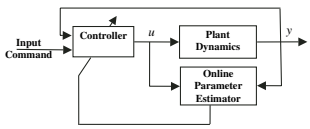

Chapter 3
Parameter Identification: Continuous Time

3.1 Introduction

The purpose of this chapter is to present the design, analysis, and simulation of a wide class of algorithms that can be used for online parameter identification of continuous-time plants. The online identification procedure involves the following three steps.

Step 1. Lump the unknown parameters in a vector θ^* and express them in the form of the parametric model SPM, DPM, B-SPM, or B-DPM.

Step 2. Use the estimate θ of θ^* to set up the *estimation model* that has the same form as the parametric model. The difference between the outputs of the estimation and parametric models, referred to as the *estimation error*, reflects the distance of the estimated parameters $\theta(t)$ from the unknown parameters θ^* weighted by some signal vector. The estimation error is used to drive the adaptive law that generates $\theta(t)$ online. The adaptive law is a differential equation of the form

$$\dot{\theta} = H(t)\varepsilon,$$

where ε is the estimation error that reflects the difference between $\theta(t)$ and θ^* and $H(t)$ is a time-varying gain vector that depends on measured signals. A wide class of adaptive laws with different $H(t)$ and ε may be developed using optimization techniques and Lyapunov-type stability arguments.

Step 3. Establish conditions that guarantee that $\theta(t)$ converges to θ^* with time. This step involves the design of the plant input so that the signal vector $\phi(t)$ in the parametric model is *persistently exciting* (a notion to be defined later on), i.e., it has certain properties that guarantee that the measured signals that drive the adaptive law carry sufficient information about the unknown parameters. For example, for $\phi(t) = 0$, we have $z = \theta^{*T}\phi = 0$, and the measured signals ϕ, z carry no information about θ^*. Similar arguments could be made for ϕ that is orthogonal to θ^* leading to $z = 0$ even though $\theta \neq \theta^*$, etc.

We demonstrate the three design steps using the following example of a scalar plant.

3.2 Example: One-Parameter Case

Consider the first-order plant model

$$y = \frac{a}{s+2}u, \tag{3.1}$$

where a is the only unknown parameter and y and u are the measured output and input of the system, respectively.

Step 1: Parametric Model We write (3.1) as

$$y = a\frac{1}{s+2}u = au_f, \tag{3.2}$$

where $u_f = \frac{1}{s+2}u$. Since u is available for measurement, u_f is also available for measurement. Therefore, (3.2) is in the form of the SPM

$$z = \theta^*\phi, \tag{3.3}$$

where $\theta^* = a$ and $z = y$, $\phi = u_f$ are available for measurement.

Step 2: Parameter Identification Algorithm This step involves the development of an estimation model and an estimation error used to drive the adaptive law that generates the parameter estimates.

Estimation Model and Estimation Error The estimation model has the same form as the SPM with the exception that the unknown parameter θ^* is replaced with its estimate at time t, denoted by $\theta(t)$, i.e.,

$$\hat{z} = \theta(t)\phi, \tag{3.4}$$

where $\hat{z}$ is the estimate of z based on the parameter estimate $\theta(t)$ at time t. It is obvious that the difference between z and $\hat{z}$ is due to the difference between $\theta(t)$ and θ^*. As $\theta(t)$ approaches θ^* with time we would expect that $\hat{z}$ would approach z at the same time. (Note that the reverse is not true, i.e., $\hat{z}(t) = z(t)$ does not imply that $\theta(t) = \theta^*$; see Problem 1.) Since θ^* is unknown, the difference $\tilde{\theta} = \theta(t) - \theta^*$ is not available for measurement. Therefore, the only signal that we can generate, using available measurements, that reflects the difference between $\theta(t)$ and θ^* is the error signal

$$\varepsilon = \frac{z - \hat{z}}{m_s^2}, \tag{3.5}$$

which we refer to as the *estimation error*. $m_s^2 \geq 1$ is a normalization signal[1] designed to guarantee that $\frac{\phi}{m_s}$ is bounded. This property of m_s is used to establish the boundedness of the estimated parameters even when ϕ is not guaranteed to be bounded. A straightforward choice for m_s in this example is $m_s^2 = 1 + \alpha\phi^2$, $\alpha > 0$. If ϕ is bounded, we can take $\alpha = 0$,

[1]Note that any $m_s^2 \geq$ nonzero constant is adequate. The use of a lower bound 1 is without loss of generality.

i.e., $m_s^2 = 1$. Using (3.4) in (3.5), we can express the estimation error as a function of the parameter error $\tilde{\theta} = \theta(t) - \theta^*$, i.e.,

$$\varepsilon = -\frac{\tilde{\theta}\phi}{m_s^2}. \tag{3.6}$$

Equation (3.6) shows the relationship between the estimation error ε and the parameter error $\tilde{\theta}$. It should be noted that ε cannot be generated using (3.6) because the parameter error $\tilde{\theta}$ is not available for measurement. Consequently, (3.6) can be used only for analysis.

Adaptive Law A wide class of adaptive laws or parameter estimators for generating $\theta(t)$, the estimate of θ^*, can be developed using (3.4)–(3.6). The simplest one is obtained by using the SPM (3.3) and the fact that ϕ is scalar to write

$$\theta(t) = \frac{z(t)}{\phi(t)}, \tag{3.7}$$

provided $\phi(t) \neq 0$. In practice, however, the effect of noise on the measurements of $\phi(t)$, especially when $\phi(t)$ is close to zero, may lead to erroneous parameter estimates. Another approach is to update $\theta(t)$ in a direction that minimizes a certain cost of the estimation error ε. With this approach, $\theta(t)$ is adjusted in a direction that makes $|\varepsilon|$ smaller and smaller until a minimum is reached at which $|\varepsilon| = 0$ and updating is terminated. As an example, consider the cost criterion

$$J(\theta) = \frac{\varepsilon^2 m_s^2}{2} = \frac{(z - \theta\phi)^2}{2m_s^2}, \tag{3.8}$$

which we minimize with respect to θ using the gradient method to obtain

$$\dot{\theta} = -\gamma \nabla J(\theta), \tag{3.9}$$

where $\gamma > 0$ is a scaling constant or step size which we refer to as the *adaptive gain* and where $\nabla J(\theta)$ is the gradient of J with respect to θ. In this scalar case,

$$\nabla J(\theta) = \frac{dJ}{d\theta} = -\frac{(z - \theta\phi)}{m_s^2}\phi = -\varepsilon\phi,$$

which leads to the adaptive law

$$\dot{\theta} = \gamma\varepsilon\phi, \quad \theta(0) = \theta_0. \tag{3.10}$$

Step 3: Stability and Parameter Convergence The adaptive law should guarantee that the parameter estimate $\theta(t)$ and the speed of adaptation $\dot{\theta}$ are bounded and that the estimation error ε gets smaller and smaller with time. These conditions still do not imply that $\theta(t)$ will get closer and closer to θ^* with time unless some conditions are imposed on the vector $\phi(t)$, referred to as the *regressor vector*.

Let us start by using (3.6) and the fact that $\dot{\tilde{\theta}} = \dot{\theta} - \dot{\theta}^* = \dot{\theta}$ (due to θ^* being constant) to express (3.10) as

$$\dot{\tilde{\theta}} = -\gamma \frac{\phi^2}{m_s^2}\tilde{\theta}, \quad \tilde{\theta}(0) = \tilde{\theta}_0. \tag{3.11}$$

This is a scalar linear time-varying differential equation whose solution is

$$\tilde{\theta}(t) = e^{-\gamma \int_0^t \frac{\phi^2(\tau)}{m_s^2(\tau)} d\tau} \tilde{\theta}_0, \tag{3.12}$$

which implies that for

$$\int_0^t \frac{\phi^2(\tau)}{m_s^2(\tau)} d\tau \geq \alpha_0 t \tag{3.13}$$

and some $\alpha_0 > 0$, $\tilde{\theta}(t)$ converges to zero exponentially fast, which in turn implies that $\theta(t) \to \theta^*$ exponentially fast. It follows from (3.12) that $\theta(t)$ is always bounded for any $\phi(t)$ and from (3.11) that $\dot{\theta}(t) = \dot{\tilde{\theta}}(t)$ is bounded due to $\frac{\phi(t)}{m_s(t)}$ being bounded.

Another way to analyze (3.11) is to use a Lyapunov-like approach as follows: We consider the function

$$V = \frac{\tilde{\theta}^2}{2\gamma}.$$

Then

$$\dot{V} = \frac{\tilde{\theta}}{\gamma}\frac{d\tilde{\theta}}{dt} = -\frac{\phi^2}{m_s^2}\tilde{\theta}^2 \leq 0$$

or, using (3.6),

$$\dot{V} = -\frac{\phi^2}{m_s^2}\tilde{\theta}^2 = -\varepsilon^2 m_s^2 \leq 0. \tag{3.14}$$

We should note that $\dot{V} = -\varepsilon^2 m_s^2 \leq 0$ implies that $\dot{V}$ is a negative semidefinite function in the space of $\tilde{\theta}$. $\dot{V}$ in this case is not negative definite in the space of $\tilde{\theta}$ because it can be equal to zero when $\tilde{\theta}$ is not zero. Consequently, if we apply the stability results of the Appendix, we can conclude that the equilibrium $\tilde{\theta}_e = 0$ of (3.11) is uniformly stable (u.s.) and that the solution of (3.11) is uniformly bounded (u.b.). These results are not as useful, as our objective is asymptotic stability, which implies that the parameter error converges to zero. We can use the properties of V, $\dot{V}$, however, to obtain additional properties for the solution of (3.11) as follows.

Since $V > 0$ and $\dot{V} \leq 0$, it follows that (see the Appendix) V is bounded, which implies that $\tilde{\theta}$ is bounded and V converges to a constant, i.e., $\lim_{t\to\infty} V(t) = V_\infty$. Let us now integrate both sides of (3.14). We have

$$\int_0^t \dot{V}(\tau)d\tau = -\int_0^t \varepsilon^2(\tau)m_s^2(\tau)d\tau$$

or

$$V(t) - V(0) = -\int_0^t \varepsilon^2(\tau)m_s^2(\tau)d\tau. \tag{3.15}$$

Since $V(t)$ converges to the limit V_∞ as $t \to \infty$, it follows from (3.15) that

$$\int_0^\infty \varepsilon^2(\tau) m_s^2(\tau) d\tau = V(0) - V_\infty < \infty,$$

i.e., εm_s is square integrable or $\varepsilon m_s \in \mathcal{L}_2$. Since $m_s^2 \geq 1$, we have $\varepsilon^2 \leq \varepsilon^2 m_s^2$, which implies $\varepsilon \in \mathcal{L}_2$. From (3.6) we conclude that $\frac{\tilde{\theta}\phi}{m_s} \in \mathcal{L}_2$ due to $\varepsilon m_s \in \mathcal{L}_2$. Using (3.10), we write

$$\dot{\theta} = \gamma \varepsilon m_s \frac{\phi}{m_s}.$$

Since $\frac{\phi}{m_s}$ is bounded and $\varepsilon m_s \in \mathcal{L}_2 \cap \mathcal{L}_\infty$, it follows (see Problem 2) that $\dot{\theta} \in \mathcal{L}_2 \cap \mathcal{L}_\infty$. In summary, we have established that the adaptive law (3.10) guarantees that (i) $\theta \in \mathcal{L}_\infty$ and (ii) $\varepsilon, \varepsilon m_s, \dot{\theta} \in \mathcal{L}_2 \cap \mathcal{L}_\infty$ independent of the boundedness of ϕ. The $\mathcal{L}_2$ property of ε, εm_s, and $\dot{\theta}$ indicates that the estimation error and the speed of adaptation $\dot{\theta}$ are bounded in the $\mathcal{L}_2$ sense, which in turn implies that their average value tends to zero with time.

It is desirable to also establish that ε, εm_s, and $\dot{\theta}$ go to zero as $t \to \infty$, as such a property will indicate the end of adaptation and the completion of learning. Such a property can be easily established when the input u is bounded (see Problem 3).

The above properties still do not guarantee that $\theta(t) \to \theta^*$ as $t \to \infty$. In order to establish that $\theta(t) \to \theta^*$ as $t \to \infty$ exponentially fast, we need to restrict $\frac{\phi}{m_s}$ to be persistently exciting (PE), i.e., to satisfy

$$\frac{1}{T} \int_t^{t+T} \frac{\phi^2(\tau)}{m_s^2} d\tau \geq \alpha_0 > 0 \tag{3.16}$$

$\forall t \geq 0$ and some constants $T, \alpha_0 > 0$. The PE property of $\frac{\phi}{m_s}$ is guaranteed by choosing the input u appropriately. Appropriate choices of u for this particular example include (i) $u = c > 0$, (ii) $u = \sin \omega t$ for any $\omega \neq 0$ and any bounded input u that is not vanishing with time. The condition (3.16) is necessary and sufficient for exponential convergence of $\theta(t) \to \theta^*$ (see Problem 4).

The PI algorithm for estimating the constant a in the plant (3.1) can now be summarized as

$$\begin{aligned}
&\dot{\theta} = \gamma \varepsilon \phi, && \theta(0) = \theta_0, \\
&\varepsilon = \frac{(z - \hat{z})}{m_s^2}, && \hat{z} = \theta \phi, \\
&z = y, && \phi = \frac{1}{s+2} u, \qquad m_s^2 = 1 + \phi^2,
\end{aligned}$$

where $\theta(t)$ is the estimate of the constant a in (3.1).

The above analysis for the scalar example carries over to the vector case without any significant modifications, as demonstrated in the next section. One important difference, however, is that in the case of a single parameter, convergence of the Lyapunov-like function V to a constant implies that the estimated parameter converges to a constant. Such a result cannot be established in the case of more than one parameter for the gradient algorithm.

3.3 Example: Two Parameters

Consider the plant model

$$y = \frac{b}{s+a}u, \tag{3.17}$$

where a, b are unknown constants. Let us assume that $y, \dot{y}, u$ are available for measurement. We would like to generate online estimates for the parameters a, b.

Step 1: Parametric Model Since y, $\dot{y}$ are available for measurement, we can express (3.17) in the SPM form

$$z = \theta^{*T}\phi,$$

where $z = \dot{y}$, $\theta^* = [b, a]^T$, $\phi = [u, -y]^T$, and z, ϕ are available for measurement.

Step 2: Parameter Identification Algorithm

Estimation Model

$$\hat{z} = \theta^T\phi,$$

where $\theta(t)$ is the estimate of θ^* at time t.

Estimation Error

$$\varepsilon = \frac{z - \hat{z}}{m_s^2} = \frac{z - \theta^T\phi}{m_s^2}, \tag{3.18}$$

where m_s is the normalizing signal such that $\frac{\phi}{m_s} \in \mathcal{L}_\infty$. A straightforward choice for m_s is $m_s^2 = 1 + \alpha\phi^T\phi$ for any $\alpha > 0$.

Adaptive Law We use the gradient method to minimize the cost,

$$J(\theta) = \frac{\varepsilon^2 m_s^2}{2} = \frac{(z - \theta^T\phi)^2}{2m_s^2} = \frac{(z - \theta_1\phi_1 - \theta_2\phi_2)^2}{2m_s^2},$$

where $\phi_1 = u$, $\phi_2 = -y$, and set

$$\dot{\theta} = -\Gamma\nabla J,$$

where

$$\nabla J = \left[\frac{\partial J}{\partial\theta_1}, \frac{\partial J}{\partial\theta_2}\right]^T,$$

$\Gamma = \Gamma^T > 0$ is the adaptive gain, and θ_1, θ_2 are the elements of $\theta = [\theta_1, \theta_2]^T$. Since

$$\frac{\partial J}{\partial\theta_1} = -\frac{(z - \theta^T\phi)}{m_s^2}\phi_1 = -\varepsilon\phi_1, \qquad \frac{\partial J}{\partial\theta_2} = -\frac{(z - \theta^T\phi)}{m_s^2}\phi_2 = -\varepsilon\phi_2,$$

we have

$$\dot{\theta} = \Gamma\varepsilon\phi, \qquad \theta(0) = \theta_0, \tag{3.19}$$

which is the adaptive law for updating $\theta(t)$ starting from some initial condition $\theta(0) = \theta_0$.

Step 3: Stability and Parameter Convergence As in the previous example, the equation for the parameter error $\tilde{\theta} = \theta - \theta^*$ is obtained from (3.18), (3.19) by noting that

$$\varepsilon = \frac{z - \theta^T \phi}{m_s^2} = \frac{\theta^{*T}\phi - \theta^T \phi}{m_s^2} = -\frac{\tilde{\theta}^T \phi}{m_s^2} = -\frac{\phi^T \tilde{\theta}}{m_s^2} \tag{3.20}$$

and $\dot{\tilde{\theta}} = \dot{\theta}$, i.e.,

$$\dot{\tilde{\theta}} = \Gamma \phi \varepsilon = -\Gamma \frac{\phi \phi^T}{m_s^2} \tilde{\theta}. \tag{3.21}$$

It is clear from (3.21) that the stability of the equilibrium $\tilde{\theta}_e = 0$ will very much depend on the properties of the time-varying matrix $-\frac{\Gamma \phi \phi^T}{m_s^2}$, which in turn depends on the properties of ϕ. For simplicity let us assume that the plant is stable, i.e., $a > 0$. If we choose $m_s^2 = 1$, $\Gamma = \gamma I$ for some $\gamma > 0$ and a constant input $u = c_0 > 0$, then at steady state $y = c_1 \triangleq \frac{c_0 b}{a} \neq 0$ and $\phi = [c_0, -c_1]^T$, giving

$$-\frac{\Gamma \phi \phi^T}{m_s^2} = -\gamma \begin{bmatrix} c_0^2 & -c_0 c_1 \\ -c_0 c_1 & c_1^2 \end{bmatrix} \triangleq A,$$

i.e.,

$$\dot{\tilde{\theta}} = A \tilde{\theta},$$

where A is a constant matrix with eigenvalues 0, $-\gamma(c_0^2 + c_1^2)$, which implies that the equilibrium $\theta_e = 0$ is only marginally stable; i.e., $\tilde{\theta}$ is bounded but does not necessarily converge to 0 as $t \to \infty$. The question that arises in this case is what properties of ϕ guarantee that the equilibrium $\tilde{\theta}_e = 0$ is exponentially stable. Given that

$$\phi = H(s)u,$$

where for this example $H(s) = [1, -\frac{b}{s+a}]^T$, the next question that comes up is how to choose u to guarantee that ϕ has the appropriate properties that imply exponential stability for the equilibrium $\tilde{\theta}_e = 0$ of (3.21). Exponential stability for the equilibrium point $\tilde{\theta}_e = 0$ of (3.21) in turn implies that $\theta(t)$ converges to θ^* exponentially fast. As demonstrated above for the two-parameter example, a constant input $u = c_0 > 0$ does not guarantee exponential stability. We answer the above questions in the following section.

3.4 Persistence of Excitation and Sufficiently Rich Inputs

We start with the following definition.

Definition 3.4.1. *The vector $\phi \in \mathcal{R}^n$ is PE with level α_0 if it satisfies*

$$\int_t^{t+T_0} \phi(\tau)\phi^T(\tau)d\tau \geq \alpha_0 T_0 I \tag{3.22}$$

for some $\alpha_0 > 0$, $T_0 > 0$ and $\forall t \geq 0$.

Since $\phi\phi^T$ is always positive semidefinite, the PE condition requires that its integral over any interval of time of length T_0 is a positive definite matrix.

Definition 3.4.2. *The signal $u \in \mathcal{R}$ is called sufficiently rich of order n if it contains at least $\frac{n}{2}$ distinct nonzero frequencies.*

For example, $u = \sum_{i=1}^{10} \sin \omega_i t$, where $\omega_i \neq \omega_j$ for $i \neq j$ is sufficiently rich of order 20. A more general definition of sufficiently rich signals and associated properties may be found in [95].

Let us consider the signal vector $\phi \in \mathcal{R}^n$ generated as

$$\phi = H(s)u, \tag{3.23}$$

where $u \in \mathcal{R}$ and $H(s)$ is a vector whose elements are transfer functions that are strictly proper with stable poles.

Theorem 3.4.3. *Consider* (3.23) *and assume that the complex vectors $H(j\omega_1), \ldots, H(j\omega_n)$ are linearly independent on the complex space $\mathcal{C}^n$ $\forall \omega_1, \omega_2, \ldots, \omega_n \in \mathcal{R}$, where $\omega_i \neq \omega_j$ for $i \neq j$. Then ϕ is PE if and only if u is sufficiently rich of order n.*

Proof. The proof of Theorem 3.4.3 can be found in [56, 87]. □

We demonstrate the use of Theorem 3.4.3 for the example in section 3.3, where

$$\phi = H(s)u$$

and

$$H(s) = \begin{bmatrix} 1 \\ -\frac{b}{s+a} \end{bmatrix}.$$

In this case $n = 2$ and

$$H(j\omega_1) = \begin{bmatrix} 1 \\ -\frac{b}{j\omega_1+a} \end{bmatrix}, \qquad H(j\omega_2) = \begin{bmatrix} 1 \\ -\frac{b}{j\omega_2+a} \end{bmatrix}.$$

We can show that the matrix $[H(j\omega_1), H(j\omega_2)]$ is nonsingular, which implies that $H(j\omega_1), H(j\omega_2)$ are linearly independent for any ω_1, ω_2 different than zero and $\omega_1 \neq \omega_2$.

Let us choose

$$u = \sin \omega_0 t$$

for some $\omega_0 \neq 0$ which is sufficiently rich of order 2. According to Theorem 3.4.3, this input should guarantee that ϕ is PE for the example in section 3.3. Ignoring the transient terms that converge to zero exponentially fast, we can show that at steady state

$$\phi = \begin{bmatrix} \sin \omega_0 t \\ c_0 \sin(\omega_0 t + \varphi_0) \end{bmatrix},$$

where

$$c_0 = \frac{|b|}{\sqrt{\omega_0^2 + a^2}}, \qquad \varphi_0 = \arg\left(\frac{-b}{j\omega_0 + a}\right).$$

Now

$$\phi\phi^T = \begin{bmatrix} \sin^2 \omega_0 t & c_0 \sin \omega_0 t \sin(\omega_0 t + \varphi_0) \\ c_0 \sin \omega_0 t \sin(\omega_0 t + \varphi_0) & c_0^2 \sin^2(\omega_0 t + \varphi) \end{bmatrix}$$

and

$$\int_t^{t+T_0} \phi(\tau)\phi^T(\tau)d\tau = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix},$$

where

$$\begin{aligned} a_{11} &= \frac{T_0}{2} - \frac{\sin 2\omega_0(t+T_0) - \sin 2\omega_0 t}{4\omega_0}, \\ a_{12} &= c_0 \frac{T_0}{2} \cos\varphi_0 + c_0 \frac{\sin\varphi_0}{4\omega_0}(\cos 2\omega_0 t - \cos 2\omega_0(t+T_0)), \\ a_{22} &= c_0^2 \frac{T_0}{2} - c_0^2 \frac{\sin 2(\omega_0(t+T_0)+\varphi_0) - \sin 2(\omega_0 t + \varphi_0)}{4\omega_0}. \end{aligned}$$

Choosing $T_0 = \frac{\pi}{\omega_0}$ it follows that

$$a_{11} = \frac{T_0}{2}, \qquad a_{12} = \frac{T_0 c_0}{2}\cos\varphi_0, \qquad a_{22} = \frac{T_0}{2}c_0^2$$

and

$$\int_t^{t+T_0} \phi(\tau)\phi^T(\tau)d\tau = \frac{T_0}{2}\begin{bmatrix} 1 & c_0\cos\varphi_0 \\ c_0\cos\varphi_0 & c_0^2 \end{bmatrix},$$

which is a positive definite matrix. We can verify that for $\alpha_0 = \frac{1}{2}\frac{(1-\cos^2\varphi_0)c_0^2}{1+c_0^2} > 0$,

$$\int_t^{t+T_0} \phi(\tau)\phi^T(\tau)d\tau \geq T_0\alpha_0 I,$$

which implies that ϕ is PE.

Let us consider the plant model

$$y = \frac{b(s^2+4)}{(s+5)^3}u,$$

where b is the only unknown parameter. A suitable parametric model for estimating b is

$$z = \theta^*\phi,$$

where

$$z = y, \qquad \theta^* = b, \qquad \phi = \frac{s^2+4}{(s+5)^3}u.$$

In this case $\phi \in \mathcal{R}$ and $H(s) = \frac{s^2+4}{(s+5)^3}$; i.e., $n = 1$ in Theorem 3.4.3. Let us use Theorem 3.4.3 to choose a sufficiently rich signal u that guarantees ϕ to be PE. In this case,

according to the linear independence condition of Theorem 3.4.3 for the case of $n = 1$, we should have

$$|H(j\omega_0)| = \frac{4 - \omega_0^2}{(25 + \omega_0^2)^{3/2}} \neq 0$$

for any $\omega_0 \neq 0$. This condition is clearly violated for $\omega_0 = 2$, and therefore a sufficiently rich input of order 1 may not guarantee ϕ to be PE. Indeed, the input $u = \sin 2t$ leads to $y = 0$, $\phi = 0$ at steady state, which imply that the output y and regressor ϕ carry no information about the unknown parameter b. For this example $u = \sin \omega_0 t$ will guarantee ϕ to be PE, provided $\omega_0 \neq 2$. Also, $u = \text{constant} \neq 0$ and $u = \sum_{i=1}^{m} \sin \omega_i t$, $m \geq 2$, will guarantee that ϕ is PE. In general, for each two unknown parameters we need at least a single nonzero frequency to guarantee PE, provided of course that $H(s)$ does not lose its linear independence as demonstrated by the above example.

