Parameter Estimation:

- The final objective is to have a recursive technique to calculate the system or the controller parameters, and that will be online and in real time.

- The basic algorithm is based on the "Least Square" principles.

- However, to give an overview and understanding of the problem, we will start by talking about system identification problem as a big problem and go down to the parameter estimation part.

* How to get models:

⊛ **Physics:** by using some principles like electrical systems as Ohm's Law, Faraday's Law ... etc, in mechanical systems as Newton Law, in thermal systems as energy conservation and so on.

⊛ **Experiments:** Based on experimental data.

⊛ **Combinations:** between the above two techniques.

Remarks:

- If we decide to use experiments, then the planning of the experiments will include the signals to be measured and most importantly it will include the type of input that we should apply at the system, so that the output of the system and the measured signal will expose the different dynamics of the system.

- This problem is called in (system identification / parameter identification / adaptive control) "presistent excitation" condition. The condition that the input must satisfied in order to be capable of identifing correctly the unknown system.

* Choice of Model Structure:

- Transfer Functions.
- Impulse Response.
- State Space Models.

⇒ Then, the next step will be to implement an algorithm to estimate or calculate the unknown parameter of the system.

* Validation:

Then, the next step will be to validate that model by comparing it's output with the output of a true system. If that satisfactory, then we take that parameter estimation and model as representation for the system and we depend on them in control design. But, if not, we have to re-design our experiment in order to get another model that fit the data.

* Identification Techniques:

- Non-parametric Methods:

@- Frequency Response Analysis: i.e. Bode plot, that we can plot the magnitude and phase versus frequency and from that we can conclude

the corner frequency, the gain of the system and we can deduce the transfer function of the system.

(b)- Transient Response Analysis: We can conclude the transfer function of the system. However, that technique is maybe limited to low order systems In stochastic systems, we can use correlation and spectrum analysis.

(c)- Correlation Analysis:

(d)- Spectrum Analysis:

- Parameteric Methods:

(a) Least Squares.

(b) Maximum likelihood.

* Model Reduction:

This is an important topic in system identification. We all know that real systems should be represented by high order models. However, we normally can neglect fast dynamics or non dominant poles and so on to reduce the system's order to make the design easier.

As we see before, that parameter estimation is one compact in system identification. Our ultimate goal is to have a recursion algorithm for parameter estimation.

* Least Square (LS) Principle:

- The sum of the squares of the difference between the actual and computed values (multiplied by numbers that reflect the degree of precision) is minimum.

- Normally, the actual values will be the observed measurements and the computed values will be values based on the computed parameter estimates. The difference between the two values is called the errors and we normally try to find some positive functions. The case of LS will be the squares of the errors and we will try to find the unknown parameters that would minimize that cost function.

- Sometimes, we multiply the errors by different weights and the reason for that could be due to different accurecies of the different errors or because we would like to give more weight to more recent data.

* Problem Formulation:

Let $y(i) = \varphi_1(i)\theta_1 + \varphi_2(i)\theta_2 + \dots + \varphi_n(i)\theta_n$

where: $y \equiv$ the observed variables.

$\theta_i \equiv$ the true model parameters.

$\varphi_i \equiv$ the known functions.

$$\phi^T(i) = [\varphi_1(i) \quad \varphi_2(i) \quad \dots \quad \varphi_n(i)]$$

$$\theta = [\theta_1 \quad \theta_2 \quad \dots \quad \theta_n]$$

$$\Rightarrow \boxed{y(i) = \phi^T(i)\theta}$$

* LS Objective:

The problem is to determine the parameter θ such that the output computed from $\hat{y}(i) = \phi^T(i)\theta$ agrees as closely as possible with the measurement $y(i)$ in the least square sense.

- It will be perfectly agree with $y(i)$ if we substitude θ for θ. Since we don't know the true values of θ, then there will be an error and that error which we would like to minimize it by the sum of squares by choosing suitable θ.

* Objective Function:

Then, we assume a cost function:

$$V(\theta, t) = \frac{1}{2} \sum_{i=1}^{t} [y(i) - \underbrace{\phi^T(i)\theta}_{\hat{y}(i),\ \text{predicted value}}]^2$$

and as we can see this function is quadratic depends on the error between the observed daten and the predicted value. We have here "$\frac{1}{2}$" this just for convinous as we are going to give gradient of V and that because of the quadratic term will get two time derivatives and the will cancel "$\frac{1}{2}$", this will be just easier for manipulation.

