Modeling of dynamic systems

PLANT

$\Downarrow$

$\Downarrow$ Abstraction

$\Downarrow$

SYSTEM

$\Downarrow$

$\Downarrow$ Simplification

$\Downarrow$

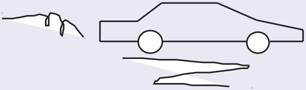
MATHEMATICAL MODEL

$\swarrow$ $\qquad$ $\searrow$

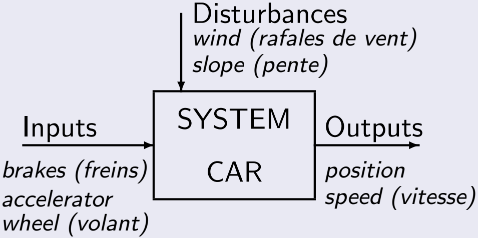
Identified model	Physical model
model for control	model for simulation

Example : driving a car

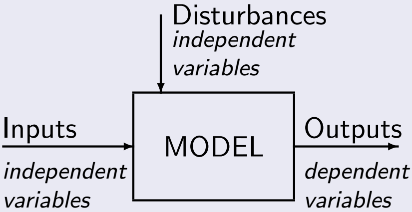
PLANT



SYSTEM



MODEL



Why Do We Need a Model ?

- **For Simulation :** Analyze the system's output for a given input.
  Example : Thermal analysis of a space shuttle during atmospheric re-entry.
- **For Design :** Determine system parameters to achieve a desired output for a given input.
  Example : Design of electrical, mechanical, or chemical installations.
- **For Prediction :** Forecast future values of the system's output.
  Example : Weather forecasting ; flood prediction.
- **For Control :** Model-based controller design.
  Example : Pole placement controller design for tracking and disturbance rejection.

How can we find a model ?

First principle modeling

Based on physical laws, physical models

$$G(s) = \frac{K}{s(\tau s + 1)}$$ (continuous-time models, pedagogical interest)

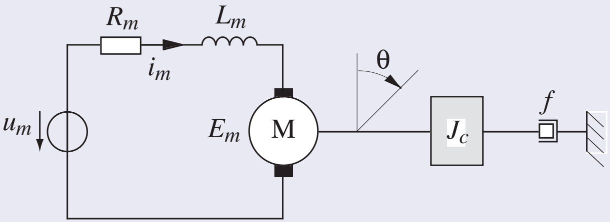
System identification

Based on input/output measured data

$$G(z) = \frac{bz}{z^2 + a_1 z + a_2}$$ (discrete-time models, practical interest)

How can we find a model ?

Physical modeling



$$E_m = K_m \dot{\theta}$$
$$M_m = K_m i_m(t)$$

Assumptions : only viscous friction, backlash and self inductance L_m neglected

$$(J_m + J_c)\ddot{\theta}(t) + f\dot{\theta}(t) = K_m i_m(t) \qquad u_m(t) = R_m i_m(t) + K_m \dot{\theta}$$

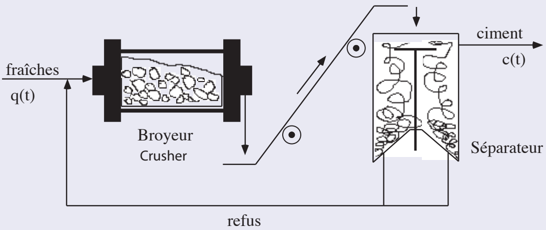
$$\left[(J_m + J_c)R_m s^2 + (fR_m + K_m^2)s\right]\Theta(s) = K_m U_m(s)$$

$$\frac{\Theta(s)}{U_m(s)} = \frac{K_m}{(J_m + J_c)R_m s^2 + (fR_m + K_m^2)s} \quad \text{model : } G(s) = \frac{K}{s(\tau s + 1)}$$

How can we find a model ?

System identification

Cement Plant : Crusher & Separator Loop



Physical model ?

Identification : Apply a specific input $q(t)$ and measure the output $c(t)$ and represent it as a function of preceding values of input and output.

$$c(t) = G[q(t), q(t-1), q(t-2), \ldots, c(t-1), c(t-2), \ldots]$$

$$G(z) = \frac{bz}{z^2 + a_1 z + a_2}$$

Physical or identified model ?

1) Physical models

$$G(s) = \frac{K_m}{(J_m + J_c)R_m s^2 + (fR_m + K_m^2)s}$$

- $\smile$ Direct relation with physical parameters
- $\frown$ High order, approximative, need complete process knowledge, physical parameters should be known

2) Identified models

$$G(z) = \frac{bz}{z^2 + a_1 z + a_2}$$

- $\smile$ Appropriate for controller design, simple and efficient
- $\frown$ Limited validity (operating point, type of input), sensors, measurement noise, unknown model structure

Types of models and representations

Models :

- dynamic/static
- monovariable/multivariable
- deterministic/stochastic
- linear/nonlinear
- time-invariant /time-variant

Representations :

- parametric/nonparametric
- continuous-time/discrete-time
- time-domain/frequency-domain
- input-output/state-space

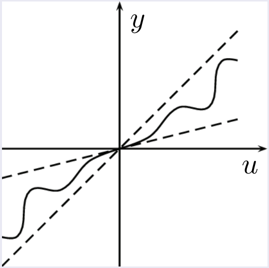
Types of models (Static/Dynamic)

Static Model

The output of the system at instant t depends only on the input at instant t and system parameters at instant t.

$$y(t) = H[u(t)]$$

Example : $y(t) = Ku(t)$ or static nonlinearity :



Dynamic Model

The output of the system at instant t depends on the input at instant t and its past values : $y(t) = H[u(\tau)] \qquad 0 < \tau \leq t$

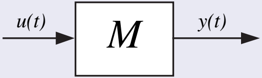
Example : All systems that are presented by differential equations :

$$\dot{y}(t) + ay(t) = Ku(t)$$

Types of models (Monovariable, Multivariable)

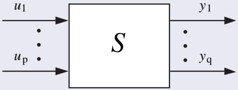
Monovariable Model

Models with a single input and a single output are called SISO models or monovariable models.



Multivariable Model

Systems with more than one input and one output are called multivariable systems or MIMO systems.



Example : Consider a state space model with $n > 2$ state variables :

$$\dot{x}(t) = Ax(t) + Bu(t) \quad , \quad y(t) = Cx(t)$$

with $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times 1}, C \in \mathbb{R}^{1 \times n}$.
Is this model monovariable or multivariable ?

Types of models (Deterministic/Stochastic)

Deterministic Model

The output of a deterministic model can be *exactly* computed based on the input signal and the model parameters. For example :

$$y(k) = -ay(k-1) + bu(k-1)$$

Stochastic Model

The output of a stochastic model contains random terms such that it cannot be *exactly* computed. The random terms are usually described by a random disturbance at the output of the system. For example :

$$y(k) = -ay(k-1) + bu(k-1) + e(k)$$

where $e(k)$ is a random process.

