

DISEÑO DE FILTROS FIR MEDIANTE MÍNIMOS CUADRADOS

Esequiel Zavala Huavel¹ y Luis González Segura²
 E-mail : d270037@unmsm.edu.pe

- 1.- *Universidad Nacional Mayor de San Marcos
 Facultad de Ingeniería Electrónica, Lima-Perú*
- 2.- *Telefónica del Perú, Departamento de Larga Distancia
 Nuevos Servicios e Interconexión*

Resumen: Este trabajo presenta el método de diseño de filtro digital de respuesta finita al impulso (FIR, Finite Impulse Response) mediante la aproximación de mínimos cuadrados. Este método permite múltiples pasos, bandas suprimidas y bandas de transición arbitraria, todos con control explícito de los bordes de banda y es tan simple como usar el método de ventanas. Este método de diseño acumula una respuesta de frecuencia multibanda óptima mediante la adición y sustracción secuencial de filtros pasabajos óptimos con bandas de transición. Sin embargo, este método asume que la respuesta ideal tiene una función de transición spline y no permite ponderaciones con error.

Abstract: This paper features the method for the design of least squares FIR digital filter. This method allow multipass stop bands and arbitrary transition bands together with explicit control of band edges and is so simple as using the windowing method. This method of design accumulates an optimal multiband frequency response by means of sequential addition and subtraction of optimal low pass filters with transition bands. However, this method assumes that the ideal response contains a spline transition function and there fore error weighting are not allowed.

Palabras Claves: FIR, Spline, IIR, Pasabajo, Pasa Banda.

I. INTRODUCCIÓN

Este trabajo presenta el método de diseño de filtros digitales (FIR, *Finite Impulse Response*) de fase lineal de mínimo error cuadrático que permite control explícito de bordes de banda y ganancia de la banda para múltiples bandas pasantes y control independiente de cada transición del borde de la banda. Este es un método analítico el cual deduce fórmulas para los coeficientes del filtro, y de este modo filtros multibandas óptimos muy largas para ser diseñadas rápidamente en computadoras y asume que la respuesta ideal tiene funciones de transición spline, y no permite ponderaciones diferentes de la banda pasante, banda suprimida o banda de transición.

Las especificaciones se dan en un formato similar al que dan para el algoritmo PM [Parks y Burrus, 1987], [Moler, Little y Bangert, 1989]. En realidad este método junto con el algoritmo PM dan dos aproximaciones de diseño optimo, uno con el criterio de error Chebychev y los otros con el criterio de error cuadrático.

La primera contribución de este trabajo es demostrar, como un filtro banda pasante simple puede ser combinado tal que las truncaciones simples terminan en una aproximación óptima a la respuesta ideal deseada con bandas de transición de la función spline. Una segunda contribución es demostrar cuan limitante es la ponderación de error que es constante sobre cada banda separada permitiendo el diseño de filtros FIR de mínimo error cuadrático de óptima ponderación de banda múltiple. Una tercera contribución es la inclusión de programas MATLAB, para ambos tipos de diseños de filtros.

La razón por la que un filtro multibanda óptimo sin ponderación, puede ser formado a partir de una suma o diferencia de filtros pasabajos óptimos depende del uso del criterio de error cuadrático ponderado uniformemente.

II. FILTROS PASABAJO BÁSICO

La respuesta en frecuencia de un filtro FIR de longitud N puede ser calculado desde la transformada discreta de Fourier (DTFT) de la respuesta al impulso como

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

Asumiendo la fase lineal, puede ser escrito

$$H(\omega) = A(\omega) e^{-jM\omega} \quad (2)$$

$A(\omega)$: amplitud de valor real (puede ser positivo o negativo) y $M = (N - 1) / 2$ es el retardo constante de grupo del filtro [Parks y Burrus, 1987]. Para N impar, esto da

$$A(\omega) = \sum_{n=-M}^M \hat{h}(n) \cos(\omega n) \quad (3)$$

$\hat{h}(n) = h(n + M)$ es $h(n)$ desplazado para ser simétrico alrededor de $n = 0$.

