EE2S11 Signals and Systems

Ch.12 Digital filter design

Alle-Jan van der Veen

–

17 January 2024

Summary

Several techniques are available to design a digital filter:

- Specify a desired amplitude characteristic $|H(\omega)|$, and find the corresponding $h[n]$. For the phase, we require a linear phase characteristic. We will obtain an (anti-)symmetric FIR filter.
- First design an analog filter based on the given specifications (pass-band, damping in the stop-band). Then transform to the digital domain. This can be done by
  - sampling of the analog impulse response ("method of impulse invariance")
  - bilinear transform $s \to z$.

  This results in an IIR filter

IIR filters usually have a lower order for the same specifications, but they do not have linear phase (possibly resulting in pulse deformation in the pass-band).

Linear phase

Definition: a filter with frequency response

$$H(\omega) = A(\omega)e^{-j(\omega\alpha-\beta)}, \qquad -\pi \leq \omega \leq \pi$$

with $A(\omega)$ real, has *generalized linear phase*.

- The filter is similar to a delay for signals in the pass-band (where $A(\omega) \approx 1$), and does not distort these signals.
- Only FIR filters can have linear phase. Moreover, they must satisfy the symmetry property: $h[n] = \epsilon h[N-n]$, where $\epsilon = \pm 1$, and N is the filter order.
- The center of the impulse response is $N/2$. If N is even, this corresponds to a coefficient $h[N/2]$, else it doesn't. If $\epsilon = -1$ then $h[N/2] = 0$.

This results in four possibilities (Type I – Type IV).

Examples

- *Type I*: $\epsilon = 1$, $N = 4$ is even:

$$
\begin{aligned}
h[n] &= [\cdots, 0, \boxed{1}, 2, 3, 2, 1, 0, \cdots] \\
H(z) &= 1 + 2z^{-1} + 3z^{-2} + 2z^{-3} + z^{-4} \\
H(\omega) &= 1 + 2e^{-j\omega} + 3e^{-j2\omega} + 2e^{-j3\omega} + e^{-j4\omega} \\
&= [3 + 4\cos(\omega) + 2\cos(2\omega)] \cdot e^{-j2\omega}
\end{aligned}
$$

- *Type IV*: $\epsilon = -1$, $N = 3$ is odd:

$$
\begin{aligned}
h[n] &= [\cdots, 0, \boxed{1}, 2, -2, -1, 0, \cdots] \\
H(z) &= 1 + 2z^{-1} - 2z^{-2} - z^{-3} \\
H(\omega) &= 1 + 2e^{-j\omega} - 2e^{-j2\omega} - e^{-j3\omega} \\
&= [4\sin(\tfrac{1}{2}\omega) + 2\sin(1\tfrac{1}{2}\omega)] \cdot j\, e^{-j1\frac{1}{2}\omega}
\end{aligned}
$$

Linear phase

- *Type I*: $\epsilon = 1$, N is even: $H(\omega) = e^{-j\omega N/2} \sum_k a_k \cos(\omega k)$
  $A(\omega) = A(-\omega)$, $\beta = 0$
- *Type II*: $\epsilon = 1$, N is odd: $H(\omega) = e^{-j\omega N/2} \sum_k a_k \cos(\omega(k - 1/2))$
  $H(\omega) = 0$ for $\omega = \pi$: cannot be a high-pass filter
  $A(\omega) = A(-\omega)$, $\beta = 0$
- *Type III*: $\epsilon = -1$, N is even: $H(\omega) = je^{-j\omega N/2} \sum_k a_k \sin(\omega k)$
  $H(\omega) = 0$ for $\omega = 0, \pi$: cannot be a low-pass nor a high-pass filter
  $A(\omega) = -A(-\omega)$, $\beta = \frac{\pi}{2}$
- *Type IV*: $\epsilon = -1$, N is odd: $H(\omega) = je^{-j\omega N/2} \sum_k a_k \sin(\omega(k - 1/2))$
  $H(\omega) = 0$ for $\omega = 0$: cannot be a low-pass filter
  $A(\omega) = -A(-\omega)$, $\beta = \frac{\pi}{2}$

The phase delay is $\alpha = N/2$, always equal to half the filter order.

