$$R_{yu}(t) := \lim_{T \to \infty} \frac{1}{T} \int_{t_0}^{t_0 + T} y(\tau)\, u^T(t + \tau)\, d\tau \in \mathbb{R}^{m \times n} \qquad (1.6.19)$$

It may be verified that the relationship between $R_{uy}$ and $R_{yu}$ is

$$R_{yu}(t) = R_{uy}^T(-t) \qquad (1.6.20)$$

and the Fourier transform of the cross correlation is referred to as the *cross spectral measure* of $u$ and $y$.

The cross correlation between the input and the output of a stable LTI system can be related in much the same way as in proposition 1.6.2.

### Proposition 1.6.3  Linear Filter Lemma—Cross Correlation

Let $y = \hat{H}(u)$, where $\hat{H}$ is a proper stable $m \times n$ matrix transfer function, with impulse response $H(t)$.

*If*        $u$ is stationary

*Then*      the cross correlation between $u$ and $y$ is given by

$$R_{yu}(t) = \int_{-\infty}^{\infty} H(\tau_1) R_u(t + \tau_1)\, d\tau_1 \qquad (1.6.21)$$

and the cross spectral measure is

$$S_{yu}(d\omega) = \hat{H}^*(j\omega) S_u(d\omega) \qquad (1.6.22)$$

### Proof of Proposition 1.6.3

The proof is analogous to the proof of proposition 1.6.2 and is omitted here.

# CHAPTER 2
# IDENTIFICATION

## 2.0  INTRODUCTION

In this chapter, we review some identification methods for single-input single-output (SISO), linear time invariant (LTI) systems. To introduce the subject, we first informally discuss a simple example. We consider the identification problem for a first order SISO LTI system described by a transfer function

$$\frac{\hat{y}_p(s)}{\hat{r}(s)} = \hat{P}(s) = \frac{k_p}{s + a_p} \qquad (2.0.1)$$

The parameters $k_p$ and $a_p$ are unknown and are to be determined by the identification scheme on the basis of measurements of the input and output of the plant. The plant is assumed to be stable, i.e. $a_p > 0$.

### Frequency Domain Approach

A standard approach to identification is the *frequency response* approach. Let the input $r$ be a sinusoid

$$r(t) = \sin(\omega_0 t) \qquad (2.0.2)$$

The steady-state response is then given by

$$y(t) = m_p \sin(\omega_0 t + \phi) \qquad (2.0.3)$$

where

$$m = |\hat{P}(j\omega_0)| = \frac{k_p}{\sqrt{\omega_0^2 + a_p^2}}$$

$$\phi = \arg \hat{P}(j\omega_0) = -\arctan\left[\frac{\omega_0}{a_p}\right] \qquad (2.0.4)$$

Measurements of the gain $m$ and phase $\phi$ at a single frequency $\omega_0 \neq 0$ uniquely determine $k_p$ and $a_p$ by inversion of the above relationships. At $\omega_0 = 0$, that is when the input signal is constant, phase information is lost. Only one equation is left, giving the DC gain. Then, only the ratio of the parameters $k_p$ and $a_p$ is determined. Conversely, if several frequencies are used, each contributes two equations and the parameters are overdetermined.

Frequency response methods will not be further discussed in this book, because our goal is adaptive control. We will therefore concentrate on *recursive* approaches, where parameter estimates are updated in real-time. However, we will still *analyze* these algorithms in the frequency domain and obtain similar results as above for the recursive schemes.

**Time Domain Approach**

We now discuss schemes based on a time-domain expression of the plant (2.0.1), that is

$$\dot{y}_p(t) = -a_p y_p(t) + k_p r(t) \qquad (2.0.5)$$

Measurements of $y_p$, $\dot{y}_p$ and $r$ at *one* time instant $t$ give us one equation with two unknown $a_p$ and $k_p$. As few as two time instants may be sufficient to determine the unknown parameters from

$$\begin{bmatrix} -a_p \\ k_p \end{bmatrix} = \begin{bmatrix} y_p(t_1) & r(t_1) \\ y_p(t_2) & r(t_2) \end{bmatrix}^{-1} \begin{bmatrix} \dot{y}_p(t_1) \\ \dot{y}_p(t_2) \end{bmatrix} \qquad (2.0.6)$$

assuming that the inverse exists. Note that, as in the frequency-domain approach, a constant input $r(t_1) = r(t_2)$ with constant output $y_p(t_1) = y_p(t_2)$ will prevent us from uniquely determining $a_p$ and $k_p$.

We may refine this approach to avoid the measurement of $\dot{y}_p(t)$. Consider (2.0.1) and divide both sides by $s + \lambda$ for some $\lambda > 0$

$$\frac{s + a_p}{s + \lambda} \hat{y}_p = \frac{k_p}{s + \lambda} \hat{r} \qquad (2.0.7)$$

This is equivalent to

$$\hat{y}_p = \frac{\lambda - a_p}{s + \lambda} \hat{y}_p + \frac{k_p}{s + \lambda} \hat{r} \qquad (2.0.8)$$

Define the signals

$$\hat{w}^{(1)} = \frac{1}{s + \lambda} \hat{r} \qquad \hat{w}^{(2)} = \frac{1}{s + \lambda} \hat{y}_p \qquad (2.0.9)$$

or, in the time domain

$$\dot{w}^{(1)} = -\lambda w^{(1)} + r \qquad \dot{w}^{(2)} = -\lambda w^{(2)} + y_p \qquad (2.0.10)$$

Then, in the time domain, (2.0.8) reads

$$y_p(t) = k_p w^{(1)}(t) + (\lambda - a_p) w^{(2)}(t) \qquad (2.0.11)$$

The signals $w^{(1)}$, $w^{(2)}$ may be obtained by stable filtering of the input and of the output of the plant. We have assumed zero initial conditions on $w^{(1)}$ and $w^{(2)}$. Nonzero initial conditions would only contribute exponential decaying terms with rate $\lambda$ (arbitrary), but are not considered in this simplified derivation. Equation (2.0.11) is to be compared with (2.0.5). Again, measurements at one time instant give us one equation with two unknowns. However, we do not require differentiation, but instead stable filtering of available signals.

In the sequel, we will assume that measurements of $r$ and $y_p$ are made continuously between 0 and $t$. We will therefore look for algorithms that use the complete information and preferably update estimates only on the basis of new data, without storing the entire signals. But first, we transform (2.0.11) into the standard framework used later in this chapter. Define the vector of *nominal* identifier parameters

$$\theta^* := \begin{bmatrix} k_p \\ \lambda - a_p \end{bmatrix} \qquad (2.0.12)$$

Knowledge of $\theta^*$ is clearly equivalent to the knowledge of the unknown parameters $k_p$ and $a_p$. Similarly, define $\theta(t)$ to be a vector of identical dimension, called the *adaptive* identifier parameter. $\theta(t)$ is the estimate of $\theta^*$ based on input-output data up to time $t$. Letting

$$w(t) := \begin{bmatrix} w^{(1)}(t) \\ w^{(2)}(t) \end{bmatrix} \qquad (2.0.13)$$

equation (2.0.11) may be written

$$y_p(t) = \theta^{*T} w(t) = w^T(t) \theta^* \qquad (2.0.14)$$

Based on measurements of $r(t)$ and $y_p(t)$ up to time $t$, $w(t)$ may be

calculated, and an estimate $\theta(t)$ derived. Since each time instant gives us one equation with two unknowns, it makes sense to consider the estimate that minimizes the *identification error*

$$e_1(t) = \theta^T(t) w(t) - y_p(t) = \left[ \theta^T(t) - \theta^{*T} \right] w(t) \qquad (2.0.15)$$

Note that the identification error is linear in the parameter error $\theta - \theta^*$. We will therefore call (2.0.15) a *linear error equation*. The purpose of the identification scheme will be to calculate $\theta(t)$, on the basis of measurements of $e_1(t)$ and $w(t)$ up to time $t$.

### Gradient and Least-Squares Algorithms

The *gradient algorithm* is a steepest descent approach to minimize $e_1^2(t)$. Since

$$\frac{\partial e_1^2}{\partial \theta} = 2 e_1 \frac{\partial e_1}{\partial \theta} = 2 e_1 w \qquad (2.0.16)$$

we let the parameter update law

$$\dot{\theta} = - g e_1 w \qquad g > 0 \qquad (2.0.17)$$

where $g$ is an arbitrary gain, called the *adaptation gain*.

Another approach is the *least-squares algorithm* which minimizes the integral-squared-error (ISE)

$$ISE = \int_0^t e_1^2(\tau) \, d\tau \qquad (2.0.18)$$

$$(\theta^T(t) w(\tau) - y_p(\tau))^2$$

Owing to the linearity of the error equation, the estimate may be obtained directly from the condition

$$\frac{\partial}{\partial \theta} \left[ \int_0^t e_1^2(\tau) \, d\tau \right] = 2 \int_0^t w(\tau) \left[ w^T(\tau) \theta(\tau) - y_p(\tau) \right] d\tau = 0 \qquad (2.0.19)$$

so that the *least-squares estimate* is given by

$$\theta_{LS}(t) = \left[ \int_0^t w(\tau) w^T(\tau) \, d\tau \right]^{-1} \left[ \int_0^t w(\tau) y_p(\tau) \, d\tau \right] \qquad (2.0.20)$$

Plugging (2.0.14) into (2.0.20) shows that $\theta_{LS}(t) = \theta^*$, *assuming that the inverse in (2.0.20) exists.*

For adaptive control applications, we are interested in *recursive formulations* such as (2.0.17), where parameters are updated continuously on the basis of input-output data. Such an expression may be obtained for the least-squares algorithms by defining

$$P(t) = \left[ \int_0^t w(\tau) w^T(\tau) \, d\tau \right]^{-1} \qquad (2.0.21)$$

so that

$$\frac{d}{dt} \left[ P^{-1}(t) \right] = w(t) w^T(t) \qquad (2.0.22)$$

Since

$$0 = \frac{d}{dt} (I) = \frac{d}{dt} \left[ P(t) P^{-1}(t) \right]$$

$$= \frac{d}{dt} \left[ P(t) \right] P^{-1}(t) + P(t) \frac{d}{dt} \left[ P^{-1}(t) \right] \qquad (2.0.23)$$

it follows that

$$\frac{d}{dt} \left[ P(t) \right] = - P(t) \frac{d}{dt} \left[ P^{-1}(t) \right] P(t)$$

$$= - P(t) w(t) w^T(t) P(t) \qquad (2.0.24)$$

On the other hand, (2.0.20) may be written

$$\theta_{LS}(t) = P(t) \int_0^t w(\tau) y_p(\tau) \, d\tau \qquad (2.0.25)$$

so that, using (2.0.24)

$$\frac{d}{dt} \left[ \theta_{LS}(t) \right] = - P(t) w(t) w^T(t) \theta_{LS}(t) + P(t) w(t) y_p(t)$$

$$= - P(t) w(t) \left[ w^T(t) \theta_{LS}(t) - y_p(t) \right]$$

$$= - P(t) w(t) e_1(t) \qquad (2.0.26)$$

Note that the recursive algorithm (2.0.24), (2.0.26) should be started with the correct initial conditions at some $t_0 > 0$ such that

$$P(t_0) = \left[ \int_0^{t_0} w(\tau) \, w^T(\tau) \, d\tau \right]^{-1} \qquad (2.0.27)$$

exists. In practice, the *recursive least-squares* algorithm is started with arbitrary initial conditions at $t_0 = 0$ so that

$$\dot{\theta}(t) = -P(t) \, w(t) \left[ \theta^T(t) \, w(t) - y_p(t) \right] \qquad \theta(0) = \theta_0$$

$$\dot{P}(t) = -P(t) \, w(t) \, w^T(t) \, P(t) \qquad F(0) = P_0 > 0 \qquad (2.0.28)$$

$$L = P(0)^T$$

It may be verified that the solution of (2.0.28) is

$$\theta(t) = \left[ P_0^{-1} + \int_0^t w(\tau) \, w^T(\tau) \, d\tau \right]^{-1} \left[ P_0^{-1}\theta_0 + \int_0^t y_p(\tau) \, w(\tau) \, d\tau \right] \qquad (2.0.29)$$

instead of (2.0.20). Since $y_p = \theta^{*T}w$, the *parameter error* is given by

$$\theta(t) - \theta^* = \left[ P_0^{-1} + \int_0^t w(\tau) \, w^T(\tau) \, d\tau \right]^{-1} P_0^{-1}(\theta_0 - \theta^*) \qquad (2.0.30)$$

It follows that $\theta(t)$ *converges asymptotically* to $\theta^*$ if $\int_0^t w(\tau) \, w^T(\tau) \, d\tau$ is

unbounded as $t \to \infty$. In this chapter, we will study conditions that guarantee *exponential convergence* of the parameter estimates to the nominal parameter. These are called *persistency of excitation* conditions and are closely related to the above condition. It is not obvious at this point how to relate the time domain condition on the vector $w$ to frequency domain conditions on the input. This will be a goal of this chapter.

**Model Reference Identification**

We now discuss another family of identification algorithms, based on the so-called *model reference approach*. The algorithms have similarities to the previous ones, and are useful to introduce adaptive control techniques.

We first define a *reference model* transfer function

$$\frac{\hat{y}_m}{\hat{u}} = \hat{M}(s) = \frac{k_m}{s + a_m} \qquad (2.0.31)$$

where $a_m, k_m > 0$. In the time domain

$$\dot{y}_m(t) = -a_m y_m(t) + k_m u(t) \qquad (2.0.32)$$

Let the input $u$ to the model be given by

$$u(t) = a_0(t) r(t) + b_0(t) y_m(t) \qquad (2.0.33)$$

where $a_0(t)$, $b_0(t)$ are adaptive parameters and $r$ is the input to the plant. The motivation is that there exist *nominal values* of the parameters, denoted $a_0^*$, $b_0^*$, such that the closed-loop transfer function matches any first order transfer function. Specifically, (2.0.32) and (2.0.33) give the closed-loop system

$$\dot{y}_m(t) = -(a_m - k_m b_0(t)) y_m(t) + k_m a_0(t) r(t) \qquad (2.0.34)$$

so that the nominal values of $a_0$, $b_0$ are

$$a_0^* = \frac{k_p}{k_m} \qquad b_0^* = \frac{a_m - a_p}{k_m} \qquad (2.0.35)$$

Clearly, knowledge of $a_0^*$, $b_0^*$ is equivalent to knowledge of $a_p$, $k_p$. We define the following vectors

$$\theta(t) = \begin{bmatrix} a_0(t) \\ b_0(t) \end{bmatrix} \qquad \theta^* = \begin{bmatrix} a_0^* \\ b_0^* \end{bmatrix} \qquad w(t) = \begin{bmatrix} r(t) \\ y_m(t) \end{bmatrix} \qquad (2.0.36)$$

and the *identification error*

$$e_1(t) = y_m(t) - y_p(t) \qquad (2.0.37)$$

so that

$$\dot{e}_1(t) = -(a_m - k_m b_0(t)) y_m(t) + k_m a_0(t) r(t) + a_p y_p(t) - k_p r(t)$$

$$= -a_m e_1(t) + k_m ((a_0(t) - a_0^*) r(t) + (b_0(t) - b_0^*) y_p(t)$$

$$= -a_m e_1(t) + k_m (\theta^T(t) - \theta^{*T}) w(t) \qquad (2.0.38)$$

In short

$$e_1(t) = \hat{M} \left[ (\theta^T(t) - \theta^{*T}) w(t) \right] \qquad (2.0.39)$$

which may be compared to the linear error equation (2.0.15). This equation is still linear, but it involves the dynamics of the transfer function $\hat{M}$. Can we still find update algorithms for $\theta(t)$, based on the new error equation? The answer is yes; but the approach is now based on the Lyapunov function

$$v(e_1, \theta) = \frac{e_1^2}{2} + \frac{k_m}{2} (\theta^T - \theta^{*T})(\theta - \theta^*) \qquad (2.0.40)$$

so that

$$\dot{v} = -a_m e_1^2 + k_m e_1 (\theta^T - \theta^{*T}) w + k_m (\theta^T - \theta^{*T}) \dot{\theta} \qquad (2.0.41)$$

By letting

$$\dot{\theta} = -e_1 w \qquad (2.0.42)$$

it follows that

$$\dot{v} = -a_m e_1^2 \leq 0 \qquad (2.0.43)$$

Therefore, $e_1$ and $\theta$ are bounded. It may also be shown that $e_1 \to 0$ as $t \to \infty$ and that $\theta \to \theta^*$ under further conditions on the reference input.

