



9 Método de Cuadrados Mínimos

- Se discutirá un procedimiento para resolver el problema de filtrado lineal, dependiente del modelo, denominado **método de cuadrados mínimos**, sin invocar las suposiciones sobre la estadística de la entrada aplicada al filtro.
- El método de cuadrados mínimos puede verse como una alternativa a la teoría del filtro de Wiener. Basicamente, los filtros de Wiener se obtienen de **promedios estadísticos**, con el resultado que un filtro (óptimo en sentido probabilístico) se obtiene para todas las realizaciones del entorno operacional, suponiendo ESA.
- Por otro lado, el método de cuadrados mínimos es una solución **determinística**. Específicamente, involucra el uso de **promedios temporales**, con el resultado que el filtro depende del número de muestras utilizadas en el cálculo.
- Se extenderá también el uso del método de cuadrados mínimos, desarrollando un algoritmo recursivo para el diseño de filtros adaptivos transversales. El algoritmo resultante se denomina **cuadrados mínimos recursivo (RLS)**.
- El algoritmo RLS puede verse como un caso especial del filtro de Kalman.
- En el desarrollo del algoritmo RLS se analizará, revisando algunas relaciones básicas asociadas al método de cuadrados mínimos, el **lema de inversión de matrices**. Una característica importante del algoritmo RLS es que utiliza información que posee memoria desde las condiciones iniciales.

9.1 Problema de cuadrados mínimos

- Consideremos un fenómeno físico caracterizado por dos conjuntos de variables, $d(i)$ y $u(i)$. Esta relación funcional, **como hipótesis**, se supone lineal.
- En particular, la respuesta $d(i)$ es modelada como

$$d(i) = \sum_{k=0}^{M-1} \sum_{k=0}^{M-1} w_{ok}^* u(i - k) + e_o(i)$$

donde w_{ok} son los **parámetros desconocidos** del modelo, y $e_o(i)$ representa el **error de medición** asociado a la naturaleza estadística del fenómeno. Este modelo se denomina **modelo de regresión lineal múltiple**.

- El **error de medición** $e_o(i)$ es una **variable aleatoria no observable** que se introduce al modelo para tener en cuenta imprecisiones. Es usual suponer que el proceso de error de medición $e_o(i)$ es blanco con media cero y varianza σ^2 . O sea,

$$E[e_o(i)] = 0, \quad \text{para todo } i \quad \text{y} \quad E[e_o(i)e_o^*(k)] = \begin{cases} \sigma^2, & i = k \\ 0, & i \neq k \end{cases}$$

lo que implica que

$$E[d(i)] = \sum_{k=0}^{M-1} \sum_{k=0}^{M-1} w_{ok}^* u(i - k)$$

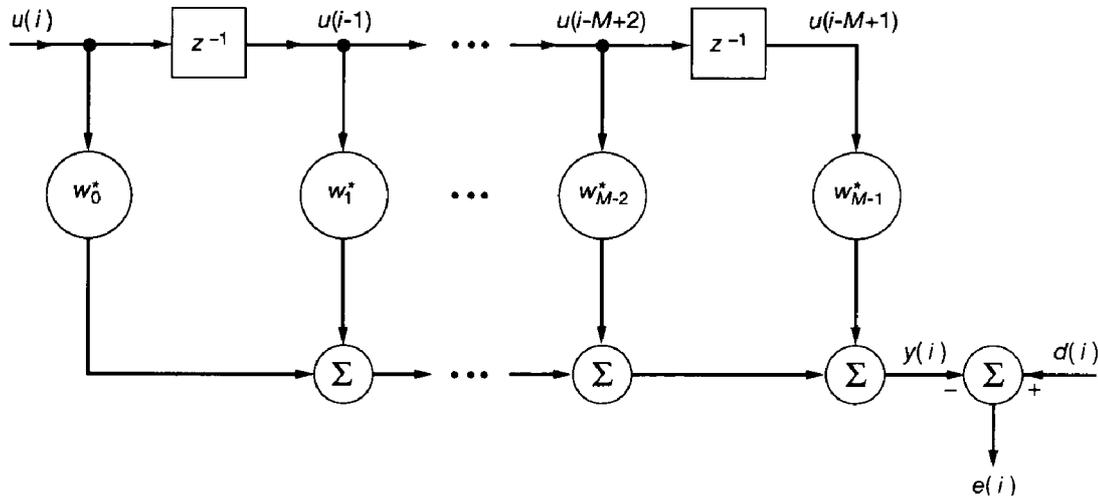


Figura 36: Modelo de un filtro transversal lineal.

- Con el filtro lineal transversal de la figura obtenemos un **error residual** $e(i)$ dado por

$$e(i) = d(i) - y(i) = d(i) - \sum_{k=0}^{M-1} w_k^* u(i-k)$$

- En el método de cuadrados mínimos elegimos los coeficientes del filtro transversal w_k para minimizar una función costo que consiste en la **suma de los errores al cuadrado**:

$$\mathcal{E}(w_0, \dots, w_{M-1}) = \sum_{i=i_1}^{i_2} |e(i)|^2 \quad (117)$$

donde i_1 e i_2 definen los índices límite entre los cuales ocurre la minimización del error. Esta suma puede verse también como **energía de error**.

- Los valores asignados a esos límites dependen del tipo de **ventana de datos** utilizada.

9.2 Ventanas de datos

- Si M es el número de coeficientes utilizados en el filtro transversal modelado en la figura 36, la ventana rectangular construida de los datos de entrada $u(1), u(2), \dots, u(N)$ puede tomar diferentes formas, dependiendo de los valores asignados a los límites i_1 y i_2 en (117).
- En particular, podemos distinguir cuatro métodos diferentes de **ventanear** los datos de entrada
 1. **Método covarianza**, el cual no hace ninguna suposición sobre los datos fuera del intervalo $[1, N]$. De esta forma, definiendo los límites de interés como $i_1 = M$ y $i_2 = N$, los datos de entrada pueden ordenarse en la siguiente forma matricial

$$\begin{bmatrix} u(M) & u(M+1) & \cdots & u(N) \\ u(M-1) & u(M) & \cdots & u(N-1) \\ & & \vdots & \\ u(1) & u(2) & \cdots & u(N-M+1) \end{bmatrix}$$

2. **Método autocorrelación**, el cual hace uso de la suposición que los datos previos al instante $i = 1$ y los datos después de $i = N$ son cero. Así, usando $i_1 = 1$ y $i_2 = N + M - 1$, la matriz de datos de entrada toma la forma

$$\begin{bmatrix} u(1) & u(2) & \cdots & u(M) & u(M+1) & \cdots & u(N) & 0 & \cdots & 0 \\ 0 & u(1) & \cdots & u(M-1) & u(M) & \cdots & u(N-1) & u(N) & \cdots & 0 \\ & & \vdots & & & & & & \vdots & \\ 0 & 0 & \cdots & u(1) & u(2) & \cdots & u(N-M+1) & u(N-M) & \cdots & u(N) \end{bmatrix}$$

3. **Método de pre-ventana**, el cual hace uso de la suposición que los datos previos a $i = 1$ son cero, pero no hace suposiciones sobre los datos después de $i = N$. De esta forma usando $i_1 = 1$ y $i_2 = N$, la matriz de entradas toma la forma

$$\begin{bmatrix} u(1) & u(2) & \cdots & u(M) & u(M+1) & \cdots & u(N) \\ 0 & u(1) & \cdots & u(M-1) & u(M) & \cdots & u(N-1) \\ & & \vdots & & & & \\ 0 & 0 & \cdots & u(1) & u(2) & \cdots & u(N-M+1) \end{bmatrix}$$

4. **Método de post-ventana**, el cual no hace suposiciones sobre los datos previos a $i_1 = 1$, pero supone que los datos posteriores a $i = N$ son cero. De esta forma, usando $i_1 = M$ e $i_2 = N + M - 1$, la matriz de entradas toma la forma

$$\begin{bmatrix} u(M) & u(M+1) & \cdots & u(N) & 0 & \cdots & 0 \\ u(M-1) & u(M) & \cdots & u(N-1) & u(N) & \cdots & 0 \\ & & & & & \vdots & \\ u(1) & u(2) & \cdots & u(N-M+1) & u(N-M) & \cdots & u(N) \end{bmatrix}$$

- Los métodos covarianza y autocorrelación son usualmente utilizados en la literatura de procesamiento de voz. La denominación proviene de la forma en que es posible interpretar los **parámetros conocidos** contenidos en el sistema de ecuaciones que resultan de minimizar la función costo (117).
- En el resto de esta discusión trabajaremos con el método covarianza o el método de pre-ventana.

