AUDIOVISUAL COMMUNICATIONS LCAV

Mathematical Foundations of Signal Processing

Mathematical Foundations of Signal Processing

Benjamín Béjar Haro
Mihailo Kolundžija
Reza Parhizkar
Martin Vetterli

November 24, 2014

Where are we now?

1. Geometrical Tools
   - Hilbert spaces, projections etc.
2. Modeling and Analysis
   - Transforms, etc.
3. Measuring and Processing
   - Sampling and Interpolation
   - *Approximation and Compression*
   - Localization and Uncertainty
   - Compressed Sensing
4. Applications

Approximation and Compression

1. Why approximate and compress?
2. Approximation of Functions by Polynomials (finite interval)
    - Least square approximation
    - Lagrange interpolation
    - Taylor series expansion
    - Minimax approximation
    - FIR filter design
3. Approximation of Functions by Splines
    - B-Splines
    - Shift Invariant Subspaces
    - Expansions in Spline Bases
4. Approximation by Series Truncation

Why approximate and compress?

Why dealing with approximation and compression?

In many applications one faces the problem of not being able to exactly represent a measurement (a function, a discrete sequence, a vector) or not being able to exactly transform it for analysis purposes. For example:

- The function may not be known everywhere, but only at a number of specific points. The question is then how to optimally *approximate* the measurement given the available values. E.g., observing a function only on a finite interval, or only at a finite number of points in an interval.
- The function may be observable everywhere, but we may be able to afford to measure it or compute its coefficients at a finite rate for complexity and storage reasons. Then how can we obtain an *approximate* expansion or a *projection* on a subspace (e.g. projection onto bandlimited space, Fourier series truncation).
- Finite precision and storage capability. How can we *quantize* and *compress* the information, *i.e.*, how to efficiently represent it.

Approximation of Functions by Polynomials (finite interval)

Approximation of Functions by Polynomials (finite interval)

Consider a function $x(t)$ describing an physical quantity, that is observed over a finite interval $[a, b]$. A degree-K polynomial

$$p_K(t) = \sum_{k=0}^{K} \alpha_k t^k \,,$$

enables to approximate the function

- with a simple parametric representation involving only a few parameters
- based on only a few measurements

Higher degree polynomials are "wigglier" and hence can better approximate "wigglier" functions. The approximation error is given by

$$e_K(t) = x(t) - p_K(t), \; t \in [a, b].$$

Approximation of Functions by Polynomials (finite interval)

Choosing the approximating polynomial

Various criteria for choosing the approximating polynomials exist. We will study the following cases:

- Minimization of the error in the $\mathcal{L}^2([a,b])$ sense: *Least square approximation*.

  $$\text{i.e., minimize } \|e_K\|_2$$

- Matching function values at given points: *Lagrange interpolation*.

  $$\text{i.e., set } e_K(t_i) = 0 \text{ at points } t_0, t_1, \ldots, t_K$$

- Matching derivatives at a point: *Taylor series expansion*.

  $$\text{i.e., set } e_K^{(k)}(t_0) = x^{(k)}(t_0) - p_K^{(k)}(t_0) = 0 \text{ for } k = 0, 1, \ldots, K.$$

- Minimization of the error in the $\mathcal{L}^\infty([a,b])$ sense: *Minimax approximation*.

  $$\text{i.e., minimize } \|e_K\|_\infty$$

Least square approximation

Here we assume that $x(t) \in \mathcal{L}^2([a,b])$ and we choose the polynomial that minimizes

$$\|e_K\|_2^2 = \int_a^b |x(t) - p_K(t)|^2 dt\,.$$

Define

$$\mathcal{P}_K([a,b]) = \{\text{polynomials of degree} \leq K \text{ in } [a,b]\}\ .$$

Observations:

- $\mathcal{L}^2([a,b])$ is a *Hilbert space*
- $\mathcal{P}_K([a,b])$ is a *subspace* of $\mathcal{L}^2([a,b])$!

  $$\mathcal{P}_K([a,b]) = \text{span}\left(\{1, t, t^2, \ldots, t^K\}\right)$$

  i.e. $p_K(t)$ is of the form $p_K(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_K t^K,$ for some $\alpha_i \in \mathbb{C}$

- Therefore?

