written as linear regression models. Notice that the regressors do not need to be linear in the inputs and outputs. An example illustrates the idea.

EXAMPLE 2.3 Nonlinear system

Consider the model

$$y(t) + ay(t-1) = b_1 u(t-1) + b_2 \sin(u(t))$$

By introducing

$$\theta = (a \quad b_1 \quad b_2)^T$$

and

$$\varphi^T(t) = (-y(t) \quad u(t) \quad \sin(u(t)))$$

the model can be written as

$$y(t) = \varphi^T(t-1)\theta$$

The model is linear in the parameters, and the least-squares method can be used to estimate θ. □

Stochastic Models

The least-squares estimate is biased when it is used on data generated by Eq. (2.12), where the errors $e(i)$ are correlated. The reason is that $E\varphi^T(i)e(i) \neq 0$ (compare Eq. (2.13)). A possibility to cope with this problem is to model the correlation of the disturbances and to estimate the parameters describing the correlations. Consider the model

$$A(q)y(t) = B(q)u(t) + C(q)e(t) \tag{2.38}$$

where $A(q), B(q)$, and $C(q)$ are polynomials in the forward shift operator and $\{e(t)\}$ is white noise. The parameters of tbe polynomial C describe the correlation of the disturbance. The model of Eq. (2.38) cannot be converted directly to a regression model, since the variables $\{e(t)\}$ are not known. A regression model can, however, be obtained by suitable approximations. To describe these, introduce

$$\varepsilon(t) = y(t) - \varphi^T(t-1)\hat{\theta}(t-1)$$

where

$$\theta = (a_1 \quad \dots \quad a_n \quad b_1 \quad \dots \quad b_n \quad c_1 \quad \dots \quad c_n)$$

$$\varphi^T(t-1) = (-y(t-1) \ \dots \ -y(t-n) \ u(t-1) \ \dots \ u(t-n) \ \varepsilon(t-1) \ \dots \ \varepsilon(t-n))$$

The variables $e(t)$ are approximated by the prediction errors $\varepsilon(t)$. The model can then be approximated by

$$y(t) = \varphi^T(t-1)\theta + e(t)$$

and standard recursive least squares can be applied. The method obtained is called *extended least squares* (ELS). The equations for updating the estimates are given by

$$\begin{aligned}\hat{\theta}(t) &= \hat{\theta}(t-1) + P(t)\varphi(t-1)\varepsilon(t)\\ P^{-1}(t) &= P^{-1}(t-1) + \varphi(t-1)\varphi^T(t-1)\end{aligned} \tag{2.39}$$

(Compare with Theorem 2.3.) Another method of estimating the parameters in Eq. (2.38) is to use Eqs. (2.39) and let the residual be defined by

$$\hat{C}(q)\varepsilon(t) = \hat{A}(q)y(t) - \hat{B}(q)u(t) \tag{2.40}$$

and regression vector φ in Eqs. (2.39) be replaced by φ_f, where

$$\hat{C}(q)\varphi_f(t) = \varphi(t) \tag{2.41}$$

The most recent estimates should be used in these updates. The method obtained is then not truly recursive, since Eqs. (2.41) and (2.40) have to be solved from $t = 1$ for each measurement. The following approximations can be made:

$$\varepsilon(t) = y(t) - \varphi_f^T(t-1)\hat{\theta}(t-1)$$

This algorithm is called the *recursive maximum likelihood (RML) method*.

It is advantageous for both ELS and RML to replace the residual in the regression vector by the *posterior residual* defined as

$$\varepsilon_p(t) = y(t) - \varphi^T(t-1)\hat{\theta}(t)$$

that is, the latest value of $\hat{\theta}$ is used to compute ε_p.

Another possibility to model the correlated noise is to use the model

$$y(t) = \frac{B(q)}{A(q)}u(t) + \frac{C(q)}{D(q)}e(t)$$

instead of Eq. (2.38). Recursive parameter estimates for this model can be derived in the same way as for Eq. (2.38).

Details about the extended least-squares method and the recursive maximum likelihood method are found in the references at the end of the chapter.

Unification

The different recursive algorithms discussed are quite similar. They can all be described by the equations

$$\begin{aligned}\hat{\theta}(t) &= \hat{\theta}(t-1) + P(t)\varphi(t-1)\varepsilon(t)\\ P(t) &= \frac{1}{\lambda}\left(P(t-1) - \frac{P(t-1)\varphi(t-1)\varphi^T(t-1)P(t-1)}{\lambda + \varphi^T(t-1)P(t-1)\varphi(t-1)}\right)\end{aligned}$$

where θ, φ, and ε are different for the different methods.