FIR digital filter design

- Filter design specifications (*Porat ch.8*)
- Linear phase filters
- Design of FIR filters via truncated response, Gibb’s phenomenon (*Porat ch.9*)
- Implementation of long FIR filters by FFT (*Porat ch. 4.7, 5.6*)
Digital filter design (Porat ch. 8)

- Rational filter:

\[ H^z(z) = \sum_{k=-\infty}^{\infty} h[k]z^{-k} = \frac{b(z)}{a(z)} = \frac{b_0 + b_1z^{-1} + \cdots + b_qz^{-q}}{1 + a_1z^{-1} + \cdots + a_pz^{-p}} \]

- \( p \geq 1 \): IIR filter; \( p = 0 \): FIR filter

- Causal if \( h[n] = 0 \) for \( n < 0 \).

  Stable+causal if the zeros of \( a(z) \) are located inside unit circle.

- Frequency response: choose \( z = e^{j\theta} \)

\[ H^f(\theta) = H^z(e^{j\theta}), \quad -\pi \leq \theta < \pi \]

- Write \( H^f(\theta) = |H^f(\theta)|e^{j\psi(\theta)} \), with \( -\pi \leq \psi(\theta) < \pi \)

  \( |H^f(\theta)| \): Magnitude response;
  \( \psi(\theta) \): Phase response
Digital filter design

Example: \( H^z(z) = 1 - z^{-1}, \quad z = e^{j\theta} \)

Note: phase is discontinuous
**Digital filter design**

Discontinuous phase for two reasons:

– jumps of $2\pi$ due to modulo $2\pi$ definition of angle

– jumps of $\pi$ due to positivity of magnitude $|H^f(\theta)|$

**Phase unwrapping procedure (to make it continuous):**

– at a discontinuity at $\theta_0$ of $2\pi$: add/subtract $2\pi$ to $\psi(\theta)$ for $\theta > \theta_0$.

– at a discontinuity at $\theta_0$ of $\pi$: add/subtract $\pi$ to $\psi(\theta)$ for $\theta > \theta_0$, multiply magnitude by $-1$ for $\theta > \theta_0$.

This leads to

$$H^f(\theta) = A(\theta)e^{j\phi(\theta)}$$

$A(\theta)$ is real: *Amplitude response*

$\phi(\theta)$: *Continuous phase response* (not unique; constrain $0 \leq \phi(0) < \pi$)
For $H(z) = 1 - z^{-1}$, the amplitude and continuous phase response are:
Various techniques for digital filter design:

- Specify a desired magnitude response $|H(\theta)|$, and find the corresponding impulse response $h[n]$. For the phase we require a linear response. We will obtain an FIR filter.

- First design an analog filter based on the given specs. Then transform this to the digital domain, e.g. by
  - sampling of the analog impulse response ("impulse invariance")
  - bilinear transformation $s \rightarrow z$.

This yields an IIR filter.

IIR filters usually require a lower order, but they have a nonlinear phase (possible distortion).
Digital filter design specifications

Magnitude specifications

- Low-pass filter specifications:
  - Frequency range: $0 \leq \omega \leq \pi$
  - Passband ripple: $1 \pm \delta_p$
  - Transition band: $\omega_p < \omega < \omega_s$
  - Stopband: $\omega_s < \omega < \pi$

- Stopband ripple: $\delta_s$

Similarly: high-pass, band-pass (filterbank), band-stop, multi-band filters

An ideal filter does not have ripples and no transition band. However, a causal filter has a finite number of zeros and cannot be ideal.

Usually only the magnitude response is specified, because this almost completely determines the phase response (spectral factorization).
Linear phase

- Pure delay \( \{ h \ast x \}[n] = x[n-L] \), with \( L \) integer:

\[
H^Z(z) = z^{-L}, \quad H^{f}(\theta) = e^{-jL\theta}
\]

Its continuous phase response is \( \phi(\theta) = -L\theta \) is linear.

A delay does not distort the input.

