NEW STABILITY THEOREMS FOR AVERAGING AND THEIR APPLICATIONS TO THE CONVERGENCE ANALYSIS OF ADAPTIVE IDENTIFICATION AND CONTROL SCHEMES

by

L-C. Fu, M. Bodson, and S. S. Sastry

Memorandum No. UCB/ERL M85/21

14 March 1985
NEW STABILITY THEOREMS FOR AVERAGING AND THEIR APPLICATIONS TO THE CONVERGENCE ANALYSIS OF ADAPTIVE IDENTIFICATION AND CONTROL SCHEMES

by

L-C. Fu, M. Bodson, and S. S. Sastry

Memorandum No. UCB/ERL M85/21

14 March 1985

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720
New Stability Theorems for Averaging and Their Application to the Convergence Analysis of Adaptive Identification and Control Schemes

Li-Chen Fu, Marc Bodson and Shanker Sastry*
Electronics Research Laboratory
Department of Electrical Engineering & Computer Science
University of California, Berkeley CA 94720

ABSTRACT

We develop new stability theorems for the convergence analysis of a class of one and two-time scale time-varying nonlinear systems using averaging theory. These theorems are applied to a class of continuous time adaptive identifiers and model reference adaptive controllers to obtain estimates of the parameter rate of exponential convergence.

March 14, 1985

Keywords: Averaging, Stability Analysis, Lyapunov Converse Theorems, Adaptive Identification and Control, Parameter Convergence.
New Stability Theorems for Averaging and Their Application to the Convergence Analysis of Adaptive Identification and Control Schemes

Li-Chen Fu, Marc Bodson and Shanker Sastry

Electronics Research Laboratory
Department of Electrical Engineering & Computer Science
University of California, Berkeley CA 94720

1. Introduction

The method of averaging is concerned with differential equations of the form

\[ \dot{x} = \varepsilon f(t,x) \]  

and relates the properties of solutions of system (1.1) with solutions of the autonomous "averaged" system

\[ \dot{x}_a = \varepsilon f_{av}(x_a) \]

for sufficiently small values of the parameter \( \varepsilon \). The method was proposed originally by Kryloff-Bogoliuboff-Mitropolskii [1], reformulated by Hale [2,3], developed subsequently by Sethna [4,5] and stated in a geometric form in Arnold [6], Guckenheimer and Holmes [7]. These results constitute a generalization of classical singular perturbation techniques such as those in Hoppensteadt [8] and are an extremely important tool for the state space analysis of systems with multiple time scales. These results have been used extensively in mathematical physics. From our viewpoint, as control theorists, we feel that the technique bears the promise of evolving into a "frequency domain" technique for the state-space trajectory analysis of some classes of nonlinear systems— to be distinguished from the Volterra approach for input-output functional expansions for non-linear systems. This theme will be developed in later work.

The current paper has two sets of contributions:

(A) We develop new theorems for averaging. With the exception of [4,5], all the aforementioned references make the assumption of almost periodicity for the right hand side of (1.2). We relax this assumption in Section 2 of this paper. Our theorems are rather different in hypothesis...
and rather simpler than those of [4,5] in this regard. Another important contribution of Section 2 is to relate the stability of the averaged system (1.2) to the stability of the unaveraged system (1.1), using a converse theorem of Lyapunov. As such, these theorems are a considerable extension of the local stability theorems of Hale.

In Section 4, we extend all of these results to two-time scale state space systems and the results are generalizations in the sense mentioned above of those of Hale and Sethna.

(B) Our development of these theorems on averaging was heavily motivated by recent literature on the application of averaging techniques to adaptive control—notably the work of Krause, et al [9], Riedle and Kokotovic [10], Astrom [11]. Averaging methods have been more prevalent in the stochastic adaptive control literature, eg. Ljung [12] and the first attempts to apply averaging were made heuristically in [9], and increasingly rigorously in [11] and [10]. The primary focus of the efforts in [9-11] is to use averaging to explain instability mechanisms in adaptive control arising from unmodelled dynamics, a phenomenon popularized by Rohrs et al [13]. In this paper, we content ourselves with applying our results on averaging theory along with techniques of generalized harmonic analysis introduced in Boyd-Sastry [14]. We study convergence rates of adaptive identification schemes and linearized adaptive control schemes without unmodelled dynamics and in the presence of persistent excitation. Estimates of convergence rates are of interest in the determination of optimal input signals for identification. In earlier work (Bodson-Sastry [15]), we also showed how persistent excitation guarantees a margin of robustness to unmodelled dynamics and established connections between the rate of convergence of the adaptive schemes and their robustness margins. A more detailed study of instability theorems for averaging and their application to understanding the mechanism of slow drift instability pointed out by Riedle-Kokotovic [16] will be presented in forthcoming work.

The results of Section 3 on the application of averaging theory to obtaining estimates of the convergence rates for adaptive identifiers are to our knowledge new, while those of Section 5 on convergence rates for adaptive control schemes in the relative degree 1 case are a small generalization of the results of [10] with a somewhat different focus.

*After this manuscript was written, new and related work of Koo and Anderson [17] was communicated to us for system (1.1) with \( f(t, x) \) linear in \( x \), but with weaker conditions in the limit in (1.3).
2. Basic Averaging Theory

In this section, we consider differential equations of the form:

\[ \dot{x} = \epsilon f(t, x, \epsilon) \quad x(0) = x_0 \quad (2.1) \]

where \( x \in \mathbb{R}^n \), \( t \geq 0 \), \( 0 < \epsilon \leq \epsilon_0 \), and \( f \) is piecewise continuous with respect to time. We will concentrate our attention on the behavior of the solutions in some closed ball \( B_h \) of radius \( h \), centered at the origin.

For small \( \epsilon \), the variation of \( x \) with time is slow, as compared to the rate of time variation of \( f \). Such systems can be conveniently studied using the method of averaging (see e.g. [1], [3], [6], [7]). The theory relies on the assumption of the existence of the mean value of \( f(t, x, 0) \) defined by the limit:

\[ f_{av}(x) = \lim_{T \to \infty} \frac{1}{T} \int_{t}^{t+T} f(r, x, 0) dr \quad (2.2) \]

assuming that the limit exists uniformly in \( t \) and \( x \). This is formulated more precisely in the following definition:

Definition 2.1 Mean Value of a Function, Convergence Function

The function \( f(t, x, 0) \) is said to have mean value \( f_{av}(x) \) if there exists a continuous function \( \gamma(T): \mathbb{R}_+ \to \mathbb{R}_+ \), strictly decreasing, such that \( \gamma(T) \to 0 \) as \( T \to \infty \), and:

\[ \frac{1}{T} \int_{t}^{t+T} f(r, x, 0) dr - f_{av}(x) \leq \gamma(T) \quad (2.3) \]

for all \( t, T \geq 0 \), \( x \in B_h \).

The function \( \gamma(T) \) will be called the convergence function.

Note that the function \( f(t, x, 0) \) has mean value \( f_{av}(x) \) if and only if the function:

\[ d(t, x) = f(t, x, 0) - f_{av}(x) \quad (2.4) \]

has zero mean value.

The following definition ([20], p 7) will also be useful:

Definition 2.2 Class K Function

A function \( \alpha(e): \mathbb{R}_+ \to \mathbb{R}_+ \) belongs to class \( K \) ( \( \alpha(e) \in K \) ), if it is continuous, strictly increasing, and \( \alpha(0) = 0 \).

It is common, in the literature on averaging, to assume that the function \( f(t, x, \epsilon) \) is periodic in \( \epsilon \), or almost periodic in \( t \). Then, the existence of the mean value is guaranteed, without
further assumption ([3], theorem 6, p 344). We do not make the assumption of (almost) periodicity, but consider instead the assumption of the existence of the mean value as the starting point of our analysis.

Note that if the function $d(t,x)$ is periodic in $t$, and is bounded, then the integral of the function $d(t,x)$ is also a bounded function of time. This is equivalent to saying that there exists a convergence function $\gamma(T) = a/T$ (i.e. of the order of $1/T$) such that (2.3) is satisfied. On the other hand, if the function $d(t,x)$ is bounded, but is only required to be almost periodic, then the integral of the function $d(t,x)$ need not be a bounded function of time, even if its mean value is zero ([3], p 346). The function $\gamma(T)$ is bounded (by the same bound as $d(t,x)$), and converges to zero as $T \to \infty$, but the convergence function need not be bounded by $a/T$ as $T \to \infty$ (it may be of order $1/\sqrt{T}$ for example). In general, a zero mean function need not have a bounded integral, although the converse is true. In this paper, we do not make the distinction between the periodic, and the almost periodic case, but we do distinguish the bounded integral case from the general case, and indicate the importance of the function $\gamma(T)$ in the subsequent development.

Assuming the existence of the mean value for the original system (2.1), the averaged system is defined to be:

$$\dot{x}_{av} = \epsilon f_{av}(x_{av}) \quad x_{av}(0) = x_0 \quad (2.5)$$

Note that the averaged system is autonomous and, for $T$ fixed and $\epsilon$ varying, the solutions over intervals $[0, T/\epsilon]$ are identical, modulo a simple time scaling by $\epsilon$.

We address the following two questions:

(i) the closeness of the response of the original and averaged systems,

(ii) the relationships between the stability properties of the two systems.

To compare the solutions of the original and of the averaged system, it is convenient to transform the original system in such a way that it becomes a perturbed version of the averaged system. An important lemma that leads to this result is attributed to Bogoliuboff and Mitropol'skii ([1], p 450, and [3], lemma 4, p 346). We state a generalized version of this lemma.

**Lemma 2.1** Approximate Integral of a Zero Mean Function

If $d(t,x): \mathbb{R}_+ \times B_1 \to \mathbb{R}^n$ is a bounded function, piecewise continuous with respect to $t$, and has zero mean value with convergence function $\gamma(T)$,

Then: There exists $\xi(\epsilon) \in K$, and a function $w_{t}(t,x): \mathbb{R}_+ \times B_1 \to \mathbb{R}^n$ such that:

$$||w_{t}(t,x)|| \leq \xi(\epsilon) \quad (2.6)$$

$$||\frac{\partial w_{t}(t,x)}{\partial t} - d(t,x)|| \leq \xi(\epsilon) \quad (2.7)$$
for all $t \geq 0$, $z \in B_A$. Moreover, $w_0(0, z) = 0$, for all $z \in B_A$.

If, moreover: $\gamma(T) = \sigma / T'$ for some $\sigma \geq 0$, $r \in (0, 1]$, 

Then: The function $\xi(\epsilon)$ can be chosen to be $2a \epsilon'$. 

