

Mathematical foundations of Machine Learning 2023 – lesson 2

Shai Dekel



Banach Spaces

Definition Banach space is a complete normed vector space B over a field $F = \{\mathbb{R}, \mathbb{C}\}$,

Vector space: $\exists 0 \in B, \forall f, g \in B, \alpha, \beta \in F \Rightarrow \alpha f + \beta g \in B$.

Complete: Every Cauchy sequence in B converges to an element of B .

Norm:

- i. $f \neq 0 \Rightarrow \|f\| > 0$
- ii. $\|\alpha f\| = |\alpha| \|f\|, \quad \forall \alpha \in F,$
- iii. Triangle inequality $\|f + g\| \leq \|f\| + \|g\|$

Measure

In this course we shall mostly use the standard Lebesgue measure – the volume of a (measurable) set.

Examples: $\Omega = [0, 2]^n \subset \mathbb{R}^n, \mu(\Omega) = |\Omega| = 2^n$.

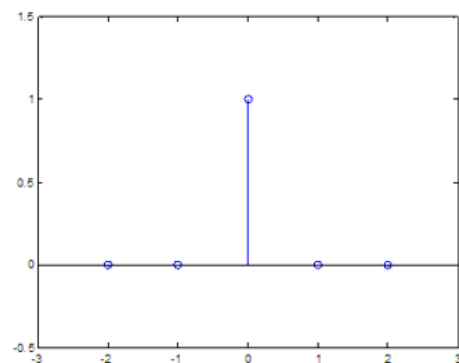
We will need the notion of zero measure (volume). Example: a set of discrete points.

Lp Spaces

$\Omega \subseteq \mathbb{R}^n$ domain. Examples: $\Omega = [a, b] \subset \mathbb{R}$, $\Omega = [0, 1]^n \subset \mathbb{R}^n$, $\Omega = \mathbb{R}^n$.

$$\|f\|_{L_p(\Omega)} := \begin{cases} \left(\int_{\Omega} |f(x)|^p dx \right)^{1/p}, & 0 < p < \infty, \\ \operatorname{ess\,sup}_{x \in \Omega} |f(x)|, & p = \infty. \end{cases}$$

$$\operatorname{ess\,sup}_x |f(x)| := \sup_{A > 0} \left\{ A > 0 : \left| \{x : |f(x)| \geq A\} \right| > 0 \right\}.$$



$1 \leq p \leq \infty$ Banach spaces

$0 < p < 1$ Quasi-Banach spaces (quasi-triangle inequality holds)

$$\|f + g\|_p^p \leq \|f\|_p^p + \|g\|_p^p.$$

Theorem [Hölder] $1 \leq p \leq \infty$, $f \in L_p(\Omega)$, $g \in L_{p'}(\Omega)$

$$\left| \int_{\Omega} fg \right| \leq \int_{\Omega} |fg| = \|fg\|_1 \leq \|f\|_p \|g\|_{p'}, \quad \frac{1}{p} + \frac{1}{p'} = 1.$$

Lemma Young's inequality for products,

$$ab \leq \frac{a^p}{p} + \frac{b^{p'}}{p'}, \quad \frac{1}{p} + \frac{1}{p'} = 1, \quad \forall a, b \geq 0.$$

Proof of lemma The logarithmic function is concave. Therefore

$$\begin{aligned} \log \left(\frac{1}{p} a^p + \frac{1}{p'} b^{p'} \right) &= \log \left(\frac{1}{p} a^p + \left(1 - \frac{1}{p} \right) b^{p'} \right) \\ &\geq \frac{1}{p} \log(a^p) + \frac{1}{p'} \log(b^{p'}) \\ &= \log(a) + \log(b) = \log(ab). \end{aligned}$$

Since the logarithmic function is increasing, we are done (or we take exp on both sides).

□

Proof of theorem If $p = \infty$

$$\int_{\Omega} |fg| \leq \|f\|_{\infty} \int_{\Omega} |g| = \|f\|_{\infty} \|g\|_1.$$

The proof is similar for $p = 1$. So, assume now $1 < p < \infty$, $\|f\|_p = \|g\|_{p'} = 1$.

Integrating pointwise and applying Young's inequality almost everywhere, gives

$$\begin{aligned}\int_{\Omega} |f(x) g(x)| dx &\leq \int_{\Omega} \left(\frac{|f(x)|^p}{p} + \frac{|g(x)|^{p'}}{p'} \right) dx \\ &= \frac{1}{p} \int_{\Omega} |f(x)|^p dx + \frac{1}{p'} \int_{\Omega} |g(x)|^{p'} dx \\ &= \frac{1}{p} + \frac{1}{p'} = 1.\end{aligned}$$

Now assuming $f, g \neq 0$ (else, we're done)

$$\int_{\Omega} \frac{|f(x)|}{\|f\|_p} \frac{|g(x)|}{\|g\|_{p'}} dx \leq 1 \Rightarrow \int_{\Omega} |fg| \leq \|f\|_p \|g\|_{p'}$$

Schwartz inequality $p = 2$

$$|\langle f, g \rangle_2| = \left| \int_{\Omega} f \overline{g} \right| \leq \int_{\Omega} |fg| = \|fg\|_1 \leq \|f\|_2 \|g\|_2.$$

The L_p spaces not comparable on unbounded domains

Example We'll use $\Omega = \mathbb{R}$. Assume $0 < q < p < \infty$

Choose

$$f(x) := \begin{cases} 0 & |x| \leq 1 \\ \frac{1}{|x|^{1/q}} & |x| > 1 \end{cases}$$

We have $f \in L_p(\mathbb{R})$, $f \notin L_q(\mathbb{R})$

Now choose

$$f(x) := \begin{cases} \frac{1}{|x|^{1/p}} & |x| \leq 1 \\ 0 & |x| > 1 \end{cases}$$

We have $f \in L_q(\mathbb{R})$, $f \notin L_p(\mathbb{R})$

Theorem If $|\Omega| < \infty$, $0 < q < p$, $f \in L_p(\Omega)$ then

$$\|f\|_{L_q(\Omega)} \leq |\Omega|^{1/q-1/p} \|f\|_{L_p(\Omega)}.$$

Proof Define $r := p/q \geq 1$

$$\begin{aligned} \|f\|_q^q &= \int_{\Omega} |f|^q = \int_{\Omega} |f|^q \mathbf{1} \underset{\text{Holder}}{\leq} \left(\int_{\Omega} (|f|^q)^r \right)^{1/r} \left(\int_{\Omega} \mathbf{1}^{r'} \right)^{1/r'} \\ &= \left(\int_{\Omega} |f|^p \right)^{q/p} |\Omega|^{1-q/p} \end{aligned}$$

Theorem Minkowski for L_p spaces $1 \leq p \leq \infty$, $f, g \in L_p$,

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p .$$

Proof for $1 < p < \infty$ ($p = 1, \infty$ is easier). W.l.g $f, g \geq 0$. We apply Hölder twice,

$$\begin{aligned} \int (f + g)^p &= \int f (f + g)^{p-1} + \int g (f + g)^{p-1} \\ &\leq (\|f\|_p + \|g\|_p) \left(\int (f + g)^{(p-1)p'} \right)^{1/p'} \\ &= (\|f\|_p + \|g\|_p) \left(\int (f + g)^p \right)^{1-1/p} \\ &= (\|f\|_p + \|g\|_p) \underbrace{\left(\int (f + g)^p \right)^{-1/p}}_{\|f+g\|_p^{-1}} . \end{aligned}$$

Theorem For $0 < p < 1$, we have

$$(i) \quad \left\| \sum_k f_k \right\|_p^p \leq \sum_k \|f_k\|_p^p$$

$$(ii) \quad \|f + g\|_p \leq 2^{1/p-1} (\|f\|_p + \|g\|_p) \quad \text{or in general} \quad \left\| \sum_{k=1}^N f_k \right\|_p \leq N^{1/p-1} \sum_{j=1}^N \|f_k\|_p$$

Proof The quasi-triangle inequality (ii) is derived from (i), by using $1 \leq p^{-1} < \infty$,

$$\left\| \sum_{k=1}^N f_k \right\|_p \stackrel{(i)}{\leq} \left(\sum_{k=1}^N \|f_k\|_p^p \right)^{1/p} = \left(\sum_{j=1}^N 1 \cdot \|f_k\|_p^p \right)^{1/p} \leq \underbrace{\left(\sum_{k=1}^N 1^{\frac{1}{1-p}} \right)}_N^{(1-p)1/p} \left(\sum_{k=1}^N \|f_k\|_p^p \right)^{1/p} = N^{1/p-1} \sum_{k=1}^N \|f_k\|_p$$

To prove (i), we need the following lemma

Lemma I For $0 < p \leq 1$ and any sequence of non-negative $a = \{a_k\}$,

$$\left(\sum_k a_k \right)^p \leq \sum_k a_k^p$$

Proof We first prove $(a_1 + a_2)^p \leq a_1^p + a_2^p$ and then apply induction.

