

COM303: Digital Signal Processing

Lecture 12: Filter design

Overview

- ▶ filter design: problem statement
- ► IIR design
- two more ideal filters



The filter design problem

You are given a set of requirements:

- ► frequency response: passband(s) and stopband(s)
- phase: overall delay, linearity
- some limit on computational resources and/or numerical precision

You must determine N, M, a_k 's and b_k 's in

$$H(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_{M-1} z^{-M}}{1 + a_1 z^{-1} + \dots + a_{N-1} z^{-N}}$$

in order to best fulfill the requirements

The filter design problem

You are given a set of requirements:

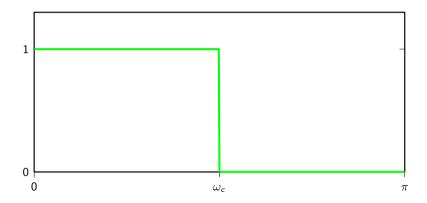
- ► frequency response: passband(s) and stopband(s)
- phase: overall delay, linearity
- some limit on computational resources and/or numerical precision

You must determine N, M, a_k 's and b_k 's in

$$H(z) = \frac{b_0 + b_1 z^{-1} + \ldots + b_{M-1} z^{-M}}{1 + a_1 z^{-1} + \ldots + a_{N-1} z^{-N}}$$

in order to best fulfill the requirements

Example: lowpass specs



- passband/stopband transitions cannot be infinitely sharp
 use transition bands
- magnitude response cannot be constant over an interval
 specify magnitude tolerances over bands

- ▶ in general:
 - smaller transition bands ⇒ higher filter order
 - smaller error tolerances ⇒ higher filter order
 - higher filter order ⇒ more expensive, larger delay

- ▶ passband/stopband transitions cannot be infinitely sharp ⇒ use transition bands
- magnitude response cannot be constant over an interva
 specify magnitude tolerances over bands

- ▶ in general:
 - smaller transition bands ⇒ higher filter order
 - smaller error tolerances ⇒ higher filter order
 - higher filter order ⇒ more expensive, larger delay

- ▶ passband/stopband transitions cannot be infinitely sharp ⇒ use transition bands
- ▶ magnitude response cannot be constant over an interval
 ⇒ specify magnitude tolerances over bands

- ► in general:
 - smaller transition bands ⇒ higher filter order
 - smaller error tolerances ⇒ higher filter order
 - higher filter order ⇒ more expensive, larger delay

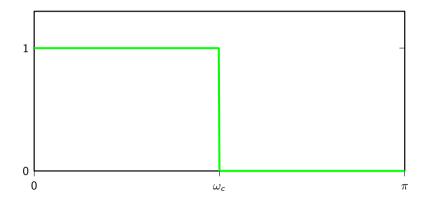
- ▶ passband/stopband transitions cannot be infinitely sharp
 ⇒ use transition bands
- ▶ magnitude response cannot be constant over an interval
 ⇒ specify magnitude tolerances over bands

- ▶ in general:
 - smaller transition bands ⇒ higher filter order
 - smaller error tolerances ⇒ higher filter order
 - ullet higher filter order \Rightarrow more expensive, larger delay

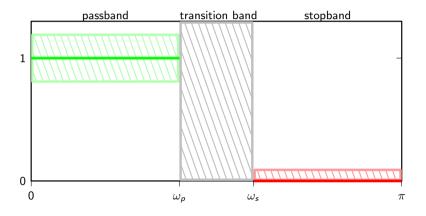
- ▶ passband/stopband transitions cannot be infinitely sharp ⇒ use transition bands
- ▶ magnitude response cannot be constant over an interval
 ⇒ specify magnitude tolerances over bands

- ▶ in general:
 - smaller transition bands ⇒ higher filter order
 - smaller error tolerances ⇒ higher filter order
 - higher filter order ⇒ more expensive, larger delay

Example: lowpass specs



Realistic specs



Why we can't have a "vertical" transition

$$H(z) = \frac{B(z)}{A(z)}$$
 is a rational function with $A, B \in C^{\infty}$

polynomial rational functions cannot have jump discontinuities

$$H(z) = \frac{B(z)}{A(z)}$$
, with A and B polynomials

$$H(e^{j\omega})=c$$
 over an interval $\Rightarrow B(z)-cA(z)=0$ over an interval $\Rightarrow B(z)-cA(z)$ has an infinite number of roots $\Rightarrow B(z)-cA(z)=0$ for all values of z $\Rightarrow H(e^{j\omega})=c$ over the entire $[-\pi,\pi]$ interval.

$$H(z) = \frac{B(z)}{A(z)}$$
, with A and B polynomials

$$H(e^{j\omega})=c$$
 over an interval $\Rightarrow B(z)-cA(z)=0$ over an interval $\Rightarrow B(z)-cA(z)$ has an infinite number of roots $\Rightarrow B(z)-cA(z)=0$ for all values of z $\Rightarrow H(e^{j\omega})=c$ over the entire $[-\pi,\pi]$ interval.

$$H(z) = \frac{B(z)}{A(z)}$$
, with A and B polynomials

$$H(e^{j\omega})=c$$
 over an interval $\Rightarrow B(z)-cA(z)=0$ over an interval $\Rightarrow B(z)-cA(z)$ has an infinite number of roots $\Rightarrow B(z)-cA(z)=0$ for all values of z $\Rightarrow H(e^{j\omega})=c$ over the entire $[-\pi,\pi]$ interval.