The proof of Lemma 2.1 is provided in the appendix. The construction of the function $w_r(t, x)$ is identical to that in [2.1], but the proof of (2.6), (2.7) is different, and leads to the relationship between the convergence function $\gamma(T)$ and the function $\xi(\epsilon)$.

The main point of Lemma 2.1 is that, although the exact integral of $d(t, z)$ may be an unbounded function of time, there exists a bounded function $w_r(t, x)$, whose first partial derivative with respect to $t$ is arbitrarily close to $d(t, z)$. Although the bound on $w_r(t, x)$ may increase as $\epsilon \to 0$, it increases slower than $1/\epsilon$, as indicated by (2.6).

It is necessary to obtain a function $w_r(t, x)$, as in Lemma 2.1, that has some additional smoothness properties. A useful lemma is given by Hale in [3] (lemma 5, p 349). For the price of additional assumptions on the function $d(t, z)$, the following lemma leads to stronger conclusions that are useful in the sequel.

**Lemma 2.2 Smooth Approximate Integral of a Zero Mean Function**

If: $d(t, z): R_+ \times B_A \to R^n$ is piecewise continuous with respect to $t$, has bounded and continuous first partial derivatives with respect to $x$, and $d(t, 0) = 0$ for all $t \geq 0$. Moreover, $d(t, z)$ has zero mean value, with convergence function $\gamma(T)||z||$, and $\frac{\partial d(t, z)}{\partial z}$ has zero mean value, with convergence function $\gamma(T)$,

Then: There exists $\xi(\epsilon) \in K$, and a function $w_r(t, z): R_+ \times B_A \to R^n$, such that:

$$||\xi w_r(t, z)|| \leq \xi(\epsilon)||z|| \quad (2.8)$$

$$||\frac{\partial w_r(t, x)}{\partial t} - d(t, x)|| \leq \xi(\epsilon)||z|| \quad (2.9)$$

$$||\epsilon \frac{\partial w_r(t, z)}{\partial z}|| \leq \xi(\epsilon) \quad (2.10)$$

for all $t \geq 0$, $z \in B_A$. Moreover, $w_0(0, z) = 0$, for all $z \in B_A$.

If, moreover: $\gamma(T) = \sigma / T'$ for some $\sigma \geq 0$, $r \in (0, 1]$, 

Then: the function $\xi(\epsilon)$ can be chosen to be $2a \epsilon'$.

The proof of Lemma 2.2 is provided in the appendix. The difference from Lemma 2.1 is in the condition on the partial derivative of $w_r(t, x)$ with respect to $x$ in (2.10), and the dependence on $||z||$ in (2.8), (2.9). These results will be necessary to derive the following theorems.
Note that if the original system is linear, i.e.:
\[ \dot{z} = A(t)\dot{z} \quad z(0) = z_0 \]  
for some \( A(t) : R^+ \rightarrow R^{n \times n} \), then the main assumption of Lemma 2.2 is that there exists \( A_{av} \) such that \( A(t) - A_{av} \) has zero mean value.

The following assumptions will henceforth be in effect.

(A1) \( z = 0 \) is an equilibrium point of system (2.1), i.e. \( f(t,0,0) = 0 \) for all \( t \geq 0 \). \( f(t,x,\epsilon) \) is Lipschitz in \( x \), i.e.:
\[ ||f(t,x_1,\epsilon) - f(t,x_2,\epsilon)|| \leq l_1||x_1 - x_2|| \]  
for all \( t \geq 0 \), \( x_1, x_2 \in B_k \), \( \epsilon \leq \epsilon_0 \).

(A2) \( f(t,x,\epsilon) \) is Lipschitz in \( \epsilon \), linearly in \( z \), i.e.:
\[ ||f(t,x,\epsilon_1) - f(t,x,\epsilon_2)|| \leq l_2||\epsilon|||\epsilon_1 - \epsilon_2| \]  
for all \( t \geq 0 \), \( x \in B_k \), \( \epsilon_1, \epsilon_2 \leq \epsilon_0 \).

(A3) \( f_{av}(0) = 0 \), and \( f_{av}(z) \) is Lipschitz in \( z \), i.e.:
\[ ||f_{av}(x_1) - f_{av}(x_2)|| \leq l_{av}||x_1 - x_2|| \]  
for all \( x_1, x_2 \in B_k \).

(A4) the function \( d(t,z) = f(t,x,0) - f_{av}(z) \) satisfies the conditions of Lemma 2.2.

Lemma 2.3 Perturbation Formulation of Averaging

If the systems (2.1), and (2.5) satisfy assumptions (A1)-(A4),

Then there exist functions \( w(t,z) \) and \( \xi(\epsilon) \), as in Lemma 2.2, and a transformation of the form:
\[ z = z + \epsilon w(t,z) \]  
under which system (2.1) becomes:
\[ \dot{z} = \epsilon f_{av}(z) + \epsilon p(t,z,\epsilon) \quad z(0) = z_0 \]  
where \( p(t,z,\epsilon) \) satisfies:
\[ ||p(t,z,\epsilon)|| \leq \psi(\epsilon)||z|| \]  
for some \( \psi(\epsilon) \in K \), \( \epsilon_z > 0 \), and for all \( \epsilon \leq \epsilon_z \). Further, \( \psi(\epsilon) \) is of the order of \( \epsilon + \xi(\epsilon) \).
Comments

The proof of Lemma 2.8 is provided in the appendix. A similar lemma can be found in [3] (lemma 3.2, p 192). Inequality (2.17) is a Lipschitz type of condition on \( p(t,x,\epsilon) \), which is not found in [3], and results from the stronger conclusions of Lemma 2.2.

Lemma 2.8 is fundamental to the theory of averaging presented hereafter. It separates the error in the approximation of the original system by the averaged system \( x-x_{av} \) into two components: \( x-z \) and \( z-x_{av} \). The first component results from a pointwise (in time) transformation of variable. This component is guaranteed to be small by inequality (2.8). For \( \epsilon \) sufficiently small \( (\epsilon \leq \epsilon_1) \), the transformation \( z \rightarrow x \) is invertible, and as \( \epsilon \rightarrow 0 \), it tends to the identity transformation. The second component is due to the perturbation term \( p(t,x,\epsilon) \). Inequality (2.17) guarantees that this perturbation is small as \( \epsilon \rightarrow 0 \).

At this point, we can relate the convergence of the function \( \gamma(T) \) to the order of the two components of the error \( x-x_{av} \) in the approximation of the original system by the averaged system. The relationship between the functions \( \gamma(T) \) and \( \xi(\epsilon) \) was indicated in Lemma 2.1. Lemma 2.8 relates the function \( \xi(\epsilon) \) to the error due to the averaging. If \( d(t,x) \) has a bounded integral (i.e. \( \gamma(T) = 1/T \)), then both \( x-z \) and \( p(t,x,\epsilon) \) are of the order of \( \epsilon \) with respect to the main term \( f_{av}(x) \). In general, these terms go to zero as \( \epsilon \rightarrow 0 \), but possibly more slowly than linearly (as \( \sqrt{\epsilon} \) for example). The proof of Lemma 2.1 provides a direct relationship between the order of the convergence to the mean value, and the order of the error terms.

We now focus attention on the approximation of the original system by the averaged system. Consider first the following assumption:

\[ (A5) \quad x_0 \text{ is sufficiently small so that, for fixed } T, \text{ and some } h' < h, x_{av}(t) \in B_{h'} \text{ for all } t \in [0,T/\epsilon] \text{ (this is possible, from (A3)).} \]

Theorem 2.4 Basic Averaging Theorem

If: The original system (2.1), and the averaged system (2.5) satisfy assumptions (A1)-(A5),

Then: There exists \( \psi(\epsilon) \) as in Lemma 2.8 such that, given \( T \geq 0 \):

\[ ||x(t)-x_{av}(t)|| \leq \psi(\epsilon) b_T \]  

(2.18)

for some \( b_T, \epsilon_T > 0 \), and for all \( t \in [0,T/\epsilon] \), \( \epsilon \leq \epsilon_T \).

Proof: From Lemma 2.2 and Lemma 2.8, we have that:

\[ ||x-z|| \leq \xi(\epsilon)||x|| \leq \psi(\epsilon)||x|| \]  

(2.19)

for \( \epsilon \leq \epsilon_1 \). On the other hand, we have that:
\[
\frac{d}{dt}(z-x) = \epsilon(f(x,t) - f(z_{av},t)) + \epsilon p(t, z, \epsilon) \quad z(0) - z_{av}(0) = 0
\]  
(2.20)

for all \(t \in [0, T/\epsilon]\), \(z_{av} \in B_h\), \(h < \epsilon\). We will now show that, on this time interval, and for as long as \(z, x \in B_h\), the errors \((z - x_{av})\) and \((z - z_{av})\) can be made arbitrarily small by reducing \(\epsilon\).

Integrating (2.19):

\[
||z(t) - z_{av}(t)|| \leq \epsilon \int_0^t ||z(\tau) - z_{av}(\tau)|| d\tau + \epsilon \int_0^t ||z(\tau)|| d\tau
\]  
(2.21)

Using the Generalized Bellman-Gronwall Lemma (see appendix):

\[
||z(t) - z_{av}(t)|| \leq \psi(\epsilon) \int_0^t ||z(\tau)|| e^{\epsilon \int_0^\tau - 1} d\tau \leq \psi(\epsilon) t
\]  
(2.22)

Combining these results:

\[
||z(t) - z_{av}(t)|| \leq ||z(t) - z(t)|| + ||z(t) - z_{av}(t)||
\]

\[
\leq \psi(\epsilon) ||z_{av}(t)|| + (1 + \psi(\epsilon)) ||z(t) - z_{av}(t)||
\]

\[
\leq \psi(\epsilon) (h + (1 + \psi(\epsilon)) \alpha_T)
\]

\[
:= \psi(\epsilon) b_T
\]  
(2.23)

By assumption, \(||z_{av}(t)|| \leq h < \epsilon\). Let \(\epsilon_T\) (with \(0 < \epsilon_T \leq \epsilon_t\)) such that \(\psi(\epsilon_T) b_T < \epsilon - h^\prime\). It follows, from a simple contradiction argument, that the estimate in (2.23) is valid for all \(t \in [0, T/\epsilon]\), whenever \(\epsilon \leq \epsilon_T\).

