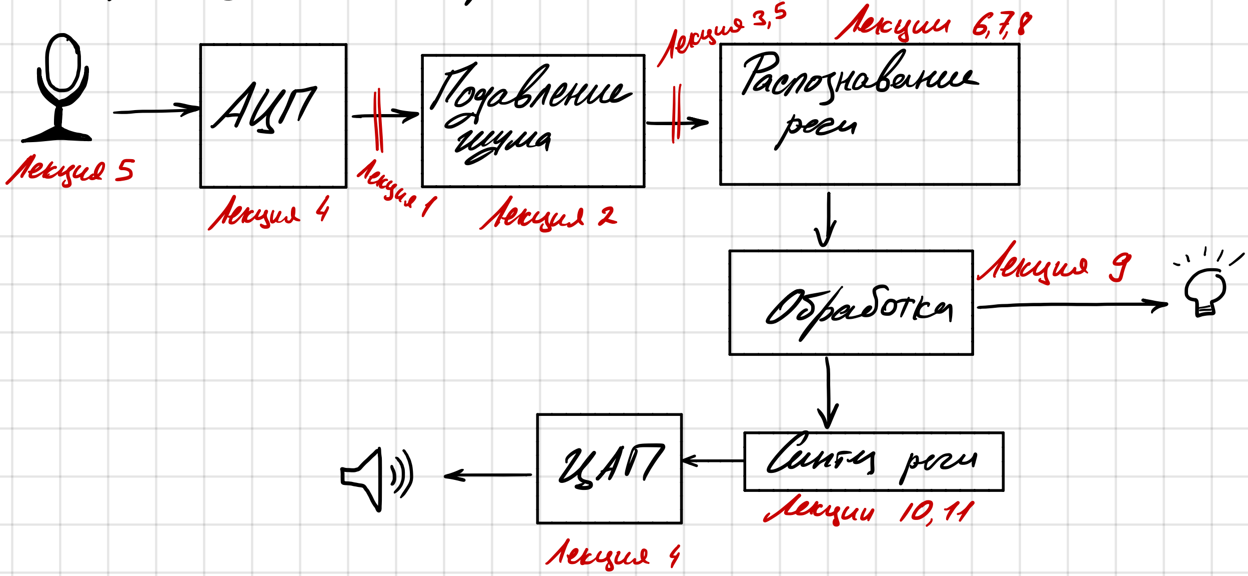
Современные методы распознавания и синтеза речи
Лекция 2

Лекция 2. Цифровые фильтры

§0 DSP и распознавание речи



Могут быть доп. шаги: e.g. использование интонации при обработке.

§1 Линейные стационарные системы

Опр. Linear time-invariant system: $y[n] = \mathcal{F}(x[n])$ $\qquad x \to \boxed{\mathcal{F}} \to y$

- $\mathcal{F}(\alpha x_1[n] + \beta x_2[n]) = \alpha \mathcal{F}(x_1[n]) + \beta \mathcal{F}(x_2[n])$
- $y[n] = \mathcal{F}(x[n]) \iff y[n-T] = \mathcal{F}(x[n-T])$

импульсная характеристика

$$y[n] = \mathcal{F}(x[n]) = \mathcal{F}\Big(\sum_{k=-\infty}^{+\infty} x[k]\delta[n-k]\Big) = \sum_{k=-\infty}^{+\infty} x[k]\,\mathcal{F}(\delta[n-k]) = \{h[n] = \mathcal{F}(\delta[n])\} =$$

$$= \sum_{k=-\infty}^{+\infty} x[k]\,h[n-k] = (x * h)[n]$$

свертка

Свойства:

$$\left.\begin{aligned} &x[n]*(\alpha y[n] + \beta w[n]) = \alpha\cdot x[n]*y[n] + \beta\cdot x[n]\cdot w[n] \\ &w[n] = x[n]*y[n] \iff x[n]*y[n-k] = w[n-k] \end{aligned}\right\} \text{Из LTI}$$

