**Solution for the exam in the course "Modeling and Learning for Dynamical System" (TSRT92) 2021-10-29**

1. (a) Three possible reasons are

   i. Bad parametrization (ex. in $y(t) = (a+b)u(t-1) + e(t)$ cannot distinguish between $a$ and $b$);

   ii. Bad choice of input (ex. concentrated in too few frequencies);

   iii. Presence of a feedback loop.

   (b) Alice with a program for solving linear systems of equations can only use ARX, as this class of models is linear in the parameters. Bob can use all classes of models.

   (c) For stationary stochastic processes, since $u$ and $e$ are uncorrelated, the spectrum can be computed as

$$\begin{aligned}\Phi_y(\omega) &= |G(i\omega)|^2\Phi_u(\omega) + \Phi_e(\omega) = G(i\omega)G(-i\omega)\Phi_u(\omega) + \Phi_e(\omega)\\ &= \frac{\omega^2+\alpha^2}{\omega^2+\beta^2} + 2 = \frac{3\omega^2+\alpha^2+2\beta^2}{\omega^2+\beta^2}\end{aligned}$$

   (d) The system is asymptotically stable, hence at equilibrium it is $\dot{x} = 0$, from which in correspondence of the input $u_o$, we get the stationary value $x_o = -\frac{u_o^2}{3}$. Hence the input-output static relationship is $y_o = x_o^2 = \frac{u_o^4}{9}$.

   (e) For a linear ODE $\dot{x} = \lambda x$, the forward Euler method is stable if $|1 + h\lambda| < 1$. In this case, the most restrictive eigenvalue is $\lambda = -4$, hence

$$|1-4h| < 1 \qquad \Longrightarrow \qquad 0 < h < \frac{1}{2}$$

2. (a) Denote $\theta = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}$. Since the true system lies in the chosen model class and the excitation is sufficiently rich ($u$ is a white noise), it is $\hat{\theta}_N \to \theta^* = \theta_0 = \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix}$, i.e., the problem is unbiased and the model is identifiable. $\theta^*$ should be obtained by explicitly minimizing $\bar{V}(\theta) = \lim_{N\to\infty} V_N(\theta)$ (we do this in point (b) below). The formula for the variance of the estimates is $P_N \approx \frac{1}{N}\lambda_v \bar{R}^{-1}$ where $\bar{R} = E[\psi(t,\theta_0)\psi^T(t,\theta_0)]$ and $\psi(t,\theta) = \frac{d}{d\theta}\hat{y}(t|\theta)$ is the gradient of the predictor. In our case $\psi(t,\theta) = \begin{bmatrix} -y(t-1) \\ u(t-1) \end{bmatrix}$, which gives

$$\bar{R} = E\begin{bmatrix} y^2(t-1) & -y(t-1)u(t) \\ -y(t-1)u(t) & u^2(t) \end{bmatrix} = \begin{bmatrix} R_y(0) & -R_{yu}(-1) \\ -R_{yu}(-1) & R_u(0) \end{bmatrix}$$

Computing the terms:

$$\begin{aligned} R_y(0) = E[y^2(t)] &= E[(-0.4y(t-1) + 0.2u(t) + v(t))^2] \\ &= 0.16R_y(0) + 0.04\lambda_u + \lambda_v \\ &\quad - 0.16\underbrace{E[y(t-1)u(t)]}_{=R_{yu}(-1)=0} - 0.8\underbrace{E[y(t-1)v(t)]}_{=0} + 0.4\underbrace{E[u(t)v(t)]}_{=0} \end{aligned}$$

i.e., $R_y(0) = \frac{2.04}{0.84} = 2.42$. Hence

$$\bar{R} = \begin{bmatrix} 2.42 & 0 \\ 0 & 1 \end{bmatrix} \implies \bar{R}^{-1} = \begin{bmatrix} 0.41 & 0 \\ 0 & 1 \end{bmatrix}$$

and $P_N \approx \frac{2}{N}\bar{R}^{-1}$, meaning that $\text{Var}[\hat{a}_1] = \frac{0.82}{N}$ and $\text{Var}[\hat{b}_1] = \frac{2}{N}$.

