Procesamiento de señales I

86.51

Cecilia G. Galarza

FIUBA
Laboratorio de Procesamiento de Señales y Comunicaciones

2do Cuatrimestre 2022

Estimación Lineal

Estimación de Menor Error Cuadrático Medio

Sean $\mathbf{x} \in \mathbb{C}^n$ e $\mathbf{y} \in \mathbb{C}^m$ dos vectores aleatorios, cuya estadística conjunta es conocida. Se desea estimar $\mathbf{x}$ a partir de la observación de una realización de $\mathbf{y}$, es decir, se busca una función

$$\begin{aligned} g(\mathbf{y}) \quad &: \mathbb{C}^m \rightarrow \mathbb{C}^n \text{ tal que} \\ &\hat{\mathbf{x}} = g(\mathbf{y}) \quad \textit{esté cercano a } \mathbf{x} \end{aligned}$$

Un posible criterio de *cercanía* es el *Error Cuadrático Medio*[1] mas

$$MSE(\underline{\hat{\mathbf{x}}}) = \mathbb{E}\left[\|\mathbf{x} - \underline{\hat{\mathbf{x}}}\|^2\right]$$

usado

Sea $\hat{\mathbf{x}}_{opt}$ el estimador que minimiza $MSE(\hat{\mathbf{x}})$. Se dice que $\hat{\mathbf{x}}_{opt}$ es el estimador MMSE (*Minimun Mean Square Error*).

[1] MSE, Mean Square Error, por sus siglas en inglés.

El MMSE es la esperanza condicional

arg q minimiza

Estimador MMSE

$$\hat{\mathbf{x}}_{opt} = \arg \min_{\hat{\mathbf{x}}=g(\mathbf{y})} \mathbb{E}\left[\|\mathbf{x} - \hat{\mathbf{x}}\|^2\right]$$

el q resulta es:

$$\hat{\mathbf{x}}_{opt} = \mathbb{E}\left[\mathbf{x}|\mathbf{y}\right]$$

el computo no es sencillo

El MMSE es la esperanza condicional: demostración I

Para simplificar notación, llamamos:

$$\mu(\mathbf{y}) = \mathbb{E}[\mathbf{x}|\mathbf{y}] \qquad \text{y} \qquad \Delta(\mathbf{y}) = \mu(\mathbf{y}) - \hat{\mathbf{x}}.$$

Luego,

$$\mathbf{x} - \hat{\mathbf{x}} = \mathbf{x} - \mu(\mathbf{y}) + \mu(\mathbf{y}) - \hat{\mathbf{x}} = \mathbf{x} - \mu(\mathbf{y}) + \Delta(\mathbf{y}).$$

Retomando la función MSE

$$\begin{aligned} MSE = \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}})^*(\mathbf{x} - \hat{\mathbf{x}})] &= \mathbb{E}\left[(\mathbf{x} - \mu(\mathbf{y}) + \Delta(\mathbf{y}))^*(\mathbf{x} - \mu(\mathbf{y}) + \Delta(\mathbf{y}))\right] \\ = \mathbb{E}\left[\|\mathbf{x} - \mu(\mathbf{y})\|^2\right] &\neq f(\hat{\mathbf{x}}) \qquad ① \\ + \mathbb{E}[(\mathbf{x} - \mu(\mathbf{y}))^*\Delta(y)] + \mathbb{E}[\Delta(\mathbf{y})^*(\mathbf{x} - \mu(\mathbf{y}))] &= 0 \qquad ② \\ + \mathbb{E}[\Delta(\mathbf{y})^*\Delta(\mathbf{y})] & \qquad ③ \end{aligned}$$

El MMSE es la esperanza condicional: demostración II

- ①: No depende de $\hat{\mathbf{x}}$.
- ②: Recordemos que $\forall\ \mathbf{x}, \mathbf{y}$, $\mathbb{E}[\mathbf{x}] = \mathbb{E}[\mathbb{E}[\mathbf{x}|\mathbf{y}]]$.

$$\begin{aligned}
\mathbb{E}[(\mathbf{x}-\mu(\mathbf{y}))^*\Delta(\mathbf{y})] &= \mathbb{E}\left[\mathbb{E}[(\mathbf{x}-\mu(\mathbf{y}))^*\Delta(\mathbf{y})|\mathbf{y}]\right] \qquad \mu(\mathbf{y})^*\Delta(\mathbf{y}) \\
&= \mathbb{E}\left[\mathbb{E}[\mathbf{x}^*\underbrace{\Delta(\mathbf{y})}_{\text{cte en } y}|\mathbf{y}]\right] - \mathbb{E}\left[\mathbb{E}[\mu(\mathbf{y})^*\Delta(\mathbf{y})|\mathbf{y}]\right] \\
&= \mathbb{E}\left[\underbrace{\mathbb{E}[\mathbf{x}^*|\mathbf{y}]}_{\mu(\mathbf{y})^*}\Delta(\mathbf{y})\right] - \mathbb{E}\left[\mu(\mathbf{y})^*\Delta(\mathbf{y})\right] = 0
\end{aligned}$$

- ③:

$$\mathbb{E}[\Delta(\mathbf{y})^*\Delta(\mathbf{y})] = \mathbb{E}\left[\|\mathbb{E}[\mathbf{x}|\mathbf{y}] - \hat{\mathbf{x}}\|^2\right] \geq 0$$

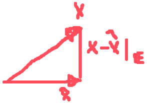
Luego, minimizar MSE con respecto a $\hat{\mathbf{x}}$ equivale a minimizar ③. Para ello,

$$\hat{\mathbf{x}}_{opt} = \mathbb{E}[\mathbf{x}|\mathbf{y}].$$

Principio de ortogonalidad

$\hat{X}_{opt} = E(x|y)$

Los vectores $\hat{\mathbf{x}}_{opt}$ y $(\mathbf{x} - \hat{\mathbf{x}}_{opt})$ son ortogonales entre sí, es decir



$$\langle \hat{\mathbf{x}}_{opt}, (\mathbf{x} - \hat{\mathbf{x}}_{opt}) \rangle = \mathbb{E}\left[\hat{\mathbf{x}}_{opt}^{*} (\mathbf{x} - \hat{\mathbf{x}}_{opt}) \right] = 0$$

Esta propiedad es muy importante ya que muchas veces nos va a permitir computar $\hat{\mathbf{x}}_{opt}$ sin tener que resolver el MSE.