The two-parameter case example presented in section 3.3 leads to the differential equation (3.21), which has exactly the same form as in the case of an arbitrary number of parameters. In the following section, we consider the case where θ^*, ϕ are of arbitrary dimension and analyze the convergence properties of equations of the form (3.21).

3.5 Example: Vector Case

Consider the SISO system described by the I/O relation

$$y = G(s)u, \qquad G(s) = \frac{Z(s)}{R(s)} = k_p \frac{\bar{Z}(s)}{R(s)}, \tag{3.24}$$

where u and y are the plant scalar input and output, respectively,

$$\begin{aligned} R(s) &= s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0, \\ Z(s) &= b_m s^m + \cdots + b_1 s + b_0, \end{aligned}$$

and $k_p = b_m$ is the high-frequency gain.[2]

We can also express (3.24) as an nth-order differential equation given by

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1\dot{y} + a_0 y = b_m u^{(m)} + \cdots + b_1\dot{u} + b_0 u. \tag{3.25}$$

Parametric Model Lumping all the parameters in (3.25) in the vector

$$\theta^* = [b_m, \ldots, b_0, a_{n-1}, \ldots, a_0]^T,$$

we can rewrite (3.25) as

$$y^{(n)} = \theta^{*T}[u^{(m)}, \ldots, u, -y^{(n-1)}, \ldots, -y]^T. \tag{3.26}$$

Filtering each side of (3.26) with $\frac{1}{\Lambda(s)}$, where $\Lambda(s) = s^n + \lambda_{n-1}s^{n-1} + \cdots + \lambda_1 s + \lambda_0$ is a monic Hurwitz polynomial, we obtain the parametric model

$$z = \theta^{*T}\phi, \tag{3.27}$$

[2]At high frequencies or large s, the plant behaves as $\frac{k_p}{s^{n-m}}$; therefore, k_p is termed high-frequency gain.

where

$$
\begin{aligned}
z &= \frac{1}{\Lambda(s)} y^{(n)} = \frac{s^n}{\Lambda(s)} y, \\
\theta^* &= [b_m, \ldots, b_0, a_{n-1}, \ldots, a_0]^T \in \mathcal{R}^{n+m+1}, \\
\phi &= \left[\frac{s^m}{\Lambda(s)} u, \ldots, \frac{1}{\Lambda(s)} u, -\frac{s^{n-1}}{\Lambda(s)} y, \ldots, -\frac{1}{\Lambda(s)} y \right]^T .
\end{aligned}
$$

If $\bar{Z}(s)$ is Hurwitz, a bilinear model can be obtained as follows: Consider the polynomials $P(s) = p_{n-1}s^{n-1} + \cdots + p_1 s + p_0$, $Q(s) = s^{n-1} + q_{n-2}s^{n-2} + \cdots + q_1 s + q_0$ which satisfy the Diophantine equation (see the Appendix)

$$k_p \bar{Z}(s) P(s) + R(s) Q(s) = \bar{Z}(s) \Lambda_0(s),$$

where $\Lambda_0(s)$ is a monic Hurwitz polynomial of order $2n - m - 1$. If each term in the above equation operates on the signal y, we obtain

$$k_p \bar{Z}(s) P(s) y + Q(s) R(s) y = \bar{Z}(s) \Lambda_0(s) y.$$

Substituting for $R(s)y = k_p \bar{Z}(s) u$, we obtain

$$k_p \bar{Z}(s) P(s) y + k_p \bar{Z}(s) Q(s) u = \bar{Z}(s) \Lambda_0(s) y.$$

Filtering each side with $\frac{1}{\Lambda_0(s)\bar{Z}(s)}$, we obtain

$$y = k_p \left[\frac{P(s)}{\Lambda_0(s)} y + \frac{Q(s)}{\Lambda_0(s)} u \right].$$

Letting

$$
\begin{aligned}
z &= y, \\
\rho^* &= k_p, \\
\theta^* &= [q_{n-2}, \ldots, q_0, p_{n-1}, \ldots, p_0]^T, \\
\phi &= \left[\frac{s^{n-2}}{\Lambda_0(s)} u, \ldots, \frac{1}{\Lambda_0(s)} u, \frac{s^{n-1}}{\Lambda_0(s)} y, \ldots, \frac{1}{\Lambda_0(s)} y \right]^T , \\
z_0 &= \frac{s^{n-1}}{\Lambda_0(s)} u,
\end{aligned}
$$

we obtain the B-SPM

$$z = \rho^* (\theta^{*T} \phi + z_0). \tag{3.28}$$

We should note that in this case θ^* contains not the coefficients of the plant transfer function but the coefficients of the polynomials $P(s)$, $Q(s)$. In certain adaptive control systems such as MRAC, the coefficients of $P(s)$, $Q(s)$ are the controller parameters, and parameterizations such as (3.28) allow the direct estimation of the controller parameters by processing the plant I/O measurements.

If some of the coefficients of the plant transfer function are known, then the dimension of the vector θ^* in the SPM (3.27) can be reduced. For example, if a_1, a_0, b_m are known, then (3.27) can be rewritten as

$$z = \theta^{*T}\phi, \tag{3.29}$$

where

$$\begin{aligned} z &= \frac{s^n + a_1 s + a_0}{\Lambda(s)} y - \frac{b_m s^m}{\Lambda(s)} u, \\ \theta^* &= [b_{m-1}, \ldots, b_0, a_{n-1}, \ldots, a_2]^T, \\ \phi &= \left[\frac{s^{m-1}}{\Lambda(s)} u, \ldots, \frac{1}{\Lambda(s)} u, -\frac{s^{n-1}}{\Lambda(s)} y, \ldots, -\frac{s^2}{\Lambda(s)} y\right]^T. \end{aligned}$$

Adaptive Law Let us consider the SPM (3.27). The objective is to process the signals $z(t)$ and $\phi(t)$ in order to generate an estimate $\theta(t)$ for θ^* at each time t. This estimate may be generated as

$$\dot{\theta}(t) = H(t)\varepsilon(t),$$

where $H(t)$ is some gain vector that depends on $\phi(t)$ and $\varepsilon(t)$ is the estimation error signal that represents a measure of how far $\theta(t)$ is from θ^*. Different choices for $H(t)$ and $\varepsilon(t)$ lead to a wide class of adaptive laws with, sometimes, different convergence properties, as demonstrated in the following sections.

3.6 Gradient Algorithms Based on the Linear Model

The gradient algorithm is developed by using the gradient method to minimize some appropriate functional $J(\theta)$. Different choices for $J(\theta)$ lead to different algorithms. As in the scalar case, we start by defining the estimation model and estimation error for the SPM (3.27).

The estimate $\hat{z}$ of z is generated by the estimation model

$$\hat{z} = \theta^T \phi, \tag{3.30}$$

where $\theta(t)$ is the estimate of θ^* at time t. The estimation error is constructed as

$$\varepsilon = \frac{z - \hat{z}}{m_s^2} = \frac{z - \theta^T \phi}{m_s^2}, \tag{3.31}$$

where $m_s^2 \geq 1$ is the normalizing signal designed to bound ϕ from above. The normalizing signal often has the form $m_s^2 = 1 + n_s^2$, where $n_s \geq 0$ is referred to as the *static normalizing signal* designed to guarantee that $\frac{\phi}{m_s}$ is bounded from above. Some straightforward choices for n_s include

$$n_s^2 = \alpha \phi^T \phi, \quad \alpha > 0,$$

or

$$n_s^2 = \phi^T P \phi, \quad P = P^T > 0,$$

where α is a scalar and P is a matrix selected by the designer.

The estimation error (3.31) and the estimation model (3.30) are common to several algorithms that are generated in the following sections.

3.6.1 Gradient Algorithm with Instantaneous Cost Function

The cost function $J(\theta)$ is chosen as

$$J(\theta) = \frac{\varepsilon^2 m_s^2}{2} = \frac{(z - \theta^T\phi)^2}{2m_s^2}, \tag{3.32}$$

where m_s is the normalizing signal given by (3.31). At each time t, $J(\theta)$ is a convex function of θ and therefore has a global minimum. The gradient algorithm takes the form

$$\dot{\theta} = -\Gamma\nabla J, \tag{3.33}$$

where $\Gamma = \Gamma^T > 0$ is a design matrix referred to as the *adaptive gain*. Since $\nabla J = -\frac{(z-\theta^T\phi)\phi}{m_s^2} = -\varepsilon\phi$, we have

$$\dot{\theta} = \Gamma\varepsilon\phi. \tag{3.34}$$

The adaptive law (3.34) together with the estimation model (3.30), the estimation error (3.31), and filtered signals z, ϕ defined in (3.29) constitute the gradient parameter identification algorithm based on the instantaneous cost function whose stability properties are given by the following theorem.

Theorem 3.6.1. *The gradient algorithm* (3.34) *guarantees the following:*

(i) $\varepsilon, \varepsilon m_s, \dot{\theta} \in \mathcal{L}_2 \cap \mathcal{L}_\infty$ *and* $\theta \in \mathcal{L}_\infty$.

(ii) *If* $\frac{\phi}{m_s}$ *is PE, i.e.,* $\int_t^{t+T_0} \frac{\phi\phi^T}{m_s^2} d\tau > \alpha_0 T_0 I \ \forall t \geq 0$ *and for some* $T_0, \alpha_0 > 0$, *then* $\theta(t) \to \theta^*$ *exponentially fast. In addition,*

$$(\theta(t) - \theta^*)^T \Gamma^{-1}(\theta(t) - \theta^*) \leq (1 - \gamma_1)^n (\theta(0) - \theta^*)^T \Gamma^{-1}(\theta(0) - \theta^*),$$

where $0 \leq t \leq nT_0$, $n = 0, 1, 2, \ldots,$ *and*

$$\gamma_1 = \frac{2\alpha_0 T_0 \lambda_{\min}(\Gamma)}{2 + \beta^4 \lambda_{\max}^2(\Gamma) T_0^2}, \quad \beta = \sup_t \left| \frac{\phi}{m_s} \right|.$$

(iii) *If the plant model* (3.24) *has stable poles and no zero-pole cancellations and the input* u *is sufficiently rich of order* $n + m + 1$, *i.e., it consists of at least* $\frac{n+m+1}{2}$ *distinct frequencies, then* ϕ, $\frac{\phi}{m_s}$ *are PE. Furthermore,* $|\theta(t) - \theta^*|$, ε, εm_s, $\dot{\theta}$ *converge to zero exponentially fast.*

Proof. (i) Since θ^* is constant, $\dot{\tilde{\theta}} = \dot{\theta}$ and from (3.34) we have

$$\dot{\tilde{\theta}} = \Gamma\varepsilon\phi = -\Gamma\frac{\phi\phi^T}{m_s^2}\tilde{\theta}. \tag{3.35}$$

We choose the Lyapunov-like function

$$V(\tilde{\theta}) = \frac{\tilde{\theta}^T \Gamma^{-1} \tilde{\theta}}{2}.$$

Then along the solution of (3.35), we have

$$\dot{V} = \tilde{\theta}^T \phi \varepsilon = -\varepsilon^2 m_s^2 \leq 0, \tag{3.36}$$

where the second equality is obtained by substituting $\tilde{\theta}^T \phi = -\varepsilon m_s^2$ from (3.31). Since $V > 0$ and $\dot{V} \leq 0$, it follows that $V(t)$ has a limit, i.e.,

$$\lim_{t \to \infty} V(\tilde{\theta}(t)) = V_\infty < \infty,$$

and $V, \tilde{\theta} \in \mathcal{L}_\infty$, which, together with (3.31), imply that $\varepsilon, \varepsilon m_s \in \mathcal{L}_\infty$. In addition, it follows from (3.36) that

$$\int_0^\infty \varepsilon^2 m_s^2 d\tau \leq V(\tilde{\theta}(0)) - V_\infty,$$

from which we establish that $\varepsilon m_s \in \mathcal{L}_2$ and hence $\varepsilon \in \mathcal{L}_2$ (due to $m_s^2 = 1 + n_s^2$). Now from (3.35) we have

$$|\dot{\tilde{\theta}}| = |\dot{\theta}| \leq \|\Gamma\| |\varepsilon m_s| \frac{|\phi|}{m_s},$$

which together with $\frac{|\phi|}{m_s} \in \mathcal{L}_\infty$ and $|\varepsilon m_s| \in \mathcal{L}_2$ imply that $\dot{\theta} \in \mathcal{L}_2 \cap \mathcal{L}_\infty$, and the proof of (i) is complete.

The proof of parts (ii) and (iii) is longer and is presented in the web resource [94]. □

Comment 3.6.2 The rate of convergence of θ to θ^* can be improved if we choose the design parameters so that $1 - \gamma_1$ is as small as possible or, alternatively, $\gamma_1 \in (0, 1)$ is as close to 1 as possible. Examining the expression for γ_1, it is clear that the constants β, α_0, T_0 depend on each other and on ϕ. The only free design parameter is the adaptive gain matrix Γ. If we choose $\Gamma = \lambda I$, then the value of λ that maximizes γ_1 is

$$\lambda^* = \left(\frac{2}{2 - \beta^4 T_0^2} \right)^{1/2},$$

provided $\beta^4 T_0^2 < 2$. For $\lambda > \lambda^*$ or $\lambda < \lambda^*$, the expression for γ_1 suggests that the rate of convergence is slower. This dependence of the rate of convergence on the value of the adaptive gain Γ is often observed in simulations; i.e., very small or very large values of Γ lead to slower convergence rates. In general the convergence rate depends on the signal input and filters used in addition to Γ in a way that is not understood quantitatively.

Comment 3.6.3 Properties (i) and (ii) of Theorem 3.6.1 are independent of the boundedness of the regressor ϕ. Additional properties may be obtained if we make further assumptions about ϕ. For example, if $\phi, \dot{\phi} \in \mathcal{L}_\infty$, then we can show that $\varepsilon, \varepsilon m_s, \dot{\theta} \to 0$ as $t \to \infty$ (see Problem 6).

Comment 3.6.4 In the proof of Theorem 3.6.1(i)–(ii), we established that $\lim_{t\to\infty} V(t) = V_\infty$, where V_∞ is a constant. This implies that

$$\lim_{t\to\infty} V(\tilde{\theta}(t)) = \lim_{t\to\infty} \frac{\tilde{\theta}^T(t)\Gamma^{-1}\tilde{\theta}(t)}{2} = V_\infty.$$

We cannot conclude, however, that $\tilde{\theta} = \theta - \theta^*$ converges to a constant vector. For example, take $\Gamma = I$, $\tilde{\theta}(t) = [\sin t, \cos t]^T$. Then

$$\frac{\tilde{\theta}^T\tilde{\theta}}{2} = \frac{\sin^2 t + \cos^2 t}{2} = \frac{1}{2},$$

and $\tilde{\theta}(t) = [\sin t, \cos t]^T$ does not have a limit.

Example 3.6.5 Consider the nonlinear system

$$\dot{x} = af(x) + bg(x)u,$$

where a, b are unknown constants, $f(x)$, $g(x)$ are known continuous functions of x, and x, u are available for measurement. We want to estimate a, b online. We first obtain a parametric model in the form of an SPM by filtering each side with $\frac{1}{s+\lambda}$ for some $\lambda > 0$, i.e.,

$$\frac{s}{s+\lambda}x = a\frac{1}{s+\lambda}f(x) + b\frac{1}{s+\lambda}g(x)u.$$

Then, for

$$z = \frac{s}{s+\lambda}x, \qquad \theta^* = [a, b]^T, \qquad \phi = \frac{1}{s+\lambda}[f(x), g(x)u]^T,$$

we have

$$z = \theta^{*T}\phi.$$

The gradient algorithm (3.34),

$$\begin{aligned} \dot{\theta} &= \Gamma\varepsilon\phi, \\ \varepsilon &= \frac{z - \theta^T\phi}{m_s^2}, \quad m_s^2 = 1 + n_s^2, \quad n_s^2 = \phi^T\phi, \end{aligned}$$

can be used to generate $\theta(t) = [\hat{a}(t), \hat{b}(t)]^T$ online, where $\hat{a}(t)$, $\hat{b}(t)$ are the online estimates of a, b, respectively. While this adaptive law guarantees properties (i) and (ii) of Theorem 3.6.1 independent of ϕ, parameter convergence of $\theta(t)$ to θ^* requires $\frac{\phi}{m_s}$ to be PE. The important question that arises in this example is how to choose the plant input u so that $\frac{\phi}{m_s}$ is PE. Since the plant is nonlinear, the choice of u that makes $\frac{\phi}{m_s}$ PE depends on the form of the nonlinear functions $f(x)$, $g(x)$ and it is not easy, if possible at all, to establish conditions similar to those in the LTI case for a general class of nonlinearities. ∎

Example 3.6.6 Consider the dynamics of a hard-disk drive servo system [96] given by

$$y = \frac{k_p}{s^2}(u + d),$$

where y is the position error of the head relative to the center of the track, k_p is a known constant, and

$$d = A_1 \sin(\omega_1 t + \varphi_1) + A_2 \sin(\omega_2 t + \varphi_2)$$

is a disturbance that is due to higher-order harmonics that arise during rotation of the disk drive. In this case, ω_1, ω_2 are the known harmonics that have a dominant effect and A_i, φ_i, $i = 1, 2$, are the unknown amplitudes and phases. We want to estimate d in an effort to nullify its effect using the control input u.

Using $\sin(a + b) = \sin a \cos b + \cos a \sin b$, we can express d as

$$d = \theta_1^* \sin \omega_1 t + \theta_2^* \cos \omega_1 t + \theta_3^* \sin \omega_2 t + \theta_4^* \cos \omega_2 t,$$

where

$$\begin{aligned} \theta_1^* &= A_1 \cos \varphi_1, \qquad & \theta_2^* &= A_1 \sin \varphi_1, \\ \theta_3^* &= A_2 \cos \varphi_2, \qquad & \theta_4^* &= A_2 \sin \varphi_2 \end{aligned} \tag{3.37}$$

are the unknown parameters. We first obtain a parametric model for

$$\theta^* = [\theta_1^*, \theta_2^*, \theta_3^*, \theta_4^*]^T.$$

We have

$$s^2 y = k_p u + k_p \theta^{*T} \psi,$$

where

$$\psi(t) = [\sin \omega_1 t, \cos \omega_1 t, \sin \omega_2 t, \cos \omega_2 t]^T.$$

Filtering each side with $\frac{1}{\Lambda(s)}$, where $\Lambda(s) = (s + \lambda_1)(s + \lambda_2)$ and $\lambda_1, \lambda_2 > 0$ are design constants, we obtain the SPM

$$z = \theta^{*T} \phi,$$

where

$$\begin{aligned} z &= \frac{s^2}{\Lambda(s)} y - k_p \frac{1}{\Lambda(s)} u, \\ \phi &= k_p \frac{1}{\Lambda(s)} \psi(t) = k_p \frac{1}{\Lambda(s)} [\sin \omega_1 t, \cos \omega_1 t, \sin \omega_2 t, \cos \omega_2 t]^T. \end{aligned}$$

Therefore, the adaptive law

$$\begin{aligned} \dot{\theta} &= \Gamma \varepsilon \phi, \\ \varepsilon &= \frac{z - \theta^T \phi}{m_s^2}, \quad m_s^2 = 1 + n_s^2, \quad n_s^2 = \alpha \phi^T \phi, \end{aligned}$$

where $\Gamma = \Gamma^T > 0$ is a 4×4 constant matrix, may be used to generate $\theta(t)$, the online estimate of θ^*. In this case, $\phi \in \mathcal{L}_\infty$ and therefore we can take $\alpha = 0$, i.e., $m_s^2 = 1$. For

$\omega_1 \neq \omega_2$, we can establish that ϕ is PE and therefore $\theta(t) \to \theta^*$ exponentially fast. The online estimate of the amplitude and phase can be computed using (3.37) as follows:

$$\tan \hat{\varphi}_1(t) = \frac{\theta_2(t)}{\theta_1(t)}, \qquad \tan \hat{\varphi}_2(t) = \frac{\theta_4(t)}{\theta_3(t)},$$
$$\hat{A}_1(t) = \sqrt{\theta_1^2(t) + \theta_2^2(t)}, \qquad \hat{A}_2(t) = \sqrt{\theta_3^2(t) + \theta_4^2(t)},$$

provided of course that $\theta_1(t) \neq 0$, $\theta_3(t) \neq 0$. The estimated disturbance

$$\hat{d}(t) = \hat{A}_1(t)\sin(\omega_1 t + \hat{\varphi}_1(t)) + \hat{A}_2(t)\sin(\omega_2 t + \hat{\varphi}_2(t))$$

can then be generated and used by the controller to cancel the effect of the actual disturbance d. ∎

3.6.2 Gradient Algorithm with Integral Cost Function

The cost function $J(\theta)$ is chosen as

$$J(\theta) = \frac{1}{2}\int_0^t e^{-\beta(t-\tau)}\varepsilon^2(t,\tau)m_s^2(\tau)d\tau,$$

where $\beta > 0$ is a design constant acting as a forgetting factor and

$$\varepsilon(t,\tau) = \frac{z(\tau) - \theta^T(t)\phi(\tau)}{m_s^2(\tau)}, \quad \varepsilon(t,t) = \varepsilon, \quad \tau \leq t,$$

is the estimation error that depends on the estimate of θ at time t and on the values of the signals at $\tau \leq t$. The cost penalizes all past errors between $z(\tau)$ and $\hat{z}(\tau) = \theta^T(t)\phi(\tau)$, $\tau \leq t$, obtained by using the current estimate of θ at time t with past measurements of $z(\tau)$ and $\phi(\tau)$. The forgetting factor $e^{-\beta(t-\tau)}$ is used to put more weight on recent data by discounting the earlier ones. It is clear that $J(\theta)$ is a convex function of θ at each time t and therefore has a global minimum. Since $\theta(t)$ does not depend on τ, the gradient of J with respect to θ is easy to calculate despite the presence of the integral. Applying the gradient method, we have

$$\dot{\theta} = -\Gamma\nabla J,$$

where

$$\nabla J = -\int_0^t e^{-\beta(t-\tau)}\frac{z(\tau) - \theta^T(t)\phi(\tau)}{m_s^2(\tau)}\phi(\tau)d\tau.$$

This can be implemented as (see Problem 7)

$$\dot{\theta} = -\Gamma(R(t)\theta + Q(t)), \quad \theta(0) = \theta_0,$$
$$\dot{R} = -\beta R + \frac{\phi\phi^T}{m_s^2}, \qquad R(0) = 0,$$
$$\dot{Q} = -\beta Q - \frac{z\phi}{m_s^2}, \qquad Q(0) = 0,$$

where $R \in \mathcal{R}^{n\times n}$, $Q \in \mathcal{R}^{n\times 1}$; $\Gamma = \Gamma^T > 0$ is the adaptive gain; n is the dimension of the vector θ^*; and m_s is the normalizing signal defined in (3.31).

Theorem 3.6.7. *The gradient algorithm with integral cost function guarantees that*

(i) $\varepsilon, \varepsilon m_s, \dot{\theta} \in \mathcal{L}_2 \cap \mathcal{L}_\infty$ *and* $\theta \in \mathcal{L}_\infty$.

(ii) $\lim_{t\to\infty} |\dot{\theta}(t)| = 0$.

(iii) *If* $\frac{\phi}{m_s}$ *is PE, then* $\theta(t) \to \theta^*$ *exponentially fast. Furthermore, for* $\Gamma = \gamma I$*, the rate of convergence increases with* γ.

(iv) *If* u *is sufficiently rich of order* $n+m+1$*, i.e., it consists of at least* $\frac{n+m+1}{2}$ *distinct frequencies, and the plant is stable and has no zero-pole cancellations, then* ϕ, $\frac{\phi}{m_s}$ *are PE and* $\theta(t) \to \theta^*$ *exponentially fast.*

Proof. The proof is presented in the web resource [94]. □

Theorem 3.6.7 indicates that the rate of parameter convergence increases with increasing adaptive gain. Simulations demonstrate that the gradient algorithm based on the integral cost gives better convergence properties than the gradient algorithm based on the instantaneous cost. The gradient algorithm based on the integral cost has similarities with the least-squares (LS) algorithms to be developed in the next section.

3.7 Least-Squares Algorithms

The LS method dates back to the eighteenth century, when Gauss used it to determine the orbits of planets. The basic idea behind LS is fitting a mathematical model to a sequence of observed data by minimizing the sum of the squares of the difference between the observed and computed data. In doing so, any noise or inaccuracies in the observed data are expected to have less effect on the accuracy of the mathematical model.

