Estimación Lineal

Cecilia Galarza

Procesos Estocásticos
Facultad de Ingeniería, Universidad de Buenos Aires

1er cuatrimestre 2025

Estimación de Menor Error Cuadrático Medio

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

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

Si $\hat{\boldsymbol{x}} = g(\boldsymbol{y})$ es visto como la transformación del VeA $\mathbf{Y}$, entonces $\hat{\mathbf{X}}$ es un VeA. Un posible criterio de *cercanía* es el *Error Cuadrático Medio* (MSE, *Mean Square Error*)

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

Estimador MMSE

Definición (Estimador MMSE)

El estimador MMSE (Minimun Mean Square Error) es el que minimiza $MSE(\hat{\mathbf{X}})$*,*

$$\hat{\mathbf{X}}_{mmse} = \arg \min_{\hat{\mathbf{X}}=g(\mathbf{Y})} \mathbb{E}\left[\|\mathbf{X} - \hat{\mathbf{X}}\|^2\right]$$

Teorema

El estimador MMSE es la esperanza condicional

$$\hat{\mathbf{X}}_{mmse} = \mathbb{E}\left[\mathbf{X}|\mathbf{Y}\right]$$

$\hat{\mathbf{X}}_{mmse}$ es la esperanza condicional: demostración

Para simplificar notación, llamamos:

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

Luego,

$$\mathbf{X} - \hat{\mathbf{X}}(\mathbf{Y}) = \mathbf{X} - \mu(\mathbf{Y}) + \mu(\mathbf{Y}) - \hat{\mathbf{X}}(\mathbf{Y}) = \mathbf{X} - \mu(\mathbf{Y}) + \Delta(\mathbf{Y}).$$

Retomando el MSE

$$\begin{aligned} MSE &= \mathbb{E}\left[\|\mathbf{X} - \hat{\mathbf{X}}\|^2\right] = \mathbb{E}\left[(\mathbf{X} - \hat{\mathbf{X}})^t(\mathbf{X} - \hat{\mathbf{X}})\right] \\ &= \mathbb{E}\left[(\mathbf{X} - \mu + \Delta)^t (\mathbf{X} - \mu + \Delta)\right] \\ &= \underbrace{\mathbb{E}\left[\|\mathbf{X} - \mu\|^2\right]}_{①} + \underbrace{\mathbb{E}\left[(\mathbf{X} - \mu)^t\Delta\right] + \mathbb{E}\left[\Delta^t(\mathbf{X} - \mu)\right]}_{②} + \underbrace{\mathbb{E}\left[\Delta^t\Delta\right]}_{③} \end{aligned}$$

$\hat{\mathbf{X}}_{mmse}$ es la esperanza condicional: demostración

$$① = \mathbb{E}\left[\|\mathbf{X} - \mu(\mathbf{Y})\|^2\right]$$

No depende de $\hat{\mathbf{X}}$, no lo considero para la minimización

$\hat{\mathbf{X}}_{mmse}$ es la esperanza condicional: demostración

$$② = \mathbb{E}\left[(\mathbf{X}-\mu)^t \underbrace{\left(\mu - \hat{\mathbf{X}}\right)}_{\Delta(\mathbf{Y})}\right] + \mathbb{E}\left[\underbrace{\left(\mu - \hat{\mathbf{X}}\right)^t}_{\Delta(\mathbf{Y})^t}(\mathbf{X}-\mu)\right]$$

$$= 2\mathbb{E}\left[(\mathbf{X}-\mu(\mathbf{Y}))^t \Delta(\mathbf{Y})\right]$$

Recordemos que $\forall$ $\mathbf{X}$, $\mathbf{Y}$, $\mathbb{E}[\mathbf{X}] = \mathbb{E}_{\mathbf{X}}[\mathbb{E}_{\mathbf{X}|\mathbf{Y}}[\mathbf{X}|\mathbf{Y}]]$.

$$\begin{aligned}
\mathbb{E}\left[(\mathbf{X}-\mu(\mathbf{Y}))^t \Delta(\mathbf{Y})\right] &= \mathbb{E}\left[\mathbb{E}[(\mathbf{X}-\mu(\mathbf{Y}))^t \Delta(\mathbf{Y}) \mid \mathbf{Y}]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbf{X}^t\Delta(\mathbf{Y}) - \mu(\mathbf{Y})^t\Delta(\mathbf{Y}) \mid \mathbf{Y}\right]\right] \\
&= \mathbb{E}\left[\underbrace{\mathbb{E}[\mathbf{X}^t \mid \mathbf{Y}]}_{\mu(\mathbf{Y})^t}\Delta(\mathbf{Y}) - \mu(\mathbf{Y})^t\Delta(\mathbf{Y})\right] = 0
\end{aligned}$$

$\hat{\mathbf{X}}_{mmse}$ es la esperanza condicional: demostración

$$③ = \mathbb{E}\left[\|\Delta(\mathbf{Y})\|^2\right] = \mathbb{E}\left[\left\|\mu(\mathbf{Y}) - \hat{\mathbf{X}}\right\|^2\right] = \mathbb{E}\left[\left\|\mathbb{E}[\mathbf{X} \mid \mathbf{Y}] - \hat{\mathbf{X}}\right\|^2\right]$$

$\hat{\mathbf{X}}_{mmse}$ es la esperanza condicional: demostración

Recapitulando

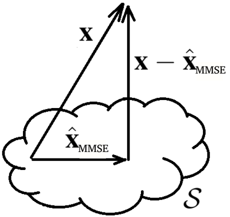
- ①, no depende de $\hat{\mathbf{X}}$
- ② es nulo
- ③ $= \mathbb{E}\left[\left\|\mathbb{E}[\mathbf{X}|\mathbf{Y}] - \hat{\mathbf{X}}\right\|^2\right]$

Minimizar MSE con respecto a $\hat{\mathbf{X}}(\mathbf{Y})$ equivale a minimizar ③. Luego,

$$\hat{\mathbf{X}}_{mmse} = \mathbb{E}[\mathbf{X}|\mathbf{Y}].$$

Principio de ortogonalidad

Recordemos que $\mathbf{X}$ y $\mathbf{Y}$ son ortogonales sii $\mathbb{E}[\mathbf{X}^t\mathbf{Y}] = 0$.



