Recursive Parameter Estimation:

Today, will show how to recursively updates the value of parameters based on your measurements that we received from the system.

Before do that, let us introduce the basic models that we will deal with in our analysis and design.

* Models of dynamic systems:

1 Finite Impulse Response (FIR) Models:

$$y(t) = b_1 u(t-1) + b_2 u(t-2) + b_3 u(t-3) + \dots + b_n u(t-n)$$

$$= \begin{bmatrix} u(t-1) & u(t-2) & \dots & u(t-n) \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \varphi^T(t-1)\,\Theta$$

2 Transfer Function Model:

$$A(q)\, y(t) = B(q)\, u(t)$$

polynomial "A" polynomial "B"

where:

$$q\, u(t) \equiv u(t+1)$$

$$q^{-1} u(t) \equiv u(t-1)$$

note that:

q : Forward shift operator

q^{-1} : Backward shift operator

Let:

$$A(q) = q^n + a_1 q^{n-1} + \cdots + a_n$$

$$B(q) = b_1 q^{m-1} + b_2 q^{m-2} + \cdots + b_m$$

if we apply the previous law $A(q)\,y(t) = B(q)\,u(t)$

$$\Rightarrow y(t+n) + a_1 y(t+n-1) + \cdots + a_n y(t)$$

$$= b_1 u(t+m-1) + b_2 u(t+m-2) + \cdots + b_m u(t)$$

However, when we form the eqn for parameter estimation, we would like the regressor to be in terms of previous values of y and u. So, to do that just multiply both sides with q^{-n}:

$$y(t) + a_1 y(t-1) + \cdots + a_n y(t-n) = b_1 u(t+m-n-1) + b_2 u(t+m-n-2) + \cdots + b_m u(t-n)$$

$$\Rightarrow y(t) = -a_1 y(t-1) - a_2 y(t-2) - \cdots - a_n y(t-n) + b_1 u(t+m-n-1) + b_2 u(t+m-n-2) + \cdots + b_m u(t-n)$$

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

$$\varphi^T(t-1) = [-y(t-1) \cdots -y(t-n) \;\; u(t+m-n-1) \cdots u(t-n)]$$

$$\boxed{y(t) = \varphi^T(t-1)\,\theta}$$

3 Nonlinear Models:

Example:

nonlinear term "sin of input"

$$y(t) = -a\,y(t-1) + b_1 u(t-1) + b_2 \sin(u(t-1))$$

This can be put in the form: $y(t) = \varphi^T(t-1)\,\theta$

where: $\varphi^T(t-1) = [-y(t-1) \quad u(t-1) \quad \sin(u(t-1))]$

$\theta^T = [a \quad b_1 \quad b_2]$

4 Stochastic Models:

$$A(q^{-1})\,y(t) = B(q^{-1})\,u(t) + C(q^{-1})\,e(t)$$

Stochastic since we have polynomial "c" multiplied by error, we can think here "e" as a white noise and "c" as a filter.

- "c" describes the correlation of the disturbances/noise.
- 'e' is an immeasurable sequence of random independent signal of zero mean.
  - So in our estimation, we have to find some ways to calculate approximate value of "e"
- It is not possible to put the above model in the regressor form since the sequence $e(t)$ is unknown.
- $e(t)$ can be approximated by $\boxed{\varepsilon(t) = y(t) - \varphi^T(t)\,\hat{\theta}}$

Recursive Least Squares (RLS):

Theorem: Assume $\phi(t)$ has full rank for $t \geq t_0$. Given $p(t_0) = [\phi^T(t_0)\,\phi(t_0)]^{-1}$ and $\theta(t_0)$, the least squares estimates $\hat{\theta}(t)$ satisfies the recursive equation:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \underbrace{k(t)\left[y(t) - \varphi^T(t)\,\hat{\theta}(t-1)\right]}_{\text{correction term}}$$

$$k(t) = p(t)\,\varphi(t) = p(t-1)\,\varphi(t)\left[I + \varphi^T(t)\,p(t-1)\,\varphi(t)\right]^{-1}$$

$$p(t) = p(t-1) - \frac{p(t-1)\,\varphi(t)\,\varphi^T(t)\,p(t-1)}{I + \varphi^T(t)\,p(t-1)\,\varphi(t)}$$

OR $\quad p(t) = \left[I - k(t)\,\varphi^T(t)\right] p(t-1)$

- With the assumption that we have an initial guess for the covariance matrix p ($p(t_0) = [\phi^T(t_0)\,\phi(t_0)]^{-1}$) and an initial guess for the parameter estimates $\theta(t_0)$. This initial guess could be produced as we did before by having a patch of data and use that patch data to calculate the estimate $\hat{\theta}$ and the corresponding covariance matrix.