Design example

Design a low-pass filter with $\omega_p = 0.2\pi$, $\omega_s = 0.3\pi$, $\delta_p = \delta_s = 0.01$.

Approach (*"truncated impulse response design technique"*):

- Define the required amplitude response $A(\omega)$
- Select the type of filter: symmetric ($\epsilon = 1$) or anti-symmetric ($\epsilon = -1$), even or odd filter order
- Choose the filter order N; the corresponding phase characteristic is $e^{-j\omega N/2}$ of $j\, e^{-j\omega N/2}$
- Apply an Inverse Discrete-Time Fourier Transform to obtain the impulse response
- Truncate at order N; hence we obtain an approximation

Design example (2)

For our example: *the requested amplitude response* $A(\omega)$ *is:*

$$A(\omega) = \begin{cases} 1\,, & |\omega| < 0.25\pi \qquad = \frac{1}{2}(\omega_p + \omega_s) \\ 0\,, & \text{elsewhere} \end{cases}$$

- *Select the phase characteristic*

  We select a filter with a symmetric impulse response, i.e., Type I or II. (The other types cannot give a low-pass filter.)

  The resulting phase function is $\phi(\omega) = \omega N/2$.

- *The desired transfer function is:*

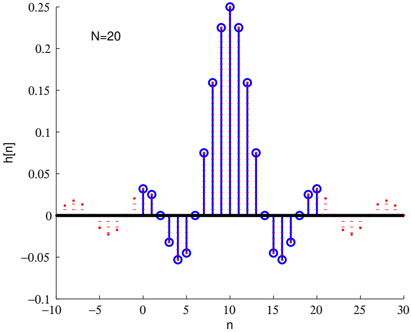
$$H_d(\omega) = \begin{cases} e^{-j\omega N/2}\,, & |\omega| < 0.25\pi \\ 0\,, & \text{elsewhere} \end{cases}$$

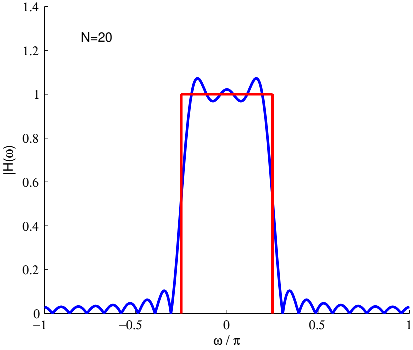
- The resulting impulse response is (IDTFT):

$$h_d[n] \;=\; \frac{1}{2\pi}\int_{-0.25\pi}^{0.25\pi} e^{j\omega(n-N/2)}d\omega \;=\; 0.25\,\mathrm{sinc}[0.25(n-N/2)]$$

Design example (3)

We select for example a filter order $N = 20$. The impulse response will then be truncated to $n \in [0, 20]$. This will change the transfer function; we apply a DTFT to see what the resulting filter is in frequency domain.

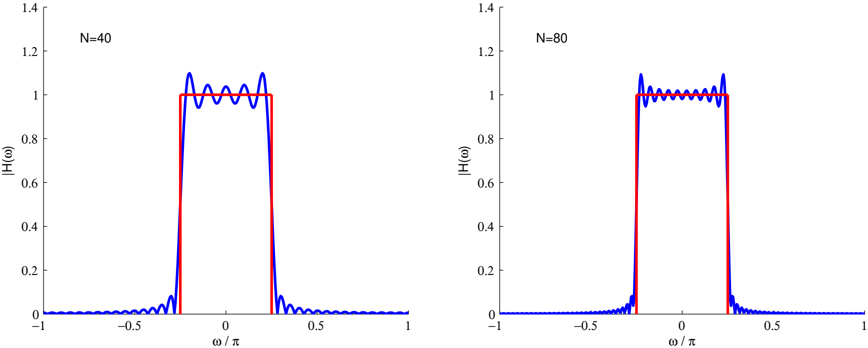




The pass-band and stop-band are OK, but the ripples are too large.

Design example (4)

We can try to reduce the ripples by selecting a larger filter order (e.g., $N = 40$ of $N = 80$)



A larger N results in a smaller transition band, but the ripples have equal magnitude! This is the Gibb's effect.

