

4 Predicción lineal

- El problema de **predicción** puede establecerse considerando un proceso estocástico $u(n), u(n-1), \dots, u(n-M)$. Involucrando solo las muestras pasadas es posible realizar una estimación de $u(n)$.
- Si U_{n-1} denota el espacio M dimensional barrido $u(n-1), u(n-2), \dots, u(n-M)$, utilizamos $\hat{u}(n|U_{n-1})$ como notación para la **predicción** de $u(n)$.
- En **predicción lineal** es posible expresar el valor estimado como una combinación lineal de las muestras en U_{n-1} . Este procedimiento se denomina **predicción lineal forward**.
- Otra forma de predicción es utilizar $u(n), u(n-1), \dots, u(n-M+1)$ para realizar una predicción de $u(n-M)$. Esta forma de predicción se denomina **predicción lineal backward**.
- Se discutirán en esta sección ambos predictores en el contexto de la teoría de filtrado de Wiener y sus aplicaciones básicas.

4.1 Predicción lineal forward

- La figura 9 muestra un **predictor forward**, consistente en un filtro transversal con M coeficientes $\mathbf{w}_f = [w_{f,1}, w_{f,2}, \dots, w_{f,M}]^T$.

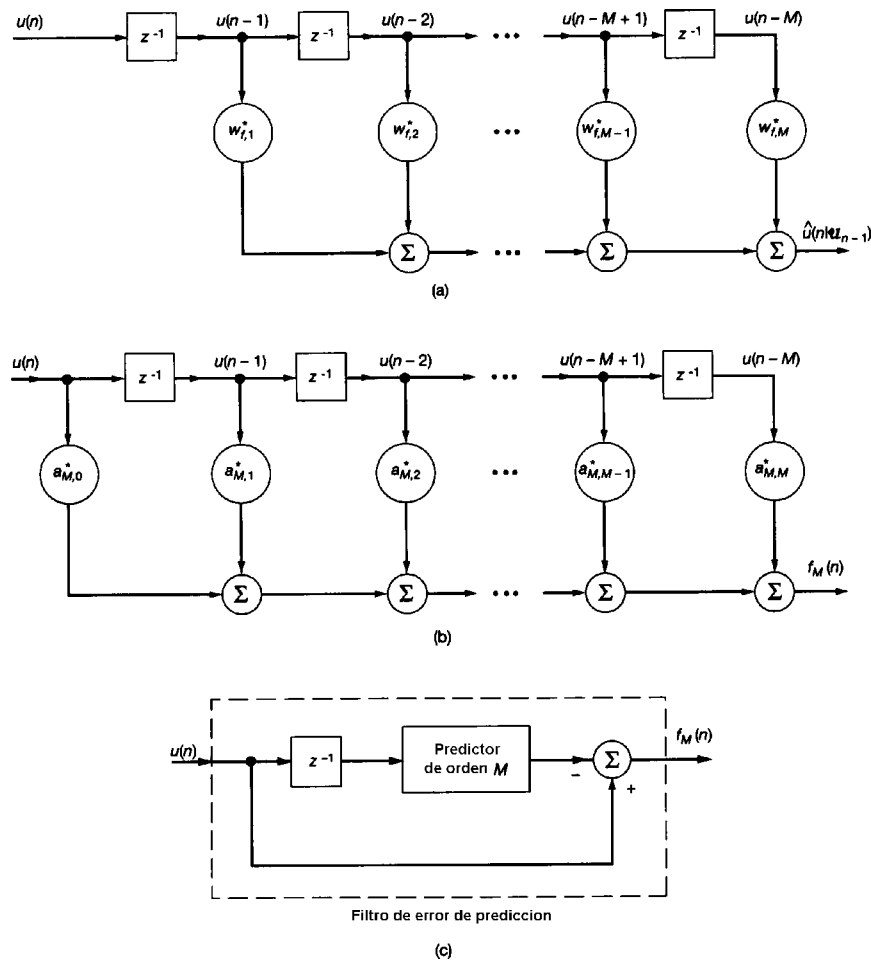


Figura 9: a) Predictor de un paso, b) filtro de error de predicción, c) relación entre el predictor y el filtro de error de predicción.

- Asumimos que $u(n)$ es un proceso ESA de media cero, y se optimizarán los valores de los coeficientes en sentido medio cuadrático siguiendo la teoría de filtrado de Wiener.

- La predicción de $u(n)$ será

$$\hat{\mathbf{u}}(n|U_{n-1}) = \sum_{k=1}^M w_{f,k}^* u(n-k)$$

- El **error de predicción forward** $f_M(n)$ es la diferencia entre la respuesta deseada $d(n) = u(n)$ y el valor estimado, o sea

$$f_M(n) = u(n) - \hat{\mathbf{u}}(n|U_{n-1})$$

donde el subíndice M indica el número de retardos necesarios para realizar la predicción.

- Con los coeficientes del predictor en el valor óptimo, el **error medio cuadrático mínimo** estará dado por

$$P_M = E[|f_M(n)|^2]$$

- Como los coeficientes se supuso tenían media cero, también la tendrá $f_M(n)$ y en consecuencia P_M es también la varianza de $f_M(n)$.
- Para obtener la solución de Wiener para los coeficientes $\mathbf{w}_f$ es necesario obtener la matrix correlación, el vector de correlación cruzada y además para obtener P_M , la varianza de $u(n)$, entonces

1. Para el vector de entradas definido por

$$\mathbf{u}(n-1) = [u(n-1), u(n-2), \dots, u(n-M)]^T$$

la matriz correlación será

$$\begin{aligned} \mathbf{R} &= E[\mathbf{u}(n-1)\mathbf{u}^H(n-1)] \\ &= \begin{bmatrix} r(0) & r(1) & \cdots & r(M-1) \\ r^*(1) & r(0) & \cdots & r(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(M-1) & r^*(M-2) & \cdots & r(0) \end{bmatrix} \end{aligned}$$

de forma que coincide con la matriz correlación de un proceso estacionario, como discutido en la sección anterior.

2. El vector de correlación cruzada entre $u(n-1), \dots, u(n-M)$ y la respuesta deseada $u(n)$ será

$$\begin{aligned} \mathbf{r} &= E[\mathbf{u}(n-1)u^*(n-1)] \\ &= \begin{bmatrix} r^*(1) \\ r^*(2) \\ \vdots \\ r^*(M) \end{bmatrix} = \begin{bmatrix} r(-1) \\ r(-2) \\ \vdots \\ r(-M) \end{bmatrix} \end{aligned}$$

3. La varianza de $u(n)$ es $r(0)$, dado que $u(n)$ tiene media cero.
- De esta forma, la solución óptima de Wiener-Hopf para resolver el problema de predicción lineal forward puede escribirse como

$$\mathbf{R} \mathbf{w}_f = \mathbf{r} \quad (19)$$

y el error medio cuadrático mínimo toma la forma

$$P_M = r(0) - \mathbf{r}^H \mathbf{w}_f \quad (20)$$

4.1.1 Relación entre predicción lineal y modelado AR

- Es sumamente informativo comparar las ecuaciones de Wiener-Hopf para el problema de predicción lineal (19) con las ecuaciones de Youle-Walker para un modelo AR.
- Las dos tienen la misma forma matemática. Además, la varianza del error de predicción forward tiene también la misma forma matemática que la varianza del proceso de ruido blanco utilizado para excitar el modelo AR.
- Para el caso de un proceso AR entonces esto implica que si el predictor es optimizado en sentido medio cuadrático sus coeficientes corresponderán a los de aquel modelo.
- Esto no es sorprendente si se tiene en cuenta que la ecuación que define el error de predicción es igual a la ecuación a diferencias que define un proceso AR.
- Cuando el proceso de entrada al predictor no es AR, sin embargo, la utilización del predictor solo provee una aproximación al proceso de entrada.