Least square approximation

Use *projection theorem*: The polynomial minimizing the error in the $\mathcal{L}^2([a,b])$ sense is given by the projection of $x(t)$ onto

$$\mathcal{P}_K([a,b]) = \{\text{polynomials of degree} \leq K \text{ in } [a,b]\} \subset \mathcal{L}^2([a,b])\,.$$

Least square approximation polynomial is given by

$$p_K(t) = \sum_{k=0}^{K} \langle x, \varphi_k \rangle \varphi_k(t)\,.$$

where $\{\varphi_0(t)\,, \varphi_1(t)\,, \ldots\,, \varphi_K(t)\}$ is an orthonormal basis of $\mathcal{P}_K([a,b])$ (e.g.,

obtained by Gram-Schmidt orthogonalization.

Orthonormal basis for $\mathcal{P}_K([a,b])$: Legendre polynomials

The Legendre polynomials

$$L_k(t) = \frac{(-1)^k}{2^k k!} \frac{d^k}{dt^k} (1-t^2)^k \,, \quad k \in \mathbb{N},\ t \in [-1,1],$$

are mutually orthogonal over $[-1,1]$ and have $\|L_k\| = \sqrt{2/(2k+1)}$.

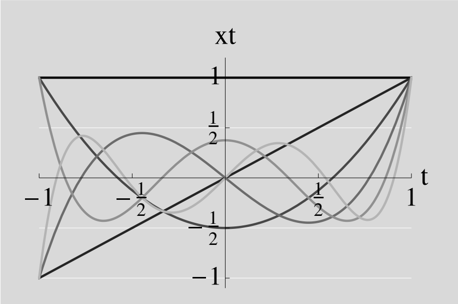
Hence

$$\varphi_k(t) = \sqrt{\frac{2k+1}{2}} L_k(t), k = 0, 1, \ldots, K$$

forms an orthonormal basis for $\mathcal{P}_K([-1,1])$!

In fact, shifted and scaled version of L_k forms orthonormal basis for $\mathcal{P}_K([a,b])$.

Least square approximation: Legendre polynomials



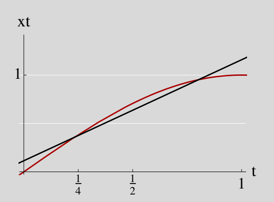
The first six Legendre polynomials $\{L_k\}_{k=0}^{5}$ (from darkest to lightest), which are orthogonal on the interval $[-1, 1]$.

Least square approximation

Approximation with Legendre polynomials

Let's approximate $x(t) = \sin(\pi t/2)$ on $[0, 1]$ with a degree-1 polynomial (a line!). An orthonormal basis of $\mathcal{P}_1([0, 1])$ is $\left\{1, \sqrt{3}(2t - 1)\right\}$. So least square approximation polynomial is

$$\begin{aligned} p_1(t) &= \langle \sin(\pi t/2), 1\rangle 1 + \langle \sin(\pi t/2), \sqrt{3}(2t - 1)\rangle \sqrt{3}(2t - 1) \\ &= \frac{12(4 - \pi)}{\pi^2} t + \frac{8(\pi - 3)}{\pi^2} . \end{aligned}$$



Least square approximation

Higher order approximations

Again let

$$x_1(t) \;=\; t \sin 5t.$$

Orthonormal polynomials on $\mathcal{L}^2([0,1])$ obtained by shifting and scaling the Legendre polynomials:

$$\begin{aligned} \varphi_0(t) &= 1 & \qquad \varphi_2(t) &= \sqrt{5}(6t^2 - 6t + 1) \\ \varphi_1(t) &= \sqrt{3}(2t-1) & \qquad \varphi_3(t) &= \sqrt{7}(20t^3 - 30t^2 + 12t - 1) \end{aligned}$$

