Mapping properties of neural networks

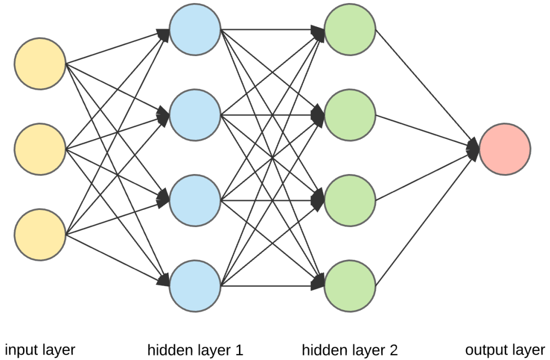
Matti Lassas

in collaboration with
Maarten de Hoop
Ivan Dokmanic
Takashi Furuya
Konik Kothari
Michael Puthawala

Finnish Centre of Excellence in Inverse Modelling and Imaging 2018-2025
fame MATHEMATICS PHYSICS COMPUTING
UNIVERSITY OF HELSINKI
Funded by the European Union
European Research Council Established by the European Commission

Outline:

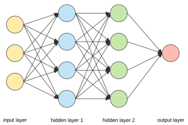
- Neural networks and the manifold hypothesis
- Networks with injective layers
- Injective networks and approximation of manifolds
- Flow networks and approximation of probability distributions
- Applications in inverse problems
- Neural operators



Deep neural network with relu-functions

A deep neural network (DNN) is a function $f_\theta : \mathbb{R}^{d_0} \to \mathbb{R}^{d_{L+1}}$ defined by

$$
\begin{aligned}
y_0 &= x, \\
y_{\ell+1} &= \phi\left(W_\theta^\ell y_\ell + b_\theta^\ell\right), \quad \ell = 0, 1, \dots, L-1, \\
f_\theta(x) &= W_\theta^L y_L + b_\theta^L.
\end{aligned}
$$



We denote the set of such functions f_θ by $\mathcal{N}\mathcal{N}(d_0, d_{L+1})$. Above,

- $\ell = 0, 1, 2, \dots, L$: the layer index.
- $y_\ell \in \mathbb{R}^{d_\ell}$: the state at layer ℓ.
- Weight matrixes $W_\theta^\ell \in \mathbb{R}^{d_{\ell+1} \times d_\ell}$ and bias vectors $b_\theta^\ell \in \mathbb{R}^{d_{\ell+1}}$
- ϕ is the activation function, the Rectified Linear Unit (relu)

$$\phi : \mathbb{R}^d \to \mathbb{R}^d, \quad \phi(x_1, \dots, x_d) = (\max(0, x_1), \dots, \max(0, x_d))$$

Basic mapping properties of neural networks

We consider the following topics

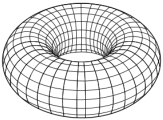
- Examples of injective neural networks $f_\theta : \mathbb{R}^n \to \mathbb{R}^m$ that are one-to-one, that is,

$$f_\theta(z_1) = f_\theta(z_2) \implies z_1 = z_2$$

- Let $K \subset \mathbb{R}^n$ be a compact set and $f_\theta : \mathbb{R}^n \to \mathbb{R}^m$ be injective (and continuous). Then,

$$f_\theta : K \to f_\theta(K)$$

  is a homeomorphism, and hence, K and $f_\theta(K)$ have the same topology.

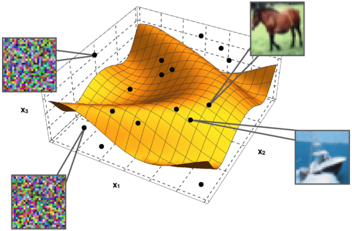


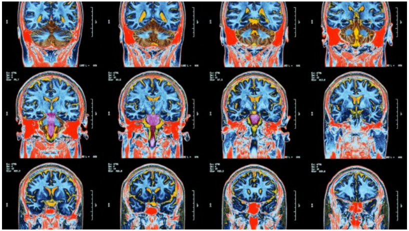
$\xrightarrow{\mathbf{f}_\theta}$

Manifold hypothesis and a generative model

Let $M \subset \mathbb{R}^m$ be a d-dimensional surface (that is, a manifold) and $\mu_0 = \mu_{data}$ be a probability distribution supported on M.
Let us assume that the topology of M is known and let $K \subset \mathbb{R}^n$ be a 'model space' that has the same topology as M. Let $Z \sim \text{Unif}(K)$ be a random variable. Here, $d \leq n < m$. Our aims are to

1. Find a neural network $f_\theta : \mathbb{R}^n \to \mathbb{R}^m$ such that $f_\theta : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one and $f_\theta(K) = M$ (or $f_\theta(K) \approx M$).
2. Find $f_\theta : \mathbb{R}^n \to \mathbb{R}^m$ such that $f_\theta(Z)$ has the distribution μ_0.



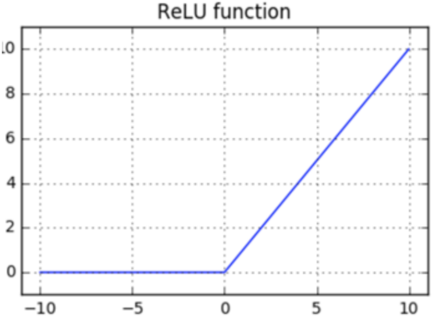


Figures: Goldt et al Phys. Rev. X 2020 and Bukhari et al Sci. Rep. 2022.

