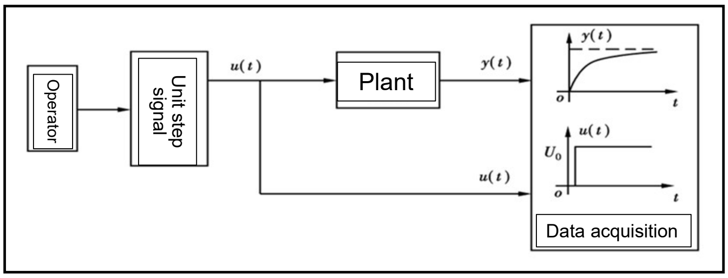
Chapter 3

System Identification Methods

3.1 Step Response Method of System Identification

- **Step response method**
  - ✓ Step response of a system measured experimentally
  - ✓ Calculate the transfer function of the system from the step response
  - ✓ Approximate method for calculating parameters of the system

♦ **Experimental schematic diagram of measuring the system step response**



♦ **Transient response of a first-order system**

A general first-order system

$$G(s) = \frac{Y(s)}{U(s)} = \frac{K}{Ts+1}$$

where K and T are constants

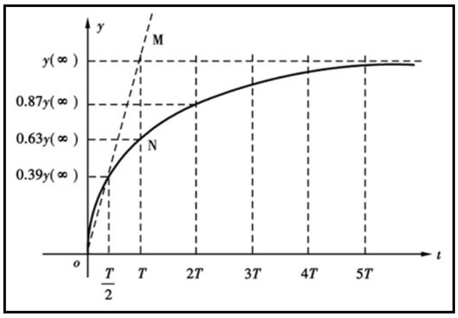
The response of this system to a unit step signal

$$Y(s) = \frac{K}{Ts+1}\,U(s) = \frac{K}{Ts+1}\frac{1}{s} = \frac{KT^{-1}}{s(s+T^{-1})} = KT^{-1}\frac{1}{s(s+T^{-1})}$$

$$y(t) = L^{-1}[Y(s)] = KT^{-1}L^{-1}\left[\frac{1}{s(s+T^{-1})}\right]$$

$$= KT^{-1}\frac{1}{T^{-1}}\left(1-e^{-T^{-1}t}\right) = K\left(1-e^{-\frac{t}{T}}\right)$$

♦ **Step response curve of the first-order system**



$$G(s) = \frac{K}{Ts+1}$$

$$y(t) = K(1 - e^{-t/T})$$

$$\left.\frac{dy}{dt}\right|_{t=0} = \frac{K}{T}e^{-t/T} = K/T$$

Identification algorithm

1) Read the steady-state value. That is the gain $K = y(\infty)$.

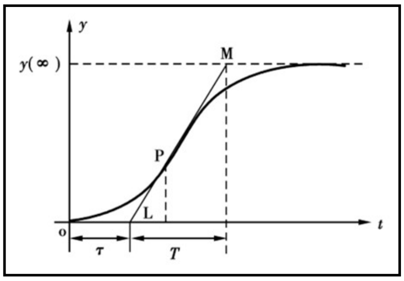
2) Determine the time where the output reaches 0.632 of its steady state. That is the time constant T.

♦ Approximate method for calculating parameters of a first-order system with delay

The first-order system with time delay: $G(s) = \frac{Ke^{-\tau s}}{Ts+1}$, where τ is the time delay.

The response of this system to a unit step signal: $y(t) = K - Ke^{-\frac{t-\tau}{T}}$

Identification algorithm:

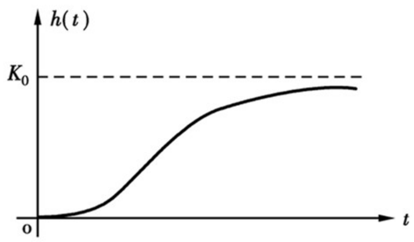


1) Read the steady-state value to obtain $K = y(\infty)$.
2) Choose two points on the unit step response:

$$y(t_1) = K - Ke^{-\frac{t_1-\tau}{T}}$$
$$y(t_2) = K - Ke^{-\frac{t_2-\tau}{T}}$$

3) Solve the above equations to get T and τ.

♦ Step response of a high-order system



Use Laplace transform method to find the transfer function of the identified system

$$G(s) = \frac{Y(s)}{U(s)} = \frac{b_m s^m + b_{m-1}s^{m-1} + \cdots + b_1 s + b_0}{a_n s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + 1}$$

Let the unit step response of the system $G(s)$ be $y(t)$.

$$K_0 = \lim_{t\to\infty} y(t) = b_0$$

Construct the unit step response of a new system $G_1(s)$ be $y_1(t)$.

$$y_1(t) = \int_0^t \big(K_0 - y(\tau)\big) d\tau$$

$$L[y_1(t)] = \frac{1}{s^2}\big(K_0 - G(s)\big) = \frac{1}{s} G_1(s) \quad \Longrightarrow \quad G_1(s) = \frac{1}{s}(K_0 - G(s))$$

$$K_1 = \lim_{t\to\infty} y_1(t) = \lim_{s\to 0} G_1(s) = K_0 a_1 \text{ - } b_1$$

where

$$\boxed{G_1(s) = \frac{1}{s}(K_0 - G(s)) = \frac{K_0 a_n s^{n-1} + \cdots + (K_0 a_m - b_m) s^{m-1} + \cdots + (K_0 a_1 - b_1)}{a_n s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + 1}}$$

Construct the unit step response of a new system $G_2(s)$ be $y_2(t)$.

$$y_2(t) = \int_0^t \big(K_1 - y_1(\tau)\big) d\tau$$

$$L[y_2(t)] = \frac{1}{s^2}\big(K_1 - G_1(s)\big) = \frac{1}{s} G_2(s) \quad \Longrightarrow \quad G_2(s) = \frac{1}{s}(K_1 - G_1(s))$$

$$K_2 = \lim_{t\to\infty} y_2(t) = \lim_{s\to 0} G_2(s) = K_1 a_1 - K_0 a_2 + b_2$$

where

$$\boxed{G_2(s) = \frac{1}{s}(K_1 - G_1(s)) = \frac{K_1 a_n s^{n-1} + \cdots + (K_1 a_m - K_0 a_{m+1} + b_{m+1}) s^{m-1} + \cdots + (K_1 a_1 - K_0 a_2 + b_2)}{a_n s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + 1}}$$

