Predictive Control – Exercise Session 2
Real-Time Parameter Estimation

1.

a. The linear regression model is

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

where the regressor vector is,

$$\varphi(t)^T = \begin{bmatrix} u(t) & u(t-1) \end{bmatrix},$$

and the parameter vector is,

$$\theta^T = \begin{bmatrix} b_0 & b_1 \end{bmatrix}.$$

b. The set of regressor vectors for N set of measurements are collected in the matrix

$$\Phi = \begin{bmatrix} u(2) & u(1) \\ \vdots & \vdots \\ u(N) & u(N-1) \end{bmatrix},$$

and the measured output signal is collected in

$$Y = \begin{bmatrix} y(2) \\ \vdots \\ y(N) \end{bmatrix}.$$

Note the indices in these vectors; there are N data points, from $\{1, \ldots, N\}$. The first output $y(1)$ must therefore be neglected since it is determined by the unknown input $u(0)$. The row size of these vectors and matrices is $N-1$. The $2x2$ normal equation is

$$\left(\Phi^T \Phi\right) \hat{\theta} = \Phi^T Y,$$

where $\hat{\theta}$ is the estimate of the parameter vector θ and the covariance matrix is given by:

$$\left(\Phi^T \Phi\right)^{-1} = \begin{bmatrix} \sum_{i=2}^{N} u^2(i) & \sum_{i=2}^{N} u(i)u(i-1) \\ \sum_{i=2}^{N} u(i)u(i-1) & \sum_{i=2}^{N} u^2(i-1) \end{bmatrix}^{-1}.$$

2.

a. One obtains

$$\left(\Phi^T\Phi\right)^{-1} = \begin{bmatrix} N-1 & N-2 \\ N-2 & N-2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & \frac{N-1}{N-2} \end{bmatrix}.$$

For $Y^T = [y(2) \; ... \; y(N)]$ one obtains

$$\Phi^T Y = \begin{bmatrix} \sum\limits_{i=2}^{N} u(i)y(i) \\ \sum\limits_{i=2}^{N} u(i-1)y(i) \end{bmatrix} = \begin{bmatrix} \sum\limits_{i=2}^{N} y(i) \\ \sum\limits_{i=3}^{N} y(i) \end{bmatrix}$$

The estimates of the parameters are the solution to the normal equation and given by:

$$\hat{\theta} = \left(\Phi^T\Phi\right)^{-1}\Phi^T Y = \begin{bmatrix} y(2) \\ -y(2) + \frac{1}{N-2}\sum\limits_{i=3}^{N} y(i) \end{bmatrix}$$

b. The estimation error is

$$\hat{\theta} - \theta^o = \begin{bmatrix} e(2) \\ -e(2) + \frac{1}{N-2}\sum\limits_{i=3}^{N} e(i) \end{bmatrix},$$

and the covariance of the estimate is

$$\operatorname{cov}\hat{\theta} = \mathrm{E}(\hat{\theta}-\theta^o)(\hat{\theta}-\theta^o)^T = \sigma^2\left(\Phi^T\Phi\right)^{-1} = \sigma^2\begin{bmatrix} 1 & -1 \\ -1 & \frac{N-1}{N-2} \end{bmatrix} \stackrel{N\to\infty}{\longrightarrow} \sigma^2\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

Tha variance does not go to zero as $N \to \infty$, thus, no consistent estimates of the parameters can be obtained.

c. The sum of the parameters can be written as:

$$b_0 + b_1 = [1 \quad 1]\,\theta.$$

Hence,

$$E(\hat{b}_0 + \hat{b}_1 - b_0^o - b_1^o)^2 = \sigma^2\,[1 \quad 1]\,E(\hat{\theta}-\theta^o)(\hat{\theta}-\theta^o)^T\,[1 \quad 1]^T = \sigma^2\frac{1}{N-2},$$

which goes to zero as $N \to \infty$, thus, the sum of the parameters b_0 and b_1 can be estimated consistently.

d. With a step input it is thus possible to determine the combination $b_0 + b_1$ consistently, i.e., the static gain. The individual values of b_0 and b_1 can, however, not be determined consistently. Note that for a step $u(t)$ and $u(t-1)$ differ only for $t = 1$. There is thus only one data point which can separate the effects of b_1 and b_2. A step is persistently exciting of order one. It can therefore be used to estimate only a single parameter. Note that this applies to finite impulse response (FIR) systems, in which the current output depends solely on past inputs (no autoregression).

