Minimax-optimal and Locally-Adaptive Online Nonparametric Regression

ALT2025, Milan, Italy

Paul Liautaud[1], Pierre Gaillard[2], Olivier Wintenberger[1,3]

February 24, 2025

[1]Sorbonne Université, CNRS, LPSM, Paris, France
[2]Université Grenoble Alpes, INRIA, CNRS, Grenoble INP, LJK, Grenoble, France
[3]Institut Pauli CNRS, Vienna University, Oskar Morgenstern Platz 1, Wien, Austria

Setting: online regression with individual sequences (1/2)

Online prediction scenario: at each round $t \in \mathbb{N}^*$, the forecaster

1. observes an input $x_t \in \mathcal{X}$;
2. chooses a prediction $\widehat{f}_t(x_t) \in \mathbb{R}$; — Choose $\widehat{f}_t$ before observing ℓ_t
3. incurs a loss $\ell_t(\widehat{f}_t(x_t))$ — No assumptions on how ℓ_t is generated
4. updates his prediction function $\widehat{f}_t \to \widehat{f}_{t+1}$ — Based on observed gradients

$\cdots \to x_t =$  $\to$  $\to$ 'zebra' $\to$  $\to x_{t+1} =$  $\to \cdots$

Goal: given some large (nonparametric) function set $\mathcal{F} \subset \mathbb{R}^{\mathcal{X}}$, we want to minimize the regret against any competitor $f \in \mathcal{F}$

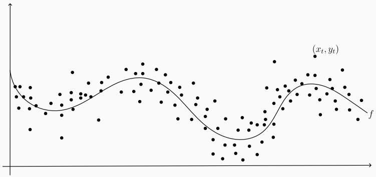
$$\mathrm{Reg}_T(f) = \underbrace{\sum_{t=1}^{T} \ell_t(\widehat{f}_t(x_t))}_{\text{our performance}} - \underbrace{\sum_{t=1}^{T} \ell_t(f(x_t))}_{\text{reference performance}} \underbrace{= o(T)}_{\text{goal}} .$$

Setting: online regression with individual sequences (2/2)

Individual sequences: no stochastic assumption on data (x_t, ℓ_t)!
$\widehat{f}_1, \dots, \widehat{f}_T$ have to perform **well** with all **arbitrary** and possibly **adversarial** sequences.

Assumptions:

- $\ell_1, \dots, \ell_T$ are G-Lipschitz convex losses, with $G > 0$;
- $\mathcal{X} \subset \mathbb{R}^d$ bounded compact subset;
- $\mathcal{F} \subset [-B, B]^{\mathcal{X}}$ for some $B > 0$;
- $\mathcal{F} \subset \mathscr{C}^\alpha(L)$ the set of α-Hölder continuous functions, with $\alpha \in (0, 1]$, $L > 0$ **unknown**.

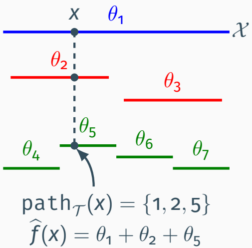


Contribution 1: parameter-free online approach with chaining trees

We present a **parameter-free** online learning method that leverages a **chaining tree** structure and achieves a regret over α-Hölder continuous functions $\mathscr{C}^\alpha(L)$:

🌲 **Chaining tree of depth $M = 3$:**

$$\sup_{f \in \mathscr{C}^\alpha(L)} \mathrm{Reg}_T(f) \lesssim GB\sqrt{T} + GL(f) \begin{cases} \sqrt{T}, & \text{if } d < 2\alpha, \\ \log_2 T\sqrt{T}, & \text{if } d = 2\alpha, \\ T^{1-\frac{\alpha}{d}}, & \text{if } d > 2\alpha. \end{cases}$$



★ **Adaptivity** to both L and α!

✔ Our rates are **minimax** over $\mathscr{C}^\alpha(L)$ for general convex losses (Rakhlin et al., 2015)

💻 Our algorithm is **computationally tractable**: we update $O(\frac{1}{d}\log_2(T))$ parameters at each round.

Main intuitions behind our algorithm

Decompose regret: $\mathrm{Reg}_T(f) = \underbrace{\sum_{t=1}^T \ell_t(\widehat{f}_t(x_t)) - \ell_t(\widehat{f}_M(x_t))}_{R_1\text{:estimation regret}} + \underbrace{\sum_{t=1}^T \ell_t(\widehat{f}_M(x_t)) - \ell_t(f(x_t))}_{R_2\text{:approximation regret}}$

Multi-scale approximation process of a chaining tree $\widehat{f}_M$:

1. *Control of the coefficient decay:*

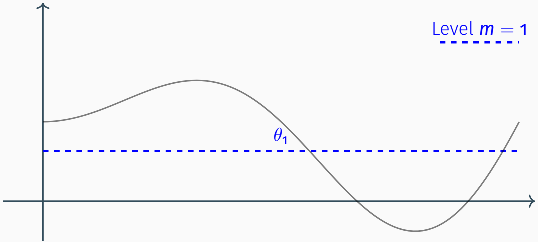
$$|\theta_{\text{level } m}| \leqslant L2^{-\alpha m}$$