To prove the inequality use $h(t) := t^p + 1 - (t+1)^p$. $h(0) = 0$ and $h'(t) = pt^{p-1} - p(t+1)^{p-1} \geq 0$. Therefore, $h(t) \geq 0$, for $t \geq 0$. This gives $t^p + 1 \geq (t+1)^p$. Setting $t = a_1 / a_2$ gives

$$\left(\frac{a_1}{a_2} \right)^p + 1 \geq \left(\frac{a_1}{a_2} + 1 \right)^p \Rightarrow a_1^p + a_2^p \geq (a_1 + a_2)^p.$$

□

Proof of Theorem (i): Simply apply the lemma pointwise for $x \in \Omega$

$$\left\| \sum_k f_k \right\|_p^p \leq \int_{\Omega} \left(\sum_k |f_k(x)| \right)^p dx \leq \int_{\Omega} \left(\sum_k |f_k(x)|^p \right) dx = \sum_k \int_{\Omega} |f_k(x)|^p dx = \sum_k \|f_k\|_p^p.$$

□

Definition The space $l_p(\mathbb{Z})$, $0 < p \leq \infty$, is the space of sequences $a = \{a_k\}_{k \in \mathbb{Z}}$, for which the norm is finite

$$\|a\|_{l_p} := \begin{cases} \left(\sum_k |a_k|^p \right)^{1/p}, & 0 < p < \infty, \\ \sup_k |a_k|, & p = \infty. \end{cases}$$

Lemma II $l_p \subset l_q$ for $p \leq q$. That is, for any sequence $a = \{a_k\}$

$$\|a\|_{l_q} \leq \|a\|_{l_p}.$$

Proof Case of $q = \infty$, for any $j \in \mathbb{Z}$,

$$|a_j| = \left(|a_j|^p \right)^{1/p} \leq \left(\sum_k |a_k|^p \right)^{1/p} = \|a\|_{l_p}.$$

Therefore,

$$\|a\|_{l_\infty} = \sup_j |a_j| \leq \|a\|_{l_p}.$$

For $q < \infty$, we have

$$\left(\sum_k |a_k|^q \right)^{p/q} \leq \sum_k \left(|a_k|^q \right)^{p/q} = \sum_k |a_k|^p \Rightarrow \left(\sum_k |a_k|^q \right)^{1/q} \leq \left(\sum_k |a_k|^p \right)^{1/p}.$$

Hilbert spaces and $L_2(\Omega)$

Def Hilbert space H : Complete metric vector space induced by an inner product $\langle, \rangle: H \times H \rightarrow \mathbb{C}$.

Properties of the inner product:

- i. symmetric $\langle f, g \rangle = \overline{\langle g, f \rangle}$,
- ii. linear $\langle \alpha f_1 + \beta f_2, g \rangle = \alpha \langle f_1, g \rangle + \beta \langle f_2, g \rangle$,
- iii. Positive definite $\langle f, f \rangle \geq 0$, with $\langle f, f \rangle = 0 \Leftrightarrow f = 0$.

The natural norm $\|f\|_H := \langle f, f \rangle^{1/2}$ satisfies

- (i) Cauchy-Schwartz

$$|\langle f, g \rangle| \leq \|f\|_H \|g\|_H$$

- (ii) Triangle inequality

$$\|f + g\|^2 = \|f\|^2 + 2 \operatorname{Re} \langle f, g \rangle + \|g\|^2 \leq \|f\|^2 + 2 \|f\| \|g\| + \|g\|^2 = (\|f\| + \|g\|)^2$$

So, a Hilbert space is a Banach space.

Examples

- (i) $l_2(\mathbb{Z}) : \langle \alpha, \beta \rangle_{l_2} := \sum_{i \in \mathbb{Z}} \alpha_i \bar{\beta}_i, \| \alpha \|_2 := \left(\sum_{i \in \mathbb{Z}} |\alpha_i|^2 \right)^{1/2}$
- (ii) $L^2(\Omega) : f, g \text{ measurable}, \langle f, g \rangle := C_\Omega \int_\Omega f(x) \overline{g(x)} dx,$
- $$\|f\|_{L_2(\Omega)} = \|f\|_2 = \langle f, f \rangle^{1/2} = \left(C_\Omega \int_\Omega |f(x)|^2 dx \right)^{1/2}.$$

For $\Omega = \mathbb{R}^n, C_\Omega = 1$. For $\Omega = [-\pi, \pi]^n, C_\Omega = \frac{1}{(2\pi)^n}$.

Multivariate algebraic polynomials

We define $\Pi_{r-1}(\mathbb{R}^n)$: polynomials of degree $r-1$.

Let $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{Z}_+^n$, $|\alpha| := \sum_{i=1}^n \alpha_i$.

Monomial $x^\alpha := \prod_{i=1}^n x_i^{\alpha_i}$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.

Polynomial $P \in \Pi_{r-1}$

$$P(x) = \sum_{|\alpha| < r} a_\alpha x^\alpha$$

Spaces of smooth functions

Multivariate derivatives: A partial derivative of order r

$$\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{Z}_+^n, \quad D^\alpha f = \frac{\partial^r f}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}}, \quad |\alpha| := \sum_{i=1}^n \alpha_i = r.$$

Definition $C^r(\Omega)$: The space of all continuously differentiable functions of order r in the classical sense.

$$\|f\|_{C^r(\Omega)} := \sum_{|\alpha| \leq r} \|D^\alpha f\|_{L_\infty(\Omega)},$$

The *semi-norm*

$$|f|_{C^r(\Omega)} := \sum_{|\alpha|=r} \|D^\alpha f\|_\infty$$

Examples $C^r(\mathbb{R})$ Then $\|f\|_{C^r(\mathbb{R})} = \sum_{k=0}^r \|f^{(k)}\|_\infty$ is a norm $|f|_{C^r(\mathbb{R})} = \|f^{(r)}\|_\infty$ is a semi-norm with the polynomials of degree $r-1$ as a null-space

Sobolev spaces

Definition We define the space of *test-functions* $C_0^r(\Omega)$ - continuously r -differentiable with compact support in Ω .