4.1.2 Ecuaciones de Wiener-Hopf aumentadas para predicción forward

- Dado el error de predicción forward dado por

$$f_M(n) = u(n) - \sum_{k=1}^M w_{f,k}^* u(n-k)$$

y en base a la siguiente redefinición de coeficientes

$$a_{M,k} = \begin{cases} 1, & k = 0 \\ -w_{f,k}, & k = 1, 2, \dots, M \end{cases}$$

sirve para definir el **filtro de error de predicción forward**, que opera sobre $u(n), u(n-1), \dots, u(n-M)$ y tiene como salida

$$f_M(n) = \sum_{k=0}^M a_{M,k}^* u(n-k)$$

- En base a la solución matricial para el problema de predicción forward es posible escribir

$$\begin{bmatrix} r(0) & \mathbf{r}^H \\ \mathbf{r} & \mathbf{R} \end{bmatrix} \begin{bmatrix} 1 \\ -\mathbf{w}_f \end{bmatrix} = \begin{bmatrix} P_M \\ \mathbf{0} \end{bmatrix} \quad \begin{bmatrix} r(0) & \mathbf{r}^H \\ \mathbf{r} & \mathbf{R} \end{bmatrix} \mathbf{a}_M = \begin{bmatrix} P_M \\ \mathbf{0} \end{bmatrix} \quad (21)$$

donde ahora la matriz asociada es también de correlación pero de orden $(M+1 \times M+1)$.

- En forma expandida

$$\sum_{k=0}^M a_{M,i} r(k-i) = \begin{cases} P_M, & i = 0 \\ 0, & i = 1, 2, \dots, M \end{cases}$$

que se denominan **ecuaciones de Wiener-Hopf aumentadas**.

Ejemplo:

- Para el caso de un filtro de error de predicción de orden $M = 1$, las ecuaciones aumentadas serán

$$\begin{bmatrix} r(0) & r(1) \\ r^*(1) & r(0) \end{bmatrix} \begin{bmatrix} a_{1,0} \\ a_{1,1} \end{bmatrix} = \begin{bmatrix} P_1 \\ \mathbf{0} \end{bmatrix}$$

tal que

$$a_{1,0} = \frac{P_1}{\Delta_r} r(0) \quad a_{1,1} = -\frac{P_1}{\Delta_r} r^*(1)$$

donde $\Delta_r = r^2(0) - |r(1)|^2$ (determinante de $\mathbf{R}$).

- Como $a_{1,0} = 1$, entonces: $P_1 = \Delta_r / r(0)$ y $a_{1,1} = -r^*(1) / r(0)$.
- De forma similar para el caso de un filtro de error de predicción de orden $M = 2$ se tendrá que

$$\begin{aligned} P_2 &= \frac{\Delta_r}{r^2(0) - |r(1)|^2} \\ a_{2,1} &= -\frac{r^*(1)(r(0) - r(1)r^*(2))}{r^2(0) - |r(1)|^2} \\ a_{2,2} &= \frac{(r^*(1))^2 - r(0)r^*(2)}{r^2(0) - |r(1)|^2} \end{aligned}$$

4.2 Predicción lineal backward

- Para U_n el espacio M dimensional barrido por $u(n)$, $u(n - 1)$, $\dots$, $u(n - M + 1)$ es posible realizar la **predicción backward** (hacia atrás) de $u(n - M)$ en forma lineal como

$$\hat{u}(n - M | U_n) = \sum_{k=1}^M e_{b,k}^* u(n - k + 1)$$

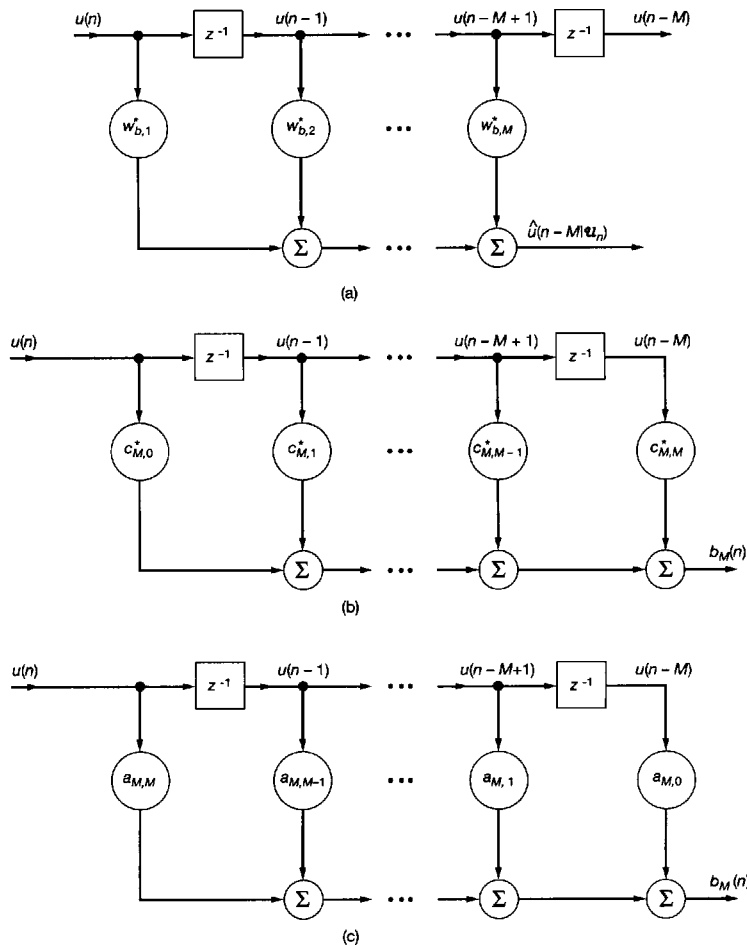


Figura 10: a) predictor backward de un paso, b) filtro de error de predicción forward, filtro de error de predicción backward definido en términos de los coeficientes del filtro correspondiente de predicción forward

- Para obtener la solución de Wiener de $\mathbf{w}_b = [w_{b,1}, w_{b,2}, \dots, w_{b,M}]^T$, es necesario considerar que la respuesta deseada es ahora $d(n) = u(n - M)$.