$x[n]*y[n] = y[n]*x[n]$ ← коммутативность ⇒ любой порядок

$(x[n]*y[n])*w[n] = x[n]*(y[n]*w[n])$ ← ассоциативность (только для l_2)

$$x \to \boxed{\mathcal{H}} \to \boxed{\mathcal{F}} \to \;\Leftrightarrow\; x \to \boxed{\mathcal{H}*\mathcal{F}} \to \;\Leftrightarrow\; x \to \boxed{\mathcal{F}*\mathcal{H}} \to$$

Общий случай: $x[n]*y[n] = \langle h^*[n-k], x[k]\rangle$

$$X(e^{j\omega}) * Y(e^{j\omega}) = \frac{1}{2\pi}\langle X^*(e^{j(\omega-\sigma)}), Y(e^{j\sigma})\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{j(\omega-\sigma)})\,Y(e^{j\sigma})\,d\sigma$$

Классификация:

- Конечная / Бесконечная импульсная характеристика
  FIR IIR

- Каузальный фильтр (выход зависит только от прошлого)

Некаузальные фильтры могут применяться для обработки полностью записанного сигнала. Каузальные фильтры: детекция ключевой фразы, конца фразы.

- BIBO стабильность (Bounded-Input Bounded-output):

$$\forall x,\ y = x*h:\ \exists L\ \forall n\ |x[n]| < L \Rightarrow \exists M:\ \forall n\ |y[n]| < M$$

Th BIBO $\Longleftrightarrow$ импульсная характеристика абс. суммируема

Proof

$\Leftarrow$ $$|x*h|[n] = \left|\sum_{k=-\infty}^{+\infty} x[k]h[n-k]\right| \le \sum_{k=-\infty}^{+\infty} |x[k]|\cdot|h[n-k]| \le L\sum_{k=-\infty}^{+\infty}|h[k]| < \infty$$

$\Rightarrow$ от противного: пусть $\sum|h[k]| = \infty$ $\quad x[n] = \operatorname{sgn} h[-n] \Rightarrow (x*h)[0] = \sum_{k=-\infty}^{+\infty} x[k]h[-k] = \sum_{k=-\infty}^{+\infty}|h[-k]| = \infty$

Зам FIR $\Rightarrow$ BIBO

Пример: $$h[n] = \frac{1}{N}\sum_{k=0}^{N-1}\delta[n-k] = \begin{cases}\frac{1}{N}, & 0 \le n < N\\ 0, & \text{иначе}\end{cases}$$ Moving Average

Для больших N:

$$y_N[n] = \frac{1}{N}\sum_{k=0}^{N-1} x[n-k] = \frac{N-1}{N}y_N[n-1] + \frac{1}{N}x[n] = \lambda y_{N-1}[n-1] + (1-\lambda)x[n] \qquad N \gg 1 \Rightarrow N \approx N-1$$

$y[n] = \lambda y[n-1] + (1-\lambda)x[n]$ $\longleftarrow$ разностное уравнение с постоянными коэффициентами

$y[n] = 0,\ n < N_0.$ $\quad x[n] = \delta[n] \Rightarrow h[n] = (1-\lambda)\lambda^n u[n]$

BIBO? $$\sum_{k=-\infty}^{+\infty}|h[k]| = \lim_{n\to\infty}|1-\lambda|\frac{1-|\lambda|^{n+1}}{1-|\lambda|} \Rightarrow \text{BIBO при } |\lambda| < 1$$

§2 Фильтрация в домене частот

Как фильтр влияет на спектр?

- Синусоида $e^{j\omega_0 n}$ $$\mathcal{H}\{e^{j\omega_0 n}\} = \sum_{k=-\infty}^{+\infty} e^{j\omega_0 k}h[n-k] = \sum_{k=-\infty}^{+\infty} h[k]e^{j\omega_0(n-k)} =$$

$$= \underbrace{H(e^{j\omega_0})}_{\in\mathbb{C}}\cdot e^{j\omega_0 n}$$ $\Rightarrow$ поменялась амплитуда и фаза, но не частота!

