Approximating **Tame functions** (NNs) by piecewise polynomials: **Theoretical bound** & **Numerical results**

Piecewise Polynomial Regression of Tame Functions via Integer Programming

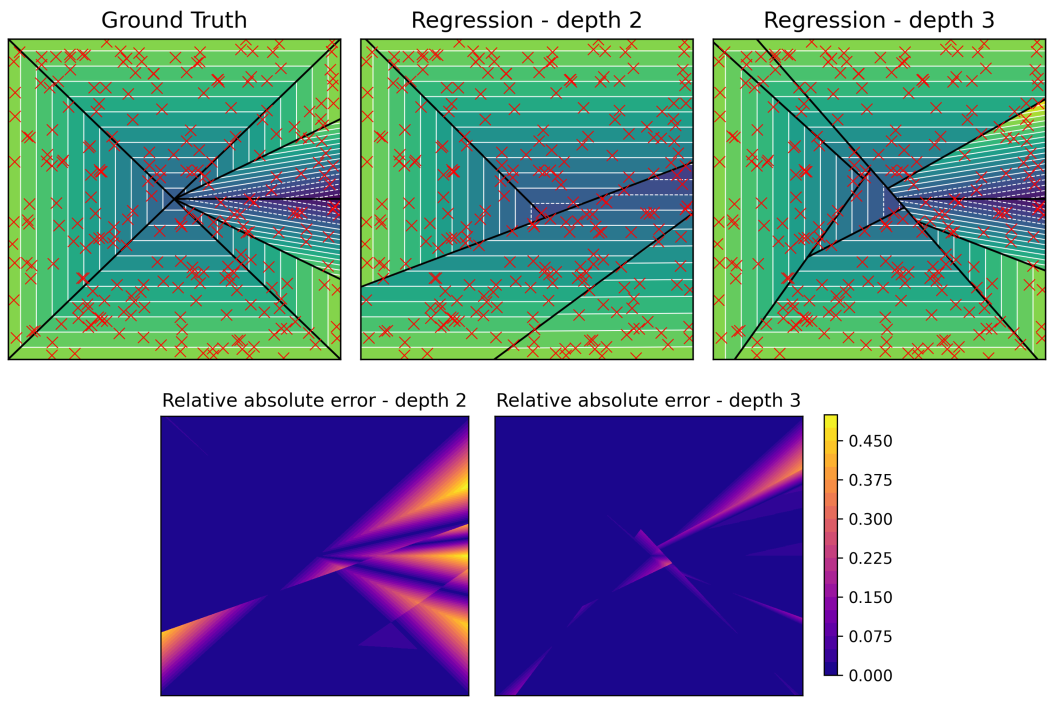
Gilles Bareilles, Johannes Aspman, Jiří Němeček, Jakub Mareček
Faculty of Electrical Engineering, Czech Technical University in Prague

Tame Functions

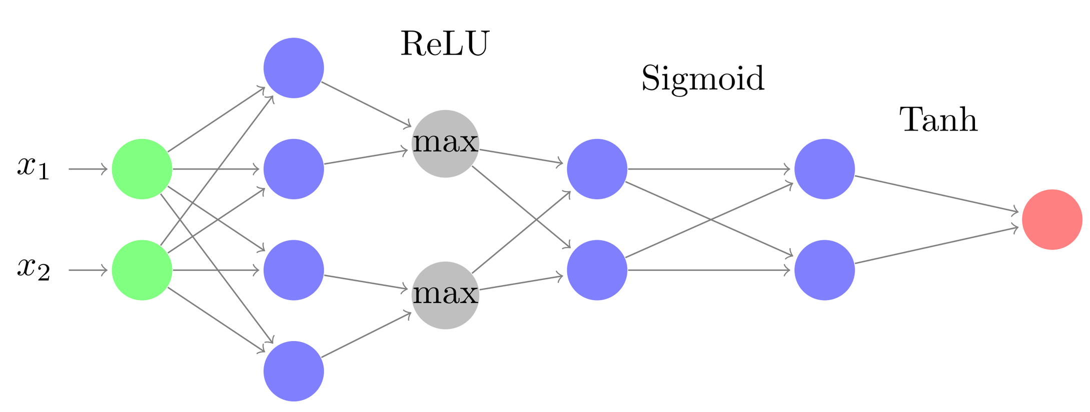
Tame functions are a general class of nonsmooth and nonconvex functions. They appear in a broad range of useful and difficult applications, such as:

- common deep learning architectures [8, 6],
- empirical risk minimization frameworks [9],
- in mixed-integer optimization, with the value function and the solution to the so-called subadditive dual [2],
- in quantum information theory, with approximations of the matrix exponential for a k-local Hamiltonian [7, 1],
- and in quantum chemistry, with functions describing the electronic structure of molecules.