- Por otro lado el error de predicción backward será

$$b_M(n) = u(n - M) - \hat{u}(n - M | U_n)$$

donde el subíndice M está relacionado al número de muestras necesarias para obtener la predicción. En particular, para el valor óptimo de $\mathbf{w}_b$, el **error medio cuadrático mínimo** será $P_M = E[|b_M|^2]$.

- Para resolver el problema de filtrado de Wiener son necesarias: la matriz correlación y el vector de correlación cruzada. Para obtener P_M será necesaria la varianza de $u(n)$. Considerando esas tres cantidades:

1. Para el vector de entradas definido por

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

la matriz correlación será

$$\mathbf{R} = E[\mathbf{u}(n)\mathbf{u}^H(n)]$$

2. El vector de correlación cruzada entre $\mathbf{u}(n)$ y la respuesta deseada $u(n - M)$ será

$$\begin{aligned} \mathbf{r}^{B*} &= E[\mathbf{u}(n)u^*(n - M)] \\ &= \begin{bmatrix} r(M) \\ r(M - 1) \\ \vdots \\ r(1) \end{bmatrix} \end{aligned}$$

donde el subíndice B indica el ordenamiento hacia atrás de los coeficientes de correlación.

3. La varianza de $u(n - M)$ es $r(0)$, dado que $u(n)$ es estacionario y tiene media cero.
- De esta forma, la solución óptima de Wiener-Hopf para resolver el problema de predicción lineal backward será

$$\mathbf{R} \mathbf{w}_b = \mathbf{r}^{B*} \quad (22)$$

y el error medio cuadrático mínimo toma la forma

$$P_M = r(0) - \mathbf{r}^{BT} \mathbf{w}_b \quad (23)$$

La relación entre las ecuaciones de Wiener-Hopf para predicción backward y forward y filtrado son

	Filtro de Wiener	Predictor Forward	Predictor Backward
Entrada	$\mathbf{u}(n)$	$\mathbf{u}(n-1)$	$\mathbf{u}(n)$
Respuesta deseada	$d(n)$	$u(n)$	$u(n - M)$
Coefficientes	$\mathbf{w}_o$	$\mathbf{w}_f$	$\mathbf{w}_b$
Error de estimación	$e(n)$	$f_M(n)$	$b_M(n)$
Matriz Correlación	$\mathbf{R}$	$\mathbf{R}$	$\mathbf{R}$
Correlación cruzada	$\mathbf{p}$	$\mathbf{r}$	$\mathbf{r}^{B*}$
Error medio cuadrático mínimo	J_{min}	P_M	P_M

Relación entre predictores backward y forward

- En base a las ecuaciones de Wiener Hopf para los predictores forward y backward, y como $\mathbf{R}^H = \mathbf{R}$,

$$\mathbf{w}_b^{B*} = \mathbf{w}_f \quad (24)$$

- Con propósito similar y considerando el error medio cuadrático mínimo para el predictor backward reescrito como

$$P_M = r(0) - \mathbf{r}^T \mathbf{w}_b^B$$

y en base a la ecuación (24) se verifica que es igual al error medio cuadrático mínimo para el caso del predictor forward.

Filtro de error de predicción backward.

- En forma similar al filtro de error de predicción forward es posible definir el **filtro de error de predicción backward**, en este caso a partir de

$$b_M(n) = u(n - M) - \sum_{k=1}^M w_{b,k}^* u(n - k + 1) = \sum_{k=0}^M c_{M,k}^* u(n - k + 1)$$

donde $c_{M,k} = \begin{cases} -w_{b,k+1} & k = 0, 1, \dots, M - 1 \\ 1, & k = M \end{cases}$ son los coeficientes del filtro de error de predicción backward.

- En base a (24) en forma escalar se tiene que $w_{b,M-k+1}^* = w_{f,k}$, para $k = 1, 2, \dots, M$, ó en forma equivalente

$$w_{b,k} = w_{f,M-k+1}^*, \quad k = 1, 2, \dots, M$$

de donde sustituyendo en la definición de los coeficientes del filtro de error de predicción backward se tiene que

$$c_{M,k} = \begin{cases} -w_{f,M-k}^* & k = 0, 1, \dots, M-1 \\ 1, & k = M \end{cases}$$

y finalmente usando la relación entre los coeficientes del filtro de error de predicción forward y los del predictor forward se obtiene que

$$c_{M,k} = a_{M,M-k}^*, \quad k = 0, 1, \dots, M$$

- De esta forma es posible expresar la relación entrada salida del filtro de error de predicción backward como

$$b_M(n) = \sum_{k=0}^M a_{M,M-k} u(n-k)$$

Ecuaciones de Wiener-Hopf aumentadas para predicción backward.

- Las ecuaciones de Wiener-Hopf para predicción backward y el error medio cuadrático mínimo pueden reescribirse (en forma similar al caso de predicción forward) como

$$\begin{bmatrix} \mathbf{R} & \mathbf{r}^{B*} \\ \mathbf{r}^{BT} & r(0) \end{bmatrix} \begin{bmatrix} -\mathbf{w}_b \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ P_M \end{bmatrix}$$

y en forma escalar, usando los coeficientes del filtro de error de predicción forward

$$\sum_{k=0}^M a_{M,M-k}^* r(k-i) = \begin{cases} 0, & i = 0, \dots, M-1 \\ P_M, & i = M \end{cases}$$

4.3 Algoritmo de Levinson Durbin

- Se discutirá ahora un método para calcular los coeficientes del filtro de error de predicción y el error de predicción mínimo resolviendo las ecuaciones aumentadas de Wiener-Hopf. El método es recursivo por sí mismo y hace uso particular de la estructura Toeplitz de la matriz correlación de la entrada.
- Basicamente el procedimiento, denominado **Algoritmo de Levinson - Durbin**, utiliza la solución de las ecuaciones aumentadas de Wiener-Hopf del filtro de error de predicción de orden $m - 1$ para obtener la solución correspondiente para el filtro de error de predicción de orden m . El orden va de $M = 1, 2, \dots$, hasta M , donde M es el orden final del filtro.
- La característica importante de este algoritmo es su eficiencia computacional para resolver las ecuaciones aumentadas como contrapartida a un método clásico como el de eliminación de Gauss.
- Si $\mathbf{a}_m$ ($m + 1 \times 1$) es el vector de coeficientes del filtro de error de predicción de orden m , vimos que el vector correspondiente de coeficientes del filtro de error de predicción backward puede obtenerse de él reordenando y conjugando sus componentes, o sea $\mathbf{a}_m^{B*}$. Consideremos además los vectores $\mathbf{a}_{m-1}$ y $\mathbf{a}_{m-1}^{B*}$ de $(m \times 1)$ los coeficientes correspondientes a los filtros de error de predicción forward y backward de orden $m - 1$.

