

00365



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE MÉXICO

---

---

POSGRADO EN CIENCIAS  
MATEMÁTICAS

FACULTAD DE CIENCIAS

REMUESTREO POR MEDIO DE ONDULETAS

T E S I S

QUE PARA OBTENER EL GRADO ACADÉMICO DE  
MAESTRO EN CIENCIAS  
(MATEMÁTICAS)

P R E S E N T A

CARLOS ERNESTO MARTINEZ RODRIGUEZ

DIRECTOR DE TESIS:  
DR. ALBERTO CONTRERAS CRISTAN

MEXICO, D. F.

2005

M: 350794



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**


Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo ~~recepional~~.

NOMBRE: CARLOS FERNANDO  
MARTINEZ RODRIGUEZ

FECHA: 22 / NOVIEMBRE / 2005

FIRMA: 

Quiero Agradecer a

Yesica y Leonardo por estar conmigo y darle sentido a mi vida.

Mi familia.

La familia de Yesica

Alberto por su enorme paciencia para conmigo y por haberme permitido trabajar con él .

José Guerrero por sus consejos y ayuda siempre oportunos.

Arturo Olvera por ser parte de mi formación personal y profesional.

Antonio, Gaby, Héctor, Irery.

Angélica, Carla y Lars.

La gente del programa de supercomputo de la DGSCA por habernos permitido utilizar su supercomputadora BKLIZ.

Hermínio, Roberto y Carmen.

Especial y principalmente a Dios.

# Índice general

Introducción	ix
<b>1. Series de Tiempo Estacionarias.</b>	<b>1</b>
1.1. Representación Espectral de Procesos Estacionarios . . . . .	5
<b>2. Teoría de Filtros y Onduletas.</b>	<b>19</b>
2.1. Transformada Discreta de Fourier . . . . .	21
2.2. Convolución . . . . .	25
2.3. Teoría de Onduletas. . . . .	35
<b>3. Transformadas Discretas de Onduletas</b>	<b>37</b>
3.1. Transformada Discreta de Fourier . . . . .	37
3.2. Transformada Discreta de Onduletas . . . . .	45
3.2.1. El filtro de Onduletas . . . . .	55
3.2.2. El filtro de Escala . . . . .	65
3.3. Paquetes de Onduletas . . . . .	77
3.4. Algoritmo para la selección de base. . . . .	85
<b>4. Bootstrapping vs. Wavestrapping.</b>	<b>89</b>



---

4.1. Caso Paramétrico . . . . .	89
4.1.1. Bootstrap de residuales . . . . .	89
4.2. Caso No paramétrico . . . . .	91
4.2.1. Bootstrap por Bloques . . . . .	92
4.2.2. Bootstrap por medio de Onduletas . . . . .	93
4.3. Ejemplos numéricos . . . . .	95
4.3.1. Caso Paramétrico . . . . .	97
4.3.2. Caso No Paramétrico, (Bootstrap vía Onduletas) . . . . .	100
4.4. Resultados . . . . .	101
4.4.1. Caso $N = 64$ . . . . .	102
4.4.2. Caso $N = 128$ . . . . .	109
4.4.3. Caso $n = 256$ . . . . .	117
<b>5. Conclusiones</b>	<b>125</b>
<b>A. Apéndice A</b>	<b>133</b>
A.1. Teorema de Kolmogorov . . . . .	133
A.2. Teoremas de Helly y Helly-Bray . . . . .	133

# Índice de figuras

3.1. Gráfica de la función de ganancia cuadrada . . . . .	68
3.2. Gráfica de las funciones de Ganancia cuadrada para la cascada de filtros utilizando el filtro AL(8) con $N = 128 = 2^7$ . . . . .	75
3.3. Ilustración de la partición $DWT$ para $N = 16 = 2^4$ . . . . .	78
3.4. Ilustración de la partición $\mathcal{P}_2 = \{\mathcal{I}_{2,0}, \mathcal{I}_{2,1}, \mathcal{I}_{2,2}, \mathcal{I}_{2,3}\}$ . . . . .	79
3.5. Sucesiones binarias para determinar los índices de la sucesión ordenada de los filtros de onduletas . . . . .	83
3.6. Esquema de una Wavelet Packet Table para $j = 3$ . . . . .	84
4.1. Gráfica de los ordenes de los modelos tipo Autoregresivo ajustados a un proceso proveniente de un modelo autoregresivo, método paramétrico, ( $N=64$ ). . . . .	103
4.2. Gráfica de los promedios y desviaciones estándar de los coeficientes de correlación parcial al lag 1, método paramétrico, ( $N=64$ ). . . . .	104
4.3. Gráfica de los promedios de los coeficientes de correlación parcial al lag 1, método no paramétrico, ( $N=64$ ) . . . . .	104
4.4. Gráfica de las desviaciones estándar de los coeficientes de correlación parcial al lag 1, método no paramétrico, ( $N=64$ ) . . . . .	105
4.5. Gráfica de los ordenes de los modelos tipo Autoregresivo ajustados a un proceso proveniente de un modelo de promedios móviles, método paramétrico, ( $N=64$ ). . . . .	106

4.6. Gráfica de los promedios y las desviaciones estándar de los coeficientes de correlación parcial al lag 1 para procesos de tipo  $MA(1)$ , método paramétrico, (N=64). 107

4.7. Gráfica de los promedios y las desviaciones estándar junto con los valores asintóticos estimados tanto para la media, como para la desviación estándar para el caso en que el proceso original es de tipo  $MA(1)$ , método no paramétrico, (N=64). . . . 108

4.8. Gráfica de los ordenes de los modelos tipo Autoregresivo ajustados a un proceso proveniente de un modelo autoregresivo, método paramétrico, (N=128). . . . . 110

4.9. Gráfica de los promedios y desviaciones estándar de los coeficientes de correlación parcial al lag 1, método paramétrico, (N=128). . . . . 111

4.10. Gráfica de los promedios y las desviaciones estándar junto con los valores asintóticos estimados tanto para la media, como para la desviación estándar para el caso en que el proceso original es de tipo  $AR(1)$ , método no paramétrico, (N=128). . . . 112

4.11. Gráfica de los ordenes de los modelos tipo Autoregresivo ajustados a un proceso proveniente de un modelo de promedios móviles, método paramétrico, (N=128). 114

4.12. Gráfica de los promedios y las desviaciones estándar de los coeficientes de correlación parcial al lag 1 para procesos de tipo  $MA(1)$ , método paramétrico, (N=128). . . . . 114

4.13. Gráfica de los promedios y las desviaciones estándar junto con los valores asintóticos estimados tanto para la media, como para la desviación estándar para el caso en que el proceso original es de tipo  $MA(1)$ , método no paramétrico, (N=128). . . 115

4.14. Gráfica de los ordenes de los modelos tipo Autoregresivo ajustados a un proceso proveniente de un modelo autoregresivo, método paramétrico, (N=256). . . . . 117

4.15. Gráfica de los promedios y desviaciones estándar de los coeficientes de correlación parcial al lag 1, método paramétrico, (N=256). . . . . 118

4.16. Gráfica de los promedios y las desviaciones estándar junto con los valores asintóticos estimados tanto para la media, como para la desviación estándar para el caso en que el proceso original es de tipo  $AR(1)$ , método no paramétrico, (N=256). . . . 119

4.17. Gráfica de los ordenes de los modelos tipo Autoregresivo ajustados a un proceso proveniente de un modelo de promedios móviles, método paramétrico, (N=128). 121

---

4.18. Gráfica de los promedios y las desviaciones estándar de los coeficientes de correlación parcial al lag 1 para procesos de tipo $MA(1)$ , método paramétrico, ( $N=256$ ) . . . . .	121
4.19. Gráfica de los promedios y las desviaciones estándar junto con los valores asintóticos estimados tanto para la media, como para la desviación estándar para el caso en que el proceso original es de tipo $MA(1)$ , método no paramétrico, ( $N=256$ ). . .	122
5.1. Gráfica de los promedios de los coeficientes de correlación entre dos muestras consecutivas para el estimador número 28 del valor asintótico de la media de los coeficientes de autocorrelación parcial a lag 1. . . . .	130



# Índice de cuadros

4.1. Caso $n=64$ , para un proceso Autoregresivo de orden 1, el porcentaje de veces que se eligió el modelo $AR(1)$ fue del 84 % . . . . .	105
4.2. Caso $n=64$ , en el que el proceso original es un $MA(1)$ , el porcentaje de veces que se eligió el modelo $AR(1)$ fue del 28 % . . . . .	108
4.3. Caso $n=128$ , para un proceso Autoregresivo de orden 1, el porcentaje de veces que se eligió el modelo $AR(1)$ fue del 76 % . . . . .	112
4.4. Caso $n=128$ , para un proceso de tipo promedios móviles de orden 1, el porcentaje de veces que se eligió el modelo $AR(1)$ fue de 8 % . . . . .	116
4.5. Caso $n=256$ , para un proceso Autoregresivo de orden 1, el porcentaje de veces que se eligió el modelo $AR(1)$ fue del 86 % . . . . .	120
4.6. Caso $n=256$ , para un proceso de tipo promedios móviles de orden 1, El porcentaje de veces que se eligió el modelo $AR(1)$ fue de 2 % . . . . .	123
5.1. Caso en que el proceso es de tipo autoregresivo, se muestra el porcentaje en que ambos métodos sobreestiman (subestiman) el valor verdadero de la media o de la desviación estándar del coeficiente de autocorrelación parcial a lag 1, considerando como estimador el promedio de los estimadores generados por ambos métodos. . . . .	126
5.2. Caso en que el proceso es de tipo autoregresivo, se muestra el porcentaje en que ambos métodos sobreestiman (subestiman) el valor verdadero de la media o de la desviación estándar del coeficiente de autocorrelación parcial a lag 1, considerando como estimador la mediana de los estimadores generados por ambos métodos. . . . .	127

- 5.3. Caso en que el proceso es de tipo de promedios móviles, se muestra el porcentaje en que ambos métodos sobreestiman (ó subestiman) el valor asintótico de la media y de la desviación estándar del coeficiente de autocorrelación parcial a lag 1, considerando como estimador el promedio de los estimadores generados por ambos métodos. . . . . 128
  
- 5.4. Caso en que el proceso original es de tipo  $MA(1)$ , en este cuadro se está mostrando el porcentaje en que ambos métodos sobreestiman (subestiman) el valor asintótico de la media o de la desviación estándar del coeficiente de autocorrelación parcial a lag 1, considerando como estimador el promedio de los estimadores generados por ambos métodos. . . . . 128

# Introducción

El presente trabajo está basado principalmente en el artículo *Nonlinear and Nonstationary Signal Processing* publicado por Percival, Sardy y Davison (2000), en el libro *Wavelet Methods for Time Series Analysis* de Percival y Walden y en la tesis doctoral *Advances in the analysis of time series using wavelets* de Contreras-Cristán.

La idea central del trabajo es abordar el problema de obtener un estimador confiable para el coeficiente de correlación a lag 1 asociado a una serie de tiempo dada. Nosotros nos enfocamos a revisar dos métodos de remuestreo para abordar este problema. El primer método revisado es un método paramétrico. El segundo método que se revisó es no paramétrico.

Como se plantea en Percival, Sardy y Davison (2000), la comparación de los métodos se hizo en el contexto de modelos lineales de series de tiempo,  $AR(1)$  y  $MA(1)$ .

De la literatura es sabido que la inferencia estadística (en un sentido clásico) de estos modelos requiere del supuesto de estacionariedad del modelo, por ejemplo en el caso  $AR(1)$  la condición de estacionariedad es  $|\phi_1| < 1$ . Por estas razones el método de remuestreo paramétrico está pensado para el caso estacionario, de hecho el método requiere de estimar modelos  $AR(p)$  ó  $MA(q)$ , en forma iterativa y los estimadores de máxima verosimilitud para estos modelos están bien determinados en el caso estacionario.

Por otro lado, el método de remuestreo no paramétrico que estudiamos no requiere de estimar parámetros del modelo, por tanto es posible ubicarlo aún en contextos de series de tiempo no estacionarias. Este método de remuestreo está basado en el uso de la transformada discreta de onduletas (DWT, del inglés Discrete Wavelets Transform). Más específicamente, se propone el uso de una generalización de la DWT conocida como la Transformada Discreta de Paquetes de Onduletas (DWPT, del inglés, Discrete Wavelets Paquets Transform).

Una vez realizada la revisión, se procedió a realizar un ejercicio numérico con la finalidad de comparar los estimadores que generan ambos métodos. El ejercicio consistió en ambos casos en realizar remuestreo ajustando un modelo específico que satisfaga ciertos criterios referentes a



la selección del modelo que mejor ajusta a los datos. Una vez realizado este ajuste se procede a calcular los residuales y utilizando un esquema de remuestreo aleatorio con reemplazo se generan muestras aleatorias (muestras *bootstrap*) de los residuales.

A continuación estas muestras fueron usadas como ruido blanco en el modelo propuesto para así generar nuevas series de datos (muestras *bootstrap* de los residuales) de los cuales se obtendrá el estimador  $\hat{\phi}_1$  del parámetro  $\phi_1$ . Como al final se generan un número  $m$  de muestras bootstrap del modelo propuesto, podemos calcular la estimación  $\hat{\rho}_1$  del parámetro  $\rho_1$  para cada muestra  $l$ ,  $l = 1, 2, \dots, m$ .

Entonces con la muestra  $\hat{\rho}_1^{(1)}, \hat{\rho}_1^{(2)}, \dots, \hat{\rho}_1^{(m)}$  podemos estudiar propiedades distribucionales de la variable aleatoria  $\hat{\rho}_1$ . Este es el caso del método paramétrico, ver Davidson y Hinkley (1997).

Para el caso del método no paramétrico, la idea detrás del bootstrap basado en la DWT es utilizar el hecho de que para procesos Gaussianos estacionarios fraccionalmente diferenciados (Stationary Gaussian Fractionally differenced (FD) processes)  $\{X_t\}$ , la DWT es una transformación que *descorrelaciona* a  $\{X_t\}$ , es decir, mientras la serie de tiempo original puede presentar un alto grado de correlación, los coeficientes de la DWT pueden ser considerados como una serie de tiempo con elementos no correlacionados, vease Percival y Constantine (2002), Percival, Sardy y Davidson (2000), y Percival y Walden (2000). La DWPT nos ayuda a aplicar esta idea para procesos  $\{X_t\}$  de segundo orden en general.

Los ejemplos numéricos se realizaron con tres tamaños distintos de muestra, a saber, 64, 128 y 256. Se estudió el caso en el que la longitud del proceso considerado era de 64 dado que para tamaños de muestra pequeños, como el que nosotros elegimos, hay posibilidades de equivocarse al momento de elegir el modelo apropiado además del orden del mismo cuando se usa el método paramétrico. Aunque no se reportan los resultados para tamaños de muestra mayores de 256 estos se comportan de manera muy similar al caso  $n = 256$  es por esto que se considero que solamente se reportaran los resultados obtenidos para tamaños de muestra de 64, 128 y 256.

Para poder realizar la comparación entre ambos métodos, se procedió a calcular la distribución asintótica del coeficiente de correlación parcial a lag 1. En el artículo publicado por Percival, Sardy y Davidson (2000), se menciona que el bootstrap no paramétrico vía onduletas no es adecuado para estimar la desviación estándar de  $\rho_1$  en comparación con el bootstrap por bloques, y nuevamente es adecuado cuando se compara con el bootstrap paramétrico en los residuales. Además, en el mismo artículo mencionan que el bootstrap no paramétrico vía onduletas no es apropiado para estimar la media de  $\rho_1$ .

Sin embargo nuestras simulaciones nos permiten dudar de esta afirmación. Una conclusión derivada de esta tesis, es que el método paramétrico tiende a generar muestras bootstrap que están más correlacionadas entre sí. Por otra parte, el método no paramétrico tiende a generar muestras bootstrap con menor correlación entre sí. Por las razones anteriores, la convergencia a valores asintóticos de los valores estimados al usar el método no paramétrica es más rápida. En ambas situaciones, tanto cuando se especifica correctamente un modelo para los datos, como cuando no se especifica adecuadamente un modelo (es decir, se comete un error en la especificación del modelo), el método paramétrico tiende a generar muestras bootstrap más correlacionadas entre sí. Es por esta situación que el método no paramétrico puede resultar más eficiente para estimar la media de  $\rho_1$  inclusive cuando el modelo está bien especificado.



# Capítulo 1

## Series de Tiempo Estacionarias.

Un proceso estocástico es una familia de variables aleatorias  $\{X_t, t \in T\}$ , definidas en un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$  donde  $T$  es un conjunto de índices. En el contexto de series de tiempo el conjunto de índices  $T$  considerado usualmente es un conjunto discreto, por tanto en esta tesis, trabajaremos con procesos a tiempo discreto y más específicamente con  $T \subseteq \mathbb{Z}$ . Así entonces, para cada  $t \in T$  fijo se tiene la variable aleatoria

$$X_t : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})),$$

mientras que para cada  $\omega \in \Omega$  fijo tenemos la sucesión en  $\mathbb{R}$  dada por  $X_{t_1}(\omega), X_{t_2}(\omega), \dots$

**Definición 1.1.** Si  $\{X_t, t \in T\}$  es un proceso tal que  $\mathbb{E}(X_t^2) < \infty$  para cada  $t \in T$ , entonces la función de autocovarianza  $\gamma_X(\cdot, \cdot)$  de  $X_t$  se define por

$$\gamma_X(r, s) = \text{cov}(X_r, X_s) = \mathbb{E}[(X_r - \mathbb{E}(X_r))(X_s - \mathbb{E}(X_s))],$$

para cada  $r, s \in T$ .

Dado que el conjunto de índices  $T$  está representando los tiempos en que se observa la colección de variables aleatorias  $\{X_t : t \in T\}$  nos referiremos indistintamente a esta colección como serie de tiempo ó proceso estacionario.

**Definición 1.2.** Sea  $\{X_t, t \in T\}$  una serie de tiempo con conjunto de índices  $T = \{\dots, -1, 0, 1, \dots\}$ , decimos que  $\{X_t, t \in T\}$  es estacionaria de segundo orden si

(i)  $\mathbb{E}(|X_t|^2) < \infty$  para todo  $t \in T$ .

(ii)  $\mathbb{E}(X_t) = \mu$  para todo  $t \in T$ .

(iii)  $\gamma_X(r, s) = \gamma_X(r + t, s + t)$  para toda  $r, s, t \in T$ .

**Nota 1.1.** Si  $\{X_t, t \in T\}$  es estacionario de segundo orden, por (iii) si tomamos a  $t$  como  $-s$ , entonces obtenemos

$$\gamma_X(r, s) = \gamma_X(r - s, 0)$$

para toda  $r, s \in T$ . Análogamente si  $t = -r$ ,

$$\gamma_X(r, s) = \gamma_X(0, s - r).$$

Es decir, si  $\{X_t, t \in T\}$  es estacionario de segundo orden, su función de autocovarianza,  $\gamma_X(t, s)$  sólo depende de  $t$  y  $s$  a través de  $h = t - s$ .

Por lo tanto, si  $\{X_t, t \in T\}$  es un proceso estacionario de segundo orden, podemos escribir su función de autocovarianza como una función que depende de un sólo parámetro  $h \in \mathbb{Z}$

$$\gamma_X(h) \equiv \gamma_X(h, 0) = \text{cov}(X_{t+h}, X_t),$$

para toda  $t, h \in \mathbb{Z}$ .

A lo largo de este trabajo la frase:  $\{X_t, t \in T\}$  es un proceso estacionario, se usará para referirnos a que el proceso  $\{X_t, t \in T\}$  es estacionario de segundo orden. La función de autocovarianza de un proceso estacionario satisface tres propiedades importantes que la caracterizan:

**Proposición 1.1.** Si  $\gamma_X(\cdot)$  es la función de autocovarianza de un proceso estacionario  $\{X_t, t \in T\}$ , entonces

$$(i) \gamma_X(0) \geq 0$$

$$(ii) |\gamma_X(h)| \leq \gamma_X(0)$$

$$(iii) \gamma(h) = \gamma(-h)$$

*Demostración.* A saber

$$\gamma_X(0) = \text{cov}(X_t, X_t) = \text{var}(X_t) \geq 0.$$

Además, por la desigualdad de *Cauchy-Schwarz* y dado que el proceso es estacionario se tiene lo siguiente:

$$|\gamma_X(h)| = |\text{cov}(X_{t+h}, X_t)| \leq \text{var}(X_{t+h})^{1/2} \text{var}(X_t)^{1/2} = \gamma_X(0) = \text{cov}(X_t, X_t).$$

Finalmente

$$\gamma_X(-h) = \text{cov}(X_{t-h}, X_t) = \text{cov}(X_t, X_{t+h}) = \gamma_X(h)$$

□

**Definición 1.3.** Una función real definida en los enteros,  $\kappa : \mathbb{Z} \rightarrow \mathbb{R}$ , se dice que es no negativa definida si y sólo si

$$\sum_{i=1}^n \sum_{j=1}^n a_i \kappa(t_i - t_j) a_j \geq 0$$

para todo entero positivo  $n$ , para todo vector  $\mathbf{a} = (a_1, a_2, \dots, a_n)^T \in \mathbb{R}^n$  y  $\mathbf{t} = (t_1, t_2, \dots, t_n)^T \in \mathbb{Z}^n$ .

**Nota 1.2.** Una matriz simétrica  $\mathbf{K} = \{k_{ij}\}_{i,j=1}^n$  es no negativa definida si

$$\sum_{i=1}^n \sum_{j=1}^n a_i k_{ij} a_j = \mathbf{a}^T \mathbf{K} \mathbf{a} \geq 0$$

para toda  $\mathbf{a} \in \mathbb{R}^n$ .

Un teorema importante que caracteriza a la covarianza es el siguiente:

**Teorema 1.1.** Una función real definida en los enteros es la función de autocovarianza de una serie de tiempo estacionaria si y sólo si es par y no negativa definida.

*Demostración.* Primero probaremos que la función de autocovarianza  $\gamma_X$  de un proceso estacionario  $\{X_t, t \in T\}$  es no negativa definida.

Sean  $\mathbf{a} = (a_1, a_2, \dots, a_n)^T \in \mathbb{R}^n$  y  $\mathbf{t} = (t_1, t_2, \dots, t_n)^T \in \mathbb{Z}^n$ ; sea

$$\mathbf{Z}_t = (X_{t_1} - \mathbb{E}(X_{t_1}), X_{t_2} - \mathbb{E}(X_{t_2}), \dots, X_{t_n} - \mathbb{E}(X_{t_n}))^T,$$

entonces se tiene que

$$0 \leq \text{var}(\mathbf{a}^T \mathbb{Z}_t) = \mathbf{a}^T \mathbb{E}[\mathbb{Z}_t \mathbb{Z}_t^T] \mathbf{a} = \mathbf{a}^T \Gamma_n \mathbf{a}, \quad (1.1)$$

donde  $\Gamma_n = [\gamma_X(t_i - t_j)]_{i,j=1}^n$  es la matriz de varianzas y covarianzas del vector  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})^T$ ; por lo tanto la varianza se puede reescribir como:

$$\text{var}(\mathbf{a}^T \mathbb{Z}_t) = \sum_{i=1}^n \sum_{j=1}^n a_i \gamma_X(t_i - t_j) a_j.$$

Además por la proposición (1.1 (iii))  $\gamma_X$  es par.

Conversamente, sea  $k : \mathbb{Z} \rightarrow \mathbb{R}$  función par y no negativa definida. Sean  $n \in \mathbb{N}$ ,  $\mathbf{t} \in \mathbb{Z}^n$ ,  $\mathbf{t} = (t_1, t_2, \dots, t_n)$  donde  $t_1 < t_2 < \dots < t_n$  y  $F_{\mathbf{t}}$  la función de distribución en  $\mathbb{R}^n$  cuya función característica es de la forma

$$\Phi_{\mathbf{t}}(\mathbf{u}) = \exp(-\mathbf{u}^T \mathbf{K} \mathbf{u} / 2),$$

donde  $\mathbf{u} = (u_1, u_2, \dots, u_n)^T \in \mathbb{R}^n$  y  $\mathbf{K} = [k(t_i - t_j)]_{i,j=1}^n$ . Como  $k$  es no negativa definida, entonces la matriz  $\mathbf{K}$  también lo es y por tanto  $\Phi_{\mathbf{t}}$  es la función característica de una distribución normal *n-variada* con media cero y matriz de covarianza  $\mathbf{K}$ . Para un vector  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T \in \mathbb{R}^n$  denotemos por  $\mathbf{y}(i)$  al vector  $n-1$  dimensional dado por  $\mathbf{y}(i) = (y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_n)^T$ , entonces

$$\lim_{u_i \rightarrow 0} \Phi_{\mathbf{t}}(\mathbf{u}) = e^{-(\mathbf{u}(i)^T \mathbf{K}(i) \mathbf{u}(i)) / 2} = \Phi_{\mathbf{t}(i)}(\mathbf{u}(i)),$$

donde  $\mathbf{K}(i)$  es la matriz de dimensión  $(n-1) \times (n-1)$  obtenida de la eliminación del *i-ésimo* renglón y de la *i-ésima* columna en la matriz  $\mathbf{K}$ . Se concluye que las distribuciones  $F_{\mathbf{t}}$  son un sistema consistente de distribuciones de dimensión finita, entonces por el teorema de *Kolmogorov*, (Teorema (A.1)), se tiene que existe una serie de tiempo  $\{X_t, t \in T\}$  con distribuciones de dimensión finita  $F_{\mathbf{t}}$  y correspondientes funciones características  $\Phi_{\mathbf{t}}$  para cada  $n \in \mathbb{N}$ . En particular, la función de distribución conjunta de  $X_i$  y  $X_j$  es normal bivariada con  $\mu = 0$  y matriz de varianzas y covarianzas

$$\begin{pmatrix} k(0) & k(i-j) \\ k(i-j) & k(0) \end{pmatrix},$$

es decir,  $\text{cov}(X_i, X_j) = k(i-j)$ . □

**Definición 1.4.** El proceso  $\{X_t, t \in T\}$  se dice que es ruido blanco con media 0 y varianza  $\sigma^2$  si y sólo si para toda  $t \in T$ ,  $\tau \neq 0$ ,  $\tau \in \mathbb{Z}$ , se tiene que  $\mathbb{E}[X_t] = 0$ ,  $\text{Cov}(X_t, X_{t+\tau}) = 0$  y  $\text{Cov}(X_t, X_t) = \sigma^2$ .

## 1.1. Representación Espectral de Procesos Estacionarios

La representación espectral de un proceso estacionario  $\{X_t, t \in T\}$  consiste en la descomposición del mismo en la suma de componentes sinusoidales con coeficientes aleatorios no correlacionados. A la par de esta descomposición, se tiene análogamente una para la función de autocovarianza de  $\{X_t, t \in T\}$ . El análisis de procesos estacionarios por medio de su representación espectral se conoce como análisis en el *dominio de frecuencias* de la serie de tiempo.

**Nota 1.3.** Al conjugado del número complejo  $z$  se le denotará por  $\bar{z}$ .

**Definición 1.5.** Sea  $\{X_t, t \in T\}$  un proceso con valores en los números complejos, decimos que  $\{X_t, t \in T\}$  es un proceso estacionario de segundo orden si  $\mathbb{E}(|X_t|^2) < \infty$ ,  $\mathbb{E}(X_t)$  es independiente de  $t$  y además  $\mathbb{E}(X_{t+h}X_t)$  es independiente de  $t$ .

**Definición 1.6.** La función de autocovarianza  $\gamma(\cdot)$  de un proceso estacionario de segundo orden con valores en los complejos  $\{X_t, t \in T\}$  se define como

$$\gamma(h) = \mathbb{E}(X_{t+h}\overline{X_t}) - \mathbb{E}(X_{t+h})\mathbb{E}(\overline{X_t}).$$

Luego entonces

$$\gamma(0) = \mathbb{E}[|X_t - \mathbb{E}(X_t)|^2] = \text{var}(X_t)$$

es real.

Las propiedades de la función de autocovarianza dadas anteriormente se pueden reenunciar en términos de funciones de autocovarianza de valores complejos como:

- (i)  $\gamma_X(0) \geq 0$ ,
- (ii)  $|\gamma_X(h)| \leq \gamma_X(0)$ ,
- (iii)  $\gamma_X(h) = \overline{\gamma_X(-h)}$ .

Entonces el teorema de caracterización para las funciones de autocovarianza se puede reenunciar de la siguiente manera:

**Teorema 1.2.** Una función  $K(\cdot)$  definida sobre los enteros y con valores en los números complejos es la función de autocovarianza de una serie de tiempo estacionaria, probablemente con valores complejos, si y sólo si  $K(\cdot)$  es Hermitiana y no negativa definida, es decir, si y sólo si  $K(n) = \overline{K(-n)}$  y además



$$\sum_{i=1}^n \sum_{j=1}^n a_i K(i-j) \bar{a}_j \geq 0 \quad (1.2)$$

para todos los enteros positivos  $n$  y todos los vectores  $\mathbf{a} = (a_1, a_2, \dots, a_n)^T \in \mathbb{C}^n$ .

*Demostración.* Primero demostremos la necesidad

Supongamos que  $K$  es la función de autocovarianza de una serie de tiempo estacionaria,  $\{X_t, t \in T\}$ , entonces hay que demostrar que  $K(h) = \overline{K(-h)}$ . Recordemos que

$$K(h) = \mathbb{E}[X_{t+h}\overline{X_t}] - \mathbb{E}[X_{t+h}]\mathbb{E}[\overline{X_t}],$$

por otra parte,

$$\overline{K(-h)} = \mathbb{E}[\overline{X_{t-h}X_t}] - \mathbb{E}[\overline{X_{t-h}}]\mathbb{E}[X_t] \quad (1.3)$$

además tenemos que

$$K(-h) = \mathbb{E}[X_{t-h}\overline{X_t}] - \mathbb{E}[X_{t-h}]\mathbb{E}[\overline{X_t}]$$

y esta expresión no depende del valor de  $t$ . Luego podemos substituir  $t = \tau + h$ , de donde

$$K(-h) = \mathbb{E}[X_\tau\overline{X_{\tau+h}}] - \mathbb{E}[X_\tau]\mathbb{E}[\overline{X_{\tau+h}}],$$

de la expresión (1.3) y del hecho de que la función  $K$  no depende de  $t$  obtenemos

$$\overline{K(-h)} = \mathbb{E}[X_{\tau+h}\overline{X_\tau}] - \mathbb{E}[X_{\tau+h}]\mathbb{E}[\overline{X_\tau}] = K(h).$$

El hecho de que  $K$  sea no negativa definida se sigue de aplicar un argumento como en la ecuación (1.1). En este caso se tendría que

$$0 \leq \text{var}(a^T \mathbf{Z}) = \mathbb{E}[a^T \mathbf{Z} \overline{\mathbf{Z}^T a}] = a^T \mathbb{E}[\mathbf{Z} \overline{\mathbf{Z}^T}] a = \sum_{i=1}^n \sum_{j=1}^n a_i K(i, j) \bar{a}_j$$

Ahora probemos que para toda  $n \in \mathbb{N}$  y  $\mathbf{a} \in \mathbb{C}^n$ ,

$$\sum_{i,j=1}^n a_i k(i-j) \overline{a_j} \geq 0.$$

Sea  $K(h) = K_1(h) + iK_2(h)$  función hermitiana y no negativa definida, y sea  $L^{(n)}$  matriz de  $2n \times 2n$  definida por

$$L^{(n)} = \frac{1}{2} \begin{bmatrix} K_1^{(n)} & K_2^{(n)} \\ -K_2^{(n)} & K_1^{(n)} \end{bmatrix} \quad (1.4)$$

donde  $K_1^{(n)} = K_1(i-j)$  y  $K_2^{(n)} = K_2(i-j)$  para  $i, j = 1, 2, \dots, n$ . Entonces  $L^{(n)}$  es una matriz real, simétrica y no negativa definida.

Dado que  $K_1(h)$  y  $K_2(h)$  son la parte real e imaginaria respectivamente de  $K(h)$ , entonces  $K_1^{(n)}$  y  $K_2^{(n)}$  son funciones reales, por lo tanto efectivamente  $L^{(n)}$  es una matriz real.

Ahora, dado que  $K$  es una función hermitiana se tiene que

$$K(h) = \overline{K(-h)}$$

por lo tanto

$$K_1(h) + iK_2(h) = \overline{K_1(-h) + iK_2(-h)} = \overline{K_1(-h)} - \overline{iK_2(-h)} = K_1(-h) - iK_2(-h)$$

entonces tenemos que

$$K_1(h) = K_1(-h) \quad (1.5)$$

y además

$$K_2(h) = -K_2(-h) \quad (1.6)$$

Entonces de la ecuación (1.5) se tiene que  $K_1^{(n)}(i-j) = K_1^{(n)}(j-i)$  es decir,  $K_1^{(n)}$  es una matriz simétrica, además de la ecuación (1.6) se tiene que  $K_2^{(n)}(i-j) = -K_2^{(n)}(j-i)$ , por lo tanto  $K_2$  es una matriz antisimétrica y además con ceros en la diagonal principal.

Por lo tanto se tiene que la matriz  $L^{(n)}$  definida en (1.4) es real y simétrica, ahora demostremos que es no negativa definida:

Sean  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ , entonces verifiquemos que se satisface la ecuacion (1.7)

$$\begin{bmatrix} \mathbf{a}^T & \mathbf{b}^T \end{bmatrix} L^{(n)} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \geq 0 \tag{1.7}$$

A saber

$$\begin{aligned} \begin{bmatrix} \mathbf{a}^T & \mathbf{b}^T \end{bmatrix} L^{(n)} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} &= \begin{bmatrix} \mathbf{a}^T & \mathbf{b}^T \end{bmatrix} \left( \frac{1}{2} \begin{bmatrix} K_1^{(n)} & K_2^{(n)} \\ -K_2^{(n)} & K_1^{(n)} \end{bmatrix} \right) \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} \mathbf{a}^T & \mathbf{b}^T \end{bmatrix} \begin{bmatrix} K_1^{(n)} & K_2^{(n)} \\ -K_2^{(n)} & K_1^{(n)} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} \mathbf{a}^T K_1^{(n)} - \mathbf{b}^T K_2^{(n)} & \mathbf{a}^T K_2^{(n)} + \mathbf{b}^T K_1^{(n)} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \\ &= \frac{1}{2} \left( \mathbf{a}^T K_1^{(n)} \mathbf{a} - \mathbf{b}^T K_2^{(n)} \mathbf{a} + \mathbf{a}^T K_2^{(n)} \mathbf{b} + \mathbf{b}^T K_1^{(n)} \mathbf{b} \right) \end{aligned}$$

es decir,

$$\begin{bmatrix} \mathbf{a}^T & \mathbf{b}^T \end{bmatrix} L^{(n)} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \frac{1}{2} \left( \mathbf{a}^T K_1^{(n)} \mathbf{a} + 2\mathbf{a}^T K_2^{(n)} \mathbf{b} + \mathbf{b}^T K_1^{(n)} \mathbf{b} \right). \tag{1.8}$$

Consideremos a una matriz  $\mathcal{M}$ , cuyas entradas pueden ser números complejos. Denotaremos la operación transpuesta *Hermitiana* aplicada a  $\mathcal{M}$ , la cual consiste en conjugar cada elemento de la transpuesta de la matriz  $\mathcal{M}$ , por  $\mathbb{H}$ .

Por ejemplo para

$$\mathcal{M} = \begin{bmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \end{bmatrix}$$

tenemos que

$$\mathcal{M}^{\Re} = \begin{bmatrix} \overline{m_{1,1}} & \overline{m_{2,1}} \\ \overline{m_{1,2}} & \overline{m_{2,2}} \end{bmatrix}.$$

Entonces de acuerdo a Strang (1985),  $(\mathbf{a} + i\mathbf{b})^H (K_1^{(n)} + iK_2^{(n)}) (\mathbf{a} + i\mathbf{b})$  es real, es decir

$$\begin{aligned} (\mathbf{a} + i\mathbf{b})^H (K_1^{(n)} + iK_2^{(n)}) (\mathbf{a} + i\mathbf{b}) &= \overline{(\mathbf{a} + i\mathbf{b})}^T (K_1^{(n)} + iK_2^{(n)}) (\mathbf{a} + i\mathbf{b}) \\ &= (\mathbf{a} - i\mathbf{b})^T (K_1^{(n)} + iK_2^{(n)}) (\mathbf{a} + i\mathbf{b}) = (\mathbf{a}^T - i\mathbf{b}^T) (K_1^{(n)} + iK_2^{(n)}) (\mathbf{a} + i\mathbf{b}) \\ &= (\mathbf{a}^T K_1^{(n)} \mathbf{a} + \mathbf{b}^T K_2^{(n)} \mathbf{a} - \mathbf{a}^T K_2^{(n)} \mathbf{b} + \mathbf{b}^T K_1^{(n)} \mathbf{b}) + i (\mathbf{a}^T K_1^{(n)} \mathbf{b} + \mathbf{b}^T K_2^{(n)} \mathbf{b} + \mathbf{a}^T K_2^{(n)} \mathbf{a} - \mathbf{b}^T K_1^{(n)} \mathbf{a}) \\ &= \mathbf{a}^T K_1^{(n)} \mathbf{a} + \mathbf{b}^T K_2^{(n)} \mathbf{a} - \mathbf{a}^T K_2^{(n)} \mathbf{b} + \mathbf{b}^T K_1^{(n)} \mathbf{b} = \mathbf{a}^T K_1^{(n)} \mathbf{a} - \mathbf{b}^T K_2^{(n)} \mathbf{a} + \mathbf{a}^T K_2^{(n)} \mathbf{b} + \mathbf{b}^T K_1^{(n)} \mathbf{b} \end{aligned}$$

por ser  $\mathbf{a}^T K_2^{(n)} \mathbf{b}$  y  $\mathbf{a}^T K_1^{(n)} \mathbf{b}$  reales, entonces lo que se tiene es

$$(\mathbf{a} + i\mathbf{b})^H (K_1^{(n)} + iK_2^{(n)}) (\mathbf{a} + i\mathbf{b}) = \mathbf{a}^T K_1^{(n)} \mathbf{a} + 2\mathbf{a}^T K_2^{(n)} \mathbf{b} + \mathbf{b}^T K_1^{(n)} \mathbf{b} \quad (1.9)$$

Por lo tanto, de las ecuaciones (1.8) y (1.9) se tiene que

$$\begin{bmatrix} \mathbf{a}^T & \mathbf{b}^T \end{bmatrix} L^{(n)} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \frac{1}{2} (\mathbf{a} + i\mathbf{b})^H (K_1^{(n)} + iK_2^{(n)}) (\mathbf{a} + i\mathbf{b}) = \frac{1}{2} (\mathbf{a} + i\mathbf{b})^H K (\mathbf{a} + i\mathbf{b}) \geq 0 \quad (1.10)$$

Por ser  $K$  función Hermitiana y positiva definida.

Ahora, utilizando un argumento análogo al desarrollado en la demostración del teorema (1.1), por el teorema de Kolmogorov se tiene que existe un proceso Gaussiano bivariado  $\{(Y_t, Z_t)^T, t = 0, \pm 1, \pm 2, \dots\}$  tal que

$$(Y_{t+1}, Y_{t+2}, \dots, Y_{t+n}, Z_{t+1}, Z_{t+2}, \dots, Z_{t+n})^T \sim N(0, L^{(n)}).$$

Definiendo  $X_t = Y_t + iZ_t$  se tiene que  $\mathbb{E}[X_t] = \mathbb{E}[Y_t] + i\mathbb{E}[Z_t] = 0$  para toda  $t$ .

Además

$$\begin{aligned}
 \mathbb{E}[X_t X_{t+h}] &= \mathbb{E}[(Y_t + iZ_t)(Y_{t+h} + iZ_{t+h})] \\
 &= \mathbb{E}[Y_t Y_{t+h} - Z_t Z_{t+h} + iY_t Z_{t+h} - iZ_t Y_{t+h}] \\
 &= \mathbb{E}[Y_t Y_{t+h}] - \mathbb{E}[Z_t Z_{t+h}] + i\mathbb{E}[Y_t Z_{t+h}] - i\mathbb{E}[Z_t Y_{t+h}] \\
 &= \frac{1}{2}(K_1(h) + K_1(h) + iK_2(h) - i(-K_2(h))) \\
 &= K_1(h) + iK_2(h) = K(h).
 \end{aligned}$$

□

Ahora avancemos un poco hacia la representación espectral de un proceso estacionario, para esto consideremos el siguiente proceso que será retomado más adelante.

Sea  $\{X_t, t \in T\}$  tal que

$$X_t = \sum_{j=1}^n A(\lambda_j) e^{it\lambda_j}, \quad (1.11)$$

donde  $-\pi \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n = \pi$  y  $A(\lambda_1), A(\lambda_2), \dots, A(\lambda_n)$  son variables aleatorias no correlacionadas con valores en los complejos y tales que

$$\mathbb{E}[A(\lambda_j)] = 0$$

y

$$\mathbb{E}\left[A(\lambda_j) \overline{A(\lambda_j)}\right] = \sigma_j^2, \text{ para toda } j = 1, 2, \dots, n.$$

Entonces se tiene que

$$\mathbb{E}[X_t] = \mathbb{E}\left[\sum_{j=1}^n A(\lambda_j) e^{it\lambda_j}\right] = \sum_{j=1}^n \mathbb{E}[A(\lambda_j) e^{it\lambda_j}] = \sum_{j=1}^n e^{it\lambda_j} \mathbb{E}[A(\lambda_j)] = 0,$$

es decir,

$$\mathbb{E}[X_t] = 0.$$

Además

$$\mathbb{E}[X_{t+h} \overline{X_t}] = \mathbb{E}\left[\sum_{j=1}^n A(\lambda_j) e^{i(t+h)\lambda_j} \overline{\sum_{j=1}^n A(\lambda_j) e^{it\lambda_j}}\right]$$

$$\begin{aligned}
 &= \mathbb{E} \left[ \sum_{j=1}^n A(\lambda_j) e^{i(t+h)\lambda_j} \sum_{j=1}^n \overline{A(\lambda_j)} e^{-it\lambda_j} \right] \\
 &= \mathbb{E} \left[ \sum_{j=1}^n \sum_{k=1}^n A(\lambda_j) \overline{A(\lambda_k)} e^{it(\lambda_j - \lambda_k)} e^{ih\lambda_j} \right] \\
 &= \sum_{j=1}^n \sum_{k=1}^n e^{it(\lambda_j - \lambda_k)} e^{ih\lambda_j} \mathbb{E} \left[ A(\lambda_j) \overline{A(\lambda_k)} \right] \\
 &= \sum_{j=1}^n e^{ih\lambda_j} \sigma_j^2 + \sum_{k \neq j} e^{it(\lambda_j - \lambda_k)} e^{ih\lambda_j} \mathbb{E} \left[ A(\lambda_j) \overline{A(\lambda_k)} \right] = \sum_{j=1}^n e^{ih\lambda_j} \sigma_j^2
 \end{aligned}$$

por ser  $A(\lambda_j)$  y  $\overline{A(\lambda_k)}$  no correlacionadas, además la última expresión es independiente de  $t$ .

El teorema (1.2) nos da la caracterización de las funciones de autocovarianza definidas en los enteros con valores complejos, como aquellas que son hermitianas y no negativa definidas. El teorema de Herglotz, (Teorema (1.3)), que a continuación se enuncia y demuestra, da una caracterización de la función de autocovarianza muy especial y de gran interés para nosotros. Esta caracterización será utilizada posteriormente.

