Department of Engineering Cybernetics

TTK4215 System Identification and Adaptive Control

Assignment 1

Non-parametric Methods

Consider the time-discrete linear time invariant system

$$y(t) = \sum_{k=1}^{\infty} g_0(k) u(t-k) + v(t), \tag{1}$$

where $g_0(k)$, $k = 1, 2, 3, \ldots$, is the impulse response, $u(t)$ is the input, $y(t)$ is the measured output, and $v(t)$ is measurement noise. Note that by introducing the forward shift operator q and the backward shift operator q^{-1}, that is

$$qu(t) = u(t+1), \tag{2}$$

$$q^{-1}u(t) = u(t-1), \tag{3}$$

we can define

$$G_0(t) = \sum_{k=1}^{\infty} g_0(k) q^{-k} \tag{4}$$

and write (1) as

$$y(t) = G_0(q) u(t) + v(t). \tag{5}$$

$G_0(t)$ is called the transfer function of the linear time-invariant system (1).

Problem 1 (Impulse-Response Analysis)

a) Let $u(t)$ be the impulse

$$u(t) = \begin{cases} a, & t = 0 \\ 0, & t \neq 0 \end{cases}. \tag{6}$$

Show that the output will be

$$y(t) = a g_0(t) + v(t). \tag{7}$$

b) Using (7), how would you propose to estimate the impulse response $\{g_0(t)\}$?

c) Find an expression for the estimation error. Do you see any practical problems with this estimation scheme?

Problem 2 (Step-Response Analysis)

a) Let $u(t)$ be the step

$$u(t) = \begin{cases} a, & t \geq 0 \\ 0, & t < 0 \end{cases}, \tag{8}$$

and write an expression for $y(t)$.

b) Using the expression for $y(t)$ from a), show that an estimate of g_0 can be obtained as

$$\hat{g}_0(t) = \frac{y(t) - y(t-1)}{a}. \tag{9}$$

c) Find an expression for the estimation error. Do you see any practical problems with this approach?

d) Despite any practical problems you might have mentioned in c), step-responses are commonly used in practice for deriving key parameters in simple transfer functions. Suppose you know that your system is stable, is of dynamic order 1, and has a time delay. Such systems can be parameterized in terms of three unknowns. Select a reasonable parameterization and use a graph to explain how you would go about estimating the three parameters using a single step response.

Problem 3 (Correlation Analysis)

Definition. A signal $\{s(t)\}$ is *quasi-stationary* if

$$\begin{array}{lc} i) & Es(t) = m_s(t), \ \ |m_s(t)| \le c, \ \forall t \\ & Es(t)s(r) = \breve{R}_s(t,r), \ \ \left|\breve{R}_s(t,r)\right| \le c, \\ ii) & \\ & \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \breve{R}_s(t,t-\tau) = R_s(\tau), \ \forall \tau. \end{array}$$

a) Show that stationary signals are quasi-stationary (Recall that for a stationary signal $\{v(t)\}$, $Ev(t)$ and $Ev(t)\,v(t-\tau) \triangleq R_v(\tau)$ are independent of t).

For simplicity, we introduce the notation

$$\bar{E}s(t) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} Es(t). \tag{10}$$

b) Show that

$$\bar{E}s(t)\,s(t-\tau) = R_s(\tau). \tag{11}$$

We call $R_s(\tau)$ the covariance function of s.

Definition. Two quasi-stationary signals $\{s(t)\}$ and $\{w(t)\}$ are *jointly quasi-stationary* if the cross-covariance function

$$R_{sw}(\tau) = \bar{E}s(t)w(t-\tau) \tag{12}$$

exists. $\{s(t)\}$ and $\{w(t)\}$ are *uncorrelated* if $R_{sw}(\tau) \equiv 0$.

