Breaking Under Stress?
Scalability and Robustness of Kolmogorov-Arnold Networks

TU Delft — Delft University of Technology

Runan Han

1.Introduction

What is **KAN** (Kolomogrov-Arnold Networks)?

- **Like MLP** (Multi-Layer Perceptron), but with **learnable functions on edges** instead of fixed activation functions on nodes.
- Good for discovering **mathematical, physical laws**

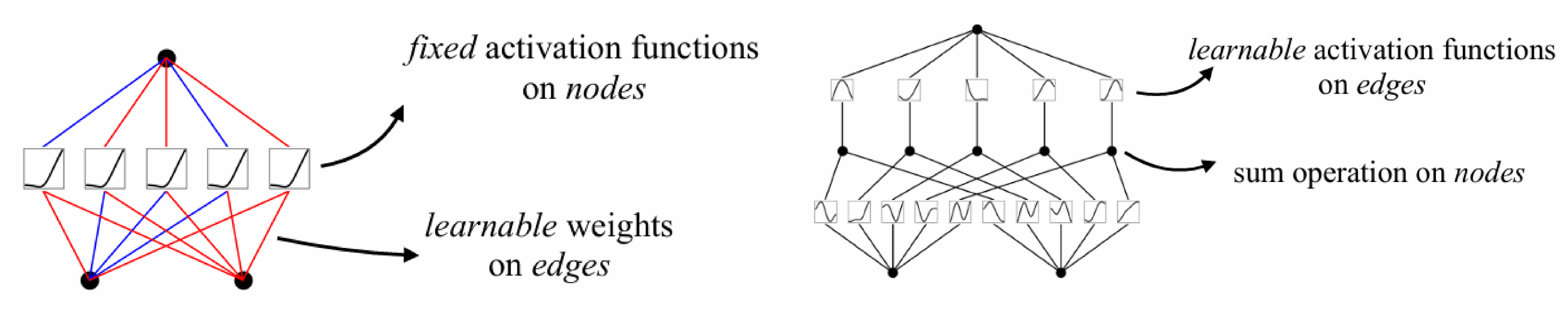


Table 1. MLP v.s. KAN, adapted from [1]

Problems:

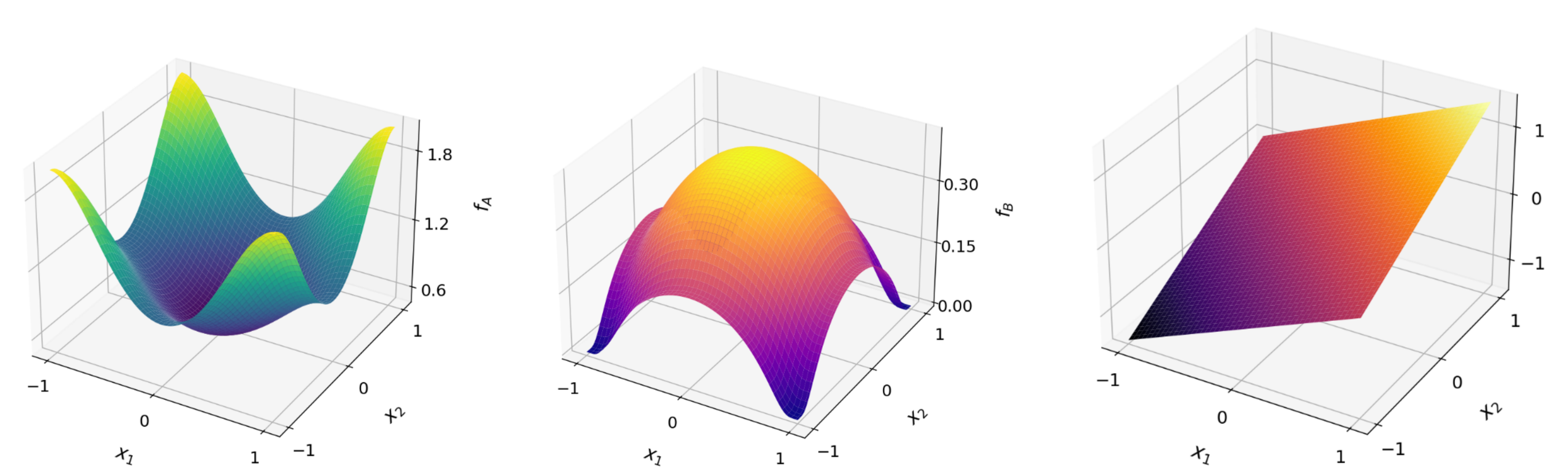
- Original KAN paper [1] claimed KANs "beat **curse of dimensionality**", **true in practice?**
- Assumes perfect data, **real world data not perfect!**

2.Research Question

How does the **empirical performance** of Kolmogorov-Arnold Networks (**KANs**) compare to Multi-Layer Perceptrons (**MLPs**) as data **dimensionality** and **noise levels increase**?