9.3 Revisión del principio de ortogonalidad

- Discutiremos el principio de ortogonalidad en forma similar a la que permitió obtener la base para los filtros de Wiener. El desarrollo de esta teoría se realizará en base al método covarianza.
- La función costo para el método covarianza se define como

$$\mathcal{E}(w_0, \dots, w_{M-1}) = \sum_{i=M}^N e(i)e^*(i) = \sum_{i=M}^N |e(i)|^2 \quad (118)$$

- Eligiendo los índices i de esta forma es posible garantizar que todas las entradas del filtro transversal no tienen valores cero. Teniendo en cuenta que $w_k = a_k + jb_k$, $k = 0, 1, \dots, M - 1$ el gradiente respecto al error $\nabla \mathcal{E}$, teniendo en cuenta que $e(i) = d(i) - \sum_{k=0}^{M-1} (a_k - jb_k)u(i-k)$, será

$$\begin{aligned} \nabla_k \mathcal{E} &= - \sum_{i=M}^N \left[e(i) \frac{\partial e^*(i)}{\partial a_k} + e^*(i) \frac{\partial e(i)}{\partial a_k} + je(i) \frac{\partial e^*(i)}{\partial b_k} + je(i) \frac{\partial e(i)}{\partial b_k} \right] \\ &= -2 \sum_{i=M}^N u(i-k)e^*(i) = 0, \quad k = 0, 1, \dots, M-1 \end{aligned}$$

- Si $e_{min}(i)$ es el valor de $e(i)$ que minimiza $\mathcal{E}(w_0, \dots, w_{M-1})$, entonces

$$\sum_{i=M}^N u(i-k)e_{min}^*(i) = 0, \quad k = 0, 1, \dots, M-1 \quad (119)$$

que es la descripción matemática de la versión temporal del **principio de ortogonalidad**:

La secuencia de error mínimo $e_{min}(i)$ es ortogonal a la secuencia $u(i-k)$ aplicada al tap k de un filtro transversal de longitud M para $k = 0, 1, \dots, M-1$, cuando el filtro está operando en su condición de cuadrados mínimos.

Corolario Suponiendo $\hat{w}_0, \dots, \hat{w}_{M-1}$ los valores de los coeficientes que optimizan la operación del filtro transversal en su condición de cuadrados mínimos, la salida del filtro $y_{min}(i)$ será

$$y_{min}(i) = \sum_{k=0}^{M-1} \hat{w}_k^* u(i - k)$$

Esta salida provee una **estimación de cuadrados mínimos** de la respuesta deseada $d(i)$. Si $\hat{d}(i|U_i)$ es la estimación de cuadrados mínimos de la respuesta deseada $d(i)$, entonces

$$\hat{d}(i|U_i) = \sum_{k=0}^{M-1} \hat{w}_k^* u(i - k) = y_{min}(i)$$

Si ahora multiplicamos ambos lados de (119) por \hat{w}_k^* y sumamos sobre el intervalo $[0, M - 1]$, entonces

$$\sum_{i=M}^N \hat{d}(i|U_i) e_{min}^*(i) = 0$$

esta ecuación es **corolario del principio de ortogonalidad**:

Cuando un filtro transversal opera en su condición de cuadrados mínimos, la estimación de cuadrados mínimos de la respuesta deseada, producida por la salida del filtro y representada por $\hat{d}(i|U_i)$ y el error de estimación mínimo $e_{min}(i)$ son ortogonales sobre el intervalo de índices de i .

9.4 Suma mínima de los errores al cuadrado

- Para obtener el valor mínimo de la función costo es posible escribir

$$\underbrace{d(i)}_{\substack{\text{respuesta} \\ \text{deseada}}} = \underbrace{\hat{d}(i|U_n)}_{\substack{\text{estimación} \\ \text{de la} \\ \text{respuesta} \\ \text{deseada}}} + \underbrace{e_{min}(i)}_{\substack{\text{error de} \\ \text{estimación}}}$$

- De donde, evaluando la energía de la secuencia $d(i)$ para valores de i en $[M, N]$ y usando el principio de ortogonalidad se tiene que

$$\begin{aligned}
 \mathcal{E}_d &= \mathcal{E}_{est} + \mathcal{E}_{min} \\
 \text{ó} \\
 \sum_{i=M}^N |d(i)|^2 &= \sum_{i=M}^N |\hat{d}(i|U_i)|^2 + \sum_{i=M}^N |e_{min}(i)|^2
 \end{aligned}$$

de donde

$$\mathcal{E}_{min} = \mathcal{E}_d - \mathcal{E}_{est} \tag{120}$$

- Entonces, dada la especificación de la respuesta deseada $d(i)$ es posible evaluar su energía y luego evaluar \mathcal{E}_d . Dado que \mathcal{E}_{min} no es negativo, \mathcal{E}_{est} nunca puede ser mayor que \mathcal{E}_d .

9.5 Filtros lineales de cuadrados mínimos

- Una forma equivalente de describir la condición de cuadrados mínimos del filtro está representada por las **ecuaciones normales**.
- El principio de ortogonalidad se formula en términos del conjunto de entradas y $e_{min}(i)$. Como

$$e_{min}(i) = d(i) - \sum_{n=0}^{M-1} \hat{w}_n^* u(i-n)$$

entonces sustituyendo y reordenando se tiene un sistema de M ecuaciones simultáneas

$$\sum_{n=0}^{M-1} \hat{w}_n \sum_{i=M}^N u(i-k)u^*(i-n) = \sum_{i=M}^N u(i-k)d^*(i), \quad k = 0, \dots, M-1 \quad (121)$$

- Con las siguientes interpretaciones

1. función autocorrelación de promedio temporal

$$\phi(n, k) = \sum_{i=M}^N u(i-k)u^*(i-n), \quad 0 \leq (n, k) \leq M-1$$

2. correlación cruzada

$$z(-k) = \sum_{i=M}^N u(i-k)d^*(i), \quad 0 \leq k \leq M-1$$

- De esta forma

$$\sum_{n=0}^{M-1} \hat{w}_n \phi(n, k) = z(-k), \quad k = 0, 1, \dots, M-1$$

que representa el **sistema expandido de las ecuaciones normales** para el filtro lineal de cuadrados mínimos.

Formulación matricial de las ecuaciones normales

- El sistema de ecuaciones puede reescribirse en forma matricial. Para ello es necesario introducir algunas definiciones. En particular

1. La matriz de **correlación de promedios temporales** de $u(i), u(i - 1), \dots, u(i - M + 1)$

$$\Phi = \begin{bmatrix} \phi(0, 0) & \phi(1, 0) & \cdots & \phi(M - 1, 0) \\ \phi(0, 1) & \phi(1, 1) & \cdots & \phi(M - 1, 1) \\ & & \vdots & \\ \phi(0, M - 1) & \phi(1, M - 1) & \cdots & \phi(M - 1, M - 1) \end{bmatrix}$$

2. El vector de **correlación cruzada de promedios temporales** entre las entradas $u(i), u(i - 1), \dots, u(i - M + 1)$ y la respuesta deseada $d(i)$

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

3. El vector de coeficientes del filtro de cuadrados mínimos

$$\hat{\mathbf{w}} = [\hat{w}_0, \hat{w}_1, \dots, \hat{w}_{M-1}]^T$$

- De esta forma las ecuaciones normales pueden escribirse como

$$\Phi \hat{\mathbf{w}} = \mathbf{z}$$

- Suponiendo que Φ no sea singular, el vector de coeficientes para el filtro de cuadrados mínimos lineal será

$$\hat{\mathbf{w}} = \Phi^{-1} \mathbf{z} \quad (122)$$

Suma mínima de los errores al cuadrado

- Para completar la evaluación de \mathcal{E}_{min} , a partir de la (120), es posible escribir para \mathcal{E}_{est} ,

$$\begin{aligned}
 \mathcal{E}_{est} &= \sum_{i=M}^N |\hat{d}(i|U_i)|^2 \\
 &= \sum_{i=M}^N \sum_{n=0}^{M-1} \sum_{k=0}^{M-1} \hat{w}_n \hat{w}_k^* u(i-k) u^*(i-n) \\
 &= \sum_{n=0}^{M-1} \sum_{k=0}^{M-1} \hat{w}_n \hat{w}_k^* \sum_{i=M}^N u(i-k) u^*(i-n) \\
 &= \sum_{n=0}^{M-1} \sum_{k=0}^{M-1} \hat{w}_n \phi(n, k) \hat{w}_k^* \\
 &= \hat{\mathbf{w}}^H \mathbf{\Phi} \hat{\mathbf{w}}
 \end{aligned}$$

donde $\hat{\mathbf{w}}$ es el vector de coeficientes de cuadrados mínimos y $\mathbf{\Phi}$ es la matriz de correlación de promedios temporales.

