Global Parameter Identification and Control of Nonlinearly Parameterized Systems

by

Aleksandar M. Kojić

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2002

Author ....
Department of Mechanical Engineering
September 21, 2001

Certified by ....
Anuradha M. Annaswamy
Principal Research Scientist
Thesis Supervisor

Accepted by ....
Ain A. Sonin
Chairman, Department Committee on Graduate Students

Global Parameter Identification and Control of Nonlinearly Parameterized Systems

by

Aleksandar M. Kojić

Submitted to the Department of Mechanical Engineering
on September 21, 2001, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

Nonlinearly parameterized (NLP) systems are ubiquitous in nature and many fields of science and engineering. Despite the wide and diverse range of applications, there exist relatively few results in control systems literature which exploit the structure of the nonlinear parameterization. A vast majority of presently applicable global control design approaches to systems with NLP, make use of either feedback-linearization, or assume linear parameterization, and ignore the specific structure of the nonlinear parameterization. While this type of approach may guarantee stability, it introduced three major drawbacks. First, they produce no additional information about the nonlinear parameters. Second, they may require large control authority and actuator bandwidth, which makes them unsuitable for some applications. Third, they may simply result in unacceptably poor performance. All of these inadequacies are amplified further when parametric uncertainties are present. What is necessary is a systematic adaptive approach to identification and control of such systems that explicitly accommodates the presence of nonlinear parameters that may not be known precisely.

This thesis presents results in both adaptive identification and control of NLP systems. An adaptive controller is presented for NLP systems with a triangular structure. The presence of the triangular structure together with nonlinear parameterization makes standard methods such as back-stepping, and variable structure control inapplicable. A concept of bounding functions is combined with min-max adaptation strategies and recursive error formulation to result in a globally stabilizing controller. A large class of nonlinear systems including cascaded LNL (linear-nonlinear-linear) systems are shown to be controllable using this approach.

In the context of parameter identification, results are derived for two classes of NLP systems. The first concerns systems with convex/concave parameterization, where min-max algorithms are essential for global stability. Stronger conditions of persistent excitation are shown to be necessary to overcome the presence of multiple equilibrium points which are introduced due to the stabilization aspects of the min-max algorithms. These conditions imply that the min-max estimator must periodically employ the local gradient information in order to guarantee parameter convergence.

The second class of NLP systems considered in this concerns monotonically parameterized systems, of which neural networks are a specific example. It is shown that a simple algorithm based on local gradient information suffices for parameter identification. Conditions on the external input under which the parameter estimates converge to the desired set starting from arbitrary values are derived. The proof makes direct use of the monotonicity in the parameters, which in turn allows local gradients to be self-similar and therefore introduces a desirable invariance property. By suitably exploiting this invariance property and defining a sequence of distance metrics, global convergence is proved. Such a proof of global convergence is in contrast to most other existing results in the area of nonlinear parameterization, in general, and neural networks in particular.

Thesis Supervisor: Anuradha M. Annaswamy
Title: Principal Research Scientist

Брату и родитељима

Никола, Гордана, Милош

Acknowledgments

In his endeavors, fortunately, no man is alone. Even in the difficult and trying times when it seems most convincingly that one is left alone, one can take comfort knowing that, in reality, those are the times when quite the opposite is true. A sense of that spirit can be observed in Sir Isaac Newton's quote: "If I've seen further, it is because I've stood on shoulders of giants". Although Newton was referring solely to his scientific achievements, such is the case for any accomplishment in life. And if we are to witness a tree bearing any fruit of accomplishment, it surely must have deep and strong roots, roots which sustain it and make the fruits possible. It is with great pleasure that I will here attempt to acknowledge my roots.

At the time of this writing, it has been exactly six years since I've started working with my advisor, *Dr. Anuradha Annaswamy*. I would like to thank her for her constant support throughout those years. I would also like to thank her for presenting me with the opportunity to work on a number of interesting and challenging analytical problems, and thus with the possibility of taking an interesting and challenging path. From today's perspective, looking back at the time just before starting at MIT, I see that it was a long journey in terms of the distance covered, and without her skillful guidance and patience along the way getting this far would have not been possible.

During the six years spent at MIT, I had immeasurable and constant support and help from the *Mikić* and *Mijailović* families. Their support, which has been so persistently manifested in numerous diverse ways, has been of vital significance. I hereby acknowledge the depth of my gratitude for such help, and pledge to, should the opportunity present itself, attempt to provide similar support to others.

I am very grateful to *Srdjan Divac* and *Zoran Spasojević* for their friendship and lending me some of their mathematical expertise. *Vladimir Bozin* has had the patience to listen to my mathematical ideas and sketches of proofs at several critical crossroads, and I thank him for the time and effort he has put in clarifying these and exposing some of their pitfalls. The members of two generations of MIT Active-Adaptive Control Laboratory have been instrumental in not only providing a pleasant and motivational work environment and aid in difficult research problems, but in easying the, at times, heavy burden of graduate student life: *Ashish Krupadanam*, *Jean-Pierre Hathout*, *Ssu-Hsin Yu*, *Fredrik Skantze*, *Rienk Bakker*, *Steingrimur Karason*, *Kathy Misovec*, *Chayakorn Thanomsat*, *Jennifer Rumsey*, *Mehmet Yunt*, *Chengyu Cao*, *Sungbae Park*, *Youssef Marzouk*.

I would also like to thank all of my friends at MOST (http://web.mit.edu/most/www) and the wider Boston area. During the past several years, we have been through a lot of joyous and sorrowful times. In the good times, having you to share the joy has given that joy a meaning, and in bad times, having you to lean on has made the sadness easier to bear. The same is true for my friends and former roommates, *Brandon Gordon*, *Petter Skantze*, *Georgios Kotsalis*, *Petros Komodromos*, and *Ivan Celanović*. Your contribution is noted and appreciated.

Thanks are also in order for *Kate Melvin*, and *Leslie Regan* and the staff of the MIT Mechanical Engineering Graduate Students Office for their help and assistance throughout the past six years.

Last, but by no means least, my deepest thanks are to my immediate and extended family, both here and at home. It goes without saying that their support and encouragement at every step has been the foundation of the journey taken thus far. I am more than happy to dedicate this work to them, since it is as much their product as it is mine.

Contents

1 Introduction **9**

1.1 Motivation 10

1.2 Thesis Goals 11

1.3 Organization of the Thesis 14

2 Parameter Convergence in systems with Convex/Concave Parameterization **15**

2.1 Introduction 15

2.2 Statement of the Problem 17

2.3 Continuous-time Parameter Convergence 19

2.3.1 Preliminaries 19

2.3.2 A Condition for Parameter Convergence 20

2.4 A case study of persistent excitation 27

2.5 Discrete-time Parameter Convergence 30

2.5.1 Parameter convergence in the presence of concave/convex nonlinear parameterization 31

2.5.2 An example of NLP-persistent excitation 38

2.6 Concluding Remarks and Future Work 40

3 Adaptive Control of Nonlinearly Parameterized Systems with a Triangular Structure **42**

3.1 Introduction 42

3.2 The Adaptive Controller for Systems in Chain Form 45

3.2.1 Preliminaries 45
3.2.2 The controller structure 50
3.2.3 A continuous controller 54
3.2.4 Tracking 57
3.2.5 Robustness 59
3.2.6 Control of L-N-L systems 62
3.2.7 n second-order systems in chain form 64
3.2.8 Numerical example 67
3.3 Adaptive Control of Systems in Triangular Form 70
3.4 Concluding Remarks and Future Work 74

4 Convergence conditions for parameter identification with the gradient algorithm in nonlinearly parameterized systems **79**

4.1 Introduction 79
4.2 Main results 80
4.3 Concluding Remarks and Future Work 104

5 Conclusions **107**

Chapter 1

Introduction

Since the earliest times, man has ben fascinated by his environment. To help explain the observed phenomena, abstract models were constructed. With the advancement of general science, especially mathematics, the models used grew increasingly complex. It is interesting to note that this relationship between the mathematical models and mathematics as a science has had a sizeable degree of feedback throughout history. Particular examples range from Newton's development of calculus and the derived laws of motion in *Principia* in 1687, Fourier theory used in his *Theory of Heat* in 1807, to the modern times when research in higher-dimensional physics theories drives advances in differential geometry and topology (for example, see [19]).

The growth in complexity in the models used is, in part, due to the fact that often satisfactory models based on existing data were used to gain further understanding and design new experiments. The possible discrepancies between the newly obtained data and a-priori predictions resulted in the revision of original models. When confronted with systems whose behavior is not fully understood, obtaining good purely mathematical models is a very important first step in gaining further insight. It has been said that

"A mathematical model is neither a hypothesis nor a theory. Unlike scientific hypotheses, a model is not verifiable directly by an experiment. For all models are both true and false.... The validation of a model is not that it is "true" but that it generates good testable hypotheses relevant to important problems. " ([33])

1.1 Motivation

From a control systems perspective, interest in modeling can be two-fold. First, by carefully modeling the system of interest, further insight can be gained that influences the later control design. For example, the type and placement of actuators and sensors can determine the type of mathematical model and the required controller. Secondly, it is crucial that the model used accurately reflect the modes of system behavior that are of particular concern. Out of the two aspects of the modeling issue in control design, in this thesis we do not concern ourselves with the former. We assume that the model for the system of interest is given. We also assume that the given model description may contain unknown, but constant, parameters.

The parameterization of the model can be either linear or nonlinear. This thesis is centered on examining identification and control issues related to nonlinearly parameterized (NLP) models. There are two main reasons for this. One is the fact that results for linearly parameterized models are well-established for a large variety of systems (for example, see [38, 45, 17, 31]. However, the results for NLP systems are scarce ([14, 41]). The other, and more important, reason for studying NLP systems is that in many branches of science nonlinear parameterization is required in order to accurately model systems of interest. Particular examples of NLP systems include:

- <u>mechanical systems</u>: friction dynamics [5], magnetic bearings [54],
- <u>electrical/electronic systems</u>: DC converters [25], CMOS modeling [16],
- <u>aerospace systems</u>: jet aircraft [25], spacecraft robotics [53],
- <u>chemical processes</u>: pH regulation [13], bioreactors [11, 12], fermentation process [10, 55],
- <u>biological systems</u>: see [30, 7],
- <u>"neural networks"</u>: these structures are inherently nonlinearly parameterized, see [21, 20, 44],

It can be seen that NLP models are applicable, and in fact required, in diverse and wide types of systems. Hence, further study is warrantied.

1.2 Thesis Goals

The scope of the thesis consists of two goals. The first is to develop control methodologies so that the output of the NLP systems behaves in a prescribed fashion, and the values of the unknown parameters are estimated in a stable manner. Once this is achieved, the second goal is identification of unknown parameters by investigating conditions under which the parameter estimates converge to their true values. The first goal is presented for NLP systems with a triangular structure. The second goal is presented in two types of NLP systems. One contains convex/concave parameterization, and parameter convergence conditions are derived for the min-max estimator of [2]. The other type of NLP system is monotonic in its parameterization. Here, parameter convergence conditions are derived for the standard gradient algorithm.

There exist a number of results in nonlinear system control. For example, see [52, 26, 49, 23, 40, 31]. The way these results can handle nonlinear parametric uncertainties is by preforming some type of feedback linearization (see [48, 35]). Feedback linearization requires a-priori knowledge on the bounds of the magnitudes of uncertainties. On-line, it does not acquire new information about the parametric uncertainties by observing system behavior and estimating the values of the unknown parameters. Since only a-priori information is used, and no on-line adaptation is carried out, this procedure may result in unnecessarily large control effort. Due to this, feedback linearization as a stand-alone technique may require large variations in control magnitude in short periods of time. Actuator bandwidth/energy issues, especially in extreme environments ([53]) then may constrain further the applicability of this approach.

There are relatively few results available for the adaptive control of NLP systems. These include the theoretical developments in [14, 41, 2, 34, 47, 12, 55]. A novel development was the introduction of the min-max estimator in [2, 34, 47]. The min-max estimator differs from traditional approaches in the sense that it does not use all the time the local gradient information for adjusting the parameter estimates. Instead, it uses both local and global gradient information, and switches between the two in a prescribed manner. The results in [12, 55, 39, 13] deal with control of biochemical processes. In [12, 55], an approach based on modifying the standard adaptive algorithms based on the plant model and choosing a suitable Lyapunov function. The result in [13] is a direct application of the min-max estimator, while [39] develops a methodology for convexifying the nonlinear parameterization so that the min-max estimator is applicable. Other theoretical results, [41, 14] are inadequate for tracking or general parameterization.

In this thesis, we propose an approach to solving the problem of adaptive control of NLP systems with a triangular structure. Our approach consists of coupling the feedback-linearization procedure described above with a suitable parameter estimation algorithm. We employ the min-max algorithm developed in [2]). A direct application of the min-max estimator is not possible, since the system in triangular form does not satisfy the so-called "matching conditions". By using the feedback-linearization procedure in a similar fashion to [48], we transform the system to a suitable form, and apply the min-max estimation technique. It is shown that coupling the two methods results in global stability of the overall system.

For the second goal of parameter identification, we propose two results. The results involve different types of parameterizations, and different estimation algorithms. Virtually no global results exist when dealing with the identification of NLP parameters for general types of nonlinearities. Since parameter identification is equivalent to solving a set of nonlinear equations, the lack of results is hardly surprising. Considering that

" *We make an extreme, but wholly defensible, statement: There are good, general methods for solving systems of more than one nonlinear equation. Furthermore, it is not hard to see why (very likely) there* never will *be any good, general methods....([43], section 9.6)*

we can, at best, hope to derive parameter convergence results on a case-by-case basis. In fact, the linear parameterization for which well-known results exists is a subset of the general NLP. The thesis presents two additional case studies in NLP identification. One deals with convex/concave parameterization, and employs the min-max algorithm of [2]. It should be noted that in this case results are obtained for the case when there are dynamics between the parametric uncertainty and the output ("error model 3" of [38]). The second deals with monotonic parameterization, employs the gradient-algorithm, and does not involve dynamics between the nonlinearity and the output ("error model 1") In both cases, we derive the necessary conditions on system signals and the nonlinearities involved to guarantee global convergence of parameter estimates.

The monotonic parameterization case can be of special interest to the neural network applications. First, many neural networks are realized using monotonic nonlinearities, usually of the sigmoid type. Second, a neural network must be trained for the task at hand, and neural network training is essentially a parameter identification process. By using a novel set of tools, we show that under relatively mild conditions on the input, the gradient method can be globally convergent for a single layer neural net .

Since neural network architectures have been applied to numerous diverse tasks, the problem of neural network training has gathered a lot of attention. For a survey of the topic, a good recent reference is [44]. To the best of our knowledge, the issues of neural network training have not been adequately addressed. Some of the existing approaches [18, 9, 8] impose strict conditions on network structure, which is limited in to applicability binary classifier tasks. Although it is well known that neural-network structures are controllable (see [50, 26]) when their parameters are fixed a-priori, some approaches (see [42, 27]) claim "semi-global" stability results when adjusting the parameters on-line, in a closed-loop setting. The stability result in these approaches, however, stems from the fact that they either overcome the nonlinearity in parameters by essentially employing linearization, or by just adjusting the linear parameters. In either case, the comments in [41] are applicable. Further, since linearization is used, it is questionable whether the use of the term "global" is justified.

By using novel stability analysis tools, we are able to show the conditions under which the gradient algorithm is globally convergent. Our results presented here pertain to the error model 1 case. Preliminary investigations suggest that by reinterpreting the standard quadratic Lyapunov arguments, it might be possible to construct new Lyapunov functions and demonstrate stability for the error model 3 case as well.

1.3 Organization of the Thesis

The thesis is organized in the following manner. We first present the convergence c conditions for the min-max estimator in Chapter 2. Convergence conditions are presented for both the continuous and discrete time versions of the min-max algorithm. A case study of parameter convergence conditions on an example system is carried out. In Chapter 3 we present the design of an adaptive controller for NLP systems in triangular form. It is shown how such a controller has the desired tracking and robustness properties. A numerical example is presented, as well. Comments on the design are given and future work is suggested. In Chapter 4 presents our discussions on the parameter convergence conditions for monotonically parameterized systems. Finally, concluding remarks are given in Chapter 5.

Chapter 2

Parameter Convergence in systems with Convex/Concave Parameterization

2.1 Introduction

Based on observation and physical laws, for many systems of interest the general form of the function which can adequately represent observed behavior is known. However, for a specific case, the known general function can depend on one or several constant parameters, whose exact values cannot be determined precisely. The question then arises how such classes of systems can be controlled to behave in a desired fashion, and whether in doing so, it is possible to gain an accurate estimate of the values of the underlying unknown parameters. The field of adaptive control and estimation has addressed these issues. Currently, many powerful techniques have been developed for the aforementioned problems (for example, see [17, 38, 45, 31]). In all of these results, the common feature is a fundamental assumption that the unknown parameters in the system occur linearly. Furthermore, this assumption is required to hold for both linear and nonlinear systems (see [31, 38, 49]).

The requirement for linear parameterization constrains the applicability of adaptive control, since many of the dynamical systems in nature exhibit such behavior which can only be accurately captured and represented by nonlinearly parameterized models. These nonlinear models can, perhaps, be converted to linearly parameterized ones by a suitable transformation. However, deriving such a transformation can be a nontrivial task, and may introduce further inaccuracies into the model. Hence, in order to accurately model complex systems, nonlinear parameterization seems unavoidable.

Recently, a stability framework built around the "min-max" algorithm has been established for studying identification and control of nonlinearly parameterized (NLP) systems [2, 4, 29, 34, 39, 47]. In all these papers various NLP systems were considered and the conditions for global stability, regulation and tracking were derived. Both continuous-time [2, 34] and discrete time [47] versions of the results have been developed. In this chapter we address the issue of parameter convergence in such systems. We derive conditions under which parameter estimates converge to their true values once a stable estimator has been established. These conditions are related to persistent excitation relevant for convergence in linearly parameterized systems, and are shown to be stronger, with the additional complexity being a function of the underlying nonlinearity. The derived results are presented for both the continuous-time and discrete-time versions of the min-max estimators.

This chapter is divided in two parts. The continuous time case is considered in Sections 2.2-2.4. Statement of the problem is given in Section 2.2. Next, convergence conditions are derived in 2.3. This section presents the sufficient conditions for accurate parameter estimation, similar to the persistent excitation conditions for linearly parameterized systems. A discussion of the obtained results on a case study is carried on in Section 2.4. In the second part of the chapter, Section 2.5 presents the analogous results for the discrete-time min-max estimator.

2.2 Statement of the Problem

Our goal is to study parameter convergence in a class of systems where the unknown parameter occurs nonlinearly. This class is of the form

$$\dot{y} = \alpha\, y + f(\phi(t), \theta) \tag{2.1}$$

where y is the measurable system output, $\theta \in \mathbb{R}^n$ is an unknown parameter, ϕ is a scalar function of time and f is a scalar function that is nonlinear both in ϕ and θ. In [2], a globally stable adaptive controller was developed for systems of the form of (2.1), and was shown to result in asymptotic stability and tracking. In this paper we derive an estimator for (2.1) and show that the parameter estimates converge to θ under appropriate conditions (of persistent excitation) on ϕ and f.

Further developments are based on the following assumptions about the above system:

(A1) $\theta \in \Theta$, where $\Theta = [\theta_{\min}, \theta_{\max}]$ is a known compact set defined by its lower ($\theta_{\min}$) and upper ($\theta_{\max}$) bounds.

(A2) $\phi(t)$ is a measurable and bounded function of time.

(A3) For any $\phi(t)$, only one of the following is true

 (i) f is concave for all $\theta \in \Theta$

 (ii) f is convex for all $\theta \in \Theta$

(A5) f is a known smooth and bounded function of its arguments.

(A6) The plant in (2.1) has globally bounded solutions for bounded ϕ.

The goal of the estimator is to closely track the output of the system and, in doing so, provide estimates of the value of the unknown parameter θ. To accomplish that task, the following estimator has been proposed:

$$
\begin{aligned}
\dot{\widehat{y}} &= -k\widetilde{y}_\epsilon + f(\phi,\widehat{\theta}) - a sat\left(\frac{\widetilde{y}}{\epsilon}\right) + \widehat{\alpha}\, y & (2.2)\\
\dot{\widehat{\alpha}} &= -\widetilde{y}_\epsilon y & (2.3)\\
\dot{\widehat{\theta}} &= -\widetilde{y}_\epsilon \omega & (2.4)\\
J(\omega,\theta) &= \left[f(\phi(t),\widehat{\theta}) - f(\phi(t),\theta) - \omega(\widehat{\theta}-\theta)\right] & (2.5)\\
a &= \min_{\omega\in\mathbb{R}} \max_{\theta\in\Theta} sign(\widetilde{y}) J(\omega,\theta) & (2.6)\\
\omega &= arg \min_{\omega\in\mathbb{R}} \max_{\theta\in\Theta} sign(\widetilde{y}) J(\omega,\theta) & (2.7)\\
\widetilde{y}_\epsilon &= \widetilde{y} - \epsilon sat\left(\frac{\widetilde{y}}{\epsilon}\right) & (2.8)
\end{aligned}
$$

where $\widetilde{y}$ is the tracking error defined as $\widetilde{y} = \widehat{y} - y$, ϵ is an arbitrary positive constant and $\widehat{\theta}$ represents an estimate of θ. The stability feature of this estimator is stated in the following theorem.

Theorem 2.1 *For the system in eq. (2.1), under the assumptions (A1)- (A4), the estimator given in eqs. (2.2)- (2.7), assures that the estimator outputs $\widehat{y}$, $\widehat{\theta}$, and $\widehat{\alpha}$ are bounded and that $|\widetilde{y}(t)| \to \epsilon$ as $t \to \infty$, provided that $\widehat{\theta}(t) \in \Theta$ for $t \geq t_0$.*

Proof Using the results of [2], it can be shown that if $V = \widetilde{y}_\epsilon^2 + (\widehat{\theta}-\theta)^T(\widehat{\theta}-\theta)$, then $\dot{V}$ along the system trajectories is given by

$$\dot{V} \leq -2\,k\,\widetilde{y}_\epsilon^2. \tag{2.9}$$

Hence, $V(t) \leq V(t_0)$ for all $t \geq t_0$. Therefore, Barbalat's Lemma can be used to show that $\dot{V} \to 0$ as $t \to \infty$, and hence $|\widetilde{y}(t)| \to \epsilon$. ●

2.3 Continuous-time Parameter Convergence

We now study the parameter convergence of the estimator in (2.2)– (2.7). For ease of exposition, we assume that α is known, and set $\widehat{\alpha} = \alpha$, and $\dot{\widehat{\alpha}} = 0$ in (2.3). In this case, the error dynamics of the estimator can be written as:

$$\begin{aligned} \dot{\widetilde{y}} &= -k\widetilde{y}_\epsilon - a sat\left(\frac{\widetilde{y}}{\epsilon}\right) + \widetilde{f} \\ \dot{\widetilde{\theta}} &= -\widetilde{y}_\epsilon\, \omega \end{aligned} \tag{2.10}$$

where $\widetilde{f} = f(\phi(t), \widehat{\theta}) - f(\phi(t), \theta)$, $\widetilde{\theta} = \widehat{\theta} - \theta$, and the quantities $\widetilde{y}_\epsilon$, a, ω were defined in eqs. (2.6)- (2.8). By inspecting the above system, we note that adaptation stops if $\widetilde{y}_\epsilon = 0$. In fact, it can be seen that all the system equilibrium points are contained within the region where $\widetilde{y}_\epsilon = 0$. The system dynamics in this region are:

$$\begin{aligned} \dot{\widetilde{y}} &= -a\,\frac{\widetilde{y}}{\epsilon} + \widetilde{f} \\ \dot{\widetilde{\theta}} &= 0 \end{aligned} \tag{2.11}$$

Since a can be a nonzero quantity, it follows that under certain conditions adaptation can stop even when $\widetilde{f}$ is large, implying unsatisfactory convergence of parameter estimates. Therefore, in order to enable parameter convergence, the probing signal ϕ must ensure that once the system enters the adaptation dead-zone with a large $\widetilde{f}$, it leaves the dead-zone in a finite amount of time.

2.3.1 Preliminaries

In facilitating the study of parameter convergence for the system in eq. (2.10), we introduce a new quantity $\beta(t)$ and establish a useful property of the system signals. The value of $\beta(t)$ quantifies the convexity/concavity of the function f and is defined as follows:

$$\beta(t) = \begin{cases} 1, & f(\phi(t), \theta) \ \text{ is convex} \\ -1, & f(\phi(t), \theta) \ \text{ is concave} \end{cases} \tag{2.12}$$

The following Lemma states a property of the min-max algorithm signals.

Lemma 2.1 *For $a(t)$ defined as in eq. (2.6), and $\beta(t)$ as in eq. (2.12), under the assumptions (A1)-(A5), the following holds:*

$$a(t)\,\beta(t)\,sign\,(\tilde{y}(t)) \;\leq\; 0 \tag{2.13}$$

Proof The solutions of the min-max problem in (2.6), as given in [2], are such that $a \neq 0$ only when $\beta sign(\tilde{y}) < 0$. It was shown in [29] that $a(t) \geq 0\ \forall t$. Combining these two elements establishes the Lemma. •

2.3.2 A Condition for Parameter Convergence

In what follows let:

$$\tilde{f}(x,y) = f\left(\phi(x), \hat{\theta}(y)\right) - f\left(\phi(x), \theta\right), \quad \text{and} \quad \tilde{f}(x) = \tilde{f}(x,x) \tag{2.14}$$

Let x denote the state vector, $x \triangleq [\tilde{y}, \tilde{\theta}]^T$.

The sufficient conditions for parameter convergence are stated in the following theorem.

Theorem 2.2 *The origin of the state space given by $[\tilde{y}_\epsilon, \tilde{\theta}]^T$ is uniformly asymptotically stable if, for every $t_1 > t_0$, there exist positive constants c, δ, T and a subinterval $[t_2, t_2+\delta] \in [t_1, t_1+T]$ such that*

$$\beta_2 \int_{t_2}^{t_2+\delta} \tilde{f}(\tau, t_2)d\tau \;\geq 2\epsilon + c\|\tilde{\theta}(t_2)\| \tag{2.15}$$

with

$$\beta(t) = \beta_2 = const\ \forall t \in [t_2, t_2+\delta] \tag{2.16}$$

If we compare the differential equations that arise in the nonlinearly-parameterized adaptive system in eq. (2.10) with those of the linearly-parameterized systems studied extensively in [37], we see that their structures exhibit a large degree of similarity. This degree of similarity is reinforced by examining and comparing the conditions for uniform asymptotic stability of state-space origin for the two types of systems. Taking into account the specific properties of the min-max algorithm , it can be seen that the condition stated in eqs. (2.15)- (2.16) is more restrictive than the persistent excitation conditions given in [37] for linearly parameterized systems. However, we shall use the same general path for establishing Theorem 2.2 as the one employed in [37]. In particular, three Lemmas, as in [37], are used in the proof of Theorem 2.2. All of the Lemmas stated below assume that the conditions (2.15) and (2.16) of Theorem 2.2 hold. The Lemmas are followed by their respective proofs.

