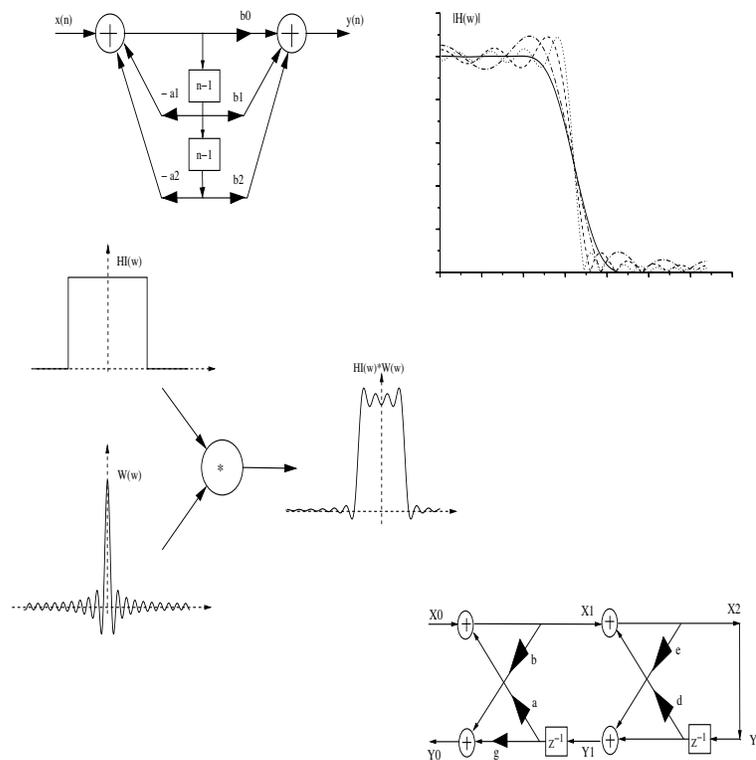


Diseño de Filtros Digitales



M.I. Larry Escobar Salguero

Marzo de 2006

Diseño de filtros digitales

Larry H. Escobar S.

Facultad de Ingeniería, UNAM

Índice general

Introducción	VIII
1. Los filtros digitales	1
1.1. Clasificación y técnicas de diseño de FD	3
1.1.1. Clasificación general	3
1.1.2. Clasificación de acuerdo al uso e implementación de los FD	3
1.2. Técnicas de diseño de filtros digitales lineales	4
1.3. Características de los FD lineales	4
1.4. Metodología para el diseño de un filtro digital	5
1.4.1. Especificaciones de un FD	6
1.5. Resumen	8
2. Filtros digitales de respuesta finita al impulso	9
2.1. Características de los filtros FIR	9
2.2. Estructuras de los filtros FIR	10
2.3. Diseño de filtros FIR	13
2.3.1. Aproximación de un FD FIR a un filtro ideal PB	15
2.3.2. Soluciones de un filtro FIR de fase lineal	19
2.4. Diseño de filtros FIR por series de Fourier y ventanas	20
2.4.1. Metodología de diseño de filtros FIR utilizando ventanas	22
2.4.2. Efecto de una ventana rectangular	22
2.4.3. Tipos de ventanas	24
2.4.4. Ejemplos de diseño	31
2.5. Diseño de filtros FIR paso altas, paso banda y supresor de banda	34
2.5.1. Diseño de filtros FIR FPA a partir de un FIR FPB	34
2.5.2. Diseño de filtros FIR paso banda	36
2.6. Diseño de filtros FIR por muestreo en frecuencia	38
2.6.1. Procedimiento de diseño utilizando soluciones de fase lineal	39
2.6.2. Procedimiento de diseño utilizando la DFTI	44
2.6.3. Estructura de muestreo en frecuencia	45
2.6.4. Localización de los polos y ceros de un filtro FIR	46

2.7. Método Remez y algoritmo Parks McClellan	49
2.8. Otros tipos de filtros FIR	56
2.8.1. Filtro FIR tipo peine	56
2.8.2. Filtro FIR diferenciador	59
2.8.3. Filtros FIR por transformada de Hilbert	59
2.9. Diseño de filtros FIR con programas simuladores	60
2.10. Resumen	60
3. Filtros de respuesta infinita al impulso (IIR)	61
3.1. Características de los filtros IIR	61
3.2. Estructuras de filtros digitales IIR	62
3.2.1. Filtro IIR forma directa	63
3.2.2. Filtro IIR en cascada	66
3.2.3. Filtro IIR en paralelo	66
3.3. Métodos de diseño de filtros IIR	67
3.3.1. Transformación de filtros analógicos-digitales	68
3.3.2. Prototipos analógicos o aproximaciones polinómicas	71
3.3.3. Prototipo Butterworth paso bajas	72
3.3.4. Prototipo Chebyshev	75
3.3.5. Diseño de filtros elípticos	84
3.3.6. Filtros de Bessel	88
3.4. Transformaciones de filtros analógicos	89
3.4.1. Diseño de filtros IIR por respuesta al impulso invariante	89
3.4.2. Transformada Z - matched	94
3.4.3. Diseño por aproximación numérica de la ecuación en diferencias	96
3.4.4. Diseño de filtros IIR por Transformada Bilineal	97
3.4.5. Metodología de diseño de un filtro IIR por TB	104
3.4.6. Método de transformaciones en la frecuencia	106
3.5. Diseño por localización de los polos y ceros en el plano z	116
3.6. Diseño de filtro IIR utilizando programas	122
3.7. Resumen	122
4. Predicción lineal	123
4.1. Predicción lineal	123
4.2. Predicción hacia adelante y ecuación normal	125
4.2.1. Relación entre predicción lineal y el modelo autorregresivo	130
4.3. Predicción lineal hacia atrás	135
4.4. Algoritmos para solución de la ecuación normal	141
4.4.1. Ecuación normal para predicción hacia adelante y hacia atrás	144
4.4.2. Relación entre la secuencia de autocorrelación y los coeficientes de reflexión	149

4.4.3. Recursión de “step-down” o Levinson inverso	150
4.5. Resumen	152
5. Filtros lattice	153
5.1. Características de los filtros lattice	153
5.2. Deducción de la estructura lattice	154
5.3. Comparación de estructuras lattice con estructuras FIR	156
5.3.1. Filtro lattice para diferentes órdenes	157
5.3.2. Filtro lattice de primer orden	157
5.3.3. Filtro lattice de segundo orden	157
5.3.4. Filtro lattice de orden m	158
5.4. Filtro lattice todo polo IIR	161
5.4.1. Filtro lattice todo polo de orden uno	163
5.4.2. Filtro lattice todo polo de orden dos	164
5.4.3. Filtro lattice todo polo de orden p	165
5.4.4. Filtro lattice de segundo orden con polos y ceros	165
5.5. Filtros lattice-ladder	167
5.6. Resumen	171
6. Aplicación en síntesis de voz	173
6.1. Síntesis de voz	173
6.2. Diseño del sistema de síntesis de voz	178
6.2.1. Resultados de síntesis de voz	182
6.3. Resumen	183
A. Glosario	185
B. Nomenclatura	187
C. Demostración de la convolución	189
D. Espectro de la ventana de Hamming	190
E. Algoritmo de Levinson-Durbin	192
F. Algoritmo de Leroux - Gueguen	196
Bibliografía	198

Índice de figuras

1.1. Plantilla de especificaciones de un filtro	8
2.1. Forma directa de un filtro FIR	12
2.2. Forma cascada de un filtro FIR	12
2.3. Forma fase lineal de un filtro FIR	14
2.4. $H_I(\omega)$ de un filtro paso bajas ideal	16
2.5. Resultado de aplicar un filtro FIR promediador	19
2.6. Efecto de convolucionar el espectro de una ventana cuadrada con $H_I(\omega)$	23
2.7. Efecto de una ventana rectangular al truncar $h(n)$	24
2.8. Comparación de ventanas: rectangular, triangular y Hanning	26
2.9. Comparación de ventanas: de Hamming y Blackman	26
2.10. Respuesta en frecuencia de un filtro FIR para diferente N	33
2.11. Diseño de un filtro FIR ideal paso altas	35
2.12. Diseño de un filtro FIR ideal paso banda	37
2.13. Espectro para el ejemplo	43
2.14. Estructura de muestreo en la frecuencia con $M=4$	47
2.15. Condiciones equirrizo para el método de Herrmann	53
2.16. Relación de $H(\omega)$ y $H_L(\omega)$ de un filtro peine	58
2.17. Filtro FIR tipo peine, $L = 4$	58
3.1. IIR forma directa I	64
3.2. IIR forma directa II o canónica	65
3.3. Forma directa transpuesta	65
3.4. Filtro IIR de cuarto orden en cascada	66
3.5. Filtro IIR de cuarto orden en paralelo	67
3.6. Transformación de $H(s)$ a $H(z)$	69
3.7. Especificaciones de un FPB analógico	70
3.8. Prototipo de un filtro Butterworth paso bajas en función del orden n	72
3.9. Parámetros para el cálculo del orden n de un filtro Butterworth	75
3.10. Respuesta en frecuencia de la aproximación Chebyshev I	78
3.11. Cálculo de orden n en un filtro con aproximación Chebyshev I	83
3.12. Respuesta en frecuencia de un filtro elíptico	88

3.13. Mapeo del plano s al plano z por diseño de respuesta al impulso invariante	92
3.14. Aproximación por diferencias finitas	98
3.15. La transformada bilineal del plano s al plano z	101
3.16. Mapeo entre las frecuencias ω y Ω por TB	103
3.17. $ H(\omega) $ de un FPBW con $f_0 = 1$ hz y $\Delta BW = 0.4$ hz.	115
3.18. Filtro IIR en cascada	115
3.19. Polos y ceros en el plano z	118
3.20. Diseño de filtros por ubicación de polos y ceros en el plano z	119
3.21. Respuesta del filtro FSBW diseñado	121
4.1. Predicción lineal	124
4.2. Predicción lineal sobre una señal $x(n)$	125
4.3. Predicción lineal hacia adelante	125
4.4. Error de predicción hacia adelante	126
4.5. Condición de ortogonalidad	128
4.6. Filtro blanqueador para el análisis del proceso AR	131
4.7. Proceso autorregresivo, filtro de síntesis	131
4.8. Error de predicción hacia adelante con parámetros a_i	136
4.9. Filtro de predicción de error hacia atrás	137
4.10. Error de predicción hacia atrás con parámetros h_i	138
4.11. Filtro de predicción de error hacia atrás con parámetros b_i	140
4.12. Filtros de predicción de error hacia adelante, orden $p = 1, 2, 3$	142
4.13. Relaciones de coeficientes por Levinson-Durbin	150
5.1. Filtro lattice de orden m	156
5.2. Filtro FIR de orden N	157
5.3. Filtro lattice de primer orden	158
5.4. Filtro lattice de segundo orden	158
5.5. Filtro lattice de orden m	159
5.6. Filtro lattice todo polo de orden uno	163
5.7. Filtro lattice todo polo de orden uno	164
5.8. Filtro lattice todo polo de orden dos	164
5.9. Filtro lattice todo polo de orden p	166
5.10. Filtro lattice IIR con polos y ceros	166
5.11. Filtro lattice-ladder	172
6.1. Sistema productor de voz	174
6.2. Modelo acústico del tracto vocal	175
6.3. Modelo lattice del tracto vocal para una sección L	176
6.4. Modelo de producción de voz sintética	178
6.5. Sistema general de síntesis de voz.	178
6.6. Espectro de voz verdadera y sintética	182

6.7. a) Señal de voz real. b) Señal de voz sintética	183
6.8. Comparación de espectros: a) voz verdadera y b) voz sintética	184

Índice de tablas

2.1. Atenuaciones utilizando ventana de Kaiser	28
2.2. Espectros de Ventanas	29
2.3. Desempeños de ventanas [32], [9]	29
2.4. Resumen de fórmulas de diseño de FD FIR utilizando ventanas	38
2.5. Forma compacta de Filtros FIR con fase lineal [9], [3]	42
2.6. Filtros FIR con fase lineal [9], [3], [32]	50
3.1. Polinomios de Butterworth	76
3.2. Polinomios de Chebyshev $C_n(\Omega)$	77
3.3. Polinomios de Chebyshev	84
3.4. TB y constantes para filtros FPB, FPA y FPBW	107
3.5. Transformación $H(s)$ a $H(s)$ [3], [32]	108
3.6. Transformación de un FPB $H(s)$ a $H(s)$ [3]	108
3.7. Transformación de filtros digitales en el plano z	110
4.1. Solución de ecuación normal en forma recursiva	144
4.2. Algoritmo de Levinson-Durbin	147
4.3. Recursión step down	152
6.1. Formato de salida de parámetros	182

Introducción

En la actualidad los filtros digitales están presentes en la solución a problemas de medición, control, filtrado, análisis de señales, análisis espectral, comunicaciones y en general en cualquier aplicación relacionada con procesamiento digital de señales (PDS), esto hace que un filtro digital (FD) sea un elemento fundamental en muchas aplicaciones.

Tomando en consideración que el PDS es un área de la ingeniería que se ha convertido en un campo esencial para la tecnología moderna y ha alcanzado una gran influencia en comunicaciones, procesamiento de imágenes, biomedicina, control, procesamiento de voz, radar, sonar, electrónica, etc, el autor ha tenido la intención de aportar este trabajo a la comunidad de la facultad de Ingeniería con el objeto de apoyar varias materias donde es necesario el estudio y análisis de los filtros digitales, así como recopilar metodologías de diseño que sean útiles a los ingenieros interesados en este campo.

En este trabajo se presenta una introducción general a los filtros digitales, los filtros lineales e invariantes en el tiempo de respuesta finita al impulso (FIR), de respuesta infinita al impulso (IIR), a la vez se pretende ir más allá del filtrado clásico, ya que el concepto de filtrado digital puede ser mucho más amplio, se introduce el concepto de predicción lineal, los filtros tipo lattice, ejemplos y aplicaciones.

El autor recomienda que para un mejor aprovechamiento de este material, el lector tenga conocimientos básicos de PDS, además se hace la aclaración que este trabajo tiene como objetivo analizar y profundizar en el estudio de los filtros digitales, metodologías de diseño y algoritmos, por esa razón se hace énfasis en los desarrollos matemáticos y no se recurre a herramientas de software que ya realizan los diseños sin necesidad de tener un conocimiento profundo de los filtros digitales. El utilizar algún programa de aplicación se considera como un paso posterior al haber adquirido las metodologías de cálculo y diseño, en el transcurso del trabajo sólo se hará mención de las herramientas existentes.

Este trabajo está organizado de la siguiente forma:

- En el *primer capítulo* se presenta una introducción general a los filtros digitales, sus ventajas, clasificación general y una metodología de diseño.

- En el *segundo capítulo* el diseño de los filtros de respuesta finita al impulso FIR.
- En el *tercer capítulo* las técnicas de diseño de filtros digitales con respuesta infinita al impulso IIR.
- En el *cuarto capítulo* predicción lineal utilizando filtros FIR.
- En el *quinto capítulo* se introducen los filtros “lattice” y “ladder”.
- Para finalizar en el *sexto capítulo* se ha realizado una aplicación de los filtro digitales en el proceso de síntesis de voz.

Este material es producto de la experiencia de impartir cursos y materias relacionadas con el PDS, el aporte hecho por el autor además de recopilar el material, ha sido agregar análisis y desarrollos, la realización de pruebas, diseño y comprobación de las aplicaciones. Se espera que este trabajo sirva de apoyo a la comunidad interesada en este tema, además se sugiere al lector que haga llegar al autor sus sugerencias, observaciones o correcciones necesarias para enriquecer y mejorar en un futuro este material. No existe ningún inconveniente en la reproducción total o parcial siempre que se cite la fuente.

El autor agradece a la Lic. Laura Montiel, por su valiosa colaboración en la revisión y corrección, así como al departamento de Publicaciones de la Facultad de Ingeniería, UNAM por su apoyo de esta obra.

M.I. Larry Escobar Salguero
Profesor Asociado, Facultad de Ingeniería, UNAM
Marzo de 2006
©Les
esco_lar@yahoo.com
larryesc@gmail.com

Capítulo 1

Los filtros digitales

El término *filtro* es utilizado para describir un sistema o un dispositivo que discrimina algunas características o atributos sobre las señales de entrada, es decir, que puede dejar pasar o modificar ciertas entradas y suprimir otras. Comúnmente en el filtrado clásico de señales continuas, el concepto de filtro es aplicado para dejar pasar señales con ciertas frecuencias y suprimir otras señales con otras frecuencias.

Típicamente un *filtro* describe a un sistema lineal e invariante en el tiempo el cual es utilizado para seleccionar un intervalo de frecuencias. En procesamiento digital de señales (PDS) los filtros son utilizados en múltiples aplicaciones y tienen un carácter más amplio, en general un filtro digital (FD) realiza funciones similares a un filtro analógico (FA), entre las que se pueden resaltar:

- 1.- Separar dos señales que han sido combinadas o mezcladas.
- 2.- Restaurar señales que han sido distorsionadas.
- 3.- Extraer información de interés de una señal con rasgos conocidos a priori.
- 4.- Remover ruido indeseable.
- 5.- Detección de señales.
- 6.- Compensación en frecuencia.
- 7.- Análisis espectral.
- 8.- Modelado de señales.

Si se piensa en la forma clásica de filtrado, un FD efectúa las mismas funciones que un FA, con la diferencia que éste es un algoritmo matemático implementado en hardware o software y opera sobre una señal de entrada digital para producir una señal de salida digital y está compuesto de coeficientes que caracterizan al filtro. Mientras que un FA opera sobre señales continuas y contiene elementos como resistencias, capacitores, inductores y amplificadores.

La forma más simple de expresar un FD es a través de un sistema con una señal de entrada que es procesada para producir una salida, el proceso a realizar debe cumplir con ciertas características especificadas de antemano. En esencia se puede decir que un FD es una

ecuación diferencial que debe ser evaluada por alguna máquina digital y producir resultados en el tiempo, es decir, que un FD efectúa operaciones sencillas como multiplicaciones, sumas y retardos. En la actualidad, con el avance tecnológico los FD se han convertido en una alternativa económica y flexible sobre los filtros analógicos.

Dentro de las *ventajas* principales de los filtros digitales sobre los filtros analógicos se pueden mencionar:

- Los FD permiten pendientes en la banda de transición, superiores a los filtros analógicos.
- Los FD son muy flexibles, esto implica que pueden ser programados para diferentes respuestas en frecuencia, con solo cambiar los coeficientes se puede obtener otro FD. En un filtro analógico para variar su respuesta en frecuencia, es necesario rediseñarlo y rearmarlo.
- Un FD puede ser de fase lineal. En aplicaciones como procesamiento de señales de audio y transmisión de datos, se prefieren filtros de fase lineal debido a que éstos evitan distorsiones por fase [21].
- Los FD proveen una mejor precisión debido a la tolerancia de los circuitos digitales.
- El desempeño de los FA depende de la tolerancia de los componentes.
- Los circuitos digitales son más inmunes al ruido.

Los filtros digitales también exhiben algunas *desventajas*:

- Requieren de muchos recursos para su procesamiento, sin embargo, con los avances tecnológicos actuales estos inconvenientes se han ido reduciendo.
- Banda limitada: como resultado del proceso de muestreo, el ancho de banda para las señales discretas está limitado a la mitad de la frecuencia de muestreo ($f_s/2$).
- Los efectos de precisión numérica pueden afectar su desempeño debido a la longitud finita de palabra de la arquitectura donde se implementen. En FD su desempeño está limitado por la precisión de la aritmética utilizada, en particular el desempeño de un FD es extremadamente sensible al “overflow”, el cual ocurre cuando la longitud del acumulador es insuficiente para representar los bits resultantes de sumas consecutivas.

1.1. Clasificación y técnicas de diseño de FD

1.1.1. Clasificación general

- De acuerdo a las frecuencias en bandas de paso
 - Filtro paso bajas (FPB)
 - Filtro paso altas (FPA)
 - Filtro paso banda (FPBW)
 - Filtro supresor de banda (FSBW)
- De acuerdo a la aproximación polinómica
 - Butterworth
 - Chebyshev
 - Elípticos
 - Bessel
- Lineales
 - Filtros FIR
 - Filtros IIR
 - Filtros “lattice”
 - Filtros “lattice-ladder”
- Filtros variantes en el tiempo
 - Filtros adaptables
 - Redes neuronales

1.1.2. Clasificación de acuerdo al uso e implementación de los FD

- En el dominio del tiempo

Son utilizados cuando la información está codificada en la forma de la señal y se aplican en suavizamiento, remover formas de señales, recorte de señales, reconocimiento de patrones, etc. Las especificaciones están dadas sobre señal o la respuesta al impulso.
- En el dominio de la frecuencia

Son utilizados cuando la información está contenida en la amplitud, frecuencia y fase de las componentes senoidales de la señal. El objetivo de éstos es separar bandas de frecuencia o separar señales.

- **De usuario**

Son utilizados cuando se requiere de una acción especial de filtrado, un ejemplo sería un filtro de deconvolución.

Se hace notar que un filtro puede estar dentro de varias clasificaciones, por ejemplo un filtro FPB FIR o un filtro FPBW IIR con aproximación Butterworth.

1.2. Técnicas de diseño de filtros digitales lineales

Existen varias técnicas para el diseño de filtros digitales y dependen del tipo de filtro requerido, dentro de las más utilizadas para filtros lineales están:

- Para filtros FIR

- Diseño por uso de diferentes ventanas. Aproxima la respuesta en frecuencia $H(e^{j\omega})$ del filtro a una $H_I(e^{j\omega})$ ideal utilizando Fourier.
- Diseño por muestreo en frecuencia.
- Diseño por aproximación numérica.

- Para filtros IIR

- Indirectos o transformaciones análogo-digital
 - Transformada bilineal.
 - Respuesta al impulso o escalón invariante.
 - Aproximación por diferencias finitas.
 - Transformación Z-matched.
- Directos
 - Diseño en el dominio de z .
 - Minimización del error.
 - Por fórmulas.
 - Solución numérica de la ecuación en diferencias.

1.3. Características de los FD lineales

De la teoría de sistemas, un sistema lineal e invariante en el tiempo discreto (SLITD) está totalmente definido si se conoce su respuesta al impulso $h(n)$.

- En filtros FIR

$h(n)$ es finita y está definida en el intervalo $0 \leq n \leq N - 1$.

- En filtros IIR
 $h(n)$ es infinita y definida en el intervalo $0 \leq n < \infty$.

Es decir, que en general los FD lineales se pueden implementar de dos formas:

1. Para un filtro FIR, como la convolución de una señal de entrada $x(n)$ con la respuesta al impulso $h(n)$.
2. Si el filtro es IIR, se efectúa una convolución de las salidas retardadas con los coeficientes y se suman a la convolución de la entrada con sus respectivos coeficientes.

En resumen, el objetivo del diseño de FD es el cálculo de los coeficiente $h(n)$ para un filtro tipo FIR y coeficientes $b(n)$ y $a(n)$ para un IIR que satisfagan las especificaciones del diseño. Durante todo este trabajo siempre que hablemos de filtros FIR o IIR se considera que los sistemas son del tipo SLITD y causales.

1.4. Metodología para el diseño de un filtro digital

En general, para el diseño de un FD se puede proceder considerando los siguientes pasos:

1. **Especificación:** las especificaciones de un FD dependen de la aplicación y comúnmente están dadas en frecuencia, en términos de la magnitud y fase deseada de la respuesta del filtro. Sin embargo, las especificaciones también pueden estar dadas en el tiempo o sobre alguna aplicación muy específica.
2. **Aproximación:** se utilizan conceptos matemáticos que describen un filtro para cumplir con las especificaciones dadas. De acuerdo a las especificaciones se selecciona un tipo de aproximación a un prototipo.
3. **Cálculo:** se determinan los coeficientes de un filtro causal FIR o IIR que generen las características especificadas en frecuencia.
4. **Simulación:** una vez calculados los coeficientes, sustituirlos en la respuesta en frecuencia $H(e^{j\omega})$ y verificar que se cumplan las especificaciones, de lo contrario volver al paso anterior.
5. **Implementación:** un filtro se puede describir como una ecuación en diferencias, una función de transferencia $H(z)$ o su respuesta al impulso $h(n)$. Para la implementación de un filtro digital es necesario programar la ecuación en diferencias, ya sea por software o hardware. En este punto, de acuerdo a los recursos (hardware o software) el diseñador decide la mejor estructura para la implementación, ya que no es lo mismo programar una ecuación en diferencias en una arquitectura con aritmética de punto flotante que una de punto entero.

En los capítulos siguientes se detallará la metodología a seguir en el diseño e implementación de FD.

1.4.1. Especificaciones de un FD

En teoría, un filtro ideal no es realizable, ya que tiene un respuesta infinita al impulso y es no causal. La causalidad de un sistema implica [32]:

- Que las características de la respuesta en frecuencia $|H(e^{j\omega})|$ no sea cero en un conjunto finito de puntos en frecuencia.
- Que $|H(e^{j\omega})|$ no tenga pendiente infinita de la banda de paso a la suprimida.
- Que $|H(e^{j\omega})|$ no puede ser constante en un intervalo finito de frecuencias.

Es decir, que en el diseño de un filtro realizable no es posible tener respuestas planas en la banda de paso, ni pendiente infinita en la banda de transición, ni características de fase lineal, por lo que debe existir un compromiso para que el diseño real se aproxime lo mejor posible a las especificaciones ideales.

La especificaciones en frecuencia de un filtro pueden estar dadas en dos formas, como se observa en la figura 1.1:

1. Especificaciones absolutas

Proveen los requerimientos de la magnitud de la respuesta en frecuencia:

$|H(e^{j\omega})|$ para un FPB

$0 < \omega < \omega_p$, banda de paso

$\omega_s < \omega < \pi$, banda suprimida

$\omega_p < \omega < \omega_s$, banda de transición

δ_1 , tolerancia o rizo aceptable en la banda de paso

$1 - \delta_1 \leq |H(e^{j\omega})| \leq 1 + \delta_1$, donde $0 < \omega < \omega_p$

δ_2 , tolerancia o rizo aceptable en la banda suprimida

$|H(e^{j\omega})| \leq \delta_2$, donde $\omega_s < \omega < \pi$

$\omega_c = (\omega_s + \omega_p)/2$, frecuencia de corte

$\Delta f = (\omega_s - \omega_p)/2$, banda de transición

Donde [21], [9]

$$A_p = 20 \log_{10} \frac{1 + \delta_1}{1 - \delta_1} \quad (\approx 0) \quad (1.1)$$

si δ_1 es muy pequeña, se puede aproximar

$$A_p \approx 8.686 \delta_1$$

y la atenuación en la banda de paso

$$A_s = -20 \log_{10}(\delta_2) > 0 \quad (\gg 1) \quad (1.2)$$

despejando los valores de los rizados, se pueden poner en función de las ganancias A_p y A_s en decibeles

$$\delta_1 = \frac{10^{A_p/20} - 1}{10^{A_p/20} + 1} \quad (1.3)$$

$$\delta_2 = 10^{A_s/20} \quad (1.4)$$

y el valor de los rizados en decibeles

$$R_p = -20 \log_{10} \frac{1 - \delta_1}{1 + \delta_1} \quad (\approx 0) \quad (1.5)$$

$$R_s = -20 \log_{10} \frac{\delta_2}{1 + \delta_1} \quad (>> 0) \quad (1.6)$$

2. Especificaciones relativas

Proveen los requerimientos de frecuencia en la escala de decibeles¹.

R_p : Rizo en la banda de paso $20 \log_{10} \delta_1$ en db.

R_s : Rizo en la banda suprimida $20 \log_{10} \delta_2$ en db.

A_s : Atenuación de la banda suprimida.

A_p : Ganancia en la banda de paso.

Para especificaciones relativas con requerimientos en db:

$$escala(db) = -20 \log_{10} \frac{|H(e^{j\omega})|}{|H(e^{j\omega})|_{max}} \geq 0 \quad (1.7)$$

Es importante recordar que la magnitud de la respuesta en frecuencia $|H(e^{j\omega})|$ de la figura 1.1, para un filtro digital es periódica cada 2π y simétrica respecto de π , por lo tanto el intervalo de las frecuencias de interés está entre $0 < \omega < \pi$.

$$db = 10 \log_{10} \left(\frac{P_2}{P_1} \right) = 4.342945 \lg_e \left(\frac{P_2}{P_1} \right)$$

$$db = 20 \log_{10} \left(\frac{A_2}{A_1} \right) = 8.685890 \lg_e \left(\frac{A_2}{A_1} \right)$$

¹**Decibel o Bel:** Es un concepto en honor a Graham Bell, significa que la potencia cambia por un factor de 10. Por ejemplo 3 bels de amplificación producen una señal de salida con ganancia de $10 \times 10 \times 10 = 1000$ veces la potencia de la entrada. Entonces un decibel, es la décima parte de un bel. Por tanto, los valores en decibeles: -20db, -10db, 0, 10db y 20db, significan la potencia de 0.01, 0.1, 1, 10 y 100 respectivamente. En otras palabras cada 10 decibeles significa que la potencia ha cambiado por un factor de 10. Mientras que 20db significa un factor de potencia de 100, en amplitud esto significa un factor de 10 [36]

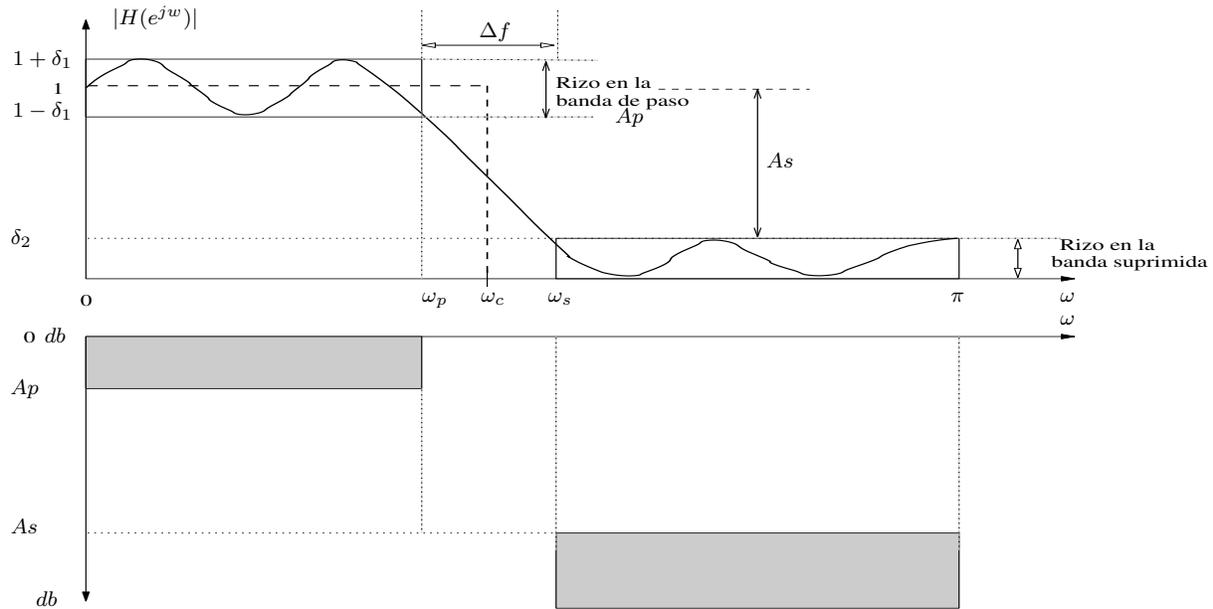


Figura 1.1: Plantilla de especificaciones de un filtro

1.5. Resumen

En este capítulo se ha dado una introducción general a la idea de los filtros digitales, la forma de clasificarlos, se ha hecho una comparación con los filtros analógicos, se explicó una metodología de diseño y las características a tomar en consideración en el diseño. En los siguientes capítulos se hará un análisis de diferentes técnicas de aproximación y diseño dependiendo del tipo de filtro.

Capítulo 2

Filtros digitales de respuesta finita al impulso

Los filtros digitales de respuesta finita al impulso (FIR) son utilizados ampliamente en el procesamiento digital de señales por su estabilidad y sus características de fase lineal que son de importancia en algunas aplicaciones como voz y audio. Debido a que la salida de este tipo de filtro sólo depende de la muestra actual de entrada y $N - 1$ retardos de la entrada, también son conocidos como no recursivos, donde N es la longitud del filtro. Por las características de estos filtros, su implementación es precisamente la operación convolución entre la respuesta al impulso del filtro y una ventana de tiempo de la señal de entrada de longitud N , por lo que su implementación en un procesador de señales digitales (DSP) es bastante eficiente.

En este capítulo se muestran las características de estos filtros, sus estructuras, los diversos métodos de diseño, ejemplos y al final se mencionan algunas funciones de Matlab que permiten un diseño rápido y fácil.

2.1. Características de los filtros FIR

Dentro de sus características más importantes se pueden mencionar:

- Son no recursivos.
- Contienen un polo múltiple en el origen de orden N .
- Se consideran sólo ceros.
- Siempre son estables.
- Son de memoria finita de longitud N , ya que no necesitan las entradas anteriores al tiempo discreto $N - 1$.

- Permiten la obtención de una respuesta con fase lineal. La fase lineal es importante en aplicaciones donde la distorsión debida a la fase no lineal puede degradar el desempeño.
- Son óptimos para tareas comunes como la reducción de ruido mientras mantienen los cambios abruptos, esto es para señales codificadas en el tiempo [36].
- Los filtros FIR no son muy eficaces para separar bandas de frecuencias.
- Los errores en la aritmética de precisión finita no son tan severos como en los filtros IIR [21].
- La cuantización del ruido debido a la precisión aritmética puede despreciarse.
- Una desventaja es que para cumplir con las especificaciones dadas se requieren muchos coeficientes en comparación con la cantidad de coeficientes utilizados en los filtros IIR.

2.2. Estructuras de los filtros FIR

La estructura de un filtro digital es la forma de seleccionar cómo se van a efectuar las operaciones matemáticas, por tanto, la selección de una estructura es de especial interés cuando los filtros se van a implementar en arquitecturas DSP, ya que se debe considerar la cantidad de memoria, el número de operaciones matemáticas, el orden de las operaciones y los efectos de la longitud finita de palabra de la arquitectura.

El *problema básico en el diseño de filtros digitales* es el cálculo de los coeficientes que aproximen las características ideales de una respuesta en frecuencia de un sistema SLITD ideal $H_I(\omega)$ con la respuesta en frecuencia de un sistema $H(\omega)$ como la ecuación (2.1)

$$H(\omega) = \frac{Y(\omega)}{X(\omega)} = \frac{B(\omega)}{A(\omega)} = \frac{\sum_{k=0}^q b_k e^{-j\omega k}}{1 + \sum_{k=1}^p a_k e^{-j\omega k}} \quad (2.1)$$

De la ecuación (2.1) un filtro FIR sólo tiene coeficientes $b_k=h_k$, es decir, que un filtro FIR tiene la característica de ser un sistema no recursivo, por lo tanto siempre es estable.

$$H(\omega) = \frac{Y(\omega)}{X(\omega)} = \sum_{k=0}^q h_k e^{-j\omega k} \quad (2.2)$$

Su diseño se lleva a cabo mediante la localización de ceros con respecto al círculo unitario en el plano z , ya que todos los polos se encuentran en el origen. Una desventaja de estos filtros es la necesidad de utilizar un mayor orden para lograr pendientes de corte pronunciadas, esto implica un mayor tiempo de procesamiento y mayor retardo en la respuesta. La respuesta al impulso de estos filtros normalmente son funciones similares a $\text{sen}(x)/x$ con

longitud finita como se verá más adelante.

Como la función de transferencia de un filtro FIR tiene la forma

$$H(z) = h_0 + h_1z^{-1} + \cdots + h_{N-1}z^{-N+1} = \sum_{i=0}^{N-1} h(i)z^{-i} \quad (2.3)$$

con respuesta al impulso

$$h(n) = \begin{cases} h_i & 0 \leq i \leq N - 1 \\ 0 & \text{otro } i \end{cases} \quad (2.4)$$

y ecuación en diferencias

$$y(n) = h_0x(n) + h_1x(n-1) + \cdots + h_{N-1}x(n-N+1) = \sum_{i=0}^{N-1} h(i)x(n-i) \quad (2.5)$$

la cual es una convolución lineal de la entrada $x(n)$ con la respuesta al impulso $h(n)$ del filtro FIR, N es la longitud del filtro y es igual al número de coeficientes del filtro.

La estructura de un FD se entiende como la forma de llevar a cabo la implementación de la ecuación en diferencias del filtro (2.5), o la función de transferencia (2.3), o cómo se efectúan las operaciones matemáticas. Dependiendo de la implementación de las ecuaciones, se pueden tener varias estructuras o formas. La mayoría de métodos para generar estructuras se basan en algoritmos específicos, por tanto, se puede decir que existen muchas estructuras equivalentes para una misma función de transferencia. A continuación se analizan las estructuras comúnmente utilizadas [24].

Forma directa

Es la estructura más sencilla y como su nombre lo indica, la ecuación (2.5) se implementa directamente utilizando los bloques básicos de un sistema discreto, para cualquier longitud N se presenta un diagrama de bloques en la figura 2.1.

Forma en cascada

La función $H(z)$ del sistema se puede descomponer en factores de segundo y primer orden dependiendo si la longitud N es par o impar

$$H(z) = h_0 + h_1z^{-1} + \cdots + h_{N-1}z^{-N+1} = \sum_{i=0}^{N-1} h(i)z^{-i} \quad (2.6)$$

$$H(z) = h_0 \left(1 + \frac{h_1}{h_0}z^{-1} + \cdots + \frac{h_{N-1}}{h_0}z^{-N+1} \right) \quad (2.7)$$

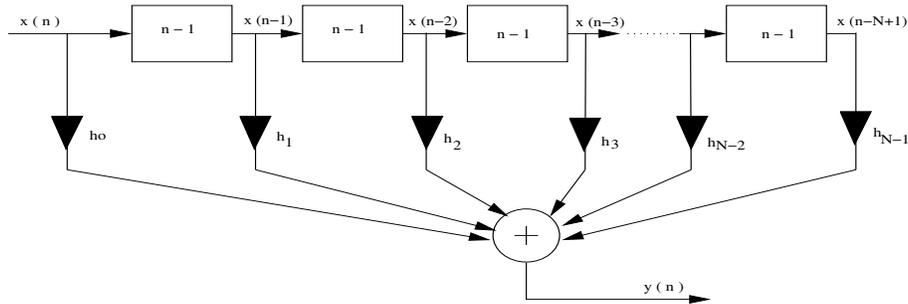


Figura 2.1: Forma directa de un filtro FIR

$$H(z) = h_0 \prod_{k=1}^K (1 + B_{k,1}z^{-1} + B_{k,2}z^{-2}) \quad (2.8)$$

Donde

$K = \frac{N-1}{2}$ si N es impar, es decir, se tienen K factores de segundo orden.

Si N es par, $K = \frac{N}{2} - 1$, se tienen K factores de segundo orden y uno de primer orden.

$B_{k,1}$ y $B_{k,2}$ son números reales que representan los coeficientes de las secciones de segundo orden.

Este tipo de estructura reduce errores de redondeo en las operaciones aritméticas que pueden ser críticos en algunas aplicaciones, sin embargo, la implementación de la forma directa es más eficiente en arquitecturas DSP [21].

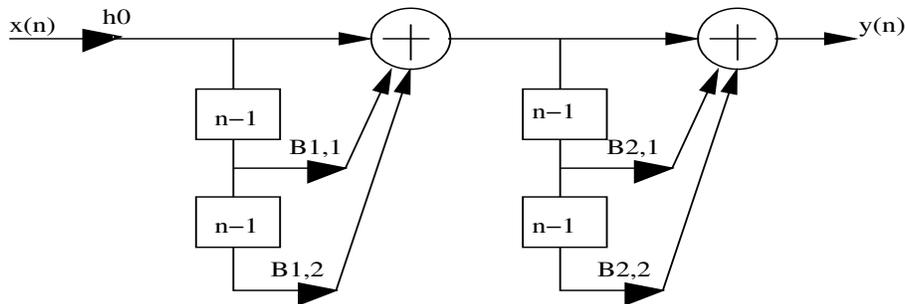


Figura 2.2: Forma cascada de un filtro FIR

Forma de fase lineal

Esta forma explota la simetría de la respuesta al impulso $h(n)$ de un filtro FIR y reduce el número de multiplicadores a la mitad. Para filtros selectivos en frecuencia, es deseable

tener una respuesta en fase lineal

$$\angle H(e^{j\omega}) = \beta - \alpha\omega; \quad -\pi < \omega < \pi \quad (2.9)$$

Para filtros causales con respuesta finita al impulso sobre el intervalo $[0, N - 1]$, la condición de fase lineal ecuación (2.9) impone condiciones de simetría en la respuesta al impulso $h(n)$

$$h(n) = h(N - 1 - n); \quad \beta = 0; \quad 0 \leq n \leq N - 1 \quad (\text{simetría}) \quad (2.10)$$

$$h(n) = -h(N - 1 - n); \quad \beta = \pm\pi/2; \quad 0 \leq n \leq N - 1 \quad (\text{antisimetría}) \quad (2.11)$$

Cuando $H(\omega)$ cambia de signo de positivo a negativo (o viceversa), la fase sufre un cambio abrupto de π radianes. Si los cambios de fase ocurren fuera de la banda de paso del filtro no tiene importancia [32].

En general para la solución simétrica con N impar:

$$y(n) = \sum_{i=0}^{\frac{N-1}{2}} h_i \left(x(n-i) + x(n-N+1+i) \right) \quad (2.12)$$

solución antisimétrica con N impar:

$$y(n) = \sum_{i=0}^{\frac{N-1}{2}} h_i \left(x(n-i) - x(n-N+1+i) \right) \quad (2.13)$$

Para la respuesta al impulso con simetría, se factorizan los coeficientes $h(n) = h(N - 1 - n)$

$$y(n) = h_0x(n) + h_1x(n-1) + h_2x(n-2) + \dots + h_{N-2}x(n-N+2) + h_{N-1}x(n-N+1) \quad (2.14)$$

$$y(n) = h_0 \left(x(n) + x(n-N+1) \right) + h_1 \left(x(n-1) + x(n-N+2) \right) + \dots \quad (2.15)$$

Como se observa de la ecuación (2.15) y la figura 2.3, en esta estructura se requiere 50% menos multiplicadores que en la forma directa, es decir, que se utilizan sólo la mitad de los coeficientes, lo que puede ser útil en el ahorro de memoria en la implementación en una arquitectura DSP. En la actualidad existen arquitecturas de DSPs que implementan este tipo de estructuras eficientemente en ciclos de repetición [40].

2.3. Diseño de filtros FIR

Dada una señal discreta de entrada $x(n)$, donde $x(n)$ representa la muestra actual y $x(n-i)$ un conjunto de muestras pasadas, $h(n)$ es una secuencia de coeficientes de la respuesta al impulso del filtro $h(0), h(1), \dots, h(N - 1)$, y $y(n)$ es la salida del filtro, como se explicó en

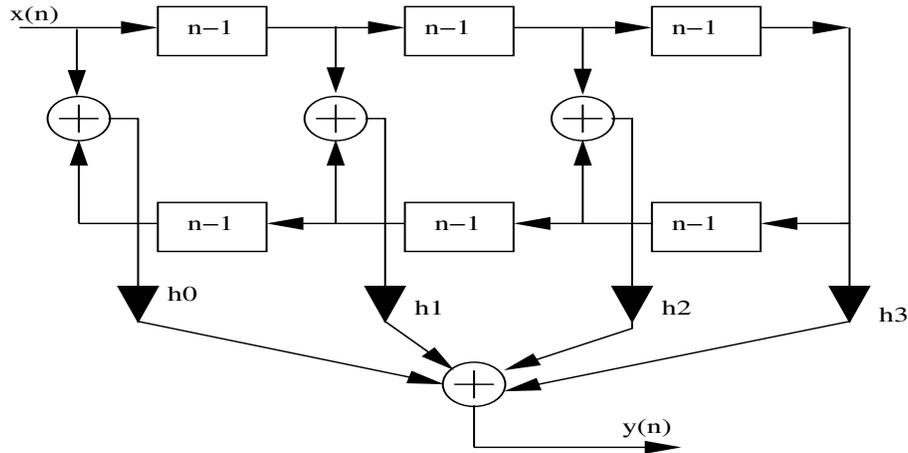


Figura 2.3: Forma fase lineal de un filtro FIR

la forma directa. De la teoría del PDS, la salida $y(n)$ de un filtro FIR se puede representar matemáticamente como la convolución¹ de la entrada $x(n)$ y $h(n)$

$$y(n) = \sum_{i=0}^{N-1} h(i)x(n-i) \quad (2.16)$$

Utilizando la transformada Z (TZ), se obtiene $Y(z)$

$$Y(z) = \sum_{i=0}^{N-1} h(i)TZ(x(n-i)) = \sum_{i=0}^{N-1} h(i)X(z)z^{-i} \quad (2.17)$$

donde se define la función de transferencia del filtro $H(z)$ como el cociente $Y(z)/X(z)$

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{i=0}^{N-1} h(i)z^{-i} \quad (2.18)$$

Con respuesta en frecuencia $H(e^{j\omega}) = H(z)|_{z=e^{j\omega}}$

$$H(e^{j\omega}) = \sum_{n=0}^{N-1} h(n)e^{-jn\omega} \quad (2.19)$$

en forma polar

$$H(e^{j\omega}) = |H(e^{j\omega})|e^{j\theta(\omega)} \quad (2.20)$$

¹Para la demostración de este teorema ver apéndice C.

Por conveniencia y facilidad se considera $H(e^{j\omega}) \approx H(\omega)$, aunque siendo estrictos en la nomenclatura, la respuesta en frecuencia $H(\omega)$ corresponde a un sistema continuo y la respuesta en frecuencia $H(e^{j\omega})$ a un sistema discreto.

La ecuación $H(e^{j\omega})$ también representa la transformada de Fourier de tiempo discreto de $h(n)$ la cual es periódica en 2π . Una vez desarrollada la respuesta en frecuencia $H(\omega)$ en términos de los coeficientes $h(n)$, se debe graficar y verificar si efectivamente corresponde a las especificaciones establecidas en frecuencia.

La fase es una característica importante en el diseño de un filtro, porque afecta los tiempos de retardo de las diferentes componentes de frecuencia que pasan a través del filtro. Por lo tanto, para especificaciones de un filtro con respuesta de fase lineal se requiere que $\theta(\omega)$ cumpla con la condición

$$\theta(\omega) = \beta - \alpha\omega, \quad -\pi \leq \omega \leq \pi$$

donde α es una constante que representa el retraso en número de muestras que sufre la secuencia al procesarla. Únicamente los filtros digitales FIR pueden ser diseñados con fase lineal, mientras que los filtros IIR no pueden tener fase lineal [32], en algunas aplicaciones no se admiten distorsiones por fase, entonces los filtros FIR son los más adecuados.

Debe enfatizarse que cuando la magnitud de la respuesta en frecuencia del filtro se vuelve negativa, se debe agregar π radianes a la función de respuesta en fase $\theta(\omega)$ [3].

Para el análisis de fase se define el *retraso de fase* τ_p

$$\tau_p = -\frac{\theta(\omega)}{\omega}$$

y el *retraso de grupo* τ_g

$$\tau_g = -\frac{d\theta(\omega)}{d\omega}$$

si τ_p y τ_g son constantes, entonces los filtros son de fase lineal [3]. Significa que los filtros de fase lineal evitan la distorsión por fase, ya que todas las componentes senoidales son retrasadas en la misma cantidad [21].

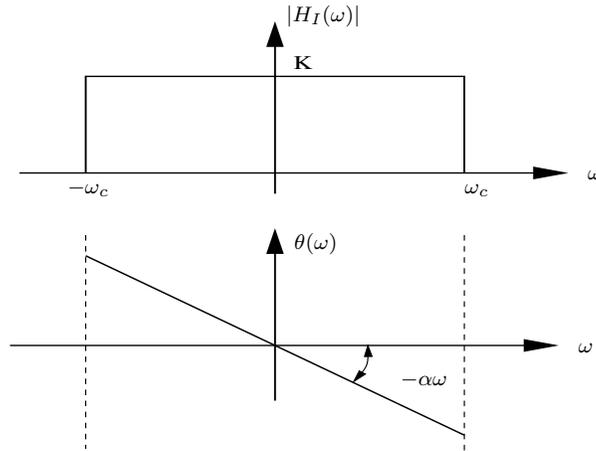
2.3.1. Aproximación de un FD FIR a un filtro ideal PB

En este procedimiento la respuesta en frecuencia $H(e^{j\omega})$ de un filtro FIR se aproxima a una respuesta en frecuencia ideal $H_I(e^{j\omega})$ como se observa en la figura 2.4, con representación matemática

$$H_I(e^{j\omega}) = \begin{cases} Ke^{-j\alpha\omega} & n = 1, 2, 3, \dots, N - 1 \\ 0 & \text{otro } n \end{cases} \quad (2.21)$$

igualando la ecuación (2.19) con la respuesta en frecuencia de un filtro ideal

$$|H(e^{j\omega})|e^{j\theta(\omega)} = Ke^{-j\alpha\omega} \quad (2.22)$$


 Figura 2.4: $H_I(\omega)$ de un filtro paso bajas ideal

desarrollando los términos exponenciales por Euler en ambos lados y si la ganancia $K = 1$

$$\sum_{n=0}^{N-1} h(n)(\cos \omega n - j \operatorname{sen} \omega n) = \cos \omega \alpha - j \operatorname{sen} \omega \alpha \quad (2.23)$$

igualando partes reales e imaginarias

$$\sum_{n=0}^{N-1} h(n) \cos \omega n = \cos \omega \alpha \quad (2.24)$$

$$\sum_{n=0}^{N-1} h(n) \operatorname{sen} \omega n = \operatorname{sen} \omega \alpha \quad (2.25)$$

dividiendo la ecuación (2.25) entre la ecuación (2.24) y separando términos

$$\sum_{n=0}^{N-1} h(n) \operatorname{sen} \omega n \cos \omega \alpha = \sum_{n=0}^{N-1} h(n) \cos \omega n \operatorname{sen} \omega \alpha \quad (2.26)$$

se tiene la sumatoria

$$\sum_{n=0}^{N-1} h(n) (\cos \omega n \operatorname{sen} \omega \alpha - \operatorname{sen} \omega n \cos \omega \alpha) = 0 \quad (2.27)$$

por identidad trigonométrica $\operatorname{sen}(A \pm B) = \operatorname{sen}(A)\cos(B) \pm \cos(A)\operatorname{sen}(B)$

$$\sum_{n=0}^{N-1} h(n) \operatorname{sen}(\omega \alpha - \omega n) = 0 \quad (2.28)$$

esta última ecuación tiene solución si $h(n)$ satisface

$$h(n) = h(N - 1 - n) \quad (\text{simetría}) \quad (2.29)$$

$$\alpha = (N - 1)/2 \quad (\text{causalidad}) \quad (2.30)$$

Es decir, que un filtro con fase lineal presenta una respuesta al impulso simétrica, real y causal.

Demostración de la solución

Sustituyendo las soluciones en la ecuación (2.28)

$$\sum_{n=0}^{N-1} h(n) \operatorname{sen} \left(\omega \frac{N-1}{2} - \omega n \right) = 0 \quad (2.31)$$

Desarrollando la sumatoria

$$h(0) \operatorname{sen} \left(\omega \frac{N-1}{2} \right) + h(1) \operatorname{sen} \left(\omega \frac{N-1}{2} - \omega \right) + \dots + h(N-1) \operatorname{sen} \left(\omega \frac{N-1}{2} - \omega(N-1) \right) \quad (2.32)$$

Por la condición de la solución que $h(n) = h(N - n - 1)$ para todo $0 \leq n \leq N - 1$

$$h(0) \left(\operatorname{sen} \left(\omega \frac{N-1}{2} \right) + \operatorname{sen} \left(-\omega \frac{N-1}{2} \right) \right) + h(1) \left(\operatorname{sen} \left(\omega \frac{N-3}{2} \right) + \operatorname{sen} \left(-\omega \frac{N-3}{2} \right) \right) + \dots = 0 \quad (2.33)$$

Como la función seno es impar, entonces la ecuación anterior es cero, con lo que se demuestra que la solución descrita en las ecuaciones (2.29) y (2.30) es válida.

Los coeficientes $h(n)$ son obtenidos por la teoría de Fourier, y para una $H_I(\omega)$ son infinitos, sin embargo, para su manejo digital se han truncado a N términos, lo que hace que la respuesta $H(\omega)$ no sea exactamente igual a la ideal. Para hacer causal la respuesta al impulso $h(n)$ se le aplica un retardo $\alpha = \frac{(N-1)}{2}$, esta modificación no afecta la amplitud de la respuesta del filtro sino la fase, y el truncamiento abrupto de la serie de Fourier ocasiona oscilaciones en la banda de paso y en la banda suprimida. Estas oscilaciones son debido a la convergencia lenta de la serie de Fourier, particularmente cercana a los puntos de discontinuidad, este efecto es llamado *fenómeno de Gibbs*. Si la longitud de N es incrementada, la amplitud de las oscilaciones decrece y la amplitud del rizo cerca a las discontinuidades tiende a permanecer sin cambios [3] (ver figura 2.10).

Ejemplo:

De la demostración anterior, si $N = 11$, entonces $\alpha = 5$, se comprueba que la ecuación (2.33) se cumple:

$$h(0) \sen(5\omega) + h(1) \sen(4\omega) + h(2) \sen(3\omega) + h(3) \sen(2\omega) + h(4) \sen(\omega) + h(5) \sen(0) + h(6) \sen(-\omega) + h(7) \sen(-2\omega) + h(8) \sen(-3\omega) + h(9) \sen(-4\omega) + h(10) \sen(-5\omega) = 0$$

esta ecuación es igual a cero por la condición de simetría de los coeficientes $h(n)$ y la antisimetría de la función seno. Es decir, $h(0) = h(5)$ y $\sen(-5\omega) = -\sen(5\omega)$ y así sucesivamente.

Filtro FIR promediador

Si se considera una señal en el tiempo con ruido agregado, entonces una manera simple de eliminar el ruido o suavizar la señal, sería ir calculando el promedio de las muestras sobre una ventana de tiempo y recorrer la ventana. Es decir, que la $h(n)$ propuesta es $1/N$ en toda la ventana y cumple con la condición de simetría propuesta en la ecuación (2.29).

Este tipo de filtro también es conocido como *Moving average (MA)*, ya que realiza el promedio sobre la muestra actual de entrada $x(n)$ y $N - 1$ muestras pasadas. Si se visualiza una señal a baja frecuencia con ruido agregado (figura 2.5), una forma fácil de suavizarla (eliminar el ruido) es ir tomando promedios de la señal en un intervalo de tiempo (una ventana de longitud N). Aplicando promedios a una señal de entrada $x(n)$ en una ventana de tiempo de longitud N se puede obtener la salida $y(n)$

$$y(n) = \sum_{i=0}^{N-1} \frac{1}{N} x(n-i) \quad (2.34)$$

si $h(n) = 1/N$ se tiene un filtro que realiza promedios sobre N muestras, aplicando transformada Z

$$Y(z) = \frac{1}{N} \sum_{i=0}^{N-1} X(z)z^{-i} \quad (2.35)$$

Entre los filtros digitales, los filtros MA producen el menor ruido para bordes muy agudos, la cantidad de reducción de ruido es igual a la raíz cuadrada del número de puntos promediados, por ejemplo si $N = 100$, un filtro FIR MA reduce el ruido por un factor de 10. La función de transferencia es

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{N} \frac{1 - z^{-N}}{1 - z^{-1}} \quad (2.36)$$

y si $z = e^{j\omega}$ su respuesta en frecuencia es

$$H(\omega) = \frac{\sen(\pi\omega N/2)}{N \sen(\pi\omega/2)} e^{-j(N-1)\omega/2} \quad (2.37)$$

Debido a la respuesta en frecuencia, este tipo de filtros no puede separar eficientemente una frecuencia de otra [36]. En la figura 2.5 se muestra una señal senoidal con ruido agregado y el resultado de utilizar un filtro FIR promediador con $N=30$ y $N=100$. Para los filtros FIR

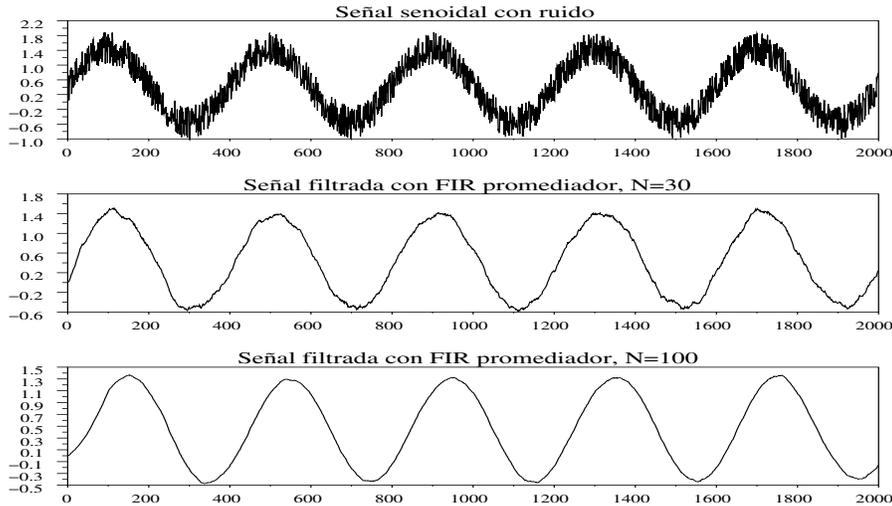


Figura 2.5: Resultado de aplicar un filtro FIR promediador

y de la figura 2.5 se observa que las primeras $N - 1$ primeras salidas corresponden al retardo del filtro o el comportamiento transitorio, y para tiempos $n > N - 1$ el buffer de entrada está completamente lleno y el filtro opera en su estado estable.

2.3.2. Soluciones de un filtro FIR de fase lineal

Algunas aplicaciones prácticas como voz, audio, video, etc., son sensibles a la distorsión por fase, por lo que pueden requerir del diseño de filtros con fase lineal. De la solución general para $h(n)$ simétrica, si N es impar $\alpha = (N - 1)/2$ y definiendo $h(n) = b(n + \alpha)$, entonces en el dominio de z

$$B(z) = \sum_{n=0}^{2\alpha} b(n)z^{-n} = \sum_{n=-\alpha}^{\alpha} b(n + \alpha)z^{-n+\alpha}$$

reacomodando términos

$$B(z) = z^{-\alpha} \sum_{n=0}^{2\alpha} h(n)z^{-n} = z^{-\alpha} H(z) \quad (2.38)$$

por propiedades de simetría $h(n) = h(-i)$ para $n = 0, 1, 2, 3, 4, \dots, \alpha$

$$B(z) = z^{-\alpha} \left[h(0) + \sum_{n=1}^{\alpha} h(n)(z^n + z^{-n}) \right] \quad (2.39)$$

evaluando para $z = e^{j\omega}$

$$B(\omega) = e^{-j\alpha\omega} \left[h(0) + 2 \sum_{i=1}^{\alpha} h(i) \cos(i\omega) \right] \quad (2.40)$$

Si $h(n)$ es real, entonces $H(\omega)$ es una función de ω , si $H(\omega) > 0$, entonces la fase de $B(\omega)$ es

$$\phi(\omega) = -\alpha\omega \quad (2.41)$$

siendo una función lineal de ω . Sin embargo, si $H(\omega) < 0$, la fase de $B(\omega)$ es $\pi - \alpha\omega$, es decir, que existe un cambio de signo de $H(\omega)$ produciendo un corrimiento de fase de 180° de $B(\omega)$. Cuando $h(n)$ es real, entonces $|H(\omega)|$ es simétrica y $\phi(\omega)$ es antisimétrica. También se puede comprobar que si z_0 es un cero de un filtro FIR de fase lineal, entonces $1/z_0$ también es un cero de $H(z)$.

La solución general para $h(n)$ antisimétrica y N es impar para $n = 0, 1, 2, 3, \dots, \alpha$, además $h(0) = 0$. Siguiendo el mismo procedimiento que para el caso simétrico

$$B(z) = z^{-\alpha} \sum_{n=0}^{2\alpha} h(n)z^{-n} = z^{-\alpha} H(z) \quad (2.42)$$

$$B(z) = z^{-\alpha} \left[h(0) + \sum_{n=1}^{\alpha} h(n)(z^n - z^{-n}) \right] \quad (2.43)$$

evaluando para $z = e^{j\omega}$

$$B(\omega) = e^{-j\alpha\omega} \left[-2j \sum_{n=1}^{\alpha} h(n) \sen(n\omega) \right] \quad (2.44)$$

si $h(n)$ es real, entonces $H(\omega)$ es puramente imaginaria y la fase de $B(\omega)$ es una función lineal de ω . Existen cuatro posibilidades de diseño de los filtros de fase lineal dependiendo si la longitud N es par o impar, si $h(n)$ es simétrica o antisimétrica, estos casos son estudiados en este capítulo.

2.4. Diseño de filtros FIR por series de Fourier y ventanas

Lo atractivo de este método es su simplicidad y flexibilidad para el cálculo de los coeficientes, sin embargo, no permite al diseñador un control adecuado sobre las especificaciones del diseño. La idea básica del método de Fourier es el diseño de un filtro FIR que aproxime la respuesta en frecuencia a una respuesta en frecuencia deseada por el cálculo de su respuesta al impulso. Además no existe una regla bien definida para el cálculo de la longitud N .