- La recursión de Levinson-Durbin puede establecerse en cualquiera de las siguientes formas equivalentes:

1. La **recursión de orden** para obtener el vector de coeficientes del filtro de error de predicción forward es

$$\mathbf{a}_m = \begin{bmatrix} \mathbf{a}_{m-1} \\ 0 \end{bmatrix} + \kappa_m \begin{bmatrix} 0 \\ \mathbf{a}_{m-1}^{B*} \end{bmatrix} \quad (25)$$

donde κ_m es una constante, con versión escalar

$$a_{m,l} = a_{m-1,l} + \kappa_m a_{m-1,m-l}^*, \quad l = 0, 1, \dots, m \quad (26)$$

donde $a_{m,l}$ es el coeficiente l -ésimo del filtro de error de predicción forward de orden m , similarmente para $a_{m-1,l}$ pero de orden $m - 1$ y $a_{m-1,m-l}^*$ es el l -ésimo coeficiente del filtro de error de predicción backward de orden $m - 1$. Notar que $a_{m-1,0} = 1$ y $a_{m-1,m} = 0$.

2. La **recursión de orden** para obtener el vector de coeficientes del filtro de error de predicción backward es

$$\mathbf{a}_m^{B*} = \begin{bmatrix} 0 \\ \mathbf{a}_{m-1}^{B*} \end{bmatrix} + \kappa_m^* \begin{bmatrix} \mathbf{a}_{m-1} \\ 0 \end{bmatrix}$$

y la versión escalar

$$a_{m,m-l}^* = a_{m-1,m-l}^* + \kappa_m^* a_{m-1,l}, \quad l = 0, 1, \dots, m \quad (27)$$

donde $a_{m,m-l}^*$ es el coeficiente l -ésimo del filtro de error de predicción backward de orden m , similarmente para $a_{m-1,m-l}^*$ pero de orden $m - 1$ y $a_{m-1,l}$ es el l -ésimo coeficiente del filtro de error de predicción forward de orden $m - 1$.

- El algoritmo de recursión de orden puede formularse en base a uno u otro tipo de filtro de error de predicción ya que una formulación puede obtenerse de la otra.
- Para establecer la relación entre este algoritmo y la solución de las ecuaciones aumentadas de Wiener-Hopf, es posible analizar la constante κ_m . Con ese objetivo, premultiplicando la recursión de la ecuación (25) por R_{m+1} , la matriz de correlación de $(m + 1 \times m + 1)$, se tiene que
 1. El término de la izquierda toma la forma (ver ecuacion aumentada (21))

$$\mathbf{R}_{m+1} \mathbf{a}_m = \begin{bmatrix} P_m \\ \mathbf{0}_m \end{bmatrix} \quad (28)$$

donde la notación para $\mathbf{R}_{m+1}$ y $\mathbf{0}_m$ hace referencia a sus dimensiones y en cambio $\mathbf{a}_m$ y P_m hace referencia al orden del predictor.

2. En el primer término de la derecha usamos una de las particiones posibles de la matriz correlación $\mathbf{R}_{m+1}$, de forma que

$$\mathbf{R}_{m+1} \begin{bmatrix} \mathbf{a}_{m-1} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_m & \mathbf{r}_m^{B*} \\ \mathbf{r}_m^{BT} & r(0) \end{bmatrix} \begin{bmatrix} \mathbf{a}_{m-1} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_m \mathbf{a}_{m-1} \\ \mathbf{r}_m^{BT} \mathbf{a}_{m-1} \end{bmatrix} \quad (29)$$

Notar que las ecuaciones aumentadas del filtro de error de predicción de orden $m - 1$ son

$$\mathbf{R}_m \mathbf{a}_{m-1} = \begin{bmatrix} P_{m-1} \\ \mathbf{0}_{m-1} \end{bmatrix}$$

y definiendo la constante

$$\Delta_{m-1} = \mathbf{r}_m^{BT} \mathbf{a}_{m-1} = \sum_{l=0}^{m-1} r(l-m) a_{m-1,l} \quad (30)$$

la ecuación (29) puede reescribirse finalmente como

$$\mathbf{R}_{m+1} \begin{bmatrix} \mathbf{a}_{m-1} \\ 0 \end{bmatrix} = \begin{bmatrix} P_{m-1} \\ \mathbf{0}_{m-1} \\ \Delta_{m-1} \end{bmatrix} \quad (31)$$

3. En el segundo término de la derecha usamos otra de las particiones posibles de la matriz correlación $\mathbf{R}_{m+1}$, de forma que

$$\mathbf{R}_{m+1} \begin{bmatrix} 0 \\ \mathbf{a}_{m-1}^{B*} \end{bmatrix} = \begin{bmatrix} r(0) & \mathbf{r}_m^H \\ \mathbf{r}_m & \mathbf{R}_m \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{a}_{m-1}^{B*} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_m^H \mathbf{a}_{m-1}^{B*} \\ \mathbf{R}_m \mathbf{a}_{m-1}^{B*} \end{bmatrix} \quad (32)$$

Para este caso las ecuaciones aumentadas del filtro de error de predicción de orden $m-1$ son

$$\mathbf{R}_m \mathbf{a}_{m-1}^{B*} = \begin{bmatrix} \mathbf{0}_{m-1} \\ P_{m-1} \end{bmatrix}$$

y teniendo en cuenta que

$$\begin{aligned} \mathbf{r}_m^H \mathbf{a}_{m-1}^{B*} &= \sum_{k=1}^m r^*(-k) a_{m-1,m-k}^* \\ &= \sum_{l=0}^{m-1} r^*(l-m) a_{m-1,l}^* = \Delta_{m-1}^* \end{aligned}$$

la ecuación (32) puede reescribirse finalmente como

$$\mathbf{R}_{m+1} \begin{bmatrix} 0 \\ \mathbf{a}_{m-1}^{B*} \end{bmatrix} = \begin{bmatrix} \Delta_{m-1}^* \\ \mathbf{0}_{m-1} \\ P_{m-1} \end{bmatrix} \quad (33)$$

De esta forma el resultado es

$$\begin{bmatrix} P_m \\ \mathbf{0}_m \end{bmatrix} = \begin{bmatrix} P_{m-1} \\ \mathbf{0}_{m-1} \\ \Delta_{m-1} \end{bmatrix} + \kappa_m \begin{bmatrix} \Delta_{m-1}^* \\ \mathbf{0}_{m-1} \\ P_{m-1} \end{bmatrix} \quad (34)$$

De esta ecuación puede deducirse que:

1.

$$P_m = P_{m-1} + \kappa \Delta_{m-1}^*$$

2.

$$0 = \Delta_{m-1} + \kappa_m P_{m-1} \quad (35)$$

de donde $\kappa_m = -\frac{\Delta_{m-1}}{P_{m-1}}$.