**Nota 1.4.** Se hará la siguiente convención:  $\gamma_X(\cdot) \equiv \gamma(\cdot)$

**Teorema 1.3. (Teorema de Herglotz)** Una función compleja definida en los enteros  $\gamma(\cdot)$ , es no negativa definida si y sólo si

$$\gamma(h) = \int_{[-\pi, \pi]} e^{ih\nu} dF(\nu) \tag{1.12}$$

para toda  $h = 0, \pm 1 \pm 2, \dots$ , donde  $F(\cdot)$  es una función real acotada, no decreciente, continua por la derecha en  $[-\pi, \pi]$  y tal que  $F(-\pi) = 0$ . A la función  $F$  se le llama **Función de Distribución Espectral** de  $\gamma$ . Si además

$$F(\lambda) = \int_{-\pi}^{\lambda} f(\nu) d\nu,$$

donde  $-\pi \leq \lambda \leq \pi$ , entonces a  $f$  se le llama **Densidad Espectral** de  $\gamma(\cdot)$ .

*Demostración.* Supongamos que

$$\gamma(h) = \int_{[-\pi, \pi]} e^{ih\nu} dF(\nu),$$

donde  $F$  es una función como se explicita arriba, entonces,

$$\gamma(-h) = \int_{[-\pi, \pi]} e^{-ih\nu} dF(\nu).$$

Por otro lado tenemos que

$$\overline{\gamma(h)} = \overline{\int_{[-\pi, \pi]} e^{ih\nu} dF(\nu)} = \int_{[-\pi, \pi]} e^{-ih\nu} \overline{dF(\nu)} = \int_{[-\pi, \pi]} e^{-ih\nu} dF(\nu),$$

puesto que  $F$  es real. Por lo tanto se tiene que  $\gamma(-h) = \overline{\gamma(h)}$ . Ahora consideremos  $\mathbf{a} = (a_1, a_2, \dots, a_n) \in \mathbb{C}^n$ , luego,

$$\begin{aligned} \sum_{r=1}^n \sum_{s=1}^n a_r \gamma(r-s) \overline{a_s} &= \sum_{r=1}^n \sum_{s=1}^n a_r \overline{a_s} \int_{[-\pi, \pi]} e^{i(r-s)\nu} dF(\nu) \\ &= \int_{[-\pi, \pi]} \sum_{r=1}^n \sum_{s=1}^n a_r \overline{a_s} e^{i(r-s)\nu} dF(\nu) = \int_{[-\pi, \pi]} \left( \sum_{r=1}^n a_r e^{ir\nu} \right) \overline{\left( \sum_{s=1}^n \overline{a_s} e^{-is\nu} \right)} dF(\nu) \\ &= \int_{[-\pi, \pi]} \left( \sum_{r=1}^n a_r e^{ir\nu} \right) \overline{\left( \sum_{s=1}^n a_s e^{is\nu} \right)} dF(\nu) = \int_{[-\pi, \pi]} \left| \sum_{r=1}^n a_r e^{ir\nu} \right|^2 dF(\nu) \geq 0; \end{aligned}$$

por lo tanto  $\gamma(\cdot)$  es no negativa definida y por el teorema (1.2) se tiene que  $\gamma$  es una función de autocovarianza.

Conversamente supongamos que  $\gamma(\cdot)$  es no negativa definida,  $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$  y definamos a  $f_N$  de la siguiente manera:

$$f_N(\nu) = \frac{1}{2\pi N} \sum_{r=1}^N \sum_{s=1}^N e^{-ir\nu} \gamma(r-s) e^{is\nu} = \frac{1}{2\pi N} \sum_{r=1}^N \sum_{s=1}^N e^{-i(r-s)\nu} \gamma(r-s)$$

los sumandos de la última expresión se pueden escribir como

$$\begin{array}{cccccc}
 & r = 1 & r = 2 & r = 3 & \dots & r = N \\
 s = 1 & e^{-i(0)\nu}\gamma(0) & e^{-i(1)\nu}\gamma(1) & e^{-i(2)\nu}\gamma(2) & \dots & e^{-i(N-1)\nu}\gamma(N-1) \\
 s = 2 & e^{-i(-1)\nu}\gamma(-1) & e^{-i(0)\nu}\gamma(0) & e^{-i(1)\nu}\gamma(1) & \dots & e^{-i(N-2)\nu}\gamma(N-2) \\
 s = 3 & e^{-i(-2)\nu}\gamma(-2) & e^{-i(-1)\nu}\gamma(-1) & e^{-i(0)\nu}\gamma(0) & \dots & e^{-i(N-3)\nu}\gamma(N-3) \\
 \vdots & \vdots & \dots & \ddots & \vdots & \\
 s = N & e^{-i(1-N)\nu}\gamma(1-N) & e^{-i(2-N)\nu}\gamma(2-N) & \dots & \dots & e^{-i(0)\nu}\gamma(0)
 \end{array}$$

es decir

$$\begin{array}{cccccc}
 & r = 1 & r = 2 & r = 3 & \dots & r = N \\
 s = 1 & e^{-i(0)\nu}\gamma(0) & e^{-i(1)\nu}\gamma(1) & e^{-i(2)\nu}\gamma(2) & \dots & e^{-i(N-1)\nu}\gamma(N-1) \\
 s = 2 & e^{-i(-1)\nu}\gamma(1) & e^{-i(0)\nu}\gamma(0) & e^{-i(1)\nu}\gamma(1) & \dots & e^{-i(N-2)\nu}\gamma(N-2) \\
 s = 3 & e^{-i(-2)\nu}\gamma(2) & e^{-i(-1)\nu}\gamma(1) & e^{-i(0)\nu}\gamma(0) & \dots & e^{-i(N-3)\nu}\gamma(N-3) \\
 \vdots & \vdots & \dots & \ddots & \vdots & \\
 s = N & e^{-i(1-N)\nu}\gamma(N-1) & e^{-i(2-N)\nu}\gamma(N-2) & \dots & \dots & e^{-i(0)\nu}\gamma(0)
 \end{array}$$

entonces, sumamos sobre las diagonales de esta matriz cada

$$m = r - s = 1 - N, 2 - N, \dots, -1, 0, 1, \dots, N - 1 = -(N - 1), \dots, -1, 0, 1, \dots, (N - 1).$$

Por ejemplo, para  $m = 0$  se tienen  $N$  sumandos en la diagonal principal, para  $m = \pm 1$ , o sea,  $|m| = 1$ , se tienen  $N - 1$  sumandos en cada diagonal arriba y abajo de la principal, y así sucesivamente. Por lo tanto podemos escribir

$$f_N(\nu) = \frac{1}{2\pi N} \sum_{|m| < N} (N - |m|) e^{-im\nu} \gamma(m) = \frac{1}{2\pi} \sum_{|m| < N} \left(1 - \frac{|m|}{N}\right) e^{-im\nu} \gamma(m).$$

Notemos que a partir de este hecho se puede ver que  $f_N$  es real:

$$f_N(\nu) = \frac{1}{2\pi} \sum_{|m| < N} \left(1 - \frac{|m|}{N}\right) e^{-im\nu} \gamma(m) = \frac{1}{2\pi} \sum_{m=-(N-1)}^{N-1} \left(1 - \frac{|m|}{N}\right) e^{-im\nu} \gamma(m)$$



$$\begin{aligned}
 &= \frac{1}{2\pi} e^{-i(0)} \gamma(0) + \frac{1}{2\pi} \sum_{m=1}^{N-1} \left(1 - \frac{|m|}{N}\right) e^{-im\nu} \gamma(m) + \frac{1}{2\pi} \sum_{m=-(N-1)}^{-1} \left(1 - \frac{|m|}{N}\right) e^{-im\nu} \gamma(m) \\
 &= \frac{1}{2\pi} e^{-i(0)} \gamma(0) + \frac{1}{2\pi} \sum_{m=1}^{N-1} \left(1 - \frac{|m|}{N}\right) e^{-im\nu} \gamma(m) + \frac{1}{2\pi} \sum_{m=1}^{N-1} \left(1 - \frac{|m|}{N}\right) e^{im\nu} \gamma(-m) \\
 &= \frac{1}{2} \gamma(0) + \frac{1}{2\pi} \sum_{m=1}^{N-1} \left(1 - \frac{|m|}{N}\right) e^{-im\nu} \gamma(m) + \frac{1}{2\pi} \sum_{m=1}^{N-1} \left(1 - \frac{|m|}{N}\right) \overline{e^{-im\nu} \gamma(m)} \\
 &= \frac{1}{2} \gamma(0) + 2\operatorname{Re} \left\{ \frac{1}{2\pi} \sum_{m=1}^{N-1} \left(1 - \frac{|m|}{N}\right) e^{-im\nu} \gamma(m) \right\},
 \end{aligned}$$

donde como  $\gamma(0) = \overline{\gamma(-0)} = \overline{\gamma(0)}$ , entonces se tiene que  $\gamma(0) \in \mathbb{R}$ .

Por ser  $\gamma(\cdot)$  no negativa definida, se tiene que  $f_N(\nu) \geq 0$  para toda  $\nu \in [-\pi, \pi]$ . Además,  $f_N : (-\pi, \pi) \rightarrow \mathbb{R}$  es una función continua por ser combinación lineal de funciones continuas de  $\nu$ , es decir, por ser combinación lineal de funciones exponenciales de la forma  $e^{-im\nu}$ .

Sea  $F_N$  la función de distribución correspondiente a

$$f_N(\cdot) \mathbb{1}_{(-\pi, \pi]}(\cdot),$$

entonces, se tiene que  $F_N(\lambda) = 0$  para  $\lambda \leq -\pi$ ,  $F_N(\lambda) = F_N(\pi)$  para  $\lambda \geq \pi$  y para  $-\pi \leq \lambda \leq \pi$

$$F_N(\lambda) = \int_{-\pi}^{\lambda} f_N(\nu) d\nu, \quad (1.13)$$

entonces, para toda  $h \in \mathbb{Z}$

$$\begin{aligned}
 \int_{(-\pi, \pi]} e^{ih\nu} dF_N(\nu) &= \int_{(-\pi, \pi]} e^{ih\nu} f_N(\nu) d\nu = \int_{(-\pi, \pi]} e^{ih\nu} \frac{1}{2\pi N} \sum_{|m| < N} e^{-im\nu} \gamma(m) (N - |m|) d\nu \\
 &= \frac{1}{2\pi N} \sum_{|m| < N} (N - |m|) \gamma(m) \int_{(-\pi, \pi]} e^{-i(m-h)\nu} d\nu = \frac{1}{2\pi} \sum_{|m| < N} \left(1 - \frac{|m|}{N}\right) \gamma(m) \int_{(-\pi, \pi]} e^{-i(m-h)\nu} d\nu,
 \end{aligned}$$

donde

$$\frac{1}{2\pi} \int_{(-\pi, \pi]} e^{-i(m-h)\nu} d\nu = \begin{cases} 0 & m \neq h, \\ 1 & m = h \end{cases} \quad (1.14)$$

lo anterior es cierto puesto que:

(i) Para  $m = h$  se tiene

$$\frac{1}{2\pi} \int_{(-\pi, \pi]} e^{-i(0)\nu} d\nu = \int_{(-\pi, \pi]} \frac{1}{2\pi} d\nu = 1.$$

(ii) Para  $m \neq h$  se tiene que para toda  $k = m - h$

$$\begin{aligned} \frac{1}{2\pi} \int_{(-\pi, \pi]} e^{-i(m-h)\nu} d\nu &= \frac{1}{2\pi} \int_{(-\pi, \pi]} e^{-i(k)\nu} d\nu \\ &= \frac{1}{2\pi} \int_{(-\pi, \pi]} \cos(k\nu) d\nu - i \frac{1}{2\pi} \int_{(-\pi, \pi]} \text{sen}(k\nu) d\nu \\ &= \frac{1}{2\pi} \int_{(-\pi, \pi]} \cos(k\nu) d\nu \\ &= \frac{1}{2\pi} \int_{(-\pi, \pi]} \frac{d \text{sen}(k\nu)}{k} d\nu \\ &= \frac{1}{2\pi k} \text{sen}(k\nu) \Big|_{-\pi}^{\pi} = 0 \end{aligned}$$

puesto que la función  $\text{sen}(\cdot)$  es impar. Por lo tanto

$$\int_{(-\pi, \pi]} e^{ih\nu} dF_N(\nu) = \begin{cases} \left(1 - \frac{|h|}{N}\right) \gamma(h) & |h| < N \\ 0 & \text{en otro caso.} \end{cases} \quad (1.15)$$

Ahora, usando (1.15) se tiene que

$$F_N(\pi) = \int_{(-\pi, \pi]} dF_N(\nu) = \gamma(0) < \infty,$$

para toda  $N$ . Además como

$$F_N(\pi) = \int_{(-\pi, \pi]} dF_N(\nu)$$

podemos utilizar el teorema de Helly y el teorema de Helly-Bray<sup>1</sup>. Así dado que por construcción la sucesión de funciones  $F_N$  resultan ser de Distribución Generalizadas, podemos aplicar tanto

---

<sup>1</sup>ver apéndice A

el teorema (A.2) como el teorema (A.3) para garantizar la existencia de una subsucesión  $\{F_{n_k}\}$  de  $\{F_n\}$  tal que para cualquier función continua y acotada  $g$  con  $g(\pi) = g(-\pi)$ , se tiene que

$$\lim_{k \rightarrow \infty} \int_{(-\pi, \pi]} g(\nu) dF_{n_k}(\nu) = \int_{(-\pi, \pi]} g(\nu) dF(\nu).$$

Entonces sustituyendo  $N$  por  $N_k$  en la ecuación (1.15) y haciendo  $k \rightarrow \infty$ , tenemos que

$$\lim_{k \rightarrow \infty} \int_{(-\pi, \pi]} e^{ih\nu} dF_{N_k}(\nu) = \gamma(h),$$

es decir,

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ih\nu} dF(\nu) \tag{1.16}$$

que es la representación espectral de  $\gamma(\cdot)$  buscada. □

**Corolario 1.1.** *Una función  $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$  es la función de autocovarianza de un proceso estacionario  $\{X_t, t \in T\}$  si y sólo si, se cumple una de las dos proposiciones siguientes:*

(i)  $\gamma(h) = \int_{(-\pi, \pi]} e^{ih\nu} dF(\nu)$ , para toda  $h = 0, \pm 1, \pm 2, \dots$ , con  $F$  función continua por la derecha, no decreciente y acotada en  $[-\pi, \pi]$  y  $F(-\pi) = 0$ .

ó

(ii)  $\sum_{i,j=1}^n a_i \gamma(i-j) \overline{a_j} \geq 0$  para toda  $n \in \mathbb{N}$  y para toda  $\mathbf{a} = (a_1, a_2, \dots, a_n)^t \in \mathbb{C}^n$ .

*Demostración.* Por el Teorema de Herglotz (Teorema (1.3)) se tiene la equivalencia entre (i) y (ii). Además (i) implica que  $\gamma(\cdot)$  es Hermitiana, por lo tanto se puede aplicar el teorema (1.1) si  $\gamma$  es tal que satisface (i) y (ii). □

Finalmente veremos un teorema que explicita las condiciones para que exista la función de densidad espectral asociada a un proceso estacionario  $\{X_t, t \in T\}$ .

**Teorema 1.4.** *Sea  $K : \mathbb{Z} \rightarrow \mathbb{C}$  tal que  $\sum_{n=-\infty}^{\infty} |K(n)| < \infty$  entonces*

$$K(h) = \int_{-\pi}^{\pi} e^{ih\nu} f(\nu) d\nu$$

para  $h = 0, \pm 1, \pm 2, \dots$ , donde

$$f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} K(n),$$

para toda  $\lambda \in [-\pi, \pi]$ .

*Demostración.*

$$\begin{aligned} \int_{-\pi}^{\pi} e^{ih\nu} f(\nu) d\nu &= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\nu} K(n) e^{ih\nu} d\nu = \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-i(n-h)\nu} K(n) d\nu \\ &= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} K(n) \int_{-\pi}^{\pi} e^{-i(n-h)\nu} d\nu = K(h), \end{aligned}$$

la última igualdad se obtiene utilizando la ecuación (1.14) de la demostración del teorema de Herglotz, (teorema (1.3)), y el teorema de Fubini.  $\square$

**Corolario 1.2.** Una función con valores en los complejos, absolutamente sumable y definida en los enteros,  $\gamma(\cdot) : \mathbb{Z} \rightarrow \mathbb{C}$ , es la función de autocovarianza de un proceso estacionario si y sólo si

$$f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} \gamma(n) \tag{1.17}$$

para toda  $\lambda \in (-\pi, \pi]$ . En cuyo caso  $f(\cdot)$  es la densidad espectral de  $\gamma(\cdot)$ .

*Demostración.* Primero supongamos que  $\gamma(\cdot)$  es la función de autocovarianza de un proceso estacionario, entonces  $\gamma(\cdot)$  es no negativa definida, Hermitiana y absolutamente sumable, por lo tanto

$$0 \leq f_N(\gamma) = \frac{1}{2\pi N} \sum_{r,s=1}^N e^{-ir\lambda} \gamma(r-s) e^{is\lambda} = \frac{1}{2\pi} \sum_{|m| \leq N} \left(1 - \frac{|m|}{N}\right) e^{-im\lambda} \gamma(m)$$

la cual converge a  $f(\lambda)$  cuando  $N \rightarrow \infty$ , por un argumento similar al que se dio en el teorema de Herglotz (Teorema (1.3)), por lo tanto tenemos que  $f(\lambda) \geq 0$ ; además por el teorema (1.4)

$$\gamma(h) = \int_{-\pi}^{\pi} e^{-ih\nu} f(\nu) d\nu$$

para toda  $h = 0, \pm 1, \pm 2, \dots$ , luego  $f$  es la densidad espectral de  $\gamma$  deseada.

Conversamente, supongamos que  $f(\cdot)$  es la densidad espectral de  $\gamma(\cdot)$ , la cual es absolutamente sumable, entonces por el teorema (1.4) se tiene que

$$\gamma(\cdot) = \int_{-\pi}^{\pi} e^{ih\nu} f(\nu) d\nu$$

para toda  $h = 0, \pm 1, \pm 2, \dots$ . Dado que  $f(\lambda) \geq 0$  entonces  $\gamma(\cdot)$  satisface el teorema de Herglotz, (1.3), de donde se tiene que

$$F(\lambda) = \int_{-\pi}^{\pi} f(\nu) d\nu,$$

y por el corolario (1.2),  $\gamma$  es la función de autocovarianza de un proceso estacionario con densidad espectral  $f$ . □

## Capítulo 2

### Teoría de Filtros y Onduletas.

A principios del siglo XIX el matemático francés Jean Baptiste Joseph Fourier afirmaba que cualquier función periódica se podía expresar como una suma infinita de senos y cosenos de diversas frecuencias, esta afirmación con el tiempo se extendió a funciones no periódicas que cambian con el tiempo en vez de repetir su forma una y otra vez, esto se hacía por medio de la *Transformada de Fourier*. Esta herramienta fue la preferida para resolver varios problemas de física e ingeniería. Un análisis más profundo del método, hizo notar que la teoría de Fourier tenía ciertos problemas para algunas aplicaciones.

Aunque cada científico intentaba resolver los problemas específicos de su respectivo campo, todos comenzaron a llegar a la misma conclusión: el origen de estos problemas son las transformaciones de Fourier en sí. También llegaron en esencia a la misma solución: quizás al dividir una señal en componentes que no fueran ondas sinusoidales puras sería posible condensar la información tanto en el dominio del tiempo como en el de la frecuencia. Esta es la idea que finalmente se denominaría *onduleta*. La primera aparición de las *onduletas* fue en 1909 en la tesis del matemático Alfred Haar, quien introdujo a las funciones ahora conocidas como *onduletas* como un conjunto ortonormal de funciones con soporte compacto, es decir, funciones que se anulan fuera de un intervalo finito, con el tiempo llamadas *bases de Haar*, sin embargo debido a que estas funciones no eran continuamente diferenciables sus aplicaciones se vieron limitadas. En 1946 el físico Dennis Gabor introdujo la transformada de Gabor, análoga a la transformada de Fourier, la cual consistía en una base no-ortogonal de *onduletas* con soporte no acotado, basado en funciones gaussianas trasladadas.

El término *onduleta* proviene del campo de la sismología, bautizado por Ricker en 1940 para describir el efecto de un impulso sísmico agudo o una carga explosiva. En 1982, Morlet, ingeniero de Elf-Aquitania, demostró cómo estos efectos sísmicos se podían modelar perfectamente con las

funciones definidas por Gabor. Morlet desarrolló su propia forma de analizar las señales sísmicas para crear componentes que estuvieran localizados en el espacio, a los que denominó *onduletas de forma constante* las cuales posteriormente se conocerían como *onduletas de Morlet*.

Alex Grossmann, físico del *Centre de Physique Théorique* de Marsella con quien Morlet trabajó durante un año confirmaron que las señales sísmicas se podían reconstruir a partir de sus descomposiciones en *onduletas*. De hecho, las transformaciones de *onduletas* resultaron funcionar mucho mejor, en la mayoría de los casos, que las transformaciones de Fourier. El artículo, publicado en 1984, de Morlet y Grossmann fue el primero en el que se utilizó la palabra *onduleta*.

Yves Meyer, actualmente en la *École Normale Supérieure de Cachan*, uno de los fundadores de la teoría de las *onduletas*, fue el primero en relacionar las *onduletas* de Morlet con las *onduletas* matemáticas anteriores. Meyer continuó su trabajo para descubrir un nuevo tipo de *onduleta* con una propiedad matemática denominada ortogonalidad, la cual hacía que manipular y trabajar con la transformación de *onduletas* resultara tan fácil como con una transformación de Fourier. En 1986, Stéphane Mallat, un antiguo alumno de Meyer vinculó la teoría de *onduletas* a la literatura existente sobre codificación de sub-bandas y filtros de duplicación de cuadratura, que son las versiones de las *onduletas* usadas por la comunidad de procesamiento de imágenes. La idea del análisis multiresolución, es decir, la observación de una señal a distintas escalas de resolución, ya era familiar para los expertos en procesamiento de imágenes. Mallat, en colaboración con Meyer, demostró que las *onduletas* están implícitas en el proceso del análisis multiresolución.

El lenguaje de las *onduletas* también resultaba más cómodo para los ingenieros eléctricos, quienes adoptaron términos familiares tales como filtros, altas y bajas frecuencias. El último gran avance se logró en 1987, cuando Ingrid Daubechies mientras visitaba el Courant Institute de la Universidad de Nueva York y posteriormente durante su trabajo en los laboratorios AT & T Bell, descubrió una clase completamente nueva de *onduletas*, que no sólo eran ortogonales (como las de Meyer) sino que también se podían implementar mediante sencillas ideas de filtrado digital, estas nuevas *onduletas* eran casi tan sencillas de programar y utilizar como las *onduletas* de Haar, suaves, sin los saltos de las *onduletas* de Haar. Al combinar las ideas de Daubechies y Mallat, se disponía de una transformación ortogonal y sencilla que se podía calcular rápidamente en las modernas computadoras digitales. Las *onduletas* de Daubechies convirtieron a la teoría en una herramienta práctica que cualquier científico con una formación matemática mínima puede programar y utilizar con facilidad.

## 2.1. Transformada Discreta de Fourier

Sea  $\{a_t : t = 0, 1, 2, \dots\}$  una sucesión de números reales o complejos tales que

$$\sum_{t=-\infty}^{\infty} |a_t|^2 < \infty.$$

La transformada discreta de Fourier (*DFT*) de  $\{a_t\}$  es la función  $A(f)$  con valores en los complejos definida por

$$A(f) \equiv \sum_{t=-\infty}^{\infty} a_t e^{-i2\pi f t}, \quad (2.1)$$

donde  $-\infty < f < \infty$  es un número real llamado *frecuencia*. A menudo a la función  $A(f)$  también se le llama análisis de Fourier de  $\{a_t\}$ . Esta función tiene varias propiedades importantes que vale la pena enunciar dado que se utilizarán constantemente más adelante:

- (i) La función  $A(\cdot)$  es periódica y con período 1, es decir, para cualquier  $j \in \mathbb{Z}$  se tiene que  $A(f + j) = A(f)$ :

$$A(f + j) = \sum_{t=-\infty}^{\infty} a_t e^{-i2\pi(f+j)t} = \sum_{t=-\infty}^{\infty} a_t e^{-i2\pi f t} e^{-i2\pi j t} = \sum_{t=-\infty}^{\infty} a_t e^{-i2\pi f t} = A(f).$$

- (ii) Si  $\{a_t\}$  es real, entonces  $A(-f) = \overline{A(f)}$ , a saber

$$A(-f) = \sum_{t=-\infty}^{\infty} a_t e^{i2\pi f t},$$

y por otra parte tenemos que

$$\overline{A(f)} = \overline{\sum_{t=-\infty}^{\infty} a_t e^{-i2\pi f t}} = \sum_{t=-\infty}^{\infty} \overline{a_t e^{-i2\pi f t}} = \sum_{t=-\infty}^{\infty} a_t e^{i2\pi f t},$$

por lo tanto se tiene la igualdad deseada.



(iii) La sucesión  $a_t$  se puede recuperar de  $A(f)$  usando la siguiente relación:

$$a_t = \int_{-1/2}^{1/2} A(f) e^{i2\pi f t} df \tag{2.2}$$

para  $t = \pm 1, \pm 2, \dots$

$$\begin{aligned} \int_{-1/2}^{1/2} A(f) e^{i2\pi f t} df &= \int_{-1/2}^{1/2} \left( \sum_{s=-\infty}^{\infty} a_s e^{-i2\pi f s} \right) e^{i2\pi f t} df \\ &= \int_{-1/2}^{1/2} \sum_{s=-\infty}^{\infty} a_s e^{-i2\pi f (s-t)} df = \sum_{s=-\infty}^{\infty} a_s \int_{-1/2}^{1/2} e^{-i2\pi f (s-t)} df \end{aligned}$$

para esta última igualdad se utilizaron los siguientes hechos:

Sea  $S : \mathbb{R} \rightarrow \mathbb{C}$  función periódica con periodo 1, tal que sea cuadrado integrable es decir,

$$\int_{-1/2}^{1/2} |S(f)|^2 df < \infty.$$

Y sea la sucesión de funciones  $\{S_N(f)\}$  definidas en  $L^2(\mathcal{B}(-1/2, 1/2], (-1/2, 1/2], \lambda)$  donde  $S_N(f) = \sum_{k=-N}^N a_k e^{i2\pi f k}$ , entonces demostremos que

$$\lim_{N \rightarrow \infty} \int_{-1/2}^{1/2} |S(f) - S_N(f)|^2 df = 0$$

equivalentemente, demostremos que la sucesión de funciones  $\{S_N(f)\}$  es una sucesión de Cauchy, a saber si  $m < n$

$$\begin{aligned} \|S_n(f) - S_m(f)\|^2 &= \left\| \sum_{k=-n}^n a_k e^{i2\pi f k} - \sum_{k=-m}^m a_k e^{i2\pi f k} \right\|^2 \\ &= \left\| \sum_{k=m+1}^n a_k e^{i2\pi f k} \right\|^2 \leq \sum_{k=m+1}^n \|a_k e^{i2\pi f k}\|^2 \\ &= \sum_{k=m+1}^n \|a_k\|^2 \|e^{i2\pi f k}\|^2 = \sum_{k=m+1}^n \|a_k\|^2 < \infty \end{aligned}$$

por lo tanto es una sucesión de Cauchy en  $L^2$  y por lo tanto se tiene que la sucesión converge en  $L^2$ . Ahora definamos el siguiente producto interno en  $L^2(-1/2, 1/2) \times L^2(-1/2, 1/2)$  en  $\mathbb{C}$

$$\langle f, g \rangle = \int_{-1/2}^{1/2} f(t) \bar{g}(t) dt$$

entonces este producto interno así definido es una función continua en  $L^2$ . Luego

$$\begin{aligned} \int_{-1/2}^{1/2} A(f) e^{i2\pi f t} df &= \int_{-1/2}^{1/2} \sum_{s=-\infty}^{\infty} a_s e^{-i2\pi f s} e^{i2\pi f t} df \\ &= \int_{-1/2}^{1/2} \sum_{s=-\infty}^{\infty} a_s e^{-i2\pi f (t-s)} df, \end{aligned}$$

entonces aplicando lo realizado anteriormente, se tiene que

$$\begin{aligned} \int_{-1/2}^{1/2} \sum_{n=-m}^m a_n e^{-i2\pi f n} e^{i2\pi f t} df &= \int_{-1/2}^{1/2} \sum_{n=-m}^m a_n e^{i2\pi f (t-n)} df \nearrow \int_{-1/2}^{1/2} \sum_{n=-\infty}^{\infty} a_n e^{-i2\pi f n} e^{i2\pi f t} df \\ &= \int_{-1/2}^{1/2} \sum_{n=-\infty}^{\infty} a_n e^{i2\pi f (t-n)} df \end{aligned}$$

por lo tanto

$$\sum_{n=-\infty}^{\infty} \int_{-1/2}^{1/2} a_n e^{i2\pi f (t-n)} df = \int_{-1/2}^{1/2} \sum_{n=-\infty}^{\infty} a_n e^{i2\pi f (t-n)} df.$$

Esta última propiedad es muy importante porque nos dice que podemos reconstruir a  $\{a_t\}$  por medio de su Transformada Discreta de Fourier. Cuando  $\{a_t\}$  y  $A(f)$  satisfacen (2.1) y (2.2) decimos que son un *par de Fourier* y lo denotamos por

$$\{a_t\} \leftrightarrow A(f). \quad (2.3)$$

Hay otras dos propiedades importantes de las cuales se va a hacer uso constante en el desarrollo de la teoría de filtrado y necesarias para la teoría de *onduletas*.

- (iv) Supongamos que los pares  $\{a_t\} \leftrightarrow A(\cdot)$  y  $\{b_t\} \leftrightarrow B(\cdot)$ , son tales que  $\sum_{t=-\infty}^{\infty} |a_t|^2 < \infty$  y  $\sum_{t=-\infty}^{\infty} |b_t|^2 < \infty$ , entonces se tiene la versión para dos sucesiones del teorema de Parseval, es decir,

$$\sum_{t=-\infty}^{\infty} a_t \bar{b}_t = \int_{-1/2}^{1/2} A(f) \bar{B}(f) df. \quad (2.4)$$

Un corolario inmediato es el siguiente

$$\sum_{t=-\infty}^{\infty} |a_t|^2 = \int_{-1/2}^{1/2} |A(f)|^2 df. \quad (2.5)$$

(v) Consideremos el par de Fourier  $\{a_t\} \leftrightarrow \{A(f)\}$ . Entonces la Transformada Discreta de Fourier de la sucesión  $\{a_{2n} : n = 0, \pm 1, \pm 2, \dots\}$  es la función definida por:

$$C(f) = \frac{1}{2} \left[ A\left(\frac{f}{2}\right) + A\left(\frac{f}{2} + \frac{1}{2}\right) \right]. \quad (2.6)$$

A saber

$$A\left(\frac{f}{2}\right) = \sum_{t=-\infty}^{\infty} a_t e^{-i2\pi\left(\frac{f}{2}\right)t} = \sum_{t=-\infty}^{\infty} a_t e^{-i\pi f t}.$$

Por otra parte tenemos que

$$\begin{aligned} A\left(\frac{f}{2} + \frac{1}{2}\right) &= \sum_{t=-\infty}^{\infty} a_t e^{-i2\pi\left(\frac{f}{2} + \frac{1}{2}\right)t} = \sum_{t=-\infty}^{\infty} a_t e^{-i\pi(f+1)t} \\ &= \sum_{t=-\infty}^{\infty} a_t e^{-i\pi f t} e^{-i\pi t} = \sum_{t=-\infty}^{\infty} (-1)^t a_t e^{-i\pi f t}, \end{aligned}$$

de donde

$$A\left(\frac{f}{2}\right) + A\left(\frac{f}{2} + \frac{1}{2}\right) = \sum_{t=-\infty}^{\infty} a_t e^{-i\pi f t} + \sum_{t=-\infty}^{\infty} (-1)^t a_t e^{-i\pi f t} = 2 \sum_{t=-\infty}^{\infty} a_{2t} e^{-i\pi f(2t)}.$$

por lo tanto concluimos que

$$\frac{1}{2} \left[ A\left(\frac{f}{2}\right) + A\left(\frac{f}{2} + \frac{1}{2}\right) \right] = \sum_{t=-\infty}^{\infty} a_{2t} e^{-i\pi f(2t)}.$$

## 2.2. Convolución

Sean  $\{a_t\}$  y  $\{b_t\}$  sucesiones tales que  $\sum_{t=-\infty}^{\infty} |a_t|^2 < \infty$  y  $\sum_{t=-\infty}^{\infty} |b_t|^2 < \infty$ , entonces se define la *convolución* entre  $\{a_t\}$  y  $\{b_t\}$  como la sucesión infinita cuyo  $t$ -ésimo elemento está dado por

$$(a * b)_t \equiv \sum_{u=-\infty}^{\infty} a_u b_{t-u} \quad (2.7)$$

para  $t = 0, \pm 1, \pm 2, \dots$

Supóngase que  $\{a_t\} \leftrightarrow A(\cdot)$  y  $\{b_t\} \leftrightarrow B(\cdot)$ , entonces la Transformada Discreta de Fourier de la convolución  $\{(a * b)_t, t = 0, \pm 1, \pm 2, \dots\}$  es la función dada por  $A(f)B(f)$  es decir, para  $|f| \leq 1/2$

$$\sum_{t=-\infty}^{\infty} (a * b)_t e^{-i2\pi ft} = A(f)B(f). \quad (2.8)$$

Para probar lo anterior tenemos que

$$\begin{aligned} \sum_{t=-\infty}^{\infty} (a * b)_t e^{-i2\pi ft} &= \sum_{t=-\infty}^{\infty} \sum_{u=-\infty}^{\infty} a_u b_{t-u} e^{-i2\pi ft} = \sum_{u=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} a_u b_{t-u} e^{-i2\pi ft} \\ &= \sum_{u=-\infty}^{\infty} a_u \sum_{t=-\infty}^{\infty} b_{t-u} e^{-i2\pi ft} = \sum_{u=-\infty}^{\infty} a_u \sum_{s=-\infty}^{\infty} b_s e^{-i2\pi f(s+u)} \\ &= \sum_{u=-\infty}^{\infty} a_u e^{-i2\pi fu} \sum_{s=-\infty}^{\infty} b_s e^{-i2\pi fs} = A(f)B(f) \end{aligned}$$

donde la segunda igualdad se obtiene aplicando el teorema de Fubini, y en la tercer igualdad se realiza el cambio de variable  $s = t - u$ , así,

$$\{(a * b)_t\} \leftrightarrow A(f)B(f). \quad (2.9)$$

Otro concepto relacionado con el de convolución es el de *correlación cruzada compleja* definida por la relación

$$(\bar{a} * b)_t \equiv \sum_{u=-\infty}^{\infty} \bar{a}_u b_{u+t}, \quad (2.10)$$

el cual, como se podrá ver, es la convolución de  $\{\bar{a}_t\}$  con una versión de  $\{b_t\}$  donde el tiempo ha sido invertido. Entonces por un argumento similar al que se usó para probar (2.9), se tiene que

$$\{(\bar{a} * b)_t\} \leftrightarrow \overline{A(f)} B(f). \quad (2.11)$$

Haciendo  $b_t = a_t$ , tenemos el concepto de *autocorrelación*

$$(\bar{a} * a)_t \equiv \sum_{u=-\infty}^{\infty} \bar{a}_u a_{u+t}, \text{ con } t = 0, \pm 1, \pm 2, \dots,$$

por (2.9) tenemos que

$$\{(\bar{a} * a)_t\} \leftrightarrow \overline{A(f)} A(f) = |A(f)|^2. \quad (2.12)$$

A la función  $|A(f)|^2$  se le llama *función de ganancia cuadrada*; un enfoque más relacionado con el tema principal de este trabajo, es en el cual podemos ver a la convolución como una operación de filtrado, es decir,  $\{a_t\}$  juega el papel del filtro y  $\{b_t\}$  el papel de la sucesión a filtrar, así,  $\{(a * b)_t\}$  sería el resultado de filtrar a  $\{b_t\}$  usando  $\{a_t\}$ .

En terminología de filtrado, a la sucesión  $\{a_t\}$  se le denomina representación en el dominio del tiempo del filtro o sucesión de impulso-respuesta del filtro, mientras que a la función  $A(f)$  se le llama representación en dominio de frecuencias del filtro o función de transferencia del filtro. Un último concepto a introducir en esta sección es el de *filtros en cascada*. Considérense la sucesión de filtros dados por

$$\{\{a_{m,t} : t = 0, \pm 1, \pm 2, \dots\} : m = 1, 2, \dots, M\} \quad (2.13)$$

con sus correspondientes funciones de transferencia  $A_m(\cdot)$ . Supongamos que la sucesión  $\{b_t\}$  es la entrada a esta cascada de filtros, es decir,  $\{b_t\}$  es la entrada a  $\{a_{1,t}\}$  con salida  $\{a_t^1\}$  con su respectivo par de fourier  $A^{(1)}(f) = A_1(f) B(f)$ , realizando la convolución con  $\{a_{2,t}\}$  obtenemos

$$\{a_t^2\} = \{a_{2,t} * a_t^1\} = \{a_{2,t} * a_{1,t} * b_t\} \leftrightarrow A_2(f) A_1(f) B(f) = A^{(2)}(f).$$

Al continuar este proceso de forma inductiva obtenemos que en la  $m$ -ésima convolución

$$\{a_{m,t} * a_t^{m-1}\} \leftrightarrow A_m(f) \left( \prod_{k=1}^{m-1} A_k(f) \right) B(f) = \left( \prod_{k=1}^m A_k(f) \right) B(f) = A(f) B(f), \quad (2.14)$$

donde

$$A(f) = \prod_{k=1}^m A_k(f). \quad (2.15)$$

Sea  $\{a_t\}$  la sucesión impulso-respuesta que es el par de Fourier de  $A(f)$ , es decir, de acuerdo con la ecuación (2.2)

$$a_t = \int_{-1/2}^{1/2} A(f) e^{i2\pi f t} df.$$

Si  $\{c_t\}$  es la sucesión resultante de aplicar los  $m$  filtros dados en la ecuación (2.13) a la sucesión  $\{b_t\}$ , entonces por las ecuaciones (2.14) y (2.15),  $\{c_t\}$  es el par de Fourier de  $A(f) B(f)$ , es decir,

$$c_t = \sum_{u=-\infty}^{\infty} a_u b_{t-u}$$

para  $t = 0, \pm 1, \pm 2, \dots$

Ahora supongamos que el  $m$ -ésimo filtro es tal que  $a_{m,k_m} \neq 0$  y  $a_{m,k_m+L_m-1} \neq 0$ , y además para toda  $t \leq K_m - 1$  y  $t \geq K_m + L_m$ , se tiene  $a_{m,t} = 0$ , donde  $L_m \in \mathbb{Z}$ . Entonces se dice que el filtro  $\{a_{m,t}\}$  tiene longitud  $L_m$ . La función de transferencia de  $\{a_{m,t}\}$  está dada por

$$A_m(f) = \sum_{t=K_m}^{L_m+K_m-1} a_{m,t} e^{-i2\pi f t}. \quad (2.16)$$

Probamos que la longitud del filtro  $\{a_{m,t}\}$  resultante de aplicar los  $m$  filtros en la cascada, está dada por

$$L^{(M)} = \sum_{m=1}^M L_m - M + 1. \quad (2.17)$$

Primero verifiquemos para  $M = 2$ , consideremos los dos filtros  $\{a_{1,t}\}$  y  $\{a_{2,t}\}$ , entonces la convolución de estos dos filtros está dada por

$$b_{2,t} = \sum_{u=-\infty}^{\infty} a_{1,u} a_{2,t-u}. \quad (2.18)$$

La más pequeña  $t$  para la cual (2.18) no se anula es  $t = k_1 + k_2$ , entonces el valor de  $b_{2,k_1+k_2}$  es  $a_{1,k_1} a_{2,k_2}$ , el valor de  $b_{2,k_1+k_2+1}$  es  $a_{1,k_1} a_{2,k_2+1} + a_{1,k_1+1} a_{2,k_2}$ , y así sucesivamente. La más grande  $t$  para la cual (2.18) no se anula es  $t = k_1 + L_1 - 1 + k_2 + L_2 - 1$ , entonces el valor de  $b_{2,k_1+L_1-1+k_2+L_2-1}$  es  $a_{1,k_1+L_1-1} a_{2,k_2+L_2-1}$ .

Entonces para  $M = 2$  filtros tenemos que la longitud de  $\{b_{2,t}\}$  es

$$L^{(2)} = k_1 + L_1 - 1 + k_2 + L_2 - 1 - (k_1 + k_2 - 1) = L_1 + L_2 - 1 = L_1 + L_2 - M + 1.$$

Si repetimos este proceso para  $M = 3$  filtros tenemos que  $b_{3,t} = \{a_{1,t} * a_{2,t} * a_{3,t}\}$ , o bien,

$$b_{3,t} = \sum_{u=-\infty}^{\infty} b_{2,u} a_{3,t-u}. \quad (2.19)$$

De manera similar a como se hizo para encontrar  $L^{(2)}$ , la convolución de los 2 filtros en (2.19) tiene longitud  $L^{(2)} + L_3 - 1$ , es decir,

$$L^{(3)} = L^{(2)} + L_3 - 1 = (L_1 + L_2 - 1) + L_3 - 1 = L_1 + L_2 + L_3 - 2 = \sum_{i=1}^3 L_i - M + 1.$$

Asumiendo  $m = M - 1$ , entonces tenemos que  $\{b_{M-1,t}\}$  tiene longitud

$$L^{(M-1)} = \sum_{i=1}^{M-1} L_i - (M - 1) + 1, \quad (2.20)$$

entonces para probar la ecuación (2.17) basta ver que la expresión (2.20) vale para  $m = M$ . Lo anterior se sigue de aplicar el razonamiento de arriba y la hipótesis de inducción (2.20) a los filtros  $b_{m-1,t}$  y  $a_{m,t}$  cuya convolución tendrá longitud

$$L^{(M-1)} + L_M - 1 = \sum_{i=1}^M L_i - M + 1.$$

Consideremos ahora a las sucesiones finitas  $\{a_t : t = 0, 1, 2, \dots, N-1\}$ , entonces su Transformada Discreta de Fourier es la sucesión de  $N$  variables  $\{A_k : k = 0, 1, 2, \dots, N-1\}$  dada por

$$A_k = \sum_{t=0}^{N-1} a_t e^{-i2\pi tk/N}, \quad k = 0, 1, 2, \dots, N-1.$$

Notemos que si  $k$  es un entero tal que  $0 \leq k \leq N-1$  y  $N$  es un entero distinto, mayor o igual que 1, entonces  $A_{k+nN} = A_k$ , en efecto

$$A_{k+nN} = \sum_{t=0}^{N-1} a_t e^{-i2\pi t(k+nN)/N} = \sum_{t=0}^{N-1} a_t e^{-i2\pi tk/N} e^{-i2\pi tn} = \sum_{t=0}^{N-1} a_t e^{-i2\pi tk/N} = A_k.$$

Entonces podemos considerar a la sucesión  $\{A_k : k \in \mathbb{Z}\}$  como una sucesión infinita, periódica y con periodo  $N$ . Estudiemos ahora como invertir el operador *DFT* en sucesiones finitas:

A saber

$$\begin{aligned} \frac{1}{N} \sum_{k=0}^{N-1} A_k e^{i2\pi tk/N} &= \frac{1}{N} \sum_{k=0}^{N-1} \left( \sum_{l=0}^{N-1} a_l e^{-i2\pi lk/N} \right) e^{i2\pi tk/N} = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} a_l e^{-i2\pi(l-t)k/N} \\ &= \frac{1}{N} \sum_{l=0}^{N-1} \sum_{k=0}^{N-1} a_l e^{-i2\pi(l-t)k/N} = \frac{1}{N} \sum_{l=0}^{N-1} a_l \sum_{k=0}^{N-1} e^{-i2\pi(l-t)k/N}. \end{aligned}$$

Si  $l = t$  entonces,

$$\sum_{k=0}^{N-1} e^{-i2\pi(l-t)k/N} = N;$$

por otra parte, si  $l \neq t$  entonces

$$\sum_{k=0}^{N-1} e^{-i2\pi(l-t)k/N} = \sum_{k=0}^{N-1} e^{-i2\pi uk/N} = \frac{1 - e^{-i2\pi u}}{1 - e^{-i2\pi u/N}} = 0$$



es decir,

$$\frac{1}{N} \sum_{k=0}^{N-1} e^{-i2\pi(t-l)k/N} = \begin{cases} 1 & t = l \\ 0 & t \neq l \end{cases}; \quad (2.21)$$

como  $0 \leq l \leq N-1$  entonces si  $0 \leq t \leq N-1$  tenemos que  $u = t - l \leq N-1$ , por lo tanto,  $0 < 2\pi u/N \leq 2\pi \frac{N-1}{N} < 2\pi$ , luego  $e^{i2\pi u/N} \neq 1$ .

