These are classnotes for Math/CS 887, Spring '03 by Carl de Boor
corrections are welcome!

## Approximation Theory overview

One considers special instances of the following

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

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

*Abbreviation:*

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

## Basic questions

- Existence: $\#\mathcal{P}_M(g) \ge 1$ ?
- Uniqueness: $\#\mathcal{P}_M(g) \le 1$ ? More generally, $\#\mathcal{P}_M(g) =$?
- 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 following problem, of approximating a curve, shows, there are important practical instances in which linearity plays no role, hence 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., $\|x\| := \|x\|_2 := \big(\int_T |x(t)|^2\,\mathrm{d}t\big)^{1/2}$ **Least-squares**.
  More generally: inner product spaces.
- $C(T)$, i.e., $\|x\| := \|x\|_\infty := \sup_{t\in T} |x(t)|$ **uniform** (or, **Chebyshev**).
- $\mathbf{L}_1(T)$, i.e., $\|x\| := \|x\|_1 := \int_T |x(t)|\,\mathrm{d}t$ **Least-mean**.
- $\mathbf{L}_p(T, w)$, i.e., $\|x\| := \|x\|_{p,w} := \big(\int_T w(t)|x(t)|^p\,\mathrm{d}t\big)^{1/p}$ **weighted $\mathbf{L}_p$**

## Specific choices for $M$

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

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

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

- $\Pi_k := \Pi_{\le k} := \mathrm{ran}[()^j : j = 0,\dots,k]$ (algebraic) polynomials of degree $\le k$. Here, $()^j : t \mapsto t^j$ is the way I'll indicate the power function until someone comes along with a better notation. More generally,

  $$\Pi_k = \Pi_k(\mathbb{R}^d) := \mathrm{ran}[()^\alpha : |\alpha| \le k]$$

  with
  $$()^\alpha : \mathbb{R}^d \to \mathbb{F} : x \mapsto x^\alpha := x(1)^{\alpha(1)}\cdots x(d)^{\alpha(d)}, \quad \alpha \in \mathbb{Z}_+^d,$$
  and $|\alpha| := \|\alpha\|_1 = \sum_i \alpha(i)$. Also, for $\alpha \in \mathbb{Z}_+^d$,

  $$\Pi_\alpha := \Pi_{\le\alpha} := \mathrm{ran}[()^\beta : \beta \le \alpha].$$

- $\mathcal{R}_{k,\ell} := \Pi_k/\Pi_\ell := \{p/q : p \in \Pi_k, q \in \Pi_\ell\}$ rational functions of degrees $k$, $\ell$.
- $\overset{\circ}{\Pi}_k := \mathrm{ran}[\sin(\nu\cdot), \cos(\nu\cdot) : \nu = 0,\dots,k]$ trigonometric polynomials of degree $k$. The natural domain for trigonometric polynomials is the circle, i.e., the interval $[0\,..\,2\pi]$ or $[-\pi\,..\,\pi]$ 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** $\theta$, a definition which even makes sense for $x \in \mathbb{F}^d$ with $\theta$ also in $\mathbb{F}^d$ and
  $$\theta\cdot x := \sum_i \theta(i)x(i)$$

- $\mathrm{Exp}_\Theta$ := $\mathrm{ran}[e_\theta : \theta \in \Theta]$ exponentials with frequencies $\theta \in \Theta$. We already noted that $\mathring{\Pi}_k(\mathbb{R}) \subseteq \mathrm{Exp}_{\mathrm{i}\{-k,\ldots,k\}}$. Also, for any $\Theta \subset \mathbb{R}$, $\Pi_k = \lim_{h\to 0} \mathrm{Exp}_{h\Theta}$ whenever $\#\Theta = k+1$.
- $\Pi^\rho_{k,\boldsymbol{\xi}}$ := piecewise polynomials (=: **pp**) in $C^{(\rho)}$, of degree $\le k$, with **break sequence** $\boldsymbol{\xi} = (\cdots < \xi_j < \xi_{j+1} < \cdots)$. (**splines**)

The first and third are linear. The second is nonlinear. The last two are linear or nonlinear depending on whether the frequencies $\Theta$ (resp. the break sequence $\boldsymbol{\xi}$) is fixed or variable.

## Degree of approximation

considers the behavior of $h \mapsto \mathrm{dist}\,(g, M_h)$ as a function of the (discrete or continuous) parameter $h$. E.g., $k \mapsto \mathrm{dist}\,(g, \Pi_k)$ as $k \to \infty$, or $h \mapsto \mathrm{dist}\,(g, \Pi^\rho_{k,h\boldsymbol{\xi}})$ as $h \to 0$. Usually, one considers only

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

with $K$ a class of functions sharing with the particular $g$ 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: $g \in K \quad\Longrightarrow\quad \mathrm{dist}\,(g, M_h) = O(h^\alpha)$

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

Saturation theorems: $\mathrm{dist}\,(g, M_h) = o(h^\alpha) \quad\Longrightarrow\quad g \in K_0$ (for some appropriate $\alpha$ and with $K_0$ some very 'small' set). E.g., $\mathrm{dist}\,(g, \Pi^0_{1,\boldsymbol{\xi}}) = o(|\boldsymbol{\xi}|^2) \implies g \in \Pi_1$.

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

$$K_f : t \mapsto \inf_q (\|f - q\| + t\|q\|'),$$

which plays a major role in the precise description of $h \mapsto \mathrm{dist}\,(K, M_h)$ for such $K$.

Related question: Is $M$ a good choice for approximating $g$, given that we know that $g \in K$? 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, i.e., $Y$ with $d_{\dim Y}(Y) = \mathrm{dist}\,(K, Y)$, one can often find a 'ladder' $(M_n)$ for which $\mathrm{dist}\,(K, M_n) \sim d_n(K)$ and $\dim M_n \sim n$.

Here,

$$\begin{aligned} A(t) = O(B(t)) \ &:= \ \limsup{}_t |A(t)/B(t)| < \infty; \\ A(t) = o(B(t)) \ &:= \ \lim{}_t |A(t)/B(t)| = 0; \\ A(t) \sim B(t) \ &:= \ A(t) = O(B(t)) \text{ and } B(t) = O(A(t)), \end{aligned}$$

with the limiting value of $t$ usually clear from the context. E.g., the order of $A(n)$ or $A_n$ will always be considered as the natural number, $n$, goes to $\infty$, while the order of $A(h)$ or $A_h$ will always be considered as the real positive number, $h$, goes to zero.

## 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 $\|(1-A)_{|K}\| := \sup_{f\in K} \|f - Af\|$ may well be close to $\operatorname{dist}(K, M)$.

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

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

Any such $A$ is necessarily a projector (onto $M$). 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

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

## 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

"weierstrass **(2) Weierstraß (1885).** *For any (finite) interval* $I = [a\,..\,b]$, $\Pi|_I$ *is dense in* $C(I)$.

Since both $\Pi$ and $\|\cdot\|_\infty$ are invariant under **translation**

$$f \mapsto f(\cdot + t)$$

and **dilation**

$$f \mapsto f(\cdot/\sigma),$$

it is sufficient to consider just one nontrivial interval, e.g., the special case $I = [0\,..\,1]$.

I 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.

"korovkin **(3) Korovkin (1957).** *$\mathbb{F} = \mathbb{R}$, $T$ compact, $(U_n)$ in $L(C(T))$, $U_n$ positive, $U_n \xrightarrow{pw} 1$ on some finite set $F \not\subseteq \{0\}$. If there exists $(a_f : f \in F)$ in $C(T)$ so that*

$$p(t,s) := \sum_{f \in F} f(t)a_f(s) \geq 0 \quad \text{with equality iff } t = s,$$

*then $U_n \xrightarrow{pw} 1$.*

Explanations:

(i) With $U_n$ and $U$ being maps from 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

$$f \leq g :\iff \forall t \in T \; f(t) \leq g(t).$$

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

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

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

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

hence,

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

**Proof of Korovkin (3):** By assumption, $U_n \xrightarrow{pw} 1$ on $F$, hence ($U_n$ being linear), also on

$$\text{ran}[F] := \{\sum_{f \in F} fc(f) : c \in \mathbb{R}^F\} =: \text{span}\, F.$$

The latter is finite-dimensional, therefore $U_n \to 1$ *uniformly on bounded subsets of* ran$[F]$.

The rest is a very nice trick, in which the arbitrary $g \in C(T)$ to be approximated from $\Pi$ is locally related to some element of ran$[F]$, as follows. From the assumption, ran$[F]$ contains (strictly) positive functions, e.g., the function $t \mapsto p(t,s) + p(t,s')$ for any $s \neq s'$ in case $\#T > 1$ and $|f|$ for any $f \in F\backslash 0$ otherwise. Let $p^*$ be one such. For $s \in T$, set

$$g =: \frac{g(s)}{p^*(s)} p^* + h(\cdot, s).$$

Then

$$(4) \qquad (U_n g)(s) = \frac{g(s)}{p^*(s)}(U_n p^*)(s) + (U_n h(\cdot, s))(s),$$

and

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

Since $T$ is compact, $\|1/p^*\|_\infty < \infty$, hence

$$\frac{g(s)}{p^*(s)}(U_n p^*)(s) \to \frac{g(s)}{p^*(s)} p^*(s) = g(s)$$

*uniformly in* $s$. Korovkin's result therefore follows from the following claim. □

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

**Proof:** By the positivity of $U_n$, $|U_n h(\cdot, s)| \le U_n(|h(\cdot, s)|)$. Take any $\varepsilon > 0$. Then $|h| \le \varepsilon +$ a bound for $|h|$ on the set $\Delta_\varepsilon := \{(t, s) : |h(t, s)| \ge \varepsilon\}$. This set is closed since $h$ is continuous, hence compact; it also does not contain the zero-set of $p$, i.e., the set $\{(t, t) : t \in T\}$, since $h$ vanishes there. Therefore,

$$\delta := \inf p(\Delta_\varepsilon) \; > 0,$$

hence, $|h| \le (\|h\|_\infty/\delta)p$ on $\Delta_\varepsilon$. So,

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

Consequently, for any $s$,

$$|U_n(h(\cdot, s))| \le U_n(|h(\cdot, s)|) \le \varepsilon M + (\|h\|_\infty/\delta) U_n(p(\cdot, s)),$$

with

$$M := \sup_n \|U_n()^0\|_\infty$$

finite since, by the strict positivity of $p^*$, there is some positive $b$ so that $()^0 \le bp^*$, therefore

$$|U_n()^0| \le U_n()^0 \le bU_n(p^*) \to bp^*.$$

Now, $\{p(\cdot, s) : s \in T\}$ is a bounded subset of the finite-dimensional linear space ran$[F]$, hence, on it, $U_n$ converges to 1 *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 $n \ge n_\varepsilon$,

$$|U_n(h(\cdot, s))(s)| \le \varepsilon(M + 1).$$

Since $\varepsilon$ was arbitrary, this proves the Claim. □

From this, (2)Weierstraß for $I = [0 \mathinner{..} 1]$ follows, with the choices $F = \{()^0, ()^1, ()^2\}$ hence $\operatorname{ran}[F] = \Pi_2$, $p : (t, s) \mapsto (t - s)^2$, and $U_n = B_n$, $n > 0$, the **Bernstein operator** (introduced by Bernstein in 1912 for a different proof of Weierstraß)

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

with

$$\beta_{r,s} : t \mapsto \binom{r+s}{r} t^r (1-t)^s.$$

Indeed, $B_n$ is linear, and is positive as a map on $C(I)$. Further,

$$DB_n f = \sum_{j=1}^{n} \binom{n}{j} j()^{j-1}(1-\cdot)^{n-j} f(\frac{j}{n}) - \sum_{j=0}^{n-1} \binom{n}{j} (n-j)()^j (1-\cdot)^{n-j-1} f(\frac{j}{n}),$$

therefore (and this is of independent interest)

$$DB_n f = n \sum_{j=0}^{n-1} \binom{n-1}{j} ()^j (1-\cdot)^{n-1-j} \Delta f(\frac{j}{n}), \qquad (5)$$

“diffbernstein

with

$$\Delta f := f(\cdot + \frac{1}{n}) - f,$$

using the facts that $\binom{n}{j} j = n\binom{n-1}{j-1}$ and $\binom{n}{j}(n-j) = n\binom{n-1}{j}$. Now note that $\Delta(\Pi_k) \subseteq \Pi_{k-1}$. Therefore, $\Delta^{k+1} := \Delta\Delta^k$ vanishes identically on $\Pi_k$, hence

$$B_n(\Pi_k) \subseteq \Pi_k$$

(for $k \le n$; it's trivial for $k > n$). Since also

$$B_n f(t) = f(t), \quad t = 0, 1,$$

it follows that $B_n = 1$ on $\Pi_1$, therefore $B_n()^j \to ()^j$ for $j = 0, 1$ trivially, and we are done once we show that $B_n f \to f$ for some $f \in \Pi_2 \backslash \Pi_1$, e.g., for $f := \beta_{1,1} : t \mapsto t(1-t)$. This $f$ vanishes at 0 and 1 and is quadratic, hence $B_n f$ must have the same properties, and therefore must equal $\alpha_n f$ for some $\alpha_n$. It follows that $DB_n f(0) = \alpha_n Df(0) = \alpha_n$, while, by (5), $DB_n f(0) = n(f(1/n) - f(0)) = nf(1/n) = 1 - \frac{1}{n}$. Therefore $\alpha_n = \frac{n-1}{n}$, i.e., $B_n f = f - f/n \longrightarrow f$ as $n \to \infty$. □

"ddimweierstrass **(6) $d$-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 that, after translation and dilation, 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 $f_I \in C(I)$ is given by the rule

$$f_I : x \mapsto \begin{cases} f(x), & x \in T; \\ \alpha + \inf_{t\in T}(f(t) - \alpha)\,\mathrm{dist}\,(x,t)/\,\mathrm{dist}\,(x,T), & x \notin T, \end{cases}$$

with $\alpha := \inf f(T) - 1$, say, i.e., so that $f - \alpha()^0$ is strictly positive. 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. This means that

$$U_n f := \sum_{0\le j\le(n,\ldots,n)} \beta_j f(j/n),$$

with $j \in \mathbb{Z}^d$ and

$$\beta_j : x \mapsto \beta_{j(1),n-j(1)}(x(1)) \cdots \beta_{j(d),n-j(d)}(x(d)).$$

Evidently, $U_n$ is linear and positive. Moreover,

$$\forall\{(f_i) \in (C([0\,..\,1]))^d\}\ \ U_n \otimes_{i=1}^d f_i = \otimes_i B_n f_i : x \mapsto \prod_i (B_n f_i)(x(i)).$$

From the univariate argument, $U_n \xrightarrow{pw} 1$ on $\Pi_{2,\ldots,2}$ since $U_n()^\alpha = \otimes_{i=1}^d B_n()^{\alpha(i)}$, 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

$$p : (t,s) \mapsto 1 - \cos(t-s) = (e_0 - (e_{\mathrm{i}} + e_{-\mathrm{i}})/2)(t-s),$$

i.e., $\mathrm{ran}[F] \subset \overset{\circ}{\Pi}_1$, 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

$$s_j f := \sum_{|\mu| \le j} e_{i\mu} \int_{\mathbb{T}} e^{-i\mu t} f(t)\, dt/(2\pi).$$

In particular, $\sigma_n$ is a positive operator (which $s_n$ is not), and

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

hence $\sigma_n(e_{ik})$ converges to $e_{ik}$ for any $k$, in particular for $|k| \le 1$. On the other hand, recall that $\|s_j\| \sim \ln j$ as $j \to \infty$, hence $s_j$ fails to converge to 1 on $C(\mathbb{T})$.

**Lebesgue's proof (1898) of (2)Weierstraß** Let $I := [a\,.\,.b]$. For any $f \in C(I)$ and any finite sequence $\boldsymbol{\xi} := (a = \xi_1 < \xi_2 < \cdots < \xi_{\ell+1} = b)$, the broken-line interpolant to $f$ is, by definition, the unique element $P_{\boldsymbol{\xi}} f$ in $\Pi^0_{1,\boldsymbol{\xi}}$ that agrees with $f$ at $\boldsymbol{\xi}$. For $\xi_j \le t \le \xi_{j+1}$,

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

hence

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

with $\omega_f$ the *(uniform) modulus of continuity of $f$* and

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

Conclusion: The collection

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

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

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

“claimapproxabs **(7) Claim.** *For finite $[a\,.\,.\,b]$ and any $s$,* $\operatorname{dist}_\infty(|\cdot - s|, \Pi)[a\,.\,.\,b] = 0$.

**Proof:** Only the case $s \in (a\,.\,.b)$ needs proof. For such $s$, we may choose $\sigma \in \Pi_1 \backslash \Pi_0$ that 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 that, on $[a\,.\,.\,b]$, is that close to $c|\cdot - s|$ for some nonzero $c$ that depends only on $\sigma$. Hence it is sufficient to prove the Claim for $[a\,.\,.\,b] = [-1\,.\,.\,1]$ and $s = 0$. Since $|t| = (t^2)^{1/2}$, i.e.,

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

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

**(8) Claim.** $\operatorname{dist}_\infty(()^{1/2}, \Pi)[0\,..\,1] = 0$.

**Proof:** A standard proof uses the Taylor series expansion

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

which converges uniformly on any compact subset of $(-1\,..\,1)$ (since the power coefficients are all bounded by 1). I prefer the following proof, from Dieudonné's Analysis.

Define the sequence $(u_n : n = 0, 1, 2\cdots)$ by the iteration

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

with

$$u_0 := 0.$$

Claim: for all $n$, $u_n \in \Pi$ and

$$u_n \le ()^{1/2}, \quad \text{on} \ \ [0\,..\,1].$$

Indeed, assuming this already to be true for $n$ (as it is for $n = 0$), we observe that then also $u_{n+1} \in \Pi$ and compute

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

on $[0\,..\,1]$.

It follows that $()^1 - u_n^2 \ge 0$ on $[0\,..\,1]$, hence $u_{n+1} = u_n + (()^1 - (u_n)^2)/2 \ge u_n$. Thus, $(u_n)$ is monotone increasing, yet bounded on $[0\,..\,1]$, therefore pointwise convergent there, and its limit is necessarily a fixed point of the iteration used to define it, hence its limit is $()^{1/2}$. But since this limit function is continuous and $[0\,..\,1]$ is compact, $u_n \to ()^{1/2}$ uniformly (by Dini's Theorem). □

## 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 $\mathbb{F}^T$, and $\mathbb{F}^T$ is a **ring** wrto these two operations. In fact, it is a **ring with identity** since it contains the multiplicative identity, i.e., the function $1 : t \mapsto 1$. 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**.

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 (since its range contains the vector $(1, 1)$ as well as some vector $(\alpha, \beta)$ with $\alpha \neq \beta$), hence the map has right inverses. In particular, for any $f \in \mathbb{F}^T$ and any $s, t \in T$, there exists $a_{f,s,t} \in A$ that agrees with $f$ at $s$ and $t$. It is this conclusion that is needed.

"stoneweierstrass **(9) Stone(-Weierstraß) (1937).** *Let $T$ be compact metric. The only closed subalgebra $A$, of* real $C(T)$*, that separates points and contains 1 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 7, $|\cdot|$ can be approximated, uniformly on $a(T)$, by polynomials $p$, and $A$, being an algebra, contains $p \circ a : t \mapsto p(a(t))$ 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 $\varepsilon$ is arbitrary and $A$ is closed, $f \in A$ follows. □

From this, even the multivariate Weierstraß theorem, (6), follows since $\Pi$ is an algebra with identity that 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.

On the other hand, Stone's theorem does have the following (weaker) immediate consequence for complex-valued functions which relies on nothing more than the fact that $C_{\mathbb{C}}(T) = C_{\mathbb{R}}(T) + \mathrm{i}C_{\mathbb{R}}(T)$.

"stoneweierstrass **(10) complex Stone.** *Let $T$ be compact metric. The only closed subalgebra $A$ of* complex $C(T)$ *with identity which separates points and is closed under conjugation is $C(T)$ itself.*

Here, being closed under conjugation means that $A$ contains, with $f$, also its conjugate, $\overline{f} : T \to \mathbb{C} : t \mapsto \overline{f(t)}$.

In particular, $\Pi|_D$ is dense in $C_{\mathbb{C}}(D)$ for $D = \{x \in \mathbb{R}^2 : \|x\|_2 \le 1\}$.

## Existence

$M \subset X$ is called an **existence set** in case $\forall\{g \in X\}\ \mathcal{P}_M(g) \ne \emptyset$. 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 $\lambda$ which **does not take on its norm**, i.e., for which there is no *nontrivial* $g \in X$ with $|\lambda g| = \|\lambda\|\|g\|$. For,

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

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

$$\|\lambda\|\,\mathrm{dist}\,(g, \ker\lambda) = |\lambda g| = |\lambda(g-m)| \le \|\lambda\|\|g-m\| = \|\lambda\|\,\mathrm{dist}\,(g, \ker\lambda),$$

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

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

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

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

of $M$ with the closed ball around $g$ with radius twice the distance of $g$ from $M$, we have $M_g$ not empty and

$$\mathcal{P}_M(g) = \mathcal{P}_{M_g}(g).$$

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

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 $\le k$ 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 $\le k$ 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 $g$, i.e., $\lim_n \|g - s_n\| = \operatorname{dist}(g, M)$. 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 $\xi$.

Assume first that $\xi$ 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, $\|g\|(U)$ 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 $\Pi_k$, and $\Pi_k$ 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 $\xi$ and polynomial pieces $(p_j : j = 1, \ldots, l)$ is, indeed, the norm limit of $(s_n)$. This implies that

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

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

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 $\xi$ 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

"defnearconv (12) $$\forall\{y \in X\} \quad \liminf_n \|x_n - y\| \ge \|x - y\|.$$

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:

"propnearlyexist **(13) 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 $g \in X$ is necessarily bounded, hence, AGTASMAT, there exists $m \in M$ so that

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

hence $m \in \mathcal{P}_M(g)$. □

(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.