The resulting update law (2.0.42) is identical to the gradient algorithm (2.0.17) obtained for the linear error equation (2.0.15). In this case however, it is *not* the gradient algorithm for (2.0.39), due to the presence of the transfer function $\hat{M}$. The motivation for the algorithm lies only in the Lyapunov stability proof.

Note that the derivation requires $a_m > 0$ (in (2.0.43)) and $k_m > 0$ (in (2.0.40)). In general, the conditions to be satisfied by $\hat{M}$ are that

• $\hat{M}$ is stable

• $\text{Re}(\hat{M}(j\omega)) > 0$ for all $\omega \geq 0$.

These are very important conditions in adaptive control, defining *strictly positive real* transfer functions. They will be discussed in greater detail later in this chapter.

## 2.1 IDENTIFICATION PROBLEM

We now consider the general identification problem for single-input single-output (SISO) linear time invariant (LTI) systems. But first, a few definitions. A polynomial in $s$ is called *monic* if the coefficient of the highest power in $s$ is 1 and *Hurwitz* if its roots lie in the open left-half plane. Rational transfer functions are called *stable* if their denominator polynomial is Hurwitz and *minimum phase* if their numerator polynomial is Hurwitz. The *relative degree* of a transfer function is by definition the difference between the degrees of the denominator and numerator polynomials. A rational transfer function is called *proper* if its relative degree is at least 0 and *strictly proper* if its relative degree is at least 1.

In this chapter, we consider the identification problem of SISO LTI systems, given the following assumptions.

**Assumptions**

**(A1)      Plant Assumptions**
The plant is a SISO LTI system, described by a transfer function

$$\frac{\hat{y}_p(s)}{\hat{r}(s)} = \hat{P}(s) = k_p \frac{\hat{n}_p(s)}{\hat{d}_p(s)} \qquad (2.1.1)$$

where $\hat{r}(s)$ and $\hat{y}_p(s)$ are the Laplace transforms of the input and output of the plant, respectively, and $\hat{n}_p(s)$ and $\hat{d}_p(s)$ are monic, coprime polynomials of degrees $m$ and $n$ respectively. $m$ is unknown, but the plant is strictly proper ($m \leq n - 1$).

**(A2)      Reference Input Assumptions**
The input $r(.)$ is piecewise continuous and bounded on $\mathbb{R}_+$.

The objective of the identifier is to obtain estimates of $k_p$ and of the coefficients of the polynomials $\hat{n}_p(s)$ and $\hat{d}_p(s)$ from measurements of the input $r(t)$ and output $y_p(t)$ only. Note that we do not assume that $\hat{P}$ is stable.

## 2.2 IDENTIFIER STRUCTURE

The identifier structure presented in this section is generally known as an *equation error* identifier. (cf. Ljung & Soderstrom [1983]). The transfer function $\hat{P}(s)$ can be explicitly written as

$$\frac{\hat{y}_p(s)}{\hat{r}(s)} = \hat{P}(s) = \frac{\alpha_n s^{n-1} + \cdots + \alpha_1}{s^n + \beta_n s^{n-1} + \cdots + \beta_1} \qquad (2.2.1)$$

where the $2n$ coefficients $\alpha_1 \ldots \alpha_n$ and $\beta_1 \ldots \beta_n$ are unknown. This expression is a *parameterization* of the unknown plant, that is a model in which only a finite number of parameters are to be determined. For identification purposes, it is convenient to find an expression which depends linearly on the unknown parameters. For example, the expression

$$s^n \hat{y}_p(s) = (\alpha_n s^{n-1} + \cdots + \alpha_1) \hat{r}(s)$$
$$- (\beta_n s^{n-1} + \cdots + \beta_1) \hat{y}_p(s) \qquad (2.2.2)$$

is linear in the parameters $\alpha_i$ and $\beta_i$. However, it would require explicit differentiations to be implemented. To avoid this problem, we

introduce a monic $n$th order polynomial denoted $\hat{\lambda}(s) = s^n + \lambda_n s^{n-1} + \cdots + \lambda_1$. This polynomial is assumed to be Hurwitz but is otherwise arbitrary. Then, using (2.1.1)

$$\hat{\lambda}(s)\hat{y}_p(s) = k_p \hat{n}_p(s)\hat{r}(s) + (\hat{\lambda}(s) - \hat{d}_p(s))\hat{y}_p(s) \qquad (2.2.3)$$

or, with (2.2.1)

$$\hat{y}_p(s) = \frac{\alpha_n s^{n-1} + \cdots + \alpha_1}{\hat{\lambda}(s)} \hat{r}(s)$$

$$+ \frac{(\lambda_n - \beta_n)s^{n-1} + \cdots + (\lambda_1 - \beta_1)}{\hat{\lambda}(s)} \hat{y}_p(s) \qquad (2.2.4)$$

This expression is a new parameterization of the plant. Let

$$\hat{a}^*(s) = \alpha_n s^{n-1} + \cdots + \alpha_1 = k_p \hat{n}_p(s)$$

$$\hat{b}^*(s) = (\lambda_n - \beta_n)s^{n-1} + \cdots + (\lambda_1 - \beta_1) = \hat{\lambda}(s) - \hat{d}_p(s) \qquad (2.2.5)$$

so that the new representation of the plant can be written

$$\hat{y}_p(s) = \frac{\hat{a}^*(s)}{\hat{\lambda}(s)} \hat{r}(s) + \frac{\hat{b}^*(s)}{\hat{\lambda}(s)} \hat{y}_p(s) \qquad (2.2.6)$$

The transfer function from $r \rightarrow y_p$ is given by

$$\frac{\hat{y}_p(s)}{\hat{r}(s)} = \frac{\hat{a}^*(s)}{\hat{\lambda}(s) - \hat{b}^*(s)} \qquad (2.2.7)$$

and it is easy to verify that this transfer function is $\hat{P}(s)$ when $\hat{a}^*(s)$ and $\hat{b}^*(s)$ are given by (2.2.5). Further, this choice is unique when $\hat{n}_p(s)$ and $\hat{d}_p(s)$ are coprime: indeed, suppose that there exist $\hat{a}^*(s) + \delta\hat{a}(s)$, $\hat{b}^*(s) + \delta\hat{b}(s)$, such that the transfer function was still $k_p \hat{n}_p(s)/\hat{d}_p(s)$. The following equation would then have to be satisfied

$$\frac{\delta\hat{a}(s)}{\delta\hat{b}(s)} = -k_p \frac{\hat{n}_p(s)}{\hat{d}_p(s)} = -\hat{P}(s) \qquad (2.2.8)$$

However, equation (2.2.8) has no solution since the degree of $\hat{d}_p$ is $n$, and $\hat{n}_p$, $\hat{d}_p$ are coprime, while the degree of $\delta\hat{b}$ is at most $n-1$.

**State-Space Realization**

A state-space realization of the foregoing representation can be found by choosing $\Lambda \in \mathbb{R}^{n \times n}$, $b_\lambda \in \mathbb{R}^n$ in controllable canonical form, such that

$$\Lambda = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & 1 \\ -\lambda_1 & \cdot & \cdot & \cdot & \cdot & -\lambda_n \end{bmatrix} \qquad b_\lambda = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 1 \end{bmatrix}$$

$$(sI - \Lambda)^{-1} b_\lambda = \frac{1}{\hat{\lambda}(s)} \begin{bmatrix} 1 \\ s \\ \cdot \\ \cdot \\ \cdot \\ s^{n-1} \end{bmatrix} \qquad (2.2.9)$$

In analogy with (2.2.5), define

$$a^{*^T} := (\alpha_1, \ldots, \alpha_n) \qquad b^{*^T} := (\lambda_1 - \beta_1, \ldots, \lambda_n - \beta_n) \qquad (2.2.10)$$

and the vectors $w_p^{(1)}(t)$, $w_p^{(2)}(t) \in \mathbb{R}^n$

$$\dot{w}_p^{(1)} = \Lambda w_p^{(1)} + b_\lambda r$$

$$\dot{w}_p^{(2)} = \Lambda w_p^{(2)} + b_\lambda y_p \qquad (2.2.11)$$

with initial conditions $w_p^{(1)}(0)$, $w_p^{(2)}(0)$. In Laplace transforms

$$\hat{w}_p^{(1)}(s) = (sI - \Lambda)^{-1} b_\lambda \hat{r}(s) + (sI - \Lambda)^{-1} w_p^{(1)}(0)$$

$$\hat{w}_p^{(2)}(s) = (sI - \Lambda)^{-1} b_\lambda \hat{y}_p(s) + (sI - \Lambda)^{-1} w_p^{(2)}(0) \qquad (2.2.12)$$

With this notation, the description of the plant (2.2.6) becomes

$$\hat{y}_p(s) = a^{*^T} \hat{w}_p^{(1)}(s) + b^{*^T} \hat{w}_p^{(2)}(s) \qquad (2.2.13)$$

and, since the plant parameters $a^*$, $b^*$ are constant, the same expression is valid in the time domain

$$y_p(t) = a^{*^T} w_p^{(1)}(t) + b^{*^T} w_p^{(2)}(t) := \theta^{*^T} w_p(t) \qquad (2.2.14)$$

where

$$\theta^{*^T} := (a^{*^T}, b^{*^T}) \in \mathbb{R}^{2n}$$

$$w_p(t)^T := (w_p^{(1)^T}(t), w_p^{(2)^T}(t)) \in \mathbb{R}^{2n} \qquad (2.2.15)$$

Equations (2.2.10)–(2.2.14) define a realization of the new parameterization. The vector $w_p$ is the *generalized state* of the plant and has dimension $2n$. Therefore, the realization of $\hat{P}(s)$ is not minimal, but the unobservable modes are those of $\hat{\lambda}(s)$ and are all stable.

The vector $\theta^*$ is a vector of unknown parameters related linearly to the original plant parameters $\alpha_i$, $\beta_i$ by (2.2.10)–(2.2.15). Knowledge of a set of parameters is equivalent to the knowledge of the other and each corresponds to one of the (equivalent) parameterizations. In the last form, however, the plant output depends linearly on the unknown parameters, so that standard identification algorithms can be used. This plant parameterization is represented in Figure 2.1.
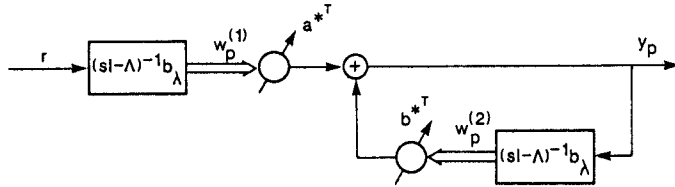


**Figure 2.1:**   Plant Parameterization

### Identifier Structure

The purpose of the identifier is to produce a recursive estimate $\theta(t)$ of the *nominal* parameter $\theta^*$. Since $r$ and $y_p$ are available, we define the *observer*

$$\dot{w}^{(1)} = \Lambda w^{(1)} + b_\lambda r$$

$$\dot{w}^{(2)} = \Lambda w^{(2)} + b_\lambda y_p \tag{2.2.16}$$

to reconstruct the states of the plant. The initial conditions in (2.2.16) are arbitrary. We also define the identifier signals

$$\theta^T(t) := (a^T(t), b^T(t)) \in \mathbb{R}^{2n}$$

$$w^T(t) := (w^{(1)^T}(t), w^{(2)^T}(t)) \in \mathbb{R}^{2n} \tag{2.2.17}$$

By (2.2.11) and (2.2.16), the *observer error* $w(t) - w_p(t)$ decays exponentially to zero, *even when the plant is unstable.* We note that the generalized state of the plant $w_p(t)$ is such that it can be reconstructed from available signals, *without knowledge of the plant parameters.*

The plant output can be written

$$y_p(t) = \theta^{*^T} w(t) + \epsilon(t) \tag{2.2.18}$$

where the signal $\epsilon(t)$ is to remind one of the presence of an additive exponentially decaying term, given here by

$$\epsilon(t) = \theta^{*^T}(w_p(t) - w(t)) \tag{2.2.19}$$

This term is due to the initial conditions in the observer. We will first neglect the presence of the $\epsilon(t)$ term but later show that it does not affect the properties of the identifier.

In analogy with the expression of the plant output, the output of the identifier is defined to be

$$y_i(t) = \theta^T(t) w(t) \quad \in \mathbb{R} \tag{2.2.20}$$

We also define the *parameter error*

$$\phi(t) := \theta(t) - \theta^* \quad \in \mathbb{R}^{2n} \tag{2.2.21}$$

and the *identifier error*

$$e_1(t) := y_i(t) - y_p(t) = \phi^T(t) w(t) + \epsilon(t) \tag{2.2.22}$$

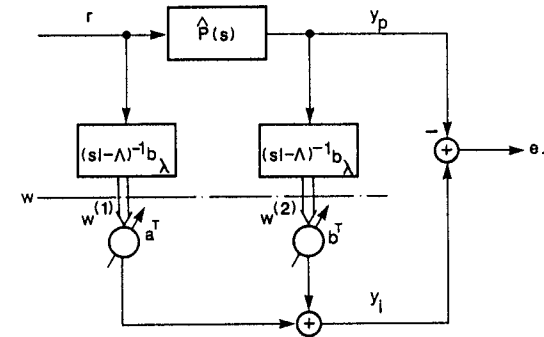These signals will be used by the identification algorithm, and are represented in Figure 2.2.



