Lecture 2: Real-Time Parameter Estimation

- Least Squares and Regression Models
- Estimating Parameters in Dynamical Systems
- Examples

Preliminary Comments

Real-time parameter estimation is one of methods of system identification

Preliminary Comments

Real-time parameter estimation is one of methods of system identification

---

The key elements of system identification are:

Preliminary Comments

Real-time parameter estimation is one of methods of system identification

---

The key elements of system identification are:

- Selection of model structure (linear, nonlinear, linear in parameters, . . .)

Preliminary Comments

Real-time parameter estimation is one of methods of system identification

---

The key elements of system identification are:

- Selection of model structure (linear, nonlinear, linear in parameters, . . .)
- Design of experiments (input signal, sampling period, . . .)

Preliminary Comments

Real-time parameter estimation is one of methods of system identification

---

The key elements of system identification are:

- Selection of model structure (linear, nonlinear, linear in parameters, . . .)
- Design of experiments (input signal, sampling period, . . .)
- Parameter estimation (off-line, on-line, . . .)

Preliminary Comments

Real-time parameter estimation is one of methods of system identification

---

The key elements of system identification are:

- Selection of model structure (linear, nonlinear, linear in parameters, . . .)
- Design of experiments (input signal, sampling period, . . .)
- Parameter estimation (off-line, on-line, . . .)
- Validation mechanism (depends on application)

All these step should be present

in Real-Time Parameter Estimation Algorithms!

Parameter Estimation: Problem Formulation

Suppose that a system (a model of experiment) is

$$\begin{aligned} y(i) &= \phi_1(i)\theta_1^0 + \phi_2(i)\theta_2^0 + \cdots + \phi_n(i)\theta_n^0 \\ &= \underbrace{\left[\phi_1(i),\ \phi_2(i),\ \ldots,\ \phi_n(i)\right]}_{\phi(i)^T \ \longleftarrow\ 1\times n} \underbrace{\begin{bmatrix} \theta_1^0 \\ \vdots \\ \theta_n^0 \end{bmatrix}}_{\theta^0 \ \longleftarrow\ n\times 1} = \phi(i)^T\,\theta^0 \end{aligned}$$

Parameter Estimation: Problem Formulation

Suppose that a system (a model of experiment) is

$$\begin{aligned} y(i) &= \phi_1(i)\theta_1^0 + \phi_2(i)\theta_2^0 + \cdots + \phi_n(i)\theta_n^0 \\ &= \underbrace{\left[\phi_1(i),\ \phi_2(i),\ \ldots,\ \phi_n(i)\right]}_{\phi(i)^T \ \longleftarrow\ 1\times n} \underbrace{\begin{bmatrix} \theta_1^0 \\ \vdots \\ \theta_n^0 \end{bmatrix}}_{\theta^0 \ \longleftarrow\ n\times 1} = \phi(i)^T\,\theta^0 \end{aligned}$$

Given the data

$$\Big\{[y(i),\ \phi(i)]\Big\}_{i=1}^{N} = \Big\{[y(1),\ \phi(1)],\ \ldots,[y(N),\ \phi(N)]\Big\}$$

The task is to find (estimate) the n unknown values

$$\theta^0 = \left[\theta_1^0,\ \theta_2^0,\ \ldots,\ \theta_n^0\right]^T$$

What kind of solution is expected?

Algorithm solving the estimation problem should be as follows.

- Estimates should be computed in a numerically efficient way (an analytical formula is preferable!)

What kind of solution is expected?

Algorithm solving the estimation problem should be as follows.

- Estimates should be computed in a numerically efficient way (an analytical formula is preferable!)
- Computed estimates (the mean value) should be close to the real estimated values (be unbiased), provided that the model is correct.

What kind of solution is expected?

Algorithm solving the estimation problem should be as follows.

- Estimates should be computed in a numerically efficient way (an analytical formula is preferable!)
- Computed estimates (the mean value) should be close to the real estimated values (be unbiased), provided that the model is correct.
- Estimates should have small variances, preferably decreasing with $N \to \infty$ (more data –> better estimate).

What kind of solution is expected?

Algorithm solving the estimation problem should be as follows.

- Estimates should be computed in a numerically efficient way (an analytical formula is preferable!)
- Computed estimates (the mean value) should be close to the real estimated values (be unbiased), provided that the model is correct.
- Estimates should have small variances, preferably decreasing with $N \to \infty$ (more data –> better estimate).
- The algorithm, if possible, should have intuitive clear motivation.

Least Squares Algorithm:

Consider the function (loss-function) to be minimized

$$V_N(\theta) = \frac{1}{2}\left\{(y(1) - \phi(1)^T\theta)^2 + \cdots + (y(N) - \phi(N)^T\theta)^2\right\}$$

Least Squares Algorithm:

Consider the function (loss-function) to be minimized

$$\begin{aligned} V_N(\theta) &= \frac{1}{2}\Big\{\underbrace{(y(1)-\phi(1)^T\theta)}_{\varepsilon(1)}{}^2 + \cdots + \underbrace{(y(N)-\phi(N)^T\theta)}_{\varepsilon(N)}{}^2\Big\} \\ &= \frac{1}{2}\Big\{\varepsilon(1)^2 + \varepsilon(2)^2 + \cdots + \varepsilon(N)^2\Big\} = \frac{1}{2}E^T E \end{aligned}$$

Here

$$E = \Big[\varepsilon(1),\, \varepsilon(2),\, \ldots,\, \varepsilon(N)\Big]^T = Y - \Phi\,\theta$$

and

$$Y = \Big[y(1),\, y(2),\, \ldots,\, y(N)\Big]^T, \quad \Phi = \begin{bmatrix} \phi(1)^T \\ \phi(2)^T \\ \vdots \\ \phi(N)^T \end{bmatrix} \longleftarrow \quad N \times n$$