The previous function can be written as:

$$V(\theta, t) = \frac{1}{2} \sum_{i=1}^{t} \varepsilon^2(i) = \frac{1}{2} E^T E = \frac{1}{2} \|E\|^2$$

$\varepsilon^2(i)$: squares of errors

$E^T E$: or in vector notation

Let $Y = [\, y(1) \;\; y(2) \;\; \cdots \;\; y(t) \,]^T$

$E = [\, \varepsilon(1) \;\; \varepsilon(2) \;\; \cdots \;\; \varepsilon(t) \,]^T$

$\varepsilon(i) = Y(i) - \phi^T(i)\,\theta$, $\Phi(t) = \begin{bmatrix} \phi(1) \\ \phi(2) \\ \vdots \\ \phi(t) \end{bmatrix}$ → normally, this vector is called the "regressor" or the "data vector"

* A mathmatical Background :

Now, in order to obtain the estimates θ that minimizes V, we should differentiate the cost function with respect to a vector θ. That needs a breif mathmatical background, because we are going to differentiate with respect to a vector rathar than to a scalar. So, consider first a case where we have a scalar function 'f' :

Let $f = X^T X = x_1^2 + x_2^2 + \cdots + x_n^2$

$$\frac{\partial f}{\partial X} = 2 \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = 2X$$

on the other hand, if we have a scalar function 'f' :

Let $f = a^T X = X^T a = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$

$$\frac{\partial f}{\partial X} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a$$

Thus, if we apply this simple rules to our cost function V:

$$V(\theta, t) = \frac{1}{2} \sum_{i=1}^{t} \varepsilon^2(i) = \frac{1}{2} E^T(t)\, E(t)$$

where: $E = Y - \Phi\theta$

$\Rightarrow$ The minimum will be obtained from $\frac{\partial V}{\partial \theta} = 0$

$$\frac{\partial V}{\partial \theta} = \left(\frac{\partial E}{\partial \theta}\right)^T E = (-\Phi^T)(Y - \Phi\theta) = 0$$

if we put the previous function equal to 0, to obtain the minimum of the function:

Least square formula $\longrightarrow$ $\hat{\theta} = \theta = (\Phi^T\Phi)^{-1}\Phi^T Y$

notes:

- We notice that the result is a minimum because V is quadratic and we know for quadratic fn which is positive all the time, if we get an extremum, that extremum will be minimum and it is a global minimum.
  - $\rightarrow$ $\hat{\theta}$ minimizes V since V is quadratic in θ.
- We can write:

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

- We notice that, in order to calculate $\hat{\theta}$, we need to get $(\Phi^T\Phi)^{-1}$, than $(\Phi^T\Phi)$ must be invertible (It is called the excitation condition), we will see next lectures that $(\Phi^T\Phi)$ depends on how much the input is capable of

exposing the dynamic of the system as the observed data.

* Statistical Interpretation:

Theorem:

Consider the estimates $\hat{\theta} = (\Phi^T\Phi)^{-1}\Phi^T Y$. Let the data be generated from $y(i) = \phi^T(i)\theta + \overset{\text{noise}}{e(i)}$, $e(i)$, $i = 1, 2, 3, \dots, t$ is a sequence of independent equally distributed random variables with zero mean and variance σ^2. Let "$\mathbb{E}$" denotes the mathmatical expectation and "cov" denotes the convariance of a random variable. If $\Phi^T\Phi$ is non-singular, then:

1- $\mathbb{E}(\hat{\theta}) = \theta$

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

3- $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{t}[y(i) - \phi^T(i)\hat{\theta}]^2}{t-n} = \dfrac{2V(\hat{\theta}, t)}{t-n}$ is an unbiased estimate of σ^2.

where, "n" is the number of parameter and "t" is the number of data points.

* Proof of $\mathbb{E}(\hat{\theta}) = \theta$:

$$Y = \Phi\theta + E$$

↑ which has sequance of noise

"Multiply both sides by $(\Phi^T\Phi)^{-1}\Phi^T$

$$(\Phi^T\Phi)^{-1}\Phi^T Y = \underbrace{(\Phi^T\Phi)^{-1}\Phi^T\Phi}_{I}\theta + (\Phi^T\Phi)^{-1}\Phi^T E$$

$$\hat{\theta} = \theta + \underbrace{(\Phi^T\Phi)^{-1}\Phi^T E}_{\text{some error}}$$

it will give θ + some error

note that, Φ and E are independent and E has zero mean.