Como la respuesta en frecuencia de un filtro digital es continua y periódica, y la respuesta en frecuencia ideal $H_I(e^{j\omega})$ se puede representar por la serie exponencial de Fourier [12], [38]

$$H(\omega) = \sum_{n=-\infty}^{\infty} h(n)e^{-jn\omega} \quad (2.45)$$

los coeficientes $h(n)$ se calculan en un período de 2π como

$$h(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(\omega)e^{j\omega n} d\omega \quad (2.46)$$

Para una respuesta ideal $|H(\omega)| = K = Constante = 1$ en el intervalo $[-\omega_c, \omega_c]$

$$h(n) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{-j\omega n} d\omega \quad (2.47)$$

$$h_{FPB-NC}(n) = \frac{e^{j\omega_c n} - e^{-j\omega_c n}}{2j\pi n} = \frac{\text{sen}(\omega_{CD}n)}{\pi n} \quad (2.48)$$

donde

$$\omega_{CD} = \frac{2\pi f_{ca}}{f_s} \quad (2.49)$$

- ω_{CD} : frecuencia de corte digital en función de frecuencias analógicas.
- f_{ca} : frecuencia de corte analógica, que viene de las especificaciones analógicas.
- f_s : frecuencia de muestreo.
- NC : no causal.
- $h_I(n)$: es la respuesta (infinita) al impulso del filtro, y representa a los coeficientes de la serie de Fourier.
- $H(e^{j\omega})$: es la respuesta ideal o aproximada para el filtro a diseñar.
- $F = \frac{f}{f_s}$: frecuencia digital normalizada, $-1/2 \leq F \leq 1/2$ y $F = fT$
- f : cualquier frecuencia analógica en Hertz.

Como la $h(n)$ calculada es infinita y no causal, para procesar digitalmente al filtro, se necesita que la $h(n)$ sea de longitud finita N y causal. Esto es equivalente a la multiplicación de $h(n)$ por una ventana rectangular $W_R(n)$ de longitud finita N , es decir, $h(n) = h_I(n)W_R(n)$, donde

$$W_R(n) = \begin{cases} 1 & n = 1, 2, 3, \dots, N - 1 \\ 0 & \text{otro } n \end{cases} \quad (2.50)$$

Para que $h(n)$ sea causal es necesario retrasarla en un número de muestras igual a α . Espectralmente la respuesta en frecuencia del filtro $h(n)$ calculado es la convolución de $H_I(\omega)$ con $W_R(\omega)$. Como se verá posteriormente, el tipo de ventana elegido tiene una influencia muy importante en la respuesta espectral del filtro.

2.4.1. Metodología de diseño de filtros FIR utilizando ventanas

- Dar las especificaciones de diseño
- Seleccionar la respuesta en frecuencia deseada $H_I(e^{j\omega})$.
- Calcular la $h_I(n)$ correspondiente utilizando fórmulas de diseño.
- Como la $h_I(n)$ de la $H_I(e^{j\omega})$ es una respuesta infinita al impulso, es necesario truncarla en N puntos para generar un filtro práctico.
- Multiplicar $h_I(n)$ por una ventana elegida de longitud N , con esto se logra una mejor respuesta espectral.
- La $h(n)$ obtenida es no causal (NC), ya que existen valores de $h(n)$ para $n < 0$, entonces es necesario desplazarla a la derecha α puntos para obtener una versión causal.
- El efecto de recortar $h_I(n)$, ocasiona que su respuesta en frecuencia sea muy diferente a la ideal.
- Graficar $|H(\omega)|$, si no cumple con los requerimientos incrementar N o cambiar la ventana y recalcular los coeficientes.

2.4.2. Efecto de una ventana rectangular

El truncamiento de la secuencia $h(n)$ hasta N puntos puede entenderse como la multiplicación de la secuencia $h(n)$ por una ventana cuadrada de amplitud uno, que solo permitirá conservar aquellos elementos que estén dentro de la ventana

$$h_R(n) = h(n)W_R(n)$$

donde $W_R(n)$ es la función de la ventana rectangular y su utilización significa que en los extremos $h_R(n)$ cambia abruptamente a cero.

En el dominio de la frecuencia $H_R(\omega)$ corresponde a la convolución de la respuesta en frecuencia ideal $H_I(\omega)$ con la respuesta en frecuencia de la ventana $W_R(\omega)$ como se ilustra en las figuras 2.6 y 2.7.

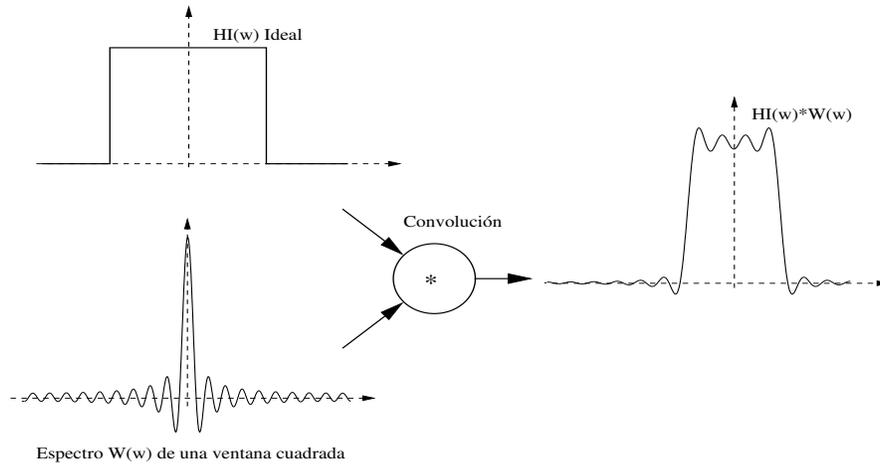


Figura 2.6: Efecto de convolucionar el espectro de una ventana cuadrada con $H_I(\omega)$

$$H_R(e^{j\omega}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(e^{j\theta})W_R(e^{j(\omega-\theta)})d\theta \quad (2.51)$$

Calculando la transformada de Fourier inversa se obtiene la respuesta real que corresponde a los coeficientes $h_R(n)$.

En la figura 2.7 se observa que:

- Debido a que la ventana $W_R(n)$ tiene una longitud finita igual a N , su respuesta en frecuencia contiene un lóbulo principal cuyo ancho es proporcional a $1/N$.
- La convolución de $W_R(\omega)$ y la respuesta en frecuencia ideal $H_I(\omega)$, produce una versión $H_R(\omega)$ real con rizo en las bandas de paso y suprimida.
- El lóbulo principal de la ventana $W_R(\omega)$ es el responsable del ancho de la banda de transición de $H(\omega)$, entre más ancho sea este lóbulo más ancha será la banda de transición.
- Los lóbulos laterales producen rizados similares en las bandas de paso y suprimida.

De los análisis anteriores se puede concluir que para evitar los efectos ocasionados por una ventana cuadrada, es necesario que $N \rightarrow \infty$ entonces $W_R(\omega)$ se aproxima a un impulso, esto implica que su lóbulo principal sea muy angosto y alto comparado con los lóbulos laterales, sin embargo, los lóbulos laterales aumentan en número y siguen manteniendo una amplitud

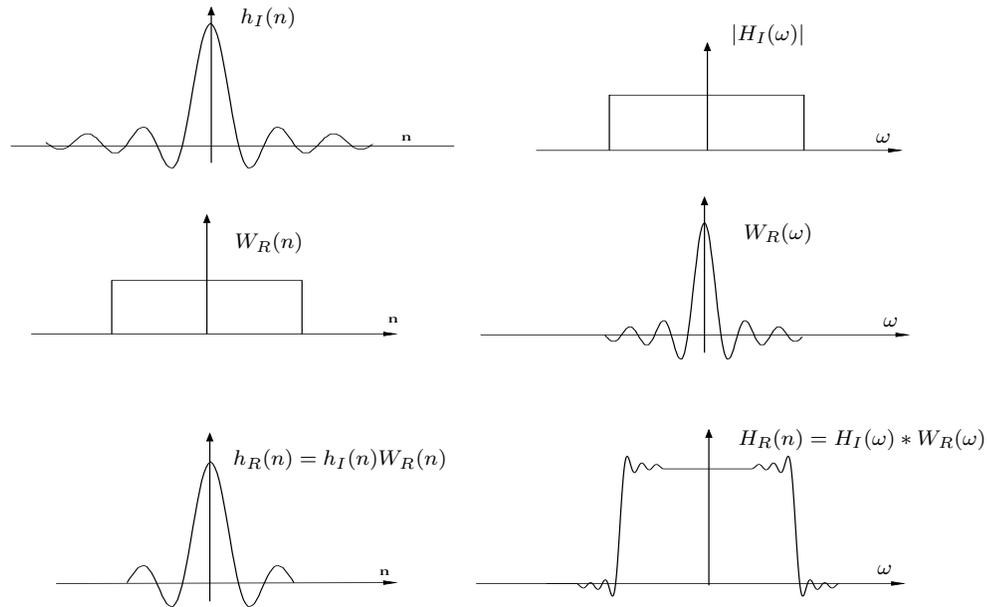


Figura 2.7: Efecto de una ventana rectangular al truncar $h(n)$

significativa respecto del principal, esto ocasiona que al efectuar la convolución en frecuencia todavía se presenten rizos.

Si la longitud de una ventana rectangular se incrementa, el número de rizos en la banda de paso y en la banda suprimida se incrementa y el ancho de los rizos decrece, pero su altura permanece casi constante. Los rizos mayores ocurren cerca de la zona de transición y su amplitud es independiente de N , este efecto es llamado *fenómeno de Gibbs* [2], [17]. Conforme se incrementa N el ancho del lóbulo principal decrece, la altura de los lóbulos laterales crece y el área bajo cada lóbulo permanece constante.

Para lograr requerimientos aceptables de la ventana a utilizar, se han propuesto otras ventanas con caídas mas suaves en los extremos y en frecuencia el lóbulo principal sea angosto y muy superior a los lóbulos laterales, esto garantiza una transición abrupta de la banda de paso a la banda suprimida y la eliminación de lóbulos en la banda de paso y suprimida [18], [21], [32].

2.4.3. Tipos de ventanas

Como se comentaba anteriormente, dependiendo de la ventana que se utilice, el filtro diseñado tendrá un mejor comportamiento espectral, a continuación se describen las ventanas más utilizadas²:

²Un tutorial amplio sobre una gran cantidad de ventanas se puede encontrar en [18].

- **Ventana rectangular:** La ventana rectangular es la que presenta el lóbulo principal más angosto, sin embargo, el primer lóbulo lateral se encuentra a sólo 13 db por debajo del lóbulo principal con pendiente de -6 db por octava [3]. La ventana rectangular presenta transiciones más abruptas, ya que para la respuesta en frecuencia la duración de las transiciones dependen del ancho del lóbulo principal. Desafortunadamente también provoca oscilaciones pronunciadas en las regiones paso banda y rechazo banda.

$$W(n) = \begin{cases} 1 & 0 \leq n \leq N - 1 \\ 0 & \text{otro } n \end{cases} \quad (2.52)$$

- **Ventana triangular o Bartlett:** Se caracteriza por ser un triángulo simétrico centrado en las muestras y en los extremos la caída a cero es suave. El primer lóbulo lateral se encuentra a -26 db por debajo del lóbulo principal.

$$W(n) = \begin{cases} \frac{2n}{N-1} & 0 \leq n \leq \frac{N-1}{2} \\ 2 - \frac{2n}{N-1} & \frac{N-1}{2} < n \leq N - 1 \\ 0 & \text{otro } n \end{cases} \quad (2.53)$$

- **Ventana de Hanning:** Para este caso se tiene la forma general:

$$W(n) = \alpha - (1 - \alpha)\cos\left(\frac{2\pi n}{N-1}\right) \quad 0 \leq n \leq N - 1$$

si $\alpha=0.5$

$$W(n) = \begin{cases} 0.5 - 0.5\cos\left(\frac{2\pi n}{N-1}\right) & 0 \leq n \leq N - 1 \\ 0 & \text{otro } n \end{cases} \quad (2.54)$$

El primer lóbulo lateral se encuentra a -31 db por debajo del lóbulo principal y el ancho del lóbulo principal es el doble del ancho del lóbulo principal de la ventana cuadrada (ver figura 2.8).

- **Ventana de Hamming:** En esta ventana, su respuesta en frecuencia presenta lóbulos laterales menores, el primer lóbulo lateral se encuentra a 41 db por abajo del lóbulo principal con pendiente de -6 db por octava [3], pero el lóbulo principal es por lo menos dos veces más ancho que el de la ventana cuadrada. A pesar de obtener un número menor de oscilaciones en las regiones pasa banda y rechaza banda de $H_w(\omega)$, su banda de transición se vuelve menos abrupta y más ancha. La ventana de Hamming es muy buena elección en muchas aplicaciones prácticas.

$$W(n) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) & 0 \leq n \leq N - 1 \\ 0 & \text{otro } n \end{cases} \quad (2.55)$$

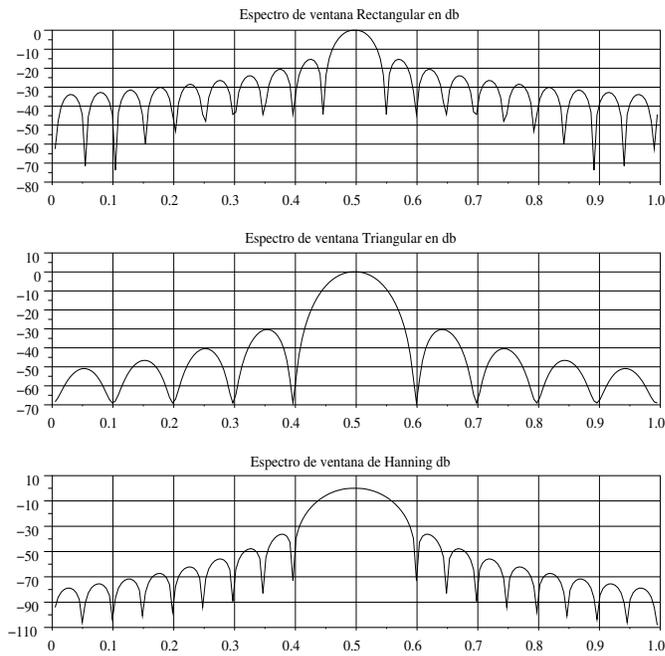


Figura 2.8: Comparación de ventanas: rectangular, triangular y Hanning

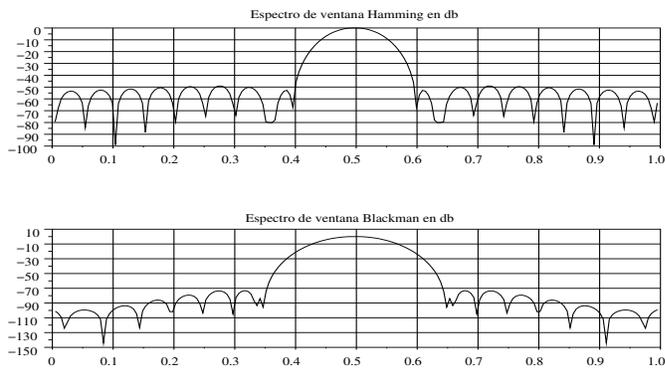


Figura 2.9: Comparación de ventanas: de Hamming y Blackman

- **Ventana de Blackman:** En este caso el primer lóbulo lateral se encuentra a -57 db por debajo del lóbulo principal.

$$W(n) = \begin{cases} 0.42 - 0.5\cos\left(\frac{2\pi n}{N-1}\right) + 0.08\cos\left(\frac{4\pi n}{N-1}\right) & 0 \leq n \leq N-1 \\ 0 & \text{otro } n \end{cases} \quad (2.56)$$

En esta ventana, el segundo término coseno tiene el efecto de incrementar en un 50 % el ancho del lóbulo principal (figura 2.9), pero a la vez mejora la amplitud de este lóbulo, sin embargo, el hecho de que el lóbulo principal se haga más ancho ocasiona que la banda de transición crezca 6 veces más respecto la ventana rectangular.

En las ventanas anteriores el ancho del lóbulo principal es inversamente proporcional a N , es decir, que un aumento en el ancho de la ventana decrece el ancho del lóbulo principal ocasionando un decremento en la banda de transición. Por otro lado, los mínimos en la banda suprimida son independientes del valor de N y sólo dependen de la ventana seleccionada [3].

- **Ventana de Kaiser:** Kaiser (1974) plantea el uso de una secuencia de ventanas, utilizando la función Bessel de orden cero modificada, que ajusta la amplitud de los lóbulos laterales con respecto al ancho del lóbulo principal mediante la modificación de la longitud de la ventana y su forma. Esta ventana tiene dos parámetros a manejar: la longitud N y el parámetro de forma β , para $\beta = 0$ la ventana de Kaiser se reduce a la ventana rectangular. Kaiser desarrolló expresiones matemáticas que permiten obtener de antemano el valor de N y de β para lograr las especificaciones deseadas de la respuesta en frecuencia.

$$W(n) = \begin{cases} \frac{I_0\left[\beta\sqrt{1-\left(1-\frac{2n}{N-1}\right)^2}\right]}{I_0(\beta)} & 0 \leq n \leq N-1 \\ 0 & \text{otro } n \end{cases} \quad (2.57)$$

Esta ventana cumple con $W(0) = W(N-1) = 1/I_0(\beta)$, es simétrica respecto del centro y es óptima en el sentido de tener la máxima energía en el lóbulo principal.

La función modificada de Bessel de orden cero se puede calcular

$$I_0(x) = 1 + \sum_{k=1}^{\infty} \left(\frac{(x/2)^k}{k!}\right)^2$$

y la constante β está dada de acuerdo a la tolerancia al rizo en la banda de paso, utilizando la fórmulas empíricas [3]

$$\beta = \begin{cases} 0.1102(A_s - 8.7) & A_s > 50 \\ 0.5842(A_s - 21)^{0.4} + 0.07886(A_s - 21) & 21 \leq A_s \leq 50 \\ 0.0 & A_s < 21 \end{cases} \quad (2.58)$$

y la constante $A_s = -20\log_{10}(\delta_2)$.

Cuando se utiliza una ventana diferente a la cuadrada, se calculan los coeficientes de la ventana $W(n)$ causal para $0 \leq n \leq N - 1$ y se multiplican punto a punto con $h(n)$ causal, para obtener los nuevos coeficientes afectados por la ventana.

La ventana de Kaiser es muy cercana a un diseño óptimo en el sentido de concentrar la mayor energía en el lóbulo principal, aunque el cálculo es más complicado. De lo anterior se concluye que el espectro de una buena ventana debe contener un lóbulo principal grande y angosto que concentre la máxima energía y lóbulos laterales muy pequeños. El ancho del lóbulo principal es definido como dos veces el primer cruce por cero de $W(\omega)$.

Las características generales de las ventanas mencionadas se han resumido en las tablas 2.1, 2.2, 2.3, además en el apéndice D se puede encontrar un desarrollo para obtener el espectro de la ventana de Hamming.

Mínima atenuación en la Banda suprimida (db)	β
25.0	1.333
30.0	2.117
40.0	3.395
50.0	4.551
60.0	5.653
70.0	6.755
80.0	7.857
90.0	8.959
100.0	10.061

Tabla 2.1: Atenuaciones utilizando ventana de Kaiser

Fórmulas para el cálculo de N

Estas fórmulas dan un valor aproximado utilizando las especificaciones de diseño [9], [32], [3], [21]. Una práctica usual de diseño es calcular $\delta = \min(\delta_1, \delta_2)$ para tener un rizo igual en las bandas de paso y suprimida.

Fórmula de Kaiser

$$N - 1 = \frac{A_s - 7.95}{14.36\Delta\omega} + 1 \tag{2.59}$$

Ventana	Espectro
Rectangular	$S(\omega) = \frac{\text{sen}(\omega N/2)}{\text{sen}(\omega/2)}$
Triangular	$(S(\omega))^2$
Hanning	$0.5S(\omega) + 0.25S(\omega - 2\pi/N) + 0.25S(\omega + 2\pi/N)$
Hamming	$0.54S(\omega) + 0.23S(\omega - 2\pi/N) + 0.23S(\omega + 2\pi/N)$
Blackman	$0.42S(\omega) + 0.25S(\omega - 2\pi/N) + 0.25S(\omega + 2\pi/N) + 0.04S(\omega - 4\pi/N) + 0.04S(\omega + 4\pi/N)$

Tabla 2.2: Espectros de Ventanas

Ventana	Ancho del lóbulo principal de $W(\omega)$	Lóbulo principal a 2do. lóbulo	A_s mínima
Rectangular	$4\pi/N$	-13	-21
Triangular	$8\pi/N$	-26	-25
Hanning	$8\pi/N$	-31	-44
Hamming	$8\pi/N$	-41	-53
Blackman	$12\pi/N$	-57	-74

Tabla 2.3: Desempeños de ventanas [32], [9]

donde

$$A_s = -20 \log_{10}(\delta)$$

Fórmula de Herrmann

$$N = \frac{-20 \log_{10} \sqrt{\delta_1 \delta_2} - 13}{14.6 \Delta f} + 1 \quad (2.60)$$

Fórmula de Herrmann más precisa para el orden N [32]

$$N = \frac{D - F(\Delta f)^2}{\Delta \omega} + 1 \quad (2.61)$$

donde

$$\Delta \omega = \frac{\omega_{sup} - \omega_c}{2\pi} < 1$$

$$D = [a_1(\log_{10}\delta_1)^2 + a_2 \log_{10}\delta_1 + a_3] \log_{10}\delta_2 + [a_4(\log_{10}\delta_1)^2 + a_5 \log_{10}\delta_1 + a_6]$$

$$F = b_1 + b_2[\log_{10}\delta_1 - \log_{10}\delta_2]$$

$$a_1 = 0.005309 \quad a_2 = 0.07114$$

$$a_3 = -0.4761 \quad a_4 = -0.00266$$

$$a_5 = -0.5941 \quad a_6 = -0.4278$$

$$b_1 = 11.01217 \quad b_2 = 0.51244$$

$$\Delta \omega = \frac{\omega_{sup} - \omega_c}{2\pi}, \quad \omega_c = \frac{2\pi f_c}{f_s}, \quad \omega_{sup} = \frac{2\pi f_{sup}}{f_s}$$

Fórmula para filtros paso banda

F. Mitzer y B. Liu han desarrollado una fórmula empírica para el cálculo del orden N para filtros FIR paso banda [3]

$$N = \frac{C_\infty(\delta_p, \delta_{sup})f_s}{\Delta f} + g(\delta_p, \delta_{sup})\frac{\Delta f}{f_s} + 1 \quad (2.62)$$

donde

$$C_\infty(\delta_p, \delta_{sup}) = \log_{10}\delta_{sup}[b_1(\log_{10}\delta_p)^2 + b_2\log_{10}\delta_p + b_3] \\ + [b_4(\log_{10}\delta_p)^2 + b_5\log_{10}\delta_p + b_6]$$

$$b_1 = 0.01201 \quad b_2 = 0.09664 \quad b_3 = -0.51325$$

$$b_4 = 0.00203 \quad b_5 = -0.57054 \quad b_6 = -0.44314$$

$$g(\delta_p, \delta_{sup}) = -14.6\log_{10}(\delta_p/\delta_{sup}) - 16.9$$

$$\Delta = f_p - f_{sup}$$

Resumen

En el diseño de filtro FIR, el uso de ventanas puede generar una mejor respuesta en frecuencia con menor número de oscilaciones en la banda de paso y en la banda rechazada. Sin embargo, las transiciones entre ambas bandas no son tan abruptas como las que se obtienen con una ventana rectangular, esto implica que es necesario incrementar el orden del filtro en el diseño [9]. La limitación principal de las ventanas de Hanning, Hamming y Blackman es que producen un factor de rizo δ fijo, con estas ventanas sólo se logran las especificaciones deseadas en las bandas de paso y suprimida. Mientras que en Kaiser no existe una limitante en cuanto a un diseño de δ fijo, sino que variando N y β se logra un δ deseable. Una desventaja del método de ventanas es que no se pueden diseñar filtros con atenuaciones diferentes en las diferentes bandas.

2.4.4. Ejemplos de diseño

- **Ejemplo 1:**

Diseñar un filtro digital paso bajas que se aproxime a una respuesta ideal con frecuencia de corte 125 hz, si la frecuencia de muestreo es 1000 hz y ganancia en la banda de paso igual a uno.

Recordando que la ω que se está utilizando corresponde a la frecuencia digital de corte, es decir ω_{cd} , esta es igual a

$$\omega_{cd} = \frac{2\pi f_{ca}}{f_s}$$

donde f_s es la frecuencia de muestreo y f_{ca} es la frecuencia de corte analógica.

Solución:

Considerando $N=11$

$$\omega_{cd} = \frac{2\pi(125)}{1000} = 0.25\pi$$

$$h_{NC}(n) = \frac{\text{sen}(n\omega_{cd})}{n\pi} = \{-0.045, 0, 0.075, 0.159, 0.225, 0.25, 0.225, 0.159, 0.075, 0, -0.045\}$$

Haciendo el corrimiento para que sea causal

$$h(n) = \{-0.045, 0, 0.075, 0.159, 0.225, 0.25, 0.225, 0.159, 0.075, 0, -0.045\}$$

Para $N=21$ y utilizando ventana de Hanning se tiene

$$h(n) = \{0.032, 0.025, 0, -0.032, -0.053, -0.045, 0, 0.075, 0.159, 0.225, 0.25, 0.225, 0.159, 0.075, 0, -0.045, -0.053, -0.032, 0, 0.025, 0.032\}$$

En la figura 2.10 se ilustra una comparación del diseño de filtros FIR con ventana cuadrada para diferente longitud N y utilizando ventana de Hamming, es obvio que el desempeño utilizando ventana de Hamming es muy superior al uso de una ventana rectangular.

■ **Ejemplo 2:**

Dado un filtro FIR causal con $N=11$, determinar la magnitud de la respuesta en frecuencia y mostrar que el retardo en fase y retardo de grupo de un filtro causal son constantes.

Solución:

Dado que

$$H(z) = \sum_{n=0}^{N-1} h(n)z^{-n} \quad \text{y} \quad \alpha = \frac{N-1}{2} = 5$$

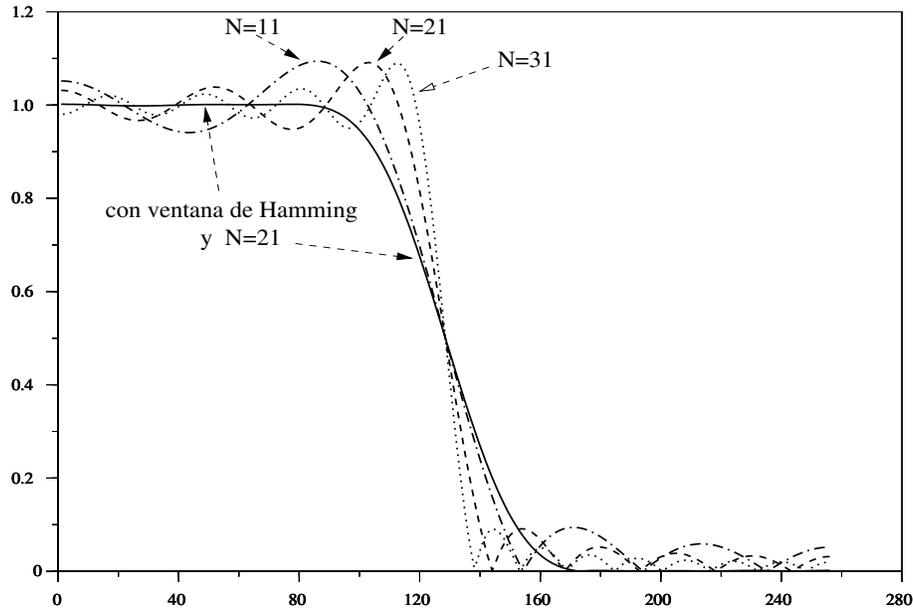


Figura 2.10: Respuesta en frecuencia de un filtro FIR para diferente N

entonces

$$H(z) = h(0) + h(1)z^{-1} + h(2)z^{-2} + h(3)z^{-3} + h(4)z^{-4} + h(5)z^{-5} + h(6)z^{-6} + h(7)z^{-7} + h(8)z^{-8} + h(9)z^{-9} + h(10)z^{-10}$$

factorizando el término z^{-5}

$$H(z) = z^{-5} \left[h(0)z^5 + h(1)z^4 + h(2)z^3 + h(3)z^2 + h(4)z^1 + h(5)z + h(6)z^{-1} + h(7)z^{-2} + h(8)z^{-3} + h(9)z^{-4} + h(10)z^{-5} \right]$$

por condición de simetría $h(n) = h(N - 1 - n)$, entonces

$$H(z) = z^{-5} \left[h(0)(z^5 + z^{-5}) + h(1)(z^4 + z^{-4}) + h(3)(z^2 + z^{-2}) + h(4)(z^1 + z^{-1}) + h(5) \right]$$

el factor z^{-n} implica un retardo de nT segundos.

Si sustituimos $z = e^{j\omega T}$ y se desarrolla por Euler se tiene

$$H(e^{j\omega}) = e^{-5j\omega T} \left[2(h(0)\cos(5\omega T) + h(1)\cos(4\omega T) \right]$$

$$+h(2)\cos(3\omega T) + h(3)\cos(2\omega T) + h(4)\cos(\omega T) + h(5)]$$

$$H(e^{j\omega}) = e^{-5j\omega T} \left[h(5) + 2 \sum_{n=1}^5 h(n)\cos(n\omega T) \right]$$

donde:

la respuesta en fase es $\theta(\omega) = -5\omega T$, es decir, que $\alpha = -5T$ y el retardo de grupo $\tau_g = -\frac{d\theta(\omega)}{d\omega} = -5T$

2.5. Diseño de filtros FIR paso altas, paso banda y supresor de banda

Para el diseño de estos filtros se puede hacer una aproximación de una función de transferencia a un filtro ideal o partir de un filtro paso bajas (FPB) utilizando transformaciones.

2.5.1. Diseño de filtros FIR FPA a partir de un FIR FPB

Existen dos métodos para convertir un filtro FIR FPB con respuesta en frecuencia $H_{FPB}(\omega)$ a un filtro FIR FPA con respuesta en frecuencia $H_{FPA}(\omega)$

- **Inversión espectral**

En este caso si se tiene una $H_{FPB}(\omega)$ con f_c se puede diseñar un filtro FPA con f_c a través de la operación en frecuencia $H_{FPA}(\omega) = 1 - H_{FPB}(\omega)$, es decir, que en el tiempo se le cambia de signo a cada muestra de $h_{FPB}(n)$ y se le suma una función $\delta(n)$. El efecto en frecuencia de esta operación consiste en hacer un giro de $H_{FPB}(\omega)$ sobre la horizontal respecto de la mitad de la amplitud de $H_{FPB}(\omega)$.

Para realizar el diseño de un filtro FIR FPA por el método de inversión espectral, si consideramos un FPB con ganancia unitaria y frecuencia de corte ω_{c1} y función de transferencia $H(\omega)$, y si el FPA tiene una función de transferencia $G(\omega)$ y frecuencia de corte ω_{c1} , entonces

$$G(\omega) = 1 - H(\omega)$$

como se conoce la respuesta al impulso $h(n)$ del FPB, entonces la respuesta al impulso del FPA sería

$$g(n) = \delta(n) - h(n)$$

- **Corrimiento espectral**

Este método se basa en la propiedad de traslación de la transformada de Fourier, el diseño de un FPA se logra cambiando de signo a cada muestra de $h_{FPB}(n)$.

Considerando el segundo método, si se desplaza la respuesta en frecuencia de un filtro paso bajas en π veces, se obtiene la respuesta en frecuencia del filtro paso altas deseado como se ve en la figura 2.11. La respuesta en frecuencia debe ser simétrica con respecto a π y periódica en 2π .

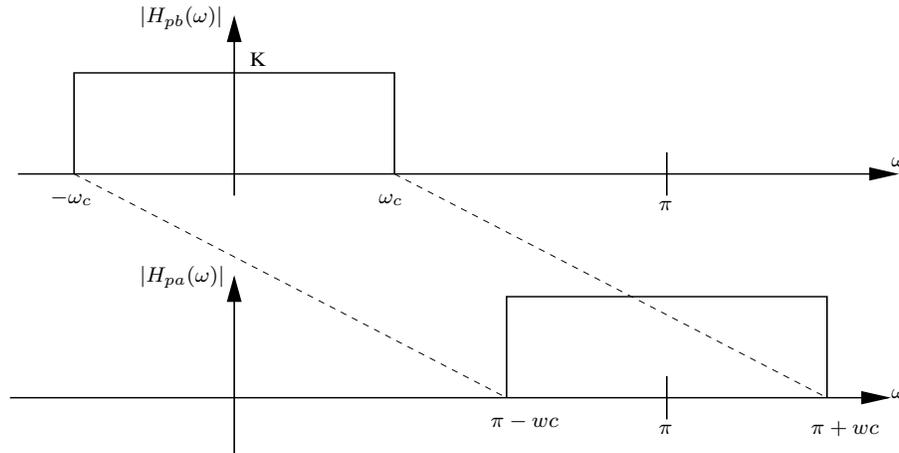


Figura 2.11: Diseño de un filtro FIR ideal paso altas

$$H_{FPA}(e^{j\omega}) = H_{FPB}(e^{j(\omega-\pi)}) \quad (2.63)$$

donde

$$H_{FPA}(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h_{FPB}(n)e^{-jn(\omega-\pi)} \quad (2.64)$$

dado que un desplazamiento en la frecuencia de π es equivalente a multiplicar la respuesta al impulso $h_{FPB}(n)$ por $e^{jn\pi}$. Para cualquier n que pertenezca a los enteros positivos $e^{jn\pi} = (-1)^n$, entonces

$$H_{FPA}(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h_{FPB}(n)(-1)^n e^{-jn\omega} \quad (2.65)$$

por lo tanto un FPA se puede diseñar con la fórmula

$$h_{FPA-NC}(n) = h_{FPB-NC}(n)(-1)^n \quad (2.66)$$

donde la frecuencia de corte digital para el cálculo sería

$$\omega_{cd} = \pi - \omega_{c-pa}$$

El cambio de signo de las muestras en el tiempo es equivalente a multiplicar $h_{FPB}(n)$ por una función cosenoidal con una frecuencia de $0.5f_s$ (recordar que el espectro está normalizado respecto de 2π), es decir

$$h_{FPA-NC}(n) = \frac{\text{sen}(\omega_{cd}n)}{n\pi} \cos(n\pi) \quad (2.67)$$

2.5.2. Diseño de filtros FIR paso banda

De manera similar se puede transformar un filtro paso bajas a un filtro paso banda (FPBW) y a un filtro supresor de banda (FSBW). Para el diseño de un filtro paso banda se parte de un filtro paso bajas desplazando la respuesta en frecuencia del FPB tanto a la izquierda como a la derecha en ω_o como se observa en la figura 2.12, es decir que se tiene

$$H_{FPBW}(\omega) = H_{FPB}(\omega - \omega_o) + H_{FPB}(\omega + \omega_o) \quad (2.68)$$

aplicando la transformada de Fourier inversa y propiedades de corrimiento en la frecuencia

$$h_{FPBW}(n) = h_{FPB}(n)e^{j\omega_o n} + h_{FPB}(n)e^{-j\omega_o n} = h_{FPB}(n)(e^{j\omega_o n} + e^{-j\omega_o n}) \quad (2.69)$$

Entonces se tienen las fórmulas de diseño

Filtro paso banda

$$h_{FPBW}(n) = 2\cos(n\omega_o)h_{FPB}(n) \quad (2.70)$$

$$2\omega_c = \omega_h - \omega_l \quad \text{y} \quad 2\omega_o = \omega_h + \omega_l$$

Filtro supresor de banda

$$h_{FSBW}(n) = K\delta(n) - h_{FPBW}(n) \quad (2.71)$$

$$h_{FSBW}(n) = -h_{FPBW}(n) \quad (2.72)$$

$$2\omega_c = \omega_h - \omega_l \quad \text{y} \quad 2\omega_o = \omega_h + \omega_l$$

■ Ejemplo 1:

Diseñar un filtro digital que se aproxime a la respuesta de un filtro paso altas ideal cuya frecuencia de corte $f_c = 500$ hz, frecuencia de muestreo es $f_s = 2500$ hz, y la ganancia en la banda de paso es 1, y $N=21$.

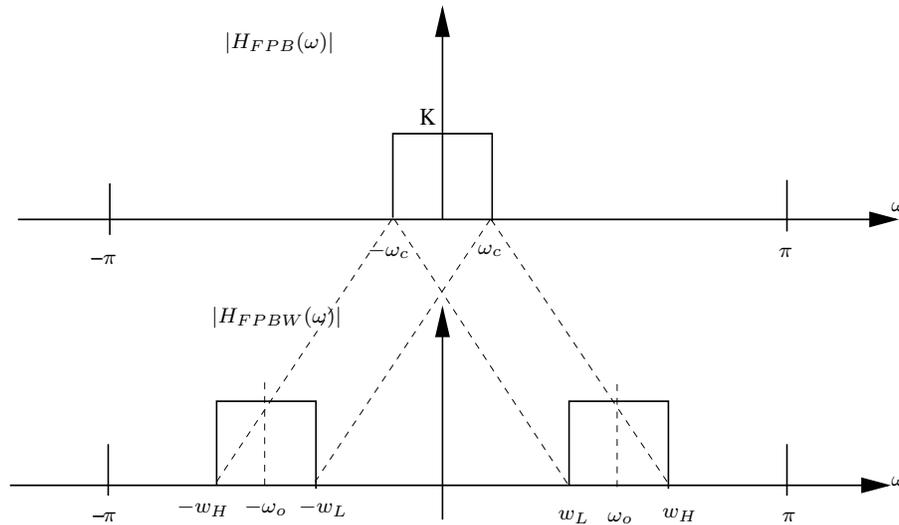


Figura 2.12: Diseño de un filtro FIR ideal paso banda

La frecuencia de corte digital se calcula

$$\omega_{cd} = \frac{2\pi 500}{2500} = 0.4\pi$$

$$0.4\pi = \pi - \omega_{cd} \Rightarrow \omega_{cd} = \pi - 0.4\pi = 0.6\pi$$

con ω_{cd} se deben calcular los valores de $h(n)$ y luego multiplicar por $(-1)^n$ para obtener los coeficientes de $h(n)$ del filtro paso altas, obteniendo

$$h(n) = \{0, 0.034, 0.023, -0.027, -0.050, 0, 0.0760, 0.062, -0.094, -0.303, 0.6, -0.303, -0.094, 0.062, 0.0760, 0, -0.050, -0.027, 0.023, 0.034, 0\}$$

■ **Ejemplo 2:**

Diseñar un filtro digital que aproxime su respuesta en frecuencia a un filtro paso banda. La región paso banda está en el intervalo de 20 KHz a 30 KHz y la frecuencia de muestreo es de 100 KHz.

Efectuando los cálculos

$$\omega_l = 2\pi 20/100 = 0.4\pi$$

$$\omega_h = 2\pi 30/100 = 0.6\pi$$

$$\omega_0 = \frac{\omega_H + \omega_L}{2} = 0.5\pi$$

$\omega_c = \frac{\omega_H - \omega_L}{2} = 0.1\pi$ con $\omega_c = 0.1\pi$, se calculan los coeficientes para el filtro paso bajas, y $\omega_0 = 0.5\pi$ se utiliza para la función coseno que aparece en la fórmula para el cálculo del FPBW

$$h_{FPBW-NC}(n) = 2\cos(n\omega_0)h_{PB}(n) \tag{2.73}$$

entonces

$$h(n) = \{0, 0, 0.04677, 0, -0.1009, 0, 0.151, 0, 0, -0.187, 0, 0.2, 0, -0.187, 0, 0.151, 0,$$

-0.1009,0,0.04677,0,0 }

En estos dos últimos ejemplos no se ha aplicado ninguna ventana, pero se puede utilizar cualquiera de las ventanas mencionadas para mejorar el desempeño del filtro.

Filtro	Fórmula de diseño	Frecuencias
Paso bajas	$h_{pb}(n) = \frac{K}{\pi n} \text{sen}(n\omega_{cd})$ $n = 0, \pm 1, \pm 2, \dots, \pm(\frac{N}{2} - 1)$	$\omega_{cd} = \frac{2\pi f_{ca}}{f_s}$
Paso altas	$h_{pa}(n) = (-1)^n h_{pb}(n)$	$\omega_{cd-pa} = \pi - \omega_{cd-pb}$ $\omega_{cd-pb} = \pi - \omega_{cd-pa}$
Paso banda	$h_{pbw}(n) = 2\cos(n\omega_0)h_{pb}(n)$	$2\omega_{cd} = \omega_{Hd} - \omega_{Ld}$ $2\omega_0 = \omega_{Hd} + \omega_{Ld}$
Supresor de banda	$h_{sbw}(n) = K - h_{pbw}(n)$ $h_{sbw}(n) = -h_{pbw}(n)$	$2\omega_{cd} = \omega_{Hd} - \omega_{Ld}$ $2\omega_0 = \omega_{Hd} + \omega_{Ld}$

Tabla 2.4: Resumen de fórmulas de diseño de FD FIR utilizando ventanas

2.6. Diseño de filtros FIR por muestreo en frecuencia

Este método se basa en el muestreo de la amplitud deseada en la frecuencia, para el cálculo de los coeficientes de $h(n)$, donde la respuesta en frecuencia deseada $H(\omega)$ es muestreada a razón de $\omega_k = 2\pi k/N$ para $k = 0, 1, 2, \dots, N - 1$

La mayor desventaja del método de ventanas para el diseño de filtros FIR es la falta del control de precisión de las frecuencias críticas, tales como ω_c y ω_{sup} , por ejemplo en el diseño de un filtro FPB los valores de estas frecuencias dependen del tipo de ventana y de la longitud N de la ventana [32]. El método del muestreo en frecuencia provee una mejora sobre el de ventanas, dado que la respuesta en frecuencia $H(\omega)$ es especificada como $\omega_k = 2\pi k/N$ o $\omega_k = \pi(2k + 1)/N$ y la banda de transición es un múltiplo de $2\pi/N$.

El método de muestreo en frecuencia es atractivo cuando el filtro FIR es diseñado en el dominio de la frecuencia, ya que la respuesta en frecuencia deseada $H(e^{j\omega})$ se define con magnitud uno para la banda de paso y cero para la banda suprimida, excepto en la banda de transición [32].

Una ventaja del método es que el diseñador puede definir una función $H(e^{j\omega})$ deseada en el dominio de la frecuencia discreta como en el caso que se requieren bandas simultáneas. Una desventaja es que la magnitud verdadera de la respuesta del filtro diseñado sólo coincide con los puntos muestreados de la respuesta deseada, para tener una buena aproximación es necesario utilizar un gran número de N muestras.

En este método no se requiere una expresión matemática para la respuesta en frecuencia deseada $H(e^{j\omega})$. El procedimiento de diseño involucra el muestreo de la respuesta en frecuencia $H(e^{j\omega})$ para formar una secuencia $H(k)$ discreta (donde $\omega_k = 2\pi k/N$) y después obtener la respuesta al impulso $h(n)$.

Para el cálculo de $h(n)$ se pueden seguir dos procedimientos:

- Utilizando soluciones de fase lineal y proponer la discretización de $H(e^{j\omega})$, donde se desarrolla para valores de n y se resuelve un sistema matricial.
- Se propone una respuesta en la frecuencia discreta $H(k)$ y se le aplica la Transformada Discreta de Fourier Inversa (DFTI).

2.6.1. Procedimiento de diseño utilizando soluciones de fase lineal

Si se sustituye la respuesta al impulso $h(n)$ suponiendo que es simétrica o antisimétrica en la función $H(\omega)$ (recordando que $H(\omega) \approx H(e^{j\omega})$, y suponiendo que $H_r(\omega)$ es la respuesta en frecuencia verdadera, se tienen cuatro soluciones dependiendo si $h(n)$ es simétrica o antisimétrica y si N es par o impar.

Solución simétrica

Para este caso $h(n) = h(N - 1 - n)$ y $\alpha = (N - 1)/2$

$$H(\omega) = H_r(\omega)e^{-j\omega(N-1)/2} \quad (2.74)$$

- Si $N = \text{impar}$

$$H_r(\omega) = h\left(\frac{N-1}{2}\right) + 2 \sum_{n=0}^{\frac{N-3}{2}} h_n \cos \omega \left(\frac{N-1}{2} - n\right) \quad (2.75)$$

- Si $N = \text{par}$

$$H_r(\omega) = 2 \sum_{n=0}^{\frac{N}{2}-1} h_n \cos \omega \left(\frac{N-1}{2} - 1 - n\right) \quad (2.76)$$

Características en fase

$$\theta(\omega) = \begin{cases} -\omega\left(\frac{N-1}{2}\right) & \text{si } H_r(\omega) > 0 \\ -\omega\left(\frac{N-1}{2}\right) + \pi & \text{si } H_r(\omega) < 0 \end{cases} \quad (2.77)$$

Cuando $H(\omega)$ cambia de signo de positivo a negativo o viceversa, la fase sufre un cambio de π radianes, si estos cambios de fase ocurren fuera de las bandas de paso o suprimida no es problemático, dado que la señal deseada tiene contenido frecuencial despreciable en la banda suprimida [32].

Solución antisimétrica

Es decir que $h(n) = -h(N-1-n)$ y $h\left(\frac{N-1}{2}\right) = 0$. Se aplica para grupos de retardo constante, esto es en diferenciadores de banda ancha y transformada de Hilbert [3].

$$H(\omega) = H_r(\omega)e^{j(-\omega(N-1)/2+\pi/2)} \quad (2.78)$$

- Si $N = \text{impar}$

$$H_r(\omega) = 2 \sum_{n=0}^{\frac{N-3}{2}} h_n \text{sen } \omega \left(\frac{N-1}{2} - n \right) \quad (2.79)$$

- Si $N = \text{par}$

$$H_r(\omega) = 2 \sum_{n=0}^{\frac{N}{2}-1} h_n \text{sen } \omega \left(\frac{N-1}{2} - n \right) \quad (2.80)$$

Características en fase

$$\theta(\omega) = \begin{cases} \frac{\pi}{2} - \omega\left(\frac{N-1}{2}\right) & \text{si } H_r(\omega) > 0 \\ \frac{3\pi}{2} - \omega\left(\frac{N-1}{2}\right) & \text{si } H_r(\omega) < 0 \end{cases} \quad (2.81)$$

Las ecuaciones (2.75), (2.76), (2.79) y (2.80) constituyen un conjunto de ecuaciones lineales para determinar la respuesta al impulso $h(n)$ del filtro FIR, siendo necesario conocer la respuesta en frecuencia en cada punto de interés de ω y resolver el sistema para $h(n)$.

Para frecuencia discreta

Con las fórmulas anteriores puede diseñarse un filtro FIR si se considera la frecuencia discreta ω_k en el intervalo $0 \leq \omega_k \leq \pi$ y considerando los valores discretos de $H_r(\omega_k)$, definiendo

Caso simétrico

$$\omega_k = 2\pi k/N \quad \alpha = \frac{N-1}{2}$$

donde

$$2h_n \cos \omega_k \left(\frac{N-1}{2} - n \right) = 1 \quad \text{si} \quad n = \frac{N-1}{2} \quad \text{para todo } k$$

$$H_r(\omega_k) = 2 \sum_{n=0}^{\frac{N-1}{2}} h_n \cos \omega_k \left(\frac{N-1}{2} - n \right) \quad k = 0, 1, 2, \dots, \frac{N-1}{2} \quad N = \text{impar} \quad (2.82)$$

$$H_r(\omega_k) = 2 \sum_{n=0}^{\frac{N}{2}-1} h_n \cos \omega_k \left(\frac{N-1}{2} - n \right) \quad k = 0, 1, 2, \dots, \frac{N}{2} - 1 \quad N = \text{par} \quad (2.83)$$

Caso antisimétrico

$$\omega_k = \frac{2\pi(k+1/2)}{N}$$

$$H_r(\omega_k) = 2 \sum_{n=0}^{\frac{N-3}{2}} h_n \sen \omega_k \left(\frac{N-1}{2} - n \right) \quad k = 1, 2, \dots, \frac{N-1}{2} \quad N = \text{par} \quad (2.84)$$

$$H_r(\omega_k) = 2 \sum_{n=0}^{\frac{N}{2}-1} h_n \sen \omega_k \left(\frac{N-1}{2} - n \right) \quad k = 1, 2, \dots, \frac{N}{2} \quad N = \text{impar} \quad (2.85)$$

Estas fórmulas generales pueden ser utilizadas para el diseño de filtros FIR de fase lineal con respuesta al impulso simétrica o antisimétrica. Aunque el diseño de filtros FIR antisimétricos no es muy conveniente para FPB y FPA, ya que $H_r(0) = 0$ y $H_r(\pi) = 0$.

Con las últimas fórmulas tanto para el caso de simetría como la antisimetría, en el proceso de diseño se proponen valores en frecuencia discreta de $H_r(\omega_k)$ entre 0 y 1, se desarrollan las sumatorias, quedando un sistema de una matriz de coeficientes $\mathbf{M}_{k,n}$ por un vector de incógnitas \mathbf{h}_n igual a un vector de valores. El sistema se resuelve para \mathbf{h}_n .

$$\mathbf{h}_n = \mathbf{M}_{k,n}^{-1} \mathbf{H}_r(\omega_k) \quad (2.86)$$

siempre y cuando exista la inversa de la matriz $\mathbf{M}_{k,n}$.

Una problemática de esta técnica es insertar en los valores de las especificaciones en la banda de transición, Rabiner y otros (1970) han propuesto y probado varias amplitudes para una variedad de valores de N , éstos valores son óptimos debido a que minimizan el lóbulo mayor

Orden N	$h(n)$	$ H(e^{j\omega}) $	$\theta(\omega)$	b_n
Impar	Simétrica	$\sum_{n=0}^{\frac{N-1}{2}} b_n \cos(n\omega)$	$e^{-j\omega(N-1)/2}$	$b_0 = h(\frac{N-1}{2})$ $b_n = 2h(\frac{N-1}{2} - n)$ $n \neq 0$
Par	Simétrica	$\sum_{n=1}^{N/2} b_n \cos \omega(n - 1/2)$	$e^{-j\omega(N-1)/2}$	$b_n = 2h(N/2 - n)$
Impar	Antisimétrica	$\sum_{n=0}^{\frac{N-1}{2}} b_n \sen(n\omega)$	$e^{-j(\omega(N-1)/2 - \pi/2)}$	$b_n = 2h(\frac{N-1}{2} - n)$
Par	Antisimétrica	$\sum_{n=1}^{N/2} b_n \sen \omega(n - 1/2)$	$e^{-j(\omega(N-1)/2 - \pi/2)}$	$b_n = 2h(\frac{N}{2} - n)$

Tabla 2.5: Forma compacta de Filtros FIR con fase lineal [9], [3]

de la banda suprimida [32].

Ejemplo:

Utilizar las ecuaciones anteriores para diseñar un filtro digital FIR paso bajas con las características deseadas en frecuencia discreta k

$$H_r(k) = \begin{cases} 1 & k = 0, 1, 2, 3, 4, 5 & \text{(banda de paso)} \\ 0.95 & k = 6 \\ 0.5 & k = 7 \\ 0.05 & k = 8 \\ 0 & k = 9, 10, 11, 12, 13, 14, 15 & \text{(banda suprimida)} \end{cases} \quad (2.87)$$

Este ejemplo se ha realizado resolviendo la ecuación (2.86), obteniendo los coeficientes:
 $h(n) = \{ 0.0006405, 0.0025547, -0.0020591, -0.0074974, 0.0010848, 0.0142960, 0.0037689, -0.0227127, -0.0151407, 0.0318656, 0.0381549, -0.0402713, -0.0889977, 0.0462296, 0.3122776, 0.4516125 \}$

En la figura 2.13 se puede apreciar el espectro real de estos coeficientes calculados, aquí se puede observar una diferencia de -50 db de la banda de paso a la banda suprimida, que se puede considerar un muy buen diseño.

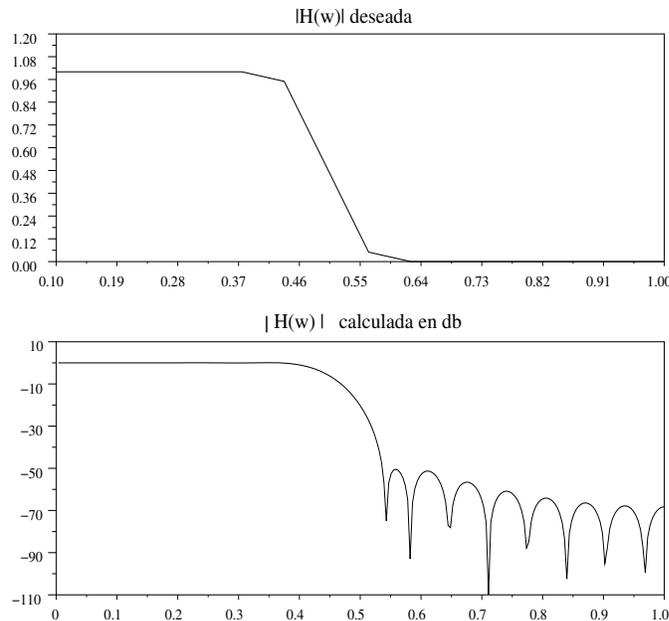


Figura 2.13: Espectro para el ejemplo

2.6.2. Procedimiento de diseño utilizando la DFTI

Cuando se utiliza este método es necesario calcular la transformada de Fourier de tiempo discreto de $h(n)$ para comprobar si el grado de aproximación de la respuesta en frecuencia deseada $H(e^{j\omega})$ es adecuado, de lo contrario hay que incrementar el número de muestras. Los pasos a seguir son:

1. Definir la función de transferencia discreta deseada $H(k)$ que representa la transformada discreta de Fourier de $h(n)$, esto es

$$H(k) = \sum_N h(n)e^{-jkn2\pi/N} \quad (2.88)$$

El objetivo de este método es aproximar la respuesta en frecuencia $H(e^{j\omega})$ utilizando para ello sólo N muestras de la respuesta al impulso, es decir

$$H(e^{j\omega}) \approx \sum_N h(n)e^{-j\omega n} \quad (2.89)$$

2. Definir la frecuencia de muestreo f_s
3. Considerar en la respuesta en frecuencia su magnitud y fase
 $H(e^{j\omega}) = H(\omega) = |H(\omega)|e^{j\theta(\omega)}$
4. Se selecciona un período de N muestras en frecuencia y muestrear $H(e^{j\omega})$ a intervalos dados por $\omega = \frac{2\pi k}{N}$. La secuencia resultante se define como

$$H(k) = H(e^{j\omega})|_{\omega=2\pi k/N}$$

5. Evaluando la DFTI, se obtienen N muestras de la respuesta al impulso $h(n)$ a partir de $H(k)$

$$h(n) = \frac{1}{N} \sum_N H(k)e^{jkn2\pi/N} \quad (2.90)$$

6. Se puede considerar la opción de utilizar una ventana $W(n)$ para suavizar la función de transferencia propuesta $H(e^{j\omega})$, es decir, en el tiempo discreto $h_w(n) = h(n)W(n)$.

El método de muestreo en la frecuencia considera transiciones abruptas de la banda de paso a la banda de transición, para suavizar esta transición y tener una mejor aproximación de la respuesta en frecuencia resultante en el muestreo a una respuesta deseada, se ha planteado el uso de muestras en la región de transición para suavizamiento de la pendiente. Sin embargo, esto obviamente incrementa la duración de la región de transición, pero aumenta la similitud con la respuesta en frecuencia deseada tanto en la banda de paso como en la banda suprimida. Este método es particularmente útil cuando sólo un pequeño porcentaje de las muestras en frecuencia son diferentes de cero, o cuando se desean bandas simultáneas.

2.6.3. Estructura de muestreo en frecuencia

Esta forma está basada en la DFT de la respuesta al impulso $h(n)$ y su implementación se lleva a una estructura paralela. Esta forma parte del hecho que $H(z)$ es la función de transferencia de un filtro FIR

$$H(z) = \sum_{n=0}^{N-1} h(n)z^{-n} \quad (2.91)$$

$$H(z) = TZ\{h(n)\} = TZ\{IDFT(H(k))\} \quad (2.92)$$

$$H(z) = \sum_{n=0}^{N-1} \left(\frac{1}{N} \sum_{k=0}^{N-1} H(k)W_N^{-kn} \right) z^{-n} = \frac{1}{N} \sum_{k=0}^{N-1} H(k) \left(\sum_{n=0}^{N-1} W_N^{-kn} z^{-n} \right) \quad (2.93)$$

utilizando la serie geométrica

$$\sum_{n=0}^{N-1} a^n = \frac{1 - a^N}{1 - a}$$

$$H(z) = \frac{1}{N} \sum_{k=0}^{N-1} H(k) \sum_{n=0}^{N-1} \left(W_N^{-k} z^{-1} \right)^n = \frac{1}{N} \sum_{k=0}^{N-1} H(k) \left(\frac{1 - W_N^{-kN} z^{-N}}{1 - W_N^{-k} z^{-1}} \right) \quad (2.94)$$

como $W_N^{-kN} = 1$

entonces la última ecuación se reduce a

$$H(z) = \frac{1 - z^{-N}}{N} \sum_{k=0}^{N-1} \frac{H(k)}{1 - W_N^{-k} z^{-1}} \quad (2.95)$$

Esta manera de expresar a un filtro FIR (ecuación 2.95) tiene una forma recursiva similar a la de un filtro de respuesta infinita al impulso (ver siguiente capítulo) ya que contiene polos y ceros, el filtro resultante es un filtro FIR tipo peine $1 - z^{-N}$ cuyos ceros se cancelan en teoría con los polos en W_N^{-k} del filtro IIR.

La implementación directa de la ecuación (2.95) requiere una aritmética compleja por lo que se puede modificar de la siguiente forma

$$H(z) = \frac{1 - z^{-N}}{N} \left(\sum_{k=1}^L 2|H(k)|H_k(z) + \frac{H(0)}{1 - z^{-N}} + \frac{H(N/2)}{1 + z^{-N}} \right) \quad (2.96)$$

donde $L = (N - 1)/2$ para N impar y $L = (N/2) - 1$ para N par.

Debido a que el diseño de FD FIR con fase lineal por muestreo en la frecuencia introduce polos y ceros equiespaciados en el círculo unitario del plano z , e idealmente los polos son eliminados por los ceros correspondientes, sin embargo, en la implementación de estos filtros en una arquitectura digital de aritmética de punto fijo, los efectos de cuantización numérica

afectan la cancelación de los polos y ceros. La ubicación de los polos sobre el círculo unitario no cancelan el ruido que se introduce en las operaciones matemáticas, por lo que este ruido puede incrementarse y destruir la operación normal del filtro llevándolo a la inestabilidad [33]. Una solución efectiva al problema de la cuantización numérica es mover los polos ligeramente a valores con magnitud cercanos a uno modificando la función de transferencia a [21]

$$H(z) = \frac{1 - z^{-N}}{N} \sum_{k=0}^{N-1} (-1)^k H(k) \left[\frac{1}{1 - e^{j2\pi k/N} z^{-1}} + \frac{1}{1 - e^{-j2\pi(N-k)/N} z^{-1}} \right]$$

$$H(z) = \frac{1 - z^{-N}}{N} \sum_{k=0}^{N/2} (-1)^k H(k) \frac{2(1 - \cos(2\pi k/N) z^{-1})}{1 - 2\cos(2\pi k/N) z^{-1} + z^{-2}} \quad (2.97)$$

Teóricamente la cancelación de polos y ceros es exacta, sin embargo, en la implementación de este tipo de filtros en aritmética de punto entero esto no es cierto, el término $\cos(2\pi k/N)$ introduce imprecisiones. Entonces moviendo los polos y ceros dentro del círculo unitario, es decir $r < 1$, $H(z)$ se puede escribir

$$H(z) = \frac{1 - r^N z^{-N}}{N} \sum_{k=0}^{N/2} (-1)^k H(k) \frac{2(1 - r\cos(2\pi k/N) z^{-1})}{1 - 2r\cos(2\pi k/N) z^{-1} + r^2 z^{-2}} \quad (2.98)$$

es evidente que este tipo de filtro se puede implementar como dos filtros en cascada, uno tipo FIR

$$P(z) = \frac{1 - r^N z^{-N}}{N} \quad (2.99)$$

y el otro como un banco de resonadores

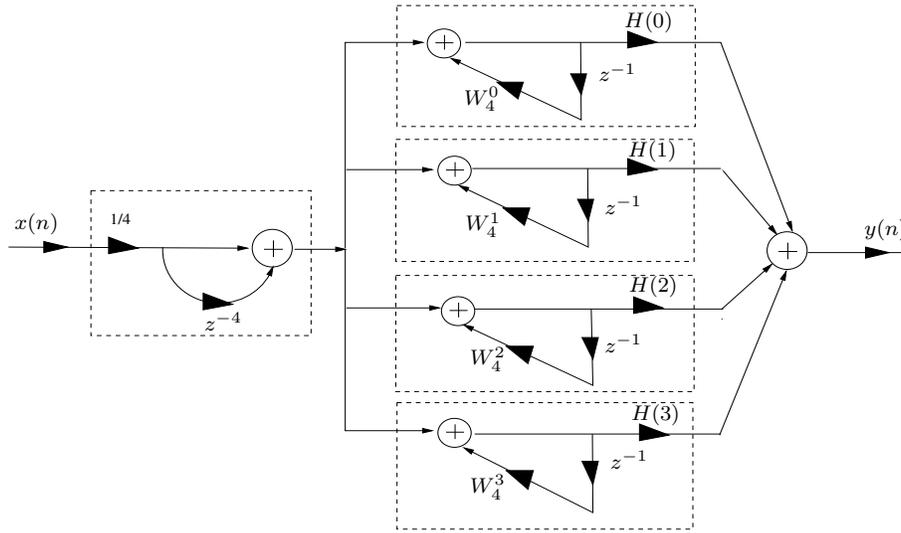
$$R_k(z) = \frac{1 - r\cos(2\pi k/N) z^{-1}}{1 - 2r\cos(2\pi k/N) z^{-1} + r^2 z^{-1}} \quad (2.100)$$

además sólo se requiere un coeficiente $r\cos(2\pi k/N)$ por resonador, los resonadores actúan como filtros de banda angosta y cada uno deja pasar sólo las frecuencias centradas en su respectiva frecuencia de resonancia $2\pi k/N$ y excluye otras fuera de esa banda. Cada resonador se multiplica a la salida por el factor $(-1)^k H(k)$ y la suma de las salidas de cada filtro contribuye a la salida total del filtro. Una estructura para $M=4$ se ilustra en la figura 2.14.

