# Filtrado lineal

Lorena Álvarez Pérez

Universidad Carlos III de Madrid

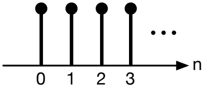
Curso académico 2024/2025

# Índice

# Señal discreta

- Secuencia indexada de números reales o complejos.
- Una secuencia es función de una variable entera $n \Rightarrow x[n]$.
- En esta asignatura, la variable independiente $n$ es el tiempo ($x[n]$ sólo está definida para **valores enteros de** $n$).
- Ejemplos:
  - Impulso unitario: $\delta[n] = \begin{cases} 1 & n = 0 \\ 0 & \text{en otro caso} \end{cases}$


  - Escalón unitario: $u[n] = \begin{cases} 1 & n \geq 0 \\ 0 & n < 0 \end{cases}$



# Clasificación de las señales discretas

- **Duración**
  - Secuencia finita
    - $x[n] = 0$ fuera de un intervalo $[N_1, N_2]$ de duración finita.
  - Secuencia infinita
    - Las que no son de duración finita.
    - Secuencia de lado izquierdo: $x[n] = 0$ para $n > n_0$
    - Secuencia de lado derecho: $x[n] = 0$ para $n < n_0$
- **Aleatoridad**
  - Secuencias deterministas
    - Pueden reproducirse exactamente con medidas repetidas
    - Ejemplo: respuesta al impulso de un sistema LTI
  - Secuencias aleatorias
    - No pueden repetirse de manera predecible
    - Ejemplo: ruido de un canal de comunicaciones

# Filtros discretos

- Un **filtro** es un **sistema** que transforma una señal de entrada ($x[n]$) en otra secuencia discreta ($y[n]$), denominada **salida del filtro**, mediante un conjunto fijo de reglas o funciones (Ejemplo: $y[n] = \left(x[n]\right)^2 + 3$).
- Propiedades de un sistema:
  - Memoria: La salida no depende de valores pasados ni futuros de la entrada.
  - Causal: La salida no depende de valores futuros de la entrada.
  - Invertible: Es posible recuperar la entrada al filtro conociendo la salida.
  - Estabilidad: Para cualquier $x[n]$ acotada, $y[n]$ también está acotada.
  - Linealidad: Dadas dos señales de entrada: $x_1[n]$ y $x_2[n]$, cuyas salidas respectivas son $y_1[n]$ e $y_2[n]$, cuando se presenta a la entrada del sistema $z[n] = \alpha x_1[n] + \beta x_2[n]$, la salida se puede obtener como $y[n] = \alpha y_1[n] + \beta y_2[n]$.
  - Invarianza: Dadas $x[n]$ e $y[n]$, se cumple que $\mathcal{T}\{x[n-n_0]\} = y[n-n_0]$, para cualquier valor de $n_0$.
- Si un filtro es **lineal e invariante en tiempo** (LTI), puede caracterizarse mediante su **respuesta al impulso** $h[n]$.

$$y[n] = x[n] * h[n] = \sum_{k=-\infty}^{\infty} h[k]x[n-k]$$

# Tipos de filtros LTI

Según sea $h[n]$, existen dos tipos de filtros:

- Filtros FIR (*Finite Impulse Response*)
  - El número de coeficientes no-ceros de $h[n]$ es una secuencia finita de $M$ instantes consecutivos de tiempo. De esta forma, se puede calcular la salida del filtro como se muestra a continuación:

$$y[n] = x[n] * h[n] = \sum_{k=-\infty}^{\infty} h[k]x[n-k] = \sum_{k=0}^{M-1} h[k]x[n-k]$$

- Filtros IIR (*Infinite Impulse Response*)
  - El número de coeficientes no-ceros de $h[n]$ es infinito ($h[n] \neq 0$ para $n \geq 0$). La convolución no se puede implementar físicamente, y se describe el filtro IIR con ecuaciones en diferencia:

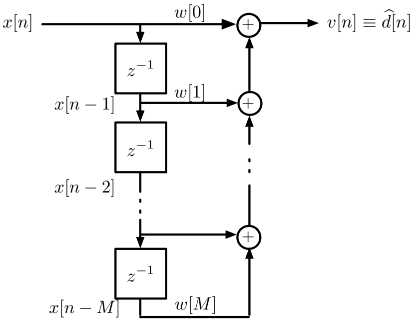
$$\sum_{k=0}^{M} a[k]y[n-k] = \sum_{j=0}^{N} b[j]x[n-j]$$

Para obtener $h[n]$, hay que recurrir a la transformada Z.

# Filtros FIR

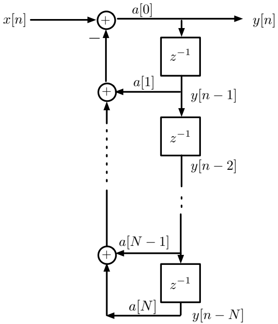
$w[n]$ ($\equiv h[n]$) tiene duración finita (supongamos, $M+1$ coeficientes distintos de cero).

$$\widehat{d}[n] = x[n] * w[n] = \sum_{k=0}^{M} w[k]x[n-k]$$



# Modelo autorregresivo

- Autorregresivo (AR): $x[n] = \sum_{k=0}^{N} a[k]y[n-k]$

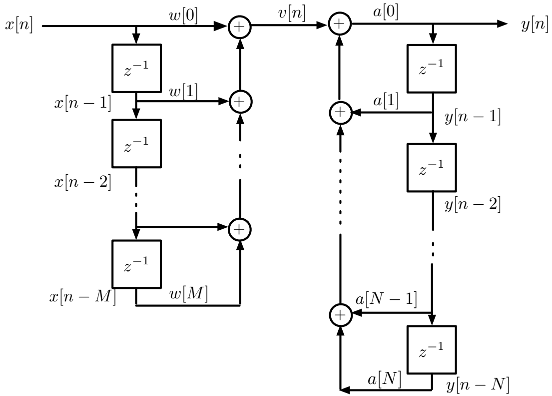



$$(a[0]y[n] + a[1]y[n-1] + a[2]y[n-2] + \ldots + a[N]y[n-N] = x[n])$$

# Filtros IIR

- Filtro (ARMA): Parte MA conectada en serie con una parte AR.