**Figure 2.2:**   Identifier Structure

### 2.3 LINEAR ERROR EQUATION AND IDENTIFICATION ALGORITHMS

Many identification algorithms (çf. Eykhoff [1974], Ljung & Soderstrom [1983]) rely on a linear expression of the form obtained above, that is

$$y_p(t) = \theta^{*^T} w(t) \tag{2.3.1}$$

where $y_p(t)$, $w(t)$ are known signals and $\theta^*$ is unknown. The vector $w(t)$ is usually called the *regressor* vector. With the expression of $y_p(t)$ is associated the standard *linear error equation*

$$e_1(t) = \phi^T(t) w(t) \tag{2.3.2}$$

We arbitrarily separated the identifier into an *identifier structure* and an *identification algorithm*. The identifier structure constructs the regressor $w$ and other signals, related by the identifier error equation. The identification algorithm is defined by a differential equation, called the *update law*, of the form

$$\dot{\theta} = \dot{\phi} = F(y_p, e_1, \theta, w) \tag{2.3.3}$$

where $F$ is a causal operator explicitly independent of $\theta^*$, which defines the evolution of the identifier parameter $\theta$.

### 2.3.1 Gradient Algorithms
The update law

$$\dot{\theta} = -g e_1 w \qquad g > 0 \tag{2.3.4}$$

defines the standard *gradient algorithm*. The right-hand side is proportional to the gradient of the output error squared, viewed as a function of $\theta$, that is

$$\frac{\partial}{\partial \theta} ( e_1^2(\theta) ) = 2 e_1 w \tag{2.3.5}$$

This update law can thus be seen as a *steepest descent* method. The parameter $g$ is a *fixed*, strictly positive gain called the *adaptation gain*, and it allows us to vary the rate of adaptation of the parameters. The initial condition $\theta(0)$ is arbitrary, but it can be chosen to take any a priori knowledge of the plant parameters into account.

An alternative to this algorithm is the *normalized gradient algorithm*

$$\dot{\theta} = -g \frac{e_1 w}{1 + \gamma w^T w} \qquad g, \gamma > 0 \tag{2.3.6}$$

where $g$ and $\gamma$ are constants. This update law is equivalent to the previous update law, with $w$ replaced by $w / \sqrt{1 + \gamma w^T w}$ in $\dot{\theta} = -g w w^T \phi$. The new regressor is thus a normalized form of $w$. The right-hand side of the differential equation (2.3.6) is globally Lipschitz in $\phi$ (using

(2.3.2)), even when $w$ is unbounded.

When the nominal parameter $\theta^*$ is known *a priori* to lie in a set $\Theta \in \mathbb{R}^{2n}$ (which we will assume to be closed, convex and delimited by a smooth boundary), it is useful to modify the update law to take this information into account. For example, the *normalized gradient algorithm with projection* is defined by

$$\dot{\theta} = -g \frac{e_1 w}{1 + \gamma w^T w} \qquad \theta \in \text{int}(\Theta)$$

$$= \text{Pr}\left[ -g \frac{e_1 w}{1 + \gamma w^T w} \right] \text{ if } \theta \in \partial\Theta \text{ and } e_1 w^T \theta_{perp} < 0 \tag{2.3.7}$$

where $\text{int}\,\Theta$ and $\partial\Theta$ denote the interior and boundary of $\Theta$, $\text{Pr}(z)$ denotes the projection of the vector $z$ onto the hyperplane tangent to $\partial\Theta$ at $\theta$ and $\theta_{perp}$ denotes the unit vector perpendicular to the hyperplane, pointing outward.

A frequent example of projection occurs when *a priori* bounds $p_i^-, p_i^+$ are known, that is

$$\theta_i^* \in [p_i^-, p_i^+] \tag{2.3.8}$$

The update law is then modified to

$$\dot{\theta}_i = 0 \qquad \text{if } \theta_i = p_i^- \text{ and } \dot{\theta}_i < 0$$

$$\text{or } \theta_i = p_i^+ \text{ and } \dot{\theta}_i > 0 \tag{2.3.9}$$

The gradient algorithms can be used to identify the plant parameters with the identifier structure described in Section 2.2. Using the normalized gradient algorithm, for example, the implementation is as follows.

**Identifier with Normalized Gradient Algorithm—Implementation**

*Assumptions*
   (A1)–(A2)
*Data*
   $n$
*Input*
   $r(t), y_p(t) \in \mathbb{R}$
*Output*

$\theta(t), y_i(t) \in \mathbb{R}$

*Internal Signals*

$w(t) \in \mathbb{R}^{2n}(w^{(1)}(t), w^{(2)}(t) \in \mathbb{R}^n)$

$\theta(t) \in \mathbb{R}^{2n}(a(t), b(t) \in \mathbb{R}^n)$

$y_i(t), e_1(t) \in \mathbb{R}$

Initial conditions are arbitrary.

*Design Parameters*

Choose

- $\Lambda \in \mathbb{R}^{n \times n}, b_\lambda \in \mathbb{R}^n$ in controllable canonical form such that

  $\det(sI - \Lambda) = \hat{\lambda}(s)$ is Hurwitz.

- $g, \gamma > 0$.

*Identifier Structure*

$\dot{w}^{(1)} = \Lambda w^{(1)} + b_\lambda r$

$\dot{w}^{(2)} = \Lambda w^{(2)} + b_\lambda y_p$

$\theta^T = (a^T, b^T)$ estimates of $(\alpha_1, \ldots, \alpha_n, \lambda_1 - \beta_1, \ldots, \lambda_n - \beta_n)$

$w^T = (w^{(1)^T}, w^{(2)^T})$

$y_i = \theta^T w$

$e_1 = y_i - y_p$

*Normalized Gradient Algorithm*

$$\dot{\theta} = -g \frac{e_1 w}{1 + \gamma w^T w}$$

□

**Comment**

We leave it to the reader to check that other choices of $(\Lambda, b_\lambda)$ are possible, with minor adjustments. Indeed, all that is required for identification is that there exist unique $a^*, b^*$ in the corresponding parameterization. Alternate choices of $(sI - \Lambda)^{-1}b_\lambda$ include

$$(sI - \Lambda)^{-1}b_\lambda = \begin{bmatrix} \dfrac{1}{s+a} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{1}{(s+a)^n} \end{bmatrix} \qquad a > 0$$

and

$$(sI - \Lambda)^{-1}b_\lambda = \begin{bmatrix} \dfrac{b_1}{s+a_1} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{b_n}{(s+a_n)} \end{bmatrix} \qquad a_i \neq a_j > 0, b_i \neq 0$$

### 2.3.2 Least-Squares Algorithms

Least-squares (LS) algorithms can be derived by several methods. One approach was presented in the introduction. Another interesting approach is to connect the parameter identification problem to the *stochastic state estimation* problem of a linear time varying system. The parameter $\theta^*$ can be considered to be the unknown state of the system

$$\dot{\theta}^*(t) = 0 \qquad\qquad (2.3.10)$$

with output

$$y_p(t) = w^T(t)\theta^*(t) \qquad\qquad (2.3.11)$$

Assuming that the right-hand sides of (2.3.10)–(2.3.11) are perturbed by zero mean white gaussian noises of spectral intensities $Q \in \mathbb{R}^{2n \times 2n}$ and $1/g \in \mathbb{R}$, respectively, the least-squares estimator is the so-called *Kalman filter* (Kalman & Bucy [1961])

$$\dot{\theta} = -g P w e_1$$

$$\dot{P} = Q - g P w w^T P \qquad Q, g > 0 \qquad (2.3.12)$$

$Q$ and $g$ are fixed design parameters of the algorithm. The update law for $\theta$ is very similar to the gradient update law, with the presence of the so-called *correlation* term $w e_1$. The matrix $P$ is called the

*covariance matrix* and acts in the $\theta$ update law as a time-varying, *directional* adaptation gain. The covariance update law in (2.3.12) is called the *covariance propagation equation.* The initial conditions are arbitrary, except that $P(0) > 0$. $P(0)$ is usually chosen to reflect the confidence in the initial estimate $\theta(0)$.

In the identification literature, the least-squares algorithm referred to is usually the algorithm with $Q = 0$, since the parameter $\theta^*$ is assumed to be constant. The covariance propagation equation is then replaced by

$$\frac{dP}{dt} = -gPww^TP \quad \text{or} \quad \frac{d(P^{-1})}{dt} = gww^T \quad g > 0 \qquad (2.3.13)$$

where $g$ is a constant.

The new expression for $P^{-1}$ shows that $dP^{-1}/dt \geq 0$, so that $P^{-1}$ may grow without bound. Then $P$ will become arbitrarily small in some directions and the adaptation of the parameters in those directions becomes very slow. This so-called *covariance wind-up* problem can be prevented using the *least-squares with forgetting factor* algorithm, defined by

$$\frac{dP}{dt} = -g(-\lambda P + Pww^TP)$$

$$\text{or} \quad \frac{d(P^{-1})}{dt} = g(-\lambda P^{-1} + ww^T) \quad \lambda, g > 0 \qquad (2.3.14)$$

Another possible remedy is the *covariance resetting,* where $P$ is reset to a predetermined positive definite value, whenever $\lambda_{\min}(P)$ falls under some threshold.

The *normalized least-squares* algorithm is defined (cf. Goodwin & Mayne [1987]) by

$$\dot{\theta} = -g\frac{Pwe_1}{1 + \gamma w^TPw} \qquad g, \gamma > 0$$

$$\frac{dP}{dt} = -g\frac{Pww^TP}{1 + \gamma w^TPw}$$

$$\text{or} \quad \frac{d(P^{-1})}{dt} = g\frac{ww^T}{1 + \gamma w^T(P^{-1})^{-1}w} \qquad (2.3.15)$$

Again $g, \gamma$ are fixed parameters and $P(0) > 0$. The same modifications can also be made to avoid covariance wind-up.

The least-squares algorithms are somewhat more complicated to implement but are found in practice to have faster convergence

properties.

**Identifier with Normalized LS Algorithm and Covariance Resetting—Implementation**

The algorithm is the same as for the normalized gradient algorithm, except

*Internal Signals*

In addition

$$P(t) \in \mathbb{R}^{2n \times 2n}$$

*Design Parameters*

Choose, in addition

- $k_0 > k_1 > 0$

*Normalized LS Algorithm with Covariance Resetting*

$$\dot{\theta} = -g\frac{Pwe_1}{1 + \gamma w^TPw} \qquad \dot{P} = -g\frac{Pww^TP}{1 + \gamma w^TPw}$$

$$P(0) = P(t_r^+) = k_0 I > 0 \text{ where } t_r = \{t \mid \lambda_{\min}(P(t)) \leq k_1\}$$

$\square$

## 2.4 PROPERTIES OF THE IDENTIFICATION ALGORITHMS—IDENTIFIER STABILITY

### 2.4.1 Gradient Algorithms

In this section, we establish properties of the gradient algorithm

$$\dot{\phi} = \dot{\theta} = -ge_1w \qquad g > 0 \qquad (2.4.1)$$

and the normalized gradient algorithm

$$\dot{\phi} = \dot{\theta} = -g\frac{e_1w}{1 + \gamma w^Tw} \qquad g, \gamma > 0 \qquad (2.4.2)$$

assuming the linear error equation

$$e_1 = \phi^Tw \qquad (2.4.3)$$

Theorems 2.4.1–2.4.4 establish general properties of the gradient algorithms and concern solutions of the differential equations (2.4.1)–(2.4.2), with $e_1$ defined by (2.4.3). The properties do not require that the vector $w$ originates from the identifier described in Section 2.2, but only require that $w$ be a piecewise continuous function of time, to guarantee the existence of the solutions. The theorems are also valid for vectors $w$ of any dimension, not necessarily even.

**Theorem 2.4.1    Linear Error Equation with Gradient Algorithm**

Consider the linear error equation (2.4.3), together with the gradient algorithm (2.4.1). Let $w : \mathbb{R}_+ \to \mathbb{R}^{2n}$ be piecewise continuous.

*Then*    (a)    $e_1 \in L_2$

          (b)    $\phi \in L_\infty$

**Proof of Theorem 2.4.1**

The differential equation describing $\phi$ is

$$\dot{\phi} = -g w w^T \phi$$

Let the Lyapunov function $v = \phi^T \phi$ so that the derivative along the trajectories is given by

$$\dot{v} = -2g(\phi^T w)^2 = -2g e_1^2 \le 0$$

Hence, $0 \le v(t) \le v(0)$ for all $t \ge 0$, so that $v, \phi \in L_\infty$. Since $v$ is a positive, monotonically decreasing function, the limit $v(\infty)$ is well-defined and

$$-\left(\frac{1}{2}g\right) \int_0^\infty \dot{v}\, dt = \int_0^\infty e_1^2\, dt < \infty$$

that is $e_1 \in L_2$.    □

**Theorem 2.4.2    Linear Error Equation with Normalized Gradient Algorithm**

Consider the linear error equation (2.4.3) together with the normalized gradient algorithm (2.4.2). Let $w : \mathbb{R}_+ \to \mathbb{R}^{2n}$ be piecewise continuous.

*Then*    (a)    $\dfrac{e_1}{\sqrt{1 + \gamma w^T w}} \in L_2 \cap L_\infty$

       (b)    $\phi \in L_\infty, \dot{\phi} \in L_2 \cap L_\infty$

       (c)    $\beta = \dfrac{\phi^T w}{1 + \| w_t \|_\infty} \in L_2 \cap L_\infty$

**Proof of Theorem 2.4.2**

Let $v = \phi^T \phi$, so that

$$\dot{v} = -\frac{2g e_1^2}{1 + \gamma w^T w} \le 0$$

Hence, $0 \le v(t) \le v(0)$ for all $t \ge 0$, so that $v, \phi, e_1/\sqrt{1 + \gamma w^T w}$,

$\beta \in L_\infty$. Using the fact that $x/1+x \le 1$ for all $x \ge 0$, we get that $|\dot{\phi}| \le (g/\gamma)|\phi|$, and $\dot{\phi} \in L_\infty$. Since $v$ is a positive, monotonically decreasing function, the limit $v(\infty)$ is well defined and

$$-\int_0^\infty \dot{v}\, dt < \infty$$

implies that $e_1/\sqrt{1 + \gamma w^T w} \in L_2$. Note that

$$\beta = \frac{e_1}{\sqrt{1 + \gamma w^T w}} \frac{\sqrt{1 + \gamma w^T w}}{1 + \| w_t \|_\infty}$$

where the first term is in $L_2$ and the second in $L_\infty$, so that $\beta \in L_2$. Since

$$|\dot{\phi}|^2 \le \frac{g^2}{\gamma} \frac{e_1^2}{1 + \gamma w^T w}$$

it follows that $\dot{\phi} \in L_2$.    □

**Effect of Initial Conditions and Projection**

In the derivation of the linear error equation in Section 2.2, we found exponentially decaying terms, such that (2.4.3) is replaced by

$$e_1(t) = \phi^T(t) w(t) + \epsilon(t) \tag{2.4.4}$$

where $\epsilon(t)$ is an exponentially decaying term due to the initial conditions in the observer.

It may also be useful, or necessary, to replace the gradient algorithms by the algorithms with projection. The following theorem asserts that these modifications do not affect the previous results.

**Theorem 2.4.3    Effect of Initial Conditions and Projection**

*If*    the linear error equation (2.4.3) is replaced by (2.4.4) and/or the gradient algorithms are replaced by the gradient algorithms with projection,

*Then*    the conclusions of theorems 2.4.1 and 2.4.2 are valid.