Estimador Lineal de Menor Error Cuadrático Medio (LMMSE)

Es interesante limitar la búsqueda de $\hat{\mathbf{x}} = g(\mathbf{y})$ a una familia de funciones específica. Por ejemplo, podemos buscar entre los estimadores lineales,

$$g(\mathbf{y}) = \mathbf{A}\mathbf{y} + \mathbf{b}.$$

Para encontrar la matriz $\mathbf{A} \in \mathbb{C}^{n \times m}$ y el vector $\mathbf{b} \in \mathbb{C}^{n}$ que minicen el MSE planteamos el principio de ortogonalidad, es decir

$$\mathbb{E}\left[\underbrace{(\mathbf{A}\mathbf{y} + \mathbf{b})^*}_{\hat{\mathbf{x}}^*}\underbrace{(\mathbf{x} - \mathbf{A}\mathbf{y} - \mathbf{b})}_{\mathbf{x} - \hat{\mathbf{x}}}\right] = 0 \quad (1)$$

Notación:

$$\begin{cases} \mu_{\mathbf{x}} = \mathbb{E}[\mathbf{x}] \quad , \quad \mathbf{C}_{\mathbf{x}} = \mathbb{E}\left[(\mathbf{x} - \mu_{\mathbf{x}})(\mathbf{x} - \mu_{\mathbf{x}})^*\right] \\ \mu_{\mathbf{y}} = \mathbb{E}[\mathbf{y}] \quad , \quad \mathbf{C}_{\mathbf{y}} = \mathbb{E}\left[(\mathbf{y} - \mu_{\mathbf{y}})(\mathbf{y} - \mu_{\mathbf{y}})^*\right] \quad , \quad \mathbf{R}_{\mathbf{y}} = \mathbb{E}[\mathbf{y}\mathbf{y}^*] \\ \mathbf{R}_{\mathbf{x}\mathbf{y}} = \mathbb{E}[\mathbf{x}\mathbf{y}^*] \quad , \quad \mathbf{C}_{\mathbf{x}\mathbf{y}} = \mathbb{E}\left[(\mathbf{x} - \mu_{\mathbf{x}})(\mathbf{y} - \mu_{\mathbf{y}})^*\right] \end{cases}$$

LMMSE

Desarrollando (1) obtenemos

$\mathrm{Tr}(A.B) = \mathrm{T}(BA)$

$$
\begin{aligned}
0 &= \mathrm{tr}\left\{\mathbb{E}\left[(\mathbf{A}\mathbf{y}+\mathbf{b})^*(\mathbf{x}-\mathbf{A}\mathbf{y}-\mathbf{b})\right]\right\}\\
&= \mathrm{tr}\left\{\mathbb{E}\left[\overbrace{\mathbf{y}^*\mathbf{A}^*\mathbf{x}}^{\text{escalar}}+\mathbf{b}^*\mathbf{x}-\overbrace{\mathbf{y}^*\mathbf{A}^*\mathbf{A}\mathbf{y}}^{\text{escalar}}-\mathbf{b}^*\mathbf{A}\mathbf{y}-\mathbf{y}^*\mathbf{A}^*\mathbf{b}-\mathbf{b}^*\mathbf{b}\right]\right\}\\
\text{tr: lineal} \quad &= \mathbb{E}\{\mathrm{tr}\,[\overbrace{\mathbf{A}^*\mathbf{x}\mathbf{y}^*}^{\text{ahora es matriz}}]\}+\mathbf{b}^*\mu_{\mathbf{x}}-\mathbb{E}\{\mathrm{tr}[\mathbf{A}^*\mathbf{A}\mathbf{y}\mathbf{y}^*]\}-\mathbf{b}^*\mathbf{A}\mu_{\mathbf{y}}-\mu_{\mathbf{y}}^*\mathbf{A}^*\mathbf{b}-\mathbf{b}^*\mathbf{b}\\
&= \mathrm{tr}\left\{\mathbf{A}^*\mathbb{E}\left[\mathbf{x}\mathbf{y}^*\right]\right\}+\mathbf{b}^*\mu_{\mathbf{x}}-\mathrm{tr}\left\{\mathbf{A}^*\mathbf{A}\mathbb{E}[\mathbf{y}\mathbf{y}^*]\right\}-\mathbf{b}^*\mathbf{A}\mu_{\mathbf{y}}-\mu_{\mathbf{y}}^*\mathbf{A}^*\mathbf{b}-\mathbf{b}^*\mathbf{b}\\
&= \mathrm{tr}\left[\mathbf{A}^*\left(\mathbf{R}_{\mathbf{xy}}-\mathbf{A}\mathbf{R}_{\mathbf{y}}-\mathbf{b}\mu_{\mathbf{y}}^*\right)\right]+\mathbf{b}^*\left[\mu_{\mathbf{x}}-\mathbf{A}\mu_{\mathbf{y}}-\mathbf{b}\right]. \quad = \emptyset
\end{aligned}
$$