Definition Sobolev spaces $W_p^r(\Omega)$, $1 \leq p \leq \infty$

Def I For $1 \leq p < \infty$, completion of $C_0^r(\Omega)$ with respect to the norm $\sum_{|\alpha| \leq r} \|\partial^\alpha f\|_{L_p(\Omega)}$. For $p = \infty$, we take

$$W_\infty^r(\Omega) := C^r(\Omega).$$

Def II Let $f \in L_p(\Omega)$. Now for $\alpha \in \mathbb{Z}_+^n$, $|\alpha| \leq r$, $g := \partial^\alpha f$ is the *distributional (generalized) derivative* of f if it is a function and for all $\phi \in C_0^r(\Omega)$

$$\int_\Omega g \phi = (-1)^{|\alpha|} \int_\Omega f \partial^\alpha \phi.$$

The Sobolev norm and semi-norm. We require that the distributional derivatives exist as functions(!) in $L_p(\Omega)$ and

$$\|f\|_{W_p^r(\Omega)} := \sum_{|\alpha| \leq r} \|\partial^\alpha f\|_{L_p(\Omega)} < \infty \qquad |f|_{W_p^r(\Omega)} := \sum_{|\alpha|=r} \|\partial^\alpha f\|_{L_p(\Omega)}.$$

Theorem $W_p^r(\Omega)$ is a Banach space

Modulus of smoothness

Def The *difference operator* Δ_h^r . For $h \in \mathbb{R}^n$ we define $\Delta_h(f, x) = f(x+h) - f(x)$. For general $r \geq 1$ we define

$$\Delta_h^r(f, x) = \underbrace{\Delta_h \circ \dots \circ \Delta_h}_r(f, x) = \sum_{k=0}^r \binom{r}{k} (-1)^{r-k} f(x+kh).$$

Remarks

1. For $\Omega \subset \mathbb{R}^n$, we modify to $\Delta_h^r(f, x) := \Delta_h^r(f, x, \Omega)$, where $\Delta_h^r(f, x) = 0$, in the case $[x, x+rh] \not\subset \Omega$. So for $\Omega = [a, b]$, $\Delta_h^r(f, x) = 0$ on $[b-rh, b]$, for any function.
2. As an operator on $L_p(\Omega)$, $1 \leq p \leq \infty$, we have that $\|\Delta_h^r\|_{L_p \rightarrow L_p} \leq 2^r$. Assume $\Omega = \mathbb{R}^n$, then

$$\|\Delta_h^r(f, \cdot)\|_p \leq \sum_{k=0}^r \binom{r}{k} \|f(\cdot+kh)\|_p = \sum_{k=0}^r \binom{r}{k} \|f\|_p = 2^r \|f\|_p$$

Def The *modulus of smoothness* of order r of a function $f \in L_p(\Omega)$, $0 < p \leq \infty$, at the parameter $t > 0$

$$\omega_r(f, t)_p := \sup_{|h| \leq t} \|\Delta_h^r(f, x)\|_{L_p(\Omega)}.$$

Example non continuous function. Let $\Omega = [-1, 1]$. $f(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 \leq x \end{cases}$

Let's compute $\omega_r(f, t)_{L_p([-1, 1])}$, $0 < t < 1$. For $0 < h \leq t$

$$\Delta_h(f, x) = \begin{cases} 0 & -1 \leq x \leq -h \\ 1 & -h < x \leq 0 \\ 0 & 0 < x \leq 1 \end{cases}$$

For $p = \infty$ we get $\omega_1(f, t)_{L_\infty([-1, 1])} = \sup_{|h| \leq t} \|\Delta_h f\|_{L_\infty([-1, 1])} = 1$.

For $p \neq \infty$ we get $\omega_1(f, t)_{L_p([-1, 1])} = \sup_{|h| \leq t} \|\Delta_h f\|_{L_p([-1, 1])} = t^{1/p}$.

$$\Delta_h^2(f, x) = \Delta_h(\Delta_h f, x) = \begin{cases} 0 & -1 \leq x \leq -2h \\ 1 & -2h < x \leq -h \\ -1 & -h < x \leq 0 \\ 0 & 0 \leq x \leq 1 \end{cases}$$

We get $\omega_2(f, t)_{L_p([-1, 1])} = (2t)^{1/p}$

In general, we get $\omega_r(f, t)_{L_p([-1, 1])} \leq C(r, p) t^{1/p}$

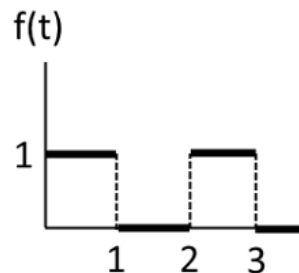
Quick jump ahead (Generalized Lipschitz / Besov smoothness) ... for $\alpha \leq 1/\tau$, $r = \lfloor \alpha \rfloor + 1$,

$$|f|_{B_{\tau,\infty}^\alpha} := \sup_{t>0} t^{-\alpha} \omega_r(f, t)_\tau \leq \sup_{0<t\leq 2} t^{-\alpha} \omega_r(f, t)_\tau \leq c \sup_{0<t\leq 2} t^{1/\tau-\alpha} < \infty.$$

We then say that f has α (weak-type) smoothness. Observe that in this example α can be arbitrarily large as long as the integration takes place with τ sufficiently small.

Machine learning perspective Let f be a ‘binary classification’ step function with M steps.

You will compute (assignment I) for $0 < \alpha < 1$, $|f|_{B_{\tau,\infty}^\alpha} \sim (2M)^{1/\tau}$.



- The feature space is ‘problematic’ for a simple ML model such as logistic regression.
- As a discontinuous function, ‘simpler’ smoothness function spaces do not contain it.
- Decision trees will find the clusters, so no need for DL.
- DL? For $M = 2^j$, the function can be realized/learnt by a neural network with $\sim j$ blocks,
- Each block has 4 neurons/features (2 layers with 2 neurons each)
- After the k -th block the function f_k has 2^{j-k} ‘steps’ with $|f_k|_{B_{\tau,\infty}^\alpha} \sim 2^{(j-k)/\tau}$.
- We can realize that the last representation layer as $\tilde{f}(\tilde{x}) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$, so it can be easily consumed by a logistic model.

Properties

1. $\omega_r(f, t)_p \leq 2^r \|f\|_{L_p(\Omega)}, 1 \leq p \leq \infty.$
2. $\omega_r(f, t)_p$ is non-decreasing in t
3. For $1 \leq p \leq \infty$ the **sub-linearity** property

$$\begin{aligned} \left| \Delta_h^r(f + g, x) \right| &= \left| \sum_{k=0}^r \binom{r}{k} (-1)^{r-k} (f + g)(x + kh) \right| \\ &\leq \left| \sum_{k=0}^r \binom{r}{k} (-1)^{r-k} f(x + kh) \right| + \left| \sum_{k=0}^r \binom{r}{k} (-1)^{r-k} g(x + kh) \right| \\ &= \left| \Delta_h^r(f, x) \right| + \left| \Delta_h^r(g, x) \right|. \end{aligned}$$

gives

$$\omega_r(f + g, t)_p \leq \omega_r(f, t)_p + \omega_r(g, t)_p.$$

4. For $N \geq 1$, $\omega_r(f, Nh)_p \leq N^r \omega_r(f, t)_p$, $1 \leq p \leq \infty$. We prove this using the property (**assignment**)

$$\Delta_{Nh}^r(f, x) = \sum_{k_1=0}^{N-1} \cdots \sum_{k_r=0}^{N-1} \Delta_h^r(f, x + k_1 h + \cdots + k_r h).$$

Let's see the case $r = 1$,

$$\begin{aligned} \Delta_{Nh}(f, x) &= f(x + Nh) - f(x) \\ &= f(x + Nh) - f(x + (N-1)h) + f(x + (N-1)h) - \cdots + f(x + h) - f(x) \\ &= \sum_{k=0}^{N-1} \Delta_h(f, x + kh) \end{aligned}$$

Then, for any $h \in \mathbb{R}^n$, $|h| \leq t$

$$\begin{aligned} \|\Delta_{Nh}^r(f, \cdot)\|_p &\leq \sum_{k_1=0}^{N-1} \cdots \sum_{k_r=0}^{N-1} \|\Delta_h^r(f, \cdot + k_1 h + \cdots + k_r h)\|_p \\ &= \sum_{k_1=0}^{N-1} \cdots \sum_{k_r=0}^{N-1} \|\Delta_h^r(f, \cdot)\|_p \leq N^r \omega_r(f, t)_p. \end{aligned}$$

Taking supremum over all $h \in \mathbb{R}^n$, $|h| \leq t$, gives $\omega_r(f, Nh)_p \leq N^r \omega_r(f, t)_p$. It is easy to see that for $0 < p < 1$, the same proof yields $\omega_r(f, Nh)_p \leq N^{r/p} \omega_r(f, t)_p$.

