Lecture 7

Wednesday, February 2, 2022 10:45 AM

Admin

① HW01 due next weds.

② OH 3-4pm today.

Recursive Least Squares - Part II

Other Important Properties of RLS

- Convergence. $\lim_{t\to\infty} \|\hat{\theta}(t) - \hat{\theta}(t-1)\| = 0$

  → see Goodwin and Sin's book, ch. 3. } Ljung, 1999 is my favorite.

  → The change to the P matrix: $\Delta P = P_t - P_{t-1}$ is negative semi definite, i.e.,

$$\Delta P = \frac{P_{t-1}\varphi(t)\varphi^T(t)P_{t-1}}{1+\varphi^T(t)P_{t-1}\varphi(t)} \leq 0$$

sloppy.

for and $\varphi(t) \in \mathbb{R}^m$ & pos. definite P_{t-1}

Estimation of Time-varying Parameters

- In some systems, the ~~model~~ system parameters can change w/ time

  Ex 1: Shocks in a car traveling over rough terrain.

  Ex 2: Output of a muscle can decrease w/ fatigue.

  Ex 3: Mechanical components wearing w/ time.

- We want the RLS algorithm to preferentially consider the most recent data and "forget", or reduce importance of older data.

- Exponentially weigting the data facilitates this:
  - Relies on a "forgetting factor", α, where

$$0 < \alpha \leq 1$$

  - α is large for slowly changing processes
  - α is small for rapidly changing processes
  - Goal: minimize weighted squared error:

$$J(\alpha) = \sum_{i}^{t} \alpha^{t-i} e^2(i) \leftarrow \text{minimize given } \alpha$$

- Goal: Minimize weighted squared error:

$$J_t(\theta) = \sum_{i=1}^{t} \alpha^{t-i} e^2(i) \leftarrow \text{minimize given } \alpha$$

$e(i) = [\hat{y}(t|\theta) - y(t)]$

$$\hat{\theta}(t) = \arg\min_{\theta} J_t(\theta)$$

- $\hat{\theta}(t)$ is given by the folloing recursive eqn:

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

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

- Drawbacks to using forgetting factors
  - when system being considered reaches "steady state", the matrix

    $[P_{t-1}\varphi(t)\varphi^T(t)P_{t-1}]$ tends toward the null matrix

$\Rightarrow P_t \simeq \frac{1}{\alpha} P_{t-1}$

- Since $\alpha < 1$, the $\frac{1}{\alpha}$ term makes P_t larger than P_{t-1}, so that P_t begins to increase exponentially.
  - $\Rightarrow$ called the "Blow-up" problem.

- Remedy for the "Blow up" problem
  - We use the "covariance re-setting method" to refresh the P matrix

$$P_t^* = \underset{\text{gain}}{K} \; \underset{\text{identity matrix}}{I} \quad ; \quad 0 < K \le \infty$$

: Typically reset every, 10, 100, 1000 time steps.

$\rightarrow$ "Revitalizes" the algorithm.

Orthogonal Projection Algorithm

- The RLS algorithm iteratively converges on its final parameter value, but this can take more than m time steps

  m: # of parameters.

The orthogonal projection algorithm provides a least squares soln. in exactly m time steps.

- Assume: $\Phi_m = [\varphi_1 \ \varphi_2 \ \cdots \ \varphi_m]$ $\underbrace{\text{span the m-dim space}}_{(\Phi\Phi^T) \text{ is full rank.}}$
- set $P_0 = I$ $(m \times m)$ and $\hat{\theta}(0)$ is arbitrary.

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

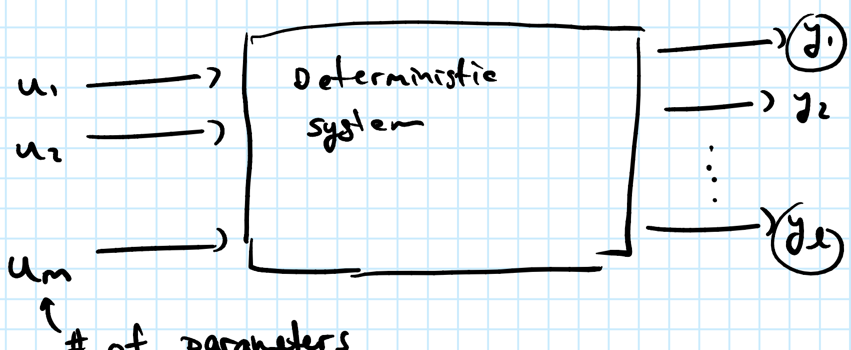
missing "1+"

- This enables the correction term to be larger
  => converges faster.
- But, $\varphi^T(t) P_{t-1} \varphi(t)$ is ill conditioned. (≈ 0), the gain becomes very large. & can amplify noise.

* orthogonal projection is more efficient, but less robust than RLS.

Multi-output Weighted Least Squares Estimation

- In reality, there will be systems w/ multiple outputs.



$\uparrow$ # of parameters

- The output becomes $\bar{y}(t) = [y_1 \; y_2 \; \cdots \; y_\ell] \in \mathbb{R}^\ell$
- For each output $\hat{y}_i(t) = \varphi_i^T(t)\theta$ so that

$$\bar{y}(t) = \underbrace{\begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_m \end{bmatrix}}_{m \times \ell} \theta \Rightarrow \psi^T(t)\theta; \; \psi \in \mathbb{R}^{m \times \ell}, \; \theta \in \mathbb{R}^{m \times 1}$$

- The estimation error $\bar{e}(t) = \begin{bmatrix} e_1 \\ \vdots \\ e_\ell \end{bmatrix} = \bar{y}(t) - \psi^T(t)\theta$

• In some instances, we may want to prioritize errors differently.

$\Rightarrow$ Use weighted multi-output squared error approach

$$J_t(\theta) = \sum_{i=1}^{t} \bar{e}^T(t) \underset{\uparrow}{W} \bar{e}(t)$$

output weighting matrix.

- RLS problem has similar form:

$$\hat{\theta}(t) = \underset{\theta}{\operatorname{argmin}} \, J_t(\theta) \Rightarrow \hat{\theta}(t) = P_t B_t$$

$$P_t = \left[\sum_{i=1}^{t} \psi(i) W \psi^T(i) \right]^{-1}, \quad B_t = \sum_{i=1}^{t} \psi^T(i) W \bar{y}(i)$$

RLS:

$$\boxed{\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) + P_t \psi(t) W \left[\bar{y}(t) - \psi^T(t) \hat{\theta}(t-1) \right] \\ P_t &= P_{t-1} - P_{t-1}\psi(t) \left[W^{-1} + \psi^T(t) P_{t-1} \psi(t) \right]^T \psi^T(t) P_{t-1} \end{aligned}}$$