(b) Also in this case the true system is contained in the model class, and the excitation is rich, hence we have still an unbiased problem and an identifiable model. From this we already know that it must be $\hat{a}_1 = 0.4$, $\hat{a}_2 = 0$ and $\hat{b}_1 = 0.2$. Let us however compute explicitly these values through the prediction error minimization, as requested in the exercise.

$$\begin{aligned} \bar{V}(\theta) =& E[(y(t) - \hat{y}(t|\theta))^2] = E[((a_1 - 0.4)y(t-1) + a_2 y(t-2) + (0.2 - b_1)u(t) + v(t))^2] \\ =& 2.42(a_1 - 0.4)^2 + 2.42a_2^2 + (0.2 - b_1)^2 + 2 - 1.94(a_1 - 0.4)a_2 \end{aligned}$$

since $R_y(0) = 2.42$ (same as before) and

$$R_y(1) = E[y(t)y(t-1)] = E[(-0.4y(t-1) + 0.2u(t) + v(t))y(t-1)] = -0.4R_y(0) + 0 = -0.97$$

Differentiating w.r.t. the parameters:

$$\begin{aligned} \frac{d\bar{V}(\theta)}{da_1} &= 4.84(a_1 - 0.4) - 1.94a_2 = 0 \\ \frac{d\bar{V}(\theta)}{da_2} &= -1.94(a_1 - 0.4) + 4.84a_2 = 0 \\ \frac{d\bar{V}(\theta)}{db_1} &= -0.4 + 2b_1 = 0 \end{aligned}$$

from which it is straightforward to see that $\hat{a}_1 = 0.4$, $\hat{a}_2 = 0$ and $\hat{b}_1 = 0.2$ is indeed a solution, in correspondence of which

$$\frac{d^2\bar{V}(\theta)}{da_1^2} > 0, \qquad \frac{d^2\bar{V}(\theta)}{da_2^2} > 0, \qquad \frac{d^2\bar{V}(\theta)}{db_1^2} > 0$$

i.e., a minimum of $\bar{V}(\theta)$. Let us compute the variance of these estimates.

$$\begin{aligned}\bar{R} =& E[\psi(t,\theta_0)\psi^T(t,\theta_0)] = E\begin{bmatrix}-y(t-1)\\ -y(t-2)\\ u(t)\end{bmatrix}\begin{bmatrix}-y(t-1) & -y(t-2) & u(t)\end{bmatrix}\\ =& \begin{bmatrix}R_y(0) & R_y(1) & -R_{yu}(-1)\\ R_y(1) & R_y(0) & -R_{yu}(-2)\\ -R_{yu}(-1) & -R_{yu}(-2) & R_u(0)\end{bmatrix}\end{aligned}$$

$R_y(1) = -0.97$ (already computed). $R_{yu}(-1) = E[y(t-1)u(t)] = 0$, $R_{yu}(-2) = 0$. Therefore

$$\bar{R} = \begin{bmatrix}2.42 & -0.97 & 0\\ -0.97 & 2.42 & 0\\ 0 & 0 & 1\end{bmatrix} \implies \bar{R}^{-1} = \begin{bmatrix}0.49 & 0.19 & 0\\ 0.19 & 0.49 & 0\\ 0 & 0 & 1\end{bmatrix}$$

or $\mathrm{Var}[\hat{a}_1] = \mathrm{Var}[\hat{a}_2] = \frac{0.98}{N}$, $\mathrm{Var}[\hat{b}_1] = \frac{2}{N}$.

(c) For any $n_a > 1$ and $n_b > 1$, it will be $\hat{a}_1 = 0.4$, $\hat{a}_2 = \ldots = \hat{a}_{n_a} = 0$ and $\hat{b}_1 = 0.2$, $\hat{b}_2 = \ldots = \hat{b}_{n_b} = 0$ (model is still unbiased and identifiable), but the variance of the estimate will increase with $n_a$ and $n_b$ (compare the change in $\mathrm{Var}[\hat{a}_1]$ in (a) and (b) above).