$$\sum_{l=0}^{N} a[l]y[n-l] = \sum_{k=0}^{M} w[k]x[n-k]$$



# Representación vectorial de los filtros

La salida del filtro ARMA se puede obtener:

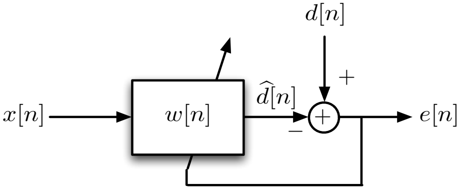
$$y[n] = \frac{1}{a[0]}\left(\sum_{k=0}^{M} w[k]x[n-k] - \sum_{l=1}^{N} a[l]y[n-l]\right)$$

$$= \frac{1}{a_0}\left[w_0, \ldots, w_M, a_1, \ldots, a_N\right] \times \begin{bmatrix} x[n] \\ \vdots \\ x[n-M] \\ y[n-1] \\ \vdots \\ y[n-N] \end{bmatrix} = \mathbf{w}^{\mathsf{T}} \times \begin{bmatrix} x[n] \\ \vdots \\ x[n-M] \\ y[n-1] \\ \vdots \\ y[n-N] \end{bmatrix}$$

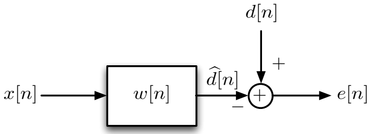
donde $\mathbf{w}^{\mathsf{T}} = \frac{1}{a_0}\left[w_0, \ldots, w_M, a_1, \ldots, a_N\right]^{\mathsf{T}}$ son los coeficientes del filtro.

# Filtros estacionarios y adaptativos

- Filtros estacionarios
  - Los coeficientes del filtro se obtienen una única vez y se mantienen fijos durante toda la vida operativa del filtro.
  - El vector de pesos es constante en el tiempo.
- Filtros adaptativos
  - Los coeficientes del filtro se actualizan constantemente cada vez que se realiza una predicción.
  - El vector de pesos varía con el tiempo (sus valores dependen del instante *n*).



# Elementos en un problema de diseño de filtros




- **Señal de entrada** ($x[n]$). Es la señal que va a ser "transformada" con el filtro.
- **Señal de referencia** ($d[n]$). El objetivo del filtro es "transformar" $x[n]$ para que sea una buena aproximación de $d[n]$. Es necesario establecer un criterio para evaluar la calidad de esta aproximación.
- **Señal de salida** ($y[n] = \widehat{d}[n]$).
- **Respuesta al impulso del filtro** ($w[n]$ o $h[n]$). Es la función que tenemos que aprender (en esta asignatura, esta función es un conjunto de coeficientes —los filtros con los que trabajamos son lineales).
- **Señal de error**: $e[n] = d[n] - \left(x[n] * w[n]\right)$. El objetivo del filtro es minimizar la energía de la señal de error tal que $\widehat{d}[n] = x[n] * w[n]$ sea un buena aproximación de $d[n]$.

# Problemas que se resuelven con filtros

- **Filtrado de ruido**: ésta es una de las aplicaciones más utilizadas de los filtros. Los elementos son:
  - $d[n]$ es la señal sin ruido (señal deseada).
  - $x[n]$ es $d[n]$ contaminada con ruido.

  El objetivo del filtro es aprender $w[n]$ de tal forma que transforme la señal $x[n]$ en una buena aproximación de su versión **sin ruido**.

  Una vez que el filtro es diseñado, la señal de referencia ($d[n]$) se "desconecta" y el filtro puede usarse para *limpiar* cualquier otra señal de entrada (distinta de $x[n]$ que es la que se utilizó para diseñar el filtro).

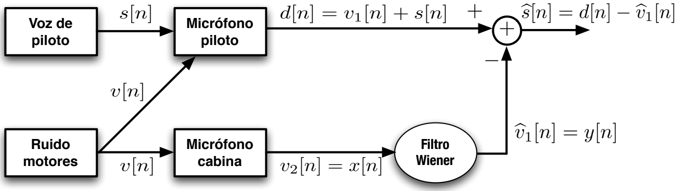
  (En la página 12 se muestra una configuración de un filtro de eliminación de ruido)

# Problemas que se resuelven con filtros

- **Predicción**:
  - La señal de referencia ($d[n]$) es una versión desplazada (en tiempo) de $x[n]$. De esta forma, el filtro **predice** los valores de $x[n]$ en instantes de tiempo futuros como combinación de los valores de $x[n]$ en instantes de tiempo pasados.
  - Este esquema puede usarse para comprimir la señal de entrada. Si el predictor es bueno, sólo se necesita transmitir el predictor y la señal de error. Además, si dicho predictor es bueno, los valores de la señal de error serán pequeños y, por tanto, pueden codificarse con un menor número de bits.

# Problemas que se resuelven con filtros

- **Cancelación de interferencias**: La señal de entrada $v[n]$ es una señal interferente o ruido, y la señal deseada $d[n]$ está formada por la suma de una señal de interés $s[n]$ y una componente interferente $v_1[n]$ relacionada con $v[n]$. El objetivo del filtro es estimar esa componente interferente a partir de $v[n]$ y restarla de $d[n]$ para recuperar la señal de interés ($s[n]$, voz del piloto en el ejemplo).