**Proof of Theorem 2.4.3**

*(a) Effect of initial conditions*

Modify the Lyapunov function to

$$v = \phi^T\phi + \frac{g}{2}\int_t^\infty \epsilon^2(\tau)\,d\tau$$

Note that the additional term is bounded and tends to zero as $t$ tends to infinity. Consider first the gradient algorithm (2.4.1), so that

$$\dot{v} = -2g(\phi^T w)^2 - 2g(\phi^T w)\epsilon - \frac{g}{2}\epsilon^2$$

$$= -2g(\phi^T w + \frac{\epsilon}{2})^2 \leq 0 \tag{2.4.5}$$

The proof can be completed as in theorem 2.4.1, noting that $\epsilon \in L_2 \cap L_\infty$, and similarly for theorem 2.4.2.

*(b) Effect of projection*

Denote by $z$ the right-hand side of the update law (2.4.1) or (2.4.2). When $\theta \in \partial\Theta$ and $z$ is directed outside $\Theta$, $z$ is replaced by $\text{Pr}(z)$ in the update law. Note that it is sufficient to prove that the derivative of the Lyapunov function on the boundary is less than or equal to its value with the original differential equation. Therefore, denote by $z_{perp}$ the component of $z$ perpendicular to the tangent plane at $\theta$, so that $z = \text{Pr}(z) + z_{perp}$. Since $\theta^* \in \Theta$ and $\Theta$ is convex, $(\theta^T - \theta^{*T})\cdot z_{perp} = \phi^T z_{perp} \geq 0$. Using the Lyapunov function $v = \phi^T\phi$, we find that, for the original differential equation, $\dot{v} = 2\phi^T z$. For the differential equation with projection, $\dot{v}_{\text{Pr}} = 2\phi^T\text{Pr}(z) = \dot{v} - 2\phi^T\cdot z_{perp}$ so that $\dot{v}_{\text{Pr}} \leq \dot{v}$, ~~that is the projection can only improve the convergence of the algorithm.~~ The proof can again be completed as before.  □

### 2.4.2  Least-Squares Algorithms

We now turn to the *normalized LS algorithm with covariance resetting,* defined by the following update law

$$\dot{\phi} = \dot{\theta} = -g\frac{Pwe_1}{1+\gamma w^T P w} \qquad g,\gamma > 0 \tag{2.4.6}$$

and a *discontinuous* covariance propagation

$$\frac{dP}{dt} = -g\frac{Pww^T P}{1+\gamma w^T P w} \quad \text{or} \quad \frac{d(P^{-1})}{dt} = g\frac{ww^T}{1+\gamma w^T(P^{-1})^{-1}w}$$

$$P(0) = P(t_r^+) = k_0 I > 0$$

$$\text{where } t_r = \{t\,|\,\lambda_{\min}(P(t)) \leq k_1 < k_0\}. \tag{2.4.7}$$

This update law has similar properties as the normalized gradient update

law, as stated in the following theorem.

**Theorem 2.4.4    Linear Error Equation with Normalized LS Algorithm and Covariance Resetting**

Consider the linear error equation (2.4.3), together with the normalized LS algorithm with covariance resetting (2.4.6)–(2.4.7).

Let $w : \mathbb{R}_+ \to \mathbb{R}^{2n}$ be piecewise continuous.

*Then*    (a)  $\dfrac{e_1}{\sqrt{1+\gamma w^T P w}} \in L_2 \cap L_\infty$

(b)  $\phi \in L_\infty,\, \dot{\phi} \in L_2 \cap L_\infty$

(c)  $\beta = \dfrac{\phi^T w}{1+\|w_t\|_\infty} \in L_2 \cap L_\infty$

**Proof of Theorem 2.4.4**

The covariance matrix $P$ is a discontinuous function of time. Between discontinuities, the evolution is described by the differential equation in (2.4.7). We note that $d/dt\,P^{-1} \geq 0$, so that $P^{-1}(t_1) - P^{-1}(t_2) \geq 0$ for all $t_1 \geq t_2 \geq 0$ between covariance resettings. At the resettings, $P^{-1}(t_r^+) = k_0^{-1}I$, so that $P^{-1}(t) \geq P^{-1}(t_0) = k_0^{-1}I$, for all $t \geq 0$.

On the other hand, due to the resetting, $P(t) \geq k_1 I$ for all $t \geq 0$, so that

$$k_0 I \geq P(t) \geq k_1 I \qquad k_1^{-1}I \geq P^{-1}(t) \geq k_0^{-1}I \tag{2.4.8}$$

where we used results of Section 1.3.

Note that the interval between resettings is bounded below, since

$$\left\|\frac{d(P^{-1})}{dt}\right\| \leq g\frac{|w|^2}{1+\gamma\lambda_{\min}(P)|w|^2}$$

$$\leq \frac{g}{\gamma}\|P^{-1}\| \tag{2.4.9}$$

where we used the fact that $x/1+x \leq 1$ for all $x \geq 0$. Thus, the differential equation governing $P^{-1}$ is globally Lipschitz. It also follows that $\{t_r\}$ is a set of measure zero.

Let now $v = \phi^T P^{-1}\phi$, so that

$$\dot{v} = -g\frac{e_1^2}{1+\gamma w^T P w} \leq 0$$

between resettings. At the points of discontinuity of $P$,

$$v(t_r^+) - v(t_r) = \phi^T(P^{-1}(t_r^+) - P^{-1}(t_r))\phi \leq 0$$

It follows that $0 \leq v(t) \leq v(0)$, for all $t \geq 0$, and, from the bounds on $P$, we deduce that $\phi, \dot{\phi}, \beta \in L_\infty$. Also

$$-\int_0^\infty \dot{v}\, dt < \infty \quad \text{implies} \quad \frac{e_1}{\sqrt{1 + \gamma w^T P w}} \in L_2$$

Note that

$$\frac{\phi^T w}{1 + \|w_t\|_\infty} = \frac{\phi^T w}{\sqrt{1 + \gamma w^T P w}} \frac{\sqrt{1 + \gamma w^T P w}}{1 + \|w_t\|_\infty} \tag{2.4.10}$$

$$\dot{\phi} = -g \frac{e_1}{\sqrt{1 + \gamma w^T P w}} \frac{P w}{\sqrt{1 + \gamma w^T P w}} \tag{2.4.11}$$

where the first terms in the right-hand sides of (2.4.10)–(2.4.11) are in $L_2$ and the last terms are bounded. The conclusions follow from this observation. □

**Comments**

a)    Theorems 2.4.1–2.4.4 state general properties of differential equations arising from the identification algorithms described in Section 2.3. The theorems can be directly applied to the identifier with the structure described in Section 2.2, and the results interpreted in terms of the parameter error $\phi$ and the identifier error $e_1$.

b)    The conclusions of theorems 2.4.1–2.4.4 may appear somewhat weak, since none of the errors involved actually converge to zero. The reader should note however that the conclusions are valid under very general conditions regarding the input signal $w$. In particular, no assumption is made on the boundedness or on the differentiability of $w$.

c)    The conclusions of theorem 2.4.2 can be interpreted in the following way. The function $\beta(t)$ is defined by

$$\beta(t) = \frac{\phi^T(t) w(t)}{1 + \|w_t\|_\infty} = \frac{e_1(t)}{1 + \|w_t\|_\infty} \tag{2.4.12}$$

so that

$$|e_1(t)| = |\phi^T(t) w(t)| = |\beta(t)| \|w_t\|_\infty + |\beta(t)| \tag{2.4.13}$$

The purpose of the identification algorithms is to reduce the parameter error $\phi$ to zero or at least the error $e_1$. In (2.4.12), $\beta$ can be interpreted

as a *normalized error*, that is $e_1$ normalized by $\|w_t\|_\infty$. In (2.4.13), $|\beta|$ can be interpreted as the *gain* from $w$ to $\phi^T w$. From theorem 2.4.2, this gain is guaranteed to become small as $t \to \infty$ in an $L_2$ sense.

### 2.4.3  Stability of the Identifier

We are not guaranteed the convergence of the *parameter error* $\phi$ to zero. Since only one output $y_p$ is measured to determine a vector of unknown parameters, some additional condition on the signal $w$ (see Section 2.5) must be satisfied in order to guarantee parameter convergence. In fact, we are not even guaranteed the convergence of the identifier error $e_1$ to zero. This can be obtained under the following additional assumption

**(A3)    Bounded Output Assumption**
        Assume that the plant is either stable or located in a control loop such that $r$ and $y_p$ are bounded.

**Theorem 2.4.5    Stability of the Identifier**

Consider the identification problem, with (A1)–(A3), the identifier structure of Section 2.2 and the gradient algorithms (2.4.1), (2.4.2) or the normalized LS algorithm with covariance resetting (2.4.6), (2.4.7).

*Then*    the output error $e_1 \in L_2 \cap L_\infty$, $e_1 \to 0$ as $t \to \infty$ and the parameter error $\phi, \dot{\phi} \in L_\infty$.
        The derivative of the parameter error $\dot{\phi} \in L_2 \cap L_\infty$ and $\dot{\phi} \to 0$ as $t \to \infty$.

**Proof of Theorem 2.4.5**

Since $r$ and $y_p$ are bounded, it follows from (2.2.16), (2.2.17), and the stability of $\Lambda$, that $w$ and $\dot{w}$ are bounded. By theorems 2.4.1–2.4.4, $\phi$ and $\dot{\phi}$ are bounded so that $e_1$ and $\dot{e}_1$ are bounded. Also $e_1 \in L_2$, and by corollary 1.2.2, $e_1, \dot{e}_1 \in L_\infty$ and $e_1 \in L_2$ implies that $e_1 \to 0$ as $t \to \infty$. Similar conclusions follow directly for $\dot{\phi}$. □

**Regular Signals**

Theorem 2.4.5 relies on the boundedness of $w, \dot{w}$, guaranteed by (A3). It is of interest to relax this condition and to replace it by a weaker condition. We will present such a result using a *regularity* condition on the regressor $w$. This condition guarantees a certain degree of smoothness of the signal $w$ and seems to be, fundamental in excluding pathological signals in the course of the proofs presented in this book. In discrete time, such a condition is not necessary, because it is automatically

verified. The definition presented here corresponds to a definition in Narendra, Lin, & Valavani [1980].

**Definition    Regular Signals**

Let $z : \mathbb{R}_+ \to \mathbb{R}^n$, such that $z, \dot{z} \in L_{\infty e}$.

$z$ is called *regular* if, for some $k_1, k_2 \geq 0$

$$|\dot{z}(t)| \leq k_1 \| z_t \|_\infty + k_2 \qquad \text{for all } t \geq 0 \qquad (2.4.14)$$

The class of regular signals includes bounded signals with bounded derivatives, but also unbounded signals (e.g., $e^t$). It typically excludes signals with "increasing frequency" such as $\sin(e^t)$. We will also derive some properties of regular signals in Chapter 3. Note that it will be sufficient for (2.4.14) to hold everywhere except on a set of measure zero. Therefore, piecewise differentiable signals can also be considered.

This definition allows us to state the following theorem, extending the properties derived in theorems 2.4.2–2.4.4 to the case when $w$ is regular.

**Proposition 2.4.6**

Let $\phi, w : \mathbb{R}_+ \to \mathbb{R}^{2n}$ be such that $w, \dot{w} \in L_{\infty e}$ and $\phi, \dot{\phi} \in L_\infty$.

*If*      (a)  $w$ is regular

$$(b) \quad \beta = \frac{\phi^T w}{1 + \| w_t \|_\infty} \in L_2$$

*Then*      $\beta, \dot{\beta} \in L_\infty$, and $\beta \to 0$ as $t \to \infty$.

**Proof of Proposition 2.4.6**

Clearly, $\beta \in L_\infty$ and since $\beta, \dot{\beta} \in L_\infty$, $\beta \in L_2$ implies that $\beta \to 0$ as $t \to \infty$ (corollary 1.2.2), we are left to show that $\dot{\beta} \in L_\infty$.

We have that

$$|\dot{\beta}| \leq \left| \dot{\phi}^T \frac{w}{1 + \| w_t \|_\infty} \right| + \left| \phi^T \frac{\dot{w}}{1 + \| w_t \|_\infty} \right|$$

$$+ \left| \frac{\phi^T w}{1 + \| w_t \|_\infty} \frac{(d / dt \, \| w_t \|_\infty)}{1 + \| w_t \|_\infty} \right| \qquad (2.4.15)$$

The first and second terms are bounded, since $\phi, \dot{\phi} \in L_\infty$ and $w$ is regular. On the other hand

---

$$\left| \frac{d}{dt} \| w_t \|_\infty \right| = \left| \frac{d}{dt} \sup_{\tau \leq t} | w(\tau)| \right|$$

$$\leq \left| \frac{d}{dt} | w(t)| \right| \leq \left| \frac{d}{dt} w(t) \right| \qquad (2.4.16)$$

The regularity assumption then implies that the last term in (2.4.15) is bounded, and hence $\dot{\beta} \in L_\infty$.  $\square$

**Stability of the Identifier with Unstable Plant**

Proposition 2.4.6 shows that when $w$ is possibly unbounded, but nevertheless satisfies the regularity condition, the relative error $e_1 / 1 + \| w_t \|_\infty$ or gain from $w \to \phi^T w$ tends to zero as $t \to \infty$.

The conclusions of proposition 2.4.6 are useful in proving stability in adaptive control, where the boundedness of the regressor $w$ is not guaranteed a priori. In the identification problem, we are now allowed to consider the case of an unstable plant with bounded input, that is, to relax assumption (A3).

**Theorem 2.4.7    Stability of the Identifier—Unstable Plant**

Consider the identification problem with (A1) and (A2), the identifier structure of Section 2.2 and the normalized gradient algorithm (2.4.2), or the normalized LS with covariance resetting (2.4.6), (2.4.7).

*Then*      The normalized error $\beta = \dfrac{\phi^T w}{1 + \| w_t \|_\infty} \in L_2 \cap L_\infty$, $\beta \to 0$ as

$t \to \infty$ and $\phi, \dot{\phi} \in L_\infty$.

**Proof of Theorem 2.4.7**

It suffices to show that $w$ is regular, to apply theorem 2.4.2 or 2.4.4 followed by proposition 2.4.6. Combining (2.2.16)–(2.2.18), it follows that

$$\dot{w}(t) = \begin{bmatrix} \Lambda & 0 \\ b_\lambda a^{*T} & \Lambda + b_\lambda b^{*T} \end{bmatrix} w(t) + \begin{bmatrix} b_\lambda \\ 0 \end{bmatrix} r(t) \qquad (2.4.17)$$

Since $r$ is bounded by (A2), (2.4.17) shows that $w$ is regular.  $\square$

**2.5  PERSISTENT EXCITATION AND EXPONENTIAL PARAMETER CONVERGENCE**

In the previous section, we derived results on the stability of the identifiers and on the convergence of the output error $e_1 = \theta^T w - \theta^{*T} w = \phi^T w$ to zero. We are now concerned with the convergence of the

parameter $\theta$ to its nominal value $\theta^*$, that is the convergence of the parameter error $\phi$ to zero.

