Lecture 6

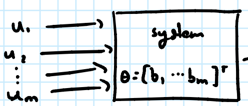
Monday, January 31, 2022 9:26 AM

Admin: HW01 due Feb 9th.

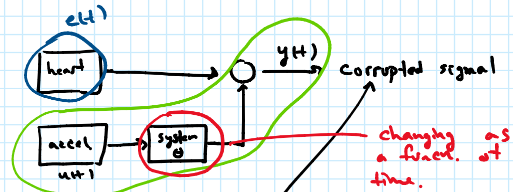
Question: about co-labs.

Recursive Least Squares Parameter Estimation

Last time: Least squares Estimation

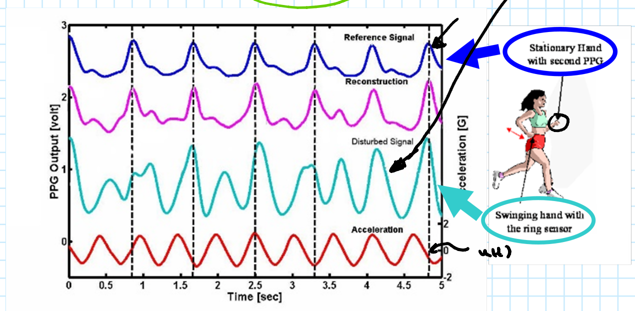


$y = \varphi^T \theta$, $\varphi^T = [u_1 \cdots u_m]$



- Find $\hat{\theta}$ by minimizing $V_N(\hat{\theta}) = \frac{1}{N}\sum_{t=1}^{N}[\hat{y}(t|\hat{\theta}) - y(t)]^2$

$\Rightarrow \hat{\theta} = PB$ where $P = \Big[\underbrace{\sum_{t=1}^{N}(\varphi(t)\varphi^T(t))}_{\text{auto-covariance}}\Big]^{-1}$ and $B = \Big[\sum_{t=1}^{N} y(t)\varphi(t)\Big]$

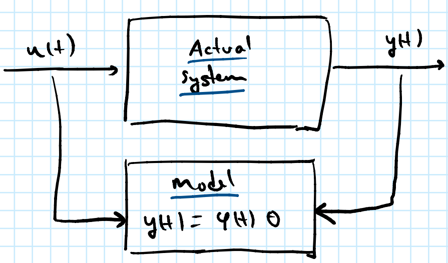


1. Important to update $\hat{\theta}$
2. we want an efficient way to do so.

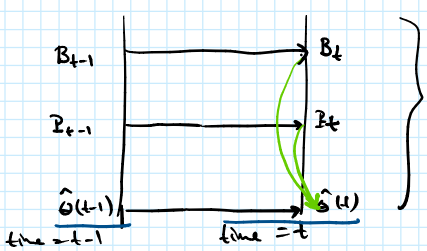
$\Rightarrow$ Recursive least squares

Recursive Least Squares Algorithm

- Least squares (Above) is useful for batch processing where all data are already collected
  $\Rightarrow$ no more to consider for estimating $\hat{\theta}$
- when we are estimating parameters in real-time (i.e., online updating $\hat{\theta}$ is required), a recursive computing method is preferred.
  - Reduce computational overhead. $\Rightarrow$ faster computation
  - No need to store all of the data. $\Rightarrow$ higher frequency.



$\Downarrow$ process steps for updating.



Goal: update our variables and estimated parameters based only on (t-1)

- Three steps for obtaining a recursive least squares (RLS) algorithm.

algorithm.