- Es posible simplificar esta ecuación si se tiene en cuenta las ecuaciones normales, o sea

$$\mathcal{E}_{est} = \hat{\mathbf{w}}^H \mathbf{z} = \mathbf{z}^H \hat{\mathbf{w}}$$

de forma que sustituyendo esta ecuación en la (120) se tiene que

$$\mathcal{E}_{min} = \mathcal{E}_d - \mathbf{z}^H \hat{\mathbf{w}} = \mathcal{E}_d - \mathbf{z}^H \mathbf{\Phi}^{-1} \mathbf{z}$$

9.6 Matriz Correlación de promedios temporales Φ

- La matriz de correlación Φ , tiene como elementos $\phi(n, k)$, tal que el índice k indica la fila y n se refiere a la columna. Con el vector $\mathbf{u}(i)$ de $(M \times 1)$ definido por

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

la matriz correlación puede escribirse como

$$\Phi = \sum_{i=M}^N \mathbf{u}(i)\mathbf{u}^H(i)$$

- Algunas propiedades de esta matriz son las siguientes
 1. La matriz correlación Φ es Hermitiana, o sea $\Phi^H = \Phi$.
 2. La matriz correlación Φ es no negativa definida, o sea, para un vector x de $M \times 1$: $x^H \Phi x \geq 0$.
 3. Los autovalores de la matriz correlación Φ son todos reales y no negativos.
 4. La matriz correlación es el producto de dos matrices rectangulares Toeplitz que son la transpuesta Hermitiana una de otra.
- La matriz correlación no es en general Toeplitz. Sin embargo, está compuesta por el producto de dos matrices Toeplitz, ya que

$$\Phi = [\mathbf{u}(M), \mathbf{u}(M + 1), \dots, \mathbf{u}(N)] \begin{bmatrix} \mathbf{u}^H(M) \\ \mathbf{u}^H(M - 1) \\ \dots \\ \mathbf{u}^H(N) \end{bmatrix}$$



- Por conveniencia de notación introducimos la **matriz de datos** \mathbf{A} ($N - M + 1 \times M$), rectangular Toeplitz, cuya transpuesta Hermitiana es

$$\begin{aligned} \mathbf{A}^H &= [\mathbf{u}(M), \mathbf{u}(M + 1), \dots, \mathbf{u}(N)] \\ &= \begin{bmatrix} u(M) & u(M + 1) & \dots & u(N) \\ u(M - 1) & u(M) & \dots & u(N - 1) \\ & & \vdots & \\ u(1) & u(2) & \dots & u(N - M + 1) \end{bmatrix} \end{aligned}$$

tal que

$$\Phi = \mathbf{A}^H \mathbf{A} \tag{123}$$

9.7 Ecuaciones normales y matriz de datos

- El sistema de ecuaciones normales puede reformularse en términos de las matrices de datos usando la ecuación anterior para Φ y z .
- Para ello, introducimos un vector de datos **deseado** \mathbf{d} , consistente en la **respuesta deseada** $d(i)$ para los valores de i en $[M, N]$, o sea

$$\mathbf{d}^H = [d(M), d(M + 1), \dots, d(N)]$$

con estas definiciones se tiene que

$$\mathbf{z} = \mathbf{A}^H \mathbf{d}$$

además las ecuaciones normales serán

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

entonces la solución de cuadrados mínimos toma la forma

$$\hat{\mathbf{w}} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{d} \quad (124)$$

de forma que

$$\mathcal{E}_{min} = \mathbf{d}^H \mathbf{d} - \mathbf{d}^H \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{d}$$

- A pesar que esta ecuación se muestra algo complicada, su característica interesante es que está expresada en términos de la matriz de datos \mathbf{A} y el vector de datos deseado \mathbf{d} .

Operador proyección

- La ecuación (124) define al vector de cuadrados mínimos $\hat{\mathbf{w}}$ en términos de la matriz de datos \mathbf{A} y el vector deseado \mathbf{d} .
- La estimación de cuadrados mínimos de \mathbf{d} está dada por

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

- De acuerdo a esto, es posible interpretar el producto $\mathbf{A} \hat{\mathbf{w}} = \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{d}$ como un **operador proyección** sobre el espacio lineal barrido por las columnas de la matriz de datos \mathbf{A} , el cual es el mismo espacio U_i , mencionado anteriormente, para $i = N$. Si introducimos la notación \mathbf{P} para este operador podemos escribir

$$\mathbf{P} = \mathbf{A} \hat{\mathbf{w}} = \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$$

- La diferencia de matrices

$$\mathbf{I} - \mathbf{A} \hat{\mathbf{w}} = \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H = \mathbf{I} - \mathbf{P}$$

es el **operador complemento ortogonal**. Notar que el operador proyección y su complemento ortogonal están únivocamente determinados por la matriz de datos \mathbf{A} .

- El operador proyección \mathbf{P} , aplicado sobre el vector dato deseado \mathbf{d} , produce la estimación correspondiente $\hat{\mathbf{d}}$. $\mathbf{I} - \mathbf{P}$ aplicado al vector dato deseado \mathbf{d} produce el vector de error de estimación $\mathbf{e}_{min} = \mathbf{d} - \hat{\mathbf{d}}$.

Ejemplo Consideremos el ejemplo de un filtro de cuadrados mínimos con dos taps ($M = 2$) y una entrada real de cuatro muestras ($N = 4$), o sea $N - M + 1 = 3$. La matriz de datos \mathbf{A} y el vector dato deseado \mathbf{d} son

$$\mathbf{A} = \begin{bmatrix} u(2) & u(1) \\ u(3) & u(2) \\ u(4) & u(3) \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 2 \\ -1 & 1 \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} d(2) \\ d(3) \\ d(4) \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1/34 \end{bmatrix}$$

El objetivo de este ejemplo es evaluar el operador proyección y su complemento ortogonal y utilizarlos para ilustrar el principio de ortogonalidad.

El operador proyección será

$$\mathbf{P} = \mathbf{A} \hat{\mathbf{w}} = \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H = \frac{1}{35} \begin{bmatrix} 26 & 15 & -2 \\ 15 & 10 & 5 \\ -3 & 5 & 34 \end{bmatrix}$$

El operador complemento ortogonal es

$$\mathbf{I} - \mathbf{P} = \frac{1}{35} \begin{bmatrix} 9 & -15 & 3 \\ -15 & 25 & -5 \\ 3 & -5 & 1 \end{bmatrix}$$

de forma que

$$\hat{\mathbf{d}} = \mathbf{P} \mathbf{d} = \begin{bmatrix} 1.91 \\ 1.15 \\ 0 \end{bmatrix} \quad \mathbf{e}_{min} = (\mathbf{I} - \mathbf{P}) \mathbf{d} = \begin{bmatrix} 0.09 \\ -0.15 \\ 0.03 \end{bmatrix}$$

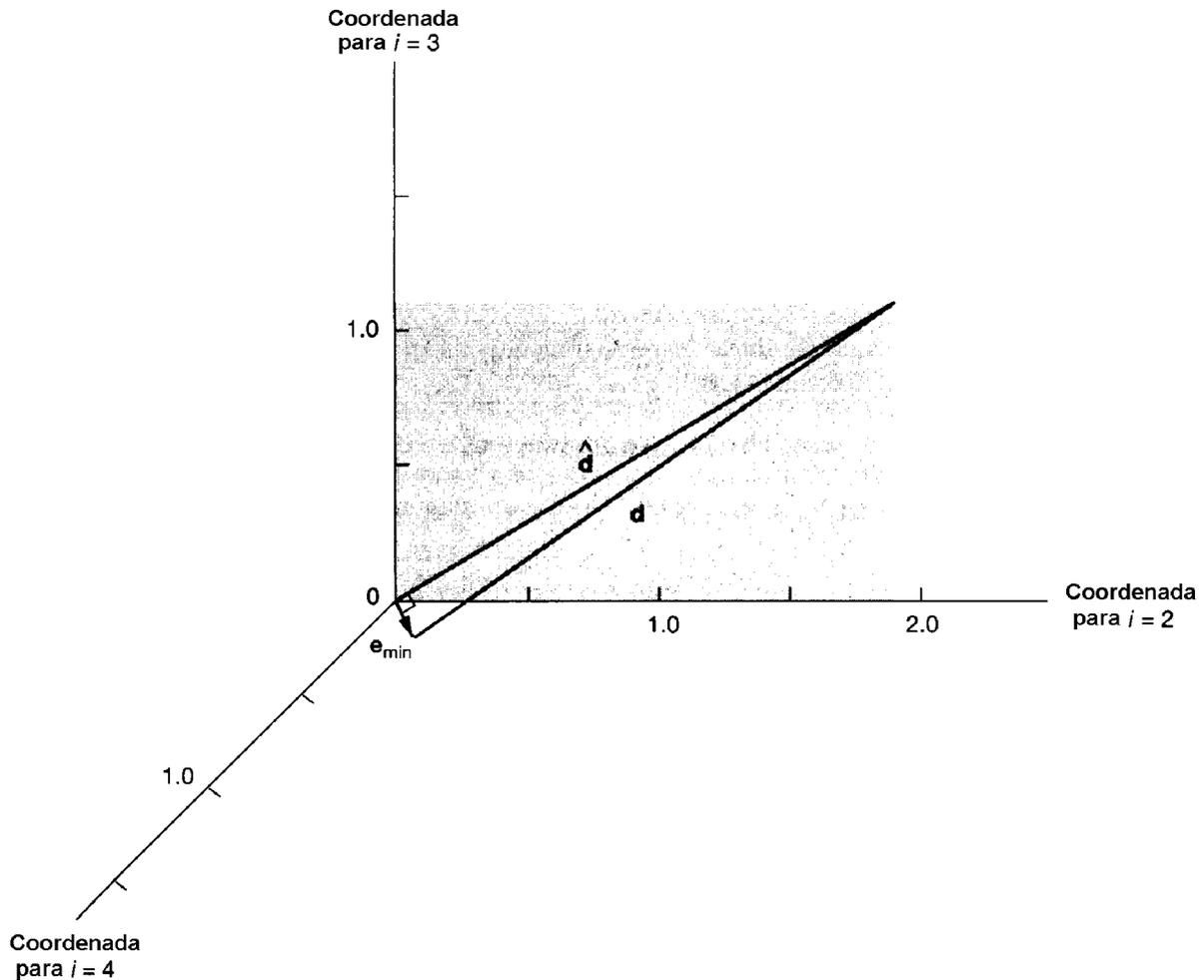


Figura 37: Representación geométrica tridimensional de \mathbf{d} , $\hat{\mathbf{d}}$ y \mathbf{e}_{min} .

