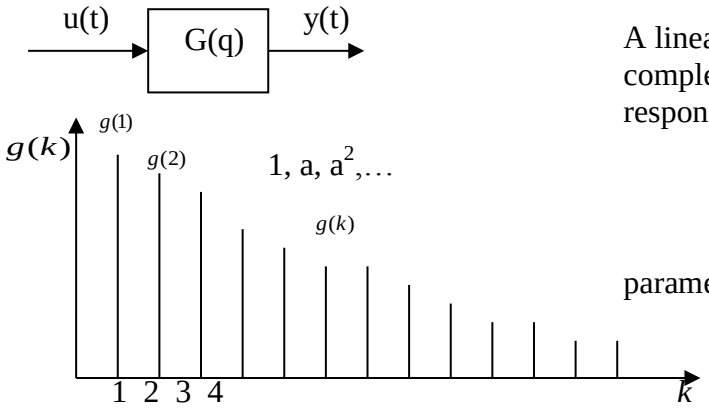
2.160 IDENTIFICATION, ESTIMATION, AND LEARNING

LECTURE NOTES NO. 12

13. Parametric Identification of Linear Time Invariant Systems

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

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

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

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

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.

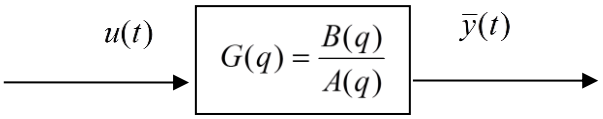
13.2 Auto-Regressive Model with eXogenous Input (ARX)

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:

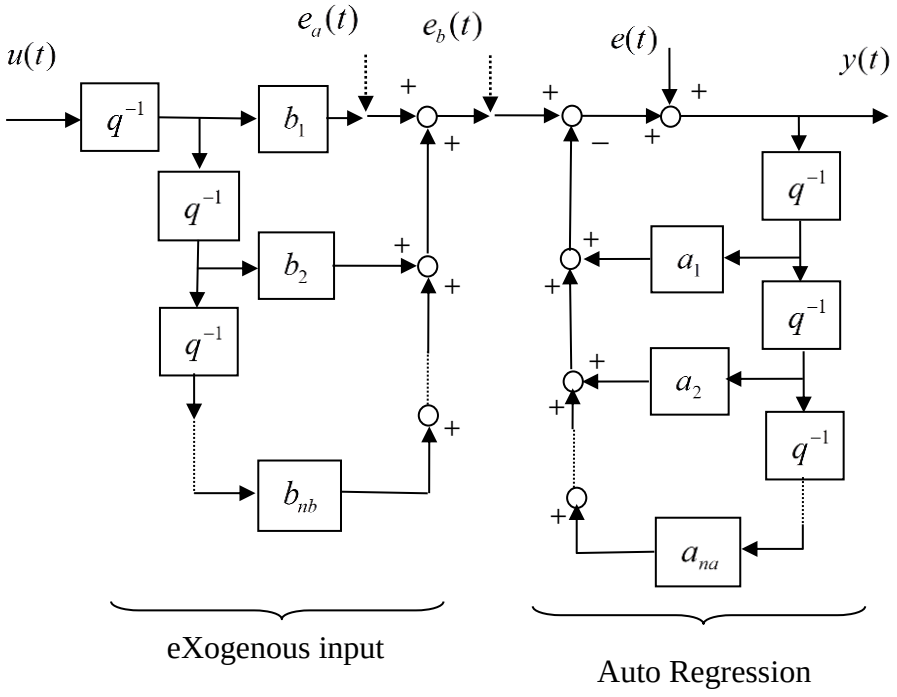
$$A(q) \equiv 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}, \tag{2}$$
$$B(q) \equiv b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}$$



The input-output relationship is then described as

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

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:

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

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

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

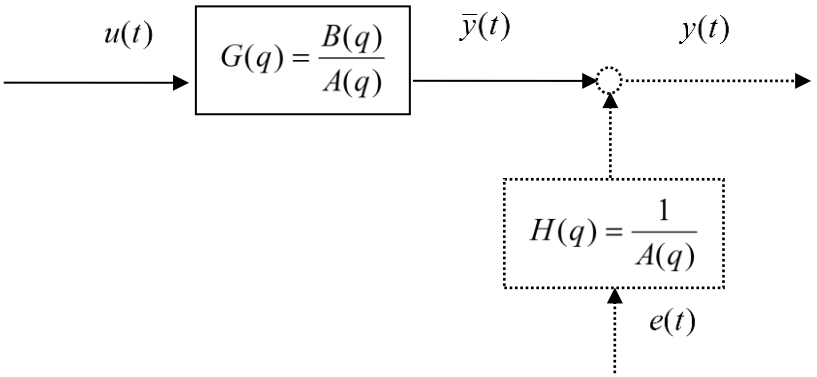
The adjustable parameters involved in the ARX model are

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

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

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

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.

13.3 Linear Regression and Optimal Parameter Estimation

Based on the ARX model given by (4), let us consider the best prediction of output $\hat{y}(t)$ given all exogenous inputs up to time *t* and output observations up to time *t*-1. Noting that the expected value of noise $e(t)$ is zero, our best prediction is given by

$$\hat{y}(t \mid \theta) = -a_1 y(t-1) - ... - a_n y(t-n_a) + b_1 u(t-1) + ... + b_{n_b} u(t-n_b) \tag{8}$$

where $\hat{y}(t|\theta)$ indicates that this predictor depends on the parameters involved in the model. Note that all the parameters to be tuned are linearly involved in this predictor. Defining the regression vector associated with this predictor of ARX model as:

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

(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 components are separated, and θ is linearly involved in the predictor. This is referred to as a Linear Regression.

The parameter vector θ can be tuned easily from experiment data. Suppose that data $\varphi(1),\cdots,\varphi(N)$, $N \gg n_a + n_b$ have been given. We want to find the parameter vector θ such that it minimizes the squared error:

$$\hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{N} (y(t) - \varphi^T(t)\theta)^2 = \left[\sum_{t=1}^{N} \varphi(t)\varphi^T(t)\right]^{-1} \sum_{t=1}^{N} \varphi(t)y(t) \tag{11}$$

This is a standard Least Squares solution.

13.4 Auto-Regressive Moving Average Model with eXogenous Input (ARMAX) and Noise Dynamics

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:

$$\begin{aligned} y(t) + a_1 y(t-1),...,+a_{n_a} y(t-n_a) &= b_1 u(t-1),...,b_{n_b} u(t-n_b) \\ &+ e(t) + c_1 e(t-1) + ... + c_{n_c} e(t-n_c) \end{aligned} \tag{12}$$

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

$$A(q)y(t) = B(q)u(t) + C(q)e(t) \tag{13}$$

Therefore

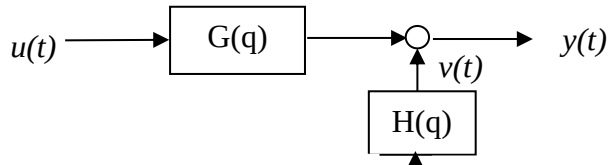
$$G(q,\theta) = \frac{B(q)}{A(q)} \qquad H(q,\theta) = \frac{C(q)}{A(q)} \tag{14}$$

$$\theta = [a_1, a_2, \cdots, a_{na}, b_1, b_2, \cdots, b_{nb}, c_1, c_2, \cdots, c_{nc}]^T \tag{15}$$

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.

One of the major differences between ARX and ARMAX is that the uncorrelated noise $e(t)$ directly enters the system along the forward path in the ARX model, while it goes through its own dynamics (moving average) in the ARMAX model. It is important to quantify such noise dynamics by extracting some structure elucidating how noise influences the system. This can be accomplish by separating such a structure from a totally unpredictable uncorrelated random process $e(t)$.