Outline:

- Neural networks and the manifold hypothesis
- Networks with injective layers
- Injective networks and approximation of manifolds
- Flow networks and approximation of probability distributions
- Applications in inverse problems

$$\phi(x) = \mathrm{relu}(x) = \begin{cases} x, & x > 0, \\ 0, & x \le 0, \end{cases} \qquad \text{for } x \in \mathbb{R}.$$



Injective of layer of a relu-neural network

Consider function $N : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$N(x) = \phi\,(Wx + b)$$

where

- $W \in \mathbb{R}^{m \times n}$ is a weight matrix and $b \in \mathbb{R}^n$ is a bias vector,
- ϕ is the Rectified Linear Unit (relu)

$$\begin{aligned} &\phi : \mathbb{R}^m \to \mathbb{R}^m, \\ &\phi(x_1, \dots, x_m) = (\max(0, x_1), \dots, \max(0, x_m)) = \max(x, 0). \end{aligned}$$

Examples of injective layers

1. The map $N : \mathbb{R}^n \to \mathbb{R}^{2n}$,

$$N(x) = \begin{bmatrix} \phi(x) \\ \phi(-x) \end{bmatrix} = \begin{bmatrix} \max(x, 0) \\ \max(-x, 0) \end{bmatrix}$$

   is injective.

2. Let $W \in \mathbb{R}^{2n \times n}$ be

$$W = \begin{bmatrix} B \\ -B \end{bmatrix}$$

   where $B \in \mathbb{R}^{n \times n}$ is an invertible matrix. Then

$$N : \mathbb{R}^n \to \mathbb{R}^{2n}$$
$$N(x) = \phi(Wx)$$

   is injective.

Conditions for injectivity

Theorem (Conditions for injectivity of relu(Wx))

Let $W \in \mathbb{R}^{m\times n}$ be a matrix with row vectors $\mathcal{W} = \{w_j\}_{j=1}^m$, and $\mathrm{relu}(y) = \max(y,0)$*. The function* $\mathrm{relu}(W(\cdot))\colon \mathbb{R}^n \to \mathbb{R}^m$ *is injective if and only if*

$$\mathit{span}\{w \in \mathcal{W} : \ w\cdot x \geq 0\} = \mathbb{R}^n \quad \text{for all } x \in \mathbb{R}^n \ . \tag{1}$$

If W is injective, then $m \geq 2n$.

Theorem (Conditions for injectivity of relu($Wx + b$))

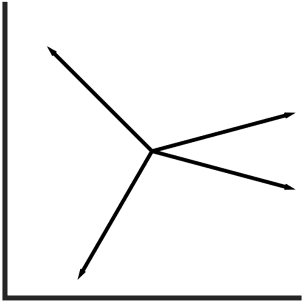
Let $W \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $\mathcal{W} = \{w_j\}_{j=1}^m$ be the row vectors of W. The function $\mathrm{relu}(W(\cdot) + b)\colon \mathbb{R}^n \to \mathbb{R}^m$ *is injective if and only if* $\{w_j \in \mathcal{W} : \ b_j \geq 0, \ w_j \cdot x \geq 0\}$ *spans $\mathbb{R}^n$ for every $x \in \mathbb{R}^n$.*

Reference: [Puthawala-Kothari-L.-Dokmanic-de Hoop JMLR 2022]

Example with a specific weight matrix W

Consider $N : \mathbb{R}^2 \to \mathbb{R}^4$, $N(x) = \mathrm{relu}(Wx)$, where

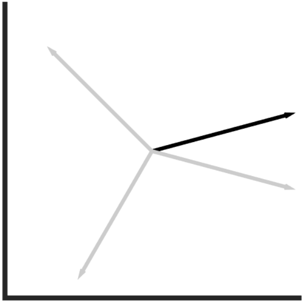
$$W = \begin{bmatrix} \cos(\pi/8) & \sin(\pi/8) \\ \cos(5\pi/8) & \sin(5\pi/8) \\ \cos(-5\pi/8) & \sin(-5\pi/8) \\ \cos(-\pi/8) & \sin(-\pi/8) \end{bmatrix}$$



Example with a specific weight matrix W

Consider $N : \mathbb{R}^2 \to \mathbb{R}^4$, $N(x) = \text{relu}(Wx)$, where

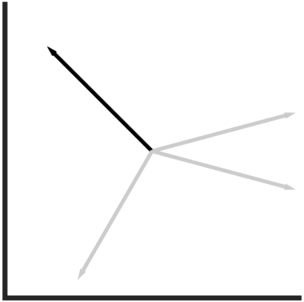
$$W = \begin{bmatrix} \color{red}{\cos(\pi/8)} & \color{red}{\sin(\pi/8)} \\ \cos(5\pi/8) & \sin(5\pi/8) \\ \cos(-5\pi/8) & \sin(-5\pi/8) \\ \cos(-\pi/8) & \sin(-\pi/8) \end{bmatrix}$$



Example with a specific weight matrix W

Consider $N : \mathbb{R}^2 \to \mathbb{R}^4$, $N(x) = \text{relu}(Wx)$, where

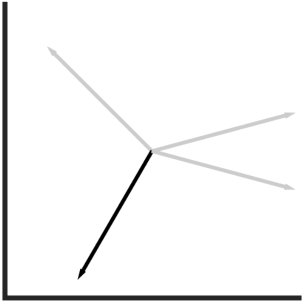
$$W = \begin{bmatrix} \cos(\pi/8) & \sin(\pi/8) \\ \cos(5\pi/8) & \sin(5\pi/8) \\ \cos(-5\pi/8) & \sin(-5\pi/8) \\ \cos(-\pi/8) & \sin(-\pi/8) \end{bmatrix}$$