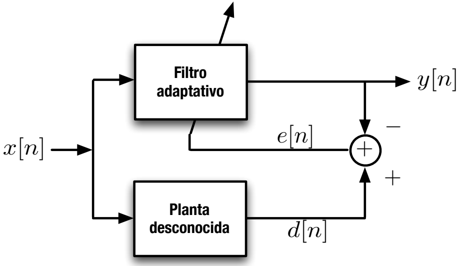
# Problemas que se resuelven con filtros

- **Identificación de planta (o de sistema)**:
  El objetivo es construir una aproximación a un sistema *desconocido* para que éste se comporte como una **caja negra** para nosotros.
  - La señal de referencia se construye introduciendo una señal de entrada $x[n]$ al sistema desconocido. Es decir, $d[n]$ es la salida del sistema desconocido cuando la entrada es $x[n]$.
  - La señal de entrada al filtro es $x[n]$.

  El objetivo del filtro es encontrar $w[n]$ tal que $y[n] = x[n] * w[n]$ sea una buena aproximación de $d[n]$.

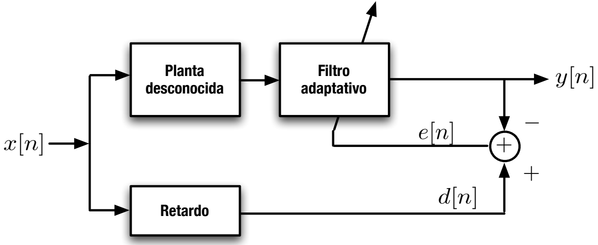
# Problemas que se resuelven con filtros

- **Identificación de planta (o de sistema)**:
  Cuanto más se parezca $y[n]$ a $d[n]$, mejor será la suposición de que el filtro lineal $w[n]$ es una buena aproximación al sistema desconocido.



# Problemas que se resuelven con filtros

- **Ecualizador**:
  - El objetivo del filtro es aprender el sistema inverso al sistema desconocido, de modo que pueda contrarrestarse. Éste es un problema fundamental en comunicaciones: ecualización del canal.



# Problemas que se resuelven con filtros

- **Ecualizador**:
  - Asumamos que la señal original transmitida es $s[n]$. Durante su transmisión, el canal (que es un sistema propiamente dicho) transforma $s[n]$ en $x[n]$.
  - El diseño del ecualizador usa:
    - Señal de entrada: $x[n]$
    - Señal de referencia: $d[n] = x[n - M]$ es una versión retardada de $s[n]$ ($s[n - M]$, siendo $M$ el número de muestras), donde el valor de $M$ depende del canal.

# Problemas que se resuelven con filtros

- **Ecualizador**:

  La razón de ser ésta la señal de referencia es la siguiente. Imaginemos que el canal es un sistema LTI FIR cuya $h[n]$ tiene 3 coeficientes.

$$x[n] = s[n] \cdot h[0] + s[n-1] \cdot h[1] + s[n-2] \cdot h[2]$$

Cada valor de la señal $x[n]$ se construye como combinación lineal de las muestras de $s[n]$. En otras palabras, cada muestra particular de $s[n]$, $s[n_*]$, participa en el valor de algunas muestras de $x[n]$:

$$\begin{aligned} x[n_*] &= s[n_*] \cdot h[0] + s[n_* - 1] \cdot h[1] + s[n_* - 2] \cdot h[2] \\ x[n_* + 1] &= s[n_* + 1] \cdot h[0] + s[n_*] \cdot h[1] + s[n_* - 1] \cdot h[2] \\ x[n_* + 2] &= s[n_* + 2] \cdot h[0] + s[n_* + 1] \cdot h[1] + s[n_*] \cdot h[2] \end{aligned}$$

Así que $x[n_*]$, $x[n_* + 1]$ y $x[n_* + 2]$ podrían usarse para encontrar el valor de $s[n_*]$.

# Problemas que se resuelven con filtros

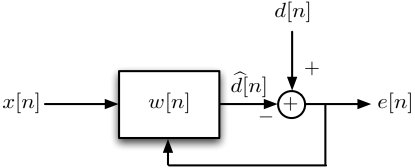
- **Ecualizador**:
  - Dado que viajar al futuro todavía está en desarrollo, es necesario esperar a la llegada de $x[n_* + 1]$ y $x[n_* + 2]$ antes de averiguar el valor de $s[n_*]$. Ésta es la razón por la que para invertir el canal, la señal de referencia debe ser una versión retardada de la señal original $s[n]$.
  - En otras palabras, el objetivo del filtro igualador es construir una buena aproximación de $s[n]$ como combinación lineal de $x[n]$, $x[n+1]$ ... $x[n+M-1]$. O, si usamos $x[n]$ como referencia, construir una buena aproximación de $d[n] = s[n-M]$ con $x[n]$, $x[n-1]$ ... $x[n-M+1]$.

# Filtrado lineal óptimo - Filtro de Wiener

- El filtro de Wiener es el **mejor filtro lineal de mínimos cuadrados** (óptimo).
- Pueden ser usados para tareas de predicción, estimación, interpolación, filtrado de señal y ruido, etc.
- Para diseñarlos, se necesita tener un conocimiento previo apropiado de las propiedades estadísticas de la señal de entrada y de la señal de referencia. El problema reside en que este conocimiento generalmente no se puede obtener.