2.6.4. Localización de los polos y ceros de un filtro FIR

La respuesta en frecuencia de un filtro FIR está dada por

$$H(e^{j\omega}) = H(z)|_{z=e^{j\omega}} = \sum_{n=0}^{N-1} h(n) e^{-j\omega n} = |H(e^{j\omega})| e^{j\theta(\omega)} \quad (2.101)$$


 Figura 2.14: Estructura de muestreo en la frecuencia con $M=4$

haciendo cambios de límites la sumatoria se puede evaluar por intervalos considerando la simetría de $h(n)$ y si N es impar

$$H(e^{j\omega}) = \sum_{n=0}^{(N-3)/2} h(n)e^{-j\omega n} + h\left(\frac{N-1}{2}\right)e^{-j\omega\left(\frac{N-1}{2}\right)} + \sum_{n=(N+1)/2}^{N-1} h(n)e^{-j\omega n} \quad (2.102)$$

sustituyendo los límites de las sumas por $m = N - 1 - n$, por simetría $h(n) = h(N - n - 1)$, por causalidad $\alpha = (N - 1)/2$, y si $m = n$

$$\sum_{n=(N+1)/2}^{N-1} h(n)e^{-j\omega n} = \sum_{n=0}^{(N-3)/2} h(n)e^{-j\omega(N-1-n)} \quad (2.103)$$

donde los límites

$$(N - 3)/2 = (N - 1)/2 - 1 = \alpha - 1$$

$$\text{y } (N + 1)/2 = (N - 1)/2 + 1 = \alpha + 1$$

entonces

$$H(e^{j\omega}) = h\left(\frac{N-1}{2}\right)e^{-j\omega\left(\frac{N-1}{2}\right)} + \sum_{n=0}^{(N-3)/2} h(n)[e^{-j\omega n} + e^{-j\omega(N-1-n)}] \quad (2.104)$$

factorizando el término $e^{-j\omega\left(\frac{N-1}{2}\right)}$

$$H(e^{j\omega}) = e^{-j\omega\left(\frac{N-1}{2}\right)} \left[h\left(\frac{N-1}{2}\right) + \sum_{k=1}^{\alpha} h(\alpha - k)[e^{j\omega k} + e^{-j\omega k}] \right] \quad (2.105)$$

$$H(e^{j\omega}) = e^{-j\omega(\frac{N-1}{2})} \left[\sum_{k=0}^{\alpha} 2h(\alpha - k) \cos \omega k \right] \quad (2.106)$$

Para el caso de simetría de un filtro FIR causal en el plano Z

$$H(z) = z^{-(N-1)/2} \sum_{k=0}^{\alpha} 2h(\alpha - k)(z^k + z^{-k}) \quad (2.107)$$

Para la función de transferencia $H(z)$ de un filtro FIR se puede comprobar que

$$H(z) = z^{N-1} H(z^{-1})$$

donde $H(z^{-1})$ y $H(z)$ son idénticos excepto por el retardo de $N - 1$ muestras, además se observa que en los polinomios $H(z^{-1})$ y $H(z)$ los ceros complejos son imágenes espejo y ocurren en pares conjugados.

De la expresión $H(z)$ (ecuación 2.107) se observa que:

- Todos los polos de un filtro FIR con fase lineal se encuentran localizados en el origen del plano Z.
- Los ceros satisfacen la condición

$$\sum_{n=0}^{\frac{N-1}{2}} a(n)(z^2 \pm z^{-2}) = 0 \quad (2.108)$$

Si $z_i = r_i e^{j\theta_i}$ es un cero de $H(z)$, entonces $z_i^{-1} = r_i^{-1} e^{-j\theta_i}$ también es un cero.

De lo anterior se puede resumir que cada cero complejo de la forma $z_i = r_i e^{j\theta_i}$, se presenta con su conjugado complejo z^* y sus recíprocos, obteniendo cuatro casos:

- Ceros complejos con magnitud uno
 $z_i = r_i e^{j\theta_i}$ con $r_i = 1$ y $\theta_i \neq 0 \neq \pi$, se presenta su conjugado complejo z^* y sus recíprocos son los mismos.
- Ceros complejos con magnitud diferente de uno
 $z_i = r_i e^{j\theta_i}$ con $r_i \neq 1$ y $\theta_i \neq 0 \neq \pi$, existe el conjugado complejo y sus recíprocos.
- Ceros reales con magnitud uno
 $z_i = r_i e^{j\theta_i}$ con $r_i = 1$ y $\theta_i = 0$ o π , los ceros se localizan en $+1$ o -1 .
- Ceros reales con magnitud diferente de uno
 $z_i = r_i e^{j\theta_i}$ con $r_i \neq 1$ y $\theta_i = 0$ o π , no existe el conjugado complejo, pero si existen sus recíprocos.

2.7. Método Remez y algoritmo Parks McClellan

Uno de los problemas del diseño por ventanas y muestreo en la frecuencia es la falta de un control preciso sobre las frecuencias críticas ω_p y ω_{sup} . Los métodos de *diseño óptimo* se basan en la formulación como una aproximación al problema de *Chebyshev*, este diseño es visto como un criterio óptimo ya que el error entre la respuesta en frecuencia deseada y la respuesta verdadera se distribuye de igual forma en toda la banda de paso y banda suprimida minimizando el máximo error, el resultado del diseño genera rizados en la banda de paso y suprimida y el orden N es mínimo.

Este método es llamado algunas veces *diseño minimax*, *diseño de Chebyshev* o *diseño óptimo* y es óptimo debido a que dadas las especificaciones para ω_p , ω_{sup} , δ_1 y δ_2 genera el orden mínimo. Sin embargo, utilizando el algoritmo Parks McClellan se pueden diseñar una serie de filtros FIR con una variedad de requerimientos [9].

Del diseño de filtros FIR de fase lineal se tiene la solución simétrica y antisimétrica, que en resumen son cuatro casos y están en la tabla 2.5, las ecuaciones de esta tabla se pueden manipular para llegar a expresiones generales para el diseño de filtros óptimos. En la tabla 2.6 se pueden observar las expresiones generales donde para $H(\omega)$ en la sumatoria se tiene la función $\cos(\omega n)$ común a los cuatro casos.

Tipo	$H(\omega)$	Relación b_n y \tilde{b}_n
1	$\sum_{n=0}^{\frac{N}{2}-1} \tilde{b}_n \cos(n\omega)$	$\tilde{b}_k = b_k$ $b_1 = \tilde{b}_0 + \tilde{b}_1/2$
2	$\cos(\omega/2) \sum_{n=0}^{\frac{N-1}{2}} \tilde{b}_n \cos(n\omega)$	$b_k = \frac{1}{2}(\tilde{b}_{k-1} + \tilde{b}_k)$ $k = 2, 3, 4, \dots, (N-1)/2$ $b_{N/2} = \frac{1}{2}\tilde{b}_{(N-1)/2}$ $b_1 = \tilde{b}_0 - \tilde{b}_2/2$
3	$\sin(\omega) \sum_{n=0}^{\frac{N-3}{2}} \tilde{b}_n \cos(n\omega)$	$b_k = \frac{1}{2}(\tilde{b}_{k-1} - \tilde{b}_{k+1})$ $k = 2, 3, 4, \dots, (N-1)/2 - 2$ $b_{(N-1)/2-1} = \frac{1}{2}\tilde{b}_{(N-1)/2-2}$ $b_{(N-1)/2} = \frac{1}{2}\tilde{b}_{(N-1)/2-1}$ $b_1 = \tilde{b}_0 - \tilde{b}_1/2$
4	$\sin(\omega/2) \sum_{n=0}^{\frac{N}{2}-1} \tilde{b}_n \cos(n\omega)$	$b_k = \frac{1}{2}(\tilde{b}_{k-1} - \tilde{b}_k)$ $k = 2, 3, 4, \dots, (N-1)/2 - 1$ $b_{N/2} = \frac{1}{2}\tilde{b}_{(N-2)/2}$

Tabla 2.6: Filtros FIR con fase lineal [9], [3], [32]

Procedimiento de diseño

Dadas las especificaciones en frecuencia para una longitud N

- En banda de paso $1 - \delta_1 \leq H(\omega) \leq 1 + \delta_1$
- En banda suprimida $-\delta_2 \leq H(\omega) \leq \delta_2$

Definiendo una función de error en la frecuencia

$$E(\omega) = W(\omega)[H_I(\omega) - H(\omega)] = -(-1)^k \rho \quad k = 0, 1, 2, 3, \dots, N + 1 \quad (2.109)$$

donde

ρ representa el máximo error $E(\omega)$ y comúnmente se selecciona como $\rho = \delta_2$.

$H_I(\omega)$ es la respuesta en frecuencia ideal del filtro.

$H(\omega)$ es la respuesta en frecuencia verdadera del filtro.

$W(\omega)$ es una función de peso sobre la aproximación del error dada por

$$W(\omega) = \begin{cases} \delta_2/\delta_1 & \omega \text{ en la banda de paso} \\ 1 & \omega \text{ en la banda suprimida} \end{cases} \quad (2.110)$$

es decir que $W(\omega)$ se selecciona de tal forma que la banda de paso refleje un énfasis del tamaño del rizo en la banda suprimida con respecto al rizo en la banda de paso.

El problema de aproximación de Chebyshev consiste básicamente en la determinación de los parámetros del filtro que minimicen el máximo del valor absoluto del error $E(\omega)$ sobre las bandas de frecuencia donde se efectúa la aproximación. Matemáticamente la solución del problema es

$$\min [max E(\omega)]_{\omega \in S} = \min [W(\omega)[H_I(\omega) - H(\omega)]_{\omega \in S} \quad (2.111)$$

Donde S representa un conjunto en las bandas de frecuencia sobre el que se realiza la optimización, básicamente de las bandas de paso y suprimida del filtro deseado. La solución de este problema fué obtenido por Parks y McClellan (1972) quienes aplicaron un teorema a la teoría de aproximación de Chebyshev [32], este teorema es llamado *de alternancia*.

Como resultado del teorema de la alternancia para un intervalo cerrado $0 \leq \omega \leq \pi$, definiendo la función $P(e^{j\omega})$ como una combinación lineal de funciones cosenos

$$P(e^{j\omega}) = \sum_{n=0}^M \alpha(n) \cos(n\omega) \approx P(\omega) \quad (2.112)$$

redefiniendo la función de peso del error

$$E(\omega) = W(\omega)[H_I(\omega) - P(\omega)] \quad (2.113)$$

donde $P(\omega)$ es la mejor aproximación de Chebyshev a la función ideal $H_I(\omega)$ con la ponderación de $W(\omega)$ si la función

$$\max_{\omega} |E(\omega)| \tag{2.114}$$

es lo más pequeño posible sobre los valores de $\alpha(n)$.

El teorema de alternancia

Enuncia que de la ecuación (2.113), la función $P(\omega)$ es la mejor aproximación única a $H_I(\omega)$ si y sólo si existen al menos $N + 2$ puntos de ω_i tal que para los valores $\omega_1 < \omega_2 < \dots < \omega_{N+2}$ entre $0 \leq \omega_i \leq \pi$, se deben cumplir las condiciones

$$E(\omega_i) = -E(\omega_{i+1}) \quad i = 1, 2, 3, \dots, N + 1 \tag{2.115}$$

$$|E(\omega_i)| = \max_{\omega} |E(\omega)| \quad i = 1, 2, 3, \dots, N + 2 \tag{2.116}$$

La función de error $E(\omega)$ alterna el signo entre dos frecuencias ω_i sucesivas, por eso se le llama de alternancia. El teorema de alternancia garantiza una solución única para el problema de Chebyshev (ecuación 2.111). La mayoría de las técnicas del diseño óptimo son esencialmente esquemas de iteración para satisfacer las condiciones del teorema de alternancia en la aproximación de la función de peso de error $E(\omega)$.

Método de Herrmann

Herrmann (1970) propuso una solución equirrizo que cumple con el teorema de la alternancia [9], el método consiste en:

1. Proponer δ_1 , δ_2 y N , donde $N = N_p + N_r - 1$
2. Seleccionar una solución $H(\omega)$ de los cuatros casos para un filtro FIR, esta función puede tener $N - 1$ extremos en el intervalo $0 < \omega < \pi$, ver figura 2.15.
3. N_p representan los máximos o mínimos (extremos) en la banda de paso $0 < \omega < \omega_p$.
 N_r representan los máximos o mínimos (extremos) en la banda suprimida $\omega_{sup} < \omega < \pi$.
donde $N_p + N_r - 1 \leq N + 1$.
4. Herrmann calcula los coeficientes tal que existan $N+1$ máximos de $H(\omega)$ en el intervalo $0 < \omega < \pi$, $H'(\omega)$ es la derivada de $H(\omega)$, bajo las restricciones:

$$\begin{aligned} H(\omega_k) &= 1 - (-1)^k \delta_1 & k = 1, 2, 3, \dots, N_p \\ H'(\omega_k) &= 0 & k = 1, 2, 3, \dots, N_p - 1 \\ H(\theta_k) &= (-1)^k \delta_2 & k = 1, 2, 3, \dots, N_r \\ H'(\theta_k) &= 0 & k = 1, 2, 3, \dots, N_r - 1 \end{aligned} \tag{2.117}$$

5. Se resuelve el sistema anterior y se obtienen las $2N$ soluciones para:

$$\omega_1, \omega_2, \omega_3, \dots, \omega_{N_p-1}$$

$$\theta_1, \theta_2, \theta_3, \dots, \theta_{N_r-1}$$

$$b_0, b_1, b_2, b_3, \dots, b_N$$

6. Los valores de $h(n)$ se calculan de la solución $b(n)$.

De la solución anterior se observa que el error se alterna entre máximos y mínimos, lo que satisface el teorema de alternancia y el resultado del diseño es óptimo en el sentido de Chebyshev. La solución genera un rizo más que el número mínimo para satisfacer el teorema de la alternancia, por eso a la solución de Herrmann se le llama *extrarrizo*.

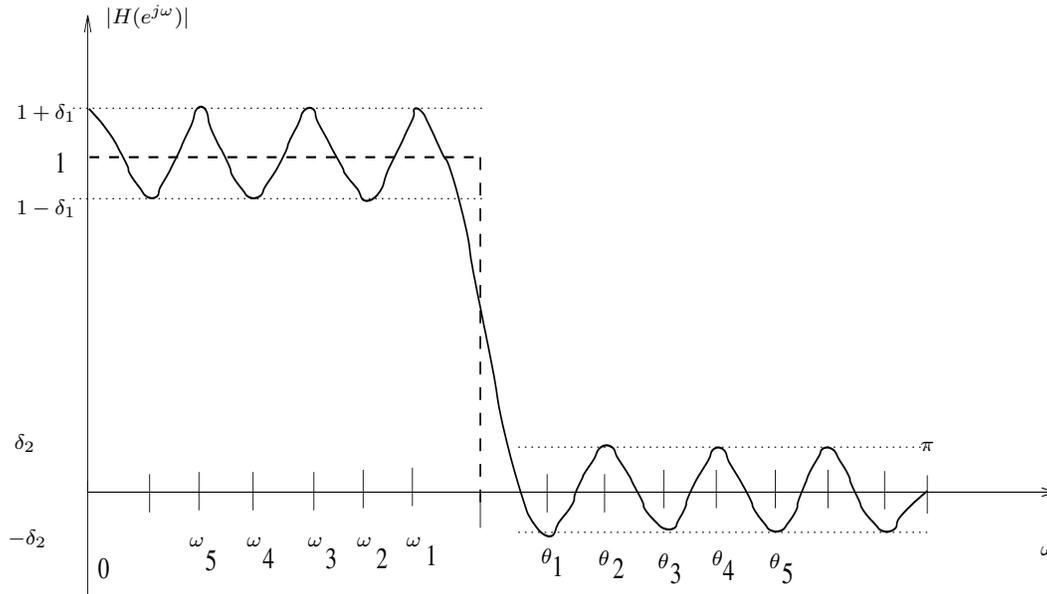


Figura 2.15: Condiciones equirrizo para el método de Herrmann

Algoritmo Parks McClellan

El algoritmo Parks McClellan deja fijos los parámetros N , ω_p , ω_{sup} , varía δ_1 y δ_2 , que son determinadas por el algoritmo. El procedimiento es iterativo y estima las frecuencias en las que ocurre un máximo en la banda de paso o suprimida, utiliza la fórmula de interpolación de Lagrange para obtener un polinomio que tenga los valores extremos preescritos de $1 \pm \delta_1$ en la banda de paso y $\pm\delta_2$ en la banda suprimida. El proceso es terminado cuando los valores máximos y mínimos del polinomio están dentro de los límites de tolerancia especificados. Este es por tanto, el problema de aproximación de Chebyshev. Si el error no es lo suficientemente pequeño, se puede incrementar N y rediseñar la función de transferencia.

Resumen del algoritmo Parks McClellan

Con base a la discusión anterior se pueden ordenar los pasos del algoritmo PM

1. Dar las especificaciones N , ω_p , ω_{sup} , δ_1 , δ_2 y $K = \delta_1/\delta_2$
2. Minimizar la ecuación $E(\omega) = W(\omega)[H_I(\omega) - P(\omega)]$, donde

$$P(\omega) = \sum_{n=0}^M \alpha(n) \cos(n\omega)$$

3. Hallar los coeficientes $b_n = \alpha(n)$ tal que el error máximo

$$\max_{\omega} |E(\omega)|$$

sea mínimo, esto involucra el procedimiento de intercambio Remez que sigue.

4. Suponer que se han dado S de $N+2$ pruebas en las frecuencias extremas $\omega_1, \omega_2, \dots, \omega_{N+2}$ y se fuerza la función de peso del error que debe satisfacer las condiciones de alternancia del error a esas frecuencias:

$$\rho = E(\omega_i) = -E(\omega_{i+1}) \quad i = 1, 2, 3, \dots, N + 1$$

donde ρ todavía es desconocido

5. Resolver para $N+2$ valores desconocidos $\alpha(0), \alpha(1), \dots, \alpha(N), \rho$. Si el error $E(\omega)$ queda dado por

$$-(-1)^k \rho = W(\omega)[H_I(\omega) - P(\omega)]$$

la solución conduce a resolver el sistema matricial:

$$P(\omega) + \frac{-(-1)^k \rho}{W(\omega)} = H_I(\omega)$$

6. Después de obtener los valores de $\alpha(i)$ para $i = 0, 1, \dots, N$, calcular la función de error $E(\omega)$. Hasta aquí el conjunto de frecuencias ω_i pueden estar fuera de los extremos (puntos con derivada cero) y la maximización del error puede no ser igual a ρ . Ahora $E(\omega)$ es conocido y se puede calcular un nuevo conjunto de frecuencias ω_i donde $E(\omega)$ verdaderamente sea extremo.
7. Con el nuevo $E(\omega)$ resolver para $\alpha(i)$ y ρ tal que se satisfaga $\rho = E(\omega_i) = -E(\omega_{i+1})$ para el nuevo conjunto de frecuencias.
8. El proceso se repite hasta que converja donde las frecuencias obtenidas satisfacen las frecuencias extremas. Entonces la segunda condición del teorema de alternancia se debe cumplir.

El algoritmo Parks McClellan fuerza el incremento de ρ en cada iteración hasta que converja a un límite superior donde se encuentra la solución óptima del problema de aproximación de Chebyshev. Una de las características interesantes del algoritmo es su flexibilidad al ofrecer una gran cantidad de respuestas de filtros. Además, el resultado del diseño genera una solución equirrizo. En la práctica no se necesitan resolver explícitamente $N + 2$ ecuaciones simultáneas de $-(-1)^k \rho = W(\omega)[H_I(\omega) - P(\omega)]$, se calcula el valor de ρ y se obtienen los valores de $P(\omega)$ a frecuencias ω_k y la función completa de $P(\omega)$ se calcula por interpolación. Antoniou ha desarrollado un procedimiento significativo que reduce más del 8% de cómputo de este algoritmo [9].

2.8. Otros tipos de filtros FIR

Existen otros tipos de filtros digitales con características muy especiales que no necesariamente siguen los procedimientos mencionados en este capítulo, sin embargo, se puede diseñar utilizando algunos conceptos de filtros digitales FIR.

2.8.1. Filtro FIR tipo peine

Este es un caso especial de filtros multibanda y su respuesta en frecuencia se parece a un peine, es útil para dejar pasar o eliminar frecuencias específicas y sus armónicas cuando existe la presencia de señales periódicas no deseadas, por ejemplo eliminar la frecuencia de 60 hz de energía eléctrica y armónicas de 120 hz, 240 hz, etc., también se utiliza en la separación de componentes solares y lunares de las medidas ionosféricas de la concentración de electrones [32]. La ecuación en diferencias de un filtro peine es

$$y(n) = x(n) - x(n - L) \quad (2.118)$$

donde L es el número de retardos y es un valor entero. Su función de transferencia es

$$H(z) = 1 - z^{-L} = \frac{z^L - 1}{z^L} \quad (2.119)$$

con respuesta en frecuencia

$$H(e^{j\omega}) = 2\text{sen}(\omega L/2), \quad \Phi(\omega) = \frac{\pi}{2} - \frac{\omega L}{2} \quad (2.120)$$

es decir, que el filtro peine tiene L polos múltiples en el origen y L ceros igualmente espaciados sobre el círculo unitario del plano z , en

$$z_l = e^{j(2\pi/L)l} \quad ; \quad l = 0, 1, 2, 3, \dots, L - 1 \quad (2.121)$$

$$y(n) = x(n) - x(n - L) \quad (2.122)$$

el centro de las bandas de paso está ubicado en las frecuencias

$$\frac{(2l + 1)\pi}{L} \quad ; \quad l = 0, 1, 2, 3, \dots, L - 1$$

y la posición de las armónicas a atenuar están en

$$\omega_l = \frac{2\pi l}{L} \quad ; \quad l = 0, 1, 2, 3, \dots, L - 1$$

Se puede seleccionar la cancelación de un cero en el filtro peine agregando un polo similar, la cancelación del cero provee una banda de paso, mientras los ceros no cancelados proveen

atenuación en la banda suprimida. Por ejemplo si se introduce un polo en $z = 1$ se tiene un filtro FIR con

$$H(z) = \frac{1 - z^{-L}}{1 - z^{-1}} \quad (2.123)$$

produce un filtro paso bajas, si el polo está en $z = \pm j$ produce un filtro paso banda y si el polo está en $z = -1$ produce un filtro paso altas.

Otra forma general de diseño de un filtro FIR peine es considerando la función de transferencia

$$H(z) = \sum_{k=0}^M h(k)z^{-k} \quad (2.124)$$

y reemplazar a z por z^L , donde L es un entero positivo

$$H(z) = \sum_{k=0}^M h(k)z^{-kL} \quad (2.125)$$

si la respuesta original del filtro FIR es $H(\omega)$, entonces la respuesta en frecuencia del filtro tipo peine es

$$H_L(\omega) = \sum_{k=0}^M h(k)e^{-jkL\omega} = H(L\omega) \quad (2.126)$$

la respuesta $H_L(\omega)$ es una repetición de orden L de $H(\omega)$ en el intervalo $0 \leq \omega \leq 2\pi$.

Si el sistema original FIR con función de transferencia $H(z)$ tiene ceros a frecuencia ω_0 , entonces la función de transferencia $H_L(\omega)$ tiene valores nulos (los respectivos ceros) a frecuencias $\omega_k = \omega_0 + 2\pi k/L$ para $k = 0, 1, 2, 3, 4, \dots, L-1$, como se ve en figura 2.16. Se puede cancelar selectivamente uno o más ceros en el filtro peine con un polo correspondiente. La cancelación de un cero provee una banda de paso, mientras que los ceros restantes proveen atenuación en la banda suprimida [21].

Un ejemplo típico se presenta en la figura 2.17, donde se ve claramente la respuesta en frecuencia de un filtro FIR tipo peine con $L = 4$ que suprime múltiples bandas.

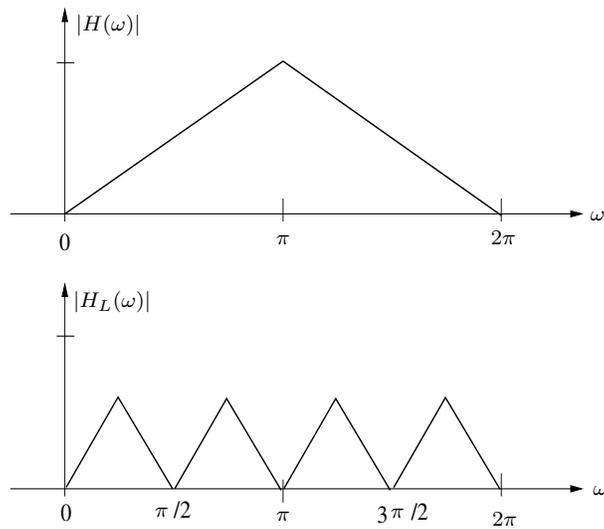


Figura 2.16: Relación de $H(\omega)$ y $H_L(\omega)$ de un filtro peine

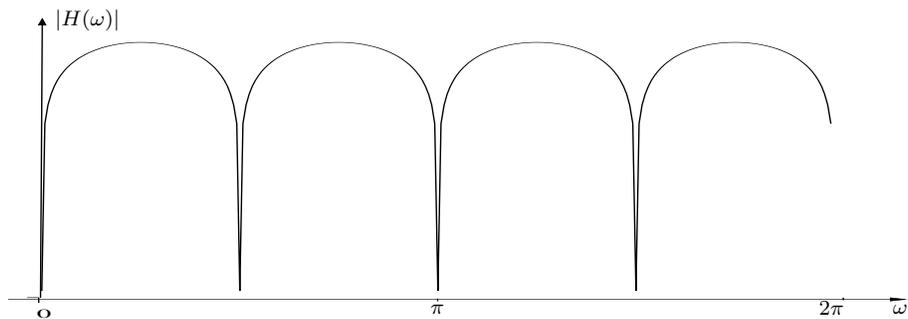


Figura 2.17: Filtro FIR tipo peine, $L = 4$

2.8.2. Filtro FIR diferenciador

Un filtro diferenciador ideal tiene una respuesta en frecuencia que es proporcional a la frecuencia y se define

$$H(\omega) = j\omega \quad -\pi \leq \omega \leq \pi \quad (2.127)$$

con respuesta al impulso

$$h(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} j\omega e^{j\omega n} d\omega$$

$$h(n) = \frac{\cos\omega n}{n} \quad -\infty \leq n \leq \infty, n \neq 0 \quad (2.128)$$

su respuesta al impulso antisimétrica, $h(n) = -h(-n)$, es decir, que su diseño puede basarse en la consideración de los filtros FIR antisimétricos del tipo tres y cuatro.

Un filtro diferenciador debe tener una respuesta en frecuencia igual a cero a frecuencia cero ($H(0) = 0$) para que se satisfaga su condición verdadera de diferenciador. Sin embargo, si se desea un diferenciador de banda completa es imposible lograrlo con un filtro FIR de longitud N impar, dado que $H(\pi) = 0$ para N es impar. En la práctica los diferenciadores se definen en un intervalo de frecuencia entre $0 \leq \omega \leq 2\pi f_p$, donde f_p es la frecuencia de paso y el intervalo $2\pi f_p \leq \omega \leq \pi$ se puede forzar a cero.

2.8.3. Filtros FIR por transformada de Hilbert

Es un caso especial de un filtro paso todo, el cual produce un corrimiento de fase de 90° a las señales de entrada. El uso principal de los filtros paso todo es la corrección de la distorsión de fase introducida por un sistema físico u otro sistema [21]. La transformada de Hilbert se utiliza frecuentemente en comunicaciones como en generación de señales moduladas con banda lateral única, en procesamiento de señales de radar y en procesamiento de voz [32].

Para el diseño de este tipo de filtros se puede hacer a través de los tipos 3 y 4 de la tabla 2.6 utilizando el algoritmo de Parks McClellan.

La transformada de Hilbert se especifica

$$H(\omega) = \begin{cases} -j & 0 < \omega \leq \pi \\ j & -\pi < \omega < 0 \end{cases} \quad (2.129)$$

Aplicando la transformada inversa de Fourier, se puede calcular la respuesta al impulso de la transformada de Hilbert:

$$h(n) = \begin{cases} \frac{2}{\pi} \frac{\text{sen}^2(\pi n/2)}{n} & n \neq 0 \\ 0 & n = 0 \end{cases} \quad (2.130)$$

donde $h(n)$ es antisimétrica, infinita y no causal, para el caso causal con una $h(n)$ de longitud N , $h(n) = -h(N-n-1)$. También se observa que para una respuesta al impulso antisimétrica

su respuesta en frecuencia es puramente imaginaria.

Para el diseño de los casos tres y cuatro de filtros FIR se requiere que la $H(\omega)$ sea cero para $\omega = 0$ para N par e impar y también para N impar en $\omega = \pi$. Por lo que sería imposible diseñar un filtro paso todo tipo Hilbert, sin embargo, en la práctica los transformadores paso todo de Hilbert son innecesarios, sólo se necesita cubrir una banda de frecuencias de interés. Entonces el valor real de la respuesta de Hilbert sería:

$$H(\omega) = 1 \quad \omega_L \leq \omega \leq \omega_H \quad (2.131)$$

2.9. Diseño de filtros FIR con programas simuladores

Como resumen final se sugiere al lector que consulte el software libre y comercial que existe en el mercado tales como Octave, Scilab, Matlab, Ptolemy, etc. Por ejemplo en Matlab se pueden encontrar las funciones para el cálculo, prueba y diseño de filtros digitales FIR:

```
fir1, fir2, fircls1, firls, fircles,  
barlett, hanning, hamming, blackman,  
kaiser, kaiserord, remez, hilbert,  
freqz, filter
```

2.10. Resumen

En este capítulo se han analizado la características principales de los filtros digitales FIR, la teoría de análisis y diseño, los métodos y procedimientos más utilizados, sus estructuras de implementación, también se han señalado en algunos casos las aplicaciones de estos filtros. Cabe mencionar que en las metodologías de diseño algunas veces resultan muy tediosas cuando se realizan manualmente o en calculadora, sin embargo, con las herramientas modernas de cómputo, una vez entendida esta teoría, se pueden desarrollar programas personales en lenguajes de alto nivel que nos permitan realizar nuestros propios diseños. O bien si se quieren utilizar herramientas de cálculo y diseño más ilustrativas podemos acudir a programas de aplicación que ya se han mencionado. Sabiendo de las virtudes de los filtros FIR, en el siguiente capítulo se abordará el diseño de filtros IIR, que son otra alternativa de filtros digitales con sus ventajas y desventajas.

Capítulo 3

Filtros de respuesta infinita al impulso (IIR)

Los filtros de respuesta infinita al impulso (IIR) también son llamados recursivos o autorregresivos de movimiento promedio (ARMA), ya que constan de una parte que efectúa la suma ponderada de la entrada $x(n)$ y retardos de $x(n - i)$, y la otra que efectúa una suma ponderada de las salidas $y(n - i)$ retrasadas. Recordando que todo filtro analógico (FA) genera una respuesta infinita al impulso, parece obvio que un sistema discreto con respuesta infinita al impulso pueda emular las características de un filtro analógico. La respuesta infinita de un sistema discreto IIR se debe a su carácter recursivo que retroalimenta la salida retrasada. Debido a la similitud con los requerimientos de un FA, en el diseño de un filtro digital IIR, la técnica más popular es el mapeo analógico a digital, es decir, mapear la función de transferencia $H(s)$ en el plano s a la función de transferencia $H(z)$ en el plano z .

En este capítulo se abordará el análisis de los filtros digitales IIR, sus características más importantes, estructuras y metodologías de diseño. Cabe mencionar que con base al análisis realizado y los conceptos manejados para el diseño de filtros FIR, el estudio de los filtros IIR puede ser más fácil.

3.1. Características de los filtros IIR

Los filtros IIR tienen algunas ventajas y desventajas sobre los filtros FIR que se enuncian a continuación:

- Debido a su recursividad con pocos coeficientes pueden generar filtros de gran pendiente en la banda de transición.
- Los filtros IIR pueden generar respuestas con un buen grado de aproximación al comportamiento de un filtro analógico, esto es, emulando la respuesta al impulso o la

respuesta en frecuencia, ya que cuando la frecuencia de muestreo es muy alta en comparación con las frecuencias de interés el comportamiento de un filtro IIR es similar a un filtro analógico.

- Los filtros IIR pueden ser inestables, por lo que hay que ser cuidadosos en el diseño. Los errores introducidos en el cálculo de la salida $y(n)$ debido a la precisión finita de los cálculos matemáticos son difíciles de predecir y como consecuencia pueden hacer inestable el sistema al mover los polos fuera del círculo unitario.
- Como $h(n)$ es infinita, no es posible tomar una suma de convolución desde un punto de vista práctico.
- No pueden ser de fase lineal como en los filtros FIR, pero esto se compensa al tener una mejor respuesta de la magnitud en frecuencia.

Del capítulo anterior se sabe que un filtro de fase lineal debe tener una función de transferencia que satisfaga la condición [32]

$$H(z) = \pm z^{N-1} H(z^{-1})$$

donde z^{N-1} representa un retardo de $N - 1$ unidades en el tiempo. Para implementar este tipo de filtro como IIR debe tener polos imagen (inversos) fuera del círculo unitario por cada polo dentro del círculo unitario, que lo haría inestable. Por tanto, un filtro IIR causal y estable no puede implementarse con fase lineal.

3.2. Estructuras de filtros digitales IIR

Recordando que la salida de un sistema lineal y discreto puede escribirse como la convolución de la entrada con su respuesta al impulso, y si un filtro IIR su respuesta al impulso es infinita, entonces

$$y(n) = x(n) * h(n) = \sum_{i=0}^{\infty} h(i)x(n-i) \quad (3.1)$$

se observa que esta operación para un sistema IIR computacionalmente no es realizable ya que se necesitan infinitos coeficientes, sumas y productos. Por tanto, para que un sistema discreto de componentes finitas pueda generar una salida infinita con una entrada impulso, el sistema debe ser recursivo, entonces para poder realizar un filtro o sistema IIR es necesaria una ecuación en diferencias que retroalimente la salida retardada, es decir

$$y(n) = \sum_{i=0}^q b(i)x(n-i) - \sum_{i=1}^p a(i)y(n-i) \quad (3.2)$$

De forma similar a los filtros digitales FIR, se tienen diferentes estructuras de implementación de filtros IIR, considerando la ecuación en diferencias que se utiliza para implementar

un filtro IIR y aplicando la Transformada Z (TZ), se obtiene la función de transferencia $H(z)$ del filtro

$$H(z) = \frac{Y(z)}{X(z)} = \frac{B(z)}{A(z)} = \frac{\sum_{i=0}^q b(i)z^{-i}}{\sum_{i=1}^p a(i)z^{-i}} = \frac{b_0 + b_{-1}z^{-1} + \dots + b_qz^{-q}}{a_0 + a_1z^{-1} + \dots + a_pz^{-p}}, \quad a_0 = 1 \quad (3.3)$$

donde p es el número de polos, q es el número de ceros de $H(z)$, a_i y b_i son los coeficientes del filtro.

Cualquier tipo de estructura que se seleccione debe ser equivalente al mismo sistema. Dependiendo como se traten las dos ecuaciones anteriores se pueden obtener cuatro formas o estructuras de implementación más utilizadas:

- Forma directa I
- Forma directa II o canónica
- Forma cascada
- Forma paralela

Otras estructuras de interés como “lattice” y “lattice-ladder” se estudiarán en los capítulos siguientes.

3.2.1. Filtro IIR forma directa

En esta forma la ecuación (3.2) es implementada directamente en un diagrama de bloques básicos sin efectuar ningún cambio. Existen dos partes en este filtro denominadas movimiento promedio y parte autorregresiva (ARMA). A la vez esta implementación conduce a dos formas: directa I y directa II.

Forma directa I

Si la ecuación en diferencias (3.2) se desarrolla directamente, entonces se tiene:

$$y(n) = b_0x(n) + b_1x(n-1) + \dots + b_qx(n-q) - a_1y(n-1) - a_2x(n-2) - \dots - a_py(n-p) \quad (3.4)$$

Se observa que existen dos líneas de retardo, una para la señal de entrada $x(n)$ y la otra para la señal de salida $y(n)$, es decir, que se tendrán $p + q$ bloques de retardo como se aprecia en la figura 3.1.

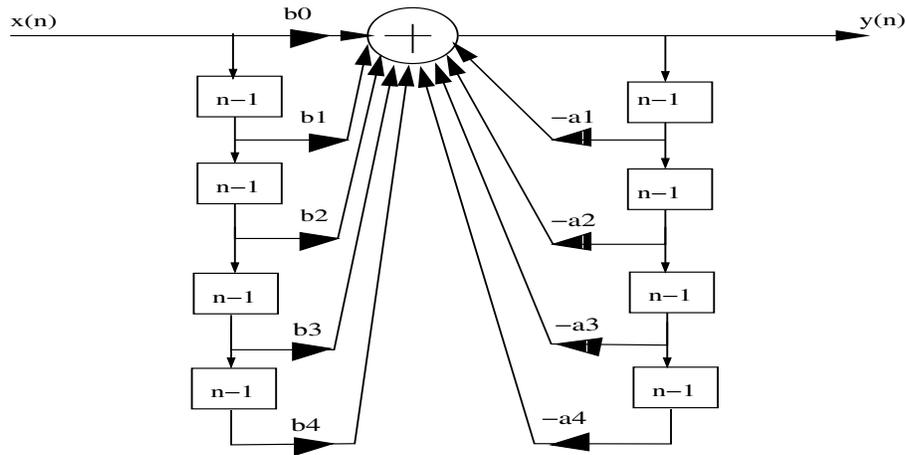


Figura 3.1: IIR forma directa I

Forma directa II o canónica

Manipulando la ecuación (3.3), las líneas de retardo se pueden convertir en una sola, lo que conduce a la forma directa II de un filtro IIR o estructura canónica. Esta forma se deduce al descomponer la ecuación (3.3) en dos factores

$$H(z) = \frac{Y(z)}{X(z)} = \frac{B(z)}{1 + A(z)} = B(z)(1 + A(z))^{-1} = B(z)C(z) \quad (3.5)$$

$$\frac{Y(z)}{X(z)} = B(z)C(z)$$

para un punto intermedio $l(n)$

$$\Rightarrow Y(z) = B(z)C(z)X(z) = B(z)L(z) \quad (3.6)$$

$$L(z) = C(z)X(z)$$

$$L(z) = X(z) - a_1L(z)z^{-1} - a_2L(z)z^{-2} + \dots \quad (3.7)$$

entonces por transformada Z inversa

$$l(n) = x(n) - a_1l(n-1) - a_2l(n-2) + \dots \quad (3.8)$$

para la salida

$$Y(z) = b_0L(z) + b_1L(z)z^{-1} + b_2L(z)z^{-2} + \dots \quad (3.9)$$

$$y(n) = b_0l(n) + b_1l(n-1) + b_2l(n-2) + \dots \quad (3.10)$$

Esta estructura se expresa gráficamente en la figura 3.2.

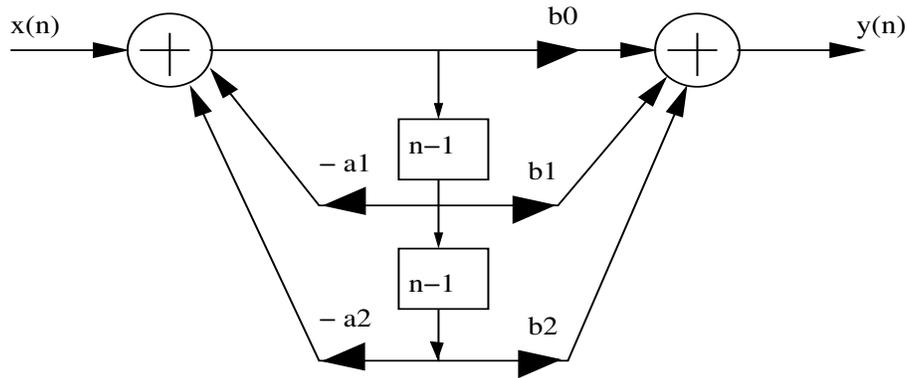


Figura 3.2: IIR forma directa II o canónica

A esta forma se le llama canónica, dado que la implementación de la función de transferencia se realiza con la menor cantidad de bloques de retardo. Una estructura de un filtro digital se dice que es canónica si el número de retardos de su representación en diagramas de bloques es igual al orden de la ecuación en diferencias [21], [24].

Otra forma canónica

Si se manipulan las operaciones en la ecuación en diferencia, se puede obtener otra forma canónica conocida como *forma directa transpuesta*, que se ilustra en la figura 3.3 y también tiene un mínimo de retardos.

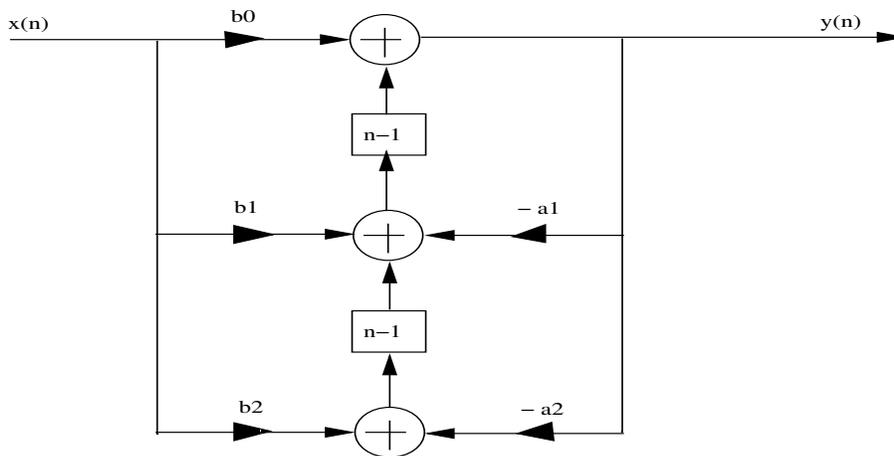


Figura 3.3: Forma directa transpuesta

3.2.2. Filtro IIR en cascada

En esta forma, la ecuación (3.3) es factorizada en secciones de primero y/o segundo orden, llamadas secciones “*biquads*”. La función de transferencia del sistema es representada como el producto de funciones biquads. Cada función biquad es implementada en forma directa o canónica y la función completa del sistema como secciones *biquad* en cascada. Es decir, que se factorizan el numerador y denominador de la ecuación (3.3) para obtener factores de segundo orden, los polos y ceros se agrupan en secciones segundo orden con polos o ceros complejos conjugados para obtener coeficientes reales. La ecuación a implementar es

$$H(z) = b_0 \prod_{k=1}^K \frac{1 + B_{k,1}z^{-1} + B_{k,2}z^{-2}}{1 + A_{k,1}z^{-1} + A_{k,2}z^{-2}} \quad (3.11)$$

Donde $K = \frac{N}{2}$, y $B_{k,1}$, $B_{k,2}$, $A_{k,1}$ y $A_{k,2}$ son números reales que representan los coeficientes de las secciones de segundo orden. Cada sección biquad se implementa en forma directa o canónica y la entrada de la sección k -ésima es la salida de la sección $(k - 1)$ (figura 3.4). Esta forma permite disminuir la acumulación de errores numéricos cuando los filtros se programan en aritmética de punto fijo, ya que en la forma directa la variación de uno de los parámetros afectará la localización de todos los polos de $H(z)$, por el contrario, en forma cascada la variación de uno de los parámetros sólo afectará al respectivo polo de esa sección. Por lo tanto, la realización en cascada es menos sensible a la variación de los parámetros comparada con la forma directa.

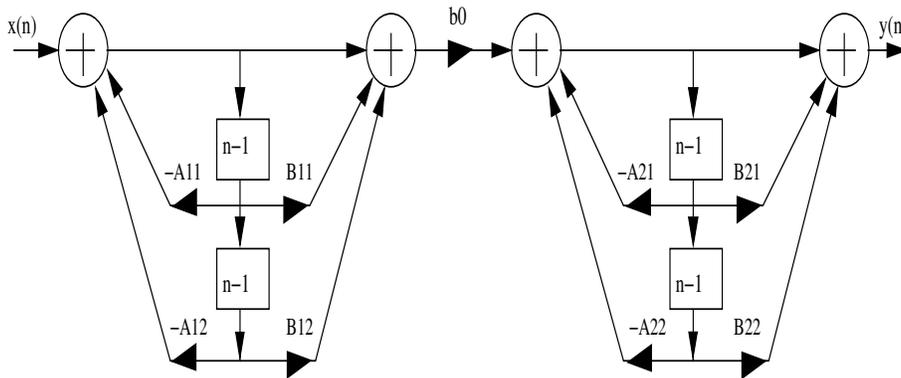


Figura 3.4: Filtro IIR de cuarto orden en cascada

3.2.3. Filtro IIR en paralelo

Es similar a la forma cascada, pero con la diferencia que la ecuación (3.3) se representa como una suma de fracciones parciales de primero y/o segundo orden. De nuevo cada sección

es implementada en forma directa o canónica y la función total del sistema como una red paralela de las secciones. Desarrollando la ecuación (3.3) en fracciones parciales:

$$H(z) = \frac{\hat{b}_0 + \hat{b}_1 z^{-1} + \dots + \hat{b}_q z^{-q}}{1 + a_1 z^{-1} + \dots + a_p z^{-p}} + \sum_{k=0}^{q-p} C_k z^{-k} \quad q \geq p \quad (3.12)$$

$$H(z) = \sum_{k=1}^K \frac{B_{k,0} + B_{k,1} z^{-1}}{1 + A_{k,1} z^{-1} + A_{k,2} z^{-2}} + \sum_{k=0}^{q-p} C_k z^{-k} \quad q \geq p \quad (3.13)$$

Donde $K = \frac{N}{2}$, y $B_{k,0}$, $B_{k,1}$, $A_{k,1}$ y $A_{k,2}$ son números reales que representan los coeficientes de las secciones de segundo orden. La entrada $x(n)$ del filtro está disponible para toda sección *biquad*, la salida de cada sección es sumada para formar la salida $y(n)$ del filtro (figura 3.5).

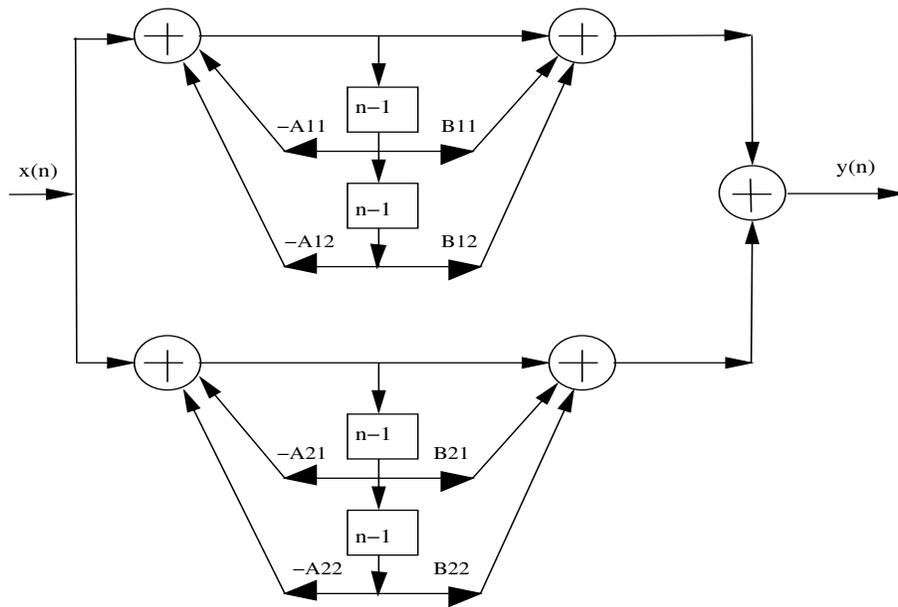


Figura 3.5: Filtro IIR de cuarto orden en paralelo

3.3. Métodos de diseño de filtros IIR

Existen varios métodos para el diseño de filtros IIR, cada uno tiene sus ventajas y desventajas como se mencionará en adelante, sin embargo, uno de los métodos más utilizados es la transformación analógica a digital a través de transformada bilineal. Entre los métodos de diseño tenemos:

■ **Indirectos o transformaciones de filtros analógicos a digitales.**

Estos métodos están basados en el desarrollo de funciones de transferencia digitales $H(z)$ a partir de funciones de transferencia analógicas $H(s)$ normalizadas que corresponden a especificaciones de un filtro analógico. Estas transformaciones se deducen en tal sentido que se preserven las características de los filtros analógicos.

- **Transformada bilineal (TB).** Es el método más utilizado en la práctica, la cual preserva las características de la función de transferencia analógica $H(s)$ al convertirla en la función de transferencia $H(z)$.
- **Transformaciones por impulso invariante.** En esta transformación se preserva la forma de la respuesta al impulso $h(t)$ de un filtro analógico, esta función se muestrea para obtener la respuesta al impulso $h(nT)$ del filtro digital correspondiente.
- **Aproximación por diferencias finitas.** En este caso la representación de la ecuación diferencial de un sistema analógico se convierte a una ecuación en diferencias de un sistema discreto correspondiente utilizando diferencias finitas.
- **Transformación Z-matched.** Se mapean directamente los polos y ceros de la función $H(s)$ en el plano z .

■ **Directos:**

Con estos métodos, el diseño se realiza en el plano z para obtener directamente la función de transferencia del filtro digital $H(z)$.

- **Diseño en el plano z :** Se realiza por la localización de polos y ceros en el plano z y luego se obtiene la función de transferencia $H(z)$. Para este método se requiere que el diseñador tenga mucha experiencia en el diseño de filtros digitales.
- **Minimización del error:** Se propone una función de transferencia del sistema discreto $H(z)$ tal que su respuesta en frecuencia se aproxime a un filtro ideal $H_I(z)$, el error entre estas dos funciones se minimiza a través de algún criterio de optimización.
- **Por fórmulas:** En algunos manuales de diseño de filtros digitales se pueden encontrar fórmulas parametrizadas, que se utilizan directamente introduciendo las especificaciones y calculando los parámetros.

3.3.1. Transformación de filtros analógicos-digitales

Las técnicas más sistemáticas de aproximación de filtros IIR están basadas en obtener una función adecuada de un filtro analógico y transformarlo al dominio discreto, procedimiento que no es posible cuando se diseña un filtro FIR. Recordando que un sistema analógico lineal e invariante con función de transferencia $H(s)$ es estable si todos sus polos están ubicados

en la mitad izquierda del plano s , y un sistema discreto con función de transferencia $H(z)$ es estable si sus polos están dentro del círculo unitario en el plano z , entonces al aplicar alguna técnica de transformación del plano s al plano z debe cumplir con las propiedades deseables:

- El eje $j\Omega$ en el plano s debe ser mapeado sobre el círculo unitario en el plano z , donde Ω se refiere a la frecuencia analógica.
- La mitad izquierda del plano s debe quedar mapeada dentro del círculo unitario en el plano z

El procedimiento de diseño clásico de filtros analógicos a filtros digitales incluye [3]:

1. Utilización de prototipos analógicos normalizados: Butterworth, Chebyshev, Elípticos o Bessel.
2. Fórmulas de transformación de bandas de frecuencia de filtros paso bajas a filtros FPA, FPBW y FSBW.
3. Utilización de algún mapeo o transformación de los polos en el plano s a los polos localizados en el plano z . En este método, dada una función de transferencia analógica $H(s)$, sus polos son mapeados directamente en los polos de la función $H(z)$, haciendo una simple sustitución.

Relación entre funciones de transferencia $H(s)$ y $H(z)$

En la figura 3.6 se observan las posibilidades de llegar a obtener la función de transferencia de $H(z)$ de un sistema discreto a partir de conocer las características de un sistema $H(s)$ continuo. Esto implica que existe una relación directa entre las transformadas de Laplace (TL), Fourier (TF) y TZ .

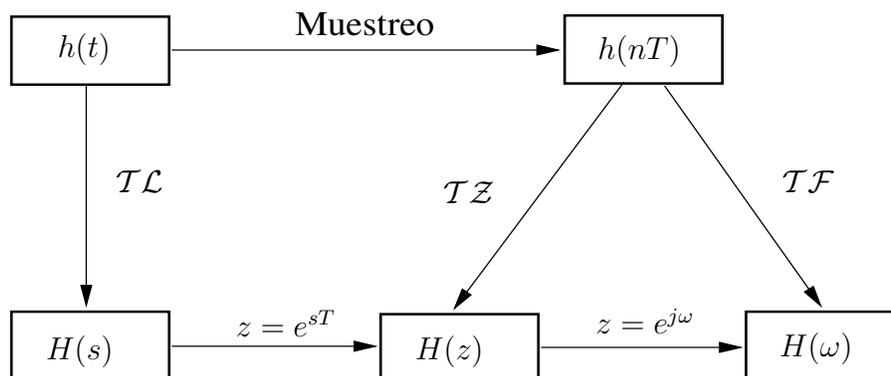


Figura 3.6: Transformación de $H(s)$ a $H(z)$

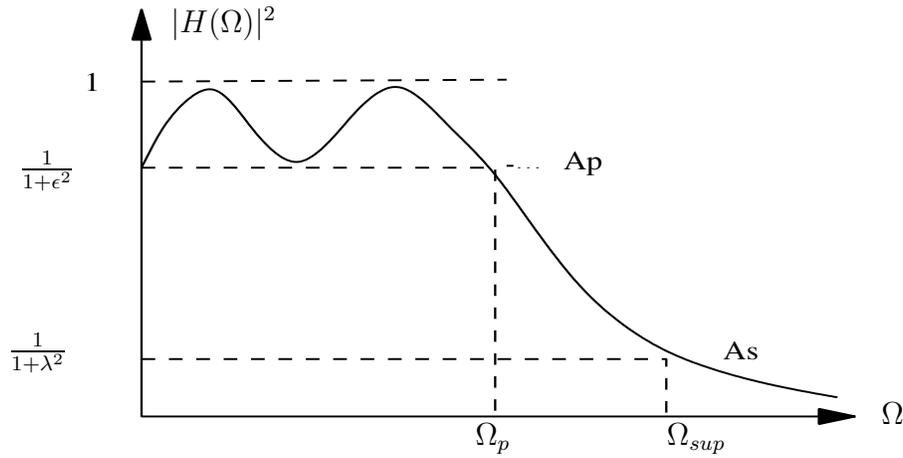


Figura 3.7: Especificaciones de un FPB analógico

En la figura 3.7 se ilustran las especificaciones de la magnitud al cuadrado de la respuesta en frecuencia de un filtro analógico (Ω se refiere a la frecuencia analógica).

Los parámetros a tomar en cuenta en el diseño son:

- ϵ : Rizo en la banda de paso.
- λ : Rizo en la banda suprimida.
- Ω_p : Frecuencia de corte en la banda de paso.
- Ω_{sup} : Frecuencia de corte en la banda suprimida.

Donde se debe satisfacer:

$$|H_a(j\Omega)|^2 = \frac{1}{1 + \epsilon^2} \quad \text{en} \quad \Omega = \Omega_p$$

$$|H_a(j\Omega)|^2 = \frac{1}{A^2} \quad \text{en} \quad \Omega = \Omega_s$$

$$\frac{1}{1 + \epsilon^2} \leq |H_a(j\Omega)|^2 \leq 1, \quad |\Omega| \leq |\Omega_p|$$

$$0 \leq |H_a(j\Omega)|^2 \leq \frac{1}{A^2}, \quad \Omega_s \leq |\Omega|$$

Relaciones en decibeles

$$A_p = -10 \log_{10} \left(\frac{1}{1 + \epsilon^2} \right) \quad \Rightarrow \quad \epsilon = \sqrt{10^{A_p/10} - 1} \quad (3.14)$$

$$A_s = -10 \log_{10} \left(\frac{1}{A^2} \right) \quad \Rightarrow \quad A = \sqrt{10^{A_s/10}} \quad (3.15)$$

Los rizados δ_1 y δ_2 en la banda de paso y banda suprimida respectivamente están en escalas absolutas:

$$\frac{1 - \delta_1}{1 + \delta_1} = \sqrt{\frac{1}{1 + \epsilon^2}} \quad \Rightarrow \quad \epsilon = \frac{2\sqrt{\delta_1}}{1 - \delta_1} \quad (3.16)$$

$$\frac{\delta_2}{1 + \delta_1} = \frac{1}{A} \quad \Rightarrow \quad A = \frac{1 + \delta_1}{\delta_2} \quad (3.17)$$

Procedimiento de diseño

1. Proponer especificaciones en frecuencia analógica A_s , A_p , Ω_c , Ω_{sup} y f_s .
2. Calcular el orden n del filtro de acuerdo al prototipo a utilizar.
3. Buscar el polinomio $H(s)$ en tablas.
4. Efectuar la transformación de $H(s)$ a $H(z)$.
5. Normalizar $H(z)$ para un filtro IIR.
6. Seleccionar una estructura para implementar el filtro.
7. Encontrar $y(n) = TZI\{Y(z)\} = TZI\{X(z)H(z)\}$.
8. Implementar la ecuación o ecuaciones en diferencias en algún simulador, un lenguaje de alto nivel o en una arquitectura para Procesamiento Digital de Señales.

3.3.2. Prototipos analógicos o aproximaciones polinómicas

Los prototipos analógicos existentes son funciones de transferencia normalizadas $H(s)$ y nos dan una muy buena aproximación de la función de transferencia $H(\Omega)$ a la función ideal $H_I(\Omega)$. En el diseño de filtros analógicos se utilizan ampliamente las aproximaciones polinómicas siguientes:

- Butterworth
- Chebyshev tipo I y II
- Elípticos o de Cauer
- Bessel

3.3.3. Prototipo Butterworth paso bajas

La aproximación Butterworth es conocida de máxima respuesta plana, y es caracterizada porque su respuesta es plana en la banda de paso y en la banda suprimida. La magnitud de $H(\Omega)$ decrece monótonamente en $\Omega \geq \Omega_p$, para obtener una banda de transición angosta se requiere un orden alto, sin embargo, algunos diseños con un orden $n \geq 3$ son bastante aceptables como se observa en la figura 3.8. La magnitud de su respuesta en frecuencia está dada por [9]

$$|H(j\Omega)|^2 = \frac{1}{1 + (\frac{\Omega}{\Omega_c})^{2n}} \quad (3.18)$$

Donde

n : es el orden del filtro

Ω_p : es la frecuencia de corte en la banda de paso, comúnmente se diseña a -3db

$A_p = 1/(1 + \epsilon^2)$: magnitud de la banda de paso a -3db

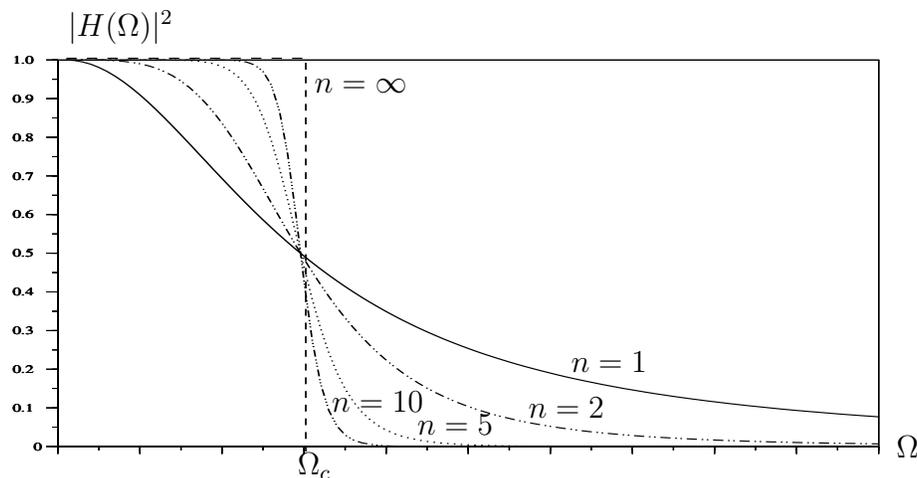


Figura 3.8: Prototipo de un filtro Butterworth paso bajas en función del orden n

De la figura 3.8 se observa que en:

- $\Omega = 0$, $|H(j\Omega)|^2 = 1$ para todo n
- $\Omega = \Omega_c$, $|H(j\Omega)|^2 = 0.5$ para todo n con una atenuación de -3db
- $|H(j\Omega)|$ es una función monótona decreciente para todo Ω
- $|H(j\Omega)|$ se aproxima a un filtro ideal si $n \rightarrow \infty$
- $1 \geq |H(\Omega_p)| \geq 1 - \delta_p$ para $|\Omega| \leq |\omega_p|$

- $|H(\Omega_{sup})| \leq \delta_{sup}$ para $|\Omega| \geq |\omega_{sup}|$

La función de transferencia está definida

$$H(s) = \frac{1}{(s - s_1)(s - s_2) \cdots (s - s_n)}$$

$$H(s) = \frac{1}{s^n + a_{n-1}s^{n-1} + \cdots + a_1s + 1} \quad (3.19)$$

donde s_i son los polos en el plano s y los coeficientes a_i son números reales.