The LS method has been widely used in parameter estimation both in recursive and nonrecursive forms mainly for discrete-time systems [46, 47, 77, 97, 98]. The method is simple to apply and analyze in the case where the unknown parameters appear in a linear form, such as in the linear SPM

$$z = \theta^{*T}\phi. \tag{3.38}$$

We illustrate the use and properties of LS by considering the simple scalar example

$$z = \theta^*\phi + d_n,$$

where $z, \theta^*, \phi \in \mathcal{R}$, $\phi \in \mathcal{L}_\infty$, and d_n is a noise disturbance whose average value goes to zero as $t \to \infty$, i.e.,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t d_n(\tau)d\tau = 0.$$

In practice, d_n may be due to sensor noise or external sources, etc. We examine the following estimation problem: Given the measurements of $z(\tau)$, $\phi(\tau)$ for $0 \le \tau < t$, find a "good" estimate $\theta(t)$ of θ^* at time t. One possible solution is to calculate $\theta(t)$ as

$$\theta(t) = \frac{z(\tau)}{\phi(\tau)} = \theta^* + \frac{d_n(\tau)}{\phi(\tau)} \tag{3.39}$$

by using the measurements of $z(\tau)$, $\phi(\tau)$ at some $\tau < t$ for which $\phi(\tau) \neq 0$. Because of the noise disturbance, however, such an estimate may be far off from θ^*. For example, at the particular time τ at which we measured z and ϕ, the effect of $d_n(\tau)$ may be significant, leading to an erroneous estimate for $\theta(t)$ generated by (3.39).

A more intelligent approach is to choose the estimate $\theta(t)$ at time t to be the one that minimizes the square of all the errors that result from the mismatch of $z(\tau) - \theta(t)\phi(\tau)$ for $0 \le \tau \le t$. Hence the estimation problem above becomes the following LS problem: Minimize the cost

$$J(\theta) = \frac{1}{2}\int_0^t |z(\tau) - \theta(t)\phi(\tau)|^2 d\tau \tag{3.40}$$

w.r.t. $\theta(t)$ at any given time t. The cost $J(\theta)$ penalizes all the past errors from $\tau = 0$ to t that are due to $\theta(t) \neq \theta^*$. Since $J(\theta)$ is a convex function over $\mathcal{R}$ at each time t, its minimum satisfies

$$\nabla J(\theta) = -\int_0^t z(\tau)\phi(\tau)d\tau + \theta(t)\int_0^t \phi^2(\tau)d\tau = 0,$$

which gives the LS estimate

$$\theta(t) = \left(\int_0^t \phi^2(\tau)d\tau\right)^{-1}\int_0^t z(\tau)\phi(\tau)d\tau,$$

provided of course that the inverse exists. The LS method considers all past data in an effort to provide a good estimate for θ^* in the presence of noise d_n. For example, when $\phi(t) = 1$, $\forall t \ge 0$, we have

$$\lim_{t\to\infty}\theta(t) = \lim_{t\to\infty}\frac{1}{t}\int_0^t z(\tau)d\tau = \theta^* + \lim_{t\to\infty}\frac{1}{t}\int_0^t d_n(\tau)d\tau = \theta^*;$$

i.e., $\theta(t)$ converges to the exact parameter value despite the presence of the noise disturbance d_n.

Let us now extend this problem to the linear model (3.38). As in section 3.6, the estimate $\hat{z}$ of z and the normalized estimation are generated as

$$\hat{z} = \theta^T\phi, \qquad e = \frac{z - \hat{z}}{m_s^2} = \frac{z - \theta^T\phi}{m_s^2},$$

where $\theta(t)$ is the estimate of θ^* at time t, and $m_s^2 = 1 + n_s^2$ is designed to guarantee $\frac{\phi}{m_s} \in \mathcal{L}_\infty$. Below we present different versions of the LS algorithm, which correspond to different choices of the LS cost $J(\theta)$.

3.7.1 Recursive LS Algorithm with Forgetting Factor

Consider the function

$$J(\theta) = \frac{1}{2}\int_0^t e^{-\beta(t-\tau)}\frac{[z(\tau)-\theta^T(t)\phi(\tau)]^2}{m_s^2(\tau)}d\tau + \frac{1}{2}e^{-\beta t}(\theta-\theta_0)^T Q_0(\theta-\theta_0), \qquad (3.41)$$

where $Q_0 = Q_0^T > 0$, $\beta \geq 0$ are design constants and $\theta_0 = \theta(0)$ is the initial parameter estimate. This cost function is a generalization of (3.40) to include possible discounting of past data and a penalty on the initial error between the estimate θ_0 and θ^*. Since $\frac{z}{m_s}$, $\frac{\phi}{m_s} \in \mathcal{L}_\infty$, $J(\theta)$ is a convex function of θ over $\mathcal{R}^n$ at each time t. Hence, any local minimum is also global and satisfies

$$\nabla J(\theta(t)) = 0 \quad \forall t \geq 0.$$

The LS algorithm for generating $\theta(t)$, the estimate of θ^*, in (3.38) is therefore obtained by solving

$$\nabla J(\theta) = e^{-\beta t}Q_0(\theta(t)-\theta_0) - \int_0^t e^{-\beta(t-\tau)}\frac{z(\tau)-\theta^T(t)\phi(\tau)}{m_s^2(\tau)}\phi(\tau)d\tau = 0 \qquad (3.42)$$

for $\theta(t)$, which yields the *nonrecursive LS algorithm*

$$\theta(t) = P(t)\left[e^{-\beta t}Q_0\theta_0 + \int_0^t e^{-\beta(t-\tau)}\frac{z(\tau)\phi(\tau)}{m_s^2(\tau)}d\tau\right], \qquad (3.43)$$

where

$$P(t) = \left[e^{-\beta t}Q_0 + \int_0^t e^{-\beta(t-\tau)}\frac{\phi(\tau)\phi^T(\tau)}{m_s^2(\tau)}d\tau\right]^{-1} \qquad (3.44)$$

is the so-called *covariance matrix*. Because $Q_0 = Q_0^T > 0$ and $\phi\phi^T$ is positive semidefinite, $P(t)$ exists at each time t. Using the identity

$$\frac{d}{dt}PP^{-1} = \dot{P}P^{-1} + P\frac{d}{dt}P^{-1} = 0$$

and $\varepsilon m_s^2 = z - \theta^T\phi$, and differentiating $\theta(t)$ w.r.t. t, we obtain the *recursive LS algorithm with forgetting factor*

$$\begin{aligned} \dot{\theta} &= P\varepsilon\phi, & \theta(0) &= \theta_0, \\ \dot{P} &= \beta P - P\frac{\phi\phi^T}{m_s^2}P, & P(0) &= P_0 = Q_0^{-1}. \end{aligned} \qquad (3.45)$$

The stability properties of (3.45) depend on the value of the *forgetting factor* β, as discussed in the following sections. If $\beta = 0$, the algorithm becomes the *pure LS* algorithm discussed and analyzed in section 3.7.2. When $\beta > 0$, stability cannot be established unless $\frac{\phi}{m_s}$ is PE. In this case (3.45) is modified, leading to a different algorithm discussed and analyzed in section 3.7.3.

The following theorem establishes the stability and convergence of θ to θ^* of the algorithm (3.45) in the case where $\frac{\phi}{m_s}$ is PE.

Theorem 3.7.1. *If $\frac{\phi}{m_s}$ is PE, then the recursive LS algorithm with forgetting factor* (3.45) *guarantees that $P, P^{-1} \in \mathcal{L}_\infty$ and that $\theta(t) \to \theta^*$ as $t \to \infty$. The convergence of $\theta(t) \to \theta^*$ is exponential when $\beta > 0$.*

Proof. The proof is given in [56] and in the web resource [94]. □

Since the adaptive law (3.45) could be used in adaptive control where the PE property of $\frac{\phi}{m_s}$ cannot be guaranteed, it is of interest to examine the properties of (3.45) in the absence of PE. In this case, (3.45) is modified in order to avoid certain undesirable phenomena, as discussed in the following sections.

3.7.2 Pure LS Algorithm

When $\beta = 0$ in (3.41), the algorithm (3.45) reduces to

$$\begin{aligned} \dot{\theta} &= P\varepsilon\phi, & \theta(0) &= \theta_0, \\ \dot{P} &= -P\frac{\phi\phi^T}{m_s^2}P, & P(0) &= P_0, \end{aligned} \tag{3.46}$$

which is referred to as the *pure LS algorithm*.

Theorem 3.7.2. *The pure LS algorithm* (3.46) *guarantees that*

(i) $\varepsilon, \varepsilon m_s, \dot{\theta} \in \mathcal{L}_2 \cap \mathcal{L}_\infty$ *and* $\theta, P \in \mathcal{L}_\infty$.

(ii) $\lim_{t\to\infty} \theta(t) = \bar{\theta}$, *where $\bar{\theta}$ is a constant vector.*

(iii) *If $\frac{\phi}{m_s}$ is PE, then $\theta(t) \to \theta^*$ as $t \to \infty$.*

(iv) *If* (3.38) *is the SPM for the plant* (3.24) *with stable poles and no zero-pole cancellations, and u is sufficiently rich of order $n+m+1$, i.e., consists of at least $\frac{n+m+1}{2}$ distinct frequencies, then ϕ, $\frac{\phi}{m_s}$ are PE and therefore $\theta(t) \to \theta^*$ as $t \to \infty$.*

Proof. From (3.46) we have that $\dot{P} \le 0$, i.e., $P(t) \le P_0$. Because $P(t)$ is nonincreasing and bounded from below (i.e., $P(t) = P^T(t) \ge 0\ \forall t \ge 0$) it has a limit, i.e.,

$$\lim_{t\to\infty} P(t) = \bar{P},$$

where $\bar{P} = \bar{P}^T \ge 0$ is a constant matrix. Let us now consider the identity

$$\frac{d}{dt}(P^{-1}\tilde{\theta}) = -P^{-1}\dot{P}P^{-1}\tilde{\theta} + P^{-1}\dot{\tilde{\theta}} = \frac{\phi\phi^T\tilde{\theta}}{m_s^2} + \varepsilon\phi = 0,$$

where the last two equalities are obtained using $\dot{\theta} = \dot{\tilde{\theta}}$, $\frac{d}{dt}P^{-1} = -P^{-1}\dot{P}P^{-1}$, and $\varepsilon = -\frac{\tilde{\theta}^T\phi}{m_s^2} = -\frac{\phi^T\tilde{\theta}}{m_s^2}$. Hence, $P^{-1}(t)\tilde{\theta}(t) = P_0^{-1}\tilde{\theta}(0)$ and therefore $\tilde{\theta}(t) = P(t)P_0^{-1}\tilde{\theta}(0)$ and $\lim_{t\to\infty}\tilde{\theta}(t) = \bar{P}P_0^{-1}\tilde{\theta}(0)$, which implies that $\lim_{t\to\infty}\theta(t) = \theta^* + \bar{P}P_0^{-1}\tilde{\theta}(0) \triangleq \bar{\theta}$.

Because $P(t) \leq P_0$ and $\tilde{\theta}(t) = P(t)P_0^{-1}\tilde{\theta}(0)$ we have $\theta, \tilde{\theta} \in \mathcal{L}_\infty$, which, together with $\frac{\phi}{m_s} \in \mathcal{L}_\infty$, implies that $\varepsilon m_s = -\frac{\tilde{\theta}^T\phi}{m_s}$ and $\varepsilon, \varepsilon m_s \in \mathcal{L}_\infty$. Let us now consider the function

$$V(\tilde{\theta}, t) = \frac{\tilde{\theta}^T P^{-1}(t)\tilde{\theta}}{2}.$$

The time derivative $\dot{V}$ of V along the solution of (3.46) is given by

$$\dot{V} = \varepsilon\tilde{\theta}^T\phi + \frac{\tilde{\theta}^T\phi\phi^T\tilde{\theta}}{2m_s^2} = -\varepsilon^2 m_s^2 + \frac{\varepsilon^2 m_s^2}{2} = -\frac{\varepsilon^2 m_s^2}{2} \leq 0,$$

which implies that $V \in \mathcal{L}_\infty$, $\varepsilon m_s \in \mathcal{L}_2$, and therefore $\varepsilon \in \mathcal{L}_2$. From (3.46) we have

$$|\dot{\theta}| \leq \|P\|\frac{|\phi|}{m_s}|\varepsilon m_s|.$$

Since $P, \frac{\phi}{m_s}, \varepsilon m_s \in \mathcal{L}_\infty$, and $\varepsilon m_s \in \mathcal{L}_2$, we have $\dot{\theta} \in \mathcal{L}_\infty \cap \mathcal{L}_2$, which completes the proof for (i) and (ii). The proofs of (iii) and (iv) are included in the proofs of Theorems 3.7.1 and 3.6.1, respectively. □

The pure LS algorithm guarantees that the parameters converge to some constant $\bar{\theta}$ without having to put any restriction on the regressor ϕ. If $\frac{\phi}{m_s}$, however, is PE, then $\bar{\theta} = \theta^*$. Convergence of the estimated parameters to constant values is a unique property of the pure LS algorithm. One of the drawbacks of the pure LS algorithm is that the covariance matrix P may become arbitrarily small and slow down adaptation in some directions. This is due to the fact that

$$\frac{d(P^{-1})}{dt} = \frac{\phi\phi^T}{m_s^2} \geq 0,$$

which implies that P^{-1} may grow without bound, which in turn implies that P may reduce towards zero. This is the so-called *covariance wind-up* problem. Another drawback of the pure LS algorithm is that parameter convergence cannot be guaranteed to be exponential.

Example 3.7.3 In order to get some understanding of the properties of the pure LS algorithm, let us consider the scalar SPM

$$z = \theta^*\phi,$$

where $z, \theta^*, \phi \in \mathcal{R}$. Let us assume that $\phi \in \mathcal{L}_\infty$. Then the pure LS algorithm is given by

$$\begin{aligned} \dot{\theta} &= p\varepsilon\phi, & \theta(0) &= \theta_0, \\ \dot{p} &= -p^2\phi^2, & p(0) &= p_0 > 0, \\ \varepsilon &= z - \theta\phi = -\tilde{\theta}\phi. \end{aligned}$$

Let us also take $\phi = 1$, which is PE, for this example. Then we can show by solving the differential equation via integration that

$$p(t) = \frac{p_0}{1 + p_0 t}$$

and $\tilde{\theta}(t) = \frac{\tilde{\theta}(0)}{1+p_0 t}$, i.e.,

$$\theta(t) = \theta^* + \frac{\theta(0) - \theta^*}{1 + p_0 t}.$$

It is clear that as $t \to \infty$, $p(t) \to 0$, leading to the so-called covariance wind-up problem. Since $\phi = 1$ is PE, however, $\theta(t) \to \theta^*$ as $t \to \infty$ with a rate of $\frac{1}{t}$ (not exponential) as predicted by Theorem 3.7.2. Even though $\theta(t) \to \theta^*$, the covariance wind-up problem may still pose a problem in the case where θ^* changes to some other value after some time. If at that instance $p(t) \cong 0$, leading to $\dot{\theta} \cong 0$, no adaptation will take place and $\theta(t)$ may not reach the new θ^*.

For the same example, consider $\phi(t) = \frac{1}{1+t}$, which is not PE since

$$\int_t^{t+T} \phi^2(\tau) d\tau = \int_t^{t+T} \frac{1}{(1+\tau)^2} d\tau = \frac{1}{1+t} - \frac{1}{1+t+T}$$

goes to zero as $t \to \infty$, i.e., it has zero level of excitation. In this case, we can show that

$$\begin{aligned} p(t) &= \frac{p_0(1+t)}{1+(1+p_0)t}, \\ \theta(t) &= \theta^* + (\theta(0) - \theta^*)\frac{1+t}{1+(1+p_0)t} \end{aligned}$$

by solving the differential equations above. It is clear that $p(t) \to \frac{p_0}{1+p_0}$ and $\theta(t) \to \frac{p_0\theta^* + \theta(0)}{1+p_0}$ as $t \to \infty$; i.e., $\theta(t)$ converges to a constant but not to θ^* due to lack of PE. In this case $p(t)$ converges to a constant and no covariance wind-up problem arises. ∎

3.7.3 Modified LS Algorithms

One way to avoid the covariance wind-up problem is to modify the pure LS algorithm using a covariance resetting modification to obtain

$$\begin{aligned} \dot{\theta} &= P\varepsilon\phi, & \theta(0) &= \theta_0, \\ \dot{P} &= -P\frac{\phi\phi^T}{m_s^2}P, & P(t_r^+) &= P_0 = \rho_0 I, \\ m_s^2 &= 1 + n_s^2, & n_s^2 &= \alpha\phi^T\phi, \quad \alpha > 0, \end{aligned} \tag{3.47}$$

where t_r^+ is the time at which $\lambda_{\min}(P(t)) \le \rho_1$ and $\rho_0 > \rho_1 > 0$ are some design scalars. Due to covariance resetting, $P(t) \ge \rho_1 I \ \forall t \ge 0$. Therefore, P is guaranteed to be positive definite for all $t \ge 0$. In fact, the *pure LS algorithm with covariance resetting* can be viewed as a gradient algorithm with time-varying adaptive gain P, and its properties are very similar to those of a gradient algorithm analyzed in the previous section. They are summarized by Theorem 3.7.4 in this section.

When $\beta > 0$, the covariance wind-up problem, i.e., $P(t)$ becoming arbitrarily small, does not exist. In this case, $P(t)$ may grow without bound. In order to avoid this phe-

nomenon, the following *modified LS algorithm with forgetting factor* is used:

$$\dot{\theta} = P\varepsilon\phi,$$
$$\dot{P} = \begin{cases} \beta P - \frac{P\phi\phi^T P}{m_s^2} & \text{if } \|P(t)\| \le R_0, \\ 0 & \text{otherwise,} \end{cases} \tag{3.48}$$

where $P(0) = P_0 = P_0^T > 0$, $\|P_0\| \le R_0$, R_0 is a constant that serves as an upper bound for $\|P\|$, and $m_s^2 = 1 + n_s^2$ is the normalizing signal which satisfies $\frac{\phi}{m_s} \in \mathcal{L}_\infty$.

The following theorem summarizes the stability properties of the two modified LS algorithms.

Theorem 3.7.4. *The pure LS algorithm with covariance resetting* (3.47) *and the modified LS algorithm with forgetting factor* (3.48) *guarantee that*

(i) $\varepsilon, \varepsilon m_s, \dot{\theta} \in \mathcal{L}_2 \cap \mathcal{L}_\infty$ *and* $\theta \in \mathcal{L}_\infty$.

(ii) *If* $\frac{\phi}{m_s}$ *is PE, then* $\theta(t) \to \theta^*$ *as* $t \to \infty$ *exponentially fast.*

(iii) *If* (3.38) *is the SPM for the plant* (3.24) *with stable poles and no zero-pole cancellations, and* u *is sufficiently rich of order* $n+m+1$, *then* ϕ, $\frac{\phi}{m_s}$ *are PE, which guarantees that* $\theta(t) \to \theta^*$ *as* $t \to \infty$ *exponentially fast.*

Proof. The proof is presented in the web resource [94]. □

3.8 Parameter Identification Based on DPM

Let us consider the DPM

$$z = W(s)[\theta^{*T}\psi].$$

This model may be obtained from (3.27) by filtering each side with $W(s)$ and redefining the signals z, ϕ. Since θ^* is a constant vector, the DPM may be written as

$$z = W(s)L(s)[\theta^{*T}\phi], \tag{3.49}$$

where $\phi = L^{-1}(s)\psi$, $L(s)$ is chosen so that $L^{-1}(s)$ is a proper stable transfer function, and $W(s)L(s)$ is a proper strictly positive real (SPR) transfer function.

$$\hat{z} = W(s)L(s)[\theta^T\phi].$$

We form the normalized estimation error

$$\varepsilon = z - \hat{z} - W(s)L(s)[\varepsilon n_s^2], \tag{3.50}$$

where the static normalizing signal n_s is designed so that $\frac{\phi}{m_s} \in \mathcal{L}_\infty$ for $m_s^2 = 1 + n_s^2$. If $W(s)L(s) = 1$, then (3.50) has the same expression as in the case of the gradient algorithm. Substituting for z in (3.50), we express ε in terms of the parameter error $\tilde{\theta} = \theta - \theta^*$:

$$\varepsilon = W(s)L(s)[-\tilde{\theta}^T\phi - \varepsilon n_s^2]. \tag{3.51}$$

For simplicity, let us assume that $W(s)L(s)$ is strictly proper and rewrite (3.51) in the minimum state-space representation form

$$\begin{aligned} \dot{e} &= A_c e + b_c(-\tilde{\theta}^T \phi - \varepsilon n_s^2), \\ \varepsilon &= c_c^T e, \end{aligned} \tag{3.52}$$

where $W(s)L(s) = c_c^T(sI - A_c)^{-1}b_c$. Since $W(s)L(s)$ is SPR, it follows that (see the Appendix) there exist matrices $P_c = P_c^T > 0$, $L_c = L_c^T > 0$, a vector q, and a scalar $\nu > 0$ such that

$$\begin{aligned} P_c A_c + A_c^T P_c &= -qq^T - \nu L_c, \\ P_c b_c &= c_c. \end{aligned} \tag{3.53}$$

The adaptive law for θ is generated using the Lyapunov-like function

$$V = \frac{e^T P_c e}{2} + \frac{\tilde{\theta}^T \Gamma^{-1} \tilde{\theta}}{2},$$

where $\Gamma = \Gamma^T > 0$. The time derivative $\dot{V}$ of V along the solution of (3.52) is given by

$$\dot{V} = -\frac{1}{2} e^T q q^T e - \frac{\nu}{2} e^T L_c e + e^T P_c b_c(-\tilde{\theta}^T \phi - \varepsilon n_s^2) + \tilde{\theta}^T \Gamma^{-1} \dot{\tilde{\theta}}.$$

Since $e^T P_c b_c = e^T c_c = \varepsilon$, it follows that by choosing $\dot{\tilde{\theta}} = \dot{\theta}$ as

$$\dot{\theta} = \Gamma \varepsilon \phi, \tag{3.54}$$

we get

$$\dot{V} = -\frac{1}{2} e^T q q^T e - \frac{\nu}{2} e^T L_c e - \varepsilon^2 n_s^2 \leq 0.$$

As before, from the properties of V, $\dot{V}$ we conclude that $e, \varepsilon, \theta \in \mathcal{L}_\infty$ and $e, \varepsilon, \varepsilon n_s \in \mathcal{L}_2$. These properties in turn imply that $\dot{\theta} \in \mathcal{L}_2$. Note that without the use of the second equation in (3.53), we are not able to choose $\dot{\tilde{\theta}} = \dot{\theta}$ using signals available for measurement to make $\dot{V} \leq 0$. This is because the state e in (3.52) cannot be generated since it depends on the unknown input $\tilde{\theta}^T \phi$. Equation (3.52) is used only for analysis.

The stability properties of the adaptive law (3.54) are summarized by the following theorem.

Theorem 3.8.1. *The adaptive law* (3.54) *guarantees that*

(i) $\varepsilon, \theta \in \mathcal{L}_\infty$ *and* $\varepsilon, \varepsilon n_s, \dot{\theta} \in \mathcal{L}_2$.

(ii) *If* $n_s, \phi, \dot{\phi} \in \mathcal{L}_\infty$ *and* ϕ *is PE, then* $\theta(t) \to \theta^*$ *exponentially fast.*

Proof. The proof for (i) is given above. The proof of (ii) is a long one and is given in [56] as well as in the web resource [94]. □

The adaptive law (3.54) is referred to as the adaptive law based on the SPR-Lyapunov synthesis approach.

Comment 3.8.2 The adaptive law (3.54) has the same form as the gradient algorithm even though it is developed using a Lyapunov approach and the SPR property. In fact, for $W(s)L(s) = 1$, (3.54) is identical to the gradient algorithm.

3.9 Parameter Identification Based on B-SPM

Consider the bilinear SPM described by (3.28), i.e.,

$$z = \rho^*(\theta^{*T}\phi + z_0), \tag{3.55}$$

where z, z_0 are known scalar signals at each time t and ρ^*, θ^* are the scalar and vector unknown parameters, respectively. The estimation error is generated as

$$\hat{z} = \rho(\theta^T\phi + z_0),$$
$$\varepsilon = \frac{z - \hat{z}}{m_s^2},$$

where $\rho(t), \theta(t)$ are the estimates of ρ^*, θ^*, respectively, at time t and where m_s is designed to bound ϕ, z_0 from above. An example of m_s with this property is $m_s^2 = 1 + \phi^T\phi + z_0^2$.