5. From (4) we get for $1 \leq p \leq \infty$,

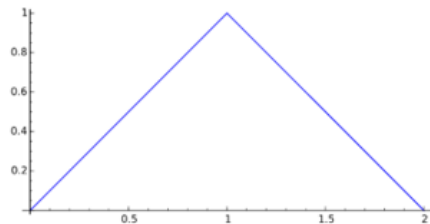
$$\omega_r(f, \lambda t)_p \leq (\lambda + 1)^r \omega_r(f, t)_p, \quad \lambda > 0$$

proof $\omega_r(f, \lambda t)_p \leq \omega_r(f, \lfloor \lambda + 1 \rfloor t)_p \leq (\lfloor \lambda + 1 \rfloor)^r \omega_r(f, t)_p \leq (\lambda + 1)^r \omega_r(f, t)_p.$

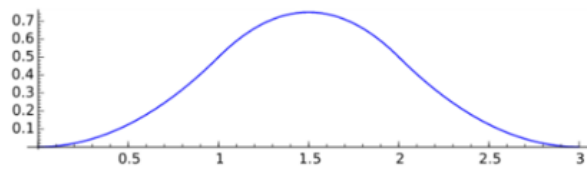
Theorem [connection between Sobolev and modulus] For $g \in W_p^r(\Omega)$, $1 \leq p \leq \infty$, we have that

$$\omega_r(g, t)_{L_p(\Omega)} \leq C(r, n) t^r |g|_{W_p^r(\Omega)}, \quad \forall t > 0.$$

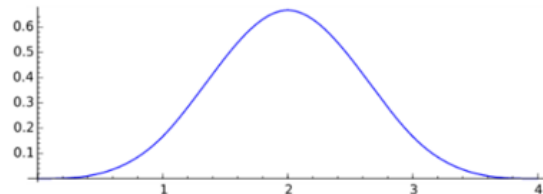
Proof for $\Omega = \mathbb{R}$. Recall the B-Splines, $N_1 = \mathbf{1}_{[0,1]^n}$. In general, $N_r := N_{r-1} * N_1 = \int_{\mathbb{R}^n} N_{r-1}(x-t) N_1(t) dt$.



N_2



N_3



N_4

- Properties:
 - Order r
 - Support $[0, r]^n$
 - Piecewise polynomial of degree $r-1$ with breakpoints (knots) at the integers
 - Smoothness $r-2$, thus in Sobolev W_p^{r-1} .
 - Tensor-product in multivariate case $N_r(x) := \tilde{N}_r(x_1) \times \cdots \times \tilde{N}_r(x_n)$, where $\tilde{N}_r$ is the univariate B-spline.
 - $\int_{\mathbb{R}^n} N_r(x) dx = 1$

Here, we use the fact that for $h \in \mathbb{R}^n$, $|\Delta_{-h}^r(f, x)| = |\Delta_h^r(f, x - rh)|$. So, w.l.g., for any $t > 0$, we can work with $0 < h \leq t$. Define $N_r(x, h) := h^{-1} N_r(h^{-1}x)$, $h > 0$. Let $g \in C^1(\mathbb{R})$. Then

$$\begin{aligned}
 h^{-1} \Delta_h(g, x) &= h^{-1} (g(x+h) - g(x)) \\
 &= h^{-1} \int_x^{x+h} g'(u) du \\
 &= \int_{\mathbb{R}} g'(x+u) N_1(u, h) du
 \end{aligned}$$

We claim that for $g \in C^r(\mathbb{R})$

$$h^{-r} \Delta_h^r(g, x) = \int_{\mathbb{R}} g^{(r)}(x+u) N_r(u, h) du$$

To see this, we apply induction

$$\begin{aligned} h^{-r} \Delta_h^r(g, x) &= h^{-1} h^{-(r-1)} \left(\Delta_h^{r-1}(g, x+h) - \Delta_h^{r-1}(g, x) \right) \\ &= h^{-1} \left(\int_{\mathbb{R}} g^{(r-1)}(x+h+u) N_{r-1}(u, h) du - \int_{\mathbb{R}} g^{(r-1)}(x+u) N_{r-1}(u, h) du \right) \\ &= h^{-1} \int_x^{x+h} \int_{-\infty}^{\infty} g^{(r)}(v+u) N_{r-1}(u, h) du dv \\ &= \int_{-\infty}^{\infty} N_{r-1}(u, h) \left(h^{-1} \int_x^{x+h} g^{(r)}(v+u) dv \right) du \end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} N_{r-1}(u, h) \left(h^{-1} \int_x^{x+h} g^{(r)}(v+u) dv \right) du \\
&= \int_{-\infty}^{\infty} N_{r-1}(u, h) \left(\int_{-\infty}^{\infty} g^{(r)}(v+u) N_1(v-x, h) dv \right) du \\
&\stackrel{v+u=x+y}{=} \int_{-\infty}^{\infty} N_{r-1}(u, h) \left(\int_{-\infty}^{\infty} g^{(r)}(x+y) N_1(y-u, h) dy \right) du \\
&= \int_{-\infty}^{\infty} g^{(r)}(x+y) \left(\int_{-\infty}^{\infty} N_{r-1}(u, h) N_1(y-u, h) du \right) dy \\
&= \int_{-\infty}^{\infty} g^{(r)}(x+y) N_r(y, h) dy
\end{aligned}$$

Now, let's see the proof for $p = 1$. Assume $g \in W_1^r(\mathbb{R}) \cap C^r(\mathbb{R})$. Let $0 < h \leq t$

$$\begin{aligned} \int_{\mathbb{R}} \left| \Delta_h^r(g, x) \right| dx &\leq h^r \int_{\mathbb{R}} \int_{\mathbb{R}} \left| g^{(r)}(x+u) \right| \left| N_r(u, h) \right| du dx \\ &\leq h^r \int_{\mathbb{R}} \left| N_r(u, h) \right| du \int_{\mathbb{R}} \left| g^{(r)}(x+u) \right| dx \\ &\leq t^r \int_{\mathbb{R}} \left| g^{(r)}(x) \right| dx \\ &\leq t^r \|g\|_{W_1^r(\mathbb{R})}. \end{aligned}$$

For general $1 \leq p < \infty$ we need Minkowski's inequality (**assignment**). It says that for measurable non-negative functions $\varphi: B \rightarrow \mathbb{R}$, $\rho: A \times B \rightarrow \mathbb{R}$

$$\left\{ \int_A \left(\int_B \varphi(y) \rho(x, y) dy \right)^p dx \right\}^{1/p} \leq \int_B \varphi(y) \left(\int_A \rho(x, y)^p dx \right)^{1/p} dy$$

Or written differently (as an integral generalization of the 'discrete' Minkowski inequality)

$$\left\| \int_B \varphi(y) \rho(\cdot, y) dy \right\|_{L_p(A)} \leq \int_B \varphi(y) \left\| \rho(\cdot, y) \right\|_{L_p(A)} dy \quad \Leftrightarrow \quad \left\| \sum_k \rho_k(\cdot) \right\|_p \leq \sum_k \left\| \rho_k(\cdot) \right\|_p$$

Using it we have for $g \in W_p^r(\mathbb{R}) \cap C^r(\mathbb{R})$

$$\begin{aligned}
 \int_{\mathbb{R}} \left| \Delta_h^r(g, x) \right|^p dx &\leq h^{pr} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \left| g^{(r)}(x+u) \right| \left| N_r(u, h) \right| du \right)^p dx \\
 &\leq h^{pr} \left(\int_{\mathbb{R}} \left| N_r(u, h) \right| \left\| g^{(r)}(\cdot + u) \right\|_{L_p(\mathbb{R})} du \right)^p \\
 &\leq h^{pr} \left(\int_{\mathbb{R}} \left| N_r(u, h) \right| \left\| g^{(r)} \right\|_{L_p(\mathbb{R})} du \right)^p \\
 &\leq t^{pr} \left\| g^{(r)} \right\|_{L_p(\mathbb{R})}^p \\
 &= t^{pr} \left| g \right|_{W_p^r(\mathbb{R})}^p.
 \end{aligned}$$

For a general function $g \in W_p^r(\mathbb{R})$ we use the density of $C^r(\mathbb{R}) \cap W_p^r(\mathbb{R})$ in $W_p^r(\mathbb{R})$ (**assignment**)

□

Corollary For any $P \in \Pi_{r-1}(\mathbb{R})$, $P(x) = \sum_{k=0}^{r-1} a_k x^k$,

$$h^{-r} \Delta_h^r(P, x) = \int_{\mathbb{R}} P^{(r)}(x+u) N_r(u, h) du = 0 \Rightarrow \Delta_h^r(P, x) = 0 \Rightarrow \omega_r(P, t)_p = 0$$

Marchaud inequalities

We know that for any $1 \leq k < r$, $1 \leq p \leq \infty$,

$$\omega_r(f, t)_p = \sup_{|h| \leq t} \|\Delta_h^r(f)\|_p = \sup_{|h| \leq t} \|\Delta_h^{r-k} \Delta_h^k(f)\|_p \leq 2^{r-k} \sup_{|h| \leq t} \|\Delta_h^k(f)\|_p = 2^{r-k} \omega_k(f, t)_p.$$

The direct inverse cannot be true. If we take $\Omega = [a, b]$ and a polynomial $P \in \Pi_{r-1}$, then $\omega_r(P, t)_p = 0$, but we don't necessarily have $\omega_k(P, t)_p = 0$ for $0 \leq k < r$.

