Global parameter identification in systems with a sigmoidal activation function[1]

Aleksandar Kojić and Anuradha M. Annaswamy
Adaptive Control Laboratory, Department of Mechanical Engineering,
Massachusetts Institute of Technology, Cambridge, MA 02139;
email: alko@mit.edu, aanna@mit.edu
Corresponding author: Prof. A.M. Annaswamy, Rm 3-461b, MIT, 77 Mass. Ave, Cambridge, MA 02139

Abstract

Parameter identification in a 2-node network with sigmoidal activation functions is considered. Given the nonlinearity in the weights, standard estimation algorithms based on linear parametrization are inadequate tools for studying global parameter convergence. In this paper, we provide an alternative approach for studying parameter identification in the presence of sigmoidal parametrization. Conditions under which a simple back propagation algorithm can lead to global convergence are considered.

1 Introduction

Neural networks have occupied the attention of the engineering community since their introduction less than a decade ago. Both analytically and in practice, it has been proven that neural network architectures have the powerful ability to approximate a wide class of nonlinear functions and thus can be useful tools in many engineering applications [1, 2, 3]. The use of neural networks in identification and control of engineering systems has been intensely debated. Significant progress has been made regarding the statement of the problem, possible ways in which neural networks can be used for identification and control, and demonstrated in seminal numerical and experimental studies [4, 5, 6]. Several stability results have been derived in the literature concerning the use of identification and control (for example,[7, 8]). However, most of them are local in nature and/or include fairly restrictive conditions under which the stability is valid. It should be noted that despite the local stability nature, the actual demonstration in applications and numerical simulations reports just the contrary: Neural networks indeed serve as powerful numerical computational units that are capable of very good approximations of nonlinear maps and provide complex functionalities of estimation, control, and optimization over a large region of operation. In this paper, we take a first, modest step towards closing this glaring gap of explaining the true scope of operation of a neural network when used for nonlinear control. The main idea behind our approach is to directly address and exploit the distinguishing feature of nonlinear regression in neural networks and derive the underlying convergence and stability properties.

In our convergence analysis, we make explicit use of the nonlinear sigmoidal activation function and exploit its properties for establishing global convergence of gradient-based methods in neural network training. We focus on a simple 2-node neural network model to derive the underlying properties.

[1] This work was supported by a grant from the the Ford SRL University Research Program.

2 Main result

The system of interest is:

$$y = \sum_{i=1}^{2} g(\theta_i, u(t)) \tag{1}$$

where u is a scalar function of time, θ_1 and θ_2 are scalar parameters with unknown values, and $g(\theta, u)$ is a sigmoidal function given by:

$$g(\theta, u) = \frac{1 - e^{-u\theta}}{1 + e^{-u\theta}}. \tag{2}$$

We make the following assumptions regarding the system in (1).

(A1) $u > 0 \ \forall t$

(A2) $\theta_i > 0$ for $i = 1, 2$.

Our goal is to design an estimator that estimates θ_1 and θ_2 as $\widehat{\theta}_1$ and $\widehat{\theta}_2$, respectively, such that if

$$\begin{aligned} A &= (\theta_1, \theta_2),\ B = (\theta_2, \theta_1),\ L = \{A, B\} \\ \pi &= \left\{ (\widehat{\theta}_1, \widehat{\theta}_2) | \widehat{\theta}_1 > 0, \widehat{\theta}_2 > 0 \right\} \\ \widehat{\theta} &= (\widehat{\theta}_1, \widehat{\theta}_2) \end{aligned}$$

then for all $\widehat{\theta} \in \pi$ the convergence of $\widehat{\theta}$ to L is guaranteed.

