 3.1, it is clear that by adding the extra hidden layer, the RMS values of the identification errors ς_x and ς_y are largely reduced, which means the identification accuracy is greatly improved by using the multilayer neural network.

Table 3.1: RMS values of the identification errors

	ς_x	ς_y
RMS(proposed)	0.0175	0.0336
RMS([26])	0.1391	0.0944

3.5.2 System Control

The control goal is to force the outputs of the nonlinear singularly perturbed system (3.40) to follow the given reference model:

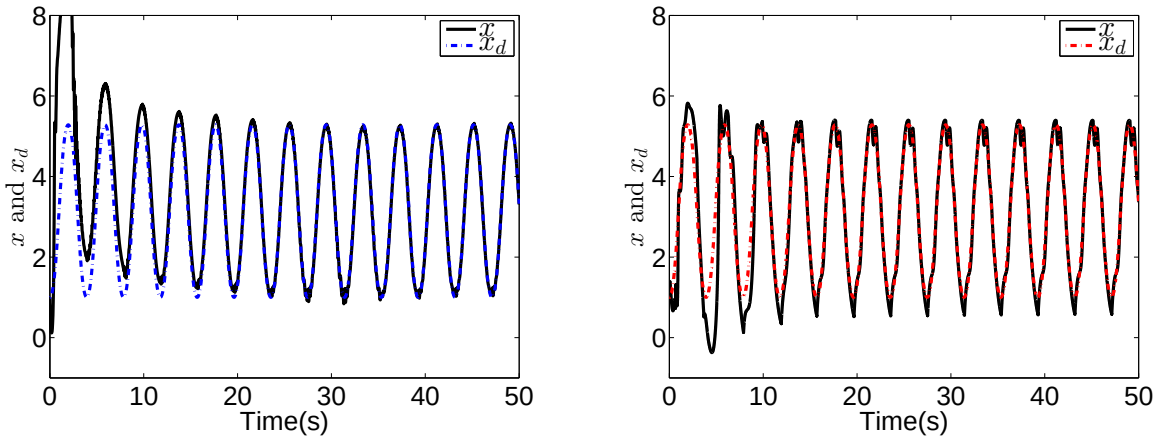
$$\begin{aligned} \dot{x}_d &= y_d, \\ \varepsilon \dot{y}_d &= \sin(x_d). \end{aligned} \tag{3.42}$$

The trajectory tracking results are shown in Figs. 3.7-3.10. In Fig. 3.7, it is clear that when the identification and control algorithm proposed in this chapter is used, the state x can track the reference signal x_d more closely. However, when the identification and control algorithm proposed in [26] is used, large spikes can be observed at the peaks and the lowest points of the sine wave. Also, in Fig. 3.8, it is clear that after 10s, the tracking error E_x is very small when the new identification and control algorithm is used. Nevertheless, larger fluctuation in E_x exists when the adaptive controller proposed in [26] is used.

The control results of y are much better when the new identification and control algorithm is used. In Fig. 3.9(a), the system state y can track the reference signal y_d precisely since the very beginning, but it takes about 15s for y to converge to y_d when the adaptive controller proposed in [26] is used, as shown in 3.9(b). When the new identification and control scheme is used, the tracking error is always small as depicted in 3.10(a), whereas E_y in 3.10(b) is still much larger after 20s.

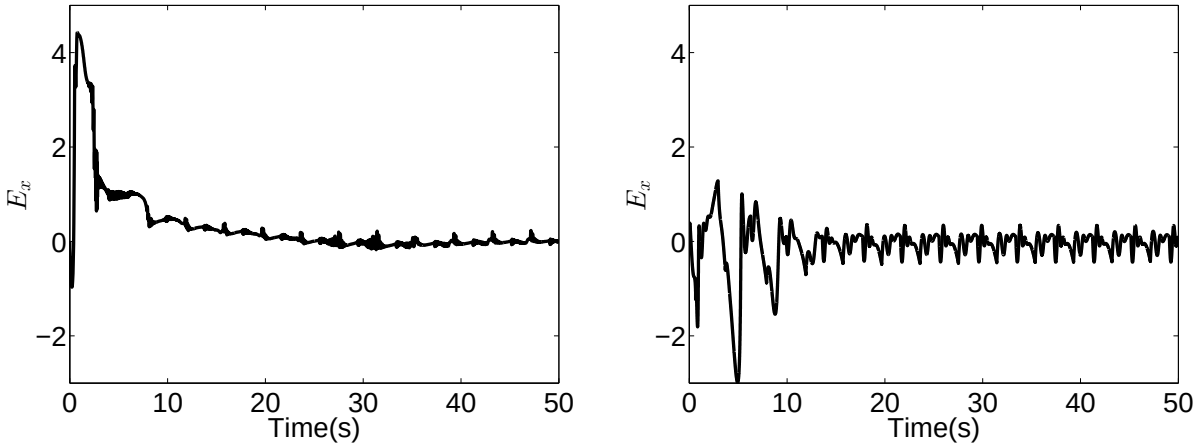
Hence, it can be concluded that by using the identification and control scheme proposed in this chapter, the closed-loop system can track the given reference signal more precisely. The tracking errors are greatly reduced when compared with the control results using the method proposed in [26]. Meanwhile, Fig. 3.8 and Fig. 3.10 show that it takes relatively more time for state x to track the reference signal than state y, because the small parameter ε accelerates both the identification and trajectory tracking process of state y.

To better illustrate the effectiveness of the proposed algorithms, the RMS values of the tracking errors E_x and E_y are also calculated, as presented in Table 3.2. The RMS values of all tracking errors demonstrate that the proposed identification and control algorithm has better performance



(a) Proposed in Section 3.4. (b) Proposed in [26].

Figure 3.7: Tracking results of x.



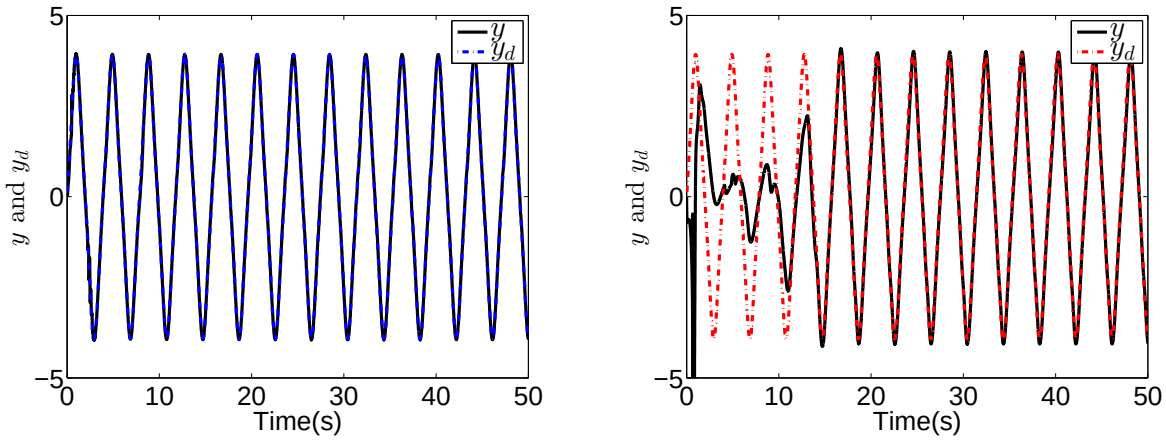
(a) Proposed in Section 3.4. (b) Proposed in [26].

Figure 3.8: Tracking errors of x.

than the method proposed [26], since smaller tracking errors can be achieved.

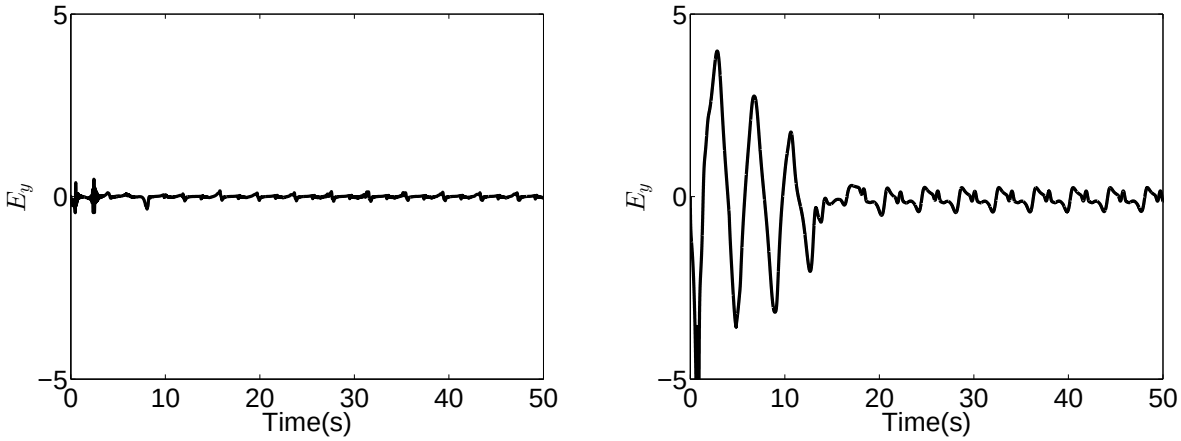
3.6 Conclusion

In this chapter, the adaptive nonlinear identification and trajectory tracking problem via recurrent multilayer neural network with different time-scales is discussed. The stability conditions for the on-line identification is determined by means of a Lyapunov-like analysis. Then a feedback



(a) Proposed in Section 3.4. (b) Proposed in [26].

Figure 3.9: Tracking results of y.



(a) Proposed in Section 3.4. (b) Proposed in [26].

Figure 3.10: Tracking errors of y.

controller is designed for trajectory tracking with consideration of the modeling error and disturbance. Simulation results show that the proposed identification and control algorithms containing both hidden layers and output layers have better performance than that in [26] which only includes a single layer of neurons.

Table 3.2: RMS values of the tracking errors

	E_x	E_y
RMS(proposed)	0.0187	0.0329
RMS([26])	0.2368	0.1596

Chapter 4

System Identification Based on Optimal Bounded Ellipsoid Algorithm

4.1 Introduction

So far, most training methods for NNs are gradient-like learning laws, such as backpropagation (BP) method. The main drawback of these methods is that the convergence speed is relatively slow. To accelerate the learning process, one solution is to use Extended Kalman Filter (EKE) in weights updating of NNs [28, 65, 66, 106, 107]. Regardless of its wide application, a main drawback of Kalman filter based training methods is that they require the modeling uncertainty of NNs to be Gaussian process in the theoretical analysis, which may not be true in real application.

Ellipsoid method for linear programing was first proposed in 1979 and widely studied thereafter [67, 108, 109]. Recently, weights updating using bounded ellipsoid method seems to be very effective in improving the learning speed for NNs. In [69], optimal bounded ellipsoid (OBE) algorithm was utilized to update the weights of the feedforward NN. Recurrent NN using OBE for weights updating was proposed in [68]. The stability property of NN based on OBE was discussed in [29]. It should be pointed out that all previous research were focused on discrete time single time scale NNs. The investigation on system identification based on OBE using continuous time

multi-time scales neural network was rare.

In this chapter, the OBE based system identification scheme using discrete time multi-time-scale neural network is studied first. Subsequently, the application of OBE to system identification problem using continuous time multi-time-scale NN is investigated. Lastly, a modified OBE based identification algorithm is proposed to further improve the performance of the identification scheme.

4.2 Identification of Discrete Systems Using Optimal Bounded Ellipsoid Algorithm

In this section, the optimal bounded ellipsoid algorithm based identification scheme for nonlinear discrete time singularly perturbed systems using multi-time-scale recurrent high order neural network is presented. The main difference between gradient descent (GD) based weight's updating laws and the OBE based weight's updating laws is that, in GD based methods, the learning gain of the weight's updating laws is fixed, while in the OBE based weight's updating laws, the learning gain can be adjusted adaptively. Because of the adaptively adjusted learning gain, the identification scheme proposed in this section can achieve faster convergence with higher accuracy. This will be demonstrated by simulation results.

4.2.1 Identification Algorithm

In this Section, the following singularly perturbed discrete-time nonlinear system is considered:

$$\begin{aligned} x(k+1) &= f(x(k), y(k), u(k)), \\ \varepsilon y(k+1) &= g(x(k), y(k), u(k), \varepsilon), \end{aligned} \tag{4.1}$$

where $x(k) \in \Re^n$, $y(k) \in \Re^m$ are the slow and fast state vectors, respectively, $u(k) \in \Re^p$ is the control input vector, $0 < \varepsilon < 1$ is a small parameter, $f(\cdot)$ and $g(\cdot)$ are unknown general nonlinear

smooth functions.

In order to model the discrete-time nonlinear system (4.1), the recurrent high order neural network (RHONN) with two time scales is used:

$$\begin{aligned} \hat{x}(k+1) &= A\hat{x}(k) + W_1\Psi_1(x,y) + W_2\Psi_2(x,y)u(k), \\ \varepsilon\hat{y}(k+1) &= \varepsilon B\hat{y}(k) + W_3\Psi_3(x,y) + W_4\Psi_4(x,y)u(k), \end{aligned} \tag{4.2}$$

where $\hat{x}(k) \in \Re^n$ and $\hat{y}(k) \in \Re^m$ are the slow and fast state vectors of the NN, $A \in \Re^{n\times n}$, $B \in \Re^{m\times m}$ are diagonal stable matrices. $W_1 \in \Re^{n\times q}$, $W_2 \in \Re^{n\times q}$, $W_3 \in \Re^{m\times q}$, $W_4 \in \Re^{m\times q}$ are the weight matrices of the NN, q is the number of neuron, the activation function vectors $\Psi_1(\cdot),\ \Psi_3(\cdot)$ are defined as [58]:

$$\begin{aligned} \Psi_i(\cdot) &= [\Psi_{i,1}, \Psi_{i,2}, \cdots, \Psi_{i,q}]^T \in \Re^{q\times 1}, i = 1,3, \\ \Psi_{i,c} &= \prod_{j\in J_c} [\psi_i(\cdot)]^{d_c(j)}, c = 1, \cdots, q, \end{aligned}$$

where J_c are the collections of l not ordered subsets of $1, 2, \cdots, n+m$ and $d_c(j)$ are non-negative integers, the activation functions $\psi_i(z)$ are chosen as:

$$\psi_i(z) = \frac{\alpha_{i,1}}{1 + e^{-\alpha_{i,2}z}} + \alpha_{i,3},\ i = 1,3. \tag{4.3}$$

The activation function matrices $\Psi_2(\cdot),\ \Psi_4(\cdot)$ are defined as:

$$\begin{aligned} \Psi_i(\cdot) &= \begin{bmatrix} \Psi_{i,1,1} & \cdots & \Psi_{i,1,p} \\ \vdots & \ddots & \vdots \\ \Psi_{i,q,1} & \cdots & \Psi_{i,q,p} \end{bmatrix} \in \Re^{q\times p}, \\ \Psi_{i,c,r} &= \prod_{j\in J_{c,r}} [\psi_i(\cdot)]^{d_{c,r}(j)}, c = 1, \cdots, q, r = 1, \cdots, p, \end{aligned}$$

where $J_{c,r}$ are the collections of l not ordered subsets of $1, 2, \cdots, n+m$ and $d_{c,r}(j)$ are non-negative integers. The activation functions $\psi_2(z),\ \psi_4(z)$ are chosen as:

$$\psi_i(z) = \frac{\alpha_{i,1}}{1 + e^{-\alpha_{i,2} z}} + \alpha_{i,3},\ i = 2, 4. \tag{4.4}$$

The parameters $\alpha_{i,j},\ i = 1, \cdots, 4,\ j = 1, \cdots, 3$ in (4.3) and (4.4) can be chosen a priori. The structure of the identification scheme is shown in Fig. 4.1.

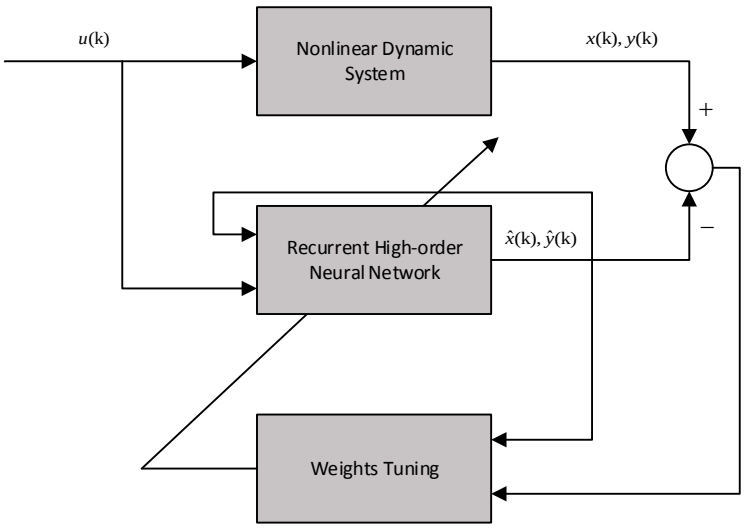


Figure 4.1: Structure of the identification scheme.

Assume that system (4.1) can be modeled by a nominal neural network with modeling error as:

$$\begin{aligned} x(k+1) &= Ax(k) + W_1^* \Psi_1(x, y) + W_2^* \Psi_2(x, y) u(k) + \zeta_x(k), \\ \varepsilon y(k+1) &= \varepsilon B y(k) + W_3^* \Psi_3(x, y) + W_4^* \Psi_4(x, y) u(k) + \zeta_y(k), \end{aligned} \tag{4.5}$$

where $W_1^* \in \Re^{n \times q},\ W_2^* \in \Re^{n \times q},\ W_3^* \in \Re^{m \times q},\ W_4^* \in \Re^{m \times q}$ are unknown optimal weights that minimize the modeling error $\zeta_x \in \Re^n$ and $\zeta_y \in \Re^m$.

Let $\zeta_{xi}(k)$ denote the i^{th} term of $\zeta_x(k),\ i = 1, \cdots, n$, and $\zeta_{yj}(k)$ denote the j^{th} term of $\zeta_y(k),\ j = 1, \cdots, m$, then the following assumption can be made.

Assumption 4.1. *The modeling error $\zeta_{xi}(k)$ and $\zeta_{yj}(k)$ are bounded by known bounds $\bar{\zeta}_{xi}$ and $\bar{\zeta}_{yj}$*

as:

$$|\zeta_{xi}(k)| \leq \bar{\zeta}_{xi},\ |\zeta_{yj}(k)| \leq \bar{\zeta}_{yj}.$$

Denote $\Theta^*_{x,k} = [W^*_1, W^*_2] = [\theta^*_{x1,k}; \theta^*_{x2,k}; \cdots; \theta^*_{xn,k}] \in \Re^{n\times 2q}$, $\theta^*_{xi,k} \in \Re^{1\times 2q}$, $i = 1, \cdots, n$, $\Theta^*_{y,k} = [W^*_3, W^*_4] = [\theta^*_{y1,k}; \theta^*_{y2,k}; \cdots; \theta^*_{ym,k}] \in \Re^{m\times 2q}$, $\theta^*_{yj,k} \in \Re^{1\times 2q}$, $j = 1, \cdots m$. $H_{x,k} = [\Psi_1(x,y); \Psi_2(x,y)u(k)] \in \Re^{2q\times 1}$, $H_{y,k} = [\Psi_3(x,y); \Psi_4(x,y)u(k)] \in \Re^{2q\times 1}$. Then model (4.5) can be written as:

$$\begin{aligned} x(k+1) &= Ax(k) + \Theta^*_{x,k}H_{x,k} + \zeta_x(k), \\ \varepsilon y(k+1) &= \varepsilon By(k) + \Theta^*_{y,k}H_{y,k} + \zeta_y(k). \end{aligned} \tag{4.6}$$

Similarly, model (4.2) can be represented as:

$$\begin{aligned} \hat{x}(k+1) &= A\hat{x}(k) + \Theta_{x,k}H_{x,k}, \\ \varepsilon\hat{y}(k+1) &= \varepsilon B\hat{y}(k) + \Theta_{y,k}H_{y,k}, \end{aligned} \tag{4.7}$$

where $\Theta_{x,k} = [W_1, W_2] = [\theta_{x1,k}; \theta_{x2,k}; \cdots; \theta_{xn,k}] \in \Re^{n\times 2q}$, $\Theta_{y,k} = [W_3, W_4] = [\theta_{y1,k}; \theta_{y2,k}; \cdots; \theta_{ym,k}] \in \Re^{m\times 2q}$.

Define the identification errors as:

$$\begin{aligned} \varsigma_x(k) &= x(k) - \hat{x}(k), \\ \varsigma_y(k) &= y(k) - \hat{y}(k). \end{aligned} \tag{4.8}$$

Define the auxiliary system outputs as:

$$\begin{aligned} \tau_x(k) &= \Theta^*_{x,k}H_{x,k} + \zeta_x(k) = x(k+1) - Ax(k), \\ \tau_y(k) &= \Theta^*_{y,k}H_{y,k} + \zeta_y(k) = \varepsilon y(k+1) - \varepsilon By(k). \end{aligned} \tag{4.9}$$

and the estimated auxiliary system output as:

$$\begin{aligned}\hat{\tau}_x(k) &= \Theta_{x,k}H_{x,k},\\ \hat{\tau}_y(k) &= \Theta_{y,k}H_{y,k}.\end{aligned} \tag{4.10}$$

Then the output errors are defined as:

$$\begin{aligned}e_x(k) =&\tau_x(k) - \hat{\tau}_x(k) = \tilde{\Theta}_{x,k}H_{x,k} + \zeta_x(k),\\ e_y(k) =&\tau_y(k) - \hat{\tau}_y(k) = \tilde{\Theta}_{x,k}H_{x,k} + \zeta_y(k),\end{aligned} \tag{4.11}$$

where $\tilde{\Theta}_{x,k} = \Theta^*_{x,k} - \Theta_{x,k}$, $\tilde{\Theta}_{y,k} = \Theta^*_{y,k} - \Theta_{y,k}$. Subtracting (4.7) from (4.6), it follows that:

$$\begin{aligned}\varsigma_x(k+1) &= A\varsigma_x(k) + e_x(k),\\ \varepsilon\varsigma_y(k+1) &= \varepsilon B\varsigma_y(k) + e_y(k).\end{aligned} \tag{4.12}$$

Remark 4.1. *The main purpose of this section is to train the recurrent high order neural network* (4.2) *such that the identification errors* $\varsigma_x(k)$, $\varsigma_y(k)$ *are bounded and minimized. This can be achieved by minimizing the auxiliary system output errors* $e_x(k)$ *and* $e_y(k)$.

From (4.12), it can be obtained that:

$$\begin{aligned}\varsigma_{xi}(k+1) &= a_i\varsigma_{xi}(k) + e_{xi}(k),\\ \varepsilon\varsigma_{yj}(k+1) &= \varepsilon b_j\varsigma_{yj}(k) + e_{yj}(k),\end{aligned} \tag{4.13}$$

where $\varsigma_{xi}(k),\ a_i,\ e_{xi}(k),\ i = 1,\cdots,n$ are the i^{th} element in $\varsigma_x(k),\ A,\ e_x(k)$, respectively, and $\varsigma_{yj}(k),\ b_j,\ e_{yj}(k),\ j = 1,\cdots,m$ are the m^{th} element in $\varsigma_y(k),\ B,\ e_y(k)$, respectively. Hence, it follows that:

$$\begin{aligned}\varsigma_{xi}(2) &= a_i\varsigma_{xi}(1) + e_{xi}(1),\\ \varsigma_{xi}(3) &= a_i\varsigma_{xi}(2) + e_{xi}(2) = a_i^2\varsigma_{xi}(1) + a_ie_{xi}(1) + e_{xi}(2),\end{aligned}$$

$$\vdots$$

$$\varsigma_{xi}(k) = a_i\varsigma_{xi}(k-1) + e_{xi}(k-1) = a_i^k\varsigma_{xi}(1) + \sum_{r=1}^{k-1} a_i^{k-1-r}e_{xi}(r).$$

Similarly, it is easy to show that:

$$\varepsilon\varsigma_{yj}(k) = \varepsilon b_j\varsigma_{yj}(k-1) + e_{yj}(k-1) = \varepsilon b_j^k\varsigma_{yj}(1) + \sum_{r=1}^{k-1} b_j^{k-1-r}e_{yj}(r).$$

Because $|a_i| < 1,\ |b_j| < 1,\ 0 < \varepsilon < 1$, it can be obtained that:

$$\varsigma_{xi}(k) < |\varsigma_{xi}(1)| + \sum_{r=1}^{k-1} |e_{xi}(r)|,\ \varepsilon\varsigma_{yj}(k) < \varepsilon|\varsigma_{yj}(1)| + \sum_{r=1}^{k-1} |e_{yj}(r)|.$$

It is noticed that $\varsigma_{xi}(1),\ \varsigma_{yj}(1)$ are constants. If the output errors $e_{xi}(k),\ e_{yj}(k)$ are minimized, the upper bounds of the identification errors $\varsigma_{xi}(k),\ \varsigma_{yj}(k)$ are minimized. Therefore, the analysis and identification algorithm design in this paper are based on minimizing the output errors $e_x(k)$ and $e_y(k)$.

Definition 4.1. *Define a real n-dimensional ellipsoid set, centered on r^* as:*

$$S(r^*, P) = \{r \in \Re^{1\times n} | (r - r^*)P^{-1}(r - r^*)^T \leq 1\}, \tag{4.14}$$

where $P \in \Re^{n\times n}$ is a positive definite symmetric matrix.

Based on (4.14), the discrete-time RHONN weight error ellipsoids $S_{x,k},\ S_{y,k}$ are defined as:

$$\begin{aligned} S_{x,k} &= \{\theta_{xi,k} | \tilde{\theta}_{xi,k}P_{x,k}^{-1}\tilde{\theta}_{xi,k}^T \leq 1\}, \\ S_{y,k} &= \{\theta_{yj,k} | \tilde{\theta}_{yj,k}P_{y,k}^{-1}\tilde{\theta}_{yj,k}^T \leq 1\}, \end{aligned} \tag{4.15}$$

where $\tilde{\theta}_{xi,k} = \theta_{xi,k}^* - \theta_{xi,k},\ i = 1, \cdots, n,\ \tilde{\theta}_{yj,k} = \theta_{yj,k}^* - \theta_{yj,k},\ j = 1, \cdots, m.\ P_{x,k},\ P_{y,k}$ are symmetric positive definite matrices.

Let $\tau_{xi}(k)$ be the i^{th} term of $\tau_x(k)$, and $\tau_{yj}(k)$ is the j^{th} term of $\tau_y(k)$. According to (4.9),

$\|\tau_{xi}(k) - \theta^*_{xi,k}H_{x,k}\|_2 = \|\zeta_{xi}(k)\|_2 \leq \bar{\zeta}_{xi}$, and $\|\tau_{yj}(k) - \theta^*_{yj,k}H_{y,k}\|_2 = \|\zeta_{yj}(k)\|_2 \leq \bar{\zeta}_{yj}$, then it is easy to know that $[\tau_{xi}(k) - \theta^*_{xi,k}H_{x,k}]$ belongs to an ellipsoid $E_{x,k}$, and $[\tau_{yj}(k) - \theta^*_{yj,k}H_{y,k}]$ belongs to an ellipsoid $E_{y,k}$, which are defined as:

$$\begin{aligned} E_{x,k}(\theta^*_{xi,k}H_{x,k}, \bar{\zeta}^2_{xi}) &= \{\tau_{xi}(k)|\frac{1}{\bar{\zeta}^2_{xi}}\|\tau_{xi}(k) - \theta^*_{xi,k}H_{x,k}\|_2^2 \leq 1\}, \\ E_{y,k}(\theta^*_{yj,k}H_{y,k}, \bar{\zeta}^2_{yj}) &= \{\tau_{yj}(k)|\frac{1}{\bar{\zeta}^2_{yj}}\|\tau_{yj}(k) - \theta^*_{yj,k}H_{y,k}\|_2^2 \leq 1\}, \end{aligned} \tag{4.16}$$

where $\|\cdot\|_2$ denotes the L_2-norm.

Assumption 4.2. *The initial weights $\theta_{xi,1}$ and $\theta_{yj,1}$ are assumed to be in the ellipsoid $S_{x,1}$ and $S_{y,1}$:*

$$\begin{aligned} S_{x,1} &= \{\theta_{xi,1}|\tilde{\theta}_{xi,1}P^{-1}_{x,1}\tilde{\theta}^T_{xi,1} \leq 1\}, \\ S_{y,1} &= \{\theta_{yj,1}|\tilde{\theta}_{yj,1}P^{-1}_{y,1}\tilde{\theta}^T_{yj,1} \leq 1\}. \end{aligned} \tag{4.17}$$

Assumption 4.2 can be satisfied by choosing the bounded initial weights $\theta_{xi,1}$, $\theta_{yj,1}$and suitable $P_{x,1}$, $P_{y,1}$.

According to (4.15) and (4.16), one gets to know that:

$$\begin{aligned} (1 - \mu_{xi,k})\tilde{\theta}_{xi,k}P^{-1}_{x,k}\tilde{\theta}^T_{xi,k} &\leq 1 - \mu_{xi,k}, \\ \frac{\mu_{xi,k}}{\bar{\zeta}^2_{xi}}\|\tau_{xi}(k) - \theta^*_{xi,k}H_{x,k}\|_2^2 &\leq \mu_{xi,k}. \end{aligned} \tag{4.18}$$

Adding up the two inequalities in (4.18) gives that

$$(1 - \mu_{xi,k})\tilde{\theta}_{xi,k}P^{-1}_{x,k}\tilde{\theta}^T_{xi,k} + \frac{\mu_{xi,k}}{\bar{\zeta}^2_{xi}}\|\tau_{xi}(k) - \theta^*_{xi,k}H_{x,k}\|_2^2 \leq 1. \tag{4.19}$$

Following the same analysis procedure, it can be obtained that:

$$(1 - \mu_{yj,k})\tilde{\theta}_{yj,k}P^{-1}_{y,k}\tilde{\theta}^T_{yj,k} + \frac{\mu_{yj,k}}{\bar{\zeta}^2_{yj}}\|\tau_{yj}(k) - \theta^*_{yj,k}H_{y,k}\|_2^2 \leq 1, \tag{4.20}$$

where $0 \leq \mu_{xi,k} < 1$ and $0 \leq \mu_{yj,k} < 1$, which will be defined later.

The following weight's updating laws are proposed to train the RHONN:

$$P_{x,k+1} = \frac{1}{1-\mu_{xi,k}}\Big[P_{x,k} - \frac{\mu_{xi,k}P_{x,k}H_{x,k}H_{x,k}^T P_{x,k}}{(1-\mu_{xi,k})\bar{\zeta}_{xi}^2 + \mu_{xi,k}H_{x,k}^T P_{x,k}H_{x,k}}\Big], \tag{4.21}$$

$$\theta_{xi,k+1}^T = \theta_{xi,k}^T + \frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^2}P_{x,k}H_{x,k}e_{xi}(k), \tag{4.22}$$

$$\mu_{xi,k} = \begin{cases} \frac{\lambda_{xi}\bar{\zeta}_{xi}^2}{1+H_{x,k}^T P_{x,k}H_{x,k}}, & \text{if } e_{xi}^2(k) \geq \frac{\bar{\zeta}_{xi}^2}{1-\lambda_{xi}} \ \&\ H_{x,k}^T P_{x,k}H_{x,k} \geq \bar{\zeta}_{xi}^2 \\ 0, & \text{if } e_{xi}^2(k) < \frac{\bar{\zeta}_{xi}^2}{1-\lambda_{xi}} \text{ or } H_{x,k}^T P_{x,k}H_{x,k} < \bar{\zeta}_{xi}^2 \end{cases}, \tag{4.23}$$

$$P_{y,k+1} = \frac{1}{1-\mu_{yj,k}}\Big[P_{y,k} - \frac{\mu_{yj,k}P_{y,k}H_{y,k}H_{y,k}^T P_{y,k}}{(1-\mu_{yj,k})\bar{\zeta}_{yj}^2 + \mu_{yj,k}H_{y,k}^T P_{y,k}H_{y,k}}\Big], \tag{4.24}$$

$$\theta_{yj,k+1}^T = \theta_{yj,k}^T + \frac{\mu_{yj,k}}{\bar{\zeta}_{yj}^2}P_{y,k}H_{y,k}e_{yj}(k), \tag{4.25}$$

$$\mu_{yj,k} = \begin{cases} \frac{\lambda_{yj}\bar{\zeta}_{yj}^2}{1+H_{y,k}^T P_{y,k}H_{y,k}}, & \text{if } e_{yj}^2(k) \geq \frac{\bar{\zeta}_{yj}^2}{1-\lambda_{yj}} \ \&\ H_{y,k}^T P_{y,k}H_{y,k} \geq \bar{\zeta}_{yj}^2 \\ 0, & \text{if } e_{yj}^2(k) < \frac{\bar{\zeta}_{yj}^2}{1-\lambda_{yj}} \text{ or } H_{y,k}^T P_{y,k}H_{y,k} < \bar{\zeta}_{yj}^2 \end{cases}, \tag{4.26}$$

where $\lambda_{xi},\ \lambda_{yj}$ are designed parameters such that $0 < \lambda_{xi} < 1,\ 0 < \lambda_{xi}\bar{\zeta}_{xi}^2 < 1,\ 0 < \lambda_{yj} < 1,\ 0 < \lambda_{yj}\bar{\zeta}_{yj}^2 < 1$.

Theorem 4.1. *Consider the model* (4.2) *for system* (4.1)*. If* $P_{x,1},\ P_{y,1}$ *are symmetric diagonal positive definite matrices,* $S_{x,k}$ *and* $S_{y,k}$ *belongs to the ellipsoid sets defined in* (4.15)*, then by using the weight's updating laws* (4.21)-(4.26)*, it can be guaranteed that* $P_x,\ P_y$ *and the identification errors* $\varsigma_x,\ \varsigma_y$ *will be bounded, and* $S_{x,k+1}$ *and* $S_{y,k+1}$ *will also be ellipsoids satisfying*

$$\begin{aligned} S_{x,k+1} &= \{\theta_{xi,k+1} | \tilde{\theta}_{xi,k+1}P_{x,k+1}^{-1}\tilde{\theta}_{xi,k+1}^T \leq 1\}, \\ S_{y,k+1} &= \{\theta_{yj,k+1} | \tilde{\theta}_{yj,k+1}P_{y,k+1}^{-1}\tilde{\theta}_{yj,k+1}^T \leq 1\}. \end{aligned} \tag{4.27}$$

Proof. Case I: $e_{xi}^2(k) \geq \frac{\bar{\zeta}_{xi}^2}{1-\lambda_{xi}},\ H_{x,k}^T P_{x,k}H_{x,k} \geq \bar{\zeta}_{xi}^2$ and $e_{yj}^2(k) \geq \frac{\bar{\zeta}_{yj}^2}{1-\lambda_{yj}},\ H_{y,k}^T P_{y,k}H_{y,k} \geq \bar{\zeta}_{yj}^2$. Using matrix inversion lemma [29]:

$$(I_1 - I_2 I_4^{-1} I_3)^{-1} = I_1^{-1} + I_1^{-1}I_2(I_4 - I_3 I_1^{-1} I_2)^{-1} I_3 I_1^{-1},$$

where $I_i,\ i = 1, \cdots, 4$ are matrices with proper dimensions, it can be obtained that:

$$\begin{aligned} P_{x,k+1}^{-1} &= (1-\mu_{xi,k})\Big[P_{x,k} - \frac{\mu_{xi,k}P_{x,k}H_{x,k}H_{x,k}^T P_{x,k}}{(1-\mu_{xi,k})\bar{\zeta}_{xi}^2 + \mu_{xi,k}H_{x,k}^T P_{x,k}H_{x,k}}\Big]^{-1} \\ &= (1-\mu_{xi,k})P_{x,k}^{-1} + \frac{\mu_{xi,k}}{\bar{\zeta}_{xi}}H_{x,k}H_{x,k}^T. \end{aligned} \tag{4.28}$$

Since $H_{x,k}^T P_{x,k}H_{x,k} > 0$ is a scalar, it is easy to show that $0 \leq \mu_{xi,k} < 1$. Also, it can be verified that $H_{x,k}H_{x,k}^T$ is semi-positive definite. Hence, according to (4.28), if $P_{x,k} > 0$, then $P_{x,k+1} > 0$ is valid. Since $P_{x,1} > 0$ is given, then it can be obtained that $P_{x,k+1} > 0$.

Following the same approach, it is easy to obtain that:

$$P_{y,k+1}^{-1} = (1-\mu_{yj,k})P_{y,k}^{-1} + \frac{\mu_{yj,k}}{\bar{\zeta}_{yj}}H_{y,k}H_{y,k}^T. \tag{4.29}$$

Let $z \in \Re^n$ be an arbitrary vector, and $V_{x,k+1}$ be defined as:

$$V_{x,k+1} = tr\{z^T P_{x,k+1} z\}. \tag{4.30}$$

Hence, it follows that:

$$\begin{aligned} \Delta V_x &= V_{x,k+1} - V_{x,k} \\ &= tr\{z^T\Big[\frac{1}{1-\mu_{xi,k}}\Big[P_{x,k} - \frac{\mu_{xi,k}P_{x,k}H_{x,k}H_{x,k}^T P_{x,k}}{(1-\mu_{xi,k})\bar{\zeta}_{xi}^2 + \mu_{xi,k}H_{x,k}^T P_{x,k}H_{x,k}}\Big]\Big]z\} - tr\{z^T P_{x,k} z\} \\ &= \frac{\mu_{xi,k}(1-\mu_{xi,k})[\bar{\zeta}_{xi}^2 - H_{x,k}^T P_{x,k}H_{x,k}]tr\{z^T P_{x,k} z\}}{(1-\mu_{xi,k})[(1-\mu_{xi,k})\bar{\zeta}_{xi}^2 + \mu_{xi,k}H_{x,k}^T P_{x,k}H_{x,k}]}. \end{aligned} \tag{4.31}$$

Since $0 \leq \mu_{xi,k} < 1$, $H_{x,k}^T P_{x,k}H_{x,k} \geq \bar{\zeta}_{xi}^2$, it is obvious that $\Delta V_x \leq 0$. This indicates that P_x is bounded. Following the same procedure, it is easy to show that P_y is also bounded.