- Another non-distorting filter acting as delay:

\[
H^{f}(\theta) = \begin{cases} 
  e^{-jL\theta}, & \theta_1 \leq |\theta| \leq \theta_2 \\
  \text{arbitrary} & \text{otherwise}
\end{cases}
\]

If \( X^f(\theta) \neq 0 \) for \( \theta_1 \leq \theta \leq \theta_2 \), and \( = 0 \) otherwise, then \( x[n] \rightarrow x[n-L] \).

\( H \) is a delay for signals limited to a band.
Linear phase

Non-integer delay $\tau = L + \delta$: Let

$$H^f(\theta) = \begin{cases} 
e^{-j\tau \theta}, & \theta_1 \leq |\theta| \leq \theta_2 \\ \text{arbitrary} & \text{otherwise} \end{cases}$$

$$y[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X^f(\theta)e^{-j\tau \theta} e^{j\theta n} d\theta$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ \sum_{m=-\infty}^{\infty} x[m]e^{-j\theta m} \right] e^{j\theta (n-\tau)} d\theta$$

$$= \sum_{m=-\infty}^{\infty} x[m] \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\theta (n-m-\tau)} d\theta \right]$$

$$= \sum_{m=-\infty}^{\infty} x[m] \text{sinc}(n-m-\tau)$$

This is an interpolated-delay filter
Linear phase

**Definition**: a filter with frequency response

\[ H_f(\theta) = A(\theta)e^{-j\theta\tau_p} \]

has *linear phase*.

The filter acts as a delay for signals in the pass-band (where \( A(\theta) \approx 1 \)) and does not distort such signals.

In the stop-band, the phase response is not important.

**Definition** of phase delay:

\[ \tau_p(\theta) = -\frac{\phi(\theta)}{\theta} \]

A linear-phase filter has a constant phase delay.
Linear phase: filter constraints

Generalized linear phase (GLP) filter: \( H^f(\theta) = A(\theta) e^{j(\phi_0 - \theta \tau)} \)

In general, \( \tau_g(\theta) := -\frac{d\phi(\theta)}{d\theta} \) is called the *group delay*

**Constraints on realizable GLP filters:**

- \( H^f(\theta) \) must be periodic (period \( 2\pi \))
  \[
  A(\theta) e^{j(\phi_0 - \theta \tau)} = A(\theta + 2\pi) e^{j(\phi_0 - \theta \tau - 2\pi \tau_g)}
  \]
  \[\Rightarrow A(\theta) = A(\theta + 2\pi) e^{-j2\pi \tau_g} \]
  \[\Rightarrow 2\tau_g \text{ is integer: } \tau_g = M \text{ or } \tau_g = M + 0.5 \]

- \( h[n] \) must be real
  \[
  H^f(-\theta) = -\overline{H^f(\theta)}
  \]
  \[\Rightarrow A(-\theta) e^{j(\phi_0 + \theta \tau)} = A(\theta) e^{-j(\phi_0 - \theta \tau)} \]
  \[\Rightarrow e^{j2\phi_0} = \frac{A(-\theta)}{A(\theta)} \text{ is real} \]
  \[\Rightarrow \phi_0 = 0 \text{ and } A(-\theta) = A(\theta) \text{ or } \phi_0 = \frac{\pi}{2} \text{ and } A(-\theta) = -A(\theta) \]
Linear phase: filter constraints

- \( h \) must be causal: \( h[n] = 0, \ n < 0 \).

