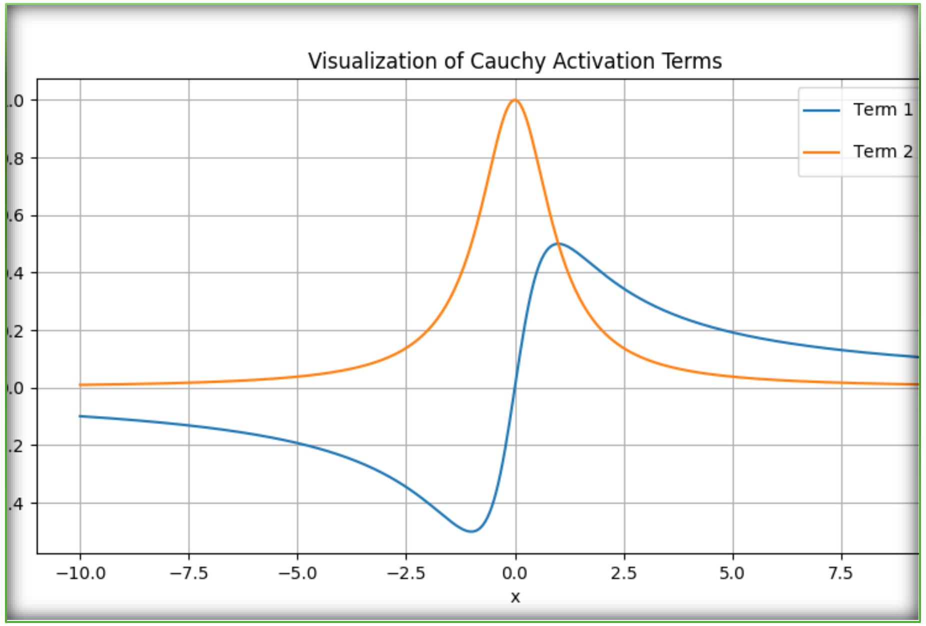
From Kolmogorov to Cauchy: Shallow XNet Surpasses KANs

Xin Li · Xiaotao Zheng · Zhihong Xia

How does a computer algorithm express a function?

- Express a function in some basis functions, or activation functions.
- Cauchy activation and Cauchy approximation, derived from Cauchy integral formula in complex analysis.

In theory, Cauchy activation can achieve an approximation of arbitrarily high order. The error is in the order of $o(\mathrm{e}^{-\mathrm{cn}})$ for some c.



High-Order Approximation of XNet

Theorem (High-Order Approximation of XNet)

Let $K \subset \mathbb{R}^d$ *be compact, and let* $f \in C^\omega(K)$ *be real analytic.*

Let $\Phi = \left\{ \frac{\lambda_1 x + \lambda_2}{x^2 + d^2} \right\}$ *be the* **Cauchy activation family***.*

Then for any $r > 0$*, there exists a one-hidden-layer XNet function* $f_N \in \mathrm{span}(\Phi)$ *such that*

$$\|f - f_N\|_{C^0(K)} = \mathcal{O}(N^{-r}).$$

Super-polynomial approximation: $\|f - f_N\| = \mathcal{O}(N^{-r})$ **for any** $r > 0$**.**

- **New:** explicit super-polynomial rate with any r.
- Previous work: XNet is universal *(rate not quantified)*.

Heaviside Step Function (Discontinuity)

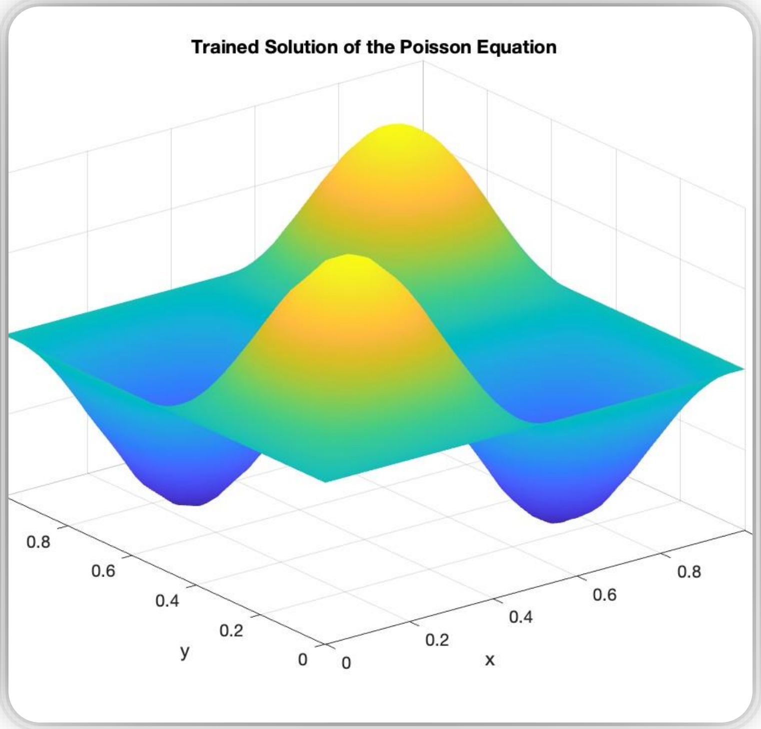
Model	MSE
XNet [1,64,1]	**8.99e-08**
ReLU (shallow)	2.05e-03
ReLU (deep)	6.81e-05
KAN (200 grids)	5.98e-04

$\sim$1000$\times$ lower MSE

(8.99e-08 vs 5.98e-04); no overshoot at $x=0$.

APPLICATION TO PINN

Compare XNet with KAN



KAN uses Kolmogorov-Arnold Representation Theorem. We compared XNet with various applications. This is a typical comparison.

Table: Comparison of XNet and KAN on the Poisson equation.

Metric	MSE	RMSE	MAE	Time (s)
PINN [2,20,20,1]	1.7998e-05	4.2424e-03	2.3300e-03	48.9
XNet (20)	1.8651e-08	1.3657e-04	1.0511e-04	57.2
KAN [2,10,1]	5.7430e-08	2.3965e-04	1.8450e-04	286.3
XNet (200)	1.0937e-09	3.3071e-05	2.1711e-05	154.8

100 dimensional Allen-Cahn Equation

Allen-Cahn Equation

a reaction-diffusion equation for the modeling of phase separation and transition in physics. Here we consider a typical Allen-Cahn equation with the “double-well potential” in 100-dimensional space

$$\frac{\partial u}{\partial t}(t,x) = \Delta u(t,x) + u(t,x) - [u(t,x)]^3,$$

with initial condition $u(0,x) = g(x)$.

Thank You!

Questions & Discussion welcome.