Universal Joint Approximation of Manifolds and Densities by Simple Injective Flows

Michael Puthawala[1]

Maarten de Hoop[1], Matti Lassas[2], Ivan Dokmanić[3]

[1]Rice University, [2]University of Helsinki, [3]University of Basel

Network Definition

If for $\ell = 1, \dots, L$, $n_\ell \in \mathbb{N}$,

1. $\mathcal{T}_\ell^{n_\ell} \subset C(\mathbb{R}^{n_\ell}, \mathbb{R}^{n_\ell})$ is a flow network,
2. $\mathcal{R}_\ell^{n_{\ell-1}, n_\ell} \subset C(\mathbb{R}^{n_{\ell-1}}, \mathbb{R}^{n_\ell})$ is an injective ReLU layer,

then

$$\mathcal{E} = \mathcal{T}_L^{n_L} \circ \mathcal{R}_L^{n_{L-1}, n_L} \circ \cdots \circ \mathcal{T}_1^{n_1} \circ \mathcal{R}_1^{n_0, n_1} \circ \mathcal{T}_0^{n_0}$$

is always a family of injective mappings, where
$\mathcal{H} \circ \mathcal{G} \coloneqq \{h \circ g \colon h \in \mathcal{H}, g \in \mathcal{G}\}$.

Network Definition

If for $\ell = 1, \ldots, L$, $n_\ell \in \mathbb{N}$,

1. $\mathcal{T}_\ell^{n_\ell} \subset C(\mathbb{R}^{n_\ell}, \mathbb{R}^{n_\ell})$ is a flow network,
2. $\mathcal{R}_\ell^{n_{\ell-1}, n_\ell} \subset C(\mathbb{R}^{n_{\ell-1}}, \mathbb{R}^{n_\ell})$ is an injective ReLU layer,

then

$$\mathcal{E} = \mathcal{T}_L^{n_L} \circ \mathcal{R}_L^{n_{L-1}, n_L} \circ \cdots \circ \mathcal{T}_1^{n_1} \circ \mathcal{R}_1^{n_0, n_1} \circ \mathcal{T}_0^{n_0}$$

is always a family of injective mappings, where
$\mathcal{H} \circ \mathcal{G} := \{h \circ g \colon h \in \mathcal{H}, g \in \mathcal{G}\}$.
When are these networks universal approximators?

Embedding Gap

We call a function f an embedding and denote if by $f \in \text{emb}(X, Y)$ if $f : X \to Y$ is continuous, injective, and $f^{-1} \colon f(X) \to X$ is continuous.

Definition (Embedding Gap)

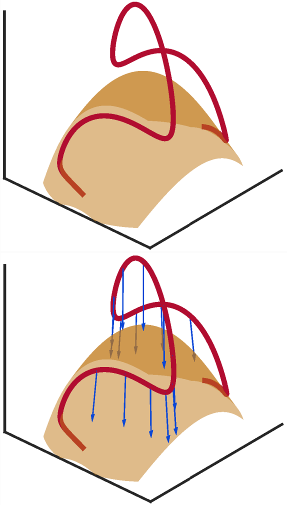
If,

- $K \subset \mathbb{R}^n$ and $W \subset \mathbb{R}^o$, both compact,
- $f \in \text{emb}(K, \mathbb{R}^m)$, and $g \in \text{emb}(W, \mathbb{R}^m)$

then we define

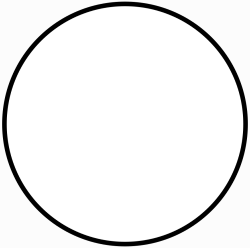
$$B_{K,W}(f, g) = \inf_{r \in \text{emb}(f(K), g(W))} \|I - r\|_{L^\infty(f(K))}$$

where $I \colon f(K) \to f(K)$ is the identity function.

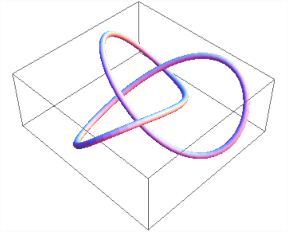


Non-approximable single-chart manifold

Let $K = S^1$ be a circle, and $f \in \mathrm{emb}(K, \mathbb{R}^3)$ an embedding of a trefoil knot into $\mathbb{R}^3$. There are no $E \in \mathcal{E} := \mathcal{T} \circ \mathcal{R}$ so that $E(K) = f(K)$.



(a) $K = S^1$



(b) Trefoil knot embedded in $\mathbb{R}^3$.

The trivial and trefoil knots are not equivalent.

Extendable Embeddings

Definition (Extendable Embedding)

With the above topological difficulty in mind, we define the set of extendable embeddings as

$$\mathcal{I}(\mathbb{R}^n, \mathbb{R}^m) := \{\Phi \circ R \in C(\mathbb{R}^n, \mathbb{R}^m) \colon R \in C(\mathbb{R}^n, \mathbb{R}^m), \Phi \in \mathbb{R}^m \to \mathbb{R}^m\}.$$

where R is linear full-rank, and ϕ is a C^1-smooth diffeomorphism.