Example with a specific weight matrix W

Consider $N : \mathbb{R}^2 \to \mathbb{R}^4$, $N(x) = \mathrm{relu}(Wx)$, where

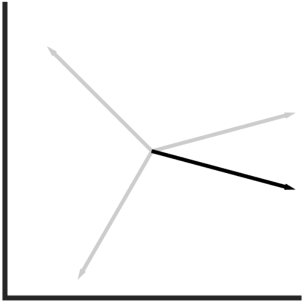
$$W = \begin{bmatrix} \cos(\pi/8) & \sin(\pi/8) \\ \cos(5\pi/8) & \sin(5\pi/8) \\ \cos(-5\pi/8) & \sin(-5\pi/8) \\ \cos(-\pi/8) & \sin(-\pi/8) \end{bmatrix}$$



Example with a specific weight matrix W

Consider $N : \mathbb{R}^2 \to \mathbb{R}^4$, $N(x) = \text{relu}(Wx)$, where

$$W = \begin{bmatrix} \cos(\pi/8) & \sin(\pi/8) \\ \cos(5\pi/8) & \sin(5\pi/8) \\ \cos(-5\pi/8) & \sin(-5\pi/8) \\ \color{red}{\cos(-\pi/8)} & \color{red}{\sin(-\pi/8)} \end{bmatrix}$$

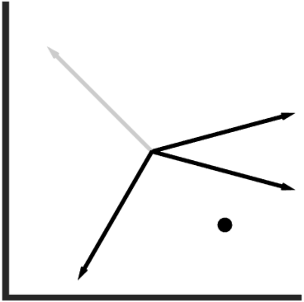


Spanning condition for the row vectors w_j of W

Let $N(x) = \text{relu}(Wx)$, where

$$W = \begin{bmatrix} \cos(\pi/8) & \sin(\pi/8) \\ \cos(5\pi/8) & \sin(5\pi/8) \\ \cos(-5\pi/8) & \sin(-5\pi/8) \\ \cos(-\pi/8) & \sin(-\pi/8) \end{bmatrix}$$

The dot is x, the black vectors are $\mathcal{W}_x = \{w_j : \ w_j \cdot x \geq 0\}$

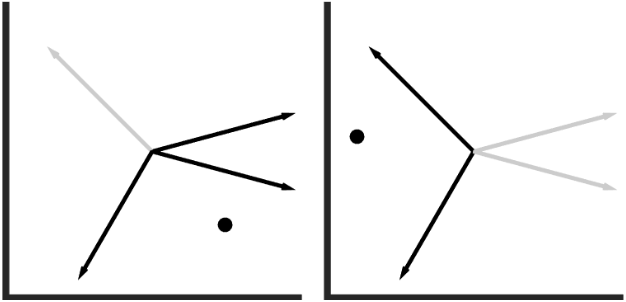


Spanning condition for the row vectors w_j of W

Let $N(x) = \text{relu}(Wx)$, where

$$W = \begin{bmatrix} \cos(\pi/8) & \sin(\pi/8) \\ \cos(5\pi/8) & \sin(5\pi/8) \\ \cos(-5\pi/8) & \sin(-5\pi/8) \\ \cos(-\pi/8) & \sin(-\pi/8) \end{bmatrix}$$

The dot is x, the black vectors are $\mathcal{W}_x = \{w_j : \; w_j \cdot x \geq 0\}$

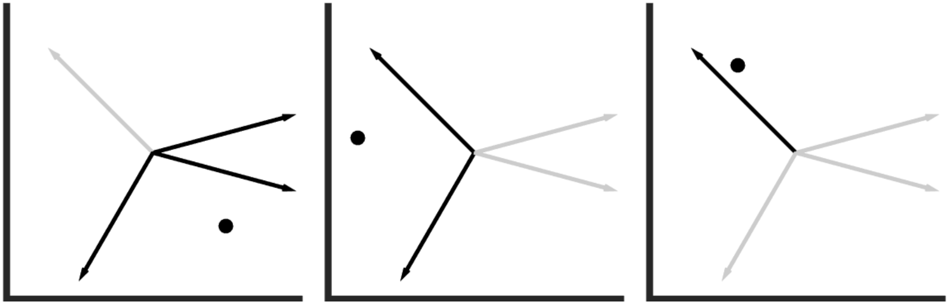


Spanning condition for the row vectors w_j of W

Let $N(x) = \text{relu}(Wx)$, where

$$W = \begin{bmatrix} \cos(\pi/8) & \sin(\pi/8) \\ \cos(5\pi/8) & \sin(5\pi/8) \\ \cos(-5\pi/8) & \sin(-5\pi/8) \\ \cos(-\pi/8) & \sin(-\pi/8) \end{bmatrix}$$

The dot is x, the black vectors are $\mathcal{W}_x = \{w_j : \ w_j \cdot x \geq 0\}$