# Planteamiento del problema

- Es un sistema al que le llegan dos señales: $x[n]$ y $\widehat{d}[n]$ y el objetivo es **determinar la respuesta al impulso** $w[n]$ de tal manera que el filtro produzca a su salida una señal $\widehat{d}[n]$, que sea una **reconstrucción óptima de** $d[n]$.




- $x[n]$ y $d[n]$ son realizaciones de procesos estocásticos conjuntamente WSS[1], de media nula (si no, basta restársela).
- $e[n] = d[n] - \widehat{d}[n]$: error de reconstrucción en el instante $n$.
- **Objetivo**: minimizar $e[n]$ en algún sentido estadístico.

[1] Sentido amplio (WSS, de *Wide Sense Stationary*)

# Restricciones y opciones de diseño

- **Sobre el filtro**:
  - Restricciones: lineal y discreto
  - Opciones: FIR o IIR (causal o no causal). Trataremos la teoría para FIRs por ser inherentemente estables.
- **Sobre las señales**:
  - Procesos estocásticos estacionarios con características espectrales conocidas o autocorrelación y correlación cruzada conocidas.
- **Sobre el criterio estadístico de optimalidad**
  - $\mathbb{E}\left\{e[n]\right\}$: convexo, no diferenciable.
  - $\mathbb{E}\left\{|e[n]|^2\right\}$: convexo, diferenciable. Es matemáticamente más sencillo de tratar (se pueden calcular gradientes).
  - Otros.
- **Objetivo**: encontrar $w[0], w[1], w[2], \ldots, w[P-1]$ tales que se minimice $J_n(\mathbf{w}) = \mathbb{E}\{|e[n]|^2\} = \mathbb{E}\left\{\left|d[n] - x[n] * w[n]\right|^2\right\}$ .

# Obtención del filtro óptimo

- Imponemos la restricción de que $w[n]$ sea un filtro FIR
  - Número finito $P$ de valores distintos de 0
  - $P$ es el **orden del filtro**.
- Consideramos que $w[n]$ es casual y definido entre $n = 0$ y $n = P - 1$.
- La salida del filtro $\widehat{d}[n]$ se calculará:

$$\widehat{d}[n] = x[n] * w[n] = \sum_{k=0}^{P-1} w[k]x[n-k]$$

- Notación vectorial:

$$\mathbf{w} = \begin{bmatrix} w[0] \\ w[1] \\ \vdots \\ w[P-1] \end{bmatrix} ; \mathbf{x}[n] = \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-P+1] \end{bmatrix}$$

# Obtención del filtro óptimo

- A partir de esta notación vectorial:

$$\widehat{d}[n] = \mathbf{w}^{\mathsf{T}}\mathbf{x}[n]$$

- La señal $e[n]$ de error de reconstrucción entre $d[n]$ y $\widehat{d}[n]$ se puede escribir:

$$e[n] = d[n] - \widehat{d}[n] = d[n] - \mathbf{w}^{\mathsf{T}}\mathbf{x}[n]$$

- El vector de pesos $\mathbf{w}$ del filtro óptimo será el que minimiza el error cuadrático medio:

$$\mathbf{w} = \arg\min_{\mathbf{w}} \mathbb{E}\left\{(e[n])^2\right\} = \arg\min_{\mathbf{w}} \mathbb{E}\left\{(d[n] - \mathbf{w}^{\mathsf{T}}\mathbf{x}[n])^2\right\}$$

De ahora en adelante, asumiremos que las señales son de valor real (... y nos deshacemos del valor absoluto)

# Obtención del filtro óptimo

- Para encontrar el vector de pesos óptimo, calculamos el gradiente con respecto a **w** e igualamos a cero.

$$\nabla_{\mathbf{w}}\mathbb{E}\left\{\left(d[n]-\mathbf{w}^{\mathsf{T}}\mathbf{x}[n]\right)^2\right\}=\mathbb{E}\left\{\nabla_{\mathbf{w}}\left(d[n]-\mathbf{w}^{\mathsf{T}}\mathbf{x}[n]\right)^2\right\}$$
$$-2\mathbb{E}\left\{\mathbf{x}[n]\left(d[n]-\mathbf{w}^{\mathsf{T}}\mathbf{x}[n]\right)\right\}=-2\mathbb{E}\{\mathbf{x}[n]e[n]\}$$

- Al igualar a cero, aparece el *Principio de Ortogonalidad* (el error de reconstrucción es estadísticamente ortogonal a todas las componentes de $x[n]$).

$$\mathbb{E}\{\mathbf{x}[n]e[n]\}=\mathbf{0}\Rightarrow\mathbb{E}\{x[n-k]e[n]\}=0,\qquad k=0,1,\ldots,P-1$$

(La señal de error y la señal de entrada están incorreladas)

- Sabiendo que $e[n]=d[n]-\mathbf{w}^{\mathsf{T}}\mathbf{x}[n]$

$$\mathbb{E}\left\{\mathbf{x}[n]\left(d[n]-\mathbf{w}^{\mathsf{T}}\mathbf{x}[n]\right)\right\}=\mathbb{E}\left\{\mathbf{x}[n]\left(d[n]-\mathbf{x}[n]^{\mathsf{T}}\mathbf{w}\right)\right\}$$
$$\mathbb{E}\left\{\mathbf{x}[n]d[n]\right\}=\mathbb{E}\left\{\mathbf{x}[n]\mathbf{x}[n]^{\mathsf{T}}\right\}\mathbf{w}$$

# Obtención del filtro óptimo

Vamos a poner "nombre" a las cosas...

$$\mathbb{E}\left\{\mathbf{x}[n]d[n]\right\} = \begin{bmatrix} \mathbb{E}\{x[n]d[n]\} \\ \mathbb{E}\{x[n-1]d[n]\} \\ \vdots \\ \mathbb{E}\{x[n-P+1]d[n]\} \end{bmatrix} = \begin{bmatrix} r_{X,D}[0] \\ r_{X,D}[1] \\ \vdots \\ r_{X,D}[P-1] \end{bmatrix}$$

donde $\mathbf{x}[n]$ y $d[n]$ son dos procesos estocásticos estacionarios WSS.