The best degree-K approximation is

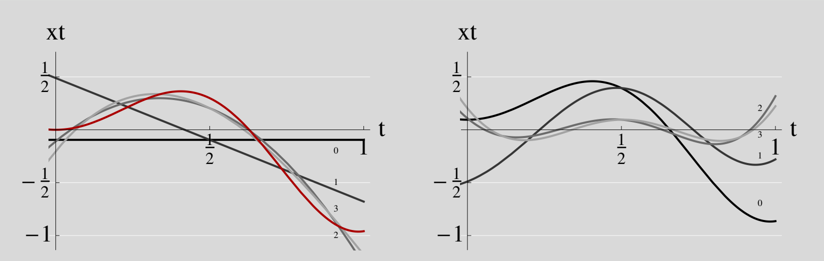
$$p_K(t) = \sum_{k=0}^{K} \langle x_1, \varphi_k \rangle \varphi_k(t)\,.$$

Least square approximation

Higher order approximations

Substituting we get:

$$\begin{aligned} p_0(t) &\approx -0.10 \\ p_1(t) &\approx 0.49 - 1.17t \end{aligned} \qquad \begin{aligned} p_2(t) &\approx -0.11 + 2.42t - 3.59t^2 \\ p_3(t) &\approx -0.20 + 3.56t - 6.43t^2 + 1.90t^3 \end{aligned}$$



Approximation of Functions by Polynomials (finite interval)

Least square approximation

Same idea can be generalized to other intervals.

The disadvantages of least square approximation: Measurements required for obtaining the approximation parameters, are *inner products* of $x(t)$ with the basis functions. Not always easy to obtain!

Easier measurements: Function values at some points. In other words, *samples* of the function.

Lagrange interpolation

Suppose we observe values of the function $x(t)$ only at $K+1$ points $t_0, \ldots, t_K$ (called *nodes*) and we constrain the approximating polynomial $p_K(t)$ to match these values at these points:

$$p_K(t_i) = x(t_i), \text{for all } i \in \{0, 1, \ldots, K\}.$$

Remark: *Remember that a degree-K polynomial has $K+1$ unknown variables, which justifies the need for $K+1$ values of $x(t)$.*

$$p_K(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_K t^K.$$

This leads to a system of $K+1$ equations with $K+1$ unknowns

$$\begin{bmatrix} 1 & t_0 & t_0^2 & \cdots & t_0^K \\ 1 & t_1 & t_1^2 & \cdots & t_1^K \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_K & t_K^2 & \cdots & t_K^K \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_K \end{bmatrix} = \begin{bmatrix} x(t_0) \\ x(t_1) \\ \vdots \\ x(t_K) \end{bmatrix}$$

A *Vandermonde* system! Invertible if and only if t_i are distinct.

Approximation of Functions by Polynomials (finite interval)

Lagrange interpolation

Unique solution is given by a *Lagrange* polynomial

$$p_K(t) = \sum_{k=0}^{K} x(t_k) \prod_{i=0, i\neq k}^{K} \frac{t - t_i}{t_k - t_i}\,.$$

Theorem (Lagrange interpolation error)

Let $x(t) \in C^{K+1}([a, b])$ ($x(t)$ has $K+1$ continuous derivatives) and assume the observed points (nodes) $t_0\,, t_1\,, \ldots\,, t_K$ are distinct. Then

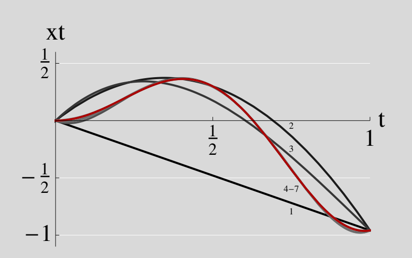
$$|e_K(t)| \leq \frac{\prod_{k=0}^{K} |t - t_k|}{(K+1)!} \max_{\eta\in[a,b]} |x^{(K+1)}(\eta)|\,.$$