Similarly, the unit step response $y_r(t)$ of a new system $G_r(s)$ can be constructed. Following the above procedure leads to

$$y_r(t) = \int_0^t \left(K_{r-1} - y_{r-1}(\tau)\right) d\tau$$

$$L[y_r(t)] = \frac{1}{s^2}\left(K_{r-1} - G_{r-1}(s)\right) = \frac{1}{s} G_r(s) \implies G_r(s) = \frac{1}{s}(K_{r-1} - G_{r-1}(s))$$

$$K_r = \lim_{t\to\infty} y_r(t) = \lim_{s\to 0} G_r(s) = K_{r-1}a_1 - K_{r-2}a_2 + \cdots + (-1)^{r-1}K_0 a_r + (-1)^r b_r$$

where

$$\boxed{G_r(s) = \frac{1}{s}(K_{r-1} - G_{r-1}(s)) = \frac{K_{r-1}a_n s^{n-1} + \cdots + (K_{r-1}a_1 - K_{r-2}a_2 + \cdots + (-1)^{r-1}K_0 a_r + (-1)^r b_r)}{a_n s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + 1}}$$

Therefore, all the parameters of the system $G(s)$ can be estimated by the equations below.

$$K_0 = b_0$$

$$K_1 = K_0 a_1 \text{-} \, b_1$$

$$K_2 = K_1 a_1 - K_0 a_2 + b_2$$

$$\vdots$$

$$K_r = K_{r-1}a_1 - K_{r-2}a_2 + \cdots + (-1)^{r-1}K_0 a_r (-1)^r b_r$$

3.2 Maximum Likelihood Parameter Identification Method

- Ronald Aylmer Fisher (1890 ~ 1962)

  British experimental geneticist and statistician, born in London in 1890, graduated from the Department of Mathematics of Cambridge University in 1912, and became a professor of genetics at Cambridge University in 1943.

  He regarded asymptotic consistency, asymptotic validity, etc. as the basic properties that parameter estimators should have, and proposed the maximum likelihood method in 1912.

In essence, the maximum likelihood method is applied to the parameter identification of stochastic systems.

According to the fact that observation data generally have random statistical characteristics, the maximum likelihood method

– introduces the conditional probability density of random variables (observation output) or conditional probability distribution $p(y|\theta)$,

– construct a likelihood function $L(Y_N|\theta)$ with observed data and unknown parameters as independent variables, and

– obtain the parameter estimates $\hat{\theta}_{ML}$ of the system model by maximizing the likelihood function

♦ **Maximum likelihood method**

Example: Estimate the expected value of a normal stochastic process using the maximum likelihood parameter identification method.

$$p(y) = \frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2\sigma^2}(y-\mu)^2]$$

The sequence of observations for a random variable y is $Y_N = [y(1), y(2), \cdots, y(N)]^T$

The joint probability density function of Y_N is:

$$p(Y_N) = \frac{1}{(\sqrt{2\pi})^N\sqrt{|\Sigma_y|}}\exp[-\frac{1}{2}(Y - E[Y_N])^T{\Sigma_y}^{-1}(Y_N - E[Y_N])]$$

where $\sum_y$ is the covariance matrix of Y_N.

The components of Y_N are independently and identically distributed.
Then

$$\Sigma_y = \sigma^2 I, \quad |\Sigma_y| = (\sigma^2)^N$$

13

$$\Longrightarrow \quad p(Y_N) = \frac{1}{(\sqrt{2\pi})^N\sigma^N}\exp\left[-\frac{1}{2}\sum_{i=1}^{N}\frac{[y_i-\mu]^2}{\sigma^2}\right]$$

Define θ ------- Parameter μ to be estimated

$\hat{\theta}$ ------- Estimated value of the parameter μ

$$\Longrightarrow \quad p(Y_N|\theta) = \frac{1}{(\sqrt{2\pi})^N\sigma^N}\exp\left[-\frac{1}{2}\sum_{i=1}^{N}\frac{[y_i-\theta]^2}{\sigma^2}\right]$$

For a certain set of observation sequences $Y_N = [y(1), y(2), \dots, y(N)]^T$

$L(Y_N|\theta)$ -------- Likelihood function $L(Y_N|\theta)$=$p(Y_N|\theta)$

$\hat{\theta}_{ML}$ --------- Maximize the likelihood function $L(Y_N|\theta)|_{\theta=\hat{\theta}_{ML}} = Max(p(Y_N|\theta))$

14

$$\left.\frac{\partial L(Y_N|\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}_{ML}} = 0$$

or $\Rightarrow$ $\hat{\theta}_{ML}$

$$\left.\frac{\partial \ln L\,(Y_N|\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}_{ML}} = 0$$

$$\ln L\,(Y_N|\theta) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\frac{[y_i - \boxed{\theta}]^2}{\sigma^2}$$

$$\Rightarrow \frac{d}{d\theta}\ln L\,(Y_N|\theta) = \frac{1}{2\sigma^2}\sum_{i=1}^{N}2(y_i-\theta) = \frac{1}{\sigma^2}\left[\left(\sum_{i=1}^{n}y_i\right) - N\theta\right]$$

$$\Rightarrow \hat{\theta}_{ML} = \frac{1}{N}\sum_{i=1}^{n}y_i$$

15

Summary:

y: Random variable (observed output)

$p(y|\theta)$: Probability density function under the condition of parameter θ

$Y_N = [y(1), y(2), \cdots, y(N)]^T$: observation sequence

$p(Y_N|\theta)$: the joint probability density of Y_N

$\Downarrow$ For a certain set of Y_N,

$L(Y_N|\theta)$

$\Downarrow$ monotonicity of logarithmic functions

$\ln L\,(Y_N|\theta)$

$$\left.\frac{\partial \ln L\,(Y_N|\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}_{ML}} = 0$$

Log likelihood equation,

? $\Downarrow$

$\hat{\theta}_{ML}$

16

♦ **Physical meaning of the maximum likelihood identification method**

According to a set of determined random sequences Y_N, try to find parameter estimates $\hat{\theta}_{ML}$, which make the probability density function of the random variable y under the condition $\hat{\theta}_{ML}$ most likely approach the one of the random variable y under condition θ (true value) , namely:

$$p\left(y|\hat{\theta}_{ML}\right) \xrightarrow{\max} p(y|\theta)$$

The maximum likelihood method is a method of estimating unknown parameters based on the observation of the observed quantity y (output), which necessarily requires prior knowledge of the joint probability density function $p(Y_N|\theta)$ of the observable variable y (involving the determination of the likelihood function)

17

♦ **Determination of Likelihood Function $\mathbf{L}(\boldsymbol{Y_L}|\boldsymbol{\theta})$**