"propseminearly **(14) Proposition.** *Assume that* $\Phi$ *is a collection of seminorms on the nls* $X$*, and that*

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

*If* $(x_n)$ **$\Phi$-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 provide a translation-invariant characterization of convergence, it is sufficient to prove (12) 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 14, hence also its conclusion. The above selection process is sure to provide a subsequence which, for some $\varepsilon > 0$, $\varphi_\varepsilon$-converges to a certain element of $\Pi_{k,l}$. Since $\Pi_k$ is finite-dimensional, this implies that the subsequence $\varphi_\eta$-converges to that element for every $0 < \eta \le \varepsilon$, hence nearly converges to it, and the proof of Proposition 13 does the rest.

Notice that the near-limit of that subsequence of $(m_n)$ only depends on the polynomial pieces $p_{j,n}$ for which $I_j$ is not trivial. Hence, if some $I_j$ is trivial, we may change the corresponding $p_{j,n}$ arbitrarily without changing the near-limit. In particular, we can arrange such changes that $m_n$ will fail to converge in norm, thus providing an explicit example of the fact that near-convergence is truly more general than convergence. □

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

$$\Pi^{\rho}_{k,\xi},$$

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

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

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

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 $\Pi^{\rho}_{k,\xi}$ can appear. It is useful to discuss this question in the more general context of $\gamma$-polynomials.

## $\gamma$-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**

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

shows that $k(\tau - \cdot)_+^{k-1}$ represents $\delta_\tau$ on $Y := \{f \in C^{(k)}[a\,..\,b] : f \ \text{ vanishes } k\text{-fold at } 0\}$ 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 $\gamma$ 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 $\gamma$ 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 $M^-_{\gamma,l}$.

We may pick up this discussion later, after we have recalled well-known facts about divided differences.

## Uniqueness

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

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

Hence, if $M$ is convex, then so is $\mathcal{P}_M(g)$, and may well contain nontrivial line segments if, e.g., $M$ is a linear subspace and the boundary $\partial B$ 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 $\ell_p$-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., $X = L_1[0\,..\,1]$, $M = \Pi_0$, $g = \chi_{[1/2..1]}$.) However, nonuniqueness is not guaranteed; it depends on the attitude of $M$ with respect to $B_{\operatorname{dist}(g,M)}(g)$. (In the example just given, one gets uniqueness for any $g = \chi_{[x..1]}$ *except* when $x = 1/2$.) Draw pictures in $\ell_p(2)$ for $p = 1, 2, \infty$.

A set $M$ that provides exactly one ba for every $g \in X$ 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

“defmetricP (15)
$$\mathcal{P}_M(g) =: \{P_M g\}, \quad g \in X.$$

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

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

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

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

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

$$d(m', g_n) - d(g_n, g) \leq d(m', g) < d(m, g) \leq d(m, m_n) + d(m_n, g_n) + d(g_n, g).$$

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

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

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

Note that, even when $M$ is a linear subspace, the metric projector is usually *non-linear* 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 $g$ is 'very far' from $M$. Draw the picture. More than that, there are perfectly reasonable $M$ that have 'corners' in the sense that $\#\mathcal{P}_M(g) > 1$ for certain $g$ arbitrarily close to $M$. Here is one example.

**(18) Example** Consider $M := \Pi_{0,2}$ in $L_2[-1\,..\,1]$, say, and take $g = \alpha()^2$. 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 $g = ()^2$. Let $m \in \mathcal{P}_M(g)$, and let $\zeta$ be its sole breakpoint. Since $g$ is even, also $m' : t \mapsto m(-t)$ is in $\mathcal{P}_M(g)$. 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, $g - m$ must be orthogonal to $m_\zeta$ for every such $\zeta$. However, $\mathrm{ran}[m_\zeta : -1 < \zeta < 1]$ is dense in $L_2[-1\,..\,1]$, therefore $g = m$ would follow, contradicting the fact that $g \notin \Pi_0$. □

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).

Note that $\gamma : [a\,..\,b] \to L_2([a\,..\,b]) : t \mapsto \chi_{[t..b]}$ is a continuous curve which, at every point $\gamma(s)$, 'turns 90°' in the sense that, for all $r < s < t$, the secant directions $\gamma(s) - \gamma(r)$ and $\gamma(t) - \gamma(s)$ are perpendicular to each other. It is this counter-intuitive example that led to the concept of $\gamma$-polynomials.

## 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 $\lambda$ on a nls $X$ over the *real* scalars is very simple: The kernel of $\lambda$ cuts $X$ into two halfspaces, on one $\lambda$ is positive, on the other it is negative. Further, $\lambda$ is constant on **hyperplane**s parallel to the kernel, i.e., on

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

for any $x_\alpha \in H(\lambda, \alpha)$.

Let $M$ be a closed subset of $X$, let $g \notin M$, hence

$$r := \mathrm{dist}\,(g, M) > 0$$

and $\inf \|M - B_r(g)\| = 0$, therefore

“supgeinf (19)
$$\sup \lambda(M) \geq \inf \lambda(B_r(g)), \quad \forall \lambda \in X^*.$$

Now suppose that, in fact, we have equality here, i.e., suppose that

"condgl (20) $$\lambda \in X^*\backslash 0 \quad \text{s.t.} \quad \alpha := \sup \lambda(M) = \inf \lambda(B_r(g)).$$

Then the hyperplane $H(\lambda, \alpha)$ has $M$ on its negative side and $B_r(g)$ on its positive side. For that reason, such $\lambda$ is called a **separating linear functional** for $M$ and $B_r(g)$. More than that, $H(\lambda, \alpha)$ is a **supporting hyperplane** for both sets, meaning that each set is on one side of it, but with 0 distance.

Finally, suppose that $m \in \mathcal{P}_M(g)$. Then

$$\begin{aligned} \lambda m \le \sup \lambda(M) = \inf \lambda(B_r(g)) \ &= \lambda g - \sup \lambda(B_r) \\ &= \lambda g - \|\lambda\|\, \|g - m\| \\ &\le \lambda g - \lambda(g - m) \ \le\ \lambda m, \end{aligned}$$

hence there must be equality throughout. In particular, then

"condba (21) $$\lambda m = \sup \lambda(M), \quad \lambda(g - m) = \|\lambda\|\, \|g - m\|.$$

One says that $\lambda$ is **parallel to** $g - m$, in symbols

$$\lambda \| g - m,$$

if both $\lambda$ and $g - m$ are nonzero and satisfy the second condition in (21). This language is derived from the special situation in a Hilbert space, since then $\lambda$ is necessarily of the form $\langle \cdot, y\rangle$ for some $y \in X$ and now $\lambda \| g - m$ implies that $y$ and $g - m$ are positive multiples of each other. Whether or not $X$ is a Hilbert space, the condition $\lambda \| g - m$ also says that $g - m$ is an **extremal** for $\lambda$, meaning that it is a nonzero vector at which $\lambda$ takes on its norm.

Either way, (21) provides very useful necessary conditions for $m \in M$ to be in $\mathcal{P}_M(g)$. Moreover, *these conditions must hold, not only for every $m \in \mathcal{P}_M(g)$ but also for every $\lambda$ satisfying (20), i.e., separating $M$ and $B_{\text{dist}\,(g,M)}(g)$.*

More than that, if $m \in M$ satisfies (21) for any such $\lambda$, then, for any $m' \in M$,

$$\|\lambda\|\|g - m\| = \lambda g - \lambda m \le \lambda g - \lambda m' \le \|\lambda\|\|g - m'\|,$$

hence (since $\lambda \ne 0$ by assumption), $m \in \mathcal{P}_M(g)$.

This proves

"propcharsep **(22) Proposition.** *Let $M$ be a closed subset of the nls $X$, $g \in X\backslash M$, hence $r := \text{dist}\,(g, M) > 0$, and let $m \in M$. If $\lambda$ separates $M$ and $B_r(g)$ (i.e., satisfies (20)), then $m \in \mathcal{P}_M(g)$ iff $m$ satisfies (21).*

In general there may be no separating linear functionals. However, if we know, in addition, that $M$ is convex, then the Separation Theorem assures us, for each $g \in X\backslash M$, of the existence of $\lambda$ satisfying (20) and, with that, we have proved “thmcharbafromconvex **(23) Characterization Theorem for ba from convex set.** *Let $M$ be closed, convex in the real nls $X$, let $g \in X\backslash M$ and $m \in M$. Then, $m \in \mathcal{P}_M(g)$ iff there is some $\lambda \| (g-m)$ with* $\sup \lambda(M) \le \lambda m$.

Since $\operatorname{dist}(g, M) = r = (\lambda g - \inf \lambda(B_r(g))/\|\lambda\|$, while, by (19), for any $\mu \in X^*$, $\mu g - \|\mu\| r = \inf \lambda(B_r(g)) \le \sup \lambda(M)$ with equality possible, it follows that

“dualdistance
$$\operatorname{dist}(g, M) = \max_{\mu \ne 0} (\mu g - \sup \mu(M))/\|\mu\|, \tag{24}$$

which provides a useful and sharp *lower bound* for the distance of $g$ 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 $\infty$. For the $\lambda$ in our theorem, this leaves only the value 0 (since $\sup \lambda(M)$ must equal $\lambda m$). Thus the condition $\sup \lambda(M) = \lambda m$ is replaced by the condition

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

“thmcharbaM **(25) Characterization Theorem for ba from lss.** *Let $M$ be a linear subspace of the nls $X$, let $g \in X$ and $m \in M$. Then, $m \in \mathcal{P}_M(g)$ iff there is some $\lambda \| (g-m)$ with $\lambda \perp M$.*

Since $\sup \lambda(M) \in \{0, \infty\}$ in case $M$ is a linear subspace, the lower bound (24) simplifies in that case to

“eqlowerboundls
$$\operatorname{dist}(g, M) = \max_{\lambda \perp M} \lambda g / \|\lambda\|. \tag{26}$$

## Construction of ba

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

In the best of circumstances, the norm in question is **smooth at** the point $(g-m)/\|g-m\|$, meaning that the condition $\lambda \| (g-m)$ determines $\lambda$ uniquely, up to nonnegative 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 $\lambda$ in the characterization theorem to be of norm 1. Hence, in smooth norms (i.e., norms that are smooth every boundary point of $B$, and for a finite-dimensional $M$, the characterizing conditions

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

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 $\ell_2$ or $L_2$), with inner product $\langle \cdot, \cdot \rangle$. In such a space, the statement

“eqextremal
$$0 < \lambda g = \|\lambda\| \, \|g\| \tag{27}$$

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

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

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

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

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

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

In $L_p$ for $p \neq 2$ but $1 < p < \infty$, the statement (27) still determines a unique $\lambda$ of norm 1, given in the form $\langle \cdot, g_\lambda \rangle$ for some $g_\lambda$ in the dual space, $L_{p^*}$, i.e., with $1/p + 1/p^* = 1$. But $g_\lambda$ depends *nonlinearly* on $g$, 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

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

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

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

In $L_\infty$, not only are ba's in general not unique, also (27) 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 $\lambda$ 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 consequence, of an observation during the proof of Proposition 22, is of use when the Characterization Theorem 25 is used in a nls with a non-smooth and/or non-strictly convex norm (such as $L_\infty$), in which case there may be many linear functionals of norm 1 parallel to the error and/or many ba's.

“lembasmustagree **(29) Lemma.** *$X$ nls, $M$ lss, $g \in X$, $m \in M$. If $M \perp \lambda \| (g-m)$, then, for any $m' \in \mathcal{P}_M(g)$, also $\lambda \| (g-m')$.*

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 $g$, must also be parallel to the error in any other ba to that $g$. 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 $C(T)$.

“thmextendtoCT **(30) Theorem.** *If $\lambda$ 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 $\mathcal{P}_M(g)$ involves linear functionals only as they act on $Y := M + \operatorname{ran}[g]$. Indeed, all the characterization demands is that $M \perp \lambda \| (g-m)$. 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 30. Since the original $\lambda$ took on its norm on $g - m \in Y$, therefore $\|\lambda|_Y\| = \|\lambda\|$, hence the replacement functional has the same norm as the original one. This gives the following.

“thmcharbaonCT **(31) 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 $g \in X$ and $m \in M$. Then, $m \in \mathcal{P}_M(g)$ iff there exists $U \subset T$ with $\#U \le n+1$ and $w \in \mathbb{R}^U$ so that $M \perp \lambda_{U,w} \| g - m$.*

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

“deLaValleePoissin **(32) de LaVallée-Poissin lower bound.** *If $0 \ne \lambda_{U,w} \perp M$ and $m \in M$ with $(g-m)(u)w(u) \ge 0$, all $u \in U$, then*

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

**Proof:** Set $\lambda := \lambda_{U,w}$ and recall, e.g. from (26), that $\lambda \perp M$ implies $|\lambda g| \le \|\lambda\| \operatorname{dist}(g, M)$, and certainly also $\lambda(g) = \lambda(g-m)$. Hence, with $e := g - m$, we have

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

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

This lower bound is the more effective, of course, the closer $\min|(g-m)(U)|$ is to $\|g-m\|$. 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 31.

The fact that the linear functional $\lambda := \lambda_{U,w}$ in the theorem is to take on its norm on the error, $e := g - m$, 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

“parconds (33) $$e(u) = \|e\| \operatorname{signum} w(u), \quad u \in U.$$

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

“eqMperp (34) $$w * Q_U V = 0,$$

with

$$Q_U : f \mapsto f|_U$$

and

$$V := [v_1, \ldots, v_n]$$

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 $g$ has its absolutely largest local extrema.

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

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

Suppose that $(a, r) \in \mathbb{R}^n \times \mathbb{R}$ solves the *linear* system

“eqbaonU (35) $$Q_U V a + r\sigma = Q_U g.$$

Then $|r| = \|g - Va\|_U$, and $\lambda(g - Va) = r\|w\|_1$, hence, by Theorem 31 applied to $C(U)$, $m_U := Va$ is then a ba to $g$ wrto the norm $\|\cdot\|_U$, with $\|g - m_U\|_U = |r|$. (If $r$ happens to be negative, simply replace $w$ by $-w$, as that won't change the fact that it is a solution of (35) but will change $r$ to $|r|$.)

(iii) *Update $U$, making it the set of $n+1$ points at which $g - m_U$ 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 (32)) that

$$|r| \le \operatorname{dist}(g, M) \le \|g - m_U\|,$$

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 $g$ 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 (35).

## The Haar condition

Since (35) is a square system, its solvability is equivalent to its unique solvability. Hence, as $g$ is arbitrary, we are asking that the pointset $U$ be total for $M$, i.e. $Q_U$ be 1-1 on $M$. But we are asking more, we are asking for the solvability of (35), i.e., for the invertibility of the matrix $[Q_U V, \sigma]$ and this is the same as its being 1-1 since it is square.

**(36) Proposition.** *If the $n+1$-set $U \subset T$ is total for the $n$-dimensional lss $M$ of $C(T)$, then (35) has exactly one solution (for every $g \in C(T)$).*

**Proof:** Let $w$ be a nontrivial solution of the homogeneous linear system (34), set $\sigma = \operatorname{signum}(w)$, and let $[Q_U V, \sigma](a, r) = 0$, with $a$ an $n$-vector and $r$ a scalar. Then $Va = -r\sigma$ on $U$, hence $0 = \lambda Va = \lambda Q_U Va = -r\|w\|_1$, which is only possible if $r = 0$. However, now $Q_U Va = 0$ follows which, by the assumption that $U$ is total for $M$, implies that $Va = 0$, therefore $a = 0$ (since $V$ is a basis). □

Since $Q_U$ maps into the $(\#U)$-dimensional space $\mathbb{R}^U$, it cannot be 1-1 unless $\#U \ge n = \dim M$. Further, if $U$ contains exactly $n$ points, then $Q_U$ also maps $M$ *onto* $\mathbb{R}^U$. This says that $M$ contains, for each $g$ defined at least on $U$, exactly one $m$ that **interpolates** $g$ **at** $U$, i.e., that agrees with $g$ on $U$.

**(37) 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 $M$.*

**(38) 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 $g \in C(T)$, and let $m \in \mathcal{P}_M(g)$ (such $m$ exists since $M$ is a finite-dimensional lss). By the Characterization Theorem 31, there exists $U \subset T$ with $\#U \le n+1$ and $w \in \mathbb{R}^U$, with $w(u) \ne 0$ for all $u \in U$, so that

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

Moreover, by Lemma 29, every $m \in \mathcal{P}_M(g)$ must satisfy this condition, hence must satisfy

$$Q_U m + \text{dist}\,(g, M)\sigma = Q_U g.$$

In particular, if also $m' \in \mathcal{P}_M(g)$, then $Q_U m = Q_U m'$. Now, since $M$ is Haar, the condition $0 \neq \lambda_{U,w} \perp M$ cannot hold unless $\#U > n$ (see Lemma 39 below). But, since $M$ is Haar, this implies that $U$ is total for $M$, therefore $m = m'$.

'$\Longleftarrow$': If $M$ fails to be Haar, then there exists an $n$-set $U$ in $T$ and some $m \in M$ with $\|m\| = 1$ that 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 $f$ in

$$G := \{g \in C(T) : \|g\| \le 1,\, Q_U g = \sigma := \text{signum}(w)\}$$

has 0 as a best approximation from $M$, and, by the Tietze Extension Theorem, there is at one such $f$. In fact, for any such $f$, also $g := (1 - |m|)g \in G$, since $\|m\| = 1$ and $m|_U = 0$. More than that, for any $\alpha \in [-1\,..\,1]$, $g - \alpha m \in G$ since (i) it agrees with $g$ on $U$, hence on $\lambda_{U,w}$, and (ii)

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

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

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

The proof took for granted the following

“cornpone **(39) Lemma.** *If the $n$-dimensional linear subspace $M$ of the $C(T)$ is Haar, then $0 \neq \lambda_{U,w} \perp M$ implies $\#U > n$.*

**Proof:** If $\#U < n+1$, then $Q_U V$ would have full row rank since we could then extend $U$ to an $n$-set $U'$ making $Q_U V$ a submatrix of the matrix $Q_{U'} V$ which is invertible since $M$ is Haar and, with that, $w * Q_U V = 0$ would imply $w = 0$. □

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, $\sigma := \text{signum}(w)$, since, as we saw earlier, the characterizing condition merely demands that the error, $e := g - m$, satisfy

“eqcharact (40)
$$e(u)\varepsilon = \|e\|\sigma(u), \quad u \in U,$$

with $\varepsilon$ some fixed (nontrivial) sign. We now investigate the possible sign vectors $\sigma$.

For this, we now think of $U$ as a sequence, i.e., order the points in $U$ somehow: $U = (u_1, \ldots, u_{n+1})$ and write, correspondingly, $w_j := w(u_j)$. 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

$$? * Q_{u_1,\ldots,u_n} V = Q_{u_{n+1}} V.$$

By Cramer's rule, this implies that

$$w_j = \det(Q_{u_1,\ldots,u_{j-1},u_{n+1},u_{j+1},\ldots,u_n}V)/\det(Q_{u_1,\ldots,u_n}V),$$

or, using column interchanges to restore order here,

“eqwoscil (41) $$w_j = (-1)^{n-j}\det(Q_{u_1,\ldots,u_{j-1},u_{j+1},\ldots,u_{n+1}}V)/\det(Q_{u_1,\ldots,u_n}V).$$

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 $\det(Q_{\tau(t)}V)$ 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 (41) have the same (positive or negative) sign*. In particular, in this case $w_j w_{j+1} < 0$, all $j$. This gives

“thmChebyshev **(42) Chebyshev's Alternation Theorem.** *If $M$ is an $n$-dimensional Chebyshev subspace of the real $C([a\,..\,b])$, then $m \in \mathcal{P}_M(g)$ iff the error, $g-m$,* **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*

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

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

**(43) 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.

"thmstrongunique **(44) Theorem (Newman-Shapiro).** *Best approximation from a finite-dimensional linear subspace $M$ of the real $X := C(T)$ is strongly unique. Explicitly, for every $g \in X$, there exists $\gamma > 0$ so that*

"eqwanted (45)
$$\forall\, f \in M \quad \|f - g\| \ge \|f - P_M g\| + \gamma \|f - P_M g\|.$$

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

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

hence

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

with

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

Since $K_\sigma$ is positive homogeneous, (45) follows with

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

It remains to show that $\gamma > 0$. Since $K_\sigma$ is continuous and $M$ is finite-dimensional, can find $f \in M$ with $\|f\| = 1$ and $\gamma = K_\sigma(f)$. Since $0 = \lambda_{U,w} f = \sum_{u \in U} w(u) f(u) = \sum_{u \in U} |w(u)| \sigma(u) f(u)$, $K_\sigma(f) = 0$ would imply $f(u) = 0$ for all $u \in U$. However, $M$ is Chebyshev, hence Haar, hence, by Corollary 39, $U$ is total for $M$, and now $f = 0$ would follow, contradicting the assumption that $\|f\| = 1$. □

## Complex $C(T)$

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

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

"lemdirectderiv **(46) Lemma.** *Let $X = C(T)$ with $T$ compact Hausdorff, and $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $e, f \in X\backslash 0$. Then, $\|e + \alpha f\| < \|e\|$ for some positive $\alpha$ iff*

$$K_e(f) := \max_{t \in \text{extr}(e)} \operatorname{Re}(e(t)\overline{f(t)}) < 0.$$

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

Conversely, if $K_e(f) < 0$, then

$$G := \{t \in T : \text{Re}(e(t)\overline{f(t)}) < K_e(f)/2\}$$

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

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

is less than $\|e\|$ for all $\alpha$ close to 0. But, by the strict negativity of $K_e(f)$, also

$$\|e+\alpha f\|^2(G) \le \|e\|^2 + 2\alpha K_e(f)/2 + |\alpha|^2\|f\|^2$$

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

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

"thmKolmogorov **(47) 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 $g \in X$ and $m \in M$. Then, $m \in \mathcal{P}_M(g)$ if and only if $\forall f \in M$, $K_{g-m}(f) \ge 0$.*

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 $g = 1/(\cdot - \alpha)$, 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

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

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

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

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

## $L_1$

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 $\ell_1$, 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.