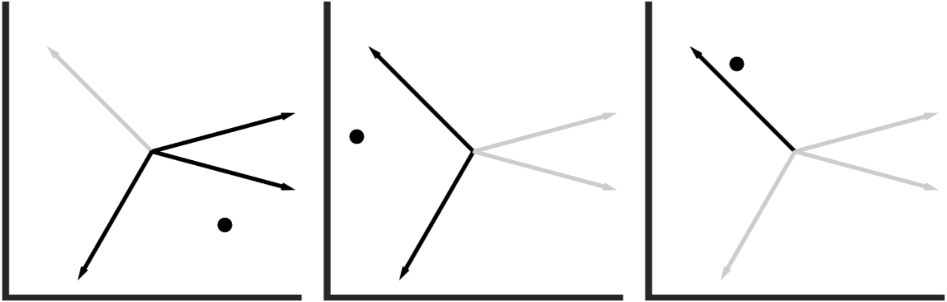


Spanning condition for the row vectors w_j of W

Let $N(x) = \mathrm{relu}(Wx)$, where

$$W = \begin{bmatrix} \cos(\pi/8) & \sin(\pi/8) \\ \cos(5\pi/8) & \sin(5\pi/8) \\ \cos(-5\pi/8) & \sin(-5\pi/8) \\ \cos(-\pi/8) & \sin(-\pi/8) \end{bmatrix}$$

The dot is x, the black vectors are $\mathcal{W}_x = \{w_j : \ w_j \cdot x \geq 0\}$



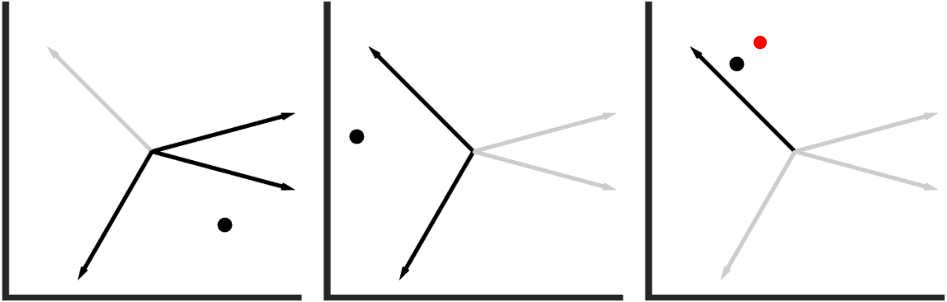
$\mathrm{span}(\mathcal{W}_x) = \mathbb{R}^2$ $\qquad$ $\mathrm{span}(\mathcal{W}_x) = \mathbb{R}^2$ $\qquad$ $\mathrm{span}(\mathcal{W}_x) \neq \mathbb{R}^2$

Spanning condition for the row vectors w_j of W

Let $N(x) = \text{relu}(Wx)$, where

$$W = \begin{bmatrix} \cos(\pi/8) & \sin(\pi/8) \\ \cos(5\pi/8) & \sin(5\pi/8) \\ \cos(-5\pi/8) & \sin(-5\pi/8) \\ \cos(-\pi/8) & \sin(-\pi/8) \end{bmatrix}$$

The dot is x, the black vectors are $\mathcal{W}_x = \{w_j : \ w_j \cdot x \geq 0\}$

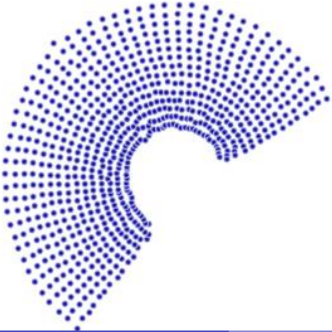


$\text{span}(\mathcal{W}_x) = \mathbb{R}^2$ $\quad$ $\text{span}(\mathcal{W}_x) = \mathbb{R}^2$ $\quad$ $\text{span}(\mathcal{W}_x) \neq \mathbb{R}^2$

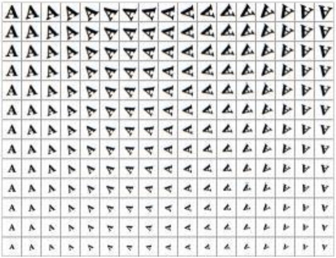
Outline:

- Neural networks and the manifold hypothesis
- Networks with injective layers
- Injective networks and approximation of manifolds
- Flow networks and approximation of probability distributions
- Applications in inverse problems

$$f_\theta(K) = M, \quad K \subset \mathbb{R}^2,\ M \subset \mathbb{R}^m$$



$\xrightarrow{\mathbf{f}_\theta}$



Universal Approximation by Injective Neural Networks

By using injective neural networks we are able to approximate any continuous function that maps a low dimensional space into a higher one.

Theorem (Puthawala-Kothari-L.-Dokmanic-de Hoop JMLR 2022)

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a continuous function, where $m \geq 2n + 1$, and $L \geq 2$. Then for any $\varepsilon > 0$ and compact subset $K \subset \mathbb{R}^n$ there exists a relu-neural network $N_\theta \in \mathcal{N}\mathcal{N}(n, m)$ of depth L such that $N_\theta : \mathbb{R}^n \to \mathbb{R}^m$ is injective (that is, a one-to-one function) and

$$|f(x) - N_\theta(x)| < \varepsilon, \quad \text{for all } x \in K.$$

Universal Approximation by Injective Neural Networks

By using injective neural networks we are able to approximate any continuous function that maps a low dimensional space into a higher one.