Definition. The *spectrum* of $\{s(t)\}$ is given as

$$\Phi_s(\omega) = \sum_{\tau=-\infty}^{\infty} R_s(\tau)\, e^{-i\tau\omega} \tag{13}$$

and the *cross spectrum* between $\{s(t)\}$ and $\{w(t)\}$ as

$$\Phi_{sw}(\omega) = \sum_{\tau=-\infty}^{\infty} R_{sw}(\tau) e^{-i\tau\omega}. \tag{14}$$

Let $\{w(t)\}$ be quasi-stationary and let

$$s(t) = G(q) w(t)$$

where $G(q)$ is a stable transfer function. Then $\{s(t)\}$ is also quasi-stationary and

$$\Phi_s(\omega) = \left|G\left(e^{i\omega}\right)\right|^2 \Phi_w(\omega) \tag{15}$$
$$\Phi_{sw}(\omega) = G\left(e^{i\omega}\right) \Phi_w(\omega). \tag{16}$$

Since the spectrum of $\{s(t)\}$ is defined as the Fourier transform of the covariance function of $\{s(t)\}$, we have

$$R_s(\tau) = \mathcal{F}^{-1}\{\Phi_s(\omega)\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_s(\omega) e^{i\tau\omega} d\omega. \tag{17}$$

Similarly, for the cross-spectrum

$$R_{sw}(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{sw}(\omega) e^{i\tau\omega} d\omega. \tag{18}$$

c) Let

$$y(t) = G_0(q) u(t) + v(t), \tag{19}$$

where $\{u(t)\}$ and $\{v(t)\}$ are uncorrelated, quasi-stationary signals. Show that

$$R_{yu}(\tau) = \sum_{k=1}^{\infty} g_0(k) R_u(\tau - k). \tag{20}$$

(Recall that $G_0(q) = \sum_{k=1}^{\infty} g_0(k) q^{-k}$.)

The relationship between the covariance function/cross-covariance function and single realizations of the signals are given in the following result:

Let $\{s(t)\}$ be quasi-stationary with $Es(t) = m(t)$. Assume that

$$s(t) - m(t) = v(t) = H_t(q) e(t)$$

where $\{e(t)\}$ is a sequence of independent random variables with $Ee(t) = 0$ and $Ee^2(t) = \lambda_t$. Then, with probability 1 as $N \to \infty$,

$$\frac{1}{N}\sum_{t=1}^{N} s(t) s(t-\tau) \to \bar{E}s(t)s(t-\tau) = R_s(\tau), \tag{21}$$

$$\frac{1}{N}\sum_{t=1}^{N}[s(t)m(t-\tau) - Es(t)m(t-\tau)] \to 0, \tag{22}$$

$$\frac{1}{N}\sum_{t=1}^{N}[s(t)v(t-\tau) - Es(t)v(t-\tau)] \to 0. \tag{23}$$

The convergence holds "with probability 1", which means that it holds for almost every realization. The result means that, provided $\{s(t)\}$ is filtered white noise, the spectrum of an observed single realization of $\{s(t)\}$ coincides with that of the process $\{s(t)\}$ defined by ensamble averages.

Let

$$y(t) = G_0(q)\,u(t) + v(t), \tag{24}$$

where v is filtered white noise. Suppose you are given the data series $\{u(0), u(2), ..., u(N)\}$ and $\{y(0), y(2), ..., y(N)\}$. From the results above, it is reasonable to estimate $R_u(\tau)$ and $R_{yu}(\tau)$ as

$$\begin{aligned} \hat{R}_u^N(\tau) &= \frac{1}{N}\sum_{t=\tau}^{N} u(t)\,u(t-\tau), & (25) \\ \hat{R}_{yu}^N(\tau) &= \frac{1}{N}\sum_{t=\tau}^{N} y(t)\,u(t-\tau). & (26) \end{aligned}$$

d) Using (20), and the estimates $\hat{R}_u^N(\tau)$ and $\hat{R}_{yu}^N(\tau)$, can you propose a way of estimating the impulse response coefficients $\{g_0(1), g_0(2), ..., g_0(M)\}$?

