

Estimación y Predicción en Series Temporales

Filtro de Wiener

Departamento de Procesamiento de Señales

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería

2022

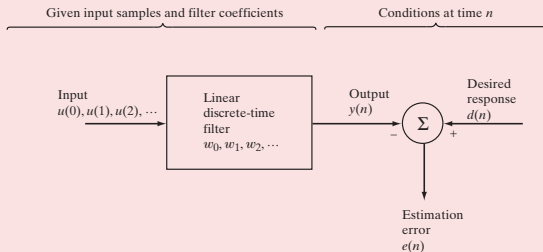
Biblio: Haykin, Adaptive Filter Theory 4^a edición, Capítulo 2.

Hayes, Statistical Digital Signal Processing and Modeling (1996), Cap. 7.

Introducción

Estudiaremos una clase de filtros en tiempo discreto, óptimos en un sentido que definiremos más adelante, llamados *Filtros de Wiener*.

Filtrado lineal óptimo: planteo del problema



- $\{u(n)\}, \{d(n)\}$: realizaciones de procesos estocásticos conjuntamente WSS, de media nula (si no, basta con restárselas).
- $y(n)$: estimación de la respuesta deseada en tiempo n .
- $e(n) = d(n) - y(n)$: error de estimación en tiempo n .

- **Sobre el filtro**

- *Restricciones:* lineal y discreto.
- *Opciones:* FIR o IIR. Trataremos la teoría para IIRs (siendo los FIRs un caso particular), pero nos focalizaremos en los FIRs por ser inherentemente estables.
- **Obs:** Cuando tratemos con filtros adaptivos, tener que lidiar con IIRs y sus posibles inestabilidades hacen del problema difícil de manejar. Por esta razón en general en filtros adaptivos se utilizan FIRs, a pesar que los IIR son mucho menos demandantes computacionalmente.

- **Sobre el criterio estadístico de optimalidad**

- $\mathbb{E}[|e(n)|]$: convexo, no diferenciable. Corresponde a distribuciones de Laplace.
- $\mathbb{E}[|e(n)|^2]$: convexo, diferenciable. Corresponde a distribuciones Gaussianas. Es matemáticamente más sencillo de tratar (podemos calcular gradientes).
- Otros.

Posibles Aplicaciones

Dependiendo de cómo están vinculadas $u(n)$ y $d(n)$, varios problemas fundamentales pueden ser planteados como un filtrado de Wiener:

- **Filtrado:** $u(n) = d(n) + v(n)$, con $v(n)$ ruido. Estimar $d(n)$ a partir de muestras ruidosas con un filtro causal, i.e. a partir de las muestras $u(n), u(n-1), u(n-2), \dots$.
- **Suavizado:** idem salvo que el filtro puede ser no causal. Por ejemplo, estimar $d(n)$ offline a partir de todo el registro de datos disponibles (previos y posteriores a n).
- **Predicción:** si $d(n) = u(n+1)$ y el filtro es causal, se busca estimar $u(n+1)$ a partir de $u(n)$ y sus muestras previas.
- **Deconvolución:** $u(n) = h * d(n) + v(n)$. El objetivo es encontrar algún filtro de respuesta al impulso $g(n)$ tal que podamos estimar $\hat{d}(n) = g * u(n)$.
- Otros.

- **Objetivo:** encontrar w_0, w_1, w_2, \dots tales que se minimice $J_n(\mathbf{w}) = \mathbb{E}[|e(n)|^2] = J(\mathbf{w})$.
- Desarrollaremos la solución matemática a este problema de optimización estadística abordando dos enfoques diferentes pero complementarios:
 - **El principio de ortogonalidad** (característicos de los espacios de Hilbert)
 - **La superficie de performance del error $e(n)$.**

Las ecuaciones normales o de Wiener-Hopf (1)

- Entrada: $u(0), u(1), u(2), \dots$
Resp. al impulso del filtro: w_0, w_1, w_2, \dots } complejas e in

- Salida del filtro:

$$\begin{aligned} y(n) &= \sum_{k=0}^{+\infty} w_k^* u(n-k), \quad n = 0, 1, 2, \dots \\ &= \mathbf{w}^H \mathbf{u}(n), \quad \text{con } \mathbf{w}^T = [w_0, w_1, w_2, \dots], \\ &\quad \mathbf{u}(n)^T = [u(n), u(n-1), u(n-2), \dots] \end{aligned}$$

- $e(n) = d(n) - y(n)$.
- Recordemos que $\{u(n)\}, \{d(n)\}$ son realizaciones de procesos estocásticos conjuntamente WSS, de media nula

Las ecuaciones normales o de Wiener-Hopf (2)

Queremos ver bajo qué condiciones en \mathbf{w} (o cual es el filtro tal que) $J(\mathbf{w}) = \mathbb{E}[e^*(n)e(n)]$ alcanza el mínimo.