Remark: *Notice that, if $x(t)$ is a polynomial of degree K the error is zero,* i.e., *the interpolant p_K matches everywhere $x(t)$: An expected result since the Lagrange polynomial is unique.*

Lagrange interpolation

Lagrange approximation

As before let $x(t) = t \sin 5t$, $t \in [0,1]$. Choose nodes uniformly as $t_k = \frac{k}{K}$.



Good match over most of the interval, but more error towards the end-points. Matches intuition since bound on error $|e_K(t)|$ is proportional to $\prod_{k=0}^{K} |t - \frac{k}{K}|$, which tends to be higher near the end-points.

Approximation of Functions by Polynomials (finite interval)

Taylor series expansion

Here we assume $x(t) \in C^K([a,b])$ and find a degree K polynomial that has the same derivatives as function $x(t)$ at some point $t_0 \in [a,b]$.

$$p_K(t) = \sum_{k=0}^{K} \frac{(t-t_0)^k}{k!} x^{(k)}(t_0)\,.$$

Also called the K-th order *Taylor series approximation* of $x(t)$ around t_0

Theorem (Taylor series expansion error)

Let $x(t) \in C^{K+1}([a,b])$ ($x(t)$ has $K+1$ continuous derivatives). Then

$$|e_K(t)| \le \frac{|t-t_0|^{K+1}}{(K+1)!} \max_{\eta \in [a,b]} |x^{(K+1)}(\eta)|\,.$$

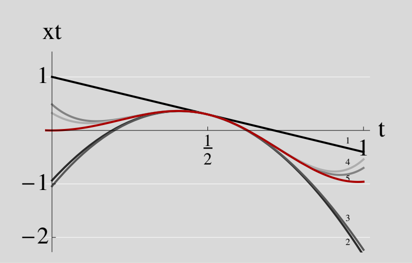
Observe: Bound on $|e_K(t)|$ proportional to $|t-t_0|^{K+1}$.

Taylor approximation

Taylor approximation

As before let

$$x(t) = t \sin 5t, t \in [0, 1]$$



Perfect match around $t_0 = 0.5$.

Approximation of Functions by Polynomials (finite interval)

Lagrange vs Taylor

Both Lagrange and Taylor have upper bound on error proportional to

$$\frac{1}{(K+1)!} \max_{\eta \in [a,b]} \left| x^{(K+1)}(\eta) \right| .$$

Intuition: Functions that have "higher" magnitude derivatives are "wigglier", and hence less suitable for approximating with low-degree polynomials.

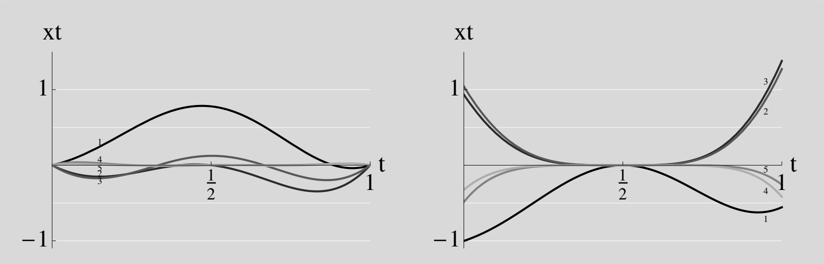
Recall that functions with higher frequency components need to be sampled at higher rate. Such functions are also "wigglier" and hence less suitable for approximating with low-bandwidth projections.

Nevertheless, the nature of Lagrange and Taylor approximations are quite different. Taylor's expansion is *accurate around t_0* whereas Lagrange's accuracy *depends on the nodes*.

