227-0689-00L System Identification **22 November 2023**

Solution 7: PEM and ARX Models

Problem 1:

Consider the MA model

$$v(t) = e(t) + 2\,e(t-1),$$

in which $e(t)$ has zero-mean value and unit variance, and $E[e(t)e(s)] = 0$ for any $t \neq s$ (i.e., $e(t)$ is uncorrelated over time).

1. Suppose we have a record of v up to time t; namely, $\{v(t), v(t-1), v(t-2), v(t-3), \dots\}$. Is it possible to construct $e(t)$ based on the data record by a stable, causal filter?
2. Now consider a model $\bar{v}(t) = \bar{e}(t) + h\,\bar{e}(t-1)$, in which $|h| < 1$, $E[\bar{e}^2(t)] = \lambda$, and $E[e(t)e(s)] = 0$ for any $t \neq s$. Find a model for the noise $\bar{v}(t)$ with the same mean value and correlation function as $v(t)$, by choosing appropriate values for h and λ.
3. Show that $\bar{v}$ can be predicted from $\{\bar{v}(t-1), \bar{v}(t-2), \bar{v}(t-3), \dots\}$ by a stable, causal predictor.

This problem shows that, using only the second order properties (mean and variance/covariance) of the noise, we cannot distinguish between v and $\bar{v}$.

Solution

1. Notice that the given model can be written as

$$v(t) = H(z)e(t) = (1 + 2z^{-1})e(t)$$

The noise model $H(z)$ is not stably invertible: it has one unstable zero and no poles; the inverse model $\frac{1}{H(z)} = \frac{1}{1+2z^{-1}}$ is causal but not stable. This becomes clear when writing the solution of the general case $v(t) = e(t) + \theta\,e(t-1)$ for some constant real value θ:

$$\begin{aligned} v(t) &= e(t) + \theta e(t-1) \\ &= e(t) + \theta(v(t-1) - \theta e(t-2)) \\ &= e(t) + \theta v(t-1) + \theta^2 e(t-2) \\ &\ \ \vdots \\ &= e(t) + \sum_{k=1}^{\infty} (-1)^{k-1} \theta^k v(t-k) \end{aligned}$$

Then it is possible to write

$$e(t) = v(t) - \sum_{k=1}^{\infty} (-1)^{k-1} \theta^k v(t-k)$$

But the infinite sum converges only for value $|\theta| < 1$. For the given case, to solve for e, one can write

$$\begin{aligned} e(t) &= -0.5e(t+1) + 0.5v(t+1) \qquad \Longrightarrow \\ &= -0.5\left(-0.5e(t+2) + 0.5v(t+2)\right) + 0.5v(t+1) \\ &= 0.5^2 e(t+2) - 0.5^2 v(t+2) + 0.5v(t+1) \\ &\quad \vdots \\ &= \sum_{k=1}^{\infty} (-1)^{k-1} 0.5^k v(t+k) \end{aligned}$$

which shows that e depends on future values of v, and therefore cannot be constructed by a causal filter.

2. Notice that $v(t)$ has zero mean, and variance $E[v^2(t)] = 5$. The covariance $E[v(t)v(t-k)]$ equals 2 when $k = 1$ and zero for all other $k > 1$. It is straightforward to show that $\bar{v}(t)$ has the same second-order properties as $v(t)$ when $h = 0.5$ and $\lambda = 4$. The variance $E[\bar{v}^2(t)] = 4 + (0.5^2 \times 4) = 5$, and $E[\bar{v}(t)\bar{v}(t-k)]$ equals $h \times \lambda = 2$ when $k = 1$ and zero for all other $k > 1$.