3. De las anteriores

$$P_m = P_{m-1}(1 - |\kappa_m|^2) \quad (36)$$

Cuando el orden del filtro de error de predicción aumenta, P_m decrece (no puede ser nunca negativo), de forma que $0 \leq P_m \leq P_{m-1}$, $m \geq 1$. Como $P_0 = r(0)$, el error medio cuadrático mínimo **final** será

$$P_M = P_0 \prod_{m=1}^M (1 - |\kappa_m|^2)$$

Interpretación de los parámetros κ_m y Δ_{m-1}

- κ_m , $1 \leq m \leq M$ se denominan **coeficientes de reflexión**.
- Notar que, en base a (36)

$$|\kappa_m| \leq 1, \quad \text{para todo } n$$

- De (26) es posible concluir además que para un filtro de error de predicción de orden m , κ_m es igual al último coeficiente del filtro $a_{m,m}$, o sea

$$\kappa_m = a_{m,m}$$

- Por otro lado, el parámetro Δ_{m-1} puede interpretarse como la correlación cruzada entre el error de predicción forward $f_{m-1}(n)$ y una versión retrasada del error de predicción backward $b_{m-1}(n-1)$, ya que

$$TZ[b_m(n)] = TZ[b_{m-1}(n-1)] + \kappa_m TZ[f_{m-1}(n)]$$

y de esta forma

$$\begin{aligned} E[f_{m-1}^*(n)b_{m-1}(n-1)] &= E[f_{m-1}^*(n)(b_m(n) - \kappa_m f_{m-1}(n))] \\ &= \Delta_{m-1} \end{aligned}$$

- Notar que $f_0(n) = b_0(n) = u(n)$, para $u(n)$ la entrada del filtro de error de predicción en el instante n y en particular

$$\Delta_0 = E[b_0(n-1)f_0^*(n)] = E[u(n-1)u^*(n)] = r^*(1)$$

Utilización del algoritmo de Levinson-Durbin

- Existen varias formas de utilizar el algoritmo de Levinson-Durbin para calcular los coeficientes del filtro de error de predicción $a_{M,k}$, $k = 0, 1, \dots, M$ y el error medio cuadrático mínimo para el predictor de orden M , P_M , como por ejemplo:

1. Usando un estimador de $r(0), r(1), \dots, r(M)$, como por ejemplo

$$\hat{r}(k) = \frac{1}{N} \sum_{n=1}^N u(n)u^*(n-k), \quad k = 0, 1, \dots, M$$

donde N es la longitud total de la secuencia de entrada $N \gg M$. Existen lógicamente otros estimadores que es posible usar. En cualquier caso es necesario obtener Δ_{m-1} y P_m . Luego, la recursión comienza con $m = 0$, tal que $P_0 = r(0)$ y $\Delta_0 = r^*(1)$ ($a_{m,0} = 1$ para todo m y $a_{m,k} = 0$ para $k > m$). Las estimaciones obtenidas por este método se denominan de **estimaciones de Youle-Walker**.

2. Usando un estimador de $\kappa_1, \kappa_2, \dots, \kappa_M$ y $r(0)$. En ese caso se pueden utilizar

$$\begin{aligned} a_{m,k} &= a_{m-1,k} + \kappa_m a_{m-1,m-k}^*, \quad k = 0, 1, \dots, m \\ P_m &= P_{m-1}(1 - |\kappa_m|^2) \end{aligned}$$

4.3.1 Algoritmo de Levinson Durbin inverso

- Si se dispone de una estimación de $a_{M,1}, a_{M,2}, \dots, a_{M,M}$ ($a_{M,0} = 1$) y se desea obtener $\kappa_1, \kappa_2, \dots, \kappa_M$ es posible utilizar el algoritmo de Levinson Durbin en forma invertida, o simplemente la **recursión inversa**.
- Usando (26) y (27) en forma matricial como

$$\begin{bmatrix} a_{m,k} \\ a_{m,m-k}^* \end{bmatrix} = \begin{bmatrix} 1 & \kappa_m \\ \kappa_m^* & 1 \end{bmatrix} \begin{bmatrix} a_{m-1,k} \\ a_{m-1,m-k}^* \end{bmatrix}, \quad k = 0, 1, \dots, m$$

donde $m = 1, 2, \dots, M$. Luego, suponiendo que $|\kappa_m| < 1$ y resolviendo la ecuación anterior para $a_{m-1,k}$ (con $\kappa_m = a_{m,m}$) se tiene que

$$a_{m-1,k} = \frac{a_{m,k} - a_{m,m} a_{m,m-k}^*}{1 - |a_{m,m}|^2}, \quad k = 0, 1, \dots, m$$

- El procedimiento entonces sería comenzar con la recursión anterior, comenzando con los coeficientes $\{a_{M,k}\}$ y luego decreciendo el orden $m = M, M - 1, \dots, 2$ para obtener los coeficientes del filtro de error de predicción de orden $M - 1, M - 2, \dots, 1$.
- Luego es posible utilizar:

$$\kappa_m = a_{m,m}, \quad m = M, M - 1, \dots, 1$$

para estimar el conjunto de coeficientes de reflexión $\kappa_M, \kappa_{M-1}, \dots, \kappa_1$.

Ejemplo.

- Suponiendo tener los coeficientes del filtro de error de predicción de orden 3 y es necesario obtener $\kappa_1, \kappa_2, \kappa_3$. Con el procedimiento anterior, para $m = 3, 2$ se obtiene

1. Para el filtro de error de predicción de orden 2 ($m = 3$):

$$a_{2,1} = \frac{a_{3,1} - a_{3,3}a_{3,2}^*}{1 - |a_{3,3}|^2} \quad a_{2,2} = \frac{a_{3,2} - a_{3,3}a_{3,1}^*}{1 - |a_{3,3}|^2}$$

2. Para el filtro de error de predicción de orden 1 ($m = 2$):

$$a_{1,1} = \frac{a_{2,1} - a_{2,2}a_{2,1}^*}{1 - |a_{2,2}|^2}$$

3. De esta forma: $\kappa_3 = a_{3,3}$, $\kappa_2 = a_{2,2}$ y $\kappa_1 = a_{1,1}$

4.4 Propiedades de filtros de error de predicción

1. **Relación entre la función autocorrelación y los coeficientes de reflexión.** Una forma alternativa de describir la estadística de segundo orden de un proceso estacionario es a través de los coeficientes de reflexión, o sea $\kappa_1, \kappa_2, \dots, \kappa_M$ y P_0 ($= r(0)$).
2. **Función transferencia de un filtro de error de predicción forward.**

$$H_{f,m}(z) = H_{f,m-1}(z) + \kappa_m^* z^{-1} H_{b,m-1}(z) \quad (37)$$

3. **Un filtro de error de predicción forward es de fase mínima.**
4. **Un filtro de error de predicción backward es de fase máxima.**
5. **Un filtro de error de predicción forward es un filtro de blanqueo.**

4.5 Factorización de Cholesky

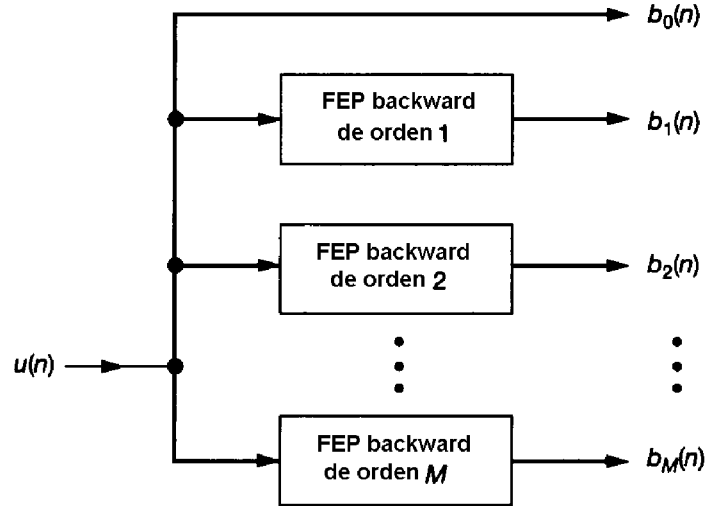


Figura 11: Conexión en paralelo de un banco de filtros de error de predicción backward de ordenes de 0 a M .