Theorem (Least Square Formula):

The function $V_N(\theta)$ is minimal at $\theta = \theta$, which satisfies the equation

$$\Phi^T \Phi \, \theta = \Phi^T \, Y$$

Theorem (Least Square Formula):

The function $V_N(\theta)$ is minimal at $\theta = \theta$, which satisfies the equation

$$\Phi^T \Phi\, \theta = \Phi^T\, Y$$

If $\det\left(\Phi^T \Phi\right) \neq 0$, then

- the minimum of the function $V_N(\theta)$ is unique,
- the minimum value is $\frac{1}{2} Y^T Y - \frac{1}{2} Y^T \Phi \left(\Phi^T \Phi\right)^{-1} \Phi^T Y$ and is attained at

$$\boxed{\theta = \left(\Phi^T \Phi\right)^{-1} \Phi^T Y}$$

Proof:

The loss-function can be written as

$$\boldsymbol{V_N(\theta)} \;=\; \frac{\mathbf{1}}{\mathbf{2}}\left(\boldsymbol{Y} - \boldsymbol{\Phi\theta}\right)^{\boldsymbol{T}}\left(\boldsymbol{Y} - \boldsymbol{\Phi\theta}\right)$$

It reaches its minimal value at $\boldsymbol{\theta}$ if

$$\boldsymbol{\nabla_\theta V_N(\theta)} = \left[\frac{\boldsymbol{\partial V_N(\theta)}}{\boldsymbol{\partial\theta_1}}, \ldots, \frac{\boldsymbol{\partial V_N(\theta)}}{\boldsymbol{\partial\theta_n}}\right] = \mathbf{0} \quad \text{with } \boldsymbol{\theta} = \boldsymbol{\theta}$$

Proof:

The loss-function can be written as

$$V_N(\theta) \quad = \quad \frac{1}{2}\left(Y - \Phi\theta\right)^T \left(Y - \Phi\theta\right)$$

It reaches its minimal value at θ if

$$\nabla_\theta V_N(\theta) = \left[\frac{\partial V_N(\theta)}{\partial \theta_1}, \ldots, \frac{\partial V_N(\theta)}{\partial \theta_n}\right] = 0 \quad \text{with } \theta = \theta$$

This equation is

$$\nabla_\theta V_N(\theta) = \frac{1}{2} \cdot 2 \cdot \nabla_\theta \left(-\Phi\theta\right)^T \left(Y - \Phi\theta\right) = -\Phi^T \left(Y - \Phi\theta\right) = 0$$

$$\left(\quad \nabla_\theta \left(\theta^T a\right) = \nabla_\theta \left(a^T \theta\right) = a^T, \qquad \nabla_\theta \left(\theta^T A\, \theta\right) = \theta^T \left(A + A^T\right) \quad \right)$$

Proof:

The loss-function can be written as

$$V_N(\theta) \;=\; \frac{1}{2}\,(Y - \Phi\theta)^T\,(Y - \Phi\theta)$$

It reaches its minimal value at θ if

$$\nabla_\theta V_N(\theta) = \left[\frac{\partial V_N(\theta)}{\partial \theta_1}, \ldots, \frac{\partial V_N(\theta)}{\partial \theta_n}\right] = 0 \quad \text{with } \theta = \theta$$

This equation is

$$\nabla_\theta V_N(\theta) = \frac{1}{2}\cdot 2\cdot \nabla_\theta\,(-\Phi\theta)^T\;(Y - \Phi\theta) = -\Phi^T\;(Y - \Phi\theta) = 0$$

$\Big(\;\; \nabla_\theta\,(\theta^T\,a) = \nabla_\theta\,(a^T\,\theta) = a^T, \qquad \nabla_\theta\,(\theta^T\,A\,\theta) = \theta^T\,(A + A^T)\;\;\Big)$ Hence

$$\Phi^T\,\Phi\,\theta = \Phi^T\,Y$$

Proof (con'd):

If the matrix $\boldsymbol{\Phi}^T\,\boldsymbol{\Phi}$ is invertible, then we solve can find the solution $\boldsymbol{\theta}$ of $\quad \boldsymbol{\Phi}^T\,\boldsymbol{\Phi}\,\boldsymbol{\theta} = \boldsymbol{\Phi}^T\,\boldsymbol{Y}$:

$$\boldsymbol{\theta} = \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^T\boldsymbol{Y}$$

Proof (con'd):

If the matrix $\boldsymbol{\Phi}^T \boldsymbol{\Phi}$ is invertible, then we solve can find the solution $\boldsymbol{\theta}$ of $\quad \boldsymbol{\Phi}^T \boldsymbol{\Phi} \boldsymbol{\theta} = \boldsymbol{\Phi}^T \boldsymbol{Y}$:

$$\boldsymbol{\theta} = \left(\boldsymbol{\Phi}^T \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^T \boldsymbol{Y}$$

$$V_N(\theta) = \frac{1}{2} (Y - \Phi\theta)^T (Y - \Phi\theta) = \frac{1}{2} Y^T Y - Y^T \Phi\theta + \frac{1}{2} \theta^T \Phi^T \Phi\theta$$

Proof (con'd):

If the matrix $\Phi^T\,\Phi$ is invertible, then we solve can find the solution θ of $\quad \Phi^T\,\Phi\,\theta = \Phi^T\,Y$:

$$\theta = \left(\Phi^T\Phi\right)^{-1}\Phi^T Y$$

$$V_N(\theta) = \frac{1}{2}\left(Y - \Phi\theta\right)^T\left(Y - \Phi\theta\right) = \frac{1}{2}Y^T Y - Y^T\Phi\theta + \frac{1}{2}\theta^T\Phi^T\Phi\theta$$

and since

$$Y^T\Phi\theta = Y^T\Phi\left(\Phi^T\Phi\right)^{-1}\left(\Phi^T\Phi\right)\theta = \theta^T\left(\Phi^T\Phi\right)\theta$$

by completing the square

$$V_N(\theta) = \underbrace{\frac{1}{2}Y^T Y - \frac{1}{2}\theta^T\left(\Phi^T\Phi\right)\theta}_{\text{independent of }\theta} + \underbrace{\frac{1}{2}(\theta - \theta)^T\left(\Phi^T\Phi\right)(\theta - \theta)}_{\geq 0,\ \cdots > 0 \text{ for } \theta \neq \theta}$$