$\mathbf{r}_{X,D}$ es el **vector de correlación cruzada** entre $\mathbf{x}[N]$ y $d[N]$ hasta $P$ instantes de tiempo.

# Obtención del filtro óptimo

Por otro lado, tenemos que:

$$\mathbb{E}\left\{\mathbf{x}[n]\mathbf{x}[n]^{\mathsf{T}}\right\} =$$

$$= \begin{bmatrix} \mathbb{E}\{x[n]x[n]\} & \mathbb{E}\{x[n]x[n-1]\} & \dots & \mathbb{E}\{x[n]x[n-P+1]\} \\ \mathbb{E}\{x[n-1]x[n]\} & \mathbb{E}\{x[n-1]x[n-1]\} & \dots & \mathbb{E}\{x[n-1]x[n-P+1]\} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}\{x[n-P+1]x[n]\} & \mathbb{E}\{x[n-P+1]x[n-1] & \dots & \mathbb{E}\{x[n-P+1]x[n-P+1]\} \end{bmatrix} =$$

$$= \begin{bmatrix} r_X[0] & r_X[-1] & \dots & r_X[1-P] \\ r_X[1] & r_X[0] & \dots & r_X[2-P] \\ \dots & \dots & \ddots & \dots \\ r_X[P-1] & r_X[P-2] & \dots & r_X[0] \end{bmatrix} = \mathbf{R}_X$$

Con esto:

$$\mathbf{R}_X \mathbf{w} = \mathbf{r}_{X,D} \Rightarrow \boxed{\mathbf{w} = \mathbf{R}_X^{-1}\mathbf{r}_{X,D}}$$

# Señal de error de reconstrucción

La potencia de la señal de error de reconstrucción será:

$$\mathbb{E}\left\{(e[n])^2\right\} = \mathbb{E}\left\{e[n](d[n] - \mathbf{w}^\intercal \mathbf{x}[n])\right\} = \mathbb{E}\left\{e[n]d[n]\right\} - \mathbf{w}^\intercal \mathbb{E}\left\{\mathbf{x}[n]e[n]\right\}$$

Aplicando el principio de ortogonalidad,

$$\mathbb{E}\left\{(e[n])^2\right\} = \mathbb{E}\left\{e[n]d[n]\right\} = \mathbb{E}\left\{d[n]d[n]\right\} - \mathbf{w}^\intercal \mathbb{E}\left\{d[n]\mathbf{x}[n]\right\}$$

Si tenemos en cuenta que $d[n]$ y $x[n]$ son dos procesos estocásticos WSS:

$$\mathbb{E}\left\{(e[n])^2\right\} = r_D[0] - \mathbf{w}^\intercal \mathbf{r}_{X,D}$$

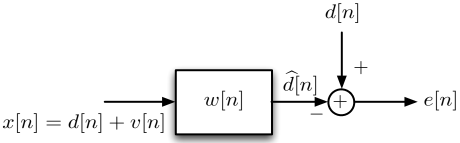
Sabiendo que $\mathbf{w} = \mathbf{R}_X^{-1}\mathbf{r}_{X,D}$, se tiene:

$$\boxed{\mathbb{E}\left\{(e[n])^2\right\} = r_D[0] - \mathbf{r}_{X,D}^\intercal \mathbf{R}_X^{-1} \mathbf{r}_{X,D}}$$

El término $r_D[0] - \mathbf{r}_{X,D}^\intercal \mathbf{R}_X^{-1} \mathbf{r}_{X,D}$ supone un límite inferior al mínimo error de reconstrucción del filtro que intente aproximar las señales generadas por el proceso $D_n$ transformando las señales generadas con el proceso estocástico $\mathbf{X}_n$.

# Planteamiento del problema

- La señal de entrada al filtro ($x[n]$) es una versión de la señal deseada ($d[n]$) corrompida con ruido aditivo ($v[n]$).




- El ruido tiene media nula e **incorrelacionado con la señal deseada** $\Rightarrow$ $\mathbb{E}\left\{d[n]v[n-k] = 0\right\}$, para valores enteros de $k$.
- La autocorrelación entre $D_n$ y $X_n$ es:

$$\begin{aligned} r_{X,D}[k] &= \mathbb{E}\left\{d[n]x[n-k]\right\} \\ &= \mathbb{E}\left\{d[n]d[n-k]\right\} + \mathbb{E}\left\{d[n]v[n-k]\right\} \\ &= r_D[k] \end{aligned}$$

# Obtención de los coeficientes

- La autocorrelación de la señal de entrada se puede expresar como:

$$\begin{aligned} r_X[k] &= \mathbb{E}\Big\{x[n]x[n-k]\Big\} = \mathbb{E}\Big\{(d[n]+v[n])(d[n-k]+v[n-k])\Big\} \\ &= \mathbb{E}\Big\{d[n]d[n-k]\Big\} + \mathbb{E}\Big\{d[n]v[n-k]\Big\} + \mathbb{E}\Big\{v[n]d[n-k]\Big\} + \\ &+ \mathbb{E}\Big\{v[n]v[n-k]\Big\} \\ &= r_D[k] + r_V[k] \end{aligned}$$