An example tame function ("cone") with seven strata (left) and its approximation and error



A Neural Network example of a tame function:



Approximation bound

When approximating a m-times continuously *differentiable function* f, the best degree N polynomial incurs an $\mathcal{O}(1/N^m)$ error [3].

For *nonsmooth functions*, the rate is notably worse. For example, approximating the absolute value over the interval $[-1, 1]$ incurs *exactly* the slow rate $1/N$ [4]:

$$\inf_{p \in \mathcal{P}_N} \| |\cdot| - p \|_{\infty,[-1,1]} = \frac{\beta}{N} + o\left(\frac{1}{N}\right).$$

Thus, to approximate **Tame functions**, polynomials are not efficient. Denote $\widetilde{\mathcal{P}}_N^l(A)$ a set of *piecewise* polynomials on A, such that

1. each piece is a polyhedron defined as the intersection of l halfspaces, represented as a leaf of a complete binary tree of depth l,
2. the restriction of the function to each piece is a polynomial of degree at most N.

Theorem 1. *Consider a function $f : A \to \mathbb{R}$, a qualified* (*Asm. 1 in the paper*) *$\mathcal{C}^m$-stratification of f for some $m \geq 2$ such that:*

- *f is a **tame** function,*
- *A is a connected compact subset of $[0, 1]^n$,*
- *f is continuous over A.*

*Then f can be approximated by **piecewise polynomial functions**:*

$$\inf_{p \in \widetilde{\mathcal{P}}_N^l(A)} \|f - p\|_{\infty,A} \leq C_1 N^{-m} + C_2 l^{-\frac{2}{n-1}}.$$

where $C_1 = c_1(n, m, A, f)$ and $C_2 = c_2(n, m, f)$.

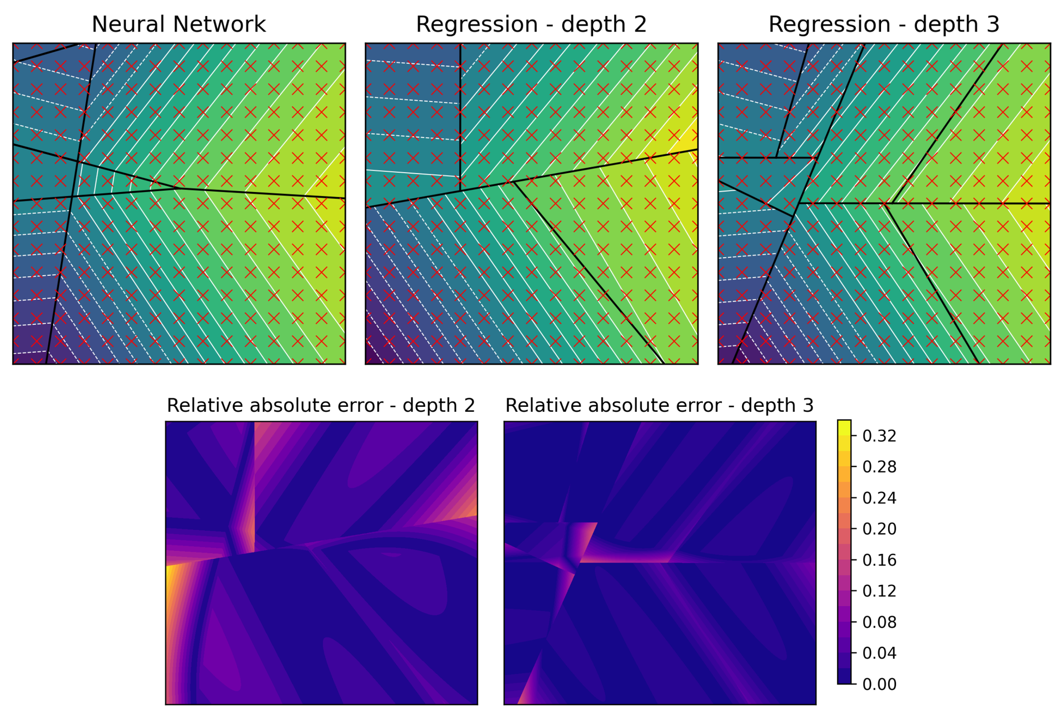
Stratification

We say that f admits a definable $\mathcal{C}^m$-stratification if there exists a finite partition of $\mathbb{R}^n$ such that

- each stratum is a definable connected $\mathcal{C}^m$-manifold,
- f is $\mathcal{C}^m$-smooth on each stratum,
- any two strata $\mathcal{M}$ and $\mathcal{M}'$ satisfy the frontier condition: $\mathcal{M} \cap \operatorname{cl} \mathcal{M}' \neq \emptyset \Rightarrow \mathcal{M} \subset \operatorname{cl} \mathcal{M}'$.