Gibb's effect

- Truncating an 'ideal' impulse response $h_d[n]$ to $h[n]$ corresponds to multiplication of $h[n]$ with a rectangular window $w_r[n]$:

$$h[n] = h_d[n]\, w_r[n]\,, \qquad w_r[n] = \begin{cases} 1\,, & n = 0, \cdots, N \\ 0\,, & \text{elsewhere} \end{cases}$$

- Multiplication in time domain corresponds to convolution in frequency domain: $H(\omega) = H_d(\omega) * W_r(\omega)$,

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

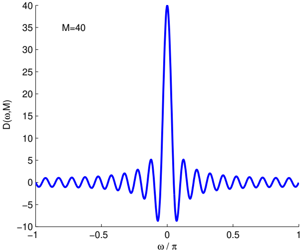
Gibb's effect

The function

$$D(\omega, M) = \frac{\sin(0.5\omega M)}{\sin(0.5\omega)}$$

is called the Dirichlet kernel.

The magnitude of the first side lobe is approximately 9% of the peak value (-13 dB), independent of M.
The main lobe has width $\approx 4\pi/M$ (first zero crossings).



Other windows

We can try to reduce the Gibb's effect by selecting another window (not rectangular).

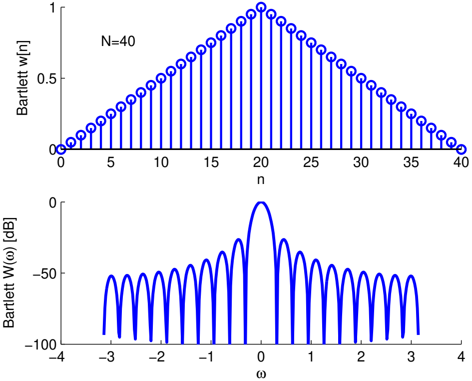
- Time domain: $h[n] = h_d[n]w[n]$.
  The window has to be symmetric: $w[n] = w[N-n]$, to keep the required symmetry of $h[n]$.
- Frequency domain: $H(\omega) = H_d(\omega) * W(\omega)$.
  Design criteria are:
  - The width of the main lobe of $W(\omega)$ should be as small as possible: this determines the width of the transition band.
    It usually is a multiple of $4\pi/M$, with $M = N+1$. Thus, we can control this using the filter order.
  - The amplitude of the first side lobe should be as small as possible.

Examples of windows

- **Bartlett window**: convolve the rectangular window with itself:

$$w_b = w_r * w_r \quad \Leftrightarrow \quad W_b(\omega) = W_r(\omega)^2$$

Width: $8\pi/M$, side lobe level $\delta_p, \delta_s = 0.05$ (-27 dB)

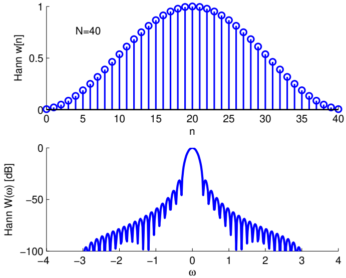


Examples of windows (cont'd)

- **Hann window**: the weighted sum of three Dirichlet kernels, designed to reduce the side lobes.

$$w[n] = \frac{1}{2}(1 - \cos\frac{2\pi n}{N})w_r[n]$$

Width: $8\pi/M$, side lobe level $\delta_p, \delta_s = 0.0063$

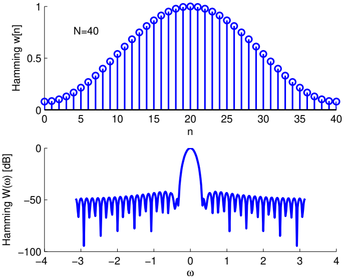


Examples of windows (cont'd)

- **Hamming window**: empirically better weighted sum of three Dirichlet kernels.

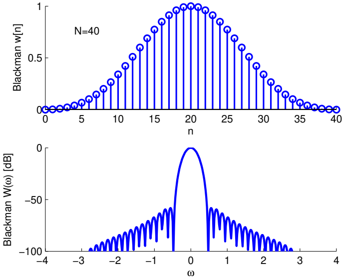
$$w[n] = (0.54 - 0.46\cos\frac{2\pi n}{N})w_r[n]$$

Width: $8\pi/M$, side lobe level $\delta_p, \delta_s = 0.0022$.



