

Estimación y Predicción en Series Temporales

Convergencia de RLS

Departamento de Procesamiento de Señales

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería

2022

Biblio:

- Haykin, *Adaptive Filter Theory*, 4ta edición, 2002.
Capítulo 9: “Recursive Least Squares Adaptive Filters”.

Estudiaremos la convergencia del algoritmo en los siguientes aspectos:

- Convergencia de $\mathbb{E}[\hat{\mathbf{w}}(n)]$
- Convergencia de $\hat{\mathbf{w}}(n)$ en media cuadrática
- Convergencia en media de la media cuadrática de $\alpha(n)$

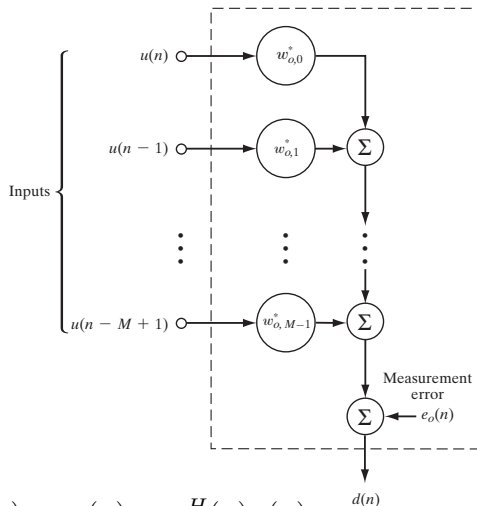
Recordemos que

$$\alpha(n) = d(n) - \hat{\mathbf{w}}^H(n-1)\mathbf{u}(n) \quad (\text{error a priori})$$

$$\hat{\mathbf{w}}(n) = \hat{\mathbf{w}}(n-1) + \mathbf{k}(n)\alpha^*(n) \quad (\text{actualización óptima})$$

$$\mathcal{E}_{min}(n) = d(n) - \hat{\mathbf{w}}^H(n)\mathbf{u}(n) \quad (\text{error a posteriori})$$

Introducción (2)



Tenemos $d(n) = e_o(n) + \mathbf{w}_0^H(n)\mathbf{u}(n)$. Observation

Suponemos $\mathbf{w}_0(n) = \mathbf{w}_0$ constante (estacionario).

En este caso el menor error se obtiene con $\lambda = 1$.

Convergencia del RLS en valor medio

Comenzamos con $\Phi(0) = \delta \mathbf{I}$ con delta $\delta > 0$ pequeño para asegurar que $\Phi(n) \succ 0$. La ecuación normal nos da

$$\hat{\mathbf{w}}(n) = \Phi^{-1}(n) \boldsymbol{\theta}(n), \quad n > M \quad (1)$$

$$\text{Si } \lambda = 1, \quad \Phi(n) = \sum_{i=1}^n \mathbf{u}(i) \mathbf{u}(i)^H + \Phi(0)$$

$$\begin{aligned} \text{y } \boldsymbol{\theta}(n) &= \sum_{i=1}^n \mathbf{u}(i) d^*(i) = \sum_{i=1}^n \mathbf{u}(i) (e_0(i) + \mathbf{w}_0^H \mathbf{u}(i))^* \\ &= \sum_{i=1}^n \mathbf{u}(i) \mathbf{u}^H(i) \mathbf{w}_0 + \sum_{i=1}^n \mathbf{u}(i) e_0^*(i) \\ &= \Phi(n) \mathbf{w}_0 - \Phi(0) \mathbf{w}_0 + \sum_{i=1}^n \mathbf{u}(i) e_0^*(i) \end{aligned} \quad (2)$$

Sustituyendo (2) en (1),

$$\hat{\mathbf{w}}(n) = \mathbf{w}_0 - \Phi^{-1}(n) \Phi(0) \mathbf{w}_0 + \Phi^{-1}(n) \sum_{i=1}^n \mathbf{u}(i) e_0^*(i).$$

Convergencia del RLS en valor medio (2)

$$\hat{\mathbf{w}}(n) = \mathbf{w}_0 - \Phi^{-1}(n)\Phi(0)\mathbf{w}_0 + \Phi^{-1}(n) \sum_{i=1}^n \mathbf{u}(i)e_0^*(i).$$

Para calcular $\mathbb{E}[\hat{\mathbf{w}}(n)]$ necesitamos la **propiedad**: $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$.