3. From part 1, we know that we can write

$$\bar{v}(t) = \bar{e}(t) + \sum_{k=1}^{\infty} (-1)^{k-1} h^k \bar{v}(t-k)$$

where the infinite sum converges. A best linear predictor of $\bar{v}(t)$ based on its history is

$$\hat{\bar{v}}(t) = \sum_{k=1}^{\infty} (-1)^{k-1} h^k \bar{v}(t-k)$$

This is exactly the same value obtained by "inverting" the noise model; namely

$$\hat{\bar{v}}(t) = \Big(1 - \underbrace{\frac{1}{1 - 0.5z^{-1}}}_{H^{-1}}\Big)\bar{v}(t)$$

Problem 2:

Consider the following ARX model

$$y(t) = ay(t-1) + u(t) + e(t)$$

in which $|a| < 1$, and $e(t) \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, 1)$. Assume that you are given the values of the output $y(0), \ldots, y(N)$, and the values of the input $u(1), \ldots, u(N)$ for some integer $N > 1$, as data.

1. Find the Maximum-Likelihood (ML) estimator of a as a closed-form (explicit) expression in the data.
2. Derive the one-step ahead predictor of y.
3. Construct a Prediction-Error (PE) estimator of a as the minimizer of the squared 2-norm of the prediction error acquired by the one-step ahead predictor. Write the estimator as a closed-form (explicit) expression in the data.
4. How are the two estimators, the ML estimator and the PE estimator, related in this case?

This problem clarifies the relationship between the ML and the PE estimators using a simple model – which estimator is based on stronger assumptions on the data?

Solution

1. To compute the ML estimator we need to construct the likelihood function of a given the data $y(1), \ldots, y(N)$, conditioned on the known initial condition $y(0)$ and the given inputs $u(1), \ldots, u(N)$; namely the quantity $p(y(N), \ldots, y(1)|u(N), \ldots, u(1), y(0); a)$ seen as a function of a. It can be written as a product of conditional likelihoods

$$\begin{aligned} p(y(N), \ldots, y(1)|u(N), \ldots, u(1), y(0); a) &= p(y(1)|u(N), \ldots, u(1), y(0); a) \\ &\cdot p(y(2)|y(1), u(N), \ldots, u(1), y(0); a) \\ &\cdot p(y(3)|y(2), y(1), u(N), \ldots, u(1), y(0); a) \\ &\quad \vdots \\ &\cdot p(y(N), |y(N-1), \ldots, y(1), u(N), \ldots, u(1), y(0); a) \end{aligned}$$

But for the given ARX model, notice that

$$\begin{aligned} p(y(t)|y(t-1), \ldots, y(0), u(t), \ldots, u(1); a) &= p(y(t)|y(t-1), u(t); a) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y(t) - ay(t-1) - u(t))^2\right) \end{aligned}$$

where the second equality holds because the noise $e(t)$ is standard Gaussian. Therefore,

$$p(y(N), \ldots, y(1)|u(N), \ldots, u(1), y(0); a) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\sum_{t=1}^{N}(y(t) - ay(t-1) - u(t))^2\right)$$

Maximizing this function is the same as minimizing the negative log likelihood function. Ignoring the constant factor,

$$-\log p(y(N), \ldots, y(1)|u(N), \ldots, u(1), y(0); a) \propto \sum_{t=1}^{N}(y(t) - ay(t-1) - u(t))^2$$

The maximum-likelihood estimator is then

$$\hat{a}_{ML} = \arg\min_{a} \sum_{t=1}^{N}(y(t) - ay(t-1) - u(t))^2$$

This is a quadratic minimization problem in a which can be solved explicitly. Differentiating and using the optimality condition, we find that

$$-\sum_{t=1}^{N} y(t-1)(y(t) - \hat{a}_{ML}y(t-1) - u(t)) = 0$$

and

$$a_{ML} = \frac{\sum_{t=1}^{N} y(t-1)(y(t) - u(t))}{\sum_{t=1}^{N} y^2(t-1)}$$

2. The one-step ahead predictor can be read immediately from the given ARX form as

$$\hat{y}(t) = ay(t-1) + u(t)$$

Alternatively, the model can be written in the form

$$y = \underbrace{\frac{1}{1-az^{-1}}}_{G} u + \underbrace{\frac{1}{1-az^{-1}}}_{H} e$$