Theorem (P. et al. 2022)

When $m \geq 3n+1$ and $k \geq 1$, for any C^k embedding $f \in \mathrm{emb}^k(\mathbb{R}^n, \mathbb{R}^m)$ and compact set $K \subset \mathbb{R}^n$, there is a map in the closures of the flow type neural network $E \in \mathcal{I}^k(\mathbb{R}^n, \mathbb{R}^m)$ such that $E(K) = f(K)$. Moreover,

$$\mathcal{I}^k(K, \mathbb{R}^m) = \mathrm{emb}^k(K, \mathbb{R}^m)$$

Manifold Universality

Let $\mathcal{F} = \text{emb}(K, \mathbb{R}^m)$, or $\mathcal{F} = \mathcal{I}(K, \mathbb{R}^m)$.

Theorem (P. et al. 2022)

Let $\mu \in \mathcal{P}(K)$ be an absolutely continuous measure w.r.t. Lebesgue measure and

1. $\mathcal{R}_\ell^{n_{\ell-1}, n_\ell}$ *is injective,*
2. $\mathcal{T}_\ell^{n_\ell}$ *is injective, universal approximator of diffeomorphisms,*
3. $\mathcal{T}_0^{n}$ *is distributionally universal and injective*

Then, there is a sequence of $\{E_i\}_{i=1,\ldots,\infty} \subset \mathcal{E} := \mathcal{T}_L^{n_L} \circ \mathcal{R}_L^{n_{L-1}, n_L} \circ \cdots \circ \mathcal{R}_1^{n_0, n_1} \circ \mathcal{T}_0^{n_0}$ *such that*

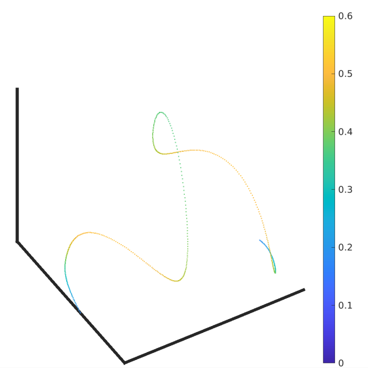
$$\lim_{i \to \infty} \mathrm{W}_2 \left(F_{\#}\mu, E_{i\#}\mu \right) = 0.$$

Decoupling

Optimality of layers of these deep neural networks can be established layer-by-layer.

Consider the problem of learning $\nu = f_{\#}\mu$ the following 1 dimensional distribution embedded in $\mathbb{R}^3$ with a network of the form