So, the expected value of the second term in the right hand is zero. The expected value of both sides leads to :

$$E(\hat{\theta}) = \theta$$

$$E(\hat{\theta}) = E\left[\theta + (\phi^T\phi)^{-1}\phi^T E\right]$$

$$= E(\theta) + E\left((\phi^T\phi)^{-1}\phi^T E\right)$$

ϕ, E are independent

$$= \theta + E(E) = \theta$$

zero from the definition since it has zero mean

Thats why we said that the least squares produces an unbiased estimates of parameters.

Example:

The data shown in the table below are collected from a second order process. The general model is:

k	0	1	2	3	4	5	6	7	8	9
u	0	1	1	1	1	1	1	1	1	1
y	0	0	0.24	0.77	1.4	2	2.5	3	3.4	3.7

$$y(k) = \alpha_1 y(k-1) + \alpha_2 y(k-2) + \beta_1 u(k-1) + \beta_2 u(k-2) + \varepsilon(k)$$

as you see here, $y(k)$ is related to the previous value of $y(k-1)$, $y(k-2)$, $u(k-1)$, and $u(k-2)$.

(a) Form the data matrices X and Y.

(b) Calculate the estimates of the unknown parameters.

Solution:

Since we have two sample delay, we will start from $k=2$.

As we derived before:

$$Y = [\, y(1) \;\; y(2) \;\cdots\; y(t) \,]$$

$$y(1) = \varphi_1(1)\,\theta_1 + \varphi_2(1)\,\theta_2 + \cdots + \varphi_n(1)\,\theta_n$$

$$\varphi(1) = [\, \varphi_1 \;\; \varphi_2 \;\cdots\; \varphi_n \,]$$

$$\theta = [\, \theta_1 \;\; \theta_2 \;\cdots\; \theta_n \,], \quad \Phi = \begin{bmatrix} \varphi(1) \\ \varphi(2) \\ \vdots \\ \varphi(t) \end{bmatrix}$$

$$y(2) = \alpha_1 y(1) + \alpha_2 y(0) + \beta_1 u(1) + \beta_2 u(0) + \varepsilon(2)$$

$$\Rightarrow \quad \varphi(2) = [\, y(1) \;\; y(0) \;\; u(1) \;\; u(0) \,] = [\, 0 \;\; 0 \;\; 1 \;\; 0 \,]$$

$$y(3) = \alpha_1 y(2) + \alpha_2 y(1) + \beta_1 u(2) + \beta_1 u(1) + \varepsilon(3)$$

$$\Rightarrow \varphi(3) = [\, y(2) \;\; y(1) \;\; u(2) \;\; u(1) \,] = [\, 0.24 \;\; 0 \;\; 1 \;\; 1 \,]$$

$$\vdots$$

$$y(9) = \alpha_1 y(8) + \alpha_2 y(7) + \beta_1 u(8) + \beta_2 u(7) + \varepsilon(9)$$

$$\Rightarrow \varphi(9) = [\, y(8) \;\; y(7) \;\; u(8) \;\; u(7) \,] = [\, 3.4 \;\; 3 \;\; 1 \;\; 1 \,]$$

$$\Rightarrow \quad Y = [\, y(2) \;\; y(3) \;\cdots\; y(9) \,]^T$$

$$= [\, 0.24 \;\; 0.77 \;\; 1.4 \;\; 2 \;\; 2.5 \;\; 3 \;\; 3.4 \;\; 3.7 \,]^T$$

$$\phi^T = \begin{bmatrix} 0 & 0.24 & 0.77 & 1.4 & 2 & 2.5 & 3 & 3.4 \\ 0 & 0 & 0.24 & 0.77 & 1.4 & 2 & 2.5 & 3 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\theta = [\alpha_1 \quad \alpha_2 \quad \beta_1 \quad \beta_2]$$

$$\hat{\theta} = (\phi^T\phi)^{-1}\phi^T y = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1.35 \\ -0.444 \\ 0.24 \\ 0.213 \end{bmatrix}$$

$$\Rightarrow y(k) = 1.35\,y(k-1) - 0.444\,y(k-2) + 0.24\,u(k-1) + 0.213\,u(k-2)$$

* Weighting Measurements differently:

If we want to give different weight for different errors or different residues, then we can write our cost function as:

$$V = \frac{1}{2} E^T W E \qquad \text{instead of} \qquad V = \frac{1}{2} E^T E$$

W is a diagonal matrix, the LS estimate is:

$$\boxed{\hat{\theta} = (\phi^T W \phi)^{-1}\phi^T W y}$$