The following estimator is proposed for such a task:

$$\widehat{y} = \sum_{i=1}^{2} g(u(t), \widehat{\theta}_i) \quad (3)$$

$$\dot{\widehat{\theta}}_i = -\widetilde{y}\, \frac{\partial g(\widehat{\theta}_i, u)}{\partial \widehat{\theta}_i} \qquad i = 1, 2 \quad (4)$$

where

$$\widetilde{y}(\widehat{\theta}_1, \widehat{\theta}_2, u) = g(\widehat{\theta}_1, u) + g(\widehat{\theta}_2, u) - g(\theta_1, u) - g(\theta_2, u) \quad (5)$$

In what follows, we assume that u is a time-varying switching sequence, which is defined as follows:

Definition 2.1 *A time varying function $u(t)$ is called a switching sequence function between u_1 and u_2 if*

(i) $u_1 \neq u_2$,

(ii) for any t_a such that $u(t_a) = u_1$, there exists a t_b, $t_b > t_a$ such that $u(t_b) = u_2$ and

(iii) for any t_l such that $u(t_l) = u_2$, there exists a t_m, $t_m > t_l$ such that $u(t_m) = u_1$.

(iv) for each t_1,t_2 such that $u(t_1) = u_1$, $u(t_2) = u_2$ there exists a ρ such that for all ϵ, $0 < \epsilon \leq \rho$, the sets $B_i = \{t \mid |t - t_i| \leq \epsilon, u(t) = u_i\}$, $i = 1, 2$ are nonempty.

It can be easily verified that the sigmoidal function in (2) has the following properties

(P1) $g(0, u) = 0$, $\lim_{\theta \to \infty} g(\theta, u) = 1$,

(P2) $\frac{\partial g(\theta, u)}{\partial \theta} > 0\ \forall \theta \in \mathbb{R}$,

(P3) $\theta\, \frac{\partial^2 g(\theta, u)}{\partial \theta^2} \leq 0, \forall \theta \in \mathbb{R}$.

An interesting property of symmetry of the estimator can be derived by examining (4). We note that eq. (4) can be rewritten as

$$\dot{\widehat{\theta}}_i\, (\widehat{\theta}_1, \widehat{\theta}_2, u) = -\widetilde{y}\, u\, \frac{2e^{-u\widehat{\theta}_i}}{\left(1 + e^{-u\widehat{\theta}_i}\right)^2}, \quad i = 1, 2 \quad (6)$$

Defining

$$E = \left\{(\widehat{\theta}_1, \widehat{\theta}_2) \mid \widehat{\theta}_1 = \widehat{\theta}_2\right\}, \quad (7)$$

it can be easily seen that the trajectories of (6) are symmetric with respect to E in π. Symmetry follows if for any value of u the following holds:

$$\dot{\widehat{\theta}}_1\, (\widehat{\theta}_1, \widehat{\theta}_2, u) = \dot{\widehat{\theta}}_2\, (\widehat{\theta}_2, \widehat{\theta}_1, u) \quad (8)$$

$$\dot{\widehat{\theta}}_2\, (\widehat{\theta}_1, \widehat{\theta}_2, u) = \dot{\widehat{\theta}}_1\, (\widehat{\theta}_2, \widehat{\theta}_1, u) \quad (9)$$

From eq. (5) we also have that $\widetilde{y}(\widehat{\theta}_1, \widehat{\theta}_2, u) = \widetilde{y}(\widehat{\theta}_2, \widehat{\theta}_1, u)$. This implies that eqs. (8) and (9) hold. Thus, the system is symmetric with respect to the set E, and therefore we focus all of our discussions that follow on H in π, where

$$H = \left\{(\widehat{\theta}_1, \widehat{\theta}_2) \mid \widehat{\theta}_1 < \widehat{\theta}_2\right\}. \quad (10)$$

The main result that we shall establish in this paper is that all trajectories that begin in $\pi \backslash E$ converge to L. [1] To establish this global result, the following set definitions are required:

$$M(u) = \left\{(\widehat{\theta}_1, \widehat{\theta}_2) \mid \widetilde{y}(\widehat{\theta}_1, \widehat{\theta}_2, u) = 0\right\} \quad (11)$$

$$M_1 = M(u_1) \qquad M_2 = M(u_2)$$

$$S_D = \{M(u) \mid u_1 \leq u \leq u_2\}$$

$$S_E = \pi \backslash S_D$$

$$S_D^h = S_D \bigcap H, \quad S_D^l = S_D \backslash S_D^h$$

$$S_D^{h-} = \left\{(\widehat{\theta}_1, \widehat{\theta}_2) \mid (\widehat{\theta}_1, \widehat{\theta}_2) \in S_D, \widehat{\theta}_2 > \theta_2\right\}$$

$$S_D^{h+} = \left\{(\widehat{\theta}_1, \widehat{\theta}_2) \mid (\widehat{\theta}_1, \widehat{\theta}_2) \in S_D, \widehat{\theta}_2 < \theta_2\right\}$$