- La figura muestra la representación geométrica de $\hat{\mathbf{d}}$ y \mathbf{e}_{min} , donde claramente esos dos vectores son normales.
- Esta figura muestra también el vector \mathbf{d} como la suma entre $\hat{\mathbf{d}}$ y \mathbf{e}_{min} .
- También \mathbf{e}_{min} es ortogonal a el espacio barrido por las columnas de \mathbf{A} , definido como el conjunto de todas las combinaciones lineales de vectores columna de \mathbf{A} (en este caso un plano). La estimación $\hat{\mathbf{d}}$ es solo un vector en ese espacio.

Teorema de unicidad

- El problema de cuadrados mínimos siempre tiene solución. O sea, para valores dados de la matriz de datos \mathbf{A} y el vector de datos deseados \mathbf{d} , siempre es posible encontrar un vector $\hat{\mathbf{w}}$ que satisface la ecuación normal. Es por eso importante saber en que condiciones la solución es **única**.

- Para ello introducimos el siguiente **teorema de unicidad**

La estimación de cuadrados mínimos $\hat{\mathbf{w}}$ es única si y solo si la dimensión del espacio nulo de \mathbf{A} , $\mathcal{N}(\mathbf{A})$, es cero.

- Sea \mathbf{C} una matriz de $K \times M$ (para la matriz de datos $K = N - M + 1$). Definimos el **espacio nulo** de \mathbf{C} , $\mathcal{N}(\mathbf{C})$, como el espacio de los vectores \mathbf{x} tal que $\mathbf{A}\mathbf{x} = \mathbf{0}$. La dimensión del espacio nulo de \mathbf{C} , $nulidad(\mathbf{C})$, verifica en general $nulidad(\mathbf{C}) \neq nulidad(\mathbf{C}^H)$.
- En base al teorema de unicidad es posible esperar una solución única al problema de cuadrados mínimos **solamente** cuando \mathbf{A} tiene **columnas linealmente independientes**, o sea, cuando \mathbf{A} tiene **rango de columnas completo**. Esto implica que la matriz \mathbf{A} tiene por los menos tantas filas como columnas, o de otra forma $N - M + 1 \geq M$.
- Esta última condición significa que el sistema de ecuaciones representado por $\mathbf{A}\hat{\mathbf{w}} = \mathbf{d}$ utilizado en la minimización es **sobre-determinado**, porque tiene más ecuaciones que incógnitas. De esta forma, si \mathbf{A} tiene rango de columnas completo, $\mathbf{A}^H\mathbf{A}$ ($M \times M$) **no es singular**, y la estimación de cuadrados mínimos tiene un único valor.

- Sin embargo, cuando \mathbf{A} tiene **columnas linealmente dependientes**, o sea, es deficiente en rango, la nulidad de la matriz \mathbf{A} es distinta de cero y el resultado es que puede obtenerse un número infinito de soluciones para minimizar la suma de errores al cuadrado. En tal situación, el problema de cuadrados mínimos lineal se complica en el sentido que debemos elegir cual solución particular adoptar.
- Dejaremos este caso de lado y supondremos en general que la solución de cuadrados mínimos $\hat{\mathbf{w}}$ es única.

9.8 Propiedades de la estimación

Las siguientes propiedades tienen como suposición básica que la matriz de datos \mathbf{A} es conocida **exactamente**.

1. La estimación de cuadrados mínimos $\hat{\mathbf{w}}$ no es sesgada, siempre que el proceso de error de medición $e_o(i)$ tenga media cero.
2. Cuando el proceso de error de medición $e_o(i)$ es blanco con media cero y varianza σ^2 , la matriz covarianza de la estimación de cuadrados mínimos $\hat{\mathbf{w}}$ es $\sigma^2 \Phi^{-1}$.
3. Cuando el proceso de error de medición $e_o(i)$ es blanco de media cero, la estimación de cuadrados mínimos $\hat{\mathbf{w}}$ es el mejor estimador lineal no sesgado.
4. Cuando el proceso de error de medición $e_o(i)$ es blanco y Gaussiano, con media cero, la estimación de cuadrados mínimos $\hat{\mathbf{w}}$ alcanza el límite inferior de Cramér-Rao para estimaciones no sesgadas.

9.9 Cuadrados Mínimos Recursivo (RLS)

- En implementaciones recursivas del método de cuadrados mínimos se comienza el cálculo con **condiciones iniciales** conocidas y se utiliza información contenida en las muestras nuevas para **adaptar** las estimaciones viejas. De esta forma encontramos que la longitud de los datos observables es variable.
- En consecuencia, es posible expresar la **función costo** a ser minimizada como $\mathcal{E}(n)$, donde n es la longitud variable de los datos observables. También es usual introducir un **factor de ponderación** en la definición de la función costo $\mathcal{E}(n)$, dada por

$$\mathcal{E}(n) = \sum_{i=1}^n \beta(n, i) |e(i)|^2$$

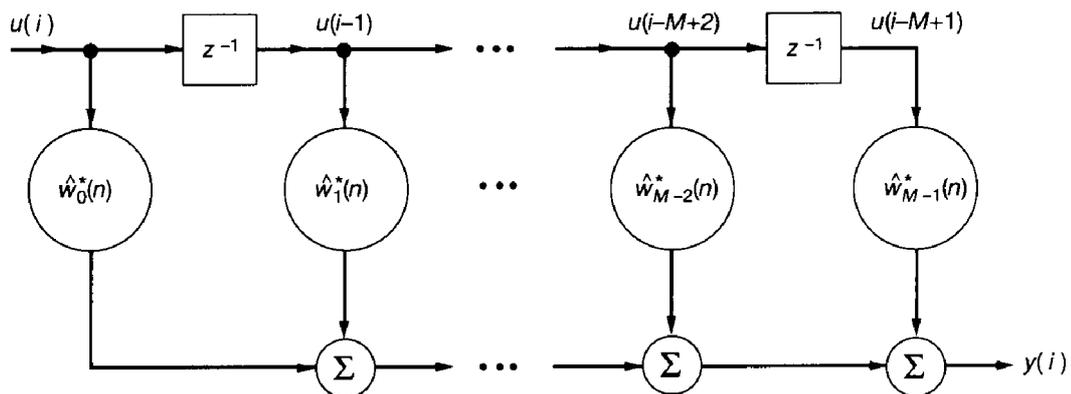


Figura 38: Filtro transversal.

donde $e(i)$ es la diferencia entre la **respuesta deseada** $d(i)$ y la salida $y(i)$ producida por un filtro transversal de las entradas $u(i), u(i - 1), \dots, u(i - M + 1)$, como se muestra en la figura 38. O sea

$$e(i) = d(i) - y(i) = d(i) - \mathbf{w}^H(n)\mathbf{u}(i)$$

donde $\mathbf{u}(i)$ es el vector de entrada en el instante i , definido por

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

y $\mathbf{w}(n)$ es el vector de coeficientes en el instante n , definido por

$$\mathbf{w}(n) = [w_o(n), w_1(n), \dots, w_{M-1}(n)]^T$$

- Notar además que los coeficientes del filtro transversal permanecen **fijos** durante el intervalo de observación $1 \leq i \leq n$ para el cual la función costo $\mathcal{E}(n)$ está definida.
- El factor de ponderación $\beta(n, i)$ tiene la propiedad $0 < \beta(n, i) \leq 1$, $i = 1, 2, \dots, n$. El uso de este factor de ponderación tiene como objetivo asegurar que los datos en el pasado distante sean "olvidados" para, de esta manera, permitir el seguimiento de las variaciones estadísticas de los datos observables cuando el filtro opera en un contexto no estacionario.
- Una forma especial de ponderación, conocida como **factor de ponderación exponencial** o **factor de olvido** está definida por

$$\beta(n, i) = \lambda^{n-i}, \quad i = 1, 2, \dots, n$$

donde λ es una constante positiva próxima, pero menor que, 1. Cuando $\lambda = 1$ se obtiene el método de cuadrados mínimos ordinario. La inversa de $1 - \lambda$, rigurosamente hablando, es una medida de la **memoria** del algoritmo. El caso especial de $\lambda = 1$ corresponde a **infinita memoria**.