Comments:

The formula can be re-written in terms of $\boldsymbol{y(i)}$, $\boldsymbol{\phi(i)}$ as

$$\boldsymbol{\theta} \;=\; \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^T\boldsymbol{Y} = \left(\sum_{i=1}^{N}\boldsymbol{\phi(i)\phi(i)}\right)^{-1}\left(\sum_{i=1}^{N}\boldsymbol{\phi(i)y(i)}\right)$$

Comments:

The formula can be re-written in terms of $y(i)$, $\phi(i)$ as

$$\boldsymbol{\theta} = \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^T\boldsymbol{Y} = \boldsymbol{P}(\boldsymbol{N})\left(\sum_{i=1}^{N}\phi(i)y(i)\right)$$

where $$\boldsymbol{P}(\boldsymbol{N}) = \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right)^{-1} = \left(\sum_{i=1}^{N}\phi(i)\phi(i)\right)^{-1}$$

Comments:

The formula can be re-written in terms of $y(i)$, $\phi(i)$ as

$$\theta = \left(\Phi^T\Phi\right)^{-1}\Phi^T Y = P(N)\left(\sum_{i=1}^{N}\phi(i)y(i)\right)$$

where $$P(N) = \left(\Phi^T\Phi\right)^{-1} = \left(\sum_{i=1}^{N}\phi(i)\phi(i)\right)^{-1}$$

The condition $\det\left(\Phi^T\Phi\right) \neq 0$ is called an excitation condition.

Comments:

The formula can be re-written in terms of $y(i)$, $\phi(i)$ as

$$\theta = \left(\Phi^T \Phi\right)^{-1} \Phi^T Y = P(N) \left(\sum_{i=1}^{N} \phi(i) y(i) \right)$$

where $$P(N) = \left(\Phi^T \Phi\right)^{-1} = \left(\sum_{i=1}^{N} \phi(i) \phi(i) \right)^{-1}$$

The condition $\det \left(\Phi^T \Phi\right) \neq 0$ is called an excitation condition.

To assign different weights for different time instants, consider

$$\begin{aligned} V_N(\theta) &= \tfrac{1}{2} \left\{ w_1 (y(1) - \phi(1)^T \theta)^2 + \cdots + w_N (y(N) - \phi(N)^T \theta)^2 \right\} \\ &= \tfrac{1}{2} \left\{ w_1 \varepsilon(1)^2 + w_2 \varepsilon(2)^2 + \cdots + w_N \varepsilon(N)^2 \right\} = \frac{1}{2} E^T W E \end{aligned}$$

Comments:

The formula can be re-written in terms of $y(i)$, $\phi(i)$ as

$$\theta = \left(\Phi^T\Phi\right)^{-1}\Phi^T Y = P(N)\left(\sum_{i=1}^{N}\phi(i)y(i)\right)$$

where $$P(N) = \left(\Phi^T\Phi\right)^{-1} = \left(\sum_{i=1}^{N}\phi(i)\phi(i)\right)^{-1}$$

The condition $\det\left(\Phi^T\Phi\right) \neq 0$ is called an excitation condition.

To assign different weights for different time instants, consider

$$V_N(\theta) = \tfrac{1}{2}\left\{w_1\left(y(1)-\phi(1)^T\theta\right)^2+\cdots+w_N\left(y(N)-\phi(N)^T\theta\right)^2\right\}$$

$$= \tfrac{1}{2}\left\{w_1\varepsilon(1)^2+w_2\varepsilon(2)^2+\cdots+w_N\varepsilon(N)^2\right\} = \frac{1}{2}E^T W E$$

$$\Rightarrow \quad \boxed{\theta = \left(\Phi^T W\Phi\right)^{-1}\Phi^T W Y}$$

Exponential forgetting

The common way to assign the weights is Least Squares with Exponential Forgetting:

$$V_N(\theta) = \frac{1}{2}\sum_{i=1}^{N} \lambda^{N-i}\Big(y(i) - \phi(i)^T\,\theta\Big)^2$$

with $0 < \lambda < 1$. What is the solution formula?

Stochastic Notation

Uncertainty is often convenient to represent stochastically.

Stochastic Notation

Uncertainty is often convenient to represent stochastically.

- e is random if it is not known in advance and could take values from the set $\{e^1, \ldots, e^i, \ldots\}$ with probabilities p^i.
- The mean value or expected value is

$$E\,e = \sum_i p^i\, e^i, \qquad E \text{ is the expectation operator.}$$

For a realization $\{e_i\}_{i=1}^{N}$: $\qquad E\,e \approx \frac{1}{N}\sum_{i=1}^{N} e_i$

Stochastic Notation

Uncertainty is often convenient to represent stochastically.

- e is random if it is not known in advance and could take values from the set $\{e^1, \dots, e^i, \dots\}$ with probabilities p^i.
- The mean value or expected value is

$$E\, e = \sum_i p^i\, e^i, \qquad E \text{ is the expectation operator.}$$

For a realization $\{e_i\}_{i=1}^N$: $\qquad E\, e \approx \frac{1}{N} \sum_{i=1}^N e_i$

- The variance or dispersion characterizes possible deviations from the mean value:

$$\sigma^2 = \mathrm{var}(e) = E\,(e - E\, e)^2.$$

Stochastic Notation

Uncertainty is often convenient to represent stochastically.

