

# DISEÑO Y SIMULACIÓN DE FILTROS DIGITALES TIIR

Augusto Riedinger<sup>1</sup>

<sup>1</sup>Universidad Tecnológica Nacional, Facultad Regional de Bahía Blanca, Departamento de Electrónica - Bahía Blanca, Argentina  
email: ariedinger4@duck.com<sup>1</sup>

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

9 de agosto de 2022

*Index Terms*—Filtros digitales, Filtros TIIR, Filtros FIR, Filtros TIIR.

## I. INTRODUCCIÓN

UN filtro es un sistema, que dependiendo de algunos parámetros, realiza un proceso de discriminación de una señal de entrada obteniendo variaciones en su salida. Los filtros digitales tienen como entrada una señal digital y a su salida tienen otra señal digital, pudiendo haber cambiado en amplitud, frecuencia o fase dependiendo de las características del filtro.

El filtrado digital es parte del procesamiento de señal digital. Se le da la denominación de digital más por su funcionamiento interno que por su dependencia del tipo de señal a filtrar, así podríamos llamar filtro digital tanto a un filtro que realiza el procesamiento de señales digitales como a otro que lo haga de señales analógicas.

El filtrado digital consiste en la realización interna de un procesamiento de datos de entrada. El valor de la muestra de la entrada actual y algunas muestras anteriores (que previamente habrían sido almacenadas) son multiplicadas, por unos coeficientes definidos. También podría tomar valores de la salida en instantes pasados y multiplicarlos por otros coeficientes. Finalmente todos los resultados de todas estas multiplicaciones son sumados, dando una salida para el instante actual. Esto implica que internamente tanto la salida como la entrada del filtro serán digitales, por lo que puede ser necesario una conversión analógico-digital o digital-analógico para uso de filtros digitales en señales analógicas.

Los filtros digitales se usan frecuentemente para tratamiento digital de la imagen o para tratamiento del sonido digital.

### I-A. Tipos de filtros

Hay varios tipos de filtros así como distintas clasificaciones para estos filtros:

- De acuerdo con la parte del espectro que dejan pasar y atenúan hay:
  - 
  - Filtros pasa alto.
  - Filtros pasa bajo.

- Filtros pasa banda.
- Banda eliminada.
- Multibanda.
- Pasa todo.
- Resonador.
- Oscilador.
- Filtro peine (Comb filter).
- Filtro ranura o filtro rechaza banda (Notch filter).

- De acuerdo con su orden:
  - Primer orden.
  - Segundo orden.
- De acuerdo con el tipo de respuesta ante la entrada unitaria:
  - FIR (Respuesta Finita al Impulso o /Finite Impulse Response/).
  - IIR (Respuesta Infinita al Impulso o /Infinite Impulse Response/).
  - TIIR (Respuesta Infinita Truncada al Impulso o /Truncated Infinite Impulse Response/).
- De acuerdo con la estructura con que se implementa:
  - Directa.
  - Transpuesta.
  - Cascada.
  - Fase lineal.
  - Lattice.

### I-B. Comparación entre los distintos tipos de filtros

Los filtros IIR son ampliamente utilizados debido a su bajo costo computacional. Los filtros FIR, en cambio, permiten la posibilidad de implementar filtros lineales digitales con un retraso de grupo constante para todas las frecuencias. La contrapartida es que, para alcanzar funciones de transferencia de magnitud similar, los filtros FIR requieren un orden mucho mayor que su contraparte IIR. Por ejemplo, La contrapartida es que, para alcanzar funciones de transferencia de magnitud similar, los filtros FIR requieren un orden mucho mayor que su contraparte IIR. Por ejemplo, un filtro general FIR de orden  $N$  requieren  $N + 1$  múltiplos y  $N$  sumandos.

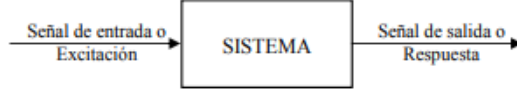


Figura 1: Esquema del sistema, señal de entrada y respuesta o salida del sistema

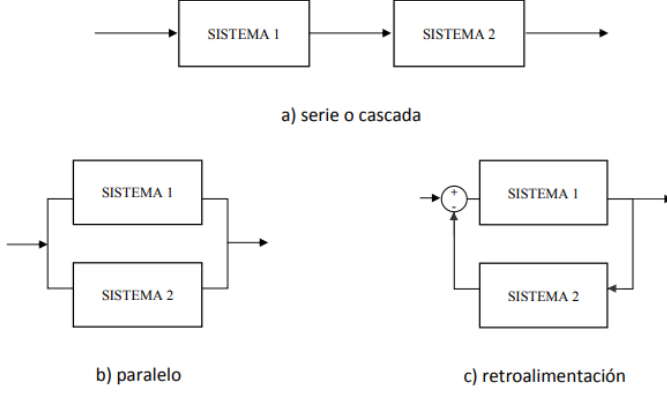


Figura 2: Interconexión de sistemas

En ciertos casos, sin embargo, es posible diseñar filtros FIR con costo computacional comparable al de los filtros IIR mientras se mantienen las ventajas que proporcionan los filtros FIR. Este tipo de filtros FIR puede ser implementado de forma eficiente mediante secuencias truncadas de bajo orden de filtros IIR. A este tipo de filtros se los conoce como filtro TIIR.