Problem 4 (Fourier Analysis)

By Fourier transforming the output $y(t)$ and the input $u(t)$, we obtain

$$\begin{aligned} Y_N(\omega) &= \frac{1}{\sqrt{N}}\sum_{t=1}^{N} y(t)\,e^{-i\omega t}, & (27) \\ U_N(\omega) &= \frac{1}{\sqrt{N}}\sum_{t=1}^{N} u(t)\,e^{-i\omega t}, & (28) \end{aligned}$$

and we can estimate the transfer function as

$$\hat{G}_N\left(e^{i\omega}\right) = \frac{Y_N(\omega)}{U_N(\omega)}, \tag{29}$$

at every frequency contained in $u(t)$. This estimate is called Empirical Transfer-function Estimate (EFTE). The properties of the EFTE can be stated mathematically as follows.

$$E\hat{G}_N\left(e^{i\omega}\right) = G_0\left(e^{i\omega}\right) + \frac{\rho_1(N)}{U_N(\omega)}, \tag{30}$$

where $|\rho_1(N)| \leq c/\sqrt{N}$ for some constant c, and

$$E[\hat{G}_N\left(e^{i\omega}\right) - G_0\left(e^{i\omega}\right)][\hat{G}_N\left(e^{i\xi}\right) - G_0\left(e^{i\xi}\right)]$$
$$= \begin{cases} \frac{1}{|U_N(\omega)|^2}\left(\Phi_v(\omega) + \rho_2(N)\right) & \text{for } \xi = \omega \\ \frac{\rho_2(N)}{U_N(\omega)U_N(-\xi)} & \text{for } |\xi-\omega| = \frac{2\pi k}{N} \\ & k = 1, 2, ..., N-1 \end{cases}, \tag{31}$$

where $|\rho_2(N)| \leq c/N$ for some constant c.

a) Based on the mathematical properties of the EFTE stated above, what can you say (in words) about the properties of the EFTE?

Department of Engineering Cybernetics

TTK4215 System Identification and Adaptive Control

Assignment 2

Prediction-Error Methods – The Least-Squares Method

Suppose a system is given in the form

$$y(t) = G(q, \theta^*) u(t) + H(q, \theta^*) e(t), \tag{32}$$

where $e(t)$ is white noise with $Ee(t) = 0$ and $Ee^2 = \lambda$. θ^* is a vector of unknown parameters assumed to take values in $D_{\mathcal{M}} \in \mathbb{R}^d$, and is sometimes omitted from the list of arguments for notational simplicity. The noise model $H(q, \theta^*)$ is stable and inversely stable (the inverse exists and is stable). Prediction-Error Methods rely on being able to predict $y(t)$ from data up to and including $t-1$. We denote this prediction $\hat{y}(t|t-1)$, $\hat{y}(t|\theta)$, or simply $\hat{y}(t)$. The prediction errors is denoted $\varepsilon(t)$ or $\varepsilon(t|\theta)$, that is

$$\varepsilon(t) = y(t) - \hat{y}(t). \tag{33}$$

While a predictor can be defined in many ways, a particularly desirable choice is the one that minimizes the variance of the prediction error.

Problem 1 (Optimal Predictions)

a) Consider the first order ARMAX model

$$y(t) + ay(t-1) = bu(t-1) + e(t) + ce(t-1), \tag{34}$$

Show that the variance of the prediction error

$$\varepsilon(t) = y(t) - \hat{y}(t) \tag{35}$$

is bounded below by λ.

b) Show that the predictor given by the recursive formula

$$\hat{y}(t) = -c\hat{y}(t-1) + (c-a) y(t-1) + bu(t-1) \tag{36}$$

is optimal in the sense that it minimizes the prediction error variance.

c) Show that the predictor can be written

$$\hat{y}(t) = W_u(q) u(t) + W_y(q) y(t), \tag{37}$$

where

$$W_u(q) = \frac{bq^{-1}}{1+cq^{-1}}, \tag{38}$$

$$W_y(q) = \frac{(c-a)\,q^{-1}}{1+cq^{-1}}. \tag{39}$$

The prediction sequence is the output resulting from feeding the output sequence $y(t)$ and the input sequence $u(t)$ through linear filters. Thus, we call $[W_u, W_y]$ a *prediction filter*. It makes sense that the optimal prediction error should be equal to $e(t)$, which is the only uncertain part since $e(t-1), e(t-2), ...$ can be computed from data by exploiting the fact that the noise model is invertible.

