Appendix

A Proof of Proposition 1 in Section 2

Proof. We use the notation $T|_{S(v,T+b)}(v) = (T_i v)_{i \in S(v,T+b)}$. Assume that $T+b$ has a DSS with respect to every $v \in L^2(D)^n$ in the sense of Definition 2, and that

$$\mathrm{ReLU}(Tv^{(1)} + b) = \mathrm{ReLU}(Tv^{(2)} + b) \quad \text{in } D, \tag{A.1}$$

where $v^{(1)}, v^{(2)} \in L^2(D)^n$. Since $T+b$ has a DSS with respect to $v^{(1)}$, we have for $i \in S(v^{(1)}, T+b)$

$$0 < \mathrm{ReLU}(T_i v^{(1)} + b_i) = \mathrm{ReLU}(T_i v^{(2)} + b_i) \text{ in } D,$$

which implies that

$$T_i v^{(1)} + b_i = T_i v^{(2)} + b_i \text{ in } D.$$

Thus,

$$v^{(1)} - v^{(2)} \in \mathrm{Ker}\left(T|_{S(v^{(1)},T+b)} \right). \tag{A.2}$$

By assuming (A.1), we have for $i \notin S(v^{(1)}, T)$,

$$\{x \in D \mid T_i v^{(1)}(x) + b_i(x) \le 0\} = \{x \in D \mid T_i v^{(2)}(x) + b_i(x) \le 0\}.$$

Then, we have

$$T_i(v^{(1)} - (v^{(1)} - v^{(2)}))(x) + b_i(x) = T_i v^{(2)}(x) + b_i(x) \le 0 \text{ if } T_i v^{(1)}(x) + b_i(x) \le 0,$$

that is,

$$T_i v^{(1)}(x) + b_i(x) \le T_i \left(v^{(1)} - v^{(2)} \right)(x) \text{ if } T_i v^{(1)}(x) + b_i(x) \le 0.$$

In addition,

$$T_i(v^{(1)} - v^{(2)})(x) = T_i v^{(1)}(x) + b_i(x) - \left(T v^{(2)}(x) + b_i(x) \right) = 0 \text{ if } T_i v^{(1)}(x) + b_i(x) > 0.$$

Thus,

$$v^{(1)} - v^{(2)} \in X(v, T+b). \tag{A.3}$$

Combining (A.2) and (A.3), and (2.1) as $v = v^{(1)}$, we conclude that

$$v^{(1)} - v^{(2)} = 0.$$

Conversely, assume that there exists a $v \in L^2(D)^n$ such that

$$\mathrm{Ker}\left(T|_{S(v,T+b)} \right) \cap X(v, T+b) \ne \{0\}.$$

Then there is $u \ne 0$ such that

$$u \in \mathrm{Ker}\left(T|_{S(v,T+b)} \right) \cap X(v, T+b).$$

For $i \in S(v, T+b)$, we have by $u \in \mathrm{Ker}(T_i)$,

$$\mathrm{ReLU}\left(T_i(v - u) + b_i(x) \right) = \mathrm{ReLU}\left(T_i v + b_i(x) \right).$$

For $i \notin S(v, T+b)$, we have by $u \in X(v, T+b)$,

$$\begin{aligned} \mathrm{ReLU}\left(T_i(v-u)(x) + b_i(x) \right) &= \begin{cases} 0 & \text{if } T_i v(x) + b_i(x) \le 0 \\ T_i v(x) + b_i(x) & \text{if } T_i v(x) + b_i(x) > 0 \end{cases} \\ &= \mathrm{ReLU}\left(T_i v(x) + b_i(x) \right). \end{aligned}$$

Therefore, we conclude that

$$\mathrm{ReLU}\left(T(v-u) + b \right) = \mathrm{ReLU}\left(Tv + b \right),$$

where $u \ne 0$, that is, $\mathrm{ReLU} \circ (T+b)$ is not injective. $\square$

B Details of Sections 3.1 and 3.2

B.1 Proof of Lemma 1

Proof. The restriction operator, $\pi_\ell : L^2(D)^m \to L^2(D)^\ell$ ($\ell < m$), acts as follows,

$$\pi_\ell(a,b) := b, \quad (a,b) \in L^2(D)^{m-\ell} \times L^2(D)^\ell. \tag{B.1}$$