- Step 1: Write B_t and P_t in terms of B_{t-1} & P_{t-1}

$$\underline{B_t} = \sum_{i=1}^{t} y(i)\,\varphi(i) = \underbrace{\sum_{i=1}^{t-1} y(i)\,\varphi(i)}_{B_{t-1}} + y(t)\,\varphi(t)$$

$$\Rightarrow B_t = B_{t-1} + y(t)\,\varphi(t)$$

For P_t

$$P_t^{-1} = \sum_{i=1}^{t}\left[\varphi(i)\,\varphi^T(i)\right] = \underbrace{\sum_{i=1}^{t-1}\left[\varphi(i)\,\varphi^T(i)\right]}_{P_{t-1}^{-1}} + \varphi(t)\,\varphi^T(t)$$

$$\Rightarrow P_t^{-1} = P_{t-1}^{-1} + \varphi(t)\,\varphi^T(t)$$

$\rightarrow$ Need to compute inverse of P_t^{-1}

- Step 2: Use Matrix Inversion Lemma to find $\underline{\underline{P_t}}$

First, premultiply by P_t and postmultiply by P_{t-1}

$$\Rightarrow \overset{\text{Ident.}}{\cancel{P_t P_t^{-1}}} P_{t-1} = P_t \overset{\text{Ident.}}{\cancel{P_{t-1}^{-1} P_{t-1}}} + P_t\,\varphi(t)\,\varphi^T(t)\,P_{t-1}$$

$$\Rightarrow P_{t-1} = P_t + P_t\,\varphi(t)\,\varphi^T(t)\,P_{t-1}$$

Next, postmultiply by $\varphi(t)$:

$$P_{t-1}\,\varphi(t) = P_t\,\varphi(t) + P_t\,\varphi(t)\,\varphi^T(t)\,P_{t-1}\,\varphi(t)$$

$$= P_t\,\varphi(t)\left[1 + \underbrace{\varphi^T(t)\,P_{t-1}\,\varphi(t)}\right]$$

Solve for $P_t\,\varphi(t)$:

$(1\times m)(m\times m)(m\times 1)$

$\Rightarrow m\times 1$

$\Rightarrow (1\times 1)$

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

Finally, postmultiply by $\varphi^T(t)\,P_{t-1}$:

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

$P_{t-1} - P_t$

$$\Rightarrow \text{so, } \boxed{P_t = P_{t-1} - \frac{P_{t-1}\,\varphi(t)\,\varphi^T(t)\,P_{t-1}}{1 + \varphi^T(t)\,P_{t-1}\,\varphi(t)}}$$

update formula for inverted covariance matrix

* matrix inversion is <u>not required</u>

- Step 3 Reduce $\hat{\theta} = P_t B_t$ to the following recursive form:

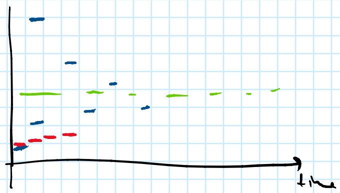
$$\hat{\theta}(t) = \hat{\theta}(t-1) + K_t\,\underbrace{\left(y(t) - \varphi^T(t)\,\hat{\theta}(t-1)\right)}\Big]$$

Error Correction Gain

prediction Error using prior estimates.

$\hat{\theta}(t) = P_t B_t$; $\hat{\theta}(t-1) = P_{t-1} B_{t-1}$ so,

$$\hat{\theta}(t) - \hat{\theta}(t-1) = \underbrace{P_t B_t}_{\text{we have update formulas for these}} - P_{t-1} B_{t-1}$$



Algebra

$$\Rightarrow \quad \hat{\theta}(t) - \hat{\theta}(t-1) = \underbrace{\frac{P_{t-1}\varphi(t)}{1+\varphi^T(t)P_{t-1}\varphi(t)}}_{K_t} \left[y(t) - \underbrace{\varphi^T(t)\hat{\theta}(t-1)}_{\text{scalar}} \right]$$

Now, we have all eqns needed for RLS:

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

$$P_t = P_{t-1} - \frac{P_{t-1}\varphi(t)\varphi^T(t)P_{t-1}}{1+\varphi^T(t)P_{t-1}\varphi(t)}, \quad \text{for } t = 1, 2, \ldots$$

Assuming initial conditions:

$\hat{\theta}(0)$: arbitrary.

P_0 : usually identity matrix [can be any positive definite matrix]

Initial Conditions for RLS

* How do ICs $\hat{\theta}(0)$ and P_0 affect the quality of the parameter estimation?