Let us consider the cost

$$J(\rho, \theta) = \frac{\varepsilon^2 m_s^2}{2} = \frac{(z - \rho^*\theta^T\phi - \rho\xi + \rho^*\xi - \rho^* z_0)^2}{2m_s^2},$$

where

$$\xi = \theta^T\phi + z_0$$

is available for measurement. Applying the gradient method we obtain

$$\dot{\theta} = -\Gamma_1 \nabla J_\theta = \Gamma_1 \varepsilon \rho^* \phi,$$
$$\dot{\rho} = -\gamma \nabla J_\rho = \gamma \varepsilon \xi,$$

where $\Gamma_1 = \Gamma_1^T > 0$, $\gamma > 0$ are the adaptive gains. Since ρ^* is unknown, the adaptive law for θ cannot be implemented. We bypass this problem by employing the equality

$$\Gamma_1 \rho^* = \Gamma_1 |\rho^*| \operatorname{sgn}(\rho^*) = \Gamma \operatorname{sgn}(\rho^*),$$

where $\Gamma = \Gamma_1 |\rho^*|$. Since Γ_1 is arbitrary any $\Gamma = \Gamma^T > 0$ can be selected without having to know $|\rho^*|$. Therefore, the adaptive laws for θ, ρ, may be written as

$$\begin{aligned} \dot{\theta} &= \Gamma \varepsilon \phi \operatorname{sgn}(\rho^*), \\ \dot{\rho} &= \gamma \varepsilon \xi, \\ \varepsilon &= \frac{z - \rho\xi}{m_s^2}, \quad \xi = \theta^T\phi + z_0. \end{aligned} \tag{3.56}$$

Theorem 3.9.1. *The adaptive law* (3.56) *guarantees that*

(i) $\varepsilon, \varepsilon m_s, \dot{\theta}, \dot{\rho} \in \mathcal{L}_2 \cap \mathcal{L}_\infty$ *and* $\theta, \rho \in \mathcal{L}_\infty$.

(ii) *If* $\frac{\xi}{m_s} \in \mathcal{L}_2$, *then* $\rho(t) \to \bar{\rho}$ *as* $t \to \infty$, *where* $\bar{\rho}$ *is a constant.*

(iii) *If* $\frac{\xi}{m_s} \in \mathcal{L}_2$ *and* $\frac{\phi}{m_s}$ *is PE, then* $\theta(t)$ *converges to* θ^* *as* $t \to \infty$.

(iv) *If the plant* (3.24) *has stable poles with no zero-pole cancellations and* u *is sufficiently rich of order* $n+m+1$, *then* ϕ, $\frac{\phi}{m_s}$ *are PE and* $\theta(t)$ *converges to* θ^* *as* $t \to \infty$.

Proof. Consider the Lyapunov-like function

$$V = \frac{\tilde{\theta}^T \Gamma^{-1} \tilde{\theta}}{2} |\rho^*| + \frac{\tilde{\rho}^2}{2\gamma}.$$

Then

$$\dot{V} = \tilde{\theta}^T \phi \varepsilon |\rho^*| \operatorname{sgn}(\rho^*) + \tilde{\rho} \varepsilon \xi.$$

Using $|\rho^*| \operatorname{sgn}(\rho^*) = \rho^*$ and the expression

$$\begin{aligned} \varepsilon m_s^2 &= \rho^* \theta^{*T} \phi + \rho^* z_0 - \rho \theta^T \phi - \rho z_0 \\ &= -\tilde{\rho} z_0 + \rho^* \theta^{*T} \phi - \rho \theta^T \phi + \rho^* \theta^T \phi - \rho^* \theta^T \phi \\ &= -\tilde{\rho}(z_0 + \theta^T \phi) - \rho^* \tilde{\theta}^T \phi = -\tilde{\rho} \xi - \rho^* \tilde{\theta}^T \phi, \end{aligned}$$

we have

$$\dot{V} = \varepsilon(\rho^* \tilde{\theta}^T \phi + \tilde{\rho} \xi) = -\varepsilon^2 m_s^2 \le 0,$$

which implies that $V \in \mathcal{L}_\infty$ and therefore $\rho, \theta \in \mathcal{L}_\infty$. Using similar analysis as in the case of the gradient algorithms for the SPM, we can establish (i) from the properties of V, $\dot{V}$ and the form of the adaptive laws.

(ii) We have

$$\begin{aligned} \rho(t) - \rho(0) = \int_0^t \dot{\rho} d\tau \le \int_0^t |\dot{\rho}| d\tau &\le \gamma \int_0^t |\varepsilon m_s| \frac{|\xi|}{m_s} d\tau \\ &\le \gamma \left(\int_0^t \varepsilon^2 m_s^2 d\tau \right)^{1/2} \left(\int_0^t \frac{|\xi|^2}{m_s^2} d\tau \right)^{1/2} < \infty, \end{aligned}$$

where the last inequality is obtained using the Schwarz inequality (see (A.14)). Since εm_s, $\frac{\xi}{m_s} \in \mathcal{L}_2$, the limit as $t \to \infty$ exists, which implies that $\dot{\rho} \in \mathcal{L}_1$ and $\lim_{t\to\infty} \rho(t) = \bar{\rho}$ for some constant $\bar{\rho}$. The proof of (iii) is long and is presented in the web resource [94]. The proof of (iv) is included in the proof of Theorem 3.6.1. □

The assumption that the sign of ρ^* is known can be relaxed, leading to an adaptive law for θ, ρ with additional nonlinear terms. The reader is referred to [56, 99–104] for further reading on adaptive laws with unknown high-frequency gain.

3.10 Parameter Projection

In many practical problems, we may have some a priori knowledge of where θ^* is located in $\mathcal{R}^n$. This knowledge usually comes in terms of upper and/or lower bounds for the elements of θ^* or in terms of location in a convex subset of $\mathcal{R}^n$. If such a priori information is available, we want to constrain the online estimation to be within the set where the unknown parameters are located. For this purpose we modify the gradient algorithms based on the unconstrained minimization of certain costs using the gradient projection method presented in section A.10.3 as follows.

The gradient algorithm with projection is computed by applying the gradient method to the following minimization problem with constraints:

$$\begin{aligned} &\text{minimize } J(\theta) \\ &\text{subject to } \theta \in S, \end{aligned}$$

where S is a convex subset of $\mathcal{R}^n$ with smooth boundary almost everywhere. Assume that S is given by

$$S = \{\theta \in \mathcal{R}^n | g(\theta) \leq 0\},$$

where $g : \mathcal{R}^n \to \mathcal{R}$ is a smooth function.

The adaptive laws based on the gradient method can be modified to guarantee that $\theta \in S$ by solving the constrained optimization problem given above to obtain

$$\dot{\theta} = \Pr(-\Gamma\nabla J) = \begin{cases} -\Gamma\nabla J & \text{if } \theta \in S^0 \\ & \text{or } \theta \in \delta(S) \text{ and } -(\Gamma\nabla J)^T\nabla g \leq 0, \\ -\Gamma\nabla J + \Gamma\frac{\nabla g\nabla g^T}{\nabla g^T\Gamma\nabla g}\Gamma\nabla J & \text{otherwise,} \end{cases} \tag{3.57}$$

where $\delta(S) = \{\theta \in \mathcal{R}^n | g(\theta) = 0\}$ and $S^0 = \{\theta \in \mathcal{R}^n | g(\theta) < 0\}$ denote the boundary and the interior, respectively, of S and $\Pr(\cdot)$ is the projection operator as shown in section A.10.3.

The *gradient algorithm based on the instantaneous cost function with projection* follows from (3.57) by substituting for $\nabla J = -\varepsilon\phi$ to obtain

$$\dot{\theta} = \Pr(\Gamma\varepsilon\phi) = \begin{cases} \Gamma\varepsilon\phi & \text{if } \theta \in S^0 \\ & \text{or } \theta \in \delta(S) \text{ and } (\Gamma\varepsilon\phi)^T\nabla g \leq 0, \\ \Gamma\varepsilon\phi - \Gamma\frac{\nabla g\nabla g^T}{\nabla g^T\Gamma\nabla g}\Gamma\varepsilon\phi & \text{otherwise,} \end{cases} \tag{3.58}$$

where $\theta(0) \in S$.

The *pure LS algorithm with projection* becomes

$$\dot{\theta} = \Pr(P\varepsilon\phi) = \begin{cases} P\varepsilon\phi & \text{if } \theta \in S^0 \\ & \text{or if } \theta \in \delta(S) \text{ and } (P\varepsilon\phi)^T\nabla g \leq 0, \\ P\varepsilon\phi - P\frac{\nabla g\nabla g^T}{\nabla g^T P\nabla g}P\varepsilon\phi & \text{otherwise,} \end{cases} \tag{3.59}$$

where $\theta(0) \in S$,

$$\dot{P} = \begin{cases} \beta P - P\frac{\phi\phi^T}{m_s^2}P & \text{if } \theta \in S^0 \\ & \text{or if } \theta \in \delta(S) \text{ and } (P\varepsilon\phi)^T\nabla g \leq 0, \\ 0 & \text{otherwise,} \end{cases}$$

and $P(0) = P_0 = P_0^T > 0$.

Theorem 3.10.1. *The gradient adaptive laws of section* 3.6 *and the LS adaptive laws of section* 3.7 *with the projection modifications given by* (3.57) *and* (3.59)*, respectively, retain all the properties that are established in the absence of projection and in addition guarantee that* $\theta(t) \in S\ \forall t \geq 0$*, provided* $\theta(0) \in S$ *and* $\theta^* \in S$.

Proof. The adaptive laws (3.57) and (3.59) both guarantee that whenever $\theta \in \delta(S)$, the direction of $\dot{\theta}$ is either towards S^0 or along the tangent plane of $\delta(S)$ at θ. This property together with $\theta(0) \in S$ guarantees that $\theta(t) \in S\ \forall t \geq 0$.

The gradient adaptive law (3.57) can be expressed as

$$\dot{\theta} = -\Gamma\nabla J + (1 - \operatorname{sgn}(|g(\theta)|)) \max(0, \operatorname{sgn}(-(\Gamma\nabla J)^T\nabla g))\Gamma\frac{\nabla g\nabla g^T}{\nabla g^T\Gamma\nabla g}\Gamma\nabla J, \quad (3.60)$$

where $\operatorname{sgn}|x| = 0$ when $x = 0$. Hence, for the Lyapunov-like function V used in section 3.6, i.e., $V = \frac{\tilde{\theta}^T\Gamma^{-1}\tilde{\theta}}{2}$, we have

$$\dot{V} = -\tilde{\theta}^T\nabla J + V_p,$$

where

$$V_p = (1 - \operatorname{sgn}(|g(\theta)|)) \max(0, \operatorname{sgn}(-(\Gamma\nabla J)^T\nabla g))\tilde{\theta}^T\frac{\nabla g\nabla g^T}{\nabla g^T\Gamma\nabla g}\Gamma\nabla J.$$

The term V_p is nonzero only when $\theta \in \delta(S)$, i.e., $g(\theta) = 0$ and $-(\Gamma\nabla J)^T\nabla g > 0$. In this case

$$\begin{aligned} V_p &= \tilde{\theta}^T\frac{\nabla g\nabla g^T}{\nabla g^T\Gamma\nabla g}\Gamma\nabla J = \frac{1}{\nabla g^T\Gamma\nabla g}(\tilde{\theta}^T\nabla g)(\nabla g^T\Gamma\nabla J) \\ &= \frac{1}{\nabla g^T\Gamma\nabla g}(\tilde{\theta}^T\nabla g)((\Gamma\nabla J)^T\nabla g). \end{aligned}$$

Since $\tilde{\theta}^T\nabla g = (\theta - \theta^*)^T\nabla g \geq 0$ for $\theta^* \in S$ and $\theta \in \delta(S)$ due to the convexity of S, $(\Gamma\nabla J)^T\nabla g < 0$ implies that $V_p \leq 0$. Therefore, the projection due to $V_p \leq 0$ can only make $\dot{V}$ more negative. Hence

$$\dot{V} \leq -\tilde{\theta}^T\nabla J, \quad (3.61)$$

which is the same expression as in the case without projection except for the inequality. Hence the results in section 3.6 based on the properties of V and $\dot{V}$ are valid for the adaptive law (3.57) as well. Moreover, since $\Gamma\frac{\nabla g\nabla g^T}{\nabla g^T\Gamma\nabla g} \in \mathcal{L}_\infty$, from (3.60) we have

$$|\dot{\theta}|^2 \leq c|\Gamma\nabla J|^2$$

for some constant $c > 0$, which can be used together with (3.61) to show that $\dot{\theta} \in \mathcal{L}_2$. The same arguments apply to the case of the LS algorithms. □

Example 3.10.2 Let us consider the plant model

$$y = \frac{b}{s+a}u,$$

where a, b are unknown constants that satisfy some known bounds, e.g., $b \geq 1$ and $20 \geq a \geq -2$. For simplicity, let us assume that y, $\dot{y}$, u are available for measurement so that the SPM is of the form

$$z = \theta^{*T}\phi,$$

where $z = \dot{y}$, $\theta^* = [b, a]^T$, $\phi = [u, -y]^T$. In the unconstrained case the gradient adaptive law is given as

$$\dot{\theta} = \Gamma\varepsilon\phi, \quad \varepsilon = \frac{z - \theta^T\phi}{m_s^2},$$

where $m_s^2 = 1 + \phi^T\phi$; $\theta = [\hat{b}, \hat{a}]^T$; $\hat{b}$, $\hat{a}$ are the estimates of b, a, respectively. Since we know that $b \geq 1$ and $20 \geq a \geq -2$, we can constrain the estimates $\hat{b}$, $\hat{a}$ to be within the known bounds by using projection. Defining the sets for projection as

$$\begin{aligned} S_b &= \{\hat{b} \in \mathcal{R} | 1 - \hat{b} \leq 0\}, \\ S_a^l &= \{\hat{a} \in \mathcal{R} | -2 - \hat{a} \leq 0\}, \\ S_a^u &= \{\hat{a} \in \mathcal{R} | \hat{a} - 20 \leq 0\} \end{aligned}$$

and applying the projection algorithm (3.57) for each set, we obtain the adaptive laws

$$\dot{\hat{b}} = \begin{cases} \gamma_1\varepsilon u & \text{if } \hat{b} > 1 \text{ or } (\hat{b} = 1 \text{ and } \varepsilon u \geq 0), \\ 0 & \text{if } \hat{b} = 1 \text{ and } \varepsilon u < 0, \end{cases}$$

with $\hat{b}(0) \geq 1$, and

$$\dot{\hat{a}} = \begin{cases} -\gamma_2\varepsilon y & \text{if } 20 > \hat{a} > -2 \text{ or } (\hat{a} = -2 \text{ and } \varepsilon y \leq 0) \\ & \text{or } (\hat{a} = 20 \text{ and } \varepsilon y \geq 0), \\ 0 & \text{if } (\hat{a} = -2 \text{ and } \varepsilon y > 0) \text{ or } (\hat{a} = 20 \text{ and } \varepsilon y < 0), \end{cases}$$

with $\hat{a}(0)$ satisfying $20 \geq \hat{a}(0) \geq -2$. ∎

Example 3.10.3 Let us consider the gradient adaptive law

$$\dot{\theta} = \Gamma\varepsilon\phi$$

with the a priori knowledge that $|\theta^*| \leq M_0$ for some known bound $M_0 > 0$. In most applications, we may have such a priori information. We define

$$S = \left\{\theta \in \mathcal{R}^n | g(\theta) = \frac{\theta^T\theta}{2} - \frac{M_0^2}{2} \leq 0\right\}$$

and use (3.57) together with $\nabla g = \theta$ to obtain the adaptive law with projection

$$\dot{\theta} = \begin{cases} \Gamma\varepsilon\phi & \text{if } |\theta| < M_0 \text{ or } (|\theta| = M_0 \text{ and } \phi^T\Gamma\theta\varepsilon \leq 0), \\ \Gamma\varepsilon\phi - \Gamma\frac{\theta\theta^T}{\theta^T\Gamma\theta}\Gamma\varepsilon\phi & \text{if } |\theta| = M_0 \text{ and } \phi^T\Gamma\theta\varepsilon > 0 \end{cases}$$

with $|\theta(0)| \leq M_0$. ∎

3.11 Robust Parameter Identification

In the previous sections we designed and analyzed a wide class of PI algorithms based on the parametric models

$$z = \theta^{*T}\phi \quad \text{or} \quad W(s)\theta^{*T}\phi.$$

These parametric models are developed using a plant model that is assumed to be free of disturbances, noise, unmodeled dynamics, time delays, and other frequently encountered uncertainties. In the presence of plant uncertainties we are no longer able to express the unknown parameter vector θ^* in the form of the SPM or DPM where all signals are measured and θ^* is the only unknown term. In this case, the SPM or DPM takes the form

$$z = \theta^{*T}\phi + \eta \quad \text{or} \quad z = W(s)\theta^{*T}\phi + \eta, \tag{3.62}$$

where η is an unknown function that represents the modeling error terms. The following examples are used to show how (3.62) arises for different plant uncertainties.

Example 3.11.1 Consider the scalar constant gain system

$$y = \theta^* u + d, \tag{3.63}$$

where θ^* is the unknown scalar and d is a bounded external disturbance due to measurement noise and/or input disturbance. Equation (3.60) is already in the form of the SPM with modeling error given by (3.62). ∎

Example 3.11.2 Consider a system with a small input delay τ given by

$$y = \frac{b}{s+a}e^{-\tau s}u, \tag{3.64}$$

where a, b are the unknown parameters to be estimated and $u \in \mathcal{L}_\infty$. Since τ is small, the plant may be modeled as

$$y = \frac{b}{s+a}u \tag{3.65}$$

by assuming $\tau = 0$. Since a parameter estimator for a, b developed based on (3.65) has to be applied to the actual plant (3.64), it is of interest to see how $\tau \neq 0$ affects the parametric model for a, b. We express the plant as

$$y = \frac{b}{s+a}u + \frac{b}{s+a}(e^{-\tau s} - 1)u,$$

which we can rewrite in the form of the parametric model (3.62) as

$$z = \theta^{*T}\phi + \eta,$$

where

$$z = \frac{s}{s+\lambda}y, \qquad \theta^* = [b, a]^T, \qquad \phi = \begin{bmatrix} \frac{1}{s+\lambda}u \\ -\frac{1}{s+\lambda}y \end{bmatrix}, \qquad \eta = \frac{b}{s+\lambda}(e^{-\tau s} - 1)u,$$

and $\lambda > 0$. It is clear that $\tau = 0$ implies $\eta = 0$ and small τ implies small η. ∎

Example 3.11.3 Let us consider the plant

$$y = \theta^*(1 + \mu\Delta_m(s))u, \tag{3.66}$$

where μ is a small constant and $\Delta_m(s)$ is a proper transfer function with poles in the open left half s-plane. Since μ is small and $\Delta_m(s)$ is proper with stable poles, the term $\mu\Delta_m(s)$ can be treated as the modeling error term which can be approximated with zero. We can express (3.66) in the form of (3.62) as

$$y = \theta^* u + \eta,$$

where

$$\eta = \mu\theta^*\Delta_m(s)u$$

is the modeling error term. ∎

For LTI plants, the parametric model with modeling errors is usually of the form

$$\begin{aligned} z &= \theta^{*T} u + \eta, \\ \eta &= \Delta_1(s)u + \Delta_2(s)y + d, \end{aligned} \tag{3.67}$$

where $\Delta_1(s)$, $\Delta_2(s)$ are proper transfer functions with stable poles and d is a bounded disturbance. The principal question that arises is how the stability properties of the adaptive laws that are developed for parametric models with no modeling errors are affected when applied to the actual parametric models with uncertainties. The following example demonstrates that the adaptive laws of the previous sections that are developed using parametric models that are free of modeling errors cannot guarantee the same properties in the presence of modeling errors. Furthermore, it often takes only a small disturbance to drive the estimated parameters unbounded.

3.11.1 Instability Example

Consider the scalar constant gain system

$$y = \theta^* u + d,$$

where d is a bounded unknown disturbance and $u \in \mathcal{L}_\infty$. The adaptive law for estimating θ^* derived for $d = 0$ is given by

$$\dot{\theta} = \gamma\varepsilon u, \quad \varepsilon = y - \theta u, \tag{3.68}$$

where $\gamma > 0$ and the normalizing signal is taken to be 1. If $d = 0$ and $u, \dot{u} \in \mathcal{L}_\infty$, then we can establish that (i) $\dot{\theta}, \theta, \varepsilon \in \mathcal{L}_\infty$, (ii) $\varepsilon(t) \to 0$ as $t \to \infty$ by analyzing the parameter error equation

$$\dot{\tilde{\theta}} = -\gamma u^2\tilde{\theta},$$

which is a linear time-varying differential equation. When $d \neq 0$, we have

$$\dot{\tilde{\theta}} = -\gamma u^2\tilde{\theta} + \gamma du. \tag{3.69}$$

In this case we cannot guarantee that the parameter estimate $\theta(t)$ is bounded for any bounded input u and disturbance d. In fact, for $\theta^* = 2$, $\gamma = 1$,

$$u = (1+t)^{-1/2},$$
$$d(t) = (1+t)^{-1/4}\left(\frac{5}{4} - 2(1+t)^{-1/4}\right) \to 0 \quad \text{as } t \to \infty,$$

we have

$$y(t) = \frac{5}{4}(1+t)^{-1/4} \to 0 \quad \text{as } t \to \infty,$$
$$\varepsilon(t) = \frac{1}{4}(1+t)^{-1/4} \to 0 \quad \text{as } t \to \infty,$$
$$\theta(t) = (1+t)^{-1/4} \to \infty \quad \text{as } t \to \infty;$$

i.e., the estimated parameter drifts to infinity with time even though the disturbance $d(t)$ disappears with time. This instability phenomenon is known as *parameter drift*. It is mainly due to the pure integral action of the adaptive law, which, in addition to integrating the "good" signals, integrates the disturbance term as well, leading to the parameter drift phenomenon.

Another interpretation of the above instability is that, for $u = (1+t)^{-1/2}$, the homogeneous part of (3.69), i.e., $\dot{\tilde{\theta}} = -\gamma u^2 \tilde{\theta}$, is only uniformly stable, which is not sufficient to guarantee that the bounded input γdu will produce a bounded state $\tilde{\theta}$. If u is persistently exciting, i.e., $\int_t^{t+T_0} u^2(\tau)d\tau \geq \alpha_0 T_0$ for some $\alpha_0, T_0 > 0$ and $\forall t \geq 0$, then the homogeneous part of (3.69) is e.s. and the bounded input γdu produces a bounded state $\tilde{\theta}$ (show it!). Similar instability examples in the absence of PE may be found in [50–52, 105].

If the objective is parameter convergence, then parameter drift can be prevented by making sure the regressor vector is PE with a level of excitation higher than the level of the modeling error. In this case the plant input in addition to being sufficiently rich is also required to guarantee a level of excitation for the regressor that is higher than the level of the modeling error. This class of inputs is referred to as *dominantly rich* and is discussed in the following section.

3.11.2 Dominantly Rich Excitation

Let us revisit the example in section 3.11.1 and analyze (3.69), i.e.,

$$\dot{\tilde{\theta}} = -\gamma u^2 \tilde{\theta} + \gamma du, \tag{3.70}$$

when u is PE with level $\alpha_0 > 0$. The PE property of u implies that the homogeneous part of (3.70) is e.s., which in turn implies that

$$|\tilde{\theta}(t)| \leq e^{-\gamma\alpha_1 t}|\tilde{\theta}(0)| + \frac{1}{\alpha_1}(1 - e^{-\gamma\alpha_1 t}) \sup_{\tau \leq t} |u(\tau)d(\tau)|$$

for some $\alpha_1 > 0$ which depends on α_0. Therefore, we have

$$\lim_{t\to\infty} \sup_{\tau \geq t} |\tilde{\theta}(\tau)| \leq \frac{1}{\alpha_1} \lim_{t\to\infty} \sup_{\tau \geq t} |u(\tau)d(\tau)| = \frac{1}{\alpha_1} \sup_{\tau} |u(\tau)d(\tau)|. \tag{3.71}$$

The bound (3.71) indicates that the PI error at steady state is of the order of the disturbance; i.e., as $d \to 0$ the parameter error also reduces to zero. For this simple example, it is clear that if we choose $u = u_0$, where u_0 is a constant different from zero, then $\alpha_1 = \alpha_0 = u_0^2$; therefore, the bound for $|\tilde{\theta}|$ is $\sup_t \frac{|d(t)|}{u_0}$. Thus the larger the value of u_0 is, the smaller the parameter error. Large u_0 relative to $|d|$ implies large signal-to-noise ratio and therefore better accuracy of identification.