and then the formula for the one-step ahead predictor,

$$\hat{y} = H^{-1}Gu + (1 - H^{-1})y,$$

can be used to arrive at the same expression

3. We will solve this part using vectors, which for some models may be an easier option. Let $Y = \begin{bmatrix} y(1) & \dots & y(N) \end{bmatrix}^T$, $Y_- = \begin{bmatrix} y(0) & \dots & y(N-1) \end{bmatrix}^T$, and $U = \begin{bmatrix} u(1) & \dots & u(N) \end{bmatrix}^T$. Morover define the vector of one step ahead predictors $\hat{Y} = \begin{bmatrix} \hat{y}(1) & \dots & \hat{y}(N) \end{bmatrix}^T$, such that by definition

$$\hat{Y} = aY_- + U$$

A PEM estimator minimizing the squared 2-norm of the prediction errors is

$$\begin{aligned} \hat{a}_{PEM} &= \arg\min_{a} \|Y - \hat{Y}\|_2^2 = \arg\min_{a} \|Y - aY_- - U\|_2^2 \\ &= \frac{\sum_{t=1}^{N} y(t-1)(y(t) - u(t))}{\sum_{t=1}^{N} y^2(t-1)} \end{aligned}$$

4. In this case, because the model is linear and the noise is Gaussian, the ML estimator coincides with the PEM estimator. The ML is based on stronger assumptions on the data, as it needs the probability density function of the noise in order to construct the likelihood function.

MATLAB **Exercise:**

Consider the (unknown) ARX system with input u and output y

$$\left(1 + 0.25z^{-1} - 0.2z^{-2} + 0.1z^{-3} + 0.05z^{-4}\right) y(k) = \left(0.6z^{-1} + 0.3z^{-2} - 0.05z^{-3}\right) u(k) + e(k),$$

and $e(k)$ a random white external disturbance.

Identify this system considering ARX model structures of the type:

$$A(z)y(k) = B(z)u(k) + e(k), \tag{7.1}$$

with

$$A(z) = 1 + a_1 z^{-1} + ... + a_{n_a} z^{-n_a}, \tag{7.2}$$
$$B(z) = b_1 z^{-1} + ... + b_{n_b} z^{-n_b}. \tag{7.3}$$

1. Generate an identification dataset of length 200 using a a random Gaussian input signal u_{id} with $\sigma_u = 1$ and a random Gaussian noise vector e_{id} with $\sigma_e = 0.1$ to generate a sequence of outputs y_{id}.
2. Identify the system for different ARX orders. Use $n_a = n_b = \{1, \ldots, 8\}$.
3. Generate a validation dataset of length 200 using a random Gaussian input signal u_{val} with $\sigma_u = 1$ and a random Gaussian noise vector e_{val} with $\sigma_e = 0.1$ to generate a sequence of outputs y_{val}.
4. Use the inputs u_{id} and u_{val} to obtain predictions of the outputs for each of the ARX models that you identified. Plot the predicted outputs for the validation dataset.
5. For each ARX models, compute and plot the mean squared prediction error on the identification and on the validation dataset. Run your code a few times to examine the behaviour for different noise realizations.
6. Compute and plot the residuals on the validation dataset for the different ARX orders.
7. Compare the frequency responses of the identified models with the true one.

Solution hints

1. The Gaussian random input and noise signals can be generated with the function `randn`.
2. For each order, construct the regression matrix and then use the backslash command in Matlab to solve the LS problem, assuming known zero initial conditions.
3. Repeat step 1.
4. Compute the predicted outputs for the identified ARX models on the validation dataset. (See Fig. 7.1)
5. The mean squared error can be computed as $\frac{1}{200}\sum_{i=1}^{200}\left(y_{val}^i - y_{ARX}^i\right)^2$. Fig. 7.2 shows that for models of order 1 and 2 are too simple to capture the true ARX model dynamics. For high model complexities, the extra flexibility of the models of order 5, 6, 7 and 8 results in an improvement for the MSE on the identification dataset that is not matched in the MSE on the validation dataset. The models of order 3 and 4 provide an overall good balance between flexibility and model complexity.