3. (a) There is no sign of resonances, see the frequency function in Fig. 1, computed with SPA (M=100).

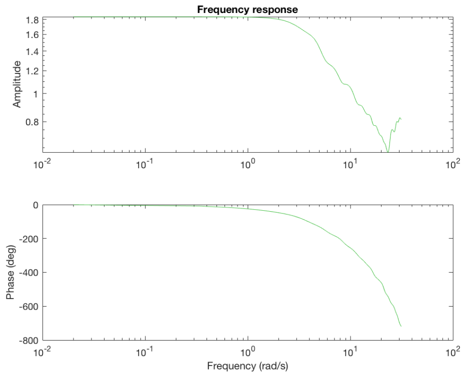


Figure 1: Frequency function.

(b) The AXR order selection tool suggests a delay $n_k = 3$. Keeping this delay, a model that satisfies the constraint of max 3 poles is for instance the following OE(2,3,3):

```
oe233 =
Discrete-time OE model:  y(t) = [B(z)/F(z)]u(t) + e(t)
  B(z) = 0.008766 (+/- 0.008597) z^-3 + 0.9868 (+/- 0.01131) z^-4

  F(z) = 1 - 0.4045 (+/- 0.01094) z^-1 - 0.1036 (+/- 0.0141) z^-2
            - 0.001491 (+/- 0.008743) z^-3

Name: oe233
Sample time: 0.1 seconds

Parameterization:
   Polynomial orders:   nb=2   nf=3   nk=3
   Number of free coefficients: 5
   Use "polydata", "getpvec", "getcov" for parameters and their uncertainties.

Status:
Termination condition: Near (local) minimum, (norm(g) < tol).
Number of iterations: 2, Number of function evaluations: 5

Estimated using PEM on time domain data "mydatade".
Fit to estimation data: 93.77\%
FPE: 0.01221, MSE: 0.01197
```

It provides a fit to validation data of 92.69%.
Parameter uncertainty is reasonably small, although for $B(z)$ the first coefficient is of the same magnitude of the error (other models with max 3 poles seems to have a similar problem).

The model fit is shown in cyan in Fig. 2. Residuals are in Fig. 3 and are within

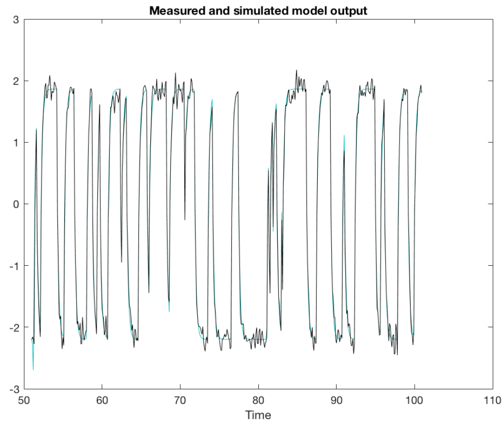


Figure 2: Model fit for OE(2,3,3).

ranges. Zeros/poles are in Fig. 5. The poles are all stable. Confidence interval are

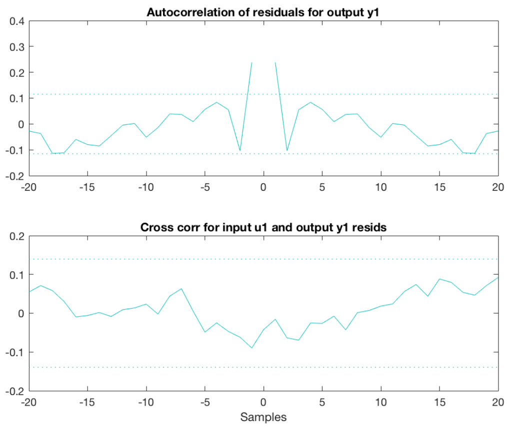


Figure 3: Residuals

not completely disjoint, but again this appears due to the pole-order constraint. The frequency function is similar to the SPA frequency function, see Fig. 5, up to a difference in the DC gain.

