α-Hölder smoothness

Chanwoo Lee, August 16, 2021

For one variate function, many papers follow definition of α-Hölder class in [2]. The following properties focuse on $\alpha \in (0, \infty)$.

Definition 1 (Univariate case [2])**.** Let $\alpha \in (0, \infty)$. A function $f\colon \mathcal{D} \to \mathbb{R}$ is in α-Hölder class if

1. $\sup_{x\in\mathcal{D}} |f^{(j)}(x)| < \infty$, for all $j = 0, 1, \ldots, \lfloor\alpha\rfloor$.
2. $\sup_{x\neq x'\in\mathcal{D}} \frac{|f^{(\lfloor\alpha\rfloor)}(x) - f^{(\lfloor\alpha\rfloor)}(x')|}{|x-x'|^{\alpha-\lfloor\alpha\rfloor}} < \infty$.

Define α-Hölder norm

$$\|f\|_{\mathcal{H}_\alpha} = \max_{j\leq\lfloor\alpha\rfloor} \sup_{x\in\mathcal{D}} |f^{(j)}(x)| + \sup_{x\neq x'\in\mathcal{D}} \frac{|f^{(\lfloor\alpha\rfloor)}(x) - f^{(\lfloor\alpha\rfloor)}(x')|}{|x-x'|^{\alpha-\lfloor\alpha\rfloor}}.$$

Remark 1. When $\mathcal{D}$ is a compact set (e.g. closed interval), the first condition in Definition (1) can be dropped. The existence of $f^{(\lfloor\alpha\rfloor)}$ directly implies the boundness of lower-order derivatives.

Notice we have equivalent definition of Hölder class with Definition 1 using α-Hölder norm.

Definition 1* (Variation for univariate case)**.** A function $f\colon \mathcal{D} \to \mathbb{R}$ is in Hölder class if

$$\|f\|_{\mathcal{H}_\alpha} \leq M,$$

for some $M = M(\alpha) > 0$.

Now consider multivariate function $f\colon \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} = \mathcal{D}^d$. For multi-index κ, we denote partial derivatives,

$$\nabla_\kappa f(x) = \frac{\partial^{|\kappa|} f(x)}{(\partial x)^\kappa}.$$

Then we define multivariate version of α-Hölder class as in [3],

Definition 2 (Multivariate case [3])**.** Define α-Hölder norm with respect to norm $\|\cdot\|$,

$$\|f\|_{\mathcal{H}_\alpha} = \max_{\boldsymbol{j}\colon |\boldsymbol{j}|\leq\lfloor\alpha\rfloor} \sup_{\boldsymbol{x}\in\mathcal{X}} |\nabla_{\boldsymbol{j}} f(\boldsymbol{x})| + \max_{\kappa=\lfloor\alpha\rfloor} \sup_{\boldsymbol{x}\neq\boldsymbol{x}'\in\mathcal{D}} \frac{|\nabla_\kappa f(\boldsymbol{x}) - \nabla_\kappa f(\boldsymbol{x}')|}{\|\boldsymbol{x}-\boldsymbol{x}'\|^{\alpha-\lfloor\alpha\rfloor}}.$$

Then f is in α-Hölder class if $\|f\|_{\mathcal{H}_\alpha} \leq M$, for some $M > 0$.

Remark 2. In [3], they choose to use $\|\cdot\|$ as zero norm when they define α-Hölder norm.

In a relation to α-Hölder class, we define α-Hölder smooth function. This smoothness definition is equivalent to α-Hölder class [1] (see Lemma 1 for one way) and is used in [4].

Definition 2* (α-Hölder smooth [4])**.** A function $f\colon \mathcal{X} \to \mathbb{R}$ is α-Hölder smooth with respect to $\|\cdot\|$ if there exists a polynomial $P_k(\cdot - \boldsymbol{x}_0)$ of degree $k = \lfloor\alpha\rfloor$, such that

$$|f(\boldsymbol{x}) - P_k(\boldsymbol{x}-\boldsymbol{x}_0)| \leq c\|\boldsymbol{x}-\boldsymbol{x}_0\|^\alpha, \text{ for all } \boldsymbol{x}, \boldsymbol{x}_0 \in \mathcal{X}.$$

Lemma 1. A function f is in α-Hölder class, if and only if f is a α-Hölder smooth.

Proof. Here we prove only one way. Let $P_k(\cdot - \boldsymbol{x}_0)$ be the Taylor polynomial of degree $\lfloor \alpha \rfloor$,

$$P_k(\cdot - \boldsymbol{x}_0) = \sum_{\kappa:\, |\kappa| \le \lfloor \alpha \rfloor} \frac{\nabla_\kappa f(\boldsymbol{x}_0)}{\kappa} (\boldsymbol{x} - \boldsymbol{x}_0)^\kappa.$$

Then, $P_k(\cdot - \boldsymbol{x}_0)$ satisfies,

$$\begin{aligned}
|f(\boldsymbol{x}) - P_k(\boldsymbol{x} - \boldsymbol{x}_0)| &= \sum_{\kappa:\, |\kappa| = \lfloor \alpha \rfloor} \frac{|\nabla_\kappa f(\boldsymbol{z}) - \nabla_\kappa f(\boldsymbol{x}_0)|}{\kappa!} (\boldsymbol{x} - \boldsymbol{x}_0)^\kappa, \text{ where } \boldsymbol{z} = \boldsymbol{x}_0 + c(\boldsymbol{x} - \boldsymbol{x}_0), c \in (0,1), \\
&\lesssim \sum_{\kappa:\, |\kappa| = \lfloor \alpha \rfloor} \frac{|\nabla_\kappa f(\boldsymbol{z}) - \nabla_\kappa f(\boldsymbol{x}_0)|}{\kappa!} \|\boldsymbol{x} - \boldsymbol{x}_0\|^{\lfloor \alpha \rfloor} \\
&\le \sum_{\kappa:\, |\kappa| = \lfloor \alpha \rfloor} \frac{\|\boldsymbol{x} - \boldsymbol{x}_0\|^{\alpha - \lfloor \alpha \rfloor}}{\kappa!} \|\boldsymbol{x} - \boldsymbol{x}_0\|^{\lfloor \alpha \rfloor} \\
&\le M_\alpha \|\boldsymbol{x} - \boldsymbol{x}_0\|^\alpha.
\end{aligned}$$

