Imperial College London

DSP Design of FIR Filters using the Remez Exchange Algorithm

Lecture 7

DR TANIA STATHAKI

READER (ASSOCIATE PROFESSOR) IN SIGNAL PROCESSING

IMPERIAL COLLEGE LONDON

Imperial College London

Computer-Aided Design of Linear-Phase FIR Filters

- In this section, we consider the application of computer-aided optimization techniques for the design of FIR filters.
- The basic idea behind the computer-based technique is to minimize iteratively an error measure that is function of the difference between the desired frequency response $D(e^{j\omega})$ and the frequency response $H(e^{j\omega})$ of the filter being designed.
- In the case of linear-phase FIR filter design, $H(e^{j\omega})$ and $D(e^{j\omega})$ are zero-phase frequency responses.
- For IIR filter design, these functions are replaced with their magnitude functions.

Computer-Aided Design of Linear-Phase FIR Filters

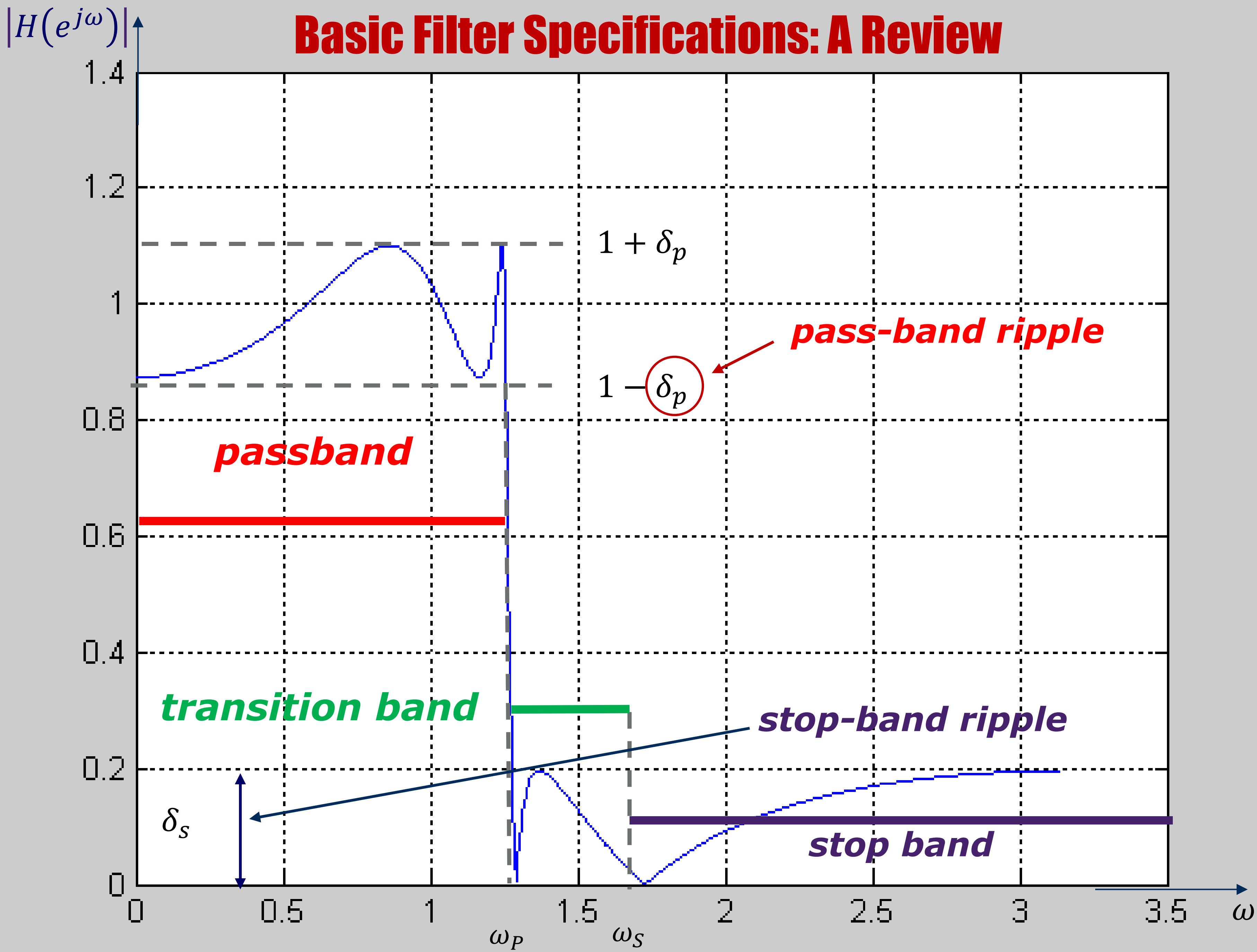
Previous part

- The windowing method and the frequency-sampling method are relatively simple techniques for designing linear-phase FIR filters.
- Here, a major problem, is a lack of precise control of the critical frequencies such cut-off frequencies of pass band and stop band.