Theorem (Puthawala-Kothari-L.-Dokmanic-de Hoop JMLR 2022)

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a continuous function, where $m \geq 2n+1$, and $L \geq 2$. Then for any $\varepsilon > 0$ and compact subset $K \subset \mathbb{R}^n$ there exists a neural network $N_\theta \in \mathcal{N}\mathcal{N}(n,m)$ of depth L such that $N_\theta : \mathbb{R}^n \to \mathbb{R}^m$ is injective (that is, a one-to-one function) and

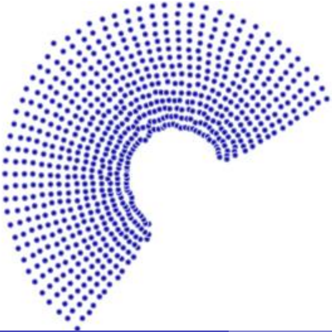
$$|f(x) - N_\theta(x)| < \varepsilon, \quad \text{for all } x \in K.$$

Combining our results with [Yarotsky, Neural Networks 2017], see also [Guhring-Raslan, Neural Networks 2022], we can estimate the depth, the width, and the inverse Lipschitz constant of N_θ in terms of ε, K and the C^1-norm of f.

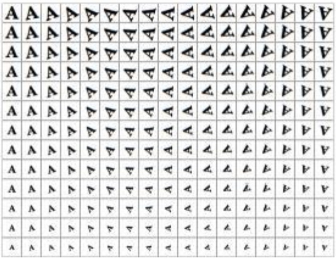
Outline:

- Neural networks and the manifold hypothesis
- Networks with injective layers
- Injective networks and approximation of manifolds
- Flow networks and approximation of probability distributions
- Applications in inverse problems

$$f_\theta(K) = M, \quad f_\theta(Z) \sim \mu_0$$



$\xrightarrow{\mathbf{f}_\theta}$



Diffeomorphic neural networks

The affine coupling flows are the maps $g : \mathbb{R}^n \to \mathbb{R}^n$,

$$g(x_1, \dots, x_{n-1}, x_n) = \Big(x_1, \dots, x_{n-1}, e^{a_\theta(x_1,\dots,x_{n-1})} \cdot x_n + b_\theta(x_1, \dots, x_{n-1})\Big)$$

where a_θ and b_θ are neural networks, e.g., relu-neural networks.
It is easy to compute the inverse function of g,

$$g^{-1}(x_1, \dots, x_n) = \Big(x_1, \dots, x_{n-1}, e^{-a_\theta(x_1,\dots,x_{n-1})} \cdot (x_n - b_\theta(x_1, \dots, x_{n-1}))\Big)$$

Functions $g : \mathbb{R}^n \to \mathbb{R}^n$ and $g^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ are continuous bijective maps.

References: Dinh et al ICLR 2015, Teshima et al NeurIPS 2020.

Universal approximation of diffeomorphisms

The diffeo-networks are finite compositions of linear invertible maps $L_j : \mathbb{R}^n \to \mathbb{R}^n$ and affine coupling flows $g_j : \mathbb{R}^n \to \mathbb{R}^n$, that is,

$$T_\theta = L_k \circ g_k \circ \cdots \circ L_1 \circ g_1 : \mathbb{R}^n \to \mathbb{R}^n.$$

We denote the set of diffeo-networks by $\mathcal{T}_n$. Observe that

$$T_\theta^{-1} = g_1^{-1} \circ L_1^{-1} \circ \cdots \circ g_k^{-1} \circ L_k^{-1}.$$

Teshima et al, NeurIPS 2020 (c.f. Bogachev 2007), proved the following:

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a C^k-smooth diffeomorphism, $k \geq 2$, and $1 \leq p < \infty$. Then for any $\varepsilon > 0$ and compact subset $K \subset \mathbb{R}^n$ there exists a neural network $T_\theta \in \mathcal{T}_n$

$$\|f - T_\theta\|_{L^p(K)} < \varepsilon.$$

Similar result holds for deep sigmoidal flows with $p = \infty$.

Definition of injective flow networks

Let $n, m \in \mathbb{N}$. An injective flow network is function

$$f = T_2 \circ R \circ T_1 : \mathbb{R}^n \to \mathbb{R}^m, \quad \text{where}$$

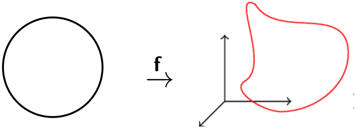
1. $T_1 : \mathbb{R}^n \to \mathbb{R}^n$ is a diffeo-network,
2. $R : \mathbb{R}^n \to \mathbb{R}^m$ be an injective linear map, e.g., $R(x) = (x, 0)$,
3. $T_2 : \mathbb{R}^m \to \mathbb{R}^m$ is a diffeo-network.

We denote the set of injective flow networks by $\mathcal{IN} = \mathcal{IN}_{n,m}$

When are these networks universal approximators for all embeddings?

