Department of
AUTOMATIC CONTROL

FRTN35 System Identification

Final Exam October 25, 2017, 8am - 13pm

General Instructions

This is an open book exam. You may use any book you want, including the slides from the lecture, but no exercises, exams, or solution manuals are allowed. Solutions and answers to the problems should be well motivated. The exam consists of 5 problems. The credit for each problem is indicated in the problem. The total number of credits is 25 points. Preliminary grade limits:

Grade 3:   12 – 16 points
Grade 4:   17 – 21 points
Grade 5:   22 – 25 points

Results

The results of the exam will be posted at the latest November 8 on the notice-board on the first floor of the M-building.

1. Derive the ML estimate of the parameters b and c in the model

$$y_k = bu_{k-1} + c + v_k$$

based on a set of N measurements of y and u and where every element in the disturbance process v_k is normally distributed with mean zero and a unique and known variance σ_k^2

Show that the estimate is consistent.

(5 p)

Solution

The likelihood function is given by

$$L(\bar{\theta}) = f_e(\varepsilon_n, \ldots, \varepsilon_N) = \prod_{k=n}^{N} f_e(\varepsilon_k) = \left(\frac{1}{\sqrt{2\pi}}\right)^{N-n} \left(\prod_{k=n}^{N} \frac{1}{\sigma_k^2}\right) \exp\left(-\frac{1}{2}\sum_{k=n}^{N}\frac{\varepsilon_k^2}{\sigma_k^2}\right)$$

where $\varepsilon_k = y_k - bu_{k-1} - c$. Further on, the log-likelihood function (without terms not depending on the parameters) is given by

$$\log L(\bar{\theta}) = -\frac{1}{2}\sum_{k=n}^{N}\left(\frac{y_k - bu_{k-1} - c}{\sigma_k}\right)^2$$

Hence, in order to maximize the log-likelihood with respect to a and b, it is sufficient to minimize the loss function

$$J(b,c) = \sum_{k=n}^{N}\left(\frac{y_k - bu_{k-1} - c}{\sigma_k}\right)^2$$

This loss function is the same as we minimize when solving the least-squares problem, with an additional weighting by σ_k. Let

$$R = E(vv^T) = \text{diag}(\sigma_n^2, \ldots, \sigma_N^2)$$

then we can write the cost function like

$$J(b,c) = (Y - \Phi\theta)^\top R^{-1}(Y - \Phi\theta)$$

with standard notation Y, Φ and θ. The minimizing θ is given by $\hat{\theta} = \left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top R^{-1}Y$. To show consistency, we start with determining the bias:

$$\hat{\theta} = \left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top R^{-1}Y$$

$$= \left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top R^{-1}(\Phi\theta + v)$$

$$= \theta + \left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top R^{-1}v$$

As $E\left\{\left(\Phi^\top R^{-1}\Phi\right)^{-1}\Phi^\top R^{-1}v\right\} = 0$ since the regressors and noise are uncorrelated we can conclude that $E\{\hat{\theta}\} = \theta$ and the estimate is unbiased. For consistency, we also prove vanishing variance:
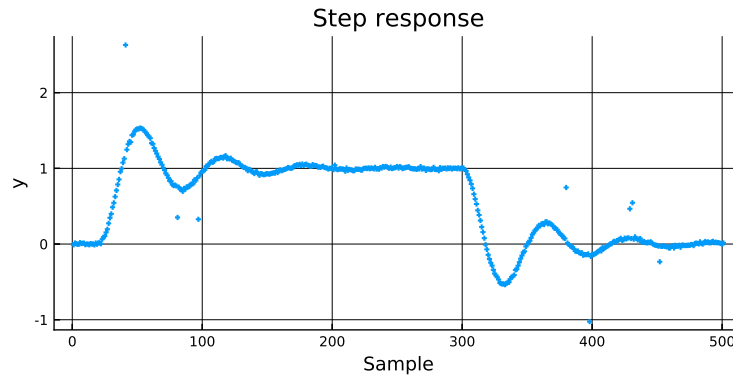
$$E\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^\top\} = E\{(\Phi^\top R^{-1}\Phi)^{-1}\Phi^\top R^{-1}vv^\top R^{-1}\Phi(\Phi^\top R^{-1}\Phi)^{-\top}\}$$