□

Important fact:

1. α-Hölder smoothness is the same as α-Lip smoothness only if $\alpha \in (0,1)$. These two smoothness notions are different when $\alpha \ge 1$.
2. The smoothness index depends on the $\mathcal{D}$. E.g. function $f(x) = x^2$ is ∞-Hölder smooth at $x \in [0,1]$, but it is 0-Hölder smooth over $x \in \mathbb{R}$.

   $f(x) = x^2$ is ∞-smooth for all $x \in \mathbb{D}$, where $\mathcal{D}$ is bounded. This is because by Taylor's theorem, we have

   $$|f(x) - P_k(x - x_0)| \le o(\|x - x_0\|^k),$$

   for all positive integers k and $x, x_0 \in \mathbb{R}$. Consider when $f(x) = x^\alpha$, $\alpha > 0$. $\alpha \in \mathbb{R}_+/\mathbb{N}_+$? Then $f(x)$ is ∞-smooth when we consider $x \in \mathcal{D} - \{0\}$ while $\lfloor \alpha \rfloor$-smooth when $x \in \mathbb{R}$. This is because $f(x)$ is infinitely differentiable at $x \ne 0$ while $f(x)$ is not $(\lfloor \alpha \rfloor + 1)$-differentiable at $x = 0$.
3. For a given function $f(x)$, an easy way to determine $\alpha \in (0, \infty)$ is to take derivative until it blows up. E.g. $f(x) = \log x$ is 0-Hölder at $x = 0$ and ∞-Hölder everywhere else.

Relation to smoothness condition in [6, 5]:

1. Smoothness assumption in [6]:

   First, they define a α Hölder class function f as

   $$\sum_{\kappa:\, |\kappa| = \lfloor \alpha \rfloor} \frac{1}{\kappa!} |\nabla_\kappa f(x) - \nabla_\kappa f(x')| \le L \|x - x'\|_\infty^{\alpha - \lfloor \alpha \rfloor}.$$

   Then, assume that $f(\cdot, y)$ is Hölder smooth function for all $y \in \mathcal{X}$. In the proof of Proposition

1, α smoothness is used only to prove that there exists a piece-wise $\lfloor \alpha \rfloor$-degree polynomial function that satisfies Definition 2*. So I think [6] uses the same definition of usual Hölder class here.

However, in the later part of the paper, [6] defines another smoothness concept in Theorem 4 independently of Proposition 1. There exists positive constants a and b such that for all multi-indices κ and $y \in \mathcal{X}$,

$$\sup_{x \in \mathcal{X}} \frac{\partial^{|\kappa|} f(x, y)}{(\partial x)^{\kappa}} \leq b a^{|\kappa|} \kappa!. \tag{1}$$

This assumption says that $f(\cdot, y)$ is infinitely many times differentiable and the partial derivative satisfies (1). In the proof of Theorem 4, they first set arbitrary $(\ell - 1)$-degree Taylor polynomial and choose suitable ℓ with respect to thresholded rank r so that the final bound has simple form (bias and variance tradeoff). Notice that (1) is simply used to bound approximation error (residual term of $(\ell - 1)$-degree Taylor polynomial is bounded with respect to (1)).

2. Smoothness assumption in [5]:

   [5] assume the infinitely analytic function $f(\cdot, y)$ for all $\{\|y\| \leq R\}$. That means that for all multi-indices κ and for all $x \in \{\|x\| \leq R\}$, the following holds,

$$\left\| \frac{\partial^{|\kappa|} f(x, y)}{(\partial x)^{\kappa}} \right\| \leq C M^{|\kappa|} \|f\|,$$

   for some $C, M > 0$ and any fixed $y \in \{\|y\| \leq R\}$. This is almost similar condition with (1) except $\kappa!$ part. This $\kappa!$ part is included in $\hat{X}_{ij}$ term when they find an low rank approximation of $X_{ij} = f(x_i, y_j)$ in their Lemma 4.1. such that

$$\|X - \hat{X}\|_{\max} \leq \epsilon \|f\|.$$

   So their assumption is almost the same as (1).

References

[1] Patrik Andersson. Characterization of pointwise hölder regularity. *Applied and Computational Harmonic Analysis*, 4(4):429–443, 1997.

[2] Lawrence C Evans. Partial differential equations and monge-kantorovich mass transfer. *Current developments in mathematics*, 1997(1):65–126, 1997.

[3] Chao Gao, Yu Lu, and Harrison H Zhou. Rate-optimal graphon estimation. *The Annals of Statistics*, 43(6):2624–2652, 2015.

[4] Olga Klopp, Alexandre B Tsybakov, and Nicolas Verzelen. Oracle inequalities for network models and sparse graphon estimation. *The Annals of Statistics*, 45(1):316–354, 2017.

[5] Madeleine Udell and Alex Townsend. Why are big data matrices approximately low rank? *SIAM Journal on Mathematics of Data Science*, 1(1):144–160, 2019.

[6] Jiaming Xu. Rates of convergence of spectral methods for graphon estimation. In *International Conference on Machine Learning*, pages 5433–5442. PMLR, 2018.