Definamos $\mathbf{p}^T = [p(0), p(-1), p(-2), \dots] = \mathbb{E}[\mathbf{u}(n)d^*(n)]$.

Ejercicio:

- 1 Escribir $J(\mathbf{w})$ como función de \mathbf{w} , $\mathbf{R} = \mathbb{E}[\mathbf{u}(n)\mathbf{u}(n)^H]$, \mathbf{p} y $\sigma_d^2 = \mathbb{E}[|d(n)|^2]$, y verificar que es cuadrática y convexa en \mathbf{w} .
- 2 Escribir el sistema de ecuaciones que debe verificar $\mathbf{w}_o = \arg \min_{\mathbf{w}} J(\mathbf{w})$, conocido como *ecuaciones normales*.

$$\begin{aligned} J(\mathbf{w}) &= \mathbb{E}[(d^*(n) - \mathbf{u}^H(n)\mathbf{w})(d(n) - \mathbf{w}^H\mathbf{u}(n))] \\ &= \mathbb{E}[|d(n)|^2] + \mathbf{w}^H \mathbb{E}[\mathbf{u}(n)\mathbf{u}(n)^H] \mathbf{w} - \mathbf{w}^H \mathbb{E}[\mathbf{u}(n)d^*(n)] - \mathbb{E}[d(n)\mathbf{u}^H(n)] \mathbf{w} \\ &= \mathbf{w}^H \mathbf{R} \mathbf{w} - \mathbf{p}^H \mathbf{w} - \mathbf{w}^H \mathbf{p} + \sigma_d^2. \end{aligned}$$

Es convexa ya que \mathbf{R} es def. positiva, por lo que admite un único mínimo local.

Sabemos que $\nabla_{\mathbf{w}} = 2 \frac{\partial}{\partial \mathbf{w}^*}$ (ver Haykin, Apéndice B).

Breve paréntesis: $\mathbb{E}[X^H Y]$ como producto interno

- X_1, X_2, Y vectores aleatorios, $\mathbb{E}[X_1] = \mathbb{E}[X_2] = \mathbb{E}[Y] = 0$, $a \in \mathbb{C}$.
- $\langle X, Y \rangle := \mathbb{E}[X^H Y]$ define un producto interno:
 - $\langle X, X \rangle \geq 0$, y $\langle X, X \rangle = 0 \Leftrightarrow X = \mathbf{0}$ c.s.
 - $\langle X, Y \rangle = \langle Y, X \rangle^*$
 - $\langle aX_1 + X_2, Y \rangle = a^* \langle X_1, Y \rangle + \langle X_2, Y \rangle$ (linealidad de $\mathbb{E}[\cdot]$).
- $\langle X, Y \rangle = \mathbb{E}[X^H Y] = 0 \Leftrightarrow X, Y$ no correlacionados (o ortogonales para este p. i.).

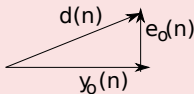
$$\begin{aligned} \mathbf{R} \mathbf{w}_o = \mathbf{p} &\Leftrightarrow \mathbb{E}[\mathbf{u}(n) \mathbf{u}(n)^H] \mathbf{w}_o = \mathbb{E}[\mathbf{u}(n) d^*(n)] \\ &\Leftrightarrow \mathbb{E}[\mathbf{u}(n) (d^*(n) - \mathbf{u}(n)^H \mathbf{w}_o)] = \mathbb{E}[\mathbf{u}(n) e_o^*(n)] = \mathbf{0}. \\ &\Leftrightarrow \mathbb{E}[u(n-k) e_o^*(n)] = 0 \quad \forall k = 0, 1, 2, \dots \end{aligned}$$

\Rightarrow La CNS para que J alcance el mínimo es que el error $e(n) = e_o(n)$ sea ortogonal a todas las muestras involucradas al tiempo n .