- De esta forma, en el **método de cuadrados mínimos ponderado exponencialmente**, minimizamos la función costo dada por

$$\mathcal{E}(n) = \sum_{i=1}^n \lambda^{n-i} |e(i)|^2 \quad (125)$$

- El valor óptimo del vector de coeficientes $\hat{\mathbf{w}}(n)$, para el cual la función costo $\text{calE}(n)$ alcanza el valor mínimo está definido por las **ecuaciones normales**, dadas por

$$\Phi(n)\hat{\mathbf{w}}(n) = \mathbf{z}(n) \quad (126)$$

donde la matriz de correlación $\Phi(n)$ ($M \times M$) está definida por

$$\Phi(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{u}(i)\mathbf{u}^H(i) \quad (127)$$

y el vector de correlación cruzada $\mathbf{z}(n)$ es

$$\mathbf{z}(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{u}(i)d^*(i)$$

- La matriz correlación $\Phi(n)$ así definida difiere de la versión de promedios temporales de (123) en dos aspectos
 1. La matriz producto $\mathbf{u}(n)\mathbf{u}^H(n)$ dentro de la sumatoria del lado derecho de (127) está ponderada por el factor exponencial λ^{n-i} , que aparece naturalmente de la función costo (125).
 2. Se supone la utilización de **pre-ventana**, por lo cual los datos de entrada previos a $i = 1$ son iguales a cero, de ahí el límite inferior de la sumatoria.

- Consideraciones similares se aplican al vector de correlación cruzada $\mathbf{z}(n)$ comparado con su contraparte de promedios temporales.
- Separando el término correspondiente a $i = n$ del resto de la sumatoria sobre el lado derecho de (127), es posible escribir

$$\begin{aligned}\Phi(n) &= \lambda \left[\sum_{i=1}^{n-1} \lambda^{n-1-i} \mathbf{u}(i) \mathbf{u}^H(i) \right] + \mathbf{u}(n) \mathbf{u}^H(n) \\ &= \lambda \Phi(n-1) + \mathbf{u}(n) \mathbf{u}^H(n)\end{aligned}\quad (128)$$

donde $\Phi(n)$ es el valor antiguo de la matriz correlación y el producto $\mathbf{u}(n) \mathbf{u}^H(n)$ juega el papel de término de corrección en la operación de adaptación.

- En forma similar es posible obtener para $\mathbf{z}(n)$,

$$\mathbf{z}(n) = \lambda \mathbf{z}(n-1) + \mathbf{u}(n) \mathbf{d}^*(n) \quad (129)$$

- Para obtener la estimación de cuadrados mínimos $\hat{\mathbf{w}}(n)$ es necesario determinar la inversa de $\Phi(n)$. En la práctica, sin embargo, es posible evitar esa operación usando un resultado básico en álgebra matricial conocido como **lema de inversión de matrices**.



9.10 El lema de inversión de matrices

- Para \mathbf{A} y \mathbf{B} dos matrices de $M \times M$ positivas definidas definidas por

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C} \mathbf{D}^{-1} \mathbf{C}^H$$

donde \mathbf{D} es otra matriz positiva definida de $N \times M$ y \mathbf{C} es una matriz de $M \times N$. De acuerdo al **lema de inversión de matrices**, es posible expresar la inversa de la matriz \mathbf{A} como

$$\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B} \mathbf{C} (\mathbf{D} + \mathbf{C}^H \mathbf{B} \mathbf{C})^{-1} \mathbf{C}^H \mathbf{B}$$

- La prueba de este lema es constructiva, o sea, multiplicando las ecuaciones anteriores para obtener la matriz identidad. El lema de inversión de matrices se usará para obtener una forma recursiva de la ecuación de solución de cuadrados mínimos.

9.11 Algoritmo RLS ponderado exponencialmente

- Con $\Phi(n)$ supuesta positiva definida (o sea no singular), es posible aplicar el lema de inversión de matrices a (128). Relacionaremos primero las siguientes variables:

$$\mathbf{A} = \Phi(n) \quad \mathbf{B}^{-1} = \lambda \Phi(n-1) \quad \mathbf{C} = \mathbf{u}(n) \quad \mathbf{D} = 1$$

y sustituyendo estas definiciones en el lema de inversión de matrices obtenemos la siguiente ecuación recursiva para la matriz correlación

$$\Phi^{-1}(n) = \lambda^{-1} \Phi^{-1}(n-1) - \frac{\lambda^{-2} \Phi^{-1}(n-1) \mathbf{u}(n) \mathbf{u}^H(n) \Phi^{-1}(n-1)}{1 + \lambda^{-1} \mathbf{u}^H(n) \Phi^{-1}(n-1) \mathbf{u}(n)} \quad (130)$$

- Por conveniencia de notación definimos $\mathbf{P}(n) = \Phi^{-1}(n)$, la **matriz correlación inversa**, y

$$\mathbf{k}(n) = \frac{\lambda^{-1} \mathbf{P}(n-1) \mathbf{u}(n)}{1 + \lambda^{-1} \mathbf{u}^H(n) \mathbf{P}(n-1) \mathbf{u}(n)} \quad (131)$$

denominado el **vector de ganancia**. De esta manera la (130) queda

$$\mathbf{P}(n) = \lambda^{-1} \mathbf{P}(n-1) - \lambda^{-1} \mathbf{k}(n) \mathbf{u}^H(n) \mathbf{P}(n-1) \quad (132)$$

(ecuación de Riccati para el RLS). Reordenando en (131)

$$\begin{aligned} \mathbf{k}(n) &= \lambda^{-1} \mathbf{P}(n-1) \mathbf{u}(n) - \lambda^{-1} \mathbf{k}(n) \mathbf{u}^H(n) \mathbf{P}(n-1) \mathbf{u}(n) \\ &= \mathbf{P}(n) \mathbf{u}(n) \end{aligned}$$

o sea, el vector ganancia puede interpretarse como el vector de $\mathbf{u}(n)$ transformado por la inversa de la matriz correlación $\Phi(n)$.

Adaptación (temporal) del vector de coeficientes

- Para desarrollar una forma recursiva de la ecuación de adaptación de la estimación de cuadrados mínimos de $\hat{\mathbf{w}}(n)$ y suando (126), (129) y $\mathbf{P}(n) = \Phi^{-1}(n)$ se obtiene para la iteración n :

$$\begin{aligned}\hat{\mathbf{w}} &= \Phi^{-1}(n)\mathbf{z}(n) = \mathbf{P}(n)\mathbf{z}(n) \\ &= \lambda\mathbf{P}(n)\mathbf{z}(n-1) + \mathbf{P}\mathbf{u}(n)d^*(n)\end{aligned}$$

y sustituyendo $\mathbf{P}(n)$ en el primer término del lado derecho de esta ecuación por (132) se obtiene

$$\begin{aligned}\hat{\mathbf{w}} &= \mathbf{P}(n-1)\mathbf{z}(n-1) - \mathbf{k}(n)\mathbf{u}^H(n)\mathbf{P}(n-1)\mathbf{z}(n-1) + \mathbf{P}(n)\mathbf{u}(n)d^*(n) \\ &= \hat{\mathbf{w}}(n-1) - \mathbf{k}(n)\mathbf{u}^H(n)\hat{\mathbf{w}}(n-1) + \mathbf{P}(n)\mathbf{u}(n)d^*(n)\end{aligned}$$

- Finalmente, teniendo en cuenta que $\mathbf{k}(n) = \mathbf{P}(n)\mathbf{u}(n)$, se tiene que

$$\begin{aligned}\hat{\mathbf{w}}(n) &= \hat{\mathbf{w}}(n-1) + \mathbf{k}(n)[d^*(n) - \mathbf{u}^H(n)\hat{\mathbf{w}}(n-1)] \\ &= \hat{\mathbf{w}}(n-1) + \mathbf{k}(n)\xi^*(n)\end{aligned}\quad (133)$$

donde $\xi(n)$ es el **error de estimación a priori** dado por

$$\begin{aligned}\xi(n) &= d(n) - \mathbf{u}^T(n)\hat{\mathbf{w}}^*(n-1) \\ &= d(n) - \hat{\mathbf{w}}^H(n-1)\mathbf{u}(n)\end{aligned}$$

que difiere del **error a posteriori** dado por

$$e(n) = d(n) - \hat{\mathbf{w}}^H(n)\mathbf{u}(n)$$

- Notar que en la optimización de cuadrados mínimos que conducen al algoritmo recursivo de (133) la función costo está basada en $e(n)$ y no en $\xi(n)$.