Theorem. For any $1 \leq k < r$, $1 \leq p \leq \infty$,

$$\text{On } \Omega = \mathbb{R}, \quad \omega_k(f, t)_p \leq ct^k \int_t^\infty \frac{\omega_r(f, s)_p}{s^{k+1}} ds, \quad t > 0.$$

$$\text{On } \Omega = [a, b], \quad \omega_k(f, t)_p \leq ct^k \left(\int_t^{b-a} \frac{\omega_r(f, s)_p}{s^{k+1}} ds + \frac{\|f\|_p}{(b-a)^k} \right), \quad 0 < t \leq \frac{b-a}{r}.$$

Lip spaces

Def For a domain $\Omega \subset \mathbb{R}^n$ and $0 < \alpha \leq 1$, we shall say that $f \in Lip(\alpha) = Lip(\alpha, \infty)$, if there exists $M > 0$, such that $|f(x) - f(y)| \leq M|x - y|^\alpha$, for all $x, y \in \Omega$. We shall denote $|f|_{Lip(\alpha)}$ by the infimum over all M satisfying the condition. Observe that we can replace the condition by

$$|\Delta_h(f, x)| \leq M|h|^\alpha, \quad \forall h \in \mathbb{R}^n \Rightarrow$$

$$\omega_1(f, t)_\infty \leq Mt^\alpha, \quad \forall t > 0 \Rightarrow$$

$$t^{-\alpha} \omega_1(f, t)_\infty \leq M, \quad \forall t > 0.$$

For $1 \leq p \leq \infty$, we define

$$|f|_{lip(\alpha, p)} := \sup_{t>0} t^{-\alpha} \omega_1(f, t)_p.$$

Example For $f(x) = x^\alpha$, $0 < \alpha \leq 1$, $f \in Lip(\alpha)$, $f \notin Lip(\beta)$, $\beta > \alpha$.

Proof

(i) Assume $f \in Lip(\beta)$, $\beta > \alpha$. Then for $0 < x \leq 1$,

$$x^\alpha - 0^\alpha = x^\alpha \leq M(x - 0)^\beta = Mx^\beta \Rightarrow x^{\alpha-\beta} \leq M \Rightarrow \text{contradiction}$$

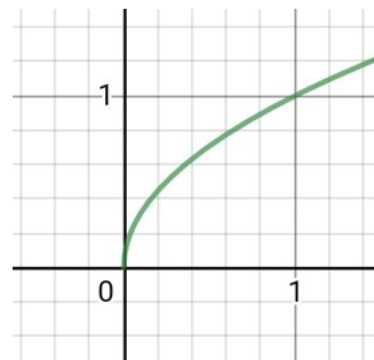
(ii) We use the inequality $(a+b)^\alpha \leq a^\alpha + b^\alpha$. Assume w.l.g $x \geq y$, we set $a = y, b = x - y$ and obtain

$$x^\alpha \leq y^\alpha + (x - y)^\alpha \Rightarrow x^\alpha - y^\alpha \leq (x - y)^\alpha, |f|_{Lip(\alpha)} = 1.$$

□

However, for any $0 < \alpha \leq 1$, $f(x) = x^\alpha \in Lip(1,1)$, because

$$\begin{aligned} \int_0^1 |f'(x)| dx &= 1 \Rightarrow f' \in L_1 \Rightarrow f \in W_1^1([0,1]) \\ &\Rightarrow \omega_1(f, t)_1 \leq t |f|_{1,1} = t, \quad \forall t > 0 \\ &\Rightarrow |f|_{Lip(1,1)} = \sup_{t>0} t^{-1} \omega_1(f, t)_1 = 1. \end{aligned}$$



Generalized Lip are a special case of Besov spaces. For any $\alpha > 0$, let $r := \lfloor \alpha \rfloor + 1$,

$$|f|_{B_{p,\infty}^\alpha} := \sup_{t>0} t^{-\alpha} \omega_r(f, t)_p.$$

Approximation using uniform piecewise constants (numerical integration)

The B-Spline of order one (degree zero, smoothness -1) $N_1(x) = \mathbf{1}_{[0,1]}(x)$.

Let $\Omega = \mathbb{R}$ or $\Omega = [a, b]$. We approximate from the space

$$S(N_1)^h := \left\{ \sum_{k \in \mathbb{Z}} c_k N_1(h^{-1}x - k) \right\} = \left\{ \sum_{k \in \mathbb{Z}} c_k \mathbf{1}_{[kh, (k+1)h]}(x) \right\}.$$

Theorem For $f \in W_p^1(\mathbb{R})$, $1 \leq p \leq \infty$,

$$E\left(f, S(N_1)^h\right)_{L_p(\mathbb{R})} := \inf_{g \in S(N_1)^h} \|f - g\|_{L_p(\mathbb{R})} \leq h |f|_{W_p^1(\mathbb{R})}.$$

Proof First assume $f \in C^1(\mathbb{R}) \cap W_p^1(\mathbb{R})$. Let's take the interval $[kh, (k+1)h]$. Then, for $p = \infty$

$$|f(x) - f(kh)| = \left| \int_{kh}^x f'(u) du \right| \leq h \max_{kh \leq u \leq (k+1)h} |f'(u)|.$$

So, select $c_k := f(kh)$ and you get the theorem for $p = \infty$. For $1 \leq p < \infty$ we do something similar

$$|f(x) - f(kh)|^p \leq \left(\int_{kh}^{(k+1)h} |f'(u)| du \right)^p, \quad x \in [kh, (k+1)h].$$

Then

$$\begin{aligned}
 \int_{kh}^{(k+1)h} |f(x) - f(kh)|^p dx &\leq h \left(\int_{kh}^{(k+1)h} |f'(u)| du \right)^p \\
 &\stackrel{\text{Holder}}{\leq} h \left(\|f'\|_{L_p([kh, (k+1)h])} \|1\|_{L_{p'}([kh, (k+1)h])} \right)^p \\
 &= hh^{p/p'} \|f'\|_{L_p([kh, (k+1)h])}^p \\
 &= h^p \|f'\|_{L_p([kh, (k+1)h])}^p.
 \end{aligned}
 \qquad
 \begin{aligned}
 1 + \frac{p}{p'} &= 1 + p \left(1 - \frac{1}{p} \right) \\
 &= 1 + p - 1 = p
 \end{aligned}$$

Therefore, with $g(x) := \sum_k f(kh) N_1(h^{-1}x - k)$, we get

$$\|f - g\|_p^p = \int_{-\infty}^{\infty} |f(x) - g(x)|^p dx = \sum_k \int_{kh}^{(k+1)h} |f(x) - f(kh)|^p dx \leq \sum_k h^p \|f'\|_{L_p([kh, (k+1)h])}^p = h^p \|f'\|_p^p.$$

Now assume $f \in W_p^1(\mathbb{R})$, $1 \leq p < \infty$. There exist sequences $\{f_k\}$, $f_k \in C^1(\mathbb{R}) \cap W_p^1(\mathbb{R})$, $\{g_k\}$, $g_k \in S(N_1)^h$, such that $\|f - f_k\|_{W_p^1(\mathbb{R})} \xrightarrow{k \rightarrow \infty} 0$ and $\|f_k - g_k\|_{L_p(\mathbb{R})} \leq h|f_k|_{W_p^1(\mathbb{R})}$. This gives

$$\begin{aligned}\|f - g_k\|_p &\leq \|f - f_k\|_p + \|f_k - g_k\|_p \\ &\leq \|f - f_k\|_p + h|f_k|_{1,p} \xrightarrow{k \rightarrow \infty} 0 + h|f|_{1,p}\end{aligned}$$

□

Linear approximation of Lip functions

Why linear? There is a ‘near best’ (possibly up to a constant) linear realization of the approximation.

