# Lecture 12: DNN Approximation - KST

Haizhao Yang

Department of Mathematics
University of Maryland College Park

2022 Summer Mini Course
Tianyuan Mathematical Center in Central China

Topic 3: Results by Kolmogorov-Arnold Representtion Theory

- Kolmogorov-Arnold representation theorem (KST):

**Theorem**
$\forall f(\mathbf{x}) \in C^0([0,1]^d)$, $\exists \psi_p(x)$ *and* $\phi(x)$ *defined on* $\mathbb{R}$ *such that*

$$f(\mathbf{x}) = \sum_{q=1}^{2d+1} a_q \phi(\sum_{p=1}^{d} b_{pq} \psi_p(x_p)).$$

But $\psi_p(x)$ and $\phi(x)$ are pathological.

# Kolmogorov-Arnold Representtion Theory

- Kolmogorov-Arnold representation theorem (KST):

**Theorem**

*$\forall f(\mathbf{x}) \in C^0([0,1]^d)$, $\exists \psi_p(x)$ and $\phi(x)$ defined on $\mathbb{R}$ such that*

$$f(\mathbf{x}) = \sum_{q=1}^{2d+1} a_q \phi(\sum_{p=1}^{d} b_{pq}\psi_p(x_p)).$$

But $\psi_p(x)$ and $\phi(x)$ are pathological.

- What if we use DNN to approximate $\psi_p(x)$ and $\phi(x)$?

# Kolmogorov-Arnold Representtion Theory

- Kolmogorov-Arnold representation theorem (KST):

**Theorem**
*$\forall f(\mathbf{x}) \in C^0([0,1]^d)$, $\exists \psi_p(x)$ and $\phi(x)$ defined on $\mathbb{R}$ such that*

$$f(\mathbf{x}) = \sum_{q=1}^{2d+1} a_q \phi(\sum_{p=1}^{d} b_{pq} \psi_p(x_p)).$$

But $\psi_p(x)$ and $\phi(x)$ are pathological.

- What if we use DNN to approximate $\psi_p(x)$ and $\phi(x)$?
- $\psi_p(x)$ and $\phi(x)$ might be exponentially bad in $d$.

# Key ideas

- J. Braun, M. Griebel, On a constructive proof of Kolmogorov's superposition theorem, Constr. Approx. 30 (2009) 653-675.
- Montanelli, Y., *Error bounds for deep ReLU networks using the Kolmogorov-Arnoldsuperposition theorem*, Neural Networks, 2020.
  - Apply ReLU in the constructive proof.
  - Check the moduli of continuity of inner and outer functions.

# Kolmogorov-Arnold Representtion Theory

Observation: Hölder continuity of the inner function with constants $(\nu, \alpha) = (O(\log d), \frac{1}{O(\log d)})$.

**Theorem (Approximation of the inner function)**

*Let $d \geq 2$ and $\psi$ be the inner function in KST. $\forall \epsilon \in (0,1)$, there is a ReLU DNN $\widetilde{\psi}$ that has a size*

$$W \leq c_1(d)\epsilon^{-[1+\log_2(d+1)]/2} \ll O(\epsilon^{-d}),$$

*such that $\|\psi - \widetilde{\psi}\|_{L^\infty} \leq \epsilon$, with a constant $c_1(d)$.*

# Kolmogorov-Arnold Representtion Theory

Observation:

$$\phi_j^r(x) = \frac{1}{m+1} \sum_{\ell=1}^{r} \sum_{\boldsymbol{d} \in (D_\ell)^d} e_{\ell-1}(\boldsymbol{d}) \mathcal{N}\mathcal{N}\left(\boldsymbol{d} + jO(d^{-2}); x\right),$$

**Theorem (Approximation of the outer functions)**

*Let $f : [0,1]^d \to \mathbb{R}$ be a continuous function, and $\phi_j^r$ be the $(2d+1)$ outer functions in KST at iteration $r$. $\forall \epsilon \in (0,1)$, there are ReLU DNNs $\widetilde{\phi}_j^r$ that have a size*

$$W \leq c_2(d,f)\epsilon^{-1/2},$$

*such that $\|\phi_j^r - \widetilde{\phi}_j^r\|_{L^\infty} \leq \epsilon$, with a constant $c_2(d,f)$.*