1) Independent observation

Each observation consists of independent samples of the random variable y, and the observations $y_1, y_2, \cdots, y_N$ are independent.

$$L(Y_N|\theta) = p(y_1|\theta)p(y_2|\theta)\cdots p(y_N|\theta) = \prod_{i=1}^{N} p(y_i|\theta)$$

18

2) Sequential observation

Sequential observations mean that $y_1, y_2, \cdots, y_N$ are not independent each other.

$$L(Y_N|\theta) = p(y_1, y_2, \cdots, y_N|\theta) = p(Y_N|\theta)$$

Conditional probability multiplication rule

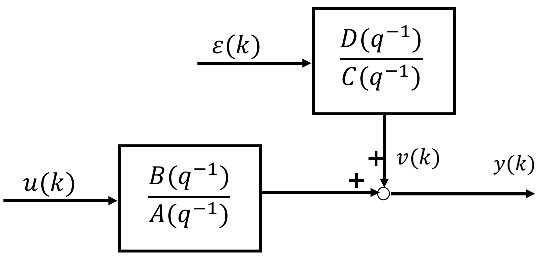
$$p(Y_N|\theta) = p(y_N|Y_{N-1}, \theta) \cdot p(Y_{N-1}|\theta)$$

$$\boxed{\begin{aligned} L(Y_N|\theta) &= p(y_N|Y_{N-1}, \theta) \cdot p(y_{N-1}|Y_{N-2}, \theta) \cdots p(y_2|y_1, \theta) \cdot p(y_1|\theta) \\ &= \prod_{k=1}^{N} p(y_k|Y_{k-1}, \theta) \end{aligned}}$$

■ Maximum likelihood estimation of parameters in dynamic system models

Dynamic system model

$$\begin{cases} A(q^{-1})y(k) = B(q^{-1})u(k) + v(k) \\ v(k) = D(q^{-1})\varepsilon(k) \end{cases}$$



where

$$\begin{cases} A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_n q^{-n} \\ B(q^{-1}) = b_1 q^{-1} + b_2 q^{-2} + \cdots + b_n q^{-n} \\ D(q^{-1}) = 1 + d_1 q^{-1} + d_2 q^{-2} + \cdots + d_n q^{-n} \end{cases}$$

$$C(q^{-1}) = A(q^{-1})$$

Dynamic system difference equation:

$$y(k) = -\sum_{i=1}^{n} a_i y(k-i) + \sum_{i=1}^{n} b_i u(k-i) + v(k)$$

The noise sequence $v(k)$ is divided into two cases:

Gaussian white noise

$$E\{v(k)\} = 0$$

$$E\{v(k)v(j)\} = \begin{cases} \sigma^2 & k = j \\ 0 & k \neq j \end{cases}$$

Colored noise

$$v(k) = \varepsilon_k + \sum_{i=1}^{n} d_i \varepsilon_{k-i}$$

$\{\varepsilon_k\}$ is a white noise

Gaussian white noise $\{v_k\}$

$$y_k = -\sum_{i=1}^{n} a_i y_{k-i} + \sum_{i=1}^{n} b_i u_{k-i} + v_k = \psi_k^T \theta + v_k$$

where

$$\psi_k = [-y_{k-1} \quad \cdots \quad -y_{k-n} \quad u_{k-1} \quad \cdots \quad u_{k-n}]^T$$

$$\theta = [a_1 \quad \cdots \quad a_n \quad b_1 \quad \cdots \quad b_n]^T$$

Let

$$Y_N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \qquad \Psi_N = \begin{bmatrix} \psi^T(1) \\ \vdots \\ \psi^T(N) \end{bmatrix} \qquad V_N = \begin{bmatrix} v(1) \\ \vdots \\ v(N) \end{bmatrix}$$

Ψ_N and θ are deterministic quantities, Y_N and V_N have the same random properties.

$$\boxed{V_N = Y_N - \Psi_N \theta} \Longrightarrow Y_N \sim N(\Psi_N \theta, \Sigma_v)$$

The covariance matrix of the noise sequence $\{v_k\}$ is written as

$$\sum_v = E\{V_N V_N^T\} = \begin{bmatrix} E\{v(1)v(1)\} & E\{v(1)v(2)\} & \cdots & E\{v(1)v(N)\} \\ E\{v(2)v(1)\} & E\{v(2)v(2)\} & \cdots & E\{v(2)v(N)\} \\ & \cdots\cdots & & \\ E\{v(N)v(1)\} & E\{v(N)v(2)\} & \cdots & E\{v(N)v(N)\} \end{bmatrix}$$

23

$$Y_N \sim N(\Psi_N \theta, \Sigma_v)$$

Joint probability distribution of N-dimensional random vector Y_N

$$p(Y_N|\theta) = (2\pi)^{-\frac{N}{2}}(\det\textstyle\sum_v)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}[Y_N - \Psi_N\theta]^T{\textstyle\sum_v}^{-1}[Y_N - \Psi_N\theta]\right\}$$

The corresponding log-likelihood function

$$\ln L\,(Y_N|\theta) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln(\det\textstyle\sum_v) - \frac{1}{2}[Y_N - \Psi_N\theta]^T{\textstyle\sum_v}^{-1}[Y_N - \Psi_N\theta]$$

$$\Rightarrow \quad \left.\frac{\partial \ln L\,[Y_N|\theta]}{\partial\theta}\right|_{\theta=\hat{\theta}_{ML}} = \left[Y_N - \Psi_N\hat{\theta}_{ML}\right]^T\textstyle\sum_v^{-1}\Psi_N = 0$$

24

$$\hat{\theta}_{ML} = [{\Psi_N}^T \textstyle\sum_v^{-1} \Psi_N]^{-1} {\Psi_N}^T \textstyle\sum_v^{-1} Y_N$$

In conclusion:

$$\textstyle\sum_v = \sigma^2 I \qquad \hat{\theta}_{ML} = [{\Psi_N}^T \Psi_N]^{-1} {\Psi_N}^T Y_N$$

The maximum likelihood estimation method is equivalent to least squares estimation.

But, the noise variance estimation of the two identification methods is slightly different.