Types of models (Linear/Nonlinear)

Linear Model

A model is considered linear if its output is linearly dependent on its inputs, or in other words, if it adheres to the superposition principle, characterized by two properties : additivity and homogeneity. For example :

$$y(k) = -ay(k-1) + bu(k-1)$$

Nonlinear Model

A system is nonlinear if, and only if, it does not obey the superposition principle. For example :

$$y(k) = -ay(k-1) + bu(k-1)y(k-2)$$

Superposition Principle : Assume that the output of a system when excited by $u_1(t)$ and $u_2(t)$ is $y_1(t)$ and $y_2(t)$, respectively. Then, the superposition principle is valid if and only if the output of the system for $\alpha_1 u_1(t) + \alpha_2 u_2(t)$ is equal to $\alpha_1 y_1(t) + \alpha_2 y_2(t)$ for any constant $\alpha_1, \alpha_2 \in \mathbb{R}$.

Types of models (Time-Invariant/Time-Varying)

Time Invariant Model

In time-invariant models with $u(t)$ as input and $y(t)$ as output, the output of the system to $u(t-\tau)$ will be $y(t-\tau)$. In other words, the model parameters are constant. For example :

$$\dot{y}(t) + ay(t) = bu(t)$$

Time-Varying Model

In time-varying models, the model parameters change with time. For example :

$$\dot{y}(t) + a(t)y(t) = b(t)u(t)$$

Remark : In all systems, the input $u(t)$, output $y(t)$, and state variables $x(t)$ vary over time. However, the model parameters may either be time-varying or constant. In a mass-spring-damper system, the force, position and velocity are time-varying variables whereas the mass, spring constant and friction coefficient are model parameters.

Types of representations (Parametric/Nonparametric)

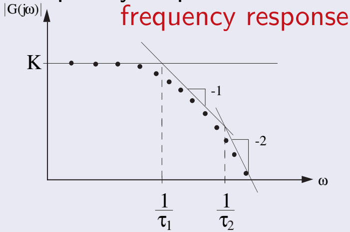
Parametric Model

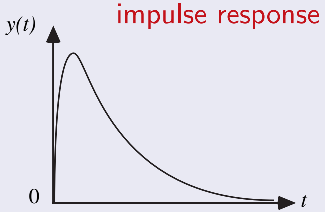
A parametric model is described with a structure and a set of parameters. For example :

$$\dot{y}(t) + ay(t) = bu(t) \quad , \quad G(s) = \frac{K}{\tau s + 1}$$

Nonparametric Model

A nonparametric model is described by a graph like the step response or the frequency response.





Remark : In practice, the nonparametric models can be viewed as a parametric model with a large number of parameters.

Types of representations (Continuous-time/Discrete-time)

Continuous-time Model

A continuous-time model describes the relation between inputs and outputs at all t. For example :

$$\dot{y}(t) + ay^2(t) = bu(t)$$

For LTI systems $y(t) = g(t) * u(t)$, where $g(t)$ is the impulse response of the system, is a continuous-time model that can also be represented by its Laplace transform : $Y(s) = G(s)U(s)$.

Discrete-time Model

A discrete-time model describes the relation between inputs and outputs at discrete time points k. For example :

$$y(k) + ay^2(k-1) = bu(k-1)$$

For LTI systems $y(k) = g(k) * u(k)$, where $g(k)$ is the impulse response of the system, is a discrete-time model that can also be represented by its z-transform : $Y(z) = G(z)U(z)$.

Types of representations (Time-domain/Frequency-domain)

Time-domain Model

A time-domain model is presented by a set of difference or differential equations. It can be parametric and represented by a state-space model or nonparametric and represented by a time-domain graph like the impulse response. The nonlinear systems are usually represented in time domain.

Frequency-domain Model

A frequency-domain model is usually obtained from the transfer function of a linear model in a parametric representation $G(s)$ or is given by a graph in Bode diagram in a nonparametric representation $G(j\omega)$.

Types of representations (Input-Output/State-Space)

Input-Output Model

The inputs and outputs of the system are the only variables in the model. For example :

$$\dot{y}(t) + ay^2(t) = bu(t) \qquad , \qquad Y(s) = G(s)U(s)$$

State-Space Model

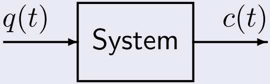
Other internal variables are involved in the mathematical model that are called the *states* of the system. For example :

$$\begin{aligned}
\dot{x}_1(t) &= 2x_1^2(t)x_2(t) + x_1(t)u(t) \\
\dot{x}_2(t) &= 2x_1(t) - 3x_1(t)x_2(t) + u(t) \\
y(t) &= 5x_1(t)x_2^2(t)u(t)
\end{aligned}$$

or for LTI systems :

$$\dot{x}(t) = Ax(t) + Bu(t) \quad , \quad y(t) = Cx(t) + Du(t)$$

Example : Cement mill (*Moulin à ciment*)



$$\dot{c}(t) + 2c^2(t) = 3q(t-\theta) \qquad c(0) = c_0$$

System : dynamic, monovariable, deterministic, parametric, nonlinear, time-invariant, continuous-time, time-domain.

dynamic/statique ? monovariable/multivariable ? deterministic/stochastic ? parametric/nonparametric ? linear/nonlinear ?
time-variant/time-invariant ? continuous-time/discrete-time ?
time-domain/frequency-domain ? Linearization ?

Linearization

Taylor Series : $f(x) \approx f(x_0) + f'(x_0)(x - x_0)$

$$c^2(t) = c_0^2 + 2c_0(c(t) - c_0), \qquad \Delta c = c(t) - c_0, \qquad \Delta q(t) = q(t) - q_0$$

$$\dot{\Delta c} + 2c_0^2 + 4c_0\Delta c = 3q(t-\theta) = 3\Delta q(t-\theta) + 3q_0$$

Input/output models

Time-domain continuous-time representation

$$y(t) = H[u(\tau)] \quad -\infty < \tau \leq t \qquad H : \text{a causal linear operator}$$

For an LTI system, define the impulse response : $g(t) \equiv H[\delta(t)]$

Convolution integral :

$$y(t) = \int_0^{\infty} g(t-\tau)u(\tau)d\tau = g(t) * u(t) = u(t) * g(t)$$

Time-domain discrete-time representation

For a discrete-time LTI system, define the impulse response $g(k)$ as the response to the Kronecker delta : $\delta(k) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$

Convolution sum :

$$y(k) = \sum_{j=0}^{k} g(k-j)u(j) = g(k) * u(k)$$

Input/output models

Frequency-domain representation

For continuous-time LTI systems :

$$y(t) = g(t) * u(t) \Rightarrow Y(s) = G(s)U(s) \Rightarrow Y(j\omega) = G(j\omega)U(j\omega)$$

Frequency-domain representation

For discrete-time LTI systems :

$$y(k) = g(k) * u(k) \Rightarrow Y(z) = G(z)U(z) \Rightarrow Y(e^{j\omega}) = G(e^{j\omega})U(e^{j\omega})$$

State-space models

Definition

State of a dynamic system at t_0 is the minimum necessary information such that knowing $u(t)$ for $t \geq t_0$ the system output can be determined uniquely. In other words :

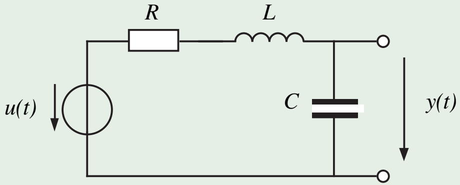
- States summarize the past of a system and are required to determine its future.
- States of a system of differential equations are its initial conditions.