$$= E\{\Phi^{-1}vv^\top\Phi^{-\top}\}$$

$$= E\{(\Phi^\top R^{-1}\Phi)^{-1}\}$$

$$= (\Phi^\top R^{-1}\Phi)^{-1}$$

where the last equality holds if v and Φ are uncorrelated. As $N \to \infty$, we have

$$E\{(\Phi^\top R^{-1} \Phi)^{-1}\} \to 0$$

with calculations analogous to those performed in home assignment 2. (5 p)

2. You have performed a double step-response experiment and obtained output data as shown in Figure 1. ((4 - number of insufficient answers) p)

Figure 1 Output data from the experiment performed in problem 2.

a. Discuss if the sampling rate has been sufficiently high to capture the dynamics.

b. How complex model can you reasonably fit to the data from the experiment performed?

c. What measures would you take before you fit a model to the data?

d. Cross validation is an important part of the model validation process. Discuss how you would perform cross validation with the data obtained.

- How would you divide the data?
- What metrics would you calculate for validation purposes if the intended usage of the model is
  1. Prediction
  2. Simulation

e. Discuss how to choose the amplitude of the input signal for an identification experiment.

Solution

a. The closed loop system step response should be at least sampled 5-10 times during it's rise time. In our case we have plenty of samples per time constant and thus sufficiently fast sampling.

b. From a single step-response experiment, one can estimate the dominating time constant, the static gain and the damping. The experiment performed contains two step responses, but since the transient response of the first is allowed to vanish before the second step is applied, the second step only provides better statistical properties to the

estimate from the first, and doesn't really allow for estimation of more parameters. Hence, a transfer function model on the form

$$K\frac{\omega^2}{s^2+2\zeta\omega+\omega^2}$$

would be a good choice. Indeed, the system that generated the step response is given by

$$\frac{1}{s^2+2\cdot0.2\cdot1s+1}$$

  c. Start by checking the data manually via plots. In this case the data exhibits severe outliers which has to be removed, either manually or by some automatic outlier removal procedure. Failure to remove outliers will cause a severe bias if models are fit with a quadratic cost function (noise assumed to be Gaussian). This should be done before the identification procedure starts.

  d. The data can be divided into two parts

- identification data
- validation data

The response from the first input step can be used for identification. After 300 samples, the oscillations from the first step have vanished and the second step begins. It is thus advisable to use data before 300 samples for identification and data after 300 samples for validation.

If the model is to be used for the prediction, a natural metric to use is prediction error variance. If, however, the model is to be used for simulation, a more suitable metric is the mean squared simulation error, i.e., the mean of squared errors between the output data and the data obtained when the model is simulated forward in time, starting from the first data point with access to the input signal only.

  e. The amplitude should be chosen as large as possible in order to achieve a good signal-to-noise ratio and to overcome problems with friction. However, the amplitude may not be chosen larger than the range in which the linearity assumption holds. (See the section on preliminary experiments above.) Typically saturations give an upper bound on the amplitude of the input signal. The mean value is in many cases non-zero in order to reduce friction problems or to give a linearized model around a stationary point with $u^0 \neq 0$.   ((4 - number of insufficient answers) p)

3. After having performed a system identification experiment with a highly exciting input signal, you observe a coherence function that is very close to 1 for all frequencies of interest. What does this tell you about the system?   (1 p)

How do you determine an adequate model order for the system if you are fitting

1. An ARMAX model by means of pseudo-linear regression.
2. A state-space model by means of subspace-based identification.

Name at least a few methods in each case.   (1 p)

Solution

a. A coherence function close to one indicates that the relationship between input and output data is linear, and that most of the energy contained in the output spectrum is explained by the input, i.e., there is no significant noise influence. (1 p)

b. Since the coherence is close to one, we can assume that the system is linear.