### I-C. Expresión general de un filtro digital

En muchas aplicaciones del procesamiento de señales es necesario diseñar dispositivos o algoritmos que realicen operaciones sobre las señales y que los englobaremos bajo la denominación genérica de sistemas.

Un sistema opera sobre una señal de entrada o excitación según una regla preestablecida, para generar otra señal llamada salida o respuesta del sistema a la excitación propuesta y que puede simbolizarse

$$y[n] = T(x[n]) \quad (1)$$

donde  $T$  simboliza la transformación, operador o procesamiento realizado por el sistema sobre la señal  $x$  para producir la señal  $y$  (ver Fig. 1). Una de las motivaciones más fuertes para el desarrollo de herramientas generales para el análisis y diseño de sistemas es que proviniendo a menudo de aplicaciones muy diferentes tienen descripciones matemáticas similares.

Existen varias maneras de representar un sistema, ya que muchos sistemas reales están contruidos como interconexiones de varios subsistemas, tal como se grafica en la Fig. 2.

Existen dos métodos básicos para el análisis del comportamiento o respuesta de un sistema lineal a una determinada entrada. Un primer camino se basa en obtener la solución de la ecuación entrada-salida del sistema que en general tiene la forma de las ecuaciones en diferencias lineales a coeficientes constantes  $a_m, b_k$

$$\sum_{m=0}^{N_a-1} a_m y[n-m] = \sum_{k=0}^{N_b-1} b_k x[n-k] \quad (2)$$

siendo  $N_a$  y  $N_b$  los órdenes máximos de las diferencias en la ecuación correspondiente a la entrada y a la salida del sistema.

El segundo método para el análisis del comportamiento del sistema reside en la aplicación del principio de superposición y consiste en descomponer la señal de entrada en una suma pesada de señales elementales las cuales se escogen de manera que sea conocida la respuesta del sistema a las mismas. Siguiendo esta línea, una señal a tiempo discreto puede visualizarse como una secuencia pesada de impulsos unitarios

$$x[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot \delta[n-k] \quad (3)$$

Aplicando la propiedad de superposición de los SLIT (Sistemas Lineales e Invariantes en el Tiempo) (Oppenheim y Willsky, 1983), se puede determinar la salida del sistema ante una cierta entrada de la siguiente manera

$$y[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot h[n-k] \quad (4)$$

siendo  $h[n]$  la respuesta o salida del sistema ante una entrada equivalente a un impulso unitario  $\delta[n]$  denominada /respuesta al impulso del sistema/. El segundo miembro de la expresión representa el producto de convolución de la señal de entrada  $x[n]$  y la respuesta al impulso del sistema  $h[n]$ ; esto es

$$y[n] = x[n] * h[n] = h[n] * x[n] \quad (5)$$

Tanto en el caso continuo como en el caso discreto, la respuesta al impulso del sistema LTI presenta las siguientes propiedades

- Sin memoria:  $h[n] = 0$  para  $n \neq 0$ .
- Causal:  $h[n] = 0$  para  $n < 0$ .
- Invertible: dado  $h[n] \exists h'[n] : h[n] * h'[n] = \delta[n]$ .
- Estable:  $\sum_{k=-\infty}^{\infty} |h[k]| < \infty$

Existen otras formas de representar un filtro, todas estas equivalentes a la respuesta al impulso unitario de sistema SLIT, sin embargo muchas veces conviene más una u otra representación. En el caso aplicar la transformada Z, a la 2 se obtiene la función de transferencia del sistema (Oppenheim y Willsky, 1983; Proakis y Manolakis; 1996; Oppenheim y Schaffer, 1999)

$$H(z) = \frac{\sum_{k=0}^{N_b-1} b_k z^{-k}}{\sum_{m=0}^{N_a-1} a_m z^{-m}} \quad (6)$$

donde  $z = A \exp(j\Omega)$  es la variable compleja en forma polar. Particularmente si el modulo  $A = 1$ , la expresión de la Ec. 11 se reduce a la respuesta en frecuencia del sistema a través de la transformada de Fourier a tiempo discreto (Oppenheim y Willsky, 1983; Proakis y Manolakis; 1996; Oppenheim y Schaffer, 1999):

$$y[n] = \sum_{k=0}^{N_b-1} b_k x[n-k] - \sum_{m=1}^{N_a-1} a_m y[n-m] \quad (7)$$

donde los coeficientes  $a_m$  y  $b_k$  son los coeficientes que definen el filtro, por lo tanto el diseño consiste en calcularlos. Como regla general se suele dejar el término  $a_0 = 1$ .