Theorem: Let $f \in Lip(\alpha)$. Approximation with uniform piecewise constants gives

$$E_N(f)_{L_\infty([0,1])} := \inf_{\phi \in S(N_1)^{1/N}} \|f - \phi\|_\infty \leq CN^{-\alpha} |f|_{Lip(\alpha)}.$$

Inverse Theorem: Assume $E_N(f)_\infty \leq MN^{-\alpha}$, $N \geq 1$. Then, $f \in Lip(\alpha)$.

Example $E_N(x^\alpha) \sim N^{-\alpha}$, $0 < \alpha \leq 1$.

Non-linear approximation of Lip functions

$$\Sigma_N := \left\{ \sum_{j=0}^{N-1} c_j \mathbf{1}_{[t_j, t_{j+1})} : T = \{t_j\}, 0 = t_0 < t_1 < \dots < t_N = 1 \right\}, \quad \sigma_N(f)_p := \inf_{g \in \Sigma_N} \|f - g\|_p.$$

This is the theoretical model of a univariate decision tree!

$$Var(f) := \sup_T \left\{ \sum |f(t_{j+1}) - f(t_j)| \right\}.$$

If f' exists a.e., $Var(f) = \|f'\|_1$. Why?

$$\int_0^1 |f'(x)| dx = \lim_{h \rightarrow 0} \sum_k h \frac{|f((k+1)h) - f(kh)|}{h}.$$

Let's go back to the examples $f(x) = x^\alpha$. In our case $\|f'\|_1 = \int_0^1 f'(x) dx = f(1) - f(0) = 1$.

Now, create a partition where $Var_{[t_j, t_{j+1}]}(f) \leq \frac{Var(f)}{N}$. If a_j is the median value in $[t_j, t_{j+1}]$, then

$$|f(x) - a_j| \leq \frac{Var_{[t_j, t_{j+1}]}(f)}{2} \leq \frac{Var(f)}{2N}, \quad \forall x \in [t_j, t_{j+1}].$$

For $f(x) = x^\alpha$, $0 < \alpha \leq 1$, this gives a free knot spline $g \in \Sigma_N$ with

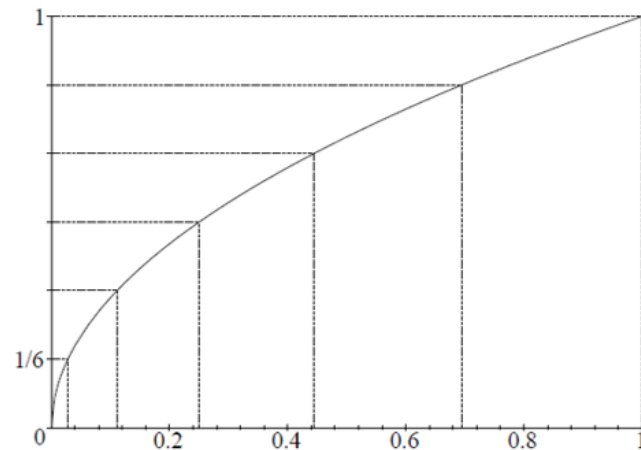
$$\|f - g\|_\infty \leq \frac{\text{Var}(f)}{2N} \leq \frac{1}{2N}.$$

To obtain an equidistant partition of the range, we choose

$$t_j = \left(\frac{j}{N}\right)^{1/\alpha}.$$

So, we see the advantage of nonlinear approximation for the family $f(x) = x^\alpha$, $0 < \alpha < 1$,

$$f \in \text{Lip}(\alpha, \infty) \Rightarrow E_N(f)_\infty \sim N^{-\alpha}, \quad f \in \text{Lip}(1, 1) \Rightarrow \sigma_N(f)_\infty \sim N^{-1}.$$



Besov Spaces

Let $\alpha > 0$, $0 < q, p \leq \infty$. Let $r \geq \lfloor \alpha \rfloor + 1$. The Besov space $B_q^\alpha(L_p(\Omega))$ is the collection of functions $f \in L_p(\Omega)$ for which

$$|f|_{B_q^\alpha(L_p(\Omega))} := \begin{cases} \left(\int_0^\infty \left[t^{-\alpha} \omega_r(f, t)_p \right]^q \frac{dt}{t} \right)^{1/q}, & 0 < q < \infty, \\ \sup_{t>0} t^{-\alpha} \omega_r(f, t)_p, & q = \infty. \end{cases}$$

is finite. The norm is

$$\|f\|_{B_q^\alpha(L_p(\Omega))} := \|f\|_{L_p(\Omega)} + |f|_{B_q^\alpha(L_p(\Omega))}.$$

Why are we asking for the condition $r \geq \lfloor \alpha \rfloor + 1$? Otherwise, the space is ‘trivial’

Theorem (univariate case) For $r < \alpha$, $1 \leq p \leq \infty$, we get that $B_q^\alpha(L_p(\Omega)) = \Pi_{r-1}$ if $\Omega = [a, b]$ and $B_q^\alpha(L_p(\Omega)) = \{0\}$ if $\Omega = \mathbb{R}$.

Theorem The space $B_q^\alpha(L_p(\Omega))$ does not depend on the choice of $r \geq \lfloor \alpha \rfloor + 1$ (application of the Marchaud inequality).

Theorem For a bounded domain we can equivalently integrate the semi-norm on $[0,1]$. That is,

$$|f|_{B_q^\alpha(L_p(\Omega))} \sim \begin{cases} \left(\int_0^1 \left[t^{-\alpha} \omega_r(f, t)_p \right]^q \frac{dt}{t} \right)^{1/q}, & 0 < q < \infty, \\ \sup_{0 < t \leq 1} t^{-\alpha} \omega_r(f, t)_p, & q = \infty. \end{cases}$$

Proof If Ω is bounded, then we have $\omega_r(f, t)_p \equiv \text{const}$ for $t \geq \text{diam}(\Omega)$. Therefore for $1/2 \leq t \leq \infty$,

$$\omega_r(f, 1/2)_p \leq \omega_r(f, t)_p \leq \omega_r(f, \text{diam}(\Omega))_p = \omega_r\left(f, \frac{2\text{diam}(\Omega)}{2}\right)_p \leq (1 + 2\text{diam}(\Omega))^r \omega_r(f, 1/2)_p.$$

This gives

$$\begin{aligned} \int_1^\infty \left[t^{-\alpha} \omega_r(f, t)_p \right]^q \frac{dt}{t} &\leq C \left(\omega_r(f, 1/2)_p \right)^q \int_1^\infty t^{-q\alpha-1} dt \\ &\leq C \left(\omega_r(f, 1/2)_p \right)^q \\ &\leq C(\alpha, q, \Omega) \int_{1/2}^1 \left[t^{-\alpha} \omega_r(f, t)_p \right]^q \frac{dt}{t} \end{aligned}$$

Lemma For any domain taking the integral over $[0,1]$ gives a quasi-norm equivalent to $\|f\|_{B_q^\alpha(L_p(\Omega))}$

Proof We replace the integral over $[1,\infty]$ by

$$\begin{aligned} \int_1^\infty \left[t^{-\alpha} \omega_r(f,t)_p \right]^q \frac{dt}{t} &\leq C \|f\|_p^q \int_1^\infty t^{-q\alpha-1} dt \\ &= C(\alpha, q) \|f\|_p^q. \end{aligned}$$

Therefore

$$\|f\|_{B_q^\alpha(L_p(\Omega))} \sim \|f\|_p + \left(\int_0^1 \left[t^{-\alpha} \omega_r(f,t)_p \right]^q \frac{dt}{t} \right)^{1/q}.$$

Theorem $B_{q_1}^{\alpha_1}(L_p) \subseteq B_{q_2}^{\alpha_2}(L_p)$ if $\alpha_2 < \alpha_1$.