- Si $\mathbf{R}_D$ y $\mathbf{R}_V$ son las matrices de autocorrelación de los procesos $V_n$ y $D_n$, se tiene:

$$\Big[\mathbf{R}_D + \mathbf{R}_V\Big]\mathbf{w} = \mathbf{r}_D \Rightarrow \boxed{\mathbf{w} = \Big[\mathbf{R}_D + \mathbf{R}_V\Big]^{-1}\mathbf{r}_D}$$

# Obtención de los coeficientes

- El problema de predicción lineal de una señal $x[n]$ (en un escenario libre de ruido) consiste en encontrar los coeficientes $w[k]$ del modelo:

$$\widehat{x}[n+1] = \sum_{k=0}^{P-1} w[k]x[n-k]$$

- Si reescribimos el problema definiendo $d[n] = x[n+1]$, la correlación cruzada entre la señal deseada y señal de entrada se reduce a:

$$r_{D,X}[k] = \mathbb{E}\{d[n]x[n-k]\} = \mathbb{E}\{x[n+1]x[n-k]\} = r_X[k+1]$$

y el error del predictor se puede calcular como:

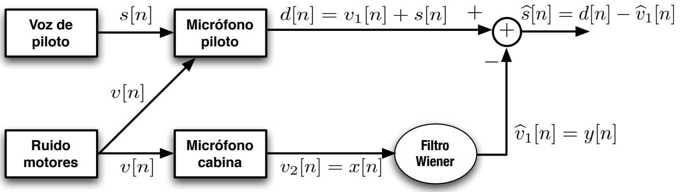
$$\mathbb{E}\{(e[n])^2\} = r_X[0] - \mathbf{w}^\intercal \mathbf{r}_X[k+1]$$

donde $\mathbf{r}_X[k+1] = \begin{bmatrix} r_X[1] & r_X[2] & \dots & r_X[P] \end{bmatrix}^\intercal$

# Objetivo

- El objetivo es cancelar el ruido de motores, maquinaria, etc. en la señal recogida en el micrófono de un piloto en una cabina de aeroplanos.
- Diferencia entre *filtrar* el ruido y *cancelar* el ruido:
  - Filtrar ruido: la señal “contaminada” pasa por un filtro donde se espera que se elimine la mayor parte del ruido.
  - Cancelar ruido: se utiliza un filtro para aprender la señal de ruido y restársela a la señal “contaminada”.

# Escenario




- $s[n]$: voz del piloto
- $v[n]$: señal de ruido de motores, viento, etc. Se transforma en:
  - $v_1[n]$: señal de ruido tras "pasar" por el micrófono del piloto
  - $v_2[n]$: señal de ruido tras "pasar" por el micrófono de la cabina
- $d[n]$: señal compuesta por la suma de la voz del piloto y el ruido en el micrófono del piloto
- $\widehat{s}[n]$: salida del cancelador de ruido. Se espera que sea una reproducción lo más exacta posible de $s[n]$

# Obtención de los coeficientes

- Diseño de un filtro FIR de Wiener que obtenga una aproximación de $v_1[n]$ a partir de $v_2[n]$ ($v_1[n]$ y $v_2[n]$ están correlados).
- Identificamos señales:
  - Señal de entrada ($x[n]$): señal de ruido registrado en el micrófono de la cabina, es decir, $v_2[n]$
  - Señal deseada ($d[n]$): señal de salida del micrófono del piloto, $d[n]$
- En este caso, se tiene que:

$$\mathbf{w} = \mathbf{R}_{V_2}^{-1}\mathbf{r}_{D,V_2}$$

  (donde $\mathbf{r}_{D,V_2} = \mathbf{r}_{V_1,V_2}$)
- Si la voz del piloto está incorrelada con el ruido de los motores, se tiene:

$$\begin{aligned} r_{D,V_2}[k] &= \mathbb{E}\{d[n]v_2[n-k]\} = \mathbb{E}\{v_1[n]v_2[n-k]\} + \mathbb{E}\{s[n]v_2[n-k]\} \\ &= \mathbb{E}\Big\{v_1[n]v_2[n-k]\Big\} = r_{V_1,V_2}[k] \end{aligned}$$

- Como $s[n]$ y $v_2[n]$ están incorreladas, el filtro sólo aprende la parte de $d[n]$ correlacionada con $v_2[n]$, y ésta es precisamente, $v_1[n]$.

# Filtrado adaptativo

Lorena Álvarez Pérez

Universidad Carlos III de Madrid

Curso académico 2023/2024

# Índice

# Hasta ahora...

- Habíamos asumido que la señal de entrada al filtro ($x[n]$) y la señal deseada ($d[n]$) eran **procesos estocásticos estacionarios en sentido amplio** (WSS).
  - El filtro se podía estudiar como un sistema LTI
- ¿Cómo se diseña el filtro cuando las señales no son estacionarias?
  - Si dividimos la señal en tramos que puedan "considerarse" estacionarios y calcular los coeficientes del filtro para cada trama de forma independiente...
    - Si los procesos estocásticos son de variación muy rápida, los intervalos donde asumimos estacionariedad serán demasiados cortos $\Rightarrow$ estimaciones poco eficientes de los parámetros de estudio.
    - Tratar de modo independiente los intervalos, dificulta el análisis de la señal de modo completo.
- **¡Solución! Filtrado adaptativo**

# Punto de partida

- Filtro FIR óptimo cuando la señal de entrada ($x[n]$) y señal deseada ($d[n]$) son procesos estocásticos estacionarios WSS:

$$\mathbf{w} = \mathbf{R}_X^{-1}\mathbf{r}_{X,D}$$

- Si los procesos varían con el tiempo, la matriz de autocorrelación de la señal de entrada $\mathbf{R}_X$, y el vector de correlación cruzada entre la señal de entrada y la señal deseada $\mathbf{r}_{X,D}$, serán variantes en el tiempo.

$$\begin{cases} \mathbf{R}_X[n] \\ \mathbf{r}_{X,D}[n] \end{cases} \Rightarrow \mathbf{w}[n] \text{ (el filtro también es función del tiempo)}$$

¿Un **filtro de Wiener** para **cada instante** $n$?