For an ARMAX model fit by least-squares, we can

- Perform an F-test on the prediction error variance of two models of varying complexity.
- Calculate the confidence bounds of the estimated parameters. Parameters that are not significantly different from zero are not motivated by the data.

For the state-space model, the model order is indicated by the number of significant singular values of the Hankel matrix constructed from the Markov parameters. The Markov parameters can be calculated from an initial high-order transfer function fit to the data. Furthermore, the F-test can be used also for state-space models obtained through subspace-based identification.

Cross validation can always be used to select between two models. The model that performs the best on the validation data is to be preferred.

(1 p)

4. Consider the following identification problem

$$\underset{a,b}{\text{minimize}} \quad ||y_{k+1} - ay_k - bu_k||$$

$$\text{subject to} \quad u_k = -Ly_k$$

where $u_k = -Ly_k$ is a feedback controller with a known constant feedback L and the vector $\theta = [a, b]$.

a. Determine a vector v in parameter space, such that the addition of αv to θ

$$\bar{\theta} = \alpha v + \theta$$

where α is any real scalar, is invisible in the data, i.e., it is impossible to distinguish the parameter vector θ from $\bar{\theta}$ using the observed data. Comment on how this affects the estimation of the parameter vector, especially, what is the worst-case scenario?

(1 p)

b. Consider the case where you add the term $\varepsilon||\theta||$ to the cost function, where ε is a very small number. Explain intuitively how this eliminates the problem of unidentifiability encountered in **a.**, draw a figure to illustrate the level curves of the original cost function and the level curves of $\varepsilon||\theta||$. What relation will hold between θ^* and v in this case? Indicate this relation in the figure.

(2 p)

Solution

a. By inserting the feedback law $u = -Ly$ in the model equation, we get

$$y_{k+1} = (a - bL)y_k$$

and we see that we can only estimate the linear combination $(a - bL)$, i.e., the addition of any vector on the form $v = \alpha [L, 1]$ for any scalar α to the parameter vector is made invisible by the feedback.

Due to the feedback, there is no guarantee that the correct parameter vector will be recovered. In the worst case, the estimate of θ may go to infinity along the line v causing the estimation to fail.

(1 p)

b. The added penalty term causes the solution to be the point along the line $\theta + v$ with shortest distance to the origin, i.e., the vector θ^* will be perpendicular to v. The perpendicular to v is given by $Lx_1 + x_2 = 0 \Rightarrow \theta^* = \beta [L, -1]$ where the constant β is chosen such that θ^* reaches $\theta + v$. The relation between θ^* and v is $\theta^* \cdot v = 0$, i.e., they are perpendicular.

The level curves of the original cost function are lines parallel to $\theta + v$ whereas the level curves of $\varepsilon||\theta||$ are concentric circles centered at the origin. At θ^*, the tangents of the circles will be parallel to the line v.

(2 p)

5. Consider a dynamical model on the form

$$A(z)Y(z) = B(z)U(z) + C(z)N(z) \tag{1}$$

where Y, U, N are the z-transforms of the output, input and noise respectively.

(2 p)

a. Modify the block diagram of Figure 2 such that it reflects a control loop closed around the system (1) with a controller G. Make sure the polynomials are visible in your figure. Use your figure to reason about whether the noise in (1) can be characterized as a load, input or measurement disturbance.

b. Suggest a method for estimating the parameters in all polynomials A, B, C. Argue why the suggested method is unbiased.
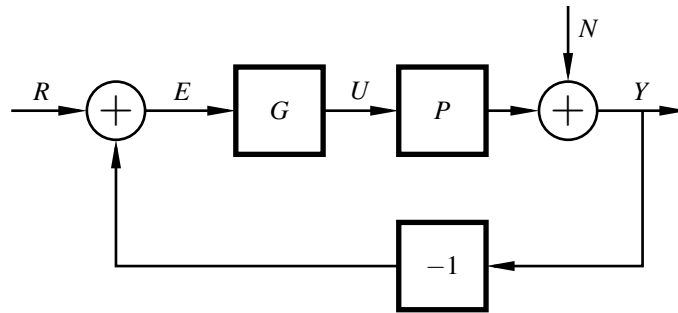
Figure 2 Block diagram of control loop. R, E, U, N, Y are the reference, error, input, noise and output signals respectively.