Using (4.21)-(4.23), $\tilde{\theta}_{xi,k+1}P_{x,k+1}^{-1}\tilde{\theta}_{xi,k+1}^T$ can be calculated as:

$$\tilde{\theta}_{xi,k+1}P_{x,k+1}^{-1}\tilde{\theta}_{xi,k+1}^T = (1-\mu_{xi,k})\tilde{\theta}_{xi,k}P_{x,k+1}^{-1}\tilde{\theta}_{xi,k}^T - 2\frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^2}\tilde{\theta}_{xi,k}H_{x,k}e_{xi}(k)$$

$$+ \frac{\mu_{xi,k}^2}{\bar{\zeta}_{xi}^4} H_{x,k}^T P_{x,k+1} H_{x,k} e_{xi}^2(k). \tag{4.32}$$

Substituting (4.28) into (4.32) yields:

$$\begin{aligned}\tilde{\theta}_{xi,k+1} P_{x,k+1}^{-1} \tilde{\theta}_{xi,k+1}^T =& (1-\mu_{xi,k}) \tilde{\theta}_{xi,k} P_{x,k}^{-1} \tilde{\theta}_{xi,k}^T + \frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^2} \tilde{\theta}_{xi,k} H_{x,k} H_{x,k}^T \tilde{\theta}_{xi,k}^T \\ & - 2\frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^2} \tilde{\theta}_{xi,k} H_{x,k} e_{xi}(k) + \frac{\mu_{xi,k}^2}{\bar{\zeta}_{xi}^4} H_{x,k}^T P_{x,k+1} H_{x,k} e_{xi}^2(k). \end{aligned} \tag{4.33}$$

From (4.19), it is known that

$$(1-\mu_{xi,k}) \tilde{\theta}_{xi,k} P_{x,k}^{-1} \tilde{\theta}_{xi,k}^T \leq 1 - \frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^2} \|\tau_{xi}(k) - \theta_{xi,k}^* H_{x,k}\|_2^2. \tag{4.34}$$

Substituting (4.34) into (4.33), one obtains that

$$\begin{aligned}\tilde{\theta}_{xi,k+1} P_{x,k+1}^{-1} \tilde{\theta}_{xi,k+1}^T \leq & 1 - \frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^2} \|\tau_{xi}(k) - \theta_{xi,k}^* H_{x,k}\|_2^2 + \frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^2} \tilde{\theta}_{xi,k} H_{x,k} H_{x,k}^T \tilde{\theta}_{xi,k}^T \\ & - 2\frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^2} \tilde{\theta}_{xi,k} H_{x,k} e_{xi}(k) + \frac{\mu_{xi,k}^2}{\bar{\zeta}_{xi}^4} H_{x,k}^T P_{x,k+1} H_{x,k} e_{xi}^2(k) \\ = & 1 + \frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^2} [-\|\tau_{xi}(k) - \theta_{xi,k}^* H_{x,k}\|_2^2 + \tilde{\theta}_{xi,k} H_{x,k} H_{x,k}^T \tilde{\theta}_{xi,k}^T \\ & - 2\tilde{\theta}_{xi,k} H_{x,k} e_{xi}(k)] + \frac{\mu_{xi,k}^2}{\bar{\zeta}_{xi}^4} H_{x,k}^T P_{x,k+1} H_{x,k} e_{xi}^2(k). \end{aligned} \tag{4.35}$$

Since $e_{xi}(k) = \tau_{xi}(k) - \theta_{xi,k} H_{x,k}$, and $\tilde{\theta}_{xi,k} = \theta_{xi,k}^* - \theta_{xi,k}$, the equations in $[\cdot]$ in (4.35) can be rewritten as:

$$\begin{aligned} & - \|\tau_{xi}(k) - \theta_{xi,k}^* H_{x,k}\|_2^2 + \tilde{\theta}_{xi,k} H_{x,k} H_{x,k}^T \tilde{\theta}_{xi,k}^T - 2\tilde{\theta}_{xi,k} H_{x,k} e_{xi}(k) \\ = & - \|\tau_{xi}(k) - \theta_{xi,k}^* H_{x,k}\|_2^2 + (\theta_{xi,k}^* - \theta_{xi,k}) H_{x,k} H_{x,k}^T (\theta_{xi,k}^* - \theta_{xi,k})^T \\ & - 2(\theta_{xi,k}^* - \theta_{xi,k}) H_{x,k} (\tau_{xi}(k) - \theta_{xi,k} H_{x,k}) \\ = & - [\tau_{xi}^2(k) - 2\theta_{xi,k} H_{x,k} \tau_{xi}(k) + \theta_{xi,k} H_{x,k} H_{x,k}^T \theta_{xi,k}^T] = -e_{xi}^2(k). \end{aligned} \tag{4.36}$$

Hence, equation (4.35) becomes

$$\tilde{\theta}_{xi,k+1}P_{x,k+1}^{-1}\tilde{\theta}_{xi,k+1}^{T} \leq 1 - \frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^{2}}e_{xi}^{2}(k) + \frac{\mu_{xi,k}^{2}}{\bar{\zeta}_{xi}^{4}}H_{x,k}^{T}P_{x,k+1}H_{x,k}e_{xi}^{2}(k). \tag{4.37}$$

According to (4.23), $\mu_{xi,k} = \frac{\lambda_{xi}\bar{\zeta}_{xi}^{2}}{1+H_{x,k}^{T}P_{x,k}H_{x,k}}$ holds in current case. Hence, (4.37) can be written as:

$$\tilde{\theta}_{xi,k+1}P_{x,k+1}^{-1}\tilde{\theta}_{xi,k+1}^{T} \leq 1 - \frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^{2}}e_{xi}^{2}(k) + \frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^{2}}\frac{\lambda_{xi}H_{x,k}^{T}P_{x,k+1}H_{x,k}e_{xi}^{2}(k)}{1+H_{x,k}^{T}P_{x,k}H_{x,k}}. \tag{4.38}$$

Because $\Delta V_{x,k} < 0$, it is easy to know that $H_{x,k}^{T}P_{x,k+1}H_{x,k} \leq H_{x,k}^{T}P_{x,k}H_{x,k}$. Thus it can be obtained that $\frac{H_{x,k}^{T}P_{x,k+1}H_{x,k}}{1+H_{x,k}^{T}P_{x,k}H_{x,k}} < 1$. Also, it is noticed that $0 < \lambda_{xi} < 1,\ 0 < \mu_{xi,k} < 1$, then one has:

$$\tilde{\theta}_{xi,k+1}P_{x,k+1}^{-1}\tilde{\theta}_{xi,k+1}^{T} \leq 1 - \frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^{2}}e_{xi}^{2}(k) + \frac{\mu_{xi,k}\lambda_{xi}}{\bar{\zeta}_{xi}^{2}}e_{xi}^{2}(k) = 1 - \frac{\mu_{xi,k}}{\bar{\zeta}_{xi}^{2}}(1-\lambda_{xi})e_{xi}^{2}(k) \leq 1 \tag{4.39}$$

Following the same approach, one obtains that

$$\tilde{\theta}_{yj,k+1}P_{y,k+1}^{-1}\tilde{\theta}_{yj,k+1}^{T} \leq 1 - \frac{\mu_{yj,k}}{\bar{\zeta}_{yj}^{2}}(1-\lambda_{yj})e_{yj}^{2}(k) \leq 1. \tag{4.40}$$

Case 2: $e_{xi}^{2}(k) < \frac{\bar{\zeta}_{xi}^{2}}{1-\lambda_{xi}}$, or $H_{x,k}^{T}P_{x,k}H_{x,k} < \bar{\zeta}_{xi}^{2}$, or $e_{yj}^{2}(k) < \frac{\bar{\zeta}_{yj}^{2}}{1-\lambda_{yj}}$, or $H_{y,k}^{T}P_{y,k}H_{y,k} \geq \bar{\zeta}_{yj}^{2}$.

In this case, according to (4.23) or (4.26), $\mu_{xi,k} = 0$ or $\mu_{yj,k} = 0$. Then $P_{x,k+1} = P_{x,k},\ \theta_{xi,k+1} = \theta_{xi,k}$, or $P_{xy,k+1} = P_{y,k},\ \theta_{yj,k+1} = \theta_{yj,k}$. Therefore, it follows:

$$\tilde{\theta}_{xi,k+1}P_{x,k+1}^{-1}\tilde{\theta}_{xi,k+1}^{T} = \tilde{\theta}_{xi,k}P_{x,k}^{-1}\tilde{\theta}_{xi,k}^{T} \leq 1, \tag{4.41}$$

or

$$\tilde{\theta}_{yj,k+1}P_{y,k+1}^{-1}\tilde{\theta}_{yj,k+1}^{T} = \tilde{\theta}_{yj,k}P_{y,k}^{-1}\tilde{\theta}_{yj,k}^{T} \leq 1. \tag{4.42}$$

Since P_x and P_y are bounded, then (4.41) and (4.42) indicate that θ_{xi} and θ_{yj} are also bounded.

Because $\zeta_{xi}(k)$ and $\zeta_{yj}(k)$ are bounded, then according to (4.11), e_{xi} and e_{yj} are bounded. This means the identification errors ς_x and ς_y are also bounded. Theorem 4.1 is thus proved. □

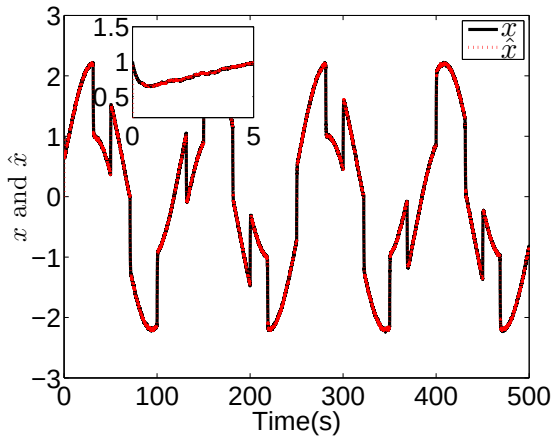
4.2.2 Simulation

In order to illustrate the effectiveness of the identification scheme proposed for the discrete time systems, the following system will be used:

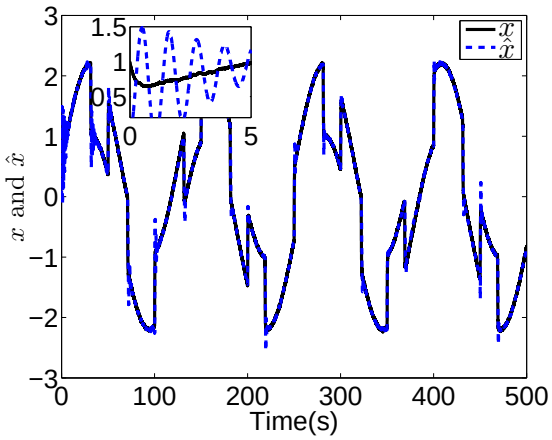
$$x(k+1) = 0.995x(k) + 0.003\text{sign}(y(k)) + 0.001u_1(k) + r_1,$$
$$\varepsilon y(k+1) = \varepsilon y(k) - 0.01y(k) + 0.002\text{sign}(x(k)) + 0.001u_2(k) + r_2,$$

where $\varepsilon = 0.2$ is a known parameter. The input signals are chosen as $u_1(k) = 8\text{sin}(5 \times 10^{-5}k)$ and $u_2(k) = 8 - 16(2 \times 10^{-5}k - \llcorner 2 \times 10^{-5}k \lrcorner)$ ($\llcorner \cdot \lrcorner$ is the floor function). r_1, r_2 are pseudorandom noise with mean of 0 and standard deviation of 0.001. The above nonlinear system is the discrete form of the system used in [16]. The neural network given in (4.2) is used to identify this nonlinear multi-time scale system. 8 neurons are used in the neural network, and $W_1 \in \Re^8$, $W_2 \in \Re^8$, $W_3 \in \Re^8$, $W_4 \in \Re^8$. The activation functions are chosen as $\psi_1(z) = \frac{10}{1+e^{-0.2z}} - 0.5$, $\psi_2(z) = \frac{2}{1+e^{-2z}} - 0.5$, $\psi_3(z) = \frac{8}{1+e^{-0.1z}} - 0.5$, $\psi_4(z) = \frac{4}{1+e^{-0.1z}} - 0.5$, $A = 0.01$, $B = 0.01$, $\lambda_{xi} = \lambda_{yj} = 0.5$, $\bar{\zeta}_{xi} = \bar{\zeta}_{yj} = 0.001$, $P_x(1) = P_y(1) = \text{diag}([100, \cdots, 100]) \in \Re^{16\times 16}$. The results of the identification method proposed in Section 4.2 and the identification method proposed in [16] are presented in Figs. 4.2-4.6.

In Figs. 4.2-4.5, the sub-figures on the left are the results of the method proposed in Section 4.2, and the sub-figures on the right are the results of the method proposed in [16]. In Fig. 4.2(a), x and $\hat{x}$ overlap each other since the very beginning. Meanwhile, in Fig. 4.3(a), the identification error ς_x is almost 0 since the beginning of the identification process, which means the identification accuracy of the identification algorithm proposed in this section is very high, and the convergence speed is very fast. However, when the identification algorithm proposed in [16] is used, lots of oscillations on $\hat{x}$ can be observed at the beginning of the identification process, as depicted in

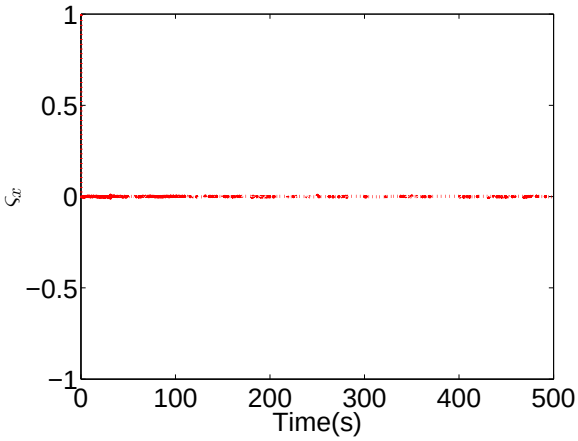


(a) Proposed in Section 4.2.

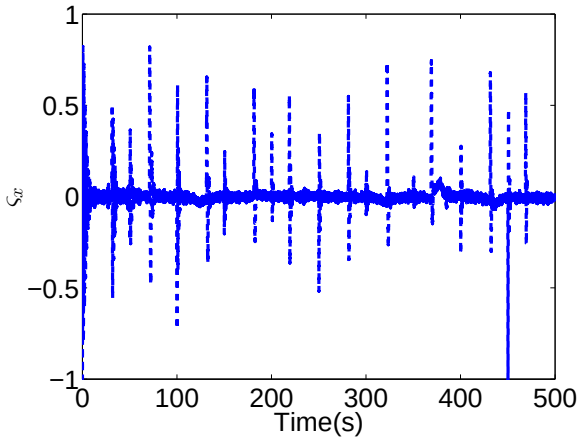


(b) Proposed in [16].

Figure 4.2: Identification results of x.



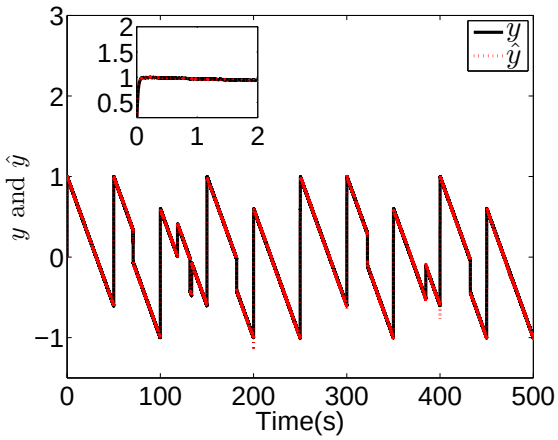
(a) Proposed in Section 4.2.



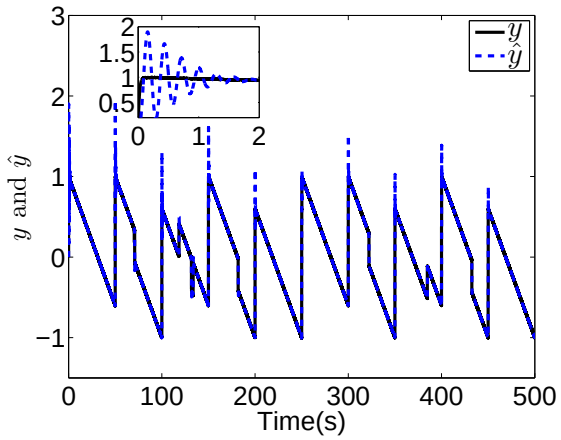
(b) Proposed in [16].

Figure 4.3: Identification errors of x.

Fig. 4.2(b). Also, the identification error ς_x has large spikes, as shown in Fig. 4.3(b). Similar phenomena exist in Fig. 4.4 and Fig. 4.5. Hence, it can be concluded that the identification results of x and y are much more accurate when using method proposed in Section 4.2. The identification errors are greatly reduced compared with the results using the method proposed in [16]. Also, when using the method proposed in Section 4.2, the NN converges much faster than using the identification algorithm proposed in [16]. This is due to that fact that the learning gain of the weight's updating law in the method proposed in Section 4.2 can vary adaptively. However, in [16] and in many other widely used learning algorithms, the learning gain is fixed. The convergence of

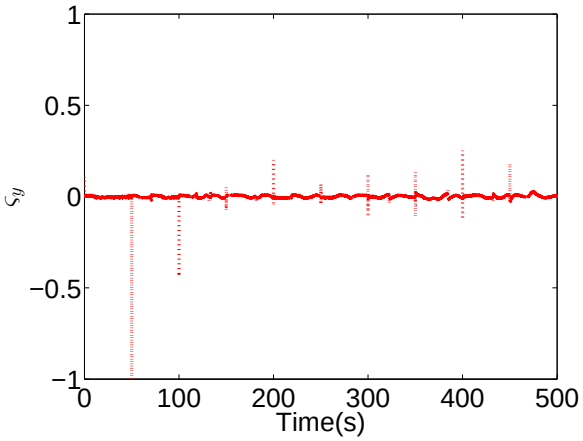


(a) Proposed in Section 4.2.

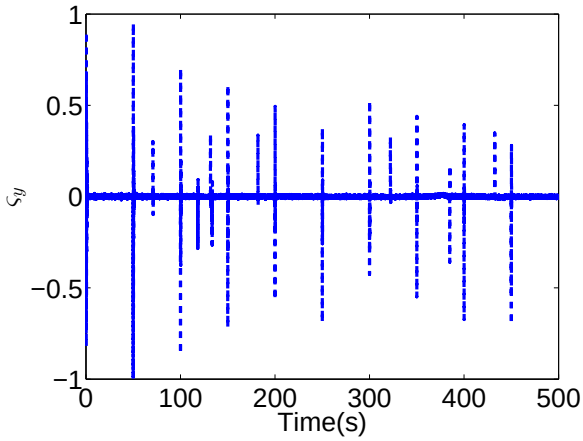


(b) Proposed in [16].

Figure 4.4: Identification results of y.



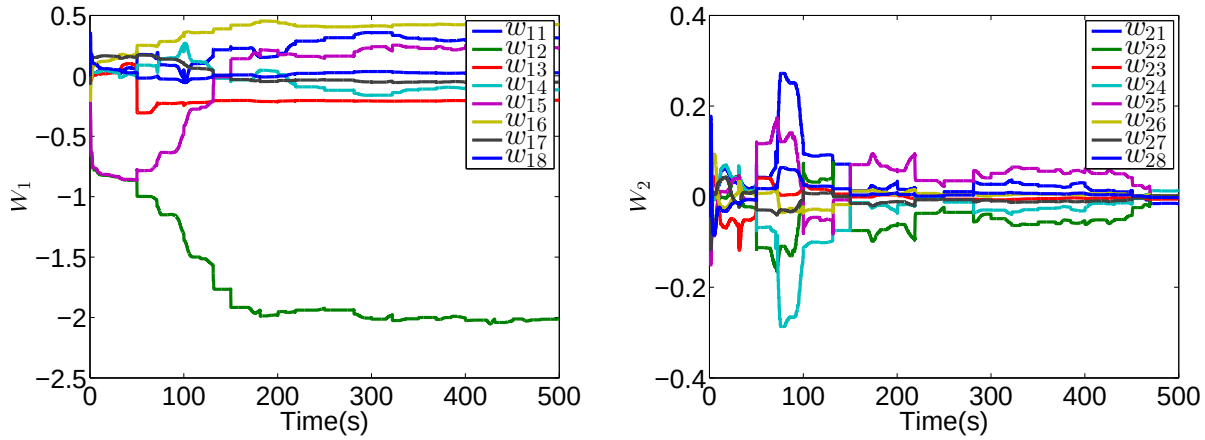
(a) Proposed in Section 4.2.



(b) Proposed in [16].

Figure 4.5: Identification errors of y.

the weight vectors W_1, W_2 are shown in Fig. 4.6. In Fig. 4.6, the w_{1i}, w_{2i} are the ith elements in W_1 and W_2, respectively. Due to the space limitation, the convergence of W_3 and W_4 are not presented here.



(a) W_1. (b) W_2.

Figure 4.6: Neural network weights.

4.3 Identification of Continuous Systems Using Optimal Bounded Ellipsoid Algorithm

In this section, the optimal bounded ellipsoid algorithm, which was designed for the discrete time systems, is extended to the continuous systems. The optimal bounded ellipsoid algorithm based identification scheme for continuous nonlinear singularly perturbed systems is established. In this novel identification scheme, the learning gain of the weight's updating laws can also be adaptively adjusted during the identification process. Thus this method can also achieve faster convergence with smaller identification errors. The effectiveness is also demonstrated by simulations.

4.3.1 Identification Algorithm

Considering the following nonlinear singularly perturbed system:

$$\begin{aligned} \dot{x} &= f(x, y, u), \\ \varepsilon \dot{y} &= g(x, y, u, \varepsilon), \end{aligned} \tag{4.43}$$

where $x \in \Re^n$, $g \in \Re^m$ are slow and fast state variables, respectively. $u \in \Re^p$ is the control input signal vector, $0 < \varepsilon < 1$ is a small parameter. $f \in C^\infty$ and $g \in C^\infty$ are unknown general nonlinear smooth functions.

In this paper, the following recurrent high-order NN with two time scales is used to identify the nonlinear plant (4.43):

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + W_1\Psi_1(x,y) + W_2\Psi_2(x,y)u, \\ \varepsilon\dot{\hat{y}} &= \varepsilon B\hat{y} + W_3\Psi_3(x,y) + W_4\Psi_4(x,y)u, \end{aligned} \tag{4.44}$$

where $\hat{x} \in \Re^n$ and $\hat{y} \in \Re^m$ are the slow and fast state variables of the NN, respectively. $A \in \Re^{n\times n}$, $B \in \Re^{m\times m}$ are diagonal stable matrices. $W_1 \in \Re^{n\times q}$, $W_2 \in \Re^{n\times q}$, $W_3 \in \Re^{m\times q}$, $W_4 \in \Re^{m\times q}$ are the weight matrices of the NN, q is the number of neuron. The activation function vectors $\Psi_1(\cdot)$, $\Psi_3(\cdot)$ are defined as:

$$\begin{aligned} \Psi_i(\cdot) &= [\Psi_{i,1}, \Psi_{i,2}, \cdots, \Psi_{i,q}]^T \in \Re^{q\times 1}, i = 1, 3, \\ \Psi_{i,c}(z) &= \prod_{j\in J_c} [\psi_i(\cdot)]^{d_c(j)}, c = 1, \cdots, q, \end{aligned} \tag{4.45}$$

where J_c are the collections of l not ordered subsets of $1, 2, \cdots, n+m$ and $d_c(j)$ are non-negative integers.

$\Psi_2(\cdot)$, $\Psi_4(\cdot)$ in (4.44) are defined as:

$$\Psi_i(\cdot) = \begin{bmatrix} \Psi_{i,1,1} & \cdots & \Psi_{i,1,p} \\ \vdots & \ddots & \vdots \\ \Psi_{i,q,1} & \cdots & \Psi_{i,q,p} \end{bmatrix} \in \Re^{q\times p}, i = 2, 4,$$

$$\Psi_{i,c,r} = \prod_{j\in J_{c,r}} [\psi_i(\cdot)]^{d_{c,r}(j)}, c = 1, \cdots, q, r = 1, \cdots, p,$$

where $J_{c,r}$ are the collections of l not ordered subsets of $1, 2, \cdots, n+m$ and $d_{c,r}(j)$ are non-negative integers. The activation functions $\psi_i(z)$ are chosen as

$$\psi_i(z) = \frac{\alpha_{i,1}}{1 + e^{-\alpha_{i,2}}} + \alpha_{i,3},\ i = 1, \cdots, 4. \tag{4.46}$$

The parameters $\alpha_{i,1}$,$\alpha_{i,2}$,$\alpha_{i,3}$ in (4.46) can be chosen a priori. The structure of the identification scheme is shown in Fig. 4.7.

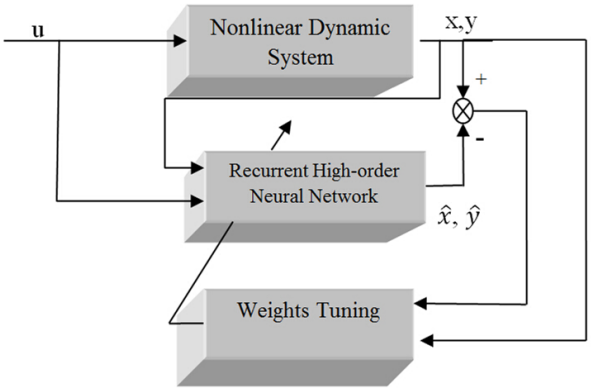


Figure 4.7: Structure of the identification scheme for continuous system.

The main goal in this section is to develop an on-line identification scheme and a weight's updating algorithm for two-time scales recurrent high-order neural network identifier (4.44) of the continuous time nonlinear system (4.43).

Assume a nominal NN with modeling errors can be used to approximate the nonlinear system (4.43) as:

$$\begin{aligned} \dot{x} &= Ax + W_1^* \Psi_1(x, y) + W_2^* \Psi_2(x, y) u + \zeta_x, \\ \varepsilon \dot{y} &= \varepsilon B y + W_3^* \Psi_2(x, y) + W_4^* \Psi_4(x, y) u + \zeta_y, \end{aligned} \tag{4.47}$$

where $W_1^* \in \Re^{n \times q}$,$W_2^* \in \Re^{n \times q}$,$W_3^* \in \Re^{m \times q}$,$W_4^* \in \Re^{n \times q}$ are the unknown optimal weights that minimize the modeling errors $\zeta_x \in \Re^{n \times 1}$ and $\zeta_y \in \Re^{m \times 1}$. Let ζ_{xi} denote the i^{th} term of ζ_x, $i = 1, \cdots, n$, and ζ_{yj} denote the j^{th} term of ζ_y, $j = 1, \cdots, m$, the following assumption is made.

Assumption 4.3. *The modeling error ζ_{xi} and ζ_{yj} are bounded by known bounds $\bar{\zeta}_{xi}$ and $\bar{\zeta}_{yj}$ as:*

$$|\zeta_{xi}| \leq \bar{\zeta}_{xi}, |\zeta_{yj}| \leq \bar{\zeta}_{yj}. \tag{4.48}$$

Denote $\Theta_x^* = [W_1^*, W_2^*] = [\theta_{x1}^*; \theta_{x2}^*; \cdots; \theta_{xn}^*] \in \Re^{n\times 2q}$, $\theta_{xi}^* \in \Re^{1\times 2q}$, $\Theta_y^* = [W_3^*, W_4^*] = [\theta_{y1}^*; \theta_{y2}^*; \cdots; \theta_{ym}^*] \in \Re^{m\times 2q}$, $\theta_{yj}^* \in \Re^{1\times 2q}$, $i = 1, \cdots, n$, $j = 1, \cdots m$. $H_x = [\Psi_1(x,y); \Psi_2(x,y)u]$, $H_y = [\Psi_3(x,y); \Psi_4(x,y)u]$. Then model (4.47) can be written as:

$$\begin{aligned} \dot{x} &= Ax + \Theta_x^* H_x + \zeta_x, \\ \varepsilon\dot{y} &= \varepsilon By + \Theta_y^* H_y + \zeta_y. \end{aligned} \tag{4.49}$$

Similarly, model (4.44) can be reformulated as:

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + \Theta_x H_x, \\ \varepsilon\dot{\hat{y}} &= \varepsilon B\hat{y} + \Theta_y H_y, \end{aligned} \tag{4.50}$$

where $\Theta_x = [W_1, W_2] = [\theta_{x1}; \theta_{x2}; \cdots; \theta_{xn}] \in \Re^{n\times 2q}$, $\Theta_y = [W_3, W_4] = [\theta_{y1}; \theta_{y2}; \cdots; \theta_{ym}] \in \Re^{m\times 2q}$. Define the identification errors as:

$$\begin{aligned} \varsigma_x &= x - \hat{x}, \\ \varsigma_y &= y - \hat{y}. \end{aligned} \tag{4.51}$$

Define the auxiliary system outputs as:

$$\begin{aligned} \tau_x &= \Theta_x^* H_x + \zeta_x = \dot{x} - Ax, \\ \tau_y &= \Theta_y^* H_y + \zeta_y = \varepsilon\dot{y} - \varepsilon By. \end{aligned} \tag{4.52}$$

The estimated auxiliary system outputs are defined as:

$$\begin{aligned} \hat{\tau}_x &= \Theta_x H_x, \\ \hat{\tau}_y &= \Theta_y H_y. \end{aligned} \tag{4.53}$$

Then the auxiliary output errors are defined as:

$$\begin{aligned} e_x =& \tau_x - \hat{\tau}_x = \tilde{\Theta}_x H_x + \zeta_x, \\ e_y =& \tau_y - \hat{\tau}_y = \tilde{\Theta}_y H_y + \zeta_y, \end{aligned} \tag{4.54}$$

where $\tilde{\Theta}_x = \Theta_x^* - \Theta_x$, $\tilde{\Theta}_y = \Theta_y^* - \Theta_y$. Subtracting (4.50) from (4.49), and using (4.51) and (4.54), it follows that:

$$\begin{aligned} e_x &= \dot{\varsigma}_x - A\varsigma_x, \\ e_y &= \varepsilon\dot{\varsigma}_y - \varepsilon B\varsigma_y, \end{aligned} \tag{4.55}$$

Remark 4.2. *The auxiliary output error* (4.55) *can be viewed as a "augmented error" or "filtered error" [110]. The main purpose of this paper is to design a weight's updating algorithm for the recurrent high order nueral networks such that the identification errors ς_x, ς_y are bounded and minimized. Equation* (4.55) *implies that this goal can be achieved if only the output errors e_x, e_y are bounded and miminized.*

According to (4.14) in Definition 4.1, the weight ellipsoids of the recurrent HONN can be defined as:

$$\begin{aligned} S_{xi}(\theta_{xi}^*, P_{xi}) &= \{\theta_{xi} | \tilde{\theta}_{xi} P_{xi}^{-1} \tilde{\theta}_{xi}^T \leq 1\}, \\ S_{yj}(\theta_{yj}^*, P_{yj}) &= \{\theta_{yj} | \tilde{\theta}_{yj} P_{yj}^{-1} \tilde{\theta}_{yj}^T \leq 1\}, \end{aligned} \tag{4.56}$$

where $\tilde{\theta}_{xi} = \theta_{xi}^* - \theta_{xi}$, $\tilde{\theta}_{yj} = \theta_{yj}^* - \theta_{yj}$, $i = 1, \cdots, n$, $j = 1, \cdots, m$. P_{xi}, P_{yj} are symmetric positive definite matrices.

From (4.52), it is easy to know that $\|\tau_{xi} - \theta_{xi}^* H_x\|_2 \leq \bar{\zeta}_{xi}$, and $\|\tau_{yj} - \theta_{yj}^* H_y\|_2 \leq \bar{\zeta}_{yj}$. Hence,

$[\tau_{xi} - \theta_{xi}^* H_x]$ and $[\tau_{yj} - \theta_{yj}^* H_y]$ belong to two ellipsoids E_{yj} and E_{yj}, which are defined as:

$$\begin{aligned} E_{xi}(\theta_{xi}^* H_x, \bar{\zeta}_{xi}^2) &= \{\tau_{xi} | \frac{1}{\bar{\zeta}_{xi}^2} \|\tau_{xi} - \theta_{xi}^* H_x\|_2^2 \leq 1\}, \\ E_{yj}(\theta_{yj}^* H_x, \bar{\zeta}_{yj}^2) &= \{\tau_{yj} | \frac{1}{\bar{\zeta}_{yj}^2} \|\tau_{yj} - \theta_{yj}^* H_y\|_2^2 \leq 1\}, \end{aligned} \tag{4.57}$$

where $\|\cdot\|_2$ denotes the L_2-norm, τ_{xi} is the i^{th} term of τ_x, and τ_{yj} is the j^{th} term of τ_y.

Assumption 4.4. *It is assumed that the initial weights $\theta_{xi}(t_0)$ and $\theta_{yj}(t_0)$ are in the ellipsoids $S_{xi}(t_0)$ and $S_{yj}(t_0)$:*

$$\begin{aligned} S_{xi}(t_0) &= \{\theta_{xi}(t_0) | \tilde{\theta}_{xi}(t_0) P_{xi}^{-1}(t_0) \tilde{\theta}_{xi}^T(t_0) \leq 1\}, \\ S_{yj}(t_0) &= \{\theta_{yj}(t_0) | \tilde{\theta}_{yj}(t_0) P_{yj}^{-1}(t_0) \tilde{\theta}_{yj}^T(t_0) \leq 1\}. \end{aligned} \tag{4.58}$$

Assumption 4.4 can be easily satisfied by choosing the bounded initial weights $\theta_{xi}(t_0)$, $\theta_{yj}(t_0)$ and suitable $P_{xi}(t_0)$ and $P_{yj}(t_0)$.

According to (4.57), it gives that:

$$\begin{aligned} \frac{\mu_{xi}}{\bar{\zeta}_{xi}^2} \|\tau_{xi} - \theta_{xi}^* H_x\|_2^2 &\leq \mu_{xi}, \\ \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2} \|\tau_{yj} - \theta_{yj}^* H_y\|_2^2 &\leq \mu_{yj}, \end{aligned} \tag{4.59}$$

where $0 \leq \mu_{xi} < 1$ and $0 \leq \mu_{yj} < 1$, which will be defined later.

Lemma 4.1. *If $P \in \Re^{n \times n}$ is a symmetric positive definite matrix, $z \in \Re^{1 \times n}$ and $h \in \Re^{1 \times n}$ are any given row vectors, then the following inequality holds:*

$$0 \leq z P h^T h P z^T \leq z P z^T h P h^T. \tag{4.60}$$

Proof. Case 1: Consider the case $P = \text{diag}(p_1, p_2, \cdots, p_n) \in \Re^{n \times n}$ is a diagonal matrix. It is obvious that

$$z P h^T h P z^T = z P h^T (z P h^T)^T = (z P h^T)^2 \geq 0.$$

Assume that for a diagonal positive definite matrix $\bar{P} = \text{diag}(p_1, p_2, \cdots, p_{n-1})$ and any given row vectors $\bar{z} \in \Re^{1\times(n-1)}$ and $\bar{h} \in \Re^{1\times(n-1)}$, the following inequality holds:

$$\bar{z}\bar{P}\bar{h}^T\bar{h}\bar{P}\bar{z}^T \leq \bar{z}\bar{P}\bar{z}^T\bar{h}\bar{P}\bar{h}^T. \tag{4.61}$$

Then for $P = \begin{bmatrix} \bar{P} & \mathbf{0} \\ \mathbf{0} & p_n \end{bmatrix} \in \Re^{n\times n}$, $z = [\bar{z}, z_n] \in \Re^{1\times n}$, and $h = [\bar{h}, h_n] \in \Re^{1\times n}$, it is shown that:

$$\begin{aligned}
&zPz^ThPh^T - zPh^ThPz^T \\
=&(\bar{z}\bar{P}\bar{z}^T + z_np_nz_n)(\bar{h}\bar{P}\bar{h}^T + h_np_nh_n) - (\bar{z}\bar{P}\bar{h}^T + z_np_nh_n)^2 \\
=&\bar{z}\bar{P}\bar{z}^T\bar{h}\bar{P}\bar{h}^T - (\bar{z}\bar{P}\bar{h}^T)^2 + z_np_nz_n\bar{h}\bar{P}\bar{h}^T + h_np_nh_n\bar{z}\bar{P}\bar{z}^T - 2\bar{z}\bar{P}\bar{h}^Tz_np_nh_n \\
&+ z_np_nz_nh_np_nh_n - (z_np_nh_n)^2.
\end{aligned} \tag{4.62}$$

Using (4.61), it can be obtained that:

$$\bar{z}\bar{P}\bar{z}^T\bar{h}\bar{P}\bar{h}^T - (\bar{z}\bar{P}\bar{h}^T)^2 \geq 0 \tag{4.63}$$

$$\begin{aligned}
&z_np_nz_n\bar{h}\bar{P}\bar{h}^T + h_np_nh_n\bar{z}\bar{P}\bar{z}^T - 2\bar{z}\bar{P}\bar{h}^Tz_np_nh_n \\
\geq&2z_np_nh_n\sqrt{\bar{h}\bar{P}\bar{h}^T\bar{z}\bar{P}\bar{z}^T} - 2\bar{z}\bar{P}\bar{h}^Tz_np_nh_n \geq 0.
\end{aligned} \tag{4.64}$$

Substituting (4.63) and (4.64) into (4.62) yields:

$$zPz^ThPh^T - zPh^ThPz^T \geq 0. \tag{4.65}$$

When $P = p_1 \in \Re,\ p_1 > 0$, $z = z_1 \in \Re$ and $h = h_1 \in \Re$ are scalars, it is obvious $zPz^ThPh^T = zPh^ThPz^T \geq 0$ is valid. When $P = \text{diag}(p_1, p_2),\ p_1 > 0,\ p_2 > 0,\ z = [z_1\ z_2],\ h = [h_1\ h_2]$, it is also easy to verify that $zPz^ThPh^T \geq zPh^ThPz^T \geq 0$ is

true. Using the analysis result presented above, it can be concluded that, for all positive definite $P = \text{diag}(p_1, p_2, \cdots, p_n) \in \Re^{n\times n}$, and any given vectors $z \in \Re^{1\times n}$ and $h \in \Re^{1\times n}$, $zPz^ThPh^T \geq zPh^ThPz^T \geq 0$ is always valid.

