2.160 System Identification, Estimation, and Learning

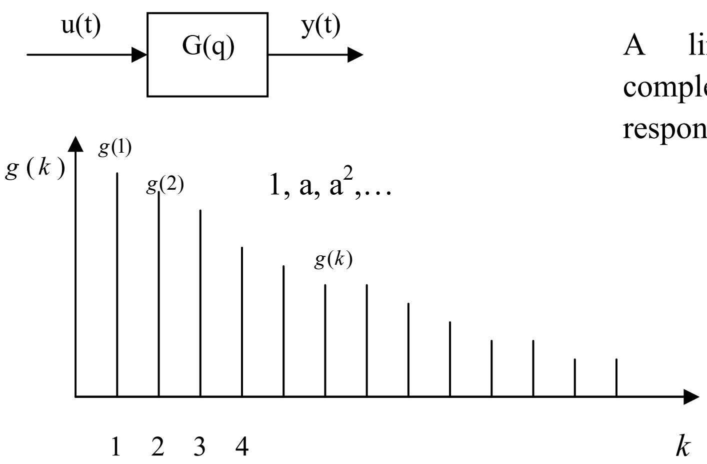
Lecture Notes No. 10

March 13, 2006

6 Model Structure of Linear Time Invariant Systems

6.1 Model Structure

In representing a dynamical system, the first step is to find an appropriate structure of the model. Depending on the choice of model structure, efficiency and accuracy of modeling are significantly different. The following example illustrates this. Consider the impulse response of a stable, linear time-invariant system, as shown below. Impulse Response is a generic representation that can represent a large class of systems, but is not necessarily efficient, i.e. it often needs a lot of parameters for representing the same input-output relationship than other model structures.



A linear time-invariant system is completely characterized by its impulse response

$$G(q) = \sum_{k=1}^{\infty} g(k) q^{-k}$$

$g(1), g(2), \ldots$ Too many parameters although truncated.

Can we represent the system with fewer parameters?

Consider $g(k) = a^{k-1}$ $\quad k = 1,2,3,\ldots$

$$G(q) = \sum_{k=1}^{\infty} a^{k-1} q^{-k}$$

Multiplying $\frac{a}{q}$: $\quad \frac{a}{q} G(q) = \sum_{k=1}^{\infty} a^k q^{-k-1} = \sum_{k=2}^{\infty} a^{k-1} q^{-k} = G(q) - \frac{1}{q}$

$$\left(1 - \frac{a}{q}\right) G(q) = \frac{1}{q} \qquad \therefore G(q) = \frac{q^{-1}}{1 - aq^{-1}} = \frac{1}{q-a}$$

Therefore, $G(q)$ is represented by only one parameter: one pole when using a rational function.

The number of parameters reduces if one finds a proper model structure. The following section describes various structures of linear time-invariant systems.

6.2 A Family of Transfer Function Models

6.2.1 ARX Model Structure

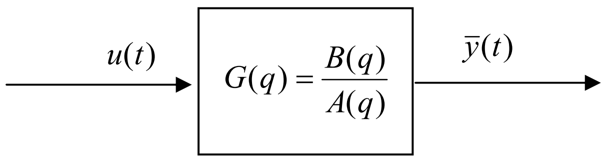
Consider a rational function for $G(q)$:

$$\bar{y}(t) = \frac{B(q)}{A(q)} u(t) \tag{1}$$

where $A(q)$ and $B(q)$ are polynomials of q:

$$A(q) \equiv 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}\ , \tag{2}$$