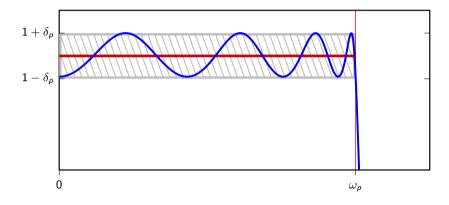
$$H(z) = \frac{B(z)}{A(z)}$$
, with A and B polynomials

$$H(e^{j\omega})=c$$
 over an interval $\Rightarrow B(z)-cA(z)=0$ over an interval $\Rightarrow B(z)-cA(z)$ has an infinite number of roots $\Rightarrow B(z)-cA(z)=0$ for all values of z $\Rightarrow H(e^{j\omega})=c$ over the entire $[-\pi,\pi]$ interval.

$$H(z) = \frac{B(z)}{A(z)}$$
, with A and B polynomials

$$H(e^{j\omega})=c$$
 over an interval $\Rightarrow B(z)-cA(z)=0$ over an interval $\Rightarrow B(z)-cA(z)$ has an infinite number of roots $\Rightarrow B(z)-cA(z)=0$ for all values of z $\Rightarrow H(e^{j\omega})=c$ over the entire $[-\pi,\pi]$ interval.

Important case: equiripple error



The big questions

- ► IIR or FIR?
- ▶ how to determine the coefficients?
- ▶ how to evaluate the performance?

IIRs: pros and cons

Pros:

- computationally efficient
- strong attenuation easy
- "natural sounding" in audio applications

Cons:

- stability and numerical precision issues
- ► limit cycles
- difficult to design for arbitrary response
- nonlinear phase

FIRs: pros and cons

Pros:

- can be designed with linear phase
- always stable
- numerically precise
- optimal design techniques exist
- efficient in multirate processing

Cons:

- computationally much more expensive
- because of length, significant delay (hard to use in live audio)

- \blacktriangleright finding N, M, a_k 's and b_k 's from specs is a hard nonlinear problem
- established methods:
 - IIR: conversion of analog design
 - FIR: optimal minimax filter design

- \blacktriangleright finding N, M, a_k 's and b_k 's from specs is a hard nonlinear problem
- established methods:
 - IIR: conversion of analog design
 - FIR: optimal minimax filter design

- \blacktriangleright finding N, M, a_k 's and b_k 's from specs is a hard nonlinear problem
- established methods:
 - IIR: conversion of analog design
 - FIR: optimal minimax filter design

- \blacktriangleright finding N, M, a_k 's and b_k 's from specs is a hard nonlinear problem
- established methods:
 - IIR: conversion of analog design
 - FIR: optimal minimax filter design



- lots of nice analog filters exist
- ▶ methods exist to "translate" the analog design into a rational transfer function
- most numerical packages (Matlab, Numpy, etc.) provide ready-made routines
- design involves specifying some parameters and testing that the specs are fulfilled

- lots of nice analog filters exist
- ▶ methods exist to "translate" the analog design into a rational transfer function
- most numerical packages (Matlab, Numpy, etc.) provide ready-made routines
- design involves specifying some parameters and testing that the specs are fulfilled

- lots of nice analog filters exist
- ▶ methods exist to "translate" the analog design into a rational transfer function
- most numerical packages (Matlab, Numpy, etc.) provide ready-made routines
- design involves specifying some parameters and testing that the specs are fulfilled

- lots of nice analog filters exist
- ▶ methods exist to "translate" the analog design into a rational transfer function
- most numerical packages (Matlab, Numpy, etc.) provide ready-made routines
- design involves specifying some parameters and testing that the specs are fulfilled

Butterworth lowpass

Magnitude response:

- maximally flat
- ightharpoonup monotonic over $[0, \pi]$

Design parameters:

- ▶ order *N* (*N* poles and *N* zeros)
- cutoff frequency

Design test criterion:

- width of transition band
- passband error

Butterworth lowpass

Magnitude response:

- maximally flat
- ightharpoonup monotonic over $[0, \pi]$

Design parameters:

- ▶ order N (N poles and N zeros)
- cutoff frequency

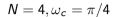
Design test criterion:

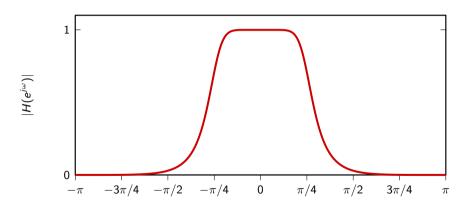
- width of transition band
- passband error

Butterworth lowpass in NumPy

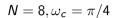
```
import scipy.signal as sp
b, a = sp.butter(4, 0.25)
wb, Hb = sp.freqz(b, a, 1024);
plt.plot(wb/np.pi, np.abs(Hb));
```

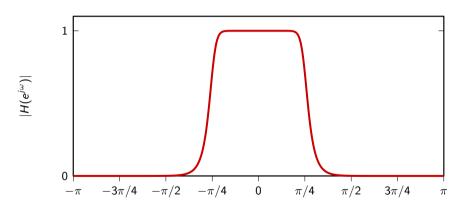
Butterworth lowpass example





Butterworth lowpass example





Chebyshev lowpass

Magnitude response:

- equiripple in passband
- ► monotonic in stopband
- ► (or vice-versa)

Design parameters:

- order N (N poles and N zeros)
- passband max error
- cutoff frequency

Design test criterion:

- width of transition band
- stopband error

Chebyshev lowpass

Magnitude response:

- equiripple in passband
- monotonic in stopband
- ► (or vice-versa)

Design parameters:

- order N (N poles and N zeros)
- passband max error
- cutoff frequency

Design test criterion:

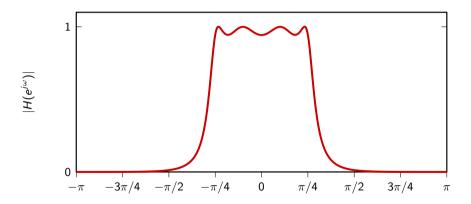
- width of transition band
- stopband error

Chebyshev lowpass in NumPy

b,
$$a = sp.cheby1(4, .12, 0.25)$$

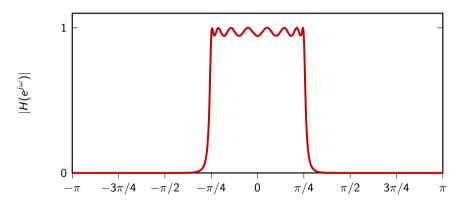
Chebyshev lowpass example

$$N = 4, \omega_c = \pi/4, e_{\sf max} = 12\%$$



Chebyshev lowpass example

$$N = 8, \omega_c = \pi/4, e_{\sf max} = 12\%$$



Elliptic lowpass

Magnitude response:

equiripple in passband and stopband

Design parameters:

- order N
- cutoff frequency
- passband max error
- ▶ stopband min attenuation

Design test criterion

width of transition band

Elliptic lowpass

Magnitude response:

equiripple in passband and stopband

Design parameters:

- ► order *N*
- cutoff frequency
- passband max error
- ▶ stopband min attenuation

Design test criterion:

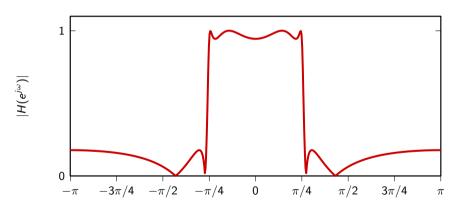
width of transition band

Elliptic lowpass in NumPy

b,
$$a = sp.ellip(4, .1, 50, 0.25)$$

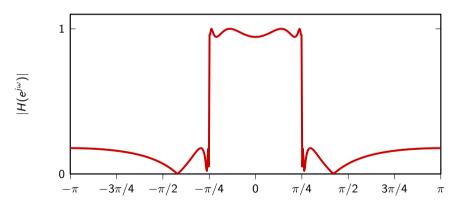
Elliptic lowpass example

$$N = 4, \omega_c = \pi/4, e_{\sf max} = 12\%, {\sf att_{\sf min}} = 0.03$$

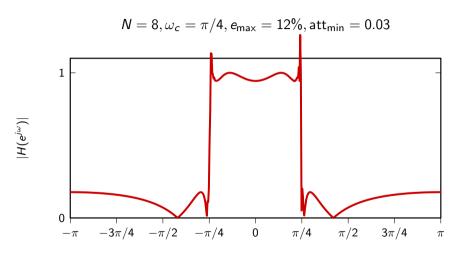


Elliptic lowpass example

$$N = 6, \omega_c = \pi/4, e_{\sf max} = 12\%, {\sf att}_{\sf min} = 0.03$$



Elliptic lowpass example



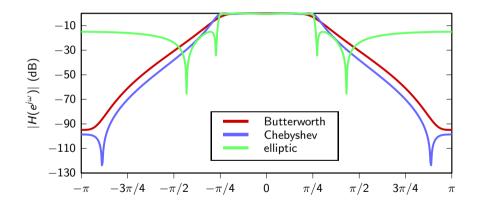
Magnitude response in decibels

- ▶ filter max passband magnitude *G*
- ▶ filter attenuation expressed in decibels as:

$$A_{\mathsf{dB}} = 20 \log_{10}(|H(e^{j\omega})|/G)$$

useful to compare attenuations between filters

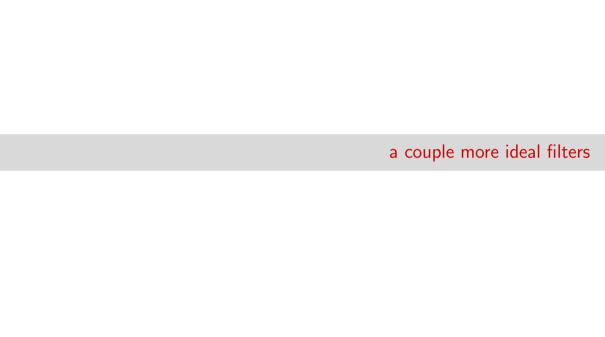
4-th order lowpass comparison



Qualitative comparison

For a given order N

- ▶ sharpness of transition band: Elliptic > Chebyshev > Butterworth
- ▶ phase distortion: Butterworth < Chebyshev < Elliptic
- ▶ pasband ripples Butterworth < Chebyshev < Elliptic
- ▶ stopband attenuation: Elliptic > Chebyshev > Butterworth



Overview

- ▶ the fractional delay
- ▶ the Hilbert filter

consider a simple delay...