- e is random if it is not known in advance and could take values from the set $\{e^1, \ldots, e^i, \ldots\}$ with probabilities p^i.
- The mean value or expected value is

$$E\, e = \sum_i p^i\, e^i, \qquad E \text{ is the expectation operator.}$$

For a realization $\{e_i\}_{i=1}^N$: $\qquad E\, e \approx \frac{1}{N} \sum_{i=1}^N e_i$

- The variance or dispersion characterizes possible deviations from the mean value:

$$\sigma^2 = \mathrm{var}(e) = E\,(e - E\, e)^2.$$

- Two random variables e and f are independent if

$$E\,(e\, f) = (E\, e) \cdot (E\, f) \qquad \text{or} \qquad \mathrm{cov}(e, f) = 0.$$

where $\mathrm{cov}(e, f) = E\big((e - E\, e)(f - E\, f)\big)$ is covariance.

Statistical Properties of LS Estimate:

Assume that the model of the system is stochastic, i.e.

$$y(i) = \phi(i)^T\theta^0 + e(i) \qquad \left\{ \Leftrightarrow Y(N) = \Phi(N)\theta^0 + E(N) \right\}.$$

Here θ^0 is vector of true parameters, $0 \le i \le N \le +\infty$,

- $\left\{ e(i) \right\} = \left\{ e(0),\, e(1),\, \ldots \right\}$ is a sequence of independent equally distributed random variables with $Ee(i) = 0$,
- $\left\{ \phi(i) \right\}$ is either a sequence of random variables independent of e or deterministic.

Statistical Properties of LS Estimate:

Assume that the model of the system is stochastic, i.e.

$$y(i) = \phi(i)^T\theta^0 + e(i) \qquad \left\{ \Leftrightarrow Y(N) = \Phi(N)\theta^0 + E(N) \right\}.$$

Here θ^0 is vector of true parameters, $0 \leq i \leq N \leq +\infty$,

- $\left\{e(i)\right\} = \left\{e(0),\, e(1),\, \ldots\right\}$ is a sequence of independent equally distributed random variables with $Ee(i) = 0$,
- $\left\{\phi(i)\right\}$ is either a sequence of random variables independent of e or deterministic.

Pre-multiplying the model by $\left(\Phi(N)^T\Phi(N)\right)^{-1}\Phi(N)^T$, gives $\left(\Phi(N)^T\Phi(N)\right)^{-1}\Phi(N)^T\times Y(N) = \hat{\theta}$

$$\hat{\theta} = \left(\Phi(N)^T\Phi(N)\right)^{-1}\Phi(N)^T \times \left(\Phi(N)\theta^0 + E(N)\right)$$

Statistical Properties of LS Estimate:

Assume that the model of the system is stochastic, i.e.

$$y(i) = \phi(i)^T\theta^0 + e(i) \qquad \left\{ \Leftrightarrow Y(N) = \Phi(N)\theta^0 + E(N) \right\}.$$

Here θ^0 is vector of true parameters, $0 \le i \le N \le +\infty$,

- $\left\{e(i)\right\} = \left\{e(0),\, e(1),\, \ldots\right\}$ is a sequence of independent equally distributed random variables with $Ee(i) = 0$,
- $\left\{\phi(i)\right\}$ is either a sequence of random variables independent of e or deterministic.

Pre-multiplying the model by $\left(\Phi(N)^T\Phi(N)\right)^{-1}\Phi(N)^T$, gives

$$\hat{\theta} \;=\; \theta^0 + \left(\Phi(N)^T\Phi(N)\right)^{-1}\Phi(N)^T \times E(N)$$

Theorem: Suppose the data are generated by

$$y(i) = \phi(i)^T \theta^0 + e(i) \qquad \left\{ \Leftrightarrow Y(N) = \Phi(N)\theta^0 + E(N) \right\}$$

where $\left\{ e(i) \right\} = \left\{ e(0),\, e(1),\, \ldots \right\}$ is a sequence of independent random variables with zero mean and variance σ^2.

Theorem: Suppose the data are generated by

$$y(i) = \phi(i)^T \theta^0 + e(i) \qquad \left\{ \Leftrightarrow Y(N) = \Phi(N)\theta^0 + E(N) \right\}$$

where $\left\{e(i)\right\} = \left\{e(0),\, e(1),\, \ldots\right\}$ is a sequence of independent random variables with zero mean and variance σ^2.

Consider the least square estimate

$$\hat{\theta} = \left(\Phi(N)^T \Phi(N)\right)^{-1} \Phi(N)^T Y(N).$$

Theorem: Suppose the data are generated by

$$y(i) = \phi(i)^T\theta^0 + e(i) \qquad \left\{ \Leftrightarrow Y(N) = \Phi(N)\theta^0 + E(N) \right\}$$

where $\left\{e(i)\right\} = \left\{e(0),\, e(1),\, \dots\right\}$ is a sequence of independent random variables with zero mean and variance σ^2.