Para N par la formulación es un tanto diferente y da:

$$A(\omega) = \sum_{n=-N/2}^{N/2-1} \hat{h}(n) \cos(\omega(n + 1/2)) \quad (4)$$

donde $\hat{h}(n) = h(n + N/2)$ es $h(n)$ desplazado para ser simétrico alrededor de $n = -\frac{1}{2}$. La inversa de la DTFT para N impar da

$$\hat{h}(n) = \frac{1}{\pi} \int_0^{\pi} A(\omega) \cos(\omega n) d\omega \quad (5)$$

y para N par

$$\hat{h}(n) = \frac{1}{\pi} \int_0^{\pi} A(\omega) \cos(\omega(n + \frac{1}{2})) d\omega \quad (6)$$

debido a las simetrías en $h(n)$ y $\hat{h}(n)$, impuestas por la fase lineal, si definimos.

$$\hat{h}(n) = \begin{cases} 1/2 \hat{h}(n) & \text{Para } n=0 \\ \hat{h}(n) & \text{Para } 0 \leq n \leq M \\ 0 & \text{Para otros casos} \end{cases} \quad (7)$$

Para N impar tenemos:

$$A(\omega) = \sum_{n=0}^M 2 \hat{h}(n) \cos(\omega n) \quad (8)$$

El cual tiene la mitad del N° de ecuaciones. Para N par tenemos:

$$A(\omega) = \sum_{n=0}^{N/2-1} 2 \hat{h}(n) \cos(\omega(n+1/2)) \quad (9)$$

El cual también tiene la mitad del N° de ecuaciones.

Para una aproximación de mínimo error cuadrático ponderado, la escala puede ser minimizada y definida como:

$$\varepsilon = \frac{1}{\pi} \int_0^{\pi} W(\omega) |A(\omega) - A_d(\omega)|^2 d\omega \quad (10)$$

Se convierte a full banda con medida de error promediado uniformemente fijando $W(\omega) = 1$, dando

$$\varepsilon = \frac{1}{\pi} \int_0^{\pi} |A(\omega) - A_d(\omega)|^2 d\omega \quad (11)$$

El problema de diseño de filtros FIR es determinar una respuesta en frecuencia cuya $h(n)$ sea de longitud finita, o en el caso de fase lineal, la respuesta de amplitud se aproxima a la deseada. Un truncamiento simple de la secuencia de coeficientes calculados a partir de (5) o (6) minimiza (11) pero no minimizará el error cuadrático ponderado en (10) ni ninguna definida sobre las partes de la banda.

Para algún caso especial pero muy útil podemos evaluar analíticamente la integral de la inversa DTFT en (5) y (6), el cual después de la truncación se hace finito y el desplazamiento se hace causal, para que el diseño sea óptimo.

2.1 Ganancia constante en la banda pasante del filtro pasa bajo ideal

El filtro FIR pasabajo ideal tiene una banda pasante de ganancia unitaria para $0 < \omega < \omega_0$ y ganancia cero de la banda alta para $\omega_0 < \omega < \pi$. Esto

$$A_d(\omega) = \begin{cases} 1 & 0 \leq \omega \leq \omega_o \\ 0 & \omega_o \leq \omega \leq \pi \end{cases} \quad (12)$$

Como la respuesta de amplitud deseada ideal. Los coeficientes del filtro ideal desplazado es la inversa de DTFT a partir (5). Los cuales para N impar están dados por

$$\hat{h}_d(n) = \frac{1}{\pi} \int_0^\pi A_d(\omega) \cos(\omega n) d\omega \quad (13)$$

$$= \frac{1}{\pi} \int_0^{\omega_o} \cos(\omega n) d\omega = \frac{\omega_o}{\pi} \frac{\sin(\omega_o n)}{\omega_o n} \quad (14)$$

