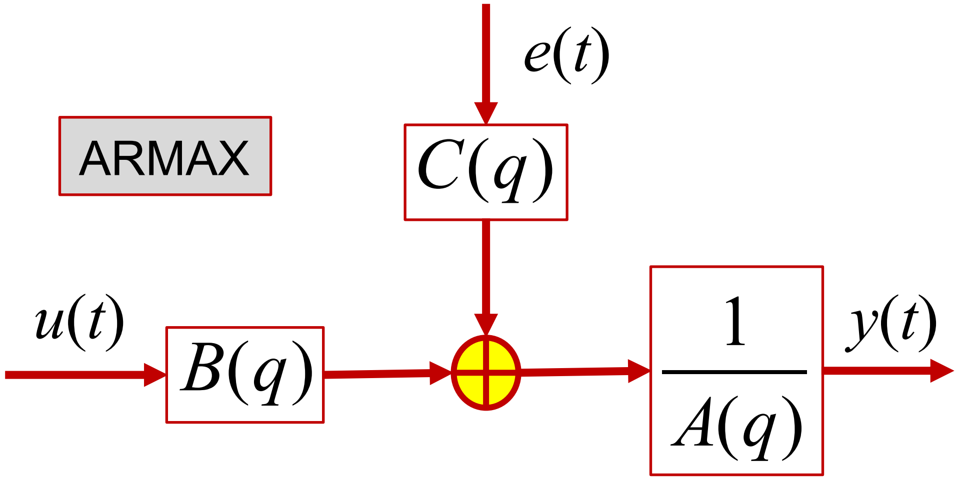
2.160 Identification, Estimation, and Learning

Part 3 Linear System Identification

Lecture 14

Parametric Linear System Identification

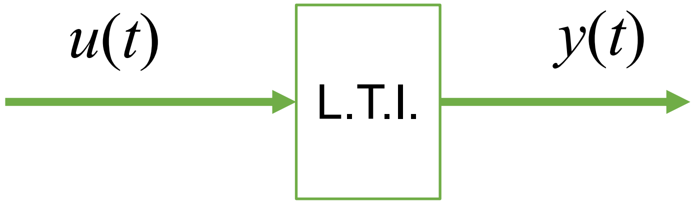


Auto-Regressive Moving Average model with eXogenous input

H. Harry Asada
Department of Mechanical Engineering
MIT

What is Parametric Model? Why do we care?

Non-parametric model,
e.g. Bode Plot, Impulse Response Model



$$y(t)=\sum_{k=0}^{\infty} g(k)u(t-k)=\underbrace{\sum_{k=0}^{\infty} g(k)q^{-k}}_{G(q)}u(t)$$

For example,

$$G(q)=1+\frac{1}{2}q^{-1}+\frac{1}{4}q^{-2}+\cdots$$

$$\frac{1}{2}q^{-1}G(q)=\underbrace{\frac{1}{2}q^{-1}+\frac{1}{4}q^{-2}+\cdots+1}_{G(q)}-1=G(q)-1$$

$$\therefore G(q)=\frac{1}{1-\frac{1}{2}q^{-1}}$$

- ❑ If we know this particular model structure, we need to identify only one parameter c.

$$G(q)=\frac{1}{1+cq^{-1}}$$

- ❑ In general, a system can be represented with a succinct model containing fewer parameters, if we know the model structure.
- ❑ For L.T.I. systems, model order, numbers of poles and zeros, etc. determine model structure.

$$G(q)=\frac{b_1q^{-1}+b_2q^{-2}+\cdots+b_{n_b}q^{-n_b}}{1+a_1q^{-1}+a_2q^{-2}+\cdots+a_{n_a}q^{-n_a}}=\frac{B(q)}{A(q)}$$

- ❑ The problem is to identify parameters

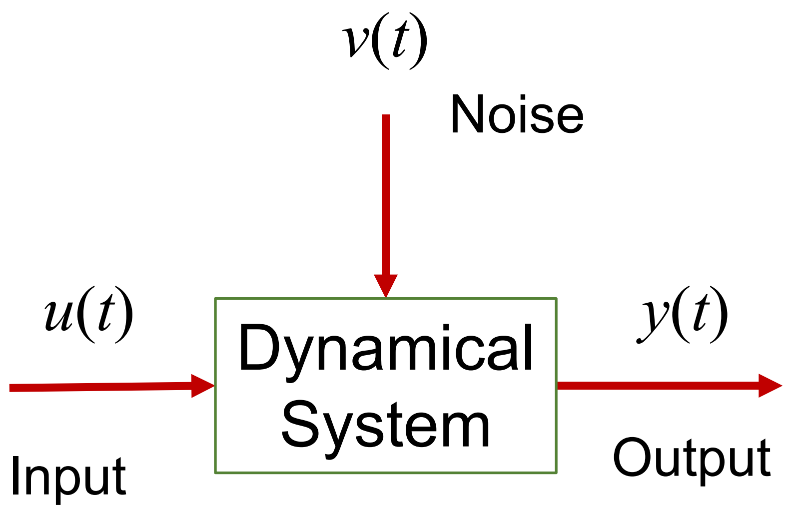
$$a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b}$$

from input-output data.

Identifying Noise Dynamics

- In control design and estimation/prediction, it is often important to identify not only the dynamics from input to output, but also the noise dynamics; how noise and disturbance perturb the system; where noise comes in; whether the noise is colored and correlated.

- In system identification, various techniques are available for identifying both input-output dynamics and noise dynamics.



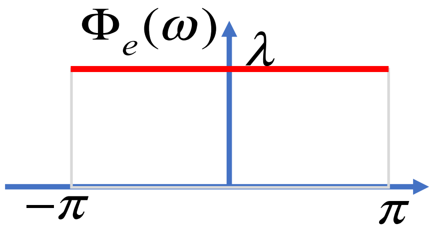
Dynamic Modeling of Noise

- The fundamental difference between input-output dynamics and noise dynamics is that the input is accessible and even manipulatable for the former, but is not for the latter.

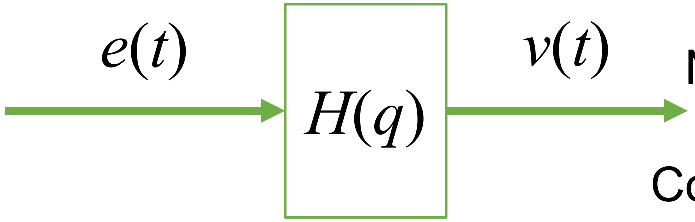
- Although not accessible, noise has statistical properties and some “structure”;

- To capture these properties, we consider the following noise dynamic model.