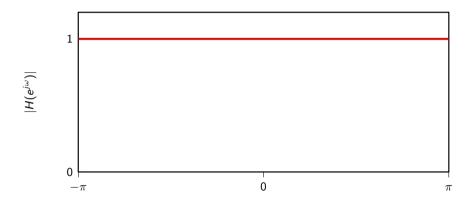
$$x[n] \longrightarrow z^{-d} \longrightarrow x[n-d]$$

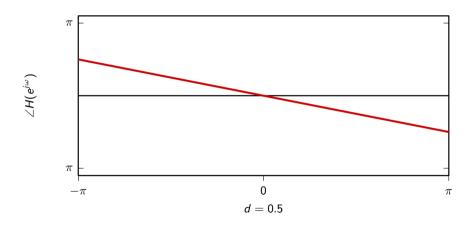
$$H(e^{j\omega})=e^{-j\omega d} \qquad d\in\mathbb{Z}$$

question

what happens if, in $H(e^{j\omega})$ we use a non-integer $d\in\mathbb{R}$?

Fractional delay: magnitude response





$$h[n] = IDTFT \left\{ e^{-j\omega d} \right\}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j\omega d} e^{j\omega n} d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega(n-d)} d\omega$$

$$= \frac{1}{\pi(n-d)} \frac{e^{j\pi(n-d)} - e^{-j\pi(n-d)}}{2j}$$

$$= \frac{\sin \pi(n-d)}{\pi(n-d)}$$

$$= \operatorname{sinc}(n-d)$$

$$h[n] = IDTFT \left\{ e^{-j\omega d} \right\}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j\omega d} e^{j\omega n} d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega(n-d)} d\omega$$

$$= \frac{1}{\pi(n-d)} \frac{e^{j\pi(n-d)} - e^{-j\pi(n-d)}}{2j}$$

$$= \frac{\sin \pi(n-d)}{\pi(n-d)}$$

$$= \operatorname{sinc}(n-d)$$

$$h[n] = IDTFT \left\{ e^{-j\omega d} \right\}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j\omega d} e^{j\omega n} d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega(n-d)} d\omega$$

$$= \frac{1}{\pi(n-d)} \frac{e^{j\pi(n-d)} - e^{-j\pi(n-d)}}{2j}$$

$$= \frac{\sin \pi(n-d)}{\pi(n-d)}$$

$$= \operatorname{sinc}(n-d)$$

$$h[n] = \text{IDTFT} \left\{ e^{-j\omega d} \right\}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j\omega d} e^{j\omega n} d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega(n-d)} d\omega$$

$$= \frac{1}{\pi(n-d)} \frac{e^{j\pi(n-d)} - e^{-j\pi(n-d)}}{2j}$$

$$= \frac{\sin \pi(n-d)}{\pi(n-d)}$$

$$= \operatorname{sinc}(n-d)$$

$$h[n] = IDTFT \left\{ e^{-j\omega d} \right\}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j\omega d} e^{j\omega n} d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega(n-d)} d\omega$$

$$= \frac{1}{\pi(n-d)} \frac{e^{j\pi(n-d)} - e^{-j\pi(n-d)}}{2j}$$

$$= \frac{\sin \pi(n-d)}{\pi(n-d)}$$

$$= \operatorname{sinc}(n-d)$$

$$h[n] = IDTFT \left\{ e^{-j\omega d} \right\}$$

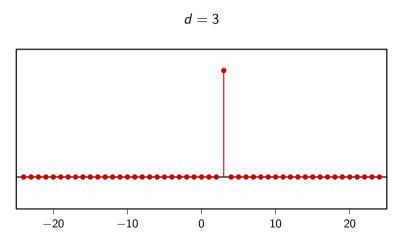
$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j\omega d} e^{j\omega n} d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega(n-d)} d\omega$$

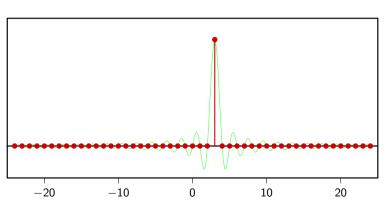
$$= \frac{1}{\pi(n-d)} \frac{e^{j\pi(n-d)} - e^{-j\pi(n-d)}}{2j}$$

