These are classnotes for Math/CS 887, Spring '94
corrections are welcome!

Approximation Theory overview

One considers special instances of the following

Problem. *Given some element f in a metric space X (with metric d), find a best approximation (=:*__ba__*) m^* to f from some given subset M of X, i.e., find*

$$m^* \in M \ \text{ s.t. } d(f, m^*) = \inf_{m \in M} d(f, m) =: \operatorname{dist}(f, M).$$

Abbreviation:

$$m^* \in \mathcal{P}_M(f).$$

Basic questions

- Existence: $\#\mathcal{P}_M(f) \geq 1$?
- Uniqueness: $\#\mathcal{P}_M(f) \leq 1$? More generally, $\#\mathcal{P}_M(f) = ?$
- Characterization: how would one recognize a ba (other than by comparing it with all other candidates)? This is important for
- Construction:

The metric

The metric is almost always a **norm metric**, i.e., $d(x, y) := \|x - y\|$, and the set M is usually a finite-dimensional linear subspace. But, as the problem of approximating a curve shows, there are important practical instances in which linearity plays no role and there is no suitable norm in which to pose the problem.

curve approximation problem X is the set of 'smooth' closed curves, of finite length, say, in $\mathbb{R}^2$; it is a metric space with the **Hausdorff metric**

$$d(A, B) := \max\{\operatorname{dist}(A, B), \operatorname{dist}(B, A)\},$$

with

$$\operatorname{dist}(A, B) := \sup_{a \in A} \inf_{b \in B} \|a - b\|_2.$$

M is the set of ellipses in $\mathbb{R}^2$, say (or some other class of 'simple' curves).

Specific choices for normed X

(T is an interval or a suitable subset of $\mathbb{R}^d$)

- $\mathbf{L}_2(T)$, i.e., $\|x\| := \|x\|_2 := \left(\int_T |x(t)|^2\right)^{1/2} \mathrm{d}t$ **Least-squares.**
  More generally: inner product spaces.
- $C(T)$, i.e., $\|x\| := \|x\|_\infty := \sup_{t\in T} |x(t)|$ **uniform** (or, **Chebyshev**).
- $\|x\| := \|x\|_1 := \int_T |x(t)| \,\mathrm{d}t$ **Least-mean.**
- (weighted) $\mathbf{L}_p$ ...

Specific choices for M

Usually, M is a finite-dimensional linear subspace, i.e., of the form

$$M = \mathrm{ran}[f_1, \ldots, f_n] := \{\sum_{j=1}^n f_j\, a(j) : a(j) \in \mathbb{F}\}$$

with $\mathbb{F}$ either $\mathbb{R}$ or $\mathbb{C}$.

- $\Pi_k := \mathrm{ran}[()^j : j = 0, \ldots, k]$ (algebraic) polynomials of degree $\le k$. Here, $()^j : t \mapsto t^j$. More generally,
  $$\Pi_k = \Pi_k(\mathbb{R}^d) := \mathrm{ran}[()^\alpha : |\alpha| \le k]$$
  with $()^\alpha : \mathbb{R}^d \to \mathbb{F} : x \mapsto x^\alpha := x(1)^{\alpha(1)} \cdots x(d)^{\alpha(d)}$, $\alpha \in \mathbb{Z}_+^d$, and $|\alpha| := \|\alpha\|_1 = \sum_i \alpha(i)$. Also, for $\alpha \in \mathbb{Z}_+^d$, $\Pi_\alpha := \Pi_{\le\alpha} := \mathrm{ran}[()^\beta : \beta \le \alpha]$.
- $\mathcal{R}_{k,h} := \Pi_k/\Pi_h := \{p/q : p \in \Pi_k, q \in \Pi_h\}$ rational functions of degrees k, h.
- $\overset{\circ}{\Pi}_k := \mathrm{ran}[1, \sin(\nu\cdot), \cos(\nu\cdot) : \nu = 1, \ldots k]$ trigonometric polynomials of degree k. The natural domain for trigonometric polynomials is the circle, i.e., the interval $[0 \mathinner{..} 2\pi]$ or $[-\pi \mathinner{..} \pi]$ with the endpoints identified. Customary notation for this set:
  $$\mathbb{T} = \mathbb{T}^1.$$
  If also complex scalars are admitted, we get the simpler description
  $$\overset{\circ}{\Pi}_k = \mathrm{ran}[e_{\mathrm{i}j} : |j| \le k],$$
  with $\mathrm{i} := \sqrt{-1}$, and
  $$e_\theta : x \mapsto \exp(\theta\cdot x)$$
  the **exponential with frequency** θ, a definition which even makes sense for $x \in \mathbb{F}^d$ with θ also in $\mathbb{F}^d$ and
  $$\theta\cdot x := \sum_i \theta(i)x(i)$$
  the ordinary scalar product.
- $\mathrm{Exp}_\Theta := \mathrm{ran}[e_\theta : \theta \in \Theta]$ exponentials with frequencies $\theta \in \Theta$. We already noted that $\overset{\circ}{\Pi}_k(\mathbb{R}) \subseteq \mathrm{Exp}_{\mathrm{i}\{-k,\ldots,k\}}$. Also, for any $\Theta \subset \mathbb{R}$, $\Pi_k = \lim_{h\to 0} \mathrm{Exp}_{h\Theta}$ whenever $\#\Theta = k + 1$.
- $\Pi_{k,\xi}^\rho :=$ piecewise polynomials (=:pp) in C^ρ, of degree $\le k$, with **break sequence** $\xi = (\ldots < \xi_j < \xi_{j+1} < \ldots)$. (**splines**)

The first and third are linear. The second is nonlinear. The last two are linear or nonlinear depending on whether the frequencies Θ (resp. the break sequence ξ) is fixed or variable.

Degree of approximation

considers the behavior of $\mathrm{dist}(f, M_h)$ as a function of the (discrete or continuous) parameter h. E.g., $k \mapsto \mathrm{dist}(f, \Pi_k)$ as $k \to \infty$, or $h \mapsto \mathrm{dist}(f, \Pi^\rho_{k,h\mathbf{t}})$. Usually, one considers only

$$\mathrm{dist}(K, M_h) := \sup_{f \in K} \mathrm{dist}(f, M_h)$$

with K a class of functions sharing with the particular f of interest certain characteristics (e.g., all functions whose 14th derivatives are no bigger than 7 in absolute value). Only the behavior in the limit, as $h \to 0$ or $h \to \infty$ or whatever, is usually considered. If nothing is said, then $h \to 0$.

Jackson type theorems: $f \in K \quad \Longrightarrow \quad \mathrm{dist}(f, M_h) = O(h^\alpha)$

Bernstein type (or, inverse) theorems: $\mathrm{dist}(f, M_h) = O(h^\alpha) \quad \Longrightarrow \quad f \in K$.

Typically, $K := \{f \in X : \|f\|' \le 1\}$ for some stronger norm $\|\cdot\|'$. This leads to consideration of the **K-functional**

$$K_f : t \mapsto \inf_g (\|f - g\| + t\|g\|').$$

related question: Is M a good choice for approximating f, given that we know that $f \in K$? Typical criterion involves the dimension of M, i.e., the degrees of freedom to be used. If $\dim M = n$, then one compares $\mathrm{dist}(K, M)$ with

$$d_n(K) := \inf_{\dim Y \le n} \mathrm{dist}(K, Y),$$

the n-**width** of K (in the sense of Kolmogorov). While it is not easy to find **optimal** subspaces, one can often find a 'scale' (M_n) for which $\mathrm{dist}(K, M_n) \sim O(d_n(K))$ and $\dim M_n \sim n$.

Good approximation

In practice, best approximation is rarely used. Instead, one looks for (cheap but good) approximation schemes. E.g., if A is a linear map into M, then $\|(\mathrm{id} - A)_{|K}\| := \sup_{f \in K} \|f - Af\|$ may well be close to $\mathrm{dist}(K, M)$.

Special case: a **near-best** A is one for which, for some const and for all f,

$$\|f - Af\| \le \mathrm{const}\, \mathrm{dist}(f, M).$$

Any such A is necessarily a projector. Conversely, if A is a bounded linear projector onto M, then, for any f, and any $m \in M$, $\|f - Af\| = \|(1 - A)(f - m)\| \le \|1 - A\|\|f - m\|$, therefore

$$\|f - Af\| \le \|1 - A\|\, \mathrm{dist}(f, M).$$

course

The intent is to give a quick reading of these basics of AT, illustrated with the help of splines, thereby giving also a quick introduction to (univariate) splines.

Weierstraß, Korovkin, Lebesgue, Bernstein

Start off the course the way Lorentz starts off his book (the nicest book in classical AT) and the way Tikhomirov starts off his survey of AT, namely with

(1) Weierstraß (1885). *For any (finite) interval* $[a \,..\, b]$, $\Pi|_{[a..b]}$ *is dense in* $C([a \,..\, b])$.

We will give three proofs (at least). The first proof is Bernstein's, but done with Korovkin's theorem. The second is Lebesgue's, done with broken lines. The third is Stone's.

(2) Korovkin (1957). *T compact, (U_n) in $L(C(T))$, U_n positive, $U_n \xrightarrow{pw} \mathrm{id}$ on some finite set G. If there exists $(a_g : g \in G)$ in $C(T)$ so that*

$$p(t,s) := \sum_{g \in G} g(t) a_g(s) \ge 0 \quad \text{equality iff } t = s,$$

then $U_n \xrightarrow{pw} \mathrm{id}$.

Explanations:

(i) For the U_n and U all maps on the same domain and into the same target, $U_n \xrightarrow{pw} U$ indicates that U_n converges **pointwise** to U, i.e., $\forall\{x \in X\}\ \ \lim_{n\to\infty} U_n x = Ux$ (in the topology of the common target of the U_n and U). This is much weaker than the more elusive **uniform** convergence, denoted $U_n \xrightarrow{u} U$, which presupposes that the common target is normed and means that

$$\lim_{n\to\infty} \sup_{x \in X} \|U_n x - Ux\| = 0.$$

(ii) It is assumed here that $C(T)$ is the set of *real-valued* continuous maps on T. For those, there is a **natural (partial) order**, namely

$$f \le g :\Longleftrightarrow \forall t \in T\ f(t) \le g(t).$$

$U : C(T) \to C(T)$ is called **positive** (or, more precisely, **nonnegative**) if

$$0 \le f \quad \Longrightarrow \quad 0 \le Uf.$$

Observation: Assume that U is positive and linear. Then $f \le g \Longrightarrow Uf \le Ug$. Further, with $|f| : T \to \mathbb{R} : t \mapsto |f(t)|$,

$$-|f| \le f \le |f| \quad \Longrightarrow \quad -U(|f|) \le Uf \le U(|f|),$$

hence,

$$|U(f)| \leq U(|f|).$$

Proof of (2)Korovkin: By assumption, $U_n \xrightarrow{pw}$ id on G, hence converges pointwise on

$$\operatorname{ran}[G] := \{\sum_{g\in G} gc(g) : c \in \mathbb{R}^G\} =: \operatorname{span} G.$$

The latter is finite-dimensional, therefore $U_n \to$ id *uniformly on bounded subsets of* $\operatorname{ran}[G]$.

The rest is a trick. From the assumption, $\operatorname{ran}[G]$ contains positive functions, e.g., the function $t \mapsto p(t,s) + p(t,s')$ for any $s \neq s'$. Let p^* be one such. Then

$$U_n p^* \to p^*.$$

Since we can choose a positive b so that $1 \leq bp^*$, we also have

$$|U_n(1)| \leq U_n(1) \leq bU_n(p^*) \to bp^*,$$

hence

$$M_1 := \sup_n \|U_n(1)\|_\infty < \infty.$$

Now let $f \in C(T)$ be the function to be approximated and consider the function

$$h : T \times T \to \mathbb{R} : (t,s) \mapsto f(t) - \frac{f(s)}{p^*(s)} p^*(t).$$

$h \in C(T \times T)$ and vanishes on

$$\Delta := \{(t,t) : t \in T\}.$$

This is sufficient for the following

Claim. *$U_n(h(\cdot,s))(s) \to 0$ uniformly in s.*

Proof: Take any $\varepsilon > 0$. Since h is continuous, the set $\Delta_\varepsilon := \{(t,s) : |h(t,s)| \geq \varepsilon\}$ is compact; it also does not contain Δ. Since p is continuous and vanishes only on Δ,

$$\delta := \inf_{\Delta_\varepsilon} p(t,s) \ > 0.$$

Therefore, $|h| \leq (\|h\|_\infty/\delta)p$ on Δ_ε and $|h| \leq \varepsilon$ otherwise; hence

$$|h| \leq \varepsilon + (\|h\|_\infty/\delta)p.$$

Consequently, for any s,

$$|U_n(h(\cdot,s))| \leq U_n(|h(\cdot,s)|) \leq \varepsilon M_1 + (\|h\|_\infty/\delta) U_n(p(\cdot,s)).$$

Now, $\{p(\cdot,s) : s \in T\}$ is a bounded subset of $\operatorname{ran}[G]$, hence, on it, U_n converges to id *uniformly*. Since $p(s,s) = 0$, this means that, for large n, $U_n(p(\cdot,s))(s)$ is close to zero uniformly in s. Therefore, for all large n,

$$|U_n(h(\cdot,s))(s)| \leq \varepsilon(M_1 + 1).$$

□

Consequently,

$$U_n(f)(s) - \frac{f(s)}{p^*(s)} U_n(p^*)(s) \to 0$$

uniformly in s. Since $U_n(p^*) \to p^*$, this shows that $U_n(f) \to f$. $\square$

From this, (1)Weierstraß follows, with the choices $G = \{()^0, ()^1, ()^2\}$, $p : (t,s) \mapsto (t-s)^2$ and $U_n = B_n$, the **Bernstein operator** (1912, introduced by Bernstein for a different proof of Weierstraß)

$$B_n : f \mapsto \sum_{j=0}^{n} \beta_{j,n-j} f(j/n),$$

with

$$\beta_{j,n-j} : t \mapsto \binom{n}{j} ()^j (1 - \cdot)^{n-j}.$$

Note that

$$B_n()^0(t) = \sum_j \binom{n}{j} t^j (1-t)^{n-j} = (t + (1-t))^n = 1,$$

i.e., $B_n()^0 = ()^0$. Also,

$$B_n()^1(t) = \sum_j \frac{j}{n} \frac{n!}{j!(n-j)!} t^j (1-t)^{n-j} = \sum_j \binom{n-1}{j} t^j (1-t)^{n-j-1} t = t,$$

i.e., $B_n()^1 = ()^1$. In other words, $B_n = \mathrm{id}$ on Π_1. Finally,

$$\sum_j j(j-1) \binom{n}{j} t^j (1-t)^{n-j} = n(n-1) \sum_j \binom{n-2}{j} t^j (1-t)^{n-j-2} t^2 = n(n-1)t^2,$$

i.e., $B_n(()^2 - ()^1/n) = \frac{n-1}{n}()^2$, hence

$$B_n()^2(t) = \frac{n-1}{n} t^2 + \frac{1}{n} t = t^2 + \frac{1}{n}(t - t^2),$$

therefore

$$\|B_n()^2 - ()^2\|_\infty \le \frac{1}{n} \longrightarrow 0.$$

$\square$

(3) d-dim. Weierstraß. *The restriction $\Pi|_T$ of the polynomials in d arguments to any compact subset T of $\mathbb{R}^d$ is dense in $C(T)$.*

Proof: It is sufficient to prove the theorem for the special case $T = I := [0\,..\,1]^d$ since T is compact, hence contained in some axi-parallel box, which, after translation and scaling, we may assume to be I, and, by Tietze's extension theorem, $C(T)$ can be isometrically imbedded into $C(I)$.

Remember: The **Tietze** extension of $f \in C(T)$ to an f_I is given by the rule

$$f_I : x \mapsto \begin{cases} f(x) & x \in T; \\ \inf_{t \in T} f(t) d(x,t)/d(x,T) & x \notin T. \end{cases}$$

A proof is usually given on the way to proving Urisohn's Lemma.

Now choose U_n as the **tensor product** $B_n \otimes \cdots \otimes B_n$ of d copies of the Bernstein operator, i.e.,

$$U_n f : x \mapsto \sum_{k \in \{0,\ldots,n\}^d} f(k/n) \prod_{i=1}^{d} \beta_{k(i),n-k(i)}(x(i)), \quad x \in [0\,..\,1]^d.$$

From the univariate argument, $U_n \xrightarrow{pw}$ id on $\Pi_{2,\ldots,2}$ since $U_n()^\alpha = \otimes_{i=1}^d B_n()^{\alpha_i} : x \mapsto \prod_i B_n()^{\alpha_i}(x(i))$, and a suitable p is

$$p : (x, y) \mapsto \sum_i (x(i) - y(i))^2.$$

Korovkin also supplies a proof that the trigonometric polynomials are dense in $C(\mathbb{T})$, the space of continuous functions on the circle. In this case, the role of p is played by

$$p : (t, s) \mapsto 1 - \cos(t - s),$$

and the maps U_n are the **Fejér** operators, i.e.,

$$\sigma_n : f \mapsto \frac{1}{\pi} \int_{\mathbb{T}} F_n(\cdot - t) f(t)\, \mathrm{d}t$$

with

$$F_n(\theta) := \frac{1}{2(n+1)} \left(\frac{\sin((n+1)\theta/2)}{\sin(\theta/2)} \right)^2.$$

The Fejér operator associates with f the average of its truncated Fourier series of orders $0, 1, \ldots, n$, i.e.,

$$\sigma_n f = \frac{1}{n+1} \sum_{j=0}^{n} s_j f$$

with

$$s_j f := \sum_{-j}^{j} e_{\mathrm{i}j} \int_{\mathbb{T}} \mathrm{e}^{-\mathrm{i}jt} f(t)\, \mathrm{d}t/(2\pi).$$

In particular, σ_n is a positive operator (which s_n is not), and

$$\sigma_n(e_{\mathrm{i}k}) = \frac{(n-k+1)_+}{n+1} e_{\mathrm{i}k},$$

hence $\sigma_n(e_{\mathrm{i}k})$ converges uniformly to $e_{\mathrm{i}k}$ for any k.

Lebesgue's proof (1898) of (1)Weierstraß Let $I := [a\,..\,b]$. For any $f \in C(I)$ and any finite sequence $\xi := (a = \xi_1 < \xi_2 < \cdots < \xi_{\ell+1} = b)$, the broken-line interpolant to f is, by definition, the unique element $P_\xi f$ in $\Pi^0_{1,\xi}$ which agrees with f at ξ. For $\xi_j \le t \le \xi_{j+1}$,

$$f(t) - P_\xi f(t) = (f(t) - f(\xi_j))\frac{\xi_{j+1} - t}{\xi_{j+1} - \xi_j} + (f(t) - f(\xi_{j+1}))\frac{t - \xi_j}{\xi_{j+1} - \xi_j},$$

hence

$$\|f - P_\xi f\|_\infty \le \omega_f(|\xi|),$$

with ω_f the *(uniform) modulus of continuity of* f and

$$|\xi| := \max_j |\xi_{j+1} - \xi_j|.$$

Conclusion: The collection

$$\Pi^0_1(I) := \bigcup_{\xi \text{ in } I} \Pi^0_{1,\xi}$$

of continuous broken lines on I is dense in $C(I)$.

Since each $\Pi^0_{1,\xi}$ is contained in $\Pi_1 + \operatorname{ran}[|\cdot - \xi_j| : j = 2, \ldots, \ell]$, the following Claim therefore finishes the proof. □

(4) Claim. *On each* $[a \,..\, b]$ *and for any* s, $|\cdot - s|$ *can be approximated uniformly by polynomials.*

Proof: For such s, we may choose $\sigma \in \Pi_1 \backslash \Pi_0$ which carries $[a \,..\, b]$ into $[-1 \,..\, 1]$ in such a way that $\sigma(s) = 0$. Further, if $p \in \Pi$ is close to $|\cdot|$ on $[-1 \,..\, 1]$, then $p \circ \sigma$ is a polynomial which, on $[a \,..\, b]$, is that close to $c|\cdot - s|$ for some nonzero c. Hence it is sufficient to prove the Claim for $[a \,..\, b] = [-1 \,..\, 1]$ and $s = 0$. Since $|t| = (t^2)^{1/2}$, i.e.,

$$|\cdot| = \sqrt{\cdot} \circ ()^2,$$

and $()^2([-1 \,..\, 1]) = [0 \,..\, 1]$, and $\Pi \circ ()^2 \subset \Pi$, the following Claim finishes the proof. □

(5) Claim. *The square-root function,* $\sqrt{\cdot}$, *can be uniformly approximated on* $[0 \,..\, 1]$ *by polynomials.*

Proof: Here are two proofs from the literature.

The Taylor series expansion

$$\sqrt{1 - \cdot} = 1 - \sum_{n>0} ()^n \prod_{k=1}^{n} |3/2 - k|/k$$

converges uniformly on any compact subset of $(-1 \,..\, 1)$ (since the power coefficients are all bounded by 1). Further, on $[0 \,..\, 1]$ (which is our range of interest) all terms but the first decrease monotonely from their value, 0, at 0. This implies that the n-th partial sum,

p_n, of the series is a monotonely decreasing polynomial whose minimum value therefore occurs at 1 and whose difference from $\sqrt{\cdot}$ is nonnegative and largest at 1. Therefore,

$$\|\sqrt{1-\cdot}-p_n\| = \max_{0\le t\le 1}(p_n(t)-\sqrt{1-t}) = p_n(1) \le p_n(t) = \sqrt{1-t} + (p_n(t)-\sqrt{1-t})$$

for any t, and both terms on the right can be made small, the first by choosing t close to 1, and the second by then choosing n sufficiently large. We conclude that $p_n(1-\cdot)$ converges to $\sqrt{\cdot}$ in the norm of $C[0\,..\,1]$.

The following proof can be found in Dieudonné's Analysis. Define the sequence $(u_n : n=0,1,2\ldots)$ of polynomials by the iteration

$$u_{n+1}(t) := u_n(t) + (t-u_n(t)^2)/2, \quad n=0,1,2,\ldots$$

with

$$u_0 := 0.$$

Claim: for all n,

$$u_n(t) \le \sqrt{t}, \quad t\in[0\,..\,1].$$

Indeed, assuming this already to be true for n (as it is for $n=0$), we compute

$$\sqrt{t}-u_{n+1}(t) = (\sqrt{t}-u_n(t))(1-(\sqrt{t}+u_n(t))/2) \ge (\sqrt{t}-u_n(t))(1-\sqrt{t}) \ge 0$$

for $t\in[0\,..\,1]$.

It follows that $t-u_n(t)^2\ge 0$ on $[0\,..\,1]$, hence (u_n) is monotone increasing, yet bounded on $[0\,..\,1]$, therefore uniformly convergent there, and its limit is necessarily a fixed point of the iteration used to define it, hence its limit is $\sqrt{\cdot}$. □

Stone-Weierstraß

For an arbitrary set T, the collection $\mathbb{F}^T$ of all scalar-valued maps (whether real or complex) is not only closed under (pointwise) addition and multiplication by a scalar, but also closed under (pointwise) multiplication of two elements, i.e., for $f,g\in\mathbb{F}^T$, also

$$fg : T\to\mathbb{F} : t\mapsto f(t)g(t)$$

is in $\mathbb{F}^T$. This makes $\mathbb{F}^T$ a **ring with identity**. Since also

$$f(\alpha g) = (\alpha f)g = \alpha(fg), \quad \alpha\in\mathbb{F},\, f,g\in\mathbb{F}^T,$$

$\mathbb{F}^T$ is an **algebra with identity**. Being an algebra, $\mathbb{F}^T$ is closed under formation of polynomials; i.e.,

$$\Pi\times\mathbb{F}^T\to\mathbb{F}^T : (p,f)\mapsto p\circ f$$

(with $p \circ f$ **map composition**, i.e., $p \circ f : T \to \mathbb{F} : t \mapsto p(f(t))$) is well-defined. Now assume that T is compact Hausdorff. Then $C(T)$ is a closed subalgebra of $\mathbb{F}^T$. Further, as the composition of continuous maps, $p \circ f \in C(T)$ if $f \in C(T)$.

The Stone(-Weierstraß) theorem employs the following notion: $A \subset \mathbb{F}^T$ is said to **separate points** if, for any two distinct points $s, t \in T$, there exists $a \in A$ with $a(s) \neq a(t)$. If A is a linear subspace and also contains the identity, then this is equivalent to the statement that the linear map

$$A \to \mathbb{F}^2 : a \mapsto (a(s), a(t))$$

is onto, hence has right inverses. In particular, for any $f \in \mathbb{F}^T$ and any $s, t \in T$, there exists $a_{f,s,t} \in A$ which agrees with f at s and t. It is this conclusion that is needed.

(6) Stone(-Weierstraß) (1937). *Let T be compact metric. The only closed subalgebra A of* real *$C(T)$ with identity which separates points is $C(T)$ itself.*

Proof: The range $a(T)$ of any $a \in A$ is a bounded subset of $\mathbb{R}$ (since T is compact), hence, by Claim 4, $|\cdot|$ can be approximated, uniformly on $a(T)$, by polynomials p, and A, being an algebra, contains $p \circ a$ for any polynomial. This implies that A, being closed, is closed under formation of absolute values; i.e.,

$$a \in A \quad \Longrightarrow \quad |a| \in A.$$

Since

$$\max\{f, g\} : T \to \mathbb{R} : t \mapsto \max\{f(t), g(t)\}$$

can also be written

$$\max\{f, g\} = ((f+g) + |f-g|)/2,$$

A is also closed under formation of the maximum of *finitely* many functions. Since

$$\min\{f, g\} = -\max\{-f, -g\},$$

A is also closed under formation of minimum.

Take $f \in C(T)$, $\varepsilon > 0$. Let $t \in T$. For each $s \in T$, there is $a_{s,t} \in A$ agreeing with f at s and t, hence, there is some neighborhood U_s of s on which $f - \varepsilon < a_{s,t}$. T being compact, there exists a *finite* set S for which $T = \cup_{s \in S} U_s$. Hence, the function

$$a_t := \max_{s \in S} a_{s,t}$$

is in A and satisfies

$$f - \varepsilon < a_t \quad \text{on } T, \quad a_t(t) = f(t).$$

The latter implies that, on some neighborhood V_t of t, $f + \varepsilon > a_t$. Hence, with S a finite set for which $T = \cup_{t \in S} V_t$, the function

$$a := \min_{t \in S} \ a_t$$

is in A and satisfies

$$f + \varepsilon > a$$

as well as $f - \varepsilon < a$. Consequently, $\|f - a\|_\infty < \varepsilon$. As ε is arbitrary and A is closed, $f \in A$ follows. □

From this, even (3) follows since Π is an algebra with identity which separates points.

Another consequence is the density of the *even* trigonometric polynomials, i.e., of $\mathrm{ran}[\cos(j\cdot) : j = 0, 1, 2, \ldots]$ in $C([0 \,..\, \pi]$ (the function cos alone is enough to separate points). Note that this algebra fails to be dense in $C([0 \,..\, \alpha])$ for any $\alpha > \pi$.

The restriction to *real*-valued functions is essential here. Not only does the proof of Stone make explicit use of it (it relies on the total ordering of the reals), but the polynomials fail to be dense in $C(\{z \in \mathbb{C} : |z| \le 1\})$ of all complex-valued continuous functions of one complex argument on the **unit disk** (even though they continue to form there an algebra with identity that separates points), since their closure consists of functions *analytic* in the interior of the disk.

Existence

$M \subset X$ is called an **existence set** in case $\forall \{x \in X\}\ \mathcal{P}_M(x) \ne \emptyset$. Such a set is necessarily closed, hence M is assumed to be closed from now on.

Having M closed does not ensure existence, even when M is a linear subspace and X is a Banach space. A standard example is the kernel $M := \ker \lambda$ of any continuous linear functional λ which **does not take on its norm**, i.e., for which there is no *nontrivial* $g \in X$ with $|\lambda g| = \|\lambda\| \|g\|$. For,

$$|\lambda f| = \|\lambda\| \operatorname{dist}(f, \ker \lambda), \quad \forall\, f \in X, \lambda \in X^*.$$

Hence, if $\|f - m\| = \operatorname{dist}(f, \ker \lambda)$ with $m \in \ker \lambda$, then

$$\|\lambda\| \operatorname{dist}(f, \ker \lambda) = |\lambda f| = |\lambda(f - m)| \le \|\lambda\| \|f - m\| = \|\lambda\| \operatorname{dist}(f, \ker \lambda),$$

therefore, necessarily, $|\lambda(f - m)| = \|\lambda\| \|f - m\|$. Hence, if λ fails to take on its norm, then this can only happen if $f = m$, i.e., if $f \in M$ (in which case $\mathcal{P}_M(f) = \{f\}$ trivially).

Specifically, take

$$X = \ell_1 := \{a \in \mathbb{R}^{\mathbb{N}} : \|a\|_1 := \sum_j |a(j)| < \infty\}.$$

Take

$$\lambda : a \mapsto \sum_j (1 - 1/j) a(j).$$

Then

$$|\lambda a| \le \sum_j (1 - 1/j) |a(j)| \le \sum_j |a(j)| = \|a\|_1,$$

and the last inequality is sharp since, for any j, $\lambda \mathbf{i}_j = 1 - 1/j \xrightarrow[j \to \infty]{} 1 = \|\mathbf{i}_j\|_1$. However, $|\lambda a| = \|a\|_1$ implies equality throughout and that's not possible unless, for each j, $|a(j)| = 0$, i.e., $a = 0$.

In effect, in the example, the sequence $(\mathbf{i}_j : j \in \mathbb{N})$ is a maximizing sequence for

$$\|\lambda\| = \sup_{x \in X} |\lambda x|/\|x\|,$$

but this sequence fails to have limit points, hence the sup is not taken on.

Put positively, existence of ba's from M is usually proved by establishing that, in some weak enough topology, closed and bounded subsets of M are compact while the topology is still strong enough to have $x \mapsto \|x\|$ at least lower semicontinuous (i.e., $x_n \to x \implies \liminf_n \|x_n\| \geq \|x\|$).

(7) Proposition. *Any finite-dimensional linear subspace M of any nls X is an existence set.*

Proof: Since $\mathcal{P}_M(f)$ necessarily lies in the intersection

$$M_f := M \cap B^-_{2\,\mathrm{dist}(f,M)}(f)$$

of M with the closed ball around f with radius twice the distance of f from M, we have M_f not empty and

$$\mathcal{P}_M(f) = \mathcal{P}_{M_f}(f).$$

M is closed (since it is finite-dimensional), therefore M_f is a closed and bounded subset of a finite-dimensional ls, hence compact. In particular, the continuous function $m \mapsto \|f - m\|$ takes on its infimum on M_f. □

The same proof supports the claim that any closed set M which is finite-dimensional in the sense that its affine hull is finite-dimensional, is an existence set. More than that, if M is merely finite-dimensional in the sense that it is the image of some closed subset of $\mathbb{F}^n$ under a continuous map with continuous inverse (as a map to M), then M is an existence set.

However, here is another example. Let $X = L_p[0\,..\,1]$, for some $p \in [1\,..\,\infty]$, and

$$M = \Pi_{k,l}[0\,..\,1]$$

the collection of all pp functions of degree $\leq k$ on $[0\,..1]$ with l pieces. For $k = 0$, this would be the space of **splines of order 1 with $l - 1$ free knots**. This collection includes also all pp functions of degree $\leq k$ with fewer than l breakpoints since the latter are obtained from the former description when some adjacent polynomial pieces happen to come from the same polynomial. This, however, indicates the technical difficulty to be overcome here: M appears to be finite-dimensional in the sense that each of its elements is specified by finitely many parameters or degrees of freedom, namely the $l - 1$ (interior) breakpoints and the $(k + 1)l$ polynomial coefficients. However, this parametrization is badly flawed since any element with fewer than $l - 1$ **active** breakpoints can be parametrized this way in infinitely many different ways, as one can think of the inactive breakpoints as being located anywhere. In effect, the natural parametrization describes this M as the image of

$$\{\xi : 0 = \xi_1 < \cdots < \xi_{l+1} = 1\} \times \mathbb{R}^{(k+1)l},$$

with the map failing to be 1-1 on the boundary of this domain, yet no better parametrization is at hand.

Let $(s_n : n \in \mathbb{N})$ be a minimizing sequence from $M = \Pi_{k,l}$ for f, i.e., $\lim_n \|f - s_n\| = \mathrm{dist}(f, M)$. Without loss, we assume that each s_n consists of exactly l pieces and, AGTASMAT (:= After Going To A Subsequence May Assume That) the corresponding sequence $(\xi^{(n)} : n \in \mathbb{N})$ of breakpoint sequences $\xi^{(n)} := (0 = \xi_1^{(n)} < \cdots < \xi_{l+1}^{(n)} = 1)$ converges to some $(l+1)$-vector ξ.

Assume first that ξ is strictly increasing. Let $(p_{j,n} : j = 1, \ldots, l)$ be the sequence of polynomial pieces which make up s_n, with $p_{j,n}$ the piece corresponding to the interval

$$I_{j,n} := (\xi_j^{(n)} \,..\, \xi_{j+1}^{(n)}).$$

Since s_n is a minimizing sequence, it is, in particular, bounded, and this implies that, for each j, the sequence $(\|p_{j,n}\|(I_{j,n}) : n \in \mathbb{N})$ is bounded. Here and below, and for any domain-dependent norm $\|\cdot\|$ such as the L_p-norms on some domain T, and for any subset U of that domain, $\|f\|(U)$ is the same norm, but for the domain U. Since the endpoints of $I_{j,n}$ converge to the endpoints of $I_j := [\xi_j \,..\, \xi_{j+1}]$, it follows that, for some slightly smaller, but still nontrivial, interval $\hat{I}_j$, the sequence $(\|p_{j,n}\|(\hat{I}_j) : n \text{ large })$ is bounded. Since $p \mapsto \|p\|(\hat{I}_j)$ is a norm on Π_k, and Π_k is finite-dimensional, it follows AGTASMAT that $p_{j,n}$ converges to some $p_j \in \Pi_k$, and since there are only finitely many j involved, we can make this assumption for all j. The argument is finished by verifying that the pp s with break sequence ξ and polynomial pieces $(p_j : j = 1, \ldots, l)$ is, indeed, the norm limit of (s_n). This implies that

$$\mathrm{dist}(f, M) \le \|f - s\| = \lim_n \|f - s_n\| = \mathrm{dist}(f, M),$$

hence $s \in \mathcal{P}_M(f)$.