#### I-D. Definición de filtros FIR e IIR

Un filtro FIR causal de orden  $N$  convencional se puede representar según:

$$y[n] = \sum_{k=0}^N h_k x[n-k] \quad (8)$$

Donde la función de transferencia tiene la siguiente forma:

$$H_{FIR}(z) = h_0 + h_1 z^{-1} + \dots + h_N z^{-N} = z^{-N} C(z) \quad (9)$$

Y se define a  $C(z)$  como el polinomio de orden  $N$  formado por los coeficientes  $h_k$ .

En cambio, un filtro IIR causal de orden  $P$  tiene la relación:

$$y[n] = - \sum_{k=1}^P a_k y[n-k] + \sum_{l=0}^P b_l x[n-l] \quad (10)$$

Con la función transferencia correspondiente es:

$$H_{IIR}(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_P z^{-P}}{1 + a_1 z^{-1} + \dots + a_P z^{-P}} = \frac{B(z)}{A(z)} \quad (11)$$

Por tanto,  $A(z)$  y  $B(z)$  serán:

$$A(z) = z^P + a_1 z^{P-1} + \dots + a_P \quad (12)$$

$$B(z) = b_0 z^P + b_1 z^{P-1} + \dots + b_P \quad (13)$$

Y pueden ser asumidos como los polinomios de grado  $P$  en  $z$ .

El retardo de grupo se define como:

$$\tau_d(\omega) = - \frac{d \arg\{H(e^{j\omega})\}}{d\omega} \quad (14)$$

El retardo de grupo a frecuencia normalizada  $\omega = 2\pi f/f_s$ ; donde  $f_s$  es la frecuencia de muestreo; es el número de muestras con retraso experimentadas por la amplitud de la envolvente de banda ancha centrada en  $\omega$ .

Un filtro de fase lineal es aquel cuya respuesta en fase a una determinada frecuencia es una función lineal de la frecuencia; esto es,  $\arg\{H(e^{j\omega})\} = K_1 \omega + K_2$  para constantes  $K_1$  y  $K_2$ . A partir de esta propiedad, se nota inmediatamente que el retardo de grupo es constante a todas las frecuencias. Filtro con una respuesta en fase lineal son usualmente los buscados, debido a que no poseen distorsión temporal dependiente de la frecuencia. Un filtro IIR con polos distintos de cero puede tener una fase lineal. Sin embargo, un filtro FIR con coeficiente  $h_0, \dots, h_N$  tiene fase lineal si existe un  $\varphi$  tal que para todo  $k \in \{0, \dots, N\}$ :

$$h_{N-k} = e^{j\varphi} h_k^* \quad (15)$$

Esto es, si los coeficientes invertidos son conjugados complejos de la próxima secuencia sumados a una constante de cambio de fase, el retardo de grupo será entonces:

$$\tau_d(f) = \frac{N}{2} \quad (16)$$

## II. FILTROS IIR TRUNCADOS (TIIR)

Considere un filtro FIR con una secuencia geoméricamente truncada  $h_0, h_0 p, \dots, h_0 p^N$  como respuesta al impulso. Este filtro tiene la misma respuesta al impulso para los primeros  $N+1$  términos que un filtro IIR de un solo polo con función de transferencia

$$H_{IIR}(z) = \frac{h_0}{1 - pz^{-1}} \quad (17)$$

Si se subtrae la "cola" de la respuesta al impulso, se obtiene:

$$H_{FIR} = h_0 + h_0 p z^{-1} + \dots + h_0 p^N z^{-N} \quad (18)$$

$$H_{FIR} = h_0 \frac{p^{N-1} z^{-(N+1)}}{1 - pz^{-1}} \quad (19)$$

La recursión en el dominio del tiempo para este filtro estará dada como:

$$y[n] = \sum_{k=0}^N h_0 p^k x[n-k] \quad (20)$$

$$y[n] = p y[n-1] + h_0 \{x[n] - p^{N+1} x[n - (N+1)]\} \quad (21)$$

Se nota que la primer formulación (Ec. 20) requiere  $N+1$  múltiplos y  $N$  sumandos para ser implementada de forma directa; mientras que la segunda ecuación (Ec. 21) requiere de solo tres múltiplos y dos sumandos, independientemente de  $N$ . Por tanto, se ve que, si se puede representar un FIR como una secuencia exponencial truncada, se encontraría una forma reducida de computar el filtro. Notar que el término  $x[n - (N+1)]$  en la Ec. 21 todavía requiere un retardo para ser mantenido; y por tanto, no hay un reducción en cuanto a almacenamiento.