"thmcharbaLone **(48) Characterization Theorem.** *Let $M$ be a linear subspace of the nls $X = L_1(T;\mu)$, with $\mu$ a non-atomic measure, let $g \in X$ and $m \in M$. Then, $m \in \mathcal{P}_M(g)$ iff there exists a $h \in L_\infty(T)$ with $\|h\|_\infty = 1$ that is perpendicular to $M$ and agrees with* $\operatorname{signum}(g-m)$ *off $Z_{g-m}$.*

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

The theorem suggests a quick try at constructing ba's from an $n$-dimensional lss of $X = L_1([a\,..\,b])$: Assuming that $Z_{g-m}$ 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 $g$ and the elements of $M$ are continuous (or, at least piecewise continuous), then we can now hope to *interpolate to $g$ at $\xi$ by elements of $M$*. Let $m$ be the resulting interpolant. Then, 'usually', the error, $g - m$, changes sign at the points of interpolation, hence 'usually', $\operatorname{signum}(g-m) = h$, and we have found a ba to $g$ 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.

"borsuk **(49) Borsuk's Theorem.** *Let $f$ be a continuous map from the unit sphere $S^n$ in $\mathbb{R}^{n+1}$ to $\mathbb{R}^n$. If $f$ is* **odd**, *i.e., $f(-x) = -f(x)$, all $x \in S^n$, then $0 \in f(S^n)$.*

"hobbyrice **(50) 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 $\lambda_s$ on $X$ as follows: With $\xi_j(s) := \sum_{i<j} s_i^2$,

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

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 $[0\,..\,1]^2 \to \mathbb{F} : (u,v) \mapsto \int_u^v f$ is continuous. Hence, with $V = [v_1, \ldots, v_n]$ any basis for $M$, the map

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

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

**(51) 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 $f \in M\backslash 0$ and set $g := h|f|$. Then, for any $\alpha \in [0\,..\,1]$,

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

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

Of course, this does *not* imply that *every* $g \in X\backslash M$ has many ba's from $M$. E.g., $g := \chi_{[\alpha..1]}$ has exactly one ba from $M := \Pi_0$, – except when $\alpha = 1/2$ in which case $[0\,..\,1]()^0 \subset \mathcal{P}_M(g)$.

## near-best approximation

While the questions concerning *best* approximations raised in the first lecture (such as existence, uniqueness and characterization) occupy a good part of a standard course in Approximation Theory, best approximations are hardly ever calculated because they can usually only be obtained as the limit of a sequence of approximations, and because of the easy availability of near-best approximations. To recall, the bounded linear map $A$ on the nls $X$ provides near-best approximations from the subset $M$ if

$$\|g - Ag\| \le \mathrm{const}\,\mathrm{dist}\,(g, M), \quad g \in X.$$

Such $A$ is, necessarily, a linear projector onto $M$, i.e., $\mathrm{ran}\, A = M$ and $A^2 = A$, i.e., $A|_M = 1$. In particular, $M$ is a linear subspace.

Conversely, if $A$ is a bounded linear projector on $X$ with range $M$, then $g - Ag = (1-A)g = (1-A)(g-m)$ for all $m \in M$, hence

$$(1) \qquad \mathrm{dist}\,(g, M) \le \|g - Ag\| \le \|1 - A\|\,\mathrm{dist}\,(g, M), \quad g \in X.$$

This is Lebesgue's inequality.

This raises (at least) two questions:

(i) How small can one make $\|1 - A\|$ by proper choice of $A$?

(ii) What can be said about $\operatorname{dist}(g, M)$ in terms of some information we might have about $g$?

**Example: broken line interpolation** With $X = C[a\,..\,b]$ for some finite interval $[a\,..\,b]$, recall the broken line interpolant $A = P_{\boldsymbol{\xi}} g$ for given $\boldsymbol{\xi} = (a = \xi_1 < \cdots < \xi_{l+1} = b)$. Its range is the space

$$M = \Pi^0_{1,\boldsymbol{\xi}}$$

of broken lines with break sequence $\boldsymbol{\xi}$. Since $\|P_{\boldsymbol{\xi}}\| = 1$, and always $\|1 - A\| \le 1 + \|A\|$, we get in this case

$$\operatorname{dist}(g, \Pi^0_{1,\boldsymbol{\xi}}) \le \|g - P_{\boldsymbol{\xi}} g\|_\infty \le 2 \operatorname{dist}(g, \Pi^0_{1,\boldsymbol{\xi}}).$$

In particular, the construction of a ba from $\Pi^0_{1,\boldsymbol{\xi}}$ will, at best, cut the maximum error in half.

Incidentally, the bound used here, $\|1 - P_{\boldsymbol{\xi}}\| \le 1 + \|P_{\boldsymbol{\xi}}\|$, is sharp as can be seen by looking at an $g$ that is the broken line with breakpoints $\xi_1, \ldots, \xi_{l+1}$ at which it has the value 1 and one additional breakpoint at which it has the value $-1$.

In particular, we can get a good estimate for $\operatorname{dist}(g, \Pi^0_{1,\boldsymbol{\xi}})$ by looking at $\|g - P_{\boldsymbol{\xi}} g\|_\infty$. In Lebesgue's proof of Weierstraß, we already observed that

$$\|g - P_{\boldsymbol{\xi}} g\|_\infty \le \omega_g(|\boldsymbol{\xi}|).$$

In particular, if $g \in L^{(1)}_\infty[a\,..\,b]$, i.e., $g$ is absolutely continuous with essentially bounded first derivative, then $\omega_g(h) \le h\|Dg\|_\infty$, hence

$$\|g - P_{\boldsymbol{\xi}} g\|_\infty \le |\boldsymbol{\xi}| \|Dg\|_\infty, \quad g \in L^{(1)}_\infty[a\,..\,b].$$

Further, since $P_{\boldsymbol{\xi}}$ reproduces all elements of $\Pi^0_{1,\boldsymbol{\xi}}$, we can replace here $g$ by an arbitrary element of $\Pi^0_{1,\boldsymbol{\xi}}$ and so obtain, more precisely,

$$\|g - P_{\boldsymbol{\xi}} g\|_\infty \le |\boldsymbol{\xi}| \operatorname{dist}(Dg, \Pi_{0,\boldsymbol{\xi}}), \quad g \in L^{(1)}_\infty[a\,..\,b].$$

Since $\operatorname{dist}(g, \Pi_{0,\boldsymbol{\xi}}) \le \omega_g(|\boldsymbol{\xi}|)$, we therefore obtain

$$\|g - P_{\boldsymbol{\xi}} g\|_\infty \le |\boldsymbol{\xi}| \omega_{Dg}(|\boldsymbol{\xi}|), \quad g \in C^{(1)}[a\,..\,b],$$

and, so, finally,

$$\|g - P_{\boldsymbol{\xi}} g\|_\infty \le |\boldsymbol{\xi}|^2 \|D^2 g\|_\infty, \quad g \in L^{(2)}_\infty[a\,..\,b].$$

Actually, since, for $\xi_j \le t \le \xi_{j+1}$, $g(t) - P_{\boldsymbol{\xi}} g(t) = (t - \xi_j)(t - \xi_{j+1})[\xi_j, \xi_{j+1}, t]g$ (with $[\alpha, \ldots, \omega]g$ the divided difference of $g$ at the point sequence $(\alpha, \ldots, \omega)$), we have the sharper estimate

$$\|g - P_{\boldsymbol{\xi}} g\|_\infty \le |\boldsymbol{\xi}|^2 \|D^2 g\|_\infty / 8, \quad g \in L^{(2)}_\infty[a\,..\,b].$$

You see how, as a function of the mesh size, $|\boldsymbol{\xi}|$, these estimates improve, i.e., go to zero faster as $|\boldsymbol{\xi}| \to 0$, when we restrict $g$ to smoother and smoother function classes. However, further more restrictive smoothness assumptions will not lead to an increase in the rate at which the interpolation error goes to zero with $|\boldsymbol{\xi}|$. E.g., if $\|g - P_{\boldsymbol{\xi}} g\| = o(|\boldsymbol{\xi}|^2)$ for arbitrary $\boldsymbol{\xi}$ (or even just for $\boldsymbol{\xi}_N = (a, a+h, \cdots, b-h, b)$ with $h := (b-a)/N$ and $N \in \mathbb{N}$) as $|\boldsymbol{\xi}| \to 0$, then $g \in \Pi_1$. Proof idea: for any collection $\Xi$ of 'quasi-uniform' $\boldsymbol{\xi}$, i.e., $\sup_{\boldsymbol{\xi}} \sup_{i,j} \Delta\xi_i/\Delta\xi_j < \infty$, must have $[\xi_j, t, \xi_{j+1}]g \to 0$ uniformly for $t \in (\xi_j \,.\,.\, \xi_{j+1})$ and $j$ as $|\boldsymbol{\xi}| \to 0$ while, for any refinement $s$ of the sequence $(a, t, b)$, the second divided difference $[a, t, b]g$ can be written as a *convex* combination of the $[s_i, s_{i+1}, s_{i+2}]g$, hence must be zero.

This is a simple illustration of our next topic, degree of approximation.

If there is time, I may come back to item (i), i.e., the question of how small one can make $\|A\|$ or $\|1 - A\|$ for given $M$ by appropriate choice of the linear projector $A$ onto $M$. To whet your appetite, I mention that every linear projector on $C[a\,.\,.\,b]$ onto $\Pi_n$ has norm no smaller than $\sim \ln n$. This is related to the fact that the projector $s_n$ providing the truncated Fourier series (mentioned earlier) is the unique projector of minimal norm on $C(\mathbb{T})$ onto $\overset{\circ}{\Pi}_n$, and $\|s_n\| \sim (2/\pi)\ln n + \text{const}$.

**Remark** I am getting tired of adapting earlier notes by making sure that the given element of $X$ to be approximated is denoted by $g$. So, from now on, the given element to be approximated will be denoted by $f$. Life is short.

## Degree of Approximation

Given a *sequence* $(M_n) = (M_n : n \in \mathbb{N})$ of approximating sets in the nls $X$ and $f \in X$, one is interested in

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

as $n \to \infty$.

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

**(i)** $M_1 = \{0\}$, *and* $(M_n)$ *is increasing,* i.e., $(M_n)$ is a **ladder**.
This guarantees that $E_n(f)$ is nonincreasing, with $E_1(f) = \|f\|$.
**(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)** $\mathbb{F}M_n := \{\alpha m : \alpha \in \mathbb{F}, m \in M_n\} \subseteq M_n$.
**(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.

To be sure, Assumption (iv) is somewhat restrictive and precludes some practically important ladders such as the following: for some $T \subset \mathbb{R}^2$,

$$M_n := \Pi_{k,n}|_T$$

consists of all piecewise polynomial functions, on $T$, of some degree $k$ and involving no more than $n$ polynomial pieces. In other words, for each $f \in \Pi_{k,n}$ there is a partition of $T$ into at most $n$ subsets and, on each such subset, $f$ agrees with some polynomial of degree $\le k$. For this example, we have only $M_n + M_n \subset M_{n^2}$. Finding the degree of approximation from this ladder $(\Pi_{k,n} : n \in \mathbb{N})$ is one of the outstanding problems in nonlinear approximation. (Strictly speaking, this particular example lacks that trivial initial space, $\{0\}$, which is really only used in the general theory to simplify notation. For that, we might have to set here $M_n = \Pi_{k,n-1}$. Such a switch, from $n$ to $n \pm 1$, does not change the orders $n^{-r}$ of interest here. Also, its 'subladder' $(\Pi_{k,2^n} : n \in \mathbb{N})$ does satisfy (i)-(v).)

Even 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 $\alpha$. Thus we are looking for conditions on $f$ that 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., $\lim_n E_n(f) n^r = 0$, or

$$\sum_n n^r E_n(f) < \infty.$$

Such conditions may single out a rather 'small' subset of $X$ in case $X$ is complete but not equal to any of the $M_n$. This is certainly so if the constant in (iv) is $c = 1$, i.e., in case the $M_n$ are linear subspaces.

**(52) Proposition (H. S. Shapiro).** *If $(M_n)$ is a sequence of proper closed linear subspaces (i.e., only (iii) as is, (i) is weakened to 'proper', and (v) is weakened to 'closed', but (iv) with $c = 1$), then, for any real sequence $(\alpha_n)$ 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.*

**Proof:** $A_\alpha = \cup_{n_0, N \in \mathbb{N}} \cap_{n \ge n_0} B^-_{N\alpha_n}(M_n)$, with

$$B^-_{Nr}(M_n) := \{f \in X : \mathrm{dist}\,(f, M_n) \le Nr\} = N B^-_r(M_n)$$

since $\mathbb{F}M_n \subset M_n$. If now, for some $r > 0$ and some $x$, $B_r(x) \subset B_{\alpha_n}(M_n)$, then $B_r(0) \subseteq (B_r(x) + B_r(-x))/2 \subset (B_{\alpha_n}(M_n) + B_{\alpha_n}(-M_n))/2$ and, even if we only knew (iv), we could now conclude that $B_r(0) \subset B_{\alpha_n}(M_{cn})$ and if, as we actually assume, the $M_n$ are proper closed linear subspaces, this implies (by Riesz' Lemma) that $r \le \alpha_n$, hence $B_r(x) \subset \cap_{n \ge n_0} B^-_{\alpha_n}(M_n)$ implies $r \le \lim_{n\to\infty} \alpha_n = 0$, hence $A_\alpha$ is of first category. However, if the $M_n$ are not linear spaces, then no such conclusion can be drawn. □

**Example** Here is a nice example, provided by Olga Holtz, to show that properties (i)-(v) by themselves are not strong enough to imply this proposition's conclusion. Let $X := \ell_\infty$. For any closed $V \subset \mathbb{F} := \mathbb{R}$, let

$$F_V := \{f \in X : \operatorname{ran} f \subset V\}.$$

Then, for any $f \in X$, $\operatorname{dist}(f, F_V) = \|f - f_V\|$, with $f_V : n \mapsto \operatorname{argmin}_{v\in V} |f(n) - v|$. Therefore, for any closed $W$,

$$\operatorname{dist}(f, F_V) \le \operatorname{dist}(f, F_W) + \operatorname{dist}(W, V).$$

Let now

$$F_k := \cup_{\#V \le k} F_V,$$

and let $(f_m)$ be a minimizing sequence in $F_k$ for $\|f - \cdot\|$. Then $(\operatorname{ran} f_m)$ is a bounded sequence of subsets in $\mathbb{R}$ of cardinality $\le k$, hence, AGTASMAT (:= After Going To A Subsequence May Assume That) there is some $V \subset \mathbb{R}$ with $\#V \le k$ for which $\lim_{m\to\infty} \operatorname{dist}(\operatorname{ran} f_m, V) = 0$. Therefore

$$\operatorname{dist}(f, F_k) \le \operatorname{dist}(f, F_V) \le \operatorname{dist}(f, F_{\operatorname{ran} f_m}) + \operatorname{dist}(\operatorname{ran} f_m, V) \xrightarrow[m\to\infty]{} \operatorname{dist}(f, F_k),$$

showing $f_V$ to be a ba to $f$ from $F_k$, hence showing $F_k$ to be an existence set. Also, with $V_k := \{1 - (2j-1)/k : j = 1, \ldots, k\}$, we have

"eqolga (53)
$$\operatorname{dist}(f, F_k) \le \operatorname{dist}(f, F_{\|f\|V_k}) \le \|f\|/k,$$

and, since also $F_k$ is homogeneous, and is increasing with $k$, the sequence $(F_k)$ satisfies conditions (i)-(v), except that, at best, $F_k + F_k \subset F_{k^2}$. But (iv) is satisfied (with $c = 2$) by its subsequence $M_n := F_{2^n}$, $n = 1, 2, \ldots$, (with $M_0 = \{0\}$, of course), along with the other conditions in (i)-(v), yet, by (53),

$$\{f \in X : \operatorname{dist}(f, M_n) = O(2^{-n})\} = X,$$

and this is not of first category since $X$ is a Bs.

Put positively, this example also shows the power of *non*linear approximation, i.e., approximation from nonlinear $M_n$. □

## Degree of Approximation quantified

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 $\|\cdot\|_Y$ that is continuously imbedded in $X$. The standard example has $M_n = \mathring{\Pi}_n \subset X := X_p$, with

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

and

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

with $\rho$ some positive integer.

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

“jackson (54) $$\exists\{C_J\}\forall\{n \in \mathbb{N},\ f \in Y\}\ \ E_n(f) \le C_J\, n^{-r}\|f\|_Y$$

and

“bernstein (55) $$\exists\{C_B\}\forall\{n \in \mathbb{N},\ g \in M_n\}\ \|g\|_Y \le C_B\, n^r\|g\|.$$

For historical reasons, the former is called a **Jackson**, or **direct**, inequality or estimate, the latter a **Bernstein**, 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 (historically earlier!) Bernstein inequality provides an upper bound, at least indirectly. The most natural bridge between Jackson and Bernstein 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).

“thmpeetre **(56) Peetre's Theorem.**
*(i) Jackson (54) implies* $\sup_n E_n(f)/K(f,n^{-r}) < \infty$.
*(ii) Bernstein (55) implies*

“peetrebernstein (57) $$K(f,n^{-r}) \le \mathrm{const}_r\, n^{-r}\|(E_k(f) : k \le n)\|^{(r)}.$$

Here, we use 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

"standard (58)
$$\|(k^{-(r-1+\alpha)} : k \le n)\|^{(r)} = \sum_{k=1}^{n} k^{-\alpha} \sim n^{1-\alpha} \quad \text{for } 0 \le \alpha < 1,$$

with the convenient abbreviation

$$b_n \sim c_n \quad :\Longleftrightarrow \quad b_n = O(c_n) \text{ and } c_n = O(b_n).$$

(In fact, this suggests the nonstandard abbreviation:

$$b_n \lesssim c_n \quad :\Longleftrightarrow \quad b_n = O(c_n),$$

which I will use occasionally, e.g., right now.) Here is a proof of (58):

$$n^{1-\alpha} \lesssim \int_1^{n+1} ()^{-\alpha} \le \sum_{k=1}^{n} k^{-\alpha} \le \int_0^n ()^{-\alpha} \lesssim n^{1-\alpha}.$$

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 $\mu$ and all $x$ and $y$. In such a case, one considers, more generally, the **weighted sequence norm**

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

With these notations now fully explained, we read off from Peetre's theorem that, for $\alpha \in [0\,..\,1)$, $E_n(f) \lesssim n^{-(r-1+\alpha)}$ if and only if $K(f, n^{-r}) \lesssim n^{-(r-1+\alpha)}$. 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 $\mathrm{Lip}_\alpha$.

**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 $k$, we choose a sequence

$$1 = m_0 < m_1 < \cdots < m_\nu = n$$

and consider

$$h_\mu := f_{m_\mu} - f_{m_{\mu-1}} \in M_{cm_\mu}.$$

Then

$$\|h_\mu\| \le \|f - f_{m_\mu}\| + \|f - f_{m_{\mu-1}}\| \le 2E_{m_{\mu-1}}(f),$$

while $f_n = \sum_{\mu=1}^{\nu} h_\mu$. Therefore,

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

with $\mathrm{const}_r := 2C_B c^r\ \max_\mu (m_\mu/m_{\mu-1})^r$ a constant depending only on $r$ (and the constants $C_B$ and $c$) if we restrict attention to sequences $(m_\mu)$ for which $\max_\mu m_\mu/m_{\mu-1} \le \mathrm{const}$.

Now, for any positive nonincreasing sequence $(a_k)$ (such as $a_k = E_k(f)$),

$$a_J \sum_{j<k\le J} k^{r-1} \le \sum_{j<k\le J} k^{r-1} a_k \le a_{j+1} \sum_{j<k\le J} k^{r-1},$$

while

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

Hence, applying this with $j = m_\mu$, $J = m_{\mu+1}$, all $\mu$, we get

$$\sum_{\mu=0}^{\nu-1} m_\mu^r E_{m_\mu}(f) \sim \mathrm{const}_r \sum_{k=1}^{n} k^{r-1} E_k(f)$$

provided we also insist that $\min_\mu (m_\mu/m_{\mu+1}) \le \mathrm{const} < 1$, i.e., the ratio $m_\mu/m_{\mu-1}$ required earlier to be bounded must, on the other hand, be bounded away from 1. Choosing the sequence $(m_\mu : \mu = 1, \ldots, \nu)$ to be roughly dyadic, i.e., $m_\mu \sim 2^{\mu-1}$, all $\mu$, will do for both requirements. With this, we get (57), with const depending on $C_B$ and $r$ only. □

For later use, I record the following just proved.

“tightestimate **(59) Lemma.** *If $a : \mathbb{N} \to \mathbb{R}_+$ is nonincreasing, then, for any $r \ge 1$,*

$$\sum_{k=1}^{2^n} k^{r-1} a_k \ \sim\ \sum_{\mu=1}^{n} 2^{\mu r} a_{2^\mu},$$

*with constants that only depend on $r$.*

We will make use of (59) in a moment, in Bernstein's proof that only smooth functions can be approximated well by trigonometric polynomials.

### Bernstein estimates for trig.pols.

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 $o(n^{-r}) \neq E_n(f) = O(n^{-r})$. 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 Bernstein, using

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

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

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

“szegoineq (60) $$(Df)^2 + (nf)^2 \le (n\|f\|)^2, \quad f \in \mathring{\Pi}_n.$$

Since $\mathring{\Pi}_n$ is translation-invariant and differentiation commutes with translation, it is sufficient to prove that

“szegoineqatzero (61) $$(Df(0))^2 + (nf(0))^2 \le (n\|f\|)^2, \quad f \in \mathring{\Pi}_n.$$

As the inequality is homogeneous in $f$, we may assume that $Df(0) \ge 0$. Now let $r > \|f\|$, let $\alpha$ be the unique point in $(-\pi \,..\, \pi)/(2n)$ at which $r \sin n\alpha = f(0)$, and consider the trig.pol.

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

Since $\|f\| < r$, $s$ 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 $\alpha$ to make it so. If now $Df(0) > D(r \sin n(\cdot + \alpha))(0)$, then $f$ would rise above $r \sin n(\cdot + \alpha)$ 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 $\sin n(\cdot + \alpha)$ and this would imply an impossible second zero of $s$ between the two extrema that bracket 0. Consequently,

$$0 \le D(r \sin n(\cdot + \alpha))(0) = Df(0) \le rn \cos n\alpha = rn\sqrt{1 - (f(0)/r)^2},$$

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

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

and, since $r > \|f\|$ is arbitrary here, (61) follows. □

We are ready to prove a sample inverse theorem.