Comments

Theorem 2.4 establishes that the trajectories of the original and the averaged system are arbitrarily close on intervals \([0, T/\epsilon]\), as \(\epsilon\) is reduced. The error is of the order of \(\psi(\epsilon)\), and the order is related to the order of convergence of \(\gamma(T)\). If \(d(t, z)\) has a bounded integral (i.e. \(\gamma(T) \sim 1/T\)), then the error is of the order of \(\epsilon\).

It is important to remember that, although the intervals \([0, T/\epsilon]\) are unbounded, Theorem 2.4 does not state that:

\[
||z(t) - z_{av}(t)|| \leq \psi(\epsilon) b
\]  
(2.24)

for all \(t \geq 0\), and some \(b\). Consequently, Theorem 2.4 does not allow us to relate the stability of the original and of the averaged system. This relationship is investigated in Theorem 2.5, after a preliminary definition.
**Definition 2.3 Exponential Stability, Rate of Convergence**

The equilibrium point \( z=0 \) of a differential equation is said to be *exponentially stable*, with *rate of convergence* \( \alpha (\alpha > 0) \), if:

\[
\|x(t)\| \leq m\|x(t_0)\| e^{-\alpha (t-t_0)}
\]

(2.25)

for all \( t \geq t_0 \geq 0 \), \( x(t_0) \in B_{h_0} \), and some \( m \geq 1 \).

We assume that \( h_0 < h/m \), so that all trajectories are guaranteed to remain in \( B_h \).

**Theorem 2.5 Exponential Stability Theorem**

If:
The original and averaged systems satisfy assumptions (A1)-(A5), the function \( f_{av}(z) \) has continuous and bounded first order partial derivatives in \( z \), and \( z=0 \) is an exponentially stable equilibrium point of the averaged system,

Then:

There exists \( \epsilon_2 > 0 \) such that the equilibrium point \( z=0 \) of the original system is exponentially stable for all \( \epsilon \leq \epsilon_2 \).

**Proof:**

The proof relies on a converse theorem of Lyapunov for exponentially stable systems (see for example [20], p 273). Under the hypotheses, there exists a function \( v(z_{av}):R^n \rightarrow R_+ \), and strictly positive constants \( \alpha_1, \alpha_2, \alpha_3, \alpha_4 \) such that, for all \( z_{av} \in B_{h_0} \):

\[
\alpha_1 \|z_{av}\|^2 \leq v(z_{av}) \leq \alpha_2 \|z_{av}\|^2
\]

(2.26)

\[
\dot{v}(z_{av}) \leq -\epsilon \alpha_2 \|z_{av}\|^2
\]

(2.27)

\[
\|\frac{\partial v}{\partial z_{av}}\| \leq \alpha_4 \|z_{av}\|
\]

(2.28)

The derivative in (2.27) is to be taken along the trajectories of the averaged system (2.5). The function \( v \) is now used to study the stability of the perturbed system (2.16). Considering \( v(z) \), inequalities (2.26) and (2.28) are still verified, with \( z \) replacing \( z_{av} \). The derivative of \( v(z) \) along the trajectories of (2.16) is given by:

\[
\dot{v}(z) = \dot{v}(z) + (\frac{\partial v}{\partial z})(\epsilon P(t, z, \epsilon))
\]

(2.29)

and, using previous inequalities (including those from **Lemma 2.8**):

\[
\dot{v}(z) \leq -\epsilon \alpha_2 \|z\|^2 + \epsilon \alpha_4 \psi(\epsilon) \|x\|^2
\]

\[
\leq -\epsilon \left( \frac{\alpha_2 \psi(\epsilon) \alpha_4}{\alpha_2} \right) v(z)
\]

(2.30)

for all \( \epsilon \leq \epsilon_1 \). Let \( \epsilon' \) be such that \( \alpha_2 \psi(\epsilon') \alpha_4 > 0 \), and define \( \epsilon_2 = \min(\epsilon_1, \epsilon') \). Denote:

\[
\alpha(\epsilon) := \frac{\alpha_2 \psi(\epsilon) \alpha_4}{2 \alpha_2}
\]

(2.31)
Consequently, (2.30) implies that:

$$v(z) \leq v(z(t_0)) e^{-2\alpha(t)(t-t_0)}$$  \hspace{1cm} (2.32)

and:

$$\|z(t)\| \leq \sqrt{\frac{\alpha_2}{\alpha_1}} \|z(t_0)\| e^{-\alpha(t)(t-t_0)}$$  \hspace{1cm} (2.33)

Since $\alpha(c_i)>0$ for all $c_i$, system (2.15) is exponentially stable. Using (L.16), it follows that:

$$\|z(t)\| \leq \sqrt{\frac{\alpha_2}{\alpha_1}} \|z(t_0)\| e^{-\alpha(c_i)(t-t_0)}$$  \hspace{1cm} (2.34)

for all $t \geq t_0 \geq 0$, $c_i \leq c_f$, and $z(t_0)$ sufficiently small that all signals remain in $B_1$. In conclusion, the original system is exponentially stable, with rate of convergence (at least) $\alpha(c_i)$.

Comments

1) Theorem 2.5 is a local exponential stability result. The original system will be globally exponentially stable, if the averaged system is globally exponentially stable, and provided that all assumptions are valid globally.

2) The proof of Theorem 2.5 gives a useful bound on the rate of convergence of the original system. As $\varepsilon$ tends to zero, $\alpha(\varepsilon)$ tends to $\frac{\varepsilon}{2} \frac{\alpha_3}{\alpha_2}$, which is the bound on the rate of convergence of the averaged system that one would obtain using (2.26)-(2.27). In other words, the proof provides a bound on the rate of convergence, and this bound gets arbitrarily close to the corresponding bound for the averaged system, provided that $\varepsilon$ is sufficiently small. This is a useful conclusion because it is in general very difficult to obtain a guaranteed rate of convergence for the original, nonautonomous system. The proof assumes the existence of a Lyapunov function satisfying (2.26)-(2.28), but does not depend on the specific function chosen. Since the averaged system is autonomous, such a function is usually easier to find than for the original system, and any such function will provide a bound on the rate of convergence of the original system for $\varepsilon$ sufficiently small.

3) The conclusion of Theorem 2.5 is quite different from the conclusion of Theorem 2.4. Since both $z$ and $z_e$ go to zero exponentially with $t$, the error $z-z_e$ also goes to zero exponentially with $t$. Yet, Theorem 2.5 does not relate the bound on the error to $\varepsilon$. It is possible, however, to combine Theorem 2.4 and Theorem 2.5 to obtain a uniform approximation result, with an estimate similar to (2.24).
3. Averaging Theory Applied to Adaptive Identifiers

Brief Review of a Simple Adaptive Identifier

We consider an unknown plant, described by a single-input single-output, exponentially stable transfer function:

\[ \hat{p}(s) = \frac{\hat{p}_n(s)}{\hat{d}_n(s)} \]

where \( \hat{p}_n(s) \) is a monic polynomial of degree \( n \) (\( n \) is assumed to be known), and \( \hat{d}_n(s) \) is a polynomial of degree less than or equal to \( n \). The coefficients of the polynomials are unknown, and are to be obtained from the identifier.

The identifier considered here is an adaptive observer/identifier (see e.g. [21], [22]), and its structure is shown in Fig. 3.1. The filter blocks \( F_1 \) and \( F_2 \) generate smoothed derivatives of the input \( r \), and of the output \( y_p \) of the plant. Each of these blocks has a transfer function:

\[ \begin{pmatrix} \hat{A} & 0 \\ 1 \\ \vdots \\ \hat{A}^{n-1} \end{pmatrix} \]

where \( \hat{A} \in \mathbb{R}^{n \times n} \), \( \hat{b} \in \mathbb{R}^n \), and \( \det(sI-A) \) is a Hurwitz polynomial. The outputs of the filters are respectively \( v^{(1)} \), \( v^{(2)} \) \( \in \mathbb{R}^n \). The signal \( y_0 \) is obtained through the adaptive gains \( c(t), d(t) \in \mathbb{R}^n \), and \( c_{n+1}(t) \in \mathbb{R} \):

\[ y_0 = c^Tv^{(1)} + d^Tv^{(2)} + c_{n+1}r \]

and it may be verified that there exists a unique choice of the adaptive gains, denoted \( \hat{c}, \hat{d}, \) and \( \hat{c}_{n+1} \) such that the transfer function from the input \( r \) to the output \( y_0 \) is identical to the plant transfer function \( \hat{p}(s) \). We define the parameter vector \( \theta \in \mathbb{R}^{2n+1} \):

\[ \theta^T = (\hat{c}^T, \hat{d}^T, \hat{c}_{n+1}) \]

and the signal vector \( w \in \mathbb{R}^{2n+1} \):

\[ w^T = (v^{(1)}r, v^{(2)}r, r) \]

so that:

\[ y_0 = \theta^Tw \]

In the sequel, we will neglect the effect of the initial conditions of the plant, and of the filter blocks \( F_1, F_2 \). The results can be modified to take them into account, without any fundamental differences in the conclusions. We simply assume that the dynamics of the observers (determined by the eigenvalues of \( \hat{A} \)) are faster than those of the identifier. The output of the plant is then
given by an equation similar to that of the identifier:
\[ y_p = \theta^* T_w \]  
(3.7)
where \( \theta^* \) is the vector of "true" parameters corresponding to \( \rho(s) \). Defining the parameter error:
\[ \phi = \theta - \theta^* \]  
(3.8)
the output error \( e_1 = y_p - y_0 \) is given by:
\[ e_1 = \phi^T w \]  
(3.9)

It can be shown ([21], [22]) that, with the following adaptation law:
\[ \gamma = -\Gamma e_1 w \]  
(3.10)
where \( \Gamma \in \mathbb{R}^{n \times n} > 0 \), the following propositions are true:

(i) if \( r \) is bounded, then \( \lim_{t \to \infty} e_1(t) = 0 \)

(ii) if, moreover, \( w \) is persistently exciting, that is, if there exist constants \( \alpha_1, \alpha_2, \delta > 0 \), such that:
\[ \alpha_1 I \leq \int_t^{t+\delta} w w^T dt \leq \alpha_2 I \text{ for all } s > 0 \]  
(3.11)
then the parameter error also tends to zero, i.e.:
\[ \lim_{t \to \infty} \phi(t) = 0 \]  
(3.12)
and the convergence is exponential.