3.

a. As shown in Problem 2, the least-square estimates of the parameters are given by:

$$\hat{\theta} = \begin{bmatrix} \sum_{i=2}^{N} u^2(i) & \sum_{i=2}^{N} u(i)u(i-1) \\ \sum_{i=2}^{N} u(i)u(i-1) & \sum_{i=2}^{N} u^2(i-1) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=2}^{N} u(i)y(i) \\ \sum_{i=2}^{N} u(i-1)y(i) \end{bmatrix}$$

It is known that for a stochastic variable $v(t)$, the variance is given by $1/N \sum_{i=1}^{N} v^2(i) \to \mathrm{E}v^2(t)$ for $N \to \infty$. Thus,

$$\mathrm{E}\hat{\theta} \stackrel{N\to\infty}{\longrightarrow} \begin{bmatrix} (N-1)\mathrm{E}u^2(t) & (N-1)\mathrm{E}u(t)u(t-1) \\ (N-1)Eu(t)u(t-1) & (N-1)Eu^2(t-1) \end{bmatrix}^{-1} \begin{bmatrix} (N-1)\mathrm{E}y(t)u(t) \\ (N-1)\mathrm{E}y(t)u(t-1) \end{bmatrix}.$$

When the input u is white noise with unit variance, independent of e, we have

$$\mathrm{E}u^2(t) = 1, \quad \mathrm{E}u(t)u(t-1) = 0, \quad \mathrm{E}y(t)u(t) = b_0, \quad \mathrm{E}y(t)u(t-1) = b_1.$$

Hence,

$$\mathrm{E}\hat{\theta} \stackrel{N\to\infty}{\longrightarrow} \begin{bmatrix} (N-1) & 0 \\ 0 & N-1 \end{bmatrix}^{-1} \begin{bmatrix} (N-1)b_0 \\ (N-1)b_1 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}.$$

b. The covariance of the estimate

$$\operatorname{cov}\hat{\theta} = \sigma^2 \begin{bmatrix} N-1 & 0 \\ 0 & N-1 \end{bmatrix}^{-1},$$

which tends to zero as $N \to \infty$. In this case it is thus possible to determine both parameters consistently.

c. Both parameters can be estimated consistently since the input signal is white noise, which is persistently exciting of any order.

4 a. The LS algorithm assumes a set of input-output data of sufficient size to produce statistically reliable results is available. In real-time, such a data set does not exist; data is collected at each sampling instant. The standard LS algorithm therefore cannot be used until a sufficient number of data are collected. After this point, the algorithm could be run at each sampling instant as new data becomes available, but this would be extremely inefficient.

b. A recursive version of the LS algorithm may be formulated. This is known as *Recursive Least Squares* (RLS). In this algorithm, the parameter estimates and covariance are updated as new data becomes avaialble. The RLS algorithm is presented in Chapter 3 of *Predictive and Adaptive Control*, equations 3.1-3.3.

c. When parameters vary with time, 'older' information retained by either the standard LS algorithm or the RLS algorithm described above will cease to be relevant. Some means of neglecting older data is required. The RLS algorithm can be modified to include a so-called ' forgetting factor' λ, which provides this function. The modified algorithm is presented in equations 3.29-3.31 in *Predictive and Adaptive Control*. The parameter lambda can be seen as an exponentially decreasing weight on old data. In particular, the number of data points with non-negligible weights using this strategy can be approximated by $\frac{1}{1-\lambda}$.