Existe una cancelación de polo-cero en la representación dada en la Ec. 19. Si  $|p| < 1$  no hay problemas, ya que el sistema será inherentemente estable. Si  $|p| > 1$ , sin embargo, entonces existe un problema potencial debido al "modo oculto".

La idea de esta sección funciona solo para casos en los que la multiplicidad de cada polo es uno; tal que cada modo exhiba un decaimiento exponencial simple.

### II-A. Extensión a secuencias de alto orden

Se puede extender la idea ilustrada en la sección anterior para el caso de un polo simple con el objetivo de computar secuencias TIIR de cualquier denominador de  $H(z)$ . El procedimiento general es el de encontrar una la cola la función transferencia IIR:

$$H'_{IIR}(z) = h'_0 z^{-1} + h'_1 z^{-2} + \dots = h_{N+1} z^{-1} + h_{N+2} z^{-2} + \dots \quad (22)$$

cuya respuesta al impulso, excepto por un cambio temporal de  $N$  pasos, es igual a la cola de la función transferencia  $H_{IIR}(z)$ , la cual se pretende trunca luego del paso  $N$ .

Entonces, multiplicando  $H_{IIR}(z)$  por  $z^N$  se obtiene:

$$z^N H_{IIR}(z) = h_0 z^N + \dots + h_{N-1} z + h_N + h_{N+1} z^{-1} + \dots \quad (23)$$

$$= C(z) + H'_{IIR}(z) \quad (24)$$

$$= \frac{z^N B(z)}{A(z)} \quad (25)$$

$$= C(z) + \frac{B'(z)}{A(z)} \quad (26)$$

donde  $O[B'(z)] < O[A(z)] = P$ . Se puede asumir que  $O[B'(z)] = P - 1$ .  $B'(z)$  es única y se puede obtener al realizar la división sintética de  $z^N B(z)$  por  $A(z)$  y encontrar el sobrante.

Una vez obtenida  $B'(z)$ , se tiene que  $H'_{IIR} = B'(z)/A(z)$ , y se puede escribir:

$$H_{FIR}(z) = H_{IIR}(z) - z^{-N} H'_{IIR}(z) = \frac{B(z) - z^{-N} B'(z)}{A(z)} \quad (27)$$

El sistema correspondiente será entonces:

$$y[n] = - \sum_{k=1}^P a_k y[n-k] + \sum_{l=0}^P b_l x[n-l] - \sum_{m=0}^{P-1} b'_m x[n-m-(N+1)] \quad (28)$$

El hecho de que los denominadores de las funciones transferencia  $H_{IIR}(z)$  y  $H'_{IIR}(z)$  son los mismos permite un menor costo computacional debido al hecho que el IIR original y la cola IIR dinámica son las mismas y no se necesita realizar procedimientos duplicados. El costo computacional de este sistema IIR de orden  $P$  es de  $3P + 1$  multiplicandos y  $3P - 2$  sumandos, independiente de  $N$ . Así, una ganancia computacional con esta clase de filtros FIR es conseguida si  $N > 3P$ .

El costo de almacenamiento para este filtro son  $P$  muestras de salida para la realimentación dinámica IIR,  $N$  muestras de entrada para el filtro FIR, y un adicional de  $P$  muestras de entrada para la cancelación de la cola; significando  $N + P$  muestras de entradas con retraso, de las cuales solo las últimas  $P$  son utilizadas, y  $P$  muestras de salida con retraso. Así, el algoritmo FIR requiere  $2P$  más de almacenamiento que una implementación FIR directa.

### II-B. Otras arquitecturas

La implementación directa especificada según la Ec. 28 puede ser no deseada por varias razones. Por ejemplo, uno puede optar por utilizar una estructura factorizada tal como la forma bicuadrática cascada o la forma de fracciones paralelas parciales. Esta última está dada como:

$$H(z) = \sum_{k=1}^{N_p} \sum_{l=1}^{M_k} \frac{C_{k,l}}{(1 - p_k z^{-1})^l} \quad (29)$$

donde  $N_p$  es el número de polos distintos, y  $M_k$  es la multiplicidad del polo número  $k$ . El término  $(k, l)$  es la expansión por fracciones parciales correspondiente al filtro con respuesta al impulso