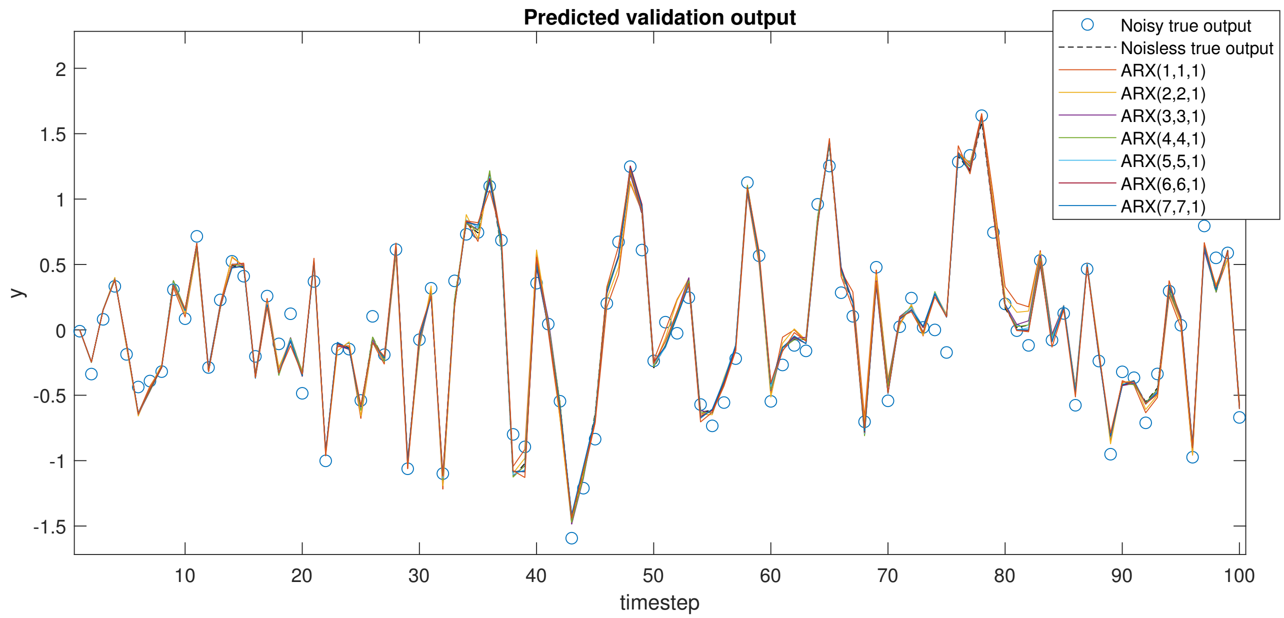

Figure 7.1: Example of predicted outputs for the identified ARX models on a validation dataset with 100 samples.

6. The residuals are computed as $r^i = y^i_{val} - y^i_{ARX}$. They are displayed in Fig. 7.3.
7. The frequency response can be computed using the Matlab function `bode`. In Fig. 7.4 we can see that the model of order 4 results in a better fit to the true plant transfer function with respect to the models of order 2 and 6.