“thmbernstein **(62) Theorem.** *Let* $M_n = \overset{\circ}{\Pi}_n$*, as a subset of* $C(\mathbb{T})$*. Let* $f \in C(\mathbb{T})$ *and assume that, for some* $r$*,*

“boundednormr (63)
$$\|(E_n(f) : n \in \mathbb{N})\|^{(r)} < \infty$$

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

**Proof:** Since differentiation is a closed operator, it is sufficient to show that $f$ is the uniform limit of a sequence that 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.

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

the last inequality by (59). By assumption, this last sum goes to zero as $j, J \to \infty$. □

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)

“algebJackson (64)
$$\mathrm{dist}_\infty(f, \Pi_n) =: E_n(f)_\infty \le \mathrm{const} n^{-1} \|Df\|_\infty.$$

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

“markovineq (65)
$$\|Df\|_\infty \le \frac{2}{|b-a|} n^2 \|f\|_\infty,$$

(note that, because of the possibility of dilating, the interval length must figure in exactly the position at which it appears, hence the inequality is sharp since it becomes equality for $[a\,..\,b] = [-1\,..\,1]$ and $f : t \mapsto \cos(n \cos^{-1} t)$ the Chebyshev polynomial of degree $n$) and these two inequalities do not at all match in the sense of Theorem 56. 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 $E_n(f)_\infty = O(n^{-r-\varepsilon})$ for some positive $\varepsilon$.

## Jackson estimates for trig.pols.

Because $\mathring{\Pi}_n$ is translation-invariant, it is easier to prove Jackson inequalities for $M_n = \mathring{\Pi}_n \subset C(\mathbb{T})$ than for $\Pi_n \subset C([a\,..\,b])$ ($\Pi_n$ is translation-invariant only on $\mathbb{R}$ or $\mathbb{C}$). The standard argument uses integral operators of **convolution type**. With $L_n$ functions still to be specified, one considers the approximation

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

One makes the following assumptions:

(i) $L_n \in \mathring{\Pi}_n$.

Then, $\mathring{\Pi}_n$ being **translation-invariant**, i.e., $\forall\{p \in \mathring{\Pi}_n,\ s \in \mathbb{T}\}\ p(\cdot - s) \in \mathring{\Pi}_n$, 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 $\mathring{\Pi}_n$. Therefore

$$L_n{*}f = \sum_j \varphi_j \int_{\mathbb{T}} \psi_j f \quad \in \mathring{\Pi}_n.$$

(ii) $\lambda_n := \int_{\mathbb{T}} L_n \ne 0$, hence wlog (i.e., after replacing $L_n$ by $L_n/\lambda_n$, $\int_{\mathbb{T}} L_n = 1$, therefore $L_n{*}()^0 = ()^0$.

It follows that

$$(f - L_n{*}f)(t) = \int_{\mathbb{T}} \big(f(t) - f(t-s)\big)L_n(s)\,\mathrm{d}s = \int_{\mathbb{T}} (\nabla_s f)(t)\,L_n(s)\,\mathrm{d}s,$$

with

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

Note that

$$\|\nabla_s f\| \le \sup_{0\le h<|s|} \|\Delta_h f\| =: \omega(f,|s|),$$

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

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

(using that $\omega(f,\cdot)$ is nondecreasing and subadditive, hence $\omega(f,|s|) = \omega(f,(n|s|)/n) \le \omega(f,\lceil n|s|\rceil/n) \le \lceil n|s|\rceil\omega(f,1/n) \le (1+n|s|)\omega(f,1/n)$), hoping that

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

However, the $L_n$ usually employed satisfy the following additional assumption:
(iii) $L_n$ even.
and, with this, we even have

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

with

$$\Delta_h \nabla_h f = \Delta^2 f(\cdot - h),$$

and

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

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 $r$th modulus of smoothness, for any natural $r$, as are the bounds

$$\omega_r(f,h) := \sup_{k<h} \|\Delta_k^r f\| \le 2^r \omega(f,h)$$

and

$$\omega_r(f,|s|) \le (1+n|s|)^r \omega_r(f,1/n),$$

with the latter using the fact that $\Delta_{mh} f = \sum_{j=0}^{m-1} \Delta_h f(\cdot + jh)$, hence

$$\Delta_{mh}^r f = \sum_{j_1=0}^{m-1} \cdots \sum_{j_r=1}^{m-1} \Delta_h^r f(\cdot + (j_1 + \cdots + j_r)h).$$

Since, in particular, $\omega_2(f,h) = \omega_2(f,(nh)/n) \le (1+nh)^2 \omega_2(f,1/n)$, a natural assumption now is the following **moment condition**:
(iv) $\sup_n \int_0^\pi (ns)^k \,|L_n(s)|\ \mathrm{d}s \ < \infty$, i.e., $\int_0^\pi ()^k |L_n| = O(n^{-k})$, for $k = 0,1,2$.
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 $L_n = D_n / \int_{\mathbb{T}} D_n$ with

$$D_n(t) := \sum_{|j|\le n} e_{\mathrm{i}j}/2 = \frac{\sin((n+1/2)t)}{2\sin(t/2)}$$

**Dirichlet's kernel**. For it, $L_n \in \mathring{\Pi}_n$, $D_n{*}()^0 = \pi$, and $L_n$ even. However, $L_n$ decays too slowly away from 0 to satisfy (iv); in fact, already $k = 0$ gives trouble since $\|D_n\| \sim$

$\int_{\mathbb{T}} |D_n| \sim \int_{1/n}^{\pi} ()^{-1} \sim \ln n$, hence, as already mentioned earlier, $L_n*$ cannot converge to 1 pointwise on $C(\mathbb{T})$, hence a more sophisticated analysis wouldn't help here, either.

The next choice is $L_n = F_n / \int_{\mathbb{T}} F_n$, 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

$$F_n = \sum_0^n D_k/(n+1),$$

hence $F_n \in \mathring{\Pi}_n$, and $F_n*()^0 = \pi$. Also, $F_n \ge 0$, hence we 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**

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

normalized to have $\int_{\mathbb{T}} J_n = 1$. It is a special case of the **generalized Jackson kernel**s:

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

Since

$$\left( \frac{\sin(mt/2)}{\sin(t/2)} \right)^2 / m = 1 + 2 \sum_{k=1}^{m-1} (1 - k/m) \cos kt = \sum_k B_2((k/m)+1) e_{\mathrm{i}k}(t)$$

(as one verifies, with $B_2$ the 'cardinal B-spline of order 2', a fact to be explored later), it follows that $J_{n,r}$ is an even, nonnegative, trig.pol. of degree $\le n$. In particular $J_{n,r}*()^0 > 0$, hence the 0th moment condition is trivially satisfied. For the others, one may prove that $J_{n,r}$ satisfies the moment conditions

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

as follows.

Since $t/\pi \le \sin(t/2) \le t/2$ on $[0\,..\,\pi]$,

$$\begin{aligned}\int_0^\pi t^k\Big(\frac{\sin(mt/2)}{\sin(t/2)}\Big)^{2r}\,\mathrm{d}t \sim \int_0^\pi t^k\Big(\frac{\sin(mt/2)}{t}\Big)^{2r}\,\mathrm{d}t \\ = \Big(\frac{2}{m}\Big)^{k-2r+1}\int_0^{m\pi/2} u^k\Big(\frac{\sin u}{u}\Big)^{2r}\,\mathrm{d}u \\ \sim n^{2r-1-k}\int_0^\infty u^k\Big(\frac{\sin u}{u}\Big)^{2r}\,\mathrm{d}u \\ \sim n^{2r-1-k}\end{aligned}$$

as long as $k-2r<-1$, hence, in particular, $\lambda_{n,r}\sim n^{2r-1}$ and so

$$\int_0^\pi (ns)^k J_{n,r}(s)\,\mathrm{d}s \sim n^k n^{2r-1-k}/n^{2r-1}\sim 1,$$

for $k=0,\ldots,2r-2$.

Thus, for $r\ge 2$, the sequence $L_n = J_{n,r}$ 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 the inequalities (due to Stechkin):

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

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

The constant in the resulting estimate $E_n(f)\le \mathrm{const}_{2r-2}\|D^{2r-2}f\|_p/n^{2r-2}$ is far from sharp. Favard showed that, with $S$ the unit ball of the semi-normed lss $Y := \mathbf{L}_p^{(r)}(\mathbb{T})$ 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.

We now know that (54)Jackson and (55)Bernstein hold for $M_n = \mathring{\Pi}_n$ and $Y = C^{(r)}(\mathbb{T})\subset C(\mathbb{T}) = X$. Hence, (56) Peetre's Theorem now tells us that

$$K(f,n^{-r};C(\mathbb{T}),C^{(r)}(\mathbb{T}))\sim E_n(f)\sim \|D^r f\|_\infty/n^r,\quad f\in C^{(r)}(\mathbb{T}).$$

## Jackson estimates for alg.pols.

It is standard to prove Jackson's theorem for *algebraic* polynomials from the one for trigonometric polynomials. For this, one first translates and dilates $[a\,..\,b]$ 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 that carries $C[-1\,..\,1]$ isometrically into

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

the space of all **even** $2\pi$-periodic functions. In particular, $F$ maps $\Pi_n$ onto

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

Also,

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

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

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

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

“jacksonforalgpols (66) $$\operatorname{dist}(f, \Pi_n) \le \|f - F^{-1}m\| = \operatorname{dist}(Ff, \mathring{\Pi}_n) \le \text{const}\ \omega(f, 1/n),$$

*provided* $F^{-1}m$ is defined and in $\Pi_n$. 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 $g \in C(\mathbb{T})$ and even (as is the case for $g = Ff$) and $m \in \mathcal{P}_{\mathring{\Pi}_n}(g)$, then $m_e$ is an even trig.pol. in $\mathring{\Pi}_n$ and $\|Ff - m_e\| = \|g_e - m_e\| \le \|g - m\| = \operatorname{dist}(g, \mathring{\Pi}_n)$, hence also $m_e \in \mathcal{P}_{\mathring{\Pi}_n}(g)$, i.e., we may take $m \in \mathcal{P}_{\mathring{\Pi}_n}(Ff)$ to be even, hence in $F(\Pi_n)$. In fact, $\Pi_n$ is Haar, hence we have just proved that the ba from $\mathring{\Pi}_n$ 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.

**(67) 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,

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

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

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

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

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

hence

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

Thus, by induction, for $n \gg r$,

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

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

"thmKeqomega **(68) 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* $f \in \mathbf{L}_p(T)$*,*

$$c\omega_r(f,t)_p \le K(f, t^r; \mathbf{L}_p, W_p^{(r)}) \le C\omega_r(f,t)_p, \quad t > 0. \tag{69}$$

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.

## Good approximation, especially by splines

As we have seen, a process of near-best approximation, from some subset $M$ of the nls $X$, is necessarily a projector (i.e., idempotent). It is particularly easy to construct and use if $M$ is a finite-dimensional linear subspace of $X$ in which case the projector can even be chosen to be linear.

Linear projectors onto finite-dimensional linear subspaces arise naturally when considering the information that is readily available in the representation of $m \in M$ with respect to a particular basis for $M$, i.e., any particular 1-1 linear map

$$V : \mathbb{F}^n \to X : a \mapsto \sum_j a(j)v_j =: [v_1, \ldots, v_n]a$$

with $\operatorname{ran} V = M$.

Given such $V$, the abstract equation $V? = m$ for the coefficients $V^{-1}m$ of $m \in M$ wrto $V$, is turned into an equivalent numerical equation $\Lambda' V? = \Lambda' m$ by any linear map $\Lambda' : X \to \mathbb{F}^n$ that is 1-1 on $M$. For, with that assumption, $\Lambda' M$ maps $\mathbb{F}^n$ linearly and 1-1 into itself, hence is an invertible matrix, and this gives the solution

$$V^{-1}m = (\Lambda' V)^{-1}\Lambda' m.$$

Now notice that we may assume, in addition, that $\Lambda' V = 1$. For, if our initial choice of $\Lambda'$ does not satisfy this, simply replace it by $(\Lambda' V)^{-1}\Lambda'$. With this additional assumption, $m = V\Lambda' m$ for any $m \in M$, i.e., $\Lambda' m$ is the desired coordinate vector. But any *linear* map $\Lambda' : X \to \mathbb{F}^n$ is necessarily of the form

$$\Lambda' f =: (\lambda_i f : i = 1, \ldots, n) =: [\lambda_1, \ldots, \lambda_n]' f$$

for certain linear functionals $\lambda_i$. Hence, we conclude that, for $m = Va$, the coefficient $a(j)$ gives us the value $\lambda_j m$ of the linear functional $\lambda_j$ at $m$.

Of course, if $M$ is a proper subspace of $X$, then there is nothing unique about these $\lambda_j$. Each choice of $\Lambda'$ with $\Lambda' V = 1$ corresponds to a particular extension of the coordinate functionals for the basis $V$. In that sense, $a(j) = (V^{-1}m)(j)$ gives us all kinds of information about $m$.

Assuming merely that $\Lambda'$ is 1-1 on $M$, it follows that, for any $f \in X$,

$$m = Pf := V(\Lambda' V)^{-1}\Lambda' f$$

is the unique element of $M$ that agrees with $f$ on $\Lambda'$ in the sense that

$$\Lambda' m = \Lambda' f.$$

In particular, $P$ is a linear projector, onto $M$, and every linear projector onto $M$ arises in this way.

## An example: polynomial interpolation and divided differences

Consider the map

$$W_c : \mathbb{F}_0^{\mathbb{N}} \to \Pi : a \mapsto \sum_{j=1}^{\infty} a(j) w_j,$$

with

"newtonpols (70) $$w_j : t \mapsto \prod_{i<j} (t - c_i), \quad j = 1, 2, \ldots,$$

the **Newton polynomial**s for the (arbitrary) sequence $c : \mathbb{N} \to \mathbb{F}$, and

$$\mathbb{F}_0^T := \{a \in \mathbb{F}^T : \#\,\mathrm{supp}(a) < \infty\}.$$

Since any $a \in \mathbb{F}_0^{\mathbb{N}}$ has only finitely many nonzero entries, $W_c a$ is, indeed, well-defined for any such $a$. Also, since $\deg w_j = j - 1$ for all $j$, $W_c$ is necessarily 1-1. It is also onto since its range contains all the $w_j$ and, for each $n$, $(w_1, \ldots, w_n)$ is linearly independent and in the $n$-dimensional space $\Pi_{<n}$, hence a basis for it.

Thus, each $p \in \Pi$ can be uniquely written as $W_c a$, and this particular representation for $p$ is called its **Newton form with (respect to) centers** $c$.

The question now is: what is $W_c^{-1} p$ for given $p \in \Pi$? Related to this is the question: What information about $p = W_c a$ is readily provided by the coefficient $a(n)$?

To answer these questions, observe that

$$p = W_c a =: p_n + w_{n+1} q_n,$$

with

$$p_n := \sum_{j \le n} a(j) w_j,$$

and with $q_n$ some polynomial. It follows that $p_n$ is the remainder after division of $p$ by $w_{n+1}$, i.e., $p_n \in \Pi_{<n}$ and $w_{n+1}$ divides $p - p_n$. But this uniquely pins down $p_n$: Indeed, if also $r \in \Pi_{<n}$ and also $w_{n+1}$ divides $p - r$, then $w_{n+1}$ also divides $p_n - r$. But since $\deg(p_n - r) < n = \deg w_{n+1}$, this implies that $p_n = r$.

Using once more the fact that $\deg w_j = j - 1$ for all $j$, it follows that $a(n)$ *is the* **leading coefficient**, *i.e., the coefficient of* $()^{n-1}$, *in the power expansion*

$$p_n = a(n)()^{n-1} + \text{l.o.t.} =: P_{c_1,\ldots,c_n} p$$

*of the unique polynomial* $p_n \in \Pi_{<n}$ *that* **agrees with** $p$ **at** $c_1, \ldots, c_n$ *in the sense that* $p - p_n$ *vanishes at that sequence, counting multiplicities.* Note that $p_n$ agrees with $p$ at $c_1, \ldots, c_n$ exactly when

"interpatc (71) $$D^\rho (p - p_n)(z) = 0, \quad 0 \le \rho < \#\{1 \le i \le n : z = c_i\}, \;\; z \in \mathbb{F}.$$

We now know that $a(n)$ only depends on $p$ 'at' $c_1, \ldots, c_n$, hence a notation like $a(n) = a(p; c_1, \ldots, c_n)$ would be appropriate. In fact, here are three notations,

“defdivdif (72)
$$a(n) = p[c_1, \ldots, c_n] = [c_1, \ldots, c_n]p = \Delta(c_1, \ldots, c_n)p,$$

the first two quite standard. The last, due to V. Kahan, will be adopted here, since $[c_1, \ldots, c_n]$ is otherwise occupied.

The view of $\Delta(c_1, \ldots, c_n)p$ as the leading coefficient in the power form for the interpolating polynomial $P_{c_1,\ldots,c_n}p$ often is the fastest way to specific results concerning divided differences. For example, if $c_1 < \cdots < c_n$, then, by Rolle's Theorem and for $j = 1{:}n-1$, $D(p - P_{c_1,\ldots,c_n}p)$ must vanish at some $\zeta_j \in (c_j \,..\, c_{j+1})$. It then follows that $DP_{c_1,\ldots,c_n}p = P_{\zeta_1,\ldots,\zeta_{n-1}}Dp$ and so, in particular,

$$(n-1)\Delta(c_1, \ldots, c_n)p = \Delta(\zeta_1, \ldots, \zeta_{n-1})Dp.$$

By induction, this gives that *if the $c_j$ are real, then there is $\xi$ in the smallest interval containing $c_1, \ldots, c_n$ so that*

“eqmeanvalue (73)
$$(n-1)!\,\Delta(c_1, \ldots, c_n)p = D^{n-1}p(\xi).$$

Consider now the computation of specific divided differences. If $c$ is a constant sequence, $c = (\zeta, \zeta, \ldots)$ say, then, by (71) or (73),

$$\Delta(\underbrace{\zeta, \ldots, \zeta}_{n \text{ terms}})p = D^{n-1}p(\zeta)/(n-1)!.$$

For general $c$, the answer is a little bit more subtle. However, we already observe the very important fact that $W_c a$ is a continuous, in fact an infinitely differentiable, function of $c$, hence so is $\Delta(c_1, \ldots, c_n)p$ for each $n$. Indeed we get, on differentiating the identity $W_c W_c^{-1} = 1$ as a function of $c$ and rearranging, that

$$D(W_c^{-1}) = -W_c^{-1} D(W_c) W_c^{-1}$$

is continuous and smooth.

Further, *$\Delta(c_1, \ldots, c_n)p$ is linear in $p$ and symmetric in $c_1, \ldots, c_n$.*

The efficient way to construct $\Delta(c_1, \ldots, c_n)p$ is obtained as a byproduct of the efficient evaluation of the polynomial $p = W_c a$ from its **Newton coefficients** $a$ which, in turn, is based on writing the Newton form in a nested way, using the fact that each $w_j$ is a factor of each $w_k$ for $k > j$:

$$p(z) = a(1) + (z - c_1)(a(2) + (z - c_2)(\cdots + (z - c_{n-2})(a(n-1) + (z - c_{n-1})a(n))\cdots))$$

in case $p \in \Pi_{<n}$. Evaluation of this expression from the inside out results in the following algorithm.

**Nested Multiplication aka Horner's Method.** *Input: The sequence* $a \in \mathbb{F}^n$ *of (essential) coefficients in the Newton form* $W_c a$ *for some* $p \in \Pi_{<n}$*, the relevant center sequence* $c = (c_i : i = 1, \ldots, n-1)$*, and some scalar* $z$*.*

$b(n) := a(n)$
`for` $j = n-1, n-2, \ldots, 1$
  $b(j) := a(j) + (z - c_j) b(j+1)$
`endfor`

*Output: The number* $b(1) = p(z)$*.*

The algorithm provides the value of $p$ at $z$, at a cost of only $3n$ flops. More than that, the entire sequence $b$ generated by this algorithm is valuable since it provides the (essential) coefficients in the Newton form for $p$ centered at $(z, c) = (z, c_1, c_2, \ldots)$, i.e.,

“newnewtonform (74)
$$p = W_{(z,c)} b.$$

This says, in particular, that

$$p = b(1) + (\cdot - z) \sum_{j>1} w_{j-1} b(j),$$

illustrating the fact that Horner's method can be viewed as a means for dividing $p$ by the linear polynomial $(\cdot - z)$. Now, to prove (74), observe that, directly from the algorithm,

“basicfact (75)
$$a(j) = b(j) + \begin{cases} 0 & j = n; \\ (c_j - z) b(j+1) & j < n. \end{cases}$$

On substituting these expressions for the $a(j)$ into $p = W_c a$, we find

$$\begin{aligned} p &= w_n b(n) + w_{n-1} \langle b(n-1) + (c_{n-1} - z) b(n) \rangle + w_{n-2} \langle \cdots \rangle + \cdots \\ &= w_{n-1} \langle (\cdot - c_{n-1}) + (c_{n-1} - z) \rangle b(n) + w_{n-1} b(n-1) + w_{n-2} \langle \cdots \rangle + \cdots \\ &= w_{n-1} (\cdot - z) b(n) + w_{n-1} b(n-1) + w_{n-2} \langle \cdots \rangle + \cdots, \end{aligned}$$

with terms 2 and 3 in the last line looking exactly like terms 1 and 2 in the first line, except that $n$ is replaced by $n-1$, hence continuation of the process eventually leads to (74). In particular, with $c_0 := z$, we have $b(j) = \Delta(c_0 \ldots, c_{j-1}) p$, all $j$, hence (75) can be rewritten

$$\Delta(c_1, \ldots, c_j) p = \Delta(c_0 \ldots, c_{j-1}) p + (c_j - c_0) \Delta(c_0, \ldots, c_j) p.$$

For $c_j \neq c_0$, this gives the **recurrence relation**

“ddrecurrence (76)
$$\frac{\Delta(c_1, \ldots, c_j) p - \Delta(c_0, \ldots, c_{j-1}) p}{c_j - c_0} = \Delta(c_0, \ldots, c_j) p$$

which holds for arbitrary $c_0, \ldots, c_j$ (as long as $c_0 \neq c_j$) and is the reason why $\Delta(c_i, \ldots, c_j) p$ is called the **divided difference of** $p$ **at** $c_i, \ldots, c_j$.