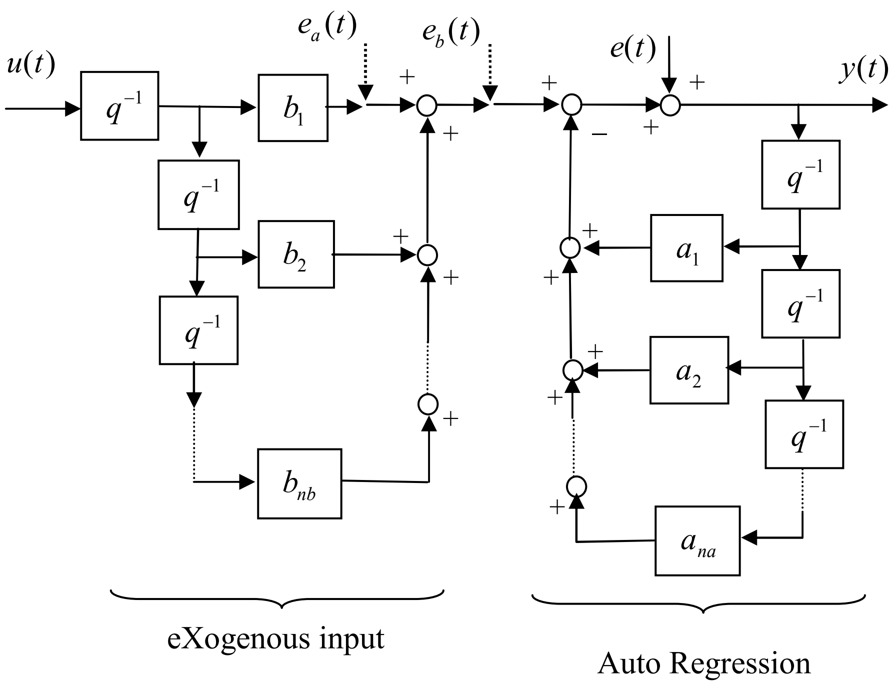
$$B(q) \equiv b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}$$



The input-output relationship is then described as

$$\begin{aligned} &\bar{y}(t) + a_1 \bar{y}(t-1) + \ldots + a_n \bar{y}(t-n_a) \\ &\quad = b_1 u(t-1) + \ldots + b_{n_b} u(t-n_b) \end{aligned} \tag{3}$$

See the block diagram below.



Now let us consider an uncorrelated noise input $e(t)$ entering the system. As long as the noise enters anywhere between the output $y(t)$ and the block of b_1, i.e. $e(t), e_a(t), e_b(t)$ in the above block diagram, the dynamic equation remains the same and is given by:

$$\begin{aligned} y(t) + a_1 y(t-1) + \ldots + a_n y(t-n_a) \\ = b_1 u(t-1) + \ldots + b_{n_b} u(t-n_b) + e(t) \end{aligned} \tag{4}$$

Including the noise term, this model is called "Auto Regressive with eXogenous input" model, or ARX Model for short. Using the polynomials $A(q)$ and $B(q)$, (4) reduces to

$$A(q)y(t) = B(q)u(t) + e(t) \tag{5}$$

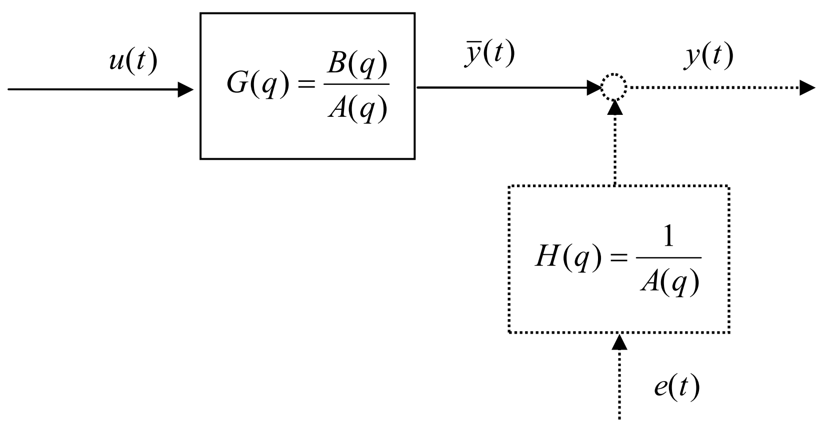
The adjustable parameters involved in the ARX model are

$$\theta = \left(a_1, a_2, \ldots, a_{n_a},\ b_1,\ b_2, \ldots, b_{n_b}\right)^T \tag{6}$$

Comparing (5) with (11) of Lecture Notes 9 yields

$$G(q,\theta) = \frac{B(q)}{A(q)} \qquad H(q,\theta) = \frac{1}{A(q)} \tag{7}$$

See the block diagram below.