Consider that output $y(t)$ comes from both deterministic input $u(t)$ through a deterministic system $G(q)$ and a random process $v(t)$.



v(t) includes measurement noise and process noise (disturbances)

e(t)= uncorrelated zero mean value random process (statistically independent)

To model *v(t)* we assume that *v(t)* comes from a dynamic process *H(q)* driven by a totally unpredictable random process *e(t)*

H(q) = deterministic; to be identified
e(t) = uncorrelated, zero mean

$$E[e(t)]=0 \qquad E[e(t)e(s)]=\begin{cases}0 & t\neq s\\ \lambda & t=s\end{cases} \tag{16}$$

$$y(t)=G(q)u(t)+H(q)e(t) \tag{17}$$

Depending on model structure the two transfer functions $G(q), H(q)$ may take different forms.

13.5 Prediction

We address system identification problems based on the framework of Prediction-Error approach. To this end let us first obtain a predictor for the model given by (17).

If we know $G(q)$ and $H(q)$, how can we best predict output $y(t)$ based on output observations up to and including $t-1$? Since $G(q)u(t)$ is deterministic, this prediction problem is basically equivalent to predict $v(t)$

$$v(t)\equiv H(q)e(t)=y(t)-G(q)u(t) \tag{18}$$

based on observation of v up to $t-1$, $\{v(s)|s \le t-1\}$.

Prediction of v

Without loss of generality we assume $H(q)$ is monic.

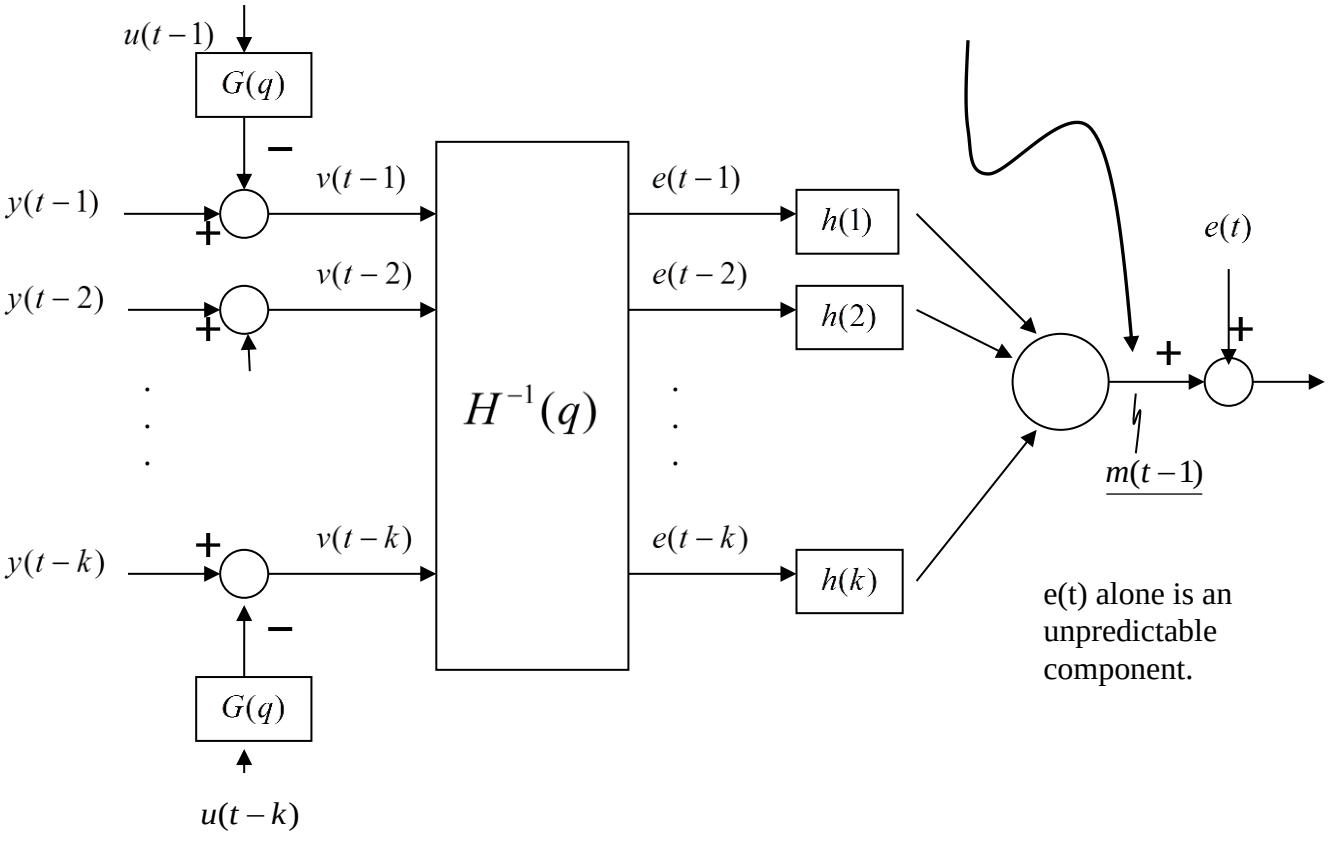
$$H(q) = 1 + \sum_{k=1}^{\infty} h(k) q^{-k}$$