Lemma 2.2 *Let $\eta > 0$ and $\epsilon_1 > 0$ be given. Then there exist positive numbers ϵ_2 and T_0 such that if $\|x(t_1)\| \leq \epsilon_1$ and if $\|\tilde{\theta}(t)\| \geq \eta$, for $t \in [t_1, t_1 + T_0]$, then there is a $t_3 \in [t_1, t_1 + T_0]$ such that $|\tilde{y}_\epsilon(t_3)| \geq \epsilon_2$.*

Proof of Lemma 2.2:

The solution of the differential equation

$$\dot{\tilde{y}} = -k\tilde{y}_\epsilon - a sat\left(\frac{\tilde{y}}{\epsilon}\right) + \tilde{f} \tag{2.17}$$

for all $t \geq t_1$ is given by

$$\tilde{y}(t+\delta) = \tilde{y}(t) + \int_t^{t+\delta} -k\tilde{y}_\epsilon(\tau) - a(\tau) sat\left(\frac{\tilde{y}(\tau)}{\epsilon}\right) + \tilde{f}(\tau)\, d\tau \tag{2.18}$$

Multiplying both sides by $\beta(t_2) = \beta_2$, for some $t_2 \in [t_1, t_1 + T + \delta]$ and separating the terms we have

$$\beta_2\tilde{y}(t+\delta) = \beta_2\tilde{y}(t) - k\beta_2 \int_t^{t+\delta} \tilde{y}_\epsilon(\tau) d\tau - \beta_2 \int_t^{t+\delta} a(\tau) sat\left(\frac{\tilde{y}(\tau)}{\epsilon}\right) d\tau + \beta_2 \int_t^{t+\delta} \tilde{f}(\tau)\, d\tau \tag{2.19}$$

We now assume that there is an $\epsilon_2 < \epsilon_1$ such that $|\tilde{y}_\epsilon(t)| < \epsilon_2$ for all $t \in [t_1, t_1 +$

$T+\delta]$. This assumption implies that $\widetilde{y}(t) > -\epsilon - \epsilon_2$ and that $|\beta_2 \widetilde{y}_\epsilon(t)| < \epsilon_2$. Hence,

$$-k\beta_2 \int_{t_2}^{t_2+\delta} \widetilde{y}_\epsilon(\tau)\, d\tau \;\; \geq \;\; -k\epsilon_2\delta \tag{2.20}$$

For a t_2 for which the Theorem hypothesis hold, and from Lemma 2.1, it follows that

$$\begin{aligned} -\beta_2 \int_t^{t+\delta} a(\tau) sat\left(\frac{\widetilde{y}(\tau)}{\epsilon}\right) d\tau \;\; &= \;\; - \int_t^{t+\delta} a(\tau)\ \beta(\tau)\ sat\left(\frac{\widetilde{y}(\tau)}{\epsilon}\right) d\tau \\ &\geq \;\; 0 \end{aligned} \tag{2.21}$$

The third integral on the right-hand side of eq. (2.19) can be expressed as

$$\beta_2 \int_t^{t+\delta} \widetilde{f}(\tau)\, d\tau = \beta_2 \int_t^{t+\delta} \widetilde{f}(\tau, t_2)\, d\tau + \beta_2 \int_t^{t+\delta} \left[\widetilde{f}(\tau) - \widetilde{f}(\tau, t_2)\right] d\tau \tag{2.22}$$

Using the condition in (2.15) which states that $\beta_2 \displaystyle\int_{t_2}^{t_2+\delta} \widetilde{f}(\tau, t_2) d\tau \;\geq 2\epsilon + c\|\widetilde{\theta}(t_2)\|$, the second integral on the right-hand side of (2.22) can be expressed as:

$$\begin{aligned} \beta_2 \int_t^{t+\delta} \widetilde{f}(\tau) - \widetilde{f}(\tau, t_2)\, d\tau \;\; &= \;\; \beta_2 \int_{t_2}^{t_2+\delta} \widehat{f}(\tau) - f(\tau) - \left(\widehat{f}(\tau, t_2) - f(\tau)\right) d\tau \\ &= \;\; \beta_2 \int_{t_2}^{t_2+\delta} \widehat{f}(\tau) - \widehat{f}(\tau, t_2) d\tau \\ &\geq \;\; -\delta M_\theta \left(\sup_{\tau \in [t_2, t_2+\delta]} \|\widehat{\theta}(t_2) - \widehat{\theta}(\tau)\| \right), \end{aligned}$$

where $M_\theta = \displaystyle\max_{\widehat{\theta} \in \Theta, \tau \in [t_2, t_2+\delta]} \nabla_{\widehat{\theta}} f(\phi(\tau), \widehat{\theta}(\tau))$. From eq. (2.4), the properties of the min-max algorithm, and the assumed bound on $\widetilde{y}_\epsilon$, it follows that

$$\begin{aligned} \beta_2 \int_t^{t+\delta} \left[\widetilde{f}(\tau) - \widetilde{f}(\tau, t_2)\right] d\tau \;\; &\geq \;\; -\delta M_\theta \left(\sup_{\tau \in [t_2, t_2+\delta]} \|\widehat{\theta}(t_2) - \widehat{\theta}(\tau)\| \right) \\ &\geq \;\; -\delta^2 M_\theta^2 \epsilon_2. \end{aligned} \tag{2.23}$$

Substituting (2.23) into eq. (2.19), we obtain:

$$\begin{aligned} \beta_2 \widetilde{y}(t_2 + \delta) \;\; &\geq \;\; -\epsilon - \epsilon_2 - k\epsilon_2\delta + 2\epsilon + c\eta - \delta^2 M_\theta^2 \epsilon_2 \\ &\geq \;\; \epsilon + \left[c\eta - \epsilon_2 \left(1 + k\delta + \delta^2 M_\theta^2\right)\right] \end{aligned}$$

For all ϵ_2 such that $0 \geq \epsilon_2 \geq \dfrac{c\eta}{2 + k\delta + \delta^2 + M_\theta^2}$ we have that

$$\beta_2 \widetilde{y}(t_2 + \delta) \geq \epsilon + \epsilon_2 \tag{2.24}$$

implying that $|\widetilde{y}_\epsilon(t_2 + \delta)| \geq \epsilon_2$. This establishes the Lemma. •

Lemma 2.3 *Let $\epsilon_1 > \epsilon_2 > 0$. Then there is an $n = n(\epsilon_1, \epsilon_2)$, such that if $\|x(t_1)\| \leq \epsilon_1$, and $S = \{t \in [t_1, \infty) | |\widetilde{y}_\epsilon(t)| > \epsilon_2\}$, then $\mu(S) \leq n$, where μ denotes Lebesgue measure.*

Proof of Lemma 2.3:

The proof directly follows by setting $n(\epsilon_1, \epsilon_2) = \epsilon_1^2 / 2k\widetilde{y}_\epsilon^2$ and using eq. (2.9). •

Lemma 2.4 *Let ϵ_1 and η be given positive numbers. Then there is a $T_0 = T_0(\epsilon_1, \eta)$ such that if $\|x(t_1)\| \leq \epsilon_1$, then there exists some $t_3 \in [t_1 + T_0]$ such that $\|\widetilde{\theta}(t_3)\| \leq \eta$.*

Proof of Lemma 2.4:

The proof of this Lemma is obtained by combining Lemmas 2.3 and 2.2. For a given ϵ_1 and η, assume that there exists a T_0 such that $\|\widetilde{\theta}(t)\| \geq \eta$ for all $t \in [t_1, t_1 + T_0]$. From Lemma 2.2, this implies that there exists a $\epsilon_2 > 0$ and a $t_3 \in [t_1, t_1 + T_0]$ such that $|\widetilde{y}_\epsilon(t_3)| > \epsilon_2$. For this value of ϵ_2, use Lemma 2.3 to obtain $n(\epsilon_1, \epsilon_2$. Set $t_1' = t_1 + n$, and set $T_0 = n + T$, where T is obtained from the theorem hypothesis for $t_1 = t_1'$. Now, we examine the value of $\|\widetilde{\theta}(t)\|$ over the interval $[t_1', t_1 + T_0]$. Over this interval, by the choice of T_0, eq. (2.15) holds. We have assumed that $\|\widetilde{\theta}(t)\| \geq \eta$ for all $t \in [t_1', t_1 + T_0]$. Invoking Lemma 2.2 once again, we obtain that there must be a $t_3' \in [t_1', t_1 + T_0]$ such that $|\widetilde{y}_\epsilon(t_3')| \geq \epsilon_2$. But, Lemma 2.3 states that $|\widetilde{y}_\epsilon(t)| < \epsilon_2$ for all $t \geq t_1'$. Hence, our assumption is contradicted, and the lemma established.

Proof of Theorem 2.2: Having established the Lemmas, the proof of theorem goes along similar lines to the proof of its counterpart in [37]. Assume that for given $\epsilon_1 > \epsilon_2 > 0$, $M > 0$:

$$\epsilon_2 \leq V(t) = \widetilde{y}_\epsilon^2 + \widetilde{\theta}^T \widetilde{\theta} \leq \epsilon_1 \qquad \text{for } t \in [t_1, t_1 + M]$$

From Theorem 2.1 we have that $V(t)$ is nonincreasing. Therefore, to demonstrate uniform asymptotic stability, it suffices to show that there exist such an M, $0 < \gamma < 1$, and a time instant $t_3 \in [t_1, t_1 + M]$ such that $V(t_3) \leq \gamma\, V(t_1)$. From Lemma 2.4 we have that for every $\epsilon_1 > 0$, $\eta > 0$, and every t_1, there is a T_0 and a $t_2 \in [t_1, t_1 + T_0]$ such that $\|\widetilde{\theta}(t_2)\| \leq \eta$. Let $\eta = \sqrt{\epsilon_2 c_1}$, for some positive number c_1 such that $1 - c_1 > 0$. Then

$$|\widetilde{y}_\epsilon(t_2)|^2 \geq V(t_2) - \eta^2 \geq V(t_2)(1 - c_1).$$

From eq. (2.24) we have that $\beta_2 \widetilde{y}(t_2) \geq \epsilon + \epsilon_2 \geq 0$, implying that $a(t_2) = 0$. Using this property and integrating the first equation in (2.10) over we have that

$$\beta_2 \widetilde{y}_\epsilon(t) = \beta_2 \widetilde{y}_\epsilon(t_2) + \beta_2 \int_{t_2}^{t} \left| -k\widetilde{y}_\epsilon(\tau) + \widetilde{f}(\tau) \right| d\tau \ \ \geq \ \ \beta_2 \widetilde{y}_\epsilon(t_2) - k\|x(t_2)\| - \int_{t_2}^{t} \left| \widetilde{f}(\tau) \right| d\tau$$

where $\beta_2 = \beta(t_2)$. Since f is a bounded function, over an interval $[t_2, t_2 + c_2]$, where c_2 is a positive number, there exists a $B_\theta > 0$ such that $|\widetilde{f}(\tau)| \leq B_\theta \|\widetilde{\theta}(\tau)\|$ for all $\tau \in [t_2, t_2 + c_2]$. Set

$$c_2 = \frac{1}{2} \frac{\sqrt{1 - c_1}}{k + B_\theta}. \tag{2.25}$$

Then, for all $t \in [t_2, t_2 + c_2]$,

$$\begin{aligned} \beta_2 \widetilde{y}_\epsilon(t) &\geq \beta_2 \widetilde{y}_\epsilon(t_2) - c_2(k + B_\theta)\|x(t_2)\| \\ &\geq \sqrt{1 - c_1}\sqrt{V(t_2)} - c_2(k + B_\theta)\|x(t_2)\| \\ &\geq \sqrt{V(t_2)}\left(\sqrt{1 - c_1} - c_2(k + B_\theta)\right) \end{aligned} \tag{2.26}$$

Let $t_3 = t_2 + c_2$. Hence,

$$\begin{aligned} V(t_2) - V(t_3) &= \int_{t_2}^{t_3} -V(\tau) d\tau \\ &\geq 2k \int_{t_2}^{t_3} |\widetilde{y}_\epsilon(\tau)|^2 d\tau \\ &\geq 2kc_2 V(t_2)\left(\sqrt{1 - c_1} - c_2(k + B_\theta)\right)^2. \end{aligned}$$

Let $\gamma = 2kc_2\left(\sqrt{1 - c_1} - c_2(k + B_\theta)\right)^2$. Clearly, γ is positive. But, we want $\gamma < 1$

as well. Thus, by solving this inequality in terms of c_1, we obtain that c_1 must be chosen such that the following relation is satisfied

$$c_1 > 1 - \left(\sqrt{\frac{1}{2kc_2}} + (k + B_\theta)c_2\right)^2 \tag{2.27}$$

However, not all values of c_1 that satisfy (2.27) fall in the allowed interval $(0, 1)$. The right hand side of (2.27) is certainly strictly less than one, but it also might be less than zero as well. To exclude this possibility, we choose c_1 as:

$$c_1 \in \left(\max\left(0, 1 - \left(\sqrt{\frac{1}{2kc_2}} + (k + B_\theta)c_2\right)^2\right),\ 1\right),$$

This guarantees that $c_1 \in (0, 1)$, and $\gamma \in (0, 1)$. Using eq. (2.25), it can be seen that the above relation is an implicit inequality in c_1, since c_2 depends on c_1. However, for given system characteristics k and B_θ, this inequality has an infinite number of solutions. Therefore, $V(t_3) \leq \gamma V(t_2) \leq \gamma V(t_1)$, establishes the theorem. •

We shall call the requirements stated in eq. (2.15) as nonlinearly-parameterized persistent excitation (NLP-PE) conditions, in contrast to its linear counterpart(LP-PE) as defined in [37]. When comparing NLP-PE and LP-PE two main differences can be observed. Both NLP-PE and LP-PE impose certain conditions on the values of the integrals of certain system signals. These conditions are required so that the only possible equilibrium points for the system are those where the parameter error is zero. The differences between the two cases concern the different requirements on the sign and magnitude of certain system integrals.

The linear condition, in terms of $\tilde{f}$, can be stated as

$$\left|\int_{t_2}^{t_2+\delta} \tilde{f}(\tau)d\tau\right| > c\|\tilde{\theta}(t_2)\| \tag{2.28}$$

Comparing (2.28) and (2.15), it can be seen that, in the case f is linear in parameters, NLP-PE directly implies LP-PE. The converse claim is not true.

From eq. (2.28) it follows that the sign of the integral in the linear case is irrelevant.

However, in the case of the NLP-PE, the sign of the integral is crucial. The sign does not solely depend on the sign of $\tilde{f}$, but on the convexity/concavity of f as well. This coupling is introduced through the the value of β which can take on either positive or negative value. The coupling arises from the features of the min-max adaptive algorithm and is required in order to enable the algorithm to escape the adaptation dead-zone specified by the region of the state-space where $|\tilde{y}_\epsilon| \leq \epsilon$. Also, the coupling between the sign of the integral and β is necessary, but not sufficient for the algorithm to leave the dead-zone. Therefore, a second difference between NLP-PE and LP-PE is introduced. Since the dead-zone has a finite size, the persistent excitation integral must have a large enough value to overcome the dead-zone. Because the dead-zone is finite, this value must be finite and bounded away from zero. That is the reason why there is a term containing ϵ, the size of the dead-zone, on the right hand side of (2.15). Even though the dead-zone exists in the linear case, it has a measure of zero, and therefore, bounding away the integral from zero is not necessary as in NLP-PE case.

Simply relying on the norm of the parameter error, $||\tilde{\theta}||$, for the magnitude of the integral may not be sufficient in the cases when $\tilde{\theta}$ becomes small. Therefore, a finite lower bound was introduced. However, some classes of nonlinearly parameterized functions, may behave in a linear fashion when $||\tilde{\theta}||$ becomes small. To satisfy NLP-PE in this case would mean that δ, the length of persistent-excitation interval would have to increase as $||\tilde{\theta}||$ becomes smaller. This may not be realizable in practical terms. What this implies is that parameter convergence for these types of functions can be guaranteed only to within a certain precision. This precision is of the order of ϵ, which is a feature of the min-max adaptive algorithm that is user specified depending. In practical terms, the size of ϵ depends on the available actuator bandwidth for a particular system. The higher the bandwidth, a more precise estimation of the parameters can be guaranteed.

2.4 A case study of persistent excitation

The NLP-PE condition in Theorem 2.2 specifies certain requirements on f in order to achieve parameter convergence. Since f depends on the time-varying signal ϕ, the persistent excitation conditions ultimately are translated to requirements on ϕ. For a given f, theorem 2.2 does not state how ϕ should behave in order to satisfy the requirements, or even whether such a ϕ is possible. In this section, we first state some observations about how ϕ should behave in order to satisfy NLP-PE for a generic function f that satisfies assumptions (A1)-(A6). Next, we apply these observations to a specific case of f, and give an example of ϕ that enables parameter convergence.

When examining NLP-PE as given in eq. (2.15), it can be considered that it consists of two separate components. By taking the absolute value on both sides of (2.15), we have that $\left|\int_{t_2}^{t_2+\delta} \tilde{f}(\tau, t_2)\, d\tau\right| \geq 2\epsilon + c||\tilde{\theta}(t_2)||$. This we will view as the first component of NLP-PE. Also, from (2.15), it follows that $\beta_2 \int_{t_2}^{t_2+\delta} \tilde{f}(\tau, t_2)\, d\tau > 0$. This we will call the second component of NLP-PE.

The first component of NLP-PE states that for a large parameter error, there must be a large error in f. It is straightforward to demonstrate that this condition is equivalent to LP-PE. This is essentially an identifiability condition, since it is not possible to estimate the parameter values exactly if there does not exist a ϕ such that, for a large parameter error, it produces a noticeable error in the system output. That is, parameter values for which all possible values of ϕ produce an equivalent output are, for all practical purposes, equivalent and indistinguishable from each other.

The second component of NLP-PE states what the sign of $\tilde{f}$ should be in relation to the convexity/concavity of f. In case that f is convex, $\tilde{f}$ should be positive, and conversely, in case f is concave, $\tilde{f}$ should be negative. This implies that the system should periodically move to the region of the phase-plane where the gradient information is used for updating the parameter estimates. Hence, the min-max feature of the algorithm is necessary to ensure stability, but is not sufficient to guarantee parameter convergence. Parameter convergence is ensured by repeatedly turning on the gradient component of the min-max algorithm.

The coupling of convexity/concavity and the sign of the integral of $\widetilde{f}$ has the following practical implications. Suppose that ϕ is such that f is always identifiable, and that for a certain value of $\widetilde{\theta}$, the integral in (2.15) is negative. To ensure parameter convergence, ϕ must be such that one of the following occurs:

(a) For the given $\widetilde{\theta}$, ϕ must change in such a way that the sign of $\widetilde{f}$ is reversed, while keeping the convexity/concavity of f the same, *or*

(b) For the given $\widetilde{\theta}$, ϕ must reverse the convexity/concavity of f, while preserving the sign of $\widetilde{f}$

For ϕ to be persistently exciting, it must be able to achieve either *(a)* or *(b)* for any combination of $\widehat{\theta} \in \Theta$ and $\theta \in \Theta$.

We illustrate the above comments with a discussion of persistent excitation for the following specific example of f:

$$f = e^{-\phi^T \theta} \tag{2.29}$$

where $\phi : \mathbb{R} \rightarrow \mathbb{R}^n$, $\theta \in \Theta \subset \mathbb{R}^n$. It can be checked that f given in (2.29) is always convex with respect to θ for all ϕ. Therefore, option *(b)* is not possible. Hence, ϕ must be such that $\widetilde{f}$ can switch sign for any $\widetilde{\theta}$ as required by option *(a)*. The following definition states the desired property of the probing signal ϕ.

Definition 2.1 *Let $w \in \mathbb{R}^n$ be any unit vector. A bounded function $\phi : \mathbb{R} \rightarrow \mathbb{R}^n$ is said to belong to class K^n if for any t_1 there exist positive constants ϵ_ϕ, δ and T, and a subinterval $[t_2, t_2 + \delta] \in [t_1, t_1 + T]$ such that*

$$\phi^T(\tau)\, w \;\geq \epsilon_\phi \qquad \forall \tau \in [t_2, t_2 + \delta] \tag{2.30}$$

Definition 2.1 states that, periodically, the vector ϕ should have a *positive* component along every w in $\mathbb{R}^n$. This is more restrictive than the linear case requirements in [37], since the latter requires ϕ to merely have a nonzero component periodically along every vector in $\mathbb{R}^n$. An example of a function that belongs to class K^2 is

$$\phi = [sin\omega t,\ cos\omega t]^T, \qquad \omega > 0 \tag{2.31}$$

Since such a ϕ represents a rotating vector in $\mathbb{R}^2$ with a constant velocity, it follows that it aligns itself with every w in $\mathbb{R}^2$ periodically.

Lemma 2.5 *Let $h(\epsilon)$ be a constant on the order of ϵ. For f defined as in eq. (2.29), and for $\widetilde{\theta} > h(\epsilon)$, $\phi \in K^n$ implies that ϕ is NLP-persistently exciting.*

Proof of Lemma 2.5 From eq. (2.29) it can be seen that f is convex in θ for all ϕ, i.e. $\beta(t) = 1\ \forall t$. This reduces eq. (2.15) to the form

$$\int_{t_2}^{t_2+\delta} \widetilde{f}(\tau, t_2)\, d\tau \geq 2\epsilon + c\|\widetilde{\theta}(t_2)\| \tag{2.32}$$

The integrand can be expressed as:

$$\begin{aligned}\widetilde{f}(\tau, t_2) &= f(\phi(\tau), \widehat{\theta}(t_2)), - f(\phi(\tau), \theta) = e^{-\phi(\tau)^T\widehat{\theta}(t_2)} - e^{-\phi(\tau)^T\theta} \\ &= e^{-\phi(\tau)^T\widehat{\theta}(t_2)}\left[1 - e^{-\phi(\tau)^T(\theta - \widehat{\theta}(t_2))}\right]\end{aligned}$$

Let $w_2 = \dfrac{\theta - \widehat{\theta}(t_2)}{\|\theta - \widehat{\theta}(t_2)\|}$. From Definition 2.1, it follows that there exists an ϵ_ϕ and a time interval $[t_2, t_2 + \delta]$ such that $\phi(\tau)^T w_2 \geq \epsilon_\phi$ for all $\tau \in [t_2, t_2 + \delta]$. Thus, $e^{-\phi(\tau)^T(\theta - \widehat{\theta}(t_2))} < 1$ over this interval, and, hence, $\widetilde{f}(\tau, t_2)$ can be expressed as

$$\begin{aligned}\widetilde{f}(\tau, t_2) &\geq Me^{-\epsilon_\phi r}\epsilon_\phi\|\theta - \widehat{\theta}(t_2)\| \\ &\geq \epsilon_f\|\widetilde{\theta}(t_2)\|\end{aligned} \tag{2.33}$$

where $r = \max\limits_{\widehat{\theta}, \theta \in \Theta} \|\widehat{\theta} - \theta\|$, $M = e^{\max\limits_{\theta \in \Theta}\|\widehat{\theta}\| \sup\limits_{t} \|\phi(t)\|}$. Therefore,

$$\int_{t_2}^{t_2+\delta} \widetilde{f}(\tau, t_2)\, d\tau \geq \delta\epsilon_f\|\widetilde{\theta}(t_2)\|. \tag{2.34}$$

and as long as $\|\tilde{\theta}(t)\| > \dfrac{2\epsilon}{\delta\epsilon_f}$, there exists a $c > 0$ such that

$$\int_{t_2}^{t_2+\delta} \tilde{f}(\tau, t_2)\, d\tau \geq 2\epsilon + c\|\tilde{\theta}(t_2)\|.$$

•

Lemma 2.5 states that, using an appropriate $\phi(t)$, it is possible to estimate the parameters up to a desired precision on the order of ϵ. However, the magnitude of precision can be modified and reduced by a proper choice of ϕ. One way to reduce the uncertainty in the parameter estimates is to increase δ, the interval of persistent excitation. For a particular choice of ϕ as in (2.31), this would mean choosing a low value of ω, corresponding to slowing the rotational velocity of the vector in phase-plane. This, in turn, might suggest that there is a possible tradeoff in the convergence rate and the guaranteed precision of parameter estimates.

2.5 Discrete-time Parameter Convergence

In this section, we present the parameter convergence results for the discrete-time min-max estimator developed in [47]. The results presented here are analogous to the ones presented in Section 2.3.

In [47], it was shown that for discrete-time nonlinearly parameterized (NLP) systems of the form

$$y_t = f(\phi_{t-1},\ \theta) + \varphi_{t-1}^T \alpha, \tag{2.35}$$

an adaptive estimator with a min-max algorithm leads to global stability. Specifically, for a Lyapunov function candidate of the form

$$V_t = \tilde{\theta}_t^T \Gamma_\theta^{-1} \tilde{\theta}_t + \tilde{\alpha}_t^T \Gamma_\alpha^{-1} \tilde{\alpha}_t \tag{2.36}$$

the min-max estimator ensures that $\Delta V_t = V_t - V_{t-1} \leq 0$. Thus, the desired stability property is ensured. However, since ΔV_t is not strictly less than zero, parameter convergence is not assured.

In this section, we investigate this problem of parameter convergence for the min-max adaptive algorithm. Since it is a requirement for the min-max estimator, we assume that the bounds on possible parameter values are known a priori. Although the min-max estimator in [47] can be applied for a general nonlinearity, we restrict our discussions to the systems where the nonlinear parameterization is concave or convex. We also assume that only the nonlinear parameters are present in the system, implying $\widetilde{\alpha}_t = 0$. In Section 2.5.1, we present sufficient conditions on the input ϕ and the nonlinearity f under which parameter convergence results. In section 2.5.2, a specific example of f and ϕ that satisfy these conditions is presented.

2.5.1 Parameter convergence in the presence of concave/convex nonlinear parameterization

The discrete min-max estimator presented in [47] results in an adaptive system of the form:

$$
\begin{aligned}
y_t &= f(\phi_{t-1}, \theta) \\
\widehat{y}_t &= f(\phi_{t-1}, \widehat{\theta}_{t-1}) \\
\widehat{\theta}_t &= \widehat{\theta}_{t-1} - \Gamma_\theta k_t \widetilde{y}_t \rho_t \omega_t \qquad \Gamma_\theta^T = \Gamma_\theta > 0 \\
k_t &= \frac{1}{\lambda + \omega_t^T \Gamma_\theta \omega_t} \qquad \lambda > 0 \\
\rho_t &= \max\{0, a_t\} \\
a_t &= 2 - \frac{2}{|\widetilde{y}_t|} J_0 \\
\omega_t &= arg \min_{\omega \in \mathrm{IR}^n} \max_{\theta \in \Theta} \operatorname{sgn}(\widetilde{y}_t) J(\omega, \theta) \\
J(\omega, \theta) &= \widetilde{f}_{t-1} - \omega^T(\widehat{\theta}_{t-1} - \theta). \\
J_0 &= \min_{\omega \in \mathrm{IR}^n} \max_{\theta \in \Theta} \operatorname{sgn}(\widetilde{y}_t) J(\omega, \theta)
\end{aligned}
\tag{2.37}
$$

where $\phi : N \to \mathrm{IR}^n$. For any ϕ and all $\theta \in \Theta \subset \mathrm{IR}^n$, where Θ is a compact set in IR^n, f is assumed to be either concave or convex. In this case, as derived in [2], the resulting min-max solution for J_0 is of the form

$$J_0 = \begin{cases} \beta\left[\widehat{f}_{t-1} - f_{min} - \dfrac{f_{max}-f_{min}}{\theta_{\max}-\theta_{\min}}(\widehat{\theta}_{t-1}-\theta_{\min})\right] & \text{if } \beta f \text{ is concave} \\ 0 & \text{if } \beta f \text{ is convex} \end{cases} \tag{2.38}$$

$$\omega_0 = \begin{cases} \dfrac{f_{max}-f_{min}}{\theta_{\max}-\theta_{\min}} & \text{if } \beta f \text{ is concave} \\ \nabla f_{\widehat{\theta}_{t-1}} & \text{if } \beta f \text{ is convex} \end{cases}$$

where $\beta = sgn(\widetilde{y}_t)$, $f_{max} = f(\phi_{t-1}, \theta_{\max})$ and $f_{min} = f(\phi_{t-1}, \theta_{\min})$. The problem is to find conditions on ϕ_t under which $\widehat{\theta}_t$ converges to θ asymptotically.

It is assumed that the function f at any time instant can be either concave or convex with respect to the parameter θ. This property of f shall be called as the curvature of f. It should be noted that the case when f is linear in θ represents the transition between concavity and convexity or vice versa, and in such a case, the curvature can be labeled as either being convex or concave.

In LP systems, the term "persistently exciting" was used to characterize a signal which was rich enough in content to enable the convergence of parameter estimates to their true values by using the standard linear gradient-update algorithms (see [17, 38]). In order to distinguish it from its counterpart for LP systems, we will use the term "NLP persistent excitation" to specify a signal which allows convergence of parameter estimates to their true values in an NLP system, using the min-max algorithm. The required conditions for a signal to be NLP persistently exciting are stated in Definition 2.2. For the sake of completeness and comparison purposes, we also define LP persistent excitation in Definition 2.3.

Definition 2.2 *A function $\phi : N \to \mathbb{R}^n$ is NLP-persistently exciting with respect to $f(\phi, \theta)$, where $f : \mathbb{R}^n \times \Theta \to \mathbb{R}$, if at any time instant t_a, given any θ_1, $\theta_2 \in \Theta$ there exist positive constants T and ϵ_f and a time instant $t_p \in [t_a + 1, t_a + T]$, such that*

(NLP-I) $\left|f\left(\phi_{t_p}, \theta_2\right) - f\left(\phi_{t_p}, \theta_1\right)\right| \geq \epsilon_f \|\theta_2 - \theta_1\|$*;*

and *at $t = t_p$,* **either**

(NLP-IIa) $sign\left(f(\phi_{t_p},\theta_2)-f(\phi_{t_p},\theta_1)\right) \neq sign\left(f(\phi_{t_a},\theta_2)-f(\phi_{t_a},\theta_1)\right)$ *while the curvature of f at t_a and curvature of f at t_p are the same* **or**

(NLP-IIb) $sign\left(f(\phi_{t_p},\theta_2)-f(\phi_{t_p},\theta_1)\right) = sign\left(f(\phi_{t_a},\theta_2)-f(\phi_{t_a},\theta_1)\right)$, *while the curvature of f at t_a and curvature of f at t_p are different.*

Definition 2.3 *A function $\phi : N \rightarrow \mathbb{R}^n$ is LP-persistently exciting if for all t there exist positive constants l and α such that [1]*

$$\sum_{i=1}^{l} \phi_{t+i}\, \phi_{t+i}^T \;\geq\; \alpha\, I \tag{2.39}$$

where I is an $n \times n$ identity matrix.

The requirements for NLP-persistent excitation consist of two components. The first component is condition *(NLP-I)* and, when f is linear, it is equivalent to the LP persistent excitation, as shown below. Condition *(NLP-II)* is a second component of NLP-persistent excitation and is needed to overcome the presence of the dead-zone which in turn was required in the min-max algorithm to ensure stability. Condition *(NLP-II)* essentially states that, periodically, the probing input ϕ should be such that f is appropriately dithered resulting in a change of either its curvature or its magnitude.