Consider the least square estimate

$$\hat{\theta} = \left(\Phi(N)^T\Phi(N)\right)^{-1}\Phi(N)^T Y(N).$$

If $\det \Phi(N)^T\Phi(N) \neq 0$, then

- $E\,\hat{\theta} = \theta^0$, i.e. the estimate is unbiased;
- the covariance of the estimate is

$$\operatorname{cov}\hat{\theta} = E\left(\hat{\theta} - \theta^0\right)\left(\hat{\theta} - \theta^0\right)^T = \sigma^2\left(\Phi(N)^T\Phi(N)\right)^{-1}$$

Regression Models for FIR

FIR (Finite Impulse Response) or MA (Moving Average Model):

$$y(t) = b_1\, u(t-1) + b_2\, u(t-2) + \cdots + b_n\, u(t-n)$$

Regression Models for FIR

FIR (Finite Impulse Response) or MA (Moving Average Model):

$$y(t) = b_1\, u(t-1) + b_2\, u(t-2) + \cdots + b_n\, u(t-n)$$

$$y(t) = \underbrace{\left[u(t-1),\ u(t-2),\ \ldots,\ u(t-n)\right]}_{=\phi(t-1)^T} \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}}_{=\theta}$$

Regression Models for FIR

FIR (Finite Impulse Response) or MA (Moving Average Model):

$$y(t) = b_1\, u(t-1) + b_2\, u(t-2) + \cdots + b_n\, u(t-n)$$

$$y(t) = \underbrace{\left[u(t-1),\ u(t-2),\ \ldots,\ u(t-n)\right]}_{=\phi(t-1)^T} \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}}_{=\theta}$$

Note the change of notation for the typical case when the RHS does no have the term $b_0\, u(t)$: $\phi(i) \longrightarrow \phi(t-1)$ to indicate dependence of date on input signals up to $t-1$th.

However, we keep: $\qquad Y(N) = \Phi(N)\,\theta$

Regression Models for ARMA

IIR (Infite Impulse Response) or ARMA (Autoregressive Moving Average Model):

$$\underbrace{\left(q^n + a_1\, q^{n-1} + \cdots + a_n\right)}_{A(q)}\, y(t) = \underbrace{\left(b_1\, q^{m-1} + b_2\, q^{m-2} + \cdots + b_m\right)}_{B(q)}\, u(t)$$

Regression Models for ARMA

IIR (Infite Impulse Response) or ARMA (Autoregressive Moving Average Model):

$$\underbrace{\left(q^n + a_1\, q^{n-1} + \cdots + a_n\right)}_{A(q)}\, y(t) = \underbrace{\left(b_1\, q^{m-1} + b_2\, q^{m-2} + \cdots + b_m\right)}_{B(q)}\, u(t)$$

q is the forward shift operator:

$$q\, u(t) = u(t+1), \qquad q^{-1}\, u(t) = u(t-1)$$

Regression Models for ARMA

IIR (Infite Impulse Response) or ARMA (Autoregressive Moving Average Model):

$$\underbrace{\left(q^n + a_1\, q^{n-1} + \cdots + a_n\right)}_{A(q)}\, y(t) = \underbrace{\left(b_1\, q^{m-1} + b_2\, q^{m-2} + \cdots + b_m\right)}_{B(q)}\, u(t)$$

q is the forward shift operator:

$$q\, u(t) = u(t+1), \qquad q^{-1}\, u(t) = u(t-1)$$

$$y(t) = \underbrace{\left[-y(t-1), \ldots, -y(t-n), u(t-n+m-1), \ldots, u(t-n)\right]}_{=\phi(i-1)^T} \underbrace{\begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_m \end{bmatrix}}_{=\theta}$$

Problem 2.2

Consider the FIR model

$$y(t) = b_0 u(t) + b_1 u(t-1) + e(t), \quad t = 1, 2, 3, \ldots, N$$

where $\{e(t)\}$ is a sequence of independent normal $\mathcal{N}(0, \sigma)$ random variables

Problem 2.2

Consider the FIR model

$$y(t) = b_0 u(t) + b_1 u(t-1) + e(t), \quad t = 1, 2, 3, \ldots, N$$

where $\{e(t)\}$ is a sequence of independent normal $\mathcal{N}(0, \sigma)$ random variables

---

- Determine LS estimates for b_0, b_1 when $u(t)$ is a step. Analyze the covariance of the estimate, when $N \to \infty$
- Make the same investigation when $u(t)$ is white noise with unit variance.

Solution (Problem 2.2):

For the model

$$y(t) = b_0 u(t) + b_1 u(t-1) + e(t)$$

the regression form is readily seen

$$y(t) = \big[u(t),\ u(t-1)\big] \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} + e(t) = \phi(t)^T \theta + e(t)$$

Solution (Problem 2.2):

For the model

$$y(t) = b_0 u(t) + b_1 u(t-1) + e(t)$$

the regression form is readily seen

$$y(t) = \big[u(t),\, u(t-1)\big] \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} + e(t) = \phi(t)^T \theta + e(t)$$

Then LS estimate is

$$\hat{\theta} \;=\; \Big(\Phi(N)^T \Phi(N)\Big)^{-1} \Phi(N)^T Y(N)$$

Solution (Problem 2.2):

For the model

$$y(t) = b_0 u(t) + b_1 u(t-1) + e(t)$$

the regression form is readily seen

$$y(t) = \big[u(t),\, u(t-1)\big] \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} + e(t) = \phi(t)^T \theta + e(t)$$

Then LS estimate is

$$\hat{\theta} = \left(\begin{bmatrix} u(1) & u(0) \\ u(2) & u(1) \\ \vdots & \\ u(N) & u(N-1) \end{bmatrix}^T \begin{bmatrix} u(1) & u(0) \\ u(2) & u(1) \\ \vdots & \\ u(N) & u(N-1) \end{bmatrix} \right)^{-1} \begin{bmatrix} u(1) & u(0) \\ u(2) & u(1) \\ \vdots & \\ u(N) & u(N-1) \end{bmatrix}^T \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix}$$

Solution (Con'd):

The formula for an arbitrary input signal $u(t)$

$$\hat{\theta} = \left(\begin{bmatrix} u(1) & u(0) \\ u(2) & u(1) \\ \vdots & \\ u(N) & u(N-1) \end{bmatrix}^T \begin{bmatrix} u(1) & u(0) \\ u(2) & u(1) \\ \vdots & \\ u(N) & u(N-1) \end{bmatrix} \right)^{-1} \begin{bmatrix} u(1) & u(0) \\ u(2) & u(1) \\ \vdots & \\ u(N) & u(N-1) \end{bmatrix}^T \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix}$$