Así, finalmente tenemos que

$$a_t = \frac{1}{N} \sum_{k=0}^{N-1} A_k e^{i2\pi tk/N}. \quad (2.22)$$

Pasemos ahora a estudiar la relación de Parseval para sucesiones finitas, supóngase que  $\{a_t : t = 0, 1, \dots, N-1\}$  y  $\{b_t : t = 0, 1, \dots, N-1\}$  son dos sucesiones tales que  $\{a_t\} \leftrightarrow \{A_k\}$  y  $\{b_t\} \leftrightarrow \{B_k\}$  entonces

$$\begin{aligned} \sum_{t=0}^{N-1} a_t \bar{b}_t &= \sum_{t=0}^{N-1} \left( \frac{1}{N} \sum_{k=0}^{N-1} A_k e^{i2\pi tk/N} \right) \left( \frac{1}{N} \sum_{l=0}^{N-1} \bar{B}_l e^{-i2\pi tl/N} \right) = \frac{1}{N^2} \sum_{t=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} A_k \bar{B}_l e^{i2\pi(k-l)t/N} \\ &= \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \sum_{t=0}^{N-1} A_k \bar{B}_l e^{i2\pi(k-l)t/N} = \frac{1}{N^2} \sum_{k=0}^{N-1} A_k \sum_{l=0}^{N-1} \bar{B}_l \sum_{t=0}^{N-1} e^{i2\pi(k-l)t/N} = \frac{1}{N} \sum_{k=0}^{N-1} A_k \bar{B}_k \end{aligned}$$

la última igualdad se obtiene utilizando el mismo razonamiento que antes. Si hacemos  $b_t = a_t$  en la relación anterior, obtenemos el teorema de Parseval (enunciado al final de la sección (2.1), expresión (2.4)), para sucesiones finitas, el cual establece que

$$\sum_{t=0}^{N-1} |a_t|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |A_k|^2. \quad (2.23)$$

Ahora definamos la *convolución circular* de dos sucesiones finitas. Sean  $\{a_t\}$  y  $\{b_t\}$  dos sucesiones finitas de longitud  $N$  tales que  $\{a_t\} \leftrightarrow \{A_k\}$  y  $\{b_t\} \leftrightarrow \{B_k\}$ . Se define la *convolución circular* de  $\{a_t\}$  y  $\{b_t\}$  como la sucesión de  $N$  números dada por

$$a * b_t \equiv \sum_{u=0}^{N-1} a_u b_{t-u}, \text{ para } t = 0, 1, 2, \dots, N-1, \quad (2.24)$$

donde  $b_t$  se define para  $t \leq -1$  utilizando su extensión periódica, es decir,  $b_{-1} = b_{N-1}$ ,  $b_{-2} = b_{N-2}, \dots, b_{1-N} = b_1$ ,  $b_{-N} = b_0$ , etc., una forma equivalente de expresar lo anterior es escribir

$$a * b_t \equiv \sum_{u=0}^{N-1} a_u b_{t-u \bmod N},$$

donde  $t - u \bmod N$  significa el entero  $t - u$  módulo  $N$ . Recordemos que si  $0 \leq j \leq N - 1$ , entonces  $j \bmod N = j$  y si  $j$  es un entero fuera de este rango, entonces  $j \bmod N = j + nN$ , donde  $nN$  es el único múltiplo entero de  $N$  tal que  $0 \leq j + nN \leq N - 1$ .

En forma análoga a como se hizo para sucesiones infinitas, se puede verificar que la *DFT* de una convolución circular preserva las siguientes relaciones con las *DFT* de  $\{a_t\}$  y  $\{b_t\}$ .

Para comenzar se tiene la siguiente relación de Parseval

$$\sum_{t=0}^{N-1} a * b_t e^{-i2\pi tk/N} = A_k B_k, \quad (2.25)$$

es decir,

$$\{a * b_t\} \leftrightarrow \{A_k B_k\}.$$

Lo anterior es cierto puesto que

$$\begin{aligned} \sum_{t=0}^{N-1} a * b_t e^{-i2\pi tk/N} &= \sum_{t=0}^{N-1} \left( \sum_{u=0}^{N-1} a_u b_{t-u \bmod N} \right) e^{-i2\pi tk/N} = \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} a_u b_{t-u \bmod N} e^{-i2\pi tk/N} \\ &= \sum_{u=0}^{N-1} \sum_{t=0}^{N-1} a_u b_{t-u \bmod N} e^{-i2\pi tk/N} = \sum_{u=0}^{N-1} a_u \sum_{t=0}^{N-1} b_{t-u \bmod N} e^{-i2\pi tk/N} \\ &= \sum_{u=0}^{N-1} a_u \sum_{s=-u}^{N-1-u} b_s \bmod N e^{-i2\pi(s+u)k/N} = \sum_{u=0}^{N-1} \left( a_u e^{-i2\pi uk/N} \sum_{s=-u}^{N-1-u} b_s \bmod N e^{-i2\pi sk/N} \right) \\ &= \sum_{u=0}^{N-1} a_u e^{-i2\pi uk/N} \sum_{s=0}^{N-1} b_s e^{-i2\pi sk/N} = A_k B_k, \end{aligned}$$

donde se hizo la sustitución  $s = t - u$ , y además se utilizó el hecho de que la sucesión  $\{b_s \bmod N e^{-i2\pi sk/N}\}_s$  es periódica con periodo  $N$ .

De igual manera se tiene la definición análoga para la correlación cruzada circular:

$$\bar{a} * a_t \equiv \sum_{u=0}^{N-1} \bar{a}_u a_{u+t \bmod N}.$$

Por lo tanto se tiene que la DFT asociada a  $\{\bar{a} * b_t\}$  está dada por

$$\sum_{t=0}^{N-1} \bar{a} * b_t e^{-i2\pi tk/N} = \bar{A}_k B_k$$

para  $k = 0, 1, 2, \dots, N-1$ , es decir,

$$\{\bar{a} * b_t\} \leftrightarrow \bar{A}_k B_k.$$

A saber:

$$\begin{aligned} \sum_{t=0}^{N-1} \bar{a} * b_t e^{-i2\pi tk/N} &= \sum_{t=0}^{N-1} \left( \sum_{u=0}^{N-1} \bar{a}_u b_{u+t \bmod N} \right) e^{-i2\pi tk/N} \\ &= \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} \bar{a}_u b_{u+t \bmod N} e^{-i2\pi tk/N} = \sum_{u=0}^{N-1} \sum_{t=0}^{N-1} \bar{a}_u b_{u+t \bmod N} e^{-i2\pi tk/N} \\ &= \sum_{u=0}^{N-1} \bar{a}_u \sum_{t=0}^{N-1} b_{u+t \bmod N} e^{-i2\pi tk/N} = \sum_{u=0}^{N-1} \left( \bar{a}_u \sum_{s=u}^{N-1+u} b_s \bmod N e^{-i2\pi(s-u)k/N} \right) \\ &= \sum_{u=0}^{N-1} \bar{a}_u e^{i2\pi uk/N} \sum_{s=0}^{N-1} b_s e^{-i2\pi sk/N} = \bar{A}_k B_k, \end{aligned}$$

donde nuevamente se hizo la sustitución  $s = u + t$  y se utilizó el hecho de que la sucesión  $\{b_s \bmod N e^{-i2\pi sk/N}\}_s$  es periódica con periodo  $N$ . Por tanto como consecuencia se tiene que

$$\sum_{t=0}^{N-1} \bar{a} * a_t e^{-i2\pi tk/N} = \bar{A}_k A_k = |A_k|^2$$

es decir,

$$\{\bar{a} * a_t\} \leftrightarrow |A_k|^2. \quad (2.26)$$

Ahora consideremos a la sucesión finita  $\{b_t : t = 0, 1, 2, \dots, N-1\}$  y la sucesión  $\{a_t : t = 0, \pm 1, \pm 2, \dots\}$  no necesariamente finita, pero tal que  $\sum_{t \in \mathbb{Z}} |a_t| < \infty$ , entonces se define una nueva sucesión:

$$c_t \equiv \sum_{v=-\infty}^{\infty} a_v b_{t-v \bmod N}, \quad t = 0, 1, \dots, N-1.$$

Lo anterior se puede escribir como la operación de aplicar a  $\{b_t\}$  un filtro circular de longitud  $N$ , veamos a que nos referimos con lo anterior:

$$c_t = \sum_{v=-\infty}^{\infty} a_v b_{t-v \bmod N} = \sum_{n=-\infty}^{\infty} \sum_{v=nN}^{(n+1)N-1} a_v b_{t-v \bmod N} = \sum_{n=-\infty}^{\infty} \sum_{v=0}^{N-1} a_{v+nN} b_{t-(v+nN) \bmod N}.$$

pero como  $b_{t-(v+nN) \bmod N} = b_{t-v \bmod N}$ , para todo entero  $n$ , entonces podemos escribir

$$\begin{aligned} c_t &= \sum_{n=-\infty}^{\infty} \sum_{v=0}^{N-1} a_{v+nN} b_{t-v \bmod N} = \sum_{v=0}^{N-1} \sum_{n=-\infty}^{\infty} a_{v+nN} b_{t-v \bmod N} \\ &= \sum_{v=0}^{N-1} \left( \sum_{n=-\infty}^{\infty} a_{v+nN} \right) b_{t-v \bmod N} = \sum_{v=0}^{N-1} a_v^{\circ} b_{t-v \bmod N}, \end{aligned}$$

donde

$$a_v^{\circ} = \sum_{n=-\infty}^{\infty} a_{v+nN}. \quad (2.27)$$

El filtro anterior  $\{a_v^{\circ}\}$  está construido a partir de cortar el filtro  $\{a_t\}$  en vectores de longitud  $N$ , es decir

$$\dots, \underline{a_{-N}, a_{-N+1}, \dots, a_{-2}, a_{-1}}_{n=-1}, \underline{a_0, a_1, \dots, a_{N-2}, a_{N-1}}_{n=0}, \underline{a_N, a_{N+1}, \dots, a_{2N-1}}_{n=1}, \dots$$

para obtener el  $t$ -ésimo elemento de  $\{a_t^{\circ}\}$  sumamos la  $t$ -ésima componente de todos los vectores

$$\begin{array}{cccccc}
 \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\
 + & + & + & & + & + \\
 a_{-N} & a_{-N+1} & a_{-N+2} & & a_{-2} & a_{-1} \\
 + & + & + & & + & + \\
 a_0 & a_1 & a_2 & & a_{N-2} & a_{N-1} \\
 + & + & + & & + & + \\
 a_N & a_{N+1} & a_{N+2} & & a_{2(N-1)} & a_{2N-1} \\
 + & + & + & & + & + \\
 \vdots & \vdots & \vdots & & \vdots & \vdots \\
 \hline
 a_0^\circ & a_1^\circ & a_2^\circ & \dots & a_{N-2}^\circ & a_{N-1}^\circ
 \end{array}$$

Sea  $\{A_k^\circ : k = 0, 1, 2, \dots, N-1\}$  la transformada discreta de Fourier asociada a  $\{a_t^\circ\}$  y sea  $\{A_k : k = 0, 1, 2, \dots, N-1\}$  la correspondiente a  $\{a_t\}$ , entonces se tiene

$$A_k^\circ = A\left(\frac{k}{N}\right),$$

es decir,

$$\{a_t^\circ\} \leftrightarrow \left\{ A_k^\circ = A\left(\frac{k}{N}\right) \right\}. \quad (2.28)$$

Veamos porque se afirma lo anterior:

$$A_k^\circ = \sum_{u=0}^{N-1} a_u^\circ e^{-i2\pi uk/N} = \sum_{u=0}^{N-1} \left( \sum_{n=-\infty}^{\infty} a_{u+nN} e^{-i2\pi uk/N} \right). \quad (2.29)$$

Dado que

$$e^{-i2\pi nNk/N} = e^{-i2\pi nk} = 1,$$

entonces se tiene

$$e^{-i2\pi(u+nN)k/N} = e^{-i2\pi uk/N} e^{-i2\pi nNk/N} = e^{-i2\pi uk/N},$$

de donde

$$\begin{aligned} A_k^o &= \sum_{u=0}^{N-1} \sum_{n=-\infty}^{\infty} a_{u+nN} e^{-i2\pi(u+nN)k/N} = \sum_{n=-\infty}^{\infty} \sum_{u=0}^{N-1} a_{u+nN} e^{-i2\pi(u+nN)k/N} \\ &= \sum_{n=-\infty}^{\infty} \sum_{t=nN}^{(n+1)N-1} a_t e^{-i2\pi tk/N} = \sum_{t=-\infty}^{\infty} a_t e^{-i2\pi tk/N} = A\left(\frac{k}{N}\right). \end{aligned}$$

## 2.3. Teoría de Onduletas.

Sea  $\psi_{a,b}(x)$ , para  $a \in \mathbb{R}$ , con  $a \neq 0$  y  $b \in \mathbb{R}$  una familia de funciones definidas como traslaciones y re-escalaciones de la función  $\psi(x)$  cuadrado integrable, tal que

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right). \quad (2.30)$$

La función  $\psi$ , *wavelet madre*, se supone que satisface la condición de admisibilidad

$$C_\psi = \int_{\mathbb{R}} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \quad (2.31)$$

donde  $\Psi(\omega)$  es la transformada de Fourier de  $\psi(\omega)$ . Si suponemos la condición de admisibilidad (2.31) entonces se tiene que

$$0 = \Psi(0) = \int_{-\infty}^{\infty} \psi(x) dx. \quad (2.32)$$

Esta última propiedad es la que motiva el nombre de onduleta. Las funciones de onduletas usualmente consideradas son aquellas que son normalizadas, es decir, aquellas con la propiedad  $\|\Psi_{a,b}(x)\| = 1$ . Para cualquier función cuadrado integrable  $f(x)$  la transformada continua de onduletas se define como la función de dos variables

$$CWT_f = \int_{-\infty}^{\infty} f(x) \overline{\psi_{a,b}(x)} dx \quad (2.33)$$

donde  $a$  se le llama parámetro de dilatación y  $b$  parámetro de traslación, los cuales varían en  $\mathbb{R}$  excepto en 0 y  $\mathbb{R}$  respectivamente. Algunas propiedades básicas que a continuación sólo enunciaremos se pueden ver con más detalle en Vidakovic (1999)

- (i) Si  $f(x)$  tiene transformada continua de onduletas  $CWT_f(a, b)$ , entonces  $g(x) = f(x - \beta)$  tiene transformada continua de onduletas  $CWT_g(a, b) = CWT_f(a, b - \beta)$ .
- (ii) Si  $f(x)$  tiene transformada continua de onduleta  $CWT_f(a, b)$ , entonces  $g(x) = \frac{1}{\sqrt{s}}f\left(\frac{x}{s}\right)$  tiene transformada continua de onduletas  $CWT_g(a, b) = CWT_f\left(\frac{a}{s}, \frac{b}{s}\right)$ .
- (iii) Sea  $f(x) = \delta(x - x_0)$  la delta de Dirac en  $x_0$ , entonces la transformada continua de onduletas de  $f$  está dada por

$$CWT_f(a, b) = \frac{1}{\sqrt{a}}\psi\left(\frac{x_0 - b}{a}\right). \quad (2.34)$$

La primer onduleta que se conoce es la Onduleta de Haar, la cual está definida de la siguiente manera:

$$\psi_{a,b}^H(x) = \frac{1}{\sqrt{a}} \left[ \mathbb{1}\left(b \leq x \leq \frac{a}{2} + b\right) - \mathbb{1}\left(\frac{a}{2} + b \leq x \leq a + b\right) \right], \quad (2.35)$$

donde  $a = 1$  y  $b = 0$ . Es decir,

$$\psi(x) = \begin{cases} -1 & \frac{1}{2} \leq x, \\ 1 & 0 \leq x < \frac{1}{2}, \\ 0 & \text{en otro caso.} \end{cases} \quad (2.36)$$

Dado que nuestro principal interés es la Transformada Discreta de Onduletas, la cual se abordará en el siguiente capítulo, en vías de enfocar el trabajo en esa dirección diremos que la manera usual de considerar los valores de  $a, b$  es por medio de términos diádicos, es decir, valores de la forma  $a = 2^{-j}$  y  $b = k2^{-j}$ , donde  $j, k \in \mathbb{Z}$ . Bajo condiciones apropiadas en la función  $\psi$  se obtiene la base ortonormal

$$\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), j, k \in \mathbb{Z}\}$$

para más detalles sobre lo anterior ver Vetterli y Kovacevic (1995).

# Capítulo 3

## Transformadas Discretas de Onduletas

En la primera parte de este capítulo se abordarán las principales propiedades y conceptos necesarios para comprender el funcionamiento de la transformada discreta de onduletas. Posteriormente estudiaremos una colección de transformaciones discretas conocidas como Paquetes de Transformadas de Onduletas (DWPT, del inglés *Discrete Wavelet Packet Transform*). Este tipo de transformación nos será útil posteriormente para instrumentar esquemas de remuestreo en series de tiempo, los cuales se abordarán en el capítulo 4.

### 3.1. Transformada Discreta de Fourier

Consideremos la matriz  $\mathcal{O}$  de dimensión  $N \times N$  con valores reales, tal que satisface la siguiente propiedad:

$$\mathcal{O}^T \mathcal{O} = I_N,$$

donde  $I_N$  es la matriz identidad de dimensiones  $N \times N$ . Entonces para  $j, k = 0, 1, 2, \dots, N-1$  sean  $\mathcal{O}_{j\bullet}^T$  el  $j$ -ésimo vector renglón y  $\mathcal{O}_{\bullet k}$  el  $k$ -ésimo vector columna de  $\mathcal{O}$ , entonces podemos escribir a  $\mathcal{O}$  en la siguiente forma:

$$\mathcal{O} = \begin{bmatrix} \mathcal{O}_{0\bullet}^T \\ \mathcal{O}_{1\bullet}^T \\ \vdots \\ \mathcal{O}_{N-1\bullet}^T \end{bmatrix} = \begin{bmatrix} \mathcal{O}_{\bullet 0} & \mathcal{O}_{\bullet 1} & \mathcal{O}_{\bullet 2} & \cdots & \mathcal{O}_{\bullet N-1} \end{bmatrix}.$$



Aplicando la ortonormalidad de  $\mathcal{O}$  se tiene como consecuencia inmediata que

$$\mathcal{O}^T = \mathcal{O}^{-1},$$

es decir

$$\mathcal{O}^T \mathcal{O} = \mathcal{O}^{-1} \mathcal{O}.$$

Entonces se tiene que los renglones y las columnas de  $\mathcal{O}$  forman dos conjuntos ortonormales de vectores en  $\mathbb{R}^N$ . Con base en lo anterior podemos hacer una descomposición de  $\mathbf{X} = \{X_0, X_1, \dots, X_{N-1}\}^T$  por medio de la siguiente expresión

$$\mathbf{O} = \mathcal{O}\mathbf{X} = \begin{bmatrix} \mathcal{O}_{0\bullet}^T \\ \mathcal{O}_{1\bullet}^T \\ \vdots \\ \mathcal{O}_{N-1\bullet}^T \end{bmatrix} \mathbf{X} = \begin{bmatrix} \mathcal{O}_{0\bullet}^T \mathbf{X} \\ \mathcal{O}_{1\bullet}^T \mathbf{X} \\ \vdots \\ \mathcal{O}_{N-1\bullet}^T \mathbf{X} \end{bmatrix}, \quad (3.1)$$

a esta ecuación se le llama *ecuación de análisis*.

A los elementos del vector columna  $\mathbf{O}$ , de dimensión  $N$ , se les conoce como los coeficientes de transformación para  $\mathbf{X}$  con respecto a la transformación ortonormal  $\mathcal{O}$ ; el  $j$ -ésimo coeficiente de transformación es  $O_j$ , es decir, el  $j$ -ésimo elemento de  $\mathbf{O}$  está dado por

$$O_j = \sum_{k=0}^{N-1} \mathcal{O}_{jk} X_k. \quad (3.2)$$

Si multiplicamos por la izquierda a la ecuación (3.1) por  $\mathcal{O}^T$  y utilizamos la propiedad de ortonormalidad de la matriz  $\mathcal{O}$  tenemos

$$\begin{aligned} \mathbf{X} &= \mathcal{O}^T \mathbf{O} = [\mathcal{O}_{0\bullet}, \mathcal{O}_{1\bullet}, \dots, \mathcal{O}_{N-1\bullet}] \begin{bmatrix} O_0 \\ O_1 \\ \vdots \\ O_{N-1} \end{bmatrix} \\ &= \mathcal{O}_{0\bullet} O_0 + \mathcal{O}_{1\bullet} O_1 + \dots + \mathcal{O}_{N-1\bullet} O_{N-1} = \sum_{j=0}^{N-1} O_j \mathcal{O}_{j\bullet}, \end{aligned}$$

es decir,

$$\mathbf{X} = \sum_{j=0}^{N-1} \mathbf{O}_j \mathcal{O}_{j\bullet}, \quad (3.3)$$

a esta ecuación se le llama *ecuación de síntesis*.

Entonces lo que se tiene es que se puede recuperar a  $\mathbf{X}$  por medio de sus coeficientes de transformación  $\mathbf{O}$ . Utilizando la ecuación (3.2) podemos reescribir la expresión dada en (3.3) como

$$\mathbf{X} = \sum_{j=0}^{N-1} \left( \sum_{k=0}^{N-1} \mathcal{O}_{jk} X_k \right) \mathcal{O}_{j\bullet}. \quad (3.4)$$

Otra propiedad importante es la siguiente:

$$\begin{aligned} \|\mathbf{O}\|^2 &= \mathbf{O}^T \mathbf{O} = (\mathcal{O}\mathbf{X})^T (\mathcal{O}\mathbf{X}) \\ &= \mathbf{X}^T \mathcal{O}^T \mathcal{O} \mathbf{X} = \mathbf{X}^T \mathbf{X} = \|\mathbf{X}\|^2. \end{aligned}$$

Además tenemos que

$$\|\mathbf{X}\|^2 = \sum_{j=0}^{N-1} \|\mathbf{O}_j \mathcal{O}_{j\bullet}\|^2, \quad (3.5)$$

lo anterior es cierto puesto que:

$$\begin{aligned} \|\mathbf{X}\|^2 &= \left\| \sum_{j=0}^{N-1} \mathbf{O}_j \mathcal{O}_{j\bullet} \right\|^2 = \left( \sum_{j=0}^{N-1} \mathbf{O}_j \mathcal{O}_{j\bullet} \right)^T \left( \sum_{i=0}^{N-1} \mathbf{O}_i \mathcal{O}_{i\bullet} \right) \\ &= (\mathbf{O}_0^T \mathcal{O}_{0\bullet}^T + \mathbf{O}_1^T \mathcal{O}_{1\bullet}^T + \dots + \mathbf{O}_{N-1}^T \mathcal{O}_{N-1\bullet}^T) (\mathcal{O}_{0\bullet} \mathbf{O}_0 + \mathcal{O}_{1\bullet} \mathbf{O}_1 + \dots + \mathcal{O}_{N-1\bullet} \mathbf{O}_{N-1}) \\ &= \mathbf{O}_0^T \mathcal{O}_{0\bullet}^T \mathcal{O}_{0\bullet} \mathbf{O}_0 + \mathbf{O}_0^T \mathcal{O}_{0\bullet}^T \mathcal{O}_{1\bullet} \mathbf{O}_1 + \dots + \mathbf{O}_0^T \mathcal{O}_{0\bullet}^T \mathcal{O}_{N-1\bullet} \mathbf{O}_{N-1} \\ &+ \mathbf{O}_1^T \mathcal{O}_{1\bullet}^T \mathcal{O}_{0\bullet} \mathbf{O}_0 + \mathbf{O}_1^T \mathcal{O}_{1\bullet}^T \mathcal{O}_{1\bullet} \mathbf{O}_1 + \dots + \mathbf{O}_1^T \mathcal{O}_{1\bullet}^T \mathcal{O}_{N-1\bullet} \mathbf{O}_{N-1} \\ &+ \dots + \mathbf{O}_{N-1}^T \mathcal{O}_{N-1\bullet}^T \mathcal{O}_{0\bullet} \mathbf{O}_0 + \mathbf{O}_{N-1}^T \mathcal{O}_{N-1\bullet}^T \mathcal{O}_{1\bullet} \mathbf{O}_1 + \dots + \mathbf{O}_{N-1}^T \mathcal{O}_{N-1\bullet}^T \mathcal{O}_{N-1\bullet} \mathbf{O}_{N-1} \\ &= \sum_{j=0}^j \mathbf{O}_j^T \mathcal{O}_{j\bullet}^T \mathcal{O}_{j\bullet} \mathbf{O}_j = \sum_{j=0}^j (\mathcal{O}_{j\bullet} \mathbf{O}_j)^T (\mathcal{O}_{j\bullet} \mathbf{O}_j) = \sum_{j=0}^{N-1} \|\mathcal{O}_{j\bullet} \mathbf{O}_j\|^2 = \sum_{j=0}^{N-1} \|\mathbf{O}_j \mathcal{O}_{j\bullet}\|^2 \end{aligned}$$

en la quinta igualdad se utilizó el hecho de que  $\mathcal{O}$  es una matriz ortonormal.

Con miras de avanzar hacia nuestro objetivo principal que es la transformada discreta de onduletas consideremos al vector  $X = \{X_0, X_1, \dots, X_{N-1}\}$  para la cual se define su transformada discreta de Fourier  $\{F_k\}$  como:

$$F_k \equiv \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi tk/N}, \quad k = 0, 1, 2, \dots, N-1, \quad (3.6)$$

al número complejo  $F_k$  se le llama el  $k$ -ésimo *coeficiente de Fourier* de  $X$ . Notemos que

$$F_0 = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi t(0)/N} = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t = \sqrt{N}\bar{X},$$

por lo tanto,  $F_0$  es real si  $\{X_t : t = 0, 1, \dots, N-1\}$  es una serie de tiempo con valores reales. Bajo esta misma hipótesis consideremos que para  $k$  tal que  $1 \leq k \leq N/2$ :

$$\begin{aligned} F_{N-k} &= \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi t(N-k)/N} = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi tN/N} e^{i2\pi tk/N} = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi t} e^{i2\pi tk/N} \\ &= \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t (1)^t e^{i2\pi tk/N} = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi tk/N} = \bar{F}_k, \end{aligned}$$

Finalmente supongamos que  $N$  es par, es decir,  $N = 2k$  para algún valor de  $k$  en  $\{1, 2, \dots\}$ , entonces se tiene que  $\frac{N}{2} = k$ , por lo tanto

$$\begin{aligned} F_{\frac{N}{2}} &= \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi t(N/2)/N} = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi t(1/2)} \\ &= \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i\pi t} = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t (-1)^t \end{aligned}$$

el cual nuevamente es un número real. Concluimos que los coeficientes de Fourier  $\{F_k\}$  quedan determinados por  $N$  valores reales, a saber,  $F_0, F_{N/2}$  (cuando  $N$  es par) y las componentes real e imaginaria de los números complejos  $F_k$ ,  $1 \leq k < N/2$ .

La ecuación (3.6) se puede reescribir como

$$\mathbf{F} = \mathcal{F}\mathbf{X}, \quad (3.7)$$

donde  $\mathbf{F}$  es un vector columna de dimensión  $N$  con elementos

$$\mathbf{F} = \begin{bmatrix} F_0 \\ F_1 \\ \vdots \\ F_{N-1} \end{bmatrix}$$

y  $\mathcal{F}$  es una matriz de dimensiones  $N \times N$ , cuyo  $(k_0, t_0)$ -ésimo elemento está determinado por

$$\frac{1}{\sqrt{N}} e^{-i2\pi t_0 k_0 / N},$$

para  $0 \leq k_0, t_0 \leq N - 1$ . Por ejemplo

$$F_0 = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi t(0)/N} = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t = \left[ \frac{1}{\sqrt{N}}, \frac{1}{\sqrt{N}}, \dots, \frac{1}{\sqrt{N}} \right] \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{N-1} \end{bmatrix},$$

$$F_1 = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi t(1)/N} = \frac{1}{\sqrt{N}} [e^{-i2\pi(0)/N}, e^{-i2\pi(1)/N}, \dots, e^{-i2\pi(N-1)/N}] \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{N-1} \end{bmatrix},$$

$$F_2 = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi t(2)/N} = \frac{1}{\sqrt{N}} [e^{-i4\pi(0)/N}, e^{-i4\pi(1)/N}, \dots, e^{-i4\pi(N-1)/N}] \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{N-1} \end{bmatrix},$$

así sucesivamente hasta  $N - 1$

$$\begin{aligned}
 F_{N-1} &= \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi t(N-1)/N} \\
 &= \frac{1}{\sqrt{N}} \left[ e^{-i2\pi(N-1)(0)/N}, e^{-i2\pi(N-1)(1)/N}, \dots, e^{-i2\pi(N-1)(N-1)/N} \right] \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{N-1} \end{bmatrix},
 \end{aligned}$$

es decir,

$$\begin{aligned}
 \begin{bmatrix} F_0 \\ F_1 \\ \vdots \\ F_{N-1} \end{bmatrix} &= \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{-i2\pi(0)/N} & e^{-i2\pi(1)/N} & \dots & e^{-i2\pi(N-1)/N} \\ e^{-i2\pi(0)2/N} & e^{-i2\pi(1)2/N} & \dots & e^{-i2\pi(N-1)2/N} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-i2\pi(\frac{N}{2}-1)(0)/N} & e^{-i2\pi(\frac{N}{2}-1)(1)/N} & \dots & e^{-i2\pi(\frac{N}{2}-1)(N-1)/N} \\ e^{-i2\pi(\frac{N}{2})(0)/N} & e^{-i2\pi(\frac{N}{2})(1)/N} & \dots & e^{-i2\pi(\frac{N}{2})(N-1)/N} \\ e^{-i2\pi(\frac{N}{2}+1)(0)/N} & e^{-i2\pi(\frac{N}{2}+1)(1)/N} & \dots & e^{-i2\pi(\frac{N}{2}+1)(N-1)/N} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-i2\pi(N-1)(0)/N} & e^{-i2\pi(N-1)(1)/N} & \dots & e^{-i2\pi(N-1)(N-1)/N} \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{N-1} \end{bmatrix} \\
 &= \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & e^{-i2\pi/N} & \dots & e^{-i2\pi(N-1)/N} \\ 1 & e^{-i2\pi 2(1)/N} & \dots & e^{-i2\pi 2(N-1)(N-1)/N} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-i2\pi(\frac{N}{2}-1)/N} & \dots & e^{-i2\pi(\frac{N}{2}-1)(N-1)/N} \\ 1 & e^{-i2\pi(\frac{N}{2})/N} & \dots & e^{-i2\pi(\frac{N}{2})(N-1)/N} \\ 1 & e^{-i2\pi(\frac{N}{2}+1)/N} & \dots & e^{-i2\pi(\frac{N}{2}+1)(N-1)/N} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-i2\pi(N-1)/N} & \dots & e^{-i2\pi(N-1)(N-1)/N} \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{N-1} \end{bmatrix}.
 \end{aligned}$$

En particular estaremos interesados en longitudes de la forma  $N = 2^J$  dado que como se verá más adelante la descomposición de la transformada discreta de onduletas se realizará por

medio de cierto tipo de vectores de longitud  $2^J$ . A continuación presentamos un ejemplo para cuando  $J = 4$

$$\begin{bmatrix} F_0 \\ F_1 \\ \vdots \\ F_{15} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & e^{-i2\pi(1)1/16} & \dots & e^{-i2\pi(15)1/16} \\ 1 & e^{-i2\pi(1)2/16} & \dots & e^{-i2\pi(15)2/16} \\ 1 & e^{-i2\pi(1)3/16} & \dots & e^{-i2\pi(15)3/16} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-i2\pi(1)7/16} & \dots & e^{-i2\pi(15)7/16} \\ 1 & e^{-i2\pi(1)8/16} & \dots & e^{-i2\pi(15)8/16} \\ 1 & e^{-i2\pi(1)9/16} & \dots & e^{-i2\pi(15)9/16} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-i2\pi(1)15/16} & \dots & e^{-i2\pi(15)15/16} \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{15} \end{bmatrix}$$

De acuerdo a la descripción dada para la matriz  $\mathcal{F}$  que representa a la DFT, tenemos que  $\mathcal{F}^H \mathcal{F} = \mathbb{I}_{N \times N}$ , es decir,  $\mathcal{F}$  es una matriz ortonormal. Lo anterior se sigue de que el producto del  $k$ -ésimo renglón de  $\mathcal{F}^H$  con la  $l$ -ésima columna de  $\mathcal{F}$  equivale a

$$\frac{1}{N} \sum_{t=0}^{N-1} e^{-i2\pi t(l-k)/N} = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$$

lo cual ya se demostró en el capítulo 2, ecuación (2.21). A raíz de lo anterior tenemos

$$\mathbf{X} = \mathcal{F}^H \mathbf{F}, \quad (3.8)$$

de donde se tienen las siguientes igualdades:

$$\begin{aligned} \|\mathbf{F}\|^2 &= \|\mathcal{F}\mathbf{X}\|^2 = (\mathcal{F}\mathbf{X})^H (\mathcal{F}\mathbf{X}) \\ &= \mathbf{X}^H \mathcal{F}^H \mathcal{F} \mathbf{X} = \mathbf{X}^H \mathbf{X} = \|\mathbf{X}\|^2 \end{aligned}$$

por lo tanto

$$\|\mathbf{F}\|^2 = \|\mathbf{X}\|^2. \quad (3.9)$$

Se puede escribir a  $\|\mathbf{F}\|$  en la siguiente forma:

$$\|\mathbf{F}\|^2 = \mathbf{F}^H \mathbf{F} = \sum_{j=0}^{N-1} |\mathbf{F}_j|^2 = |\mathbf{F}_0|^2 + |\mathbf{F}_{N/2}|^2 + 2 \sum_{j=1}^{N/2-1} |\mathbf{F}_j|^2, \quad (3.10)$$

si además utilizamos el hecho de que

$$F_0 = \sqrt{N\bar{X}}$$

podemos obtener una expresión en estos términos para la varianza del vector  $\mathbf{X} = \{X_0, X_1, \dots, X_{N-1}\}$  como se muestra a continuación

$$\begin{aligned} \hat{\sigma}_X^2 &= \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \bar{X})^2 = \frac{1}{N} \sum_{t=0}^{N-1} X_t^2 - \bar{X}^2 = \frac{1}{N} \|\mathbf{X}\|^2 - \bar{X}^2 \\ &= \frac{1}{N} \|\mathbf{F}\|^2 - \bar{X}^2 = \frac{1}{N} \left( |\mathbf{F}_0|^2 + |\mathbf{F}_{N/2}|^2 + 2 \sum_{j=1}^{N/2-1} |\mathbf{F}_j|^2 \right) - \bar{X}^2 \\ &= \frac{1}{N} \left( |\sqrt{N\bar{X}}|^2 + |\mathbf{F}_{N/2}|^2 + 2 \sum_{j=1}^{N/2-1} |\mathbf{F}_j|^2 \right) - \bar{X}^2 = \frac{1}{N} |\mathbf{F}_{N/2}|^2 + \frac{2}{N} \sum_{j=1}^{N/2-1} |\mathbf{F}_j|^2 \end{aligned}$$

Así,  $2 |\mathbf{F}_k|^2 / N$  representa la contribución a la varianza muestral de  $\{X_t\}$  asociada a la frecuencia  $f_k$ , para  $0 < k < N/2$ .

Escribamos la ecuación (3.8) en forma análoga a la ecuación (3.4)

$$\mathbf{X} = \mathcal{F}^H \mathbf{F} = \sum_{j=0}^{N-1} \left( \sum_{k=0}^{N-1} \mathcal{F}_{jk} X_k \right) \mathcal{F}_{j\bullet} = \sum_{j=0}^{N-1} F_j \mathcal{F}_{j\bullet}, \quad (3.11)$$

es decir,

$$\mathbf{X} = \sum_{j=0}^{N-1} F_j \mathcal{F}_{j\bullet}, \quad (3.12)$$

como  $\mathcal{F}_{kt} = \frac{1}{\sqrt{N}} e^{-i2\pi tk/N}$ , entonces tenemos que

$$X_t = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} F_k e^{i2\pi tk/N} \quad \text{para } t = 0, 1, 2, \dots, N-1, \quad (3.13)$$

es decir, hemos desarrollado una expresión para la Transformada Inversa de Fourier Discreta Ortonormal (*IODFT*).

## 3.2. Transformada Discreta de Onduletas

Ahora estudiaremos otro tipo de transformada con propiedades similares a la DFT, esta es la *Transformada Discreta de Onduletas*, *DWT*, del inglés *Discrete Wavelet Transform*; la cual al igual que las de las secciones anteriores es una transformada ortonormal para  $\mathbf{X} = \{X_0, X_1, \dots, X_{N-1}\}^T$ .

La DWT también puede representarse como el producto de una matriz  $N \times N$  dimensional  $\mathcal{W}$  por el vector de datos  $\mathbf{X}$ . Se estudiará a detalle la construcción de los renglones de  $\mathcal{W}$  y se demostrará que bajo tal construcción la matriz resulta ser una matriz ortonormal, pero por el momento sólo se procederá a hacer una descripción cualitativa de esta clase de transformada. Se puede definir una transformada de onduletas dependiendo de una colección finita de números, para nuestra descripción se considerarán reales,  $\{h_0, h_1, \dots, h_{L-1}\}$  a la cual llamaremos *filtro de onduletas*.

En la literatura existen diferentes filtros de onduletas y cada uno de ellos a su vez define una transformada de onduletas, de hecho veremos que los renglones de la matriz  $\mathcal{W}$  serán construidos en función de  $\{h_0, h_1, \dots, h_{L-1}\}$ . Para nuestra descripción, sean  $N = 2^J$ , donde  $J$  es un entero positivo y  $\mathcal{W}$  la matriz  $N \times N$  dimensional que representa a la DWT. Para una muestra de la serie de tiempo  $\mathbf{X} = \{X_0, X_1, \dots, X_{N-1}\}^T$ , podemos escribir al vector  $\mathbf{W}$  resultante de aplicar la transformación a  $\mathbf{X}$  como

$$\mathbf{W} = \mathcal{W}\mathbf{X}. \quad (3.14)$$

En forma similar a como se hizo anteriormente, sean  $\{W_n : n = 0, 1, 2, \dots, N-1\}$  los coeficientes de la DWT, podemos escribir a la transformada discreta de onduletas en la forma (3.14), donde  $\mathbf{W}$  es un vector columna de longitud  $N = 2^J$ , cuyo  $n$ -ésimo elemento es el  $n$ -ésimo coeficiente de la DWT,  $W_n$ . La matriz con entradas reales  $\mathcal{W}$  cumple con la propiedad de ortonormalidad, es decir, satisface la expresión

$$\mathcal{W}^t \mathcal{W} = I_N,$$

por lo tanto análogamente a las expresiones (3.8) y (3.9) tenemos las siguientes ecuaciones



$$\mathbf{X} = \mathcal{W}^t \mathbf{W}, \quad (3.15)$$

$$\|\mathbf{W}\|^2 = \|\mathbf{X}\|^2. \quad (3.16)$$

Consideremos el efecto que se tiene al trasladar de manera circular el vector  $\mathbf{X}$ , con la matriz  $\mathcal{T}$  definida como

$$\mathcal{T} = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{bmatrix}. \quad (3.17)$$

Notemos que

$$\mathcal{T}^{-1} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix}. \quad (3.18)$$

Ambas matrices de dimensión  $N \times N$  satisfacen lo siguiente:

$$\begin{aligned} \mathcal{T}\mathbf{X} &= [X_{N-1}, X_0, X_1, \dots, X_{N-3}, X_{N-2}]^T & \text{y} \\ \mathcal{T}^{-1}\mathbf{X} &= [X_1, X_2, X_3, \dots, X_{N-2}, X_{N-1}, X_0]^T. \end{aligned}$$

Además se tiene que

$$\begin{aligned} \mathcal{T}^2\mathbf{X} &= \mathcal{T}\mathcal{T}\mathbf{X} & \text{y} \\ \mathcal{T}^{-2}\mathbf{X} &= \mathcal{T}^{-1}(\mathcal{T}^{-1}\mathbf{X}). \end{aligned}$$

De aquí se tiene por lo tanto que

$$\mathcal{T}^{-1}\mathcal{T} = I_N,$$

así,

$$\mathcal{T}^{-1} = \mathcal{T}^T,$$

es decir,  $\mathcal{T}$  es una transformación ortonormal.