Noise is originated in white (totally uncorrelated) exogenous input.



Uncorrelated
“White” noise



Noise
Colored

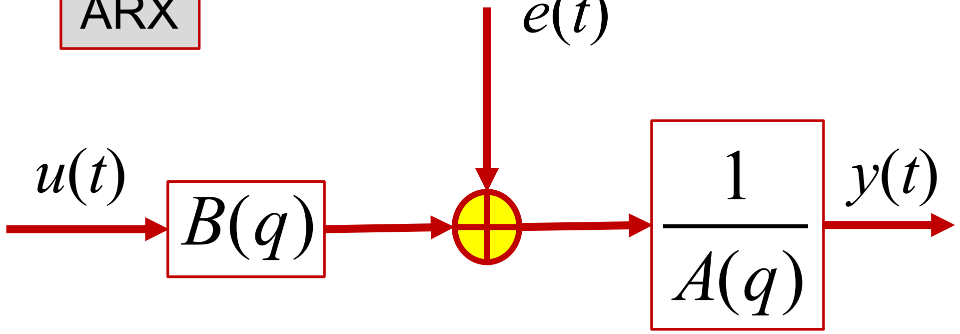
- We assume that noise is originated in a completely uncorrelated “White” stochastic process e(t) with covariance λ, and that it goes through a transfer function H(q), resulting in colored noise v(t).

- White noise e(t) is completely unpredictable, while colored noise v(t) is to some extent predictable if we know H(q). We aim to identify this noise transfer function H(q) representing the “structure” or the “channel” of the noise through which its unpredictable input e(t) is transmitted, resulting in the colored noise.

Various Model Structures of Stochastic LTI Systems

- Depending on how input and output are connected and how noise comes into the system, various model structures have been used. In the following A(q), B(q), C(q), … represent polynomials of q.

ARX

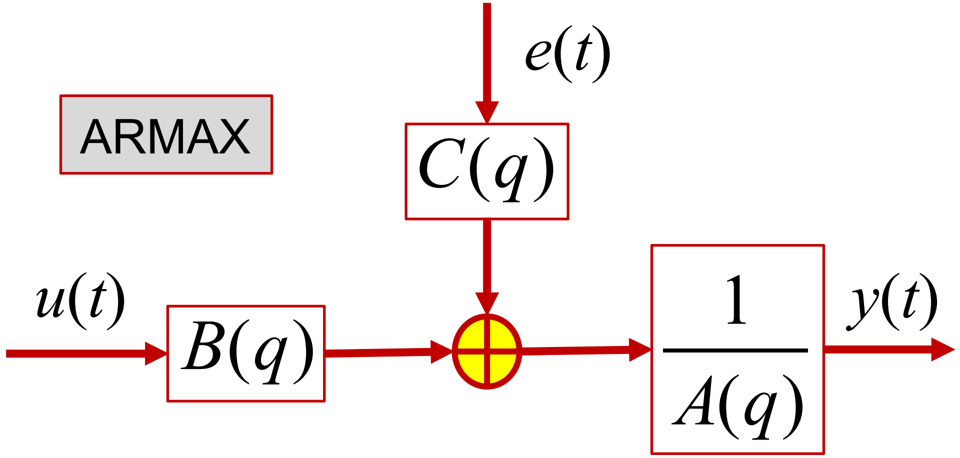


Auto-Regressive model with eXogenous input

$$A(q) = 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a},$$

$$B(q) = b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}$$

ARMAX

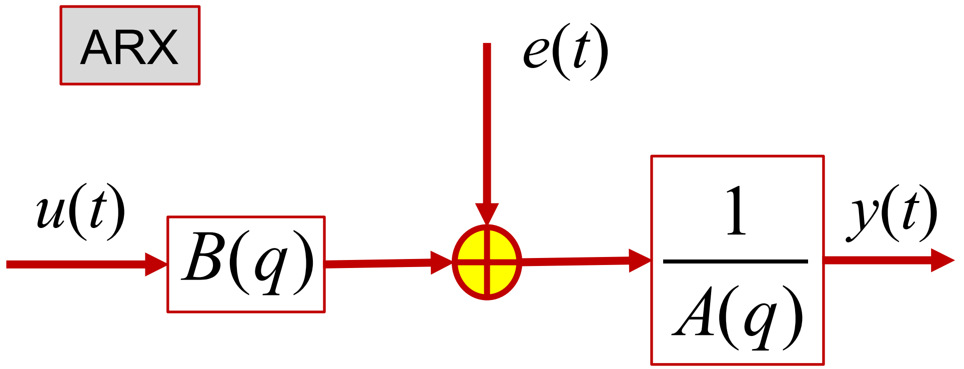


Auto-Regressive Moving Average model with eXogenous input

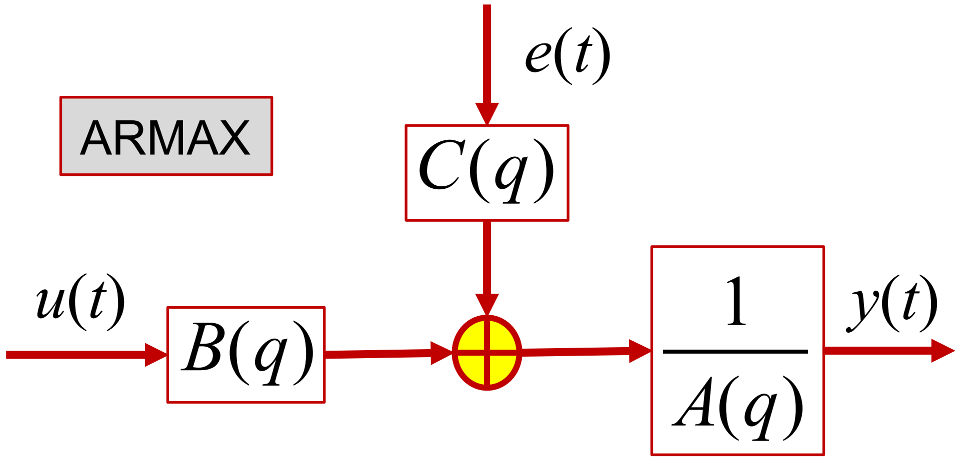
$$C(q) = 1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c}$$

