Mathematical foundations of Machine Learning 2024 – lesson 2

Shai Dekel

TEL AVIV UNIVERSITY

Banach Spaces

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

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

i. $\|f\| \geq 0$ and $\|f\| = 0 \Leftrightarrow f = 0$.

ii. $f \in B$ only if $\|f\|_B < \infty$.

iii. $\|\alpha f\| = |\alpha| \|f\|$, $\forall \alpha \in F, \forall f \in B$.

iv. Triangle inequality $\|f + g\| \leq \|f\| + \|g\|$.

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

Measure

In this course we only use the standard Lebesgue measure $\leftrightarrow$ the volume of a (measurable) set.

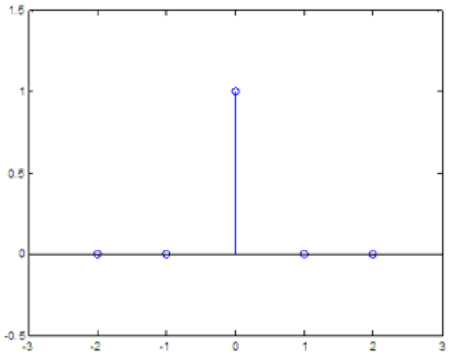
Example: $\Omega = [0,2]^n \subset \mathbb{R}^n$, $\mu(\Omega) = |\Omega| = 2^n$.

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

Lp Spaces

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

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



$$\underset{x}{ess\sup} |f(x)| := \sup_{A>0} \left\{ A > 0 : \ \left| \{ x : |f(x)| \ge A \} \right| > 0 \right\}.$$

For $1 \le p \le \infty$, $L_p(\Omega)$ are Banach spaces.

For $0 < p < 1$, $L_p(\Omega)$ are Quasi-Banach spaces (quasi-triangle inequality holds)

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

Theorem [Hölder] $1 \le p \le \infty$, $f \in L_p, g \in L_{p'}$

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

Lemma Young's inequality for products,

$$ab \le \frac{a^p}{p} + \frac{b^{p'}}{p'}, \qquad \frac{1}{p} + \frac{1}{p'} = 1, \ \forall a,b \ge 0.$$

Proof of lemma The logarithmic function is concave. Therefore

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

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

Proof of theorem If $p=\infty$

$$\int_\Omega |fg| \le \|f\|_\infty \int_\Omega |g| \le \|f\|_\infty \|g\|_1 .$$

The proof is similar for $p=1$. So, assume now $1<p<\infty$ and $\|f\|_p = \|g\|_{p'} = 1$.
Integrating pointwise and applying Young's inequality almost everywhere, gives

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

Now assuming $f, g \ne 0$ (else, we are done)

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

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

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

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

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

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

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

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

Proof The quasi-triangle inequality (ii) is derived from (i). Observe first

$$\left.\begin{array}{c} 1 < r := \dfrac{1}{p} < \infty \\ \dfrac{1}{r} + \dfrac{1}{r'} = 1 \end{array}\right\} \Rightarrow r' = \frac{1}{1-p}$$

Then

$$\left\|\sum_{k=1}^N f_k\right\|_p \underset{(i)}{\leq} \left(\sum_{k=1}^N \|f_k\|_p^p\right)^{1/p} = \left(\sum_{k=1}^N 1 \cdot \|f_k\|_p^p\right)^{1/p} \underset{\substack{\text{Discrete} \\ \text{Holder}}}{\leq} \left(\sum_{k=1}^N 1^{\frac{1}{1-p}}\right)^{(1-p)1/p} \left(\sum_{k=1}^N \|f_k\|_p\right) = N^{1/p-1}\sum_{k=1}^N \|f_k\|_p$$

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 Observe that it is sufficient to prove $(a_1 + a_2)^p \leq a_1^p + a_2^p$ and then apply induction.