25

$$\ln L\,(Y_N|\theta,\sigma^2) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}[Y_N - \Psi_N\theta]^T[Y_N - \Psi_N\theta]$$

Partial derivative

$$\Longrightarrow \left.\frac{\partial \ln L\,[Y_N|\theta,\sigma^2]}{\partial\sigma^2}\right|_{\substack{\theta=\hat{\theta}_{ML}\\ \sigma^2=\hat{\sigma}_{ML}^2}} = \frac{-N}{2\hat{\sigma}_{ML}^2} + \frac{[Y_N - \Psi_N\hat{\theta}_{ML}]^T[Y_N - \Psi_N\hat{\theta}_{ML}]}{2\hat{\sigma}_{ML}^4} = 0$$

$$\Downarrow$$

$$\hat{\sigma}_{ML}^2 = \frac{[Y_N - \Psi_N\hat{\theta}_{ML}]^T[Y_N - \Psi_N\hat{\theta}_{ML}]}{N}$$

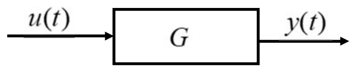
The noise variance of the least squares estimate:

$$\hat{\sigma}_{LS}^2 = \frac{[Y_N - \Psi_N\hat{\theta}_{LS}]^T[Y_N - \Psi_N\hat{\theta}_{LS}]}{N - \dim\theta}$$

26

3.3 Correlation analysis method of system identification

■ Impulse response method



The impulse response method is to use the input and output information of a linear system to identify the mathematical model of the system through the impulse response.

Although this method is simple and practical, it has a certain scope of application (systems with high signal-to-noise ratio).

It is not only an identification method of a non-parametric model (impulse response), but also an identification method of a parametric model (transfer function) obtained from an impulse response.

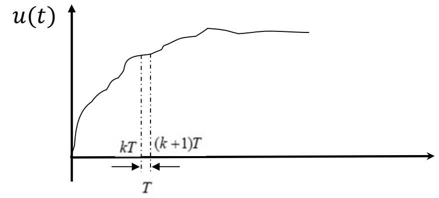
- ✓ From the input $u(t)$ and output $y(t)$, find the impulse response $g(t)$
- ✓ Find the impulse transfer function $G(q^{-1})$ from the impulse response $g(k)$.

♦ From input $\boldsymbol{u(t)}$ and output $\boldsymbol{y(t)}$, find the impulse response $\boldsymbol{g(t)}$

The convolution integral of a linear time invariant system

$$y(t) = \int_0^t g(\tau)u(t-\tau)d\tau$$

Assume that $u(t)$ and $y(t)$ are periodically sampled by a sampler with a sampling period of T, and set T small enough, $u(t)$ and g(t) can be replaced with a staircase signal by a piecewise constant approximation



$$\begin{aligned} u(t) &= u(kT) \\ y(t) &= y(kT) \quad kT \le t < (k+1)T \\ g(t) &= g(kT) \quad k = 0,1,2,\cdots \end{aligned}$$

$$\Longrightarrow \quad y(1T) = \int_0^T g(\tau)u(T-\tau)\,d\tau = Tu(0)g(0)$$

$$y(2T) = \int_0^{2T} g(\tau)u(2T-\tau)\,d\tau = \int_0^{T} g(\tau)u(2T-\tau)\,d\tau + \int_T^{2T} g(\tau)u(2T-\tau)\,d\tau$$
$$= T[g(0)u(T) + g(T)u(0)]$$

$$y(3T) = \int_0^{3T} g(\tau)u(3T-\tau)\,d\tau$$
$$= \int_0^{T} g(\tau)u(3T-\tau)\,d\tau + \int_T^{2T} g(\tau)u(3T-\tau)\,d\tau + \int_{2T}^{3T} g(\tau)u(3T-\tau)\,d\tau$$
$$= T[g(0)u(2T) + g(T)u(T) + g(2T)u(0)]$$

$$\vdots$$

$$y(NT) = T\sum_{i=0}^{N-1} g(iT)u(NT-iT-T)$$

29

Let

$$Y = \begin{bmatrix} y(T) \\ y(2T) \\ \vdots \\ y(NT) \end{bmatrix} \qquad G = \begin{bmatrix} g(0) \\ g(T) \\ \vdots \\ g(NT-T) \end{bmatrix}$$

$$U = \begin{bmatrix} u(0) & 0 & \cdots & 0 \\ u(T) & u(0) & \ddots & \vdots \\ u(2T) & u(T) & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ u(NT)-T & \cdots & u(T) & u(0) \end{bmatrix}$$

$$\Longrightarrow \quad Y = TUG \quad \Longrightarrow \quad \boxed{G = \frac{1}{T}U^{-1}Y}$$

30

♦ **Find the transfer function $G(q^{-1})$ from the impulse response $g(k)$**

The transfer function is determined by the impulse response, and there are many specific methods, such as semi-logarithmic method, order moment method, difference equation method, Hankel matrix method, etc.

Determination of system transfer function by Hankel matrix method

Let the impulse transfer function of the system be

$$G(q^{-1}) = \frac{b_1 q^{-1} + \cdots + b_n q^{-n}}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}} = g(1)q^{-1} + g(2)q^{-2} + \cdots = \sum_{i=1}^{\infty} g(i)q^{-i}$$

$$b_1 q^{-1} + \cdots + b_n q^{-n}$$

$$\Longrightarrow \quad = g(1)q^{-1} + [g(2) + a_1 g(1)]q^{-2} + \cdots + \left[g(n) + \sum_{i=1}^{n-1} a_{n-i} g(i)\right] q^{-n} +$$

$$\left[g(n+1) + \sum_{i=1}^{n} a_{n+1-i} g(i)\right] q^{-(n+1)} + \cdots + \left[g(2n) + \sum_{i=n}^{2n-1} a_{2n-i} g(i)\right] q^{-2}$$

The same power terms on both sides above should have the same coefficients, *i.e.*,

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ a_1 & 1 & \cdots & 0 & 0 \\ & & \cdots & & \\ a_{n-1} & a_{n-2} & \cdots & a_1 & 1 \end{bmatrix} \begin{bmatrix} g(1) \\ g(2) \\ \vdots \\ g(n) \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} g(1) & g(2) & \cdots & g(n) \\ g(2) & g(3) & \cdots & g(n+1) \\ \cdots & \cdots & \cdots & \cdots \\ g(n) & g(n+1) & \cdots & g(2n-1) \end{bmatrix}}_{H(n,1)} \begin{bmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_1 \end{bmatrix} = - \begin{bmatrix} g(n+1) \\ g(n+2) \\ \vdots \\ g(2n) \end{bmatrix}$$

Define the Hankel matrix:

$$H(l,k) = \begin{bmatrix} g(k) & g(k+1) & \cdots & g(k+l-1) \\ g(k+1) & g(k+2) & \cdots & g(k+l) \\ & \cdots & & \\ g(k+l-1) & g(k+l) & \cdots & g(k+2l-2) \end{bmatrix}$$

Find out a_i first, then b_i.