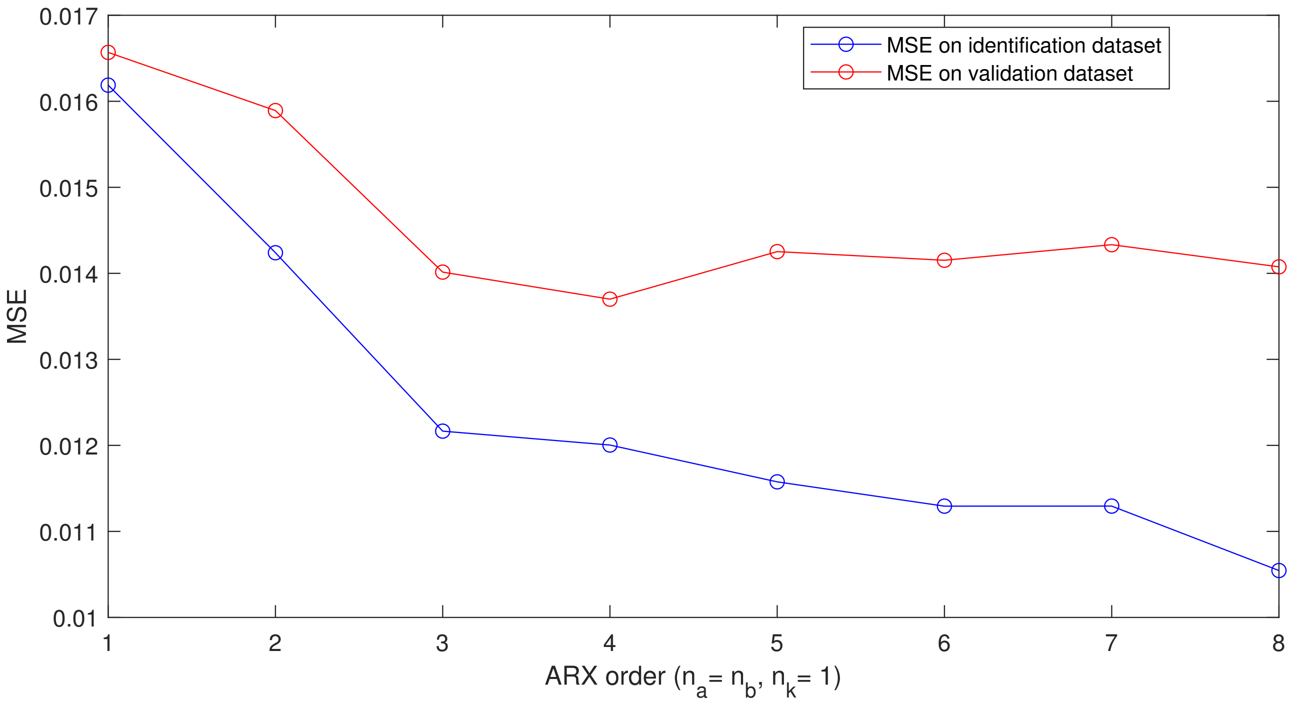

Figure 7.2: MSE on identification and validation datasets for different ARX orders.