d) Starting from $y(t) - \hat{y}(t) = e(t)$, show that the optimal predictor for system (32) can be written in the form (37) with

$$W_u(q) = H^{-1}(q)\,G(q), \tag{40}$$

$$W_y(q) = I - H^{-1}(q). \tag{41}$$

The basic idea behind the prediction-error method is to select θ such that the prediction errors become small in some sense. Let $\mathbb{Z}_N$ denote the set of data obtained from the system, that is

$$\mathbb{Z}_N = \{y(1), u(1), y(2), u(2), ..., y(N), u(N)\}. \tag{42}$$

Based on the data, we can compute the prediction errors $\varepsilon(t,\theta) = y(t) - \hat{y}(t)$, $t = 1, 2, 3, ..., N$ (the argument θ is included to emphasize that the prediction error is a function of the unknown parameters). It is common to allow for the flexibility of filtering this sequence, so we define

$$\varepsilon_F(t,\theta) = L(q)\,\varepsilon(t,\theta), \ 1 \le t \le N, \tag{43}$$

where $L(q)$ is a (linear) stable filter. The (filtered) prediction-error sequence can be viewed as a vector in $\mathbb{R}^N$, whose size can be measured using any norm in $\mathbb{R}^N$. We can formulate this as

$$V_N\left(\theta, \mathbb{Z}^N\right) = \frac{1}{N}\sum_{t=1}^{N} l\left(\varepsilon_F(t,\theta)\right), \tag{44}$$

where $l(\cdot)$ is a scalar-valued positive function. V_N is a scalar-valued function of $\theta \in D_{\mathcal{M}}$ for a given dataset $\mathbb{Z}^N$, and a measure for the size of the prediction-error sequence. The natural way to estimate θ^* is then to find the θ that minimizes (44), which we denote $\hat{\theta}_N$, that is

$$\hat{\theta}_N\left(\mathbb{Z}^N\right) = \arg\min_{\theta \in D_M} V_N\left(\theta, \mathbb{Z}^N\right). \tag{45}$$

While (45) is in general a nonlinear optimization problem, a particularly simple case arises when the underlying model structure is linear in the unknown parameters. In this case, the system takes the form

$$y(t) = \varphi^T(t)\,\theta^* + v_0(t), \tag{46}$$

where $\varphi(t)$ is a vector of known signals, called the *regression vector* (with components called *regressors*), and θ^* is the unknown parameter vector. The signal $v_0(t)$ may have a known, data-dependent component, as well as noise. That is $v_0(t) = \mu(t) + v(t)$, where $\mu(t)$ is known and $v(t)$ is not. $\mu(t)$ can easily be removed from the problem setup by defining a new output as $y(t) - \mu(t)$, therefore we can assume that $\mu(t) = 0$ without loss of generality. Taking the predictor as

$$\hat{y}(t) = \varphi^T(t)\,\theta, \tag{47}$$

the prediction-error is

$$\varepsilon(t,\theta) = y(t) - \varphi^T(t)\,\theta. \tag{48}$$

We now select $L(q) = 1$ (no filtering of the prediction-error sequence) and $l(\varepsilon) = \varepsilon^2/2$. From (44), we then get

$$V_N\left(\theta, \mathbb{Z}^N\right) = \frac{1}{2N}\sum_{t=1}^{N}\left(y(t) - \varphi^T(t)\,\theta\right)^2, \tag{49}$$

which is quadratic in θ, allowing the optimization problem (45) to be solved in closed form. The method is a special case of the prediction-error method and is called the *least-squares method*. Due to the linear model (46), the identification problem reduces to a linear regression problem. We denote the estimate $\hat{\theta}_N^{LS}\left(\mathbb{Z}^N\right)$, and it is given by

$$\hat{\theta}_N^{LS}\left(\mathbb{Z}^N\right) = \arg\min_{\theta\in\mathbb{R}^d}\frac{1}{2N}\sum_{t=1}^{N}\left(y(t) - \varphi^T(t)\,\theta\right)^2. \tag{50}$$

Problem 2 (Least-Squares Method)

a) Find the solution to the optimization problem (50). Are there any conditions on the data that need to be satisfied for the solution to exist?