Example

Consider an RLC circuit

$$u_L = L\frac{di_L(t)}{dt}$$
$$i_c = C\frac{du_C}{dt}$$

States : $-$inductor current $i_L(t)$
$-$capacitor voltage $u_C(t)$



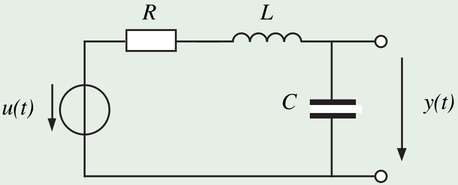
State-space models

Example

RLC circuit

$u_R(t) + u_L(t) + u_C(t) = u(t)$
$i_L(t) = i_C(t)$
$u_R(t) = R i_L(t)$

$u_L(t) = L\dfrac{di_L(t)}{dt} \qquad i_C = C\dfrac{du_C}{dt}$



$$\begin{matrix} x_1(t) \equiv i_L(t) \\ x_2(t) \equiv u_C(t) \end{matrix} \Rightarrow \begin{matrix} R x_1(t) + L\dot{x}_1(t) + x_2(t) = u(t) \\ x_1(t) = C\dot{x}_2(t) \end{matrix}$$

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} -R/L & -1/L \\ 1/C & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 1/L \\ 0 \end{bmatrix} u(t)$$

$$y(t) = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

State-space models

Continuous-time nonlinear model

$$\dot{x}(t) = f[x(t), u(t), t] \quad \text{state equation} \qquad x(t_0) = x_0$$
$$y(t) = g[x(t), u(t), t] \quad \text{output equation}$$

Continuous-time linear time-varying model

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \qquad x(t_0) = x_0$$
$$y(t) = C(t)x(t) + D(t)u(t)$$

Continuous-time linear time-invariant model

$$\dot{x}(t) = Ax(t) + Bu(t) \qquad x(t_0) = x_0$$
$$y(t) = Cx(t) + Du(t)$$

Discrete-time linear time-invariant model

$$x(k+1) = Ax(k) + Bu(k) \qquad x(k_0) = x_0$$
$$y(k) = Cx(k) + Du(k)$$

State-space models

Discrete-time nonlinear model

$$x(k+1) = f[x(k), u(k), k] \quad \text{state equation} \qquad x(k_0) = x_0$$
$$y(k) = g[x(k), u(k), k] \quad \text{output equation}$$

Frequency-domain model

continuous-time

$$sX(s) = AX(s) + BU(s)$$
$$Y(s) = CX(s) + DU(s)$$

discrete-time

$$zX(z) = AX(z) + BU(z)$$
$$Y(z) = CX(z) + DU(z)$$

From state-space to input-output model

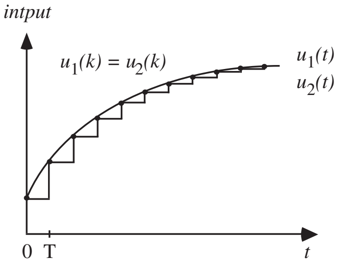
continuous-time

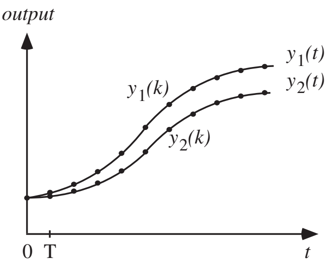
$$Y(s) = [C(sI - A)^{-1}B + D]U(s) \Rightarrow G(s) = C(sI - A)^{-1}B + D$$

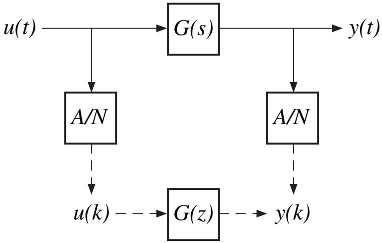
Discrete-time

$$Y(z) = [C(zI - A)^{-1}B + D]U(z) \Rightarrow G(z) = C(zI - A)^{-1}B + D$$

Relation between $G(s)$ and $G(z)$







$$u_1(k) = u_2(k) \quad \Rightarrow \quad y_1(k) \neq y_2(k)$$

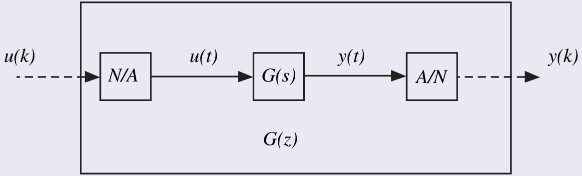
$$G_1(z) = \frac{Y_1(z)}{U_1(z)} \quad \neq \quad G_2(z) = \frac{Y_2(z)}{U_2(z)}$$

For a given $G(s)$, $G(z)$ depends on input $u(t)$

Find $G(z)$ from $G(s)$

Conversion with Zero-Order-Hold

Input is a linear combination of steps.



$$G(z) = \frac{\mathcal{Z}\{\mathcal{L}^{-1}[G(s)/s]|_{kT}\}}{\frac{1}{1-z^{-1}}} = (1 - z^{-1})\mathcal{Z}\{\mathcal{L}^{-1}[G(s)/s]|_{kT}\}$$

$$G(s) \to \frac{G(s)}{s} \xrightarrow{\mathcal{L}^{-1}} \gamma(t) \xrightarrow{\text{sampling}} \gamma(k) \xrightarrow{\mathcal{Z}} \frac{G(z)}{1 - z^{-1}} \xrightarrow{(1-z^{-1})} G(z)$$

Find $G(z)$ from $G(s)$

Example

Find $G(z)$ from $G(s) = \dfrac{1}{s+1}$ $\quad T = 0.2$ by ZOH method.

$$\begin{aligned} G(z) &= (1 - z^{-1})\mathcal{Z}\left\{\mathcal{L}^{-1}\left[\frac{1}{s(s+1)}\right]\Big|_{kT}\right\} \\ &= (1 - z^{-1})\mathcal{Z}\left\{\mathcal{L}^{-1}\left[\frac{1}{s} - \frac{1}{(s+1)}\right]\Big|_{kT}\right\} \\ &= (1 - z^{-1})\left[\frac{1}{1 - z^{-1}} - \frac{1}{1 - e^{-T}z^{-1}}\right] \\ &= 1 - \frac{1 - z^{-1}}{1 - e^{-T}z^{-1}} = \frac{(1 - e^{-T})z^{-1}}{1 - e^{-T}z^{-1}} = \frac{0.18z^{-1}}{1 - 0.82z^{-1}} \end{aligned}$$