33

Example:

Given that the system to be identified is a third-order system, that is, the structural parameter n=3.

Take the step size T=0.05s, 2n=6

The sampling values of the impulse response are

t (s)	0.05	0.1	0.15	0.2	0.25	0.3
$g(t)$	7.1570	9.4911	8.5639	5.9305	2.8460	0.1446

34

Let $G(q^{-1}) = \dfrac{b_1 q^{-1} + b_2 q^{-2} + b_3 q^{-3}}{1 + a_1 q^{-1} + a_2 q^{-2} + a_3 q^{-3}}$

According to

$$\begin{bmatrix} g(1) & g(2) & \cdots & g(n) \\ g(2) & g(3) & \cdots & g(n+1) \\ \cdots & \cdots & \cdots & \cdots \\ g(n) & g(n+1) & \cdots & g(2n-1) \end{bmatrix} \begin{bmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_1 \end{bmatrix} = - \begin{bmatrix} g(n+1) \\ g(n+2) \\ \vdots \\ g(2n) \end{bmatrix}$$

then
$$\begin{bmatrix} 7.1570 & 9.4911 & 8.5639 \\ 9.4911 & 8.5639 & 5.93059 \\ 8.5639 & 5.93059 & 2.8460 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = - \begin{bmatrix} 5.9305 \\ 2.8460 \\ 0.1446 \end{bmatrix}$$

Solution:

$$a_1 = -2.2326, \quad a_2 = 1.7641, \quad a_3 = -0.4966$$

35

Substitute a_i into the following

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ a_1 & 1 & \cdots & 0 & 0 \\ & & \cdots & & \\ a_{n-1} & a_{n-2} & \cdots & a_1 & 1 \end{bmatrix} \begin{bmatrix} g(1) \\ g(2) \\ \vdots \\ g(n) \end{bmatrix}$$

then
$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2.2326 & 1 & 0 \\ 1.7641 & -2.2326 & 1 \end{bmatrix} \begin{bmatrix} 7.1570 \\ 9.4911 \\ 8.5639 \end{bmatrix}$$

Solution:

$$b_1 = 7.1570, \quad b_2 = -6.4875, \quad b_3 = 0.0$$

36

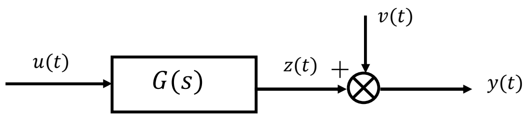
The transfer function of the system is

$$G(q^{-1}) = \frac{b_1 q^{-1} + \cdots + b_n q^{-n}}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}}$$

$$= \frac{7.1570q^{-1} - 6.4875q^{-2}}{1 - 2.2326q^{-1} + 1.7641q^{-2} - 0.4966q^{-3}}$$

■ Correlation analysis method

When the above-mentioned impulse response method contains a noise in the system, the uncertainty of the output result will lead to identification errors.



To avoid the influence of random noise $v(t)$, the correlation analysis method can be used to obtain the unit impulse response $g(t)$.

The correlation analysis method proposed in 1951 is an identification method to obtain the system impulse response, based on the correlation function between the stationary random input and output information of the plant.

According to the convolution relationship: $z(t) = \int_0^{\infty} g(\lambda)u(t-\lambda)d\lambda$
$(\text{zero initial condition: } u(t) = 0, \text{for } t \le 0)$

Since the output $y(t)$ is polluted by noise $v(t)$, there is

$$y(t) = z(t) + v(t) = \int_0^{\infty} g(\lambda)u(t-\lambda)d\lambda + v(t)$$

Multiply both sides of the equation by $u(t-\tau)$

$$\Longrightarrow \quad y(t)u(t-\tau) = \int_0^{\infty} g(\lambda)u(t-\lambda)u(t-\tau)d\lambda + v(t)u(t-\tau)$$

$\Downarrow$ Take the mathematical expectation

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T y(t)u(t-\tau)\,dt = \int_0^{\infty} g(\lambda)\left[\lim_{T\to\infty}\frac{1}{T}\int_0^T u(t-\lambda)u(t-\tau)dt\right]d\lambda + \lim_{T\to\infty}\frac{1}{T}\int_0^T v(t)u(t-\tau)\,dt$$

If $v(t)$ and $u(t)$ are uncorrelated and defined by the correlation function, then

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T v(t)u(t-\tau)\,dt = 0$$

The cross-correlation function $R_{uy}(\tau)$ between input $u(t)$ and output $y(t)$ is

$$R_{uy}(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_0^T y(t)u(t-\tau)\,dt$$

The autocorrelation function $R_{uu}(\tau)$ for input $u(t)$ is:

$$R_{uu}(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_0^T u(t)u(t-\tau)\,dt$$

$$\underbrace{\lim_{T\to\infty}\int_0^T y(t)u(t-\tau)\,dt}_{R_{uy}(\tau)} = \underbrace{\int_0^\infty g(\lambda)\left[\lim_{T\to\infty}\frac{1}{T}\int_0^T u(t-\lambda)u(t-\tau)dt\right]d\lambda}_{\int_0^\infty g(\lambda)\,R_{uu}(\tau-\lambda)d\lambda} + \underbrace{\lim_{T\to\infty}\frac{1}{T}\int_0^T v(t)u(t-\tau)\,dt}_{0}$$

$$\Rightarrow \boxed{R_{uy}(\tau) = \int_0^\infty g(\lambda)\,R_{uu}(\tau-\lambda)d\lambda}$$

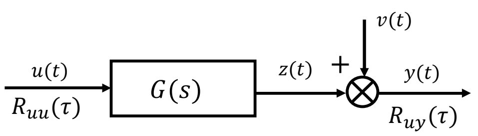
Wiener-Hopf Equation

The Wiener-Hopf equation has the same form as the convolution integral, which can be interpreted as:

For a system with an impulse response function of $g(t)$, if the input signal is the autocorrelation function of $u(t)$, then the impulse response of the system $g(t)$ is equal to the cross-correlation function between the input signal $R_{uu}(\tau)$ and the corresponding output signal $R_{uy}(\tau)$.

Therefore, the correlation analysis method has the effect of avoiding noise interference.

As long as the data of the input and output signals of the system are used to calculate their autocorrelation function and cross-correlation function, the impulse response function of the identified system may be obtained by solving the Wiener-Hopf equation.



How to solve the Wiener-Hopf equation?