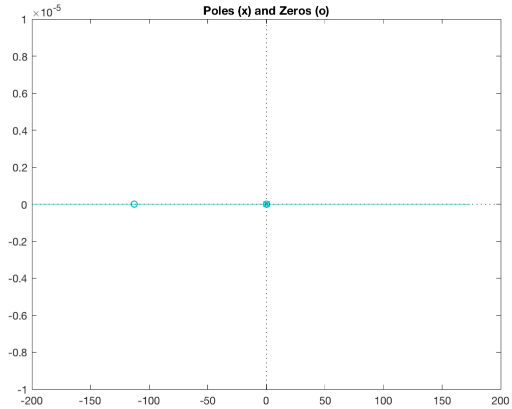


Figure 4: Zeros and poles

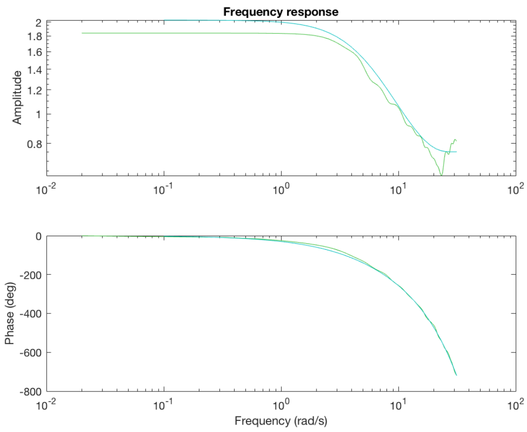


Figure 5: Frequency function of OE and SPA.

4. (a) Formulate a linear regression

$$\begin{aligned} y(t) &= -ay(t-1) + bu(t-1) + e(t) \\ &= [-y(t-1) \quad u(t-1)] \begin{bmatrix} a \\ b \end{bmatrix} + e(t) \\ &= \varphi^T(t)\theta + e(t) \end{aligned}$$

with the estimator

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

Thus, the regressor is $\varphi^T(t) = [-y(t-1) \quad u(t-1)]\,,\ t = 1, \ldots, N$. We formulate the multivariable case in matrix terms (i. e. we stack the data):

$$\begin{aligned} Y(t) &= [y(1) \ \ldots \ y(N)]^T \\ \Phi(t) = X &= \left[\varphi^T(0) \ \ldots \ \varphi^T(N-1)\right]^T \end{aligned}$$

Then the estimator (1) becomes

$$\hat{Y}(t) = \begin{bmatrix} \hat{y}(1) \\ \vdots \\ \hat{y}(N) \end{bmatrix} = \Phi(t)\theta = \begin{bmatrix} -y(0) & u(0) \\ \vdots & \vdots \\ -y(N-1) & u(N-1) \end{bmatrix} \theta$$

(b) The prediction error

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

The least-squares criterion for the linear regression is given by

$$V(\theta) = \frac{1}{N} \sum_{t=1}^{N} \frac{1}{2} \left[y(t) - \varphi^T(t)\theta\right]^2$$

and can be minimized analytically

$$0 = \frac{d}{d\theta} V(\theta)$$

$$\hat{\theta} = \left[\frac{1}{N} \sum_{t=1}^{N} \varphi(t)\varphi(t)^T\right]^{-1} \frac{1}{N} \sum_{t=1}^{N} \varphi(t) y(t)$$