Principio de ortogonalidad (cont.)

Cuando el filtro opera en condición óptima $\mathbf{w} = \mathbf{w}_o$, la estimación de la respuesta deseada y el error de estimación son ortogonales:

$$\mathbb{E}[y_o(n)e_o^*(n)] = \mathbb{E}[\mathbf{w}_o^H \mathbf{u}(n)e_o^*(n)] = \mathbf{w}_o^H \mathbb{E}[\mathbf{u}(n)e_o^*(n)] = 0.$$



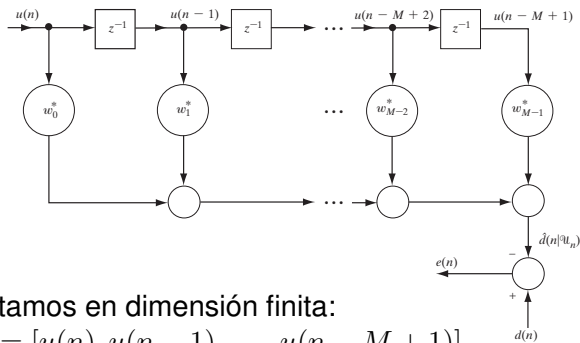
Error cuadrático medio mínimo

$$e_o(n) = d(n) - y_o(n), \quad y_o(n) = \hat{d}(n), \quad J_{min} = \mathbb{E}[|e_o(n)|^2]$$

Por el principio de ortogonalidad, sabemos que $e_o(n) \perp y_o(n)$, de dónde

$$d(n) = e_o(n) + y_o(n) \Rightarrow \mathbb{E}[|d(n)|^2] = \mathbb{E}[|e_o(n)|^2] + \mathbb{E}[|y_o(n)|^2]$$

Solución de las ecs. normales para filtros transversales



Ahora estamos en dimensión finita:

$$\mathbf{u}(n)^T = [u(n), u(n-1), \dots, u(n-M+1)]$$

$$\mathbf{w}^T = [w_0, w_1, \dots, w_{M-1}], \quad \mathbf{w}_o^T = [w_{o,0}, w_{o,1}, \dots, w_{o,M-1}]$$

$$\mathbf{p}^T = [p(0), p(-1), \dots, p(-M+1)]$$

$\mathbf{w}_o = \mathbf{R}^{-1} \mathbf{p}$ \rightarrow Necesitamos conocer: matriz de correlación de $\mathbf{u}(n)$ y las correlaciones cruzadas entre $u(n-k)$ y $d(n)$, $k = 0, 1, \dots, M-1$.

Superficie de performance del error

$$J(\mathbf{w}) = \mathbf{w}^H \mathbf{R} \mathbf{w} - \mathbf{p}^H \mathbf{w} - \mathbf{w}^H \mathbf{p} + \sigma_d^2.$$

$$\begin{aligned} J_{min} = J(\mathbf{w}_o) &= \mathbf{p}^H \mathbf{R}^{-H} \mathbf{R} \mathbf{R}^{-1} \mathbf{p} - \mathbf{p}^H \mathbf{R}^{-1} \mathbf{p} - \mathbf{p}^H \mathbf{R}^{-H} \mathbf{p} + \sigma_d^2. \\ &= \sigma_d^2 - \mathbf{p}^H \mathbf{R}^{-1} \mathbf{p} \\ &= \sigma_d^2 - \mathbf{p}^H \mathbf{w}_o. \end{aligned}$$

Se puede ver fácilmente que

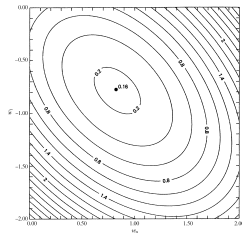
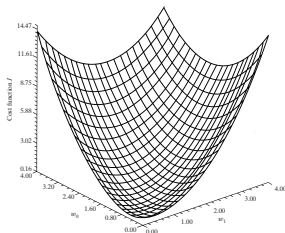
$$J(\mathbf{w}) = J_{min} + (\mathbf{w} - \mathbf{w}_o)^H \mathbf{R} (\mathbf{w} - \mathbf{w}_o)$$

