4.5 Algoritmo RLS (*Recursive Least Squares*)

- Método de mínimos cuadrados (LS)
  - Ecuaciones normales
  - Pseudoinversa
  - Variantes del LS
- Algoritmo RLS (*Recursive Least Squares*)
  - Introducción
  - Cálculo recursivo de la matriz de autocorrelación y la estima LS
  - Convergencia
  - Comparación de prestaciones con el LMS
- Conclusiones

Método de mínimos cuadrados (LS)

- LS: Solución determinista a problemas de estimación lineal
- Planteamiento del problema
  Determinar los coeficientes óptimos de un filtro FIR dados los patrones de entrada $x(n)$ y las salidas deseadas $d(n)$
- Solución estocástica: minimizar

$$J(\mathbf{w}) = E\left[|e(n)|^2\right] \Rightarrow \begin{cases} \text{Filtro de Wiener} \\ \text{LMS} \end{cases}$$

- Solución determinista: minimizar

$$J(\mathbf{w}) = \sum_{n=0}^{N-1} |e(n)|^2 \Rightarrow \begin{cases} \text{Mínimos cuadrados} \\ \text{RLS} \end{cases}$$

Principio de ortogonalidad

- Problema a resolver: encontrar el mínimo para $J = \sum_{n=0}^{N-1} e(n)e^*(n)$
- Gradiente

$$\nabla_k J = -2 \sum_{n=0}^{N-1} x(n-k)e^*(n)$$

- $\nabla_k J = 0 \rightarrow$ Principio de ortogonalidad
  La serie temporal de errores mínimos, $e_{min}(n)$, es ortogonal con la serie temporal de entrada del filtro $x(n-k)$

$$\sum_{n=0}^{N-1} x(n-k)e^*_{min}(n) = 0, \qquad k = 0, 1, 2, \cdots, M-1$$

- Corolario: La salida del filtro óptimo, $y_{min}(n)$, es ortogonal al error $e_{min}(n)$

$$\sum_{i=0}^{N-1} y_{min}(n)e^*_{min}(n) = 0$$

Ecuaciones normales

- Principio de ortogonalidad

$$\sum_{n=0}^{N-1} x(n-k)\left(d^*(n) - \sum_{i=0}^{M-1} \hat{\omega}_i x^*(n-i)\right) = 0$$

- Sistema de ecuaciones: Ecuaciones Normales

$$\sum_{i=0}^{M-1} \hat{\omega}_i \sum_{n=0}^{N-1} x(n-k)x^*(n-i) = \sum_{n=0}^{N-1} x(n-k)d^*(n), \;\; k = 0, \cdots, M-1$$

$$\left(\mathbf{X}^H \mathbf{X}\right) \hat{\mathbf{w}} = \mathbf{X}^H \mathbf{d}$$

- Notación Problema: $\mathbf{X}_{N\times M}\mathbf{w}^*_{M\times 1} + \mathbf{e}_{M\times 1} = \mathbf{d}_{M\times 1}$ (Normalmente $N > M$)

$$\begin{bmatrix} x(0) & x(-1) & \cdots & x(-M+1) \\ x(1) & x(0) & \cdots & x(-M+2) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-1) & x(N-2) & \cdots & x(N-M) \end{bmatrix} \begin{bmatrix} \omega_0^* \\ \omega_1^* \\ \vdots \\ \omega_{M-1}^* \end{bmatrix} + \begin{bmatrix} e(0) \\ e(1) \\ \vdots \\ e(N-1) \end{bmatrix} = \begin{bmatrix} d(0) \\ d(1) \\ \vdots \\ d(N-1) \end{bmatrix}$$

Solución LS

- Solución única

  a) $N \geq M$ (Sistema sobredeterminado)

  b) Rank($\mathbf{X}^H\mathbf{X}$)=M (columnas linealmente independientes)

$$\boxed{\hat{\mathbf{w}} = \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H\mathbf{d}}$$

- Infinitas soluciones
  - $N < M$ (Sistema indeterminado)
  - Solución de norma mínima

$$\boxed{\hat{\mathbf{w}} = \mathbf{X}^H\left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{d}}$$

