Modulus of smoothness

Def The ***difference operator*** Δ_h^r. For $h \in \mathbb{R}^d$ we define $\Delta_h(f,x) = f(x+h) - f(x)$. For general $r \ge 1$ 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 in fact 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 $L_p(\Omega)$, $1 \le p \le \infty$, we have that $\left\|\Delta_h^r\right\|_{L_p \to L_p} \le 2^r$. Assume $\Omega = \mathbb{R}^n$, then

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

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

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

For $r = 1$ the modulus of smoothness is called the ***modulus of continuity***.

Properties

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

$$\left|\Delta_h^r(f+g,x)\right| \le \left|\Delta_h^r(f,x)\right| + \left|\Delta_h^r(g,x)\right|,$$

gives

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

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

$$\begin{aligned}\left\|\Delta_{Nh}^r(f,\cdot)\right\|_p &\le \sum_{k_1=0}^{N-1}\cdots\sum_{k_r=0}^{N-1}\left\|\Delta_h^r(f,\cdot+k_1h+\cdots+k_rh)\right\|_p \\ &= \sum_{k_1=0}^{N-1}\cdots\sum_{k_r=0}^{N-1}\left\|\Delta_h^r(f,\cdot)\right\|_p \le N^r\omega_r(f,t)_p .\end{aligned}$$

Taking supremum over all $h \in \mathbb{R}^n$, $|h| \le t$, gives $\omega_r(f,Nt)_p \le N^r \omega_r(f,t)_p$.

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

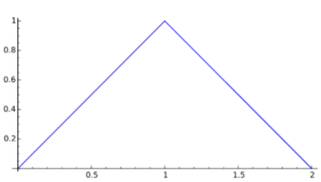
Theorem [connection between Sobolev and modulus] For $g \in W_p^r(\Omega)$, $1 \le p \le \infty$, we have that

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

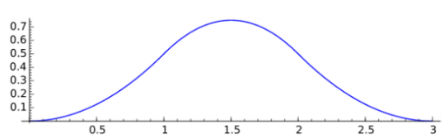
B-splines

- Definition by convolution $N_1 = \mathbf{1}_{[0,1]^n}$. In general, $N_r := N_{r-1} * N_1 = \int_{\mathbb{R}^n} N_{m-1}(x-t)N_1(t)\,dt$.

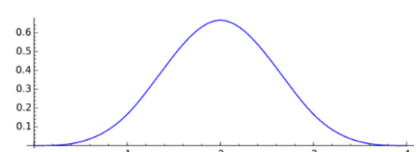
$$N_2(x) = N_1 * N_1(x) = \int_0^1 1_{[0,1]}(x-t)\,dt = \int_{\max(x-1,0)}^{\min(x,1)} dt = \begin{cases} 0 & x<0 \\ x & 0 \le x < 1 \\ 2-x & 1 \le x \le 2 \\ 0 & x > 2 \end{cases}$$



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}.
- $\int_{\mathbb{R}^n} N_r(x)\,dx = 1$
- Tensor-product in multivariate case

Let's see how we get the property of $\int_{\mathbb{R}^n} N_r(x)\,dx = 1$. Let $f,g \in L_1(\mathbb{R}^n)$ then

$$\int_{\mathbb{R}^n} f * g(x)\,dx = \int_{\mathbb{R}^n} f \int_{\mathbb{R}^n} g$$

Proof of Theorem ($\Omega = \mathbb{R}$). Here, we use the fact that for $h>0$, $\Delta_{-h}^r(f,x) = \Delta_h^r(f,x-rh)$. So W.L.G, for any $0 \le t$, we can work with $0 < h \le t$.

Define $N_r(x,h) := h^{-1} N_r(x/h)$, $h>0$. Let $g \in C^1(\mathbb{R})$. Then

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

More generally, $g \in C^r(\mathbb{R})$

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

To see that we apply induction

$$\begin{aligned} h^{-r}\Delta_h^r(g,x) &= h^{-1}h^{-(r-1)}\left(\Delta_h^{r-1}(g,x+h) - \Delta_h^{r-1}(g,x)\right) \\ &= h^{-1}\left(\int_{\mathbb{R}} g^{(r-1)}(x+h+u) N_{r-1}(u,h)\,du - \int_{\mathbb{R}} g^{(r-1)}(x+u) N_{r-1}(u,h)\,du\right) \\ &= h^{-1}\int_x^{x+h}\int_{-\infty}^{\infty} g^{(r)}(v+u) N_{r-1}(u,h)\,du dv \\ &= \int_{-\infty}^{\infty} N_{r-1}(u,h)\left(h^{-1}\int_x^{x+h} g^{(r)}(v+u) dv\right) du \\ &= \int_{-\infty}^{\infty} N_{r-1}(u,h)\left(\int_{-\infty}^{\infty} g^{(r)}(v+u) N_1(v-x,h) dv\right) du \\ &\overset{v+u=x+y}{=} \int_{-\infty}^{\infty} N_{r-1}(u,h)\left(\int_{-\infty}^{\infty} g^{(r)}(x+y) N_1(y-u,h)\,dy\right) du \\ &= \int_{-\infty}^{\infty} g^{(r)}(x+y)\left(\int_{-\infty}^{\infty} N_{r-1}(u,h) N_1(y-u,h)\,du\right) dy \\ &= \int_{-\infty}^{\infty} g^{(r)}(x+y)\,\mathrm{N}_r(y,h)\,\mathrm{dy} \end{aligned}$$