3.Experimental Setup

Generate 3 different types of regression data:



f_A: **Non-linear & compositional.** Expectations: KAN excels because learnable functions; MLP struggles.

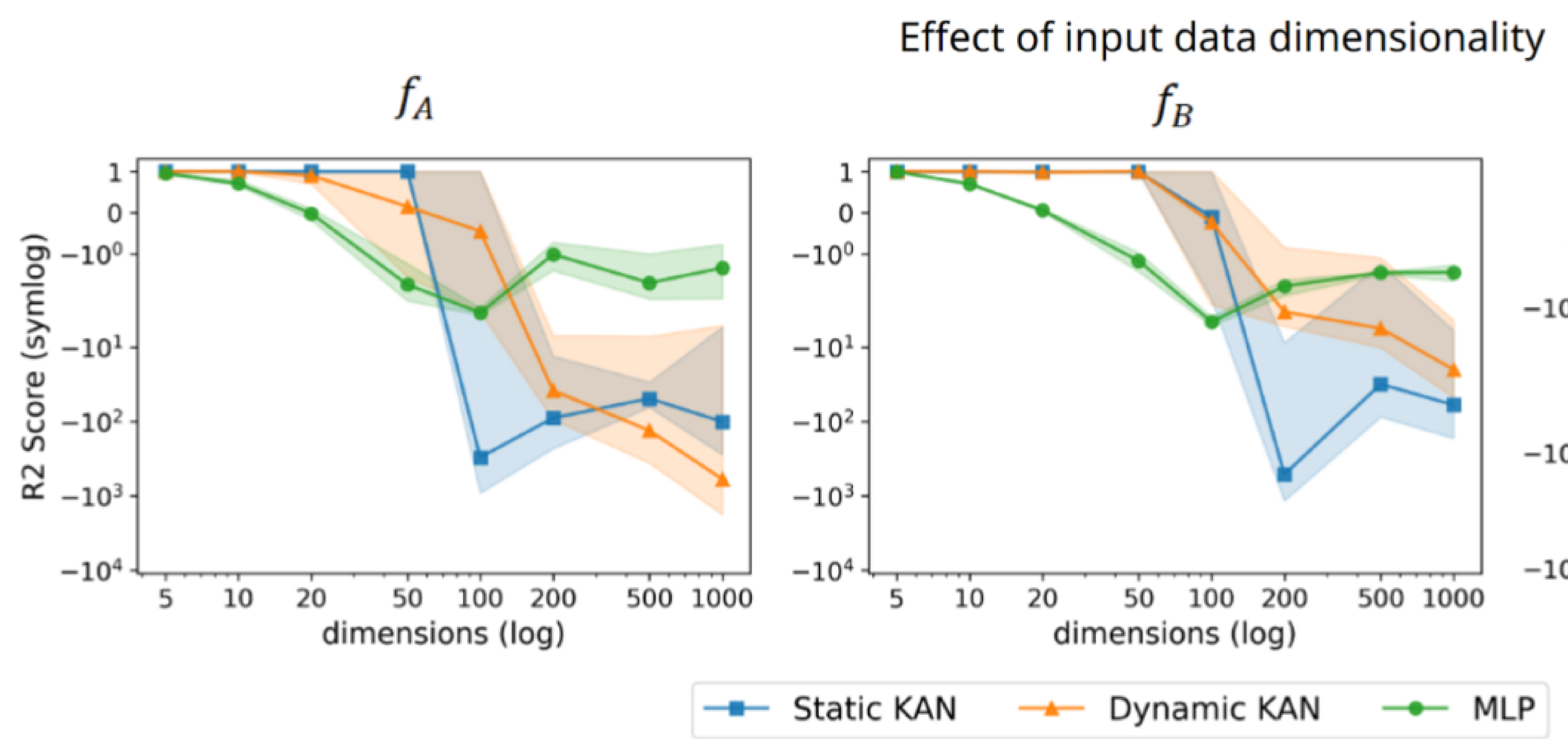
f_B: **Smooth & non-analytic.** Expectations: KAN excels because local splines; MLP because global approximation.

f_C: **Purely linear & simple.** Expectations: Baseline test; all models excel.