Теоремы о свёртке и модуляции:

$$y[n] = x[n] * h[n] \iff Y(e^{j\omega}) = X(e^{j\omega}) H(e^{j\omega})$$

$$y[n] = x[n] \cdot h[n] \iff Y(e^{j\omega}) = X(e^{j\omega}) * H(e^{j\omega})$$

Proof (т. о свёртке)

$$Y(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} x[k] h[n-k] e^{-j\omega n} = \sum_{k=-\infty}^{+\infty} x[k] e^{-j\omega k} \sum_{n=-\infty}^{+\infty} h[n-k] e^{-j\omega(n-k)} =$$

$$\stackrel{k-}{=} \sum_{-\infty}^{+\infty} x[k] e^{-j\omega k} \sum_{n=-\infty}^{+\infty} h[n] e^{-j\omega n} = X(e^{j\omega}) H(e^{j\omega})$$

Пример: Единичная задержка $h[n] = \delta[n-1]$

$$y[n] = \sum_{k=-\infty}^{+\infty} x[k] h[n-k] = \sum_{k=-\infty}^{+\infty} x[k] \delta[n-k-1] = x[n-1]$$

Дробная задержка: $e^{j\omega_0 n} \to e^{j\omega_0 (n+\varphi)} = \underbrace{e^{j\omega_0 \varphi}}_{\text{линейная фаза}} \cdot e^{j\omega_0 n}$

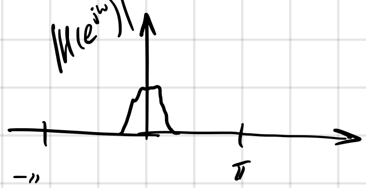
обычно у фильтров лин. фаза, чтобы сохранить форму

$$H(e^{j\omega}) = e^{j\omega\varphi}$$

$$h[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(e^{j\omega}) e^{j\omega n} d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega(n+\varphi)} d\omega = 0 \quad ?!$$

$\mathcal{H}(\delta[n]) = 0$? Это не LTI система? На практике всё равно можем вычислить преобразование

Фильтр низких частот: сохраняет низкие частоты ($\omega \approx 0$)

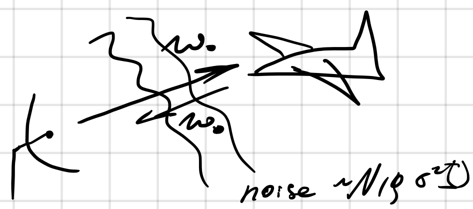


Фильтр высоких частот:

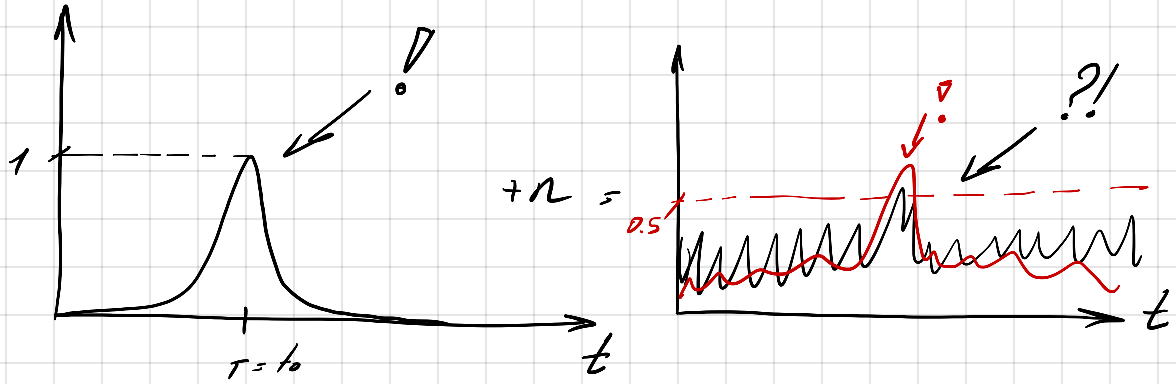
Полосовой фильтр: $\omega = \pm\omega_p$