$$\mathbb{E}\left[\hat{\mathbf{X}}^t_{mmse}(\mathbf{Y})\left(\mathbf{X} - \hat{\mathbf{X}}_{mmse}(\mathbf{Y})\right)\right] = \mathbb{E}_{\mathbf{Y}}\left[\mathbb{E}_{\mathbf{X}|\mathbf{Y}}\left[\hat{\mathbf{X}}^t_{mmse}(\boldsymbol{y})\mathbf{X} - \hat{\mathbf{X}}^t_{mmse}(\boldsymbol{y})\hat{\mathbf{X}}_{mmse}(\boldsymbol{y}) \mid \boldsymbol{y}\right]\right]$$
$$= \mathbb{E}_{\mathbf{Y}}\left[\hat{\mathbf{X}}^t_{mmse}\mathbb{E}[\mathbf{X}|\mathbf{Y}] - \hat{\mathbf{X}}^t_{mmse}\hat{\mathbf{X}}_{mmse}\right] = 0$$

Los vectores $\hat{\mathbf{X}}_{mmse}$ y $(\mathbf{X} - \hat{\mathbf{X}}_{mmse})$ son ortogonales entre sí

Ejemplo

Sea la VA $Y = X + W$ donde $X \sim \text{Ber}(p)$ y $W \sim \mathcal{N}(0,1)$, X y W independientes. Halle el estimador MMSE de X al observar Y.

Sabemos que

$$\hat{X}_{mmse} = \mathbb{E}[X|Y].$$

Planteamos

$$\varphi(y) = \mathbb{E}[X|Y = y].$$

Esta función se conoce como función de *regresión*.

Ejemplo (cont.)

Como $X \sim \text{Ber}(p)$,

$$\varphi(y) = 0\,\mathbb{P}(X = 0|Y = y) + 1\,\mathbb{P}(X = 1|Y = y) = \mathbb{P}(X = 1|Y = y).$$

Usando la definición de probabilidad condicional y Bayes:

$$\mathbb{P}(X = 1|Y = y) = \frac{f_{Y|X=1}(y)\mathbb{P}(X = 1)}{f_Y(y)}.$$

Por definición Y es una mezcla de gaussianas:

$$f_Y(y) = (1 - p)\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}} + p\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-1)^2}{2}}.$$

Por otro lado,

$$f_{Y|X=1}(y) = f_W(y - 1) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(y-1)^2}{2}}.$$

Ejemplo (cont.)

Combinando todo se tiene:

$$\varphi(y) = \frac{pe^{-\frac{(y-1)^2}{2}}}{(1-p)e^{-\frac{y^2}{2}} + pe^{-\frac{(y-1)^2}{2}}} = \frac{p}{(1-p)e^{-\frac{1}{2}\left[y^2-(y-1)^2\right]} + p}$$
$$= \frac{p}{p + (1-p)e^{\frac{1}{2}-y}}.$$

$$\longrightarrow \hat{X}_{mmse} = \frac{p}{p + (1-p)e^{\frac{1}{2}}e^{-Y}}.$$

Observaciones:

- $X \in \{0, 1\}$, pero $\hat{X}_{mmse} \in [0, 1]$.
- $\phi(y) = \frac{p}{p+(1-p)e^{\frac{1}{2}}e^{-y}} \longrightarrow$ función no-lineal en y.

Notación

$$\mu_{\mathbf{X}} = \mathbb{E}[\mathbf{X}] \quad , \quad \mathbf{C}_{\mathbf{X}} = \mathbb{E}\left[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})^t\right]$$

$$\mu_{\mathbf{Y}} = \mathbb{E}[\mathbf{Y}] \quad , \quad \mathbf{C}_{\mathbf{Y}} = \mathbb{E}\left[(\mathbf{Y} - \mu_{\mathbf{Y}})(\mathbf{Y} - \mu_{\mathbf{Y}})^t\right] \quad , \quad \mathbf{R}_{\mathbf{Y}} = \mathbb{E}[\mathbf{Y}\mathbf{Y}^t]$$

$$\mathbf{R}_{\mathbf{XY}} = \mathbb{E}[\mathbf{X}\mathbf{Y}^t] \quad , \quad \mathbf{C}_{\mathbf{XY}} = \mathbb{E}\left[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{Y} - \mu_{\mathbf{Y}})^t\right]$$

Estimador Lineal de Menor Error Cuadrático Medio (LMMSE)

Es interesante limitar la búsqueda de $\hat{\mathbf{X}} = g(\mathbf{Y})$ a estimadores del tipo,

$$g(\mathbf{Y}) = \mathbf{AY} + \mathbf{b}.$$

donde $\mathbf{A} \in \mathbb{C}^{n\times m}$ y $\mathbf{b} \in \mathbb{C}^{n}$ son constantes. Para minimizar el MSE planteamos el principio de ortogonalidad, es decir

$$\mathbb{E}\left[\underbrace{(\mathbf{AY}+\mathbf{b})^t}_{\hat{\mathbf{X}}^t}\underbrace{(\mathbf{X}-\mathbf{AY}-\mathbf{b})}_{\mathbf{X}-\hat{\mathbf{X}}}\right] = 0 \qquad (1)$$

LMMSE

Por el principio de ortogonalidad, tenemos

$$\mathbb{E}\left[\underbrace{(\mathbf{AY}+\mathbf{b})^t}_{\hat{\mathbf{X}}^t}\underbrace{(\mathbf{X}-\mathbf{AY}-\mathbf{b})}_{\mathbf{X}-\hat{\mathbf{X}}}\right]=0.$$

Como $\hat{\mathbf{X}}^t(\mathbf{X}-\hat{\mathbf{X}})\in\mathbb{R}\Longrightarrow\hat{\mathbf{X}}^t(\mathbf{X}-\hat{\mathbf{X}})=\operatorname{tr}\left[\hat{\mathbf{X}}^t(\mathbf{X}-\hat{\mathbf{X}})\right].$

Entonces, buscamos **A** y **b** tal que

$$\mathbb{E}\left\{\operatorname{tr}\left[(\mathbf{AY}+\mathbf{b})^t(\mathbf{X}-\mathbf{AY}-\mathbf{b})\right]\right\}=0$$