This argument is not only a little bit shaky (since the verification mentioned two sentences ago was not explicitly carried out), but runs into trouble in case there is **coalescence**, i.e., in case ξ is not strictly increasing. This implies that, for some j, $I_j = \lim_n I_{j,n}$ has no interior, hence the boundedness of the sequence $(\|p_{j,n}\|(I_{j,n}) : n \in \mathbb{N})$ does not force convergence of some subsequence of $(p_{j,n} : n)$. At the same time, the fact that I_j is trivial would suggest that, somehow, we don't care about this polynomial sequence. On the other hand, we cannot simply ignore it. Or can we?

Here is a soft-analysis approach around this. A sequence (x_n) in a nls X is said to **nearly converge to** $x \in X$ if

$$\forall \{y \in X\} \quad \liminf_n \|x_n - y\| \ge \|x - y\|. \tag{8}$$

Further, a subset Y of X is **nearly compact in** Z if every sequence in Y has a nearly converging subsequence, with 'near-limit' in Z.

This is certainly a useful notion here. For:

(9) Proposition. *If bounded subsets of $M \subset X$ are nearly compact in M, then M is an existence set.*

Proof: Any minimizing sequence (m_n) in M for $f \in X$ is necessarily bounded, hence, AGTASMAT, there exists $m \in M$ so that

$$\mathrm{dist}(f, M) \le \|f - m\| \le \liminf_n \|f - m_n\| = \mathrm{dist}(f, M),$$

hence $m \in \mathcal{P}_M(f)$. □

(It would have sufficed here to work with the even weaker notion of demanding only that

$$\forall\{y \in X\} \quad \limsup_n \|x_n - y\| \ge \|x - y\|.$$

However, such a notion of "convergence" isn't even preserved by going to subsequences.)

Here is a ready source for nearly convergent sequences.

(10) Proposition. *Assume that Φ is a collection of seminorms on the nls X, and that*

$$\sup_{\varphi \in \Phi} \varphi(x) = \|x\|, \quad x \in X.$$

If (x_n) **Φ-converges to** *x, i.e.,*

$$\forall\{\varphi \in \Phi\} \quad \lim_n \varphi(x - x_n) = 0,$$

then (x_n) nearly converges to x.

Proof: Since seminorms are translation-invariant, it is sufficient to prove (8) for just $y = 0$. For this, since $\varphi(x_n) \le \|x_n\|$, we have $\varphi(x) = \lim_n \varphi(x_n) \le \liminf_n \|x_n\|$, hence

$$\|x\| = \sup_{\varphi \in \Phi} \varphi(x) \le \liminf_n \|x_n\|.$$

□

Returning to the example, choose $\Phi = \{\varphi_\varepsilon : \varepsilon > 0\}$, with

$$\varphi_\varepsilon : x \mapsto \|x\|([0\,..\,1] \backslash B_\varepsilon(\xi)).$$

For any $p \in [1\,..\,\infty]$, these are seminorms on $X = L_p[0\,..\,1]$ satisfying the assumptions of Proposition 10, hence also its conclusion. The above selection process is sure to provide a subsequence which, for some $\varepsilon > 0$, φ_ε-converges to a certain element of $\Pi_{k,l}$. Since Π_k is finite-dimensional, this implies that the subsequence φ_η-converges to that element for every $0 < \eta \le \varepsilon$, hence nearly converges to it, and the proof of Proposition 9 does the rest.

□

The same arguments also settle existence questions when dealing with *smooth* piecewise polynomials. These are the elements in the space

$$\Pi_{k,\xi}^{\rho},$$

consisting of all $s \in \Pi_{k,\xi}$ which are in $C^{(\rho_i)}(\xi_i)$, all i. If ρ is just an integer, it is taken to stand for the corresponding *constant* sequence. In any event, each entry of ρ is restricted to be no greater than $k+1$ since $\rho_i = k+1$ implies infinite smoothness at ξ_i.

To be precise, the above arguments guarantee that any minimizing sequence from

$$\bigcup_{0=\xi_1<\cdots<\xi_{l+1}=1} \Pi_{k,\xi}^{\rho},$$

leads to an element in some $\Pi_{k,\xi}$, but not all elements of $\Pi_{k,\xi}$ can appear. For example, if $l = 2$ and $\rho = k - 1 > 0$, then only elements of some $\Pi_{k,\xi}^{\rho}$ can appear. It is useful to discuss this question in the more general context of γ-polynomials.

γ-polynomials

Let $\gamma : I \to X$ be a **curve in the nls** X, i.e., a continuous map from some (finite or infinite) interval I into X. We are interested in the approximating family

$$M_{\gamma,l} := \bigcup_{\xi_1<\cdots<\xi_l} \operatorname{ran}[\gamma(\xi_1), \ldots, \gamma(\xi_l)].$$

Our particular concern is the special case

$$\gamma : x \mapsto (\cdot - x)_+^k,$$

with

$$()_+ : \mathbb{R} \to \mathbb{R} : s \mapsto (s + |s|)/2$$

the **truncation map**, and with X a normed linear space of functions on I, such as $L_p(I)$, as this leads to splines. However, here are other very useful examples.

- *exponentials*, where
$$\gamma(t) := e_t;$$
- *rationals*, where
$$\gamma(t) := 1/(\cdot - t).$$
- *point functionals*, where
$$\gamma : I \to X^* : t \mapsto \delta_t,$$

with X some space of functions on I.

The last example is dear to Numerical Analysts who like to approximate linear functionals, like $f \mapsto \int_I f$, by linear combinations of values (and, perhaps, derivatives). It is prototypical, since, for all the others, there exists a space Y of functions on I and a pairing $\langle\cdot,\cdot\rangle$ so that $\gamma(t)$ is (up to a factor) the **representer** of evaluation at t, i.e., so that

$$y(t) = \langle y, \gamma(t)\rangle, \quad y \in Y,\, t \in I.$$

For example, the **truncated Taylor series with remainder**

$$f = \sum_{j<k} D^j f(a)()^j/j! + \int_a^b (\cdot - t)_+^{k-1}/(k-1)!\, D^k f(t)\,\mathrm{d}t \quad \text{on } [a \mathinner{.\,.} b]$$

shows that $k(\tau - \cdot)_+^{k-1}$ represents δ_τ on $Y := \{f \in C^{(k)}[a \mathinner{.\,.} b] : f \ \text{ vanishes } k\text{-fold at } 0\}$ with respect to the pairing $(g, f) \mapsto \int_a^b g\, D^k f/k!$.

In the context of *existence*, it becomes important to know whether $M_{\gamma,l}$ is closed. More precisely, it becomes important to know the possible near-limits of minimizing sequences in $M_{\gamma,l}$. Since near convergence is a weaker concept than (norm) convergence, we may have to go beyond the (norm-)closure of $M_{\gamma,l}$ in order to get an existence set. However, we need to know the norm-closure $M_{\gamma,l}^-$ of $M_{\gamma,l}$ in any case.

If γ is a *smooth* curve, in the sense that the (norm-)limit

$$D\gamma(x) := \lim_{y\to x} (\gamma(x) - \gamma(y))/(x - y)$$

exists, then also $D\gamma(x)$, being a norm-limit of elements of $M_{\gamma,l}$, must be in $M_{\gamma,l}^-$. If γ is even smoother, in the sense that the 'tangent' curve $D\gamma$ is itself smooth, then also all points on the second derived curve $D^2\gamma$ must be in $M_{\gamma,l}^-$. For a discussion of the general case, divided differences are handy.

Divided differences

To recall (see almost any textbook on Numerical Analysis), the **divided difference of f at the point sequence** $\tau := (\tau_0, \ldots, \tau_k)$ is denoted by

$$\delta_\tau f$$

(other notations are $f[\tau_0, \ldots, \tau_k]$ and $[\tau_0, \ldots, \tau_k]f$), and is, by definition, the coefficient of $()^k$ in the **power form** $I_\tau f =: \sum_{j=0}^k ()^j a(j) =: p$ of the unique polynomial p of degree $\le k$ which **agrees with f at τ** in the sense that, for any z which appears at least j-fold in τ,

$$D^{j-1}(f - p)(z) = 0.$$

E.g., if $\tau = (z, \ldots, z)$, then $I_\tau f$ is Taylor's expansion at $a = z$, truncated after the term of degree k, and the divided difference is the highest coefficient, i.e., the number $\delta_z D^k f/k!$. In particular,

$$\delta_{(\tau_0)} = \delta_{\tau_0}.$$

In general, the divided difference equals $D^k f(x)/k!$ for some x in the interval spanned by the points in τ. Note that, directly from its definition, δ_τ is a *symmetric* function of its

$k+1$ arguments. Also, as a linear functional on $C^{(k)}$, δ_τ is a continuous function of its arguments. Also, for $x \neq y$,

$$\delta_{x,y} = \frac{\delta_y - \delta_x}{y - x}.$$

More generally, if x, y are any two *distinct* points in τ, then

$$\delta_\tau = \frac{\delta_{\tau\backslash x} - \delta_{\tau\backslash y}}{y - x} \tag{11}$$

(with $\tau\backslash x$ the sequence obtained from τ by omission of (one occurrence of) x). By induction on k, this shows that a divided difference at $k+1$ points can be obtained from divided differences involving just one point by k-fold formations of difference quotients. For this reason, $\delta_\tau f$ is also called **a k-th divided difference**.

The divided differences appear naturally as the coefficients in the **Newton form** of the interpolating polynomial:

$$I_\tau f = \sum_j \prod_{i<j} (\cdot - \tau_i)\ \delta_{\tau_0,\ldots,\tau_j} f.$$

Divided differences of the truncated power 'curve'

Return to the discussion of the γ-polynomials, with γ from now on the specific curve

$$\gamma : x \mapsto (\cdot - x)_+^j,$$

say as a map into $L_p(I)$ for some finite interval I (to be specific). Assume first that $j > 0$. If $x < y$, then

$$\delta_{x,y}\gamma = \frac{(\cdot - x)_+^j - (\cdot - y)_+^j}{x - y},$$

and this function is pp of degree j, with two breaks, at x and at y. It is zero at any $t < x$. Further, at any $t > y$, we can drop the subscript $_+$, i.e., the value equals

$$\delta_{x,y}(t - \cdot)^j = \frac{(t-x)^j - (t-y)^j}{x - y}.$$

This is a polynomial in t of degree $< j$, since it is the difference of two polynomials of degree j with the same leading coefficient. In any case, since $(t - \cdot)^j$ is a smooth function, $\delta_{x,y}(t - \cdot)^j = -j(t - \zeta(t))^{j-1}$ for some $\zeta(t) \in [x\,..\,y]$. Hence, as $y \to x$, $\delta_{x,y}\gamma$ near-converges (in L_p) to

$$D\gamma(x) = -j(\cdot - x)_+^{j-1}.$$

Thus, if $x = y$, then

$$\delta_{x,y}\gamma = D\gamma(x) = -j(\cdot - x)_+^{j-1}.$$

In fact, this convergence is even in norm, except for $j = 1$ and $p = \infty$. For, when $j = 1$, then

$$\delta_{x,y}\gamma : t \mapsto \begin{cases} 0 & t < x; \\ \frac{t-x}{x-y} & x \le t < y; \\ -1 & y \le t, \end{cases}$$

and, as $y \to x$, this function goes *pointwise* to the function $-(\cdot - x)_+^0$, at least for any $t \ne x$. At $t = x$, the function $(\cdot - x)_+^0$ is, offhand, not defined; by custom, such piecewise continuous functions are taken, in spline theory, to be continuous from the right (i.e., $(0)_+^0 := 1$), except at the right endpoint of their domain of definition. However, for $p = \infty$,

$$\|\delta_{x,y}\gamma - (-(\cdot - x)_+^0)\| = 1$$

unless $y = x$.

To summarize: if $x \le y$, then $\delta_{x,y}\gamma$ is pp of degree $\le j$, it vanishes to the left of x, is a polynomial of degree $j - 1$ to the right of y, has breaks just at x and y, is in $C^{(j-1)}$ at its breakpoints in case they are different, but is only $C^{(j-2)}$ at x if also $x = y$.

What about $j = 0$? In this case,

$$\delta_{x,y}\gamma : t \mapsto \begin{cases} 0 & t < x; \\ \frac{1}{y-x} & x \le t < y; \\ 0 & y \le t, \end{cases}$$

and this near-converges to 0 in any L_p, hence fails to converge in any L_p since

$$\|\delta_{x,y}\gamma\|_p = (y - x)^{1/p-1}.$$

It follows that this curve has no tangent anywhere. In particular, for $p = 2$, this γ provides a curve in $L_p(I)$ with the property that, for any $x < y < z$, the secant $\gamma(y) - \gamma(x)$ is perpendicular to the secant $\gamma(z) - \gamma(y)$. (The whole concept of γ-polynomials had its beginning in this observation which was the answer to a homework question in a Hilbert space course by Halmos asking for just such a curve.) On the other hand, if we think of γ as a curve in $C[a \mathinner{..} b]^*$, via

$$\gamma(x) : f \mapsto \int_a^b (\cdot - x)_+^0 f,$$

then $\delta_{x,y}\gamma : f \mapsto \int_x^y f/(y - x)$, hence $\delta_{x,y}\gamma$ converges, in the norm of $C[a \mathinner{..} b]^*$, to δ_x.

Next, consider $\delta_{x,y,z}\gamma$ with $x \le y \le z$. If $x < z$, then, by (11), this is

$$\delta_{x,y,z}\gamma = \frac{\delta_{x,y}\gamma - \delta_{y,z}\gamma}{x - z}.$$

In particular, it vanishes for $t < x$ and is a polynomial of degree $j - 2$ for $t > z$ (since both $\delta_{x,y}\gamma$ and $\delta_{y,z}\gamma$ are polynomials of degree $j - 1$ there with the same leading coefficient). In particular, $\delta_{x,y,z}\gamma$ is zero for $t > z$ in case $j < 2$. Altogether, it is pp of degree j with breaks at x, y, and z. If $x < y < z$, then it is in $C^{(j-1)}$ at all its breaks. However, if, e.g., $x = y < z$, then, because of the term $\delta_{x,y}\gamma = -j(\cdot - x)^{j-1}$, it is only $C^{(j-2)}$ at x. If $j > 1$, then $\delta_{x,y,z}\gamma$ near-converges to $\delta_{x,x,x}\gamma = D^2\gamma(x) = j(j-1)(\cdot - x)_+^{j-2}$ and this convergence is in the L_p-norm except when $p = \infty$ and $j = 2$.

Induction and the divided difference formula (11) now give the following.

(12) Fact. *If $\tau = (\tau_0 \le \cdots \le \tau_k)$, and $\gamma : x \mapsto (\cdot - x)_+^j$, then $\delta_\tau \gamma$ vanishes to the left of τ_0 and is a polynomial of degree $\le j - k$ to the right of τ_k. In particular, it also vanishes to the right of τ_k in case $j < k$. Further, it is pp of degree j, with breaks at $\tau_0, \ldots, \tau_k$, and in $C^{(j-1-\mu)}(z)$ in case z occurs no more than μ times in τ.*

Splines

I spent all this time on the truncated power because of its simple connection to splines. If p, q are two polynomials of degree $< k$, the piecewise polynomial function

$$s : t \mapsto \begin{cases} p(t) & t < x; \\ q(t) & x \le t, \end{cases}$$

is in $C^{(\rho)}(x)$ iff $D^r p = D^r q$ at x for $r = 0, \ldots, \rho$, i.e., iff

$$q - p = \sum_{\rho < r < k} (\cdot - x)^r a(r)$$

For some coefficients a, i.e., iff

$$s = p + \sum_{\rho < r < k} (\cdot - x)_+^r a(r). \tag{13}$$

Now, with

$$\gamma : x \mapsto (\cdot - x)_+^{k-1}$$

(note the switch from j to $k-1$), we can think of $(\cdot - x)_+^r$ as a (nontrivial) scalar multiple of $D^{k-1-r}\gamma(x)$, hence, in particular, as the limit of $\delta_{(\zeta_1,\ldots,\zeta_{k-r})}\gamma$ as the $k - r$ points ζ_ν all coalesce at x. This implies that we can think of s as having been obtained in the limit from elements of

$$\Pi_{<k} + \mathrm{ran}[(\cdot - \zeta_j)_+^{k-1} : j = 1, \ldots, \mu], \tag{14}$$

as the

$$\mu := k - \rho - 1$$

points $\zeta_1 < \cdots < \zeta_\mu$ all coalesce at x. The elements of (14) are pp functions which are *as smooth as possible*, meaning that they have breaks but could not be any smoother at any break without, in effect, losing that break. These are the original **splines**, so-called by Schoenberg in his fundamental '46 paper. We follow him in calling the points ζ_j the **knots** of such a spline. Because s in (13) is the limit of such splines as the μ knots coalesce, we now call s a spline with a knot of **multiplicity** μ at x.

It follows that

$$\Pi^\rho_{<k,\xi} = \Pi_{<k} + \mathrm{ran}[(\cdot - \xi_i)_+^j : \rho_i < j < k;\ i = 2, \ldots, l]$$

becomes, in this way, the space of splines of degree $< k$ on $[\xi_1 \,..\, \xi_{l+1}]$ with (interior) knots ξ_i of multiplicity $\mu_i := k - \rho_i - 1$, $i = 2, \ldots, l$.

We will show soon that this space, considered as a space of functions on the **basic interval**

$$[\xi_1 \,..\, \xi_{l+1}] = [t_k \,..\, t_{n+1}],$$

has a basis of the form

$$\delta_{t_j,\ldots,t_{j+k}}\gamma, \quad j = 1, \ldots, n,$$

with $\mathbf{t} = (t_j : j = 1, \ldots, n + k)$, the so-called **knot sequence** for this spline space, the shortest nondecreasing sequence which contains, for $i = 2, \ldots, l$, the point ξ_i exactly $\mu_i = k - \rho_i - 1$ times. For this reason, we will write that space

$$S_{k,\mathbf{t}}$$

and call it the space of splines of **order** k with knot sequence $\mathbf{t}$. Following Schoenberg, the function $\delta_{t_j,\ldots,t_{j+k}}\gamma$ is called a **B-spline with knots** $t_j, \ldots, t_{j+k}$. It is usually *normalized*, i.e., suitably scaled, in some way. Already at this point, we can infer, from Fact 12, a major attraction of such a B-spline, namely its 'local support': *A B-spline vanishes outside the interval spanned by its knots.*

Uniqueness

The question of uniqueness is, more generally, the question of the nature of $\mathcal{P}_M(f)$. By definition,

$$\mathcal{P}_M(f) = M \cap B^-_{\operatorname{dist} f,M}(f).$$

Hence, if M is convex, then so is $\mathcal{P}_M(f)$, and may well contain nontrivial line segments if, e.g., M is a linear subspace and the boundary ∂B of the unit ball, B, contains line segments, i.e., the norm fails to be *strictly convex*. The latter is the case for $\ell_p(n)$ (which is $\mathbb{F}^n$ with the ℓ_p-norm) iff $p = 1, \infty$. It is also the case for L_p with $p = 1, \infty$. In these spaces, we expect nonuniqueness of best approximation even from finite-dimensional spaces. (E.g., $X = L_1[0 \,..\, 1]$, $M = \Pi_0$, $f = \chi_{[.5..1]}$.) However, nonuniqueness is not guaranteed; it depends on the attitude of M with respect to $B_{\operatorname{dist}(f,M)}(f)$. (In the example just given, one gets uniqueness for any $f = \chi_{[x..1]}$ *except* when $x = .5$.) Draw pictures in $\ell_p(2)$ for $p = 1, 2, \infty$.

A set M which provides exactly one ba for every $f \in X$ is called by some a **Chebyshev** set. Any Chebyshev set M in a nls X induces a map P_M, called the **metric projector for** M, by the rule

$$\mathcal{P}_M(f) =: \{P_M f\}, \quad f \in X. \tag{15}$$

Any finite-dimensional Chebyshev subspace M of $C[a \,..\, b]$ (e.g., $M = \Pi_k$) provides not only a *unique* best approximation, but something called **strong uniqueness**. This means that, for every $f \in C[a \,..\, b]$, there exists a positive constant c so that, for all $m \in M$,

$$\|f - m\| \geq \|f - P_M f\| + c\|m - P_M f\|.$$

This reflects the geometric fact that the unit ball in C has corners.

(16) Proposition. *The metric projector of any finite-dimensional Chebyshev subspace M of any nls X is continuous.*

Proof: Assume that $f_n \to f$. We are to prove that $P_M f_n \to P_M f$. For this, it is sufficient to prove that $P_M f$ is the unique limit point of $(P_M f_n)$. For this, observe that $\|P_M g\| \le \|g\|$ for any g, hence the convergence of (f_n) implies that $(P_M f_n)$ is bounded, hence has limit points, as a bounded subset of a finite-dimensional ls. However, any such limit point necessarily equals $P_M f$, by the following Lemma. □

(17) Lemma. *If X is ms, and (f_n) is a sequence in X with limit f, and (m_n) is a corresponding sequence with $m_n \in \mathcal{P}_M(f_n)$ which converges to some m, then $m \in \mathcal{P}_M(f)$.*

Proof: By an earlier assumption, M is closed, hence $m \in M$. Therefore, if not $m \in \mathcal{P}_M(f)$, then there would exist $m' \in M$ with $d(f, m') < d(f, m)$. We use the triangle inequality to conclude that

$$d(m', f_n) - d(f_n, f) \le d(m', f) < d(m, f) \le d(m, m_n) + d(m_n, f_n) + d(f_n, f).$$

Since $d(f_n, f)$, $d(m, m_n)$, and $d(f_n, f)$ all go to zero as $n \to \infty$, this would imply that, for sufficiently large n,

$$d(m', f_n) < d(m_n, f_n),$$

contradicting the fact that $m_n \in \mathcal{P}_M(f_n)$. □

Note that, even when M is a linear subspace, the metric projector is usually *nonlinear* hence not all that easy to construct or apply. Inner product spaces (or, equivalently, *least-squares approximation*) are so popular precisely because the resulting metric projector is linear. (In fact, one of the simpler characterizations of inner product spaces describes them as the normed linear spaces in which the metric projector for every three-dimensional linear subspace is linear.)

Much paper has been used in attacking the **metric selection problem**, which is the problem of understanding in what circumstances an existence set permits the construction of a *continuous* map P_M satisfying (15). Since the resulting constructions are often not practical, it is, from a practical point of view, much more interesting to construct *near-best* projectors.

If M is a *nonconvex* existence set in the nls X, then nonuniqueness is guaranteed, particularly if f is 'very far' from M. Draw the picture. More than that, there are perfectly reasonable M which have 'corners' in the sense that $\#\mathcal{P}_M(f) > 1$ for certain f arbitrarily close to M. Here is one example.

(18) Example Consider $M := \Pi_{0,2}$ in $L_2[-1 \,..\, 1]$, say, and take $f = \alpha()^2$. Since M is closed under multiplication by scalars, we have $\mathcal{P}_M(\alpha()^2) = |\alpha|\mathcal{P}_M(()^2)$, hence it is sufficient to consider $f = ()^2$. Let $m \in \mathcal{P}_M(f)$, and let ζ be its sole breakpoint. Since f is even, also $m' : t \mapsto m(-t)$ is in $\mathcal{P}_M(f)$. Hence, uniqueness would imply that $m = m'$, therefore, $m \in \Pi_0$. This would imply that, for any $\zeta \in [-1 \,..\, 1]$, $m + \mathrm{ran}[m_\zeta] \subset M$, with $m_\zeta := \chi_{[\zeta..1]}$, hence, necessarily, $f - m$ must be orthogonal to m_ζ for every such ζ. However, $\mathrm{ran}[m_\zeta : -1 < \zeta < 1]$ is dense in $L_2[-1 \,..\, 1]$, therefore $f = m$ would follow, contradicting the fact that $f \notin \Pi_0$. □

The same argument shows that, in any L_2-best approximation by splines with l simple (free or variable) knots, all l knots are active (though coalescence is, of course, possible).

Characterization

The standard characterization theorems for ba's are in terms of linear functionals (which is not too surprising since the derivative of the scalar-valued map $m \mapsto \|f - m\|$ at some $m \neq f$ is necessarily a linear functional if it exists).

The action of a continuous linear functional λ on a nls X over the *real* scalars is very simple: The kernel of λ cuts X into two halfspaces, on one λ is positive, on the other it is negative. Further, λ is constant on **hyperplane**s parallel to the kernel, i.e.,

$$H(\lambda, \alpha) := \{x \in X : \lambda x = \alpha\} = x_\alpha + \ker \lambda$$

for any $x_\alpha \in H(\lambda, \alpha)$. Hence, the minimum distance between $H(\lambda, \alpha)$ and $H(\lambda, \beta)$ is proportional to $|\alpha - \beta|$, and the proportionality factor is necessarily $\|\lambda\|$, since this is certainly so when $\alpha = \|\lambda\|$ and $\beta = 0$.

Indeed, in this particular case, $\alpha = \|\lambda\| = \sup \lambda(B)$, with

$$B := \{x \in X : \|x\| < 1\}$$

the (open) unit ball of X. Consequently,

$$H := H(\lambda, \|\lambda\|)$$

necessarily **touches** or **supports** B, meaning that (i) B lies entirely to one side of it (the negative side in this particular case), and (ii) $\inf \|B - H\| = 0$. This says that

$$\inf \|H(\lambda, \|\lambda\|) - \ker \lambda\| = 1 = (\|\lambda\| - 0)/\|\lambda\|,$$

as claimed.

This is also verified by the following formula (which itself follows from it by scaling):

$$|\lambda x| = \|\lambda\| \operatorname{dist}(x, \ker \lambda), \quad \lambda \in X^*, \ x \in X,$$

which we already made use of.

Draw the picture of the unit ball B, of $\ker \lambda$, and of the hyperplane $H := H(\lambda, \|\lambda\|)$, and note that H must lie on the positive side of $\ker \lambda$. Further, if there is $x \in H \cap B^-$, then $\|\lambda\| \operatorname{dist}(x, \ker \lambda) = \lambda x = \|\lambda\| = \|\lambda\|\|x\| \neq 0$. This shows that

$$0 \neq \lambda x = \|\lambda\|\|x\|. \tag{19}$$

If now X were just 2-dim. Euclidean space, $\ell_2(2)$, then λ would be given by the scalar product with some 2-vector x_λ, and (19) would imply that this 2-vector is a positive scalar multiple of x. Because of this familiar simple case, one says, even for any arbitrary nls X, that λ **and** x **are parallel** and writes

$$\lambda \| x$$

in case (19) holds.

More generally, if $m \in H(\lambda, \sup \lambda(B_r(f))) \cap B_r^-(f)$ for some $f \in X$ and some $r > 0$, then $\lambda \| (m-f)$. For, translation by f will shift f to 0 and m to $m-f$, and $\sup \lambda(B_r(f))$ to $\sup \lambda(B_r) = \|\lambda\| r$, while $\|m-f\| = r$. On replacing here sup by inf, we get, for arbitrary $m \neq f$,

$$(20) \qquad m \in H(\lambda, \inf \lambda(B_r(f))) \cap B_r^-(f) \implies \lambda \| (f-m).$$

However, in general, the equations

$$(21) \qquad ?\|x, \quad \|?\| = 1$$

may have many solutions in X^*. For example, if $x = (1,1) \in \ell_\infty$, then $0 < \lambda x = \|\lambda\|\|x\|$ for any λ of the form $f \mapsto \alpha f(1) + \beta f(2)$ with $\alpha, \beta \geq 0$ and, certainly, these are not all scalar multiples of each other. Note that the solution set for (21) is convex, hence the norm of X^* fails to be strictly convex if (21) has more than one solution for some x. In any case, the geometric imagery suggested by the statement 'λ is parallel to x' needs to be taken with a grain of salt. A point $x \in \partial B$ for which (21) has a unique solution is called a **smooth** point. If all points in ∂B are smooth, then the norm or the nls X itself are called 'smooth'. L_p and ℓ_p are smooth for $1 < p < \infty$ and fail to be smooth for $p = 1, \infty$.

If $\lambda \| x$, then λ takes on its norm at x since then there is equality in the standard inequality

$$|\lambda x| \leq \|\lambda\|\|x\|.$$

For this reason, an x satisfying (19) is also called an **extremal for** λ. (Note that, in general, λ could have many extremals, even if we restricted attention to points of norm 1, unless the norm, i.e., the closed unit ball B^-, is strictly convex.)

However, as already noted, a continuous linear functional need not take on its norm. Hence, while we may draw $H(\lambda, \|\lambda\|)$, this hyperplane may have no point in common with ∂B.

(22) Characterization Theorem for ba from convex set. *Let M be closed, convex in the real nls X, let $f \in X \backslash M$ and $m \in M$. Then, $m \in \mathcal{P}_M(f)$ iff there is some $\lambda \| (f-m)$ with* $\sup \lambda(M) \leq \lambda m$.

Proof: Draw the closed convex set, M, a point f outside M, and the (open) ball with center f and largest possible radius which fails to intersect M, i.e., the ball $B_r(f)$ with

$$r := \operatorname{dist}(f, M).$$

Note that $r > 0$ since $f \notin M^- = M$. Choose any $\lambda \in X^* \backslash 0$ and draw $H_+ := H(\lambda, \sup \lambda(M))$. Note that H_+ supports M.

Also draw the hyperplane $H_- := H(\lambda, \inf \lambda(B_r(f)))$. The latter is always on the negative side of the former, since $\inf \|B_r(f) - M\| = 0$ hence

$$\inf \lambda(B_r(f)) \le \sup \lambda(M).$$

Moreover, we have equality here exactly when the corresponding two hyperplanes, H_+ and H_-, coincide, i.e., exactly when they simultaneously support both M and $B_r(f)$, with M on the negative side and $B_r(f)$ on the positive side. In other words, such a hyperplane **separates** M and $B_r(f)$, and the Separation Theorem ensures the existence of such a hyperplane, i.e., of $\lambda \in X^* \backslash 0$ with

$$\sup \lambda(M) = \inf \lambda(B_r(f)).$$

Let now λ be so chosen. If $m \in \mathcal{P}_M(f)$, then, necessarily $m \in \partial M \cap \partial B_r(f)$, hence also $m \in H_+ = H_- = H(\lambda, \inf \lambda(B_r(f))$, and this, by (20), implies that $\lambda \| (f - m)$ *for any such* λ.

Conversely, if $m \in M$ and $\lambda \in X^*$ are such that $\lambda \| (f - m)$ while $\lambda m = \sup \lambda(M)$, then the hyperplane $H := H(\lambda, \lambda m)$ supports the ball $B_{\|f-m\|}(f)$ on its positive side while, simultaneously, it supports M on its negative side. In particular,

$$\|f - m\| = \operatorname{dist}(f, H) \le \operatorname{dist}(f, M) \le \|f - m\|,$$

showing that $\|f - m\| = \operatorname{dist}(f, M)$, therefore $m \in \mathcal{P}_M(f)$. □

Since $\operatorname{dist}(f, M) = r = (\lambda f - \inf \lambda(B_r(f))/\|\lambda\|$, while $\inf \lambda(B_r(f)) \le \sup \lambda(M)$ with equality possible, it follows that

$$\operatorname{dist}(f, M) = \max_{\lambda \ne 0} (\lambda f - \sup \lambda(M))/\|\lambda\|, \tag{23}$$

which provides a useful and sharp *lower bound* for the distance of f from the convex set M.

The theorem applies, in particular, to best approximation from a closed linear subspace M of a nls X. However, for any $\lambda \in X^* \backslash 0$, $\sup \lambda(M)$ on a lss M can only take on the values 0 and ∞. For the λ in our theorem, this leaves only the value 0 (since $\sup \lambda(M)$ must equal λm). Thus the condition $\sup \lambda(M) = \lambda m$ is replaced by the condition

$$\lambda \perp M := \ker \lambda \supset M.$$

(24) Characterization Theorem for ba from lss. *Let M be a linear subspace of the nls X, let $f \in X$ and $m \in M$. Then, $m \in \mathcal{P}_M(f)$ iff there is some $\lambda \| (f - m)$ with $\lambda \perp M$.*

Since $\sup \lambda(M) \in \{0, \infty\}$ in case M is a linear subspace, the lower bound (23) simplifies in that case to

$$\operatorname{dist}(f, M) = \max_{\lambda \perp M} \lambda f / \|\lambda\|. \tag{25}$$

Construction of ba

Characterization theorems are used in the construction of ba's.

If the norm in question is smooth, at least at the point $(f-m)/\|f-m\|$, then (by the very definition of 'smooth'), the condition $\lambda\|(f-m)$ determines λ uniquely, up to scalar multiples. On the other hand, the condition $\lambda \perp M$ holds iff it hold for any nontrivial scalar multiple, hence one may without loss restrict λ in the characterization theorem to be of norm 1. Hence, in smooth norms and for a finite-dimensional M, the characterizing conditions

$$M \perp \lambda \parallel (f-m)$$

constitute finitely many equations in the coefficients of the sought-for ba, m, with respect to some convenient basis for M, i.e., in as many unknowns as there are equations.

The classical example (from which the entire geometric language used here derives) is **least-squares approximation**, i.e., best approximation in an inner product space (such as ℓ_2 or L_2), with inner product $\langle\cdot,\cdot\rangle$. In such a space, the statement

$$0 < \lambda f = \|\lambda\|\|f\| \tag{26}$$

implies that $\lambda = c\langle\cdot, f\rangle$ for some positive constant c. The characterization theorem therefore specializes to the familiar

(27) Proposition. *The element m of the closed linear subspace M of the inner product space X is a ba to a given $f \in X$ iff $f - m \perp M$.*

The condition $f - m \perp M$, when expressed in terms of a basis $V = [v_1, \ldots, v_n]$ for M, becomes the so-called **normal equations**,

$$\sum_j \langle v_j, v_i\rangle a(j) = \langle f, v_i\rangle, \quad i = 1, \ldots, n,$$

a *linear* system in the coefficient vector a for the ba $m = Va$ with respect to that basis. A careless choice of the basis V may lead to numerical disaster (as would be the case if, e.g., $X = L_2[100\,..\,101]$, $M = \Pi_k$, and one were to choose the power basis, $[()^j : j = 0, \ldots, k]$). However, if M is a spline space, then it is usually acceptable to choose for V the B-spline basis for that space.