Example 3.11.4 (unmodeled dynamics) Let us consider the plant

$$y = \theta^*(1 + \Delta_m(s))u,$$

where $\Delta_m(s)$ is a proper transfer function with stable poles. If $\Delta_m(s)$ is much smaller than 1 for small s or all s, then the system can be approximated as

$$y = \theta^* u$$

and $\Delta_m(s)$ can be treated as an unmodeled perturbation. The adaptive law (3.68) that is designed for $\Delta_m(s) = 0$ is used to identify θ^* in the presence of $\Delta_m(s)$. The parameter error equation in this case is given by

$$\dot{\tilde{\theta}} = -\gamma u^2 \tilde{\theta} + \gamma u \eta, \quad \eta = \theta^* \Delta_m(s) u. \tag{3.72}$$

Since u is bounded and $\Delta_m(s)$ is stable, it follows that $\eta \in \mathcal{L}_\infty$ and therefore the effect of $\Delta_m(s)$ is to introduce the bounded disturbance term η in the adaptive law. Hence, if u is PE with level $\alpha_0 > 0$, we have, as in the previous example, that

$$\lim_{t\to\infty} \sup_{\tau \geq t} |\tilde{\theta}(\tau)| \leq \frac{1}{\alpha_1} \sup_t |u(t)\eta(t)|$$

for some $\alpha_1 > 0$. The question that comes up is how to choose u so that the above bound for $|\tilde{\theta}|$ is as small as possible. The answer to this question is not as straightforward as in the example of section 3.11.1 because η is also a function of u. The bound for $|\tilde{\theta}|$ depends on the choice of u and the properties of $\Delta_m(s)$. For example, for constant $u = u_0 \neq 0$, we have $\alpha_0 = \alpha_1 = u_0^2$ and $\eta = \theta^* \Delta_m(s) u_0$, i.e., $\lim_{t\to\infty} |\eta(t)| = |\theta^*||\Delta_m(0)||u_0|$, and therefore

$$\lim_{t\to\infty} \sup_{\tau \geq t} |\tilde{\theta}(\tau)| \leq |\Delta_m(0)||\theta^*|.$$

If the plant is modeled properly, $\Delta_m(s)$ represents a perturbation that is small in the low-frequency range, which is usually the range of interest. Therefore, for $u = u_0$, we should have $|\Delta_m(0)|$ small if not zero leading to the above bound, which is independent of u_0. Another choice of a PE input is $u = \cos \omega_0 t$ for some $\omega_0 \neq 0$. For this choice of u, since

$$e^{-\gamma \int_0^t \cos^2 \omega_0 \tau d\tau} = e^{-\frac{\gamma}{2}\left(t + \frac{\sin 2\omega_0 t}{2\omega_0}\right)} = e^{-\frac{\gamma}{4}\left(t + \frac{\sin 2\omega_0 t}{\omega_0}\right)} e^{-\frac{\gamma}{4} t} \leq e^{-\frac{\gamma}{4} t}$$

(where we used the inequality $t + \frac{\sin 2\omega_0 t}{\omega_0} \geq 0 \ \forall t \geq 0$) and $\sup_t |\eta(t)| \leq |\Delta_m(j\omega_0)||\theta^*|$, we have

$$\lim_{t\to\infty} \sup_{r \geq t} |\tilde{\theta}(r)| \leq c|\Delta_m(j\omega_0)|,$$

where $c = \frac{4|\theta^*|}{\gamma}$. This bound indicates that for small parameter error, ω_0 should be chosen so that $|\Delta_m(j\omega_0)|$ is as small as possible. If $\Delta_m(s)$ is due to high-frequency unmodeled dynamics, then $|\Delta_m(j\omega_0)|$ is small, provided ω_0 is a low frequency. As an example, consider

$$\Delta_m(s) = \frac{\mu s}{1 + \mu s},$$

where $\mu > 0$ is a small constant. It is clear that for low frequencies $|\Delta_m(j\omega)| = O(\mu)$[3] and $|\Delta_m(j\omega)| \to 1$ as $\omega \to \infty$. Since

$$|\Delta_m(j\omega_0)| = \frac{|\mu\omega_0|}{\sqrt{1 + \mu^2\omega_0^2}},$$

it follows that for $\omega_0 = \frac{1}{\mu}$ we have $|\Delta_m(j\omega_0)| = \frac{1}{\sqrt{2}}$ and for $\omega_0 \gg \frac{1}{\mu}$, $\Delta_m(j\omega_0) \approx 1$, whereas for $\omega_0 < \frac{1}{\mu}$ we have $|\Delta_m(j\omega_0)| = O(\mu)$. Therefore, for more accurate PI, the input signal should be chosen to be PE, but the PE property should be achieved with frequencies that do not excite the unmodeled dynamics. For the above example of $\Delta_m(s)$, $u = u_0$ does not excite $\Delta_m(s)$ at all, i.e., $\Delta_m(0) = 0$, whereas for $u = \sin\omega_0 t$ with $\omega_0 \ll \frac{1}{\mu}$, the excitation of $\Delta_m(s)$ is small leading to an $O(\mu)$ steady-state error for $|\tilde{\theta}|$. ∎

The above example demonstrates the well-known fact in control systems that the excitation of the plant should be restricted to be within the range of frequencies where the plant model is a good approximation of the actual plant. We explain this statement further using the plant

$$y = G_0(s)u + \Delta_a(s)u + d, \tag{3.73}$$

where $G_0(s)$, $\Delta_a(s)$ are proper and stable, $\Delta_a(s)$ is an additive perturbation of the modeled part $G_0(s)$, and d is a bounded disturbance. We would like to identify the coefficients of $G_0(s)$ by exciting the plant with the input u and processing the I/O data.

Because $\Delta_a(s)u$ is treated as a disturbance, the input u should be chosen so that at each frequency ω_i contained in u, we have $|G_0(j\omega_i)| \gg |\Delta_a(j\omega_i)|$. Furthermore, u should be rich enough to excite the modeled part of the plant that corresponds to $G_0(s)$ so that y contains sufficient information about the coefficients of $G_0(s)$. For such a choice of u to be possible, the spectrums of $G_0(s)$ and $\Delta_a(s)u$ should be separated, or $|G_0(j\omega)| \gg |\Delta_a(j\omega)|$ at all frequencies. If $G_0(s)$ is chosen properly, then $|\Delta_a(j\omega)|$ should be small relative to $|G_0(j\omega)|$ in the frequency range of interest. Since we are usually interested in the system response at low frequencies, we would assume that $|G_0(j\omega)| \gg |\Delta_a(j\omega)|$ in the low-frequency range for our analysis. But at high frequencies, we may have $|G_0(j\omega)|$ of the same order as or smaller than $|\Delta_a(j\omega)|$. The input signal u should therefore be designed to be sufficiently rich for the modeled part of the plant, but its richness should be achieved in the low-frequency range for which $|G_0(j\omega)| \gg |\Delta_a(j\omega)|$. An input signal with these two properties is called *dominantly rich* [106] because it excites the modeled or dominant part of the plant much more than the unmodeled one.

The separation of spectrums between the dominant and unmodeled part of the plant can be seen more clearly if we rewrite (3.73) as

$$y = G_0(s)u + \Delta_a(\mu s)u + d, \tag{3.74}$$

[3] A function $f(x)$ is of $O(\mu)$ $\forall x \in \Omega$ if there exists a constant $c \geq 0$ such that $\|f(x)\| \leq c|\mu|$ $\forall x \in \Omega$.

where $\Delta_a(\mu s)$ is due to high-frequency dynamics and $\mu > 0$ is a small parameter referred to as the singular perturbation parameter. $\Delta_a(\mu s)$ has the property that $\Delta_a(0) = 0$ and $|\Delta_a(\mu j\omega)| \leq O(\mu)$ for $\omega \ll \frac{1}{\mu}$, but $\Delta_a(\mu s)$ could satisfy $|\Delta_a(\mu j\omega)| = O(1)$ for $\omega \gg \frac{1}{\mu}$. Therefore, for μ small, the contribution of Δ_a in y is small, provided that the plant is not excited with frequencies of the order of $\frac{1}{\mu}$ or higher. Let us assume that we want to identify the coefficients of $G_0(s) = \frac{Z(s)}{R(s)}$ in (3.74) which consist of m zeros and n stable poles, a total of $n + m + 1$ parameters, by treating $\Delta_a(\mu s)u$ as a modeling error. We can express (3.74) in the form

$$z = \theta^{*T}\phi + \eta,$$

where $\eta = \frac{\Delta_a(\mu s)R(s)}{\Lambda(s)}u + d$, $\frac{1}{\Lambda(s)}$ is a filter with stable poles, $\theta^* \in \mathcal{R}^{n+m+1}$ contains the coefficients of the numerator and denominator of $G_0(s)$, and

$$\phi = \frac{1}{\Lambda(s)}[s^m u, s^{m-1}u, \ldots, u, -s^{n-1}y, -s^{n-2}y, \ldots, -y]^T.$$

If ϕ is PE, we can establish that the adaptive laws of the previous sections guarantee convergence of the parameter error to a set whose size depends on the bound for the modeling error signal η, i.e.,

$$|\tilde{\theta}(t)| \leq \alpha_1 e^{-\alpha t}(|\tilde{\theta}(0)| + \eta_0) + \alpha_2\eta_0,$$

where $\eta_0 = \sup_t |\eta(t)|$ and $\alpha, \alpha_1, \alpha_2 > 0$ are some constants that depend on the level of excitation of ϕ. The problem is how to choose u so that ϕ is PE despite the presence of $\eta \neq 0$. It should be noted that it is possible, for an input that is sufficiently rich, for the modeled part of the plant not to guarantee the PE property of ϕ in the presence of modeling errors because the modeling error could cancel or highly corrupt terms that are responsible for the PE property ϕ. We can see this by examining the relationship between u and ϕ given by

$$\phi = H_0(s)u + H_1(\mu s, s)u + B_1 d, \tag{3.75}$$

where

$$\begin{aligned}
H_0(s) &= \frac{1}{\Lambda(s)}[s^m, s^{m-1}, \ldots, s, 1, -s^{n-1}G_0(s), -s^{n-2}G_0(s), \ldots, -sG_0(s), -G_0(s)]^T,\\
H_1(\mu s, s) &= \frac{1}{\Lambda(s)}[0, \ldots, 0, -s^{n-1}\Delta_a(\mu s), -s^{n-2}\Delta_a(\mu s), \ldots, -\Delta_a(\mu s)]^T,\\
B_1 &= \frac{1}{\Lambda(s)}[0, \ldots, 0, -s^{n-1}, -s^{n-2}, \ldots, -s, -1]^T.
\end{aligned}$$

It is clear that the term $H_1(\mu s, s)u + B_1 d$ acts as a modeling error term and could destroy the PE property of ϕ, even for inputs u that are sufficiently rich of order $n + m + 1$. So the problem is to define the class of sufficiently rich inputs of order $n + m + 1$ that would guarantee ϕ to be PE despite the presence of the modeling error term $H_1(\mu s, s)u + B_1 d$ in (3.75).

Definition 3.11.5. *A sufficiently rich input u of order $n + m + 1$ for the dominant part of the plant* (3.74) *is called* dominantly rich *of order $n + m + 1$ if it achieves its richness with frequencies ω_i, $i = 1, 2, \ldots, N$, where $N \geq \frac{n+m+1}{2}$, $|\omega_i| < O(\frac{1}{\mu})$, $|\omega_i - \omega_j| > O(\mu)$, $i \neq j$, and $|u| > O(\mu) + O(d)$.*

Lemma 3.11.6. *Let $H_0(s)$, $H_1(\mu s, s)$ satisfy the following assumptions:*

(a) *The vectors $H_0(j\omega_1), H_0(j\omega_2), \ldots, H_0(j\omega_{\bar{n}})$ are linearly independent on $\mathcal{C}^{\bar{n}}$ for all possible $\omega_1, \omega_2, \ldots, \omega_{\bar{n}} \in R$, where $\bar{n} \triangleq n + m + 1$ and $\omega_i \neq \omega_k$ for $i \neq k$.*

(b) *For any set $\{\omega_1, \omega_2, \ldots, \omega_{\bar{n}}\}$ satisfying $|\omega_i - \omega_k| > O(\mu)$ for $i \neq k$ and $|\omega_i| < O(\frac{1}{\mu})$, we have $|\det(\bar{H})| > O(\mu)$, where $\bar{H} \triangleq [H_0(j\omega_1), H_0(j\omega_2), \ldots, H_0(j\omega_{\bar{n}})]$.*

(c) *$|H_1(j\mu\omega, j\omega)| \leq c$ for some constant c independent of μ and $\forall \omega \in \mathcal{R}$.*

Then there exists a $\mu^* > 0$ such that for $\mu \in [0, \mu^*)$, ϕ is PE of order $n + m + 1$ with level of excitation $\alpha_1 > O(\mu)$, provided that the input signal u is dominantly rich of order $n + m + 1$ for the plant (3.74).

Proof. Since $|u| > O(d)$, we can ignore the contribution of d in (3.75) and write

$$\phi = \phi_0 + \phi_1,$$

where

$$\phi_0 = H_0(s)u, \qquad \phi_1 = H_1(\mu s, s)u.$$

Since $H_0(s)$ does not depend on μ, ϕ_0 is the same signal vector as in the ideal case. The sufficient richness of order $\bar{n}$ of u together with the assumed properties (a)–(b) of $H_0(s)$ imply that ϕ_0 is PE with level $\alpha_0 > 0$ and α_0 is independent of μ, i.e.,

$$\frac{1}{T}\int_t^{t+T} \phi_0(\tau)\phi_0^T(\tau)d\tau \geq \alpha_0 I \tag{3.76}$$

$\forall t \geq 0$ and for some $T > 0$. On the other hand, since $H_1(\mu s, s)$ is stable and $|H_1(j\mu\omega, j\omega)| \leq c\ \forall \omega \in \mathcal{R}$, we have $\phi_1 \in \mathcal{L}_\infty$,

$$\frac{1}{T}\int_t^{t+T} \phi_1(\tau)\phi_1^T(\tau)d\tau \leq \beta I \tag{3.77}$$

for some constant β which is independent of μ. Because of (3.76) and (3.77), we have

$$\begin{aligned}
\frac{1}{T}\int_t^{t+T} \phi(\tau)\phi^T(\tau)d\tau &= \frac{1}{T}\int_t^{t+T} (\phi_0(\tau) + \mu\phi_1(\tau))(\phi_0^T(\tau) + \mu\phi_1^T(\tau))d\tau \\
&\geq \frac{1}{T}\left(\int_t^{t+T} \frac{\phi_0(\tau)\phi_0^T(\tau)}{2}d\tau - \mu^2\int_t^{t+T} \phi_1(\tau)\phi_1^T(\tau)d\tau\right) \\
&\geq \frac{\alpha_0}{2}I - \mu^2\beta I = \frac{\alpha_0}{4}I + \frac{\alpha_0}{4}I - \mu^2\beta I,
\end{aligned}$$

where the first inequality is obtained by using $(x + y)(x + y)^T \geq \frac{xx^T}{2} - yy^T$. In other words, ϕ has a level of PE $\alpha_1 = \frac{\alpha_0}{4}$ for $\mu \in [0, \mu^*)$, where $\mu^* \triangleq \sqrt{\frac{\alpha_0}{4\beta}}$. $\square$

For further reading on dominant richness and robust parameter estimation, see [56].

Example 3.11.7 Consider the plant

$$y = \frac{b}{s+a}\left(1 + \frac{2\mu(s-1)}{(\mu s+1)^2}\right)u,$$

where a, b are the unknown parameters and $\mu = 0.001$. The plant may be modeled as

$$y = \frac{b}{s+a}u$$

by approximating $\mu = 0.001 \cong 0$. The input $u = \sin\omega_0 t$ with $1 \ll \omega_0 \ll 1000$ would be a dominantly rich input of order 2. Frequencies such as $\omega_0 = 0.006$ rad/sec or $\omega_0 = 900$ rad/sec would imply that u is not dominantly rich even though u is sufficiently rich of order 2. ∎

3.12 Robust Adaptive Laws

If the objective in online PI is convergence of the estimated parameters (close) to their true values, then dominantly rich excitation of appropriate order will meet the objective, and no instabilities will be present. In many applications, such as in adaptive control, the plant input is the result of feedback and cannot be designed to be dominantly or even sufficiently rich. In such situations, the objective is to drive the plant output to zero or force it to follow a desired trajectory rather than convergence of the online parameter estimates to their true values. It is therefore of interest to guarantee stability and robustness for the online parameter estimators even in the absence of persistence of excitation. This can be achieved by modifying the adaptive laws of the previous sections to guarantee stability and robustness in the presence of modeling errors independent of the properties of the regressor vector ϕ.

Let us consider the general plant

$$y = G_0(s)u + \Delta_u(s)u + \Delta_y(s)y + d, \tag{3.78}$$

where $G_0(s)$ is the dominant part, $\Delta_u(s)$, $\Delta_y(s)$ are strictly proper with stable poles and small relative to $G_0(s)$, and d is a bounded disturbance. We are interested in designing adaptive laws for estimating the coefficients of

$$G_0(s) = \frac{Z(s)}{R(s)} = \frac{b_m s^m + \cdots + b_1 s + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0}.$$

The SPM for (3.78) is

$$z = \theta^{*T}\phi + \eta, \tag{3.79}$$

where

$$\theta^* = [b_m, \ldots, b_0, a_{n-1}, \ldots, a_0]^T \in \mathcal{R}^{n+m+1},$$
$$z = \frac{s^n}{\Lambda(s)}y,$$

$$\phi = \left[\frac{s^m}{\Lambda(s)}u, \ldots, \frac{1}{\Lambda(s)}u, -\frac{s^{n-1}}{\Lambda(s)}y, \ldots, -\frac{1}{\Lambda(s)}y\right]^T,$$
$$\eta = \Delta_u(s)\frac{R(s)}{\Lambda(s)}u + \Delta_y(s)\frac{R(s)}{\Lambda(s)}y + \frac{R(s)}{\Lambda(s)}d.$$

The modeling error term η is driven by both u and y and cannot be assumed to be bounded unless u and y are bounded.

The parametric model can be transformed into one where the modeling error term is bounded by using normalization, a term already used in the previous sections. Let us assume that we can find a signal $m_s > 0$ with the property $\frac{\phi}{m_s}, \frac{\eta}{m_s} \in \mathcal{L}_\infty$. Then we can rewrite (3.79) as

$$\bar{z} = \theta^{*T}\bar{\phi} + \bar{\eta}, \tag{3.80}$$

where $\bar{z} = \frac{z}{m_s}$, $\bar{\phi} = \frac{\phi}{m_s}$, $\bar{\eta} = \frac{\eta}{m_s}$. The normalized parametric model has all the measured signals bounded and can be used to develop online adaptive laws for estimating θ^*. The bounded term $\bar{\eta}$ can still drive the adaptive laws unstable, as we showed using an example in section 3.11.1. Therefore, for robustness, we need to use the following modifications:

- Design the normalizing signal m_s to bound the modeling error in addition to bounding the regressor vector ϕ.
- Modify the "pure" integral action of the adaptive laws to prevent parameter drift.

In the following sections, we develop a class of normalizing signals and modified adaptive laws that guarantee robustness in the presence of modeling errors.

3.12.1 Dynamic Normalization

Let us consider the SPM with modeling error

$$z = \theta^{*T}\phi + \eta,$$

where

$$\eta = \Delta_1(s)u + \Delta_2(s)y.$$

$\Delta_1(s)$, $\Delta_2(s)$ are strictly proper transfer functions with stable poles. Our objective is to design a signal m_s so that $\frac{\eta}{m_s} \in \mathcal{L}_\infty$. We assume that $\Delta_1(s)$, $\Delta_2(s)$ are analytic in $\Re[s] \geq -\frac{\delta_0}{2}$ for some known $\delta_0 > 0$. Apart from this assumption, we require no knowledge of the parameters and/or dimension of $\Delta_1(s)$, $\Delta_2(s)$. It is implicitly assumed, however, that they are small relative to the modeled part of the plant in the frequency range of interest; otherwise, they would not be treated as modeling errors.

Using the properties of the $\mathcal{L}_{2\delta}$ norm (see Lemma A.5.9), we can write

$$|\eta(t)| \leq \|\Delta_1(s)\|_{2\delta}\|u_t\|_{2\delta} + \|\Delta_2(s)\|_{2\delta}\|y_t\|_{2\delta}$$

for any $\delta \in [0, \delta_0]$, where the above norms are defined as

$$\|x_t\|_{2\delta} \triangleq \left(\int_0^t e^{-\delta(t-\tau)}x^T(\tau)x(\tau)d\tau\right)^{1/2},$$

$$\|H(s)\|_{2\delta} \triangleq \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{\infty} \left| H\left(j\omega - \frac{\delta}{2}\right) \right|^2 d\omega \right)^{1/2}.$$

If we define

$$n_d = \|u_t\|_{2\delta_0}^2 + \|y_t\|_{2\delta_0}^2,$$

which can be generated by the differential equation

$$\dot{n}_d = -\delta_0 n_d + u^2 + y^2, \quad n_d(0) = 0,$$

then it follows that

$$m_s^2 = 1 + n_d$$

bounds $|\eta(t)|$ from above and

$$\frac{|\eta(t)|}{m_s} \leq \|\Delta_1(s)\|_{2\delta_0} + \|\Delta_2(s)\|_{2\delta_0}.$$

The normalizing signal m_s used in the adaptive laws for parametric models free of modeling errors is required to bound the regressor ϕ from above. In the presence of modeling errors, m_s should be chosen to bound both ϕ and the modeling error η for improved robustness properties. In this case the normalizing signal m_s has the form

$$m_s^2 = 1 + n_s^2 + n_d,$$

where n_s^2 is the static part and n_d the dynamic one. Examples of static and dynamic normalizing signals are $n_s^2 = \phi^T \phi$ or $\phi^T P \phi$, where $P = P^T > 0$,

$$\dot{n}_d = -\delta_0 n_d + \delta_1 (u^2 + y^2), \quad n_d(0) = 0, \tag{3.81}$$

or

$$n_d = n_1^2, \quad \dot{n}_1 = -\delta_0 n_1 + \delta_1 (|u| + |y|), \quad n_1(0) \geq 0, \tag{3.82}$$

or

$$n_d = n_\infty^2, \quad n_\infty = \delta_1 \max\left(\sup_{\tau \leq t} |u(\tau)|, \sup_{\tau \leq t} |y(\tau)| \right). \tag{3.83}$$

Any one of choices (3.81)–(3.83) can be shown to guarantee that $\frac{\eta}{m_s} \in \mathcal{L}_\infty$. Since $\phi = H(s)[\begin{smallmatrix} u \\ y \end{smallmatrix}]$, the dynamic normalizing signal can be chosen to bound ϕ from above, provided that $H(s)$ is analytic in $\Re[s] \geq -\frac{\delta_0}{2}$, in which case $m_s^2 = 1 + n_d$ bounds both ϕ and η from above.

3.12.2 Robust Adaptive Laws: σ-Modification

A class of robust modifications involves the use of a small feedback around the "pure" integrator in the adaptive law, leading to the adaptive law structure

$$\dot{\theta} = \Gamma\varepsilon\phi - \sigma_\ell\Gamma\theta, \tag{3.84}$$

where $\sigma_\ell \geq 0$ is a small design parameter and $\Gamma = \Gamma^T > 0$ is the adaptive gain, which in the case of LS is equal to the covariance matrix P. The above modification is referred to as the *σ-modification* [50, 51, 59] or as *leakage*.

Different choices of σ_ℓ lead to different robust adaptive laws with different properties, described in the following sections. We demonstrate the properties of these modifications when applied to the SPM with modeling error,

$$z = \theta^{*T}\phi + \eta, \tag{3.85}$$

where η is the modeling error term that is bounded from above by the normalizing signal. Let us assume that η is of the form

$$\eta = \Delta_1(s)u + \Delta_2(s)y,$$

where $\Delta_1(s)$, $\Delta_2(s)$ are strictly proper transfer functions analytic in $\Re[s] \geq -\frac{\delta_0}{2}$ for some known $\delta_0 > 0$. The normalizing signal can be chosen as

$$\begin{aligned} m_s^2 &= 1 + \alpha_0\phi^T\phi + \alpha_1 n_d, \\ \dot{n}_d &= -\delta_0 n_d + \delta_1(u^2 + y^2), \quad n_d(0) = 0, \end{aligned} \tag{3.86}$$

for some design constants $\delta_1, \alpha_0, \alpha_1 > 0$, and shown to guarantee that $\phi/m_s, \eta/m_s \in \mathcal{L}_\infty$.

Fixed σ-Modification

In this case

$$\sigma_\ell(t) = \sigma > 0 \quad \forall t \geq 0, \tag{3.87}$$

where σ is a small positive design constant. The gradient adaptive law for estimating θ^* in (3.85) takes the form

$$\begin{aligned} \dot{\theta} &= \Gamma\varepsilon\phi - \sigma\Gamma\theta, \\ \varepsilon &= \frac{z - \theta^T\phi}{m_s^2}, \end{aligned} \tag{3.88}$$

where m_s is given by (3.86). If some a priori estimate θ_0 of θ^* is available, then the term $\sigma\Gamma\theta$ may be replaced with $\sigma\Gamma(\theta - \theta_0)$ so that the leakage term becomes larger for larger deviations of θ from θ_0 rather than from zero.

Theorem 3.12.1. *The adaptive law* (3.88) *guarantees the following properties:*

(i) $\varepsilon, \varepsilon m_s, \theta, \dot{\theta} \in \mathcal{L}_\infty$.