b) Download the file "Data_Assignment2.mat" from It's learning, and load it into MATLAB by writing 'load Data_Assignment2' at the MATLAB prompt. The file contains data series of $u(t)$ and $y(t)$ from a process with unknown parameters. Assume the following model structure

$$A(q)\,y(t) = B(q)\,u(t) + v(t) \tag{51}$$

where

$$\begin{aligned} A(q) &= 1 + a_1q^{-1} + a_2q^{-2} && (52)\\ B(q) &= b_1q^{-1} + b_2q^{-2} + b_3q^{-3}, && (53) \end{aligned}$$

and assume that $v(t)$ is white noise with zero mean and unit variance. Write a predictor in the form (47). Using MATLAB, write a program that computes the least-squares estimate, $\hat{\theta}_N^{LS}$, for $N = \{10, ..., 10000\}$. Plot the estimated parameters as a function of N. Enclose this plot in your paper, along with the value $\hat{\theta}_{10000}^{LS}$.

c) If the output $y(t)$ is a p-vector, φ in (47) is a $d \times p$ matrix, and $l(\varepsilon)$ can be taken as $l(\varepsilon) = \varepsilon^T\Lambda^{-1}\varepsilon/2$, where Λ is a symmetric positive definite $p \times p$ matrix. Find an expression for the LS estimate in this case.

d) A slight modification of the LS method is obtained by assigning weights to the different measurements. In this case, the least-squares criterion changes to

$$V_N\left(\theta, \mathbb{Z}^N\right) = \sum_{t=1}^{N} \beta\left(N, t\right)\left(y\left(t\right) - \varphi^T\left(t\right)\theta\right)^2, \tag{54}$$

where $\beta\left(N, t\right)$ is the weight depending on time and N. Find an expression for the resulting estimate $\hat{\theta}_N^{WLS}\left(\mathbb{Z}^N\right)$ in this case.

It is of interest to anaylize the asymptotic behaviour of the least-square estimate as $N \to \infty$.

e) Using (46) and your result from a), show that the estimation error is given as

$$\tilde{\theta}_N = \hat{\theta}_N^{LS} - \theta^* = \left[\frac{1}{N}\sum_{t=1}^{N} \varphi\left(t\right)\varphi^T\left(t\right)\right]^{-1} \frac{1}{N}\sum_{t=1}^{N} \varphi\left(t\right) v_0\left(t\right). \tag{55}$$

Define

$$R^* = \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \varphi\left(t\right)\varphi^T\left(t\right), \tag{56}$$

and

$$f^* = \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \varphi\left(t\right) v_0\left(t\right). \tag{57}$$

Assuming, as above, that $\varphi\left(t\right)$ and $v_0\left(t\right)$ are quasi-stationary, the ergodicity results (21)–(23) state that the statistical properties of the data-series coincide with those of the ensemble with probability 1 as $N \to \infty$, so with probability 1, we have

$$R^* = \bar{E}\varphi\left(t\right)\varphi^T\left(t\right), \tag{58}$$

$$f^* = \bar{E}\varphi\left(t\right) v_0\left(t\right), \tag{59}$$

where $\varphi\left(t\right)$ and $v_0\left(t\right)$ now denote sequences of stochastic processes. We therefore obtain that

$$\lim_{N\to\infty} \tilde{\theta}_N = \left(R^*\right)^{-1} f^*. \tag{60}$$

It follows that for the estimate to be consistent (that is, $\lim_{N\to\infty} \tilde{\theta}_N = 0$), R^* must be invertible and f^* must be zero. While the first condition usually can be satisfied by the choice of input, the latter condition requires that $\varphi\left(t\right)$ and $v_0\left(t\right)$ are independent, which happens if

1. $v_0\left(t\right)$ is a sequence of independent stochastic variables with $\bar{E}v_0\left(t\right) = 0$ (white noise).
2. $u\left(t\right)$ is independent of $v_0\left(t\right)$ and past values of y do not enter into $\varphi\left(t\right)$ (such as FIR models, for instance).