- If we don't have previous data, we can have initial guess for these values of $p(t_0)$ and $\theta(t_0)$, and there are different techniques for that. We will discuss that later on.

- Now, suppose we have the initial values for $p(t_0)$ and $\theta(t_0)$, then at each instant of time at each sample t, we calculate and update the values of parameters and covariance matrix.

Notes:

- $\hat{\theta}(t)$ is obtained by adding a correction term to $\hat{\theta}(t-1)$. The correction term depends on the current measurments $y(t)$ and a predicated value $\hat{y}(t) = \varphi^T(t)\,\hat{\theta}(t-1)$.

- The least square estimate can be interpreted as a kalman filter for the process:

$$\theta(t+1) = \theta(t)$$

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

* Deriving the basic equation:

Let us derive the basic equation of recursive parameter estimation based on the least square.

The LS estimate is: $\hat{\theta}(t) = (\phi^T\phi)^{-1}\phi^T Y$ and let $p(t) = (\phi^T\phi)^{-1}$

$$\Rightarrow \hat{\theta}(t) = p(t)\sum_{i=1}^{t}\varphi(i)y(i) = p(t)\left[\left(\sum_{i=1}^{t-1}\varphi(i)y(i)\right) + \underbrace{\varphi(t)\,y(t)}_{\text{Last term}}\right] \quad — (1)$$

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

By re-arrange the last eqn:

$$\sum_{i=1}^{t-1} \varphi(i) y(i) = P^{-1}(t-1)\,\hat{\theta}(t-1) \quad \text{— (2)}$$

According to definition:

$$P^{-1}(t) = \phi^T \phi = \sum_{i=1}^{t} \varphi(i)\varphi^T(i) = \underbrace{P^{-1}(t-1)}_{\text{previous value}} + \underbrace{\varphi(t)\varphi^T(t)}_{\text{last term}}$$

$$P^{-1}(t-1) = P^{-1}(t) - \varphi(t)\varphi^T(t) \quad \text{— (3)}$$

From (2) and (3):

$$\sum_{i=1}^{t-1} \varphi(i) y(i) = \left[P^{-1}(t) - \varphi(t)\varphi^T(t)\right]\hat{\theta}(t-1) \quad \text{— (4)}$$

From (4) and (1):

$$\hat{\theta}(t) = P(t)\left[\left(P^{-1}(t) - \varphi(t)\varphi^T(t)\right)\hat{\theta}(t-1) + \varphi(t) y(t)\right]$$

$$= \underbrace{P(t)\,P^{-1}(t)}_{1}\,\hat{\theta}(t-1) - P(t)\varphi(t)\varphi^T(t)\,\hat{\theta}(t-1) + P(t)\varphi(t)\,y(t)$$

$$\Rightarrow \boxed{\hat{\theta}(t) = \hat{\theta}(t-1) + P(t)\varphi(t)\left[y(t) - \varphi^T(t)\,\hat{\theta}(t-1)\right]}$$

Hint: Matrix Inverse Lemma:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1}$$

From ③:

$$p(t) = \left[p^{-1}(t-1) + \varphi(t)\varphi^T(t) \right]^{-1} \quad — \quad ⑤$$

By applying the matrix inverse lemma: $A = p^{-1}(t-1)$, $B = \varphi(t)$

$C = 1$, $D = \varphi^T(t)$

Hence (5) can be written as:

$$\boxed{p(t) = p(t-1) - p(t-1)\varphi(t)\left[I + \varphi^T(t)p(t-1)\varphi(t)\right]^{-1}\varphi^T(t)p(t-1)}$$

$$k(t) = p(t)\varphi(t)$$

$$= p(t-1)\varphi(t) - p(t-1)\varphi(t)\left[I + \varphi^T(t)p(t-1)\varphi(t)\right]^{-1}\varphi^T(t)p(t-1)\varphi(t)$$