Vemos que $\mathbf{A}_{opt}$ y $\mathbf{b}_{opt}$ deben cumplir el siguiente sistema de ecuaciones:

$$
\begin{cases}
\mathbf{A}\mathbf{R}_y+\mathbf{b}\mu_{\mathbf{y}}^* = \mathbf{R}_{\mathbf{xy}}\\
\mathbf{A}\mu_{\mathbf{y}}+\mathbf{b} = \mu_{\mathbf{x}}
\end{cases}
$$

M es única

cuya solución es $A(\underbrace{R_y - \mu_y \mu_y^*}_{C_y}) = \underbrace{R_{xy} - \mu_x \mu_y^*}_{C_{xy}}$

$$
\mathbf{A}_{opt} = \left[\mathbf{R}_{\mathbf{xy}}-\mu_{\mathbf{x}}\mu_{\mathbf{y}}^*\right]\left[\mathbf{R}_{\mathbf{y}}-\mu_{\mathbf{y}}\mu_{\mathbf{y}}^*\right]^{-1} = \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{y}}^{-1}
$$

$$
\mathbf{b}_{opt} = \mu_{\mathbf{x}}-\left[\mathbf{R}_{\mathbf{xy}}-\mu_{\mathbf{x}}\mu_{\mathbf{y}}^*\right]\left[\mathbf{R}_{\mathbf{y}}-\mu_{\mathbf{y}}\mu_{\mathbf{y}}^*\right]^{-1}\mu_{\mathbf{y}} = \mu_{\mathbf{x}}-\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{y}}^{-1}\mu_{\mathbf{y}}
$$

LMMSE

Estimador Lineal de Menor Error Cuadrático Medio (LMMSE)

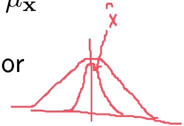
$$\hat{\mathbf{x}}_{opt} = \mu_{\mathbf{x}} + \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{y}}^{-1}(\mathbf{y} - \mu_{\mathbf{y}}). \quad (2)$$

$\hat{x}_{BLU}$ el mejor est. insesgado

LMMSE: Propiedades

- El estimador (2) es insesgado. Se lo conoce también como el estimador BLU (Best Linear Unbiased).

$$\mathbb{E}[\hat{\mathbf{x}}_{blu}] = \mathbb{E}[\mu_{\mathbf{x}} + \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{y}}^{-1}(\mathbf{y} - \mu_{\mathbf{y}})] = \mu_{\mathbf{x}}$$



- El error de estimación $\mathbf{x} - \hat{\mathbf{x}}_{blu}$ queda caracterizado por

$$\text{Cov}[\mathbf{x} - \hat{\mathbf{x}}_{blu}] = \mathbf{C}_{\mathbf{x}} - \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{C}_{\mathbf{yx}}$$

La dispersion de x_sombero sera menor que simplemente tomar la media de la señal

- Cuando $\mathbf{x}$ e $\mathbf{y}$ son dos vectores conjuntamente gaussianos,

$$\hat{\mathbf{x}}_{blu} = \hat{\mathbf{x}}_{MMSE}$$

Estimación de procesos aleatorios

Juntos x,y est. bivit. alcanza NO

Sean $x[n]$ e $y[n]$ dos procesos aleatorios.

Problema

A partir de $y[m]$, $m = m_0, m_0 + 1, m_0 + 2, \cdots$, estimar $x[n]$ de forma lineal

$$\hat{x}[n] = \sum_m k_{nm} y[m]$$

el instante q' estimo

el instante q' observo

tal que minimice el error cuadrático medio

la esperanza de la Energía del error

$$\|x - \hat{x}\|^2 = \mathbb{E}\left[\sum_n |x[n] - \hat{x}[n]|^2\right]$$

El problema de estimación es entonces obtener los coeficientes k_{nm}.

Hipótesis

$\mathbb{E}\left[x(m)\, y(m+k)\right] = f(k)$

Nva ahora

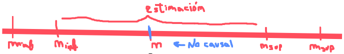
Asumimos lo siguientes:

- $x[n]$ e $y[n]$ son conjuntamente ESA
- $\mathbb{E}\{x[n]\} = \mathbb{E}\{y[n]\} = 0$ para todo n.
- La correlación cruzada $R_{xy}(k) = \mathbb{E}\{x[n]y[n+k]\}$ y la autocorrelación $R_y(k) = \mathbb{E}\{y[n]y[n+k]\}$ son funciones conocidas.

Tres posibles esquemas de estimación

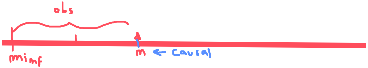
Distinguimos 3 situaciones de interés:

1. **Suavizado** (*smoothing*)
   Dado $\{y[m], m_{inf} \leq m \leq m_{sup}\}$ *obtener* $\{\hat{x}[n], n_{inf} \leq n \leq n_{sup}\}$ *donde* $n_{inf} \geq m_{inf}$, $n_{sup} \leq m_{sup}$.

   Estimador *no causal* ya que utiliza observaciones futuras.

2. **Filtrado**
   Dado $\{y[m],\, m \leq n\}$ *obtener* $\hat{x}[n]$.

   Estimador *causal*. Sólo utiliza observaciones pasadas y presentes.

3. **Predicción**

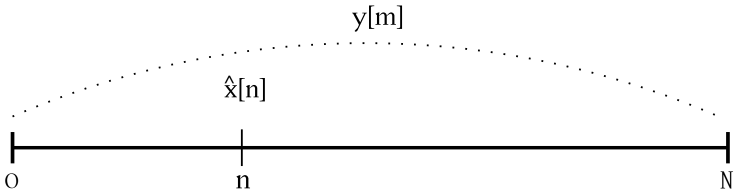
   

   Dado $\{y[m],\, m \leq n\}$ *obtener* $\hat{x}[n+\tau]$ *donde* $\tau > 0$.