To prove the inequality use $h(t) := t^p + 1 - (t+1)^p$ for $t \geq 0$. $h(0) = 0$ 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 $x \in \Omega$ and then Tonelli's theorem for the exchange of integration and sum

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

Multivariate algebraic polynomials

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

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

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

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

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

Spaces of smooth functions

Multivariate derivatives: A partial derivative of order r

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

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

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

The ***semi-norm***

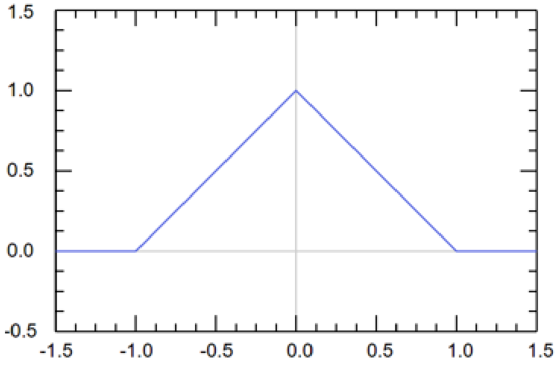
$$|f|_{C^r(\Omega)} := \sum_{|\alpha| = r} \|D^\alpha f\|_\infty$$

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 $W_p^r(\Omega)$, $1 \le p \le \infty$

Def For $1 \le p < \infty$, the closure of the compactly supported smooth functions $C_0^r(\Omega)$ with respect to the norm $\sum_{|\alpha| \le r} \|D^\alpha f\|_{L_p(\Omega)}$. For $p = \infty$, we take $W_\infty^r := C^r$.

$$H(x) := \begin{cases} x+1, & -1 \le x < 0, \\ 1-x, & 0 \le x \le 1, \\ 0, & else. \end{cases},$$



So, in this sense $H \in W_p^1(\mathbb{R})$, $1 \le p < \infty$.

The Sobolev norm and semi-norm.

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

Theorem W_p^r 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

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

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 $\|\Delta_h^r\|_{L_p \to L_p} \leq 2^r$. Assume $\Omega = \mathbb{R}^n$, then

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

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 $\Omega = [-1,1]$. $f(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 \le x \end{cases}$

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

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

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

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

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

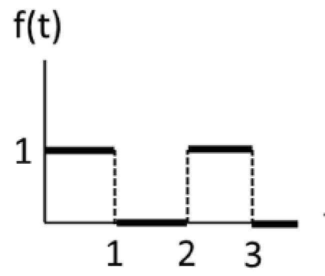
We get $\omega_2(f,t)_{L_p([-1,1])} = (2t)^{1/p}$

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

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

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

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 $0 < \alpha < 1$, $|f|_{B^{\alpha}_{\tau,\infty}} \sim (2M)^{1/\tau}$.

- 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.
- Deep Learning? For $M = 2^j$, the function can be approximated by a neural network with $\sim j$ blocks,
- After each k-th block (2 layers) the function f_k has 2^{j-k} 'steps' with $|f_k|_{B^{\alpha}_{\tau,\infty}} \sim 2^{(j-k)/\tau}$.

Properties

1. $\omega_r(f,t)_p \le 2^r \|f\|_{L_p(\Omega)}$, $1 \le p \le \infty$.

2. $\omega_r(f,t)_p$ is non-decreasing in t

3. For $1 \le p \le \infty$ the ***sub-linearity*** property

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

gives

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

4. For $N \geq 1$, $\omega_r(f, Nt)_p \leq N^r \omega_r(f,t)_p$, $1 \leq p \leq \infty$. 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 $h \in \mathbb{R}^n$, $|h| \leq t$

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

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

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

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

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

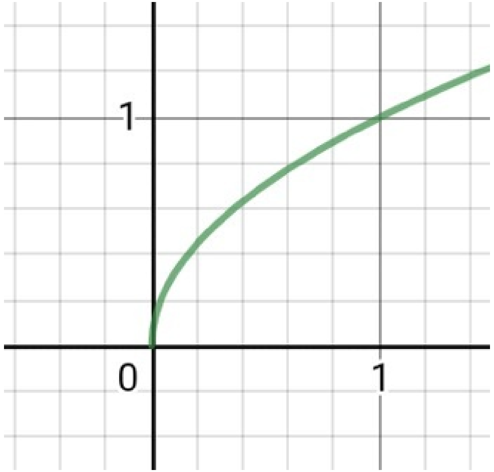
Theorem [connection between Sobolev and modulus] For $g \in W_p^r(\Omega)$, $1 \le p \le \infty$, we have that