$$= p(t-1)\varphi(t)\left[1 - \left[1 + \varphi^T(t)p(t-1)\varphi(t)\right]^{-1}\varphi^T(t)p(t-1)\varphi(t)\right]$$

$$= p(t-1)\varphi(t)\left[1 + \varphi^T(t)p(t-1)\varphi(t)\right]^{-1}\left(1 + \underbrace{\varphi^T(t)p(t-1)\varphi(t) - \varphi^T(t)p(t-1)\varphi(t)}_{0}\right)$$

$$\boxed{k(t) = p(t-1)\varphi(t)\left[1 + \varphi^T(t)p(t-1)\varphi(t)\right]^{-1}}$$

* Normalized Projection Algorithm:

- The purpose is to simplify recursive estimation by avoiding calculations of $p(t)$.
- Let $y(t) = \varphi^T(t)\theta$. Assume an estimate $\hat{\theta}(t-1)$ is available and new measurement is obtained.

- choose $\hat{\theta}(t)$ to minimize $\|\hat{\theta}(t) - \hat{\theta}(t-1)\|$ subject to the constraint

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

- Let $V = \frac{1}{2}\left[\hat{\theta}(t) - \hat{\theta}(t-1)\right]^T\left[\hat{\theta}(t) - \hat{\theta}(t-1)\right] + \alpha\left[y(t) - \varphi^T(t)\,\hat{\theta}(t)\right]$

- $$\frac{\partial V}{\partial \hat{\theta}(t)} = \hat{\theta}(t) - \hat{\theta}(t-1) - \alpha\,\varphi(t) = 0 \quad — (7)$$

- $$\frac{\partial V}{\partial \alpha} = y(t) - \varphi^T(t)\,\hat{\theta}(t) = 0 \quad — (8)$$

Multiply both sides of (7) by $\varphi^T(t)$,

$$\varphi^T(t)\,\hat{\theta}(t) - \varphi^T(t)\,\hat{\theta}(t-1) = \alpha\,\varphi^T(t)\,\varphi(t) \quad — (9)$$

note from (8), $y(t) = \varphi^T(t)\,\hat{\theta}(t)$ $\Rightarrow$ substitution in (9), then:

$$\alpha = \frac{1}{\varphi^T(t)\,\varphi(t)}\left[y(t) - \varphi^T(t)\,\hat{\theta}(t-1)\right] \quad — (10)$$

substitution for α from (10) into (7) gives:

$$\boxed{\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\varphi(t)}{\varphi^T(t)\,\varphi(t)}\left[y(t) - \varphi^T(t)\,\hat{\theta}(t-1)\right]}$$

Kaczmarz's Algorithm

note here, we calculated $\hat{\theta}(t)$ without calculate of $P(t)$.

- Let $\alpha \geq 0$ and $0 < \gamma < 2$. To normalize the algorithm and avoid division by zero, it is modified to:

$$\boxed{\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\gamma\, \varphi(t)}{\alpha + \varphi^T(t)\, \varphi(t)} \left[y(t) - \varphi^T(t)\, \hat{\theta}(t-1) \right]}$$

Projection Algorithm (PA)

- The projection algorithm assumes that the data is generated by $\boxed{y(t) = \varphi^T(t)\theta}$ with no error. When the data is generated by $\boxed{y(i) = \varphi^T(t)\theta + e(i)}$ with additional error, a simplified algorithm is given by:

$$\boxed{\hat{\theta}(t) = \hat{\theta}(t-1) + p(t)\, \varphi(t) \left[y(t) - \varphi^T(t)\, \hat{\theta}(t-1) \right]}$$

Stochastic Alogorithm (SA)

where: $p(t) = \left(\sum_{i=1}^{t} \varphi^T(i)\, \varphi(i) \right)^{-1}$

This is the stochastic approximation (SA) algorithm. Notice that $p(t) = \phi \phi^T$ is now scalar when $y(t)$ is a scalar.

- A further simplification is the "Least Mean Square (LMS)" algorithm in which the parameter updating is done by using:

$$\boxed{\hat{\theta}(t) = \hat{\theta}(t-1) + \gamma\, \varphi(t) \left[y(t) - \varphi^T(t)\, \hat{\theta}(t-1) \right]}$$

Least Mean Square (LMS) Algorithm

where γ is constant.