Proof ($q_1 = q_2$) We may use $r_1 = \lfloor \alpha_1 \rfloor + 1 \geq \lfloor \alpha_2 \rfloor + 1 = r_2$ to equivalently define $B_{q_2}^{\alpha_2}(L_p)$

For $0 < t \leq 1$, $t^{-\alpha_2} \leq t^{-\alpha_1}$. So,

$$\begin{aligned} \|f\|_{B_q^{\alpha_2}(L_p)} &\leq C \left(\|f\|_p + \left(\int_0^1 \left[t^{-\alpha_2} \omega_{r_1}(f, t)_p \right]^q \frac{dt}{t} \right)^{1/q} \right) \\ &\leq C \left(\|f\|_p + \left(\int_0^1 \left[t^{-\alpha_1} \omega_{r_1}(f, t)_p \right]^q \frac{dt}{t} \right)^{1/q} \right) \\ &\leq C \|f\|_{B_q^{\alpha_1}(L_p)} \end{aligned}$$

Theorem $W_p^m \subseteq B_q^\alpha(L_p)$, $\forall \alpha < m$, $1 \leq p \leq \infty$, $0 < q \leq \infty$.

Proof Let $g \in W_p^m(\Omega)$. This implies $g \in L_p(\Omega)$. We have that $r := \lfloor \alpha \rfloor + 1 \leq m$. It is sufficient to take the integral over $[0, 1]$.

$$\begin{aligned} \int_0^1 \left[t^{-\alpha} \omega_r(g, t)_p \right]^q \frac{dt}{t} &\leq C \int_0^1 \left[t^{-\alpha} t^r |g|_{r,p} \right]^q \frac{dt}{t} \\ &\leq C |g|_{r,p}^q \int_0^1 t^{(r-\alpha)q-1} dt \\ &\leq C |g|_{r,p}^q. \end{aligned}$$

Discretization of the Besov semi-norm

Theorem One has the following equivalent form of the Besov semi-norm

$$|f|_{B_q^\alpha(L_p(\Omega))} \sim \begin{cases} \left(\sum_{k=-\infty}^{\infty} \left[2^{k\alpha} \omega_r(f, 2^{-k})_p \right]^q \right)^{1/q}, & 0 < q < \infty. \\ \sup_{k \in \mathbb{Z}} 2^{k\alpha} \omega_r(f, 2^{-k})_p, & q = \infty. \end{cases}$$

Proof Define $\varphi(t) := t^{-\alpha} \omega_r(f, t)_p$. Then we claim that for $t \in [2^{-k-1}, 2^{-k}]$, $k \in \mathbb{Z}$, we have

$$2^{-r} \varphi(2^{-k}) \leq \varphi(t) \leq 2^{\alpha} \varphi(2^{-k}).$$

To see that, we use the following properties:

- (i) $\omega_r(f, t)_p$ is non-decreasing
- (ii) For $N \in \mathbb{N}$, $1 \leq p \leq \infty$, $\omega_r(f, Nt)_p \leq N^r \omega_r(f, t)_p$

The left-hand side

$$\begin{aligned} 2^{-r} \varphi(2^{-k}) &= 2^{k\alpha-r} \omega_r(f, 2^{-k})_p = 2^{k\alpha-r} \omega_r(f, 2^{-(k-1)})_p \\ &\stackrel{(ii)}{\leq} 2^{k\alpha-r} 2^r \omega_r(f, 2^{-(k-1)})_p \stackrel{(i)}{\leq} 2^{k\alpha} \omega_r(f, t)_p \leq t^{-\alpha} \omega_r(f, t)_p \end{aligned}$$

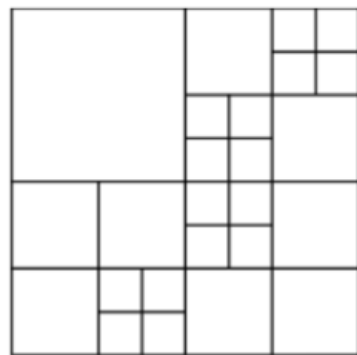
The right-hand side

$$t^{-\alpha} \omega_r(f, t)_p \stackrel{(i)}{\leq} t^{-\alpha} \omega_r(f, 2^{-k})_p \leq 2^{(k+1)\alpha} \omega_r(f, 2^{-k})_p \leq 2^\alpha \varphi(2^{-k})$$

This gives us for $0 < q < \infty$, $k \in \mathbb{Z}$

$$\int_{2^{-(k-1)}}^{2^{-k}} \varphi(t)^q \frac{dt}{t} \sim \varphi(2^{-k})^q \int_{2^{-(k-1)}}^{2^{-k}} \frac{dt}{t} \sim \varphi(2^{-k})^q \Rightarrow \int_{2^{-(k-1)}}^{2^{-k}} \left(t^{-\alpha} \omega_r(f, t)_p \right)^q \frac{dt}{t} \sim \left[2^{k\alpha} \omega_r(f, 2^{-k})_p \right]^q.$$

Discretization over cubes



Definition [Dyadic cubes] Let $D := \{D_k : k \in \mathbb{Z}\}$

$$D_k := \left\{ Q = 2^{-kn} [m_1, m_1 + 1] \times \cdots \times [m_n, m_n + 1] : m \in \mathbb{Z}^n \right\}.$$

Observe that $Q \in D_k \Rightarrow |Q| = 2^{-kn}$.

For nonlinear/adaptive/sparse approximation in $L_p(\Omega)$, $\Omega \subseteq \mathbb{R}^n$, it is useful to use the special cases of Besov spaces

$$B_\tau^\alpha := B_\tau^\alpha(L_\tau(\Omega)), \quad \frac{1}{\tau} = \frac{\alpha}{n} + \frac{1}{p}.$$

Theorem $\Omega = \mathbb{R}^n$. We have the equivalence

$$|f|_{B_\tau^\alpha} \sim \left(\sum_{k \in \mathbb{Z}} \left(2^{k\alpha} \omega_r(f, 2^{-k})_\tau \right)^\tau \right)^{1/\tau} \sim \left(\sum_{Q \in D} \left(|Q|^{-\alpha/n} \omega_r(f, Q)_\tau \right)^\tau \right)^{1/\tau},$$

$$\omega_r(f, Q)_\tau := \sup_{h \in \mathbb{R}^n} \left\| \Delta_h^r(f, Q, \cdot) \right\|_{L_\tau(Q)}.$$

The following theorem generalizes what we showed for the univariate case

Theorem Let $f(x) = \mathbf{1}_{\tilde{\Omega}}(x)$, $\tilde{\Omega} \subset [0,1]^n$, a domain with smooth boundary. Then $f \in B_\tau^\alpha$, $\alpha < 1/\tau$.