- Consideraremos un banco de filtros de error de predicción backward de orden 0 a M conectados en paralelo. Es posible expresar la salida de los filtros en la forma

$$\begin{aligned}
 b_0(n) &= u(n) \\
 b_1(n) &= a_{1,1}u(n) + a_{1,0}u(n-1) \\
 b_2(n) &= a_{2,2}u(n) + a_{2,1}u(n-1) + a_{2,0}u(n-2) \\
 &\vdots \\
 b_M(n) &= a_{M,M}u(n) + a_{M,M-1}u(n-1) + \cdots + a_{M,0}u(n-M)
 \end{aligned}$$

definiendo los vectores de entradas y de errores (de $(M+1) \times 1$) como

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

$$\mathbf{b}(n) = [b_0(n), b_1(n), \dots, b_M(n)]^T$$

entonces el sistema puede expresarse como

$$\mathbf{b}(n) = \mathbf{L} \mathbf{u}(n) \quad (38)$$

donde

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{1,1} & 1 & \cdots & 0 \\ & & \vdots & \\ a_{M,M} & a_{M,M-1} & \cdots & 1 \end{bmatrix}$$

tiene tres propiedades útiles

1. $\mathbf{L}$ es **triangular inferior** con unos sobre la diagonal principal.
 2. El determinante de $\mathbf{L}$ es uno, en consecuencia no es singular.
 3. Los elementos no nulos de cada fila de $\mathbf{L}$, excepto por la conjugación compleja, son iguales a los pesos del filtro de error de predicción backward cuyo orden corresponde a la posición de esa fila.
- La transformación de (38) se conoce como **algoritmo de ortogonalización de Gram-Schmidt**. Muestra que existe una correspondencia uno a uno entre $\mathbf{u}(n)$ y $\mathbf{b}(n)$. Por otro lado,

$$\mathbf{u}(n) = \mathbf{L}^{-1} \mathbf{b}(n)$$

Ortogonalidad entre errores de predicción backward.

- Los errores de predicción backward $b_0(n), b_1(n), \dots, b_M(n)$ tienen una propiedad muy importante. En particular,

$$E[b_m(n)b_i^*(n)] = \begin{cases} P_m, & i = m \\ 0, & i \neq m \end{cases}$$

- Notar que esta propiedad es válida solo para un proceso de entrada ESA. Esto muestra además que el proceso de ortogonalización transforma la entrada correlacionada en una secuencia de errores no correlacionados.

Factorización de $\mathbf{R}^{-1}$

- Teniendo en cuenta la ortogonalidad de los errores de predicción backward es posible escribir

$$\begin{aligned} E[\mathbf{b}(n)\mathbf{b}^H(n)] &= E[\mathbf{L}\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{L}^H] \\ &= \mathbf{L} E[\mathbf{u}(n)\mathbf{u}^H(n)] \mathbf{L}^H \end{aligned}$$

donde además es posible definir $\mathbf{D} = E[\mathbf{b}(n)\mathbf{b}^H(n)]$, tal que

$$\mathbf{D} = \mathbf{L}\mathbf{R}\mathbf{L}^H$$

- Notar que como $\mathbf{R}$ es positiva definida su inversa existe y lo mismo para $\mathbf{D}$. Esta matriz es diagonal por la ortogonalidad de los errores de predicción backward. En consecuencia

$$\mathbf{D} = \text{diag}(P_0, P_1, \dots, P_M)$$

donde P_i es el error medio cuadrático mínimo asociado al i -ésimo error de predicción backward $b_i(n)$, o sea $P_i = E[|b_i(n)|^2]$, $i = 0, 1, \dots, M$.

- Entonces

$$\mathbf{D}^{-1} = \text{diag}(P_0^{-1}, P_1^{-1}, \dots, P_M^{-1})$$

- De esta forma es posible factorizar $\mathbf{R}^{-1}$ como

$$\mathbf{R}^{-1} = \mathbf{L}^H \mathbf{D}^{-1} \mathbf{L} = (\mathbf{D}^{-1/2} \mathbf{L})^H (\mathbf{D}^{-1/2} \mathbf{L})$$

- Esta transformación se denomina **factorización de Cholesky**. De acuerdo a esta factorización, la inversa de la matriz correlación $\mathbf{R}^{-1}$ puede factorizarse en el producto de una matriz triangular superior y una triangular inferior.

4.6 Predictores Lattice

- Para implementar el algoritmo de ortogonalización de Gram-Schmidt es posible usar el banco de filtros de error de predicción backward como mostrado en la figura 11.
- Sin embargo, un método mucho más eficiente de implementar esta ortogonalización es utilizar una estructura recursiva en orden denominada **predictor lattice**.
- Específicamente el predictor lattice consiste en una conexión en cascada de etapas elementales, cada una de las cuales tiene la forma de una lattice. El número de etapas del predictor lattice determina el orden del filtro.

Recursividad para la adaptación en orden

- Las relaciones entrada-salida que caracterizan al predictor lattice pueden obtenerse de varias formas, dependiendo de la forma particular utilizada del algoritmo de Levinson-Durbin.
- Para ello comenzamos por las formas matriciales del filtro de error de predicción forward y backward que se repiten a continuación

$$\mathbf{a}_m = \begin{bmatrix} \mathbf{a}_{m-1} \\ 0 \end{bmatrix} + \kappa_m \begin{bmatrix} 0 \\ \mathbf{a}_{m-1}^{B*} \end{bmatrix} \quad (39)$$

$$\mathbf{a}_m^{B*} = \begin{bmatrix} 0 \\ \mathbf{a}_{m-1}^{B*} \end{bmatrix} + \kappa_m^* \begin{bmatrix} \mathbf{a}_{m-1} \\ 0 \end{bmatrix} \quad (40)$$

- Consideramos $u(n), u(n-1), \dots, u(n-m)$ al filtro de error de predicción forward de orden m , y $\mathbf{u}_{m+1}(n)$ ($m+1 \times 1$) particionado como

$$\mathbf{u}_{m+1}(n) = \begin{bmatrix} \mathbf{u}_m(n) \\ u(n-m) \end{bmatrix} \quad (41)$$

o en forma equivalente

$$\mathbf{u}_{m+1}(n) = \begin{bmatrix} u(n) \\ \mathbf{u}_m(n-1) \end{bmatrix} \quad (42)$$

- Una primera ecuación puede obtenerse premultiplicando $\mathbf{u}_{m+1}(n)$ por $\mathbf{a}_m^H$ definido en la ecuación (39), con lo que se obtiene

$$f_m(n) = f_{m-1}(n) + \kappa_m^* b_{m-1}(n-1)$$

- Por otro lado, una segunda ecuación de recursión puede obtenerse premultiplicando $\mathbf{u}_{m+1}(n)$ por $(\mathbf{a}_m^{BT})^H$ definido en la ecuación (40), con lo que se obtiene

$$b_m(n) = b_{m-1}(n-1) + \kappa_m f_{m-1}(n)$$

- Las ecuaciones obtenidas pueden agruparse matricialmente como

$$\begin{bmatrix} f_m(n) \\ b_m(n) \end{bmatrix} = \begin{bmatrix} 1 & \kappa_m^* \\ \kappa_m & 1 \end{bmatrix} \begin{bmatrix} f_{m-1}(n) \\ b_{m-1}(n-1) \end{bmatrix} \quad m = 1, 2, \dots, M$$

- Para el caso elemental $m = 0$ se tiene que $f_0(n) = b_0(n) = u(n)$.