$$h(0) = 1$$

Then

(19)
$$v(t) = e(t) + \underbrace{\sum_{k=1}^{\infty} h(k) e(t-k)}_{\text{Replace this by } m(t-1)}$$

Note
Inverting $H(q)$, we can relate $e(t-1)$, $e(t-2)$,… to $v(t-1), v(t-2)$,….
Since we observed, $y(t-1), y(t-2)$…,
we can reconstruct $\underline{m(t-1)}$



e(t) alone is an unpredictable component.

Let $f_e(x)$ be the probability density function (PDF) of random variable *e*.

$$P_{rob}(x \le e(t) \le x + \Delta x) \cong f_e(x)\Delta x$$

From(14) $e(t) = v(t) - m(t-1)$ (20)

Therefore,

$$\begin{aligned} f_v(x)\Delta x &\cong P_{rob}\left(x \le v(t) \le x + \Delta x \middle| v(t-1), v(t-2), \ldots\right) \\ &= P_{rob}\left(x \le m(t-1) + e(t) \le x + \Delta x\right) \\ &= P_{rob}\left(x - m(t-1) \le e(t) \le x - m(t-1) + \Delta x\right) \\ &= f_e\left(x - m(t-1)\right)\Delta x \end{aligned}$$

Therefore, the most probable value of $v(t)$, given $m(t-1)$, is the one for which the PDF $f_e\left(x - m(t-1)\right)$ takes its maximum.

….The maximum a posteriori (MAP) prediction.

If we use the mean value of the distribution,
That is given by the mean of $e(t)$,i.e. $E[e(t)] = 0$

v

$f_e\left(x - m(t-1)\right)$

mean 0

$v \leftarrow m(t-1)$

This is a conditional mean, hence we write

$$\hat{v}(t|t-1) = m(t-1) = \sum_{k=1}^{\infty} h(k)e(t-k) \tag{21}$$

$$\min_{\hat{v}(t)} E\left[v(t) - \hat{v}(t)\right]^2 \quad \Longrightarrow \quad \hat{v}(t) = \hat{v}(t|t-1)$$

$$\hat{v}(t|t-1) = e(t) + \underbrace{\sum_{k=1}^{\infty} h(k)e(t-k)}_{H(q)e(t)} - e(t) = \left(H(q) - 1\right)e(t) \tag{22}$$

Combining the deterministic (known) term $G(q)u(t)$ with $\hat{v}(t|t-1)$, we can obtain the most probable output prediction, given $m(t-1)$

$$\begin{aligned} \hat{y}(t|t-1) &= G(q)u(t) + \hat{v}(t|t-1) \\ &= G(q)u(t) + \left[H(q) - 1\right]e(t) \\ &= G(q)u(t) + \left[1 - H^{-1}(q)\right]\left[y(t) - G(q)u(t)\right] \end{aligned}$$

$e(t) = H^{-1}(q)v(t)$

$y(t) - G(q)u$

$$\hat{y}(t|t-1) = H^{-1}(q)G(q)u(t) + \underbrace{\left[1 - H^{-1}(q)\right]}_{\text{Note this does not include } y(t),\ \text{but includes } y(t-1),\ y(t-2)\ldots} y(t) \tag{23}$$

This is called "One-Step Ahead Prediction Model".