Various Model Structures of Stochastic LTI Systems

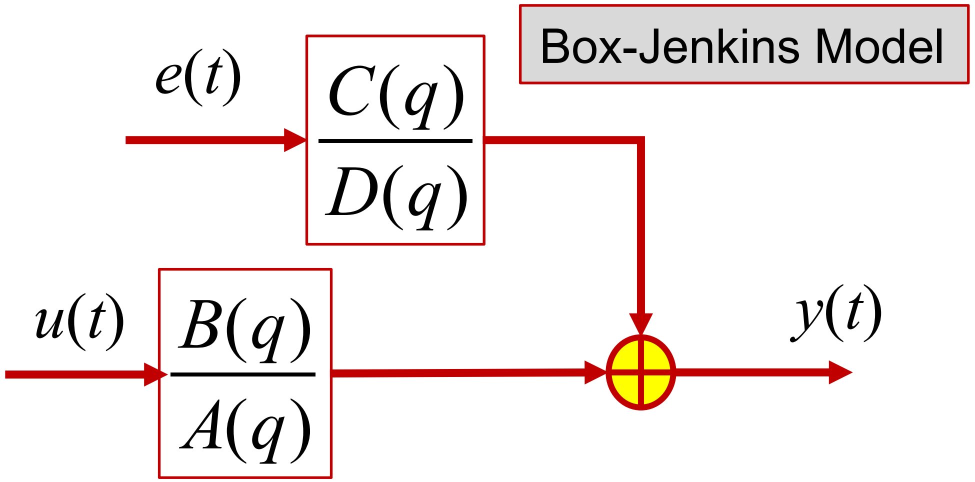
- Depending on how input and output are connected and how noise comes into the system, various model structures have been used. In the following A(q), B(q), C(q), … represent polynomials of q.

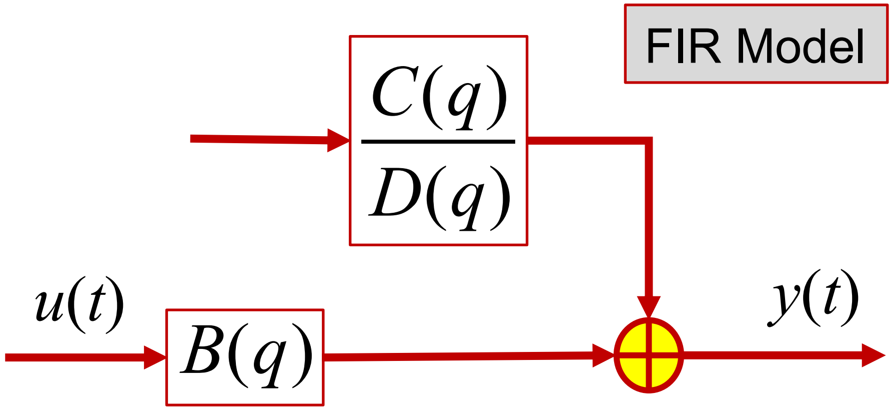


Auto-Regressive model with eXogenous input



Auto-Regressive Moving Average model with eXogenous input

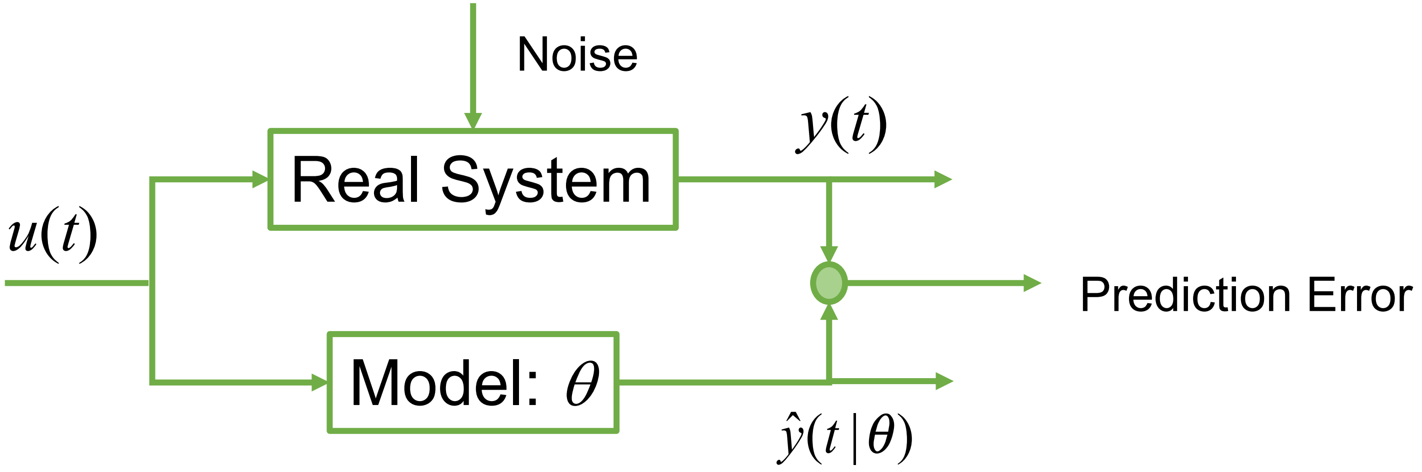




Finite Impulse Response model

Estimating Model Parameters from Input-Output Data

- How can we determine parameters $\theta = (a_1, \cdots, a_{n_a}, b_1, \cdots, b_{n_b})^T$?
- We use the Prediction Error Method first.



- Identification Data

$$D = \{(u(t), y(t)) \mid t = 1, 2, \cdots, N\}$$

- Apply Least Squares Estimate

$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum (y(t) - \hat{y}(t \mid \theta))^2$$

Auto-Regressive model with eXogenous input (ARX)

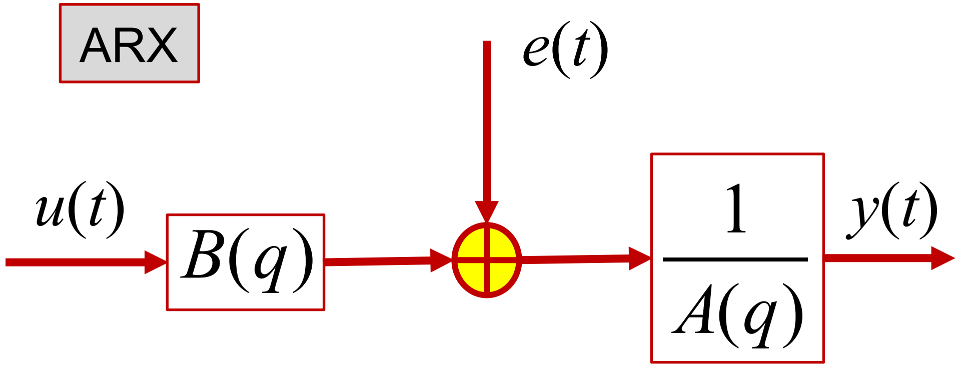
- In ARX model, White noise e(t) directly acts along the forward path of the system.
- A(q) and B(q) are polynomials given by

$$A(q) = 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a},$$

$$B(q) = b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}$$

White noise

Zero mean $E[e(t)] = 0$



- Output y(t) can be written as follows,

$$y(t) = \frac{B(q)}{A(q)} u(t) + \frac{1}{A(q)} e(t), \quad A(q) y(t) = B(q) u(t) + e(t)$$

$$y(t) = -a_1 y(t-1) - \cdots - a_{n_a} y(t-n_a) + b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b) + e(t)$$