We re-arrange the equation (assume $y(0) = 0$ and $u(0) = 0$):

$$\left(\sum_{t=1}^{N} \varphi(t)\varphi(t)^T\right) \hat{\theta} = \sum_{t=1}^{N} \varphi(t) y(t)$$

$$\left(\sum_{t=1}^{N} \begin{bmatrix} -y(t-1) \\ u(t-1) \end{bmatrix} [-y(t-1) \ \ u(t-1)]\right) \hat{\theta} = \sum_{t=1}^{N} \begin{bmatrix} -y(t-1) \\ u(t-1) \end{bmatrix} y(t)$$

$$\left(\sum_{t=0}^{N-1}\begin{bmatrix} y^2(t) & -y(t)u(t) \\ -y(t)u(t) & u^2(t)\end{bmatrix}\right)\hat{\theta} = \sum_{t=1}^{N}\begin{bmatrix} -y(t-1)y(t) \\ u(t-1)y(t)\end{bmatrix}$$

$$\begin{bmatrix} \sum_{t=0}^{N-1} y^2(t) & -\sum_{t=0}^{N-1} y(t)u(t) \\ -\sum_{t=0}^{N-1} y(t)u(t) & \sum_{t=0}^{N-1} u^2(t)\end{bmatrix}\hat{\theta} = \begin{bmatrix} -\sum_{t=1}^{N} y(t-1)y(t) \\ \sum_{t=1}^{N} u(t-1)y(t)\end{bmatrix} \tag{2}$$

Choosing

$$\alpha = \sum_{t=0}^{N-1} y^2(t); \quad \beta = \sum_{t=0}^{N-1} u(t)y(t); \quad \gamma = \sum_{t=0}^{N-1} u^2(t)$$

and

$$\delta = \sum_{t=1}^{N} y(t-1)y(t); \quad \eta = \sum_{t=1}^{N} u(t-1)y(t)$$

gives

$$\begin{bmatrix} \alpha & -\beta \\ -\beta & \gamma\end{bmatrix}\hat{\theta} = \begin{bmatrix} -\delta \\ \eta\end{bmatrix}$$

i. e.

$$X^TX = \begin{bmatrix} \alpha & -\beta \\ -\beta & \gamma\end{bmatrix}, \quad X^TY = \begin{bmatrix} -\delta \\ \eta\end{bmatrix}$$

(c) Use equation (2) and plug in $u(t) = -ky(t)$

$$\sum_{t=0}^{N-1}\begin{bmatrix} y^2(t) & ky^2(t) \\ ky^2(t) & k^2y^2(t)\end{bmatrix}\hat{\theta} = \sum_{t=1}^{N}\begin{bmatrix} -y(t-1)y(t) \\ -ky(t-1)y(t)\end{bmatrix}$$

$$\sum_{t=0}^{N-1} y^2(t)\begin{bmatrix} 1 & k \\ k & k^2\end{bmatrix}\hat{\theta} = -\sum_{t=1}^{N} y(t-1)y(t)\begin{bmatrix} 1 \\ k\end{bmatrix}$$

Choosing

$$\alpha = \sum_{t=0}^{N-1} y^2(t); \quad \delta = \sum_{t=1}^{N} y(t-1)y(t)$$

gives

$$\alpha\begin{bmatrix} 1 & k \\ k & k^2\end{bmatrix}\hat{\theta} = -\delta\begin{bmatrix} 1 \\ k\end{bmatrix} \tag{3}$$

i. e.

$$X^TX = \alpha\begin{bmatrix} 1 & k \\ k & k^2\end{bmatrix}, \quad X^TY = -\delta\begin{bmatrix} 1 \\ k\end{bmatrix}$$

Re-arrange the equation (3), assuming that $\alpha \neq 0$:

$$\begin{bmatrix} 1 & k \\ k & k^2\end{bmatrix}\hat{\theta} = -\frac{\delta}{\alpha}\begin{bmatrix} 1 \\ k\end{bmatrix} \tag{4}$$

We see that the solution of the normal equation depends on the quotient $\frac{\delta}{\alpha}$: Any $a$ and $b$ that satisfies $a + bk = -\frac{\delta}{\alpha}$: $\{(a, b) : a = -bk - \frac{\delta}{\alpha}, b \in \mathbb{R}\}$ is a solution. Since the two rows in (4) are linearly dependent (the second is $k$ times the first), and there is no unique solution.
Or: $X^TX$ is not invertible, i.e. there is no unique solution as $\hat{\theta} = (X^TX)^{-1}X^TY$.