   Predicción del futuro utilizando observaciones presentes y pasadas.

Problema de suavizado en una ventana fija

Dadas las observaciones $\{y[m], m = 0, \cdots, M\}$ obtener $\hat{x}[n]$, para $0 \leq n \leq M$.



Para cada instante n, el problema es equivalente al de estimar una v.a. $x[n]$ a partir de la observación de un vec. a. $\mathbf{y} = [y[0], \ \ y[1] \ \ \cdots, y[M]]^t$. Vimos que $\hat{x}_{blu}$ es la solución, donde

$$\hat{x}_{blu} = \mu_x + \mathbf{C}_{x\mathbf{y}}\mathbf{C}_{\mathbf{y}}^{-1}\left(\mathbf{y} - \mu_{\mathbf{y}})\right]$$

Por hipotesis, $\mu_x = 0, \mu_{\mathbf{y}} = 0$. Luego,

Notar que solo para estimar x_(n) es decir solo una muestra, un escalar, se debe de tener una ventama de Y = {y[0], y[1], ..., y[M]}

$$\hat{x}[n] = \mathbf{R}_{x\mathbf{y}}[n]\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{y}. \tag{3}$$

Problema de suavizado en una ventana fija

- Correlación cruzada: $\mathbf{R}_{x\mathbf{y}}[n] \in \mathbb{C}^{1\times(M+1)}$

$$\begin{aligned}\mathbf{R}_{x\mathbf{y}}[n] &= \mathbb{E}\left\{x[n]\left[y[0],\ \ y[1],\ \ \cdots, y[n],\ \ \cdots, y[M]\right]\right\}\\ &= \left[R_{xy}(n),\ \ R_{xy}(n-1)\ \ \cdots, R_{xy}(0),\ \ \cdots, R_{xy}(n-M)\right]\end{aligned}$$

- Autocorrelación: $\mathbf{R}_{\mathbf{y}} \in \mathbb{R}^{(M+1)\times(M+1)}$

$$\mathbf{R}_{\mathbf{y}} = \mathbb{E}\left\{\begin{bmatrix} y[0] \\ \vdots \\ y[M]\end{bmatrix}\begin{bmatrix} y[0] & \cdots & y[M]\end{bmatrix}\right\} = \begin{bmatrix} R_y(0) & \cdots & R_y(-M) \\ R_m(1) & \text{———} & R_m(1-M) \\ & \ddots & \\ R_y(M) & \cdots & R_y(0)\end{bmatrix}$$

La matrix $\mathbf{R}_{\mathbf{y}}$ es una matriz Toeplitz simétrica.

Ls diagonales constantes, simétricas

Suavizado en una ventana fija (Visión alternativa)

Un estimador lineal general tiene la forma

Se obtiene una Km para cada muestra estimada, es ecir que para obtener mucho x_[n] estimados la ganancia cambiara

$$\hat{x}[n] = \sum_{m=0}^{M} k_{nm} y[m]$$

El principio de ortogonalidad de todo estimador MMSE establece que (Por qué?)

$x - \hat{x} \perp \hat{x}$ / $\hat{x} = \sum k_{n,m} y(m)$

$$\left(x[n] - \sum_{m=0}^{M} k_{nm} y[m]\right) \perp y[l] \qquad l = 0, \cdots M$$

o en forma equivalente

$$\mathbb{E}\left\{\left(x[n] - \sum_{m=0}^{M} k_{nm} y[m]\right) y[l]\right\} = 0 \qquad l = 0, \cdots M.$$

$E(xy - \sum k\ yy)$

Suavizado en una ventana fija

Se obtiene un sistema de $M+1$ ecuaciones con $M+1$ incógnitas. Éstas son las *ecuaciones normales:*

$$R_{xy}(n-l) = \sum_{m=0}^{M} k_{nm} R_y(m-l) \qquad l = 0, \cdots M.$$

Sea

$$\mathbf{k}_n = [k_{n0}, \ \cdots \ , \ k_{nM}].$$

Luego, $\hat{x}[n] = \mathbf{k}_n \mathbf{y}$ y las ecuaciones normales resultan

$$\mathbf{R}_{x\mathbf{y}}[n] = \mathbf{k}_n \mathbf{R}_{\mathbf{y}}.$$

cuya solución coincide con la solución obtenida con el planteo inicial.

Comentario: *Este resultado era de esperarse ya que el $\hat{x}_{blu}$ se dedujo a partir del principio de ortogonalidad*

Suavizado en una ventana fija: error de estimación

Sean

- $e[n] = x[n] - \hat{x}[n]$, el error de estimación
- $\sigma_e^2[n]$, la varianza de $e[n]$
- σ_x^2, la varianza de $x[n]$.

A priori, antes de realizar observación alguna, σ_x^2 es una indicación de la incertidumbre que se tiene sobre el proceso $x[n]$. Luego de obtener $\hat{x}[n]$, la incertidumbre en $x[n]$ está dada por $\sigma_e^2[n]$. Luego, cómo es la varianza de $e[n]$?

Suavizado en una ventana fija: error de estimación

- Sabemos que $\hat{x}[n]$ es un estimador insesgado, luego $\mathbb{E}[e[n]] = 0$.
- Recordando $Var(\hat{x}_{blu})$, tenemos que para cada n

$$\mathbb{E}[e[n]^2] = \sigma_e^2[n] = \sigma_x^2 - \mathbf{R}_{x\mathbf{y}}[n]\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{y}x}[n] = \sigma_x^2 - \mathbf{k}_n\mathbf{R}_{\mathbf{y}x}[n]$$

Cuanto mas correlacionan, menor error.