Luego,

$$\begin{aligned}\mathbb{E}[\hat{\mathbf{w}}(n)] &= \mathbf{w}_0 + \mathbb{E} \left[\Phi^{-1}(n) \sum_{i=1}^n \mathbf{u}(i)e_0^*(i) - \Phi^{-1}(n)\Phi(0)\mathbf{w}_0 \right] \\ &= \mathbf{w}_0 + \mathbb{E} \left[\mathbb{E} \left[\Phi^{-1}(n) \sum_{i=1}^n \mathbf{u}(i)e_0^*(i) - \Phi^{-1}(n)\Phi(0)\mathbf{w}_0 \mid \mathbf{u}(1), \dots, \mathbf{u}(n) \right] \right]\end{aligned}$$

Obs.:

- 1 $\Phi(n)$ está totalmente determinada por $\mathbf{u}(1), \dots, \mathbf{u}(n)$.
- 2 $e_0(i)$ es independiente de $\mathbf{u}(n) \forall i$.
- 3 $\mathbb{E}[e_0(i)] = 0$.

Asumiendo ergodicidad en autocorrelación de $\{u(n)\}$,

$$\mathbf{R} \approx \frac{1}{n} \Phi(n) \quad \forall n > M,$$

$$\mathbb{E}[\hat{\mathbf{w}}(n)] = \mathbf{w}_0 - \frac{1}{n} \mathbf{R}^{-1} \Phi(0) \mathbf{w}_0 = \mathbf{w}_0 - \frac{\delta}{n} \mathbf{R}^{-1} \mathbf{w}_0 = \mathbf{w}_0 - \frac{\delta}{n} \mathbf{p}, \quad n > M.$$

Obs.: Si optamos por $\Phi(0) = \mathbf{0}$, la convergencia se da para $n > M$. no hay que esperar a $n \rightarrow +\infty$.

Convergencia de los coeficientes en media cuadrática

Ignorando los efectos de la inicialización ($n \gg \delta$ o $\delta = 0$),

$$\boldsymbol{\epsilon}(n) = \hat{\mathbf{w}}(n) - \mathbf{w}_0 = \boldsymbol{\Phi}^{-1}(n) \sum_{i=1}^n \mathbf{u}(i) e_0^*(i).$$

$$\begin{aligned} \mathbf{K}(n) &= \mathbb{E}_{u,e}[\boldsymbol{\epsilon}(n) \boldsymbol{\epsilon}^H(n)] \\ &= \mathbb{E}_{u,e}[\boldsymbol{\Phi}^{-1}(n) \sum_{i=1}^n \sum_{j=1}^n \mathbf{u}(i) e_0^*(i) e_0^*(j) \mathbf{u}^H(j) \boldsymbol{\Phi}^{-1}(n)] \\ &= \mathbb{E}_u[\boldsymbol{\Phi}^{-1}(n) \sum_{i=1}^n \sum_{j=1}^n \mathbf{u}(i) \underbrace{\mathbb{E}_e[e_0^*(i) e_0^*(j)]}_{\sigma^2 \delta_{ij}} \mathbf{u}^H(j) \boldsymbol{\Phi}^{-1}(n)] \\ &= \sigma^2 \mathbb{E}_u[\boldsymbol{\Phi}^{-1}(n) \underbrace{\sum_{j=1}^n \mathbf{u}(j) \mathbf{u}^H(j)}_{\boldsymbol{\Phi}(n)} \boldsymbol{\Phi}^{-1}(n)] \\ &= \sigma^2 \mathbb{E}[\boldsymbol{\Phi}^{-1}(n)]. \end{aligned}$$

Convergencia de los coeficientes en media cuadrática (2)

Suponemos ahora (similar a como hicimos para LMS):

- 1 $\mathbf{u}(1), \mathbf{u}(2), \dots, \mathbf{u}(n)$ i.i.d.
- 2 $\mathbf{u}(1), \mathbf{u}(2), \dots, \mathbf{u}(n)$ provienen de procesos gaussianos de matriz de autocorrelación \mathbf{R} .