Evaluando la función $|H(j\Omega)|^2$ como función de s para encontrar los polos y ceros, si $s = j\Omega$, además $|H(j\Omega)|^2 = H(s)H(-s)$ y $\Omega = s/j$, se tiene

$$|H(j\Omega)|_{\Omega=s/j}^2 = \frac{1}{1 + (\frac{s}{j\Omega_c})^{2n}} = \frac{(j\Omega_c)^{2n}}{s^{2n} + (j\Omega_c)^{2n}} \quad (3.20)$$

Las raíces del denominador son los polos de $H(j\Omega)$

$$p_k = (-1)^{\frac{1}{2n}} (j\Omega_c) = \Omega_c e^{j\frac{\pi}{2n}(2k+n+1)} \quad k = 0, 1, 2, \dots, 2n - 1$$

En resumen la aproximación Butterworth tiene las características:

- Existen $2n$ polos de $H(s)H(-s)$ distribuidos en el círculo de radio Ω_c e igualmente espaciados en π/n radianes. Los polos son simétricos respecto al eje real e imaginario, sin embargo, sólo los n polos que están ubicados en el semiplano izquierdo forman la función de transferencia $H(s)$ estable.
- Si n es par los polos son: $p_k = \Omega_c e^{j\frac{\pi k}{2n}} = \Omega_c \left(\cos(\frac{\pi k}{2n}) + j \operatorname{sen}(\frac{\pi k}{2n}) \right)$ donde $k = 0, 1, 2, 3, \dots, 2n - 1$
- Si n es impar los polos son: $p_k = \Omega_c e^{j(\frac{\pi}{2n} + \frac{\pi k}{n})}$, donde $k = 0, 1, \dots, 2n - 1$
- Los polos nunca están en el eje imaginario ($p_k \neq j\Omega_c$). En ambos casos si n es par o impar los polos son simétricos respecto al eje real.
- Si n es impar uno de los polos está en el eje real.
- Un filtro con aproximación Butterworth está completamente caracterizado por los parámetros n , ϵ , δ y la relación Ω_{sup}/Ω_c
- Los polos en la parte izquierda del plano s se ubican en

$$p_k = \Omega_c \left(-\operatorname{sen}\left(\frac{(2k+1)\pi}{2n}\right) + j \operatorname{cos}\left(\frac{(2k+1)\pi}{2n}\right) \right)$$

$$k = 0, 1, \dots, n - 1$$

- La función de transferencia se puede escribir

$$H(s) = (-1)^n \prod_{k=0}^{n-1} \frac{p_k}{s - p_k}$$

Cálculo del orden n de un filtro tipo Butterworth

Un primer paso en el diseño de un filtro IIR consiste en calcular el orden n de la función de transferencia $H(s)$ de las especificaciones analógicas. Este cálculo depende del tipo de aproximación que se desee y de las especificaciones deseadas tales como:

- A_p ganancia en decibels de la banda de paso
- A_s ganancia o atenuación en decibels de la banda suprimida
- f_c frecuencia de corte en Hertz
- f_p frecuencia de corte en banda de paso en Hertz
- f_{sup} frecuencia de corte en banda suprimida en Hertz
- f_s frecuencia de muestreo en Hertz.

De la figura 3.9 se pueden evaluar las magnitudes de las ganancias en la banda de paso justamente en la respuesta de corte utilizando la ecuación (3.18).

Para las especificaciones de un FPB, si $\Omega = \Omega_p$ en la banda de paso

$$A_p = -10 \log_{10} \left(\frac{1}{1 + \left(\frac{\Omega_p}{\Omega_c}\right)^{2n}} \right)$$

$$A_p = 10 \log_{10} \left(1 + \left(\frac{\Omega_p}{\Omega_c}\right)^{2n} \right)$$

$$\left(\frac{\Omega_p}{\Omega_c}\right)^{2n} = 10^{A_p/10} - 1$$

para la banda suprimida $\Omega = \Omega_s$

$$A_s = -10 \log_{10} \left(\frac{1}{1 + \left(\frac{\Omega_{sup}}{\Omega_c}\right)^{2n}} \right)$$

$$A_s = 10 \log_{10} \left(1 + \left(\frac{\Omega_{sup}}{\Omega_c}\right)^{2n} \right)$$

$$\left(\frac{\Omega_{sup}}{\Omega_c}\right)^{2n} = 10^{A_{sup}/10} - 1$$

Resolviendo para n

$$n > \frac{\log_{10}\left(\frac{10^{0.1A_s}-1}{10^{0.1A_p}-1}\right)}{2\log_{10}\left(\frac{\Omega_{sup}}{\Omega_c}\right)} \quad (3.21)$$

Para esta fórmula, A_p y A_s están en valores positivos de decibels, y A_p generalmente se diseña para -3 db, n también se puede escribir en términos de logaritmos naturales

$$n > \frac{\ln\left(\frac{e^{0.23A_s}-1}{e^{0.23A_p}-1}\right)}{2\ln\left(\frac{\Omega_{sup}}{\Omega_c}\right)} \quad (3.22)$$

Las ecuaciones (3.21) y (3.22) son válidas para el cálculo de filtros FPB Butterworth. Para el cálculo de filtros FPA Butterworth se debe invertir el cociente del logaritmo en el denominador de estas ecuaciones, es decir, escribir $\frac{\Omega_c}{\Omega_{sup}}$.

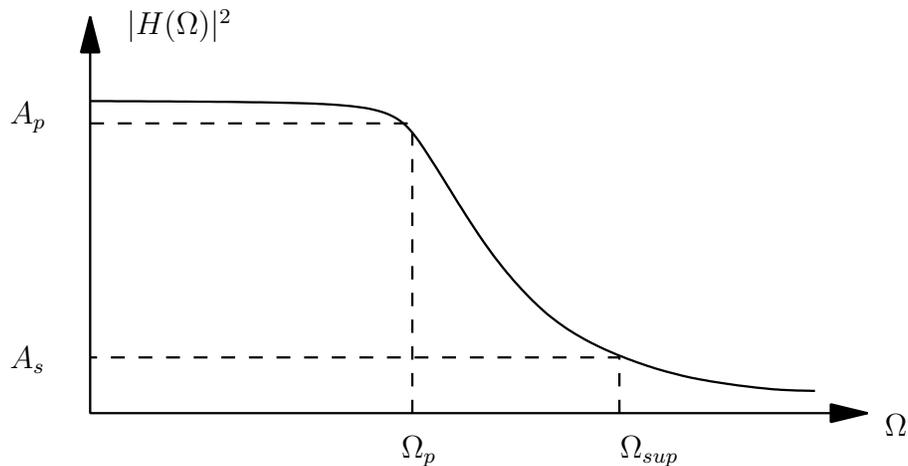


Figura 3.9: Parámetros para el cálculo del orden n de un filtro Butterworth

Polinomios de Butterworth

La tabla 3.1 no se hace muy extensa ya que un filtro IIR no utiliza valores de n muy altos, además la factorización se hace en polinomios de segundo orden para implementarlos en estructuras biquad. Esta tabla se puede generar fácilmente si se ubican los polos simétricamente sobre el semicírculo unitario izquierdo en el plano s .

3.3.4. Prototipo Chebyshev

Este tipo de aproximación contiene rizo en la banda de paso o suprimida, sin embargo, se pueden obtener filtros de menor orden que un diseño con aproximación Butterworth con las

Orden	Polinomio
1	$s + 1$
2	$s^2 + \sqrt{2}s + 1$
3	$(s^2 + s + 1)(s + 1)$
4	$(s^2 + 0.76537s + 1)(s^2 + 1.84776s + 1)$
5	$(s + 1)(s^2 + 0.61803s + 1)(s^2 + 1.61803s + 1)$
6	$(s^2 + 0.51764s + 1)(s^2 + 1.41421s + 1)(s^2 + 1.93185s + 1)$

Tabla 3.1: Polinomios de Butterworth

mismas especificaciones, es decir, que la aproximación tipo Chebyshev mejora la pendiente en la banda de transición para un mismo orden n a costa de introducir rizos en las bandas de paso o suprimida, por tanto en su diseño siempre va a existir un rizo ϵ , entre más alto sea el orden se mejora la pendiente pero crece la cantidad de rizo. Sus polos se localizan simétricamente sobre una semielipse circunscrita en un semicírculo en el plano s , esto se deducirá analíticamente. En cuanto a su respuesta en fase son más no lineales que los tipo Butterworth.

Esta aproximación se clasifica a la vez en:

- *Filtro Chebyshev I*, contiene sólo polos, exhibe rizo en la banda de paso y su función de transferencia es monótona en la banda de suprimida.
- *Filtro Chebyshev II*, contiene polos y ceros, exhibe rizo en la banda suprimida y su función de transferencia es monótona en la banda de paso. Los ceros de $H(s)$ se ubican en el eje imaginario del plano s , también es conocido como filtro inverso de Chebyshev.

Filtro Chebyshev I

Su función de transferencia sólo contiene polos y la magnitud de la respuesta en frecuencia de un filtro tipo Chebyshev I está dada por [9]

$$|H(j\Omega)|^2 = \frac{1}{1 + \epsilon^2 C_n^2\left(\frac{\Omega}{\Omega_p}\right)} \quad (3.23)$$

en variable s

$$|H(s)|^2 = \frac{1}{1 + \epsilon^2 C_n^2(-js)} \quad (3.24)$$

donde

n : es el orden del filtro.

$\epsilon = \sqrt{10^{Ap/10} - 1}$, es la magnitud del rizo en la banda de paso relacionado con la ganancia

en la banda de paso A_p .

$C_n(x)$: es el polinomio de Chebyshev de orden n , donde $x = \frac{\Omega}{\Omega_c}$ y está dado por:

$$C_n(x) = \begin{cases} \cos(n \cos^{-1}(x)) & 0 \leq x \leq 1 \\ \cosh(ncosh^{-1}(x)) & 1 < x < \infty \end{cases} \quad (3.25)$$

El rizo se debe a la naturaleza de los polinomios $C_n(x)$. De la magnitud de la función de transferencia al cuadrado se pueden resaltar algunas características:

- Para $0 < x < 1$, $C_n(x)$ oscila entre -1 y 1.
- para $1 < x < \infty$, $C_n(x)$ se incrementa monóticamente a ∞ .
- $C_n(1) = 1$ para todo n .
- Todas las raíces de los polinomios $C_n(x)$ ocurren en el intervalo $-1 \leq x \leq 1$

si $\Omega_c = 1$ y si $u = \cos^{-1}(\Omega)$, entonces $C_n(\Omega) = \cos(nu)$
de la identidad trigonométrica

$$\cos[(n+1)u] = 2 \cos(nu) \cos(u) - \cos((n-1)u)$$

desarrollando para diferentes valores de n , se puede obtener la fórmula recursiva

$$C_{n+1}(\Omega) = 2\Omega C_n(\Omega) - C_{n-1}(\Omega) \quad n = 1, 2, 3, 4... \quad (3.26)$$

que genera los polinomios $C_n(\Omega)$, para el desarrollo de polinomios de orden superior se deben considerar las condiciones iniciales $C_0(\Omega) = 1$ y $C_1(\Omega) = \Omega$ como se observa en la tabla 3.2.

Orden	Polinomio $C_n(\Omega)$
0	1
1	Ω
2	$2\Omega^2 - 1$
3	$4\Omega^3 - 3\Omega$
4	$8\Omega^4 - 8\Omega^2 + 1$
5	$16\Omega^5 - 20\Omega^3 + 5\Omega$

Tabla 3.2: Polinomios de Chebyshev $C_n(\Omega)$

Además existen dos formas posibles del comportamiento del rizo para $|H(j\Omega)|^2$ dependiendo si n es par o impar, tal como se observa en la figura 3.10, donde $x = \frac{\Omega}{\Omega_c}$ es la frecuencia normalizada a la frecuencia de corte Ω_c . De la figura 3.10 se puede deducir que:

- En $x = 0$ ($\Omega = 0$)
 $|H(j0)|^2 = 1$ si n =impar
 $|H(j0)|^2 = \frac{1}{1+\epsilon^2}$ si n =par
- En $x = 1$ ($\Omega = \Omega_c$)
 $|H(jx)|^2 = \frac{1}{1+\epsilon^2}$, para n par o impar
- Para $0 \leq x \leq 1$ ($0 \leq \Omega \leq \Omega_c$)
 $|H(jx)|^2$ oscila entre 1 y $\frac{1}{1+\epsilon^2}$
- Si $x > 1$ ($\Omega > \Omega_c$)
 $|H(jx)|^2$ decrece monótonamente a cero
- En $x = \Omega_{sup}$
 $|H(jx)|^2 = \frac{1}{A^2}$

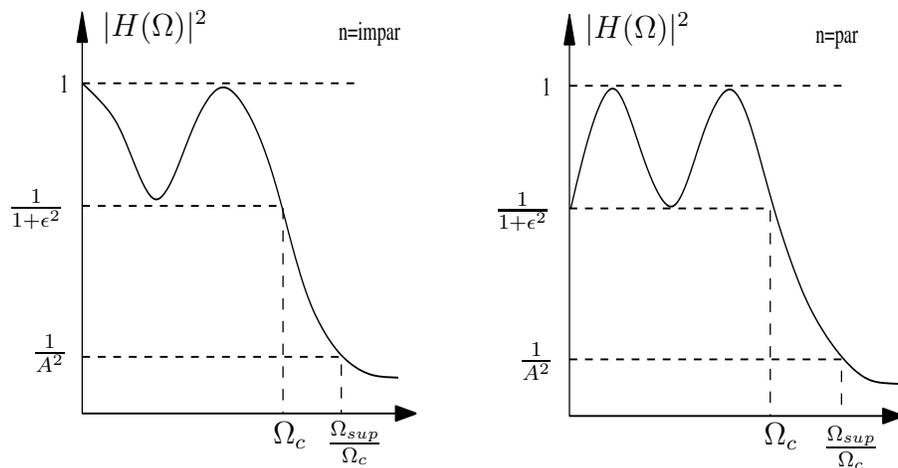


Figura 3.10: Respuesta en frecuencia de la aproximación Chebyshev I

Localización de los polos para filtros Chebyshev I

Igualando a cero el denominador de la ecuación (3.24), se obtienen las raíces

$$1 + \epsilon^2 C_n^2(-js) = 0 \implies C_n(-js) = \pm j/\epsilon \quad (3.27)$$

sustituyendo $s = \sigma + j\Omega$ y la definición de los polinomios $C_n(x)$

$$\cos(ncos^{-1}(\Omega - j\sigma)) = \pm j/\epsilon \quad (3.28)$$

definiendo $\cos^{-1}(\Omega - j\sigma) = x + jy$,

$$\cos(nx + jny) = \pm j/\epsilon \quad (3.29)$$

por identidades trigonométricas

$$\cos(nx)\cos(jny) - \operatorname{sen}(nx)\operatorname{sen}(jny) = \pm j/\epsilon \quad (3.30)$$

dado que $\cos(jx) = \cosh(x)$ y $\operatorname{sen}(jx) = j \operatorname{senh}(y)$, nos queda

$$\cos(nx)\cosh(ny) - j\operatorname{sen}(nx)\operatorname{senh}(ny) = \pm j/\epsilon \quad (3.31)$$

igualando las partes reales e imaginarias

$$\cos(nx)\cosh(ny) = 0 \quad \text{y} \quad \operatorname{sen}(nx)\operatorname{senh}(ny) = \pm 1/\epsilon \quad (3.32)$$

donde $\cosh(ny) \neq 0$ y $\operatorname{sen}(nx) = 1$, entonces $\cos(nx) = 0$, por lo que

$$x = (2k + 1)\frac{\pi}{2n}, \quad k = 0, 1, 2, \dots, 2n - 1 \quad (3.33)$$

$$y = \pm \frac{1}{n} \operatorname{senh}^{-1}(1/\epsilon) \quad (3.34)$$

Entonces las partes reales e imaginarias de la localización de los polos son

$$\sigma_k = \pm \operatorname{sen}\left((2k + 1)\frac{\pi}{2n}\right) \operatorname{senh}\left(\frac{1}{n} \operatorname{senh}^{-1}(1/\epsilon)\right) \quad (3.35)$$

$$\Omega_k = \cos\left((2k + 1)\frac{\pi}{2n}\right) \cosh\left(\frac{1}{n} \operatorname{senh}^{-1}(1/\epsilon)\right) \quad (3.36)$$

para $k = 0, 1, 2, \dots, 2n - 1$ los polos están distribuidos sobre la elipse

$$\frac{\sigma_k}{\operatorname{senh}^2 y} + \frac{\Omega_k}{\cosh^2 y} = 1 \quad \text{y} \quad y = \frac{1}{n} \operatorname{senh}^{-1}(1/\epsilon) \quad (3.37)$$

donde el eje menor es σ (real) y el mayor $j\Omega$ (imaginario).

Cálculo del orden n

De las especificaciones $A_p = 10 \log(1 + \epsilon^2)$ y $\epsilon = \sqrt{10^{0.1A_p} - 1}$

calculando la magnitud de A_{sup}

$$A_{sup} = 10 \log \left(1 + \epsilon^2 C_n^2 \left(\frac{\Omega_{sup}}{\Omega_p} \right) \right) \quad (3.38)$$

sustituyendo el polinomio C_n para $\Omega > \Omega_{sup}$

$$A_{sup} = 10 \log \left(1 + \epsilon^2 \cosh^2 \left(n \cosh^{-1} \left(\frac{\Omega_{sup}}{\Omega_p} \right) \right) \right) \quad (3.39)$$

por lo tanto

$$\cosh^2 \left(n \cosh^{-1} \left(\frac{\Omega_{sup}}{\Omega_p} \right) \right) = \frac{10^{0.1A_{sup}} - 1}{10^{0.1A_p} - 1} \quad (3.40)$$

finalmente despejando n

$$n > \frac{\cosh^{-1} \left(\frac{\sqrt{10^{0.1A_{sup}} - 1}}{\sqrt{10^{0.1A_p} - 1}} \right)}{\cosh^{-1} \left(\frac{\Omega_{sup}}{\Omega_p} \right)} \quad (3.41)$$

Filtros Chebyshev II

Esta aproximación también es conocida como Chebyshev inverso y tiene una magnitud plana en la banda de paso y rizo en la banda suprimida. La función de transferencia $H(s)$ tiene polos y ceros. La magnitud al cuadrado de la función de transferencia de un filtro tipo Chebyshev está dada por

$$|H(j\Omega)|^2 = \frac{\epsilon^2 C_n^2 \left(\frac{\Omega_{sup}}{\Omega} \right)}{1 + \epsilon^2 C_n^2 \left(\frac{\Omega_{sup}}{\Omega} \right)} \quad (3.42)$$

con atenuación

$$A_s(\Omega) = 10 \log \left(1 + \frac{1}{\epsilon^2 C_n^2 \left(\frac{\Omega_{sup}}{\Omega} \right)} \right) \text{ db} \quad (3.43)$$

calculando ϵ , en $\Omega = \Omega_{sup}$ al inicio de la banda suprimida y de la tabla 3.2 se puede comprobar que $C_n(1) = 1$

$$A_s(\Omega) = 10 \log \left(1 + \frac{1}{\epsilon^2 C_n^2(1)} \right) \implies \epsilon = \frac{1}{\sqrt{10^{0.1A_s} - 1}} \quad (3.44)$$

En la banda de paso la máxima atenuación permitida A_p no puede exceder $\Omega = \Omega_p$, entonces

$$A_p = 10 \log \left(1 + \frac{1}{\epsilon^2 C_n^2 \left(\frac{\Omega_{sup}}{\Omega_p} \right)} \right) \quad (3.45)$$

el orden se calcula de igual forma que para el filtro Chebyshev I. Los polos y los ceros pueden calcularse analíticamente por medio de

$$|H(s)|^2 = H(s)H(-s) = \frac{\epsilon^2 C_n^2 \left(\frac{j\Omega_{sup}}{s} \right)}{1 + \epsilon^2 C_n^2 \left(\frac{j\Omega_{sup}}{s} \right)} \quad (3.46)$$

Los ceros se pueden calcular igualando a cero el numerador de la función de transferencia

$$C_n\left(\frac{j\Omega_{sup}}{s}\right) = 0 \quad \text{o} \quad \cos\left(ncos^{-1}\left(\frac{j\Omega_{sup}}{s}\right)\right) = 0$$

que satisface

$$\cosh^{-1}\left(\frac{j\Omega_{sup}}{s}\right) = \frac{m\pi}{2n} \quad m \text{ impar}$$

entonces los ceros están ubicados en

$$s_m = j\Omega_{sup} \sec\left(\frac{m\pi}{2n}\right) \quad m = 1, 3, \dots, 2n - 1 \quad (3.47)$$

La localización de los polos se obtienen calculando las raíces del polinomio

$$1 + \epsilon^2 C_n^2\left(\frac{\Omega_{sup}}{\Omega}\right) = 0 \quad (3.48)$$

de forma similar a Chebyshev I, sólo que hay que reemplazar $-s$ por $1/s$. Es decir, que los polos de Chebyshev II son los recíprocos de los del I, entonces si s_p es un polo del tipo I, $1/s_p$ es un polo de tipo II. Los polos del tipo II también están ubicados sobre una elipse [9], [27].

Implementación de un filtro Chebyshev I

Como es conocido para un sistema continuo con función de transferencia $H(s)$ causal y estable, los polos de $H(s)H(-s)$ deben estar en la parte izquierda del plano s , estos polos son obtenidos al hallar las raíces del polinomio del denominador

$$p(x) = 1 + \epsilon^2 C_n^2\left(\frac{s}{j\Omega_c}\right) \quad (3.49)$$

Para su solución se puede comprobar que si $p_k = \sigma_k + j\Omega_k$, donde $k = 0, 1, 2, \dots, n - 1$, son las raíces del polinomio $p(x)$, entonces cada una de las variables son

$$\sigma_k = (a\Omega_c) \cos\left(\frac{\pi}{2} + \frac{(2k+1)\pi}{2n}\right) \quad \Omega_k = (b\Omega_c) \sin\left(\frac{\pi}{2} + \frac{(2k+1)\pi}{2n}\right)$$

para $k = 0, 1, 2, 3, \dots, n - 1$

$$a = \frac{1}{2} \left(\sqrt[n]{\alpha} - \sqrt[n]{\frac{1}{\alpha}} \right) \quad b = \frac{1}{2} \left(\sqrt[n]{\alpha} + \sqrt[n]{\frac{1}{\alpha}} \right)$$

$$\alpha = \frac{1}{\epsilon} + \sqrt{1 + \frac{1}{\epsilon^2}} = \frac{1 + \sqrt{\epsilon^2 + 1}}{\epsilon} = \beta^n$$

Las raíces están ubicadas sobre una elipse con eje mayor $b\Omega_c$ y eje menor $a\Omega_c$. Una vez calculado los polos p_k , éstos se sustituyen en la siguiente función $H(s)$ de Chebyshev I

$$H(s) = K \prod_{k=0}^{n-1} \frac{p_k}{(s - p_k)} \quad (3.50)$$

Donde K es un factor de normalización elegido de la siguiente forma

$$K = \begin{cases} -1 & , \text{ si } n \text{ impar} \\ \frac{1}{\sqrt{1+\epsilon^2}} & , \text{ si } n \text{ par} \end{cases} \quad (3.51)$$

Con las ecuaciones (3.50) y (3.51) se pueden generar los polinomios de Chebyshev I para cualquier rizo ϵ y orden n .

Dadas las especificaciones del filtro, Ω_p , Ω_{sup} , A_p y A_s , se requieren tres parámetros para determinar un filtro Chebyshev I: ϵ , Ω_c y n . Las ecuaciones que relacionan estos parámetros son

$$\epsilon = \sqrt{10^{0.1A_p} - 1} \quad y \quad A = 10^{A_s/10} \quad (3.52)$$

donde $\Omega_c = \Omega_p$ para el orden n

$$n = \frac{\log_{10}(l + \sqrt{l^2 - 1})}{\log_{10}(\Omega_r + \sqrt{\Omega_r^2 - 1})} \quad l = \sqrt{(A^2 - 1)/\epsilon^2} \quad (3.53)$$

de la gráfica 3.11 se tienen las siguientes variables

$$\epsilon = \sqrt{10^{0.1A_p} - 1} \quad , \quad \lambda = \sqrt{10^{0.1A_s} - 1}$$

$$K_0 = \frac{\Omega_p}{\Omega_{sup}} \quad y \quad A_p = 10 \log(1 + \epsilon^2)$$

Para la banda suprimida

$$A_s = 10 \log \left(1 + \epsilon^2 C_n^2 \left(\frac{\Omega_{sup}}{\Omega_p} \right) \right)$$

sustituyendo el polinomio $C_n^2(\Omega)$

$$A_s = 10 \log \left(1 + \epsilon^2 \cosh^2 \left(n \cosh^{-1} \left(\frac{\Omega_{sup}}{\omega_p} \right) \right) \right)$$

y el orden n para un FPB se puede calcular por la ecuación

$$n > \frac{\cosh^{-1} \left(\frac{\sqrt{10^{0.1A_s} - 1}}{\sqrt{10^{0.1A_p} - 1}} \right)}{\cosh^{-1} \left(\frac{\Omega_{sup}}{\Omega_p} \right)} \quad (3.54)$$

Para efectos de cálculo de las funciones seno y coseno hiperbólico se pueden utilizar las identidades

$$\begin{aligned} \operatorname{senh}^{-1}(x) &= \ln \left[x + \sqrt{x^2 + 1} \right] \\ \operatorname{cosh}^{-1}(x) &= \ln \left[x + \sqrt{x^2 - 1} \right] \\ \cos(jx) &= \operatorname{cosh}(x) \quad y \quad \operatorname{sen}(jx) = j \operatorname{senh}(y) \\ \operatorname{senh}(x) &= \frac{e^x - e^{-x}}{2} \quad y \quad \cos(x) = \frac{e^x + e^{-x}}{2} \end{aligned}$$

Los polos están ubicados en

$$\begin{aligned} s_k &= \operatorname{sen} \left[(2k + 1) \frac{\pi}{2n} \right] \operatorname{senh} \left[\frac{1}{n} \operatorname{senh}^{-1} \frac{1}{\epsilon} \right] \\ &+ j \operatorname{cos} \left[(2k + 1) \frac{\pi}{2n} \right] \operatorname{cosh} \left[\frac{1}{n} \operatorname{senh}^{-1} \frac{1}{\epsilon} \right] \end{aligned} \quad (3.55)$$

para $k = 0, 1, 2, 3, \dots, n - 1$

Es decir que se tiene calculada la ubicación de los polos de un filtro Chebyshev I y se puede obtener la función de transferencia $H(s)$.

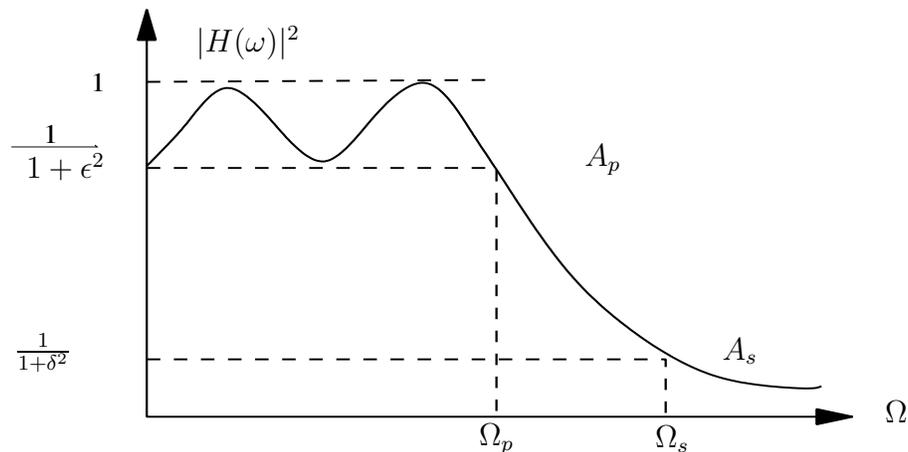


Figura 3.11: Cálculo de orden n en un filtro con aproximación Chebyshev I

Tabla de polinomios de Chebyshev

Los polinomios de Chebyshev $H(s)$ se pueden desarrollar a partir de los polinomios $C_n(\Omega)$ como se explicó anteriormente, sin embargo, ya existen tablas tabuladas para diferentes valores del rizo ϵ , en la tabla 3.3 se muestran cinco polinomios para tres rizados diferentes. Si se desea conocer los polinomios de mayor orden u otras tablas con diferente rizo ver la referencia [9].

Orden n	Polinomio para rizo de 1 db ($\epsilon = 0.50885$)
1	$s + 1.96523$
2	$s^2 + 1.09773s + 1.10251$
3	$(s^2 + 0.49417s + 0.99421)(s + 0.49417)$
4	$(s^2 + 0.27907s + 0.98651)(s^2 + 0.67374s + 0.27940)$
5	$(s + 0.28949)(s^2 + 0.17892s + 0.98832)(s^2 + 0.46841s + 0.42930)$
Orden n	Polinomio para rizo de 2 db ($\epsilon = 0.76478$)
1	$s + 1.30756$
2	$s^2 + 0.80382s + 0.82306$
3	$(s^2 + 0.36891s + 0.88610)(s + 0.36891)$
4	$(s^2 + 0.20978s + 0.92868)(s^2 + 0.50644s + 0.22157)$
5	$(s + 0.21831)(s^2 + 0.13492s + 0.95217)(s^2 + 0.35323s + 0.39315)$
Orden n	Polinomio para rizo de 3 db ($\epsilon = 0.99763$)
1	$s + 1.00238$
2	$s^2 + 0.64490s + 0.70795$
3	$(s^2 + 0.29862s + 0.83917)(s + 0.29862)$
4	$(s^2 + 0.17034s + 0.90309)(s^2 + 0.41124s + 0.19598)$
5	$(s + 0.17753)(s^2 + 0.10970s + 0.93603)(s^2 + 0.28725s + 0.37701)$

Tabla 3.3: Polinomios de Chebyshev

3.3.5. Diseño de filtros elípticos

Los filtros elípticos (o Cauer) exhiben rizados iguales tanto en la banda de paso como en la banda suprimida, su función de transferencia contiene polos y ceros, para valores de n muy bajos se logran grandes pendientes en la banda de transición, por lo tanto se puede decir que son filtros muy óptimos, pero su fase es extremadamente no lineal. Estos filtros son muy difíciles de analizar y diseñar, por lo que se utilizan tablas o programas para su diseño. La magnitud de su función de transferencia al cuadrado se escribe

$$|H(j\Omega)|^2 = \frac{1}{1 + \epsilon^2 U_n^2\left(\frac{\Omega}{\Omega_p}\right)} \quad (3.56)$$

Donde n es el orden del filtro y el parámetro ϵ está relacionado con el rizo en la banda de paso. $U(\cdot)$ es conocida como una función racional de Chebyshev y sus raíces están relacionadas a un función seno Jacobiano elíptica de orden n calculada por Zverev (1967) [32], [9]. Los ceros están ubicados en el eje $j\Omega$.

Propiedades de la función $U(\Omega)$

1. $U_n(\Omega)$ es par si n es par y es impar si n es impar
2. Los ceros de $U_n(\Omega)$ están en el intervalo $|\Omega| < 1$ y los polos en el intervalo $|\Omega| > 1$
3. La función $U_n(\Omega)$ oscila entre ± 1 en la banda de paso
4. $U_n(\Omega) = 1$ si $\Omega = 1$
5. $U_n(\Omega)$ oscila entre $\pm 1/d$ e infinito en la banda suprimida, d se define en el procedimiento de diseño.

Ecuaciones de Diseño

Cálculo del orden n

$$n = \frac{K(k)K(\sqrt{1-k_1^2})}{K(k_1)K(\sqrt{1-k^2})} \quad (3.57)$$

$$k = \frac{\Omega_p}{\Omega_{sup}} \quad (3.58)$$

$$k_1 = \frac{\epsilon}{\sqrt{A^2 - 1}} = \frac{\epsilon}{\delta} \quad (3.59)$$

$$\delta_2 = 1/\sqrt{1 + \delta} \quad (3.60)$$

la función $K(x)$ es la integral elíptica completa de primer orden definida

$$K(x) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - x^2 \text{sen}^2 \theta}} \quad (3.61)$$

los valores de esta integral ya han sido calculados y se pueden encontrar en tablas (ver Jahnke y Emde 1945 y Dwight 1957) [32]. Una forma práctica de calcular el orden n evitando la integral es [24]

$$n \approx \frac{2 \log(4/k_1)}{\log(1/q)} \quad (3.62)$$

con

$$q = q_0 + 2q_0^5 + 15q_0^9 + 150q_0^{13}$$

$$q_0 = \frac{1}{2} \left(\frac{1 - (1 - k^2)^{1/4}}{1 + (1 - k^2)^{1/4}} \right)$$

Método de cálculo de un filtro elíptico

En seguida se dan los pasos para el cálculo de un filtro elíptico realizando ciertas aproximaciones, todo este procedimiento es factible de programarlo en una computadora [9]. Para funciones de transferencia $H(s)$ normalizadas a $\Omega = 1$.

Si n par

$$H(s) = H_0 \prod_{i=1}^{n/2} \frac{s^2 + a_i}{s^2 + b_i s + c_i} \quad (3.63)$$

Si n impar

$$H(s) = \frac{H_0}{s + a} \prod_{i=1}^{(n-1)/2} \frac{s^2 + a_i}{s^2 + b_i s + c_i} \quad (3.64)$$

Calcular

1. $k = \Omega_p / \Omega_{sup}$

2.

$$q_0 = \frac{1}{2} \left(\frac{1 - (1 - k^2)^{1/4}}{1 + (1 - k^2)^{1/4}} \right) \quad (3.65)$$

3.

$$q = q_0 + 2q_0^5 + 15q_0^9 + 150q_0^{13} \quad (3.66)$$

4.

$$d = \left(\frac{10^{0.1A_p} - 1}{10^{0.1A_{sup}} - 1} \right)^{1/2} \quad (3.67)$$

5.

$$n \geq \frac{\log(16/d^2)}{\log(1/q)} \quad (3.68)$$

6.

$$\epsilon = \sqrt{10^{0.1A_p} - 1} \quad (3.69)$$

7.

$$\beta = \frac{1}{2n} \ln \frac{(1 + \epsilon^2)^{1/2} + 1}{(1 + \epsilon^2)^{1/2} - 1} \quad (3.70)$$

8.

$$a = \frac{2q^{1/4} \sum_{m=0}^{\infty} (-1)^m q^{m(m+1)} \sinh[(2m+1)\beta]}{1 + 2 \sum_{m=1}^{\infty} (-1)^m q^{m^2} \cosh[2m\beta]} \quad (3.71)$$

9.

$$U = \sqrt{(1 + ka^2)(1 + a^2/k)} \quad (3.72)$$

10.

$$\Omega_i = \frac{2q^{1/4} \sum_{m=0}^{\infty} (-1)^m q^{m(m+1)} \sinh[(2m+1)\pi l/n]}{1 + 2 \sum_{m=1}^{\infty} (-1)^m q^{m^2} \cosh[2\pi ml/n]} \quad (3.73)$$

donde

$l = i - 1/2$, $i = 1, 2, \dots, n/2$ para n par

$l = i$, $i = 1, 2, \dots, (n-1)/2$ para n impar

11.

$$V_i = \sqrt{(1 - k\Omega_i^2)(1 - \Omega_i^2/k)} \quad (3.74)$$

12.

$$a_i = \frac{1}{\Omega_i^2} \quad (3.75)$$

13.

$$b_i = \frac{2aV_i}{1 + a^2\Omega_i^2} \quad (3.76)$$

14.

$$c_i = \frac{(aV_i)^2 + (\Omega_i U)^2}{(1 + a^2\Omega_i^2)} \quad (3.77)$$

15. Finalmente

$$H_0 = a \prod_{i=1}^{(n-1)/2} \frac{c_i}{a_i} \quad n \text{ impar} \quad (3.78)$$

$$H_0 = \frac{1}{\sqrt{1 + \epsilon^2}} \prod_{i=1}^{n/2} \frac{c_i}{a_i} \quad n \text{ par} \quad (3.79)$$

para desnormalizar, sustituir s por s/Ω_0 .

Las funciones normalizadas de $U_n(\Omega)$ a la frecuencia central Ω_0 están dadas por

$$U_n(\Omega) = \Omega \prod_{i=0}^{(n-1)/2} \frac{\Omega_i^2 - \Omega^2}{1 - \Omega_i^2 \Omega^2} \quad n \text{ impar} \quad (3.80)$$

$$U_n(\Omega) = \prod_{i=0}^{n/2} \frac{\Omega_i^2 - \Omega^2}{1 - \Omega_i^2 \Omega^2} \quad n \text{ par} \quad (3.81)$$

Los polos y los ceros de $U_n(\Omega)$ son recíprocos entre sí y exhiben geometría simétrica con respecto a la frecuencia central Ω_0 , además [9]

$$U_n\left(\frac{1}{\Omega}\right) = \frac{1}{U_n(\Omega)} \quad (3.82)$$

Esto se debe a que la función fraccional de Chebyshev $U_n(\Omega)$ exhibe simetría con respecto a la frecuencia central $\Omega_0 = 1$, tal que cualquier frecuencia Ω_i en el intervalo $0 \leq \Omega \leq \Omega_0$ es el recíproco de cualquier valor de geometría simétrica con respecto a $1/\Omega_i$.

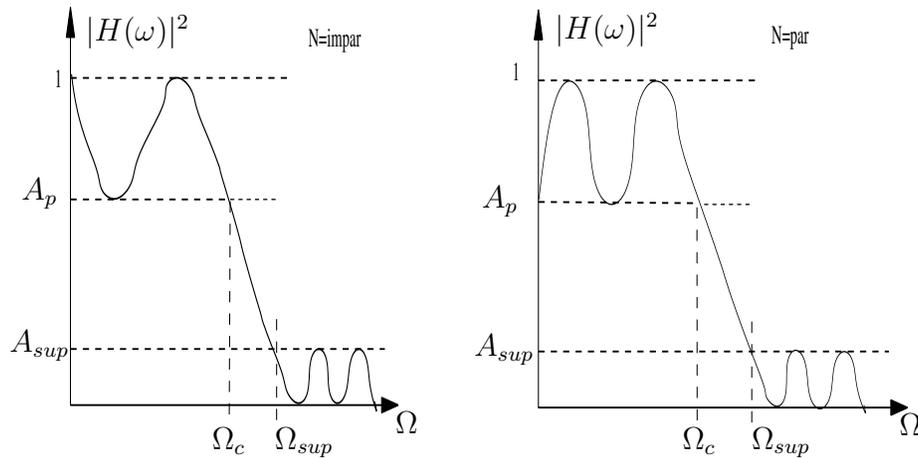


Figura 3.12: Respuesta en frecuencia de un filtro elíptico

En los filtros elípticos su respuesta en fase es todavía más no lineal en la banda de paso que los filtros tipo Butterworth y los tipo Chebyshev, por esta razón no se utilizan frecuentemente en aplicaciones prácticas, sin embargo, para un orden dado éste tipo de filtro produce una banda de transición más angosta.

3.3.6. Filtros de Bessel

Son filtros del tipo todo polo, su característica importante es su fase lineal sobre toda la banda de paso pero con una banda de transición mayor que la respuesta de un filtro Butterworth de igual orden. Este tipo de aproximación se caracteriza por la función de transferencia

$$H(s) = \frac{1}{B_N(s)} \quad (3.83)$$

donde $B(s)$ es una función polinomial de orden N que puede ser expresada como

$$B(s) = \sum_{k=0}^N a_k s^k \quad (3.84)$$

y los coeficientes a_k están dados por

$$a_k = \frac{(2N-1)!}{2^{N-k} k! (N-1)!} \quad k = 0, 1, 2, 3, \dots, N \quad (3.85)$$

Alternativamente los polinomios de Bessel pueden ser generados recursivamente de la relación

$$B_N(s) = (2N-1)B_{N-1}(s) + s^2 B_{N-2}(s) \quad (3.86)$$

En los filtros tipo Bessel se debe enfatizar que su característica de fase lineal de un filtro analógico es destruida por el proceso de conversión al dominio del filtro digital por medio de la transformación analógica a digital utilizada [32].

Como resumen de los tipos de aproximaciones se puede mencionar que para un diseño con $k = 0.75$, $A_p = 2$ db y $A_s = 60$, se necesita un orden de 25 para aproximación Butterworth, de 10 para Chebyshev y de 6 para un elíptico.

3.4. Transformaciones de filtros analógicos

En esta sección se introducen algunas transformaciones para el diseño de filtros digitales considerando que se tienen prototipos adecuados en el plano s , la idea es conocer estas transformaciones y saber sus ventajas y desventajas. Una transformación óptima del plano s al plano z , debe mapear todo el eje imaginario en s sobre el círculo unitario en el plano z , y toda la región a la izquierda del plano s dentro del círculo unitario en el plano z , esto es mapear la región estable en plano s a la región estable en el plano z . Como se verá en seguida, existen varias posibilidades de diseño de filtros IIR utilizando transformaciones, sin embargo, algunas transformaciones muestran ciertos inconvenientes.

3.4.1. Diseño de filtros IIR por respuesta al impulso invariante

Los filtros digitales IIR también se pueden diseñar si se tienen exigencias sobre la respuesta al impulso, es decir, que se busca una función de transferencia discreta de un filtro digital que emule la respuesta al impulso de un filtro analógico. Como la respuesta al impulso de un filtro analógico $h(t)$, es continua, es necesario muestrearla para obtener una respuesta al impulso discreta $h(n)$. Este procedimiento establece una correspondencia directa entre la respuesta al impulso del prototipo analógico y la respuesta al impulso del filtro digital a través del muestreo de la primera, haciendo que los valores de la respuesta al impulso $h(nT)$

del filtro digital sean iguales a los valores muestreados de la respuesta al impulso $h(t)$ del filtro analógico.

Si los coeficientes de $H(z)$ son ajustados de tal forma que su respuesta al impulso sea idéntica a la respuesta al impulso $h(n)$ especificada, entonces se dice que el filtro diseñado $H(z)$ es invariante al impulso y es equivalente al sistema analógico $H(s)$. Este método preserva la respuesta al impulso del sistema analógico, pero no la magnitud de respuesta en frecuencia debido al aliasing introducido, por lo que este método es inadecuado en el diseño de filtros paso altas y supresor de banda.

Procedimiento de diseño

1. Dado un sistema analógico con función de transferencia $H(s)$, utilizando la Transformada de Laplace inversa obtener $h(t)$

$$H(s) = \mathcal{L}\{h(t)\} \quad \text{entonces} \quad h(t) = \mathcal{L}^{-1}\{H(s)\}$$

2. Muestrear $h(t)$ a intervalos T si, $t = n/f_s = nT$

$$h(n) = h(t)|_{t=nT}$$

3. Obtener $H(z) = TZ\{h(n)\}$

4. Evaluar el desempeño de la respuesta en frecuencia $H(e^{j\omega})$

5. Por transformada Z inversa de $H(z)$ obtener la ecuación en diferencias y despejar $y(n)$

El muestreo de una señal continua tiene ciertas implicaciones, si una señal $h_a(t)$ con espectro $H_a(f_a)$ (limitado en banda), es muestreada a razón de $f_s = 1/T$ muestras por segundo, el espectro de la señal muestreada es continuo, periódico y escalado por $f_s H_a(f_a)$ con período f_s [32]

$$H(f) = f_s \sum_{k=-\infty}^{\infty} H_a[(f - k)f_s] \quad (3.87)$$

donde $f = f_a/f_s$ es la frecuencia normalizada respecto de la frecuencia de muestreo f_s . Si la frecuencia de muestreo es menor que dos veces el ancho de banda de $H_a(f_a)$ ocurre el fenómeno de "aliasing". La ecuación (3.87) se puede escribir en términos de ω

$$H(\omega) = f_s \sum_{k=-\infty}^{\infty} H_a[(\omega - 2\pi k)f_s] \quad (3.88)$$

En teoría se diseña un filtro digital con respuesta en frecuencia $H(\omega)$ con las mismas características en frecuencia de su correspondiente filtro analógico si el intervalo de muestreo T ha sido seleccionado adecuadamente para evitar los efectos de aliasing. De aquí se puede apreciar que el método de respuesta invariante al impulso es inapropiado para el diseño

de filtros paso altas debido al espectro aliasing que resulta del proceso de muestreo. En el dominio de la frecuencia, las funciones de transferencia $H(z)$ y $H(s)$ se relacionan

$$H(z) = \frac{1}{T} \sum_{k=-\infty}^{\infty} H(s - j\frac{2\pi}{T}k)$$

Esta identidad $H(\omega) = \frac{1}{T}H(\Omega)$ es válida si $H(\Omega) = 0$ para valores analógicos $|\Omega| \geq \pi/T$.

Para investigar el mapeo de los puntos entre el plano z y el plano s por el proceso del muestreo, se calcula la transformada Z de $h(n)$ aplicándole la transformada de Laplace a $h(t)$

$$H(z)|_{z=e^{sT}} = \frac{1}{T} \sum_{k=-\infty}^{\infty} H(s - j\frac{2\pi}{T}k) \quad (3.89)$$

por definición de TZ

$$H(z) = \sum_{k=0}^{\infty} h(n)z^{-n} \quad (3.90)$$

$$H(z)|_{z=e^{sT}} = \sum_{k=0}^{\infty} h(n)e^{-sTn} \quad (3.91)$$

Es decir, que los coeficientes de $H(z)$ son ajustados a los valores de la respuesta al impulso del sistema analógico $h(t)$ especificada de antemano.

Como se conoce la relación

$$z = e^{sT} \quad (3.92)$$

y

$$s = \sigma + j\Omega = e^{\sigma T} e^{j\Omega T} \text{ y } z = re^{j\omega},$$

entonces sustituyendo en la transformación

$$z = re^{j\omega} = e^{\sigma T} e^{j\Omega T}$$

e igualando términos

$$r = e^{\sigma T} \quad \text{y} \quad \omega = \Omega T$$

que cumple con:

si $\sigma < 0$ entonces $0 < r < 1$

si $\sigma > 0$ entonces $r > 1$.

si $\sigma = 0$ el eje $j\Omega$ se mapea sobre todo el círculo unitario.

Este mapeo asegura respuestas en frecuencia muy similares a los prototipos analógicos y diseños estables. Sin embargo, dado que la frecuencia digital ω está definida en el intervalo

$(-\pi, \pi)$, entonces el mapeo $\omega = \Omega T$ implica el intervalo $-\pi/T < \Omega < \pi/T$. También el intervalo $-\pi/T < \Omega < 3\pi/T$ puede mapearse en $(-\pi, \pi)$, en general cualquier intervalo $(2k - 1)\pi/T < \Omega < (2k + 1)\pi/T$, lo que significa que el mapeo de $j\Omega$ dentro del círculo unitario no es único, esto refleja los efectos de aliasing debido al muestreo. En la figura 3.13 se muestra el mapeo cíclico del plano s al plano z .

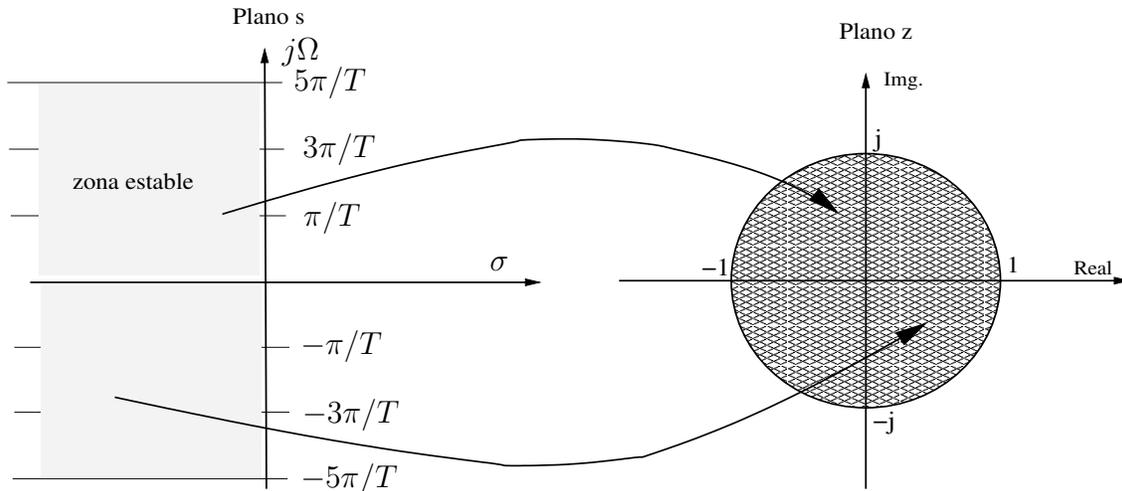


Figura 3.13: Mapeo del plano s al plano z por diseño de respuesta al impulso invariante

Adicionalmente en el diseño por el método de respuesta invariante al impulso pueden existir otros efectos, si en el dominio analógico existen varios polos p_k y si c_k son los coeficientes de la expansión en fracciones parciales

$$H(s) = \sum_{k=1}^N \frac{c_k}{s - p_k} \quad (3.93)$$

entonces por transformada de Laplace inversa

$$h(t) = \sum_{k=1}^N c_k e^{p_k t} \quad t \geq 0 \quad (3.94)$$

muestreando $h(t)$ en períodos $t = nT$

$$h(n) = \sum_{k=1}^N c_k e^{p_k T n} \quad (3.95)$$

calculando la transformada Z de $h(n)$

$$H(z) = \sum_{k=1}^{\infty} h(n) z^{-n} = \sum_{k=1}^{\infty} \left(\sum_{k=1}^N c_k e^{p_k T n} \right) z^{-n} = \sum_{k=1}^N c_k \sum_{k=1}^{\infty} (e^{p_k T} z^{-1})^n \quad (3.96)$$

como los polos $p_k < 0$, resolviendo la serie geométrica

$$\sum_{k=1}^{\infty} (e^{p_k T} z^{-1})^n = \frac{1}{1 - e^{p_k T} z^{-1}} \quad (3.97)$$

por lo tanto la función de transferencia del sistema digital queda

$$H(z) = \sum_{k=1}^N \frac{C_k}{1 - e^{p_k T} z^{-1}} \quad (3.98)$$

y los polos del filtro digital están ubicados en

$$z_k = e^{p_k T} \quad k = 1, 2, 3, \dots, N \quad (3.99)$$

Del análisis anterior se debe observar que la magnitud de la respuesta digital está escalada por $f_s = 1/T$ debido al proceso de muestreo, por lo que la función de transferencia $H(z)$ obtenida debe multiplicarse por T , por tanto

$$H(z) = T \sum_{k=1}^N \frac{C_k}{1 - e^{p_k T} z^{-1}} \quad (3.100)$$

Aunque los polos son mapeados del plano s al plano z por la ecuación (3.99) se debe observar que los ceros en los dos dominios no satisfacen la misma relación. Por lo tanto, el método de invarianza al impulso no corresponde a un mapeo uno a uno de los puntos en ambos planos dado por la ecuación (3.92).

La desventaja de este método es que al utilizar prototipos analógicos no limitados en banda generan aliasing en la respuesta en frecuencia del filtro digital, ya que $H(\omega)$ es una versión periódica de la respuesta en frecuencia del filtro analógico $H(\Omega)$

Ejemplo:

Dada una función $H(s)$ de un filtro analógico

$$H(s) = \frac{(s + 8)}{(s + 2)(s + 4)} = \frac{3}{s + 2} - \frac{2}{s + 4}$$

con polos en $s = -2$ y $s = -4$, aplicando la transformada de Laplace inversa se obtiene la respuesta al impulso

$$h(t) = \{3e^{-2t} - 2e^{-4t}\}U(t)$$

tomando la transformación directa de polos en variable z utilizando la ecuación (3.100)

$$H(z) = \frac{3T}{1 - e^{-2T} z^{-1}} - \frac{2T}{1 - e^{-4T} z^{-1}}$$

Comparado las respuestas en frecuencia:

- para el filtro analógico

$$H(\Omega) = \frac{(j\Omega + 8)}{(j\Omega + 2)(j\Omega + 4)}$$

- para el filtro digital

$$H(\omega) = \frac{3T}{1 - e^{-2T}e^{-j\omega T}} - \frac{2T}{1 - e^{-4T}e^{-j\omega T}}$$

las respuesta en DC, para el filtro analógico en $\Omega = 0$

$$H(0) = 1$$

para el filtro digital en $\omega = 0$

$$H(\omega) = \frac{3T}{1 - e^{-2T}} - \frac{2T}{1 - e^{-4T}} \neq 0$$

Los resultados son diferentes para este punto debido al aliasing en frecuencia. Con una razón de muestreo T muy pequeña se pueden hacer las aproximaciones $e^{-2T} \approx 1 - 2T$ y $e^{-4T} \approx 1 - 4T$, entonces la ecuación $H(\omega)|_{\omega=0}$ se aproxima a uno. En este ejemplo se puede concluir que utilizando una frecuencia de muestreo alta el efecto de aliasing es despreciable a frecuencia cero.

3.4.2. Transformada Z - matched

Otro método para convertir filtros analógicos en su equivalente digital es el mapeo directo de los polos y ceros de la función $H(s)$ en los polos y ceros respectivos en el plano z . Suponiendo que para el sistema analógico se tiene la función [32]

$$H(s) = \frac{\prod_{k=1}^M (s - c_k)}{\prod_{k=1}^N (s - p_k)} \quad (3.101)$$

donde c_k son los ceros y p_k son los polos del filtro, entonces la función $H(z)$ para un filtro digital

$$H(z) = \frac{\prod_{k=1}^M (1 - e^{c_k T} z^{-1})}{\prod_{k=1}^N (1 - e^{p_k T} z^{-1})} \quad (3.102)$$

donde T es el intervalo de muestreo. Cada uno de los factores $(s - a)$ de $H(s)$ son mapeados como $(1 - e^{aT} z^{-1})$ en $H(z)$. A este tipo de mapeo se le conoce como *transformación Z-matched*.

Para preservar las características de respuesta en frecuencia del filtro analógico, el intervalo T de muestreo en Z-matched debe ser seleccionado adecuadamente para producir la localización de los polos y ceros en la posición correspondiente en el plano z .

Se puede observar que los polos obtenidos por la transformación Z-matched son idénticos a los polos obtenidos con el método de respuesta invariante al impulso, aunque en estas dos técnicas resultan diferentes las posiciones de los ceros.

Procedimiento de diseño por Z - Matched

1. Dadas las especificaciones de un filtro paso bajas, $\omega_p, \omega_{sup}, A_p$ y A_s , para una determinada función de transferencia $H(z)$, se puede diseñar un filtro analógico equivalente y luego se mapea en el filtro digital deseable.
2. Elegir el período de muestreo T y determinar las frecuencias analógicas:
 Ω_p y Ω_{sup}
3. Utilizando $\Omega_p, \Omega_{sup}, A_p$ y A_s , diseñar $H(s)$ con la aproximación deseada Butterworth, Chebyshev, elíptica o Bessel.
4. Si un filtro analógico tiene p polos diferentes y q ceros, se puede utilizar expansión en productos para $H(s)$

$$H(s) = \frac{\prod_{k=1}^q (s - c_k)}{\prod_{k=1}^p (s - p_k)}$$

5. Si se muestrea cada período T , se pueden transformar los polos y los ceros analógicos en sus respectivos polos y ceros digitales directamente con $e^{p_k T}$ y $e^{c_k T}$

$$H(z) = \frac{\prod_{k=1}^q (1 - e^{c_k T} z^{-1})}{\prod_{k=1}^p (1 - e^{p_k T} z^{-1})}$$

6. Verificar la respuesta en frecuencia $H(e^{j\omega})$ del diseño

Este procedimiento de diseño se puede generalizar mostrando que $H(z)$ puede ser obtenida directamente de $H(s)$, sin tener que encontrar $h(t)$ y en consecuencia $h(nT)$. La desventaja de este método es que existe un “aliasing” en la respuesta en frecuencia y en algunos casos es intolerable. Este método es útil sólo cuando el filtro analógico es esencialmente limitado en banda y puede ser utilizado para el diseño de FD FPA y FSBW, pero no preservará las magnitudes equirrizo de los prototipos de Chebyshev y elípticos.

Ejemplo:

Si se tiene un sistema analógico representado por la función de transferencia

$$H(s) = \frac{1}{s + a}$$

aplicando transformada de Laplace inversa

$$h(t) = \mathcal{L}^{-1}\left\{\frac{1}{s+a}\right\} = e^{-at}U(t)$$

que se puede muestrear para $t = nT$ y aplicando la transformada Z

$$TZ\{e^{-anT}\} = \frac{1}{1 - e^{-aT}z^{-1}}$$

esto implica que se está efectuando una transformación de un polo analógico $s = -a$ a un polo digital $z = e^{-aT}$

$$(s + a) \rightarrow (1 - e^{-aT}z^{-1})$$

Para un polo complejo se tendría

$$(s + a - jb) \rightarrow (1 - e^{-(a-jb)T}z^{-1})$$

3.4.3. Diseño por aproximación numérica de la ecuación en diferencias

Un procedimiento simple para convertir un filtro analógico a uno digital es aproximar una ecuación diferencial a una ecuación en diferencias. Este tipo de aproximación es frecuentemente utilizado para resolver ecuaciones diferenciales con coeficientes constantes en forma numérica.

Muestreando una función continua $y(t)$ a tiempos $t = nT$, donde T es el período de muestreo y utilizando técnicas de integración numérica, la derivada de la función $y(t)$ en el tiempo continuo se aproxima por diferencias finitas (ver figura 3.14)

$$\left.\frac{dy(t)}{dt}\right|_{t=nT} = \frac{y(nT) - y(nT - T)}{T} = \frac{y(n) - y(n-1)}{T} \quad (3.103)$$

esta es la aproximación de Euler y es un caso especial de

$$\left.\frac{dy(t)}{dt}\right|_{t=nT} = \frac{1}{T} \sum_{i=0}^M a_i y(nT - iT) \quad (3.104)$$

para $y(n) = y(nT)$.

En la ecuación (3.103), para el lado del tiempo continuo la función de transferencia es $H(s) = 1/s$ y en el lado del tiempo discreto su función de transferencia es $H(z) = (1 - z^{-1})/T$, donde se obtiene la correspondencia

$$s = \frac{1 - z^{-1}}{T} = \frac{1}{T} \frac{z - 1}{z} \quad (3.105)$$

que es la aproximación de Euler. Entonces una aproximación de un sistema analógico por un sistema digital sería la transformación

$$H(z) = H(s)\Big|_{s=\frac{1}{T}\frac{z-1}{z}} \quad (3.106)$$

donde $H(s)$ es la función de transferencia del filtro analógico a aproximar digitalmente, también se puede despejar $z = 1/(1 - sT)$. Si se representan las variables complejas como $s = r + jx$ y $z = R + jI$ y si $T = 1$, entonces se llega a las ecuaciones

$$\left(R - \frac{1}{2(r-1)}\right)^2 + I^2 = \frac{1}{4(r-1)^2} \quad (3.107)$$

$$R^2 + \left(I - \frac{1}{2x}\right)^2 = \frac{1}{4x^2} \quad (3.108)$$

si se toman a r y x como parámetros, las dos ecuaciones anteriores generan las familias de círculos:

- con centro $C(\frac{1}{2(r-1)}, 0)$ y radio $\frac{1}{2(r-1)}$
- con centro $C(0, \frac{1}{2x})$ y radio $\frac{1}{2x}$

por lo que se concluye que este tipo de transformación mapea la zona estable del plano s a un círculo de radio $1/2$ circunscrito en el círculo unitario en el plano z sobre el eje real de 0 a 1. Si Ω varía de $-\infty$ a ∞ , los puntos correspondientes en el plano z están ubicados sobre un círculo de radio $1/2$ centrado en $1/2$ sobre el eje real.

Este mapeo tiene la característica deseable de ir de la zona estable del plano s al plano z , sin embargo, la región posible de localizar los polos de un filtro digital en el plano z está muy restringida al diseño de FPB de ancho de banda muy angosto.

3.4.4. Diseño de filtros IIR por Transformada Bilineal

La Transformada Bilineal (TB) aproxima la respuesta de un prototipo analógico a una versión digital a través de la conversión de la ecuación diferencial del sistema analógico en una ecuación en diferencias que represente al sistema digital. Este método es un tipo de mapeo de variable compleja que transforma el eje $j\Omega$ en el plano s al círculo unitario en el plano z y toda la zona estable en el plano s al interior del círculo unitario, por lo que evita aliasing de las componentes en frecuencia [32].