We now show that Definition 2.3 and condition *(NLP-I)* are equivalent when f is linear. Suppose ϕ is LP-persistently exciting. Then, the inequality in (2.39) is rewritten as

$$w^T \sum_{i=1}^{l} \phi_{t+i}\, \phi_{t+i}^T\, w \;\geq\; w^T \alpha\, I\, w \geq \alpha\, w^T\, I\, w \;\geq \alpha \tag{2.40}$$

where w is any unit vector in $\mathbb{R}^n$. Since l is finite, Eq. (2.40) implies that there exist a $t^* \in [t+i, t+l]$ and $\epsilon > 0$ such that

$$w^T\, \phi_{t^*} \phi_{t^*}^T\, w \;\geq\; \epsilon \tag{2.41}$$

For $\theta_1, \theta_2\ \in \mathbb{R}^n$, let $w = \dfrac{\theta_2 - \theta_1}{||\theta_2 - \theta_1||}$. Noting that for LP systems $f(\phi, \theta) = \phi^T \theta$ and using Eq. (2.41) we have that

$$[f(\phi_{t^*}, \theta_2) - f(\phi_{t^*}, \theta_1)]^2 = \|\theta_2 - \theta_1\|^2 \; w^T \phi_{t^*} \phi_{t^*}^T w \geq \epsilon \, \|\theta_2 - \theta_1\|^2 \tag{2.42}$$

This satisfies condition *(NLP-I)*.

That condition *(NLP-I)* implies LP-persistent excitation can be established as follows. Condition *(NLP-I)* for LP systems can be expressed as

$$\left\| \phi_{t_p}^T (\theta_2 - \theta_1) \right\| = \left| \phi_{t_p}^T w \right| \|(\theta_2 - \theta_1)\| \geq \epsilon_f \, \|(\theta_2 - \theta_1)\| \tag{2.43}$$

where $w = \dfrac{\theta_2 - \theta_2}{\|(\theta_2 - \theta_1)\|}$ is a unit vector in $\mathbb{R}^n$. The inequality above implies that $\left|\phi_{t_p}^T w\right| \geq \epsilon_f$ for an arbitrary w. Hence, $\exists \alpha = \epsilon_f^2$ such that

$$\left|\phi_{t_p}^T w\right|^2 = w^T \phi_{t_p} \phi_{t_p}^T w \geq \alpha.$$

If $l \geq T$, since $t_p \in [t+1, t+l]$, it follows that

$$\alpha \leq w^T \phi_{t_p} \phi_{t_p}^T w \leq w^T \sum_{i=1}^{l} \; \phi_{t+i} \phi_{t+i}^T \, w$$

and, hence, ϕ is LP-persistently exciting.

The LP-persistent excitation states that parameter convergence follows if the signal input to the system is such that for a large error in the parameter, it produces an observable difference in the output between the plant and the estimator. Essentially, this is an identifiability condition, since parameter updates are not possible if no error in the output is observed. As such, it is needed for NLP persistent excitation as well.

In order to establish parameter convergence in Eq. (2.37), we note first that $\|\tilde{\theta}\|$ is non-increasing, which follows from the fact that $\Delta V_t \leq 0$, as established in [47]. As can be seen in the adaptive system equations in (2.37), it is possible for adaptation to stop. This occurs whenever ρ_t is small. To accommodate this behavior, the following notation is introduced. Let the set Ω_D denote the set of all time such that

$$\Omega_D = \{t \mid 0 \leq \rho_t < \epsilon_\rho\}, \qquad \text{where } \epsilon_\rho \text{ is a constant} \in (0,2). \tag{2.44}$$

If ϵ_ρ is sufficiently close to zero, then Ω_D represents the time the system spends in the "dead-zone" where parameter adaptation is turned off. The complement of Ω_D is defined as $\Omega_D^C = \{t \mid \rho_t \geq \epsilon_\rho\}$. The question therefore is whether ϕ_t can be chosen so that the trajectories lie in Ω_D^C sufficiently often, which is answered in the affirmative below. As a first step, an important lemma which states a necessary condition for the system to be in the "dead-zone" is given. This is followed by Theorem 2.3 which presents the main result in parameter convergence.

Lemma 2.6 *For the adaptive system given by Eq. (2.37), if $t \in \Omega_D$ then* **either**

(D1) f_{t-1} is concave in θ, and $\tilde{y}_t > 0$ **or**

(D2) f_{t-1} is convex in θ, and $\tilde{y}_t < 0$.

Proof: It follows from Eq. (2.37) that $\rho_t < \epsilon_\rho$ if and only if $J_0 \neq 0$. From Eq.(2.38), it follows that a necessary condition for $J_0 \neq 0$ is that $sign(\tilde{y}_t) f_{t-1}$ is concave, which proves Lemma 2.6. ■

Theorem 2.3 *For the system given by Eq. (2.37), if ϕ is NLP-persistently exciting and $\hat{\theta}_t \in \Theta\ \forall t$, then $\tilde{\theta}_t \to 0$ as $t \to \infty$.*

Proof: From the fact that the min-max estimator is a stable (see [47]), we have that $\Delta V_t \leq 0$, and it follows that $\forall t \geq t_0$, $||\tilde{\theta}_t|| \leq ||\tilde{\theta}_{t_0}||$. Let t_1 be an arbitrary time instant such that $t_1 \geq t_0$. From condition *(NLP-I)*, it follows that there exists a $t_{p_1} \in [t_1, t_1 + T]$ such that

$$|\tilde{f}_{t_{p_1}}| \geq \epsilon_f ||\tilde{\theta}_{t_{p_1}}|| \tag{2.45}$$

We establish parameter convergence by showing that $\left\|\tilde{\theta}\right\|$ decreases by a finite amount over $[t_1, t_1 + 2T]$. From the system equations it follows that the value of $\tilde{\theta}$ at time t_{p_1}

determines the values of the adaptation signals ρ, ω, and $\widetilde{y}$ at time instant $t_{p_1}+1$. Defining $t_2 = t_{p_1}+1$, we consider two mutually exclusive and collectively exhaustive cases, *(a)* $t_2 \in \Omega_D^C$, and *(b)* $t_2 \in \Omega_D$.

Case (a) $t_2 \in \Omega_D^C$. In this case we have that $\rho_{t_2} \geq \epsilon_\rho$. Since $k_t \omega_t^T \Gamma_\theta \omega_t = 1 - \lambda k_t$, and $V_t = \widetilde{\theta}_t^T \Gamma_\theta^{-1} \widetilde{\theta}_t$, we have that at $t = t_2$,

$$\Delta V_{t_2} \quad = \quad -\lambda k_{t_2}^2 \rho_{t_2}^2 \widetilde{y}_{t_2}^2 + k_{t_2}\rho_{t_2}\widetilde{y}_{t_2}\left[-2\left(\omega_{t_2}^T \widetilde{\theta}_{t_2-1} - \widetilde{f}_{t_2-1}\right) + \left(\rho_{t_2}-2\right)\widetilde{y}_{t_2}\right]. \tag{2.46}$$

Also, we have that $\forall t$, $k_t \geq \dfrac{1}{\lambda + \omega_{max}}$, where $\omega_{max} \geq \omega_t^T \Gamma_\theta \omega_t$. Using these facts and Eq. (2.45), and defining

$$c = \frac{\lambda\, \epsilon_\rho^2}{(\lambda + \omega_{max})^2} \frac{\epsilon_f^2}{\gamma_{max}},$$

where γ_{max} is the maximum eigenvalue of Γ_θ^{-1}, Eq. (2.46) implies that

$$V_{t_2} \leq (1-c) V_{t_{p_1}}.$$

Since V_t is non-increasing at any t, it follows that

$$V_{t_2} \leq (1-c) V_{t_1}. \tag{2.47}$$

Case (b) $t_2 \in \Omega_D$. From Lemma 2.6, it follows that either

(A) $f_{t_{p_1}}$ is concave *and* $\widetilde{y}_{t_2} > 0$, or

(B) $f_{t_{p_1}}$ is convex *and* $\widetilde{y}_{t_2} < 0$.

We provide the proof for case *(A)* in detail below. From condition *(NLP-I)* it follows that there exists a $t_{p_2} \in [t_2, t_2 + T]$ such that

$$|\widetilde{f}_{t_{p_2}}| \geq \epsilon_f \|\widetilde{\theta}_{t_{p_2}}\|, \tag{2.48}$$

since condition *(NLP-I)* is valid for some $t_p \in [t_a, t_a + T]$ for every t_a. From requirement *(NLP-II)* and since $f_{t_{p_1}}$ is concave in case *(A)*, it follows that either of the two following cases must hold:

(A-i) $\left[f(\phi_{t_{p_2}},\widehat{\theta}_{t_{p_2}})-f(\phi_{t_{p_2}},\theta)\right]\left[f(\phi_{t_{p_1}},\widehat{\theta}_{t_p})-f(\phi_{t_{p_1}},\theta)\right]<0$ and $f_{t_{p_2}}$ is concave

(A-ii) $\left[f(\phi_{t_{p_2}},\widehat{\theta}_{t_{p_2}})-f(\phi_{t_{p_2}},\theta)\right]\left[f(\phi_{t_{p_1}},\widehat{\theta}_{t_p})-f(\phi_{t_{p_1}},\theta)\right]>0$ and $f_{t_{p_2}}$ is convex

Let $t_3=t_{p_2}+1$. Case *(A-i)* implies that $\widetilde{y}_{t_3}<0$.

Lemma 2.6 implies *(D1)* and *(D2)* are necessary conditions for t to lie in Ω_D. However, since $f_{t_{p_2}}$ is concave in case *(A-i)*, then neither *(D1)* nor *(D2)* are satisfied. Hence, it follows that in case *(A-i)* $t_3\in\Omega_D^C$. Similarly, in case *(A-ii)* we have that $\widetilde{y}_{t_3}>0$ and $f_{t_{p_2}}$ is convex. Once again, neither *(D1)* nor *(D2)* in Lemma 2.6 are satisfied, and we have that $t_3\in\Omega_D^C$. A similar argument can be given for case *(B)* as well to conclude that if $t_2\in\Omega_D$, then $t_3\in\Omega_D^C$. This implies that $\rho_{t_3}\geq\epsilon_\rho$. Therefore, similar to case *(a)*, it follows that

$$V_{t_3}\leq(1-c)V_{t_2}. \tag{2.49}$$

Combining cases *(a)* and *(b)*, and since Definition 2.2 implies that there is a finite T such that $T=max\{t_3-t_2,t_2-t_1\}$, for any t_1, we have that

$$V_{t_1+2T}\leq(1-c)V_{t_1}. \tag{2.50}$$

Since t_1 is arbitrary in all of the above arguments, Eq. (2.50) implies uniform asymptotic stability of $\widetilde{\theta}=0$. ■

Parameter convergence is established in Theorem 2.3 essentially by showing that V_t decreases by a finite amount over each interval $2T$. Twice the period of persistent excitation, i.e $2T$, is required for this decrease to occur due to the possibility of adaptation to be stopped. The proof of Theorem 2.3 shows that a fraction of $2T$ is required for the parameter estimate to leave the deadzone while the remaining fraction is required for the parameter to decrease by a finite amount. In particular, case *(b)* shows that for any t_1, it is possible for $\rho_t<\epsilon_\rho\ \forall t\in[t_1,t_{p2}]$, and $\rho_t\geq\epsilon_\rho$ at $t=t_{p_2}+1$, where $t_{p_2}\in[t_1+1,t_1+2T-1]$, which results in ΔV_t to decrease at time $t_{p_2}+1$. Conditions *(NLP-II)* and *(NLP-I)* are needed to establish that ρ_t becomes greater than or equal to ϵ_ρ at $t_{p2}+1$, and that V_t decreases at t_{p2}, respectively.

Since both properties are required at the same time instant, conditions *(NLP-I)* and *(NLP-II)* have to be satisfied by $f(\phi_t, \theta)$ at the same time instant t_p. As mentioned before, parameter convergence in LP systems does not encounter the presence of a deadzone. As a result, case *(b)* is not relevant and therefore the period over which V decreases coincides with T, the period of persistent excitation [38]. We also note that, by choosing a projection algorithm as in [6] instead of the parameter update for $\widehat{\theta}_t$ in (2.37), we can relax the requirement that $\widehat{\theta}_t$ belong to Θ $\forall t$ to the requirement that $\widehat{\theta}_{t_0}$ be in Θ.

The above discussion illustrates that for parameter convergence to occur, periodically ρ_t must become sufficiently large, which also implies that periodically, the gradient solution for ω_t is invoked. While such a gradient feature is necessary for parameter convergence, it should however be emphasized that the gradient algorithm alone cannot guarantee stability of the estimation process. The min-max component is essential to guarantee stability. The discussions in this section merely illustrate that the gradient component of the min-max algorithm has to be activated periodically for parameter convergence to occur.

2.5.2 An example of NLP-persistent excitation

By comparing the results derived for the continuous-time estimation algorithm in Section 2.3 with the discrete time version in Section 2.5.1, we see that they are analogous to each other. The results for both cases state that the NLP-PE conditions consist of two parts. The first is similar to persistent excitation conditions for linearly parameterized systems, while the second one imposes an additional requirement on the sign of the estimation error and the curvature of f.

In this section, we illustrate the similarity between the discrete and continuous time versions of the algorithm. We show that in a discrete systems with the same type of nonlinearity as studied in Section 2.4, the same type of input will be NLP-persistently exciting. The example system is given by:

$$f = e^{-\phi^T \theta} \tag{2.51}$$

where $\phi : N \to \mathbb{R}^n$, $\theta \in \Theta \subset \mathbb{R}^n$. The following definition states the desired property of the probing signal ϕ.

Definition 2.4 *Let $w \in \mathbb{R}^n$ be any unit vector. A bounded function $\phi : N \to \mathbb{R}^n$ is said to belong class K^n if for any $t_a > t_0$, there exist positive constants ϵ_ϕ and T, and a time instant $t_p \in [t_a + 1, t_a + T]$ such that*

$$\phi_{t_p}^T \, w \geq \epsilon_\phi$$

Definition 2.4 states that, periodically, the vector ϕ should have a positive component along every w in $\mathbb{R}^n$. This is more restrictive than the linear case requirements in Definition 2.3, since the latter requires ϕ to merely have a nonzero component periodically along every vector in $\mathbb{R}^n$. An example of a function that belongs to class K^2 is

$$\phi = [\sin \nu \, t \, , \cos \nu \, t \,]^T \tag{2.52}$$

Since such a ϕ represents a rotating vector $\mathbb{R}^2$ with a constant angular velocity, it follows that it aligns itself with every w in $\mathbb{R}^2$ periodically.

Lemma 2.7 *For f defined as in Eq. (2.51), $\phi \in K^n$ implies that ϕ is NLP-persistently exciting.*

Proof: For any θ_2, θ_1, and t_1, Definition 2.4 implies that $\exists t_a > t_1$ such that $\left|\phi_{t_a}^T (\theta_2 - \theta_1)\right| > 0$. Then, there there exists a unit vector $u \in \mathbb{R}^n$

$$u = -\frac{\theta_2 - \theta_1}{||\theta_2 - \theta_1||} \; sign \left(\phi_{t_a}^T (\theta_2 - \theta_1) \right) . \tag{2.53}$$

From Definition 2.4, it follows that, given ϕ_{t_a}, $\exists t_p > t_a$ such that $\phi_{t_p}^T u \geq \epsilon_\phi$. From the choice of u in (2.53), this implies that

$$\begin{aligned} |f(\phi_{t_p}, \theta_2) - f(\phi_{t_p}, \theta_1)| &= e^{-\phi_{t_p}^T \theta_2} \left| 1 - e^{-\phi_{t_p}^T (\theta_2 - \theta_1)} \right| \\ &\geq e^{-M} e^{-r} \epsilon_\phi ||\theta_2 - \theta_1|| \end{aligned}$$

$$\geq \epsilon_f ||\theta_2 - \theta_1||$$

where $M = \max_{\theta \in \Theta} ||\theta|| \sup_t ||\phi_t||$, $r = \max_{\theta_2, \theta_1 \in \Theta} ||\theta_2 - \theta_1|| \sup_t ||\phi_t||$, and $\epsilon_f = e^{-M} e^{-r} \epsilon_\phi$. Therefore, ϕ satisfies *(NLP-I)*.

Since f in Eq. (2.51) is always convex for any ϕ and θ, we only need to show that $\phi \in K^n$ implies that ϕ satisfies *(NLP-IIa)*. From the choice of u and Definition 2.4, it follows that

$$sign\left(\phi_{t_p}^T (\theta_2 - \theta_1)\right) = -\ sign\left(\phi_{t_a}^T (\theta_2 - \theta_1)\right)$$

and hence,

$$sign\left(f\left(\phi_{t_p}, \theta_2\right) - f\left(\phi_{t_p}, \theta_1\right)\right) = sign\left(1 - e^{\phi_{t_p}^T (\theta_2 - \theta_1)}\right) = -sign\left(1 - e^{\phi_{t_a}^T (\theta_2 - \theta_1)}\right)$$

for any t_a, establishing that ϕ satisfies *(NLP-II)*. ∎

2.6 Concluding Remarks and Future Work

The parameter convergence of the min-max estimator was enabled by studying an error model of the form of (2.10) and imposing conditions of NLP-persistent excitation on f. The NLP-persistent excitation conditions were presented for both the continuous and discrete time versions of the min-max estimator. The derived conditions are analogous in the two cases, and each contain two components. The first component is analogous to the persistent excitation conditions for linearly parameterized systems. It requires that the estimation error $\tilde{f} = \hat{f} - f$ be periodically large if the parameter error is large. The second component of the NLP conditions couples the sign of the estimation error $\tilde{f}$ and the curvature of f. In essence, the second component states that the estimation error and the curvature of f must be such that the gradient part of the min-max estimator is periodically activated.

It should be noted that the same error model and therefore the same conditions under which parameter convergence results is applicable for a fairly large class of problems of estimation and control. For example consider a plant and controller given by

$$\begin{aligned} \dot{x} &= A\,x + b\,(u + f(\phi,\theta)) \\ u &= -f(\phi,\widehat{\theta}) + r + \alpha^T x - asat((x - x_m)) \end{aligned} \tag{2.54}$$

where θ is unknown, $\widehat{\theta}$ is the estimate of θ, x_m is the desired state for x with r as a reference input, and $(A + b\ \alpha^T)$ is asymptotically stable. The same condition as in Theorem 2.2 enables the convergence of $\widehat{\theta}$ to θ.

Several extensions of the above approach remain to be carried out.

1. Extensions to the case when f is general will most likely have to exploit similar features in the min-max algorithm such as those used in the proof of Theorems 2.2, 2.3. ϕ may have to behave such that the regions in the parameter estimate-space where the gradient features are invoked are visited more often than others so that the parameter errors continue to decrease to zero.

2. The class of ϕ's that are NLP-persistently exciting for a general f is yet another question that remains to be addressed. Such a characterization may be difficult to obtain and may be obtainable perhaps on a problem-by-problem basis.

3. Since the results derived in this section for concave/convex parameterizations assume that θ is a vector, the results are valid when both linear and nonlinear parameters are present in the system by treating the linear component as another concave (or convex) parameter. However, if any of the parameters occur linearly, by making use of the linearity, it may be possible to relax the persistent excitation requirements of the underlying ϕ.

Chapter 3

Adaptive Control of Nonlinearly Parameterized Systems with a Triangular Structure

3.1 Introduction

One of the most common assumptions made in the context of adaptive control is that the unknown parameters occur linearly, and appear in linear [38] and nonlinear systems [24, 32, 45, 46]. Recently, a new approach has been developed [2, 34, 29, 3] to address nonlinearly parameterized (NLP) systems and their adaptive control. The main problem that is introduced due to nonlinearity in the parameterization is the failure of the gradient approach. Whether viewed from an optimization or a stabilization view-point, the gradient scheme is a powerful and simple procedure for adapting the adjustable parameters to cope with parametric uncertainty. When parameters occur linearly, the gradient scheme is sufficient to minimize the underlying cost function related to the parameter error; the gradient scheme guarantees a quadratic Lyapunov function leading to global stability. These properties are not sufficient when parameters occur nonlinearly. The approach in [2, 34, 29, 3] outlines the construction of an alternative strategy for generating adaptation laws that guarantee stability. In [2], it is assumed that the underlying parameterization is convex/concave which is made use of in constructing a quadratic Lyapunov function. In [34], the results are extended to include general parameterizations. In both cases, it is assumed that state variables are accessible and that the underlying class of nonlinear systems are of the form

$$\dot{X}_p = A_p X_p + b(f(\phi(t), \theta) + u) \tag{3.1}$$

where f is a scalar nonlinearity in the unknown parameter θ, can be globally stabilized. In [29], the class in (3.1) is extended further to include a special class of systems where matching conditions [51] are not satisfied. These systems are second-order, have a triangular structure, and are of the form

$$\begin{aligned} \dot{x}_1\,(t) &= x_2(t) + \sum_{i=1}^{n} \sigma_i f_i(x_1(t), \theta_i) \\ \dot{x}_2\,(t) &= f_0(x, \theta_{j_0}) + u(t) \end{aligned} \tag{3.2}$$

where $x = [x_1, x_2]^T$, x_1, x_2, and u are scalar functions of time, σ_i, θ_i, and θ_{j_0} are scalar parameters which are unknown, and f_0: $\mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$, f_i: $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $i = 1, ..., n$. Global stabilization and tracking to within a desired precision is established in [29].

In this Chapter, we seek to generalize these results to systems with a triangular structure and are of arbitrary order. These systems can be described as

$$\begin{aligned} \dot{x}_1 &= \gamma_1(x_2) + f_1(x_1, \theta_1) \\ \dot{x}_2 &= \gamma_2(x_3) + f_2(x_1, x_2, \theta_2) \\ &\vdots \\ \dot{x}_n &= u + f_n(x_1, x_2, \ldots, x_n, \theta_n) \end{aligned} \tag{3.3}$$

where $x = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^n$, $\theta_i \in \mathbb{R}^{m_i}$, and θ_i are unknown parameters. The goals are (i) to stabilize the system at the origin, (ii) for x_1 to track a desired trajectory x_{d_1}, and (iii) establish robustness with respect to bounded disturbances d_i. In all cases, it is assumed that the states are accessible, and that the unknown parameters belong to a compact set Θ_{m_i} in $\mathbb{R}^{m_i}$.

Since the "matching conditions" are not met, a static output-feedback linearization controller (see [40, 23, 49]) is not applicable. Hence, some form of dynamic output-feedback linearization is required to control the system in (3.3). In addressing this problem, there exist a number of results relevant to certain forms of the general class of systems described in (3.3). In the case that the functions f_i, $i = 1, \ldots, n$ are linearly parameterized, [46, 32, 31] present techniques for designing a stable adaptive controller. However, since the presented results depend on the fact that the parameterization in the system is linear, they are not directly applicable to the nonlinear parameterization we consider here. Set point regulation is considered for nonlinearly parameterized systems in [36]. Since [36] considers only the output as measurable, it is assumed that the parametric nonlinearities depend on the output only. A self-tuning controller is developed for the case when the bounds on the unknown parameter are known, but parameter estimation is not carried out. Dynamic input-output feedback linearizaton controller is developed in [48]. By using a dynamics sliding surface approach, this controller is robust with respect to bounded parametric uncertainties, and guarantees output trajectory tracking to within a set precision. However, it does not employ parameter estimation techniques. In [22], only local results are achieved for nonlinearly parameterized systems.

Both nonadaptive and adaptive controllers are proposed to accomplish the stabilization. In the nonadaptive case, global boundedness is established by making use of bounding functions that are independent of the unknown parameters. The bounding function approach in this case is similar to the techniques used in [48, 31]. In the adaptive case, the controller proposed provides a stability framework for estimating the unknown parameters. For this purpose, in addition to a bounding function, functions generated using a min-max optimization problem are introduced in the adaptive controller, as in [2].

The Chapter is organized as follows. In section 3.2, we present the adaptive controller for systems in chain-form which is a special case of (3.4). The structure of the controller as well as various properties of the closed-loop system including global stability, tracking, and robustness to bounded disturbances are established in Sections 3.2.3 - 3.2.5. We also present in section 3.2.6, extensions of systems in chain-form to LNL systems [30] and n coupled second-order systems are discussed. A simulation example is included to illustrate the controller performance in Section 3.2.8. Finally, in Section 3.3, we present the stabilizing controller for systems in general triangular form. Concluding remarks and future work is presented in Section 3.4.

3.2 The Adaptive Controller for Systems in Chain Form

We first address systems that are of the form

$$\begin{aligned}
\dot{z}_1 &= z_2 + f_1(z_1, \theta) \\
\dot{z}_2 &= z_3 \\
&\vdots \\
\dot{z}_n &= u
\end{aligned} \tag{3.4}$$

where θ is unknown, and z_i's and θ are scalars.

A few preliminary definitions and lemmas are stated in section 3.2.1. A controller structure is proposed and its rationale is discussed in section 3.2.2. In section 3.2.3, the complete controller is presented and its stability property is stated and proved.

3.2.1 Preliminaries

Definition 3.1 *A function $g(\theta) : \mathbb{R} \to \mathbb{R}$ is said to be (i) convex on a compact set Θ in $\mathbb{R}$ if $\forall \theta_1, \theta_2 \in \Theta$ it satisfies the inequality*

$$g(\lambda\theta_1 + (1-\lambda)\theta_2) \leq \lambda g(\theta_1) + (1-\lambda) g(\theta_2) \tag{3.5}$$

and (ii) concave if $\forall \theta_1, \theta_2 \in \Theta$ *it satisfies the inequality*

$$g(\lambda\theta_1 + (1-\lambda)\theta_2) \geq \lambda g(\theta_1) + (1-\lambda)g(\theta_2) \tag{3.6}$$

where $0 \leq \lambda \leq 1$.

A useful property of these functions is their relation to the gradient. When $f(\theta)$ is convex and differentiable on Θ, then it can be shown that

$$f(\theta) - f(\theta_0) \geq \nabla f_{\theta_0}(\theta - \theta_0) \qquad \forall \theta, \theta_0 \in \Theta \tag{3.7}$$

and when $f(\theta)$ is concave on Θ, then

$$f(\theta) - f(\theta_0) \leq \nabla f_{\theta_0}(\theta - \theta_0) \qquad \forall \theta, \theta_0 \in \Theta \tag{3.8}$$

where $\nabla f_{\theta_0} = \frac{\partial f}{\partial \theta}|_{\theta_0}$.

The following lemmas are useful in the development of the adaptive controllers.

Lemma 3.1 *Let* Θ *be a compact set in* $\mathbb{R}$ *specified by* $\Theta = \left[\underline{\theta}, \overline{\theta}\right]$. *For a given* $\widehat{\theta} \in \Theta$, *let*

$$J(\omega, \theta) = \beta\left[f(\phi, \theta) - f(\phi, \widehat{\theta}) + \omega(\widehat{\theta} - \theta)\right] \tag{3.9}$$

$$a_0 = \min_{\omega \in \mathbb{R}} \max_{\theta \in \Theta} J(\omega, \theta) \tag{3.10}$$

$$\omega_0 = arg \min_{\omega \in \mathbb{R}} \max_{\theta \in \Theta} J(\omega, \theta) \tag{3.11}$$

where β *and* ϕ *are known quantities independent of* θ. *Then, given* ϕ *and* β, *and defining* $g(\theta) = sign(\beta) f(\phi, \theta)$,

$$a_0 = \begin{cases} \beta\left[f_{\min} - \widehat{f} + \dfrac{f_{\max} - f_{\min}}{\overline{\theta} - \underline{\theta}}(\widehat{\theta} - \underline{\theta})\right], & \\ & \textit{if } g(\theta) \textit{ is convex on } \Theta \\ 0 & \textit{if } g(\theta) \textit{ is concave on } \Theta \end{cases} \tag{3.12}$$

$$\omega_0 = \begin{cases} \dfrac{f_{\max} - f_{\min}}{\overline{\theta} - \underline{\theta}} & \text{if } g(\theta) \text{ is convex on } \Theta \\ \nabla f_{\widehat{\theta}} & \text{if } g(\theta) \text{ is concave on } \Theta \end{cases} \tag{3.13}$$

where $\widehat{f} = f(\phi, \widehat{\theta})$, $f_{\max} = f(\phi, \overline{\theta})$, *and* $f_{\min} = f(\phi, \underline{\theta})$.

Proof: See [2, 34]. •

Lemma 3.2 *Let* α, ϵ *be arbitrary positive quantities, and let* $\alpha_{max} \geq \alpha$. *For a given* $\widehat{\theta} \in \Theta \subset \mathbb{R}$, *let* a_0 *and* ω_0 *be chosen as in eqs. (3.10)- (3.11) with* $\beta = \alpha_{max} sign(z)$, $z \in \mathbb{R}$. *If* $|z| \geq \epsilon$, *the following is then true* $\forall \phi$ *and* $\forall \theta \in \Theta$, *whether* $f(\phi, \theta)$ *is concave or convex on* Θ:

$$z \left\{ \alpha \left[f(\phi, \theta) - f(\phi, \widehat{\theta}) + (\widehat{\theta} - \theta)\omega_0 \right] - a_0 sat\left(\frac{z}{\epsilon}\right) \right\} \leq 0 \tag{3.14}$$

where the function sat(.) denotes the saturation function.

Proof: See [29]. •

In what follows, the notation $x = [x_i] \in \mathbb{R}^n$ is used to represent a vector in $\mathbb{R}^n$ whose components are x_i, $i = 1, \ldots, n$. Also, the function $\sigma(x) : \mathbb{R}^n \to \mathbb{R}^n$ is used to denote the unit vector acting in the direction x, and is defined as

$$\sigma(x) = \begin{cases} \dfrac{x}{||x||}, & ||x|| > 0 \\ \mathbf{0}, & ||x|| = 0 \end{cases} \tag{3.15}$$

We now state the definition of a Bounding Function and its key property in Lemma 3.3.