Train & Test against 3 different models:

- MLP
- Static KAN: learns functions at fixed grid size ("resolution")
- Dynamic KAN: learns functions with increasing grid size: coarsely then precisely
- ! Ensure **fairness** with matching parameter count

4. Effect of dimensionality



Low dimensions ($d \leq 10$):

- KANs superior accuracy for f_A, f_B
- ∵ Learnable splines directly fit inner functions' shapes

High dimensions ($d \geq 100$):

- KANs' accuracy degrades severely across all complexities
- ∵ Overfitting + Optimization (not architectural!) collapse

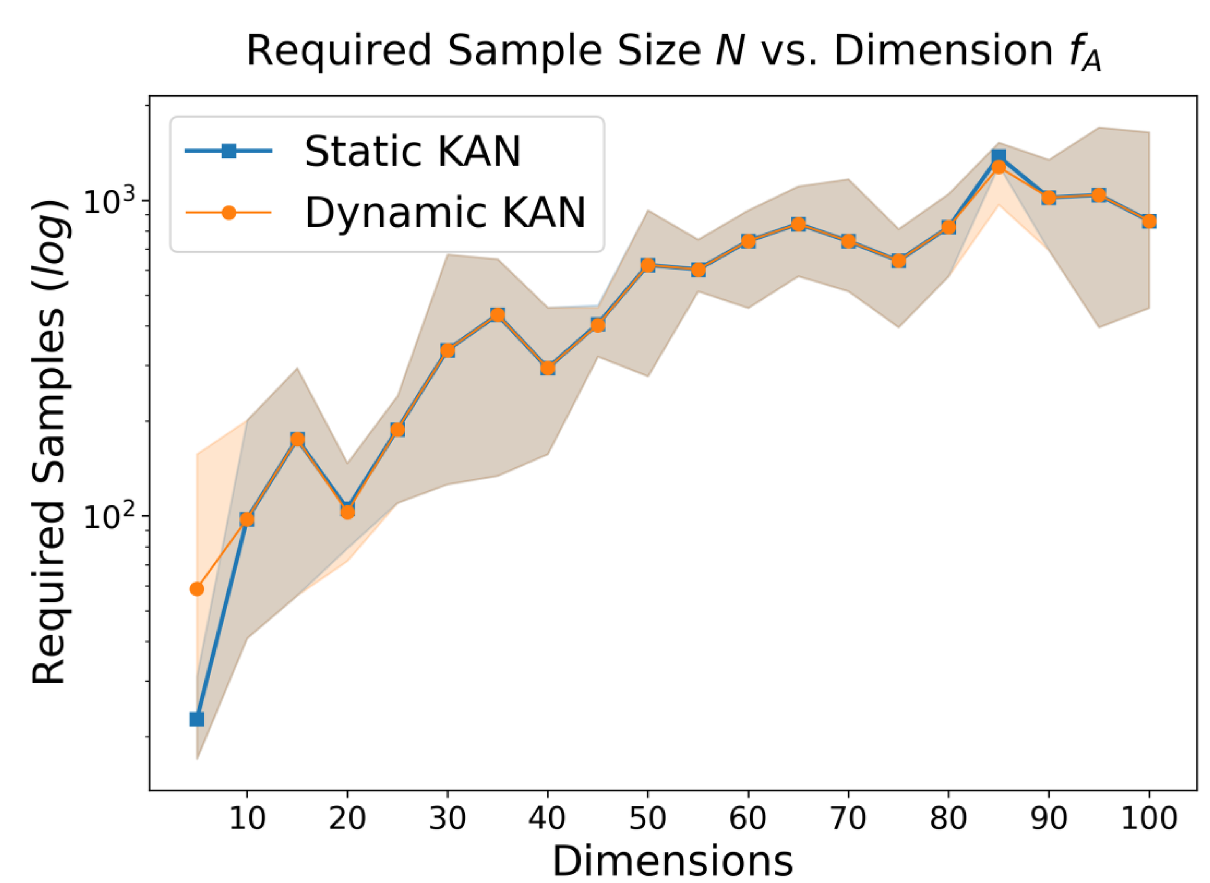
Potential "fixes":

- More frequent grid-updates (but high computational cost!)
- Hybrid optimizers (instead of current LBFGS optimizer)

5. Curse of dimensionality (CoD)?

Validate/Disprove "no CoD" claim - Given dimensionality d, with what data size N does KAN need to achieve a target R^2?

- Required N scales polynomially $O(d^x)$: no CoD!
- Required N scales exponentially $O(e^d)$ [2]: CoD!



- ? Sub-exponential growth: no CoD?
- ! Inconclusive! High variance + non-monotonous growth → Favorable initializations!
- ! Can be pre-asymptotic phase!