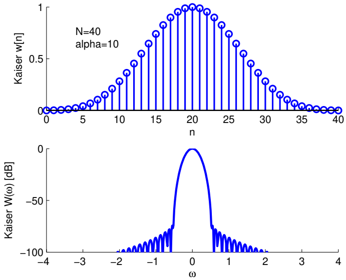
Examples of windows (cont'd)

- **Blackman window**: weighted sum of five Dirichlet kernels
  Width: $12\pi/M$, side lobe level $\delta_p, \delta_s = 0.0002$



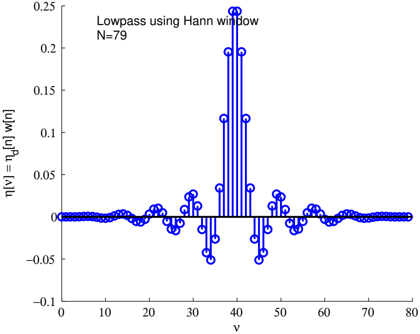
Examples of windows (cont'd)

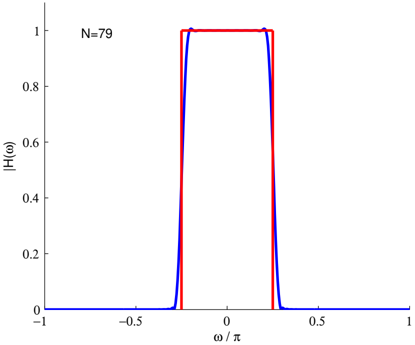
- **Kaiser window**: obtained by computer optimization (minimize the width of the peak, for a fixed energy of the side lobes).
  This design has a parameter α specifying the trade-off. E.g., $\alpha = 10$ gives width $12\pi/M$, side lobe level $\delta_p, \delta_s = 0.00001$



Returning to our filter design example

- Design a low-pass filter with
  $\omega_p = 0.2\pi, \quad \omega_s = 0.3\pi, \quad \delta_p = \delta_s = 0.01$.
  For $\delta_p = \delta_s = 0.01$, a Hann window suffices.
  The transition band is $0.1\pi = 8\pi/M$, hence we can take $M = 80$, i.e., filter order $N = 79$.





Design a digital differentiator

- *Analog:* $H_a(\Omega) = j\Omega$ *Digital:* $H(\omega) = j\dfrac{\omega}{T}$, $-\pi \le \omega \le \pi$,

with T: sample period.

Because of the 'j' we need to take Type III or IV ($\epsilon = -1$).

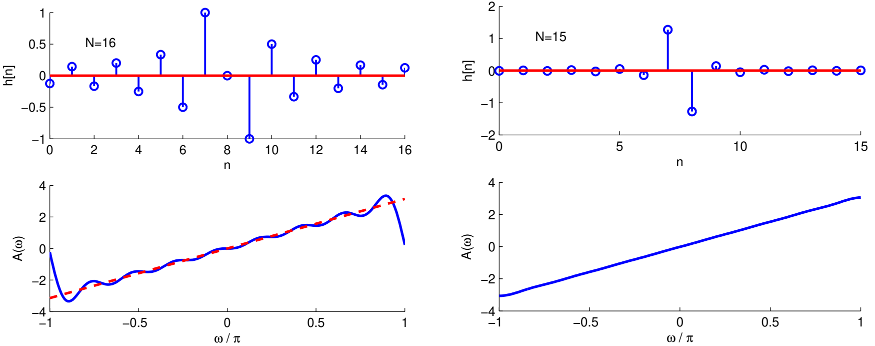
- *Resulting desired frequency response:*

$$H_d(\omega) = \frac{j\omega}{T} e^{-j(\omega N/2)} = \frac{\omega}{T} e^{j(0.5\,\pi - 0.5\,\omega N)}$$

- Corresponding impulse response:

$$h_d[n] = \frac{1}{2\pi T}\int_{-\pi}^{\pi} \omega e^{j(\omega n + \frac{1}{2}\pi - \frac{1}{2}\omega N)} \mathrm{d}\omega = \begin{cases} \dfrac{(-1)^{(n-\frac{1}{2}N)}}{(n-\frac{1}{2}N)T} & N \text{ even}, \ n \neq \frac{1}{2}N \\ 0 & N \text{ even}, \ n = \frac{1}{2}N \\ \dfrac{(-1)^{(n-\frac{1}{2}N+\frac{1}{2})}}{\pi(n-\frac{1}{2}N)^2 T} & N \text{ odd}. \end{cases}$$

Design a digital differentiator (2)



- An odd filter order N (type IV) results in a much faster decay of $h[n]$, because of the square in the denominator.