Case 2: Consider the case $P \in \Re^{n\times n}$ is any given symmetric positive definite matrix. Using matrix eigen decomposition method, P can be easily factorized as $P = U\Lambda U^T$, where $U \in \Re^{n\times n}$ is an orthonormal matrix, and $\Lambda \in \Re^{n\times n}$ is a diagonal positive definite matrix whose entries are the eigenvalues of P. For any given vectors $z \in \Re^{1\times n}$ and $h \in \Re^{1\times n}$, zPh^ThPz^T becomes $zU\Lambda U^ThU\Lambda U^Tz^T = z'\Lambda h'^Th'\Lambda z'^T$, and zPz^ThPh^T becomes $zU\Lambda U^Tz^ThU\Lambda U^Th^T = z'\Lambda z'^Th'\Lambda h'^T$, where $z' = zU$, $h' = hU$. According to the results in case 1, it can be obtained that $z'\Lambda z'^Th'\Lambda h'^T \geq z'\Lambda h'^Th'\Lambda z'^T \geq 0$, which means $zPz^ThPh^T \geq zPh^ThPz^T \geq 0$ is true for all symmetric positive definite matrix P. Lemma 4.1 is thus proved. □

The following updating laws are proposed to update the RHONN:

$$\dot{P}_{xi} = -\frac{g_{xi}\mu_{xi}P_{xi}H_xH_x^TP_{xi}}{(1-\mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi}H_x^TP_{xi}H_x}, \tag{4.66}$$

$$\dot{\theta}_{xi}^T = \frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}g_{xi}P_{xi}H_xe_{xi}, \tag{4.67}$$

$$\mu_{xi} = \frac{\lambda_{xi}\bar{\zeta}_{xi}^2}{1 + H_x^TP_{xi}H_x}, \tag{4.68}$$

$$\dot{P}_{yj} = -\frac{g_{yj}\mu_{yj}P_{yj}H_yH_y^TP_{yj}}{(1-\mu_{yj})\bar{\zeta}_{yj}^2 + \mu_{yj}H_y^TP_{yj}H_y}, \tag{4.69}$$

$$\dot{\theta}_{yj}^T = \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}g_{yj}P_{yj}H_ye_{yj}, \tag{4.70}$$

$$\mu_{yj} = \frac{\lambda_{yj}\bar{\zeta}_{yj}^2}{1 + H_y^TP_{yj}H_y}, \tag{4.71}$$

where e_{xi} is the i^{th} element in e_x, $i = 1, \cdots, n$, e_{yj} is the j^{th} element in e_y, $j = 1, \cdots, m$, λ_{xi}, λ_{yj}

are designed parameters such that $0 < \lambda_{xi}\bar{\zeta}_{xi}^2 < 1$, $0 < \lambda_{yj}\bar{\zeta}_{yj}^2 < 1$, g_{xi}, g_{yj} are defined as:

$$\begin{aligned} g_{xi} &= \begin{cases} 1 & \text{if } e_{xi}^2 > \bar{\zeta}_{xi}^2 \ \& \ H_x^T P_{xi} H_x > \bar{\zeta}_{xi}^2 \\ 0 & \text{if } e_{xi}^2 \leq \bar{\zeta}_{xi}^2 \text{ or } H_x^T P_{xi} H_x \leq \bar{\zeta}_{xi}^2 \end{cases}, \\ g_{yj} &= \begin{cases} 1 & \text{if } e_{yj}^2 > \bar{\zeta}_{yj}^2 \ \& \ H_x^T P_{yj} H_y > \bar{\zeta}_{yj}^2 \\ 0 & \text{if } e_{yj}^2 \leq \bar{\zeta}_{yj}^2 \text{ or } H_y^T P_{yj} H_y \leq \bar{\zeta}_{yj}^2 \end{cases}. \end{aligned} \tag{4.72}$$

Theorem 4.2. *Consider the NN identifier* (4.44) *for the system* (4.43)*. If the initial weights are bounded ellipsoid sets as in* (4.58)*, by using the updating laws proposed in* (4.66)-(4.71)*, it can be guaranteed that P_{xi}, P_{yj} and the identification errors ς_{xi}, ς_{yj} are bounded and θ_{xi}, θ_{yj} remain inside ellipsoids defined in* (4.56)*.*

Proof. Case 1: When $e_{xi}^2 > \bar{\zeta}_{xi}^2$, $H_x^T P_{xi} H_x > \bar{\zeta}_{xi}^2$, $e_{yj}^2 > \bar{\zeta}_{yj}^2$, and $H_x^T P_{yj} H_y > \bar{\zeta}_{yj}^2$. In this case, it can be known from (4.72) that $g_{xi} = 1$ and $g_{yj} = 1$. Let $z \in \Re^n$ be an arbitrary non-zero constant row vector, and assume $P_{xi}(t)$ is a symmetric positive definite matrix, it is shown that by using Taylor series expansion and omitting the higher order terms, the following equation is obtained:

$$z P_{xi}(t + \Delta t) z^T = z(P_{xi}(t) + \dot{P}_{xi} \Delta t) z^T, \tag{4.73}$$

where $\Delta t \to 0$ is a small time interval. Substituting (4.66) into (4.73) gives:

$$\begin{aligned} & z P_{xi}(t + \Delta t) z^T \\ =& z(P_{xi} - \frac{\mu_{xi} \Delta t P_{xi} H_x H_x^T P_{xi}}{(1 - \mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi} H_x^T P_{xi} H_x}) z^T \\ =& \frac{\mu_{xi}[H_x^T P_{xi} H_x z P_{xi} z^T - \Delta t z P_{xi} H_x H_x^T P_{xi} z^T]}{(1 - \mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi} H_x^T P_{xi} H_x} + \frac{(1 - \mu_{xi})\bar{\zeta}_{xi}^2 z P_{xi} z^T}{(1 - \mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi} H_x^T P_{xi} H_x}. \end{aligned}$$

Using lemma 4.1, one has $H_x^T P_{xi} H_x z P_{xi} z^T \geq z P_{xi} H_x H_x^T P_{xi} z^T \geq 0$. Because $0 < \mu_{xi} < 1$ and $H_x^T P_{xi} H_x > 0$, it is obtained that $z P_{xi}(t + \Delta t) z^T > 0$ since $\Delta t \to 0$ is a very small time interval satisfying $\Delta t < 1$. If the initial matrix $P_{xi}(t_0)$ is chosen to be symmetric positive definite

matrix, it can be concluded that by using updating law (4.66), P_{xi} will remain symmetric positive definite. Following the same analysis procedure, it is obvious that P_{yj} will also remain symmetric positive definite during the identification process.

Define:

$$V_{xz} = tr\{zP_{xi}z^T\}, \tag{4.74}$$

$$V_{yz} = tr\{zP_{yj}z^T\}. \tag{4.75}$$

Using (4.66), the derivative of (4.74) is obtained as:

$$\begin{aligned} \dot{V}_{xz} &= tr\{z\dot{P}_{xi}z^T\} \\ &= -\frac{tr\{zP_{xi}H_xH_x^TP_{xi}z^T\}}{(1-\mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi}H_x^TP_{xi}H_x} \\ &= -\frac{(zP_{xi}H_x)^2}{(1-\mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi}H_x^TP_{xi}H_x}. \end{aligned} \tag{4.76}$$

Since $\mu_{xi} < 1$ and $H_x^TP_{xi}H_x > 0$, it is easy to show that $\dot{V}_{xz} \leq 0$, which implies P_{xi} is bounded. Using the same approach, it can be obtained that $\dot{V}_{yz} \leq 0$, which indicates that P_{yj} is also bounded.

Define

$$L_{xi} = \tilde{\theta}_{xi}P_{xi}^{-1}\tilde{\theta}_{xi}^T, \tag{4.77}$$

$$L_{yj} = \tilde{\theta}_{yj}P_{yj}^{-1}\tilde{\theta}_{yj}^T. \tag{4.78}$$

The derivative of L_{xi} is given as:

$$\dot{L}_{xi} = 2\tilde{\theta}_{xi}P_{xi}^{-1}\dot{\tilde{\theta}}_{xi}^T - \tilde{\theta}_{xi}P_{xi}^{-1}\dot{P}_{xi}P_{xi}^{-1}\tilde{\theta}_{xi}^T = -2\tilde{\theta}_{xi}P_{xi}^{-1}\dot{\theta}_{xi}^T - \tilde{\theta}_{xi}P_{xi}^{-1}\dot{P}_{xi}P_{xi}^{-1}\tilde{\theta}_{xi}^T. \tag{4.79}$$

Substituting (4.66), (4.67) into (4.79), it follows that:

$$\begin{aligned}\dot{L}_{xi} =& -\frac{2\mu_{xi}}{\bar{\zeta}_{xi}^2}\tilde{\theta}_{xi}H_x e_{xi} + \frac{\mu_{xi}\tilde{\theta}_{xi}H_x H_x^T\tilde{\theta}_{xi}^T}{(1-\mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi}H_x^T P_{xi}H_x} - \frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}\|\tau_{xi} - \theta_{xi}^* H_x\|_2^2 \\ &+ \frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}\tilde{\theta}_{xi}H_x H_x^T\tilde{\theta}_{xi}^T + \frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}\|\tau_{xi} - \theta_{xi}^* H_x\|_2^2 - \frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}\tilde{\theta}_{xi}H_x H_x^T\tilde{\theta}_{xi}^T \\ =&\Big[-\frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}\|\tau_{xi} - \theta_{xi}^* H_x\|_2^2 + \frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}\tilde{\theta}_{xi}H_x H_x^T\tilde{\theta}_{xi}^T - \frac{2\mu_{xi}}{\bar{\zeta}_{xi}^2}\tilde{\theta}_{xi}H_x e_{xi}\Big] \\ &+ \frac{\mu_{xi}\tilde{\theta}_{xi}H_x H_x^T\tilde{\theta}_{xi}^T}{(1-\mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi}H_x^T P_{xi}H_x} - \frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}\tilde{\theta}_{xi}H_x H_x^T\tilde{\theta}_{xi}^T + \frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}\|\tau_{xi} - \theta_{xi}^* H_x\|_2^2. \end{aligned} \tag{4.80}$$

It is noticed that $e_{xi} = \tau_{xi} - \theta_{xi}H_x$ and $\tilde{\theta}_{xi} = \theta_{xi}^* - \theta_{xi}$, then the equation in $[\cdot]$ in (4.80) can be rewritten as:

$$\begin{aligned} &-\frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}\|\tau_{xi} - \theta_{xi}^* H_x\|_2^2 + \frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}\tilde{\theta}_{xi}H_x H_x^T\tilde{\theta}_{xi}^T - \frac{2\mu_{xi}}{\bar{\zeta}_{xi}^2}\tilde{\theta}_{xi}H_x e_{xi} \\ =&\frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}\big[-\|\tau_{xi} - \theta_{xi}^* H_x\|_2^2 + (\theta_{xi}^* - \theta_{xi})H_x H_x^T(\theta_{xi}^* - \theta_{xi})^T - 2(\theta_{xi}^* - \theta_{xi})H_x(\tau_{xi} - \theta_{xi}H_x)\big] \\ =&-\frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}\big[\tau_{xi}^2 - 2\theta_{xi}H_x\tau_{xi} + \theta_{xi}H_x H_x^T\theta_{xi}^T\big] \\ =&-\frac{\mu_{xi}}{\bar{\zeta}_{xi}^2}e_{xi}^2. \end{aligned} \tag{4.81}$$

Substituting (4.81) into (4.80), and using (4.59), $0 < \mu_{xi} < 1$, $H_x^T P_{xi}H_x > \bar{\zeta}_{xi}^2$, $e_{xi}^2 > \bar{\zeta}_{xi}^2$, it gives the following inequality:

$$\begin{aligned}\dot{L}_{xi} \leq& \frac{[\bar{\zeta}_{xi}^2 - (1-\mu_{xi})\bar{\zeta}_{xi}^2 - \mu_{xi}H_x^T P_{xi}H_x]\mu_{xi}\tilde{\theta}_{xi}H_x H_x^T\tilde{\theta}_{xi}^T}{\bar{\zeta}_{xi}^2[(1-\mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi}H_x^T P_{xi}H_x]} + \mu_{xi}(1 - \frac{e_{xi}^2}{\bar{\zeta}_{xi}^2}) \\ =&\frac{[\bar{\zeta}_{xi}^2 - H_x^T P_{xi}H_x]\mu_{xi}^2\tilde{\theta}_{xi}H_x H_x^T\tilde{\theta}_{xi}^T}{\bar{\zeta}_{xi}^2[(1-\mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi}H_x^T P_{xi}H_x]} + \mu_{xi}(1 - \frac{e_{xi}^2}{\bar{\zeta}_{xi}^2}) < 0. \end{aligned} \tag{4.82}$$

Since the initial NN weights are bounded as $L(t_0) = \tilde{\theta}_{xi}(t_0)P_{xi}^{-1}(t_0)\tilde{\theta}_{xi}^T(t_0) \leq 1$. From (4.82), it is easy to know that the θ_{xi} will remain bounded and belong to the ellipsoid defined in (4.56). Using (4.54), since $\tilde{\theta}_{xi}$ is bounded, the auxiliary output errors e_{xi} is bounded, which implies ς_{xi} is also bounded according to (4.55). Using the same method, it can be proved that $\dot{L}_{yj} < 0$ is also valid, and θ_{yj} will also remain bounded and belong to the ellipsoid defined in (4.56). Hence, e_{yj}

and ς_{yj} are all bounded.

Case 2: $e_{xi}^2 \leq \bar{\zeta}_{xi}^2$, or $H_x^T P_{xi} H_x \leq \bar{\zeta}_{xi}^2$, or $e_{yj}^2 \leq \bar{\zeta}_{yj}^2$, or $H_y^T P_{yj} H_y \leq \bar{\zeta}_{yj}^2$. According to (4.72), one has $g_{xi} = 0$ or $g_{yj} = 0$. In this case, it can be obtained that $\dot{P}_{xi} = 0$, $\dot{\theta}_{xi}^T = 0$, or $\dot{P}_{yj} = 0$, $\dot{\theta}_{yj}^T = 0$ based on the updating laws (4.66)-(4.71). This indicates that P_{xi} or P_{yj} will be constant and bounded. This also implies that θ_{xi} or θ_{yj} will remain constant and inside the ellipsoids defined in (4.56). Therefore, e_{xi}, ς_{xi}, e_{yj} and ς_{yi} will also remain bounded. Theorem 4.2 is thus proved. □

Remark 4.3. *From* (4.74) *and* (4.76)*, it is shown that* $V_{xi} = tr\{zP_{xi}z^T\}$ *will decrease during the identification process, which means* $zP_{xi}^{-1}z^T$ *will increase during the identification process. According to Theorem 2,* θ_{xi} *remains inside the ellipsoid defined in* (4.56)*, i.e.* $\tilde{\theta}_{xi}P_{xi}^{-1}\tilde{\theta}_{xi}^T \leq 1$*. This implies* $\tilde{\theta}_{xi}$ *will decrease during the identification process, and* θ_{xi} *will converge to its nominal value* θ_{xi}^**. Following the same analysis procedure, it can be obtained that* θ_{yj} *will also converge to its nominal value* θ_{yj}^**.*

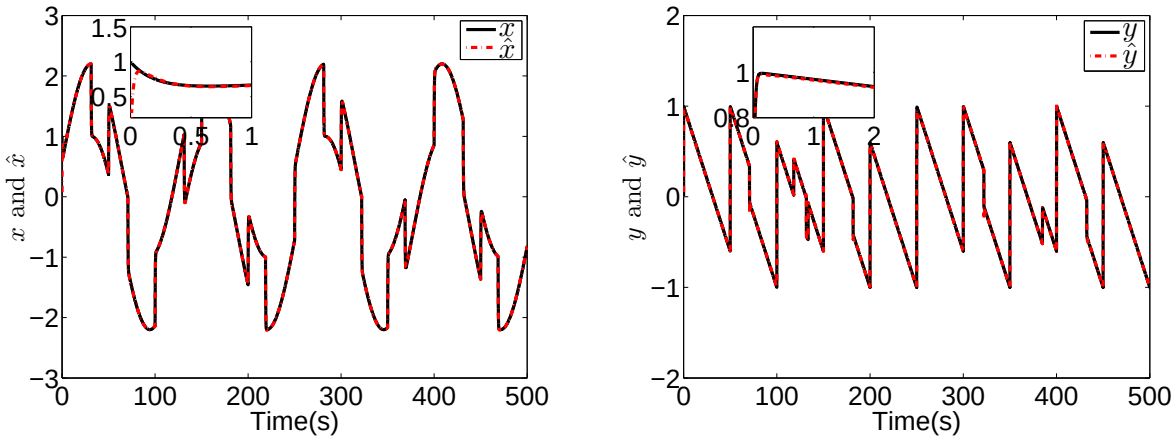
Remark 4.4. *In* (4.68) *and* (4.71)*, due to the existence of* P_{xi} *and* P_{yj}*, the "learning gain" of the weight's updating laws for* θ_{xi} *and* θ_{yj} *will be changed adaptively, whereas in many other gradient descent like algorithms, the "learning gain" is fixed. The main advantage of the weight's updating laws proposed in this section is that by using adaptively adjusted "learning gain", the identification process can achieve faster convergence with less oscillation. This is demonstrated by the simulations.*

4.3.2 Simulation

In order to demonstrate the effectiveness of the new identification algorithm proposed in this section, simulation is conducted based on the following system:

$$\begin{aligned} \dot{x} &= -5x + 3\text{sign}(y) + u_1, \\ 0.2\dot{y} &= -10y + 2\text{sign}(x) + u_2, \end{aligned} \tag{4.83}$$

with $x(t_0) = 1$ and $y(t_0) = 0$. The input signal u_1 is chosen as sinusoidal function $u_1 = 8\sin(0.05t)$, and u_2 is chosen to be a saw-tooth function with a frequency of 0.02Hz and amplitude of 8. System (4.83) is the same as the system considered in Section 3.5 for comparison purpose. In order to identify this singularly perturbed nonlinear system, the NN defined in (4.44) with node number $q = 4$ is utilized. In this case, $W_1 \in \Re^{1\times 4}$, $W_2 \in \Re^{1\times 4}$, $W_3 \in \Re^{1\times 4}$, $W_4 \in \Re^{1\times 4}$. The activation functions in (4.46) are chosen as $\psi_1(z) = 10/(1+e^{-2z}) + 1$, $\psi_2(z) = 1/(1+e^{-z}) + 1$, $\psi_3(z) = 1/(1+e^{-z}) + 1$, $\phi_4(z) = 1/(1+e^{-z}) + 1$. $A = B = -50$, $\lambda_x = 500$, $\lambda_y = 1000$, $\bar{\zeta}_x = \bar{\zeta}_y = 0.0001$, $P_{xi}(t_0) = P_{yj}(t_0) = I \in \Re^{8\times 8}$. The sampling time used in this simulation is 1 ms. The simulation results of identification algorithm proposed in this section (method 1), the identification algorithm using multilayer NN proposed in Section 3.5 (method 2) and identification algorithm using single layer NN proposed in [16] (method 3) are presented in Figs. 4.8-4.13.

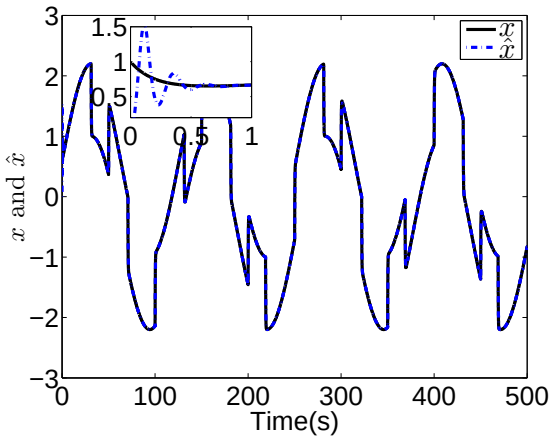


(a) Identification result of x.

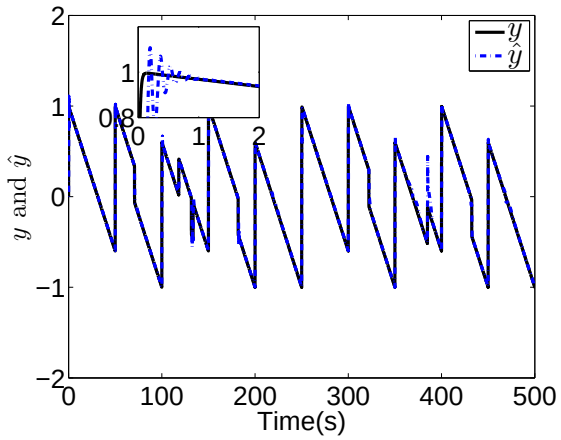
(b) Identification result of y.

Figure 4.8: Identification results using the method 1.

In Fig. 4.8, the magnification plots of the beginning phases of the identification process are presented in the top left corners. From these magnification plots, it is clear that both $\hat{x}$ and $\hat{y}$ overlap real system states x and y since the very beginning the identification process when the method 1 is used. In Fig. 4.9, the magnification plots shows that $\hat{x}$ converges to x after 0.5s and $\hat{y}$ converges to y after 1s when method 2 is used. In Fig. 4.10, large difference between $\hat{x}$ and x can be observed, and much more oscillations on $\hat{y}$ can be seen before it converges to y when method 3 is used. Meanwhile, Fig. 4.11 shows that when method 1 is used, the identification errors ς_x and

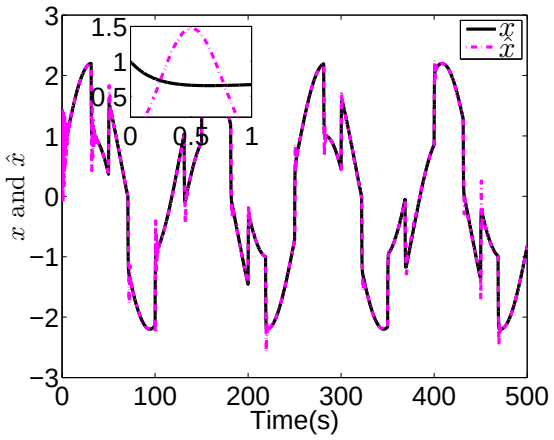


(a) Identification result of x.

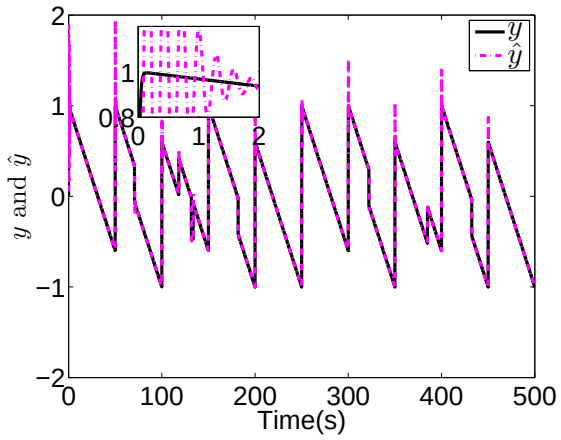


(b) Identification result of y.

Figure 4.9: Identification results using the method 2.



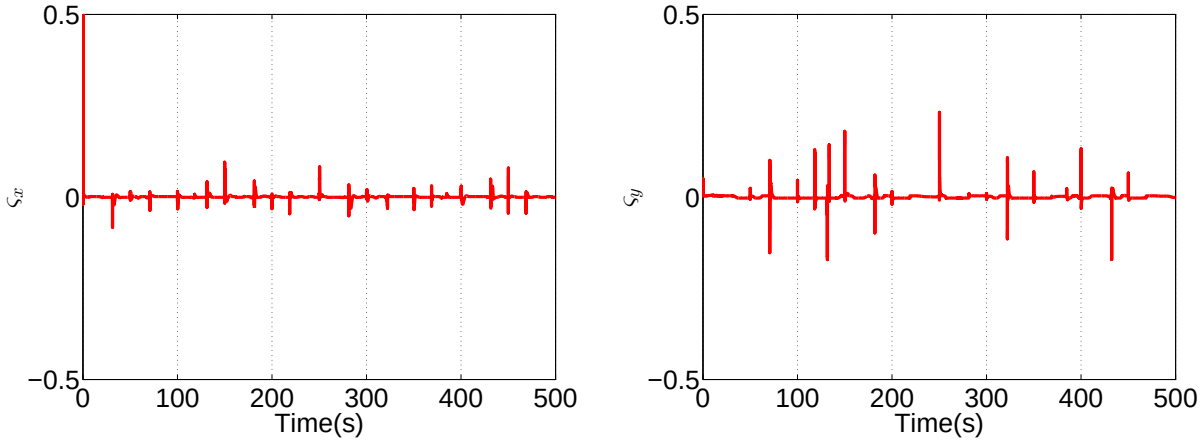
(a) Identification result of x.



(b) Identification result of y.

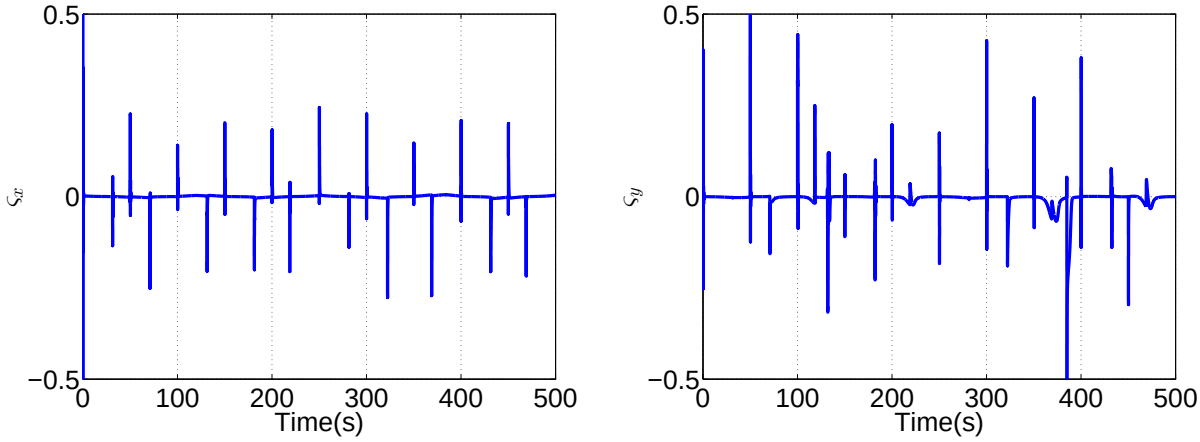
Figure 4.10: Identification results using the method 3.

ς_y are much smaller, compared with the identification errors obtained when method 2 and 3 are used, as presented in Fig. 4.12 and Fig. 4.13. Hence, it is clear that the identification results of method 1 is much better than that of method 2 and 3. Unlike method 2, 3 and many other widely used gradient-like learning algorithms which have fixed "learning gain", the "learning gain" of the updating laws in the identification algorithm proposed in this section can be changed adaptively. So the identification results of method 1 can converge to the reference signals faster, with less oscillation, compared to the results of method 2 and 3. Also, it is very clear from Fig. 4.11- 4.13 that the identification errors ς_x and ς_y of method 1 are greatly reduced compared to the errors of



(a) Identification error of x. (b) Identification error of y.

Figure 4.11: Identification errors using the method 1.



(a) Identification error of x. (b) Identification error of y.

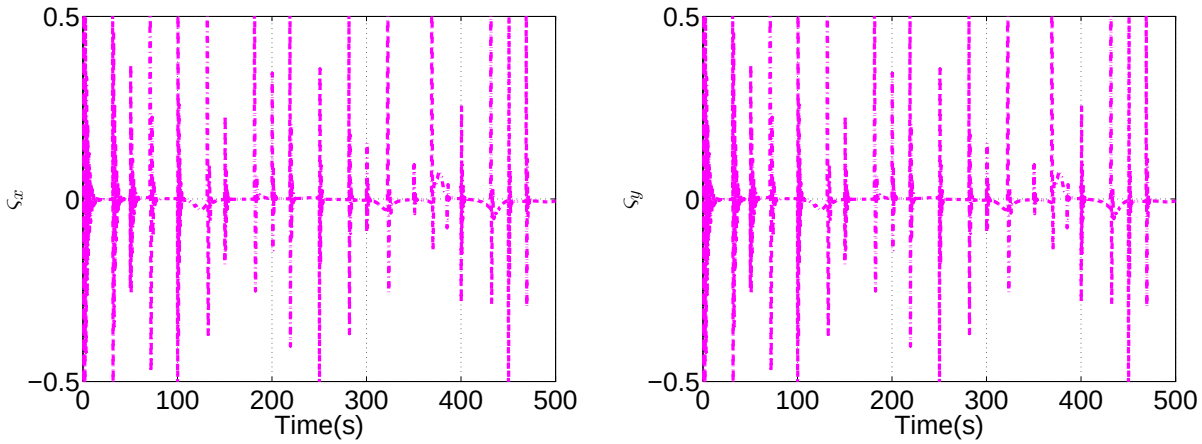
Figure 4.12: Identification errors using the method 2.

method 2 and 3.

The performance index-Root Mean Square (RMS) can also be used to illustrate the effectiveness of the identification algorithm proposed in this section. The RMS of the identification errors ς_x and ς_y are calculated as:

$$RMS_x = \sqrt{(\sum_{i=1}^{n} \varsigma_x^2(i))/n},$$

$$RMS_y = \sqrt{(\sum_{i=1}^{n} \varsigma_y^2(i))/n},$$



(a) Identification error of x.

(b) Identification error of y.

Figure 4.13: Identification errors using the method 3.

where n is the total number of simulation steps. The RMS values of ς_x and ς_y for method 1, method 2 and method 3 are given in Table 4.1. From Table 4.1, it is very clear that RMS_x and RMS_y of method 1 is much smaller than that of the method 2 and 3, which means the identification algorithm proposed in this section can achieve more accurate results.

Table 4.1: RMS values of ς_x and ς_y

	Method 1	Method 2	Method 3
RMS_x	0.0068	0.0175	0.0937
RMS_y	0.0064	0.0336	0.0382

4.4 Identification of Continuous Systems Using Modified Optimal Bounded Ellipsoid Algorithm

In Section 4.3, the OBE based identification scheme for singularly perturbed systems using continuous multi-time-scale neural networks is established. The indirect adaptive control for a regulation problem based on the identification scheme proposed in Section 4.3 is also investigated, which will be presented in Chapter 6. However, it was found that when the OBE based identification method proposed in Section 4.3 is used in a indirect adaptive trajectory tracking control

problem, the identification errors will be small at the beginning but will increase to a higher level at the end of the identification process. This is due to the fact that the learning gain of the OBE algorithm based identification scheme will keep decreasing during the identification process. When the learning gain is too small, the identification method will lose its ability to adjust the NN weights. Therefore, the NN cannot adapt to the changing dynamics of the system, and a larger identification error will occur.

In this section, by adding two additional terms in the updating laws, a modified OBE algorithm is designed to update the NN weights. By using this modified OBE algorithm, it can be guaranteed that the NN weight errors will be uniformly ultimately bounded (UUB), and the weight errors will converge to the boundary exponentially. Thus the convergence speed is much faster than that of the GD based identification scheme, which can only achieve asymptotic convergence. Meanwhile, unlike the original OBE algorithm based methods proposed in Section 4.3, the learning gain of the modified OBE algorithm will not go to $\mathbf{0}$. Instead, it will converge to an equilibrium point which can be chosen arbitrarily by the user. Hence, the modified OBE algorithm based training method will remain effective during the whole identification process.

4.4.1 Identification Algorithm

Consider the following affine in control nonlinear SPS:

$$\begin{aligned} \dot{x} &= f_1(x) + f_2(x)y, \\ \varepsilon\dot{y} &= g_1(x,y) + g_2(x,y)u, \end{aligned} \tag{4.84}$$

where $x \in \Re^n$ and $y \in \Re^m$ are the slow and fast states, respectively, $u \in \Re^p$ is the control signal vector, $0 < \varepsilon < 1$ is a small parameter, $f_i \in C^\infty, i = 1, 2$ and $g_i \in C^\infty, i = 1, 2$ are unknown general nonlinear smooth functions.

To identify the nonlinear system (4.84), the following multi-time-scale RHONN is used:

$$\dot{\hat{x}} = A\hat{x} + W_1\Psi_1(x) + W_2\Psi_2(x)y + L_x\varsigma_x, \tag{4.85a}$$

$$\varepsilon\dot{\hat{y}} = B\hat{y} + W_3\Psi_3(x,y) + W_4\Psi_4(x,y)u + L_y\varsigma_y, \tag{4.85b}$$

where $\hat{x} \in \Re^n$ and $\hat{y} \in \Re^m$ are the estimation of the slow and fast states x and y, respectively, $A = \text{diag}(a_1, \cdots, a_n) \in \Re^{n\times n}$, $B = \text{diag}(b_1, \cdots, b_m) \in \Re^{m\times m}$ are diagonal stable matrices, $W_1 \in \Re^{n\times q}, W_2 = \text{diag}(w_{21}, \cdots, w_{2n}) \in \Re^{n\times n}, W_3 \in \Re^{m\times q}, W_4 = \text{diag}(w_{41}, \cdots, w_{4m}) \in \Re^{m\times m}$ are the weight matrices of the RHONN, q is the number of neuron, $L_x = \text{diag}(l_{x1}, \cdots, l_{xn})$, $L_y = \text{diag}(l_{y1}, \cdots, l_{ym})$ are diagonal positive definite matrices, $\varsigma_x,\ \varsigma_y$ are the identification errors defined as:

$$\begin{aligned} \varsigma_x &= x - \hat{x}, \\ \varsigma_y &= y - \hat{y}. \end{aligned} \tag{4.86}$$

The activation function vectors $\Psi_1(\cdot)$, $\Psi_3(\cdot)$ are defined as:

$$\begin{aligned} \Psi_i(\cdot) &= [\Psi_{i,1}, \Psi_{i,2}, \cdots, \Psi_{i,q}]^T \in \Re^{q\times 1}, i = 1, 3, \\ \Psi_{i,c}(\cdot) &= \prod_{j\in J_c} [\psi_i(\cdot)]^{d_c(j)}, c = 1, \cdots, q, \end{aligned}$$

where J_c are the collections of l not ordered subsets of $1, 2, \cdots, n+m$ and $d_c(j)$ are non-negative integers.

The activation function matrix $\Psi_2(\cdot)$ is defined as:

$$\begin{aligned} \Psi_2(\cdot) &= \begin{bmatrix} \Psi_{2,1,1} & \cdots & \Psi_{2,1,m} \\ \vdots & \ddots & \vdots \\ \Psi_{2,n,1} & \cdots & \Psi_{2,n,m} \end{bmatrix} \in \Re^{n\times m}, \\ \Psi_{2,c,r} &= \prod_{j\in J_{c,r}} [\psi_2(\cdot)]^{d_{c,r}(j)}, c = 1, \cdots, n, r = 1, \cdots, m, \end{aligned}$$

where $J_{c,r}$ are the collections of l not ordered subsets of $1, 2, \cdots, n+m$ and $d_{c,r}(j)$ are non-negative integers. The activation function matrix $\Psi_4(\cdot)$ is defined as:

$$\Psi_4(\cdot) = \begin{bmatrix} \Psi_{4,1,1} & \cdots & \Psi_{4,1,p} \\ \vdots & \ddots & \vdots \\ \Psi_{4,m,1} & \cdots & \Psi_{4,m,p} \end{bmatrix} \in \Re^{m \times p},$$

$$\Psi_{4,c,r} = \prod_{j \in J_{c,r}} [\psi_4(\cdot)]^{d_{c,r}(j)}, c = 1, \cdots, m, r = 1, \cdots, p,$$

where $J_{c,r}$ are the collections of l not ordered subsets of $1, 2, \cdots, n+m$ and $d_{c,r}(j)$ are non-negative integers.

The activation functions $\psi_i(\cdot)$ are chosen as:

$$\psi_i(z) = \frac{\alpha_{i,1}}{1 + e^{-\alpha_{i,2} z}} + \alpha_{i,3}, i = 1, \cdots, 4. \tag{4.87}$$

Remark 4.5. *In this section, a multi-time-scale RHONN model* (4.85) *is used to identify the unknown nonlinear SPS* (4.84)*, such that the singular perturbation theory can be applied to the identified system model to decompose the original system into the fast and slow subsystems. This will reduce the system order, and simplify the structure of controller. System* (4.84) *can also be identified using a regular (one-time-scale) NN [111] with similar structure as:*

$$\dot{Z} = A_z Z + W_{z1} \Psi_{z1}(Z) + W_{z2} \Psi_{z2}(Z) u, \tag{4.88}$$

where $Z = [x; y] \in \Re^{(n+m) \times 1}$, $A_z \in \Re^{(n+m) \times (n+m)}$ is a stable matrix, $W_{z1},\ W_{z2}$ and $\Psi_{z1},\ \Psi_{z2}$ are weight matrices and activation function vectors (matrices) with appropriate dimensions, respectively. However, it will be much more difficult to apply the singular perturbation theory to separate the fast dynamic states from the slow dynamic states in the identified model (4.88)*. As a result, the system order will not be reduced and the controller design can not be simplified.*

Assume the nonlinear system (4.84) can be approximated by the following nominal RHONN:

$$\begin{aligned} \dot{x} &= Ax + W_1^*\Psi_1(x) + W_2^*\Psi_2(x)y + \zeta_x', \\ \varepsilon\dot{y} &= By + W_3^*\Psi_3(x,y) + W_4^*\Psi_4(x,y)u + \zeta_y, \end{aligned} \tag{4.89}$$

where $\zeta_x' \in \Re^{n\times 1}$, $\zeta_y \in \Re^{m\times 1}$ are the modeling errors, $W_1^* \in \Re^{n\times q}, W_2^* = \text{diag}(w_{21}^*, \cdots, w_{2n}^*) \in \Re^{n\times n}, W_3^* \in \Re^{m\times q}, W_4^* = \text{diag}(w_{41}^*, \cdots, w_{4m}^*) \in \Re^{m\times m}$ are unknown optimal weights which minimize ζ_x' and ζ_y.

Assumption 4.5. *The modeling error ζ_x' and ζ_y are bounded by upper bounds $\bar{\zeta}_x'$ and $\bar{\zeta}_y$ as:*

$$0 < \|\zeta_x'\|_2 \leq \bar{\zeta}_x', 0 < \|\zeta_y\|_2 \leq \bar{\zeta}_y, \tag{4.90}$$

where $\|\cdot\|_2$ denotes the L_2-norm.

Using (4.86) and (4.89), one can obtain that

$$\begin{aligned} \dot{x} =& Ax + W_1^*\Psi_1(x) + W_2^*\Psi_2(x)\hat{y} + \zeta_x, \\ \varepsilon\dot{y} =& By + W_3^*\Psi_3(x,y) + W_4^*\Psi_4(x,y)u + \zeta_y, \end{aligned} \tag{4.91}$$

where

$$\zeta_x = W_2^*\Psi_2(x)\varsigma_y + \zeta_x'. \tag{4.92}$$

Remark 4.6. *In* (4.92)*, $W_2^*,\ \Psi_2,(x)\ \zeta_x'$ are all bounded. Also, it can be proved later that by using the identification scheme proposed in this section, the identification error ς_y will be bounded. Hence, ζ_x will also be bounded.*

Let ζ_{xi} denotes the i^{th} term of ζ_x, $i = 1, \cdots, n$, and ζ_{yj} denotes the j^{th} term of ζ_y, $j = 1, \cdots, m$. One can obtain that:

$$0 < |\zeta_{xi}| \leq \bar{\zeta}_{xi}, 0 < |\zeta_{yj}| \leq \bar{\zeta}_{yj}, \tag{4.93}$$

where $\bar{\zeta}_{xi}$ and $\bar{\zeta}_{yj}$ are the upper bounds of ζ_{xi}, ζ_{yj}.