Entonces siguiendo con este razonamiento tenemos las siguientes igualdades

$$\begin{aligned} \mathcal{T}\mathbf{X} &= [X_{N-1}, X_0, X_1, \dots, X_{N-3}, X_{N-2}]^T, \\ \mathcal{T}^2\mathbf{X} &= [X_{N-2}, X_{N-1}, X_0, X_1, \dots, X_{N-4}, X_{N-3}]^T, \\ &\vdots \\ \mathcal{T}^m\mathbf{X} &= [X_{N-m}, X_{N-m+1}, \dots, X_0, X_1, \dots, X_{N-m-1}]^T. \end{aligned}$$

Analicemos un caso muy importante, la transformada discreta de onduletas de *Haar*, históricamente esta es la primera transformada discreta de onduletas ya que los filtros o vectores básicos que la definen son el análogo discreto de las funciones estudiadas por Haar en 1910 para construir bases ortonormales de funciones. Proporcionaremos la forma de la matriz  $\mathcal{W}$  para el caso  $N = 2^4 = 16$ , ( $J = 4$ ). Así los renglones 0, 8, 12, 14 y 15 de la matriz  $\mathcal{W}$  están dados por:

$$\begin{aligned} \mathcal{W}_{0\bullet}^T &= \left[ -1/\sqrt{2}, \quad 1/\sqrt{2}, \quad \underbrace{0, 0, \dots, 0}_{14 \text{ ceros}} \right] \\ \mathcal{W}_{8\bullet}^T &= \left[ -1/2, \quad -1/2, \quad 1/2, \quad 1/2, \quad \underbrace{0, 0, \dots, 0}_{12 \text{ ceros}} \right] \\ \mathcal{W}_{12\bullet}^T &= \left[ -\frac{1}{\sqrt{8}}, \quad -\frac{1}{\sqrt{8}}, \quad -\frac{1}{\sqrt{8}}, \quad -\frac{1}{\sqrt{8}}, \quad \frac{1}{\sqrt{8}}, \quad \frac{1}{\sqrt{8}}, \quad \frac{1}{\sqrt{8}}, \quad \frac{1}{\sqrt{8}}, \quad \underbrace{0, 0, \dots, 0}_{8 \text{ ceros}} \right] \\ \mathcal{W}_{14\bullet}^T &= \left[ \underbrace{-1/4, \dots, -1/4}_8 \text{ entradas}, \quad \underbrace{1/4, \dots, 1/4}_8 \text{ entradas} \right] \\ \mathcal{W}_{15\bullet}^T &= \left[ \underbrace{1/4, 1/4, \dots, 1/4}_{16 \text{ entradas}} \right] \end{aligned}$$

Usaremos la matriz dada en (3.17) para definir los otros renglones de  $\mathcal{W}$ . Los vectores restantes son versiones trasladadas circularmente

$$\begin{aligned}
\mathcal{W}_{1\bullet}^T &= (T^2 \mathcal{W}_{0\bullet})^T = \left[ 0, 0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \underbrace{0, 0, \dots, 0}_{12 \text{ ceros}} \right] \\
\mathcal{W}_{2\bullet}^T &= (T^4 \mathcal{W}_{0\bullet})^T = \left[ 0, 0, 0, 0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \underbrace{0, 0, \dots, 0}_{10 \text{ ceros}} \right] \\
&\vdots \\
\mathcal{W}_{7\bullet}^T &= (T^{14} \mathcal{W}_{0\bullet})^T = \left[ \underbrace{0, 0, \dots, 0}_{14 \text{ ceros}}, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right] \\
\mathcal{W}_{9\bullet}^T &= (T^4 \mathcal{W}_{8\bullet})^T = \left[ 0, 0, 0, 0, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \underbrace{0, 0, \dots, 0}_{8 \text{ ceros}} \right] \\
\mathcal{W}_{10\bullet}^T &= (T^8 \mathcal{W}_{8\bullet})^T = \left[ \underbrace{0, 0, \dots, 0}_{8 \text{ ceros}}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0 \right] \\
\mathcal{W}_{11\bullet}^T &= (T^{12} \mathcal{W}_{8\bullet})^T = \left[ \underbrace{0, 0, \dots, 0}_{12 \text{ ceros}}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2} \right] \\
\mathcal{W}_{13\bullet}^T &= (T^8 \mathcal{W}_{12\bullet})^T = \left[ \underbrace{0, 0, \dots, 0}_{8 \text{ ceros}}, -\frac{1}{\sqrt{8}}, -\frac{1}{\sqrt{8}}, -\frac{1}{\sqrt{8}}, -\frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}} \right]
\end{aligned}$$

entonces de acuerdo con la ecuación (3.14) se tienen las siguientes expresiones:

$$\begin{aligned}
W_0 &= \mathcal{W}_{0\bullet}^T X = -\frac{1}{\sqrt{2}} X_0 + \frac{1}{\sqrt{2}} X_1 = \frac{1}{\sqrt{2}} (X_1 - X_0) \\
W_1 &= \mathcal{W}_1^T X = -\frac{1}{\sqrt{2}} X_2 + \frac{1}{\sqrt{2}} X_3 = \frac{1}{\sqrt{2}} (X_3 - X_2) \\
W_2 &= \mathcal{W}_2^T X = -\frac{1}{\sqrt{2}} X_4 + \frac{1}{\sqrt{2}} X_5 = \frac{1}{\sqrt{2}} (X_5 - X_4) \\
&\vdots \\
W_7 &= \mathcal{W}_7^T X = -\frac{1}{\sqrt{2}} X_{14} + \frac{1}{\sqrt{2}} X_{15} = \frac{1}{\sqrt{2}} (X_{15} - X_{14}) \\
W_8 &= \mathcal{W}_8^T X = -\frac{1}{2} X_0 - \frac{1}{2} X_1 + \frac{1}{2} X_2 + \frac{1}{2} X_3 = \frac{1}{2} (X_3 + X_2 - X_1 - X_0) \\
W_9 &= \mathcal{W}_9^T X = -\frac{1}{2} X_4 - \frac{1}{2} X_5 + \frac{1}{2} X_6 + \frac{1}{2} X_7 = \frac{1}{2} (X_7 + X_6 - X_5 - X_4) \\
W_{10} &= \mathcal{W}_{10}^T X = -\frac{1}{2} X_8 - \frac{1}{2} X_9 + \frac{1}{2} X_{10} + \frac{1}{2} X_{11} = \frac{1}{2} (X_{11} + X_{10} - X_9 - X_8) \\
W_{11} &= \mathcal{W}_{11}^T X = -\frac{1}{2} X_{12} - \frac{1}{2} X_{13} + \frac{1}{2} X_{14} + \frac{1}{2} X_{15} = \frac{1}{2} (X_{15} + X_{14} - X_{13} - X_{12}) \\
W_{12} &= \mathcal{W}_{12}^T X = -\frac{1}{\sqrt{8}} X_0 - \frac{1}{\sqrt{8}} X_1 - \frac{1}{\sqrt{8}} X_2 - \frac{1}{\sqrt{8}} X_3 + \frac{1}{\sqrt{8}} X_4 + \frac{1}{\sqrt{8}} X_5 + \frac{1}{\sqrt{8}} X_6 + \frac{1}{\sqrt{8}} X_7
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{8}} (X_7 + X_6 + X_5 + X_4 - X_3 - X_2 - X_1 - X_0) \\
W_{13} = W_{13}^T X &= -\frac{1}{\sqrt{8}} X_8 - \frac{1}{\sqrt{8}} X_9 - \frac{1}{\sqrt{8}} X_{10} - \frac{1}{\sqrt{8}} X_{11} + \frac{1}{\sqrt{8}} X_{12} + \frac{1}{\sqrt{8}} X_{13} + \frac{1}{\sqrt{8}} X_{14} + \frac{1}{\sqrt{8}} X_{15} \\
&= \frac{1}{\sqrt{8}} (X_{15} + X_{14} + X_{13} + X_{12} - X_{11} - X_{10} - X_9 - X_8) \\
W_{14} = W_{14}^T X &= -\frac{1}{4} X_0 - \frac{1}{4} X_1 - \frac{1}{4} X_2 - \frac{1}{4} X_3 - \dots - \frac{1}{4} X_7 + \frac{1}{4} X_8 + \frac{1}{4} X_9 + \dots + \frac{1}{4} X_{15} \\
&= \frac{1}{4} (X_{15} + X_{14} + \dots + X_8 + X_7 - X_6 - X_5 - \dots - X_1 - X_0) \\
W_{15} = W_{15}^T X &= \frac{1}{4} X_0 + \frac{1}{4} X_1 + \dots + \frac{1}{4} X_{14} + \frac{1}{4} X_{15} = \frac{1}{4} (X_{15} + X_{14} + \dots + X_1 + X_0),
\end{aligned}$$

es decir, se tiene lo siguiente:

$$\mathbf{W} = \begin{bmatrix} \frac{1}{\sqrt{2}} (X_1 - X_0) \\ \frac{1}{\sqrt{2}} (X_3 - X_2) \\ \frac{1}{\sqrt{2}} (X_5 - X_4) \\ \vdots \\ \frac{1}{\sqrt{2}} (X_{15} - X_{14}) \\ \frac{1}{2} (X_3 + X_2 - X_1 - X_0) \\ \frac{1}{2} (X_7 + X_6 - X_5 - X_4) \\ \frac{1}{2} (X_{11} + X_{10} - X_9 - X_8) \\ \frac{1}{2} (X_{15} + X_{14} - X_{13} - X_{12}) \\ \frac{1}{\sqrt{8}} (X_7 + X_6 + X_5 + X_4 - X_3 - X_2 - X_1 - X_0) \\ \frac{1}{\sqrt{8}} (X_{15} + X_{14} + X_{13} + X_{12} - X_{11} - X_{10} - X_9 - X_8) \\ \frac{1}{4} (X_{15} + X_{14} + \dots + X_8 + X_7 - X_6 - X_5 - \dots - X_1 - X_0) \\ \frac{1}{4} (X_{15} + X_{14} + \dots + X_1 + X_0) \end{bmatrix} \quad (3.19)$$

Lo anterior motiva la siguiente definición de *escala*, la cual nos proporciona una interpretación alternativa de lo que hasta el momento se ha desarrollado:

Dado  $\lambda$  entero positivo cualquiera, se define

$$\bar{X}_t(\lambda) \equiv \frac{1}{\lambda} \sum_{l=0}^{\lambda-1} X_{t-l}. \quad (3.20)$$

que es el promedio entre  $\lambda$  datos consecutivos con índices entre  $t - \lambda + 1$  y  $t$ ; observando con cuidado a la ecuación (3.19) se puede ver que

$$W_{15} = 4\overline{X}_{15}(16),$$

es decir,  $W_{15}$  es proporcional al promedio de todos los datos; veamos como podemos escribir a cualquier otro renglón de  $\mathbf{W}$  en términos de la ecuación  $\overline{X}_t(\lambda)$ .

Por ejemplo, tomemos el caso de

$$W_9 = \frac{1}{2}(X_7 + X_6 - X_5 - X_4) = \frac{1}{2}(X_7 + X_6) - \frac{1}{2}(X_5 + X_4).$$

Pero  $\frac{1}{2}(X_7 + X_6) = \frac{1}{2}\sum_{l=0}^1 X_{7-l} = \overline{X}_7(2)$  y  $\frac{1}{2}(X_5 + X_4) = \frac{1}{2}\sum_{l=0}^1 X_{5-l} = \overline{X}_5(2)$ ,

es decir, para  $\lambda = 2$  y  $t = 5, 7$  tenemos

$$W_9 = \overline{X}_7(2) - \overline{X}_5(2).$$

Aplicando este tipo de razonamiento para los demás renglones de  $W$  tenemos lo siguiente:

para

$$W_1 = \frac{1}{\sqrt{2}}(X_3 - X_2),$$

$X_3$  se puede reescribir con  $\lambda = 1$  y  $t = 3$  como  $\overline{X}_3(1)$ ; análogamente para  $X_2$ ,  $\lambda = 1$  y  $t = 2$ , tenemos  $\overline{X}_2(1)$ , así

$$\begin{aligned} W_1 &= \frac{1}{\sqrt{2}}(\overline{X}_3(1) - \overline{X}_2(1)) \\ W_2 &= \frac{1}{\sqrt{2}}(\overline{X}_5(1) - \overline{X}_4(1)) \\ &\vdots \\ W_7 &= \frac{1}{\sqrt{2}}(\overline{X}_{15}(1) - \overline{X}_{14}(1)) \end{aligned}$$

Como se podrá notar, estos 8 coeficientes son proporcionales a la diferencias de promedios adyacentes de los datos  $\{X_t\}$  correspondientes a escala  $\lambda = 1$ .

Por otra parte, considerando a  $\lambda = 2$  y  $t = 3$  tenemos que

$$\frac{1}{2}(X_3 + X_2) = \bar{X}_3(2)$$

y para  $t = 1$

$$\frac{1}{2}(X_1 + X_0) = \bar{X}_1(2),$$

por lo tanto

$$W_8 = \bar{X}_3(2) - \bar{X}_1(2).$$

De manera análoga

$$W_{10} = \bar{X}_{11}(2) - \bar{X}_9(2),$$

$$W_{11} = \bar{X}_{15}(2) - \bar{X}_{13}(2).$$

Al igual que para el caso anterior, los coeficientes de  $W_8, W_9, W_{10}$ , y  $W_{11}$  son proporcionales a la diferencia entre promedios adyacentes a escala  $\lambda = 2^1 = 2$ .

Consideremos  $\lambda = 4$  y  $t = 7$ , entonces tenemos

$$\frac{1}{\sqrt{8}} \sum_{l=0}^3 X_{7-l} = \sqrt{2} \left( \frac{1}{4}(X_7 + X_6 + X_5 + X_4) \right) = \sqrt{2} \bar{X}_7(4)$$

y además

$$\frac{1}{\sqrt{8}}(X_3 + X_2 + X_1 + X_0) = \sqrt{2} \bar{X}_3(4),$$

de donde

$$W_{12} = \sqrt{2}(\bar{X}_7(4) - \bar{X}_3(4)).$$

Además

$$W_{13} = \sqrt{2}(\bar{X}_{15}(4) - \bar{X}_{11}(4)).$$

Estos coeficientes corresponden a diferencias en promedios adyacentes a escala  $\lambda = 2^2 = 4$ .

Tomemos  $W_{14}$ , con  $\lambda = 8$  y  $t = 15$ , entonces

$$W_{14} = 2(\bar{X}_{15}(8) - \bar{X}_7(8)),$$

este coeficiente corresponde a diferencias en promedios adyacentes a escala  $\lambda = 2^3 = 8$ . Finalmente si  $\lambda = 16$  y  $t = 15$ , tenemos

$$W_{15} = 4\bar{X}_{15}(16),$$

el cual corresponde a diferencias en promedios adyacentes a escala  $\lambda = 2^4 = 16$ .

En resumen, los primeros  $N/2$  elementos de  $\mathbf{W}$  corresponden a cambios (de la serie original  $X$ ) en escala  $\lambda = 1$ , los siguientes  $N/4$  corresponden a cambios en promedios adyacentes (de la serie original  $\mathbf{X}$ ) en escala  $\lambda = 2$ , los siguientes  $N/8$  corresponden a cambios en escala  $\lambda = 4$ . El penúltimo coeficiente corresponde a cambios en promedios adyacentes a escala  $\lambda = 8$  y finalmente el último coeficiente corresponde a cambios a escala  $\lambda = 16$ , lo cual es equivalente a calcular el promedio de todas las observaciones.

Para  $J$  un entero positivo,  $N = 2^J$  y para otras DWT para las cuales los renglones de  $\mathcal{W}$  no se construyen a partir de los filtros de Haar, se puede realizar un análisis similar (véase Percival y Walden, (2000)). Es decir, para las transformadas DWT que procederemos a discutir en este capítulo, es posible relacionar la información de los primeros  $N_2 \equiv N/2$  coeficientes  $W_0, W_1, \dots, W_{N/2-1}$  con cambios a escala  $\lambda = \tau_1 \equiv 1$ , los siguientes  $N_2 \equiv N/4$  coeficientes  $W_{N/2}, W_{N/2+1}, \dots, W_{N/2+N/4-1}$  con cambios a escala  $\lambda = \tau_2 \equiv 2$ , y así sucesivamente. El último coeficiente  $W_{N-1}$  es proporcional al promedio de los datos. En general para  $\tau_j = 2^{j-1}$  con  $j = 1, 2, \dots, J$ , hay  $N_j = N/(2\tau_j)$  coeficientes de la DWT asociados a cambios en la escala  $\tau_j$ , para  $j = 1, 2, \dots, J$ . Además sólo hay un coeficiente  $W_{N-1}$  asociado con el promedio a escala  $N$ .

A los primeros  $N - 1$  coeficientes en  $W$  se les llaman *coeficientes de onduletas* y al último coeficiente  $W_{N-1}$  se le llama *coeficiente de escala*.

Los renglones de  $\mathcal{W}$  que producen los coeficientes de onduletas para alguna escala particular  $\tau_j$ , son versiones circularmente trasladadas de un mismo vector. El tamaño de la traslación o desplazamiento entre dos renglones adyacentes de  $\mathcal{W}$  a escala  $\tau_j$  es  $2\tau_j = 2^j$ .

Cada coeficiente de onduletas tiene información de los datos asociada con tiempos específicos en el conjunto de índices  $\{0, 1, \dots, N - 1\}$ . Esta *localización en el tiempo* puede ilustrarse en el ejemplo visto para la DWT de Haar. En este ejemplo  $\mathbf{W}_0$  involucra sólo a los tiempos 1 y 2, mientras que  $\mathbf{W}_8$  involucra a los tiempos 0, 1, 2 y 3. En contraste con esta propiedad, los coeficientes de la DFT se construyen con la información de la serie de tiempo  $\{X_t\}$  en todo el rango de índices  $0, 1, \dots, N - 1$ . Por tanto los coeficientes de la DFT no están asociados a información de los datos a tiempos específicos.

Dado un entero  $J_0$  tal que  $J_0 \leq J$  podemos particionar a los elementos del vector  $\mathbf{W}$  en  $J_0 + 1$  subvectores, los primeros  $J_0$  subvectores serán denotados por  $\mathbf{W}_j$ , para  $j = 1, 2, \dots, J_0$  donde el  $j$ -ésimo de estos vectores es de dimensión  $\frac{N}{2^j} \times 1$ . Así entonces  $\mathbf{W}_j$  contiene a los coeficientes de onduletas a escala  $\tau_j$ ,  $j = 1, 2, \dots, J_0$ . El último subvector se denotará por  $\mathbf{V}_{J_0}$ .

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_{J_0} \\ \mathbf{V}_{J_0} \end{bmatrix}. \quad (3.21)$$

Entonces con base en lo anterior y la ecuación (3.16) tenemos la siguiente expresión:

$$\|\mathbf{X}\|^2 = \|\mathbf{W}\|^2 = \sum_{j=1}^{J_0} \|\mathbf{W}_j\|^2 + \|\mathbf{V}_{J_0}\|^2. \quad (3.22)$$

Para el caso en el que  $J_0 = 4$  y  $N = 16$  tenemos

$$\begin{aligned} \mathbf{W}_1^T &= [W_0, W_1, W_2, W_3, W_4, W_5, W_6, W_7] & \mathbf{W}_2^T &= [W_8, W_9, W_{10}, W_{11}] \\ \mathbf{W}_3^T &= [W_{12}, W_{13}] & \mathbf{W}_4^T &= [W_{14}] & \mathbf{V}_4^T &= [W_{15}]. \end{aligned}$$

De acuerdo a como se hizo con el vector  $\mathbf{W}$ , se puede particionar a la matriz  $\mathcal{W}$  de manera análoga, es decir

$$\mathcal{W} = \begin{bmatrix} \mathcal{W}_1 \\ \mathcal{W}_2 \\ \vdots \\ \mathcal{W}_{J_0} \\ \mathcal{V}_{J_0} \end{bmatrix} \quad (3.23)$$

donde  $\mathcal{W}_1$  es una matriz de dimensiones  $N/2 \times N$  formada por los renglones desde  $n = 0$  hasta  $n = \frac{N}{2} - 1$  de  $\mathcal{W}$ ,  $\mathcal{W}_2$  es una matriz de dimensiones  $\frac{N}{4} \times N$  formada por los renglones desde  $n = \frac{N}{2}$  hasta  $n = \frac{3N}{4} - 1$  de  $\mathcal{W}$ , y así sucesivamente hasta  $\mathcal{W}_{J_0}$  y  $\mathcal{V}_{J_0}$  ambas matrices de



dimensión  $N/2^{J_0} \times N$ . Para  $N = 2^J$  y en el caso de que  $J = J_0$ ,  $\mathcal{W}_{J_0}$  y  $\mathcal{V}_{J_0}$  son matrices de dimensión  $1 \times N$  las cuales contienen los dos últimos renglones de  $\mathcal{W}$ .

Así, retomando el ejemplo de Haar, nuestra matriz  $\mathcal{W}$  quedaría particionada de la siguiente manera:

$$\mathcal{W}_1 = \begin{bmatrix} \mathcal{W}_{0\bullet} \\ \mathcal{W}_{1\bullet} \\ \vdots \\ \mathcal{W}_{7\bullet} \end{bmatrix} \quad \mathcal{W}_2 = \begin{bmatrix} \mathcal{W}_{8\bullet} \\ \mathcal{W}_{9\bullet} \\ \mathcal{W}_{10\bullet} \\ \mathcal{W}_{11\bullet} \end{bmatrix} \quad \mathcal{W}_3 = \begin{bmatrix} \mathcal{W}_{12\bullet} \\ \mathcal{W}_{13\bullet} \end{bmatrix} \quad \mathcal{W}_4 = [\mathcal{W}_{14\bullet}] \quad \mathcal{V}_4 = [\mathcal{W}_{15\bullet}].$$

De acuerdo con la ecuación (3.15) podemos escribir una ecuación de síntesis para  $\mathbf{X}$

$$\mathbf{X} = \mathcal{W}^t \mathbf{W} = [\mathcal{W}_1^T, \mathcal{W}_2^T, \dots, \mathcal{W}_{J_0}^T, \mathcal{V}_{J_0}] \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_{J_0} \\ \mathbf{V}_{J_0} \end{bmatrix} = \sum_{j=1}^{J_0} \mathcal{W}_j^T \mathbf{W}_j + \mathcal{V}_{J_0}^T \mathbf{V}_{J_0}, \quad (3.24)$$

entonces para cada  $j = 1, 2, \dots, J_0$  definimos al vector de dimensión  $N \times 1$

$$\mathcal{D}_j = \mathcal{W}_j^T \mathbf{W}. \quad (3.25)$$

Podemos pensar en  $\mathcal{D}_j$  como una *proyección* de los datos  $\mathbf{X}$  sobre el espacio vectorial que generan los renglones de  $\mathcal{W}_j$  o bien como la *componente* de  $\mathbf{X}$  en la ecuación de síntesis (3.24), que está asociada a variaciones ó cambios a escala  $\tau_j$ .

Además, para el caso  $J_0 = J$ , definimos

$$\mathcal{S}_J = \mathcal{V}_J^T \mathbf{V}_J, \quad (3.26)$$

probaremos más adelante que los elementos de este vector  $N$ -dimensional son constantes, de hecho son iguales al promedio de todos los elementos de  $\{X_t\}$ , es decir,

$$\mathcal{S}_J^T = (\bar{X}, \bar{X}, \dots, \bar{X}).$$

entonces si  $J_0 = J$  tenemos la ecuación

$$\mathbf{X} = \sum_{j=1}^J \mathcal{D}_j + \mathcal{S}_J, \quad (3.27)$$

a esta expresión se le denomina *análisis multiresolución* de  $\mathbf{X}$  con respecto a la DWT  $\mathcal{W}$ .

Analizando  $\mathcal{D}_j$  obtenemos la siguiente propiedad:

$$\|\mathcal{D}_j\|^2 = \mathcal{D}_j^T \mathcal{D}_j = (\mathcal{W}_j^T \mathbf{W}_j)^T (\mathcal{W}_j^T \mathbf{W}_j) = \mathbf{W}_j^T \mathcal{W}_j \mathcal{W}_j^T \mathbf{W}_j = \mathbf{W}_j^T \mathbf{W}_j = \|\mathcal{W}_j\|^2.$$

es decir

$$\|\mathcal{D}_j\|^2 = \|\mathcal{W}_j\|^2. \quad (3.28)$$

Para comenzar el estudio de la construcción de los renglones de  $\mathcal{W}$ , será necesario empezar a hablar de los filtros de onduletas y de escala. Estas son colecciones finitas de números (para nuestra descripción los consideraremos como números reales) que surgen como soluciones de un problema llamado *factorización espectral*. En Percival y Walden, (2000), se describe el problema y se estudian varias soluciones de la factorización espectral.

### 3.2.1. El filtro de Onduletas

Sea  $L$  un entero positivo, llamaremos a la sucesión de números reales  $\{h_l : l \in \mathbb{Z}\}$  un filtro de longitud  $L$  si  $h_0 \neq 0$ ,  $h_{L-1} \neq 0$ , si para  $l < 0$  tenemos  $h_l = 0$  y si para  $l \geq L$  tenemos  $h_l = 0$ . De esta forma tenemos la colección  $h_0, h_1, \dots, h_{L-1}$  de números reales. Decimos que  $\{h_l, l = 0, 1, 2, \dots, L-1\}$  es un filtro de onduletas si satisface las siguientes tres propiedades:

$$\sum_{l=0}^{L-1} h_l = 0, \quad (3.29)$$

$$\sum_{l=0}^{L-1} h_l^2 = 1, \quad (3.30)$$

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0, \quad (3.31)$$

para todo  $n$  entero distinto de cero. El valor de  $L$  debe de ser un entero par, esto se debe de cumplir necesariamente por la siguiente razón:

Considérese el filtro  $\{h_l : l = 0, 1, \dots, L-1\}$ , donde la longitud del filtro  $L$  es un entero impar tal que satisface que  $h_l = 0$  para  $l < 0$  y  $l \geq L$  entonces se afirma que este filtro no satisface la ecuación (3.31).

Sea  $n = (L-1)/2$ , éste número entero está bien definido puesto que  $L$  es un entero impar. Por ejemplo para  $L = 2m + 1$ , con  $m \in \{1, 2, \dots\}$  se tiene

$$\sum_{l=0}^{L-1} h_l h_{l+2[(L-1)/2]} = \sum_{l=0}^{L-1} h_l h_{l+L-1} = h_0 h_{L-1},$$

pero por hipótesis  $\{h_l\}$  es un filtro de longitud  $L$  y esto implica que  $h_0 \neq 0$  y  $h_{L-1} \neq 0$ . Por lo tanto  $\{h_l\}$  no satisface la ecuación (3.31).

Notemos que si  $L = 1$ , entonces  $\{h_l : l = 0, 1, \dots, L-1\}$  no satisface (3.29) y (3.30) al mismo tiempo.

Por lo tanto la longitud del filtro debe ser un entero par para que  $\{h_l : l = 0, 1, \dots, L-1\}$  sea un filtro de onduletas.

De hecho hasta el momento el único filtro de onduletas que conocemos es el de *Haar*, donde  $\{h_0 = 1/\sqrt{2}, h_1 = -1/\sqrt{2}\}$  y  $L = 2$ , más adelante veremos más ejemplos.

Ahora definamos la función de transferencia  $H(\cdot)$  asociada al filtro  $\{h_l : l = 0, 1, \dots, L-1\}$ :

$$H(f) = \sum_{l=-\infty}^{\infty} h_l e^{-i2\pi fl} = \sum_{l=0}^{L-1} h_l e^{-i2\pi fl} \quad (3.32)$$

y su correspondiente *función de ganancia* (módulo al cuadrado)

$$\mathcal{H}(f) = |H(f)|^2.$$

La siguiente condición es equivalente a las condiciones de las ecuaciones (3.30) y (3.31):

$$\mathcal{H}(f) + \mathcal{H}\left(f + \frac{1}{2}\right) = 2, \quad \text{para toda } f. \quad (3.33)$$

Justifiquemos la aseveración anterior, primero veamos que si se supone la condición (3.33) entonces tendremos las ecuaciones (3.30) y (3.31).

Sea  $\{h_l\}$  filtro con valores en los reales, tales que su función de ganancia cuadrada satisface la condición (3.33). La correlación cruzada compleja de  $\{h_l\}$  está dada por:

$$h \star h_j \equiv \sum_{l=-\infty}^{\infty} \bar{h}_l h_{l+j}$$

para  $j = \dots, -2, -1, 0, \dots, 1, 2, \dots$ , entonces dado que es un filtro con valores reales lo anterior es equivalente a

$$h \star h_j \equiv \sum_{l=-\infty}^{\infty} h_l h_{l+j}.$$

Por la ecuación (2.26) tenemos

$$\{h \star h_j\} \leftrightarrow \mathcal{H}(\cdot).$$

Una propiedad importante es la siguiente, la cual se obtiene de aplicar la expresión dada en la ecuación (2.6), es decir,

$$\{h \star h_{2n}\} \leftrightarrow \frac{1}{2} \left[ \mathcal{H}\left(\frac{f}{2}\right) + \mathcal{H}\left(\frac{f}{2} + \frac{1}{2}\right) \right],$$

pero al calcular la Transformada Inversa de Fourier tenemos

$$h \star h_{2n} = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[ \mathcal{H}\left(\frac{f}{2}\right) + \mathcal{H}\left(\frac{f}{2} + \frac{1}{2}\right) \right] e^{i2\pi f n} df. \quad (3.34)$$

Entonces como se satisface la ecuación (3.33) para toda  $f$ , en particular se cumple para  $\frac{f}{2}$ , es decir,

$$\mathcal{H}\left(\frac{f}{2}\right) + \mathcal{H}\left(\frac{f}{2} + \frac{1}{2}\right) = 2,$$

lo cual implica que

$$\frac{1}{2} \left[ \mathcal{H}\left(\frac{f}{2}\right) + \mathcal{H}\left(\frac{f}{2} + \frac{1}{2}\right) \right] = 1,$$

por lo tanto retomando la expresión (3.34):

$$h \star h_{2n} = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi f n} df = \begin{cases} 0 & n \neq 0 \\ 1 & n = 0; \end{cases}$$

lo cuál es cierto por lo desarrollado en la ecuación (1.14), es decir, se tienen las expresiones (3.30) y (3.31).

Conversamente, ahora supongamos que se satisfacen las ecuaciones (3.30) y (3.31) y veamos que efectivamente se tiene la expresión (3.33).

Recordemos que  $H(f) = \sum_{l=0}^{L-1} h_l e^{-i2\pi fl}$ , definamos a la sucesión de números complejos  $\{a_l : l = 0, 1, 2, \dots, L-1\}$  de la siguiente manera

$$a_l = h_l e^{-i2\pi fl}, \quad l = 0, 1, 2, \dots, L-1.$$

La sucesión  $\{a_l\}$  satisface la siguiente condición

$$a_l \bar{a}_l = h_l e^{-i2\pi fl} \overline{(h_l e^{-i2\pi fl})} = h_l e^{-i2\pi fl} h_l e^{i2\pi fl} = h_l^2, \quad l = 0, 1, 2, \dots, L-1. \quad (3.35)$$

Calculemos la siguiente expresión

$$\mathcal{H}(f) = H(f) \overline{H(f)} = \left( \sum_{l=0}^{L-1} a_l \right) \left( \sum_{l=0}^{L-1} \bar{a}_l \right) \quad (3.36)$$

$$\begin{aligned} \left( \sum_{l=0}^{L-1} a_l \right) \left( \sum_{l=0}^{L-1} \bar{a}_l \right) &= (a_0 + a_1 + a_2 + \dots + a_{L-1}) (\bar{a}_0 + \bar{a}_1 + \bar{a}_2 + \dots + \bar{a}_{L-1}) \\ &= a_0 \bar{a}_0 + a_0 \bar{a}_1 + a_0 \bar{a}_2 + \dots + a_0 \bar{a}_{L-4} + a_0 \bar{a}_{L-3} + a_0 \bar{a}_{L-2} + a_0 \bar{a}_{L-1} \\ &+ a_1 \bar{a}_0 + a_1 \bar{a}_1 + a_1 \bar{a}_2 + \dots + a_1 \bar{a}_{L-4} + a_1 \bar{a}_{L-3} + a_1 \bar{a}_{L-2} + a_1 \bar{a}_{L-1} \\ &+ a_2 \bar{a}_0 + a_2 \bar{a}_1 + a_2 \bar{a}_2 + \dots + a_2 \bar{a}_{L-4} + a_2 \bar{a}_{L-3} + a_2 \bar{a}_{L-2} + a_2 \bar{a}_{L-1} \\ &+ a_3 \bar{a}_0 + a_3 \bar{a}_1 + a_3 \bar{a}_2 + \dots + a_3 \bar{a}_{L-4} + a_3 \bar{a}_{L-3} + a_3 \bar{a}_{L-2} + a_3 \bar{a}_{L-1} \\ &\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ &+ a_{L-2} \bar{a}_0 + a_{L-2} \bar{a}_1 + a_{L-2} \bar{a}_2 + \dots + a_{L-2} \bar{a}_{L-3} + a_{L-2} \bar{a}_{L-2} + a_{L-2} \bar{a}_{L-1} \\ &+ a_{L-1} \bar{a}_0 + a_{L-1} \bar{a}_1 + a_{L-1} \bar{a}_2 + \dots + a_{L-1} \bar{a}_{L-3} + a_{L-1} \bar{a}_{L-2} + a_{L-1} \bar{a}_{L-1} \end{aligned}$$

sumando sobre las diagonales obtenemos

$$= \sum_{l=0}^{L-1} a_l \bar{a}_l + \sum_{l=0}^{L-2} a_l \bar{a}_{l+1} + \sum_{l=0}^{L-3} a_l \bar{a}_{l+2} + \dots + \sum_{l=0}^2 a_l \bar{a}_{l+L-3} + \sum_{l=0}^1 a_l \bar{a}_{l+L-2} + \sum_{l=0}^0 a_l \bar{a}_{l+L-1}$$

$$+ \sum_{l=0}^{L-2} a_{l+1} \bar{a}_l + \sum_{l=0}^{L-3} a_{l+2} \bar{a}_l + \sum_{l=0}^{L-4} a_{l+3} \bar{a}_l + \dots + \sum_{l=0}^2 a_{l+L-3} \bar{a}_l + \sum_{l=0}^1 a_{l+L-2} \bar{a}_l + \sum_{l=0}^0 a_{l+L-1} \bar{a}_l$$

donde se tiene que

$$\begin{aligned} \sum_{l=0}^{L-2} a_l \bar{a}_{l+1} &= \sum_{l=0}^{L-2} h_l e^{-i2\pi f l} h_{l+1} e^{i2\pi f (l+1)} = \sum_{l=0}^{L-2} h_l h_{l+1} e^{i2\pi f} \quad y \\ \sum_{l=0}^{L-2} a_{l+1} \bar{a}_l &= \sum_{l=0}^{L-2} h_{l+1} e^{-i2\pi f (l+1)} h_l e^{i2\pi f l} = \sum_{l=0}^{L-2} h_l h_{l+1} e^{-i2\pi f} \quad , \end{aligned}$$

sumando las dos expresiones anteriores obtenemos que:

$$\sum_{l=0}^{L-2} a_l \bar{a}_{l+1} + \sum_{l=0}^{L-2} a_{l+1} \bar{a}_l = \sum_{l=0}^{L-2} h_l h_{l+1} (e^{i2\pi f} + e^{-i2\pi f}) = 2 \sum_{l=0}^{L-2} h_l h_{l+1} \cos(2\pi f) .$$

Procediendo de manera análoga para la siguiente pareja de sumas

$$\begin{aligned} \sum_{l=0}^{L-3} a_l \bar{a}_{l+2} &= \sum_{l=0}^{L-3} h_l e^{-i2\pi f l} h_{l+2} e^{i2\pi f (l+2)} = \sum_{l=0}^{L-3} h_l h_{l+2} e^{i4\pi f} \quad y \\ \sum_{l=0}^{L-3} a_{l+2} \bar{a}_l &= \sum_{l=0}^{L-3} h_{l+2} e^{-i2\pi f (l+2)} h_l e^{i2\pi f l} = \sum_{l=0}^{L-3} h_l h_{l+2} e^{-i4\pi f} \quad , \end{aligned}$$

nuevamente sumamos las dos expresiones que acabamos de obtener

$$\sum_{l=0}^{L-3} a_l \bar{a}_{l+2} + \sum_{l=0}^{L-3} a_{l+2} \bar{a}_l = \sum_{l=0}^{L-3} h_l h_{l+2} (e^{i4\pi f} + e^{-i4\pi f}) = 2 \sum_{l=0}^{L-3} h_l h_{l+2} \cos(4\pi f)$$

y así sucesivamente para el resto de los términos en la expresión resultante de realizar el producto indicado en la ecuación (3.36). Dado que por hipótesis se satisface la ecuación (3.31) entonces tenemos

$$\begin{aligned} \mathcal{H}(f) &= \sum_{l=0}^{L-1} h_l^2 + 2 \sum_{l=0}^{L-2} h_l h_{l+1} \cos(2\pi f) \\ &+ 2 \sum_{l=0}^{L-4} h_l h_{l+3} \cos(6\pi f) + \dots + 2 \sum_{l=0}^0 h_l h_{l+L-1} \cos(2\pi f (L-1)) . \end{aligned}$$

Utilizando los mismos argumentos encontremos

$$\mathcal{H}\left(f + \frac{1}{2}\right) = H\left(f + \frac{1}{2}\right) \overline{H\left(f + \frac{1}{2}\right)}, \quad (3.37)$$

donde

$$H\left(f + \frac{1}{2}\right) = \sum_{l=0}^{L-1} h_l e^{-i2\pi(f+\frac{1}{2})l} = \sum_{l=0}^{L-1} h_l e^{-i2\pi fl} e^{-i\pi l} = \sum_{l=0}^{L-1} (-1)^l h_l e^{-i2\pi fl} = \sum_{l=0}^{L-1} h'_l e^{-i2\pi fl} \quad (3.38)$$

haciendo

$$h'_l = (-1)^l h_l \quad l = 0, 1, 2, \dots, L-1.$$

La sucesión  $\{h'_l\}$  satisface las siguientes condiciones para cualquier  $l = 0, 1, 2, \dots, L-1$

$$(h'_l)^2 = ((-1)^l h_l)^2 = h_l^2, \quad (3.39)$$

$$h'_l h'_{l+2k} = (-1)^l h_l (-1)^{l+2k} h_{l+2k} = (-1)^{2l+2k} h_l h_{l+2k} = h_l h_{l+2k}. \quad (3.40)$$

Entonces la expresión (3.37) queda en la forma

$$\begin{aligned} \mathcal{H}\left(f + \frac{1}{2}\right) &= \sum_{l=0}^{L-1} (h'_l)^2 + \sum_{l=0}^{L-2} h'_l h'_{l+1} e^{i2\pi f} + \sum_{l=0}^{L-3} h'_l h'_{l+2} e^{i2\pi f(2)} + \dots + \sum_{l=0}^1 h'_l h'_{l+L-2} e^{i2\pi f(L-2)} \\ &+ \sum_{l=0}^0 h'_l h'_{l+L-1} e^{i2\pi f(L-1)} + \sum_{l=0}^{L-2} h'_l h'_{l+1} e^{-i2\pi f} + \sum_{l=0}^{L-3} h'_l h'_{l+2} e^{-i2\pi f(2)} + \dots \\ &+ \sum_{l=0}^1 h'_l h'_{l+L-2} e^{-i2\pi f(L-2)} + \sum_{l=0}^0 h'_l h'_{l+L-1} e^{-i2\pi f(L-1)} = \sum_{l=0}^{L-1} h_l^2 + \sum_{l=0}^{L-2} h'_l h'_{l+1} e^{i2\pi f} \\ &+ \sum_{l=0}^{L-3} h'_l h'_{l+2} e^{i2\pi f(2)} + \dots + \sum_{l=0}^1 h'_l h'_{l+L-2} e^{i2\pi f(L-2)} + \sum_{l=0}^0 h'_l h'_{l+L-1} e^{i2\pi f(L-1)} \\ &= \sum_{l=0}^{L-1} h_l^2 + 2 \sum_{l=0}^{L-2} h'_l h'_{l+1} \cos(2\pi f) + 2 \sum_{l=0}^{L-4} h'_l h'_{l+3} \cos(2\pi f(3)) + \dots \\ &+ \sum_{l=0}^2 h'_l h'_{l+L-3} \cos(2\pi f(L-3)) + \sum_{l=0}^0 h'_l h'_{l+L-1} \cos(2\pi f(L-1)), \end{aligned}$$

donde hemos usado la ecuación (3.31). Por lo tanto se tiene que

$$\begin{aligned}
\mathcal{H}(f) + \mathcal{H}\left(f + \frac{1}{2}\right) &= \sum_{l=0}^{L-1} h_l^2 + 2 \sum_{l=0}^{L-2} h_l h_{l+1} \cos(2\pi f) + 2 \sum_{l=0}^{L-4} h_l h_{l+3} \cos(2\pi f(3)) + \dots \\
&+ 2 \sum_{l=0}^0 h_l h_{l+L-1} \cos(2\pi f(L-1)) + \sum_{l=0}^{L-1} h_l^2 + 2 \sum_{l=0}^{L-2} (-1)^{2l+1} h_l h_{l+1} \cos(2\pi f) \\
&+ 2 \sum_{l=0}^{L-4} (-1)^{2l+3} h_l h_{l+3} \cos(2\pi f(3)) + \dots + 2 \sum_{l=0}^0 (-1)^{2l+L-1} h_l h_{l+L-1} \cos(2\pi f(L-1)) \\
&= 2.
\end{aligned}$$

Con lo cual queda demostrada la ecuación (3.33).

Los componentes del sub-vector  $\mathbf{W}_j$  resultante de multiplicar  $\mathcal{W}_j$  por  $\mathbf{X}$ , son los coeficientes de la DWT correspondientes a la escala  $\tau_j = 2^{j-1}$ . Usaremos la notación  $W_{j,t}$  para los coeficientes de la DWT donde el subíndice  $j$  en  $W_{j,t}$  indica la escala a la que pertenece el coeficiente  $W_{j,t}$  para  $t = 0, 1, 2, \dots, N/2^{j-1}$  con  $j = 1, 2, \dots, J$ .

$$\mathbf{W}_1 = \begin{bmatrix} W_{1,0} \\ W_{1,1} \\ \vdots \\ W_{1,N/2-1} \end{bmatrix}, \quad \mathbf{W}_2 = \begin{bmatrix} W_{2,0} \\ W_{2,1} \\ \vdots \\ W_{2,N/4-1} \end{bmatrix}.$$

En Percival y Walden (2000), establecen que se puede relacionar el subíndice  $t$  con tiempos en los que se observa a la serie de tiempo  $\mathbf{X}$  bajo estudio. Por ejemplo para el caso de filtros de Haar, para el coeficiente  $W_{1,0}$ , el índice  $t = 0$  guarda relación con los tiempos 1 y 0 en la serie original, (véase la ecuación (3.19)).

Sea  $N = 2^J$ , con  $J$  entero positivo, para obtener los coeficientes de onduletas asociados a escala  $\tau = 1$ , se filtra de manera circular la serie de tiempo  $\{X_t : t = 0, 1, 2, \dots, N-1\}$  con el filtro  $\{h_l\}$ , al resultado de esta operación de filtrado se le denotará por

$$2^{1/2} \widetilde{W}_{1,t} \equiv \sum_{l=0}^{L-1} h_l X_{t-l \bmod N}, \quad t = 0, 1, 2, \dots, N-1. \quad (3.41)$$

Los coeficientes de onduletas  $\mathbf{W}_1$  para escala  $\tau = 1$ , se definen como



$$W_{1,t} \equiv 2^{1/2} \widetilde{W}_{1,2t+1} = \sum_{l=0}^{L-1} h_l X_{2t+1-l \bmod N}, \quad t = 0, 1, 2, \dots, N/2 - 1, \quad (3.42)$$

el primer subíndice en  $W_{1,t}$  indica la escala (en este caso  $\tau_1 = 2^0 = 1$ ) asociada a los  $N/2$  coeficientes de onduletas. Al procedimiento de tomar solo los coeficientes con índice impar  $2^{1/2} \widetilde{W}_{1,2t+1}$  para definir a  $W_{1,t}$  en la ecuación (3.42) se le llama *subsampling* por dos ó *down-sampling* por dos. Usemos la ecuación (3.42) con  $\mathbf{X}^T = (X_0, X_1, \dots, X_{15})$  y el filtro de escala de Haar  $\{h_0 = 1/\sqrt{2}, h_1 = -1/\sqrt{2}\}$  para calcular  $W_{1,t}$  para  $t = 0, 1, \dots, 7$ .

$$W_{1,0} = \sum_{l=0}^1 h_l X_{2(0)+1-l \bmod N} = h_0 X_1 + h_1 X_0 = 1/\sqrt{2} X_1 - 1/\sqrt{2} X_0 = 1/\sqrt{2} (X_1 - X_0)$$

$$W_{1,1} = \sum_{l=0}^1 h_l X_{2(1)+1-l \bmod N} = h_0 X_3 + h_1 X_2 = 1/\sqrt{2} X_3 - 1/\sqrt{2} X_2 = 1/\sqrt{2} (X_3 - X_2)$$

$$W_{1,2} = \sum_{l=0}^1 h_l X_{2(2)+1-l \bmod N} = h_0 X_5 + h_1 X_4 = 1/\sqrt{2} X_5 - 1/\sqrt{2} X_4 = 1/\sqrt{2} (X_5 - X_4)$$

$$W_{1,3} = \sum_{l=0}^1 h_l X_{2(3)+1-l \bmod N} = h_0 X_7 + h_1 X_6 = 1/\sqrt{2} X_7 - 1/\sqrt{2} X_6 = 1/\sqrt{2} (X_7 - X_6)$$

$$W_{1,4} = \sum_{l=0}^1 h_l X_{2(4)+1-l \bmod N} = h_0 X_9 + h_1 X_8 = 1/\sqrt{2} X_9 - 1/\sqrt{2} X_8 = 1/\sqrt{2} (X_9 - X_8)$$

$$W_{1,5} = \sum_{l=0}^1 h_l X_{2(5)+1-l \bmod N} = h_0 X_{11} + h_1 X_{10} = 1/\sqrt{2} X_{11} - 1/\sqrt{2} X_{10} = 1/\sqrt{2} (X_{11} - X_{10})$$

$$W_{1,6} = \sum_{l=0}^1 h_l X_{2(6)+1-l \bmod N} = h_0 X_{13} + h_1 X_{12} = 1/\sqrt{2} X_{13} - 1/\sqrt{2} X_{12} = 1/\sqrt{2} (X_{13} - X_{12})$$

$$W_{1,7} = \sum_{l=0}^1 h_l X_{2(7)+1-l \bmod N} = h_0 X_{15} + h_1 X_{14} = 1/\sqrt{2} X_{15} - 1/\sqrt{2} X_{14} = 1/\sqrt{2} (X_{15} - X_{14})$$

Si pensamos a  $\mathbf{X}$  como un vector columna de dimensión  $N = 2^J$ , de las ecuaciones (3.14) y (3.23) los coeficientes de onduletas asociados a escala  $\tau = 1$ , serán

$$\mathbf{W}_1 = \mathcal{W}_1 \mathbf{X}$$

donde  $\mathcal{W}_1$  es la matriz de dimensión  $\frac{N}{2} \times N$  que contiene los primeros  $\frac{N}{2}$  renglones de  $\mathcal{W}$  y los elementos de  $W_1$  son los coeficientes de onduletas  $\{W_{1,t} : t = 0, 1, \dots, N/2 - 1\}$  definidos por las expresiones en (3.41) y (3.42).

