ID3
ARX Identification

3.13 KALMAN FILTERING IN RECURSIVE ESTIMATION

An alternate approach to using weighted least squares in on–line identification can be based on Kalman filtering. Assuming model (3.1.1) as non stationary, the time dependence of its parameters is described by relations

$$\alpha_i(t+1) = \alpha_i(t) + w_i(t) \qquad (i = 1, \ldots, n) \tag{3.13.1a}$$

$$\beta_i(t+1) = \beta_i(t) + w_{i+n}(t) \qquad (i = 1, \ldots, n) \tag{3.13.1b}$$

where $w_i(\cdot)$ $(i = 1, \ldots, 2n)$ denote white Gaussian processes with null expected values; every process is assumed as independent from remaining ones. Consider then the matrix

$$C(t) = \left[\, y(t-n) \ \ldots \ y(t-1)\ u(t-n) \ \ldots \ u(t-1) \,\right] \tag{3.13.2}$$

and the parameter vector

$$\theta(t) = \left[\, \alpha_1(t) \ \ldots \ \alpha_n(t)\ \beta_1(t) \ \ldots \ \beta_n(t) \,\right]^T. \tag{3.13.3}$$

Parameter variations and process observations can now be described by the dynamical system

$$\theta(t+1) = \theta(t) + w(t) \tag{3.13.4a}$$

$$y(t) = C(t)\,\theta(t) + e(t). \tag{3.13.4b}$$

Using a Kalman filter to estimate the state of system (3.13.4) we obtain, at time t, the estimate

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\big(y(t) - C(t)\,\hat{\theta}(t-1)\big) = \hat{\theta}(t-1) + K(t)\,\varepsilon(t). \tag{3.13.5}$$

Gain matrix $K(t)$ is given by

$$K(t) = \frac{P(t-1)\,C^T(t)}{C(t)P(t-1)C^T(t) + \sigma_e^2} \tag{3.13.6}$$

while $P(t)$ is obtained solving the Riccati difference equation

$$P(t) = P(t-1) - \frac{P(t-1)C^T(t)C(t)P(t-1)}{C(t)P(t-1)C^T(t) + \sigma_e^2} + Q(t). \tag{3.13.7}$$

$\sigma_e^2 = \mathrm{E}\,[e^2(t)]$ denotes the variance of the process $e(\cdot)$ while $Q(t) = \mathrm{E}\,[w(t)\,w^T(t)]$ is the covariance matrix of the process $w(\cdot)$. It can be easily verified that the most simple expression of recursive least squares (obtained updating $S(t)^{-1}$ instead of $R(t)^{-1}$) can be obtained from (3.13.7) assuming $Q(t) = 0$ (constant parameters) and $\sigma_e^2 = 1$. Assuming also $e(\cdot)$ as Gaussian, $P(t)$ can be interpreted as covariance matrix of the estimate.

When the process to be identified is stationary, relation (3.13.4a) could be substituted with $\theta(t+1) = \theta(t)$, eliminating $Q(t)$ in (3.13.7). This would, however, lead to an increase in the sensitivity of the filter to errors in formulating its equations and to possible estimation errors not described by $P(t)$; it is thus preferable to consider $w(t)$ as always present by introducing, when the process is stationary, covariance matrices very small but sufficient to avoid that $K(t) \to 0$ for $t \to \infty$.
Kalman filtering can be applied, as recursive least squares, only once a certain number of samples of $u(\cdot)$ and $y(\cdot)$ are available; from expression (3.13.2) of $C(t)$ it follows that the minimal number of samples necessary to start the algorithm equals n. It is also necessary to introduce, in the equations of the filter, estimates of σ_e^2, Q, of the initial value of $\hat{\theta}(t)$ (that should be equal to its expected value) and of the initial value of the covariance matrix, $P(t_0)$. It is common practice to set, in absence of specific information, $\hat{\theta}(t_0) = 0$, $P(t_0) = \sigma^2 I$ where σ^2 is a relatively large value. This leads, initially, to a large gain matrix which allows fast variations of $\hat{\theta}(t)$; if a better initial estimate of $\theta(t_0)$ is available, σ^2 will be proportionally reduced in order to reduce also the sensitivity to innovations $\varepsilon(t)$.

It can be observed that Kalman filtering offers more degrees of freedom than recursive weighted least squares because of the possible choices for the entries of Q. Even limiting the choice to constant and diagonal matrices, it is possible to introduce higher variances for the noises corresponding to parameters subject to larger variations. The estimate of σ_e^2 can rely on (3.10.5) or (3.14.10).