Note that the uncorrelated noise term $e(t)$ enters as a direct error in the dynamic equation. This class of model structures, called Equation Error Model, has a favorable characteristic leading to a linear regression, which is easy to identify.

Note that if $n_a = 0$ then $y(t) = B(q)u(t) + e(t)$. This is called a Finite Impulse Response (FIR) Model, as we have seen before.

6.2.2 Linear Regressions

Consider the one-step-ahead prediction model in Lecture Notes 9-(18)

$$\hat{y}(t|t-1) = H^{-1}(q)G(q)u(t) + \left[1 - H^{-1}(q)\right]y(t) \tag{9-(18)}$$

From (7), the predictor of ARX model is given by

$$\hat{y}(t|\theta) = B(q)u(t) + [1 - A(q)]y(t) \tag{8}$$

This can be more directly obtained from (4), since the noise is uncorrelated.

The regression vector associated with this prediction model of ARX is defined as:

$$\varphi(t) \equiv [-y(t-1),...,-y(t-n_a), u(t-1),...,u(t-n_b)]^T \tag{9}$$

using this regression vector, (8) reduces to

$$\hat{y}(t|\theta) = \varphi^T(t)\theta \tag{10}$$

Note that the predictor $\hat{y}(t|\theta)$ is a scalar function of

$\varphi(t)$: a known vector, and

θ : adjustable parameters.

$\varphi(t)$ does not include any information of θ.

The known and unknown parts are separated, and θ is linearly involved in the predictor.

… a Linear Regression.

6.2.3 ARMAX Model Structure

Linear regressions can be obtained only for a class of model structures. Many others cannot be written in such a manner where a parameter vector is linearly involved in the predictor model. Consider the following input-output relationship:

$$\begin{aligned} y(t) + a_1 y(t-1),...,+a_{n_a} y(t-n_a) &= b_1 u(t-1),...,b_{n_b} u(t-n_b) \\ &+ e(t) + c_1 e(t-1) + ... + c_{n_c} e(t-n_c) \end{aligned} \tag{11}$$

Using $c(q) = 1 + c_1 q^{-1} + ... + c_{n_c} q^{-n_c}$, (11) reduces to

$$A(q)y(t) = B(q)u(t) + C(q)e(t) \tag{12}$$

Therefore

$$G(q,\theta) = \frac{B(q)}{A(q)} \qquad H(q,\theta) = \frac{C(q)}{A(q)} \tag{13}$$

$$\theta = [a_1, a_2, \cdots, a_{na}, b_1, b_2, \cdots, b_{nb}, c_1, c_2, \cdots, c_{nc}]^T \qquad (14)$$

This model structure consists of the Moving Average part (MA), $C(q)e(t)$, the Auto Regressive(AR) part, $A(q)y(t)$, and the eXogenous input part (X). This model structure is called an ARMAX model for short. An ARMAX model cannot be written as a linear regression.

6.2.4 Pseudo-linear Regressions

The one-step-ahead predictor for the above ARMAX model is given by

$$\hat{y}(t|\theta) = \overset{GH^{-1}}{\frac{B(q)}{C(q)}} u(t) + \overset{1-H^{-1}}{[1 - \frac{A(q)}{C(q)}]} y(t) \qquad (15)$$

This cannot be written in the same form as the linear regression, but can be written in a similar (apparently linear) inner product. Multiplying $c(q)$ to both sides of (15) and adding $[1 - c(q)]\hat{y}(t|\theta)$ yields

$$\hat{y}(t|\theta) = B(q)u(t) + [1 - A(q)]y(t) + [c(q) - 1]\varepsilon(t,\theta) \qquad (16)$$

Define Prediction Error

$$\varepsilon(t,\theta) = y(t) - \hat{y}(t|\theta) \qquad (17)$$

and vector $\varphi(t)$ as

$$\varphi(t,\theta) \equiv [-y(t-1), \ldots, -y(t-n_a), u(t-1), \ldots, u(t-n_b)$$
$$\varepsilon(t-1,\theta), \ldots \varepsilon(t-n_c,\theta)]^T \qquad (18)$$

Then (15) reduces to $\hat{y}(t|\theta) = \varphi^T(t,\theta)\theta$ (19)