The convergence of the identification algorithms is related to the asymptotic stability of the differential equation

$$\dot{\phi}(t) = -g\, w(t)\, w^T(t)\, \phi(t) \qquad g > 0 \qquad (2.5.1)$$

which is of the form

$$\dot{\phi}(t) = -A(t)\, \phi(t) \qquad (2.5.2)$$

where $A(t) \in \mathbb{R}^{2n \times 2n}$ is a positive semidefinite matrix for all $t$. Using the Lyapunov function $v = \phi^T \phi$

$$\dot{v} = -\phi^T (A + A^T) \phi$$

When $A(t)$ is uniformly positive definite, with $\lambda_{\min}(A + A^T) \geq 2\alpha$, then $\dot{v} \leq -2\alpha v$, which implies that system (2.5.2) is exponentially stable with rate $\alpha$. For the original differential equation (2.5.1), such is never the case, however, since at any instant the matrix $w(t) w^T(t)$ is of rank 1. In fact, any vector $\phi$ perpendicular to $w$ lies in the null space of $w w^T$ and results in $\dot{\phi} = 0$. However, since $w$ varies with time, we can expect $\phi$ to still converge to 0 if $w$ completely spans $\mathbb{R}^{2n}$ as $t$ varies. This leads naturally to the following definition. For consistency, the dimension of $w$ is assumed to be $2n$, but it is in fact arbitrary.

**Definition    Persistency of Excitation (PE)**

A vector $w : \mathbb{R}_+ \to \mathbb{R}^{2n}$ is *persistently exciting (PE)* if there exist $\alpha_1$, $\alpha_2$, $\delta > 0$ such that

$$\alpha_2 I \geq \int_{t_0}^{t_0 + \delta} w(\tau)\, w^T(\tau)\, d\tau \geq \alpha_1 I \qquad \text{for all } t_0 \geq 0 \qquad (2.5.3)$$

Although the matrix $w(\tau) w^T(\tau)$ is singular for all $\tau$, the PE condition requires that $w$ rotates sufficiently in space that the integral of the matrix $w(\tau) w^T(\tau)$ is uniformly positive definite over any interval of some length $\delta$.

The condition has another interpretation, by reexpressing the PE condition in scalar form

$$\alpha_2 \geq \int_{t_0}^{t_0 + \delta} (w^T(\tau) x)^2\, d\tau \geq \alpha_1 \qquad \text{for all } t_0 \geq 0,\ |x| = 1 \qquad (2.5.4)$$

which appears as a condition on the energy of $w$ in all directions.

With this, we establish the following convergence theorem.

**Theorem 2.5.1  PE and Exponential Stability**

*Let*  $w : \mathbb{R}_+ \to \mathbb{R}^{2n}$ be piecewise continuous.

*If*       $w$ is PE

*Then*    (2.5.1) is globally exponentially stable

**Comments**

The proof of theorem 2.5.1 can be found in various places in the literature (Sondhi & Mitra [1976], Morgan & Narendra [1977a&b], Anderson [1977], Kreisselmeier [1977]). The proof by Anderson has the advantage of leading to interesting interpretations, while those by Sondhi & Mitra and Kreisselmeier give estimates of the convergence rates.

The idea of the proof of exponential stability by Anderson [1977] is to note that the PE condition is a UCO condition on the system

$$\dot{\theta}^*(t) = 0$$
$$y(t) = w^T(t)\, \theta^*(t) \qquad (2.5.5)$$

which is the system described earlier in the context of the least-squares identification algorithms (cf. (2.3.10), (2.3.11)). We recall that the identification problem is equivalent to the state estimation problem for the system described by (2.5.5). We now find that the persistency of excitation condition, which turns out to be an *identifiability condition,* is equivalent to a uniform complete observability condition on system (2.5.5).

The proof of theorem 2.5.1 uses the following lemma by Anderson & Moore [1969], which states that the UCO of the system $[C, A]$ is equivalent to the UCO of the system with output injection $[C, A + KC]$. The proof of the lemma is given in the Appendix. It is an alternate proof to the original proof by Anderson & Moore and relates the eigenvalues of the associated observability grammians, thereby leading to estimates of the convergence rates in the proof of theorem 2.5.1 given afterward.

**Lemma 2.5.2  Uniform Complete Observability Under Output Injection**

Assume that, for all $\delta > 0$, there exists $k_\delta \geq 0$ such that, for all $t_0 \geq 0$

$$\int_{t_0}^{t_0 + \delta} \| K(\tau) \|^2\, d\tau \leq k_\delta \qquad (2.5.6)$$

*Then*   the system $[C, A]$ is uniformly completely observable *if and only if* the system $[C, A + KC]$ is uniformly completely observable.

Moreover, if the observability grammian of the system $[A, C]$ satisfies

$$\beta_2 I \;\geq\; N(t_0, t_0 + \delta) \;\geq\; \beta_1 I$$

then the observability grammian of the system $[A + KC, C]$ satisfies these inequalities with identical $\delta$ and

$$\beta_1' \;=\; \beta_1 / (1 + \sqrt{k_\delta \beta_2})^2 \tag{2.5.7}$$

$$\beta_2' \;=\; \beta_2 \exp(k_\delta \beta_2) \tag{2.5.8}$$

**Proof of Lemma 2.5.2** in Appendix.

**Proof of Theorem 2.5.1**

Let $v = \phi^T \phi$, so that $\dot{v} = -2g(w^T \phi)^2 \leq 0$ along the trajectories of (2.5.1). For all $t_0 \geq 0$

$$\int_{t_0}^{t_0 + \delta} \dot{v} \, d\tau \;=\; -2g \int_{t_0}^{t_0 + \delta} (w^T(\tau) \phi(\tau))^2 \, d\tau \tag{2.5.9}$$

By the PE assumption, the system $[0, w^T(t)]$ is UCO. Under output injection with $K(t) = -gw(t)$, the system becomes $[-gw(t)w^T(t), w^T(t)]$, with

$$k_\delta \;=\; \int_{t_0}^{t_0 + \delta} |gw(\tau)|^2 \, d\tau$$

$$=\; g^2 \operatorname{tr} \left[ \int_{t_0}^{t_0 + \delta} w(\tau) w^T(\tau) \, d\tau \right] \;\leq\; 2ng^2 \beta_2 \tag{2.5.10}$$

where $2n$ is the dimension of $w$. By lemma 2.5.2, the system with output injection is UCO. Therefore, for all $t_0 \geq 0$

$$\int_{t_0}^{t_0 + \delta} \dot{v} \, d\tau \;\leq\; \frac{-2g\beta_1}{(1 + \sqrt{2ng\beta_2})^2} \, |\phi(t_0)|^2 \tag{2.5.11}$$

Exponential convergence then follows from theorem 1.5.2.  □

**Exponential Convergence of the Identifier**

Theorem 2.5.1 can be applied to the identification problem as follows.

**Theorem 2.5.3    Exponential Convergence of the Identifier**

Consider the identification problem with assumptions (A1)–(A3), the identifier structure of Section 2.2 and the gradient algorithms (2.4.1) or (2.4.2), or the normalized LS algorithm with covariance resetting (2.4.6), (2.4.7).

*If*      $w$ is PE

*Then*   the identifier parameter $\theta$ converges to the nominal parameter $\theta^*$ exponentially fast.

**Proof of Theorem 2.5.3**

This theorem follows directly from theorem 2.5.1. Note that when $w$ is bounded, $w$ PE is equivalent to $w/\sqrt{1 + \gamma w^T w}$ PE, so that the exponential convergence is guaranteed for both gradient update laws. The bounds on $P$ obtained in the proof of theorem 2.4.4 allow us to extend the proof of exponential convergence to the LS algorithm.  □

**Exponential Convergence Rates**

Estimates of the convergence rates can be found from the results in the proof of theorem 2.5.1. For the standard gradient algorithm (2.4.1), for example, the convergence rate is given by

$$\alpha \;=\; \frac{1}{2\delta} \ln \left[ \frac{1}{1 - \dfrac{2g\alpha_1}{(1 + \sqrt{2n}g\alpha_2)^2}} \right] \tag{2.5.12}$$

where $g$ is the adaptation gain, $\alpha_1$, $\alpha_2$, and $\delta$ come from the PE definition (2.5.3) and $n$ is the order of the plant. The influence of some design parameters can be studied with this relationship. The constants $\alpha_1$, $\alpha_2$ and $\delta$ depend in a complex manner on the input signal $r$ and on the plant being identified. However, if $r$ is multiplied by 2, then $\alpha_1$, $\alpha_2$ are multiplied by 4. In the limiting case when the adaptation gain $g$ or the reference input $r$ is made small, the rate of convergence $\alpha \rightarrow g\alpha_1/\delta$. In this case, the convergence rate is proportional to the adaptation gain $g$ and to the lower bound in the PE condition. Through the PE condition, it is also proportional to the square of the amplitude of the reference input $r$. This result will be found again in Chapter 4, using averaging techniques.

When the adaptation gain and reference input get sufficiently large, this approximation is no longer valid and (2.5.12) shows that above some level, the convergence rate estimate saturates and even decreases (cf. Sondhi & Mitra [1976]).

It is also possible to show that the presence of the exponentially decaying terms due to initial conditions in the observer do not affect the exponential stability of the system. The rate of convergence will, however, be as found previously only if the rate of decay of the transients is faster than the rate of convergence of the algorithm (cf. Kreisselmeier [1977]).

## 2.6  MODEL REFERENCE IDENTIFIERS—SPR ERROR EQUATION

### 2.6.1  Model Reference Identifiers

In Section 2.2, we presented an identifier structure which led to a linear error equation. This structure was based on a convenient reparametrization of the plant. It is worth pointing out that there exist several ways to reparametrize the plant and many error equations that may be used for identification. The scheme discussed in Sections 2.2–2.3 is generally known as an *equation error* identifier. Landau [1979] discussed an interesting alternative called the *output error* approach. The resulting scheme has significant advantages in terms of noise bias, although its stability may only be guaranteed from prior knowledge about the plant. Another approach, which we will call the *model reference approach* (cf. Luders & Narendra [1973]) is discussed now. We start from an arbitrary reference model, with transfer function $\hat{M}$ satisfying the following conditions

(A3)    **Reference Model Assumptions**
The reference model is a SISO LTI system, described by a transfer function

$$\hat{M}(s) = k_m \frac{\hat{n}_m(s)}{\hat{d}_m(s)} \qquad (2.6.1)$$

where $\hat{n}_m(s)$, $\hat{d}_m(s)$ are monic, coprime polynomials of degrees $l$ and $k \leq n$, respectively. Assume that the reference model is strictly proper, but that its relative degree is no greater than the relative degree of the plant, that is $1 \leq k - l \leq n - m$. The reference model is stable, i.e. $\hat{d}_m$ is Hurwitz.

A simple choice of $\hat{M}$ is $1/s + 1$. As previously, the plant is assumed to satisfy assumptions (A1) and (A2) (cf. Section 2.1).

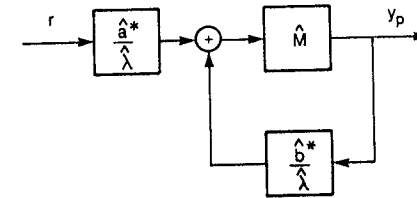Consider now the representation of the plant of Figure 2.3.



**Figure 2.3:**   Model Reference Reparameterization

Although we will have to show (see proposition 2.6.1) that the plant can indeed be represented this way, we may already interpret it as being obtained by modifying the reference model through feedback and feed-forward action, so as to match the plant transfer function. Alternatively, we might interpret this structure as a more general parametrization of the plant than the one in Figure 2.1. In that case, the model transfer function was simply the identity.

The polynomial $\hat{\lambda}$ is a monic, Hurwitz polynomial of degree $n - 1$. It serves a similar purpose in Section 2.2, and although the degree is now $n - 1$, instead of $n$ in Section 2.2, the following derivations can also be carried out with a degree equal to $n$, with only minor adjustments (the reference model may then be assumed to be proper and not strictly proper). We will have to require that the zeros of $\hat{\lambda}(s)$ contain those of $\hat{n}_m(s)$ and therefore we write

$$\hat{\lambda}(s) = \hat{n}_m(s)\hat{\lambda}_0(s) \qquad (2.6.2)$$

where $\hat{\lambda}_0(s)$ is a monic, Hurwitz polynomial of degree $n - l - 1$.

The polynomials $\hat{a}^*(s)$, $\hat{b}^*(s)$ in Figure 2.3 have degrees at most $n - 1$ and serve similar purposes as before. Therefore, we start with the following proposition.

**Proposition 2.6.1**

There exist unique $\hat{a}^*$, $\hat{b}^*$ such that the transfer function from $r \to y_p$ in Figure 2.3 is $\hat{P}(s)$.

**Proof of Proposition 2.6.1**

*Existence*

The transfer function from $r \to y_p$ in Figure 2.3 is given by

$$\frac{\hat{y}_p}{\hat{r}} = \frac{\hat{a}^*}{\hat{\lambda}} \cdot \frac{\dfrac{k_m \hat{n}_m}{\hat{d}_m}}{1 - k_m \dfrac{\hat{n}_m}{\hat{d}_m} \dfrac{\hat{b}^*}{\hat{\lambda}}} = \frac{\hat{a}^* k_m \hat{n}_m}{\hat{\lambda} \hat{d}_m - k_m \hat{n}_m \hat{b}^*}$$

$$= \frac{k_m \hat{a}^*}{\hat{\lambda}_0 \hat{d}_m - k_m \hat{b}^*} \tag{2.6.3}$$

This transfer function will be equal to $\hat{P}(s)$ if and only if

$$\hat{\lambda}_0 \hat{d}_m - k_m \hat{b}^* = \frac{k_m}{k_p} \hat{d}_p \frac{\hat{a}^*}{\hat{n}_p} \tag{2.6.4}$$

The problem is therefore to find polynomials $\hat{a}^*$, $\hat{b}^*$ of degrees at most $n - 1$ that satisfy (2.6.4).

A solution can be found by inspection. Divide $\hat{\lambda}_0 \hat{d}_m$ by $\hat{d}_p$: denote by $\hat{q}$ the quotient of degree $k - l - 1$ and let $k_m \hat{b}^*$ be the remainder of degree $n - 1$. In other words, let

$$\hat{\lambda}_0 \hat{d}_m = \hat{q} \hat{d}_p + k_m \hat{b}^* \tag{2.6.5}$$

This defines $\hat{b}^*$ appropriately. Equality (2.6.4) is satisfied if $\hat{a}^*$ is defined by

$$\hat{a}^* = \frac{k_p}{k_m} \hat{q} \hat{n}_p \tag{2.6.6}$$

The degree of the polynomial in the right-hand side is $m + k - l - 1$, which is at most $n - 1$ by assumption (A3), so that the degree requirements are satisfied.

