

KAN: Kolmogorov–Arnold Networks

Георгий Булгаков, 3 курс ФПМИ, кафедра «ПИР»
Данил Руденко, 3 курс ФПМИ, кафедра «ПИР»

Руководители:
Глеб Молодцов
Даниил Медяков,
МФТИ, Лаборатория ММО

Предпосылки

Теорема Колмогорова-Арнольда

Любая **непрерывная** функция многих переменных **на компакте**

$$f : [0,1]^n \rightarrow \mathbb{R}$$

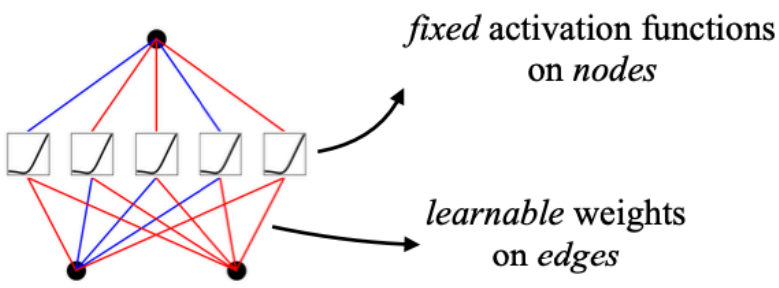
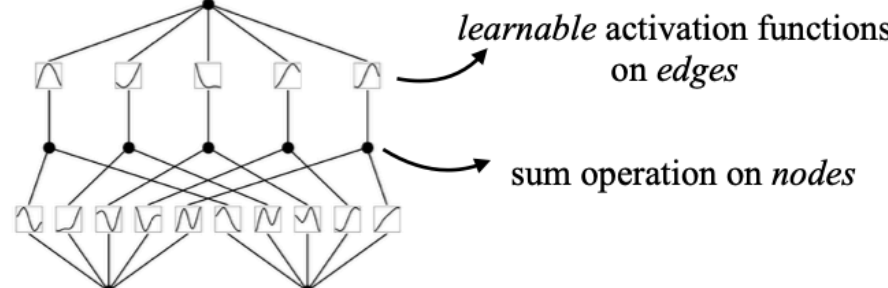
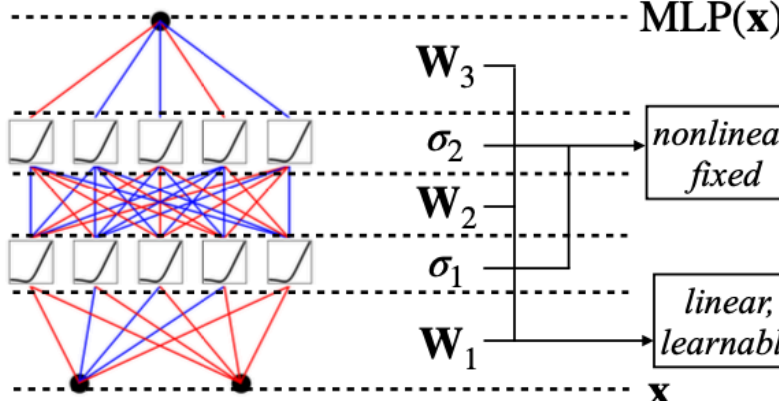
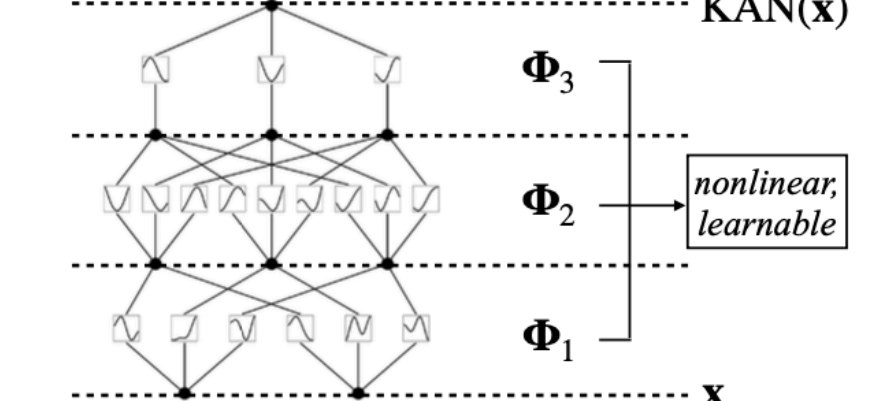
Может быть представлена в виде