Pseudoinversa

- Definiendo el operador pseudoinversa $\mathbf{X}^+$

$$\boxed{\hat{\mathbf{w}} = \mathbf{X}^+\mathbf{d}}$$

$$\mathbf{X}^+ = \begin{cases} \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H, & \text{Si } N > M \\ \mathbf{X}^H\left(\mathbf{X}^H\mathbf{X}\right)^{-1}, & \text{Si } N < M \\ \mathbf{X}^{-1}, & \text{Si } N = M \end{cases}$$

- Cuando $N \geq M$ y $\text{Rank}(\mathbf{X}^H\mathbf{X}) = M$

$$\hat{\mathbf{d}} = \mathbf{X}\hat{\mathbf{w}}^* = \underbrace{\mathbf{X}\left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H}_{\mathbf{P}_x}\mathbf{d}$$

$\mathbf{P}_x$: matriz de proyección en el sub-espacio de las columnas de $\mathbf{X}$

LS/Filtro de Wiener

- Ecuaciones normales

$$\left(\mathbf{X}^H\mathbf{X}\right)\hat{\mathbf{w}} = \mathbf{X}^H\mathbf{d}$$

$$\Phi_{M\times M}\hat{\mathbf{w}} = \theta_{M\times 1}$$

- Interpretación de Φ y θ
  - Φ: estima de la autocorrelación

$$\Phi = \mathbf{X}^H\mathbf{X} = \sum_{n=0}^{N-1}\mathbf{x}_n\mathbf{x}_n^H$$

  - θ: estima de la correlación cruzada

$$\theta = \mathbf{X}^H\mathbf{d} = \sum_{n=0}^{N-1} d(n)\mathbf{x}_n^*$$

Propiedades

- La matriz Φ
  - es hermítica ($\Phi = \Phi^H$)
  - es semidefinida positiva ($\mathbf{x}^H \Phi \mathbf{x} \geq 0$)
- El estimador LS es insesgado si el error tiene media nula
- Si el error es blanco, de media nula y varianza σ^2
  - $E\left[(\hat{\mathbf{w}} - \mathbf{w}_o)(\hat{\mathbf{w}} - \mathbf{w}_o)^H\right] = \sigma^2 \Phi^{-1}$
  - El estimador LS es el mejor estimador lineal insesgado (BLUE)
- Si además el error es gaussiano, el estimador LS alcanza el límite de Cramer-Rao (es el mejor estimador, lineal o no lineal)

LS ponderado (*Weighted Least Squares*)

- El WLS introduce una matriz de ponderación

$$J(\mathbf{w}) = (\mathbf{d} - \mathbf{X}\mathbf{w}^*)^H \mathbf{A} (\mathbf{d} - \mathbf{X}\mathbf{w}^*) = ||\mathbf{e}||_{\mathbf{A}}^2$$

  $\mathbf{A}$: hermítica positiva semidefinida

- $\mathbf{A}$ diagonal: se pondera cada error de forma distinta

$$J(\mathbf{w}) = \sum_{n=0}^{N-1} a(n)|e(n)|^2$$

- Ecuaciones normales

$$\mathbf{X}^H \mathbf{A} \mathbf{X} \mathbf{w} = \mathbf{X}^H \mathbf{A} \mathbf{d}$$

- Solución (si $N \geq M$ y Rank($\mathbf{X}^H\mathbf{A}\mathbf{X}$)=$M$)

$$\hat{\mathbf{w}} = \left(\mathbf{X}^H \mathbf{A} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{A} \mathbf{d}$$

LS regularizado

- Función de coste

$$J(\mathbf{w}) = \mathbf{w}^H \mathbf{A}\mathbf{w} + ||\mathbf{d} - \mathbf{X}\mathbf{w}^*||^2$$

  $\mathbf{A}$: hermítica positiva semidefinida

- Solución

$$\hat{\mathbf{w}} = \left(\mathbf{X}^H \mathbf{X} + \mathbf{A}\right)^{-1} \mathbf{X}^H \mathbf{d}$$