2. *Control of estimation regret $(\widehat{f}_t) \to \widehat{f}_M$:*

$$R_1 \leqslant GL \sum_{m=1}^M 2^{-\alpha m} \sqrt{2^{dm} T}\,.$$

3. *Control of approximation regret:*

$$R_2 \leqslant GT \cdot \sup_{f \in \mathscr{C}^\alpha(L)} \|\widehat{f}_M - f\|_\infty \lesssim GTL2^{-\alpha M}$$



Previous works: Gaillard and Gerchinovitz; Cesa-Bianchi et al. (2015; 2017) designed explicit chaining algorithms for square and absolute loss.

Contribution 2: locally-adaptive algorithm

We present an algorithm that **optimally** competes against any **pruning** and adapts to the **local Hölder regularities** of the competitor, achieving for $d = 1, \alpha \in [\frac{1}{2}, 1]$:

$$\sup_{f \in \mathscr{C}^\alpha(L)} \mathrm{Reg}_T(f) \lesssim \inf_{\text{prun}} \left\{ \sqrt{T|\text{prun}|} + \sum_{n \in \text{prun}} 2^{-\alpha \text{level}(n)} L_n(f) \sqrt{|T_n|} \right\},$$

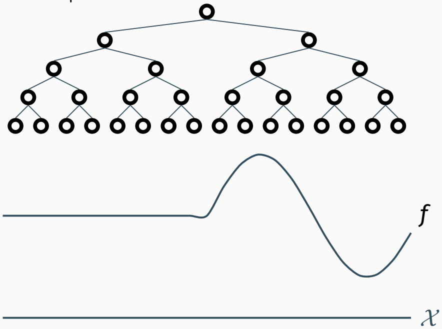
with $T_n = \{1 \leqslant t \leqslant T : x_t \in \mathcal{X}_n\}$.

★ **Adaptivity** to local regularities ($L_n(f)$) with respect to any pruning.

✔ **Adaptivity** to the loss curvature.

🏆 From global $O(L\sqrt{T})$ to **local** $O(\sum_n L_n \sqrt{|T_n|})$: low regret in low-variation regions.

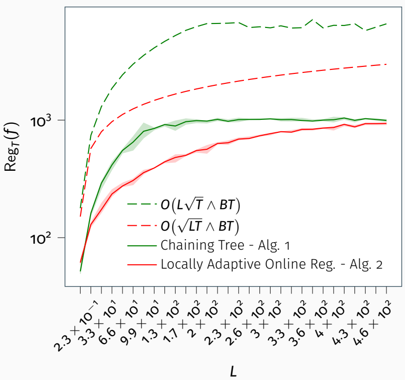
Example 1:



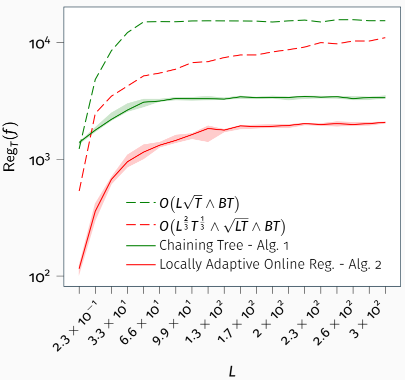
Corollary: minimax-optimality - $d = 1, \alpha \in [\frac{1}{2}, 1]$

Reference	Assumptions	Regret bound
Alg. 2	(ℓ_t) exp-concave, $L > 0$ unknown	$\min\left\{L^{\frac{1}{2\alpha}}\sqrt{T}, L^{\frac{2}{2\alpha+1}}T^{\frac{1}{2\alpha+1}}\right\}$
	(ℓ_t) convex, $L > 0$ unknown	$L^{\frac{1}{2\alpha}}\sqrt{T}$
Kuzborskij et al. (2020)	(ℓ_t) square loss, $L > 0$ unknown, $\alpha = 1$	$\sqrt{LT}$
Hazan et al. (2007)	(ℓ_t) square loss, $L > 0$ known, $\alpha = 1$	$\sqrt{LT}$

ℓ_t absolute loss, $T = 2000$



ℓ_t squared loss, $T = 2000$



Conclusion

- We propose a parameter-free online strategy on chaining tree achieving **minimax regret**;
- A unique algorithm that both adapts to **local regularities** of the competitor and **curvature** of sequential losses;
- **First** constructive algorithm to achieve **optimal locally adaptive regret**;
- Open problem: could this procedure be adapted to approach other classes of functions ($\alpha \geqslant 1$).

Thank you!

Questions?