- Therefore, the output can be predicted by using estimated parameters

$$\hat{y}(t) = -\hat{a}_1 y(t-1) - \cdots - \hat{a}_{n_a} y(t-n_a) + \hat{b}_1 u(t-1) + \cdots + \hat{b}_{n_b} u(t-n_b)$$

- Note that the unknown parameters are linearly involved in the above expression.

Derivation and Note

Example:

$$A(q)y(t) = (1 + a_1 q^{-1})y(t)$$

$$= y(t) + a_1 q^{-1} y(t) = y(t) + a_1 y(t-1)$$

Clarification:

In deterministic system identification, where e(t) is dropped, the system shown below is called ARMA (auto-regressive, moving average) model.

$$y(t) = \frac{B(q)}{A(q)}u(t) + \frac{1}{A(q)}e(t), \quad A(q)y(t) = B(q)u(t) + e(t)$$

$$y(t) = \underbrace{-a_1 y(t-1) - \cdots - a_{n_a} y(t-n_a)}_{\text{Auto-Regressive Part}} + \underbrace{b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b)}_{\text{~~Moving Average Part~~}} + e(t)$$

Auto-Regressive Part

~~Moving Average Part~~

Exogenous Input

ARX Model

Auto-Regressive model with eXogenous input (ARX)

$$\hat{y}(t) = -\hat{a}_1 y(t-1) - \cdots - \hat{a}_{n_a} y(t-n_a) + \hat{b}_1 u(t-1) + \cdots + \hat{b}_{n_b} u(t-n_b)$$

- This is a linear regression with regressor and parameter vectors defined by

$$\varphi(t) = \begin{pmatrix} -y(t-1) \\ \vdots \\ -y(t-n_a) \\ u(t-1) \\ \vdots \\ u(t-n_b) \end{pmatrix}, \quad \theta = \begin{pmatrix} a_1 \\ \vdots \\ a_{n_a} \\ b_1 \\ \vdots \\ b_{n_b} \end{pmatrix} \qquad \hat{y}(t \mid \theta) = \hat{\theta}^T \varphi(t)$$

- This is a standard least squares estimate problem. If the regressor data are persistently exciting, it has a unique solution.

$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum (y(t) - \hat{y}(t \mid \theta))^2 = \left[\sum_{t=t^*}^{N} \varphi(t)\varphi^T(t) \right]^{-1} \left[\sum_{t=t^*}^{N} \varphi(t) y(t) \right]$$

- Note that due to an edge effect, the data usable for the above calculation starts at $t^* = \max[n_a, n_b]$.

Discussion

- In the above ARX model, the output prediction was obtained simply by taking expectation of the true system:

True System: $y(t) = -a_1 y(t-1) - \cdots - a_{n_a} y(t-n_a) + b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b) + e(t)$

$$E[y(t)] = E[-a_1 y(t-1) - \cdots - a_{n_a} y(t-n_a) + b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b)] + E[e(t)]$$

$$= -\hat{a}_1 y(t-1) - \cdots - \hat{a}_{n_a} y(t-n_a) + \hat{b}_1 u(t-1) + \cdots + \hat{b}_{n_b} u(t-n_b) = \hat{\theta}^T \varphi(t)$$

- This is a special case where the white noise e(t) alone is added to the output. In general, this is not the case. For example,

$$y(t) = -a_1 y(t-1) + b_1 u(t-1) + e(t) + c_1 e(t-1)$$

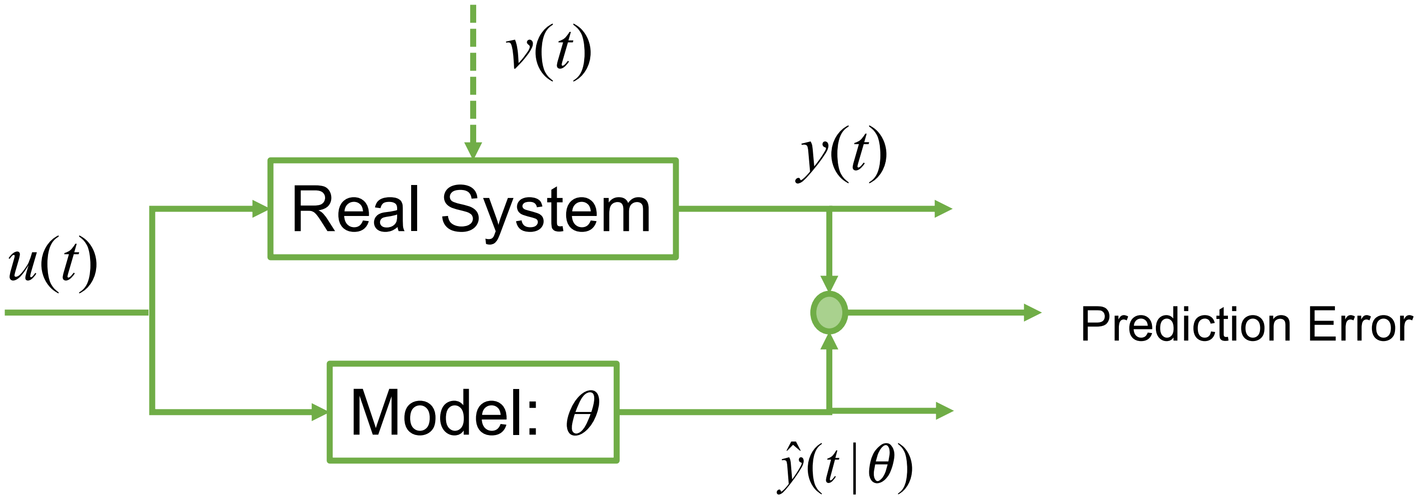
- Simply averaging out the e(t) and e(t-1) terms does not provide the best prediction:

$$\hat{y}(t \mid \theta) \neq -\hat{a}_1 y(t-1) + \hat{b}_1 u(t-1)$$

- As shown in the following, the term e(t-1) is predictable after observing y(t-1) and knowing u(t-1). In the Prediction Error Method we must make the best estimate $\hat{y}(t \mid \theta)$ to evaluate the parameters.