Definition 3.2 *Let* $x, z \in \mathbb{R}$, *and* $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. *Then,* $F(x, z)$ *is said to be a Bounding Function of* $f(x, \theta)$ *with respect to* z *with a buffer* δ_f *if*

$$F(x, z) = \sigma(z) \left(\max_{\theta \in \Theta} |f(x, \theta)| + \delta_f \right) \tag{3.16}$$

It should be noted that this definition implies that $\dfrac{\partial F}{\partial z} = 0$ when $z \neq 0$.

Lemma 3.3 *If $F(x,z)$ is a bounding function of $f(x,\theta)$ with respect to z with a buffer $\delta_f > \epsilon$, then for any y, all $\theta \in \Theta$, a compact set in $\mathbb{R}$, and all $|z| > 0$ the following holds:*

$$\sigma(z)\ \left[f(z,\theta) - F(z) + \epsilon sat(y)\right] \leq -\left(\delta_f - \epsilon\right) < 0 \tag{3.17}$$

Proof: Substituting (3.16) into (3.17) it follows:

$$\begin{aligned} \sigma(z)\left[f(z,\theta) - F(z) + \epsilon sat(y)\right] &= \sigma(z)\sigma\left(f(z,\theta)\right)|f(z,\theta)| - \max_{\theta\in\Theta}|f(z,\theta)| - \delta_f + \epsilon\sigma(z)sat(y) \\ &\leq |f(z,\theta)| - \max_{\theta\in\Theta}|f(z,\theta)| - \delta_f + \epsilon < 0 \end{aligned}$$

due to the choice of $\delta_f > \epsilon$. •

We now define a Bounding Function $F(x,z)$ when x and z are vectors in $\mathbb{R}^n$.

Definition 3.3 *Let $x, z \in \mathbb{R}^n$, and $f(x,\theta) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$. Define $\overline{f}_i : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, the components of $\overline{f} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, as*

$$\overline{f}_i(x;z_i) = \sigma(z_i)\left(\max_{\theta\in\Theta}|f_i(x,\theta)| + \overline{\delta}_{f_i}\right) \qquad i = 1,...n \tag{3.18}$$

The values $\overline{\delta}_{f_i}$, $i = 1,\ldots,n$ are the components of the vector $\overline{\delta}_f \in R^n$. Then the bounding function $F(x,z)$ of $f(x,\theta)$ with respect to z with a buffer $\overline{\delta}_f = [\delta_{f_i}] \in \mathbb{R}^n$, is defined as

$$F(x,z) = u_z u_z^T \overline{f} \tag{3.19}$$

where $u_z = \sigma(z)$.

Similar to Lemma 3.3, the key property of the vector bounding function is stated in Lemma 3.4.

Lemma 3.4 *Let $x, y, z \in \mathbb{R}^n$, $u_z = \sigma(z)$, $\epsilon = [\epsilon_i] \in \mathbb{R}^n$. If $F(x,z)$ is a bounding function of $f(x,\theta)$ with respect to z with a buffer $\epsilon_f = [\epsilon_{f_i}]$ such that $\epsilon_{f_i} > \epsilon_i$, then for any y and all $\theta \in \Theta$, a compact set in $\mathbb{R}^m$, the following holds:*

$$u_z^T\left[f(z,\theta) - F(z) + \epsilon^T sat(y)\right] \leq 0 \tag{3.20}$$

where $sat(x) = x$ *for all* $|x| \leq 1$, *and* $sat(x) = sign(x)$ *for all* $|x| > 1$.

Proof: From (3.19), it follows that $u_z^T F = u_z^T u_z u_z^T \overline{f} = u_z^T \overline{f}$. Thus, by using eq. (3.18), (3.20) can be expressed as

$$
\begin{aligned}
u_z^T \left[(f(z,\theta) - F(z)) + \epsilon^T v \right] &= \sum_{i=1}^{n} u_{z_i} \left(f_i - \sigma(z_i) \max_{\theta \in \Theta} |f_i(x,\theta)| + \epsilon_i sat(y_i) - \sigma(z_i)\epsilon_{f_i} \right) \\
&= \sum_{i=1}^{n} |u_{z_i}| \left(\sigma(z_i) f(z,\theta) - \max_{\theta \in \Theta} |f(z,\theta)| - (\epsilon_{f_i} - \sigma(z_i) sat(y_i)\epsilon_i) \right) \\
&\leq \sum_{i=1}^{n} |u_{z_i}| \left(|f(z,\theta)| - \max_{\theta \in \Theta} |f(z,\theta)| - (\epsilon_{f_i} - \epsilon_i) \right) \leq 0
\end{aligned}
$$

due to the choice of ϵ_f. •

In order to ensure the continuity of the proposed adaptive controller, a smooth bounding function is required. The following definitions and Lemma 3.5 serve this purpose.

Definition 3.4 *A smoothing function* $S(z,\epsilon) : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ *is an (n-1) times differentiable odd function which satisfies the following:*

$$
\begin{aligned}
\sigma(S(z,\epsilon)) &= \sigma(z) \\
S(z,\epsilon) &= \sigma(z) \quad \forall |z| \geq \epsilon > 0 \qquad (3.21) \\
|S(z,\epsilon)| &\leq 1
\end{aligned}
$$

One example of $S(z)$ is a $sat(z)$ function with smooth corners.

Definition 3.5 $F(x,z,\epsilon) : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *is said to be a* smooth *bounding function of* $f(x,\theta)$ *with respect to* z *and a buffer* ϵ_f *if*

$$
F(x,z,\epsilon) = S(z,\epsilon) \left(\max_{\theta \in \Theta} |f(x,\theta)| + \delta_f \right) \qquad (3.22)
$$

Analogous to Lemma 3.3, we state Lemma 3.5 for smooth bounding functions. A corresponding Lemma can be stated for smooth vector bounding functions as well.

Lemma 3.5 *If $F(x, z, \epsilon_0)$ is a bounding function of $f(x, \theta)$ with respect to z with a buffer $\delta_f > \delta$, then for any y, all $\theta \in \Theta$, a compact set in $\mathbb{R}^m$, and for all $|z| > \epsilon_0$, the following holds:*

$$\sigma(z) \ \left[(f(x, \theta) - F(x, z, \epsilon_0)) + \delta S(y, \epsilon_0) \right] \leq - (\delta_f - \delta) < 0 \tag{3.23}$$

Proof: The proof follows straightforwardly from the definition of $F(x, z, \epsilon_0)$ and is similar to the proof of Lemma 3.3. •

3.2.2 The controller structure

This section outlines the basic ideas behind our approach to designing a controller for the system in (3.4). In [29], we have derived a stabilizing controller for (3.4) when $n = 2$. The question therefore pertains to the complexities introduced by a higher order system. This section discusses these complexities, and how they can be addressed, with the starting point being the approach taken in [29]. The specifics of the controller realization are then presented in section 3.2.3.

Briefly, the results in [29] are as follows. The system under consideration is of the form

$$\begin{aligned} \dot{z}_1 &= z_2 + f_1(z_1, \theta) \\ \dot{z}_2 &= u \end{aligned} \tag{3.24}$$

and the goal is global stabilization using u when θ is unknown. The main obstacle here is that the dynamics of z_1 are not under our direct control. Rather, the control input is passed through the integrator, which introduces an inherent time lag and can potentially destabilize z_1. We overcome this by choosing errors e_0 and e_1 such that when $e_1 \to 0$ it assures that $e_0 \to 0$. The task that remains then is to choose a u such that e_1 tends to zero. In particular, a choice of

$$e_0 = z_1, \qquad e_1 = z_2 + g_1(z_1, e_0) \tag{3.25}$$

leads to error equations

$$
\begin{aligned}
\dot{e}_0 &= e_1 + f_1(z_1, \theta) - g_1(z_1, e_0) \\
\dot{e}_1 &= u + f_2(z_1, z_2, \theta)
\end{aligned}
\tag{3.26}
$$

where

$$
f_2 = \frac{\partial g}{\partial z_1}(z_2 + f_1) + \frac{\partial g}{\partial e_0}(e_1 + f_1 - g_1)
$$

By making g_1 a bounding function of f_1 with respect to e_0, we essentially stabilize e_0 in the absence of e_1. To choose u such that $e_1 \to 0$, especially due to uncertainties in a nonlinear parameterization, the min-max algorithm as in [2] is used.

A direct extension of this approach to higher dimensions requires appropriate characterizations of n errors, e_j, $j = 0, 1, \ldots, n-1$. The basic idea is to define these errors in such a way that if $e_i \to 0$, it guarantees that all errors e_j, $j = 0, \ldots, i-1$, tend to zero. Let us assume that the errors are of the form, as in (3.25),

$$
e_0 = z_1, \qquad e_j = z_{j+1} + g_j(z_1, \ldots, z_i, e_{j-1}), \qquad j = 1, \ldots, n-1 \tag{3.27}
$$

Suppose that these errors satisfy the relationship, as in (3.26),

$$
\dot{e}_j = e_{j+1} - e_{j-1} + f_{j+1}(z_1, \ldots, z_j, \theta) - g_{j+1}(z_1, \ldots, z_j, e_j) \qquad j = 0, \ldots, n-2 \tag{3.28}
$$

$$
\dot{e}_{n-1} = u + f_n(z_1, \ldots, z_n, \theta) \tag{3.29}
$$

where $e_{-1} = 0$, for suitably defined f_j and g_j. The advantage of the structure in (3.28)-(3.29) is apparent if g_i is a bounding function of f_i with respect to e_{i-1}, since the latter leads to the property

$$
(f_{i+1} - g_{i+1})\sigma(e_i) \leq 0
$$

This follows since

$$V = \sum_{i=0}^{n-1} e_i^2 \tag{3.30}$$

yields a time derivative of the form

$$\dot{V} = \sum_{i=0}^{n-1} \left[e_i \left(e_{i+1} - e_{i-1}\right) + e_i \left(f_{i+1} - g_{i+1}\right)\right] + e_{n-1}(u + f_n) \leq e_{n-1}(e_{n-2} + u + f_n)$$

which suggests that V is a Control Lyapunov Function for (3.28)-(3.29), leading to global stabilization.

The question that arises is if indeed errors e_j an be constructed as in (3.27) so that they satisfy (3.28)-(3.29). This is answered by the following recursive relationships:

$$\begin{aligned}
e_0 &= z_1 \\
e_1 &= z_2 + g_1(z_1, e_0) \\
e_2 &= e_0 + z_3 + g_2(z_1, z_2, e_1) \\
&\vdots \\
e_i &= e_{i-2} + z_{i+1} + g_i(z_1, \ldots, z_i, e_{i-1}) \quad i = 1, \ldots, n-1
\end{aligned} \tag{3.31}$$

where $z_{n+1} = u$,

$$g_i(z_1, \ldots, z_i, e_{i-1}) = k_{i-1}(z_1, \ldots, z_i) + \overline{h}_{i-1}(z_1, \ldots, z_{i-1}, e_{i-1}) \quad i = 1, \ldots, n - \tag{3.32}$$

$$\begin{aligned}
k_0(z_1) &= 0 \\
k_1(z_1) &= \frac{\partial \overline{h}_0}{\partial z_1} z_2 \\
&\vdots \\
k_i(z_1, \ldots, z_{i+1}) &= k_{i-2} + z_i + \sum_{j=1}^{i} \frac{\partial k_{i-1}}{\partial z_j} z_{j+1} + \sum_{j=1}^{i-1} \frac{\partial \overline{h}_{i-1}}{\partial z_j} z_{j+1}, \qquad i = 2, \ldots, n-1
\end{aligned} \tag{3.33}$$

$$\begin{aligned}
h_{-1} &= 0; \qquad h_0(z_1, \theta) = f_1(z_1, \theta); \\
h_i(z_1, \ldots, z_i, \theta) &= h_{i-2} + \left(\frac{\partial k_{i-1}}{\partial z_1} + \frac{\partial \overline{h}_{i-1}}{\partial z_1}\right) f_1(z_1, \theta) \qquad i = 1, \ldots, n-1
\end{aligned} \tag{3.34}$$

with $\overline{h}_i(z_1, \ldots, z_i, e_i)$ as bounding functions of h_i with respect to e_i, so that

$$(h_i - \overline{h}_i)\sigma(e_i) \;\leq\; 0 \qquad i = 0, \ldots, n-2. \tag{3.35}$$

Suppose we assume that $\dfrac{\partial g_i}{\partial e_{i-1}} = 0$ for all e_{i-1} and z_i. It can then be shown that the errors e_j satisfy (3.28) and (3.29) using the method of induction: Suppose (3.28) holds for $j = i-2$, so that

$$\dot{e}_{i-2} = z_i + f_{i-1} \tag{3.36}$$

where

$$f_i(z_1, \ldots, z_i, \theta) \;\triangleq\; k_{i-1}(z_1, \ldots, z_i) + h_{i-1}(z_1, \ldots, z_{i-1}, \theta). \tag{3.37}$$

From (3.31) and (3.36), we have that

$$\begin{aligned} \dot{e}_i &= z_{i+2} + z_i + k_{i-2} + h_{i-2} + \sum_{k=1}^{i-1}\left(\frac{\partial k_{i-1}}{\partial z_k} + \frac{\partial \overline{h}_{i-1}}{\partial z_k}\right) z_{k+1} \\ &\quad + \frac{\partial k_{i-1}}{\partial z_i} z_{i+1} + \left(\frac{\partial k_{i-1}}{\partial z_1} + \frac{\partial \overline{h}_{i-1}}{\partial z_1}\right) f_1(z_1, \theta) + \frac{\partial g_i}{\partial e_{i-1}}\, \dot{e}_{i-1} \\ &= z_{i+2} + k_i + h_i \end{aligned} \tag{3.38}$$

which establishes (3.28) for $j = i$. Defining $z_{n+1} = u$, we therefore have that (3.28)-(3.29) hold for $j = 0, \ldots, n-1$. This implies in turn that V in (3.30) is a Control Lyapunov function.

The arguments can be extended to the case when θ is unknown, by modifying V in (3.30) to

$$V \;=\; \frac{1}{2}\left(\sum_{i=0}^{n-1} e_i^2 + \tilde{\theta}^2\right) \tag{3.39}$$

where $\widehat{\theta}$ is an estimate of θ and $\tilde{\theta} = \widehat{\theta} - \theta$. As a result, we obtain that

$$\dot{V} \;=\; e_{n-1}(e_{n-2} + u + f_n) + \tilde{\theta}\,\dot{\tilde{\theta}}\,.$$

This suggests the following control strategy for u:

$$
\begin{aligned}
u &= -\widehat{f}_n(z_1, z_2, \ldots, z_n, \widehat{\theta}) - \gamma e_{n-1} - e_{n-2} - a^*\sigma(e_{n-1}), \quad \gamma > 0 && (3.40)\\
\widehat{f}_n &= k_{n-1}(z_1, \ldots, z_n) + h_{n-1}(z_1, \ldots, z_n, \widehat{\theta}) && (3.41)\\
\dot{\widehat{\theta}} &= -e_{n-1}\omega^* \qquad\qquad \gamma_n > 0 && (3.42)
\end{aligned}
$$

where a^* and ω^* are solutions of the following min-max problem:

$$
(a^*, \omega^*) = \min_{\omega} \max_{\theta} \left[f_n - \widehat{f}_n + (\widehat{\theta} - \theta)\omega\right] \tag{3.43}
$$

This leads to the relation

$$
\begin{aligned}
\dot{V} &= -\gamma e_{n-1}^2 + \left[\sum_{i=0}^{n-2} e_i(f_{i+1} - g_{i+1})\right] + e_{n-1}(f_n - \widehat{f}_n - \widetilde{\theta}\omega - a^*\mathrm{sign}(e_{n-1})) && (3.44)\\
&\leq 0
\end{aligned}
$$

from Lemmas 1 and 3, from the fact that g_i is a bounding function of f_i, and the choices of a^* and ω^*. Therefore e_i, $i = 1, \ldots, n-1$, and $\widetilde{\theta}$ are bounded. The recursive relationship between e_i's and z_i's in (3.31) implies that all z_i's are bounded. Barbalat's lemma then implies that e_i's and therefore z_i's tend to zero as $t \to \infty$.

Obviously, the above controller is predicated on the assumption that $\dfrac{\partial g_i}{\partial e_{i-1}} = 0$ for all e_{i-1} and z_i. Such an assumption cannot be guaranteed since the bounding function g_1 includes a signum function of e_0, and each g_i is recursively defined, for $i = 1, \ldots, n-1$, using g_1 and its higher-order derivatives. In the following section, we show that this difficulty can be avoided by using a smooth bounding function and that stability and asymptotic stability to within a desired precision ϵ follows.

3.2.3 A continuous controller

The controller in (3.40)-(3.41) is not defined, since it contains the derivatives of the $\sigma(\cdot)$ function. We now demonstrate how, by replacing the $\sigma(\cdot)$ function by the smooth $S(\cdot, \epsilon)$ function (in Definition 3.5 a continuous controller can be derived that guarantees all solutions converge to within a desired precision ϵ.

In order to develop the continuous controller we modify the definition of the recursive functions k_i and h_i in (3.33) and (3.34) as follows:

$$\begin{aligned}
k_0(z_1, e_0) &= 0 \\
k_1(z_1, z_2, e_0) &= \left(\frac{\partial \overline{h}_0}{\partial z_1} + \frac{\partial \overline{h}_0}{\partial e_0}\right) z_2 \\
&\vdots \\
k_i(z_1, \ldots, z_{i+1}, e_0, \ldots, e_{i-1}) &= k_{i-2} + z_i + \sum_{j=1}^{i} \frac{\partial k_{i-1}}{\partial z_j} z_{j+1} + \sum_{j=0}^{i-2} \frac{\partial k_{i-1}}{\partial e_j}\left(z_{j+2} + k_j\right) \\
&\quad + \sum_{j=1}^{i-1} \frac{\partial \overline{h}_{i-1}}{\partial z_j} z_{j+1} + \sum_{j=0}^{i-1} \frac{\partial \overline{h}_{i-1}}{\partial e_j}\left(z_{j+2} + k_j\right), \\
&\qquad i = 2, \ldots, n-1
\end{aligned} \tag{3.45}$$

$$\begin{aligned}
h_{-1} &= 0; \qquad h_0(z_1, \theta) = f_1(z_1, \theta); \\
h_i(z_1, \ldots, z_i, e_0, \ldots, e_{i-1}, \theta) &= h_{i-2} + \left(\frac{\partial k_{i-1}}{\partial z_1} + \frac{\partial \overline{h}_{i-1}}{\partial z_1}\right) f_1(z_1, \theta) + \\
&\quad \sum_{j=0}^{i-2} \left(\frac{\partial k_{i-1}}{\partial e_j} + \frac{\partial \overline{h}_{i-1}}{\partial e_j}\right) h_j + \frac{\partial \overline{h}_{i-1}}{\partial e_{i-1}} h_{i-1}, \\
&\qquad i = 1, \ldots, n-1
\end{aligned} \tag{3.46}$$

In (3.45) and (3.46), $\overline{h}_i(z_1, \ldots, z_i, e_i, \epsilon_i)$ are chosen as smooth bounding functions of h_i, with respect to e_i, with buffers δ_{f_i} such that $\delta_{f_0} \geq \epsilon_1$, and $\delta_{f_i} \geq \epsilon_{i-1} + \epsilon_{i+1}$ for $i = 2, \ldots, n-2$, and

$$e_i' = e_i - \epsilon_i \, sat(e_i/\epsilon_i), \qquad \epsilon_i > 0, \;\; i = 0, \ldots, n-1 \tag{3.47}$$

The continuous controller is given by

$$\begin{aligned}
u &= -\widehat{f}_n(z_1, z_2, \ldots, z_n, e_0, \ldots, e_{n-1}, \widehat{\theta}) - \gamma e_{n-1} - e_{n-2}' - a^* \mathrm{S}(e_{n-1}, \epsilon), \quad \gamma > 0 \qquad (3.48) \\
\widehat{f}_n &= k_{n-1}(z_1, \ldots, z_n, e_0, \ldots, e_{n-2}) + h_{n-1}(z_1, \ldots, z_{n-1}, e_0, \ldots, e_{n-1}, \widehat{\theta}) \\
\dot{\widehat{\theta}} &= -e_{n-1}' \omega^* \qquad\qquad \gamma_n > 0
\end{aligned} \tag{3.49}$$

where a^* and ω^* are chosen as in (3.43). The stabilizing property of the controller in (3.45)-(3.49) is given in the following theorem.

Theorem 3.1 *The controller defined by (3.31), (3.32), (3.43), (3.45)-(3.49) results in global stabilization of the system in (3.4) at the origin, and each $|z_i(t)|$ tends to $c_i \epsilon_i$ as $t \to \infty$ for some constants c_i, $i = 1, \ldots, n$.*

Proof: With the functions in (3.45) and (3.46), by proceeding in the same manner as in section 3.2.2 that the errors e_i defined in (3.31) can once again be shown to satisfy the equations (3.37) and (3.38) for $i = 1, \ldots, n-1$, where $z_{n+1} = u$. Choosing

$$V = \frac{1}{2}\left(\sum_{i=0}^{n-1} e_i'^2 + \tilde{\theta}^2\right) \tag{3.50}$$

and using Eqs. (3.31), (3.38), and (3.49), and the fact that $\frac{d}{dt}\left(e_i'^2\right) = 2e_i'\,\dot{e}_i$, we obtain a time derivative

$$\dot{V} = e_0'(e_1 + f_1 - g_1) + \left[\sum_{i=1}^{n-2} e_i'(e_{i+1} - e_{i-1} + f_{i+1} - g_{i+1})\right] + e_{n-1}'(u + f_n) - \tilde{\theta} e_{n-1}' \tag{3.51}$$

Substituting the control law from (3.48) into (3.51), we have

$$\begin{aligned}\dot{V} = {} & e_0'(e_1 + f_1 - g_1) + \left[\sum_{i=1}^{n-2} e_i'(e_{i+1}' - e_{i-1}' + \epsilon_{i+1} sat(y_{i+1}) - \epsilon_{i-1} sat(y_{i-1}) + f_{i+1} - g_{i+1})\right] \\ & - e_{n-2}' e_{n-1}' - \gamma\, e_{n-1}'\, e_{n-1} + e_{n-1}'\left(f_n - \hat{f}_n - \tilde{\theta}\omega^* - a^* \mathrm{S}(e_{n-1}, \epsilon)\right)\end{aligned}$$

where $y_i = e_i/\epsilon_i$. Therefore,

$$\begin{aligned}\dot{V} \leq {} & -\gamma e_{n-1}'^{\,2} + e_0'\left[e_1 + f_1 - g_1 + \epsilon_1 sat(y_1)\right] \\ & + \left[\sum_{i=1}^{n-2} e_i'\left(f_{i+1} - g_{i+1} + \epsilon_{i+1} sat(y_{i+1}) - \epsilon_{i-1} sat(y_{i-1})\right)\right].\end{aligned}$$

Since $\delta_{f_0} \geq \epsilon_1$ and $\delta_{f_i} \geq \epsilon_{i-1} + \epsilon_{i+1}$ and f_1 and $\overline{h}_i$ are bounding functions of g_1 and h_i, with buffers δ_{f_0} and δ_{f_i}, respectively, it follows that the second and third terms are nonpositive. Hence, $\dot{V} \leq 0$. Since e_i' are bounded and g_i are bounded functions of their arguments, z_i are bounded, which, by Barbalat's lemma implies that all e_i' tend to zero. This means that all $|e_i|$ tend to ϵ_i, which in turn, sets the bounds on all z_i. •

3.2.4 Tracking

We now discuss the tracking problem related to (3.4). Supppose the goal is to find u in (3.4) such that the state z_1 tracks a desired trajectory $z_d(t)$ asymptotically. We assume that z_d is sufficiently smooth, so that $\dot{z}_d, z_d^{(2)}, \ldots, z_d^{(n)}$ are bounded.

The requisite controller that accomplishes tracking while ensuring globally bounded signals in the closed-loop system is very similar to Eqs. (3.48)-(3.49). The following equations specify the adaptive controller in this case:

$$\begin{aligned}
e_0 &= z_1 - z_d \\
e_1 &= z_2 + g_1(z_1, e_0) - \dot{z}_d \\
&\vdots \\
e_i &= e_{i-2} + z_{i+1} + g_i(z_1, \ldots, z_i, e_{i-1}) - z_d^{(i)}, \;\; i = 1, \ldots, n-1
\end{aligned} \tag{3.52}$$

where

$$\begin{aligned}
g_i(z_1, \ldots, z_i, e_{i-1}) &= k_{i-1}(z_1, \ldots, z_i, e_0, \ldots, e_{i-2}, t) + \overline{h}_{i-1}(z_1, \ldots, z_{i-1}, e_{i-1}), \\
&\quad\; i = 2, \ldots, n-1 \\
k_0(z_1) &= 0 \\
k_1(z_1) &= \frac{\partial \overline{h}_0}{\partial z_1} z_2 \\
&\;\vdots \\
k_i(z_1, \ldots, z_{i+1}, t) &= k_{i-2} + (z_i - z_d^{(i-1)}) + \sum_{j=1}^{i} \frac{\partial k_{i-1}}{\partial z_j} z_{j+1} + \sum_{j=1}^{i-1} \frac{\partial \overline{h}_{i-1}}{\partial z_j} z_{j+1} \\
&\quad + \sum_{j=0}^{i-2} \frac{\partial k_{i-1}}{\partial e_j} \left(z_{j+2} + k_j - z_d^{(j+1)} \right) + \sum_{j=0}^{i-1} \frac{\partial \overline{h}_{i-1}}{\partial e_j} \left(z_{j+2} + k_j - z_d^{(j+1)} \right) \\
&\quad + \sum_{j=1}^{i} \frac{\partial k_{i-1}}{\partial z_d^{(}j)} z_d^{(j+1)},
\end{aligned}$$

$$i = 2, \ldots, n-1$$

$$\begin{aligned}
h_{-1} &= 0; \qquad h_0(z_1, \theta) = f_1(z_1, \theta); \\
h_i(z_1, \ldots, z_i, \theta) &= h_{i-2} + \left(\frac{\partial k_{i-1}}{\partial z_1} + \frac{\partial \overline{h}_{i-1}}{\partial z_1} \right) f_1(z_1, \theta) \\
&\quad + \sum_{j=0}^{i-2} \frac{\partial k_{i-1}}{\partial e_j} h_j + \sum_{j=1}^{i-1} \frac{\partial \overline{h}_{i-1}}{\partial e_j} h_j \qquad i = 1, \ldots, n-1
\end{aligned} \tag{3.53}$$

where $\overline{h}_i(\cdot)$ are smooth bounding functions of h_i with buffers δ_{f_i} such that $\delta_{f_0} > \epsilon_0$, and $\delta_{f_i} \geq \epsilon_{i-1} + \epsilon_{i+1}$ for $i = 2, \ldots, n-2$, and the control input is given by

$$\begin{aligned}
u &= -\widehat{f}_n(z_1, z_2, \ldots, z_n, \widehat{\theta}) - \gamma e_{n-1} - e'_{n-2} - a^* \mathrm{S}(e_{n-1}, \epsilon) + z_d^{(n)}, \quad \gamma > 0 \\
\widehat{f}_n &= k_{n-1}(z_1, \ldots, z_n, t) + h_{n-1}(z_1, \ldots, z_{n-1}, \widehat{\theta}) \\
\dot{\widehat{\theta}} &= -e'_{n-1} \omega^* \qquad \qquad \gamma_n > 0
\end{aligned}$$

and a^* and ω^* are defined as in (3.43).

The proof of stability with the above controller can be established in a manner identical to that of Theorem 1. It can be shown that the errors e_i satisfy the equations

$$\dot{e}_i = z_{i+2} + f_{i+1} - z_d^{(i+1)}$$

and that V as in (3.50) is a Lyapunov function. Barbalat's lemma and the definitions of the errors can be used to show that $e_0(t)$ tends to zero as $t \to \infty$.

3.2.5 Robustness

All of the above results can be extended to the case when bounded disturbances are present. For example, if (3.4) is modified to

$$\begin{aligned}
\dot{z}_1 &= z_2 + f_1(z_1, \theta) + d_1 \\
\dot{z}_2 &= z_3 + d_2 \\
&\vdots \\
\dot{z}_n &= u + d_n
\end{aligned} \tag{3.54}$$

where $\theta \in$ a compact set in $\mathbb{R}$. Global boundedness can be established under the following assumptions. (1) d_i, $i = 1, n-1$, are bounded and an upper bound $d_{i_{\max}}$ is known, and (2) d_n is bounded.

Robustness for a second order system

We first consider the system in (3.54) when $n = 2$. The requisite controller is chosen to be of the form

$$\begin{aligned}
e_0 &= z_1 \qquad (3.55)\\
e_1 &= z_2 + g_1(z_1, e_0)
\end{aligned}$$

where g_1 is such that

$$(f_1 - g_1 + d_1)\sigma(e_0) \;\le\; -\epsilon_f < 0 \tag{3.56}$$

and

$$\begin{aligned}
u \;=\; & -\left(\frac{\partial g_1}{\partial z_1} + \frac{\partial g_1}{\partial e_0}\right) z_2 - S(e_1, \epsilon)\left|\frac{\partial g_1}{\partial z_1} + \frac{\partial g_1}{\partial e_0}\right| d_{1_{max}} \\
& -\left(\frac{\partial g_1}{\partial z_1} + \frac{\partial g_1}{\partial e_0}\right) \widehat{f}_1 - \gamma e_1 - e_0 - a^* S(e_1, \epsilon)
\end{aligned} \tag{3.57}$$

where $|d_1| \leq d_{\max}$ and

$$\dot{\widehat{\theta}} \;=\; -\sigma\widehat{\theta} - e_1'\omega^* \tag{3.58}$$

and (a^*, ω^*) are min-max solutions from (3.43) for $f_n = \left(\frac{\partial g_1}{\partial z_1} + \frac{\partial g_1}{\partial e_0}\right) f_1$.