Inicialización

$$\mathbf{P}(0) = \delta^{-1} \mathbf{I}, \delta \text{ una constante positiva pequeña.}$$

$$\hat{\mathbf{w}}(0) = \mathbf{0}$$

Para cada n , $n = 1, 2, \dots$, calcular

$$\mathbf{k}(n) = \frac{\lambda^{-1} \mathbf{P}(n-1) \mathbf{u}(n)}{1 + \lambda^{-1} \mathbf{u}^H(n) \mathbf{P}(n-1) \mathbf{u}(n)}$$

$$\xi(n) = d(n) - \hat{\mathbf{w}}^H(n-1) \mathbf{u}(n)$$

$$\hat{\mathbf{w}}(n) = \hat{\mathbf{w}}(n-1) + \mathbf{k}(n) \xi^*(n)$$

$$\mathbf{P}(n) = \lambda^{-1} \mathbf{P}(n-1) - \lambda^{-1} \mathbf{k}(n) \mathbf{u}^H(n) \mathbf{P}(n-1)$$

Tabla 4: Algoritmo RLS

El algoritmo de **cuadrados mínimos recursivo** (RLS) se muestra completo en la tabla

Inicialización del algoritmo RLS

- El algoritmo RLS requiere la inicialización de la recursión de (132) mediante la elección del valor inicial de $\mathbf{P}(0)$ que asegura la no singularidad de la matrix $\Phi(n)$.
- Es posible realizar esto evaluando la inversa de

$$\left[\sum_{i=-n_0}^0 \lambda^{-i} \mathbf{u}(i) \mathbf{u}^H(i) \right]^{-1}$$

donde $\mathbf{u}(i)$ se obtiene del bloque inicial de datos para $-n_0 \leq i \leq 0$.

- Un procedimiento más simple es modificar la expresión levemente para la matriz $\Phi(n)$ escribiendo

$$\Phi(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{u}(i) \mathbf{u}^H(i) + \delta \lambda^n \mathbf{I}$$

donde δ es una constante positiva pequeña. Así para $n = 0$ en la ecuación anterior, tenemos

$$\Phi(0) = \delta \mathbf{I} \quad \text{tal que} \quad \mathbf{P}(0) = \delta^{-1} \mathbf{I}$$

- Esta inicialización es equivalente a forzar que las primeras muestras de datos desconocidos tengan el valor $\lambda^{(-M+1)/2} \delta^{1/2}$ en lugar de cero. En otras palabras, debido al período de inicialización modificamos el método de pre-ventana escribiendo

$$u(n) = \begin{cases} \lambda^{(-M+1)/2} \delta^{1/2}, & n = -M + 1 \\ 0, & n < 0, n \neq -M + 1 \end{cases}$$



- Notar que para un filtro transversal de M taps, el índice $n = M + 1$ se refiere al último tap en el filtro. Cuando entra la primera muestra $u(i)$ no cero al filtro la entrada de inicialización $u(-M + 1)$ deja el filtro y desde ese momento el algoritmo RLS comienza.
- Solo falta elegir el valor inicial del vector de coeficientes, lo que usualmente se hace con $\hat{\mathbf{w}}(0) = \mathbf{0}$. δ es el único parámetro requerido. La elección de δ es que debería ser pequeño comparado con $0.01\sigma_u^2$, donde σ_u^2 .
- Es importante notar que usando el procedimiento de inicialización anterior no se está obteniendo la solución que minimiza la función costo exactamente.
- En su lugar, se está calculando la solución que minimiza

$$\mathcal{E} = \delta \lambda^n \|\mathbf{w}(n)\|^2 + \sum_{i=1}^n \lambda^{n-i} |e(i)|^2$$

- En otras palabras, el algoritmo de la tabla produce la solución recursiva exacta de

$$\min_{\mathbf{w}(n)} [\delta \lambda^n \|\mathbf{w}(n)\|^2 + \sum_{i=1}^n \lambda^{n-i} |e(i)|^2]$$

9.12 Aspectos adicionales

- El valor mínimo de la suma de errores al cuadrado ponderada \mathcal{E}_{min} , resulta cuando el vector de coeficientes es igual a la estimación de cuadrados mínimos $\hat{\mathbf{w}}(n)$.
- Para calcular $\mathcal{E}_{min}(n)$ es necesario usar la relación

$$\mathcal{E}(n) = \mathcal{E}_d(n) - \mathbf{z}^H(n)\hat{\mathbf{w}}(n) \quad (134)$$

donde $\mathcal{E}_d(n)$ está definido por

$$\mathcal{E}_d(n) = \sum_{i=1}^n \lambda^{n-i} |d(i)|^2 = \lambda \mathcal{E}_d(n-1) + |d(n)|^2$$

- De forma que

$$\begin{aligned} \mathcal{E}_{min}(n) = & \lambda[\mathcal{E}_d(n-1) - \mathbf{z}^H(n-1)\hat{\mathbf{w}}(n-1)] \\ & + d(n)[d^*(n) - \mathbf{u}^H(n)\hat{\mathbf{w}}(n-1)] - \mathbf{z}^H(n)\mathbf{k}(n)\xi^*(n) \end{aligned} \quad (135)$$

- Por definición, la expresión dentro del primer conjunto de corchetes sobre el lado derecho de la ecuación anterior es igual a $\mathcal{E}(n-1)$. También por definición, la expresión dentro del segundo conjunto de corchetes es $\xi^*(n)$. Para el último término, usamos la definición de $\mathbf{k}(n)$ para escribir

$$\mathbf{z}^H \mathbf{k}(n) = \mathbf{z}^H(n) \Phi^{-1}(n) \mathbf{u}(n) = [\Phi^{-1}(n) \mathbf{z}(n)]^H \mathbf{u}(n) = \hat{\mathbf{w}}^H(n) \mathbf{u}(n)$$

- Entonces (135) puede escribirse como

$$\begin{aligned} \mathcal{E}_{min}(n) &= \lambda \mathcal{E}_{min}(n-1) + d(n)\xi^*(n) - \hat{\mathbf{w}}^H(n) \mathbf{u}(n) \xi^*(n) \\ &= \lambda \mathcal{E}_{min}(n-1) + \xi^*(n)[d(n) - \hat{\mathbf{w}}^H(n) \mathbf{u}(n)] \\ &= \lambda \mathcal{E}_{min}(n-1) + \xi^*(n)e(n) \end{aligned}$$



Factor de Conversión

- La fórmula anterior involucra dos errores de estimación diferentes: el **error a priori** $\xi(n)$ y el **error a posteriori** $e(n)$.
- Para establecer una relación entre ellos tenemos en cuenta que

$$\begin{aligned}e(n) &= d(n) - [\hat{\mathbf{w}}(n-1) + \mathbf{k}(n)\xi^*(n)]^H \mathbf{u}(n) \\ &= d(n) - \hat{\mathbf{w}}^H(n-1)\mathbf{u}(n) - \mathbf{k}^H(n)\mathbf{u}(n)\xi(n) \\ &= (1 - \mathbf{k}^H(n)\mathbf{u}(n))\xi(n)\end{aligned}$$

- Esta relación entre los errores $e(n)$ y $\xi(n)$ se denomina **factor de conversión**

$$\gamma(n) = \frac{e(n)}{\xi(n)} = 1 - \mathbf{k}^H(n)\mathbf{u}(n)$$

9.13 Análisis de convergencia

- Consideraremos el análisis de convergencia del algoritmo RLS operando en un ambiente estacionario, para lo que se utilizará la teoría de independencia descrita en la sección del algoritmo LMS.
- Supondremos que la respuesta deseada $d(n)$ y el vector $\mathbf{u}(n)$ están relacionados por el **modelo de regresión lineal múltiple**, al que podemos escribir como

$$d(n) = e_o(n) + \mathbf{w}_o^H \mathbf{u}(n) \quad (136)$$

donde \mathbf{w}_o es el vector de parámetros de regresión (constante) y $e_o(n)$ es el error de medición, blanco, con media cero y varianza σ^2 . La suposición de \mathbf{w}_o constante implica decir que el filtro adaptivo transversal opera en un ambiente estacionario con $\lambda = 1$.