References:

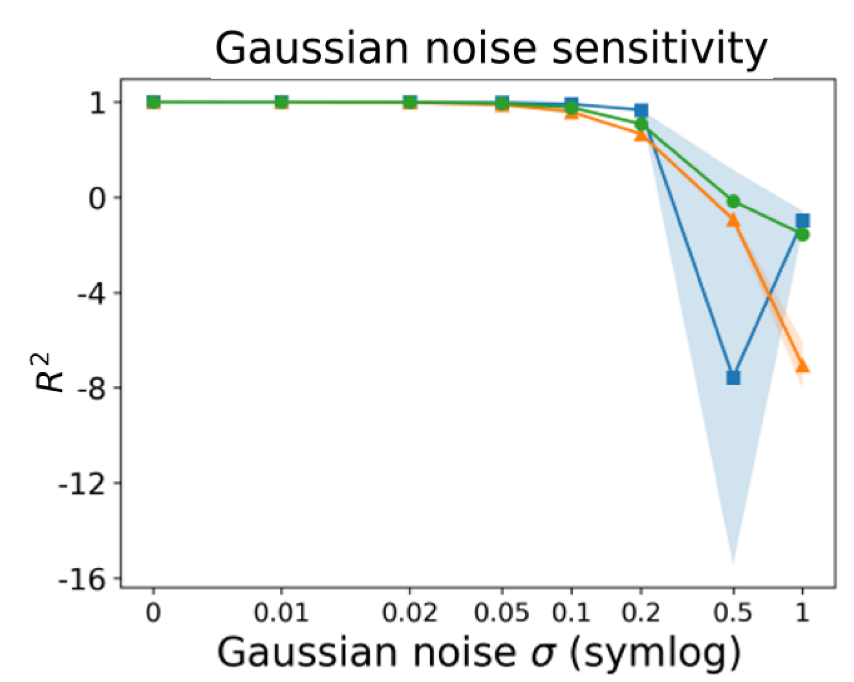
[1] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljacic, Thomas Y. Hou, and Max Tegmark. KAN: Kolmogorov-arnold networks. In *The Thirteenth International Conference on Learning Representations*, 2025.

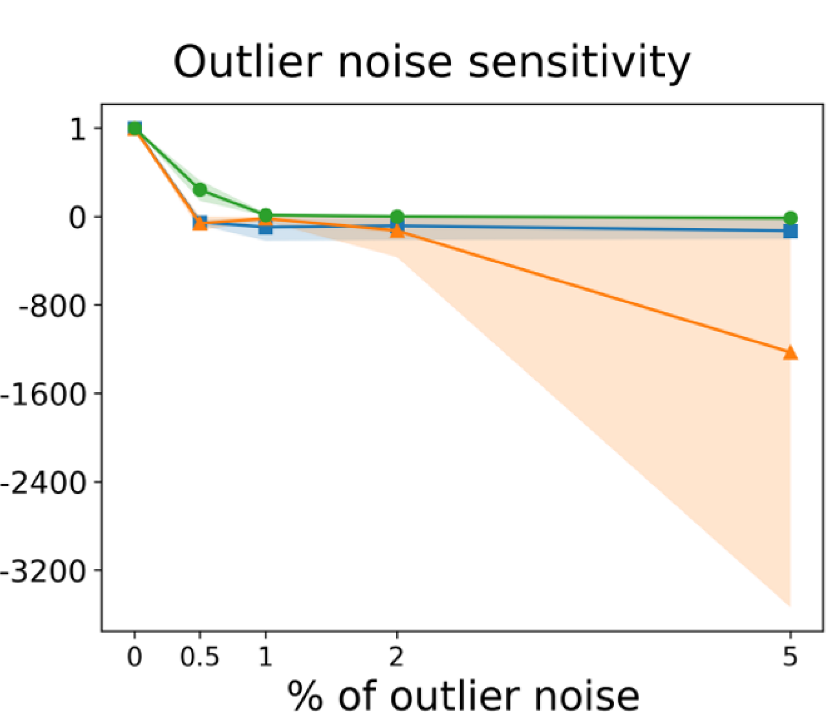
[2] Jiequn Han, Arnulf Jentzen, and Weinan E. Solving high-dimensional partial differential equations using deep learning. *Proc. Natl. Acad. Sci. U. S. A.*, 115(34):8505–8510, August 2018.

6. Effect of noise

Given $d = 100$, $f = f_C$ where all models perform competitively, what is the impact of:

- **Gaussian noise** of different σ (scaled by variance of function)?
- **Outlier noise** of different probability %?





- Accuracy smoothly degrades for Gaussian, extreme sensitivity to outliers
- ! MSE loss criterion heavily penalizes extremities
- ! Amplified in high dimension
- ! Dynamic KAN especially brittle: additional freedom spent on fitting noise, not signal!

7. Conclusion

KANs are neither a breakthrough nor a failure. They are **specialized tools** for **low-dimensional, clean** mathematical problems — **not yet a general-purpose** architecture for large-scale, noisy, real-world data.

runanhan@tudelft.nl