$$= \frac{\sin \pi(n-d)}{\pi(n-d)}$$

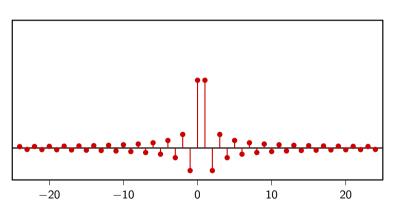
$$= \operatorname{sinc}(n-d)$$



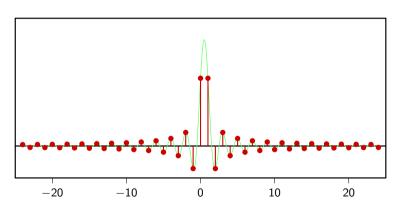




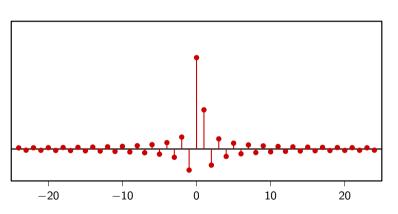




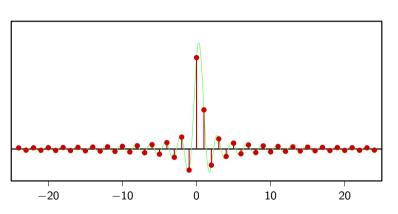


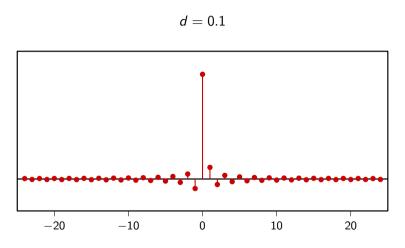




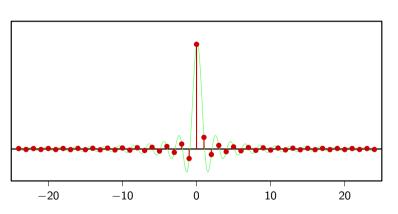










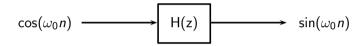


fractional delay

- ▶ fractional delay computes "in-between" values for samples
- ▶ it is an ideal filter!
- ▶ often approximated with local interpolation
- ▶ all will be clear when we study the sampling theorem



a quirky machine



can we build such a thing?

$$\begin{split} \mathsf{DTFT} \left\{ 2\cos(\omega_0 n) \right\} &= \tilde{\delta}(\omega - \omega_0) + \tilde{\delta}(\omega + \omega_0) \\ \mathsf{DTFT} \left\{ 2\sin(\omega_0 n) \right\} &= -j\tilde{\delta}(\omega - \omega_0) + j\tilde{\delta}(\omega + \omega_0) \end{split}$$

$$H(e^{j\omega})[\tilde{\delta}(\omega-\omega_0)+\tilde{\delta}(\omega+\omega_0)]=-j\tilde{\delta}(\omega-\omega_0)+j\tilde{\delta}(\omega+\omega_0)$$

$$\left\{ egin{array}{ll} H(e^{j\omega_0}) &= -1, \ H(e^{-j\omega_0}) &= +1, \end{array}
ight.$$

$$\begin{split} \mathsf{DTFT} \left\{ 2\cos(\omega_0 n) \right\} &= \tilde{\delta}(\omega - \omega_0) + \tilde{\delta}(\omega + \omega_0) \\ \mathsf{DTFT} \left\{ 2\sin(\omega_0 n) \right\} &= -j\tilde{\delta}(\omega - \omega_0) + j\tilde{\delta}(\omega + \omega_0) \end{split}$$

$$H(e^{j\omega})[\tilde{\delta}(\omega-\omega_0)+\tilde{\delta}(\omega+\omega_0)]=-j\tilde{\delta}(\omega-\omega_0)+j\tilde{\delta}(\omega+\omega_0)$$

$$egin{cases} H(e^{j\omega_0}) &= -, \ H(e^{-j\omega_0}) &= +, \end{cases}$$

$$\begin{split} \mathsf{DTFT} \left\{ 2\cos(\omega_0 n) \right\} &= \tilde{\delta}(\omega - \omega_0) + \tilde{\delta}(\omega + \omega_0) \\ \mathsf{DTFT} \left\{ 2\sin(\omega_0 n) \right\} &= -j\tilde{\delta}(\omega - \omega_0) + j\tilde{\delta}(\omega + \omega_0) \end{split}$$

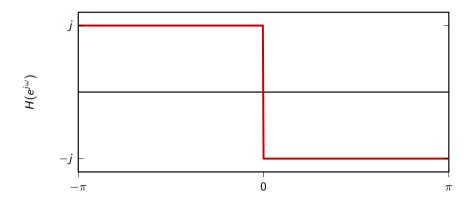
$$H(e^{j\omega})[\tilde{\delta}(\omega-\omega_0)+\tilde{\delta}(\omega+\omega_0)]=-j\tilde{\delta}(\omega-\omega_0)+j\tilde{\delta}(\omega+\omega_0)$$

$$\begin{cases} H(e^{j\omega_0}) & = -j \\ H(e^{-j\omega_0}) & = +j \end{cases}$$

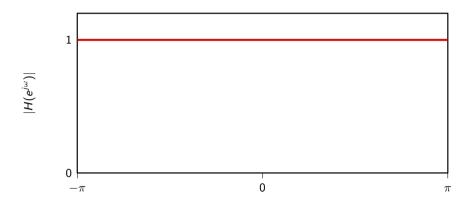
for the machine to work at all frequencies:

$$H(e^{j\omega_0}) = egin{cases} -j & ext{for } 0 \leq \omega < \pi \ +j & ext{for } -\pi \leq \omega < 0 \end{cases}$$
 (2 π -periodic)