Convergencia del algoritmo RLS en media

- Comenzando con $\Phi(0) = \mathbf{0}$ que corresponde a $\mathbf{u}(0) = \mathbf{0}$ (o sea pre-ventana de los datos de entrada), y el procedimiento de inicialización, $\hat{\mathbf{w}}(n)$ obtenido por el RLS coincide casi exactamente con el método de cuadrados mínimos para $n \geq M$, donde M es el número de coeficientes del filtro adaptivo.
- De acuerdo a esto, con las ecuaciones normales

$$\hat{\mathbf{w}}(n) = \Phi^{-1}(n) \mathbf{z}(n), \quad n \geq M \quad (137)$$

donde, para $\lambda = 1$,

$$\Phi(n) = \sum_{i=1}^n \mathbf{u}(i)\mathbf{u}^H(i) \quad \text{y} \quad \mathbf{z}(n) = \sum_{i=1}^n \mathbf{u}(i)d^*(i) \quad (138)$$

- Sustituyendo (136) en (138) se obtiene

$$\begin{aligned} \mathbf{z}(n) &= \sum_{i=1}^n \mathbf{u}(i)\mathbf{u}^H(i)\mathbf{w}_o + \sum_{i=1}^n \mathbf{u}(i)e_o^*(i) \\ &= \Phi(n)\mathbf{w}_o + \sum_{i=1}^n \mathbf{u}(i)e_o^*(i) \end{aligned}$$

- Así, (137) puede reescribirse como

$$\begin{aligned} \hat{\mathbf{w}} &= \Phi^{-1}(n)\Phi(n)\mathbf{w}_o + \Phi^{-1}(n) \sum_{i=1}^n \mathbf{u}(i)e_o^*(i) \\ &= \mathbf{w}_o + \Phi^{-1}(n) \sum_{i=1}^n \mathbf{u}(i)e_o^*(i) \end{aligned} \quad (139)$$

- Utilizaremos ahora la propiedad de esperanza estadística de una variable aleatoria

$$E[x] = E[E[x|y]] \quad (140)$$

donde $E[x|y]$ es la esperanza condicional de una variable aleatoria x dada por otra variable y . La esperanza restante en el lado derecho de esta ecuación es respecto a y . Utilizando esta propiedad en (139) se obtiene

$$\begin{aligned} E[\hat{\mathbf{w}}(n)] &= \mathbf{w}_o + E\left[\Phi^{-1}(n) \sum_{i=1}^n \mathbf{u}(i)e_o^*(i)\right] \\ &= \mathbf{w}_o + E\left[E\left[\Phi^{-1}(n) \sum_{i=1}^n \mathbf{u}(i)e_o^*(i) \mid \mathbf{u}(i), i = 1, 2, \dots, n\right]\right] \end{aligned}$$



- Teniendo en cuenta que
 - $\Phi(n)$ está unívocamente definida por $\mathbf{u}(1), \mathbf{u}(2), \dots, \mathbf{u}(n)$.
 - $e_o(i)$ es independiente de $\mathbf{u}(i)$, para todo i .
 - $e(i)$ tiene media cero

de forma que

$$E[\hat{\mathbf{w}}(n)] = \mathbf{w}_o, \quad n \geq M$$

de forma que el algoritmo RLS converge en media para $n \geq M$, donde M es el número de coeficientes del filtro adaptivo. Notar que a diferencia del algoritmo LMS, el algoritmo RLS no requiere esperar que $n \rightarrow \infty$ para que la convergencia en media sea alcanzada.

Error medio cuadrático del vector de coeficientes

- Para analizar la convergencia del error medio cuadrático del vector de coeficientes $\hat{\mathbf{w}}(n)$, es posible expresar el error en los coeficientes como

$$\begin{aligned}\boldsymbol{\epsilon} &= \hat{\mathbf{w}}(n) - \mathbf{w}_o \\ &= \boldsymbol{\Phi}^{-1}(n) \sum_{i=1}^n \mathbf{u}(i) e_o^*(i)\end{aligned}$$

- Luego usando la definición de matriz correlación del error en los coeficientes se tiene que

$$\begin{aligned}\mathbf{K}(n) &= E[\boldsymbol{\epsilon}(n)\boldsymbol{\epsilon}^H(n)] \\ &= E\left[\boldsymbol{\Phi}^{-1} \sum_{i=1}^n \sum_{j=1}^n \mathbf{u}(i) e_o^*(i) e_o(j) \mathbf{u}^H(j) \boldsymbol{\Phi}^{-1}(n)\right]\end{aligned}$$

e invocando la propiedad (140)

$$\mathbf{K}(n) = E\left[\boldsymbol{\Phi}^{-1}(n) \sum_{i=1}^n \sum_{j=1}^n \mathbf{u}(i) E[e_o^*(i) e_o(j)] \mathbf{u}^H(j) \boldsymbol{\Phi}^{-1}(n)\right]$$

dado que el error de medición $e_o(i)$ se supone ruido blanco de varianza σ^2 se tiene que

$$\begin{aligned}\mathbf{K}(n) &= \sigma^2 E\left[\boldsymbol{\Phi}^{-1}(n) \sum_{i=1}^n \sum_{j=1}^n \mathbf{u}(i) \mathbf{u}^H(j) \boldsymbol{\Phi}^{-1}(n)\right] \\ &= \sigma^2 E[\boldsymbol{\Phi}^{-1}(n) \boldsymbol{\Phi}(n) \boldsymbol{\Phi}^{-1}(n)] \\ &= \sigma^2 E[\boldsymbol{\Phi}^{-1}(n)]\end{aligned}$$

- Luego, en base a las siguientes suposiciones de la teoría de la independencia
 1. $\mathbf{u}(1), \mathbf{u}(2), \dots, \mathbf{u}(n)$ son independientes y de distribución idéntica.
 2. $\mathbf{u}(1), \mathbf{u}(2), \dots, \mathbf{u}(n)$ se obtienen de un proceso estocástico con **distribución Gaussiana multivariable** de media cero y matriz correlación \mathbf{R} .
- El resultado interesante es que $\Phi(n)$ está descrita por una **distribución compleja de Wishart**. En particular para esa distribución se verifica que

$$E[\Phi^{-1}(n)] = \frac{1}{n - M - 1} \mathbf{R}^{-1}, \quad n > M + 1$$

- Sustituyendo en la ecuación anterior se obtiene

$$\mathbf{K}(n) = \frac{\sigma^2}{n - M - 1} \mathbf{R}^{-1}, \quad n > M + 1 \quad (141)$$

para la cual

$$\begin{aligned} E[\boldsymbol{\epsilon}^H(n)\boldsymbol{\epsilon}(n)] &= \text{tr}[\mathbf{K}(n)] = \frac{\sigma^2}{n - M - 1} \text{tr}[\mathbf{R}^{-1}] \\ &= \frac{\sigma^2}{n - M - 1} \sum_{i=1}^M \frac{1}{\lambda_i}, \quad n > M + 1 \end{aligned}$$

donde λ_i son los autovalores de \mathbf{R} .



- En base a esta ecuación es posible realizar las siguientes observaciones
 1. El error medio cuadrático del error en los coeficientes se **amplifica por la inversa del menor autovalor** λ_{min} , de donde, como una primera aproximación, la sensibilidad del algoritmo RLS a la dispersión de los autovalores es inversamente proporcional al menor autovalor. En consecuencia, problemas de cuadrados mínimos mal condicionados pueden conducir a malas propiedades de convergencia.
 2. El error medio cuadrático del error en los coeficientes decae casi linealmente con el número de iteraciones n . En consecuencia, la estimación producida por el algoritmo RLS converge en media cuadrática al parámetro w_o del modelo de regresión lineal casi linealmente con el tiempo.

Curva de aprendizaje del algoritmo RLS

- Existen en el algoritmo RLS dos errores de estimación, el error a priori y el error a posteriori a ser considerados.
- En base a las condiciones iniciales discutidas el valor medio cuadrático para esos errores puede ser muy diferente en el tiempo.
- Para $n = 1$, el valor medio cuadrático de $\xi(n)$ toma valores grandes (el valor medio cuadrático de $d(n)$) y decae después cuando n se incrementa.
- El valor medio cuadrático de $e(n)$, por otro lado, tiene valores pequeños para $n = 1$ y luego aumenta con n .
- De esta forma, la elección de $\xi(n)$ como el error de interés produce una curva de aprendizaje para el algoritmo RLS con una forma general similar a la del algoritmo LMS. Para comparar estas curvas analizaremos entonces el algoritmo RLS sobre la base del error de estimación a priori $\xi(n)$.
- Eliminando la respuesta deseada de la definición de $\xi(n)$ se obtiene

$$\begin{aligned}\xi(n) &= e_o(n) - [\hat{\mathbf{w}}(n-1) - \mathbf{w}_o]^H \mathbf{u}(n) \\ &= e_o(n) - \boldsymbol{\epsilon}^H(n-1) \mathbf{u}(n)\end{aligned}\quad (142)$$

- Como un índice de **desempeño estadístico** para el algoritmo RLS es conveniente utilizar el **error medio cuadrático a priori** definido por

$$J'(n) = E[|\xi(n)|^2]$$

- Reemplazando la ecuación (142) en esta ecuación se obtiene

$$\begin{aligned}
 J'(n) = & E[|e_o|^2] + E[\mathbf{u}^H \boldsymbol{\epsilon}(n-1) \boldsymbol{\epsilon}^H(n-1) \mathbf{u}(n)] \\
 & - E[\boldsymbol{\epsilon}^H(n-1) \mathbf{u}(n) e_o^*(n)] - E[e_o(n) \mathbf{u}^H(n) \boldsymbol{\epsilon}(n-1)]
 \end{aligned}$$