This part

- The new filter design method described in this section is formulated as a so called **Chebyshev approximation problem**.
- It is viewed as an optimum design criterion in the sense that the maximum weighted approximation error between the desired frequency response and the actual frequency response is minimized.
- **The resulting filter designs have ripples in both the pass-band and the stop-band.**
- To describe the design procedure, let us recall the following basic filter specifications.

Basic Filter Specifications: A Review



Computer-Aided Design of Linear-Phase FIR Filters

- The design objective is to iteratively adjust the filter parameters so that the error function defined by the equation:

$$\varepsilon(\omega) = W(e^{j\omega})[H(e^{j\omega}) - D(e^{j\omega})]$$

  is minimum according to some criterion.

  $W(e^{j\omega})$ is some user-specified positive weighting function.

- The following criteria are popular:

  ***Minimax criterion*:**

  minimize $\max_{\omega \in R} \left| W(e^{j\omega})[H(e^{j\omega}) - D(e^{j\omega})] \right|$

  ***Least squares criterion*:**

  Minimize $\int_{\omega \in R} \left| W(e^{j\omega})\left(H(e^{j\omega}) - D(e^{j\omega})\right) \right|^p d\omega$

- R is the set of disjoint frequency bands in the range $0 \leq \omega \leq \pi$. In filtering applications, R is composed of passbands and stopbands.

Computer-Aided Design of Equiripple Linear-Phase FIR Filters

- The linear phase filter that is obtained by minimizing the peak absolute value of the weighted error ε given by

$$\varepsilon = \max_{\omega \in R} |\varepsilon(\omega)|$$

  is usually called the ***equiripple FIR filter***, since, after ε has been minimized, the weighted error function $\varepsilon(\omega)$ exhibits an equiripple behavior in the frequency range of interest.
- In this part we outline the ***weighted-Chebyshev approximation method*** advanced by Parks and McClellan for designing equiripple linear phase FIR filters.
- This method is more commonly known as the ***Parks-McClellan algorithm***.

Computer-Aided Design of Equiripple Linear-Phase FIR Filters

- The general form of the frequency response $H(e^{j\omega})$ of a causal linear-phase FIR filter of length $N+1$ is given by

$$H(e^{j\omega}) = e^{-jN\omega/2} e^{j\beta} \breve{H}(\omega)$$

where $\breve{H}(\omega)$ is the amplitude response of $H(e^{j\omega})$ and is a real function of ω.

- The weighted error function in this case involves the amplitude response and is given by

$$\varepsilon(\omega) = W(\omega)\left[\breve{H}(\omega) - D(\omega)\right]$$

A positive weighting function

The desired amplitude response

- The Parks-McClellan algorithm is based on iteratively adjusting the coefficients of the amplitude response until the peak absolute value of $\varepsilon(\omega)$ is minimized.

Computer-Aided Design of Equiripple Linear-Phase FIR Filters

- If the minimum value of the peak absolute value of $\varepsilon(\omega)$ in a band $\omega_a \leq \omega \leq \omega_b$ is ε_0, then the absolute error satisfies

$$\left|\breve{H}(\omega) - D(\omega)\right| \leq \frac{\varepsilon_0}{|W(\omega)|}, \omega_a \leq \omega \leq \omega_b$$

- In typical filter design applications, the desired amplitude response is given by

$$D(\omega) = \begin{cases} 1, & \text{in the passband} \\ 0, & \text{in the stopband} \end{cases}$$

- The amplitude response $\breve{H}(\omega)$ is required to satisfy the above desired response with a ripple of $\pm\delta_p$ in the passband and a ripple δ_s in the stopband.
- As a result, it is evident from the weighted error function that the weighting function can be chosen either as

$$W(\omega) = \begin{cases} 1, & \text{in the passband} \\ \delta_p/\delta_s, & \text{in the stopband} \end{cases}$$ or $$W(\omega) = \begin{cases} \delta_s/\delta_p, & \text{in the passband} \\ 1, & \text{in the stopband} \end{cases}$$