We shall show that all trajectories that begin in $\pi \backslash E$ converge to S_D, and that the trajectories in S_D converge to L. Noting that M defines the boundaries of the set S_D, properties of M are crucial in establishing the main result. This is carried out in Lemma 2.2. Properties of S_D and S_E are stated in Lemmas 2.3 and 2.5. Finally, an aspect of differential geometry that is needed in the main result is stated in Lemma 2.6. Due to space considerations, the detailed proofs of lemmas and sublemmas are omitted here.

We first begin with some properties of the estimator in H. To derive these, the following quantities are useful.

$$\beta(u, \widehat{\theta}) = \frac{\dot{\widehat{\theta}}_2}{\dot{\widehat{\theta}}_1} \quad (12)$$

$$\beta_1 = \beta(u_1, \widehat{\theta}),\ \beta_2 = \beta(u_2, \widehat{\theta}), \quad \gamma = \frac{\beta_1}{\beta_2} \quad (13)$$

$$e_1(\widehat{\theta}) = \sum_{j=1}^{2} g(\widehat{\theta}_j, u_1) - \sum_{j=1}^{2} g(\theta_j, u_1) \quad (14)$$

$$e_2(\widehat{\theta}) = \sum_{j=1}^{2} g(\widehat{\theta}_j, u_2) - \sum_{j=1}^{2} g(\theta_j, u_2) \quad (15)$$

Lemma 2.1 *For the estimator in eqs. (3)-(4), the following properties hold $\forall\, \widehat{\theta} \in H$:*

(A) $\beta > 0, \gamma < 1$, *(B)* $\frac{\partial \beta}{\partial u} < 0$,

(C) $0 < c_{1i} < \frac{\partial e_i}{\partial \widehat{\theta}_1} < c_{2i}$ *for* $i = 1, 2$, *and for bounded* $\widehat{\theta}_1, u$; *and*

(D) $\frac{\partial \beta}{\partial \widehat{\theta}_2}|_{\widehat{\theta}_1, u=const} < 0$, $\frac{\partial \beta}{\partial \widehat{\theta}_1}|_{\widehat{\theta}_2, u=const} > 0$.

[1] Trajectories that begin in E, however, stay in E. Since these are of measure zero, the result is "almost global".

We now proceed to derive properties of M. In particular, we shall show that M_1 and M_2 intersect only at two points, A and B. This is proved by establishing the shape of M_i in the π plane for any given u. The properties of M_1 and M_2 are stated below in Lemmas 2.2(a) and (b).

Lemma 2.2 *(a) M_1 and M_2 are represented by two monotonically decreasing convex curves in π.*

(b) $M_1 \bigcap M_2 = L$.

We now introduce a metric d_i which we shall use to derive the convergence properties of the estimator. This is defined as follows. For any $P \in H$,

$$d_i(P) = e_i(P)^2, \quad v_i(P) = -e_i(P)\frac{\partial g(\widehat{\theta}_P, u_i)}{\partial \widehat{\theta}}, \quad i = 1, 2 \tag{16}$$

Thus, the functions d_1 and d_2 are two non-negative functions, which are zero on M_1 and M_2 respectively, and positive everywhere else. Therefore, they can be thought of as a measure of distance of a point to the curves M_1, and M_2, respectively. The functions v_1 and v_2 represent velocities that are associated with each point in π. As the trajectory moves in π, at each point P, it will have a velocity specified either by $v_1(P)$ or $v_2(P)$, depending on the particular value of the input u at the time the trajectory is at point P. Properties of e_i and d_i are presented in the following lemmas. Since their properties vary depending upon whether they are in S_D, S_E, or in the neighborhood of A, they are classified into three distinct Lemmas 2.3 –2.5.