Solution (Con'd):

The formula for an arbitrary input signal $u(t)$

$$\hat{\theta} = \left(\begin{bmatrix} u(1) & u(0) \\ u(2) & u(1) \\ \vdots & \\ u(N) & u(N-1) \end{bmatrix}^T \begin{bmatrix} u(1) & u(0) \\ u(2) & u(1) \\ \vdots & \\ u(N) & u(N-1) \end{bmatrix} \right)^{-1} \begin{bmatrix} u(1) & u(0) \\ u(2) & u(1) \\ \vdots & \\ u(N) & u(N-1) \end{bmatrix}^T \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix}$$

becomes

$$\hat{\theta} = \begin{bmatrix} \sum_{t=1}^{N} u(t)^2 & \sum_{t=1}^{N} u(t)u(t-1) \\ \sum_{t=1}^{N} u(t)u(t-1) & \sum_{t=1}^{N} u(t-1)^2 \end{bmatrix}^{-1} \times$$

$$\times \begin{bmatrix} \sum_{t=1}^{N} u(t)y(t) \\ \sum_{t=1}^{N} u(t-1)y(t) \end{bmatrix}$$

Solution (Con'd):

Suppose that $u(t)$ is a unit step applied at $t = 1$

$$u(t) = \begin{cases} 1, & t = 1,\, 2,\, \ldots \\ 0, & t = 0,\, -1,\, -2,\, -3,\, \ldots \end{cases}$$

Solution (Con'd):

Suppose that $u(t)$ is a unit step applied at $t = 1$

$$u(t) = \begin{cases} 1, & t = 1, 2, \ldots \\ 0, & t = 0, -1, -2, -3, \ldots \end{cases}$$

Substituting this signal into the general formula, we obtain

$$\hat{\theta} = \begin{bmatrix} \sum_{t=1}^{N} u(t)^2 & \sum_{t=1}^{N} u(t)u(t-1) \\ \sum_{t=1}^{N} u(t)u(t-1) & \sum_{t=1}^{N} u(t-1)^2 \end{bmatrix}^{-1} \times$$

$$\times \begin{bmatrix} \sum_{t=1}^{N} u(t)y(t) \\ \sum_{t=1}^{N} u(t-1)y(t) \end{bmatrix}$$

Solution (Con'd):

Suppose that $u(t)$ is a unit step applied at $t = 1$

$$u(t) = \begin{cases} 1, & t = 1, 2, \ldots \\ 0, & t = 0, -1, -2, -3, \ldots \end{cases}$$

Substituting this signal into the general formula, we obtain

$$\hat{\theta} = \begin{bmatrix} N & \sum_{t=1}^{N} u(t)u(t-1) \\ \sum_{t=1}^{N} u(t)u(t-1) & \sum_{t=1}^{N} u(t-1)^2 \end{bmatrix}^{-1} \times$$

$$\times \begin{bmatrix} \sum_{t=1}^{N} u(t)y(t) \\ \sum_{t=1}^{N} u(t-1)y(t) \end{bmatrix}$$

Solution (Con'd):

Suppose that $u(t)$ is a unit step applied at $t = 1$

$$u(t) = \begin{cases} 1, & t = 1, 2, \ldots \\ 0, & t = 0, -1, -2, -3, \ldots \end{cases}$$

Substituting this signal into the general formula, we obtain

$$\hat{\theta} = \begin{bmatrix} N & N-1 \\ N-1 & \sum_{t=1}^{N} u(t-1)^2 \end{bmatrix}^{-1} \times$$

$$\times \begin{bmatrix} \sum_{t=1}^{N} u(t)y(t) \\ \sum_{t=1}^{N} u(t-1)y(t) \end{bmatrix}$$

Solution (Con'd):

Suppose that $u(t)$ is a unit step applied at $t = 1$

$$u(t) = \begin{cases} 1, & t = 1, 2, \ldots \\ 0, & t = 0, -1, -2, -3, \ldots \end{cases}$$

Substituting this signal into the general formula, we obtain

$$\hat{\theta} = \begin{bmatrix} N & N-1 \\ N-1 & N-1 \end{bmatrix}^{-1} \times$$

$$\times \begin{bmatrix} \sum_{t=1}^{N} u(t)y(t) \\ \sum_{t=1}^{N} u(t-1)y(t) \end{bmatrix}$$

Solution (Con'd):

Suppose that $u(t)$ is a unit step applied at $t = 1$

$$u(t) = \begin{cases} 1, & t = 1,\, 2,\, \ldots \\ 0, & t = 0,\, -1,\, -2,\, -3,\, \ldots \end{cases}$$

Substituting this signal into the general formula, we obtain

$$\hat{\theta} = \begin{bmatrix} N & N-1 \\ N-1 & N-1 \end{bmatrix}^{-1} \times$$

$$\times \begin{bmatrix} \sum_{t=1}^{N} y(t) \\ \sum_{t=2}^{N} y(t) \end{bmatrix}$$

Solution (Con'd):

Suppose that $u(t)$ is a unit step applied at $t = 1$

$$u(t) = \begin{cases} 1, & t = 1, 2, \ldots \\ 0, & t = 0, -1, -2, -3, \ldots \end{cases}$$

Substituting this signal into the general formula, we obtain

$$\hat{\theta} = \frac{1}{N(N-1) - (N-1)^2} \begin{bmatrix} N-1 & -(N-1) \\ -(N-1) & N \end{bmatrix} \times$$

$$\times \begin{bmatrix} \sum_{t=1}^{N} y(t) \\ \sum_{t=2}^{N} y(t) \end{bmatrix}$$

Solution (Con'd):

Suppose that $u(t)$ is a unit step applied at $t = 1$

$$u(t) = \begin{cases} 1, & t = 1, 2, \ldots \\ 0, & t = 0, -1, -2, -3, \ldots \end{cases}$$