- Como $\mathbf{R}_{\mathbf{y}} > 0$, tenemos que

$$\mathbf{R}_{x\mathbf{y}}[n]\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{y}x}[n] = \mathbf{R}_{x\mathbf{y}}[n]\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{x\mathbf{y}}^{t}[n] > 0 \qquad \forall n.$$

Luego, $\sigma_e^2[n] < \sigma_x^2$ para todo n. Es decir, la estimación reduce la incertidumbre sobre $x[n]$.

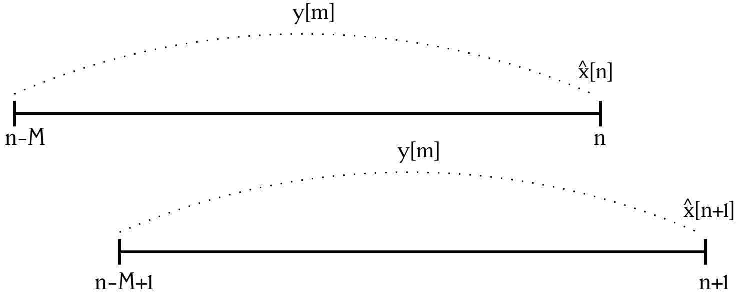
Pregunta: El proceso $e[n]$ es ESA? NO

Rxy = E[x[n].{y[0], y[1], ..., y[M]}]
= [Rxy(n), Rxy(n-1), ..., Rxy(n-M)]

Depende de "n"

Filtrado en una ventana fija

Dadas las observaciones $\{y[m], n - M \leq m \leq n\}$, obtener $\hat{x}[n]$.



La sútil diferencia entre el problema de suavizado y el de filtrado, es que el segundo requiere de un sistema *causal* y por ende sólo utiliza observaciones pasadas y presente[2]. Para cada valor de n, podemos hacer un planteo similar al que realizamos en el suavizado, es decir, para cada n planteo

$$\hat{x}[n] = \sum_{m=n-M}^{n} k_{nm} y[m]$$

Filtrado en una ventana fija

Dado que buscamos un estimador MMSE, planteamos la ortogonalidad entre el error y las observaciones:

$$(x[n] - \sum_{m=n-M}^{n} k_{nm} y[m]) \perp y[l] \qquad l = n - M, \cdots n.$$

Recordando que $x[n]$ e $y[n]$ son dos procesos conjuntamente ESA, obtenemos las ecuaciones normales

$$R_{xy}(n-l) = \sum_{m=n-M}^{n} k_{nm} R_y(m-l) \qquad l = n - M, \cdots n.$$

Filtrado en una ventana fija

Rxy = E[x[n].{y[n-M], y[n-M+1], ..., y[n]}]
= [Rxy(M), Rxy(M-1), ..., Rxy(0)]

Como antes, definimos

No depende de "n"

$$\mathbf{R}_{x\mathbf{y}} = \begin{bmatrix} R_{xy}(M) & \cdots & R_{xy}(0) \end{bmatrix} = \mathbf{R}_{x\mathbf{y}}[M]$$

$$\mathbf{R}_{\mathbf{y}} = \begin{bmatrix} R_y(0) & \cdots & R_y(-M) \\ & \ddots & \\ R_y(M) & \cdots & R_y(0) \end{bmatrix}$$

$$\mathbf{k}_n = \begin{bmatrix} k_{n,n-M} & \cdots & k_{n,n} \end{bmatrix}$$

En este caso, la matriz $\mathbf{R}_{x\mathbf{y}}$ no depende de n.

Filtrado en una ventana fija

Expresando el sistema de ecuaciones en forma explícita, tenemos

$$\mathbf{R}_{x\mathbf{y}} = \mathbf{k}_n \mathbf{R}_{\mathbf{y}}$$

La solución es

$$\mathbf{k}_n = \mathbf{R}_{x\mathbf{y}} \mathbf{R}_y^{-1}$$

El vector $\mathbf{k}_n$ sólo depende de M, el largo de la ventana de filtrado. Luego,

$$\mathbf{k}_n = \mathbf{k} = \begin{bmatrix} k_0 & \cdots & k_M \end{bmatrix}$$

Vemos que los coeficientes k_{nm} sólo dependen de la diferencia $n - m$.

Filtrado en una ventana fija

El estimador para resolver el problema de filtrado fue planteado como

$$\hat{x}[n] = \sum_{m=n-M}^{n} k_{nm} y[m]$$

Pero vimos que k_{nm} no depende de n sino de $n-m$. Luego,

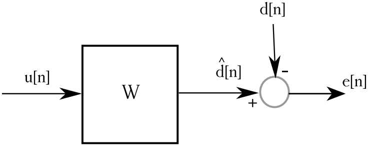
$$\hat{x}[n] = \sum_{m=n-M}^{n} k_{n-m} y[m] = \sum_{l=0}^{M} k[l] y[n-l]$$

donde $k[l] = k_l$ es un filtro FIR de longitud $M+1$ que realiza el filtrado causal.

Pregunta para ir pensando: Cómo se compara este filtro con los que estudiamos en la primer parte del curso?

Notación

En cierta bibliografía del tema, es costumbre plantear el problema de filtrado como el mecanismo que permite transformar una señal en otra señal deseada.

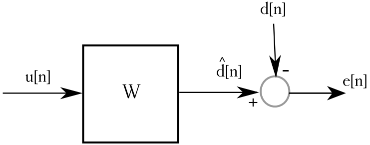


El planteo es el mismo si consideramos la siguiente notación:

- Las observaciones $y[n]$ son las *entradas* del filtro $u[n]$.
- El proceso $x[n]$ es la señal *deseada* $d[n]$.
- W es un filtro FIR de longitud $M+1$, lo que equivale a trabajar con una ventana de longitud finita M

El filtro óptimo se obtiene planteando las ecuaciones normales a partir del principio de ortogonalidad del estimador MMSE.