Now, let's see the proof for $p=1$. Let $|h| \le t$

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

For general $1 \le p < \infty$ we need Minkowski's inequality (**assignment**). It says that for measurable non-negative functions φ, ρ

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

Or written differently

$$\left\|\int_B \varphi(y)\rho(\cdot,y)dy\right\|_{L_p(A)} \le \int_B \varphi(y)\left\|\rho(\cdot,y)\right\|_{L_p(A)} dy$$

Using it we have

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

For a general function $g \in W_p^r(\mathbb{R})$ we use a density argument.

♦

Corollary From the theorem we get that $\Delta_h^r(P,x) = 0$ for any $P \in \Pi_{r-1}$ and therefore $\omega_r(P,t)_p = 0$.

Marchaud inequalities

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

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

The direct inverse cannot be true. If we take $\Omega = [a,b]$ and a polynomial $P \in \Pi_{r-1}$, then for any $h > 0$,

$$\Delta_h^r(P,x) = h^r \int_{-\infty}^{\infty} P^{(r)}(x+y) \mathrm{N}_r(y,h) \mathrm{d}y = 0.$$

So $\omega_r(P,t)_p = 0$, but we don't necessarily have $\omega_k(P,t)_p = 0$ for $0 \le k < r$.

Theorem $\Omega = \mathbb{R}$. For any $1 \le k < r$

$$\omega_k\left(f,t\right)_p \le ct^k \int_t^{\infty} \frac{\omega_r\left(f,s\right)_p}{s^{k+1}}\,ds\,, \qquad t>0\,.$$

The K-functional

For two Banach spaces $X_1 \subset X_0$ the corresponding K-functional

$$K\left(f,t,X_0,X_1\right) := \inf_{f=f_0+f_1} \left\|f_0\right\|_{X_0} + t\left|f_1\right|_{X_1}$$

$$K\left(f,t,L_p\left(\Omega\right),W_p^r\left(\Omega\right)\right) := \inf_{g\in W_p^r(\Omega)} \left\|f-g\right\|_{L_p(\Omega)} + t\left|g\right|_{W_p^r(\Omega)}\,, \qquad 1\le p\le\infty\,.$$

Theorem [Equivalence of K-functional and modulus] For $\Omega \subseteq \mathbb{R}^n$ There exist $C_1, C_2 > 0$, such that for any $t>0$

$$C_1 K_r\left(f,t^r\right)_p \le \omega_r\left(f,t\right)_p \le C_2 K_r\left(f,t^r\right)_p\,. \tag{1.1}$$

It is easy to show that C_2 depends only on r, but the constant C_1 further depends on the geometry of Ω.

Proof of the easy direction Let $f \in L_p\left(\Omega\right)$ and let $g \in W_p^r\left(\Omega\right)$. Then

$$\begin{aligned}\omega_r\left(f,t\right)_p &\le \omega_r\left(f-g,t\right)_p + \omega_r\left(g,t\right)_p \\ &\le 2^r\left\|f-g\right\|_{L_p(\Omega)} + C\left(r\right)t^r\left|g\right|_{W_p^r(\Omega)} \\ &\le C\left(r\right)\left(\left\|f-g\right\|_{L_p(\Omega)} + t^r\left|g\right|_{W_p^r(\Omega)}\right)\end{aligned}$$

Taking infimum over all possible $g \in W_p^r\left(\Omega\right)$ we obtain the right hand side of (1.1).

♦

Applications of K-functionals

The K-functional appears in many applications such as denoising. Balance between approximation and smoothness.

1. Least squares of 'soft version'

$$\min_{g=\sum \alpha_k N_r(\cdot-k)} \left\|f-g\right\|_2^2 + t\left\|g^{(2)}\right\|_2^2\,.$$

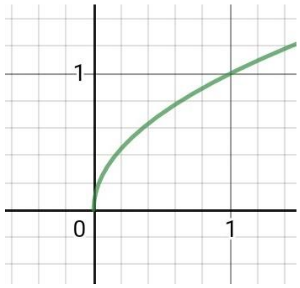
2. Denoising with TV minimization

$$\min_g \left\|f-g\right\|_2 + t\left\|g\right\|_{TV}$$

Lip spaces

Def $0<\alpha\le 1$, $\left|f(x)-f(y)\right|\le M\left|x-y\right|^{\alpha}$. $\left|f\right|_{Lip(\alpha)}:=\inf M$. For $1\le p\le\infty$,