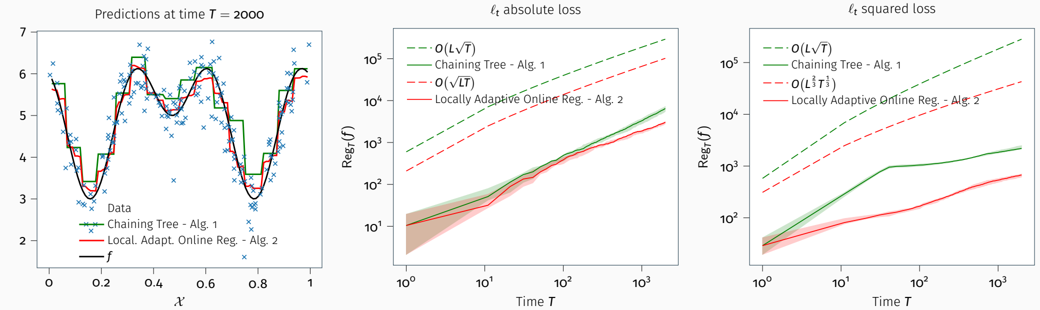
Comparison with the litterature

Ref.	Assumptions	Upper bound
[1]	(ℓ_t) exp-concave, $L > 0$ unknown	$\min\left\{\sqrt{LT}, L^{\frac{2}{3}}T^{\frac{1}{3}}\right\}$
	(ℓ_t) convex, $L > 0$ unknown	$\sqrt{LT}$
[2]	(ℓ_t) square loss, $L > 0$ unknown	$\sqrt{LT}$
[3]	(ℓ_t) absolute loss, $L > 0$ known	$L^{\frac{1}{3}}T^{\frac{2}{3}}$
	(ℓ_t) square loss, $L > 0$ known	$\sqrt{LT}$
[4]	(ℓ_t) square loss, $L = 1$ known	$T^{\frac{1}{3}}$
[5]	(ℓ_t) convex, $L = 1$ known	$\sqrt{T}$

[1] Liautaud, Gaillard, and Wintenberger, “Minimax-optimal and Locally-adaptive Online Nonparametric Regression”.
[2] Kuzborskij and Cesa-Bianchi, “Locally-adaptive nonparametric online learning”.
[3] Hazan, Agarwal, and Kale, “Logarithmic regret algorithms for online convex optimization”.
[4] Gaillard and Gerchinovitz, “A Chaining Algorithm for Online Nonparametric Regression”.
[5] Cesa-Bianchi et al., “Algorithmic chaining and the role of partial feedback in online nonparametric learning”.

Experiments

Regression setting: $y_t = f(x_t) + \varepsilon_t$, where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.5, f(x) = \sin(10x) + \cos(5x) + 5$, for $x \in \mathcal{X} = [0, 1]$ and $\sup_x |f'(x)| \leqslant 15 =: L$.



References

Cesa-Bianchi, Nicolò and Gábor Lugosi (2006). *Prediction, Learning, and Games.* Cambridge University Press.

Cesa-Bianchi, Nicolò et al. (2017). "Algorithmic chaining and the role of partial feedback in online nonparametric learning". In: *Conference on Learning Theory*. PMLR, pp. 465–481.

Cutkosky, Ashok and Francesco Orabona (2018). "Black-box reductions for parameter-free online learning in banach spaces". In: *Conference On Learning Theory*. PMLR, pp. 1493–1529.

Gaillard, Pierre and Sebastien Gerchinovitz (2015). "A Chaining Algorithm for Online Nonparametric Regression". In: *COLT*.

Hazan, Elad, Amit Agarwal, and Satyen Kale (2007). "Logarithmic regret algorithms for online convex optimization". In: *Machine Learning* 69.2, pp. 169–192.

Kuzborskij, Ilja and Nicolo Cesa-Bianchi (2020). "Locally-adaptive nonparametric online learning". In: *Advances in Neural Information Processing Systems* 33, pp. 1679–1689.

Liautaud, Paul, Pierre Gaillard, and Olivier Wintenberger (2024). "Minimax-optimal and Locally-adaptive Online Nonparametric Regression". In: *arXiv preprint arXiv:2410.03363*.

Mhammedi, Zakaria and Wouter M Koolen (2020). “Lipschitz and comparator-norm adaptivity in online learning”. In: *Conference on Learning Theory*. PMLR, pp. 2858–2887.

Orabona, Francesco and Dávid Pál (2016). “Coin betting and parameter-free online learning”. In: *Advances in Neural Information Processing Systems* 29.

Rakhlin, Alexander and Karthik Sridharan (2014). “Online non-parametric regression”. In: *Conference on Learning Theory*. PMLR, pp. 1232–1264.

— (2015). “Online nonparametric regression with general loss functions”. In: *arXiv preprint arXiv:1501.06598*.

Zinkevich, Martin (2003). “Online convex programming and generalized infinitesimal gradient ascent”. In: *Proceedings of the 20th international conference on machine learning (icml-03)*, pp. 928–936.