Existen varios métodos para el diseño de filtros digitales IIR, sin embargo, uno de los más utilizados es la transformación de bandas de frecuencias por TB, que convierten un filtro paso baja (FPB) analógico normalizado a filtro paso alta (FPA), filtro paso banda (FPBW) y filtro supresor de banda (FSBW) no normalizados. Existen dos técnicas para efectuar la transformación y obtener filtros digitales [3], [32], [9],[38].

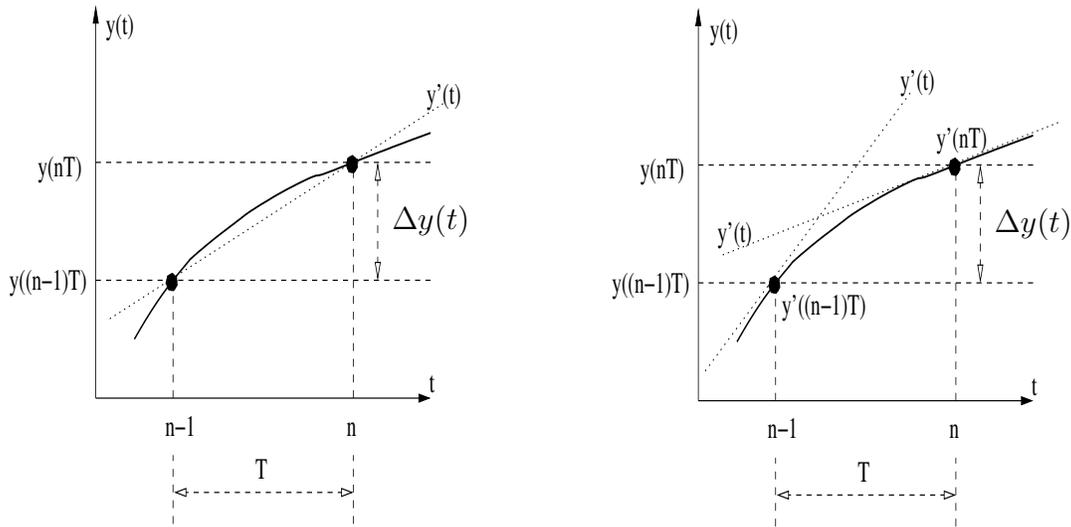


Figura 3.14: Aproximación por diferencias finitas

1. Mapeo de un FPB analógico normalizado en el plano s al plano z aplicando una transformación en la banda de frecuencias.
2. Transformación de un FPB analógico normalizado a un filtro analógico no normalizado que puede ser digitalizado para obtener el tipo del filtro digital deseado.

Se ha comprobado que la digitalización de un filtro analógico utilizando la Transformada Bilineal (TB) produce resultados similares usando los dos procedimientos anteriores [3], [19]. A continuación se discutirán ambos métodos y se comprobará que se llega a los mismos resultados. La TB tiene la característica de hacer un mapeo no lineal de la región estable del plano complejo s a la región estable del plano complejo z . Es decir, que la TB produce siempre filtros digitales estables a partir de prototipos analógicos estables.

La TB está profundamente ligada a la relación que existe entre los dominios s y z

$$z = e^{sT} \tag{3.109}$$

despejando s

$$s = \frac{1}{T} \ln(z) \tag{3.110}$$

De estas dos últimas ecuaciones se puede verificar que la parte izquierda del plano s se mapea dentro del círculo unitario en el plano z y la parte derecha del plano s se mapea fuera del círculo unitario en el plano z , por otro lado el eje imaginario en el plano s se mapea sobre el círculo unitario de manera periódica cada $\Omega = 2\pi/T$. Si se transforma una función

$H(s)$ utilizando la ecuación (3.110) se obtiene una función $H(z)$ irracional en términos de logaritmos, la cual no es útil para realizar un sistema. Para obtener una función racional de $H(z)$ es necesario racionalizar la función logarítmica de z . La ecuación (3.110) se puede expresar desarrollando la serie del logaritmo [37]

$$s = \frac{2}{T} \left[\frac{z-1}{z+1} + \frac{1}{3} \left(\frac{z-1}{z+1} \right)^3 + \frac{1}{5} \left(\frac{z-1}{z+1} \right)^5 \dots \right] \quad (3.111)$$

$$s = \frac{1}{T} \left[\frac{z-1}{z} + \frac{1}{2} \left(\frac{z-1}{z} \right)^2 + \frac{1}{3} \left(\frac{z-1}{z} \right)^3 \dots \right] \quad (3.112)$$

Como z se aproxima a uno, entonces la serie (3.111) converge muy rápido, por lo que se pueden despreciar los términos de orden superior, aunque se agrega un ligero error, entonces

$$s = \frac{2}{T} \left(\frac{z-1}{z+1} \right) \quad (3.113)$$

y para la serie (3.112)

$$s = \frac{2}{T} \left(\frac{z-1}{z} \right) \quad (3.114)$$

La utilización de la ecuación (3.112) lleva a un mapeo de la región estable del plano s a un círculo circunscrito en la parte derecha del círculo unitario en el plano z (figura 3.15 región sombreada) lo que ocasiona mayor deformación en la transformación y menos posibilidades en la selección de anchos de bandas y frecuencias.

Otra forma de llegar a las ecuaciones (3.113) y (3.114) es con la técnica de integración numérica utilizada para simular un integrador de un filtro analógico o la integración por el método trapezoidal de una ecuación en diferencias [32], [26]. La TB puede relacionarse con la fórmula trapezoidal de integración numérica, si se tiene un sistema representado por ecuaciones en diferencias de primer orden

$$\frac{dy(t)}{dt} + ay(t) = bx(t) \quad (3.115)$$

aplicando la transformada de Laplace se tiene la función de transferencia de un filtro lineal analógico

$$H(s) = \frac{b}{s+a} \quad (3.116)$$

si en lugar de utilizar derivadas, se utilizan diferencias finitas

$$y'(t) = \frac{\Delta y}{\Delta t} = \frac{y(t+\Delta t) - y(t)}{\Delta t} \quad (3.117)$$

$$y(t) = -y'(t)\Delta t + y(t+\Delta t) \quad (3.118)$$

si se integra la derivada y aproximando la integral por la regla del trapecio

$$y(t) = \int_{t_o}^t y'(\tau) d\tau + y(t_o) \quad (3.119)$$

donde $y'(t)$ es la derivada de $y(t)$, tomando $t = nT$ y $t_o = nT - T$

$$y(nT) = y(nT - T) + \frac{T}{2}[y'(nT) + y'(nT - T)] \quad (3.120)$$

evaluando la ecuación diferencial en tiempos $t = nT$ (ver figura 3.14)

$$y'(nT) = -ay(nT) + b(x(nT)) \quad (3.121)$$

si $y(n) = y(nT)$ y $x(n) = x(nT)$ y sustituyendo $y'(nT)$ en (3.120) se tendría

$$y(nT) = y(nT - T) + \frac{T}{2}[-ay(nT) + bx(nT) - ay((n-1)T) + bx((n-1)T)] \quad (3.122)$$

reagrupando términos se obtiene

$$\left(1 + \frac{aT}{2}\right)y(n) - \left(1 - \frac{aT}{2}\right)y(n-1) = \frac{bT}{2}[x(n) + x(n-1)] \quad (3.123)$$

aplicando transformada Z

$$\left(1 + \frac{aT}{2}\right)Y(z) - \left(1 - \frac{aT}{2}\right)z^{-1}Y(z) = \frac{bT}{2}(1 + z^{-1})X(z) \quad (3.124)$$

equivale a la función de transferencia de un filtro digital

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\frac{bT}{2}(1 + z^{-1})}{1 + \frac{aT}{2} - (1 - \frac{aT}{2})z^{-1}} \quad (3.125)$$

reordenando términos

$$H(z) = \frac{b}{\frac{2}{T}\left(\frac{1-z^{-1}}{1+z^{-1}}\right) + a} \quad (3.126)$$

se observa que se ha realizado un mapeo del plano s al plano z de la función $H(s)$ de primer orden a $H(z)$ por medio de la transformación

$$s = \frac{2}{T}\left(\frac{z-1}{z+1}\right) \quad (3.127)$$

esta transformación es conocida como *Transformada Bilineal* y es igual al resultado obtenido en la ecuación (3.113).

Otro nombre de esta transformación es "*Linear Fraccional*", porque cuando se ordena la fracción de la ecuación (3.113) se obtiene la función de s y z :

$$\frac{T}{2}sz + \frac{T}{2}s - z + 1 = 0$$

Que es lineal si s o z son constantes, o bilineal en s y z .

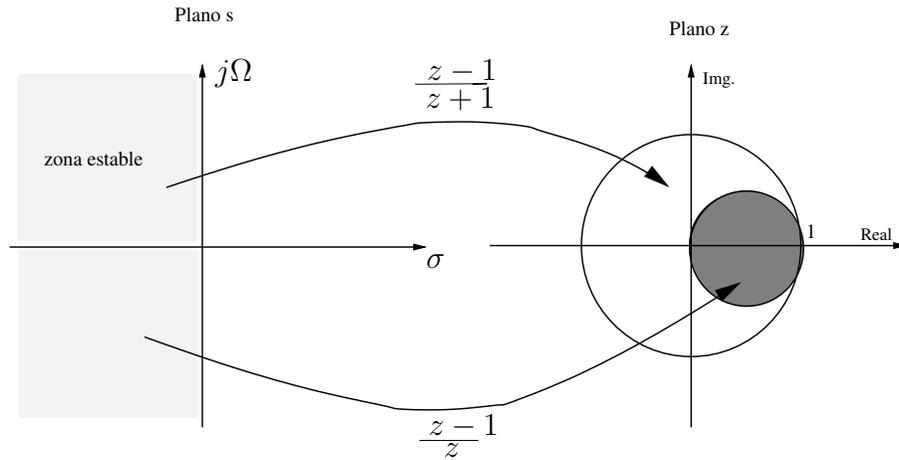


Figura 3.15: La transformada bilineal del plano s al plano z

Mapeo del plano s al plano z por la TB

Si se despeja z de la ecuación (3.113)

$$z = \frac{1 + sT/2}{1 - sT/2} = \frac{2/T + s}{2/T - s} \quad (3.128)$$

Como las variables s y z son complejas, se define

$$z = re^{j\omega} = R + jI \quad \text{y} \quad s = \sigma + j\Omega$$

entonces se obtiene la variable z en función de los parámetros del plano s

$$z = \frac{1 + \sigma T/2 + j\Omega T/2}{1 - \sigma T/2 - j\Omega T/2} \quad (3.129)$$

cuyo módulo $|z|$ depende del valor de σ , por tanto se tienen tres posibilidades del mapeo:

1. La parte izquierda del plano s es mapeada dentro del círculo unitario del plano z . Es decir, que la transformación para un diseño analógico estable produce un diseño digital estable

$$\sigma < 0 \quad |z| = \frac{1 + \sigma T/2 + j\Omega T/2}{1 - \sigma T/2 - j\Omega T/2} < 1 \quad (3.130)$$

2. El eje imaginario del plano s es mapeado sobre la periferia del círculo unitario en el plano z . Esto implica que no existe aliasing en el dominio de la frecuencia

$$\sigma = 0 \quad |z| = \frac{1 + j\Omega T/2}{1 - j\Omega T/2} = 1 \quad (3.131)$$

3. La parte derecha del plano s (zona inestable de $H(s)$) es mapeada fuera del círculo unitario del plano z . Se concluye que la transformación para un diseño analógico inestable produce un diseño digital inestable

$$\sigma > 0 \quad |z| = \frac{1 + \sigma T/2 + j\Omega T/2}{1 - \sigma T/2 - j\Omega T/2} > 1 \quad (3.132)$$

Si $\sigma = 0$

$$z = \frac{1 + j\Omega T/2}{1 - j\Omega T/2} = e^{j\omega T} \quad (3.133)$$

cuya magnitud es 1, y si resolvemos para ω en función de Ω o viceversa, se tiene

$$\omega = \frac{2}{T} \operatorname{tg}^{-1}\left(\frac{\Omega T}{2}\right) \quad \text{o} \quad \Omega = \frac{2}{T} \operatorname{tg}\left(\frac{\omega T}{2}\right) \quad (3.134)$$

es decir que Ω está relacionada no linealmente con ω , sin embargo, no existe aliasing.

Sustituyendo las variables s y z en la ecuación (3.127), se pueden obtener las familias de curvas que definen la transformación o que realacionan ambos planos

$$s = \frac{2}{T} \left(\frac{r e^{j\omega T} - 1}{r e^{j\omega T} + 1} \right) = \frac{2}{T} \left(\frac{r(\cos\omega T + j \operatorname{sen}\omega T) - 1}{r(\cos\omega T + j \operatorname{sen}\omega T) + 1} \right)$$

$$s = \frac{2}{T} \left(\frac{r^2 - 1}{1 + r^2 + 2r \cos\omega T} + j \frac{2r \operatorname{sen}\omega T}{1 + r^2 + 2r \cos\omega T} \right)$$

igualando términos reales e imaginarios

$$\sigma = \frac{2}{T} \left(\frac{r^2 - 1}{1 + r^2 + 2r \cos\omega T} \right) \quad (3.135)$$

$$\Omega = \frac{2}{T} \left(\frac{2r \operatorname{sen}\omega T}{1 + r^2 + 2r \cos\omega T} \right) \quad (3.136)$$

se verifica que si $r < 1$, entonces $\sigma < 0$ y si $r > 1$, entonces $\sigma > 0$, por lo que toda la región izquierda del plano s es mapeada al interior del círculo unitario en el plano z . Cuando $r = 1$ $\sigma = 0$, entonces

$$\Omega = \frac{2}{T} \frac{\operatorname{sen}\omega T}{1 + \cos\omega T} = \frac{2}{T} \frac{2 \operatorname{sen}(\omega T/2) \cos(\omega T/2)}{2 \cos^2(\omega T/2)} = \frac{2}{T} \tan(\omega T/2) \quad (3.137)$$

despejando

$$\omega = \frac{2}{T} \tan^{-1}\left(\frac{\Omega T}{2}\right) \quad (3.138)$$

de estas dos últimas ecuaciones y por la definición de la función tangente se observa que el intervalo completo de la frecuencia analógica Ω ($-\infty < \Omega < \infty$) es mapeada dentro del

intervalo de la frecuencia digital $\pi \leq \omega \leq \pi$, como se observa en la figura 3.16. Este mapeo es no lineal, por lo que se produce una compresión en la frecuencia o un *doble* debido a la no linealidad de la función arco tangente, un punto en $s = \infty$ genera un polo en $z = -1$ y filtro analógico con polo en $s = -\infty$ produce un filtro digital con un cero en $z = -1$. La TB conserva la amplitud de los rizados aproximadamente constantes y la forma de la magnitud, por el fenómeno de *doble* la escala de las frecuencias analógicas y digitales son diferentes y están relacionadas por la ecuación (3.138).

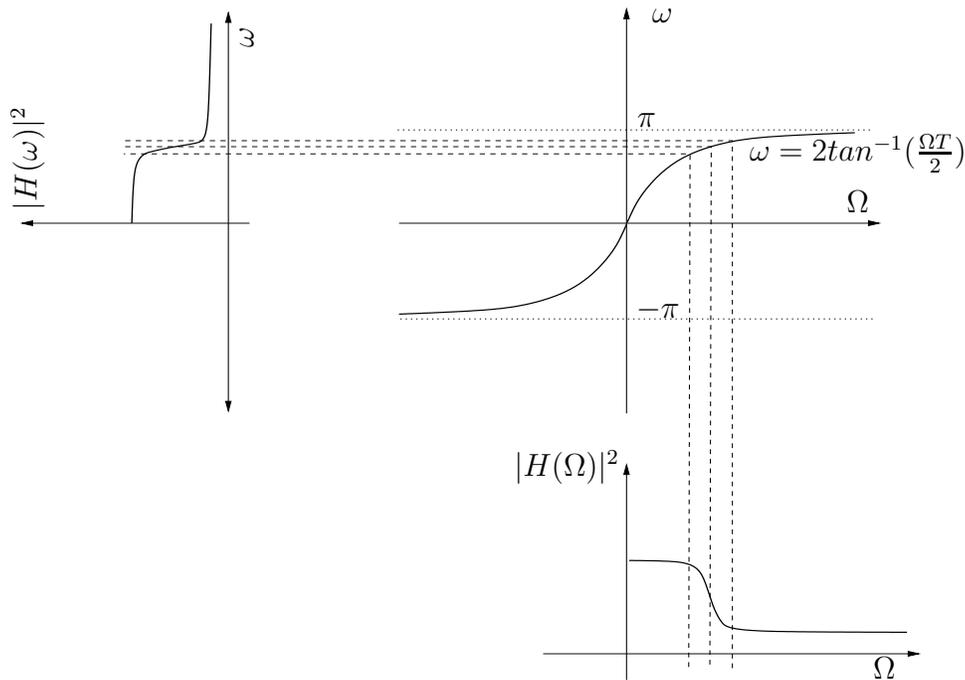


Figura 3.16: Mapeo entre las frecuencias ω y Ω por TB

De forma similar se pueden obtener las ecuaciones paramétricas que describen las curvas (círculos) en el plano z generados por la transformación, si se considera en la ecuación (3.127) a la constante $\frac{2}{T} = 1$ y se sustituyen la variables s y z en forma cartesiana, entonces

$$\sigma + j\Omega = \frac{R + jI - 1}{R + jI + 1} = \frac{R^2 + I^2 - 1}{(R + 1)^2 + I^2} + j \frac{2I}{(R + 1)^2 + I^2} \quad (3.139)$$

igualando partes reales e imaginarias

$$\sigma = \frac{R^2 + I^2 - 1}{(R + 1)^2 + I^2} \quad (3.140)$$

$$\Omega = \frac{2I}{(R+1)^2 + I^2} \quad (3.141)$$

desarrollando éstas ecuaciones y completando cuadrados se obtienen dos familias de círculos en el plano z con coordenadas R y I , con parámetros σ y Ω . Para la parte real

$$I^2 + \left(R + \frac{r}{r-1}\right)^2 = \frac{1}{(r-1)^2} \quad (3.142)$$

que representa los círculos con centro en $C(\frac{r}{r-1}, 0)$ y radio $\frac{1}{(r-1)}$ para la parte imaginaria¹.

$$\left(R+1\right)^2 + \left(I - \frac{1}{x}\right)^2 = \frac{1}{x^2} \quad (3.143)$$

que representa los círculos con centro en $C(-1, \frac{1}{x})$ y radio $\frac{1}{x}$

3.4.5. Metodología de diseño de un filtro IIR por TB

Partiendo de las especificaciones de diseño de un filtro digital, ω_p frecuencia de corte en la banda de paso, ω_{sup} frecuencia de corte en la banda suprimida, A_p ganancia en la banda de paso y A_s atenuación en la banda suprimida para determinar la función de transferencia $H(z)$. La aplicación de la TB se puede realizar de dos formas:

1. Forma directa, que transforma la función de transferencia $H(s)$ normalizada a $H(z)$ directamente.
2. Se transforma una función de transferencia $H(s)$ normalizada a otra $H(s)$ deseada y luego a esta última $H(s)$ se le aplica la TB con $T = 1$ para obtener $H(z)$.

Método de diseño de filtros IIR por TB forma directa $H(s) \implies H(z)$

Se tiene la siguiente metodología de diseño:

1. Elegir el tipo de aproximación polinómica: Butterworth, Chebyshev, elíptico o Bessel.
2. Calcular el orden n con base en las especificaciones de las frecuencias de corte y las ganancias en las bandas de paso. Para un FPB con aproximación polinómica tipo Butterworth y Chebyshev

$$n_{Butt} > \frac{\log_{10}\left(\frac{10^{0.1A_s}-1}{10^{0.1A_p}-1}\right)}{2\log_{10}\left(\frac{\Omega_{sup}}{\Omega_c}\right)} \quad n_{Cheby} > \frac{\cosh^{-1}\left(\frac{\sqrt{10^{0.1A_s}-1}}{\sqrt{10^{0.1A_p}-1}}\right)}{\cosh^{-1}\left(\frac{\Omega_{sup}}{\Omega_p}\right)}$$

con especificaciones analógicas:

¹Un desarrollo similar se puede encontrar en la construcción de la carta Smith cuando se realiza una transformación de los complejos a los complejos [6], [43].

- A_p ganancia en decibels de la banda de paso
 - A_s atenuación en decibels de la banda suprimida
 - Ω_c frecuencia de corte en Hertz
 - Ω_{sup} frecuencia de corte en banda suprimida en Hertz
3. Elegir el período de muestreo T , es decir, $T = \frac{1}{f_s} \rightarrow \omega_d = \frac{2\pi f}{f_s}$
 4. Con el orden n elegir $H(s)$ de tablas de polinomios
 5. Diseñar un filtro analógico $H(s)$ con especificaciones Ω_p , Ω_s , A_p y A_s
 6. Aplicar la TB para obtener $H(z)$ dependiendo del tipo de filtro a diseñar

$$H_{FPB}(z) = H(s) \Big|_{s=C_1 \frac{1-z^{-1}}{1+z^{-1}}} \quad (3.144)$$

$$H_{FPA}(z) = H(s) \Big|_{s=C_2 \frac{1+z^{-1}}{1-z^{-1}}} \quad (3.145)$$

$$H_{FPBW}(z) = H(s) \Big|_{s=C_1 \frac{1-z^{-1}}{1+z^{-1}} + C_2 \frac{1+z^{-1}}{1-z^{-1}}} \quad (3.146)$$

La constante C_1 se calcula para filtros paso bajas y C_2 para filtros paso altas

7. De $H(z)$ obtener $y(n)$ como una ecuación en diferencias utilizando la transformada Z inversa
8. Comprobar el desempeño de la $H(z)$ diseñada sustituyendo los coeficientes de $H(z)$ y calcular $|H(\omega)|$
9. Seleccionar la estructura adecuada de acuerdo a la aplicación para implementar $H(z)$
10. Implementar el filtro:
 - Implementación software: en lenguajes de alto nivel o programas de simulación
 - Implementación micro-software: en lenguajes de bajo nivel o lenguaje ensamblador
11. Implementar un prototipo donde va a operar el filtro diseñado, es decir, una arquitectura dedicada

Constantes para diferentes tipos de filtros

De la fórmula de transformada bilineal ecuación (3.127), la constante $2/T$ se puede calcular sustituyendo los valores normalizados para s y z , para el caso de un filtro paso baja calculamos C_1

$$s = C_1 \left(\frac{z - 1}{z + 1} \right) \quad (3.147)$$

Esta contante C_1 corrige la distorsión de frecuencias en la transformación de la función $H(s)$ a $H(z)$, sustituyendo $s = j\Omega_a$, $z = e^{jT\omega_d}$ y utilizando identidades trigonométricas, se obtiene el valor de C_1 , que nos sirve para obtener un filtro digital

$$s = C_1 \frac{e^{jT\omega_d} - 1}{e^{jT\omega_d} + 1} = C_1 \frac{e^{jT\omega_d/2} (e^{jT\omega_d/2} - e^{-jT\omega_d/2})}{e^{jT\omega_d/2} (e^{jT\omega_d/2} + e^{-jT\omega_d/2})} \quad (3.148)$$

entonces

$$j\Omega_a = C_1 \frac{j \operatorname{sen}(T\omega_d/2)}{2 \operatorname{cos}(T\omega_d/2)} = C_1 j \operatorname{tg}(T\omega_d/2) \quad (3.149)$$

Para valores normalizados, donde $\Omega_a = 1$ y $T = 1$, despejando la constante C_1 .

$$C_1 = \operatorname{ctg}(T\omega_d/2) = \operatorname{ctg}\left(\frac{2\pi f_{ca}}{2f_s}\right) = \operatorname{ctg}\left(\frac{\pi f_{ca}}{f_s}\right) \quad (3.150)$$

Esta constante calculada se utiliza para el diseño de filtros digitales por TB tipo paso baja, para diseño de filtros paso alta se puede proceder de forma similar invirtiendo ambos lados de la TB sustituyendo s y z , y despejando la constante C_2 , estas constantes se muestran en la tabla 3.4. Otro método fácil para el cálculo de filtros paso altas y paso bandas, consiste en utilizar transformaciones en frecuencias como se mencionará más adelante².

3.4.6. Método de transformaciones en la frecuencia

Este método consiste en mapear una función $H(s)$ de un filtro paso bajas con frecuencia de corte Ω_c a una función $H(s)$ del filtro deseado, paso bajas con frecuencia de corte Ω_{PB} , paso altas con frecuencia de corte Ω_{PA} y filtro paso banda o supresor de banda con frecuencias de corte Ω_L y Ω_H . Después se transforma la $H(s)$ obtenida a una respuesta $H(z)$ de un filtro digital, donde en general la transformación $H(s)$ a $H(z)$ se hace de un filtro normalizado paso bajas a la función del filtro deseado, y luego emplear la TB para obtener $H(z)$. Este proceso produce el mismo resultado que al aplicar la TB en forma directa de la sección anterior.

²Para mayor detalle en el procedimiento del cálculo de estas constantes se puede consultar en [34], [3], la referencia MATLAB de Math-Works, el programa SCILAB de INRIA [20].

Tipo de filtro	TB	Constante
Filtro paso bajas (FPB)	$s = C_1 \left(\frac{z-1}{z+1} \right)$	$C_1 = \cotg \left(\frac{\pi f_{ca}}{f_s} \right)$
Filtro paso altas (FPA)	$s = C_2 \left(\frac{z+1}{z-1} \right)$	$C_2 = tg \left(\frac{\pi f_{ca}}{f_s} \right)$
Filtro paso banda (FPBW)	$s = C_1 \left(\frac{z-1}{z+1} \right) + C_2 \left(\frac{z+1}{z-1} \right)$	C_1 y C_2 de FPB y FPA

Tabla 3.4: TB y constantes para filtros FPB, FPA y FPBW

En la tabla 3.5 se muestran las transformaciones necesarias, donde Ω_H es la frecuencia de corte alta y Ω_L la frecuencia de corte baja en la banda de paso del filtro analógico FPBW o FSBW. Para estos filtros la frecuencia central se calcula $\Omega_0 = \sqrt{\Omega_L \Omega_H}$ y el factor de calidad $Q = \Omega_0 / (2\pi(\Omega_H - \Omega_L))$, en estos filtros hay que tener en cuenta que las transformaciones son no lineales y puede aparecer una distorsión de la respuesta característica del filtro FPB, sin embargo, los efectos de no linealidad en frecuencia son mínimos y preservan la amplitud de la respuesta característica del filtro. Así un filtro FPB equirrizo es transformado en otro filtro equirrizo en las bandas de paso y suprimida de un FPA [32].

Transformaciones en la frecuencia en el dominio de $H(s)$ normalizada

Si la función de transferencia $H(s)$ del filtro FPB con frecuencia de corte Ω_c está normalizada, entonces sustituyendo $\Omega_c = 1$ en las ecuaciones de la tabla 3.5, por lo que se tienen las transformaciones de la tabla 3.6.

Tipo de filtro	Transformación
FPB \rightarrow FPB	$s \rightarrow \Omega_c \frac{s}{\Omega_{PB}}$
FPB \rightarrow FPA	$s \rightarrow \Omega_c \frac{\Omega_{PA}}{s}$
FPB \rightarrow FPBW	$s \rightarrow \Omega_c \frac{s^2 + \Omega_L \Omega_H}{s(\Omega_H - \Omega_L)}$
FPB \rightarrow FSBW	$s \rightarrow \Omega_c \frac{s(\Omega_H - \Omega_L)}{s^2 + \Omega_L \Omega_H}$

Tabla 3.5: Transformación $H(s)$ a $H(s)$ [3], [32]

Tipo de filtro	Transformación
FPB \rightarrow FPB	$s \rightarrow \frac{s}{\Omega_{PB}}$
FPB \rightarrow FPA	$s \rightarrow \frac{\Omega_{PA}}{s}$
FPB \rightarrow FPBW	$s \rightarrow \frac{s^2 + \Omega_L \Omega_H}{s(\Omega_H - \Omega_L)}$
FPB \rightarrow FSBW	$s \rightarrow \frac{s(\Omega_H - \Omega_L)}{s^2 + \Omega_L \Omega_H}$

Tabla 3.6: Transformación de un FPB $H(s)$ a $H(s)$ [3]

Metodología de diseño de filtros IIR por transformación $H(s) \implies H(s) \implies H(z)$

1. Calcular el orden n del polinomio de la aproximación polinómica elegida
2. Transformar ω_p y ω_{sup} utilizando las ecuaciones

$$\Omega_p = tg\left(\frac{\omega_p}{2}\right) \tag{3.151}$$

$$\Omega_{sup} = tg\left(\frac{\omega_s}{2}\right) \tag{3.152}$$

3. Con base al tipo de filtro, de la tabla (3.6) se elige la transformación y se obtiene la nueva $H(s)$ para las especificaciones dadas
4. Se aplica la transformada bilineal

$$H(z) = H(s)\Big|_{s=\frac{1-z^{-1}}{1+z^{-1}}} \tag{3.153}$$

5. Comprobar el desempeño de la $H(z)$ diseñada, es decir, sustituir los coeficientes de $H(z)$ y calcular $|H(\omega)|$
6. De $H(z)$, $Y(z) = H(z)X(z)$ obtener $y(n)$ como una ecuación en diferencias utilizando la transformada Z inversa
7. Seleccionar la estructura adecuada de implementación de $H(z)$
8. Implementar el filtro:
 - Implementación software: en lenguajes de alto nivel o programas de simulación
 - Implementación micro-software: en lenguajes de bajo nivel o lenguaje ensamblador
9. Implementar un prototipo donde va a operar el filtro diseñado, es decir, una arquitectura dedicada

Transformaciones en la frecuencia en el dominio digital

El diseño de filtros digitales por medio de transformaciones en frecuencia se puede realizar de forma similar al diseño de filtros analógicos, es decir, teniendo el prototipo digital de un filtro paso bajas se pueden obtener otros tipos de filtros. La transformación a utilizar implica reemplazar la variable z^{-1} por una función racional de z^{-1} que satisfaga las propiedades:

1. Los puntos dentro del círculo unitario en el plano z se deben mapear a puntos dentro del círculo unitario, esta es una condición de estabilidad.

2. Los puntos sobre el círculo unitario en el plano z se deben mapear a puntos sobre el círculo unitario.

Es decir que la función de mapeo debe ser un filtro paso todo de la forma

$$g(z^{-1}) = \pm \prod_{k=1}^n \frac{z^{-1} - a_k}{1 - a_k z^{-1}} \quad (3.154)$$

donde $|a_k| < 1$ para asegurar la estabilidad del filtro. Las transformaciones respectivas se muestran en la tabla 3.7 [32].

Tipo de filtro	Transformación	Parámetros
FPB \rightarrow FPB	$z^{-1} \rightarrow \frac{z^{-1}-a}{1-az^{-1}}$	$\omega_p = \omega_c$ original del FPB $\omega_1 =$ nueva ω_c $a = \frac{\text{sen}[(\omega_p-\omega_1)/2]}{\text{sen}[(\omega_p+\omega_1)/2]}$
FPB \rightarrow FPA	$z^{-1} \rightarrow -\frac{z^{-1}+a}{1+az^{-1}}$	$\omega_1 =$ nueva ω_{cpa} $a = -\frac{\cos[(\omega_p-\omega_1)/2]}{\cos[(\omega_p+\omega_1)/2]}$
FPB \rightarrow FPBW	$z^{-1} \rightarrow -\frac{z^{-2}-a_1z^{-1}+a_2}{a_2z^{-2}-a_1z^{-1}+1}$	$a_1 = -2\alpha K/(K+1)$ $a_2 = (K-1)/(K+1)$ $K = \cot(\frac{\omega_H-\omega_L}{2}) \text{tg}(\frac{\omega_o}{2})$ $\alpha = \frac{\cos[(\omega_H+\omega_L)/2]}{\cos[(\omega_H-\omega_L)/2]}$
FPB \rightarrow FSBW	$z^{-1} \rightarrow \frac{z^{-2}-a_1z^{-1}+a_2}{a_2z^{-2}-a_1z^{-1}+1}$	$a_1 = -2\alpha K/(K+1)$ $a_2 = (1-K)/(K+1)$ $K = \text{tg}(\frac{\omega_H-\omega_L}{2}) \text{tg}(\frac{\omega_o}{2})$ $\alpha = \frac{\cos[(\omega_H+\omega_L)/2]}{\cos[(\omega_H-\omega_L)/2]}$

Tabla 3.7: Transformación de filtros digitales en el plano z

Ejemplo 1:

Diseñar un filtro paso bajas tipo IIR que aproxime una respuesta analógica cuyas especificaciones son:

$f_c=1700$ Hz, $f_{sup}=4250$ Hz, $A_p=-3$ db, $A_{sup}= -12$ db y $f_s=12000$ muestras/seg.

Solución:

Utilizando una aproximación tipo Butterworth, se calcula el orden por la fórmula

$$n > \frac{\log_{10}\left(\frac{10^{0.1*12}-1}{10^{0.1*3}-1}\right)}{2\log_{10}\frac{4250}{1700}} > 1.4738$$

Por lo que el orden del filtro es $n=2$, entonces se busca el polinomio correspondiente (tabla 3.1), y se construye la función de transferencia correspondiente $H(s)$

$$H(s) = \frac{1}{s^2 + \sqrt{2}s + 1}$$

donde $C_1 = \text{ctg}\left(\frac{\pi 1700}{12000}\right) = 2.096$, es decir, que la TB es $s = 2.096\left(\frac{z-1}{z+1}\right)$ desarrollando $H(z) = H(s)|_{s=2.096\left(\frac{z-1}{z+1}\right)}$, se tiene:

$$H(z) = \frac{1}{(2.096)^2\left(\frac{z-1}{z+1}\right)^2 + (1.41)2.096\left(\frac{z-1}{z+1}\right) + 1}$$

$$H(z) = \frac{z^2 + 2z + 1}{8.357z^2 - 6.786z + 2.429}$$

normalizando (dividiendo numerador y denominador por $8.357z^2$)

$$H(z) = \frac{0.1196 + 0.2393z^{-1} + 0.1196z^{-2}}{1 - 0.8120z^{-1} + 0.2906z^{-2}} = \frac{Y(z)}{X(z)}$$

como $Y(z) = H(z)X(z)$, se despeja $Y(z)$

$$Y(z) = 0.8120z^{-1}Y(z) - 0.2906z^{-2}Y(z) + 0.1196X(z) + 0.2393z^{-1}X(z) + 0.1196z^{-2}X(z)$$

aplicando la transformada Z inversa, se obtiene $y(n) = TZI(Y(z))$

$$y(n) = 0.1196x(n) + 0.2393x(n-1) + 0.1196x(n-2) + 0.8120y(n-1) - 0.2906y(n-2)$$

De la última ecuación se deduce que los coeficientes del filtro IIR son:

$$a_0 = 1 \quad a_1 = -0.8120 \quad a_2 = 0.2906$$

$$b_0 = 0.1196 \quad b_1 = 0.2393 \quad b_2 = 0.1196$$

Ejemplo 2:

Diseñar un filtro paso altas IIR con aproximación Butterworth y especificaciones analógicas: $A_s = -12\text{db}$, $A_p = -3\text{d}$, $f_p = 3400\text{hz}$, $f_{sup} = 1360\text{hz}$, $f_s = 16000\text{hz}$.

Solución:

Para los datos

$$n > \frac{\log_{10}\left(\frac{10^{0.1*12} - 1}{10^{0.1*3} - 1}\right)}{2\log_{10}\left(\frac{3400}{1360}\right)} = 1.473$$

entonces $n=2$.

$$H(s) = \frac{1}{s^2 + 1.41s + 1}$$

la constante $C_2 = \text{tg}\left(\frac{\pi*3400}{16000}\right) = 0.7883$, entonces

$$s = 0.7883\left(\frac{z+1}{z-1}\right)$$

sustituyendo en $H(s)$

$$H(z) = \frac{1}{\left(0.7883\left(\frac{z+1}{z-1}\right)\right)^2 + 1.41 * 0.7883\left(\frac{z+1}{z-1}\right) + 1}$$

$$H(z) = \frac{0.3654 - 0.7309z^{-1} + 0.3654z^{-2}}{1 - 0.2767z^{-1} + 0.1851z^{-2}}$$

entonces

$$y(n) = 0.3654x(n) - 0.7309x(n-1) + 0.3654x(n-2) + 0.2767y(n-1) - 0.1851y(n-2)$$

Ejemplo 3:

Encontrar $H(z)$ para un filtro digital paso banda Butterworth de orden cuatro ($n=4$), con frecuencia de muestreo $f_s=20\text{ hz}$ y frecuencias de corte $f_L=0.8\text{ hz}$ y $f_H=1.2\text{ hz}$, la ganancia debe ser unitaria.

Solución por TB de $H(s)$ a $H(z)$

Como el orden del diseño del filtro digital solicitado es $n=4$ y la aproximación polinómica es del tipo Butterworth, entonces se elige el polinomio $P(s)$ de Butterworth $n=4/2=2$, ya que por este método para el cálculo de un filtro FPBW se duplica el orden en el dominio de z .

Para el cálculo del filtro digital pedido se procede a encontrar $H(z)$ a través de

$$H(z) = H(s) \Big|_{s=C_1\left(\frac{z-1}{z+1}\right) + C_2\left(\frac{z+1}{z-1}\right)} \tag{3.155}$$

donde

$$C_1 = \cot(1.2\pi/20) = 5.2421 \quad C_2 = \operatorname{tg}(0.8\pi/20) = 0.1263$$

$$H(z) = \frac{1}{s^2 + \sqrt{2}s + 1} \Big|_{s=5.2421\left(\frac{z-1}{z+1}\right)+0.1263\left(\frac{z+1}{z-1}\right)} \quad (3.156)$$

$$H(z) = \frac{1}{\left(5.2421\left(\frac{z-1}{z+1}\right) + 0.1263\left(\frac{z+1}{z-1}\right)\right)^2 + \sqrt{2}\left(5.2421\left(\frac{z-1}{z+1}\right) + 0.1263\left(\frac{z+1}{z-1}\right)\right) + 1} \quad (3.157)$$

después de efectuar el desarrollo algebraico se tiene

$$H(z) = \frac{0.0009 - 0.0019z^{-2} + 0.0009z^{-4}}{1 - 3.8650z^{-1} + 5.6474z^{-2} - 3.6969z^{-3} + 0.915z^{-4}} \quad (3.158)$$

Cálculo de la ecuación en diferencias $y(n)$:

Como $H(z) = \frac{Y(z)}{X(z)}$ entonces $Y(z) = H(z)X(z)$, aplicando transformada Z inversa

$$\begin{aligned} y(n) &= 0.0009x(n) - 0.0019x(n-2) + 0.0009x(n-4) \\ &+ 3.8650y(n-1) - 5.6474y(n-2) + 3.6969y(n-3) - 0.915y(n-4) \end{aligned} \quad (3.159)$$

Solución por transformación de $H(s) \rightarrow H(s) \rightarrow H(z)$

Cálculo de la transformación de las frecuencias de corte

$$\Omega_{iL} = \operatorname{tg}\left(\frac{0.8\pi}{20}\right) = 0.1263 \quad \Omega_{iH} = \operatorname{tg}\left(\frac{1.2\pi}{20}\right) = 0.1907 \quad (3.160)$$

entonces de la tabla (3.6):

$$s \rightarrow \frac{s^2 + 0.02409}{s(0.0644)}$$

sustituyendo en $H(s)$ de orden $n=2$ de los polinomios de Butterworth

$$H_{FPBW}(s) = \frac{1}{s^2 + \sqrt{2}s + 1} \Big|_{s=\left(\frac{s^2+0.02409}{s(0.0644)}\right)} \quad (3.161)$$

$$H_{FPBW}(s) = \frac{1}{\left(\frac{s^2+0.02409}{s(0.0644)}\right)^2 + \sqrt{2}\left(\frac{s^2+0.02409}{s(0.0644)}\right) + 1} \quad (3.162)$$

después de efectuar las reducciones algebraicas

$$H_{FPBW}(s) = \frac{0.004147s^2}{s^4 + 0.09107s^3 + 0.052147s^2 + 0.002186s + 0.000576} \quad (3.163)$$

aplicando la transformada bilineal con $T=1$

$$H(z) = H(s) \Big|_{s=\frac{1-z^{-1}}{1+z^{-1}}} \quad (3.164)$$

se obtiene

$$H(z) = \frac{0.001 - 0.002z^{-2} + 0.001z^{-4}}{1 - 3.8629z^{-1} + 5.6419z^{-2} - 3.6919z^{-3} + 0.9135z^{-4}} \quad (3.165)$$

Del ejemplo se puede concluir que la utilización de la TB por los dos métodos expuestos dan resultados iguales de $H(z)$ a excepción de errores numéricos al despreciar decimales en los cálculos algebraicos. El método del mapeo de $H(s)$ a $H(z)$ por TB tiene la ventaja que el valor de n se calcula directamente de las especificaciones del diseño en frecuencia y de la aproximación polinómica a utilizar, es decir, que este método es más directo.

Respuesta espectral gráfica

Para la $H(z)$ calculada en el ejemplo anterior, se obtiene la Transformada en el Tiempo Discreto de Fourier (DTFT), es decir que la función $H(z)$ se evalúa en $e^{j\omega}$ lo que daría la función $H(\omega)$

$$H(\omega) = \frac{0.0009 - 0.0019e^{-2j\omega} + 0.0009e^{-4j\omega}}{1 - 3.8650e^{-j\omega} + 5.6474e^{-2j\omega} - 3.6969e^{-3j\omega} + 0.915e^{-4j\omega}} \quad (3.166)$$

Utilizando Euler esta función se puede expresar en términos de funciones seno y coseno, para tener una gráfica representativa de la magnitud hay que programar $H(\omega)$ o recurrir a algún programa para Procesamiento Digital de Señales, los resultados obtenidos se muestran en la figura (3.17). Donde se puede observar que los cálculos desarrollados corresponden a las especificaciones solicitadas para este diseño.

Implementación en estructura cascada

En esta forma, la ecuación (3.165) es factorizada en secciones de segundo orden, llamadas secciones *biquads*. La función de transferencia del sistema es representada como el producto de funciones biquads. Cada función biquad es implementada en forma directa o canónica y la función completa del sistema como secciones *biquad* en cascada. Esta forma de implementación permite disminuir la acumulación de errores numéricos cuando los filtros se programan en aritmética de punto fijo. De la ecuación (3.165) y factorizando la $H(z)$ calculada

$$H(z) = 0.0009 \left(\frac{1 + 2z^{-1} + z^{-2}}{1 - 1.9196589z^{-1} + 0.950475z^{-2}} \right) \left(\frac{1 - 2z^{-1} + z^{-2}}{1 - 1.9452z^{-1} + 0.96265z^{-2}} \right) \quad (3.167)$$

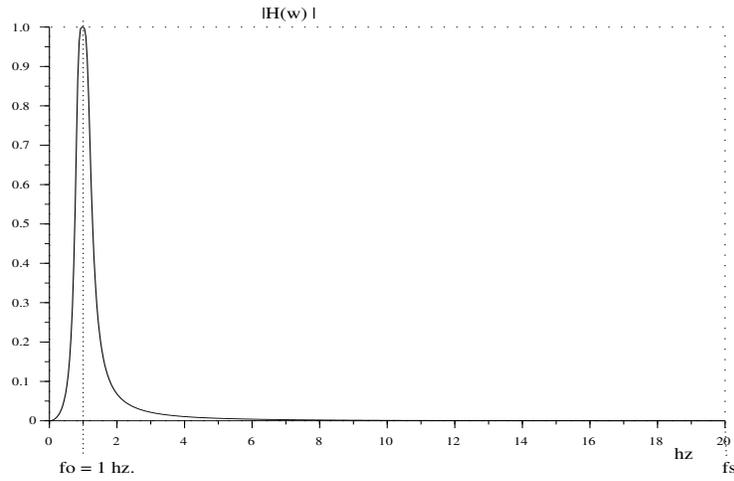


Figura 3.17: $|H(\omega)|$ de un FPBW con $f_0 = 1$ hz y $\Delta BW = 0.4$ hz.

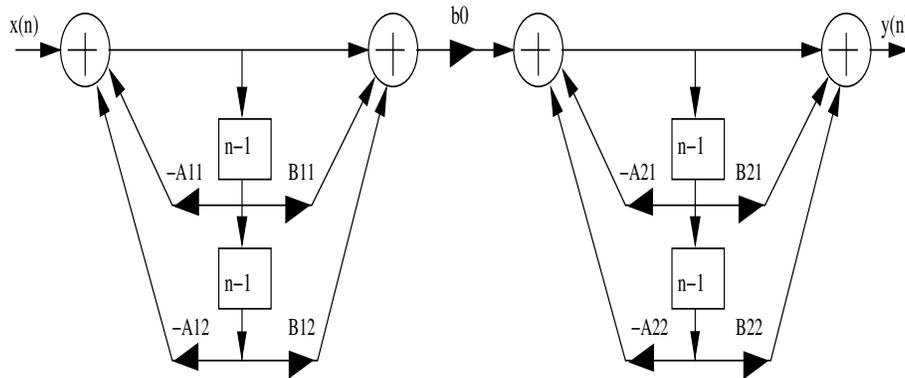


Figura 3.18: Filtro IIR en cascada

Los coeficientes para la forma cascada de la figura 3.18 son:

$$\begin{aligned}
 b_0 &= 0.0009, B_{10} = 1, B_{11} = 2.00, B_{12} = 1.00 \\
 B_{20} &= 1, B_{21} = -2.00, B_{22} = 1.00 \\
 A_{10} &= 1, A_{11} = -1.9196589, A_{12} = 0.95047585 \\
 A_{20} &= 1, A_{21} = -1.9452919, A_{22} = 0.96265027
 \end{aligned}$$

Diseño de filtros IIR con especificaciones digitales

Si las especificaciones del diseño se dan en frecuencia digital ω , entonces para utilizar el diseño por transformación analógica a digital se transforma la frecuencia de corte (especificaciones en frecuencia) por medio de la fórmula

$$\Omega = \frac{2}{T} \tan\left(\frac{\omega}{2}\right)$$

a especificaciones analógicas. Posteriormente se diseña el filtro analógico con base al polinomio normalizado de la aproximación, se desnormaliza el filtro y se convierte a filtro digital por medio de la TB, durante este proceso el parámetro T es transparente y se puede fijar arbitrariamente $T = 1$.

Si se tiene una señal discreta $x(n)$ de N puntos y se quiere diseñar un filtro que seleccione ciertas frecuencias del espectro $X(e^{j\omega})$, otro procedimiento de análisis es aplicando a $x(n)$ la Transformada Discreta de Fourier (DFT) para obtener su espectro discreto $X(k)$, el diseñador puede seleccionar la o las frecuencias de corte digital de interés con la relación en el intervalo $0 < \omega < \pi$, se calcula

$$\omega_{cd} = \frac{\pi k_i}{N/2}$$

donde k_i representa un punto en la frecuencia discreta.

3.5. Diseño por localización de los polos y ceros en el plano z

Este método consiste en ir colocando los polos y ceros en plano z con base a la experiencia del diseñador y luego evaluar el desempeño de $H(z)|_{z=e^{j\omega}}$, hasta obtener una respuesta en frecuencia deseada, en general se puede decir que es un método de prueba y error.

Representación de $H(z)$ en el plano z

Un sistema discreto se puede representar en el dominio de z utilizando el patrón de polos y ceros, y si se utiliza la relación $z = e^{j\omega T}$ se tiene la representación en el dominio de la

frecuencia por medio de la amplitud $|H(e^{j\omega T})|$ y la fase $\phi(\omega)$ [3]. La respuesta de un sistema se puede escribir

$$H(z) = \frac{Y(z)}{X(z)} = K_0 \frac{\prod_{i=1}^N (z - c_i)}{\prod_{i=1}^N (z - p_i)} \quad (3.168)$$

donde las constantes c_i son los ceros, p_i los polos de $H(z)$ y K_0 una ganancia.

Si sustituimos $z = e^{j\omega T}$

$$\begin{aligned} H(e^{j\omega T}) &= |H(e^{j\omega T})| e^{j\phi(\omega)} \\ H(e^{j\omega T}) &= K_0 \frac{\prod_{i=1}^N (e^{j\omega T} - c_i)}{\prod_{i=1}^N (e^{j\omega T} - p_i)} \\ H(e^{j\omega T}) &= K_0 \frac{\prod_{i=1}^N M_{c_i} e^{j\omega c_i}}{\prod_{i=1}^N M_{p_i} e^{j\omega p_i}} \end{aligned} \quad (3.169)$$

la amplitud $|H(e^{j\omega T})|$ es un cociente de productos de amplitudes y la fase $\phi(\omega)$ es un resta de las fases de los ceros y los polos:

$$|H(\omega)| = K_0 \frac{\prod_{i=1}^N M_{c_i}}{\prod_{i=1}^N M_{p_i}} \quad (3.170)$$

$$\phi(\omega) = \sum_{i=1}^N \phi_{c_i} - \sum_{i=1}^N \phi_{p_i} \quad (3.171)$$

Las respuestas de magnitud y fase del sistema pueden obtenerse para cada frecuencia ω determinando la magnitud y el ángulo de fase de los polos y ceros respecto del punto $e^{j\omega T}$ sobre el círculo unitario, como se observa en la figura 3.19. Es decir, que se puede obtener un cálculo geométrico.

De la figura 3.19, los puntos marcados tienen el siguiente significado:

A representa la frecuencia cero

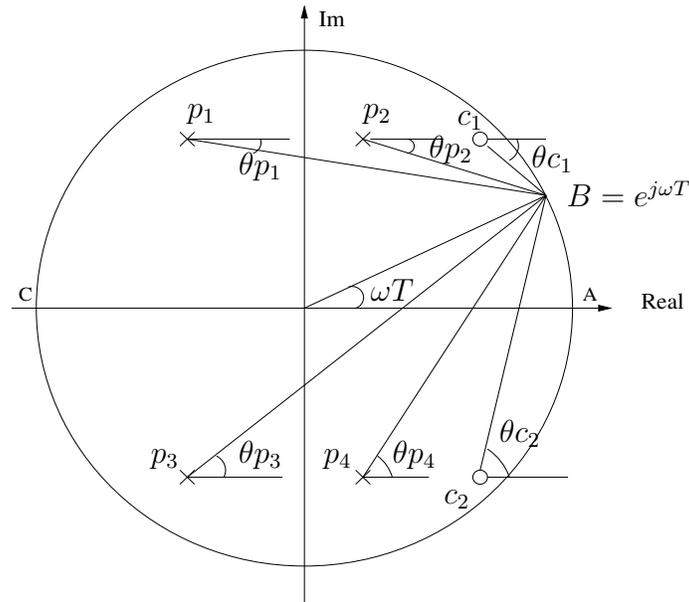
B cualquier punto sobre el círculo unitario donde se evalúa $H(e^{j\omega T})$

C la frecuencia $f_s/2$, donde f_s es la frecuencia de muestreo.

El comportamiento de $H(e^{j\omega T})$ está determinado en el intervalo $0 \leq f \leq f_s/2$ donde $H(e^{j\omega T}) = H(e^{-j\omega T})$ es simétrica respecto de $f_s/2$.

Es posible obtener la magnitud de $H(e^{j\omega T})$ en un punto dado $\omega = \omega_0$ a través de la construcción de todos los vectores que unen el punto $z = e^{j\omega_0 T}$ (sobre el círculo unitario) con cada cero y polo de $H(z)$. De la evaluación de los productos de magnitudes, resulta la magnitud de $H(e^{j\omega T})$ y $\phi(\omega)$ para una ω particular.

Este método consiste en ubicar los polos en los puntos cercanos al círculo unitario en las frecuencias que se quieran enfatizar y los ceros cerca de las frecuencias a desenfatar. Para este método se imponen las siguientes limitaciones


 Figura 3.19: Polos y ceros en el plano z

- Todos los polos deben localizarse dentro del círculo unitario para garantizar la estabilidad del sistema.
- Por cada polo o cero complejos debe existir su conjugado con el objeto de que los coeficientes del diseño sean reales.

La función de transferencia $H(z)$ se expresa

$$H(z) = \frac{\sum_{k=0}^q b_k z^{-k}}{1 + \sum_{k=0}^p a_k z^{-k}} = b_0 \frac{\prod_{k=0}^q (1 - c_k z^{-1})}{\prod_{k=0}^p (1 - p_k z^{-1})} \quad (3.172)$$

donde b_0 es una constante de ganancia seleccionada para normalizar $|H(\omega)|$ a una frecuencia específica tal que $|H(\omega_0)| = 1$, ω_0 es una frecuencia dentro de la banda de paso del filtro, c_k son los ceros de $H(z)$ y p_k son los polos de $H(z)$. Usualmente el número de polos p se selecciona mayor o igual que el número de ceros q .

Para el diseño de **filtros digitales paso bajas** los polos se deben ubicar cerca al círculo unitario correspondiente a los puntos de la frecuencia de corte ω_c cerca a $\omega = 0$, esto es bajas frecuencias. Los ceros deben localizarse cerca al círculo unitario correspondiente a los puntos de la frecuencia alta, $\omega = \pi$.

Para el diseño de **filtros digitales paso altas** la localización de polos y ceros es lo opuesto a filtros paso bajas, en general pueden darse varias posibilidades como se observa

en la figura 3.20. El método es similar para el diseño de FPBW y FSBW, un filtro FPBW debe contener uno o mas pares de polos conjugados complejos cerca del círculo unitario en la vecindad de las frecuencias que constituyen la banda de paso del filtro.

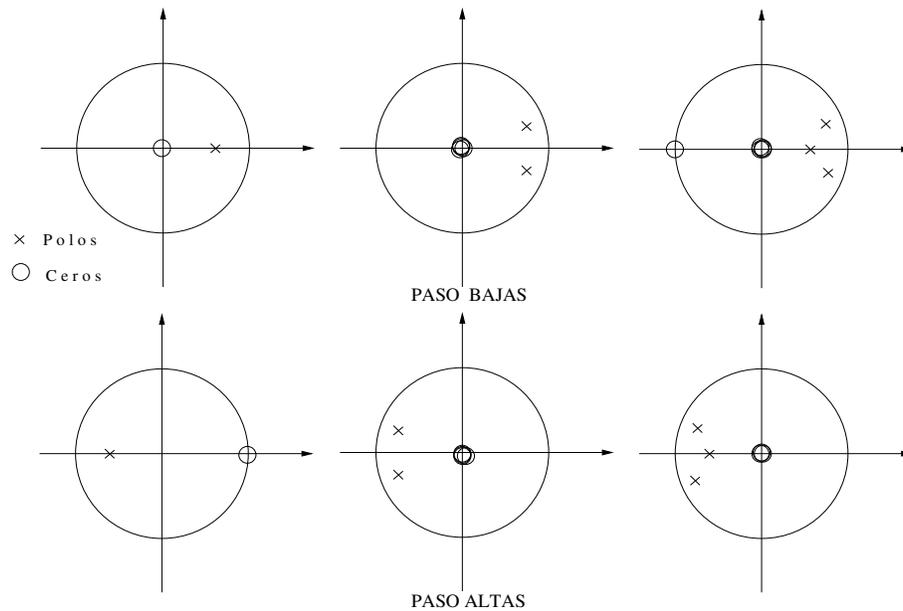


Figura 3.20: Diseño de filtros por ubicación de polos y ceros en el plano z

Metodología de diseño por ubicación de polos y ceros en plano z

1. Determinar las frecuencias ω de interés con base a las especificaciones.
2. Ubicar los polos y ceros en el plano z .
3. Escribir la función de transferencia $H(z)$.
4. Evaluar el desempeño de $H(e^{j\omega})$, hasta obtener una respuesta en frecuencia deseada, si no cumple con especificaciones regresar al paso dos.
5. Normalizar la función de transferencia y obtener $y(n) = TZI\{H(z)X(z)\}$.
6. Identificar los coeficientes e implementar el filtro.

Este método de diseño no es muy práctico ya que no se puede tener mucho control sobre las especificaciones.

Diseño de un filtro supresor de banda o ranura

Para el diseño de un filtro supresor de banda a frecuencia ω_0 simplemente se introducen un par de ceros conjugados sobre el círculo unitario a un ángulo ω_0 , $z_{1,2} = e^{\pm j\omega_0}$, cuyo diseño es del tipo solo ceros y por consecuencia un filtro tipo FIR. Esto produce un ancho de banda muy grande, lo que implica que las frecuencias vecinas alrededor de ω_0 sean atenuadas fuertemente. Para evitar este efecto se debe recurrir a otras técnicas utilizando un filtro tipo IIR introduciendo dos polos conjugados en la vecindad de ω_0 , es decir

$$p_{1,2} = r_p e^{\pm j\omega_0}$$

El efecto de estos polos es producir resonancias cerca de los ceros reduciendo el ancho de banda del filtro de ranura y el rizo en la banda de paso. La función de transferencia resultante sería:

$$H(z) = b_0 \frac{(1 - r_c e^{j\omega_0} z^{-1})(1 - r_c e^{-j\omega_0} z^{-1})}{(1 - r_p e^{j\omega_0} z^{-1})(1 - r_p e^{-j\omega_0} z^{-1})} = b_0 \frac{1 - 2r_c z^{-1} \cos\omega_0 + r_c^2 z^{-2}}{1 - 2r_p z^{-1} \cos\omega_0 + r_p^2 z^{-2}} \quad (3.173)$$

este diseño también genera filtros paso banda:

- si $r_c < r_p$ se obtiene un filtro FPBW
- si $r_c > r_p$ se obtiene un filtro FSBW
- si $r_c \approx r_p$ el efecto de los polos y los ceros se anulan, se tiene un filtro paso todo.

La cantidad de ganancia y atenuación es controlada por la diferencia entre r_p y r_c , la distancia r_p al círculo unitario determinará el ancho de banda del filtro.

Una desventaja del método es que puede conducir a ensayos de prueba y error para alcanzar un buen desempeño.

Ejemplo de un filtro de ranura

Diseñar un filtro digital FSBW con $\omega_0 = \pi/4$ con ancho de banda muy reducido. La calidad de un filtro tipo ranura es controlada por la distancia de los polos a los ceros. Para los polos y ceros:

$$c_1 = 0.7071 + j0.7071$$

$$c_2 = 0.7071 - j0.7071$$

$$p_1 = 0.6364 + j0.6364$$

$$p_2 = 0.6364 - j0.6364$$

por la ubicación de los polos, este filtro puede ser visto como un filtro de ranura con frecuencia $\pi/4$ o 0.125 de la frecuencia de muestreo. Sustituyendo los polos y ceros en la ecuación (3.172)

$$H(z) = \frac{(z - (0.7071 + j0.7071))(z - (0.7071 - j0.7071))}{(z - (0.6364 + j0.6364))(z - (0.6364 - j0.6364))}$$

la función de transferencia normalizada

$$H(z) = \frac{1 - 1.414z^{-1} + z^{-2}}{1 - 1.273z^{-1} + 0.810z^{-2}}$$

con coeficientes

$$b_i = \{ 1, -1.41, 1 \}$$

$$a_i = \{ 1, -1.273, 0.810 \}$$

La magnitud de la respuesta en frecuencia $|H(\omega)|$ en decibeles para este diseño se observa en la figura 3.21.

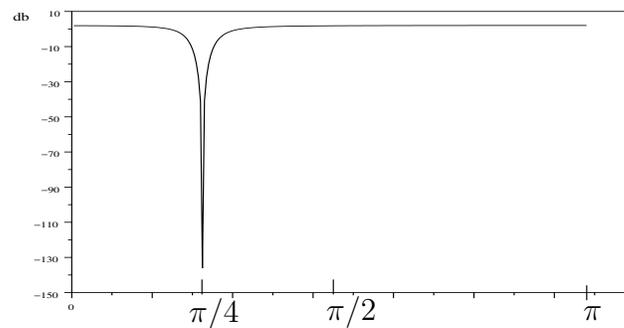


Figura 3.21: Respuesta del filtro FSBW diseñado

Sistemas de fase mínima

Para finalizar este capítulo se agrega una característica de interés en filtros digitales, ya que el concepto de fase mínima va a ser útil en el siguiente capítulo y en muchas aplicaciones cuando se pretende encontrar una función de transferencia inversa, lo que implica que el sistema sea invertible. La importancia de los sistemas o filtros de fase mínima, es que al invertirlos su respuesta en frecuencia producen sistemas estables, la invertibilidad de un sistema es de mucho interés en el estudio de la deconvolución. Entonces para un sistema o un filtro con una función de transferencia inversa estable, debe ser de fase mínima y no contener ceros sobre el círculo unitario.

Se dice que un sistema es de fase mínima si su función de transferencia satisface:

1. Los polos de $H(z)$ están dentro del círculo unitario en el plano z .
2. Los ceros de $H(z)$ están dentro del círculo unitario.

3. El cambio de fase $\Phi(\omega)$ de $\Phi(0)$ a $\Phi(\pi)$ es cero [32].

Para una respuesta en frecuencia $|H(e^{j\omega})|$, la respuesta en fase del filtro asume el mínimo valor posible. Si conocemos la respuesta en amplitud o la respuesta en fase del filtro, se puede determinar la otra componente. Un filtro $H(z)$ se dice que es de fase mínima cuando el retardo de grupo es el mínimo posible para todas las frecuencias.

Sistemas de fase no mínima o fase máxima

Se le llama así a un sistema, cuando todos los ceros de la función de transferencia del sistema están localizados fuera del círculo unitario en el plano z . Para una respuesta en amplitud dada, la respuesta en fase del sistema asume el máximo valor posible.