Sea  $\{h_l^\circ, l = 0, 1, 2, \dots, N-1\}$  el filtro periodizado de longitud  $N$  asociado al filtro  $\{h_l : l = 0, 1, \dots, L-1\}$  (véase la definición (2.27)), entonces podemos re-escribir la ecuación (3.42) como

$$W_{1,t} = \sum_{l=0}^{N-1} h_l^\circ X_{2t+1-l \bmod N}, \quad t = 0, 1, 2, \dots, \frac{N}{2} - 1. \quad (3.43)$$

Pero para cada  $t = 0, 1, 2, \dots, N/2 - 1$ , podemos obtener  $W_{1,t}$  como el producto del  $t$ -ésimo renglón de  $\mathcal{W}$  por  $\mathbf{X}$

$$W_{1,t} = \mathcal{W}_{t,\bullet}^T \mathbf{X} = \sum_{l=0}^{N-1} h_l^\circ X_{2t+1-l \bmod N} = \sum_{l=0}^{N-1} h_{2t+1-l \bmod N}^\circ X_l. \quad (3.44)$$

Por ejemplo, para  $t = 0$ ,

$$\begin{aligned} W_{1,0} &= \mathcal{W}_{0,\bullet} \mathbf{X} = \sum_{l=0}^{N-1} h_{1-l \bmod N}^\circ X_l = h_1^\circ X_0 + h_0^\circ X_1 + h_{-1 \bmod N}^\circ X_2 + \dots + h_{3-N \bmod N}^\circ X_{N-2} \\ &+ h_{2-N \bmod N}^\circ X_{N-1} = h_1^\circ X_0 + h_0^\circ X_1 + h_{N-1}^\circ X_2 + h_{N-2}^\circ X_3 + \dots + h_3^\circ X_{N-2} + h_2^\circ X_{N-1}, \end{aligned}$$

por lo tanto

$$\mathcal{W}_{0,\bullet}^T = [h_1^\circ, h_0^\circ, h_{N-1}^\circ, h_{N-2}^\circ, \dots, h_3^\circ, h_2^\circ].$$

Notemos que los elementos de  $\{h_l^\circ\}$  aparecen invertidos en el tiempo y trasladados de manera periódica.

Entonces como ya se vió anteriormente y de la ecuación (3.44), los restantes  $N/2 - 1$  renglones de la matriz  $\mathcal{W}_1$  son copias de  $\mathcal{W}_{0,\bullet}^T$  desplazadas circularmente, es decir,

$$\mathcal{W}_{t,\bullet}^T = [T^{2t} \mathcal{W}_{0,\bullet}]^T, \quad t = 1, 2, \dots, \frac{N}{2} - 1.$$

Por ejemplo, para  $t = 1$  se tiene:

$$\begin{aligned} W_{1,1} &= \mathcal{W}_{1,\bullet}^T \mathbf{X} = \sum_{l=0}^{N-1} h_{2+1-l \bmod N}^\circ X_l = \sum_{l=0}^{N-1} h_{3-l \bmod N}^\circ X_l \\ &= h_3^\circ X_0 + h_2^\circ X_1 + h_1^\circ X_2 + h_0^\circ X_3 + h_{-1 \bmod N}^\circ X_4 + \dots + h_{5-N \bmod N}^\circ X_{N-2} + h_{4-N \bmod N}^\circ X_{N-1}, \end{aligned}$$

de donde

$$\mathcal{W}_{1\bullet} = [h_3^\circ, h_2^\circ, h_1^\circ, h_0^\circ, h_{N-1}^\circ, h_{N-2}^\circ, \dots, h_5^\circ, h_4^\circ]. \quad (3.45)$$

A continuación, probaremos para el caso  $L \leq N$  que los renglones de  $\mathcal{W}_1$  forman un conjunto de  $N/2$  vectores ortonormales. En este caso el filtro periodizado toma la forma

$$h_l^\circ = \begin{cases} h_l & 0 \leq l \leq L-1; \\ 0 & L \leq l \leq N-1, \end{cases}$$

por lo tanto el primer renglón de  $\mathcal{W}_1$  está dado por

$$\mathcal{W}_{0\bullet}^t = [h_1^\circ, h_0^\circ, h_{N-1}^\circ, h_{N-2}^\circ, \dots, h_L^\circ, h_{L-1}^\circ, h_2^\circ] = \left[ h_1, h_0, \underbrace{0, \dots, 0}_{N-L \text{ ceros}}, h_{L-1}, h_2 \right],$$

utilizando el hecho de que  $\mathcal{T}^{2T}$  es una transformación ortonormal, tenemos

$$\|\mathcal{W}_{t\bullet}\|^2 = \|\mathcal{T}^{2t}\mathcal{W}_{0\bullet}\|^2 = \sum_{l=0}^{L-1} h_l^2 = 1$$

dado que  $\{h_l\}$  es un filtro de onduletas, (véase la ecuación (3.30)). Por otra parte para  $j' \neq j$  se tiene

$$\langle \mathcal{W}_{j'\bullet}, \mathcal{W}_{j\bullet} \rangle = \mathcal{W}_{j'\bullet}^T \mathcal{W}_{j\bullet} = \mathcal{W}_{0\bullet}^T \mathcal{T}^{-2j'} \mathcal{T}^{2j} \mathcal{W}_{0\bullet} = \sum_{l=0}^{L-1} h_l h_{l+2(j-j')} = 0$$

lo anterior es cierto utilizando la ecuación (3.31) referente a filtros de onduletas, por lo tanto tenemos que los renglones de  $\mathcal{W}_1$  son ortonormales. Ahora daremos una interpretación alternativa a los coeficientes de onduletas de escala  $\tau = 1$  en términos de teoría de filtrado.

Primero defínase la transformada discreta de Fourier de  $\{X_t\}$

$$\mathcal{X}_k \equiv \sum_{t=0}^{N-1} X_t e^{-i2\pi tk/N}, \quad t = 0, 1, 2, \dots, N-1. \quad (3.46)$$

Entonces aplicando el teorema de Parseval (dado en el capítulo 2, ecuación (2.23)) y la ecuación (3.9) tenemos la siguiente expresión:

$$\sum_{t=0}^{N-1} X_t^2 = \frac{1}{N} \sum_{k=0}^{N-1} |\mathcal{X}_k|^2. \quad (3.47)$$

Si ahora filtramos a  $\{X_t\}$  con  $\{h_l\}$  obtenemos

$$2^{1/2} \widetilde{\mathcal{W}}_{1,t} = \sum_{l=0}^{L-1} h_l X_{t-l \bmod N} = \sum_{l=0}^{N-1} h_l^\circ X_{t-l \bmod N}$$

donde la primera igualdad se obtiene aplicando la ecuación (3.41) y la segunda igualdad es cierta puesto que  $L \leq N$ . La sucesión  $\{2^{1/2} \widetilde{\mathcal{W}}_{1,t}\}$  es el resultado de hacer la convolución circular entre  $\{X_t\}$  y el filtro  $\{h_l^\circ\}$  de longitud  $N$ ; de acuerdo con la ecuación (2.25), el par de Fourier correspondiente es

$$\{2^{1/2} \widetilde{\mathcal{W}}_{1,t}\} \leftrightarrow \left\{ H \left( \frac{k}{N} \right) \mathcal{X}_k \right\}.$$

Luego entonces el teorema de Parseval nos permite escribir:

$$2 \sum_{t=0}^{N-1} \widetilde{\mathcal{W}}_{1,t}^2 = \frac{1}{N} \sum_{k=0}^{N-1} \left| H \left( \frac{k}{N} \right) \mathcal{X}_k \right|^2 = \frac{1}{N} \sum_{k=0}^{N-1} \left| H \left( \frac{k}{N} \right) \right|^2 |\mathcal{X}_k|^2 = \frac{1}{N} \sum_{k=0}^{N-1} \mathcal{H} \left( \frac{k}{N} \right) |\mathcal{X}_k|^2 \quad (3.48)$$

Así, el efecto de filtrar  $\{X_t\}$  para obtener  $\{2^{1/2} \widetilde{\mathcal{W}}_{1,t}\}$  puede entonces evaluarse estudiando  $\mathcal{H} \left( \frac{k}{N} \right)$  en las frecuencias de la forma  $k/N$ .

### 3.2.2. El filtro de Escala

Ahora definamos un segundo filtro,  $\{g_l\}$ , este será necesario para construir los renglones de las matrices  $\mathcal{W}_2, \mathcal{W}_3, \dots, \mathcal{W}_J$ , así como para  $\mathcal{V}_J$  en la ecuación (3.23). Recordemos que si  $N = 2^J$  y  $J = J_0$  en la ecuación (3.23) entonces  $\mathcal{V}_J$  es un vector  $N$ -dimensional cuyo producto interno con  $\mathbf{X}$  nos da el coeficiente de escala  $\mathcal{V}_J$ . Por esta razón, al filtro  $\{g_l\}$  dado por la expresión

$$g_l \equiv (-1)^{l+1} h_{L-1-l} \quad (3.49)$$

se le llamará *filtro de escala*. Se puede recuperar  $\{h_l\}$  a partir de los coeficientes de  $\{g_l\}$  por medio de la relación

$$h_l = (-1)^l g_{L-1-l}. \quad (3.50)$$

Claramente (3.49) y (3.50) son equivalentes. Por ejemplo, si aceptamos la ecuación (3.50) tenemos

$$h_{L-1-l} = (-1)^{L-1-l} g_l = (-1)^L (-1)^{-(l+1)} g_l$$

de donde

$$g_l = (-1)^{l+1} h_{L-1-l}.$$

Como ya se mencionó, los coeficientes  $\{h_l : l = 0, 1, \dots, L-1\}$  surgen como solución a un problema de factorización espectral, (para mayores referencias consulte Percival y Walden, (2000)).

Retomando nuestro ejemplo de Haar, dado por  $h_0 = 1/\sqrt{2}$  y  $h_1 = -1/\sqrt{2}$ , los términos distintos de cero del filtro de escala son

$$g_0 = -h_1 = 1/\sqrt{2} \quad y \quad g_1 = h_0 = 1/\sqrt{2}.$$

Otro filtro conocido es el de Asimetría Mínima de Daubechies,  $LA(8)$ , con  $L = 8$  coeficientes:

$$\begin{aligned} g_0 &= -0.0757657147893407 & g_1 &= -0.02963552764595541 \\ g_2 &= 0.4976186676324578 & g_3 &= 0.8037387518052163 \\ g_4 &= 0.2978577956055422 & g_5 &= -0.0992195435769354 \\ g_6 &= -0.0126039672622612 & g_7 &= 0.0322231006040713 \end{aligned}$$

y así de acuerdo con la ecuación (3.50)

$$\begin{aligned} h_0 &= g_7 = 0.03222310060407 & h_1 &= -g_6 = -0.01260396726226 \\ h_2 &= g_5 = -0.09921954357694 & h_3 &= -g_4 = -0.29785779560554 \\ h_4 &= g_3 = 0.80373875180522 & h_5 &= -g_2 = -0.49761866763246 \\ h_6 &= g_1 = -0.02963552764595 & h_7 &= -g_0 = -0.07576571478934 \end{aligned}$$

Sean el filtro de onduletas  $\{h_l\}$  con función de transferencia  $H(\cdot)$ ,  $\{g_l\}$  el filtro de escala con su correspondiente función de transferencia  $G(\cdot)$ . Entonces la función de transferencia  $G(\cdot)$  está dada por

$$G(f) \equiv \sum_{l=-\infty}^{\infty} g_l e^{-i2\pi fl} = \sum_{l=0}^{L-1} g_l e^{-i2\pi fl} = e^{-i2\pi f(L-1)} H\left(\frac{1}{2} - f\right) \quad (3.51)$$

y por lo tanto

$$\mathcal{G}(f) = \mathcal{H}\left(\frac{1}{2} - f\right), \quad (3.52)$$

donde  $\mathcal{G}(f) = |G(f)|^2$  es la función de ganancia cuadrada del filtro  $\{g_l\}$ .

Para probar (3.51) notemos que

$$H\left(\frac{1}{2} - f\right) = \sum_{l=0}^{L-1} h_l e^{-i2\pi(\frac{1}{2}-f)l} = \sum_{l=0}^{L-1} h_l e^{i2\pi fl} e^{-i\pi l} = \sum_{l=0}^{L-1} (-1)^l h_l e^{i2\pi fl},$$

por lo tanto

$$\begin{aligned} e^{-i2\pi f(L-1)} H\left(\frac{1}{2} - f\right) &= \sum_{l=0}^{L-1} (-1)^l h_l e^{i2\pi fl} e^{-i2\pi f(L-1)} = \sum_{l=0}^{L-1} (-1)^l h_l e^{-i2\pi f(L-1-l)} \\ &= \sum_{l=0}^{L-1} (-1)^l (-1)^l g_{L-1-l} e^{-i2\pi f(L-1-l)} = \sum_{l=0}^{L-1} g_{L-1-l} e^{-i2\pi f(L-1-l)} \\ &= \sum_{l=0}^{L-1} g_l e^{-i2\pi fl} = G(f). \end{aligned}$$

Por lo tanto se tiene que

$$G(f) = e^{-i2\pi f(L-1)} H\left(\frac{1}{2} - f\right)$$

y como además

$$|e^{-i2\pi f(L-1)}| = 1$$

entonces

$$|G(f)|^2 = \left| H\left(\frac{1}{2} - f\right) \right|^2,$$

es decir,

$$\mathcal{G}(f) = \mathcal{H}\left(\frac{1}{2} - f\right).$$

Debido a la definición (3.50), a  $\{g_l\}$  se le conoce como *filtro de cuadratura en espejo* asociado a  $\{h_l\}$ , este nombre es debido al hecho de que si graficamos las funciones  $\mathcal{H}(f) = |H(f)|^2$  y  $\mathcal{G}(f) = |G(f)|^2$  para  $f$  en  $[0, 1/2)$  obtenemos lo siguiente

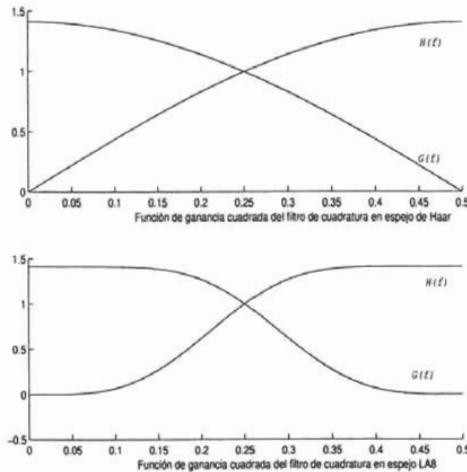


Figura 3.1: Gráfica de la función de ganancia cuadrada

Notemos que de la expresión (3.52) tenemos que

$$\mathcal{G}(1/4) = \mathcal{H}(1/4).$$

Si imaginamos a la recta  $f = 1/4$  como un espejo, entonces  $\mathcal{G}(f)$  se obtiene reflejando a  $\mathcal{H}(f)$  sobre este espejo. La razón por la cuál sólo analizamos el comportamiento de  $\mathcal{G}(f)$  y  $\mathcal{H}(f)$  en el intervalo  $(0, 1/2]$  es porque ambas son funciones pares periódicas (con periodo 1) de  $f$ .

Una expresión alternativa a (3.33) utilizando lo recién demostrado es la siguiente:

$$\mathcal{G}(f) + \mathcal{G}\left(f + \frac{1}{2}\right) = 2 \quad \text{para toda } f \quad (3.53)$$

alternativamente

$$\mathcal{G}(f) + \mathcal{H}(f) = 2 \quad \text{para toda } f. \quad (3.54)$$

Supongamos primero que  $\mathcal{H}(f) + \mathcal{H}(f + \frac{1}{2}) = 2$  entonces

$$\mathcal{G}(f) = \mathcal{H}\left(\frac{1}{2} - f\right) = \mathcal{H}\left(\frac{1}{2} - f - 1\right) = \mathcal{H}\left(-\left(f + \frac{1}{2}\right)\right) = \mathcal{H}\left(f + \frac{1}{2}\right)$$

en la primera igualdad se utilizó el hecho de que la función de ganancia tiene periodo uno, para la última igualdad se utilizó el hecho de que  $\mathcal{H}(f) = \mathcal{H}(-f)$ , por otra parte,

$$\mathcal{G}\left(f + \frac{1}{2}\right) = \mathcal{H}\left(\frac{1}{2} - \left(f + \frac{1}{2}\right)\right) = \mathcal{H}(-f) = \mathcal{H}(f)$$

por lo tanto de estas dos últimas igualdades tenemos que

$$\mathcal{G}\left(f + \frac{1}{2}\right) + \mathcal{G}(f) = 2$$

alternativamente,

$$\mathcal{H}(f) + \mathcal{G}(f) = 2.$$

Recíprocamente, supongamos ahora que para toda  $f$  se cumple

$$\mathcal{G}\left(f + \frac{1}{2}\right) + \mathcal{G}(f) = 2$$

o

$$\mathcal{H}(f) + \mathcal{G}(f) = 2,$$

entonces, dado que

$$\mathcal{G}(f) = \mathcal{H}\left(\frac{1}{2} - f\right) = \mathcal{H}\left(\frac{1}{2} + f\right)$$

tenemos

$$\mathcal{G}\left(f + \frac{1}{2}\right) = \mathcal{H}\left(\left(f + \frac{1}{2}\right) + \frac{1}{2}\right) = \mathcal{H}(f + 1) = \mathcal{H}(f)$$

por lo tanto de estas dos expresiones se tiene lo que se buscaba demostrar, es decir,

$$2 = \mathcal{G}(f) + \mathcal{G}\left(f + \frac{1}{2}\right) = \mathcal{H}(f) + \mathcal{H}\left(f + \frac{1}{2}\right).$$



Al filtro  $\{h_l\}$  se le denomina filtro de frecuencias altas y a  $\{g_l\}$  filtro de frecuencias bajas, a su vez a ambos se les llama filtros de media banda ya que dividen la banda de frecuencias  $[0, 1/2]$  a la mitad.

El filtro de escala  $\{g_l\}$  satisface las siguientes tres propiedades:

$$\sum_{l=0}^{L-1} g_l = \sqrt{2} \quad \text{ó} \quad \sum_{l=0}^{L-1} g_l = -\sqrt{2}, \quad (3.55)$$

$$\sum_{l=0}^{L-1} g_l^2 = 1 \quad (3.56)$$

$$\sum_{l=0}^{L-1} g_l g_{l+2n} = \sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0, \quad \text{para toda } n \neq 0. \quad (3.57)$$

A continuación justificaremos estas propiedades. Como la condición (3.53) se satisface para toda  $f$ , entonces en particular se cumple para  $f = 0$ , es decir:

$$G(0) + G\left(\frac{1}{2}\right) = 2,$$

es decir,

$$\begin{aligned} 2 &= |G(0)|^2 + |G(1/2)|^2 = \left| \sum_{l=0}^{L-1} g_l e^{-i2\pi l(0)} \right|^2 + \left| \sum_{l=0}^{L-1} g_l e^{-i2\pi l(1/2)} \right|^2 \\ &= \left| \sum_{l=0}^{L-1} g_l \right|^2 + \left| \sum_{l=0}^{L-1} g_l e^{-i\pi l} \right|^2 = \left| \sum_{l=0}^{L-1} g_l \right|^2 + \left| \sum_{l=0}^{L-1} (-1)^l g_l \right|^2 \\ &= \left| \sum_{l=0}^{L-1} g_l \right|^2 + \left( (-1)^{L-1} \right)^2 \left| \sum_{l=0}^{L-1} (-1)^l g_l \right|^2 = \left| \sum_{l=0}^{L-1} g_l \right|^2 + \left| \sum_{l=0}^{L-1} (-1)^{L-1+l} g_l \right|^2 \\ &= \left| \sum_{l=0}^{L-1} g_l \right|^2 + \left| \sum_{l=0}^{L-1} (-1)^l g_{L-1-l} \right|^2 = \left| \sum_{l=0}^{L-1} g_l \right|^2 + \left| \sum_{l=0}^{L-1} h_l \right|^2 \\ &= \left| \sum_{l=0}^{L-1} g_l \right|^2 \end{aligned}$$

por lo tanto

$$\sum_{l=0}^{L-1} g_l = \pm\sqrt{2}.$$

Verifiquemos que  $\sum_{l=0}^{L-1} g_l^2 = 1$ ; utilizando la ecuación (3.49)

$$g_l^2 = \left( (-1)^l h_{L-1-l} \right)^2 = (-1)^{2l} h_{L-1-l}^2 = h_{L-1-l}^2$$

por lo tanto

$$\sum_{l=0}^{L-1} g_l^2 = \sum_{l=0}^{L-1} h_{L-1-l}^2 = \sum_{l=0}^{L-1} h_l^2 = 1.$$

Finalmente probemos que

$$\sum_{l=0}^{L-1} g_l g_{l+2n} = 0.$$

A saber, tenemos que

$$g_l = (-1)^{l+1} h_{L-1-l}$$

por lo tanto

$$\begin{aligned} g_{l+2n} &= (-1)^{l+2n} h_{L-1-(l+2n)} = (-1)^l h_{L-1-(l+2n)}, \text{ entonces} \\ \sum_{l=0}^{L-1} g_l g_{l+2n} &= - \sum_{l=0}^{L-1} (-1)^l h_{L-1-l} (-1)^l h_{L-1-(l+2n)} = - \sum_{l=0}^{L-1} (-1)^{2l} h_{L-1-l} h_{L-1-(l+2n)} \\ &= - \sum_{l=0}^{L-1} h_l h_{l+2n} = 0 \end{aligned}$$

por lo tanto

$$\sum_{l=0}^{L-1} g_l g_{l+2n} = - \sum_{l=0}^{L-1} h_l h_{l+2n} = 0.$$

ya que se satisface la ecuación (3.31) por ser  $\{h_l\}$  filtro de onduletas.

Al filtro  $\{g_l\}$  se le llama *filtro de bajas* puesto que filtra frecuencias en el intervalo  $[-1/4, 1/4]$ , obsérvese la figura (3.1). El soporte de  $\mathcal{G}(f)$  es aproximadamente este intervalo.

Definamos el filtro  $\{h_l^\dagger\}$ , que se obtiene de agregar un cero entre dos elementos consecutivos del filtro original  $\{h_l\}$ , es decir, dado el filtro  $\{h_0, h_1, h_2, \dots, h_{L-2}, h_{L-1}\}$  sea

$$\{h_l^\dagger\} = \{h_0, 0, h_1, 0, \dots, h_{L-1}\},$$

la longitud del filtro resultante es  $2L - 1$ ; se define el filtro  $\{h_{2,l}\} \equiv \{g \star h_l^\dagger\}$  resultante de la convolución entre los filtros  $\{g_l\}$  de longitud  $L$  y  $\{h_l^\dagger\}$  filtro de longitud  $2L - 1$ , entonces por la ecuación (2.17) se tiene que la longitud del filtro  $\{h_{2,l}\}$  es:  $L_2 = L + (2L - 1) - 1 = 3L - 2$ .

Consideremos el caso en el que al filtro  $\{h_l\}$  se le agregan  $m$  ceros entre dos elementos consecutivos del mismo, es decir,

$$h_0, \underbrace{0, 0, \dots, 0}_m, h_1, \underbrace{0, 0, \dots, 0}_m, \dots, \underbrace{0, 0, \dots, 0}_m, h_{L-1}. \quad (3.58)$$

entonces su función de transferencia asociada está dada por

$$H((m+1)f). \quad (3.59)$$

Para justificar la aseveración anterior, utilicemos un razonamiento inductivo:

Primero veamos que esto es cierto para  $m = 0$ , es decir, el caso en el que no se agregaron ceros, entonces la ecuación (3.59) se convierte en  $H(f)$  lo cual es cierto. Ahora veamos que también es cierto para  $m = 1$ :

$$\{h_0, 0, h_1, 0, \dots, h_{L-1}\} = \{h_0^1, h_1^1, h_2^1, \dots, h_{2L-2}^1\}$$

entonces su función de transferencia es:

$$\sum_{l=0}^{2(L-1)} h_l^1 e^{-i2\pi fl} = \sum_{l=0}^{L-1} h_{2l}^1 e^{-i2\pi(2f)l} = \sum_{l=0}^{L-1} h_l e^{-i2\pi(2f)l} = H(2f).$$

Para  $m = 2$ :

$$\{h_0, 0, 0, h_1, 0, 0, \dots, h_{L-1}\} = \{h_0^2, h_1^2, h_2^2, \dots, h_{3L-3}^2\}$$

su función de transferencia es

$$\sum_{l=0}^{3(L-1)} h_l^1 e^{-i2\pi fl} = \sum_{l=0}^{L-1} h_{3l}^1 e^{-i2\pi(3f)l} = \sum_{l=0}^{L-1} h_l e^{-i2\pi(3f)l} = H(3f).$$

Procediendo de manera sucesiva se tiene que si se insertan  $m$  ceros, el filtro quedaría de la forma:

$$\left\{ h_0, \underbrace{0, 0, \dots, 0}_m, h_1, \underbrace{0, 0, \dots, 0}_m, \dots, \underbrace{0, 0, \dots, 0}_m, h_{L-1} \right\} = \{h_0^m, h_1^m, h_2^m, \dots, h_{(m+1)(L-1)}^m\}$$

con función de transferencia

$$\sum_{l=0}^{(m+1)(L-1)} h_l^m e^{-i2\pi fl} = \sum_{l=0}^{L-1} h_{(m+1)l}^m e^{-i2\pi((m+1)f)l} = \sum_{l=0}^{L-1} h_l e^{-i2\pi((m+1)f)l} = H((m+1)f).$$

Entonces, haciendo  $m = 1$  y usando la ecuación (2.14) se tiene que la función de transferencia para  $\{h_{2,l}\}$  es

$$H_2(f) = H(2f)G(f).$$

Así, entonces tenemos que dado un entero  $j$  en el conjunto de índices  $\{2, 3, \dots, J\}$ , para calcular los coeficientes de onduletas  $\mathbf{W}_j$  a escala  $\tau_j = 2^{j-1}$  necesitamos el filtro  $\{h_{j,l}\}$  resultante de la convolución de los siguientes  $j$  filtros:

$$\begin{aligned} k &= 1, & g_0, g_1, \dots, g_{L-1}; \\ k &= 2, & g_0, 0, g_1, 0, \dots, 0, g_{L-1}; \\ &\vdots \\ k &= j-1, & g_0, \underbrace{0, 0, \dots, 0}_{2^{j-2}-1}, g_1, \underbrace{0, 0, \dots, 0}_{2^{j-2}-1}, \dots, \underbrace{0, 0, \dots, 0}_{2^{j-2}-1}, g_{L-1}; \end{aligned}$$

y finalmente para  $k = j$  se insertan  $2^{j-1} - 1$  ceros en el filtro  $\{h_l\}$ , es decir,

$$h_0, \underbrace{0, 0, \dots, 0}_{2^{j-1}-1}, h_1, \underbrace{0, 0, \dots, 0}_{2^{j-1}-1}, \dots, \underbrace{0, 0, \dots, 0}_{2^{j-1}-1}, h_{L-1}. \quad (3.60)$$

Notemos que para cada  $k = 2, 3, \dots, j-1$  el  $k$ -ésimo filtro de esta cascada se obtiene de insertar un cero entre cada dos coeficientes consecutivos del  $(k-1)$ -ésimo filtro.

Así entonces, la longitud del  $k$ -ésimo filtro en la cascada es:

$$L^{(k)} = L + (2^{k-1} - 1)(L - 1), \quad (3.61)$$

entonces de acuerdo a la ecuación (2.17) la longitud del filtro  $\{h_{j,l}\}$  está dada por:

$$L_j \equiv (2^j - 1)(L - 1) + 1. \quad (3.62)$$

A saber, por la ecuación dada en (2.17) se tiene:

$$\begin{aligned} L_j &= \sum_{m=1}^j L^{(m)} - j + 1 = \sum_{m=1}^j (L + (2^{m-1} - 1)(L - 1)) - j + 1 \\ &= jL + (L - 1) \sum_{m=1}^j (2^{m-1} - 1) - j + 1 = (L - 1) \sum_{m=1}^j 2^{m-1} + 1 \\ &= (L - 1)(2^j - 1) + 1. \end{aligned}$$

Además el filtro en la cascada resultante,  $\{h_{j,l}\}$ , tiene función de transferencia dada por:

$$H_j(f) \equiv H(2^{j-1}f) \prod_{l=0}^{j-2} G(2^l f). \quad (3.63)$$

Para verificar esto, utilizamos la expresión dada en la ecuación (3.59), entonces tenemos que para  $k = 1, 2, \dots, j-1$  la función de transferencia correspondiente al  $k$ -ésimo filtro es tal que

$$\left\{ g_0, \underbrace{0, 0, \dots, 0}_{2^{k-1}-1}, g_1, \underbrace{0, 0, \dots, 0}_{2^{k-1}-1}, \dots, \underbrace{0, 0, \dots, 0}_{2^{k-1}-1}, g_{L-1} \right\} \leftrightarrow G([2^{k-1} - 1 + 1]f) = G(2^{k-1}f)$$

y para  $k = j$  se tiene

$$\left\{ \underbrace{h_0, 0, 0, \dots, 0}_{2^{j-1}-1}, \underbrace{h_1, 0, 0, \dots, 0}_{2^{j-1}-1}, \dots, \underbrace{0, 0, \dots, 0}_{2^{j-1}-1}, h_{L-1} \right\} \leftrightarrow H([2^{j-1} - 1 + 1] f) = H(2^{j-1} f),$$

por lo tanto, utilizando la ecuación (3.59) y la ecuación (2.9), finalmente tenemos la expresión (3.63).

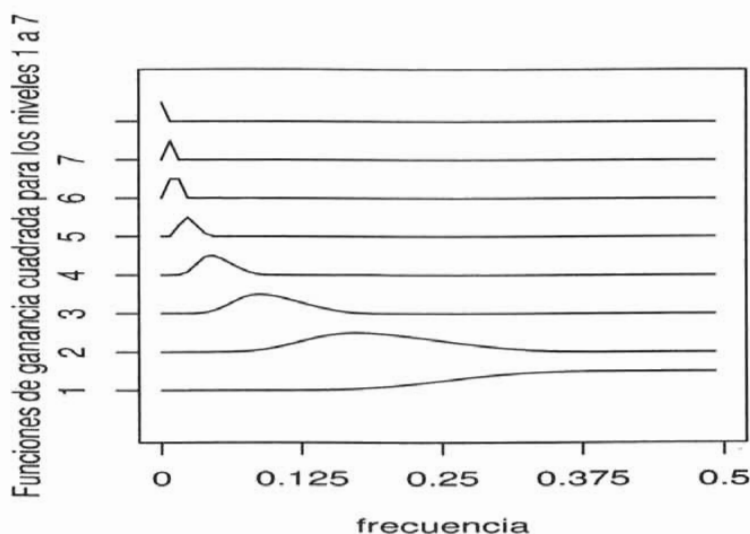


Figura 3.2: Gráfica de las funciones de Ganancia cuadrada para la cascada de filtros utilizando el filtro AL(8) con  $N = 128 = 2^7$ .

Se puede verificar que la función de ganancia cuadrada  $\mathcal{H}_j(f) = |H_j(f)|^2$  tiene soporte aproximadamente sobre el conjunto de frecuencias  $1/2^{j+1} \leq |f| \leq 1/2^j$ . Por esta razón se dice que el filtro  $\{h_{j,t}\}$  filtra frecuencias en la banda de frecuencias  $(\frac{1}{2^{j+1}}, \frac{1}{2^j})$ .

Los coeficientes de onduletas  $W_{j,t}$  asociados a la escala  $\tau_j$  para  $j = 2, 3, \dots, J$  pueden calcularse utilizando el filtro iterado  $\{h_{j,t}\}$

$$2^{j/2}\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l \bmod N} \quad t = 0, 1, 2, \dots, N-1, \quad (3.64)$$

y luego entonces

$$W_{j,t} = 2^{j/2}\widetilde{W}_{j,2^j(t+1)-1} = \sum_{l=0}^{L_j-1} h_{j,l} X_{2^j(t+1)-1-l \bmod N} \quad t = 0, 1, 2, \dots, N_j-1, \quad (3.65)$$

donde  $N_j = N/2^j$ . Dado que se tiene la igualdad  $W_j = \mathcal{W}_j \mathbf{X}$ , se tiene que los renglones de  $\mathcal{W}_j$  están conformados por los elementos del filtro  $\{h_{j,l}^\circ\}$  que es el filtro  $\{h_{j,l}\}$  periodizado a longitud  $N$  e invertido en el tiempo. Por ejemplo, para  $N = 2^3 = 8$ , la matriz  $\mathcal{W}_2$  de dimensión  $2 \times 8$  tiene como primer renglón el vector  $[h_{2,3}^\circ, h_{2,2}^\circ, h_{2,1}^\circ, h_{2,0}^\circ, h_{2,7}^\circ, h_{2,6}^\circ, h_{2,5}^\circ, h_{2,4}^\circ]$  y los renglones restantes son versiones circulares del mismo.

De manera equivalente el filtro  $\{g_{j,t}\}$  se obtiene de hacer la convolución de los  $j$  filtros

$$\begin{aligned} k &= 1, & g_0, g_1, \dots, g_{L-1}; \\ k &= 2, & g_0, 0, g_1, 0, \dots, 0, g_{L-1}; \\ &\vdots \\ k &= j-1, & g_0, \underbrace{0, 0, \dots, 0}_{2^{j-2}-1}, g_1, \underbrace{0, 0, \dots, 0}_{2^{j-2}-1}, \dots, \underbrace{0, 0, \dots, 0}_{2^{j-2}-1}, g_{L-1}; \end{aligned}$$

y finalmente para  $k = j$  se insertan  $2^{j-1} - 1$  ceros en el filtro  $\{g_l\}$ , es decir,

$$g_0, \underbrace{0, 0, \dots, 0}_{2^{j-1}-1}, g_1, \underbrace{0, 0, \dots, 0}_{2^{j-1}-1}, \dots, \underbrace{0, 0, \dots, 0}_{2^{j-1}-1}, g_{L-1}. \quad (3.66)$$

El filtro  $\{g_{j,t}\}$  es tal que  $G_j(f) = |G_j(f)|^2$  tiene soporte aproximadamente sobre el intervalo  $0 \leq |f| \leq 1/2^{j+1}$ . Así como se argumentó para el filtro  $\{h_{j,t}\}$ , se puede verificar que el filtro  $\{g_{j,t}\}$  tiene función de transferencia dada por

$$G_j(f) \equiv \prod_{l=0}^{j-1} G(2^l f). \quad (3.67)$$

Los coeficientes de escala  $\{V_{j,t}\}$  se pueden calcular utilizando el filtro

$$2^{j/2}\tilde{V}_{j,t} \equiv \sum_{l=0}^{L_j-1} g_{j,l}X_{t-l \bmod N}, \quad t = 0, 1, 2, \dots, N-1, \quad (3.68)$$

por tanto

$$V_{j,t} = 2^{j/2}\tilde{V}_{j,2^{j(t+1)-1}} = \sum_{l=0}^{L_j-1} g_{j,l}X_{2^{j(t+1)-1-l} \bmod N}, \quad t = 0, 1, 2, \dots, N_j-1. \quad (3.69)$$

Dado que se tiene que  $V_j = \mathcal{V}_j \mathbf{X}$ , los renglones de  $\mathcal{V}_j$  están conformados por los elementos del filtro  $\{g_{j,l}^\circ\}$  que es el filtro  $\{g_{j,l}\}$  periodizado a longitud  $N$ . Una justificación de que para  $j = 2, 3, \dots, J$ , las matrices  $\mathcal{W}_j$  y  $\mathcal{V}_j$  así construidas son matrices ortogonales puede encontrarse en Percival y Walden (2000).

### 3.3. Paquetes de Onduletas

La transformada discreta de paquetes de onduletas, en inglés *Discrete Wavelet Packet Transform* (DWPT), se puede considerar como una colección de transformaciones ortonormales donde cada una de las cuales se puede calcular como hasta ahora se ha venido haciendo. La DWT tiene asociada una partición  $\mathcal{P}_{DWT}$  del conjunto de frecuencias  $[0, 1/2]$ . De la figura (3.2) se puede ver que los filtros básicos  $\{h_{j,t}\}$  que definen la transformación, son tales que sus funciones de ganancia cuadrada  $\mathcal{H}_j(f)$  tienen soporte aproximadamente sobre el intervalo  $[1/2^{j+1}, 1/2^j]$  para  $j = 1, 2, \dots, 7$ ; finalmente el filtro  $\{g_{7,t}\}$  es tal que  $\mathcal{G}_7$  tiene soporte aproximadamente sobre el intervalo  $[0, 1/2^8]$ .

Para  $N = 2^J$ , donde  $J$  es un entero positivo, la partición asociada a la DWT es

$$\mathcal{P}_{DWT} = \left\{ [1/4, 1/2], [1/8, 1/4], \dots, [1/2^{J+1}, 1/2^J], [0, 1/2^{J+1}] \right\}.$$

Por ejemplo, para  $N = 16 = 2^4$  los renglones de la matriz  $\mathcal{W}_j$  están asociados al intervalo de frecuencias  $[1/2^{j+1}, 1/2^j]$  para  $j = 1, 2, 3, 4$ , mientras que el vector  $\mathcal{V}_4$  está asociado al intervalo  $[0, 1/2^5]$ . La figura (3.3) es un esquema de la partición del intervalo  $[0, 1/2]$ , cuya transformación correspondiente dentro de la colección *DWPT*, es la *DWT*.



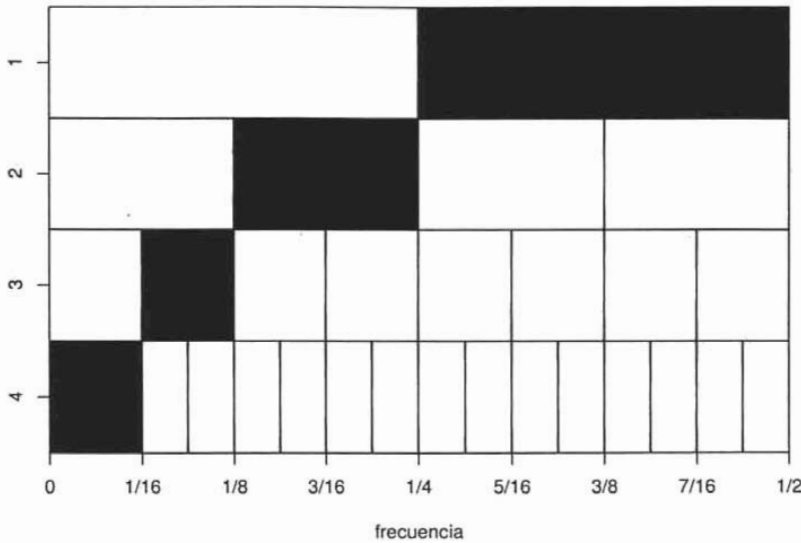


Figura 3.3: Ilustración de la partición  $DWT$  para  $N = 16 = 2^4$

Para cada nivel  $j = 1, 2, \dots, J$  y para cada banda o escala de frecuencias  $n \equiv n_j = 0, 1, 2, \dots, 2^j - 1$ , se define el intervalo  $\mathcal{I}_{j,n} = [n/2^{j+1}, (n+1)/2^{j+1}]$ . Una forma de entender a la  $DWPT$  es como una colección de transformaciones ortogonales, tal que a cada partición  $\mathcal{P}$ , del intervalo  $[0, 1/2]$  en subintervalos  $\mathcal{I}_{j,n}$  le corresponde una (y sólo una) transformación dentro de esta colección.

En el ejemplo anterior para  $N = 2^4$  y para la partición  $\mathcal{P}_{DWT}$  de la figura (3.3), la transformación correspondiente dentro de la  $DWPT$  es la  $DWT$ , pero una partición  $\mathcal{P}$  diferente da origen a otra transformación ortogonal.

Nuestro objetivo es describir los filtros o sucesiones básicas para calcular, por ejemplo, la transformación asociada a la partición  $\mathcal{P}_1 = \{\mathcal{I}_{1,0}, \mathcal{I}_{2,2}, \mathcal{I}_{2,3}\}$  o los filtros asociados a la partición  $\mathcal{P}_2 = \{\mathcal{I}_{2,0}, \mathcal{I}_{2,1}, \mathcal{I}_{2,2}, \mathcal{I}_{2,3}\}$  que se ilustra en la figura (3.4).

En la sección (3.2) construimos los filtros  $\{h_{j,l}\}$  y  $\{g_{j,l}\}$  para  $j = 1, 2, \dots, J$ , tales que sus funciones de ganancia cuadrada  $\mathcal{H}_j(f)$  y  $\mathcal{G}_j(f)$  tienen soporte aproximadamente sobre los elementos de la partición  $\mathcal{P}_{DWT}$ , compárense las figuras (3.2) y (3.3).