El cual es algunas veces llamado una función "sinc". Note que $\hat{h}_d(n)$ es generalmente infinita en longitud. Este es simétricamente truncado y desplazado por $M = (N-1)/2$ para dar los óptimos coeficientes del filtro FIR de longitud N como:

$$h(n) = \frac{\omega_o}{\pi} \frac{\sin(\omega_o(n - M))}{\omega_o(n - M)} \quad \text{para } 0 \leq n \leq N-1 \quad (15)$$

$$h(n) = 0 \text{ en otro caso}$$

La deducción correspondiente para una longitud par empieza con la inversa de DTFT en (6) para un filtro de longitud par desplazado.

$$\begin{aligned} \hat{h}_d &= \frac{1}{\pi} \int_0^\pi A_d(\omega) \cos(\omega(n + \frac{1}{2})) d\omega \\ &= \frac{\omega_o}{\pi} \frac{\sin(\omega_o(n + \frac{1}{2}))}{\omega_o(n + \frac{1}{2})} \end{aligned} \quad (16)$$

El cual cuando es truncada y desplazada por $N/2$ da la misma fórmula que para el diseño de longitud impar en (15), pero notará que $M = (N-1)/2$ no es un entero para N par.

2.2 Filtro pasabaja de banda pasante lineal ideal

Deducimos ahora, la fórmula de diseño para un filtro con una respuesta de amplitud ideal que tiene una función incrementante lineal en la banda pasante, algo que es una constante como se asumió anteriormente. Esta respuesta de amplitud ideal esta dado por:

$$A_d(\omega) = \begin{cases} \frac{1}{\pi} \omega & 0 \leq \omega \leq \omega_o \\ 0 & \omega_o \leq \omega \leq \pi \end{cases} \quad (17)$$

Para N impar los coeficientes del filtro ideal desplazado de longitud infinitamente grande son la inversa de DTFT de esta amplitud dada por

$$\begin{aligned}\hat{h}_d(n) &= \frac{1}{\pi} \int_0^{\omega_o} \left(\frac{1}{\pi} \omega \right) \cos(\omega n) d\omega \\ &= \frac{\cos(\omega_o n) - 1}{\pi^2 n^2} + \frac{\omega_o \sin(\omega_o n)}{\pi^2 n}\end{aligned}\quad (18)$$

Con el indeterminado $\hat{h}_d(0) = \omega_o^2 / 2\pi^2$. Este es truncado y desplazado por $M = (N-1)/2$ para obtener los coeficientes del filtro FIR de óptima causal y longitud N como:

$$h(n) = \frac{\cos[\omega_o(n-M)] - 1}{\pi^2(n-M)^2} + \frac{\omega_o \sin[\omega_o(n-M)]}{\pi^2(n-M)} \quad 0 \leq n \leq N-1 \quad (19)$$

$$h(n) = 0 \text{ en otro caso.}$$

La deducción correspondiente para una longitud uniforme comienza con la inversa DTFT correspondiente a un filtro desplazado de longitud uniforme dado en (5) y después por $N/2$ para dar el mismo resultado (19).

2.3 Filtro pasa bajo más diferenciador ideal

Afortunadamente la inversa de DTFT para un diferenciador ideal combinado con un filtro pasabajo puede también ser analíticamente evaluado. La respuesta de amplitud ideal es la misma que la dada en (17), pero que en este caso tiene una respuesta al impulso que es simétrico impar, la inversa de DTFT utiliza funciones senoidales los cuales para N impar da:

$$\begin{aligned}h_d(n) &= \frac{1}{\pi} \int_0^{\omega_o} \left(\frac{1}{\pi} \omega \right) \sin(\omega n) d\omega \\ &= \frac{\sin(\omega_o n)}{\pi^2 n^2} - \frac{\omega_o \cos(\omega_o n)}{\pi^2 n}\end{aligned}\quad (20)$$

Con el indeterminado $\hat{h}_d(0) = 0$. Este es truncado y desplazado por $M = (N-1)/2$ para obtener los coeficientes del filtro FIR óptimo, causal y de longitud N como

$$h(n) = \frac{\sin(\omega_o(n-M))}{\pi^2(n-M)^2} - \frac{\omega_o \cos(\omega_o(n-M))}{\pi^2(n-M)}, \text{ para } 0 \leq n \leq N-1 \quad (21)$$

$$h(n) = 0 \text{ para otros casos}$$

Nuevamente la deducción correspondiente para una longitud par da el mismo resultado que en (21). Note que esta es una fórmula general que incluye como caso especial la longitud par e impar de un diferenciador de banda total ($\omega_o = \pi$) dado en [Parks y Burrus, 1987].