$$h_{k,l}[n] = C_{k,l} \binom{n+l-1}{n-1} p_k^n \quad (30)$$

Para formar el filtro TIIR, una cola del filtro IIR es derivada por cada fracción parcial utilizando división sintética como se demostró en la sección anterior. Cada respuesta TIIR es calculada de forma separada, y los resultados son sumados para formar la respuesta completa. La factorización no debe ser necesariamente completa como se muestra en la Ec. 29. Es posible optar por un nivel intermedio de factorización, dejando algunos factores juntos y otros separados. Por ejemplo, se puede optar por agrupar los pares conjugados complejos para evitar aritmética compleja. Alternativamente, se puede dejar los terminos con los mismos polos juntos, según:

$$H(z) = \sum_{k=1}^{N_p} \frac{B_k(z)}{(z - p_k)^{M_k}} \quad (31)$$

ya que para cacular la cola de la respuesta IIR para cada término de orden  $n$  se necesita un polinomio de orden  $n - 1$  en el numerador de todas formas. La respuesta al impulso de cada fracción parcial  $k$  en este caso es:

$$h_k[n] = \sum_{l=0}^{M_k} b_{k,l} \binom{n+l+M_k-1}{M_k-1} p_k^{n-l} \quad (32)$$

donde  $b_{k,l} = 0, \dots, M_k$  son los coeficientes de  $B_k(z)$ .

Otro ejemplo es agrupar a los factores estables juntos; esto es, aquellos con polos  $p_k$  tales que  $|p_k| < 1$ , e implementar de forma separada polos inestables para los cuales  $|p_k| > 1$ .

## III. FILTROS FAST FIR (FFIR) DE FASE LINEAL

La implementación de filtros como sistemas TIIR no tendría mucho sentido a menos que exista una ventaja que no sea posible con filtros IIR no-truncados. Esta ventaja radica en la posibilidad de realizar filtros de fase-lineal exacte. Se presentarán dos estrategias básicas para el diseño basado en filtros TIIR.

### III-A. Método de diseño por factorización aditiva

Se vió que un filtro FIR tendrá fase lineal si se cumple la Ec. 15. Es posible mantener dicha relación utilizando filtros TIIR. Si  $H_{FIR}^+$  es una función transferencia TIIR tal que:

$$H_{FIR}^+(z) = \sum_{k=0}^N h_k^+ z^{-k} = \frac{B^+(z) - z^{-N} B^{+'}(z)}{A^+(z)} \quad (33)$$

Formando la función transferencia de tiempo-contrario truncada:

$$H_{FIR}^-(z) = \sum_{k=0}^N h_k^- z^{-k} \quad (34)$$

$$= \sum_{k=0}^N h_k^{+*} z^{k-N} \quad (35)$$

$$= z^{-N} \left[ H_{FIR}^+ \left( \frac{1}{z^*} \right) \right]^* \quad (36)$$

$$= \frac{z^{-N} \left[ B^+ \left( \frac{1}{z^*} \right) - B^{+'} \left( \frac{1}{z^*} \right) \right]^*}{\left[ A^+ \left( \frac{1}{z^*} \right) \right]^*} \quad (37)$$

$$= \frac{-z B^{-'}(z) + z^{-N} B^-(z)}{A^-(z)} \quad (38)$$

donde los superíndices + y - denotan 'adelante' y filtro 'reverso-conjugado' respectivamente; y el superíndice \* denota conjugación compleja. De esta forma, comparando con las Ec. 12 y 13 se tiene:

$$A^-(z) = 1 + a_1^* z + \dots + a_P^* z^P \quad (39)$$

$$B^-(z) = b_0 + b_1^* z + \dots + b_P^* z^P \quad (40)$$

Entonces:

$$B^{-'}(z) = b_0^* + b_1^* z + \dots + b_{P-1}^* z^{P-1} \quad (41)$$

Se asume que  $B^+(z)$  y  $B^{-'}(z)$  tienen ordenes  $P$  y  $P-1$  respectivamente. Si se asume que  $H_{FIR}^+(z)$  es un filtro TIIR estable, entonces  $H_{FIR}^-(z)$  es un filtro TIIR inestable cuyos modos escondidos son conjugados recíprocos de los de  $H_{FIR}^+(z)$ .