When the input of the system to be identified adopts a white noise, the solution is easy, because the autocorrelation function of white noise is a δ function, namely $R_{uu}(\tau) = \sigma^2\delta(\tau)$

$$R_{uy}(\tau) = \int_0^\infty g(\lambda)\,R_{uu}(\tau-\lambda)d\lambda = \int_0^\infty g(\lambda)\,\sigma^2\delta(\tau-\lambda)d\lambda = \sigma^2 g(\tau)$$

Hence

$$g(\tau) = \frac{1}{\sigma^2} R_{uy}(\tau)$$

When the input of the system to be identified is a white noise, as long as the cross-correlation function between the input and output signals is determined, the impulse response function $g(\tau)$ of the system to be identified can be obtained

■ Identification of an impulse response using an M-sequence

Correlation functions:

$$R_{uy}(m) = \sum_{j=0}^{N_p-1} g(j) R_{uu}(m-j)$$

$$R_{uy}(m) = \frac{1}{N_p} \sum_{k=0}^{N_p-1} u(k-m) y(k)$$

$$\boxed{R_{uu}(m-j) = \frac{1}{N_p} \sum_{k=0}^{N_p-1} u(k-m) u(k-j)}$$

For an M-sequence,

$$R_{uu}(m-j) = \begin{cases} a^2 & m-j=0 \\ -\dfrac{a^2}{N_p} & 1 \le (m-j) \le N_p - 1 \end{cases}$$

$$
\begin{aligned}
R_{u\mathrm{y}}(m) &= \sum_{j=0}^{N_p-1} g(j) R_{uu}(m-j) \\
&= a^2 g(m) - \frac{a^2}{N_p}\left(\sum_{j=0}^{m-1} g(j) + \sum_{j=m+1}^{N_p-1} g(j)\right) \\
&= a^2 g(m) + \frac{a^2}{N_p} g(m) - \frac{a^2}{N_p}\sum_{j=0}^{N_p-1} g(j) \\
&= \frac{N_p+1}{N_p} a^2 g(m) - \frac{a^2}{N_p}\sum_{j=0}^{N_p-1} g(j) \\
&= \frac{N_p+1}{N_p} a^2 g(m) - c \qquad (3.3.1)
\end{aligned}
$$

where

$$
c = \frac{a^2}{N_p}\sum_{j=0}^{N_p-1} g(j)
$$

45

The parameter c can also be calculated by

$$
\begin{aligned}
\sum_{j=0}^{N_p-1} R_{u\mathrm{y}}(j) &= \sum_{j=0}^{N_p-1}\left(\frac{N_p+1}{N_p} a^2 g(j) - c\right) \\
&= \sum_{j=0}^{N_p-1}\left(\frac{N_p+1}{N_p} a^2 g(j)\right) - N_p c \\
&= \sum_{j=0}^{N_p-1}\left(\frac{N_p+1}{N_p} a^2 g(j)\right) - a^2\sum_{j=0}^{N_p-1} g(j) \\
&= a^2\left(\frac{N_p+1}{N_p} - 1\right)\sum_{j=0}^{N_p-1} g(j) \\
&= \frac{a^2}{N_p}\sum_{j=0}^{N_p-1} g(j) = c
\end{aligned}
$$

46

From eq. (3.3.1),

$$\hat{g}(m) = \frac{N_p}{N_p + 1}\frac{1}{a^2}\left(R_{uy}(m) + c\right)$$

As $m \to \infty,\ g(m) = 0.$ Then,

$$c = -R_{uy}(\infty)$$

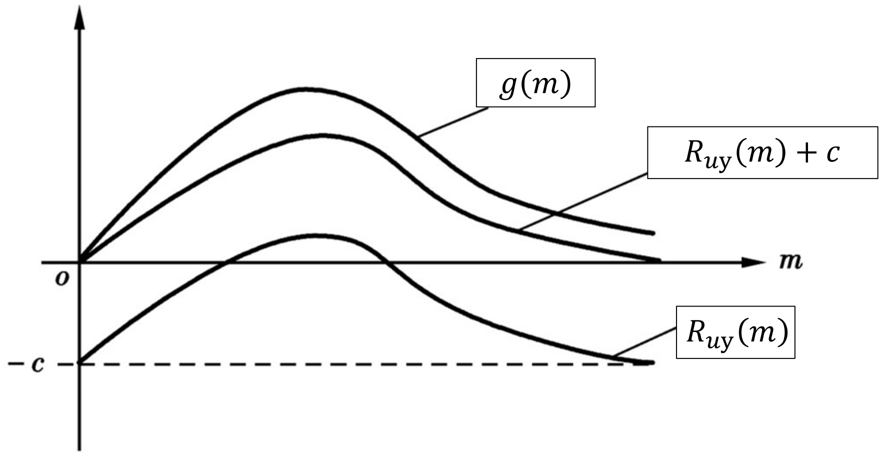
When N_p is very large, *c* can be ignored. So,

$$\hat{g}(m) = \frac{N_p}{(N_p + 1)a^2} R_{uy}(m)$$

To improve the identification accuracy of the impulse response, the M-sequence can periodically be used for r times. Then,

$$R_{uy}(m) = \frac{1}{rN_p}\sum_{k=0}^{rN_p - 1} u(k-m)y(k), \qquad m = 0, 1, \ldots, N_p - 1$$

Impulse response and cross-correlation function of a system with an M-sequence



■ Application of the correlation analysis method

The application of relevant identification technology in engineering can be summarized as follows:

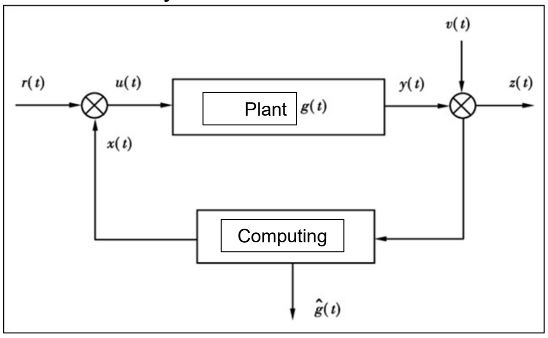
① On-line test of system dynamic characteristics.

② Conduct online tuning of a control system to optimize the parameters of the system;

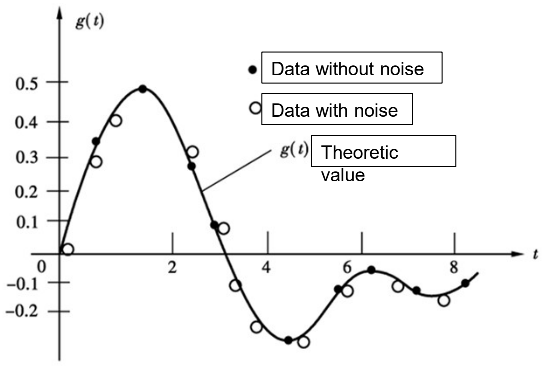
③ Non-parametric model identification in adaptive control, etc.