LMMSE

Recordando que $\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB)$, desarrollamos

$$
\begin{aligned}
0 &= \mathbb{E}\left\{\operatorname{tr}\left[(\mathbf{AY}+\mathbf{b})^t(\mathbf{X}-\mathbf{AY}-\mathbf{b})\right]\right\} \\
&= \mathbb{E}\left\{\operatorname{tr}\left[\mathbf{Y}^t\mathbf{A}^t\mathbf{X}+\mathbf{b}^t\mathbf{X}-\mathbf{Y}^t\mathbf{A}^t\mathbf{AY}-\mathbf{b}^t\mathbf{AY}-\mathbf{Y}^t\mathbf{A}^t\mathbf{b}-\mathbf{b}^t\mathbf{b}\right]\right\} \\
&= \mathbb{E}\left\{\operatorname{tr}\left[\mathbf{Y}^t\mathbf{A}^t\mathbf{X}\right]\right\}+\mathbb{E}\left[\mathbf{b}^t\mathbf{X}\right]-\mathbb{E}\left\{\operatorname{tr}[\mathbf{Y}^t\mathbf{A}^t\mathbf{AY}]\right\}-\mathbb{E}\left[\mathbf{b}^t\mathbf{AY}\right]-\mathbb{E}\left[\mathbf{Y}^t\mathbf{A}^t\mathbf{b}\right]-\mathbf{b}^t\mathbf{b} \\
&= \mathbb{E}\left\{\operatorname{tr}\left[\mathbf{A}^t\mathbf{XY}^t\right]\right\}+\mathbf{b}^t\mu_{\mathbf{X}}-\mathbb{E}\left\{\operatorname{tr}[\mathbf{A}^t\mathbf{AYY}^t]\right\}-\mathbf{b}^t\mathbf{A}\mu_{\mathbf{Y}}-\mu_{\mathbf{Y}}^t\mathbf{A}^t\mathbf{b}-\mathbf{b}^t\mathbf{b} \\
&= \operatorname{tr}\left\{\mathbb{E}\left[\mathbf{A}^t\mathbf{XY}^t\right]\right\}+\mathbf{b}^t\mu_{\mathbf{X}}-\operatorname{tr}\left\{\mathbb{E}[\mathbf{A}^t\mathbf{AYY}^t]\right\}-\mathbf{b}^t\mathbf{A}\mu_{\mathbf{Y}}-\mu_{\mathbf{Y}}^t\mathbf{A}^t\mathbf{b}-\mathbf{b}^t\mathbf{b} \\
&= \operatorname{tr}\left\{\mathbf{A}^t\mathbb{E}\left[\mathbf{XY}^t\right]\right\}+\mathbf{b}^t\mu_{\mathbf{X}}-\operatorname{tr}\left\{\mathbf{A}^t\mathbf{A}\mathbb{E}[\mathbf{YY}^t]\right\}-\mathbf{b}^t\mathbf{A}\mu_{\mathbf{Y}}-\mu_{\mathbf{Y}}^t\mathbf{A}^t\mathbf{b}-\mathbf{b}^t\mathbf{b} \\
&= \operatorname{tr}\left[\mathbf{A}^t\left(\mathbf{R}_{\mathbf{XY}}-\mathbf{AR}_{\mathbf{Y}}-\mathbf{b}\mu_{\mathbf{Y}}^t\right)\right]+\mathbf{b}^t\left[\mu_{\mathbf{X}}-\mathbf{A}\mu_{\mathbf{Y}}-\mathbf{b}\right].
\end{aligned}
$$

LMMSE

La matriz **A** y el vector **b** que corresponden al estimador LMMSE deben cumplir el siguiente sistema de ecuaciones:

$$\begin{cases} \mathbf{A}\mathbf{R}_{\mathbf{Y}} + \mathbf{b}\mu_{\mathbf{Y}}^{t} = \mathbf{R}_{\mathbf{XY}} \\ \mathbf{A}\mu_{\mathbf{Y}} + \mathbf{b} = \mu_{\mathbf{X}} \end{cases}$$

cuya solución es

$$\mathbf{A}_{lmmse} = \left[\mathbf{R}_{\mathbf{XY}} - \mu_{\mathbf{X}}\mu_{\mathbf{Y}}^{t}\right]\left[\mathbf{R}_{\mathbf{Y}} - \mu_{\mathbf{Y}}\mu_{\mathbf{Y}}^{t}\right]^{-1} = \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{Y}}^{-1}$$

$$\mathbf{b}_{lmmse} = \mu_{\mathbf{X}} - \left[\mathbf{R}_{\mathbf{XY}} - \mu_{\mathbf{X}}\mu_{\mathbf{Y}}^{t}\right]\left[\mathbf{R}_{\mathbf{Y}} - \mu_{\mathbf{Y}}\mu_{\mathbf{Y}}^{t}\right]^{-1}\mu_{\mathbf{Y}} = \mu_{\mathbf{X}} - \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{Y}}^{-1}\mu_{\mathbf{Y}}$$

LMMSE

Estimador Lineal de Menor Error Cuadrático Medio (LMMSE)

$$\hat{\mathbf{X}}_{lmmse} = \mu_{\mathbf{X}} + \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{Y}}^{-1}\left(\mathbf{Y} - \mu_{\mathbf{Y}}\right). \tag{2}$$

Cuando **X** e **Y** son dos vectores conjuntamente gaussianos,

$$\hat{\mathbf{X}}_{lmmse} = \hat{\mathbf{X}}_{mmse}$$

LMMSE: Propiedades

- $\hat{\mathbf{X}}_{lmmse}$ es un estimador *insesgado*. Se lo conoce también como el estimador BLU (Best Linear Unbiased).

$$\mathbb{E}[\hat{\mathbf{X}}_{lmmse}] = \mathbb{E}[\mu_{\mathbf{X}} + \mathbf{C_{XY}C_Y^{-1}}(\mathbf{Y} - \mu_{\mathbf{Y}})] = \mu_{\mathbf{X}}$$

LMMSE: Propiedades

El error de estimación es:

$$\mathbf{E}_{lmmse} = \mathbf{X} - \hat{\mathbf{X}}_{lmmse} = \underbrace{(\mathbf{X} - \mu_{\mathbf{X}})}_{\delta_{\mathbf{X}}} - \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{Y}}^{-1} \underbrace{(\mathbf{Y} - \mu_{\mathbf{Y}})}_{\delta_{\mathbf{Y}}}.$$