These conditions are strong, and therefore one must in general expect bias in linear least-squares estimation. Many applications can tolerate a small bias (feedback control, for instance), while others may require consistent estimates. Let's analyze consistency for an example. Assume that data is generated by the process

$$y(t) = -a_0 y(t-1) + b_0 u(t-1) + v_0(t), \tag{61}$$

where

$$v_0(t) = e_0(t) + c_0 e(t-1) \tag{62}$$

and $\{u(t)\}$ and $\{e(t)\}$ are independent white noise processes with zero mean and unit variance ($Eu(t) = Ee(t) = 0$, $Eu^2(t) = Ee^2(t) = 1$). This is an ARMAX model. Suppose we do not know the noise model (62), and instead model the process as ARX. That is

$$y(t) = -ay(t-1) + bu(t-1) + e(t), \tag{63}$$

where $e(t)$ is assumed to be white noise. The choice of predictor would then be

$$\hat{y}(t) = -ay(t-1) + bu(t-1), \tag{64}$$

which corresponds to (47) with

$$\varphi(t) = \begin{bmatrix} -y(t-1) \\ u(t-1) \end{bmatrix}, \ \theta = \begin{bmatrix} a \\ b \end{bmatrix}. \tag{65}$$

f) Show that the LS estimate for b_0 is consistent while the estimate for a_0 is not by computing the bias as a function of a_0, b_0 and c_0 (Find the limit (60)).

Department of Engineering Cybernetics

TTK4215 System Identification and Adaptive Control
Assignment 3
Instrumental Variables Method

As demonstrated above, the least-squares estimate will in general lead to biased estimates, since the noise model can not be expected to be correct so that optimal predictors can be constructed. The least-squares method does not really take advantage of statistical properties, and might just as well be applied to deterministic systems. The method we will look at now seeks to avoid bias by rendering the estimation error independent of the data. Considering again the predictor (37), and the least-squares estimation method (50), it gives rise to the equation

$$\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\left[y(t)-\varphi^{T}(t)\,\hat{\theta}_{N}^{LS}\right]=0 \tag{66}$$

for determining the least-sqare estimate $\hat{\theta}_{N}^{LS}$, where data $y(t)$ is generated by (46). Substituting (46) into (66), we get

$$\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\left[-\varphi^{T}(t)\,\tilde{\theta}_{N}^{LS}+v_{0}(t)\right]=0. \tag{67}$$

In order to avoid bias, we now replace regression vector $\varphi(t)$ in (67) with a general correlation vector $\xi(t)$ to obtain

$$\frac{1}{N}\sum_{t=1}^{N}\xi(t)\left[-\varphi^{T}(t)\,\tilde{\theta}_{N}^{IV}+v_{0}(t)\right]=0. \tag{68}$$

If $\xi(t)$ can be chosen to be independent of $v_0(t)$, yet correlated with $\varphi(t)$ so that the matrix

$$\frac{1}{N}\sum_{t=1}^{N}\xi(t)\,\varphi^{T}(t) \tag{69}$$

is invertible, the estimate will be consistent. The elements of the vector $\xi(t)$ are called instrumental variables.

Problem 1 (Instrumental Variables Method)

a) Show that

$$\lim_{N\to\infty}\tilde{\theta}_{N}^{IV}=\left(\bar{E}\xi(t)\,\varphi^{T}(t)\right)^{-1}\bar{E}\xi(t)\,v_{0}(t). \tag{70}$$

It remains to select the instrumental variables. We will illustrate how this can be done for an ARMAX type system in open loop (u independent of v). Assume the process is

$$y(t)+a_{1}^{0}y(t-1)+\cdots+a_{n}^{0}y(t-n)=b_{1}^{0}u(t-1)+\cdots+b_{m}^{0}u(t-m)+v(t), \tag{71}$$

and take the predictor as (37) with

$$\begin{aligned}\varphi(t) &= \begin{bmatrix} -y(t-1) & \cdots & -y(t-n) & u(t-1) & \cdots & u(t-m) \end{bmatrix}^T \quad (72)\\ \theta &= \begin{bmatrix} a_1 & \cdots & a_n & b_1 & \cdots & b_m \end{bmatrix}^T. \quad (73)\end{aligned}$$

