

Estimación y Predicción en Series Temporales

Recursive Least Squares (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”.

En los métodos anteriores hemos utilizado el **error cuadrático medio (MSE)** como índice de performance.

En el algoritmo LMS se utilizó una estimación muy rudimentaria del gradiente para optimizar el MSE:

- En algunos casos este procedimiento puede no converger suficientemente rápido.
- En algunos casos puede producir un error en exceso demasiado alto.

Una alternativa es tomar una medida del error que no se base en valores esperados, sino que se calcule directamente sobre los datos.

Una función de este tipo es el **error cuadrático mínimo**

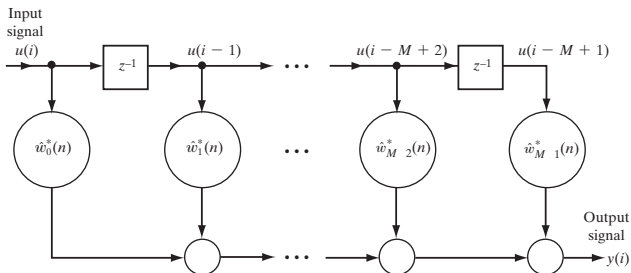
$$\mathcal{E}(n) = \sum_{i=1}^n |e(i)|^2.$$

La diferencia entre los dos criterios es fundamental:

- En el primer caso, el filtro se diseña según promedios en los ensambles \Rightarrow No depende de una realización particular.
- En el segundo caso, el filtro es óptimo para la realización en particular, y no en un sentido estadístico para una cierta clase de procesos.

Hoy estudiaremos el algoritmo adaptivo RLS, que se puede considerar como una versión determinística del filtro de Kalman.

Introducción



Notación:

$$\mathbf{u}^T(i) = [u(i), \dots, u(i - M + 1)], \quad \mathbf{w}^T(n) = [w_1(n), \dots, w_M(n)].$$

Índice de performance a minimizar: $\mathcal{E}(n) = \sum_{i=1}^n \beta(n, i) |e(i)|^2$,
donde:

- $\beta(n, i) \in (0, 1]$, $i = 1, 2, \dots, n$: factor de ponderación u “olvido”.
- $e(i) = d(i) - y(i) = d(i) - \mathbf{w}^H(n) \mathbf{u}(i)$, diferencia entre la señal deseada $d(i)$ y la salida del filtro.

Introducción (2)

$\beta(n, i)$ puede tomar muchas formas. Una forma muy común es $\beta(n, i) = \lambda^{n-i}$ con $\lambda > 0$ cercano (menor) a 1.

$$\Rightarrow \mathcal{E}(n) = \sum_{i=1}^n \lambda^{n-i} |e(i)|^2.$$

El valor óptimo de \mathbf{w} que minimiza $\mathcal{E}(n)$ viene dado por la ecuación normal determinística $\Phi(n)\hat{\mathbf{w}}(n) = \boldsymbol{\theta}(n)$, donde

$$\Phi(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{u}(i) \mathbf{u}^H(i), \quad \text{matriz } M \times M.$$

$$\boldsymbol{\theta}(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{u}(i) d^*(i), \quad \text{vector } M \times 1.$$

Ambos se pueden escribir recursivamente:

$$\Phi(n) = \lambda \left(\sum_{i=1}^{n-1} \lambda^{n-i-1} \mathbf{u}(i) \mathbf{u}^H(i) \right) + \mathbf{u}(n) \mathbf{u}^H(n) = \lambda \Phi(n-1) + \mathbf{u}(n) \mathbf{u}^H(n),$$

$$\boldsymbol{\theta}(n) = \lambda \left(\sum_{i=1}^{n-1} \lambda^{n-i-1} \mathbf{u}(i) d^*(i) \right) + \mathbf{u}(n) d^*(n) = \lambda \boldsymbol{\theta}(n-1) + \mathbf{u}(n) d^*(n).$$

Algoritmo recursivo de mínimos cuadrados exponenciales

Preliminares

Objetivo: calcular $\hat{\mathbf{w}}(n)$ recursivamente, según las ecs. normales.