$$f(\mathbf{x}) = f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \Phi_{q,p} (x_p) \right), \text{ причем } \textbf{непрерывны ф-ции}$$

$$\Phi_{q,p} : [0,1] \rightarrow \mathbb{R}, \quad \Phi_q : \mathbb{R} \rightarrow \mathbb{R} \quad \text{для } p = 1, \dots, n, \quad q = 1, \dots, 2n + 1$$

Архитектура KAN vs MLP

Слой - матрица $n_{out} \times n_{in}$ из одномерных сплайнов $\|\Phi_{q,p}\|$

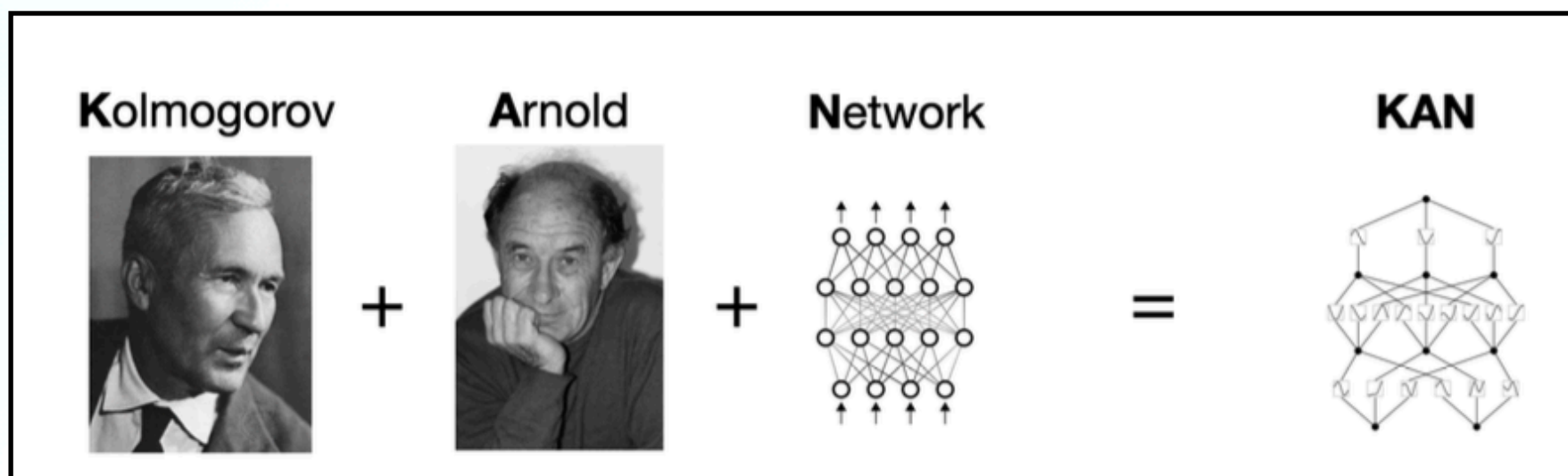
Model	Multi-Layer Perceptron (MLP)	Kolmogorov-Arnold Network (KAN)
Theorem	Universal Approximation Theorem	Kolmogorov-Arnold Representation Theorem
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(\epsilon)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right)$
Model (Shallow)	<p>(a)</p>  <p>fixed activation functions on nodes</p> <p>learnable weights on edges</p>	<p>(b)</p>  <p>learnable activation functions on edges</p> <p>sum operation on nodes</p>
Formula (Deep)	$\text{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$	$\text{KAN}(\mathbf{x}) = (\Phi_3 \circ \Phi_2 \circ \Phi_1)(\mathbf{x})$
Model (Deep)	<p>(c)</p>  <p>\mathbf{W}_3</p> <p>σ_2</p> <p>\mathbf{W}_2</p> <p>σ_1</p> <p>\mathbf{W}_1</p> <p>\mathbf{x}</p> <p>nonlinear, fixed</p> <p>linear, learnable</p>	<p>(d)</p>  <p>Φ_3</p> <p>Φ_2</p> <p>Φ_1</p> <p>\mathbf{x}</p> <p>nonlinear, learnable</p>

<https://arxiv.org/pdf/2404.19756>,

Ziming Li et al., KAN : Kolmogorov-Arnold Networks

Мотивация

- ❖ Internal & External Degrees of freedom
- ❖ Интерпретируемость
- ❖ Continual Learning
- ❖ Обучаемые функции активации
- ❖ Применение в науке!