(ii) $\varepsilon, \varepsilon m_s, \dot{\theta} \in \mathcal{S}(\sigma + \frac{\eta^2}{m_s^2})$.[4]

(iii) *If* $\frac{\phi}{m_s}$ *is PE with level* $\alpha_0 > 0$, *then* $\theta(t)$ *converges exponentially to the residual set* $D_\sigma = \{\tilde{\theta} \mid |\tilde{\theta}| \le c(\sigma + \bar{\eta})\}$, *where* $\bar{\eta} = \sup_t \frac{|\eta(t)|}{m_s(t)}$ *and* $c \ge 0$ *is some constant.*

Proof. (i) From (3.88), we obtain

$$\dot{\tilde{\theta}} = \Gamma\varepsilon\phi - \sigma\Gamma\theta, \quad \varepsilon m_s^2 = -\tilde{\theta}^T\phi + \eta.$$

Consider the Lyapunov-like function

$$V(\tilde{\theta}) = \frac{\tilde{\theta}^T\Gamma^{-1}\tilde{\theta}}{2}.$$

We have

$$\dot{V} = \tilde{\theta}^T\phi\varepsilon - \sigma\tilde{\theta}^T\theta$$

or

$$\dot{V} = -\varepsilon^2 m_s^2 + \varepsilon\eta - \sigma\tilde{\theta}^T(\tilde{\theta} + \theta^*) \le -\varepsilon^2 m_s^2 + |\varepsilon m_s|\left|\frac{\eta}{m_s}\right| - \sigma\tilde{\theta}^T\tilde{\theta} + \sigma|\tilde{\theta}||\theta^*|.$$

Using $-a^2 + ab = -\frac{a^2}{2} - \frac{1}{2}(a-b)^2 + \frac{b^2}{2} \le -\frac{a^2}{2} + \frac{b^2}{2}$ for any a, b, we write

$$\dot{V} \le -\frac{\varepsilon^2 m_s^2}{2} - \frac{\sigma}{2}|\tilde{\theta}|^2 + \frac{1}{2}\frac{|\eta|^2}{m_s^2} + \frac{\sigma}{2}|\theta^*|^2.$$

Since $\frac{\eta}{m_s} \in \mathcal{L}_\infty$, the positive terms in the expression for $\dot{V}$ are bounded from above by a constant. Now

$$V(\tilde{\theta}) \le |\tilde{\theta}|^2\frac{\lambda_{\max}(\Gamma^{-1})}{2}$$

and therefore

$$-\frac{\sigma}{2}|\tilde{\theta}|^2 \le -\frac{\sigma}{\lambda_{\max}(\Gamma^{-1})}V(\tilde{\theta}),$$

which implies

$$\dot{V} \le -\frac{\varepsilon^2 m_s^2}{2} - \frac{\sigma}{\lambda_{\max}(\Gamma^{-1})}V(\tilde{\theta}) + c_0\left(\sigma + \frac{\eta^2}{m_s^2}\right),$$

where $c_0 = \max\{\frac{|\theta^*|^2}{2}, \frac{1}{2}\}$. Hence for

$$V(\tilde{\theta}) \ge V_0 \triangleq c_0(\sigma + \bar{\eta}^2)\frac{\lambda_{\max}(\Gamma^{-1})}{\sigma},$$

where $\bar{\eta} = \sup_t \frac{|\eta(t)|}{m_s(t)}$, we have $\dot{V} \le -\frac{\varepsilon^2 m_s^2}{2} \le 0$, which implies that V is bounded and therefore $\tilde{\theta} \in \mathcal{L}_\infty$. From $\varepsilon m_s = -\frac{\tilde{\theta}^T\phi}{m_s} + \frac{\eta}{m_s}$, $\tilde{\theta} \in \mathcal{L}_\infty$, and the fact that m_s bounds

[4] As defined in the Appendix, $\mathcal{S}(\mu) = \{x : [0,\infty) \to \mathcal{R}^n \mid \int_t^{t+T} x^T(\tau)x(\tau)d\tau \le c_0\mu T + c_1\ \forall t, T \ge 0\}$.

ϕ, η from above, we have $\varepsilon m_s, \varepsilon \in \mathcal{L}_\infty$. Furthermore, using $\dot{\tilde{\theta}} = \Gamma \varepsilon m_s \frac{\phi}{m_s} - \sigma \Gamma \theta$ and $\theta, \varepsilon m_s, \frac{\phi}{m_s} \in \mathcal{L}_\infty$, we obtain $\dot{\tilde{\theta}} = \dot{\theta} \in \mathcal{L}_\infty$.

(ii) The inequality

$$\dot{V} \leq -\frac{\varepsilon^2 m_s^2}{2} - \frac{\sigma}{\lambda_{\max}(\Gamma^{-1})} V(\tilde{\theta}) + c_0 \left(\sigma + \frac{\eta^2}{m_s^2}\right)$$

obtained above implies

$$\dot{V} \leq -\frac{\varepsilon^2 m_s^2}{2} + c_0 \left(\sigma + \frac{\eta^2}{m_s^2}\right).$$

Integrating both sides, we obtain

$$\frac{1}{2} \int_0^t \varepsilon^2 m_s^2 d\tau \leq V(0) - V(t) + c_0 \int_0^t \left(\sigma + \frac{\eta^2}{m_s^2}\right) d\tau,$$

which implies that $\varepsilon m_s \in \mathcal{S}(\sigma + \frac{\eta^2}{m_s^2})$. This together with $\theta, \frac{\phi}{m_s} \in \mathcal{L}_\infty$ implies that $\varepsilon, \dot{\theta} \in \mathcal{S}(\sigma + \frac{\eta^2}{m_s^2})$.

(iii) Using $\varepsilon m_s = -\frac{\tilde{\theta}^T \phi}{m_s} + \frac{\eta}{m_s} = -\frac{\phi^T \tilde{\theta}}{m_s} + \frac{\eta}{m_s}$, we have

$$\dot{\tilde{\theta}} = -\Gamma \frac{\phi}{m_s} \frac{\phi^T}{m_s} \tilde{\theta} - \sigma \Gamma \theta + \frac{\eta}{m_s}.$$

Since $\frac{\phi}{m_s}$ is PE, it follows that the homogenous part of the above equation is e.s., which implies that

$$|\tilde{\theta}(t)| \leq \alpha_1 e^{-\alpha_0 t} |\tilde{\theta}(t)| + \alpha_1 \int_0^t e^{-\alpha_0 (t-\tau)} \left(\sigma \|\Gamma\| |\theta(\tau)| + \frac{|\eta|}{m_s}\right) d\tau$$

for some $\alpha_1, \alpha_0 > 0$. Using $\frac{\eta}{m_s}, \theta \in \mathcal{L}_\infty$, we can write

$$|\tilde{\theta}(t)| \leq \alpha_1 e^{-\alpha_0 t} |\tilde{\theta}(0)| + c(\sigma + \bar{\eta})(1 - e^{-\alpha_0 t}),$$

where $\bar{\eta} = \sup_t \frac{|\eta(t)|}{m_s(t)}$, $c = \max(\frac{\alpha_1}{\alpha_0} \|\Gamma\| \sup_t |\theta(t)|, \frac{\alpha_1}{\alpha_0})$, which concludes that $\tilde{\theta}(t)$ converges to D_σ exponentially fast. $\square$

One can observe that if the modeling error is removed, i.e., $\eta = 0$, then the fixed σ-modification will not guarantee the ideal properties of the adaptive law since it introduces a disturbance of the order of the design constant σ. This is one of the main drawbacks of the fixed σ-modification that is removed in the next section. One of the advantages of the fixed σ-modification is that no assumption about bounds or location of the unknown θ^* is made.

Switching σ-Modification

The drawback of the fixed σ-modification is eliminated by using a switching σ_ℓ term which activates the small feedback term around the integrator when the magnitude of the parameter

vector exceeds a certain value M_0. The assumptions we make in this case are that $|\theta^*| \leq M_0$ and M_0 is known. Since M_0 is arbitrary it can be chosen to be high enough in order to guarantee $|\theta^*| \leq M_0$ in the case where limited or no information is available about the location of θ^*. The switching σ-modification is given by

$$\sigma_\ell(t) = \sigma_s = \begin{cases} 0 & \text{if } |\theta| \leq M_0, \\ \left(\frac{|\theta|}{M_0} - 1\right)^{q_0} \sigma_0 & \text{if } M_0 < |\theta| \leq 2M_0, \\ \sigma_0 & \text{if } |\theta| > 2M_0, \end{cases} \tag{3.89}$$

where $q_0 \geq 1$ is any finite integer and M_0, σ_0 are design constants satisfying $M_0 > |\theta^*|$ and $\sigma_0 > 0$. The switching from 0 to σ_0 is continuous in order to guarantee the existence and uniqueness of solutions of the differential equation. If an a priori estimate θ_0 of θ^* is available, it can be incorporated in (3.89) to replace $|\theta|$ with $|\theta - \theta_0|$. In this case, M_0 is an upper bound for $|\theta^* - \theta_0| \leq M_0$.

The gradient algorithm with the switching σ-modification given by (3.89) is described as

$$\begin{aligned} \dot{\theta} &= \Gamma \varepsilon \phi - \sigma_s \Gamma \theta, \\ \varepsilon &= \frac{z - \theta^T \phi}{m_s^2}, \end{aligned} \tag{3.90}$$

where m_s is given by (3.86).

Theorem 3.12.2. *The gradient algorithm* (3.90) *with the switching σ-modification guarantees the following:*

(i) $\varepsilon, \varepsilon m_s, \theta, \dot{\theta} \in \mathcal{L}_\infty$.

(ii) $\varepsilon, \varepsilon m_s, \dot{\theta} \in \mathcal{S}(\frac{\eta^2}{m_s^2})$.

(iii) *In the absence of modeling error, i.e., for $\eta = 0$, the properties of the adaptive law* (3.90) *are the same as those of the respective unmodified adaptive laws* (*i.e., with* $\sigma_\ell = 0$).

(iv) *If $\frac{\phi}{m_s}$ is PE with level $\alpha_0 > 0$, then*

(a) *$\tilde{\theta}$ converges exponentially fast to the residual set*

$$D_s = \{\tilde{\theta} \mid |\tilde{\theta}| \leq c(\sigma_0 + \bar{\eta})\},$$

where $c \geq 0$ is some constant;

(b) *there exists a constant $\bar{\eta}^* > 0$ such that for $\bar{\eta} < \bar{\eta}^*$, $\tilde{\theta}$ converges exponentially to the residual set*

$$\bar{D}_s = \{\tilde{\theta} \mid |\tilde{\theta}| \leq c\bar{\eta}\}.$$

Proof. We choose the same Lyapunov-like function as in the case of no modeling error, i.e.,

$$V = \frac{\tilde{\theta}^T \Gamma^{-1} \tilde{\theta}}{2}.$$

Then

$$\dot{V} = \tilde{\theta}^T \phi \varepsilon - \sigma_s \tilde{\theta}^T \theta.$$

Substituting $\varepsilon m_s^2 = -\tilde{\theta}^T \phi + \eta$, we obtain

$$\dot{V} = -\varepsilon^2 m_s^2 + \varepsilon \eta - \sigma_s \tilde{\theta}^T \theta.$$

Note that $-\varepsilon^2 m_s^2 + \varepsilon \eta \leq -\frac{\varepsilon^2 m_s^2}{2} + \frac{\eta^2}{2m_s^2}$ and

$$\sigma_s \tilde{\theta}^T \theta = \sigma_s (|\theta|^2 - \theta^{*T} \theta) \geq \sigma_s |\theta| (|\theta| - M_0 + M_0 - |\theta^*|).$$

Since $\sigma_s(|\theta| - M_0) \geq 0$ and $M_0 > |\theta^*|$, it follows that

$$\sigma_s \tilde{\theta}^T \theta \geq \sigma_s |\theta| (|\theta| - M_0) + \sigma_s |\theta| (M_0 - |\theta^*|) \geq \sigma_s |\theta| (M_0 - |\theta^*|) \geq 0,$$

i.e.,

$$\sigma_s |\theta| \leq \sigma_s \frac{\tilde{\theta}^T \theta}{M_0 - |\theta^*|}. \tag{3.91}$$

Therefore, the inequality for $\dot{V}$ can be written as

$$\dot{V} \leq -\frac{\varepsilon^2 m_s^2}{2} - \sigma_s \tilde{\theta}^T \theta + \frac{\eta^2}{2m_s^2}. \tag{3.92}$$

Since for $|\theta| = |\tilde{\theta} + \theta^*| > 2M_0$ the term $-\sigma_s \tilde{\theta}^T \theta = -\sigma_0 \tilde{\theta}^T \theta \leq -\frac{\sigma_0}{2} |\tilde{\theta}|^2 + \frac{\sigma_0}{2} |\theta^*|^2$ behaves as the equivalent fixed σ term, we can follow the same procedure as in the proof of Theorem 3.12.1 to show the existence of a constant $V_0 > 0$ for which $\dot{V} \leq 0$ whenever $V \geq V_0$ and conclude that $V, \varepsilon, \theta, \dot{\theta} \in \mathcal{L}_\infty$, completing the proof of (i).

Integrating both sides of (3.92) from t_0 to t, we obtain that $\varepsilon, \varepsilon m_s, \sqrt{\sigma_s \tilde{\theta}^T \theta} \in \mathcal{S}(\frac{\eta^2}{m_s^2})$. From (3.91), it follows that

$$\sigma_s^2 |\theta|^2 \leq c_2 \sigma_s \tilde{\theta}^T \theta$$

for some constant $c_2 > 0$ that depends on the bound for $\sigma_0 |\theta|$, and therefore $|\dot{\theta}|^2 \leq c(|\varepsilon m_s|^2 + \sigma_s \tilde{\theta}^T \theta)$ for some $c \geq 0$. Since $\varepsilon m_s, \sqrt{\sigma_s \tilde{\theta}^T \theta} \in \mathcal{S}(\frac{\eta^2}{m_s^2})$, it follows that $\dot{\theta} \in \mathcal{S}(\frac{\eta^2}{m_s^2})$, which completes the proof of (ii).

The proof of part (iii) follows from (3.92) by setting $\eta = 0$, using $-\sigma_s \tilde{\theta}^T \theta \leq 0$, and repeating the above calculations for $\eta = 0$.

The proof of (iv)(a) is almost identical to that of Theorem 3.12.1(iii) and is omitted.

To prove (iv)(b), we follow the same arguments used in the proof of Theorem 3.12.1(iii) to obtain the inequality

$$\begin{aligned} |\tilde{\theta}| &\leq \beta_0 e^{-\beta_2 t} + \beta_1' \int_0^t e^{-\beta_2 (t-\tau)} (|\eta| + \sigma_s |\theta|) d\tau \\ &\leq \beta_0 e^{-\beta_2 t} + \frac{\beta_1'}{\beta_2} \bar{\eta} + \int_0^t e^{-\beta_2 (t-\tau)} \sigma_s |\theta| d\tau \end{aligned} \tag{3.93}$$

for some positive constants β_0, β_1', β_2. From (3.91), we have

$$\sigma_s|\theta| \le \frac{1}{M_0 - |\theta^*|}\sigma_s\tilde{\theta}^T\theta \le \frac{1}{M_0 - |\theta^*|}\sigma_s|\theta||\tilde{\theta}|. \tag{3.94}$$

Therefore, using (3.94) in (3.93), we have

$$|\tilde{\theta}| \le \beta_0 e^{-\beta_2 t} + \frac{\beta_1'}{\beta_2}\bar{\eta} + \beta_1''\int_0^t e^{-\beta_2(t-\tau)}\sigma_s|\theta||\tilde{\theta}|d\tau, \tag{3.95}$$

where $\beta_1'' = \frac{\beta_1'}{M_0-|\theta^*|}$. Applying the Bellman–Gronwall (B–G) Lemma 3 (Lemma A.6.3) to (3.95), it follows that

$$|\tilde{\theta}| \le \left(\beta_0 + \frac{\beta_1'}{\beta_2}\bar{\eta}\right)e^{-\beta_2(t-t_0)}e^{\beta_1''\int_{t_0}^t \sigma_s|\theta|ds} + \beta_1'\bar{\eta}\int_{t_0}^t e^{-\beta_2(t-\tau)}e^{\beta''\int_{t_0}^t \sigma_s|\theta|ds}d\tau. \tag{3.96}$$

Note from (3.91)–(3.92) that $\sqrt{\sigma_s|\theta|} \in \mathcal{S}(\frac{\eta^2}{m_s^2})$, i.e.,

$$\int_{t_0}^t \sigma_s|\theta|ds \le c_1\bar{\eta}^2(t-t_0) + c_0$$

$\forall t \ge t_0 \ge 0$ and for some constants c_0, c_1. Therefore,

$$|\tilde{\theta}| \le \bar{\beta}_1 e^{-\bar{\alpha}(t-t_0)} + \bar{\beta}_2\bar{\eta}\int_{t_0}^t e^{-\bar{\alpha}(t-\tau)}d\tau, \tag{3.97}$$

where $\bar{\alpha} = \beta_2 - \beta_2''c_1\bar{\eta}^2$ and $\bar{\beta}_1, \bar{\beta}_2 \ge 0$ are some constants that depend on c_0 and on the constants in (3.96). Hence, for any $\bar{\eta} \in [0, \bar{\eta}^*)$, where $\bar{\eta}^* = \sqrt{\frac{\beta_2}{\beta_1''c_1}}$, we have $\bar{\alpha} > 0$, and (3.97) implies that

$$|\tilde{\theta}| \le \frac{\bar{\beta}_2}{\bar{\alpha}}\bar{\eta} + ce^{-\bar{\alpha}(t-t_0)}$$

for some constant c and $\forall t \ge t_0 \ge 0$, which completes the proof of (iv). □

Another class of σ-modification involves leakage that depends on the estimation error ε, i.e.,

$$\sigma_s = |\varepsilon m_s|\nu_0,$$

where $\nu_0 > 0$ is a design constant. This modification is referred to as the *ε-modification* and has properties similar to those of the fixed σ-modification in the sense that it cannot guarantee the ideal properties of the adaptive law in the absence of modeling errors. For more details on the ε-modification, see [56, 86, 107].

The use of σ-modification is to prevent parameter drift by forcing $\theta(t)$ to remain bounded. It is often referred to as *soft projection* because it forces $\theta(t)$ not to deviate far from the bound $M_0 \ge |\theta|$. That is, in the case of soft projection $\theta(t)$ may exceed the M_0 bound, but it will still be bounded.

In the following section, we use projection to prevent parameter drift by forcing the estimated parameters to be within a specified bounded set.

3.12.3 Parameter Projection

The two crucial techniques that we used in sections 3.12.1 and 3.12.2 to develop robust adaptive laws are the dynamic normalization m_s and leakage. The normalization guarantees that the normalized modeling error term $\frac{\eta}{m_s}$ is bounded and therefore acts as a bounded input disturbance in the adaptive law. Since a bounded disturbance may cause parameter drift, the leakage modification is used to guarantee bounded parameter estimates. Another effective way to guarantee bounded parameter estimates is to use projection to constrain the parameter estimates to lie inside a bounded convex set in the parameter space that contains the unknown θ^*. Adaptive laws with projection have already been introduced and analyzed in section 3.10. By requiring the parameter set to be bounded, projection can be used to guarantee that the estimated parameters are bounded by forcing them to lie within the bounded set. In this section, we illustrate the use of projection for a gradient algorithm that is used to estimate θ^* in the parametric model

$$z = \theta^{*T}\phi + \eta.$$

In order to avoid parameter drift, we constrain θ to lie inside a bounded convex set that contains θ^*. As an example, consider the set

$$\wp = \{\theta \mid g(\theta) = \theta^T\theta - M_0^2 \le 0\},$$

where M_0 is chosen so that $M_0 \ge |\theta^*|$. Following the results of section 3.10, we obtain

$$\dot{\theta} = \begin{cases} \Gamma\varepsilon\phi & \text{if } \theta^T\theta < M_0^2 \\ & \text{or if } \theta^T\theta = M_0^2 \text{ and } (\Gamma\varepsilon\phi)^T\theta \le 0, \\ \left(I - \frac{\Gamma\theta\theta^T}{\theta^T\Gamma\theta}\right)\Gamma\varepsilon\phi & \text{otherwise,} \end{cases} \tag{3.98}$$

where $\theta(0)$ is chosen so that $\theta^T(0)\theta(0) \le M_0^2$ and $\varepsilon = \frac{z-\theta^T\phi}{m_s^2}$, $\Gamma = \Gamma^T > 0$.

The stability properties of (3.98) for estimating θ^* in (3.85) in the presence of the modeling error term η are given by the following theorem.

Theorem 3.12.3. *The gradient algorithm with projection described by* (3.98) *and designed for the parametric model $z = \theta^{*T}\phi + \eta$ guarantees the following:*

(i) $\varepsilon, \varepsilon m_s, \theta, \dot{\theta} \in \mathcal{L}_\infty$.

(ii) $\varepsilon, \varepsilon m_s, \dot{\theta} \in \mathcal{S}(\frac{\eta^2}{m_s^2})$.

(iii) *If $\eta = 0$, then $\varepsilon, \varepsilon m_s, \dot{\theta} \in \mathcal{L}_2$.*

(iv) *If $\frac{\phi}{m_s}$ is PE with level $\alpha_0 > 0$, then*

(a) *$\tilde{\theta}$ converges exponentially to the residual set*

$$D_p = \{\tilde{\theta} \mid |\tilde{\theta}| \le c(f_0 + \bar{\eta})\},$$

where $\bar{\eta} = \sup_t \frac{|\eta|}{m}$, $c \ge 0$ is a constant, and $f_0 \ge 0$ is a design constant;

(b) *there exists a constant* $\bar{\eta}^* > 0$ *such that if* $\bar{\eta} < \bar{\eta}^*$, *then* $\bar{\theta}$ *converges exponentially fast to the residual set*

$$D_p = \{\tilde{\theta} \mid |\tilde{\theta}| \leq c\bar{\eta}\}$$

for some constant $c \geq 0$.

Proof. As established in section 3.10, projection guarantees that $|\theta(t)| \leq M_0 \ \forall t \geq 0$, provided $|\theta(0)| \leq M_0$. Let us choose the Lyapunov-like function

$$V = \frac{\tilde{\theta}^T \Gamma^{-1} \tilde{\theta}}{2}.$$

Along the trajectory of (3.98), we have

$$\dot{V} = \begin{cases} -\varepsilon^2 m_s^2 + \varepsilon\eta & \text{if } \theta^T\theta < M_0^2 \\ & \text{or if } \theta^T\theta = M_0^2 \text{ and } (\Gamma\varepsilon\phi)^T\theta \leq 0, \\ -\varepsilon^2 m_s^2 + \varepsilon\eta - \frac{\tilde{\theta}^T\theta}{\theta^T\Gamma\theta}\theta^T\Gamma\varepsilon\phi & \text{if } \theta^T\theta = M_0^2 \text{ and } (\Gamma\varepsilon\phi)^T\theta > 0. \end{cases}$$

For $\theta^T\theta = M_0^2$ and $(\Gamma\varepsilon\phi)^T\theta = \theta^T\Gamma\varepsilon\phi > 0$, we have $\operatorname{sgn}\{\frac{\tilde{\theta}^T\theta}{\theta^T\Gamma\theta}\theta^T\Gamma\varepsilon\phi\} = \operatorname{sgn}\{\tilde{\theta}^T\theta\}$. For $\theta^T\theta = M_0^2$, we have

$$\tilde{\theta}^T\theta = \theta^T\theta - \theta^{*T}\theta \geq M_0^2 - |\theta^*||\theta| = M_0(M_0 - |\theta^*|) \geq 0,$$

where the last inequality is obtained using the assumption that $M_0 \geq |\theta^*|$. Therefore, it follows that $\frac{\tilde{\theta}^T\theta\theta^T\Gamma\varepsilon\phi}{\theta^T\Gamma\theta} \geq 0$ when $\theta^T\theta = M_0^2$ and $(\Gamma\varepsilon\phi)^T\theta = \theta^T\Gamma\varepsilon\phi > 0$. Hence the term due to projection can only make $\dot{V}$ more negative, and therefore

$$\dot{V} = -\varepsilon^2 m_s^2 + \varepsilon\eta \leq -\frac{\varepsilon^2 m_s^2}{2} + \frac{\eta^2}{2m_s^2}.$$

Since V is bounded due to $\theta \in \mathcal{L}_\infty$, which is guaranteed by the projection, it follows that $\varepsilon m_s \in \mathcal{S}(\frac{\eta^2}{m_s^2})$, which implies that $\varepsilon \in \mathcal{S}(\frac{\eta^2}{m_s^2})$. From $\tilde{\theta} \in \mathcal{L}_\infty$ and $\frac{\phi}{m_s}, \frac{\eta}{m_s} \in \mathcal{L}_\infty$, we have $\varepsilon, \varepsilon m_s \in \mathcal{L}_\infty$. Now, for $\theta^T\theta = M_0^2$, we have $\frac{\|\Gamma\theta\theta^T\|}{\theta^T\Gamma\theta} \leq c$ for some constant $c \geq 0$, which implies that

$$|\dot{\theta}| \leq c|\varepsilon\phi| \leq c|\varepsilon m_s|.$$

Hence $\dot{\theta} \in \mathcal{S}(\frac{\eta^2}{m_s^2})$, and the proof of (i) and (ii) is complete. The proof of (iii) follows by setting $\eta = 0$ and has already been established in section 3.10.

