7: Optimal FIR filters

We restrict ourselves to zero-phase filters of odd length $M+1$, symmetric around $h[0]$, i.e. $h[-n] = h[n]$.

$$\overline{H}(\omega) = H(e^{j\omega}) = \sum_{-\frac{M}{2}}^{\frac{M}{2}} h[n]e^{-jn\omega} = h[0] + 2\sum_{1}^{\frac{M}{2}} h[n]\cos n\omega$$

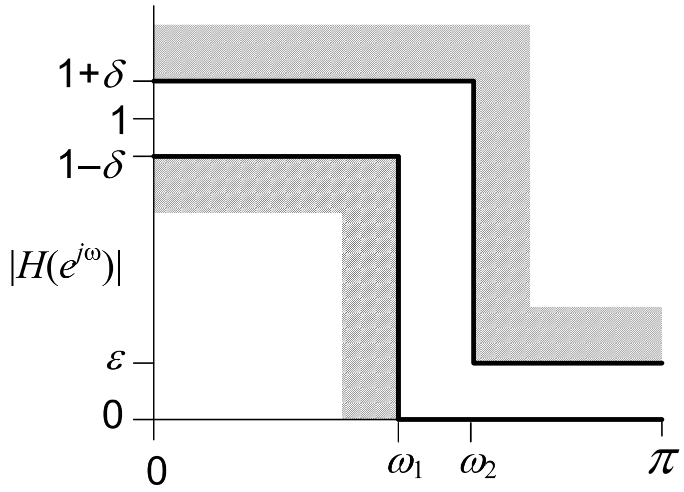
$\overline{H}(\omega)$ is real but not necessarily positive (unlike $\left|H(e^{j\omega})\right|$).

Weighted error: $e(\omega) = s(\omega)\left(\overline{H}(\omega) - d(\omega)\right)$ where $d(\omega)$ is the target.
Choose $s(\omega)$ to control the error variation with ω.

Example: lowpass filter

$$d(\omega) = \begin{cases} 1 & 0 \le \omega \le \omega_1 \\ 0 & \omega_2 \le \omega \le \pi \end{cases}$$

$$s(\omega) = \begin{cases} \delta^{-1} & 0 \le \omega \le \omega_1 \\ \epsilon^{-1} & \omega_2 \le \omega \le \pi \end{cases}$$



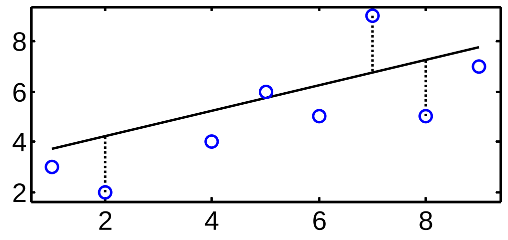
$e(\omega) = \pm 1$ when $\overline{H}(\omega)$ lies at the edge of the specification.

Minimax criterion: $h[n] = \arg\min_{h[n]} \max_{\omega} |e(\omega)|$: minimize max error

Alternation Theorem

Want to find the best fit line: with the smallest maximal error.

Best fit line always attains the maximal error three times with alternate signs



Proof:

Assume the first maximal deviation from the line is negative as shown.
There must be an equally large positive deviation; or else just move the line downwards to reduce the maximal deviation.
This must be followed by another maximal negative deviation; or else you can rotate the line and reduce the deviations.

Alternation Theorem:

A polynomial fit of degree n to a set of bounded points is minimax if and only if it attains its maximal error at $n+2$ points with alternating signs.
There may be additional maximal error points.
Fitting to a continuous function is the same as to an infinite number of points.

$$\overline{H}(\omega) = H(e^{j\omega}) = h[0] + 2\sum_{1}^{\frac{M}{2}} h[n]\cos n\omega$$

But $\cos n\omega = T_n(\cos\omega)$: Chebyshev polynomial of 1st kind

$$\cos 2\omega = 2\cos^2\omega - 1 = T_2(\cos\omega) \qquad T_2(x) = 2x^2 - 1$$
$$\cos 3\omega = 4\cos^3\omega - 3\cos\omega = T_3(\cos\omega) \qquad T_3(x) = 4x^3 - 3x$$

Recurrence Relation:

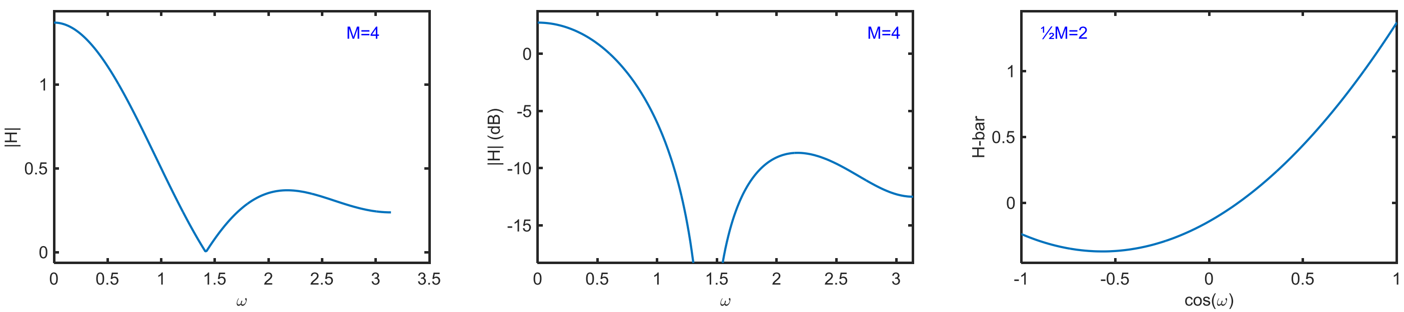
$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$ with $T_0(x) = 1$, $T_1(x) = x$

Proof: $\cos(n\omega + \omega) + \cos(n\omega - \omega) = 2\cos\omega\cos n\omega$

So $\overline{H}(\omega)$ is an $\frac{M}{2}$ order polynomial in $\cos\omega$: alternation theorem applies.

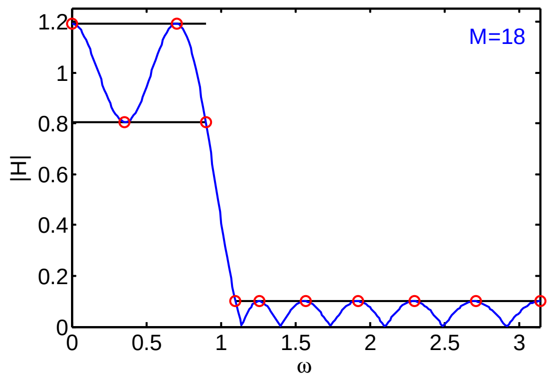
Example: Symmetric lowpass filter of order $M = 4$

$$H(z) = 0.1766z^2 + 0.4015z + 0.2124 + 0.4015z^{-1} + 0.1766z^{-2}$$



Maximal Error Locations

Maximal error locations occur either at band edges or when $\frac{d\overline{H}}{d\omega}=0$



$$\begin{aligned}\overline{H}(\omega) &= h[0]+2\sum_1^{\frac{M}{2}} h[n]\cos n\omega \\ &= P(\cos\omega)\end{aligned}$$

where $P(x)$ is a polynomial of order $\frac{M}{2}$.

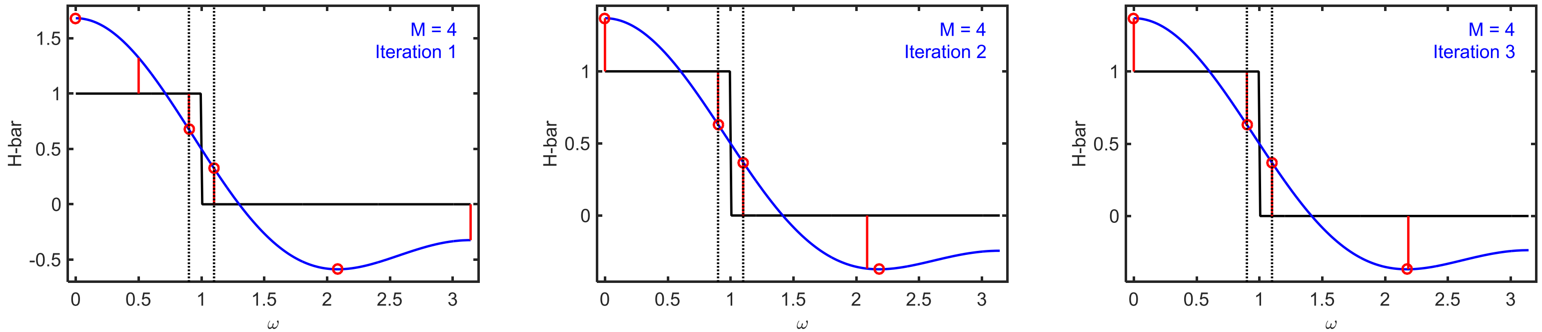
$$\begin{aligned}\frac{d\overline{H}}{d\omega} &= -P'(\cos\omega)\sin\omega \\ &= 0 \text{ at } \omega=0,\ \pi \text{ and at most } \frac{M}{2}-1 \text{ zeros of polynomial } P'(x).\end{aligned}$$