The Prediction Error:

$$\begin{aligned} y(t) - \hat{y}(t|t-1) &= -H^{-1}(q)G(q)u(t) + H^{-1}(q)y(t) \\ &= H^{-1}(q)\underbrace{\left[y(t) - G(q)u(t)\right]}_{v(t)} \\ &= e(t) \end{aligned} \tag{24}$$

This variable $e(t)$ represents the part of $y(t)$ that cannot be predicted from past data. So, the one-step ahead prediction model exploits all usable information contained in the observations, leaving only the "garbage" $e(t)$.

Now we define the goal of our system identification problem is to find *G(q)* and *H(q)* from input-output data.

In Kalman filter, two separate noise sources, i.e. process noise and measurement noise, were considered. However, identifying the two noise characteristics separately is difficult or infeasible for many practical applications. In the following we will consider an aggregated noise model originated in a single random process.

13.6 Pseudo-linear Regressions

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

$$\hat{y}(t|\theta) = \underbrace{\frac{B(q)}{C(q)}}_{GH^{-1}} u(t) + \underbrace{\left[1 - \frac{A(q)}{C(q)}\right]}_{1-H^{-1}} y(t) \tag{25}$$

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 (25) and adding $[1-c(q)]\hat{y}(t|\theta)$ yields

$$\hat{y}(t|\theta) = B(q)u(t) + [1 - A(q)]y(t) + [c(q) - 1]\varepsilon(t,\theta) \tag{26}$$

Define Prediction Error

$$\varepsilon(t,\theta) = y(t) - \hat{y}(t|\theta) \tag{27}$$

and vector $\varphi(t)$ as

$$\varphi(t,\theta) \equiv [-y(t-1),...,-y(t-n_a),u(t-1),...,u(t-n_b) \quad \varepsilon(t-1,\theta),...\varepsilon(t-n_c,\theta)]^T \tag{28}$$

Then (25) reduces to $\hat{y}(t|\theta) = \varphi^T(t,\theta)\theta$ (29)

Note that $\varepsilon(t,\theta)$ includes θ and therefore φ depends on θ. Strictly speaking (25) is not a linear function of θ: A Pseudo Linear Regression

In system identification we want to find θ from data $Z^N = \{y(t),u(t)\,|\,1 \le t \le N\}$. This problem is now nonlinear, and there is no closed-form solution, unlike the least squares solution given by (11) for linear regressions. However, the pseudo-linear formulation allows us to solve the problem by repeatedly using the least squares solution.

This is called Extended Least Squares Algorithm (We assume $n_a \ge n_b, n_c$ for simplicity):

i). Set i = 0 and $c_1,\cdots,c_{n_c} = 0$, solve the $(n_a + n_b)$-dimensional Least Squares problem to obtain initial estimates of $A(q), B(q)$.

ii). Using these estimates obtain the initial estimate of prediction error

$$\varepsilon^{(0)}(t-1,\hat{\theta}),\cdots,\varepsilon^{(0)}(t-n_c,\hat{\theta}), \quad n_a \le t \le N. \tag{30}$$

iii). Form the pseudo-linear regression using the prediction error in the previous step:

$$\phi(t,\theta) \equiv [-y(t-1),...,-y(t-n_a),u(t-1),...,u(t-n_b) \quad \varepsilon^{(i)}(t-1,\theta),...\varepsilon^{(i)}(t-n_c,\theta)]^T \tag{31}$$

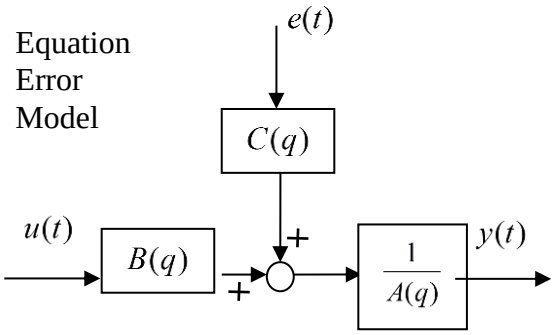
and solve the $(n_a + n_b + n_c)$-dimensional Least Squares problem. Using the updated $A(q), B(q), C(q)$, update the prediction error estimate $\varepsilon^{(i+1)}(t-n_c,\theta),\cdots,\varepsilon^{(i+1)}(t-n_c,\theta)$.

iv). Set $i = i+1$, and repeat the computation for *M* times.