Application of the Averaging Theory

To apply the averaging theory developed in section 2, we will study the case when \( r = \epsilon I \), i.e. when the update law (3.10) is given by:
\[ \dot{\phi} = -\epsilon_1 w \]  
(3.13)
or, using eq. (3.9):
\[ \dot{\phi} = -\epsilon w w^T \phi \]  
(3.14)
Eq. (3.14) leads us to the following definition:

Definition 3.1 (Stationarity, Autocovariance): A signal \( u : \mathbb{R}^+ \to \mathbb{R}^n \) is said to be stationary if the following limit exists uniformly in \( s \):
\[ R_u(r) := \lim_{T \to 0} \frac{1}{T} \int_{s}^{s+T} u(t)u^T(t+r) dt \in \mathbb{R}^{n \times n} \]  
(3.15)
in which instance, the limit \( R_u(r) \) is called the autocovariance of \( u \).
It may be verified that the autocovariance matrix of a stationary signal \( w \) is a positive semidefinite function \( R_w(\tau) \), and that \( w \) is persistently exciting if and only if the autocovariance at 0 is positive definite \([14]\). Also, \( R_w(\tau) \) can be written as the inverse Fourier transform of a positive spectral measure \( S_w(d\nu) \):

\[
R_w(\tau) = \int_{-\infty}^{\infty} e^{i\pi \nu \tau} S_w(d\nu)
\]  

(3.16)

Further, if the input \( r \) is also stationary, \( S_w(d\nu) \) can be computed, using the fact that the transfer function from \( r \) to \( w \) is given by:

\[
q(\nu) := \left( (e^{j\nu} - \Lambda)^{-1} \right) \in \mathbb{R}^{2n+1}(\nu)
\]  

(3.17)

so that:

\[
S_w(d\nu) = q(j\nu)q^*(j\nu)\pi_r(d\nu)
\]  

(3.18)

Using eqns. (3.16) and (3.17), we can conclude that:

\[
R_w(0) = \int_{-\infty}^{\infty} q(j\nu)q^*(j\nu)\pi_r(d\nu) > 0
\]  

(3.19)

This in turn is assured \([14]\) if the support of \( \pi_r(d\nu) \) is greater than or equal to \( 2n+1 \) points (the dimension of \( w = \) the number of unknown parameters = \( 2n+1 \)).

With these definitions, the averaged system corresponding to (3.14) is simply:

\[
\dot{\varphi}_w = -\epsilon R_w(0)\varphi_w
\]  

(3.20)

This system is particularly easy to study, since it is linear, and when \( w \) is persistently exciting, \( R_w(0) \) is a positive definite matrix.

A natural Lyapunov function for (3.14) is:

\[
V(\varphi_w) = \frac{1}{2}\|\varphi_w\|^2
\]  

(3.21)

and:

\[
-\epsilon \lambda_{\min}(R_w(0))\|\varphi_w\|^2 \leq -\dot{V}(\varphi_w) \leq -\epsilon \lambda_{\max}(R_w(0))\|\varphi_w\|^2
\]  

(3.22)

where \( \lambda_{\min} \) and \( \lambda_{\max} \) are respectively the minimum and maximum eigenvalues of \( R_w(0) \). Thus, the rate of exponential convergence of the averaged system is at least \( \epsilon \lambda_{\min}(R_w(0)) \), and at most \( \epsilon \lambda_{\max}(R_w(0)) \). By the comments after theorem 2.5, we can conclude that the rate of convergence of the unaveraged system for \( \epsilon \) small enough is close to the interval \([\epsilon \lambda_{\min}(R_w(0)), \epsilon \lambda_{\max}(R_w(0))\]).
Eq. (3.19) gives an interpretation of \( R_v(0) \) in the frequency domain, and also a mean of computing an estimate of the rate of convergence of the adaptive algorithm, given the spectral content of the reference input. If the input \( r \) is periodic or almost periodic, the integral in (3.19) may be replaced by a summation. Since the transfer function \( g(s) \) depends on the unknown plant being identified, the use of the averaged equation to determine the rate of convergence is more conceptual than practical. If, however, bounds on parameter values are available, then some bounds on \( R_v(0) \) and on the corresponding rates of convergence can be deduced. These in turn can be used to determine the spectral content of the reference input that will optimize the rate of convergence of the identifier, given the physical constraints on \( r \). Such a procedure is very reminiscent of the procedure indicated in [23] (chapter 6) for the design of input signals in identification. The autocovariance matrix defined here is similar to the average information matrix defined in [23] (p 134). Our interpretation is, however, in terms of rates of parameter convergence of the averaged system rather than in terms of parameter covariance.

To illustrate the conclusions of this section, we consider the following example:

\[
\begin{align*}
\hat{\phi}(s) &= \frac{2s+2}{s+3} \\
\end{align*}
\]

The filter is chosen to be \( \text{det}(sI-A) = (s+5) \). The "true" values of the parameters \( c_1, d_1, d_2 \) are -1.6, 0.4, and 2. Denote the parameter error as

\[
\phi_1 = c_1 - c_1^*, \quad \phi_2 = d_1 - d_1^*, \quad \phi_3 = c_2 - c_2^*
\]

Since the number of unknown parameters is 3, parameter convergence will occur when the support of \( s_\alpha(d\dot{u}) \) is greater than or equal to 3 points. For the simulations, we considered an input of the form \( a_0 + 2\sin(\omega t) \). By virtue of (3.18) and (3.19), (3.20) now becomes

\[
\begin{pmatrix}
\dot{\phi}_{r_1} \\
\dot{\phi}_{r_2} \\
\dot{\phi}_{r_3}
\end{pmatrix} = -\epsilon
\begin{pmatrix}
\frac{2}{3}a_1^2 + \frac{25a_1^2}{2(25+\omega^2)} & \frac{2}{9}a_1^2 + \frac{25(3+\omega^2)a_1^2}{(9+\omega^2)(25+\omega^2)} & \frac{2}{3}a_1 + \frac{25a_1^3}{2(25+\omega^2)} \\
\frac{25(3+\omega^2)a_1^2}{(9+\omega^2)(25+\omega^2)} & \frac{4}{9}a_1^2 + \frac{50(1+\omega^2)a_1^2}{(9+\omega^2)(25+\omega^2)} & \frac{2}{9}a_1^2 + \frac{5(15+7\omega^2)a_1^2}{(9+\omega^2)(25+\omega^2)} \\
\frac{25a_1^2}{2(25+\omega^2)} & \frac{2}{9}a_1^2 + \frac{5(15+7\omega^2)a_1^2}{(9+\omega^2)(25+\omega^2)} & \frac{2}{9}a_1^2 + \frac{a_1}{2}
\end{pmatrix}
\]

With \( a_0 = 2, a_1 = 2 \) and \( \omega = 4 \), the eigenvalues of the averaged system (3.24) are computed to be -0.28\( \epsilon \), -0.64\( \epsilon \) and -15.39\( \epsilon \). Figs 3.2 and 3.3 show the plots of the parameter errors of \( c_1 \) and \( d_1 \) for both the original and averaged systems with three different adaptation gains \( \epsilon = 0.1, 0.5, 1 \). Fig 3.4 is a plot of the Lyapunov function of (3.21) for both systems using a log scale. It illustrates
the closeness of the rate of convergence of the two systems.
4. Averaging of Two-Time Scale Systems

Systems of the form (2.1) studied in section 2 are to be thought of as one time scale systems in that the entire state variable \( x \) is varying slowly in comparison with the rate of time variation of the right hand side of the differential equation. In this section, we will study averaging for the case when only some of the state variables are slowly varying.

Consider, for example, the system:

\[
\dot{x} = \varepsilon f(t,x,y) \quad x(0) = x_0 \tag{4.1}
\]

\[
\dot{y} = Ay + \varepsilon g(t,x,y) \quad y(0) = y_0 \tag{4.2}
\]

where \( x \in \mathbb{R}^n \) is called the slow state, \( y \in \mathbb{R}^m \) is called the fast state, and \( f, g \) are piecewise continuous functions of time.

The goal of averaging will be to approximate the evolution of the slow state. The system (4.1), (4.2) is not the most general two-time scale system. In fact, it is easily seen to be decoupled and linear at \( \varepsilon = 0 \). The study of this special form is motivated by several applications. We will also study another special form later in this section. It is easy to see, from the proofs of this section and those of section 2, that \( f \) and \( g \) may be allowed to depend smoothly on \( \varepsilon \) as in (A2).

The averaged system for the slow state is:

\[
\dot{x}_{av} = \varepsilon f_{av}(x_{av}) \quad x_{av}(0) = x_0 \tag{4.3}
\]

where \( f_{av} \) is defined by the limit:

\[
f_{av}(x) = \lim_{T \to \infty} \frac{1}{T} \int_0^T f(t,x,0)dt \tag{4.4}
\]

assuming that the limit exists uniformly in \( x \).

The following assumptions will be in effect for (4.1), (4.2):

(B1) \( x=0, y=0 \) is an equilibrium point of system (4.1), (4.2), i.e. \( f(t,0,0)=0 \) and \( g(t,0,0)=0 \) for all \( t \geq 0 \). Both \( f \) and \( g \) are Lipschitz in \( x \) and \( y \), i.e.:

\[
||f(t,x_1,y_1)-f(t,x_2,y_2)|| \leq l_1 ||x_1-x_2|| + l_2 ||y_1-y_2|| \tag{4.5}
\]

\[
||g(t,x_1,y_1)-g(t,x_2,y_2)|| \leq l_3 ||x_1-x_2|| + l_4 ||y_1-y_2|| \tag{4.6}
\]

for all \( t \geq 0, x_1,x_2 \in B_1, y_1,y_2 \in B_1 \).

(B2) \( f_{av}(0) = 0 \), and \( f_{av} \) is Lipschitz in \( x \), i.e.:

\[
||f_{av}(x_1)-f_{av}(x_2)|| \leq l_{av} ||x_1-x_2|| \tag{4.7}
\]

for all \( x_1,x_2 \in B_1 \).
(B3) the function \( d(t,x) = f(t,x,0) - f_{sv}(x) \) satisfies the conditions of Lemma 2.2

(B4) \( A \in \mathbb{R}^{n \times m} \) is Hurwitz.

(B5) \( x_0 \) is sufficiently small that, for \( T \) fixed, and some \( h' < h, x_{sv}(t) \in B_{h'} \) for all \( t \in [0, T/h'] \) (this is possible, from (B2)). We will also assume that \( y_0 \in B_{h'} \), the corresponding closed ball in \( \mathbb{R}^m \).

**Theorem 4.1 Basic Averaging Theorem for Two-Time Scale System**

If: The original system (4.1), (4.2), and the averaged system (4.3), satisfy assumptions (B1)-(B5),

Then: There exists \( \psi(\epsilon) \in K \) such that, given \( T \geq 0 \):

\[
\|z(t) - z_{av}(t)\| \leq \psi(\epsilon) b_T
\]

for some \( b_T, \epsilon_T > 0 \), and for all \( t \in [0, T/h'] \), \( \epsilon \leq \epsilon_T \), and \( y_0 \) sufficiently small. Further, \( \psi(\epsilon) \) is of the order of \( \epsilon + \xi(\epsilon) \) (as defined in Lemma 2.2).