$\therefore$ With two bands, we have at most $\frac{M}{2}+3$ maximal error frequencies.
We require $\frac{M}{2}+2$ of alternating signs for the optimal fit.

Only three possibilities exist (try them all):

(a) $\omega=0$ + two band edges + all $\left(\frac{M}{2}-1\right)$ zeros of $P'(x)$.
(b) $\omega=\pi$ + two band edges + all $\left(\frac{M}{2}-1\right)$ zeros of $P'(x)$.
(c) $\omega=\{0 \text{ and } \pi\}$ + two band edges + $\left(\frac{M}{2}-2\right)$ zeros of $P'(x)$.

1. Guess the positions of the $\frac{M}{2}+2$ maximal error frequencies and give alternating signs to the errors (e.g. choose evenly spaced ω).
2. Determine the error magnitude, ϵ, and the $\frac{M}{2}+1$ coefficients of the polynomial that passes through the maximal error locations.
3. Find the local maxima of the error function by evaluating $e(\omega) = s(\omega)\left(\overline{H}(\omega) - d(\omega)\right)$ on a dense set of ω.
4. Update the maximal error frequencies to be an alternating subset of the local maxima + band edges + $\{0$ and/or $\pi\}$.
   If maximum error is $> \epsilon$, go back to step 2. (typically 15 iterations)
5. Evaluate $\overline{H}(\omega)$ on $M+1$ evenly spaced ω and do an IDFT to get $h[n]$.



Remex Step 2: Determine Polynomial

For each extremal frequency, ω_i for $1 \le i \le \frac{M}{2}+2$

$$d(\omega_i) = \overline{H}(\omega_i) + \frac{(-1)^i \epsilon}{s(\omega_i)} = h[0] + 2\sum_{n=1}^{\frac{M}{2}} h[n]\cos n\omega_i + \frac{(-1)^i \epsilon}{s(\omega_i)}$$

Method 1: (Computation time $\propto M^3$)

Solve $\frac{M}{2}+2$ equations in $\frac{M}{2}+2$ unknowns for $h[n] + \epsilon$.

In step 3, evaluate $\overline{H}(\omega) = h[0] + 2\sum_{n=1}^{\frac{M}{2}} h[n]\cos n\omega_i$

Method 2: Don't calculate $h[n]$ explicitly (Computation time $\propto M^2$)

Multiply the ω_i equation by $c_i = \prod_{j\ne i} \frac{1}{\cos\omega_i - \cos\omega_j}$ and add them:

$$\sum_{i=1}^{\frac{M}{2}+2} c_i \left(h[0] + 2\sum_{n=1}^{\frac{M}{2}} h[n]\cos n\omega + \frac{(-1)^i \epsilon}{s(\omega_i)}\right) = \sum_{i=1}^{\frac{M}{2}+2} c_i d(\omega_i)$$

All terms involving $h[n]$ sum to zero leaving

$$\sum_{i=1}^{\frac{M}{2}+2} \frac{(-1)^i c_i}{s(\omega_i)} \epsilon = \sum_{i=1}^{\frac{M}{2}+2} c_i d(\omega_i)$$

Solve for ϵ then calculate the $\overline{H}(\omega_i)$ then use Lagrange interpolation:

$$\overline{H}(\omega) = P(\cos\omega) = \sum_{i=1}^{\frac{M}{2}+2} \overline{H}(\omega_i) \prod_{j\ne i} \frac{\cos\omega - \cos\omega_j}{\cos\omega_i - \cos\omega_j}$$