También debe observarse que para un diferenciador de toda la banda, una longitud par es preferido en lugar de una longitud impar debido al cero que se obtiene a $\omega = \pi$. Sin embargo para el diferenciador con un filtro pasabajo, el cero facilita en el filtrado pasabajo y de este modo puede ser una ventaja.

2.4 Transformada de Hilbert

La inversa de la DTFT correspondiente a una transformada de Hilbert ideal [Oppenheim y Schaffer, 1989] combinado con un filtro pasa bajo puede también ser analíticamente evaluado.

La respuesta de amplitud ideal es la misma que (12) pero con un desplazamiento de fase constante de $\phi = \frac{\pi}{2}$.

Puesto que este caso tiene una respuesta al impulso que es simétrico impar, la inversa de la DTFT utiliza funciones senoidales, los cuales para N impar da:

$$\hat{h}_d(n) = \frac{1}{\pi} \int_0^{\omega_o} \sin(\omega n) d\omega = \frac{1 - \cos(\omega_o n)}{\pi n} \quad (22)$$

Con la indeterminada $\hat{h}_d(0) = 0$, este es truncado y desplazado por $M = \frac{(N-1)}{2}$ para obtener los coeficientes del filtro FIR óptimo causal y de longitud N como

$$h(n) = \frac{1 - \cos(\omega_o(n-M))}{\pi(n-M)} \quad 0 \leq n \leq N-1 \quad (23)$$

$$h(n) = 0 \quad \text{en otros casos}$$

Nuevamente la deducción correspondiente para una longitud par da el mismo resultado que en (23)

2.5 Diseño de la banda de transición spline

Los cuatro filtros pasabajo descritos arriba presentan el fenómeno de Gibbs (para cualquier valor finito de N la amplitud pico de los rizados permanece constante) cuando es truncado a una longitud finita. Para eliminar este efecto y dar una especificación más explícita de los bordes pasante y de la banda suprimida se inserta una banda de transición entre los bordes pasante y la banda suprimida. Una banda de transición se ubica en esta banda para hacer que la respuesta de amplitud total deseada sea una función continua.

Si usamos una función spline de p -ésimo orden como la función de transición, el efecto de añadir esta banda de transición a la respuesta de amplitud ideal del filtro pasabajo básico ideal dado en (12) consiste en multiplicar la respuesta del impulso ideal en (14) por una función SINC para dar:

$$\hat{h}_d(n) = \frac{\sin(\omega_o n)}{\pi n} \left(\frac{\sin(\Delta n / p)}{\Delta n p} \right)^p \quad (24)$$

donde $\omega_o = (\omega_s + \omega_p) / 2$ es el promedio del borde de la banda y $\Delta = (\omega_s - \omega_p) / 2$ es la mitad del ancho de la banda de transición normalízalo en radianes/segundo para el caso de una velocidad de muestreo de una muestra por segundo [Schüssler y Steffen, 1984], [Parks y Burrus, 1987], [Burrus, Soewit y Gopinath, 1992]. El spline produce una función de transición el cual consiste de P segmentos de polinomios de pe-ésimo orden conectado todos juntos tal que P-1 deducciones son continuas en las uniones. El valor óptimo del exponente P es elegido tal que P es igual a $1.624 (f_s - f_p)N$ (para una velocidad de muestra unitaria) el cual minimiza el error de aproximación [Burrus, Soewit y Gopinath, 1992]. Cada uno de los cuatro filtros ideales deducido previamente pueden tener una simple banda de transición añadida mediante la multiplicación de sus respuestas al impulso por la función ponderada Sinc como se ilustra (24).