Linear-Phase FIR Transfer Functions

- It is nearly impossible to design a linear-phase IIR transfer function.
- It is always possible to design an FIR transfer function with an exact linear-phase response.
- Consider a causal FIR transfer function $H(z)$ of length $N+1$, i.e., of order N as follows:

$$H(z) = \sum_{n=0}^{N} h[n]z^{-n}$$

- The above transfer function has a linear phase, if its impulse response $h[n]$ is either ***symmetric***, i.e.,

$$h[n] = h[N-n], 0 \leq n \leq N$$

  or is ***antisymmetric***, i.e.,

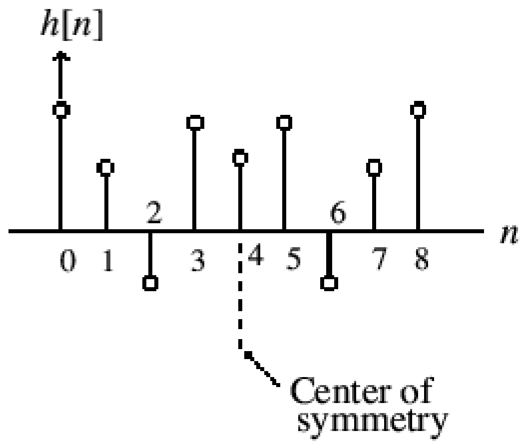
$$h[n] = -h[N-n], 0 \leq n \leq N$$

- Since the length of the impulse response can be either even or odd, we can define four types of linear-phase FIR transfer functions.
- For an antisymmetric FIR filter of odd length, i.e., N even

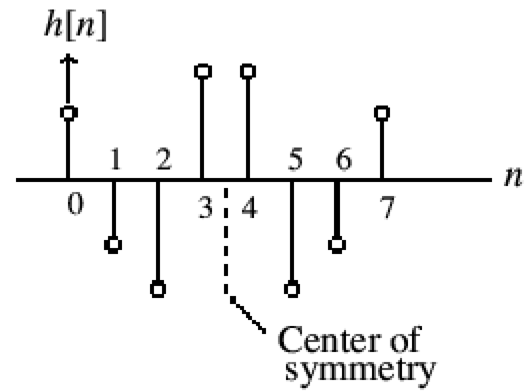
$$h[N/2] = 0$$

Imperial College
London

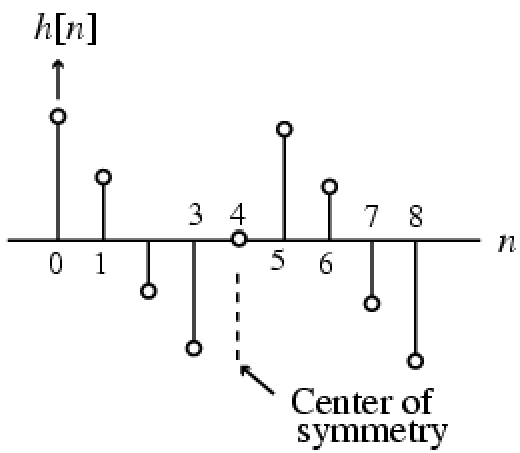
4 Types of Linear-Phase FIR Transfer Functions



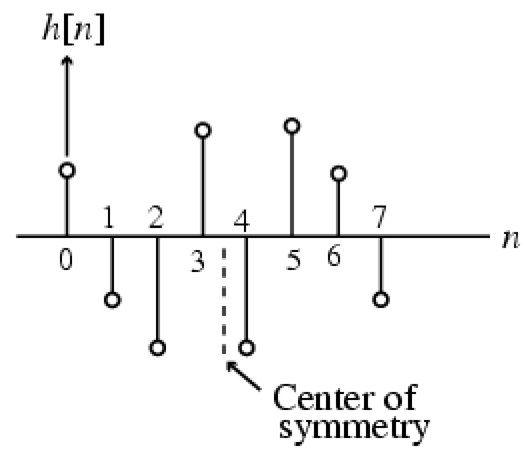
Type 1: $N = 8$



Type 2: $N = 7$



Type 3: $N = 8$



Type 4: $N = 7$

4 Types of Linear-Phase FIR Transfer Functions
Amplitude Response of Type 1

- By a clever manipulation, the expression for the amplitude response for each of the four types of linear-phase FIR filters can be expressed in the same form.
- The same algorithm can be adapted to design any one of the four types of filters.
- To develop this general form for the amplitude response expression, we consider each of the four types of filters separately.
- For the **Type 1 linear-phase FIR filter**, the amplitude response can be rewritten using the notation $N = 2M$ in the form

$$\breve{H}(\omega) = \sum_{k=0}^{M} a[k] \cos(\omega k)$$

$$a[0] = h[M], \quad a[k] = 2h[M-k], \quad 1 \le k \le M$$

Imperial College London

4 Types of Linear-Phase FIR Transfer Functions
Amplitude Response of Type 2

- For the **Type 2 linear-phase FIR filter**, the amplitude response can be rewritten using the notation $N = 2M$ in the form