♦ Identification simulation

Simulation system structure



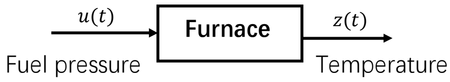
Identification results via simulation

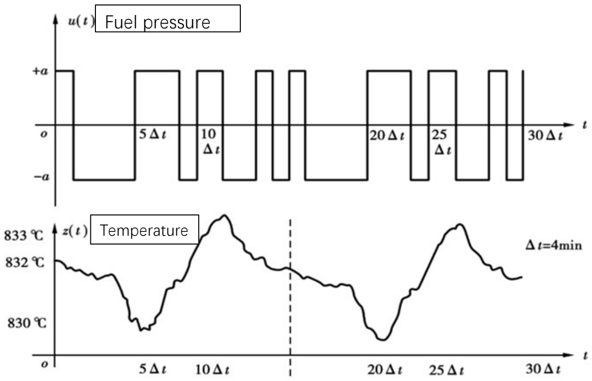


where the plant to be identified is

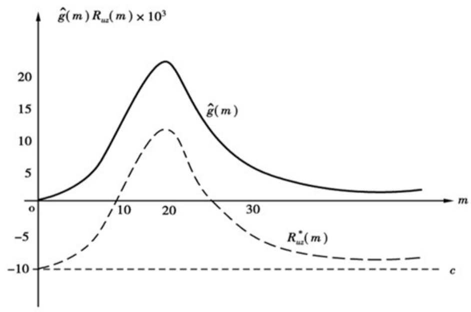
$$G(s) = \frac{2}{s^2 + s + 2}$$

♦ **Identification application to a furnace**





Identification results via measurements



3.4 Gradient correction parameter estimation method

The gradient correction parameter estimation algorithm has the same structure as the least squares recursive algorithm, namely

New estimation $\hat{\boldsymbol{\theta}}(\boldsymbol{k})$ = Old estimation $\hat{\boldsymbol{\theta}}(\boldsymbol{k}-\mathbf{1})$ + Innovation

But its principle is completely different from the least squares method.

The basic idea of the gradient correction method is: along the cost function (objective function), and gradually revise the estimated value of the model parameters until the cost function reaches its minimum.

This type of parameter estimation algorithms is simple and easy to understand, and the amount of real-time computation is small, but its convergence is slow.

■ Gradient correction parameter estimation method

The deterministic system is described as follows:

$$A(q^{-1})y(k) = B(q^{-1})u(k-d) \tag{3.4.1}$$

where $u(k)$ and $y(k)$ represent the input and output of the system, respectively, and

$$A(q^{-1}) = 1 + a_1q^{-1} + a_2q^{-2} + \cdots + a_{n_a}q^{-n_a}$$
$$B(q^{-1}) = b_0 + b_1q^{-1} + b_2q^{-2} + \cdots + b_{n_b}q^{-n_b}$$

Equation (3.4.1) can also be written as

$$y(k) = \varphi^T(k)\theta \tag{3.4.2}$$

where

$$\varphi(k) = [-y(k-1) \quad \cdots \quad -y(k-n_a) \quad u(k-d) \quad \cdots \quad u(k-d-n_b)]^T \in \mathcal{R}^{(n_a+n_b+1)\times 1}$$
$$\theta = \begin{bmatrix} a_1 & \cdots & a_{n_a} & b_0 & \cdots & b_{n_b} \end{bmatrix}^T \in \mathcal{R}^{(n_a+n_b+1)\times 1}$$

Assume that the parameters of the system are estimated as

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \gamma\varphi(k) \tag{3.4.3}$$

The parameter γ should be able to make $y(k) = \hat{y}(k)$ hold, *i.e.,*

$$y(k) = \varphi^T(k)\hat{\theta}(k) = \varphi^T(k)\hat{\theta}(k-1) + \gamma\varphi^T(k)\varphi(k)$$

It can be deduced from the above

$$\gamma = \frac{1}{\varphi^T(k)\varphi(k)}\left(y(k) - \varphi^T(k)\hat{\theta}(k-1)\right)$$

Substitute the above equation into Equation (3.4.3) to obtain

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{\varphi(k)}{\varphi^T(k)\varphi(k)}\left(y(k) - \varphi^T(k)\hat{\theta}(k-1)\right) \tag{3.4.4}$$

Equation (3.4.4) is essentially for a gradient correction method.

When $\varphi(k) = 0$, the algorithm is not feasible.

Equation (3.4.4) is often modified as

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{\alpha\varphi(k)}{c + \varphi^T(k)\varphi(k)}\left(y(k) - \varphi^T(k)\hat{\theta}(k-1)\right) \qquad (3.4.5)$$

where c>0 and 0< α< 2.

For the above values of c and α, it can be proved that the algorithm is convergent.

Being compared with the least squares method, the recursive gradient correction method is characterized by the scalar operation in the recursive process.

The load of computation was significantly reduced.

This kind of algorithms converges slowly and can be applied to adaptive control systems with low convergence speed.

♦ **Algorithm 3.4.1** (recursive gradient correction estimation, RGC)

Given n_a, n_b and d.

Step1: Set initial value $\hat{\theta}(0)$ and parameters c and α, enter initial data.

Step 2: Sample the current output $y(k)$ and input $u(k)$.

Step 3: Use Equation (3.4.5) to calculate $\hat{\theta}(k)$ 。

Step 4: $k \rightarrow k+1$, return to Step 2 and continue the loop.

♦ **Example**

Consider deterministic systems:

$$y(k) - 1.5y(k-1) + 0.7y(k-2) = u(k-3) + 0.5u(k-4)$$

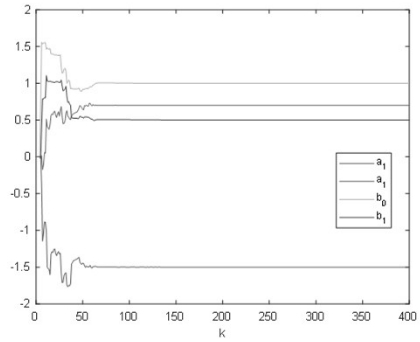
Note:

— Take initial value $\hat{\theta}(0) = 0$, and take $\alpha = 1$ and $c = 0.1$.