Lagrange vs Taylor

Who wins?

Approximation for $x(t) = t\sin(5t)$ over $[0, 1]$. Nodes chosen at uniform spacing for Lagrange approximation. For Taylor's $t_0 = 0.5$.



Observe: Lagrange error (figure on left) vanishes at the nodes. Taylor's error (figure on right) is low around 0.5.

Minimax approximation

Here we would like to find the polynomial p_K minimizing

$$\|e_K\|_\infty = \max_{t \in [a,b]} |e_K(t)| = \max_{t \in [a,b]} |x(t) - p_K(t)| .$$

Such a minimization is really not trivial given that WE ARE NOT in a Hilbert space, hence, we cannot exploit Hilbert space tools such as the projection theorem! Finding exact minimax approximation is difficult, however it is easier to approximate!

Minimax approximation

Theorem (Chebyshev equioscillation theorem)

Let $x(t)$ be continuous on $[a, b]$. Let ε_K denote the minimax error when approximating with polynomial of degree at most K. The minimax approximation p_K is unique and determined by the existence of at least $K + 2$ points

$$a \leq s_0 < s_1 < \ldots < s_{K+1} \leq b \,,$$

for which

$$x(s_k) - p_K(s_k) = \sigma(-1)^k \|e_K\|_\infty = \sigma(-1)^k \varepsilon_K \,,$$

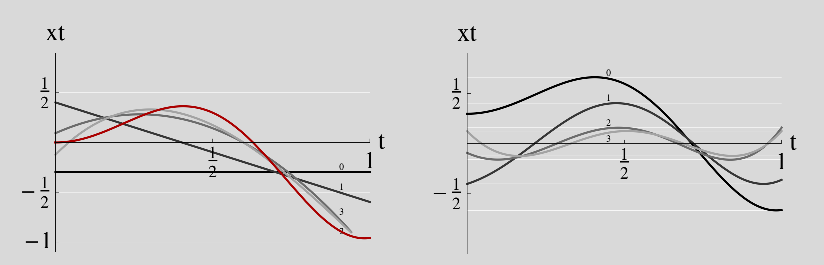
where $\sigma = \pm 1$ is independent of k.

Note: Not a constructive theorem to find the minimax approximation

Minimax approximation: Example

Minimax error

Approximation for $x(t) = t\sin(5t)$ over $[0, 1]$.



Observe: Lower errors for larger K. Equioscillations in error function.

In fact error can be made arbitrarily low for large K.

Approximation of Functions by Polynomials (finite interval)

Theorem (Weierstrass approximation theorem)

Let x be continuous on $[a, b]$ and let $\varepsilon > 0$. Then, there exists a polynomial p for which

$$|e(t)| \; = \; |x(t) - p(t)| \; \leq \; \varepsilon \qquad \textit{for every } t \in [a, b].$$

In other words, a *continuous* function can be approximated arbitrarily well on a *finite interval* by a polynomial!

(nearly) Minimax approximation over $[-1, 1]$

If we go back to the Lagrange interpolation, we recall that the error is bounded by

$$|e_K(t)| \leq \frac{\prod_{k=0}^{K} |t - t_k|}{(K+1)!} \max_{\eta \in [-1,1]} |x^{(K+1)}(\eta)|$$

Hence maximum error is bounded by

$$\|e_K\|_\infty \leq \max_{t \in [-1,1]} \frac{\prod_{k=0}^{K} |t - t_k|}{(K+1)!} \max_{\eta \in [-1,1]} |x^{(K+1)}(\eta)|$$

Idea: choose the nodes $t_0, \ldots, t_K$ to minimize right hand side. In other words instead of minimizing the maximum error we are minimizing the maximum bound on the error.