**Proof:** We first apply Lemma 2.2, and obtain a result similar to Lemma 2.3. Consider the transformation of variable:

\[
z = z + \epsilon w(t, x)
\]

with \( \epsilon \leq \epsilon_1 \). This transformation leads to:

\[
\dot{z} = (I + \epsilon \frac{\partial w_1}{\partial z})^{-1} \epsilon \left\{ f_{sv}(z) + (f(t,x,0) - f_{sv}(z) - \frac{\partial w_1}{\partial t}) + (f(t,z + \epsilon w_0,0) - f(t,z,0)) + (f(t,z + \epsilon w_0,0) - f(t,z,0)) \right\}
\]

or:

\[
\dot{z} = \epsilon f_{sv}(z) + \epsilon p_1(t,x,\epsilon) + \epsilon p_2(t,x,y,\epsilon)
\]

\( z(0) = x_0 \)

where:

\[
\|p_1(t,z,\epsilon)\| \leq \frac{1}{1 - \xi(\epsilon)} (\xi(\epsilon) \epsilon_0 + \xi(\epsilon) + \xi(\epsilon)t_1) \|z\| := \xi(\epsilon) k_1 \|z\|
\]

and:

\[
\|p_2(t,z,y,\epsilon)\| \leq \frac{1}{1 - \xi(\epsilon)} t_2 \|v\| := \xi(\epsilon) k_2 \|v\|
\]
We now estimate the error \( z - x_{av} \), following a proof similar to the proof of Theorem 2.4.

First, we have that:

\[
\| z - z \| \leq \epsilon(\| z \|)
\]  

(4.14)

Then, the error \( z - x_{av} \) can be estimated from:

\[
\frac{d}{dt}(z - x_{av}) = \epsilon(f_{av}(z) - f_{av}(x_{av})) + \epsilon \phi_1(t, z, \epsilon) + \epsilon \phi_2(t, z, \epsilon, \epsilon) \\
(0 - x_{av}(0)) = 0
\]  

(4.15)

for all \( t \in [0, T/\epsilon] \), \( x_{av}(t) \in B_{h'} \), \( h' < h \). As in the proof of Theorem 2.4, we will show that, on this interval, and for as long as \( z, x \in B_h \), the errors \( z - x_{av} \) and \( x - x_{av} \) can be made arbitrarily small by reducing \( \epsilon \).

Integrating (4.15):

\[
\| z(t) - x_{av}(t) \| \leq \epsilon \int_0^t \| z(\tau) - x_{av}(\tau) \| d\tau + \epsilon \xi(\epsilon) k_1 \int_0^t \| z(\tau) \| d\tau \\
+ \epsilon k_2 \int_0^t \| y(\tau) \| d\tau
\]  

(4.16)

Further, \( y(t) \) can be calculated from (4.2):

\[
y(t) = e^{At} y_0 + \epsilon \int_0^t e^{A(t-\tau)} g(\tau, z, \epsilon) d\tau
\]  

(4.17)

Since \( A \) is Hurwitz, we have that:

\[
\| e^{At} \| \leq m \ e^{-\lambda t}
\]  

(4.18)

for some \( m, \lambda > 0 \), and:

\[
\| y(t) \| \leq m \| y_0 \| e^{-\lambda t} + \epsilon m \int_0^t e^{-\lambda (t-\tau)} (l_3 \| z(\tau) \| + l_4 \| y(\tau) \|) d\tau
\]  

(4.19)

or:

\[
\| e^{At} y(t) \| \leq m \| y_0 \| + \epsilon m l_3 \int_0^t e^{-\lambda \tau} \| z(\tau) \| d\tau + \epsilon m l_4 \int_0^t \| e^{At} y(\tau) \| d\tau
\]  

(4.20)

Applying the Generalized Bellman-Gronwall Lemma:

\[
\| e^{At} y(t) \| \leq m \| y_0 \| e^{\lambda t} + \int_0^t \| e^{At} y(\tau) \| e^{\lambda (t-\tau)} d\tau
\]

Define \( \lambda(\epsilon) = \lambda - \epsilon m l_4 \), and \( \epsilon_6 (0 < \epsilon_6 \leq \epsilon_4) \) so that \( \lambda(\epsilon) > 0 \) for \( \epsilon \leq \epsilon_6 \). It follows that:

\[
\| y(t) \| \leq m h e^{-\lambda(\epsilon)t} + \epsilon m l_3 h / \lambda(\epsilon)
\]  

(4.21)

Using this estimate in (4.16), and using the Generalized Bellman-Gronwall Lemma again:
As in Theorem 2.4, it follows that, for some $b_T$:

$$||x(t)-z_{av}(t)|| \leq \int_0^t \left( \xi(\epsilon) k_1 h + m k_2 h e^{-\lambda(\epsilon) t} + \frac{\epsilon m k_2 h}{\lambda(\epsilon)} \right) \epsilon e^{\lambda(\epsilon) t} e^{-\epsilon \lambda(\epsilon) t} e^{l_{av} (t-\epsilon)} d\tau$$

$$\leq (\epsilon + \xi(\epsilon)) \left( k_1 h + \frac{m k_2 h l_{av}}{\lambda(\epsilon) + \epsilon l_{av}} + \frac{m k_2 h}{\lambda(\epsilon)} \right) e^{l_{av} t}$$

$$:= \psi(\epsilon) b_T$$  \hspace{1cm} (4.23)

By assumption, $||z_{av}(t)|| \leq h'< h$. Let $\epsilon_T (0<\epsilon_T \leq \epsilon_0)$ such that $\psi(\epsilon_T) b_T < h - h'$. Further, let $y_0$, and $\epsilon_T$ sufficiently small that, by (4.22), $y(t) \in B_1$, for all $t \in [0,t/c]$. It follows, from a simple contradiction argument, that the estimate in (4.24) is valid for all $t \in [0,T/c]$, whenever $\epsilon \leq \epsilon_T$.

Theorem 4.2 Exponential Stability Theorem for Two-Time Scale Systems

If: The original system (4.1), (4.2), and the averaged system (4.3) satisfy assumptions (B1)-(B4), the function $f_{av}(x)$ has continuous and bounded first partial derivatives in $x$, and $x=0$ is an exponentially stable equilibrium point of the averaged system,

Then: There exists $\epsilon_4>0$ such that the equilibrium point $x=0$ of the original system is exponentially stable for all $\epsilon \leq \epsilon_4$.

Proof: Since $z_{av}=0$ is an exponentially stable equilibrium point of the averaged system, there exists a function $v(x_{av})$ satisfying (2.26)-(2.28). On the other hand, since $A$ is Hurwitz, there exist matrices $P, Q > 0$, such that $A^T P + PA = -Q$. Denote by $\rho_1, \rho_2, \sigma_1, \sigma_2$ the minimum and maximum eigenvalues of the $P$ and $Q$ matrices. We now study the stability of the system (4.11), (4.2), and consider the following Lyapunov function:

$$v_1(x,y) = v(x) + \frac{\alpha_2}{\rho_2} y^T P y$$  \hspace{1cm} (4.25)

so that:

$$\alpha'_1 (||x||^2 + ||y||^2) \leq v_1(x,y) \leq \alpha_2 (||x||^2 + ||y||^2)$$  \hspace{1cm} (4.26)

where $\alpha'_1 = \min(\alpha_1, \frac{\alpha_2}{\rho_2} \rho_1)$. The derivative of $v_1$ along the trajectories of (4.11), (4.2) can be estimated, using the previous results:

$$\dot{v}_1(x,y) \leq -\epsilon \alpha_0 ||x||^2 + \epsilon k_1 \xi(\epsilon) \alpha_0 ||x||^2$$

$$+ \epsilon k_2 \alpha_4 ||x|| ||y|| - \frac{\alpha_2}{\rho_2} \sigma_1 ||y||^2$$
for $\epsilon \leq \epsilon_1$ (so that, in particular, $||z|| \leq 2||z||$). Note that since $ab \leq (a^2 + b^2)/2$ for all $a, b \in \mathbb{R}$, we have:

$$
\epsilon ||z|| ||y|| \leq \frac{1}{2} \left( \epsilon^{\sqrt{2}} ||z||^2 + \epsilon^{2/3} ||y||^2 \right)
$$

so that:

$$
\dot{v}_1(z, y) \leq -\epsilon \left( \alpha_3 - \xi(\epsilon) k_1 \alpha_4 - \epsilon^{1/3} \frac{k_2 \alpha_4}{2} - 2\epsilon^{1/3} \tau_1 \alpha_2 \right) ||z||^2
$$

$$
- \left( \frac{\alpha_2}{p_2} q_1 - 2\epsilon_4 \alpha_2 - \epsilon^{2/3} \frac{k_2 \alpha_4}{2} - 2\epsilon^{2/3} \tau_1 \alpha_2 \right) ||y||^2
$$

$$
= -2\epsilon \alpha_2 \alpha(\epsilon) ||z||^2 - q(\epsilon) ||y||^2
$$

(4.29)

Note that, with this definition, $\alpha(\epsilon) \to \frac{1}{2} \frac{\alpha_3}{\alpha_2}$ as $\epsilon \to 0$.

Let $\epsilon_4 (0 < \epsilon_4 \leq \epsilon_1)$ be sufficiently small that $\alpha(\epsilon) > 0$, $q(\epsilon) > 0$, and $2\epsilon \alpha_2 \alpha(\epsilon) \leq q(\epsilon)$ whenever $\epsilon \leq \epsilon_4$. Consequently:

$$
\dot{v}_1(z, y) \leq -2\epsilon \alpha(\epsilon) v_1(z, y)
$$

(4.30)

and:

$$
v_1(z, y) \leq v_1(z(t_0), y(t_0)) e^{-2\epsilon \alpha(\epsilon)(t-t_0)}
$$

(4.31)

As in Theorem 2.5, this implies the exponential convergence of the original system, with rate of convergence $\epsilon \alpha(\epsilon)$. Also, for $z(t_0)$, $y(t_0)$ sufficiently small, all signals are guaranteed to remain in $B_k$, so that all assumptions are applicable.

Comments

The comments of Theorem 2.5 apply similarly to Theorem 4.2. In particular, the proof gives a useful bound on the rate of convergence of the original system, and this bound again tends to the bound on the rate of convergence of the averaged system.