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



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

$$\int_0^1\left|f'(x)\right|dx=1\Rightarrow f'\in L_1$$
$$\Rightarrow \omega_1(f,t)_1\le t\left|f'\right|_1=t \quad .$$
$$\Rightarrow \left|f\right|_{Lip(1,1)}=\sup_{t>0} t^{-1}\omega_1(f,t)_1=1$$

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

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

Linear approximation of Lip functions

Theorem: Let $f\in Lip(\alpha)$. With the linear method of uniform piecewise constants, we have $E_N(f)_{L_\infty([0,1])}:=\inf_{\phi\in S(N_1)^{1/N}}\left\|f-\phi\right\|_\infty\le N^{-\alpha}\left|f\right|_{Lip(\alpha)}$.

Proof [Classic technique] Recall that for $g\in C^1[0,1]$, we constructed $\phi_g\in S(N_1)^{1/N}$, such that $E_N(g)_\infty\le\left\|g-\phi_g\right\|_\infty\le N^{-1}\left|g\right|_{1,\infty}$. Therefore,

$$\left\|f-\phi_g\right\|_\infty\le\left\|f-g\right\|_\infty+\left\|g-\phi_g\right\|_\infty$$
$$\le\left\|f-g\right\|_\infty+N^{-1}\left|g\right|_{1,\infty}$$

For a sequence $\{g_k\}$, with $K_1(f,N^{-1})_\infty=\lim_{k\to\infty}\left\|f-g_k\right\|_\infty+N^{-1}\left|g_k\right|_{1,\infty}$, we get

$$\left\| f-\phi_{g_k} \right\|_\infty \le \left\| f-g_k \right\|_\infty + N^{-1}\left| g_k \right|_{1,\infty} \underset{k\to\infty}{\to} K_1\left(f,N^{-1}\right)_\infty .$$

Using the equivalence of the modulus of smoothness and K-functional,

$$\begin{aligned} E_N\left(f\right)_\infty &\le K_1\left(f,N^{-1}\right)_\infty \\ &\le \omega_1\left(f,N^{-1}\right)_\infty \\ &\le N^{-\alpha}\left|f\right|_{Lip(\alpha)} . \end{aligned}$$

♦

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

Intuition $0\le y<x\le 1$. Let $x=y+h$, $\left(N+1\right)^{-1}\le h\le N^{-1}$. If $x,y\in\left[kN^{-1},\left(k+1\right)N^{-1}\right]$, then with the approximation constant approximation c_k in that interval,

$$\begin{aligned} \left|f\left(x\right)-f\left(y\right)\right| &\le \left|f\left(x\right)-c_k\right|+\left|f\left(y\right)-c_k\right| \\ &\le 2MN^{-\alpha} \\ &\le 2M\left|x-y\right|^\alpha \end{aligned}$$

However, since they might not fall in the same interval, there is a mixing argument.

So linear approximation is kind of limited when α is small. The problem is that we're not spending enough 'budget' in the vicinity of zero.

Adaptive / Nonlinear / Sparse approximation

Remark This is the univariate version of the ML decision trees and random forests we will review later

Denote

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

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

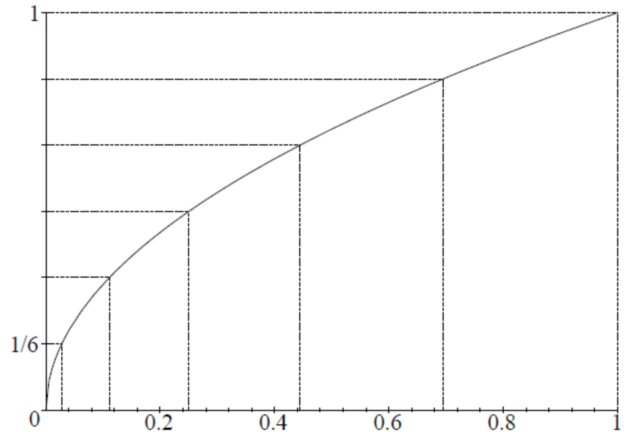
If f' exists a.e., $Var\left(f\right)=\left\|f'\right\|_1$. Why?

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

Let's go back to the examples $f\left(x\right)=x^\alpha$. In our case $\left\|f'\right\|_1=Var\left(f\right)=1$.

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

$\left|f(x)-a_j\right| \le \frac{Var(f)}{2N}$, $\forall x \in \left[t_j, t_{j+1}\right]$. This gives a free knot spline $g \in \Sigma_N$ with $\|f-g\|_\infty \le \frac{Var(f)}{2N} \le \frac{1}{2N}$, $f(x)=x^\alpha$, $0<\alpha\le 1$.



In our example, 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 $f(x)=x^\alpha$, $0<\alpha<1$,

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