(nearly) Minimax approximation over $[-1, 1]$

The optimal $K+1$ nodes $t_0, \ldots, t_K$ (interpolation points) are given by the roots of a $K+1$ degree Chebyshev polynomial

$$t_k = \cos\left(\frac{2k+1}{2(K+1)}\pi\right), \quad k = 0, 1, \ldots, K.$$

Better than choosing interpolation points uniformly. Uniform choice leads to larger errors at the end-points of the interval.

Application to FIR filter design

Given: a desired frequency response H^{d} assumed to be *real (zero phase) and even*. Objective: Design an FIR filter with *real coefficients and zero phase*. Assume length $L = 2K + 1$ and *even symmetry*:

$$h_n = \begin{cases} h_{-n}, & \text{for } |n| \le K; \\ 0, & \text{otherwise.} \end{cases}$$

The frequency response of this filter is

$$\begin{aligned} H(e^{j\omega}) &= \sum_{n=-K}^{K} h_n e^{-j\omega n} = h_0 + \sum_{n=1}^{K} h_n \left(e^{-j\omega n} + e^{j\omega n} \right) \\ &= h_0 + 2 \sum_{n=1}^{K} h_n \cos(n\omega). \end{aligned}$$

Note: $H(e^{j\omega})$ is a polynomial of degree n in $t = \cos\omega$. Useful for minimax approximation but not least squares.

FIR filter design: Least-squares approximation

Least-squares criterion

$$\underset{\{h_0,\, h_1,\, \ldots,\, h_K\}}{\arg\min} \|H^{\mathrm{d}}(e^{j\omega}) - H(e^{j\omega})\|_2^2 \;=\; \underset{\{h_0,\, h_1,\, \ldots,\, h_K\}}{\arg\min} \int_{-\pi}^{\pi} |H^{\mathrm{d}}(e^{j\omega}) - H(e^{j\omega})|^2 \, d\omega.$$

Equivalently, by Parseval's equality:

$$\underset{\{h_0,\, h_1,\, \ldots,\, h_K\}}{\arg\min} \|h^{\mathrm{d}} - h\|_2^2 \;=\; \underset{\{h_0,\, h_1,\, \ldots,\, h_K\}}{\arg\min} \sum_{n\in\mathbb{Z}} |h_n^{\mathrm{d}} - h_n|^2.$$

Best solution is to ensure zero contribution from each of the terms

$n \in \{-K, -K+1, \ldots, K\}$.
Hence *optimal choice* is

$$h_n \;=\; h_n^{\mathrm{d}}, \qquad \text{for } n = -K, -K+1, \ldots, K.$$

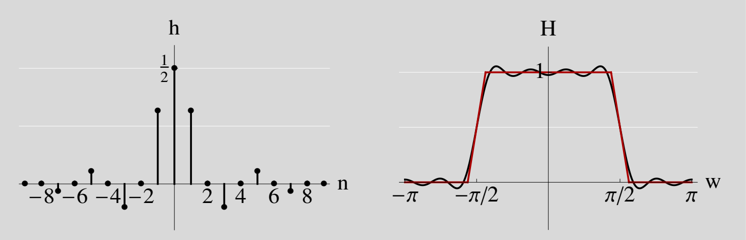
FIR filter design: Least-squares approximation

Ideal lowpass filter

An ideal halfband lowpass filter with unit passband gain:

$$h_n^{\mathrm{d}} = \tfrac{1}{2}\,\mathrm{sinc}(\tfrac{1}{2}\pi n) \quad \overset{\mathrm{DTFT}}{\longleftrightarrow} \quad H^{\mathrm{d}}(e^{j\omega}) = \begin{cases} 1, & \text{for } \omega \in [0,\, \tfrac{1}{2}\pi]; \\ 0, & \text{for } \omega \in (\tfrac{1}{2}\pi,\, \pi]. \end{cases}$$

Choose $K = 7$ (i.e., length 15)



Observe: *Gibbs phenomenon* leads to large absolute error at points of discontinuity.

FIR filter design: Minimax approximation

Minimax criterion in Fourier domain is better justified for FIR filter design.