Notación



Principio de ortogonalidad:

$$< u[n], e[n]) >= 0$$

Filtro con horizonte finito resultante:

$$\begin{bmatrix} w_0 \\ \vdots \\ w_M \end{bmatrix} = \begin{bmatrix} R_u(0) & \cdots & R_u(M) \\ & \ddots & \\ R_u(-M) & \cdots & R_u(0) \end{bmatrix}^{-1} \begin{bmatrix} R_{du}(M) \\ \vdots \\ R_{du}(0) \end{bmatrix}$$

$$\hat{d}[n] = \sum_{m=0}^{M} w[l]u[n-l]$$

Problemas de estimación con horizonte infinito

Supongamos ahora que la ventana de observación es infinita. Entonces, los estimadores para el suavizado y el filtrado son:

1. Suavizado: $\hat{x}[n] = \sum_{m=-\infty}^{+\infty} k_{nm} y[m]$
2. Filtrado: $\hat{x}[n] = \sum_{m=-\infty}^{n} k_{nm} y[m]$

En lo que sigue, vamos a analizar ambos problemas.

Entonces horizonte finito o infinito hace referencia al ancho de la ventana de observacion.
osea el rango de 'm'

Como siempre, partimos del principio de ortogonalidad para obtener las ecuaciones normales:

$$\left(x[n] - \sum_{m=-\infty}^{+\infty} k_{nm} y[m]\right) \quad \perp \quad y[l] \qquad -\infty < l < +\infty$$

Aplicando la definición de producto interno entre variables aleatorias, tenemos

$$\mathbb{E}\left[\left(x[n] - \sum_{m=-\infty}^{+\infty} k_{nm} y[m]\right)^t y[l]\right] = 0 \qquad -\infty < l < +\infty$$

Suavizado con horizonte infinito

Las ecuaciones normales en este caso, resultan

$$R_{xy}(n-l) = \sum_{m=-\infty}^{+\infty} k_{nm} R_y(m-l) \qquad -\infty < l < +\infty$$

Haciendo dos cambios de variables

$$\begin{aligned} R_{xy}(n') &= \sum_{m'=-\infty}^{+\infty} k_{(n'+l)(m'+l)} R_y(m') \qquad -\infty < l < +\infty \\ &= \sum_{m'=-\infty}^{+\infty} w[n'-m'] R_y(m'), \end{aligned}$$

donde $w[n'-m'] = k_{n'm'}$. Éste es el *filtro de Wiener* utilizado en un problema de suavizado.

Suavizado con horizonte infinito

Las infinitas ecuaciones normales plantean una operación de convolución. Supongamos que $w[n]$ tiene transformada de Fourier $W(\omega) = \sum_{l=-\infty}^{+\infty} w[l]e^{-\jmath\omega l}$ y sean

$$S_{xy}(\omega) = \sum_{l=-\infty}^{+\infty} R_{xy}[l]e^{-\jmath\omega l} \qquad S_y(\omega) = \sum_{l=-\infty}^{+\infty} R_y[l]e^{-\jmath\omega l}$$

Luego, planteamos la transformada de Fourier de las ecuaciones normales.

$$S_{xy}(\omega) = W(\omega)S_y(\omega)$$

Si $S_y(\omega) > 0$ para todo ω,

$$W(\omega) = \frac{S_{xy}(\omega)}{S_y(\omega)}.$$

Suavizado con horizonte infinito: Ejemplo

Supongamos que los procesos $x[n]$ e $y[n]$ están relacionados del siguiente modo:

$$y[n] = x[n] + v[n]$$

donde $x[n]$ tiene densidad espectral de potencia $S_x(\omega)$ y $v[n]$ es una señal de ruido que está descorrelacionada con $x[n]$ y tiene densidad espectral de potencia $S_v(\omega)$. Con estas hipótesis, tenemos que

$$S_y(\omega) = S_x(\omega) + S_v(\omega) \qquad S_{xy}(\omega) = S_x(\omega).$$

Luego,

$$W(\omega) = \frac{S_{xy}(\omega)}{S_y(\omega)} = \frac{S_x(\omega)}{S_x(\omega) + S_v(\omega)} = \frac{1}{1 + \frac{S_v(\omega)}{S_x(\omega)}}$$

Los pesos y la potencia de ruido son inversamente proporcionales, a mayor potencia de ruido menor sera la contribucion de la señal ruidosa a la estimacion.

Suavizado con horizonte infinito: Ejemplo

Sea $Y(\omega)$ la DTFT de la realización particular de $y[n]$ que es observada, y $\hat{X}(\omega)$ la DTFT de $\hat{x}[n]$. Luego,

$$\hat{X}(\omega) = W(\omega)Y(\omega)$$

Observaciones:

- En cada frecuencia, la intensidad del filtro depende de la relación entre la potencia de ruido y la de señal en esa frecuencia
- Las frecuencias en las cuales el ruido tiene mayor potencia que la señal son atenuadas, para evitar la contaminación del ruido
- Por otro lado, cuando el ruido es muy bajo, el filtro deja pasar la medición casi sin modificarla.

Suavizado con horizonte infinito: Conclusiones

Suavizado óptimo

El suavizado de Wiener establece el compromiso óptimo entre el ruido y la señal que permite limpiar las observaciones ruidosas. Filtra el ruido de forma selectiva de modo de obtener la menor varianza del error de estimación.