The proof for parameter error convergence is completed as follows: Define the function

$$f \triangleq \begin{cases} \frac{\theta^T\Gamma\varepsilon\phi}{\theta^T\Gamma\theta} & \text{if } \theta^T\theta = M_0^2 \text{ and } (\Gamma\varepsilon\phi)^T\theta > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.99}$$

It is clear from the analysis above that $f(t) \geq 0 \ \forall t \geq 0$. Then (3.98) may be written as

$$\dot{\theta} = \Gamma\varepsilon\phi - \Gamma f\theta. \tag{3.100}$$

We can establish that $f\tilde{\theta}^T\theta$ has properties very similar to $\sigma_s\tilde{\theta}^T\theta$, i.e.,

$$f|\tilde{\theta}||\theta| \geq f\tilde{\theta}^T\theta \geq f|\theta|(M_0 - |\theta^*|), \quad f \geq 0,$$

and $|f(t)| \leq f_0\ \forall t \geq 0$ for some constant $f_0 \geq 0$. Therefore, the proof of (iv)(a)–(b) can be completed following exactly the same procedure as in the proof of Theorem 3.12.2. □

The gradient algorithm with projection has properties identical to those of the switching σ-modification, as both modifications aim at keeping $|\theta| \leq M_0$. In the case of the switching σ-modification, $|\theta|$ may exceed M_0 but remain bounded, whereas in the case of projection $|\theta| \leq M_0\ \forall t \geq 0$, provided $|\theta(0)| \leq M_0$.

3.12.4 Dead Zone

Let us consider the estimation error

$$\varepsilon = \frac{z - \theta^T\phi}{m_s^2} = \frac{-\tilde{\theta}^T\phi + \eta}{m_s^2} \tag{3.101}$$

for the parametric model

$$z = \theta^{*T}\phi + \eta.$$

The signal ε is used to "drive" the adaptive law in the case of the gradient and LS algorithms. It is a measure of the parameter error $\tilde{\theta}$, which is present in the signal $\tilde{\theta}^T\phi$, and of the modeling error η. When $\eta = 0$ and $\tilde{\theta} = 0$, we have $\varepsilon = 0$ and no adaptation takes place. Since $\frac{\eta}{m_s}, \frac{\phi}{m_s} \in \mathcal{L}_\infty$, large εm_s implies that $\frac{\tilde{\theta}^T\phi}{m_s}$ is large, which in turn implies that $\tilde{\theta}$ is large. In this case, the effect of the modeling error η is small, and the parameter estimates driven by ε move in a direction which reduces $\tilde{\theta}$. When εm_s is small, however, the effect of η on εm_s may be more dominant than that of the signal $\tilde{\theta}^T\phi$, and the parameter estimates may be driven in a direction dictated mostly by η. The principal idea behind the dead zone is to monitor the size of the estimation error and adapt only when the estimation error is large relative to the modeling error η, as shown below.

We consider the gradient algorithm for the linear parametric model (3.85). We consider the same cost function as in the ideal case, i.e.,

$$J(\theta, t) = \frac{\varepsilon^2 m_s^2}{2} = \frac{(z - \theta^T\phi)^2}{2m_s^2},$$

and write

$$\dot{\theta} = \begin{cases} -\Gamma\nabla J(\theta) & \text{if } |\varepsilon m_s| > g_0 > \frac{|\eta|}{m_s}, \\ 0 & \text{otherwise,} \end{cases} \tag{3.102}$$

where g_0 is a known upper bound of the normalized modeling error $\frac{|\eta|}{m_s}$. In other words, we move in the direction of the steepest descent only when the estimation error is large relative to the modeling error, i.e., when $|\varepsilon m_s| > g_0$. In view of (3.102) we have

$$\dot{\theta} = \Gamma\phi(\varepsilon + g), \quad g = \begin{cases} 0 & \text{if } |\varepsilon m_s| > g_0, \\ -\varepsilon & \text{if } |\varepsilon m_s| \leq g_0. \end{cases}$$

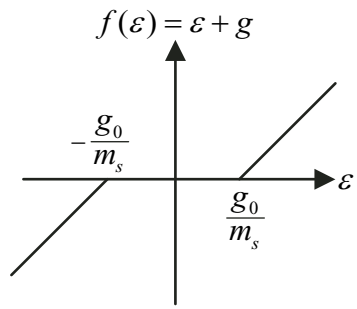

Figure 3.1. *Normalized dead zone function.*

To avoid any implementation problems which may arise due to the discontinuity in (3.102), the dead zone function is made continuous as shown in Figure 3.1; i.e.,

$$\dot{\theta} = \Gamma\phi(\varepsilon + g), \quad g = \begin{cases} \frac{g_0}{m_s} & \text{if } \varepsilon m_s < -g_0, \\ -\frac{g_0}{m_s} & \text{if } \varepsilon m_s > g_0, \\ -\varepsilon & \text{if } -g_0 \le \varepsilon m_s \le g_0. \end{cases} \tag{3.103}$$

Since the size of the dead zone depends on m_s, this dead zone function is often referred to as the *variable* or *relative dead zone*.

Theorem 3.12.4. *The adaptive law* (3.103) *guarantees the following properties:*

(i) $\varepsilon, \varepsilon m_s, \theta, \dot{\theta} \in \mathcal{L}_\infty$.

(ii) $\varepsilon, \varepsilon m_s, \dot{\theta} \in \mathcal{S}(g_0 + \frac{\eta^2}{m_s^2})$.

(iii) $\dot{\theta} \in \mathcal{L}_1 \cap \mathcal{L}_2$.

(iv) $\lim_{t\to\infty} \theta(t) = \bar{\theta}$, *where* $\bar{\theta}$ *is a constant vector.*

(v) *If* $\frac{\phi}{m_s}$ *is PE with level* $\alpha_0 > 0$, *then* $\tilde{\theta}(t)$ *converges to the residual set*

$$D_d = \{\tilde{\theta} \in \mathcal{R}^n \mid |\tilde{\theta}| \le c(g_0 + \bar{\eta})\},$$

where $\bar{\eta} = \sup_t \frac{|\eta(t)|}{m_s(t)}$ *and* $c \ge 0$ *is a constant.*

Proof. The proof can be found in [56] and in the web resource [94]. □

The dead zone modification has the following properties:

- It guarantees that the estimated parameters always converge to a constant. This is important in adaptive control employing an adaptive law with dead zone because at steady state the gains of the adaptive controller are almost constant, leading to an LTI closed-loop system when the plant is LTI.

- As in the case of the fixed σ-modification, the ideal properties of the adaptive law are destroyed in an effort to achieve robustness. That is, if the modeling error term becomes zero, the ideal properties of the adaptive law cannot be recovered unless the dead zone modification is also removed.

The robust modifications that include leakage, projection, and dead zone are analyzed for the case of the gradient algorithm for the SPM with modeling error. The same modifications can be used in the case of LS and DPM, B-SPM, and B-DPM with modeling errors. For further details on these modifications the reader is referred to [56]. One important property of these modifications is that $\varepsilon, \varepsilon m_s, \dot{\theta} \in \mathcal{S}(\lambda_0 + \frac{\eta^2}{m_s^2})$ for some constant $\lambda_0 \geq 0$. This means that the estimation error and speed of adaptation are guaranteed to be of the order of the modeling error and design parameter in the mean square sense. In other words intervals of time could exist at steady state where the estimation error and speed of adaptation could assume values higher than the modeling error. This phenomenon is known as "bursting," and it may take considerable simulation time to appear [105]. The dead zone exhibits no bursting as the parameters converge to constant values. Bursting can be prevented in all the other modifications if they are combined with a small-size dead zone at steady state. The various combinations of modifications as well as the development of new ones and their evaluation using analysis and simulations are a good exercise for the reader.

3.13 State-Space Identifiers

Let us consider the state-space plant model

$$\dot{x} = A_p x + B_p u, \tag{3.104}$$

where $x \in \mathcal{R}^n$ is the state, $u \in \mathcal{R}^m$ is the input vector, and $A_p \in \mathcal{R}^{n\times n}$, $B_p \in \mathcal{R}^{n\times m}$ are unknown constant matrices. We assume that x, u are available for measurement. One way to estimate the elements of A_p, B_p online is to express (3.104) as a set of n scalar differential equations and then generate n parametric models with scalar outputs and apply the parameter estimation techniques covered in the previous sections. Another more compact way of estimating A_p, B_p is to develop online estimators based on an SSPM model for (3.104) as follows.

We express (3.104) in the form of the SSPM:

$$\dot{x} = A_m x + (A_p - A_m)x + B_p u,$$

where A_m is an arbitrary stable matrix. The estimation model is then formed as

$$\dot{\hat{x}} = A_m \hat{x} + (\hat{A}_p - A_m)x + \hat{B}_p u = A_m(\hat{x} - x) + \hat{A}_p x + \hat{B}_p u,$$

where $\hat{A}_p(t)$, $\hat{B}_p(t)$ are the estimates of A_p, B_p at time t, respectively. The above estimation model has been referred to as the *series-parallel model* in the literature [56]. The estimation error vector is defined as

$$\varepsilon = x - \hat{x} - (sI - A_m)^{-1}\varepsilon n_s^2$$

or

$$\varepsilon = x - \hat{x} - w,$$

$$\dot{w} = A_m w + \varepsilon n_s^2, \quad w(0) = 0,$$

where n_s^2 is the static normalizing signal designed to guarantee $\frac{x}{\sqrt{1+n_s^2}}, \frac{u}{\sqrt{1+n_s^2}} \in \mathcal{L}_\infty$. A straightforward choice for n_s is $n_s^2 = x^T x + u^T u$. It is clear that if the plant model (3.104) is stable and the input u is bounded, then n_s can be taken to be equal to zero.

It follows that the estimation error satisfies

$$\dot{\varepsilon} = A_m \varepsilon - \tilde{A}_p x - \tilde{B}_p u - \varepsilon n_s^2,$$

where $\tilde{A}_p \triangleq \hat{A}_p - A_p$, $\tilde{B}_p \triangleq \hat{B}_p - B_p$ are the parameter errors.

The adaptive law for generating $\hat{A}_p$, $\hat{B}_p$ is developed by considering the Lyapunov function

$$V(\varepsilon, \tilde{A}_p, \tilde{B}_p) = \varepsilon^T P \varepsilon + \operatorname{tr}\left(\frac{\tilde{A}_p^T P \tilde{A}_p}{\gamma_1}\right) + \operatorname{tr}\left(\frac{\tilde{B}_p^T P \tilde{B}_p}{\gamma_2}\right), \tag{3.105}$$

where $\operatorname{tr}(A)$ denotes the trace of matrix A; $\gamma_1, \gamma_2 > 0$ are constant scalars; and $P = P^T > 0$ is chosen as the solution of the Lyapunov equation

$$P A_m + A_m P^T = -Q \tag{3.106}$$

for some $Q = Q^T > 0$, whose solution is guaranteed by the stability of A_m (see the Appendix). The time derivative $\dot{V}$ is given by

$$\dot{V} = \dot{\varepsilon}^T P \varepsilon + \varepsilon^T P \dot{\varepsilon} + \operatorname{tr}\left(\frac{\dot{\tilde{A}}_p^T P \tilde{A}_p}{\gamma_1} + \frac{\tilde{A}_p^T P \dot{\tilde{A}}_p}{\gamma_1}\right) + \operatorname{tr}\left(\frac{\dot{\tilde{B}}_p^T P \tilde{B}_p}{\gamma_2} + \frac{\tilde{B}_p^T P \dot{\tilde{B}}_p}{\gamma_2}\right).$$

Substituting for $\dot{\varepsilon}$, using (3.106), and employing the equalities $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$ and $\operatorname{tr}(A^T) = \operatorname{tr}(A)$ for square matrices A, B of the same dimension, we obtain

$$\dot{V} = -\varepsilon^T Q \varepsilon - 2\varepsilon^T P \tilde{A}_p x - 2\varepsilon^T P \tilde{B}_p u + 2 \operatorname{tr}\left(\frac{\tilde{A}_p^T P \dot{\tilde{A}}_p}{\gamma_1} + \frac{\tilde{B}_p^T P \dot{\tilde{B}}_p}{\gamma_2}\right) - 2\varepsilon^T P \varepsilon n_s^2. \tag{3.107}$$

Using the equality $v^T y = \operatorname{tr}(v y^T)$ for vectors v, y of the same dimension, we rewrite (3.107) as

$$\dot{V} = -\varepsilon^T Q \varepsilon + 2 \operatorname{tr}\left(\frac{\tilde{A}_p^T P \dot{\tilde{A}}_p}{\gamma_1} - \tilde{A}_p^T P \varepsilon x^T + \frac{\tilde{B}_p^T P \dot{\tilde{B}}_p}{\gamma_2} - \tilde{B}_p^T P \varepsilon u^T\right) - 2\varepsilon^T P \varepsilon n_s^2. \tag{3.108}$$

The obvious choice for $\dot{\hat{A}}_p = \dot{\tilde{A}}_p$, $\dot{\hat{B}}_p = \dot{\tilde{B}}_p$ to make $\dot{V}$ negative is

$$\dot{\hat{A}}_p = \gamma_1 \varepsilon x^T, \qquad \dot{\hat{B}}_p = \gamma_2 \varepsilon u^T, \tag{3.109}$$

which gives us

$$\dot{V} = -\varepsilon^T Q \varepsilon - 2\varepsilon^T P \varepsilon n_s^2 \le 0.$$

This implies that V, $\hat{A}_p$, $\hat{B}_p$, ε are bounded. We can also write

$$\dot{V} \leq -|\varepsilon|^2\lambda_{\min}(Q) - 2|\varepsilon n_s|^2\lambda_{\min}(P) \leq 0$$

and use similar arguments as in the previous sections to establish that $\varepsilon, \varepsilon n_s \in \mathcal{L}_2$. From (3.109) we have that

$$\|\dot{\hat{A}}_p\| \leq \gamma_1|\varepsilon m_s|\frac{|x^T|}{m_s}, \qquad \|\dot{\hat{B}}_p\| \leq \gamma_2|\varepsilon m_s|\frac{|u^T|}{m_s},$$

where $m_s^2 = 1 + n_s^2$. Since $\frac{|x|}{m_s}, \frac{|u|}{m_s} \in \mathcal{L}_\infty$ and $|\varepsilon m_s| \in \mathcal{L}_2$, we can also conclude that $\|\dot{\hat{A}}_p\|, \|\dot{\hat{B}}_p\| \in \mathcal{L}_2$.

We have, therefore, established that independent of the stability of the plant and boundedness of the input u, the adaptive law (3.109) guarantees that

- $\|\hat{A}_p\|, \|\hat{B}_p\|, \varepsilon \in \mathcal{L}_\infty$,
- $\|\dot{\hat{A}}_p\|, \|\dot{\hat{B}}_p\|, \varepsilon, \varepsilon n_s \in \mathcal{L}_2$.

These properties are important for adaptive control where the adaptive law is used as part of the controller and no a priori assumptions are made about the stability of the plant and boundedness of the input. If the objective, however, is parameter estimation, then we have to assume that the plant is stable and the input u is designed to be bounded and sufficiently rich for the plant model (3.104). In this case, we can take $n_s^2 = 0$.

Theorem 3.13.1. *Consider the plant model* (3.104) *and assume that* (A_p, B_p) *is controllable and* A_p *is a stable matrix. If each element* u_i, $i = 1, 2, \ldots, m$, *of vector* u *is bounded, sufficiently rich of order* $n+1$, *and uncorrelated, i.e., each* u_i *contains different frequencies, then* $\hat{A}_p(t)$, $\hat{B}_p(t)$ *generated by* (3.109) (*where* n_s *can be taken to be zero*) *converge to* A_p, B_p, *respectively, exponentially fast.*

The proof of Theorem 3.13.1 is long, and the reader is referred to [56] for the details.

Example 3.13.2 Consider the second-order plant

$$\dot{x} = Ax + Bu,$$

where $x = [x_1, x_2]^T$, $u = [u_1, u_2]^T$ is a bounded input vector, the matrices A, B are unknown, and A is a stable matrix. The estimation model is generated as

$$\dot{\hat{x}} = \begin{bmatrix} -a_m & 0 \\ 0 & -a_m \end{bmatrix}(\hat{x} - x) + \begin{bmatrix} \hat{a}_{11} & \hat{a}_{12} \\ \hat{a}_{21} & \hat{a}_{22} \end{bmatrix} x + \begin{bmatrix} \hat{b}_{11} & \hat{b}_{12} \\ \hat{b}_{21} & \hat{b}_{22} \end{bmatrix} u,$$

where $\hat{x} = [\hat{x}_1, \hat{x}_2]^T$ and $a_m > 0$. The estimation error is given by

$$\varepsilon = x - \hat{x},$$

where $n_s = 0$ due to the stability of A and boundedness of u. The adaptive law (3.109) can be written as

$$\dot{\hat{a}}_{ij} = \gamma_1\varepsilon_i x_j, \qquad \dot{\hat{b}}_{ij} = \gamma_2\varepsilon_i u_j$$

for $i = 1, 2$, $j = 1, 2$, and adaptive gains γ_1, γ_2. An example of a sufficiently rich input for this plant is

$$\begin{aligned} u_1 &= c_1 \sin 2.5t + c_2 \sin 4.6t, \\ u_2 &= c_3 \sin 7.2t + c_4 \sin 11.7t \end{aligned}$$

for some nonzero constants c_i, $i = 1, 2, 3, 4$. ∎

The class of plants described by (3.104) can be expanded to include more realistic plants with modeling errors. The adaptive laws in this case can be made robust by using exactly the same techniques as in the case of SISO plants described in previous sections, and this is left as an exercise for the reader.

3.14 Adaptive Observers

Consider the LTI SISO plant

$$\begin{aligned} \dot{x} &= Ax + Bu, \quad x(0) = x_0, \\ y &= C^T x, \end{aligned} \tag{3.110}$$

where $x \in \mathcal{R}^n$. We assume that u is a piecewise continuous bounded function of time and that A is a stable matrix. In addition, we assume that the plant is completely controllable and completely observable. The problem is to construct a scheme that estimates both the plant parameters, i.e., A, B, C, as well as the state vector x using only I/O measurements. We refer to such a scheme as the *adaptive observer*.

A good starting point for designing an adaptive observer is the Luenberger observer used in the case where A, B, C are known. The Luenberger observer is of the form

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + Bu + K(y - \hat{y}), \quad \hat{x}(0) = \hat{x}_0, \\ \hat{y} &= C^T \hat{x}, \end{aligned} \tag{3.111}$$

where K is chosen so that $A - KC^T$ is a stable matrix, and guarantees that $\hat{x} \to x$ exponentially fast for any initial condition x_0 and any input u. For $A - KC^T$ to be stable, the existence of K is guaranteed by the observability of (A, C).

A straightforward procedure for choosing the structure of the adaptive observer is to use the same equation as the Luenberger observer (3.111), but replace the unknown parameters A, B, C with their estimates $\hat{A}$, $\hat{B}$, $\hat{C}$, respectively, generated by some adaptive law. The problem we face with this procedure is the inability to estimate uniquely the n^2+2n parameters of A, B, C from the I/O data. The best we can do in this case is to estimate the parameters of the plant transfer function and use them to calculate $\hat{A}$, $\hat{B}$, $\hat{C}$. These calculations, however, are not always possible because the mapping of the $2n$ estimated parameters of the transfer function to the n^2+2n parameters of $\hat{A}$, $\hat{B}$, $\hat{C}$ is not unique unless (A, B, C) satisfies certain structural constraints. One such constraint is that (A, B, C) is in

the observer form, i.e., the plant is represented as

$$\begin{aligned} \dot{x}_\alpha &= \left[\begin{array}{c|c} -a_p & \begin{array}{c} I_{n-1} \\ \hline 0 \end{array} \end{array}\right] x_\alpha + b_p u, \\ y &= [1, 0, \ldots, 0] x_\alpha, \end{aligned} \tag{3.112}$$

where $a_p = [a_{n-1}, a_{n-2}, \ldots, a_0]^T$ and $b_p = [b_{n-1}, b_{n-2}, \ldots, b_0]^T$ are vectors of dimension n, and $I_{n-1} \in \mathcal{R}^{(n-1)\times(n-1)}$ is the identity matrix. The elements of a_p and b_p are the coefficients of the denominator and numerator, respectively, of the transfer function

$$\frac{y(s)}{u(s)} = \frac{b_{n-1}s^{n-1} + b_{n-2}s^{n-2} + \cdots + b_0 s}{s^n + a_{n-1}s^{n-1} + a_{n-2}s^{n-2} + \cdots + a_0 s} \tag{3.113}$$

and can be estimated online from I/O data using the techniques presented in the previous sections.

Since both (3.110) and (3.112) represent the same plant, we can focus on the plant representation (3.112) and estimate x_α instead of x. The disadvantage is that in a practical situation x may represent some physical variables of interest, whereas x_α may be an artificial state vector.

The adaptive observer for estimating the state x_α of (3.112) is motivated from the Luenberger observer structure (3.111) and is given by

$$\begin{aligned} \dot{\hat{x}} &= \hat{A}(t)\hat{x} + \hat{b}_p(t)u + K(t)(y - \hat{y}), \quad \hat{x}(0) = \hat{x}_0, \\ \hat{y} &= [1, 0, \ldots, 0]\hat{x}, \end{aligned} \tag{3.114}$$

where $\hat{x}$ is the estimate of x_α,

$$\hat{A}(t) = \left[\begin{array}{c|c} -\hat{a}_p(t) & \begin{array}{c} I_{n-1} \\ \hline 0 \end{array} \end{array}\right], \qquad K(t) = a^* - \hat{a}_p(t),$$

$\hat{a}_p(t)$ and $\hat{b}_p(t)$ are the estimates of the vectors a_p and b_p, respectively, at time t, and $a^* \in \mathcal{R}^n$ is chosen so that

$$A^* = \left[\begin{array}{c|c} -a^* & \begin{array}{c} I_{n-1} \\ \hline 0 \end{array} \end{array}\right] \tag{3.115}$$

is a stable matrix that contains the eigenvalues of the observer.

A wide class of adaptive laws may be used to generate $\hat{a}_p(t)$ and $\hat{b}_p(t)$ online. As in Chapter 2, the parametric model

$$z = \theta^{*T}\phi \tag{3.116}$$

may be developed using (3.113), where

$$\theta^* = [b_p^T, a_p^T]^T$$

and z, ϕ are available signals, and used to design a wide class of adaptive laws to generate $\theta(t) = [\hat{b}_p^T(t), a_p^T(t)]^T$, the estimate of θ^*. As an example, consider the gradient algorithm

$$\dot{\theta} = \Gamma\varepsilon\phi, \quad \varepsilon = \frac{z - \theta^T\phi}{m_s^2}, \tag{3.117}$$

where $m_s^2 = 1 + \alpha\phi^T\phi$ and $\alpha \geq 0$.

Theorem 3.14.1. *The adaptive observer described by* (3.114)–(3.117) *guarantees the following properties:*

(i) *All signals are bounded.*

(ii) *If u is sufficiently rich of order $2n$, then the state observation error $|\hat{x} - x_\alpha|$ and the parameter error $|\theta - \theta^*|$ converge to zero exponentially fast.*

Proof. (i) Since A is stable and u is bounded, we have $x_\alpha, y, \phi \in \mathcal{L}_\infty$ and hence $m_s^2 = 1+\alpha\phi^T\phi \in \mathcal{L}_\infty$. The adaptive law (3.117) guarantees that $\theta \in \mathcal{L}_\infty$ and $\varepsilon, \varepsilon m_s, \dot{\theta} \in \mathcal{L}_2 \cap \mathcal{L}_\infty$. The observer equation may be written as

$$\dot{\hat{x}} = A^*\hat{x} + \hat{b}_p(t)u + (\hat{A}(t) - A^*)x_\alpha.$$

Since A^* is a stable matrix and $\hat{b}_p$, $\hat{A}$, u, x_α are bounded, it follows that $\hat{x} \in \mathcal{L}_\infty$, which in turn implies that all signals are bounded.

(ii) The state observation error $\tilde{x} = \hat{x} - x_\alpha$ satisfies

$$\dot{\tilde{x}} = A^*\tilde{x} + \tilde{b}_p u - \tilde{a}_p y, \tag{3.118}$$

where $\tilde{b}_p = \hat{b}_p - b_p$, $\tilde{a}_p = \hat{a}_p - a_p$ are the parameter errors. Since for u sufficiently rich we have that $\theta(t) \to \theta^*$ as $t \to \infty$ exponentially fast, it follows that $\tilde{b}_p, \tilde{a}_p \to 0$ exponentially fast. Since $u, y \in \mathcal{L}_\infty$, the error equation consists of a homogenous part that is exponentially stable and an input that is decaying to zero. This implies that $\tilde{x} = \hat{x} - x_\alpha \to 0$ as $t \to \infty$ exponentially fast. □

For further reading on adaptive observers, the reader is referred to [56, 86, 108–110].