Para un sistema de fase máxima, el cambio de fase $\Phi(\omega)$ de $\Phi(0)$ a $\Phi(\pi)$ es π . Las características en fase de un sistema es una medida del tiempo de retardo que las componentes en frecuencia de la señal de entrada sufren al pasar por el sistema. Una característica de fase mínima implica una función de retardo mínimo, y lo contrario una de fase máxima [32]. Un filtro $H(z)$ se dice que es de fase máxima cuando el retardo de grupo es el máximo para ciertas frecuencias.

3.6. Diseño de filtro IIR utilizando programas

Como se mencionó para el caso de los filtros FIR, existe software libre y comercial para el diseño de filtro IIR. En el paquete MATLAB en el “toolbox” para procesamiento de señales se tienen las funciones de interés que el lector puede verificar a través de las ayudas.

butterord	butter	bilinear	buttap
cheb1ord	cheby1	chevy2	chev1ap
ellipord	ellip	residuez	chev2ap
besself	besselap	poly2rc	ellipap
impinvar	tf2zp	zp2sos	filter
filter2	filtfilt	freqz	freqs

3.7. Resumen

En este capítulo se han analizado los métodos más utilizados en el diseño de filtros digitales IIR, sus características, ventajas y desventajas. Se hace notar que todavía existen otras técnicas para el diseño de filtros IIR que el lector interesado puede consultar y entenderlas con las bases que se han dado anteriormente. Sin embargo, con los conceptos abordados en este capítulo el lector está en capacidad de diseñar filtros de muy buen desempeño e incluso realizar diseños que van a operar en tiempo real.

Capítulo 4

Predicción lineal

Una vez estudiado y analizado los filtros digitales FIR e IIR, se está en la posibilidad de aplicar estos filtros al análisis de la codificación por predicción lineal (LPC) que se utiliza ampliamente en estimación, compresión de señales, procesamiento de voz, detección, imágenes, cancelación de ruido en comunicaciones, instrumentación, geofísica, etc.

En la estimación de señales, el diseño de filtros se realiza con base a un criterio de optimización y comúnmente el más utilizado es la minimización del error cuadrático medio (MSE), para este criterio sólo se requieren estadísticos de segundo orden, es decir, el valor medio y correlaciones de un proceso estacionario.

En este capítulo se realizará un estudio de la predicción lineal con base a los filtros ya estudiados, a la predicción se le denota lineal ya que los filtros FIR a utilizar son lineales, de este análisis se deducirán los filtros “lattice” cuyas estructuras las relacionaremos con los diseños abordados. Dentro de los análisis se obtendrá la ecuación normal y la de Yule-Walker que son de amplia aplicación en el procesamiento de señales y se expondrán algunos algoritmos de solución.

4.1. Predicción lineal

Un problema de interés en el procesamiento digital de señales es predecir el valor de una señal en un tiempo dado conociendo un conjunto de valores de un proceso estacionario. Esto es, hacer una estimación de una muestra particular de la señal con base a valores conocidos de la misma en otros tiempos. Si la muestra a estimar es la presente o futura con base a muestras anteriores conocidas, a la predicción se le conoce “*forward*” o *hacia adelante*, si se va a realizar una estimación de una muestra pasada con un retardo cualquiera con base a muestras conocidas, a la predicción se le conoce “*backward*” o *hacia atrás*.

Para analizar ambos procesos empezamos por el modelo de predicción de la figura 4.1.

Dado un conjunto de valores $x(n), x(n-1), x(n-2), \dots, x(n-M)$ de un proceso estacionario como se muestra en la figura 4.2, donde:

- El valor a estimar o predecir puede ser el valor futuro $\hat{x}(n + \Delta)$

$$\hat{x}(n + \Delta) = f(x(n), x(n-1), x(n-2), \dots, x(n-M))$$

donde:

Δ es un tiempo futuro

$\hat{x}(\cdot)$ es el valor estimado de $x(\cdot)$

$d(n)$ es el valor de una señal deseada o muestra a estimar

$e(n)$ es el error de predicción

- Si se conocen las muestras anteriores a tiempo actual se puede predecir el valor actual $\hat{x}(n)$, obteniendo el *predictor hacia adelante* (FLP), un diagrama de este predictor se ilustra en la figura 4.1, donde

$$\hat{x}(n) = f(x(n-1), x(n-2), \dots, x(n-M))$$



Figura 4.1: Predicción lineal

- Si se conoce la muestra actual $x(n)$ y las $M - 1$ muestras anteriores $x(n), x(n-1), x(n-2), \dots, x(n-M+1)$, se puede estimar la muestra $\hat{x}(n-M)$, a este predictor se le conoce como *predictor hacia atrás* (BLP)

$$\hat{x}(n-M) = f(x(n), x(n-1), x(n-2), \dots, x(n-M+1))$$

Se dice que la predicción es lineal, si el valor a estimar se puede escribir como una combinación lineal de los valores conocidos. Los requerimientos de la predicción se deben especificar de tal forma que el error de predicción sea el mínimo.

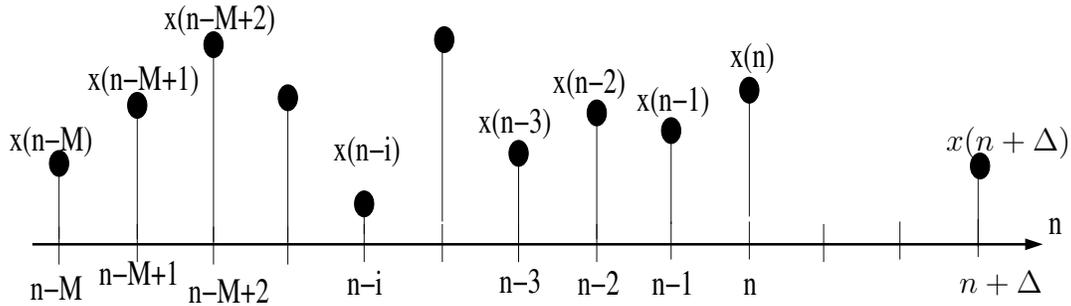


Figura 4.2: Predicción lineal sobre una señal $x(n)$

4.2. Predicción hacia adelante y ecuación normal

Considerando la secuencia de datos $x(n-1), x(n-2), \dots, x(n-M)$ obtenidas de un proceso estacionario, el problema de predicción lineal hacia adelante (FLP) consiste en utilizar este conjunto de muestras y predecir la muestra actual $x(n)$ utilizando un filtro FIR, como se observa en la figura 4.3.

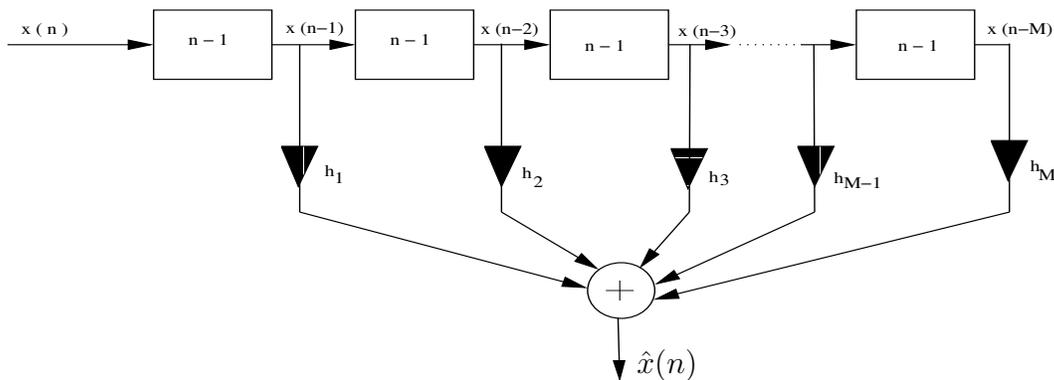


Figura 4.3: Predicción lineal hacia adelante

Definiendo el *error de predicción hacia adelante* $e_M^f(n)$

$$e_M^f(n) = x(n) - \hat{x}(n) \tag{4.1}$$

donde $x(n)$ es equivalente a la respuesta deseada, el subíndice M significa el orden del predictor o la cantidad de muestras que utiliza el predictor, y el superíndice f quiere decir hacia adelante como se ve en la figura 4.4. Para la implementación del predictor se utiliza un filtro FIR algunas veces llamado transversal con líneas de retardo (TDL) como se ilustra en las figuras 4.3 y 4.4.

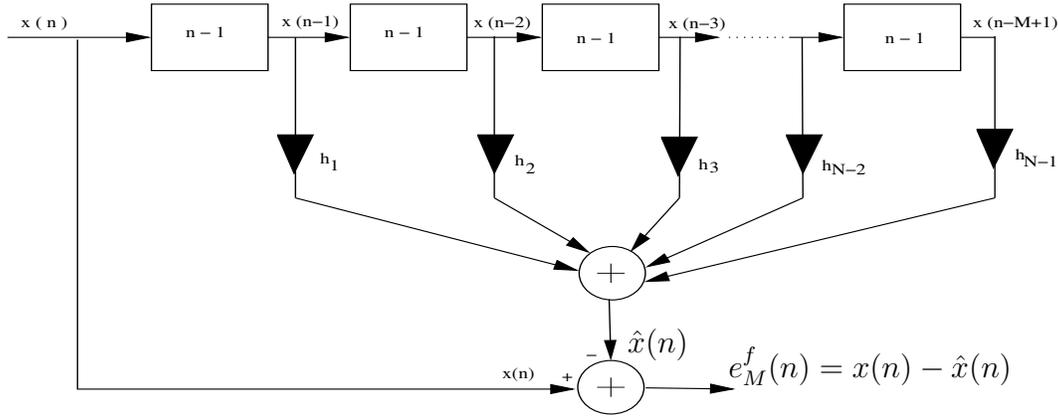


Figura 4.4: Error de predicción hacia adelante

La salida estimada $\hat{x}(n)$ se puede calcular como la convolución entre los retardos de $x(n)$ y los coeficientes $h(n)$ del filtro

$$\hat{x}(n) = \sum_{k=1}^M h(k)x(n-k) \quad (4.2)$$

entonces

$$e_M^f(n) = x(n) - \sum_{k=1}^M h(k)x(n-k) \quad (4.3)$$

Para encontrar los coeficientes $h_0(n)$ que minimizan el error entre la señal deseada $x(n)$ y la estimada $\hat{x}(n)$ es necesario aplicar algún criterio de minimización. Un criterio comúnmente utilizado es el de los mínimos cuadrados (MSE).

Desarrollando el error hacia delante $e_M^f(n)$ al cuadrado

$$|e_M^f(n)|^2 = [x(n) - \hat{x}(n)]^2 = \left[x(n) - \sum_{i=1}^M h(i)x(n-i) \right]^2 \quad (4.4)$$

calculando la esperanza matemática y definiendo una función de costo $J_M^f(n)$ como la energía de error de predicción hacia adelante

$$J_M^f(n) = |e_M^f(n)|^2 = E\{x^2(n)\} - 2E\{x(n)\hat{x}(n)\} + E\{\hat{x}^2(n)\} \quad (4.5)$$

sustituyendo la salida del filtro $\hat{x}(n)$

$$J_M^f(n) = E\{x^2(n)\} - 2 \sum_{i=1}^M h(i)E\{x(n)x(n-i)\} + \sum_{i=1}^M \sum_{k=1}^M h(i)h(k)E\{x(n-i)x(n-k)\} \quad (4.6)$$

Considerando que la entrada $x(n)$ y la señal deseada $d(n)$ o estimada $\hat{x}(n)$ son conjuntamente estacionarias en el sentido amplio, además significa que éstas son estacionarias en el sentido amplio e independientes, por lo tanto, sus funciones de correlación son independientes. Si los valores medios de las señales de entrada $x(n)$ y estimada $\hat{x}(n)$ son cero y se conoce la autocorrelación $r_{xx}(i)$ de $x(n)$, entonces definiendo los términos:

- El valor medio cuadrático de la señal deseada

$$r_{xx}(0) = E\{x^2(n)\} \quad (4.7)$$

- La esperanza matemática de la correlación entre la señal deseada y la señal retrasada

$$r_{xx}(i) = E\{x(n)x(n-i)\} \quad (4.8)$$

entonces el segundo término de la derecha de la ecuación (4.6) puede ser escrito como

$$\sum_{i=1}^M h(i)E\{x(n)x(n-i)\} = \sum_{i=1}^M h(i)r_{xx}(i) \quad (4.9)$$

- La esperanza matemática de la autocorrelación de la entrada

$$r_{xx}(k-i) = E\{x(n-i)x(n-k)\} \quad (4.10)$$

Sustituyendo estas definiciones en la función de costo

$$J_M^f(n) = r_{xx}(0) - 2 \sum_{i=1}^M h(i)r_{xx}(i) + \sum_{i=1}^M \sum_{k=1}^M h(i)h(k)r_{xx}(k-i) \quad (4.11)$$

para minimizar el error cuadrático se deriva $J_M^f(n)$ respecto de los parámetros $h(n)$ y se iguala a cero

$$\frac{\partial J_M^f(n)}{\partial h(n)} = -2r_{xx}(k) + 2 \sum_{k=1}^M h(k)r_{xx}(k-i) \quad k = 0, 1, 2, 3, \dots, \infty \quad (4.12)$$

obteniendo la *ecuación de Wiener-Hopf* [14], [15], [30], [32], [41], de donde se puede despejar la respuesta al impulso $h(n)$ del filtro en el tiempo discreto

$$\sum_{m=1}^M h(m)r_{xx}(k-m) = r_{xx}(k) \quad k = 1, 2, 3, 4, \dots, M \quad (4.13)$$

sustituyendo esta solución en la función de costo, se obtiene el error cuadrático promedio mínimo

$$J_{M-min}^f(n) = r_{xx}(0) - \sum_{k=1}^M h(k)r_{xx}(k) \quad (4.14)$$

Ecuación normal y principio de ortogonalidad

La ecuación (4.13) es conocida como *ecuación normal* y consiste de un sistema de M ecuaciones simultáneas cuyas incógnitas son los coeficientes $h(n)$. De los resultados anteriores también se puede deducir que

$$E\{e(n)x(n)\} = 0 \quad k = 0, 1, 2, 3, \dots, M - 1 \quad (4.15)$$

Esta ecuación es conocida como *el principio de ortogonalidad*, y es un resultado directo de la minimización del error medio cuadrático. La condición de ortogonalidad establece que el producto escalar del error con los datos en la predicción es cero, es decir, que la secuencia del error no está correlacionada con los datos. El principio de ortogonalidad establece que la longitud de $J_M^f(n)$ es mínima cuando el error $e(n)$ es perpendicular a los datos $x(n - i)$ como se observa en la figura 4.5, una interpretación geométrica permite observar que la señal estimada $\hat{x}(n)$ se obtiene al proyectar la señal deseada $x(n)$ en el espacio de los datos de $x(n - i)$ como una combinación lineal de éstos con los coeficientes $h(n)$.

La solución de la ecuación normal para los parámetros $h(n)$ significa encontrar un conjunto de coeficientes del filtro predictor (ecuación 4.1) tal que minimice el error cuadrático medio, si este error es mínimo significa que se ha obtenido una muy buena estimación de $x(n)$.

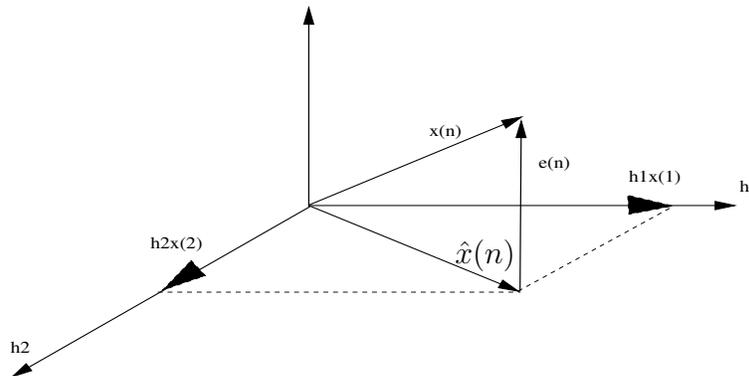


Figura 4.5: Condición de ortogonalidad

Desarrollando la sumatoria de la ecuación (4.13) para el índice m

$$\begin{aligned} h(1)r_{xx}(k - 1) + h(2)r_{xx}(k - 2) + h(3)r_{xx}(k - 3) + \dots \\ \dots + h(M)r_{xx}(k - M) = r_{xx}(k) \end{aligned} \quad (4.16)$$

y para $k = 1, 2, \dots, M$, se tiene el sistema matricial

$$\begin{bmatrix} r_{xx}(0) & r_{xx}(-1) & r_{xx}(-2) & \cdots & r_{xx}(M-1) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(M-2) \\ r_{xx}(2) & r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(M-3) \\ \vdots & \vdots & \vdots & & \vdots \\ r_{xx}(M-1) & r_{xx}(M-2) & r_{xx}(M-3) & \cdots & r_{xx}(0) \end{bmatrix} \begin{bmatrix} h(1) \\ h(2) \\ h(3) \\ \vdots \\ h(M) \end{bmatrix} = \begin{bmatrix} r_{xx}(1) \\ r_{xx}(2) \\ r_{xx}(3) \\ \vdots \\ r_{xx}(M) \end{bmatrix} \quad (4.17)$$

Por tanto, para calcular los coeficientes $h(n)$ del filtro predictor se puede utilizar la ecuación normal adecuada al problema de la predicción hacia adelante.

La ecuación normal (4.17) se puede escribir en forma matricial como

$$\mathbf{R}_x \mathbf{h} = \mathbf{r}_x \quad (4.18)$$

y si existe la inversa de la matriz de autocorrelaciones de \mathbf{R}_x , se tiene la solución

$$\mathbf{h} = \mathbf{R}_x^{-1} \mathbf{r}_x \quad (4.19)$$

y el error medio cuadrático mínimo

$$J_{M-min}^f(n) = r_{xx}(0) - \mathbf{r}_x^T \mathbf{R}_x^{-1} \mathbf{r}_x = r_x(0) - \mathbf{r}_x^T \mathbf{h} \quad (4.20)$$

donde el subíndice T significa transpuesta.

Propiedades de la matriz de autocorrelaciones \mathbf{R}

- La matriz de autocorrelación \mathbf{R} es simétrica o hermitiana en el caso complejo, es decir, que $\mathbf{R} = \mathbf{R}^T$, entonces sus vectores característicos son ortogonales, si \mathbf{Q} es una matriz de vectores característicos de la matriz \mathbf{R} , entonces $\mathbf{Q}_m^T \mathbf{Q}_n = 0$ [41].
- Si la matriz \mathbf{R} es real y simétrica, todos sus valores característicos deben ser reales o iguales a cero [41].
- La matriz \mathbf{Q} de vectores característicos puede ser normalizada, de tal forma que $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ [41].
- Para procesos estacionarios la matriz \mathbf{R} es del tipo Toeplitz, esto es, que todos los elementos de cada diagonal son iguales [15].
- La matriz \mathbf{R} es positivamente definida, implica que cumple con $\mathbf{x}^T \mathbf{R} \mathbf{x} > 0$ para cualquier vector \mathbf{x} .
- Cuando los elementos de un vector del proceso estacionario en observación son arreglados en forma “backward”, el efecto es equivalente a la transposición de la matriz de autocorrelación [15].

- La matriz \mathbf{R}_{p+1} de orden $p+1$ para un proceso estacionario se puede escribir en función de la matriz \mathbf{R}_p de orden p [15].

4.2.1. Relación entre predicción lineal y el modelo autorregresivo

La predicción lineal y el modelado autorregresivo (AR) de un proceso aleatorio están íntimamente relacionados. Considerando un proceso estacionario $x(n)$ que es la entrada a un filtro transversal con función de transferencia $A(z)$, coeficientes a_i y salida $v(n)$ del tipo ruido blanco con media cero y varianza σ^2 , por sus características al filtro $A(z)$ se conoce como *filtro blanqueador*, y su función de transferencia está dada por

$$A(z) = 1 + \sum_{i=1}^M a_i z^{-i} = \frac{V(z)}{X(z)} \quad (4.21)$$

entonces la respuesta al impulso $h(n)$ del filtro blanqueador es de *duración finita FIR* como se muestra en la figura 4.6.

Por otro lado se puede obtener un modelo con una señal $x(n)$ como la salida y una señal de ruido blanco $v(n)$ como la entrada, esto significa un sistema inverso al filtro blanqueador con respuesta al impulso $g(n)$ de *duración infinita*. Considerando un sistema tipo IIR sólo polo, como se observa en la figura 4.7 que también se le conoce como *autorregresivo (AR)*, con función de densidad de potencia

$$S_x(z) = \frac{\sigma^2}{\prod_{i=1}^M (1 - \alpha_i z^{-i})(1 - \alpha_i^* z)} \quad (4.22)$$

donde los coeficientes α_i son los polos del sistema y $|\alpha_i| < 1$ para toda i .

Del sistema FIR, si la función de transferencia $A(z)$ es de fase mínima existe su inversa, y si definimos al sistema inverso como

$$G(z) = \frac{1}{A(z)} = \frac{X(z)}{V(z)} = \frac{1}{\prod_{i=1}^M (1 - \alpha_i z^{-i})} = \frac{1}{1 + \sum_{i=1}^M a_i z^{-i}} \quad (4.23)$$

Entonces los coeficientes a_i de ambos sistemas son los mismos.

De las ecuaciones (4.23), (4.21) y de un modelo autorregresivo como se observa en la figura 4.7 se pueden obtener las ecuaciones en diferencias en el dominio del tiempo

$$x(n) + a_1 x(n-1) + a_2 x(n-2) + a_3 x(n-3) + \dots + a_M x(n-M) = v(n) \quad (4.24)$$

despejando $x(n)$

$$x(n) = -a_1 x(n-1) - a_2 x(n-2) - a_3 x(n-3) - \dots - a_M x(n-M) + v(n) \quad (4.25)$$

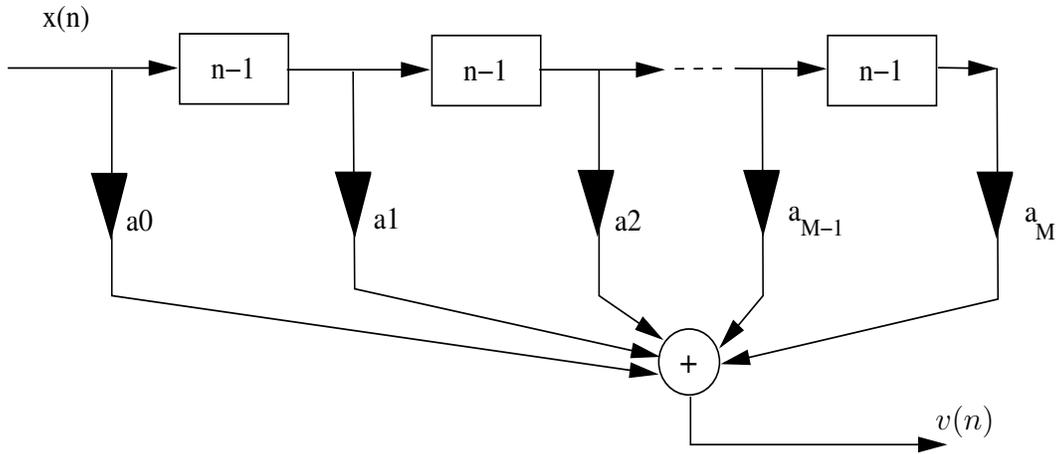


Figura 4.6: Filtro blanqueador para el análisis del proceso AR

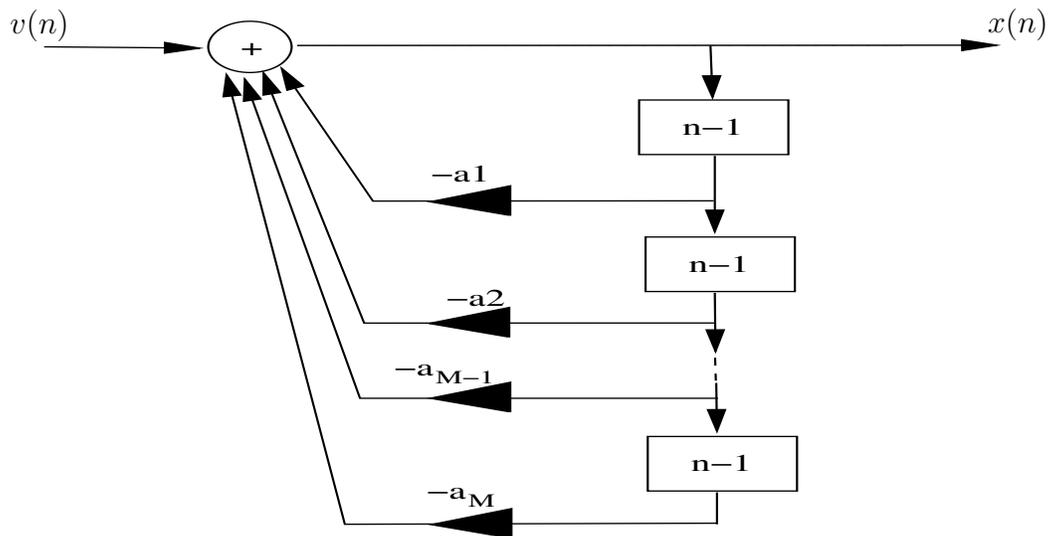


Figura 4.7: Proceso autorregresivo, filtro de síntesis

$v(n)$ es ruido blanco con media cero y varianza σ^2 .

Este proceso es llamado *Autorregresivo de orden M* porque el valor actual $x(n)$ del proceso es una regresión de las muestras pasadas de $x(n)$. Los coeficientes a_i de la combinación lineal son llamados *parámetros autorregresivos* del proceso $x(n)$.

De la ecuación (4.23) se observa que el proceso AR es un sistema del tipo todo polo, y para que sea estable todos los polos deben estar dentro del círculo unitario, por tanto la función $A(z)$ tiene todas sus raíces dentro del círculo unitario.

La salida $x(n)$ se puede aproximar como un proceso estacionario en el sentido amplio (WSS), lo que implica que la entrada $v(n)$ sea ruido blanco. En teoría el filtro predictor de error de orden M puede blanquear (producir una salida de ruido blanco) para cualquier entrada estacionaria representada por la secuencia $x(n), x(n-1), x(n-2), \dots, x(n-M)$ con M suficientemente grande. El incremento del orden del predictor M reduce la correlación entre muestras adyacentes del proceso aplicado a la entrada del filtro.

La relación entre el filtro predictor de error hacia adelante y el predictor mismo se ilustra en las figuras 4.6 y 4.7, donde se nota que la longitud de filtro predictor de error es mayor en uno que el predictor. Sin embargo, ambos filtros son de orden M , entonces ambos involucran la misma cantidad de retardos.

Modelo autorregresivo

Considerando la ecuación en diferencias (4.24), donde se tiene un sistema SLITD sólo polos de orden M , es decir $A(z) = 1/G(z)$, si $A(z)$ es de *fase mínima*, y si $v(n)$ es un proceso de ruido blanco WSS, con media cero, entonces su función de autocorrelación está definida

$$r_{vv}(l) = \sigma_v^2 \delta(l) \quad (4.26)$$

Donde las ecuaciones (4.24) y (4.23) describen un proceso AR de orden M (AR(M)) con entrada $v(n)$ y salida $x(n)$. Multiplicando la ecuación (4.24) por la señal $x(n)$ retrasada en m

$$\begin{aligned} x(n)x(n-m) + a_1x(n-1)x(n-m) + a_2x(n-2)x(n-m) + \\ a_3x(n-3)x(n-m) + \dots + a_Mx(n-p)x(n-m) = v(n)x(n-m) \end{aligned} \quad (4.27)$$

la autocorrelación de $x(n)$ se puede definir como el valor esperado

$$r_{xx}(l) = E\{x(n)x(n-l)\} \quad (4.28)$$

tomando el valor esperado de (4.27), y si $x(n)$ y $v(n)$ no están correlacionadas, se tiene

$$r_{xx}(m) + a_1r_{xx}(m-1) + a_2r_{xx}(m-2) + a_3r_{xx}(m-3) + \dots + a_Mr_{xx}(m-p) = 0 \quad (4.29)$$

en forma más compacta

$$\sum_{i=0}^M a_i r_{xx}(m-i) = 0 \quad (4.30)$$

al desarrollar esta última ecuación para un orden M se obtiene un sistema de ecuaciones adecuado para el cálculo de los coeficientes a_1, a_2, \dots, a_M , para $m = 1, 2, 3, \dots, M$.

En forma matricial se tiene lo que se conoce como ecuaciones de *Yule-Walker* [9], [16], [25]

$$\begin{bmatrix} r_{xx}(0) & r_{xx}(-1) & r_{xx}(-2) & \cdots & r_{xx}(M-1) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(M-2) \\ r_{xx}(2) & r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(M-3) \\ \vdots & \vdots & \vdots & & \vdots \\ r_{xx}(M-1) & r_{xx}(M-2) & r_{xx}(M-3) & \cdots & r_{xx}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_M \end{bmatrix} = - \begin{bmatrix} r_{xx}(1) \\ r_{xx}(2) \\ r_{xx}(3) \\ \vdots \\ r_{xx}(M) \end{bmatrix} \quad (4.31)$$

esto es, el sistema $\mathbf{R}_M \mathbf{a}_M = \mathbf{r}_M$ y si la matriz \mathbf{R}_M es no singular, existe la solución $\mathbf{a}_M = \mathbf{R}_M^{-1} \mathbf{r}_M$, cuya forma es similar a la ecuación normal o de Wiener-Hopf [41], [15], [30].

En el contexto de *predicción lineal*, si $x(n)$ es un valor al tiempo n y $\hat{x}(n)$ el valor estimado

$$\hat{x}(n) = - \sum_{m=1}^M a_m x(n-m) \quad (4.32)$$

$\hat{x}(n)$ es la predicción lineal de $x(n)$ basado en las M muestras anteriores de $x(n)$ y el error de predicción

$$e(n) = x(n) - \hat{x}(n) \quad (4.33)$$

entonces los coeficientes a_m minimizan el error de predicción medio cuadrático (MSE)

$$E\{e^2(n)\} = E\{[x(n) + \hat{x}(n)]^2\} \quad (4.34)$$

Dentro del campo de identificación de sistemas, los coeficientes a_m representan el proceso AR y son los mismos que aparecen en la ecuación de Yule-Walker, en algunos eventos existe una correspondencia uno a uno entre sistemas de identificación AR y predicción lineal [9].

Relación entre coeficientes $h(n)$ y $a(n)$

Para la predicción hacia adelante se han utilizado los coeficientes $h(m)$, donde $p = M$ es el orden del predictor, del análisis de predicción lineal y del modelo AR se pueden relacionar

los coeficientes $h(m)$ y los coeficientes $a_p(i)$. La función de transferencia del filtro predictor de error hacia adelante está dada por

$$H(z) = \sum_{k=0}^p h(k)z^{-k} \quad (4.35)$$

y la función de transferencia de modelo AR visto como un sistema lineal retroalimentado

$$H_{AR}(z) = \frac{1}{1 - H(z)} = \frac{1}{1 - \sum_{k=0}^p h(k)z^{-k}} \quad (4.36)$$

los coeficientes del filtro predictor de error están relacionados con el proceso AR

$$H_{AR}(z) = \frac{1}{\sum_{k=0}^p a_p(k)z^{-k}} = \frac{1}{A_p(z)} \quad (4.37)$$

Esto muestra que la función de transferencia del modelo AR es igual al inverso de la función de transferencia del filtro predictor de error hacia adelante, por lo que al proceso AR se le llama frecuentemente *filtro inverso* [14], y los coeficientes se relacionan

$$a_p(i) = \begin{cases} 1, & m = 0 \\ -h(m), & i = 1, 2, 3, \dots, p \\ 0, & m > p \end{cases} \quad (4.38)$$

Por lo tanto el análisis de predicción lineal de un proceso estacionario en el sentido amplio y el modelado AR de un proceso pueden ser vistos como un par de operaciones complementarias.

- *Análisis*: a través de un filtro predictor de error sobre una señal $x(n)$ WSS se pueden calcular los coeficientes $h(n)$ y por tanto los coeficientes $a_p(n)$, (figura 4.8).
- *Síntesis*: conociendo los coeficientes $a_p(n)$ para un modelo AR y teniendo ruido blanco como una señal de entrada se tiene una señal de salida $x(n)$, (figura 4.7).

Reformulando la ecuación normal e introduciendo el término de la derecha $r_{xx}(i)$ en la sumatoria

$$\sum_{m=0}^p a_p(m)r_{xx}(k - m) = 0 \quad k = 1, 2, 3, 4, \dots, p \quad (4.39)$$

también se puede reescribir la ecuación del error medio cuadrático del predictor hacia adelante introduciendo el valor de $r_{xx}(0)$ en la sumatoria, es decir iniciando el índice en $m = 0$

$$\sum_{m=0}^p a_p(m)r_{xx}(m) = J_p^f(n) \quad (4.40)$$

se observa que ambas sumatorias del lado izquierdo de las ecuaciones (4.39 y 4.40) son iguales, por lo que se pueden escribir en una ecuación de orden p

$$\sum_{m=0}^p a_p(m)r_{xx}(k-m) = \begin{cases} J_p^f(k), & k = 0 \\ 0, & k = 1, 2, 3, \dots, p \end{cases} \quad (4.41)$$

Esta última ecuación es conocida como la *ecuación normal aumentada para predicción lineal hacia adelante*.

En forma matricial

$$\begin{bmatrix} r_{xx}(0) & \mathbf{r}_x^T \\ \mathbf{r}_x & \mathbf{R}_p \end{bmatrix} \begin{bmatrix} 1 \\ -\mathbf{h}(\mathbf{n}) \end{bmatrix} = \begin{bmatrix} J_p^f \\ \mathbf{0} \end{bmatrix} \quad (4.42)$$

También el error de predicción lineal hacia adelante (ecuación 4.40) se puede escribir como

$$e_p^f(n) = \sum_{m=0}^p a_p(m)r_{xx}(n-m) \quad (4.43)$$

Esta ecuación es claramente una relación entrada/salida que lleva a una estructura transversal en función de los parámetros a_i , cuya salida es $e_p^f(n)$ llamada *filtro de predicción de error*.

Para un predictor de orden p , la secuencia de coeficientes del filtro predictor de error $a_p(i)$ y la energía del error de predicción $J_p^f(n)$ están relacionados por la secuencia de autocorrelación $r_{xx}(i)$ al orden $p+1$, es decir, la ecuación normal aumentada. El modelo AR está relacionado a la secuencia de autocorrelación del proceso por la ecuación normal aumentada, esta estructura del modelo AR es referido como la ecuación de Yule-Walker que tiene una estructura matemática igual a la ecuación normal aumentada pero con orígenes físicos diferentes [14].

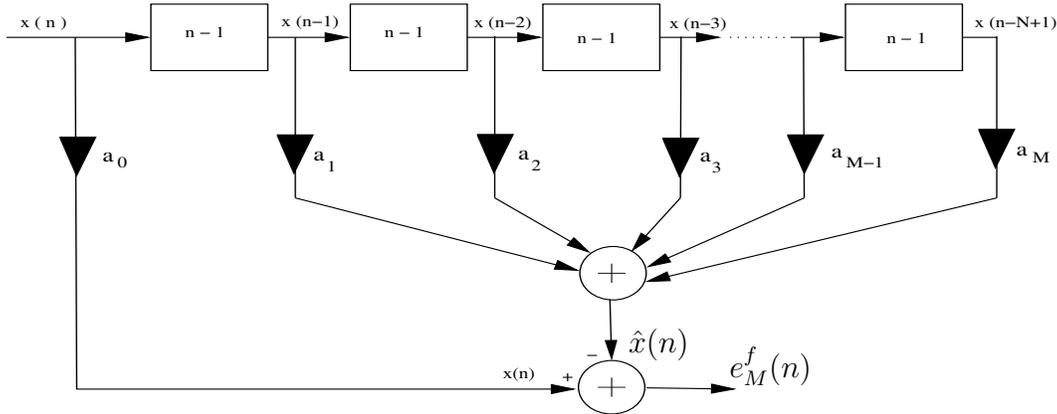
4.3. Predicción lineal hacia atrás

En este caso se pretende estimar la muestra $x(M-1)$ considerando que se conoce la muestra actual $x(n)$ y las $M-1$ muestras anteriores $x(n-1), x(n-2), \dots, x(n-M+1)$, la muestra a estimar $\hat{x}(n-M)$ es una función

$$\hat{x}(n-M) = f(x(n), x(n-1), x(n-2), \dots, x(n-M+1))$$

si $x(n-M)$ es el valor verdadero de la muestra a estimar, entonces se tiene un error de predicción hacia atrás dado por

$$e_M^b(n) = x(n-M) - \hat{x}(n-M) \quad (4.44)$$


 Figura 4.8: Error de predicción hacia adelante con parámetros a_i

donde $x(n - M)$ juega el papel de la *respuesta deseada*, M indica el número de muestras del predictor y el superíndice b hacia atrás.

Utilizando una estructura de líneas de retardo TDL como en la figura 4.9 similar a un filtro FIR, el valor estimado se puede escribir como la convolución de los coeficientes del predictor $b(k)$ y la señal $x(n)$ con sus retardos.

$$\hat{x}(n - M) = \sum_{i=1}^M b(i)x(n - i + 1) \quad (4.45)$$

Para encontrar los coeficientes $b(i)$ en términos de la media de los mínimos cuadrados se puede utilizar la ecuación normal adecuándola al problema de la predicción hacia atrás

$$\sum_{m=1}^M b(m)r_{xx}(m - i) = r_{xx}(M - i + 1) \quad i = 1, 2, 3, 4, \dots, M \quad (4.46)$$

donde $r_x(m - i)$ es la correlación entre muestras en la línea de retardo definida por

$$r_{xx}(m - i) = E\{x(m - i + 1)x(n - m + 1)\} \quad i = 1, 2, 3, 4, \dots, M \quad (4.47)$$

y $r_{xx}(M - i + 1)$ es la correlación entre la *señal deseada* y los retardos de la entrada dados por

$$r_{xx}(M - i + 1) = E\{x(n - i + 1)x(n - M)\} \quad i = 1, 2, 3, 4, \dots, M \quad (4.48)$$

de forma similar el error medio cuadrático mínimo del predictor hacia atrás está dado por

$$J_M^b(n) = E\{|e_M^b(n)|^2\} = r_{xx}(0) - \sum_{m=1}^M b(k)r_{xx}(M - m + 1) \quad (4.49)$$

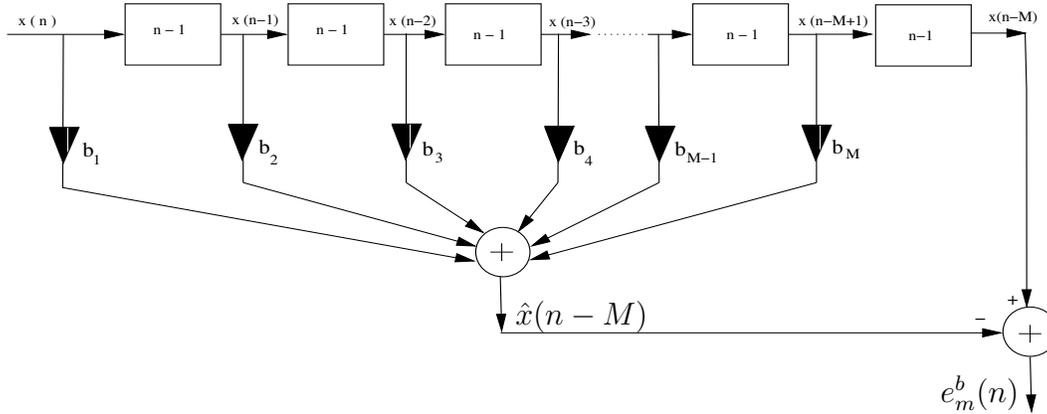


Figura 4.9: Filtro de predicción de error hacia atrás

donde $r_{xx}(0)$ es igual al valor medio cuadrático de la respuesta deseada.

Si en la ecuación (4.46) se reemplazan los índices m por $M - m + 1$, e i por $M - i + 1$ y considerando que para un proceso estacionario $r_{xx}(i - m) = r_{xx}(m - i)$, la ecuación normal para predicción hacia atrás se puede escribir

$$\sum_{m=1}^M b(M - m + 1) r_{xx}(i - m) = r_{xx}(i) \quad i = 1, 2, 3, 4, \dots, M \quad (4.50)$$

comparando la ecuación (4.30) con (4.50), se observa que tienen la misma forma matemática, por lo que los coeficientes de predicción hacia adelante y hacia atrás se pueden relacionar

$$a(m) = b(M - m + 1); \quad b(m) = a(M - m + 1) \quad m = 1, 2, 3, \dots, M \quad (4.51)$$

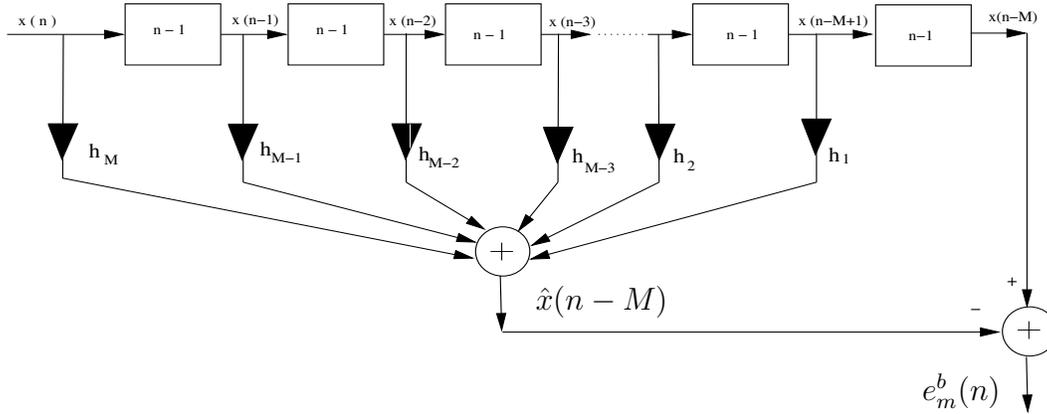
Esto sugiere que en un proceso estacionario se pueden usar los coeficientes del predictor hacia adelante en orden inverso para obtener los coeficientes de predicción hacia atrás. Reemplazando m por $M - m + 1$ en la ecuación (4.49), se observa que

$$J_M^f(n) = J_M^b(n) \quad (4.52)$$

Lo que implica que para un proceso estacionario el error de predicción hacia atrás $e_M^b(n)$ y el error de predicción hacia adelante $e_M^f(n)$ tienen el mismo valor medio cuadrático, entonces $e_M^b(n) = e_M^f(n) = e_M(n)$.

Filtro de predicción de error hacia atrás

Combinando la solución de la ecuación normal para la predicción hacia atrás (4.46) y la ecuación de la media del error cuadrático hacia atrás en función de los coeficientes hacia


 Figura 4.10: Error de predicción hacia atrás con parámetros h_i

adelante

$$\sum_{m=1}^M h(M-m+1)r_{xx}(m-i) = r_{xx}(M-i+1) \quad i = 1, 2, 3, 4, \dots, M \quad (4.53)$$

reemplazando $i-1$ por j y $m-1$ por l

$$\sum_{m=0}^{M-1} h(M-l)r_{xx}(l-j) = r_{xx}(M-j) \quad j = 0, 1, 2, 3, 4, \dots, M-1 \quad (4.54)$$

para mayor claridad sin afectar el significado de esta ecuación, se sustituye l por m y j por i

$$\sum_{m=0}^{M-1} h(M-m)r_{xx}(m-i) = r_{xx}(M-i) \quad i = 0, 1, 2, 3, 4, \dots, M-1 \quad (4.55)$$

introduciendo la relación entre los coeficientes $h(n)$ y $a_M(n)$ del modelo AR, se puede reescribir la ecuación normal

$$\sum_{m=0}^M a_M(M-m)r_{xx}(m-i) = 0 \quad i = 0, 1, 2, 3, 4, \dots, M-1 \quad (4.56)$$

Reformulando la media del error cuadrático hacia atrás en términos de coeficientes $h(n)$ (figura 4.10)

$$J_M^b(n) = r_{xx}(0) - \sum_{m=0}^M h(M-m+1)r_{xx}(M-m+1) \quad (4.57)$$

reemplazando $m - 1$ por m

$$J_M^b(n) = r_{xx}(0) - \sum_{m=0}^{M-1} h(M-m)r_{xx}(M-m) \quad (4.58)$$

definiendo el orden $M = p$, igual al número de polos del predictor y observando que $r_{xx}(0) = r_{xx}(p-m)$ para $m = p$, es decir, que se puede mover a $r_{xx}(0)$ dentro de la sumatoria y a la vez introduciendo los coeficientes $a_p(n)$

$$J_p^b(n) = \sum_{m=0}^p a_p(p-m)r_{xx}(p-m) = \sum_{m=0}^p a_p(p-m)r_{xx}(m-p) \quad (4.59)$$

en un proceso estacionario $r_{xx}(M-m) = r_{xx}(m-M)$.

Para el orden $M = p$, la ecuación normal aumentada para un proceso de predicción lineal hacia atrás se pueden escribir

$$\sum_{m=0}^p a_p(p-m)r_{xx}(m-i) = \begin{cases} 0 & i = 0, 1, 2, \dots, p-1 \\ J_p^b(i) & i = p \end{cases} \quad (4.60)$$

en forma matricial

$$\begin{bmatrix} \mathbf{R}_p & \mathbf{r}_x^b \\ \mathbf{r}_x^{bT} & r_{xx}(0) \end{bmatrix} \begin{bmatrix} -\mathbf{b} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ J_p^b(p) \end{bmatrix} \quad (4.61)$$

El error de predicción se puede escribir como un filtro transversal con coeficientes $h(n)$, $b(n)$ o $a(n)$

$$e_p^b(n) = x(n-p) - \sum_{i=1}^p b_p(i)x(n-i+1) = x(n-p) - \sum_{i=1}^p h(p-i+1)x(n-i+1) \quad (4.62)$$

donde

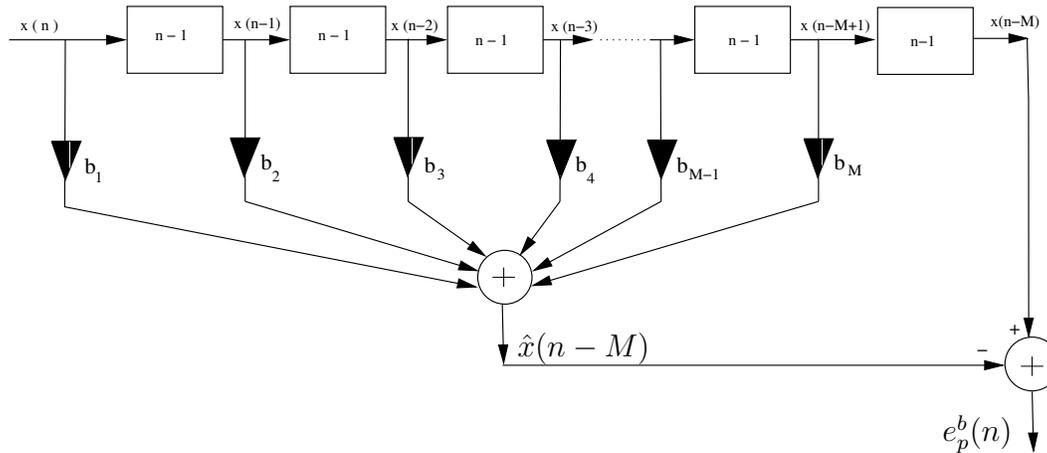
$$a_p(p-i+1) = \begin{cases} 1 & i = p+1 \\ -h(p-i+1), & i = 1, 2, \dots, p \end{cases} \quad (4.63)$$

entonces

$$e_p^b(n) = \sum_{i=1}^{p+1} a_p(p-i+1)x(n-i+1) \quad (4.64)$$

reemplazando $i - 1$ por i , se tiene la expresión para el filtro predictor hacia atrás en términos de los coeficientes hacia adelante

$$e_p^b(n) = \sum_{i=0}^M a_p(M-i)x(n-i) \quad (4.65)$$


 Figura 4.11: Filtro de predicción de error hacia atrás con parámetros b_i

como se ilustra en la figura 4.11.

Como la función de transferencia del predictor hacia adelante de orden p está dada por

$$A_p^f(z) = \sum_{n=0}^p a_p(n) z^{-n} \quad (4.66)$$

para el filtro predictor hacia atrás se tiene una función de transferencia en términos de los coeficientes $a_p(n)$

$$A_p^b(z) = a_p(p) + a_p(p-1)z^{-1} + \dots + a_p(0)z^{-p} = \sum_{n=0}^p a_p(p-n)z^{-n} \quad (4.67)$$

si se cambia el índice $l = p - n$, entonces estas dos funciones de transferencia se pueden relacionar

$$\begin{aligned} A_p^b(z) &= \sum_{l=0}^p a_p(l) z^{-(p-l)} = z^{-p} \sum_{l=0}^p a_p(l) z^l \\ A_p^b(z) &= z^{-p} A_p^f(z^{-1}) \end{aligned} \quad (4.68)$$

Se concluye

- Que la función de transferencia $A_p^b(z)$ no puede tener ceros dentro del círculo unitario, es decir que se tiene un *filtro de fase máxima*.
- En un filtro de fase máxima su respuesta al impulso es retardada lo máximo posible con respecto al instante de tiempo cuando se aplica un impulso al filtro. Esto es que la energía contenida en la respuesta al impulso de un filtro de fase máxima está concentrada cerca de la parte final hacia atrás.

4.4. Algoritmos para solución de la ecuación normal

Como se ha notado en la exposición de este capítulo que la ecuación normal (4.31) aparece en varias ocasiones, y su solución matemática consiste en la inversión de la matriz de autocorrelaciones \mathbf{R}_p que podría realizarse a través de la eliminación gaussiana, sin embargo, la carga computacional se hace excesiva (orden de operaciones $O(p^3)$), lo que implica la búsqueda de soluciones computacionalmente óptimas. Para un proceso AR de orden p (AR(p)), la salida $x(n)$ puede estar completamente definida con los primeros p retardos de la función de autocorrelación $r_{xx}(m)$.

Recordando la ecuación normal de orden p o de Yule-Walker para un modelo autorregresivo:

$$\begin{bmatrix} r_{xx}(0) & r_{xx}(-1) & r_{xx}(-2) & \cdots & r_{xx}(p-1) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(p-2) \\ r_{xx}(2) & r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{xx}(p-1) & r_{xx}(p-2) & r_{xx}(p-3) & \cdots & r_{xx}(0) \end{bmatrix} \begin{bmatrix} a_p(1) \\ a_p(2) \\ a_p(3) \\ \vdots \\ a_p(p) \end{bmatrix} = - \begin{bmatrix} r_{xx}(1) \\ r_{xx}(2) \\ r_{xx}(3) \\ \vdots \\ r_{xx}(p) \end{bmatrix} \quad (4.69)$$

Para la solución de los coeficientes a_i de diferentes órdenes es necesario introducir un índice que indique el orden de la ecuación normal, por lo que se utiliza la nomenclatura $a_m(i)$, donde m representa el orden e i el índice del coeficiente, $i = 1, 2, \dots, m$.

Considerando un proceso AR de **orden** $p = 1$, es decir AR(1), entonces al desarrollar la ecuación de (4.69) se convierte en

$$r_{xx}(0)a_1(1) = -r_{xx}(1) \quad (4.70)$$

y definiendo la constante K_1 de orden 1

$$K_1 = a_1(1) = -\frac{r_{xx}(1)}{r_{xx}(0)} \quad (4.71)$$

donde $r_{xx}(0) > r_{xx}(1)$, entonces $|K_1 = a_1(1)| < 1$
y el error cuadrático para un estimador hacia adelante

$$P_1 = J_1^f(n) = r_{xx}(0) + a_1(1)r_{xx}(-1) = r_{xx}(0)[1 - K_1^2] \quad (4.72)$$

definiendo la energía

$$P_0 = r_{xx}(0) \quad (4.73)$$

entonces

$$P_1 = P_0[1 - K_1^2] \quad (4.74)$$

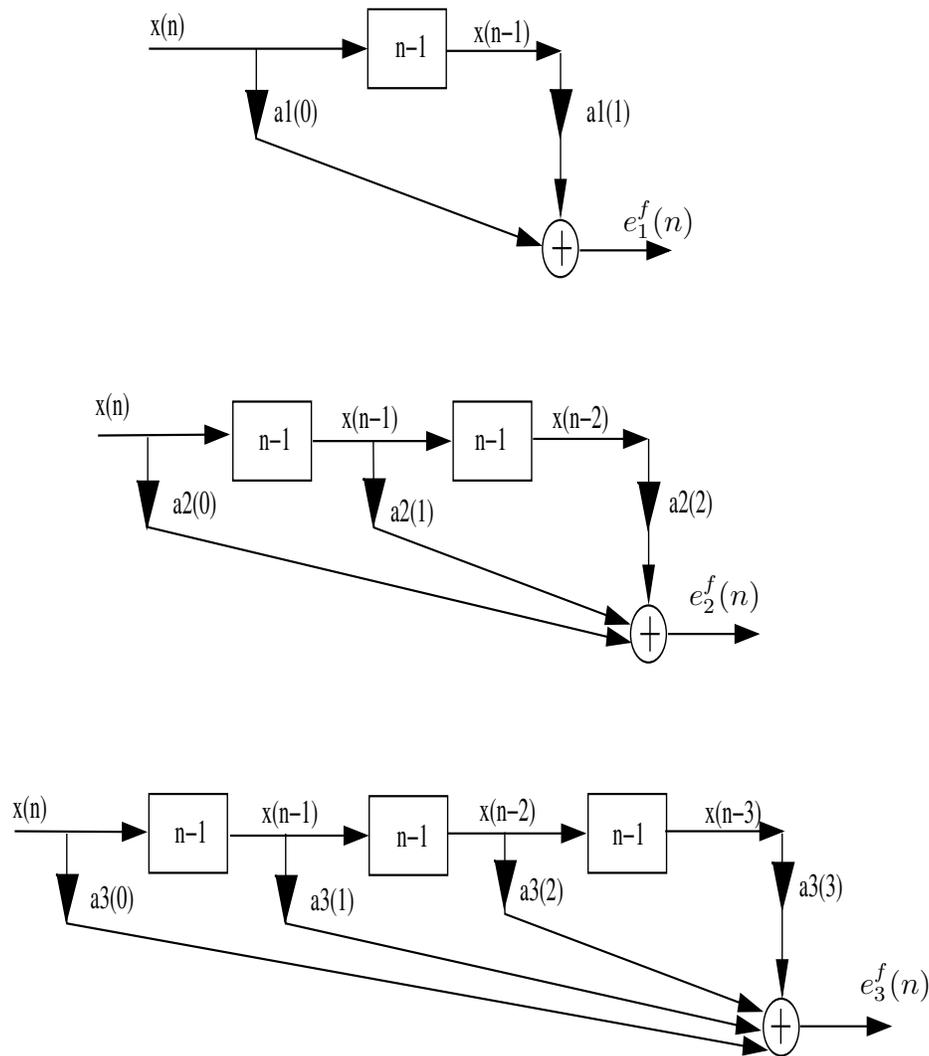


Figura 4.12: Filtros de predicción de error hacia adelante, orden $p = 1, 2, 3$

En la figura 4.12 se observa la implementación de un filtro de predicción de error hacia adelante para varios los órdenes $p = 1, 2, 3$.

Para un modelo de **orden** $p = 2$, AR(2), desarrollando la ecuación (4.69) para $p = 2$ se tiene

$$\begin{aligned} r_{xx}(0)a_2(1) + r_{xx}(1)a_2(2) &= -r_{xx}(1) \\ r_{xx}(1)a_2(1) + r_{xx}(0)a_2(2) &= -r_{xx}(2) \end{aligned} \quad (4.75)$$

y utilizando eliminación gaussiana para resolver el sistema de ecuaciones

$$K_2 = a_2(2) = -\frac{r_{xx}(2) + K_1 r_{xx}(1)}{r_{xx}(0) - K_1 r_{xx}(1)} = -\frac{r_{xx}(2) + K_1 r_{xx}(1)}{r_{xx}(0)(1 - K_1^2)}$$

$$K_2 = a_2(2) = -\frac{r_{xx}(2) + K_1 r_{xx}(1)}{J_1^f(n)} = -\frac{r_{xx}(2) + K_1 r_{xx}(1)}{P_1} \quad (4.76)$$

calculando P_1 en forma recursiva

$$P_1 = P_0(1 - K_1^2) \quad (4.77)$$

por tanto

$$a_2(1) = K_1 + K_1 a_2(2) = a_1(1) + K_2 a_1(1) \quad (4.78)$$

también se puede comprobar que $K_2 = |a_2(2)| < 1$.

Para un **orden** $p = 3$, AR(3), se tiene el sistema de ecuaciones

$$\begin{aligned} r_{xx}(0)a_3(1) + r_{xx}(1)a_3(2) + r_{xx}(2)a_3(3) &= -r_{xx}(1) \\ r_{xx}(1)a_3(1) + r_{xx}(0)a_3(2) + r_{xx}(1)a_3(3) &= -r_{xx}(2) \\ r_{xx}(2)a_3(1) + r_{xx}(1)a_3(2) + r_{xx}(0)a_3(3) &= -r_{xx}(3) \end{aligned} \quad (4.79)$$

con solución para los coeficientes $a_3(m)$

$$K_3 = a_3(3) = -\frac{r_{xx}(3) + a_2(1)r_{xx}(2) + a_2(2)r_{xx}(1)}{r_{xx}(0)(1 - K_1^2)(1 - K_2^2)}$$

$$K_3 = -\frac{r_{xx}(3) + a_2(1)r_{xx}(2) + a_2(2)r_{xx}(1)}{P_1(1 - k_2^2)} \quad (4.80)$$

donde

$$P_2 = P_1(1 - K_2^2) \quad (4.81)$$

$$a_3(2) = a_2(2) + a_2(1)a_3(3) = a_2(2) + K_3 a_2(1) \quad (4.82)$$

$$a_3(1) = a_2(1) + a_2(2)a_3(3) = a_2(1) + K_3 a_2(2) \quad (4.83)$$

y

$$K_1 = a_1(1) = -\frac{r_{xx}(1)}{r_{xx}(0)} \quad (4.84)$$

$$K_2 = a_2(2) = -\frac{r_{xx}(2) + a_1(1)r_{xx}(1)}{r_{xx}(0)(1 - K_1^2)} \quad (4.85)$$

$$a_2(1) = a_1(1) + K_2 a_1(1) \quad (4.86)$$

Del desarrollo anterior se puede observar que para encontrar los coeficientes $a_p(i)$ al orden p , existe una relación con los coeficiente $a_{p-1}(i)$ al orden $p - 1$. Los coeficientes K_i también se pueden escribir en función de los coeficientes al orden anterior K_{i-1} y $a_{i-1}(j)$.

Por lo tanto, las fórmulas generales se pueden escribir para el cálculo de estos coeficientes para órdenes m superiores. La energía del error cuadrático satisface que

$$P_0 \geq P_1 \geq \dots \geq E_p$$

variable	ecuación
K_m	$K_m = a_m(m) = -\frac{r_{xx}(m) + \sum_{n=1}^{m-1} a_{m-1}(n)r_{xx}(m-n)}{P_{m-1}}$
$a_m(n)$	$a_m(n) = a_{m-1}(n) + [a_{m-1}(m-n)]K_m$ $n = 1, 2, 3, \dots, m-1$
P_m	$J_m^f = P_m = P_{m-1}[1 - K_m^2]$

Tabla 4.1: Solución de ecuación normal en forma recursiva

Este algoritmo es conocido de Levinson-Durbin y será explicado con mayor detalle a continuación.