- Con $e_o(n)$ supuesto de media cero, el primer término de la derecha de la ecuación anterior es simplemente su varianza, σ^2 . Para los siguientes términos y teniendo en cuenta las suposiciones de la teoría de la independencia discutidas previamente se obtiene que

1. $\hat{\mathbf{w}}(n-1)$ y en consecuencia $\boldsymbol{\epsilon}(n-1)$ es independiente de $\mathbf{u}(n)$, este último obtenido de un proceso ESA de media cero. Usando la independencia estadística, junto con resultados conocidos de álgebra, es posible obtener para el segundo término

$$E[\mathbf{u}^H(n) \boldsymbol{\epsilon}(n-1) \boldsymbol{\epsilon}^H(n-1) \mathbf{u}(n)] = \text{tr}[\mathbf{R} \mathbf{K}(n-1)]$$

2. $e_o(n)$ depende de $\mathbf{u}(n)$. En consecuencia $\boldsymbol{\epsilon}(n-1)$ es independiente de $\mathbf{u}(n)$ y $e_o(n)$. Así, el tercer término del lado derecho de esa ecuación es cero teniendo en cuenta que

$$E[\boldsymbol{\epsilon}^H(n-1) \mathbf{u}(n) e_o^*(n)] = E[\boldsymbol{\epsilon}^H(n-1)] E[\mathbf{u}(n) e_o^*(n)]$$

tal que por el principio de ortogonalidad,

$$E[\boldsymbol{\epsilon}^H(n-1) \mathbf{u}(n) e_o^*(n)] = 0$$

3. El cuarto término de la derecha tiene la misma forma que el considerado en el punto anterior, tal que

$$E[e_o(n)\mathbf{u}^H(n)\boldsymbol{\epsilon}(n-1)] = 0$$

- Usando estos resultados es posible reescribir la ecuación (143) como sigue

$$J'(n) = \sigma^2 + \text{tr}[\mathbf{R} \mathbf{K}(n-1)]$$

y usando (141) (con $\lambda = 1$)

$$J'(n) = \sigma^2 + \frac{M\sigma^2}{n-M-1}, \quad n \geq M+1$$

- Con base en este resultado es posible hacer las siguientes observaciones
 1. La curva de aprendizaje promedio del algoritmo RLS converge en alrededor de $2M$ iteraciones, o sea, la velocidad de convergencia del algoritmo RLS es típicamente un orden de magnitud más rápida que la del algoritmo LMS.
 2. Para $n \rightarrow \infty$, $J'(n)$ tiende a un valor final igual a la varianza σ^2 de $e_o(n)$. En otras palabras, el algoritmo RLS teóricamente produce un exceso de error medio cuadrático cero (ó desajuste cero) cuando opera en un ambiente estacionario.
 3. La convergencia del algoritmo RLS en media cuadrática es independiente de los autovalores de \mathbf{R} .
- Debe enfatizarse que la mencionada mejora en la velocidad de convergencia del algoritmo RLS en relación al algoritmo LMS se mantiene solamente cuando $e_o(n)$ es pequeño en comparación con $d(n)$, o sea cuando la relación señal a ruido es alta. También, la propiedad de desajuste cero supone que $\lambda = 1$ (memoria infinita).



9.14 Formulación por variables de estado

- El algoritmo RLS exponencialmente ponderado fue obtenido a partir del lema de inversión de matrices, fundamentalmente en base a un **modelo matemático determinístico**.
- El algoritmo RLS puede también obtenerse en **forma exacta** directamente del algoritmo de filtrado Kalman tipo covarianza.
- Este análisis alternativo de la solución del problema de RLS es importante porque nos permite establecer una importante lista de correspondencias entre el algoritmo RLS y el filtro de Kalman.

Una comparación de modelos estocásticos y determinísticos

- Considerando el modelo dinámico no forzado

$$\mathbf{x}(n+1) = \lambda^{-1/2} \mathbf{x}(n) \quad (144)$$

$$y(n) = \mathbf{u}^H(n) \mathbf{x}(n) + \nu(n) \quad (145)$$

- De la primera ecuación es simple concluir que $\mathbf{x}(n) = \lambda^{-n/2} \mathbf{x}(0)$
- Evaluando (145) para $n = 0, 1, \dots$ y utilizando la ecuación anterior es posible obtener el siguiente sistema de ecuaciones estocásticas lineales

$$\begin{aligned} y(0) &= \mathbf{u}^H(0) \mathbf{x}(0) + \nu(0) \\ y(1) &= \lambda^{-1/2} \mathbf{u}^H(1) \mathbf{x}(0) + \nu(1) \\ &\vdots \\ y(n) &= \lambda^{-n/2} \mathbf{u}^H(n) \mathbf{x}(0) + \nu(n) \end{aligned}$$

o en forma equivalente

$$\begin{aligned}y(0) &= \mathbf{u}^H(0)\mathbf{x}(0) + \nu(0) \\ \lambda^{1/2}y(1) &= \mathbf{u}^H(1)\mathbf{x}(0) + \nu(1) \\ &\vdots \\ \lambda^{n/2}y(n) &= \mathbf{u}^H(n)\mathbf{x}(0) + \nu(n)\end{aligned}$$

- Este sistema representa una caracterización estocástica del modelo dinámico no forzado asociado desde el punto de vista de filtrado de Kalman.
- Considerando ahora la formulación determinística del problema desde el punto de vista del RLS y adecuando el modelo de la ecuación (136) es posible escribir el siguiente sistema de ecuaciones lineales simultáneas

$$\begin{aligned}d^*(0) &= \mathbf{u}^H(0)\mathbf{w}_o + e_o^*(0) \\ d^*(1) &= \mathbf{u}^H(1)\mathbf{w}_o + e_o^*(1) \\ &\vdots \\ d^*(n) &= \mathbf{u}^H(n)\mathbf{w}_o + e_o^*(n)\end{aligned}$$

- Se tiene entonces dos sistemas de ecuaciones lineales simultáneas para resolver esencialmente el mismo problema. Es de esperar que ambos conduzcan exactamente a la misma solución del problema.
- Además, teniendo en cuenta que esos dos sistemas tienen la misma forma matemática, es razonable hacer $\mathbf{x}(0) = \mathbf{w}_o$. Sobre esta base y los sistemas anteriores es posible concluir las siguientes correspondencias

$$y(n) = \lambda^{-n/2}d^*(n) \quad \nu(n) = \lambda^{-n/2}e_o^*(n)$$

Una comparación del filtrado de Kalman y el algoritmo RLS

- La comparación de las tablas que resumen los algoritmos de Kalman y RLS arrojan las siguientes correspondencias

$$\begin{aligned}
 \mathbf{K}(n-1) &= \lambda^{-1} \mathbf{P}(n-1) \\
 \mathbf{g}(n) &= \lambda^{-1/2} \mathbf{k}(n) \\
 \alpha(n) &= \lambda^{-n/2} \xi^*(n) \\
 \hat{\mathbf{x}}(n+1|Y_n) &= \lambda^{-(n+1)/2} \hat{\mathbf{w}}(n)
 \end{aligned}$$

- Además, es posible concluir que el factor de conversión $r^{-1}(n)$ en la forma especializada del filtro de Kalman tipo covarianza y el factor de conversión $\gamma(n)$ para el algoritmo RLS coinciden, o sea $r^{-1}(n) = \gamma(n)$.

Un resumen de las correspondencias entre variables de filtrado de Kalman y RLS se muestra a continuación.

Kalman		RLS	
Descripción	Variable	Descripción	Variable
Estado inicial	$\mathbf{x}(0)$	Coefficientes de regresión	\mathbf{w}_o
Estados	$\mathbf{x}(n)$	Coefficientes de regresión exponencialmente ponderados	$\lambda^{-n/2} \mathbf{w}_o$
Observaciones	$y(n)$	Respuesta deseada	$\lambda^{-n/2} d^*(n)$
Ruido de medición	$\nu(n)$	Error de medición	$\lambda^{-n/2} e_o^*(n)$
Estados del predictor de un paso	$\hat{\mathbf{x}}(n+1 Y_n)$	Coefficientes estimados	$\lambda^{-(n+1)/2} \hat{\mathbf{w}}(n)$
Correlación del error de predicción de estados	$\mathbf{K}(n)$	Inversa de la correlación de la entrada	$\lambda^{-1} \mathbf{P}(n)$
Ganancia Kalman	$\mathbf{g}(n)$	Vector ganancia	$\lambda^{-1/2} \mathbf{k}(n)$
Innovación	$\alpha(n)$	Error a priori	$\lambda^{-n/2} \xi^*(n)$
Factor de conversión	$r^{-1}(n)$	Factor de conversión	$\gamma(n)$
Condiciones iniciales	$\hat{\mathbf{x}}(1 y_0) = \mathbf{0}$ $\mathbf{K}(\mathbf{0})$	Condiciones iniciales	$\hat{\mathbf{w}}(0) = \mathbf{0}$ $\lambda^{-1} \mathbf{P}(0)$