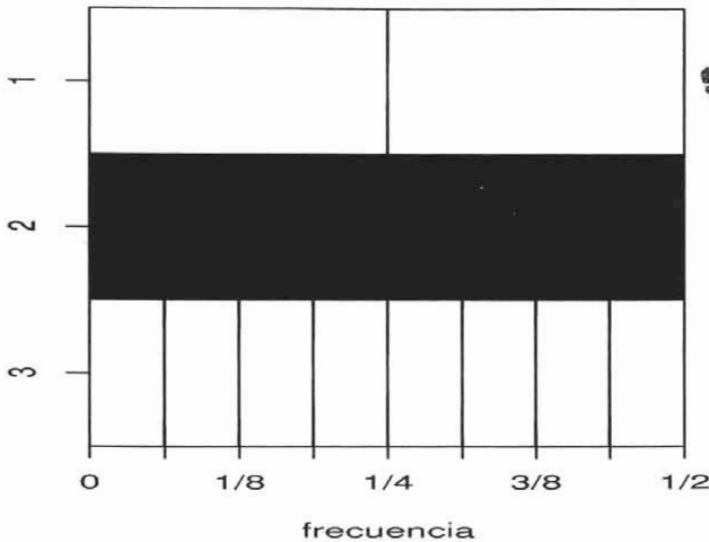


Figura 3.4: Ilustración de la partición  $\mathcal{P}_2 = \{\mathcal{I}_{2,0}, \mathcal{I}_{2,1}, \mathcal{I}_{2,2}, \mathcal{I}_{2,3}\}$

Para facilitar la notación nos referiremos a cada nodo (o localidad) en los esquemas de las figuras (3.3) y (3.4) con el índice  $(j, n)$ . Por ejemplo en el esquema (3.4) están iluminados los nodos  $(2, 0)$ ,  $(2, 1)$ ,  $(2, 2)$  y  $(2, 3)$ .

Para  $j = 1, 2, \dots, J$  y  $n \equiv n_j = 0, 1, 2, \dots, 2^j - 1$  dado un nodo  $(j, n)$ , se puede determinar un filtro  $\{u_{j,n,l}, l = 0, 1, \dots, L_j - 1\}$  correspondiente, es decir, un filtro tal que su función de ganancia al cuadrado

$$U_{j,n}(f) = \left| \sum_{l=0}^{L_j-1} u_{j,n,l} e^{-i2\pi fl} \right|^2,$$

tenga soporte aproximadamente sobre el intervalo  $\mathcal{I}_{j,n}$ .

Denotemos por  $[x]$  a la parte entera del número real  $x$ . Dado un nivel  $j = 1, 2, \dots, J$  y  $n \in \{0, 1, \dots, 2^j - 1\}$  sea

$$u_{n,l} \equiv \begin{cases} g_l, & n = 0 \text{ ó } 3 \text{ mod } 4 \\ h_l, & n = 1 \text{ ó } 2 \text{ mod } 4. \end{cases}$$

Para  $j = 1$ ,  $n = 0, 1$  los filtros correspondientes a  $\mathcal{I}_{1,0}$  y  $\mathcal{I}_{1,1}$  son

$$u_{1,0,l} \equiv g_l, \quad l = 0, 1, \dots, L-1; \quad y$$

$$u_{1,1,l} \equiv h_l, \quad l = 0, 1, \dots, L-1.$$

Para los nodos  $(j, n)$ , con  $j > 1$ , sea

$$u_{j,n,l} \equiv \begin{cases} \sum_{k=0}^{L-1} u_{n,k} u_{j-1, \lfloor \frac{n}{2} \rfloor, l-2^{j-1}k} & l = 0, 1, \dots, L_j - 1, \\ 0, & l < 0, l \geq L_j. \end{cases}$$

donde  $L_j \equiv (2^j - 1)(L - 1) + 1$ .

Por ejemplo, para  $j = 2$ ,

$$\begin{aligned} u_{2,0,l} &= \sum_{k=0}^{L-1} u_{0,k} u_{1,0,l-2k} = \sum_{k=0}^{L-1} g_k g_{l-2k}, \\ u_{2,1,l} &= \sum_{k=0}^{L-1} u_{1,k} u_{1,0,l-2k} = \sum_{k=0}^{L-1} h_k g_{l-2k}, \\ u_{2,2,l} &= \sum_{k=0}^{L-1} u_{2,k} u_{1,1,l-2k} = \sum_{k=0}^{L-1} h_k h_{l-2k}, \\ u_{2,3,l} &= \sum_{k=0}^{L-1} u_{3,k} u_{1,1,l-2k} = \sum_{k=0}^{L-1} g_k h_{l-2k}. \end{aligned}$$

Estos filtros son los necesarios para definir la transformación ortogonal asociada a la partición  $\mathcal{P}_2$  que se muestra en la figura (3.4).

Dado el conjunto de datos  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  se puede construir una matriz  $\mathcal{W}_{\mathcal{P}_2}$  de dimensiones  $N \times N$ , la cuál estará particionada en 4 bloques de dimensiones  $\frac{N}{4} \times N$ , donde los primeros  $\frac{N}{4}$  renglones son versiones circulares invertidas en el tiempo del filtro periodizado  $\{u_{2,3,l} : l = 0, 1, \dots, N-1\}$ . Los siguientes  $\frac{N}{4}$  renglones  $\frac{N}{4} + 1, \dots, N/2$  están conformados por versiones circulares invertidas en el tiempo del filtro periodizado  $\{u_{2,2,l} : l = 0, 1, \dots, N-1\}$  así sucesivamente hasta completar la matriz ortonormal  $\mathcal{W}_{\mathcal{P}_2}$ .

Entonces, el igual que como se ha venido haciendo hasta ahora, tenemos la expresión

$$\mathbf{W}_2 = \mathcal{W}_{P_2} \mathbf{X}.$$

Dado que el filtro  $\{u_{j,n,l}\}$  está determinado por los filtros  $\{h_l\}$  y  $\{g_l\}$ , entonces su función de transferencia  $U_{j,n}$  depende de las funciones de transferencia  $H(f)$  y  $G(f)$  correspondientes al filtro de onduletas y al filtro de escala respectivamente.

Para hacer ver esta dependencia defínase

$$M_0 \equiv G(\cdot) \quad \text{y} \quad M_1 \equiv H(\cdot).$$

A cada nodo  $(j, n)$  en los esquemas de las figuras (3.3) y (3.4) se le asocia un vector  $C_{j,n}$  de longitud  $j$ , donde sus elementos son iguales a uno o cero.

Para  $j = 1$ , se define  $C_{1,0} \equiv [0]$  y  $C_{1,1} \equiv [1]$ . Para el caso  $j > 1$ , se define  $c_{j,n}$ , con  $n = 0, 1, \dots, 2^j - 1$ , de manera recursiva en términos de  $C_{j-1,n'}$  con  $n' = 0, 1, \dots, 2^{j-1} - 1$  aplicando la siguiente regla:

- (a) Si  $n = 0$  ó  $3 \bmod 4$ , entonces  $C_{j,n}$  se obtiene agregando un cero al final de  $c_{j-1, \lfloor \frac{n}{2} \rfloor}$ .
- (b) Si  $n = 1$  ó  $2 \bmod 4$ , entonces  $C_{j,n}$  se obtiene agregando un uno al final de  $c_{j-1, \lfloor \frac{n}{2} \rfloor}$ .

Ilustremos la regla anterior con un ejemplo:

**Ejemplo 3.1.** Sean  $n = 7$  y  $j = 4$ ; dado que

$$7 = 3 \bmod 4, \text{ se tiene que } C_{4,7} = (C_{3,3}, 0);$$

además

$$3 = 3 \bmod 4, \text{ por lo tanto } C_{3,3} = (C_{2,1}, 0)$$

aplicando la regla nuevamente tenemos que

$$1 = 1 \bmod 4, \text{ por lo cual } C_{2,1} = (C_{1,0}, 1).$$

Por lo tanto dado que

$$C_{1,0} \equiv [0],$$

entonces

$$C_{2,1} = (C_{1,0}, 1) = (0, 1);$$

luego

$$C_{3,3} = (C_{2,1}, 0) = (0, 1, 0),$$

finalmente,

$$C_{4,7} = (C_{3,3}, 0) = (0, 1, 0, 0).$$

Por ejemplo para la figura (3.3), se tiene  $j = 4$  y para cada  $n = 0, 1, \dots, 2^4 - 1 = 15$ , los vectores necesarios son los siguientes:

$$\begin{aligned} C_{1,1} &= [1], \\ C_{2,1} &= (C_{1,0}, 1) = (0, 1), \\ C_{3,1} &= (C_{2,0}, 1) = (0, 0, 1), \\ C_{4,1} &= (C_{3,0}, 1) = (0, 0, 0, 1). \end{aligned}$$

Consideremos la figura (3.5), donde se tienen los vectores  $C_{j,n}$  con  $j = 1, 2, 3$  y  $n = 0, 1, \dots, 2^j - 1 = 7$ . Para poder determinar su función de transferencia  $U_{j,n}(\cdot)$ , sea  $C_{j,n,m}$  el  $m$ -ésimo elemento de  $C_{j,n}$ , entonces se define la función de transferencia como sigue

$$U_{j,n}(f) \equiv \prod_{m=0}^{j-1} M_{C_{j,n,m}}(2^m f). \quad (3.70)$$

$$U_{1,0}(f) = \prod_{m=0}^0 M_{C_{1,0,m}}(2^m f) = M_0(f) = G(f)$$

$$U_{1,1}(f) = \prod_{m=0}^0 M_{C_{1,1,m}}(2^m f) = M_1(f) = H(f)$$

$$U_{2,0}(f) = \prod_{m=0}^1 M_{C_{2,0,m}}(2^m f) = M_{C_{2,0,0}}(f) M_{C_{2,0,1}}(2f) = M_0(f) M_0(2f) = G(f) G(2f)$$

$$U_{2,1}(f) = \prod_{m=0}^1 M_{C_{2,1,m}}(2^m f) = M_{C_{2,1,0}}(f) M_{C_{2,1,1}}(2f) = M_0(f) M_1(2f) = G(f) H(2f)$$

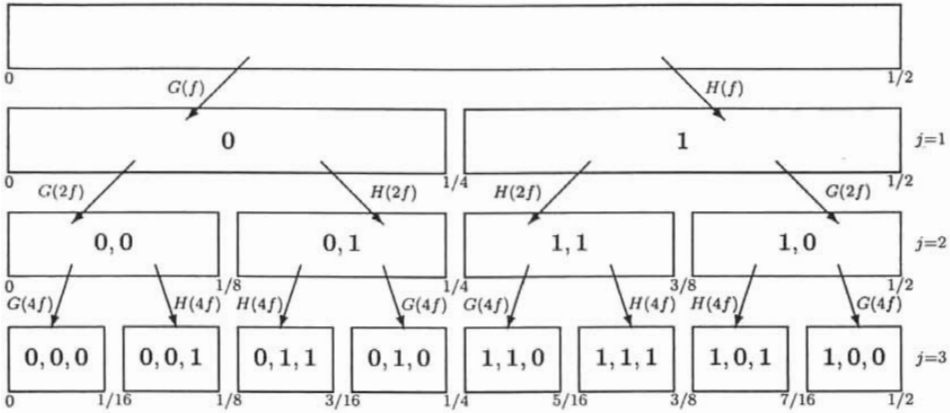


Figura 3.5: Ilustración de las secuencias binarias  $\{c_{j,n}\}$ , con  $j = 3$ .

$$U_{2,2}(f) = \prod_{m=0}^1 M_{C_{2,2,m}}(2^m f) = M_{C_{2,2,0}}(f) M_{C_{2,2,1}}(2f) = M_1(f) M_1(2f) = H(f) H(2f)$$

$$U_{2,3}(f) = \prod_{m=0}^1 M_{C_{2,3,m}}(2^m f) = M_{C_{2,3,0}}(f) M_{C_{2,3,1}}(2f) = M_1(f) M_0(2f) = H(f) G(2f)$$

$$\begin{aligned} U_{3,0}(f) &= \prod_{m=0}^2 M_{C_{3,0,m}}(2^m f) = M_{C_{3,0,0}}(f) M_{C_{3,0,1}}(2f) M_{C_{3,0,2}}(4f) \\ &= M_0(f) M_0(2f) M_0(4f) = G(f) G(2f) G(4f) \end{aligned}$$

$$\begin{aligned} U_{3,1}(f) &= \prod_{m=0}^2 M_{C_{3,1,m}}(2^m f) = M_{C_{3,1,0}}(f) M_{C_{3,1,1}}(2f) M_{C_{3,1,2}}(4f) \\ &= M_0(f) M_0(2f) M_1(4f) = G(f) G(2f) H(4f) \end{aligned}$$

$$\begin{aligned} U_{3,2}(f) &= \prod_{m=0}^2 M_{C_{3,2,m}}(2^m f) = M_{C_{3,2,0}}(f) M_{C_{3,2,1}}(2f) M_{C_{3,2,2}}(4f) \\ &= M_0(f) M_1(2f) M_1(4f) = G(f) H(2f) H(4f) \end{aligned}$$

$$\begin{aligned} U_{3,3}(f) &= \prod_{m=0}^2 M_{C_{3,3,m}}(2^m f) = M_{C_{3,3,0}}(f) M_{C_{3,3,1}}(2f) M_{C_{3,3,2}}(4f) \\ &= M_0(f) M_1(2f) M_0(4f) = G(f) H(2f) G(4f) \end{aligned}$$

$$\begin{aligned}
U_{3,4}(f) &= \prod_{m=0}^2 M_{C_{3,4,m}}(2^m f) = M_{C_{3,4,0}}(f) M_{C_{3,4,1}}(2f) M_{C_{3,4,2}}(4f) \\
&= M_1(f) M_1(2f) M_0(4f) = H(f) H(2f) G(4f) \\
U_{3,5}(f) &= \prod_{m=0}^2 M_{C_{3,5,m}}(2^m f) = M_{C_{3,5,0}}(f) M_{C_{3,5,1}}(2f) M_{C_{3,5,2}}(4f) \\
&= M_1(f) M_1(2f) M_0(4f) = H(f) H(2f) H(4f) \\
U_{3,6}(f) &= \prod_{m=0}^2 M_{C_{3,6,m}}(2^m f) = M_{C_{3,6,0}}(f) M_{C_{3,6,1}}(2f) M_{C_{3,6,2}}(4f) \\
&= M_1(f) M_0(2f) M_1(4f) = H(f) G(2f) H(4f) \\
U_{3,7}(f) &= \prod_{m=0}^2 M_{C_{3,7,m}}(2^m f) = M_{C_{3,7,0}}(f) M_{C_{3,7,1}}(2f) M_{C_{3,7,2}}(4f) \\
&= M_1(f) M_0(2f) M_0(4f) = H(f) G(2f) G(4f)
\end{aligned}$$

Entonces en general se tendría un esquema como el que se muestra en la figura (3.6)

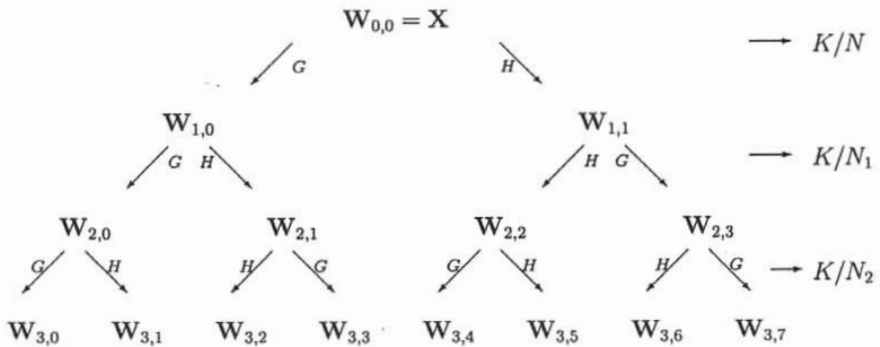


Figura 3.6: Esquema de una Wavelet Packet Table para  $j = 3$

Recordemos que en la sección (3.2) se le llamo *downsampling* al procedimiento de dar de baja la mitad de los coeficientes (a escala  $\tau = 1$ ) resultantes de filtrar los datos  $\mathbf{X}$ . Para escalas  $\tau_j$  con  $j > 1$ , las ecuaciones (3.64) y (3.65) establecen que el *downsampling apropiado* corresponde a conservar únicamente  $N_j = N/2^j$  coeficientes de onduletas. De hecho, de los coeficientes resultantes de filtrar a  $\mathbf{X}$  con el filtro  $\{h_{j,l}\}$ , solamente nos quedamos con aquellos coeficientes cuyos índices son  $2^j - 1, 2^j \cdot 2 - 1, 2^j \cdot 3 - 1, \dots, 2^j \cdot N_j - 1$ .

Alternativamente al uso de matrices, dada una partición  $\mathcal{P} = \{\mathcal{I}_{j_1, n_1}, \dots, \mathcal{I}_{j_k, n_k}\}$  del intervalo  $[0, 1/2)$ , para cada  $(j, n) \in \{(j_1, n_1), \dots, (j_k, n_k)\}$  podemos calcular los coeficientes correspondientes a la sub-banda  $n$  del nivel  $j$  en términos de una operación de filtrado (con un *downsampling apropiado*):

$$\mathbf{W}_{j,n,t} = \sum_{l=0}^{L_j-1} u_{j,n,t} \mathbf{X}_{2^j[t+1]-1-l \bmod N}, \quad t = 0, 1, 2, 3, \dots, N_j - 1. \quad (3.71)$$

Si procedemos de manera análoga para cualquier  $j$  y  $n = 0, 1, 2, \dots, 2^{j-1} - 1$  siempre podremos determinar el orden de filtrado para obtener el  $\mathbf{W}_{j,n}$  correspondiente. Para terminar esta sección recordemos que cada  $\mathbf{W}_{j,n}$  en la *DWPT* está asociado a bandas de longitud  $1/2^{j+1}$ , es decir a la longitud del intervalo  $\mathcal{I}_{j,n}$ .

### 3.4. Algoritmo para la selección de base.

Como se ha venido mencionando, se pueden extraer distintas transformaciones ortonormales de una tabla de paquetes de onduletas a un cierto nivel  $J_0$ ; recordemos que cada transformación ortonormal está asociada a una partición del intervalo de frecuencias  $[0, 1/2]$ , usando intervalos diádicos de la forma  $[\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}})$ . La pregunta es ¿de que manera podemos elegir a una de estas particiones ó transformaciones de modo que sea óptima en algún sentido para una serie de tiempo en particular?.

Sean  $N = 2^J$ , para  $J$  un entero positivo y  $J_0$  un entero en  $\{1, 2, \dots, J\}$ . Consideremos a la colección de parejas  $(j, n)$  que denotan los índices de los nodos en una tabla de paquetes de onduletas (es decir, un esquema como el de las figuras (3.3) y (3.4)),

$$\mathcal{N} = \{(j, n) : j = 1, 2, \dots, J_0; n = 0, 1, \dots, 2^j - 1\}$$

Dada una partición  $\mathcal{P}$  del intervalo  $[0, 1/2)$  en intervalos diádicos  $\mathcal{I}_{j,n}$ , donde  $j \in \{1, 2, \dots, J_0\}$  y  $n \in \{0, 1, \dots, 2^j - 1\}$ , sea  $\mathcal{C}_{\mathcal{P}} = \{(j, n) : \mathcal{I}_{j,n} \in \mathcal{P}\}$ , luego entonces  $\mathcal{C}_{\mathcal{P}} \subset \mathcal{N}$ .

Coiffman y Wickerhauser (1992) abordaron este problema bajo el nombre de *Algoritmo para la selección de la mejor base*, que consiste en los siguientes dos pasos



- (a) Dada la tabla de paquete de onduletas a nivel  $J_0$  entonces para cada  $(j, n) \in \mathcal{N}$  asociamos a cada  $\mathbf{W}_{j,n}$  un costo  $M(\mathbf{W}_{j,n})$ , donde  $M(\cdot)$  es una función de costo aditiva de la forma:

$$M(\mathbf{W}_{j,n}) \equiv \sum_{t=0}^{N_j-1} m(|W_{j,n,t}|), \quad (3.72)$$

y  $m$  es una función con valores reales definida en  $[0, \infty)$  con  $m(0) = 0$ .

- (b) La transformación ortonormal,  $\mathcal{P}$ , óptima que se puede extraer de la tabla del paquete de onduletas es la solución del problema

$$\min_{\mathcal{C}_{\mathcal{P}}} \sum_{(j,n) \in \mathcal{C}_{\mathcal{P}}} M(\mathbf{W}_{j,n}), \quad (3.73)$$

es decir, se busca la transformación ortonormal dada por  $\mathcal{C}_{\mathcal{P}} \subset \mathcal{N}$  tal que la suma de los costos sobre el conjunto de pares  $(j, n) \in \mathcal{C}_{\mathcal{P}}$  se minimice.

En Percival y Walden (2000) se dan tres ejemplos de funciones costo aditivas las cuales a continuación enunciamos:

- (a) La norma  $-\ell^2 \log(\ell^2)$

$$m(|\bar{W}_{j,n,t}|) = \begin{cases} -\bar{W}_{j,n,t}^2 \log(\bar{W}_{j,n,t}^2), & \text{si } W_{j,n,t} \neq 0; \\ 0, & \text{si } W_{j,n,t} = 0, \end{cases} \quad (3.74)$$

donde  $\bar{W}_{j,n,t} \equiv W_{j,n,t}/\|\mathbf{X}\|$ .

- (b) La funcional *threshold* (umbral), es decir, la cantidad de términos que exceden cierto umbral  $\delta$

$$m(|W_{j,n,t}|) = \begin{cases} 1, & \text{si } |W_{j,n,t}| > \delta; \\ 0, & \text{e.o.c.} \end{cases} \quad (3.75)$$

- (c) La funcional de costo en  $\ell_p$  definida por

$$m(|W_{j,n,t}|) = |W_{j,n,t}|^p, \quad (3.76)$$

para  $0 < p < 2$ .

Finalmente en Contreras y Walden (2002), se propone un algoritmo para seleccionar la mejor base, por medio de una prueba de hipótesis, para determinar si los coeficientes del paquete de onduletas son ruido blanco o no lo son. Al final se puede obtener una base ortonormal de paquetes de onduletas cuya correlación sea mínima.

El método se describe en el siguiente algoritmo:

**Algoritmo 1.** (a) Para el nivel  $J = 1$  se calculan los coeficientes del paquete de onduletas  $\mathbf{W}_{1,0}$  y  $\mathbf{W}_{1,1}$ .

(b) Se aplica una prueba de hipótesis de ruido blanco en los coeficientes en  $\mathbf{W}_{1,0}$  y  $\mathbf{W}_{1,1}$ :

$H_0$  : los coeficientes en  $\mathbf{W}_{j,n}$  son una muestra de ruido blanco,

$H_1$  : No  $H_0$ .

En el caso de que  $H_0$  sea rechazada se calculan los coeficientes del paquete de onduletas del siguiente nivel, por ejemplo, para el caso  $j = 1, n = 1$  supongamos que  $H_0$  es rechazada, entonces en el siguiente nivel,  $j = 2$  tenemos que calcular el vector  $\mathbf{W}_{2,2}$  y  $\mathbf{W}_{2,3}$ . Por otra parte, en el caso de que  $H_0$  no se rechace, entonces el intervalo  $\mathcal{I}_{j,n}$  se agrega a la partición del intervalo  $[0, 1/2]$  en sub-intervalos diádicos disjuntos, es decir, en el siguiente nivel no habrá necesidad de calcular  $\mathbf{W}_{j+1,2n}$  y  $\mathbf{W}_{j+1,2n+1}$ .

(c) Para  $j = 2, 3, \dots, J_0 - 1$ , donde  $J_0$  es el nivel más alto de la transformación, se repite el paso (b) en todos los coeficientes de los vectores calculados en el paso anterior.

El periodograma asociado a una muestra  $Y_1, Y_2, \dots, Y_n$  de un proceso estacionario de segundo orden,  $\{Y_t\}_t$  se define como:

$$\widehat{S}(f) \equiv \sum_{t=-(N-1)}^{N-1} \widehat{\gamma}(t) e^{-i2\pi ft} \quad (3.77)$$

donde  $\widehat{\gamma}(t) = \widehat{cov}(Y_k, Y_{k+t})$

El periodograma puede ser un buen estimador de la densidad espectral. En Brockwell y Davis (1998) se justifica el porqué utilizar el periodograma para realizar la prueba de hipótesis de

ruido blanco por medio de la prueba de bondad de ajuste de Kolmogorov para una distribución completamente especificada. Esta prueba nos servirá para contrastar las hipótesis  $H_0$  y  $H_1$  en el inciso (b) del algoritmo (1). Entonces si definimos a  $C_l$  como

$$C_l = \frac{\sum_{k=1}^l \hat{S}(f_k)}{\sum_{k=1}^{M'_{j-1}} \hat{S}(f_k)} \quad (3.78)$$

para  $l = 1, 2, \dots, M'_{j-1}$ , donde  $f_k = K/M_{j-1}$  y denotamos al número de estas frecuencias que satisfacen la relación  $0 < k/M_{j-1} < 1/2$ , por  $M'_{j-1} = (M_{j-1}/2) - 1$ . Además  $\hat{S}(f_k)$  es el periodograma asociado a los coeficientes de onduletas  $\mathbf{W}_{j,n}$  correspondientes al nodo  $(j, n)$ .

Definiendo

$$D^+ = \max_{1 \leq l \leq M'_{j-1}} \left( \frac{l}{M'_{j-1} - 1} - C_l \right) \quad (3.79)$$

$$D^- = \max_{1 \leq l \leq M'_{j-1}} \left( C_l - \frac{l-1}{M'_{j-1} - 1} \right) \quad (3.80)$$

se puede rechazar la hipótesis nula de que los elementos de  $\mathbf{W}_{j,n}$  son una muestra de ruido blanco a un nivel de confianza  $\alpha$  si

$$D = \max \{ D^+, D^- \}$$

excede el valor de  $D(\alpha)$  de la distribución original. Una aproximación de  $D(\alpha)$  se da en Stephens (1974)([18]) y en Fitzgerald *et al* (2000):

$$\tilde{D}(\alpha) = \frac{C(\alpha)}{(M_j - 1)^{1/2} + 0,12 + \frac{0,11}{(M_j - 1)^{1/2}}}, \quad (3.81)$$

donde  $C(0,05) = 1,358$ .

## Capítulo 4

# Bootstrapping vs. Wavestrapping.

En este capítulo se estudiarán las distintas maneras de realizar remuestreo en los datos originales, este remuestreo se puede realizar de manera directa, es decir, utilizando un método no paramétrico tal como el bootstrap por bloques, o ajustando al conjunto de datos un modelo específico que satisface ciertos criterios referentes a la selección del modelo que mejor ajusta a los datos. La idea del remuestreo es generar conjuntos de datos donde la variabilidad de la estadística en estudio puede ser estimada. A este tipo de remuestreo se le conoce como *bootstrap* o *Método de Computo intensivo*. Básicamente se puede clasificar en dos tipos: *bootstrap Paramétrico* y *bootstrap No Paramétrico*; a continuación damos una breve descripción de ambos tipos de bootstrap.

### 4.1. Caso Paramétrico

#### 4.1.1. Bootstrap de residuales

La idea central en el contexto de series de tiempo de este tipo de remuestreo es ajustar un modelo apropiado a los datos, ya sea un modelo tipo *Autoregresivo* de orden  $p$ ,  $AR(p)$  una vez ajustado el modelo a los datos se procede a calcular los residuales. Utilizando un esquema de remuestreo con reemplazo sobre el conjunto de residuales se producen muestras aleatorias (muestras bootstrap de los mismos).

Por ejemplo supongase que  $\varepsilon_{1,l}^*, \varepsilon_{2,l}^*, \dots, \varepsilon_{n,l}^*$  es la  $l$ -ésima muestra bootstrap de residuales producida donde  $l = 1, 2, \dots, m$ . Al utilizar esta muestra de ruido blanco en el modelo ajustado podemos obtener una muestra bootstrap  $X_1^{(l)}, X_2^{(l)}, \dots, X_n^{(l)}$  de la serie de tiempo de los datos.

Entonces usamos estas muestras para calcular  $\hat{\rho}_1^{(l)}$ . Al final del proceso tenemos una muestra  $\hat{\rho}_1^{(1)}, \hat{\rho}_1^{(2)}, \dots, \hat{\rho}_1^{(m)}$  con las cuales podemos estudiar propiedades distribucionales de la variable aleatoria  $\hat{\rho}_1$ .

Usualmente estos residuales son recentrados con la finalidad de que tengan la misma media que el proceso de ruido blanco. Para ejemplificar lo hasta ahora mencionado consideremos el siguiente algoritmo

**Algoritmo 2.** (a) Sea  $\{Y_t : t = 1, \dots, n\}$  un proceso estacionario de segundo orden, para poder determinar el modelo que mejor ajusta a los datos un criterio útil es el criterio de información corregido de Akaike (AICC). El criterio de AICC (Criteria Akaike's Information Criterion) se utiliza para prevenir una sobreparametrización de un modelo, asignando un costo a cada parámetro adicional introducido en el modelo. El modelo que se seleccione debe de ser aquel que minimize el valor del coeficiente, para mayores referencias ver Brockwell y Davis (1998).

Supongamos, por cuestiones de simplicidad, que el modelo encontrado es un modelo de tipo AR(p), es decir, es un modelo de la forma

$$Y_t = \sum_{u=1}^p \phi(u) Y_{t-u} + \xi_t. \quad (4.1)$$

para el caso  $p = 1$ , nuestro modelo es de la forma:

$$Y_t = \phi Y_{t-1} + \xi_t, \quad (4.2)$$

donde  $\{\xi_t\}$  es un proceso de ruido blanco.

(b) Una vez realizado el ajuste y encontrado el estimador de  $\hat{\phi}$ , se calculan los estimadores de los residuales por medio de la siguiente ecuación:

$$e_t = Y_t - \hat{\phi} Y_{t-1}. \quad t = 2, 3, 4, \dots, n. \quad (4.3)$$

(c) Enseguida se calcula el valor de  $\bar{e}$  para recentrar los residuales, es decir, a cada uno de ellos se les resta su media:

$$e_1 - \bar{e}, e_2 - \bar{e}, \dots, e_n - \bar{e}.$$

(d) Realizando muestreo independiente con reemplazo en los residuales obtenemos simulaciones independientes de ruido blanco, es decir

$$\xi_0^*, \xi_1^*, \xi_2^*, \dots, \xi_n^*.$$

(e) Finalmente definimos a  $Y_0^* \equiv \xi_0^*$  y así se procede a calcular

$$Y_t^* = \widehat{\phi} Y_{t-1}^* + \xi_t^*,$$

para toda  $t = 1, 2, \dots, n$ .

En el algoritmo anterior se está considerando  $|\widehat{\phi}| < 1$  aunque este hecho no garantiza la estacionariedad del proceso así generado. Para poder garantizar que efectivamente se esta generando un proceso estacionario, se sugiere repetir (d) un número grande de veces para después realizar el paso (e)  $n + k$  veces, donde a  $k$  se le llama de periodo de *quema*. Dicho periodo debe de elegirse lo suficientemente grande de modo tal que efectivamente se asegure la estacionariedad de la serie  $\{Y_1^*, Y_2^*, \dots, Y_n^*\}$ , es decir se desecharán los valores  $\{Y_{-k}^*, Y_{-k+1}^*, \dots, Y_0^*\}$ .

Para mayores referencias sobre este método revisar el libro publicado por Davison y Hinkley (1997), quienes aplican dicho método a un conjunto de datos de conteos de manchas en el sol por año, estos datos son observaciones del año 1771 al año 1993. Al realizar el ajuste a los datos con modelos de tipo autoregresivo resultó que el orden del modelo que mejor ajustaba a los datos era uno de orden 9; después de realizar el ajuste y obtener los estimadores de los parámetros calcularon los residuales, se realizó bootstrap sobre estos y luego se calcularon nuevas series bootstrap de datos; a cada serie de datos se les ajustó un modelo de tipo autoregresivo, los autores obtuvieron que en el 74% de los casos el orden del modelo seleccionado fue 9. En un intento nuestro por reproducir el ejemplo y los resultados, se crearon rutinas para ser ejecutadas en el programa **R** que hicieran todo lo descrito anteriormente y así poder entender de mejor manera todo el método de realizar bootstrap en los residuales.

En la repetición del ejemplo propuesto por el texto, dado que no se contaba con los datos en ese mismo periodo de tiempo sino con datos del año 1893 al año 1990, resultó que el modelo autoregresivo que mejor ajustaba a los datos era un autoregresivo de orden 2.

## 4.2. Caso No paramétrico

A continuación revisaremos brevemente algunos métodos no paramétricos para realizar bootstrap.

### 4.2.1. Bootstrap por Bloques

Una segunda propuesta es en la que el remuestreo se realiza no sobre los residuales sino sobre bloques de observaciones consecutivas, a continuación analizaremos el caso más sencillo de estos.

Supongamos que se tiene el conjunto de datos  $\{y_1, y_2, \dots, y_n\}$ , entonces a continuación se procede a construir  $b$  bloques de longitud  $l$ , sin que se superpongan, es decir, que no tengan elementos en común uno con otro, de modo tal que  $n = bl$ , es decir:

$$\begin{aligned} z_1 &= (y_1, y_2, \dots, y_l) \\ z_2 &= (y_{l+1}, y_{l+2}, \dots, y_{2l}) \\ &\vdots \\ z_b &= (y_{l(b-1)+1}, y_{l(b-1)+2}, \dots, y_{l(b-1)+l}) \\ &= (y_{n-l+1}, y_{n-l+2}, \dots, y_n). \end{aligned}$$

Enseguida se realiza el remuestreo con reemplazo equiprobable, pero ahora este se realiza sobre los bloques  $\{z_1, z_2, \dots, z_b\}$ , las cuales se concatenan para así conformar una nueva serie; por ejemplo veamos el caso  $b = 3$  y  $l = 5$ , por lo tanto  $n = bl = 15$ :

$$\{y_1, y_2, y_3, \dots, y_{15}\},$$

entonces tenemos los 3 bloques:

$$\{y_1, y_2, y_3, y_4, y_5\}, \{y_6, y_7, y_8, y_9, y_{10}\}, \{y_{11}, y_{12}, y_{13}, y_{14}, y_{15}\}.$$

Ahora supongamos que como resultado del remuestreo obtuvimos  $z_1^* = z_1$  y  $z_2^* = z_3$  y  $z_3^* = z_3$ , entonces tenemos que la muestra bootstrap es

$$y_t^* = \{y_1^*, y_2^*, y_3^*, y_4^*, y_5^*, y_{11}^*, y_{12}^*, y_{13}^*, y_{14}^*, y_{15}^*, y_{11}^*, y_{12}^*, y_{13}^*, y_{14}^*, y_{15}^*\}.$$

Con el fin de preservar la estructura de dependencia de la serie original, se propone considerar bloques de tamaño lo suficientemente grandes, así al calcular una estadística,  $T^*$ , de la serie  $\{y_t^*\}$ , esta será más parecida en distribución a la que estadística  $T$  que se obtuviera con  $\{y_t\}$ . Por otra parte, si  $T$  es la cantidad a estimar, para poder hacer buenas estimaciones de la variabilidad de  $T$ , las cantidades  $T^*$  obtenidas deben de exhibir un mínimo de variabilidad,

es decir, no se puede considerar  $b$  demasiado grande, por ejemplo si  $b = n$  entonces  $T^*$  no presentará variabilidad alguna. Es por esto que para para lograr este fin, se requiere que el tamaño de los bloques debe de ser pequeño.

En Davison y Hinkley (1997) se sugiere trabajar con bloques de longitud  $l$ , del orden  $n^\gamma$ , con  $\gamma \in (0, 1)$  para poder satisfacer las dos condiciones anteriores. Para obtener más información referente a este hecho, se sugiere revisar los artículos de Bühlmann (2001), Bühlmann y Künsch (1995) y (1999) así como el artículo publicado por Franke *at al* (1998).

#### 4.2.2. Bootstrap por medio de Onduletas

La idea detrás del bootstrap basado en la transformada discreta de onduletas (DWT) es utilizar el hecho de que para procesos Gaussianos estacionarios fraccionalmente diferenciados (Stationary Gaussian fractionally differenced (FD) processes)  $\{X_t\}$ , la DWT es una transformación que *descorrelaciona* a  $\{X_t\}$ , es decir, mientras la serie de tiempo original puede presentar un alto grado de correlación, los coeficientes de la DWT pueden ser considerados como una serie de tiempo con elementos no correlacionados, vease Percival y Constantine (2002), Percival, Sardy y Davison (2000), y Percival y Walden (2000).

Para series de tiempo tales que la DWT elimina la correlación entre sus componentes, podemos realizar bootstrap usando el siguiente algoritmo:

**Algoritmo 3.** (a) Dada la serie de tiempo  $\mathbf{X}$  de longitud  $2^J$  se calcula la DWT correspondiente al nivel  $J_0 = J - 2$  para así obtener los coeficientes de onduletas, las entradas de los vectores  $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_{J_0}$  y el vector con los coeficientes de escala  $\mathbf{V}_{J_0}$ . Considerando a  $J_0 = J - 2$  tenemos que por lo visto en la sección (3.2) tanto el vector  $\mathbf{W}_{J_0}$  como el vector  $\mathbf{V}_{J_0}$  son de dimensión

$$N_j \times 1 = \frac{N}{2^{\tau_{J_0}}} \times 1 = \frac{2^J}{2^{(2^{J_0-1})}} \times 1 = \frac{2^J}{2^{J-2}} \times 1 = 2^2 \times 1 = 4 \times 1.$$

Si el valor considerado para  $J_0$  es menor, entonces, se tienen más coeficientes de onduletas en  $\mathbf{W}_{J_0}$  en donde realizar bootstrap, pero también se tendría la misma cantidad de coeficientes de escala en el respectivo vector  $\mathbf{V}_{J_0}$ .

(b) Realizar remuestreo con reemplazo  $N_j = N/2^{\tau_j}$  veces en los elementos de  $\mathbf{W}_j$  para así obtener la versión bootstrap  $\mathbf{W}_j^{(b)}$  del mismo para  $j = 1, 2, \dots, J_0$ ; hacer lo mismo para  $\mathbf{V}_{J_0}$ .



(c) Aplicar la transformada inversa de la DWT a las versiones bootstrap  $\mathbf{W}_1^{(b)}, \mathbf{W}_2^{(b)}, \dots, \mathbf{W}_{J_0}^{(b)}$  para obtener la versión bootstrap de la serie de tiempo  $\mathbf{X}^{(b)}$  de la cual se calculará la estadística de interés.

Repetir este proceso una y otra vez las veces que sea necesario, nos permite construir una distribución muestral de las estadísticas bootstrap, las cuales se utilizarán como un sustituto para la distribución de la estadística muestral.

El bootstrap basado en la DWT funciona adecuadamente para procesos estacionarios Gaussianos fraccionalmente diferenciados (ver Percival, Sardy y Davison, 2000), no siendo el caso para procesos de memoria corta, por ejemplo un proceso de promedios móviles de orden 1, ver Percival, Sardy y Davison (2000).

Una solución a este problema es utilizar la generalización de la transformada discreta de onduletas, es decir, los paquetes de onduletas (DWPT).

Como ya se vio en el capítulo anterior, ahora se tiene toda una colección de transformaciones ortonormales, las cuales están definidas en sub-intervalos de la forma  $\mathcal{I}_{j,n} = [\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}})$  al nivel  $j$ -ésimo de la transformación, todos ellos elementos de una partición diádica del intervalo  $[0, 1/2]$ .

Para seleccionar una transformación adecuada (una partición del intervalo  $[0, 1/2]$  adecuada) se puede utilizar el método propuesto por Contreras y Walden (2002). El cual consiste en realizar una prueba de hipótesis de que los elementos de  $W_{j,n}$  conforman parte de una muestra de ruido blanco, esto nos da un proceso iterativo que se ilustra en el siguiente esquema (vease secciones (3.3) y (3.4)):

**Algoritmo 4.** (a) Dada la serie de tiempo  $\mathbf{X}$  de longitud  $2^J$ , calcular los coeficientes de los paquetes de onduletas  $W_{1,n} = \mathbf{W}\mathbf{X}$  para  $n = 0, 1$  asociados a las bandas  $\mathcal{I}_{0,1}$  e  $\mathcal{I}_{1,1}$ .

(b) Se aplica una prueba de hipótesis

$H_0$  : Las componentes del vector  $W_{j,n}$  son una muestra de ruido blanco,

$H_1$  : Las componentes del vector  $W_{j,n}$  no son una muestra de ruido blanco.

Si no se rechaza  $H_0$ , entonces la banda de frecuencias  $\mathcal{I}_{j,n}$  es un elemento de la partición diádica  $\mathcal{P}$  del intervalo  $[0, 1/2]$  que definirá la transformación de paquete de onduletas a usarse

Si se rechaza  $H_0$ , entonces se descarta  $W_{j,n}$  y se transforma el mismo, mediante su respectiva transformación ortonormal, en los subvectores  $W_{j+1,2n}$  y  $W_{j+1,2n+1}$ , es decir, hay que subdividir la banda de frecuencias  $\mathcal{I}_{j,n}$  en  $\mathcal{I}_{j+1,2n} \cup \mathcal{I}_{j+1,2n+1}$

- (c) Repetir el paso (b) para cada uno de los vectores de coeficientes de onduletas que hayan tenido que calcularse en la iteración anterior, para cada  $j = 2, 3, \dots, J_0 - 1$  (donde  $J_0$  es el nivel máximo de la transformación).

Este algoritmo determina una partición  $\mathcal{P} = \{\mathcal{I}_{j_1, n_1}, \dots, \mathcal{I}_{j_m, n_m}\}$  del intervalo  $[0, 1/2)$ , y consecuentemente una transformación ortonormal asociada. Denotemos por  $W_{j_1, n_1}, W_{j_2, n_2}, \dots, W_{j_m, n_m}$  a los coeficientes resultantes de aplicar la transformación al proceso  $\mathbf{X}$ .

En virtud de que las componentes de cada  $W_{j_i, n_i}$  son una muestra de ruido blanco, el siguiente paso es realizar remuestreo con reemplazo en las componentes de este vector para así obtener una versión bootstrap  $W_{j_i, n_i}^{(b)}$  y repetir esto para cada  $i = 1, 2, \dots, m$ .

Al aplicar la inversa de la transformación asociada a  $\mathcal{P}$  al vector  $N \times 1$  ( $N = 2^J$ ) dimensional dado por

$$\mathbf{W}^{(b)} = \begin{bmatrix} W_{j_1, n_1}^{(b)} \\ \vdots \\ W_{j_m, n_m}^{(b)} \end{bmatrix}, \quad (4.4)$$

obtenemos una muestra bootstrap  $\mathbf{X}^{(b)}$  del proceso  $\mathbf{X}$ . Mediante este procedimiento podemos generar un número  $B$  de muestras bootstrap  $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(B)}$  del proceso original  $\mathbf{X}$ .

La prueba de hipótesis para ver si las componentes de  $W_{j_i, n_i}$  son ruido blanco puede llevarse a cabo usando el estadístico  $D$  definido por las ecuaciones (3.79) y (3.80), prueba de hipótesis de periodograma acumulativo, ver Percival y Walden (1994).

### 4.3. Ejemplos numéricos

Con el fin de comparar ambos métodos de remuestreo, primeramente planteamos un ambiente favorable para el método paramétrico al simular de un proceso autoregresivo y luego utilizar un modelo autoregresivo para ajustar los datos simulados, es de esperarse que el remuestreo paramétrico funcione mejor que el remuestreo no paramétrico.

Por otra parte, si el verdadero modelo no es autoregresivo pero se utiliza un modelo autoregresivo para ajustar los datos este *error* tendería a propiciar que el método no paramétrico sea mejor que el paramétrico. Se tendría así un segundo ambiente de comparación esta vez favorable al método no paramétrico.