— Select a white noise with variance of 1 as the input signal $u(k)$.

— Use the RGC algorithm to estimate the parameters.

— The simulation results are shown on the left.



At $k = 400$, the estimated values of the parameters are

$\hat{a}_1 = -1.5$, $\hat{a}_2 = 0.7$, $\hat{b}_0 = 1.0$, $\hat{b}_1 = 0.5$

57

♦ **Simulation codes**

example341.m

Self Assessment Question

The conditions are exactly the same as the previous example. But, the estimation model form is assumed to be

$$y(k) - 1.2y(k-1) + 0.6y(k-2) = 0.3u(k-2) + 0.9u(k-4) + v(t)$$

where $v(t)$ is a random noise with mean of 0 and variance of 0.1. Determine the parameters of the above model at k=300 using the RGC method.

saq341.m

58

■ Stochastic Newton method

If the input and (or) output of the system contain noises, the RGC method is not applicable. The gradient correction method of the stochastic system is needed.

♦ Recursive stochastic Newton method (RSN)

Consider the model:

$$A(q^{-1})x(k) = B(q^{-1})v(k-d)$$

where

$$A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_{n_a} q^{-n_a}$$
$$B(q^{-1}) = b_0 + b_1 q^{-1} + b_2 q^{-2} + \cdots + b_{n_b} q^{-n_b}$$

Both input and output data are polluted by noises, *i.e.*

$$y(k) = x(k) + \xi(k)$$
$$u(k) = v(k) + \eta(k)$$

where $\xi(k)$ and $\eta(k)$ are uncorrelated random noises with mean of 0 and variance of σ^2.

The model with noises

$$A(q^{-1})(y(k) - \xi(k)) = B(q^{-1})(u(k-d) - \eta(k-d))$$

which can be rewritten as

$$A(q^{-1})y(k) = B(q^{-1})u(k-d) + A(q^{-1})\xi(k) - B(q^{-1})\eta(k-d)$$

The above model can be converted into the least squares format:

$$y(k) = \varphi^T(k)\theta + e(k)$$

where

$$\varphi(k) = [-y(k-1) \quad \cdots \quad -y(k-n_a) \quad u(k-d) \quad \cdots \quad u(k-d-n_b)]^T \in \mathcal{R}^{(n_a+n_b+1)\times 1}$$

$$\theta = \left[a_1 \quad \cdots \quad a_{n_a} \quad b_0 \quad \cdots \quad b_{n_b}\right]^T \in \mathcal{R}^{(n_a+n_b+1)\times 1}$$

$$e(k) = A(q^{-1})\xi(k) - B(q^{-1})\eta(k-d)$$

If the mean value of noise $e(k)$ is 0, the system parameters can be estimated by stochastic Newton algorithm.

The parameter estimation using the recursive stochastic Newton algorithm

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \rho(k)R^{-1}(k)\varphi(k)\left(y(k) - \varphi^T(k)\hat{\theta}(k-1)\right) \tag{3.4.6}$$

$$R(k) = R(k-1) + \rho(k)\left(\varphi(k)\varphi^T(k) - R(k-1)\right) \tag{3.4.7}$$

where the initial value of $R(k)$ can be taken as $R(0) = \mathrm{I}$, $\rho(k)$ is the convergence factor and is required to meet

$$\rho(k) > 0, \forall k; \qquad \lim_{k\to\infty} \rho(k) = 0$$

$$\sum_{k=1}^{\infty} \rho(k) = \infty; \qquad \sum_{k=1}^{\infty} \rho^2(k) = \sigma^2$$

□□ □ □□ □□

The algorithm is detailed in the books:

[1]方崇智,萧德云.过程辨识[M].北京:清华大学出版社,1988.

[2]冯培悌.系统辨识[M].杭州:浙江大学出版社,1999.

♦ **Algorithm 3.4.2** (recursive stochastic Newton algorithm, RSNA)

Given n_a, n_b and d.

Step 1: Set initial value $\hat{\theta}(0)$ and $R(0)$, enter the initial data.

Step 2: Sample the current output $y(k)$ and input $u(k)$.

Step 3: Select the convergence factor $\rho(k)$, calculate $R(k)$ and $\hat{\theta}(k)$ using (3.4.7) and (3.4.6), respectively.

Step 4: k → k+1, return to Step 2 and continue the loop.

♦ **Example**

Consider the system:

$$x(k) - 1.5x(k-1) + 0.7x(k-2) = v(k-3) + 0.5v(k-4)$$
$$u(k) = v(k) + \eta(k)$$
$$y(k) = x(k) + \xi(k)$$

where $\xi(k)$ and $\eta(k)$ are white noises with variance of 0.1 and 0.25, respectively.

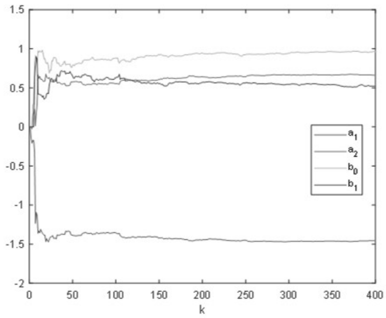
Note:

— The initial value is $R(0) = I, \hat{\theta}(0) = 0$, and take $\rho(k) = 1, \forall k$.

— Select white noise with variance of 1 as the input signal $u(k)$.

— RSNA is used for parameter estimation.

— The simulation results are shown on the right.



At $k = 400$, the estimated values of the parameters are

$\hat{a}_1 = -1.4771, \quad \hat{a}_2 = 0.6853, \quad \hat{b}_0 = 0.9906, \quad \hat{b}_1 = 0.5435$

♦ **Simulation codes**

example342.m

Exercise 3.1

Let the difference equation of the single-input-single-output system be

$$y(k) + a_1 y(k-1) + a_2 y(k-2) = b_1 u(k-1) + b_2 u(k-2) + V(k)$$
$$V(k) = c_1 v(k) + c_2 v(k-1) + c_3 v(k-2)$$

Take

$$a_1 = 1.6, \ \ a_2 = 0.7, \ b_1 = 1.0, b_2 = 0.4, c_1 = 0.9, c_2 = 1.2, \ c_3 = 0.3$$

and the input signal adopts a 4th-order M-sequence with an amplitude of 1.

When the mean and variance of Gaussian noise $v(k)$ are 0 and 0.5, the parameters are estimated by the maximum likelihood identification method, correlation analysis method, recursive gradient correction estimation method, and recursive stochastic Newton method, respectively.