Mixed Time Scales

We now discuss a more general class of two-time scale systems, arising in adaptive control:

$$
\dot{z} = \epsilon f'(t, z, y')
$$

(4.32)

$$
\dot{y}' = Ay' + h(t, z) + \epsilon g'(t, z, y')
$$

(4.33)
We will show that system (4.32)-(4.33) can be transformed into the system described in the previous section. In this case, $x$ is a slow variable, but $y'$ has both a fast, and a slow component.

The averaged system corresponding to (4.32), (4.33) is obtained as follows. Define the function:

$$v(t,x) = \int_0^t e^{A(t-\tau)}h(\tau,x)\,d\tau$$

and assume that the following limit exists uniformly in $t$ and $x$:

$$I'(x) = \lim_{T \to \infty} \frac{1}{T} \int_0^T f'(t,x,v(t,x))\,dt$$

Intuitively, $v(t,x)$ represents the steady-state value of the variable $y$ with $x$ frozen and $\epsilon=0$ in (4.33).

To show that the averaged system of (4.35) is the right one, we transform the system (4.32), (4.33) to the form (4.1), (4.2), using the transformation:

$$y = y' - v(t,x)$$

From (4.34), $v(t,x)$ satisfies:

$$\frac{\partial}{\partial t} v(t,x) = Av(t,x) + h(t,x) \quad v(t,0) = 0$$

Differentiating (4.36), we have that:

$$\dot{y} = Ay + \epsilon \left[ \frac{\partial v(t,x)}{\partial x} f'(t,x,y+v(t,x)) + g'(t,x,y+v(t,x)) \right]$$

so that system (4.32), (4.33) is of the form (4.1), (4.2), with:

$$f(t,x,y) = f'(t,x,y+v(t,x))$$

$$g(t,x,y) = -\frac{\partial v(t,x)}{\partial x} f'(t,x,y+v(t,x)) + g'(t,x,y+v(t,x))$$

The averaged system is obtained by averaging the right-hand side of (4.39) with $\epsilon=0$, so that the definitions (4.4), and (4.35) agree.

To apply Theorem 4.1, and Theorem 4.2, we require that assumptions (B1)-(B5) be satisfied. In particular, we assume similar Lipschitz conditions on $f'$, $g'$, and the following assumption on $h(t,x)$:

(B6) $h(t,0) = 0$ for all $t \geq 0$, and:

*This choice of transformation was pointed out to us by B. Riedle & P. Kokotovic.
\[ \left\| \frac{\partial h(t,x)}{\partial x} \right\| \leq k \]  
\hfill (4.41)

for all \( t \geq 0, x \in B_k \).

This new assumption implies that \( v(t,0) = 0 \), and:

\[ \left\| \frac{\partial v(t,x)}{\partial x} \right\| \leq k' \]  
\hfill (4.42)

for all \( t \geq 0, x \in B_k \).

This condition is sufficient to guarantee Lipschitz conditions for the system (4.1), (4.2), given Lipschitz conditions for the system (4.32), (4.33). The theory developed earlier in this section can therefore be directly applied to systems of the form (4.32), (4.33). The key to the preceding transformation is the fact that the new state variable \( y \) is truly a fast variable, so that the two time scales have been separated.
5. Two-Time Scale Averaging Applied to Model Reference Adaptive Controller

To apply the theory of Section 4 to model reference adaptive controllers (our results here are a small extension of those of Riedle, Kokotovic [11]), we review the model reference adaptive system of Narendra, Valavani [18] for the relative degree 1 case (our notation is however consistent with Sastry [19]). Consider a plant with transfer function

\[ n(s) = \frac{n_0}{s^{\delta_0}} \]

where \( n_0, \delta_0 \) are relatively prime monic polynomials of degree \( n-1, n \) respectively and \( k_p \) is a scalar (the representation (5.1) is assumed minimal). The following are assumed to be known about the plant transfer function:

(C1) The degrees of the polynomials \( \delta_p, n_p \) are known.
(C2) The sign of \( k_p \) is known (say \( k_p > 0 \)).
(C3) The plant transfer function is assumed to be minimum phase.

The objective is to build a compensator so that the plant output asymptotically matches that of a stable reference model \( \hat{m}(s) \) with input \( r(t) \), output \( y_m(t) \) and transfer function

\[ \hat{m}(s) = k_m \frac{\hat{n}_m(s)}{\hat{d}_m(s)} \]

where \( k_m > 0 \) and \( \hat{n}_m, \hat{d}_m \) are monic polynomials of degree \( n-1, n \) respectively (not necessarily relatively prime but both Hurwitz). If we denote the input and output of the plant \( u(t) \) and \( y_p(t) \) respectively, the objective may be stated as: find \( u(t) \) so that \( y_p(t) - y_m(t) \rightarrow 0 \) as \( t \rightarrow \infty \). By using suitable prefiltering of the reference signal if necessary, we may assume that the model \( \hat{m}(s) \) is strictly positive real.

The scheme is shown in Figure 5.1. The dynamical compensator blocks \( F_1 \) and \( F_2 \) (reminiscent of those in Section 3) are identical one input, \( n-1 \) output systems, each with transfer function \( (sI - A)^{-1}b \); \( A \in \mathbb{R}^{n-1 \times n-1} , b \in \mathbb{R}^{n-1} \) where \( A \) is chosen so that its eigenvalues are the zeros of \( \hat{n}_m \). The pair \( A, b \) is assumed controllable and, for ease of book-keeping (in the algorithm proof alone), we assume that they are in controllable form so that

\[ (sI - A)^{-1} = \frac{1}{\hat{n}_m(s)} \begin{bmatrix} 1 & 0 \\ \vdots & \ddots \\ 0 & \cdots & 1 \end{bmatrix} \]

The parameters \( c \in \mathbb{R}^{n-1} \) in the precompensator block serve to tune the closed loop plant zeros; \( d \in \mathbb{R}^{n-1}, d_0 \in \mathbb{R} \) in the feedback compensator assign the closed loop plant poles. The parameter
$c_0$ adjusts the overall gain of the closed loop plant. Thus, the vector of $2n$ adjustable parameters denoted $\theta$ is

$$\theta^T = [c_0, c^T, d_0, d^T]$$

with the signal vector $w \in \mathbb{R}^{2n}$ defined by

$$w^T = [r, \nu^{(1)}_r, \nu_r, \nu^{(2)}_r]$$

The input to the plant is seen to be

$$u = \theta^T w$$

and the state equations of the plant loop are given by

$$\begin{bmatrix}
  \dot{z}_p \\
  \dot{\nu}^{(1)}_r \\
  \dot{\nu}^{(2)}_r
\end{bmatrix} =
\begin{bmatrix}
  A_p & 0 & 0 \\
  0 & A & 0 \\
  b c^T & 0 & A
\end{bmatrix}
\begin{bmatrix}
  z_p \\
  \nu^{(1)}_r \\
  \nu^{(2)}_r
\end{bmatrix} +
\begin{bmatrix}
  b_p \\
  b \\
  0
\end{bmatrix} \theta^T w$$

(5.2)

It may be verified that there is a unique constant $\theta^* \in \mathbb{R}^{2n}$ such that, when $\theta = \theta^*$, the transfer function of the plant plus controller equals $\hat{m}(s)$. It can also be shown [18] that when $r$ is bounded and the parameter update law is given by

$$\dot{\theta} = -\Gamma \epsilon_1 w = -\Gamma (\nu_r - y_m) w$$

(5.3)

with $\Gamma \in \mathbb{R}^{2n \times 2n}$, a positive definite matrix, all signals in the loop, i.e. $u, \nu, \nu^{(1)}, \nu^{(2)}, \nu_r, y_m$ are bounded. In addition, $\lim_{t \to \infty} \epsilon_1(t) = 0$ so that asymptotically $\nu_r(t)$ approaches $y_m(t)$. The proof of this fact uses the following procedure: represent the model (in non-minimal form) as the plant loop with $\theta$ set equal to $\theta^*$. The state equations for the model loop are given by

$$\begin{bmatrix}
  \dot{x}_m \\
  \dot{\nu}^{(1)}_m \\
  \dot{\nu}^{(2)}_m
\end{bmatrix} =
\begin{bmatrix}
  A_p + b_p d_0 c^T & b_p c^T & b_p d^T \\
  b d_0 c^T & A + b c^T & b d^T \\
  b c^T & 0 & A
\end{bmatrix}
\begin{bmatrix}
  x_m \\
  \nu^{(1)}_m \\
  \nu^{(2)}_m
\end{bmatrix} +
\begin{bmatrix}
  b_p \\
  b \\
  0
\end{bmatrix} c_0 r$$

(5.4)

The $3n-2 \times 3n-2$ matrix in (5.4) is henceforth referred to as $\tilde{A}$, and the $3n-2$ vector in (5.4) as $\tilde{\nu}$. Then, subtracting (5.4) from (5.2) with

$$\epsilon^T = [x^T_p, \nu^{(1)}_r, \nu^{(2)}_r] - [x^T_m, \nu^{(1)}_m, \nu^{(2)}_m]$$

we have that

$$\dot{\epsilon} = \tilde{A} \epsilon + \tilde{\nu} \phi^T w$$

(5.5)

and

$$\epsilon_1 := y_r - y_m = [c^T_p, 0, 0] \epsilon = : \tilde{c}^T \epsilon$$

(5.6)

