## 1. Best approximants

Interpolation of data $\mathscr{D} = \{(x_i, y_i = f(x_i)), i = 0, \ldots, n\}$ by an approximant $p(t)$ corresponds to the minimization problem

$$\min_p \|f - p\|,$$

in the discrete one-norm at the sample points $x_i$

$$\|f\| = \|\boldsymbol{f}\|_1 = \sum_{i=0}^{n} |f(x_i)|.$$

Different approximants are obtained upon changing the norm.

THEOREM (EXISTENCE OF BEST APPROXIMANT. *For any element $f \in F$ in a normed vector space $\mathscr{F} = (F, S, +, \cdot)$, there exists a best approximant $g \in G$ within a finite dimensional subspace $G \subset F$ that is a solution of*

$$\min_{g \in G} \|f - g\|.$$

The argument underlying the above theorem is based upon constructing the closed and bounded subset of $G$

$$K = \{g \in G \mid \|g - f\| \leqslant \|0 - f\| = \|f\|\} \subset G.$$

Since $G$ is finite dimensional, $K$ is compact, and the continuous mapping $g \to \|g - f\|$ attains is extrema.

The two main classes of approximants $g$ of real functions $f: [a, b] \to \mathbb{R}$ that arise are:

**Approximants based upon sampling.** The vectors $\boldsymbol{f} = f(\boldsymbol{x}), \boldsymbol{g} = g(\boldsymbol{x})$ are constructed at sample points $\boldsymbol{x} \in \mathbb{R}^m$ and the best approximant solves the problem

$$\min_{g \in G} \|\boldsymbol{f} - \boldsymbol{g}\|.$$

Note that the minimization is carried out over the members of the subset $G$, not over the vectors $\boldsymbol{g}$. The norm can include information on derivatives as in the norm

$$\|f\|_H = \|\boldsymbol{f}\|_1 + \|\boldsymbol{f}'\|_1,$$

arising in Hermite interpolation.

**Approximants over the function domain.** The norm is now expressed through an integral such as the $p$-norms

$$\|f\|_p = \left(\int_a^b |f(t)|^p \, \mathrm{d}t\right)^{1/p}.$$

In general, the best approximant in a normed space is not unique. However, the best approximant is unique in a Hilbert space, and is further characterized by orthogonality of the residual to the approximation subspace.

THEOREM (BEST APPROXIMANT IN HILBERT SPACE). *For any element $f \in F$ in a Hilbert space $\mathscr{F} = (F, S, +, \cdot)$, there exists a unique approximant $g \in G$ within a finite dimensional subspace $G \subset F$ that is a solution of*

$$\min_{g \in G} \|f - g\|,$$

*and the residual $f - g$ is orthogonal to $G$, $\forall h \in G$*

$$(f - g, h) = 0.$$

Note that orthogonality of the residual $(f - g, h) = 0$ implies $(f, h) = (g, h)$ or that the best approximant is the projection of $f$ onto $G$.

## 2. Two-norm approximants in Hilbert spaces

For Hilbert spaces with a norm is induced by the scalar product

$$\|f\| = (f, f)^{1/2},$$

finding the best approximant reduces to a problem within $\mathbb{R}^m$ (or $\mathbb{C}^m$). Introduce a basis $\mathscr{B} = \{b_1, b_2, \ldots\}$ for $\mathscr{F}$ such that any $f \in F$ has an expansion

$$f(t) = \sum_{j=1}^{\infty} f_j b_j(t), f_j = (f, b_j)$$

Since $G$ is finite dimensional, say $n = \dim(G)$, an approximant has expansion

$$g(t) = \sum_{j=1}^{n} g_j b_{s(j)}(t).$$

Note that the approximation may lie in an arbitrary finite-dimensional subspace of $\mathscr{F}$. Choosing the appropriate subset through the function $s: \mathbb{N} \to \mathbb{N}$ is an interesting problem in itself, leading to the goal of selecting those basis functions that capture the largest components of $f$, i.e., the solution of

$$\min_{s \in \mathbb{N}^n} \sum_{j=1}^{n} |(f, b_{s(j)})|.$$

Approximate solutions of the basis component selection are obtained by processes such as greedy approximation or clustering algorithms. The approach typically adopted is to exploit the Bessel inequality

$$\sum_{i=1}^{n} f_{s(i)}^2 \leqslant \|f\|^2,$$

and select

$$s(1) = \arg\max_{i \in S} f_i^2,$$

eliminate $s(1)$ from $S$, and search again. The $k^{\text{th}}$-step is

$$s(k) = \arg\max_{i \in S} f_i^2,$$

with $S_k = S - \{s(1), \ldots, s(k-1)\}$.