- Utilizamos la identidad de Woodbury:

Sean $\mathbf{A} \succ 0$ y $\mathbf{B} \succ 0$ dos matrices $M \times M$, $\mathbf{D} \succ 0$ matriz $N \times N$, \mathbf{C} matriz $M \times N$, tales que $\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^H$. Entonces, se cumple $\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}(\mathbf{D} + \mathbf{C}^H\mathbf{B}\mathbf{C})^{-1}\mathbf{C}^H\mathbf{B}$.

- Como $\Phi(n) \succ 0 \forall n$, aplicamos el lema anterior a

$$\Phi(n) = \lambda\Phi(n-1) + \mathbf{u}(n)\mathbf{u}^H(n),$$

identificando $\mathbf{A} = \Phi(n)$, $\mathbf{B}^{-1} = \lambda\Phi(n-1)$, $\mathbf{C} = \mathbf{u}(n)$ y $\mathbf{D} = 1$:

$$\Phi^{-1}(n) = \lambda^{-1}\Phi^{-1}(n-1) - \frac{\lambda^{-2}\Phi^{-1}(n-1)\mathbf{u}(n)\mathbf{u}^H(n)\Phi^{-1}(n-1)}{1 + \lambda^{-1}\mathbf{u}^H(n)\Phi^{-1}(n-1)\mathbf{u}(n)}.$$

Algoritmo recursivo de mínimos cuadrados exponenciales

Preliminares (2)

Reescribiendo (1) usando (2):

$$\mathbf{P}(n) = \lambda^{-1} \mathbf{P}(n-1) - \lambda^{-1} \mathbf{k}(n) \mathbf{u}^H(n) \mathbf{P}(n-1). \quad (3)$$

Obs: $\mathbf{P}(n)$ es $M \times M$, y $\mathbf{k}(n)$ (vector de ganancia) es $M \times 1$.

Reescribiendo (2):

$$\begin{aligned} \mathbf{k}(n) (1 + \lambda^{-1} \mathbf{u}^H(n) \mathbf{P}(n-1) \mathbf{u}(n)) &= \lambda^{-1} \mathbf{P}(n-1) \mathbf{u}(n) \\ \Leftrightarrow \mathbf{k}(n) &= \lambda^{-1} \mathbf{P}(n-1) \mathbf{u}(n) - \lambda^{-1} \mathbf{k}(n) \mathbf{u}^H(n) \mathbf{P}(n-1) \mathbf{u}(n) \\ &= (\lambda^{-1} \mathbf{P}(n-1) - \lambda^{-1} \mathbf{k}(n) \mathbf{u}^H(n) \mathbf{P}(n-1)) \mathbf{u}(n) \\ &\stackrel{(3)}{=} \mathbf{P}(n) \mathbf{u}(n) \\ &= \Phi^{-1}(n) \mathbf{u}(n). \end{aligned}$$

Algoritmo recursivo de mínimos cuadrados exponenciales

Actualización de los coeficientes

Para encontrar la relación de recurrencia en los coeficientes, utilizamos las siguientes ecuaciones:

$$\Phi(n)\hat{\mathbf{w}} = \boldsymbol{\theta}(n)$$

$$\boldsymbol{\theta}(n) = \lambda\boldsymbol{\theta}(n-1) + \mathbf{u}(n)d^*(n)$$

$$\mathbf{P}(n) = \Phi^{-1}(n)$$

$$\Rightarrow \hat{\mathbf{w}}(n) = \Phi^{-1}(n)\boldsymbol{\theta}(n)$$

$$= \mathbf{P}(n)\boldsymbol{\theta}(n)$$

$$= \lambda\mathbf{P}(n)\boldsymbol{\theta}(n-1) + \mathbf{P}(n)\mathbf{u}(n)d^*(n)$$