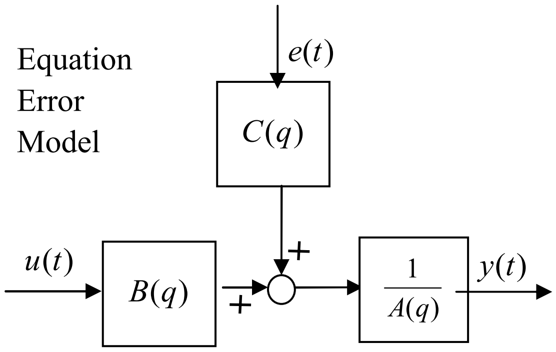
Note that $\varepsilon(t,\theta)$ includes θ and therefore φ depends on θ. Strictly speaking (15) is not a linear function of θ…

….. A Pseudo Linear Regression

This will make a significant difference in parameter estimation as discussed later.

6.2.5 Output Error Model Structure

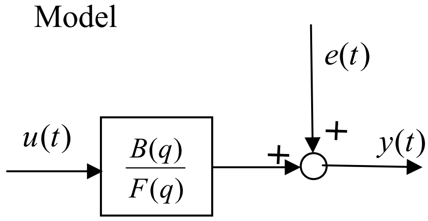
Equation Error Model



Noise enter the process.

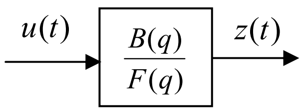
This resembles Preocess Noise in the Kalman Filter.

Output Error Model



Noise dynamics is independent of the process dynamics
This resembles the measurement noise of KF

Let $z(t)$ be undisturbed output driven by $u(t)$ alone,



(20)
$$z(t)+f_1 z(t-1),\ldots,+f_{n_f} z(t-n_f)$$
$$=b_1 u(t-1),\ldots,b_{n_b} u(t-n_f)$$

$$F(q)=1+f_1 q^{-1}+\ldots+f_{n_c} q^{-n_c}$$

Note that $z(t)$ is not an observable output. What we observed is $y(t)$

(21) $$y(t)=\frac{B(q)}{F(q)}u(t)+e(t)$$

The parameters to be determined are collectively given by

(22) $$\theta=[b_1 \quad b_2 \ldots b_{n_b} \; f_1 \quad f_2 \ldots \; f_{n_f} \quad]^{\mathrm{T}}$$

Note that $z(t)$ is a variable to be computed (estimated) based on the parameter vector θ; therefore, $z(t,\theta)$. The one-step-ahead predictor is

(23) $$\hat{y}(t|\theta)=\frac{B(q)}{F(q)}u(t)=z(t,\theta)$$

Which is nothing but $z(t,\theta)$. Therefore, $\hat{y}(t-1|\theta)=z(t-1,\theta)$

$$
\begin{aligned}
\hat{y}(t|\theta) &= -f_1 z(t-1,\theta) - -f_2 z(t-2,\theta) - \cdots - f_{n_f} z(t-n_f,\theta) \\
&\quad + b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b) \qquad (24) \\
&= \varphi^T(t,\theta)\cdot\theta
\end{aligned}
$$

where $\varphi(t,\theta)$ is

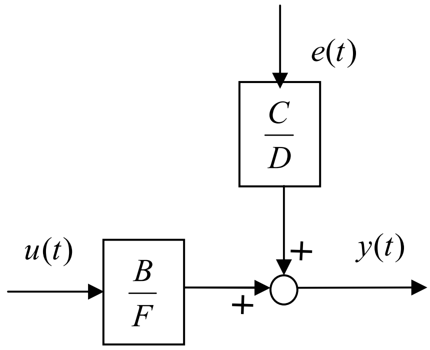
(25) $$\varphi(t,\theta) = [u(t-1)\cdots u(t-n_b) - z(t-1,\theta)\cdots - z(t-n_f,\theta)]^T$$

Therefore this is a Pseudo-Linear Regression.