Recordemos que

$$\mathbf{C}_{\mathbf{X}} = \mathbb{E}[\delta_{\mathbf{X}}\delta_{\mathbf{X}}^t] \quad , \quad \mathbf{C}_{\mathbf{Y}} = \mathbb{E}[\delta_{\mathbf{Y}}\delta_{\mathbf{Y}}^t] \quad , \quad \mathbf{C}_{\mathbf{XY}} = \mathbb{E}[\delta_{\mathbf{X}}\delta_{\mathbf{Y}}^t].$$

Luego, la covarianza del error resulta:

$$\begin{aligned}
\mathbf{Cov}[\mathbf{E}_{lmmse}] &= \mathbb{E}\left[\mathbf{E}_{lmmse}\mathbf{E}_{lmmse}^t\right] = \mathbb{E}\left[\left(\delta_{\mathbf{X}} - \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{Y}}^{-\mathbf{1}}\delta_{\mathbf{Y}}\right)\left(\delta_{\mathbf{X}} - \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{Y}}^{-\mathbf{1}}\delta_{\mathbf{Y}}\right)^t\right] \\
&= \mathbb{E}\left[\delta_{\mathbf{X}}\delta_{\mathbf{X}}^{\mathbf{t}}\right] - \mathbb{E}\left[\delta_{\mathbf{X}}\delta_{\mathbf{Y}}^{\mathbf{t}}\right]\mathbf{C}_{\mathbf{Y}}^{-1}\mathbf{C}_{\mathbf{XY}}^t - \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{Y}}^{-1}\mathbb{E}\left[\delta_{\mathbf{Y}}\delta_{\mathbf{X}}^{\mathbf{t}}\right] \\
&\qquad + \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{Y}}^{-1}\mathbb{E}\left[\delta_{\mathbf{Y}}\delta_{\mathbf{Y}}^{\mathbf{t}}\right]\mathbf{C}_{\mathbf{Y}}^{-1}\mathbf{C}_{\mathbf{XY}}^t \\
&= \mathbf{C}_{\mathbf{X}} - \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{Y}}^{-1}\mathbf{C}_{\mathbf{YX}}
\end{aligned}$$

LMMSE: Propiedades

Finalmente, el error cuadrático medio es

$$\begin{aligned}
MMSE &= \mathbb{E}\left[\|\mathbf{E}_{lmmse}\|^2\right] = \mathbb{E}[\mathbf{E}_{lmmse}^t \mathbf{E}_{lmmse}] \\
&= \mathbb{E}\left[\mathrm{tr}(\mathbf{E}_{lmmse}^t \mathbf{E}_{lmmse})\right] = \mathbb{E}\left[\mathrm{tr}(\mathbf{E}_{lmmse} \mathbf{E}_{lmmse}^t)\right] \\
&= \mathrm{tr}\left\{\mathbb{E}\left[\mathbf{E}_{lmmse} \mathbf{E}_{lmmse}^t\right]\right\} \\
&= \mathrm{tr}\left\{\mathbf{Cov}[\mathbf{E}_{lmmse}]\right\}
\end{aligned}$$

LMMSE: resumen

$$\hat{\mathbf{X}}_{lmmse} = \mu_{\mathbf{X}} + \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{Y}}^{-1}\left(\mathbf{Y} - \mu_{\mathbf{Y}}\right).$$

$$\mathbb{E}[\hat{\mathbf{X}}_{lmmse}] = \mu_{\mathbf{X}}.$$

$$\mathbf{Cov}[\mathbf{E}_{lmmse}] = \mathbf{C}_{\mathbf{X}} - \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{Y}}^{-1}\mathbf{C}_{\mathbf{YX}}.$$

$$MMSE = \operatorname{tr}\left\{\mathbf{C}_{\mathbf{X}} - \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{Y}}^{-1}\mathbf{C}_{\mathbf{YX}}\right\}.$$

Estimación lineal de procesos aleatorios

Sean $X(n)$ e $Y(n)$ dos procesos aleatorios.

Problema

A partir de muestras de una realización $y(m)$, recolectadas en el intervalo $m_{init} \leq m \leq m_{fin}$, hallar el estimador

$$\hat{x}(n) = \sum_{m=m_{init}}^{m_{fin}} k_{nm} y(m)$$

que minimice el error cuadrático medio

$$\mathbb{E}\left[\left\|X - \hat{X}_{lmmse}\right\|^2\right] = \mathbb{E}\left[\sum_n \left|X(n) - \hat{X}_{lmmse}(n)\right|^2\right]$$

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

Dos posibles esquemas de estimación

1. **Suavizado** (*smoothing*) $m_{init} = m_0$, $m_{fin} = m_0 + M$.

   Dado $\{y(m), m_0 \leq m \leq m_0 + M\}$ *obtener* $\{\hat{x}(n), n_1 \leq n \leq n_2\}$ *donde* $m_0 \leq n_1 \leq n_2 \leq m_0 + M$*.*

   Estimador *no causal* ya que utiliza observaciones futuras.

2. **Filtrado** $m_{init} = n - M$, $m_{fin} = n$.

   Dado $\{y(m), n - M \leq m \leq n\}$ *obtener* $\hat{x}(n)$*.*

   Estimador *causal* que utiliza pasado y presente.

Análisis del problema

- Vamos a trabajar con observaciones en una ventana de duración finita es decir que todas las observaciones se pueden agrupar en un vector
$$\mathbf{Y} = \begin{bmatrix} Y(m_{init}) \\ \vdots \\ Y(m_{fin}) \end{bmatrix}$$
- Para cada valor de n, se puede considerar un problema de estimación de una VA, $X_n = X(n)$ a partir de un VeA, $\mathbf{Y}$.
- La solución LMMSE a este problema la conocemos
$$\hat{X}_{lmmse}(n) = \mu_{X_n} + \mathbf{C}_{X_n\mathbf{Y}}\mathbf{C}_{\mathbf{Y}}^{-1}\left(\mathbf{Y} - \mu_{\mathbf{Y}}\right).$$
- El problema es entonces poder implementar esta solución en forma eficiente para cada instante de tiempo n.