$$\stackrel{(3)}{=} \mathbf{P}(n-1)\boldsymbol{\theta}(n-1) - \mathbf{k}(n)\mathbf{u}^H(n)\mathbf{P}(n-1)\boldsymbol{\theta}(n-1) + \mathbf{P}(n)\mathbf{u}(n)d^*(n)$$

$$= \Phi^{-1}(n-1)\boldsymbol{\theta}(n-1) - \mathbf{k}(n)\mathbf{u}^H(n)\Phi^{-1}(n-1)\boldsymbol{\theta}(n-1) + \mathbf{P}(n)\mathbf{u}(n)d^*(n)$$

$$= \hat{\mathbf{w}}(n-1) - \mathbf{k}(n)\mathbf{u}^H(n)\hat{\mathbf{w}}(n-1) + \mathbf{P}(n)\mathbf{u}(n)d^*(n)$$

$$\stackrel{\mathbf{k}(n)=\mathbf{P}(n)\mathbf{u}(n)}{=} \hat{\mathbf{w}}(n-1) + \mathbf{k}(n)(d^*(n) - \mathbf{u}^H(n)\hat{\mathbf{w}}(n-1))$$

$$= \hat{\mathbf{w}}(n-1) + \mathbf{k}(n)\alpha^*(n),$$

Algoritmo recursivo de mínimos cuadrados exponenciales

Actualización de los coeficientes (2)

Retomando, tenemos entonces dos nuevas ecuaciones:

$$\hat{\mathbf{w}}(n) = \hat{\mathbf{w}}(n-1) + \mathbf{k}(n)\alpha^*(n), \quad (4)$$

$$\alpha(n) = d(n) - \hat{\mathbf{w}}^H(n-1)\mathbf{u}(n) \quad (5)$$

Observaciones:

- El término $\hat{\mathbf{w}}^H(n-1)\mathbf{u}(n)$ corresponde a una predicción de $d(n)$ basada en el estimador de coeficientes anterior.
- Por esta razón $\alpha(n)$ se llama error *a priori*.
- El error *a posteriori* es $e(n) = d(n) - \hat{\mathbf{w}}^H(n)\mathbf{u}(n)$.
- Notar el paralelismo con Kalman en (4).

Resumiendo, las ecuaciones (2) a (5) conforman el algoritmo RLS:

$$\mathbf{k}(n) = \frac{\lambda^{-1}\mathbf{P}(n-1)\mathbf{u}(n)}{1 + \lambda^{-1}\mathbf{u}^H(n)\mathbf{P}(n-1)\mathbf{u}(n)}$$

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

Algoritmo recursivo de mínimos cuadrados exponenciales

Inicialización

Para aplicar el algoritmo es necesario inicializar $\mathbf{P}(0)$ de forma que asegure que $\Phi(n)$ no sea singular:

- Podríamos evaluar $\left(\sum_{i=-n_0}^0 \lambda^{-i} \mathbf{u}(i) \mathbf{u}^H(i)\right)^{-1}$.
- Una alternativa más sencilla consiste en modificar la matriz de correlación determinística para asegurar que sea definida positiva:

$$\Phi(n) = \sum_{i=1}^n \lambda^{n-1} \mathbf{u}(i) \mathbf{u}(i)^H + \delta \lambda^n \mathbf{I}, \quad \text{con } \delta > 0 \text{ pequeño.}$$

Obtenemos $\Phi(0) = \delta \mathbf{I} \Rightarrow \mathbf{P}(0) = \delta^{-1} \mathbf{I}$.