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

Lip spaces

Def For a domain $\Omega \subset \mathbb{R}^n$ and $0 < \alpha \le 1$, we shall say that $f \in Lip(\alpha) = Lip(\alpha, \infty)$, if there exists $M > 0$, such that $|f(x) - f(y)| \le M|x - y|^\alpha$, 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

$$|\Delta_h(f,x)| \le M|h|^\alpha, \quad \forall h \in \mathbb{R}^n \Rightarrow \omega_1(f,t)_\infty \le Mt^\alpha \Rightarrow t^{-\alpha}\omega_1(f,t)_\infty \le M .$$



For $1 \le p \le \infty$, we can generalize by

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

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

Proof

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

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

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

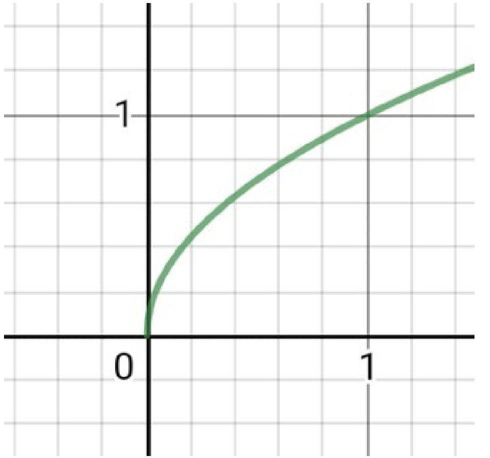
□

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

$$\int_0^1 |f'(x)| dx = 1 \Rightarrow f' \in L_1$$

$$\Rightarrow \omega_1(f,t)_1 \le t|f|_1 = t\|f'\|_1 = t$$

$$\Rightarrow |f|_{Lip(1,1)} = \sup_{t>0} t^{-1}\omega_1(f,t)_1 \le 1.$$



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

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

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

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

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

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

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

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

First glimpse to Adaptive / Nonlinear / Sparse approximation

Approximation using free-knot splines / non-uniform piecewise constants in $L_\infty([0,1])$

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

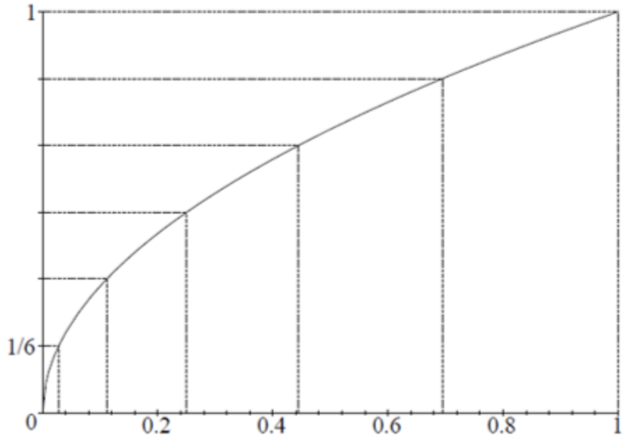
Assume f' exists a.e., Now, create a partition where

$$\int_{t_j}^{t_{j+1}} |f'| \le \frac{\|f'\|_1}{N} .$$

For the example $f(x) = x^\alpha,\ 0 < \alpha < 1$,

$$\int_0^1 f' = 1 \Rightarrow \int_{t_j}^{t_{j+1}} f'(u)\,du = \frac{1}{N}, \quad 0 \le j \le N-1 .$$

This equidistant partition of the range is achieved by choosing $t_j = \left(\frac{j}{N}\right)^{1/\alpha}$.



If a_j is the median value in $\left[t_j, t_{j+1}\right]$, then

$$\left|f(x)-a_j\right| \le \frac{\int_{t_j}^{t_{j+1}}\left|f'\right|}{2} \le \frac{\left\|f'\right\|_1}{2N}, \ \forall x \in \left[t_j, t_{j+1}\right].$$