- Even if ICs are sufficient, they may be suboptimal
- Comparing the "optimality" of different ICs is best done w/ a specific data set, we can theoretically show that optimal $\hat{\theta}(0)$ & P_0 can be derived from prelim. data.

Theorem: RLS algorithm minimizes the following cost func:

$$J_t(\theta) = \frac{1}{2}\sum_{i=1}^{t} \underbrace{\left[y(i) - \varphi^T(i)\theta \right]^2}_{\text{Squared Estimation Error (A)}} + \underbrace{\frac{1}{2}\left(\theta - \hat{\theta}(0)\right)^T P_0^{-1}\left(\theta - \hat{\theta}(0)\right)}_{\text{weighted squared distance of parameters from } \hat{\theta}(0) \text{ (B)}}$$

Proof: See addendum

1) As time t gets larger, more data obtained and term (A) gets much larger than (B)
   $\Rightarrow$ Influence of ICs fades.
2) Early in time t, θ is pulled toward $\hat{\theta}(0)$ especially when the eigenvalues of P_0^{-1} are large.
3) In contrast, if the eigenvalues of P_0^{-1} are small, θ tends to respond more quickly to prediction error: $[y(t) - \varphi^T(t)\theta]$
4) The initial matrix P_0 represents the level of confidence in initial parameter value, $\hat{\theta}(0)$.

Addendum

Proof: Differiate $J_t(\theta)$ and set to zero:

$$\frac{dJ_t(\theta)}{d\theta} = -\sum_{i=1}^{t}(y(i)-\varphi^T(i)\theta)\varphi(i) + P_0^{-1}(\theta-\hat{\theta}(0)) = 0 \qquad (19)$$

– Collect terms: $\underbrace{\left[\sum_{i=1}^{t}\varphi(i)\varphi^T(i) + P_0^{-1}\right]}_{=P_t^{-1}}\theta = \sum_{i=1}^{t} y(i)\varphi(i) + P_0^{-1}\hat{\theta}(0)$

– Solving for θ: $\hat{\theta}(t) = P_t\left[\sum_{i=1}^{t} y(i)\varphi(i) + P_0^{-1}\hat{\theta}(0)\right]$

$$= P_t\left[\underbrace{y(t)\varphi(t)}_{\text{current data}} + \underbrace{\sum_{i=1}^{t-1} y(i)\varphi(i) + P_0^{-1}\hat{\theta}(0)}_{\text{Previous data}}\right]$$

$\Rightarrow$ The term for the previous data can be written as:

$$\sum_{i=1}^{t-1} y(i)\varphi(i) + \underbrace{P_0^{-1}\hat{\theta}(0)}_{B_0} = \sum_{i=1}^{t-1} y(i)\varphi(i) + B_0 \qquad (20)$$

$$B_0 \xrightarrow[\text{from}]{} \hat{\theta}(0) = P_0 B_0$$

Since we're using prelim data, B_0 is the same as B_{t-1}, and $B_{t-1} = P_{t-1}^{-1}\hat{\theta}(t-1)$

– Now recall that $P_t^{-1} = \varphi(t)\varphi^T(t) + P_{t-1}^{-1}$. We can write:

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

$$= \hat{\theta}(t-1) + \underbrace{P_t\varphi(t)}\left[y(t) - \varphi^T(t)\hat{\theta}(t-1)\right] \qquad (21)$$

– If we postmultiply $\varphi(t)$ to both sides of (14) in Lecture 4 we get the expression for $P_t\varphi(t)$ from prior data

$$P_t\varphi(t) = P_{t-1}\varphi(t) - \frac{P_{t-1}\varphi(t)\varphi^T(t)P_{t-1}\varphi(t)}{(1+\varphi^T(t)P_{t-1}\varphi(t))} \qquad (22)$$

$$= \frac{P_{t-1}\varphi(t)}{(1+\varphi^T(t)P_{t-1}\varphi(t))}$$

– Plugging this into (21) we get:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{P_{t-1}\varphi(t)}{(1+\varphi^T(t)P_{t-1}\varphi(t))}\left[y(t) - \varphi^T(t)\hat{\theta}(t-1)\right] \qquad (23)$$