Интерпретируемость

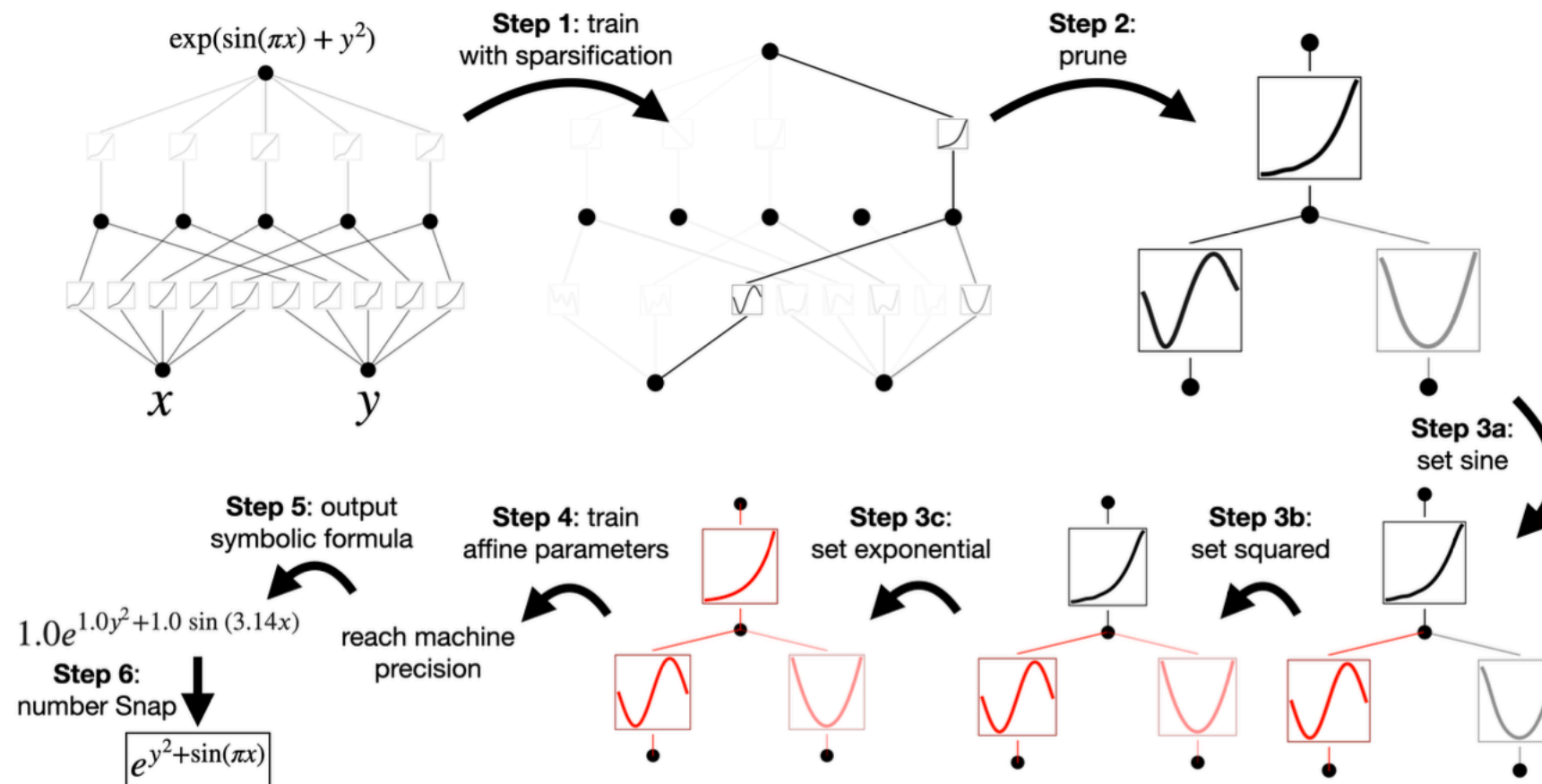
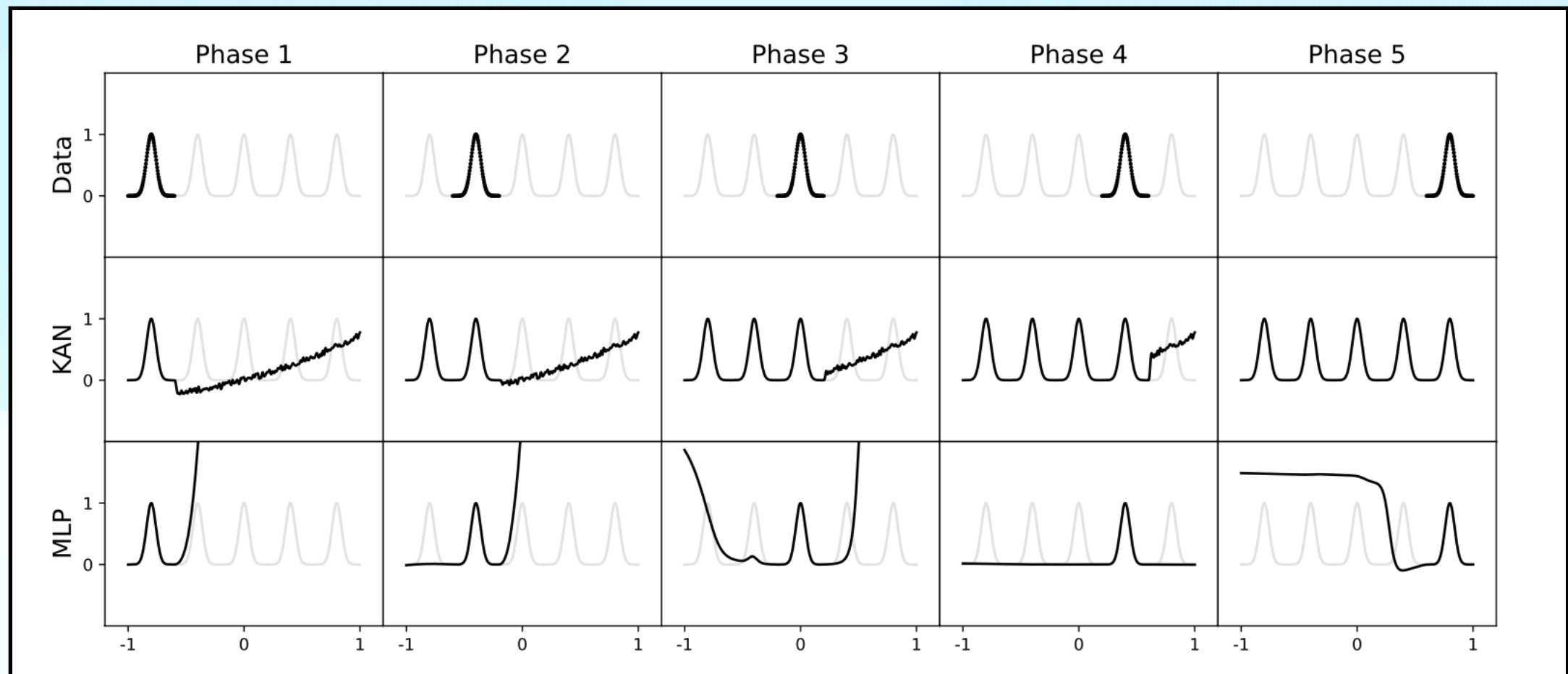


Figure 2.4: An example of how to do symbolic regression with KAN.