This gives a free knot piecewise constant $g \in \Sigma_N$ with

$$\left\|f-g\right\|_\infty \le \frac{\left\|f'\right\|_1}{2N} .$$

Recall that earlier on, we promised that 'integration' of differences will be meaningful. Indeed, for the family $f(x)=x^\alpha$, $0<\alpha<1$, we see the smoothness

$$\left|f\right|_{Lip(1,1)} = \sup_{t>0} t^{-1}\omega_1(f,t)_1 \le \sup_{t>0} t^{-1} t \left\|f'\right\|_1 = 1,$$

comes into play to show the advantage of nonlinear approximation over linear approximation

$$f \in Lip(\alpha,\infty), \qquad f \in Lip(1,1),$$

$$E_N(f)_\infty \sim N^{-\alpha}, \qquad \sigma_N(f)_\infty \sim N^{-1}.$$

Besov Spaces

Let $\alpha > 0$, $0 < q, p \le \infty$. Let $r \ge \lfloor \alpha \rfloor + 1$. The Besov space $B_q^\alpha\left(L_p(\Omega)\right)$ 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 $r \ge \lfloor \alpha \rfloor + 1$? Otherwise, the space is 'trivial'

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

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

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

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

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,

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

$$\leq C\left(\|f\|_p + \left(\int_0^1 \left[t^{-\alpha_1}\omega_{r_1}(f,t)_p\right]^q \frac{dt}{t}\right)^{1/q}\right)$$

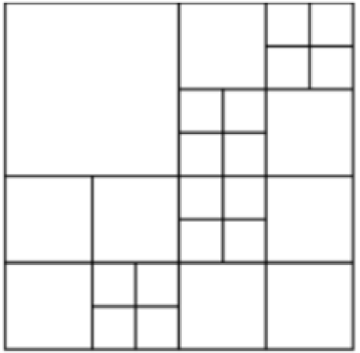
$$\leq C\|f\|_{B_q^{\alpha_1}(L_p)}$$

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

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

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

Discretization over cubes



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

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

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

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

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 $k\geq 0$

$$\int_0^1\left[t^{-\alpha}\omega_r(f,t)_p\right]^\tau\frac{dt}{t}\sim\sum_{Q\in D,l(Q)\geq 0}\left(|Q|^{-\alpha/n}\omega_r(f,Q)_\tau\right)^\tau.$$

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

$$\omega_r(f,Q)_\tau\leq C\|f\|_{L_\tau(Q)}\leq C\left(\int_Q 1^\tau\right)^{1/\tau}=C|Q|^{1/\tau}=C2^{-kn/\tau}.$$