  If \( \phi_0 = 0 \):
  \[
  h[2\tau_g - n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} H^f(\theta) e^{j\theta(2\tau_g - n)} \, d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\theta) e^{j\theta(\tau_g - n)} \, d\theta
  \]
  \[
  = \overline{h[2\tau_g - n]} = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\theta) e^{j\theta(n - \tau_g)} \, d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} H^f(\theta) e^{j\theta n} \, d\theta = h[n]
  \]
  Hence \( h[n] = 0 \) for \( n > 2\tau_g \).

  \[\Rightarrow\] \text{h must be FIR of order } N = 2\tau_g, \text{ and symmetric: } h[n] = h[N - n]\]

  Similar, if \( \phi_0 = \frac{\pi}{2} \), then \( h \) must be FIR, anti-symmetric: \( h[n] = -h[N - n] \)

In summary: a distortionless filter must have linear phase in the passband, hence must be FIR (symmetric or anti-symmetric)
Digital filter design

- **Type I**: $h[n] = h[N-n]$, $N$ is even: $H(\theta) = e^{-j\theta N/2} \sum_k a_k \cos(\theta k)$

- **Type II**: $h[n] = h[N-n]$, $N$ is odd: $H(\theta) = e^{-j\theta N/2} \sum_k a_k \cos(\theta(k-1/2))$
  
  $H(\theta) = 0$ for $\theta = \pi$: cannot be a highpass filter

- **Type III**: $h[n] = -h[N-n]$, $N$ is even: $H(\theta) = je^{-j\theta N/2} \sum_k a_k \sin(\theta k)$
  
  $A(\theta)$ is antisymmetric and periodic with $2\pi$.

  $H(\theta) = 0$ for $\theta = 0, \pi$: cannot be a lowpass nor a highpass filter.

- **Type IV**: $h[n] = -h[N-n]$, $N$ is odd: $H(\theta) = je^{-j\theta N/2} \sum_k a_k \sin(\theta(k-1/2))$
  
  $A(\theta)$ is antisymmetric and periodic with $4\pi$.

  $H(\theta) = 0$ for $\theta = 0$: cannot be a lowpass filter

The phase delay is always equal to half the filter length (possibly fractional).
Linear phase filter

A linear phase filter must have zeros that are reciprocal (for symmetry) and conjugated (to be real)

E.g.

$$H(z) = (z - \beta)(1 - \beta z) = -\beta + (1 + \beta^2)z - \beta z^2$$

**Proof:** In general

$$H(z^{-1}) = \sum_{n=0}^{N} h[n]z^n = z^N \sum_{n=0}^{N} h[n]z^{-(N-n)} = \pm z^N \sum_{k=0}^{N} h[k]z^{-k} = \pm z^N H(z)$$

Hence if \(H(\beta) = 0\), also \(H(1/\beta) = 0\).
FIR filters are popular for digital filter design because

- they are inherently stable,
- can have linear phase (distortionless response in passband),
- are easily implementable,
- however, need more coefficients to reach same spec. performance as IIR
FIR filter design

Design example:

Design a lowpass filter with $\theta_p = 0.2\pi$, $\theta_s = 0.3\pi$, $\delta_p = \delta_s = 0.01$.

Approach (*truncated impulse response design*):

1. Formulate desired amplitude response $A_d(\theta)$
2. choose phase characteristic. Usually linear phase (select initial phase $\phi_0 = 0$ or $\frac{\pi}{2}$, i.e. a symmetric or anti-symmetric filter)
3. choose filter order $N$ (set group delay $\tau_g$ equal to $N/2$)
4. IDFT to obtain ideal impulse response $h_d[n]$
5. truncate to length $N$
For our example:

\[ A_d(\theta) = \begin{cases} 
1, & |\theta| < 0.25\pi \\
0, & \text{otherwise}
\end{cases} = \frac{1}{2}(\theta_p + \theta_s) \]

**Select a phase characteristic:** We choose a symmetric impulse response (constant phase delay), hence initial phase \( \phi_0 = 0 \). Choose group delay 0.5\( N \) since we will truncate to \([0, N]\). Thus \( \phi(\theta) = 0.5N\theta \).