Utilizando un cambio de tiempo arbitrario  $M \geq 0$  y cambio de fase  $\varphi$ , se define el filtro:

$$H_{LPFIR}(z) = H_{FIR}^+(z) + e^{j\varphi} z^{-M} H_{FIR}^-(z) \quad (42)$$

$$= \sum_{k=0}^N h_k^+ z^{-k} + e^{j\varphi} \sum_{k=0}^N h_k^{+*} z^{k-N-M} \quad (43)$$

Se nota que este nuevo filtro FIR de longitud  $M+N+1$  es invariante con respecto al orden reverso, conjugando los coeficientes, y multiplicando por un factor de fase  $e^{j\varphi}$ . De esta forma, se cumple la Ec. 15; y así,  $H_{LPFIR}(z)$  es un filtro FIR de fase lineal. Se puede ver la propiedad de fase lineal de forma más directa al notar primero que en el círculo unidad

$$H_{FIR}^-(e^{j\omega}) = e^{-jN\omega} [H_{FIR}^+(e^{j\omega})]^* \quad (44)$$

tal que

$$H_{LPFIR}(e^{j\omega}) = H_{FIR}^+(e^{j\omega}) + e^{j\varphi} e^{-jM\omega} H_{FIR}^-(e^{j\omega}) \quad (45)$$

$$= e^{-j[(N+M)\omega + \varphi]/2} H_{FIR}^+(e^{j\omega}) \quad (46)$$

por lo que el retardo de grupo resulta en:

$$\tau_d = \frac{M+N}{2} \quad (47)$$

Es relativamente complicado diseñar un FFIR para un determinado set de especificaciones utilizando el método de factorización aditiva debido a que no es intuitivamente obvio como controlar la magnitud  $H_{LPFIR}(e^{j\omega})$ . Sin embargo, para determinadas respuestas al impulso que están bien caracterizadas, ésta técnica provee una herramienta útil para el diseño de filtros FFIR.

### III-B. Método de diseño por magnitud al cuadrado

El siguiente método es tal vez de mayor utilidad para el diseño de filtros FFIR en aplicaciones de procesamiento de señal. Se comienza eligiendo una función transferencia no-negativa y real  $H^2(z) > 0$  para la cual existe una función transferencia estable  $H(z)$  tal que:

$$H^2(e^{j\omega}) = |H(e^{j\omega})|^2 \quad (48)$$

$$= H(e^{j\omega}) * H(e^{j\omega}) \quad (49)$$

y con polinomios  $A^+(z)$  y  $B^+(z)$  tales que:

$$H(z) = \sum_{k=0}^{\infty} h_k z^{-k} \quad (50)$$

$$= \frac{B^+(z)}{A^+(z)} \quad (51)$$

Si los coeficientes de  $A^+(z)$  y  $B^+(z)$  son reales, entonces la Ec. 49 indica que:

$$H^2(e^{j\omega}) = H(e^{j\omega}) H(e^{-j\omega}) \quad (52)$$

Formando la función transferencia TIIR

$$H_{FIR}^+(z) = \sum_{k=0}^N h_k z^{-k} \quad (53)$$

$$= \frac{B^+(z) - z^{-N} B^{+'}(z)}{A^+(z)} \quad (54)$$

al igual que se hizo en la Ec. 33. Similarmente, se forma  $H_{FIR}^-(z)$  al igual que en la Ec. 38; y, en vistas de la Ec. 44 se ve que el filtro

$$H_{FIR}^2(z) = H_{FIR}^+(z) H_{FIR}^-(z) \quad (55)$$

tiene la propiedad de

$$H_{FIR}^2(e^{j\omega}) = e^{-jN\omega} |H_{FIR}^+(e^{j\omega})| \quad (56)$$

y que obviamente es un filtro de fase lineal con retardo de grupo igual a:

$$\tau_d = N \quad (57)$$

Este filtro puede ser implementado como una cascada de  $H_{FIR}^+(z)$  y  $H_{FIR}^-(z)$ , que a su vez pueden ser implementadas según la Ec. 28.

La relación entre  $H_{FIR}^2(z)$  y  $H^2(z)$  se puede ver considerando la convolución cíclica en la Ec. 11:

$$H_{FIR}^+(e^{j\omega}) = \frac{1}{2\pi} \int_0^{2\pi} H(e^{j\theta}) W_N[e^{j(\omega-\theta)}] d\theta \quad (58)$$

donde

$$W_N(e^{j\omega}) = \frac{\sin[(N+1/2)\omega]}{\sin(\omega/2)} \quad (59)$$

La longitud  $N$  del filtro para  $H_{FIR}^+(z)$  y  $H_{FIR}^-(z)$  debe ser elegida lo suficientemente larga como para que el error inducido por la convolución periódica entre  $H(e^{j\omega})$  y  $W_N(e^{j\omega})$  no genere distorsión en la respuesta en frecuencia. De hecho, sin tener en cuenta el ruido por cuantización en el momento, como  $N$  crece hacia el infinito,  $W_N(e^{j\omega})$  se aproxima el impulso centrado en  $\omega = 0$ , y  $H_{FIR}^+(z)$  converge a  $H(z)$ .