III. EL MÉTODO DE DISEÑO DE BANDA MÚLTIPLE ÓPTIMA

El método de diseño de banda múltiple óptima consiste de dos partes algo independientes, el primero es el diseño de un filtro pasabajo óptimo de mínimos cuadrados con una banda de transición como el escrito arriba o como se calculó mediante la inversa de FFT y la segunda parte construye un filtro multibanda óptimo desde una combinación de esos filtros pasabajos óptimos y es el punto central de este trabajo

El problema de diseño del filtro FIR de fase lineal por el método de mínimos cuadrados del error sin ponderar es determinar los coeficientes del filtro que minimice el error definido por.

$$\varepsilon = \int_0^\pi |A(\omega) - A_d(\omega)|^2 d\omega \quad (25)$$

donde:

$A(\omega)$: Es la respuesta de frecuencia en amplitud del filtro real.

$A_d(\omega)$: Es la respuesta de amplitud ideal deseada.

Esto es hecho mediante el truncamiento de la inversa de la transformada de Fourier en tiempo discreto de $A_d(\omega)$. La dificultad es la evaluación analítica de la integral en la transformada inversa [Parks y Burrus, 1987]. Si se utiliza una función de transición Spline, una fórmula analítica puede ser deducido para el filtro que minimiza (25). Los detalles de este resultado puede ser determinado en [Burrus, Soewit y Gopinath, 1992].

El diseño de los filtros de longitud infinita en base a la inversa de la transformada de Fourier en tiempo discreto de la respuesta ideal tiene una respuesta en frecuencia la cual es la misma que el ideal y de este modo no tiene error. Una respuesta de amplitud ideal puede ser formulado como la suma de filtros pasabajo ideal mas simples, diferenciadores o transformadores de Hilbert juntos con sus funciones de transición Spline dado por

$$A_d(\omega) = \sum_k K_k A_{dk}(\omega) \quad (26)$$

donde:

$A_{dk}(\omega)$: es la respuesta pasabajo deseada con una banda de transición en la k-ésima banda, tal como en (12) ó (17);

K_k : es una constante arbitrarias

Debido a la linealidad de la transformada de Fourier, una respuesta ideal de banda múltiple puede ser construido por simple adición y sustracción de las respuestas al impulso de un filtro bajo ideal apropiado.

$$\hat{h}_d(n) = \sum_k k_k \text{IDTFT}\{A_{dk}(\omega)\} \quad (27)$$

$$\hat{h}_d(n) = \sum_k k_k \hat{h}_{dk}(n) \quad (28)$$

Debido a la ortogonalidad de las funciones básicas de la transformada de Fourier, la secuencia truncada $\hat{h}_d(n)$ de la respuesta al impulso infinitamente largo dará una aproximación óptima a $A_d(\omega)$ en el sentido de mínimos cuadrados. Esta argumentación no permite bandas de transición afectadas de error de ponderación o "don't care" ni banda de transición cualesquiera, ni métodos tradicionales de tratamiento por ventanas. Sin embargo, esta argumentación sí permite la formulación de funciones de transición SPLINE [Burrus, Soewit y Gopinath, 1992]. Utilizando los hechos mencionados, un filtro multibanda óptimo puede ser construido mediante sucesivas sumas y restas de respuestas al impulso de filtros pasabajo óptimos, como los dados en (27). Por ejemplo un filtro pasabanda que se aproxima a cero para $0 < \omega < \omega_1$ tiene una banda de transición SPLINE para $\omega_1 < \omega < \omega_2$, se aproxima a uno (o alguna otra constante) para $\omega_2 < \omega < \omega_3$, tiene una 2ª banda de transición independiente para $\omega_3 < \omega < \omega_4$ y finalmente se aproxima a cero para $\omega_4 < \omega < \pi$, puede ser diseñado primeramente mediante un filtro pasabajo simple con una banda de transición $\omega_3 < \omega < \omega_4$ y luego sustrayendo desde su respuesta al impulso, el impulso de un 2º filtro pasa bajo diseñado con una banda de transición $\omega_1 < \omega < \omega_2$.