**Theorem (Approximation of continuous functions using KST)**
*Let $f : [0,1]^d \to \mathbb{R}$ be a continuous function. Then, for any scalar $0 < \epsilon < 1$, there is a ReLU DNN $\widetilde{f}_r$ that has a size*

$$W \leq (2d^2 + d)c_3(d,f)\epsilon^{-1+\log_2(d+1))]/2} + (2d+1)c_4(d,f)\epsilon^{-1/2},$$

*such that $\|f - \widetilde{f}_r\|_{L^\infty([0,1]^d)} \leq \epsilon$ with $c_3(d,f)$ and $c_4(d,f)$ as constants.*

Remark: unspecified constants may results in the curse of dimensionality.

# DNNs with advanced activation function

## Elementary universal activation function (EUAF)

A continuous activation function without gradient vanishing

$$\sigma_1(x) = \left| x - 2\lfloor \tfrac{x+1}{2} \rfloor \right|,$$

$$\sigma_2(x) := \frac{x}{|x| + 1},$$

$$\sigma(x) := \begin{cases} \sigma_1(x) & \text{for } x \in [0, \infty), \\ \sigma_2(x) & \text{for } x \in (-\infty, 0). \end{cases}$$

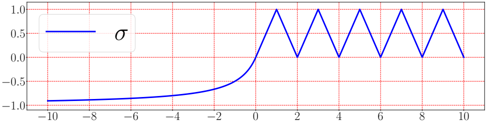



Figure: An illustration of $\sigma$ on $[-10, 10]$.

# DNNs with advanced activation function

### Theorem (EUAF approximation in $d$-dimensions)

*Arbitrarily small error with a fixed number of neurons for* $C([0,1]^d)$.

- *For any* $\epsilon > 0$, *there exists* $\phi$ *of width* $36d(2d+1)$ *and depth* $11$ *s.t.*

$$\|f(x) - \phi(x)\|_{L^\infty([0,1]^d)} \leq \epsilon$$

- *Shen, Y., and Zhang (arXiv:2107.02397)*

# DNNs with advanced activation function

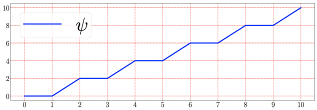
### Theorem (EUAF representation in $d$-dimensions)

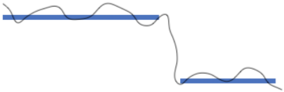
*Exact representation with a fixed number of neurons for classification functions.*

- *For any classification function* $f(x)$ *with* $K$ *classes, there exists* $\phi$ *of width* $36d(2d+1)$ *and depth* $12$ *s.t.*

$$f(x) = \phi(x)$$

  *on the supports of each class.*
- *Shen, Y., and Zhang (arXiv:2107.02397)*



# DNNs with advanced activation function

Two main ideas

- Theorem (Kolmogorov-Arnold Superposition Theorem)
  *$\forall f(\mathbf{x}) \in C([0,1]^d)$, there exist $\psi_p(x)$ and $\phi(x)$ in $C(\mathbb{R})$ and $b_{pq} \in \mathbb{R}$ s.t.*
  $$f(\mathbf{x}) = \sum_{q=1}^{2d+1} a_q \phi(\sum_{p=1}^{d} b_{pq} \psi_p(x_p)).$$

- Lemma (EUAF approximation in 1D (Shen, Y., and Zhang (arXiv:2107.02397))
  *NNs with width 36 and depth 5 constructed with EUAF is dense in $C([0,1])$.*

# DNNs with advanced activation function

EUAF is more powerful than bit extraction.