Filtrado con horizonte infinito

La estimación que buscamos en este caso es

$$\hat{x}[n] = \sum_{m=-\infty}^{n} k_{nm} y[m].$$

Siguiendo los pasos de siempre, obtenemos las ecuaciones normales

$$R_{xy}(n-l) = \sum_{m=-\infty}^{n} k_{nm} R_y(m-l) \qquad -\infty < l \leq n.$$

Consideremos el cambio de variables $n' = n - l$, $m' = m - l$.

$$R_{xy}(n') = \sum_{m'=-\infty}^{n'+l} k_{(n'+l)(m'+l)} R_y(m') \qquad n' \geq 0$$

Filtrado con horizonte infinito

Como $R_{xy}(n')$ no depende de l

$$R_{xy}(n') = \sum_{m'=-\infty}^{n'} k_{n'm'} R_y(m')$$

$$= \sum_{m'=-\infty}^{n'} k_{n'-m'} R_y(m') \qquad n' \geq 0 \qquad = \sum_{l=0}^{+\infty} k_l R_y(n'-l)$$

Si imponemos la condición $k_l = 0$ para $l < 0$, entonces obtenemos la ecuación discreta de Wiener-Hopf:

$$R_{xy}(n) = \sum_{l=-\infty}^{+\infty} k_l R_y(n-l) \qquad n \geq 0 \tag{4}$$

Si definimos una operación de convolución, la igualdad en (4) sólo se da para $n \geq 0$. No podemos resolver la ecuación utilizando Fourier como lo hicimos en el caso del suavizado.

Solución de la ecuación de Wiener-Hopf discreta: Preliminares

Asumimos que:

- $S_{xy}(z) = \sum_{l=-\infty}^{+\infty} R_{xy}(l) z^{-l}$ existe.
- $S_y(z) = \sum_{l=-\infty}^{+\infty} R_y(l) z^{-l}$ existe y no tiene ceros en el círculo unitario.
- $S_y(z)$ admite la *factorización espectral*

$$S_y(z) = \alpha L(z) L^*(z^{-*})$$

donde

1. $L(z)$ tiene fase mínima, es decir, sus polos y ceros están dentro del círculo unitario
2. $L(\infty) = 1$, es decir, $L(z) = 1 + \sum_{i=1}^{\infty} l_i z^{-i}$
3. $\alpha > 0$.

Solución de la ecuación de Wiener-Hopf discreta: Preliminares

- Sea $F(z) = \sum_{l=-\infty}^{+\infty} f(l) z^{-l}$. Definimos la *parte causal* de $F(z)$ como

$$\{F(z)\}_+ = \sum_{l=0}^{+\infty} f(l) z^{-l}$$

Nota: Recordemos que si $K(z)$ es la transferencia de un sistema LTI causal, entonces $K(z) = \sum_{l=0}^{+\infty} k_l z^{-l}$.

Volviendo a (4), planteamos la diferencia entre los dos miembros como

$$g(n) = R_{xy}(n) - \sum_{m=-\infty}^{+\infty} k_m R_y(n-m).$$

La función $g(n)$ está definidad $\forall n$. Sólo cuando $n \geq 0$, $g(n) = 0$. Es decir,

$g(n)$ **es una secuencia anticausal**.

Usando las hipótesis preliminares para calcular $G(z)$ obtenemos

$$\begin{aligned} G(z) &= S_{xy}(z) - K(z)S_y(z) \\ &= S_{xy}(z) - \alpha K(z)L(z)L^*(z^{-*}). \end{aligned}$$

Luego,

$$\frac{G(z)}{\alpha L^*(z^{-*})} = \frac{S_{xy}(z)}{\alpha L^*(z^{-*})} - K(z)L(z). \tag{5}$$

Solución de la ecuación de Wiener-Hopf discreta

En (5) tenemos:

- La transformada de secuencia anticausal

$$\frac{G(z)}{\alpha L^*(z^{-*})}$$

- La transformada de secuencia causal

$$K(z)L(z)$$

- Un factor con términos causales y anticausales

$$\frac{S_{xy}(z)}{\alpha L^*(z^{-*})}.$$

Solución de la ecuación de Wiener-Hopf discreta

Para obtener la igualdad en (5), necesitamos que

$$\left\{ \frac{S_{xy}(z)}{\alpha L^*(z^{-*})} \right\}_+ = K(z)L(z)$$

El filtro óptimo resulta:

$$K(z) = \frac{1}{L(z)} \left\{ \frac{S_{xy}(z)}{\alpha L^*(z^{-*})} \right\}_+$$

Para obtener $K(z)$ debemos obtener:

1. La factorización espectral de $S_y(z)$
2. La parte causal del cociente de arriba.

Factorización espectral canónica

Definición

Sea $S(z) = \sum_{n=-\infty}^{+\infty} r[n] z^{-n}$, donde $r[n]$ es la autocorrelación de un proceso de potencia finita. Asumimos que $S(z)|_{z=e^{\jmath\omega}} > 0$ para todo $\omega \in \mathbb{R}$. La *factorización espectral canónica* de $S(z)$ es

$$S(z) = \alpha L(z) L^*(z^{-*}),$$

donde $L(z)$ es de fase mínima y tiene la forma $L(z) = 1 + \sum_{i=1}^{\infty} l_i z^{-i}$, y $\alpha > 0$.

Factorización espectral canónica

$L(z)$ puede ser vista como la transferencia de un sistema causal y estable cuya respuesta impulsiva es $l[n]$.