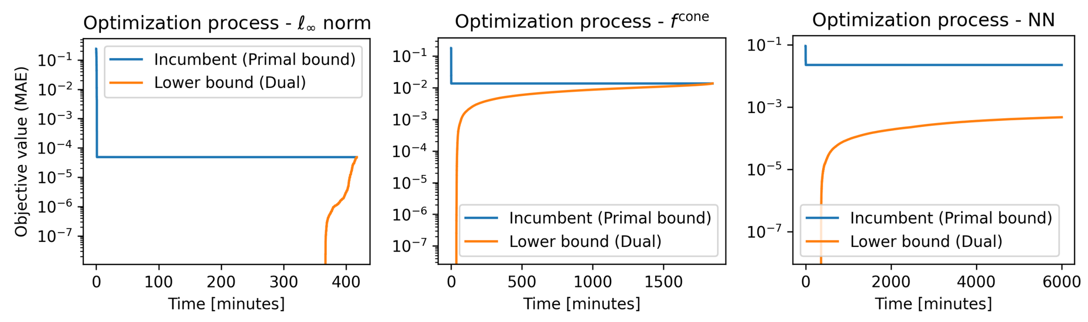
Mixed-Integer Formulation

We slightly modify the affine-hyperplane tree Mixed-Integer Linear formulation (OCT-H, [5]) to separate the input space into 2^l polyhedral regions and fit a polynomial to each region.



Solving process

The primal objective value converges quickly; most of the computation time is spent on improving lower bound.



References

[1] Christos Aravanis, Johannes Aspman, Georgios Korpas, and Jakub Marecek. Polynomial matrix inequalities within tame geometry. *arXiv preprint arXiv:2206.03941*, 2022.

[2] Johannes Aspman, Georgios Korpas, and Jakub Marecek. Taming binarized neural networks and mixed-integer programs. *arXiv preprint arXiv:2310.04469*, 2023.

[3] T. Bagby, L. Bos, and N. Levenberg. Multivariate simultaneous approximation. *Constructive Approximation*, 18(4):569–577, December 2002. ISSN 1432-0940. doi: 10.1007/s00365-001-0024-6.

[4] Serge Bernstein. Sur la meilleure approximation de $|x|$ par des polynomes de degrés donnés. *Acta Mathematica*, 37(1):1–57, December 1914. ISSN 1871-2509. doi: 10.1007/BF02401828.

[5] Dimitris Bertsimas and Jack Dunn. Optimal classification trees. *Machine Learning*, 106(7):1039–1082, July 2017. ISSN 1573-0565. doi: 10.1007/s10994-017-5633-9.

[6] Jérôme Bolte and Edouard Pauwels. A mathematical model for automatic differentiation in machine learning. In *Advances in Neural Information Processing Systems*, volume 33, pages 10809–10819. Curran Associates, Inc., 2020.

[7] Denys I Bondar, Zakhar Popovych, Kurt Jacobs, Georgios Korpas, and Jakub Marecek. Recovering models of open quantum systems from data via polynomial optimization: Towards globally convergent quantum system identification. *arXiv preprint arXiv:2203.17164*, 2022.

[8] Damek Davis, Dmitriy Drusvyatskiy, Sham Kakade, and Jason D. Lee. Stochastic Subgradient Method Converges on Tame Functions. *Foundations of Computational Mathematics*, 20(1):119–154, February 2020. ISSN 1615-3383. doi: 10.1007/s10208-018-09409-5.

[9] Franck Iutzeler and Jérôme Malick. Nonsmoothness in Machine Learning: Specific Structure, Proximal Identification, and Applications. *Set-Valued and Variational Analysis*, 28(4):661–678, December 2020. ISSN 1877-0533, 1877-0541. doi: 10.1007/s11228-020-00561-1.

Scan the QR code to see the full paper

CTU CZECH TECHNICAL UNIVERSITY IN PRAGUE