Find $G(z)$ from $G(s)$

Tustin method (*Trapezoid integration*)

Suppose that $G(s) = \dfrac{Y(s)}{U(s)} = \dfrac{1}{s}$ then $\dot{y}(t) = u(t)$.

$$\begin{cases} \frac{y(k+1)-y(k)}{T} = u(k) \\ \frac{y(k+1)-y(k)}{T} = u(k+1) \end{cases} \Rightarrow \frac{y(k+1)-y(k)}{T} = \frac{1}{2}[u(k)+u(k+1)]$$

$$\Rightarrow \frac{z-1}{T}Y(z) = \frac{1}{2}(1+z)U(z) \Rightarrow \frac{Y(z)}{U(z)} = \frac{T(z+1)}{2(z-1)} \Rightarrow \boxed{s : \frac{2}{T}\frac{z-1}{z+1}}$$

Example

Find $G(z)$ from $G(s) = \dfrac{1}{s+1}$ and $T = 0.2$ by trapezoid integration.

$$s : \frac{2}{T}\frac{z-1}{z+1} \Rightarrow G(z) = \frac{1}{\frac{2(z-1)}{T(z+1)} + 1} = \frac{Tz+T}{2z-2+Tz+T} = \frac{0.2+0.2z^{-1}}{2.2-1.8z^{-1}}$$

Summary of models, representations and identification

- By some abstraction and simplification, physical systems can be represented by mathematical models.
- There are several types of models (e.g. linear/nonlinear, SISO/MIMO, etc.)
- For each type of model there are several representation models (parametric/nonparametric, input-output/state-space, discrete-time/continuous-time, etc.)
- There are some methods to get one representation from another.
- System identification is the art of fitting a model to the data.
- System identification is an experimental method so the design of experiment (input design) is very important.

The method presented in this course are applicable to all types of models, however, discrete-time LTI models are emphasized.

Input Signals

- Quality of identified models depends on the quality of the excitation signal (richness) which is related to its spectrum (power spectral density function).

Power Spectral Density

For a discrete-time signal $u(k)$ is defined as :

$$\Phi_{uu}(\omega) = \mathcal{F}(R_{uu}(h)) = \sum_{h=-\infty}^{\infty} R_{uu}(h) e^{-j\omega h}$$

or

$$\Phi_{uu}(\omega) = |\mathcal{F}(u(k))|^2 = |U(e^{j\omega})|^2$$

where $R_{uu}(h)$ is the **autocorrelation function** of $u(k)$ and $U(e^{j\omega})$ is the Fourier transform of $u(k)$

Richness of a signal

Degree of excitation

The number of non-zero values of $\Phi_{uu}(\omega)$ in the interval $[0\,,\,2\pi[$ is a measure of its richness and is related to the maximum number of parameters that can be estimated by the signal.

Example

Consider a discrete sinusoidal signal $u(k) = \sin \omega_0 k$. The power spectral density of this signal has only two non zero values at ω_0 and $-\omega_0$, so its degree of excitation is 2.
This signal applied to an LTI system $G(s)$ leads to the identification of $G(j\omega_0)$ (one complex value). Suppose that

$$G(s) = \frac{K}{\tau s + 1} \quad \Rightarrow \quad \Re[G(j\omega_0)] = \frac{K}{1 + \tau^2 \omega_0^2}, \Im[G(j\omega_0)] = \frac{-\tau\omega_0}{1 + \tau^2 \omega_0^2}$$

These two equations let us compute two parameters K and τ. If $G(s)$ had more than two parameters we could not identify them using just one sinusoidal signal.

Richness of a signal

Signal Energy :

The integral of the spectrum is equal to the energy of the signal :

$$\frac{1}{2\pi}\int_0^{2\pi} \Phi_{uu}(\omega)d\omega = \sum_{k=-\infty}^{\infty} u^2(k)$$

The above equality (Parseval's relation) can be proved easily by evaluating the inverse Fourier transform of the spectrum at $h = 0$:

$$R_{uu}(h) = \frac{1}{2\pi}\int_0^{2\pi} \Phi_{uu}(\omega)e^{j\omega h}d\omega$$

It shows that more signal energy leads to a better signal-to-noise ratio and a more accurate estimation of the parameters.

Spectrum Shape :

The shape of the spectrum shows in which frequencies better estimation will be achieved.

Autocorrelation Function

Energy signals : For bounded energy signals, i.e.

$$\lim_{N\to\infty}\sum_{k=-N}^{N}u^2(k)<\infty \quad \Rightarrow \quad R_{uu}(h)=\lim_{N\to\infty}\sum_{k=-N}^{N}u(k)u(k-h)$$

Power signals : For power bounded signals, i.e.

$$\frac{1}{2N+1}\lim_{N\to\infty}\sum_{k=-N}^{N}u^2(k)<\infty \Rightarrow R_{uu}(h)=\lim_{N\to\infty}\frac{1}{2N+1}\sum_{k=-N}^{N}u(k)u(k-h)$$

Periodic signals : For this class of power signals, we have :

$$u(k)=u(k+M) \quad \Rightarrow \quad R_{uu}(h)=\frac{1}{M}\sum_{k=0}^{M-1}u(k)u(k-h)$$

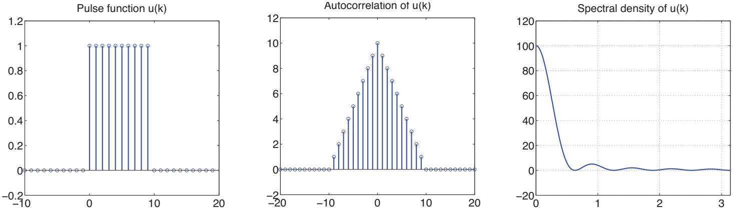
Step signal

The Fourier transform of a step signal is not defined. Therefore, its spectrum can be computed as the limit for a pulse signal ($M \to \infty$) :

$$u(k) = \begin{cases} \alpha & 0 \le k < M \\ 0 & k < 0 \text{ and } k \ge M \end{cases}$$

This is an energy signal so $R_{uu}(h)$ and $\Phi_{uu}(\omega)$ are defined as :

$$\begin{aligned} R_{uu}(h) &= \lim_{N \to \infty} \sum_{k=-N}^{N} u(k)u(k-h) = (M - |h|)\alpha^2 \\ \Phi_{uu}(\omega) &= \sum_{h=-\infty}^{\infty} R_{uu}(h) e^{-j\omega h} = \sum_{h=-M}^{M} (M - |h|)\alpha^2 e^{-j\omega h} \end{aligned}$$