Substituting this signal into the general formula, we obtain

$$\hat{\theta} = \begin{bmatrix} 1 & -1 \\ -1 & \dfrac{N}{N-1} \end{bmatrix} \times$$

$$\times \begin{bmatrix} \sum_{t=1}^{N} y(t) \\ \sum_{t=2}^{N} y(t) \end{bmatrix}$$

Solution (Con'd):

Suppose that $u(t)$ is a unit step applied at $t = 1$

$$u(t) = \begin{cases} 1, & t = 1,\, 2,\, \ldots \\ 0, & t = 0,\, -1,\, -2,\, -3,\, \ldots \end{cases}$$

Substituting this signal into the general formula, we obtain

$$\begin{aligned} \hat{\theta} &= \begin{bmatrix} \sum_{t=1}^{N} y(t) - \sum_{t=2}^{N} y(t) \\ -\sum_{t=1}^{N} y(t) + \dfrac{N}{N-1} \displaystyle\sum_{t=2}^{N} y(t) \end{bmatrix} \\ &= \begin{bmatrix} y(1) \\ \dfrac{1}{N-1} \displaystyle\sum_{t=1}^{N} y(t) - y(1) \end{bmatrix} \end{aligned}$$

Solution (Con'd):

How to compute the covariance of θ? Consider

$$\theta - \theta^0 = \begin{bmatrix} y(1) \\ \frac{1}{N-1}\sum_{t=1}^{N} y(t) - y(1) \end{bmatrix} - \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

Solution (Con'd):

How to compute the covariance of θ? Consider

$$\theta - \theta^0 = \begin{bmatrix} y(1) \\ \frac{1}{N-1}\sum_{t=1}^{N} y(t) - y(1) \end{bmatrix} - \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

$$= \begin{bmatrix} \left\{ b_0 \cdot 1 + b_1 \cdot 0 + e(1) \right\} - b_0 \\ \dfrac{\sum_{t=2}^{N} \left\{ b_0 \cdot 1 + b_1 \cdot 1 + e(t) \right\} + y(1)}{N-1} - y(1) - b_1 \end{bmatrix}$$

Solution (Con'd):

How to compute the covariance of $\boldsymbol{\theta}$? Consider

$$\boldsymbol{\theta} - \theta^0 = \begin{bmatrix} y(1) \\ \dfrac{1}{N-1}\sum_{t=1}^{N} y(t) - y(1) \end{bmatrix} - \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

$$= \begin{bmatrix} e(1) \\ \dfrac{1}{N-1}\sum_{t=2}^{N} e(t) - e(1) \end{bmatrix}$$

Solution (Con'd):

The estimation error is

$$\theta - \theta^0 = \begin{bmatrix} e(1) \\ \frac{1}{N-1}\sum_{t=2}^{N} e(t) - e(1) \end{bmatrix}$$

The covariance $E(\theta - \theta^0)(\theta - \theta^0)^T$ of estimate is then

$$\text{cov}(\theta) = \begin{bmatrix} Ee(1)^2 & E\left\{\frac{\sum_{t=2}^{N} e(t)}{N-1} - e(1)\right\} e(1) \\ E\left\{\frac{\sum_{t=2}^{N} e(t)}{N-1} - e(1)\right\} e(1) & E\left\{\frac{\sum_{t=2}^{N} e(t)}{N-1} - e(1)\right\}^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & -\sigma^2 \\ -\sigma^2 & \sigma^2\left(\frac{N-1}{(N-1)^2} - 1\right) \end{bmatrix}$$

Solution (Con'd):

The estimation error is

$$\theta - \theta^0 = \begin{bmatrix} e(1) \\ \frac{1}{N-1}\sum_{t=2}^{N} e(t) - e(1) \end{bmatrix}$$

The covariance $E(\theta - \theta^0)(\theta - \theta^0)^T$ of estimate is then

$$\text{cov}(\theta) = \begin{bmatrix} Ee(1)^2 & E\left\{\frac{\sum_{t=2}^{N} e(t)}{N-1} - e(1)\right\} e(1) \\ E\left\{\frac{\sum_{t=2}^{N} e(t)}{N-1} - e(1)\right\} e(1) & E\left\{\frac{\sum_{t=2}^{N} e(t)}{N-1} - e(1)\right\}^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & -\sigma^2 \\ -\sigma^2 & \sigma^2 \frac{N}{N-1} \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & -1 \\ -1 & \frac{N}{N-1} \end{bmatrix}$$

Solution (Con'd):

To conclude with the unit step input signal LS estimate is

$$\hat{\theta} = \begin{bmatrix} b_0 \\ \\ b_1 \end{bmatrix} = \begin{bmatrix} y(1) \\ \\ \dfrac{1}{N-1} \displaystyle\sum_{t=1}^{N} y(t) - y(1) \end{bmatrix}$$

Solution (Con'd):

To conclude with the unit step input signal LS estimate is

$$\hat{\theta} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} y(1) \\ \dfrac{1}{N-1}\displaystyle\sum_{t=1}^{N} y(t) - y(1) \end{bmatrix}$$

The covariance of such estimate looks as

$$E(\theta - \theta^0)(\theta - \theta^0)^T = \sigma^2 \begin{bmatrix} 1 & -1 \\ -1 & \frac{N}{N-1} \end{bmatrix}$$

Solution (Con'd):

To conclude with the unit step input signal LS estimate is

$$\hat{\theta} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} y(1) \\ \frac{1}{N-1}\sum_{t=1}^{N} y(t) - y(1) \end{bmatrix}$$

The covariance of such estimate looks as

$$E(\theta - \theta^0)(\theta - \theta^0)^T = \sigma^2 \begin{bmatrix} 1 & -1 \\ -1 & \frac{N}{N-1} \end{bmatrix}$$

As a number of measured data increases $N \to \infty$

$$E(\theta - \theta^0)(\theta - \theta^0)^T \to \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

and does NOT IMPROVE!