That the controller specified by eqs. (3.55)-(3.58) leads to global boundedness can be shown by noting that

$$\begin{aligned}
\dot{e}_0 &= e_1 + f_1 - g_1 + d_1 \\
\dot{e}_1 &= u + d_2 + \left(\frac{\partial g_1}{\partial z_1} + \frac{\partial g_1}{\partial e_0}\right)(z_2 + f_1 + d_1)
\end{aligned}$$

and that for $V = \frac{1}{2}({e_0'}^2 + {e_1'}^2 + \widetilde{\theta}^2)$, $\dot{V} \leq 0$ in D^c where D is a compact set in the $(e_0', e_1', \widetilde{\theta})$ space. The details are deferred to the nth order case discussed below.

Robustness for an nth order system

We now state the controller that ensures robustness for the nth order system in (3.54).

Theorem 3.2 *With the errors defined as in (3.31), g_i, k_i, and h_i as in Eq. (3.32) with $\overline{h}_i$ as a bounding function for $h_i + h_{d_i}$ with respect to e_i and a buffer $d_{i_{\max}} + \delta_{f_i} + \epsilon_{i-1} + \epsilon i + 1$, $i = 0, \ldots, n-1$, where*

$$h_{d_i} = d_{i-1} + \sum_{j=1}^{i} \frac{\partial k_{i-1}}{\partial z_j} d_j + \sum_{j=0}^{i-2} \frac{\partial k_{i-1}}{\partial e_j} d_{j+1} + \sum_{j=1}^{i-1} \left(\frac{\partial \overline{h}_{i-1}}{\partial z_j} d_j + \frac{\partial \overline{h}_{i-1}}{\partial e_j} d_{j+1} \right) \tag{3.59}$$

and $\epsilon_{-1} = \epsilon_n = 0$, choosing the control input u as

$$u = -f_n(z_1, z_2, \ldots, z_n, \theta) - \gamma e_{n-1}' - e_{n-2}', \;\; \gamma > 0 \tag{3.60}$$

when θ is known, or

$$u \;=\; -f_n(z_1, z_2, \ldots, z_n, \widehat{\theta}) - \gamma e_{n-1}' - e_{n-2}' - a^* S(e_{n-1}, \epsilon), \qquad \gamma > 0 \tag{3.61}$$

$$\dot{\widehat{\theta}} \;=\; e_{n-1}'\omega^* - \sigma\widehat{\theta} \tag{3.62}$$

when θ is unknown assures that $e_i' \to 0$ as $t \to \infty$ for all $i = 0, \ldots, n-1$.

Proof: Using the method of induction, and the proof of the above theorem for $n = 2$, it can be shown that if e_{i-2} satisfies the differential equation

$$\dot{e}_{i-2} \;=\; z_i + f_{i-1} + d_{i-1}, \qquad \text{for } \; i = 2, \ldots, n-1,$$

then,

$$\dot{e}_i \;=\; z_{i+2} + f_{i+1} + d_{i+1}, \qquad i = 2, \ldots, n-1$$

where $z_{n+1} = u$.

<u>Case (i) θ known:</u> Here, for $V = \frac{1}{2}\sum_{i=0}^{n-1} e_i'^2$,

$$\begin{aligned}
\dot{V} \;&=\; e_0'\left(e_1' + f_1 - g_1 + \epsilon_0 sat(y_0)\right) + e_{n-1}'(u + f_n + d_n) \\
&\quad + \sum_{i=0}^{n-2} e_i' d_{i+1} + \sum_{i=1}^{n-2} e_i' \left(e_{i+1}' - e_{i-1}' + \epsilon_{i+1} sat(y_{i+1}) - \epsilon_{i-1} sat(y_{i-1}) - g_{i+1} + f_{i+1}\right) \\
&=\; \sum_{i=0}^{n-2} e_i' \left(f_{i+1} - g_{i+1} + d_{i+1} + \epsilon_{i+1} sat(y_{i+1}) - \epsilon_{i-1} sat(y_{i-1})\right) \\
&\quad + e_{n-2}' e_{n-1}' + e_{n-1}'(-\gamma e_{n-1}' - e_{n-2}' + d_n)
\end{aligned} \tag{3.63}$$

where $y_i = e_i/\epsilon_i$. From the choices of g_1 and $\overline{h}_i$, it follows that

$$\left(h_i - \overline{h}_i + h_{d_i} + \epsilon_{i+1} sat(y_{i+1}) - \epsilon_{i-1} sat(y_{i-1})\right)\sigma(e_i) \;\leq\; 0 \qquad i = 0, \ldots, n-1 \tag{3.64}$$

Therefore, we have that

$$\sum_{i=0}^{n-2} e_i'(f_{i+1} - g_{i+1} + \epsilon_{i+1} sat(y_{i+1}) - \epsilon_{i-1} sat(y_{i-1}) + d_{i+1}) \;\leq\; -\delta_f \left|e_i'\right|.$$

Hence, Eq. (3.63) can be written as

$$\begin{aligned} \dot{V} &\leq -\sum_{i=0}^{n-2} \delta_{f_i} |e_i'| - \gamma {e_{n-1}'}^2 + |e_{n-1}'| d_{n_{\max}} \\ &\leq 0 \qquad \forall\, (e_0', \ldots, e_{n-1}') \in D^c \end{aligned} \tag{3.65}$$

where D is a compact set in the $(e_0', \ldots, e_{n-1}')$ space.

Case (ii) θ unknown: For $V = \frac{1}{2}\sum_{i=0}^{n-1}\left({e_i'}^2 + \tilde{\theta}^2\right)$, using (3.61) and (3.62) and proceeding as in case (i), we have

$$\begin{aligned} \dot{V} &\leq -\sum_{i=0}^{n-2} \delta_{f_i} |e_i'| - \gamma {e_{n-1}'}^2 + |e_{n-1}'| d_{n_{\max}} + e_{n-1}'\left(f_n - \hat{f}_n + \tilde{\theta}\omega^* - a^* S(e_{n-1}, \epsilon)\right) - \sigma\tilde{\theta}\hat{\theta} \\ &\leq 0 \qquad \forall\, (e_0,' \ldots, e_{n-1}', \tilde{\theta}) \in D^c \end{aligned} \tag{3.66}$$

where D is a compact set in the $(e_0', \ldots, e_{n-1}', \tilde{\theta})$ space containing the origin.

Therefore, in both cases (i) and (ii), global boundedness follows. •

We note that the choice of g_1 in (3.56) differs from the choice of the bounding function as in (3.35). However (3.56) can be satisfied using a similar procedure by choosing the buffer to be $d_{1\max} + \delta_f$.

In the above discussions, we have used a $\sigma-$modification scheme to update $\hat{\theta}$. Similar adaptive laws such as a dead-zone, and e_1-modification schemes [38] can also be employed to result in global boundedness.

3.2.6 Control of L-N-L systems

A special class of chain-form systems has three systems in cascade, which include Linear dynamics, followed by static Nonlinearities, and Linear dynamics, and referred to as LNL systems [ref.?]. One such form is given by

$$\begin{aligned} x^{(m)} &= f(z, \theta) \\ z^{(n)} &= u \end{aligned} \tag{3.67}$$

where the unknown parameter θ lies in a compact set in $\mathbb{R}^p$ and the goal is to stabilize this system and enable x to track a desired trajectory. Eq. (3.67) can be considered to be an extension of (3.4) with z_1 in (3.4) replaced by an mth order system with an output x. In what follows, we present a stabilizing controller for the case when m is arbitrary $n = 1$. The approach presented can be extended in a straightforward manner to include the systems where $n \geq 2$. Define

$$\begin{aligned} e_0 &= D(s) \int_0^t x(\tau) d\tau \\ e_1 &= D_1(s)[x] + g(z, e_0) \end{aligned} \tag{3.68}$$

where $D(s) = s^m + a_1 s^{m-1} + \ldots + a_m$ is a Hurwitz polynomial, $D_1(s) = D(s) - s^m$ and $g(z, e_0)$ is a bounding function with a buffer $\delta_f + \epsilon_0$ of $f(z, \theta)$ with respect to e_0. Then,

$$\begin{aligned} \dot{e}_0 &= e_1 + f(z, \theta) - g(z, e_0) \\ \dot{e}_1 &= \frac{\partial g}{\partial z} u + (a_1 + \frac{\partial g}{\partial e_0}) f(z, \theta) + D_2(s)[x] + \frac{\partial g}{\partial e_0} D_1(s)[x] \end{aligned} \tag{3.69}$$

with $D_2(s) = s(D_1(s) - a_1 s^{m-1})$. The adaptive controller is designed as

$$\begin{aligned} u &= \left(\frac{\partial g}{\partial z}\right)^{-1} \left[-e_0' - \gamma e_1' - D_2(s)[x] - \frac{\partial g}{\partial e_0} D_1(s)[x] - (a_1 + \frac{\partial g}{\partial e_0}) f(z, \hat{\theta}) - a^* S(e_1, \epsilon) \right], \quad \gamma > 0 \\ \dot{\hat{\theta}} &= e_1' \omega^* \end{aligned} \tag{3.70}$$

with a^* and ω^* adjusted according to the min-max algorithm. This leads to a time derivative of $V = \frac{1}{2}(e_0'^2 + e_1'^2)$ being nonpositive,

$$\dot{V} \leq -|e_0'| \delta_f - \gamma e_1'^2,$$

indicating global stability of the system. From the definition of e_0, if $|e_0| \to \epsilon_0$ implies that the state x and all its time derivatives $x^{(i)}$, $i = 1, \ldots, m$ are bounded as well. It should be noted that in order to be able to compute the control input u as in eq. (3.70), it is required that the inverse of $\frac{\partial g}{\partial z}$ always exist. Since g is a constructed feature of the controller, and not of the physical system, it can be designed in such a way that $\left(\frac{\partial g}{\partial z}\right)^{-1}$ always exists.

3.2.7 n second-order systems in chain form

We consider in this section, yet another class of systems in chain form. This includes a set of coupled nonlinear systems, each of which is a second-order system, and includes nonlinear parameterizations. Such a class of nonlinear systems can be described as

$$\begin{aligned} \ddot{x}_1 &= x_2 + f_1(x_1, \theta) \\ \ddot{x}_2 &= x_3 \\ &\vdots \\ \ddot{x}_n &= u \end{aligned} \tag{3.71}$$

where u, $x_i \in \mathbb{R}^p$, $i = 1, \ldots, n$, $\theta \in$ a compact set Θ in $\mathbb{R}^\ell$. The assumptions are that (A1) x_i, $\dot{x}_i$ are measurable for $i = 1$ to n, (A2) $f_1(x_1, \theta)$ is bounded for all bounded x_1 and $\theta \in \Theta$, and (A3) that $x_i = 0$, $i = 1, \ldots, n$ is an equilibrium point. The goal is to choose u so that both stabilization and tracking can be accomplished globally.

We begin with the case when $n = 2$ in Eq. (3.71). The following controller can be shown to lead to global stability, when θ is known.

$$e_0 = \dot{x}_1 + 2\Lambda x_1 \qquad \Lambda > 0 \tag{3.72}$$

$$e_1 = x_2 + 2\Lambda \dot{x}_1 + g_1(x_1, e_0) \tag{3.73}$$

$$e_2 = e_0 + \dot{x}_2 + k_1(x_1, x_2, \dot{x}_1, e_0) + \overline{h}_1(x_1, e_0, e_1) \tag{3.74}$$

$$k_1 = 2\Lambda x_2 + \frac{\partial g_1}{\partial x_1} \dot{x}_1 + \frac{\partial g_1}{\partial e_0}(x_2 + 2\Lambda \dot{x}_1)$$

$$h_1 = \left(2\Lambda + \frac{\partial g_1}{\partial e_0}\right) f_1$$

$$
\begin{aligned}
k_2 &= \left(\frac{\partial k_1}{\partial x_1} + \frac{\partial \overline{h}_1}{\partial x_1}\right)\dot{x}_1 + \frac{\partial k_1}{\partial x_2}\dot{x}_2 + \frac{\partial k_1}{\partial \dot{x}_1} x_2 \\
&\quad + \left(1 + \frac{\partial k_1}{\partial e_0} + \frac{\partial \overline{h}_1}{\partial e_0}\right)(x_2 + 2\Lambda\,\dot{x}_1) + \frac{\partial \overline{h}_1}{\partial e_1}(x_2 + k_1) \\
h_2 &= \left(\frac{\partial k_1}{\partial \dot{x}_1} + \frac{\partial k_1}{\partial e_0} + \frac{\partial \overline{h}_1}{\partial e_0} + 1\right) f_1 + \frac{\partial \overline{h}_1}{\partial e_1} h_1 \\
u &= -e_1' - \Gamma e_2' - k_2 - \overline{h}_2
\end{aligned} \tag{3.75}
$$

where Λ and Γ are diagonal, positive-definite matrices in $\mathbb{R}^{p\times p}$ and g_1, $\overline{h}_1$, $\overline{h}_2$ are bounding functions of f_1, h_1, and h_2 with respect to e_0, e_1, and e_2 with buffers δ_0, δ_1 and δ_2, respectively.

Theorem 3.3 *The system in (3.71) for $n = 2$ can be globally stabilized by the controller in Eqs. (3.73)-(3.75) and $|x_i|$ tend to $\frac{\epsilon_i}{2\lambda_i}$ as $t \to \infty$ for $i = 1, 2$, where $\lambda_i = \Lambda_{ii}$.*

Proof: From the choice of e_0, it follows that

$$
\dot{e}_0 = x_2 + f_1 + 2\Lambda\,\dot{x}_1
$$

which can be rewritten, using the definition of e_1 as

$$
\dot{e}_0 = e_1 + f_1 - g_1.
$$

Noting that

$$
\dot{e}_1 = \dot{x}_2 + k_1 + h_1
$$

a Lyapunov function candidate

$$
V_1 = \frac{1}{2}\left({e_0'}^T e_0' + {e_1'}^T e_1'\right)
$$

yields

$$\dot{V}_1 = e_0'^T(f_1 - g_1) + e_1'^T(e_0' + \dot{x}_2 + k_1 + h_1).$$

The choice of g_1 and $\overline{h}_1$ as bounding functions of f_1 and h_1 guarantees that

$$(f_1 - g_1)^T e_0' \leq \delta_0 ||e_0'|| \tag{3.76}$$

$$(h_1 - \overline{h}_1)^T e_1' \leq \delta_1 ||e_1'|| \tag{3.77}$$

$$(h_2 - \overline{h}_2)^T e_2' \leq \delta_2 ||e_2'|| \tag{3.78}$$

From Eqs. (3.76) and (3.74), it follows that

$$\dot{V}_1 \leq e_1'^T(e_2' + h_1 - \overline{h}_1) \leq e_1'^T e_2'. \tag{3.79}$$

Equation (3.79) suggests that V_1 needs to be updated as

$$V_2 = V_1 + \frac{1}{2} e_2'^T e_2'.$$

Noting that

$$\dot{e}_2 = u + k_2 + h_2$$

and using Eq. (3.75), it follows that

$$\begin{aligned} \dot{V}_2 &\leq e_1'^T e_2' + e_2'^T(-e_1' - \Gamma e_2') \\ &\leq -e_2'^T \Gamma e_2' \leq 0 \qquad \text{from Eq. (3.78)} \end{aligned}$$

which implies that $||e_i||$ tend to ϵ_i, $i = 0, 1, 2$. This in turn establishes the theorem. •

As an alternative to choosing $\overline{h}_3$ as dictated by the bounding function, one can estimate h_3 and still ensure global boundedness. This is shown below:

Theorem 3.4 *The adaptive controller given by eqs. (3.73)-(3.75) with g_1 and g_2*

chosen as bounding functions for f_1 and f_2, and

$$\begin{aligned} \overline{h}_3 &= \widehat{h}_2(x_1, \dot{x}_1, x_2, e_0, e_1, \widehat{\theta}) - a^* \\ \dot{\widehat{\theta}}_i &= -e'_{2_i}\omega_i^* \end{aligned}$$

where $\theta_i \in \mathbb{R}^m$ is the argument of f_{1_i}, the ith element of f_1, the ith element of a^* is $a_i^* S(e_{2_i}, \epsilon)$, e_{2_i} is the ith element of e_2, and $a_i^* \in \mathbb{R}$ and $\omega_i^* \in \mathbb{R}^m$ are chosen as min-max solutions of (3.43) for $f_n = f_3$, leads to global boundedness of the closed-loop adaptive system.

Proof: Proceeding in the same manner as above, we can show that for a scalar function

$$V = \frac{1}{2}\left({e'_0}^T e'_0 + {e'_1}^T e'_1 + {e'_2}^T e'_2 + \sum_{i=1}^{p} \widetilde{\theta}_i^T \widetilde{\theta}_i \right)$$

where $\widetilde{\theta}_i = \widehat{\theta}_i - \theta_i$, the time-derivative is of the form

$$\begin{aligned} \dot{V} &\leq -{e'_2}^T \Gamma e'_2 + {e'_2}^T (h_3 - \widehat{h}_3 - a^*) - \sum_{i=1}^{p} \widetilde{\theta}_i^T {e_{2_i}}' \omega_i \\ &= -{e'_2}^T \Gamma e'_2 + \sum_{i=1}^{p} {e_{2_i}}' \left(h_{3_i} - \widehat{h}_{3_i} + \widetilde{\theta}_i^T \omega_i - a_i^* S(e_{2_i}, \epsilon) \right) \leq 0 \end{aligned}$$

from Lemma 2. The definitions of e_0, e_1, e_2, and u imply that all signals are bounded. •

A similar extension of the controller in (3.73)-(3.75) can be carried out for the system in (3.71) for an arbitrary n using recursive formulations of errors e_i and bounding functions g_i.

3.2.8 Numerical example

Using the methodology discussed in the previous sections, we present here an adaptive controller for a particular class of low-velocity friction compensation systems. Friction compensation plays a important part in control of high-precision positioning systems.

As with any controller design, accurate models of the underlying physical system are required in order to improve performance. The model which we'll use incorporates two features: compliance among the elements of the system and an accurate low-velocity model of the friction force. The first feature allows for non-ideal transmission elements by assuming small strains, and thus modelling them as linear springs. The friction force model is based on evidence([5]) that, at low velocities, the friction force exhibits such behavior which can be best characterized by a nonlinearly-parameterized model. For this example, we employ the following simplified version of the model suggested in [5]

$$F = Ksgn(\dot{x})e^{-\left(\frac{\dot{x}}{v_s}\right)^2} \tag{3.80}$$

where $\dot{x}$ represents the relative velocities of the bodies in contact, v_s is the Stribeck parameter, and K is the friction coefficient. Thus, we are interested in the control of the following system

$$\begin{aligned} \ddot{x}_1 &= k_{x_1}(x_2 - x_1) - K_1 sgn(\dot{x}_1)e^{-\left(\frac{\dot{x}_1}{v_s}\right)^2} \\ \ddot{x}_2 &= u + k_{x_1}(x_1 - x_2) \end{aligned} \tag{3.81}$$

For the sake of simplicity, we take the linear coefficients k_{x_1} and K_1 as unity. Furthermore, the exact value of the Stribeck parameter v_s is unknown, but it is assumed to lie in the range of $[v_{s_{min}}, v_{s_{max}}]$. The goal was to develop an adaptive controller that can track a trajectory specified by $x_{1_d} = 10\ sin(t)$ in the face of pronounced nonlinear effects.

First, we put the above system into the following form:

$$\begin{aligned} \dot{z}_1 &= z_2 \\ \dot{z}_2 &= z_3 - z_1 - sgn(\dot{z}_2)e^{-(\theta_1 \dot{x}_1)^2} \\ \dot{z}_3 &= z_4 \\ \dot{z}_4 &= u + (z_1 - z_3) \end{aligned} \tag{3.82}$$

Then, using the strategies presented in sections 3.2.7 and 3.2.3 the following errors are defined

$$\begin{aligned}
e_0 &= z_1 - x_{1_d} \\
e_1 &= e_0 + z_2 - \dot{x}_{1_d} \qquad (3.83) \\
e_2 &= e_1 + z_3 - z_1 - \ddot{x}_{1_d} + g_1(z_2, e_1) \\
e_3 &= e_2 + z_4 - x_{1_d}^{(3)} - g_1(z_2, e_1) - \dot{x}_{1_d} + \frac{\partial g_1}{\partial z_2}(z_3 - z_1) + \frac{\partial g_1}{\partial e_1}(e_2 - e_0 - g_1) + g_2(z_2, e_1, e_2)
\end{aligned}$$

The bounding functions g_1 and g_2 are constructed as

$$g_1 = S(e_1, \epsilon_1)\left(e^{-v_{s_{min}} z_2^2} + \delta_{f_1}\right) \qquad g_2 = S(e_2, \epsilon_2)\left(e^{-v_{s_{min}} z_2^2}\,|\psi| + \delta_{f_2}\right) \qquad (3.84)$$

where $S(.)$ is an odd-powered fifth-order polynomial that satisfies the smoothing function conditions given in Definition3.4, and $\psi = 1 + \frac{\partial g_1}{\partial z_2} + \frac{\partial g_1}{\partial e_1}$.

The values of all the dead-zone coefficients ϵ_i, $i = 0, 3$ are set to unity, $\delta_{f_1} = \delta_{f_2} = 2$, and it is assumed that $v_s \in [0, 200]$. The adaptive controller is then chosen as

$$\begin{aligned}
u &= -e_2 - e_3 - z_4 - \frac{\partial g_1}{\partial z_2}(z_4 - z_2) - y_1^T\, K_x + \hat{f}_1 - a^* S(e_3, \epsilon_3) \\
&\quad + x_{1_d}^{(4)} + x_{1_d}^{(3)} + 2\ddot{x}_{1_d} + \dot{x}_{1_d} \\
\dot{\hat{v}}_s &= -e_3'\, \omega_1 \qquad (3.85)
\end{aligned}$$

where

$$
\begin{aligned}
y &= \begin{bmatrix} -\dfrac{\partial g_1}{\partial e_1} \\ \dfrac{\partial^2 g_1}{\partial e_1^2}(e_2 - e_0 - g_1) - \left(\dfrac{\partial g_1}{\partial e_1}\right)^2 + \dfrac{\partial g_1}{\partial z_2}(z_3 - z_1) + \sigma(\psi) g_2 \left(\dfrac{\partial^2 g_1}{\partial e_1^2} + \dfrac{\partial g_1}{\partial e_1})\right) \\ \dfrac{\partial g_1}{\partial e_1} + |\psi| \dfrac{\partial g_2}{\partial e_2} \\ \dfrac{\partial g_1}{\partial e_1}(e_2 - e_0 - g - 1) + \dfrac{\partial g_1}{\partial e_1}\dfrac{\partial g_1}{\partial z_2} + \dfrac{\partial^2 g_1}{\partial z_2^2}(z_3 - z_1) + \dfrac{\partial g_2}{\partial e_1} + \dfrac{\partial g_2}{\partial z_2} \end{bmatrix} \\
K_x &= \begin{bmatrix} -z_2 \\ e_2 - e_0 - g_1 \\ e_3 - e_1 - g_2 \\ z_3 - z_1 \end{bmatrix} \\
\widehat{f} &= \sigma(z_2) \left(1 + y^T \begin{bmatrix} 0 \\ 1 \\ \psi \\ 1 \end{bmatrix} \right) e^{-\widehat{v}_s z_2^2}
\end{aligned} \tag{3.86}
$$

and a^*, ω_1 are obtained through the min-max algorithm. The above controller and system was numerically simulated and the resulting positions are shown in Figure 3-1, indicating satisfactory convergence of the system to the desired trajectory.

3.3 Adaptive Control of Systems in Triangular Form

In the previous section, we considered systems in chain form, given by Eq. (3.4). In order to extend the controller presented in section 3.2.3 to the triangular system in Eq. (3.3), one has to ensure that the presence of the nonlinearities γ_i and the presence of multiple unknown parameters θ_i do not introduce any insurmountable difficulties. This is indeed the case. A stabilizing controller with a few additional terms in the functions k_i, h_i, and $\widehat{f}_n$ can be constructed, and is shown below:

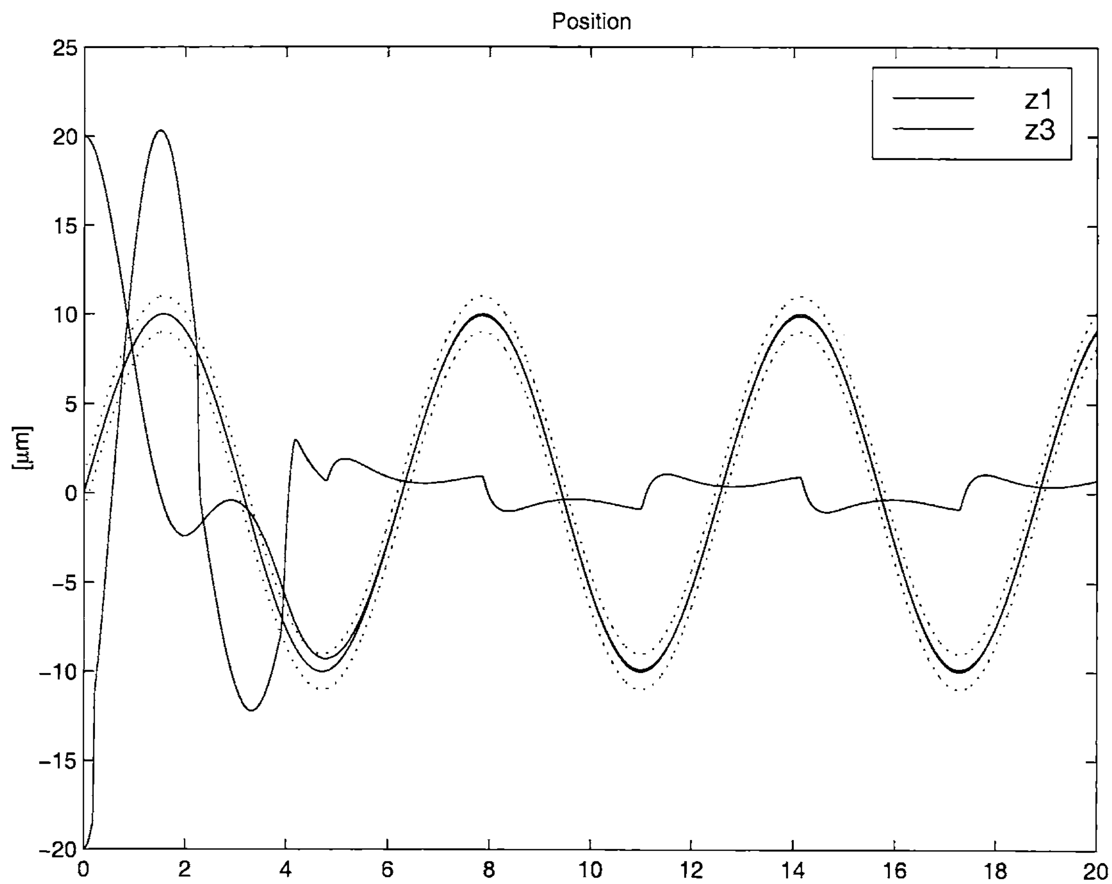

Figure 3-1: Time behavior of the position of two system elements, z_1 and z_3. The ϵ bounds of guaranteed convergence to the desired trajectory are given by dotted lines.

As before, the idea is to construct errors e_i, $i = 0, \ldots, n-1$, which are such that if e_i tends to zero, it guarantees that e_j, $j = 0, \ldots, i-1$ tend to zero. These errors are constructed as follows:

$$\begin{aligned} e_0 &= x_1 \qquad\qquad (3.87) \\ e_i &= e_{i-2} + \Gamma_{i-1}(x_2, \ldots, x_i)\gamma_i(x_{i+1}) + g_i(x_1, \ldots, x_i), \qquad i = 1, \ldots, n-1 \end{aligned}$$

where

$$\begin{aligned} g_i &= k_{i-1} + \overline{h}_{i-1} \\ k_i &= k_{i-2} + \sum_{j=1}^{i} \left(\frac{\partial k_{i-1}}{\partial x_j} + \frac{\partial \overline{h}_{i-1}}{\partial x_j} \right) \gamma_j(x_{j+1}) + \Gamma_{i-2}\gamma_{i-1} + \sum_{j=2}^{i} \frac{\partial \Gamma_{i-1}}{\partial x_j} \gamma_j \gamma_i \\ h_i(x_1, \ldots, x_{i+1}, \theta_1, \ldots, \theta_{i+1}) &= h_{i-2} + \sum_{j=1}^{i} \left(\frac{\partial k_{i-1}}{\partial x_j} + \frac{\partial \overline{h}_{i-1}}{\partial x_j} \right) f_j(x_1, \ldots, x_j, \theta_k) \\ & \sum_{j=2}^{i} \frac{\partial \Gamma_{i-1}}{\partial x_j} f_j(x_1, \ldots, x_j, \theta_k)\gamma_i + \Gamma_i f_{i+1}(x_1, \ldots, x_{i+1}, \theta_{i+1}) \qquad (3.88) \end{aligned}$$

for $i = 2, \ldots, n-1$,

$$\Gamma_i = \Gamma_{i-1} \frac{\partial \gamma_i}{\partial x_{i+1}}, \qquad i = 1, \ldots, n-1 \qquad (3.89)$$

with

$$\begin{aligned} k_0 &= 0; \qquad k_1 = \frac{\partial \overline{h}_0}{\partial x_1} \\ h_{-1} &= 0; \qquad h_0 = f_1(x_1, \theta); \\ \Gamma_0 &= 1, \qquad e_{-1} = 0, \end{aligned}$$

and $\overline{h}_i(x_1, \ldots, x_{i+1})$ as bounding functions of h_i with respect to e_i, so that

$$(h_i - \overline{h}_i)\sigma(e_i) \leq 0 \qquad i = 0, \ldots, n-1.$$

The errors in (3.88) can be shown to satisfy the differential equations

$$\dot{e}_{i+1} = e_{i+1} - e_{i-1} + \overline{f}_{i+1} - g_{i+1} \qquad i = 0, \ldots, n-1, \tag{3.90}$$

where $\gamma_n = u$, and

$$\overline{f}_i = k_{i-1} + h_{i-1}, \qquad i = 1, \ldots, n.$$

This suggests that the function V in (3.30) has a time-derivative

$$\dot{V} \;=\; \sum_{i=0}^{n-2} e_i(e_{i+1} - e_{i-1}) + \sum_{i=0}^{n-2} e_i(\overline{f}_{i+1} - g_{i+1}) + e_{n-1}(\Gamma_{n-1}u + \overline{f}_n) \tag{3.91}$$

From the definition of e_{-1}, the choice of $\overline{h}_i$, and a choice of u as

$$u \;=\; \Gamma_{n-1}^{-1}(-e_{n-2} - ke_{n-1} - g_n) \qquad k > 0 \tag{3.92}$$

it follows that V is a Lyapunov function.