13.7 Output Error Model Structure

There is another set of model structures called Output Error Model.

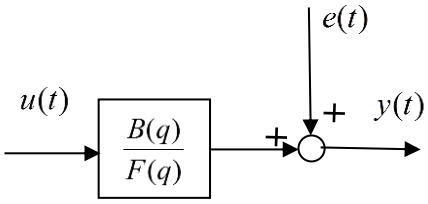
Equation Error Model



Noise enter the process.

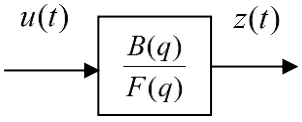
This resembles
Preocess Noise
in the Kalman Filter.

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,



$$z(t) + f_1 z(t-1),...,+f_{n_f} z(t-n_f) = b_1 u(t-1),...,b_{n_b} u(t-n_f) \quad (32)$$

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

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

$$y(t) = \frac{B(q)}{F(q)} u(t) + e(t) \quad (33)$$

The parameters to be determined are collectively given by

$$\theta = [\, b_1 \;\; b_2 \ldots b_{n_b} \;\; f_1 \;\; f_2 \ldots \;\; f_{n_f} \,]^{\mathrm{T}} \quad (34)$$

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

$$\hat{y}(t|\theta) = \frac{B(q)}{F(q)} u(t) = z(t,\theta) \quad (35)$$

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

$$\begin{aligned}\hat{y}(t|\theta) &= -f_1 z(t-1,\theta) - -f_2 z(t-2,\theta) - \cdots - f_{n_f} z(t-n_f,\theta) \\ &\quad + b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b) \\ &= \varphi^T(t,\theta)\cdot\theta\end{aligned} \tag{36}$$

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

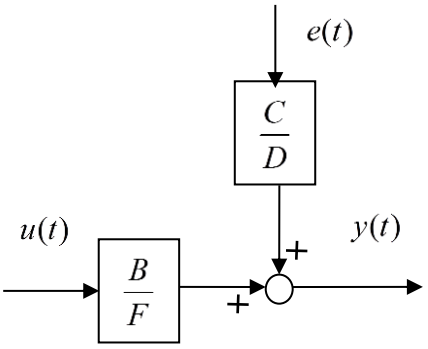
$$\varphi(t,\theta) = [u(t-1)\cdots u(t-n_b) - z(t-1,\theta)\cdots - z(t-n_f,\theta)]^T \tag{37}$$

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

$$y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t) \tag{38}$$



13.8 Consistent and Unbiased Estimation: Preview System ID Analysis

The Extended Least Squares algorithm works well for most of ARMAX systems. However, for more complex systems, including output error model, a question arises whether we can obtain the correct estimate. This section briefly describes some important issues on model structure in estimating parameters involved in the model. Details will be discussed later.

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 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 $\hat{y}(t\,|\,\theta)$.

Consider the following squared norm of prediction error:

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

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

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

Let us apply the Least Square Estimate (LSE) minimizing the mean squared error $V_N(\theta, Z^N)$:

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

where

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

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 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

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

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

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 Excite" 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, if ARMAX model is used, the linear regression cannot be used, and the output sequence involved in $\varphi(t)$ may be correlated with $v_0(t)$:

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

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

(45)

For output error models, however, the pseudo-linear regressor does not include the output $y(t)$, thereby, no correlation with the noise. The variables there are estimated noise-free state variables, $z(t)$, called "Instrumental Variables". It is known that estimation using instrumental variables provides no bias.