Lemma 2.3 *Let $\left(\widehat{\theta}_1^0(u), \widehat{\theta}_2^0\right) \in M(u) \bigcap H$. For $i = 1, 2$, let $\widehat{\theta}_{1i}^0 = \widehat{\theta}_1^0(u_i)$. If $u_1 < u_2$, then*

(i) (a) *if $\underline{\widehat{\theta} \in S_D^{h-}}$, then $\widehat{\theta}_{11}^0 < \widehat{\theta}_1 < \widehat{\theta}_{12}^0$, and* (b) *if $\underline{\widehat{\theta} \in S_D^{h+}}$, then $\widehat{\theta}_{11}^0 > \widehat{\theta}_1 > \widehat{\theta}_{12}^0$,*

(ii) $\underline{\forall \widehat{\theta} \in S_D^h}$, $e_1(\widehat{\theta})\ e_2(\widehat{\theta}) < 0$,

(iii) $\underline{\forall \widehat{\theta} \in S_E \bigcap H}$, $e_1(\widehat{\theta}) e_2(\widehat{\theta}) > 0$,

Lemma 2.4 *If $\widehat{\theta} \in S_E$, then the following holds*

$$\frac{d}{dt} d_i(\widehat{\theta}(t)) < 0, \qquad i = 1, 2 \tag{17}$$

Lemma 2.5 *Let $N_\delta(A)$ be the δ neighborhood of point A defined as $N_\delta(A) = \{Q \mid d(Q, A) < \delta\}$, with $d(Q, A)$ being a Euclidean distance metric (see [9]) between two points Q and A. For all $P \notin N_\delta(A) \bigcup N_\delta(B)$, there exist $\epsilon_2 > \epsilon_1 > 0$ such that*

$$d_i(P) < \epsilon_1 \rightarrow d_j(P) > \epsilon_2; \quad i, j \in \{1, 2\}, \text{ and } i \neq j. \tag{18}$$

The following lemma is useful in describing the properties of the system trajectories in S_D.

Lemma 2.6 *Let K_1 and K_2 be two curves in the $O\xi\eta$ coordinate system described by $\eta_1 = \psi_1(\xi)$ and $\eta_2 = \psi_2(\xi)$, respectively. Let $K = K_1 \bigcap K_2$. If*

(i) the functions ψ_1 and ψ_2 are twice differentiable, with bounded derivatives ψ_1', ψ_1'', and ψ_2', ψ_2''; and

(ii) $\psi_1'(\xi) > \psi_2'(\xi)$ for all ξ such that $\psi_1(\xi) = \psi_2(\xi)$,

then (A) K is a set consisting of at most one point $Q(q, \psi_1(q))$, and (B) $\forall \xi > q$, $\psi_1(\xi) > \psi_2(\xi)$ and $\forall\ \xi < q$, $\psi_1(\xi) < \psi_2(\xi)$.

We now establish the main result that L is a globally attractive equilibrium point for almost all points in π.

Theorem 2.1 *If $u(t)$ is a switching sequence function between u_1 and u_2, under assumptions* (A1)-(A2)*, the solutions $\widehat{\theta}_i(t)$ of (3)-(4) satisfy the following:*

(i) if $\widehat{\theta}(0) \in E$, then $\widehat{\theta}(t) \in E\ \forall t \geq 0$.

(ii) if $\widehat{\theta}(0) \in \pi \backslash E$, then $\widehat{\theta}(t)$ converges to L as $t \rightarrow \infty$.

Proof:

The proof of (i) is immediate since $\forall \widehat{\theta} \in E$, we have that $\dfrac{\partial g(\widehat{\theta}_1, u)}{\partial \widehat{\theta}_1} = \dfrac{\partial g(\widehat{\theta}_2, u)}{\partial \widehat{\theta}_2}$, and it follows that $\dot{\widehat{\theta}}_1 = \dot{\widehat{\theta}}_2$. Hence, $\widehat{\theta}(0) \in E$ implies that $\widehat{\theta}(t) \in E$ for all $t \geq 0$.

(ii) To prove (ii), we proceed in the following three steps.

Step 1: Trajectories starting in $\pi \backslash E$ either converge to L as $t \rightarrow \infty$ or enter S_D (see figure 1) in a finite time T.