Un filtro con dos o más pasabandas puede ser diseñado sumando las respuestas al impulso de dos o más filtros pasa banda individuales.

En realidad, un método de diseño completamente general puede ser formulado mediante las sumas y sustracciones alternados de filtros pasa bajo comenzando en la frecuencia más alta de la banda de transición y desplazándose secuencialmente y descendiendo hacia la más baja. Si la respuesta en frecuencia ideal no es cero a $\omega = \pi$, luego uno comienza con una respuesta de frecuencia constante (un impulso en el dominio del tiempo) y sustrae el filtro pasabajo (recuerde que la longitud debe ser no uniforme para este caso). Aplicando consideraciones de escala a cada filtro pasabajo, son obtenidas diferentes ganancias en cada banda.

Debe tenerse cuidado que la función de transición spline construido se adapte convenientemente a las bandas en ambos lados. Esto no ayudará automáticamente si allí están dos bandas adyacentes con diferentes pendientes conectadas por una función de transición las cuales están simplemente sumadas entre sí. Esto ayudará automáticamente si cada banda pasante está separada por una banda suprimida o si bandas adyacentes tienen la misma pendiente.

3.1 Un programa de diseño de filtros mediante MATLAB

Un Programa en MATLAB denominado FIR 3.m, dado en el apéndice de este trabajo, diseñará filtros óptimos aplicando el método descrito en el apartado anterior. Este programa particular requiere ganancia pasabanda constante pero arbitraria y utiliza un formato para especificaciones similar a la función llamada *remez* () de MATLAB. Este construye el filtro de banda múltiple basándose en (28) mediante la adición y sustracción de filtros pasabajos óptimos, diseñados según la fórmula (24) y calculados en el segundo programa denominado *Fir31* p.m.

El programa principal se le asigna un vector f de longitud par que contiene los bordes de banda suprimida y banda pasante normalizados, incluyendo $f=0$ y $f=1$. A este programa se le asigna también un vector m de longitud par que contiene la respuesta ideal a cada frecuencia en f . Debido a que el filtro pasabajos tiene un pasabanda constante la respuesta ideal del filtro multibanda tendrá pasabandas constantes. Esto significa que m

VI. REFERENCIAS

- Parks, T.W., y C.S. Burrus. *Digital Filter Design*. New York: Wiley 1987.
- Moler, C., J. Little, y S. Bangert. *Matlab User's Guide*. South Natick, MA: Math Works. 1989.
- Oppenheim, A.V., y R. W. Schafer. *Discrete-Time Signal Processing* Englewood Cliffs, NJ: Prentice-Hall 1989.
- Schüssler, H.W., y P. Steffen, "A hybrid system for the reconstruction of a smooth function from its samples," *Circuits, Systems, and Signal Processing*, vol. 3, pp. 295-314, 1984.
- Burus, C.S., A. W. Soewit, y R. A. Gopinath, "Least squared error FIR filter design with transition bands," *IEEE Trans. Signal Processing*, vol. 40, N° 6 pp. 1327-1340, June 1992.
- Fleischer, P.E., "Digital realization of complex transfer functions," *simulation*, vol. 6, pp. 171-180, Mar, 1966.
- Tufts, D.W., D.W. Rorabacher, y W.E. Moiser, "Design of nonrecursive digital Filters", *IEEE Trans., Audio Electroacoust.*, vol. 18, N° 2, pp. 142-158, June 1970.
- Burrus, C.S., J.A. Barreto, y I. W. Selesnick, "Iterative reweighted least squares design of FIR filters," *IEEE Trans. Signal Processing*, vol. 42, N° 11, Nov. 1994.
- Selesnik, I.W., M. Lang, y C.S. Burrus, "Constrained least square design of FIR filters without explicitl specified transition bands," *IEEE Trans. Signal Processing*, Submitted June 1994.