If V is chosen to be orthogonal, i.e., $\langle v_j, v_i\rangle = 0$ iff $i \neq j$, then the normal system becomes diagonal, and the best approximation is given by

$$m = \sum_j v_j \frac{\langle f, v_j\rangle}{\langle v_j, v_j\rangle}.$$

In L_p for $p \neq 2$ but $1 < p < \infty$, the statement (26) still determines a unique λ of norm 1, given in the form $\langle\cdot, g_\lambda\rangle$ for some g_λ in the dual space, L_{p^*}, i.e., with $1/p + 1/p^* = 1$. But g_λ depends *nonlinearly* on f, and this makes the resulting 'normal equations' harder to solve. One successful technique consists in converting the problem into a **weighted** L_2-problem, with the weight determined iteratively. Specifically, if $X = L_p[0\,..\,1]$, then

$$\|f-m\|_p^p = \int_0^1 |f-m|^2 w_{f,m},$$

with the weight function $w_{f,m} := |f-m|^{p-2}$.

Such techniques even work for $p=\infty$ if one uses the Pólya Algorithm, which obtains a ba in L_∞ as the limit of L_p-approximations as $p\to\infty$. Since ba's in L_∞ need not be unique, this procedure is also a way to *select* a particular ba from among the possibly many.

In L_∞, not only are ba's in general not unique, also (26) may have many (linearly independent) solutions, and this makes the application of the characterization theorem a bit harder. On the other hand, if M is finite-dimensional and $X = C(T)$, then it is possible to restrict considerably the set of linear functionals λ to be considered, namely to those which are linear combinations of no more than $(\dim M)+1$ point evaluations. This, and its pleasant consequences, are detailed in the next section.

The following observation is of use when the Characterization Theorem 24 is used in a nls with a non-smooth and/or non-strictly convex norm (such as L_∞), in which case there may be many linear functionals of norm 1 parallel to the error and/or many ba's.

(28) Lemma. *X nls, M lss, $f\in X$, $m\in M$. If $M\perp\lambda\|(f-m)$, then, for any $m'\in\mathcal{P}_M(f)$, also $\lambda\|(f-m')$.*

Proof: Assume WLOG that $\|\lambda\|=1$. Then, $\operatorname{dist}(f,M) = \|f-m\| = \lambda(f-m) = \lambda(f-m') \le \|\lambda\|\|f-m'\| = \operatorname{dist}(f,M)$. Hence, $\lambda\|(f-m')$. □

In short, any **characterizing** linear functional, i.e., any linear functional perpendicular to the approximating space and parallel to the error in some ba to the given f, must also be parallel to the error in any other ba to that f. At times, this leads to a proof of uniqueness of the ba.

Best approximation in $C(T)$

The tool here is the following representation theorem for linear functionals on finite-dimensional linear subspaces of $C(T)$.

(29) Theorem. *If λ is a linear functional on a lss Y, of dimension n, of the real nls $C(T)$, with T compact Hausdorff, then there exist $U\subset T$ with $\#U = n$ and $w\in\mathbb{R}^U$ so that $\lambda = \sum_{u\in U} w(u)\delta_u|_Y$ and $\|\lambda\| = \|w\|_1$.*

In other words, every linear functional on an n-dimensional subspace of the real $C(T)$ has a norm-preserving extension to all of $C(T)$ in the form

$$\lambda_{U,w} := \sum_{u\in U} w(u)\delta_u$$

of a linear combination of no more than n point evaluations.

This representation theorem is germane because our characterization of the elements m of $\mathcal{P}_M(f)$ involves linear functionals only as they act on $Y := M + \operatorname{ran}[f]$. Indeed, all the characterization demands is that $M \perp \lambda \| (f - m)$. Hence, whatever $\lambda \in C(T)^*$ the characterization theorem might have dragged in here, we may replace it by the extension of $\lambda|_Y$ to $C(T)$ guaranteed by Theorem 29. Since the original λ took on its norm on $f-m$, its restriction to Y also takes on its norm on $f - m$, hence the replacement functional has the same norm as the original one. This gives the following.

(30) Characterization Theorem for ba in $C(T)$. *Let $X = C(T)$ with $\mathbb{F} = \mathbb{R}$ and T compact Hausdorff, let M be an n-dimensional lss of X, let $f \in X$ and $m \in M$. Then, $m \in \mathcal{P}_M(f)$ iff there exists $U \subset T$ with $\#U \le n+1$ and $w \in \mathbb{R}^U$ so that $M \perp \lambda_{U,w} \,\|\, f-m$.*

Before exploiting this theorem for the construction of such ba's, here is a useful aspect of such linear functionals $\lambda_{U,w}$.

(31) de LaValleé-Poissin lower bound. *If $0 \ne \lambda_{U,w} \perp M$ and $m \in M$ with $\operatorname{signum}((f - m)(u)) = \operatorname{signum}(w(u))$, all $u \in U$, then*

$$\min |(f - m)(U)| \le \operatorname{dist}(f, M) \le \|f - m\|.$$

Proof: Set $\lambda := \lambda_{U,w}$ and recall, e.g. from (25), that $\lambda \perp M$ implies $|\lambda f| \le \|\lambda\| \operatorname{dist}(f, M)$, while trivially $\lambda(f) = \lambda(f - m)$. Hence, with $e := f - m$, we have

$$\min_{u \in U} |e(u)| \|\lambda\| \le \sum_u |e(u)||w(u)| = |\lambda e| = |\lambda f| \le \|\lambda\| \operatorname{dist}(f, M),$$

and division by $\|\lambda\| = \|w\|_1$ does the rest. □

This lower bound is the more effective, of course, the closer $\min |(f - m)(U)|$ is to $\|f - m\|$. The process of constructing such U for given m and, further, such m for given U is formalized in the Remes Algorithm below. For it, we now discuss related consequences of the characterization theorem 30.

The fact that the linear functional $\lambda := \lambda_{U,w}$ in the theorem is to take on its norm on the error, $e := f - m$, forces equality in the following string of inequalities:

$$\lambda e = \sum_u w(u)e(u) \le \sum_u |w(u)e(u)| \le \|w\|_1 \max_{u \in U} |e(u)| \le \|\lambda\| \|e\|.$$

In particular, assuming as we may WLOG that none of the $w(u)$ is zero, this implies that

$$e(u) = \|e\| \operatorname{signum} w(u), \quad u \in U.$$

This says that the error must take on its norm at every point in U, with the sign determined by the signature of the corresponding weight. On the other hand, these weights are not arbitrary. Rather, they are determined by the condition that $M \perp \lambda$. Since M is n-dimensional, the statement $M \perp \lambda$ is, in effect, a homogeneous linear system of n equations, namely the linear system

$$\sum_{u \in U} v_j(u) \, ?(u) = 0, \quad j = 1, \ldots, n, \tag{32}$$

with $V := [v_1, \ldots, v_n]$ any basis for M, and with the weight vector w the solution sought. In particular, there are nontrivial solutions for any choice of U with $\#U = n + 1$ (since then this homogeneous system has more unknowns than equations).

The theorem suggests the following *numerical procedure* (associated with the name **Remes**).

Remes Algorithm

(i) *Pick any $(n+1)$-set U in T.*

Usually, one picks U as the set at which some approximation $m \in M$ to the given f has its absolutely largest local extrema.

(ii) *Compute a best approximation m_U to f from M in the discrete norm*

$$\|\cdot\|_U : g \mapsto \max |g(U)|.$$

This is often possible by solving the *linear* system

$$\sum_j v_j(u)a(j) + \sigma(u)r = f(u), \quad u \in U \tag{33}$$

for the coefficient vector a and the $\|\cdot\|_U$-distance r of f from M, with $\sigma = \text{signum}(w)$ for some nontrivial solution of the linear system (32). If a is a solution of (33), then we *know* from the characterization theorem that the resulting $m_U := Va$ is, indeed, a ba to f from M in the norm $\|\cdot\|_U$, with $\|f - m_U\|_U = |r|$. (If r happens to be negative, simply replace w by $-w$, as that won't change the fact that it is a solution of (33) but will change r to $|r|$.)

(iii) *Update U, making it the set of $n+1$ points at which $f - m_U$ has its absolutely largest local extrema, and go to (ii).*

The algorithm is quite attractive since, at the end of step (iii), we know (as in (31)) that

$$|r| \le \text{dist}(f, M) \le \|f - m_U\|,$$

hence can gauge whether or not it is worth our while to continue. Indeed, the second inequality is obvious; the first follows from the fact that $|r|$ is the error in the ba to f from M when we only consider the maximum absolute error on the subset U of T.

The only fly in the ointment is uncertainty about the solvability of (33).

The Haar condition

Since (33) is a square system, its solvability is equivalent to its unique solvability. Hence, as f is arbitrary, we are asking that the pointset U be total for M, i.e., that $m \in M$ and $m|_U = 0$ imply $m = 0$. But we are asking more, we are asking for the solvability of (33). This means that the matrix

$$W := [v_1|_U, \ldots, v_n|_U, \sigma]$$

be 1-1.

(34) Proposition. *If some subset of the $n+1$-set $U \subset T$ is total for the n-dimensional lss M of $C(T)$, then (33) has exactly one solution (for every $f \in C(T)$).*

Proof: Let w be a nontrivial solution of the homogeneous linear system (32), hence $\sigma = \mathrm{signum}(w)$, and let $W(a,r) = 0$, with a an n-vector and r a scalar. Then $(Va)(u) = -r\,\mathrm{signum}(W(u))$, all $u \in U$, hence $0 = \sum_{u\in U} w(u)(Va)(u) = -r\|w\|_1$, which is only possible if $r = 0$. However, now $Va = 0$ on U follows, which by the assumption that some subset of U is total for M, implies that $a = 0$. □

Now, U being total for M is the same as the linear map

$$M \to \mathbb{R}^U : m \mapsto m|_U$$

being 1-1. Hence, if U is total for an n-dimensional lss M, then U must contain at least n points. Further, if U contains exactly n points, then this restriction map is also onto. This says that M contains, for each g defined at least on U, exactly one m which **interpolates** f **at** U, i.e., which agrees with f on U.

(35) Definition. *The n-dimensional linear space M of functions on some domain T is a* **Haar space** $:\Longleftrightarrow$ *every n-set U in T is total for M.*

(36) Theorem (Haar). *If M is a finite-dimensional lss of $C(T)$ (with T compact Hausdorff), then M is Haar iff M is Chebyshev.*

Proof: '$\Longrightarrow$': Let $f \in C(T)$, and let $m \in \mathcal{P}_M(f)$ (such m exists since M is a finite-dimensional lss). By the Characterization Theorem 30, there exists $U \subset T$ with $\#U \le n+1$ and $w \in \mathbb{R}^U$ so that

$$M \perp \lambda_{U,w} \| (f - m).$$

Now let $m' \in \mathcal{P}_M(f)$. By Lemma 28, we must have

$$\sum_{u\in U} w(u)(f-m')(u) = \lambda_{U,w}(f-m') = \|w\|_1\|f-m'\|$$
$$\ge \sum_{u\in U} |w(u)(f-m')(u)| \ge |\sum_{u\in U} w(u)(f-m')(u)|,$$

which is possible only if

$$(f - m')(u) = \operatorname{signum}(w(u)) \operatorname{dist}(f, M), \quad u \in U.$$

By assumption, M is Haar, hence U must contain at least $n+1$ points, and none of the entries of w is zero. This means that m' must agree with m on all of U and, as U is total for M, this implies $m' = m$.

'$\Longleftarrow$': If M fails to be Haar, then there exists an n-set U in T and some $m \in M$ with $\|m\| = 1$ which vanishes on U. Further, there exists $w \in \mathbb{R}^U \backslash 0$ with $\|w\|_1 = 1$ so that

$\lambda := \lambda_{U,w} \perp M$. It follows that, by the characterization theorem, any $g \in C(T)$ of norm ≤ 1 with $g = \mathrm{signum}(w)$ on U (hence, actually, $\|g\| = 1$) has 0 as a best approximation from M, and, by the Tietze Extension Theorem, there exists at least one such. Further, since $\|m\| = 1$ and $m|_U = 0$, the function $f := (1 - |m|)g$ is one such. But then, so is any $f - \alpha m$ with $\alpha \in [-1\,..\,1]$ since (i) it agrees with g on U, hence on $\lambda_{U,w}$, and (ii)

$$|f(t) - \alpha m(t)| \le |f(t)| + |\alpha||m(t)| \le 1 - |m(t)| \;+\; |m(t)|,$$

hence $\|f - \alpha m\| \le 1$. In particular, 0 is a ba to $f - \alpha m$ from M, therefore $\alpha m \in \mathcal{P}_M(f)$. Thus, $[-1\,..\,1]m \subset \mathcal{P}_M(f)$. □

Note that $\Pi_k(\mathbb{R})$ is Haar, hence we now know that it is also Chebyshev.

The proof took for granted the following

(37) Lemma. *If the n-dimensional linear subspace M of the $C(T)$ is Haar, then $0 \ne \lambda_{U,w} \perp M$ implies $\#U > n$.*

Proof: Let $V = [v_1, \ldots, v_n]$ be any basis for M, order the elements of U: $U = (u_1, \ldots, u_r)$, say, and set $w_j := w(u_j)$, all j. Then (w_j) solves the homogeneous linear system

$$\begin{bmatrix} v_1 & \cdots & v_n \\ u_1 & \cdots & u_r \end{bmatrix} ? = 0,$$

whose coefficient matrix is the $n \times r$-matrix whose (i,j)-entry is the number $v_i(u_j)$. If $r \le n$, then, since M is Haar, this matrix could be extended to an invertible matrix, hence must itself be 1-1, hence $w = 0$ would follow, a contradiction to the assumption that $\lambda_{U,w} \ne 0$. □

Since the characterizing $\lambda_{U,w}$ can be chosen with $\#U \le \dim M + 1$, it follows that, for a Chebyshev space M, $\#U$ must be equal to $\dim M + 1$, meaning in particular that all the $n+1$ entries of the weight vector w must be nonzero. Further, once we know this, we don't really care about the weight vector itself anymore, all we need for checking a proposed ba $m \in M$ is the sign vector, $\sigma := \mathrm{signum}(w)$, since, as we saw earlier, the characterizing condition merely demands that the error, $e := f - m$, satisfy

$$e(u)\varepsilon = \|e\|\sigma(u), \quad u \in U, \tag{38}$$

with ε some fixed (nontrivial) sign. We now investigate the possible sign vectors σ.

Let $V = [v_1, \ldots, v_n]$ be a basis for M, and order the points in U somehow, $U = (u_1, \ldots, u_{n+1})$. Then, for the specific normalization $w_{n+1} = -1$, the n-vector $(w_j : j = 1, \ldots, n)$ is the unique solution of the linear system

$$\begin{bmatrix} v_1 & \cdots & v_n \\ u_1 & \cdots & u_n \end{bmatrix} ? = \begin{bmatrix} v_1 & \cdots & v_n \\ & u_{n+1} & \end{bmatrix}.$$

By Cramer's rule, this implies that

$$w_j = \det \begin{bmatrix} v_1 & \cdots & v_n \\ u_1 & \ldots u_{j-1}, u_{n+1}, u_{j+1} \ldots & u_n \end{bmatrix} \Big/ \det \begin{bmatrix} v_1 & \cdots & v_n \\ u_1 & \cdots & u_n \end{bmatrix},$$

or, using column interchanges to restore order here,

$$(39) \qquad w_j = (-1)^{n-j} \det \begin{bmatrix} v_1 & \cdots & v_n \\ u_1 & \ldots u_{j-1}, u_{j+1} \ldots & u_{n+1} \end{bmatrix} / \det \begin{bmatrix} v_1 & \cdots & v_n \\ u_1 & \ldots & u_n \end{bmatrix}.$$

If now T is an interval or, more generally, 'interval-like', i.e., a connected totally ordered set, then it makes sense to choose the ordering of the points in U accordingly, i.e., to choose $u_1 < \cdots < u_{n+1}$. Moreover, each of the ordered sequences

$$(u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_{n+1})$$

can be connected to the sequence $(u_1, \ldots, u_n)$ by a *continuous* transformation $[0\,..\,1] : t \mapsto \tau(t) := (\tau_1(t) < \cdots < \tau_n(t))$, with $\tau(1)$ the former and $\tau(0)$ the latter. It follows that the corresponding determinants $\det \begin{bmatrix} v_1 & \cdots & v_n \\ \tau_1(t) & \ldots & \tau_n(t) \end{bmatrix}$ depend continuously on t and none can be zero since M is Haar. Hence, if the v_j are all real, then *all the determinants in (39) have the same (positive or negative) sign.* In particular, in this case $w_j w_{j+1} < 0$, all j. This gives

(40) Chebyshev's Alternation Theorem. *If M is an n-dimensional Chebyshev subspace of the real $C([a\,..\,b])$, then $m \in \mathcal{P}_M(f)$ iff the error, $f - m$,* **alternates at least** n **times***, i.e., iff there are points $u_1 < \cdots < u_{n+1}$ in $[a\,..\,b]$ and an $\varepsilon \in \{-1, 1\}$ so that*

$$(f - m)(u_j)\varepsilon \;=\; (-1)^j \|f - m\|, \quad j = 1, \ldots, n+1.$$

The argument leading up to this classical theorem brings with it a somber consequence, called by some frivolous people the 'Loss of Haar'. As soon as T 'contains a fork', i.e., contains three open arcs which have exactly one point in common, then no linear subspace of dimension > 1 can be Haar. For, in such a setting, we can arrange the continuous map $[0\,..\,1]t \mapsto (\tau_1(t), \ldots, \tau_n(t))$, from the unit interval to n-sequences with n *distinct* entries in T, in such a way that

$$\tau(0) = (r, s, u_3, \ldots, u_n), \quad \tau(1) = (s, r, u_3, \ldots, u_n).$$

This implies that the determinant corresponding to $\tau(0)$ is the negative of the determinant corresponding to $\tau(1)$. Hence, if the v_j are real, then the determinant must be zero for some $\tau(t)$, and the corresponding n-set fails to be total for M. We have proved

(41) Loss of Haar. *If M is a finite-dimensional real linear subspace of dimension > 1 of $C(T)$ (with T compact Hausdorff), and T 'contains a fork', then M cannot be Chebyshev (since it cannot be Haar).*

In particular, no polynomial space of dimension > 1 in more than one variable can be Haar. This has made the construction of uniform best approximations to functions of several arguments, even by polynomials of low degree, something of an art. Correspondingly, it has encouraged the development of near-best methods of approximation.

(42) Theorem (Newman-Shapiro). *Best approximation from a finite-dimensional linear subspace M of the real $X := C(T)$ is strongly unique. Explicitly, for every $f \in X$, there exists $\gamma > 0$ so that*

$$\forall\, g \in M \quad \|f - g\| \geq \|f - P_M f\| + \gamma \|g - P_M f\|. \tag{43}$$

Proof: For $f \in M$, $\gamma = 1$ will do. For $f \in X \backslash M$, let $M \perp \lambda \| f - P_M f$, with $\lambda = \lambda_{U,w}$ for some $U \subset T$ with $\#U \leq n+1$. Set $\sigma := \mathrm{signum}(w)$. For any $g \in M$ and any $u \in U$,

$$\|f - g\| \geq \sigma(u)(f-g)(u) = \sigma(u)(f - P_M f)(u) + \sigma(u)(P_M f - g)(u),$$

hence

$$\|f - g\| \geq \|f - P_M f\| + K_\sigma(P_M f - g),$$

with

$$K_\sigma(g) := \max_{u \in U} \sigma(u) g(u).$$

Since K_σ is positive homogeneous, (43) follows with

$$\gamma := \inf\{K_\sigma(m) : m \in M, \|m\| = 1\}.$$

It remains to show that $\gamma > 0$. Since K_σ is continuous and M is finite-dimensional, can find $g \in M$ with $\|g\| = 1$ and $\gamma = K_\sigma(g)$. Since $0 = \lambda_{U,w} g = \sum_{u \in U} w(u) g(u) = \sum_{u \in U} |w(u)| \sigma(u) g(u)$, $K_\sigma(g) = 0$ would imply $g(u) = 0$ for all $u \in U$. However, M is Chebyshev, hence Haar, hence, by Corollary 37, U is total for M, and now $g = 0$ would follow, contradicting the assumption that $\|g\| = 1$. □

Complex $C(T)$

We pointed out earlier that the general characterization theorem for ba's involves linear functionals parallel to the error $f - m$ because they provide the gradient of the map $g \mapsto \|f - g\|$ at the point m. Kolmogorov's characterization theorem is more explicitly based on this idea of a directional derivative of the map $g \mapsto \|f - g\|$. The criterion is of interest because it is also valid when $\mathbb{F} = \mathbb{C}$. It formalizes the following idea: if, in the max-norm, we want $\|f + g\| < \|f\|$, then, at all points t at which f takes on its norm, $f(t) + g(t)$ needs to be less than $|f(t)|$ in absolute value. Hence, even f and g are complex-valued, we need that $\mathrm{Re}\, f(t)\overline{g(t)} < 0$. Moreover, this must be so uniformly over the set

$$\mathrm{extr}(f) := \{t \in T : |f(t)| = \|f\|\} = \mathrm{argmax}\, |f(T)|.$$

(44) Lemma. *Let $X = C(T)$ with T compact Hausdorff, and $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $f, g \in X \backslash 0$. Then, $\|f + \alpha g\| < \|f\|$ for some positive α iff*

$$K_f(g) := \max_{t \in \mathrm{extr}(f)} \mathrm{Re}(f(t)\overline{g(t)}) < 0.$$

Proof: $\|f+g\|^2 \ge \|f+g\|^2(\mathrm{extr}(f)) \ge \|f\|^2 + 2K_f(g)$, hence, as K_f is real homogeneous, $\|f+\alpha g\| < \|f\|$ for some $\alpha > 0$ implies $K_f(g) < 0$.

Conversely, if $K_f(g) < 0$, then

$$G := \{t \in T : \mathrm{Re}(f(t)\overline{g(t)}) < K_f(g)/2\}$$

is open and contains $\mathrm{extr}(f)$ since, for $t \in \mathrm{extr}(f)$, we even have $\mathrm{Re}(f(t)\overline{g(t)}) \le K_f(g)$. Therefore, $\|f\|(T\backslash G) < \|f\|$, and this implies that

$$\|f+\alpha g\|(T\backslash G) \le \|f\|(T\backslash G) + |\alpha|\|g\|(T\backslash G)$$

is less than $\|f\|$ for all α close to 0. But, by the strict negativity of $K_f(g)$, also

$$\|f+\alpha g\|^2(G) \le \|f\|^2 + 2\alpha K_f(g)/2 + |\alpha|^2\|g\|^2$$

is less than $\|f\|^2$ for all *positive* α close to 0. So, altogether, $\|f+\alpha g\| < \|f\|$ for all positive α near 0. □

Since $\|f-(m-g)\| = \|(f-m) + g\|$, this lemma has the following very useful consequence.

(45) Kolmogorov Criterion. *Let $X = C(T)$ with T compact Hausdorff and $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and let M be a lss of X, let $f \in X$ and $m \in M$. Then, $m \in \mathcal{P}_M(f)$ if and only if $\forall g \in M$, $K_{f-m}(g) \ge 0$.*

Here is **Alper**'s example: $T = \{z \in \mathbb{C} : |z| \le 1\}$ is the unit disc in the complex plane; we take $M = \Pi_k$, and $f = 1/(\cdot - \alpha)$, with $\alpha \in \mathbb{C}\backslash T$. Consider the function

$$m : z \mapsto \frac{1}{z-\alpha} - cz^k\frac{\overline{\alpha}z - 1}{z-\alpha} = \frac{1 - cz^k\,(\overline{\alpha}z-1)}{z-\alpha}.$$

This is a polynomial (in z) iff the numerator of the last expression vanishes at $z = \alpha$, i.e., iff $c = \alpha^{-k}/(|\alpha|^2 - 1)$. With that choice, $m \in \Pi_k$, and

$$e := f - m = z^k\frac{\overline{\alpha}z-1}{z-\alpha},$$

hence $\mathrm{extr}\, e = \{z : |z| = 1\}$. Let $g \in \Pi_k$ and consider $K_e(g)$. Having it nonnegative requires that

$$\mathrm{Arg}(e(z)\overline{g(z)}) = \mathrm{Arg}(e(z)) - \mathrm{Arg}(g(z)) \in [-\pi/2\,..\,\pi/2]$$

for some $|z| = 1$. Now, $\mathrm{Arg}(e(z)) = \mathrm{const} + k\,\mathrm{Arg}(z) + \mathrm{Arg}(z - \overline{\alpha}^{-1}) - \mathrm{Arg}(z-\alpha)$, and this increases by $2\pi(k+1)$ as we circumnavigate the unit disc. On the other hand, g, being a polynomial of degree $\le k$, can have at most k zeros in the unit disc, hence $\mathrm{Arg}(g(z))$ can increase by at most $2\pi k$. And that does it. □

The continuous linear functionals on $X := L_1(T;\mu)$ are of the form $f \mapsto \int_T hf$, with $h \in L_\infty(T)$. Further,

$$\int_T hf \le \|h\|_\infty \|f\|_1,$$

with equality iff $h = \|h\|_\infty \operatorname{signum}(f)$ off the set

$$Z_f := \{t \in T : f(t) = 0\}.$$

(In particular, 'most' continuous linear functionals on $L_1(T)$ do not take on their norm, and, even on ℓ_1, some don't (cf. p.11).) In general, the **zero-set** Z_f is only determined up to sets of measure 0 since it depends, of course, on the particular representer of f's equivalence class used for its determination.

(46) Characterization Theorem. *Let M be a linear subspace of the nls $X = L_1(T;\mu)$, with μ a non-atomic measure, let $f \in X$ and $m \in M$. Then, $m \in \mathcal{P}_M(f)$ iff there exists a $h \in L_\infty(T)$ with $\|h\|_\infty = 1$ which is perpendicular to M and agrees with* $\operatorname{signum} f - m$ *off Z_{f-m}.*

In particular, if Z_{f-m} has measure zero, then $m \in \mathcal{P}_M(f)$ iff $\operatorname{signum}(f-m) \perp M$.

The theorem suggests a quick try at constructing ba's from an n-dimensional lss of $X = L_1([a\,..\,b])$: Assuming that Z_{f-m} has measure zero, we look for a **sign function**, i.e., a real function h with $|h| = 1$, the simpler the better, i.e., the fewer breakpoints the better. The condition $h \perp M$ is, in effect, a system of equations for the breakpoints of h, hence, in general, we would not expect to succeed with fewer than n (interior) breakpoints. Now suppose we have succeeded, and now have in hand such a sign function with exactly n interior breakpoints, $a < \xi_1 < \cdots < \xi_n < b$ say. If f and the elements of M are continuous (or, at least piecewise continuous), then we can now hope to *interpolate to f at ξ by elements of M*. Let m be the resulting interpolant. Then, 'usually', the error, $f - m$, changes sign at the points of interpolation, hence 'usually', $\operatorname{signum} f - m = h$, and we have found a ba to f from M.

The fact that, for every n-dimensional lss M of $L_1([a\,..\,b])$, there exists a sign function orthogonal to it with $\le n$ breakpoints is called the Hobby-Rice Theorem. Its original proof was quite involved. While a postdoc here in Madison, Allan Pinkus pointed out that it is a ready consequence of Borsuk's Theorem. Since the latter has played a major role in the discussion of n-width, here is its statement.

(47) Borsuk's Theorem. *Let g be a continuous map from the unit sphere S^n in $\mathbb{R}^{n+1}$ to $\mathbb{R}^n$. If g is **odd**, i.e., $g(-x) = -g(x)$, all $x \in S^n$, then $0 \in g(S^n)$.*

(48) Hobby-Rice Theorem. *For every n-dimensional lss M of $X = L_1([a\,..\,b])$, there exists a sign function orthogonal to it with $\le n$ breakpoints.*

Proof: Assume WLOG that $[a\,..\,b] = [0\,..\,1]$, and, for s in

$$S^n := \{(s_0, \ldots, s_n) \in \mathbb{R}^{n+1} : \sum_j s_j^2 = 1\},$$

define the linear functional λ_s on X as follows: With $\xi_j(s) := \sum_{i<j} s_i^2$,

$$\lambda_s f := \sum_{j=0}^{n} \operatorname{signum}(s_j) \int_{\xi_j(s)}^{\xi_{j+1}(s)} f \ .$$

Then $\lambda_s f = \int_0^1 h_s f$, with h_s a sign function with at most n breakpoints (since we may ignore $\xi_j(s)$ if $s_{j-1}s_j \ge 0$). Further, $\lambda_{-s} = -\lambda_s$ and, for fixed f, the map $S^n \to \mathbb{R} : s \mapsto \lambda_s f$ is continuous, since, for any $f \in L_1[0\,..\,1]$, the map $[0\,..\,1] \to \mathbb{F} : (u,v) \mapsto \int_u^v f$ is continuous. Hence, with $V = [v_1, \ldots, v_n]$ any basis for M, the map

$$g : S^n \to \mathbb{R}^n : s \mapsto (\lambda_s v_j : j = 1, \ldots, n)$$

is continuous and odd, therefore Borsuk tells us that $g(s) = 0$ for some $s \in S^n$. The corresponding sign function, h_s, provides what we are looking for. □

(49) Corollary (Krein). *No (nontrivial) finite-dimensional lss M of $X = L_1([a\,..\,b])$ is a Chebyshev space.*

Proof: Let h be a sign function orthogonal to M. Take any $g \in M\backslash 0$ and set $f := h|g|$. Then, for any $\alpha \in [0\,..\,1]$,

$$h(f - \alpha g) = |g| - \alpha h g \ge (1-\alpha)|g| \ge 0,$$

showing that $M \perp \int h \cdot \parallel (f - \alpha g)$, hence $[0\,..\,1]g \in \mathcal{P}_M(f)$. □

Of course, this does *not* imply that *every* $f \in X\backslash M$ has many ba's from M. E.g., while $\chi_{[0..1/2]}$ has many ba's from $\Pi_0 \subset L_1([0\,..\,1])$, $()^1$ has only 1 (as does, in fact, any $\chi_{[0..\alpha]}$ *except* when $\alpha = 1/2$).

Degree of Approximation

Given a *sequence* $(M_n) = (M_n : n = 0, 1, \ldots)$ of approximating sets in the nls X and $f \in X$, one is interested in

$$n \mapsto E_n(f) := \operatorname{dist}(f, M_n)$$

as $n \to \infty$.

In this generality, nothing can be said. The following general model (from Chapter 7 of DeVore and Lorentz) covers all situations of practical, and even most situations of theoretical, interest.

(i) *$M_0 = \{0\}$, and (M_n) is increasing.*
This guarantees that $E_n(f)$ is nonincreasing, with $E_0(f) = \|f\|$.
(ii) *$\cup_n M_n$ is dense in X.*
This is equivalent to having $\lim_n E_n(f) = 0$ for all $f \in X$.

In view of approximation by rationals, or exponentials with free frequencies, or splines with free knots, it would be too much to assume that the M_n are *linear* subspaces. But it is ok to assume

(iii) $\mathbb{F}M_n \subset M_n$.

(iv) $\exists\{c\}\forall\{n\}\ M_n + M_n \subset M_{cn}$.

Finally, although clearly not essential, the following assumption will avoid a certain amount of epsilontics.

(v) *Each M_n is an existence set.*

In particular, each M_n is closed.

With these assumptions in place, the best we can say about $E_n(f)$ for given f has already been said: $E_n(f)$ converges monotonely to 0. The question of interest is just how 'fast' or 'slow' this convergence is. As with all measuring, these terms are made precise by comparing $E_n(f)$ with certain *standard* sequences. The most popular of these are the sequences $(n^{-\alpha} : n \in \mathbb{N})$ for some real α. Thus we are looking for conditions on f which guarantee that

$$E_n(f) = O(n^{-r}),$$

i.e., $\limsup_n E_n(f)n^r < \infty$, or, perhaps, even

$$E_n(f) = o(n^{-r}),$$

i.e., $\lim_n E_n(f)n^r = 0$. Such conditions necessarily single out a rather 'small' subset of X in case X is complete but not equal to any of the M_n.

(50) Proposition. *If $M_n \neq X$, all n, then, for any real sequence α_n converging monotonely to 0, the set*

$$A_\alpha := \{f \in X : E_n(f) = O(\alpha_n)\}$$

is 'thin' in the sense that it is of first category, i.e., the countable union of nowhere dense sets. In particular, it cannot equal X in case X is complete.

As is customary in mathematics, we express suitable conditions on f as membership in some set Y. A typical choice for Y is a *semi*-normed ls whose semi-norm we denote by $\|\cdot\|_Y$ which is continuously imbedded in X. The standard example has $M_n = \mathring{\Pi}_n \subset X := X_p$, with

$$X_p := \begin{cases} L_p(\mathbb{T}), & p < \infty; \\ C(\mathbb{T}), & p = \infty, \end{cases}$$

and

$$Y = X_p^r := \{f \in X_p : \|f\|_Y := \|D^r f\|_p < \infty\},$$

with r some positive integer.

One hopes to choose Y and r in such a way that, simultaneously,

$$\exists\{C_J\}\forall\{n \in \mathbb{N},\ f \in Y\}\ \ E_n(f) \le C_J\, n^{-r}\|f\|_Y \tag{51}$$

and

$$\exists\{C_B\}\forall\{n \in \mathbb{N},\ m \in M_n\}\|m\|_Y \le C_B\, n^r\|m\|. \tag{52}$$

For historical reasons, the former is called a **Jackson**, or **direct**, inequality or estimate, the latter a **Bernshtein**, or **inverse**, inequality or estimate. The Jackson inequality gives a *lower* bound on the speed with which $E_n(f)$ goes to zero, and, as we shall see, the Bernshtein inequality provides an upper bound, at least indirectly. The most natural bridge between Jackson and Bernshtein inequalities turns out to be **Peetre's K-functional**:

$$K(f,t) := K(f,t;X,Y) := \inf_{g\in Y}(\|f-g\| + t\|g\|_Y).$$

This functional measures how 'smooth' f is in the sense that it tells us how closely we can approximate f by 'smooth' elements. The larger t, the more stress we lay on the smoothness of g. As a function of t, $K(f,t)$ is the infimum of a collection of straight lines, all with nonnegative slope, hence $K(f,.)$ is also weakly increasing (i.e., nondecreasing), and, further, is *concave* (as would be the infimum of any collection of straight lines increasing or not). This implies that $t \mapsto K(f,t)/t$ *is (weakly) decreasing* (i.e., nonincreasing).