$$\breve{H}(\omega) = \sum_{k=1}^{(2M+1)/2} b[k] \cos\left(\omega\left(k - \frac{1}{2}\right)\right)$$

$$b[k] = 2h\left[\frac{2M+1}{2} - k\right], \quad 1 \leq k \leq \frac{2M+1}{2}$$

- The above can also be expressed in the form:

$$\breve{H}(\omega) = \cos\left(\frac{\omega}{2}\right) \sum_{k=1}^{(2M-1)/2} \tilde{b}[k] \cos(\omega k)$$

where

$$b[1] = \frac{1}{2}\left(\tilde{b}[1] + 2\tilde{b}[0]\right)$$

$$b[k] = \frac{1}{2}\left(\tilde{b}[k] + \tilde{b}[k-1]\right), \quad 2 \leq k \leq \frac{2M-1}{2}$$

$$b\left[\frac{2M+1}{2}\right] = \frac{1}{2}\tilde{b}\left[\frac{2M-1}{2}\right]$$

4 Types of Linear-Phase FIR Transfer Functions
Amplitude Response of Type 3

- For the **Type 3 linear-phase FIR filter**, the amplitude response can be rewritten using the notation $N = 2M$ in the form

$$\breve{H}(\omega) = \sum_{k=1}^{M} c[k]\sin(\omega k)$$

$$c[k] = 2h[M-k], \quad 1 \le k \le M$$

- The above can also be expressed in the form:

$$\breve{H}(\omega) = \sin(\omega)\sum_{k=0}^{M-1} \tilde{c}[k]\cos(\omega k)$$

where

$$c[1] = \tilde{c}[0] - \frac{1}{2}\tilde{c}[1]$$

$$c[k] = \frac{1}{2}(\tilde{c}[k-1] - \tilde{c}[k]), \quad 2 \le k \le M-1$$

$$c[M] = \frac{1}{2}\tilde{c}[M-1]$$

4 Types of Linear-Phase FIR Transfer Functions
Amplitude Response of Type 4

- For the **Type 4 linear-phase FIR filter**, the amplitude response can be rewritten using the notation $N = 2M$ in the form

$$\breve{H}(\omega) = \sum_{k=1}^{(2M+1)/2} d[k] \sin\left(\omega\left(k - \frac{1}{2}\right)\right)$$

$$d[k] = 2h\left[\frac{2M+1}{2} - k\right], \quad 1 \leq k \leq \frac{2M+1}{2}$$

- The above can also be expressed in the form:

$$\breve{H}(\omega) = \sin\left(\frac{\omega}{2}\right) \sum_{k=1}^{(2M-1)/2} \tilde{d}[k] \cos(\omega k)$$

where

$$d[1] = \tilde{d}[0] - \frac{1}{2}\tilde{d}[1]$$

$$d[k] = \frac{1}{2}\left(\tilde{d}[k-1] - \tilde{d}[k]\right), \quad 2 \leq k \leq \frac{2M-1}{2}$$

$$d\left[\frac{2M+1}{2}\right] = \tilde{d}\left[\frac{2M-1}{2}\right]$$

Amplitude response of linear-phase FIR filters: Generic Form

- The amplitude response for all four types of linear-phase FIR filters can be expressed in the form

$$\breve{H}(\omega) = Q(\omega)A(\omega)$$

$$Q(\omega) = \begin{cases} 1, & \text{for Type 1} \\ \cos(\omega/2), & \text{for Type 2} \\ \sin(\omega), & \text{for Type 3} \\ \sin(\omega/2), & \text{for Type 4} \end{cases}$$

$$A(\omega) = \sum_{k=0}^{L} \tilde{a}[k]\cos(\omega k)$$

$$\tilde{a}[k] = \begin{cases} a[k], & \text{for Type 1} \\ \tilde{b}[k], & \text{for Type 2} \\ \tilde{c}[k], & \text{for Type 3} \\ \tilde{d}[k], & \text{for Type 4} \end{cases} \qquad L = \begin{cases} M, & \text{for Type 1} \\ \dfrac{2M-1}{2}, & \text{for Type 2} \\ M-1, & \text{for Type 3} \\ \dfrac{2M-1}{2}, & \text{for Type 4} \end{cases}$$