where $\phi$, the parameter error $:= \theta - \theta^*$. Note from (5.4) that $\tilde{c}^T (sI - \tilde{A})^{-1} \tilde{c} \phi^*$ is equal to the model
transfer function and that \( c_0^t = k_p/k_m \) is the ratio of the high frequency gain (positive by assumption). Now the update law (5.3) is
\[
\dot{\phi} = -\Gamma w(t)\bar{e}^T e
\]
(5.7)
To apply averaging, we consider slow adaptation, i.e. \( \Gamma = \epsilon I \) resulting in
\[
\dot{e} = \bar{A} e + \bar{b} w^T \phi
\]
(5.5)
and
\[
\dot{\phi} = -\epsilon w \bar{e}^T e
\]
(5.8)
Equations (5.5) and (5.8) are superficially of the form (4.32), (4.33) with \( h(t, \phi) := bw^T(t)\phi \) and \( f'(t, \phi, e) = -w(t)e. \) Difficulty, however, arises from the fact that \( w(t) \) in (5.5), (5.8) is not independently and exogenously specified, but in fact depends on \( e. \) To show this dependence explicitly, we set
\[
w := \sum_{\alpha=0}^{\infty} w^{(\alpha)}
\]
an exogenously defined 3n-2 dimensional vector that can be obtained either from \( r(t) \) alone or as linear combinations of the state variables of (5.4), and rewrite
\[
w = w_m + Q e
\]
(5.9)
where
\[
Q = \begin{bmatrix}
0 & 0 & 0 \\
0 & I & 0 \\
0 & 0 & I
\end{bmatrix}
\]
Using (5.9), the equations (5.3) and (5.8) are
\[
\dot{e} = \bar{A} e + \bar{b} w_m^T \phi + \bar{b} \epsilon e^T Q^T \phi
\]
(5.10)
\[
\dot{\phi} = -\epsilon w_m \bar{e}^T e - \epsilon Q e \bar{e}^T e
\]
(5.11)
With the exception of the last terms (quadratic in \( e \) and \( \phi \)), equations (5.10), (5.11) are linear time varying equations describing the linearized adaptive control system, around the equilibrium \( e = 0, \phi = 0. \) In this section, we apply averaging to the linearized equations (5.10), (5.11) corresponding to small \( e \) and \( \phi. \) Averaging of the full non-linear equations (5.10), (5.11) is more subtle and is not considered here. We consider
\[
\dot{e} = \bar{A} e + \bar{b} w_m^T \phi
\]
(5.12)
\[
\dot{\phi} = -\epsilon w_m \bar{e}^T e
\]
(5.13)
Since \( r \) is bounded and \( \bar{A} \) is stable (its eigenvalues are the union of the zeros of \( \bar{a}_m, \bar{n}_p \) and the
eigenvalues of $\Lambda$), $w_m$ is bounded. Hence it is easy to see that the equations (5.12), (5.13) are of the form of (4.33), (4.32) with the functions $f'$ and $h$ satisfying the conditions of Section 4.

To establish the averaging results, we assume that $\nu$ is stationary. This implies, as has been shown in Boyd and Sastry [14], that $w_m$ is stationary. Its spectral measure is related to that of $\nu$ by

$$S_{w_m}(d\nu) = \hat{n}(j\nu) \hat{n}^*(j\nu) s_c(d\nu)$$  \hspace{1cm} (5.14)

with

$$\hat{n}(s) := \begin{pmatrix} \frac{1}{\hat{m}^{-1}(sI - \Lambda)^{-1}b} \\ \hat{m} \\ \hat{m}(sI - \Lambda)^{2}b \end{pmatrix}$$

an exponentially stable transfer function.

The function $v(t,\phi)$ of Section 4 for the systems (5.12), (5.13) is

$$v(t,\phi) := \int_0^t e^{\tilde{A}(t-\tau)} \tilde{b}w_m^T(\tau) d\tau \phi$$

and the averaged $f$ is given by

$$f_{\nu}(\phi) = -\lim_{T\to\infty} \frac{1}{T} \int_0^T w_m(t) \tilde{e}^T \left[ \int_0^t e^{\tilde{A}(t-\tau)} \tilde{b}w_m^T(\tau) d\tau \right] dt \phi$$  \hspace{1cm} (5.15)

Since $w_m$ is stationary, the limit in (5.15) may be shown to exist as follows. Define a filtered version of $w_m$ to be

$$w_{m_f}(t) = \int_0^t \tilde{e}^T e^{\tilde{A}(t-\tau)} \tilde{b}w_m(\tau) d\tau$$  \hspace{1cm} (5.16)

Since $\tilde{e}^T(sI - \tilde{\Lambda})^{-1}\tilde{b} = \frac{1}{c_0} \hat{m}(s)$ is stable so that it follows that $w_{m_f}(t)$ is also stationary. The quantity inside the square brackets in (5.15) is

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T w_m(t) w_{m_f}(t) dt = R_{v_m,v_{m_f}}(0)$$  \hspace{1cm} (5.17)

i.e. the cross correlation between $w_m$ and $w_{m_f}$ evaluated at 0. Consequently, we may use (5.14) and (5.16) to obtain a formula for $R_{v_m,v_{m_f}}(0)$ as

$$R_{v_m,v_{m_f}}(0) = \frac{1}{c_0} \int_0^\infty \hat{n}(j\nu) \hat{n}^*(j\nu) \hat{m}(j\nu) s_c(d\nu)$$  \hspace{1cm} (5.18)

and the averaged system is a linear time-invariant system.
Since \( \dot{m}(0) \) is strictly positive real, the matrix \( R_{\dot{m} m}(0) \) is a positive definite matrix. Unlike the matrix \( R_v(0) \) of Section 3, \( R_{\dot{m} m}(0) \) need not be symmetric, and its eigenvalues need not be real. However, the real parts are guaranteed to be positive. When the reference input \( r \) is almost periodic, i.e.

\[
r(t) \sim \sum_k r_k e^{j\nu_k t}
\]
a simple formula for \( R_{\dot{m} m}(0) \) is given by

\[
\frac{1}{c_0^2} \sum_k \tilde{m}(j\nu_k) \tilde{m}^*(j\nu_k) m(j\nu_k) r_k
\]

(5.20)

As an illustration of the preceding results, we consider the following example of a first order plant with an unknown pole and an unknown gain:

\[
\dot{\phi}(s) = \frac{k_p}{s + a_p}
\]

The adaptive process is to adjust the feedforward gain \( c_0 \) and the feedback gain \( d_0 \) so as to make the closed loop transfer function match the model transfer function

\[
\tilde{m}(s) = \frac{k_m}{s + a_m}
\]

To guarantee persistency of excitation, we use a sinusoidal input signal of the form

\[
r(t) = a \sin(\omega t)
\]

Thus, equations (5.5), (5.7) become

\[
\dot{e} = -a_m e + k_p (\phi_1 r + \phi_2 y_m)
\]

\[
\dot{\phi}_1 = -e r
\]

\[
\dot{\phi}_2 = -e y_m
\]

where

\[
\phi_1 = c_0 - c_0^* \quad , \quad \phi_2 = d_0 - d_0^*
\]

Consequently, the averaged system defined in (5.19) now is

\[
\begin{bmatrix}
\dot{\phi}_{av1} \\
\dot{\phi}_{av2}
\end{bmatrix}
= -e^2 \begin{bmatrix}
k_p \\
k_m
\end{bmatrix}
\begin{bmatrix}
\frac{18}{(9 + \omega^2)} & \frac{18(9 - \omega^2)}{(9 + \omega^2)^2} \\
\frac{18}{(9 + \omega^2)} & \frac{162}{(9 + \omega^2)^2}
\end{bmatrix}
\begin{bmatrix}
\phi_{av1} \\
\phi_{av2}
\end{bmatrix}
\]
using equation (5.18). With \( a_\pi = 3, k_\pi = 3, a_p = 1, k_p = 2, a = 3, w = 2 \), the two eigenvalues of the averaged system are computed to be \(-3.10c\) and \(-0.43c\), both real negative. Fig 5.2, 5.3 show the plots of the parameter errors of \( c_0 \) and \( d_0 \) for the original and averaged system, with three different adaptation gains. Fig. 4 corresponds to a higher frequency input signal \( \omega = 4 \) such that the eigenvalues of the matrix \( R_{e_m e_m}(0) \) are complex \((-0.49 \pm 0.30i)c\), and explains the oscillatory behavior of the original and averaged systems.

Using the results of Boyd-Sastry [14], it is easy to verify the following facts

(i) \( R_{e_m e_m}(0) \) is singular unless \( R_w(0) > 0 \), i.e. \( w(t) \) is persistently exciting. Thus persistent excitation of \( w \) is a necessary condition for stability of (5.19).

(ii) If \( m(s) \) is strictly positive real and \( w(t) \) is persistently exciting, then \( R_{e_m e_m}(0) \) is Hurwitz. Hence \( m(s) \) being strictly positive real is a sufficient condition for stability of (5.19), given that \( w(t) \) is persistently exciting.

It is intuitive that if \( w \) is persistently exciting and \( m(s) \) is close in some sense to being strictly positive real that \( R_{e_m e_m}(0) \) will be Hurwitz (in particular, this is the case if \( \text{Re} m(j\nu) \) fails to be positive at frequencies where \( m(j\nu) \) is small enough. ). More specific results in this context are in [11,17].
6. Concluding Remarks

We have presented in this paper new stability theorems for averaging analysis of one and two time scale systems. We have applied these techniques to obtain bounds on the rates of convergence of adaptive identifiers and controllers of relative degree 1.

We feel that the techniques presented here can be extended to obtain instability theorems for averaging. Such theorems could be used to study the mechanism of slow drift instability in adaptive schemes in the presence of unmodelled dynamics, in a framework resembling that of [10]. Also, our analysis of the use of averaging in the study of adaptive control required that the scheme be linearized in a certain sense made precise in Section 5. In order to relax this requirement, averaging techniques for a wider class of two-time scale systems than that discussed in Section 4 will be needed.
References


APPENDIX

Proof of Lemma 2.1

Define:

\[ w_t(t,x) = \int_0^t d(r,x) e^{-\gamma(t-r)} d\tau \]  

(L.1)

and:

\[ w_d(t,x) = \int_0^t d(r,x) d\tau \]  

(L.2)

From the assumptions:

\[ ||w_d(t+t_0,x)-w_d(t_0,x)|| \leq \gamma(t).t \]  

(L.3)

for all \( t,t_0 \geq 0 \), \( x \in B_1 \). Integrating (L.1) by parts:

\[ w_t(t,x) = w_d(t,x) - \epsilon \int_0^t e^{-\gamma(t-r)} w_d(r,x) d\tau \]  

(L.4)

Using the fact that:

\[ \epsilon \int_0^t e^{-\gamma(t-r)} w_d(r,x) d\tau = w_d(t,x) - w_d(t,x) e^{-\gamma t} \]  

(L.5)

(L.4) can be rewritten as:

\[ w_t(t,x) = w_d(t,x) e^{-\gamma t} + \epsilon \int_0^t e^{-\gamma(t-r)} (w_d(t,x) - w_d(r,x)) d\tau \]  

(L.6)

and, using (L.3):

\[ ||w_t(t,x)|| \leq \gamma(t) t e^{-\gamma t} + \epsilon \int_0^t e^{-\gamma(t-r)} (t-r) \gamma(t-r) d\tau \]  

(L.7)

Consequently,

\[ ||\epsilon w_t(t,x)|| \leq \sup_{r \geq 0} \gamma \left( \frac{r}{\epsilon} \right) t^r e^{-\gamma} + \int_0^t \gamma \left( \frac{r}{\epsilon} \right) t^r e^{-\gamma} d\tau \]  