Hipótesis

Asumimos lo siguiente:

- $X(n)$ e $Y(n)$ son conjuntamente ESA
- Las correlaciones siguientes son conocidas:

$$R_{XY}(k) = \mathbb{E}[X(n)Y(n+k)] \quad \text{y} \quad R_Y(k) = \mathbb{E}[Y(n)Y(n+k)].$$

- Como el estimador LMMSE es insesgado, sin falta de generalidad

$$\mathbb{E}[X(n)] = \mathbb{E}[Y(n)] = 0.$$

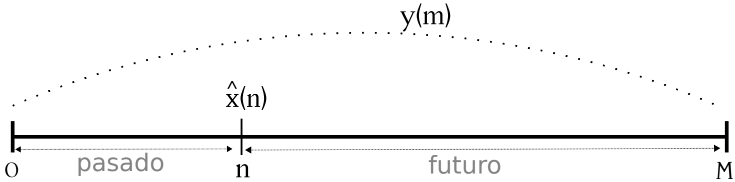
Entonces,

$$\hat{X}_{lmmse}(n) = \mathbf{R}_{X_n\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{Y}.$$

Problema de suavizado en una ventana fija

Como los procesos son CESA, sin falta de generalidad $m_0 = 0$.

$$\{y(m), m = 0, \cdots, M\} \quad \longrightarrow \quad \hat{x}(n),\ 0 \leq n \leq M.$$



$$\hat{X}_{lmmse}(n) = \hat{X}(n) = \mathbf{R}_{X_n\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{Y}.$$

Problema de suavizado en una ventana fija (cont)

- *Matriz de autocorrelación* : $\mathbf{R_Y} \in \mathbb{R}^{(M+1)\times(M+1)}$

$$\mathbf{R_Y} = \mathbb{E}\left[\mathbf{YY}^t\right] = \mathbb{E}\left\{ \begin{bmatrix} Y(0) \\ \vdots \\ Y(M) \end{bmatrix} \begin{bmatrix} Y(0) & \cdots & Y(M) \end{bmatrix} \right\}$$

$$= \begin{bmatrix} R_Y(0) & \cdots & R_Y(M) \\ & \ddots & \\ R_Y(M) & \cdots & R_Y(0) \end{bmatrix}$$

- Matriz Toeplitz simétrica.
- Se forma a partir de $R_Y(k)$, *función de autocorrelación* de $Y(n)$.
- Si $Y(n)$ es ESA, $\mathbf{R_Y}$ no depende de n.

Problema de suavizado en una ventana fija (cont)

- *Vector de correlación cruzada* : $\mathbf{R}_{X_n\mathbf{Y}} \in \mathbb{R}^{1\times(M+1)}$,

$$\begin{aligned}\mathbf{R}_{X_n\mathbf{Y}} &= \mathbb{E}[X(n)\mathbf{Y}^t] \\ &= \mathbb{E}\left\{X(n)\left[Y(0),\ \ Y(1),\ \ \cdots, Y(n),\ \ \cdots, Y(M)\right]\right\} \\ &= [R_{XY}(-n),\ \ R_{XY}(-n+1)\ \ \cdots, R_{XY}(0),\ \ \cdots, R_{XY}(M-n)]\end{aligned}$$

  $R_{XY}(k)$: *correlación cruzada* entre los procesos $X(n)$ e $Y(n)$.

- Pese a que $X(n)$ e $Y(n)$ son CESA, el vector $\mathbf{R}_{X_n\mathbf{Y}}$ depende de n.

Suavizado en una ventana fija

- Luego, en cada instante,

$$\hat{X}(n) = \mathbf{k}_n \mathbf{Y} \quad \text{con} \quad \mathbf{k}_n = \mathbf{R}_{X_n\mathbf{Y}} \mathbf{R}_{\mathbf{Y}}^{-1}.$$

- El procesamiento para resolver el problema de suavizado es no-causal y variante en el tiempo

Suavizado en una ventana fija: error de estimación

Sea $E(n) = X(n) - \hat{X}(n)$, el error de estimación.

A priori, antes de realizar observación alguna, $\mathbb{V}[X(n)] = \sigma_X^2$ 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 $\mathbb{V}[E(n)] = \sigma_E^2(n)$.

Suavizado en una ventana fija: error de estimación

A partir del resultado del estimador LMMSE tenemos para cada valor de n:

- $\mathbb{E}[E(n)] = 0$ por ser un estimador insesgado.
- $\mathbb{V}[E(n)] = \mathbb{E}[|E(n)|^2] = \sigma_X^2 - \mathbf{R}_{X_n\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{X_n\mathbf{Y}}^t$ (ver [1] al pie)
- Como $\mathbf{R}_{\mathbf{Y}} > 0$, tenemos que

$$\mathbf{R}_{X_n\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{X_n\mathbf{Y}}^t > 0 \qquad \forall n.$$

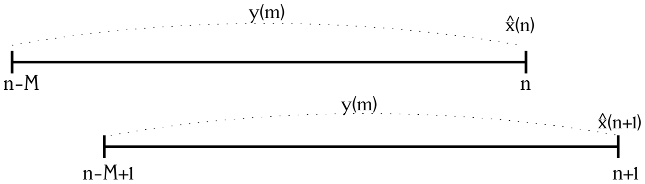
Luego, $\mathbb{V}[E(n)] < \sigma_X^2$ para todo n. Es decir, la estimación reduce la incertidumbre sobre $X(n)$.

Pregunta: El proceso $E(n)$ es ESA?

[1] Recordemos que en el caso general, $MMSE = \text{tr}\left\{\mathbf{C}_{\mathbf{X}} - \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{Y}}^{-1}\mathbf{C}_{\mathbf{YX}}\right\}$

Filtrado en una ventana fija

$$\{y(m), n - M \leq m \leq n\} \quad \longrightarrow \quad \hat{x}(n).$$



El filtrado tiene una solución *causal* que sólo utiliza observaciones pasadas y presentes.

Filtrado en una ventana fija: solución LMMSE

Para cada valor de n, queremos estimar $\hat{X}(n)$ a partir de la realización

$$y(n-M), y(n-M+1), \cdots y(n).$$

A diferencia del problema de suavizado, la ventana de observación del filtrado se desplaza con n.