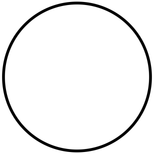
When $K \subset \mathbb{R}^n$ is compact, we say that a function $f : K \to \mathbb{R}^m$ is an embedding if $f : K \to \mathbb{R}^m$ is continuous and injective.
We denote embeddings by

$$\mathrm{emb}^k(K, \mathbb{R}^m) = \mathrm{emb}(K, \mathbb{R}^m) \cap C^k(K, \mathbb{R}^m).$$

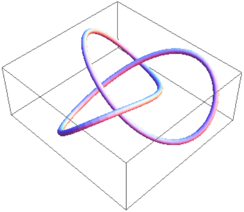


All embeddings are not given by flow networks

Let $K = S^1 \subset \mathbb{R}^2$ be a circle, and $f : K \to \mathbb{R}^3$ be an embedding of K to a trefoil knot to $\mathbb{R}^3$. There are no map $E = T_2 \circ R \circ T_1$, where $T_1 : \mathbb{R}^2 \to \mathbb{R}^2$ is a diffeomorphism, $R : \mathbb{R}^2 \to \mathbb{R}^3$ is an injective linear map and $T_2 : \mathbb{R}^3 \to \mathbb{R}^3$ is a diffeomorphism, so that $E(K) = f(K)$.



(a) $K = S^1$



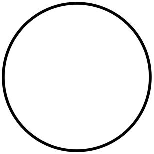
(b) Trefoil knot in $\mathbb{R}^3$.

Example: the manifold of 3×3 patches of pixels of natural images can be well modelled by a Klein bottle [Carlsson et al 2008].

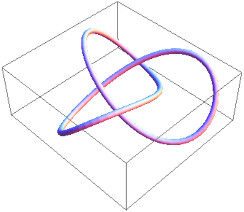
Practical implication: It is difficult to learn a prior distribution that is supported on a general submanifold. Next we show that this difficulty can be avoided by choosing a finer discretization in the inverse problem.

All embeddings are not given by flow networks

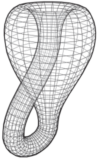
Let $K = S^1 \subset \mathbb{R}^2$ be a circle, and $f : K \to \mathbb{R}^3$ be an embedding of K to a trefoil knot to $\mathbb{R}^3$. There are no map $E = T_2 \circ R \circ T_1$, where $T_1 : \mathbb{R}^2 \to \mathbb{R}^2$ is a diffeomorphism, $R : \mathbb{R}^2 \to \mathbb{R}^3$ is an injective linear map and $T_2 : \mathbb{R}^3 \to \mathbb{R}^3$ is a diffeomorphism, so that $E(K) = f(K)$.



(a) $K = S^1$



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



(c) Klein's bottle.

Example: the manifold of 3×3 patches of pixels of natural images can be well modelled by a Klein bottle [Carlsson et al 2008].

Practical implication: It is difficult to learn a prior distribution that is supported on a general submanifold. Next we show that this difficulty can be avoided by choosing a finer discretization in the inverse problem.

All embeddings are extendable when $m \geq 3n+1$

Definition (Extendable Embedding)

Let $K \subset \mathbb{R}^n$ be compact set. The set of extendable embeddings is

$$\mathcal{I}^k(K,\mathbb{R}^m) := \{\Phi \circ R \in C(K,\mathbb{R}^m) \colon \Phi \in \mathrm{Diff}^k(\mathbb{R}^m \to \mathbb{R}^m),\\ R : \mathbb{R}^n \to \mathbb{R}^m \text{ is injective and linear}\}.$$

Theorem (Puthawala-L.-Dokmanic-de Hoop ICML 2022)

When $m \geq 3n+1$, $k \geq 1$, and $K \subset \mathbb{R}^n$ is compact, it holds that

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

When $p < \infty$, for any C^k-embedding $f \in emb^k(\mathbb{R}^n,\mathbb{R}^m)$ there is a map E in the L^p-closure of the injective flow neural networks $\mathcal{IN}$ such that $E(K) = f(K)$.

The proof is based on the "clean trick" in algebraic topology, see e.g. Madsen-Tornehave 1997

Approximation of probability measures

Theorem (Puthawala-L.-Dokmanic-de Hoop ICML 2022)

Suppose $m \geq 3n+1$. Let ν be a probability measure on a compact set $K \subset \mathbb{R}^n$ that is absolutely continuous w.r.t. Lebesgue measure, $f : K \to \mathbb{R}^m$ a C^1-embedding, and

1. *$R : \mathbb{R}^n \to \mathbb{R}^m$ is an injective linear map*
2. *$\mathcal{T}_m$ is universal approximator of diffeomorphisms in L^∞,*
3. *$\mathcal{T}'_n$ is universal approximator of diffeomorphisms in L^p, $p \geq 1$.*

Then, there are $E_i = T_i \circ R \circ T'_i$, $T_i \in \mathcal{T}_m$, $T'_i \in \mathcal{T}'_n$, $i = 1, 2, \dots$ such that

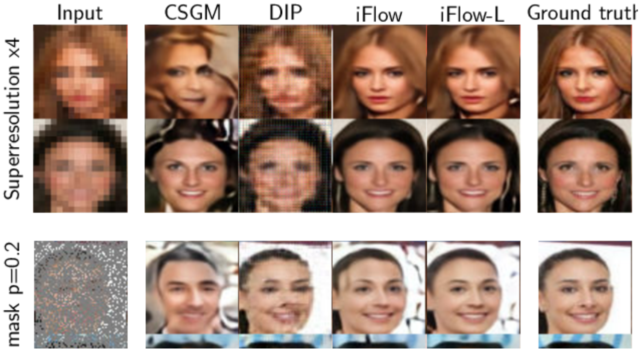
$$\lim_{i \to \infty} \mathrm{W}_2 \left(f_\# \nu, (E_i)_\# \nu \right) = 0,$$

where W_2 is the Wasserstein distance and $F_\# \nu(A) = \nu(F^{-1}(A))$.