13.9 State Space Model

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

Stationary Time-invariant

$$x(t+1)=A(\theta)x(t)+B(\theta)u(t)+w(t) \qquad A(\theta)\in R^{n\times n} \qquad (46)$$

$$y(t)=C(\theta)x(t)+v(t) \qquad B(\theta)\in R^{n\times m} \qquad (47)$$

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

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:

$$E[w(t)w^T(t)]=R_1(\theta)$$

$$E[v(t)v^T(t)]=R_2(\theta) \qquad (48)$$

$$E[w(t)v^T(t)]=R_{12}(\theta)$$

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

Using forward shift operation q, we can rewrite (46) as

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

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

$$y(t) = C(\theta)\left[qI - A(\theta)\right]^{-1} B(\theta)u(t) + C(\theta)\left[qI - A(\theta)\right]^{-1} w(t) + v(t) \quad (49)$$

from (47)

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

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

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

$$y(t) - C(\theta)\hat{x}(t,\theta) = C(\theta)[x(t) - \hat{x}(t,\theta)] + v(t) \equiv e(t) \quad (51)$$

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

represents the part of $y(t)$ that cannot be predicted form past data. This part is called, the "innovation", denoted $e(t)$. Using this innovation, K-F is written as

$$\hat{x}(t+1,\theta) = A(\theta)\hat{x}(t,\theta) + B(\theta)u(t) + K(\theta)e(t) \quad (52)$$
$$y(t) = C(\theta)\hat{x}(t,\theta) + e(t) \quad (53)$$

The covariance of innovation $e(t)$ is

$$E[e(t)e^T(t)] = c(\theta)P(\theta)C^T(\theta) + R,(\theta) \quad (54)$$

Error covariance of state estimation

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

Combining (52) and (53), and comparing it with (2),

$$y(t) = G(q,\theta)u(t) + H(q,\theta)e(t) \tag{55}$$

$$\begin{aligned} G(q,\theta) &= C(\theta)[qI - A(\theta)]^{-1}B(\theta) \\ H(q,\theta) &= C(\theta)[qI - A(\theta)]^{-1}K(\theta) + I \end{aligned} \tag{56}$$

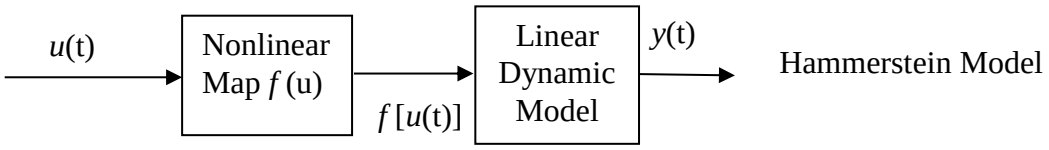
This shows the relationship between the state space model and the input-output model. They are connected through the innovation process.

$$\begin{aligned} &(qI - A)\hat{x}(t,\theta) = Bu(t) + Ke(t) \\ &\hat{x}(t,\theta) = (qI - A)^{-1}Bu(t) + (qI - A)^{-1}Ke(t) \\ &y(t) = C(qI - A)^{-1}Bu(t) + C(qI - A)^{-1}Ke(t) + e(t) \end{aligned} \tag{57}$$

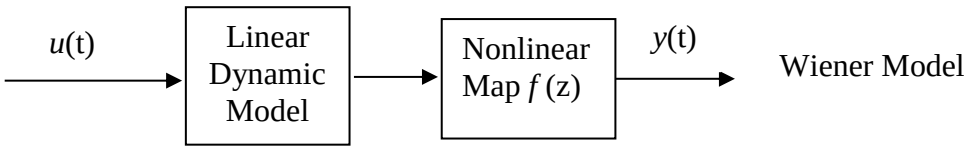
13.10 Combined Linear and Nonlinear Models

We often need to deal with nonlinear dynamical systems, to which the above linear time-invariant model structures cannot be applied. However, a class of nonlinear dynamical systems can be represented as a combination of the above LTI model and nonlinear algebraic functions discussed in Part 2. In particular, if the system can be divided into a linear dynamical subsystem and a nonlinear algebraic subsystem, the following model structures can be used:

a) Hammerstein Model



b) Wiener Model



Furthermore, if unknown parameters are linearly involved in the nonlinear algebraic maps, such as radial-basis functions with fixed center points, the predictor is given by a linear regression.