Filtrado en una ventana fija: solución LMMSE

Armamos el vector **Y** que contiene las observaciones y calculamos la matriz de correlación

$$\mathbf{R_Y} = \mathbb{E}\left[\mathbf{Y}\mathbf{Y}^t\right] = \mathbb{E}\left\{\begin{bmatrix} Y(n-M) \\ \vdots \\ Y(n) \end{bmatrix}\begin{bmatrix} Y(n-M) & \cdots & Y(n) \end{bmatrix}\right\}$$

$$= \begin{bmatrix} R_Y(0) & \cdots & R_Y(M) \\ & \ddots & \\ R_Y(M) & \cdots & R_Y(0) \end{bmatrix}$$

Filtrado en una ventana fija: solución LMMSE

Como en el caso del suavizado, también armamos el vector de correlación cruzada a partir de $R_{XY}(k)$, la función de correlación cruzada entre los procesos $X(n)$ e $Y(n)$:

$$\begin{aligned}\mathbf{R}_{X_n\mathbf{Y}} &= \mathbb{E}\left[X(n)\mathbf{Y}^t\right] \\ &= \mathbb{E}\left\{X(n)\left[Y(n-M), \ \cdots \ Y(n-1), \ \ Y(n)\right]\right\} \\ &= \left[R_{XY}(M), \ \cdots R_{XY}(1) \ \ , R_{XY}(0)\right] = \mathbf{R}_{X\mathbf{Y}}.\end{aligned}$$

En este caso, el vector $\mathbf{R}_{X_n\mathbf{Y}} = \mathbf{R}_{X\mathbf{Y}}$ no depende de n.

Filtrado en una ventana fija

Finalmente, $\hat{X}(n) = \underbrace{\mathbf{R}_{X\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}}_{\boldsymbol{w}^t}\,\mathbf{Y}$

donde

$$\boldsymbol{w} = \mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{X\mathbf{Y}}^{t} = \begin{bmatrix} w(M) \\ w(M-1) \\ \vdots \\ w(0) \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} y(n-M) \\ y(n-M+1) \\ \vdots \\ y(n) \end{bmatrix}.$$

Luego,

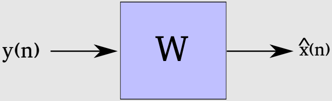
$$\begin{aligned} \hat{X}(n) &= w(M)y(n-M) + w(M-1)y(n-M+1) + \cdots + w(0)y(n) \\ &= \sum_{s=0}^{M} w(s)y(n-s). \end{aligned}$$

Filtrado en una ventana fija (cont)

$$\hat{X}(n) = \sum_{s=0}^{M} w(s)y(n-s) = \sum_{m=n-M}^{n} w(n-m)y(m).$$

Asociamos los coeficientes $w(s)$ a la respuesta impulsiva de un filtro FIR de longitud $M+1$ que realiza el filtrado causal. Este filtro se lo conoce como *Filtro de Wiener*.

Filtro de Wiener



$\hat{X}_{lmmse}(n)$ resulta al filtrar linealmente las observaciones $Y(n)$.

- La respuesta impulsiva FIR del filtro W se obtiene como los componentes del vector

$$\boldsymbol{w} = \mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{X\mathbf{Y}}^{t}$$

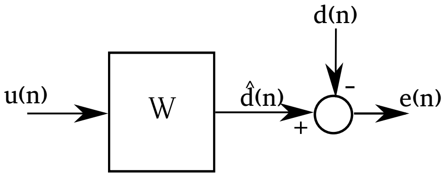
  donde $\mathbf{R}_{\mathbf{Y}}$ y $\mathbf{R}_{X\mathbf{Y}}$ depende del largo del filtro M pero no del instante de filtrado n.

- El *costo de Wiener* es igual a

$$J_{Wiener} = \mathbb{E}[|E(n)|^2] = \sigma_X^2 - \mathbf{R}_{X\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{X\mathbf{Y}}^{t}$$

Notación

En cierta bibliografía del tema, se plantea el problema de filtrado con la siguiente nomenclatura, donde W es un filtro FIR de longitud $M+1$

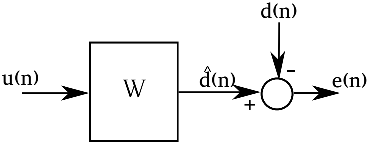


$$\hat{d}(n) = \sum_{s=0}^{M} w(s)u(n-s)$$

Este problema es equivalente al ya visto si se considera que:

- Las observaciones $y(n)$ son las *entradas* del filtro $u(n)$.
- El proceso $x(n)$ es la señal *deseada* $d(n)$.

Notación



Principio de ortogonalidad:

$$\mathbb{E}\left[U^t(n)E(n)\right] = 0$$

Planteando las ecuaciones normales:

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

Retomamos el diseño de W

Hipótesis:

1. $X(n)$ e $Y(n)$ son conjuntamente ESA,
2. $R_Y(k)$ y $R_{XY}(k)$ son funciones conocidas.

Qué sucede si

- Los procesos no son estacionarios?
  - Ejemplo 1: $X(n)$ tiene una deriva en el tiempo y $\mathbb{E}[X(n)] = \mu_X(n)$.
  - Ejemplo 2: $X(n)$ es una constante, pero la fuente del ruido que afecta la observación varía con el tiempo.
- No se conoce la estadística de $X(n)$ o de $Y(n)$?
  - Por ejemplo, se sabe que $Y(n) = X(n) + V(n)$, pero no se tienen observaciones previas que permitan estimar R_Y y/o R_{XY}.

Encaramos estos problemas estimando $X(n)$ a medida que observamos $Y(n)$, es decir, *adaptando* el diseño del filtro.