Solution

(2 p)
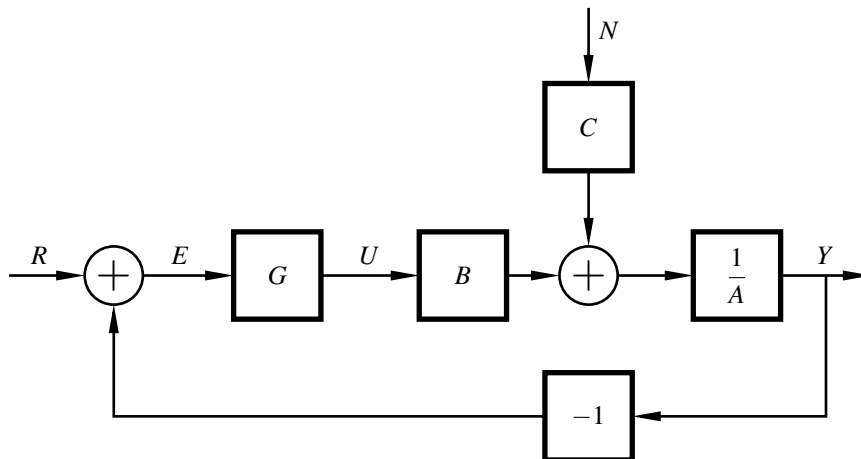
a. The noise is clearly filtered through the $1/A$ transfer function before it is visible in the output, it can hence not be characterized as measurement noise. It is acting before the B-polynomial and is thus not acting on the process input, thus we conclude that N in this case is likely describing a wide-spectrum load disturbance, e.g., the fluctuations of road inclination disturbing the velocity of a car, where the C polynomial is of low-pass character, indicating that the road is not changing with too high frequency.

b. One can utilize either pseudo-linear regression or the prediction error method. PLR works since any rational function can be approximated by its Taylor series expansion. The estimation of the noise process is asymptotically unbiased as the order of the expansion increases. The prediction error method will in this case search a non-convex loss surface and may suffer from local minima, but have the potential to find the global optimum, in which case the estimate will be unbiased.

6. While estimating a model for a resonant system, you observe the impulse response shown in Fig. 3.
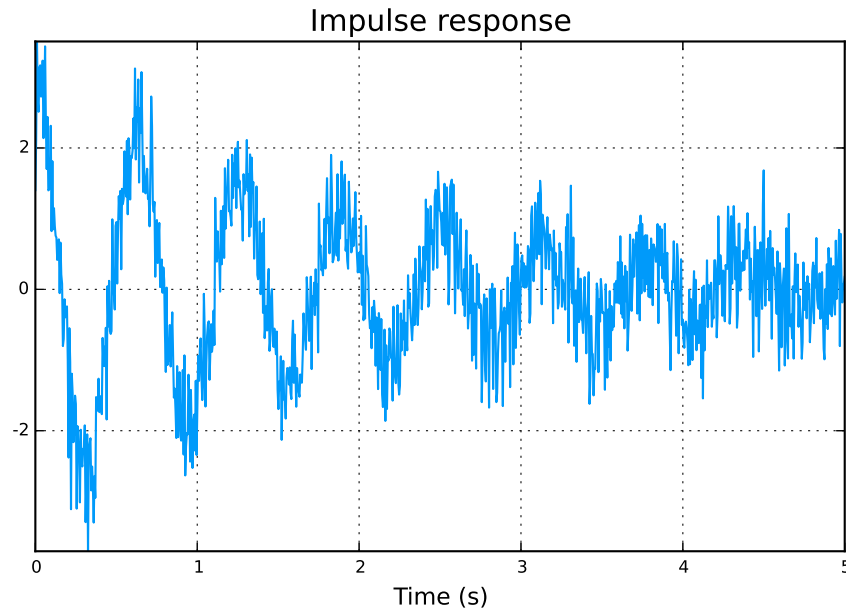
Figure 3 Impulse response in problem 6.