Clearly, $\varphi(t)$ is correlated with $v(t)$ (unless v is white) due to the $y-$ elements appearing in it. Therefore, we can not allow y to appear in the instrumental variables ξ. However, it seems reasonable to produce ξ in a similar manner as φ in order to ensure that matrix (69) is invertible. We propose to generate a signal that resembles y by using the same structure as (71) as follows

$$x(t) + a_1 x(t-1) + \cdots + a_n x(t-n) = b_1 u(t-1) + \cdots + a_m u(t-m), \quad (74)$$

where the input sequence $u(t)$ is the same as that applied to the real process. The sequence $x(t)$ can be expected to resemble $y(t)$, while being independent of $v(t)$. A natural choice for the instruments is now

$$\xi(t) = \begin{bmatrix} -x(t-1) & \cdots & -x(t-n) & u(t-1) & \cdots & u(t-m) \end{bmatrix}^T. \quad (75)$$

In computing the $x(t)$ sequence, we have to make an a priori choice of the parameter vector θ (which of course is unknown, since the objective of the IV method is to find it!). A straight forward choice is to use the least-squares estimate, to obtain

$$x(t) = \xi^T(t)\,\hat{\theta}_N^{LS}. \quad (76)$$

b) Starting from equation (66) for the least-squares estimate, show that the instrumental-variables estimate, once the instrumental variables have been produced, is given by

$$\hat{\theta}_N^{IV} = \left(\frac{1}{N}\sum_{t=1}^{N}\xi(t)\,\varphi^T(t)\right)^{-1}\frac{1}{N}\sum_{t=1}^{N}\xi(t)\,y(t). \quad (77)$$

c) Download the file "Data_Assignment3.mat" from It's learning, and load it into MATLAB by writing 'load Data_Assignment3' at the MATLAB prompt. The file contains data series of $u(t)$ and $y(t)$ from a process with unknown parameters. Assume the following model structure

$$A(q)\,y(t) = B(q)\,u(t) + v(t) \quad (78)$$

where

$$\begin{aligned}A(q) &= 1 + a_1 q^{-1} + a_2 q^{-2} \quad (79)\\ B(q) &= b_1 q^{-1} + b_2 q^{-2} + b_3 q^{-3}, \quad (80)\end{aligned}$$

and assume that $v(t)$ is white noise with zero mean and unit variance. Using MATLAB, write a program that computes the least-squares estimate, $\hat{\theta}_N^{LS}$, for $N = \{10, ..., 10000\}$. Plot the estimated parameters as a function of N. Enclose this plot in your paper, along with

$\hat{\theta}_{10000}^{LS}$ (if you did Assignment 2, you already have the program, but you need to run it again because of the new data).

d) Suppose now that $v(t)$ cannot be assumed white. Suggest a procedure for computing an estimate using the instrumental variables method. Using MATLAB, write a program that computes $\hat{\theta}_N^{IV}$, for $N = \{10, ..., 10000\}$. Plot the estimated parameters as a function of N. Enclose this plot in your paper, along with $\hat{\theta}_{10000}^{IV}$.

For rigorous analysis of the prediction error method, LS method and IV method, see [1, 2].

Problem 2 (Extremum Seeking) *And now, for something completely different*: Read sections 1-3 in [3], implement the peak-seeking algorithm in MATLAB and apply it to the static function $f : \mathbb{R} \rightarrow \mathbb{R}$

$$f(\theta) = -\theta^2 + 2\theta + 5. \tag{81}$$

Enclose a plot of $\hat{\theta}$ as a function of time for at least 3 different initial conditions $\hat{\theta}(0)$ in your paper.

References

[1] L. Ljung, *System Identification, Theory for the User*, 2nd. Ed., Prentice-Hall, 1999.

[2] T. Söderström and P. Stoica, *System Identification*, Prentice-Hall, 1989.

[3] M. Krstic and H.-H. Wang, "Stability of extremum seeking feedback for general nonlinear dynamic systems," *Automatica*, vol. 36, pp. 595–601, 2000.