Let us examine the last statement.

In the Prediction Error Method we must make the best estimate $\hat{y}(t|\theta)$ to evaluate the parameters.

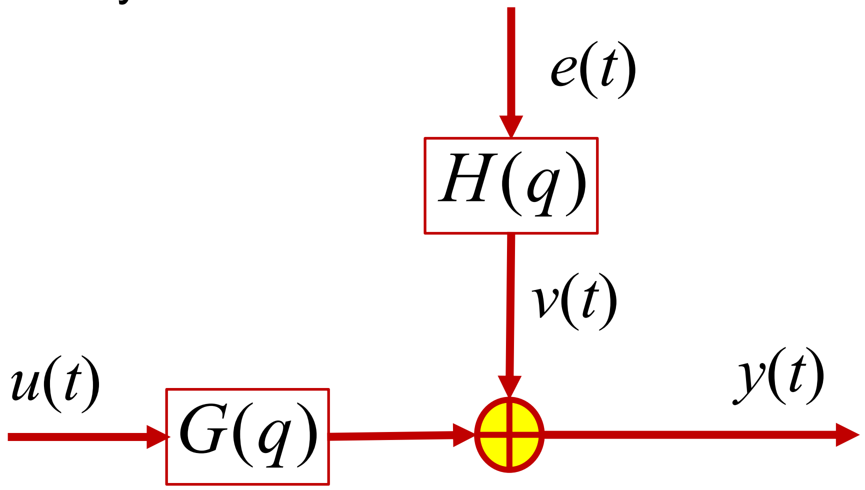


- If the prediction $\hat{y}(t|\theta)$ is not an optimal estimate for a given parameter θ, the prediction error is due to the inappropriate prediction rather than the inaccuracy of the parameter θ.
- If the prediction is the best, the error is largely due to the error of the parameter values.

- The estimate of parameters may be biased, if we ignore the noise, which is colored.

Noise Dynamics Transfer Function

- Consider a general case where input u(t) goes through a forward path transfer function G(q) and White noise e(t) through noise dynamics transfer function H(q). This model subsumes all the model structures discussed previously.



- White noise e(t) has variance λ;

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

- In the block diagram, v(t) represents colored noise:

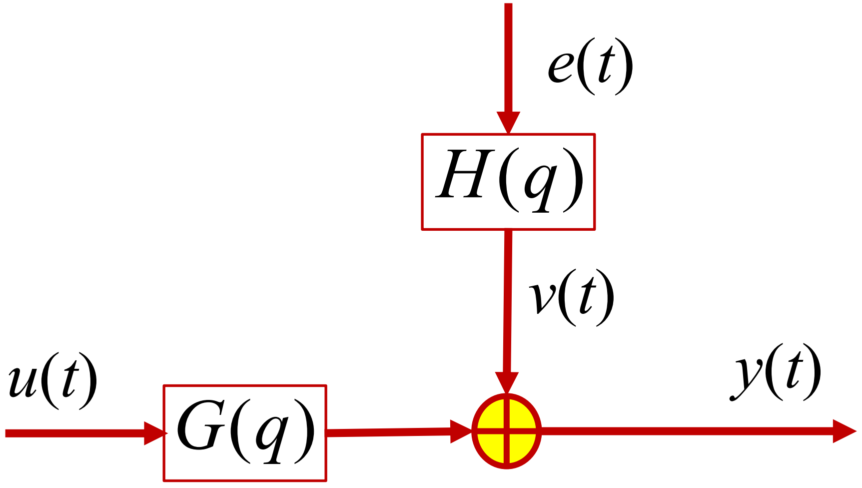
$$v(t) = H(q)e(t)$$

- Without loss of generality, we can assume that the DC gain of the transfer function H(q) is unity. Because the variance of the White noise represents the scale (magnitude) of the noise strength, we can set the DC gain of H(q) to be unity.

- Transfer function with a unit DC gain* is called monic.

* The coefficient of the highest order term is 1 in a polynomial: $H(q) = 1 + c_1 q^{-1} + c_2 q^{-2} + \cdots$

Predictor



- In an attempt to find the optimal output prediction $\hat{y}(t\,|\,\theta)$, consider to utilize the observation of prior output y(t-1) and input u(t-1),

- The output time sequence of the system is given by

$$y(t) = G(q)u(t) + H(q)e(t),$$
$$y(t-1) = G(q)u(t-1) + H(q)e(t-1),$$
$$\ldots$$

- Since we know y(t-1) and u(t-1), we can estimate e(t-1) from the second equation above.

$$e(t-1) = \frac{y(t-1) - G(q)u(t-1)}{H(q)}$$

In general, $$e(t-k) = \frac{y(t-k) - G(q)u(t-k)}{H(q)}$$

Predictor

- Let us express the noise transfer function H(q) in impulse response form:

$$H(q) = 1 + h(1)q^{-1} + h(2)q^{-2} + \cdots$$

Note that H(q) is monic.

- The colored noise v(t) is then expressed as

$$v(t) = H(q)e(t) = e(t) + h(1)e(t-1) + h(2)e(t-2) + \cdots$$

Recall that these terms can be predicted: $e(t-k) = \dfrac{y(t-k) - G(q)u(t-k)}{H(q)}$

- Define m(t-1) to be the sum of all these predictable terms in v(t),

$$m(t-1) \triangleq h(1)e(t-1) + h(2)e(t-2) + \cdots$$

$$= h(1)e(t-1) + h(2)e(t-2) + \cdots + e(t) - e(t)$$

$$= H(q)e(t) - e(t) = (H(q)-1)e(t)$$

Since H(q) is monic, this does not include e(t), but e(t-1), e(t-2),…

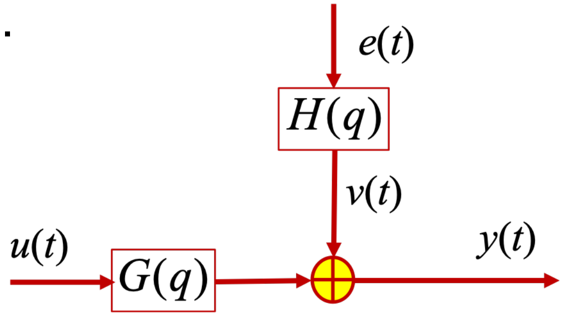
Predictor

- Estimation of the colored noise v(t):

$$\hat{v}(t) = E[v(t)] = E[e(t) + m(t-1)] = m(t-1) = (H(q)-1)e(t)$$

$$= (H(q)-1)H(q)^{-1}(y(t) - G(q)u(t)) = (1 - H(q)^{-1})(y(t) - G(q)u(t))$$

Note that $H(q)e(t) = y(t) - G(q)u(t)$ is used in the above computation.