- Típicamente: $\mathbf{A} = \alpha \mathbf{I}$

$$\hat{\mathbf{w}} = \left(\mathbf{X}^H \mathbf{X} + \alpha \mathbf{I}\right)^{-1} \mathbf{X}^H \mathbf{d}$$

- Si la matriz $\mathbf{X}^H\mathbf{X}$ está mal condicionada el LS regularizado reduce la amplificación de ruido (a cambio de sesgar el estimador)

$$\underbrace{\frac{\lambda_{max} + \alpha}{\lambda_{min} + \alpha}}_{\text{cond}(\mathbf{X}^H\mathbf{X}+\alpha\mathbf{I})} < \underbrace{\frac{\lambda_{max}}{\lambda_{min}}}_{\text{cond}(\mathbf{X}^H\mathbf{X})}$$

Algoritmo RLS: Introducción

- Estima recursiva de la solución LS
- Problema: estimar la media de N muestras $x(n)$, $n = 1, \cdots, N$.

$$\hat{x}_N = \frac{1}{N} \sum_{n=1}^{N} x(n)$$

- Si se dispone de una nueva muestra $x(N+1)$

$$\hat{x}_{N+1} = \frac{1}{N+1} \left(N\hat{x}_N + x(N+1) \right)$$

- Algoritmo RLS: resuelve de modo similar el caso del estimador LS
  - ¿Como se actualiza la estima LS obtenida con N datos cuando se dispone de un nuevo dato, $x(N+1)$,$d(N+1)$?

Función de coste y solución LS

- Función de coste (determinista) en el instante n

$$J_n(\mathbf{w}) = \sum_{i=1}^{N} \lambda^{n-i}|e(i)|^2 = |e(n)|^2 + \lambda|e(n-1)|^2 + \cdots + \lambda^{n-i}|e(1)|^2$$

$e(n) = d(n) - \mathbf{w}^H\mathbf{x}_n$

λ: factor de olvido exponencial ($0 < \lambda < 1$)

- La solución cumple las ecuaciones normales

$$\Phi_n\mathbf{w}_n = \theta_n \quad \Rightarrow \quad \mathbf{w}_n = \Phi_n^{-1}\theta_n$$

$$\Phi_n = \sum_{i=1}^{N} \lambda^{n-i}\mathbf{x}_i\mathbf{x}_i^H, \qquad \theta_n = \sum_{i=1}^{N} \lambda^{n-i}\mathbf{x}_i d(n)$$

Cuestiones preliminares

- En la solución obtenida para cada instante n intervienen todos los datos hasta ese instante (aunque ponderados de distinta manera)
- La estima LS es determinista; no obstante, si $\lambda = 1$ y los procesos que intervienen son ergódicos

$$\left.\begin{array}{l} \lim_{n\to\infty} \frac{1}{n}\Phi_n = \mathbf{R} \\ \lim_{n\to\infty} \frac{1}{n}\theta_n = \mathbf{p} \end{array}\right. \Rightarrow \lim_{n\to\infty} \text{LS} = \text{Wiener}$$

- La inversión de la matriz de autotocorrelación para cada n necesitaría $O(M^3)$ operaciones y $O(M^3)$ posiciones de memoria
  - ¿Se puede hacer el cálculo de forma recursiva?

Cálculo recursivo de Φ_n^{-1}

$$\Phi_{n+1} = \sum_{i=1}^{N+1} \lambda^{n+1-i} \mathbf{x}_i \mathbf{x}_i^H = \lambda \Phi_n + \mathbf{x}_{n+1} \mathbf{x}_{n+1}^H$$

$$\Phi_{n+1}^{-1} = \left(\lambda \Phi_n + \mathbf{x}_{n+1} \mathbf{x}_{n+1}^H \right)^{-1}$$

- Aplicando el *"Matrix Inversion Lemma"*

$$\Phi_{n+1}^{-1} = \lambda^{-1} \Phi_n^{-1} - \frac{\lambda^{-2} \Phi_n^{-1} \mathbf{x}_{n+1} \mathbf{x}_{n+1}^H \Phi_n^{-1}}{1 + \lambda^{-1} \mathbf{x}_{n+1}^H \Phi_n^{-1} \mathbf{x}_{n+1}}$$