<https://arxiv.org/pdf/2404.19756>,

Ziming Li et al., KAN : Kolmogorov-Arnold Networks

Continual Learning



<https://arxiv.org/pdf/2404.19756>,

Ziming Li et al., KAN : Kolmogorov-Arnold Networks

Основное утверждение

Theorem 2.1 (Approximation theory, KAT). *Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$. Suppose that a function $f(\mathbf{x})$ admits a representation*

$$f = (\Phi_{L-1} \circ \Phi_{L-2} \circ \dots \circ \Phi_1 \circ \Phi_0)\mathbf{x}, \quad (2.14)$$

as in Eq. (2.7), where each one of the $\Phi_{l,i,j}$ are $(k+1)$ -times continuously differentiable. Then there exists a constant C depending on f and its representation, such that we have the following approximation bound in terms of the grid size G : there exist k -th order B-spline functions $\Phi_{l,i,j}^G$ such that for any $0 \leq m \leq k$, we have the bound

$$\|f - (\Phi_{L-1}^G \circ \Phi_{L-2}^G \circ \dots \circ \Phi_1^G \circ \Phi_0^G)\mathbf{x}\|_{C^m} \leq CG^{-k-1+m}. \quad (2.15)$$

Here we adopt the notation of C^m -norm measuring the magnitude of derivatives up to order m :

$$\|g\|_{C^m} = \max_{|\beta| \leq m} \sup_{x \in [0,1]^n} |D^\beta g(x)|.$$

<https://arxiv.org/pdf/2404.19756>,

Ziming Li et al., KAN : Kolmogorov-Arnold Networks

Neural scaling loss

Test RMSE: $l \sim N^{-\alpha}$

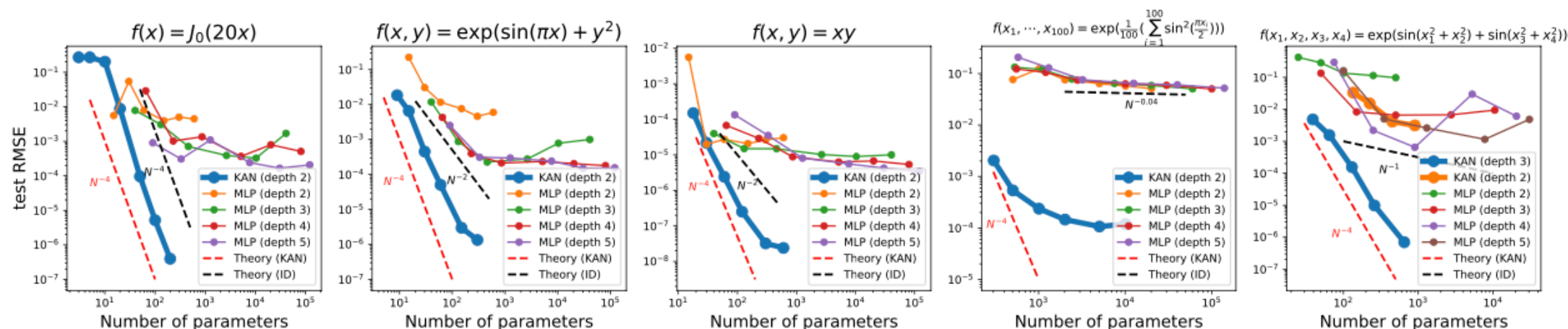
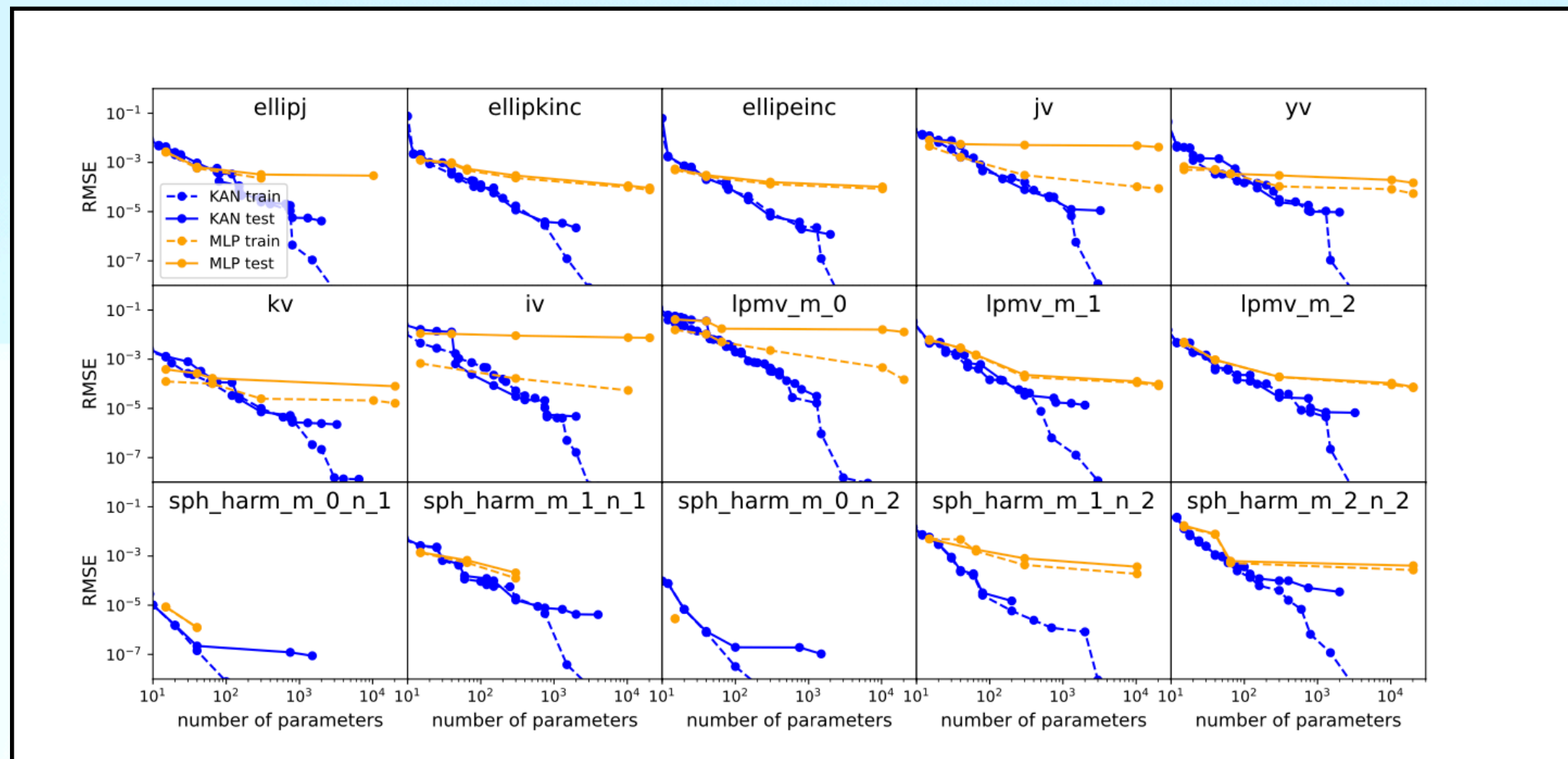