Linear-Phase FIR Filter Design by Optimisation

- The amplitude response for all 4 types of linear-phase FIR filters can be expressed as

$$\breve{H}(\omega) = Q(\omega)A(\omega)$$

- Before, we gave the weighted error function as

$$\varepsilon(\omega) = W(\omega)\left[\breve{H}(\omega) - D(\omega)\right]$$

- The modified form of the weighted error function is now

$$\varepsilon(\omega) = W(\omega)[Q(\omega)A(\omega) - D(\omega)] = W(\omega)Q(\omega)\left[A(\omega) - \frac{D(\omega)}{Q(\omega)}\right]$$
$$= \widetilde{W}(\omega)[A(\omega) - \widetilde{D}(\omega)]$$

where

$$\widetilde{W}(\omega) = W(\omega)Q(\omega)$$
$$\widetilde{D}(\omega) = D(\omega)/Q(\omega)$$

Optimisation Problem

- **Problem formulation**

  Determine $\tilde{a}[k]$ which minimise the peak absolute value of

$$\varepsilon(\omega) = \widetilde{W}(\omega)[\sum_{k=0}^{L} \tilde{a}[k]\cos(\omega k) - \widetilde{D}(\omega)]$$

  over the specified frequency bands $\omega \in R$.
- After $\tilde{a}[k]$ has been determined, construct the original $A(e^{j\omega})$ and hence $h[n]$.
- Solution is obtained via the so called ***Alternation Theorem***.
- The optimal solution has equiripple behavior, consistent with the total number of available parameters.
- Parks and McClellan used the ***Remez*** algorithm to develop a procedure for designing linear FIR digital filters.

The Parks-McClellan Algorithm

- **Problem formulation**

  Determine $\tilde{a}[k]$ which minimise the peak absolute value of

$$\varepsilon(\omega) = \widetilde{W}(\omega)[\sum_{k=0}^{L} \tilde{a}[k] \cos(\omega k) - \widetilde{D}(\omega)]$$

- Parks and McClellan solved the above problem applying the following theorem from the theory of Chebyshev approximation.

  Alternation Theorem: The amplitude function $A(\omega)$ is the best unique approximation of the desired amplitude response obtained by minimizing the peak absolute value ε of $\varepsilon(\omega)$, if and only if there exist at least $L+2$ extremal angular frequencies $\omega_0, \omega_1, \ldots, \omega_{L+1}$, in a closed subset R of the frequency range $0 \leq \omega \leq \pi$ such that $\omega_0 < \omega_1 < \cdots < \omega_L < \omega_{L+1}$ and $\varepsilon(\omega_i) = -\varepsilon(\omega_{i+1})$, with $|\varepsilon(\omega_i)| = \varepsilon$ for all i in the range $0 \leq i \leq L+1$.

The Parks-McClellan Algorithm

- Let us examine the behaviour of the amplitude response for a Type I equiripple lowpass FIR filter whose approximation error $\varepsilon(\omega)$ satisfies the condition of the alternation theorem.
- The peaks of $\varepsilon(\omega)$ are at $\omega = \omega_i$, $0 \leq i \leq L+1$, where

$$\frac{d\varepsilon(\omega)}{d\omega} = 0$$

- Since in the passband and the stopband, $\widetilde{W}(\omega)$ and $\widetilde{D}(\omega)$ are piecewise constant, we see that

$$\left.\frac{d\varepsilon(\omega)}{d\omega}\right|_{\omega=\omega_i} = \left.\frac{dA(\omega)}{d\omega}\right|_{\omega=\omega_i} = 0$$

  or, in other words, the amplitude response $A(\omega)$ also has peaks at $\omega = \omega_i$.
- We use the relation $\cos(\omega k) = T_k(\cos\omega)$ where $T_k(x)$ is the kth order Chebyshev polynomial defined by

$$T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x),\ T_0(x) = 0, T_1(x) = 1$$

  The amplitude response $A(\omega)$ can be expressed as a power series in $\cos\omega$

$$A(\omega) = \sum_{k=0}^{L} a[k](\cos\omega)^k$$

Chebyshev Polynomial Revision

- Chebyshev polynomials of 1st kind:
  $T_0(x) = 0$
  $T_1(x) = 1$
  $T_2(x) = 2x^2 - 1$
  $T_3(x) = 4x^3 - 3x$
  $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$

  We know that
  $\cos 2\omega = 2\cos^2\omega - 1 = T_2(\cos\omega)$
  $\cos 3\omega = 4\cos^3\omega - 3\cos\omega = T_3(\cos\omega)$

  It is proven that
  $\cos k\omega = T_k(\cos\omega)$

  The amplitude response $A(\omega)$ can be expressed as a power series in $\cos\omega$.