- Definiciones

$$\mathbf{P}_n = \Phi_n^{-1} \text{ Inversa de la autocorrelación}$$

$$\mathbf{k}_{n+1} = \frac{\lambda^{-1} \mathbf{P}_n \mathbf{x}_{n+1}}{1 + \lambda^{-1} \mathbf{x}_{n+1}^H \mathbf{P}_n \mathbf{x}_{n+1}} \text{ Vector de ganancia}$$

Cálculo recursivo de Φ_n^{-1} (II)

- Ecuación de Ricatti para el RLS

$$\mathbf{P}_{n+1} = \lambda^{-1}\left(\mathbf{P}_n - \mathbf{k}_{n+1}\mathbf{x}_{n+1}^H\mathbf{P}_n\right)$$

- Vector de ganancia

$$\mathbf{k}_{n+1} = \mathbf{P}_{n+1}\mathbf{x}_{n+1}$$

  son los datos blanqueados por la inversa de la matriz de autocorrelación

Actualización del filtro

- Solución de mínimos cuadrados

$$\mathbf{w}_{n+1} = \Phi_{n+1}^{-1}\theta_{n+1}$$

que puede desarrollarse como

$$\mathbf{w}_{n+1} = \mathbf{P}_{n+1}\theta_{n+1} = \lambda\mathbf{P}_{n+1}\theta_n + \mathbf{P}_{n+1}\mathbf{x}_{n+1}d^*(n+1)$$

- Teniendo en cuenta la recursión de la ecuación de Ricatti

$$\mathbf{w}_{n+1} = \underbrace{\mathbf{P}_n\theta_n}_{\mathbf{w}_n} - \mathbf{k}_{n+1}\mathbf{x}_{n+1}^H\mathbf{P}_n\theta_n + \underbrace{\mathbf{P}_{n+1}\mathbf{x}_{n+1}}_{\mathbf{k}_{n+1}} d^*(n+1)$$

- Expresión final

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mathbf{k}_{n+1}\left[d^*(n+1) - \mathbf{x}_{n+1}^H\mathbf{w}_n\right]$$

Errores *"a priori"* y *"a posteriori"*

- Error *a priori* (innovación)

$$\alpha(n+1) = d(n+1) - \mathbf{w}_n^H \mathbf{x}_{n+1}$$

error que comete el filtro estimado sin usar el nuevo dato

- Error *a posteriori*

$$e(n+1) = d(n+1) - \mathbf{w}_{n+1}^H \mathbf{x}_{n+1}$$

error utilizando el nuevo dato

- En la función de coste se minimizan los errores *a posteriori*
- En la recursión del RLS aparecen los errores *a priori*

Convergencia del RLS

- Convergencia en media
  - El RLS converge a la solución de mínimos cuadrados
  - Si $\lambda = 1$ y las señales son ergódicas el RLS converge en media a Wiener
- Convergencia en error cuadrático ($\lambda = 1$, ergodicidad)

$$J(n) \approx J_{min}\left(1 + \frac{M}{n}\right)$$

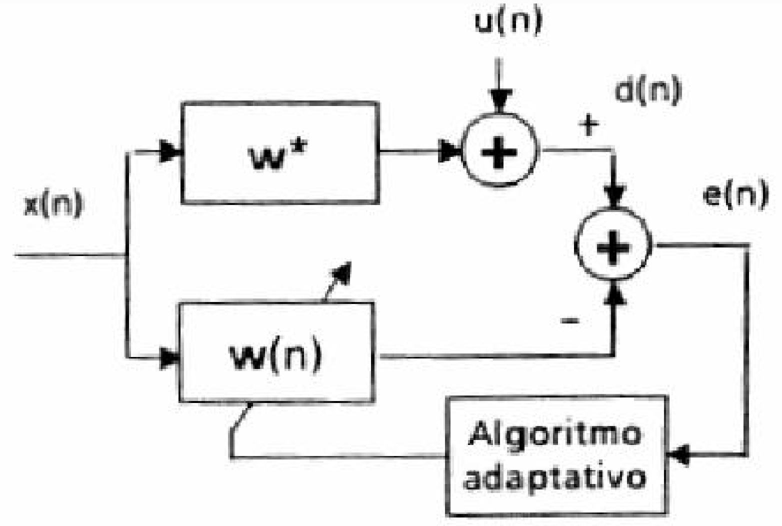
  - El RLS converge en aproximadamente $2M$ iteraciones
  - El RLS no tiene desajuste (si $\lambda = 1$)