An adaptive version of the controller is simply given by

$$u \;=\; \Gamma_{n-1}^{-1}(-e_{n-2} - ke_{n-1} - \widehat{f}_n - a^*\sigma(e_{n-1})) \qquad k > 0 \tag{3.93}$$

$$\begin{aligned} \widehat{f}_n &= k_{n-1} + \widehat{h}_{n-1}(x_1, \ldots, x_n, \widehat{\theta}_1, \ldots, \widehat{\theta}_n) \\ h_{n-1} &= \sum_{j=1}^{n} h_{jn}(x, \theta_j) \\ \widehat{h}_{n-1} &= \sum_{j=1}^{n} h_{jn}(x, \widehat{\theta}_j) \\ \dot{\widehat{\theta}}_i &= e_{n-1}\omega_i^* \\ a^* &= \sum_{j=1}^{n} a_j^* \end{aligned} \tag{3.94}$$

$$(a_j^*, \omega_j^*) \;=\; \min_{\omega_j} \max_{\theta} \left[h_{jn}(x, \theta_j) - h_{jn}(x, \widehat{\theta}_j) + (\widehat{\theta}_j - \theta_j)\omega_j \right] \tag{3.95}$$

As mentioned earlier, the closed-form solutions to (3.95) can be found along the lines outlined in [2, 34], with the solutions being considerably simpler when h_{jn} is convex/concave with respect to θ_j. In addition, as outlined in Section 3.2.3, a smooth version of the controller in (3.93) can be obtained as

$$u = \Gamma_{n-1}^{-1}(-e_{n-2}' - ke_{n-1} - \widehat{f}_n - a^* S(e_{n-1}, \epsilon)) \qquad k > 0$$

by appropriately modifying the functions k_i and h_i in (3.88).

3.4 Concluding Remarks and Future Work

The stabilizing controller proposed here requires the availability of two functions a^* and ω^*. These in turn imply that closed-form solutions of (3.43) are needed. These can be constructed in a simple manner, as outlined in [2], when f_n is a convex/concave function of θ. Convexity/concavity of the underlying nonlinearity has also been exploited in [14, 41]. The computational burden increases when f_n is a general function, and is discussed in [34]. Special classes of functions f_i which can be reparameterized so as to result in concavity (or convexity) are described in [39]. For all such functions, the controller proposed above results in global boundedness.

The proof of theorem 3.1 and the preceding discussions also demonstrate that stabilization of systems in chain form can be accomplished without adapting to the parameter θ, as was also demonstrated in [48]. Instead of estimating f_n as $\widehat{f}_n$, one could simply construct yet another bounding function and stabilize (3.4). However, the advantage of using $\widehat{f}_n$ is that it enables the unknown parameter to be estimated in addition to stabilization. Once such a stable framework is generated for parameter estimation, conditions related to persistent excitation can be invoked to obtain parameter convergence. In Chapter 2, it has been shown that for a class of error models of the form

$$\dot{e} = -e + f(\phi, \theta) - f(\phi, \widehat{\theta}) - a^* S(e, \epsilon) \tag{3.96}$$

$$\dot{\widehat{\theta}} = e' \omega^* \tag{3.97}$$

conditions of persistent excitation of ϕ with respect to f for nonlinearly parameterized systems can be derived so as to result in the convergence of the parameter $\widehat{\theta}$ to θ to within a desired precision ϵ. It is worth noting that the error e_n satisfies a differential equation that is quite similar in form to that of (3.96). Hence, an extension of the derived nonlinear persistent excitation results to the case of parameter convergence using the controller presented should be quite feasible.

The stability result in Theorem 3.1 can be viewed as an extension of the parametric-strict-feedback systems considered in [24] to the case when the unknown parameters occur nonlinearly. Examples of such systems abound in several applications [2, 3, 39, 12]. In contrast to the back-stepping approach suggested in [24], we use a Bounding Function to generate the errors e_i in the system. We note that in contrast to linear adaptive control where parameter adaptation is proposed at each of the ith level of back-stepping [24, 46], we determine the adaptive laws at the final step so as to generate a Lyapunov function. This enables us to achieve a stabilizing adaptive controller without using over-parameterization of the controller.

We now present some possible avenues for further research. As was the case in deriving the general triangular structure results, we present the preliminary ideas on a much simplified second-order system of the form

$$\begin{aligned} \dot{x}_1 &= x_2 + f(x_1, \theta) \\ \dot{x}_2 &= u \end{aligned} \tag{3.98}$$

The hope is that the ideas presented for this case can be extended to the general n^{th} order system. Since for the results presented above, this extension was not a trivial task, we anticipate that there will be a number of issues to be resolved.

The first topic is to attempt to extend the results above to the case when only the output x_1 is available for measurement. Since the min-max algorithm of [2] requires that the time-varying part of the nonlinear function f be measurable, the time-varying part of the nonlinearity f must depend only on the output, just as in the system considered in [36]. Let

$$e = \widehat{x_2} + F(x_1) \tag{3.99}$$

where $F(x_1)$ is the bounding function, and $\widehat{x_2}$ is our estimate of x_2. Then, the system dynamics can be written as

$$\begin{aligned} \dot{x}_1 &= e - \widetilde{x_2} - F + f \\ \dot{e} &= \dot{\widehat{x_2}} + \frac{d\,F}{d\,x_1}(\widehat{x_2} - \widetilde{x_2} + f) \end{aligned} \tag{3.100}$$

Letting

$$\begin{aligned} u &= x_1 - e - \frac{d\,F}{d\,x_1}\widehat{x_2} + u_f \\ \dot{\widehat{x_2}} &= u + x_1 + \frac{d\,F}{d\,x_1}e \end{aligned} \tag{3.101}$$

where u_f comes from the min-max algorithm ensures that $V = x_1^2 + e^2 + \widetilde{x_2}^2$ is a Lyapunov function, and hence the stability and boundedness of the system. The main reason this output-feedback was possible is because the dynamics of the unknown state, x_2 are entirely driven by u, and thus known. It is questionable whether this simplistic approach can work in higher-order systems when the dynamics driving the unobservable states depend on other unobservable states, and thus are not completely known.

The second topic deals with the issue of the magnitude of the control effort required. From the analysis of the proposed controller and simulation experience, it can be observed that the magnitude of the control effort is related to the magnitude of the bounding function. The magnitude of the bounding function F depends on how much information is available about the parametric uncertainty f. In the current controller design procedure, the bounding function F does not depend on the estimated parameter value $\widehat{\theta}$. Hence, the magnitude of the bounding function F is determined a-priori during the design phase, and does not incorporate the on-line information gathered during the adaptation process. It should be noted that the magnitude of u does benefit from the adaptation process, as can be seen in the fact that u does depend on the on-line estimates of the parametric uncertainties. However, F does not depend on these on-line estimates, and indirectly, some control effort might be needlessly exerted.

We propose that

$$e = x_2 + \widehat{F} \tag{3.102}$$

Using (3.102) to substitute for x_2 in (3.98), and differentiating (3.102) we obtain that

$$\begin{aligned} \dot{x}_1 &= e - \widehat{F} + f \\ \dot{e} &= u + \frac{\partial \widehat{F}}{\partial x_1}(x_2 + f) + \frac{\partial \widehat{F}}{\partial x_1}\dot{\widehat{\theta}} \end{aligned} \tag{3.103}$$

Let

$$\begin{aligned} u &= u_f - x1 - e\frac{\partial \widehat{F}}{\partial x_1}\dot{\widehat{\theta}} - \frac{\partial \widehat{F}}{\partial x_1}x_2 \\ \dot{\widehat{\theta}} &= -x_1\omega_1 - e\omega_2 \end{aligned} \tag{3.104}$$

with u_f obtained from the min-max algorithm, and ω_1, ω_2 are to be specified. We note that the first term in the product $f_2(x_1, \widehat{\theta}, \theta) = \dfrac{\partial \widehat{F}}{\partial x_1} f$ depends only on measurable time-varying quantities and not in θ. Thus, the function f_2 inherits the convexity/concavity property from f.

Taking $V = x_1^2 + e^2 + \widetilde{\theta}^2$, the derivative is obtained as

$$\dot{V} = -e^2 + e(u_f - \frac{\partial \widehat{F}}{\partial x_1}f - \widetilde{\theta}\omega_2) + x_1(f - \widehat{F} - \widetilde{\theta}\omega_1) \tag{3.105}$$

If ω_2 is chosen according to the min-max algorithm, the second term in (3.105) is non-positive. To make the last term in (3.105) non-negative, we adopt the following strategy. First, from the min-max algorithm we observe that, based on the values of other signals, ω_1 can have the value of the local or the global gradient of $\widehat{F}$ with respect to $\widehat{\theta}$. In the case ω_1 has the value of the local gradient, the third term in (3.105) is non-positive. Otherwise, we set $\omega_1 = 0$, and let $\widehat{F}$ be specified as the standard bounding function described in Section 3.2.1. That suffices to make the third term of (3.105) non-positive. Thus, $\dot{V} \leq -e^2$ and stability is assured.

We note that in this design $\widehat{F}$ is chosen as either the bounding function or the estimated value of the unknown function. In either case, stability is ensured, and the use of $\widehat{F}$ as the estimate does not destroy the convexity/concavity property of f. In the dynamics of e it introduces an additional term, $\dfrac{\partial \widehat{F}}{\partial x_1} \dot{\hat{\theta}}$. However, this is a known term. Hence, this approach could possibly be extended to the full n-dimensional system. The approach does not require over-parameterization, and since $\widehat{F}$ is only partially a bounding function, fully exploits the properties of the min-max estimator.

Chapter 4

Convergence conditions for parameter identification with the gradient algorithm in nonlinearly parameterized systems

4.1 Introduction

In general, stability analysis of parameter identification in nonlinearly parameterized systems is currently a difficult task. A higher degree of difficulty stems from the fact that if the well-known techniques for linear systems [38, 45] are applied, only local results can be achieved. Thus, new methods and techniques are needed. Presently, very few results [2, 34] exist which offer new techniques and deal with the problem of parameter identification. The results in [2, 34] give global stability of the identification process by employing a modified version of the widely-used gradient algorithm. For such a modified algorithm, parameter convergence conditions were derived in [28]. Other available results [3, 12, 14, 22, 32, 29, 39, 41] primarily deal with the task of control of nonlinearly parameterized systems, and not with the problems of identifying the values of the nonlinear parameters. However, identifying the parameter values can also be a valuable resource in system analysis, control and monitoring, and in expanding the applicability of adaptive control to new types of systems. For example, NLP systems are abundant in biological models (see [7, 30]). Another important type of NLP systems are the various types of neural network architectures, which are inherently nonlinear. Clearly, to accomplish the task of accurately identifying parameters in these various types of NLP systems, new techniques for the study of the identification algorithms are needed.

In this Chapter, we examine the gradient algorithm in more detail. We develop new techniques and requirements for parameter convergence for showing global stability and convergence. It is shown, using the developed techniques, what conditions need to be satisfied so that the gradient algorithm can lead to convergence in a wide class of nonlinearly parameterized systems. Such conditions are placed on the input, and the exact type of conditions depends on the type of nonlinear parameterization present in the system. Current work is focused on showing that the parameter identification conditions are generally satisfied in a class of neural-network systems, which inherently are nonlinearly parameterized systems. The main results of the Chapter are presented in Section 4.2. Concluding remarks and future work directions are presented in 4.3.

4.2 Main results

We consider the following system

$$y(u,\theta^*) = \sum_{i=1}^{N} g(u,\theta_i^*) = h(u,\theta^*) \tag{4.1}$$

where $u, y : \mathbb{R} \to \mathbb{R}$, $\theta_i^* \in \mathbb{R}$.

Assumption 4.1 *We assume that the function $h(u,\theta)$ is differentiable and the magnitudes of the first derivatives $\nabla_\theta h(u,\theta)$ are bounded.*

Due to the structure of the system, it follows that without loss of generality we can assume that $\theta^* \in \Omega_\theta$, where

$$\Omega_\theta = \{\theta \mid \theta = [\theta_1, \ldots, \theta_N], \quad \theta_1 < \theta_2 < \theta_3 < \ldots \theta_N\} \tag{4.2}$$

The goal is to derive and algorithm and establish conditions under which it is possible to identify $\theta^* = [\theta_1^*, \ldots, \theta_N^*]$. To achieve that result, we examine the gradient-based algorithm given by:

$$\begin{aligned} \widehat{y}(u, \widehat{\theta}) &= \sum_{i=1}^{N} g(u, \widehat{\theta}_i) = h(u, \widehat{\theta}) \qquad e(u, \widehat{\theta}) = \widehat{y}(u, \widehat{\theta}) - y(u, \theta^*) \\ \dot{\widehat{\theta}} &= -e \, \nabla_\theta h(u, \theta) \end{aligned} \tag{4.3}$$

where $\widehat{\theta} = \left[\widehat{\theta}_1, \ldots, \widehat{\theta}_N\right]$ is our estimate of θ^*. Since for a given problem θ^* is fixed, e is thus considered only a function of u and $\widehat{\theta}$.

Let $\nu(\theta)$ be vector which is orthogonal to some hyper-plane $\theta_i = \theta_j$, $i \neq j$, where θ_i are the components of θ. From (4.3), it follows that $\dot{\widehat{\theta}}^T (u, \theta)\nu(\theta) = 0$ for any u. From (4.2) and the definition of ν we have that ν is orthogonal to the boundary of the set Ω_θ, and thus if $\widehat{\theta}(t_a) \in \Omega_\theta$, for some t_a, it follows that $\widehat{\theta}(t) \in \Omega_\theta$ for all $t > t_a$. Thus, we can, without loss of generality, restrict our analysis to the case when $\theta \in \Omega_\theta$, and assume that $\theta^*, \widehat{\theta} \in \Omega_\theta$.

In what follows, we use the notation $\lambda(u, \theta)$ to denote the gradient of h with respect to θ, i.e.

$$\lambda(u, \theta) = \nabla_\theta h(u, \theta). \tag{4.4}$$

We restrict our treatment of the problem to the case when $h(u, \theta)$ is monotonically parameterized in θ. Specifically, we require that the following assumption holds.

Assumption 4.2 *We assume that* $\lambda(u_1, \theta)\lambda(u_2, \theta) \geq 0$ *for any* u_1,u_2 *and* θ.

The δ-neighborhood of a set A is denoted as $N_\delta(\theta_a)$. That is,

$$N_\delta(A) = \{ \ \theta \mid \exists \, \theta_a \in A, \ \ ||\theta - \theta_a|| \leq \delta \ \} \tag{4.5}$$

Definition 4.1 *A basis B for a vector space A is a set of linearly independent vectors such that each vector $v \in A$ can be uniquely expressed as a linear combination of vectors in B. This relation between A and B is expressed by the operator $\mathcal{L}(\cdot)$ as*

$$B = \mathcal{L}\{A\}$$

Definition 4.2 *Let A be a set whose elements are vectors in $\mathbb{R}^N$, and let*

$$\mathcal{N}\{A\} = \left\{ b \mid b^T a = 0, \quad a \in A \right\}. \tag{4.6}$$

If $B = \mathcal{N}\{A\}$, then B is orthogonal to A.

In what follows, we utilize the standard concepts of a manifold and a curve in $\mathbb{R}^N$ (for example, see [23]). It is assumed that a curve Q in $\mathbb{R}^N$ permits a scalar parameterization so that $Q : \mathbb{R} \to \mathbb{R}^N$.

Definition 4.3 *Let L be a manifold in $\mathbb{R}^N$, and let $\theta \in L$. Let Q be a curve on L, and let $q(\theta)$ be the tangent vector to Q at θ. Let $\Sigma(\theta)$ be the set of tangent vectors $q(\theta)$ for any curve Q which contains the point θ. Then,* the tangent plane $\Pi(L, \theta)$ *is defined as*

$$\Pi(L, \theta) = span\left\{\Sigma(\theta)\right\}.$$

If $\nu(\theta)$ is a gradient vector to L at the point θ, then

$$\nu^T(\theta)\eta(\theta) = 0, \qquad \eta(\theta) \in \Pi(L, \theta). \tag{4.7}$$

Definition 4.4 *Let $T_u > 0$, and let $\lambda(u, \theta) = \nabla_\theta h(u, \theta)$, with $h(u, \theta) : \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}$.*

Let

$$
\begin{aligned}
\Omega_t &= [t_0, t_0 + T_u], \qquad\qquad T_u > 0 \\
\Psi &= \{t_i \in \Omega_t, \quad i = 1, \ldots, M \quad | \quad |t_{i+1} - t_i| > \varepsilon_0\}, \qquad\qquad M, \varepsilon_u > 0 \qquad (4.8) \\
\Lambda(\Psi, \theta) &= \{\lambda(x(t_k), \theta) \quad | \quad t_k \in \Psi\}
\end{aligned}
$$

A function $u(t) : \mathbb{R} \to \mathbb{R}$ is said to belong to the class U_{PE}^N on the interval $t \in [t_0, t_0 + T_u]$ if it satisfies the two properties defined as follows:

(P1) linear independence is invariant*: If the set $\Lambda(\Psi_a, \theta_a)$ is linearly independent for some set $\Psi_a \in \Omega_t$ and $\theta_a \in \Omega_\theta$, then $\Lambda(\Psi_a, \theta)$ is linearly independent for all $\theta \in \Omega_\theta$.*

(P2) sufficient degree of excitation exists*: There exists a set $\Psi_b \in \Omega_t$ consisting of N elements such that $\Lambda(\Psi_b, \theta_a)$ is linearly independent.*

Definition 4.4 implies Lemmas 4.1,4.2 and Corollaries 4.1 -4.3. For ease of exposition, we denote $\lambda(u(t_i), \theta)$ as $\lambda_i(\theta)$.

Lemma 4.1 *Let $u(t) \in U_{PE}^N$ over some Ω_t. Suppose that there exists a point θ_a, a set $\Psi_a \in \Omega_t$, defined as in (4.8), and a time instant $t_b \in \Omega_t$, such that $\lambda_b(\theta_a) \notin span\{\Lambda(\Psi_a, \theta_a)\}$. Then, $\lambda_b(\theta) \notin span\{\Lambda(\Psi_a, \theta)\}$ for all θ.*

Proof. Let

$$
\begin{aligned}
\Psi_{al} &= \{ \; t_j \; | \; \lambda_j(\theta_a) \in \mathcal{L}\{\Lambda(\Psi_a, \theta_a) \; \} \\
W_a(\theta) &= span\,\{\Lambda(\Psi_a, \theta)\}, \qquad W_{al}(\theta) = span\,\{\Lambda(\Psi_{al}, \theta)\} \qquad (4.9)
\end{aligned}
$$

We note that Ψ_a is distinct from Ψ_{al} if any of the $\lambda's$ in $\Lambda(\Psi_a, \theta_a)$ are linearly dependent. First, we will show that by property (P1) we have that

$$W_a(\theta) = W_{al}(\theta) \qquad (4.10)$$

Clearly, (4.9) and Definition 4.1 imply that (4.10) holds for $\theta = \theta_a$. Let $t_a \in \Psi_a \setminus \Psi_{al}$. Thus, $\lambda_a(\theta) \in W_a(\theta)$ for all θ. Since (4.10) holds for $\theta = \theta_a$, we also have that $\lambda_a(\theta_a) \in W_{al}(\theta_a)$. Suppose that there exists a θ_b such that $\lambda_a(\theta_b) \notin W_{al}(\theta_b)$, implying that $W_{al}(\theta_b) \neq W_a(\theta_b)$. Letting $\Psi_b = \Psi_{al} \bigcup t_a$, this implies that $\Lambda(\Psi_b, \theta_b)$ is linearly independent. But, we have already shown that $\Lambda(\Psi_b, \theta_a)$ is linearly dependent. This contradicts (P1), and hence (4.10) holds for all θ. To establish Lemma 4.1, it now suffices to show that if $\lambda_b(\theta_a) \notin W_{al}(\theta_a)$, then $\lambda_b(\theta) \notin W_{al}(\theta)$ for all θ. By letting $\Psi_c = \Psi_{al} \bigcup t_b$, the supposition in Lemma 4.1 implies that $\Lambda(\Psi_c, \theta_a)$ is linearly independent. If there existed a θ_c such that $\lambda_b(\theta_c) \in W_{al}(\theta_c)$, it would imply that $\Lambda(\Psi_c, \theta_c)$ is linearly dependent. This contradicts (P1), and hence such a θ_c does not exist. This establishes the Lemma. •

The following corollaries can be derived using Lemma 4.1. The first corollary is the converse of Lemma 4.1, while the rest are direct consequences of that Lemma.

Corollary 4.1 *Let $u(t) \in U_{PE}^N$ over some Ω_t. Suppose there exists a point θ_a, a set $\Psi_a \in \Omega_t$, defined as in (4.8), and a time instant $t_b \in \Omega_t$, such that $\lambda_b(\theta_a) \in span\{\Lambda(\Psi_a, \theta_a)\}$. Then, $\lambda_b(\theta) \in span\{\Lambda(\Psi_a, \theta)\}$ for all θ.*

Proof. Let θ_b be such that $\lambda_b(\theta_b) \notin span\{\Lambda(\Psi_a, \theta_b)\}$. Then, Lemma 4.1 implies that θ_a as in the Corollary statement cannot exist. Hence, this is a contradiction, and θ_b in our initial assumption cannot exist. This establishes Corollary 4.1. •

Throughout the rest of the paper, we use $\dim\{A\}$ to denote the number of elements of a set A.

Corollary 4.2 *Let $V(\theta) = span\{\Lambda(\Psi, \theta)\}$ for any θ and for some Ψ such that $\dim\{\mathcal{L}\{V(\theta)\}\} = i$, $i < N$. Let $W(\theta) = \mathcal{N}\{V(\theta)\}$. If $u \in U_{PE}^N$, and if there exist a time instant t_a, point θ_a, and constant $\varepsilon_a > 0$ such that*

$$|\lambda_a^T(\theta_a)\, w(\theta_a)| \geq \varepsilon_a, \qquad w(\theta_a) \in W(\theta_a), \tag{4.11}$$

then for any θ, there exist an $\varepsilon_1 > 0$ and a $w(\theta) \in W(\theta)$ such that

$$|\lambda_a^T(\theta)\, w(\theta)| \geq \varepsilon_1 \tag{4.12}$$

Proof. For any θ, $V(\theta)$ is an i-dimensional subspace, and $i < N$. Hence it follows that $W(\theta)$ is nonempty. Since (4.11) holds, and since W is orthogonal to V, it follows that

$$\lambda_a(\theta_a) \notin span\{\Lambda(\Psi, \theta_a)\} \tag{4.13}$$

By Lemma 4.1, we have that (4.13) holds for all θ. Suppose that at some point θ_b, there does not exist an ε_1 such that (4.12) holds. This implies that $|\lambda_a^T(\theta_b)\, w(\theta)| = 0$, $w(\theta) \in W(\theta)$. Since $W(\theta)$ is the collection of all vectors which are orthogonal to $V(\theta)$, this implies that at θ_b

$$\lambda_a(\theta_b) \in span\{\Lambda(\Psi, \theta_b)\} \tag{4.14}$$

This contradicts Lemma 4.1, and hence such a point θ_b does not exist. •

Corollary 4.3 *If there exists an ε_a and a point θ_a such that*

$$\left|\lambda^T(u_1, \theta_a)\lambda(u_2, \theta_a)\right| \geq \varepsilon_a \tag{4.15}$$

for some u_1, u_2, then there exists an ε_1 such that

$$\left|\lambda^T(u_1, \theta)\lambda(u_2, \theta)\right| \geq \varepsilon_1 \tag{4.16}$$

for all θ.

Proof. Let $L = \{\ \theta\ \mid\ e(u_1, \theta_a) = c_1\}$, for some value of c_1. Suppose that for some Ψ,

$$\Pi(L, \theta_a) = span\{\Lambda(\Psi, \theta_a)\}, \tag{4.17}$$

where $\Pi(L, \theta_a)$ is the tangent plane as defined in Definition 4.3. In the rest of the proof, we will use $\Pi(\theta)$ to denote $\Pi(L, \theta)$.

Since $\lambda(u_1, \theta)$ is the gradient L at θ, by definition it is orthogonal to any vector in $\Pi(\theta)$. Hence,

$$\lambda(u_1, \theta) \in \mathcal{N}\{\Pi(\theta)\} \tag{4.18}$$

Since L is an $N-1$ dimensional manifold, it follows that $\Pi(\theta)$ is an $N-1$ dimensional manifold. Thus,

$$\left|\lambda^T(u_1, \theta) w(\theta)\right| = |\lambda(u_1, \theta)|\,|w(\theta)|, \qquad w(\theta) \in \mathcal{N}\{\Pi(\theta)\} \tag{4.19}$$

implying that $\lambda(u_1, \theta)$ and any $w(\theta) \in \mathcal{N}\{\Pi(\theta)\}$ are collinear. Since (4.15) and (4.18) hold, Corollary 4.2 implies that there exists an ε_b and a $w(\theta) \in \mathcal{N}\{\Pi(\theta)\}$ such that

$$\left|\lambda^T(u_2, \theta) w(\theta)\right| \geq \varepsilon_b \tag{4.20}$$

Since $\lambda(u_1, \theta)$ and $w(\theta)$ are collinear, (4.20) establishes the Corollary. •

Corollary 4.3 states that if $\lambda(u_2, \theta_a)$ is not in the tangent plane of the surface $e(u_1, \theta) = c_1$ at the point θ_a, for some constant c_1, then it is not in the tangent plane of $e(u_1, \theta) = c_2$ for any θ and c_2.

Lemma 4.2 *If $u(t) \in U^N_{PE}$ over some Ω_t, then there exist an $\varepsilon_1 > 0$ and a $t_1 \in \Omega_t$ such that for any unit vector $w \in \mathbb{R}^N$ and any $\theta \in \Omega_\theta$, the following holds*

$$\left|\, \lambda^T(u(t_1), \theta)\, w \,\right| \geq \varepsilon_0 \tag{4.21}$$

Proof. Any set of N linearly independent vectors spans an N-dimensional space (see [15]). Property (P2) implies that there exists a set Λ consisting of N linearly independent vectors. Thus, $span\{\Lambda\} = \mathbb{R}^N$, and hence, for every unit vector $w \in \mathbb{R}^N$, there exists a $t_1 \in \Omega_t$ for which (4.21) holds.

Definition 4.5 *Let $q(a) = l(a) \otimes \{\eta_1(a), \eta_2(a), \ldots, \eta_n(a)\}$ denote the orthogonal projection of a vector l at a point* a *onto the surface whose tangent plane at* a *is defined* *by normals $\{\eta_1(a),\ldots,\eta_i(a)\}$. The orthogonal projection is defined as*

$$
\begin{aligned}
q(a) &= l(a) - \sum_{j=1}^{n} \frac{l^T \nu_j}{||\nu_j||^2}\nu_j, \qquad \textit{where} \\
\nu_j &= \eta_j(a) - \sum_{k=1}^{j-1} \frac{\eta_j^T \nu_k}{||\nu_k||^2}\nu_k, \qquad j \in \{1,\ldots,n\}
\end{aligned}
\tag{4.22}
$$

Some properties of Definition 4.5 are summarized in the following Lemma.

Lemma 4.3 *The orthogonal projection operation in Definition 4.5 has the following properties.*

(A) $\nu_i^T \nu_j = 0$ if $i \neq j$; $\nu_i^T \eta_j = 0$ if $i \geq j$,

(B) For some vector r, and for all $i = 1,\ldots,n$, $r^T \eta_i = 0$ if and only if $r^T \nu_i = 0$.

(C) $q^T \nu_i = 0$, $q^T \eta_i = 0$ for $i = 1,\ldots,n$.