*Uniqueness*

Assume that there exist $\hat{a}^* + \delta \hat{a}$, $\hat{b}^* + \delta \hat{b}$ satisfying

$$\hat{\lambda}_0 \hat{d}_m - k_m (\hat{b}^* + \delta \hat{b}) = \frac{k_m}{k_p} \hat{d}_p \frac{(\hat{a} + \delta \hat{a})}{\hat{n}_p} \tag{2.6.7}$$

Subtracting (2.6.4) from (2.6.7), we find that

$$\frac{\delta \hat{a}}{\delta \hat{b}} = -k_p \frac{\hat{n}_p}{\hat{d}_p} = -\hat{P} \tag{2.6.8}$$

Recall that $\hat{n}_p$, $\hat{d}_p$ are assumed to be coprime, while the degree of $\hat{d}_p$ and $\delta \hat{b}$ are $n$ and at most $n - 1$, respectively. Therefore, equation (2.6.8) cannot have any solution. $\square$

**Identifier Structure**

The identifier structure is obtained in a similar way as in Section 2.2. We let $\Lambda \in \mathbb{R}^{n-1 \times n-1}$, $b_\lambda \in \mathbb{R}^{n-1}$ in controllable canonical form such that $\det (sI - \Lambda) = \hat{\lambda}(s)$ and

$$(sI - \Lambda)^{-1} b_\lambda = \frac{1}{\hat{\lambda}(s)} \begin{bmatrix} 1 \\ s \\ \cdot \\ \cdot \\ \cdot \\ s^{n-2} \end{bmatrix} \tag{2.6.9}$$

Given any plant and model transfer functions satisfying (A1) and (A3), we showed that there exist unique polynomials $\hat{a}^*(s)$, $\hat{b}^*(s)$. We now let $a_0^*, b_0^* \in \mathbb{R}$ and $a^*, b^* \in \mathbb{R}^{n-1}$ such that

$$\frac{\hat{a}^*(s)}{\hat{\lambda}(s)} = a_0^* + a^{*T} (sI - \Lambda)^{-1} b_\lambda$$

$$\frac{\hat{b}^*(s)}{\hat{\lambda}(s)} = b_0^* + b^{*T} (sI - \Lambda)^{-1} b_\lambda \tag{2.6.10}$$

We define the observer through

$$\dot{w}^{(1)} = \Lambda w^{(1)} + b_\lambda r$$

$$\dot{w}^{(2)} = \Lambda w^{(2)} + b_\lambda y_p \tag{2.6.11}$$

where $w^{(1)}, w^{(2)} \in \mathbb{R}^{n-1}$ and the initial conditions are arbitrary. The regressor vector is now

$$w^T(t) := \left[ r(t), w^{(1)T}(t), y_p(t), w^{(2)T}(t) \right] \in \mathbb{R}^{2n} \tag{2.6.12}$$

and the nominal parameter is

$$\theta^{*T} := \left[ a_0^*, a^{*T}, b_0^*, b^{*T} \right] \in \mathbb{R}^{2n} \tag{2.6.13}$$

Using proposition 2.6.1, we find that any plant can be represented uniquely as

$$y_p(t) = \hat{M} \left[ \theta^{*T} w(t) \right] \tag{2.6.14}$$

Rigorously, one should add an exponentially decaying term due to the initial conditions in the observer and in the reference model transfer function, but we will neglect this term for simplicity.

In similarity with (2.6.14), the *identifier output* is defined to be

$$y_i(t) := \hat{M} \left[ a_0(t) r(t) + a^T(t) w^{(1)}(t) + b_0(t) y_p(t) \right.$$

$$\left. + b^T(t) w^{(2)}(t) \right]$$

$$:= \hat{M} \left[ \theta^T(t) w(t) \right] \tag{2.6.15}$$

where

$$\theta^T(t) := \left[ a_0(t), a^T(t), b_0(t), b^T(t) \right] \in \mathbb{R}^{2n} \tag{2.6.16}$$

is the vector of adaptive parameters. We define the parameter error

$$\phi(t) := \theta(t) - \theta^* \tag{2.6.17}$$

so that the *identifier error*

$$e_1(t) = y_i(t) - y_p(t) \tag{2.6.18}$$

is also given, for the purpose of the analysis, by

$$e_1(t) = \hat{M} \left[ \phi^T(t) w(t) \right] \tag{2.6.19}$$

The error equation is linear in the parameter error, but it now involves the dynamics of the transfer function $\hat{M}$. We will show in the next section that the same gradient algorithm as used previously can still be used for identification, provided that the following additional assumption is satisfied.

**(A4)   Positive Real Model**

$\hat{M}$ is strictly positive real.

We will define and discuss this assumption in the next section. The overall identifier structure is represented in Figure 2.4 and we summarize the implementation of the algorithm hereafter.

**Model Reference Identifier—Implementation**

*Assumptions*

    (A1)–(A4)

*Data*

    $n$, upper bound on $n - m$ (e.g., $n$)
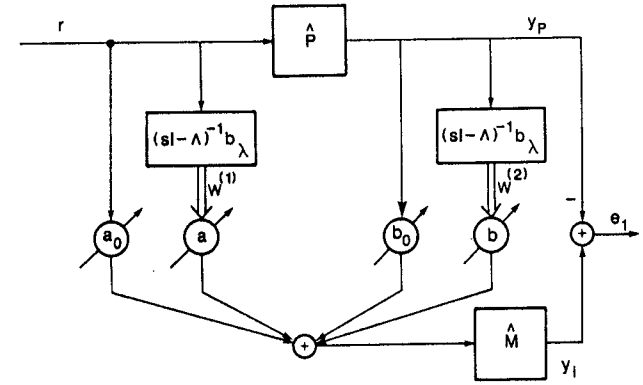
*Input*



**Figure 2.4:**   Model Reference Identifier Structure

    $r(t), y_p(t) \in \mathbb{R}$

*Output*

    $\theta(t), y_i(t) \in \mathbb{R}$

*Internal Signals*

    $w(t) \in \mathbb{R}^{2n} [ w^{(1)}(t), w^{(2)}(t) \in \mathbb{R}^{n-1} ]$

    $\theta(t) \in \mathbb{R}^{2n} [ a_0(t) \in \mathbb{R}, a(t) \in \mathbb{R}^{n-1}, b_0(t) \in \mathbb{R}, b(t) \in \mathbb{R}^{n-1} ]$

    $y_i(t), e_1(t) \in \mathbb{R}$

    Initial conditions are arbitrary.

*Design Parameters*

- $\hat{M}$ satisfying (A3)–(A4)
- $\Lambda \in \mathbb{R}^{n-1 \times n-1}, b_\lambda \in \mathbb{R}^{n-1}$ in controllable canonical form such that $\det(sI - \Lambda)$ is Hurwitz, and contains the zeros of $\hat{n}_m(s)$.
- $g > 0$

*Identifier Structure*

$$\dot{w}^{(1)} = \Lambda w^{(1)} + b_\lambda r$$

$$\dot{w}^{(2)} = \Lambda w^{(2)} + b_\lambda y_p$$

$$\theta^T = [ a_0, a^T, b_0, b^T ]$$

$$w^T = [ r, w^{(1)^T}, y_p, w^{(2)^T} ]$$

$$y_i = \hat{M}(\theta^T w)$$

$$e_1 = y_i - y_p$$

*Gradient Algorithm*

$$\dot{\theta} = -g\,e_1 w$$

☐

## Comments

The model reference identifier is an alternative to the equation error scheme discussed previously. It may result in savings in computations. Indeed, the state variable filters are of order $n - 1$, instead of $n$ previously. If the transfer function $\hat{M}$ is of order 1, the realization requires one less state. The presence of $\hat{M}$, however, precludes the use of the least-squares algorithms, which are usually faster. We will also show in Chapter 4 how the transfer function $\hat{M}$ influences the convergence properties of the gradient algorithm.

### 2.6.2 Strictly Positive Real Error Equation and Identification Algorithms

In Section 2.6.1, we encountered a more general error equation, which we will call the *strictly positive real* (SPR) error equation

$$e_1(t) = \hat{M}\left[\phi^T(t)w(t)\right]$$

where $\hat{M}$ is a strictly positive real transfer function. This error equation is still linear, but it now involves additional dynamics contained in $\hat{M}$. In this section, we will establish general properties involving this error equation. For uniformity with previous discussions, we assume that $w : \mathbb{R}_+ \rightarrow \mathbb{R}^{2n}$, but the dimension of $w$ is in fact completely arbitrary.

The definitions of *positive real* (PR) and *strictly positive real* (SPR) functions originate from network theory. A rational transfer function is the driving point impedance of a *passive* network if and only if it is PR. Similarly, it is the driving point impedance of a *dissipative* network if and only if it is SPR. The following definitions are deduced from these properties.

### Definition   Positive Real (PR) and Strictly Positive Real (SPR) Functions

A rational function $\hat{M}(s)$ of the complex variable $s = \sigma + j\omega$ is *positive real* (PR), if $\hat{M}(\sigma) \in \mathbb{R}$ for all $\sigma \in \mathbb{R}$ and $\text{Re}(\hat{M}(\sigma + j\omega)) \geq 0$ for all $\sigma > 0$, $\omega \geq 0$. It is *strictly positive real* (SPR) if, for some $\epsilon > 0$, $\hat{M}(s - \epsilon)$ is PR.

It may be shown (cf. Ioannou & Tao [1987]) that a *strictly proper* transfer function $\hat{M}(s)$ *is SPR if and only if*

- $\hat{M}(s)$ is stable
- $\text{Re}(\hat{M}(j\omega)) > 0$, for all $\omega \geq 0$
- $\lim\limits_{\omega \to \infty} \omega^2 \text{Re}(\hat{M}(j\omega)) > 0$

For example, the transfer function

$$\hat{M}(s) = \frac{s + c}{(s + a)(s + b)}$$

is SPR if and only if $a > 0$, $b > 0$, $a + b > c > 0$.

SPR *transfer functions* form a rather restricted class. In particular, an SPR transfer function must be minimum phase and its phase may never exceed 90°. An important lemma concerning SPR transfer functions is the *Kalman-Yacubovitch-Popov lemma* given next.

### Lemma 2.6.2   Minimal Realization of an SPR Transfer Function

Let $[A, b, c^T]$ be a minimal realization of a strictly proper, stable, rational transfer function $\hat{M}(s)$. Then, the following statements are equivalent

(a)     $\hat{M}(s)$ is *SPR*

(b)     There exist symmetric positive definite matrices $P$, $Q$, such that

$$PA + A^T P = -Q$$

$$Pb = c \qquad (2.6.20)$$

**Proof of Lemma 2.6.2**   cf. Anderson & Vongpanitlerd [1973].

### SPR Error Equation with Gradient Algorithm

A remarkable fact about SPR transfer functions is that the gradient update law

$$\dot{\phi}(t) = \dot{\theta}(t) = -g e_1(t) w(t) \qquad g > 0 \qquad (2.6.21)$$

has similar properties when $e_1$ is defined by the SPR error equation (2.6.19), and when $e_1$ is defined by the linear error equation (2.4.3). Note that in the case of the SPR error equation, the algorithm is *not* the true gradient anymore, although we will keep using this terminology for the similarity.

Using lemma 2.6.2, a state-space realization of $\hat{M}(s)$ with state $e_m$ can be obtained so that

$$\dot{e}_m(t) = A e_m(t) + b\phi^T(t)w(t)$$

$$e_1(t) = c^T(t) e_m(t)$$

$$\dot{\phi}(t) = -gc^T e_m(t) w(t) \quad g > 0 \qquad (2.6.22)$$

$$\stackrel{\wedge}{(t)}$$

### Theorem 2.6.3    SPR Error Equation with Gradient Algorithm

Let $w : \mathbb{R}_+ \to \mathbb{R}^{2n}$ be piecewise continuous. Consider the SPR error equation (2.6.19) with $\hat{M}(s)$ SPR, together with the gradient update law (2.6.21). Equivalently, consider the state-space realization (2.6.22) where $[A, b, c^T]$ satisfy the conditions of lemma 2.6.2.

*Then*

(a)        $e_m, e_1 \in L_2$

(b)        $e_m, e_1, \phi \in L_\infty$

### Proof of Theorem 2.6.3

Let $P, Q$ be as in lemma 2.6.2 and $v = ge_m^T P e_m + \phi^T \phi$. Along the trajectories of (2.6.22)

$$\dot{v} = ge_m^T PA e_m + ge_m^T Pb\phi^T w + ge_m^T A^T Pe_m$$

$$+ g\phi^T wb^T Pe_m - 2gc^T e_m \phi^T w$$

$$= -ge_m^T Q e_m \leq 0 \qquad (2.6.23)$$

where we used (2.6.20). The conclusions follow as in theorem 2.4.1, since $P$ and $Q$ are positive definite.    □

### Modified SPR Error Equation

The normalized gradient update law presented for the linear error equation is not usually applied to the SPR error equation. Instead, a *modified SPR error equation* is considered

$$e_1(t) = \hat{M}\left[ \phi^T(t) w(t) - \gamma w^T(t) w(t) e_1(t) \right] \quad \gamma > 0 \qquad (2.6.24)$$

where $\gamma$ is a constant. The same gradient algorithm may be applied with this error equation, so that in state-space form

$$\dot{e}_m(t) = A e_m(t) + b \left[ \phi^T(t) w(t) - \gamma w^T(t) w(t) c^T e_m(t) \right]$$

$$e_1(t) = c^T e_m(t)$$

$$\dot{\phi}(t) = -gc^T e_m(t) w(t) \qquad g, \gamma > 0 \qquad (2.6.25)$$

### Theorem 2.6.4    Modified SPR Error Equation with Gradient Algorithm

Let $w : \mathbb{R}_+ \to \mathbb{R}^{2n}$ be piecewise continuous. Consider the modified SPR error equation (2.6.24) with $\hat{M}(s)$ SPR, together with the gradient update law (2.6.21). Equivalently, consider the state-space realization (2.6.25), where $[A, b, c^T]$ satisfy the conditions of lemma 2.6.2.

*Then*

(a)        $e_m, e_1, \dot{\phi} \in L_2$

(b)        $e_m, e_1, \phi \in L_\infty$

### Proof of Theorem 2.6.4

Let $P, Q$ be as in lemma 2.6.2 and $v = ge_m^T P e_m + \phi^T \phi$. Along the trajectories of (2.6.25)

$$\dot{v} = -ge_m^T Q e_m - 2g\gamma (e_1 w)^T (e_1 w) \leq 0 \qquad (2.6.26)$$

Again, it follows that $e_m, e_1, \phi$ are bounded, and $e_m, e_1 \in L_2$. Moreover, it also follows now that $e_1 w \in L_2$, so that $\dot{\phi} \in L_2$.    □

### 2.6.3    Exponential Convergence of the Gradient Algorithms with SPR Error Equations

As stated in the following theorem, the gradient algorithm is also exponentially convergent with the SPR error equations, under the PE condition.

### Theorem 2.6.5 Exponential Convergence of the Gradient Algorithms with SPR Error Equations

Let $w : \mathbb{R}_+ \to \mathbb{R}^{2n}$. Let $[A, b, c^T]$ satisfy the conditions of lemma 2.6.2.

*If*        $w$ is PE and $w, \dot{w} \in L_\infty$

*Then*    (2.6.22) and (2.6.25) are globally exponentially stable.

The proof given hereafter is similar to the proof by Anderson [1977] (with some significant differences however). The main condition for exponential convergence is the PE condition, as required previously, and again, the main idea is to interpret the condition as a UCO condition. The additional boundedness requirement on $\dot{w}$ guarantees that PE is not lost through the transfer function $\hat{M}$ (cf. lemma 2.6.7 hereafter). It is sufficient that the boundedness conditions hold almost everywhere, so that piecewise differentiable signals may be considered.