This recurrence relation is also the simple tool for the calculation, in a **divided difference table**, of the coefficients

"newtoncoeffs (77)
$$a(j) = \Delta(c_1, \ldots, c_j)p, \quad j = 1, \ldots, n,$$

needed for the Newton form $W_c a$ for $p \in \Pi_{<n}$, starting with the numbers

(78) "allequal
$$\Delta(\underbrace{\zeta, \ldots, \zeta}_{\rho \text{ terms}})p \;=\; D^{\rho-1}p(\zeta)/(\rho-1)!, \quad 0 < \rho \le \#\{1 \le i \le n : \zeta = c_i\}.$$

The table is triangular and contains eventually all the numbers

$$t(i,j) := \Delta(c_i, \ldots, c_j)p, \quad 1 \le i \le j \le n.$$

Assuming for simplicity that

"goodfortable (79)
$$c_i = c_j \quad \Longrightarrow \quad c_i = c_{i+1} = \cdots = c_j,$$

each such entry $t(i,j) = \Delta(c_i, \ldots, c_j)p$ either has $c_i = c_j$, and in that case

$$t(i,j) = D^{j-i}p(c_i)/(j-i)!,$$

i.e., one of the numbers we started with, or else $c_i \ne c_j$, in which case

$$t(i,j) = \frac{t(i+1,j) - t(i,j-1)}{c_j - c_i},$$

hence computable from entries $t(r,s)$ with $s - r < j - i$.

Having generated this table in the manner described (and assuming still for simplicity that (79) holds), we obtain from the table the coefficients of the Newton form $p = W_c a$ of the polynomial $p$ for which the values

"interpconds (80)
$$y_j = D^\rho p(c_j)/\rho!, \quad \rho = \max\{j - i : c_i = c_j\}, \quad j = 1, \ldots, n,$$

are as entered into the table. But this means that, by entering arbitrary numbers $y_j$ at these places in the table, and then completing the table via the recurrence, we obtain the unique polynomial $p$ in $\Pi_{<n}$ that satisfies (80). In other words, we have completely solved the problem of polynomial interpolation, including osculatory (or Hermite) interpolation. If the $y_j$ are computed from some (sufficiently smooth) function $f$ (polynomial or not), then the table entries are denoted by

$$t(i,j) =: \Delta(c_i, \ldots, c_j)f$$

and called divided differences of the function $f$. Thus the "defdivdiff **(81) Definition.** *The* **divided difference of** $f$ **at** $c_i, \ldots, c_j$, *denoted by*

$$\Delta(c_i, \ldots, c_j)f,$$

*is defined for any sufficiently smooth function* $f$ *as the leading coefficient of the unique polynomial of degree* $\le j - i$ *that agrees with* $f$ *at the sequence* $(c_i, \ldots, c_j)$.

"thmpolint **(82) Theorem.** *For any sufficiently smooth function* $f$, *and any sequence* $(c_1, \ldots, c_n)$,

$$p := \sum_{j=1}^{n} \Delta(c_1, \ldots, c_j)f \prod_{i=1}^{j-1}(\cdot - c_i)$$

*is the unique polynomial of degree* $< n$ *that agrees with* $f$ *at the sequence* $(c_1, \ldots, c_n)$.

Of the many wonderful properties that divided differences possess, I mention here only one, as this one served as the portal through which we first glanced a theory of multivariate splines, and since, strangely, there is only one Numerical Analysis text (Isaacson-Keller) that actually provides it. Also, it makes many other properties of the divided differences quite evident.

"genocchi **(83) Genocchi-Hermite formula.** *For every* $c = (c_0, \ldots, c_k)$ *and every* $p \in \Pi$,

$$\Delta(c_0, \ldots, c_k)p \;=\; \int_{(c_0,\ldots,c_k)} D^k p,$$

*which uses the* **Genocchi functional**

$$f \mapsto \int_{(c_0,\ldots,c_k)} f := \int_{\Sigma_k} f(c_0 + \sum_{i=1}^{k} s_i \nabla c_i)\;\mathrm{d}s$$

*in which integration is done over the standard* $k$*-simplex,* $\Sigma_k := \{s \in \mathbb{R}^k : 1 \ge s_1 \ge \cdots \ge s_k \ge 0\}$.

**Proof:** The proof, like many that involve divided differences, is done by induction on $k$. To be sure, since $\int_{(c_0,\ldots,c_k)} 1 = \mathrm{vol}(\Sigma_k) = 1/k!$, hence

$$\int_{\Sigma_k} f(c_0 + \sum_{i=1}^{k} s_i \nabla c_i)\;\mathrm{d}s = f(\xi)/k!$$

for some $\xi \in \mathrm{conv}\{c_0, \ldots, c_k\}$, the Genocchi formula is immediate for the special case $c_0 = \cdots = c_k$, and so, in particular, for $k = 0$. Assume that $c_{k-1} \ne c_k$. Then, for any $t$,

$$\int_0^{s_{k-1}} D^k g(t + s_k(c_k - c_{k-1}))\,\mathrm{d}s_k = \frac{D^{k-1}g(t + s_{k-1}(c_k - c_{k-1})) - D^{k-1}g(t)}{c_k - c_{k-1}}.$$

Hence, with

$$t := c_0 + s_1 \nabla c_1 + \cdots + s_{k-1} \nabla c_{k-1},$$

we compute

$$\begin{aligned}\int_{\Sigma_k} D^k g(c_0 + \sum_{i=1}^k s_i \nabla c_i)\,\mathrm{d}s = &\int_0^1 \cdots \int_0^{s_{k-1}} D^k g(t + s_k \nabla c_k)\,\mathrm{d}s_k \cdots \mathrm{d}s_1 \\ = &\int_0^1 \cdots \int_0^{s_{k-2}} \frac{D^{k-1} g(t + s_{k-1}\nabla c_k) - D^{k-1} g(t)}{c_k - c_{k-1}}\,\mathrm{d}s_{k-1} \cdots \mathrm{d}s_1 \\ = &\frac{\Delta(c_0,\ldots,c_{k-2},c_k) - \Delta(c_0,\ldots,c_{k-1})}{c_k - c_{k-1}} g = \Delta(c_0,\ldots,c_k) g,\end{aligned}$$

the second last equality by induction hypothesis and since

$$t + s_{k-1}\nabla c_k = c_0 + \cdots + s_{k-2}\nabla c_{k-2} + s_{k-1}(c_k - c_{k-2}).$$

□

To be sure, any conclusion derived from the Genocchi formula (including the formula itself) can be extended to any function $g$ for which it makes sense and that can be approximated suitably by polynomials, e.g., to all $g \in C^{(k)}(T)$ for a suitable $T$.

## Univariate B-splines

Here is a very swift introduction to B-splines, and thereby to splines *aka* smooth piecewise polynomials. (A picturesque version of this material is available, by anonymous ftp, from the site `/ftp.cs.wisc.edu` in the subdirectory `Approx` as the `postscript` file `bsplbasic.ps`.)

Let $\mathbf{t} := \cdots \le t_{i-1} \le t_i \le t_{i+1} \le \cdots$ 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) $\mathbf{t} = \mathbb{B} := (\cdots, 0, 0, 0, 1, 1, 1, \ldots)$, 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

“defBj (84)
$$B_j := B_{j,k} := B_{j,k,\mathbf{t}} : x \mapsto \Delta(t_j, \ldots, t_{j+k})(\cdot - x)_+^{k-1}(t_{j+k} - t_j),$$

with $\Delta(T)$ the divided difference functional introduced in the preceding section. The normalization ensures that the $B_j$ form a (nonnegative) **partition of unity** (see (98) below). The B-spline $B_j$ was originally denoted by $N_j = N_{j,k,\mathbf{t}}$ in order to distinguish it from the differently normalized B-spline

$$M_j = M_{j,k,\mathbf{t}} := \frac{k}{t_{j+k} - t_j}\, B_{j,k,\mathbf{t}},$$

for which

$$\int M_j = 1.$$

The latter normalization arises naturally when applying the divided difference to both sides of the Taylor identity:

$$f = \sum_{r<k} D^r f(a)(\cdot - a)^r/r! + \int_a^b k(\cdot - s)_+^{k-1} D^k f(s)\,\mathrm{d}s/k!$$

to obtain (under the assumption that $t_j, \ldots, t_{j+k} \in [a\,..\,b]$)

“peano (85)
$$\Delta(t_j, \ldots, t_{j+k})f = \int_{\mathbb{R}} M_{j,k,\mathbf{t}}\, D^k f/k!,$$

showing that $M_{j,k,\mathbf{t}}$ is the Peano kernel for the divided difference $\Delta(t_j, \ldots, t_{j+k})$.

By using Leibniz' Rule

$$\Delta(\tau_0, \ldots, \tau_k)(fg) = \sum_{i=0}^{k} \Delta(\tau_0, \ldots, \tau_i)f \;\; \Delta(\tau_i, \ldots, \tau_k)g$$

for the divided difference of a product, applied to the particular product $(\cdot - x)_+^{k-1} = (\cdot - x)(\cdot - x)_+^{k-2}$, one readily obtains the **recurrence relations**

“bsplrec (86a)
$$B_{i,k} = \omega_{i,k} B_{i,k-1} \ + \ (1 - \omega_{i+1,k}) B_{i+1,k-1}$$

with

(86b)
$$B_{i,1} := \chi_{[t_i \,..\, t_{i+1})}, \qquad \omega_{i,k}(x) := \frac{x - t_i}{t_{i+k-1} - t_i}.$$

However, once one knows these recurrence relations, it is most efficient for the development of the B-spline theory to take (86) as the *starting point*, i.e., to *define* B-splines by (86). Equivalently, the development about to be given will make no use of divided differences, but will rely entirely on (86).

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

(87)
$$B_{i,k} = \sum_{j=i}^{i+k-1} b_{j,k} \chi_{[t_j \,..\, t_{j+1})},$$

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

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

while

“bsplsupport (88)
$$B_{i,k} = 0 \qquad \text{off } [t_i \,..\, t_{i+k}].$$

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

“defBjalter (89)
$$B_{i,k} =: B(\cdot | t_i, \ldots, t_{i+k})$$

is also customary. The actual smoothness of $B_{i,k}$ depends on the multiplicity with which each of the knots $t_j$, $i \le j \le i+k$, appears in its knot sequence $(t_i, \ldots, t_{i+k})$, 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 (88) 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

“recureval (90) $$\sum a_i B_{i,k} = \sum \bigl(a_i\omega_{i,k} + a_{i-1}(1-\omega_{i,k})\bigr) B_{i,k-1},$$

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 (90) 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 matters 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

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

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

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

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

“inductionstep (91) $$\sum \psi_{i,k}(\tau) B_{i,k} = (\cdot - \tau) \sum \psi_{i,k-1}(\tau) B_{i,k-1},$$

hence, by induction, that

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

This proves the following identity.

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

$$(\cdot - \tau)^{k-1} = \sum_i \psi_{i,k}(\tau) B_{i,k} \tag{94}$$

"marsden

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

Since $\tau$ in (94) 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 (94) with respect to $\tau$, we obtain the identities

$$\frac{(\cdot - \tau)^{k-\nu}}{(k-\nu)!} = \sum_i \frac{(-D)^{\nu-1}\psi_{i,k}(\tau)}{(k-1)!} B_{i,k}, \quad \nu > 0. \tag{95}$$

"difmarsden

On using these identities 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

$$p = \sum_i \lambda_{i,k} p \; B_{i,k}, \tag{96}$$

"polrepro

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

$$\lambda_{i,k} f := \lambda_{i,k,\mathbf{t}} f := \sum_{\nu=1}^{k} \frac{(-D)^{\nu-1}\psi_{i,k}(\tau)}{(k-1)!} D^{k-\nu} f(\tau). \tag{97}$$

"dualfunct

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

$$1 = \sum_i B_{i,k} \tag{98}$$

"eqpartition

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, anticipating that $\lambda_{jk} p$ is independent of $\tau$ in case $p \in \Pi_{<k}$, since $D^{k-2}\psi_{i,k}$ is a linear polynomial that vanishes at

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

$$\ell = \sum_i \ell(t^*_{i,k}) B_{i,k}, \qquad \ell \in \Pi_1. \tag{99}$$

"exactforlinear

The identity (95) 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 $\nu \le \#t_j := \#\{s : t_s = t_j\}$, the choice $\tau = t_j$ in (95) leaves only terms with support either entirely to the left or else entirely to the right of $t_j$. This implies that

“difmarsdenplus (100)
$$\frac{(\cdot - t_j)_+^{k-\nu}}{(k-\nu)!} = \sum_{i \ge j} \frac{(-D)^{\nu-1}\psi_{i,k}(t_j)}{(k-1)!}\, B_{i,k}, \quad 0 < \nu \le \#t_j.$$

Consequently,

“eqtrunc (101)
$$(\cdot - t_j)_+^{k-\nu} \;\in\; S_{k,\mathbf{t}} \;\text{ for }\; 1 \le \nu \le \#t_j,$$

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

**(102) Theorem.** *If $t_i < t_{i+k}$ for all $i$, then the B-spline sequence $(B_{i,k} : i)$ is 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 $t_i$ that are $\rho(i) := k - 1 - \#t_i$ times continuously differentiable at $t_i$, all $i$. In particular, each $f \in S_{k,\mathbf{t}}$ is in $C^{(\rho(i))}$ near $t_i$, 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

“eqjunk (103)
$$(\cdot - a)^{k-\nu},\ \nu = 1, \cdots, k;\ (\cdot - t_i)_+^{k-\nu},\ \nu = 1, \cdots, \#t_i,\ \text{for}\ \ a < t_i < b.$$

This follows from the observations that (i) the sequence of functions in (103) is linearly independent; and (ii) a piecewise polynomial function $f$ with a breakpoint at $t_i$ that is $k - 1 - \#t_i$ times continuously differentiable there can be written uniquely as

$$f = p + \sum_{\nu=1}^{\#t_i} a_\nu (\cdot - t_i)_+^{k-\nu},$$

with $p$ a suitable polynomial of degree $< k$ and suitable coefficients $a_\nu$. Since each of the functions in (103) lies in $S_{k,\mathbf{t}}$, by (95) and (101), we conclude that

(104)
$$S_{|I} \subset (S_{k,\mathbf{t}})_{|I}.$$

On the other hand, the dimension of $S_{|I}$, i.e., the number of functions in (103), equals the number of B-splines with some support in $I$ (since it equals $k + \#\{i : a < t_i < b\}$), hence is an upper bound on the dimension of $(S_{k,\mathbf{t}})_{|I}$. This implies that equality must hold in (104), and that the set of B-splines having some support in $I$ must be linearly independent over $I$. □

“corone **(105) Corollary.** *The sequence* $(B_{j,k} : B_{j,k}|_I \neq 0)$ *of B-splines having some support in a given (proper) interval* $I$ *is linearly independent over that interval.*

“corconst **(106) Corollary.** *For all* $p \in \Pi_{<k}$, $D_\tau \lambda_{j,k} p = 0$.

Indeed, since the B-spline sequence is linearly independent, the coefficients $\lambda_{i,k}p$ appearing in (96) are uniquely determined, hence cannot change with $\tau$. When convenient later on, we will choose $\tau$ in (97) in dependence on $i$, i.e., as $\tau_i$.

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

“corderiv **(108) 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.,* $DS_{k,\mathbf{t}} \subseteq S_{k-1,\mathbf{t}}$.

**(109) 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 (108) 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. For, by (85) and (73),

“pointmass (110)
$$\lim_{t_i,\ldots,t_{i+k}\to\tau} M(\cdot|t_i,\ldots,t_{i+k}) \;=\; \delta_\tau = \lfloor\!\Delta(\tau)$$

as a distribution, and that limit is quite different from 0. □

The identity (96) can be extended to all spline functions. For this, we agree, consistent with (86b), that all derivatives in (97) are to be taken as limits from the right in case $\tau$ coincides with a knot.

“thmdualfunct **(111) Theorem.** *If* $\tau = \tau_i$ *in definition (97) 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* $\nu$ *and any pp* $f$,

“duality (112)
$$\lambda_{i,k}\Big(\sum_j a_j B_{j,k}\Big) = a_i.$$

It is remarkable that $\tau$ can be chosen arbitrarily in the interval $[t_i \,.\,.\, t_{i+k})$. The reason behind this is Corollary (106).

**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 that 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 (96) on $[t_l \,.\,.\, t_{l+1})$, while, by Corollary (105) or directly from (96), $(p_{l+1-k}, \cdots, p_l)$ is 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 (112), 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

$$\lambda_{i,k} f = \sum_{\nu=1}^{k} \frac{(-D)^{\nu-1}\psi_{i,k}(\tau)}{(k-1)!} D^{k-\nu} f(\tau) \tag{113}$$

with

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

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

“dualdiff
$$\lambda_{i,k-1} D = \frac{\lambda_{i,k} - \lambda_{i-1,k}}{(t_{i+k-1} - t_i)/(k-1)}. \tag{114}$$

Consequently, we get the **differentiation formula**

“splinediff
$$D \sum_i a_i B_{i,k} = \sum_i \frac{a_i - a_{i-1}}{(t_{i+k-1} - t_i)/(k-1)} B_{i,k-1}. \tag{115}$$

We now consider how $\lambda_{i,k}$ depends on the knot sequence $\mathbf{t}$. Perhaps surprisingly, although $B_{i,k}$ involves the knots $t_i, \cdots, t_{i+k}$, $\lambda_{i,k}$ 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 $t_{i+1}, \cdots, t_{i+k-1}$. Indeed, for any $\alpha, x, y$,

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

Hence, with

$$\lambda_k : \mathbb{R}^{k-1} \to (C^{(k-1)})' : t \mapsto \lambda_{0,k},$$

we have

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

as well as

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

for any permutation $\sigma$ of order $k-1$.

Consider, in particular, $t_1 = \cdots = t_{k-1} = x$. 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

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

In particular

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

This meshes entirely with the fact that, if $x = t_{i+1} = \cdots = t_{i+k-1}$, then $B_{i,k}$ is the only $B_{j,k}$ that 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.

**side issue: 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

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

E.g., the blossom of $(\cdot - \tau)^r \in \Pi_r$ is $s \mapsto (s_1 - \tau)\cdots(s_r - \tau)$. If $p = \sum_j ()^j c_j \in \Pi_r$, then

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

with $\binom{M}{j} := \{K \subset M : \#K = j\}$. We deduce from the above that

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

In particular, the $i$th B-spline coefficient of a $k$th order spline with knot sequence $\mathbf{t}$ is the value at $(t_{i+1},\ldots,t_{i+k-1})$ of the blossom of *any* of the $k$ polynomial pieces associated with the intervals $[t_j \,..\, t_{j+1})$, $j = i,\ldots,i+k-1$. 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.

## 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

$$\hat{\lambda}_{i,k} = (1 - \widetilde{\omega}_{i,k}(x))\lambda_{i-1,k} + \widetilde{\omega}_{i,k}(x)\lambda_{i,k}, \quad \text{all } i,$$

with

“defgot (116)
$$\widetilde{\omega}_{i,k}(x) := \max\{0, \min\{1, \omega_{i,k}(x)\}\}.$$

Correspondingly,

(117) “recurinsert
$$\sum_i a_i B_{i,k} = \sum_i \hat{a}_i \hat{B}_{i,k},$$

with

$$\hat{a}_i = (1 - \widetilde{\omega}_{i,k}(x)) a_{i-1} + \widetilde{\omega}_{i,k}(x) a_i, \quad \text{all } i.$$

Note that this is exactly the way we computed coefficients by recurrence in (90). 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,

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

In CAGD, one thinks of a spline $f = \sum_j a_j B_{j,k}$ in terms of the curve $x \mapsto (x, f(x))$ that is its graph. Since $x = \sum_j t^*_{j,k} B_{j,k}(x)$ by (99), we can represent this spline *curve* as the **vector-valued** spline $\sum_j P_j B_{j,k}$ with coefficients

$$P_j := (t^*_{j,k}, a_j),$$

called its **control points**. The broken line with vertex sequence $(P_j)$ is called its **control polygon**; I'll denote it by

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

to stress its dependence on $\mathbf{t}$.

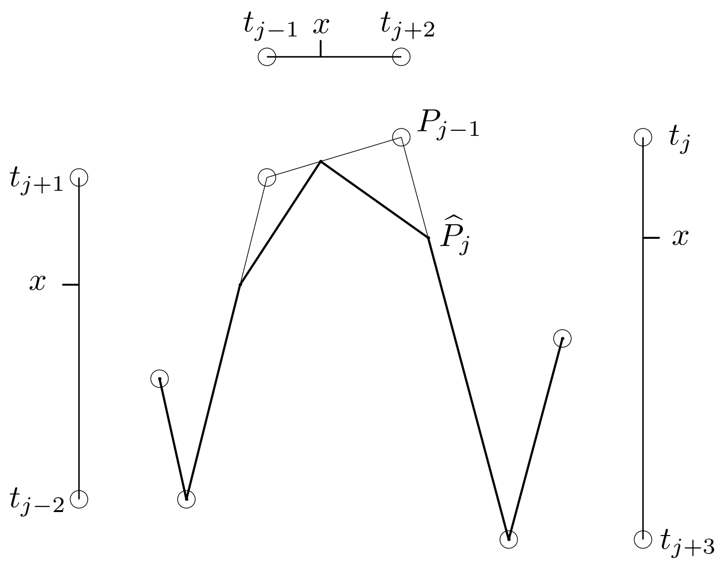
 

**(118) Proposition.** *If $\hat{\mathbf{t}}$ is obtained from $\mathbf{t}$ by the insertion of one additional knot, then, for any $f \in S_{k,\mathbf{t}}$, $C_{k,\hat{\mathbf{t}}}f$ interpolates, at its breakpoints, to $C_{k,\mathbf{t}}f$, specifically,*

$$\hat{P}_j = (1 - \widetilde{\omega}_{j,k}(x))P_{j-1} + \widetilde{\omega}_{j,k}(x)P_j,$$

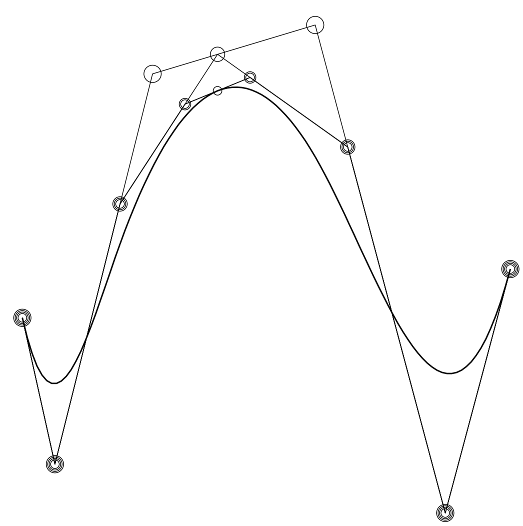
*(and is thereby uniquely determined).*

Here are two figures, to illustrate this geometric interpretation of knot insertion which, ultimately led to an entirely new and quite different way to generate curves and surfaces, namely **subdivision**.