RLS/LMS

- RLS: $\mathbf{w}_{n+1} = \mathbf{w}_n + \mathbf{k}_{n+1} \left[d^*(n+1) - \mathbf{x}_{n+1}^H \mathbf{w}_n \right]$
- LMS: $\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \mathbf{x}_n \left[d^*(n) - \mathbf{x}_n^H \mathbf{w}_n \right]$
- En ambos casos se actualiza el filtro mediante un término de error
- Para obtener $\mathbf{w}_{n+1}$
  - El LMS utiliza los datos en n
  - El RLS utiliza todos los datos (a través de $\mathbf{K}_{n+1}$)
- En el LMS el error se multiplica por $\mu \mathbf{x}_n$
- En el RLS el error se multiplica por

$$\mathbf{k}_{n+1} = \Phi_{n+1}^{-1} \mathbf{x}_{n+1} \Rightarrow \begin{cases} \text{Los datos se blanquean en cada iteración} \\ \text{Desacoplo de la convergencia de } \mathbf{w} \end{cases}$$

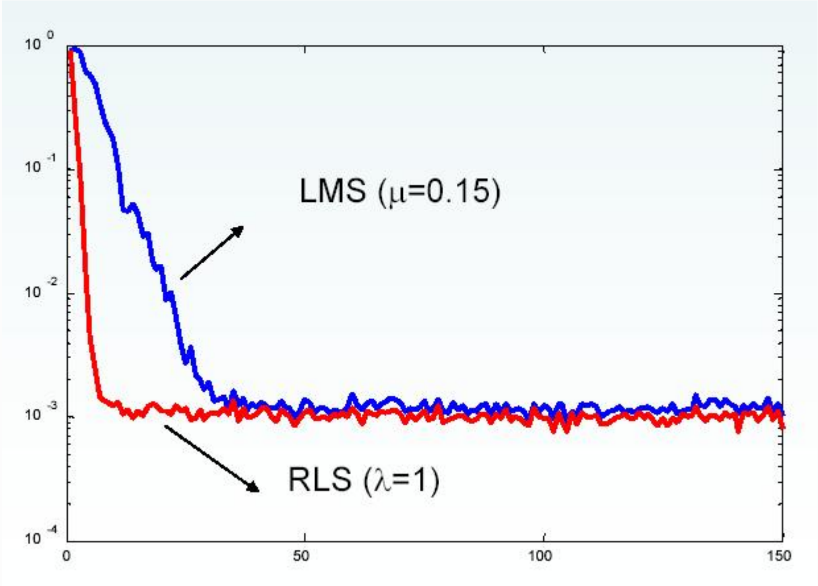
Ejemplo: Identificación



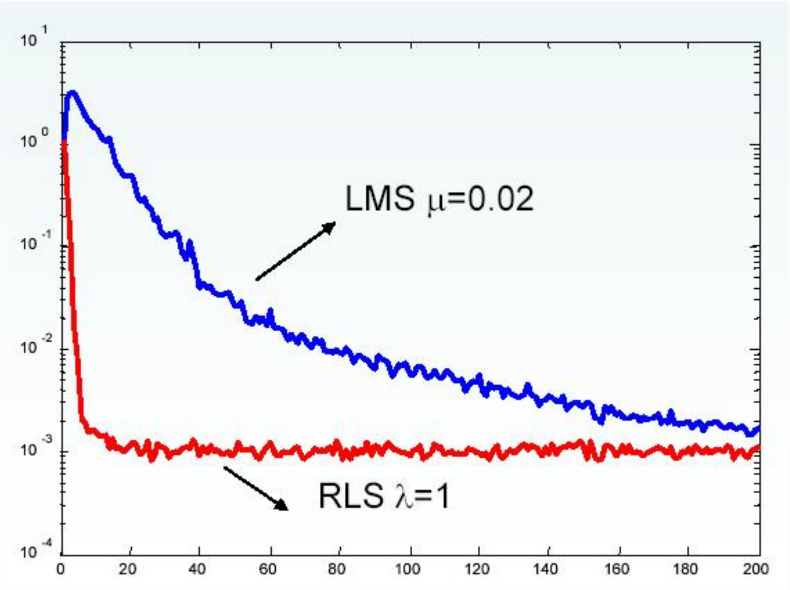
- Sistema: $W^*(z) = 1 + 0.5z^{-1}$
- Entrada correlada con un factor r, es decir, $\mathbf{R} = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$
- $u(n)$: ruido aditivo balnco, gaussiano de media cero, varianza 0.001 e independiente de la entrada