**Lemma (Curve filling in $K$-dimensions (Shen, Y., and Zhang (arXiv:2107.02397))**

*For any $K \in \mathbb{N}^+$, the following point set*

$$\left\{ \left[\sigma_1\left(\tfrac{w}{\pi+1}\right),\ \sigma_1\left(\tfrac{w}{\pi+2}\right),\ \cdots,\ \sigma_1\left(\tfrac{w}{\pi+K}\right)\right]^T : w \in \mathbb{R} \right\} \subseteq [0,1]^K$$

*is dense in $[0,1]^K$, where $\pi$ is the ratio of the circumference of a circle to its diameter.*

Proof.
Ideas:

- Transcendental number $+$ distinct rational numbers $\rightarrow$ rationally independent numbers
- Rationally independent numbers $+$ periodic functions $\rightarrow$ dense set in $[0,1]^K$

$\square$

For arbitrary $K$, NN with width 1 and depth 2 constructed with EAUF can fit $K$ points up to arbitrary accuracy.

# DNNs with advanced activation function

### Other EUAF

- $C^s$ EUAF
- Sigmod EUAF

# Topic 4: Low-Dimensional Structure

# Low-Dimensional Structure

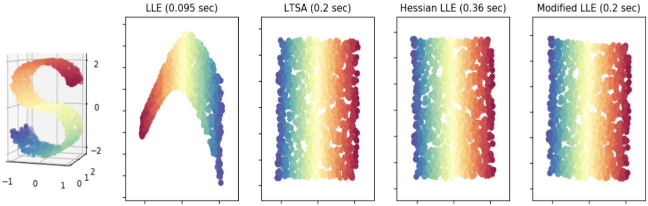



Figure: “Nearly" isometric maps.

# Low-Dimensional Structure

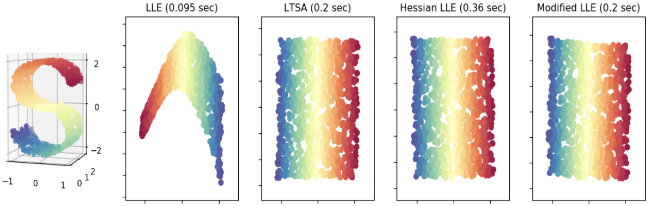



Figure: “Nearly" isometric maps.

## Theorem (Theorem 3.1 of Baraniuk et al 2009)

*Let $\mathcal{M}$ be a compact $d_{\mathcal{M}}$-dimensional Riemannian submanifold of $\mathbb{R}^d$. Fix $\delta \in (0,1)$ and $\gamma \in (0,1)$. Let $\boldsymbol{A} = \sqrt{\frac{d}{d_\delta}}\Phi$, where $\Phi \in \mathbb{R}^{d_\delta \times d}$ is a random orthoprojector with*

$$d_\delta = \mathcal{O}\left(\frac{d_{\mathcal{M}} \ln(d\delta^{-1}) \ln(1/\gamma)}{\delta^2}\right).$$

*If $d_\delta \leq d$, then with probability at least $1-\gamma$, the following statement holds: For every $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{M}$,*

$$(1-\delta)|\boldsymbol{x}_1 - \boldsymbol{x}_2| \leq |\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{A}\boldsymbol{x}_2| \leq (1+\delta)|\boldsymbol{x}_1 - \boldsymbol{x}_2|.$$

# Continuous functions

Idea: data in $\mathbb{R}^d \xrightarrow[\text{projection}]{\text{random}}$ data in $\mathbb{R}^{d_\delta} \xrightarrow[\text{network}]{\text{neural}}$ function values

**Theorem (Shen, Y., Zhang, CiCP, 2020)**

*$\mathcal{M}$ is a $d_{\mathcal{M}}$-manifold in $[0,1]^d$. For any $L, N \in \mathbb{N}^+$, let $f : \mathcal{M}_\epsilon \to \mathbb{R}$ be a Lip. cont. function with constant 1, $p \in [1, \infty)$, then there exists $\phi \in \mathcal{NN}(\#\text{input} = 1;\ \text{maxwidth} \leq \max\{8d_\delta \lfloor N^{1/d_\delta} \rfloor + 4d_\delta,\ 12N + 14\};\ \#\text{layer} \leq 9L + 12)$ such that*

$$\|f - \phi\|_{L^p([0,1]^d, \mu_\varrho)} \leq 3\frac{4\epsilon}{1-\delta}\sqrt{\frac{d}{d_\delta}} + 5\frac{16d}{(1-\delta)\sqrt{d_\delta}} N^{-2/d_\delta} L^{-2/d_\delta}.$$