Proof For $\Omega = [0,1]^n$, with $l(Q)$ denoting the level of the cube Q , we may take the sum over $k \geq 0$

$$|f|_{B_\tau^\alpha} \sim \left(\sum_{k=0}^{\infty} \left(2^{k\alpha} \omega_r(f, 2^{-k})_\tau \right)^\tau \right)^{1/\tau} \sim \left(\sum_{Q \in D, l(Q) \geq 0} \left(|Q|^{-\alpha/n} \omega_r(f, Q)_\tau \right)^\tau \right)^{1/\tau}.$$

For any Q , we have that $\omega_r(f, Q)_\tau = 0$, if $\partial\tilde{\Omega} \cap Q = \emptyset$. Otherwise, if $l(Q) = k$,

$$\omega_r(f, Q)_\tau \leq C \|f\|_{L_\tau(Q)} \leq C \left(\int_Q 1^\tau \right)^{1/\tau} = C |Q|^{1/\tau} = C 2^{-kn/\tau}.$$

Therefore,

$$\begin{aligned} |f|_{B_\tau^\alpha}^\tau &\leq C \sum_{l(Q) \geq 0} \left(|Q|^{-\alpha/n} \omega_r(f, Q)_\tau \right)^\tau \\ &\leq C \sum_{k=0}^{\infty} \left(2^{k\alpha} 2^{-kn/\tau} \right)^\tau \# \{Q : l(Q) = k, Q \cap \partial\tilde{\Omega} \neq \emptyset\} \\ &= C \sum_{k=0}^{\infty} 2^{k(\alpha\tau - n)} \# \{Q : l(Q) = k, Q \cap \partial\tilde{\Omega} \neq \emptyset\} \end{aligned}$$

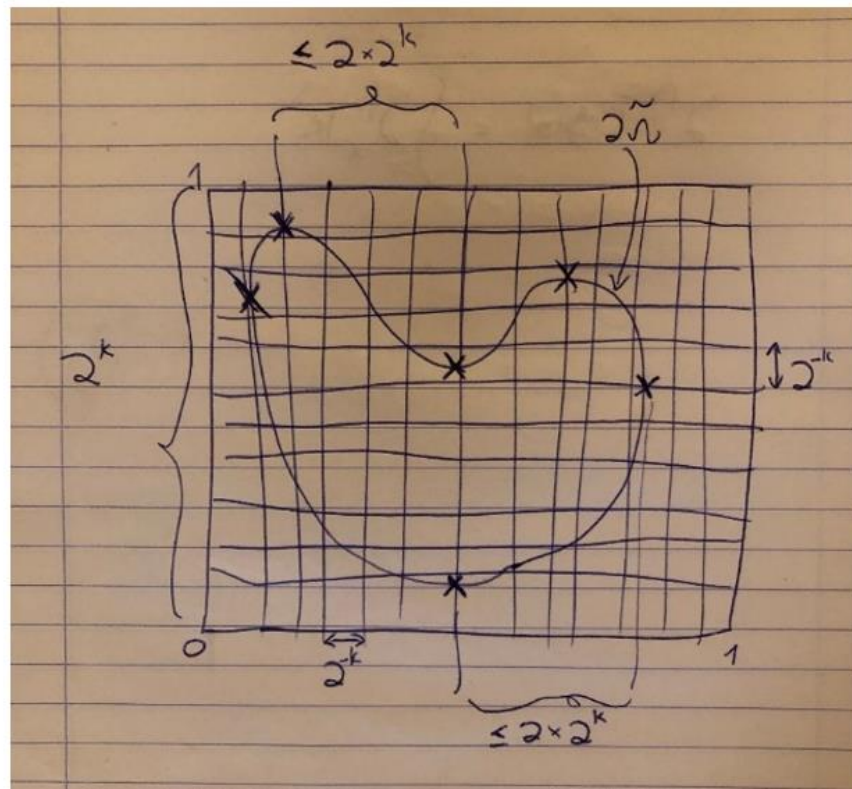
We argue that

$$\#\{Q : l(Q) = k, Q \cap \partial\tilde{\Omega} \neq \emptyset\} \leq c(\tilde{\Omega}) 2^{k(n-1)}. \quad (*)$$

This implies that if $\alpha < 1/\tau$

$$|f|_{B_r^\alpha}^\tau \leq C \sum_{k=0}^{\infty} 2^{k(\alpha\tau-n)} 2^{k(n-1)} = C \sum_{k=0}^{\infty} 2^{k(\alpha\tau-1)} < \infty.$$

Let's get back to the estimate (*). Let us show a picture argument for $\tilde{\Omega} \subset [0,1]^2$. There is a finite number of points where the gradient of the boundary of the domain is aligned with one of the main axes. Between these points, the boundary segments are monotone in x_1 and x_2 , and therefore can only intersect at most 2×2^k dyadic cubes.



Approximation Spaces

Let $\Phi := \{\Phi_N\}_{N \geq 0}$, each Φ_N is a set of functions in a (quasi) Banach space X , satisfying:

- (i) $0 \in \Phi_N$, $\Phi_0 := \{0\}$,
- (ii) $\Phi_N \subset \Phi_{N+1}$,
- (iii) $a\Phi_N = \Phi_N$, $\forall a \neq 0$,
- (iv) $\Phi_N + \Phi_N \subset \Phi_{cN}$, for some constant $c(\Phi)$,
- (v) $\overline{\bigcup_N \Phi_N} = X$,
- (vi) Each $f \in X$ has a near best approximation from Φ_N . That is, there exists a constant $C(\Phi)$, such that for any N , one has $\varphi_N \in \Phi_N$,

$$\|f - \varphi_N\|_X \leq C E_N(f)_X, \quad E_N(f)_X := \inf_{\varphi \in \Phi_N} \|f - \varphi\|_X.$$

Examples for Φ_N

Linear

- Trigonometric polynomials of degree $\leq N$, $X = L_p \left([-\pi, \pi]^n \right)$.
- Algebraic polynomials of degree $\leq N$, $X = L_p [-1, 1]$.
- Uniform dyadic knot piecewise polynomials over pieces of length 2^{-N} , of fixed order r , $X = L_p [0, 1]$.
- Shift invariant refinable spaces $\Phi_N := S(\phi)^{2^{-N}}$, $S(\phi) \subset S(\phi)^{1/2}$, $X = L_p(\mathbb{R}^n)$.

Nonlinear/Adaptive

- Rational functions of degree $\leq N$, $X = L_p [-1, 1]$,
- Free knot piecewise polynomials of fixed order r over N non-uniform intervals, $X = L_p [0, 1]$.
- N-term wavelets $\Phi_N = \Sigma_N := \left\{ \sum_{\#I \leq N} c_I \psi_I \right\}$, $X = L_2(\mathbb{R}^n)$.

Def Approximation spaces for $\alpha > 0$, $0 < q \leq \infty$, $f \in X$,

$$|f|_{A_q^\alpha} := \begin{cases} \left(\sum_{N=1}^{\infty} \left[N^\alpha E_N(f) \right]^q \frac{1}{N} \right)^{1/q}, & 0 < q < \infty, \\ \sup_{N \geq 1} N^\alpha E_N(f), & q = \infty. \end{cases}$$

$$\|f\|_{A_q^\alpha} := \|f\|_X + |f|_{A_q^\alpha}.$$

One can show

$$|f|_{A_q^\alpha} \sim \begin{cases} \left(\sum_{m=0}^{\infty} \left[2^{m\alpha} E_{2^m}(f) \right]^q \right)^{1/q}, & 0 < q < \infty, \\ \sup_{m \geq 0} 2^{m\alpha} E_{2^m}(f), & q = \infty. \end{cases}$$

Goal: Fully characterize approximation spaces by smoothness spaces (iff)

Characterization of approximation spaces

1. Trigonometric polynomials

$X = L_p[-\pi, \pi]$, $1 \leq p \leq \infty$, Φ_N trigonometric polynomials of degree N

$$A_q^\alpha(L_p) \sim B_q^\alpha(L_p).$$

2. Dyadic univariate piecewise polynomials

$X = L_p[0, 1]$, Φ_N piecewise polynomials of degree $d \geq 0$, over uniform subdivision of 2^N intervals.

For $1 \leq p \leq \infty$, $\alpha < r - 1 + 1/p$, $0 < q \leq \infty$,

$$A_q^\alpha(L_p) \sim B_q^\alpha(L_p).$$

3. Adaptive non-uniform univariate piecewise polynomials

$$A_\tau^\alpha(L_p) \sim B_\tau^\alpha, \quad \frac{1}{\tau} = \alpha + \frac{1}{p}.$$

The mathematical foundations of decision trees

For the theory of geometric approximation in higher dimensions we generalize to anisotropic partitions of trees over $[0,1]^n$ (replacing dyadic cubes!)

$$|f|_{B_r^\alpha(\mathcal{T})} := \left(\sum_{\Omega \in \mathcal{T}} \left(|\Omega|^{-\alpha} \omega_r(f, \Omega)_\tau \right)^\tau \right)^{1/\tau}$$