Ejemplo: Identificación (II)

Coeficiente de correlación de la entrada $r = 0.1$



Coeficiente de correlación de la entrada $r = 0.8$



Resumen del RLS

- Parámetros iniciales

$$M = \text{n}^{\text{o}} \text{ coef.}, \;\; \mathbf{P}_0 = \delta^{-1}\mathbf{I}, \;\; \delta < 0.01\sigma_x^2, \;\; \mathbf{w}_0 = \mathbf{0}_{M\times 1}, \;\; \lambda \leq 1$$

- Iteraciones
  - $\mathbf{k}_{n+1} = \dfrac{\lambda^{-1}\mathbf{P}_n\mathbf{x}_{n+1}}{1 + \lambda^{-1}\mathbf{x}_{n+1}^H\mathbf{P}_n\mathbf{x}_{n+1}}$
  - $\alpha(n+1) = d(n+1) - \mathbf{w}_n^H\mathbf{x}_{n+1}$
  - $\mathbf{w}_{n+1} = \mathbf{w}_n + \mathbf{k}_{n+1}\alpha^*(n+1)$
  - $\mathbf{P}_{n+1} = \lambda^{-1}\left(\mathbf{P}_n - \mathbf{k}_{n+1}\mathbf{x}_{n+1}^H\mathbf{P}_n\right)$

El RLS en ambientes no estacionarios

- Aproximadamente la memoria del RLS viene dada por

$$\text{Memoria} = \frac{1}{1-\lambda}$$

- Si $\lambda = 1$ la memoria es infinita (ambientes estacionarios)
- Si $\lambda < 1$
  - Aumenta la capacidad de seguimiento
  - Aumenta el deasjuste

$$D = \frac{1-\lambda}{1+\lambda} M \text{ (para valores de } \lambda \text{ cercanos a 1)}$$

Capacidad de seguimiento LMS/RLS

- LMS: capacidad de seguimiento determinada por el paso de adaptación μ

$$\mu \uparrow \quad \Rightarrow \quad \begin{cases} \text{mejor seguimiento} \\ \text{mayor desajuste} \end{cases}$$

- RLS: capacidad de seguimiento determinada por el factor de olvido λ

$$\lambda \downarrow \quad \Rightarrow \quad \begin{cases} \text{mejor seguimiento} \\ \text{mayor desajuste} \end{cases}$$

- Para un mismo desajuste (que fija μ y λ) el LMS tiene habitualmente un mejor comportamiento en ambientes no estacionarios.

Conclusiones

- El RLS obtiene de manera recursiva el estimador LS
- Utiliza todos los datos pasados (ponderados por un factor de olvido)
- Evita la inversión de la matriz de autocorrelación en cada paso: se actualiza Φ_{n+1}^{-1} a partir de Φ_n^{-1}
- Insensible a la dispersión de autovalores de la matriz de autocorrelación
- El valor de ganancia $\mathbf{k}_n$ desacopla la convergencia de los coeficientes
- Para $\lambda = 1$ y procesos ergódicos el RLS converge a Wiener y no existe desajuste
- En comparación con el LMS
  - Velocidad de convergencia muy superior (del orden de $2M$)
  - Gasto computacional superior (versiones rápidas: Fast-RLS)
  - Para un mismo desajuste se comporta peor en ambientes no estacionarios
  - El RLS puede presentar problemas de estabilidad numérica