We also need the following somewhat complicated weighted sequence norm

$$\|a\|^{(r)} := \sum_{k=1}^{n} |k^r a(k)|/k, \quad a \in \mathbb{F}^n.$$

This norm is **monotone**, meaning that $|a| \le |b|$ implies $\|a\|^{(r)} \le \|b\|^{(r)}$. Further, for the following 'standard' null-sequence $a = (a(k) = k^{-(r-1+\alpha)} : k \le n)$, one computes

$$\|a\|^{(r)} = \sum_{k=1}^{n} k^{-\alpha} \le \int_0^n ()^{-\alpha} = n^{1-\alpha}/(1-\alpha), \tag{53}$$

as long as $\alpha \in [0..1)$. Here is a quick comment concerning the fact that, in the description of $\|a\|^{(r)}$, I did not combine the two powers of k appearing there. The reason is the following. The argument to follow remains valid even when $\|\cdot\|$ and/or $\|\cdot\|_Y$ are only **quasi-norms**, meaning, in effect, that, instead of the triangle inequality, we only have

$$\|x+y\|^\mu \le \|x\|^\mu + \|y\|^\mu$$

for some μ and all x and y. In such a case, one considers, more generally, the weighted sequence norm

$$\|a\|_\mu^{(r)} := \left(\sum_{k=1}^{n} |k^r a(k)|^\mu/k\right)^{1/\mu}.$$

(54) Peetre's Theorem.
(i) (51) implies $\sup_n E_n(f)/K(f,n^{-r}) < \infty$.
(ii) (52) implies

$$K(f,n^{-r}) \le \mathrm{const}_r\, n^{-r}\|(E_k(f) : k \le n)\|^{(r)}. \tag{55}$$

In particular, we have $E_n(f) = O(n^{-(r-1+\alpha)})$ for some $\alpha \in [0\,..\,1)$ if and only if $K(f,n^{-r}) = O(n^{-(r-1+\alpha)})$. In the standard situation, the latter can be shown to be equivalent to having $f \in C^{(r-1)}$ with $D^{r-1}f$ in Lip_α.

Proof: (i): For any $g \in Y$, $E_n(f) \le \|f-g\| + E_n(g) \le C(\|f-g\| + n^{-r}\|g\|_Y)$, with $C := \max\{1, C_J\}$, and taking the inf over $g \in Y$ does it.

(ii): The trick here is the use of a telescoping series. (The proof in DeVore and Lorentz is a touch terse.) With $f_k \in \mathcal{P}_{M_k(f)}$, all k, and $\nu := \lceil \log_2 n \rceil$ the natural number for which $2^{\nu-1} < n \le 2^\nu$, we set

$$n_k := \begin{cases} 0 & k=-1; \\ 2^k & 0 \le k < \nu; \\ n & k = \nu, \end{cases}$$

and consider

$$h_k := f_{n_k} - f_{n_{k-1}} \in M_{cn_k}.$$

Then

$$\|h_k\| \le \|f - f_{n_k}\| + \|f - f_{n_{k-1}}\| \le 2E_{n_{k-1}}(f),$$

while $f_n = \sum_{k=0}^{\nu} h_k$. Therefore,

$$\begin{aligned} K(f, n^{-r}) &\le \|f - f_n\| + n^{-r}\|f_n\|_Y \\ &\le E_n(f) + n^{-r}\sum_{k=0}^{\nu}\|h_k\|_Y \\ &\le n^{-r}\big(n^r E_n(f) + C_B\, c^r\, 2\sum_{k=0}^{\nu} n_k^r E_{n_{k-1}}(f)\big) \\ &\le n^{-r}\mathrm{const}_r \sum_{k=0}^{\nu} n_k^r E_{n_k}(f), \end{aligned}$$

with $\mathrm{const}_r := 2C_B c^r\, 2^r$, the last factor to bound $\max_k (n_k/n_{k-1})^r$. Since $k \mapsto E_k(f)$ is monotone (decreasing), we have

$$\sum_{j<k\le J} k^{r-1}E_k(f) \ge E_J(f) \sum_{j<k\le J} k^{r-1},$$

while

$$\sum_{j<k\le J} k^{r-1} \ge \int_j^J ()^{r-1} = (1/r)(J^r - j^r) \ge \mathrm{const}_r J^r,$$

with $\mathrm{const}_r := (1-(1/2)^r)/r$ if, as is the case in our application, $J \ge 4$ and $j \ge J/2$. This implies (55), with const depending on C_B and r only. □

The last argument proved that

$$\sum_{k=0}^{\nu} n_k^r E_{n_k}(f) \le \mathrm{const}_r \sum_{j=n_0}^{n_\nu} j^{r-1}E_j(f). \tag{56}$$

However, since also

$$\sum_{j\le k<J} k^{r-1} \le \int_j^J ()^{r-1} = (1/r)(J^r - j^r),$$

we also get

$$\sum_{j=n_0}^{n_\nu} j^{r-1}E_j(f) \le \text{const}_r \sum_{k=0}^{\nu} n_k^r E_{n_k}(f)$$

with $\text{const}_r = 2^r - 1$. Thus, up to r-dependent constants, we haven't lost anything when going from the dyadic sum to the full sum.

Remark The argument just given is very free and easy with constants. Its only purpose is to establish statements about the degree of convergence, such as the assertion that $o(n^{-r}) \ne E_n(f) = O(n^{-r})$. Rather different arguments are used to establish statements of the sort that

$$E_n(f) \le \text{const}\, n^{-r}\|f\|_Y, \quad f \in Y,$$

with const the smallest possible constant independent of f and n.

The first such theorem was proved by Bernshtein, using

Bernshtein Inequality. *For any* $f \in \overset{\circ}{\Pi}_n \subset C(\mathbb{T})$,

$$\|Df\| \le n\|f\|.$$

Proof: Here is v.Golitschek's proof of Szegö's stronger inequality

$$(Df)^2 + (nf)^2 \le (n\|f\|)^2, \quad f \in \overset{\circ}{\Pi}_n. \tag{57}$$

As the inequality is homogeneous in f, we may assume WLOG that $\|f\| < 1$. (This assumption is a convenience, to avoid discussion of double zeros.) Further, since all points in $\mathbb{T}$ are alike, it is sufficient to prove that

$$(Df(0))^2 + (nf(0))^2 \le (n\|f\|)^2, \quad f \in \overset{\circ}{\Pi}_n. \tag{58}$$

We may even assume WLOG that $Df(0) \ge 0$. Now let α be the unique point in $(-\pi\,..\,\pi)/(2n)$ at which $\sin n\alpha = f(0)$, and consider the trig.pol.

$$s := \sin n(\cdot + \alpha) - f.$$

Since $\|f\| < 1$, s alternates in sign at the extrema of $\sin n(\cdot+\alpha)$ and, as there are $2n$ such, s has exactly one zero between any two adjacent extrema of $\sin n(\cdot+\alpha)$. In particular, one of these zeros must be the point 0 since we chose α to make it so. If now $Ds(0) < 0$, then f would rise above $\sin n(\cdot+\alpha)$ as we move to the right from 0, yet is certain to be below it again by the time we reach the first extremum of $\sin n(\cdot+\alpha)$ to the right of 0 which is a maximum, and this would imply an impossible second zero of s between the two extrema which bracket 0. Consequently,

$$0 \le Df(0) = n\cos n\alpha - Ds(0) \le n\cos n\alpha = n\sqrt{1-(f(0))^2},$$

using the fact that, by our choice of α, $\sin n\alpha = f(0)$. Squaring and reordering terms gives

$$(Df(0))^2 + (nf(0))^2 \le n^2,$$

which isn't quite what we hoped for. However, for any $r > \|f\|$, we also have $\|f/r\| < 1$, and therefore $(Df(0))^2 + (nf(0))^2 \le (nr)^2$ for all such r, and taking now the infinum over all such r gives (58). $\square$

We are ready to prove a sample inverse theorem.

(59) Theorem. *Let $M_n = \mathring{\Pi}_n$, as a subset of $C(\mathbb{T})$. Let $f \in C(\mathbb{T})$ and assume that, for some r,*

$$\|(E_n(f) : n \in \mathbb{N})\|^{(r)} < \infty \tag{60}$$

(as it would be the case in case $E_n(f) = O(n^{-r-\varepsilon})$ for some positive ε). Then $f \in C^{(r)}(\mathbb{T})$.

Proof: It is sufficient to show that f is the uniform limit of a sequence which is Cauchy in $C^{(r)}(\mathbb{T})$. With $p_n \in \mathcal{P}_{M_n}(f)$, all n, we have $f = \lim_n p_n$; however, we have no way of knowing that the whole sequence $(p_n : n \in \mathbb{N})$ is Cauchy in $C^{(r)}$. The sequence $(p_{2^n} : n \in \mathbb{N})$, on the other hand, is seen to be Cauchy, as follows.

$$\begin{aligned}
\|D^r(p_{2^j} - p_{2^J})\| &\le \sum_{k=j+1}^{J} \|D^r(p_{2^k} - p_{2^{k-1}})\| \\
&\le \sum_{k=j+1}^{J} (2^k)^r \|p_{2^k} - p_{2^{k-1}}\| \\
&\le \sum_{k=j+1}^{J} (2^k)^r 2E_{2^{k-1}}(f) \\
&\le 2\text{const}_r \sum_{k=2^j}^{2^J} k^{r-1} E_k(f),
\end{aligned}$$

the last inequality by (56). By assumption, this last sum goes to zero as $j \to \infty$. □

This theorem is quite remarkable since it says that even if f is very smooth *except* on a small subinterval of $\mathbb{T}$, it will be hard to approximate f well by trig.pols.

There are similar results for approximation by *algebraic* polynomials on $[a \,..\, b]$, except for the following hitch. In contrast to $\mathbb{T}$, the points in $[a \,..\, b]$ are *not* all equal. Taking for simplicity the max-norm on $[a \,..\, b]$, one has the Jackson inequality (see below)

$$\text{dist}_\infty(f, \Pi_n) =: E_n(f)_\infty \le \text{const} n^{-1} \|Df\|_\infty. \tag{61}$$

However, the best inequality relating $\|f\|$ and $\|Df\|$ for $f \in \Pi_n \in C([a..b])$ is the **Markov Inequality**

$$\|Df\|_\infty \le \frac{2}{|b-a|} n^2 \|f\|_\infty, \tag{62}$$

(note that, because of the possibility of dilating, the interval length must figure in exactly the position it appears), and these two inequalities do not at all match in the sense of Theorem 54. It is possible, though, to prove that $f \in C([a \,..\, b])$ must be in $C^{(r)}(I)$ for all closed subintervals I of $(a \,..\, b)$ in case $E_n(f)_\infty = O(n^{-n-\varepsilon})$ for some positive ε.

Because $\mathbb{T}$ has no endpoints, it is easier to prove Jackson inequalities for $M_n = \overset{\circ}{\Pi}_n \subset C(\mathbb{T})$ than for $\Pi_n \subset C([a\,..\,b])$. The standard argument uses integral operators of **convolution type**. With L_n functions still to be specified, one considers the approximation

$$L_n{*}f : t \mapsto \int_{-\pi}^{\pi} L_n(t-s)f(s)\,\mathrm{d}s = \int_{-\pi}^{\pi} f(t+s)L_n(s)\,\mathrm{d}s.$$

One makes the following assumptions:
(i) $L_n \in \overset{\circ}{\Pi}_n$.
Then, $\overset{\circ}{\Pi}_n$ being **translation-invariant**, i.e., $\forall\{p \in \overset{\circ}{\Pi}_n,\ s \in \mathbb{T}\}\ p(\cdot - s) \in \overset{\circ}{\Pi}_n$, we necessarily have

$$L_n(t-s) = \sum_j \varphi_j(t)\psi_j(s),$$

with $[\varphi_j : -n \le j \le n]$ any basis for $\overset{\circ}{\Pi}_n$. Therefore

$$L_n{*}f = \sum_j \varphi_j \int_{-\pi}^{\pi} \psi_j f \quad \in\ \overset{\circ}{\Pi}_n.$$

(ii) $L_n{*}1 = 1$.
Assuming that

$$\lambda_n := L_n{*}()^0 \ne 0,$$

this can always be achieved by then replacing L_n by L_n/λ_n. After that,

$$(f - L_n{*}f)(t) = \int_{-\pi}^{\pi} \big(f(t) - f(t+s)\big)L_n(s)\,\mathrm{d}s = \int_{-\pi}^{\pi} (-\Delta_s f)(t)\,L_n(s)\,\mathrm{d}s,$$

with

$$\Delta_h f := f(\cdot + h) - f.$$

Note that

$$\|\Delta_h f\| \le \sup_{k<h} \|\Delta_k f\| =: \omega(f,h),$$

with $\omega(f,\cdot)$ the **(uniform) modulus of continuity of** f. One could now try to get an error estimate involving $\omega(f,1/n)$ by making use of the fact that

$$\omega(f,s) = \omega(f, ns{\cdot}1/n) \le (1+ns)\omega(f,1/n),$$

hoping that

$$\sup_n \int_{-\pi}^{\pi} (1+n|s|)|L_n(s)|\,\mathrm{d}s < \infty.$$

However, the L_n usually employed satisfy the following additional assumption:
(iii) L_n even.

and, with this, we even have

$$f(t) - L_n{*}f(t) = \int_0^\pi \big(f(t) - f(\cdot + s)\big) + \big(f(t) - f(\cdot - s)\big) L_n(s)\,\mathrm{d}s = \int_0^\pi (-\Delta_s^2 f)(t + \cdot) L_n,$$

with

$$\Delta_s^2 f := \Delta_s(\Delta_s f),$$

and

$$\|\Delta_h^2\| \le \sup_{k<h} \|\Delta_k^2 f\| =: \omega_2(f, s),$$

and $\Omega_2(f, \cdot)$ called the **second (uniform) modulus of smoothness of** f. In this language, $\omega(f, \cdot)$ is called the **first (uniform) modulus of smoothness of** f. It is clear how one would define the rth modulus of smoothness, for any natural r.

Since $\omega_2(f, h) = \omega_2(f, (nh)/n) \le (1 + nh)^2 \omega_2(f, 1/n)$, natural assumption now is the following **moment condition**:
(iv) $\sup_n \int_0^\pi (ns)^k\,|L_n(s)|\,\mathrm{d}s \ < \infty$, $k = 0, 1, 2$.
and with it, we get the typical **Jackson inequality**

$$\|f - L_n{*}f\| \le c\omega_2(f; 1/n).$$

It remains to find suitable L_n.
The simplest choice is $L_n = D_n / D_n{*}1$ with

$$D_n(t) := \frac{\sin(n + 1/2)t)}{2\sin(t/2)}$$

Dirichlet's kernel. For it, $L_n \in \overset{\circ}{\Pi}_n$, $D_n{*}1 = \pi$, and L_n even. However, L_n decays too slowly away from 0 to satisfy (iv). Nor would a more sophisticated analysis help here. For, in this case, $L_n{*}f$ is the truncated Fourier series for f and this is well-known not to converge to f uniformly for every $f \in C(\mathbb{T})$.

The next choice is $L_n = F_n / F_n{*}1$, with

$$F_n(t) := \frac{1}{2(n+1)} \left(\frac{\sin(n+1)t/2}{\sin t/2}\right)^2$$

Fejér's kernel. One checks that

$$F_n = \sum_0^n D_k,$$

hence $F_n \in \overset{\circ}{\Pi}_n$, and $F_n{*}1 = \pi$. Also, $F_n \ge 0$, hence only need

$$\int_0^\pi (1 + ns)^2 F_n(s)\,\mathrm{d}s = O(1).$$

However this, unfortunately, is not true. Thus, while, by Korovkin, $F_n{*}f$ converges to f uniformly for every $f \in C(\mathbb{T})$, we don't, offhand, get a Jackson estimate from it since (iv) does not hold for it.

Jackson's choice is the **Jackson kernel**

$$K_n(t) := \left(\frac{\sin mt/2}{\sin t/2}\right)^4 / \lambda_n, \quad m := \lfloor n/2 \rfloor + 1,$$

normalized to have $K_n{*}1 = 1$. It is a special case of the **generalized Jackson kernel**s:

$$K_{n,r}(t) := \left(\frac{\sin mt/2}{\sin t/2}\right)^{2r} / \lambda_{n,r}, \quad m := \lfloor n/r \rfloor + 1.$$

Since

$$\left(\frac{\sin mt/2}{\sin t/2}\right)^2 / m = 1/2 + \sum_{k=0}^{m-1} (1 - k/m) \cos kt$$

(as one verifies!), it follows that $K_{n,r}$ is an even, nonnegative, trig.pol. of degree $\le n$. In particular $K_{n,r}{*}1 > 0$. Further, one verifies that $K_{n,r}$ satisfies the **moment conditions**

$$\sup_n \int_0^\pi (nt)^k K_{n,r}(t)\,\mathrm{d}t < \infty, \quad k = 0, 1, \ldots, 2r-2.$$

Thus, for $r \ge 2$, the sequence $L_n = K_{n,r}/K_{n,r}{*}1$ satisfies all four conditions (i)–(iv) above.

For $r = 2$, Jackson got in this way the Jackson Inequality. For $r > 2$, the additional moment conditions provide, for $L_n := K_{n,r}/K_{n,r}{*}1$, the inequalities (due to Stechkin):

$$\|f - L_n{*}f\|_p \le C_r \omega_{2r-2}(f, 1/n)_p,$$

for any $f \in L_p(\mathbb{T})$ and any $1 \le p \le \infty$.

The resulting constants are far from sharp. Favard showed that, with S the unit ball in the semi-normed lss $Y := W_p^r(\mathbb{T})$ of $X_p(\mathbb{T})$, with semi-norm

$$\|f\|_Y := \|D^r f\|_p,$$

$$\sup_{f \in S} E_{n-1}(f) = E_{n-1}(B_p^r) \le K_r/n^r,$$

with this inequality exact for $p = 1, \infty$, and the numbers K_r, the so-called **Favard constants** given as the value of a fast converging series whose value, for large r, is indistinguishable from $\pi/4$.

The extremal function, B_p^r, is a **Bernoulli spline**, of which, perhaps, more anon.

It is standard to prove Jackson's theorem for *algebraic* polynomials from the one for trigonometric polynomials. For this, one first shifts and scales $[a \,..\, b]$ into the interval $[-1 \,..\, 1]$. Then one considers the map

$$F : C[-1 \,..\, 1] \to C(\mathbb{T}) : f \mapsto Ff : \theta \mapsto f(\cos(\theta)).$$

(Draw the picture.) This is a linear map which carries $C[-1\,..\,1]$ isometrically into

$$C(\mathbb{T})_e := \{g \in C(\mathbb{T}) : g(-\cdot) = g\},$$

the space of all **even** 2π-periodic functions. In particular, F maps Π_n onto

$$\overset{\circ}{\Pi}_{n,e} := \overset{\circ}{\Pi}_n \cap C(\mathbb{T})_e.$$

Also,

$$\omega(Ff, \cdot) \le \omega(f, \cdot)$$

since the map $[-1\,..\,1] \to [0\,..\,\pi] : t \mapsto \cos^{-1}(t)$ has slope ≥ 1 in absolute value everywhere. Hence

$$\mathrm{dist}(Ff, \overset{\circ}{\Pi}_n) \le \text{const}\ \omega(Ff, 1/n) \le \text{const}\ \omega(f, 1/n).$$

Therefore, with $m \in \mathcal{P}_{\overset{\circ}{\Pi}_n}(Ff)$,

$$\mathrm{dist}(f, \Pi_n) \le \|f - F^{-1}m\| = \mathrm{dist}(Ff, \overset{\circ}{\Pi}_n) \le \text{const}\ \omega(f, 1/n), \tag{63}$$

provided $F^{-1}m$ is defined and in Π_n. For this, observe that the map

$$C(\mathbb{T}) \to C(\mathbb{T})_e : g \mapsto g_e := (g + g(-\,\cdot))/2$$

is norm-reducing since $g \mapsto g(-\,\cdot)$ is an isometry. Hence, if $g \in C(\mathbb{T})$ (as is the case for $g = Ff$) and $m \in \mathcal{P}_{\overset{\circ}{\Pi}_n}(g)$, then m_e is an even trig.pol. in $\overset{\circ}{\Pi}_n$ and $\|Ff - m_e\| = \|g_e - m_e\| \le \|g - m\| = \mathrm{dist}(g, \overset{\circ}{\Pi}_n)$, hence also $m_e \in \mathcal{P}_{\overset{\circ}{\Pi}_n}(g)$, i.e., we may take $m \in \mathcal{P}_{\overset{\circ}{\Pi}_n}(Ff)$ to be even, hence in $F(\Pi_n)$. In fact, Π_n is Haar, hence we have just proved that the ba from $\overset{\circ}{\Pi}_n$ to an even function is even.

Note, though, that we cannot bound $\omega(f, \cdot)$ in terms of $\omega(Ff, \cdot)$, and this is another indication that we cannot expect the same kind of paired direct/inverse theorems concerning the degree of approximation by algebraic polynomials.

Instead, one has theorems of the following kind.

(64) Theorem (Nikol'skii, Timan). *For some* const *and for all* $f \in C[-1\,..\,1]$ *and all* n*, there exists* $p_n \in \Pi_n$ *so that*

$$|f(t) - p_n(t)| \le \text{const}\ \omega(f, \Delta_n(t)),$$

with

$$\Delta_n : t \mapsto \max\{\frac{1}{n^2}, \frac{\sqrt{1-t^2}}{n}\}.$$

If one is not too worried about the constants involved, then Jackson's theorem provides the right order of the degree of approximation by polynomials to smooth functions, as follows: Starting with Jackson's theorem,

$$\mathrm{dist}(f, \Pi_n) \le \text{const}\ \omega(f, 1/n),$$

and adding to it the fact that, for $f \in C^{(1)}[a\,..\,b]$,

$$\omega(f, h) \le h\|Df\|,$$

we get, for any $p \in \Pi_n$,

$$\operatorname{dist}(f, \Pi_n) = E_n(f - p, \Pi_n) \le \text{const } (1/n)\, \|D(f - p)\|,$$

hence

$$\operatorname{dist}(f, \Pi_n) \le \text{const } (1/n) \operatorname{dist}(f, \Pi_{n-1}).$$

Thus, by induction,

$$\operatorname{dist}(f, \Pi_n) \le \text{const}_r \ (1/n)^r \omega(D^r f, 1/n).$$

Finally, here is a standard result relating a specific K-functional to the modulus of smoothness of a certain order.

(65) Theorem (Johnen). *If T is a closed interval (finite, infinite or all of $\mathbb{R}$) or $\mathbb{T}$, and $r \in \mathbb{N}$, then there exist positive constants c and C such that, for all $p \in [1\,..\,\infty]$ and all $f \in L_p(T)$,*

$$c\omega_r(f, t)_p \le K(f, t^r; L_p, W_p^r) \le C\omega_r(f, t)_p, \quad t > 0. \tag{66}$$

Here, to recall,

$$\omega_r(f, t)_p := \sup_{h \le t} \|\Delta_h^r f\|_p,$$

with

$$\Delta_h^r f := \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} f(\cdot + kh).$$

If T has an endpoint, then $\Delta_h^r f$ may, offhand, only be defined on some proper subset of T; it is taken to be zero wherever it is not defined.

Univariate B-splines

Here is a very swift introduction to B-splines, and thereby to splines *aka* smooth piecewise polynomials.

Let $\mathbf{t} := \ldots \le t_{i-1} \le t_i \le t_{i+1} \le \ldots$ be a nondecreasing sequence. This sequence may be finite, infinite, or even bi-infinite. Two important (and extreme) special cases are

(i) $\mathbf{t} = \mathbb{Z}$, leading to **cardinal splines**;

(i) $\mathbf{t} = \mathbb{B} := (\ldots, 0, 0, 0, 1, 1, 1, \ldots)$, leading to the **Bernstein-Bézier** form (or, BB-form) for polynomials (restricted to $[0\,..\,1]$).

The associated (normalized) B-splines of **order** k for the **knot sequence t** are, by definition, the functions

$$B_j := B_{j,k} := B_{j,k,\mathbf{t}} : x \mapsto \delta_{t_j,\ldots,t_{j+k}} (\cdot - x)_+^{k-1} (t_{j+k} - t_j), \tag{67}$$

with δ_T the divided difference functional introduced earlier (see pp.17ff). The normalization ensures that the B_j form a (nonnegative) **partition of unity** (see (79) below). By using Leibniz' Rule

$$\delta_{\tau_0,\dots,\tau_k}(fg) = \sum_{i=0}^{k} (\delta_{\tau_0,\dots,\tau_i} f)(\delta_{\tau_i,\dots,\tau_k} g)$$

for the divided difference of a product, one readily obtains the **recurrence relations**

$$a(68) \qquad B_{i,k} = \omega_{i,k} B_{i,k-1} \ + \ (1 - \omega_{i+1,k}) B_{i+1,k-1}$$

with

$$(68b) \qquad \begin{aligned} B_{i,1} &:= \chi_{[t_i \,..\, t_{i+1})}; \\ \omega_{i,k}(x) &:= \begin{cases} \frac{x - t_i}{t_{i+k-1} - t_i}, & \text{if } t_i < t_{i+k-1} \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

However, once one knows these recurrence relations, it is most efficient for the development of the B-spline theory to take (68) as the *starting point*, i.e., to *define* B-splines by (68). Equivalently, the development about to be given will make no use of divided differences, but will rely entirely on (68).

It follows at once that $B_{i,k}$ can be written in the form

$$(68) \qquad B_{i,k} = \sum_{j=i}^{i+k-1} b_{j,k} \chi_{[t_j \,..\, t_{j+1})},$$

with each $b_{j,k}$ a polynomial of degree $< k$ since it is the sum of products of $k-1$ linear polynomials. Therefore,

$$B_{i,k} \in \Pi_{<k,(t_i,\dots,t_{i+k})} \qquad \text{on } [t_i \,..\, t_{i+k}],$$

while

$$(69) \qquad B_{i,k} = 0 \qquad \text{off } [t_i \,..\, t_{i+k}].$$

Further, $B_{i,k}$ depends only on the knots $t_i, \dots, t_{i+k}$. For this reason, the alternative notation

$$(70) \qquad B_{i,k} =: B(\cdot | t_i, \dots, t_{i+k})$$

is also customary. The actual smoothness of $B_{i,k}$ depends on the multiplicity with which each of its **interior** knots t_j, $i < j < i + k$, appears in its knot sequence $(t_i \,..\, t_{i+k})$, as we will see in a moment.

Since both $\omega_{i,k}$ and $1 - \omega_{i+1,k}$ are positive on $(t_i \,..\, t_{i+k})$, it follows from (69) by induction on k that $B_{i,k}$ *is positive on* $(t_i \,..\, t_{i+k})$.

Since at most k of the $B_{i,k}$ are nonzero at any one point $x \in \mathbb{R}$, the definition

$$\sum a_i B_{i,k} : \mathbb{R} \to \mathbb{R} : x \mapsto \sum a_i B_{i,k}(x)$$

of $\sum_i a_i B_{i,k}$ as a pointwise sum makes sense for arbitrary a even when the sum has infinitely many terms. We call any such function a **spline of order k with knot sequence $\mathbf{t}$** and denote the collection of all such functions by

$$S_{k,\mathbf{t}}.$$

We deduce from the recurrence relation that

$$\sum a_i B_{i,k} = \sum \bigl(a_i \omega_{i,k} + a_{i-1}(1-\omega_{i,k})\bigr) B_{i,k-1}, \tag{71}$$

showing that the coefficients on the right are affine combinations of neighboring coefficients on the left.

Before proceeding with this, note the following technical difficulty. If $\mathbf{t}$ has a first knot, e.g., $\mathbf{t} = (t_1, \ldots)$, then the sum on the left in (71) starts with $i = 1$, while the sum on the right can only start at $i = 2$ since the coefficient a_{i-1} appears in it. To avoid such difficulty, we agree in this case (and the corresponding case when $\mathbf{t}$ has a last entry) to extend $\mathbf{t}$ in any way whatsoever to a bi-infinite knot sequence, denoting the extension again by $\mathbf{t}$. However, this increases the number of available B-splines, hence also increases the spline space. Since we are still interested only in our original spline space, we further agree to make use of results obtained from the extended situation only to the extent that they don't explicitly involve any of the additional knots. We can be sure none of the additional knots matter if we restrict attention to the largest interval not intersected by the interior of the support of any of the additional B-splines. We call this the **basic interval for $S_{k,\mathbf{t}}$** and denote it by

$$I_{k,\mathbf{t}} := (\mathbf{t}_- \,..\, \mathbf{t}_+)$$

with

$$\mathbf{t}_- := \begin{cases} t_k, & \text{if } \mathbf{t} = (t_1, \ldots); \\ \inf_i t_i, & \text{otherwise,} \end{cases} \qquad \mathbf{t}_+ := \begin{cases} t_{n+1}, & \text{if } \mathbf{t} = (\ldots, t_{n+k}); \\ \sup_i t_i, & \text{otherwise,} \end{cases}$$

In practice, the knot sequence is finite, having both a first and a last knot. In that case, one chooses $I_{k,\mathbf{t}}$ to be closed, and this is fine for the left endpoint, since the definition of B-splines makes them all continuous from the right. For this to work properly at the right endpoint, one modifies the above definition of B-splines to make them continuous from the left at the right endpoint of $I_{k,\mathbf{t}}$.

In summary, even if the given knot sequence is not biinfinite, we may always assume it to be biinfinite as long as we apply the results so obtained only to functions on the basic interval $I_{k,\mathbf{t}}$ determined by the given knot sequence.

With this, consider the special sequence

$$a_i := \psi_{i,k}(\tau) := (t_{i+1} - \tau) \cdots (t_{i+k-1} - \tau)$$

(with $\tau \in \mathbb{R}$). We find for $B_{i,k-1} \neq 0$, i.e., for $t_i < t_{i+k-1}$, that

$$\begin{aligned} a_i\omega_{i,k} + a_{i-1}(1-\omega_{i,k}) &= \psi_{i,k-1}(\tau)\Big((t_{i+k-1}-\tau)\omega_{i,k} + (t_i-\tau)(1-\omega_{i,k})\Big) \\ &= \psi_{i,k-1}(\tau)(\cdot-\tau) \end{aligned}$$

since $f(t_{i+k-1})\omega_{i.k} + f(t_i)(1-\omega_{i,k})$ is the straight line which agrees with f at t_{i+k-1} and t_i. This shows that

$$\sum \psi_{i,k}(\tau)B_{i,k} = (\cdot-\tau)\sum \psi_{i,k-1}(\tau)B_{i,k-1}, \tag{72}$$

hence, by induction, that

$$\sum \psi_{i,k}(\tau)B_{i,k} = (\cdot-\tau)^{k-1}\sum \psi_{i,1}(\tau)B_{i,1}. \tag{73}$$

This proves the following identity.

(74) Marsden's Identity. *For any $\tau \in \mathbb{R}$,*

$$(\cdot-\tau)^{k-1} = \sum_i \psi_{i,k}(\tau)B_{i,k} \tag{75}$$

on $I_{k,\mathbf{t}}$, with $\psi_{i,k}(\tau) := (t_{i+1}-\tau)\cdots(t_{i+k-1}-\tau)$.

Since τ in (75) is arbitrary, it follows that $S_{k,\mathbf{t}}$ contains all polynomials of degree $< k$ (restricted to $I_{k,\mathbf{t}}$). More than that, we can even give an explicit expression for the required coefficients, as follows.

By differentiating (75) with respect to τ, we obtain the identities

$$\frac{(\cdot-\tau)^{k-\nu}}{(k-\nu)!} = \sum_i \frac{(-D)^{\nu-1}\psi_{i,k}(\tau)}{(k-1)!}\, B_{i,k}, \quad \nu > 0. \tag{76}$$

On using this identity in the Taylor formula

$$p = \sum_{\nu=1}^{k} \frac{(\cdot-\tau)^{k-\nu}}{(k-\nu)!} D^{k-\nu}p(\tau)$$

for a polynomial p of degree $< k$, we conclude that any such polynomial can be written in the form

$$p = \sum_i \lambda_{i,k}p\ B_{i,k}, \tag{77}$$

with $\lambda_{i,k}$ given by the rule

$$\lambda_{i,k}f := \lambda_{i,k,\mathbf{t}}f := \sum_{\nu=1}^{k} \frac{(-D)^{\nu-1}\psi_{i,k}(\tau)}{(k-1)!}\, D^{k-\nu}f(\tau). \tag{78}$$

Here are two special cases of particular interest. For $p = 1$, we get

$$1 = \sum_i B_{i,k} \tag{79}$$

since $D^{k-1}\psi_{i,k} = (-1)^{k-1}(k-1)!$, and this shows that the $B_{i,k}$ form a **partition of unity**. Further, since $D^{k-2}\psi_{i,k}$ is a linear polynomial which vanishes at

$$t_i^* := (t_{i+1} + \cdots + t_{i+k-1})/(k-1),$$

$$\ell = \sum_i \ell(t_i^*) B_{i,k}, \qquad \ell \in \Pi_1. \tag{80}$$

The identity (76) also gives us various **piecewise** polynomials contained in $S_{k,\mathbf{t}}$: Since $t_i < t_j < t_{i+k}$ implies that $D^{\nu-1}\psi_{i,k}(t_j) = 0$ in case $\nu \le \#t_j := \{t_i : t_i = t_j\}$, the choice $\tau = t_j$ in (76) leaves only terms with support either entirely to the left or else entirely to the right of t_j. This implies that

$$\frac{(\cdot - t_j)_+^{k-\nu}}{(k-\nu)!} = \sum_{i \ge j} \frac{(-D)^{\nu-1}\psi_{i,k}(t_j)}{(k-1)!} \, B_{i,k}, \quad 0 < \nu \le \#t_j. \tag{81}$$

Consequently,

$$(\cdot - t_j)_+^{k-\nu} \;\in\; S_{k,\mathbf{t}} \;\text{ for }\; 1 \le \nu \le \#t_j, \tag{82}$$

(on $I_{k,\mathbf{t}}$).