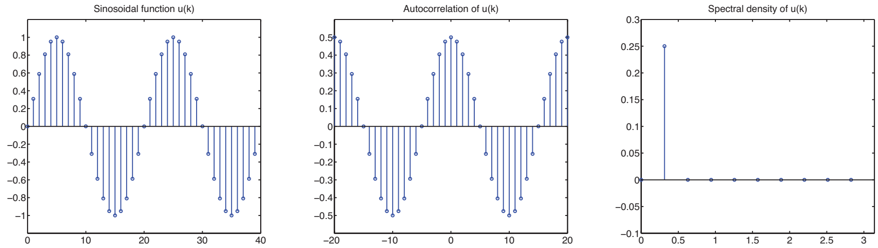


Sinusoid signal

For some periodic signals, it's easier to compute the spectrum using the Fourier series of the signal.

$$u(k) = a_1 \sin(\omega_1 k) = \frac{a_1}{2j} e^{j\omega_1 k} + \frac{-a_1}{2j} e^{-j\omega_1 k}$$

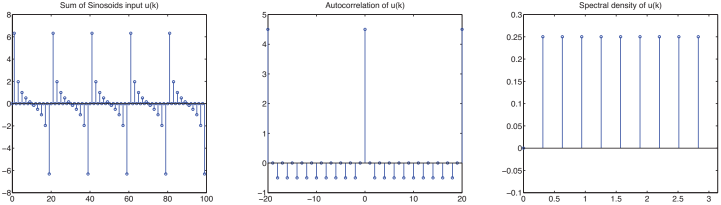
Therefore, $U(j\omega_n)$ has one component at ω_1 and another at $-\omega_1$ with an amplitude of $a_1/2$. So its power spectral density has one component with amplitude of $a_1^2/4$ for positive frequencies in the interval $[0\,,\,\pi[$.



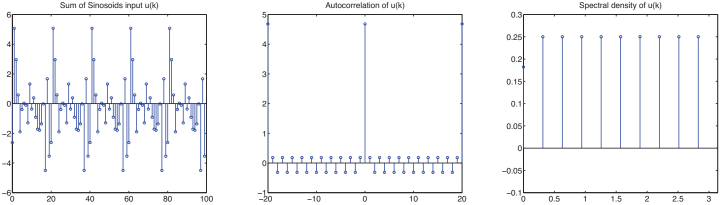
Sum of Sinusoids

By adding some sinusoids the input signal becomes richer.

$$u(k) = \sum_{i=1}^{m} a_i \sin(\omega_i k + \phi_i)$$



In practice, the phase is chosen as a random signal :

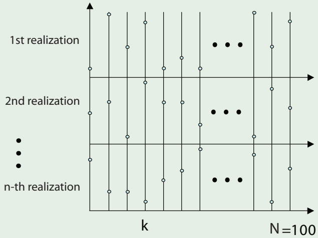


Discrete Random Process

Example (Throw a die N times)

$x(k)$: random variable
$x(k) \in \mathcal{D} = \{1, 2, 3, 4, 5, 6\}$

Expected value of $x(k)$:
$\mathbb{E}\{x(k)\} = \sum_{\mathcal{D}} x(k) P[x(k)] = 3.5$



For stationary processes, expected value can be computed from measurements :

$$\mathbb{E}\{x(k)\} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x(k,i) \qquad \text{or} \qquad \mathbb{E}\{x(k)\} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} x(k)$$

Autocorrelation function :

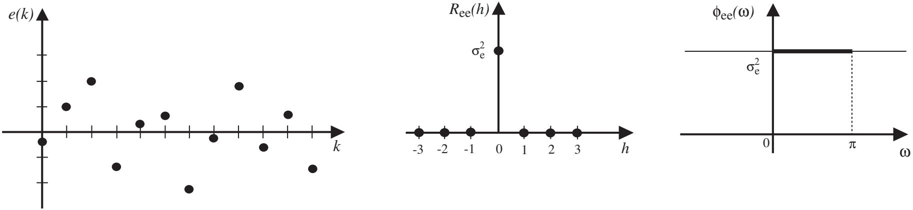
$$R_{xx}(h) = \mathbb{E}\{x(k)x(k-h)\} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} x(k)x(k-h)$$

Discrete-time zero-mean white noise

$$\mathbb{E}\{e(k)\} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} e(k) = 0$$

$$R_{ee}(h) = \mathbb{E}\{e(k)e(k-h)\} = 0 \quad \text{for } h \neq 0$$

$$\mathbb{E}\{e^2(k)\} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} e^2(k) = \sigma_e^2$$



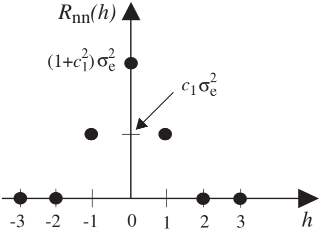
Spectral density is the Fourier Transform of the autocorrelation function :

$$\Phi_{ee}(\omega) = \mathcal{F}\{R_{ee}(h)\} = \sum_{h=-\infty}^{\infty} R_{ee}(h) e^{-j\omega h} = \sigma_e^2$$

Discrete-time zero-mean colored noise

$$n(k) = (1 + c_1 q^{-1})e(k) = C(q^{-1})e(k) \qquad \mathbb{E}\{e(k)\} = 0$$

$$\begin{aligned}
\mathbb{E}\{n(k)\} &= \mathbb{E}\{(1 + c_1 q^{-1})e(k)\} = \mathbb{E}\{e(k) + c_1 e(k-1)\} \\
&= \mathbb{E}\{e(k)\} + c_1 \mathbb{E}\{e(k-1)\} = 0 \\
R_{nn}(0) &= \mathbb{E}\{n(k)n(k)\} = \mathbb{E}\{[e(k) + c_1 e(k-1)]^2\} \\
&= \mathbb{E}\{e^2(k)\} + c_1^2 \mathbb{E}\{e^2(k-1)\} + 2c_1 \mathbb{E}\{e(k)e(k-1)\} \\
&= (1 + c_1^2)\sigma_e^2 \\
R_{nn}(1) &= \mathbb{E}\{n(k)n(k-1)\} = c_1 \sigma_e^2
\end{aligned}$$

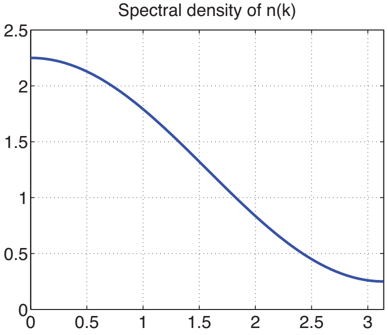


Discrete-time zero-mean colored noise

Power spectral density of colored noise :