Step 2: S_D is invariant.

Step 3: Trajectories that start in S_D converge to L.

For ease of exposition, Figure 1 shows the nature of M_1, M_2, S_D, and S_E in π for $\widehat{\theta} \in \mathbb{R}^2$.

Proof of Step 1: We prove Step 1 by showing that all trajectories that start in S_E enter the set $S_D \backslash \{N_\delta(A) \bigcup N_\delta(B)\}$ in a finite time. Since δ is arbitrary, this implies that Step 1 holds. Suppose that the trajectories always remain in S_E. Since d_1 and d_2 are decreasing along the system trajectories, there exists a time instant t_1 such that $d_1(\widehat{\theta}(t_1)) < \epsilon_1$, and hence from Lemma 2.5, $d_2(\widehat{\theta}(t_2)) > \epsilon_2$. If trajectories always remain in S_E, d_2 is decreasing, and hence, there exists a $t_2 > t_1$ such that $d_2(\widehat{\theta}(t_2)) < \epsilon_1$. But, this implies that $d_1(\widehat{\theta}(t_2)) > \epsilon_2$. This is a contradiction, since in S_E, d_1 is decreasing as well.

Proof of Step 2: Let

$$(i)\widehat{\theta}(t_0) \in S_D \quad \text{and} \quad (ii) \quad \widehat{\theta}(t_2) \in S_E \quad \text{at} \quad t_2 > t_0. \tag{19}$$

From Lemma 2.3(ii) and (iii), we have that

$$\begin{aligned} e_1\left(\widehat{\theta}(t_0)\right) e_2\left(\widehat{\theta}(t_0)\right) &< 0 \\ e_1\left(\widehat{\theta}(t_2)\right) e_2\left(\widehat{\theta}(t_2)\right) &> 0. \end{aligned}$$

This implies that e_i must have changed sign over the interval $[t_0, t_2]$ for some $i \in 1, 2$. Hence,

$$e_i(\widehat{\theta}(t_1)) = 0 \qquad \text{for some} \quad t_1 \in [t_0, t_2], \tag{20}$$

and $\widehat{\theta}(t) \in S_E$ for $t \in [t_1, t_2]$. Let (20) be true for $i = 1$. From (16),(19), and (20) it follows that there exists a $\epsilon > 0$ such that

$$(i) \;\; d_1(\widehat{\theta}(t_1)) = 0 \qquad\qquad (ii) \;\; d_1(\widehat{\theta}(t_2)) > \epsilon. \tag{21}$$

which in turn implies that d_1 increases in S_E. This contradicts Lemma 2.4. Hence no $t_2 > t_0$ exists such that $\widehat{\theta}(t_2) \in S_E$, proving Step 2.

Proof of Step 3: We now show that trajectories starting at any point $P \in S_D$ converge to L. Since the system behavior is symmetric with respect to E, we examine only the trajectories that start in S_D^h, and show that they converge to A.

From Lemma 2.4, it follows that if $u(t)$ is kept constant either at u_1 or at u_2 for all $t \geq T$, then the trajectories will converge to a point either in M_1, or in M_2, respectively. We define the solution of Eq. (6) for $u = u_1$, an initial value $\widehat{\theta}_0(t) \in S_D^h$, and for all $\tau \geq t$ as $\widehat{\theta}(\tau, \widehat{\theta}_0(t), u_1)$ and limit points $N_1(t)$ and $N_2(t)$ as

$$N_i(t) \;\; = \;\; \lim_{\tau \to \infty} \widehat{\theta}\left(\tau; \widehat{\theta}_0(t); u_i\right), \quad i = 1, 2 \tag{22}$$

and denote their individual components as $N_1(t) = [\widehat{\theta}_{11}(t) \;\; \widehat{\theta}_{21}(t)]^T$, $N_2(t) = [\widehat{\theta}_{12}(t) \;\; \widehat{\theta}_{22}(t)]^T$. Let the sets $C_1(t)$ and $C_2(t)$ represent the sequence of points in S_D^h along which the trajectory converges to $N_1(t)$ and $N_2(t)$, respectively, for $u = u_1$ and $u = u_2$. That is,

$$C_i(t) \;\; = \;\; \left\{ P \,|\, P \in S_D^h, \, \lim_{\tau \to \infty} \widehat{\theta}(\tau, P, u_i) = N_i(t) \right\}, \quad i = 1, 2. \tag{23}$$