Hilbert filter



Hilbert filter is an allpass



$$h[n] = \frac{1}{2\pi} \int_{-\pi}^{0} je^{j\omega n} d\omega + \frac{1}{2\pi} \int_{0}^{\pi} -je^{j\omega n} d\omega$$
$$= \frac{1}{2\pi n} [1 - e^{-j\pi n} - (e^{j\pi n} - 1)]$$
$$= \frac{1}{\pi n} [1 - \cos(\pi n)]$$
$$= \begin{cases} \frac{2}{\pi n} & n \text{ odd} \\ 0 & n \text{ even} \end{cases}$$

$$h[n] = \frac{1}{2\pi} \int_{-\pi}^{0} j e^{j\omega n} d\omega + \frac{1}{2\pi} \int_{0}^{\pi} -j e^{j\omega n} d\omega$$
$$= \frac{1}{2\pi n} [1 - e^{-j\pi n} - (e^{j\pi n} - 1)]$$
$$= \frac{1}{\pi n} [1 - \cos(\pi n)]$$
$$= \begin{cases} \frac{2}{\pi n} & n \text{ odd} \\ 0 & n \text{ even} \end{cases}$$

$$h[n] = \frac{1}{2\pi} \int_{-\pi}^{0} j e^{j\omega n} d\omega + \frac{1}{2\pi} \int_{0}^{\pi} -j e^{j\omega n} d\omega$$

$$= \frac{1}{2\pi n} [1 - e^{-j\pi n} - (e^{j\pi n} - 1)]$$

$$= \frac{1}{\pi n} [1 - \cos(\pi n)]$$

$$= \begin{cases} \frac{2}{\pi n} & n \text{ odd} \\ 0 & n \text{ even} \end{cases}$$

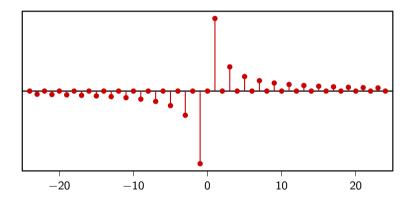
$$h[n] = \frac{1}{2\pi} \int_{-\pi}^{0} j e^{j\omega n} d\omega + \frac{1}{2\pi} \int_{0}^{\pi} -j e^{j\omega n} d\omega$$

$$= \frac{1}{2\pi n} [1 - e^{-j\pi n} - (e^{j\pi n} - 1)]$$

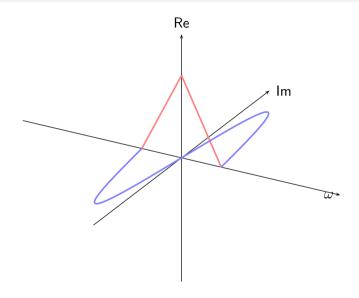
$$= \frac{1}{\pi n} [1 - \cos(\pi n)]$$

$$= \begin{cases} \frac{2}{\pi n} & n \text{ odd} \\ 0 & n \text{ even} \end{cases}$$

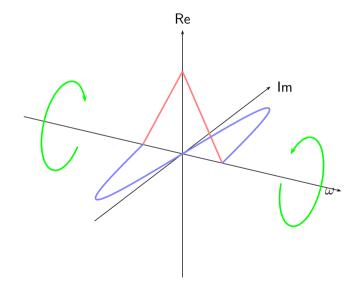
Hilbert filter

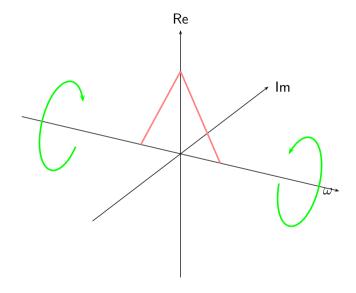


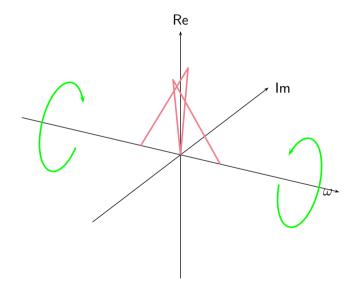
what does the Hilbert filter do?

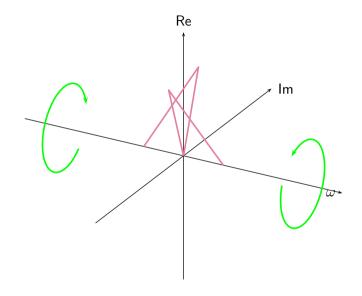


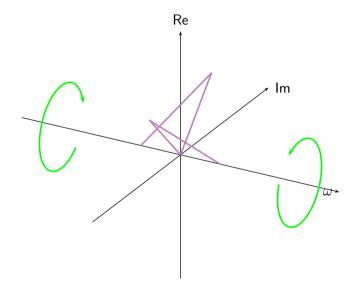
what does the Hilbert filter do?