3.15 Case Study: Users in a Single Bottleneck Link Computer Network

The congestion control problem in computer networks has been identified as a feedback control problem. The network users adjust their sending data rates, in response to congestion signals they receive from the network, in an effort to avoid congestion and converge to a stable equilibrium that satisfies certain requirements: high network utilization, small queue sizes, small delays, fairness among users, etc. Many of the proposed congestion control schemes require that at each link the number of flows, N say, utilizing the link is known. Since the number of users varies with time, N is an unknown time-varying parameter,

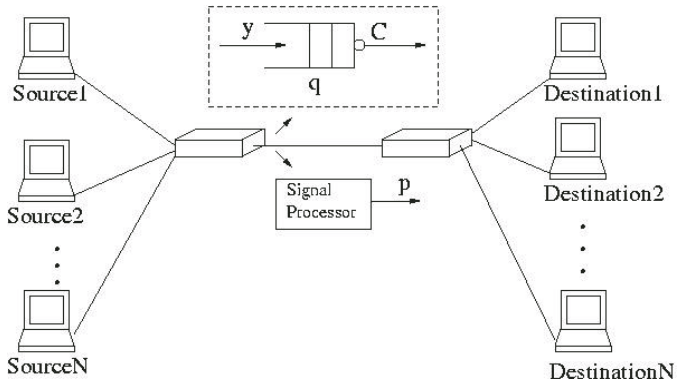

Figure 3.2. *Network topology.*

which needs to be estimated online. Estimation algorithms, which have been proposed in the literature, are based on pointwise time division, which is known to lack robustness and may lead to erroneous estimates. In this study, we consider a simple estimation algorithm, which is based on online parameter identification.

We consider the single bottleneck link network shown in Figure 3.2. It consists of N users which share a common bottleneck link through high bandwidth access links. At the bottleneck link we assume that there exists a buffer, which accommodates the incoming packets. The rate of data entering the buffer is denoted by y, the queue size is denoted by q, and the output capacity is denoted by C. At the bottleneck link, we implement a signal processor, which calculates the desired sending rate p. This information is communicated to the network users, which set their sending rate equal to p. The desired sending rate p is updated according to the following control law:

$$\dot{p} = \begin{cases} \frac{1}{\hat{N}}[k_i(C-y) - k_q q] & \text{if } 1 < p < C, \\ \frac{1}{\hat{N}}[k_i(C-y) - k_q q] & \text{if } p = 1,\ \frac{1}{\hat{N}}[k_i(C-y) - k_q q] > 0, \\ \frac{1}{\hat{N}}[k_i(C-y) - k_q q] & \text{if } p = C,\ \frac{1}{\hat{N}}[k_i(C-y) - k_q q] < 0, \\ 0 & \text{otherwise,} \end{cases} \tag{3.119}$$

where $\hat{N}$ is an estimate of N which is calculated online and k_i, k_q are design parameters. Since N is changing with time, its estimate $\hat{N}$ has to be updated accordingly. In this study we use online parameter estimation to generate $\hat{N}$. Since the sending rate of all users is equal to p, it follows that

$$y = Np. \tag{3.120}$$

Since y and p are measured at the bottleneck link, (3.120) is in the form of an SPM with N as the unknown parameter. We also know that N cannot be less than 1. Using the

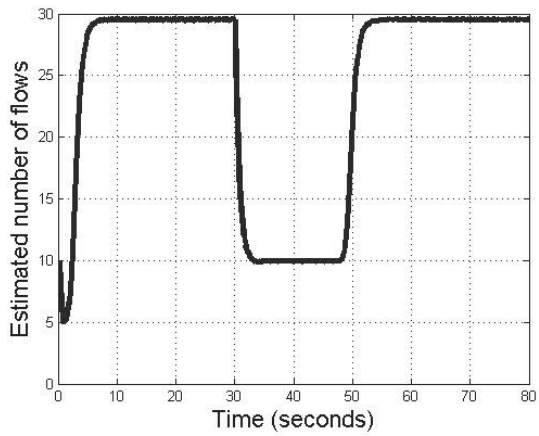

Figure 3.3. *Time response of the estimate of the number of flows.*

results of the chapter, we propose the following online parameter estimator:

$$\dot{\hat{N}} = \begin{cases} \gamma \varepsilon p & \text{if } \hat{N} > 1 \text{ or } \hat{N} = 1 \text{ and } \varepsilon p \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$
$$\varepsilon = \frac{y - \hat{N} p}{1 + p^2}, \tag{3.121}$$

where $\hat{N}(0) \geq 1$. We demonstrate the effectiveness of the proposed algorithm using simulations, which we conduct on the packet-level simulator ns-2. We consider the network topology of Figure 3.2 in our simulations. The bandwidth of the bottleneck link is set to 155 Mb/s, and the propagation delay of each link is set to 20 ms. The design parameters are chosen as follows: $\gamma = 0.1$, $k_i = 0.16$, $k_q = 0.32$. Initially 30 users utilize the network. The estimator starting with an initial estimate of 10 converges to 30. After $t = 30$ seconds 20 of these users stop sending data, while an additional 20 users enter the network at $t = 45$ seconds. The output of the estimator at the bottleneck link is shown in Figure 3.3. We observe that the estimator accurately tracks the number of flows utilizing the network. In addition we observe good transient behavior as the responses are characterized by fast convergence and no overshoots. The estimator results are obtained in the presence of noise and delays which were not included in the simple model (3.120). Since the number of parameters to be estimated is 1, the PE property is satisfied for $p \neq 0$, which is always the case in this example.

See the web resource [94] for examples using the Adaptive Control Toolbox.

Problems

1. Consider the SPM

$$z = \theta^{*T} \phi$$

and the estimation model

$$\hat{z} = \theta^T(t)\phi.$$

Find values for $\theta(t)$, $\phi(t)$ such that $z = \hat{z}$ but $\theta(t) \neq \theta^*$.

2. Consider the adaptive law

$$\dot{\theta} = \gamma\varepsilon\phi,$$

where $\theta, \phi \in \mathcal{R}^n$, $\frac{\phi}{m_s} \in \mathcal{L}_\infty$, $\varepsilon m_s \in \mathcal{L}_2 \cap \mathcal{L}_\infty$, and $m_s \geq 1$. Show that $\dot{\theta} \in \mathcal{L}_2 \cap \mathcal{L}_\infty$.

3. Show that if $u \in \mathcal{L}_\infty$ in (3.1), then the adaptive law (3.10) with $m_s^2 = 1 + \alpha\phi^2$, $\alpha \geq 0$, guarantees that $\varepsilon, \varepsilon m_s, \dot{\theta} \in \mathcal{L}_2 \cap \mathcal{L}_\infty$ and that $\varepsilon(t), \varepsilon(t)m_s(t), \dot{\theta}(t) \to 0$ as $t \to \infty$.

4. (a) Show that (3.16) is a necessary and sufficient condition for $\theta(t)$ in the adaptive law (3.10) to converge to θ^* exponentially fast.

   (b) Establish which of the following choices for u guarantee that ϕ in (3.10) is PE:

      (i) $u = c_0 \neq 0$, c_0 is a constant.
      (ii) $u = \sin t$.
      (iii) $u = \sin t + \cos 2t$.
      (iv) $u = \frac{1}{1+t}$.
      (v) $u = e^{-t}$.
      (vi) $u = \frac{1}{(1+t)^{1/2}}$.

   (c) In (b), is there a choice for u that guarantees that $\theta(t)$ converges to θ^* but does not guarantee that ϕ is PE?

5. Use the plant model (3.24) to develop the bilinear parametric model (3.28). Show all the steps.

6. In Theorem 3.6.1, assume that $\phi, \dot{\phi} \in \mathcal{L}_\infty$. Show that the adaptive law (3.34) with $m_s^2 = 1 + n_s^2$, $n_s^2 = \alpha\phi^T\phi$, and $\alpha \geq 0$ guarantees that $\varepsilon(t), \varepsilon(t)m_s(t), \dot{\theta}(t) \to 0$ as $t \to \infty$.

7. Consider the SPM $z = \theta^{*T}\phi$ and the cost function

$$J(\theta) = \frac{1}{2}\int_0^t e^{-\beta(t-\tau)}\frac{(z(\tau) - \theta^T(t)\phi(\tau))^2}{m_s^2(\tau)}d\tau,$$

where $m_s^2 = 1 + \phi^T\phi$ and $\theta(t)$ is the estimate of θ^* at time t.

   (a) Show that the minimization of $J(\theta)$ w.r.t. θ using the gradient method leads to the adaptive law

$$\dot{\theta}(t) = \Gamma\int_0^t e^{-\beta(t-\tau)}\frac{z(\tau) - \theta^T(t)\phi(\tau)}{m_s^2(\tau)}\phi(\tau)d\tau, \quad \theta(0) = \theta_0.$$

   (b) Show that the adaptive law in part (a) can be implemented as

$$\dot{\theta}(t) = -\Gamma(R(t)\theta(t) + Q(t)), \quad \theta(0) = \theta_0,$$
$$\dot{R}(t) = -\beta R(t) + \frac{\phi(t)\phi^T(t)}{m_s^2(t)}, \quad R(0) = 0,$$

$$\dot{Q}(t) = -\beta Q(t) - \frac{z(t)\phi(t)}{m_s^2(t)}, \qquad Q(0) = 0,$$

which is referred to as the integral adaptive law.

8. Consider the second-order stable system

$$\dot{x} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & 0 \end{bmatrix} x + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} u,$$

where x, u are available for measurement, $u \in \mathcal{L}_\infty$, and a_{11}, a_{12}, a_{21}, b_1, b_2 are unknown parameters. Design an online estimator to estimate the unknown parameters. Simulate your scheme using $a_{11} = -0.25$, $a_{12} = 3$, $a_{21} = -5$, $b_1 = 1$, $b_2 = 2.2$, and $u = 10 \sin 2t$. Repeat the simulation when $u = 10 \sin 2t + 7 \cos 3.6t$. Comment on your results.

9. Consider the nonlinear system

$$\dot{x} = a_1 f_1(x) + a_2 f_2(x) + b_1 g_1(x) u + b_2 g_2(x) u,$$

where $u, x \in \mathcal{R}$; f_i, g_i are known nonlinear functions of x; and a_i, b_i are unknown constant parameters and $i = 1, 2$. The system is such that $u \in \mathcal{L}_\infty$ implies $x \in \mathcal{L}_\infty$. Assuming that x, u can be measured at each time t, design an estimation scheme for estimating the unknown parameters online.

10. Design and analyze an online estimation scheme for estimating θ^* in (3.49) when $L(s)$ is chosen so that $W(s)L(s)$ is biproper and SPR.

11. Design an online estimation scheme to estimate the coefficients of the numerator polynomial

$$Z(s) = b_{n-1}s^{n-1} + b_{n-2}s^{n-2} + \cdots + b_1 s + b_0$$

of the plant

$$y = \frac{Z(s)}{R(s)} u$$

when the coefficients of $R(s) = s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0$ are known. Repeat the same problem when $Z(s)$ is known and $R(s)$ is unknown.

12. Consider the mass–spring–damper system shown in Figure 3.4, where β is the damping coefficient, k is the spring constant, u is the external force, and $y(t)$ is the displacement of the mass m resulting from the force u.

    (a) Verify that the equations of motion that describe the dynamic behavior of the system under small displacements are

    $$m\ddot{y} + \beta\dot{y} + ky = u.$$

    (b) Design a gradient algorithm to estimate the constants m, β, k when y, u can be measured at each time t.

    (c) Repeat (b) for an LS algorithm.

    (d) Simulate your algorithms in (b) and (c) assuming $m = 20$ kg, $\beta = 0.1$ kg/sec, $k = 5$ kg/sec^2, and inputs u of your choice.

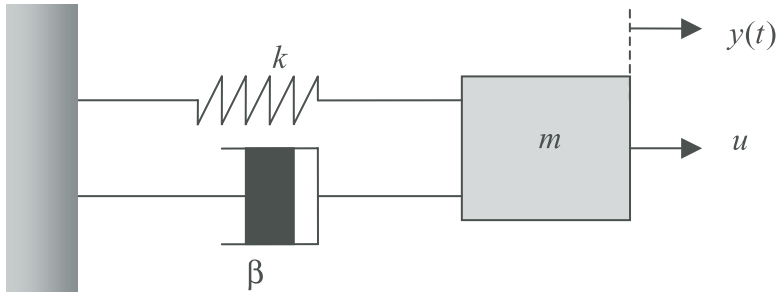

Figure 3.4. *The mass–spring–damper system for Problem* 12.

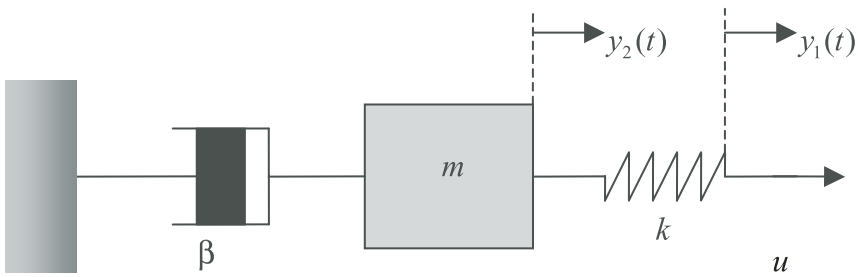

Figure 3.5. *The mass–spring–damper system for Problem* 13.

(e) Repeat (d) when $m = 20$ kg for $0 \le t \le 20$ sec and $m = 20(2 - e^{-0.01(t-20)})$ kg for $t \ge 20$ sec.

13. Consider the mass–spring–damper system shown in Figure 3.5.

(a) Verify that the equations of motion are given by

$$k(y_1 - y_2) = u,$$
$$k(y_1 - y_2) = m\ddot{y}_2 + \beta\dot{y}_2.$$

(b) If y_1, y_2, u can be measured at each time t, design an online parameter estimator to estimate the constants k, m, and β.

(c) We have the a priori knowledge that $0 \le \beta \le 1$, $k \ge 0.1$, and $m \ge 10$. Modify your estimator in (b) to take advantage of this a priori knowledge.

(d) Simulate your algorithm in (b) and (c) when $\beta = 0.2$ kg/sec, $m = 15$ kg, $k = 2$ kg/sec^2, and $u = 5 \sin 2t + 10.5$ kg $\cdot$ m/sec^2.

14. Consider the block diagram of a steer-by-wire system of an automobile shown in Figure 3.6, where r is the steering command in degrees, θ_p is the pinion angle in degrees, and $\dot{\theta}$ is the yaw rate in degree/sec.
The transfer functions $G_0(s)$, $G_1(s)$ are of the form

$$G_0(s) = \frac{k_0\omega_0^2}{s^2 + 2\xi_0\omega_0 s + \omega_0^2(1 - k_0)},$$

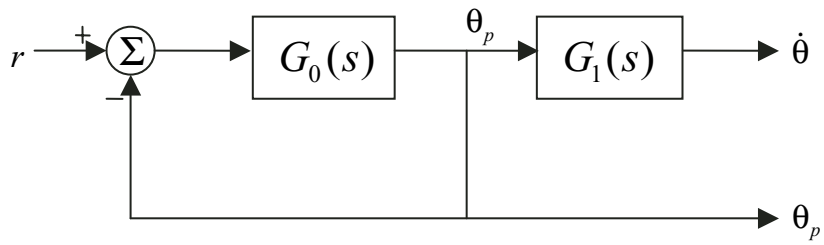

Figure 3.6. *Block diagram of a steer-by-wire system for Problem* 14.

$$G_1(s) = \frac{k_1\omega_1^2}{s^2 + 2\xi_1\omega_1 s + \omega_1^2},$$

where k_0, ω_0, ξ_0, k_1, ω_1, ξ_1 are functions of the speed of the vehicle. Assuming that r, θ_p, $\dot{\theta}$ can be measured at each time t, do the following:

(a) Design an online parameter estimator to estimate k_i, ω_i, ξ_i, $i = 0, 1$, using the measurements of θ_p, $\dot{\theta}$, r.

(b) Consider the values of the parameters shown in the following table at different speeds:

Speed V	k_0	ω_0	ξ_0	k_1	ω_1	ξ_1
30 mph	0.81	19.75	0.31	0.064	14.0	0.365
60 mph	0.77	19.0	0.27	0.09	13.5	0.505

Assume that between speeds the parameters vary linearly. Use these values to simulate and test your algorithm in (a) when

(i) $r = 10 \sin 0.2t + 8$ degrees and $V = 20$ mph.

(ii) $r = 5$ degrees and the vehicle speeds up from $V = 30$ mph to $V = 60$ mph in 40 seconds with constant acceleration and remains at 60 mph for 10 seconds.

15. Consider the equation of the motion of the mass–spring–damper system given in Problem 12, i.e.,

$$m\ddot{y} + \beta\dot{y} + ky = u.$$

This system may be written in the form

$$y = \rho^*(u - m\ddot{y} - \beta\dot{y}),$$

where $\rho^* = \frac{1}{k}$ appears in a bilinear form with the other unknown parameters m, β. Use the adaptive law based on the bilinear parametric model to estimate ρ^*, m, β when u, y are the only signals available for measurement. Since $k > 0$, the sign of ρ^* may be assumed to be known. Simulate your adaptive law using the numerical values given in (d) and (e) of Problem 12.

16. The effect of initial conditions on the SPM can be modeled as

$$z = \theta^{*T}\phi + \eta_0,$$

$$\dot{\omega}_0 = \Lambda\omega_0,$$
$$\eta_0 = C^T\omega_0,$$

where Λ is a transfer matrix with all poles in $\Re[s] < 0$ and $\omega_0 \in \mathcal{R}^n$, $\eta_0 \in \mathcal{R}$. Show that the properties of an adaptive law (gradient, LS, etc.) with $\eta_0 = 0$ are the same as those for the SPM with $\eta_0 \neq 0$.

17. Consider the system

$$y = e^{-\tau s}\frac{b}{(s-a)(\mu s+1)}u,$$

where $0 < \tau \ll 1$, $0 < \mu \ll 1$, and a, b, τ, μ are unknown constants. We want to estimate a, b online.

(a) Obtain a parametric model that can be used to design an adaptive law to estimate a, b.

(b) Design a robust adaptive law of your choice to estimate a, b online.

(c) Simulate your scheme for $a = -5$, $b = 100$, $\tau = 0.01$, $\mu = 0.001$, and different choices of the input signal u. Comment on your results.

18. The dynamics of a hard-disk drive servo system are given by

$$y = \frac{k_p}{s^2}\sum_{i=1}^{6}\frac{b_{1i}s + b_{0i}}{s^2 + 2\zeta_i\omega_i s + \omega_i^2}u,$$

where ω_i, $i = 1, \ldots, 6$, are the resonant frequencies which are large, i.e., $\omega_1 = 11.2\pi \times 10^3$ rad/sec, $\omega_2 = 15.5\pi \times 10^3$ rad/sec, $\omega_3 = 16.6\pi \times 10^3$ rad/sec, $\omega_4 = 18\pi \times 10^3$ rad/sec, $\omega_5 = 20\pi \times 10^3$ rad/sec, $\omega_6 = 23.8\pi \times 10^3$ rad/sec. The unknown constants b_{1i} are of the order of 10^4, b_{0i} are of the order of 10^8, and k_p is of the order of 10^7. The damping coefficients ζ_i are of the order of 10^{-2}.

(a) Derive a low-order model for the servo system. (*Hint*: Take $\frac{\alpha}{\omega_i} \cong 0$ and hence $\frac{\alpha^2}{\omega_i^2} \cong 0$ for α of the order of less than 10^3.)

(b) Assume that the full-order system parameters are given as follows:

$$\begin{aligned}
&b_{16} = -5.2 \times 10^4, && b_{01} = 1.2 \times 10^9, && b_{02} = 5.4 \times 10^8,\\
&b_{03} = -7.7 \times 10^8, && b_{04} = -1.6 \times 10^8, && b_{05} = -1.9 \times 10^8,\\
&b_{06} = 1.2 \times 10^9, && k_p = 3.4 \times 10^7, && \zeta_1 = 2.6 \times 10^{-2},\\
&\zeta_2 = 4.4 \times 10^{-3}, && \zeta_3 = 1.2 \times 10^{-2}, && \zeta_4 = 2.4 \times 10^{-3},\\
&\zeta_5 = 6.8 \times 10^{-3}, && \zeta_6 = 1.5 \times 10^{-2}.
\end{aligned}$$

Obtain a Bode plot for the full-order and reduced-order models.

(c) Use the reduced-order model in (a) to obtain a parametric model for the unknown parameters. Design a robust adaptive law to estimate the unknown parameters online.

19. Consider the time-varying plant

$$\dot{x} = -a(t)x + b(t)u,$$

where $a(t)$, $b(t)$ are slowly varying unknown parameters; i.e., $|\dot{a}|$, $|\dot{b}|$ are very small.

(a) Obtain a parametric model for estimating a, b.

(b) Design and analyze a robust adaptive law that generates the estimates $\hat{a}(t)$, $\hat{b}(t)$ of $a(t)$, $b(t)$, respectively.

(c) Simulate your scheme for a plant with $a(t) = 5 + \sin \mu t$, $b(t) = 8 + \cos 2\mu t$ for $\mu = 0, 0.01, 0.1, 1, 5$. Comment on your results.

20. Consider the parameter error differential equation (3.69), i.e.,

$$\dot{\tilde{\theta}} = -\gamma u^2 \tilde{\theta} + \gamma d u.$$

Show that if the equilibrium $\tilde{\theta}_e = 0$ of the homogeneous equation

$$\dot{\tilde{\theta}} = -\gamma u^2 \tilde{\theta}$$

is exponentially stable, then the bounded input $\gamma d u$ will lead to a bounded solution $\tilde{\theta}(t)$. Obtain an upper bound for $|\tilde{\theta}(t)|$ as a function of the upper bound of the disturbance term $\gamma d u$.

21. Consider the system

$$y = \theta^* u + \eta,$$
$$\eta = \Delta(s) u,$$

where y, u are available for measurement, θ^* is the unknown constant to be estimated, and η is a modeling error signal with $\Delta(s)$ being proper and analytic in $\Re[s] \geq -0.5$. The input u is piecewise continuous.

(a) Design an adaptive law with a switching σ-modification to estimate θ^*.

(b) Repeat (a) using projection.

(c) Simulate the adaptive laws in (a), (b) using the following values:

$$\theta^* = 5 + \sin 0.1t,$$
$$\Delta(s) = 10\mu \frac{\mu s - 1}{(\mu s + 1)^2}$$

for $\mu = 0, 0.1, 0.01$ and $u =$ constant, $u = \sin \omega_0 t$, where $\omega_0 = 1, 10, 100$. Comment on your results.

22. The linearized dynamics of a throttle angle θ to vehicle speed V subsystem are given by the third-order system

$$V = \frac{b p_1 p_2}{(s + a)(s + p_1)(s + p_2)} \theta + d,$$

where $p_1, p_2 > 20$, $1 \geq a > 0$, and d is a load disturbance.

(a) Obtain a parametric model for the parameters of the dominant part of the system.

(b) Design a robust adaptive law for estimating these parameters online.

(c) Simulate your estimation scheme when $a = 0.1$, $b = 1$, $p_1 = 50$, $p_2 = 100$, and $d = 0.02 \sin 5t$, for different constant and time-varying throttle angle settings θ of your choice.

23. Consider the parametric model

$$z = \theta^{*T}\phi + \eta,$$

where

$$\eta = \Delta_u(s)u + \Delta_y(s)y$$

and Δ_u, Δ_y are proper transfer functions analytic in $\Re[s] \geq -\frac{\delta_0}{2}$ for some known $\delta_0 > 0$.

(a) Design a normalizing signal m_s that guarantees $\frac{\eta}{m_s} \in \mathcal{L}_\infty$ when (i) Δ_u, Δ_y are biproper, (ii) Δ_u, Δ_y are strictly proper. In each case specify the upper bound for $\frac{|\eta|}{m_s}$.

(b) Calculate the bound for $\frac{|\eta|}{m_s}$ when (i) $\Delta_u(s) = \frac{e^{-\tau s}-1}{s+2}$, $\Delta_y(s) = \mu\frac{s^2}{(s+1)^2}$, (ii) $\Delta_u(s) = \frac{\mu s}{\mu s+2}$, $\Delta_y(s) = \frac{\mu s}{(\mu s+1)^2}$, where $0 < \mu \ll 1$ and $0 < \tau \ll 1$.

(c) Design and simulate a robust adaptive law to estimate θ^* for the following example:

$$\begin{aligned}
\theta^* &= [1, 0.1]^T, \\
\phi &= [u, y]^T, \\
z &= \frac{s}{s+5}y, \\
\Delta_u(s) &= \frac{e^{-\tau s}-1}{s+2}, \qquad \Delta_y(s) = \frac{\mu(s-1)}{(\mu s+1)^2},
\end{aligned}$$

where $\tau = 0.01$, $\mu = 0.01$.