It is worth noting that $\widehat{\theta}_0(t)$ monotonically approaching A in S_D^{h-} is equivalent to $\widehat{\theta}_{11}(t)$ increasing and $\widehat{\theta}_{21}(t)$ decreasing montonically with $u = u_1$, as well as $\widehat{\theta}_{12}(t)$ increasing and $\widehat{\theta}_{22}(t)$ decreasing monotonically with $u = u_2$. We show that trajectories starting in S_D^{h-} approach A by showing that $\widehat{\theta}_{1i}$ increases and $\widehat{\theta}_{2i}$ decreases for $i = 1, 2$ as $t \to \infty$. In a similar manner, it can be shown that all trajectories starting in S_D^{h+} approach A. Due to the symmetry of the problem, it then would follow that all trajectories starting in S_D^l approach B.

Convergence of trajectories to A in S_D^h is established by considering the two mutually exclusive and collectively exhaustive cases: (a) $\widehat{\theta} \in S_D^{h-}$ and (b) $\widehat{\theta} \in S_D^{h+}$. Using the definitions of S_D^{h-} and S_D^{h+}, it follows that $S_D^h = S_D^{h-} \bigcup S_D^{h+} \bigcup M_1 \bigcup M_2$ and $S_D^{h-} \bigcap S_D^{h+} = \emptyset$, and that A is a limit point for both S_D^{h-} and S_D^{h+}.

Case (a) $\widehat{\theta} \in S_D^{h-}$: In this case, it can be shown that $\widehat{\theta}_{21}$ and $\widehat{\theta}_{22}$ are decreasing functions of time. Let $\widehat{\theta}(t_0) \in S_D^{h-}$. Then, it follows that $N_1(t_0) \in M_1$ and $N_2(t_0) \in M_2$, and that $\widehat{\theta}(t, \widehat{\theta}(t_0), u)$ moves along $C_1(t_0)$ if $u(t) = u_1 \; \forall \, t \geq t_0$, and along $C_2(t_0)$ if $u(t) = u_2 \; \forall \; t \geq t_0$. From the definition of β in (12), it follows that the slope of the tangents to the curves $C_1(t_0)$ and $C_2(t_0)$ at any $\widehat{\theta}$ are given by β_1 and β_2, respectively. Since β_1 and β_2 are bounded, it follows that there exist functions $\psi_1(\widehat{\theta}_1)$ and $\psi_2(\widehat{\theta}_1)$ which can represent the curves $C_1(t_0)$ and $C_2(t_0)$ in S_D,respectively, where the slopes $\psi_1'(\widehat{\theta}_1)$ and $\psi_2'(\widehat{\theta}_1)$ are given by β_1 and β_2, respectively. Since $u_1 < u_2$, Lemma 2.1(B) implies that

$$\psi_2'(\widehat{\theta}_1) < \psi_1'(\widehat{\theta}_1) \; \forall \widehat{\theta}_1 \text{ such that } \; \psi_1(\widehat{\theta}_1) = \psi_2(\widehat{\theta}_1) \tag{24}$$

That is, ψ_1 and ψ_2 satisfy the conditions of Lemma 2.6, and hence (A) and (B) of Lemma 2.6 hold. Suppose that at time instant t_0, $\widehat{\theta}(t_0) = P_0(q, p) \in S_D^{h-}$. From the definition of a switching sequence, there exists a time interval $T_2 = [t_0, t_1] \subset \Omega_2$. Thus, on the interval T_2 the trajectory moves along $C_2(t_0)$. Since $\widehat{\theta} \in S_D^{h-}$, from Lemma 2.3-$(i$-$a)$, we have that $\widehat{\theta}_1 < \widehat{\theta}_{12}^0$. From Lemma 2.1$(C)$ we have that $e_2(.)$ is monotonically increasing with $\widehat{\theta}_1$, and thus we have that $e_2(\widehat{\theta}) < 0$. Therefore, for all $t \in T_2$, $\dot{\widehat{\theta}}_1 \; (t) = -e_2(\widehat{\theta}) \; \dfrac{\partial g(\widehat{\theta}_1, u_2)}{\partial \widehat{\theta}_1} > 0$. Hence, on T_2, $\widehat{\theta}_1 > q$. From Lemma 2.6(B) we have that