$$F_\theta(x) = T_2 \circ R_2 \circ T_1 \circ R_1 \circ T_0(x)$$

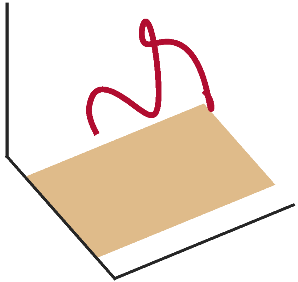


Decoupling: Step one

First, we can update $T_2 \circ R_2$ to decrease

$$B_{K,W}(f, T_2 \circ R_2)$$

$$F_\theta(x) = T_2 \circ R_2 \circ T_1 \circ R_1 \circ T_0(x)$$

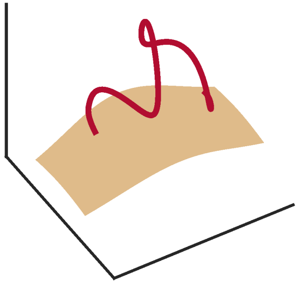


Decoupling: Step one

First, we can update $T_2 \circ R_2$ to decrease

$$B_{K,W}(f, T_2 \circ R_2)$$

$$F_\theta(x) = T_2 \circ R_2 \circ T_1 \circ R_1 \circ T_0(x)$$

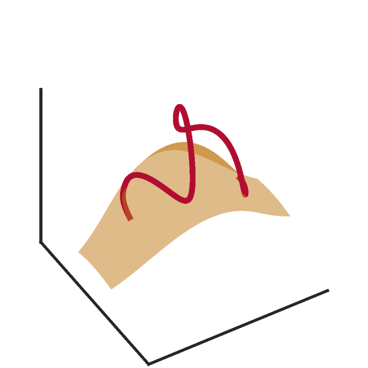


Decoupling: Step one

First, we can update $T_2 \circ R_2$ to decrease

$$B_{K,W}(f, T_2 \circ R_2)$$

$$F_\theta(x) = T_2 \circ R_2 \circ T_1 \circ R_1 \circ T_0(x)$$

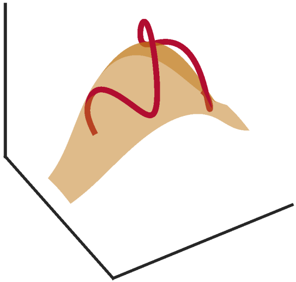


Decoupling: Step one

First, we can update $T_2 \circ R_2$ to decrease

$$B_{K,W}(f, T_2 \circ R_2)$$

$$F_\theta(x) = T_2 \circ R_2 \circ T_1 \circ R_1 \circ T_0(x)$$

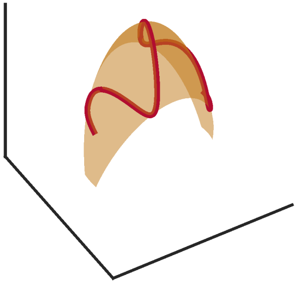


Decoupling: Step one

First, we can update $T_2 \circ R_2$ to decrease

$$B_{K,W}(f, T_2 \circ R_2)$$

$$F_\theta(x) = T_2 \circ R_2 \circ T_1 \circ R_1 \circ T_0(x)$$

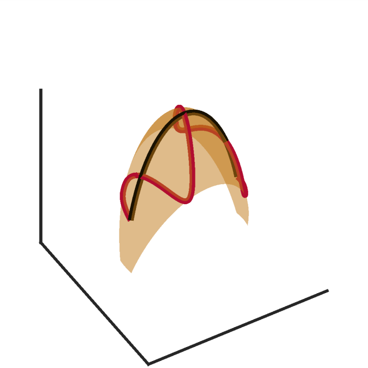


Decoupling: Step two

Now fix $T_2 \circ R_2$ and update $T_1 \circ R_1$ to decrease

$$B_{K,W}(f, T_2 \circ R_2 \circ T_1 \circ R_1)$$

$$F_\theta(x) = T_2 \circ R_2 \circ T_1 \circ R_1 \circ T_0(x)$$

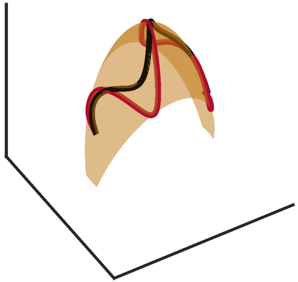


Decoupling: Step two

Now fix $T_2 \circ R_2$ and update $T_1 \circ R_1$ to decrease

$$B_{K,W}(f, T_2 \circ R_2 \circ T_1 \circ R_1)$$

$$F_\theta(x) = T_2 \circ R_2 \circ T_1 \circ R_1 \circ T_0(x)$$

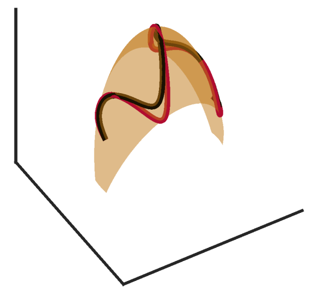


Decoupling: Step two

Now fix $T_2 \circ R_2$ and update $T_1 \circ R_1$ to decrease

$$B_{K,W}(f, T_2 \circ R_2 \circ T_1 \circ R_1)$$

$$F_\theta(x) = T_2 \circ R_2 \circ T_1 \circ R_1 \circ T_0(x)$$

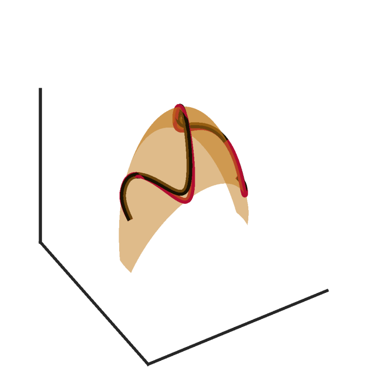


Decoupling: Step two

Now fix $T_2 \circ R_2$ and update $T_1 \circ R_1$ to decrease

$$B_{K,W}(f, T_2 \circ R_2 \circ T_1 \circ R_1)$$

$$F_\theta(x) = T_2 \circ R_2 \circ T_1 \circ R_1 \circ T_0(x)$$

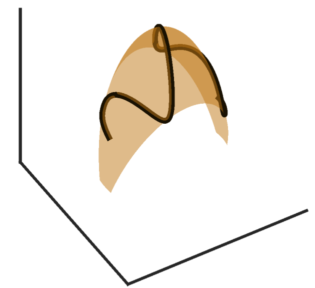


Decoupling: Step two

Now fix $T_2 \circ R_2$ and update $T_1 \circ R_1$ to decrease

$$B_{K,W}(f, T_2 \circ R_2 \circ T_1 \circ R_1)$$

$$F_\theta(x) = T_2 \circ R_2 \circ T_1 \circ R_1 \circ T_0(x)$$



Universal Joint Approximation of Manifolds and Densities by Simple Injective Flows

Michael Puthawala[1]

Maarten de Hoop[1], Matti Lassas[2], Ivan Dokmanić[3]

[1]Rice University, [2]University of Helsinki, [3]University of Basel