Therefore,

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

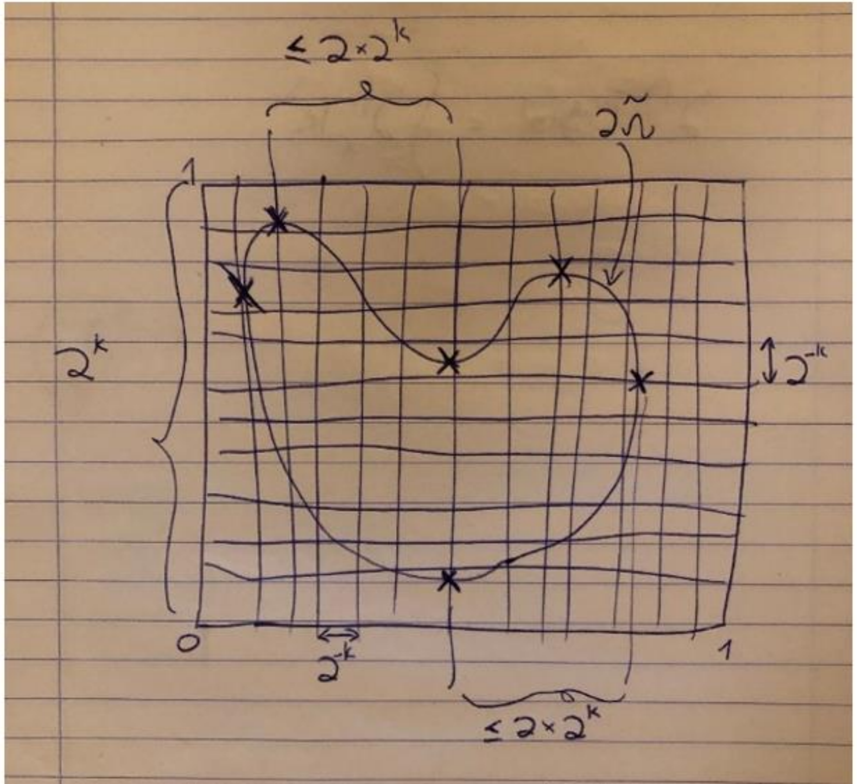
We argue that

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

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

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

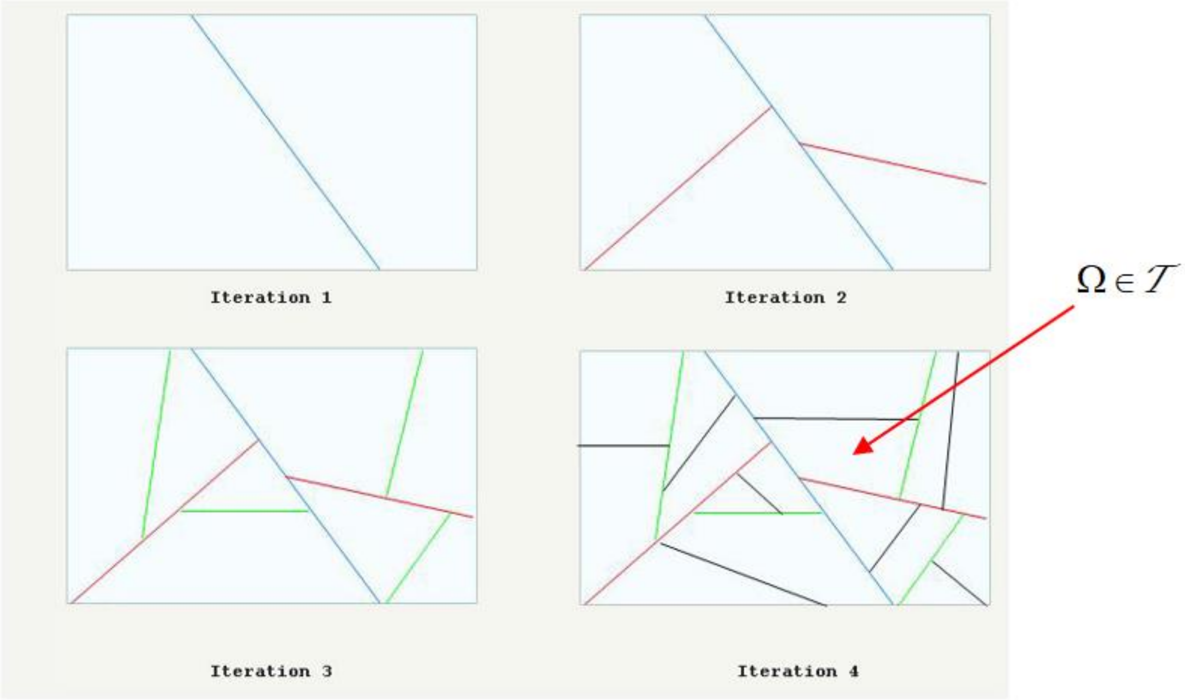
Let's get back to the estimate (*). Let use show a picture argument for $\tilde{\Omega} \subset [0,1]^2$. There is a finite number of points where the gradient of the boundary of the domain is aligned with one of the main axes. Between these points, the boundary segments are monotone in x_1 and x_2,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



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!)

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



Approximation Spaces

Let $\Phi := \{\Phi_N\}_{N \ge 0}$, 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 $\varphi_N \in \Phi_N$,

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

Examples for Φ_N

<u>Linear</u>

- 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 $\Phi_N := S(\phi)^{2^{-N}}$, $S(\phi) \subset S(\phi)^{1/2}$, $X = L_p(\mathbb{R}^n)$.

<u>Nonlinear/Adaptive</u>

- 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 $\alpha > 0$, $0 < q \le \infty$, $f \in X$,

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

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

One can show

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

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

Characterization of approximation spaces

1. Trigonometric polynomials

$X = L_p[-\pi,\pi]$, $1 \le p \le \infty$, 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 $d \ge 0$, over uniform subdivision of 2^N intervals.

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

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

3. Adaptive non-uniform univariate piecewise polynomials

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