$$\psi_2(\widehat{\theta}_1(t)) < \psi_1(\widehat{\theta}_1(t)) \quad \forall t \in T_2. \tag{25}$$

Let u switch from u_2 to u_1 at $t = t_1$, and let $\widehat{\theta}(t_1) = P_3(q_3, p_3)$. From the property of the switching sequence, there exists a $t_2 > t_1$ such that $T_3 = [t_1, t_2] \subset \Omega_1$. Proceeding as above, we note that from Lemma 2.3-$(i$-$a)$ and Lemma 2.1(c) we have that $e_1(\widehat{\theta}) > 0 \; \forall \widehat{\theta} \in S_D^{h-}$. Hence,

$$\widehat{\theta}_1 < q_3 \quad \forall t \in T_3. \tag{26}$$

Let $\psi_3(\widehat{\theta}_1)$ denote the curve $C_1(t_1)$ which represents the trajectory of $\widehat{\theta}(t)$ for $t \in T_3$. Define a curve $\psi_4(\widehat{\theta}_1)$ as

$$\psi_4(\widehat{\theta}_1) = \psi_1(\widehat{\theta}_1) - (\psi_1(q_3) - \psi_2(q_3)) \tag{27}$$

From the definition of q_3 and $\psi_3(.)$ and eq.(27), we have that

$$\psi_4(\widehat{\theta}_1(t_1)) = \psi_2(q_3) = \psi_3(q_3). \tag{28}$$

Also, from (25) and (28), it follows that

$$\psi_3(q_3) < \psi_1(q_3). \tag{29}$$

Since

$$\begin{aligned} \psi_3'(q_3) \;\; &= \;\; \beta(q_3, \psi_3(q_3), u_1) \\ \psi_1'(q_3) \;\; &= \;\; \beta(q_3, \psi_1(q_3), u_1) \end{aligned} \tag{30}$$

it follows from Lemma 2.1(D) and (29) that

$$\psi_1'(q_3) < \psi_3'(q_3). \tag{31}$$

Differentiating (27) with respect to $\widehat{\theta}_1$, we obtain that $\psi_4'(\widehat{\theta}_1) = \psi_1'(\widehat{\theta}_1)$. Eq. (31) implies that

$$\psi_4'(q_3) < \psi_3'(q_3). \tag{32}$$

Eqs. (32) and (28) imply that Lemma 2.6 holds. Therefore, using Lemma 2.6 and (26) it follows that

$$\psi_3(\widehat{\theta}_1(t)) < \psi_4(\widehat{\theta}_1(t)) \quad \forall t \in T_3 \tag{33}$$

and hence

$$\psi_3(\widehat{\theta}_1(t)) < \psi_1(\widehat{\theta}_1(t)) \quad \forall t \in T_3. \tag{34}$$

Since $N_1(t_1)$ is the limit point of $C_1(t_1)$, and $N_1(t_0)$ is a limit point of $C_1(t_0)$ it follows that that $\widehat{\theta}_{21}(t_1) < \widehat{\theta}_{21}(t_0)$. This relation holds for arbitrary t_0, t_1, and t_2, and hence it follows that $\widehat{\theta}_{21}$ is a decreasing function of time on T_2. A similar analysis can be carried out to conclude that $\widehat{\theta}_{22}$ is a decreasing function of time on T_3.

Case (b):$\widehat{\theta} \in S_D^{h+}$: Similar steps and arguments as presented in case (a) can be used to conclude that in this case $\widehat{\theta}_{21}(t)$ and $\widehat{\theta}_{22}(t)$ are increasing functions of time.