Denote $\theta^*_{xi} = [w^*_{1i}, w^*_{2i}] \in \Re^{1\times(q+1)}$, $\theta^*_{yj} = [w^*_{3j}, w^*_{4j}] \in \Re^{1\times(q+1)}$, $i = 1, \cdots, n$, $j = 1, \cdots m$, $H_{xi} = \left[\Psi_1(x); [\Psi_2(x)\hat{y}]_i\right] \in \Re^{(q+1)\times 1}$, $H_{yj} = \left[\Psi_3(x,y); [\Psi_4(x,y)u]_j\right] \in \Re^{(q+1)\times 1}$, where w^*_{1i} and w^*_{3j} are the i^{th} and j^{th} row of W^*_1 and W^*_3, respectively, $[\Psi_2(x)\hat{y}]_i$ and $[\Psi_4(x,y)u]_j$ are the i^{th} and j^{th} element of $\Psi_2(x)\hat{y}$ and $\Psi_4(x,y)u$, respectively. Then (4.91) can be written as:

$$\begin{aligned} \dot{x}_i &= a_i x_i + \theta^*_{xi} H_{xi} + \zeta_{xi}, \\ \varepsilon \dot{y}_j &= b_j y_j + \theta^*_{yj} H_{yj} + \zeta_{yj}, \end{aligned} \tag{4.94}$$

where x_i and y_j are the i^{th} and j^{th} element of x and y, respectively. Similarly, (4.85) can be represented as:

$$\begin{aligned} \dot{\hat{x}}_i &= a_i \hat{x}_i + \theta_{xi} H_{xi} + l_{xi}\varsigma_{xi}, \\ \varepsilon \dot{\hat{y}}_j &= b_j \hat{y}_j + \theta_{yj} H_{yj} + l_{yj}\varsigma_{yj}, \end{aligned} \tag{4.95}$$

where $\hat{x}_i$ and $\hat{y}_j$ are the i^{th} and j^{th} element of $\hat{x}$ and $\hat{y}$, respectively, $\theta_{xi} = [w_{1i}, w_{2i}] \in \Re^{1\times(q+1)}$, $\theta_{yj} = [w_{3j}, w_{4j}] \in \Re^{m\times(q+1)} \in \Re^{1\times(q+1)}$, w_{1i} and w_{3j} are the i^{th} and j^{th} row of W_1 and W_3, respectively.

Define the auxiliary system outputs as:

$$\begin{aligned} \tau_{xi} &= \theta^*_{xi} H_{xi} + \zeta_{xi} = \dot{x}_i - a_i x_i, \\ \tau_{yj} &= \theta^*_{yj} H_{yj} + \zeta_{yj} = \varepsilon \dot{y}_j - b_j y_j, \end{aligned} \tag{4.96}$$

and define the estimated auxiliary system outputs as:

$$\begin{aligned} \hat{\tau}_{xi} &= \theta_{xi} H_{xi}, \\ \hat{\tau}_{yj} &= \theta_{yj} H_{yj}. \end{aligned} \tag{4.97}$$

Define the auxiliary output errors as:

$$\begin{aligned} e_{xi} =& \tau_{xi} - \hat{\tau}_{xi} = \tilde{\theta}_{xi} H_{xi} + \zeta_{xi}, \\ e_{yj} =& \tau_{yj} - \hat{\tau}_{yj} = \tilde{\theta}_{yj} H_{yj} + \zeta_{yj}, \end{aligned} \tag{4.98}$$

where $\tilde{\theta}_{xi} = \theta^*_{xi} - \theta_{xi}$, $\tilde{\theta}_{yj} = \theta^*_{yj} - \theta_{yj}$. Subtracting (4.95) from (4.94), and using (4.86) and (4.98), it follows that:

$$\begin{aligned} e_{xi} &= \dot{\varsigma}_{xi} - (a_i - l_{xi})\varsigma_{xi}, \\ e_{yj} &= \varepsilon \dot{\varsigma}_{yj} - (b_j - l_{yj})\varsigma_{yj}. \end{aligned} \tag{4.99}$$

Remark 4.7. *The objective of the NN identification scheme is to train the RHONN so that the weights θ_{xi}, θ_{yj} will converge to their nominal values θ^*_{xi}, θ^*_{yj}, the outputs of the RHONN model will trace the outputs of the nonlinear SPS, and the identification errors $\varsigma_{xi}, \varsigma_{yj}$ will be bounded and minimized. This goal can be achieved by minimizing the output errors e_{xi}, e_{yj} as indicated by* (4.99).

According to (4.96), $\|\tau_{xi} - \theta^{*}_{xi} H_{xi}\|_2 = \|\zeta_{xi}\|_2 \leq \bar{\zeta}_{xi}$, and $\|\tau_{yj} - \theta^{*}_{yj} H_{yj}\|_2 = \|\zeta_{yj}\|_2 \leq \bar{\zeta}_{yj}$ are valid. Therefore, it is easy to obtain that:

$$\begin{aligned} &\frac{1}{\bar{\zeta}^2_{xi}} \|\tau_{xi} - \theta^{*}_{xi} H_{xi}\|_2^2 \leq 1, \\ &\frac{1}{\bar{\zeta}^2_{yj}} \|\tau_{yj} - \theta^{*}_{yj} H_{yj}\|_2^2 \leq 1. \end{aligned} \tag{4.100}$$

The NN weight vectors $\theta_{xi},\ \theta_{yj}$ can be updated by the following updating laws:

$$\dot{P}_{xi} = g_{xi} P_{xi} - \frac{\mu_{xi} P_{xi} \text{diag}(H_{xi})^2 P_{xi}}{(1-\mu_{xi})\bar{\zeta}^2_{xi} + \mu_{xi} H^T_{xi} P_{xi} H_{xi}}, \tag{4.101}$$

$$\dot{\theta}^T_{xi} = \frac{\mu_{xi}}{\bar{\zeta}^2_{xi}} P_{xi} H_{xi} e_{xi}, \tag{4.102}$$

$$\mu_{xi} = \frac{\lambda_{xi} \bar{\zeta}^2_{xi}}{1 + H^T_{xi} P_{xi} H_{xi}}, \tag{4.103}$$

$$\dot{P}_{yj} = g_{yj} P_{yj} - \frac{\mu_{yj} P_{yj} \text{diag}(H_{yj})^2 P_{yj}}{(1-\mu_{yj})\bar{\zeta}^2_{yj} + \mu_{yj} H^T_{yj} P_{yj} H_{yj}}, \tag{4.104}$$

$$\dot{\theta}_{yj}^T = \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2} P_{yj} H_{yj} e_{yj}, \tag{4.105}$$

$$\mu_{yj} = \frac{\lambda_{yj} \bar{\zeta}_{yj}^2}{1 + H_{yj}^T P_{yj} H_{yj}}, \tag{4.106}$$

where $\lambda_{xi},\ \lambda_{yj}$ are the designed parameters such that $0 < \lambda_{xi}\bar{\zeta}_{xi}^2 < 1,\ 0 < \lambda_{yj}\bar{\zeta}_{yj}^2 < 1$. $g_{xi} > 0$, $g_{yj} > 0$ are two designed parameters satisfying:

$$\begin{aligned} 0 < & g_{xi} < \frac{\lambda_{xi}}{(q+1)(1+\lambda_{xi})}, \\ 0 < & g_{yj} < \frac{\lambda_{yj}}{(q+1)(1+\lambda_{yj})}. \end{aligned} \tag{4.107}$$

Lemma 4.2. *By using the weight's updating laws presented in* (4.101)-(4.106)*, it can be guaranteed that P_{xi} and P_{yj} will remain diagonal positive definite and bounded during the identification process as long as the initial values $P_{xi}(t_0)$ and $P_{yj}(t_0)$ are diagonal positive definite, and the g_{xi} and g_{yj} satisfying the condition given in* (4.107)*. Also, it is easy to show that P_{xi}, P_{yj} will converge to their equilibrium points P_{xie} and P_{yje}, and $H_{xi}^T P_{xie} H_{xi}$, $H_{yj}^T P_{yje} H_{yj}$ will satisfy the following equations:*

$$\begin{aligned} H_{xi}^T P_{xie} H_{xi} &= \frac{g_{xi}(q+1)(1-\lambda_{xi}\bar{\zeta}_{xi}^2)}{\lambda_{xi} - g_{xi}(q+1)(1+\lambda_{xi})}, \\ H_{yj}^T P_{yje} H_{yj} &= \frac{g_{yj}(q+1)(1-\lambda_{yj}\bar{\zeta}_{yj}^2)}{\lambda_{yj} - g_{yj}(q+1)(1+\lambda_{yj})}. \end{aligned} \tag{4.108}$$

Proof. Assuming that P_{xie} is the equilibrium of P_{xi} and is positive definite, it follows that:

$$\begin{aligned} \dot{P}_{xie} &= g_{xi} P_{xie} - \frac{\mu_{xi} P_{xie} \text{diag}(H_{xi})^2 P_{xie}}{(1-\mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi} H_{xi}^T P_{xie} H_{xi}} = 0 \\ \Rightarrow g_{xi} P_{xie} &= \frac{\mu_{xi} P_{xie} \text{diag}(H_{xi})^2 P_{xie}}{(1-\mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi} H_{xi}^T P_{xie} H_{xi}}. \end{aligned} \tag{4.109}$$

By right multiplying both sides of (4.109) with P_{xie}^{-1}, one can obtain that:

$$\begin{aligned} g_{xi} \boldsymbol{I} &= \frac{\mu_{xi} P_{xie} \text{diag}(H_{xi})^2}{(1-\mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi} H_{xi}^T P_{xie} H_{xi}} \\ \Rightarrow tr(g_{xi} \boldsymbol{I}) &= tr\left(\frac{\mu_{xi} P_{xie} \text{diag}(H_{xi})^2}{(1-\mu_{xi})\bar{\zeta}_{xi}^2 + \mu_{xi} H_{xi}^T P_{xie} H_{xi}}\right). \end{aligned} \tag{4.110}$$

Substituting (4.103) into (4.110), and using the fact that $tr(P_{xie}\text{diag}(H_{xi})^2) = H_{xi}^T P_{xie} H_{xi}$, it can be obtained that

$$H_{xi}^T P_{xie} H_{xi} = \frac{g_{xi}(q+1)(1-\lambda_{xi}\bar{\zeta}_{xi}^2)}{\lambda_{xi} - g_{xi}(q+1)(1+\lambda_{xi})}. \tag{4.111}$$

Because $0 < \lambda_{xi}\bar{\zeta}_{xi}^2 < 1$, in order to guarantee that $H_{xi}^T P_{xie} H_{xi}$ is positive, $\lambda_{xi} - g_{xi}(q+1)(1+\lambda_{xi}) > 0$ should be satisfied, i.e.,

$$0 < g_{xi} < \frac{\lambda_{xi}}{(q+1)(1+\lambda_{xi})}. \tag{4.112}$$

Next, it can be proved that each single element in the P_{xi} will converge to its equilibrium point. Let p_{xr} denotes the r^{th} diagonal element of P_{xi}, and h_{xr} be r^{th} element of H_{xi}. Using (4.101) and (4.103), it can be obtained that

$$\dot{p}_{xr} = g_{xi} p_{xr} - \frac{\lambda_{xi} p_{xr}^2 h_{xr}^2}{(1-\lambda_{xi}\bar{\zeta}_{xi}^2) + (1+\lambda_{xi}) H_{xi}^T P_{xi} H_{xi}}.$$

Assuming that $\dot{p}_{xr} > 0$, and using the fact that $p_{xr}(t_0) > 0$, $1 - \lambda_{xi}\bar{\zeta}_{xi} > 0$, one has

$$\begin{aligned}
&g_{xi} p_{xr} - \frac{\lambda_{xi} p_{xr}^2 h_{xr}^2}{(1-\lambda_{xi}\bar{\zeta}_{xi}^2) + (1+\lambda_{xi}) H_{xi}^T P_{xi} H_{xi}} > 0 \\
\Leftrightarrow & g_{xi}(1-\lambda_{xi}\bar{\zeta}_{xi}) + g_{xi}(1+\lambda_{xi}) H_{xi}^T P_{xi} H_{xi} > \lambda_{xi} p_{xr}^2 h_{xr}^2 \\
\Leftrightarrow & \big(\lambda_{xi} - g_{xi}(1+\lambda_{xi})\big) p_{xr} h_{xr}^2 \\
& < g_{xi}(1-\lambda_{xi}\bar{\zeta}_{xi}) + g_{xi}(1+\lambda_{xi}) \sum_{k=1, k\neq r}^{q+1} p_{xk} h_{xk}^2.
\end{aligned} \tag{4.113}$$

If (4.112) is satisfied, it is easy to show that $\lambda_{xi} - g_{xi}(1+\lambda_{xi}) > 0$. Thus, (4.113) can be reformulated as:

$$p_{xr} < \frac{g_{xi}(1-\lambda_{xi}\bar{\zeta}_{xi}) + g_{xi}(1+\lambda_{xi}) \sum_{k=1,k\neq r}^{q+1} p_{xk} h_{xk}^2}{\big(\lambda_{xi} - g_{xi}(1+\lambda_{xi})\big) h_{xr}^2} \equiv p_{xre}. \tag{4.114}$$

Hence, if $p_{xr} < p_{xre}$, one has $\dot{p}_{xr} > 0$. If $p_{xr} > p_{xre}$, it can be obtained that $\dot{p}_{xr} < 0$. If $p_{xr} = p_{xre}$, $\dot{p}_{xr} = 0$ will be obtained and the convergence of p_{xr} is achieved. This implies that if only $p_{xr}(t_0) > 0$, $1 - \lambda_{xi}\bar{\zeta}_{xi} > 0$, and (4.107) are satisfied, p_{xr} will always converge to its equilibrium p_{xre} which is a positive number. Hence, it can be concluded that P_{xi} will remain diagonal positive definite. Following the same procedure, it is easy to show that if only $P_{yj}(t_0) > 0$, $1 - \lambda_{xi}\bar{\zeta}_{xi} > 0$, and (4.107) are satisfied, P_{yj} will also remain diagonal positive definite. The proof is thus completed. □

Theorem 4.3. *Consider the multi-time-scale RHONN model* (4.85) *for the nonlinear SPS* (4.84). *By using the updating laws* (4.101)-(4.106)*, it can be guaranteed that the NN weight errors $\tilde{\theta}_{xi}$, $\tilde{\theta}_{yj}$ will be uniformly ultimately bounded, and the identification errors ς_x, ς_y will also be bounded.*

Proof. Consider the following Lyapunov function:

$$V_{yj} = \tilde{\theta}_{yj} P_{yj}^{-1} \tilde{\theta}_{yj}^T. \tag{4.115}$$

The derivative of V_{yj} is given as:

$$\dot{V}_{yj} = 2\tilde{\theta}_{yj} P_{yj}^{-1} \dot{\tilde{\theta}}_{yj}^T - \tilde{\theta}_{yj} \frac{dP_{yj}^{-1}}{dt} \tilde{\theta}_{yj}^T. \tag{4.116}$$

It is noticed that

$$\frac{d}{dt}\left[P_{yj} P_{yj}^{-1}\right] = \dot{P}_{yj} P_{yj}^{-1} + P_{yj} \frac{d}{dt} P_{yj}^{-1} = \frac{dI}{dt} = 0. \tag{4.117}$$

Using (4.104), it can be obtained that:

$$\frac{d}{dt} P_{yj}^{-1} = -P_{yj}^{-1} \dot{P}_{yj} P_{yj}^{-1} = -g_{yj} P_{yj}^{-1} + \frac{\mu_{yj} \text{diag}(H_{yj})^2}{(1-\mu_{yj})\bar{\zeta}_{yj}^2 + \mu_{yj} H_{yj}^T P_{yj} H_{yj}}. \tag{4.118}$$

Substituting (4.105) and (4.118) into (4.116), it follows that:

$$\begin{aligned}\dot{V}_{yj} =& -\frac{2\mu_{yj}}{\bar{\zeta}_{yj}^2}\tilde{\theta}_{yj}H_{yj}e_{yj} + \frac{\mu_{yj}\tilde{\theta}_{yj}\text{diag}(H_{yj})^2\tilde{\theta}_{yj}^T}{(1-\mu_{yj})\bar{\zeta}_{yj}^2 + \mu_{yj}H_{yj}^T P_{yj}H_{yj}} - \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}\|\tau_{yj} - \theta_{yj}^* H_{yj}\|_2^2 \\ &+ \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}\tilde{\theta}_{yj}H_{yj}H_{yj}^T\tilde{\theta}_{yj}^T + \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}\|\tau_{yj} - \theta_{yj}^* H_{yj}\|_2^2 - \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}\tilde{\theta}_{yj}H_{yj}H_{yj}^T\tilde{\theta}_{yj}^T - g_{yj}\tilde{\theta}_{yj}P_{yj}^{-1}\tilde{\theta}_{yj}^T \\ =&\Big[-\frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}\|\tau_{yj} - \theta_{yj}^* H_{yj}\|_2^2 + \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}\tilde{\theta}_{yj}H_{yj}H_{yj}^T\tilde{\theta}_{yj}^T - \frac{2\mu_{yj}}{\bar{\zeta}_{yj}^2}\tilde{\theta}_{yj}H_{yj}e_{yj}\Big] - g_{yj}\tilde{\theta}_{yj}P_{yj}^{-1}\tilde{\theta}_{yj}^T \\ &+ \frac{\mu_{yj}\tilde{\theta}_{yj}\text{diag}(H_{yj})^2\tilde{\theta}_{yj}^T}{(1-\mu_{yj})\bar{\zeta}_{yj}^2 + \mu_{yj}H_{yj}^T P_{yj}H_{yj}} - \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}\tilde{\theta}_{yj}H_{yj}H_{yj}^T\tilde{\theta}_{yj}^T + \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}\|\tau_{yj} - \theta_{yj}^* H_{yj}\|_2^2. \end{aligned} \tag{4.119}$$

Using the fact that $e_{yj} = \tau_{yj} - \theta_{yj}H_{yj}$ and $\tilde{\theta}_{yj} = \theta_{yj}^* - \theta_{yj}$, the equation in $[\cdot]$ in (4.119) can be rewritten as:

$$\begin{aligned}&-\frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}\|\tau_{yj}-\theta_{yj}^* H_{yj}\|_2^2+\frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}\tilde{\theta}_{yj}H_{yj}H_{yj}^T\tilde{\theta}_{yj}^T-\frac{2\mu_{yj}}{\bar{\zeta}_{yj}^2}\tilde{\theta}_{yj}H_{yj}e_{yj} \\ =&\frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}\big[-\|\tau_{yj}-\theta_{yj}^* H_{yj}\|_2^2+(\theta_{yj}^*-\theta_{yj})H_{yj}H_{yj}^T(\theta_{yj}^*-\theta_{yj})^T-2(\theta_{yj}^*-\theta_{yj})H_{yj}(\tau_{yj}-\theta_{yj}H_{yj})\big] \\ =&-\frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}\big[\tau_{yj}^2-2\theta_{yj}H_{yj}\tau_{yj}+(\theta_{yj}H_{yj})^2\big]=-\frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}e_{yj}^2. \end{aligned} \tag{4.120}$$

Substituting (4.120) into (4.119), using (4.100), and noticing that $0 < \mu_{yj} < 1$, $\tilde{\theta}_{yj}\text{diag}(H_{yj})^2\tilde{\theta}_{yj}^T \leq \tilde{\theta}_{yj}H_{yj}H_{yj}^T\tilde{\theta}_{yj}^T$, the following inequality is obtained:

$$\begin{aligned}\dot{V}_{yj} \leq& \frac{\mu_{yj}(\tilde{\theta}_{yj}H_{yj})^2}{(1-\mu_{yj})\bar{\zeta}_{yj}^2 + \mu_{yj}H_{yj}^T P_{yj}H_{yj}} - \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}(\tilde{\theta}_{yj}H_{yj})^2 + \mu_{yj} - g_{yj}\tilde{\theta}_{yj}P_{yj}^{-1}\tilde{\theta}_{yj}^T - \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}e_{yj}^2 \\ =& -k_1(\tilde{\theta}_{yj}H_{yj})^2 - \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}(e_{yj}^2 - \bar{\zeta}_{yj}^2) - g_{yj}V_{yj}, \end{aligned} \tag{4.121}$$

where

$$k_1 = \frac{[H_{yj}^T P_{yj}H_{yj} - \bar{\zeta}_{yj}^2]\mu_{yj}^2}{\bar{\zeta}_{yj}^2[(1-\mu_{yj})\bar{\zeta}_{yj}^2 + \mu_{yj}H_{yj}^T P_{yj}H_{yj}]}.$$

By choosing suitable g_{yj} and activation functions $\psi_i,\ i = 1, \cdots, 4$, it can be guaranteed that $H_{yj}^T P_{yj}H_{yj} > \bar{\zeta}_{yj}^2$ is always valid. Thus, $k_1 > 0$ is always true. Substituting (4.98) into (4.121), it

can be obtained that

$$\begin{aligned}\dot{V}_{yj} &\leq -g_{yj}V_{yj} - (k_1 + \frac{\mu_{yj}}{\bar{\zeta}^2})(\tilde{\theta}_{yj}H_{yj})^2 - \frac{2\mu_{yj}}{\bar{\zeta}^2}\tilde{\theta}_{yj}H_{yj}\zeta + \mu_{yj}(1 - \frac{\zeta^2}{\bar{\zeta}^2}) \\ &= -g_{yj}V_{yj} - (k_2\tilde{\theta}_{yj}H_{yj} + k_3\zeta)^2 + k_4 \\ &\leq -g_{yj}V_{yj} + \bar{k}_4. \end{aligned} \tag{4.122}$$

where $k_2 = \sqrt{k_1 + \frac{\mu_{yj}}{\bar{\zeta}_{yj}^2}}$, $k_3 = \frac{\mu_{yj}}{\bar{\zeta}^2 k_2}$, $k_4 = \mu_{yj}(1 - \frac{\zeta^2}{\bar{\zeta}^2}) + \frac{\mu_{yj}\zeta_{yj}^2}{\bar{\zeta}^4 k_2^2}$, and $\bar{k}_4$ is the upper bound of k_4. From (4.122), it is clear that the $\tilde{\theta}_{yj}$ is UUB. Using (4.98), it can be obtained that the auxiliary output error e_{yj} will also be bounded, which implies the identification error ς_{yj} is also bounded according to (4.99). Following the same procedure, it is easy to show that $\tilde{\theta}_{xi}$ is also UUB, and the auxiliary output error e_{xi} and identification error ς_{xi} will also be bounded. Theorem 4.3 is thus proved. □

Remark 4.8. *In* (4.121)*, when* $|e_{yj}| > \bar{\zeta}_{yj}$*, one can obtain that* $\dot{V} < -g_{yj}V_{yj}$*. Thus,* $\tilde{\theta}_{yj}$ *will decrease exponentially. However, in [16, 27, 32, 112], because the identification algorithms were based on GD, only the asymptotic stability can be guaranteed, and the NN weight errors proposed in those papers would decrease asymptotically. Hence, the modified OBE based identification algorithm proposed in this section can achieve faster convergence speed than the GD based algorithms proposed in [16, 27, 32, 112]. This conclusion will be further validated later in the experiment.*

Remark 4.9. *The main difference between the modified OBE algorithm proposed in this section and the OBE algorithm proposed in Section 4.3 and in [31, 36] is that in this section, two additional terms* $g_{xi}P_{xi}$*,* $g_{yj}P_{yj}$ *are introduced into* (4.101) *and* (4.104)*. Hence, the terms* P_{xi} *and* P_{yj} *will converge to the equilibrium points* P_{xie} *and* P_{yje} *which are determined by* g_{xi}*,* H_{xi}*,* g_{yj}*, and* H_{yj}*, as shown in* (4.108)*. Thus, the NN weight's updating laws will remain effective during the identification process. In the OBE based learning laws proposed Section 4.3 and in [31, 36], the two terms* $g_{xi}P_{xi}$ *and* $g_{yj}P_{yj}$ *are not considered. Hence,* $\dot{P}_{xi}$ *and* $\dot{P}_{yj}$ *will always be negative definite, and* P_{xi}*,* P_{yj} *will converge to* $\mathbf{0}$*. As a result, the weight's updating laws will lose the ability to adjust the NN weights according to* (4.102) *and* (4.105)*.*

4.4.2 Experiment

To verify the effectiveness of the proposed identification algorithms, the experiments on a harmonic drive system are conducted. The experimental setup is shown in Fig. 4.14. As shown in Fig. 4.14, a metal disc (the load) with unknown moment of inertia is mounted to axes of the DC servo actuator. The actuator is driven by the PWM servo drive. A DC power supply (maximum output 60 volts) is used as the power source of the servo drive. The identification algorithms are running in the dSPACE. The dSPACE measures the angular velocity through an encoder (1000 PPR), and the current through an A/D converter (16 bits, $\pm$10V), and sends the control signal to the servo drive through an D/A converter (16 bits, $\pm$10V). The whole process are monitored and recorded using the ControlDesk running on a PC with Windows 7 operation system. In this harmonic drive system, the angular velocity is the slow dynamic state and the current is the fast dynamic state. The objective of the experiment is to on-line identify the unknown system model of the harmonic drive.

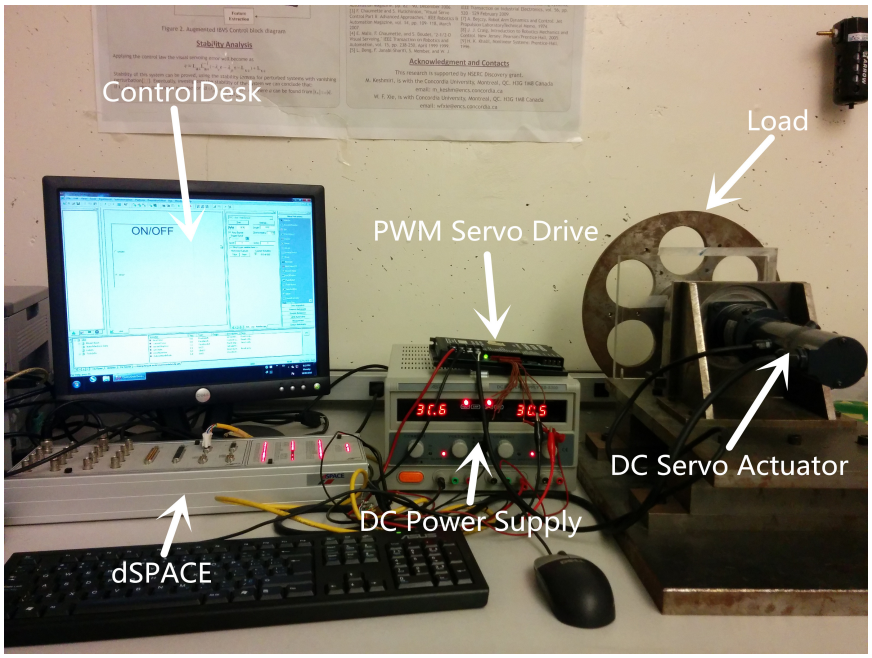


Figure 4.14: Experimental setup of the harmonic drive system.

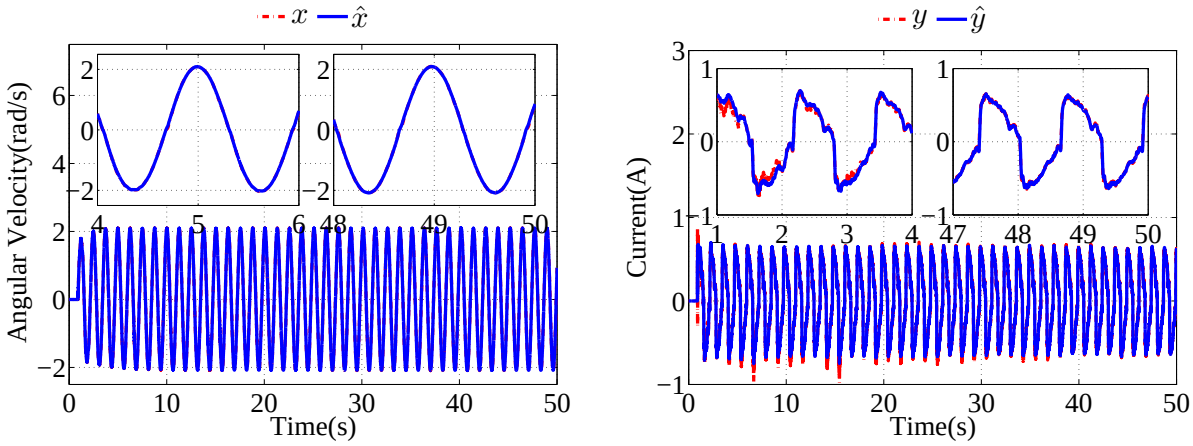
A typical DC motor can be described by the following model:

$$\begin{aligned} J\frac{d\omega}{dt} &= k_t i, \\ L\frac{di}{dt} &= -k_b\omega - Ri + V, \end{aligned} \tag{4.123}$$

where J is the moment of inertia, ω is the angular velocity, k_t is the torque force constant, i is the armature current, L is the armature inductance, k_b is the back electromotive force constant, R, V are the armature resistance and voltage, respectively. However, generally speaking, this model is not faithful because the system parameters may vary from time to time. For example, in many cases, the load is not fixed and will change during different working conditions. The condition of the connecting points of the circuit changes due to aging and rust, thus increase the circuit resistance. Also, the magnetic field strength will decrease gradually if the permanent magnet is used. Besides, there will be some nonlinearity in the real systems, such as friction and backlash [113]. Define x as the angular velocity, y as the current, and u as the control signal. Therefore, in this experiment, the NN model (4.85) with modified OBE algorithm based updating laws (4.101)-(4.106) will be used to identify the unknown DC motor system.

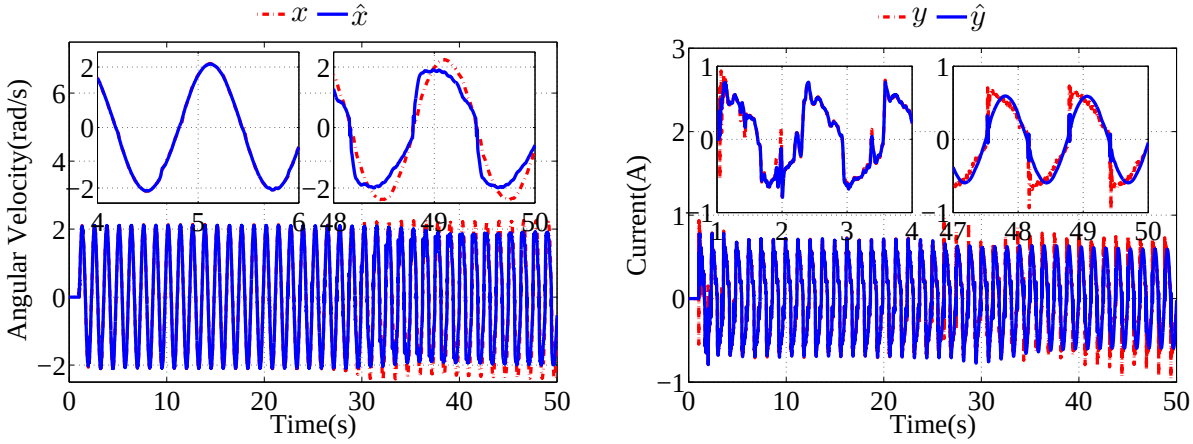
In this experiment, $\varepsilon = 0.1$ is obtained based on the prior knowledge of the setup, and the following NN parameters are used: $A = -20$, $B = -20$, $\lambda_x = 500$, $\lambda_y = 500$, $L_x = 20$, $L_y = 10$, $g_x = g_y = 0.332$, $\bar{\zeta}_x = 0.001$, $\bar{\zeta}_y = 0.001$, the number of neuron $q = 2$, $W_1 \in \Re^{1\times 2}$, $W_2 \in \Re$, $W_3 \in \Re^{1\times 2}$, $W_4 \in \Re$, $W_1(t_0) = W_3(t_0) = \mathbf{0} \in \Re^{1\times 2}$, $W_2(t_0) = W_4(t_0) = 0.5$, $P_{xi}(t_0) = \text{diag}([50, 50, 50])$, $P_{yj}(t_0) = \text{diag}([50, 50, 50])$, the activation functions are chosen as $\psi_1(z) = 2/(1 + e^{-0.5z}) + 1$, $\psi_2(z) = 1/(1 + e^{-0.2z}) + 10$, $\psi_3(z) = 2/(1 + e^{-0.1z}) + 1$, $\psi_4(z) = 1/(1 + e^{-0.1z}) + 5$. To demonstrate the superiority, the OBE based identification scheme proposed in Section 4.3, the GD based identification scheme proposed in [112] are also tested for comparison purpose. The sampling time for the experiment is 0.1 ms. The experimental results are presented in Figs. 4.15-4.20.

From Figs. 4.15-4.20, it is clear that the modified OBE based identification algorithm can achieve the best performance among all results. Fig. 4.15 shows that when the modified OBE



(a) Identification result of x. (b) Identification result of y.

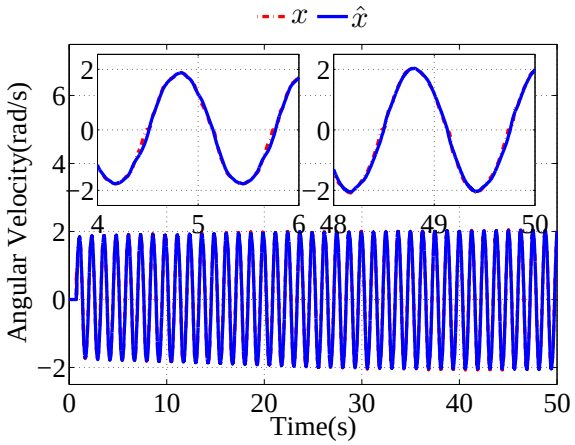
Figure 4.15: Identification results using the modified OBE.



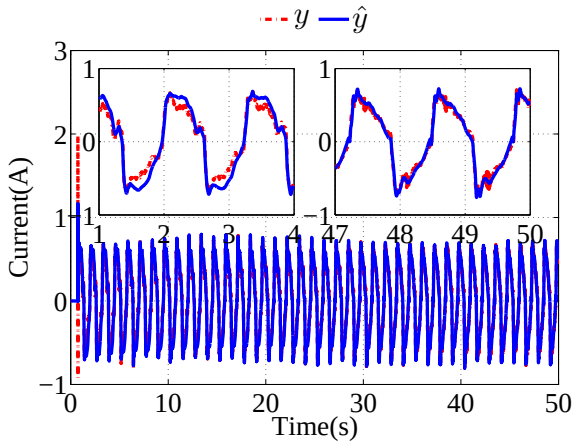
(a) Identification result of x. (b) Identification result of y.

Figure 4.16: Identification results using the original OBE.

based identification scheme is used, there is a visible difference between the real current y, and the estimation $\hat{y}$ at beginning. However, after a short learning period (about 4s), the difference among them became negligible. The difference between x and $\hat{x}$ is small since the very beginning. When the original OBE based algorithm proposed in Section 4.3 is used, the identification performances are very good at the beginning. The $\hat{y}$ can converge to y even faster than the case when the modified OBE based method is used. The difference between x and $\hat{x}$ is also very small since the very beginning. Nevertheless, after 25 seconds, because P_x and P_y converge to $\mathbf{0}^{3\times3}$, the weight's updating laws are no longer effective, and large gaps between the real signals and their estimations

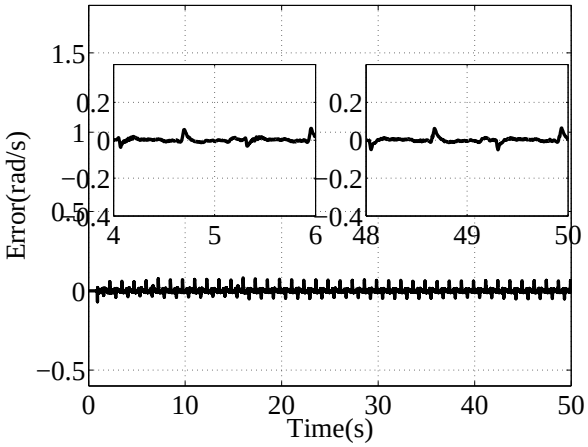


(a) Identification result of x.

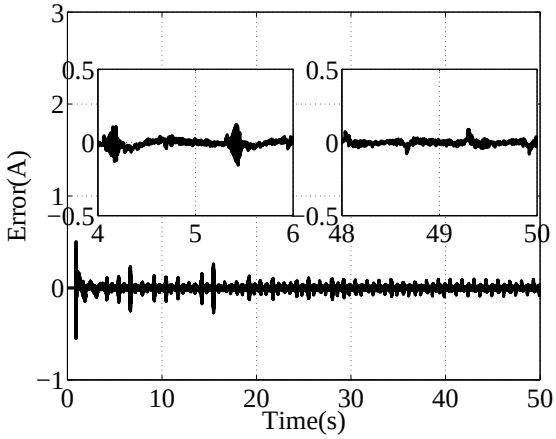


(b) Identification result of y.

Figure 4.17: Identification results using the gradient descent.



(a) Identification error of x.

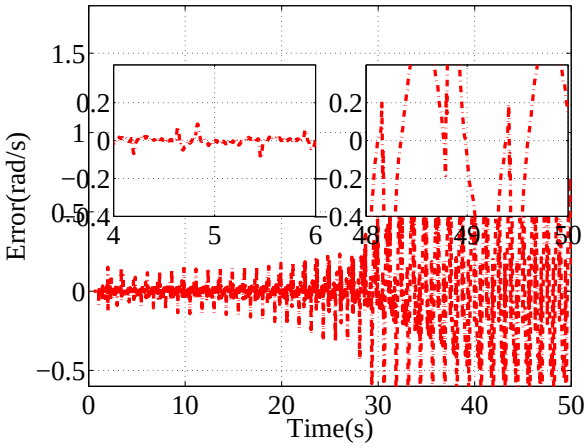


(b) Identification error of y.

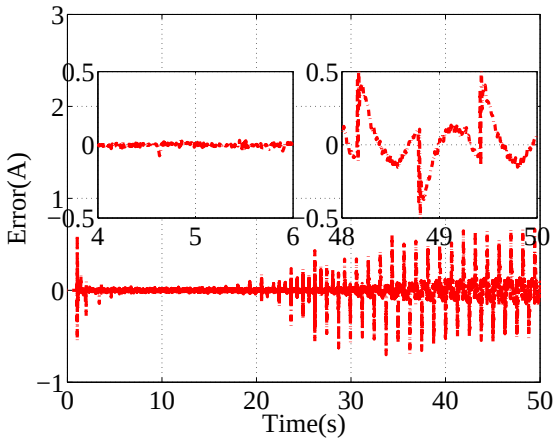
Figure 4.18: Identification errors using the modified OBE.

can be observed. When GD based algorithm is used, it takes a much longer time for the $\hat{x}$ and $\hat{y}$ to converge to x and y, as shown in Fig. 4.17. Even at the end of the experiment, the differences between the real signals and the their estimations are obvious.

From Figs. 4.18-4.20, it is also clear that the least identification errors of x and y can be obtained when the modified OBE based identification scheme is used. In Fig. 4.19, when the original OBE based identification scheme is used, the identification errors are small at the beginning, but increased significantly after about 25 seconds. When the gradient descent based identification scheme is used, larger identification errors can be observed during the whole identification process.

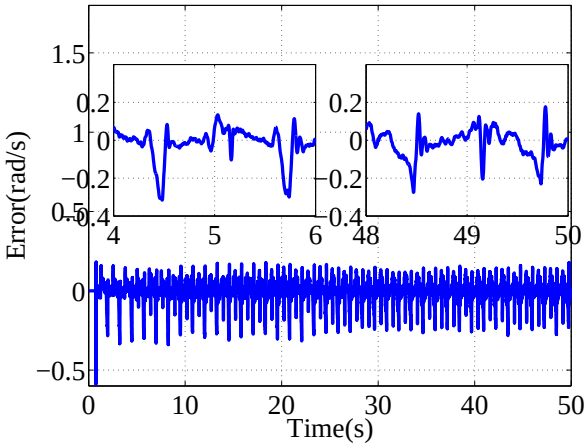


(a) Identification error of x.