## Auxiliary Lemmas on PE Signals

The following auxiliary lemmas will be useful in proving the theorem. Note that the sum of two PE signals is not necessarily PE. On the other hand, an $L_2$ signal is necessarily *not* PE. Lemma 2.6.6 asserts that PE is not altered by the addition of a signal belonging to $L_2$. In particular, this implies that terms due to initial conditions do not affect PE. Again, we assume the dimension of the vectors to be $2n$, for uniformity, but the dimension is in fact arbitrary.

## Lemma 2.6.6   PE and $L_2$ Signals

Let $w, e : \mathbb{R}_+ \to \mathbb{R}^{2n}$ be piecewise continuous.

*If*      $w$ is PE

         $e \in L_2$

*Then*    $w + e$ is PE.

**Proof of Lemma 2.6.6**    in the Appendix.

Lemma 2.6.7 shows that PE is not lost if the signal is filtered by a stable, minimum phase transfer function, provided that the signal is sufficiently smooth.

## Lemma 2.6.7   PE Through LTI Systems

Let $w : \mathbb{R}_+ \to \mathbb{R}^{2n}$.

*If*      $w$ is PE and $w, \dot{w} \in L_\infty$

         $\hat{H}$ is a stable, minimum phase, rational transfer function

*Then*    $\hat{H}(w)$ is PE.

**Proof of Lemma 2.6.7**    in the Appendix.

We now prove theorem 2.6.5.

## Proof of Theorem 2.6.5

As previously, let $v = g e_m^T P e_m + \phi^T \phi$, so that for both SPR error equations

$$\int_{t_0}^{t_0+\delta} \dot{v} \, d\tau \leq -g \int_{t_0}^{t_0+\delta} e_m^T Q e_m \, d\tau$$

$$\leq -g \frac{\lambda_{\min}(Q)}{|c|^2} \int_{t_0}^{t_0+\delta} e_1^2 \, d\tau \leq 0 \qquad (2.6.27)$$

By theorem 1.5.2, exponential convergence will be guaranteed if, for some $\alpha_3 > 0$

$$\int_{t_0}^{t_0+\delta} e_1^2(\tau) \, d\tau \geq \alpha_3 \left( |e_m(t_0)|^2 + |\phi(t_0)|^2 \right) \qquad (2.6.28)$$

for all $t_0, e_m(t_0), \phi(t_0)$.

**Derivation of 2.6.28**: This condition can be interpreted as a UCO condition on the system

$$\dot{e}_m = A e_m + b \phi^T w$$

$$\dot{\phi} = -g c^T e_m w$$

$$e_1 = c^T e_m \qquad (2.6.29)$$

An additional term $-b\gamma w^T w c^T e_m$ is added in the differential equation governing $e_m$ in the case of the modified SPR error equation. Using lemma 2.5.2 about UCO under output injection, we find that inequality (2.6.28) will be satisfied if the following system

$$\dot{e}_m = A e_m + b \phi^T w$$

$$\dot{\phi} = 0$$

$$e_1 = c^T e_m \qquad (2.6.30)$$

is UCO. For this, we let

$$K = \begin{bmatrix} 0 \\ gw \end{bmatrix} \quad \text{or} \quad K = \begin{bmatrix} b\gamma w^T w \\ gw \end{bmatrix} \qquad (2.6.31)$$

for the basic SPR and modified SPR error equations, respectively. The condition on $K$ in lemma 2.5.3 is satisfied, since $w$ is bounded.

We are thus left to show that system (2.6.30) is UCO, i.e. that

$$e_1(t) = c^T e^{A(t-t_0)} e_m(t_0) + \int_{t_0}^{t} c^T e^{A(t-\tau)} b \, w^T(\tau) \, d\tau \, \phi(t_0)$$

$$:= x_1(t) + x_2(t) \qquad (2.6.32)$$

satisfies, for some $\beta_1, \beta_2, \delta > 0$

$$\beta_2 \left[ |e_m(t_0)|^2 + |\phi(t_0)|^2 \right]$$

$$\geq \int_{t_0}^{t_0+\delta} e_1^2(\tau) \, d\tau$$

$$\geq \beta_1 \left[ |e_m(t_0)|^2 + |\phi(t_0)|^2 \right] \qquad (2.6.33)$$

for all $t_0$, $e_m(t_0)$, $\phi(t_0)$.

**Derivation of 2.6.33:** By assumption, $w$ is PE and $w$, $\dot{w} \in L_\infty$. Therefore, using lemma 2.6.7, we have that, for all $t_0 \geq 0$, the signal

$$w_f(t) = \int_{t_0}^{t} c^T e^{A(t-\tau)} b \, w(\tau) \, d\tau \qquad (2.6.34)$$

is PE. This means that, for some $\alpha_1, \alpha_2, \sigma > 0$

$$\alpha_2 |\phi(t_0)|^2 \geq \int_{t_1}^{t_1+\sigma} x_2^2(\tau) \, d\tau \geq \alpha_1 |\phi(t_0)|^2 \qquad (2.6.35)$$

for all $t_1 \geq t_0 \geq 0$ and $\phi(t_0)$.

On the other hand, since $A$ is stable, there exist $\gamma_1, \gamma_2 > 0$, such that

$$\int_{t_0+m\sigma}^{\infty} x_1^2(\tau) \, d\tau \leq \gamma_1 |e_m(t_0)|^2 e^{-\gamma_2 m\sigma} \qquad (2.6.36)$$

for all $t_0$, $e_m(t_0)$ and an arbitrary integer $m > 0$ to be defined later. Since $[A \ c^T]$ is observable, there exists $\gamma_3(m\sigma) > 0$, with $\gamma_3(m\sigma)$ increasing with $m\sigma$ such that

$$\int_{t_0}^{t_0+m\sigma} x_1^2(\tau) \, d\tau \geq \gamma_3(m\sigma) |e_m(t_0)|^2 \qquad (2.6.37)$$

for all $t_0$, $e_m(t_0)$ and $m > 0$.

Let $n > 0$ be another integer to be defined and let $\delta = (m+n)\sigma$. Using the ~~triangle inequality~~ Fact that $(a \cdot b)^2 \geq a^2/2 - b^2$

$$\int_{t_0}^{t_0+\delta} e_1^2(\tau) \, d\tau \geq \frac{1}{2} \int_{t_0}^{t_0+m\sigma} x_1^2(\tau) \, d\tau - \int_{t_0}^{t_0+m\sigma} x_2^2(\tau) \, d\tau$$

$$+ \frac{1}{2} \int_{t_0+m\sigma}^{t_0+\delta} x_2^2(\tau) \, d\tau - \int_{t_0+m\sigma}^{t_0+\delta} x_1^2(\tau) \, d\tau$$

$$\geq \frac{1}{2} \gamma_3(m\sigma) |e_m(t_0)|^2 - m\alpha_2 |\phi(t_0)|^2$$

$$+ \frac{1}{2} n\alpha_1 |\phi(t_0)|^2 - \gamma_1 e^{-\gamma_2 m\sigma} |e_m(t_0)|^2 \qquad (2.6.38)$$

Let $m$ be large enough to get

$$\frac{1}{2} \gamma_3(m\sigma) - \gamma_1 e^{-\gamma_2 m\sigma} \geq \gamma_3(m\sigma)/4 \qquad (2.6.39)$$

and $n$ sufficiently large to obtain

$$\frac{1}{2} n\alpha_1 - m\alpha_2 \geq \alpha_1 \qquad (2.6.40)$$

Further, define

$$\beta_1 = \min(\alpha_1, \gamma_3(m\sigma)/4) \qquad (2.6.41)$$

The lower inequality in (2.6.33) follows from (2.6.38), with $\beta_1$ as defined, while the upper inequality is easily found to be valid with

$$\beta_2 = \max(\gamma_1, (m+n)\alpha_2) \qquad (2.6.42)$$

$\square$

**Comments**

a) Although the proof of theorem 2.6.5 is somewhat long and tedious, it has some interesting features. First, it relies on the same basic idea as the proof of exponential convergence for the linear error equation (cf. theorem 2.5.1). It interprets the condition for exponential convergence as a uniform complete observability condition. Then, it uses lemma 2.5.2 concerning UCO under output injection to transform the UCO condition to a UCO condition on a similar system, but *where the vector $\phi$ is constant* (cf. (2.6.30)). The UCO condition leads then to a PE condition on a vector $w_f$, which is a filtered version of $w$, through the LTI system $\hat{M}(s)$.

b) The steps of the proof can be followed to obtain guaranteed rates of exponential convergence. Although such rates would be useful to the designer, the expression one obtains is quite complex and examination of the proof leaves little hope that the estimate would be tight. A more successful approach is found in Chapter 4, using averaging techniques.

c) The results presented in this section are very general. They do not rely on the structure of the identifier, but only on the SPR error equation (2.6.19). We leave it to the reader to specialize these results and obtain stability and convergence properties of the model reference identifier of Section 2.6.1.

## 2.7   Frequency Domain Conditions for Parameter Convergence

Theorems 2.5.3 and 2.6.5 give a condition on the regressor vector $w$, namely that $w$ be PE, to guarantee exponential convergence of the parameter error. The difficulty with the PE condition is that it is not an *explicit* condition on the reference input $r$. In this section, we give frequency domain conditions on the input $r$ to guarantee that $w$ is PE.

The result of this section will make precise the following intuitive argument: assume that the parameter vector *does* converge (but not necessarily to the nominal value). Then, the plant loop is "asymptotically time-invariant." If the reference input contains frequencies $\omega_1, \ldots$ , $\omega_k$, we expect that $y_p$ and $y_i$ will too. Since $y_i \to y_p$ as $t \to \infty$, the asymptotic identifier transfer function must match the plant transfer function at $s = j\omega_1, \ldots , j\omega_k$. If $k$ is large enough, this will imply that the asymptotic identifier transfer function is *precisely* the plant transfer function and, therefore, that the parameter error converges to zero. Thus, we will show that the reference signal must be "rich enough," that is, "contains enough frequencies," for the parameter error to converge to zero. Roughly speaking, we will show

> A reference input $r$ results in parameter error convergence to zero *unless* its spectrum is concentrated on $k < 2n$ points, where $2n$ is the number of unknown parameters in the adaptive scheme.

We will also discuss partial parameter convergence when the input is not sufficiently rich. We will use the results of generalized harmonic analysis developed in Section 1.6 and restrict our attention to stationary reference signals.

### 2.7.1  Parameter Convergence

From the definition of the regressor $w$ in (2.2.16), we see that $w$ is the output of a LTI system with input $r$ and transfer function

$$\hat{H}_{wr}(s) = \begin{bmatrix} (sI - \Lambda)^{-1} b_\lambda \\ (sI - \Lambda)^{-1} b_\lambda \hat{P}(s) \end{bmatrix} \tag{2.7.1}$$

We will assume that the input $r$ is stationary and that $\hat{P}$ is stable. Then, by the linear filter lemma (proposition 1.6.2), it follows that $w$ is also stationary, that is has an autocovariance. For stationary signals, persistency of excitation is directly related to the positive definiteness of the autocovariance, as stated in the following proposition. Note that $R_w(0)$ is a symmetric positive semidefinite matrix.

### Proposition 2.7.1   PE and Autocovariance

Let $w(t) \in \mathbb{R}^{2n}$ be stationary.

(a) $w$ is PE          *if and only if*          (b) $R_w(0) > 0$.

### Proof of Proposition 2.7.1

*(a) implies (b)*

From the PE condition on $w$, we have that for all $c \in \mathbb{R}^{2n}$ and for any positive integer $k$

$$\frac{1}{k\delta} \int_{t_0}^{t_0 + k\delta} (w^T(\tau)c)^2 \, d\tau \geq \frac{\alpha_1}{\delta} |c|^2 \tag{2.7.2}$$

Hence, for all $T \geq \delta$

$$\frac{1}{T} \int_{t_0}^{t_0 + T} (w^T(\tau)c)^2 \, d\tau \geq \frac{T - \delta}{T} \frac{\alpha_1}{\delta} |c|^2 \tag{2.7.3}$$

Since $w$ has an autocovariance

$$\lim_{T \to \infty} \frac{1}{T} \int_{t_0}^{t_0 + T} (w^T(\tau)c)^2 \, d\tau = c^T R_w(0) c \tag{2.7.4}$$

Combining (2.7.3) and (2.7.4) yields that $R_w(0) \geq \dfrac{\alpha_1}{\delta} I > 0$.

*(b) implies (a)*

From the definition of autocovariance, for all $\epsilon > 0$, there exists $\delta$ such that

$$\left| \frac{1}{\delta} \int_{t_0}^{t_0 + \delta} (w^T(\tau)c)^2 \, d\tau - c^T R_w(0) c \right| < \epsilon \tag{2.7.5}$$

Therefore

$$\frac{1}{\delta} \int_{t_0}^{t_0 + \delta} (w^T(\tau)c)^2 \, d\tau \geq \lambda_{\min}(R_w(0))|c|^2 - \epsilon \tag{2.7.6}$$

which implies PE for $\epsilon$ sufficiently small. $\square$

We may relate the PE condition to the frequency content of $w$ through the formula

$$R_w(0) = \int S_w(d\omega) \tag{2.7.7}$$

where $S_w(d\omega)$ is the spectral measure of $w$. In turn, using the linear

filter lemma (proposition 1.6.2) and (2.7.7), we see that

$$R_w(0) = \int \hat{H}_{wr}^*(j\omega)\hat{H}_{wr}^T(j\omega)S_r(d\omega) \qquad (2.7.8)$$

The expression (2.7.8) allows us to relate the frequency content of $r$ to the persistency of excitation of $w$. This leads us to the following definition.

### Definition  Sufficient Richness of Order k

A stationary signal $r : \mathbb{R}_+ \to \mathbb{R}$ is called *sufficiently rich of order k*, if the support of the spectral density of $r$, namely, $S_r(d\omega)$ contains at least $k$ points.

### Comment

Note that a single sinusoid in the input contributes 2 points to the spectrum: at $+\omega_0$ and at $-\omega_0$. On the other hand, a DC signal contributes only one point in the support of the spectral density.

### Theorem 2.7.2    PE and Sufficient Richness

Let $w(t) \in \mathbb{R}^{2n}$ be the output of a stable LTI system with transfer function $\hat{H}_{wr}(s)$ and stationary input $r(t)$. Assume that $\hat{H}_{wr}(j\omega_1), \ldots, \hat{H}_{wr}(j\omega_{2n})$, are linearly independent on $\mathbb{C}^{2n}$ for all $\omega_1, \omega_2, \ldots, \omega_{2n} \in \mathbb{R}$.

(a) $w$ is PE          *if and only if*          (b) $r$ is sufficiently rich of order $2n$.

### Proof of Theorem 2.7.2

By proposition 2.7.1, (a) is equivalent to $R_w(0) > 0$.

*(a) implies (b)*

We prove this by contradiction. Assume that $r$ is not sufficiently rich of order $2n$, that is the support of $S_r(d\omega)$ is $\omega_1, \ldots, \omega_k$ with $k < 2n$. Then, from (2.7.8), it follows that

$$R_w(0) = \sum_{i=1}^{k} \hat{H}_{wr}^*(j\omega_i)\hat{H}_{wr}^T(j\omega_i)S_r(\{\omega_i\}) \qquad (2.7.9)$$

The right hand side of (2.7.9) is the sum of $k$ dyads, so that the rank of $R_w(0)$ is at most $k < 2n$, contradicting the PE of $w$.