It is interesting to note that, because the set S_D^h is invariant and bounded, $\widehat{\theta}_{11}$ and $\widehat{\theta}_{12}$ remain bounded in both cases *(a)* and *(b)*. Since

$$\begin{aligned} \theta_2 &= \min_{\widehat{\theta}_2} P(\widehat{\theta}_1, \widehat{\theta}_2) \ \ \forall P \in S_D^{h-} \\ \theta_2 &= \max_{\widehat{\theta}_2} P(\widehat{\theta}_1, \widehat{\theta}_2) \ \ \forall P \in S_D^{h+} \end{aligned}$$

it then follows that when u is a switching sequence, the system will converge to the point A starting from anywhere in the set S_D^h.

In summary, for any starting point in S_E, the system converges in finite time to S_D. Starting from any point in S_D, the system converges asymptotically to either point A or B. Thus, the theorem is established. ■

Theorem 2.1 shows that in $\pi \backslash E$, all trajectories of (4) converge to L. This was established by making use of the properties of $M(u)$ which enabled in turn to show that all trajectories converge to S_D and that trajectories in S_D converge to L.

At first glance, the problem solved here appears deceptively simple. Parameter convergence in a nonlinearly parameterized system (1) has been established for very low dimensionality, i.e.$\theta \in \mathbb{R}^2$. It might then be questioned as to why such elaborate discussions and numerous lemmas and sublemmas were needed to prove this seemingly simple result. We submit that these discussions and properties were essential for proving the main result.

The complexity mainly begins with the nonlinear dependence of g in θ. The first implication of g is symmetry; both A and B are stable equilibrium points of the system in (8), and E represents a separatrix. As a result, a simple approach that consists of a quadratic positive definite function

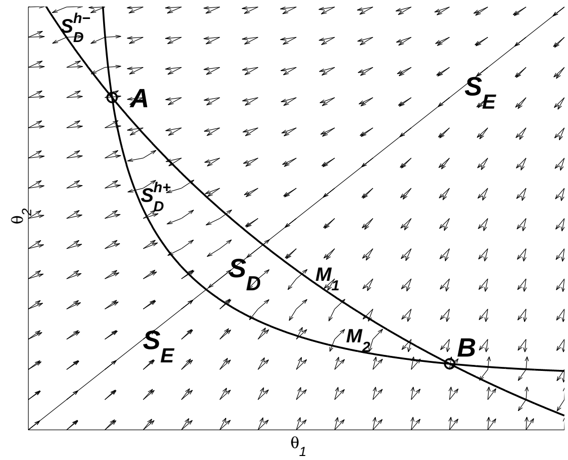

Figure 1: Phase plot for a two-parameter system with sigmoidal parameterization using two different values of u.

V and showing that its derivative is sign-definite is ruled out. The nonlinear dependence of g on θ also implies the existence of manifolds with which are associated regions with certain invariant properties. What has been carried out above is precisely the characterization of these manifolds, given by E, M_1, and M_2, and invariant regions given by S_D.

References

[1] P. J. Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences.* PhD thesis, Harvard University, 1974.

[2] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

[3] J. Park and I. W. Sandberg. Universal approximation using radial-basis function networks. *Neural Computation*, 3:246–257, 1991.

[4] K.S. Narendra and K. Parthasarathy. Gradient methods for the optimization of dynamical systems containing neural networks. *IEEE Transactions on Neural Networks*, 2:252–262, 1991.

[5] R.M. Sanner and J-J.E. Slotine. Gaussian networks for direct adaptive control. *IEEE Transactions on Neural Networks*, 3(6):837–863, November 1992.

[6] G. V. Puskorius and L. A. Feldkamp. Neurocontrol of nonlinear dynamical systems with Kalman filter trained recurrent networks. *IEEE Transactions on Neural Networks*, 5(2):279–297, 1994.

[7] S. Yu and A.M. Annaswamy. Stable neural controllers for nonlinear dynamic systems. *Automatica*, 34:669–679, May 1998.

[8] S. Jagannathan, F. L. Lewis, and O. Pastravanu. Discrete-time model reference adaptive control of nonlinear dynamical systems using neural networks. *International Journal of Control*, 64(2):217–230, 1996.

[9] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, Inc., 1976.