Box-Jenkins Model Structure

This simple output error (OE) model can be extended to the one having an ARMA model for the error dynamics

(26) $$y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t)$$



6.3 State Space Model

State variables $x(t) = [x_1(t), x_2(t), \cdots x_n(t)]^T$

Stationary Time-invariant

(27) $$x(t+1) = A(\theta)x(t) + B(\theta)u(t) + w(t) \qquad A(\theta) \in R^{n\times n}$$

(28) $$y(t) = C(\theta)x(t) + v(t) \qquad B(\theta) \in R^{n\times m}$$

$$C(\theta) \in R^{l\times n}$$

Matrix $A(\theta)$, $B(\theta)$ and $C(\theta)$ contain parameters to be determined, θ

$\{w(t)\}$ and $\{u(t)\}$ are process and output noises, respectively, with zero mean values and covariance matrices:

(29)
$$
\begin{aligned}
E[w(t)w^T(t)] &= R_1(\theta) \\
E[v(t)v^T(t)] &= R_2(\theta) \\
E[w(t)v^T(t)] &= R_{12}(\theta)
\end{aligned}
$$

Usually $R_{12}(\theta) = 0$

Using forward shift operation q, we can rewrite (27) as

$$[qI - A(\theta)]x(t) = B(\theta)u(t) + w(t)$$

Therefore the output $y(t)$ is given by

(30) $$(t) = C(\theta)[qI - A(\theta)]^{-1} B(\theta)u(t) + C(\theta)[qI - A(\theta)]^{-1} w(t) + v(t)$$

from (28)

Comparing this with

$$y(t) = G(q,\theta)u(t) + H(q,\theta)e(t) \qquad (2)$$

(31) $$H(q,\theta)e(t) \longleftrightarrow C(\theta)[qI - A(\theta)]^{-1} w(t) + v(t)$$

Equation Error Model w/ $A(q)$

OE Model

Innovations representation of the Kalman filter.

Let $\hat{x}(t,\theta)$ be the estimated state using the Kalman filter.

The prediction error given by

(32) $$y(t) - C(\theta)\hat{x}(t,\theta) = C(\theta)[x(t) - \hat{x}(t,\theta)] + v(t) \equiv e(t)$$

$$[qI - A(\theta)]^{-1} Bu + qI - A(\theta)]^{-1} w]$$

represents the part of $y(t)$ that ca This part is called, the "innovation", denoted $e(t)$. Using this innovation, K-F is ritten as

(33) $$\hat{x}(t+1,\theta) = A(\theta)\hat{x}(t,\theta) + B(\theta)u(t) + k(\theta)e(t)$$

(34) $$y(t) = C(\theta)\hat{x}(t,\theta) + e(t)$$

The covariance of innovation $e(t)$ is

(35) $$E[e(t)e^T(t)] = c(\theta)P(\theta)C^T(\theta) + R_2(\theta)$$

Error covariance of state estimation

$e(t)$ and $\hat{x}$ are not correlated

Combining (33) and (34), and comparing it with (2),

$$y(t) = G(q,\theta)u(t) + H(q,\theta)e(t) \tag{36}$$

$$\begin{aligned} G(q,\theta) &= C(\theta)[qI - A(\theta)]^{-1} B(\theta) \\ H(q,\theta) &= C(\theta)[qI - A(\theta)]^{-1} K(\theta) + I \end{aligned} \tag{37}$$

This shows the relationship between the state space model and the input-output model. They are connected through the innovation process.

$$(qI - A)\hat{x}(t,\theta) = Bu(t) + Ke(t)$$

$$\hat{x}(t,\theta) = (qI - A)^{-1} Bu(t) + (qI - A)^{-1} Ke(t) \tag{38}$$

$$y(t) = C(qI - A)^{-1} Bu(t) + C(qI - A)^{-1} Ke(t) + e(t)$$

6.4 Consistent and Unbiased Estimation: Preview of Part 3, System ID

This section briefly describes some important issues on model structure in estimating the parameters involved in the model. Details will be discussed in Part 3, System Identification. Let Z^N be a set of data obtained over the period of time: $1 \le t \le N$. One of the critical issues in system identification is whether the estimated model parameters $\hat{\theta}_N$ based on the data set Z^N approaches the true values θ_0, as the number of data points N tends to infinity. Several conditions must be met to guarantee this important property, called "Consistency". First, the model structure must be the correct one. Second, the data set Z^N must be rich enough to identify all the parameters involved in the model. Furthermore, it depends on whether the noise term $v(t)$ entering the system is correlated, which estimation algorithm is used for determining $\hat{\theta}_N$, and how the parameters of the model are involved in the predictor $\hat{y}(t \mid \theta)$.