Forma canónica de la superficie de error

Sabemos que $\mathbf{R} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H$ (\mathbf{R} hermítica, teorema espectral).
Luego,

$$\begin{aligned} J(\mathbf{w}) &= J_{min} + (\mathbf{w} - \mathbf{w}_o)^H \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H (\mathbf{w} - \mathbf{w}_o) \\ &= J_{min} + \mathbf{v}^H \mathbf{\Lambda} \mathbf{v}, \quad \text{con } \mathbf{v} = \mathbf{Q}^H (\mathbf{w} - \mathbf{w}_o) \text{ la proyección de } (\mathbf{w} - \mathbf{w}_o) \\ &\quad \text{en los ejes principales de } \mathbf{R} \text{ (} \therefore \text{ de la superficie de error).} \end{aligned}$$

Superficie de performance del error: ejemplo de un filtro con dos taps



En la forma canónica, J_{min} se alcanza en $(\mathbf{w} - \mathbf{w}_o) = (0, 0)$, y los ejes están orientados según los ejes principales de las elipses de las curvas de nivel.

Filtro de Wiener IIR (no causal)

- Dado el proceso $\{u(n)\}$, buscamos el filtro $\{h(n)\}$ que produzca la salida más cercana a $\{d(n)\}$ en el sentido MSE.
- Si bien la formulación es la misma que para los FIR, ahora tenemos un **número infinito de incógnitas $\{h(n)\}$** :

$$e(n) = d(n) - \sum_{k=-\infty}^{+\infty} h(k)u(n-k), \quad n \in \mathbb{Z}$$

$$J(\{h\}) = \mathbb{E}[|e(n)|^2] = \mathbb{E}[e(n)e^*(n)].$$

Calculando $\nabla_{h(k)} J$ tenemos

$$\nabla_{h(k)} J = 2 \frac{\partial J}{\partial h^*(k)} = -2\mathbb{E}[e(n)u^*(n-k)], \quad k \in \mathbb{Z}.$$

Igualando a cero obtenemos el **principio de ortogonalidad**:

$$\mathbb{E}[e(n)u^*(n-k)] = 0, \quad k \in \mathbb{Z}.$$

Remplazando $e(n)$ y usando la estacionaridad conjunta de $\{d(n)\}$ y $\{u(n)\}$ tenemos las **ecuaciones de Wiener-Hopf para el filtro IIR**:

$$\sum_{k=-\infty}^{+\infty} h(k)r_u(n-k) = r_{du}(n), \quad n \in \mathbb{Z}$$

Si bien no podemos proceder a expresar el sistema matricialmente, podemos expresarlo fácilmente en Fourier ya que $h * r_u(n) = r_{du}(n)$:

$$H(\omega)S_u(\omega) = S_{du}(\omega) \Rightarrow H(\omega) = \frac{S_{du}(\omega)}{S_u(\omega)}$$

De esta forma, procediendo como hicimos para el caso FIR,

$$J_{min} = r_d(0) - \sum_{k=-\infty}^{+\infty} h(k)r_{du}^*(k) = r_d(0) - \frac{1}{2\pi} \int_{-\pi}^{\pi} H(\omega)S_{du}^*(\omega)d\omega,$$

$$J_{min} = \frac{1}{2\pi} \int_{-\pi}^{\pi} [S_u(\omega) - H(\omega)S_{du}^*(\omega)] d\omega$$

Filtro de Wiener IIR causal

En este caso el filtro debe verificar $h(n) = 0$ para todo $n < 0$.
La salida del filtro viene dada por

$$y(n) = h * u(n) = \sum_{k=0}^{+\infty} h(k)u(n-k).$$

De esta forma las ecuaciones de Wiener-Hopf para un IIR causal quedan

$$\sum_{k=0}^{+\infty} h(k)r_u(n-k) = r_{du}(n), \quad n \in \mathbb{N}$$

- La restricción $n \geq 0$ es importante ya que ahora no podemos expresar $r_{du}(n)$ como la convolución de h y u .
- Las ecuaciones de Wiener-Hopf para este caso se pueden resolver, pero el procedimiento es más complejo (ver Hayes, §7.3.2).