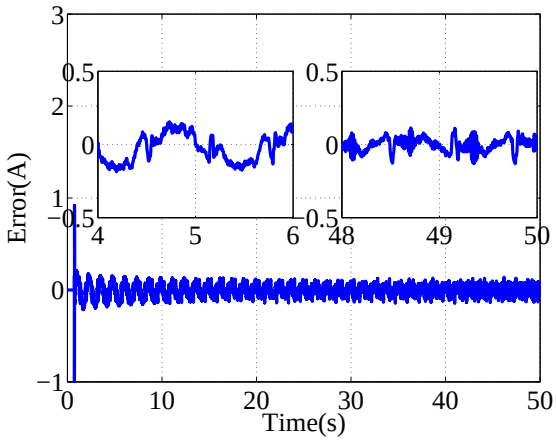


(b) Identification error of y.

Figure 4.19: Identification errors using the original OBE.



(a) Identification error of x.



(b) Identification error of y.

Figure 4.20: Identification errors using the GD.

The figures of the weight's updating process are presented in Figs. 4.21-4.23. From Fig. 4.22, it can be noticed that after 40 seconds, the original OBE based weight's updating laws stop working, and the weights of the NN remain constant. This is because P_{xi} and P_{yj} are close to $\mathbf{0}$, and the leaning gains of the original OBE based weight's updating laws are too small to adjust the NN weights. When the modified OBE is used, the weight vector Θ_y converges in less than 60 seconds, and the weight vector Θ_x almost converges in 100 seconds, as shown in Fig. 4.21. However, when GD based weight's updating laws are used, it takes more than 100 seconds for the Θ_y to converge, and there is no sign of convergence for Θ_x even after 200 seconds, as depicted in Fig. 4.23. Thus

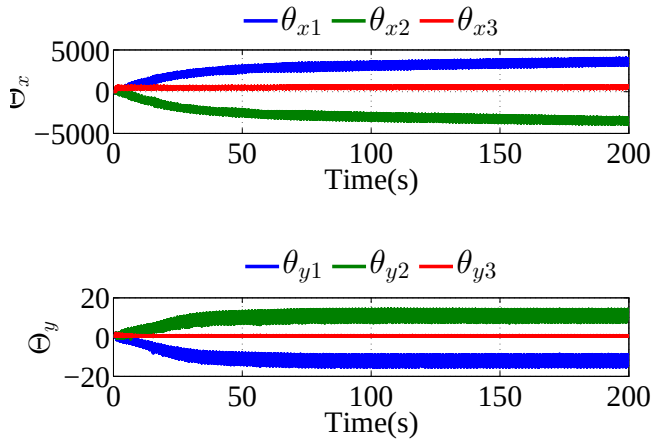


Figure 4.21: Weights updating process using modified OBE.

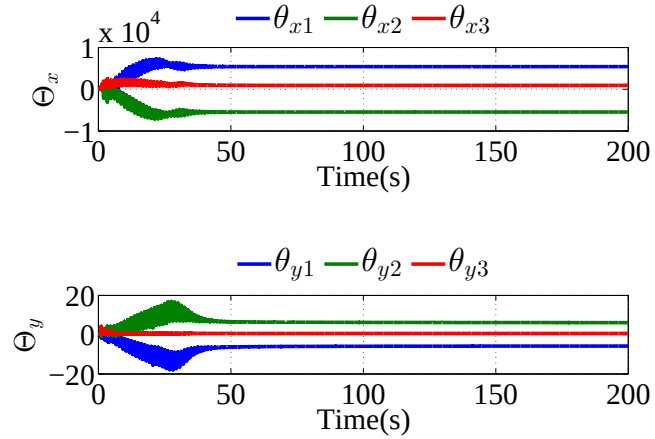


Figure 4.22: Weights updating process using OBE.

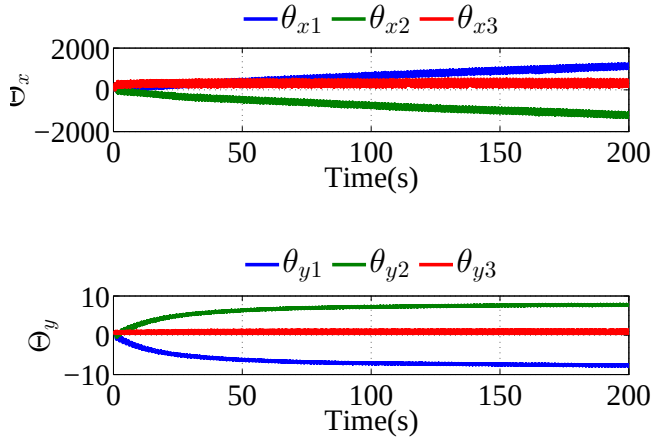


Figure 4.23: Weights updating process using GD.

by using the modified OBE based weight's updating laws, the faster convergence can be achieved, and the updating laws will remain effective during the whole identification process.

To further compare the performance of different identification methods, the performance index-ITAE is also calculated. The ITAE is defined as:

$$\text{ITAE} = \int_0^T t|e(t)|dt. \tag{4.124}$$

The results of the ITAE calculation are presented in Table 4.2, where, ITAE_{xi}, ITAE_{yi} are the ITAE values of the identification error ς_x and identification error ς_y, respectively. From Table 4.2, it is clear that the identification errors using the modified OBE proposed in this section are much smaller compared with the errors using the other methods.

Table 4.2: ITAE values of ς_x and ς_y

	Modified OBE	OBE	GD
$ITAE_{xi}$	13.8	534.9	94.2
$ITAE_{yi}$	24.1	152.8	73.3

4.5 Conclusion

In this chapter, an OBE based identification scheme is firstly proposed for the discrete nonlinear singularly perturbed systems using multi-time-scale neural networks. A faster convergence can be achieved with higher accuracy when the OBE based weight's updating laws are used, because the learning gain will be adaptively adjusted. On the contrary, the convergence speed is slow when the conventional GD based identification scheme is used because of the fixed learning gain. Subsequently, the discrete identification scheme is extended to a continuous case. However, it is found that the learning gain of the weight's updating laws using the continuous OBE based identification scheme will decrease during the identification process and thus lose the ability to

adjust the neural network weights when the learning gain is too small. To solve this problem, a modified OBE based identification algorithm is proposed where two extra terms are added to the weight's updating laws such that the P_x and P_y will converge to the equilibriums which can be set arbitrarily by the user, rather than $\mathbf{0}$.

Chapter 5

Robust Identification Scheme of Nonlinear SPSs Using Filtered Variables

5.1 Introduction

In previous chapter, the discrete and continuous OBE based identification schemes using multi-time-scale NNs are discussed. However, it should be pointed out that in these schemes, the derivatives of the identification errors are needed in the weight's updating laws, which are usually obtained by differentiating the identification errors directly. If the measurement noises are involved in the system states, the identification accuracy would be severely undermined.

In [114], a filter was adopted to generate the regressor matrix for a finite-time parameter estimation problem. By using this filter, the measurement or computation of the velocity state vector was not required. Afterwards, Na et al. proposed the robust adaptive parameter estimation algorithms for a class of nonlinear robotic systems in [115, 116], in which a set of auxiliary filtered variables was introduced to obtain an expression of the parameter estimation error. The parameter estimation problem of nonlinear system with completely unknown dynamics or sinusoidal signals were also discussed in [117, 118].

Inspired by their works, in this chapter, a new identification algorithm for a class of nonlinear

SPSs using multi-time-scale recurrent neural network is proposed. A set of filtered variables are firstly defined and incorporated into the NN. Then, the auxiliary output errors are derived, and the augmented OBE algorithm is proposed to train the RNN with filtered variables. Thus, the derivatives of the identification errors are no longer needed, and the proposed identification scheme is more robust to measurement noises.

5.2 Identification Algorithm

Consider the following unknown affine in control nonlinear SPS:

$$\begin{aligned} \dot{x} &= f_1(x) + f_2(x)y, \\ \varepsilon\dot{y} &= g_1(x,y) + g_2(x,y)u, \end{aligned} \tag{5.1}$$

where $x \in \Re^n$ and $y \in \Re^m$ are the slow and fast states, respectively, $u \in \Re^p$ is the control signal vector, $0 < \varepsilon < 1$ is a small parameter, $f(\cdot),\ f_2(\cdot),\ g_1(\cdot),\ g_2(\cdot)$ are unknown smooth functions.

The nonlinear system (5.1) can be represented by the following nominal multi-time-scale RNN:

$$\begin{aligned} \dot{x} &= Ax + W_1^*\Psi_1(x) + W_2^*\Psi_2(x)y + \zeta_x, \\ \varepsilon\dot{y} &= By + W_3^*\Psi_3(x,y) + W_4^*\Psi_4(x,y)u + \zeta_y, \end{aligned} \tag{5.2}$$

where $A \in \Re^{n\times n}$, $B \in \Re^{m\times m}$ are stable matrices, $\zeta_x \in \Re^n$, $\zeta_y \in \Re^m$ are the modeling errors, $W_1^* \in \Re^{n\times q}$, $W_2^* \in \Re^{n\times q}$, $W_3^* \in \Re^{m\times q}$, $W_4^* \in \Re^{m\times q}$ are unknown optimal weights which minimize ζ_x and ζ_y, q is the number of neuron, the activation function vectors $\Psi_i(\cdot),\ i = 1, 3$ are defined as:

$$\begin{aligned} \Psi_i(\cdot) &= [\Psi_{i,1}, \Psi_{i,2}, \cdots, \Psi_{i,q}]^T \in \Re^q, \\ \Psi_{i,c}(\cdot) &= \prod_{j\in J_c} [\psi_i(\cdot)]^{d_c(j)},\ c = 1, \cdots, q, \end{aligned}$$

the activation function matrix $\Psi_2(\cdot)$ is defined as:

$$\Psi_2(\cdot) = \begin{bmatrix} \Psi_{2,1,1} & \cdots & \Psi_{2,1,m} \\ \vdots & \ddots & \vdots \\ \Psi_{2,q,1} & \cdots & \Psi_{2,q,m} \end{bmatrix} \in \Re^{q\times m},$$
$$\Psi_{2,c,r} = \prod_{j\in J_{c,r}} [\psi_2(\cdot)]^{d_{c,r}(j)}, c = 1,\cdots,q, r = 1,\cdots,m,$$

where $J_{c,r}$ are the collections of l not ordered subsets of $1, 2, \cdots, n+m$ and $d_{c,r}(j)$ are non-negative integers, and the activation function matrix $\Psi_4(\cdot)$ is defined as:

$$\Psi_4(\cdot) = \begin{bmatrix} \Psi_{4,1,1} & \cdots & \Psi_{4,1,p} \\ \vdots & \ddots & \vdots \\ \Psi_{4,q,1} & \cdots & \Psi_{4,q,p} \end{bmatrix} \in \Re^{q\times p},$$
$$\Psi_{4,c,r} = \prod_{j\in J_{c,r}} [\psi_4(\cdot)]^{d_{c,r}(j)}, c = 1,\cdots,q, r = 1,\cdots,p,$$

where $J_{c,r}$ are the collections of l not ordered subsets of $1, 2, \cdots, n+m$ and $d_{c,r}(j)$ are non-negative integers. The activation functions $\psi_i(\cdot),\ i = 1,\cdots,4$ are chosen as:

$$\psi_i(z) = \frac{\alpha_{i,1}}{1+e^{-\alpha_{i,2}z}} + \alpha_{i,3}. \tag{5.3}$$

Denote $\Theta_x^* = [W_1^*, W_2^*] = [\theta_{x1}^*;\cdots;\theta_{xn}^*] \in \Re^{n\times 2q}$, $\Theta_y^* = [W_3^*, W_4^*] = [\theta_{y1}^*;\cdots;\theta_{ym}^*] \in \Re^{m\times 2q}$, where $\theta_{xi}^* \in \Re^{2q}$, $\theta_{yj}^* \in \Re^{2q}$, $i = 1,\cdots,n$, $j = 1,\cdots,m$ are row vectors, $H_x = [\Psi_1(x); \Psi_2(x)y] \in \Re^{2q}$, $H_y = [\Psi_3(x,y); \Psi_4(x,y)u] \in \Re^{2q}$. Then nominal model (5.2) can be written as:

$$\begin{aligned} \dot{x} &= Ax + \Theta_x^* H_x + \zeta_x, \\ \varepsilon\dot{y} &= By + \Theta_y^* H_y + \zeta_y. \end{aligned} \tag{5.4}$$

Define the filtered variables of x, y, H_x, H_y, ζ_x, ζ_y as:

$$\begin{cases} k_x\dot{x}_f + x_f = x,\ x_f(0) = 0 \\ k_x\dot{H}_{xf} + H_{xf} = H_x,\ H_{xf}(0) = 0\ , \\ k_x\dot{\zeta}_{xf} + \zeta_{xf} = \zeta_x,\ \zeta_{xf}(0) = 0 \end{cases} \tag{5.5}$$

$$\begin{cases} k_y\dot{y}_f + y_f = y,\ y_f(0) = 0 \\ k_y\dot{H}_{yf} + H_{yf} = H_y,\ H_{yf}(0) = 0\ , \\ k_y\dot{\zeta}_{yf} + \zeta_{yf} = \zeta_y,\ \zeta_{yf}(0) = 0 \end{cases} \tag{5.6}$$

where $k_x > 0$, $k_y > 0$ are two filter parameters.

Remark 2: The filtered modeling errors ζ_{xf} and ζ_{yf} are introduced only for theoretical analysis. They are not required in real application.

From (5.4), (5.5) and (5.6), it is easy to obtain that

$$\begin{aligned} \dot{x}_f &= Ax_f + \Theta_x^* H_{xf} + \zeta_{xf}, \\ \varepsilon\dot{y}_f &= By_f + \Theta_y^* H_{yf} + \zeta_{yf}, \end{aligned} \tag{5.7}$$

At this point, the following RNN is proposed to identify the unknown nonlinear system (5.1):

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + \Theta_x H_{xf} + L_x\tilde{x}, \\ \varepsilon\dot{\hat{y}} &= B\hat{y} + \Theta_y H_{yf} + L_y\tilde{y}, \end{aligned} \tag{5.8}$$

where $\hat{x}$ and $\hat{y}$ are the outputs of the RNN, $\Theta_x = [W_1, W_2] = [\theta_{x1}; \cdots; \theta_{xn}] \in \Re^{n\times 2q}$, $\Theta_y = [W_3, W_4] = [\theta_{y1}; \theta_{y2}; \cdots; \theta_{ym}] \in \Re^{m\times 2q}$, with $\theta_{xi} \in \Re^{2q}$, $\theta_{yj} \in \Re^{2q}$, $i = 1, \cdots, n$, $j = 1, \cdots m$ are row vectors, $L_x \in \Re^{n\times n}$, $L_y \in \Re^{m\times m}$ are positive definite matrices, the filtered identification errors $\tilde{x}$, $\tilde{y}$ are defined as:

$$\begin{aligned} \tilde{x} &= x_f - \hat{x}, \\ \tilde{y} &= y_f - \hat{y}. \end{aligned} \tag{5.9}$$

The aim of this section is to develop the novel weight's updating laws for the NN such that the

NN weights $W_i,\ i = 1, \cdots, 4$ will converge to their nominal values W_i^*. It is obvious the filtered identification errors $\tilde{x}$, $\tilde{y}$ will also be minimized when W_i converge to W_i^*.

Subtracting (5.8) from (5.4) yields

$$\begin{aligned} \dot{\tilde{x}} &= (A - L_x)\tilde{x} + \tilde{\Theta}_x H_{xf} + \zeta_{xf}, \\ \varepsilon\dot{\tilde{y}} &= (B - L_y)\tilde{y} + \tilde{\Theta}_y H_{yf} + \zeta_{yf}, \end{aligned} \tag{5.10}$$

where $\tilde{\Theta}_x = \Theta_x^* - \Theta_x$, $\tilde{\Theta}_y = \Theta_y^* - \Theta_y$.

Define the filtered auxiliary system outputs as:

$$\begin{aligned} \tau_{xf} &= \Theta_x^* H_{xf} + \zeta_{xf}, \\ \tau_{yf} &= \Theta_y^* H_{yf} + \zeta_{yf}, \end{aligned} \tag{5.11}$$

and the filtered estimation of the auxiliary system outputs as:

$$\begin{aligned} \hat{\tau}_{xf} &= \Theta_x H_{xf}, \\ \hat{\tau}_{yf} &= \Theta_y H_{yf}, \end{aligned} \tag{5.12}$$

Besides, define the filtered auxiliary output errors as:

$$\begin{aligned} e_{xf} &= \dot{\tilde{x}} - (A - L_x)\tilde{x}, \\ e_{yf} &= \varepsilon\dot{\tilde{y}} - (B - L_y)\tilde{y}. \end{aligned} \tag{5.13}$$

Thus, (5.10) can be rewritten as

$$\begin{aligned} e_{xf} &= \tilde{\Theta}_x H_{xf} + \zeta_{xf}, \\ e_{yf} &= \tilde{\Theta}_y H_{yf} + \zeta_{yf}. \end{aligned} \tag{5.14}$$

Let ζ_{xfi}, ζ_{yfj} denote the i^{th} and j^{th} terms of ζ_{xf} and ζ_{yf}, respectively. The properties of these two errors are assumed as follows.

Assumption 5.1. *The filtered modeling errors ζ_{xfi} and ζ_{yfj} are bounded by upper bounds $\bar{\zeta}_{xfi}$ and $\bar{\zeta}_{yfj}$ as:*

$$0 < |\zeta_{xfi}| < \bar{\zeta}_{xfi},\ 0 < |\zeta_{yfj}| < \bar{\zeta}_{yfj}. \tag{5.15}$$

According to (5.11), $|\tau_{xfi} - \theta_{xi}^* H_{xf}| = |\zeta_{xfi}| \leq \bar{\zeta}_{xfi}$, and $|\tau_{yfj} - \theta_{yj}^* H_{yf}| = |\zeta_{yfj}| \leq \bar{\zeta}_{yfj}$ are valid, where τ_{xfi} and τ_{yfj} are the i^{th} and j^{th} elements of τ_{xf} and τ_{yf}, respectively. Therefore, it is easy to obtain that:

$$\begin{aligned} &\frac{1}{\bar{\zeta}_{xfi}^2}|\tau_{xfi} - \theta_{xi}^* H_{xf}|^2 \leq 1, \\ &\frac{1}{\bar{\zeta}_{yfj}^2}|\tau_{yfj} - \theta_{yj}^* H_{yf}|^2 \leq 1. \end{aligned} \tag{5.16}$$

The following updating laws are proposed to train the RNN weight vectors $\theta_{xi},\ \theta_{yj}$:

$$\dot{P}_{xi} = g_{xi} P_{xi} - \frac{\mu_{xi} P_{xi} \text{diag}(H_{xf})^2 P_{xi}}{(1-\mu_{xi})\bar{\zeta}_{xfi}^2 + \mu_{xi} H_{xf}^T P_{xi} H_{xf}}, \tag{5.17}$$

$$\dot{\theta}_{xi}^T = \frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2} P_{xi} H_{xf} e_{xi}, \tag{5.18}$$

$$\mu_{xi} = \frac{\lambda_{xi}\bar{\zeta}_{xfi}^2}{1 + H_{xf}^T P_{xi} H_{xf}}, \tag{5.19}$$

$$\dot{P}_{yj} = g_{yj} P_{yj} - \frac{\mu_{yj} P_{yj} \text{diag}(H_{yf})^2 P_{yj}}{(1-\mu_{yj})\bar{\zeta}_{yfj}^2 + \mu_{yj} H_{yf}^T P_{yj} H_{yf}}, \tag{5.20}$$

$$\dot{\theta}_{yj}^T = \frac{\mu_{yj}}{\bar{\zeta}_{yfj}^2} P_{yj} H_{yf} e_{yfj}, \tag{5.21}$$

$$\mu_{yj} = \frac{\lambda_{yj}\bar{\zeta}_{yfj}^2}{1 + H_{yf}^T P_{yj} H_{yf}}, \tag{5.22}$$

where e_{xfi}, e_{yfj} are the i^{th} and j^{th} elements of e_{xf} and e_{yf}, respectively, $\lambda_{xi},\ \lambda_{yj}$ are designed parameters such that $0 < \lambda_{xi}\bar{\zeta}_{xfi}^2 < 1,\ 0 < \lambda_{yj}\bar{\zeta}_{yfj}^2 < 1$, g_{xi}, g_{yj} are two designed parameters satisfying

$$0 < g_{xi} < \frac{\lambda_{xi}}{2q(1+\lambda_{xi})},\ 0 < g_{yj} < \frac{\lambda_{yj}}{2q(1+\lambda_{yj})}. \tag{5.23}$$

Lemma 5.1. *By using the weight's updating laws presented in* (5.17)-(5.22)*, it can be guaranteed*

that P_{xi} and P_{yj} will remain diagonal positive definite and bounded during the identification process as long as the initial values $P_{xi}(t_0)$ and $P_{yj}(t_0)$ are diagonal positive definite, and g_{xi} and g_{yj} satisfy the conditions given in (5.23)*. Also, it is easy to show that P_{xi}, P_{yj} will converge to their equilibrium points P_{xie} and P_{yje}, and $H_{xf}^T P_{xie} H_{xf}$, $H_{yf}^T P_{yje} H_{yf}$ will satisfy the following equations:*

$$\begin{aligned} H_{xf}^T P_{xie} H_{xf} &= \frac{2g_{xi}q(1-\lambda_{xi}\bar{\zeta}_{xf}^2)}{\lambda_{xi}-2g_{xi}q(1+\lambda_{xi})}, \\ H_{yf}^T P_{yje} H_{yf} &= \frac{2g_{yj}q(1-\lambda_{yj}\bar{\zeta}_{yf}^2)}{\lambda_{yj}-2g_{yj}q(1+\lambda_{yj})}. \end{aligned} \tag{5.24}$$

Proof. Let P_{xie} denotes the equilibrium of P_{xi}, it follows that:

$$\begin{aligned} &\dot{P}_{xie} = g_{xi}P_{xie} - \frac{\mu_{xi}P_{xie}\text{diag}(H_{xf})^2 P_{xie}}{(1-\mu_{xi})\bar{\zeta}_{xfi}^2 + \mu_{xi}H_{xf}^T P_{xie}H_{xf}} = 0 \\ \Leftrightarrow &g_{xi}P_{xie} = \frac{\mu_{xi}P_{xie}\text{diag}(H_{xf})^2 P_{xie}}{(1-\mu_{xi})\bar{\zeta}_{xfi}^2 + \mu_{xi}H_{xf}^T P_{xie}H_{xf}} \\ \Leftrightarrow &g_{xi}\mathbf{I} = \frac{\mu_{xi}P_{xie}\text{diag}(H_{xf})^2}{(1-\mu_{xi})\bar{\zeta}_{xfi}^2 + \mu_{xi}H_{xf}^T P_{xie}H_{xf}} \\ \Rightarrow &\text{tr}(g_{xi}\mathbf{I}) = \text{tr}\Big(\frac{\mu_{xi}P_{xie}\text{diag}(H_{xf})^2}{(1-\mu_{xi})\bar{\zeta}_{xfi}^2 + \mu_{xi}H_{xf}^T P_{xie}H_{xf}}\Big). \end{aligned} \tag{5.25}$$

Substituting (5.19) into (5.25), and using the fact that $\text{tr}(P_{xie}\text{diag}(H_{xf})^2) = H_{xf}^T P_{xie} H_{xf}$, it can be obtained that

$$H_{xf}^T P_{xie} H_{xf} = \frac{2g_{xi}q(1-\lambda_{xi}\bar{\zeta}_{xfi}^2)}{\lambda_{xi}-2g_{xi}q(1+\lambda_{xi})}. \tag{5.26}$$

Because $0 < \lambda_{xi}\bar{\zeta}_{xfi}^2 < 1$, in order to guarantee that $H_{xf}^T P_{xie} H_{xf}$ is positive, $\lambda_{xi} - 2g_{xi}q(1+\lambda_{xi}) > 0$ should be satisfied, i.e.,

$$0 < g_{xi} < \frac{\lambda_{xi}}{2q(1+\lambda_{xi})}. \tag{5.27}$$

Let p_{xr} denotes the r^{th} diagonal element of P_{xi}, and h_{xr} be r^{th} element of H_{xf}. Using (5.17)

and (5.19), it can be obtained that

$$\dot{p}_{xr} = g_{xi}p_{xr} - \frac{\lambda_{xi}p_{xr}^2 h_{xr}^2}{(1-\lambda_{xi}\bar{\zeta}_{xfi}^2) + (1+\lambda_{xi})H_{xf}^T P_{xi} H_{xf}}.$$

Using the fact that $p_{xr}(t_0) > 0$, $1 - \lambda_{xi}\bar{\zeta}_{xfi} > 0$, one has

$$\begin{aligned}
&\dot{p}_{xr} > 0 \\
\Leftrightarrow & g_{xi}p_{xr} - \frac{\lambda_{xi}p_{xr}^2 h_{xr}^2}{(1-\lambda_{xi}\bar{\zeta}_{xfi}^2) + (1+\lambda_{xi})H_{xf}^T P_{xi} H_{xf}} > 0 \\
\Leftrightarrow & g_{xi}(1-\lambda_{xi}\bar{\zeta}_{xfi}) + g_{xi}(1+\lambda_{xi})H_{xf}^T P_{xi} H_{xf} > \lambda_{xi}p_{xr}^2 h_{xr}^2 \\
\Leftrightarrow & \big(\lambda_{xi} - g_{xi}(1+\lambda_{xi})\big)p_{xr}h_{xr}^2 \\
& < g_{xi}(1-\lambda_{xi}\bar{\zeta}_{xfi}) + g_{xi}(1+\lambda_{xi}) \sum_{k=1,k\neq r}^{2q} p_{xk}h_{xk}^2.
\end{aligned} \tag{5.28}$$

If g_{xi} satisfies (5.23), then (5.28) can be reformulated as:

$$\begin{aligned}
p_{xr} < & \frac{g_{xi}(1-\lambda_{xi}\bar{\zeta}_{xfi}) + g_{xi}(1+\lambda_{xi})\sum_{k=1,k\neq r}^{2q} p_{xk}h_{xk}^2}{\big(\lambda_{xi} - g_{xi}(1+\lambda_{xi})\big)h_{xr}^2} \\
& \equiv p_{xre}.
\end{aligned} \tag{5.29}$$

Hence, if $p_{xr} < p_{xre}$, one has $\dot{p}_{xr} > 0$. If $p_{xr} > p_{xre}$, it can be obtained that $\dot{p}_{xr} < 0$. This implies that if only $P_{xi}(t_0) > 0$, and (5.23) is satisfied, then p_{xr} will always converge to its equilibrium p_{xre} which is a positive number. Hence, it can be concluded that P_{xi} will remain diagonal positive definite. Following the same procedure, it is easy to show that P_{yj} will also remain diagonal positive definite. The proof of Lemma 5.1 is thus completed. □

Theorem 5.1. *Consider the multi-time-scale RNN model* (5.8) *for nonlinear SPS* (5.1) *with the updating laws* (5.17)-(5.22)*, it can be guaranteed that the identification errors $\tilde{x}$, $\tilde{y}$ and NN weight errors $\tilde{\theta}_{xi}$, $\tilde{\theta}_{yj}$ can be minimized and remain bounded.*

Proof. Consider the following Lyapunov function:

$$V_{xi} = \tilde{\theta}_{xi} P_{xi}^{-1} \tilde{\theta}_{xi}^T. \tag{5.30}$$

The derivative of V_{xi} is given as:

$$\dot{V}_{xi} = 2\tilde{\theta}_{xi} P_{xi}^{-1} \dot{\tilde{\theta}}_{xi}^T - \tilde{\theta}_{xi} \frac{dP_{xi}^{-1}}{dt} \tilde{\theta}_{xi}^T. \tag{5.31}$$

It should be noticed that

$$\frac{d}{dt}\left[P_{xi} P_{xi}^{-1}\right] = \dot{P}_{xi} P_{xi}^{-1} + P_{xi} \frac{d}{dt} P_{xi}^{-1} = \frac{d\mathbf{I}}{dt} = 0. \tag{5.32}$$

Using (5.17), the following equation can be obtained:

$$\frac{d}{dt} P_{xi}^{-1} = -P_{xi}^{-1} \dot{P}_{xi} P_{xi}^{-1} = -g_{xi} P_{xi}^{-1} + \frac{\mu_{xi} \text{diag}(H_{yf})^2}{(1-\mu_{xi})\bar{\zeta}_{xfi}^2 + \mu_{xi} H_{yf}^T P_{xi} H_{yf}}. \tag{5.33}$$

Substituting (5.18), (5.33) into (5.31), it follows that:

$$\begin{aligned}
\dot{V}_{xi} =& -\frac{2\mu_{xi}}{\bar{\zeta}_{xfi}^2} \tilde{\theta}_{xi} H_{xf} e_{xfi} + \frac{\mu_{xi} \tilde{\theta}_{xi} \text{diag}(H_{xf})^2 \tilde{\theta}_{xi}^T}{(1-\mu_{xi})\bar{\zeta}_{xfi}^2 + \mu_{xi} H_{xf}^T P_{xi} H_{xf}} - \frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2} |\tau_{xfi} - \theta_{xi}^* H_{xf}|^2 \\
&+ \frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2} \tilde{\theta}_{xi} H_{xf} H_{xf}^T \tilde{\theta}_{xi}^T + \frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2} |\tau_{xfi} - \theta_{xi}^* H_{xf}|^2 - \frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2} \tilde{\theta}_{xi} H_{xf} H_{xf}^T \tilde{\theta}_{xi}^T - g_{xi} \tilde{\theta}_{xi} P_{xi}^{-1} \tilde{\theta}_{xi}^T \\
=& \left[-\frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2} |\tau_{xfi} - \theta_{xi}^* H_{xf}|^2 + \frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2} \tilde{\theta}_{xi} H_{xf} H_{xf}^T \tilde{\theta}_{xi}^T - \frac{2\mu_{xi}}{\bar{\zeta}_{xfi}^2} \tilde{\theta}_{xi} H_{xf} e_{xfi} \right] - g_{xi} \tilde{\theta}_{xi} P_{xi}^{-1} \tilde{\theta}_{xi}^T \\
&+ \frac{\mu_{xi} \tilde{\theta}_{xi} \text{diag}(H_{xf})^2 \tilde{\theta}_{xi}^T}{(1-\mu_{xi})\bar{\zeta}_{xfi}^2 + \mu_{xi} H_{xf}^T P_{xi} H_{xf}} - \frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2} \tilde{\theta}_{xi} H_{xf} H_{xf}^T \tilde{\theta}_{xi}^T + \frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2} \|\tau_{xfi} - \theta_{xi}^* H_{xf}\|_2^2.
\end{aligned} \tag{5.34}$$

Using the fact that $e_{xfi} = \tau_{xfi} - \theta_{xi} H_{xf}$ and $\tilde{\theta}_{xi} = \theta_{xi}^* - \theta_{xi}$, the equation in $[\cdot]$ in (5.34) can be reformulated as:

$$-\frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2} |\tau_{xfi} - \theta_{xi}^* H_{xf}|^2 + \frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2} \tilde{\theta}_{xi} H_{xf} H_{xf}^T \tilde{\theta}_{xi}^T - \frac{2\mu_{xi}}{\bar{\zeta}_{xfi}^2} \tilde{\theta}_{xi} H_{xf} e_{xfi}$$

$$
\begin{aligned}
=&\frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2}\left[-|\tau_{xfi}-\theta_{xi}^* H_{xf}|^2+(\theta_{xi}^*-\theta_{xi})H_{xf}H_{xf}^T(\theta_{xi}^*-\theta_{xi})^T-2(\theta_{xi}^*-\theta_{xi})H_{xf}(\tau_{xfi}-\theta_{xi}H_{xf})\right]\\
=&-\frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2}\left[\tau_{xfi}^2-2\theta_{xi}H_{xf}\tau_{xfi}+(\theta_{xi}H_{xf})^2\right]\\
=&-\frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2}e_{xfi}^2. \qquad (5.35)
\end{aligned}
$$

Substituting (5.35) into (5.34), and using (5.16), along with the fact that $0<\mu_{xi}<1$, $\tilde{\theta}_{xi}\text{diag}(H_{xf})^2\tilde{\theta}_{xi}^T\leq\tilde{\theta}_{xi}H_{xf}H_{xf}^T\tilde{\theta}_{xi}^T$, the following inequality is obtained:

$$
\begin{aligned}
\dot{V}_{xi}\leq&\frac{\mu_{xi}\tilde{\theta}_{xi}H_{xf}H_{xf}^T\tilde{\theta}_{xi}^T}{(1-\mu_{xi})\bar{\zeta}_{xfi}^2+\mu_{xi}H_{xf}^TP_{xi}H_{xf}}-\frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2}\tilde{\theta}_{xi}H_{xf}H_{xf}^T\tilde{\theta}_{xi}^T+\mu_{xi}-g_{xi}\tilde{\theta}_{xi}P_{xi}^{-1}\tilde{\theta}_{xi}^T-\frac{\mu_{xi}}{\bar{\zeta}_{xfi}^2}e_{xfi}^2\\
=&\frac{[\bar{\zeta}_{xfi}^2-H_{xf}^TP_{xi}H_{xf}]\mu_{xi}^2\tilde{\theta}_{xi}H_{xf}H_{xf}^T\tilde{\theta}_{xi}^T}{\bar{\zeta}_{xfi}^2[(1-\mu_{xi})\bar{\zeta}_{xfi}^2+\mu_{xi}H_{xf}^TP_{xi}H_{xf}]}+\mu_{xi}(1-\frac{e_{xfi}^2}{\bar{\zeta}_{xfi}^2})-g_{xi}\tilde{\theta}_{xi}P_{xi}^{-1}\tilde{\theta}_{xi}^T. \qquad (5.36)
\end{aligned}
$$

It can be guaranteed that $H_{xf}^TP_{xi}H_{xf}>\bar{\zeta}_{xfi}^2$ is always true if only the g_{xi} and activation functions ψ_i are properly selected. Since P_{xi}^{-1} is positive definite, $\dot{V}_{xi}<0$ is valid whenever $e_{xfi}^2>\bar{\zeta}_{xfi}^2$. Thus, the auxiliary output error e_{xfi} will be bounded by $\bar{\zeta}_{xfi}$, and θ_{xi} will converge to its nominal value θ_{xi}^*, which implies the filtered identification error $\tilde{x}$ is also bounded according to (5.13). Following the same procedure, it is easy to show that the auxiliary output error e_y and the identification error $\tilde{y}$ will be bounded, and θ_{yj} will also converge to θ_{yj}^*. Theorem 5.1 is thus proved. □

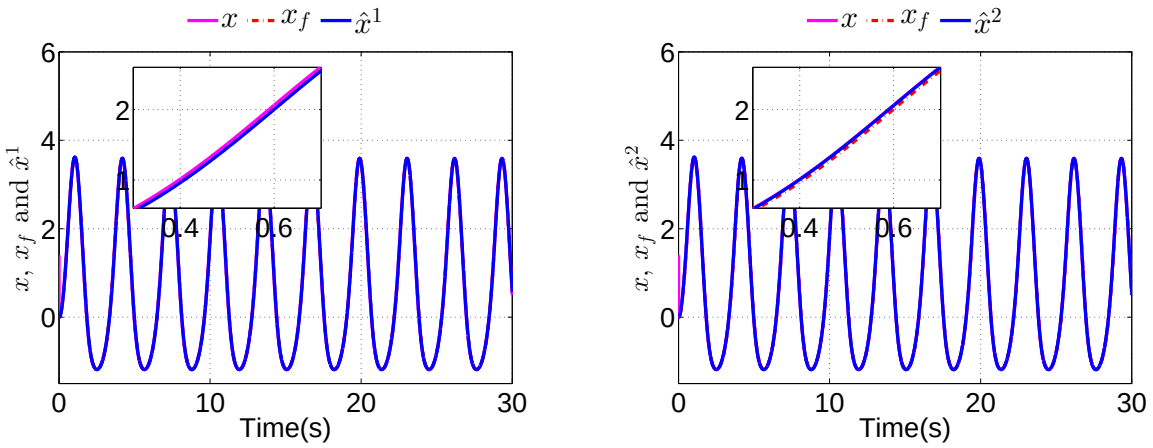
Remark 5.1. *In* (4.55) *of Section 4.3, and* (4.99) *of Section 4.4, the* $\dot{\varsigma}_x$, $\dot{\varsigma}_y$ *are involved in* e_x *and* e_y. *In these cases, one usually has to differentiate* ς_x *and* ς_y *in order to obtain* $\dot{\varsigma}_x$ *and* $\dot{\varsigma}_y$. *However, this may cause failure to the identification process when the signals (measurements of* x *and* y*) contain noises, which will be demonstrated in the simulation later. In this section, by defining the filtered variables as in* (5.5) *and* (5.6)*, the derivatives of* x *and* y *are not needed. As a result, the system identification process can achieve high precision even if the signals are very noisy.*

5.3 Simulation

In order to demonstrate the effectiveness of the robust system identification scheme proposed in Chapter 5, the following nonlinear SPS is considered in simulation:

$$\begin{aligned} \dot{x} &= -5x + xy + 3y, \\ 0.1\dot{y} &= -10y + 2\sin(x) + u, \end{aligned} \tag{5.37}$$

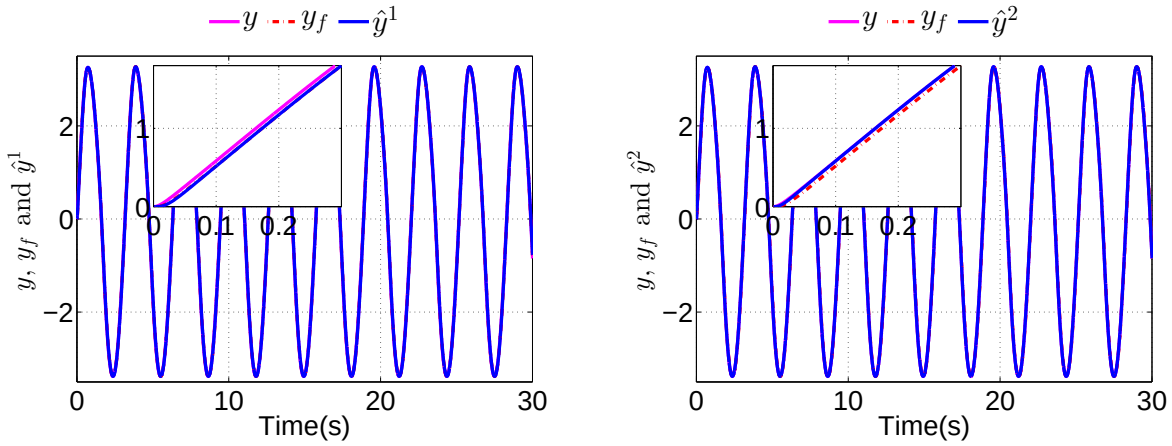
with $x(t_0) = 1.4$ and $y(t_0) = 0$. The input signal u is chosen as $u = 8\sin(2t)$. In order to identify this nonlinear SPS, the RNN defined in (5.8) with node number $q = 2$ is used. In this case, $W_1 \in \Re^2$, $W_2 \in \Re^2$, $W_3 \in \Re^2$, $W_4 \in \Re^2$. The activation functions in (5.3) are selected as $\psi_1(z) = 10/(1+e^{-z}) + 1$, $\psi_2(z) = 1/(1+e^{-z}) + 1$, $\psi_3(z) = 1/(1+e^{-z}) + 1$, $\psi_4(z) = 1/(1+e^{-z}) + 1$. The other RNN parameters are chosen as $A = B = -1$, $\lambda_x = 100$, $\lambda_y = 100$, $g_x = 0.33$, $g_y = 0.33$, $k_x = k_y = 0.01$, $L_x = 5$, $L_y = 5$, $\bar{\zeta}_{xf} = \bar{\zeta}_{yf} = 0.0001$, $P_x(t_0) = \text{diag}(100, 100, 100, 100)$, $P_y(t_0) = \text{diag}(1000, 1000, 1000, 1000)$. For comparison purpose, the identification algorithm proposed in Section 4.4 is also tested. Hereinafter, the superscript ^{1} denotes the results obtained using the robust identification algorithm proposed in Chapter 5, and ^{2} denotes the results using the algorithm proposed in Section 4.4. The sampling time for the simulation is 1 ms. When there is no measurement noises, the simulation results are presented in Figs. 5.1-5.3.