This means the following: if $Z \sim \nu$ and $X = f(Z) \sim \mu_0$ then the distribution of $X_i = E_i(Z)$ is close to μ_0.

Outline:

- Neural networks and the manifold hypothesis
- Injective networks and approximation of manifolds
- Flow networks and approximation of probability distributions
- Applications in inverse problems
- Neural operators



Credit on figure: Kothari-Khorashadizadeh-de Hoop-Dokmanic UAI 2021.

Solving an inverse problem with a learned prior

Let μ_0 be a distribution supported on manifold $M \subset \mathbb{R}^m$.
Let x_j, $j = 1, 2, \dots, N$ be independent samples from the distribution μ_0 and $Z \sim N(0, I)$ in $\mathbb{R}^n$.
We aim to find a neural network $f_\theta : \mathbb{R}^n \to \mathbb{R}^m$ such that $f_\theta(Z) \sim \mu_0$.

Then, we will consider the following inverse problem:
Let $A : \mathbb{R}^m \to \mathbb{R}^k$ and

$$\begin{aligned} Y &= AX + \varepsilon, \\ X &= f_\theta(Z) \in \mathbb{R}^m, \quad \text{where } Z \sim N(0, I) \text{ in } \mathbb{R}^n, \\ \varepsilon &\sim N(0, I). \end{aligned}$$

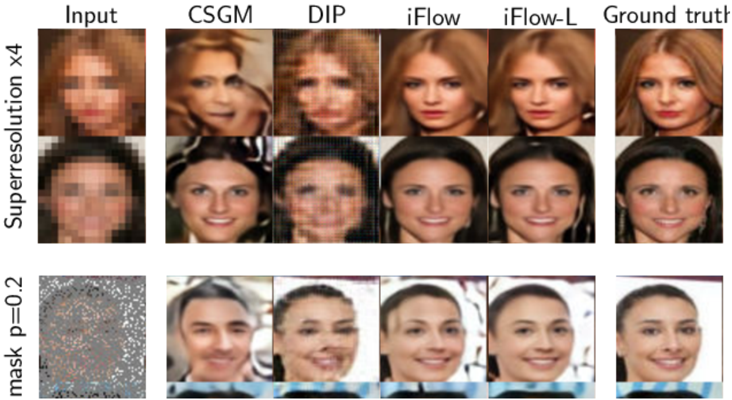
Let y be a sample of the random variable Y. The Maximum A Posteriori estimate for X is

$$x_{MAP} = \mathrm{argmax}_x \pi_{X|Y}(x|y)$$

We seek a solution of the form $x = f_\theta(z)$.

Numerical results

Figure below shows numerical the results with iFlow and iFlow-L in [Kothari-Khorashadizadeh-de Hoop-Dokmanic UAI 2021] when $Y = AX + \varepsilon$ and A is a down-sampling operator or a random mask.



Outline:

- Neural networks and the manifold hypothesis
- Networks with injective layers
- Injective networks and approximation of manifolds
- Flow networks and approximation of probability distributions
- Applications in inverse problems
- Neural operators

Definition (Neural operators, by Kovachki-Lanthaler-Mishra 2021)

Let $D \subset \mathbb{R}^d$ is bounded open set. A neural operator is a function

$$F : L^2(D)^{n_0} \to L^2(D)^{n_k}, \quad F = \mathcal{L}_{k-1} \circ \mathcal{L}_k \circ \cdots \circ \mathcal{L}_0$$

where layers $\mathcal{L}_\ell : L^2(D)^{n_\ell} \to L^2(D)^{n_{\ell+1}}$ are

$$\mathcal{L}_\ell(u)(x) = \sigma(W_\ell u(x) + K_\ell(u)(x)),$$

and $W_\ell \in C(\overline{D}, \mathbb{R}^{n_{\ell+1} \times n_\ell})$ and K_ℓ are non-linear integral operators

$$K_\ell(u)(x) = \int_D k_\ell(x, y, u(x), u(y))u(y)dy,$$

where k_ℓ are bounded and C^3-smooth (or a generalized function), and $\sigma : \mathbb{R}^n \to \mathbb{R}^n$ is pointwise activation function (e.g. a leaky relu).

A leaky relu is $\sigma_a(x) = \mathrm{relu}_a(x) = \begin{cases} x, & x > 0, \\ ax, & x \le 0, \end{cases}$ for $x \in \mathbb{R}$, $a \ge 0$.

Analogous continuous generative networks based on wavelets by Alberti, Santacesaria, and Sciutto 2022.

A non-linear operator $G : L^2(D)^n \to L^2(D)^n$ is coercive if

$$\lim_{\|u\|_{L^2(D)^n}\to\infty} \langle G(u), \frac{u}{\|u\|_{L^2(D)^n}} \rangle_{L^2(D)^n} = \infty.$$

Theorem (de Hoop-Furuya-L.-Puthawala, NeurIPS 2023)

Let $\sigma : \mathbb{R} \to \mathbb{R}$ be surjective (e.g., a leaky relu), values of $W \in C^1(\overline{D}; \mathbb{R}^{n\times n})$ be invertible matrixes, and $K : L^2(D)^n \to L^2(D)^n$ be a non-linear integral operator with a C^3-smooth kernel.
Moreover, assume that the map $u \mapsto \alpha u + W^{-1}K(u)$ is coercive with some $0 < \alpha < 1$. Then, the operator

$$F : L^2(D)^n \to L^2(D)^n,$$
$$F(u) = \sigma(Wu + K(u))$$

is surjective.