Proof.(A) From (4.22) we have that $\nu_2^T \nu_1 = \eta_2^T \eta_1 - \frac{\nu_2^T \nu_1}{||\nu_1||^2}\nu_1^T \nu_1 = 0$. Assuming $\nu_j^T \nu_1 = 0$, it follows that $\nu_{j+1}^T \nu_1 = \eta_{j+1}^T \nu_1 - \frac{\eta_{j+1}^T \nu_1}{||\nu_1||^2}\nu_1^T \nu_1 = 0$. Hence $\nu_i^T \nu_1 = 0$ holds for any $i > 1$. Similar analysis can then be carried on to conclude that $\nu_i^T \nu_2 = 0$ for any $i > 2$, and likewise obtain that $\nu_i^T \nu_j = 0$ for any $i > j$. Since $\nu_i^T \nu_j = \nu_j^T \nu_i$, it follows that $\nu_i^T \nu_j = 0$ for any $i \neq j$.

For the second part of (A) we note that $\nu_2^T \eta_1 = \eta_2^T \eta_1 - \frac{\eta_2^T \eta_1}{||\eta_1||^2}\eta_1^T \eta_1 = 0$. Observing that $\nu_i^T \eta_j \neq \eta_i^T \nu_j$, and carrying out a similar procedure as above, we conclude that $\nu_i^T \eta_j = 0$ for $i > j$. (B) From (4.22) we have that

$$
\eta_j = \nu_j + \sum_{k=1}^{j-1} \frac{\eta_j^T \nu_k}{||\nu_k||^2}\nu_k, \qquad j \in \{1,\ldots,n\} \tag{4.23}
$$

Eq. (4.23) implies that if $r^T \nu_j = 0$ for all j, then $r^T \eta_k = 0$ for any $k = 1,\ldots,n$. In the converse, we start with $r^T \eta_j = 0$ for any j. We note that $r^T \nu_1 = r^T \eta_1 = 0$. Using the induction principle, we assume that $r^T \nu_j = 0$ for some j. From (4.22) it then follows that $r^T \nu_{j+i} = r^T \eta_{j+1} = 0$. Hence, $r^T \nu_j = 0$ for any j.

(C) Since $\nu_i^T \nu_j = 0$ from part (A), we have that $q^T \nu_i = l^T \nu_i - \frac{l^T \nu_i}{||\nu_i||^2} \nu_i^T \nu_i = 0$ for any $i \in \{1, \ldots, n\}$. Using part (B), it follows that $q^T \eta_i = 0$ as well.

In what follows, we will use the notation $e_i(\theta) = e(u_i, \theta)$, and $\lambda_i(\theta) = \lambda(u(t_i), \theta)$, where $t_i \in \Omega_t$.

Definition 4.6 *Let*

$$\begin{aligned} H(\theta) &= \{\ u_i \mid e(u_i, \theta) = 0\} \\ H_\lambda(\theta) &= \{\ \lambda(u, \theta) \mid u \in H(\theta)\} \\ I(\theta) &= \dim \{\mathcal{L}\{H_\lambda(\theta)\}\} \\ K_i &= \{\ \theta \mid I(\theta) \geq i\} \end{aligned} \tag{4.24}$$

Definition 4.6 specifies the construction of several new sets which are crucial in establishing our convergence analysis. These sets are important because they indirectly allow and lead into the design of a sequence of distance metrics for convergence. The global closed form analytical expressions for these distance metrics is not derived. This is in contrast to the vast majority of present literature on nonlinear systems, where a single positive-definite Lyapunov function of state and possibly time of the form $V(x, t)$ is considered. The positive-definite function $V(x, t)$ is explicitly defined and stated, and the system is said to be stable if the the value of the function decreases with time. This is the place where we break from the traditional view of stability analysis. Our approach is to analyze the state-space points per se, irrelevant of the type of system motion in it. By analyzing the points of state-space, we wish to associate certain characteristics with each point. Then, based on those characteristics of each point, we group points with similar characteristics into larger sets, and on each one of these sets design an appropriate distance metric. The construction of distance metrics is specified locally, and it is shown that it is not necessary to derive the global closed-form solution for these metrics. Hence, the design of a stability-analysis tools is such that the dependence on the state-space is implicit.

We now provide a few qualitative comments about the sets in Definition 4.6. First, the set $H(\theta)$ associates with each point θ a set of input values u such that, at the point θ, the observed output error is zero. Since the underlying function h in the system is nonlinear and depends on both θ and u, different points in the θ-space might have different sizes of the corresponding set $H(\theta)$. It is this difference in sizes of the corresponding $H(\theta)$ for different points in the θ-space that can be exploited to form a qualitative measure of how distant a particular point θ is from the point θ^*. At θ^*, by definition, any value of u produces a zero output error e. Hence, the smaller the size of $H(\theta)$, the farther away is the point θ from θ^*.

Specifically, a quantitative tool for measuring of the size of $H(\theta)$ can be determined by using the sets $H_\lambda(\theta)$ and $I(\theta)$. First, the set $H_\lambda(\theta)$ specifies the gradient vectors at the point θ for the values of u in $H(\theta)$. Next, the set $I(\theta)$ counts how many linearly independent gradient vectors there are in each $H_\lambda(\theta)$. This discrete number which $I(\theta)$ associates with each point will be the measure of the size of $H(\theta)$, and plays a central role in categorizing and grouping the points in state-space into different regions. This grouping is done through the sets K_i, which collect all points in θ-space whose index $I(\theta)$ is greater than i. The definition of $H(\theta)$ implies that we group together all points for which there exist a certain number of inputs such that (i) the zero output error surface for those inputs passes through that point <u>and</u> (ii) the gradients to those surfaces are linearly independent. The fact that we require the gradients to be linearly independent is important. First of all, we note that the the direction of system motion is specified by the gradient. Thus, the larger number of linearly independent gradients there are, the more of the state-space is explored, which in turn is a pre-requisite for accurate parameter identification.

We will establish the proof of convergence by showing that $\widehat{\theta}$ begins in K_0, progresses through $K_1, K_2, \ldots, K_{N-1}$, and converges to K_N, which is coincidental with θ^*. Before we state the main result we use the above definitions to derive key state-space properties. Those properties are summarized in Lemmas 4.4 and 4.5.

Lemma 4.4 *If $u \in U_{PE}^N$ over some interval Ω_t, then $K_N = \theta^*$.*

Proof. First, we will show that the set K_N consists of isolated points. Then, we will show that only $\theta^* \in K_N$. Suppose that there exist a curve $L_0 \subset K_N$, with a tangent vector denoted by l_0. From Definition 4.6, K_N is the intersection of N manifolds whose normals are λ_i, $i = 1, \ldots, N$. Since $L_0 \in K_N$, and l_0 is a tangent vector to L_0, it follows that

$$\lambda_i^T(\theta)\, l_0 \;=\; 0, \qquad i = 1, \ldots, N, \qquad \theta \in L_0 \tag{4.25}$$

From Definition 4.6, it follows that λ_i are linearly independent. Hence, (4.25) allows only the trivial solution $l_0 = 0$, and thus a curve $L_0 \in K_N$ does not exist. Therefore, K_N consists only of isolated points. It also follows that $K_N \bigcap N_\delta(\theta^*) = \theta^*$.

In order to show that $K_N = \theta^*$, by applying Definition 4.6, it now suffices to show that

$$\max_{\theta \notin N_\delta(\theta^*)} I(\theta) = N - 1 \tag{4.26}$$

Let

$$\Psi \;=\; \{\, t_i \in \Omega_t,\;\; i = 1, \ldots, N-1 \;\mid\; |t_{i+1} - t_i| > \varepsilon_t,\;\; \dim\{\mathcal{L}\{\Lambda(\Psi, \theta^*)\}\} = N \tag{4.27}$$

with $\Lambda(\Psi, \theta) = \{\lambda_i(\theta) \;\mid\; t_i \in \Psi\}$. Our assumption that $u \in U_{PE}^N$ and property (P2) in Definition 4.4 guarantee that Ψ as in (4.27) exists. Next, let

$$L \;=\; \{\, \theta \mid e(u(t), \theta) = 0,\;\; t \in \Psi \,\} \tag{4.28}$$

Let $l(\theta)$ denote a tangent vector of L. Since $\Lambda(\Psi, \theta)$ is the set of gradients to the manifolds $e(u(t_i), \theta)$, $t_i \in \Psi$, it follows that

$$l^T(\theta)\lambda_i(\theta) = 0, \qquad \lambda_i(\theta) \in \Lambda(\Psi, \theta) \tag{4.29}$$

Eq. (4.27) implies that $\Lambda(\Psi, \theta^*)$ is linearly independent, and that it spans a $N-1$ dimensional subspace. Therefore, l is uniquely specified, and hence L represents a curve that maps $\mathbb{R}$ to $\mathbb{R}^N$. Since L is a curve, and since by inspection, $\theta^* \in L$, we will reparameterize L in terms of a scalar s chosen such that $\theta(s) \in L$, and $\theta(s=0) = \theta^*$. In the following exposition, we will use the parameter s to denote all points $\theta(s)$ on L. Since (4.29) holds, it follows that

$$l(s) = \mathcal{N}\{\Lambda(\Psi, s)\}. \tag{4.30}$$

We will show that (4.26) holds by contradiction. Since (4.26) considers points $\theta(s) \notin N_\delta(\theta^*)$, we only examine the points for which $|s| \geq \varepsilon_s > 0$. Specifically, to prove (4.26) we will examine what the implications are of supposing that there exists an s_1, $|s_1| \geq \varepsilon_s > 0$, such that $I(s_1) = N$. Then, we will show that these implications cannot be satisfied on L, and therefore it must be that $I(s_1) < N$.

From (4.28), it follows that $I(s) \geq N-1$. Suppose that there exists an s_1, $|s_1| \geq \varepsilon_s > 0$ such that $I(s_1) = N$. Since $\dim\{\Psi\} = N-1$, $I(s_1) = N$ implies that there exists a $t_a \notin \Psi$ such that

$$(a)\;\; \lambda_a(s_1) \notin span\,\{\Lambda(\Psi, s_1)\} \quad \text{and} \quad (b)\;\; e(u(t_a), s_1) = 0 \tag{4.31}$$

Eq (4.31)(a) stems from the fact that $I(s_1)$ is the number of linearly independent $\lambda' s$ at s_1. Since $\Lambda(\Psi, s_1)$ has only $N-1$ vectors, whereas $I(s_1) = N$, eq (4.31)(a) follows. Eq (4.31)(b) follows from the fact that the number of linearly independent $\lambda' s$ at s_1 are evaluated only for values of u_i for which $e(u_i, s_1) = 0$.

Since (4.30) and (4.31)(a) hold, it follows that there exists an $\varepsilon_0 > 0$ such that

$$|\lambda_a^T(s_1) l(s_1)| \geq \varepsilon_0 \tag{4.32}$$

Since (4.30) and (4.32) hold, Corollary 4.2 implies that there exists an ε_1 such that

$$\left|\lambda_a^T(s) l(s)\right| \geq \varepsilon_1 \qquad \forall s \quad \text{on } L. \tag{4.33}$$

Denoting $e_a(s) = e(u(t_a), s)$, we will now show that $e_a(s_1) \neq 0$. Since $\lambda_a(s)$ is the gradient of $e_a(s)$, it follows that the rate of change of $e_a(s)$, denoted by de_a, along the curve L is specified by $de_a(s) = \lambda_s^T(s)l(s)$. Since $e_a(0) = 0$, we have that

$$|e_a(s_1)| = \left| \int_0^{s_1} \lambda_a^T(s)l(s)ds \right| \tag{4.34}$$

Since (4.33) holds, it follows that $|e_a(s_1)| \geq \varepsilon_e$, where $\varepsilon_e = \varepsilon_1|s_1|$. However, this contradicts (4.31)(b). Since for a given Ψ, t_a in (4.31) is arbitrarily chosen, it follows that the contradiction result is valid for any t_a which satisfies (4.31). Thus, for any given Ψ, we have that a point s_1 as in (4.31) does not exist. Therefore, $\max I(s_1) = N - 1$ for all $|s_1| \geq \varepsilon_s$. Hence, (4.26) holds, and thus the Lemma is established. •

Lemma 4.5 *Let $\theta_a \notin N_\delta(\theta^*)$. If $u \in U_{PE}^N$ over some interval Ω_t, then there exist at least one $t_1 \in \Omega_t$ and an $\varepsilon_e > 0$ such that*

$$|e(u(t_1), \theta_a)| \geq \varepsilon_e. \tag{4.35}$$

Proof. To establish Lemma 4.5, we start with an arbitrary θ_a and time instants t' such that $e(u(t'), \theta_a)$ is small. We then show that if $u \in U_{PE}^N$, then there must be a time instant t_1 such that $e(u(t_1), \theta_a)$ becomes large. This is established in two steps. We first prove Lemma 4.5 starting from t' where $e(u(t'), \theta_a) \equiv 0$. Next, we examine the case when $e(u(t'), \theta_a)$ is small and show that Lemma 4.5 is still valid.

Let

$$\mathcal{T} \;=\; \{\; t \;|\; |e(u(t), \theta_a)| < \varepsilon_a \;\}, \qquad \varepsilon_a > 0; \tag{4.36}$$

and Ψ is a set such that

$$(i)\Psi \subset \mathcal{T}, \;\text{ and }\; (ii)\Lambda(\Psi, \theta^*) = \mathcal{L}\left\{\Lambda(\mathcal{T}, \theta^*)\right\}; \tag{4.37}$$

and a set L is defined as

$$L = \{\, \theta \mid e(u(t_i),\theta) = 0, \;\; t_i \in \Psi \,\}. \tag{4.38}$$

We will show that the definitions of $\mathcal{T}$ and L, and the fact that $u \in U^N_{PE}$, imply that θ_a must lie in a neighborhood of L. The proof of Lemma 4.5 then proceeds by considering the steps: (i) $\theta_a \in L$, and (ii)$\theta_a \in N_\delta(L)$.

For ease of notation, we will denote $\lambda(u(t),\theta)$ as $\lambda(\theta)$ when the value of either t or $u(t)$ is obvious. Definition (4.37) implies that

$$\lambda(\theta^*) \in span\{\Lambda(\Psi,\theta)\} \quad \text{for all} \quad \lambda(\theta^*) \in \Lambda(\mathcal{T},\theta^*) \tag{4.39}$$

Corollary 4.1 then implies that (4.39) holds for any θ. Hence, we have that

$$span\{\Lambda(\Psi,\theta)\} \equiv span\{\Lambda(\mathcal{T},\theta)\}, \tag{4.40}$$

Since $\Lambda(\Psi,\theta)$ is the set of gradients to L, it follows that

$$\begin{aligned} \mathcal{N}(\Lambda(\Psi,\theta)) &= \Pi(L,\theta), \quad \text{and by using (4.40) that} \\ \mathcal{N}(\Lambda(\mathcal{T},\theta)) &= \Pi(L,\theta). \end{aligned} \tag{4.41}$$

We first note that we are interested in examining the largest possible set $\mathcal{T}$ such that $\mathcal{T} \in \Omega_t$. For, if there is a single time instant $t^* \in \Omega_t$ such that $t^* \notin \mathcal{T}$, it implies that there exists an ε_a such that $|e(u(t^*),\theta_a)| \geq \varepsilon_a$, and the proof of Lemma 4.5 is done. We will use the size of $span\{\Lambda(\mathcal{T},\theta)\}$ as a tool to measure how large the set $\mathcal{T}$ is. We note that a larger set $\mathcal{T}$ would imply a larger possible $span\{\Lambda(\mathcal{T},\theta)\}$. Since we are considering the largest possible set $\mathcal{T}$, this implies that we need to consider the largest possible $span\{\Lambda(\mathcal{T},\theta)\}$. By (4.40), this directly translates to examining $\dim\{\Psi\}$. Suppose that $\dim\{\Psi\} = i$, $i < N$. Let $w \notin span\{\Lambda(\Psi,\theta)\}$. Since $u \in U^N_{PE}$, from Lemma 4.2, it follows that there exists a t_a and ε_w such that $|\lambda^T(u(t_a),\theta)\, w| \geq \varepsilon_w$. This implies that $\lambda_a(\theta) \notin span\{\Lambda(\Psi,\theta)\}$. If $t_a \notin \mathcal{T}$, Lemma

4.5 is established. Hence, suppose that $t_a \in \mathcal{T}$. Consequently, we then have that $\dim\{\Psi\} = i+1$, and this process can be repeated. However, since Lemma 4.5 postulates that $\theta_a \notin N_\delta(\theta^*)$, Lemma 4.4 specifies an upper bound on $\dim\{\Psi\}$. That is $\max\limits_{\theta \notin N_\delta(\theta^*)} I(\theta) = N-1$. Therefore, throughout the rest of the proof, we assume that $\dim\{\Psi\} = N-1$.

We now note that if $\theta \in L$, we have that $e(u(t),\theta) = 0$ for all $t \in \mathcal{T}$. This can be easily shown by considering the following two facts. First, $e(u(t),\theta^*) = 0$ for any t. Second, from (4.41) we have that

$$q^T\lambda(\theta) = 0 \quad \text{for any } q \in \Pi(L,\theta), \;\; \lambda(\theta) \in \Lambda(\mathcal{T},\theta). \tag{4.42}$$

Since $q(\theta)^T\lambda(u(t),\theta)$ specifies the rate of change of $e(u(t),\theta)$ along $q(\theta)$, it follows that the rate of change of $e(u(t),\theta)$ along any q in L is zero for any $t \in \mathcal{T}$. Hence, $e(u(t),\theta) = 0$ for all $\theta \in L$ and $t \in \mathcal{T}$.

We now proceed to show that if θ_a is such that $|e(u(t),\theta_a)| < \varepsilon_a$ for all $t \in \mathcal{T}$, then $\theta_a \in N_\delta(L)$. We have established that in L, $e(u(t),\theta) = 0$ for all $t \in \mathcal{T}$. Next, we will show that in any direction orthogonal to L, there exists a $t \in \mathcal{T}$ such that $|e(u(t),\theta)|$ becomes large.

Let $l(\theta)$ be a tangent unit vector to L, that is $l(\theta) \in \Pi(L,\theta)$, $\|l(\theta)\| = 1$. From (4.41) we have that $l^T(\theta)\lambda(\theta) = 0$ for all $\lambda(\theta) \in \Lambda(\Psi,\theta)$. Since we have that $\dim\{\Psi\} = N-1$, (4.41) implies that $\dim\{\mathcal{L}\{\Pi(L,\theta)\}\} = 1$. It follows that $l(\theta)$ is uniquely specified. Let L be parameterized by a scalar r such that $L(r=0) = \theta^*$. On L, we now pick an arbitrary point $L(r_a)$. Starting from $L(r_a)$, we construct a curve $C(s)$, parameterized by a scalar parameter s. In further text when possible, we will use the index s to denote the point $C(s)$. The curve $C(s)$ is constructed such that the following properties are satisfied:

(C1) $C(0) = L(r_a)$,

(C2) the mapping $C(s) : \mathbb{R} \to \mathbb{R}^N$ is one-to-one, $C(s)$ is smooth.

(C3) (a) Let $c(s) \in \Pi(C,s)$. Then, $c(s) \in span\{\Lambda(\Psi,s)\}$ for all s; and

(b) the sign of $c^T(s)\lambda_i(s)$ is invariant for all s, $\lambda_i(s) \in \Lambda(\Psi, s)$.

Property (C3)-(a) implies that the curve $C(s)$ is orthogonal to L at their intersection. Property (C3)-(b) implies that $C(s)$ can be thought of as a monotonic curve in the coordinate systems specified by the set $\Lambda(\Psi, s)$. The properties (C1)-(C3) are not restrictive, in that for any point $\theta_a \notin L$, we can choose an arbitrary r_a, such that $L(r_a) = C(0)$, and $\theta_a = C(s_a)$ for some s_a. Using the curve C, we show that for any θ_a such that (4.36) holds, $\theta_a \in N_\delta(L)$.

For some $t_c \in \mathcal{T}$, we have that the rate of change of $e_c(s) = e(u(t_c), s)$ is given by $de_c = c^T(s)\lambda(u(t_c), s)$. Since $e_c(u(t), 0) = 0$ for $t \in \mathcal{T}$ from (4.36) and $(C1)$, we have that

$$\left| \int_0^{s_a} c^T(s)\lambda(u(t), s)\, ds \right| < \varepsilon_a, \quad \text{for any} \quad t \in \mathcal{T}. \tag{4.43}$$

Our goal is now to show that if (4.43) is to hold, the upper limit of integration, s_a, must be bounded. We note that since $c(s) \in span\{\Lambda(\Psi, s)\}$, this implies that $c(s)$ can be expressed as a linear combination of elements of $\Lambda(\Psi, s)$, with at least one of the coefficients being non-zero. It follows then that for every s, there there exists an $\varepsilon_c(s)$ and at least one $i \in [1, \ldots, N-1]$ such that $|c^T(s)\lambda_i(s)| > \varepsilon_c(s)$. Since $c^T(s)\ c(s) = 1$ and $\|\lambda_i(s)\|$, $i =, 1 \ldots, N-1$, are bounded for all s, it follows that $\varepsilon_c(s)$ must have a lower bound. Thus, there exists an ε_{c_0} such that for every s_i there is an $i \in [1, \ldots, N-1]$ such that

$$|c^T(s_i)\lambda_i(s_i)| > \varepsilon_{c_0}. \tag{4.44}$$

Since $C(s)$ is smooth, and since $c^T(s)\lambda_i(s)$ does not change sign, this implies that for each s_i such that (4.44) holds, there exists a $\delta_s > 0$ such that

$$\left| \int_{s_i}^{s_i+\delta} c^T(s)\lambda_i(s)\, ds \right| > \varepsilon_{c_0}\delta_s. \tag{4.45}$$

Since (4.44) and (4.45) hold for an arbitrary s_i, and since the space is finite-dimensional, it follows that, in the worst case, there exists an i^* such that $|c^T(s)\lambda_{i^*}(s)|$ gets periodically large. Since by (C3), $c^T(s)\lambda_{i^*}(s)$ is a monotonic function in s for any t_{i^*}, it follows that, for a given ε_a, there exist an s_a, M_1 and an $i^* \in [1, \ldots, N-1]$ such that

$$|e(u(t_{i^*}), s_a)| = \left| \int_0^{s_a} c^T(s)\lambda_{i^*}(s)\, ds \right| \geq \varepsilon_a, \qquad s_a = M_1 \varepsilon_a \tag{4.46}$$

For $\theta_a = C(s_a)$, (4.46) implies that if $t_{i^*} \in \mathcal{T}$, then

$$s_a < M_1 \varepsilon_a \tag{4.47}$$

for an arbitrary ε_a. Thus, we have established that $\theta_a \in N_\delta(L)$.

We now consider two cases, case (i) $\theta_a \in L$, and case (ii) $\theta_a \in N_\delta(L)$. In both cases we show that starting with $t \in \mathcal{T}$, there exists a t_1 such that $|e(u(t_1), \theta_a)|$ must become large.

Case (i) $\theta_a \in L$: From Lemma 4.2, there exists a $t_1 \in \Omega_t$ such that $\lambda_1(r) = \lambda(u(t_1), L(r))$ has a non-zero projection along $l(r)$ for any r. Therefore,

$$\left| \lambda_1^T(r) l(r) \right| \geq \varepsilon_2 \quad \forall r \in [0, r_1] \tag{4.48}$$

where r_1 is such that $\theta_a = L(r_1)$. Let $e_1(r) = e(u(t_1), L(r))$. Noting that $e_1(0) = 0$ and that $\lambda_1^T l$ denotes the change in e_1 along L, we have that

$$|e_1(r_1)| = \left| \int_0^{r_1} \lambda_1^T(r)\ l(r)\, dr \right| \geq \varepsilon_2 r_1. \tag{4.49}$$

This proves case (i).

Case (ii) $\theta_a \in N_\delta(L)$: Let $e_1(s_a) = e(u(t_1), C(s_a))$, for with s_a as given in (4.47). It follows that

$$|e_1(s_a)| \geq |e_1(r_1)| - |e_1(r_1) - e_1(s_a)| \tag{4.50}$$

Integrating the change of e_1 along the curve C, the second term in (4.43) can be expressed as

$$|e_1(r_1) - e_1(s_a)| \le \left| \int_{L(r_1)}^{C(s_a)} de_1 \right| \le \left| \int_0^{s_a} c^T(s)\lambda(u(t_1), s)\, ds \right| \le s_a M_2, \qquad (4.51)$$

where $|c^T \lambda_1| \le M_2$, and $L(r_1) = C(0)$. From (4.47), (4.49) and (4.51) we have that

$$|e_1(s_a)| \ge \varepsilon_2 r_1 - M_2 M_1 \varepsilon_a \triangleq \varepsilon_e. \qquad (4.52)$$

Since ε_a is arbitrary, ε_e is guaranteed to be positive. This proves case (ii). •

Theorem 4.1 *Let Assumptions 4.1 and 4.2 hold. For the system in (4.1)-(4.3), if for every $t > 0$ there exist a $t_1 > t$ and $T > 0$ such that $u(t) \in U_{PE}^N$ over the interval $[t_1, t_1 + T]$, then $\lim_{t\to\infty} \widehat{\theta}(t) = \theta^*$.*

Proof. We will prove the theorem by showing that the index $I(\widehat{\theta})$ is monotonically increasing along system trajectories. In order to establish this, we need to show that the system exhibits the following properties.

(1.) If $\widehat{\theta}(t_a) \in K_i$, then there exists a $t_b > t_a$ such that $\widehat{\theta}(t_b) \in K_{i+1}$, for $i = 0, 1, 2, \ldots, N-2$, and

(2.) If $\widehat{\theta}(t_b) \in K_{i+1}$, then $\widehat{\theta}(t) \in K_{i+1}$ for all $t > t_b$, and $i = 0, 1, 2, \ldots, N-2$.

(3.) If $\widehat{\theta}(t) \in K_{N-1}$, then $\lim_{t\to\infty} \widehat{\theta}(t) \in K_N$.

Step 1. The goal in Step 1 is to show that starting from an arbitrary point in K_i, $\widehat{\theta}$ will be in the set K_{i+1} after a finite time. To show this, in Step 1.1 we will examine a specific manifold M which can lie both in K_i and K_{i+1}. In Step 1.2, we will construct a metric J on M that represents the distance of our starting point in K_i to the set K_{i+1}, and show that J converges to zero. Then, in Step 1.3, we will show that J converges to zero in finite time.

Step 1.1 In this step, we first establish the tools necessary for constructing a convergence metric. We do this by examining the properties of the set K_i and K_{i+1}.

Using these properties, we shall construct a metric J, which will be a measure of the distance of a point in K_i to the set K_{i+1}.

Suppose that for some t_a, $\widehat{\theta}(t_a) \in K_i$. From the definition of K_i, this implies that (a) there exists a set $\Psi_M = \{t_1, t_2, \ldots, t_i\}$ such that $\Lambda(\Psi_M, \widehat{\theta}(t_a))$ is linearly independent, and (b) that $\widehat{\theta}(t_a)$ lies on the intersection of i manifolds $e_j(\widehat{\theta}) = 0$, $j = 1, \ldots, i$. Let

$$
\begin{aligned}
M(\Psi_M) &= \{\theta \mid e(u(t), \theta) = 0, \;\; t \in \Psi_M\} \\
W(\theta) &= span\,\{\Lambda(\Psi_M, \theta)\}
\end{aligned}
\tag{4.53}
$$

Since every $\lambda(u(t), \theta) \in \Lambda(\Psi_M, \theta)$ represents the gradient to the surface $e(u(t), \theta)$, it follows that the set $\Lambda(\Psi_M, \theta)$ is the set of all normals to the manifold M at a point θ.

Next, we choose a time instant t_l such that for some θ,

$$
\lambda(u(t_l), \theta) \notin W(\theta). \tag{4.54}
$$

Since $\dim\{\Lambda(\Psi_M, \theta)\} = i$, where $i < N$, the fact that $u \in U_{PE}^N$ guarantees that such a time instant t_l exists. Let

$$
\begin{aligned}
\Psi_e &= \Psi_M \bigcup t_l, \\
M_e(\Psi_e, \theta_1) &= \{ \;\theta \mid \theta \in M(\Psi_M), \;\; e(u(t_l), \theta) = e(u(t_l), \theta_1) \;\} && (4.55) \\
M_z(u) &= \{ \;\theta \mid \theta \in M(\Psi_M), \;\; e(u(t), \theta) = 0, \;\; \lambda(u, \theta) \notin W(\theta) \;\} && (4.56)
\end{aligned}
$$

Thus, it can be seen that the manifold M_e represents the intersection of the surface $e(u(t_l), \theta) = c_1$, where c_1 is some constant, and the manifold M. Since M is an $N - i$ dimensional manifold, it follows that M_e is a manifold of dimension $N - i - 1$. Since $M_e \in M$, it follows that every $\lambda(\theta) \in \Lambda(\Psi_M, \theta)$ is linearly independent from any vector in M_e, since it is orthogonal to any vector in M_e. Since $\dim\{\Lambda(\Psi_M, \theta)\} = i$, it follows that if $\overline{M} = M_e \bigcup \Lambda(\Psi_M, \theta)$, then $\overline{M}$ spans an $N - 1$ dimensional subspace.

Comparing (4.56) and (4.55), we see that the sets M_z and M_e are similar in nature. As noted, for any particular value u_l, the set M_e represents the intersection of sets $e(u_l, \theta) = c_1$ and M, for some constant c_1. On the other hand, the set M_z represents the intersection of sets $e(u_l, \theta) = 0$ and M. Since the constant $c_1 = 0$ for the points in the set $M_z(u_l)$, it can be shown that if $\theta \in M_z(u_l)$, then $\theta \in K_{i+1}$. This follows because for any $\theta \in M$, it follows that $\theta \in K_i$. Since for any $\theta \in M_z(u_l)$ implies that $\theta \in M$, and since $e(u_l, \theta) = 0$ and $\lambda_l(\theta) \notin W(\theta)$ it follows that $\theta \in K_{i+1}$.