**Desired frequency response is**

\[ H_d(\theta) = \begin{cases} 
e^{-j0.5N\theta}, & |\theta| < 0.25\pi \\
0, & \text{otherwise}
\end{cases} \]

Corresponding desired impulse response is

\[ h_d[n] = \frac{1}{2\pi} \int_{-0.25\pi}^{0.25\pi} e^{j\theta(n-0.5N)} \, d\theta = 0.25 \text{sinc}[0.25(n-0.5N)] \]
Note: the passband and stopband are OK, but the ripples are too large.
FIR filter design

Try to reduce the ripples by increasing the filter order $N$:

Increasing the filter order narrows the transition band, but does not reduce the amplitude of the ripples, only ‘compresses’ them! This is the Gibbs’ phenomenon.
Gibbs’ phenomenon

The truncation of the ‘ideal’ impulse response $h_d[n]$ to $h[n]$ is a form of rectangular windowing:

$$h[n] = h_d[n]w_r[n], \quad w_r[n] = \begin{cases} 1, & n = 0, \ldots, N \\ 0, & \text{otherwise} \end{cases}$$

It translates to a convolution in frequency domain: $H(\theta) = H_d(\theta) * W_r(\theta)$,

$$W_r(\theta) = \sum_{n=0}^{N} e^{-j\theta n} = \frac{1 - e^{-j(N+1)}}{1 - e^{-j\theta}} = \frac{\sin(0.5\theta(N+1))}{\sin(0.5\theta)} e^{-j0.5\theta N}$$

The function $D(\theta, L) = \frac{\sin(0.5\theta L)}{\sin(0.5\theta)}$ is called the Dirichlet kernel
Gibbs' phenomenon

Set $L = N + 1$, $\theta_0 = 0.25\pi$ (in our example)

$$A(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} A_d(\lambda) D(\theta - \lambda, L) d\lambda = \frac{1}{2\pi} \int_{-\theta_0}^{\theta_0} D(\theta - \lambda, L) d\lambda$$

Around $\theta \approx \theta_0$,

$$A(\theta) = \frac{1}{2\pi} \int_{-\theta_0}^{\theta} D(\theta - \lambda, L) d\lambda + \frac{1}{2\pi} \int_{\theta}^{\theta_0} D(\theta - \lambda, L) d\lambda$$

$$\approx 0.5 + \frac{1}{\pi} \int_{0}^{0.5L(\theta_0 - \theta)} \frac{\sin \mu}{\mu} d\mu$$

$$= 0.5 + \frac{1}{\pi} \text{Si}(0.5L(\theta_0 - \theta))$$

where $\text{Si}(x) = \int_{0}^{x} \frac{\sin \mu}{\mu} d\mu$
Gibbs’ phenomenon

\[ A(\theta) = \begin{cases} 
0.5 + \frac{1}{\pi} \text{Si}(0.5L(\theta_0 - \theta)) , & \theta < \theta_0 \\
0.5 - \frac{1}{\pi} \text{Si}(0.5L(\theta - \theta_0)) , & \theta > \theta_0 
\end{cases} \]

The transition width is \( 4\pi/L \), the ripple does not depend on \( L \) (about 0.09, or 21 dB in the stop band/0.75 dB in the passband).
FIR filter design using windows

Gibbs phenomenon is due to rectangular windowing. We can try to improve using other windows with better frequency behavior (smaller ripples).

\[ h[n] = h_d[n]w[n], \quad w_r[n] = \begin{cases} 1, & n = 0, \ldots, N \\ 0, & \text{otherwise} \end{cases} \]

Window has to be symmetric: \( w[n] = w[N-n] \), to maintain desired symmetry of \( h[n] \).

\[ H(\theta) = H_d(\theta) * W(\theta) \]

Frequency response of window: important is

- Main lobe, usually a multiple of \( \frac{4\pi}{L} \), with \( L = N + 1 \)
- Magnitude of highest side lobe
FIR filter design using windows

Examples of windows:

- **Bartlett window**: \( w_b = w_r \ast w_r \) \( \iff \) \( W_b(\theta) = W_r(\theta)^2 \)

Width: \( \frac{8\pi}{L} \), sidelobe level \( \delta_p, \delta_s = 0.05 \)
**FIR filter design using windows**

- **Hann window**: sum of three Dirichlet kernels, to cancel sidelobes in frequency domain
  