$\left(\frac{M}{2}+1\right)$-polynomial going through all the $\overline{H}(\omega_i)$ [actually order $\frac{M}{2}$]

Example Design

Filter Specifications:

- Bandpass $\omega = [0.5, 1]$, Transition widths: $\Delta\omega = 0.2$
- Stopband Attenuation: -25 dB and -15 dB
- Passband Ripple: ± 0.3 dB

Determine gain tolerances for each band:

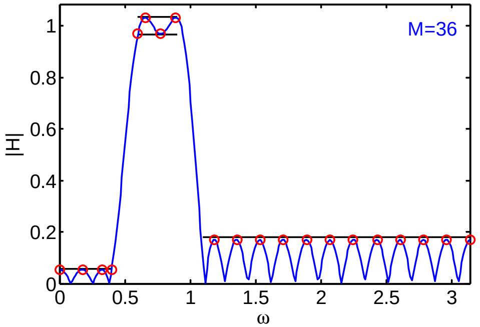
- $-25 \text{ dB} = 0.056$, $-0.3 \text{ dB} = 1 - 0.034$, $-15 \text{ dB} = 0.178$

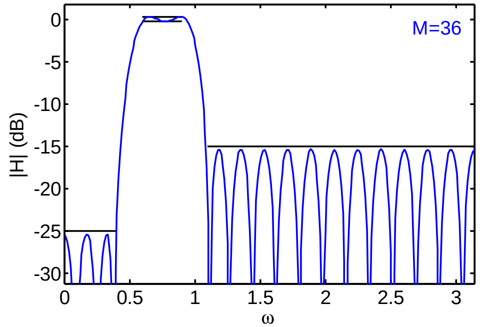
Predicted order: $M = 36$

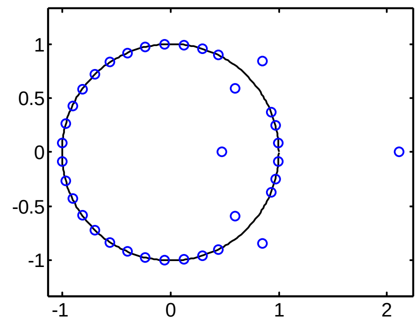
- $\frac{M}{2} + 2$ extremal frequencies are distributed between the bands
- Filter meets specs ☺; clearer on a decibel scale

Most zeros are on the unit circle + three reciprocal pairs

- Reciprocal pairs give a linear phase shift







FIR Pros and Cons

- Can have linear phase
  - no envelope distortion, all frequencies have the same delay ☺
  - symmetric or antisymmetric: $h[n] = h[-n] \forall n$ or $-h[-n] \forall n$
  - antisymmetric filters have $H(e^{j0}) = H(e^{j\pi}) = 0$
  - symmetry means you only need $\frac{M}{2} + 1$ multiplications to implement the filter.
- Always stable ☺
- Low coefficient sensitivity ☺
- Optimal design method fast and robust ☺
- Normally needs higher order than an IIR filter ☹
  - Filter order $M \approx \frac{\text{dB}_{\text{atten}}}{3.5\Delta\omega}$ where $\Delta\omega$ is the most rapid transition
  - Filtering complexity $\propto M \times f_s \approx \frac{\text{dB}_{\text{atten}}}{3.5\Delta\omega} f_s = \frac{\text{dB}_{\text{atten}}}{3.5\Delta\Omega} f_s^2$ $\propto f_s^2$ for a given specification in unscaled Ω units.

Summary

Optimal Filters: minimax error criterion

- use weight function, $s(\omega)$, to allow different errors in different frequency bands
- symmetric filter has zeros on unit circle or in reciprocal pairs
- Response of symmetric filter is a polynomial in $\cos\omega$
- Alternation Theorem: $\frac{M}{2}+2$ maximal errors with alternating signs

Remez Exchange Algorithm (also known as Parks-McLellan Algorithm)

- multiple constant-gain bands separated by transition regions
- very robust, works for filters with $M > 1000$
- Efficient: computation $\propto M^2$
- can go mad in the transition regions

Modified version works on arbitrary gain function

- Does not always converge

For further details see Mitra: 10.

MATLAB routines

firpm	optimal FIR filter design
firpmord	estimate require order for firpm
cfirpm	arbitrary-response filter design
remez	[obsolete] optimal FIR filter design