Se nota que la fase del filtro  $H(z)$  utilizada en el diseño de  $H_{FIR}^2(z)$  es irrelevante, y así, filtros IIR, que normalmente se considera que poseen una excesiva distorsión de fase cerca de las frecuencias de corte, pueden ser utilizados. Por tanto, cualquier filtro IIR de tiempo discreto tal como Chebyshev, elíptico o Butterworth puede ser elegido como  $H(z)$  en la Ec. 49 y transformado a una magnitud cuadrada con fase lineal. Adicionalmente, el hecho de que la magnitud para  $H^2(z)$  es el doble de la distancia entre 0 [dB] comparada con la magnitud de  $H(z)$  implica que las especificaciones en la banda de rechazo de  $H^2(z)$  deberían ser solo la mitad de las deseadas en cuanto a magnitud, mientras que en cambio, el ripple en la banda de paso de  $H(z)$  debe ser más de la mitad que las especificaciones de diseño deseadas.

### III-C. Algoritmo de truncamiento refinado

En ambos métodos descritos en las dos subsecciones anteriores, la longitud  $N$  está limitada por el aumento del error de cuantización en el filtro inestable  $H_{FIR}^-(z)$ . Una forma de solucionar este inconveniente es implementar un piso significativo  $\lambda_S$  sobre el piso para precisión numérica  $\lambda_P$ . Si  $H(z)$  es la respuesta al impulso IIR sin truncar asociada a  $H_{FIR}^+(z)$  como en la Ec. 54; mientras  $H(z)$  sea estable, se tiene que  $|p_k| < 1$ , donde  $p_k$  son los polos de  $H(z)$ , y consecuentemente de  $H_{FIR}^+(z)$ . Los polos de  $H_{FIR}^-(z)$  son  $1/p_k^*$  y por tanto es inestable. Se realiza una expansión por fracciones parciales de  $H(z)$  y luego se observa la respuesta al impulso de cada fracción parcial. Se define un punto de corte  $N_k$  para la fracción parcial número  $k$  en el tiempo más pequeño luego del cual la máxima respuesta al impulso se

vuelve insignificante; esto es,  $\forall n > N_k, |\mu h_{k,n}| \leq \lambda_S$ , donde  $\mu$  es la magnitud de entrada más grande. Se puede resolver

$$|h_{k,N_k}| = \frac{\lambda_S}{\mu} \quad (60)$$

numéricamente utilizando la aproximación:

$$h_{k,n} \simeq \sum_{l=0}^{M_k} b_{k,l} (n-l)^{M_k-1} p_k^{n-l} \quad (61)$$

$$\simeq n^{M_k-1} p_k^n \sum_{l=0}^{M_k} \frac{b_{k,l} p_k^{-l}}{(M_k-1)!} \quad (62)$$

$$= B_k n^{M_k-1} p_k^n \quad (63)$$

para grandes valores de  $n$ , donde  $B_k$  es definida de forma implícita. Para  $M_k = 1$  se tiene:

$$N_k = \frac{\log\left(\frac{\lambda_S}{\mu B_k}\right)}{\log(|p_k|)} \quad (64)$$

Así, si  $N_k < N$ , se puede truncar la fracción parcial  $k$  a  $N_k$  en lugar de  $N$ . Sin embargo, ya que  $H(z)$  es estable y las respuestas debidas a las fracciones parciales  $k$  luego del tiempo  $N_k$  se encuentran por debajo del piso de significancia, esto no hace ninguna diferencia. El refinamiento se da por la implementación de  $H_{FIR}^-(z)$  como la suma de las inversas de fracciones parciales truncadas, pero con la fracción parcial  $k$  truncada luego de  $n = N_k$  muestras en lugar de  $n = N$  si  $N_k < N$ . Por tanto, el modo inestable de  $H_{FIR}^-(z)$  debido al polo  $1/p_k^*$  solo necesita aumentar para tener una magnitud mayor al piso de significancia  $\lambda_S$  y entonces posee menos tiempo para acumular exponencialmente ruido por cuantización.