(L.8)

Since, for some \( \beta \), \( ||d(t,x)|| \leq \beta \), we also have that \( \gamma(t) \leq \beta \). Note that, for all \( t' \geq 0 \), \( t' e^{-\gamma t} \leq e^{-t} \), and \( t' \leq t' \), so that:

\[ ||\epsilon w_t(t,x)|| \leq \sup_{r \in [0,\sqrt{t}]} \left( \gamma \left( \frac{t'}{\epsilon} \right) \right) t' e^{-\gamma} + \sup_{r \geq \sqrt{t}} \left( \gamma \left( \frac{t'}{\epsilon} \right) \right) t' e^{-\gamma} \]  

(L.9)

\[ + \int_0^\sqrt{t} \gamma \left( \frac{r}{\epsilon} \right) r e^{-\gamma} d\tau + \int_0^\infty \gamma \left( \frac{r}{\epsilon} \right) r e^{-\gamma} d\tau \]
This, in turn, implies that

\[
\|\epsilon w_i(t,x)\| \leq \beta \sqrt{\epsilon} + \gamma \left( \frac{1}{\sqrt{\epsilon}} \right) \epsilon^{-\frac{3}{2}} + \beta \frac{\epsilon}{2} + \gamma \left( \frac{1}{\sqrt{\epsilon}} \right) (1 + \sqrt{\epsilon}) \epsilon^{-\frac{1}{2}}
\]

\[=: \xi(\epsilon) \tag{L10} \]

Clearly \( \xi(\epsilon) \in K \). From (L1), it follows that:

\[
\frac{\partial w_i(t,x)}{\partial t} - d(t,x) = -\epsilon w_i(t,x) \tag{L11}
\]

so that both (2.6) and (2.7) are satisfied.

If \( \gamma(T) = a/T' \), then the right-hand side of (L8) can be computed explicitly:

\[
\sup_{t' \geq 0} a \epsilon' (t')^{1-r} e^{-t'} = a \epsilon' (1-r)^{1-r} e^{-t'} \leq a \epsilon' \tag{L12}
\]

and, with \( \Gamma \) denoting the standard gamma function:

\[
\int_0^\infty a \epsilon' (t')^{1-r} e^{-t'} dt' = a \epsilon' \Gamma(2-r) \leq a \epsilon' \tag{L13}
\]

Defining \( \xi(\epsilon) = 2a \epsilon' \), the second part of the lemma is verified.

Proof of Lemma 2.2

Define \( w_i(t,x) \) as in Lemma 2.1. Consequently,

\[
\frac{\partial w_i(t,x)}{\partial x} = \frac{\partial}{\partial x} \left[ \int_0^t d(r,x) e^{-\gamma(t-r)} dr \right] = \int_0^t \left( \frac{\partial}{\partial x} d(r,x) \right) e^{-\gamma(t-r)} dr \tag{L14}
\]

Since \( \frac{\partial d(t,x)}{\partial x} \) is zero mean, and is bounded, Lemma 2.1 can be applied to \( \frac{\partial d(t,x)}{\partial x} \), and inequality (2.6) of Lemma 2.1 becomes inequality (2.10) of Lemma 2.2. Note that since \( \frac{\partial d(t,x)}{\partial x} \) is bounded, and \( d(t,0)=0 \) for all \( t \geq 0 \), \( d(t,x) \) is Lipschitz. Since \( d(t,x) \) is zero mean, with convergence function \( \gamma(T)||x|| \), the proof of Lemma 2.1 can be extended, with an additional factor \( ||x|| \). This leads directly to (2.8) and (2.9) (although the function \( \xi(\epsilon) \) may be different from that obtained with \( \frac{\partial d(t,x)}{\partial x} \), these functions can be replaced by a single \( \xi(\epsilon) \)).

Proof of Lemma 2.3

The proof proceeds in two steps.

Step 1: for \( \epsilon \) sufficiently small, and for \( t \) fixed, the transformation (2.15) is a homeomorphism.
Applying Lemma 2.2, and let $\epsilon_1$ such that $\xi(\epsilon_1) < 1$. Given $z \in B_k$, the corresponding $x$ such that:

$$z = x - \epsilon_1 w(t, z) \quad (L15)$$

may not belong to $B_k$. Similarly, given $z \in B_k$, the solution $z$ of (L15) may not exist in $B_k$. However, for any $x, z$ satisfying (L15), inequality (2.8) implies that:

$$(1 - \xi(\epsilon)) ||z|| \leq ||z|| \leq (1 + \xi(\epsilon)) ||z|| \quad (L16)$$

Define:

$$h'(\epsilon) = \min \left( h(1 - \xi(\epsilon)), \frac{h}{1 + \xi(\epsilon)} \right) \quad (L17)$$

and note that $h'(\epsilon) \to h$ as $\epsilon \to 0$.

We now show that:

- for all $z \in B_k$, there exists a unique $x \in B_k$ such that (L15) is satisfied,
- for all $x \in B_k$, there exists a unique $z \in B_k$ such that (L15) is satisfied.

In both cases, $||z - z|| \leq \xi(\epsilon) h$.

The first part follows directly from (L16), (L17). The fact that $||z - z|| \leq \xi(\epsilon) h$ also follows from (L16), and implies that, if a solution $z$ exists to (L15), it must lie in the closed ball $U$ of radius $\xi(\epsilon) h$ around $z$. It can be checked, using (2.10), that the mapping $F_z(z) = z - \epsilon_1 w(t, z)$ is a contraction mapping in $U$, provided that $\xi(\epsilon) < 1$. Consequently, $F$ has a unique fixed point $z$ in $U$. This solution is also a solution of (L15), and since it is unique in $U$, it is also unique in $B_k$ (and actually in $R^n$). For $x \in B_k$, but outside $B_k$, there is no guarantee that a solution $z$ exists in $B_k$, but if it exists, it is again unique in $B_k$. Consequently, the map defined by (L15) is well-defined. From the smoothness of $w(t, z)$ with respect to $z$, it follows that the map is a homeomorphism.

Step 2: the transformation of variable leads to the differential equation (2.16)

Applying (L15) to the system (2.1):

$$\left( I + \epsilon \frac{\partial w}{\partial z} \right) \dot{z} = \epsilon f_w(z) + \epsilon \left( f(t, z, 0) - f_w(z) \frac{\partial w}{\partial z} \right) + \epsilon \left( f(t, x + \epsilon w(t, z) - f(t, z, \epsilon) \right) + \epsilon \left( f(t, z, \epsilon) - f(t, z, 0) \right)$$

$$:= \epsilon f_w(z) + \epsilon p'(t, z, \epsilon) \quad (L18)$$

where, using the assumptions, and the results of Lemma 2.2:
\[ ||p'(t,z,\epsilon)|| \leq \xi(\epsilon)||z|| + \xi(\epsilon)l_{41}||z|| + \xi(\epsilon)l_{52}||z|| \] (L.19)

For \(\epsilon \leq \epsilon_1\), (2.10) implies that \((I + \epsilon \frac{\partial w_t}{\partial z})^{-1}\) has a bounded inverse for all \(t \geq 0, z \in B_k\). Consequently, \(z\) satisfies the differential equation:

\[ \dot{z} = \left( I + \epsilon \frac{\partial w_t}{\partial z} \right)^{-1} \left( \epsilon f_{av}(z) + \epsilon p'(t,z,\epsilon) \right) \]

\[ = \epsilon f_{av}(z) + \epsilon p(t,z,\epsilon) \quad z(0) = x_0 \] (L.20)

where:

\[ p(t,z,\epsilon) = \left( I + \epsilon \frac{\partial w_t}{\partial z} \right)^{-1} \left( p'(t,z,\epsilon) - \epsilon \frac{\partial w_t}{\partial z} f_{av}(z) \right) \] (L.21)

and:

\[ ||p(t,z,\epsilon)|| \leq \frac{1}{1 - \xi(\epsilon_1)} \left( \xi(\epsilon)l_{41} + \xi(\epsilon)l_{52} + \xi(\epsilon)l_{6v} \right)||z|| \]

\[ := \psi(\epsilon)||z|| \] (L.22)

for all \(t \geq 0, \epsilon \leq \epsilon_1, z \in B_k\).

**Generalised Bellman-Gronwall Lemma (cf. [7], p 169)**

If \(z(t), a(t), u(t)\) are positive functions satisfying:

\[ z(t) \leq \int_0^t a(\tau)z(\tau)\,d\tau + u(t) \] (L.23)

for all \(t \in [0,T]\), and \(u(t)\) is differentiable,

Then:

\[ z(t) \leq u(0) e^{\int_0^t a(\tau)d\tau} + \int_0^t \dot{u}(\tau) e^{\int_\tau^t a(\sigma)d\sigma} \,d\tau \] (L.24)

for all \(t \in [0,T]\).
Fig 3.1 Block diagram of adaptive identifier.
Fig 3.2 Trajectories of parameter error $\phi_1(=e_1-e_1^*)$ and $\phi_{cov}$ with three different adaptation gains (a) $\epsilon=1$ (b) $\epsilon=0.5$ (c) $\epsilon=0.1$
Fig 3.3 Trajectories of parameter error $\phi_d (= d_1 - d_1^*)$ and $\phi_{av2}$ with three different adaptation gains (a) $\epsilon=1$ (b) $\epsilon=0.5$ (c) $\epsilon=0.1$
Fig 3.4 Trajectories of Lyapunov function $V(\phi)$ and $V(\dot{\phi},\ddot{\phi})$ with three adaptation gains (a) $\epsilon=1$ (b) $\epsilon=0.5$ (c) $\epsilon=0.1$ using log scale.
Fig 5.1 Block diagram of model reference adaptive control system
Fig 5.2 Trajectories of parameter error $\phi_2(= c_0 - c_0^\circ)$ and $\phi_{s_1}$ with three adaptation gains (a) $\epsilon = 1$ (b) $\epsilon = 0.5$ (c) $\epsilon = 0.1$
Fig 5.3 Trajectories of parameter error $\phi_2 = d_{\theta} - \hat{d}_{\theta}$ and $\phi_{av2}$ with three adaptation gains (a) $\varepsilon=1$ (b) $\varepsilon=0.5$ (c) $\varepsilon=0.1$ using log scale.
Fig 5.4 Phase plot for $\phi_2(\phi_1)$ and $\phi_{av2}(\phi_{av1})$ with three different adaptation gains (a) $\epsilon=1$ (b) $\epsilon=0.5$ (c) $\epsilon=0.1$; and $r=3\sin(4t)$. 