Theorem (Minimax design criterion)

Let operator $E : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ denote effect of filtering by difference of h^{d} and h:

$$Ex \;=\; h^{\mathrm{d}} * x - h * x.$$

The operator norm of E is given by (See Solved Exercise 6.3 in book):

$$\|E\| \;=\; \max_{\omega \in [-\pi,\pi]} |H^{\mathrm{d}}(e^{j\omega}) - H(e^{j\omega})|.$$

Hence if H^{d} is real and even, the zero-phase FIR filter h with length $2K+1$ and even impulse response that minimizes the energy of the difference between filtering by h^{d} and filtering by h over unit-energy inputs is

$$\arg\min_{\{h_0, h_1, \ldots, h_K\}} \max_{\omega \in [0,\pi]} |H^{\mathrm{d}}(e^{j\omega}) - H(e^{j\omega})|.$$

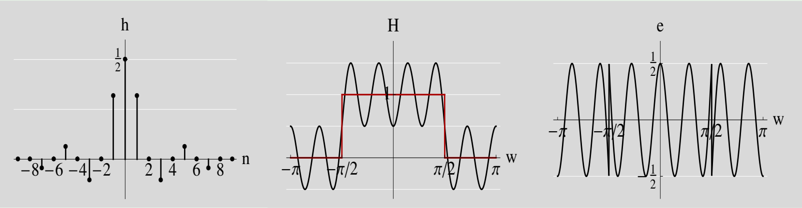
Problem is equivalent to designing a polynomial in $t = \cos\omega$ of degree K. When desired response H^{d} is continuous *equioscillation theorem* is applicable.

Ideal lowpass filter: Equiripple FIR filter design

An ideal halfband lowpass filter with unit passband gain:

$$h_n^{\mathrm{d}} \;=\; \tfrac{1}{2}\,\mathrm{sinc}(\tfrac{1}{2}\pi n) \quad \overset{\mathrm{DTFT}}{\longleftrightarrow} \quad H^{\mathrm{d}}(e^{j\omega}) \;=\; \begin{cases} 1, & \text{for } \omega \in [0,\, \tfrac{1}{2}\pi]; \\ 0, & \text{for } \omega \in (\tfrac{1}{2}\pi,\, \pi]. \end{cases}$$

Choose $K = 7$ (i.e., length 15). Solution via Parks-McClellan algorithm (an iterative algorithm that makes use of the equioscillation theorem).



Observe: *Equioscillations* in error function though equioscillation theorem is not applicable since $H^{\mathrm{d}}(e^{j\omega})$ is not continuous.

Recap: Approximation by degree K polynomials

- Minimization of the error in the $\mathcal{L}^2([a, b])$ sense: *Least square approximation*.
  - Solution via projection theorem
  - Can construct orthonormal basis using Legendre polynomials mutually orthogonal over $[-1, 1]$
- Matching function values at $K + 1$ *nodes*: *Lagrange interpolation*.
  - Inverting a Vandermonde system
- Matching K derivatives at a point t_0: *Taylor series expansion*.
  - Error is low around t_0 and increases away from t_0
- Taylor and Lagrange: Error bounded by factor proportional to $(K + 1)$-th derivative of x
  - "Wigglier" functions more difficult to approximate using low degree polynomials
- Minimization of the error in the $\mathcal{L}^\infty([a, b])$ sense: *Minimax approximation*.
  - Chebyshev equioscillation theorem. Exact solution difficult.
  - Use Lagrange method with nodes at roots of Chebyshev polynomial for nearly minimax approximation over $[-1, 1]$.

Approximation of Functions by Polynomials

Advantages and Disadvantages

+ Approximation is smooth
+ Weierstrass theorem: Can approximate continuous functions arbitrarily well over finite intervals
- Cannot approximate discontinuous functions well
- Cannot approximate over infinite interval well
- Approximating continuous functions with high degree polynomials tends to be problematic