**¡Poca aplicación práctica!** Si en escenarios estacionarios era difícil conseguir una caracterización completa de las funciones de autocorrelación, para cada instante se torna más inviable aún.

# Obtención coeficientes - diseño incremental

- Partimos de que en el instante $n$, tendremos un vector de pesos $\mathbf{w}[n]$.
- El vector de pesos correspondiente al instante $n+1$, se calcula como:

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \Delta\mathbf{w}[n]$$

- **Objetivo**: diseñar el **algoritmo para construir** $\Delta\mathbf{w}[n]$
- Propiedades "deseables" de dicho algoritmo:
  - En un problema estacionario, el filtro adaptativo debería producir una secuencia de actualizaciones de $\mathbf{w}[n]$ que converja al filtro óptimo:

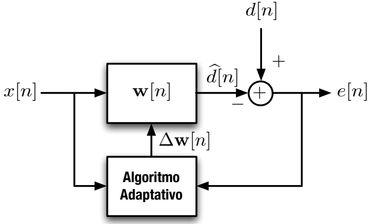
$$\lim_{n\to\infty} \mathbf{w}[n] = \mathbf{R}_X^{-1}\mathbf{r}_{X,D}$$

    (se podría usar el filtro adaptativo en escenarios estacionarios, ahorrando la inversión de la matriz de autocorrelación).
  - No deberían tener que calcularse $\mathbf{R}_X[n]$ ni $\mathbf{r}_{X,D}[n]$
  - Con señales no estacionarias, el filtro debería ser capaz de adaptarse a los cambios en los estadísticos de las señales y realizar un seguimiento de la solución a medida que ésta evoluciona con el tiempo.

# Esquema de un filtro adaptativo

El funcionamiento del filtro se divide en tres fases:

- Predicción: Cuando llega el vector $\mathbf{x}[n]$, el filtro realiza una predicción de la señal deseada $\widehat{d}[n] = \mathbf{w}[n]^\intercal \mathbf{x}[n]$
- Cálculo de error:
  Se le comunica al filtro el valor real de $d[n]$ y se calcula el error cometido por el filtro en la muestra: $e[n] = d[n] - \widehat{d}[n]$.
- Actualización:
  A partir de $e[n]$, se obtiene la actualización $\Delta\mathbf{w}[n]$ y, a continuación, los coeficientes $\mathbf{w}[n+1]$ que se usarán para procesar $x[n+1]$ que representa la observación que llega en el instante $n+1$.



# Introducción

- Método iterativo de actualización del vector de pesos que considera la potencia del error de reconstrucción como una función de los pesos del filtro.
- Las actualizaciones del vector de pesos han de ir en la dirección de encontrar el mínimo de esta superficie de error.
- Sea $\xi[n] = \mathbb{E}\{e^2[n]\}$ (donde $e[n]^2 = f(\mathbf{w})$):
  - La dirección de máxima variación de $\xi[n]$ con $\mathbf{w}[n]$ viene dada por $\nabla_{\mathbf{w}}\xi[n]$.
  - Como nos interesa minimizar $\xi[n]$, la mejor actualización (para vector de pesos) es en la dirección opuesta al gradiente, $-\nabla_{\mathbf{w}}\xi[n]$.
  - Regla de actualización:

$$\mathbf{w}[n+1] = \mathbf{w}[n] - \frac{\mu}{2}\nabla_{\mathbf{w}}\xi[n]$$

(desde la posición actual, $\mathbf{w}[n]$, se da un paso de magnitud $\frac{\mu}{2}$ en la dirección dada por $-\nabla_{\mathbf{w}}\xi[n]$).

A la constante $\mu$ se le conoce como paso de actualización.

# Cálculo del gradiente

- El cálculo del gradiente $\nabla_{\mathbf{w}}\xi[n]$ se realiza a partir de la definición del error de reconstrucción: $e[n] = d[n] - \widehat{d}[n] = d[n] - \mathbf{w}^{\intercal}\mathbf{x}[n]$.

$$\begin{aligned}
\nabla_{\mathbf{w}}\xi[n] = \nabla_{\mathbf{w}}\mathbb{E}\{e[n]^2\} &= \mathbb{E}\left\{\nabla_{\mathbf{w}} e[n]^2\right\} = \mathbb{E}\left\{\nabla_{\mathbf{w}}\left(d[n] - \mathbf{w}^{\intercal}\mathbf{x}[n]\right)^2\right\} \\
&= -2\mathbb{E}\left\{\left(d[n] - \mathbf{w}^{\intercal}\mathbf{x}[n]\right)\mathbf{x}[n]\right\} \\
&= -2\mathbb{E}\left\{e[n]\mathbf{x}[n]\right\}
\end{aligned}$$

  (Función de coste derivable $J(\mathbf{w})$, tal que para la solución óptima $\mathbf{w}_0$, se cumple que: $J(\mathbf{w}_0) < J(\mathbf{w})$).