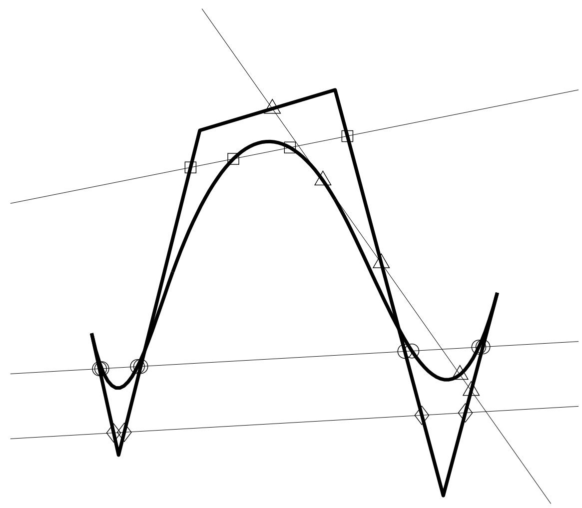

(119) Figure. Insertion of $x = 2$ into the knot sequence $\mathbf{t} = (0,0,0,0,\mathbf{1},\mathbf{3},5,5,5,5)$, with $k = 4$.

"figknotinsert



(120) Figure. Three-fold insertion of the same point.



(121) 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.

It follows 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 (an example of the Bauhaus maxim "less is more"?). It equals the number of sign changes when we ignore the zeros. The number of strong (weak) sign changes can only increase (decrease) under small perturbations. Precisely,

$$S^-(a) \le \liminf_{b \to a} S^-(b) \le \limsup_{b \to a} S^+(b) \le S^+(a).$$

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

“lemlaneriesenfeld **(122) 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

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

“propsignreduce **(123) 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

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

□

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

$$Vg := \sum_j g(t^*_{j,k}) B_{j,k}.$$

*It is usually defined only when $\#t_i < k$, all $i$.*

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

“nomorecuts (126) $$S^-(Vg-\ell) \le S^-(g-\ell).$$

*Even more precisely,*

“preservesgn (127) $$D^r g \ge 0 \quad \Longrightarrow \quad D^r Vg \ge 0, \quad r = 0,1,2,$$

*and this holds even 'locally'.*

**Proof:** By (99), $V\ell = \ell$ (on $I_{k,\mathbf{t}}$) for all $\ell \in \Pi_1$, hence $V(f-\ell) = Vf-\ell$, therefore, by Proposition (123), and by the strict increase in the sequence $t^*_i : i = 1,\ldots,n$ $S^-(Vf-\ell) \le S^-((f-\ell)(t^*_i) : i) \le S^-(f-\ell)$.

However, (126) by itself fails to imply the more precise statement (127), which follows from the nonnegativity of the B-splines along with the observation that

$$DVg = \sum_j \Delta(t^*_{j-1}, t^*_j) g\, B_{j,k-1},$$

hence

$$D^2Vg = \sum_j \frac{\Delta(t^*_{j-1},t^*_j)g - \Delta(t^*_{j-2},t^*_{j-1})g}{(t_{j+k-2}-t_j)/(k-2)}\, B_{j,k-2}.$$

□

## zeros of a spline, counting multiplicities

There is a complete theory that provides an upper bound on the number of zeros of a spline, even counting multiplicity, in terms of the number of sign changes (strong and/or weak) in its B-spline coefficients, with **multiplicity** of a zero defined as the maximal number of distinct nearby zeros in a nearby spline (from the same spline space).

For the theory to give useful information, one has to assign a finite multiplicity to zero *intervals*, and this adds a further complication.

Multiplicity considerations are important when one wishes to consider osculatory spline interpolation, i.e., interpolation at possibly repeated points. Since I will not get to that topic in this course, I am content to state and prove only the following very useful proposition.

"propzerocount **(128) Proposition.** *If* $f = \sum_j a_j B_{j,k,\mathbf{t}}$ *vanishes at* $x_1 < \cdots < x_r$, *while* $f^{\mathbf{t}} := \sum_j |a_j| B_{j,k,\mathbf{t}}$ *does not, then* $S^-(a) \ge r$.

**Proof:** Since $f^{\mathbf{t}}(x_i) > 0$, while $f(x_i) = 0$, the sequence $(a_j B_j(x_i) : B_j(x_i) \ne 0)$ must have at least one strong sign change, hence, so must the sequence $(a_j : B_j(x_i) \ne 0)$, by the nonnegativity of the B-splines. This gives altogether $r$ strong sign changes in $a$, provided we can be sure that different $x_i$ generate different sign changes. Off-hand, this may not be so, but can be guaranteed by inserting each of the points $(x_i + x_{i+1})/2$, $i = 1, \ldots, r-1$, into the knot sequence $k$ times. If $\tilde{\mathbf{t}}$ and $\tilde{a}$ are the resulting knot and coefficient sequences, respectively, then still $f^{\tilde{\mathbf{t}}} > 0 = f$ on the $x_i$ (since, from $f^{\mathbf{t}}(x_i) > 0$, we know that $x_i$ is an isolated zero of $f$), while now $\{j : \tilde{B}_j(x_i) \ne 0\} \cap \{j : \tilde{B}_j(x_h) \ne 0\} = \emptyset$ for $i \ne h$, hence

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

□

## spline interpolation

We consider **spline collocation**, i.e., interpolation from $S_{k,\mathbf{t}}$ at the increasing sequence $\mathbf{x} = (x_i)$ 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 $\ge k$, 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

$$\mathbf{x} = (x_1 < \cdots < x_n).$$

Hence, for given $g$, we are seeking $f \in S_{k,\mathbf{t}}$ with $f = g$ on $\mathbf{x}$. Equivalently, we are seeking a solution to the linear system $A? = g|_{\mathbf{x}}$, 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.

**(129) 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 $i$, $t_{i+k} \le x_i$, then the first $i$ columns of $A$ have nonzero entries only in the first $i-1$ rows, hence $A$ cannot be invertible. Again, if $x_i \le t_i$, 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 $\mathbf{x}$ and the sequence of the supports of the B-splines are increasing. In particular, we have proved

**(130) Corollary.** *If $(B_{m_j,k}(s_i) : i,j = 1, \ldots, r)$ is invertible, with both $(m_j)$ and $(s_i)$ increasing, then $B_{m_i}(s_i) \ne 0$, 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

"schoenbergwhitney **(131) Schoenberg-Whitney Theorem.** *Let $\mathbf{t} = (t_1, \ldots, t_{n+k})$ with $B_{j,k} \ne 0$, all $j$, and let $\mathbf{x} := (x_1 < \cdots < x_n)$. Then, $A_{\mathbf{x}} := (B_{j,k}(x_i))$ is invertible iff $t_i < x_i < t_{i+k}$, all $i$.*

**Proof:** We only need to prove the 'if'. Since $A_{\mathbf{x}}$ is square, it is sufficient to prove that $A_{\mathbf{x}} a = 0$ implies $a = 0$. Consider $f = \sum_j a_j B_{j,k}$ with $A_{\mathbf{x}} a = 0$. If $a \ne 0$, then, by Corollary (105), $f \ne 0$. Let $I = (\alpha \,..\, \beta)$ be a maximal open interval in

$$\operatorname{supp} \sum_j |a_j| B_{j,k} = \bigcup \{(t_j \,..\, t_{j+k}) : a_j \ne 0\}.$$

It follows that $I = (t_\nu \,..\, t_{\mu+k})$ 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, by Proposition (128),

$$S^-(a_\nu, \ldots, a_\mu) \ge \mu + 1 - \nu,$$

which is nonsense. □

The Schoenberg-Whitney Theorem has been generalized in at least two directions: (i) permission of coincidences in the sequence $(x_i)$ correspondingly to osculatory interpolation; and (ii) consideration of a subsequence $(B_{m_j} : j = 1, \ldots, n)$ instead of the whole sequence (for a suitably longer knot sequence).

Any $\mathbf{x} = (x_1 < \cdots < x_n)$ satisfying the Schoenberg-Whitney conditions for $S_{k,\mathbf{t}}$ gives rise to the corresponding projector $P_{\mathbf{x}}$ that associates $g \in C$ with the unique $f = P_{\mathbf{x}} g \in S_{k,\mathbf{t}}$ that agrees with $g$ at $\mathbf{x}$. Since $P_{\mathbf{x}}$ is a linear projector, we have

$$\operatorname{dist}(g, S_{k,\mathbf{t}}) \le \|g - P_{\mathbf{x}} g\| \le (1 + \|P_{\mathbf{x}}\|) \operatorname{dist}(g, S_{k,\mathbf{t}}),$$

hence $P_{\mathbf{x}}$ s a candidate for a "good" approximation scheme from $S_{k,\mathbf{t}}$ to the extent that $\|P_{\mathbf{x}}\|$ is "small", i.e., not much larger than 1. We pursue this question in the context of the uniform norm, i.e., in the space

$$X = C[t_1 \mathinner{..} t_{n+k}].$$

Since $P_{\mathbf{x}}g = \sum_j a_j B_j$ with $a = A_{\mathbf{x}}^{-1}(g|_{\mathbf{x}})$, while $(B_j)$ is a positive partition of unity,

$$\|P_{\mathbf{x}}g\|_\infty \le \|A_{\mathbf{x}}^{-1}(g|_{\mathbf{x}})\|_\infty \le \|A_{\mathbf{x}}^{-1}\|_\infty \|(g|_{\mathbf{x}})\|_\infty \le \|A_{\mathbf{x}}^{-1}\|_\infty \|g\|_\infty,$$

with $\|A_{\mathbf{x}}^{-1}\|_\infty = \max_i \sum |A_{\mathbf{x}}^{-1}(i,j)|$. Hence

$$\|P_{\mathbf{x}}\| \le \|A_{\mathbf{x}}^{-1}\|_\infty.$$

While this bound is not sharp (in general), it is the only bound readily available. Hence, in search for a good approximation scheme from $S_{k,\mathbf{t}}$, we now look for $\mathbf{x}$ so that $\|A_{\mathbf{x}}^{-1}\|_\infty$ is as small as possible, and this will lead us to a particularly good choice for $\mathbf{x}$, namely the Chebyshev-Demko sites $\mathbf{x}^*$, easily computable for given $\mathbf{t}$, and, for these, $\|P_{\mathbf{x}^*}\| \le \|A_{\mathbf{x}^*}^{-1}\|_\infty \le k2^k$. This bound is quite small for modest $k$ and, surprisingly, is independent of the knot sequence $\mathbf{t}$. In other words, interpolation at the Chebyshev-Demko sites is near-best independent of the knot sequence. This makes it possible to use such interpolation profitably even when one has chosen the knot sequence quite non-uniform in order to adjust to the varied behavior of the function being approximated.

The search for $\mathbf{x}$ that minimizes $\|A_{\mathbf{x}}^{-1}\|_\infty$ is aided by the knowledge that our collocation matrix $A_{\mathbf{x}}$ is totally positive, to be established next.

## total positivity

We recall that an $(m,n)$-matrix $C$ is **totally positive** if, for any strictly increasing (index) sequences $\mathbf{i} = (i_1 < \cdots < i_r)$ in $\{1,\ldots,m\}$ and $\mathbf{j} = (j_1 < \cdots < j_r)$ in $\{1,\ldots,n\}$, the determinant $\det C(\mathbf{i},\mathbf{j})$ of the $(r,r)$-submatrix

$$C(\mathbf{i},\mathbf{j}) := \big(C(i_p, j_q) : p,q = 1,\ldots,r\big)$$

is nonnegative. The most immediately important fact concerning total positivity is the following.

**(132) Fact.** *If $C$ is invertible and totally positive, then its inverse is* **checkerboard**, *meaning that $C^{-1}(i,j)(-1)^{i-j} \ge 0$, all $i,j$.*

**Proof:** By Cramer's rule,

$$C^{-1}(i,j) = (-1)^{i-j} \det C(\backslash j, \backslash i)/\det C.$$

□

“thmtp **(133) Theorem (Karlin).** *For any* $\mathbf{x} := (x_1 < \cdots < x_n)$*, any* $k \in \mathbb{N}$*, and any knot sequence* $\mathbf{t} = (t_i : i = 1, \ldots, n+k)$*, the collocation matrix*

$$A := \big(B_j(x_i) : i, j = 1, \ldots, n\big)$$

*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, by (117),

$$B_j = (1 - \alpha_{j+1})\hat{B}_{j+1} + \alpha_j \hat{B}_j,$$

with all $\alpha_j \in [0\,..\,1]$. Since the determinant of a matrix is a linear function of the columns of that matrix, we have, e.g.,

$$\det[\cdots, B_j(\mathbf{x}), \ldots] = (1 - \alpha_{j+1}) \det[\cdots, \hat{B}_{j+1}(\mathbf{x}), \ldots] + \alpha_j \det[\cdots, \hat{B}_j(\mathbf{x}), \ldots],$$

with $\cdots$ unchanged in their respective places. It follows that, for any $\mathbf{i}, \mathbf{j}$,

$$\det A(\mathbf{i}, \mathbf{j}) = \sum_{\mathbf{h}} \gamma_{\mathbf{h}} \det \hat{A}(\mathbf{i}, \mathbf{h}),$$

with all the $\gamma_{\mathbf{h}} \ge 0$, and the sum, offhand, over certain nondecreasing sequences, since only neighboring columns of $\hat{A}$ participate in a column of $A$. However, we may omit all $\mathbf{h}$ that are not strictly increasing, since the corresponding determinant is trivially zero. Therefore,

$$\det A(\mathbf{i}, \mathbf{j}) = \sum_{\mathbf{h}} \gamma_{\mathbf{h}} \det \hat{A}(\mathbf{i}, \mathbf{h}),$$

with the $\gamma_{\mathbf{h}} \ge 0$ and all $\mathbf{h}$ 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

$$\det A(\mathbf{i}, \mathbf{j}) = \sum_{\mathbf{h}} \gamma_{\mathbf{h}} \det \tilde{A}(\mathbf{i}, \mathbf{h}),$$

with the $\gamma_{\mathbf{h}} \ge 0$ and all $\mathbf{h}$ strictly increasing. However, in each row of $\tilde{A}$, there is exactly one nonzero entry, namely the entry belonging to $\tilde{B}_{m_i}$ with $\tilde{t}_{m_i} < x_i = \tilde{t}_{m_i+1}$, and that entry equals 1. In other words, $\tilde{A}$ has all its entries zero except that the submatrix $\tilde{A}(:, \mathbf{m})$ is the identity matrix. Thus $\det \tilde{A}(\mathbf{i}, \mathbf{h}) = \delta_{\mathbf{h}, \mathbf{m} \circ \mathbf{i}}$, hence $\det A(\mathbf{i}, \mathbf{j}) = \gamma_{\mathbf{m} \circ \mathbf{i}} \ge 0$. □

## spline interpolation (cont.)

We continue with the spline interpolation setup introduced earlier, considering interpolation at $\mathbf{x} = (x_1 < \cdots < x_n)$ from $S_{k,\mathbf{t}}$ with $\mathbf{t} = (t_1 \le \cdots \le t_{n+k})$ and $\#t_i < k$, all $i$.

The fact that the collocation matrix

$$A_{\mathbf{x}} := (B_j(x_i))$$

is totally positive implies that

$$\|A_{\mathbf{x}}^{-1}\|_\infty = \max_i \sum_j |A_{\mathbf{x}}^{-1}(i,j)| = \max_i \sum_j A_{\mathbf{x}}^{-1}(i,j)(-1)^{i-j} = \max_i (-1)^{i-n} a_i^{\mathbf{x}},$$

with $a^{\mathbf{x}}$ the unique solution to the equation $A_{\mathbf{x}}? = ((-1)^{n-j} : j = 1, \ldots, n)$, hence

$$\|P_{\mathbf{x}}\|_\infty \le \|A_{\mathbf{x}}^{-1}\|_\infty = \|a^{\mathbf{x}}\|_\infty,$$

with

$$f_{\mathbf{x}} := \sum_j a_j^{\mathbf{x}} B_j$$

the unique element of $S_{k,\mathbf{t}}$ satisfying

$$f_{\mathbf{x}}(x_i) = (-1)^{n-i}, \quad i = 1, \ldots, n.$$

Our search for argmin $\|A_{\mathbf{x}}^{-1}\|_\infty$ therefore is the search for the $\mathbf{x}$ that minimizes $\|a^{\mathbf{x}}\|_\infty$. For this, we use, in effect, Remez' (second) algorithm for the construction of a ba from $\Pi_k$.

First we note that

$$(-1)^{n-i} a_i^{\mathbf{x}} > 0, \text{all } i,$$

since $a_i^{\mathbf{x}} = \sum_j A_{\mathbf{x}}^{-1}(i,j)(-1)^{n-j} = (-1)^{n-i} \sum_j |A_{\mathbf{x}}^{-1}(i,j)|$ with the sum necessarily positive. Further, since $f_{\mathbf{x}}$ strictly alternates in sign at the $n$ points $x_1, \ldots, x_n$ and vanishes outside $(t_1 \,..\, t_{n+k})$, it follows that $f_{\mathbf{x}}$ has $n$ distinct local extrema $y_1 < \cdots < y_n$ with

$$(-1)^{n-i} f_{\mathbf{x}}(y_i) \ge 1, \quad \forall i.$$

This implies that $t_i < y_i < t_{i+k}$. all $i$. Indeed, if, e.g., $y_i \le t_i$ for some $i$, then $B_\mu(y_\nu) = 0$ for all $\nu \le i \le \mu$, showing that, on $y_1, \ldots, y_i$, $f_{\mathbf{x}}$ agrees with $\sum_{j<i} a_j^{\mathbf{x}} B_j$. In particular, $i - 2 \ge S^-(a_1^{\mathbf{x}}, \ldots, a_{i-1}^{\mathbf{x}}) \ge S^-(\sum_{j<i} a_j^{\mathbf{x}} B_j) = i - 1$, 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 $(-1)^{n-i} a_i^{\mathbf{x}} \ge (-1)^{n-i} a_i^{\mathbf{y}} \ge 0$, 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 $x_1^* < \cdots < x_n^*$ since $f_m$ strictly alternates in sign on $y_1^m < \cdots < y_n^m$. Let $f_* := \sum_j a_j^* B_j$. Then,

“maxoscil (134)
$$1 = (-1)^{n-i} f_*(x_i^*) = \|f_*\|_\infty, \quad i = 1, \ldots, n,$$

since $1 = (-1)^{n-i} f_m(y_i^{m-1})$, while $\|f_m\| = \max_i |f_m(y_i^m)|$ and $f_*$ is the uniform limit of $(f_m : m)$.

“chebspl **(135) Lemma.** *Let $f_* = \sum_j a_j^* B_{j,k,\mathbf{t}}$ satisfy (134) for some (strictly) increasing $\mathbf{x}^*$. Then*

$$(-1)^{n-i} a_i^* = 1/\operatorname{dist}(B_i, \operatorname{span}(B_j : j \neq i)), \quad i = 1, \ldots, n.$$

*In particular, $f_*$ is independent of $\mathbf{x}^*$ (hence of the initial $\mathbf{x}$ in the above iteration).*

**Proof:** Let $A_* := (B_j(x_i^*))$ and set

$$\lambda_i : S_{k,\mathbf{t}} \to \mathbb{R} : f \mapsto \sum_j A_*^{-1}(i,j) f(x_j^*).$$

Then $\lambda_i B_j = \delta_{ij}$, hence

$$(-1)^{n-i} a_i^* = (-1)^{n-i} \lambda_i f_* = \sum_j (-1)^{n-i} A_*^{-1}(i,j)(-1)^{n-j} = \sum_j |A_*^{-1}(i,j)| = \|\lambda_i\|,$$

the last equality by the fact that $\sum_j |A_*(i,j)|$ is obviously an upper bound for $\|\lambda_i\|$, yet it equals $|\lambda_i f_*|$ with $\|f_*\|_\infty = 1$, hence is also a lower bound for $\|\lambda_i\|$. Now, for any linear functional $\lambda$ on any nls $X$ and for any $f \in X$,

$$|\lambda f| = \|\lambda\| \operatorname{dist}(f, \ker \lambda).$$

Hence, $1 = \operatorname{dist}(f_*, \ker \lambda_i)$, while $\ker \lambda_i = \operatorname{span}(B_j : j \neq i)$ and so $|a_i^*| \operatorname{dist}(B_i, \ker \lambda_i) = \operatorname{dist}(f_*, \ker \lambda_i) = 1$. □

The function $f_*$ is, by definition, the **Chebyshev spline** for $S_{k,\mathbf{t}}$, i.e., the unique (up to scalar multiples) element that maximally equioscillates, i.e., satisfies (134). From Lemma (135), we can write it as

$$T_{k-1,\mathbf{t}} := \sum_j (-1)^{n-j} B_j / \operatorname{dist}(B_j, \operatorname{span}(B_i : i \neq j)).$$

To be sure, any $f \in S_{k,\mathbf{t}}$ can have at most $n-1$ sign changes, hence $n$ is indeed the maximal number of equioscillations possible for $f \in S_{k,\mathbf{t}}$.