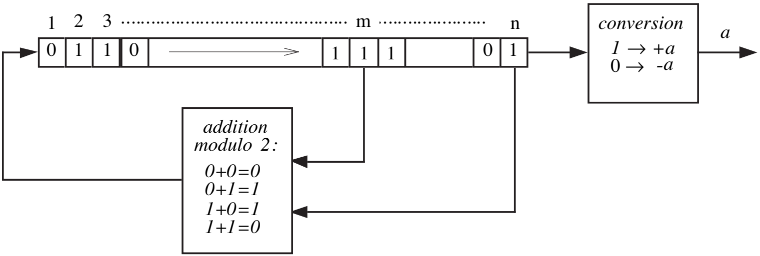
$$\Phi_{nn}(\omega) = \sum_{h=-\infty}^{\infty} R_{nn}(h)e^{-j\omega h} = c_1\sigma_e^2 e^{j\omega} + (1+c_1^2)\sigma_e^2 + c_1\sigma_e^2 e^{-j\omega}$$

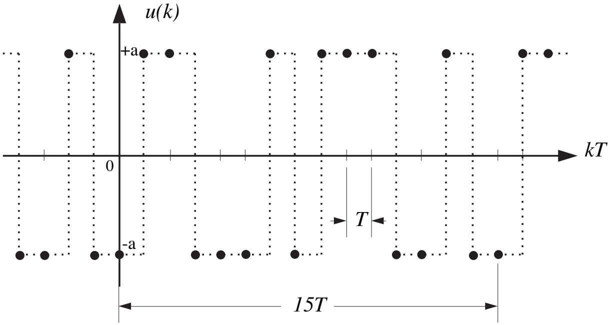
$$= |1 + c_1 e^{-j\omega}|^2 \sigma_e^2$$



In general, if $n(k) = F(q^{-1})e(k)$, then $\Phi_{nn}(\omega) = |F(e^{-j\omega})|^2\sigma_e^2$.

Pseudo Random Binary Sequence (PRBS)





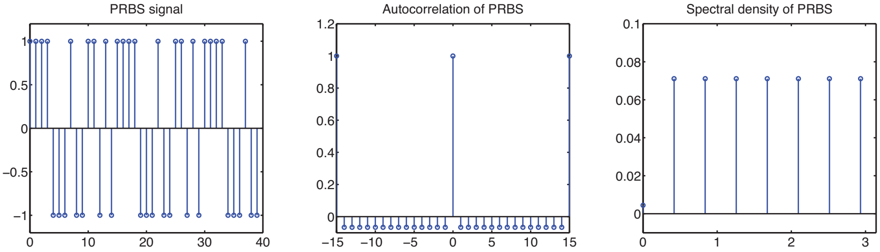
Pseudo Random Binary Sequence (PRBS)

Autocorrelation function : PRBS is periodic so :

$$R_{uu}(h) = \frac{1}{M}\sum_{k=0}^{M-1} u(k)u(k-h) = \begin{cases} a^2 & h = 0, \pm M, \pm 2M, \ldots \\ -a^2/M & \text{elsewhere} \end{cases}$$

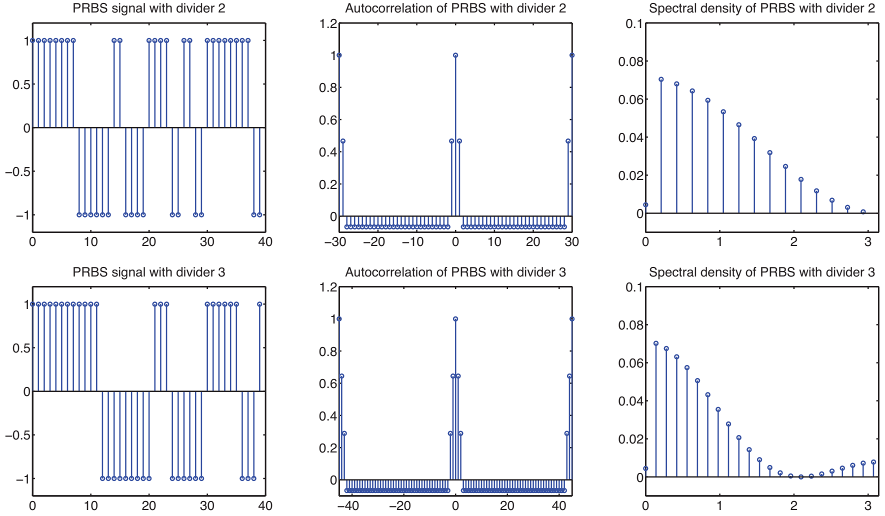
Power spectral density : Since autocorrelation function is periodic the spectrum is discrete :

$$\Phi_{uu}(\omega_n) = \frac{1}{M}\sum_{h=0}^{M-1} R_{uu}(h) e^{-j\frac{2n\pi}{M}h} = \begin{cases} a^2/M^2 & n = 0 \\ (M+1)a^2/M^2 & n = 1, \ldots, M-1 \end{cases}$$



Pseudo Random Binary Sequence (PRBS)

Frequency divider : The spectrum of a PRBS can be shaped by using a frequency divider at the input of the clock pulse to the shift register.



Summary on input signals

- The input signal should be sufficiently rich in the interesting frequency zone for identification.
- In general, good excitation around the system's bandwidth is required (one decade before and one decade after).
- The step signal excites very well the low-frequencies (degree of excitation is infinity because of its continuous spectrum).
- A random white noise has a continuous flat spectrum so excites all frequencies equivalently (degree of excitation is infinity). By a filtered white noise the spectrum can be shaped.
- By sum of sinusoids the degree of excitation and the shape of the spectrum can be easily designed (but the amplitude of the signal cannot be directly controlled). The signal is periodic and the spectrum discrete.
- A PRBS has a similar characteristic as white noise but is periodic and its spectrum discrete. The spectrum can be shaped by a frequency divider. The amplitude of signal can be directly controlled.

Least Squares Algorithm

Consider a measured signal $y(k)$ for $k = 1, \ldots, N$ and find a polynomial that fits the data such that :

$$y(k) = a_1 + a_2 k + a_3 k^2 + \cdots + a_n k^{n-1}$$

Let's define an error function as a *linear regression* :

$$\begin{aligned} \varepsilon(k, \theta) &= y(k) - (a_1 + a_2 k + a_3 k^2 + \cdots + a_n k^{n-1}) \\ &= y(k) - \phi^T(k)\theta \end{aligned}$$

where

$$\begin{aligned} \phi^T(k) &= [1 \quad k \quad k^2 \quad \ldots \quad k^{n-1}] \\ \theta^T &= [a_1 \quad a_2 \quad a_3 \quad \ldots \quad a_n] \end{aligned}$$