- We need to examine whether H^{-1} is monic. Let us write the inverse of H(q) as

$$H(q)^{-1} = \tilde{h}(0) + \tilde{h}(1)q^{-1} + \tilde{h}(2)q^{-2} + \cdots$$
$$H(q)H(q)^{-1} = 1$$

$$1 \cdot \tilde{h}(0) + (\tilde{h}(1) + h(1)\tilde{h}(0))q^{-1} + (\text{heigher-order polynomial of } q^{-1}) = 1$$

$$\therefore \tilde{h}(0) = 1;\ \text{monic}$$

- The predicted colored noise is a function of all known variables.

$$\hat{v}(t) = (1 - H(q)^{-1})(y(t) - G(q)u(t)) = \hat{v}(y(t-1), y(t-2), \cdots, u(t-1), \cdots)$$

This does not include y(t), u(t), but older output and input variables.

Predictor

- Estimation of the output:

$$\hat{y}(t\,|\,t-1) = G(q)u(t) + \hat{v}(t) = G(q)u(t) + (1 - H(q)^{-1})(y(t) - G(q)u(t))$$

$$= G(q)u(t) - G(q)u(t) + H(q)^{-1}G(q)u(t) + (1 - H(q)^{-1})y(t)$$

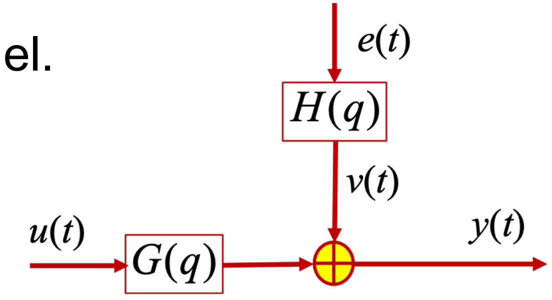
$$\therefore\ \hat{y}(t\,|\,t-1) = H(q)^{-1}G(q)u(t) + (1 - H(q)^{-1})y(t)$$

- Let us examine whether this is the *best* estimate we can make based on the model. To this end, we evaluate the prediction error:

$$y(t) - \hat{y}(t\,|\,t-1) = y(t) - H(q)^{-1}G(q)u(t) - (1 - H(q)^{-1})y(t)$$

$$= y(t) - y(t) - H(q)^{-1}G(q)u(t) + H(q)^{-1}y(t)$$

$$= H(q)^{-1}(y(t) - G(q)u(t)) = H(q)^{-1}H(q)e(t) = e(t)$$



$$y = Gu + He$$

The prediction error is nothing but the White noise, which is completely unpredictable. We have used all the predictable information in $\hat{y}(t\,|\,t-1)$.

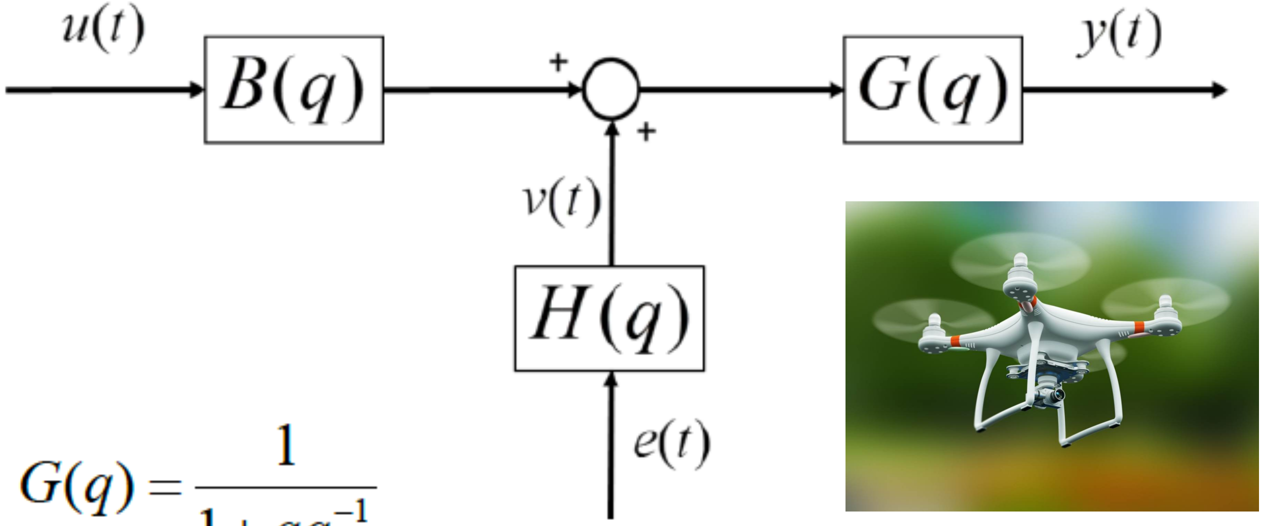
- This predictor, called One-Step Ahead Predictor, is the optimal.

$$\boxed{\hat{y}(t\,|\,t-1) = H(q)^{-1}G(q)u(t) + (1 - H(q)^{-1})y(t)}$$

- Punch line: $\hat{y}(t\,|\,t-1)$ is the optimal prediction as far as the model structure and parameters are correct. Prediction error higher than e(t) is due to the model parameters error, if the model structure is correct.