(a) Proposed in Chapter 5.

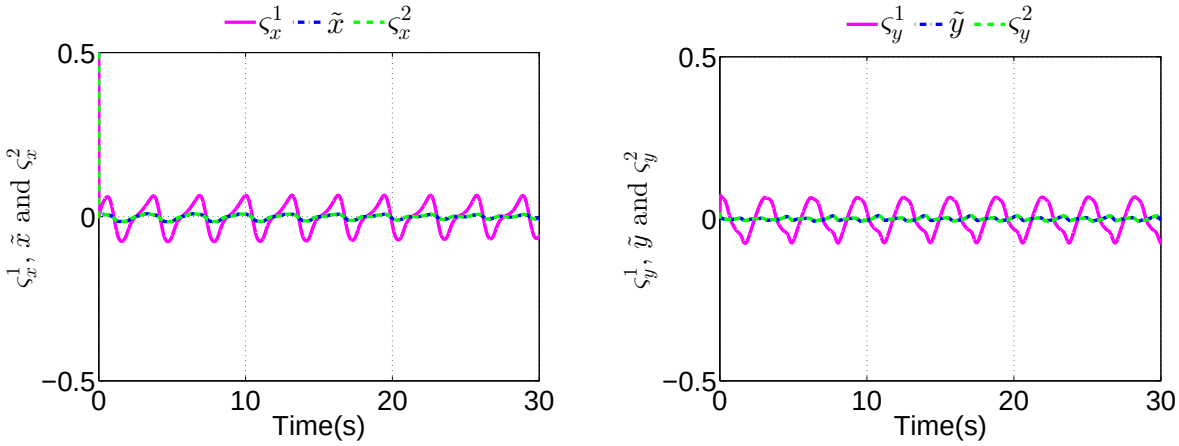
(b) Proposed in Section 4.3.

Figure 5.1: Identification results of x without noises.



(a) Proposed in Chapter 5. (b) Proposed in Section 4.3.

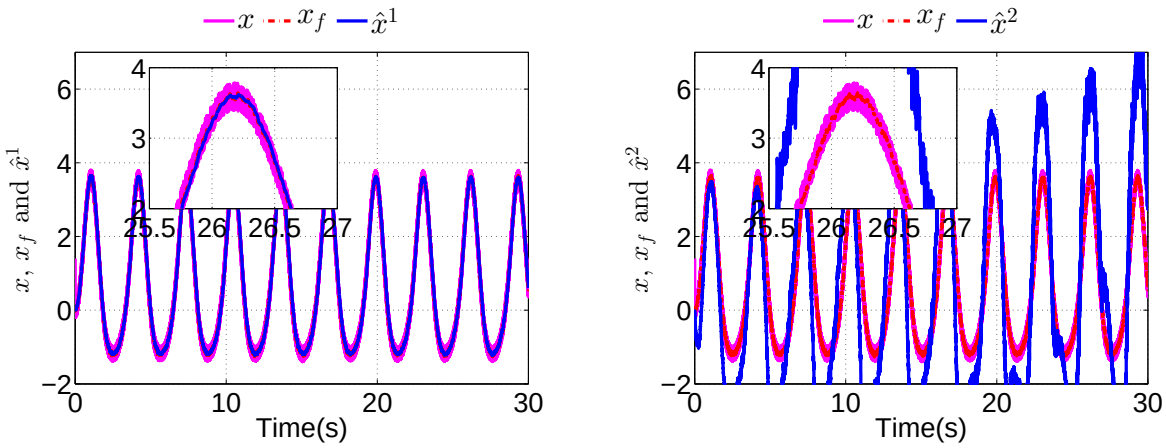
Figure 5.2: Identification results of y without noises.



(a) Identification errors of x. (b) Identification errors of y.

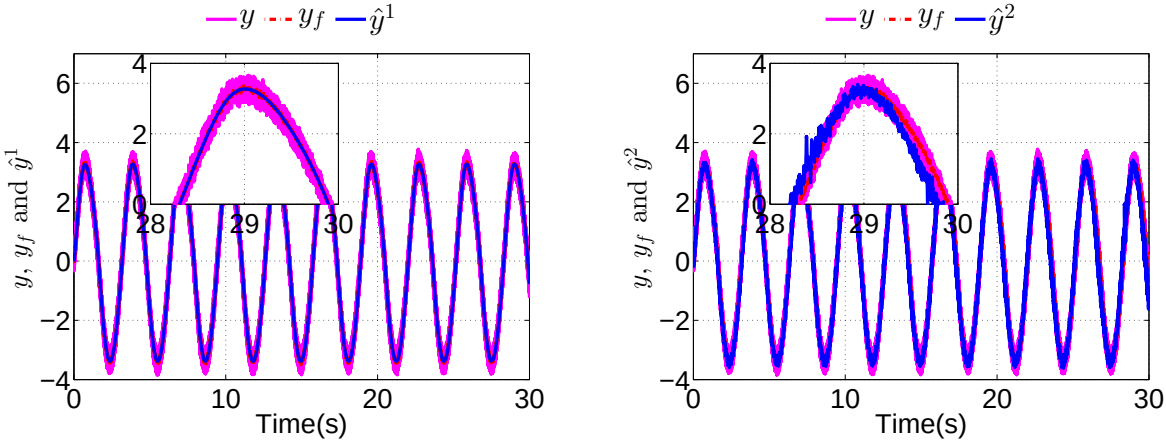
Figure 5.3: Identification errors without noises.

In Fig. 5.1 and Fig. 5.2, x and y represent the measured system outputs, x_f and y_f are the filtered outputs as defined in (5.5) and (5.6), $\hat{x}$ and $\hat{y}$ are the outputs of the RNN. It is clear that when there is no measurement noise, the RNN outputs $\hat{x}^1$, $\hat{y}^1$ of identification scheme proposed in Chapter 5 converge to x_f and y_f, respectively, and the RNN outputs $\hat{x}^2$, $\hat{y}^2$ of the identification scheme proposed in Section 4.4 will track x and y. Both identification methods can achieve satisfactory performance. Also, it can be noticed that there are small gaps between x, x_f and y, y_f. Because x_f and y_f are the filtered results of x and y, hence the "phase lag" will always exist. This can be reduced by choosing smaller k_x and k_y.



(a) Proposed in Chapter 5. (b) Proposed in Section 4.3.

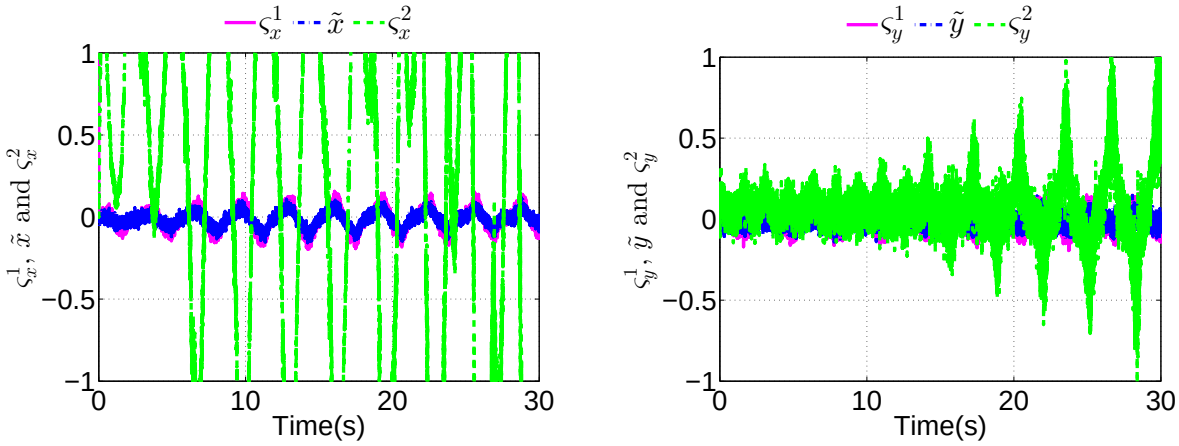
Figure 5.4: Identification results of x with noises.



(a) Proposed in Chapter 5. (b) Proposed in Section 4.3.

Figure 5.5: Identification results of y with noises.

In Fig. 5.3, ς_x and ς_y are the identification errors defined as $\varsigma_x = x - \hat{x}$, $\varsigma_y = y - \hat{y}$. The $\tilde{x}$ and $\tilde{y}$ are the filtered identification errors as defined in (5.9). Fig. 5.3 shows that the filtered identification errors $\tilde{x}$, $\tilde{y}$ are almost the same as ς_x^2 and ς_y^2, respectively, while ς_x^1 and ς_y^1 are larger. However, it should be emphasized that the overall goal of the identification scheme is to train the NN weights such that Θ_x, Θ_y will converge to there nominal value Θ_x^* and Θ_y^*. Thus the identification accuracy using the robust identification scheme proposed in Chapter 5 is determined by $\tilde{x}$ and $\tilde{y}$ instead of δ_x^1 and δ_y^1. As long as $\tilde{x}$ and $\tilde{y}$ are minimized, it implies that Θ_x and Θ_y have converged to Θ_x^* and Θ_y^*, respectively, and the goal of the identification algorithm has been achieved.



(a) Identification errors of x.

(b) Identification errors of y.

Figure 5.6: Identification errors with noises.

The identification results with measurement noises are presented in Figs. 5.4-5.6. In Fig. 5.4 and Fig. 5.5, the x and y are the measured states with noises. The random noise d_x with magnitude of 0.2 and mean of 0 is added to x and the random noise d_y with magnitude of 0.4 and and mean of 0 is added to y. In Fig. 5.4(a) and Fig. 5.5(a), it is shown that when measurement noises are involved, the outputs of the RNN ($\hat{x}^1$ and $\hat{y}^1$) always track x and y closely. However, when the identification algorithm proposed in Section 4.3 is used, large difference between $\hat{x}^2$ and x can be observed, as depicted in Fig. 5.4(b), and the $\hat{x}^2$ is diverging. In Fig. 5.6, it is clear that both ς_x^1 and ς_y^1 are small, which means the small identification errors can be achieved when the robust identification algorithm proposed in Chapter 5 is used. However, when the identification algorithm proposed in Section 4.3 is used, both ς_x^2 and ς_y^2 are very large, and the magnitude of ς_y^2 is increasing. The superiority of the identification scheme proposed in Chapter 5 is thus demonstrated.

The performance index-Root Mean Square (RMS) can also be used to illustrate the effectiveness of the identification algorithm proposed in this Chapter. The RMS of an array $z \in \Re^n$ is calculated as:

$$RMS = \sqrt{(\sum_{i=1}^{n} z^2(i))/n}. \tag{5.38}$$

The RMS values of ς_x^1, ς_y^1, ς_x^2, ς_y^2, $\tilde{x}$, $\tilde{y}$ are given in Table 5.1. From Table 5.1, it is very clear

that when there is no noise, $\tilde{x}$, $\tilde{y}$ are almost the same as ς_x^2 and ς_y^2, while ς_x^1 and ς_y^1 are larger due to the "phase lag". When measurement noises are involved, the ς_x^2 and ς_y^2 increase dramatically, while $\tilde{x}$, $\tilde{y}$, ς_x^1, ς_y^1 still remain at relative low levels.

Table 5.1: RMS values

	ς_x^1	ς_y^1	$\tilde{x}$	$\tilde{y}$	ς_x^2	ς_y^2
Without Noise	0.0533	0.0504	0.0198	0.0063	0.0200	0.0063
With Noise	0.1265	0.2486	0.0349	0.0365	3.6440	0.8917

5.4 Conclusion

In this chapter, a robust OBE based identification scheme is proposed. By using some filtered variables in the identification scheme, the new algorithm no longer need the derivatives of the system states, and thus is more robust to measurement noises.

Chapter 6

Controller Design Based on Singular Perturbation Theory

6.1 Introduction

In Chapter 4 and Chapter 5, the system identification using OBE and multi-time-scale neural network is studied. In this chapter, the adaptive control based on singular perturbation theory and the identified system model will be investigated.

The *direct* adaptive control using singular perturbation theory and single time scale NN was discussed in [23–25]. Nevertheless, the *indirect* adaptive control using singular perturbation theory and multi-time-scale neural networks is rarely studied. In Chapter 3 and [26,27], the indirect adaptive controllers were designed based on the multi-time-scale NN identification results. However, the authors did not take advantage of the identified model to design two controllers for the slow and fast subsystems respectively using singular perturbation theory. Instead, the authors treated the system as a regular system (one-time-scale system), and designed a controller for the whole system. Thus, the order of the matrices in the controller could be very high if the slow and fast system states have high dimensions, and the matrices in the controller could be ill-conditioned because the singular perturbation parameter ε is involved, and ε is usually very small.

In this chapter, two indirect adaptive controllers will be designed based on the identified system model and singular perturbation theory. Firstly, an indirect adaptive controller for a regulation problem using the practically asymptotically stability (PAS) theory will be designed. Then, a second indirect adaptive controller will be proposed to solve the trajectory tracking problem. By using singular perturbation theory, the identified system model is decomposed into the reduced slow subsystem and the reduced fast subsystem. The controller is then designed for the reduced subsystems. Hence, the order of the system is reduced, which makes it easier to design a controller for the reduced subsystems. Meanwhile, the order of the matrices in the controller is also reduced, thus the required computational resource is reduced.

6.2 Controller Design for Regulation Problem

In this section, an indirect adaptive controller based on the PAS is proposed for a regulation problem.

6.2.1 Practically Asymptotically Stability

For a singularly perturbed system

$$\begin{aligned} \dot{x} &= f_1(x,\varepsilon) + f_2(x,\varepsilon)y, \\ \varepsilon\dot{y} &= g_1(x,y,\varepsilon) + g_2(x,y,\varepsilon)u, \end{aligned} \tag{6.1}$$

with $f_i(0,0) = 0,\ i = 1,2$ and $g_i(0,0,0) = 0,\ i = 1,2$, where $x \in \Re^n$ and $y \in \Re^m$ are the slow and fast states, respectively, $u \in \Re^p$ is the control signal vector, $0 < \varepsilon < 1$ is a small parameter, $f_i \in C^\infty, i = 1,2$ and $g_i \in C^\infty, i = 1,2$ are unknown general nonlinear smooth functions.

For simplicity, denote $f_s(t,x,y,\varepsilon) = f_1(x,\varepsilon) + f_2(x,\varepsilon)y$, and $g_s(t,x,y,\varepsilon) = g_1(x,y,\varepsilon) +$

$g_2(x, y, \varepsilon)$, then (6.1) can be represented as:

$$\begin{aligned} \dot{x} &= f_s(t, x, y, \varepsilon), \quad x(t_0) = \iota_x(\varepsilon), \\ \varepsilon\dot{y} &= g_s(t, x, y, \varepsilon), \quad y(t_0) = \iota_y(\varepsilon). \end{aligned} \tag{6.2}$$

Define $\tau = (t - t_0)/\varepsilon$, and system (6.2) can be written as:

$$\begin{aligned} \frac{dx}{d\tau} &= \varepsilon f_s(t_0 + \varepsilon\tau, x, y, \varepsilon), \quad x(0) = \iota_x(\varepsilon), \\ \frac{dy}{d\tau} &= g_s(t_0 + \varepsilon\tau, x, y, \varepsilon), \quad y(0) = \iota_y(\varepsilon)y. \end{aligned} \tag{6.3}$$

By setting $\varepsilon = 0$ in (6.3), it is shown that this system is a regular system

$$\begin{aligned} \frac{dx}{d\tau} &= 0, \qquad\qquad x(0) = \iota_x(0), \\ \frac{dy}{d\tau} &= g_s(t_0, x, y, 0), \quad y(0) = \iota_y(0). \end{aligned} \tag{6.4}$$

Hence the state y of any solution of system (6.2) varies rapidly according to the equation

$$\frac{dy}{d\tau} = g_s(t_0, \iota_x(0), y, 0),\ y(0) = \iota_y(0). \tag{6.5}$$

Equation (6.5) is called *boundary layer* equation. It consists in equation

$$\frac{dy}{d\tau} = g_s(t, x, y, 0),\ y(0) = \iota_y(0), \tag{6.6}$$

where $t = t_0$ and $x = \iota_x(0)$ are fixed at their initial values. Assume that the solutions of (6.6) tend toward an equilibrium $h_s(t, x)$ where $y = h_s(t, x)$ is a root of equation

$$g_s(t, x, y, 0) = 0. \tag{6.7}$$

The manifold $\mathcal{L}$ defined by equation (6.7) is called *slow manifold*. The solutions of (6.2) have a rapid transition from $(\iota_x(0), \iota_y(0))$ to a point of the slow manifold $\mathcal{L}$ $(\iota_x(0), h_s(t_0, \iota_x(0)))$. Then

a slow motion starts on the slow manifold according to the equation

$$\dot{x} = f_s(t, x, h_s(t, x), 0). \tag{6.8}$$

Equation (6.8) is called *the reduced problem*.

The following theory is valid [14]:

Theorem 6.1. *Consider the singularly perturbed system* (6.2)*, Assume that* $f_s(t, 0, 0, 0) = 0$ *and* $g_s(t, 0, 0, 0) = 0$*. Let* $y = h_s(t, x)$ *be an isolated root of* (6.7) *such that* $h_s(t, 0) = 0$*. Assume that the equilibrium* $y = h_s(t, x)$ *of the boundary layer equation* (6.5) *is asymptotically uniformly stable in* (t, x)*, and the origin of the corresponding reduced model* (6.8) *is asymptotically stable. Then there exists a positive constant* ε^* *such that for all* $0 < \varepsilon < \varepsilon^*$*, the origin of system* (6.2) *is practically asymptotically stable (PAS).*

6.2.2 Controller Design

Consider the nonlinear singularly perturbed system given in (6.1). In order to identify the nonlinear SPS (6.1), the following multi-time-scale NN will be used:

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + W_1\Psi_1(x) + W_2\Psi_2(x)y, \\ \varepsilon\dot{\hat{y}} &= B\hat{y} + W_3\Psi_3(x, y) + W_4\Psi_4(x, y)u. \end{aligned} \tag{6.9}$$

where $\hat{x} \in \Re^n$ and $\hat{y} \in \Re^m$ are the estimation of the slow and fast states x and y, respectively, $A \in \Re^{n\times n}$, $B \in \Re^{m\times m}$ are diagonal stable matrices, $W_1 \in \Re^{n\times q}$, $W_2 \in \Re^{n\times q}$, $W_3 \in \Re^{m\times q}$, $W_4 \in \Re^{m\times q}$ are the weight matrices of the RHONN, q is the number of neuron. The activation function vectors $\Psi_i(\cdot), i = 1, 3$ are defined as:

$$\begin{aligned} \Psi_i(\cdot) &= [\Psi_{i,1}, \Psi_{i,2}, \ldots, \Psi_{i,q}]^T \in \Re^{q\times 1}, i = 1, 3, \\ \Psi_{i,c} &= \prod_{j\in J_c} [\psi_i(\cdot)]^{d_c(j)}, c = 1, \ldots, q, \end{aligned}$$

where J_c are the collections of q not ordered subsets of $1, 2, \ldots, n+m$ and $d_c(j)$ are non-negative integers. The activation function matrix $\Psi_2(\cdot)$ is defined as:

$$\Psi_2(\cdot) = \begin{bmatrix} \Psi_{2,1,1} & \ldots & \Psi_{2,1,m} \\ \vdots & \ddots & \vdots \\ \Psi_{2,q,1} & \ldots & \Psi_{2,q,m} \end{bmatrix} \in \Re^{q \times m},$$

$$\Psi_{2,c,r} = \prod_{j \in J_{c,r}} [\psi_2(\cdot)]^{d_{c,r}(j)}, c = 1, \ldots, q, r = 1, \ldots, m,$$

where $J_{c,r}$ are the collections of $q \times m$ not ordered subsets of $1, 2, \ldots, n+m$ and $d_{c,r}(j)$ are non-negative integers. The activation function matrix $\Psi_4(\cdot)$ is defined as:

$$\Psi_4(\cdot) = \begin{bmatrix} \Psi_{4,1,1} & \ldots & \Psi_{4,1,p} \\ \vdots & \ddots & \vdots \\ \Psi_{4,q,1} & \ldots & \Psi_{4,q,p} \end{bmatrix} \in \Re^{q \times p},$$

$$\Psi_{4,k,r} = \prod_{j \in J_{c,r}} [\psi_4(\cdot)]^{d_{c,r}(j)}, c = 1, \ldots, q, r = 1, \ldots, p,$$

where $J_{c,r}$ are the collections of $q \times p$ not ordered subsets of $1, 2, \ldots, n+m$ and $d_{c,r}(j)$ are non-negative integers. The activation function $\psi_i(z)$ is chosen as:

$$\psi_i(z) = \frac{\alpha_{i,1}}{1 + e^{-\alpha_{i,2} z}} + \alpha_{i,3}, \ i = 1, \cdots, 4. \tag{6.10}$$

Using the multi-time-scale NN (6.9) and the weight's updating laws proposed in Section 4.3, the identified system model can be represented as:

$$\dot{x} = Ax + W_1 \Psi_1(x) + W_2 \Psi_2(x) y + \delta'_x, \tag{6.11}$$

$$\varepsilon \dot{y} = By + W_3 \Psi_3(x, y) + W_4 \Psi_4(x, y) u + \delta'_y, \tag{6.12}$$

where $\delta'_x, \ \delta'_y$ are modeling errors.

Let $h(x)$ be the equilibrium of (6.12), and define a new system state $E_y = y - h(x)$. Therefore, the identified model given in (6.11) and (6.12) can be rewritten as:

$$\dot{x} = Ax + W_1\Psi_1(x) + W_2\Psi_2(x)(E_y + h(x)) + \delta'_x, \tag{6.13a}$$

$$\varepsilon\dot{E}_y = BE_y + W_3\Psi_3(x,y) + W_4\Psi_4(x,y)u + Bh(x) + \delta'_y - \varepsilon\frac{dh(x)}{dx}\dot{x}. \tag{6.13b}$$

For a singularly perturbed system, because ε is usually very small, the changing rate of E_y is very high, which means y will converge to its equilibrium $h(x)$ rapidly. In fact, when ε is set to be 0, the transient response of y is instantaneous whenever the right hand side of (6.13b) is not equal to 0. By setting $\varepsilon = 0$ in (6.13), and having $E_y = 0$ when $\varepsilon = 0$, the *Reduced Slow Subsystem* can be obtained as:

$$\dot{x} = Ax + W_1\Psi_1(x) + W_2\Psi_2(x)h(x) + \delta'_x, \tag{6.14}$$

$$0 = BE_y + W_3\Psi_3(x,y) + W_4\Psi_4(x,y)u + Bh(x) + \delta'_y. \tag{6.15}$$

Define a new "stretched" time variable as $\tau = (t - t_0)/\varepsilon$, where t_0 is the initial time, and set $\varepsilon = 0$ in (6.13). The *Reduced Fast Subsystem* can be written as:

$$\frac{dx}{d\tau} = 0, \tag{6.16}$$

$$\frac{dE_y}{d\tau} = BE_y + W_3\Psi_3(x,y) + W_4\Psi_4(x,y)u + Bh(x) + \delta'_y. \tag{6.17}$$

Remark 6.1. *It is challenging to design the control signal u if the order of system* (6.13) *is very high. However, by using Tiknonov's theorem [14], the singularly perturbed system* (6.13) *can be decomposed into two subsystems: the* Reduced Slow Subsystem *and the* Reduced Fast Subsystem. *According to Theorem 6.1, if the equilibrium of* (6.17) *is asymptotically uniformly stable, and* (6.14) *is asymptotically stable, then there exists a positive constant ε^* such that for all $0 < \varepsilon < \varepsilon^*$, the origin of system* (6.13) *is PAS. Hence, the control problem of the singularly perturbed system* (6.13) *is changed to that of two lower order subsystems* (6.14)-(6.17). *The complexity of*

the controller design problem is thus reduced.

To ensure the asymptotic stability of (6.14), the control signal u should be designed such that the equilibrium $h(x)$ of (6.12) is:

$$h(x) = W_2\Psi_2(x)^\dagger h'(x), \tag{6.18}$$

$$h'(x) = -Ax - W_1\Psi_1(x) - k_{x1}x - k_{x2}[\text{sign}(x_1), \cdots, \text{sign}(x_n)]^T, \tag{6.19}$$

where $W_2\Psi_2(x)^\dagger$ is the Moore-Penrose pseudo-inverse of $W_2\Psi_2(x)$, $x_i,\ i = 1, \cdots, n$ is the i^{th} element of x, $k_{x1},\ k_{x2}$ are the designed parameters satisfying $k_{x1} > 0,\ k_{x2} > \bar{\delta}_x$, where $\bar{\delta}_x$ is the upper bound of $\|\delta_x\|_1$, and $\|\cdot\|_1$ denotes the L_1-norm.

Remark 6.2. *$h(x) = W_2\Psi_2(x)^\dagger h'(x)$ is the unique best approximate solution of equation [119]*

$$W_2\Psi_2(x)h(x) = h'(x), \tag{6.20}$$

i.e., if the approximate solution error is defined as

$$\upsilon_x = h'(x) - W_2\Psi_2(x)h(x), \tag{6.21}$$

then by using $h(x)$ given in (6.18)*, $\|\upsilon_x\|_2$ is minimized, where $\|\cdot\|_2$ denotes the L_2-norm. It should be noticed that when $W_2\Psi_2(x)$ is a square nonsingular matrix, the following equations hold:*

$$W_2\Psi_2(x)^\dagger = W_2\Psi_2(x)^{-1} \tag{6.22}$$

$$h(x) = W_2\Psi_2(x)^{-1}h'(x) \tag{6.23}$$

$$\upsilon_x = 0. \tag{6.24}$$

Equation (6.21) can be rewritten as:

$$W_2\Psi_2(x)h(x) = h'(x) - \upsilon_x, \tag{6.25}$$

Substituting (6.25) and (6.19) into (6.14), it gives that

$$\dot{x} = -k_{x1}x - k_{x2}[\text{sign}(x_1), \cdots, \text{sign}(x_n)]^T + \delta_x, \tag{6.26}$$

where $\delta_x = \delta'_x - \upsilon_x$. It can be proved that the dynamic system (6.26) is asymptotically stable. The detailed proof will be given later.

In order to ensure the uniform asymptotic stability of (6.17), i.e., to ensure y will converge to its equilibrium $h(x)$ asymptotically and uniformly in x, the control signal u can be designed as:

$$u = W_4\Psi_4(x,y)^{\dagger}u', \tag{6.27}$$

$$u' = -BE_y - W_3\Psi_3(x,y) - Bh(x) - k_{y1}E_y - k_{y2}[\text{sign}(E_{y1}), \cdots, \text{sign}(E_{ym})]^T, \tag{6.28}$$

where $W_4\Psi_4(x,y)^{\dagger}$ is the Moore-Penrose pseudo-inverse of $W_4\Psi_4(x,y)$. $E_{yj},\ j = 1, \cdots, m$ is the j^{th} element of E_y. $k_{y1},\ k_{y2}$ are the design parameters satisfying $k_{y1} > 0,\ k_{y2} > \bar{\delta}_y$, where $\bar{\delta}_y$ is the upper bound of $\|\delta_y\|_1$.

Similarly, the approximate solution error υ_y is defined as:

$$\upsilon_y = u' - W_4\Psi_4(x,y)^{\dagger}u. \tag{6.29}$$

Equation (6.29) can be rewritten as:

$$W_4\Psi_4(x,y)^{\dagger}u = u' - \upsilon_y. \tag{6.30}$$

Substituting (6.30) and (6.28) into (6.17) gives that

$$\frac{dE_y}{d\tau} = -k_{y1}E_y - k_{y2}[\text{sign}(E_{y1}), \cdots, \text{sign}(E_{ym})]^T + \delta_y, \tag{6.31}$$

where $\delta_y = \delta'_y - \upsilon_y$. It can also be proved that the error dynamic system (6.31) will be asymptotically uniformly stable, which implies y will converge to $h(x)$ asymptotically and uniformly in

x.

Theorem 6.2. *Consider the nonlinear system* (6.1)*, with the identified system model* (6.12)*, control law* (6.27) *with* $h(x)$ *defined in* (6.18)*, there exists* $\varepsilon^* > 0$ *such that for all* $0 < \varepsilon < \varepsilon^*$*, the origin of system* (6.13) *is PAS, and* x *will converge to* 0*.*

Proof. Select Lyapunov function V_x as

$$V_x = \frac{1}{2} x^T x. \tag{6.32}$$

Using (6.26), the derivative of V_x can be obtained as:

$$\begin{aligned}
\dot{V}_x =& x^T \dot{x} \\
=& -k_{x1} x^T x - k_{x2} x^T [\text{sign}(x_1), \cdots, \text{sign}(x_n)]^T + x^T \delta_x \\
\leq& -k_{x1} x^T x - k_{x2} \|x\|_1 + \|x\|_1 \|\delta_x\|_1 \\
=& -k_{x1} x^T x - (k_{x2} - \|\delta_x\|_1) \|x\|_1.
\end{aligned} \tag{6.33}$$

According to (6.33), since $k_{x1} > 0,\ k_{x2} > \bar{\delta}_x \geq \|\delta_x\|_1$, it is obvious that $\dot{V}_x < 0$ is valid. Hence subsystem (6.14) is asymptotically stable.

Similarly, select Lyapunov function V_{Ey} as

$$V_{Ey} = \frac{1}{2} E_y^T E_y. \tag{6.34}$$

Using (6.31), the derivative of V_{Ey} with respective to τ is given as

$$\begin{aligned}
\frac{dV_{E_y}}{d\tau} =& E_y^T \frac{dE_y}{d\tau} \\
=& -k_{y1} E_y^T E_y - k_{y2} E_y^T [\text{sign}(E_{y1}), \cdots, \text{sign}(E_{ym})]^T + E_y^T \delta_y \\
\leq& -k_{y1} E_y^T E_y - k_{y2} \|E_y\|_1 + \|E_y\|_1 \|\delta_y\|_1 \\
=& -k_{y1} E_y^T E_y - (k_{y2} - \|\delta_y\|_1) \|E_y\|_1.
\end{aligned} \tag{6.35}$$

Since $k_{y1} > 0,\ k_{y2} > \bar{\delta}_y \geq \|\delta_y\|_1$, it can be known from (6.35) that $\frac{dV_{E_y}}{d\tau} < 0$ is also valid. Hence the boundary layer equation (6.17) is asymptotically uniformly stable, and E_y will converge to 0 asymptotically uniformly in x, which means that the states y will converge to the equilibrium $h(x)$ asymptotically uniformly in x. Since boundary layer equation (6.17) is asymptotically uniformly stable in x, and the origin of the corresponding reduced model (6.14) is asymptotically stable. According to Theorem 6.1, there exists $\varepsilon^* > 0$ such that for all $0 < \varepsilon < \varepsilon^*$, (6.13) is PAS. The proof of Theorem 6.2 is thus completed. □

A summary of the identification and control of nonlinear singularly perturbed system using two-time-scale neural networks is listed as follows:

1. Construct the two-time-scale RHONN as in (6.9). Select suitable sigmoid functions $\psi_i(z), i = 1, \cdots, 4$ and initial values of the weight matrices $W_i, i = 1, \cdots, 4$. The most commonly used sigmoid functions are logistic function and hyperbolic tangent function. The initial values of the weight matrices should be carefully selected such that the ill-conditioned problem will not occur.

2. Use singular perturbation theorem to decompose the error dynamic equations (6.13) into the reduced slow subsystem (6.14), (6.15) and the reduced fast subsystem (6.16) and (6.17).

3. Design the equilibrium $h(x)$ in (6.18) such that the reduced slow subsystem is asymptotically stable.

4. Design the control signal u in (6.27) such that the reduced fast subsystem will converge to its equilibrium $h(x)$ asymptotically and uniformly in x. Select k_{x1}, k_{y1} to be small values at the beginning and increase them gradually to achieve satisfactory control results. Generally speaking, larger k_{x1}, k_{y1} will result in faster convergence of x, y with more oscillation. k_{x2}, k_{y2} should be small enough to minimize the ripples when x, y are close to their equilibrium.

Remark 6.3. *In this section, the* indirect *adaptive controller is designed based on the identified model using* multi-time-scale *RHONN. In [23–25], NN was also adopted for adaptive control purpose. The difference is that in these papers, only* regular (one-time-scale) *NN was used to design* direct *adaptive controllers. Multi-time-scale NN was also used in Chapter 3 and [26, 27] to design*

indirect adaptive controllers. Nevertheless, in these papers, the controllers were designed by treating the SPSs as regular systems, and singular perturbation theorem was not considered. Hence, the order of the matrices in the controllers designed in [26, 27] could be very large if the dimensions of $x,\ y$ *were high. Also, ill-conditioned problem was more likely to occur because* $1/\varepsilon$ *was involved in the controller matrices. In this section, by using the singular perturbation technique, the original system is decomposed into two lower-order subsystems, and the indirect adaptive controller is designed for the subsystems. As a result, the matrices of the controller designed in this section has lower order, and the required computational resource is reduced. Besides,* ε *is not involved in the controller matrices in this paper. Hence, the ill-conditioned problem is less likely to occur.*

Remark 6.4. *In general, there is no theoretical proof that the ill-conditioned problem will not occur during the identification and control process due to the matrix inversion in controller design. Some methods such as projection operation can be included in the adaptive laws for the NN weights to avoid the potential ill-conditioned problem [48, 96–98]. However, in practice, by choosing proper initial weight matrices and selecting suitable learning and control parameters, the ill-conditioned problem can be avoided.*

6.2.3 Simulation

To demonstrate the potential application of the suggested identification and control schemes for practical systems, an armature-controlled DC motor is considered:

$$\begin{aligned} J\frac{d\omega}{dt} &= k_t i, \\ L\frac{di}{dt} &= -k_b\omega - Ri + V, \end{aligned} \tag{6.36}$$

where $R,\ L,\ i,\ V$ are the armature resistance, inductance, current, and voltage, J is the moment of inertia, ω is the angular speed, $k_t,\ k_b$ are the torque and back electromotive force constant. Define $x = \omega,\ y = Ri/k_b,\ u = v/k_b,\ T_m = JR/(k_t k_b),\ T_e = L/R,\ t_r = t/T_m$, the state equations

(6.36) can be brought into a standard form of a singularly perturbed system [120]:

$$\begin{aligned} \frac{dx}{dt_r} &= a_1 y, \\ \varepsilon \frac{dy}{dt_r} &= -b_1 x - b_2 y + u, \end{aligned} \tag{6.37}$$

where $\varepsilon = T_e/T_m,\ a_1 = b_1 = b_2 = 1$. Since the electrical time constant T_e is much smaller than the mechanical time constant T_m, the ε is very small. In this simulation, ε is assumed to be 0.05.

To identify and control the singularly perturbed system (6.37), the multi-time-scale RHONN defined in (6.9) is used and the following parameters for the RHONN are chosen: $A = -10$, $B = -2$, $\lambda_x = 200$, $\lambda_y = 50$, the neural nodes number $q = 4$, and $W_i \in \Re^{1\times 4}, i = 1, \ldots, 4$. The activation functions in (4.46) are chosen as $\psi_1(z) = \psi_2(z) = \psi_3(z) = \psi_4(z) = 1/(1 + e^{-z}) + 1$. The initial values are set to be $x(t_0) = 1.4$, $y(t_0) = 0$, $P_x(t_0) = 10^3 I \in \Re^{8\times 8}$, $P_y(t_0) = I \in \Re^{8\times 8}$, $W_1(t_0) = W_2(t_0) = [1, 1, 1, 1]$, $W_3(t_0) = W_4(t_0) = [10^{-2}, 10^{-2}, 10^{-2}, 10^{-2}]$. For the controller, the following parameters are used: $k_{x1} = 10$, $k_{x2} = 10^{-4}$, $k_{y1} = 10$, $k_{y2} = 10^{-4}$. The sampling time for the simulation is 0.001s and the numerical method is Euler method. The simulation results of the control scheme proposed in Section 6.2 and the state feedback controller (SFC) developed in [121] are presented in Figs. 6.1-6.6.

In Fig. 6.1, x represents the state controlled by the controller developed in this section. x_{sfc} denotes the state controlled by the SFC. $\hat{x}$ is the identification result of x during the identification process. In Fig. 6.2, y_r denotes the equilibrium point $h(x)$. y and $\hat{y}$ are the system state and identification result, respectively. State errors er_x and er_y in Fig. 6.3 are defined as $er_x = x - 0$ and $er_y = y - y_r$. Fig. 6.1 shows that by using the control scheme developed in this paper, x will converge to the origin faster compared to the result by using the SFC. Fig. 6.2 indicates y will converge to its equilibrium point. This is also demonstrated in Fig. 6.3 as the state errors er_x and er_y tend towards 0. Fig. 6.4 shows that during the control process, the identification errors ς_x and ς_y converge to 0, i.e., the identification results $\hat{x}$ and $\hat{y}$ are very close to their true values x and y.