- Por tanto, la regla de actualización de descenso por máxima pendiente será:

$$\boxed{\mathbf{w}[n+1] = \mathbf{w}[n] + \mu\mathbb{E}\left\{e[n]\mathbf{x}[n]\right\}}$$

# Aplicacion al filtro de Wiener

- La regla de descenso por máxima pendiente puede utilizarse en un caso de filtrado estacionario donde no convenga invertir las matrices de autocorrelación.
- Las actualizaciones vendrán dadas por:

$$\begin{aligned}\mathbb{E}\{e[n]\mathbf{x}[n]\} &= \mathbb{E}\{(d[n] - \mathbf{w}[n]^\intercal \mathbf{x}[n])\mathbf{x}[n]\} \\ &= \mathbb{E}\{(d[n]\mathbf{x}[n] - \mathbf{x}[n]^\intercal \mathbf{x}[n]\mathbf{w}[n])\} \\ &= \mathbb{E}\{d[n]\mathbf{x}[n]\} - \mathbb{E}\{\mathbf{x}[n]^\intercal \mathbf{x}[n]\}\mathbf{w}[n] \\ &= \mathbf{r}_{D,X} - \mathbf{R}_X \mathbf{w}[n]\end{aligned}$$

- El algoritmo de descenso por máxima pendiente será:

$$\boxed{\mathbf{w}[n+1] = \mathbf{w}[n] + \mu\big(\mathbf{r}_{D,X} - \mathbf{R}_X \mathbf{w}[n]\big)}$$

Cuando el algoritmo encuentre un vector de pesos $\mathbf{w}[n]$ que coincida con el filtro de Wiener, el algoritmo se detiene, ya el que el filtro de Wiener verifica que:

$$\mathbf{w}[n] = \mathbf{R}_X^{-1}\mathbf{r}_{D,X}$$

y el término de actualización sería nulo.

# Paso de actualización

- Puede demostrarse que en escenarios estacionarios, el filtro adaptativo actualizado mediante el algoritmo de descenso por máxima pendiente converge al filtro óptimo si el paso de adaptación cumple:

$$0 < \mu < \frac{2}{\lambda_{max}}$$

donde $\lambda_{max}$ es el autovalor más grande de $\mathbf{R}_X$.

# Consideraciones

- El algoritmo de descenso por máxima pendiente es de **poca utilidad práctica**.
- Es necesario calcular $\mathbb{E}\left\{e[n]\mathbf{x}[n]\right\}$ y en la mayor parte de las aplicaciones, estos estadísticos se desconocen y deben estimarse a partir de los datos:
  - En casos estacionarios, las estimaciones de la autocorrelación de $X_n$ y de la correlación cruzada entre $X_n$ y $D_n$ a partir de muestras, pueden ser más o menos, robustas.
  - En casos no estacionarios, es imposible conseguir estimaciones razonables de estos parámetros.

# Introducción

- Una estrategia razonable puede ser sustituir dicha esperanza por una media muestral, calculada a partir de $L$ observaciones previas de la señal de entrada, es decir, $x[n]$, $x[n-1]$, $x[n-2]$, ..., $x[n-L+1]$, para obtener así estimaciones muestrales de dichos estadísticos:

$$\mathbb{E}\{e[n]\mathbf{x}[n]\} \approx \frac{1}{L}\sum_{l=0}^{L-1} e[n-l]\mathbf{x}[n-l]$$

- Nótese que hay una diferencia significativa entre $L$ (número de instantes de tiempo que se utiliza para estimar los estadísticos) y $P$ (número de coeficientes del filtro). En general, $L \gg P$.
  - $P$ hace referencia al número de valores significativos de la función de autocorrelación para procesos estocásticos.
  - $L$ denota el número de instantes de tiempo en los que se puede suponer que los estadísticos de estos procesos pueden considerarse estacionarios.

# Introducción

- Caso particular: $L = 1$ (**algoritmo LMS**, del inglés *Least-Mean-Square algorithm*)[1].
  - Se utiliza el vector de entrada $\mathbf{x}[n]$ y error de reconstrucción actual $e[n]$ para realizar una estimación de rango uno de $\mathbb{E}\{E[n]X[n]\}$.
  - Ecuación de actualización:

$$\mathbf{w}[n+1] = \mathbf{w}[n] - \mu e[n]\mathbf{x}[n]$$

¡Observe que esta regla de actualización involucra vectores!

$$\begin{bmatrix} w_0[n+1] \\ w_1[n+1] \\ \vdots \\ w_{M-1}[n+1] \end{bmatrix} = \begin{bmatrix} w_0[n] \\ w_1[n] \\ \vdots \\ w_{M-1}[n] \end{bmatrix} + \begin{bmatrix} \mu e[n]x[n] \\ \mu e[n]x[n-1] \\ \vdots \\ \mu e[n]x[n-P+1] \end{bmatrix}$$

[1] Fue inventado en 1960 por Bernard Widrow (profesor de la Universidad de Standford) y su primer estudiante de doctorado (Ted Hoff).

# Parámetro de actualización

- La convergencia del algoritmo LMS al filtro óptimo, en situaciones estacionarias, depende del parámetro $\mu$ (determina la magnitud de cada paso de actualización en la dirección opuesta al gradiente).
- El LMS converge *en la media al cuadrado* a la solución de Wiener si:
  - Se da independencia estadística entre el vector de pesos $\mathbf{w}[n]$ y los datos $\mathbf{x}[n]$
  - El paso de actualización ($\mu$) verifica:

$$0 < \mu < \frac{2}{P\mathbb{E}\left\{x[n]^2\right\}}$$

donde $P$ es el orden del filtro. En aplicaciones prácticas, $\mathbb{E}\left\{x[n]^2\right\}$ puede estimarse a partir de $N$ valores de la señal $x[n]$ mediante:

$$\mathbb{E}\left\{x[n]^2\right\} \approx \frac{1}{N}\sum_{k=0}^{N-1} x[k]^2$$