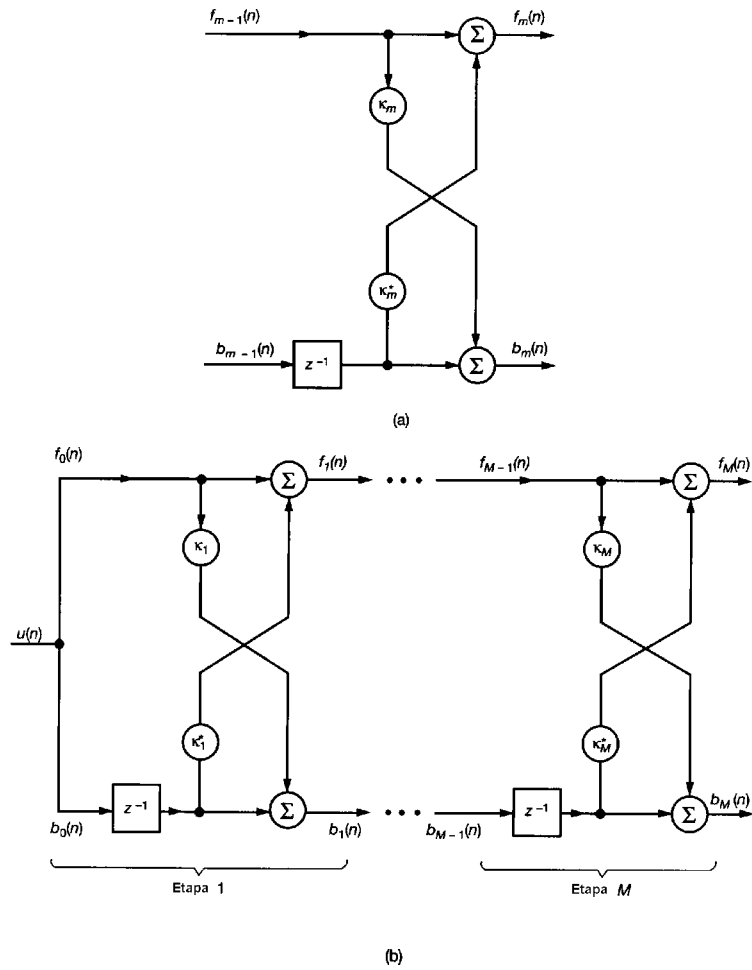


Figura 12: a) Diagrama de la etapa m de un predictor lattice, b) modelo lattice equivalente de un filtro de error de predicción de orden M .

Algunas características

1. Estructura altamente eficiente para generar errores forward y backward simultaneamente.
2. Las diferentes etapas del predictor lattice están desacopladas en el sentido de la ortogonalización de los errores de predicción backward.
3. Es una estructura **modular**, o sea, para incrementar el orden del filtro solo es necesario agregar etapas adicionales sin alterar los cálculos obtenidos.
4. Todas las etapas del predictor lattice tienen la misma estructura.

4.7 Estimación conjunta

- Discutiremos un problema similar al de filtrado de Wiener para estimar una respuesta deseada $d(n)$, excepto que ahora en lugar de utilizar directamente la secuencia de entrada $u(n)$, aplicaremos la estructura del predictor lattice anteriormente discutido y la decorrelación asociada a los errores de predicción backward.

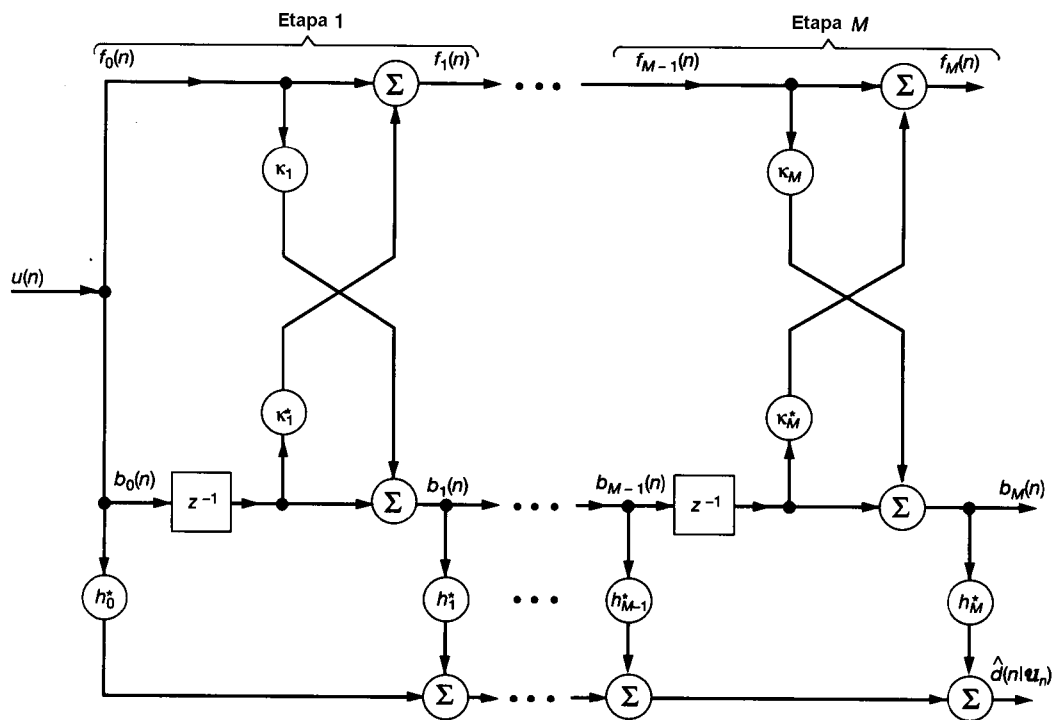


Figura 13: Estructura lattice para el estimador conjunto.

- El estimador está compuesto por un lado por el **predictor lattice** que transforma la secuencia (correlacionada) de entrada $u(n)$ en la secuencia (no correlacionada) de errores de predicción backward.
- Por otro lado, tiene un filtro de **regresión**, caracterizado por los coeficientes $h_0, h_1, \dots, h_M$ que transforma la secuencia de errores de predicción backward en una estimación de la respuesta deseada $d(n)$.