4.4.1. Ecuación normal para predicción hacia adelante y hacia atrás

Tomando en consideración las ecuaciones deducidas para predicción hacia adelante como hacia atrás, de la ecuación normal aumentada para predicción hacia adelante al orden p

$$\sum_{m=0}^p a_p(m)r_x(k-m) = \begin{cases} J_p^f(k), & k = 0 \\ 0, & k = 1, 2, 3, \dots, p \end{cases} \quad (4.87)$$

se puede desarrollar para un orden $p - 1$ y si la energía mínima del error de predicción hacia adelante $J_p^f(k) = E_{p-1}$,

$$\begin{bmatrix} r_{xx}(0) & r_{xx}(-1) & r_{xx}(-2) & \cdots & r_{xx}(1-p) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(2-p) \\ r_{xx}(2) & r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(3-p) \\ \vdots & \vdots & \vdots & & \vdots \\ r_{xx}(p-1) & r_{xx}(p-2) & r_{xx}(p-3) & \cdots & r_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_{p-1}(1) \\ a_{p-1}(2) \\ \vdots \\ a_{p-1}(p-1) \end{bmatrix} = - \begin{bmatrix} E_{p-1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.88)$$

De la ecuación normal aumentada para un proceso de predicción lineal hacia atrás al orden p se obtuvo

$$\sum_{m=0}^p a_p(p-m)r_x(m-k) = \begin{cases} 0, & k = 0, 1, 2, \dots, p-1 \\ J_p^b(k), & k = p \end{cases} \quad (4.89)$$

se puede desarrollar al orden $p-1$ y como se está trabajando sobre una misma ventana de tiempo, entonces la energía mínima del error de predicción hacia atrás se puede escribir $J_p^b(k) = J_p^f(k) = E_{p-1}$ [16]

$$\begin{bmatrix} r_{xx}(0) & r_{xx}(-1) & r_{xx}(-2) & \cdots & r_{xx}(1-p) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(2-p) \\ r_{xx}(2) & r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(3-p) \\ \vdots & \vdots & \vdots & & \vdots \\ r_{xx}(p-1) & r_{xx}(p-2) & r_{xx}(p-3) & \cdots & r_{xx}(0) \end{bmatrix} \begin{bmatrix} a_{p-1}(p-1) \\ a_{p-1}(p-2) \\ a_{p-1}(p-3) \\ \vdots \\ 1 \end{bmatrix} = - \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ E_{p-1} \end{bmatrix} \quad (4.90)$$

Efectuando la suma de la ecuación (4.88) con (4.90) multiplicada por una constante K_m , aumentando el sistema al orden p y factorizando la matriz de autocorrelaciones

$$\begin{bmatrix} r_{xx}(0) & r_{xx}(-1) & r_{xx}(-2) & \cdots & r_{xx}(1-p) & r_{xx}(-p) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(2-p) & r_{xx}(1-p) \\ r_{xx}(2) & r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(3-p) & r_{xx}(2-p) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ r_{xx}(p-1) & r_{xx}(p-2) & r_{xx}(p-3) & \cdots & r_{xx}(0) & r_{xx}(-1) \\ r_{xx}(p) & r_{xx}(p-1) & r_{xx}(p-2) & \cdots & r_{xx}(1) & r_{xx}(0) \end{bmatrix} \cdot \left\{ \begin{bmatrix} 1 \\ a_{p-1}(1) \\ a_{p-1}(2) \\ \vdots \\ a_{p-1}(p-1) \end{bmatrix} + K_m \begin{bmatrix} a_{p-1}(p-1) \\ a_{p-1}(p-2) \\ a_{p-1}(p-3) \\ \vdots \\ 1 \end{bmatrix} \right\} = \left\{ \begin{bmatrix} E_{p-1} \\ 0 \\ 0 \\ \vdots \\ \Delta_m \end{bmatrix} + K_m \begin{bmatrix} \Delta_m \\ 0 \\ 0 \\ \vdots \\ E_{p-1} \end{bmatrix} \right\} \quad (4.91)$$

la constante Δ_m sirve para ajustar los vectores de energía cuando se incrementa el orden de las ecuaciones. Igualando la ecuación (4.91) con la ecuación del filtro de predicción de error hacia adelante desarrollada al orden p

$$\begin{bmatrix} r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(1-p) & r_{xx}(-p) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(2-p) & r_{xx}(1-p) \\ r_{xx}(2) & r_{xx}(1) & \cdots & r_{xx}(3-p) & r_{xx}(2-p) \\ \vdots & \vdots & & \vdots & \vdots \\ r_{xx}(p-1) & r_{xx}(p-2) & \cdots & r_{xx}(0) & r_{xx}(1) \\ r_{xx}(p) & r_{xx}(p-1) & \cdots & r_{xx}(1) & r_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_p(1) \\ a_p(2) \\ \vdots \\ a_p(p-1) \\ a_p(p) \end{bmatrix} = \begin{bmatrix} E_p \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad (4.92)$$

igualando el vector de coeficientes $a_m(p)$ de la ecuación (4.92) con el vector equivalente de la ecuación (4.91)

$$\begin{bmatrix} 1 \\ a_p(1) \\ a_p(2) \\ \vdots \\ a_p(p-1) \\ a_p(p) \end{bmatrix} = \left\{ \begin{bmatrix} 1 \\ a_{p-1}(1) \\ a_{p-1}(2) \\ \vdots \\ a_{p-1}(p-1) \end{bmatrix} + K_m \begin{bmatrix} a_{p-1}(p-1) \\ a_{p-1}(p-2) \\ a_{p-1}(p-3) \\ \vdots \\ 1 \end{bmatrix} \right\}$$

en forma escalar las componentes de los vectores se reduce a

$$a_p(m) = a_{p-1}(m) + K_m a_{p-1}(p-m) \quad (4.93)$$

es decir, que los coeficientes $a_p(m)$ al orden p se pueden calcular de los coeficientes $a_{p-1}(m)$ al orden $p-1$.

De igual forma, igualando los vectores del lado derecho de las ecuaciones (4.92) y (4.91)

$$\begin{bmatrix} E_p \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} = \left\{ \begin{bmatrix} E_{p-1} \\ 0 \\ 0 \\ \vdots \\ \Delta_m \end{bmatrix} + K_m \begin{bmatrix} \Delta_m \\ 0 \\ 0 \\ \vdots \\ E_{p-1} \end{bmatrix} \right\}$$

de la primera línea del sistema de ecuaciones se tiene

$$E_p = E_{p-1} + K_m \Delta_m \quad (4.94)$$

de la última línea

$$\Delta_m + K_m E_{p-1} = 0 \implies \Delta_m = -K_m E_{p-1} \quad (4.95)$$

entonces

$$E_p = E_{p-1}(1 - |K_m|^2) \quad (4.96)$$

También se ha obtenido un cálculo de la energía del error E_p al orden p con base a la energía del error al orden anterior $p-1$. La energía del error E_p decrece o permanece igual conforme

el orden del filtro de predicción de error se incrementa, es decir, $E_p \leq E_{p-1}$. Entonces de la ecuación (4.96) se deduce que los coeficientes de reflexión K_m deben cumplir con la condición $|K_m| \leq 1$, esto garantiza la estabilidad del diseño del filtro predictor.

Resumen de algoritmo de Levinson-Durbin

En esta sección se ha llegado a una solución recursiva de la ecuación normal, cuyos autores de este algoritmo son Levinson y Durbin [1], [15], [16], [29]. Este algoritmo utiliza $a_1(1)$ para generar $a_2(1)$ y $a_2(2)$, con $a_2(1)$ y $a_2(2)$ genera $a_3(1)$, $a_3(2)$ y $a_3(3)$ y así sucesivamente hasta el orden p , este proceso de recursión se muestra en la tabla 4.2 y requiere haber calculado la secuencia de autocorrelación $r_{xx}(0), r_{xx}(1), r_{xx}(0), \dots, r_{xx}(p)$.

	<p>Condiciones iniciales para el orden $m = 0$</p> $a_0(0) = 1$ $a_0(1) = 0$ $P_0 = r_{xx}(0)$
Pasos	Cálculos
1	$m = 1$ $K_1 = a_1(1) = -\frac{r_{xx}(1)}{r_{xx}(0)}$
2	$m = 2, 3, \dots, p$ $K_m = a_m(m) = -\frac{r_{xx}(m) + \sum_{n=1}^{m-1} a_{m-1}(n)r_{xx}(m-n)}{P_{m-1}}$ $a_m(n) = a_{m-1}(n) + [a_{m-1}(m-n)]K_m$ $n = 1, 2, 3, \dots, m-1$ $P_m = P_{m-1}[1 - K_m^2]$

Tabla 4.2: Algoritmo de Levinson-Durbin

Interpretación física

La ecuación $P_m = P_{m-1}[1 - K_m^2]$ tiene la misma forma matemática de la ecuación de transmisión de potencia a través la terminación de una red de dos puertos. Debido a esta analogía, a los coeficientes K_i se les conoce como *coeficientes de reflexión* o PARCOR (PARTIAL CORRELATION) [29],[22],[23].

En la práctica, para estimar la autocorrelación se puede utilizar un estimador no sesgado

$$\hat{r}_{xx}(m) = \frac{1}{L_x} \sum_{n=1}^{L_x-|m|} x(n)x(n-m) \quad |m| \ll L_x, \quad m = 0, 1, 2, \dots, L_x \quad (4.97)$$

donde L_x es la longitud de $x(n)$, $p \ll L_x$, y para el cálculo de algoritmo de Levinson-Durbin se puede reemplazar $r_{xx}(m)$ por el estimador $\hat{r}_{xx}(m)$, esto asegura que la varianza de $\hat{r}_{xx}(m)$ es pequeña para grandes retardos.

Ejemplo:

Para una señal de entrada $x(n)$ con una secuencia de autocorrelación normalizada de longitud $p + 1 = 5$

$$r_{xx}(l) = \{r(0), r(1), r(2), r(3), r(4), r(5)\} \quad \text{con} \quad r(0) = 1$$

Desarrollar la recursión del algoritmo de Levinson-Durbin para $p = 4$

Solución:

Condiciones iniciales para el orden $m = 0$

$$K_0 = a_0(0) = 1, \quad a_0(1) = 0, \quad P_0 = r_0 = 1$$

para el orden $m = 1$

$$a_1(0) = 1$$

$$a_1(1) = K_0 = -r_1/r_0 = 1$$

$$P_1 = P_0(1 - K_1^2) = (1 - r_1^2)$$

para el orden $m = 2$

$$a_2(2) = K_2 = \frac{-\sum_{j=0}^1 a_1(j)r(2-j)}{P_1(1 - K_1^2)}$$

$$a_2(1) = a_1(1) + K_2 a_1(1)$$

$$a_2(0) = 1$$

$$P_2 = P_1(1 - K_2^2)$$

para el orden $m = 3$

$$a_3(3) = K_3 = \frac{-\sum_{j=0}^2 a_2(j)r(3-j)}{P_2(1 - K_2^2)}$$

$$\begin{aligned} a_3(2) &= a_2(2) + K_3 a_2(1) \\ a_3(1) &= a_2(1) + K_3 a_2(2) \\ a_3(0) &= 1 \\ P_3 &= P_2(1 - K_3^2) \end{aligned}$$

para el orden $m = 4$

$$a_4(4) = K_4 = \frac{-\sum_{j=0}^3 a_3(j)r(4-j)}{P_3(1 - K_3^2)}$$

$$\begin{aligned} a_4(3) &= a_3(3) + K_4 a_3(1) \\ a_4(2) &= a_3(2) + K_4 a_3(2) \\ a_4(1) &= a_3(1) + K_4 a_3(3) \\ a_4(0) &= 1 \\ P_3 &= P_2(1 - K_3^2) \end{aligned}$$

Con el conjunto de coeficientes $a_4(i) = \{a_4(0), a_4(1), a_4(2), a_4(3), a_4(4)\}$ se puede realizar un filtro IIR todo polo (AR) tal que para una entrada de un proceso estacionario reproduzca la estimación de la señal original $x(n)$.

4.4.2. Relación entre la secuencia de autocorrelación y los coeficientes de reflexión

En la sección anterior se obtuvieron los coeficientes K_i y $a_m(n)$ conociendo la secuencia de autocorrelación de un proceso estacionario. Existen varias formas de calcular los coeficientes conociendo algún conjunto de parámetros o viceversa, como se muestra en la figura 4.13.

Conociendo la energía promedio de la señal $P_0 = r_x(0)$, y los coeficientes

$$K_i = K_1, K_2, K_3, \dots, K_p$$

se puede obtener otra forma de representar los estadísticos de segundo orden de una señal estacionaria, la cual es equivalente a especificar la secuencia de autocorrelación. Algunas veces a este algoritmo se le conoce como *inverso de Levinson* [16].

De la ecuación

$$K_m = a_m(m) = -\frac{r_{xx}(m) + \sum_{n=1}^{m-1} a_{m-1}(n)r_{xx}(m-n)}{P_{m-1}} \quad (4.98)$$

despejando el término $r_{xx}(m)$

$$r_{xx}(m) = -K_m P_{m-1} - \sum_{n=1}^{m-1} a_{m-1}(n)r_{xx}(m-n) \quad m = 1, 2, \dots, p \quad (4.99)$$

esta ecuación se puede plantear como un filtro inverso para el cálculo de los coeficientes de autocorrelación en forma recursiva conociendo los coeficientes $a_m(n)$ y K_i , partiendo de la condición inicial $P_0 = r_{xx}(0)$

En la figura 4.13 se observa la relación entre los métodos para el cálculo de los coeficientes para un proceso estacionario.

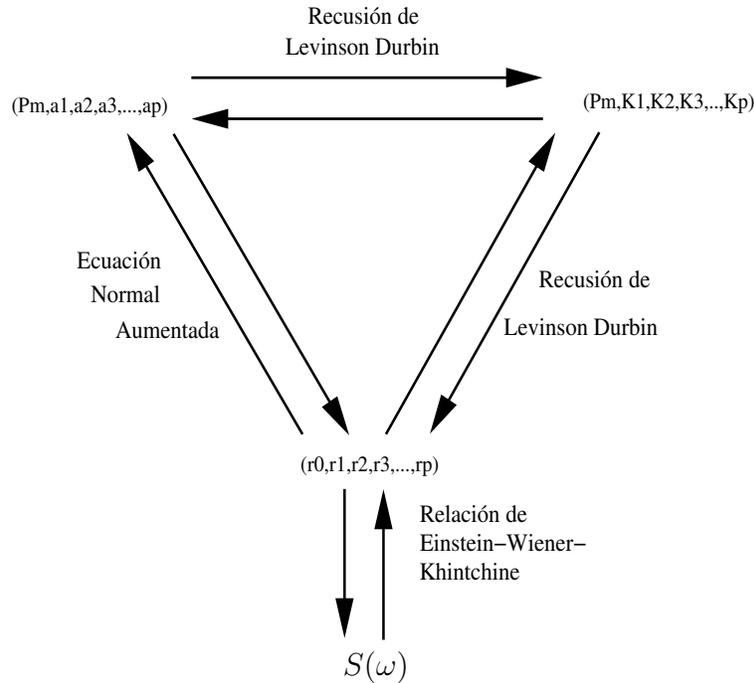


Figura 4.13: Relaciones de coeficientes por Levinson-Durbin

4.4.3. Recursión de “step-down” o Levinson inverso

Este procedimiento permite calcular los coeficientes $a_{m-1}(i)$ y K_{m-1} conociendo los coeficientes $a_m(i)$ y K_m al orden superior m , por esta razón a este tipo de recursión se le conoce como “*step down*”. Considerando la relación que existen entre los coeficientes $a_p(i)$ para un filtro FIR y K_i para un filtro “lattice”¹.

Para el filtro FIR

$$H(z) = \sum_{i=0}^{M-1} h_i z^{-i} = h_0 \left(1 + \sum_{i=1}^{M-1} \frac{h_i}{h_0} z^{-i} \right) \quad (4.100)$$

¹Estos filtros se analizarán con profundidad en el capítulo siguiente

y para el polinomio $A_{p-1}(z)$ de orden $p - 1$

$$A_{p-1}(z) = \left(1 + \sum_{i=1}^{p-1} a_{p-1}(i)z^{-i}\right) \quad (4.101)$$

$$a_{p-1}(i) = \frac{h_i}{h_0} \quad i = 1, 2, 3, \dots, p - 1$$

Para el filtro lattice

$$A_m(z) = A_{m-1}(z) + K_m z^{-1} B_{m-1}(z) \quad (4.102)$$

$$B_m(z) = K_m A_{m-1}(z) + z^{-1} B_{m-1}(z) \quad (4.103)$$

$$m = 1, 2, 3, 4, \dots, p - 1$$

Entonces

$$A_m(z) = A_{m-1}(z) + K_m [B_m(z) - K_m A_{m-1}(z)] \quad (4.104)$$

resolviendo para A_{m-1}

$$A_{m-1}(z) = \frac{A_m(z) - K_m B_m(z)}{1 - K_m^2} \quad (4.105)$$

donde

$$m = p - 1, p - 2, p - 3, \dots, 1$$

Aquí se calculan todos los polinomios de grado menor al polinomio $A_m(z)$ empezando con $A_{m-1}(z)$ y obteniendo los coeficientes “lattice” deseados de la relación $K_m = a_m(m)$ donde $|K_m| \neq 1$. La fórmula recursiva se obtiene calculando K_m empezando por $m = p - 1$ hacia $m = 1$, es decir

$$K_0 = h_0$$

$$K_m = a_m(m) \quad y \quad a_{m-1}(0) = 1$$

aplicando transformada Z inversa a ecuación (4.105)

$$a_{m-1}(n) = \frac{a_m(n) - K_m b_m(n)}{1 - K_m^2} = \frac{a_m(n) - a_m(m)a_m(m-n)}{1 - a_m^2(m)} \quad (4.106)$$

Una aplicación interesante del algoritmo “step-down” es la *prueba de estabilidad de los filtros digitales* de Schur-Cohn, ésta establece que las raíces del polinomio del filtro de predicción hacia adelante $A(z)$ deberán estar dentro del círculo unitario sí y solo sí la magnitud de los coeficientes de reflexión K_m son menores que uno [32], [16], [14]. Entonces dado un filtro lineal, causal e invariante en el tiempo con función de transferencia

$$H(z) = \frac{B(z)}{A(z)} \quad (4.107)$$

para probar su estabilidad, se le aplica el algoritmo *step down* a los coeficientes del denominador $A(z)$ y se verifica que los coeficientes K_m generados sean de magnitud menor que uno.

Condiciones iniciales $m = 0$	Fijar $K_p = a_p(p)$ $a_0(0) = 1$
para $m = p - 1, p - 2, \dots, 1$ para $n = 1, 2, \dots, m$	$a_{m-1}(n) = \frac{a_m(n) - K_m b_m(n)}{1 - K_m^2}$ $a_{m-1}(n) = \frac{a_m(n) - a_m(m) a_m(m-n)}{1 - a_m^2(m)}$ fijar $K_m = a_m(m)$ si $ K_m = 1$, salir

Tabla 4.3: Recursión step down

4.5. Resumen

En este capítulo se ha analizado la codificación por predicción lineal utilizando filtros FIR, se estudiaron los casos de predicción hacia adelante y hacia atrás, después de haber hecho una estimación sobre el error cuadrático se obtuvo la ecuación normal y se explicó el algoritmo de Levinson-Durbin que es una forma óptima de solución. Por otro lado se estudió la relación entre coeficientes del modelo de predicción y la secuencia de autocorrelación de un señal WSS. Cabe mencionar la amplia utilidad de los conceptos sobre filtros. En este capítulo también se mencionaron los filtros lattice que tienen características muy relevantes y se abordarán en el siguiente capítulo.

Capítulo 5

Filtros lattice

En el capítulo de predicción lineal se utilizó la predicción hacia adelante y hacia atrás, llegando a un sistema de ecuaciones conocido como la ecuación normal, y para la solución de esta ecuación se analizaron varios métodos donde se obtuvieron los coeficientes de reflexión K_m con los que se puede construir una nueva estructura de filtro conocida como *filtros "lattice"* que nos ofrece otras características de interés. Las estructuras de filtros "lattice" son utilizados extensamente en procesamiento digital de voz, en geología y en la implementación de filtros adaptables.

5.1. Características de los filtros lattice

- Una estructura lattice combina los filtros de predicción de error hacia adelante y hacia atrás.
- Utilizan únicamente los coeficientes de reflexión K_i y su forma tiene la apariencia "lattice" (celosía o enrejado).
- Proveen otra forma para describir las estadísticas de segundo orden de una señal estacionaria a través de los coeficientes de reflexión.
- Un predictor lattice puede ser visto como el proceso de análisis de un proceso autorregresivo AR [14].
- Sus etapas son desacopladas entre sí, es decir, que al agregar otra etapa al diseño, los coeficientes anteriores no cambian. Por el contrario, en un filtro FIR al aumentar el orden N hay que recalcular el los coeficiente para otro orden.
- Un filtro lattice es modular, esto es que los estados individuales tienen una estructura igual.

- El error de predicción hacia atrás producido en las etapas lattice son ortogonales entre sí para una entrada estacionaria en el sentido amplio.
- Los coeficientes de reflexión K_i tienen la propiedad $|K_i| < 1$, por lo que su representación numérica en aritmética de punto fijo tienen mejor precisión que una estructura transversal TDL, evitando errores de redondeo y la variación de parámetros.

5.2. Deducción de la estructura lattice

Considerando la función de transferencia de un filtro de predicción hacia adelante de orden m [16], vista en el capítulo anterior

$$F_m(z) = \sum_{n=0}^m a_m(n)z^{-n} \quad (5.1)$$

y para un filtro de predicción hacia atrás de orden m

$$G_m(z) = \sum_{n=0}^m b_m(n)z^{-n} = \sum_{n=0}^m a_m(m-n)z^{-n} \quad (5.2)$$

donde los coeficientes $a_m(n)$ y $b_m(n)$ se pueden reescribir utilizando la solución recursiva de Levinson-Durbin

$$a_m(n) = a_{m-1}(n) + K_m a_{m-1}(m-n) = a_{m-1}(n) + K_m b_{m-1}(n)$$

$$b_m(n) = b_{m-1}(n) + K_m a_{m-1}(n) = a_{m-1}(m-n) + K_m a_{m-1}(n) = a_m(m-n)$$

$$n = 0, 1, 2, 3, \dots, m$$

sustituyendo estos coeficientes en las funciones de transferencia $F_m(z)$

$$F_m(z) = \sum_{n=0}^m \left(a_{m-1}(n) + K_m a_{m-1}(m-n) \right) z^{-n}$$

$$F_m(z) = \sum_{n=0}^m a_{m-1}(n)z^{-n} + \sum_{n=0}^m K_m a_{m-1}(m-n)z^{-n} \quad (5.3)$$

en la sumatoria el coeficiente $a_{m-1}(m) = 0$, entonces se puede reducir el orden de a $m-1$, modificando las sumatorias

$$F_m(z) = \sum_{n=0}^{m-1} a_{m-1}(n)z^{-n} + K_m \sum_{n=0}^{m-1} a_{m-n-1}(m-n)z^{-n}$$

el término a_{m-n-1} temporal implica un retardo en tiempo, por lo que introduce un factor z^{-1} en la segunda sumatoria, obteniendo

$$F_m(z) = \sum_{n=0}^{m-1} a_{m-1}(n)z^{-n} + K_m z^{-1} \sum_{n=0}^{m-1} a_{m-1}(m-n)z^{-n} \quad (5.4)$$

La primera sumatoria corresponde a la función de transferencia de un filtro de predicción hacia adelante de orden $m-1$, es decir $F_{m-1}(z)$ y la segunda sumatoria corresponde a la función de transferencia de un filtro de predicción hacia atrás de orden $m-1$, $G_{m-1}(z)$, por lo tanto tenemos

$$F_m(z) = F_{m-1}(z) + K_m z^{-1} G_{m-1}(z) \quad (5.5)$$

De manera similar para un filtro de predicción hacia atrás de orden m

$$G_m(z) = \sum_{n=0}^m b_m(n)z^{-n} = \sum_{k=0}^m a_m(m-n)z^{-n} \quad (5.6)$$

$$G_m(z) = \sum_{n=0}^m \left(a_{m-1}(m-n) + K_m a_{m-1}(n) \right) z^{-n} \quad (5.7)$$

para un orden $m-1$

$$B_m(z) = z^{-1} B_{m-1}(z) + K_m F_{m-1}(z) \quad (5.8)$$

Para las funciones en el tiempo se aplica la transformada Z inversa, obteniendo el par de ecuaciones para los filtros de predicción de error hacia adelante y hacia atrás [16], [39], [31]

$$f_m(n) = f_{m-1}(n) + K_m g_{m-1}(n-1) \quad (5.9)$$

$$g_m(n) = g_{m-1}(n-1) + K_m f_{m-1}(n) \quad (5.10)$$

por lo que se tiene una estructura que incluye la predicción hacia adelante y hacia atrás como se observa en la figura 5.1

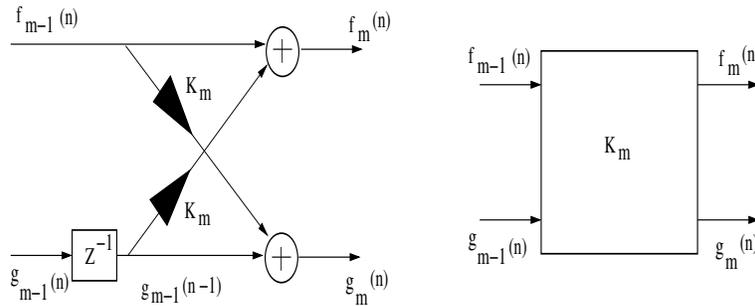


Figura 5.1: Filtro lattice de orden m

5.3. Comparación de estructuras lattice con estructuras FIR

Otra forma más ilustrativa de obtener los coeficientes de reflexión K_m de la estructura lattice es a partir de filtro FIR y comparar coeficientes. De la teoría de filtros FIR, la salida $y(n)$ de un sistema SLITD puede escribirse como una ecuación en diferencias, que es la convolución entre la entrada $x(n)$ del filtro y su respuesta finita al impulso $h(n)$ que se muestra en la figura 5.2

$$y(n) = \sum_{i=1}^m h(i)x(n-i) \quad (5.11)$$

si se hace un cambio de notación para los coeficientes $h(n) = a_m(n)$, donde m es el orden del filtro

$$y(n) = \sum_{i=0}^m a_m(i)x(n-i) = x(n) + \sum_{i=1}^m a_m(i)x(n-i) \quad (5.12)$$

entonces para un filtro FIR, se tiene su respuesta en frecuencia

$$H(z) = 1 + \sum_{n=1}^m a_m(n)z^{-n} \quad m \geq 1 \quad (5.13)$$

Donde

- $h_m(n)$ es la respuesta al impulso de orden m , $h_m(0) = a_m(0) = 1$ y $h_m(n) = a_m(n)$, donde $n = 1, 2, 3, \dots, m$
- $x(n)$ es la secuencia de entrada al filtro
- $y(n)$ es la salida del filtro

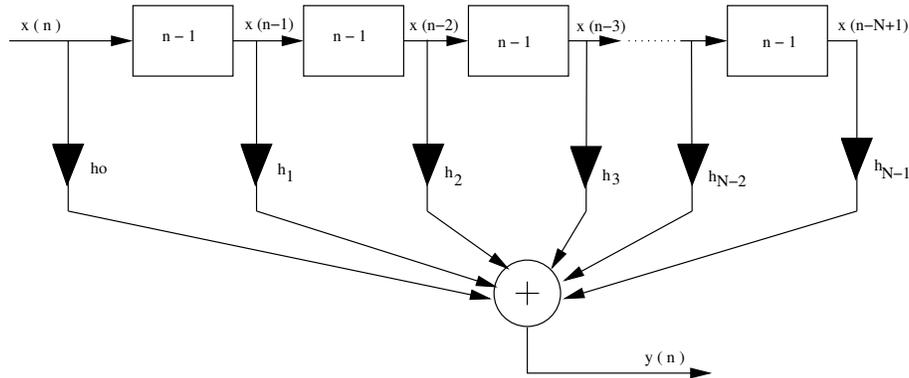


Figura 5.2: Filtro FIR de orden N

5.3.1. Filtro lattice para diferentes órdenes

Un filtro con estructura lattice (figura 5.3) puede desarrollarse a partir de un filtro FIR (figura 5.2). Si se tiene un filtro lattice de primer orden o estado simple.

5.3.2. Filtro lattice de primer orden

Para un filtro FIR de orden uno, si $m = 1$ y si $x(n) = f_0(n) = g_0(n)$ se obtiene

$$y(n) = x(n) + a_1(1)x(n - 1)$$

Para una estructura lattice de primer orden, como se observa en la figura 5.3, se tienen las salidas

$$f_1(n) = f_0(n) + K_1 g_0(n - 1) = x(n) + K_1 x(n - 1) \quad (5.14)$$

$$g_1(n) = K_1 f_0(n) + g_0(n - 1) = K_1 x(n) + x(n - 1) \quad (5.15)$$

En la parte superior de la figura $f_1(n) = y(n)$ y $K_1 = a_1(1)$.

5.3.3. Filtro lattice de segundo orden

Para un filtro FIR de segundo orden, $m = 2$, se tiene

$$y(n) = x(n) + a_2(1)x(n - 1) + a_2(2)x(n - 2) \quad (5.16)$$

Para una estructura lattice de segundo orden, como se observa en la figura 5.4, se tienen las salidas

$$f_1(n) = x(n) + K_1 x(n - 1) \quad f_2(n) = f_1(n) + K_2 g_1(n - 1) \quad (5.17)$$

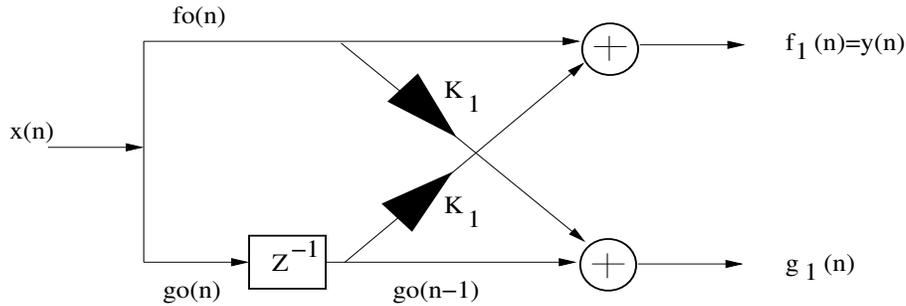


Figura 5.3: Filtro lattice de primer orden

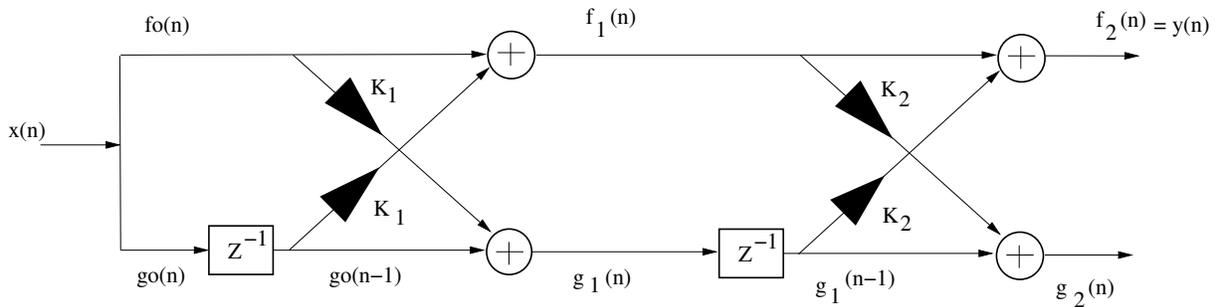


Figura 5.4: Filtro lattice de segundo orden

$$g_1(n) = K_1 x(n) + x(n-1) \quad g_2(n) = K_2 f_1(n) + g_1(n-1) \quad (5.18)$$

sustituyendo $f_1(n)$ y $g_1(n)$ en $f_2(n)$, se obtiene

$$f_2(n) = x(n) + K_1(1 + K_2)x(n-1) + K_2x(n-2) = y(n) \quad (5.19)$$

comparando coeficientes para la salida de un filtro FIR de segundo orden, se tiene que

$$a_2(1) = K_1(1 + K_2)$$

$$K_2 = a_2(2) \quad (5.20)$$

$$K_1 = \frac{a_2(1)}{1 + K_2} = \frac{a_2(1)}{1 + a_2(2)} \quad (5.21)$$

5.3.4. Filtro lattice de orden m

En general siguiendo procesos similares, se puede extender a un filtro lattice de orden m , esto conduce a un filtro lattice que se puede desarrollar escribiendo las ecuaciones recursivas en el orden. Si a la entrada de la estructura lattice

$$f_0(n) = g_0(n) = x(n)$$

entonces para cualquier sección m de la estructura lattice como se observa en la figura 5.5, los filtros de predicción son

$$f_m(n) = f_{m-1}(n) + K_m g_{m-1}(n-1) \quad m = 1, 2, 3, \dots, p-1 \quad (5.22)$$

$$g_m(n) = K_m f_{m-1}(n) + g_{m-1}(n-1) \quad m = 1, 2, 3, \dots, p-1 \quad (5.23)$$

la salida del estado $p-1$ corresponde a la salida de un filtro FIR de orden $p-1$

$$y(n) = f_{p-1}(n)$$

$m = 1, 2, 3, 4, \dots, p-1$ y $y(n)$ es la salida de estado $p-1$ del filtro lattice, por lo que existe una equivalencia con un filtro FIR. La salida $f_m(n)$ del estado m del filtro lattice puede expresarse como

$$f_m(n) = \sum_{i=0}^m a_m(i)x(n-i); \quad a_m(0) = 1 \quad (5.24)$$

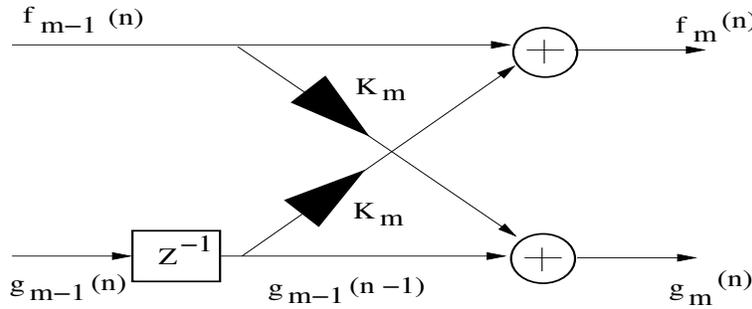


Figura 5.5: Filtro lattice de orden m

de la figura 5.5 y utilizando la transformada Z , se pueden escribir las ecuaciones (5.22) y (5.23) en variable z

$$F_m(z) = F_{m-1}(z) + K_m G_{m-1}(z)z^{-1} \quad (5.25)$$

$$G_m(z) = K_m F_{m-1}(z) + G_{m-1}(z)z^{-1} \quad (5.26)$$

$$m = 1, 2, 3, 4, \dots, p-1$$

en forma matricial

$$\begin{bmatrix} F_m(z) \\ G_m(z) \end{bmatrix} = \begin{bmatrix} 1 & K_m \\ K_m & 1 \end{bmatrix} \begin{bmatrix} F_{m-1}(z) \\ z^{-1}G_{m-1}(z) \end{bmatrix} \quad (5.27)$$

Por transformada Z , $F_m(z) = TZ\{f_m(n)\} = A_m(z)X(z)$, obteniendo la función de transferencia

$$A_m(z) = \frac{F_m(z)}{X_m(z)} = \frac{F_m(z)}{F_0(z)}$$

$g_m(n)$ también puede expresarse en forma de convolución, utilizando los coeficientes $b_m(i)$.

Donde se nota que los coeficientes para el filtro lattice de orden uno que produce $f_1(n)$ son $\{1, K_1\}$ y $\{1, a_1(n)\}$. Mientras que los coeficientes que producen la salida $g_1(n)$ son $\{K_1, 1\}$ y $\{a_1(n), 1\}$, estos dos últimos conjuntos de coeficientes están en orden inverso, entonces la salida $g_2(n)$ se puede expresar como

$$\begin{aligned} g_2(n) &= K_2 f_1(n) + g_1(n-1) \\ g_2(n) &= K_2[x(n) + K_1 x(n-1)] + K_1 x(n-1) + x(n-2) \\ g_2(n) &= K_2 x(n) + K_1(1 + K_2)x(n-1) + x(n-2) \\ g_2(n) &= a_2(2)x(n) + a_2(1)x(n-1) + x(n-2) \end{aligned} \quad (5.28)$$

Ahora los coeficientes son $\{a_2(2), a_2(1), 1\}$, de nuevo para la salida $g_2(n)$, los coeficientes $a_m(n)$ aparecen en orden inverso, entonces $g_m(n)$ para orden m del filtro lattice puede ser expresado como una suma de convolución

$$g_m(n) = \sum_{n=0}^m b_m(n)x(n-n) \quad (5.29)$$

Los coeficientes $\{b_m(n)\}$ son asociados con un filtro que produce la salida $g_m(n)$ pero en orden inverso

$$\begin{aligned} b_m(n) &= a_m(m-n); & b_m(0) &= 1 \\ n &= 0, 1, 2, 3, \dots, m \end{aligned} \quad (5.30)$$

Utilizando transformada Z para obtener $B_m(z)$, que se le puede considerar como la función de transferencia de un sistema FIR con coeficientes $\{b_m(n)\}$

$$G_m(z) = B_m(z)X(z) \quad B_m(z) = \frac{G_m(z)}{X(z)} \quad (5.31)$$

$$B_m(z) = \sum_{n=0}^m b_m(n)z^{-n} \quad (5.32)$$

como $b_m(n) = a_m(m-n)$ haciendo un cambio de variable, si $l = m-n$, $n = m-l$,

$$\begin{aligned} B_m(z) &= \sum_{n=0}^m a_m(m-n)z^{-n} = \sum_{n=0}^m a_m(l)z^{l-m} \\ B_m(z) &= z^{-m} \sum_{n=0}^m a_m(l)z^l = z^{-m} A_m(z) \end{aligned} \quad (5.33)$$

Se observa que los ceros de $B_m(z)$ son recíprocos de los ceros de $A_m(z)$, por esta razón $B_m(z)$ es llamado el recíproco de $A_m(z)$.

Tomando en cuenta las relaciones anteriores y utilizando la transformada Z , se tienen las relaciones:

$$\begin{aligned} F_0(z) &= G_0(z) = X(z) \\ F_m(z) &= F_{m-1}(z) + K_m z^{-1} G_{m-1}(z) \end{aligned} \quad (5.34)$$

$$G_m(z) = K_m F_{m-1}(z) + z^{-1} G_{m-1}(z) \quad (5.35)$$

Dividiendo por $X(z)$ y si $A_0(z) = B_0(z) = 1$

$$A_m(z) = A_{m-1}(z) + K_m z^{-1} B_{m-1}(z) \quad m = 1, 2, 3, 4, \dots, M - 1 \quad (5.36)$$

$$B_m(z) = K_m A_{m-1}(z) + z^{-1} B_{m-1}(z) \quad m = 1, 2, 3, 4, \dots, M - 1 \quad (5.37)$$

En forma matricial

$$\begin{bmatrix} A_m(z) \\ B_m(z) \end{bmatrix} = \begin{bmatrix} 1 & K_m \\ K_m & 1 \end{bmatrix} \begin{bmatrix} A_{m-1}(z) \\ z^{-1} B_{m-1}(z) \end{bmatrix} \quad (5.38)$$

Para el cálculo de los coeficientes K_i , la solución es calculada recursivamente empezando con $m = p$, así se obtiene una secuencia de un filtro FIR de longitud $(M - 1)$, uno para cada uno de los valores de m . Es importante notar que una caracterización de esta clase de m filtros FIR en forma directa requiere $m(m + 1)/2$ coeficientes.

En un filtro lattice se necesitan solo m coeficientes de reflexión K_i , es decir que este tipo de filtros provee una representación más compacta y si se requiere agregar mas etapas, no se alteran los parámetros calculados en las etapas previas. Por otro lado, en un filtro FIR al agregar mas etapas, cambian todos los coeficientes y es necesario recalcularlos.

5.4. Filtro lattice todo polo IIR

Este tipo de estructura puede verse como un filtro de síntesis o un filtro inverso, ya que dado un conjunto de coeficientes K_m se desea reproducir un proceso original AR $x(n)$ al aplicar como entrada un proceso estacionario de ruido blanco $v(n)$ con media cero.

Recordando que la función de transferencia de un filtro lattice tipo FIR con coeficientes $a_m(n)$

$$A_m(z) = \frac{F_m(z)}{X_m(z)} = \frac{F_m(z)}{F_0(z)} = 1 + \sum_{i=n}^m a_m(n) z^{-n}$$

donde la entrada al filtro es el predictor de orden cero $f_0(n)$ y las salidas son $f_m(n)$ y $g_m(n)$ para los predictores de orden m hacia adelante y hacia atrás respectivamente. Entonces para el diseño de un filtro todo polo se puede considerar como un filtro FIR inverso.

$$A_m(z) = \frac{F_0(z)}{F_m(z)} \quad (5.39)$$

donde se ha intercambiado la entrada con la salida, esto es que se tiene una entrada $f_m(n)$ que produce una salida $f_0(n)$.

Una estructura lattice para un filtro IIR todo polo, puede ser desarrollada de una estructura FIR, dada una función de un sistema todo polo

$$H(z) = \frac{1}{1 + \sum_{i=1}^p a_p(i)z^{-i}} = \frac{1}{A_p(z)} \quad (5.40)$$

esta función de transferencia corresponde a un sistema inverso de un filtro FIR lattice. Entonces se tiene un filtro IIR de orden p con estructura lattice de p etapas como se ve en la figura 5.9.

Considerando las ecuaciones en diferencias acopladas de los predictores hacia adelante y hacia atrás y desplegando $f_{m-1}(n)$

$$f_{m-1}(n) = f_m(n) - K_m g_{m-1}(n-1) \quad (5.41)$$

$$g_m(n) = K_m f_{m-1}(n) + g_{m-1}(n-1) \quad (5.42)$$

se pueden determinar el predictor hacia adelante al orden anterior $f_m(n)$ con base en $f_{m+1}(n)$ y $g_m(n-1)$ y el predictor hacia atrás $g_{m+1}(n)$ con base en $f_m(n)$ y $g_m(n-1)$

$$f_m(n) = x(n) \quad m = p \quad (5.43)$$

$$f_{m-1}(n) = f_m(n) - K_m g_{m-1}(n-1) \quad m = p, p-1, \dots, 1 \quad (5.44)$$

$$g_m(n) = K_m f_{m-1}(n) + g_{m-1}(n-1) \quad m = p, p-1, \dots, 1 \quad (5.45)$$

$$y(n) = f_0(n) = g_0(n) \quad (5.46)$$

Estas dos ecuaciones definen una red de dos puertos como se observa en la figura 5.6 y representan un filtro lattice todo polo de una etapa.

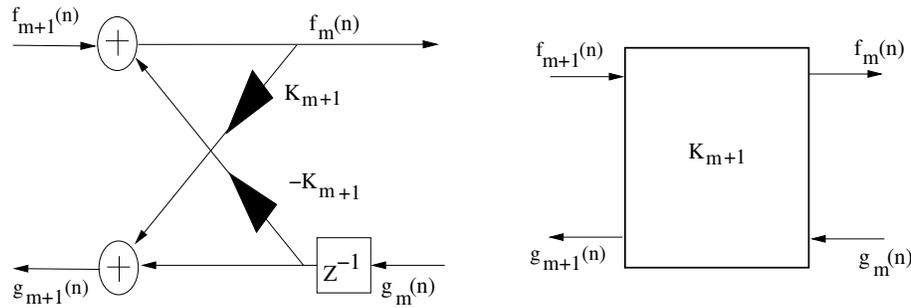


Figura 5.6: Filtro lattice todo polo de orden uno

5.4.1. Filtro lattice todo polo de orden uno

La conversión de los coeficientes $a_m(n)$ de un filtro IIR se puede realizar desarrollando las ecuaciones que resultan de las estructuras IIR y lattice e igualar las ecuaciones en diferencias.

Para un filtro IIR todo polo de orden uno, su función de transferencia es

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{1 + a_1(1)z^{-1}} \quad (5.47)$$

y la ecuación en diferencias para una entrada $x(n)$ y salida $y(n)$

$$y(n) = x(n) - a_1(1)y(n - 1) \quad (5.48)$$

despejando $x(n)$

$$x(n) = y(n) + a_1(1)y(n - 1) \quad (5.49)$$

que es similar a un filtro FIR con coeficientes $a_m(n)$ si se intercambia la entrada con la salida.

Para la estructura lattice de orden uno de la figura 5.7, si la entrada es $f_1(n) = x(n)$ y la salida $y(n) = f_0(n) = g_0(n)$

$$\begin{aligned} y(n) = f_0(n) = g_0(n) &= f_1(n) - K_1 g_0(n - 1) = x(n) - K_1 y(n - 1) \\ x(n) &= y(n) + K_1 y(n - 1) \end{aligned} \quad (5.50)$$

es decir que la relación de coeficientes es

$$K_1 = a_1(1)$$

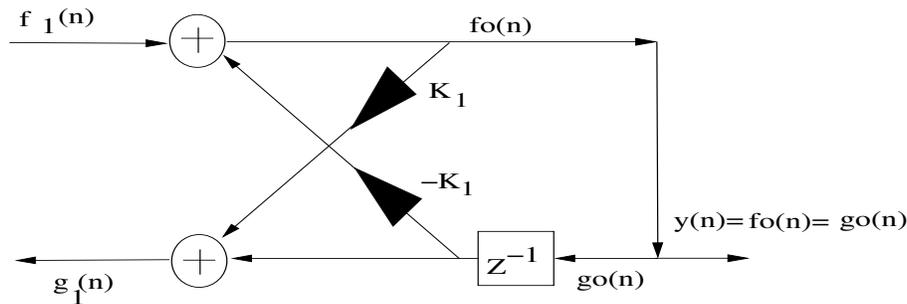


Figura 5.7: Filtro lattice todo polo de orden uno

5.4.2. Filtro lattice todo polo de orden dos

De manera similar se procede para un filtro IIR todo polo de orden dos con función de transferencia

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{1 + a_2(1)z^{-1} + a_2(2)z^{-2}} \quad (5.51)$$

y su ecuación en diferencias para una entrada $x(n)$ y salida $y(n)$

$$y(n) = x(n) - a_2(1)y(n - 1) - a_2(2)y(n - 2) \quad (5.52)$$

despejando $x(n)$

$$x(n) = y(n) + a_2(1)y(n - 1) + a_2(2)y(n - 2) \quad (5.53)$$

que es similar a un filtro FIR con coeficientes $a_m(n)$ donde se ha intercambiado la entrada con la salida para la estructura lattice de la figura 5.8, $f_2(n) = x(n)$.

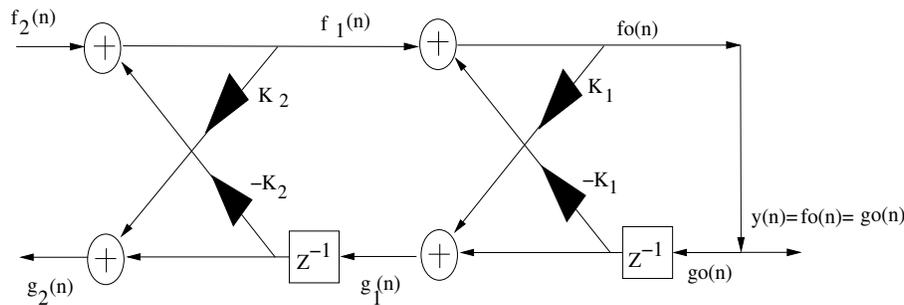


Figura 5.8: Filtro lattice todo polo de orden dos

Desarrollando las ecuaciones en diferencias

$$y(n) = f_0(n) = g_0(n) = f_1(n) - K_1g_0(n - 1) = f_1(n) - K_1y(n - 1) \quad (5.54)$$

$$\begin{aligned} f_1(n) &= f_2(n) - K_2 g_1(n-1) = x(n) - K_2 g_1(n-1) \\ g_1(n) &= g_0(n-1) + K_1 f_0(n) = y(n-1) + K_1 y(n) \end{aligned} \quad (5.55)$$

sustituyendo $g_1(n)$ en $f_1(n)$

$$\begin{aligned} f_1(n) &= x(n) - K_2 \left(y(n-2) + K_1 y(n-1) \right) \\ f_1(n) &= x(n) - K_1 K_2 y(n-1) - K_2 y(n-2) \end{aligned} \quad (5.56)$$

sustituyendo $f_1(n)$ en ecuación (5.54)

$$\begin{aligned} y(n) &= x(n) - K_1 K_2 y(n-1) - K_2 y(n-2) - K_1 y(n-1) \\ y(n) &= x(n) - K_1(1 + K_2)y(n-1) - K_2 y(n-2) \end{aligned} \quad (5.57)$$

despejando $x(n)$

$$x(n) = y(n) + K_1(1 + K_2)y(n-1) + K_2 y(n-2) \quad (5.58)$$

igualando coeficientes con los de un filtro FIR de segundo orden

$$a_2(1) = K_1(1 + K_2) \quad a_2(2) = K_2 \quad (5.59)$$

entonces

$$K_1 = \frac{a_2(1)}{1 + K_2} \quad (5.60)$$

Esta ecuación también se obtuvo para el caso de la conversión de un filtro lattice FIR de segundo orden. Comparando con la relación de coeficientes para el filtro lattice de un filtro FIR, de manera similar para un filtro todo polo tipo lattice.

5.4.3. Filtro lattice todo polo de orden p

En la figura 5.9 se observa que cada etapa tiene una entrada $f_m(n)$ y una salida $f_{m+1}(n)$ que están relacionadas por las ecuaciones recursivas (5.41) y (5.42) y $y(n) = f_0(n) = g_0(n)$.

En general los coeficientes K_i de la estructura lattice para un filtro IIR todo polo se pueden calcular por medio de un algoritmo step down (visto en capítulo anterior).

5.4.4. Filtro lattice de segundo orden con polos y ceros

En este caso se hacen algunos cambios en las estructuras lattice vistas anteriormente con el fin de comprobar que una estructura lattice también puede generar una función de transferencia de un filtro IIR con polos y ceros.

De la figura 5.10 para la primera estructura lattice se pueden escribir las ecuaciones

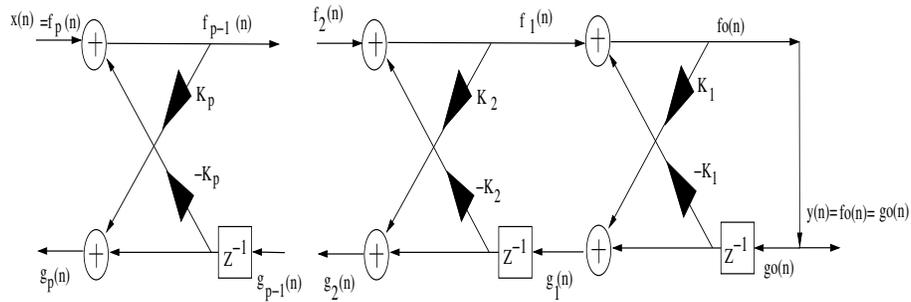


Figura 5.9: Filtro lattice todo polo de orden p

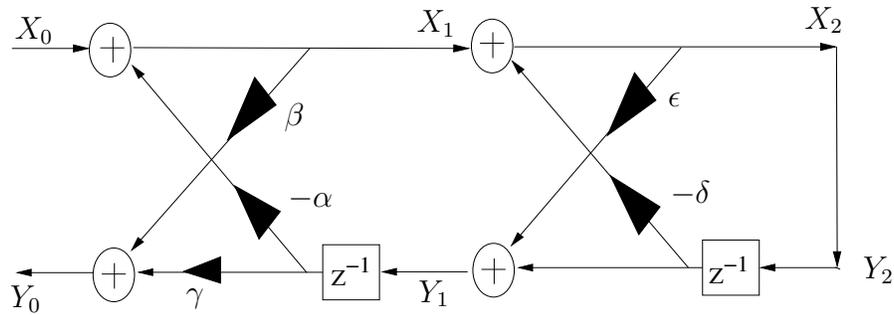


Figura 5.10: Filtro lattice IIR con polos y ceros

$$\begin{aligned} X_1 &= X_0 - \alpha z^{-1} Y_1 \\ X_0 &= X_1 + \alpha z^{-1} Y_1 \\ Y_0 &= \beta X_1 + \gamma z^{-1} Y_1 \end{aligned}$$

En forma matricial

$$\begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} = \begin{bmatrix} 1 & \alpha z^{-1} \\ \beta & \gamma z^{-1} \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} \quad (5.61)$$

para la segunda estructura lattice

$$\begin{aligned} X_2 &= X_1 - \delta z^{-1} Y_2 \\ X_1 &= X_2 + \delta z^{-1} Y_2 \\ Y_1 &= \epsilon X_2 + z^{-1} Y_2 \end{aligned}$$

En forma matricial

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} = \begin{bmatrix} 1 & \delta z^{-1} \\ \epsilon & z^{-1} \end{bmatrix} \begin{bmatrix} X_2 \\ Y_2 \end{bmatrix} \quad (5.62)$$

sustituyendo el vector con subíndices uno en ecuación (5.61)

$$\begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} = \begin{bmatrix} 1 & \alpha z^{-1} \\ \beta & \gamma z^{-1} \end{bmatrix} \begin{bmatrix} 1 & \delta z^{-1} \\ \epsilon & z^{-1} \end{bmatrix} \begin{bmatrix} X_2 \\ Y_2 \end{bmatrix} \quad (5.63)$$

desarrollando el producto entre matrices

$$\begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} = \begin{bmatrix} 1 + \alpha\epsilon z^{-1} & \delta z^{-1} + \alpha z^{-2} \\ \beta + \gamma\epsilon z^{-1} & \beta\delta z^{-1} + \gamma z^{-2} \end{bmatrix} \begin{bmatrix} X_2 \\ Y_2 \end{bmatrix} \quad (5.64)$$

al dividir la segunda línea entre la primera del arreglo matricial y como $X_2 = Y_2$, se obtiene la función de transferencia de un filtro IIR de segundo orden con polos y ceros

$$H(z) = \frac{Y_0(z)}{X_0(z)} = \frac{\beta + (\gamma\epsilon + \beta\delta)z^{-1} + \gamma z^{-2}}{1 + (\alpha\epsilon + \delta)z^{-1} + \alpha z^{-2}} \quad (5.65)$$

con coeficientes

$$b_0 = \beta \quad b_1 = \gamma\epsilon + \beta\delta \quad b_2 = \gamma$$

$$a_0 = 1 \quad a_1 = \alpha\epsilon + \delta \quad a_2 = \alpha$$

esto prueba la equivalencia entre estructuras IIR y lattice.

5.5. Filtros lattice-ladder

En general un filtro IIR contiene q ceros y p polos, comúnmente $q \leq p$, su función de transferencia se escribe

$$H(z) = \frac{\sum_{i=0}^q b_q(i)z^{-i}}{1 + \sum_{i=1}^p a_p(i)z^{-i}} = \frac{B_q(z)}{A_p(z)} = \frac{Y(z)}{X(z)} \quad (5.66)$$

$H(z)$ puede ser implementada en cascada con funciones de transferencia solo ceros $B_q(z)$ y todo polo $1/A_p(z)$

$$H(z) = \frac{Y(z)}{X(z)} = [B_q(z)] \left[\frac{1}{A_p(z)} \right] = \left[\frac{Y(z)}{W(z)} \right] \left[\frac{W(z)}{X(z)} \right] \quad (5.67)$$

separando los factores

$$[B_q(z)] = \left[\frac{Y(z)}{W(z)} \right] \quad \text{y} \quad \left[\frac{1}{A_p(z)} \right] = \left[\frac{W(z)}{X(z)} \right]$$

se obtiene el par de ecuaciones en diferencias

$$w(n) = x(n) - \sum_{i=1}^p a_p(i)w(n-i) \quad (5.68)$$

donde $w(n)$ es la salida del filtro todo polo con entrada $x(n)$ y

$$y(n) = \sum_{i=0}^q b_q(i)w(n-i) \quad (5.69)$$

es la salida de un filtro solo ceros con entrada $w(n)$.

Debido a que los ceros de $H(z)$ son introducidos al tomar una combinación lineal de las salidas retardadas del filtro todo polo, esto se puede realizar con un filtro lattice todo polo, y el filtro todo cero como una combinación lineal de señales del error hacia adelante $e_0^f(n-i)$ con $i = 0, 1, 2, \dots, q$. Sin embargo, una forma más eficiente de realizar el filtro es tomando una combinación lineal del predictor de error hacia atrás [16], [31]

$$y(n) = \sum_{i=0}^q c_q(i)e_i^b(n) \quad (5.70)$$

como se muestra en la figura 5.11, donde $p \geq q$.

Una estructura lattice del filtro IIR, puede construirse de una estructura lattice todo polo con coeficientes K_m , $1 \leq m \leq p$ y una parte "ladder" que tiene una salida como una combinación lineal de las señales $g_m(n)$ y coeficientes c_i como se muestra en la figura 5.11.

La función de transferencia del filtro solo cero se escribe

$$Y(z) = \sum_{j=0}^q c_q(j)E_j^b(z) \quad (5.71)$$

Recordando la relación que existe entre las funciones de transferencia de los coeficientes hacia adelante y hacia atrás

$$A_j^b = z^{-j}A_j^*(1/z) = \sum_{m=0}^j a_j^*(m)z^{m-j} \quad (5.72)$$

escribiendo la relación entre los errores de predicción hacia atrás y hacia adelante en variable z

$$E_j^b(z) = A_j^b(z)E_0^f(z) = \frac{A_j^b(z)}{A_p(z)}E_p^f(z) \quad (5.73)$$

entonces sustituyendo $E_j^b(z)$ en $Y(z)$

$$Y(z) = \sum_{j=0}^q c_q(j) \frac{A_j^b(z)}{A_p(z)} E_p^f(z) \quad (5.74)$$

despejando la relación entrada $E_p^f(z)$ y salida $Y(z)$

$$H(z) = \frac{E_p^f(z)}{Y(z)} = \frac{\sum_{j=0}^q c_q(j) A_j^b(z)}{A_p(z)} = \frac{B_q(z)}{A_p(z)} \quad (5.75)$$

que contiene p polos y q ceros, donde los ceros de $H(z)$ son las raíces del polinomio

$$B_q(z) = \sum_{j=0}^q c_q(j) A_j^b(z) \quad (5.76)$$

que es una función de coeficientes $c_q(j)$ relacionados con los K_i . Si se reemplaza $A_j^b(z)$

$$A_j^b(z) = z^{-j} A_j^*(1/z^*) = \sum_{m=0}^j a_j^*(m) z^{m-j}$$

$$B_q(z) = \sum_{j=0}^q c_q(j) \sum_{m=0}^j a_j^*(m) z^{m-j} \quad (5.77)$$

cambiando variables $k = j - m$

$$B_q(z) = \sum_{j=0}^q \sum_{k=0}^j c_q(j) a_j^*(j-k) z^{-k} \quad (5.78)$$

intercambiando el orden de las sumatorias y adecuando el índice de la segunda sumatoria para el índice k

$$B_q(z) = \sum_{k=0}^q \left[\sum_{j=k}^q c_q(j) a_j^*(j-k) \right] z^{-k} = \sum_{k=0}^p b_q(k) z^{-k} \quad (5.79)$$

es decir que

$$b_q(k) = \sum_{j=k}^q c_q(j) a_j^*(j-k) \quad (5.80)$$

Esta última ecuación indica que los coeficientes $b_q(k)$ se pueden obtener de los $c_q(k)$ y K_i . Los *coeficientes del filtro ladder* c_i determinan los ceros de la función $H(z)$.

De un filtro todo polo, si se conocen los coeficientes K_i se pueden encontrar los $a_j(j-k)$. Desarrollando la última ecuación donde $a_j(0) = 1$

$$b_q(k) = c_q(k) + \sum_{j=k+1}^q c_q(j)a_j^*(j-k) \quad (5.81)$$

despejando $c_q(k)$

$$c_q(k) = b_q(k) - \sum_{j=k+1}^q c_q(j)a_j^*(j-k) \quad (5.82)$$

para $k = q-1, q-2, \dots, 0$, con condición inicial

$$c_q(q) = b_q(q) \quad (5.83)$$

Ejemplo:

Para un filtro IIR orden cuatro, desarrollar las expresiones para calcular los coeficientes c_i y K_i de una estructura tipo lattice-ladder. Si se conoce la función de transferencia del filtro

$$H(z) = \frac{b_4(0) + b_4(1)z^{-1} + b_4(2)z^{-2} + b_4(3)z^{-3} + b_4(4)z^{-4}}{1 + a_4(1)z^{-1} + a_4(2)z^{-2} + a_4(3)z^{-3} + a_4(4)z^{-4}}$$

Solución:

Utilizando el algoritmo de Levinson-Durbin inverso o step-down:

De la función de transferencia para el orden $m = 4$ se conocen los coeficientes $a_4(i)$

$$a_4(i) = \{a_4(0) = 1, a_4(1), a_4(2), a_4(3), a_4(4)\}$$

entonces por definición de $K_i = a_i(i)$, $K_4 = a_4(4)$, se pueden calcular los demás coeficientes al orden inferior con la fórmula

$$a_{m-1}(n) = \frac{1}{1 - K_m^2} [a_m(n) - K_m a_m(m-n)]$$

para $m = 3, 2, 1$.

Si el orden es $m = 3$

$$a_3(1) = \frac{1}{1 - K_4^2} [a_4(1) - K_4 a_4(3)]$$

$$a_3(2) = \frac{1}{1 - K_4^2} [a_4(2) - K_4 a_4(2)]$$

$$a_3(3) = \frac{1}{1 - K_4^2} [a_4(3) - K_4 a_4(1)]$$

entonces se obtiene

$$a_3(i) = \{a_3(0) = 1, a_3(1), a_3(2), a_3(3) = K_3\}$$

para el orden $m = 2$

$$a_2(1) = \frac{1}{1 - K_3^2} [a_3(1) - K_3 a_2(2)]$$

$$a_2(2) = \frac{1}{1 - K_3^2} [a_3(2) - K_3 a_2(1)]$$

$$a_2(i) = \{a_2(0) = 1, a_2(1), a_2(2) = K_2\}$$

para el orden $m = 1$

$$a_1(1) = \frac{1}{1 - K_2^2} [a_2(1) - K_2 a_1(1)]$$

$$a_1(i) = \{a_1(0) = 1, a_1(1) = K_1\}$$

Al final se ha calculado el conjunto de coeficientes

$$K_i = \{K_1, K_2, K_3, K_4\}$$

Los coeficientes $c_4(i)$ del filtro ladder se calculan recursivamente con base en los coeficientes $a_4(i)$ y $b_4(i)$ conocidos de la función de transferencia del filtro IIR, la estructura del filtro lattice-ladder se muestra en la figura 5.11.

$$\begin{aligned} c_4(4) &= b_4(4) \\ c_4(3) &= b_4(3) - c_4(4)a_4(1) \\ c_4(2) &= b_4(2) - c_4(3)a_3(1) - c_4(4)a_4(2) \\ c_4(1) &= b_4(1) - c_4(2)a_2(1) - c_4(3)a_3(2) - c_4(4)a_4(3) \\ c_4(0) &= b_4(0) - c_4(1)a_1(1) - c_4(2)a_2(2) - c_4(3)a_3(3) - c_4(4)a_4(4) \end{aligned}$$

5.6. Resumen

En este capítulo se han analizado los filtros lattice y ladder, sus virtudes, sus formas de deducción e implementación. Además se ha hecho énfasis en la comparación de estructuras FIR e IIR, y cómo se puede convertir un tipo de filtro a otro. Hay que hacer notar que las características importantes de los filtros lattice es su modularidad, que el agregar más etapas no altera el cálculo de los otros coeficientes y algo muy importante es que sus coeficientes $|K_i| \leq 1$, esto hace que sus estructuras sean de mucho interés en la implementación de filtros digitales en arquitecturas de aritmética de punto entero.