Se puede ver (Haykin) que $\Phi^{-1}(n)$ sigue una distribución de Wishart compleja, y que cumple

$$\mathbb{E}[\Phi^{-1}(n)] = \frac{1}{n} \mathbf{R}^{-1}, \quad n > M.$$

Luego, podemos escribir $\mathbf{K}(n) = \frac{\sigma^2}{n} \mathbf{R}^{-1}$, $n > M$. El desvío cuadrático medio vale

$$\mathbb{E}[\epsilon^H(n)\epsilon(n)] = \mathbb{E}[\text{Tr}[\epsilon(n)\epsilon^H(n)]]$$

$$= \text{Tr}[\mathbf{K}(n)] = \frac{\sigma^2}{n} \text{Tr}[\mathbf{R}^{-1}(n)] = \frac{\sigma^2}{n} \sum_{i=1}^M \frac{1}{\lambda_i}, \quad n > M.$$

Obs.:

- 1 El error aumenta proporcionalmente al λ_{\min}^{-1} (problemas con situaciones mal condicionadas)

Curva de aprendizaje

Tenemos dos tipos de errores: $\alpha(n)$ error a prior, $e(n)$ error a posteriori.

Veremos que:

- $\alpha(n)$ grande al principio, luego decae.
- $e(n)$ pequeño al principio, luego crece.

Teníamos:

$$\alpha(n) = d(n) - \hat{\mathbf{w}}^H(n-1)\mathbf{u}(n)$$

$$d(n) = e_0(n) + \mathbf{w}_0^H \mathbf{u}(n).$$

De dónde

$$\alpha(n) = e_0(n) - (\hat{\mathbf{w}}(n-1) - \mathbf{w}_0)^H \mathbf{u}(n) = e_0(n) - \boldsymbol{\epsilon}^H(n-1)\mathbf{u}(n).$$

Tomamos como índice de performance $J'(n) = \mathbb{E}[|\alpha(n)|^2]$.

Curva de aprendizaje (2)

$$\begin{aligned} J'(n) &= \mathbb{E}[|\alpha(n)|^2] = \mathbb{E}[|e_0(n) - \boldsymbol{\epsilon}^H(n-1)\mathbf{u}(n)|^2] \\ &= \underbrace{\mathbb{E}[|e_0(n)|^2]}_{\sigma^2} + \underbrace{\mathbb{E}[\mathbf{u}^H(n)\boldsymbol{\epsilon}^H(n-1)\boldsymbol{\epsilon}^H(n-1)\mathbf{u}(n)]}_{(a)} \\ &\quad - \underbrace{\mathbb{E}[\boldsymbol{\epsilon}^H(n-1)\mathbf{u}(n)e_0^*(n)]}_{(b)} - \underbrace{\mathbb{E}[e_0(n)\mathbf{u}^H(n)\boldsymbol{\epsilon}(n-1)]}_{(c)}. \end{aligned}$$

Tenemos:

$$\begin{aligned} (a) &= \mathbb{E}[\text{Tr}[\mathbf{u}^H(n)\boldsymbol{\epsilon}^H(n-1)\boldsymbol{\epsilon}^H(n-1)\mathbf{u}(n)]] \\ &= \mathbb{E}[\text{Tr}[\mathbf{u}(n)\mathbf{u}^H(n)\boldsymbol{\epsilon}(n-1)\boldsymbol{\epsilon}^H(n-1)]] \\ &= \text{Tr}[\mathbb{E}[\mathbf{u}(n)\mathbf{u}^H(n)\boldsymbol{\epsilon}^H(n-1)\boldsymbol{\epsilon}^H(n-1)]] \\ &\approx \text{Tr}[\mathbb{E}[\mathbf{u}(n)\mathbf{u}^H(n)]\mathbb{E}[\boldsymbol{\epsilon}(n-1)\boldsymbol{\epsilon}^H(n-1)]] \\ &= \text{Tr}[\mathbf{R}\mathbf{K}(n-1)] = \frac{\sigma^2}{n-1} \text{Tr}[\mathbf{R}\mathbf{R}^{-1}] = \frac{M\sigma^2}{n-1}, \quad n > M. \end{aligned}$$

Curva de aprendizaje (3)

Luego,

$$(b) = \mathbb{E}[\boldsymbol{\epsilon}^H(n-1)\mathbf{u}(n)e_0^*(n)] = \mathbb{E}[\boldsymbol{\epsilon}^H(n-1)\mathbf{u}(n)] \underbrace{\mathbb{E}[e_0^*(n)]}_0 = 0,$$

y de la misma forma $(c) = 0$.

Finalmente,

$$J'(n) = \sigma^2 + \frac{M}{n-1}\sigma^2, \quad n > M.$$

Obs.:

- $J'(n) \xrightarrow{n \rightarrow +\infty} \sigma^2$. Esto es, en teoría, el algoritmo RLS no produce error en exceso o desajuste.
- La convergencia en media cuadrática es independiente de los valores propios de \mathbf{R} (matriz de autocorrelación del vector de entrada $\mathbf{u}(n)$).