  Width: \( \frac{8\pi}{L} \), sidelobe level \( \delta_p, \delta_s = 0.0063 \)

- **Hamming window**: better (empirical) sum of three Dirichlet kernels

  Width: \( \frac{8\pi}{L} \), sidelobe level \( \delta_p, \delta_s = 0.0022 \)
FIR filter design using windows

- **Blackman window**: sum of five Dirichlet kernels

  Width: $\frac{12\pi}{L}$, sidelobe level $\delta_p, \delta_s = 0.0002$

- **Kaiser window**: obtained via optimization (minimize main lobe width for fixed energy in sidelobes)

  Has a parameter $\alpha$ which gives tradeoff between width and sidelobe level.
  
  E.g. $\alpha = 10$ gives width $\frac{12\pi}{L}$, sidelobe level $\delta_p, \delta_s = 0.00001$
FIR filter design using windows

Recall design example:

Design a lowpass filter with $\theta_p = 0.2\pi$, $\theta_s = 0.3\pi$, $\delta_p = \delta_s = 0.01$.

In view of $\delta_p = \delta_s = 0.01$, we can select a Hann window.

The transition bandwidth is $0.1\pi = \frac{8\pi}{L}$, so we can take $L = 80$, i.e. $N = 79$. 
Another filter design:

Implement a digital differentiator.

*Analog:* \( H^F(\omega) = j\omega \)

*Digital:* \( H^f(\theta) = j\frac{\theta}{T}, \quad -\pi \leq \theta \leq \pi, \)

where \( T \): sample rate

Initial phase response is \( \phi_0 = 0.5\pi, -\pi \leq \theta \leq \pi. \)

Add linear phase term \(-0.5\theta N\) to center the response at a delay of \(0.5N\).

*Desired frequency response:*

\[
H^f_d(\theta) = \frac{\theta}{T} e^{j(0.5\pi - 0.5\theta N)}
\]
**Differentiators**

*Desired frequency response:*

\[ H_d^f(\theta) = \frac{\theta}{T} e^{j(0.5\pi - 0.5\theta N)} \]

*Corresponding impulse response:*

\[ h_d[n] = \frac{1}{2\pi T} \int_{-\pi}^{\pi} \theta e^{j(\theta n + 0.5\pi - 0.5\theta N)} d\theta = \begin{cases} 
\frac{(-1)^{(n-0.5N)}}{(n-0.5N)\pi T}, & N \text{ even}, \ n \neq 0.5N \\
0, & N \text{ even}, \ n = 0.5N \\
\frac{(-1)^{(n-0.5N+0.5)}}{\pi(n-0.5N)^2 T}, & N \text{ odd.}
\end{cases} \]
Note odd $N$ gives much faster decay to zero (square in denominator).

This is because for even $N$, the amplitude response is anti-symmetric and periodic with $2\pi$, therefore it must have $A(\pm \pi) = 0$. 
FIR filter implementation via FFT (Porat 4.7, 5.6)

- Linear convolution via circular convolution

\[
y[n] = \sum_{k=0}^{N_2-1} x(n-k)h(k)
\]

If \( x[n] \) and \( h[n] \) are zero-padded to length \( N \geq N_1 + N_2 - 1 \), then

\[
y[n] = \{x \ast h\}[n] = \{x \circ h\}[n], \quad 0 \leq n \leq N-1
\]
Circular convolution can be implemented via FFT

\[ y = \{x \otimes h\} = \text{IFFT} \left( \text{FFT}(x) \cdot \text{FFT}(y) \right) \]

Complexity: \(3 \cdot 2^N \log N\) versus \(N_1 N_2\) operations
FIR filter implementation via FFT

- **Overlap-add convolution** (use if $N_1$ is very large)

\[ x[n] = \sum_i x_i[n], \quad \text{with } x_i[n] \text{ nonzero only over } N_1 \text{ samples} \]

\[ y_i = \{x_i \circ h\} \text{ is obtained via FFT/IFFT. Efficient for } N_2 > 18. \]