  This is because for even N, the amplitude response is anti-symmetric and periodic in 2π, resulting in $A(\pm\pi) = 0$ (not desired for a high-pass characteristic). Hence, Type III is not suitable.

Design tools

In practice, we use a computer program for FIR filter design.

Often used: Parks/McClellan technique, also known as the Remez exchange algoritm

- Specify pass-band and stop-band, e.g., $F = [0, 0.4, 0.6, 1]$ specifies a pass-band from 0 until $\omega_p = 0.4\pi$, a stop-band from $\omega_s = 0.6\pi$ until π.
- Specify the desired reponse at these critical frequencies (F), e.g., $H_d = [1, 1, 0, 0]$.
- Specify the ripple, as a weight vector $W = [1/\delta_p, 1/\delta_s]$.
- Select the filter order N (using rules of thumb; trial and error)

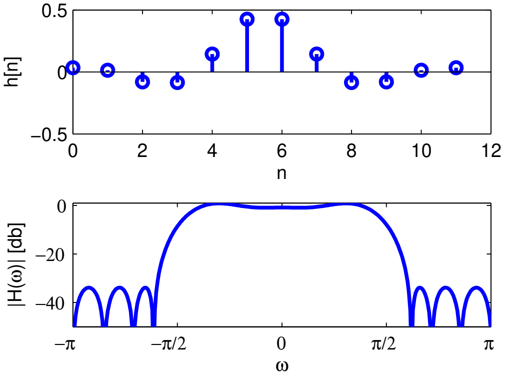
The algoritm searches $h[n]$ such that $\max_\omega \| W(\omega)\,(H_d(\omega) - A(\omega))\,\|$ is as small as possible ("minimax" optimization).

Design tools

Matlab: `h = remez(N,F,Hd,W)`

(function is now called `firpm`)

Result ($N = 11, \delta_p = 0.05, \delta_s = 0.01$):



Alternative (IIR filters): method of impulse invariance

- First design a suitable *analog* filter. E.g., a Butterworth,

$$|H_a(\Omega)|^2 = \frac{1}{1+(\Omega/\Omega_c)^{2N}} \;\Rightarrow\; H_a(s) = \frac{(\Omega_c)^N}{(s-s_0)(s-s_1)\cdots(s-s_{N-1})}$$

(Ω_c is the cut-off frequency, poles s_i are on a circle with radius Ω_c)

- Sample the corresponding impulse response $h_a(t)$ with period T_s:

$$h[n] = h_a(nT_s)$$

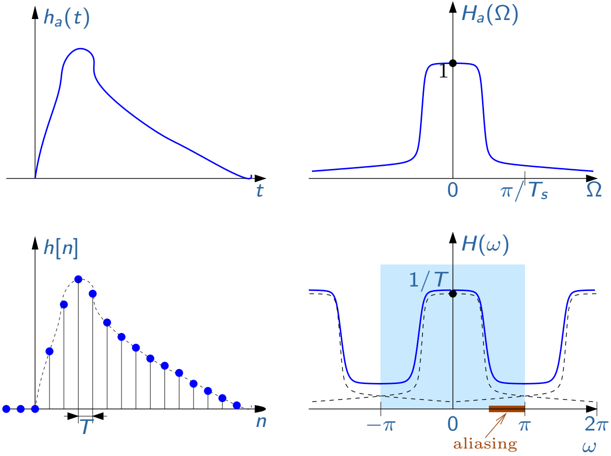
During sampling, *aliasing* can occur:

$$H(\omega) = \frac{1}{T_s}\sum_k H_a(\Omega - \frac{2\pi k}{T_s}), \qquad \Omega = \frac{\omega}{T_s}$$

The frequency response in digital domain is periodic. Without aliasing,

$$H(\omega) = \frac{1}{T_s} H_a(\frac{\omega}{T_s}), \quad |\omega| < \pi$$

Method of impulse invariance



This technique is not suitable for high-pass characteristics because of aliasing.