Volvamos al planteo del LMMSE

Vimos que el filtro de Wiener es un filtro FIR causal que obtiene el estimador LMMSE. Sea $w(l)$ la respuesta impulsiva del filtro,

$$\hat{X}(n) = \sum_{l=0}^{M} w(l)Y(n-l) = \underbrace{\begin{bmatrix} w(M) & \cdots & w(0) \end{bmatrix}}_{\mathbf{w}^t} \underbrace{\begin{bmatrix} Y(n-M) \\ \vdots \\ Y(n) \end{bmatrix}}_{\mathbf{Y}}$$

Volvamos al planteo del LMMSE

Para cada instante n, podemos calcular el error cuadrático medio,

$$\begin{aligned}\mathbb{E}\left[|X(n)-\mathbf{w}^t\mathbf{Y}|^2\right] &= \mathbb{E}[|X(n)|^2]-\mathbb{E}\left[X(n)\mathbf{Y}^t\mathbf{w}\right] \\ &\quad -\mathbb{E}\left[\mathbf{w}^t\mathbf{Y}X(n)\right]+\mathbb{E}\left[\mathbf{w}^t\mathbf{Y}\mathbf{Y}^t\mathbf{w}\right] \\ &= \sigma_X^2-\mathbf{R}_{X\mathbf{Y}}\mathbf{w}-\mathbf{w}^t\mathbf{R}_{\mathbf{Y}X}+\mathbf{w}^t\mathbf{R}_{\mathbf{Y}}\mathbf{w}\end{aligned}$$

Ésta es una función de $\boldsymbol{w}$ que llamamos $J(\boldsymbol{w})$. Esta función se la conoce como función costo. Claramente, la solución LMMSE minimiza $J(\boldsymbol{w})$, es decir,

$$\boldsymbol{w}_{lmmse}=\arg\min J(\boldsymbol{w})$$

Volvamos al planteo del LMMSE

Para minimizar $J(\boldsymbol{w})$ calculamos el gradiente.
Sea $w(l) = w_R(l) + \jmath w_I(l)$. Luego,

$$\nabla J = \begin{bmatrix} \frac{\partial J}{\partial w_R(M)} + \jmath \frac{\partial J}{\partial w_I(M)} \\ \vdots \\ \frac{\partial J}{\partial w_R(0)} + \jmath \frac{\partial J}{\partial w_I(0)} \end{bmatrix} = 2\left(\mathbf{R_Y w} - \mathbf{R}^t_{X\mathbf{Y}}\right).$$

Claramente, $\boldsymbol{w}_{lmmse}$ coincide con la solución de $\nabla J = 0$.

Pregunta: Es posible calcular $\boldsymbol{w}$ en forma recursiva?

Algoritmo *steepest descent* para optimización de funciones multivariables

- Sea $f : \mathbb{R}^L \to \mathbb{R}$ y $\bar{w}$ un mínimo local de f. Suponemos que $f(.)$ es diferenciable en todo punto de $B(\bar{w}) \subset \mathbb{R}^L$, un vecindario de $\bar{w}$.
- Para hallar $\bar{w}$ de forma numérica armamos una trayectoria que luego de n pasos resulta

$$w_{n+1} = w_n + p$$

donde p es la dirección de actualización[2].

[2] La referencia clásica para este tema es *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, ,J. E. Dennis, Jr. and Robert B. Schnabel, Society for Industrial and Applied Mathematics, donde se desarrolla este tema en el contexto general de optimización de funciones.

Algoritmo *steepest descent* para optimización de funciones multivariables

- Si $\nabla f^t p < 0$, y δ pequeño, $f(w_n + \delta p) < f(w_n)$
- La dirección de descenso más rápido es

$$p = -\frac{\nabla f}{\|\nabla f\|}$$

- Algoritmo de descenso por la mayor pendiente o *steepest descent*

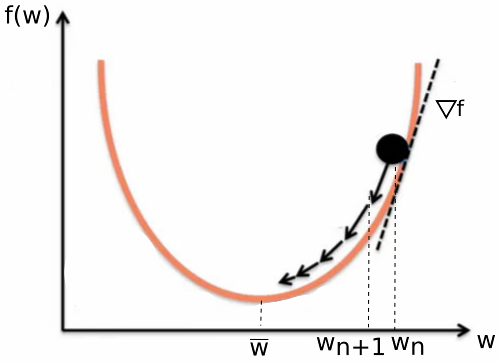
$$w_{n+1} = w_n - \mu \nabla f$$

donde $\mu > 0$ es un parámetro que define el tamaño del paso.

Algoritmo *steepest descent* para optimización de funciones multivariables

Si $\bar{w}$ es un mínimo global, el algoritmo iterativo converge, es decir

$$w_n \to \bar{w} \qquad n \to \infty.$$



Volvamos al filtro de Wiener

Vimos que el filtro de Wiener se obtiene minimizando la función costo

$$J(\mathbf{w}) = \sigma_X^2 - \mathbf{R}_{X\mathbf{Y}}\mathbf{w} - \mathbf{w}^t\mathbf{R}_{\mathbf{Y}X} + \mathbf{w}^t\mathbf{R}_{\mathbf{Y}}\mathbf{w} \qquad \boldsymbol{w} \in \mathbb{R}^{M+1}.$$

Más aún, vimos que

$$\nabla J \propto \left(\mathbf{R}_{\mathbf{Y}}\mathbf{w} - \mathbf{R}_{X\mathbf{Y}}^t\right).$$

Planteamos el esquema de steepest descent para minimizar $J(\boldsymbol{w})$.

Planteo iterativo del filtro de Wiener

Filtrado de Wiener utilizando Steepest-descent

A partir de $\mathbf{w}_n$, la siguiente actualización es

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \underbrace{\left(\mathbf{R}_{\mathbf{Y}}\mathbf{w}_n - \mathbf{R}^t_{X\mathbf{Y}}\right)}_{\nabla J} \qquad n = 0, 2, \cdots$$

En cada iteración, se obtienen las $M + 1$ componentes del vector

$$\boldsymbol{w}_n = \begin{bmatrix} w_n(M) \\ \vdots \\ w_n(0) \end{bmatrix}$$

El parámetro $\mu > 0$ controla la velocidad de convergencia.
Como $J(\boldsymbol{w})$ es convexa, $\boldsymbol{w}_{lmmse}$ es mínimo global y

$$\boldsymbol{w}_n \to \boldsymbol{w}_{lmmse}.$$

Steepest descent para obtener el filtro de Wiener

El término de correción luego del paso n es $\mathbf{R}_{\mathbf{Y}}\mathbf{w}_n - \mathbf{R}^t_{X(n)\mathbf{Y}}$.

$$\begin{aligned}
\mathbf{R}_{\mathbf{Y}}\mathbf{w}_n - \mathbf{R}^t_{X(n)\mathbf{Y}} &= \mathbb{E}\left[\mathbf{Y}\mathbf{Y}^t\right]\mathbf{w}_n - \mathbb{E}\left[\mathbf{Y}X^t(n)\right] \\
&= \mathbb{E}\left[\mathbf{Y}\underbrace{\left(\mathbf{w}_n^t\mathbf{Y} - X(n)\right)^t}_{\mathbf{E}^t(n)}\right] \\
&= \mathbb{E}\left[\mathbf{Y}\mathbf{E}(n)^t\right].
\end{aligned}$$

La correción en el paso n es igual al producto entre el error de estimación $E(n) = \mathbf{w}_n^t\mathbf{Y} - X(n)$ y el vector de observaciones $\mathbf{Y}$.