In order to illustrate the robustness of the adaptive identification and control scheme, system

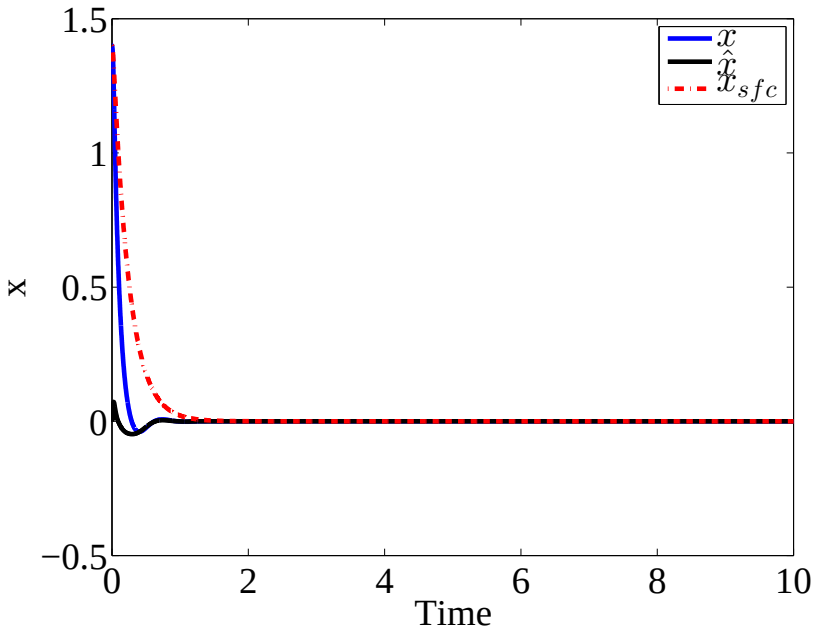


Figure 6.1: Identification and control results of x.

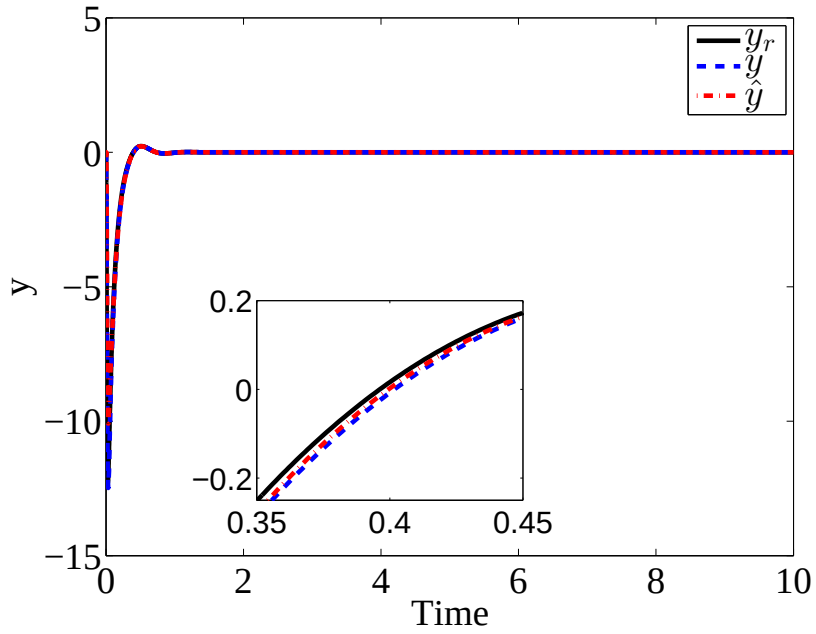


Figure 6.2: Identification and control results of y.

parameter b_2 is changed to 10 while the controller parameters remain the same. The control results are depicted in Fig. 6.5, where it is clear that when the system parameter changes, the result of the control scheme developed in this section remain almost the same. However, the performance of the SFC is deteriorated. It takes a longer time for the system state to converge to the origin when

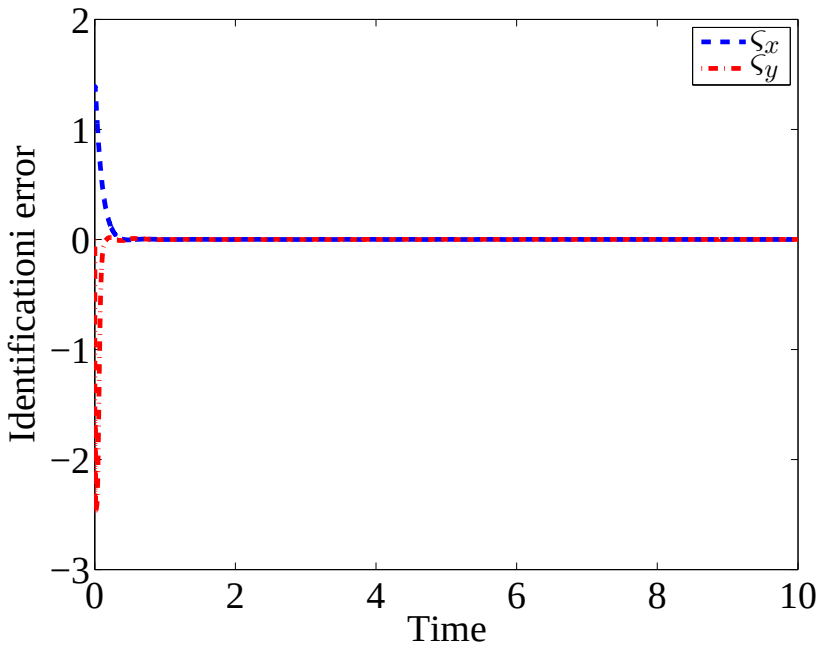


Figure 6.3: State errors of x and y.

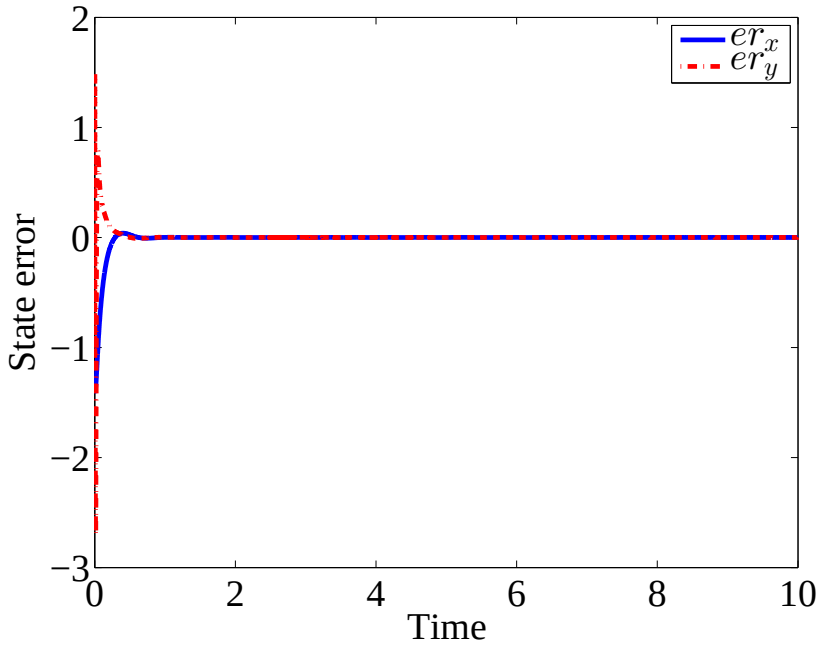


Figure 6.4: Identification errors of x and y.

using the SFC. If the system parameters are further changed to $a_1 = 0.5,\ b_2 = 10$, Fig. 6.6 shows that the performance of the controller developed in Section 6.2 remain almost the same, while the SFC is worse.

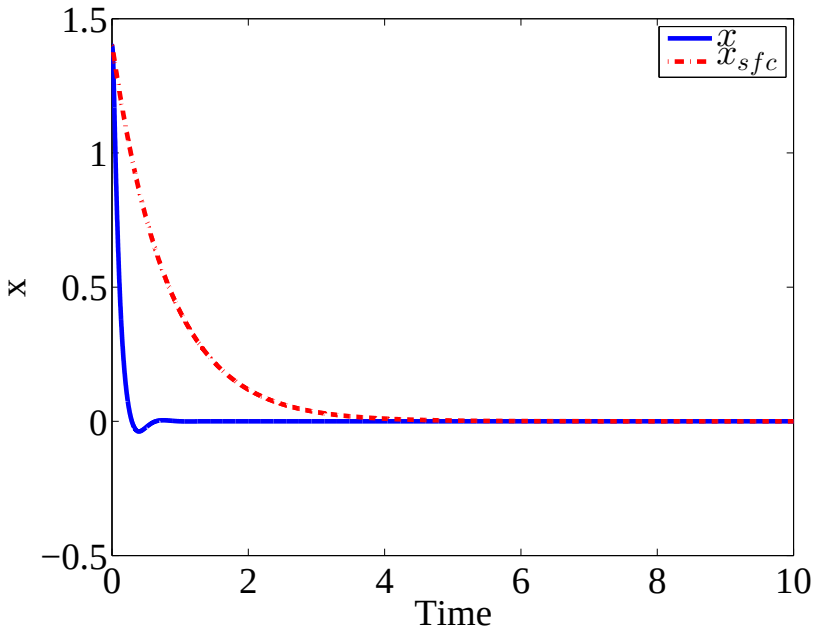


Figure 6.5: Control results of x when $b_2 = 10$.

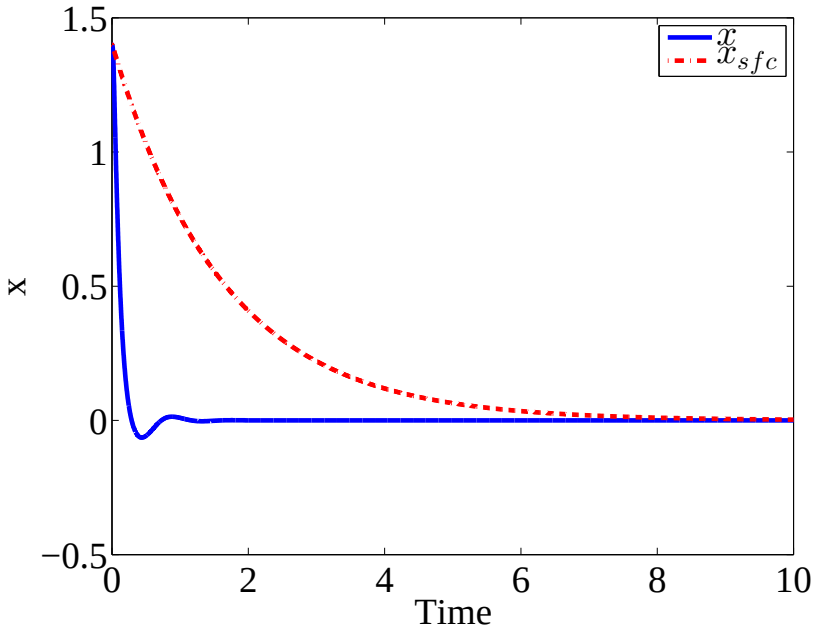


Figure 6.6: Control results of x when $a_1 = 0.5$ and $b_2 = 10$.

6.3 Controller Design for Trajectory Tracking Problem

In Section 6.2, an indirect adaptive controller is designed for a regulation problem using the PAS theory (Theorem 6.1). However, in Theorem 6.1, it is required that the system has to satisfy

the conditions $f_s(t,0,0,0) = 0$ and $g_s(t,0,0,0) = 0$. Hence it is not straightforward to apply this theory to a trajectory tracking problem. In this section, an indirect adaptive control scheme for a trajectory tracking problem is proposed without using the PAS theory. Instead, the overall stability of the closed-loop system is analyzed directly via Lyapunov approach. It is proved that the uniformly ultimately boundedness can be achieved by using the indirect adaptive controller proposed in this section.

6.3.1 Controller Design

The nonlinear singularly perturbed system (4.84) given in Section 4.4 is considered in this section:

$$\begin{aligned} \dot{x} &= f_1(x) + f_2(x)y, \\ \varepsilon\dot{y} &= g_1(x,y) + g_2(x,y)u. \end{aligned} \tag{4.84 revisited}$$

Also, the NN (4.85) defined in Section 4.4 is used to identify the nonlinear SPS (4.84). Assume the system states x of (4.84) track a given bounded reference $x_d \in C^2$. Define the tracking error as $E_x = x - x_d$ and denote the estimated tracking error as $\hat{E}_x = \hat{x} - x_d$. Then the tracking error can be rewritten as $E_x = x - \hat{x} + \hat{x} - x_d = \varsigma_x + \hat{E}_x$. This implies that the tracking error E_x can be minimized by designing a controller to control the identification result $\hat{x}$ such that the estimated tracking error $\hat{E}_x$ is minimized, since ς_x is bounded according to the identification scheme proposed in Section 4.4.

For a multi-time-scale RHONN model (4.85), because the parameter ε is very small, the changing rate of $\hat{y}$ will be very high. If $h(\hat{x})$ is the equilibrium point of (4.85b), $\hat{y}$ will converge to $h(\hat{x})$ rapidly. Specially, when $\varepsilon = 0$, the transient response of $\hat{y}$ will be instantaneous. Denote the tracking error of y as $E_y = y - h(\hat{x})$, and the estimated tracking error of y as $\hat{E}_y = \hat{y} - h(\hat{x})$. Then the estimated tracking error dynamics can be represented as:

$$\begin{aligned} \dot{\hat{E}}_x =& A\hat{E}_x + W_1\Psi_1(x) + W_2\Psi_2(x)(\hat{E}_y + h(\hat{x})) + Ax_d + L_x\varsigma_x - \dot{x}_d, \\ \varepsilon\dot{\hat{E}}_y =& B\hat{E}_y + W_3\Psi_3(x,y) + W_4\Psi_4(x,y)u + Bh(\hat{x}) + L_y\varsigma_y - \varepsilon\frac{dh}{dt}. \end{aligned} \tag{6.38}$$

By setting $\varepsilon = 0$ in (6.38), and using the fact that $\hat{E}_y = 0$ when $\varepsilon = 0$, one can obtain the *Reduced Slow Subsystem* as:

$$\dot{\hat{E}}_x = A\hat{E}_x + W_1\Psi_1(x) + W_2\Psi_2(x)h(\hat{x}) + Ax_d + L_x\varsigma_x - \dot{x}_d. \tag{6.39}$$

Define a new "stretched" time variable as $\tau = (t - t_0)/\varepsilon$, where t_0 is the initial time, and set $\varepsilon = 0$, then the *Reduced Fast Subsystem* is obtained as:

$$\frac{d\hat{E}_y}{d\tau} = B\hat{E}_y + W_3\Psi_3(x, y) + W_4\Psi_4(x, y)u + Bh(\hat{x}) + L_y\varsigma_y. \tag{6.40}$$

By decomposing the original error dynamics (6.38) into the *Reduced Slow Subsystem* (6.39) and the *Reduced Fast Subsystem* (6.40), the stabilization problem of (6.38) can be solved by designing two controllers to stabilize the slow and fast subsystems (6.39) and (6.40) separately. To guarantee the stability of the *Reduced Slow Subsystem* (6.39), $h(\hat{x})$ can be designed as:

$$\begin{aligned} h(\hat{x}) &= [W_2\Psi_2(x)]^{\dagger}h'(\hat{x}), \\ h'(\hat{x}) &= -Ax_d - W_1\Psi_1(x) - L_x\varsigma_x + \dot{x}_d, \end{aligned} \tag{6.41}$$

where † denotes the Moore-Penrose pseudo-inverse [119]. Substituting (6.41) into (6.39), one obtains:

$$\dot{\hat{E}}_x = A\hat{E}_x + \upsilon_x, \tag{6.42}$$

where $\upsilon_x = W_2\Psi_2(x)h(\hat{x}) - h'(\hat{x})$ is the bounded approximate solution error [31]. Specially, when $W_2\Psi_2(x)$ is a square nonsingular matrix, $\upsilon_x = 0$. Since A is a stable diagonal matrix, it is easy to show that $\hat{E}_x$ is UUB.

To ensure that the *Reduced Fast Subsystem* (6.40) is stable, one can design a control signal u

as:

$$\begin{aligned} u =& [W_4\Psi_4(x,y)]^{\dagger}u', \\ u' =& -W_3\Psi_3(x,y) - Bh(\hat{x}) - L_y\varsigma_y. \end{aligned} \tag{6.43}$$

Substituting (6.43) into (6.40), one obtains:

$$\frac{d\hat{E}_y}{d\tau} = B\hat{E}_y + \upsilon_y, \tag{6.44}$$

where $\upsilon_y = W_4\Psi_4(x,y)u - u'$ is the bounded approximate solution error. When $W_4\Psi_4(x,y)$ is a square nonsingular matrix, $\upsilon_x = 0$. Since B is a negative definite diagonal matrix, it is easy to show that $\hat{E}_y$ is also UUB.

Lemma 6.1. *It can be proved that as long as the system states x, y are bounded, then the perturbation term in* (6.38) *satisfies:*

$$\|\frac{dh}{dt}\|_2 \leq \beta. \tag{6.45}$$

Proof. Denote $W_2\Psi_2(x)$ as M. According to (6.41), one has:

$$\frac{dh}{dt} = \frac{dM^{\dagger}}{dt}[-Ax_d - W_1\Psi_1(x) - L_x\varsigma_x + \dot{x}_d] + M^{\dagger}[-A\dot{x}_d - \frac{dW_1\Psi_1(x)}{dt} - L_x\dot{\varsigma}_x + \ddot{x}_d], \tag{6.46}$$

where

$$\frac{dM^{\dagger}}{dt} = M^{\dagger}\frac{dM}{dt}M^{\dagger} + M^{\dagger}M^{\dagger T}\frac{dM^T}{dt}(I - MM^{\dagger}) + (I - M^{\dagger}M)\frac{dM^T}{dt}M^{\dagger T}M^{\dagger} \tag{6.47}$$

$$\frac{d[W_i\Psi_i(x)]}{dt} = \dot{W}_i\Psi_i(x) + W_i\frac{d\Psi_i(x)}{dt}, i = 1, 2 \tag{6.48}$$

$$\frac{d\Psi_i(x)}{dt} = \frac{d\Psi_i(x)}{dx}\dot{x}. \tag{6.49}$$

It should be noticed that x_d, W_1, W_2, $\Psi_1(x)$, $\Psi_2(x)$, ς_x, $\dot{x}_d$, $\ddot{x}_d$, $[W_2\Psi_2(x)]^{\dagger}$ are all bounded. Since e_{xi} is also bounded, then according to (4.99), $\dot{\varsigma}_x$ is bounded. Moreover, from (4.84), it is

obvious that when x and y are bounded, $\dot{x}$ is also bounded. Meanwhile, it is easy to known that $d\Psi_i(x)/dx$ is bounded, which means $d\Psi_i(x)/dt$ is bounded. From (4.102), since $\dot{\theta}_{xi}$ is bounded, then $\dot{W}_i$ is also bounded. Thus, all the terms in (6.46) are bounded, i.e.,

$$\|\frac{dh}{dt}\|_2 \leq \beta \tag{6.50}$$

is true for a certain positive constant β. Lemma 6.1 is thus proved. □

Theorem 6.3. *Consider the nonlinear system* (4.84) *with the RHONN given in* (4.85)*. If the control signal* u *is designed as* (6.43) *with* $h(\hat{x})$ *given in* (6.41)*, there exists a upper bound* $\varepsilon^* > 0$*, such that for all* $0 < \varepsilon < \varepsilon^*$*, the semi-global uniformly boundedness of the estimated tracking errors* $\|\hat{E}_x\|_2$ *and* $\|\hat{E}_y\|_2$ *can be guaranteed.*

Proof. Select the Lyapunov function as:

$$V = \hat{E}_x^T \hat{E}_x + \hat{E}_y^T \hat{E}_y. \tag{6.51}$$

The derivative of V is:

$$\dot{V} = 2\hat{E}_x^T \dot{\hat{E}}_x + 2\hat{E}_y^T \dot{\hat{E}}_y. \tag{6.52}$$

Substituting (6.38) into (6.52), and using (6.41) and (6.43), it follows that:

$$\dot{V} = 2\hat{E}_x^T (A\hat{E}_x + W_2\Psi_2(x)\hat{E}_y + \upsilon_x) + \frac{2}{\varepsilon}\hat{E}_y B\hat{E}_y - 2\hat{E}_y(\frac{dh}{dt} + \upsilon_y). \tag{6.53}$$

Using Lemma 6.1 and Young's inequality, it can be obtained that:

$$\begin{aligned} 2\hat{E}_x^T \upsilon_x &\leq 2\bar{\upsilon}_x \|\hat{E}_x\|_2 \leq \bar{\upsilon}_x^2 + \|\hat{E}_x\|_2^2, \\ -2\hat{E}_y \frac{dh}{dt} &\leq 2\beta \|\hat{E}_y\|_2 \leq \beta^2 + \|\hat{E}_y\|_2^2, \\ -2\hat{E}_y \upsilon_y &\leq 2\bar{\upsilon}_y \|\hat{E}_y\|_2 \leq \bar{\upsilon}_y^2 + \|\hat{E}_y\|_2^2, \end{aligned} \tag{6.54}$$

where $\bar{\upsilon}_x$ and $\bar{\upsilon}_y$ are the upper bounds of $\|\upsilon_x\|_2$ and $\|\upsilon_y\|_2$, respectively. According to Theorem 4.3, W_2 is bounded. Also, it is obvious that $\Psi_2(x)$ is bounded. Substituting (6.54) into (6.53), it can be obtained that:

$$\begin{aligned}\dot{V} \leq & -\alpha_1\|\hat{E}_x\|_2^2 + \alpha_2\|\hat{E}_x\|_2\|\hat{E}_y\|_2 - \frac{\alpha_3}{\varepsilon}\|\hat{E}_y\|_2^2 + \beta^2 + \bar{\upsilon}_x^2 + \bar{\upsilon}_y^2 \\ = & -\begin{bmatrix}\|\hat{E}_x\|_2 & \|\hat{E}_y\|_2\end{bmatrix}\Lambda\begin{bmatrix}\|\hat{E}_x\|_2 \\ \|\hat{E}_y\|_2\end{bmatrix} + \beta^2 + \bar{\upsilon}_x^2 + \bar{\upsilon}_y^2, \end{aligned} \tag{6.55}$$

where α_i, $i = 1, 2, 3$ are some positive constants, and Λ is given by

$$\Lambda = \begin{bmatrix}\alpha_1 & -\frac{\alpha_2}{2} \\ -\frac{\alpha_2}{2} & \frac{\alpha_3}{\varepsilon}\end{bmatrix}. \tag{6.56}$$

It should be pointed out that α_1, α_3 are dependent on A and B, respectively. The value of α_1, α_3 can be set arbitrarily large by increasing A, B. From (6.56), it is easy to show that Λ will be positive definite if only

$$\varepsilon < \frac{4\alpha_1\alpha_3}{\alpha_2^2} \equiv \varepsilon^*. \tag{6.57}$$

Hence, there exist a $\varepsilon^* > 0$, such that for all $0 < \varepsilon < \varepsilon^*$, $\dot{V} \leq -\gamma V + \beta^2 + \bar{\upsilon}_x^2 + \bar{\upsilon}_y^2$ for some $\gamma > 0$. Therefore it can be concluded that the semi-global uniformly boundedness of the estimated tracking errors is guaranteed, and Theorem 6.3 is thus proved. □

Remark 6.5. *According to* (6.57)*, one can always guarantee that the error dynamics* (6.38) *is semi-global uniformly stable by choosing suitable A and B such that ε^* is large enough.*

Remark 6.6. *According to Theorem 6.3, the estimated tracking error $\|\hat{E}_x\|_2$ is bounded. This implies the tracking error E_x is bounded since $E_x = \varsigma_x + \hat{E}_x$, and the identification error ς_x is bounded.*

Remark 6.7. *In Section 6.2, the controller is designed for the regulation problem only. In this*

paper, an indirect adaptive controller is designed for the trajectory tracking problem based on the identified system model and the UUB of the closed-loop system is proved.

6.3.2 Experiment

To verify the effectiveness of the proposed control algorithms, the experiments on a harmonic drive system are conducted. The harmonic drive system used in this experiment is the same as the one used in Section 4.4. As presented in Section 4.4, the system model of a typical DC motor is given as:

$$\begin{aligned} J\frac{d\omega}{dt} &= k_t i, \\ L\frac{di}{dt} &= -k_b\omega - Ri + V, \end{aligned} \tag{6.58}$$

where J is the moment of inertia, ω is the angular velocity, k_t is the torque force constant, i armature current, L is the armature inductance, k_b is the back electromotive force constant, R, V are the armature resistance and voltage, respectively. Define x as the angular velocity, y as the current, and u as the control signal. The RNN model (4.85) with modified OBE algorithm based updating laws (4.101)-(4.106) proposed in Section 4.4 will be used to identify the unknown DC motor system, and based on the identification results, the controller designed in Section 6.3 is used to control the harmonic drive system to track a given reference signal

$$x_d = 2.1\text{sin}(5t).$$

In this experiment, $\varepsilon = 0.1$ is obtained based on the prior knowledge of the setup, and the following NN parameters are used: $A = -20$, $B = -20$, $\lambda_x = 500$, $\lambda_y = 500$, $L_x = 20$, $L_y = 10$, $g_x = g_y = 0.332$, $\bar{\zeta}_x = 0.001$, $\bar{\zeta}_y = 0.001$, the number of neuron $q = 2$, $W_1 \in \Re^{1\times 2}$, $W_2 \in \Re$, $W_3 \in \Re^{1\times 2}$, $W_4 \in \Re$, $W_1(t_0) = W_3(t_0) = \mathbf{0} \in \Re^{1\times 2}$, $W_2(t_0) = W_4(t_0) = 0.5$, $P_{xi}(t_0) = \text{diag}([50, 50, 50])$, $P_{yj}(t_0) = \text{diag}([50, 50, 50])$, the activation functions are chosen as $\psi_1(z) = 2/(1 + e^{-0.5z}) + 1$, $\psi_2(z) = 1/(1 + e^{-0.2z}) + 10$, $\psi_3(z) = 2/(1 + e^{-0.1z}) + 1$, $\psi_4(z) = 1/(1 + e^{-0.1z}) + 5$. For comparison purpose, the original OBE based weight's updating laws

proposed in Section 4.3, and the GD based weight's updating laws proposed in [112] are also tested. It should be pointed out that the OBE and GD based scheme are only used to train the NN weights. The indirect adaptive controllers remain the same when using different identification algorithm. Also, the traditional PID control method is used in order to compare the control performance. The sampling time for the experiment is 0.1 ms. The experimental results are presented in Figs. 6.7-6.13.

Remark 6.8. *According to* (6.55)*, larger A and B will result in faster convergence of the estimated tracking errors. However, if A and B are too large, severe oscillation will be observed. This situation also happens to λ_{xi}, λ_{yj}, and L_x, L_y given in Section 4.4. In the experiment, one can set these parameters as small values at beginning, and then increase them gradually until the desired results are achieved. The terms g_{xi}, g_{yj} should satisfy* (4.107)*, and also should be chosen to be large enough such that $H_{xi}^T P_{xi} H_{xi} > \bar{\zeta}_{yj}^2$, $H_{yj}^T P_{yj} H_{yj} > \bar{\zeta}_{yj}^2$ can be always guaranteed, and a higher convergence speed of the weights can be achieved, as indicated in* (4.122)*. The parameters $\bar{\zeta}_{xi}$, $\bar{\zeta}_{yj}$ can be chosen as two small values, which do not affect the identification results very much. The initial values of W_1 and W_3 are usually set to be 0, and the initial values of the elements in W_2 and W_4 are usually set to be small non-zero values. The initial values of P_{xi}, P_{yj} can be chosen as arbitrary positive definite matrices. For the activation functions, usually setting $\alpha_{i,1} = \alpha_{i,2} = \alpha_{i,3}$=1 can result in an acceptable result. To improve the result, one can change these values via trial and error method.*

From Figs. 6.7-6.13, it is clear that the modified OBE based identification and control algorithm can achieve the best performance among all results. Fig. 6.7 and Fig. 6.11 show that at beginning, there are visible difference between the desired angular velocity x_d, the real velocity x and the estimation $\hat{x}$, as well as the desired current $h(\hat{x})$, the real current y and the estimation $\hat{y}$ at peak and the lowest point of the trajectory. However, after a short learning period (about 6s), the differences among them became negligible. When the original OBE based algorithm proposed in Section 4.3 is used, the identification and control performances are very good at the beginning. The x, $\hat{x}$ and y, $\hat{y}$ converge to x_d and $h(\hat{x})$ even faster than the modified OBE and GD based methods. Nevertheless,

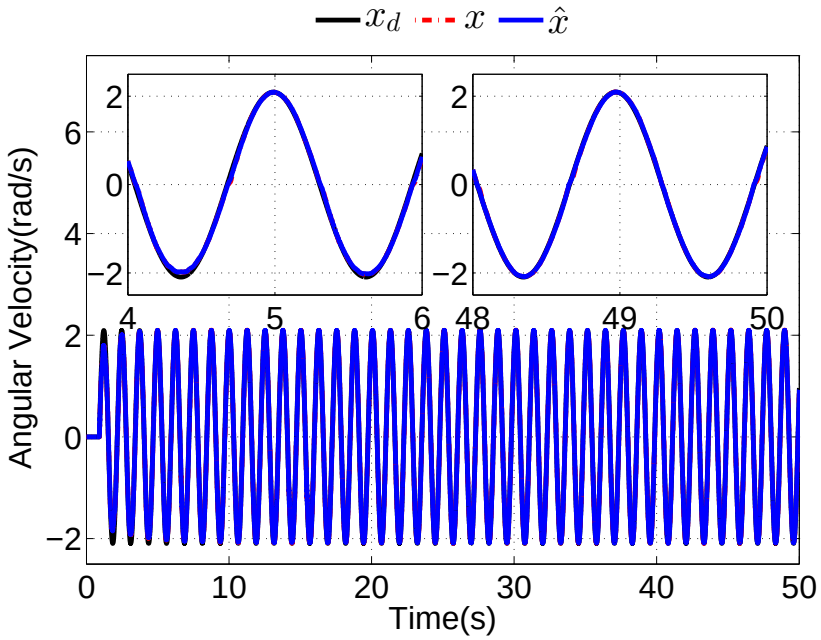


Figure 6.7: Identification and control results of velocity using the modified OBE.

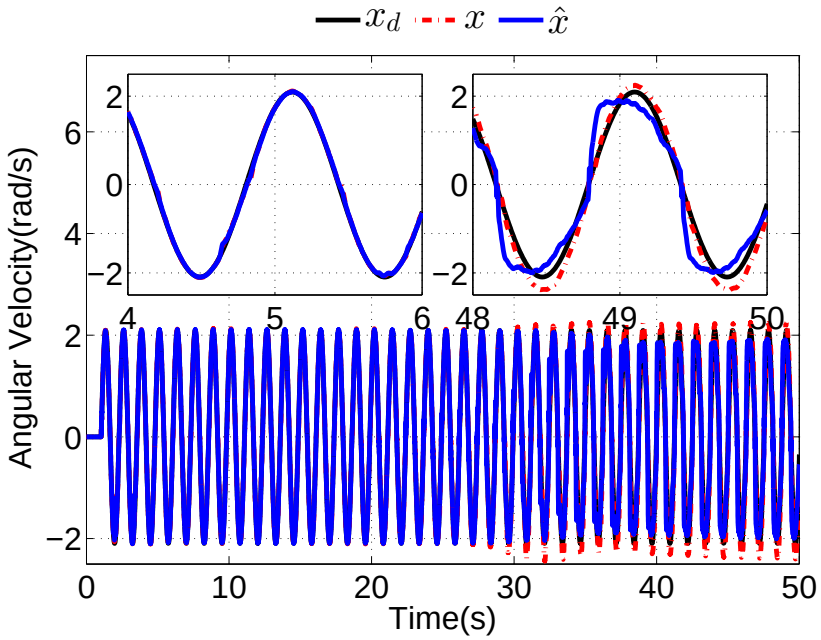


Figure 6.8: Identification and control results of angular velocity using OBE.

after 25 seconds, because P_x and P_y converge to $\mathbf{0}^{3\times3}$, the weight's updating laws are no longer effective, and large gaps between the references and the real signals can be observed, as shown in Fig. 6.8 and Fig. 6.12. When GD based algorithm is used, it takes much longer time for x, $\hat{x}$ and y, $\hat{y}$ to converge to their references x_d and $h(\hat{x})$, as shown in Fig. 6.9 and Fig. 6.13. Even at the

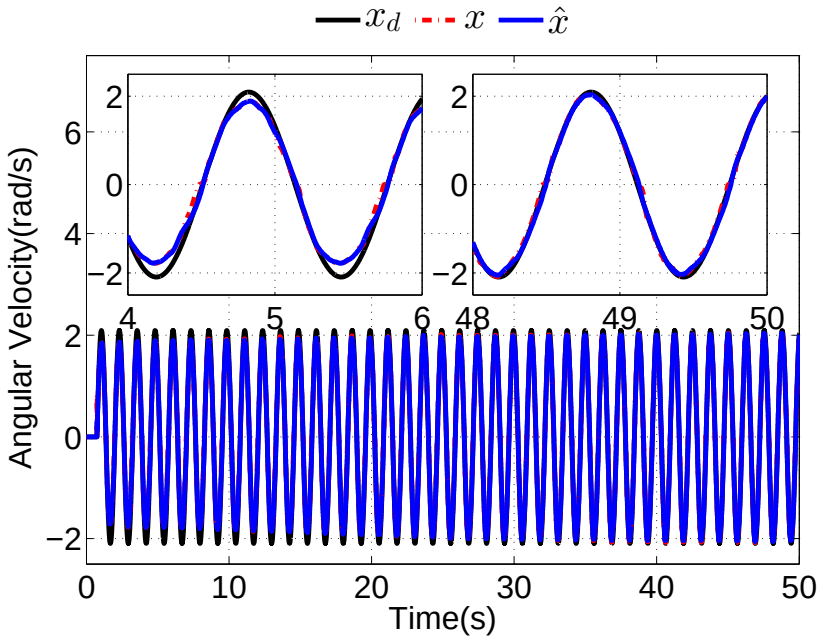


Figure 6.9: Identification and control results of angular velocity using GD.

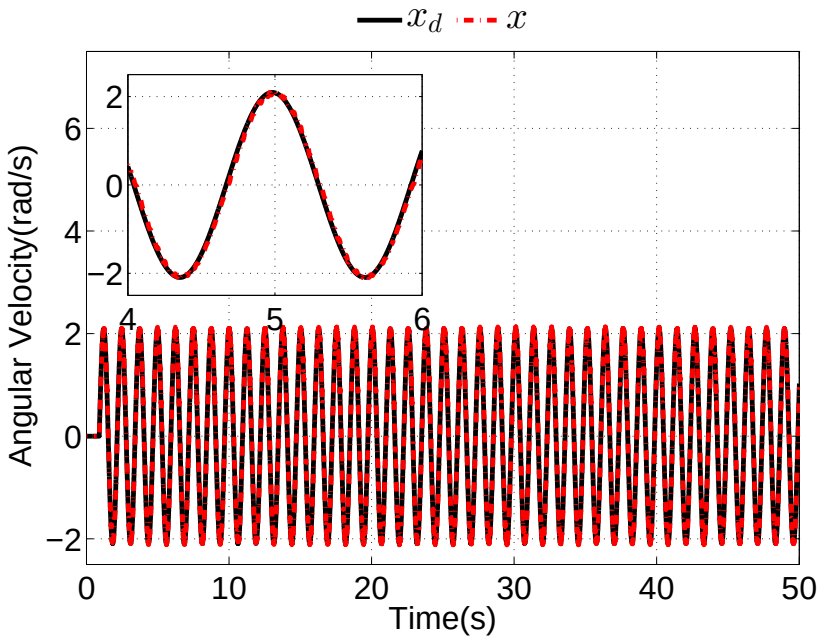


Figure 6.10: Control result of angular velocity using PID.

end of the experiment, the differences between the references and real signals are obvious. When the PID control is applied, the system state x can always track the reference signal x_d, but a larger gap between x and x_d can be observed, as shown in Fig. 6.10, compared with the results shown in Fig. 6.7.

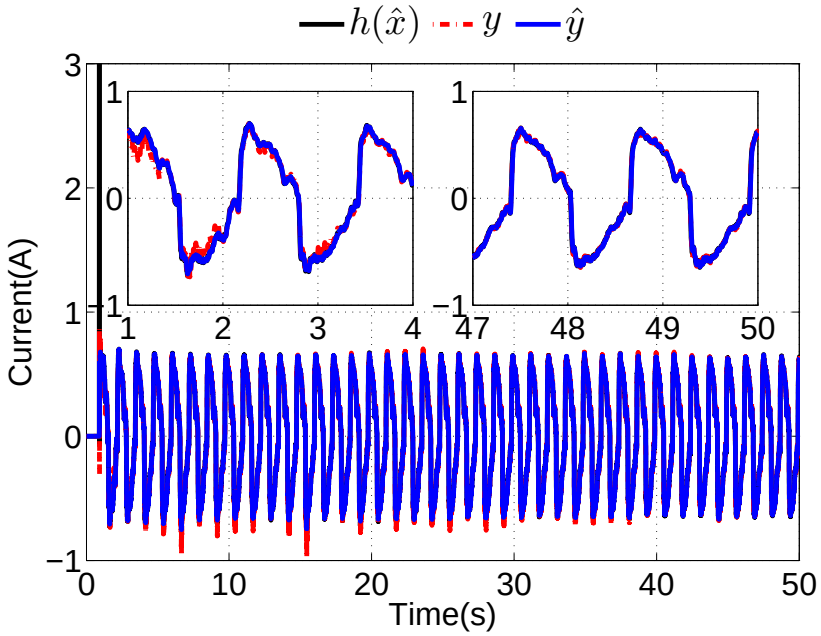


Figure 6.11: Identification and control results of current using modified OBE.

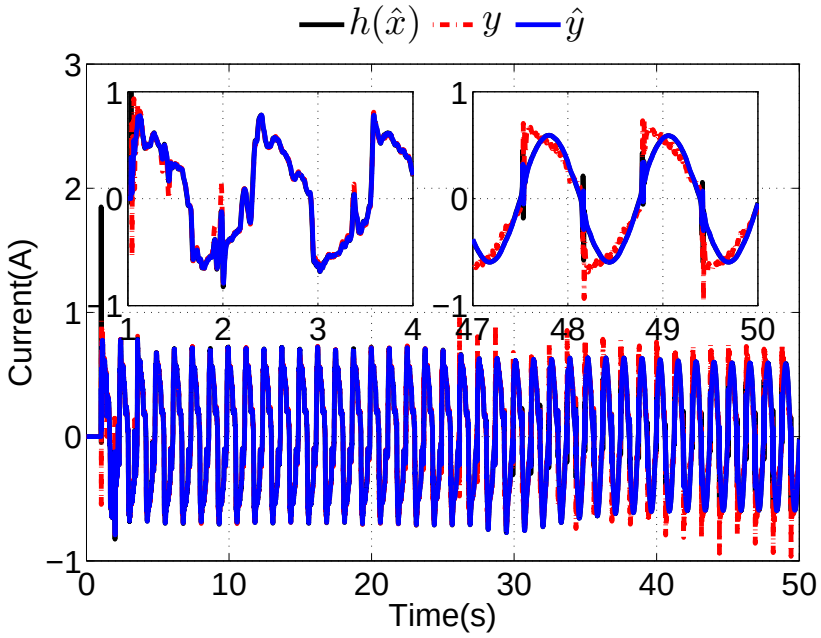


Figure 6.12: Identification and control results of current using OBE.