Se tiene que:

$$H_{FIR}^-(z) = \sum_{k=1}^{N_p} \frac{-z B_k^{-'}(z) + z^{-N'_k} B_k^-(z, N'_k)}{(1 - p_k^* z)^{M_k}} \quad (65)$$

Asumiendo que el error de cuantización ocurre en el orden del piso de precisión  $\lambda_P$ , se tiene que  $\rho_v^2 \simeq \lambda_P^2$ . Adicionalmente, el piso de significancia setea un nivel de tolerancia al ruido conveniente; por tanto, se puede asumir  $\rho_v^2 = \lambda_P^2$ . De esta manera:

$$\lambda_S^2 \geq \lambda_P^2 \frac{1 - |p_k|^{-4N'_k}}{1 - |p_k|^{-2}} \quad (66)$$

$$\geq \lambda_P^2 \frac{1 - |p_k|^{-4N_k'}}{|p_k|^{-2} - 1} \quad (67)$$

Y así:

$$\lambda_P \leq \lambda_S |p_k|^{2N_k'-1} \sqrt{1 - |p_k|^2} \quad (68)$$

Combinando las Ec. 64 y 68, y asumiendo  $M_k = 1$ , se llega a:

$$\lambda_P \leq \frac{\lambda_S^3 \sqrt{1 - |p_k|^2}}{\mu^2 |p_k| |B_k|^2} \quad (69)$$

Se nota que la precisión, en bits, para las variables de estado debe ser al menos tres veces mayor que la precisión de los datos para prevenir una acumulación de ruido significativa. Este resultado es intuitivo ya que la parte significativa de la respuesta al impulso debe trabajar en el rango dinámico especificado por la salida de mayor magnitud (que se puede asumir como 1 de forma normalizada) y  $\lambda_S$  en un periodo de  $N_k t$  muestras; mientras que el ruido en  $H_{FIR}^-(z)$  puede crecer en un periodo de hasta  $2N_k t$  muestras utilizando las mismas dinámicas. El análisis anterior no es generalmente aplicable cuando se trata con cantidades de punto flotante, debido a que el rango de significancia varía con el exponente. Asumiendo una entrada de energía constante, se ve que el análisis todavía se mantiene, ya que los pisos de significancia y precisión son aproximadamente constantes en un estado continuo.

#### III-D. Filtros TIIR en modo de magnitud unitaria

Una tercera clase de filtros lineales consiste de respuestas TIIR puras con una longitud  $N$  tal que todos sus modos esten sobre el círculo unidad y que cada uno posea multiplicidad 1. Así, es posible asegurar que  $H_\mu(z)$  tendrá una fase lineal. Esto se puede observar en un filtro que satisfaga la Ec. 15. Si no hay modos con multiplicidad mayor a 1, entonces la Ec. 15 no se satisface. Sin embargo, se pueden plantear cuidadosamente los ceros para forzar a que se cumpla la Ec. 15.

Aplicando el análisis de la subsección anterior, se ve a partir de la Ec. 64 que los  $N_k$  para este tipo de filtro serán todos infinito. En este caso, no es necesario resetear las variables de estado seguido, ya que el crecimiento en el error de cuantización será aditivo y no exponencial.

Si se considera el peor escenario, el número de muestras en las cuales el error en el bit de menor significancia (LSB) se puede llegar a acumular antes de alcanzar el piso de significancia  $\lambda_s$  es aproximadamente

$$N = \frac{\lambda_S}{\lambda_P} \quad (70)$$

asumiendo que hay 1  $b$  de error en el LSB por muestra procesada. Si la media del ruido no es cero, entonces el error acumulado es aleatorio, significando que el error estándar crece proporcionalmente a la raíz cuadrada del número de muestras procesadas. Se examina entonces el error para que se encuentre a una distancia del LSB. Para un error acumulativo que se encuentra por debajo del piso de significancia se tiene que:

$$3\lambda_P\sqrt{N} < \lambda_S \quad (71)$$

Resultando:

$$N < \left( \frac{\lambda_S}{3\lambda_P} \right)^2 \simeq \frac{4^G}{10} \quad (72)$$

donde  $G$  es el número de bits de seguridad. De esta forma, se ve que los filtros TIIR con modos escondidos en el círculo unitario tiene propiedades de estabilidad deseables en comparación con los descritos en las dos secciones anteriores.

Como la magnitud unitaria de estos filtros TIIR es cuasies- table y no necesita ser reseteada cada  $N$  ciclos, el número

de computaciones necesarias para estabilizarlos puede ser reducido significativamente. Como un filtro FIR tiene memoria infinita, es suficiente realizar dos filtros en paralelo comenzando con solo  $N$  pasos antes de la transferencia de variables de estado.

La respuesta al impulso de filtros TIIR con magnitud unitaria tiene la forma:

$$h_k[n] = P_k(n)p_k^n = P_k(n)e^{j\omega_k n} \quad (73)$$

para  $n = 0, \dots, N$ , donde  $p_k$  es el polo único  $k$  y  $P_k(n)$  es el polinomio de grado  $m_k - 1$  tal que  $m_k$  es el múltiplo del polo  $k$ .

#### IV. REFERENCIAS

- [1] H. Kopka and P. W. Daly, *A Guide to L<sup>A</sup>T<sub>E</sub>X*, 3rd ed. Harlow, England: Addison-Wesley, 1999.