**(136) Proposition (S. Demko).** $\mathbf{x} \mapsto \|a^{\mathbf{x}}\|_\infty$ *is uniquely minimized when* $\mathbf{x}$ *is the extreme-point sequence for the Chebyshev spline.*

**Proof:** We just saw that, starting with any $\mathbf{x}$, we reach the Chebyshev spline in the limit in a process during which the B-spline coefficients decrease in absolute value. Hence $\|a^*\|_\infty \leq \|a^{\mathbf{x}}\|_\infty$ for any $\mathbf{x}$. □

The terminology *Chebyshev spline*, though apt in view of the fact that, for $t_1 = \cdots = t_k = -1, t_{k+1} = \cdots = t_{2k} = 1$ it is the Chebyshev polynomial of degree $k-1$, unfortunately clashes with the standard term 'Tschebyscheffian' spline, meaning a piecewise function whose pieces all come from the same Haar space. The notation $T_{k-1,\mathbf{t}}$ is meant as a legitimate extension of the notation $T_r$ for the Chebyshev polynomial of degree $r$. Although whole books have been written on the special case $k = n$, in which case $T_{k-1,\mathbf{t}}$ on $I_{k,\mathbf{t}}$ agrees with the (suitably scaled and translated) Chebyshev polynomial $T_{k-1}$ of degree $k-1$, the Chebyshev spline is largely unexplored territory (except for K. Mørken's Ph.D. Thesis).

"corchebspl **(137) Corollary.** *For* $\mathbf{x} = (x_1 < \cdots < x_n)$ *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* $T_{k-1,\mathbf{t}}$.

Offhand, some of the extreme points of $T_{k-1,\mathbf{t}}$ may lie outside the basic interval $I_{k,\mathbf{t}} = [t_k \,..\, t_{n+1}]$. 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 $f_{\mathbf{x}}$ is strictly monotone on $[t_1 \,..\, t_2]$, hence necessarily $t_2 \le y_1$. By the same reasoning, $y_n \le t_{n+k-1}$ in case $t_{n+k-1} < t_{n+k}$. This implies that nothing in the above arguments changes if we use the interval $[t_2 \,..\, t_{n+k-1}]$ instead. That choice made, the location of $t_1$ and $t_{n+k}$ becomes irrelevant to $S_{k,\mathbf{t}}$ as restricted to $[t_2 \,..\, t_{n+k-1}]$ (and not even 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 $t_{n+k} = t_{n+1}$, hence have $[t_1 \,..\, t_{n+k}] = I_{k,\mathbf{t}}$.

Since, for all $i$, $\lambda_{i,k} = \sum_j A_*^{-1}(i,j)\Delta(x_j^*)$ on $S_{k,\mathbf{t}}$, and the latter linear functional, as we have just seen, takes on its norm on $T_{k-1,\mathbf{t}}$, it follows that *the map*

$$S_{k,\mathbf{t}} \to \mathbb{R} : \sum_j a_j B_j \mapsto \|a\|_\infty / \|\sum_j a_j B_j\|$$

*takes on its maximum at* $T_{k-1,\mathbf{t}}$, and that maximum is $\|a^*\|_\infty$. This maximum determines the condition of the B-spline basis, to be discussed next.

## 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 $\sup_a \frac{\|\sum_j a_j B_j\|_\infty}{\|a\|_\infty} = 1$.

Further,

$$\kappa_\infty([B_j : j]) = \max_i \|\lambda_{i,k}\|.$$

Hence the following lemma is of interest.

"lemdk **(138) 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 $I := [t_\ell \,..\, t_{\ell+1}]$ be any knot interval in $[t_i \,..\, t_{i+k}]$, and choose, as we may, $\tau_i \in I$. Then, as both $\psi_i$ and $f$ are polynomials of degree $< k$ on $I$, Markov's inequality (65) implies that

$$|(-D)^{\nu-1}\psi_i(\tau_i)||D^{k-\nu} f(\tau_i)| \le \mathrm{const}_{k,\nu} \|\psi_i\|_\infty(I)/|I|^{\nu-1}\; \|f\|_\infty(I)/|I|^{k-\nu}.$$

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

**(139) Claim.** $d_k = \sup_{\mathbf{s}} \|\lambda_{k-1,k}|_{S_{k,\mathbf{s}}}\|([-1\,..\,1])$, *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 $\tilde{\mathbf{t}}$ 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 $\tilde{i}$ be such that $\tilde{t}_{\tilde{i}+j} = t_{i+j}$ for $j = 1, \ldots, k-1$. The corresponding spline space $S_{k,\tilde{\mathbf{t}}}$ may well be larger than the space $S_{k,\mathbf{t}}$ we started with, but $\lambda_i = \tilde{\lambda}_{\tilde{i}}$ since, by (97), $\lambda_i$ only depends on the knots $t_{i+1}, \ldots, t_{i+k-1}$. This shows that

$$\|\lambda_i\| := \sup_{f\in S_{k,\mathbf{t}}} |\lambda_i f|/\|f\|_\infty \le \sup_{f\in S_{k,\tilde{\mathbf{t}}}} |\tilde{\lambda}_{\tilde{i}} f|/\|f\|_\infty = \|\tilde{\lambda}_{\tilde{i}}\| = \|\tilde{\lambda}_{\tilde{i}}\|([-1\,..\,1]),$$

the last equality since $\tilde{\lambda}_{\tilde{i}} f$ only depends on $f|_{[-1..1]}$.

Consequently, $d_k \le \sup_{\mathbf{s}} \|\lambda_{k-1,k}\|([-1\,..\,1])$. 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 considering the related extremum problem

“defdkone (140) $$d_{k,1} := \sup_{\mathbf{t}} \sup_i \sup_{f \in S_{k,\mathbf{t}}} |\lambda_{i,k,\mathbf{t}} f| |t_i - t_{i+k}| / \|f\|_1([t_i \,.\,.\, t_{i+k}]),$$

Karl Scherer and Aleksei Shadrin were recently able to show that $d_k \le k2^k$.

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 $\mathbf{L}_p$, 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 (140) and Hahn-Banach, we deduce the following

“propreprdualfunct **(141) Proposition.** *There exists $h_j$ 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 $[t_j \,..\, t_{j+k}]$ have some given knot interval $[t_\ell \,..\, t_{\ell+1}]$ in common. It follows that, for any $1 \le p \le \infty$,

“bsplinecondp (142)
$$\|c\|_p/d_{k,1} \le \|\sum_i B_{i,k,p}c_i\|_p \le \|c\|_p,$$

for any knot sequence $\mathbf{t}$ and any coefficient sequence $c$, with

“defbsplp (143)
$$B_{i,k,p} := (k/(t_{i+k} - t_i))^{1/p} B_{i,k}.$$

## Degree of approximation by splines

### quasiinterpolants

It follows from Theorem (111) that the linear map

$$P : g \mapsto \sum_j B_{j,k}\lambda_{j,k}g$$

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 $\tau$ in (97) equal to some $\tau_i \in [t_i \,..\, t_{i+k})$. It is also local, since $\lambda_{j,k}g$ depends only on the behavior of $g$ near $\tau_j$, hence, by our choice of $\tau_j$, only on the behavior of $g$ 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 $\mathbf{L}_1(I_{k,\mathbf{t}})$, we make use of (141) in order to obtain the linear functional

$$\mu_j g := \int h_j g$$

with $\operatorname{supp} h_j \subset [t_j \,..\, t_{j+k}]$ and $\|h_j\|_\infty \le d_{k,1}/(t_{j+k} - t_j)$ that, on $S_{k,\mathbf{t}}$, agrees with $\lambda_{j,k,\mathbf{t}}$. This implies that, for any $g \in \mathbf{L}_1$ and for $1 \le p \le \infty$,

$$|\mu_j g| \le \frac{d_{k,1}}{(t_{j+k} - t_j)^{1/p}} \|g\|_p([t_j \,..\, t_{j+k}]).$$

The corresponding linear map

$$Q : g \mapsto \sum_j \mu_j g\, B_{j,k}$$

is defined for any $g \in \mathbf{L}_1(I_{k,\mathbf{t}})$, hence for any $g \in \mathbf{L}_p(I_{k,\mathbf{t}})$. 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 $x \in [t_l \,..\, t_{l+1})$,

(144)
$$\begin{aligned} Qg(x) = \sum_{j=l+1-k}^{l} \mu_j g\, B_j(x) &\le \sum_{j=l+1-k}^{l} |\mu_j g|\, B_j(x) \\ &\le d_{k,1}\|g\|_p([t_{l+1-k} \,..\, t_{l+k}]) \sum_{j=l+1-k}^{l} B_j(x)/(t_{j+k} - t_j)^{1/p} \\ &\le d_{k,1}\|g\|_p([t_{l+1-k} \,..\, t_{l+k}]) \max_{j\in\{l+1-k,\ldots,l\}} (t_{j+k} - t_j)^{-1/p}. \end{aligned}$$
“goodlocal

Since also $g - Qg = (1-Q)(g-q)$ for any $q \in \Pi_{<k}$, we conclude that, for any $g$,

“quasierrorbound (145) $$\|g - Qg\|_p([t_l \,..\, t_{l+1}]) \le (1 + d_{k,1})\,\mathrm{dist}\,(g, \Pi_{<k})_p([t_{l+1-k} \,..\, t_{l+k}])$$

(since $(t_{l+1} - t_l)/(t_{j+k} - t_j) \le 1$ for $j = l+1-k, \ldots, l$). In fact, this conclusion can already be reached if we only take care that $\mu_j$ agree with $\lambda_j$ on $\Pi_{<k}$, all $j$, while still $|\mu_j g| \le d_{k,1}\|g\|_1([t_j \,..\, t_{j+k}])/(t_{j+k} - t_j)$.

The resulting map $Q$ is called a **good quasiinterpolant of order** $k$, with ‘good’ referring to the uniform localness expressed by (144), 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 $Q$ that 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 (145) is local; it is in terms of how well $g$ can be approximated locally from polynomials of order $k$. That local distance is best estimated with the aid of

“whitney **(146) Whitney's Theorem.** *For any finite interval $I$,*

$$\mathrm{dist}\,(g, \Pi_{<k})_p(I) \sim \omega_k(g, |I|)_p.$$

**Proof:** For all $f \in \Pi_{<k}$, $\omega_k(g, |I|)_p = \omega_k(g - f, |I|)_p \le 2^k \|g - f\|_p(I)$, hence $\omega_k(g, |I|)_p \le \mathrm{const}_k\,\mathrm{dist}\,(g, \Pi_{<k})_p(I)$.

For the converse inequality, let $I =: [a \,..\, b]$ and start with an arbitrary $f \in W_p^{(k)}(I)$. With

$$T_k f := \sum_{j<k} D^j f(a)(\cdot - a)^j / j!$$

its truncated Taylor series, we find

$$\begin{aligned} |f(x) - T_k f(x)| &= |\int_I (x - \cdot)_+^{k-1}/(k-1)!\; D^k f| \\ &\le \|D^k f\|_p/(k-1)!\;\big(\int_I |(x-\cdot)_+^{k-1}|^{p^*}\big)^{1/p^*} \\ &\le \|D^k f\|_p/(k-1)!\;|I|^{k-1/p}, \end{aligned}$$

hence, for any $1 \le q \le \infty$,

$$\|f - T_k f\|_q(I) \le \|D^k f\|_p/(k-1)!\;|I|^{k-1/p+1/q}.$$

Consequently,

$$\mathrm{dist}\,(g, \Pi_{<k})_p(I) \le \|g - T_k f\|_p(I) \le \|g - f\|_p + |I|^k \|D^k f\|_p/(k-1)!,$$

and, as $f \in W_p^{(k)}(I)$ is arbitrary here, we get

$$\mathrm{dist}\,(g, \Pi_{<k})_p(I) \le \mathrm{const}_k K(g, |I|^k; \mathbf{L}_p(I), W_p^{(k)}(I)),$$

while, from Theorem (68), we know that $\omega_k(g,t)_p \sim K(f, t^k; \mathbf{L}_p, W_p^{(k)})$. □

We conclude from Whitney's theorem and from (145) that

"splprejackson (147) $$\operatorname{dist}(g, S_{k,\mathbf{t}})_p \le \operatorname{const}_k \omega_k(g, |\mathbf{t}|)_p, \quad 1 \le p \le \infty.$$

More than that, since

$$\omega_k(g,h)_p(I) = O(h^k) \quad \Longleftrightarrow \quad g \in W_p^{(k)}(I),$$

with

$$\sup_h \frac{\omega_k(g,h)_p}{h^k} = \begin{cases} \|D^k g\|_p(I), & p > 1; \\ \operatorname{Var}(D^{k-1}g), & p = 1, \end{cases}$$

we have the local bound

$$\|g - Qg\|_p([t_l \,..\, t_{l+1}]) \le \operatorname{const}_k |t_{l+k} - t_{l+1-k}|^k \|D^k g\|_p([t_{l+1-k} \,..\, t_{l+k}]).$$

For $p = \infty$, this suggests that, in approximating some $g$ that is smooth except for some isolated singularities, the knot sequence $\mathbf{t}$ be chosen so as to make

$$l \mapsto |t_l - t_{l+1}|^k \|D^k g\|_\infty([t_l \,..\, t_{l+1}])$$

approximately constant. This is equivalent to making the map

$$l \mapsto |t_l - t_{l+1}| \big(\|D^k g\|_\infty([t_l \,..\, t_{l+1}])\big)^{1/k}$$

approximately constant, or, at least for a large knot sequence, making

$$l \mapsto |t_l - t_{l+1}| |D^k g([t_l \,..\, t_{l+1}])|^{1/k}$$

approximately constant. If there are to be $n$ knot intervals, then we can achieve this (approximately), by choosing $t_0 = 0 < t_1 < \cdots < t_n = 1$ so that

$$l \mapsto \int_{t_l}^{t_{l+1}} |D^k g|^{1/k}$$

is constant. With that choice,

$$\frac{1}{n^k} \|D^k g\|_{1/k} = \Big( \int_{t_l}^{t_{l+1}} |D^k g|^{1/k} \Big)^k \sim |t_{l+1} - t_l|^k \|D^k g\|_\infty([t_l \,..\, t_{l+1}]),$$

hence

$$\|g - Qg\|_\infty \le \operatorname{const}_k n^{-k} \|D^k g\|_{1/k}.$$

While this argument lacks some details, it makes the following essential point: *In order to achieve approximation order $n^{-k}$ from the set of splines of order $k$ with $n$ interior knots, it is sufficient to have*

$$\|D^k g\|_{1/k}$$

*finite.* For example, this norm is finite for functions such as $|\cdot|^{1/2}$ on $[-1 \,..\, 1]$, ensuring therefore 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 from $\Pi_n$ to $|\cdot|^{1/2}$ on $[-1 \,..\, 1]$ cannot be better than $O(n^{-1/2})$, by Bernstein's Inverse Theorem, hence approximation by splines with $n$ equally spaced knots cannot be better, either, as we show in the next section.

## Jackson and Bernstein for splines with uniform knot sequence

We know from (147) that, for every $g \in W_p^{(k)}$,

"spljackson (148)
$$\operatorname{dist}(g, S_{k,\mathbf{t}})_p \le \mathrm{const}_k |\mathbf{t}|^k \|D^k g\|_p.$$

However, having $\operatorname{dist}(g, S_{k,\mathbf{t}})_p = O(|\mathbf{t}|^k)$ is, in general, no guarantee that $g \in W_p^{(k)}$ unless the knot sequences $\mathbf{t}$ involved are sufficiently 'generic'. Indeed, if every knot sequence considered contains the point 1/2, then the function $g := (\cdot - 1/2)_+^{k-1}$ can be approximated without error from $S_{k,\mathbf{t}}$ even though $g$ fails to have a $k$th derivative. But it is true that (148) cannot hold for *every* knot sequence $\mathbf{t}$ unless $g \in W_p^{(k)}$. In fact, this conclusion can already be reached if (148) holds for every *uniform* knot sequence. The essential point is that, for every point in the interval of approximation, there must be infinitely many knot sequences among those considered for which that point falls somewhere in the middle of a knot interval. Here are the details.

"splbernstein **(149) Theorem (Butler, DeVore, Richards).** *For* $m \in \mathbb{N}$, *let*

$$\mathbf{t}^{(m)} := (\ldots, 0, 0, 1/m, 2/m, \ldots, 1, 1, \ldots),$$

*and set*

$$E_m(g)_p := \operatorname{dist}(g, S_{k,\mathbf{t}^{(m)}})_p(I)$$

*with* $I := [0 \,..\, 1]$. *Then,*

$$\omega_k(g,\delta)_p \;\le\; \mathrm{const}_k \begin{cases} \left(\frac{1}{n}\sum_{m=n}^{2n} E_m(g)_p^p\right)^{1/p}, & p < \infty; \\ \max_{n\le m\le 2n} E_m(g)_\infty, & p = \infty, \end{cases}$$

*with*

$$n := \lfloor 1/\delta \rfloor.$$

**Proof:** Since

$$|\Delta_h^k g(x)| = |\Delta_h^k (g-f)(x)| \le 2^k \|g-f\|_\infty (x \,..\, x+kh)$$

for every $f \in \Pi_{<k}$, the heart of the proof is in the (nontrivial) observation (known as a **mixing lemma**, see(150) below) that, for all $n$, there exists $m \in \{n, \ldots, 2n\}$ so that $\operatorname{dist}(x, \mathbf{t}^{(m)}) \ge 1/(16n)$, hence, for our $n$ and for all $0 < h \le \delta_1 := 1/(16kn)$, since $kh \le k\frac{1}{16kn} = 1/(16n)$, there is $m \in \{n, \ldots, 2n\}$ so that $[x \,..\, x+kh] \cap \mathbf{t}^{(m)} = \emptyset$, implying that any $f \in S_{k,\mathbf{t}^{(m)}}$ is a polynomial on $[x \,..\, x+kh]$, hence, with $f$ a ba to $g$ from $S_{k,\mathbf{t}^{(m)}}$, we get

$$|\Delta_h^k g(x)| \le 2^k E_m(g)_\infty.$$

It follows that

$$\omega_k(g, \delta_1) \le 2^k \max_{n \le m \le 2n} E_m(g)_\infty.$$

Since $\delta \le 1/n = 16k\delta_1$, hence $\omega_k(g,\delta) \le \omega_k(g,1/n) \le (16k)^k\omega_k(g,\delta_1)$, this implies the claimed result for $p=\infty$.

As to $p<\infty$, the mixing lemma (150) implies that

$$\sum_{m=n}^{2n} \chi_{I_m}(x) \ge \frac{n}{64} \qquad \text{on } [0\,..\,1-kh],$$

with

$$I_m := \{x \in [0\,..\,1-kh] : [x\,..\,x+1/(16n)] \cap \mathbf{t}^{(m)} = \emptyset\}.$$

Therefore,

$$\frac{1}{64}\int_0^{1-kh} |\Delta_h^k g(x)|^p \,\mathrm{d}x \;\le\; \frac{1}{n}\sum_{m=n}^{2n}\int_{I_m} |\Delta_h^k g(x)|^p \,\mathrm{d}x \le \frac{2^{kp}}{n}\sum_{m=n}^{2n} E_m(g)_p^p.$$

This proves the result for $p<\infty$. □

For the record, here is the afore-mentioned

“mixinglemma **(150) Mixing Lemma.** *For any* $x \in [0\,..\,1]$ *and any* $n \in \mathbb{N}$,

$$\#\{n \le m \le 2n : \operatorname{dist}(x, \mathbf{t}^{(m)}) \ge 1/(16n)\} \;\ge\; n/64.$$

Its proof (see, e.g., [DeVore and Lorentz, *Constructive Approximation*: pp. 356-7]) relies on the fact that, for any $i, N \in \mathbb{N}$ with $i < N$,

$$\#\{N \le m \le 2N : \operatorname{dist}(i/N, \mathbf{t}^{(m)}) \ge 1/(6N)\} \;\ge\; N/16.$$

We conclude from the Theorem that

$$E_n(g)_p \sim \omega_k(g, 1/n)_p.$$

In particular,

$$E_n(g)_p = O(n^{-k}) \quad\iff\quad g \in \begin{cases} W_p^{(k)} & 1<p\le\infty; \\ W^{(k-1)}(BV) & p=1. \end{cases}$$

Further, we get the *saturation result*:

$$E_n(g)_p = o(n^{-k}) \quad\iff\quad g \in \Pi_{<k}.$$

For the characterization of other rates of convergence, we make use of the following standard way to measure convergence behavior.

## Approximation Spaces

It has become standard to measure the decay of $E_n(g) := \operatorname{dist}(g, M_n)$ of the distance of $g \in X$ from a given sequence $(M_n : n \in \mathbb{N})$ of subsets by comparing it to the sequence $(n^{-\alpha} : n \in \mathbb{N})$ for various values of $\alpha$. Most simply, one might ask to have

$$E_n(g) = O(n^{-\alpha}),$$

with the supremum over all such $\alpha$ then being the **approximation order** for $g$ provided by $(M_n : n \in \mathbb{N})$. However, in order to be able to *characterize* the class of functions $g$ with a given approximation order, a somewhat more subtle way of measuring approximation order turns out to be often needed.