$$A(\omega) = \sum_{k=0}^{L} a[k](\cos\omega)^k$$

The Parks-McClellan Algorithm

- The amplitude response $A(\omega)$ can be expressed as a power series in $\cos\omega$

$$A(\omega) = \sum_{k=0}^{L} a[k](\cos\omega)^k$$

- It is an Lth order polynomial in $\cos\omega$.
- As a result $A(\omega)$ can have at most $L-1$ minima and maxima inside the specified passband and stopband.
- Moreover, at the band edges, $\omega = \omega_p$ and $\omega = \omega_s$, $|\varepsilon(\omega)|$ is maximum and therefore, $A(\omega)$ has extrema in these angular frequencies.
- In addition $A(\omega)$ may also have extrema at $\omega = 0$ and $\omega = \pi$.
- Therefore, there are, at most $L+3$ extremal frequencies of $\varepsilon(\omega)$.
- We can generalize and say that in the case of a linear phase FIR filter with K specified band edges and designed using the Remez exchange algorithm, there can be at most $L+K+1$ extremal frequencies.
- To arrive at the optimum solution we need to solve the set of $L+2$ equations:

$$\widetilde{W}(\omega_i)[A(\omega_i) - \widetilde{D}(\omega_i)] = (-1)^i\varepsilon,\ 0 \le i \le L+1$$

for the unknowns $\tilde{a}(i)$ and ε, provided the $L+2$ extremal angular frequencies are known.

Imperial College London

The Parks-McClellan Algorithm

- To arrive at the optimum solution we need to solve the set of $L+2$ equations:

$$\widetilde{W}(\omega_i)[A(\omega_i)-\widetilde{D}(\omega_i)] = (-1)^i\varepsilon,\ 0 \le i \le L+1$$

for the unknowns $\tilde{a}(i)$ and ε, provided the $L+2$ extremal angular frequencies are known.

- The above is rewritten in matrix form as

$$\begin{bmatrix} 1 & \cos(\omega_0) & \dots & \cos(L\omega_0) & \frac{-1}{\widetilde{W}(\omega_0)} \\ 1 & \cos(\omega_1) & \dots & \cos(L\omega_1) & \frac{-1}{\widetilde{W}(\omega_1)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ & & & & \frac{(-1)^{L-1}}{\widetilde{W}(\omega_L)} \\ 1 & \cos(\omega_L) & \dots & \vdots & \\ 1 & \cos(\omega_{L+1}) & \dots & \cos(L\omega_{L+1}) & \frac{(-1)^{L-1}}{\widetilde{W}(\omega_{L+1})} \end{bmatrix} \begin{bmatrix} \tilde{a}[0] \\ \tilde{a}[1] \\ \vdots \\ \tilde{a}[L] \\ \varepsilon \end{bmatrix} = \begin{bmatrix} \widetilde{D}(\omega_0) \\ \widetilde{D}(\omega_1) \\ \vdots \\ \widetilde{D}(\omega_L) \\ \widetilde{D}(\omega_{L+1}) \end{bmatrix}$$

- The ***Remez Exchange Algorithm*** is used to solve the above.

The Parks-McClellan Algorithm

- The Remez exchange algorithm, a highly efficient iterative procedure, is used to determine the locations of the extremal frequencies and consists of the following steps at each iteration stage.
- **Step 1:** A set of initial values for the extremal frequencies are either chosen or are available from the completion of the previous iteration.
- **Step 2:** Solving the system of equations we obtain

$$\varepsilon = \frac{c_0\widetilde{D}(\omega_0) + c_1\widetilde{D}(\omega_1) + \cdots + c_{L+1}\widetilde{D}(\omega_{L+1})}{\frac{c_0}{\widetilde{W}(\omega_0)} - \frac{c_1}{\widetilde{W}(\omega_1)} + \cdots + \frac{(-1)^{L-1}c_{L+1}}{\widetilde{W}(\omega_{L+1})}}$$

$$c_n = \prod_{\substack{i=0 \\ i\neq n}}^{L+1} \frac{1}{\cos(\omega_n) - \cos(\omega_i)}$$

The Parks-McClellan Algorithm

- **Step 3:** The values of the amplitude response $A(\omega)$ at $\omega = \omega_i$ are then computed using

$$A(\omega_i) = \frac{(-1)^i \varepsilon}{\widetilde{W}(\omega_i)} + \widetilde{D}(\omega_i),\ 0 \le i \le L+1$$

- **Step 4:** The polynomial $A(\omega)$ is determined by interpolating the above values at the $L + 2$ extremal frequencies using the Lagrange interpolation formula:

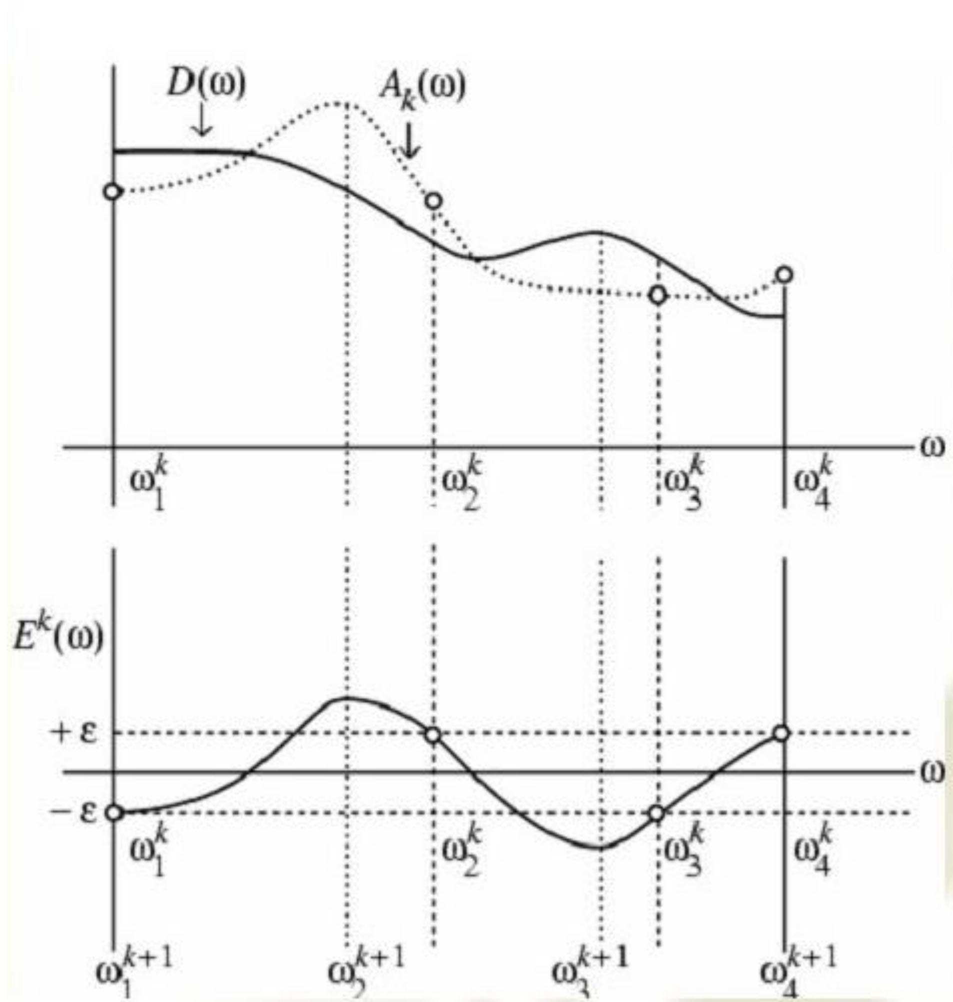
$$A(\omega) = \sum_{i=0}^{L+1} A(\omega_i)\, P_i(\cos\omega)$$

where $P_i(\cos\omega) = \prod_{\substack{l=0 \\ l\neq i}}^{L+1} \left(\frac{\cos\omega - \cos\omega_l}{\cos\omega_i - \cos\omega_l}\right),\ 0 \le i \le L+1$

- **Step 5:** The new weighted error function $\varepsilon(\omega)$ is computed at a dense set $S (S \ge L)$ of frequencies. In practice, $S = 16L$ is adequate. Determine the $L + 2$ new extremal frequencies from the values of $\varepsilon(\omega)$ evaluated at the dense set of frequencies.
- **Step 6:** If the peak values ε are equal in magnitude, the algorithm has converged. Otherwise, we go bask to Step 2.

The Parks-McClellan Algorithm

- Plots of the desired response $D(\omega)$, the amplitude response $A_k(\omega)$ and the error $\varepsilon^k(\omega)$ at the end of the kth iteration. The locations of the new extremal frequencies are given by ω_i^{k+1}.
- The iteration process is stopped after the difference between the value of the peak error ε calculated at any stage and that at the previous stage is below a present threshold value, such as 10^{-6}.
- In practice the process converges after very few iterations.