Method of impulse invariance

If we know $H_a(s)$, we do not need to first compute $h_a(t)$.

- Determine the poles of $H_a(s)$:

$$H_a(s) = \sum_k \frac{A_k}{s - s_k}$$

- The corresponding analog impulse response is

$$h_a(t) = \sum_k A_k e^{s_k t} u(t)$$

- The sampled version is

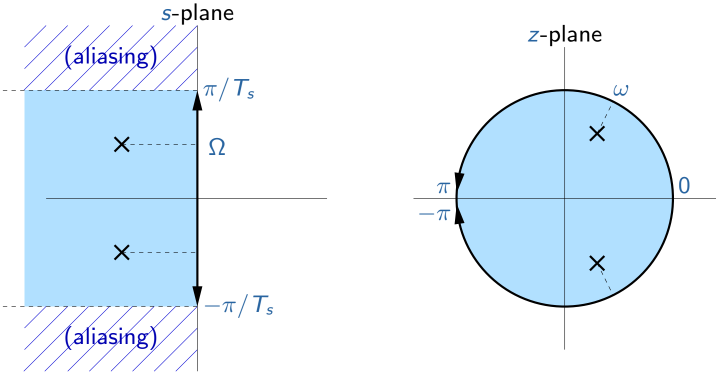
$$h[n] = \sum_k A_k e^{s_k n T_s} u[n] = \sum_k A_k (e^{s_k T_s})^n u[n]$$

- The corresponding z-transform is

$$H(z) = \sum \frac{A_k}{1 - p_k z^{-1}}, \qquad p_k = e^{s_k T_s}$$

Method of impulse invariance

- The filter order is constant,
- A 'stable pole' s_k in the left half plane ($\mathrm{Re}(s_k) < 0$) is transformed to a pole $p_k = e^{s_k T_s}$ within the unit circle ($|p_k| < 1$).
  Causality and stability are preserved.



Second alternative: bilinear transform

Given an analog filter $H_a(s)$. We can transform this into a digital filter by the mapping

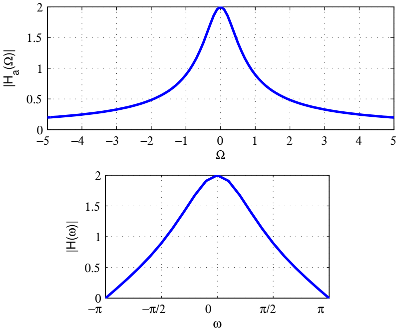
$$H(z) = H_a(s)\,, \qquad \text{with} \quad s := \frac{1 - z^{-1}}{1 + z^{-1}}$$

- Example (with $b > 0$)

$$H_a(s) = \frac{1}{s+b} \qquad \rightarrow$$

$$H(z) = \frac{1}{\frac{1-z^{-1}}{1+z^{-1}} + b} = \frac{1}{1+b} \cdot \frac{1 + z^{-1}}{1 - \frac{1-b}{1+b} z^{-1}}$$

The pole at $s = -b$ is mapped to a pole at $z = \rho = \dfrac{1-b}{1+b}$.



Derivation: bilinear transform

Simple example: consider an integrator $H_a(s) = \frac{Y(s)}{X(s)} = \frac{1}{s}$, or

$$y(nT) = y((n-1)T) + \int_{(n-1)T}^{nT} x(\tau) d\tau$$

Approximate the integral using a trapezium rule:

$$y(nT) \approx y((n-1)T) + \frac{T}{2}[x(nT) + x((n-1)T)]$$

The corresponding z-transform gives

$$Y(z) = z^{-1}Y(z) + \frac{T}{2}[X(z) + z^{-1}X(z)]$$

with transfer function

$$H(z) = \frac{Y(z)}{X(z)} = \frac{T}{2}\frac{1+z^{-1}}{1-z^{-1}}$$

The same result is obtained by substituting in $H_a(s) = 1/s$ the s by

$$s \rightarrow \frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}$$

Properties of the bilinear transform

The transformation

$$s = \frac{1 - z^{-1}}{1 + z^{-1}} \qquad \Leftrightarrow \qquad z = \frac{1+s}{1-s}$$

is called the *bilinear transform*.