You determine that the impulse response has the nonlinear form

$$y(t) = A\cos(\omega t)e^{-at} + e(t) \qquad (2)$$

for some parameters A, ω and a.

a. Determine a suitable identification procedure to find the parameters A, a, ω that minimizes the sum of squared errors between the measured impulse response data and the model predictions.

Discuss the merits and drawbacks of your suggested strategy.

Hint: You might recognize the form of Equation (2). (2 p)

b. If you are given the chance to send an arbitrary input signal to the system, determine a more suitable method of identifying a model for the system.

Discuss how and why this strategy might improve the results as compared to the strategy in the previous subproblem. (2 p)

Solution

We observe that the impulse response equation is an exponentially decaying sinusoid, more specifically the solution to the differential equation

$$\ddot{y} + 2a\dot{y} + (a^2 + \omega^2)y = A(\dot{u} + au)$$

corresponding to the Laplace transform

$$A\frac{s+a}{(s+a)^2 + \omega^2} = A\frac{s+a}{s^2 + 2as + a^2 + \omega^2}$$

The system giving rise to the observed impulse response is thus a linear system which can be estimated with any of the methods treated in the course!

8

a. The impulse response equation is nonlinear in the parameters. One approach is to formulate the least-squares cost function and optimize it using gradient descent, this requires a good initial guess, which in this case is slightly tricky to obtain since the decay rate is hard to determine due to the noise. The period of the oscillation is rather easily observable. Once having identified the parameters in the impulse response, we have a linear model of the process given by the transfer function.

Since the system is linear, any method for impulse response estimation can be used. The gradient descent method might be a poor choice due to the high noise content which might lead to the algorithm getting stuck in a local minimum. Another approach is to form a Hankel matrix with the Markov parameters which, through the singular value decomposition, can be factorized to obtain the matrices A, B and C in a linear state-space model. (2 p)

b. In this case we can send any wide-band signal as input and use, e.g., least-squares estimation or subspace-based methods to find the parameters of a linear model of second order. This will further allow for an arbitrary measurement duration which can improve the covariance of the estimate for consistent estimation methods. Especially, a longer duration experiment with continuous excitation on the input allows for more accurate identification of the system zeros and gain. (2 p)

7. Figure 4 shows the squared coherence function $\gamma^2(\omega)$ and the input and output autospectra S_{uu} and S_{yy} from an identification experiment conducted on the system given by:

$$Y(s) = P(s)U(s) + N(s)$$

where the input $U(s)$ and noise $N(s)$ are known to be uncorrelated. The aim of the identification is to design a controller for the process. The data sets $\{u_k\}$ and $\{y_k\}$ from the experiment are used to obtain an ARX model $P(z)$ of the system, shown in Figure 5.
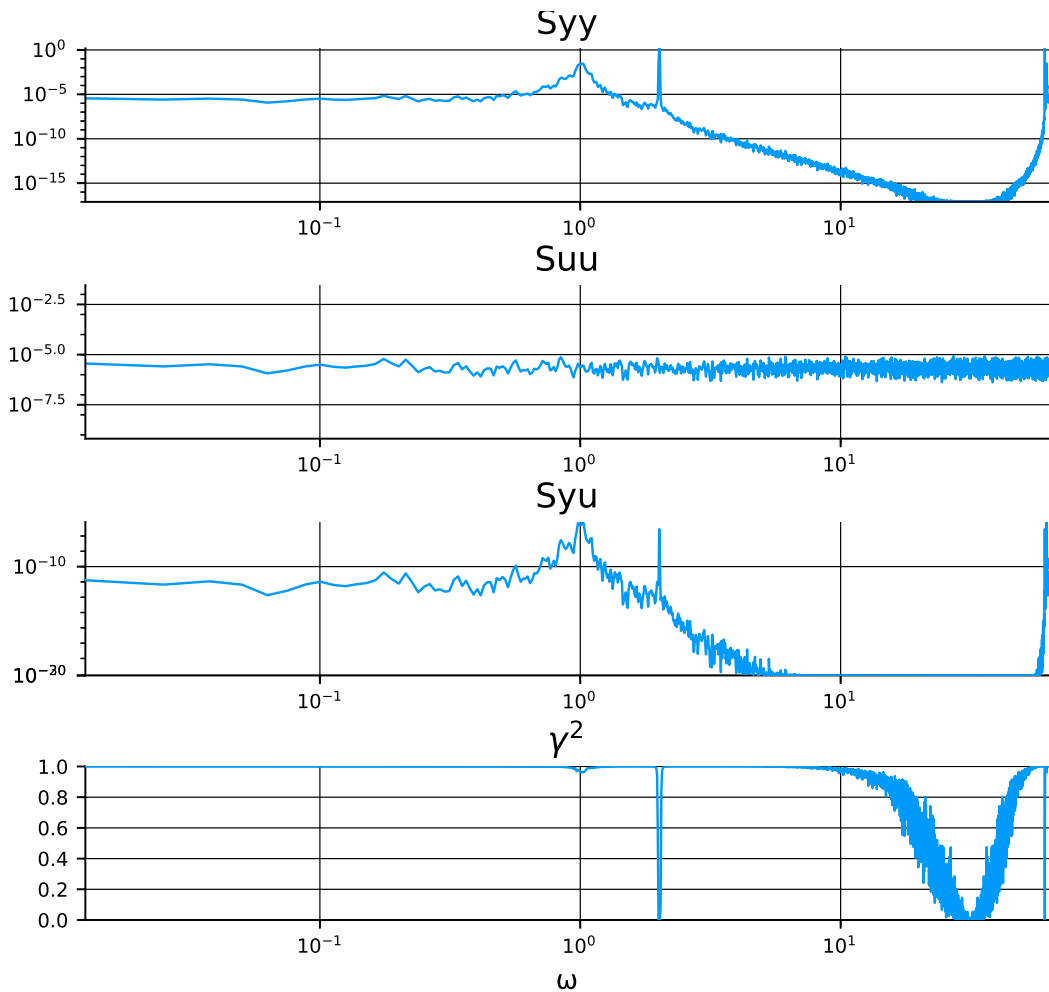

Figure 4 Input, output and cross spectra S_{uu}, S_{yy} and S_{yu} as well as squared coherence function $\gamma^2(\omega)$

After designing a simple P-controller, the sensitivity function is drawn (shown in Figure 5). Given the coherence function from the estimation, is it advisable to use the designed controller? If not, what measures can be taken to improve upon 1) The identification procedure, 2) the control design?   (3 p)

Hint: The sensitivity function is given by $\dfrac{1}{1+PG}$ with the control loop depicted in Figure 2.