Minimize the sum of the square errors defined by the criterion :

$$J(\theta) = \sum_{k=1}^{N} \varepsilon^2(k, \theta)$$

Least Squares Algorithm

The criterion can be written in matrix form as :

$$\sum_{k=1}^{N} \varepsilon^2(k,\theta) = \mathcal{E}^T\mathcal{E}$$

where : $\mathcal{E}^T = [\varepsilon(1,\theta) \quad \varepsilon(2,\theta) \quad \ldots \quad \varepsilon(N,\theta)]$. The error vector can be written as :

$$\begin{bmatrix} \varepsilon(1,\theta) \\ \vdots \\ \varepsilon(N,\theta) \end{bmatrix} = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} - \begin{bmatrix} \phi^T(1) \\ \vdots \\ \phi^T(N) \end{bmatrix} \theta$$

or as : $\mathcal{E} = Y - \Phi\theta$. Then the criterion becomes :

$$J(\theta) = \mathcal{E}^T\mathcal{E} = [Y - \Phi\theta]^T[Y - \Phi\theta] = Y^TY - 2Y^T\Phi\theta + \theta^T\Phi^T\Phi\theta$$

The minimum is obtained by setting the gradient of $J(\theta)$ to zero :

$$J'(\theta) = -2\Phi^TY + 2\Phi^T\Phi\theta = 0$$

which leads to : $\hat{\theta} = (\Phi^T\Phi)^{-1}\Phi^TY$

Least Squares Algorithm

The LS solution $\hat{\theta} = (\Phi^T\Phi)^{-1}\Phi^T Y$ exists if the information matrix $\Phi^T\Phi$ is not singular or Φ is full rank. The solution can be written as :

$$\hat{\theta} = \left[\sum_{k=1}^{N} \phi(k)\phi^T(k)\right]^{-1} \left[\sum_{k=1}^{N} \phi(k)y(k)\right]$$

Weighted least squares :

If the errors in different instance have different importance a weighted error function can be defined : $\mathcal{E}_W \equiv W[Y - \Phi\theta]$ where W is a weighting matrix (usually diagonal). The criterion becomes :

$$J(\theta) = \mathcal{E}_W^T \mathcal{E}_W = \mathcal{E}^T W^T W \mathcal{E} = [Y - \Phi\theta]^T W^T W [Y - \Phi\theta]$$

and the parameter estimates :

$$\hat{\theta} = (\Phi^T W^T W \Phi)^{-1} \Phi^T W^T W Y$$

Least Squares Algorithm (stochastic case)

Suppose that the data are generated by the following model :

$$y(k) = \phi^T(k)\theta_0 + e(k)$$

where θ_0 is the vector of true parameters and $e(k)$ a zero-mean stationary random process. In a matrix form, we have :

$$Y = \Phi\theta_0 + E$$

where, $E^T = [e(1) \quad e(2) \quad \dots \quad e(N)]$.

The least squares estimate is also a random variable :

$$\hat{\theta} = (\Phi^T\Phi)^{-1}\Phi^T Y = (\Phi^T\Phi)^{-1}\Phi^T(\Phi\theta_0 + E)$$

its quality is expressed by its mean value and its variance.

Biasedness :

An estimator is called unbiased if

$$\mathbb{E}\{\hat{\theta}\} = \theta_0$$

Least Squares Algorithm (stochastic case)

Lemma

Assume that Φ is a deterministic matrix and $e(k)$ a zero-mean stationary signal with $\text{cov}(E) = \mathbb{E}\{EE^T\} = R$, then the least squares estimate is unbiased and

$$\text{cov}(\hat{\theta}) = (\Phi^T\Phi)^{-1}\Phi^T R\Phi(\Phi^T\Phi)^{-1}$$

Proof : The estimates are unbiased because :

$$\mathbb{E}\{\hat{\theta}\} = \theta_0 + (\Phi^T\Phi)^{-1}\Phi^T\mathbb{E}\{E\} = \theta_0$$

The covariance of the estimates is computed as :

$$\begin{aligned}
\text{cov}(\hat{\theta}) &= \mathbb{E}\{(\hat{\theta}-\theta_0)(\hat{\theta}-\theta_0)^T\} \\
&= \mathbb{E}\{[(\Phi^T\Phi)^{-1}\Phi^T E][(\Phi^T\Phi)^{-1}\Phi^T E]^T\} \\
&= (\Phi^T\Phi)^{-1}\Phi^T\mathbb{E}\{EE^T\}\Phi(\Phi^T\Phi)^{-1} = (\Phi^T\Phi)^{-1}\Phi^T R\Phi(\Phi^T\Phi)^{-1}
\end{aligned}$$

Remark : If $e(k)$ is white noise with variance σ^2, then $R = \sigma^2 I$ and :

$$\text{cov}(\hat{\theta}) = \sigma^2(\Phi^T\Phi)^{-1}$$

Least Squares Algorithm (stochastic case)

Lemma : Noise variance estimate

If $e(k)$ is white, an unbiased estimate of σ^2 can be obtained by :

$$\hat{\sigma^2} = \frac{J(\hat{\theta})}{N-n}$$

Proof :

$$\begin{aligned}
\mathbb{E}\{J(\hat{\theta})\} &= \mathbb{E}\{[Y-\Phi\hat{\theta}]^T[Y-\Phi\hat{\theta}]\} \\
&= \mathbb{E}\{[Y-\Phi(\Phi^T\Phi)^{-1}\Phi^T Y]^T[Y-\Phi(\Phi^T\Phi)^{-1}\Phi^T Y]\} \\
&= \mathbb{E}\{Y^T Y - Y^T\Phi(\Phi^T\Phi)^{-1}\Phi^T Y\} \\
&= \mathbb{E}\{Y^T[I-\Phi(\Phi^T\Phi)^{-1}\Phi^T]Y\} \\
&= \mathbb{E}\{(\Phi\theta_0+E)^T[I-\Phi(\Phi^T\Phi)^{-1}\Phi^T](\Phi\theta_0+E)\} \\
&= \mathbb{E}\{E^T[I-\Phi(\Phi^T\Phi)^{-1}\Phi^T]E\} \\
&= \mathbb{E}\{tr\{E^T[I-\Phi(\Phi^T\Phi)^{-1}\Phi^T]E\}\} \\
&= \mathbb{E}\{tr\{[I-\Phi(\Phi^T\Phi)^{-1}\Phi^T]EE^T\}\} \\
&= [tr\{I_N\} - tr\{(\Phi^T\Phi)^{-1}\Phi^T\Phi\}]\sigma^2 = (N-n)\sigma^2
\end{aligned}$$

Least Squares Algorithm (stochastic case)

Minimum Variance Unbiased Estimates : If $e(k)$ is not white and has a covariance matrix R, it is reasonable to weight the errors inversely proportional to the noise variance (less weight if variance is high). A weighted least squares can be used with :

$$W^T W = R^{-1}$$

then :

$$\hat{\theta} = (\Phi^T R^{-1} \Phi)^{-1} \Phi^T R^{-1} Y$$

and the covariance of the estimates becomes :

$$\begin{aligned}
\text{cov}(\hat{\theta}) &= \mathbb{E}\{(\hat{\theta} - \theta_0)(\hat{\theta} - \theta_0)^T\} \\
&= \mathbb{E}\{[(\Phi^T R^{-1} \Phi)^{-1} \Phi^T R^{-1} E][(\Phi^T R^{-1} \Phi)^{-1} \Phi^T R^{-1} E]^T\} \\
&= (\Phi^T R^{-1} \Phi)^{-1} \Phi^T R^{-1} R R^{-1} \Phi (\Phi^T R^{-1} \Phi)^{-1} = (\Phi^T R^{-1} \Phi)^{-1}
\end{aligned}$$

Remark : It can be shown that this estimator is the best unbiased estimator (with minimum variance) for the problem.