Define

$$\|a\|_q^{(\alpha)} := \begin{cases} \big(\sum_n (n^\alpha |a(n)|)^q/n\big)^{1/q} & 0 < q < \infty; \\ \sup_n n^\alpha |a(n)| & q = \infty. \end{cases}$$

With this, we define

$$A_q^{(\alpha)} := \{a \in \mathbb{R}^{\mathbb{N}} : \Delta|a| \le 0, \|a\|_q^{(\alpha)} < \infty\},$$

and make the following observations. First,

$$\|a\|_\infty^{(\alpha)} < \infty \quad \iff \quad a(n) = O(n^{-\alpha}),$$

i.e., $A_\infty^{(\alpha)}$ consists exactly of all the antitone sequences $a$ that go to zero (at least) to order $-\alpha$. Further, $\|a\|_q^{(\alpha)} < \infty$ for some finite $q$ implies that $a(n) = o(n^{-\alpha})$. (Indeed, if $a(n) \ne o(n^{-\alpha})$, then $m(n)^\alpha |a(m(n))| \ge M > 0$ for some strictly increasing $m : \mathbb{N} \to \mathbb{N}$, and AGTASMAT $m(n-1) < m(n)/2$; thus, $(\|a\|_q^{(\alpha)})^q \ge \sum_n \sum_{m(n)/2 \le j \le m(n)} (j^\alpha |a(j)|)^q/j \ge \sum_n (2^{-\alpha} M)^q \sum_{m(n)/2 \le j \le m(n)} 1/j \to \infty$ (since $j^\alpha |a(j)| = (j/m(n))^\alpha m(n)^\alpha |a(m(n))| \ge 2^{-\alpha} M$ for $m(n)/2 \le j < m(n)$, and $\sum_{m(n)/2 \le j \le m(n)} 1/j \sim \ln 2$), a contradiction.) In particular

$$N := \|()^\alpha a\|_\infty < \infty,$$

hence

$$\infty > (\|a\|_q^{(\alpha)})^q =: N^q \sum_n |b(n)|^q/n \ge N^q \sum_n |b(n)|^r/n$$

for any $r > q$ since $|b(n)| \le 1$, all $n$. In particular,

$$\|a\|_r^{(\alpha)} = N\big(\sum_n |b(n)|^r/n\big)^{1/r}$$

is finite for any $r > q$. In other words,

$$q < r \quad \implies \quad A_q^{(\alpha)} \subset A_r^{(\alpha)}.$$

Finally, under the same assumption, and for any positive $\varepsilon$ and any $r$,

$$\sum_n |n^{\alpha-\varepsilon}a(n)|^r/n = \sum_n |n^\alpha a(n)|^r/n^{1+r\varepsilon} \le N^r \sum_n n^{-1-r\varepsilon} < \infty.$$

Hence,

$$\beta < \alpha \quad \Longrightarrow \quad A_q^{(\alpha)} \subset A_\infty^{(\alpha)} \subset A_r^{(\beta)}, \quad \text{all } q, r.$$

For a given sequence $(M_n : n \in \mathbb{N})$ of subsets of the nls $X$, one defines, correspondingly, the **approximation classes**

$$A_q^{(\alpha)}(X,(M_n)) := \{g \in X : \|(E_n(g) : n \in \mathbb{N})\|_q^{(\alpha)} < \infty\}$$

that single out all the elements $g$ of $X$ for which $E_n(g)$ goes to zero in a certain way. By the earlier discussion of $(\alpha,q) \mapsto \|a\|_q^{(\alpha)}$, we conclude that $(\alpha,q) \mapsto A_q^{(\alpha)}$ is antitone in $\alpha$ for arbitrary $q$ and isotone in $q$ for fixed $\alpha$, i.e.,

$$\beta < \alpha \Longrightarrow A_q^{(\alpha)} \subset A_r^{(\beta)},$$

$$q < r \Longrightarrow A_q^{(\alpha)} \subset A_r^{(\alpha)}.$$

There is a corresponding way to quantify and compare the speed at which, for a function $f$, $f(t)$ approaches 0 as $t \to 0$. Precisely, for such a function, one defines

$$\|f\|_q^{(\alpha)} := \begin{cases} \big(\int_I |t^{-\alpha}f(t)|^q \,\mathrm{d}t/t\big)^{1/q}, & q < \infty; \\ \sup_I |t^{-\alpha}f(t)|, & q = \infty, \end{cases}$$

with $I$ some suitable interval, e.g., $I = [0\,..\,1]$. Usually, $f$ is isotone. In that case, $\|f\|_q^{(\alpha)}$ is equivalent to (i.e., bounded above and below by certain $f$-independent positive multiples of) the following discrete versions, in which $I$ is replaced by the sequence (a) $I = (1/n : n \in \mathbb{N})$ or (b) $I = (2^{-n} : n = 0, 1, \ldots)$, and, correspondingly, on the interval $[t_{n+1}\,..\,t_n]$, $\mathrm{d}t/t$ is replaced by $\sim (t_n - t_{n+1})/t_n$, which, for (a), is $1/(n+1)$, and, for (b), is a constant. For $q < \infty$, this gives the equivalent discrete versions

$$\|f\|_q^{(\alpha)}(1/\mathbb{N}) := \|(f(1/n) : n \in \mathbb{N})\|_q^{(\alpha)}$$

and

$$\|f\|_q^{(\alpha)}(2^{-\mathbb{N}}) := \big(\sum_n (2^{n\alpha}f(2^{-n}))^q\big)^{1/q},$$

respectively.

With these definitions in place, consider again the earlier result that

$$E_n(g)_p := \mathrm{dist}\,(g, S_{k,\mathbf{t}^{(n)}})_p \sim \omega_k(g, 1/n)_p,$$

hence

$$g \in A_q^{(\alpha)}(\mathbf{L}_p, (S_{k,\mathbf{t}^{(n)}} : n \in \mathbb{N})) \quad \Longleftrightarrow \quad \|\omega_k(g,\cdot)_p\|_q^{(\alpha)} < \infty.$$

The latter condition appeared some time ago in the study of approximation order from trigonometric polynomials, leading to what we now call **Besov spaces**. Precisely,

$$B_q^{(\alpha)}(\mathbf{L}_p) := \{g \in \mathbf{L}_p : |g|_{B_q^{(\alpha)}(\mathbf{L}_p)} := \|\omega_{\lfloor\alpha\rfloor+1}(g,\cdot)_p\|_q^{(\alpha)} < \infty\}, \quad 0 < \alpha, 0 < q \le \infty.$$

These are complete metric spaces with the metric given by the (quasi-)norm

$$\|g\|_{B_q^{(\alpha)}(\mathbf{L}_p)} := \|g\|_p + |g|_{B_q^{(\alpha)}(\mathbf{L}_p)}.$$

(This is only a **quasi-**norm when $p < 1$ since then one only has $\|a+b\| \le \text{const}(\|a\| + \|b\|)$ for some const instead of the triangle inequality.)

**(151) Fact.** *For any* $r > \alpha$,

$$g \mapsto \|\omega_r(g,\cdot)_p\|_q^{(\alpha)}$$

*provides a (quasi-)seminorm on* $B_q^{(\alpha)}(\mathbf{L}_p)$ *equivalent to* $|\cdot|_{B_q^{(\alpha)}(\mathbf{L}_p)}$.

One direction of this claim is obvious since

$$\omega_{r+m}(g,t)_p \le 2^m \omega_r(g,t)_p.$$

For the other direction, one needs a result like

"marchaud **(152) Marchaud's Theorem.** *For* $g \in \mathbf{L}_p(I)$,

$$\omega_r(g,t)_p \le \text{const}_r t^r \Big( \int_t^{|I|} \frac{\omega_{r+m}(g,s)_p}{s^{r+1}} \,\mathrm{d}s + \frac{\|g\|_p}{|I|^r} \Big).$$

along with

"hardy **(153) Hardy's Inequality.** *For* $\alpha > 0$, *and* $1 \le q \le \infty$, *and* $f$ *any positive measurable function,*

$$\int_0^\infty \big[t^{-\alpha} \int_0^t f(s) \frac{\mathrm{d}s}{s}\big]^q \frac{\mathrm{d}t}{t} \le \frac{1}{\alpha^q} \int_0^\infty \big[t^{-\alpha} f(t)\big]^q \frac{\mathrm{d}t}{t}.$$

With this, we reach the conclusion that

"classisbesov (154) $$A_q^{(\alpha)}(\mathbf{L}_p(0\,..\,1), (S_{k,\mathbf{t}^{(n)}} : n \in \mathbb{N})) \;=\; B_q^{(\alpha)}(\mathbf{L}_p(0\,..\,1)), \qquad 0 < q \le \infty,\ 0 < \alpha < k.$$

But we have to question just what we have gained by this somewhat formal exercise. The gain is substantial to the extent that we understand

## Besov spaces

Such understanding comes from knowing more about 'typical' elements and/or from knowing alternative characterizations of such spaces. Among these will be those derived from the fact that Besov spaces turn out to be approximation spaces for various other sequences of approximating sets, among these $S_{k,n} :=$ splines with $n$ free knots and $\Pi_n/\Pi_n :=$ rational functions with numerator and denominator degree $\le n$, both classical examples of *nonlinear* approximating sets, to be discussed next.

As for specific examples, consider the Heaviside function

$$g := ()_+^0,$$

as an element of $\mathbf{L}_p[-1\,..\,1]$, say. For $0 < h < 1$, $\|\Delta_h g\|_p^p = h$, hence

$$\omega_1(g,h)_p = h^{1/p}.$$

Therefore, for $0 < \alpha < 1$ and $0 < q < \infty$,

$$(\|\omega(g,\cdot)_p\|_q^{(\alpha)})^q \sim \int_0^1 (t^{-\alpha}t^{1/p})^q\,\mathrm{d}t/t = \int_0^1 ()^{(-\alpha+1/p)q-1},$$

and this is finite iff $\alpha < 1/p$. For $q = \infty$, we look instead at

$$\sup_t t^{-\alpha}t^{1/p},$$

and this is finite iff $\alpha \le 1/p$. Consequently,

$$()_+^0 \in B_q^{(\alpha)}(\mathbf{L}_p) \iff \begin{cases} \alpha < 1/p, & \text{if } 0 < q < \infty; \\ \alpha \le 1/p, & \text{if } q = \infty. \end{cases}$$

Note the minor, yet decisive, role played by the parameter $q$ here. Note also that $\alpha$ can be large provided we are willing to consider $p < 1$. Finally, note the implication that any function with *finitely many* jump discontinuities but that is otherwise smooth lies in $B_q^{(\alpha)}(\mathbf{L}_p)$ for exactly the same triples $(\alpha, q, p)$.

More generally,

$$(\Delta_h^k ()_+^{k-1})(x) = k!h^k\,\boldsymbol{\Delta}(0,h,\ldots,kh)(x+\cdot)_+^{k-1} = (k-1)!h^{k-1}B(-x|0,h,\ldots,kh),$$

hence, $\|\Delta_h^k ()_+^{k-1}\|_p \sim h^{k-1+1/p}$, and so, for $0 < q < \infty$,

$$(\|\omega_k(()_+^{k-1},\cdot)_p\|_q^{(\alpha)})^q \sim \int_0^1 (t^{-\alpha}t^{k-1+1/p})^q\,\mathrm{d}t/t = \int_0^1 ()^{(-\alpha+k-1+1/p)q-1}$$

while

$$\|\omega_k(()_+^{k-1},\cdot)_\infty\|_\infty^{(\alpha)} \sim \sup_t t^{-\alpha+k-1+1/p}.$$

Therefore,

$$()_+^{k-1} \in B_q^{(\alpha)}(\mathbf{L}_p) \iff \begin{cases} \alpha < k-1+1/p, & \text{if } 0 < q < \infty; \\ \alpha \le k-1+1/p, & \text{if } q = \infty. \end{cases}$$

Other examples include:

$$B_\infty^{(r)}(\mathbf{L}_p(I)) \supset \begin{cases} W_p^{(r)}(I), & \text{if } 1 < p < \infty; \\ W^{(r-1)}BV(I), & \text{if } p = 1, \end{cases}, \quad r \in \mathbb{N},$$

i.e., for $1 \le p < \infty$ and integer $r$, $B_\infty^{(r)}(\mathbf{L}_p(I))$ contains the **Sobolev space** of all functions on $I$ with absolutely continuous $(r-1)$st derivative and (a) $r$th derivative in $\mathbf{L}_p$ if $p > 1$; (b) $(r-1)$st derivative of bounded variation, if $p = 1$. There is no equality here, the only related equality being

$$B_2^{(r)}(\mathbf{L}_2) = W_2^{(r)}.$$

For $p = \infty$, one has

$$B_\infty^{(r)}(C(I)) = \mathrm{Lip}(r+1, C(I)),$$

with the special case $B_\infty^{(1)}(C(I))$ equal the **Zygmund space**, i.e., slightly larger than $\mathrm{Lip}_1(I)$. Note the somewhat more subtle description in the extreme cases $p = 1, \infty$.

More generally, for nonintegral $\alpha$, and with $\alpha =: r-1+\beta$ for some $r \in \mathbb{N}$ and $\beta \in (0\,..\,1)$,

$$B_\infty^{(\alpha)}(\mathbf{L}_p(I)) = \mathrm{Lip}(\alpha, X_p(I)),$$

the space of all functions on $I$ with absolutely continuous $(r-1)$st derivative and $r$th derivative in $\mathrm{Lip}_\beta(I)_p$. In view of the fact that

$$\mathrm{Lip}_\beta(I)_p := \{g \in \mathbf{L}_p(I) : \sup_h \|\Delta_h g\|_p / h^\beta < \infty\},$$

this is certainly just a tautology.

Besov spaces are helpful also because they appear as the 'right' spaces in interpolation between standard spaces and in the Sobolev embedding theorem and its generalization. For the discussion of these matters, it is very helpful to follow Ron DeVore's advice and view the whole situation by representing *all* the spaces $(B_q^{(\alpha)}(\mathbf{L}_p) : 0 < q \le \infty)$ by the point

$$(1/p, \alpha)$$

in $\mathbb{R}_+^2$, as in Figure (155).

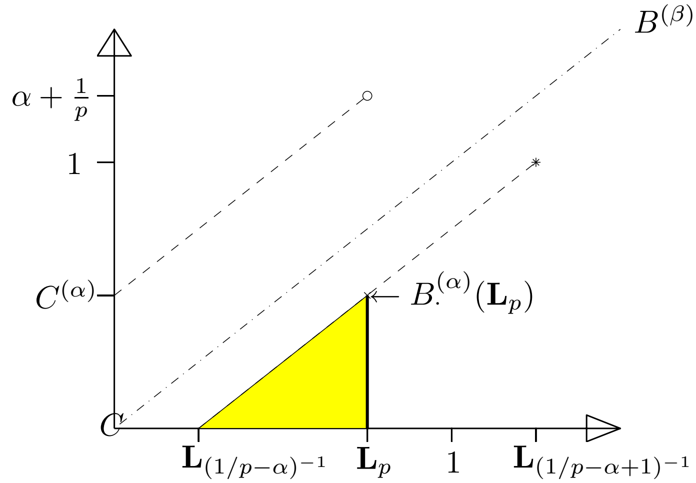
Besov spaces arise naturally in interpolation between the spaces $\mathbf{L}_p$ and $W_p^{(r)}$. Briefly, for any pair $(X_0, X_1)$ of (quasi-)normed spaces with

$$X_1 \hookrightarrow X_0,$$

i.e., $X_1$ continuously imbedded in $X_0$, one obtains a two-parameter continuum of spaces

$$X_1 \subset (X_0, X_1)_{\theta,q} \subset X_0$$




(155) Figure. DeVore's diagram associates with the point $(\frac{1}{p}, \alpha)$ the whole family $B_{\cdot}^{(\alpha)}(\mathbf{L}_p) := (B_q^{(\alpha)}(\mathbf{L}_p[0 \mathinner{..} 1]) : 0 < q \le \infty)$ of Besov spaces, to facilitate discussion (and retention) of basic facts about embeddings, interpolation in smoothness spaces, as well as the essential difference between linear *vs* nonlinear approximation.

The shaded triangle comprises all the Besov spaces into which the marked space $B_q^{(\alpha)}(\mathbf{L}_p)$ is continuously embedded, with the precise choice of the secondary parameter, $q$, of import only along the slanted edge.

"figdevore

as

$$X_{\theta,q} := (X_0, X_1)_{\theta,q} := \{g \in X_0 : |g|_{\theta,q} := \|K(g,\cdot)\|_q^{(\theta)} < \infty\},$$

with

$$K(g,t) := K(g,t;X_0,X_1) := \inf_{f \in X_1} (\|g-f\|_{X_0} + t\|f\|_{X_1})$$

the K-functional for the pair $(X_0, X_1)$.

The main result concerning this interpolation of spaces is the following

"thminterp **(156) Theorem.** *Let $X_1 \hookrightarrow X_0$ and $Y_1 \hookrightarrow Y_0$ be pairs of complete (quasi-)normed spaces and assume that the linear map $U$ maps $X_i$ boundedly into $Y_i$, $i = 0, 1$. Then, for $0 < q \le \infty$ and $0 < \theta < 1$, $U$ also maps each $X_{\theta,q}$ boundedly into $Y_{\theta,q}$ and, with*

$$M_i := \|U : X_i \to Y_i\|, \quad i = 0, 1,$$

*one has*

$$\|U : X_{\theta,q} \to Y_{\theta,q}\| \le M_0^{1-\theta} M_1^{\theta}.$$

(See Theorem 7.1 in Chapter 6 of DeVore-Lorentz.)

As an application, take $X_0 = \mathbf{L}_p(I)$ and $X_1 = W_p^{(k)}(I)$. Then, for any $0 < \alpha < k$,

$$(X_0, X_1)_{\alpha/k,q} = B_q^{(\alpha)}(\mathbf{L}_p),$$

(using Theorem (68)). Further, with $Y_0 = \mathbf{L}_p = Y_1$, we take for $U$ the error $1 - Q$ in the quasi-interpolant $Q$ introduced earlier. From (145) and Whitney's Theorem (146), we know that then

$$M_1 := \|(1-Q) : X_1 \to X_0\| \le \text{const}_k |\mathbf{t}|^k,$$

while

$$M_0 := \|(1-Q) : X_0 \to X_0\| \le 1 + d_{k,1} < \infty.$$

Therefore, for any $0 < \alpha < k$,

$$\|(1-Q) : B_q^{(\alpha)}(\mathbf{L}_p) \to \mathbf{L}_p\| \le \text{const}_k |\mathbf{t}|^\alpha,$$

thus providing (a lower bound on) the approximation order from $S_{k,\mathbf{t}}$ to elements of $B_q^{(\alpha)}(\mathbf{L}_p)$, hence, in particular, of $L_p^{(\alpha)}(I)$ in case $\alpha$ is an integer.

## Nonlinear approximation

We know from Theorem (149) that we cannot have $\text{dist}\,(g, S_{k,\mathbf{t}})_p = O(\mathbf{t}^k)$ for all $\mathbf{t}$ unless $g \in W_p^{(k)}$. Yet, we already observed that some functions without a $k$th derivative, like $g := {}^1\!/\!_2$ on $I := [0\,..\,1]$, can nevertheless be approximated to $O(n^{-k})$ by a spline of order $k$ with $n$ suitably chosen interior knots.

The basic results here are the following. Let

$$M_n := S_{k,n}$$

be the space of all splines of order $k$ on $I$ with $< n$ interior knots, hence with at most $n$ polynomial pieces. Then, $M_n$ is scale-invariant but fails to be closed under addition. However,

$$M_n + M_n \subset M_{2n},$$

hence Theorem (56) is applicable here provided we can produce compatible Jackson and Bernstein inequalities. These were obtained not all that long ago by Petrushev.

"thmpetrushev **(157) Theorem (Petrushev).** *Let $0 < p < \infty$ and, correspondingly,*

$$B^{(\alpha)} := B_\gamma^{(\alpha)}(\mathbf{L}_\gamma(0\,..\,1)), \qquad \gamma := 1/(\alpha + 1/p).$$

*then, for $0 < \alpha < k$,*

*(i)* $\forall g \in \mathbf{L}_p, \quad \text{dist}\,(g, S_{k,n}) \le \text{const}_k n^{-\alpha} |g|_{B^{(\alpha)}}$.

*(ii)* $\forall m \in S_{k,n}, \quad \|m\|_{B^{(\alpha)}} \le \text{const}_k n^\alpha \|m\|_p$.

(See Theorem 8.2 in Chapter 12 of DeVore-Lorentz.) In light of Peetre's Theorem (56), this says that these particular Besov spaces consist exactly of those functions that can be approximated in $\mathbf{L}_p$ to $O(n^{-\alpha})$ by splines with $n$ polynomial pieces. To get that kind of approximation order for splines with $n$ polynomial pieces and a uniform knot sequence (i.e., linear approximation), we need $g$ to lie in $B_\infty^{(\alpha)}(\mathbf{L}_p)$. In the DeVore diagram (155), both spaces lie on the same horizontal line, but the latter is much further to the left (i.e., much smaller) than the former, and so nicely illustrates the gain available to nonlinear approximation.

As a quick example, consider the simplest possible case, that of approximation from $M_n := S_{1,n}$, the (nonlinear) space of all step functions on $[0\,..\,1]$ with at most $n$ different values. Already in 1961, Kahane proved the following neat result:

"propkahane **(158) Proposition (Kahane).** *For* $g \in C([0\,..\,1])$, $\operatorname{dist}(g, S_{1,n}) \le M/(2n)$ *for all* $n \in \mathbb{N}$ *if and only if* $\operatorname{Var}(g) \le M$.

**Proof:** There is nothing to prove if $M$ or $\operatorname{Var}(g)$ are infinite, hence assume that both are finite.

Choose $0 = t_0 < \cdots < t_n = 1$ so that $\operatorname{Var}(g)(t_i\,..\,t_{i+1}) \le \operatorname{Var}(g)/n$, all $i$. Let $f \in S_{1,n}$ be such that, on $(t_i\,..\,t_{i+1})$, $f$ equals the midpoint of the interval $g([t_i\,..\,t_{i+1}])$, all $i$. Then, $\|g - f\|_\infty \le \frac{1}{2}\operatorname{Var}(g)/n$.

Conversely, for any $f \in S_{1,n}$, and any $0 = x_0 < \cdots < x_m = 1$,

$$|g(x_i) - g(x_{i-1})| \le 2\|g - f\|_\infty \# f([x_{i-1}\,..\,x_i]),$$

therefore

$$\sum_i |g(x_{i-1}) - g(x_i)| \le 2\|g - f\|_\infty \sum_{i=1}^m \# f([x_{i-1}\,..\,x_i]) \le 2(n+m)\|g - f\|_\infty.$$

Hence, if $\operatorname{dist}(g, S_{1,n}) \le M/(2n)$, then, for any $\varepsilon > 0$,

$$\sum_i |g(x_{i-1}) - g(x_i)| \le (M + \varepsilon)(1 + m/n),$$

therefore, by letting $n \to \infty$, $\operatorname{Var}(g) \le M + \varepsilon$. □

Thus, while $\operatorname{dist}(g, S_{1,\mathbf{t}^{(n)}})_\infty = O(1/n)$ requires $g$ to lie in $W_\infty^{(1)}$, getting $\operatorname{dist}(g, S_{1,n}) = O(1/n)$ only requires $g \in \mathrm{BV}$, i.e., in $B_\infty^{(1)}(\mathbf{L}_1)$. In the DeVore diagram (155), $W_\infty^{(1)}$ lies vertically above $C$, while the latter space lies on the 45-degree line emanating from $C$. Petrushev's result shows this to hold for splines of general order $k$. Namely, the approximation order from splines of order $k$ cannot exceed $k$, but that order is achieved by nonlinear approximation (i.e., approximation from $(S_{k,n} : n \in \mathbb{N})$) to much rougher functions than is possible by linear approximation (i.e., approximation from $(S_{k,\mathbf{t}^{(n)}} : n \in \mathbb{N})$).

See DeVore's survey article on Nonlinear Approximation, in *Acta Numerica, 1998.*