En ambos casos, tanto para el remuestreo paramétrico como el remuestreo vía onduletas, en términos de la estimación de la varianza y la media del coeficiente de correlación parcial a lag 1,  $var(\rho_1)$  y  $\bar{\rho}_1$ , se procedió a implementar el algoritmo (2) y el algoritmo (4) numéricamente. Estos algoritmos se realizaron con dos tipos de procesos, primero en un autoregresivo de orden 1, con parámetro  $\phi = \frac{1}{2}$  y después en un proceso de promedios móviles de orden 1 con parámetro  $\theta = -\frac{1}{2}$ . Para ambos métodos se estudiaron los dos procesos arriba mencionados para tres distintos tamaños de muestra, a saber:  $N = 64$ ,  $128$  y  $N = 256$ .

Con el fin de estimar  $var(\rho_1)$  y  $\bar{\rho}_1$ , en el caso paramétrico se realizó bootstrap 100 veces en los residuales para después con estos generar  $b = 1, 2, \dots, 100$  muestras tipo bootstrap de cada uno de los procesos. Para el caso no paramétrico se generaron 100 muestras tipo bootstrap en los componentes del vector  $W_{j_i, n_i}$  para cada  $i = 1, 2, \dots, m$ , a continuación se aplicó la inversa de la transformación asociada a  $\mathcal{P}$ , al vector  $\mathbf{W}^{(b)}$ ,  $b = 1, 2, \dots, 100$ , obteniendo así 100 muestras bootstrap  $\mathbf{X}^{(b)}$ ,  $b = 1, 2, \dots, 100$ , del proceso  $\mathbf{X}$ .

Finalmente, ya con las 100 muestras bootstrap del proceso  $\{X_t\}$ , en cada una de ellas se calculó el coeficiente de correlación parcial a lag 1, con estos 100 coeficientes se calculó tanto la desviación estándar como el promedio de los mismos obteniendo así un estimador para  $var(\rho_1)$  y un estimador para  $\bar{\rho}_1$ .

Todo esto se repitió 50 veces para así al final tener 50 estimadores para  $var(\rho_1)$  y 50 estimadores para  $\bar{\rho}_1$ . Para realizar comparaciones entre ambos métodos con respecto al valor verdadero de los estimadores se calcularon tanto el valor asintótico, vía simulación, como el valor teórico de  $var(\rho_1)$  y  $\bar{\rho}_1$  utilizando las ecuaciones (4.7) y (4.8). Para obtener el valor asintótico se generaron 10,000 procesos de ambos tipos, a cada uno de ellos se les calculó su autocorrelación muestral  $\hat{\rho}_1$ , es decir, se obtuvieron  $\hat{\rho}_1^{(1)}, \hat{\rho}_1^{(2)}, \dots, \hat{\rho}_1^{(10000)}$ . De este conjunto de valores se obtuvo tanto el promedio como la desviación estándar, a éstos se les denominó *promedio asintótico estimado* y *desviación estándar asintótica estimada* respectivamente.

De acuerdo con el artículo de Percival, Sardy y Davison (2000), el método de remuestreo vía onduletas genera mejores estimadores para  $var(\rho_1)$ , no siendo así el caso cuando se desea generar buenos estimadores para  $\bar{\rho}_1$ .

Dado que nosotros aplicamos ambos métodos a dos procesos distintos, es decir, tanto a un proceso autoregresivo como a un proceso de promedios móviles, en el caso del proceso autoregresivo dado que el verdadero modelo es el que se usa para ajustar los datos, el remuestreo paramétrico deberá de generar mejores estimadores para la desviación estándar que el método de remuestreo vía onduletas; por esta misma razón, para el caso de estimadores de la media del coeficiente de correlación parcial a lag 1 también esperamos que el método paramétrico genere mejores estimadores que el método de remuestreo vía onduletas.

Para el caso del proceso de promedios móviles al cuál se le está ajustando un modelo autoregresivo, dado que estamos equivocando el modelo correcto, se espera que el método de remuestreo vía onduletas genere mejores estimadores, tanto para  $var(\rho_1)$  como para  $\bar{\rho}_1$ , que el método paramétrico, puesto que en el remuestreo vía onduletas no juegan un papel determinante los parámetros del modelo en cuestión.

### 4.3.1. Caso Paramétrico

El proceso que a continuación se describe se realizó considerando el modelo autoregresivo de orden 1 con parámetro  $\phi = \frac{1}{2}$ , es decir, se tiene el proceso  $AR(1)$ :

$$X_t = \frac{1}{2}X_{t-1} + z_t, \quad (4.5)$$

donde  $z_t$  se distribuye  $N(0, 1)$ .

Se generaron variables aleatorias  $\{z_t\}$  independientes idénticamente distribuidas  $N(0, 1)$ , para mediante la ecuación (4.5), generar un proceso autoregresivo de orden 1. Para obtener la estacionariedad de segundo orden se desecharon los primeros 500 elementos del proceso  $\{X_t\}_t$ .

Una vez generado el modelo autoregresivo de orden 1 con parámetro  $\phi = \frac{1}{2}$ , se asumirá que se desconoce tanto el orden del proceso como el parámetro  $\phi$  del modelo del que proviene, para que a continuación se le ajuste un modelo de tipo  $AR$  de orden  $p$ , éste orden se seleccionará de acuerdo a aquel que minimice el valor del coeficiente de AICC.

Una vez realizado el ajuste se obtendrán tanto los residuales como el orden del modelo seleccionado. Después se realizó bootstrap en los residuales 100 veces, es decir, se generaron 100 muestras tipo bootstrap de residuales, siguiendo los pasos del bootstrap paramétrico descrito en la sección (4.1.1).

Es así que se cuenta con la siguiente información:

- El orden  $p$  del modelo ajustado.
- 100 series tipo bootstrap de los residuales.
- Los valores de los parámetros del modelo de tipo  $AR(p)$ , los cuales se guardan en un vector  $\hat{\phi}_t$ .

Con esta información se procedió a generar 100 nuevos procesos  $\{\hat{X}_t^{(i)}\}$ , con  $i = 1, 2, \dots, 100$ ,  $t = 1, 2, \dots, 64$ .

$$\begin{array}{cccc} \hat{X}_1^{(1)} & \hat{X}_2^{(1)} & \dots & \hat{X}_{64}^{(1)} \\ \hat{X}_1^{(2)} & \hat{X}_2^{(2)} & \dots & \hat{X}_{64}^{(2)} \\ \vdots & \vdots & & \vdots \\ \hat{X}_1^{(100)} & \hat{X}_2^{(100)} & \dots & \hat{X}_{64}^{(100)} \end{array}$$

Después, a cada serie se le resta su media para así convertirlos en procesos con media cero. A continuación se procedió a calcular el coeficiente de correlación parcial a lag 1 para cada una de las 100 muestras tipo bootstrap por medio de la ecuación

$$\hat{\rho}_1 = \frac{\sum_{t=1}^{N-2} X_t X_{t+1}}{\sum_{t=1}^{N-1} X_t^2}.$$

Por tanto, se tienen  $\hat{\rho}_1^{(1)}, \hat{\rho}_1^{(2)}, \dots, \hat{\rho}_1^{(100)}$  coeficientes de correlación parcial, de este conjunto de 100 valores se calculó el promedio y la desviación estándar, es decir, se obtuvieron  $\bar{\rho}_1^{(1)}$  y  $\hat{\sigma}_{\rho_1}^{(1)}$ , donde el superíndice indica que es el resultante de la primera realización de todo el proceso arriba descrito.

Por tanto, al final del procedimiento se tiene los siguientes datos:

- El orden del modelo ajustado al proceso, es decir,  $p_1$

- El promedio de los coeficientes de correlación parcial al lag 1 de las 100 muestras obtenidas al hacer bootstrap sobre los residuales,  $\bar{\rho}_1^{(1)}$ .
- La Desviación estándar de los coeficientes de correlación parcial al lag 1, de las 100 muestras obtenidas al hacer bootstrap sobre los residuales  $\hat{\sigma}_{\rho_1}^{(1)}$ .

Este mismo proceso se repitió 50 veces a partir de la ecuación (4.5), es decir, para cada  $k = 1, 2, \dots, 50$ , generamos un proceso  $AR(1)$ ,  $\{X_t\}$ , como en (4.5) y mediante bootstrap obtenemos  $\bar{\rho}_1^{(k)}$  y  $\hat{\sigma}_{\rho_1}^{(k)}$ , con lo cual obtenemos

- $\{p_1, p_2, \dots, p_{50}\}$ ,
- $\{\bar{\rho}_1^{(1)}, \bar{\rho}_1^{(2)}, \dots, \bar{\rho}_1^{(50)}\}$  y
- $\{\hat{\sigma}_{\rho_1}^{(1)}, \hat{\sigma}_{\rho_1}^{(2)}, \dots, \hat{\sigma}_{\rho_1}^{(50)}\}$ .

Para el caso de un proceso de tipo promedios móviles de orden 1 con parámetro  $\theta = -\frac{1}{2}$ , es decir el proceso  $MA(1)$  con expresión

$$X_t = -\frac{1}{2}\xi_{t-1} + \xi_t, \quad (4.6)$$

donde  $\xi_t$  se distribuye  $N(0, 1)$ , se realizó lo mismo excepto por las siguientes variaciones:

Al igual que antes se generaron variables aleatorias  $\{\xi_t\}$  independientes idénticamente distribuidas  $N(0, 1)$  para que utilizando la ecuación (4.6) generar un proceso de promedios móviles; con el fin de estudiar el método de bootstrap en sus enfoques paramétrico y no paramétrico supóngase que nos equivocamos al momento de determinar un modelo para los datos y se supone que el modelo apropiado es un proceso autoregresivo de orden  $p$ , este orden se eligió de modo tal que se minimizara el coeficiente de AICC. En la práctica, si el tamaño de la muestra es pequeño (por ejemplo  $N = 64$ ), la anterior equivocación puede ocurrir.

Una vez ajustado el modelo *apropiado* los siguientes pasos fueron los mismos que para el caso en que el modelo procedía de proceso de tipo autoregresivo.

### 4.3.2. Caso No Paramétrico, (Bootstrap vía Onduletas)

Al igual que para el caso paramétrico se estudiaron dos tipos de procesos, primero un proceso de tipo autoregresivo de orden 1 con parámetro  $\phi = \frac{1}{2}$  y después un proceso de promedios móviles de orden 1 con parámetro  $\theta = -\frac{1}{2}$ .

Primero se generó un proceso  $AR(1)$  utilizando la ecuación (4.5), para a continuación aplicar el algoritmo (4) dado en la sección (4.2.2) y así determinar una base ortogonal ó equivalentemente, una transformación  $\mathcal{P}$  en paquetes de onduletas. Esta transformación es tal que los vectores de coeficientes resultantes  $W_{j_1, n_1}, W_{j_2, n_2}, \dots, W_{j_m, n_m}$  son muestras de ruido blanco.

En seguida, para cada  $(j_i, n_i)$ , con  $i = 1, 2, \dots, m$ , se realizó muestreo independiente con reemplazo sobre las componentes de  $W_{j_i, n_i}$  para construir muestras bootstrap  $W_{j_i, n_i}^{(1)}, W_{j_i, n_i}^{(2)}, \dots, W_{j_i, n_i}^{(100)}$ .

Entonces para cada  $b = 1, 2, \dots, 100$ , se construyo una muestra bootstrap  $\mathbf{X}^{(b)}$  de  $\mathbf{X}$  como resultado de aplicar la inversa de la transformación  $\mathcal{P}$  al vector dado en la ecuación (4.4). El resultado del proceso anterior son las siguientes muestras bootstrap

$$\begin{array}{cccc} \hat{X}_1^{(1)} & \hat{X}_2^{(1)} & \dots & \hat{X}_{64}^{(1)} \\ \hat{X}_1^{(2)} & \hat{X}_2^{(2)} & \dots & \hat{X}_{64}^{(2)} \\ \vdots & \vdots & & \vdots \\ \hat{X}_1^{(100)} & \hat{X}_2^{(100)} & \dots & \hat{X}_{64}^{(100)} \end{array}$$

A cada una de estas series se les calculó el coeficiente de autocorrelación parcial a lag 1, por lo tanto tenemos  $\{\rho_1^{w,(1)}, \rho_1^{w,(2)}, \dots, \rho_1^{w,(100)}\}$ . Después se calculó el promedio y la desviación estándar de los mismos, es decir,  $\bar{\rho}_1^{w,(1)}$  y  $\hat{\sigma}_{\rho_1}^{w,(1)}$ . Todo este proceso se repitió 50 veces, es decir, para cada  $k = 1, 2, \dots, 50$  generamos un proceso  $AR(1)$ ,  $\{X_t\}$ , como en la ecuación (4.5) y usando bootstrap vía onduletas calculamos la media y la desviación estándar de los 100 coeficientes de autocorrelación a lag 1 los cuales se denotarán por  $\bar{\rho}_1^{w,(k)}$  y  $\hat{\sigma}_{\rho_1}^{w,(k)}$ , para así al final del proceso obtener:

$$\begin{array}{l} \blacksquare \left\{ \overline{\rho_1^{w,(1)}}, \overline{\rho_1^{w,(2)}}, \dots, \overline{\rho_1^{w,(50)}} \right\}. \\ \blacksquare \left\{ \hat{\sigma}_{\rho_1}^{w,(1)}, \hat{\sigma}_{\rho_1}^{w,(2)}, \dots, \hat{\sigma}_{\rho_1}^{w,(50)} \right\}. \end{array}$$

En forma análoga para el proceso tipo  $MA(1)$  dado en la ecuación (4.6).

## 4.4. Resultados

Para poder realizar una comparación entre el método paramétrico y el método no paramétrico se procedió a calcular tanto para el proceso autoregresivo como para el proceso de promedios móviles la distribución asintótica del coeficiente de correlación parcial a lag 1, es decir, se calcularon los valores asintóticos de la desviación estándar y la media del coeficiente de correlación 1.

Para un proceso autoregresivo de orden 1, el coeficiente de correlación a lag 1  $\hat{\rho}_1$  es asintóticamente normal con media  $\phi$  y varianza  $\sigma^2 = (1 - \phi^2) n^{-1}$ , donde  $n$  es la longitud del proceso, es decir,

$$\hat{\rho}_1 \sim AN \left( \phi, \frac{1 - \phi^2}{n} \right). \quad (4.7)$$

En cambio para un proceso de promedios móviles de orden 1, el coeficiente de correlación a lag 1 se distribuye asintóticamente normal con media  $\rho_1$  y varianza  $\sigma^2 = (1 + 2\rho_1^2) n^{-1}$ , donde  $n$  es la longitud del proceso, es decir,

$$\hat{\rho}_1 \sim AN \left( \frac{\theta}{1 + \theta^2}, \frac{1 + 2\rho_1^2}{n} \right). \quad (4.8)$$

De igual manera para ambos procesos se estimaron tanto la media y la desviación estándar asintótica del coeficiente de correlación parcial a lag 1. Para hacer esto se generaron 10,000 procesos de ambos tipos, a cada uno de ellos se les calculó la autocorrelación muestral  $\hat{\rho}_1$ , es decir, se obtuvieron,  $\hat{\rho}_1^{(1)}, \hat{\rho}_1^{(2)}, \dots, \hat{\rho}_1^{(10000)}$ . De este conjunto de valores se obtuvo tanto el promedio como la desviación estándar, a éstos se les denominó *promedio asintótico estimado* y *desviación estándar asintótica estimada* respectivamente.

Los resultados obtenidos por ambos métodos para los tres tamaños distintos de muestra tanto para el caso en que el proceso original era un  $AR(1)$  como el caso en que el proceso original era un  $MA(1)$  se encuentran resumidos en los cuadros (4.1), (4.2), (4.3), (4.4), (4.5) y (4.6). Dado que se cuentan con 50 estimadores para la media y la desviación estándar del coeficiente de correlación se calcularon las tres medidas de tendencia central: media, mediana y moda.

Para el caso en que el proceso original es autoregresivo de orden 1,  $AR(1)$ , estas medidas de tendencia central se calcularon para los resultados obtenidos vía bootstrap paramétrico,



(columna *Bootstp*) y vía bootstrap no paramétrico (columna *Wavest*), además se está mostrando el valor asintótico de la media (Asym) y los valores asintóticos para la desviación estándar, la teórica (Asym Est), obtenida de acuerdo a la ecuación (4.7), y la desviación estándar asintótica estimada vía simulación, (Asym).

Mientras que para el caso en que el proceso original es de promedios móviles de orden 1,  $MA(1)$ , estas medidas de tendencia central se calcularon para los resultados obtenidos vía bootstrap paramétrico, (columna *Bootstp*) y vía bootstrap no paramétrico (columna *Wavest*), además se está mostrando el valor asintótico de la media (Asym) y los valores asintóticos para la desviación estándar teórica, (Asym), obtenida de acuerdo a la ecuación (4.7).

#### 4.4.1. Caso $N = 64$

Primero presentaremos los resultados en el que el proceso original es autoregresivo para el caso paramétrico, luego los resultados obtenidos en el caso no paramétrico; haremos las comparaciones de resultados entre ambos métodos y finalmente haremos lo mismo para el caso en que el modelo original es de promedios móviles.

En la figura (4.1) se presentan los ordenes obtenidos al ajustar a las series de datos un modelo de tipo  $AR(p)$ , el valor de  $p$  se eligió de modo tal que se minimizara el valor del coeficiente de AICC; el porcentaje de veces que se eligió un modelo  $AR(1)$  fue de 84 %.

En la figura (4.2) se muestran tanto los promedios como las desviaciones estándar de los coeficientes de correlación parcial a lag 1, junto con sus respectivos valores asintóticos tanto de la media como de la desviación estándar respectivamente, obtenidos vía la ecuación (4.7) y vía simulación.

Para el caso de los promedios se puede observar que estos valores están por debajo del valor asintótico estimado vía simulación. Para el caso de la desviación estándar, se está mostrando tanto la desviación estándar asintótica teórica, representada por la línea punteada, y la desviación estándar asintótica estimada vía simulación, la línea continua ubicada por encima de la línea punteada, es decir, de la desviación estándar teórica.

Como se podrá observar los valores de las desviaciones estándar de los coeficientes de correlación parcial a lag 1 se concentran alrededor del valor de la desviación estándar asintótica estimada, mientras que la desviación estándar asintótica teórica está por debajo de los valores obtenidos vía el bootstrap paramétrico.

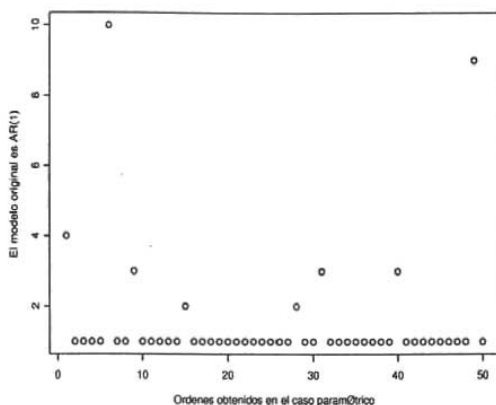


Figura 4.1: Gráfica de los ordenes de los modelos tipo Autoregresivo ajustados a un proceso proveniente de un modelo autoregresivo, método paramétrico, ( $N=64$ ).

En este caso se podrá observar que se obtienen valores asintóticos distintos, esto es consecuencia del tamaño de la muestra, el cuál fue de 64, sin embargo más adelante se podrá observar que para tamaños de muestra mayores estos valores serán muy similares.

Para el caso no paramétrico, en la figura (4.3) se muestran la gráficas de los resultados obtenidos para la media contrastados con el valor asintótico

Ahora en la figura (4.4) se muestra la gráfica de los resultados obtenidos para la desviación estándar contrastados con los valores asintóticos

A diferencia del bootstrap paramétrico se puede ver que los valores de los promedios son más cercanos al valor asintótico, sin embargo este sigue siendo subestimando por los promedios, una justificación a este hecho se da en el *capítulo 5*.

Para el caso de la desviación estándar, se observa que ambas estimaciones asintóticas están por debajo de las desviaciones estándar de los coeficientes de correlación parcial a lag 1, vía bootstrap no paramétrico a diferencia del paramétrico.

Los resultados obtenidos por ambos métodos se encuentran resumidos en el cuadro (4.1). Dado que se cuentan con 50 estimadores para la media y la desviación estándar del coeficiente de correlación se procedió a calcular las medidas de tendencia central (media, mediana y moda).

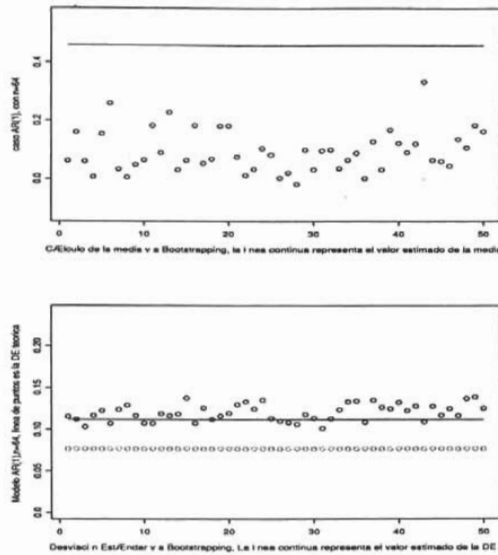


Figura 4.2: Gráfica de los promedios y desviaciones estándar de los coeficientes de correlación parcial al lag 1, método paramétrico, ( $N=64$ ).

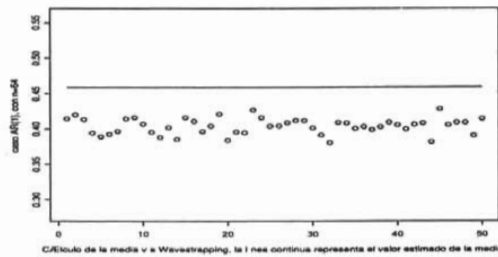


Figura 4.3: Gráfica de los promedios de los coeficientes de correlación parcial al lag 1, método no paramétrico, ( $N=64$ )

Como se podrá observar para el caso de la media es notablemente mejor el método del bootstrap

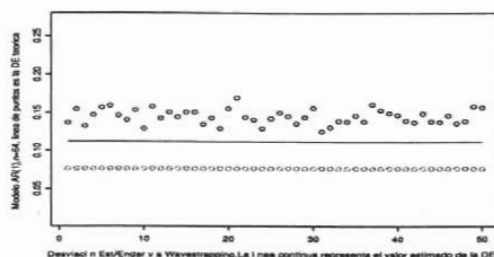


Figura 4.4: Gráfica de las desviaciones estándar de los coeficientes de correlación parcial al lag 1, método no paramétrico, (N=64)

Cuadro 4.1: Caso  $n=64$ , para un proceso Autoregresivo de orden 1, el porcentaje de veces que se eligió el modelo  $AR(1)$  fue del 84 %.

	Ordenes	Media			Desviación Estándar			
		Bootst	Wavest	Asym	Bootst	Asym Est	Wavest	Asym
Moda	1	-	-	0.458868	-	0.111659	-	0.076541
Media	-	0.094439	0.403696	-	0.119920	-	0.144612	-
Mediana	1	0.079721	0.404543	0.458868	0.118615	0.111659	0.144014	0.076541
Std Dev	-	0.072341	0.011195	-	0.010015	-	0.009667	-

no paramétrico, la diferencia tanto en la media como en la mediana con respecto al valor obtenido vía wavestrapping es mucho menor que con respecto a los obtenidos vía bootstrap paramétrico, en el capítulo 5 trataremos de explicar este hecho.

Para el caso de la desviación estándar cambian notablemente las cosas, el bootstrap genera mejores estimadores que el wavestrapping, que era lo que se esperaba dado que se conocen los parámetros del modelo del cual provienen los datos, cuyos valores son significativamente mayores que los valores asintóticos estimados.

Ahora veamos los resultados en el caso en que el proceso original es de promedios móviles y se le ajusto un modelo autoregresivo, es decir, el caso en que nos equivocamos al seleccionar el modelo.

En la figura (4.5) se muestran los ordenes elegidos en el caso paramétrico cuando el modelo original es un  $MA(1)$ . Para este caso el porcentaje de veces que se eligió un modelo  $AR(1)$  fue

del 28%, mientras que el modelo que se eligió con mayor frecuencia fue el modelo  $AR(2)$ , en un 60% de las veces.

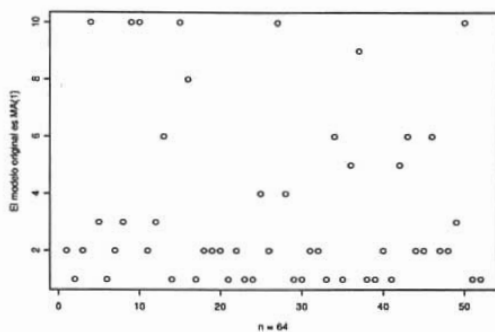


Figura 4.5: Gráfica de los ordenes de los modelos tipo Autoregresivo ajustados a un proceso proveniente de un modelo de promedios móviles, método paramétrico, ( $N=64$ ).

En la gráfica (4.6) se muestran tanto los promedios como las desviaciones estándar de los coeficientes de correlación parcial a lag 1, junto con sus respectivos valores asintóticos tanto de la media como de la desviación estándar respectivamente, obtenidos vía la ecuación (4.8) y vía simulación.

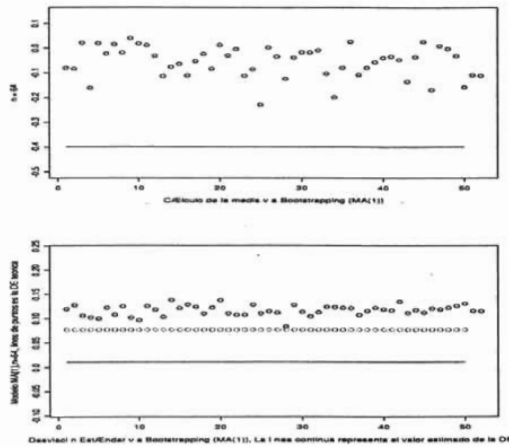


Figura 4.6: Gráfica de los promedios y las desviaciones estándar de los coeficientes de correlación parcial al lag 1 para procesos de tipo  $MA(1)$ , método paramétrico, ( $N=64$ ).

Para el caso de los promedios se puede observar que estos valores están por encima del valor asintótico. Para el caso de la desviación estándar, se está mostrando la desviación estándar asintótica teórica, que nuevamente está representado por la línea punteada, y la desviación estándar asintótica estimada vía simulación, que está representada por la línea continua ubicada por debajo de la línea punteada, es decir, de la desviación estándar teórica.

Aquí nuevamente se puede observar que los valores asintóticos son distintos, como ya se mencionó antes esto es consecuencia del tamaño de la muestra, más adelante se verá que para los casos  $n = 128$  y  $n = 256$  estos valores coincidirán.

Para el caso no paramétrico en la gráfica (4.7) se presentan los resultados obtenidos contrastados con los valores asintóticos ya mencionados con anterioridad:

A diferencia del bootstrap paramétrico, para este caso se puede ver que los valores de los promedios son más lejanos al valor asintótico, es decir sobreestiman el valor verdadero de la media.

Para el caso de la desviación estándar, se observa que ambas estimaciones asintóticas están por debajo de estimaciones de las desviaciones estándar de los coeficientes de correlación parcial al lag 1, vía bootstrap no paramétrico.

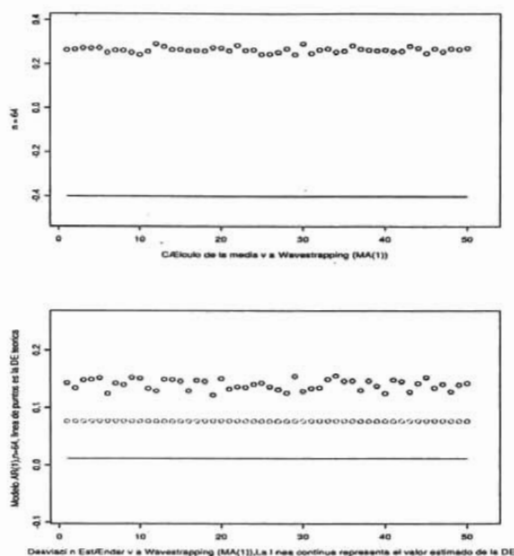


Figura 4.7: Gráfica de los promedios y las desviaciones estándar junto con los valores asíntóticos estimados tanto para la media, como para la desviación estándar para el caso en que el proceso original es de tipo  $MA(1)$ , método no paramétrico, ( $N=64$ ).

Los resultados obtenidos en ambos métodos se encuentran resumidos en el cuadro (4.2).

Cuadro 4.2: Caso  $n=64$ , en el que el proceso original es un  $MA(1)$ , el porcentaje de veces que se eligió el modelo  $AR(1)$  fue del 28%.

	Orders	Media			Desviación Estándar		
		Bootst	Wavest	Asym	Bootst	Wavest	Asym
Moda	2	-	-	-0.399624	-	-	0.143614
Media	-	-0.054178	0.265150	-	0.116434	0.140050	-
Mediana	2	-0.037418	0.264511	-0.399624	0.118002	0.141274	0.143614
Std Dev	-	0.062646	0.011290	-	0.010772	0.009064	-

Las medidas de tendencia central que se muestran en la primer columna de la tabla se calcularon para los resultados obtenidos vía bootstrap paramétrico, (columna *Bootstp*) y vía bootstrap no paramétrico (columna *Wavest*), además se está mostrando el valor asintótico de la media (*Asym*) y el valor asintótico para la desviación estándar, en este caso desviación estándar asintótica teórica, (*Asym*), obtenida de acuerdo a la ecuación (4.7).

De la gráfica (4.7) y de la tabla (4.2) podemos concluir que para este tamaño de muestra en el caso de que el proceso del cual provienen los datos es de tipo  $MA(1)$  el método paramétrico genera mejores estimadores para la media asintótica que el método no paramétrico; no siendo así para generar estimadores de la desviación estándar asintótica, donde el método no paramétrico genera estimadores mucho más cercanos que el método paramétrico, además de que la desviación estándar de los estimadores vía método no paramétrico es menor que la que se obtiene por medio del método paramétrico.

Estos resultados obtenidos eran de esperarse, dado que estamos equivocando el modelo el cuál es parte fundamental en el bootstrap paramétrico, al contrario del bootstrap no paramétrico, donde no importa el modelo, puesto que el bootstrap se realiza en los componentes de vector  $W_{j_i, m_i}$  para cada  $i = 1, 2, \dots, m$ . Recordemos que este estudio se realiza con la finalidad de comparar el método paramétrico con el método no paramétrico.

#### 4.4.2. Caso $N = 128$

A continuación presentamos los resultados obtenidos por ambos métodos para procesos de longitud  $N = 128$ . Primero presentaremos los obtenidos cuando el proceso original es un  $AR(1)$  para el método paramétrico y en seguida los obtenidos por el método no paramétrico. Finalmente haremos lo mismo cuando el proceso original es un  $MA(1)$ .

En la figura (4.8) se presentan los ordenes obtenidos al ajustar a las series de datos un modelo de tipo  $AR(p)$ , como ya se mencionó anteriormente, el valor de  $p$  se eligió de modo tal que se minimizara el valor del coeficiente de AICC; en particular, el porcentaje de veces que se eligió un modelo  $AR(1)$  fue del 76%.

En la figura (4.9) se muestran tanto los promedios como las desviaciones estándar de los coeficientes de correlación parcial a lag 1, junto con sus respectivos valores asintóticos tanto de la media como de la desviación estándar respectivamente, obtenidos vía la ecuación (4.7) y vía simulación.



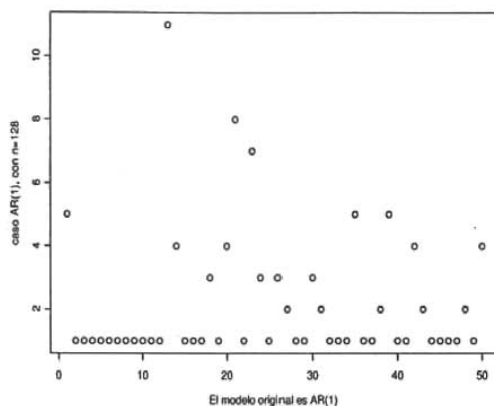


Figura 4.8: Gráfica de los ordenes de los modelos tipo Autoregresivo ajustados a un proceso proveniente de un modelo autoregresivo, método paramétrico, ( $N=128$ ).

Para el caso de los promedios se puede observar que estos valores están muy por debajo del valor asintótico estimado vía simulación. Nuevamente, para una justificación de lo anterior ver el *Capítulo 5*.

Para el caso de la desviación estándar, se está mostrando tanto la desviación estándar asintótica teórica, en este caso está representado por la línea punteada, y la desviación estándar asintótica estimada vía simulación, que es la línea continua ligeramente por encima de la línea punteada, es decir, de la desviación estándar teórica, además como se podrá observar los valores de las desviaciones estándar de los coeficientes de correlación parcial a lag 1 se concentran alrededor del valor de las desviación estándar asintóticas estimadas.

Como se mencionó en la sección anterior ambos valores asintóticos de la desviación estándar ya no son tan distintos como ocurrió en el caso  $N = 64$ .

Para el caso no paramétrico, en la figura (4.10) se muestran las gráficas de los resultados obtenidos tanto para la media como para la desviación estándar contrastados con sus respectivos valores asintóticos ya descritos anteriormente.

A diferencia del bootstrap paramétrico se puede observar que los valores de los promedios son más cercanos al valor asintótico en el caso no paramétrico, sin embargo se sigue sobreestimando

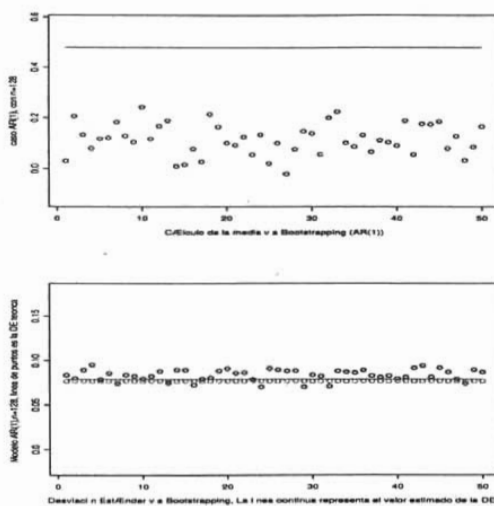


Figura 4.9: Gráfica de los promedios y desviaciones estándar de los coeficientes de correlación parcial al lag 1, método paramétrico, ( $N=128$ ).

el valor verdadero de la media, ver comentarios en el capítulo 5.

Para el caso de la desviación estándar, se observa que ambas estimaciones asintóticas están por debajo de los estimadores de las desviaciones estándar de los coeficientes de correlación parcial a lag 1, vía el método no paramétrico a diferencia del bootstrap paramétrico.

Los resultados obtenidos por ambos métodos se encuentran resumidos en el cuadro (4.3).

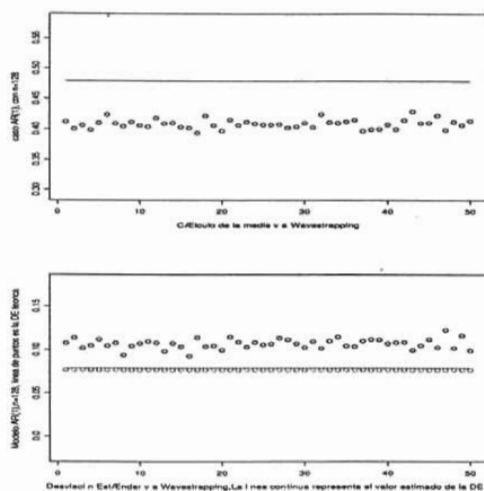


Figura 4.10: Gráfica de los promedios y las desviaciones estándar junto con los valores asintóticos estimados tanto para la media, como para la desviación estándar para el caso en que el proceso original es de tipo  $AR(1)$ , método no paramétrico, ( $N=128$ ).

Cuadro 4.3: Caso  $n=128$ , para un proceso Autoregresivo de orden 1, el porcentaje de veces que se eligió el modelo  $AR(1)$  fue del 76 %

	Orders	Media			Desviación Estándar			
		Bootst	Wavest	Asym	Bootst	Asym Est	Wavest	Asym
Moda	1	-	-	0.479512	-	0.078765	-	0.076547
Media	-	0.127736	0.408513	-	0.085627	-	0.106942	-
Mediana	1	0.102430	0.408420	0.479512	0.084214	0.078765	0.107401	0.076547
Std Dev	-	0.070304	0.007619	-	0.005776	-	0.005670	-

Como se podrá observar, para el caso de la media es notablemente mejor el método del bootstrap no paramétrico, la diferencia tanto en la media como en la mediana con respecto al valor obtenido vía wavestrapping es mucho menor que con respecto a los obtenidos vía bootstrap paramétrico; además de que la desviación estándar del estimador vía bootstrap no paramétrico

es mucho menor que la obtenida por el método paramétrico.

Para el caso de la desviación estándar cambian notablemente las cosas, el bootstrap paramétrico genera mejores estimadores que el método no paramétrico, que era lo que se esperaba dado que se conocen los parámetros del modelo del cual provienen los datos, los valores obtenidos vía el bootstrap no paramétrico son significativamente mayores que los valores asintóticos estimados aunque tienen menor variación entre ellos.

Ahora veamos los resultados en el caso en que el proceso original es de promedios móviles y se le ajusta un modelo autoregresivo, es decir, el caso en que nos equivocamos al seleccionar el modelo.

En la figura (4.11) se presentan los ordenes de los modelos de tipo autoregresivo elegidos, el porcentaje de veces que se eligió un modelo  $AR(1)$  fue de 8%; el modelo que se eligió con mayor frecuencia fue el modelo  $AR(3)$ , en un 74% de las veces.

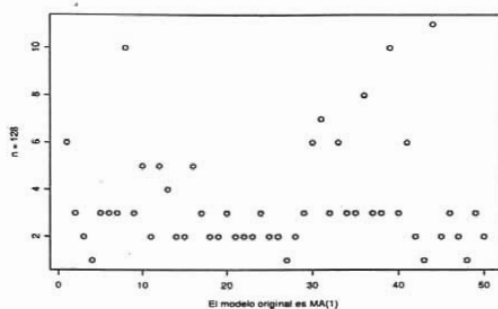


Figura 4.11: Gráfica de los órdenes de los modelos tipo Autoregresivo ajustados a un proceso proveniente de un modelo de promedios móviles, método paramétrico, ( $N=128$ ).

En la gráfica (4.12) se muestran tanto los promedios como las desviaciones estándar de los coeficientes de correlación parcial a lag 1, junto con sus respectivos valores asintóticos tanto de la media como de la desviación estándar respectivamente, obtenidos vía la ecuación (4.8) y vía simulación.

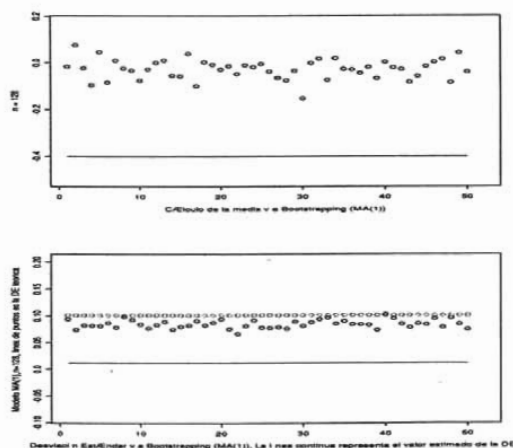


Figura 4.12: Gráfica de los promedios y las desviaciones estándar de los coeficientes de correlación parcial al lag 1 para procesos de tipo  $MA(1)$ , método paramétrico, ( $N=128$ ).

Para el caso de los promedios se puede observar que estos valores están por encima del valor asintótico. Para el caso de la desviación estándar, se está mostrando la desviación estándar asintótica teórica, que nuevamente está representado por la línea punteada, y la desviación estándar asintótica estimada vía simulación, que está representada por la línea continua ubicada por debajo de los valores de las desviaciones estándar estimadas obtenidas vía bootstrap no paramétrico.

Si consideramos como estimador del valor verdadero de la desviación estándar a la desviación estándar asintótica teórica, se puede observar que el método paramétrico genera estimadores ligeramente superiores al valor de desviación estándar asintótico en cuestión.

Para el caso no paramétrico en la gráfica (4.13) se presentan los resultados obtenidos contrastados con los valores asintóticos ya mencionados con anterioridad:

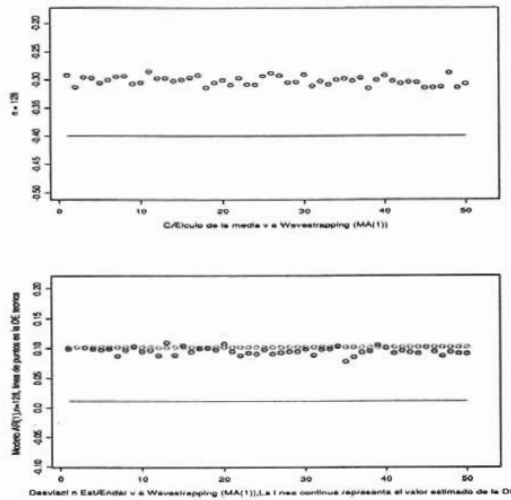


Figura 4.13: Gráfica de los promedios y las desviaciones estándar junto con los valores asintóticos estimados tanto para la media, como para la desviación estándar para el caso en que el proceso original es de tipo  $MA(1)$ , método no paramétrico, ( $N=128$ ).

En el caso de los promedios a pesar de que el valor asintótico es menor que los promedios calculados, intuitivamente la distancia que hay entre los mismos es menor que para el caso del bootstrap paramétrico. Lo mismo ocurre en el caso de las desviaciones estándar, donde se

puede observar que la mayoría de estos estimadores se agrupan alrededor del valor asintótico de la desviación estándar.

Los resultados obtenidos en ambos métodos se encuentran resumidos en el cuadro (4.4).

Cuadro 4.4: Caso  $n=128$ , para un proceso de tipo promedios móviles de orden 1, el porcentaje de veces que se eligió el modelo  $AR(1)$  fue de 8%.

	Orders	Media			Desviación Estándar		
		Bootst	Wavest	Asym	Bootst	Wavest	Asym
Moda	3	-	-	-0.399900	-	-	0.101551
Media	-	-0.028246	-0.301774	-	0.084798	0.095155	-
Mediana	3	-0.025988	-0.302158	-0.399900	0.084200	0.094629	0.101551
Std Dev	-	0.042564	0.007824	-	0.007636	0.006152	-

Las medidas de tendencia central que se muestran en la primer columna, al igual que la desviación estándar, se calcularon para los resultados obtenidos vía bootstrap paramétrico, (columna *Bootstp*) y vía bootstrap no paramétrico (columna *Wavest*), además se está mostrando el valor asintótico de la media (Asym) y el valor asintótico para la desviación estándar, en este caso desviación estándar asintótica teórica, (Asym), obtenida de acuerdo a la ecuación (4.7).

Se podrá observar que para obtener estimadores de la media es notablemente mejor el método del bootstrap no paramétrico que el bootstrap paramétrico. Ocurre lo mismo para estimar la desviación estándar, donde el bootstrap no paramétrico proporciona mejores estimadores que el bootstrap paramétrico.

Estos resultados obtenidos eran de esperarse, dado que estamos equivocando el modelo el cuál es parte fundamental en el bootstrap paramétrico, al contrario del bootstrap no paramétrico, donde no importa el modelo, puesto que el bootstrap se realiza en los componentes de vector  $W_{j_i, n_i}$  para cada  $i = 1, 2, \dots, m$ . Nuevamente recordemos que este estudio se realiza con la finalidad de comparar el método paramétrico con el método no paramétrico.