*(b) implies (a)*

We also prove this by contradiction. If $w$ is not PE, there exists $c \in \mathbb{R}^{2n}$ such that

$$c^T R_w(0) c = 0 \qquad (2.7.10)$$

Using (2.7.8), we see that (2.7.10) implies that

$$\int |\hat{H}_{wr}^T(j\omega)c|^2 S_r(d\omega) = 0 \qquad (2.7.11)$$

Since $S_r(d\omega)$ is a continuous, positive measure, we may conclude that

$$\hat{H}_{wr}^T(j\omega)c = 0 \qquad \text{for all } \omega \in \text{support } S_r(d\omega) \qquad (2.7.12)$$

Since the support of $S_r(d\omega)$ has at least $2n$ points, say $\omega_1, \ldots, \omega_{2n}$, we have that

$$\hat{H}_{wr}^T(j\omega_i)c = 0 \qquad i = 1, \ldots, 2n \qquad (2.7.13)$$

contradicting the hypothesis of linear independence of $\hat{H}_{wr}(j\omega_i)$ for $i = 1, \ldots, 2n$. $\square$

We saw that sufficient richness of order $2n$ in the reference input translates to $w$ PE, provided that the $\hat{H}_{wr}(j\omega_i)$ are independent for every set of $2n$ $\omega_i$'s. It remains to be verified that this property holds for the $\hat{H}_{wr}(s)$ given in (2.7.1).

### Theorem 2.7.3    Exponential Parameter Convergence and Sufficient Richness

Consider the identification problem of Section 2.1, with assumptions (A1) and (A2) and $\hat{P}$ stable, the identifier structure of Section 2.2 and the gradient algorithms (2.4.1) or (2.4.2), or the normalized LS algorithm with covariance resetting (2.4.6)–(2.4.7).

*If*          $r$ is stationary and sufficiently rich of order $2n$

*Then*          the identifier parameter $\theta$ converges to the nominal parameter $\theta^*$ exponentially fast.

### Proof of Theorem 2.7.3

Using theorems 2.5.3 and 2.7.2, we are left to show that, for every $\omega_1, \ldots, \omega_{2n} \in \mathbb{R}$, the vectors $\hat{H}_{wr}(j\omega_i)$ (with $\hat{H}_{wr}$ defined in (2.7.1)) are linearly independent in $\mathbb{C}^{2n}$.

We show the result by contradiction. Assume that there existed $\omega_1, \ldots, \omega_{2n}$ such that

$$\hat{H}_{wr}^T(j\omega_i)c = 0 \qquad (2.7.14)$$

for some $c \in \mathbb{C}^{2n}$ and for all $i = 1, \ldots, 2n$. Using the fact that $(\Lambda, b_\lambda)$ are in controllable canonical form and that $\hat{P}(s) = k_p \hat{n}_p(s)/\hat{d}_p(s)$,

(2.7.14) becomes

$$\left[ 1, j\omega_i, \ldots, (j\omega_i)^{n-1}, k_p \frac{\hat{n}_p(j\omega_i)}{\hat{d}_p(j\omega_i)}, \right.$$

$$\left. (j\omega_i) k_p \frac{\hat{n}_p(j\omega_i)}{\hat{d}_p(j\omega_i)}, \ldots, (j\omega_i)^{n-1} k_p \frac{\hat{n}_p(j\omega_i)}{\hat{d}_p(j\omega_i)} \right] c = 0 \qquad (2.7.15)$$

for $i = 1, \ldots, 2n$, that is

$$[\hat{d}_p(s), s\hat{d}_p(s), \ldots, s^{n-1}\hat{d}_p(s), k_p\hat{n}_p(s),$$

$$k_p s \hat{n}_p(s), \ldots, s^{n-1} k_p \hat{n}_p(s)] c = 0 \qquad (2.7.16)$$

for $s = j\omega_1, \ldots, j\omega_{2n}$. Equation (2.7.16) may be written more compactly as

$$\hat{c}^{(1)}(s)\hat{d}_p(s) + \hat{c}^{(2)}(s) k_p \hat{n}_p(s) = 0 \qquad (2.7.17)$$

for $s = j\omega_1, \ldots, j\omega_{2n}$, where

$$\hat{c}^{(1)}(s) = c_1 + c_2 s + \cdots + c_n s^{n-1}$$

and

$$\hat{c}^{(2)}(s) = c_{n+1} + c_{n+2} s + \cdots + c_{2n} s^{2n-1} \qquad (2.7.18)$$

The polynomial on the left-hand side of (2.7.17) is of degree at most $2n - 1$. Since it vanishes at $2n$ points, it must be identically zero. Consequently, we have

$$\frac{k_p \hat{n}_p(s)}{\hat{d}_p(s)} = - \frac{\hat{c}^{(1)}(s)}{\hat{c}^{(2)}(s)} \qquad \text{for all } s \qquad (2.7.19)$$

By assumption, $\hat{d}_p(s)$ is of degree $n$ and $\hat{n}_p(s)$, $\hat{d}_p(s)$ are coprime. Therefore, (2.7.19) cannot be satisfied since the degree of $\hat{c}^{(2)}$ is at most $n - 1$. This establishes the contradiction. □

**Comments**

It would appear that the developments of this section would enable us to give an explicit expression for the rate of exponential convergence: it is in fact possible to carry out this program rigorously when the rate of adaptation is slow (see Chapter 4 for a further and more detailed description of this point).

   If $w$ is *not* PE, then the parameter error need not converge to zero. In this case, $S_r$ is concentrated on $k < 2n$ frequencies $\omega_1, \ldots, \omega_k$. Intuition suggests that although $\theta$ need not converge to $\theta^*$, it should

converge to the set of $\theta$'s for which the *identifier transfer function matches the plant transfer function* at the frequencies $s = j\omega_1, \ldots, j\omega_k$. This is indeed the case.

### 2.7.2 Partial Parameter Convergence

We now consider the case when the spectrum $S_r(d\omega)$ is concentrated on $k < 2n$ points. Before stating the theorem, let us discuss the idea informally. From the structure of the identifier, we see that the plant output is given by

$$y_p = \theta^{*T} \hat{H}_{wr}(r) \qquad (2.7.20)$$

and the identifier output by

$$y_i = \theta^T(t) \hat{H}_{wr}(r) \qquad (2.7.21)$$

Consequently, if the identifier output matches the plant output at $s = j\omega_1, \ldots, j\omega_k$, the asymptotic value of $\theta$ must satisfy

$$\begin{bmatrix} \hat{H}_{wr}(j\omega_1)^T \\ \cdot \\ \cdot \\ \cdot \\ \hat{H}_{wr}(j\omega_k)^T \end{bmatrix} \theta = \begin{bmatrix} \hat{H}_{wr}(j\omega_1)^T \\ \cdot \\ \cdot \\ \cdot \\ \hat{H}_{wr}(j\omega_k)^T \end{bmatrix} \theta^* \qquad (2.7.22)$$

Let us then call $\Theta$, the set of $\theta$'s, which satisfy (2.7.22). Clearly, $\theta^* \in \Theta$ and

$$\Theta = \theta^* + \text{null space} \begin{bmatrix} \hat{H}_{wr}(j\omega_1)^T \\ \cdot \\ \cdot \\ \cdot \\ \hat{H}_{wr}(j\omega_k)^T \end{bmatrix} \qquad (2.7.23)$$

The $k$ row vectors $\hat{H}_{wr}(j\omega_i)^T$ are linearly independent. Consequently, $\Theta$ has dimension $2n - k$.

   In terms of the parameter error vector $\phi = \theta - \theta^*$, $\Theta$ has the simple description

$$\theta \in \Theta \iff R_w(0)\phi = 0 \qquad (2.7.24)$$

We leave the verification of this fact to the reader, recalling that

$$R_w(0) = \sum_{i=1}^{k} \hat{H}_{wr}(j\omega_i)\hat{H}_{wr}^*(j\omega_i)S_r(\{\omega_i\}) \qquad (2.7.25)$$

Proceeding more rigorously, we have the following theorem.

### Theorem 2.7.4 Partial Convergence Theorem

Consider the identification problem of Section 2.1, with assumptions (A1)–(A2) and $\hat{P}$ stable, the identifier structure of Section 2.2 and the gradient algorithms (2.4.1) or (2.4.2), or the normalized LS algorithm with covariance resetting (2.4.6)–(2.4.7).

*If* $r$ is stationary

*Then* $\lim_{t\to\infty} R_w(0)\phi(t) = 0$

### Proof of Theorem 2.7.4

The proof relies on the fact, proved in theorem 2.4.5, that the output error $e_1 = \phi^T w \to 0$ as $t \to \infty$, and so does the derivative of the parameter error $\dot{\phi}$. Since $\phi$ and $w$ are bounded, let $k$ be such that $|\phi(t)|$, $|w(t)| \le k$, for all $t \ge 0$.

We will show that, for all $\epsilon > 0$, there exists $t_1 \ge 0$ such that, for all $t \ge t_1$, $\phi^T(t)R_w(0)\phi(t) \le \epsilon$. This means that $\phi^T(t)R_w(0)\phi(t)$ converges to zero as $t \to \infty$ and since $R_w(0)$ is symmetric positive definite, it also implies that $|R_w(0)\phi(t)|$ tends to zero as $t \to \infty$.

Since $w$ has an autocovariance, for $T$ large enough

$$\left| R_w(0) - \frac{1}{T}\int_{t_0}^{t_0+T} w(\tau)w^T(\tau)\,d\tau \right| \le \frac{\epsilon}{3k^2} \qquad (2.7.26)$$

for all $t_0 \ge 0$ and therefore

$$\left| \phi(t)^T R_w(0)\phi(t) - \phi(t)^T\frac{1}{T}\int_{t_0}^{t_0+T} w(\tau)w(\tau)^T\,d\tau\,\phi(t) \right| \le \frac{\epsilon}{3} \qquad (2.7.27)$$

From the fact that $\dot{\phi} \to 0$ as $t \to \infty$ (cf. theorem 2.4.5) and $\phi^T w \to 0$ as $t \to \infty$, we find that there exists $t_1$ such that, for all $t \ge t_1$

$$|\phi^T(t)w(t)| \le \frac{\epsilon}{3} \qquad (2.7.28)$$

and

$$|\dot{\phi}(t)| \le \frac{\epsilon}{6k^3T} \qquad (2.7.29)$$

From (2.7.29), we have that $|\phi(\tau) - \phi(t)| \le \epsilon(\tau - t)/6k^3T$ for all $\tau \ge t \ge t_1$. Together with the boundedness of $\phi$ and $w$, this implies that, for $t \ge t_1$

$$\left| \phi(t)^T \left[\frac{1}{T}\int_{t}^{t+T} w(\tau)w^T(\tau)\,d\tau\right]\phi(t) - \frac{1}{T}\int_{t}^{t+T}\left[\phi^T(\tau)w(\tau)w^T(\tau)\phi(\tau)\right]d\tau \right|$$

$$= \left| \frac{1}{T}\int_{t}^{t+T} w^T(\tau)\,(\phi(t) - \phi(\tau))\,w^T(\tau)\,(\phi(t) + \phi(\tau))\,d\tau \right|$$

$$\le \frac{\epsilon}{3} \qquad (2.7.30)$$

Using (2.7.28), we have that, for $t \ge t_1$

$$\left| \frac{1}{T}\int_{t}^{t+T} \phi^T(\tau)w(\tau)w^T(\tau)\phi(\tau)\,d\tau \right| \le \frac{\epsilon}{3} \qquad (2.7.31)$$

Now, combining (2.7.27), (2.7.30) and (2.7.31), we have that, for $t \ge t_1$

$$\phi^T(t)R_w(0)\phi(t) \le \epsilon \qquad (2.7.32)$$

which completes the proof of the theorem. $\square$

### Comments

The proof relies on the fact that for both the identification schemes discussed *the parameter error eventually becomes orthogonal to the regressor w*, and the *updating slows down*. These are very common properties of identification schemes.

While the $2n - k$ dimensional set $\Theta$ to which $\theta(t)$ converges depends only on the frequencies $\omega_1, \ldots, \omega_k$ and not on the average powers $S_r(\{\omega_1\}), \ldots, S_r(\{\omega_k\})$ contained in the reference signal at those frequencies, the *rate of convergence* of $\theta$ to $\Theta$ depends on both.

As opposed to the proof of theorem 2.7.3, the proof of theorem 2.7.4 does not rely on theorem 2.5.3 relating PE and exponential convergence. If $w$ is PE and $R_w(0) > 0$, the proof of theorem 2.7.4 is an alternate proof of the parameter convergence results of Section 2.5, with the additional assumption of stationarity of the reference input $r(t)$.

## 2.8  CONCLUSIONS

In this chapter, we derived a simple identification scheme for SISO LTI plants. The scheme involved a generic linear error equation, relating the identifier error, the regressor and the parameter error. Several gradient and least-squares algorithms were reviewed and common properties were established, that are valid under general conditions. It was shown that for any of these algorithms and provided that the regressor was a bounded function of time, the identifier error converged to zero as $t$ approached infinity. The parameter error was also guaranteed to remain bounded. When the regressor was not bounded, but satisfied a regularity condition, then it was shown that a normalized error still converged to zero.

The exponential convergence of the parameter error to its nominal value followed from a persistency of excitation condition on the regressor. Guaranteed rates of exponential convergence were also obtained and showed the influence of various design parameters. In particular, the reference input was found to be a dominant factor influencing the parameter convergence.

The stability and convergence properties were further extended to strictly positive real error equations. Although more complex to analyze, the SPR error equation was found to have similar stability and convergence properties. In particular, PE appeared as a fundamental condition to guarantee exponential parameter convergence.

Finally, the PE conditions were transformed into conditions on the input. We assumed stationarity of the input, so that a frequency-domain analysis could be carried out. It was shown that parameter convergence was guaranteed, if the input contained the same number of spectral components as there were unknown parameters. If the input was a sum of sinusoids, for example, their number should be greater than or equal to the order of the plant.

# CHAPTER 3
# ADAPTIVE CONTROL

## 3.0   INTRODUCTION

In this chapter, we derive and analyze algorithms for adaptive control. Our attention is focused on model reference adaptive control. Then, the objective is to design an adaptive controller such that the behavior of the controlled plant remains close to the behavior of a desirable model, despite uncertainties or variations in the plant parameters. More formally, a *reference model* $\hat{M}$ is given, with input $r(t)$ and output $y_m(t)$. The unknown *plant* $\hat{P}$ has input $u(t)$ and output $y_p(t)$. The control objective is to design $u(t)$ such that $y_p(t)$ asymptotically tracks $y_m(t)$, with all generated signals remaining bounded.

We will consider linear time invariant systems of arbitrary order, and establish the stability and convergence properties of the adaptive algorithms. In this section however, we start with an informal discussion for a first order system with two unknown parameters. This will allow us to introduce the algorithms and the stability results in a simpler context.

Consider a first order single-input single-output (SISO) linear time invariant (LTI) plant with transfer function

$$\hat{P} = \frac{k_p}{s + a_p} \tag{3.0.1}$$

where $k_p$ and $a_p$ are unknown. The reference model is a stable SISO LTI system of identical order