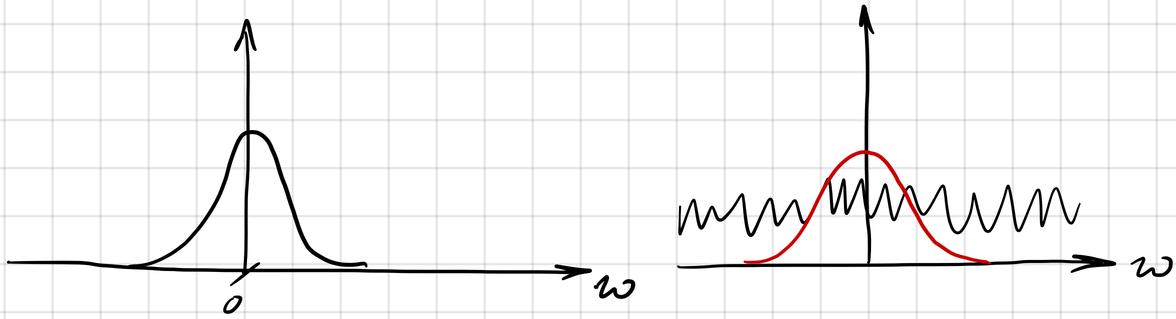
Универсальный фильтр: $[-\pi, \pi]$

Пример: радар



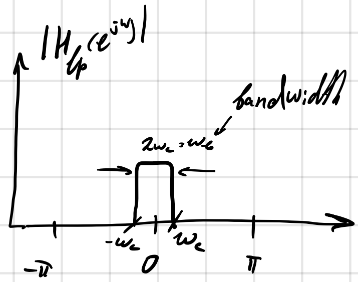
$$y[n] = x[n] + \varepsilon[n] \qquad \varepsilon[n] \sim \mathcal{N}(0, \sigma^2 I)$$





Фильтр помогает найти автокорреляцию. У шума не скоррелированы фазы.

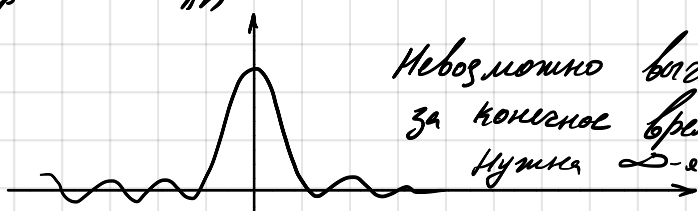
§3 Анализ фильтров



$$H_{lp}(e^{j\omega}) = \begin{cases} 1, & |\omega| \le \omega_c \\ 0, & \text{иначе} \end{cases} = \mathrm{rect}\left(\frac{\omega}{\omega_b}\right)$$

$$h_{lp}[n] = \frac{\sin(\omega_c n)}{\pi n} = \frac{\omega_b}{2\pi}\,\mathrm{sinc}\left(\frac{\omega_b}{2\pi} n\right) \qquad \left[\mathrm{sinc}(x) = \frac{\sin \pi x}{\pi x}\right]$$

$$\mathrm{rect}(x) = \begin{cases} 1, & |x| \le \frac{1}{2} \\ 0, & \text{иначе} \end{cases}$$



Невозможно вычислить за конечное время (идеальный)
Нужна ∞-я задержка

По аналогии вводится полосовой фильтр и фильтр высоких частот

Реализуемый фильтр: $\sum_{k=0}^{N-1} a_k y[n-k] = \sum_{k=0}^{M-1} b_k x[n-k]$ ← разностное уравнение с пост-ми коэф-ми

$\Updownarrow$

(CCDE)

$$y[n] = \sum_{k=1}^{N-1} \hat{a}_k y[n-k] + \sum_{k=0}^{M-1} \hat{b}_k x[n-k]$$ ← это уже можно брать и кодить!