Figure 3.1: Compare KANs to MLPs on five toy examples. KANs can almost saturate the fastest scaling law predicted by our theory ($\alpha = 4$), while MLPs scales slowly and plateau quickly.

Применение в науке

Аппроксимация более сложных функций



Недостатки KAN

- * Не оптимизированы на GPU;
- * Применимость: тестировались только на малоразмерных задачах;
- * Медленное обучение.

Что уже сделано?

- ✓ Глубокое изучение Back Propagation;
- ✓ Подробное изучение модуля PyTorch;
- ✓ Изучение алгоритмов стох. оптимизации (SGD, SAGA, SVRG, SARAH);
- ✓ Изучение адаптивных алгоритмов (AdaGrad, RMSProp, Adam, AdamW);

Текущие задачи

- ➡ Анализ Ziming Li et al., KAN : Kolmogorov-Arnold Networks и экспериментов;
- ➡ Сравнить с MLP на других задачах;
- ➡ Проанализировать применение в **AI & Science**.

Планы и цели

- ➔ Реализовать KAN и воспроизвести эксперименты;
- ➔ Замена линейных слоев?;
- ➔ Применение в **Tabular DL**;
- ➔ Применение в графовых сетях

Заключение

- **KAN's породили новый подход!**
- KAN's - перспективная замена MLP;
- Будущие исследования могут значительно расширить их применение.

Спасибо за внимание!