Observaciones

- Si se aplica a la entrada de $L(z)$ un proceso $e[n]$ que es blanco y de potencia α, la salida será $y[n] = l[n] * e[n]$.
- $S_y(\omega) = \alpha |L(\omega)|^2 = \alpha L(\omega) L^*(\omega) = S(\omega)$, es decir, implementar $L(z)$ es el modo de obtener un proceso con densidad espectral determinada.
- Como $L(z)$ es causal y mínima fase, $L^{-1}(z)$ existe y es estable y causal. Más aún, $L^{-1}(z)$ es el filtro blanqueador que obtiene un proceso blanco ($e[n]$) a partir de $y[n]$.

Factorización espectral canónica

- En general, obtener la factorización espectral de un proceso es una tarea compleja. Sin embargo, cuando el proceso es tal que $S(z)$ es una funcioń racional de z, la factorización queda determinada por sus polos y ceros.
- $S(z)$ por ser la transformada de una secuencia real y par, cumple con

$$S(z) = S^*(z^{-*})$$

  Luego, si p es un polo/cero, p^{-*} también lo es.
- Si $S(z)$ es una función racional de z, entonces tiene la forma:

$$S(z) = \alpha \frac{\prod_{i=1}^{m}(z - z_i)(z^{-1} - z_i^*)}{\prod_{i=1}^{n}(z - p_i)(z^{-1} - p_i^*)} \qquad |z_i|, |p_i| < 1, \quad \alpha > 0.$$

- Para armar $L(z)$, retenemos los ceros y polos estables de $S(z)$

$$L(z) = z^{n-m} \frac{\prod_{i=1}^{m}(z - z_i)}{\prod_{i=1}^{n}(z - p_i)}$$

Parte causal utilizando fracciones parciales

Obtener la parte causal de una función $F(z)$ racional es bastante sencillo. Para ello, partimos de la descomposición en fracciones parciales.
Supongamos que $F(z)$ tiene más polos que cero. Luego, es posible expresar $F(z)$ como

$$F(z) = r_0 + \sum_{i=1}^{p} \sum_{l=1}^{l_i} \frac{r_{pl}}{(z - z_i)^l}.$$

Es claro que la operación que extrae la parte causal es lineal, es decir,

$$\{F(z)\}_+ = \{r_0\}_+ + \sum_{i=1}^{p} \sum_{l=1}^{l_i} r_{pl} \left\{ \frac{1}{(z - z_i)^l} \right\}_+ .$$

Factorización espectral utilizando fracciones parciales

Recordando que la ROC de la parte causal y estable de un sistema contiene al círculo unitario, obtenemos las siguientes reglas:

- $\{r_0\}_+ = r_0$ para toda constante r_0.
- $$\left\{\frac{1}{(z-z_i)^l}\right\}_+ = \begin{cases} \frac{1}{(z-z_i)^l} & |z_i| < 1 \\ \frac{1}{(-z_i)^l} & |z_i| > 1 \end{cases}$$

Ojo!!!! Antes de aplicar estas reglas, poner cuidado en la expresión de $F(z)$. La descomposición en fracciones parciales debe ser en función de z, no de z^{-1}.

Retomamos el problema de una señal $x[n]$ perturbada por un ruido descorrelacionado $v[n]$.

$$y[n] = x[n] + v[n]$$

Para simplificar el tratamiento, vamos a asumir que $v[n]$ es blanco, es decir, $S_v(z) = \sigma_v^2$. Como antes, tenemos

$$S_y(z) = S_x(z) + \sigma_v^2 \qquad S_{xy}(z) = S_x(z).$$

Más aún,

$$S_y(z) = \alpha L(z) L^*(z^{-*})$$

La transferencia del filtro de Wiener resulta

$$\begin{aligned}
K(z) &= \frac{1}{L(z)}\left\{\frac{S_{xy}(z)}{\alpha L^*(z^{-*})}\right\}_+ \\
&= \frac{1}{L(z)}\left\{\frac{S_y(z)-\sigma_v^2}{\alpha L^*(z^{-*})}\right\}_+ \\
&= \frac{1}{\alpha L(z)}\left\{\alpha L(z)-\frac{\sigma_v^2}{L^*(z^{-*})}\right\}_+ \\
&= \frac{1}{\alpha L(z)}\left\{\alpha L(z)-\sigma_v^2\left\{\frac{1}{L^*(z^{-*})}\right\}_+\right\} \\
&= 1-\frac{\sigma_v^2}{\alpha}\frac{1}{L(z)} \\
&= 1-\frac{1}{\alpha\frac{L(z)}{\sigma_v^2}}
\end{aligned}$$

Bibliografía

1. *Linear Estimation*, Thomas Kailath, Ali H. Sayed, Babak Hassibi, Prentice Hall, 2000 - Mathematics
2. *Adaptive Filter Theory*, Simon Haykin, Pearson 2014

Bibliografía

1. *Discrete-Time Signal Processing*, Alan V. Oppenheim, Ronald W. Schafer, and John R. Buck, Prentice-Hall Signal Processing Series.
2. *Digital Signal Processing*, J Proakis, D. Manolakis, Prentice Hall
3. *Multirate Systems And Filter Banks*, P. P. Vaidyanathan, Prentice Hall.
4. *Introduction to Spectral Analysis,*, P. Stoica, R. Moses, Prentice Hall.
5. *Linear Estimation*, Thomas Kailath, Ali H. Sayed, and Babak Hassibi Prentice Hall.
6. L.R. Rabiner, B. Gold, C. A. McGonegal, "An Approach to the Approximation Problem for Nonrecursive Digital Filters", IEEE Trans. Audio and Electroacustics, vol. AU-18, June 1970.
7. Neuvo Y., D. Cheng-yu and S. Mitra , "Interpolated finite impulse response filters." IEEE Transaction on Acoustics Speech, and Signal Processing, Vol. ASSP-32, 1984, pp. 563-570.