Анализ фильтров: $\mathcal{Z}$-transform

$$X(e^{j\omega}) = \sum_{k=-\infty}^{+\infty} x[k] e^{-j\omega n} \to X(z) = \sum_{k=-\infty}^{+\infty} x[k] z^{-n}$$

Задержка: $\mathcal{Z}(x[n-N]) = z^{-N} X(z)$ + линейность

Решение CCDE:

$$y[n] = \sum_{k=0}^{M-1} b_k x[n-k] - \sum_{k=1}^{N-1} a_k y[n-k] \quad \Big| \mathcal{Z}$$

$$Y(z) = \sum_{k=0}^{M-1} b_k z^{-k} X(z) - \sum_{k=1}^{N-1} a_k z^{-k} Y(z)$$

$$Y(z) = \frac{\sum_{k=0}^{M-1} b_k z^{-k}}{1+\sum_{k=1}^{N-1} a_k z^{-k}} X(z) = \underbrace{H(z)}_{\text{функция переноса}} X(z)$$

=> подставляем $e^{j\omega}$ и делаем обр. DTFT

У реализуемых фильтров $H(z)$ – отношение полиномов

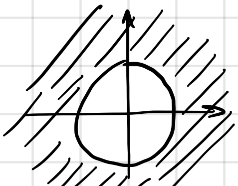
! Не нужна импульсная характеристика.

Сходимость: степенной ряд $\sum_{n=-\infty}^{+\infty} x[n] z^{-n} = \sum_{n=0}^{+\infty} \frac{x[n]}{z^n} + \sum_{n=0}^{+\infty} x[-n] z^n$

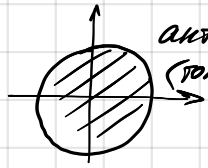
сходится в дополнении круга

сходится в круге

Th BIBO <=> единичный круг в области сходимости



каузальный



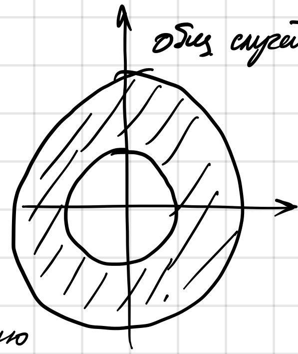
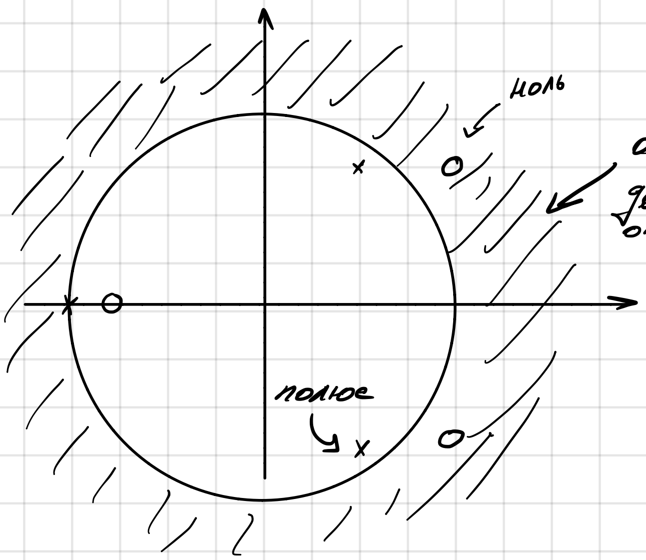


График нулей и полюсов

$$H(z) = \frac{\prod (1 - z_n z^{-1})}{\prod (1 - p_n z^{-1})}$$

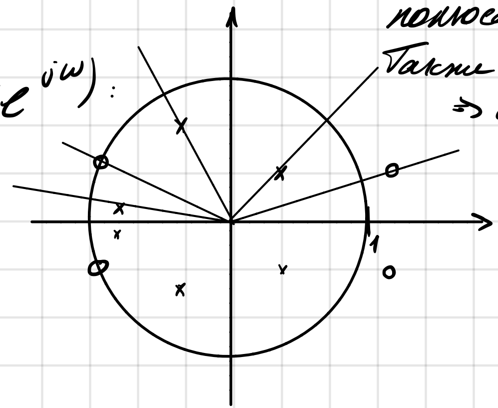


область сходимости должна содержать единичную окружность.

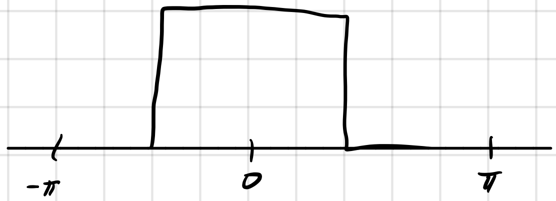
Если фильтр вещественный, то с z идет z^*, с p идет p^*.