The proof is based on Leray-Schauder degree theory.
If K is a non-linear Volterra operator, then $u \to Wu + K(u)$ is injective.

Next we consider the case $n = 1$ and the non-linear integral operator

$$K(u)(x) := \int_D k(x, y, u(x))u(y)dy, \; x \in D.$$

Assume that

$$k(x, y, t) = \sum_{j=1}^{J} c_j(x, y)\phi(a_j(x, y)t + b_j(x, y)),$$

where ϕ is a wavelet activation function $\phi_{wire} : \mathbb{R} \to \mathbb{R}$, see Saragadam et al. 2023,

$$\phi_{wire}(t) = \operatorname{Im}\left(e^{i\omega t} e^{-t^2}\right),$$

and $a, b, c \in C(\overline{D} \times \overline{D})$, $a_j(x, y) \neq 0$.
For $\alpha > 0$, the operators $u \mapsto \alpha u + W^{-1}K(u)$ are coersive and the operator

$$\begin{aligned} &F : L^2(D) \to L^2(D), \\ &F(u) = \sigma(Wu + K(u)) \end{aligned}$$

is surjective.

Let $n = 1$ and $D \subset \mathbb{R}$ be a bounded interval. Let

$$F(u) = \sigma_a \circ G(u), \quad G(u)(x) = W(x)u(x) + \int_D k(x, y, u(y))u(y)dy,$$

where $a > 0$. Assume that $W \in C^1(\overline{D})$ satisfies $0 < c_1 \leq W(x) \leq c_2$ and

$$\|k\|_{C^3(\overline{D}\times\overline{D}\times\mathbb{R})} \leq c_0, \quad \|W\|_{C^1(\overline{D})} \leq c_0 \tag{2}$$

and for all $u_0 \in H^1(D)$, the Frechet derivatives satisfy

$$DG|_{u_0} : H^1(D) \to H^1(D) \text{ is an injective operator.} \tag{3}$$

Theorem (de Hoop-Furuya-L.-Puthawala, NeurIPS 2023)

Assume that F satisfies the above assumptions (2) and (3) and that $F : H^1(D) \to H^1(D)$ is a bijection. Let $\mathcal{Y} \subset \sigma_a(\overline{B}_{C^{1,\alpha}(\overline{D})}(0, R))$ where $\alpha > 0$. Then the inverse of $F : H^1(D) \to H^1(D)$ in $\mathcal{Y}$ can written as a limit as a limit of integral neural operators having distributional kernels.

Idea of the proof

The inverse of $F = \sigma_a \circ G$ is $F^{-1} = G^{-1} \circ \sigma_{1/a}$.
When w is sufficiently near to $w_j = G(u_j)$ in $H^1(D)$, the Banach fixed point theorem implies that

$$\begin{pmatrix} G^{-1}(w) \\ w \end{pmatrix} = \lim_{m\to\infty} \mathcal{H}_j^{\circ m} \begin{pmatrix} u_j \\ w \end{pmatrix},$$

$$\mathcal{H}_j \begin{pmatrix} u \\ w \end{pmatrix} = \begin{pmatrix} u - A_{u_j}^{-1}(G(u) - w) \\ w \end{pmatrix}, \quad A_{u_j} = DG|_{u_j}.$$

When we cover $\sigma_{1/a}(\mathcal{Y})$ with small neighborhoods of w_j in $H^1(D)$, and define a partition of unity Φ_j for these neighborhoods.
The operators $\Phi_j = \psi_{j,k} \circ \cdots \circ \psi_{j,1}$ are products operators of the form

$$\psi_{j,k} \begin{pmatrix} u \\ w \end{pmatrix}(x) = \begin{pmatrix} \int_D m_{j,k}(x, y, v(x), w(y))dy \\ w(x) \end{pmatrix}$$

where $m_{j,k}(x, y, v(x), w(y)) = v(x)\mathbf{1}_{[s_{j,k}, s_{j,k}+h)}(w(y))\delta(y - z_{j,k})$.

Idea of the proof

Combining the above operators, we have for $w \in \sigma_{1/a}(\mathcal{Y})$

$$\begin{pmatrix} G^{-1}(w) \\ w \end{pmatrix} = \lim_{m\to\infty} \sum_{j\in\mathcal{I}} \Phi_j \mathcal{H}_j^{\circ m} \begin{pmatrix} u_j \\ w \end{pmatrix}.$$

Thus, $G^{-1}|_{\sigma_{1/a}(\mathcal{Y})}$ can be written as a limit of neural operators which kernels are generalized functions.

Summary

- Neural operators are generalization of neural networks where matrix multiplication is replaced by integral operators.
- Compactness of integral operators make it possible to analyze mapping properties of neural operators.
- Neural operators are non-linear operators which share some properties of pseudodifferential operators:
  - Neural operators and pseudodifferential operators are algebras
  - If pseudodifferential operator is invertible (more precisely, elliptic) then the inverse operator is a pseudodifferential operator.
    Similarly, if a neural operator is invertible then the inverse operator is a neural operator.

Thank you for your attention

Supported by the PDE-Inverse project of the European Research Council of the European Union.[1]

[1] Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the other funding organizations. The European Union can not be held responsible for them.