Detailed Mathematical Description of KAN Networks

Liu Tianyi

May 19, 2025

1 Kolmogorov-Arnold Representation Theorem

Theorem 1.1 (Kolmogorov-Arnold (1957)). *For any continuous function $f : [0,1]^n \to \mathbb{R}$, there exist continuous univariate functions:*

- *Outer functions:* $\phi_q : \mathbb{R} \to \mathbb{R},\ q = 1, \ldots, 2n+1$
- *Inner functions:* $\psi_{q,p} : [0,1] \to \mathbb{R},\ q = 1, \ldots, 2n+1;\ p = 1, \ldots, n$

such that:

$$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \phi_q \left(\sum_{p=1}^{n} \psi_{q,p}(x_p) \right) \tag{1}$$

1.1 Constructive Proof Sketch

1. **Dimensionality reduction**: Define inner functions $\psi_{q,p}$ as:

   $$\psi_{q,p}(x_p) = \lambda_q x_p + \gamma_{q,p}(x_p)$$

   where λ_q are rationally independent constants and $\gamma_{q,p}$ are Lipschitz continuous.

2. **Channel-wise aggregation**: For each channel q, compute:

   $$z_q = \sum_{p=1}^{n} \psi_{q,p}(x_p)$$

   ensuring injectivity via the mapping $\Psi : \mathbb{R}^n \to \mathbb{R}^{2n+1}$.

3. **Outer composition**: Approximate f using:

   $$f(\mathbf{x}) \approx \sum_{q=1}^{2n+1} \phi_q(z_q)$$

   where ϕ_q are constructed via iterative approximation.

2 KAN Network Architecture

2.1 Parametrization of Univariate Functions

Using cubic B-splines with K basis functions:

$$\psi_{q,p}(x) = \sum_{k=1}^{K} c_{q,p,k} B_k(x) \tag{2}$$

where basis functions $B_k(x)$ satisfy:

$$\int_0^1 B_k(x) B_{k'}(x) dx = \delta_{kk'}$$

2.2 Layer-wise Computation

For layer l with input $\mathbf{x}^{(l)} \in \mathbb{R}^{d_l}$:

$$x_j^{(l+1)} = \sum_{q=1}^{Q_l} \phi_{j,q}^{(l)} \left(\sum_{i=1}^{d_l} \psi_{j,q,i}^{(l)}(x_i^{(l)}) \right), \quad j = 1, \ldots, d_{l+1} \tag{3}$$

where Q_l is the number of channels in layer l.

2.3 Gradient Computation

For parameter $c_{q,p,k}$ in (2):

$$\frac{\partial \mathcal{L}}{\partial c_{q,p,k}} = \sum_{i=1}^{N} \frac{\partial \mathcal{L}}{\partial f(\mathbf{x}_i)} \cdot \frac{\partial \phi_q}{\partial \psi_{q,p}} \cdot B_k(x_{i,p})$$

3 Theoretical Analysis

3.1 Approximation Error Bound

For $f \in C^m([0,1]^n)$ and cubic spline parametrization:

$$\|f - f_{\text{KAN}}\|_{L^\infty} \leq C_1 h^4 + C_2 K^{-m}$$

where h is the spline spacing and K is the number of basis functions.

4 Comparison with MLPs

4.1 MLP Universal Approximation

For MLP with ReLU activation:

$$f_{\text{MLP}}(\mathbf{x}) = \mathbf{W}_2 \sigma(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2$$

requires width $\geq O(\epsilon^{-n})$ for ϵ-approximation.

4.2 KAN Parameter Efficiency

KAN parameter complexity:

$$\mathcal{P}_{\text{KAN}} = O(nQK) \quad \text{vs} \quad \mathcal{P}_{\text{MLP}} = O(n^2 H)$$

where Q is number of channels and H is MLP hidden dimension.

5 Implementation Details

5.1 Spline Node Adaptation

Optimize node positions t_k via:

$$\min_{\{t_k\},\{c_{q,p,k}\}} \sum_{i=1}^{N} \left| f(\mathbf{x}_i) - \sum_{q=1}^{Q} \phi_q \left(\sum_{p=1}^{n} \sum_{k=1}^{K} c_{q,p,k} B_k(x_{i,p}; \{t_k\}) \right) \right|^2$$

5.2 Initialization Strategy

- Inner functions: $\psi_{q,p}(x) \approx x + \mathcal{N}(0, 0.01)$
- Outer functions: $\phi_q(z) \sim \mathcal{N}(0, \sigma_z^2)$