Maximum Likelihood Estimator (MLE)

Maximum Likelihood Estimator

This estimator chooses the parameter values that make the observed data, the most likely data to have been observed. The estimated parameters are obtained by maximizing the *Likelihood* function.

Likelihood function

Consider a random variable y with the probability density function $p(y, \theta_0)$, where θ_0 is the vector of the known parameters of the *pdf*. The likelihood function is $p(y^\circ, \theta)$, where y° is a realization of the random variable and θ is the unknown parameter to be estimated.

Estimation Procedure

The MLE is obtained usually by maximizing the log-likelihood function :

$$\hat{\theta} = \arg\max_{\theta} \log p(y^\circ, \theta)$$

Maximum Likelihood Estimator (MLE)

Example (MLE of the DC level)

Consider a random variable $y = \theta_0 + e$, where e is a zero-mean Gaussian white noise with the following pdf :

$$p(y, \theta_0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\theta_0)^2/2\sigma^2}$$

If we have one observation y° what is the MLE of θ_0.

Solution :

$$\hat{\theta} = \arg\max_{\theta} \ln \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y^\circ-\theta)^2/2\sigma^2} \right] = \arg\max_{\theta} \left[-\ln\sqrt{2\pi\sigma^2} - (y^\circ - \theta)^2/2\sigma^2 \right]$$
$$= \arg\min_{\theta} (y^\circ - \theta)^2 = y^\circ$$

Maximum Likelihood Estimator (MLE)

Example (MLE of the DC level)

For the same problem, If we have n independent observations, what is the MLE of θ_0.

Solution : The pdf of $Y = [y_1, \ldots, y_n]$ is

$$p(Y, \theta_0) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_i - \theta_0)^2/2\sigma^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \theta_0)^2}$$

Therefore, the MLE is :

$$\hat{\theta} = \arg\max_{\theta} \ln p(Y, \theta) = \arg\min_{\theta} \sum_{i=1}^{n} (y_i - \theta)^2 = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Remark : For white Gaussian distribution, the MLE is the same as least squares estimator.

Maximum Likelihood Estimator (MLE)

Suppose that the data are generated by $y(k) = \phi^T(k)\theta_0 + e(k)$, where θ_0 is the vector of true parameters and $e(k)$ a zero-mean stationary random Gaussian process. In a matrix form, $Y = \Phi\theta_0 + E$, and $R = \mathbb{E}\{EE^T\}$ is the noise covariance matrix. Then, the likelihood function is :

$$p(Y,\theta) = \frac{1}{\sqrt{(2\pi)^N \det(R)}} e^{-\frac{1}{2}(Y-\Phi\theta)^T R^{-1}(Y-\Phi\theta)}$$

The MLE of the parameters is :

$$\begin{aligned}\hat{\theta} &= \arg\max_{\theta} \ln p(Y,\theta) = \arg\min_{\theta} (Y-\Phi\theta)^T R^{-1}(Y-\Phi\theta) \\ &= (\Phi^T R^{-1}\Phi)^{-1}\Phi^T R^{-1} Y\end{aligned}$$

Remarks :

- The minimum variance estimator is an MLE if the noise is Gaussian.
- The estimate $\hat{\theta}$ has also a Gaussian distribution $\mathcal{N}(\theta_0, (\Phi^T R^{-1}\Phi)^{-1})$.

Concluding Remarks

- Least Squares Estimator (LSE) is an optimal estimator in the deterministic case.
- For linear models with zero-mean white noise, LSE is unbiased and has the optimal variance.
- For linear models with zero-mean colored noise of covariance R, weighted LSE with $W^TW = R^{-1}$ is unbiased and has the optimal variance.
- For linear models with zero-mean Gaussian white noise, LSE is the same as Maximum Likelihood Estimator (MLE).
- For linear models with zero-mean Gaussian colored noise of covariance R, weighted LSE with $W^TW = R^{-1}$ is the same as MLE.

Bias-Variance Tradeoff

Consider that the data are generated by the following model :

$$y(k) = \phi^T(k)\theta_0 + e(k)$$

where $e(k)$ is a zero-mean stationary random noise with variance σ^2.
Consider an estimator $\hat{\theta}$ with (dim $\hat{\theta} <$ dim θ_0) such that :

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

The quality of the estimator cannot be assessed by the bias and the variance of the parameters, so we define :

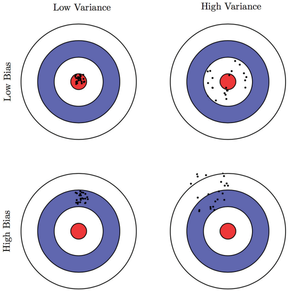
$$\text{Mean Square Error (MSE)} := \mathbb{E}\left\{[y(k) - \varphi^T(k)\hat{\theta}]^2\right\}$$

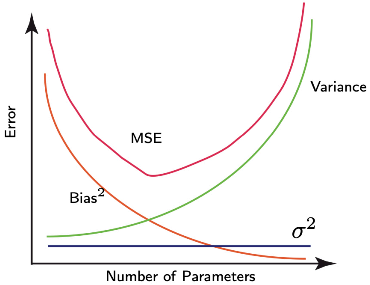
We also define :

$$\text{Bias error} := \phi^T(k)\theta_0 - \mathbb{E}\{\varphi^T(k)\hat{\theta}\} = \phi^T(k)\theta_0 - \varphi^T(k)\theta^*$$

$$\text{Variance error} := \mathbb{E}\left\{[\varphi^T(k)\hat{\theta} - \mathbb{E}\{\varphi^T(k)\hat{\theta}\}]^2\right\} = \mathbb{E}\left\{[\varphi^T(k)\hat{\theta} - \varphi^T(k)\theta^*]^2\right\}$$

Bias-Variance Tradeoff





$$\text{MSE} = \underbrace{\sigma^2}_{\text{Noise error}} + \underbrace{\mathbb{E}\left\{[\varphi^T(k)\hat{\theta} - \varphi^T(k)\theta^*]^2\right\}}_{\text{Variance error}} + \underbrace{[\phi^T(k)\theta_0 - \varphi^T(k)\theta^*]^2}_{\text{Bias error}}$$

Bias-Variance Tradeoff : By increasing the number of parameters (dim θ^*) the bias error is decreased but the variance error augments. Therefore, having a small bias may lead to a better MSE.