(83) Theorem. *If $t_i < t_{i+k}$, then the B-splines $B_{i,k}$ are linearly independent and the space $S_{k,\mathbf{t}}$ coincides with the space $S := \Pi^\rho_{<k,\mathbf{t}}$ of all piecewise polynomials of degree $< k$ with breakpoints t_i which are $\rho(i) := k - 1 - \#t_i$ times continuously differentiable at t_i, all i.*

Proof: It is sufficient to prove that, for any finite interval $I := [a\,..\,b]$, the restriction $S_{|I}$ of the space S to the interval I coincides with the restriction of $S_{k,\mathbf{t}}$ to that interval. The latter space is spanned by all the B-splines having some support in I, i.e., all $B_{i,k}$ with $(t_i\,..\,t_{i+k}) \cap I \ne \emptyset$. The space $S_{|I}$ has a basis consisting of the functions

$$(\cdot - a)^{k-\nu},\ \nu = 1, \cdots, k;\ (\cdot - t_i)_+^{k-\nu},\ \nu = 1, \cdots, \#t_i,\ \text{for }\ a < t_i < b, \tag{84}$$

as we showed on page 19. Since each of the functions in (84) lies in $S_{k,\mathbf{t}}$, by (76) and (82), we conclude that

$$S_{|I} \subset (S_{k,\mathbf{t}})_{|I}. \tag{85}$$

On the other hand, the dimension of $S_{|I}$, i.e., the number of functions in (84), equals the number of B-splines with some support in I (since it equals $k + \sum_{a<t_i<b} \#t_i$), hence is an upper bound on the dimension of $(S_{k,\mathbf{t}})_{|I}$. This implies that equality must hold in (85), and that the set of B-splines having some support in I must be linearly independent over I. □

(86) Corollary. *All B-splines having some support on a given interval are linearly independent over that interval.*

(87) Corollary. *If $\hat{\mathbf{t}}$ is a refinement of the knot sequence $\mathbf{t}$, then $S_{k,\mathbf{t}} \subset S_{k,\hat{\mathbf{t}}}$.*

(88) Corollary. *The derivative of a spline in $S_{k,\mathbf{t}}$ is a spline of degree $< k-1$ with respect to the same knot sequence, i.e., $DS_{k,\mathbf{t}} \subseteq S_{k-1,\mathbf{t}}$.*

(89) Remark The word 'derivative' is used here in the pp sense: The 'derivative' of a pp f is, by definition, the pp with the same breakpoint sequence whose polynomial pieces are the first derivative of the corresponding polynomial pieces of f. This makes it possible in (88) to ignore the possibility that $t_i = t_{i+k-1}$ for some i, hence the elements of $S_{k,\mathbf{t}}$ will, in general, fail to be differentiable at such a t_i.

More generally, here and elsewhere, we do not exclude the possibility that some of the $B_{i,k}$ are trivial, i.e., that $t_i = t_{i+k}$ for some i. However, if we were to interpret $B_{i,k,\mathbf{t}}$ as *distributions*, we would have to proceed with more caution. Since, for any smooth f,

(90) $$\delta_{t_i,\ldots,t_{i+k}} f = \int_{\mathbb{R}} M(\cdot|t_i,\ldots,t_{i+k}) D^k f/k!,$$

with

$$M(\cdot|t_i,\ldots,t_{i+k}) := \frac{k}{t_{i+k}-t_i} B_{i,k,\mathbf{t}},$$

we see that

(91) $$\lim_{t_i,\ldots,t_{i+k}\to\tau} M(\cdot|t_i,\ldots,t_{i+k}) \;=\; \delta_\tau$$

which, as a distribution, is quite different from 0. □

The identity (77) can be extended to all spline functions. For this, we agree, consistent with (68b), that all derivatives in (78) are to be taken as limits from the right in case τ coincides with a knot.

(92) Theorem. *If τ in definition (78) of $\lambda_{i,k}$ is chosen in the interval $[t_i \,..\, t_{i+k})$, then, with the* **understanding** *that $D^\nu f(\tau) := D^\nu f(\tau+)$ for any ν and any pp f,*

(93) $$\lambda_{i,k}\Big(\sum_j a_j B_{j,k}\Big) = a_i.$$

It is remarkable that τ can be chosen arbitrarily in the interval $[t_i \,..\, t_{i+k})$. The reason behind this is that $\lambda_{i,k} f$ does not depend on τ at all if f is a polynomial of degree less than k (which could be shown explicitly, but is not used explicitly in the proof here).

Proof: Assume that $\tau \in [t_i \,..\, t_{i+k})$, hence $\tau \in [t_l \,..\, t_{l+1}) \subset [t_i \,..\, t_{i+k})$ for some l, and let p_j be the polynomial which agrees with $B_{j,k}$ on $(t_l \,..\, t_{l+1})$. Then

$$\lambda_{i,k} B_{j,k} = \lambda_{i,k} p_j.$$

On the other hand,

$$p_j = \sum_{m=l+1-k}^{l} \lambda_{m,k} p_j \; p_m,$$

since this holds by (77) on $[t_l \,..\, t_{l+1})$, while, by Corollary 86 or directly from (77), $p_{l+1-k}, \cdots, p_l$ are linearly independent. Therefore necessarily $\lambda_{i,k} B_{j,k} = \lambda_{i,k} p_j$ equals 1 if $i = j$ and 0 otherwise. □

use of the dual functionals: differentiation; dependence on knots

Because of (93), the functionals $\lambda_{i,k}$ are called **the dual functionals for the B-splines**. Strictly speaking, there are many such functionals (of which more anon), but these particular ones have turned out to be quite useful in various contexts. Here are several examples.

Compare

$$\lambda_{i,k} f = \sum_{\nu=1}^{k} \frac{(-D)^{\nu-1} \psi_{i,k}(\tau)}{(k-1)!} \; D^{k-\nu} f(\tau) \tag{94}$$

with

$$\lambda_{i,k-1} D f = \sum_{\nu=1}^{k-1} \frac{(-D)^{\nu-1} \psi_{i,k-1}(\tau)}{(k-2)!} \; D^{k-1-\nu} D f(\tau).$$

Since $(t_i - \cdot)\psi_{i,k-1} = \psi_{i-1,k}$ and $(t_{i+k-1} - \cdot)\psi_{i,k-1} = \psi_{i,k}$, subtraction of the latter from the former gives

$$(t_{i+k-1} - t_i)\psi_{i,k-1} = \psi_{i,k} - \psi_{i-1,k}.$$

Hence

$$\lambda_{i,k-1} D = \frac{\lambda_{i,k} - \lambda_{i-1,k}}{(t_{i+k-1} - t_i)/(k-1)}. \tag{95}$$

Consequently, we get the **differentiation formula**

$$D \sum_i a_i B_{i,k} = \sum_i \frac{a_i - a_{i-1}}{(t_{i+k-1} - t_i)/(k-1)} \, B_{i,k-1}. \tag{96}$$

We now consider how $\lambda_{i,k}$ depends on the knot sequence $\mathbf{t}$. Perhaps surprisingly, although $B_{i,k}$ involves the knots $t_i, \ldots, t_{i+k}$, $\lambda_{i,k}$ only depends on the 'interior' knots, $t_{i+1}, \ldots, t_{i+k-1}$. Further, since $\lambda_{i,k}$ depends *linearly* on $\psi_{i,k}$, it depends *affinely* and *symmetrically* on the points $t_{i+1}, \ldots, t_{i+k-1}$. Indeed, for any α, x, y,

$$((\alpha x + (1-\alpha) y) - \cdot) = \alpha\,(x - \cdot) + (1-\alpha)(y - \cdot).$$

Hence, with

$$\lambda_k : \mathbb{R}^{k-1} \to \mathbb{R} : t \mapsto \lambda_{0,k},$$

we have

$$\lambda_k(\alpha x + (1-\alpha)y, s_2, \dots, s_{k-1}) = \alpha\lambda_k(x, s_2, \dots, s_{k-1}) + (1-\alpha)\lambda_k(y, s_2, \dots, s_{k-1})$$

as well as

$$\lambda_k(s) = \lambda_k(s \circ \sigma)$$

for any permutation σ of order $k-1$.

Consider, in particular, $t_1 = \dots = t_{k-1} = x$. Then $\psi_{0,k} = (x - \cdot)^{k-1}$, hence

$$(-D)^{\nu-1}\psi_{0,k}(\tau)/(k-1)! = (x-\tau)^{k-\nu}/(k-\nu)!$$

and therefore

$$\lambda_k(x, \dots, x) = \sum_{\nu=1}^{k} \frac{(x-\tau)^{k-\nu}}{(k-\nu)!}\delta_\tau D^{k-\nu}.$$

In particular

$$\lambda_k(x, \dots, x)p = p(x), \quad p \in \Pi_{<k}.$$

This meshes entirely with the fact that, if $x = t_{i+1} = \dots = t_{i+k-1}$, then $B_{i,k}$ is the only $B_{j,k}$ which is nonzero at x, hence $B_{i,k}(x)$ must be 1 (since the $B_{j,k}$ form a partition of unity), and therefore

$$\sum_j a_j B_{j,k}(x) = a_i$$

in this case.

blossoms

It is worthwhile to point out that, associated with each $p \in \Pi_r$, there is a unique symmetric r-affine form called its **polar form** (in Algebra) or its **blossom** (in CAGD), denoted therefore here by

$$\overset{\omega}{p},$$

for which

$$\forall\{x \in \mathbb{R}\}\ p(x) = \overset{\omega}{p}(x, \dots, x).$$

If $p = \sum_j ()^j c_j \in \Pi_r$, then

$$\overset{\omega}{p}(s_1, \dots, s_r) := \sum_j \sum \{\prod_{i \in I} s_i : I \in \binom{\{1, \dots, r\}}{j}\} / \binom{r}{j}\ c_j.$$

We deduce from the above that

$$\overset{\omega}{p}(t_1, \dots, t_{k-1}) = \lambda_k(t_1, \dots, t_{k-1})\, p, \quad p \in \Pi_{<k}.$$

In particular, the ith B-spline coefficient of a kth order spline with knot sequence $\mathbf{t}$ is the value at $(t_{i+1}, \dots, t_{i+k-1})$ of the blossom of *any* of the k polynomial pieces associated with the intervals $[t_j \,..\, t_{j+1})$, $j = i, \dots, i+k-1$. This observation was made, in language incomprehensible to the uninitiated, by de Casteljau in the sixties. It was discovered independently and made plain (and given the nice name of 'blossom') by Lyle Ramshaw in the early eighties.

I hope to get to the multivariate spline theory resulting from the lifting of these ideas to functions of several variables in the last two weeks of this semester.

knot insertion

This has led to the *evaluation of a spline by knot sequence refinement*. Such a refinement can always be reached adding one knot at a time. So, suppose that the knot sequence $\hat{\mathbf{t}}$ has been obtained from the knot sequence $\mathbf{t}$ by the insertion of just one additional term, the point x say. Thus, for some j,

$$\hat{t}_i = \begin{cases} t_i, & \text{for } i \le j; \\ x, & \text{for } i = j+1; \\ t_{i-1}, & \text{for } i > j+1. \end{cases}$$

We saw already that, with

$$x = \alpha t_{i+k-1} + (1-\alpha) t_i,$$

i.e., with

$$\alpha = \omega(x) := \omega_{i,k}(x),$$

we have

$$\lambda_k(x, t_{i+1}, \ldots, t_{i+k-2}) = \alpha \lambda_k(t_{i+1}, \ldots, t_{i+k-1}) + (1-\alpha)\lambda_k(t_i, \ldots, t_{i+k-2}).$$

While this is true for arbitrary i, it matters here only when $i < j+1 < i+k$. Altogether, we find that

$$\hat{\lambda}_{i,k} = \begin{cases} \lambda_{i,k}, & \text{for } i+k \le j+1; \\ \omega(x)\lambda_{i,k} + (1-\omega(x))\lambda_{i-1,k}, & \text{for } i < j+1 < i+k; \\ \lambda_{i-1,k}, & \text{for } j < i. \end{cases}$$

Correspondingly,

$$\sum_i a_i B_{i,k} = \sum_i \hat{a}_i \hat{B}_{i,k},$$

with

$$\hat{a}_i = \begin{cases} a_i, & \text{for } i+k \le j+1; \\ \omega(x) a_i + (1-\omega(x)) a_{i-1}, & \text{for } i < j+1 < i+k; \\ a_{i-1}, & \text{for } j < i. \end{cases}$$

Note that the middle line is exactly the way we computed coefficients by recurrence in (71). This intimate connection between the recurrence relation and knot insertion was first observed by Wolfgang Boehm. Note further that $k-1$-fold insertion of the knot x produces eventually a knot sequence $\tilde{\mathbf{t}}$, in which all the interior knots for $\tilde{B}_{j,k}$ are equal to x, hence, correspondingly,

$$\tilde{a}_j = (\sum_i a_i B_{i,k})(x).$$

Since also

$$\hat{t}^*_i = \begin{cases} t^*_i, & \text{for } i+k \le j+1; \\ \omega(x) t^*_i + (1-\omega(x)) t^*_{i-1}, & \text{for } i < j+1 < i+k; \\ t^*_{i-1}, & \text{for } j < i, \end{cases}$$

knot insertion has a very pretty geometric interpretation in terms of the **control polygon**

$$C_{k,\mathbf{t}} f$$

of the spline $f := \sum_j a_j B_{j,k}$. This is, by definition, the broken line whose vertices are the **control points** for $\sum_j a_j B_{j,k}$, i.e., the points

$$(P_i := (t_i^*, a_i) \in \mathbb{R}^2 : i).$$

We have just proved the following:

(97) Proposition. *If $\hat{\mathbf{t}}$ is obtained from $\mathbf{t}$ by the insertion of one additional knot, then, for any $f \in S_{k,\mathbf{t}}$, $C_{k,\hat{\mathbf{t}}} f$ interpolates, at its breakpoints, to $C_{k,\mathbf{t}} f$ (and is thereby uniquely determined).*

Below are two figures, to illustrate this geometric interpretation of knot insertion.

This implies that the map $a \mapsto f_a := \sum_i a_i B_{i,k}$ is **variation-diminishing**, meaning that f_a has no more variations in sign than does the sequence a, in a sense to be made precise next.

variation diminution

For a real sequence a without any zero entries,

$$S(a) := \#\{i : a(i)a(i+1) < 0\}$$

denotes the number of **sign changes** in it. It is less clear what this number should be in case a has some zero entries. The maximum number of sign changes obtainable in such a sequence by an appropriate choice in the sign of any zero entry is called the number of **weak sign change**s in it, denoted

$$S^+(a).$$

The minimum number so obtainable is denoted by

$$S^-(a)$$

and called the number of **strong sign change**s. If we think of choosing the sign of the zero entries as making very small perturbations of these entries, then we see that strong sign changes are impervious to small perturbations, while weak sign changes are not.

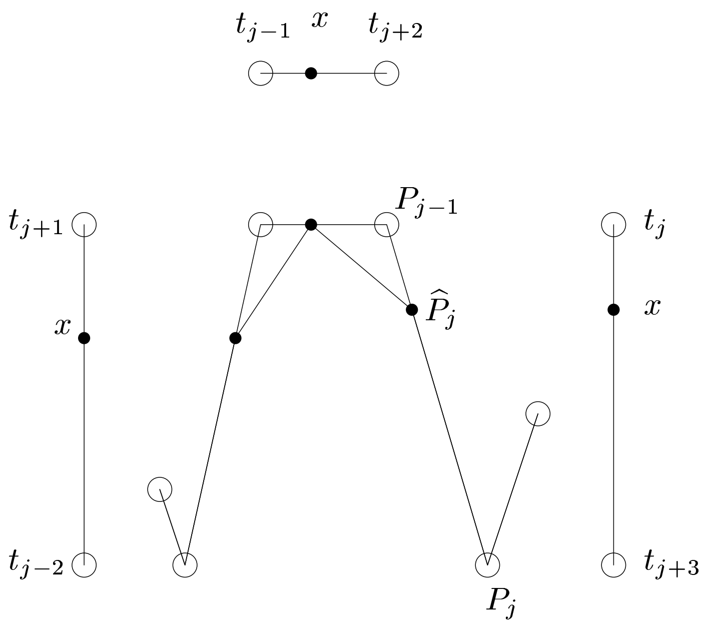
The following Lemma is immediate (from the geometric picture of knot insertion).

(101) Lemma (Lane, Riesenfeld). *If $\tilde{a}$ is the B-spline coefficient sequence obtained from the sequence a by the insertion of (zero or more) knots, then*

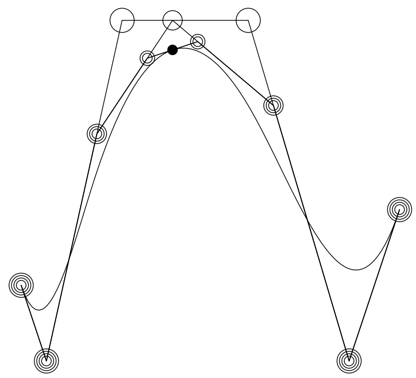
$$S^-(\tilde{a}) \le S^-(a).$$

If f is a function on an interval (including the interval $\mathbb{R}$), one defines

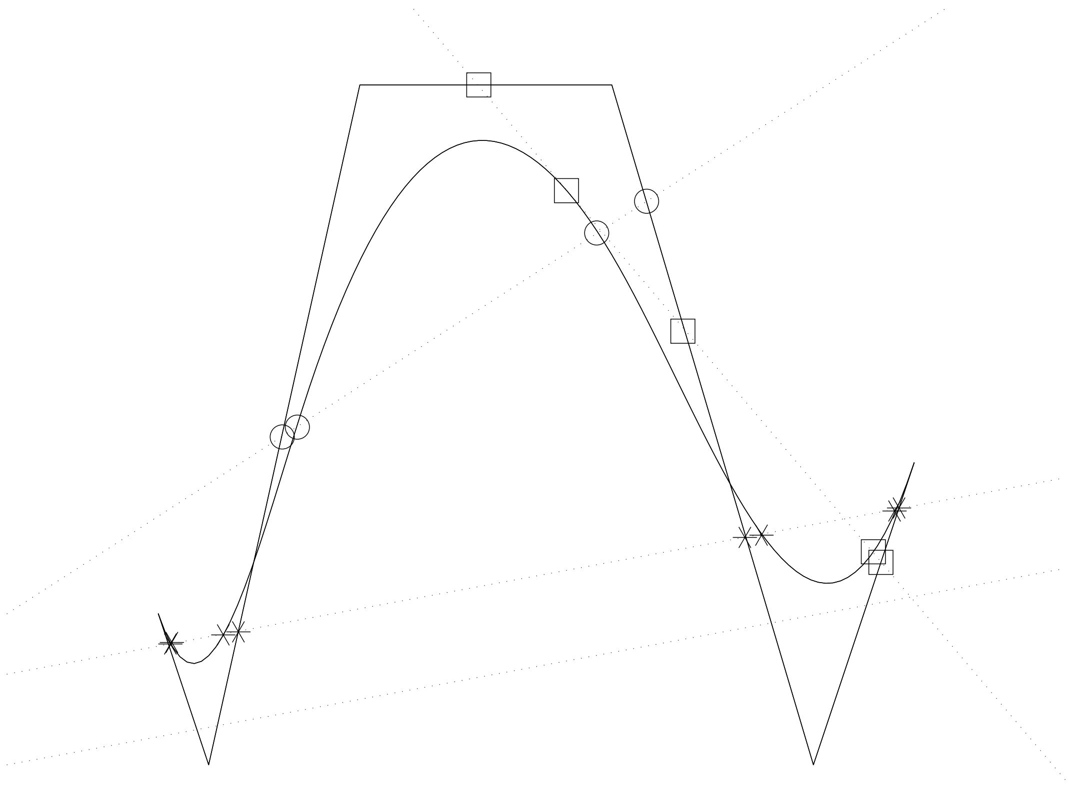
$$S^-(f) := \sup_{x_1 < \cdots < x_r} S^-(f(x_1), \ldots, f(x_r)).$$



(98) Figure Insertion of $x = 2$ into the knot sequence
$\mathbf{t} = (0,0,0,0,\mathbf{1},\mathbf{3},5,5,5,5)$, with $k = 4$.



(99) Figure The cubic spline and its control polygon from (98)Figure and the sequence of control polygons generated by three-fold insertion of the same knot. (The finest control polygon differs from its predecessor only by an additional vertex point.)



(100) Figure A cubic spline, its control polygon, and various straight lines intersecting them. The control polygon *exaggerates* the shape of the spline. The spline crossings are bracketed by the control polygon crossings.

(102) Proposition. $S^-(\sum_j a_j B_j) \le S^-(a)$.

Proof: Insert into $\mathbf{t}$ each of the entries in a given increasing sequence (x_i) enough times to have them appear in the resulting knot sequence $\tilde{\mathbf{t}}$ at least $k-1$ times. Then $(f(x_1), \ldots, f(x_r))$ is a subsequence of the resulting B-spline coefficient sequence $\tilde{a}$, hence

$$S^-(f(x_1), \ldots, f(x_r)) \le S^-(\tilde{a}) \le S^-(a).$$

□

(103) Definition. Schoenberg's (spline) operator *is, by definition, the linear map* $V = V_{k,\mathbf{t}}$ *given by the rule*

$$Vf := \sum_j f(t_j^*) B_j.$$

(104) Proposition. *Schoenberg's spline operator is* **variation-diminishing**. *Precisely, for any* f *and any* $\ell \in \Pi_1$,

$$S^-(Vf - \ell) \le S^-(f - \ell).$$

Proof: By (80), $V\ell = \ell$ (on $I_{k,\mathbf{t}}$) for all $\ell \in \Pi_1$, hence $V(f-\ell) = Vf - \ell$, therefore, by (102)Proposition, and by the strict increase in the sequence $t_i^* : i = 1, \ldots, n$ $S^-(Vf-\ell) \le S^-((f-\ell)(t_i^*) : i) \le S^-(f-\ell)$. □

zeros of a spline, counting multiplicities

The situation concerning weak sign changes is a bit more complicated. In its discussion, we refer to any maximal sequence of consecutive zeros in a sequence as a **zero interval**. We call such a zero interval **rich** if its length is less than its multiplicity as a (weak) sign change; e.g., $(+,0,+)$ and $(+,0,0,-)$ are rich while $(+,0,-)$ and $(+,0,0,+)$ are not. If we call the nonzero entries (if any) nearest to a zero interval its **fences**, then the zero interval is rich iff either there is at most one fence, or else the fences have the same sign and the length is odd, or else the fences have opposite sign and the length is even.

With this, let $f = \sum_i a_i B_{i,k,\mathbf{t}}$ and consider the case that just one knot is added, at the point x say, giving the refined knot sequence $\hat{\mathbf{t}}$, with

$$\hat{t}_j = t_j \le x = \hat{t}_{j+1} < t_{j+1} = \hat{t}_{j+2}.$$

The sequence

$$b := (\ldots, \hat{a}_{j-k+1} = a_{j-k+1}, \hat{a}_{j-k+2}, a_{j-k+2}, \ldots, a_{j-1}, \hat{a}_j = \hat{a}_{j+1}, \ldots)$$

has the same number of strong sign changes as does a, since it is a sequence of consecutive points on the control polygon

$$C_{a,\mathbf{t}} := C_{k,\mathbf{t}} f$$

and includes all the vertices of that polygon. Since $\hat{a}$ is obtained from b by the omission of certain terms, this, together with induction, is a proof of (101)Lemma. However, $S^+(b)$ may well be bigger than $S^+(a)$. Of course, this will happen only if one or more of the $\hat{a}_i$ appearing in b are zero. If such a $\hat{a}_i$ is zero, then a_{i-1} and a_i must be of opposite sign. Now, if either a_{i-1} or a_i is nonzero, then the other must be nonzero and, regardless of how we choose the sign of $\hat{a}_i = 0$, no additional sign change is created. The difficulty comes when either a_{i-1} or a_i is zero, hence both must be zero. For, now we have three consecutive zeros where before we only had two, hence have increased the length of the corresponding zero interval while not changing the sign of its fences (if any), and this may well make $S^+(b)$ bigger than $S^+(a)$. Of course, we are ultimately only interested in $S^+(\hat{a})$, and $\hat{a}$ is obtained from b by the omission of the $k-2$ terms $a_{j-k+2}, \ldots, a_{j-1}$, so the value of $S^+(\hat{a})$ is still in doubt.

Consider first the case that all k consecutive coefficients $a_{j-k+1}, \ldots, a_j$ are part of a zero interval. Then also $\hat{a}_{j-k+1}, \ldots, \hat{a}_{j+1}$ are all zero, hence $\hat{a}$ has a longer zero interval, and therefore, $S^+(\hat{a}) = 1 + S^+(a)$ provided the longer zero interval is rich.

In any other case, the resulting string of consecutive zeros in $\hat{a}$ is no bigger than it was in a and its fences (if any) have retained the sign they had in a. Consequently, $S^+(\hat{a}) \le S^+(a)$. Since $a_{j-k+1}, \ldots, a_j$ are the B-spline coefficients which determine $f = \sum_i a_i B_{i,k}$ on the interval $[t_j \,..\, t_{j+1})$ which, by definition of j, contains that additional knot $x = \hat{t}_{j+1}$, we conclude

(105) Lemma. *If $\hat{\mathbf{t}}$ is obtained from $\mathbf{t}$ by the insertion of just one knot, x, and $a_i B_{i,k}(x) \neq 0$ for some i, then $S^+(\hat{a}) \le S^+(a)$.*

Actually, more than that is true, as has first been pointed out and made use of by T.N.T. Goodman in his recent paper on the zeros of a spline: Any zero interval strictly inside the sequence $a_{j-k+1}, \ldots, a_j$ is *reduced* in length by insertion of a knot into $[t_j \mathinner{\ldotp\ldotp} t_{j+1})$. Here is the precise statement.

(106) Lemma. *If $\hat{\mathbf{t}}$ is obtained from $\mathbf{t}$ by the insertion of just one knot, x, with $t_j \le x = \hat{t}_{j+1} < t_{j+1}$, then any zero interval strictly inside the sequence $a_{j-k+1}, \ldots, a_j$ gives rise to a zero interval of lesser length in the corresponding (longer) sequence $\hat{a}_{j-k+1}, \ldots, \hat{a}_{j+1}$ with fences of the same sign as before, hence*

$$S^+(\hat{a}_{j-k+1}, \ldots, \hat{a}_{j+1}) < S^+(a_{j-k+1}, \ldots, a_j)$$

in case the former zero interval was rich.

Proof: If $(a_\nu, \ldots, a_\mu)$ is the zero interval in question, then

$$a_{\nu-1}\hat{a}_\nu > 0 = \hat{a}_{\nu+1} = \ldots = \hat{a}_{\mu-1} < \hat{a}_\mu a_{\mu+1}.$$

□

(107) Corollary. *If $f^{\mathbf{t}} := \sum_j |a_j| B_{j,k,\mathbf{t}} > 0$ on $(t_1 \mathinner{\ldotp\ldotp} t_{n+k})$, and $\#t_i < k$ (hence all B-splines involved are continuous) and $a_1 a_n \neq 0$, then $f := \sum_j a_j B_{j,k} = \sum_j \tilde{a}_j B_{j,k,\tilde{\mathbf{t}}}$, with $S^+\tilde{a} = S^-(\tilde{a})$.*

Proof: The proof is by induction on the difference $S^+(a) - S^-(a)$, there being nothing to prove when this difference is zero. In the contrary case, there is some zero interval $(a_i : \nu \le i \le \mu)$ in a with $S^+(a_i : \nu-1 \le i \le \mu+1) > S-+(a_i : \nu-1 \le i \le \mu+1)$. The length of $(a_i : \nu \le i \le \mu)$ cannot exceed $k-2$ since, otherwise, $f^{\|}(t_{\mu+1}) = 0$. Consequently, repeated insertion of some $x \in (t_\mu \mathinner{\ldotp\ldotp} t_{\nu+k}$ (not already in $\mathbf{t}$, to be sure that all B-splines stay continuous) will, by the Lemma, give a new representation $f = \sum_j \hat{a}_j \hat{B}_j$ satisfying the same hypotheses as the representation we started with and have the same coefficient sequence, except that the segment $(a_i : \nu-1 \le i \le \mu+1)$ has been replaced by the segment $b := (\hat{a}_{\nu-1} = a_\nu - 1, \hat{a}_\nu, \ldots, a_{\mu+1})$ with at most one zero coefficient, hence $S^+(b) = S^-(b) = S^+(a_i : \nu-1 \le i \le \mu+1)$, showing that $S^+(\hat{a}) - S^-(\hat{a}) < S^+(a) - S^-(a)$. Hence, by induction hypothesis, there is a knot refinement $\tilde{\mathbf{t}}$ of $\hat{\mathbf{t}}$, hence of $\mathbf{t}$, of the desired sort. □

With this, we are ready to make a definition of the multiplicity of a zero of a spline which is taylormade for the theorem that bounds the number of zeros of a spline in terms of sign changes in its coefficient sequence. The definition satisfies the requirement that multiplicity should count the number of simple zeros (i.e., strong sign changes) that some 'nearby' spline of the same degree and with the same number of knots has in the neighborhood of the zero in question.

(108) Definition. *A* **zero** *of the spline* $f = \sum_j a_j B_j$ *with knot sequence* $\mathbf{t} = (t_i : i = 1, \ldots, n+k)$ *is, by definition, any maximal closed interval* $I = [\alpha\,.\,.\beta]$ *in* $(t_1\,.\,.t_{n+k}) \cap f^{-1}\{0\}$. *Its multiplicity is the number*

$$S^+(\tilde{a}_{\nu-1}, \ldots, \tilde{a}_{\mu+1}),$$

with $\tilde{\mathbf{t}}$ *the knot sequence obtained by inserting some point of* I *until* I *is also a zero of*

$$f^{\tilde{\mathbf{t}}} = \sum_j |\tilde{a}_j| \tilde{B}_j,$$

and $(\tilde{a}_\nu, \ldots, \tilde{a}_\mu)$ *the zero interval of* $\tilde{a}$ *associated with* I.

We now show that the multiplicity of any such I is not dependent on whatever knot insertion has been used, but is entirely determined by the behavior of the spline in an open neighborhood, however small, of the zero. This is trivially so when I is already a zero of $f^{\mathbf{t}}$. In particular, this is so when I has interior. Consider therefore the case that $I = \{x\}$, but $f^{\mathbf{t}}(x) \neq 0$, let s be the multiplicity of x as a knot in $\mathbf{t}$, and let $\tilde{s}$ be its multiplicity as a knot in $\tilde{\mathbf{t}}$. Explicitly, assume that

$$\tilde{t}_j < x = \tilde{t}_{j+1} = \ldots = \tilde{t}_{j+\tilde{s}} < \tilde{t}_{j+\tilde{s}+1}.$$

Let $(\tilde{a}_\mu, \ldots, \tilde{a}_\nu)$ be the corresponding zero interval in $\tilde{a}$. Then

$$j + 1 - k < \mu \leq j + \tilde{s} + 1 - k \leq j \leq \nu < j + \tilde{s},$$

the strict inequalities because x is an isolated zero of f. Now, according to (92)Theorem, every a_i with $i = j+1, \ldots, j+\tilde{s}$ is entirely determined by the vector $(D^r f(x+) : r = 0, \ldots, k-1)$, while every a_i with $j+1-k, \ldots, j+\tilde{s}-k$ is determined by $(D^r f(x-) : r = 0, \ldots, k-1)$. In particular, $S^+(\tilde{a}_{\mu-1}, \ldots, \tilde{a}_{\nu+1})$ depends only on the right and left derivatives of f at x, while, by (107)Corollary,

$$S^+(\tilde{a}_{\mu-1}, \ldots, \tilde{a}_{\nu+1}) \leq S^-(a_{\mu-1}, \ldots, a_{\nu+1-\tilde{s}+s}).$$

Thus, with $Z^+(f)$ the sum of all the zeros of f counting multiplicity, we have

$$Z^+(f) \leq S^+(a),$$

with is a version of a Theorem due to T.N.T. Goodman. However, we can improve on this in the following way: Assuming that $a_1 a_n \neq 0$,

$$\begin{aligned}
Z^+(f) \leq{} & S^-(a_1, \ldots, a_{\nu_1-1}) + S^+(a_{\nu_1-1}, \ldots, a_{\mu_2+1}) \\
& + S^-(a_{\mu_2+1}, \ldots, a_{\nu_2-1}) + S^+(a_{\nu_2-1}, \ldots, a_{\mu_3+1}) \\
& \vdots \\
& + S^-(a_{\mu_r+1}, \ldots, a_n),
\end{aligned} \tag{109}$$

with $(a_{\nu_i}, \ldots, a_{\mu_{i+1}})$ the zero intervals belonging to zeros which are also zeros of $f^{\mathbf{t}}$.

spline interpolation

We consider **spline collocation**, i.e., interpolation from $S_{k,\mathbf{t}}$ at the increasing sequence (x_i) of points. We consider this under the assumption that $\mathbf{t} = (t_i : i = 1, \ldots, n+k)$, with $\#t_i < k$, all i, hence $S_{k,\mathbf{t}} \subset C(\mathbb{R})$. This simplifying assumption avoids discussion otherwise needed in case some x_j agrees with a knot of multiplicity $\geq k$, in which case one would have to specify, in addition, whether it is x_j- or x_j+ one wants. With the assumption, none of the B-splines $B_{i,k}$ is trivial, hence $\dim S_{k,\mathbf{t}} = n$. For this reason, we assume, more precisely, that

$$x = (x_1 < \cdots < x_n).$$

Hence, for given g, we are seeking $f \in S_{k,\mathbf{t}}$ with $f = g$ on x. Equivalently, we are seeking a solution to the linear system $A? = g|_x$, with

$$A := (B_{j,k}(x_i) : i, j = 1, \ldots, n)$$

the so-called **collocation matrix**. There is exactly one interpolant to a given g iff A is invertible iff A is 1-1.

(110) Proposition. *If $A = (B_{j,k}(x_i))$ is invertible, then*

$$t_i < x_i < t_{i+k}, \quad \forall i.$$

Proof: If, for some i, $t_{i+k} \leq x_i$, then the first i columns of A have nonzero entries only in the first $i-1$ rows, hence A cannot be invertible. Again, if $x_i \leq t_i$, then columns $i, \ldots, n$ of A have nonzero entries only in rows $i+1, \ldots, n$, hence A cannot be invertible. □

Note that the argument used nothing more than the fact that both the sequence x and the sequence of the supports of the B-splines are increasing. In particular, we have proved

(111) Corollary. *If $(B_{m_j,k}(s_i) : i, j = 1, \ldots, r)$ is invertible, with both (m_j) and (s_i) increasing, then $B_{m_i}(s_i) \neq 0$, all i.*

In other words, such a matrix is invertible only if its diagonal entries are nonzero. As it turns out, the converse also holds. The converse of the Proposition is

(112) Schoenberg-Whitney Theorem. *Let $\mathbf{t} = (t_1, \ldots, t_{n+k})$ with $B_{j,k} \neq 0$, all j, and let $x_1 < \cdots < x_n$. Then, $A = (B_{j,k}(x_i))$ is invertible iff $t_i < x_i < t_{i+k}$, all i.*

Proof: We only need to prove the 'if'. Since A is square, it is sufficient to prove that $Aa = 0$ implies $a = 0$. Consider $f = \sum_j a_j B_{j,k}$ with $Aa = 0$. If $a \neq 0$, then, by (86)Corollary, $f \neq 0$. Let $I = (\alpha \mathinner{..} \beta)$ be a maximal open interval in

$$\operatorname{supp} \sum_j |a_j| B_{j,k} = \bigcup \{(t_j \mathinner{..} t_{j+k}) : a_j \neq 0\}.$$

It follows that $I = (t_\nu \mathinner{..} t_{\mu+k})$ for some $1 \le \nu \le \mu \le n$, and that

$$f = f_I := \sum_{j=\nu}^{\mu} a_j B_{j,k} \qquad \text{on} \quad I.$$

In particular, f_I has the distinct zeros $x_\nu, \ldots, x_\mu$, therefore

$$Z(f_I) \ge \mu + 1 - \nu > \mu - \nu \ge S^+(a_\nu, \ldots, a_\mu),$$

a contradiction to (109). □

total positivity

We recall that an (m,n)-matrix C is **totally positive** if, for any strictly increasing index sets $I = (i_1 < \cdots < i_r)$ in $\{1, \ldots, m\}$ and and $J = (j_1 < \cdots < j_r)$ in $\{1, \ldots, n\}$, the determinant $\det C(I,J)$ of the (r,r)-submatrix

$$C(I,J) := \bigl(C(i_p, j_q) : p, q = 1, \ldots, r\bigr)$$

is nonnegative. The most immediately important fact concerning total positivity is the following.