En el artículo publicado por Percival, Sardy y Davison, [16], se menciona que el bootstrap no paramétrico es adecuado para estimar la desviación estándar de  $\rho_1$  en comparación con

el bootstrap por bloques, y nuevamente es adecuado cuando se compara con el bootstrap paramétrico en los residuales.

Además, en el mismo artículo mencionan que el bootstrap no paramétrico no es apropiado para estimar la media de  $\rho_1$ , sin embargo, en este trabajo se obtuvo que también para la media parece ser un método apropiado tanto cuando no se especifica bien el modelo como cuando sí es está especificando de manera adecuada el modelo.

#### 4.4.3. Caso $n = 256$

A continuación presentamos los resultados obtenidos por ambos métodos para procesos de longitud  $N = 256$ . Primero presentaremos los obtenidos cuando el proceso original es un  $AR(1)$  para el método paramétrico y en seguida los obtenidos por el método no paramétrico. Finalmente haremos lo mismo cuando el proceso original es un  $MA(1)$ .

En la figura (4.14) se presentan los ordenes obtenidos al ajustar a las series de datos un modelo de tipo  $AR(p)$ , como se ha venido mencionando, el valor de  $p$  se eligió de modo tal que se minimizara el valor del coeficiente de AICC; en particular, el porcentaje de veces que se eligió un modelo  $AR(1)$  fue del 86 %.

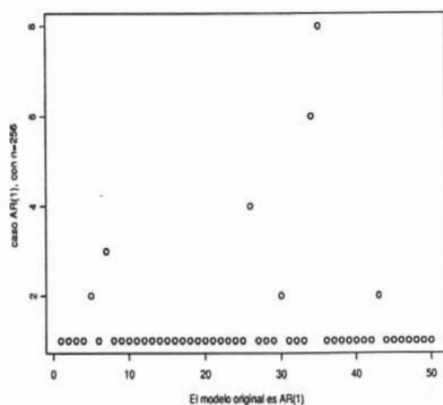


Figura 4.14: Gráfica de los ordenes de los modelos tipo Autoregresivo ajustados a un proceso proveniente de un modelo autoregresivo, método paramétrico, ( $N=256$ ).



En la figura (4.15) se muestran tanto los promedios como las desviaciones estándar de los coeficientes de correlación parcial a lag 1, junto con sus respectivos valores asintóticos tanto de la media como de la desviación estándar respectivamente, obtenidos vía la ecuación (4.7) y vía simulación.

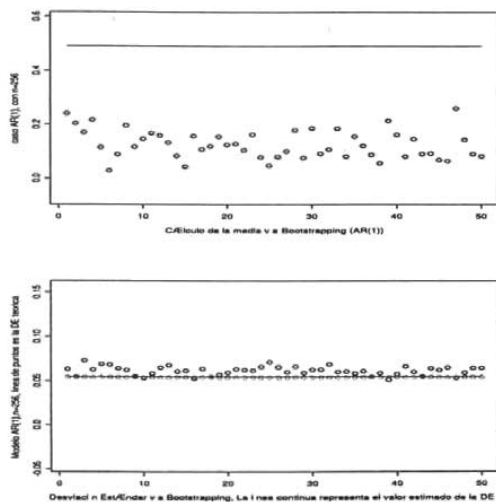


Figura 4.15: Gráfica de los promedios y desviaciones estándar de los coeficientes de correlación parcial al lag 1, método paramétrico, ( $N=256$ ).

Para el caso de los promedios se puede observar que estos valores están por debajo del valor asintótico estimado vía simulación.

Para el caso de la desviación estándar, se está mostrando tanto la desviación estándar asintótica teórica, representada por la línea punteada, y la desviación estándar asintótica estimada vía simulación, la línea continua que coincide con la línea punteada, es decir, de la desviación estándar teórica. Como se podrá observar los valores estimados para la desviación estándar son ligeramente mayores que los valores asintóticos estimados arriba descritos, es decir, el bootstrap paramétrico esta arrojando buenos estimadores que como se mencionó al principio del capítulo es lo que se esperaba, este hecho se ha venido repitiendo para los tamaños de muestra anteriores, lo cual satisface las expectativas generadas.

Para el caso no paramétrico, en la figura (4.16) se muestran las gráficas de los resultados

obtenidos tanto para la media como para la desviación estándar contrastados con sus respectivos valores asintóticos ya mencionados con anterioridad.

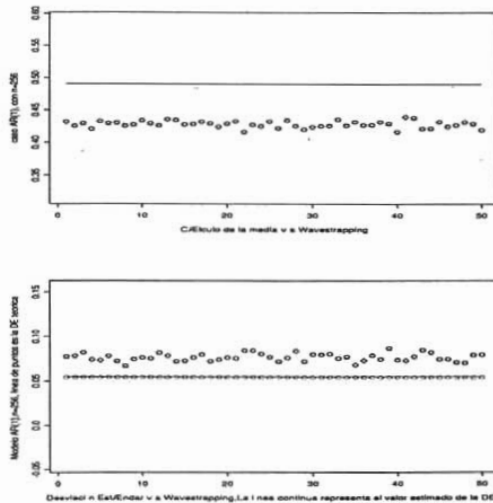


Figura 4.16: Gráfica de los promedios y las desviaciones estándar junto con los valores asintóticos estimados tanto para la media, como para la desviación estándar para el caso en que el proceso original es de tipo  $AR(1)$ , método no paramétrico, ( $N=256$ ).

Al igual que antes, tenemos que los valores de los promedios están por debajo del valor asintótico, sin embargo estos valores se encuentran más cercanos que los obtenidos utilizando bootstrap paramétrico.

Para el caso de la desviación estándar, al igual que antes se está mostrando tanto la desviación estándar asintótica teórica, como la desviación estándar asintótica estimada vía simulación. Como se podrá observar los valores de las desviaciones estándar de los coeficientes de correlación parcial a lag 1 son cercanos tanto al valor de la desviación estándar asintótica estimada, como a la Desviación estándar asintótica teórica, ya que estos prácticamente coinciden.

Los resultados obtenidos por ambos métodos se encuentran resumidos en el cuadro (4.5). Dado que se cuentan con 50 estimadores para la media y la desviación estándar del coeficiente de correlación se calcularon medidas de tendencia central: la moda, la mediana y la media.

Cuadro 4.5: Caso  $n=256$ , para un proceso Autoregresivo de orden 1, el porcentaje de veces que se eligió el modelo  $AR(1)$  fue del 86 %.

	Orders	Media			Desviación Estándar			
		Bootst	Wavest	Asym	Bootst	Asym Est	Wavest	Asym
Moda	1	-	-	0.490531	-	0.054699	-	0.054127
Media	-	0.126635	0.428367	-	0.061442	-	0.076535	-
Mediana	1	0.118233	0.428191	0.490531	0.061962	0.054699	0.076282	0.054127
Std Dev	-	0.052965	0.005194	-	0.005022	-	0.004356	-

Como se podrá observar para el caso de la media es notablemente mejor el método del bootstrap no paramétrico (ver comentarios en el capítulo 5), la diferencia tanto en la media como en la mediana con respecto al valor obtenido vía bootstrap no paramétrico es mucho menor que con respecto a los obtenidos vía bootstrap paramétrico, además de ser el que menor variación tiene, de hecho es casi cero a comparación de la variación de los estimadores obtenidos por el método paramétrico.

Para el caso de la desviación estándar cambian notablemente las cosas, el bootstrap paramétrico genera mejores estimadores que el bootstrap no paramétrico, que era lo que se esperaba dado que se conocen los parámetros del modelo del cual provienen los datos, nuevamente al igual que para los tamaños de muestra anteriores, los estimadores obtenidos por el método no paramétrico son los que menos variación tienen.

Ahora veamos los resultados en el caso en que el proceso original es de promedios móviles y se le ajusto un modelo autoregresivo, es decir, el caso en que nos equivocamos al seleccionar el modelo.

En la figura (4.17) se presentan los ordenes de los modelos de tipo autoregresivo elegidos, el porcentaje de veces que se eligió un modelo  $AR(1)$  fue de 2%; el modelo que se eligió con mayor frecuencia fue el modelo  $AR(2)$ , en un 38% de las veces.

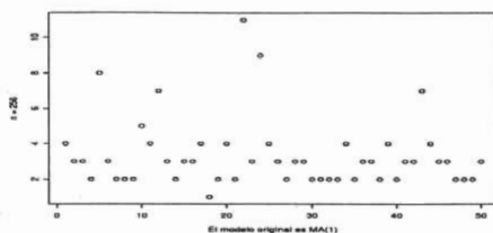


Figura 4.17: Gráfica de los ordenes de los modelos tipo Autoregresivo ajustados a un proceso proveniente de un modelo de promedios móviles, método paramétrico, ( $N=128$ ).

En la gráfica (4.18) se muestran tanto los promedios como las desviaciones estándar de los coeficientes de correlación parcial al lag 1, junto con sus respectivos valores asintóticos tanto de la media como de la desviación estándar respectivamente, obtenidos vía la ecuación (4.8) y vía simulación.

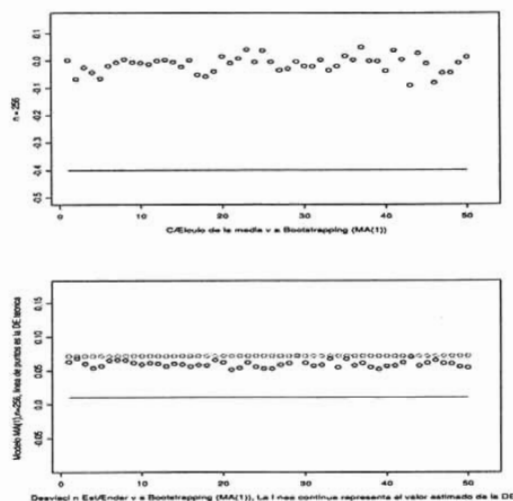


Figura 4.18: Gráfica de los promedios y las desviaciones estándar de los coeficientes de correlación parcial al lag 1 para procesos de tipo  $MA(1)$ , método paramétrico, ( $N=256$ ).

Para el caso de los promedios se puede observar que estos valores están por encima del valor

asintótico. Para el caso de la desviación estándar, se está mostrando la desviación estándar asintótica teórica, que nuevamente está representado por la línea punteada, y la desviación estándar asintótica estimada vía simulación, que está representada por la línea continua ubicada por debajo de los valores de las desviaciones estándar estimadas obtenidas vía bootstrap no paramétrico.

Para el caso no paramétrico en la gráfica (4.19) se presentan los resultados obtenidos contrastados con los valores asintóticos ya mencionados con anterioridad:

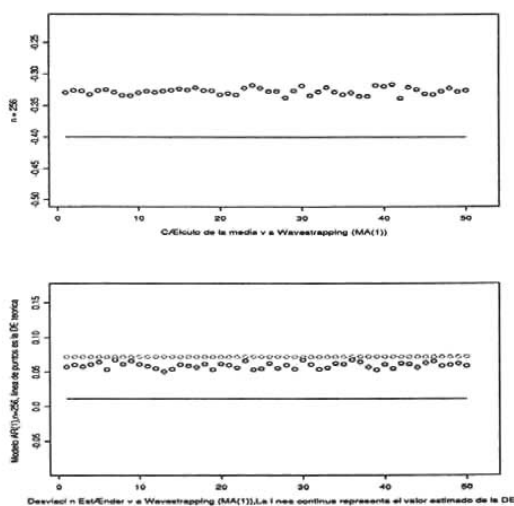


Figura 4.19: Gráfica de los promedios y las desviaciones estándar junto con los valores asintóticos estimados tanto para la media, como para la desviación estándar para el caso en que el proceso original es de tipo  $MA(1)$ , método no paramétrico, ( $N=256$ ).

En el caso de los promedios a pesar de que el valor asintótico es menor que los promedios calculados, la distancia que hay entre los mismos es menor que para el caso del bootstrap paramétrico. No hay diferencia significativa entre ambos métodos para este tamaño de muestra en los estimadores para la desviación estándar

Los resultados obtenidos en ambos métodos se encuentran resumidos en el cuadro (4.6). Dado que se cuentan con 50 estimadores para la media y la desviación estándar del coeficiente de correlación se calcularon medidas de tendencia central: media, mediana y moda.

Cuadro 4.6: Caso  $n=256$ , para un proceso de tipo promedios móviles de orden 1, El porcentaje de veces que se eligió el modelo  $AR(1)$  fue de 2%.

	Orders	Media			Desviación Estándar		
		Bootst	Wavest	Asym	Bootst	Wavest	Asym
Moda	2	-	-	-0.399888	-	-	0.071807
Media	-	-0.013429	-0.327308	-	0.059831	0.059386	-
Mediana	3	-0.007826	-0.327252	-0.399888	0.059658	0.060071	0.071807
Std Dev	-	0.030083	0.005225	-	0.004798	0.004206	-

Se podrá observar que para obtener estimadores de la media es notablemente mejor el método del bootstrap no paramétrico que el bootstrap paramétrico. Para estimar la desviación estándar, el método no paramétrico produce estimadores con menor variación que el paramétrico, por esta razón el bootstrap no paramétrico proporciona mejores estimadores que el bootstrap paramétrico.



## Capítulo 5

### Conclusiones

Como ya se mencionó en el capítulo anterior, para encontrar estimadores del valor verdadero de la media y la desviación estándar del coeficiente de autocorrelación parcial a lag 1, se obtuvieron 50 estimadores para la media y 50 estimadores de la desviación estándar del coeficiente de correlación parcial a lag 1 de tal forma que para obtener un representante de cada grupo de estimadores se calcularon las tres medidas de tendencia central (media, mediana y moda).

Primero analizaremos los resultados obtenidos cuando el proceso original es de tipo autoregresivo considerando el promedio de los estimadores, comparándolo con el valor asintótico tanto de la media como de la desviación estándar.

En el cuadro (5.1) se muestran los porcentajes en que el valor asintótico es subestimado ó sobreestimado por el promedio de los estimadores de la media, el signo negativo es un indicador de que el valor asintótico es sobreestimado mientras que el signo positivo indica que el valor asintótico está siendo subestimado por los estimadores de la media y la desviación estándar del coeficiente de correlación parcial.



Cuadro 5.1: Caso en que el proceso es de tipo autoregresivo, se muestra el porcentaje en que ambos métodos sobreestiman (subestiman) el valor verdadero de la media o de la desviación estándar del coeficiente de autocorrelación parcial a lag 1, considerando como estimador el promedio de los estimadores generados por ambos métodos.

Promedio	Media		Desviación Estándar	
	Bootstrap	Wavestrapping	Bootstrap	Wavestrapping
$N$				
64	79.42 %	12.02 %	-56.66 %	-88.92 %
128	73.36 %	14.81 %	-11.86 %	-39.71 %
256	74.18 %	12.67 %	-13.52 %	-41.40 %

Como se podrá observar en el cuadro (5.1) para estimar el valor verdadero de la media el bootstrap no paramétrico genera mejores estimadores que el bootstrap paramétrico, puesto que el primero produce valores que subestiman en menor porcentaje el valor verdadero de la media para los tres tamaños de muestra.

Para el caso de la desviación estándar se observa que el bootstrap paramétrico sobreestima en menor porcentaje al valor asintótico de la desviación estándar que el bootstrap no paramétrico, en este sentido el bootstrap paramétrico proporciona mejores estimadores para la desviación estándar del coeficiente de autocorrelación parcial a lag 1 que el método no paramétrico, hecho que se esperaba, puesto que se conocen los parámetros verdaderos y por tanto es más factible poder determinar valores más cercanos al verdadero modelo.

Aquí vale la pena hacer notar que en el caso de la desviación estándar existe una gran diferencia cuando se considera un tamaño de muestra de 64, para este caso en particular las estimaciones son muy inferiores al valor asintótico de la desviación estándar, esto se debe principalmente al tamaño de las muestras, además como ya se mencionó en el capítulo 4 para este tamaño de muestra es muy probable cometer errores al momento de seleccionar algún modelo.

Si uno considera a la mediana de los respectivos promedios y desviaciones estándar como estimador de los valores verdaderos de la media y la desviación estándar respectivamente, en el cuadro (5.2) presentamos los porcentajes en que la mediana de los estimadores para la media y la desviación estándar obtenidos vía el bootstrap paramétrico y el bootstrap no paramétrico sobreestiman ó subestiman los valores asintóticos ya mencionados.

Cuadro 5.2: Caso en que el proceso es de tipo autoregresivo, se muestra el porcentaje en que ambos métodos sobreestiman (subestiman) el valor verdadero de la media o de la desviación estándar del coeficiente de autocorrelación parcial a lag 1, considerando como estimador la mediana de los estimadores generados por ambos métodos.

Mediana	Media		Desviación Estándar	
	Bootstrap	Wavestrapping	Bootstrap	Wavestrapping
<i>N</i>				
64	82.63 %	11.84 %	-54.96 %	-88.14 %
128	78.64 %	14.83 %	-10.02 %	-40.31 %
256	75.90 %	12.71 %	-14.48 %	-40.93 %

Nuevamente se puede ver que para estimar los valores verdaderos el método no paramétrico produce mejores estimadores que el bootstrap paramétrico para la media, no siendo así cuando se quiere estimar la desviación estándar, en donde el bootstrap paramétrico se comporta mejor que el no paramétrico.

Los resultados obtenidos para los procesos de promedios móviles son particularmente interesantes, puesto que para casi todos los casos, como más adelante se verá, el bootstrap no paramétrico proporciona mejores estimadores, tanto para la media como para la desviación estándar del coeficiente de correlación parcial a lag 1, que el paramétrico. Solamente cuando el tamaño de muestra es igual a 64 y se quiere estimar el valor verdadero de la media ya sea por medio del promedio o de la mediana, el método paramétrico proporciona mejores estimadores que el no paramétrico.

En el cuadro (5.3) se presentan los porcentajes en que el promedio de los estimadores de la media y la desviación estándar sobreestiman ó subestiman a sus respectivos valores asintóticos para los tres tamaños de muestra analizados.

Cuadro 5.3: Caso en que el proceso es de tipo de promedios móviles, se muestra el porcentaje en que ambos métodos sobreestiman (ó subestiman) el valor asintótico de la media y de la desviación estándar del coeficiente de autocorrelación parcial a lag 1, considerando como estimador el promedio de los estimadores generados por ambos métodos.

Promedio	Media		Desviación Estándar	
	Bootstrap	Wavestrapping	Bootstrap	Wavestrapping
$N$				
64	86.44 %	166.35 %	18.93 %	2.48 %
128	92.94 %	24.54 %	16.50 %	6.30 %
256	96.64 %	18.15 %	16.68 %	17.30 %

Como se puede observar en el cuadro (5.3), para el caso de la media, el método no paramétrico genera mejores estimadores que el bootstrap paramétrico, excepto en el caso de tamaño de muestra  $N = 64$ , ya que los estimadores generados por este último método subestiman en menor porcentaje a los valores asintóticos, por su parte el método paramétrico genera estimadores para los tres tamaños de muestra, totalmente alejados del valor asintótico de la media.

Para la desviación estándar el método no paramétrico en general genera mejores estimadores que el paramétrico, ya que estos subestiman en menor porcentaje el valor asintótico

En el cuadro (5.4) se presentan los porcentajes en que la mediana de los estimadores de los promedios y las desviaciones estándar subestiman a los respectivos valores asintóticos para los tres tamaños de muestra.

Cuadro 5.4: Caso en que el proceso original es de tipo  $MA(1)$ , en este cuadro se está mostrando el porcentaje en que ambos métodos sobreestiman (subestiman) el valor asintótico de la media o de la desviación estándar del coeficiente de autocorrelación parcial a lag 1, considerando como estimador el promedio de los estimadores generados por ambos métodos.

Mediana	Media		Desviación Estándar	
	Bootstrap	Wavestrapping	Bootstrap	Wavestrapping
$N$				
64	90.64 %	166.19 %	17.83 %	1.63 %
128	93.50 %	24.44 %	17.09 %	6.82 %
256	98.04 %	18.16 %	16.92 %	16.34 %

Como se puede observar en el cuadro (5.4), para el caso de la media, nuevamente el método no paramétrico genera mejores estimadores que el bootstrap paramétrico, excepto en el caso de tamaño de muestra  $N = 64$ , ya que los estimadores generados por este último método subestiman en menor porcentaje a los valores asintóticos, por su parte el método paramétrico genera estimadores para los tres tamaños de muestra, totalmente alejados del valor asintótico de la media.

Para el caso de la desviación estándar el método no paramétrico en general genera mejores estimadores que el paramétrico, ya que estos subestiman en menor porcentaje el valor asintótico

Lo anterior era de esperarse dado que en el bootstrap no paramétrico, los parámetros no juegan un papel determinante en la estimación tanto de la media como de la desviación estándar. No así para el bootstrap paramétrico donde tanto el modelo seleccionado como los parámetros considerados son determinantes en los resultados obtenidos, puesto que como ya se mencionó con anterioridad, el bootstrap paramétrico se realiza en los residuales los cuales a su vez son obtenidos después de ajustar el modelo *apropiado* al proceso generado, en este caso el proceso es de tipo promedios móviles y se le está buscando el mejor modelo autoregresivo de orden  $p$ ; he aquí la razón del porqué el bootstrap paramétrico no proporciona mejores estimadores que el no paramétrico.

Finalmente retomando el caso autoregresivo, en particular veamos a los estimadores del valor asintótico de la media. Contrario a lo que se esperaba, el bootstrap no paramétrico proporciona mejores estimadores que el bootstrap paramétrico. Con la finalidad de tratar de justificar el hecho anterior, se analizó el caso  $N = 64$ , la forma en que se hizo fue estudiando el tipo de muestras que se obtienen por ambos métodos.

Como ya se mencionó en el capítulo anterior, para encontrar estimadores del valor verdadero de la media del coeficiente de autocorrelación parcial a lag 1, se obtuvieron 50 estimadores para la media del coeficiente de correlación parcial a lag 1 de estos 50 promedios se detectó aquel cuya distancia al valor verdadero fuese mayor, para nuestro ejemplo numérico esto ocurrió con el valor número 28.

Recordemos que este es el promedio de 100 coeficientes de correlación parcial a lag 1, los cuales a su vez fueron obtenidos de 100 muestras bootstrap, por ambos métodos; de estas 100 muestras tipo bootstrap (100 por cada método) correspondientes se trató de determinar que tan correlacionadas son las muestras obtenidas.

Para poder determinar lo anterior, una vez seleccionado el coeficiente, se procedió a analizar

sus 100 muestras correspondientes, esto se hizo calculando los coeficientes de correlación entre dos muestras consecutivas, por lo tanto se tienen 99 coeficientes de correlación, después a cada uno de estos coeficientes se les construyó su respectivo intervalo de confianza a un 95% de confianza, (ver Draper y Smith (1981)), para verificar la hipótesis nula de que el coeficiente de correlación es cero, es decir las muestras generadas no están correlacionadas entre sí, contra la hipótesis alternativa de que el coeficiente de correlación es distinto de cero.

En la figura (5.1) se muestran las correlaciones entre las 100 muestras bootstrap paramétrico correspondientes al promedio número 28 de los 50 estimados.

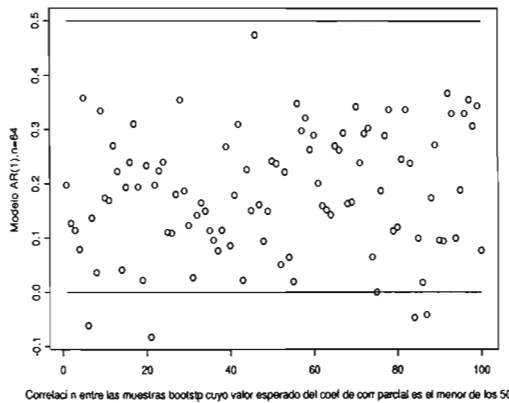


Figura 5.1: Gráfica de los promedios de los coeficientes de correlación entre dos muestras consecutivas para el estimador número 28 del valor asintótico de la media de los coeficientes de autocorrelación parcial a lag 1.

Los resultados obtenidos fueron los siguientes: para el caso del bootstrap paramétrico el 34% de las veces se rechazó la hipótesis nula, mientras que para el bootstrap no paramétrico solamente el 4% de las veces se rechazó la hipótesis nula, es por esto que se afirma que el bootstrap paramétrico genera muestras correlacionadas en mayor porcentaje a diferencia del método no paramétrico.

Lo anterior nos da una justificación al hecho de que a pesar de que se garantiza la convergencia en distribución a los valores asintóticos, para el método paramétrico esta convergencia es lenta a consecuencia de que el método mismo genera series de tiempo correlacionadas y por tanto los

valores esperados obtenidos son notablemente distintos y lejanos de los valores asintóticos. Por su parte para el método no paramétrico, dado que de las series que genera, un bajo porcentaje de ellas están correlacionadas entre sí, la convergencia en distribución es mucho más rápida que para el otro método.

Finalmente para el caso de la desviación estándar, en Perciva, Sardy y Davison (2000) se afirma que vía el método no paramétrico en general se obtienen buenos estimadores para la desviación estándar asintótica o verdadero, no así para la media, más sin embargo hay que resaltar que a pesar de lo afirmado en el citado artículo, sí genera mejores estimadores para la media que el método paramétrico.



# Apéndice A

## Apéndice A

### A.1. Teorema de Kolmogorov

**Teorema A.1.** (Teorema de Kolmogorov) *Las funciones de distribución de probabilidad*

$$\{F_t(\cdot) : t \in F\}$$

*son funciones de distribución de un proceso estocástico,  $\{X_t, t \in T\}$  si y sólo si se cumple que para toda  $n \in \mathbb{N}$ ,  $\mathbf{t} = (t_1, t_2, \dots, t_n)' \in F$  con  $1 \leq i \leq n$  se tiene que*

$$\lim_{x_i \rightarrow \infty} F_t(\mathbf{X}) = F_{t(i)}(\mathbf{X}(i))$$

*donde  $\mathbf{t}(i)$  y  $\mathbf{X}(i)$  son las  $n - 1$  componentes del vector, obtenidas al eliminar la  $i$ -ésima componente de  $\mathbf{t}$  y  $\mathbf{X}$  respectivamente; con*

$$F = \{(t_1, t_2, \dots, t_n)' \in T^n : t_1 \leq t_2 \leq \dots \leq t_n, n = 1, 2, \dots\}.$$

*Entonces las funciones de distribución finito dimensionales de  $\{X_t, t \in T\}$  son las funciones  $\{F_t(\cdot), t \in F\}$  definidas para  $\mathbf{t}$  por*

$$F_t(\mathbf{X}) = \mathbb{P}[X_{t_1} < x_1, X_{t_2} < x_2, \dots, X_{t_n} < x_n].$$

### A.2. Teoremas de Helly y Helly-Bray

Una función de distribución generalizada es una función definida sobre  $\mathbb{R} \cup \{-\infty, \infty\}$  creciente, tal que  $F(-\infty) = 0$  y  $F(\infty) = 1$ .

**Lema A.1.** *Sea  $\{F_n\}$  sucesión de funciones de distribución generalizadas tales que  $|F_n| \leq M$  para toda  $n$ ,*



supongamos que existe  $G : \mathbb{R} \rightarrow \mathbb{R}$  creciente y acotada, tal que  $\{F_n\}$  converge puntualmente a  $G$  sobre  $\mathcal{D} \subseteq \mathbb{R}$  denso. Entonces  $\{F_n(x)\}$  converge a  $G(x)$  para toda  $x \in \mathcal{D}$  punto de continuidad de  $G$ .

*Demostración.* Sea  $x$  punto de continuidad de  $G$ , sean  $w_k, y_k \in \mathcal{D}$  tales que  $w_k < x < y_k$  y  $y_k \downarrow x, w_k \uparrow x$ , entonces

$$F_n(w_k) \leq F_n(x) \leq F_n(y_k)$$

dado que  $\{F_n\}$  converge puntualmente a  $G$  sobre  $\mathcal{D}$ , tenemos que

$$G(w_k) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq G(y_k)$$

si hacemos  $k \rightarrow \infty$ , entonces

$$G(x) = G(x^-) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq G(x^+) = G(x)$$

por lo tanto

$$\lim_{n \rightarrow \infty} f_n(x) = G(x)$$

para todo  $x$  punto de continuidad de  $G$ . □

**Lema A.2.** Sea  $\{F_n\}$  sucesión de funciones de distribución generalizadas tales que  $|F_n| \leq M$  para toda  $n$  y sea  $D = \{x_k | k \geq 1\}$  conjunto numerable de números reales, entonces existe una subsucesión  $\{F_{n_j}\}$  de  $\{F_n\}$  tal que  $\{F_{n_j}(x)\}$  converge para toda  $K$  cuando  $j \rightarrow \infty$ .

*Demostración.* Sean  $\{F_n(x_1)\}_{n \geq 1} \subseteq [-M - M]$  entonces existe una subsucesión  $\{F_{n_1}^1(x_1)\}_{n_1 \geq 1}$  convergente, es decir, existe  $a_1 \in \mathbb{R}$  tal que  $\{F_{n_1}^1(x_1)\}$  converge a  $a_1$ . Para  $\{F_{n_1}^1(x_2)\}$  una subsucesión convergente  $\{F_{n_2}^2(x_2)\}$  a  $a_2 \in \mathbb{R}$  y como  $\{F_{n_2}^2(x_2)\} \subseteq \{F_{n_1}^1(x_1)\}$  entonces se tiene que  $\{F_{n_2}^2(x_1)\}$  converge a  $a_1$ , procediendo sucesivamente tenemos una sucesión  $\{F_{n_1}^1(x_1)\}, \{F_{n_2}^2(x_2)\}, \dots, \{F_{n_k}^k(x_k)\}$  tal que  $\{F_{n_k}^k(x_k)\}$  converge a  $a_k$  para toda  $k = 1, 2, \dots$ . Consideremos la diagonal  $\{F_{n_n}^n(\cdot) : n \geq 1\}$  y veamos que para toda  $x \in \mathcal{D}$ , se tiene que  $\{F_{n_n}^n(x)\}_{n \geq 1}$  converge. Sea  $k \geq 1, x_k \in \mathcal{D}$ , entonces  $\{F_{n_n}^n(x_k)\}_{n \geq k}$  es una subsucesión de  $\{F_{n_n}^k(x_k)\}_{n \geq 1}$ , la cual ya se vio, que converge a  $a_k$ , por lo tanto  $\{F_{n_n}^n(x_k)\}_{n \geq k}$  converge a  $a_k$  cuando  $n \rightarrow \infty$ . □

**Teorema A.2. (Teorema de Helly)** Sean  $\{F_n\}$  sucesión de funciones de Distribución Generalizadas tales que  $|F_n| \leq M$  para toda  $n$ , entonces  $\{F_n\}$  contiene una subsucesión  $\{F_{n_k}\}$  que converge débilmente a  $F$  función de distribución generalizada.

*Demostración.* Sea  $\mathcal{D} = \{x_k : k \geq 1\}$  subconjunto denso de  $\mathbb{R}$ , por el lema (A.2) existe una subsucesión  $\{F_{n_j}\}_{j \geq 1}$  de  $\{F_n\}_{n \geq 1}$ , tal que, para toda  $x_k \in \mathcal{D}$ ,

$$\lim_{j \rightarrow \infty} F_{n_j}(x_k) = a_k = G(x_k)$$

por el lema (A.1). Sea  $G(x) = \sup_{x_k < x} G(x_k)$  con  $x \in \mathcal{D}^c$  y  $x_k \in \mathcal{D}$ , entonces se tiene que  $G(x)$  es acotada puesto que para toda  $k$  se cumple

$$|\lim_{j \rightarrow \infty} F_{n_j}(x_k)| = |G(x_k)| \leq M.$$

Ahora veamos que  $G$  es creciente:

Caso 1: Sean  $x_k, x_l \in \mathcal{D}$  tales que  $x_k < x_l$ , entonces

$$F_{n_j}(x_k) \leq F_{n_j}(x_l)$$

para toda  $j$ , por lo tanto

$$G(x_k) = \lim_{j \rightarrow \infty} F_{n_j}(x_k) \leq \lim_{j \rightarrow \infty} F_{n_j}(x_l) = G(x_l)$$

Caso 2: Supongamos que  $x \notin \mathcal{D}$  y  $x < x_k$ , donde  $x_k \in \mathcal{D}$ , sea  $x_l \in \mathcal{D}$  tal que  $x_l < x$ , entonces  $x_l < x_k$  y por el caso 1 se tiene que  $G(x_l) \leq G(x_k)$  por lo tanto

$$G(x) = \sup_{x_l < x} G(x_l) \leq G(x_k)$$

Caso 3: Ahora supongamos  $x_k \in \mathcal{D}$ ,  $x_k < x$ , y  $x \notin \mathcal{D}$ , entonces se tiene

$$G(x) \geq \sup_{x_l < x, x_l \in \mathcal{D}} G(x_l) \geq G(x_k)$$

para  $k \in \{l : l \geq 1\}$

Caso 4:  $x < y$ ,  $x, y \notin \mathcal{D}$ , entonces

$$\{x_l \in \mathcal{D} | x_l < x\} \subset \{x_l \in \mathcal{D} | x_l < y\}$$

por lo tanto se tiene que

$$G(x) = \sup_{x_l < x, x_l \in \mathcal{D}} G(x_l) \leq \sup_{x_l < y, x_l \in \mathcal{D}} G(x_l) = G(y)$$

por lo tanto  $G$  es creciente.

Ahora, como

$$\lim_{n \rightarrow \infty} F_{n_j}(x) = G(x)$$

para  $x \in D$  conjunto Denso, se tiene que por el lema (A.1)  $\{F_{n_j}\}$  converge a  $G(x)$  para todo  $x$  punto de continuidad de  $G$ .

Sea

$$F(x) = \begin{cases} G(x), & x \in C_G = \{x \in \mathbb{R} : G \text{ es continua en } x\} \\ G(x^+) & x \notin C_G \end{cases}$$

entonces se tiene que  $|F| \leq M$ , es decir,  $F$  es acotada.

Además,  $F$  es creciente, puesto que si  $x < y$ , entonces dado que  $G$  es creciente:

- (i) Si  $x, y \in C_G$ , entonces  $F(x) = G(x) \leq G(y) \leq F(y)$
- (ii) Si  $x \in C_G, y \notin C_G$ , entonces,  $F(x) = G(x) \leq G(y) \leq G(y^+) = F(y)$
- (iii) Si  $x \notin C_G, y \in C_G$ , entonces  $F(x) = G(x^+) \leq G(y^+) = G(y) = F(y)$
- (iv) Si  $x \notin C_G, y \notin C_G$ , entonces,  $F(x) = G(x^+) \leq G(y^+) \leq F(y)$

por lo tanto se tiene que  $F$  es creciente, además de ser continua por la derecha por como está definida. Finalmente,  $F_{n_j}$  converge débilmente a  $F$ , puesto que  $x$  es punto de continuidad de  $F$  si y sólo si  $x$  es punto de continuidad de  $G$ , pero  $F_{n_j}$  converge a  $G(x) = F(x)$ , para todo  $x$  punto de continuidad de  $G$  y por lo tanto de  $F$ . □

**Teorema A.3. (Teorema de Helly-Bray)** Sea  $g : [a, b] \rightarrow \mathbb{R}$  función continua y acotada y sea  $\{F_n\}$  sucesión de funciones de distribución generalizadas tales que  $|F_n| \leq M$  y tales que  $\{F_n\}$  converge débilmente a  $F$  donde  $a, b \in C_F = \{x \in \mathbb{R} | F \text{ es continua en } x\}$ ; entonces

$$\lim_{n \rightarrow \infty} \int_a^b g dF_n = \int_a^b g dF.$$

*Demostración.* A saber,  $F$  es acotada y creciente con límite por la derecha e izquierda. Sea  $\mu_n$  medida asociada a  $F_n$ , definida por la relación:

$$\mu_n((-\infty, x]) = F_n(x)$$

y sea  $\mu$  medida asociada a  $F$ , definida

$$\mu((-\infty, x]) = F(x)$$

para toda  $x \in \mathbb{R}$ , y  $\{(-\infty, x]\}$  clase determinante. Sea  $a = x_{0,N} < x_{1,N} < \dots < x_{N,N} = b$  partición de  $[a, b]$  tal que

$$\lim_{N \rightarrow \infty} A_N = \lim_{N \rightarrow \infty} \max_{1 \leq j \leq k_N} (x_{j,N} - x_{j-1,N}) = 0,$$

entonces función  $g$  puede aproximarse uniformemente sobre el intervalo  $[a, b]$  por una sucesión de funciones simples  $\{g_n\}$  tales que

$$\lim_{N \rightarrow \infty} \sup_{a \leq x \leq b} |g_N(x) - g(x)| = 0,$$

es decir,

$$g_N(x) = \sum_{k=1}^{K_N} g(x_{K,N}) \mathbb{1}_{(x_{k-1,N}, x_{K,N}]}$$

y tal que  $|g_N| \leq |g|$  para toda  $N$ . La sucesión  $\{g_n\}$  es uniformemente acotada y por el *Teorema de Convergencia Dominada* se tiene

$$\int_a^b g dF_n = \int_{(a,b]} g d\mu_n = \lim_{N \rightarrow \infty} \int_{(a,b]} g_N d\mu_n$$

para toda  $n \geq 1$  y

$$\int_a^b g dF = \int_{(a,b]} g d\mu = \lim_{N \rightarrow \infty} \int_{(a,b]} g_N d\mu.$$

Ahora veamos que para toda  $n \geq 1$  se cumple que

$$\lim_{n \rightarrow \infty} \int_{(a,b]} g_N d\mu_n = \int_{(a,b]} g_N d\mu.$$

Sean  $a = x_{0,N} < x_{1,N} < \dots < x_{N,N} = b$  puntos de continuidad de  $F$ ; dado que la sucesión  $\{F_n\}$  converge débilmente a  $F$

$$\lim_{n \rightarrow \infty} \mu_n((x_{k-1,N}, x_{k,N}]) = \lim_{n \rightarrow \infty} [F_n(x_{k,n}) - F_n(x_{k-1,n})] = F(x_{k,n}) - F(x_{k-1,n}) = \mu((x_{k-1,N}, x_{k,N}]),$$

para toda  $k = 1, 2, \dots, k_N$ , por lo tanto

$$\lim_{n \rightarrow \infty} \int_{(a,b]} g_N d\mu_n = \lim_{n \rightarrow \infty} \sum_{k=1}^{k_N} g(x_{k,N}) \mu_n((x_{k-1,N}, x_{k,N}]) = \sum_{k=1}^{k_N} g(x_{k,N}) \mu((x_{k-1,N}, x_{k,N}]) = \int_{(a,b]} g_N d\mu.$$

Sea  $M_N = \sup_{a \leq x \leq b} |g_N(x) - g(x)|$ , tal que  $\lim_{N \rightarrow \infty} M_N = 0$ , entonces

$$\begin{aligned}
 & \left| \int_a^b g dF_n - \int_a^b g dF \right| = \left| \int_{(a,b]} g d\mu_n - \int_{(a,b]} g d\mu \right| \\
 = & \left| \int_{(a,b]} g d\mu_n - \int_{(a,b]} g_N d\mu_n + \int_{(a,b]} g_N d\mu_n + \int_{(a,b]} g_N d\mu - \int_{(a,b]} g_N d\mu - \int_{(a,b]} g d\mu \right| \\
 = & \left| \left( \int_{(a,b]} |g - g_N| d\mu_n \right) + \left( \int_{(a,b]} g_N d\mu_n - \int_{(a,b]} g_N d\mu \right) + \int_{(a,b]} (g - g_N) d\mu \right| \\
 \leq & \int_{(a,b]} |g - g_N| d\mu_n + \left| \int_{(a,b]} g_N d\mu_n - \int_{(a,b]} g_N d\mu \right| + \int_{(a,b]} |g - g_N| d\mu \\
 \leq & M_n \mu_n((a, b]) + \left| \int_{(a,b]} g_N d\mu_n - \int_{(a,b]} g_N d\mu \right| + M_n \mu((a, b])
 \end{aligned}$$

como  $\lim_{N \rightarrow \infty} M_N = 0$  y si hacemos  $n \rightarrow \infty$ , entonces,

$$\lim_{N, n \rightarrow \infty} \left| \int_a^b g dF_n - \int_a^b g dF \right| = 0$$

□

# Bibliografía

- [1] Brockwell Peter J., Davis Richard A., 1998, *Time Series: Theory and Methods*, Springer Series in Statistics.
- [2] Bühlmann Peter and Künsch H.R., 1995, The blockwise bootstrap for general parameters of stationary time series, *Scandinavian Journal of Statistic* Vol. 22, pp 35-54.
- [3] Bühlmann Peter and Künsch H.R., 1999, Block length selection in the bootstrap for time series, *Computational Statistics and Data Analysis* Vol. 31, pp 295-310.
- [4] Bühlmann Peter, 2001, Bootstraps for Time Series, *Statistical Science*, vol. 17, pp. 52-72.
- [5] Contreras Cristán Alberto, Ph. D. Thesis, Advances in the analysis of time series using wavelets, 1998.
- [6] Contreras Cristán Alberto, Walden Andrew T., 2002, Multitaper Power Spectrum Estimation and Thresholding: Wavelets Packets Versus Wavelets, *IEEE Transactions on Signal Processing*, Vol. 50, No. 12, pp 2976-2986.
- [7] Davison A.C., Hinkley D.V., 1997, *Bootstrap Methods and their Applications*, Cambridge University Press.
- [8] Draper Norman R., Smith Harry, 1981, *Applied Regression Analysis*, Wiley and Sons.
- [9] Franke, J., Kresiss, J.P., Mammen, E. and Neumann, M.H., 1998, Properties of the non-parametric autoregressive bootstrap, Preprint, *Universität Kaiserlautern*.
- [10] Graps Amara, 1995, Waveletes, *Institute of Electrical and Electronics Engineers*, Vol. 2, Num. 2, 10662.
- [11] Kresiss, J. P. and Franke J., 1992, Bootstrapping stationary autoregressive moving-average models, *Journal of Time Series Analysis* Vol. 13, pp. 297-317.

- 
- [12] Li, H. and Maddala, G.S., 1996, Bootstrapping Time Series Models, *Econometric Reviews*, Vol. 15, Num. 2, pp 115-158.
- [13] Mackenzie Dana, 2001, Wavelets: ver el bosque y los árboles, *U.S. National Academy of Sciences*.
- [14] Percival Donald B., Walden Andrew T., 1994, *Spectral Analysis for Physical Applications*, Cambridge University Press.
- [15] Percival Donald B., Walden Andrew T., 2000, *Wavelets Methods for Time Series Analysis*, Cambridge University Press.
- [16] Percival Donald B, S. Sardy, and Davison A.C., 2000, *Nonlinear and Nonstationary Signal Processing*, Cambridge University Press.
- [17] Percival Donald B., Constantine William L.B., 2002, Exact Simulation of Time-Varying Fractionally Differenced Processes, submitted article.
- [18] Stephens M. A., 1974, EDF Statistics for goodness of fit and some comparisons, *Journal of the American Statistical Association*, vol 69, pp. 730-737.
- [19] Strang Gilbert, 1985, *Algebra Lineal y Aplicaciones*, Prentice Hall.
- [20] Vetterli Martin, Kovacevic Jelena, 1995, *Wavelets and Subband Coding*, Prentice Hall.
- [21] Vidakovic Brani, Muller Peter, 1991, *Wavelets for kids*, Insitute of Statistics and Decision Sciences, Duke University.
- [22] Vidakovic Brani, 1999, Statistical Modeling by Wavelets, *Wiley Series in Probability and Statistics*.
- [23] Walden Andrew T., Percival Donald B., and McCoy E. J., 1998, Spectrum estimation by wavelet thresholding of multitaper estimators, *IEEE Transactions on Signal Processing*, Vol. 46, no. 12, pp. 3153-3165.