- La estimación de la respuesta puede formularse como

$$\hat{d}(n|U_n) = \sum_{i=0}^M h_i^* b_i(n) = \mathbf{h}^H \mathbf{b}(n)$$

- Si $\mathbf{D}$ ($M + 1 \times M + 1$) es la matriz de correlación de $\mathbf{b}(n)$ y $\mathbf{z}$ es el vector de correlación cruzada entre los errores de predicción backward y la referencia $d(n)$, $\mathbf{z} = E[\mathbf{b}(n)d^*(n)]$, entonces las ecuaciones de Wiener-Hopf para este problema en términos de la solución óptima para $\mathbf{h}$ se pueden escribir como

$$\mathbf{D} \mathbf{h}_0 = \mathbf{z}$$

tal que

$$\mathbf{h}_0 = \mathbf{D}^{-1} \mathbf{z}$$

Relación entre coeficientes de regresión óptimos y la solución de Wiener

- De la factorización de Cholesky $\mathbf{D}^{-1} = \mathbf{L}^{-H} \mathbf{R}^{-1} \mathbf{L}^{-1}$, tal que

$$\mathbf{h}_0 = \mathbf{L}^{-H} \mathbf{R}^{-1} \mathbf{L}^{-1} \mathbf{z}$$

y como $\mathbf{b}(n) = \mathbf{L} \mathbf{u}(n)$, tal que $\mathbf{z} = \mathbf{L} E[\mathbf{u}(n)d^*(n)] = \mathbf{L} \mathbf{p}$, se obtiene

$$\begin{aligned} \mathbf{h}_0 &= \mathbf{L}^{-H} \mathbf{R}^{-1} \mathbf{L}^{-1} \mathbf{L} \mathbf{p} \\ &= \mathbf{L}^{-H} \mathbf{R}^{-1} \mathbf{p} = \mathbf{L}^{-H} \mathbf{w}_o \end{aligned}$$

4.8 Estimación en bloques

- Vimos dos estructuras básicas para construir un **predictor lineal** o su extensión natural el **filtro de error de predicción**: el **filtro transversal** y el **filtro lattice**. La relación entre las dos realizaciones está asociada al algoritmo de Levinson-Durbin.
- Independientemente del interés particular en la estructura elegida es claramente necesario estimar los coeficientes del filtro. Para ello pueden utilizarse dos tipos de procedimientos:
 - Estimación en bloques.
 - Estimación adaptativa.
- En la **estimación en bloques**, los datos disponibles se dividen en bloques cada uno de longitud N . Esta longitud se elige para asegurar la pseudoestacionaridad de los datos. Bajo esta suposición, los coeficientes se calculan por cada bloque de datos. Típicamente los coeficientes del filtro varían de un bloque de datos a otro. Los algoritmos de estimación en bloques pueden categorizarse como
 1. **Métodos indirectos**. Se realiza la estimación de los coeficientes de autocorrelación para después utilizar el algoritmo de Levinson-Durbin y obtener los coeficientes correspondientes.
 2. **Métodos directos**. Para cada bloque de datos se estiman los coeficientes de reflexión de las diferentes etapas del predictor lattice.

- Consideraremos un método de estimación directa de los parámetros del predictor lattice, conocido como **algoritmo de Burg**. En ese caso la idea es utilizar la propiedad de desacoplamiento de la realización lattice para **llevar a cabo la optimización global de todas las etapas como una secuencia de optimizaciones locales, una por cada etapa**.
- Para la etapa m del predictor lattice, dada por

$$\begin{aligned} f_m(n) &= \mathbf{a}_m^H \mathbf{u}_{m+1}(n) \\ b_m(n) &= b_{m-1}(n-1) + \kappa_m f_{m-1}(n) \end{aligned}$$

- La función costo a optimizar localmente para la etapa m se define como

$$J_m = E[|f_m(n)|^2] + E[|b_m(n)|^2]$$

de forma que sustituyendo las ecuaciones de recursión de orden se obtiene

$$\begin{aligned} J_m &= \left(E[|f_m(n)|^2] + E[|b_m(n)|^2] \right) [1 + |\kappa_m|^2] \\ &\quad + 2\kappa_m E[f_{m-1}(n)b_{m-1}^*(n-1)] \\ &\quad + 2\kappa_m^* E[b_{m-1}(n-1)f_{m-1}^*(n)] \end{aligned}$$

entonces, considerando $\kappa_m = \alpha_m + j\beta_m$, y derivando respecto a la función costo J_m se obtiene

$$\begin{aligned} \nabla J_m &= \frac{\partial J_m}{\partial \alpha_m} + j \frac{\partial J_m}{\partial \beta_m} \\ &= 2\kappa_m \left(E[|f_{m-1}(n)|^2] + E[|b_{m-1}(n-1)|^2] \right) \\ &\quad + 4E[b_{m-1}(n-1)f_{m-1}^*(n)] \end{aligned}$$

de donde haciendo el gradiente igual a cero se obtiene

$$\kappa_{m,o} = -\frac{2E[b_{m-1}(n-1)f_{m-1}^*(n)]}{E[|f_m(n)|^2] + E[|b_m(n)|^2]}, \quad m = 1, 2, \dots, M$$

- Esta se conoce como la **fórmula de Burg** y tiene dos propiedades interesantes:

1. El coeficiente de reflexión satisface $|\kappa_{m,o}| \leq 1$, para todo m . o en otras palabras, produce un diseño de fase mínima.
2. El valor medio cuadrático de los errores de predicción forward y backward en la etapa m están relacionados con su entrada por

$$\begin{aligned} E[|f_m(n)|^2] &= (1 - |\kappa_{m,o}|^2)E[|f_{m-1}(n)|^2] \\ E[|b_m(n)|^2] &= (1 - |\kappa_{m,o}|^2)E[|b_{m-1}(n-1)|^2] \end{aligned}$$

- Una forma de estimar las esperanzas estadísticas de la fórmula de Burg es la siguiente

$$\hat{\kappa}_{m,o} = -\frac{2 \sum_{n=m+1}^N b_{m-1}(n-1)f_{m-1}^*(n)}{\sum_{n=m+1}^N [|f_m(n)|^2] + [|b_m(n)|^2]}, \quad m = 1, 2, \dots, M$$

donde N es la longitud del bloque de datos de entrada y $f_0(n) = b_0(n) = u(n)$.

- Con un predictor lattice de m etapas, cada una conteniendo un solo retardo y suponiendo $u(n) = 0$ para $n \leq 0$, todas las muestras de entrada contribuyen a la salida de la etapa m .
- Notar también que $\hat{\kappa}_{m,o}$ depende de la longitud N tal que la elección de N está usualmente determinada por dos factores conflictivos.
 - debe ser suficientemente grande para que los efectos del ruido en la estimación anterior se suavizen,
 - pero por otro lado, debe ser suficientemente pequeña de forma de garantizar la pseudoestacionaridad de los datos de entrada durante el cálculo.
- En general este procedimiento requiere una gran cantidad de cálculos y almacenamiento y además la estimación obtenida en el instante $n+1$ no depende de una forma simple de la estimación previa en n .