(113) Fact. *If A is invertible and totally positive, then its inverse is* **checkerboard**, *meaning that $A^{-1}(i,j)(-1)^{i-j} \ge 0$, all i, j.*

Proof: By Cramer's rule,

$$A^{-1}(i,j) = (-1)^{i-j} \det A(\backslash j, \backslash i) / \det A.$$

□

(114) Theorem. *For any $x_1 < \cdots < x_n$, any $k > 0$, and any knot sequence $\mathbf{t} = (t_i : i = 1, \ldots, n+k)$, the* **collocation matrix**

$$C := \bigl(B_j(x_i) : i, j = 1, \ldots, n\bigr)$$

is totally positive.

Proof: If $\hat{\mathbf{t}}$ is obtained from $\mathbf{t}$ by the insertion of just one knot, and $\hat{B}_j := B_{j,k,\hat{\mathbf{t}}}$, all j, then

$$\hat{B}_j = (1 - \alpha_{j-1}) \hat{B}_{j-1} + \alpha_j \hat{B}_j,$$

with all $\alpha_j \in [0 \mathinner{..} 1]$. Since the determinant of a matrix is a linear function of the columns of that matrix, we have, e.g.,

$$\det(\ldots, B_j(x), \ldots) = (1 - \alpha_{j-1}) \det(\ldots, \hat{B}_{j-1}(x), \ldots) + \alpha_j \det(\ldots, \hat{B}_j(x), \ldots),$$

with ... unchanged in their respective places. It follows that, for any I, J,

$$\det C(I,J) = \sum_H \gamma_H \det \hat{C}(I,H),$$

with all the $\gamma_H \ge 0$, and the sum, offhand, over certain nondecreasing sequences, since only neighboring columns of $\hat{C}$ participate in a column of C. However, we may omit all H which are not strictly increasing, since the corresponding determinant is trivially zero. Therefore,

$$\det C(I,J) = \sum_H \gamma_H \det \hat{C}(I,H),$$

with the $\gamma_H \ge 0$ and all H strictly increasing.

Now insert each of the x_i enough times so that the resulting refined knot sequence $\tilde{\mathbf{t}}$ contains each x_i exactly $k-1$ times. By induction, we have

$$\det C(I,J) = \sum_H \gamma_H \det \tilde{C}(I,H),$$

with the $\gamma_H \ge 0$ and all H strictly increasing. However, in each row of $\tilde{C}$, there is exactly one nonzero entry, namely the entry belonging to $\tilde{B}_j$ with $t_j < x_i = \tilde{t}_{j+1}$, and that entry equals 1. Further, the sole nonzero entry in subsequent rows occurs to the right of this one. Thus $\tilde{C}(I,H)$ has a nonzero entry in every row iff $C(I,H)$ is the identity matrix. Consequently, $\det \tilde{C}(I,H) \ge 0$ for any H, and therefore $\det C(I,J) \ge 0$. □

the Chebyshev spline

The Chebyshev spline in $S_{k,\mathbf{t}}$ is a minor miracle when defined in the following way.

Definition. *For given k and given $\mathbf{t} = (t_i : i = 1, \ldots, n+k)$ with $\#t_i < k$, all i, the* **Chebyshev spline for** $S_{k,\mathbf{t}}$ *is*

$$C_{k,\mathbf{t}} := \sum_j (-1)^{n-j} B_j / \operatorname{dist}(B_j, S_j),$$

with $S_j := \operatorname{ran}[B_i : i \ne j]$ and the distance measured in the max-norm on $\mathbb{R}$, hence $\operatorname{dist}(B_j, S_j) > 0$, *all j, by (86)Corollary.*

The Chebyshev spline is aptly named because of the following.

(115) Lemma. *For given $\mathbf{t} = (t_i : i = 1, \ldots, n+k)$ with $\#t_i < k$, all i, $C_{k,\mathbf{t}}$ is, up to multiplication by a scalar, the unique element of $S_{k,\mathbf{t}}$ which equioscillates exactly n times.*

Proof: We first prove existence of such a maximally equioscillating element of $S_{k,\mathbf{t}}$.

Let $x_1 < \cdots < x_n$ with $t_i < x_i < t_{i+k}$, all i. Then there is exactly one spline, $f_{\mathbf{x}} = \sum_j a_j^{\mathbf{x}} B_j$ of degree $\le k$ with knot sequence $\mathbf{t}$ with $f_{\mathbf{x}}(x_i) = (-1)^{n-i}$, all i. Moreover, all $a_i^{\mathbf{x}}$ must be nonzero, by (102).

Since $f_{\mathbf{x}}$ strictly alternates in sign at the n points $x_1, \ldots, x_n$ and vanishes outside $(t_1 \mathinner{..} t_{n+k})$, it follows that $f_{\mathbf{x}}$ has n distinct local extrema $y_1 < \cdots < y_n$ with

$$(-1)^i f_{\mathbf{x}}(y_i) \geq 1, \quad \forall i.$$

This implies that $t_i < y_i < t_{i+k}$. all i. Indeed, if, e.g., $y_i \leq t_i$ for some i, then $B_\mu(y_\nu) = 0$ for all $\nu \leq i \leq \mu$, showing that, on $y_1, \ldots, y_i$, $f_{\mathbf{x}}$ agrees with $\sum_{j<i} a_j^{\mathbf{x}} B_j$. In particular, $S^-(\sum_{j<i} a_j^{\mathbf{x}} B_j) = i - 1 > S^+(a_1^{\mathbf{x}}, \ldots, a_{i-1}^{\mathbf{x}})$, a contradiction.

Consequently, there is exactly one $f_{\mathbf{y}} =: \sum_j a_j^{\mathbf{y}} B_j$ with $f_{\mathbf{y}}(y_i) = (-1)^{n-i}$, all i. For any $\gamma < 1$, the difference, $f_{\mathbf{x}} - \gamma f_{\mathbf{y}}$, strictly alternates in sign on $y_1 < \cdots < y_n$, hence we must have $S^-(a^{\mathbf{x}} - \gamma a^{\mathbf{y}}) = n - 1$, and therefore $(-1)^i a_i^{\mathbf{x}} \geq (-1)^i a_i^{\mathbf{y}} \geq 0$, all i.

In this way, we obtain a sequence $f_m := \sum_j a_j^m B_j$, $m = 1, 2, \ldots$, whose coefficients converge monotonely to some sequence a. Its corresponding sequences $y_1^m < \cdots < y_n^m$ of extrema of f_m therefore also converge, necessarily to a strictly increasing sequence $y_1 < \cdots < y_n$ since f_m strictly alternates in sign on $y_1^m < \cdots < y_n^m$. Let $f := \sum_j a_j B_j$. Then, $1 = (-1)^{n-i} f(y_i) = \|f\|_\infty$ (since $1 = (-1)^{n-i} f_m(y_i^{m-1})$, while $\|f_m\| = \max_i |f_m(y_i^m)|$).

As to uniqueness, we now show that f just constructed must equal $C_{k,\mathbf{t}}$. This amounts to showing that, for each r, f/a_r is the error in the best approximation to B_r from S_r. This is, in fact, so with respect to the uniform norm taken on any interval $[\alpha \mathinner{..} \beta]$ which contains the n extreme points, $y_1, \ldots, y_n$, of f. Indeed, we know that the collocation matrix

$$C := (B_j(y_i) : i, j = 1, \ldots, n)$$

is totally positive (by (114)) and invertible, hence its inverse is checkerboard. In particular, the vector $w := C^{-1}(r, :)$ alternates in sign, hence the corresponding linear functional

$$\lambda_w := \sum_i w(i) \delta_{y_i}$$

is parallel to the error f/a_r, while also $\lambda_w B_j = 0$ for every $j \neq r$, hence λ_w vanishes on S_r. Since $\|f/a_r\| = (-1)^{n-r}/a_r$, it follows from (30) that

$$(-1)^{n-r}/a_r = \operatorname{dist}(B_r, S_r).$$

□

The terminology *Chebyshev spline* unfortunately clashes with the standard term 'Tschebyscheffian' spline, meaning a piecewise function whose pieces all come from the same Haar space. The notation $C_{k,\mathbf{t}}$ unfortunately clashes with the earlier notation concerning the control polygon of a spline. Although whole books have been written on the special case $k = n$, in which case $C_{k,\mathbf{t}}$ on $I_{k,\mathbf{t}}$ agrees with the (suitably scaled and translated) Chebyshev polynomial of degree $k - 1$, the Chebyshev spline is largely unexplored territory (except for K. Mørken's Ph.D. Thesis).

(116) Corollary. *For $x = (x_1 < \cdots < x_n)$ with $t_i < x_i < t_{i+k}$, all i, let $f_{\mathbf{x}} = \sum_j a_j^{\mathbf{x}} B_j$ be the unique element in $S_{k,\mathbf{t}}$ satisfying $f(x_i) = (-1)^{n-i}$, all i. Then* $\operatorname{argmin}_{\mathbf{x}} \|a^{\mathbf{x}}\|_\infty$ *equals the extreme-point sequence of $C_{k,\mathbf{t}}$.*

Proof: In the proof of the preceding lemma, we obtained the Chebyshev spline as the limit of an iteration that started with an arbitrary such $f_{\mathbf{x}}$ and, at each step, reduced the absolute value of every B-spline coefficient; moreover, this reduction was proper unless $f_{\mathbf{x}}$ was already equioscillating. This latter case would imply that $f_{\mathbf{x}} = C_{k,\mathbf{t}}$, by the proven uniqueness. □

Offhand, some of the extreme points of $C_{k,\mathbf{t}}$ may lie outside the basic interval $I_{k,\mathbf{t}} = [t_k \mathinner{\ldotp\ldotp} t_{n+1}]$. However, if we restrict attention to this interval, then we would choose $t_1 = t_k$ and $t_{n+1} = t_{n+k}$. This violates the assumption that $\#t_i < k$, all i. However, assume first that, e.g., $t_1 < t_2 = t_k$. Then f is strictly monotone on $[t_1 \mathinner{\ldotp\ldotp} t_2]$, hence necessarily $t_2 \le y_1$. Thus we may choose $\alpha = t_2$ in this case. For the same reason, we may choose $\beta = t_{n+k-1}$ in case $t_{n+k-1} < t_{n+k}$. That choice made, the location of t_1 and t_{n+k} becomes irrelevant to $S_{k,\mathbf{t}}$ as restricted to $I_{k,\mathbf{t}}$ (though the B-spline coefficients will change as we vary t_1 and t_{n+k}). In particular, we may choose $t_1 = t_k$ and $t_{n+k} = t_{n+1}$.

Let now $\| \ \|$ be the max-norm on the basic interval $I_{k,\mathbf{t}}$. Since S_r is the kernel (in $S_{k,\mathbf{t}}$) of the dual functional $\lambda_{r,k}$, while $\lambda_{r,k} B_r = 1$, it follows that

$$\operatorname{dist}(B_r, S_r) = 1/\|\lambda_{r,k}|_{S_{k,\mathbf{t}}}\|, \qquad r = 1, \ldots, n.$$

In particular, *the map*

$$S_{k,\mathbf{t}} \to \mathbb{R} : \sum_j a_j B_j \mapsto \|\sum_j a_j B_j\| / \|a\|_\infty$$

takes on its minimum at $C_{k,\mathbf{t}}$. This minimum determines the condition of the B-spline basis.

The condition of the B-spline basis

The **condition**

$$\kappa(V) := \|V\| \|V^{-1}\|$$

of a basis V (i.e., an invertible linear map from some $\mathbb{F}^n$ to the normed linear space $\operatorname{ran} V$) measures the extent to which the relative changes in the coordinates a of an element Va may be close to the resulting relative change in Va itself. The closer $\kappa(V)$ to 1, the more closely do these two relative sizes correspond.

For the B-spline basis, using the max-norm both in $\mathbb{R}^n$ and in $S_{k,\mathbf{t}}$, we have

$$\kappa_\infty([B_j : j]) = \sup_a \frac{\|a\|_\infty}{\|\sum_j a_j B_j\|_\infty}$$

since the B_j form a partition of unity, hence trivially $\sup_a \frac{\|\sum_j a_j \varphi_j\|_\infty}{\|a\|_\infty} = 1$.

Further,

$$\kappa_\infty([B_j : j]) = \max_i \|\lambda_{i,k}\|.$$

Hence the following lemma is of interest.

(117) Lemma. *The number*

$$d_k := \sup_{\mathbf{t}} \sup_i \sup_{f\in S_{k,\mathbf{t}}} |\lambda_{i,k,\mathbf{t}} f|/\|f\|_\infty$$

is finite.

Proof: Let $f \in S_{k,\mathbf{t}}$, hence

$$\lambda_{i,k} f = \sum_{\nu=1}^{k} (-D)^{\nu-1}\psi_i(\tau_i)/(k-1)!\, D^{k-\nu} f(\tau_i),$$

and let $I := [t_\ell \mathinner{..} t_{\ell+1}]$ be any knot interval in $[t_i \mathinner{..} t_{i+k}]$, and choose, as we may, $\tau_i \in I$. Then, as both ψ_i and f are polynomials of degree $< k$ on I, (62)Markov's inequality implies that

$$|(-D)^{\nu-1}\psi_i(\tau_i)||D^{k-\nu} f(\tau_i)| \le \mathrm{const}_{k,\nu} \|\psi\|_\infty(I)/|I|^{\nu-1}\, \|f\|_\infty(I)/|I|^{k-\nu}.$$

On the other hand, by choosing, as we may, I to be a largest such knot interval, we can ensure that

$$|t_j - \tau_i| \le k|I|, \quad j = i, \ldots, i+k,$$

therefore $\|\psi_i\|_\infty(I) \le \mathrm{const}_k |I|^{k-1}$. Therefore, altogether, $|\lambda_{i,k} f| \le \mathrm{const}'_k \|f\|_\infty(I)$, which is even stronger than the claim to be proved. □

A more careful quantitative analysis shows that $d_k = O(9^k)$. Better results can be obtained with the aid of the following

(118) Claim. $d_k = \sup_{\mathbf{s}} \|\lambda_{k-1,k}|_{S_{k,\mathbf{s}}}\|([-1\mathinner{..}1])$, *with* $\|\lambda\|(I) := \sup_{f\in \mathrm{dom}\,\lambda} |\lambda f|/\|f\|_\infty(I)$, *and* $\mathbf{s}$ *any knot sequence of the type*

$$s_1 = \cdots = s_k = -1 \le s_{k+1} \le \cdots \le s_{2k-3} \le 1 = s_{2k-2} = \cdots = s_{3k-3}.$$

Proof: Consider any particular $\lambda_i := \lambda_{i,k}|_{S_{k,\mathbf{s}}}$. If $t_{i+1} = t_{i+k-1}$, then (assuming WLOG that $t_i < t_{i+1}$), $\lambda_i = \delta_{t_{i+1}-}$, therefore $\|\lambda_i\| = 1 \le d_k$. In the contrary case, we may assume, after a suitable linear change of the independent variable, that $-1 = t_{i+1}$, $t_{i+k-1} = 1$. Let $\hat{\mathbf{t}}$ be the knot sequence obtained from $\mathbf{t}$ by inserting both -1 and 1 enough times to increase their multiplicity to $k-1$, and let $\hat{i}$ be such that $\hat{t}_{\hat{i}+j} = t_{i+j}$ for $j = 1, \ldots, k-1$. The corresponding spline space $S_{k,\hat{\mathbf{t}}}$ may well be larger than the space $S_{k,\mathbf{t}}$ we started with, but $\lambda_i = \hat{\lambda}_{\hat{i}}$ since, by (78), λ_i only depends on the knots $t_{i+1}, \ldots, t_{i+k-1}$. This shows that

$$\|\lambda_i\| := \sup_{f\in S_{k,\mathbf{t}}} |\lambda_i f|/\|f\|_\infty \le \sup_{f\in S_{k,\hat{\mathbf{t}}}} |\hat{\lambda}_{\hat{i}} f|/\|f\|_\infty = \|\hat{\lambda}_{\hat{i}}\| = \|\hat{\lambda}_{\hat{i}}\|([-1\mathinner{..}1]),$$

the last equality since $\hat{\lambda}_{\hat{i}} f$ only depends on $f|_{[-1..1]}$.

Consequently, $d_k \le \sup_{\mathbf{s}} \|\lambda_{k-1,k}\|([-1\mathinner{..}1])$. The opposite equality is trivial. □

It is believed that $d_k \sim 2^k$. However, earlier hopes that the extremal knot configuration in the Claim would have no interior knots (i.e., would have an $\mathbf{s}$ with all its entries from $\{-1, 1\}$) were dashed by the following simple counter-example: The cubic Chebyshev polynomial can be shown to have maximum B-coefficient 5 when written as an element of $S_{k,\mathbf{t}}$ with $\mathbf{t} = (-1, -1, -1, -1, 1, 1, 1, 1)$; however, the cubic Chebyshev spline for the knot sequence $(-1, -1, -1, -1, 0, 1, 1, 1, 1)$ has B-coefficients $(1, -7/2, 11/2, -7/2, 1)$. Nevertheless, by showing that, for the related extremum problem

$$d_{k,1} := \sup_{\mathbf{t}} \sup_{i} \sup_{f \in S_{k,\mathbf{t}}} |\lambda_{i,k,\mathbf{t}} f| |t_i - t_{i+k}| / \|f\|_1([t_i \,..\, t_{i+k}]), \tag{119}$$

the extremal knot configuration has no interior knots, Karl Scherer was recently able to show that $d_k/2^{k-3/2} \in [(k-1)/k \,..\, (16k-8)/\sqrt{2\pi(k-1)}]$.

It follows that

$$\|a\|_\infty / d_k \le \|\sum_i B_{i,k} a_i\|_\infty \le \|a\|_\infty$$

for any knot sequence $\mathbf{t}$ and any coefficient sequence a.

For such an estimate in L_p, observe that, for any p and with $1/p + 1/p^* = 1$ and by Hölder's Inequality,

$$|\sum_j B_j(x) a_j| \le (\sum_j B_j(x) |a_j|^p)^{1/p} (\sum_j B_j(x))^{1/p^*},$$

hence

$$\|\sum_j B_j a_j\|_p^p \le \sum_j |a_j|^p (t_{j+k} - t_j)/k$$

(using the fact that $\int B_j = (t_{j+k} - t_j)/k$). On the other hand, from (119) and Hahn-Banach, we deduce the following

(120) Proposition. *There exists h_j with with $|h_j| \le d_{k,1}/(t_{j+k} - t_j)$ and support in $[t_j \,..\, t_{j+k}]$ for which*

$$\int_{\mathbb{R}} h_j f = \int_{t_j}^{t_{j+k}} h_j f = \lambda_{j,k} f, \qquad f \in S_{k,\mathbf{t}}.$$

In particular,

$$d_k \le d_{k,1}.$$

Consequently, if $f = \sum_j B_j a_j$ and with $\int_j := \int_{t_j}^{t_{j+k}}$, then

$$a_j = \int_j h_j f \le (\int_j |h_j|^{p^*})^{1/p^*} (\int_j |f|^p)^{1/p},$$

while

$$(\int_j |h_j|^{p^*})^{1/p^*} \le \frac{d_{k,1}}{(t_{j+k} - t_j)} (t_{j+k} - t_j)^{1/p^*} = d_{k,1}/(t_{j+k} - t_j)^{1/p}.$$

Therefore,

$$\sum_j |a_j|^p (t_{j+k} - t_j)/k \le (d_{k,1})^p \sum_j \int_j |f|^p/k \ \le (d_{k,1})^p \|f\|_p^p,$$

the last inequality since at most k of the intervals $[t_j \mathinner{..} t_{j+k}]$ have some given knot interval $[t_\ell \mathinner{..} t_{\ell+1}]$ in common. It follows that, for any $1 \le p \le \infty$,

$$\|c\|_p / d_{k,1} \le \|\sum_i B_{i,k,p} c_i\|_p \le \|c\|_p, \tag{121}$$

for any knot sequence $\mathbf{t}$ and any coefficient sequence c, with

$$B_{i,k,p} := (k/(t_{i+k} - t_i))^{1/p} B_{i,k}. \tag{122}$$

good interpolation points

As a quick application of the Chebyshev spline, consider the question of 'good interpolation points'. By (112) Schoenberg-Whitney, we know that, for any choice of $x_1 < \cdots < x_n$ with $t_i < x_i < t_{i+k}$, there is, for given g, exactly one element from $S_{k,\mathbf{t}}$, denoted $P_{\mathbf{x}} g$, which agrees with g at the x_i. The smaller we can make $\|P_{\mathbf{x}}\|$, the better the following standard error estimate (see page 3) looks:

$$\|g - P_{\mathbf{x}} g\| \le (1 + \|P_{\mathbf{x}}\|) \operatorname{dist}(g, S_{k,\mathbf{t}}).$$

Let the norm be the max-norm, and let $C := (B_j(x_i) : i, j = 1, \ldots, n)$. Then $P_{\mathbf{x}} g = \sum_j a_j B_j$, with $Ca = (g(x_i) : i)$. Therefore $\|P_{\mathbf{x}} g\| \le \|a\|_\infty \le \|C^{-1}\|_\infty \|g\|$. Further, C is totally positive, therefore $\|C^{-1}\| = \|a^{\mathbf{x}}\|_\infty$ with $Ca^{\mathbf{x}} = ((-1)^{n-i} : i)$. Consequently, $\|P_{\mathbf{x}}\| \le \|a^{\mathbf{x}}\|_\infty$. By (116)Corollary, $\|a^{\mathbf{x}}\|_\infty$ is uniquely minimized when x is chosen to be the sequence of extreme points of $C_{k,\mathbf{t}}$ and, in that case, $\|a^{\mathbf{x}}\|_\infty = \kappa_\infty([B_j : j])$. This proves the following useful result.

(123) Proposition (Demko). *If Pg is the interpolant from $S_{k,\mathbf{t}}$ to g at the n extrema $y_1 < \cdots < y_n$ of $C_{k,\mathbf{t}}$, then*

$$\|Pg\|_\infty \le \kappa_\infty([B_{j,k,\mathbf{t}} : j]) \|g\|_\infty.$$

In particular, this shows that, for fixed k and for any knot sequence $\mathbf{t}$ (with at least $k+1$ entries), there exists a sequence $\mathbf{x}$ so that interpolation from $S_{k,\mathbf{t}}$ at $\mathbf{x}$ is uniquely possible and so that the resulting linear projector, $P_{\mathbf{x}}$, has its norm (in $C(\mathbb{R})$) bounded by d_k, i.e., bounded independently of $\mathbf{t}$.

Degree of approximation by splines

quasiinterpolants

It follows from (92)Theorem that the linear map

$$P : f \mapsto \sum_j \lambda_{j,k} f \, B_{j,k}$$

is a linear projector, with range $S = S_{k,\mathbf{t}} = \Pi^{\rho}_{k,\mathbf{t}}$, provided (as we have already assumed) that, for each i, we choose τ in (78) equal to some $\tau_i \in [t_i \,..\, t_{i+k})$. It is also local, since $\lambda_{j,k} f$ depends only on the behavior of f near τ_j, hence, by our choice of τ_j, only on the behavior of f on $\operatorname{supp} B_{j,k}$.

However, P is defined only for sufficiently smooth functions. In order to get such a projector on all of $L_1(I_{k,\mathbf{t}})$, we make use of (120) in order to obtain the linear functional

$$\mu_j f := \int h_j f$$

with $\operatorname{supp} h_j \subset [t_j \,..\, t_{j+k}]$ and $\|h_j\|_\infty \le d_{k,1}/(t_{j+k} - t_j)$ which, on $S_{k,\mathbf{t}}$, agrees with $\lambda_{j,k,\mathbf{t}}$. This implies that, for any essentially bounded function f,

$$|\mu_j f| \le d_{k,1} \|f\|_\infty([t_j \,..\, t_{j+k}]).$$

The corresponding linear map

$$Q : f \mapsto \sum_j \mu_j f \, B_{j,k}$$

is defined for any $f \in L_1(I_{k,\mathbf{t}})$, hence for any $f \in L_p(I_{k,\mathbf{t}})$. Further, it is the identity on its range, hence a linear projector, its norm is bounded by $d_{k,1}$, and it is **local**, in the sense that, for $x \in [t_l \,..\, t_{l+1})$,

$$
\begin{aligned}
Qf(x) = \sum_{j=l+1-k}^{l} \mu_j f \, B_j(x) &\le \sum_{j=l+1-k}^{l} |\mu_j f| \, B_j(x) \\
&\le d_{k,1} \|f\|_\infty([t_{l+1-k} \,..\, t_{l+k}]) \sum_{j=l+1-k}^{l} B_j(x) \\
&= d_{k,1} \|f\|_\infty([t_{l+1-k} \,..\, t_{l+k}]).
\end{aligned}
\tag{124}
$$

Since also $f - Qf = (1 - Q)(f - q)$ for any $q \in \Pi_{<k}$, we conclude that, for any f,

$$\|f - Qf\|_\infty([t_l \,..\, t_{l+1}]) \le d_{k,1} \operatorname{dist}(f, \Pi_{<k})_\infty (L_\infty[t_{l+1-k} \,..\, t_{l+k}]). \tag{125}$$

In fact, this conclusion can already be reached if we only take care that μ_j agree with λ_j on $\Pi_{<k}$, all j.

The resulting map Q is called a **good quasiinterpolant of order** k, with 'good' referring to the uniform localness expressed by (124), and 'order k' referring to the fact that Q reproduces $\Pi_{<k}$. The term 'quasiinterpolant' was chosen by finite-element people once they realized that approximation order could be ascertained with the aid of maps Q which did not actually interpolate at the 'nodal points' of their elements, but merely matched enough information to give reproduction of certain polynomial spaces.

The error bound (125) is local. In particular, from Jackson's inequality for approximation by algebraic polynomials (63) , we deduce, by dilation, that

$$\operatorname{dist}(f, \Pi_{<k})_\infty (L_\infty[t_{l+1-k} \mathinner{\ldotp\ldotp} t_{l+k}]) \le \operatorname{const}_r |t_{l+1-k} - t_{l+k}|^r \|D^r f\|_\infty ([t_{l+1-k} \mathinner{\ldotp\ldotp} t_{l+k}])$$

for any $r = 0, \ldots, k$. This suggests that, in approximating some f which is smooth except for some isolated singularities, the knot sequence $\mathbf{t}$ should be chosen so as make

$$l \mapsto |t_l - t_{l+1}|^k \|D^k f\|_\infty ([t_l \mathinner{\ldotp\ldotp} t_{l+1})$$

approximately constant. For functions such as $\sqrt{\cdot}$ on $[0 \mathinner{\ldotp\ldotp} 1]$, this strategy is entirely successful, giving approximations by splines of order k whose error behaves like $O(n^{-k})$, with n the degrees of freedom used. In contrast, the error in best polynomial approximation to $\sqrt{\cdot}$ on $[0 \mathinner{\ldotp\ldotp} 1]$ cannot be better than $O(n^{-1/2})$, by Bernshtein's Inverse Theorem.

The variational approach to splines

the fairy tale of the draughtman's spline

Schoenberg, in his fundamental 1946 paper, chose the word 'spline' to describe any pp function of order k in $C^{(k-2)}(\mathbb{R})$ with breakpoint sequence $\mathbb{Z} - \lfloor k/2 \rfloor$. Subsequently, he termed such functions **cardinal splines**; the beautiful theory of cardinal splines, mostly developed by Schoenberg, can be read up on in his monograph on the subject.

Schoenberg credits a certain G.D. Thomas from Ohio State University for making the connection between the graph of his cubic splines and the position assumed by a thin beam or 'spline' when it is constrained to go through certain points.

Explicitly, Schoenberg observes that classical thin-beam theory (going back to D. Bernoulli and Euler) identifies such a thin beam as the particular curve γ satisfying the given constraints which minimizes the **elastic strain energy** (or, **Bernoulli-Euler functional**)

$$J(\gamma) := \int_\gamma \kappa(s)^2 \, \mathrm{d}s,$$

with $s \mapsto \kappa(s)$ the **curvature** of the curve as a function of **arclength**; that the curvature of the curve $x \mapsto (x, f(x))$ formed by the graph of the smooth function f is

$$\kappa(x) = \frac{D^2 f(x)}{(1 + (Df)(x)^2)^{3/2}};$$

and that for a sufficiently 'flat' f, Df is essentially constant, hence can be ignored, leaving us, finally, to minimize

$$\int (D^2 f)^2,$$

whose minimizer is necessarily piecewise cubic and in $C^{(2)}$.

Schoenberg never followed up on this, but others have, to the point that this description is often given *without* the qualifying phrase 'assuming that Df is essentially constant' and is then used in textbooks to justify the use of cubic splines for data fitting of any sort.

However, G. Birkhoff pointed out the following in 1964. Suppose our curve γ is a circle. Then its curvature is constant, equal to $1/r$, with r the radius of the circle. Therefore,

$$J(\gamma) = \int_\gamma \kappa^2 = 2\pi r \, (1/r)^2 \quad = \quad 2\pi/r,$$

and this can be made arbitrarily small simply by choosing the circle sufficiently large. It follows that, with γ a smooth curve through given points $P_1, \ldots, P_n$ in the plane (in that order), we can make $J(\gamma)$ arbitrarily small by starting with the broken line interpolant to the given points, but making it a smooth curve by inserting, for each $j = 2, \ldots, n-1$, a huge loop which leaves P_j in the direction $P_j - P_{j-1}$ and ends again at P_j, but headed in the direction $P_{j+1} - P_j$. This raises the question why the minimization of the Bernoulli-Euler functional is so desirable.

It might be more desirable to minimize the modified functional

$$J_\sigma(\gamma) = \int_\gamma (\kappa^2 + \sigma),$$

as this penalizes long curves, the more the bigger the **tension parameter** σ is.

In fact, current efforts concerning nonlinear splines or elastica focus on the stationary points of J_σ (including the case $\sigma = 0$) as γ varies over the following set of curves γ.

All $\gamma \in C^{(1)}([0, L] \to \mathrm{I\!R}^2)$, with

- L variable;
- $\forall s \ \|\dot\gamma(s)\|_2 = 1$ (i.e., *arclength-parametrized*);
- $\gamma(s_i) = P_i$, all i, for some $0 \le s_1 < \cdots < s_n \le L$ (with $(P_i : i)$ a given *sequence* in $\mathrm{I\!R}^2$);
- possibly additional conditions (**pinned**, **clamped**, etc.).

To be sure, any such curve can be written in the form

$$\gamma(s) = \gamma(0) + \int_0^s \mathrm{e}^{i\theta(t)} \ \mathrm{d}t,$$

in which one interprets $\mathrm{I\!R}^2$ as the complex plane, hence $\theta(s)$ is the angle which the tangent, $\gamma(s) + \dot\gamma(s)\mathrm{I\!R}$, at $\gamma(s)$ makes with the line $(1,0)\mathrm{I\!R}$. For this reason, θ is called the **tangent indicatrix** for γ. It follows that

$$\kappa(s) = \dot\theta(s),$$

and this provides a standard way for recovering a curve from its initial point and initial tangent and its curvature.

Any critical point (if any!) of J has *continuous* curvature κ which, as a function of arclength, is piecewise an **elastica**, i.e., a solution of the ODE

$$\ddot\kappa + \kappa^3/2 = 0$$

hence expressible in terms of elliptic functions. Consequently, one can attempt to construct such a critical point by trying to piece together such elastica at the given points in such a way that a curve with continuous tangent and curvature results. However, the resulting system (in the parameters that characterize the individual elastica pieces) is nonlinear, hence one may fail to solve it simply because of an inadequate initial guess.

For this reason, Schoenberg's (implicit) suggestion to forget about the curvature and curves and be content to minimize $\int (D^2 f)^2$ over all functions which interpolate to given data is all the more welcome. However, the resulting interpolant only shares the name with the "real" spline.

Holladay

Apparently entirely unaware of Schoenberg's spline work, Holladay (a scientist at Los Alamos National Laboratory), while struggling with the numerical approximation of

$$\mu g := \int_a^b g$$

based on the given data

$$((x_i, g(x_i)) : i = 1, \ldots, n),$$

decided that a particularly good guess for μg could be had as $\mu \overset{\Omega}{g}$, with $f = \overset{\Omega}{g}$ the unique interpolant to the given data which minimizes

$$\|D^2 f\|_2^2 := \int_a^b (D^2 f)^2.$$

(126) Theorem (Holladay). *Given $((x_i, y_i) : i = 1, \ldots, n)$ with $a \le x_1 < \cdots < x_n \le a$, among all $f \in C^{(2)}[a\,..\,b]$ for which $f(x_i) = y_i$, all i, the seminorm $\|D^2 f\|_2$ is uniquely minimized by the* **natural interpolating cubic spline**, *i.e., by the unique cubic pp function f in $C^{(2)}[a..b]$ with (interior) breakpoints $x_1, \ldots, x_n$, with $f(x_i) = y_i, i = 1, \ldots, n$, and with*

$$D^2 f(x_1) = D^3 f(x_1-) = \ 0 \ = D^2 f(x_n) = D^3 f(x_n+). \tag{127}$$

To be sure, the terminology 'natural' was introduced much later, by Schoenberg, to reflect the fact that the natural endconditions (127) of the variational problem are used here, and to distinguish it from **complete cubic spline interpolation** in which the additional data $Df(x_1) = s_1$ and $Df(x_n) = s_n$ are provided, hence the resulting smoothest interpolant satisfies these endconditions (instead of the 'natural' ones). The latter approximates to a smooth function much better than the former, since the former enforces the artificial condition that $D^2 f(x_1) = 0 = D^2 f(x_n)$.