Assuming $s(j) = j$, the orthogonality relation $f - g \perp G$ leads to a linear system

$$(f - g, b_i) = 0 \Rightarrow \left( \sum_{j=1}^{n} g_j b_j, b_i \right) = \sum_{j=1}^{n} (b_i, b_j)\, g_j = (f, b_i) \Rightarrow \boldsymbol{B}\boldsymbol{g} = \boldsymbol{f}.$$

If the basis is orthonormal, then $\boldsymbol{B} = \boldsymbol{I}$, and the best approximant is simply given by the projection of $f$ onto the basis elements. Note that the scalar product need not be the Euclidean discrete or continuous versions

$$(f, g) = \sum_{i=1}^{n} f_i g_i, (f, g) = \int_a^b f(t)\, g(t)\, \mathrm{d}t.$$

A weighting function may be present as in

$$(f, g) = \boldsymbol{f}^T \boldsymbol{W} \boldsymbol{g}, (f, g) = \int_a^b f(t)\, g(t)\, w(t) \mathrm{d}t,$$

discrete and continuous versions, respectively. In essense the appropriate measure $\mu(t)$ for some specific problem

$$\mathrm{d}\mu(t) = w(t)\, \mathrm{d}t,$$

arises and might not be the Euclidean measure $w(t) = 1$.

## 3. Inf-norm approximants

In the vector space of continuous functions defined on a topological space $X$ (e.g., a closed and bounded set in $\mathbb{R}^n$), a norm can be defined by

$$\|f\| = \max_{x \in X} |f(x)|,$$

and the best approximant is found by solving the problem

$$\inf_{g\in G} \|f-g\| = \inf_{g\in G} \max_{x\in X} |f(x)-g(x)|.$$

The fact that $g$ is the best approximant of $f$ can be restated as 0 being the approximant of $f-g$ since

$$\|f-g-0\| \leqslant \|f-(g+h)\|.$$

A key role is played by the points where $f(x)=g(x)$ leading to the definition of a critical set as

$$\operatorname{crit}(f) = \mathscr{Z}(f) = \{x\in X : |f(x)| = \|f\|\}.$$

When $G=P_{n-1}$, the space of polynomials of degree at most $n-1$, with $\dim P_{n-1}=n$, the best approximant can be charaterized by the number of sign changes of $f(x)-g(x)$.

THEOREM (CHEBYSHEV ALTERNATION). *The polynomial $p\in P_{n-1}$ is the best approximant of $f:[a,b]\to\mathbb{R}$ in the inf-norm*

$$\|f-p\|_\infty = \max_{a\leqslant x\leqslant b} |f(x)-p(x)|$$

*if and only if there exist $n+1$ points $a\leqslant x_0<x_1<\cdots<x_n\leqslant b$ such that*

$$f(x_i)-p(x_i) = s\cdot(-1)^i \|f-p\|_\infty,$$

*where $|s|=1$.*

Recall that choosing $x_i=\cos[(2i-1)\pi/(2n)]$, the roots of the $T_n(\theta)=\cos(n\theta)$ Chebyshev polynomial (with $x=\cos\theta$, $a=-1$, $b=1$), leads to the optimal error bound in polynomial interpolation

$$|f(t)-p(t)| \leqslant \frac{\|f^{(n+1)}\|_\infty}{(n+1)!2^n}.$$

The error bound came about from consideration of the alternation of signs of $p(x_j)-q(x_j)$ at the extrema of the Chebyshev polynomial $T_n$, $x_i=\cos(i\pi/n)$, $i=0,1,\ldots n$, with $p,q$ monic polynomials. The Cebyshev alternation theorem generalizes this observation and allows the formulation of a general approach to finding the best inf-norm approximant known as the Remez algorithm. The idea is that rather than seeking to satisfy the interpolation conditions

$$\boldsymbol{M}\boldsymbol{a} = \boldsymbol{y}$$

in the monomial basis

$$\boldsymbol{M} = \mathscr{M}_{n-1}(\mathrm{x}) = [\ \mathbf{1}\ \ \boldsymbol{x}\ \ \ldots\ \ \boldsymbol{x}^{n-1}\ ] \in \mathbb{R}^{n\times n},$$

attempt to find $n$ alternating-sign extrema points by considering the basis set

$$\boldsymbol{R} = \mathscr{R}_n(\boldsymbol{x}) = [\ \mathbf{1}\ \ \boldsymbol{x}\ \ \ldots\ \ \boldsymbol{x}^{n-1}\ \ \pm\mathbf{1}] \in \mathbb{R}^{(n+1)\times(n+1)}$$

with $\pm\mathbf{1} = [\ +1\ \ -1\ \ +1\ \ \ldots\ ]$.

**Algorithm (Remez)**

1. Initialize $\boldsymbol{x}\in\mathbb{R}^{n+1}$ to Chebyshev maxima on interval $[a,b]$
2. Solve $\boldsymbol{R}\boldsymbol{c} = f(\boldsymbol{x})$ $\mathscr{R}(\boldsymbol{x})$, $\boldsymbol{c}^T = [\ \boldsymbol{a}^T\ \ c_{n+1}\ ]$, $\boldsymbol{a}\in\mathbb{R}^n$
3. Find the extrema $\boldsymbol{y}$ of $p(t)-f(t)$ with $p(t)=a_0+a_1t+\cdots+a_{n-1}t^{n-1}$
4. If $p(y_i)-f(y_i)$ are approximately equal in absolute value and of opposite signs, return $\boldsymbol{x}$
5. Otherwise set $\boldsymbol{x}=\boldsymbol{y}$, repeat