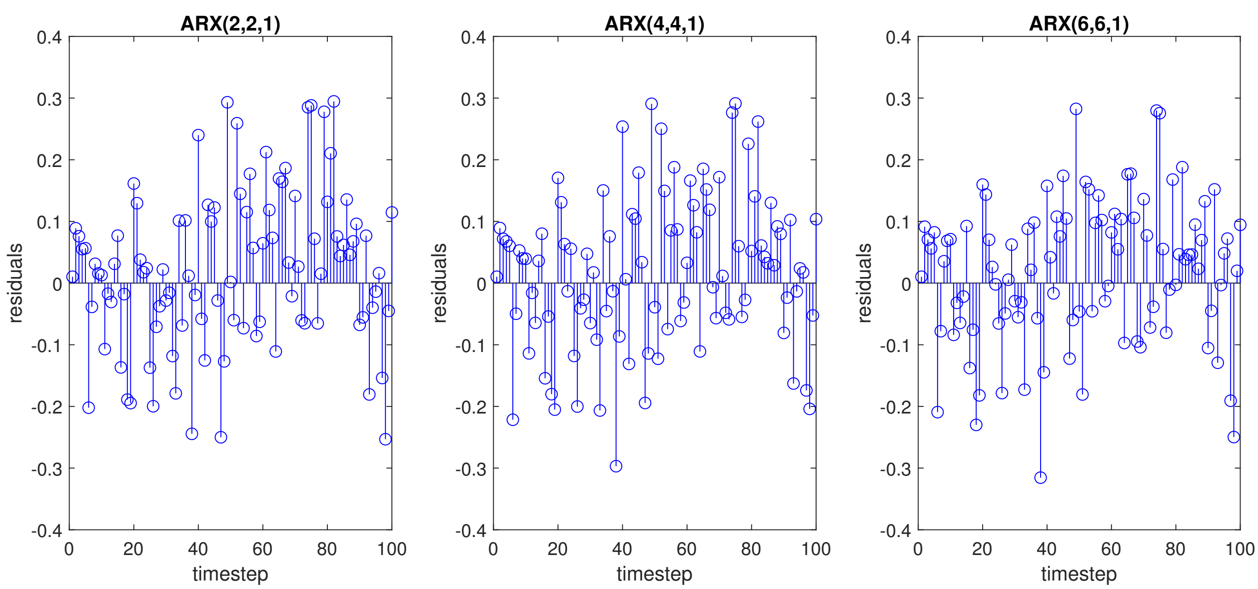

Figure 7.3: Residuals on the validation dataset for the identified ARX of order $n_a = n_b = \{2, 4, 6\}$.

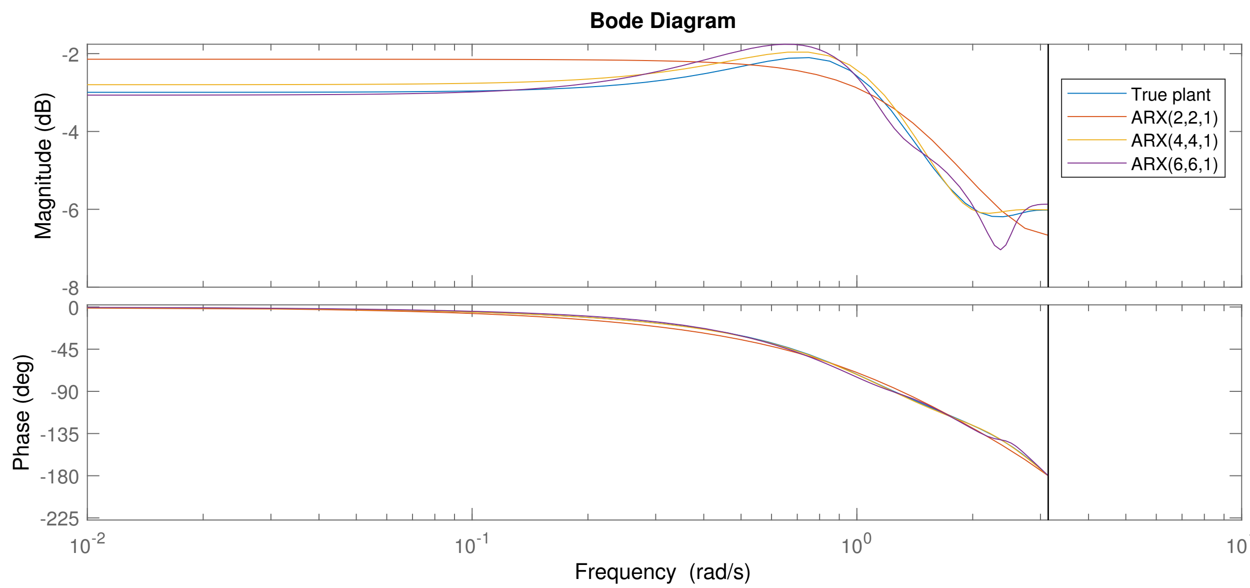

Figure 7.4: Bode diagrams for the identified ARX of order $n_a = n_b = \{2, 4, 6\}$.