Если фильтр имеет конечную импульсную характеристику и линейный фазовый сдвиг, то все полюса в 0. => сходится везде.
Также если z_0 – ноль, то и $1/z_0$ – ноль => => и z_0^* и $1/z_0^*$

Набросок $H(e^{j\omega})$:



Почему невозможно реализовать „идеальный" фильтр?



Const на интервале $\Rightarrow \infty$ число нулей $\Rightarrow \infty$ степень полинома!

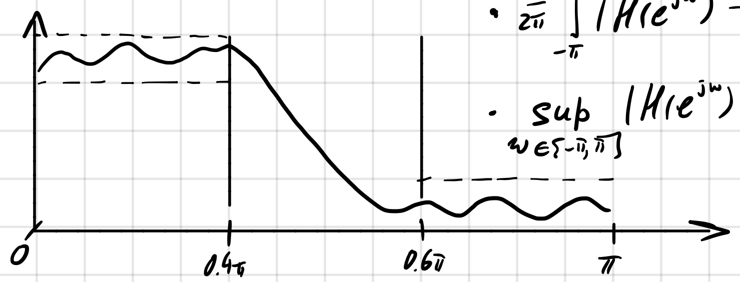
§4 Дизайн фильтров

Идеальные в Real Time сделать не можем. Почему нужно real time? Детекция ключевой фразы и конца предложения: не хотим постоянно держать включенный микрофон. Можем записывать в плохом качестве, а активировать после "Hey, Siri". Будем приближать реализуемым фильтром.

FIR или IIR (e.g. Moving average vs Exp. smoothing)

	+	−
FIR	· BIBO · точная лин. фаза · алгоритм для получения опт. фильтра · робастность к погр. выч.	· большая задержка · много вычислений
IIR	· мало вычислений · малая задержка · компактная запись	· м.б. не BIBO · проблемы с фазой · сложно строить · чувствительны к численным погрешностям

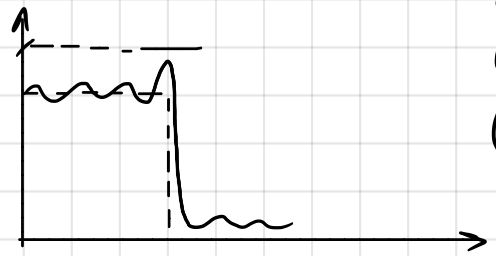
Как построить фильтр? Указать ограничения и функционал качества. Решить задачу оптимизации.



- $\frac{1}{2\pi}\int_{-\pi}^{\pi} |H(e^{j\omega}) - \hat{H}(e^{j\omega})|^2 d\omega \to \min$
- $\sup_{\omega\in[-\pi,\pi]} |H(e^{j\omega}) - \hat{H}(e^{j\omega})| \to \min$

Простой подход: $\hat{h}[n] = \begin{cases} h[n], & -N \le n \le N \\ 0, & \text{иначе} \end{cases}$

Эффект Гиббса:



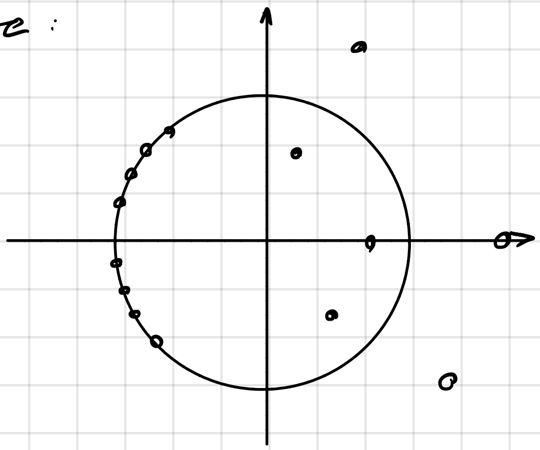
Другие окна: $\hat{h}[n] = h[n]\,w[n]$

Треугольные окна, Hamming, Blackman

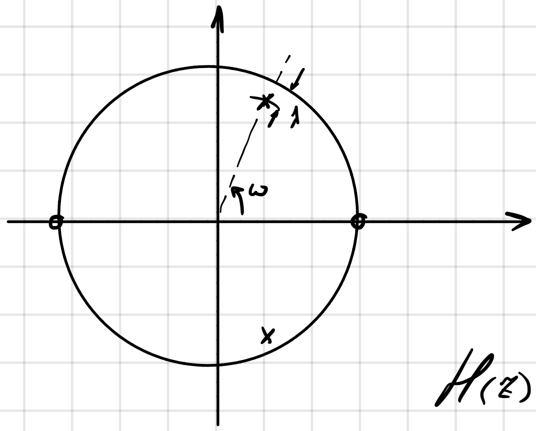
Проблема: локализованная ошибка может сильно исказить сигнал (особенно в звуке)

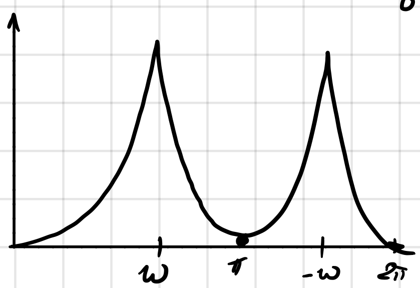
Для minmax задачи можно свести задачу к полиномиальной оптимизации (через полиномы Чебышёва): алгоритм Паркса-МакКлеллана

В результате:



Пример: фильтр, выделяющий частоту ω (e.g. определенные важные в распознавании частоты)





$$H(z) = const \cdot \frac{1 - z^{-2}}{1 - (2\lambda\cos\omega_0)z^{-1} + \lambda^2 z^{-2}}$$

$$(z - \lambda e^{j\omega})(z - \lambda e^{-j\omega}) = z^2 - (2\lambda\cos\omega_0)z + \lambda^2$$

Доп. Обращение свертки: развертка Виннера

Хотим восстановить сигнал до свертки.

$y[n] = x[n] * h[n] + \varepsilon[n]$. Ищем обратное преобразование в виде

$$\hat{x}[n] = y[n] * g[n] : \mathbb{E}\sum_{n=-\infty}^{+\infty}[x[n] - \hat{x}[n]]^2 \to \min_g$$

$\|$ [тождество Парсеваля]

$$\mathbb{E}\frac{1}{2\pi}\int_{-\pi}^{\pi}|X(e^{j\omega}) - \hat{X}(e^{j\omega})|^2 d\omega \to \min_g$$

$$\mathbb{E}\|X - \hat{X}\|^2 = \mathbb{E}\|X - GY\|^2 = \mathbb{E}\|X - G[HX + V]\|^2 = \mathbb{E}\|[1 - GH]X - GV\|^2 =$$

$$= [1-GH][1-GH]^*\mathbb{E}|X|^2 - [1-GH]G^*\mathbb{E}[XV^*] - G^*[1-GH]\mathbb{E}[VX^*] +$$

$$+ GG^*\mathbb{E}|V|^2 = \{\text{Noize and signal are independent w/ zero mean}\} \ominus$$

$$\Rightarrow \mathbb{E}[XV^*] = \mathbb{E}[VX^*] = 0$$

$S = \mathbb{E}|X|^2, \quad N = \mathbb{E}|V|^2$

$$[1 - GH][1 - GH]^* S + GG^* N \longrightarrow \min_G \qquad \Big| \frac{d}{dG}$$

$$2H[1 - GH]^* S + 2G^* N = 0$$

$$G = \frac{H^* S}{|H|^2 S + N}$$

Интерпретация: $G = \frac{1}{H}\left[\frac{|H|^2 S}{|H|^2 S + N}\right] = \frac{1}{H}\left[\frac{|H|^2}{|H|^2 + \frac{N}{S}}\right] = \frac{1}{H}\left[\frac{|H|^2}{|H|^2 + \frac{1}{SNR}}\right]$

можно оценить численно по отложенной выборке.

Применение: подавление шума.