# **LECTURE 18: BEST APPROXIMANT**

#### 1. Best approximants

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

$$\min \|f - p\|,$$

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

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

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

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

Since G is finite dimensional, K is compact, and the continuous mapping  $g \rightarrow ||g - f||$  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 f = f(x), g = g(x) are constructed at sample points  $x \in \mathbb{R}^m$  and the best approximant solves the problem

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

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

$$||f||_{H} = ||f||_{1} + ||f'||_{1},$$

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  $\mathcal{F} = (F, S, +, \cdot)$ , 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  $\mathcal{B} = \{b_1, b_2, ...\}$  for  $\mathcal{F}$  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

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

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

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 B = 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

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

A weighting function may be present as in

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

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

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

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

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

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

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

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 characterized 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 infnorm

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

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

$$f(x_i) - p(x_i) = s \cdot (-1)^{l} \|f - p\|_{\infty}$$

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

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

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, ..., 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

Ma = y

in the monomial basis

$$\boldsymbol{M} = \mathcal{M}_{n-1}(\mathbf{x}) = \begin{bmatrix} \mathbf{1} \ \boldsymbol{x} \ \dots \ \boldsymbol{x}^{n-1} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

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

$$\boldsymbol{R} = \mathcal{R}_n(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{1} & \boldsymbol{x} & \dots & \boldsymbol{x}^{n-1} \\ \pm \boldsymbol{1} \end{bmatrix} \in \mathbb{R}^{(n+1) \times (n+1)}$$

with  $\pm 1 = [+1 -1 +1 ...]$ .

#### Algorithm (Remez)

- 1. Initialize  $\mathbf{x} \in \mathbb{R}^{n+1}$  to Chebyshev maxima on interval [a, b]
- 2. Solve  $\mathbf{R}\mathbf{c} = f(\mathbf{x}) \ \mathcal{R}(\mathbf{x}), \mathbf{c}^T = [\mathbf{a}^T \ \mathbf{c}_{n+1}], \mathbf{a} \in \mathbb{R}^n$
- 3. Find the extrema 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 x
- 5. Otherwise set x = y, repeat