Algoritmo LMS

Para implementar el algoritmo de *steepest descent* se requiere conocer la estadística conjunta de $X(n)$ e $Y(n)$. Esto no siempre es posible, o puede resultar muy complejo. Una alternativa es el algoritmo iterativo conocido como LMS (*Least-Mean-Square*).

Como referencia para este tema, sugiero leer el artículo donde Bernard Widrow cuenta cómo llego a la formulación del algoritmo, "Thinking about thinking: the discovery of the LMS algorithm," B. Widrow, en IEEE Signal Processing Magazine, vol. 22, no. 1, pp. 100-106, Enero 2005

Algoritmo LMS

La idea del algoritmo LMS es aproximar el operador $\mathbb{E}[.]$ por el resultado de una realización del proceso, es decir

$$\mathbb{E}\left[\mathbf{Y}E(n)^t\right] \quad \text{pasa a ser} \quad \boldsymbol{y}(n)e(n)^t,$$

$$\boldsymbol{y}(n) = \begin{bmatrix} y(n-M) \\ \vdots \\ y(n) \end{bmatrix} \quad y(k) \text{ es la realización del proceso observado.}$$

$$e(n) = x(n) - \hat{x}(n) \quad \text{la realización del error de estimación}$$

Para hacer énfasis en el aspecto adaptativo, incluimos el tiempo n en el vector de observaciones $\boldsymbol{y}(n)$.

Adaptación de acuerdo al algoritmo *Least-Mean-Square*

Algoritmo LMS

Sea $\mathbf{w}_n$ el vector de coeficientes del filtro luego del n-ésimo paso de adaptación. En cada paso, tenemos:

$$\hat{x}(n) = \mathbf{w}_n^t \boldsymbol{y}(n)$$

y la adaptación

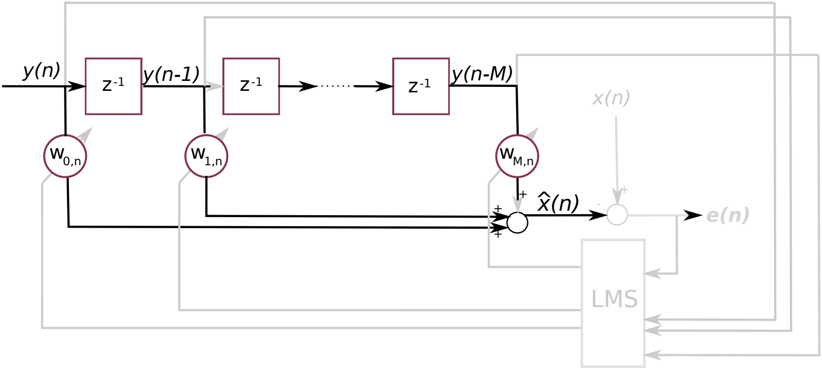
$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \left(\boldsymbol{y}(n) e(n)^t \right)$$

Adaptación de acuerdo al algoritmo *Least-Mean-Square*

Algunas observaciones preliminares

- La dirección de descenso del LMS no es determinística. Este algoritmo pertenece a la clase de algoritmos de *gradiente estocástico*.
- La convergencia del algoritmo no está asegurada, si bien su performance es muy buena bajo ciertas condiciones
- La enorme ventaja (y razón de su popularidad) del algoritmo LMS es su simplicidad.

Implementación del algoritmo



$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \left(\boldsymbol{y}(n) e(n)^t \right)$$

Convergencia del Algoritmo *Least-Mean-Square*

Vamos a analizar si $\mathbf{w}_n$ converge al filtro de Wiener a medida que aumenta n. Sea el vector de diferencias $\varepsilon_n = \mathbf{w}_n - \mathbf{w}_{lmmse}$.

- Como el límite de $\varepsilon_n \neq 0$, entonces $\mathbf{w}_n$ no converge a $\mathbf{w}_{lmmse}$.
- El valor del costo $J_{LMS}(\mathbf{w}_n)$ es superior al costo de Wiener, $J_{Wiener} = \sigma_x^2 - \mathbf{R}_{X\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{X\mathbf{Y}}^t$.
- $J(\mathbf{w}_n) = J_{Wiener} + J_{ex}(n)$ donde

$$\lambda_{min} \leq \frac{J_{ex}(n)}{\mathbb{E}[\|\varepsilon(n)\|^2]} \leq \lambda_{max} \qquad \forall n$$

Convergencia del Algoritmo

- En general, la convergencia del algoritmo LMS es *ruidosa*, ya que depende de una aproximación ruidosa de ∇J.
- El algoritmo converge a un valor mayor al costo de Wiener J_{opt}

Convergencia del Algoritmo

- Para convergencia, necesitamos

$$0 < \mu < \frac{2}{\mathrm{tr}(\mathbf{R_Y})}.$$

Cuando no se conoce $\mathbf{R_Y}$ o sus autovalores, es difícil verificar que se cumpla. Observemos que

$$\mathrm{tr}(\mathbf{R_Y}) = (M+1)R_Y(0) = (M+1)\mathbb{E}[|Y(n)|^2] \simeq \sum_{l=0}^{M} |y(n-l)|^2$$

donde $\sum_{l=0}^{M} |y(n-l)|^2$ es la energía contenida en la señal observada durante la ventana de duración $M+1$. Luego, una indicación para elegir μ puede ser

$$0 < \mu < \frac{2}{\sum_{l=0}^{M} |y(n-l)|^2}$$