- If $s = \sigma + j\Omega$ with $\sigma < 0$, then $|z| < 1$
- If $H_a(s)$ has a pole at $s = s_k$, then $H(z)$ has a pole at $p_k = \dfrac{1+s_k}{1-s_k}$

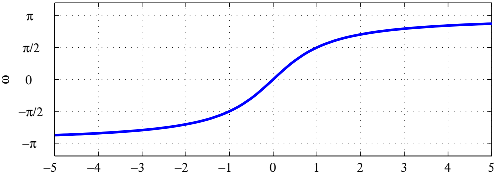
  If $H_a(s)$ is causally stable, then also $H(z)$. The filter order remains the same.
- The imaginary axis $s = j\Omega$ is mapped one-to-one to the unit circle

$$s = j\Omega \qquad \Leftrightarrow \qquad z = \frac{1+j\Omega}{1-j\Omega} \quad = \frac{A}{A^*} \quad \Rightarrow \quad |z| = 1$$

Properties of the bilinear transform (cont'd)

- With $s = j\Omega$ and $z = e^{j\omega}$ we find

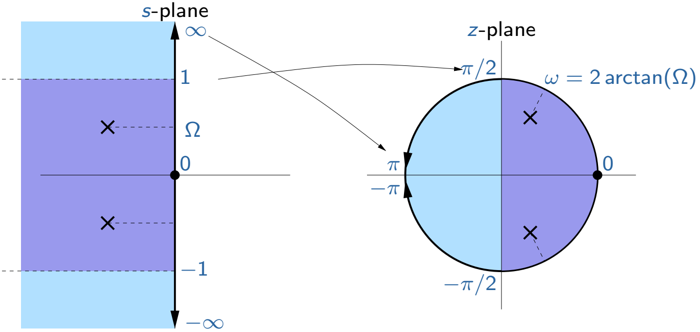
$$\omega = 2\arctan(\Omega) \qquad \Leftrightarrow \qquad \Omega = \tan(\frac{\omega}{2})$$



High frequencies ($\Omega \to \infty$) are compressed towards $\omega \to \pi$. Until $\Omega = 1$, the mapping is approximately linear.

Properties of the bilinear transform (cont'd)

The bilinear transform maps the Ω-axis non-linearly to the unit circle: no aliasing but a deformation.



Transformation from analog to digital filter

After designing an analog filter, we can transform this to a digital filter. We have seen these transforms:

- impulse invariance (not suitable for high-pass or band-stop)
- bilinear transform

Alternative: first design an analog low-pass filter, transform to a digital filter, apply a frequency transformation in digital domain.

- (We do not discuss these transforms)
- Also suitable for high-pass or band-stop
- Not equivalent, except for bilinear transform

Design specifications for a digital filter first have to be translated to specs for an analog filter. After the design we return to the digital domain via the selected transform.

Example: design using digital specifications

- Design a first-order digital low-pass filter with 3-dB band-width at $\omega_c = 0.2\pi$.
- Solution:
  - Bilinear transform of ω_c to the analog frequency domain:
    $\Omega_c = \tan(\omega_c/2) = 0.325$
  - Design a first-order Butterworth filter:

$$|H_a(s)|^2 = \frac{1}{1 + (\Omega/\Omega_c)^2} = H_a(s)H_a(-s)\Big|_{s=j\Omega} \;\Rightarrow\; H_a(s) = \frac{\Omega_c}{s + \Omega_c}$$

  - Bilinear transform of $H_a(s)$ back to $H(z)$

$$H(z) = \frac{\Omega_c}{\frac{1-z^{-1}}{1+z^{-1}} + \Omega_c} = \frac{\Omega_c(1+z^{-1})}{1 + \Omega_c - z^{-1}(1-\Omega_c)} = \frac{0.245(1+z^{-1})}{1 - 0.509\,z^{-1}}$$

  - Check: $|H(\omega = 0)| = 1, \quad |H(\omega = 0.2\pi)|^2 = 1/2$.

Example: design using digital specifications (cont'd)