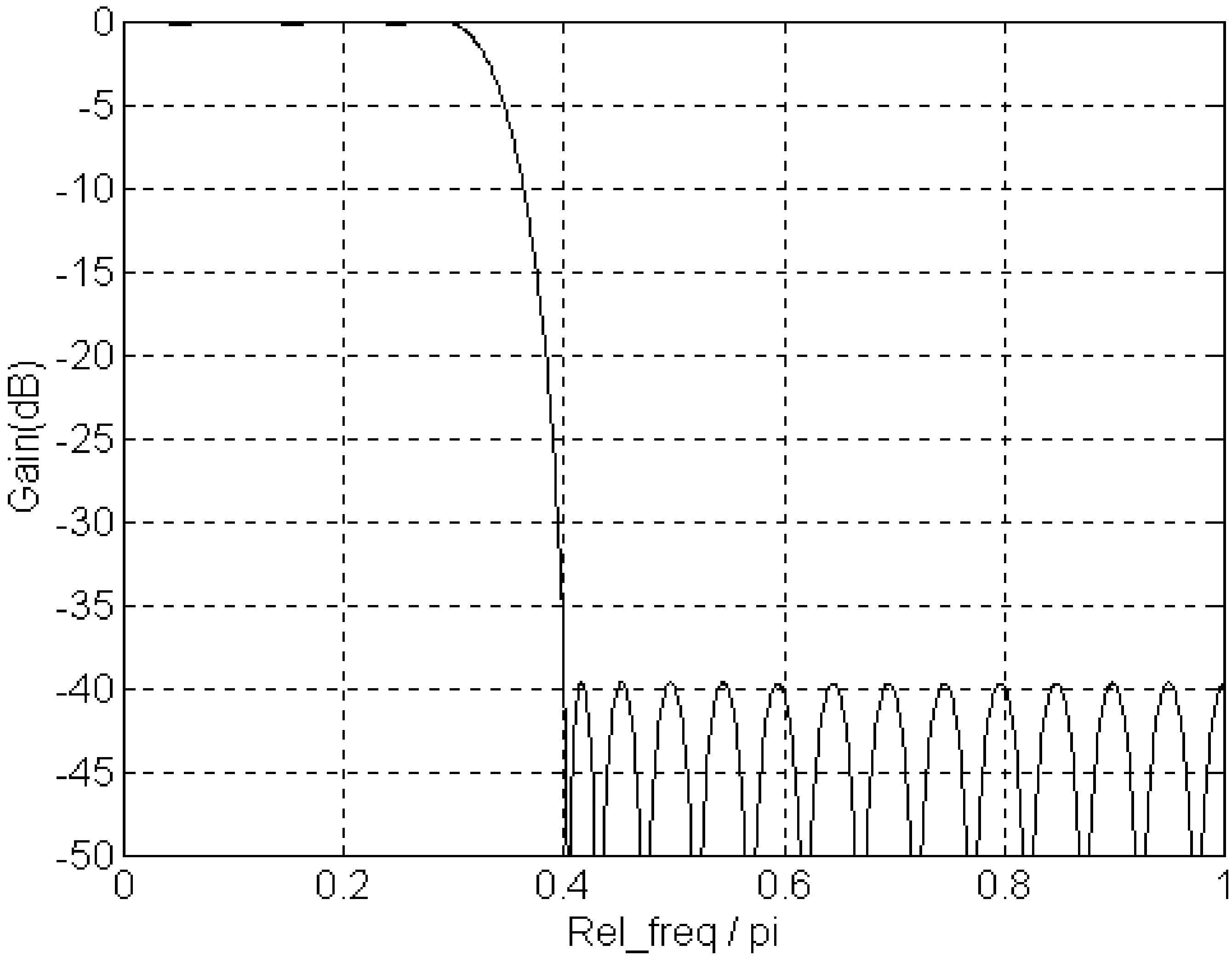


Imperial College London

Remez Exchange Algorithm

- Better than windowing technique, but more complicated.
- Available in MATLAB.
- Design 40th order FIR lowpass filter whose gain is unity (0 dB) in range 0 to 0.3π radians/sample & zero in range 0.4π to π.
- The 41 coefficients will be found in array ‘a’.
- Produces equiripple gain-responses where peaks of stop-band ripples are equal rather than decreasing with increasing frequency.
- Highest peak in stop-band lower than for FIR filter of same order designed by windowing technique to have same cut-off rate.
- There are equiripple pass-band ripples.

```
a = remez (40, [0, 0.3,  0.4,1],[1, 1,  0, 0] );
h = freqz (a,1,1000);
plot([0:999]/1000,20*log10(abs(h)),'k');
axis([0,1,-50,0]);
grid on;
xlabel('Rel_freq / pi');
ylabel('Gain(dB)');
```



Gain of 40th order FIR lowpass filter designed by “ Remez ”