Example

An improved robotic airplane that can fly against gusty winds is being developed. The research team attempted to build a stochastic dynamic model to quantify the wind properties and use the wind model to better predict the airplane behavior. The overall system is modeled as a stochastic, linear time-invariant system with discrete-time transfer functions $G(q)$, $H(q)$, and input $u(t)$, output $y(t)$, and wind disturbance $v(t)$, as shown in the figure below. Note that $e(t)$ is white noise with unknown variance λ and that $H(q)$ is a monic, inversely-stable transfer function representing the wind dynamics.



b). In the aeronautics literature gusty winds have been modeled as a parametric model with a few characteristic parameters. For simplicity, let us write the model as

$$H(q) = 1 + cq^{-1} \qquad B(q) = bq^{-1} \qquad G(q) = \frac{1}{1 + aq^{-1}}$$

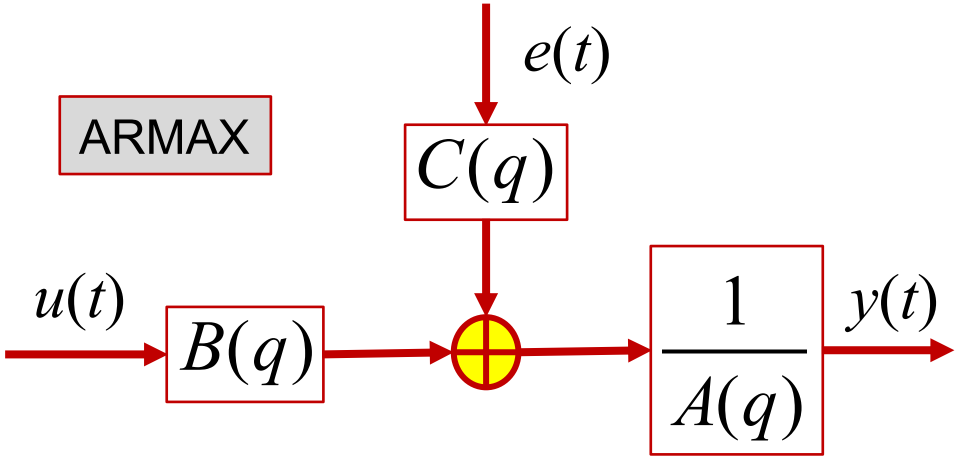
Obtain the one-step ahead predictor of the output, $\hat{y}(t \mid t-1)$, as a function of input and output variables.

Identification of ARMAX Model: Auto-Regressive Moving Average with eXogenous input

- The output of an ARMAX system is given by

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

where $C(q) = 1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c}$



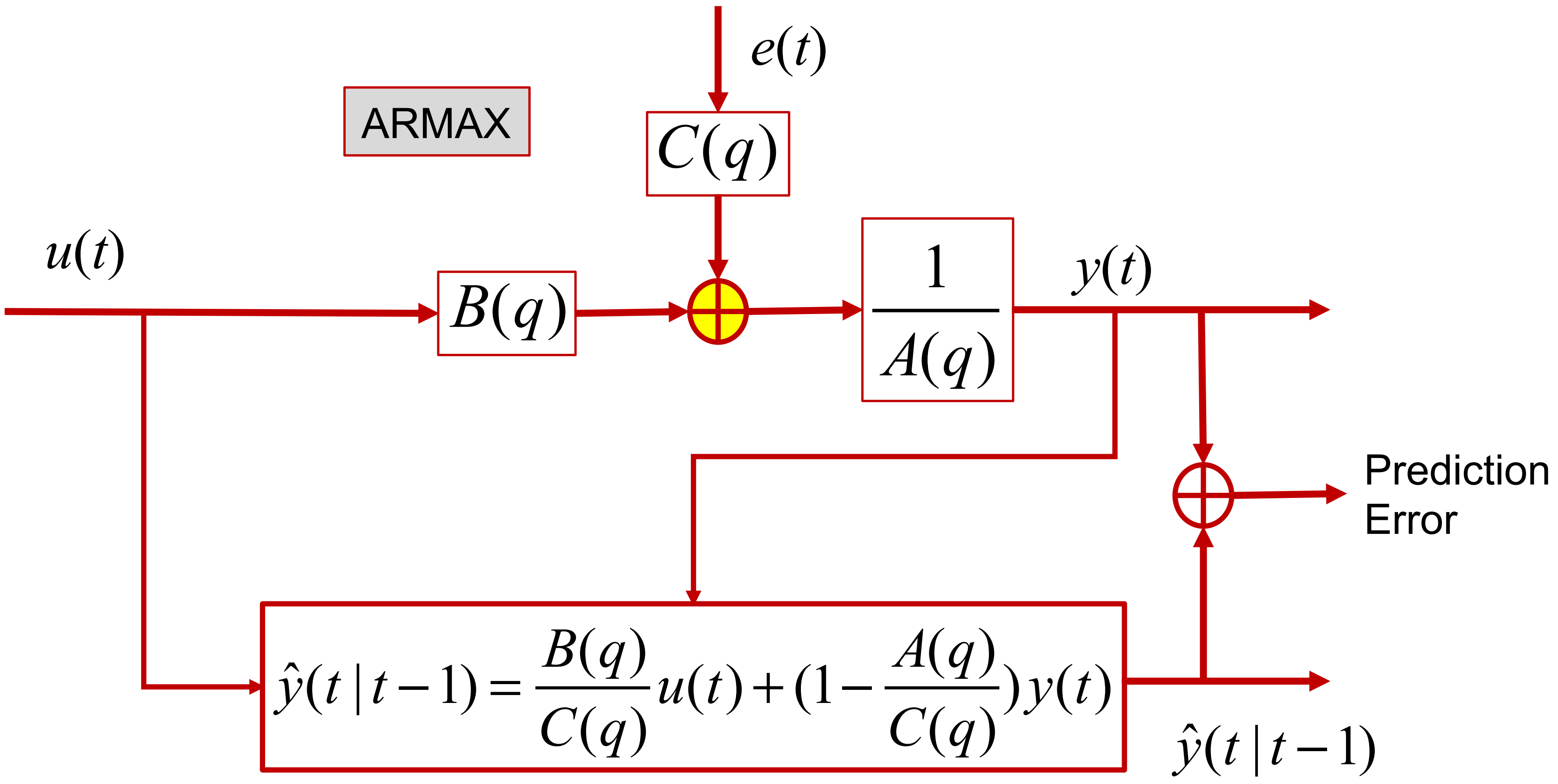
- The optimal predictor is given by

$$\hat{y}(t|t-1) = H(q)^{-1}G(q)u(t) + (1 - H(q)^{-1})y(t)$$

$$= \frac{\cancel{A(q)}}{C(q)}\frac{B(q)}{\cancel{A(q)}}u(t) + (1 - \frac{A(q)}{C(q)})y(t)$$