Using the properties of M_e and M we shall define a metric J. In particular, we will choose J as a positive definite function of a variable s, where s is the arclength of a curve Q. Q, in turn, will be specified by its tangent vector m, which will be shown to lie in M and be orthogonal to M_e.

Let θ_a be an arbitrary point in M_e, and let

$$m_u(\theta_a) = \lambda(u, \theta_a) \otimes \Lambda(\Psi_M, \theta_a) \tag{4.57}$$

For ease of notation, let $m_l(\theta_a) = \lambda(u(t_l), \theta_a) \otimes \Lambda(\Psi_M, \theta_a)$. We shall now show that *(i)* m_l is orthogonal to M_e, and that *(ii)* m_l lies in the tangent plane of M at θ_a. From (4.22) we have that

$$m_l = \lambda_l - \sum_{j=1}^{i} \frac{\lambda_l^T \nu_j}{||\nu_j||^2} \nu_j \tag{4.58}$$

where ν_j are defined as in (4.22). Let l_e be an arbitrary vector which lies in the tangent plane of M_e. Since λ_l is the gradient to $e_l(\theta) = const$, it follows that $\lambda_l^T l_e = 0$. However, $M_e \in M$, and since the elements of $\Lambda(\Psi_M, \theta)$ are orthogonal to any vector in the tangent plane of M, it follows that $\lambda_j^T l_e = 0$, for any $\lambda_j(\theta_a) \in \Lambda(\Psi_M, \theta_a)$. Applying Lemma 4.3-(B), we have that $l_e^T \nu_j = 0$ for $j = 1, \ldots, i$. Thus

$$m_l^T l_e = \lambda_l^T l_e - \sum_{j=1}^{i} \frac{\lambda_l^T \nu_j}{||\nu_j||^2} \nu_j^T l_e = 0 \tag{4.59}$$

Since l_e is arbitrary, we conclude that m_l is orthogonal to M_e. Likewise, from Lemma 4.3(C), we obtain that $m_l^T(\theta_a)\lambda_j(\theta_a) = 0$ for any $\lambda_j(\theta_a) \in \Lambda(\Psi_M, \theta_a)$.Hence, m_l lies in the tangent plane of M.

As noted above, the sets $M_z(u_l)$ and M_e represent the intersections of manifolds $e(u_l, \theta) = c_1$ with the set M. In the case of $M_z(u_l)$, $c_1 = 0$, while in the latter case c_1 is an arbitrary constant. Since c_1 is an arbitrary constant, it follows that for any $\theta \in M_z(u_l)$, we have that $m_l(\theta)$ is orthogonal to the set $M_z(u_l)$.

Having established the properties of the normals of M_e and M_z, we now establish an additional property of M_z. Since we have shown that $M_z(u) \in K_{i+1}$, let u_l be chosen such that $M_z(u_l)$ represents a boundary of K_{i+1} on the set M. The term boundary is used to specify the following. Let $\theta_1 \in M_z(u_l)$, and let $\theta_2 = \theta_1 + \varepsilon_m m_l(\theta_1)$, where ε_m is some small constant. Then, θ_1 is on the boundary of K_{i+1} if there does not exist a u_2 such that $\theta_2 \in M_z(u_2)$. We are guaranteed that such a boundary of K_{i+1} exists on M. Otherwise, it would imply that the entire set M belongs to K_{i+1}, and the proof of this step of the theorem would be done.

We shall now construct a curve Q in M to specify the necessary convergence metric J. Let Q be a curve parameterized by a scalar s, and let the tangent vector to Q at $Q(s)$ be given by $q(s)$. In what follows, where advantageous for purposes of clarity and brevity, we shall use just the index s to denote a point $Q(s)$. We construct the curve Q such that

(a) for every s_1 such that $Q(s_1) \in K_i$, $q(s_1) = m_l(s_1)$.

(b) for every s_2 such that $Q(s_2) \in K_{i+1}$, $q(s_2) = m_u(s_2)$, where u is such that (i) $e(u, s_2) = 0$, and (ii) $\lambda(u, s_2) \notin W(s_2)$. Since $Q(s_2) \in K_{i+1}$, it follows that a u exists such that (i) and (ii) are satisfied.

Noting that Q always lies in M, we observe that Q follows along m_l in K_i, and along m_u in K_{i+1}. The specific value of u and thus m_u is determined by the property (b)-(i),(ii). These properties also imply that $Q(s_2) \in M_z(u)$ for every s_2.

We now define more precisely the parameterization of Q. Without loss of generality, we assume that $\widehat{\theta}(t_a) \in Q$. That is, there exists an s_a such that $Q(s_a) = \widehat{\theta}(t_a)$. Since m_l is orthogonal to $M_z(u_l)$, and since $M_z(u_l)$ is the boundary of K_{i+1} on M, it follows that at the boundary of K_{i+1} on M, Q is orthogonal to the boundary of

K_{i+1}. Without loss of generality, we choose the parameterization of Q such that $Q(s=0) \in M_z(u_l)$, and the direction of increase of s be along the direction of $q(s)$. By Assumption 4.2, we have that $\lambda_k^T \lambda_l > 0$ for any k, l. Hence, s monotonically changes along Q. Let s_b be such that $Q(s_b) \in K_{i+1}$. The initial Step 1. assumption was that $Q(s_a) \in K_i$. Since $s = 0$ at the boundary of K_{i+1}, and since s monotonically changes along Q, this implies that

$$s_a(s_a - s_b) > 0 \tag{4.60}$$

Thus, the parameter s represents the distance of a point s_a in K_i to the set K_{i+1}. For an arbitrary s_a, we define a metric J measured along Q as

$$J = \frac{1}{2} s_a^2 \tag{4.61}$$

Step 1.2 In Step 1.1 we have constructed a distance metric J as given by (4.61). In this step, we examine the time change of the metric along the direction of $\dot{\hat{\theta}}$. Differentiating (4.61) with respect to time, we have that

$$\dot{J} = s_a \, \dot{s}_a \, . \tag{4.62}$$

Since $q(s_a)$ is the tangent vector along which s_a changes, it follows that, using (4.3), we obtain the change of s_a along the system trajectory as

$$\dot{s}_a = q^T(s_a) \, \dot{\hat{\theta}} \, (t) = -e(u(t), s_a) \, q^T(s_a) \, \lambda(u(t), s_a) \tag{4.63}$$

Since $de(u(t), s) = q^T(s) \lambda(u(t), s) \, ds$, it follows that for any given time instant t we obtain that

$$e(u(t), s_a) = \int_{s_b}^{s_a} m^T(s) \lambda(u, s) \, ds \tag{4.64}$$

where s_b is such that

$$e(u, \theta_q(s_b)) = 0. \tag{4.65}$$

Then,

$$\dot{J} = s_a \, \dot{s_a} = -\int_{s_b}^{s_a} s_a f(s_a) f(s)\, ds\,, \qquad f(s) = q^T(s)\lambda(u(t), s). \tag{4.66}$$

We will now examine the function $f(s)$ defined in (4.66) by considering two cases:

(a) There exists an ε_q such that

$$\mathcal{T}_a = \left\{ t \;\mid\; |\lambda_l^T(s_a)\ \lambda(u(t), s_a| \geq \varepsilon_q, \quad \lambda(u(t), s_a) \notin W(s_a) \right\} \tag{4.67}$$

(b) $\mathcal{T}_b = \Omega_t \setminus \mathcal{T}_a$.

We shall show that $\dot{J} < 0$ in case (a) and $\dot{J} = 0$ in case (b), concluding that J decreases since $\mathcal{T}_a$ is nonempty. Since for any $t_1 \in \mathcal{T}_a$ we have that $|\lambda_l^T(s_a)\lambda_1(s_a)| \geq \varepsilon_q$, Corollary 4.3 implies that there exists an ε_1 such that

$$\left|\lambda_l^T(s)\ \lambda_1(s)\right| > \varepsilon_1 \quad \forall s \in [s_b, s_a], \quad t_1 \in \mathcal{T}_a \tag{4.68}$$

Eq. (4.68) implies that

$$\lambda(u(t), s) \notin \Pi M_e(\Psi_e, s) \quad \forall s \in [s_b, s_a], \quad \forall t \in \mathcal{T}_a \tag{4.69}$$

Since $\lambda(u(t), s_a) \notin W(s_a)$, (P1) implies that

$$\lambda(u(t), s) \notin W(s) \quad \forall s \in [s_b, s_a]. \tag{4.70}$$

Letting v be any unit vector in M, eq (4.70) implies that

$$v^T \lambda(u(t), s) \neq 0, \quad \forall s \in [s_b, s_a]. \tag{4.71}$$

By definition, we have that $q(s) \in M$. Also, by definition, $q(s)$ is orthogonal to M_e at any s. Since $\lambda(u(t), s) \notin \Pi M_e(\Psi_e, s)$ for any s, it thus follows that there exists an ε_s such that $|q^T(s)\lambda(u(t), s)| \geq \varepsilon_s$ for any s and all $t \in \mathcal{T}_a$. Therefore, $f(s) \neq 0$ for any s. Thus, $f(s)f(s_a) > 0$ for all $s \in [s_b, s_a]$. Therefore, using the mean value theorem, we can obtain that

$$\dot{J} = -\int_{s_b}^{s_a} s_a f(s_a) f(s) ds = -s_a f(s_a) f(s_c)(s_a - s_b) \tag{4.72}$$

for some $s_c \in [s_b, s_a]$. Applying (4.60), we obtain that

$$\dot{J} = -s_a f(s_a) f(s_c)(s_a - s_b) < 0. \tag{4.73}$$

In case (b), we first note that Ω_t, and hence $\mathcal{T}_b$ are bounded. Since there does not exist an ε_q as in case (a), this implies that $q^T(s_a)\lambda(u(t), s_a) = 0$. By property $(P1)$, this implies that $q^T(s)\lambda(u(t), s)$ for all s. Hence $\dot{J} = 0$.

Since $u \in U_{PE}^N$, we have that $\mathcal{T}_a$ is nonempty. Thus, J decreases, and hence $\widehat{\theta}$ tends towards K_{i+1} from K_i.

Step 1.3 In Steps 1.1 and 1.2 we have constructed a distance metric from a point in K_i to the set K_{i+1}. We have shown that the distance metric decreases everywhere in K_i. We now need to show that J converges to zero in finite time. That is, we need to show that there exists a finite $t_b > t_a$ such that $\widehat{\theta}(t_b) \in K_{i+1}$. To establish this, it is sufficient to show that in the neighborhood of $J = 0$, the system velocity $\| \dot{\widehat{\theta}} \|$ has a lower bound. That is, we need to show that there exists an $\varepsilon_v > 0$ such that $\|\dot{\widehat{\theta}}(t)^T q(s)\| > \varepsilon_v$ for all t such that $\widehat{\theta}(t) \in K_i$. From Lemma 4.1 we have that $|\lambda^T(u(t), s)q(s)| \geq \varepsilon_2$ for all s. If $\theta \in N_\delta(\theta^*)$, then the theorem is done. We suppose that $\theta \notin N_\delta(\theta^*)$. From Lemma 4.5 we have that there always exists an ε_e such that $|e(u(t), \theta)| \geq \varepsilon_e$. Letting $\varepsilon_v = \varepsilon_2 \varepsilon_e$, we establish that a lower bound on the velocity exists, and hence that there exists a $t_b > t_a$ such that $\widehat{\theta}(t_a) \in K_i$, and $\widehat{\theta}(t_b) \in K_{i+1}$. Hence, we are done with Step 1.

Step 2. In Step 1 we have shown that for any point outside of K_{i+1}, the distance metric J to the set K_{i+1} decreases with time. Hence, there does not exist a trajectory which, starting on the boundary surface, increases the distance to set K_{i+1}. Since J is measured along a curve which is normal to the boundary surface at the boundary, it implies that the set K_{i+1} is invariant. By definition, we have that K_0 covers all state-space, and hence is invariant. Therefore, it holds that for all i, once the trajectory enters a set K_i, it always remains in K_i.

Step 3. In Step 3, we would like to show that starting in K_{N-1}, the system asymptotically converges to K_N. In substeps 1.1 and 1.2 we have established that there exists a decreasing distance metric from K_i to K_{i+1}. Since i is arbitrary, it holds for $i = N - 1$ as well. Thus, we can construct a non-negative distance metric J from any point in the set K_{N-1} to the set K_N. At any point in K_{N-1}, the distance metric J is decreasing. Hence, $\widehat{\theta}$ tends to K_N from K_{N-1}. •

Thus, we have established that if the signal $u(t)$ is sufficiently rich for the given nonlinear parameterization, it is possible to have parameter convergence with the gradient algorithm.

4.3 Concluding Remarks and Future Work

In the previous section, we have given a set of conditions under which the gradient parameter identification algorithm is globally convergent. The two convergence conditions involve both the input u, and the nonlinear function h. Here, we compare the two conditions with the existing results in linear parameterization case.

What the first condition essentially requires is a form of similarity between different points in the parameter space. By property (P1), it is required that if a set of gradient vectors λ_i be linearly independent for a set of values of u at a particular point θ, then the set of gradient vectors be for the same set of values u be linearly independent at any other point θ. Thus, all the points in the θ space are similar. It can be argued that this type of similarity stems from the fact that in monotonic functions, a neighborhood of a point is similar in terms of the slope to the neighborhood of any other point. In the linear parameterization case, the gradient vector λ is a function of the external input only. Thus, for a given set of external inputs, we have exactly the same behavior of the set of gradient vectors everywhere in the parameter space. Hence, linear parameterization identically satisfies property (P1). Thus, in the linear case, this property is inherent to the parameterization, and does not need to be specified explicitly. We can also view (P1) as a relaxation of the linear case property that all the neighborhoods of points in parameter space be exactly the same to the property that all neighborhoods are "similar".

Property (P2) has the same form as the standard linear persistent excitation property (see [17, 38]). It requires the external input to be sufficiently rich so that the vector λ periodically orients itself along any given direction. Since the direction of the parameter estimate update is given by the direction of λ, (P2) is a requirement that all of the regions of the state-space are explored. It is to be expected that such a property is necessary in establishing parameter convergence.

The results given here represent only a first step in research into the use of the gradient vector in nonlinearly parameterized systems. One of the tasks ahead is to characterize more precisely which types of functions and what type of external inputs can satisfy properties (P1) and (P2). Preliminary studies show that sigmoidal functions of the type used most often in feed-forward neural networks satisfy (P1) nad (P2).

The results presented here can have a direct impact on the neural network training issues. However, most of the neural net training in practice is done in discrete steps, and not by a time-continuous process modeled by the system in (4.1),(4.3). Thus, it would be very useful from a practical viewpoint to examine the discrete version of the algorithm discussed here. A preliminary result would suggest the following modifications to the training process done in practice

<u>Step 1.</u> Apply the input u.

<u>Step 2.</u> Calculate the error $e(u,\widehat{\theta})$. Let $e_1 = e(u,\widehat{\theta})$

<u>Step 3.</u> Let $\theta_2 = -\nu e_1 \lambda(u,\widehat{\theta})$.

<u>Step 4.</u> Let $e_2 = e(u,\theta_2)$. If $e_2 e_1 > 0$ proceed to Step 5. Otherwise, let $\nu = \frac{1}{2}\nu$.

<u>Step 5.</u> $\widehat{\theta} = \theta_2$. Evaluate the training results. If not successful, select a u and goto Step 1. Otherwise, stop.

This method of controlling the step size ν would ensure that the sets K_i remain invariant. Otherwise, since the algorithm is taking finite sized discrete steps, there exists a possibility that the K_i can be crossed and the sets be no longer invariant.

Since the results presented here can be relevant to neural networks, it is worth examining several modifications to the system considered in (4.1). Such modifications would bring closer the considered system in (4.1) to some of the neural net architectures commonly used in practice. The primary modifications of interest would be expanding (4.1) to consider the following two cases

$$\begin{aligned} y &= W\sum_{i=1}^{N} g(u,\theta_i) \\ y &= \sum_{i=1}^{N} w_i g(u,\theta_i) \end{aligned}$$

where W, w_i would be scalar unknown parameters. It seems that the the extension of the obtained results to include the first system in (4.74) is quite feasible. That is because in this case $\hat{\theta} = -e\widehat{W}\lambda(u,\hat{\theta})$, with $\widehat{W}$ being our estimate of W. Letting $\lambda' = \widehat{W}\lambda(u,\hat{\theta})$, it can be shown that if λ satisfy (P1)-(P2), then λ' satisfies these properties as well. It is not clear immediately if this equivalence would hold in the second system in (4.74). Also, symmetry of the θ weight space is broken in this case. These types of systems require further investigations and generalizations of the presented results.

Finally, since the results presented can have an impact on the training process in neural nets, it would be useful to try to extend these results to encompass the on-line training problem. This is an open and very interesting problem. No global results exist in this case for nonlinearly parameterized systems. For a critique of the shortfalls of existing theoretical results, see [41]. Yet, results have been claimed in practice. Our preliminary investigations show that it is possible to derive stability results for on-line, dynamic training by reinterpreting the current results for Lyapunov function design in linearly parameterized systems.

Chapter 5

Conclusions

In this thesis, we have constructively addressed the issue of identification and control in nonlinearly parameterized (NLP) systems. We have attempted to bridge the gap between the wide and diverse applicability of NLP system models, and the existing control design methodologies. The existing methodologies have two major drawbacks. The first is that many results are only local in nature. The other is the fact that the majority of global results is based on feedback linearization techniques. Such technique require the knowledge of bounds of the parametric uncertainties, and do not obtain any additional information about the values of the unknown parameters.

In contrast to many existing techniques, our approach is based on adaptive control and parameter estimation. All the results presented are global in nature. For certain classes of systems, two goals were achieved: *(a)* design of stable adaptive control methodologies, and *(b)* parameter identification in NLP systems.

The thesis presented three theoretical results. In Chapter 2, we have stated the conditions under which accurate parameter identification is possible for a class of convex/concave NLP systems. The derived results were presented for both the continuous and discrete versions of the min-max estimator in [2]. Chapter 3 presented the results for adaptive control of NLP systems with a triangular structure. It was shown how an coupling the min-max estimator with an approach based on a type feedback linearization (similar to, for example [48]), results in global stability. Such a coupling retains the advantage of both techniques: the approach is overall globally stable, and the required control authority and actuator bandwidth may be reduced.

Chapter 4 presents the results in analyzing the convergence properties of the gradient algorithm in monotonically parameterized systems. For example, many of the general neural network architectures are of this type. The results give relatively mild conditions under which parameter identification is obtained using the gradient algorithm. Obtaining the results was made possible by developing new techniques and tools for analyzing the properties of the gradient algorithm. The tools are based on examining the properties of the state-space in the gradient algorithm. Upon such an examination, it was possible to construct a sequence of distance metrics. Each of the distance metrics is defined in separate regions of state space. It was shown that each of the distance metrics is decreasing while the system traverses the corresponding region of state-space. Thus, it was possible to show that the corresponding regions of state-space were invariant, and that the global sequence of traversing these regions results in overall parameter convergence.

Suggested future research is along two main tracks. One is a closer scrutiny of the min-max estimation algorithm in the context of other types of NLP systems than the ones studied in the thesis. Such an examination may yield (a) new types of nonlinearities for which parameter convergence is possible, and (b) further reduce the required control effort and actuator bandwidth in control of NLP systems. The second track consists of expanding the results for the gradient identifier in monotonically parameterized systems. Of particular concern is the extension of the results to the case when dynamics are involved between the nonlinear uncertainty and the available measurements. Preliminary investigations suggest that by re-interpreting the meaning of the standard Lyapunov functions used in linearly parameterized systems, it may be possible to construct new types of Lyapunov functions for NLP systems. Such a result would have a significant impact in applying the powerful neural network function approximation capabilities in an on-line, closed-loop learning environment. Further insight into learning and adaptation could also be gained.

Bibliography

[1] B.D.O. Anderson and Jr. C.R. Johnson. "Exponential convergence of adaptive identification and control". *Automatica*, vol. 18:1–13, 1982.

[2] A. M. Annaswamy, A.P. Loh, and F.P. Skantze. Adaptive control of continuous time systems with convex/concave parametrization. *Automatica*, 34:33–49, January 1998.

[3] A. M. Annaswamy, C. Thanomsat, N. Mehta, and A.P. Loh. Applications of adaptive controllers based on nonlinear parametrization. *ASME Journal of Dynamic Systems, Measurement, and Control*, 120:477–487, December 1998.

[4] A.M. Annaswamy, N. Mehta, C. Thanomsat, and A.P. Loh. Application of adaptive controllers based on nonlinear parametrization to chemical reactors and magnetic bearings". In *Proceedings of the ASME Dynamic Systems and Control Division*, Dallas, TX, November 1997.

[5] Brian Armstrong-Hélouvry, Pierre Dupont, and Carlos Canudas de Wit. A survey of models, analysis tools and compensation methods for the control of machines with friction. *Automatica*, 30(7):1083–1138, 1994.

[6] R. Bakker and A.M. Annaswamy. Stability and robustness properties of a simple adaptive controller. *IEEE Transactions on Automatic Control*, 41:1352–1358, September 1996.

[7] D. Bates. *Nonlinear regression analysis and its applications*. Wiley, New York, 1988.

[8] M. Bianchini, P. Frasconi, and M. Gori. Learning without local minima in radial basis function networks. *IEEE Transactions on Neural Networks*, 6(3):749–755, 1995.

[9] M. Bianchini and M. Gori. Optimal learning in artificial neural networks: A review of theoretical results. *Neurocomputing*, 13:313–346, 1995.

[10] J. D. Bosković. Stable adaptive control of a class of first-order nonlinearly parameterized plants. *IEEE Transactions on Automatic Control*, 40(2):347–350, 1995.

[11] J. D. Bosković. Stable adaptive control of a class of continuous- flow bioreactor processes. *AIChE J.*, 42(1):176–186, 1996.

[12] J.D. Bosković. Adaptive control of a class of nonlinearly parametrized plants. *IEEE Transactions on Automatic Control*, 43:930–934, July 1998.

[13] D. Sankar De, A.-P. Loh, and P. R. Krishnaswamy. A nonlinear adaptive controller for a ph neutralization process. In *Proceedings of the IEEE CDC*, pages 2250–2254, Chicago, IL, 2000.

[14] V. Fomin, A. Fradkov, and V. Yakubovich. *Adaptive control of dynamical systems.* Eds. Nauka, Moscow, 1981.

[15] S. H. Friedberg, A. J. Insel, and L. W. Spence. *Linear Algebra.* Prentice-Hall, Inc., Upper Saddle River, N.J., 1997.

[16] G. J. Gomez, S. H. K Embabi, E. Sanchez-Sinencio, and M. Lefebvre. A nonlinear macromodel for cmos otas. In *Proceedings of the International Symposium on Circuits and Systems*, volume 2, pages 920–923, 1995.

[17] G. C. Goodwin and K. S. Sin. *Adaptive Filtering, Prediction, and Control.* Prentice-Hall, 1984.

[18] M. Gori and M. Maggini. Optimal convergence of on-line backpropagation. *IEEE Transactions on Neural Networks*, 7(1):251–254, 1996.

[19] Brian Green. *The Elegant Universe.* W.W. Norton and Company, 1999.

[20] S. S. Haykin. *Neural networks: A comprehensive foundation.* Prentice-Hall, 1998.

[21] John Hertz, Anders Krogh, and Richard G. Palmer. *Introduction to the Theory of Neural Computation.* Addison-Wesley Publishing Company, 1991.

[22] R. Hotzel and L. Karsenti. Tracking control scheme for uncertain systems with nonlinear parametrization. In *Proceedings of the European Control Conference*, Brussels, Belgium, May 1997.

[23] A. Isidori. *Nonlinear Control Systems.* Springer-Verlag, New York, NY, 2nd edition, 1989.

[24] I. Kanellakopoulos, P.V. Kokotovic, and A.S. Morse. Systematic design of adaptive controllers for feedback linearizable systems. *IEEE Transactions on Automatic Control*, 36:1241–1253, November 1991.

[25] L. Karsenti, H. Sira-Ramirez, and F. Lamnabhi-Lagariggue. Adaptive backstepping pwm stabilization of dc-to-dc converters towards minimum or non-minimum phase equilibria. In *Proceedings of the 34^{th} IEEE CDC*, pages 1591–1596, New Orleans, LA, 1995.

[26] H. K. Khalil. *Nonlinear Systems.* Prentice Hall, Upper Saddle River, NJ, 1996.

[27] B.S. Kim and A.J. Calise. Nonlinear flight control using neural networks. *Journal of Guidance, Control and Dynamics*, 20:26–33, 1997.

[28] A. Kojić and A.M. Annaswamy. Parameter convergence in systems with convex/concave parameterization. In *The 2000 American Control Conference*, pages 2240–2244, Chicago, IL, June 2000.

[29] A. Kojić, A.M. Annaswamy, A.-P. Loh, and R. Lozano. Adaptive control of a class of nonlinear systems with convex/concave parameterization. *Systems & Control Letters*, 37:267–274, 1999.

[30] M. J. Kornenberg and I.W. Hunter. The identification of nonlinear biological systems: LNL cascade models. *Biological Cybernetics*, 55:125–134, 1986.

[31] M. Krstić, I. Kanellakopoulos, and P. Kokotović. *Nonlinear and Adaptive Control Design*. Wiley Interscience, New York, NY, 1995.

[32] M. Krstić, I. Kanellakopoulos, and P. V. Kokotović. Adaptive nonlinear control without overparameterization. *Systems & Control Letters*, 19:177–186, March 1992.

[33] R. Levin. *American Scientist*, 54:421–431, 1966.

[34] A. P. Loh, A.M. Annaswamy, and F.P. Skantze. Adaptation in the presence of a general nonlinear parametrization: An error model approach. *IEEE Transactions on Automatic Control*, AC-44:1634–1652, September 1999.

[35] S.B. Margolis. Nonlinear stability of combustion-driven acoustic oscillations. *Journal of Fluid Mechanics*, 53:67–103, 1993.

[36] R. Marino and P. Tomei. Global adaptive output-feedback control of nonlinear systems, part ii: Nonlinear parameterization. *IEEE Transactions on Automatic Control*, 38:33–48, 1993.

[37] A. P. Morgan and K. S. Narendra. On the stability of nonautonomous differential equations $\dot{x}= [a + b(t)]\,x$, with skew symmetric matrix $b(t)$. *SIAM Journal of Control and Optimization*, 15(1):163–176, January 1997.

[38] K. S. Narendra and A. M. Annaswamy. *Stable Adaptive Systems*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1989.

[39] M.S. Netto, P. Moya, R. Ortega, and A. M. Annaswamy. Adaptive control of a class of first-order nonlinearly parametrized systems. In *Proceedings of the CDC*, pages 5170–5175, Phoenix, AZ, USA, December 1999.

[40] H. Nijmeijer and A.J. van der Schaft. *Nonlinear Dynamical Control Systems*. Springer-Verlag, New York, NY, 1991.

[41] Romeo Ortega. Some remarks on adaptive neuro-fuzzy systems. *International Journal of Adaptive Control and Signal Processing*, 10:79–83, 1996.

[42] M. Polycarpou. Stable adaptive neural control scheme for nonlinear systems. *IEEE Transactions on Automatic Control*, 41:447–451, 1996.

[43] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1992.

[44] R. D. Reed and II R. J. Marks. *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. The MIT Press, Cambridge, Massachusetts, 1999.

[45] S. S. Sastry and M. Bodson. *Adaptive Control: Stability, Convergence, and Robustness*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1989.

[46] Danbing Seto, Anuradha M. Annaswamy, and J. Baillieul. Adaptive control of nonlinear systems with a triangular structure. *IEEE Transactions on Automatic Control*, 39(7):1411–1428, July 1994.

[47] F.P. Skantze, A.P. Loh, and A.M. Annaswamy. Adaptive estimation of discrete-time systems with nonlinear parametrization. *Automatica*, 36(12):1879–1887, 2000.

[48] J.-J. E. Slotine and J. K. Hedrick. Robust input-output feedback linearization. *International Journal of Control*, 57(5):1133–1139, 1993.

[49] J.-J. E. Slotine and W. Li. *Applied Nonlinear Control*. Prentice Hall, Englewood Cliffs, NJ, 1991.

[50] E. D. Sontag. *Mathematical Control Theory*. Springer-Verlag New York, Inc., 1990.

[51] D.G. Taylor, P.V. Kokotovic, R. Marino, and I. Kanellakopoulos. "Adaptive regulation of nonlinear systems with unmodeled dynamics". *IEEE Transactions on Automatic Control*, 34:405–412, 1989.

[52] M. Vidyasagar. *Nonlinear Systems Analysis.* Prentice Hall, 2nd edition, 1993.

[53] Y. Xu, H.-Y. Shum, T. Kanade, and J.-J. Lee. Parameterization and adaptive control of space robot systems. *IEEE Transactions on Aerospace and Electronic Systems*, 30(2):435–451, 1994.

[54] Ting-Jen Yeh. *Modeling, analysis and control of magnetically levitated rotating machines.* PhD thesis, MIT, Massachusetts, MA, 1996.

[55] T. Zhang, S.S. Ge, C.C. Hang, and T.Y. Chai. Adaptive control of first-order systems with nonlinear parameterization. *IEEE Transactions on Automatic Control*, 45(8):1512–1516, 2000.