In Fig. 6.14, it is shown that when PID controller is used, a consistent tracking error occurs during the whole experiment, and the magnitude of the tracking error remains the same. When the modified OBE based control algorithm is use, the tracking error is smaller than the error obtained using the PID controller at the beginning, and decreases dramatically at the end of the experiment.

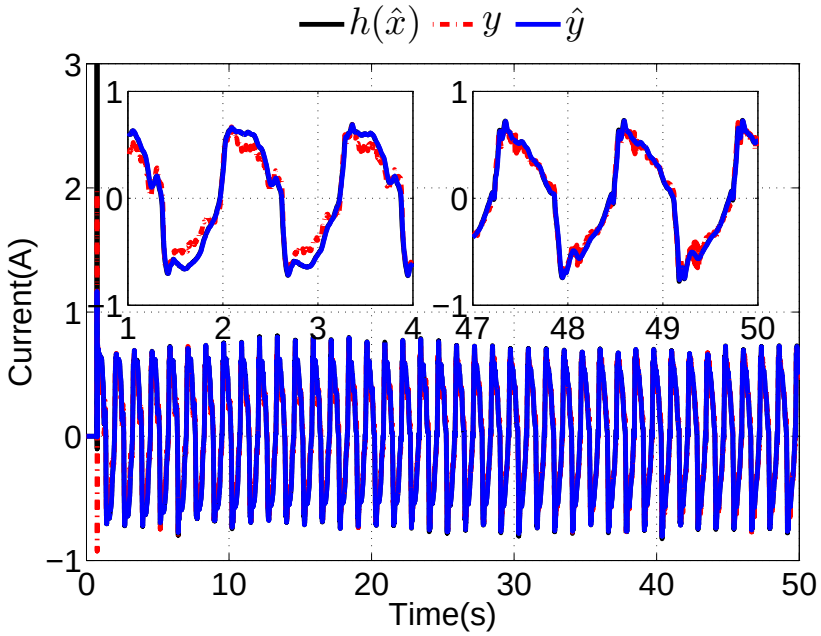


Figure 6.13: Identification and control results of current using GD.

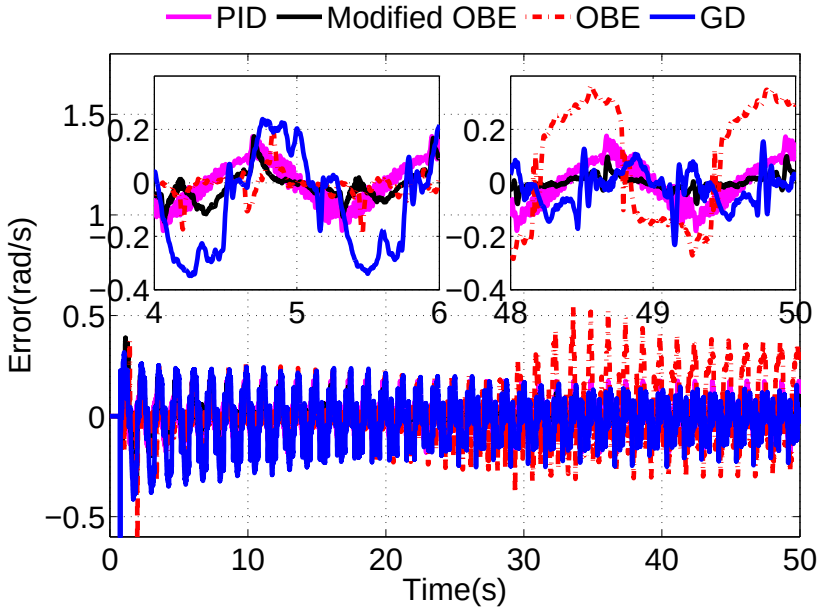


Figure 6.14: Tracking errors of angular velocity.

When the GD based control algorithm is used, the tracking error is larger than the error achieved using the PID controller at the beginning, but it becomes slightly smaller than the error obtained using the PID controller at the end of the experiment. When the original OBE based control algorithm is used, the tracking error is small at the beginning, but the magnitude increases to a

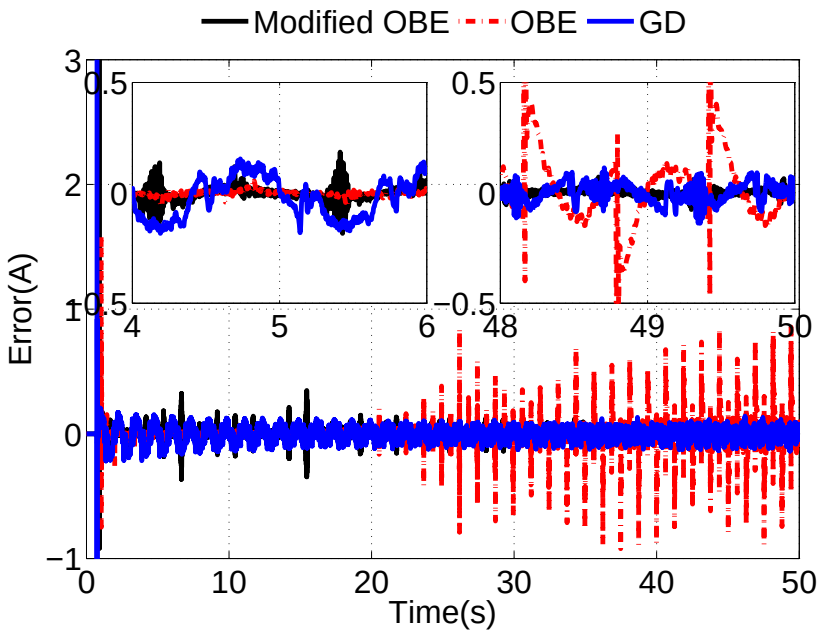


Figure 6.15: Tracking errors of current.

very large scale at the end of the experiment. In Fig. 6.15, similar phenomena is observed. At the beginning of the control process, the original OBE based controller can achieve the minimum tracking error, and the GD based controller has the largest tracking error. However, at the end of the experiment, the modified OBE based controller has the minimum tracking error, while the original OBE based controller has the largest tracking error.

To further compare the performance of different identification and control methods, the performance index-ITAE is also calculated. The ITAE is defined as:

$$\text{ITAE} = \int_0^T t|e(t)|dt. \tag{6.59}$$

The results of the ITAE calculation are presented in Table 6.1, where $ITAE_{xt}$, $ITAE_{yt}$, are the ITAE values of the tracking error E_x and E_y respectively. From Table 6.1, it is clear that the control errors are much smaller when the modified OBE proposed in Section 4.4 and the indirect adaptive controller proposed in Section 6.3 are used, compared with the errors using the other methods.

Table 6.1: ITAE values of E_x and E_y

	Modified OBE	OBE	GD	PID
$ITAE_{xt}$	35.2	277.9	119.5	122.2
$ITAE_{yt}$	24.5	208.2	71.0	N/A

6.4 Conclusion

In this chapter, two indirect adaptive controllers are proposed based on the system models obtained by using the identification schemes proposed in Chapter 4. To design these two controllers, the singular perturbation theory is used firstly to decompose the original high-order multi-time-scale nonlinear system into two reduced order subsystems. Then the controllers are designed for the reduced order subsystems instead of the original high order system. Thus the complexity of the controller design problem is simplified, and the required computational resource is also reduced. Meanwhile, because the term $1/\varepsilon$ is no longer involved in the controllers, it is less likely to have the potential singularity problem. Specifically, the first indirect adaptive controller is designed for a regulation problem based on the PAS theory. By using the Lyapunov analysis approach, the second indirect adaptive controller is proposed without using the PAS theory and thus can be applied to a trajectory tracking problem. The stability of the closed-loop systems is guaranteed, and the validity of the proposed controllers is demonstrated via simulations and experiments.

Chapter 7

Conclusions and Future Works

7.1 Conclusions

In this Ph.D research, the identification and control of the nonlinear singularly perturbed systems using multi-time-scale neural networks are investigated. Several new identification schemes are proposed to identify the nonlinear singularly perturbed systems with unknown system models or parameters. These novel identification schemes can achieve high identification accuracy with fast convergence. Based on the identified system models, three indirect adaptive controllers are proposed to control the nonlinear system adaptively. The main results of this research are listed as follows:

- An identification scheme using the multilayer neural network is proposed in Section 3.3. When this multilayer NN identification scheme is used, the outputs of the NN follow those of the nonlinear system more accurately and quickly compared to the results in [26]. The eigenvalues of the linear parameter matrices are universally smaller than zero during the identification process, which means the identification scheme remains stable.

- To achieve faster convergence, the OBE algorithm based identification scheme is proposed in Section 4.2 for a discrete time system. The identification results of the system states are much more accurate when the identification algorithm proposed in Section 4.2 is used.

The identification errors are greatly reduced compared with the results using the method proposed in [16]. Also, when using the method proposed in Section 4.2, the NN converges much faster than using the identification algorithm proposed in [16]. This is due to that fact that the learning gain of the weight's updating law in the method proposed in Section 4.2 can vary adaptively. However, in [16] and in many other widely used learning algorithms, the learning gain is fixed.

- In Section 4.3, the discrete time OBE based identification scheme is extended to a continuous case. In the continuous OBE based identification algorithm, the learning gain can also be adjusted adaptive during the identification process. The simulation results show that when the identification scheme proposed in Section 4.3 is used, the estimated system states can converge to the real system states more precisely and quickly, compared with the results achieved when the multilayer NN identification scheme proposed in Section 3.3, or the single layer NN identification scheme proposed in [16] are used. Also, the identification errors can be greatly reduced when the continuous OBE based identification algorithm is used.

- A modified OBE based identification scheme is proposed in Section 4.4, where two extra terms are added into the weight's updating laws. Although the OBE based identification algorithm proposed in Section 3.3 and Section 4.2 can achieve high identification accuracy with fast convergence, it is found that when the identification errors may increase at the end of the identification process if the identified system model is used to design indirect adaptive controller. This is due to the fact that the learning gain of the OBE algorithm based identification scheme will keep decreasing during the identification process. When the learning gain is too small, the identification method will lose its ability to adjust the NN weights. In the modified OBE based identification algorithm, two additional terms $g_{xi}P_{xi}$, $g_{yj}P_{yj}$ are introduced. Hence, the terms P_{xi} and P_{yj} will converge to the equilibrium points P_{xie} and P_{yje} which are determined by g_{xi}, H_{xi}, g_{yj}, and H_{yj}. Thus, the NN weight's updating laws will remain effective during the identification process.

- A robust identification scheme using filtered variables is proposed in Chapter 5 and the derivatives of the identification errors are no longer needed. When there is no measurement noise, both the modified identification scheme proposed in Section 4.4 and the robust identification scheme proposed in Chapter 5 can achieve satisfactory performance. When the measured system states contain noises, the identification errors remain small when the robust identification scheme is used, however, large identification errors can be observed during the identification process, and the magnitudes of the errors keep increasing, when the modified OBE based identification algorithm is used. Because in the modified OBE based algorithm, the noises will be amplified when the measured system stated are differentiated. However, in the robust OBE based identification algorithm, the derivatives of the system states are not needed, and the identification scheme will remain effective during the identification process.

- Based on the identified models, an indirect adaptive controller using feedback linearisation and sliding mode technique is designed in Section 3.4, where the identified model is treated as a regular system. By using the identification and control scheme proposed in Section 3.4, the closed-loop system can track the given reference signal more precisely. The tracking errors are greatly reduced when compared with the control results using the method proposed in [26]. Meanwhile, it can be observed that it takes relatively more time for the slow system states to track the reference signals than the fast system states, because the small parameter ε accelerates both the identification and trajectory tracking process of fast system states.

- To solve the regulation problem, an indirect adaptive controller is designed in Section 6.2 based on the PAS theory. In order to simplify the controller structure and reduce the required computational resources, the singular perturbation theory is employed to decompose the high order multi-time-scale system into two reduced order subsystems, and then the indirect controllers are designed for the reduced order subsystems. It is shown that when the indirect adaptive controller is used, the system state x can always converge to 0 fast. Even if the system parameters changes, the system response remain almost the same. However, when

the state feedback controller [121] is used, only slower convergence can be achieve. If the system parameter changes, the control results deteriorate.

- In Section 6.3, an adaptive controller is designed to solve a trajectory tracking problem. Through Lyapunov approach, the upper bound of ε is found, and the closed-loop stability is guaranteed for any $0 < \varepsilon < \varepsilon^*$. Meanwhile, it is shown in the experiment that when the modified OBE based controller is used, the best tracking performance can be achieved. The tracking errors are small at the beginning and will keep decreasing during the identification process.

- A harmonic drive DC motor system is built up. The effectiveness of the identification and control schemes proposed in this research are verified via practical experiments.

7.2 Future Works

Although this Ph.D research has achieved some remarkable results, there are still many problems that needs further investigation:

- So far, all system states are considered to be measurable. If some system states cannot be measured, an observer could be utilized to observe those immeasurable system states, then theorems and proofs in this thesis should be modified to consider the observer errors as well. This topic could be considered in the future work.

- In Chapter 6, the controllers are designed for the affine-in-control cascade systems. How to design a controller for a non-affine-in-control non-cascade system is still challenging.

- The controllers are only designed for the regulation or trajectory tracking problems. Also, no constraint on the systems is considered. It is worth discussing other kinds of control problems, such as the optimal control problem under system constraints.

Bibliography

[1] P. S. Gandhi and F. Ghorbel, "High-speed precision tracking with harmonic drive systems using integral manifold control design," *International Journal of Control*, vol. 78, no. 2, pp. 112–121, 2005.

[2] P. K. A. Menon, M. E. Badgett, and R. A. Walker, "Nonlinear flight test trajectory controllers for aircraft," *Journal of Guidance, Control, and Dynamics*, vol. 10, no. 1, pp. 67–72, 1987.

[3] M. Vakil, R. Fotouhi, and P. Nikiforuk, "End-effector trajectory tracking of a flexible link manipulator using integral manifold concept," *International Journal of Systems Science*, vol. 42, no. 12, pp. 2057–2069, 2011.

[4] J. R. Winkelman, J. H. Chow, J. J. Allemong, and P. V. Kokotovic, "Multi-time-scale analysis of a power system," *Automatica*, vol. 16, no. 1, pp. 35–43, 1980.

[5] P. Kokotovic, H. K. Khalil, and J. O'Reilly, *Singular Perturbation Methods in Control: Analysis and Design*, ser. Classics in Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics, 1999, vol. 25.

[6] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*. Strtz GmbH, Germany: Springer-Verlag Berlin Heidelberg, 2008.

[7] A. J. Calise, "Singular perturbation methods for variational problems in aircraft flight," *IEEE Transactions on Automatic Control*, vol. 21, no. 3, pp. 345–353, Jun 1976.

[8] "Robotics and mechatronics center," http://www.dlr.de/rm/en/desktopdefault.aspx/tabid-9277/15984_read-39321, accessed:01-Dec-2016.

[9] "General dynamics f-16 fighting falcon," https://en.wikipedia.org/wiki/General_Dynamics_F-16_Fighting_Falcon, accessed:01-Dec-2016.

[10] Y. Zhang, D. S. Naidu, C. Cai, and Y. Zou, "Singular perturbations and time scales in control theories and applications: An overview 2002-2012," *International Journal of Information and Systems Sciences*, vol. 9, no. 1, pp. 1–36, 2014.

[11] D. S. Naidu, *Singular Perturbation Methodology in Control Systems*, ser. IEE Control Engineering Series 34, D. P. Atherton and K. Warwich, Eds. London, United Kingdom: Peter Peregrinus Ltd., 1988.

[12] V. Saksena, J. O'Reilly, and P. Kokotovic, "Singular perturbations and time-scale methods in control theory: Survey 19761983," *Automatica*, vol. 20, no. 3, pp. 273–293, 1984.

[13] D. S. Naidu and A. J. Calise, "Singular perturbations and time scales in guidance and control of aerospace systems: A survey," *Journal of Guidance, Control, and Dynamics*, vol. 24, no. 6, pp. 1057–1078, 2001.

[14] C. Lobry and T. Sari, "Singular perturbations methods in control theory," in *Contrôle non linéaire et Applications, Les Cours du CIMPA*. Paris, France: Hermann, Editeur des Sciences et des arts, 2005, pp. 151–177.

[15] B. S. Heck, "Sliding-mode control for singularly perturbed systems," *International Journal of Control*, vol. 53, no. 4, pp. 985–1001, 1991.

[16] X. Han, W.-F. Xie, Z. Fu, and W. Luo, "Nonlinear systems identification using dynamic multi-time scale neural networks," *Neurocomputing*, vol. 74, no. 17, pp. 3428–3439, 2011.

[17] Z. Fu, W.-F. Xie, and W. Luo, "Induction motor identification using dynamic two-time scales neural networks with sliding mode learning," in *25th IEEE Canadian Conference on Electrical & Computer Engineering (CCECE 2012)*, Apr. 2012, pp. 1–6.

[18] ——, "Robust on-line nonlinear systems identification using multilayer dynamic neural networks with two-time scales," *Neurocomputing*, vol. 113, no. 0, pp. 16–26, 2013.

[19] W. Yu and A. C. Sandoval, "Some new stability properties of dynamic neural networks with different time-scales," *International Journal of Neural Systems*, vol. 16, no. 3, pp. 191–199, 2006.

[20] W. Yu and X. Li, "Passivity analysis of dynamic neural networks with different time-scales," *Neural Processing Letters*, vol. 25, no. 2, pp. 143–155, 2007.

[21] A. C. Sandoval, W. Yu, and X. Li, "Some stability properties of dynamic neural networks with different time-scales," in *Proceedings of the IEEE 2006 International Joint Conference on Neural Networks*, Vancouver, Canada, Jul. 2006, pp. 4218–4224.

[22] A. Meyer-Baese, R. Roberts, and S. Joshi, "Robust stability analysis of a class of noise perturbed two-time scale neural networks," in *Proceedings of the IEEE 2006 International Joint Conference on Neural Networks*, Vancouver, Canada, Jul. 2006, pp. 1029–1032.

[23] A. Bazaei and V. J. Majd, "Neuro-adaptive friction compensation for single-link flexible robots using serial-gray-box modeling strategy," *Journal of Dynamic Systems, Measurement, and Control*, vol. 128, no. 2, pp. 297–306, 2005.

[24] S. S. Ge, T. H. Lee, and Z. P. Wang, "Adaptive neural network control for smart materials robots using singular perturbation technique," *Asian Journal of Control*, vol. 3, no. 2, pp. 143–155, 2001.

[25] F. Lewis, S. Jagannathan, and A. Yesildirak, *Neural Network Control of Robot Manipulators and Non-Linear Systems*. Bristol, PA: Taylor & Francis, Inc., 1998.

[26] X. Han and W.-F. Xie, "Nonlinear systems identification and control using dynamic multi-time scales neural networks," in *Proceedings of the IEEE International Conference on Automation and Logistics*, Shenyang, China, Aug. 2009, pp. 96–101.

[27] Z. Fu, W.-F. Xie, X. Han, and W. D. Luo, "Nonlinear systems identification and control via dynamic multitime scales neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 11, pp. 1814–1823, Nov. 2013.

[28] Y. Iiguni, H. Sakai, and H. Tokumaru, "A real-time learning algorithm for a multilayered neural network based on the extended Kalman filter," *IEEE Transactions on Signal Processing*, vol. 40, no. 4, pp. 959–966, Apr. 1992.

[29] W. Yu and J. de Jesús Rubio, "Recurrent neural networks training with stable bounding ellipsoid algorithm," *IEEE Transactions on Neural Networks*, vol. 20, no. 6, pp. 983–991, Jun. 2009.

[30] D.-D. Zheng and W.-F. Xie, "Identification and trajectory tracking control of nonlinear singularly perturbed system," *IEEE Transactions on Industrial Electronics*, 2016, accepted.

[31] D.-D. Zheng, W.-F. Xie, X. Ren, and J. Na, "Identification and control for singularly perturbed systems using multitime-scale neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, no. 99, pp. 1–13, 2016.

[32] D.-D. Zheng, Z. Fu, W.-F. Xie, and W. Luo, "Indirect adaptive control of nonlinear system via dynamic multilayer neural networks with multi-time scales," *International Journal of Adaptive Control and Signal Processing*, vol. 29, no. 4, pp. 505–523, 2015.

[33] D.-D. Zheng and W.-F. Xie, "Indirect adaptive control of flexible joint robotic manipulator," in *Proceedings of the World Congress of the International Federation of Automatic Control*, Submitted for publication.

[34] ——, "Robust identification for nonlinear singularly perturbed systems using multi-time-scale recurrent neural network," in *Proceedings of the American Control Conference*, Submitted for publication.

[35] D.-D. Zheng, W.-F. Xie, and X. Ren, "Identification and control for singularly perturbed systems using multi-time-scale neural networks," in *Proceedings of the 2015 IEEE International Conference on Information and Automation*, Lijiang, China, Aug. 2015, pp. 1233–1239.

[36] D.-D. Zheng and W.-F. Xie, "Identification for nonlinear singularly perturbed system using recurrent high-order multi-time scales neural network," in *Proceedings of the American Control Conference*, Chicago, USA, Jul. 2015, pp. 1824–1829.

[37] D.-D. Zheng, W.-F. Xie, and S. Dai, "Identification of singularly perturbed nonlinear system using recurrent high-order neural network," in *Proceedings of the 11th World Congress on Intelligent Control and Automation*, Shenyang, China, Jul. 2014, pp. 5770–5775.

[38] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[39] P. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences," Ph.D. dissertation, Harvard University, 1974.

[40] K. Hunt, D. Sbarbaro, R. bikowski, and P. Gawthrop, "Neural networks for control systemsa survey," *Automatica*, vol. 28, no. 6, pp. 1083–1112, 1992.

[41] S. Lu and T. Basar, "Robust nonlinear system identification using neural-network models," *IEEE Transactions on Neural Networks*, vol. 9, no. 3, pp. 407–429, May 1998.

[42] J. G. Kuschewski, S. Hui, and S. Zak, "Application of feedforward neural networks to dynamical system identification and control," *IEEE Transactions on Control Systems Technology*, vol. 1, no. 1, pp. 37–49, 1993.

[43] A. Alessandri, R. Bolla, M. Gaggero, and M. Repetto, "Modeling and identification of nonlinear dynamics for freeway traffic by using information from a mobile cellular network," *IEEE Transactions on Control Systems Technology*, vol. 17, no. 4, pp. 952–959, Jul. 2009.

[44] W. He, A. O. David, Z. Yin, and C. Sun, "Neural network control of a robotic manipulator with input deadzone and output constraint," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 6, pp. 759–770, Jun. 2016.

[45] "Adaptive neural network predictive control for nonlinear pure feedback systems with input delay," *Journal of Process Control*, vol. 22, no. 1, pp. 194–206, 2012.

[46] Y.-J. Liu, J. Li, S. Tong, and C. L. P. Chen, "Neural network control-based adaptive learning design for nonlinear systems with full-state constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 7, pp. 1562–1571, Jul. 2016.

[47] X. Li and W. Yu, "Dynamic system identification via recurrent multilayer perceptrons," *Information Sciences*, vol. 147, no. 1, pp. 45–63, 2002.

[48] X. Ren, A. B. Rad, P. Chan, and W. L. Lo, "Identification and control of continuous-time nonlinear systems via dynamic neural networks," *IEEE Transactions on Industrial Electronics*, vol. 50, no. 3, pp. 478–486, Jun. 2003.

[49] A. S. Poznyak, W. Yu, E. N. Sanchez, and J. P. Perez, "Nonlinear adaptive trajectory tracking using dynamic neural networks," *IEEE Transactions on Neural Networks*, vol. 10, no. 6, pp. 1402–1411, Nov. 1999.

[50] F. L. Lewis, A. Yesildirek, and K. Liu, "Multilayer neural-net robot controller with guaranteed tracking performance," *IEEE Transactions on Neural Networks*, vol. 7, no. 2, pp. 388–399, Mar. 1996.

[51] W. Chen and L. Jiao, "Adaptive tracking for periodically time-varying and nonlinearly parameterized systems using multilayer neural networks," *IEEE Transactions on Neural Networks*, vol. 21, no. 2, pp. 345–351, 2010.

[52] J. de Villiers and E. Barnard, "Backpropagation neural nets with one and two hidden layers," *IEEE Transactions on Neural Networks*, vol. 4, no. 1, pp. 136–141, 1993.

[53] S. Bhama and H. Singh, "Single layer neural networks for linear system identification using gradient descent technique," *IEEE Transactions on Neural Networks*, vol. 4, no. 5, pp. 884–888, Sep. 1993.

[54] P. M. Patre, W. MacKunis, K. Kaiser, and W. E. Dixon, "Asymptotic tracking for uncertain dynamic systems via a multilayer neural network feedforward and rise feedback control structure," *IEEE Transactions on Automatic Control*, vol. 53, no. 9, pp. 2180–2185, Oct. 2008.

[55] S. Jagannathan, "Control of a class of nonlinear discrete-time systems using multilayer neural networks," *IEEE Transactions on Neural Networks*, vol. 12, no. 5, pp. 1113–1120, Sep. 2001.

[56] E. B. Kosmatopoulos, M. M. Polycarpou, M. A. Christodoulou, and P. A. Ioannou, "High-order neural network structures for identification of dynamical systems," *IEEE Transactions on Neural Networks*, vol. 6, no. 2, pp. 422–431, 1995.

[57] A. Y. Alanis, E. N. Sanchez, and A. G. Loukianov, "Discrete-time adaptive backstepping nonlinear control via high-order neural networks," *IEEE Transactions on Neural Networks*, vol. 18, no. 4, pp. 1185–1195, 2007.

[58] D.-D. Zheng, J. Na, X. Ren, G. Herrmann, and S. Longo, "Adaptive control of robotic servo system with friction compensation," in *Proceedings of the 5th IEEE International Conference on Robotics, Automation and Mechatronics*, Sep. 2011, pp. 285–290.

[59] J. Na, J. Yang, X. Ren, and Y. Guo, "Adaptive online estimation of time-varying parameter nonlinear systems," in *32nd ChineseControl Conference (CCC 2013)*. IEEE, 2013, pp. 4570–4575.

[60] A. Meyer-Base, F. Ohl, and H. Scheich, "Quadratic-type lyapunov functions for competitive neural networks with different time-scales," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 6, Perth, Australia, Nov. 1995, pp. 3210–3214.

[61] ——, "Stability analysis techniques for competitive neural networks with different time-scales," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 6, Perth, Australia, Nov. 1995, pp. 3215–3219.

[62] ——, "Singular perturbation analysis of competitive neural networks with different time scales," *Neural Computation*, vol. 8, no. 8, pp. 1731–1742, 1996.

[63] A. Meyer-Base, "Flow invariance for competitive neural networks with different time-scales," vol. 1, 2002, pp. 858–861.

[64] X. Wang and Y. Huang, "Convergence study in extended Kalman filter-based training of recurrent neural networks," *IEEE Transactions on Neural Networks*, vol. 22, no. 4, pp. 588–600, Apr. 2011.

[65] C.-S. Leung and L.-W. Chan, "Dual extended Kalman filtering in recurrent neural networks," *Neural Networks*, vol. 16, no. 2, pp. 223–239, 2003.

[66] G. V. Puskorius and L. A. Feldkamp, "Neurocontrol of nonlinear dynamical systems with Kalman filter trained recurrent networks," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 279–297, Mar. 1994.

[67] R. G. Bland, D. Goldfarb, and M. J. Todd, "The ellipsoid method: A survey," *Operations Research*, vol. 29, no. 6, pp. 1039–1091, 1981.

[68] J. de Jesús Rubio and W. Yu, "Neural networks training with optimal bounded ellipsoid algorithm," in *Advances in Neural Networks–ISNN 2007*, ser. Lecture Notes in Computer Science, D. Liu, S. Fei, Z.-G. Hou, H. Zhang, and C. Sun, Eds. Berlin, Germany: Springer Berlin Heidelberg, 2007, vol. 4491, pp. 1173–1182.

[69] J. D. J. R. Avila, A. F. Ramírez, and C. Avilés-Cruz, "Nonlinear system identification with a feedforward neural network and an optimal bounded ellipsoid algorithm," *WSEAS Transactions on Computers*, vol. 7, no. 5, pp. 542–551, 2008.

[70] M. Moallem, K. Khorasani, and R. Patel, "An integral manifold approach for tip-position tracking of flexible multi-link manipulators," *IEEE Transactions on Robotics and Automation*, vol. 13, no. 6, pp. 823–837, Dec. 1997.

[71] A. Narang, "Analysis and control of non-affine, non-standard, singularly perturbed systems," Ph.D. dissertation, Texas A & M University, 2012.

[72] X. Chen, M. Heidarinejad, J. Liu, and P. Christofides, "Composite fast-slow MPC design for nonlinear singularly perturbed systems: Stability analysis," in *Proceedings of the American Control Conference*, Montreal, Canada, Jun. 2012, pp. 4136–4141.

[73] C. Sueur and G. Dauphin-Tanguy, "Bond graph approach to multi-time scale systems analysis," *Journal of the Franklin Institute*, vol. 328, no. 56, pp. 1005–1026, 1991.

[74] K.-J. Lin, "Composite observer-based feedback design for singularly perturbed systems via LMI approach," in *Proceedings of the SICE Annual Conference 2010*, Taipei, Taiwan, China, Aug. 2010, pp. 3056–3061.

[75] H.-L. Choi, Y.-S. Shin, and J.-T. Lim, "Control of nonlinear singularly perturbed systems using feedback linearisation," *IEE Proceedings - Control Theory and Applications*, vol. 152, no. 1, pp. 91–94(3), Jan. 2005.

[76] K.-H. Shim and M. E. Sawan, "Linear-quadratic regulator design for singularly perturbed systems by unified approach using δ operators," *International Journal of Systems Science*, vol. 32, no. 9, pp. 1119–1125, 2001.

[77] M. Innocenti, L. Greco, and L. Pollini, "Sliding mode control for two-time scale systems: Stability issues," *Automatica*, vol. 39, no. 2, pp. 273–280, 2003.

[78] H. Yu and B. Zhang, "Stabilizability of a class of singularly perturbed systems via switched output feedback," in *Proceedings of the 2013 Chinese Intelligent Automation Conference*, ser. Lecture Notes in Electrical Engineering, Z. Sun and Z. Deng, Eds. New York: Springer Berlin Heidelberg, 2013, vol. 254, pp. 729–734.

[79] J.-W. Son and J.-T. Lim, "Robust stability of nonlinear singularly perturbed system with uncertainties," *IEE Proceedings - Control Theory and Applications*, vol. 153, no. 1, pp. 104–110, Jan. 2006.

[80] P. D. Christofides, "Robust output feedback control of nonlinear singularly perturbed systems," *Automatica*, vol. 36, no. 1, pp. 45–52, 2000.

[81] Y. Gao, B. Sun, and G. Lu, "Passivity-based integral sliding-mode control of uncertain singularly perturbed systems," *IEEE Transactions on Circuits and Systems—Part II: Express Briefs*, vol. 58, no. 6, pp. 386–390, Jun. 2011.

[82] J. Wang, J. Wang, and H. Li, "Nonlinear PI control of a class of nonlinear singularly perturbed systems," *IEE Proceedings - Control Theory and Applications*, vol. 152, no. 5, pp. 560–566, Sep. 2005.

[83] I. Lizarraga, V. Etxebarria, and A. Sanz, "Sliding-mode adaptive control for flexible-link manipulators using a composite design," *Cybernetics and Systems*, vol. 36, no. 5, pp. 471–490, 2005.

[84] Y.-R. Hu and G. Vukovich, "Position and force control of flexible joint robots during constrained motion tasks," *Mechanism and Machine Theory*, vol. 36, no. 7, pp. 853–871, 2001.

[85] F. Ghorbel, "Adaptive control of flexible joint robot manipulators: A singular perturbation approach," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 1991.

[86] K. Khorasani, "Adaptive control of flexible-joint robots," *IEEE Transactions on Robotics and Automation*, vol. 8, no. 2, pp. 250–267, Apr. 1992.

[87] B. Esakki, "Modeling and robust control of two collaborative robot manipulators handling a flexibile object," Ph.D. dissertation, Concordia University, Montreal, Canada, Mar. 2011.

[88] Y. Zheng, H. A. Abdel Fattah, and K. A. Loparo, "Non-linear adaptive sliding mode observer–controller scheme for induction motors," *International Journal of Adaptive Control and Signal Processing*, vol. 14, no. 2-3, pp. 245–273, 2000.

[89] F. Lewis, D. Dawson, and C. Abdallah, *Robot Manipulator Control: Theory and Practice*, ser. Automation and Control Engineering. Boca Raton, Florida: CRC Press, 2003.

[90] A. K. Kostarigka and G. A. Rovithakis, "Adaptive dynamic output feedback neural network control of uncertain mimo nonlinear systems with prescribed performance," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 1, pp. 138–149, Jan. 2012.

[91] Y.-J. Liu, C. L. P. Chen, G.-X. Wen, and S. Tong, "Adaptive neural output feedback tracking control for a class of uncertain discrete-time nonlinear systems," *IEEE Transactions on Neural Networks*, vol. 22, no. 7, pp. 1162–1167, Jul. 2011.

[92] B. Ren, S. S. Ge, K. P. Tee, and T. H. Lee, "Adaptive neural control for output feedback nonlinear systems using a barrier lyapunov function," *IEEE Transactions on Neural Networks*, vol. 21, no. 8, pp. 1339–1345, Aug. 2010.

[93] K. J. Astrom and B. Wittenmark, *Adaptive Control*, 2nd ed. Boston, MA: Addison-Wesley Longman Publishing Co., Inc., 1994.

[94] A. Y. Alanis, E. N. Sanchez, A. G. Loukianov, and M. A. Perez-Cisneros, "Real-time discrete neural block control using sliding modes for electric induction motors," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 1, pp. 11–21, Jan. 2010.

[95] A. S. Poznyak, W. Yu, and E. N. Sanchez, "Identification and control of unknown chaotic systems via dynamic neural networks," *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, vol. 46, no. 12, pp. 1491–1495, Dec. 1999.

[96] S. S. Ge, F. Hong, and T. H. Lee, "Adaptive neural control of nonlinear time-delay systems with unknown virtual control coefficients," *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 34, no. 1, pp. 499–516, Feb. 2004.

[97] J. Na, X. Ren, G. Herrmann, and Z. Qiao, "Adaptive neural dynamic surface control for servo systems with unknown dead-zone," *Control Engineering Practice*, vol. 19, no. 11, pp. 1328–1343, 2011.

[98] A. Yeildirek and F. L. Lewis, "Feedback linearization using neural networks," *Automatica*, vol. 31, no. 11, pp. 1659–1664, 1995.

[99] K. S. Narendra and A. M. Annaswamy, "A new adaptive law for robust adaptation without persistent excitation," *IEEE Transactions on Automatic Control*, vol. 32, no. 2, pp. 134–145, Feb. 1987.

[100] F. W. Warner, *Foundations of Differentiable Manifolds and Lie Groups*, ser. Graduate Texts in Mathematics. Strtz GmbH, Germany: Springer-Verlag Berlin Heidelberg, 1983, vol. 94.

[101] S. M. LaValle, *Planning Algorithms*. New York, USA: Cambridge University Press, 2006.

[102] R. Suckley and V. N. Biktashev, "The asymptotic structure of the hodgkin-huxley equations," *International Journal of Bifurcation and Chaos*, vol. 13, no. 12, pp. 3805–3825, 2003.

[103] Y. Boutalis, D. Theodoridis, T. Kottas, and M. A. Christodoulou, *System Identification and Adaptive Control: Theory and Applications of the Neurofuzzy and Fuzzy Cognitive Network Models*, ser. Advances in industrial control. Basel, Switzerland: Springer, 2014.

[104] R. Rojas, *Neural Networks: A Systematic Introduction*. Berlin, Germany: Springer, 1996.

[105] W. J. Terrell, *Stability and Stabilization: An Introduction*. Princeton, New Jersey: Princeton University Press, 2009.

[106] S. Sarkka, "On unscented Kalman filtering for state estimation of continuous-time nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1631–1641, 2007.

[107] Y. Zhang and X. R. Li, "A fast u-d factorization-based learning algorithm with applications to nonlinear system modeling and identification," *IEEE Transactions on Neural Networks*, vol. 10, no. 4, pp. 930–938, Jul. 1999.

[108] L. Ros, A. Sabater, and F. Thomas, "An ellipsoidal calculus based on propagation and fusion," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 32, no. 4, pp. 430–442, 2002.

[109] S. Kavuri and V. Venkatasubramanian, "Representing bounded fault classes using neural networks with ellipsoidal activation functions," *Computers & chemical engineering*, vol. 17, no. 2, pp. 139–163, 1993.

[110] D. Naso, F. Cupertino, and B. Turchiano, "Precise position control of tubular linear motors with neural networks and composite learning," *Control Engineering Practice*, vol. 18, no. 5, pp. 515–522, 2010.

[111] G. A. Rovithakis and M. A. Christodoulou, *Adaptive Control With Recurrent High-Order Neural Networks: Theory and Industrial Applications*. London, UK: Springer-Verlag London, 2000.

[112] Z. Fu, L. Liu, and W.-F. Xie, "Indirect adaptive control of nonlinear system via dynamic multilayer neural networks with multi-time scales," in *Proceedings of the 4th International Conference on Intelligent Control and Information Processing*, Beijing, China, Jun. 2013, pp. 814–819.

[113] A. Tadayoni, W.-F. Xie, and B. W. Gordon, "Adaptive control of harmonic drive with parameter varying friction using structurally dynamic wavelet network," *International Journal of Control, Automation and Systems*, vol. 9, no. 1, pp. 50–59, 2011.

[114] V. Adetola and M. Guay, "Finite-time parameter estimation in adaptive control of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, pp. 807–811, Apr. 2008.

[115] J. Na, M. N. Mahyuddin, G. Herrmann, X. Ren, and P. Barber, "Robust adaptive finite-time parameter estimation and control for robotic systems," *International Journal of Robust and Nonlinear Control*, vol. 25, no. 16, pp. 3045–3071, 2015.

[116] J. Na, J. Yang, X. Ren, and Y. Guo, "Robust adaptive estimation of nonlinear system with time-varying parameters," *International Journal of Adaptive Control and Signal Processing*, vol. 29, no. 8, pp. 1055–1072, 2015.

[117] Y. Lv, J. Na, Q. Yang, X. Wu, and Y. Guo, "Online adaptive optimal control for continuous-time nonlinear systems with completely unknown dynamics," *International Journal of Control*, vol. 89, no. 1, pp. 99–112, 2016.

[118] J. Na, J. Yang, X. Wu, and Y. Guo, "Robust adaptive parameter estimation of sinusoidal signals," *Automatica*, vol. 53, pp. 376 – 384, 2015.

[119] R. Penrose, "On best approximate solutions of linear matrix equations," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 52, pp. 17–19, Jan. 1956.

[120] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[121] C. Chen, “Global exponential stabilisation for nonlinear singularly perturbed systems,” *IEE Proceedings - Control Theory and Applications*, vol. 145, no. 4, pp. 377–382, Jul. 1998.