I prefer to discuss the proof of Holladay's theorem in the much more general (related) context of best estimates for μg, given some information about g, as developed by Golomb and Weinberger.

the GW interval

The problem considered in Golomb and Weinberger's fundamental 1959 paper is the following:

With μ some linear functional on the linear space X, what can be said about

$$\mu g$$

*given that $\Lambda g = \alpha$ for some linear $\Lambda : X \to \mathrm{I\!R}^n$ (**data map**) and some known α ?*

The answer to this question turns out to be trivial, hence uninformative: The set sought for is $\mu(\Lambda^{-1}\{\alpha\})$, with $\Lambda^{-1}\{\alpha\}$ the **flat**

$$\Lambda^{-1}\{\alpha\} = \{f \in X : \Lambda f = \alpha\} = g + \ker \Lambda.$$

If $\ker \mu \supset \ker \Lambda$, then μ is constant on $g + \ker \Lambda$, and the sought-for set consists of that constant value. In other words, we know μ exactly. Since this case will occur iff μ is a linear combination of the **row**s of Λ, i.e., iff

$$\mu \in \operatorname{ran}[\lambda_1, \ldots, \lambda_n]$$

with

$$\Lambda : f \mapsto (\lambda_i f : i = 1, \ldots, n),$$

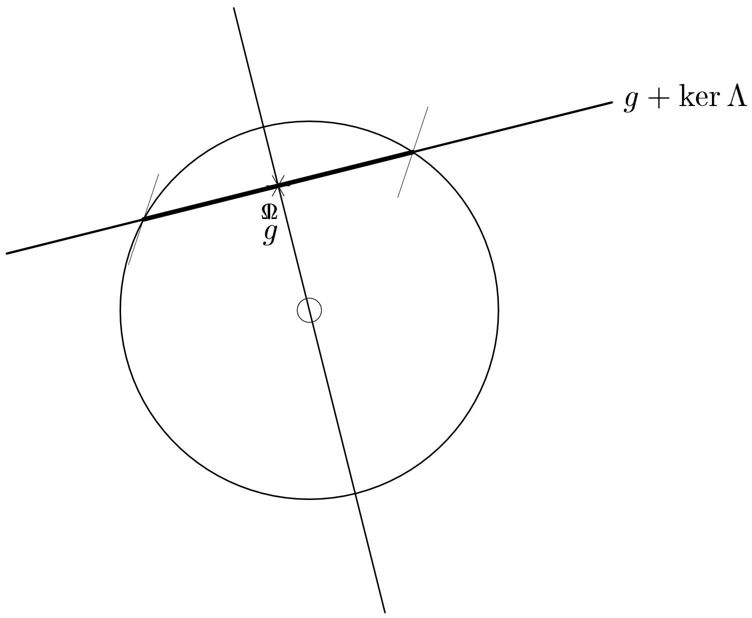
we are not surprised by it. In the contrary case, μ can take any value, i.e., the sought-for set is all of $\mathbb{R}$.

For this reason, Golomb and Weinberger actually consider the following refined problem:

With μ some continuous linear functional on the normed linear space X, what can be said about

$$\mu g$$

*given that $\Lambda g = \alpha$ for some linear $\Lambda : X \to \mathbb{R}^n$ (**data map**) and some known α, and $\|g\| \le r$ for some known r?*



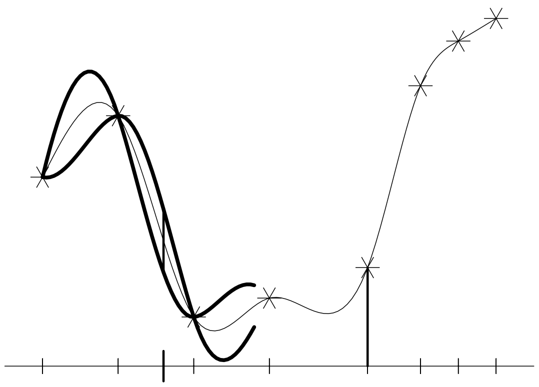
(128) Figure The element g is known to lie on the flat $g + \ker \Lambda$ and inside the ball; if X is a Hilbert space, then $*$ marks the location of $\overset{\Omega}{g}$, at the intersection of the flat with $(\ker \Lambda)^\perp$.

The answer is simple (see (128)Figure): the possible values of μg make up the set

$$\mu((g + \ker \Lambda) \cap B_r^-)$$

and this, as the image of a convex set under a linear functional, is necessarily an interval; however, it may, or may not, be closed. It is, by definition, the **GW-interval** for the given information.

This suggests the following construction of a smoothest interpolant to given data $((x_i, y_i) : i = 1, \ldots, n)$: For each $t \in [a \,..\, b]$, let $f(t)$ be the center of the GW-interval for the linear functional $\mu = \delta_t$, as computed from the conditions that $\Lambda g := (g(x_i) : i = 1, \ldots, n) = (y_i : i = 1, \ldots, n)$ and that $\|g\| \le r$ (in some suitable norm).



(129) Figure Some data, part of the envelope provided by the GW intervals, and the interpolant formed by the centers.

(129)Figure shows, for many t, the resulting GW-interval corresponding to $\mu = \delta_t$ and, for all t, the resulting function f; the norm used here is the following:

(130)
$$\|g\|^2 := g(x_1)^2 + g(x_2)^2 + \int_a^b (D^2 g(t))^2 \, dt.$$

The fact that the 'middle-function' is smooth is not at all an accident, but is the result of the fact that

$$f(t) = \delta_t \overset{\Omega}{g}, \quad a \le t \le b$$

for some fixed $C^{(2)}$-function $\overset{\Omega}{g}$.

The reason for this is already suggested in (128)Figure: If X is an inner product space, then the 'disk' $(g + \ker \Lambda) \cap B_r^-$ has at its center the unique element, $\overset{\Omega}{g}$, in the flat $g + \ker \Lambda$ of minimum norm and, for any μ, the number $\mu \overset{\Omega}{g}$ is therefore the center of the corresponding GW-interval. Here is the verification of these assertions.

The "best" interpolant

The **"best" interpolant** to the given data Λg is, by definition, the element of smallest norm in the flat $g + \ker\Lambda$. Any such element $\overset{\mathfrak{S}}{g}$ is a b.a. to 0 from that flat. Since the flat is, in particular, convex, it follows that $\overset{\mathfrak{S}}{g}$ is uniquely determined provided X is strictly convex, as is the case when X is an inner product space, and is not the case if X is, e.g., $C(T)$ or $L_1(T)$. In any case, the flat is also closed. Hence, we know from (22)Theorem that $\overset{\mathfrak{S}}{g}$ is characterized by the fact that (it is an element of the flat and) there is some $\lambda \| (-\overset{\mathfrak{S}}{g})$ with $\sup \lambda(g + \ker\Lambda) \le \lambda \overset{\mathfrak{S}}{g}$. Since $\lambda(g+\ker\Lambda) = \lambda g + \lambda(\ker\Lambda)$, the last condition holds iff $\lambda(\ker\Lambda) = \{0\}$, i.e., iff

$$\ker\lambda \supset \ker\Lambda = \cap_i \ker\lambda_i,$$

with $\lambda_1, \ldots, \lambda_n$ the rows of the data map Λ, and this, in turn, is well known to be equivalent to having

$$\lambda \in \operatorname{ran}[\lambda_1, \ldots, \lambda_n].$$

In the nicest case (true in the situation of Holladay's Theorem), X is a Hilbert space, hence the rows of Λ are of the form

$$\lambda_i = \langle f_i, \cdot\rangle, \quad i = 1, \ldots, n$$

for some elements f_i of X, and so, since $-\overset{\mathfrak{S}}{g}$ is to be parallel to some element λ of $\operatorname{ran}[\lambda_1, \ldots, \lambda_n]$, it follows that $\overset{\mathfrak{S}}{g}$ is necessarily in

$$F := \operatorname{ran}[f_1, \ldots, f_n] = (\ker\Lambda)^\perp,$$

hence is the unique element in $(g + \ker\Lambda) \cap (\ker\Lambda)^\perp$. This identifies $\overset{\mathfrak{S}}{g}$ as an element with many different interesting properties which, in the spline literature, have been discovered singly and repeatedly.

(131) Theorem. *Let X be a real Hilbert space, and let $X \to \mathbb{R}^n : \Lambda : x \mapsto (\langle f_i, x\rangle : i = 1, \ldots, n)$ be a given data map. For given $g \in X$, there is an element $\overset{\mathfrak{S}}{g}$ in X having all of the following properties (each of which determines it uniquely):*

(i) **smoothest***: $\overset{\mathfrak{S}}{g}$ uniquely minimizes $\|\cdot\|$ on $g + \ker\Lambda$;*

(ii) **best approximant***: $\overset{\mathfrak{S}}{g}$ is the unique b.a. to g from $F := \operatorname{ran}[f_1, \ldots, f_n]$;*

(iii) **interpolant***: $\overset{\mathfrak{S}}{g}$ is the unique element in F which agrees with g "on" Λ.*

(iv) **center***: For any $\mu \in X^*$, $\mu \overset{\mathfrak{S}}{g}$ is the center of the GW interval $\mu((g + \ker\Lambda) \cap B_r^-)$ for μ.*

Proof: (i) is the definition of $\overset{\mathfrak{S}}{g}$. As a consequence, we get (iii), namely the statement that $\overset{\mathfrak{S}}{g}$ is the unique element in $(g + \ker\Lambda) \cap (\ker\Lambda)^\perp$, and so, $(g - \overset{\mathfrak{S}}{g}) \perp (\ker\Lambda)^\perp = F \ni \overset{\mathfrak{S}}{g}$, which implies (ii). This also implies that

$$r^2 \ge \|g\|^2 = \|g - \overset{\mathfrak{S}}{g}\|^2 + \|\overset{\mathfrak{S}}{g}\|^2,$$

hence the disk $B_r^- \cap (g + \ker\Lambda)$ equals $\overset{\Omega}{g} + B_{r'}^- \cap \ker\Lambda$, with

$$r' := \sqrt{r^2 - \|\overset{\Omega}{g}\|^2}.$$

Consequently, the GW-interval is

$$\mu\overset{\Omega}{g} + \mu(B_{r'}^- \cap \ker\Lambda) = \mu\overset{\Omega}{g} + [-1\,..\,1]\, r'\, \|\mu|_{\ker\Lambda}\|,$$

and this implies (iv). □

Holladay once more

When looking for the "smoothest' interpolant to given data $\Lambda g := (g(x_i) : i = 1, \ldots, n)$ in Holladay's context (with $a \le x_1 < \cdots < x_n \le b$), we are seeking, for $m = 2$,

$$\overset{\Omega}{g} := \operatorname{argmin}\{\|D^m f\|_2 : f \in X, \Lambda f = \Lambda g\}$$

with

$$\|h\|_2^2 := \int_a^b h^2$$

and

$$X = L_2^{(m)}[a\,..\,b] := \{f \in C^{(m-1)}[a\,..\,b] : D^{m-1}f \text{ abs.cont.},\ D^m f \in L_2[a\,..\,b]\}.$$

However,

$$f \mapsto \|D^m f\|_2$$

is only a *semi*norm. There are two options. In one, one augments the seminorm by the addition of another seminorm which is (a) constant on flats parallel to $\ker\Lambda$; and (b) makes the resulting sum a norm, i.e., definite (see, e.g., (130)). The other option was apparently first used by Favard in his 1940 paper on smoothest interpolation in which the max-norm (rather than the 2-norm) of $D^m f$ is to be minimized (of which more anon).

In Favard's approach, one expresses the given constraints, $\Lambda f = \Lambda g$, on the interpolant f explicitly as constraints on the function $h := D^m f$ whose norm is to be minimized.

Since we are considering the special case $\Lambda : f \mapsto (f(x_i) : i = 1, \ldots, n)$, we can derive from Λf the mth divided differences

$$\delta_{x_i,\ldots,x_{i+m}} f = \int M(\cdot|x_i, \ldots, x_{i+m}) D^m f/m!, \quad i = 1, \ldots, n - m,$$

with

$$M_i := M(\cdot|x_i, \ldots, x_{i+m}) := \frac{m}{x_{i+m} - x_i} B_{i,m,\mathbf{x}}$$

the ith B-spline for the knot sequence $\mathbf{x} = (x_i : i = 1, \ldots, n)$ normalized appropriately for the 1-norm, i.e., so that

$$(132) \qquad \int M(\cdot|x_i, \ldots, x_{i+m}) \;=\; 1.$$

In effect, on the space

$$Y := L_2[a \,..\, b],$$

we are given the data map

$$(133) \qquad \mathrm{M} : Y \to \mathbb{R}^{n-m} : h \mapsto (\langle M_i, h\rangle : i = 1, \ldots, n-m)$$

and are seeking $h = D^m f$ of minimal norm among all those for which

$$\mathrm{M}h = \mathrm{M}D^m g = (\delta_{x_i,\ldots,x_{i+m}} g : i = 1, \ldots, n-m).$$

By (131)Theorem, this minimization problem has exactly one solution, namely the unique element $\overset{\Omega}{h}$ in $(D^m g + \ker \mathrm{M}) \cap (\ker \mathrm{M})^\perp$. Also by the earlier discussion (or directly),

$$\ker \mathrm{M} = \mathrm{ran}[M_i : i = 1, \ldots, n-m] = S_{m,\mathbf{x}}.$$

In other words, $\overset{\Omega}{h}$ is a spline of order m with knot sequence $\mathbf{x}$. Finally, if now f is any element in X with $D^m f = \overset{\Omega}{h}$ and p is the unique polynomial of degree $< m$ which matches $g - f$ at $(x_1, \ldots, x_m)$, then (i) $D^m(f + p) = D^m f = \overset{\Omega}{h}$, and (ii) $(f + p)(x_i) = g(x_i), i = 1, \ldots, m$, and (iii) $\delta_{x_i,\ldots,x_{i+m}}(f + p) = \delta_{x_i,\ldots,x_{i+m}} f = \delta_{x_i,\ldots,x_{i+m}} g$, $i = 1, \ldots, n - m$. This implies, by induction on j, that $(f + p)(x_i) = g(x_i)$, $i = 1, \ldots, j$, for $j = m, \ldots, n$. Consequently, $f + p = \overset{\Omega}{g}$. – This finishes the proof of the natural generalization of (126)Theorem. However, it proves a bit more since we may also treat in exactly the same way the limiting situation when there are coincidences among the x_i and, corresponding to our earlier definition of the divided difference $\delta_{x_i,\ldots,x_{i+m}}$, we are considering also the matching of certain consecutive derivatives, according to the following recipe.

One says that f **vanishes on the** *sequence* $\mathbf{x} = (x_j : j)$ and writes

$$f|_{\mathbf{x}} = 0$$

in case

$$0 \le i < \#_{\mathbf{x}} z \quad \Longrightarrow \quad D^i f(z) = 0,$$

with

$$\#_{\mathbf{x}} z := \#\{j : z = x_j\}.$$

Correspondingly, one says that f **interpolates to** g **on** $\mathbf{x}$ in case $(f - g)|_{\mathbf{x}} = 0$, and speaks of **osculatory** or **Hermite** interpolation in case $\#_{\mathbf{x}} z > 1$ for some z.

If our sequence $\mathbf{x}$ contains no point with multiplicity greater than m, then the data map M (defined in (133)) is well defined on $L_2[a\,..\,b]$, and the above argument goes through otherwise unchanged. In particular, if

$$x_1 = \cdots = x_m = a, \quad x_{n-m+1} = \cdots = x_n = b,$$

then we are, in effect, prescribing $D^r g$, $r = 0, \ldots, m-1$, at both a and b. At the same time, $S_{m,\mathbf{x}}$ is a **complete** spline space, i.e., it equals the space $\Pi^\rho_{<m,\xi}$ of all pp of order m on $[a\,..\,b]$ with ξ the strictly increasing sequence of distinct points in $\mathbf{x}$, and $\rho_i = m - 1 - \#_{\mathbf{x}}\xi_i$, all i. We have proved:

(134) Theorem. *Let* $\mathbf{x} = (x_i : i = 1, \ldots, n)$ *be any nondecreasing sequence in* $[a \mathinner{..} b]$ *with* $\#_{\mathbf{x}} z \le m$ *for any* z*. The unique interpolant* f *of smallest seminorm* $\|D^m f\|_2$ *to given* $g \in L_2^{(m)}[a \mathinner{..} b]$ *on* $\mathbf{x}$ *is characterized by the fact that* $D^m f \in S_{m,\mathbf{x}}$*.*

In particular, this 'smoothest' interpolant is pp of order $2m$, with (interior) breakpoint sequence $\mathbf{x}$, and reduces to a polynomial of order m on $[a \mathinner{..} x_1)$ and on $(x_n \mathinner{..} b]$.

Incidentally, the argument for (112)Theorem can be extended straightforwardly to cover also the case of Hermite interpolation, as follows.

(135) Theorem. *Let* $\mathbf{x} = (x_1 \le \ldots \le x_n)$ *and let* $\mathbf{t} = (t_j : j = 1, \ldots, n+k)$ *be a knot sequence and assume that no point appears in the combined sequence* $(\mathbf{x}, \mathbf{t})$ *more than* $k-1$ *times. Then, for every smooth* g*, there exists exactly one* $f \in S_{k,\mathbf{t}}$ *for which* $f - g$ *vanishes on* $\mathbf{x}$ *if and only if* $t_i < x_i < t_{i+k}$*, all* i*.*

Krein

When $\|D^m f\|_p$ for $p \ne 2$ is to be minimized, one uses an argument that goes back to M.G. Krein (1938). Krein observed (in a different context) that, for $1 < p \le \infty$ and with $1/p + 1/p^* = 1$, any $h \in L_p[a \mathinner{..} b]$ with $\mathrm{M}h = \mathrm{M}D^m g$ provides an extension to all of $L_{p^*}[a \mathinner{..} b]$, of the linear functional

$$(136) \qquad \lambda_0 : S_{m,\mathbf{x}} \to \mathbb{R} : \sum_i B_{i,m,\mathbf{x}} a_i \mapsto (m-1)! \sum_i \bigl(\delta_{x_{i+1},\ldots,x_{i+m}} g - \delta_{x_i,\ldots,x_{i+m-1}} g\bigr) a_i.$$

Here, the following calculation was used:

$$\int B_{i,m,\mathbf{x}}\, D^m f = m! \int \frac{x_{i+m} - x_i}{m} M_i\, D^m f / m! = (m-1)! \bigl(\delta_{x_{i+1},\ldots,x_{i+m}} - \delta_{x_i,\ldots,x_{i+m-1}}\bigr) f.$$

The linear functional λ_0 is well-defined since $(B_{i,m,\mathbf{x}} : i = 1, \ldots, n-m)$ is a basis for $S_{m,\mathbf{x}}$. (Also, the information about g required is available for any $g \in L_p^{(m)}$ since it does not go beyond derivatives of order $< m$.) Further, $\|h\|_p$ is the norm of the linear functional $f \mapsto \int_a^b hf$ on L_{p^*}, hence must be at least as big as $\|\lambda_0\|$. Therefore, $\|h\|_p$ is minimized by any norm-preserving extension $\int h\cdot$ of λ_0 to all of $L_{p^*}[a \mathinner{..} b]$ and, by Hahn-Banach, there is at least one such.

Any such norm-preserving extension can, in principle, be located as follows: Since $S_{m,\mathbf{x}}$ is finite-dimensional, it contains an **extremal** for λ_0, i.e., some element ψ with $\lambda_0 \| \psi$. This implies that, for any norm-preserving extension $\int h\cdot$ of λ_0, there must be equality in Hölder's Inequality

$$\|h\|_p \|\psi\|_{p^*} = \lambda_0 \psi = \int h\psi \le \|h\|_p \|\psi\|_{p^*}.$$

For $1 < p < \infty$, this determines h uniquely. In particular,

$$(137) \qquad D^m \overset{\boldsymbol{\Omega}}{g} \;=\; \overset{\boldsymbol{\Omega}}{h} \;=\; \mathrm{const} |\psi|^{p^*-1}\, \mathrm{signum}\, \psi.$$

For $p = 2$, we recover the earlier result since then $p^* = 2$, hence $\overset{\boldsymbol{\Omega}}{h} = \mathrm{const}\, \psi$.

Favard

For $p = \infty$, (137) becomes

$$(138), \qquad D^m \overset{\boldsymbol{\Omega}}{g} \;=\; \overset{\boldsymbol{\Omega}}{h} \;=\; \| \overset{\boldsymbol{\Omega}}{h} \|_\infty \operatorname{signum} \psi$$

and this determines $\overset{\boldsymbol{\Omega}}{h}$ only on the support of any extremal ψ for λ_0. On

$$Z(\psi) := \psi^{-1}\{0\},$$

$\overset{\boldsymbol{\Omega}}{h}$ is only constrained to be $\le \|\lambda_0\|$ in absolute value (in addition to the conditions that $\int M_i \overset{\boldsymbol{\Omega}}{h} = \int M_i D^m g$, all i). In particular, a smoothest interpolant $f = \overset{\boldsymbol{\Omega}}{g}$ in the sense that it minimizes $\|D^m f\|_\infty$ is, in general, not unique. (See (148)Figure.)

As a simple example, if $m = 2$ and $\mathbf{x} = (1, 2, 3, 4, 5)$, and $\int M_i D^2 g = i$, $i = 1, 2, 3$, then

$$\|\lambda_0\| = \sup_a |\sum_{i=1}^{3} a_i i| / \| \sum_i a_i M_i \|_1 = 3,$$

since $\| \sum_i a_i M_i \|_1 \le \| \sum_i |a_i| M_i \|_1 = \|a\|_1$ (as $\int |M_i| = \int M_i = 1$ by (132)), with equality if, e.g., all the a_i are nonnegative, while $\sum_i a_i i| \le 3\|a\|_1$, with equality if $a = (0, 0, 1)$. In particular, $\psi = M_3$ is an extremal, and its support is $[3\,..\,5]$, hence $\overset{\boldsymbol{\Omega}}{h}$ is only pinned down on that interval, and there it has to be the constant 3.

Favard deals with this in a constructive way. He adds the additional requirement that $\overset{\boldsymbol{\Omega}}{h}$ should also minimize $\|h\|_\infty(Z(\psi))$. In the example, this gives the additional condition that $\overset{\boldsymbol{\Omega}}{h} = 1$ on $[1\,..\,3]$, hence pins it down completely. In the general case, there may be still some uncertainty, and Favard deals with it in the same way. Such a process is guaranteed to terminate in finitely many steps since, at each step, the set of uncertainty is reduced by at least one knot interval. Moreover, at each step, if there is still some set of uncertainty, it is certain to consist of one or more disjoint intervals with the property that any $B_{i,m,\mathbf{x}}$ has support in at most one of these, hence, at any stage, one is solving nothing more complicated than a problem like the original one.

(139) Theorem. *Favard's "best" interpolant $\overset{\boldsymbol{\Omega}}{g}$ to given data $(x_i, y_i) : i = 1, \ldots, n)$ with $a \le x_1 < \cdots < x_n \le b$, is the unique element, in the set G of all interpolants from $L_\infty[a\,..\,b]$ to the given data, with the following two properties:*

(i) $|D^m \overset{\boldsymbol{\Omega}}{g}\,|$ has finite (essential) range, $\{c_1, \ldots, c_r\}$ say, with $c_1 > \cdots > c_r$ and $r \le n - m$.

(ii) With $\Omega_i := \{x \in [a\,..\,b] : |D^m \overset{\boldsymbol{\Omega}}{g}\,| \le c_i\}$, $i = 1, \ldots, r$, and $\Omega_{r+1} := \{\}$, $D^m \overset{\boldsymbol{\Omega}}{g}$ uniquely minimizes $|D^m g|$ on $\Omega_1 \backslash \Omega_{i+1}$, i.e., $g \in G$ and $|D^m g| \le |D^m \overset{\boldsymbol{\Omega}}{g}\,|$ pointwise on $\Omega_1 \backslash \Omega_{i+1}$ implies that $g = \overset{\boldsymbol{\Omega}}{g}$ on $\Omega_i \backslash \Omega_{i+1}$, for $i = 1, \ldots, r$.

Favard's "best" interpolant can be shown to have the following additional properties:

(i) Its mth derivative vanishes outside $[x_1 \,..\, x_n]$ and has fewer than $n-m$ jumps inside $(x_1 \,..\, x_n)$.
(ii) The absolute value of its mth derivative has jumps only at points from $\mathbf{x}$.
(iii) (Chui, Smith, Ward (TAMS, 1976)) It is the L_1-limit, as $p \to \infty$, of the unique interpolant f which minimizes $\|D^m f\|_p$.

perfect splines

The term **perfect** spline was given by Glaeser (JAT, 1973) to any spline whose piecewise constant derivative is absolutely constant.

(140)Theorem (Karlin, 1973). *For given nondecreasing* $\mathbf{x}$ *in* $[a \,..\, b]$ *with no multiplicity greater than* m *and given* $g \in L_\infty^{(m)}[a \,..\, b]$, *the set* $G := \{f \in L_\infty^{(m)}[a \,..\, b] : f|_{\mathbf{x}} = g|_{\mathbf{x}}\}$ *contains a perfect spline of degree* m *with fewer than* $n-m$ *interior knots. Further, any such perfect spline minimizes* $f \mapsto \|D^m f\|$ *over* G.

In view of Favard's 'best' interpolant, this perfect interpolating spline seems less than perfect. However, the theorem makes it easy to characterize the GW intervals for any $\mu = \delta_t$, and provides other sharp estimates (see below).

Proof: We start with a proof of the existence of such a perfect spline.

If λ_0 (see (136)) has an extremal ψ with $Z(\psi)$ of measure zero, then the interpolant f with minimal $\|D^m f\|_\infty$ is uniquely determined; its mth derivative is, up to a constant multiple, equal to the signature of that spline, ψ, hence cannot jump more than $n-m-1$ times.

If we knew that $S_{m,\mathbf{x}}$ is Haar (as a subspace of $C[a \,..\, b]$), then we would know that ψ, being nontrivial, could only vanish at $< \dim S_{m,\mathbf{x}}$ points, hence $Z(\psi)$ could only be a finite set. Unfortunately, in general, elements of $S_{m,\mathbf{x}}$ can vanish on entire intervals, yet not be zero. The problem is that $\det(M_j(\tau_i) : i,j = 1,..n-m)$ can be zero even if $a < \tau_1 < \cdots < \tau_{n-m} < b$. To deal with this problem, one simply applies a little heat, as follows. One uses some integral operators

$$T_\varepsilon : f \mapsto \int a_\varepsilon(\cdot, s) f(s)\, \mathrm{d}s,$$

with $T_\varepsilon \xrightarrow[\varepsilon \to 0]{} \mathrm{id}$ on $C[a \,..\, b]$ with the kernel, a_ε, **strictly totally positive**. This means that, for all strictly increasing sequences $\mathbf{i}$ and $\mathbf{j}$ with $\#\mathbf{i} = \#\mathbf{j}$,

$$\det a_\varepsilon(\mathbf{i}, \mathbf{j}) > 0.$$

A standard example is the heat operator,

$$T_\varepsilon f := \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-(\cdot - s)^2/2\varepsilon^2}}{\varepsilon\sqrt{2\pi}} f(s)\, \mathrm{d}s.$$

With such T_ε in hand, one can find nearby spaces

$$S_\varepsilon := \mathrm{ran}[T_\varepsilon M_j : j = 1, \ldots, n-m]$$

whose corresponding collocation matrices are *strictly* totally positive, i.e., satisfy

$$\det(T_\varepsilon M_j(\tau_i) : i,j = 1, \ldots, n-m) > 0$$

for any $\tau_1 < \cdots < \tau_{n-m}$. This is so because of

(141)Cauchy-Binet. *For any two matrices A and B for which AB is defined, and for any applicable strictly increasing index sequences I and J of the same length,*

$$\det(AB)(I,J) = \sum_K \det A(I,K) \det B(K,J)$$

with the sum over all applicable index sequences K.

The application here even deals with an A with infinitely many columns and a B with correspondingly infinitely many rows:

$$\det\big((T_\varepsilon M_j)(\tau_i)\big) = \int_{k_1<\cdots<k_{n-m}} \det a_\varepsilon(\tau,\mathbf{k}) \det\big((M_j)(k_i)\big)\ \mathrm{d}\mathbf{k},$$

and this is always positive since, by assumption, $\det a_\varepsilon(\tau,\mathbf{k})$ is positive for all *strictly* increasing $\mathbf{k}$, while $\det\big((M_j)(k_i)\big)$ is nonnegative for all such $\mathbf{k}$ (by (114)Theorem), and positive for any $\mathbf{k}$ with $x_i < k_i < x_{i+m}$, all i, and the set of such $\mathbf{k}$ has positive $(n-m)$-dimensional measure.

It follows that, for any $\varepsilon > 0$, we get an absolutely constant h_ε with $S^-(h_\varepsilon) < n-m$ and

$$\lambda_{0,\varepsilon} h_\varepsilon = \|\lambda_{0,\varepsilon}\| \|h_\varepsilon\|_\infty \xrightarrow[\varepsilon\to 0]{} \|\lambda_0\| \|h\|,$$

with h any limit point of $(h_\varepsilon : \varepsilon > 0)$, hence h is exactly what we want.

This settles existence of a perfect spline in G with $< n-m$ interior knots. Let now s be any such. Any $f \in G$ with $\|D^m f\|_\infty < \|D^m s\|$ would lead to an $h := D^m(f-s) \perp S_{m,\mathbf{x}}$ with fewer than $n-m$ sign changes, and this would contradict (142)Lemma below. □

(142) Lemma. *If $h \in L_\infty[a\,..\,b]$ is orthogonal to $S_{m,\mathbf{x}}$ and there are points $a =: \zeta_0 < \cdots < \zeta_{r+1} := b$ and some $\sigma \in \{-1,1\}^{r+1}$ with $\sigma_i h > 0$ on $(\zeta_{i-1}\,..\,\zeta_i)$, all i, then $r \ge n-m$, with equality only if*

(143) $$x_i < \zeta_i < x_{i+m}, \qquad \forall\{i\}.$$

Proof: WLOG (see page 77), $h = D^m s$, with $s|_{\mathbf{x}} = 0$. By Rolle's Theorem, this implies that $D^{m-1}s|_\tau = 0$ for some $\tau = (\tau_1 < \cdots < \tau_{n-m+1})$ with

(144) $$x_i \le \tau_i \le x_{i+m-1}, \qquad \forall\{i\}.$$

By assumption, $D^{m-1}s = \int h$ is strictly monotone on each of the intervals $[\zeta_{i-1}\,..\,\zeta_i]$, hence any such interval can contain at most one of the τ_j. This implies that there must be at least as many such intervals as there are τ_j, i.e., $r+1 \ge n-m+1$, and equality is possible only if there is exactly one such τ_j in each such interval, i.e., $\zeta_{i-1} < \tau_i < \zeta_i$, $i = 1,\ldots,n-m+1 = r+1$, and, in conjunction with (144), this implies (143). □

(145)Remark Let f be a best L_p-approximation from $S_{m,\mathbf{x}}$ to some $g \in L_p[a \,.\,. b]$, and let $1 \le p < \infty$. By (24)Theorem, and by the fact that $L^* = L_{p^*}$ (with $1/p+1/p^* = 1$), it follows that

$$h := |g - f|^{p^*-1} \operatorname{signum}(g - f)$$

is perpendicular to $S_{m,\mathbf{x}}$. In particular, with a slight strengthening of the lemma, it follows that signum$(g - f)$ must have at least $n - m$ strong sign changes. If f is continuous, this implies that g must interpolate to f at $n - m$ points. Further, if signum$(g - f)$ has only $n-m$ sign changes, at $\tau_1 < \cdots < \tau_{n-m}$ say, then these points of interpolation must satisfy the Schoenberg-Whitney conditions (143). □

(146)Remark The proof of this lemma does not make full use of all the assumptions made (implicitly) about $D^{m-1}s = \int h$. To be sure, it uses the strict monotonicity of $D^{m-1}s = \int h$ on the intervals $[\zeta_{i-1} \,.\,. \zeta_i]$. It also uses the fact that $D^{m-2}s$ is the integral of $D^{m-1}s$, hence continuous. But the continuity of $D^{m-1}s$ itself is used only to infer the existence of the zeros τ_j. If we agree to call z a zero of the piecewise continuous function f in case $f(z-)f(z+) \le 0$, then Rolle's theorem provides the zeros $\tau_1 < \cdots < \tau_{n-m+1}$ of $D^{m-1}s$ in $[a \,.\,. b]$ even when $D^{m-1}s$ is only piecewise continuous and, correspondingly h has finitely many point masses. We do lose the conclusion (143) in case $r = n - m$, but the rest of the lemma is valid in this more general setting (with the understanding that the condition $\sigma_i h > 0$ on $(\zeta_{i-1} \,.\,. \zeta_i)$ means that $\sigma_i m_t \delta_t > 0$ for any point mass in h at t with $\zeta_{i-1} < t < \zeta_i$).

In consequence, we have the following strengthening of Karlin's Theorem which involves the space

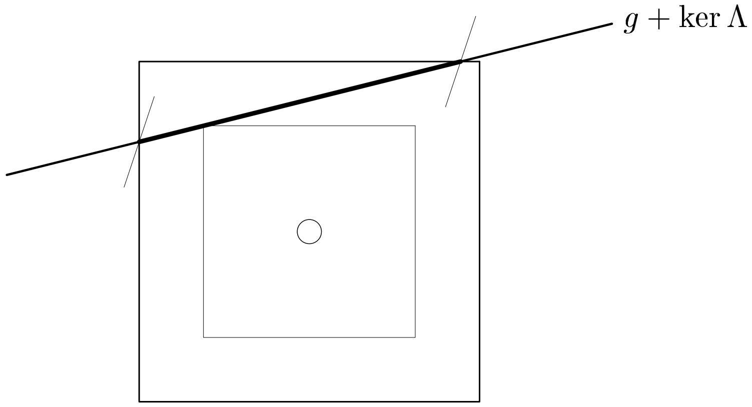
$$L^{(m)}_{\infty,\zeta}[a \,.\,. b] := L^{(m-1)}_\infty \cap \bigcap_j L^{(m)}_\infty(\zeta_{j-1} \,.\,. \zeta_j),$$

with $a = \zeta_0 < \cdots < \zeta_{r+1} = b$ for some ζ. For any f in this space, we define

$$\|D^m f\|_\infty := \max_j \|D^m f\|_\infty(\zeta_{j-1} \,.\,. \zeta_j).$$

(147)Proposition. *Let* $\mathbf{x}$ *be nondecreasing in* $[a \,.\,. b]$ *with no multiplicity greater than* m*. Then, for any perfect spline* s *of degree* m *with no more than* $r < n - m$ *interior knots,* $\zeta_1 < \cdots < \zeta_r$ *say, none in* $\mathbf{x}$*, and for any* $f \in L^{(m)}_{\infty,\zeta}$*,* $f = s$ *on* $\mathbf{x}$ *implies that* $\|D^m f\|_\infty \le \|D^m s\|\infty$.