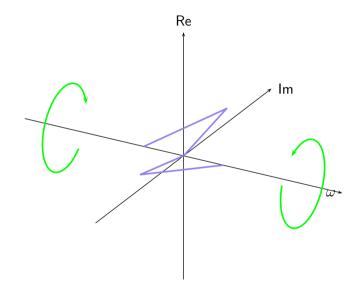


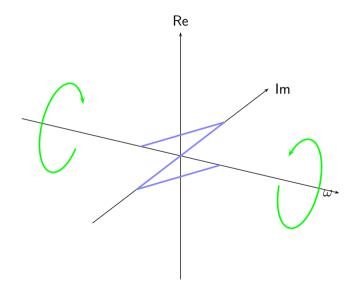


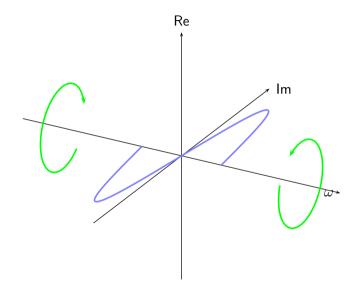


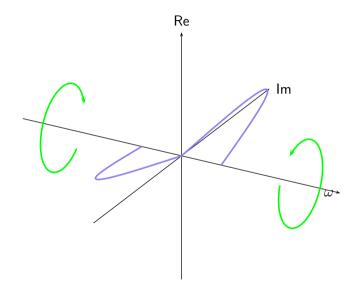


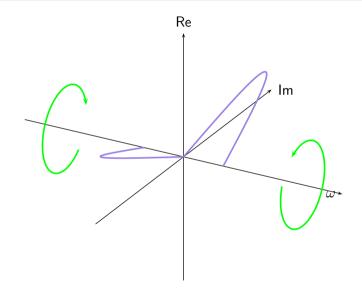


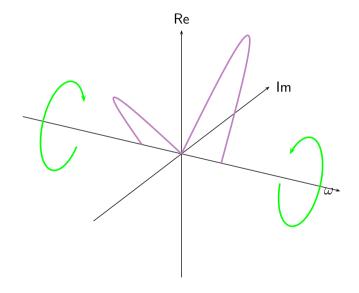


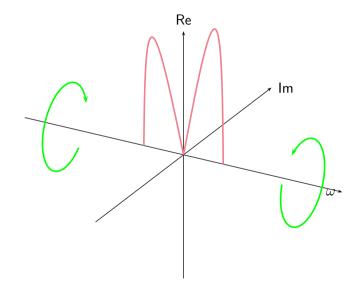


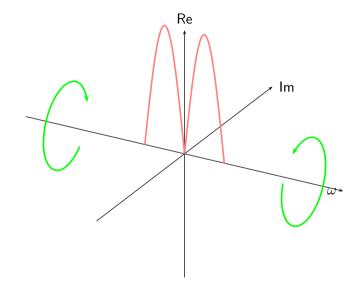


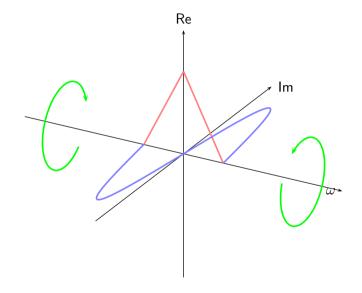


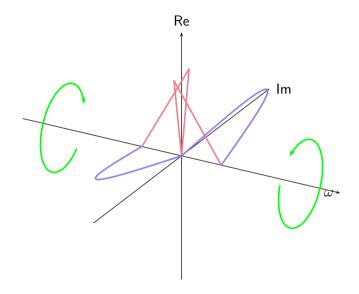


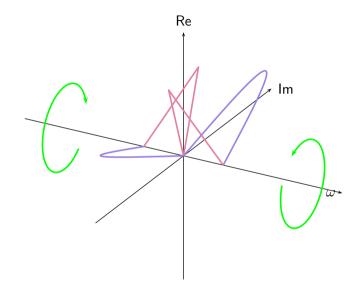


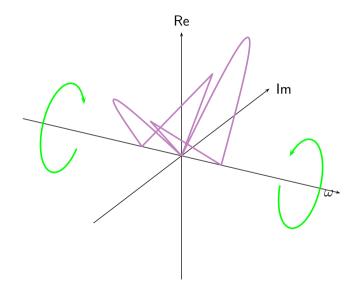


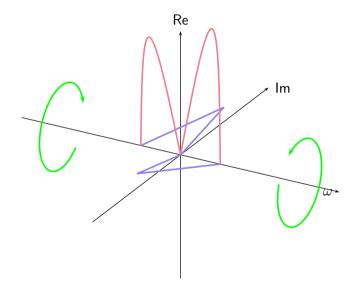


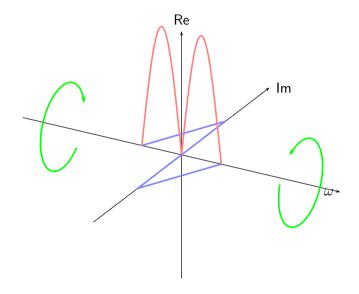




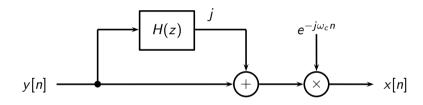






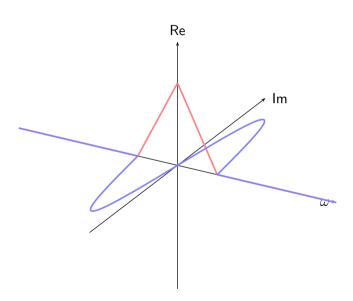


Hilbert demodulation



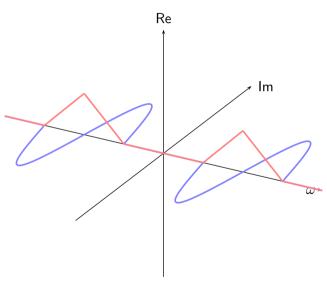
Hilbert demodulation

x[n]