Se puede demostrar que esto equivale a minimizar el funcional

$$\mathcal{E}(n) = \sum_{i=1}^n \lambda^{n-i} |e(i)|^2 + \delta \lambda^n \|\mathbf{w}(n)\|^2,$$

Resumen del algoritmo RLS exponencial

- Inicialización:

$$\mathbf{P}(0) = \delta^{-1} \mathbf{I}, \quad \delta > 0 \text{ pequeño.}$$

$$\hat{\mathbf{w}}(0) = \mathbf{0}$$

- Para cada instante $n = 1, 2, \dots$ calcular:

$$\mathbf{k}(n) = \frac{\lambda^{-1} \mathbf{P}(n-1) \mathbf{u}(n)}{1 + \lambda^{-1} \mathbf{u}^H(n) \mathbf{P}(n-1) \mathbf{u}(n)}$$

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

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

$$\mathbf{P}(n) = \lambda^{-1} \mathbf{P}(n-1) - \lambda^{-1} \mathbf{k}(n) \mathbf{u}^H(n) \mathbf{P}(n-1)$$

Algoritmo RLS exponencial

Actualización de la suma de errores

El valor mínimo del error de estimación, $\mathcal{E}_{min}(n)$, se alcanza para los coeficientes $\hat{\mathbf{w}}(n)$. Por el ppio de ortogonalidad, tenemos

$$\mathcal{E}_{min}(n) = \mathcal{E}_d(n) - \boldsymbol{\theta}^H(n) \hat{\mathbf{w}}(n), \quad (6)$$

$$\text{con } \mathcal{E}_d(n) = \sum_{i=1}^n \lambda^{n-i} |d(i)|^2 = \lambda \mathcal{E}_d(n-1) + |d(n)|^2.$$

$$\text{Teníamos: } \boldsymbol{\theta}(n) = \lambda \boldsymbol{\theta}(n-1) + \mathbf{u}(n) d^*(n) \quad (7)$$

$$\begin{aligned} \hat{\mathbf{w}}(n) &= \hat{\mathbf{w}}(n-1) + \mathbf{k}(n) (d^*(n) - \mathbf{u}^H(n) \hat{\mathbf{w}}(n-1)) \\ &= \hat{\mathbf{w}}(n-1) + \mathbf{k}(n) \alpha^*(n). \end{aligned} \quad (8)$$

$$\begin{aligned} \stackrel{(8) \Rightarrow (6)}{\implies} \mathcal{E}_{min}(n) &= \lambda \mathcal{E}_d(n-1) + |d(n)|^2 - \boldsymbol{\theta}^H(n) \hat{\mathbf{w}}(n-1) - \boldsymbol{\theta}^H(n) \mathbf{k}(n) \alpha^*(n) \\ &\stackrel{(7)}{=} \underbrace{\lambda (\mathcal{E}_d(n-1) - \boldsymbol{\theta}^H(n-1) \hat{\mathbf{w}}(n-1))}_{\mathcal{E}_{min}(n-1)} \\ &\quad + \underbrace{d(n) (d^*(n) - \mathbf{u}^H(n) \hat{\mathbf{w}}(n-1))}_{\alpha^*(n)} - \boldsymbol{\theta}^H(n) \mathbf{k}(n) \alpha^*(n). \end{aligned}$$

Algoritmo RLS exponencial

Actualización de la suma de errores (2)

Tenemos entonces:

$$\mathcal{E}_{min}(n) = \lambda \mathcal{E}_{min}(n-1) + (d(n) - \boldsymbol{\theta}^H(n) \mathbf{k}(n)) \alpha^*(n).$$

Ahora,

$$\boldsymbol{\theta}^H(n) \mathbf{k}(n) = \boldsymbol{\theta}^H(n) \boldsymbol{\Phi}^{-1}(n) \mathbf{u}(n) = (\boldsymbol{\Phi}^{-1}(n) \boldsymbol{\theta}(n))^H \mathbf{u}(n) = \hat{\mathbf{w}}^H(n) \mathbf{u}(n)$$

de donde

$$\begin{aligned} \mathcal{E}_{min}(n) &= \lambda \mathcal{E}_{min}(n-1) + \alpha^*(n) (d(n) - \hat{\mathbf{w}}^H(n) \mathbf{u}(n)) \\ &= \lambda \mathcal{E}_{min}(n-1) + \alpha^*(n) e(n). \end{aligned}$$