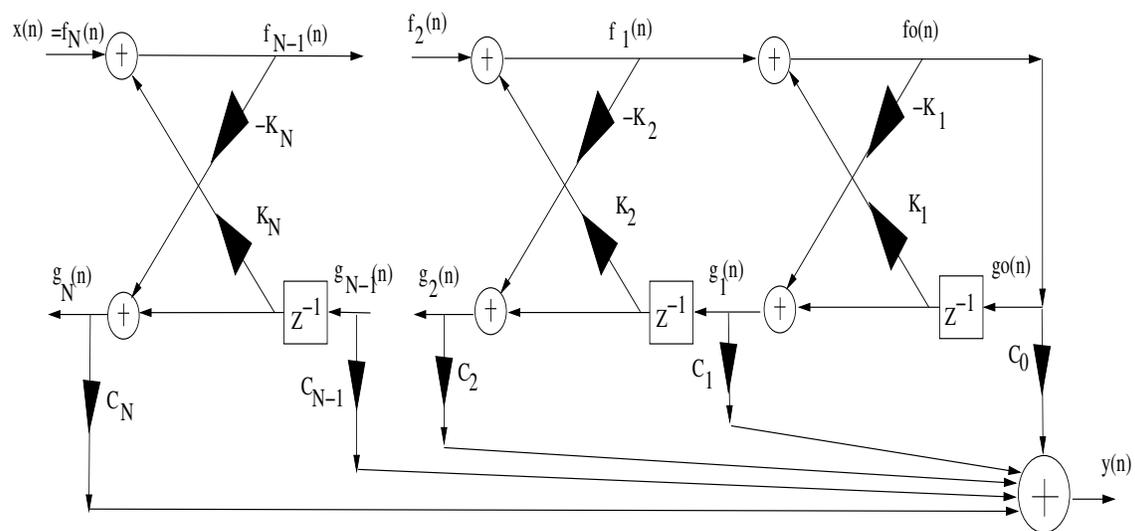


Figura 5.11: Filtro lattice-ladder

Capítulo 6

Aplicación en síntesis de voz

Este capítulo tiene como objetivo integrar los conceptos de predicción lineal, el modelo AR, filtros FIR, IIR y lattice en la realización de una aplicación particular, *síntesis de voz*, esto es, desarrollar un modelo que sea capaz de producir voz. Para realizar esta aplicación se hace una breve presentación del órgano productor de voz, su forma de modelarlo, un modelo en el tiempo discreto y la aplicación de la mayoría de los filtros vistos anteriormente, se mostrarán pruebas de la producción de síntesis de voz, que desde otro punto de vista también se puede pensar como un codificador-decodificador de voz (VOCODER), y al final se muestran los resultados obtenidos en el tiempo como en el espectro, haciendo comparaciones entre la voz sintética y real se observa la similitud de los resultados.

6.1. Síntesis de voz

El sistema humano de producción de voz está compuesto por un conjunto de órganos: los pulmones como la fuente generadora de aire, la tráquea, las cuerdas vocales, el glottis, la lengua, los dientes y los labios como elementos de conducción y vibración del sonido. En este esquema se puede encontrar una relación entrada/salida en la producción de un fonema, a cada fonema se le podrá asociar una función de transferencia que lo modele como un sistema lineal e invariante en el tiempo (SLIT), entonces se podría generar un conjunto de sistemas o filtros que adaptan sus características o una función de transferencia de acuerdo al modelo que se desea reproducir.

La producción de voz

Considerando un modelo simplificado entrada/salida, es decir, un sistema con entrada $x(t)$, respuesta al impulso $h(t)$ y salida $y(t)$, donde la excitación o entrada $x(t)$ se produce en el glottis y la salida $y(t)$ es precisamente la voz en los labios, entonces la presión del aire es forzada desde el glottis hasta los labios a través del tracto vocal para producir la voz.

En la figura 6.1 la entrada $x(t)$ corresponde a la presión del aire en el glottis, la respuesta al impulso del sistema SLIT es derivada de las propiedades físicas del tracto vocal y la salida $y(t)$ sería la voz [22]. Es decir que el proceso de producción de voz puede modelarse como la convolución de $x(t)$ con $h(t)$, en el tiempo continuo

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t - \tau)d\tau = h(t) * x(t) = x(t) * h(t) \quad (6.1)$$

donde $h(t)$ es la respuesta al impulso del sistema de producción de voz.

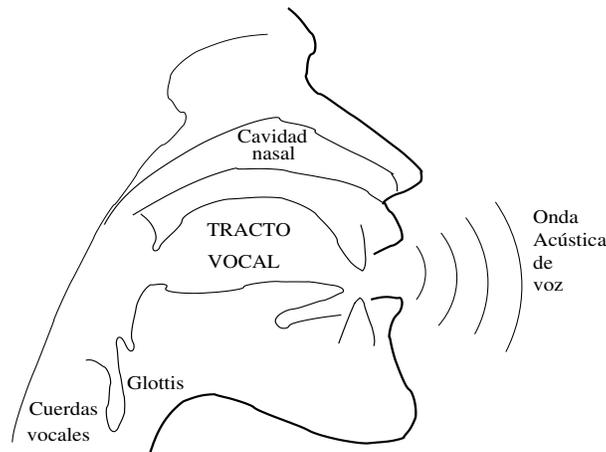


Figura 6.1: Sistema productor de voz

Modelo del tracto vocal

Trasladando la idea anterior a sistemas y señales discretas, la producción de una onda de voz consiste de una onda de excitación del glottis que viaja a través del tracto vocal hasta que se genera una onda de salida en los labios, como se ve en figura 6.1.

Para efectuar la deconvolución en un proceso físico como la voz, se necesita un modelo del sistema productor de voz, esto es, resolver el problema inverso de la producción de voz. Un modelo aproximado del tracto vocal puede ser un sistema de parámetros distribuidos visto como un tubo acústico con secciones cilíndricas diferentes acopladas en cascada, donde cada sección tiene una longitud L y un área transversal A_m como se ilustra en figura 6.2 [4], [9], [35], [11].

Desde el punto de vista físico, cuando el glottis crea una excitación acústica, se produce un sonido y en cada sección del modelo la onda de sonido cambia como una función del tiempo. Si se conoce la onda de excitación del glottis como una función del tiempo y las ondas en las secciones $A_1, A_2, A_3, \dots, A_N$, entonces se puede producir voz [9], [35], [11]. Para

un sonido dado, el área de la sección de un cilindro es denotado como A_m , donde m va de 1 a N , de la sección cercana a los labios a la sección cercana al glottis.

Considerando una onda plana que se propaga del glottis a los labios por todo el modelo acústico, la velocidad de la onda acústica satisface en cada sección el modelo de una onda unidimensional acústica

$$\frac{\partial^2 U_m(x, t)}{\partial t^2} = v^2 \frac{\partial^2 U_m(x, t)}{\partial x^2} \quad (6.2)$$

donde $U_m(x, t)$ es la velocidad de la onda en una sección cilíndrica m , v es la velocidad del sonido en el tracto vocal, x es la variable espacial, t es el tiempo. La solución de la ecuación de onda (6.2) puede expresarse como [11]

$$U_m(x, t) = C_1 U_m\left(\frac{t - x}{v}\right) + C_2 U_m^-\left(\frac{t + x}{v}\right) \quad (6.3)$$

se observa que en cada sección cilíndrica viajan dos ondas, la componente $C_1 U_m((t - x)/v)$ (onda emitida) que viaja a la derecha y la componente $C_1 U_m^-((t + x)/v)$ (onda reflejada) que viaja a la izquierda. En la frontera entre las secciones $m - 1$ y m con áreas A_{m-1} y A_m respectivamente, existe una discontinuidad que altera la onda generando una serie de turbulencias. Debido a las discontinuidades en la frontera de las secciones adyacentes, la onda emitida será parcialmente reflejada y parcialmente transmitida, esto sucede en cada frontera y produce un modelo de reflexiones múltiples. La energía acústica entre las ondas

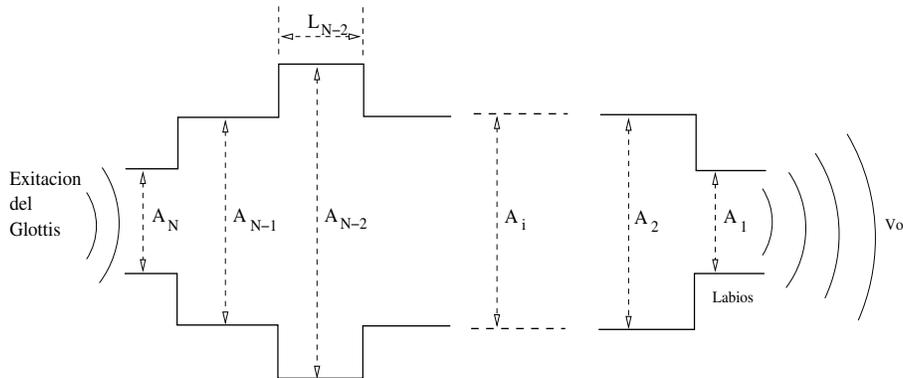


Figura 6.2: Modelo acústico del tracto vocal

emitidas y reflejadas está gobernada por los *coeficientes de reflexión* K_m . Por ejemplo si entre la secciones $m - 1$ y m se tienen las áreas A_{m-1} y A_m , de las consideraciones físicas de velocidad y presión los coeficientes de reflexión de las ondas entre secciones está dado por [4], [9], [22]¹

$$K_m = \frac{A_{m-1} - A_m}{A_{m-1} + A_m} \quad (6.4)$$

¹Se puede observar que existe una similitud entre esta expresión de los coeficientes, la definición de transformada bilineal y el coeficiente de reflexión de una línea de transmisión [6].

después de tomar en cuenta las condiciones de frontera entre secciones se obtiene [22]

$$y_m^+(t) = v_m U_m^+(t + \tau - t_m) \quad y_m^-(t) = -v_m U_m^-(t - \tau - t_m) \quad (6.5)$$

donde

$$v_m \equiv \prod_{j=1}^m (1 + K_j) \quad m = 1, 2, 3, \dots, N \quad v_0 \equiv 1 \quad (6.6)$$

considerando a $y_m^+(t)$ como la *onda emitida* en la superficie m y $y_m^-(t)$ como la *onda reflejada* en la superficie m , entonces se tiene el par de ecuaciones acopladas

$$y_m^+(t) = y_{m-1}^+(t) + K_m y_{m-1}^-(t) \quad (6.7)$$

$$y_m^-(t) = y_{m-1}^-(t - T) + K_m y_{m-1}^+(t - T) \quad m = 1, 2, 3, \dots, N \quad (6.8)$$

con $T = 4\tau = 2L/v$, que es dos veces el tiempo necesario para que una onda se propague a través de un cilindro de longitud L . Si se muestrean las ecuaciones (6.7) y (6.8) al tiempo $t = nT$ y $T = 1$, entonces se obtienen las ecuaciones en el tiempo discreto

$$y_m^+(n) = y_{m-1}^+(n) + K_m y_{m-1}^-(n) \quad (6.9)$$

$$y_m^-(n) = y_{m-1}^-(n - 1) + K_m y_{m-1}^+(n - 1) \quad m = 1, 2, 3, \dots, N \quad (6.10)$$

estas dos últimas ecuaciones relacionan las ondas emitida y reflejada en las secciones $m - 1$ y m , además esta relación se puede describir por medio de una *estructura lattice* como se muestra en la figura 6.3.

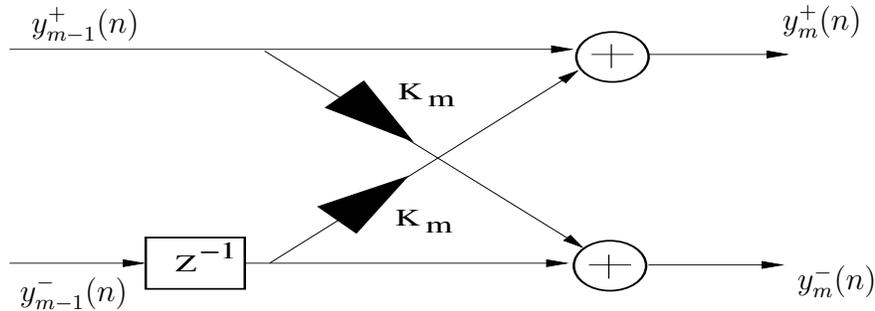


Figura 6.3: Modelo lattice del tracto vocal para una sección L

De la figura 6.3 y para un modelo lattice de orden N , se deduce que la señal $y_N^+(n)$ es proporcional a la forma de onda de voz muestreada y $y_0^+(n)$ es proporcional a la forma de onda muestreada de excitación del glottis. Si se hace $x(n) = y_0(n)$ y $y(n) = y_N(n)$, entonces

para un modelo discreto con entrada $x(n)$ y salida $y(n)$, están relacionadas por medio de un modelo autorregresivo de orden N dado por

$$y(n) + a_N(1)y(n-1) + a_N(2)y(n-2) + a_N(3)y(n-3) + \dots + a_N(N)y(n-N) = x(n) \quad (6.11)$$

donde la salida $y(n)$ se puede aproximar como un proceso estacionario en el sentido amplio (WSS) lo que implica que la entrada $x(n)$ sea ruido puro, y si se conocen los coeficientes $a_N(i)$ del modelo, la salida es la voz sintética generada.

En una señal de voz pueden identificarse tres tipos de sonidos diferentes cuya clasificación se hace de acuerdo a la forma de excitación del tracto vocal [1], [4], [11], [22], [35]:

- **Sonidos voceados**, se producen al forzar aire a través del glottis con la tensión de las cuerdas vocales ajustada para que vibren a una oscilación, este tipo de sonidos se puede generar excitando pulsos cuasi periódicos al modelo a utilizar.
- **Sonidos No-voceados**, se generan por alguna constricción en algún punto del tracto vocal forzando el aire a través de un conducto, generando una turbulencia lo cual produce una fuente de ruido que excita el tracto vocal.
- **Sonidos explosivos**, se generan cuando se produce un cierre total antes del tracto vocal y liberando abruptamente el aire y así excitar el tracto vocal.
En los procesos de síntesis de voz, se utilizan con más frecuencia la reproducción de sonidos voceados y no-voceados.

Producción de voz sintética

La síntesis de voz consiste de dos procesos:

- **Análisis** que consiste en extraer los parámetros de una señal de voz verdadera, almacenarlos o transmitirlos y posteriormente se pueda reproducir la voz. En este proceso, la cantidad de parámetros a extraer es mucho menor que la cantidad de muestras de la señal, por que se puede considerar como un proceso de compresión de voz.
- La **Síntesis de voz** se puede considerar como un proceso de deconvolución, ya que dada una señal de voz $y(n)$ desconocida, se calcula un modelo AR tal que $H(z) = 1/A(z)$. Es decir una vez conocida $H(z)$ el sistema es excitado por una señal $x(n)$ conocida para reproducir la voz sintética (figura 6.4).

En la figura 6.4 se muestra un modelo para producción de voz sintética, donde se tienen dos fuentes de señales $x(n)$, una para producir sonidos parecidos al ruido (No-voceada) y la otra fuente para producir sonidos con cierta armonía (voceadas), estas señales alimentan a un filtro AR con coeficientes a_i que depende del segmento de voz a producir y que previamente han sido extraídos de la señal de voz original.

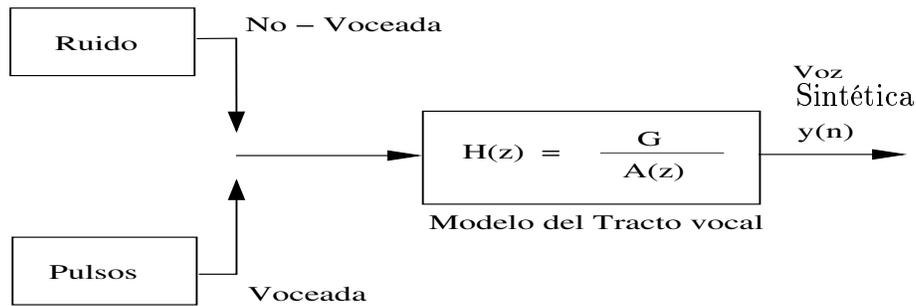


Figura 6.4: Modelo de producción de voz sintética

La señal de voz satisface el modelo $H(z)$ y si se minimiza la media del error cuadrático, se llega a un modelo donde el error $e(n)$ del proceso de análisis es igual a la excitación $x(n)$ multiplicada por la ganancia del tramo de voz y los parámetros $a_p(k)$, $k = 1, 2, 3, 4, \dots, p$.

6.2. Diseño del sistema de síntesis de voz

En la figura 6.5 se presenta un esquema general del proceso de un sistema de síntesis que engloba un *bloque de análisis*, donde se estimarán los parámetros de la señal de voz original y un *bloque de síntesis*, que mediante los parámetros obtenidos en el proceso de análisis se reconstruirá la señal. A continuación se desglosan los procesos de análisis y síntesis respectivamente [5], [7], [8].

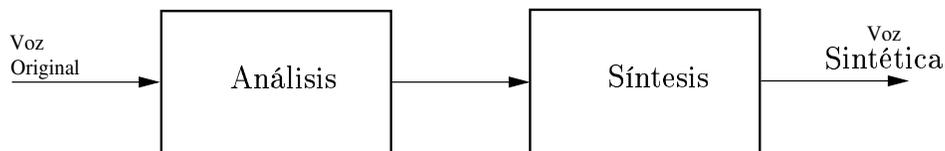


Figura 6.5: Sistema general de síntesis de voz.

■ Proceso de Análisis:

1. Estimación de la energía de la señal para decidir si el segmento es de silencio. Una forma de calcular la energía del segmento de voz en estudio es por medio de la ecuación (6.12)

$$E = \sum_{i=0}^{N-1} |s(n)|^2 \quad (6.12)$$

Donde E es la energía del segmento, $s(n)$ es el segmento de voz y N es el número de muestras para cada segmento. La elección del valor del umbral para la determinación de la existencia de un segmento de silencio dependerá de las condiciones de ruido donde se realice la adquisición de la voz. Cabe mencionar que si el segmento es considerado como silencio no será necesario estimar los parámetros de las siguientes etapas en el proceso de análisis.

2. Aplicar un proceso de ventaneo a la señal de voz original eligiendo la ventana de Hamming por ser la más utilizada para síntesis de voz, debido a que el ancho de transición aproximado del lóbulo principal de la ventana es de $8\pi/N$ y la diferencia de amplitud entre el lóbulo principal y el primer lóbulo lateral es de -41 db [32]. La ecuación (6.13) muestra la definición de la ventana de Hamming

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \quad (6.13)$$

La longitud de la ventana a utilizar es de $N = 240$ muestras, es decir, adquirimos un segmento en 30 ms a una frecuencia de muestreo de $f_s = 8$ kHz, para este tamaño de ventana, según experimentos, a la señal de voz se le puede considerar estacionaria. Si se desea experimentar con la calidad de la voz sintética se puede utilizar un traslape de ventanas, se recomienda el 30% del total del segmento.

3. Decidir la naturaleza del segmento bajo estudio y el período de “pitch”, mediante *el método del recorte central* [29], que consiste en recortar las amplitudes de una ventana entre los niveles $\pm C_L$, esto es considerar sólo las amplitudes de la ventana a partir de los niveles de umbral propuestos y posteriormente calcular su función de autocorrelación para determinar la periodicidad de la señal. En forma analítica se define la función del recorte central

$$\begin{aligned} y(n) &= s(n) - C_L && \text{Si } s(n) \geq C_L \\ y(n) &= s(n) + C_L && \text{Si } s(n) \leq -C_L \\ y(n) &= 0 && \text{Otro caso} \end{aligned} \quad (6.14)$$

Otra manera de definir a la función del recorte central, es considerando una forma más eficiente de recortar a la señal por razones de formato numérico²

$$\begin{aligned} y(n) &= 1 && \text{Si } s(n) \geq C_L \\ y(n) &= -1 && \text{Si } s(n) \leq -C_L \\ y(n) &= 0 && \text{Otro caso} \end{aligned} \quad (6.15)$$

²Evitando así desbordamientos al momento de realizar las multiplicaciones y acumulaciones en la función de autocorrelación.

El efecto de aplicar un proceso de recorte es conservar solo la información de la periodicidad de la señal. El nivel de umbral de C_L es calculado para cada ventana si dividimos a la ventana en tres subsegmentos y empleamos la siguiente estrategia:

- Encontrar las amplitudes máximas del primer y tercer subsegmento (S_1 y S_3).
- Calcular el umbral C_L

$$C_L = K \min(S_1, S_3) \quad (6.16)$$

Donde el operador $\min(\cdot)$ calcula el valor mínimo entre S_1 y S_3 y K es un parámetro de calibración que se encuentra en el intervalo de 0.6 a 0.8 [29].

4. Una vez determinado el umbral C_L y teniendo la señal recortada, aplicamos la función de autocorrelación a la señal recortada. La función de autocorrelación se define como

$$R_{ss}(l) = \sum_{n=0}^{N-1} s(n)s(n+l) \quad (6.17)$$

Donde $s(n)$ es el segmento de voz recortado, N el número de muestras del segmento y l el índice temporal de retraso.

En el caso de segmentos de sonidos voceados existe una periodicidad de la forma de onda en la señal, no así para ventanas no voceadas. Por tanto, la función de autocorrelación tiene la propiedad de ser periódica si el segmento de la señal en estudio es periódica. Una ventaja de usar la función de autocorrelación es que presenta con mayor realce la propiedad de periodicidad de las señales, en consecuencia, la propiedad de periodicidad puede ser utilizada como criterio para la estimación de la naturaleza del segmento. Sin embargo, además de resaltar la periodicidad del segmento, también muestra información de la energía de la señal, $R_{ss}(0)$. Esta información adicional puede propiciar decisiones erróneas [29]. Entonces para la determinación de la naturaleza de la ventana se calcula el valor máximo de la función de autocorrelación comprendida entre las muestras 20 a la 200, y este valor es comparado con un valor umbral igual a $0.3R_{ss}(0)$. Si el valor máximo es mayor que el valor umbral se decide como un segmento voceado, si es menor, el segmento será no voceado.

5. Aplicar al segmento (ya sea voceado o no voceado) un proceso de filtrado de pre-énfasis (filtro FIR de primer orden paso altas). En la ecuación (6.18) definimos el filtro de pre-énfasis que solo posee un cero en α

$$H_{pre}(z) = 1 - \alpha z^{-1} \quad (6.18)$$

Donde α es un valor cercano a 0.9. El valor de α comúnmente usado es de 0.9375. [29]. La ecuación en diferencias del filtro de pre-énfasis es

$$y(n) = x(n) - \alpha x(n-1) \quad (6.19)$$

Donde $y(n)$ es la salida y $x(n)$ es la entrada al filtro.

6. Estimar los parámetros LPC y calcular la ganancia del segmento por medio del *algoritmo de Levinson - Durbin* [16].
7. Genera un archivo de salida con las características del segmento de voz, los coeficientes LPC y la energía de la señal (ver tabla 6.1). En un sistema de comunicaciones esta información se podría codificar y transmitir a través de un canal.

■ **Proceso de Síntesis:**

1. Lectura de parámetros obtenidos del proceso de análisis.
2. Si el segmento en el proceso de análisis se decidió como silencio, generar una ventana de ceros.
3. Aplicar un filtro todo polo definido por los parámetros LPC y la ganancia a la entrada de excitación, según sea la naturaleza del segmento (voceado - tren de impulsos, no voceado - ruido blanco, figura 6.4). Entonces la función de transferencia del filtro todo polo está definida por la ecuación

$$H(z) = \frac{G}{1 + \sum_{i=1}^p a_i z^{-1}} \quad (6.20)$$

o bien por su ecuación en diferencias

$$\hat{y}(n) = Gx(n) - a_1\hat{y}(n-1) - a_2\hat{y}(n-2) - \dots - a_p\hat{y}(n-p) \quad (6.21)$$

Donde G es la ganancia del segmento, $x(n)$ la entrada de excitación, $\hat{y}(n)$ es la voz sintética o estimada.

4. Proceso de filtrado de de-énfasis. El filtro de de-énfasis puede ser definido como un sistema IIR (de respuesta infinita al impulso), que se muestra en la ecuación

$$H(z) = \frac{1}{1 - \alpha z^{-1}} \quad (6.22)$$

Donde α es el valor del polo en la función de transferencia. El filtro de de-énfasis tiene un comportamiento paso bajas, resaltando de esta forma las primeras frecuencias formantes. Se puede observar que el filtro de de-énfasis es un filtro inverso del filtro de pre-énfasis (proceso de análisis) por lo que es usual elegir el mismo valor de α . La ecuación en diferencias del filtro de de-énfasis es

$$y(n) = x(n) + \alpha y(n-1) \quad (6.23)$$

En el proceso de síntesis se utiliza el modelo de la figura 6.4 para producir voz sintética y se alimenta con los parámetros de la tabla 6.1, que puede estar almacenada en memoria o ser recibida por algún sistema, se va determinando el tipo de señal que debe alimentar al sistema, es decir *voceada* o *no-voceada*, detectar la generación de zonas de silencio, y con los parámetros a_i para implementar el sistema de producción de voz sintética.

Parámetros
Silencio (1/0)
Energía (G)
Voceada (1/0)
a_1
a_2
a_3
a_4
a_5
a_6
...
...
a_p

Tabla 6.1: Formato de salida de parámetros

6.2.1. Resultados de síntesis de voz

Como parte final de este trabajo se muestran las pruebas realizadas sobre un segmento de señal de voz real, estos resultados se obtuvieron en el trabajo [8], [5]. En la figura 6.6 se muestra la respuesta en frecuencia tanto para la voz real como para el modelo paramétrico de orden $p=8,10,12$. Se puede observar que la respuesta del modelo paramétrico es la envolvente del espectro verdadero y para $p \geq 12$ cambia muy poco respecto $p=10$, además en la medida que el orden p se hace más grande la carga computacional se hace mayor, por lo que un valor de $p=10$ es un modelo adecuado y utilizado en la práctica.

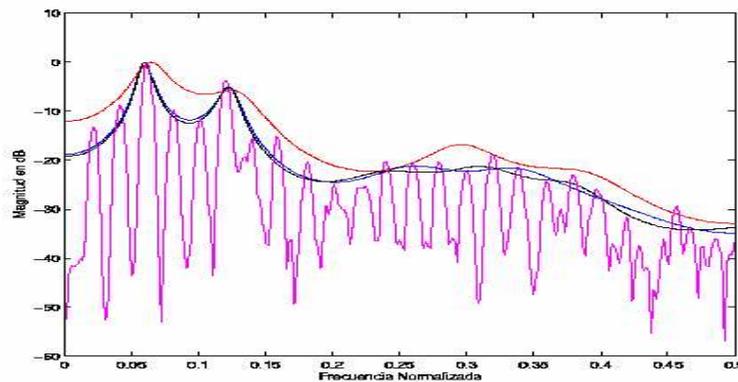


Figura 6.6: Espectro de voz verdadera y sintética

En la figura 6.7 se muestran las gráficas de las señales en el tiempo, en el inciso a) se tiene la señal real y en el inciso b) la señal sintética, se observa la similitud entre ambas señales.

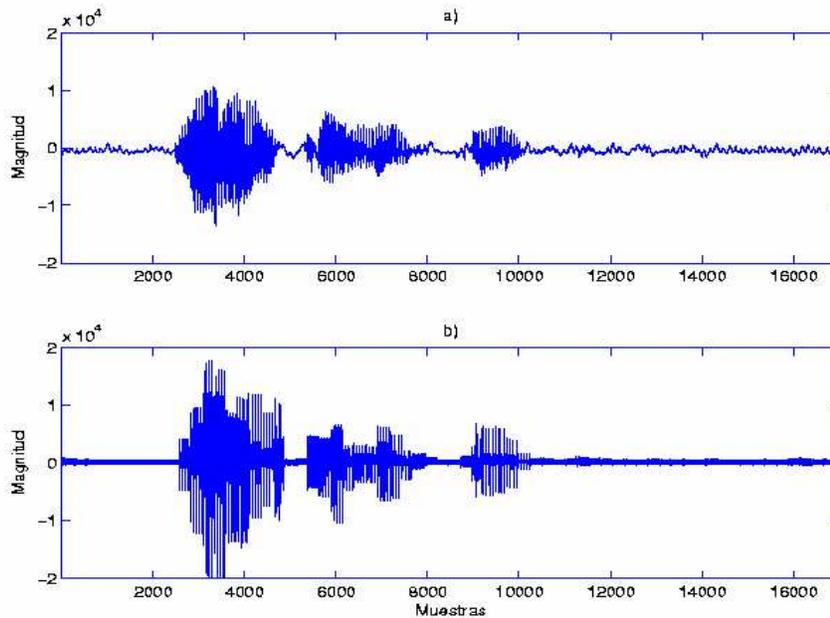


Figura 6.7: a) Señal de voz real. b) Señal de voz sintética

En la figura 6.8 se muestran las gráficas de los espectros de las señales de voz verdadera y voz sintética, de nuevo se aprecia la similitud de los resultados.

6.3. Resumen

Como se mencionó en la introducción a este trabajo que los filtros digitales son de vital importancia en los sistemas de procesamiento digital de señales, en la aplicación de voz sintética mostrada, se ha hecho un extenso uso de los filtros digitales, predicción lineal y los algoritmos abordados. Se puede resaltar en esta aplicación el buen desempeño de los resultados obtenidos, lo que verifica la importancia de la teoría expuesta durante este trabajo.

Se puede concluir que el filtrado digital es un área de aplicación en muchos campos de la ingeniería, por tanto el conocimiento, su estudio, análisis y técnicas de diseño son muy importantes para los ingenieros que trabajan en estos campos. No hay duda que existen muchas técnicas y algoritmos que se pueden seguir estudiando y generando para lograr mejores desempeños y tener nuevas aplicaciones.

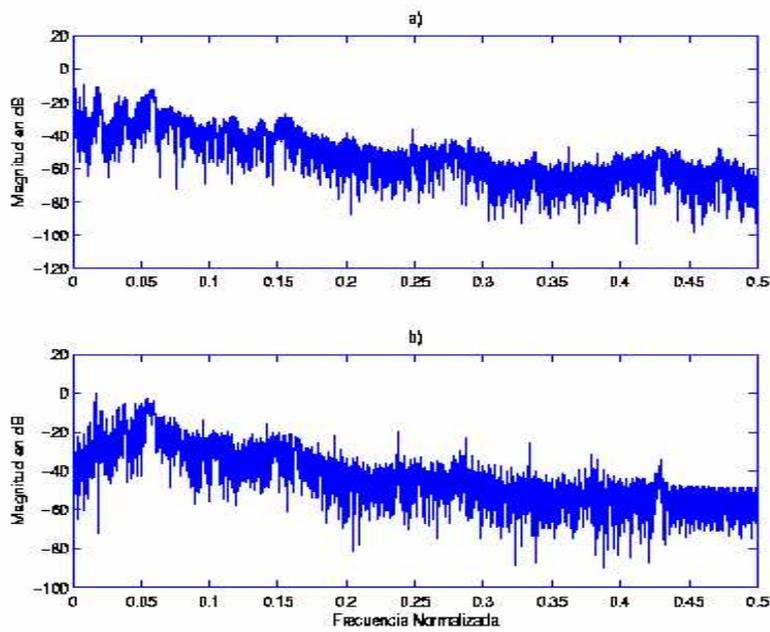


Figura 6.8: Comparación de espectros: a) voz verdadera y b) voz sintética

Apéndice A

Glosario

AR	autorregresivo
ARMA	auto regresivo de movimiento promedio
aliasing	existencia de frecuencias fantasmas o alias
backward	hacia atrás
bel	significa que la potencia cambia en un factor de 10
biquad	estructura de un filtro digital de segundo orden
BLP	predicción lineal hacia atrás
DFT	transformada discreta de Fourier
DFTI	transformada discreta de Fourier inversa
decibel	es la décima parte de un bel
DSP	procesadores digitales de señales
FA	filtro analógico
FD	filtro digital
FFT	transformada rápida de Fourier
FIR	respuesta finita al impulso
FLP	predicción lineal hacia adelante
forward	hacia adelante
FPB	filtro paso bajas
FPA	filtro paso altas
FPBW	filtro paso banda
FPSW	filtro supresor de banda
IIR	respuesta infinita al impulso
ladder	estructura escalera
lattice	celosía, enrejado, rejilla
LPC	codificación por predicción lineal
MSE	error cuadrático medio
PARCOR	partial correlation coefficients (in speech coding) o coeficientes de reflexión K_i

PDF	función de densidad de probabilidad
PDS	procesamiento digital de señales
pitch	se refiere al tono que existe en un tramo de señal
PSD	densidad espectral de potencia
NC	no causal
SLITD	sistema lineal e invariante en el tiempo discreto
T	período de muestreo
tap	retardo en un período T o etapa en un filtro FIR o TDL
TB	transformada bilineal
TDL	línea de retardo transversal
TF	transformada de Fourier
TL	transformada de Laplace
TZ	transformada Z
TZI	transformada Z inversa
VOCODER	codificador / decodificador de voz
WB	ancho de banda
WSS	proceso estacionario en el sentido amplio

Apéndice B

Nomenclatura

α	constante de retardo
$a(n)$	coeficientes a_i de un sistema IIR
$a_m(i)$	coeficientes i de un sistema AR de orden m
A_p	ganancia en la banda de paso
A_s	ganancia o atenuación en la banda suprimida
$b(n)$	coeficientes b_i de un sistema IIR
$d(n)$	señal deseada
δ	rizo
$e(n)$	error entre una señal deseada y una estimada
f_{ca}	frecuencia de corte analógica
f_H	frecuencia analógica de corte alta para FD FPBW o FSBW
f_L	frecuencia analógica de corte baja para FD FPBW o FSBW
f_s	frecuencia de muestreo
$h(t)$	respuesta al impulso de un sistema continuo
$h(n)$	respuesta al impulso de un sistema discreto
$h_I(n)$	respuesta al impulso de un filtro digital ideal
$H(s)$	función de transferencia de un sistema continuo
$H(z)$	función de transferencia de un sistema discreto
$H(\Omega)$	respuesta en frecuencia de un sistema continuo
$H(e^{j\omega})$	respuesta en frecuencia de un sistema discreto
$H_I(e^{j\omega})$	respuesta en frecuencia de un filtro ideal
$H(k)$	espectro discreto de un sistema discreto
$J(n)$	función de costo o energía del error
N	longitud de la respuesta al impulso $h(n)$
ω_{cd}	frecuencia de corte digital $2\pi f_{ca}/f_s$
ω_p	frecuencia de paso digital
ω_{sup}	frecuencia suprimida digital

$r_{xx}(i)$	autocorrelación de $x(n)$
τ_g	retardo de grupo
τ_p	retardo de fase
ω_H	frecuencia de corte alta para un FD FPBW o FSBW
ω_L	frecuencia de corte baja para un FD FPBW o FSBW
$W(n)$	función ventana
$\Phi(\omega)$	respuesta en fase de un sistema discreto
$x(n)$	señal de entrada de un sistema discreto
$y(n)$	señal de salida de un sistema discreto
$ \cdot $	valor absoluto
$E\{\cdot\}$	esperanza matemática
$(\cdot)^T$	transpuesta
$(\cdot)^*$	conjugado
$(\cdot)^H$	hermitiano (conjugado transpuesto)
$\ \mathbf{x}\ _p$	norma $l_p = (\sum_n x_n ^p)^{1/p}$
\mathbf{I}	matriz identidad
∇_w	gradiente respecto de los coeficiente w
$Re(\cdot)$	parte real
$Im(\cdot)$	parte imaginaria
$sgn(\cdot)$	operador signo, $sgn(x) = 1$ si $x \geq 0$ de lo contrario $sgn(x) = -1$
$\hat{x}(\cdot)$	valor estimado de $x(\cdot)$

Apéndice C

Demostración de la convolución

Por definición de convolución para un sistema entrada $x(n)$ y salida $y(n)$ con respuesta al impulso $h(n)$

$$\begin{aligned} y(n) &= x(n) * h(n) = h(n) * x(n) \\ y(n) &= \sum_{m=-\infty}^{\infty} x(m)h(n-m) = \sum_{m=-\infty}^{\infty} x(n-m)h(m) \end{aligned} \quad (\text{C.1})$$

por definición de transformada Z , $Y(z) = TZ(y(n))$

$$Y(z) = \sum_{n=-\infty}^{\infty} y(n)z^{-n} \quad (\text{C.2})$$

sustituyendo la ecuación (C.1) en la ecuación (C.2)

$$Y(z) = \sum_{n=-\infty}^{\infty} \left(\sum_{m=-\infty}^{\infty} x(m)h(n-m) \right) z^{-n} \quad (\text{C.3})$$

utilizando una variable auxiliar $i = n - m$, $n = i + m$, entonces

$$Y(z) = \sum_{n=-\infty}^{\infty} \left(\sum_{m=-\infty}^{\infty} x(m)h(i) \right) z^{-i} z^{-m} \quad (\text{C.4})$$

reagrupando variables

$$Y(z) = \sum_{i=-\infty}^{\infty} h(i)z^{-i} \sum_{m=-\infty}^{\infty} x(m)z^{-i} = H(z)X(z) \quad (\text{C.5})$$

se comprueba que la primera sumatoria de la última ecuación corresponde a $H(z)$

$$H(z) = \sum_{i=-\infty}^{\infty} h(i)z^{-i} \quad (\text{C.6})$$

quedando demostrado el teorema de la convolución.

Apéndice D

Espectro de la ventana de Hamming

Para la ventana de Hamming en su forma causal

$$W(n) = \{0.54 - 0.46 \cos(\frac{2\pi n}{N-1})\} \{U(n) - U(n-N)\}$$

expresando la función coseno por Euler como

$$\cos(\frac{2\pi n}{N-1}) = \frac{e^{\frac{j2\pi n}{N-1}} + e^{-\frac{j2\pi n}{N-1}}}{2}$$

Aplicando transformada Z

$$W(z) = 0.54 \frac{1-z^{-N}}{1-z^{-1}} - 0.23 \frac{1-z^{-N}}{1-e^{\frac{j2\pi}{N-1}}z^{-1}} - 0.23 \frac{1-z^{-N}}{1-e^{-\frac{j2\pi}{N-1}}z^{-1}} \quad (D.1)$$

evaluando $z = e^{j\omega}$

$$W(e^{j\omega}) = 0.54 \frac{1-e^{-j\omega N}}{1-e^{-j\omega}} - 0.23 \frac{1-e^{-j\omega N}}{1-e^{\frac{j2\pi}{N-1}}e^{j\omega}} - 0.23 \frac{1-e^{-j\omega N}}{1-e^{-\frac{j2\pi}{N-1}}e^{j\omega}} \quad (D.2)$$

realizando factorizaciones se obtiene

$$W(e^{j\omega}) = 0.54 \frac{e^{-\frac{j\omega N}{2}} [e^{\frac{j\omega N}{2}} - e^{-\frac{j\omega N}{2}}]}{e^{-\frac{j\omega}{2}} [e^{\frac{j\omega}{2}} - e^{-\frac{j\omega}{2}}]} - 0.23 \frac{e^{-\frac{j\omega N}{2}} [e^{\frac{j\omega N}{2}} - e^{-\frac{j\omega N}{2}}]}{e^{-j(\omega - \frac{2\pi}{N-1})} [e^{j(\omega - \frac{2\pi}{N-1})} - e^{-j(\omega - \frac{2\pi}{N-1})}]} \\ - 0.23 \frac{e^{-\frac{j\omega N}{2}} [e^{\frac{j\omega N}{2}} - e^{-\frac{j\omega N}{2}}]}{e^{-j(\omega + \frac{2\pi}{N-1})} [e^{j(\omega + \frac{2\pi}{N-1})} - e^{-j(\omega + \frac{2\pi}{N-1})}]}$$

de nuevo utilizando Euler para regresar a funciones seno

$$W(e^{j\omega}) = 0.54 \frac{\text{sen}(\frac{\omega N}{2})}{\text{sen}(\frac{\omega}{2})} e^{-\frac{j\omega(N-1)}{2}} - 0.23 \frac{\text{sen}(\frac{\omega N}{2})}{\text{sen} \frac{1}{2}(\omega - \frac{2\pi}{N-1})} e^{-j[\frac{\omega}{2}(N-1) + \frac{2\pi}{N-1}]} \quad (D.3)$$

$$-0.23 \frac{\text{sen}(\frac{\omega N}{2})}{\text{sen} \frac{1}{2}(\omega + \frac{2\pi}{N-1})} e^{-j[\frac{\omega}{2}(N-1) - \frac{2\pi}{N-1}]}$$

entonces la magnitud del espectro queda

$$|W(e^{j\omega})| = 0.54 \frac{\text{sen}(\frac{\omega N}{2})}{\text{sen}(\frac{\omega}{2})} + 0.23 \frac{\text{sen}(\frac{\omega N}{2})}{\text{sen} \frac{1}{2}(\omega - \frac{2\pi}{N-1})} + 0.23 \frac{\text{sen}(\frac{\omega N}{2})}{\text{sen} \frac{1}{2}(\omega + \frac{2\pi}{N-1})} \quad (\text{D.4})$$

Para una ventana centrada en $N/2$ o $N \gg 1$, la magnitud del espectro es

$$|W(e^{j\omega})| = 0.54 \frac{\text{sen}(\frac{\omega N}{2})}{\text{sen}(\frac{\omega}{2})} + 0.23 \frac{\text{sen}(\frac{\omega N}{2})}{\text{sen} \frac{1}{2}(\omega - \frac{2\pi}{N})} + 0.23 \frac{\text{sen}(\frac{\omega N}{2})}{\text{sen} \frac{1}{2}(\omega + \frac{2\pi}{N})} \quad (\text{D.5})$$

Si definimos la función

$$S(\omega) = \frac{\text{sen}(\omega N/2)}{\text{sen}(\omega/2)}$$

el espectro de la ventana de Hamming se puede resumir

$$|W(e^{j\omega})| = 0.54S(\omega) + 0.23S(\omega - 2\pi/N) + 0.23S(\omega + 2\pi/N) \quad (\text{D.6})$$

El primer valor $|W(e^{j\omega})| = 0$ ocurre cuando $\frac{4\pi}{N-1}$, por tanto el ancho del lóbulo principal para $N \gg 1$ es $\frac{8\pi}{N}$.

Apéndice E

Algoritmo de Levinson-Durbin

El método de Levinson-Durbin es uno de los algoritmos recursivos más importantes en el procesamiento digital de señales [1], [9]. Este algoritmo es de tipo recursivo y nos da una solución de la ecuación normal del orden $O(p^2)$ operaciones, el algoritmo explota las propiedades de la matriz \mathbf{R}_p y calcula la solución de la ecuación normal al orden $p + 1$ a partir de la solución al orden p .

La eliminación gaussiana para resolver un sistema de p ecuaciones lineales con p incógnitas requiere aproximadamente $p^3/3$ multiplicaciones y divisiones, el algoritmo de Levinson-Durbin requiere de $p^2 + 2p$ multiplicaciones y divisiones, y p^2 sumas. Otra ventaja de Levinson-Durbin es que requiere $2(p + 1)$ localidades de memoria, $(p + 1)$ para la secuencia de autocorrelación y $p + 1$ localidades para los coeficientes a_i , mientras que la eliminación gaussiana requiere p^2 localidades de memoria [16].

La eliminación gaussiana no toma en cuenta la simetría de la matriz \mathbf{R}_p ($r_{xx}(i) = r_{xx}(-i)$), esta matriz además de ser simétrica, los elementos de las diagonales son iguales y se le conoce como matriz tipo *Toeplitz*. Por lo que existe un método recursivo propuesto por Levinson y Durbin que explota la estructura de la matriz Toeplitz.

Existen varias formas de llegar a obtener este algoritmo, tal como se mostró en el capítulo 4, en esta sección se mostrarán otras dos posibilidades de llegar al algoritmo.

Expresando en forma particionada el vector \mathbf{a}

$$\mathbf{a}_p = \begin{bmatrix} a_1^p \\ a_2^p \\ \vdots \\ a_p^p \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{p-1} \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbf{d}_{p-1} \\ K_p \end{bmatrix} \quad (\text{E.1})$$

donde

\mathbf{a}_p : es el vector de coeficientes del filtro de predicción de error al orden p

\mathbf{a}_{p-1} : es el vector de coeficientes del filtro de predicción de error al orden $p - 1$

\mathbf{d}_{p-1} : vector de coeficientes por determinar al orden $p - 1$

K_p : constante escalar por determinar

y la matriz \mathbf{R}_p

$$\mathbf{R}_p = \begin{bmatrix} \mathbf{R}_{p-1} & \mathbf{r}_{p-1}^r \\ \mathbf{r}_{p-1}^{rT} & r_{xx}(0) \end{bmatrix} \quad (\text{E.2})$$

sustituyendo en la ecuación normal $\mathbf{R}_p \mathbf{a}_p = \mathbf{r}_p$

$$\begin{bmatrix} \mathbf{R}_{p-1} & \mathbf{r}_{p-1}^r \\ \mathbf{r}_{p-1}^{rT} & r_{xx}(0) \end{bmatrix} \left(\begin{bmatrix} \mathbf{a}_{p-1} \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbf{d}_{p-1} \\ k_p \end{bmatrix} \right) = \begin{bmatrix} \mathbf{r}_{p-1} \\ r_{xx}(p) \end{bmatrix} \quad (\text{E.3})$$

el superíndice r en el vector \mathbf{r}_{p-1}^r significa el vector escrito en orden inverso. De la primera fila de la ecuación anterior (E.3) se obtiene

$$\mathbf{R}_{p-1} \mathbf{a}_{p-1} + \mathbf{R}_{p-1} \mathbf{d}_{p-1} + K_p \mathbf{r}_{p-1}^r = \mathbf{r}_{p-1}$$

por la ecuación normal, el primer término es igual al del lado derecho, despejando \mathbf{d}_{p-1}

$$\mathbf{d}_{p-1} = -K_p \mathbf{R}_{p-1}^{-1} \mathbf{r}_{p-1}^r$$

que es el vector $\mathbf{a}_m(i)$ en orden inverso multiplicado por una constante

$$\mathbf{d}_{p-1} = -K_p \begin{bmatrix} a_{p-1}^{p-1} \\ a_{p-2}^{p-1} \\ \vdots \\ a_1^{p-1} \end{bmatrix} \quad (\text{E.4})$$

La segunda fila de la ecuación (E.3)

$$\mathbf{r}_{p-1}^{rT} \mathbf{a}_{p-1} + \mathbf{r}_{p-1}^{rT} \mathbf{d}_{p-1} + K_p r_{xx}(0) = r_{xx}(p)$$

sustituyendo en esta última ecuación el vector \mathbf{d}_{p-1} y despejando la constante K_p

$$\begin{aligned} K_p &= \frac{r_{xx}(p) - \mathbf{r}_{p-1}^{rT} \mathbf{a}_{p-1}}{r_{xx}(0) - \mathbf{r}_{p-1}^{rT} \mathbf{R}_{p-1}^{-1} \mathbf{r}_{p-1}^r} \\ K_p &= \frac{r_{xx}(p) - \mathbf{r}_{p-1}^{rT} \mathbf{a}_{p-1}}{r_{xx}(0) - \mathbf{r}_{p-1} \mathbf{a}_{p-1}} \\ K_p &= \frac{r_{xx}(p) - \mathbf{r}_{p-1}^{rT} \mathbf{a}_{p-1}}{\mathcal{E}_{p-1}} \end{aligned} \quad (\text{E.5})$$

donde \mathcal{E}_{p-1} es el error promedio cuadrático MSE dado como

$$\mathcal{E}_{p-1} = r_{xx}(0) - \mathbf{a}_{p-1} \mathbf{r}_{p-1} \quad (\text{E.6})$$

Sustituyendo \mathbf{d}_{p-1} en \mathbf{a}_p , se obtiene la expresión escalar recursiva

$$a_i^j = a_i^{(j-1)} + K_p a_{(j-i)}^{(j-1)} \quad j = 1, 2, \dots, p-1 \quad i = 1, 2, \dots, p-1 \quad (\text{E.7})$$

donde

$$a_j^j = k_j$$

el valor mínimo del MSE, puede calcularse recursivamente

$$\mathcal{E}_j = r_{xx}(0) - \sum_{i=1}^p a_i^j r_{xx}(i)$$

utilizando la ecuación (E.7) de los coeficientes a_i^j

$$\mathcal{E}_j = r_{xx}(0) - \sum_{i=1}^{p-1} a_i^{(j-1)} r_{xx}(i) - a_i^j \left[r_{xx}(0) - \sum_{i=1}^{p-1} a_{(i-j)}^{(j-1)} r_{xx}(i) \right]$$

es decir, que \mathcal{E}_j se puede calcular en forma recursiva

$$\mathcal{E}_j = \mathcal{E}_{j-1} - a_i^{(j)2} \mathcal{E}_{j-1} = \mathcal{E}_{j-1} (1 - a_i^{(j)2}) \quad (\text{E.8})$$

Ecuación de Levinson-Durbin utilizando variable z

Para los coeficientes $a_{(i)}^m$ de una estructura TDL de orden 0, se tiene

$$A_0(z) = B_0(z) = 1 \quad (\text{E.9})$$

para un orden m se obtiene una relación entre polinomios hacia adelante $A_m(z)$ y hacia atrás $B_m(z)$

$$B_m(z) = z^{-m} A_m(z) \quad (\text{E.10})$$

$$A_m(z) = A_{m-1}(z) + K_m z^{-1} B_{m-1}(z) \quad (\text{E.11})$$

donde $m = 1, 2, 3, 4, \dots, M-1$

Aplicando la definición de transformada Z

$$\sum_{i=0}^m a_m(i) z^{-i} = \sum_{i=0}^{m-1} a_{m-1}(i) z^{-i} + K_m \sum_{i=0}^{m-1} a_{m-1}(m-1-i) z^{-(i+1)} \quad (\text{E.12})$$

Efectuando las sumatorias en ambos lados de la ecuación y comparando coeficientes de las respectivas potencias en z , y recordando que $a_m(0) = 1$, se tienen las constantes

$$a_{(0)}^m = a_{(0)}^{m-1} = 1 \quad (\text{E.13})$$

$$K_m = a_{(m)}^m \tag{E.14}$$

$$a_{(i)}^m = a_{(i)}^{m-1} + K_m a_{(m-i)}^{m-1} \tag{E.15}$$

$$a_{(i)}^m = a_{(i)}^{m-1} + a_{(m)}^m a_{(m-i)}^{m-1} \tag{E.16}$$

para

$$m = 1, 2, 3, \dots, M - 1 \quad y \quad 1 \leq i \leq m - 1$$

En estos análisis se ha comprobado de forma diferente la recursión de Levinson-Durbin.

Apéndice F

Algoritmo de Leroux - Gueguen

El algoritmo de Levison-Durbin es eficiente para determinar los coeficientes LPC $a_m(i)$, pero no muy adecuado para el cálculo en aritmética de punto fijo, debido a que el intervalo dinámico de los coeficientes $a_m(i)$ es muy grande.

El algoritmo de Leroux - Gueguen calcula los coeficientes K_i sin necesidad de calcular los coeficientes $a_m(i)$, este algoritmo utiliza únicamente los valores de correlación y efectúa recursión sobre los coeficientes [1], [29].

Considerando el parámetro C_j^m de orden m

$$C_j^m = \sum_{i=0}^m a(i)^m r_{xx}(j-i) \quad (\text{F.1})$$

donde $C_j^m \leq r_{xx}(0)$.

El algoritmo de Leroux - Gueguen relaciona el parámetro C_i^m y las variables del algoritmo Levison-Durbin

$$K_m = \frac{C_m^{m-1}}{C_0^{m-1}} \quad m = 1, 2, 3, \dots, p \quad (\text{F.2})$$

introduciendo los parámetros LPC y desarrollando el parámetro C_i^{m-1} de orden $m-1$ se obtiene

$$\begin{aligned} C_j^m &= \sum_{i=0}^{m-1} [a(i)^{m-1} + K_m a_{(m-i)}^{m-1}] r_{xx}(m-i) + K_m r_{xx}(m-i) \\ &= \sum_{i=0}^{m-1} [a(i)^{m-1} + K_m \sum_{i=0}^{m-1} a_{(i)}^{m-1}] r_{xx}(i+j-m) \end{aligned}$$

se puede escribir

$$C_j^m = C_j^{m-1} + K_m C_{m-j}^{m-1} \quad (\text{F.3})$$

Esta es una forma recursiva de calcular los parámetros C_j^m al orden m con base a los parámetros al orden $m-1$, donde $C_j^0 = r_{xx}(j)$. Los parámetros deben ser calculados al orden m para el intervalo $j = 0, 1, 2, 3, \dots, m$

Para la predicción de **orden** $m = p$

$$K_p = \frac{C_p^{p-1}}{C_0^{p-1}} \quad (\text{F.4})$$

se necesita C_0^{p-1} y C_p^{p-1} .

Para la predicción de **orden** $m = p$, hay que resolver

$$K_{p-1} = \frac{C_{p-1}^{p-2}}{C_0^{p-2}}$$

$$C_p^{p-1} = C_p^{p-2} + K_{p-1}C_{p-1}^{p-2}$$

$$C_0^{p-1} = C_0^{p-2} + K_{p-1}C_{p-1}^{p-2} \quad (\text{F.5})$$

se necesita los parámetros C_{-1}^{p-2} , C_0^{p-2} , C_{p-1}^{p-2} y C_p^{p-2}

Resumen del algoritmo para un orden p

- Inicializar con el orden $m = 0$, entonces

$$C_j^0 = r_{xx}(j) \quad j = -p + 1, -p + 2, \dots, p$$

- Efectuar la recursión sobre $l = 1, 2, \dots, p$

- **Paso 1:** calcular el coeficiente K_l con

$$K_l = \frac{C_l^{l-1}}{C_0^{l-1}}$$

si $l = p$ finalizar

- **Paso 2:** calcular el coeficiente C con la ecuación recursiva

$$C_j^l = C_j^{l-1} + K_l C_{l-j}^{l-1} \quad j = -p + l + 1, \dots, 0, l + 1, \dots, p$$

fijar $l = l + 1$ y regresar a paso 1

Bibliografía

- [1] CHU W. C. *Speech coding algorithms, foundation and evolution of standardized coders*. John Wiley & Sons, San Jose, California, USA 2003.
- [2] CHURCHIL R. *Series de Fourier y problemas de contorno*. McGraw-Hill, México 1977.
- [3] DEFATTA D. J., LUCAS J. G. & HODGKISS W. S. *Digital Signal Processing, A System Design Approach*. John Wiley & Sons. Singapore 1988.
- [4] DELLER J. R. JR., PROAKIS J. G. & HANSEN J. H. *Discrete-Time Processing Speech Signals*. Prentice-Hall, USA 1987.
- [5] ESCOBAR S. L. Y PASALAGUAS P. A. *Síntesis de voz en tiempo real utilizando una arquitectura de punto flotante*. XIX Congreso Internacional Académico de Ingeniería, Instituto Tecnológico de Chihuahua, Chihuahua, octubre de 1997.
- [6] ESCOBAR S. L. Y CORTÉS S. F. *Fundamentos de Teoría Electromagnética*. Facultad de Ingeniería, UNAM, México D. F., junio del 2001. 243 pags.
- [7] ESCOBAR S. L. Y MOLERO A. M. *Sintetizador de voz en tiempo real utilizando el DSP TMS320C5402*. XXVII Congreso de Ingeniería Electrónica, Instituto Tecnológico de Chihuahua, Chihuahua, Oct. 2005.
- [8] ESCOBAR S. L., PSENICKA B. Y MOLERO A. M. *Arquitecturas de DSPs, familias TMS32054x y el TMS320C54xx, y aplicaciones*. Facultad de Ingeniería, UNAM, México D. F., Octubre del 2005. 190 pags.
- [9] EDITED BY ELLIOT F. D. *Handbook of Digital Signal Processing Engineering Applications*. Academic Press, USA 1987.
- [10] FLIEGE N. J. *Multirate Digital Signal Processing*. John Wiley & Sons, New York 1994.
- [11] FURUI S. *Digital Speech Processing, Synthesis, and Recognition*. Marcel Dekker, New York 2001.
- [12] HAMMING R. W. *Digital Filters*. Second Edition, Prentice-Hall Englewood Cliffs, New Jersey 1983.
- [13] HAYKIN S. EDITED BY. *Nonlinear Methods of Spectral Analysis Modern Filters*. Springer-Verlag. Berlin, New York, Tokio 1986.

- [14] HAYKIN S. *Modern Filters*. MacMillan, New York 1989.
- [15] HAYKIN S. *Adaptive Filter Theory*. Englewood Cliffs, NJ Prentice-Hall, 1991.
- [16] HAYES M. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, USA 1996.
- [17] HSU HWEI, *Análisis de Fourier*. Addison-Wesley Iberoamericana. USA 1987.
- [18] HARRIS F. *On the use of windows for harmonics analysis with th Discrete Fourier Transform*. Proceeding of IEEE, Vol. 66 No. 1 Jan 1978.
- [19] INGLE V. K. & PROAKIS J. G. *Digital Signal Processing Using Matlab V.4*. PWS Publishing Company. USA 1997.
- [20] INRIA SCILAB GROUP *Signal Processing with SCILAB*. PWS Publishing Company. USA 1997.
- [21] KUO S. & LEE B. *Real Time Digital Signal Processing Algorithms, implementation, applications and Experiments with th TMS320C55x*. John Wiley & Sons, England 2001.
- [22] MARKEL J. D. & GRAY A. H. *Linear prediction of speech*. Springer Verlag, New York, 1976.
- [23] MAKHOUL J. *Linear prediction: A tutorial review*. Proc. IEEE 63, pp. 561-580, 1975.
- [24] MITRA S. K. *Digital Signal Processing. A Cumputer-Based Approach..* Second Edition, McGraw-Hill, New York 2001.
- [25] MOON T. K. & STIRLING W. C. *Mathematical Methods and Algorithms for Signal Processing*. Prentice-Hall, New Jersey 2000.
- [26] OPPENHEIM A. V., SCHAFER R. W. & BUCK J. R. *Discrete time Signal Processing*. Prentice-Hall, USA 1999.
- [27] ORFANIDIS S. J. *Introduction to Signal Processing*. Prentice-Hall, New Jersey 1996.
- [28] OWENS F. J. *Signal Processing of Speech*. McGraw-Hill Inc, USA 1993.
- [29] PAPAMICHALIS E. P. *Practical Approaches to Speech Coding*. Prentice - Hall and Texas Instrument Digital Signal Processor Series. New Jersey 1987.
- [30] PAPOULIS A. *Probability, Random Variables and Stochastic Processes*. MacGraw-Hill, Singapore 1991.
- [31] PROAKIS J., LIN R. C. AND NIKIAS C. *Advanced Digital Signal Processing*. Macmillan-Maxwell, Ontario Canada 1992.
- [32] PROAKIS J.G. & MANOLAKIS. *Digital Signal Processing, Algorithms and Applications*. Macmillan. Singapore 1992.
- [33] PROAKIS J. G. & MANOLAKIS. *Tratamiento digital de señales. Principios, algoritmos y aplicaciones*, Prentice Hall, tercera edición, Madrid, España 1998.

- [34] PSENICKA B. Y ESCOBAR S. L. *Procesamiento Digital de Señales, segunda parte, Microcontroladores y realización de los filtros digitales con TMS320Cxx*. Facultad de Ingeniería, UNAM, México D.F., julio de 1998.
- [35] RABINER L. R. & SCHAFER R. W. *Digital Processing of Speech Signal*. Prentice-Hall, 1975.
- [36] SMITH S. W. *The Scientist and engineer's guide to digital signal processing*. 2d. Edition, California Technical Publishing. San Diego California, USA 1999.
- [37] SPIEGEL M. R & ABELLANAS L. *Fórmulas y tablas de matemática aplicada*. McGraw-Hill, México 1988.
- [38] TAYLOR F. J. *Digital Filter Design Handbook* . Ed. Marcel Dekker Inc. New York 1983.
- [39] THERRIEN C. W. *Discrete random signal and statistical signal processing*. Prentice-Hall, New Jersey 1992.
- [40] TEXAS INSTRUMENTS *TMS320C54x Reference Set, Vols. 1: CPU and Peripherals, 2: Mnemonic Instruction Set, 3: Algebraic Instruction Set & 4: Applications Guide*. USA 1997.
- [41] WIDROW B. AND STEARNS S. D. *Adaptive Signal Processing*. Englewood Cliffs, Prentice-Hall, NJ. 1985.
- [42] WIDROW B. AND WALACH E. *Adaptive Inverse Control*. Prentice-Hall, NJ. 1996.
- [43] WUNSCH A. D. *Variable compleja con aplicaciones*. Pearson Educación, Addison Wesley Iberoamericana, México 1997.