Proof: If we had $\|D^m f\|_\infty < \|D^m s\|_\infty$, then $D^{m-1}(s - f)$ would be strictly monotone on each of the intervals $(\zeta_{i-1} \,.\,. \zeta_i)$, yet would vanish, in the extended sense of having left and right limit weakly of opposite sign, at $n - m + 1$ distinct points, a contradiction. □



(148) Figure The element g is known to lie on the flat $g+\ker\Lambda$ and inside the 'ball'; since this 'ball' has edges, there may be many smallest elements in the flat, and none need provide the center for a GW-interval.

optimal recovery

Optimal recovery concerns the limits on the accuracy with which one can recover or reconstruct a function g from the information Λg about it. Here is a particular instance, due to Micchelli, Rivlin, Winograd (1973).

We continue to consider the data map $\Lambda : f \mapsto f|_{\mathbf{x}}$, with $\mathbf{x} = (x_1, \ldots, x_n)$ nondecreasing, in $[a\,..\,b]$, and with all multiplicities $\le m$. We consider $T : \mathbb{R}^n \to X$ which **recover** f from Λf in the sense that $\Lambda T = \mathrm{id}$. No other assumption is made about T; in particular, T needn't even be linear. We measure the error $f - T(f|_{\mathbf{x}})$ in the max-norm on $[a\,..\,b]$, and hope to bound it in terms of $\|D^m f\|_\infty$, thus restricting attention to $f \in L_\infty^{(m)}[a\,..\,b]$. Whatever T may be, we have the trivial bound

$$\|f - T(f|_{\mathbf{x}})\| \le \mathrm{const}_T \|D^m f\|,$$

with

$$\mathrm{const}_T := \sup_{f\in X} \|f - T(f|_{\mathbf{x}})\| / \|D^m f\|.$$

In particular, const_T could be infinite. One calls T an **optimal recovery** scheme (for the given Λ and the given norm-pair) if const_T is as small as possible.

There is a simple lower bound for const_T:

$$(149)\qquad \mathrm{const}_T = \sup_f \frac{\|f - T(f|_{\mathbf{x}})\|}{\|D^m f\|} \ge \sup_{f|_{\mathbf{x}}=0} \frac{\|f - T(0)\|}{\|D^m f\|} \ge \sup_{f|_{\mathbf{x}}=0} \frac{\|f\|}{\|D^m f\|},$$

the last inequality since $\max\{\|f-T(0)\|, \|-f-T(-0)\|\} \ge \|f\|$.

Now consider $f \in L_\infty^{(m)}[a\,..\,b]$ with $f|_{\mathbf{x}} = 0$. This implies that $\delta_{x_i,\ldots,x_{i+m}} f = 0$, all i, therefore $D^m f \perp S_{m,\mathbf{x}}$. Let s be any perfect spline of degree m with $\le n-m$ interior knots which vanishes on $\mathbf{x}$. Such perfect splines exist, by (140)Theorem (take any $t \notin \mathbf{x}$ and match 0 on $\mathbf{x}$ and something nonzero at t). Assume without loss that $D^m s(a+) = 1$, hence $\|D^m s\|_\infty = 1$.

Now let $t \in [a\,..\,b]\backslash\mathbf{x}$ arbitrary. Since $D^m s$ has at most $n-m$ jumps, (142)Lemma implies that s cannot vanish at t, hence $(f(t)/s(t))s$ is well-defined and a perfect spline which agrees with f at $\mathbf{x}$ *and* the additional point, t. By (140)Theorem, this implies that

$$\|D^m f\|_\infty \ge \|D^m (f(t)/s(t))s\|_\infty = |f(t)/s(t)|.$$

In other words (since also $s|_{\mathbf{x}} = 0$), for any $t \in [a\,..\,b]$,

$$\max_{f|_{\mathbf{x}}=0} f(t)/\|D^m f\|_\infty = |s(t)|, \tag{150}$$

a remarkable result.

This shows that the lower bound on const_T provided by (149) equals $\|s\|_\infty$. Next, I show that this lower bound is sharp.

By (142)Lemma, $D^m s$ must have exactly $n-m$ sign changes (since, by assumption, it has no more than that, yet, since it is perpendicular to $S_{m,\mathbf{x}}$, it must have at least that many), therefore, also by the lemma, its sign changes, call them $a < \zeta_1 < \cdots < \zeta_{n-m} < b$, must interlace the knot sequence $\mathbf{x}$ in the sense that the Schoenberg-Whitney conditions (143) hold. Therefore, with the definitions

$$\zeta_{-m+1} = \cdots = \zeta_0 = a, \qquad b = \zeta_{n-m+1} = \cdots = \zeta_n,$$

we also have

$$\zeta_{i-m} \le x_i \le \zeta_i, \qquad i = -m+1, \ldots, -n+m,$$

with strict inequality except, perhaps when $\zeta_{i-m} = a$ or $\zeta_i = b$. By (135)Theorem, this implies that the spline space $S_{m,\zeta}$ contains, for each $f \in C^{(m-1)}[a..b]$, exactly one element, denoted Pf, which agrees with f on $\mathbf{x}$.

Since $f - Pf$ vanishes on $\mathbf{x}$, as does s, and the point masses in $D^m(f-Pf)$ occur only at the interior knots, and none of these lie in $\mathbf{x}$, we conclude from (147)Proposition that, for any $t \notin \mathbf{x}$, $\|D^m(f-Pf)\|_\infty \ge |f(t) - (Pf)(t)|/|s(t)|$. Since $\|D^m Pf\|_\infty = 0$ (as an element of $L_{\infty,\zeta}^{(m)}$), we conclude that

$$|f(t) - Pf(t)| \le |s(t)| \|D^m f\|_\infty, \qquad t \in [a\,..\,b],\ f \in L_\infty^{(m)}[a\,..\,b]. \tag{151}$$

This shows that

$$\mathrm{const}_P \le \|s\|_\infty \le \inf_T \mathrm{const}_T.$$

Now, $\operatorname{ran} P = S_{m,\zeta}$ is *not* contained in $C^{(m-1)}$, let alone in $L_\infty^{(m)}$, hence P cannot function as a recovery scheme in the present setup. However, for any small positive ε, we can modify each element in some basis for $S_{m,\zeta}$ in some ε-ball around $\{\zeta_i : i = 1, \ldots, n-m\}$ so as to produce an absolutely continuous $(m-1)$st derivative and a bounded mth derivative *without any change outside that ball* and an arbitrarily small change to the element inside that ball. Since none of the interior knots of ζ coincide with any of the x_i, the map P_ε, obtained in this way from P, will still be a recovery map and, by making ε small enough, we can guarantee that $\mathrm{const}_{P_\varepsilon}$ is arbitrarily close to const_P. This shows, in particular, that $\inf_T \mathrm{const}_T = \|s\|_\infty$.

GW interval centers and optimal interpolation

well, I am running out of time. For details, look at my paper 'computational aspects of optimal recovery' in 'Optimal Estimation in Approximation Theory', C. A. Micchelli and T. J. Rivlin (eds.), Plenum Publishing; 1977; 69–91.

smoothing spline

The data $\gamma \sim \Lambda g$ are usually not exact but only approximations to the exact values Λg. Correspondingly, most often one does not actual match the given γ but, rather, merely tries for an f with Λf close to γ while, at the same time, keeping some roughness measure $\|Tf\|$ of f 'small'. The two competing goals, of keeping both $\|\gamma - \Lambda f\|$ and $\|Tf\|$ 'small', are balanced by the choice of the **smoothing parameter** σ in the combined expression

$$\|\gamma - \Lambda f\| + \sigma \|Tf\|$$

which is to be minimized.

In this generality, both Λ and T are, typically, linear maps, from the given function space X to the normed linear spaces Y and Z, respectively. Assuming $\gamma = \Lambda g$ for some $g \in X$, the function

$$\sigma \to K(g, \sigma) := \inf_f \|\Lambda(g - f)\| + \sigma \|Tf\|$$

is exactly like Peetre's K-functionals mentioned earlier in the course. In particular, $K(g, \cdot)$ (as the pointwise infimum of a collection of straight lines with nonnegative y-intercept and nonnegative slope) is continuous, nondecreasing and concave downward, and is bounded since among the lines is one (in fact, there are many) with zero slope. Therefore, with

$$\overset{\Omega}{g}_{\sigma} := \operatorname{argmin}\{\|\Lambda(g - f)\| + \sigma \|Tf\| : f \in X\},$$

the map $\sigma \mapsto \|\Lambda(g - \overset{\Omega}{g}_{\sigma})\|$ is nondecreasing, while the map $\sigma \mapsto \|T \overset{\Omega}{g}_{\sigma}\|$ is nonincreasing. In effect, as σ increases, we lay more stress on smoothness, hence tolerate a larger $\|\Lambda(g - \overset{\Omega}{g}_{\sigma})\|$ in exchange for a smaller $\|T \overset{\Omega}{g}_{\sigma}\|$.

(152) Figure The smoothing spline is close to the center of two (degenerate) balls, one around the flat $g + \ker \Lambda$, the other around the subspace $\ker T$.

Offhand, there is no guarantee that $\overset{\mho}{g}_\sigma$ is well-defined; it might not exist at all, or there might be many minimizers. I consider this question only in the situation which corresponds to Schoenberg's (and Reinsch's) **smoothing spline**. The concrete situation here is that $X = L_2^{(m)}[a\,..\,b]$, with $\Lambda : f \mapsto f|_{\mathbf{x}}$ for some strictly increasing $\mathbf{x}$ in $[a\,..\,b]$, and $T := D^m$. This is Schoenberg's adaptation of Whittaker's smoothing method for tabulated data, in which $X = \ell_2(n)$, $\Lambda = \mathrm{id}$, and $T : \mathbb{R}^n \to \ell_2(n-3) : f \mapsto \Delta^3 f$.

Thus, with both Λ and T linear maps into Hilbert spaces, it is much more suitable to minimize

$$\|f\|_\sigma^2 := \|\Lambda(g - f)\|^2 + \sigma\|Tf\|^2,$$

since $\|\cdot\|_\sigma$ so defined is certain to be a seminorm on X associated with the semi-inner product

$$\langle f, g\rangle_\sigma := \langle \Lambda f, \Lambda g\rangle + \sigma\langle Tf, Tg\rangle.$$

Moreover, this is an inner product on X and, correspondingly, $\|\cdot\|_\sigma$ is the norm on X as a subspace of the direct product $Y \times Z$ via the embedding

$$X \to Y \times Z : f \mapsto (\Lambda f, Tf).$$

Of course, for this to be an embedding, it is necessary and sufficient that

$$\ker \Lambda \cap \ker T = \{0\},$$

an *assumption* made for the remainder of this section.

With this setup, we are looking at the special case $(y, z) = (\Lambda g, 0)$ of finding a best approximation from the linear subspace X to the given element (y, z) of the Hilbert space $Y \times Z$, hence know that uniqueness is automatic, as is characterization, and only existence is, offhand, in doubt. In particular, existence for arbitrary $(y, z) \in Y \times Z$ is equivalent to having X closed (in $Y \times Z$). Since Y and Z, as naturally embedded in $Y \times Z$, are orthogonal to each other, X is closed iff $\Lambda(X)$ is closed in Y and $T(X)$ is closed in Z. Typically, as in the case of the smoothing spline, Λ has finite rank, hence $\Lambda(X)$ is trivially closed in Y, thus all depends on $T(X)$ being closed in Z. This is usually, in effect, assumed (or else is evident, as in the smoothing spline case, when T is onto). To put it differently, the issue is 'only' whether $T(X)$ is complete.

So, *assume* that $\operatorname{ran} T$ is closed. Then we know that $\overset{\Omega}{g}_{\sigma}$ exists for every $0 < \sigma < \infty$. Further,

$$\overset{\Omega}{g}_{0+} = \operatorname{argmin}\{\|Tf\| : f \in g + \ker \Lambda\}$$

while

$$\overset{\Omega}{g}_{\infty} = \operatorname{argmin}\{\|\Lambda(g - f)\| : f \in \ker T\}.$$

The computation, *for given* σ, of $\overset{\Omega}{g}_{\sigma}$ is made easy by the standard characterization of a ba from a linear subspace of an inner product space. For given $(y, z) \in Y \times Z$, the ba $\overset{\Omega}{f}$ from X satisfies

$$\langle y - \Lambda \overset{\Omega}{f}, \Lambda f\rangle + \sigma\langle z - T\overset{\Omega}{f}, Tf\rangle = 0 \quad \forall\{f \in X\}.$$

In particular,

$$\langle z - T\overset{\Omega}{f}, T(\ker \Lambda)\rangle = 0,$$

showing that

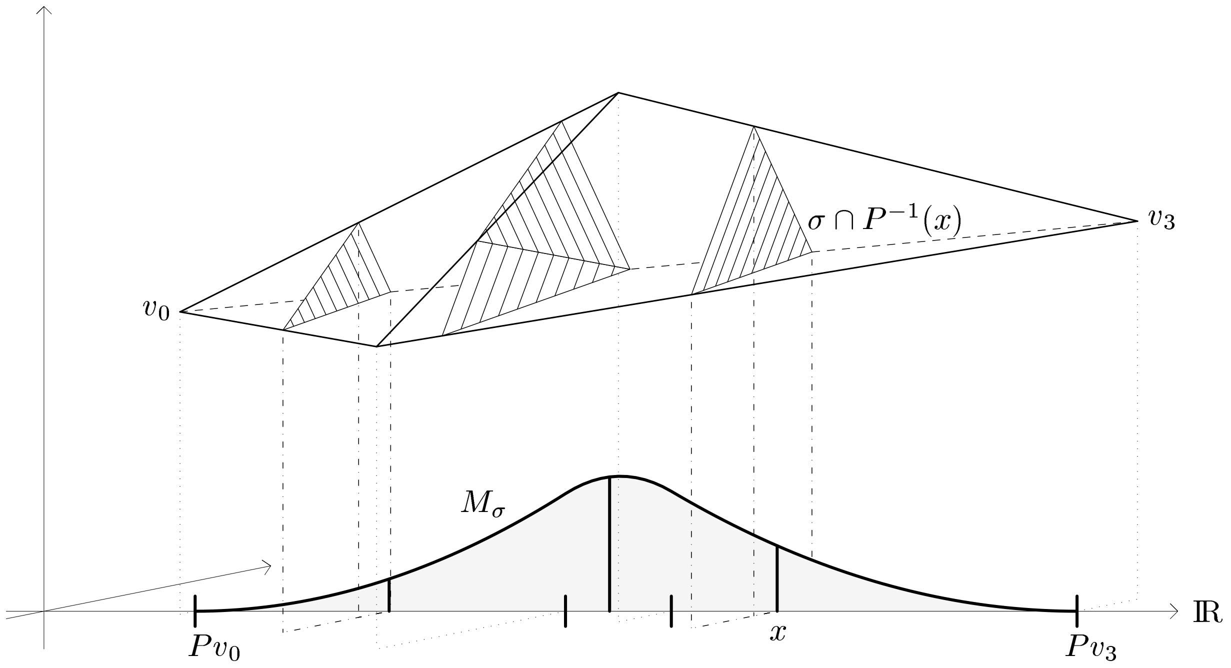
$$\overset{\Omega}{f} \in T^{-1}\{z\} + (\ker \Lambda)^{\perp}.$$

In particular, for the case $z = 0$ of present interest, $(\ker \Lambda)^{\perp} = \operatorname{ran} \Lambda^*$ and this is a finite-dimensional space when Λ maps to $\mathbb{R}^n$, being spanned by the representers of the 'rows' of Λ. In the case of the smoothing spline, this is just the space of natural splines.

The choice of σ, on the other hand, has been the subject of intensive investigations, with Grace Wahba's *Generalized Crossvalidation* approach being the method of choice. I cannot help ask, though, what would be so wrong with looking for a suitable compromise between $\overset{\Omega}{g}_{0+}$ and $\overset{\Omega}{g}_{\infty}$ on the straight-line segment between the two.

Multivariate B-splines

The definition of a multivariate B-spline as the d-dimensional 'shadow' of an n-dimensional polytope is based on Schoenberg's realization that a univariate B-spline is such a shadow, as is illustrated in (153)Figure.



(153)Figure. A quadratic B-spline as the shadow of a simplex

Here are the needed definitions. For any convex set B in $\mathbb{R}^n$, and any function φ defined at least on B, we define

$$\int_B \varphi := |\det V| \int_{V^{-1}B} \varphi V(a)\,\mathrm{d}a, \tag{154}$$

with $V : \mathbb{R}^r \to \mathbb{R}^n$ 1-1 affine and such that $\operatorname{ran} V = \flat(B)$, where

$$\flat(B)$$

denotes the **flat** spanned by B, i.e., the affine hull of B. Here, $|\det V|$ is, by definition, the r-dimensional volume of $V[0\,..\,1]^r$. Specifically, if

$$V : a \mapsto v_0 + [v_1, \ldots, v_r]a,$$

then

$$|\det V| = \sqrt{\det([v_1, \ldots, v_r]^*\,[v_1, \ldots, v_r])}.$$

The presence of the factor $|\det V|$ in (154) ensures that $\int_B \varphi$ is independent of the particular 1-1 affine map V used in its definition (154); in particular, $\int_B ()^0 = \mathrm{vol}_r(B)$.

A **polytope** is the convex hull of a finite set.

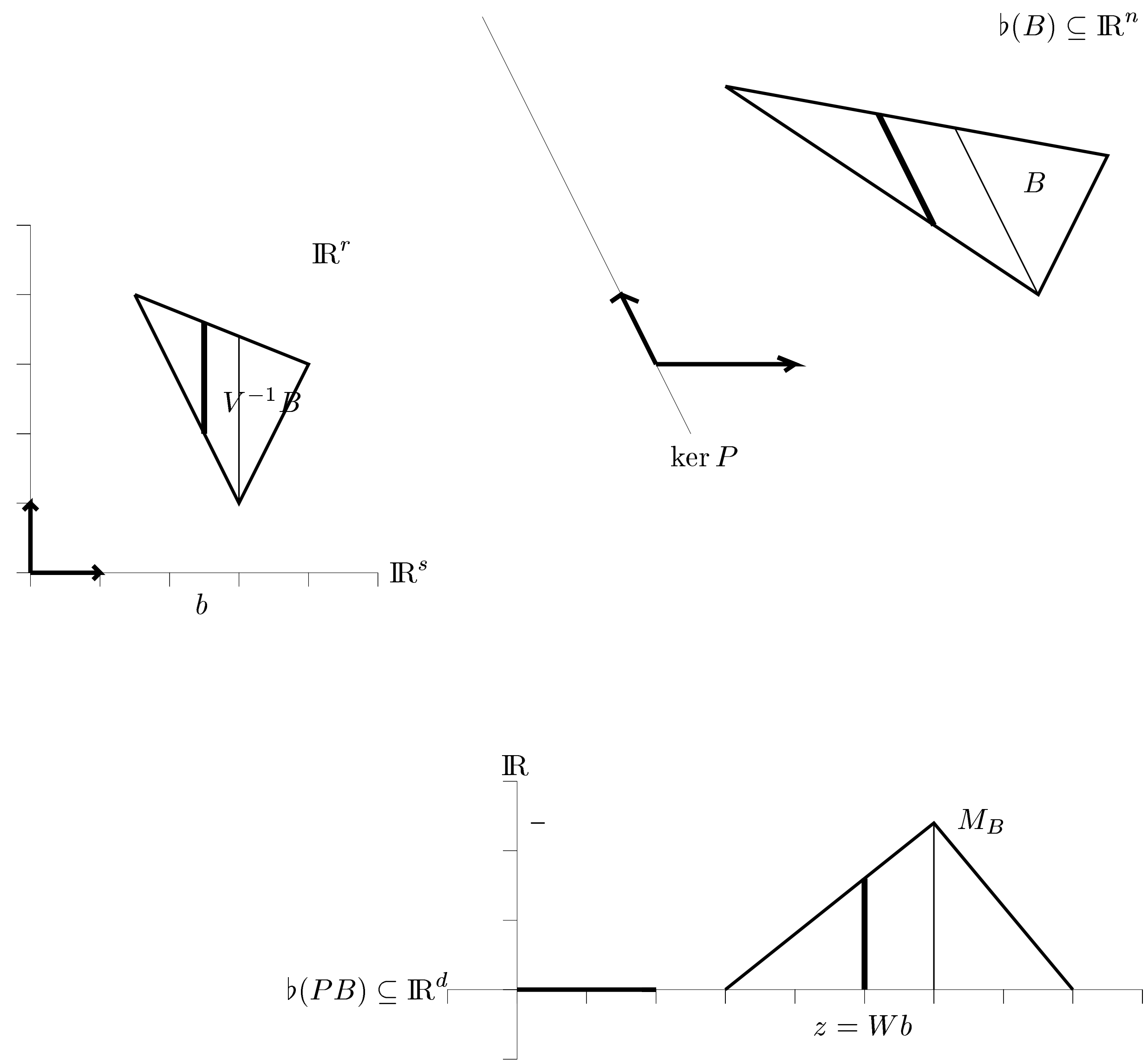
The d-dimensional B-spline determined from the polytope $B \subset \mathbb{R}^n$ by the affine map $P : \mathbb{R}^n \to \mathbb{R}^d$ is, by definition, the *distribution*

$$M_B : \varphi \mapsto \int_B \varphi P. \tag{155}$$

Equivalently, it is the *function* M_B defined on PB by the rule that

$$\int_B \varphi P = \int_{PB} M_B \varphi, \quad \forall \varphi \in C_0(\mathbb{R}^d).$$

Here is a formula for the function M_B.



(156)Figure. The three spaces involved in the formula for M_B

After translation of the origin in $\mathbb{R}^n$, we may assume that $\flat(B)$ is a linear subspace, of dimension r, hence $V = [v_1, \ldots, v_r]$. Also, after translation of the origin in $\mathbb{R}^d$, we may assume that P is linear. We also may choose V so that $[v_{s+1}, \ldots, v_r]$ is a basis for $\ker P \cap \flat(B)$. Then, $W := P[v_1, \ldots, v_s]$ is a basis for $\flat(PB)$, and, with $a = (b, c) \in \mathbb{R}^s \times \mathbb{R}^{r-s} = \mathbb{R}^r$, hence $P(Va) = Wb$,

$$\begin{aligned}\int_B \varphi P &= |\det V| \int_{V^{-1}B} \varphi(P(Va))\, \mathrm{d}a = |\det V| \int_{V^{-1}B} \varphi(Wb)\, \mathrm{d}c\, \mathrm{d}b \\ &= |\det V| \int_{W^{-1}(PB)} \varphi(Wb) \operatorname{vol}_{r-s}(V^{-1}B \cap \{(b,c) \in \mathbb{R}^r : c \in \mathbb{R}^{r-s}\})\, \mathrm{d}b \\ &= \left|\frac{\det V}{\det W}\right| \int_{PB} \varphi g,\end{aligned}$$

with

$$\begin{aligned} g(z) &= \operatorname{vol}_{r-s}(V^{-1}B \cap \{x \in \mathbb{R}^r : W(x_1, \ldots, x_s) = z\}) \\ &= \operatorname{vol}_{r-s} V^{-1}(B \cap P^{-1}\{z\}) \\ &= \operatorname{vol}_{r-s}(B \cap P^{-1}\{z\}) / |\det[v_{s+1}, \ldots, v_r]|, \end{aligned}$$

the last equality since $\operatorname{ran}[v_{s+1}, \ldots, v_r] = \ker P$. Therefore, with the additional requirement that

$$\operatorname{ran}[v_1, \ldots, v_s] \perp \ker P$$

(which implies that $|\det V| = |\det[v_1, \ldots, v_s] \det[v_{s+1}, \ldots, v_r]|$), we have, altogether,

$$M_B(z) = \left|\frac{\det[v_1, \ldots, v_s]}{\det[Pv_1, \ldots, Pv_s]}\right| \operatorname{vol}_{r-s}(B \cap P^{-1}\{z\}). \tag{157}$$

It follows that M_B is in $L_\infty(\mathbb{R}^d)$ exactly when PB has interior.

While this formula tells us that M_B can be considered a function on PB (and this was, indeed, how multivariate B-splines were first defined), the formulation (155) as a distribution provides mathematical efficiency. For example, if B has interior (i.e., if $\flat(B) = \mathbb{R}^n$), then, for any $z \in \mathbb{R}^n$,

$$D_{Pz} M_B = -\sum_{F \in B^{(n-1)}} \langle z, \mathbf{n}_F \rangle M_F, \tag{158}$$

with

$$B^{(n-1)}$$

the collection of **facet**s of B (i.e., of $n-1$-dimensional faces of B, with **face** the convex hull of any set of vertices of B) whose essentially disjoint union forms the boundary ∂B of B, and with $\mathbf{n}_F$ the outward unit normal to B on the facet F (and with $\langle , \rangle$ the scalar product on $\mathbb{R}^n$). The proof uses nothing more than the fact that, for any distribution M, $D_y M : \varphi \mapsto -M(D_y \varphi)$, hence

$$D_{Pz} M_B : \varphi \mapsto -\int_B (D_{Pz}\varphi) P = -\int_B D_z(\varphi P) = -\int_{\partial B} \langle z, \mathbf{n} \rangle \varphi P = -\sum_{F \in B^{(n-1)}} \langle z, \mathbf{n}_F \rangle \int_F \varphi P,$$

using Stokes' Theorem for the second-last equality, with $\mathbf{n}$ (perhaps confusingly) here standing for the unit outward normal for B, and using, for the last equality, the fact that the outward unit normal is constant on each of the facets F.

(159)Example. The simple example: $B = [0\,..\,1]^2 \subset \mathrm{I\!R}^2$, $P : \mathrm{I\!R}^2 \to \mathrm{I\!R} : x \mapsto x(1)$, $z = (1,1)$ illustrates all points made so far. One computes $M_B = \chi_{[0..1]}$, bearing in mind that M_B is only a distribution, hence, even if it is (given by) an L_∞ function, its value at a point may be uncertain (as it is here at 0 and at 1). $B^{(n-1)}$ has four elements, $F_l = 0 \times [0\,..\,1]$, $F_b = [0\,..\,1] \times 0$, $F_r = 1 \times [0\,..\,1]$ and $F_t = [0\,..\,1] \times 1$, and two of these are mapped by P to proper intervals, while the other two are mapped to points. E.g., $M_{F_l} = \delta_0$, while $M_{F_t} = M_{F_b} = M_B$. Thus, from (158),

$$DM_B = D_{Pz} M_B = -(-\delta_0 - M_{F_b} + \delta_1 + M_{F_t}) = \delta_1 - \delta_0.$$

□

It follows that, for any multiset $Y \subset \mathrm{I\!R}^n$, with $\eta := \#Y$,

$$D_{PY} M_B := \Big(\prod_{y \in Y} D_{Py}\Big) M_B = \sum_{F \in B^{(n-\eta)}} c_{Y,F} M_F. \tag{160}$$

Since, for $\eta > n - d$, any $F \in B^{(n-\eta)}$ has dimension $< d$, it follows that any derivative of M_B of order $> n - d$ has support only on a finite union of convex sets of dimension $< d$. This implies that M_B is pp, of degree $n - d$, in the sense that, on each connected component of the complement of $\cup PB^{(d-1)}$, M_B agrees with some polynomial of degree $\le n - d$. Further, if all PF with $F \in B^{(n-\eta)}$ have interior, then all derivatives of M_B of order $\le \eta$ are in $L_\infty(\mathrm{I\!R}^d)$, hence all derivatives of order $< \eta$ are continuous. Thus,

$$M_B \in \Pi^\rho_{n-d,\mu},$$

with

$$\mu$$

the collection of connected components of the complement of $\cup PB^{(d-1)}$ and

$$\rho := -1 + \max\{\eta : \forall \{F \in B^{(n-\eta)}\}\ \flat(PF) = \mathrm{I\!R}^d\}.$$

Any work with these multivariate B-splines is unthinkable without use of recurrence relations. The derivation of such recurrence relations relies on the operator

$$E : \varphi \mapsto \sum_j ()^{e_j} D_j \varphi$$

which was used already by Euler who proved that

$$Ef = mf \quad \forall f \in \Pi_m(\mathrm{I\!R}^d)^0.$$

(Indeed, for any multiindex α, $E()^\alpha = \sum_j ()^{e_j}\alpha_j()^{\alpha-e_j} = |\alpha|()^\alpha$.) Note that, for any y, $(Ef)(y) = (D_y f)(y)$. When applying E to a distribution, keep in mind that, for any distribution M and any continuous function f,

$$f\cdot M : \varphi \mapsto M(f\cdot\varphi).$$

With this,

$$EM_B : \varphi \mapsto \sum_j -M_B(D_j(()^{e_j}\varphi)) = -\sum_j M_B\left((1+()^{e_j}D_j)\varphi\right) = -dM_B\varphi - M_B(E\varphi).$$

Further,

$$(E\varphi)(Pz) = (D_{Pz}\varphi)(Pz) = D_z(\varphi P)(z) = E(\varphi P)(z).$$

However,

$$\sum_{j=1}^n D_j()^{e_j}(\varphi P) = n\varphi P + \sum_j ()^{e_j}D_j(\varphi P) = n\varphi P + E(\varphi P),$$

while, by Stokes',

$$\int_B \sum_{j=1}^n D_j()^{e_j}(\varphi P) = -\int_{\partial B}\langle\cdot,\mathbf{n}\rangle\varphi P,$$

using the fact that $\sum_j\langle e_j,\mathbf{n}\rangle()^{e_j} = \langle\cdot,\mathbf{n}\rangle$. Therefore,

$$M_B(E\varphi) = \int_B E(\varphi P) = -nM_B\varphi - \sum_{F\in B^{(n-1)}}\langle F,\mathbf{n}_F\rangle M_F\varphi.$$

Hence, altogether,

$$EM_B = (n-d)M_B - \sum_{F\in B^{(n-1)}}\langle F,\mathbf{n}_F\rangle M_F, \tag{161}$$

a formula first proved by Hakopian in the special case when B is a simplex.

Since $(EM_B)(Pz) = (D_{Pz}M_B)(Pz) = -\sum_{F\in B^{(n-1)}}\langle z,\mathbf{n}_F\rangle M_F z(Pz)$ by (158), this gives, finally, the **recurrence relation**

$$(n-d)M_B(Pz) = \sum_{F\in B^{(n-1)}}\langle F-z,\mathbf{n}_F\rangle M_F(Pz), \tag{162}$$

first proved by Micchelli in the special case when B is a simplex. (Here and above, I wrote $\langle F,\mathbf{n}_F\rangle$ to stand for the sole element in that set.)

Standard B-splines

While the recurrence relation makes it possible to refer the evaluation of a B-spline to the simple level where M_F is a scalar multiple of a characteristic function, it is ever so much nicer to deal with M_B for which the resulting M_F are of the same kind and, moreover, are describable entirely entirely in terms relating to $\mathbb{R}^d$ only. This is possible only for very few bodies B, namely (essentially) simplices, parallelepipeds, and polyhedral cones. Since any affine change of $\mathbb{R}^n$ can always be accounted for by the choice of P, one usually uses standard bodies, e.g., the standard simplex

$$\Delta_n := \operatorname{conv}\{\mathbf{i}_j : j = 0, \ldots, n\},$$

with

$$\mathbf{i}_0 := 0,$$

or the standard cube

$$\square := [0 \,..\, 1]^n,$$

or the positive cone

$$\mathbb{R}_+ \Delta_n = \mathbb{R}^n_+.$$

The resulting B-splines are called, correspondingly, **simplex** splines, **box** splines, and **cone** splines.

As could be surmised from the relationships between these standard bodies, there are relationships between the resulting B-splines. For example, any B-spline is a finite sum of simplex splines (since any polytope is the essentially disjoint union of simplices). Also, near an extreme point of its support, any B-spline M_B is a scalar multiple of a simplex spline.

If $\sigma = \operatorname{conv}\{p_0, \ldots, p_n\}$ and $\sigma' = \operatorname{conv}\{p'_0, \ldots, p'_n\}$ are two simplices in $\mathbb{R}^n$ with interior, then there is an affine map Q on $\mathbb{R}^n$ which maps one to the other. If, in addition, $Pp_j = Pp'_j$, all j, then $M_\sigma = cM_{\sigma'}$. For this reason, it has become customary to identify a simplex spline by the $n+1$ points $t_j := Pp_j$, $j = 0, ..1$, i.e., to write

$$M(\cdot | t_0, \ldots, t_n),$$

with the understanding that this simplex spline is always normalized to have integral 1. Note that this notation agrees with the standard one for a univariate B-spline that integrates to 1.

Similarly, with P any linear map from $\mathbb{R}^n$ to $\mathbb{R}^d$, the box spline defined by it from the standard box $\square$ is written

$$M_X := M(\cdot | X),$$

with X the matrix representation of P with respect to the standard bases (i.e., with $X = P$, if we take for granted that any linear map from $\mathbb{R}^n$ to $\mathbb{R}^d$ is, in effect, a $d \times n$ matrix).

There's not much use for just one B-spline

In approximation, one uses pp because, for a sufficiently fine partition, one expects them to approximate well locally. Of course, that means working with the span of many functions.

The bad news is that it is not possible to generate a B-spline basis for a smooth pp space on some arbitrary partition on $\mathbb{R}^d$ (for $d > 1$). Rather, the B-spline construction succeeds in providing compactly supported pp of some degree k and in $C^{(k-1)}$ (or close to it) precisely because it generates a very special partition.

For this reason, useful spaces of pp functions with a B-spline basis have been obtained for a 'general' partition only if it is a suitably constructed refinement of given general triangulation, or else is a very special regular partition.

(to be continued)

Index