$$\boxed{\hat{y}(t|t-1) = \frac{B(q)}{C(q)}u(t) + (1 - \frac{A(q)}{C(q)})y(t)}$$

Prediction Error Method



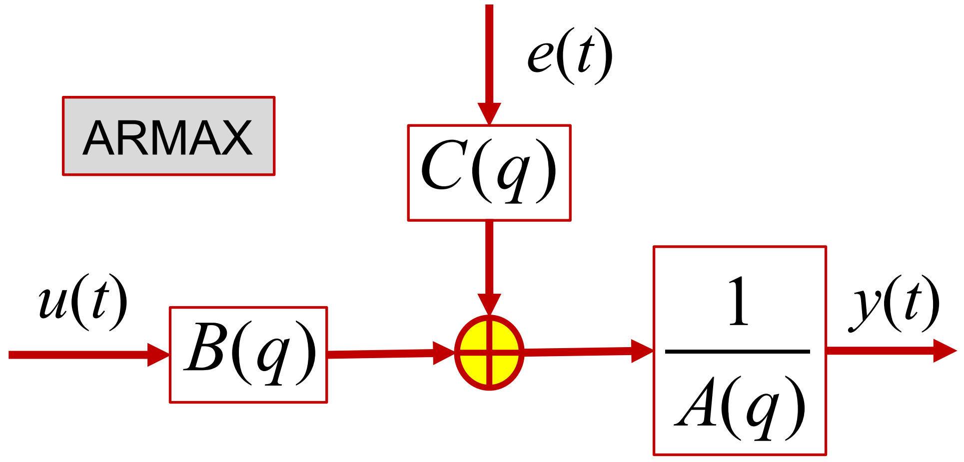
Model: Optimal Predictor

- ❑ Prediction error comes from the estimated parameters involved in polynomials A(q), B(q), and C(q).
- ❑ This allows us to estimate not only the input-output forward dynamics but also the noise dynamics.

Identification of ARMAX Model: Auto-Regressive Moving Average with eXogenous input

- To emphasize that the predictor is dependent on parameters, let us write it as $\hat{y}(t \mid \theta)$

- Multiplying C(q) to both sides,



$$\hat{y}(t \mid \theta) = \frac{B(q)}{C(q)} u(t) + \left(1 - \frac{A(q)}{C(q)}\right) y(t)$$

$$C(q)\hat{y}(t \mid \theta) = B(q)u(t) + (C(q) - A(q))y(t)$$

$$\hat{y}(t \mid \theta) = (1 - C(q))\hat{y}(t \mid \theta) + B(q)u(t) + (C(q) - A(q))y(t)$$

In this term, parameters are not linearly involved, since $\hat{y}(t \mid \theta)$ is dependent on the parameters.

Identification of ARMAX Model: Auto-Regressive Moving Average with eXogenous input

- Parameters are not linearly involved in the predictor $\hat{y}(t\,|\,\theta)$, therefore it cannot be written in linear regression form.

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

- Instead, it can be written as a Pseudo-Linear Regression.

$$\begin{aligned}\hat{y}(t\,|\,\theta) &= (1-C(q))\hat{y}(t\,|\,\theta) + B(q)u(t) + (C(q)-1)y(t) + (1-A(q))y(t)\\ &= (1-A(q))y(t) + B(q)u(t) + (C(q)-1)(y(t)-\hat{y}(t\,|\,\theta))\end{aligned}$$

- Replacing the last term by $\varepsilon(t\,|\,\theta) \triangleq y(t) - \hat{y}(t\,|\,\theta)$ we can write:

$$\begin{aligned}\hat{y}(t\,|\,\theta) = &-a_1 y(t-1) - \cdots - a_{n_a} y(t-n_a) + b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b)\\ &+ c_1 \varepsilon(t-1\,|\,\theta) + \cdots + c_{n_c} \varepsilon(t-n_c\,|\,\theta)\end{aligned}$$

- Defining parameter and regressor vectors as:

$$\theta = [a_1, \cdots, a_{n_a}, b_1, \cdots, b_{n_b}, c_1, \cdots, c_{n_c}]^T$$

$$\varphi(t\,|\,\theta) = [-y(t-1), -\cdots, -y(t-n_a), u(t-1), \cdots, u(t-n_b), \varepsilon(t-1\,|\,\theta), \cdots, \varepsilon(t-n_c\,|\,\theta)]^T$$

We can obtain the predictor in Pseudo-Linear Regression form.

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

Extended Least Squares Estimate

- With the pseudo-linear regression the Prediction Error Method becomes a nonlinear optimization problem, where the standard Least Squares Estimate cannot be applied.

$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum (y(t) - \hat{y}(t\,|\,\theta))^2 \quad \text{where} \quad \hat{y}(t\,|\,\theta) = \theta^T \varphi(t\,|\,\theta)$$

- The following Extended Least Squares Estimate can be used for obtaining the parameters that minimize the squared error index.

Step 1 Initially assuming that C(q) = 1, solve the ARX problem with the same order of A(q) and B(q). This will give an initial estimate $\bar{\theta} = (\bar{a}_1, \cdots, \bar{a}_{n_a}, \bar{b}_1, \cdots, \bar{b}_{n_b})^T \in \Re^{(n_a+n_b)\times 1}$

Step 2 Set $\theta^{(0)} = (\bar{\theta}^T, 0, \cdots, 0)^T \in \Re^{(n_a+n_b+n_c)\times 1}$, and compute: $\varepsilon(t-1\,|\,\theta^{(0)}), \cdots, \varepsilon(t-n_c\,|\,\theta^{(0)})$

Step 3 Repeat the following loop:

For i = 1 to M

Construct $\varphi(t\,|\,\theta^{(i-1)}), \quad t = 1, \cdots, N$

Compute LSE $$\hat{\theta}^{(i)} = \left[\sum_{t=t^*}^{N} \varphi(t\,|\,\hat{\theta}^{(i-1)}) \varphi^T(t\,|\,\hat{\theta}^{(i-1)}) \right]^{-1} \left[\sum_{t=t^*}^{N} \varphi(t\,|\,\hat{\theta}^{(i-1)}) y(t) \right]$$

Compute $\varepsilon(t-1\,|\,\theta^{(i)}), \cdots, \varepsilon(t-n_c\,|\,\theta^{(i)})$

end

- Unlike LSE, this is not a convex optimization problem. It may face a local minima problem.

Reflection

- ❑ Parametric linear system identification;
- ❑ Model structure; ARX, ARMAX, etc.
- ❑ ARX → Linear regression
- ❑ Noise dynamics: $v(t) = H(q)e(t)$
- ❑ Predictor $\hat{y}(t|t-1) = H(q)^{-1}G(q)u(t) + (1 - H(q)^{-1})y(t)$
- ❑ ARMAX → Pseudo-linear regression
- ❑ Extended Least Squares Estimate