Consider the following squared norm of prediction error:

$$V_N(\theta, Z^N) = \frac{1}{N} \sum_{t=1}^{N} \frac{1}{2} \left(y(t) - \hat{y}(t \mid \theta) \right)^2 \tag{39}$$

Assume that the one-step-ahead predictor is given by a linear regression:

$$\hat{y}(t|\theta) = \varphi^T(t,\theta)\theta \tag{19}$$

The Least Square Estimate (LSE) provides the optimal parameters minimizing the mean squared error $V_N(\theta, Z^N)$:

$$\hat{\theta}_N^{LS} = \arg\min_{\theta} V_N(\theta, Z^N) = \left(R(N) \right)^{-1} f(N) \tag{40}$$

where

$$R(N) = \frac{1}{N} \sum_{t=1}^{N} \varphi(t)\varphi(t)^T \quad \text{and} \quad f(N) = \frac{1}{N} \sum_{t=1}^{N} \varphi(t)\, y(t) \tag{41}$$

Suppose that the model structure is correct, and real data are generated from the true process with the true parameter values θ_0:

$$y(t) = \varphi(t)^T \theta_0 + v_0(t) \tag{42}$$

Whether the estimate $\hat{\theta}_N^{LS}$ is consistent or not depends on the data set and the stochastic properties of the noise term $v_0(t)$. Substituting the expression of the true process into $f(N)$ yields

$$f(N) = \frac{1}{N}\sum_{t=1}^{N}\varphi(t)\left(\varphi(t)^T \theta_0 + v_0(t)\right) = R(N)\theta_0 + \underbrace{\frac{1}{N}\sum_{t=1}^{N}\varphi(t)v_0(t)}_{f^*(N)} \tag{43}$$

$$\hat{\theta}_N^{LS} - \theta_0 = \left(R(N)\right)^{-1}\left[R(N)\theta_0 + f^*(N)\right] - \theta_0 = \left(R(N)\right)^{-1} f^*(N) \tag{44}$$

To be consistent, i.e. $\lim_{N\to\infty}\hat{\theta}_N^{LS} = \theta_0$, the following conditions must be met:

(I) Matrix $\lim_{N\to\infty} R(N)$ must be non-singular. The data series, $\varphi(1), \varphi(2), \varphi(3), \cdots$, must be able to "Persistently Exciting" the system.

(II) $\lim_{N\to\infty} f^*(N) = 0$. This can be achieved in two ways:

Case A: $v_0(t)$ is an uncorrelated random process with zero mean values. Then, $v_0(t)$ is not correlated with $y(t-1), y(t-2), y(t-3), \cdots$ and $u(t-1), u(t-2), u(t-3), \cdots$, i.e. all the components of $\varphi(t)$. Therefore,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)v_0(t) = 0$$ *

Case B: The model structure is FIR, i.e. $n_a = 0$, and inputs $u(t-1), u(t-2), \cdots$ are uncorrelated with $v_0(t)$. The noise term $v_0(t)$ itself may be correlated, for example, $v_0(t) = H(q,\theta_0)e(t)$. If the model structure is FIR with uncorrelated inputs, then $\varphi(t) \leftrightarrow v_0(t)$ are uncorrelated, hence $\lim_{N\to\infty} f^*(N) = 0$.

The above two are straightforward cases; Consistent estimates are guaranteed with simple LSE, as long as the data are persistently exciting. Care must be taken for other model structures and correlated noise term. For example, ARMAX model is used, the linear regression cannot be used, and the output sequence involved in $\varphi(t)$ may be correlated with $v_0(t)$:

$$\varphi(t) = [-y(t-1), -y(t-2), \cdots]$$

$$y(t-1) = \varphi(t-1)^T\theta_0 + v_0(t-1) \rightarrow$$ This may be correlated with $v_0(t)$

---

* Ergodicity of the random process is assumed.