Solution

The input autospectrum indicates that there is equal excitation for a wide band of
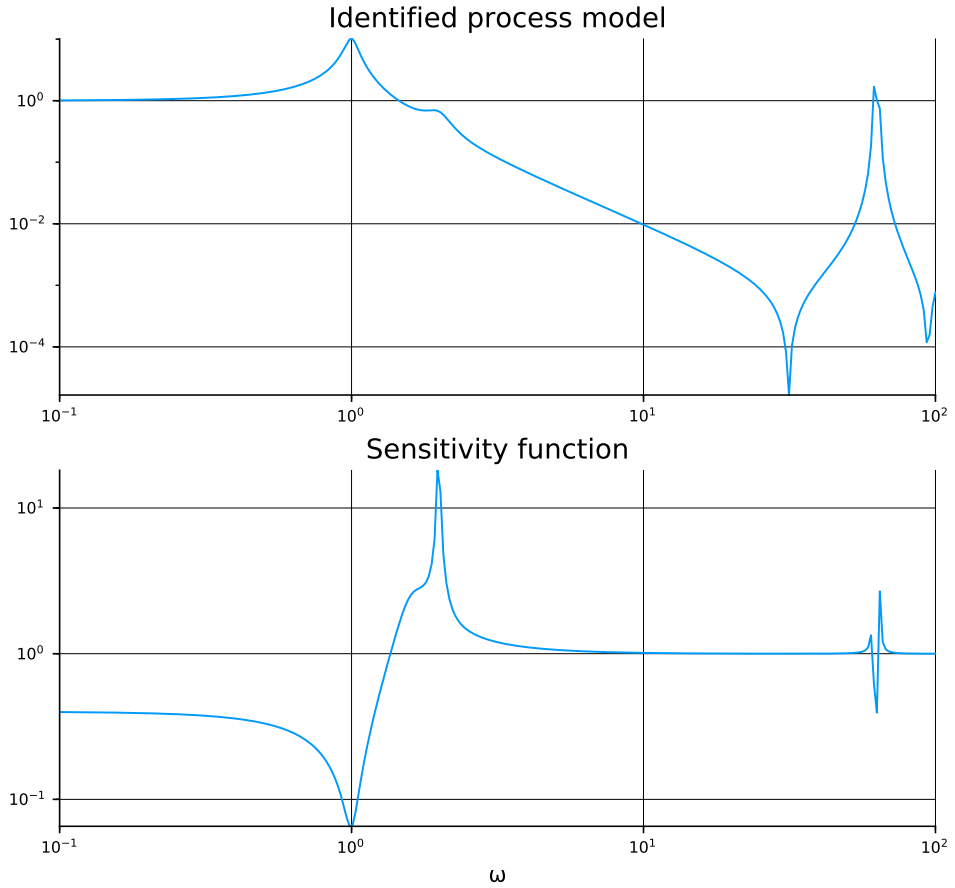
Figure 5 Bode diagrams of the identified process model and the sensitivity function obtained after controller design.

frequencies. The coherence function does however display a significant loss of coherence for frequencies around 2rad/s, indicating that there is a strong, narrow-band noise component acting on the output of the system.

To improve upon the identification procedure, one can either ensure that there is higher energy in the input signal at the affected frequency or at the very least, filter the measurement signal to get rid of some of the noise energy before performing the estimation since the peak in the bode diagram at the noise frequency indicates that the identification has been affected. Another improvement is to identify a model on the form

$$Y(s) = P(s)U(s) + G(s)N(s)$$

where $G(s)$ explicitly models the noise spectrum.

With the proposed controller design, the sensitivity function displays a very high amplification of noise around exactly 2rad/s (the sensitivity function $\dfrac{1}{1+PG}$ is the transfer function from N to Y). It is thus not advisable to use the proposed controller design. To mitigate the issue, a control design which results in a smaller sensitivity function around the noise affected frequency is required, such as a controller with a notch filter tuned to the noise frequency or a controller with significantly higher gain.   (3 p)

8. Determine if the least-squares estimates of the model parameters in the following systems will be consistent when the input signal u is white noise. In all cases, the noise processes e_k and v_k have the characteristics

$$e_k \sim N(0,\sigma^2) \qquad\qquad \mathbb{E}\{e_i e_j\} = \sigma^2 \delta_{ij}$$
$$v_k \sim N(0,\sigma^2) \qquad\qquad \mathbb{E}\{v_i v_j\} = \sigma^2 e^{-(i-j)^2}$$

You may (correctly) refer to results obtained during the course home assignments in your solutions, i.e., you do not *have* to prove the asymptotic properties, but your answer must be properly motivated.

(2 - (number of insufficient answers) p)

a.
$$y_{k+1} = ay_k + bu_k + v_k$$

b.
$$y_{k+1} = ay_k + bu_k + e_k$$

c.
$$y_k = b_1 u_{k-1} + b_2 u_{k-2} + e_k$$

d.
$$y_k = b_1 u_{k-1} + b_2 u_{k-2} + v_k$$

Solution

We start by noticing that e_k is a white noise process whereas v_k is a colored noise process. (2 - (number of insufficient answers) p)

a. In this case there will be a bias due to the feedback of colored noise.

b. This is a standard case that was proved bias free in the home assignments.

c. This is another simple case where there will be no bias.

d. This case is also bias free. The noise process is colored, but there is no feedback, hence, there will be no bias.