Solution (Problem 2.2):

Consider now the model

$$y(t) = b_0 u(t) + b_1 u(t-1) + e(t)$$

when the input signal is a white noise with unit variance

$$E u(t)^2 = 1, \quad E u(t) u(s) = 0 \text{ if } t \neq s$$

and when u and e are independent

$$E u(t) e(s) = 0 \quad \forall\, t,\ s$$

Solution (Problem 2.2):

Consider now the model

$$y(t) = b_0 u(t) + b_1 u(t-1) + e(t)$$

when the input signal is a white noise with unit variance

$$Eu(t)^2 = 1, \quad Eu(t)u(s) = 0 \text{ if } t \neq s$$

and when u and e are independent

$$Eu(t)e(s) = 0 \quad \forall\, t,\ s$$

Such assumptions imply that

$$Ey(t)u(t) = b_0, \quad Ey(t)u(t-1) = b_1$$

Solution (Problem 2.2):

The LS estimate becomes dependent on a realization of the stochastic variable $u(t)$; we need to consider its mean value

$$
\begin{aligned}
E\hat{\theta} \;&=\; E\left\{ \begin{bmatrix} \sum_{t=1}^{N} u(t)^2 & \sum_{t=2}^{N} u(t)u(t-1) \\ \sum_{t=2}^{N} u(t)u(t-1) & \sum_{t=2}^{N} u(t-1)^2 \end{bmatrix}^{-1} \times \right. \\
&\qquad\qquad \left. \times \begin{bmatrix} \sum_{t=1}^{N} u(t)y(t) \\ \sum_{t=2}^{N} u(t-1)y(t) \end{bmatrix} \right\}
\end{aligned}
$$

Solution (Problem 2.2):

The LS estimate becomes dependent on a realization of the stochastic variable $u(t)$; we need to consider its mean value

$$E\hat{\theta} = E\left\{\begin{bmatrix} N\left(\frac{1}{N}\sum_{t=1}^{N} u(t)^2\right) & (N-1)\left(\frac{1}{N-1}\sum_{t=2}^{N} u(t)u(t-1)\right) \\ (N-1)\left(\frac{1}{N-1}\sum_{t=2}^{N} u(t)u(t-1)\right) & (N-1)\left(\frac{1}{N-1}\sum_{t=2}^{N} u(t-1)^2\right) \end{bmatrix}^{-1} \times \right.$$

$$\left. \times \begin{bmatrix} N\left(\frac{1}{N}\sum_{t=1}^{N} u(t)y(t)\right) \\ (N-1)\left(\frac{1}{N-1}\sum_{t=2}^{N} u(t-1)y(t)\right) \end{bmatrix}\right\}$$

Solution (Problem 2.2):

The LS estimate becomes dependent on a realization of the stochastic variable $u(t)$; we need to consider its mean value

$$E\hat{\theta} \approx \begin{bmatrix} N\,Eu(t)^2 & (N-1)\,Eu(t)u(t-1) \\ (N-1)\,Eu(t)u(t-1) & (N-1)\,Eu(t-1)^2 \end{bmatrix}^{-1} \times$$

$$\times \begin{bmatrix} N\,Eu(t)y(t) \\ (N-1)\,Eu(t-1)y(t) \end{bmatrix}$$

$$= \begin{bmatrix} N & 0 \\ 0 & (N-1) \end{bmatrix}^{-1} \begin{bmatrix} Nb_0 \\ (N-1)b_1 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

Solution (Problem 2.2):

To compute the covariance, use the formula in Theorem

$$
\begin{aligned}
\mathrm{cov}(\theta - \theta^0) &= \sigma^2 E\left(\Phi(N)^T \Phi(N)\right)^{-1} \\
&= \sigma^2 \begin{bmatrix} N & 0 \\ 0 & (N-1) \end{bmatrix}^{-1} \\
&= \sigma \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & \frac{1}{N-1} \end{bmatrix}
\end{aligned}
$$

Solution (Problem 2.2):

To compute the covariance, use the formula in Theorem

$$\begin{aligned}\mathrm{cov}(\boldsymbol{\theta} - \theta^0) &= \boldsymbol{\sigma^2 E \left(\Phi(N)^T \Phi(N)\right)^{-1}} \\ &= \boldsymbol{\sigma^2 \begin{bmatrix} N & 0 \\ 0 & (N-1) \end{bmatrix}^{-1}} \\ &= \boldsymbol{\sigma \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & \frac{1}{N-1} \end{bmatrix}}\end{aligned}$$

It converges to zero!

Next Lecture / Assignments:

- Next lecture (April 13, 10:00-12:00, in A206Tekn): Recursive Least Squares, modifications, continuous-time systems.

Next Lecture / Assignments:

- Next lecture (April 13, 10:00-12:00, in A206Tekn): Recursive Least Squares, modifications, continuous-time systems.
- **2.5** Consider the discrete-time system

  $$y(t+1) + ay(t) = bu(t) + e(t+1)$$

  where the input signal u and the noise e are sequences of independent random variables with zero mean values and standard deviation σ and 1. Determine the covariance of the estimates obtained for large observation sets.

  2.6 Consider data generated by the least-squares model

  $$y(t+1) + ay(t) = bu(t) + e(t+1) + ce(t) \qquad t = 1, 2, \ldots$$

  where $\{u(t)\}$ and $\{e(t)\}$ are sequences of independent random variables with zero mean values and standard deviations 1 and σ. Assume that parameters a and b of the model

  $$y(t+1) + ay(t) = bu(t)$$

  are estimated by least squares. Determine the asymptotic values of the estimates.