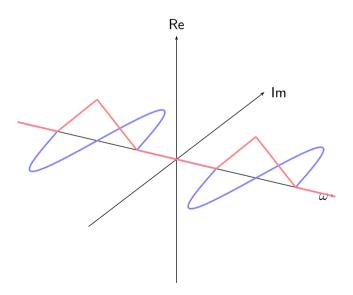


Hilbert demodulation

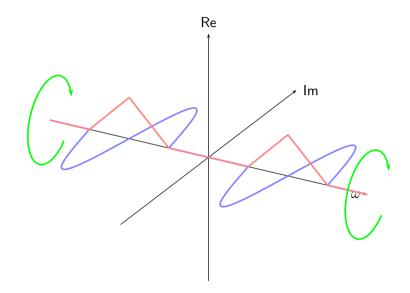
$$x[n]\cos(\omega_0 n) = y[n]$$



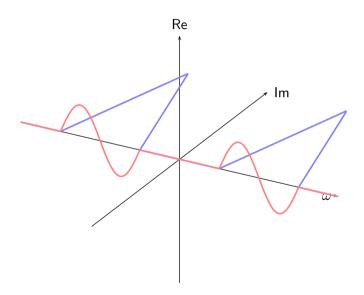
y[n]



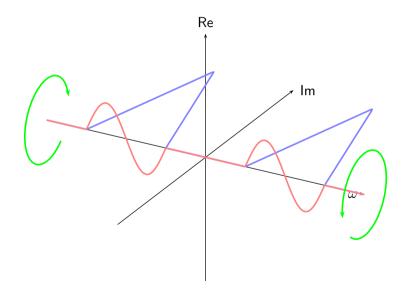
jy[n]



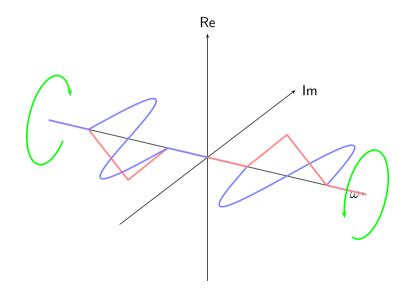
jy[n]

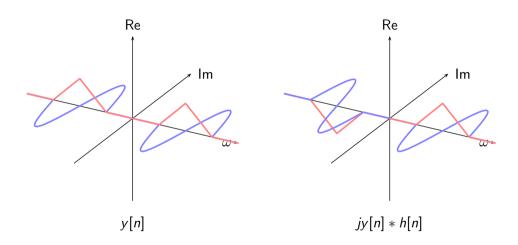


jy[n] * h[n]

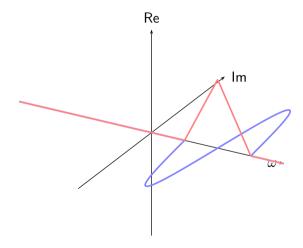


jy[n] * h[n]

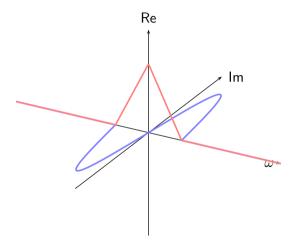




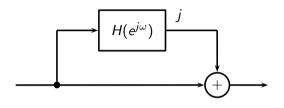
Hilbert demodulation: $jy[n] * h[n] + y[n] = x[n]e^{j\omega_0 n}$



Hilbert demodulation: $(jy[n]*h[n] + y[n])e^{-j\omega_0 n}$



Hilbert demodulator extracts the positive frequencies

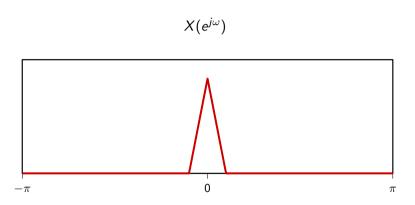


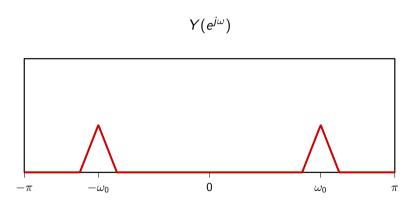
$$G(e^{j\omega}) = 1 + jH(e^{j\omega}) = egin{cases} 2 & 0 \leq \omega < \pi \ 0 & -\pi \leq \omega < 0 \end{cases}$$

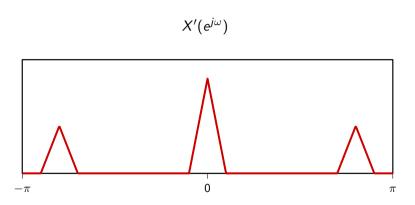
60

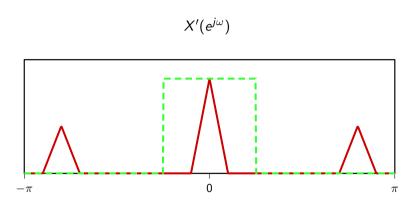
remember the classic demodulation scheme:

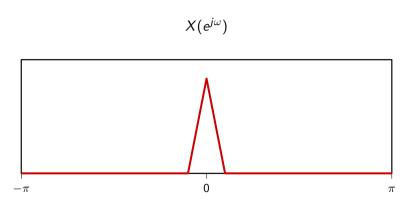
- ▶ apply sinusoidal modulation to x[n]: $y[n] = x[n] \cos \omega_0 n$
- demodulate by multiplying by the carrier $x'[n] = y[n] \cos \omega_0 n$
- remove unwanted high-frequency components via lowpass filtering











Hilbert vs Classic Demodulation

- ▶ no need to know the bandwidth of the signal
- ▶ same filter for all modulation frequencies
- ▶ good FIR approximations for the Hilbert filter exist!