Since $L^2(D)$ is a separable Hilbert space, there exists an orthonormal basis $\{\varphi_k\}_{k\in\mathbb{N}}$ in $L^2(D)$. We denote by

$$\varphi^0_{k,j} := \Bigg(0, ..., 0, \underbrace{\varphi_k}_{j-th}, 0, ..., 0\Bigg) \in L^2(D)^m,$$

for $k \in \mathbb{N}$ and $j \in [m-\ell]$. Then, $\{\varphi^0_{k,j}\}_{k\in\mathbb{N}, j\in[m-\ell]}$ is an orthonormal sequence in $L^2(D)^m$, and

$$\begin{aligned} V_0 &:= L^2(D)^{m-\ell} \times \{0\}^\ell \\ &= \operatorname{span}\left\{\varphi^0_{k,j} \,\middle|\, k \in \mathbb{N},\ j \in [m-\ell]\right\}. \end{aligned}$$

We define, for $\alpha \in (0,1)$,

$$\varphi^\alpha_{k,j} := \Bigg(0, ..., 0, \underbrace{\sqrt{(1-\alpha)}\varphi_k}_{j-th}, 0, ..., 0, \sqrt{\alpha}\xi_{(k-1)(m-\ell)+j}\Bigg) \in L^2(D)^m, \tag{B.2}$$

with $k \in \mathbb{N}$ and $j \in [m-\ell]$. We note that $\{\varphi^\alpha_{k,j}\}_{k\in\mathbb{N}, j\in[m-\ell]}$ is an orthonormal sequence in $L^2(D)^m$. We set

$$V_\alpha := \operatorname{span}\left\{\varphi^\alpha_{k,j} \,\middle|\, k \in \mathbb{N}, j \in [m-\ell]\right\}. \tag{B.3}$$

It holds for $0 < \alpha < 1/2$ that

$$\left\| P_{V_\alpha^\perp} - P_{V_0^\perp} \right\|_{\mathrm{op}} < 1.$$

Indeed, for $u \in L^2(D)^m$ and $0 < \alpha < 1/2$,

$$\begin{aligned} &\left\| P_{V_\alpha^\perp} u - P_{V_0^\perp} u \right\|^2_{L^2(D)^m} = \| P_{V_\alpha} u - P_{V_0} u \|^2_{L^2(D)^m} \\ &= \left\| \sum_{k\in\mathbb{N}, j\in[m-\ell]} (u, \varphi^\alpha_{k,j})\varphi^\alpha_{k,j} - (u, \varphi^0_{k,j})\varphi^0_{k,j} \right\|^2_{L^2(D)^m} \\ &= \left\| \sum_{k\in\mathbb{N}, j\in[m-\ell]} (1-\alpha)(u_j, \varphi_k)\varphi_k - (u_j, \varphi_k)\varphi_k \right\|^2_{L^2(D)} \\ &+ \left\| \sum_{k\in\mathbb{N}, j\in[m-\ell]} \alpha(u_m, \xi_{(k-1)(m-\ell)+j})\xi_{(k-1)(m-\ell)+j} \right\|^2_{L^2(D)} \\ &\le \alpha^2 \sum_{j\in[m-\ell]} \sum_{k\in\mathbb{N}} |(u_j, \varphi_k)|^2 + \alpha^2 \sum_{k\in\mathbb{N}} |(u_m, \xi_k)|^2 \le 4\alpha^2 \|u\|^2_{L^2(D)^m}, \end{aligned}$$

which implies that $\left\| P_{V_\alpha^\perp} - P_{V_0^\perp} \right\|_{\mathrm{op}} \le 2\alpha$.

We will show that the operator

$$P_{V_\alpha^\perp} \circ T : L^2(D)^n \to L^2(D)^m,$$

is injective. Assuming that for $a, b \in L^2(D)^n$,

$$P_{V_\alpha^\perp} \circ T(a) = P_{V_\alpha^\perp} \circ T(b),$$

is equivalent to

$$T(a) - T(b) = P_{V_\alpha}(T(a) - T(b)).$$

Denoting by $P_{V_\alpha}(T(a) - T(b)) = \sum_{k\in\mathbb{N}, j\in[m-\ell]} c_{k,j}\varphi^\alpha_{k,j}$,

$$\pi_1(T(a) - T(b)) = \sum_{k\in\mathbb{N}, j\in[m-\ell]} c_{k,j}\xi_{(k-1)(m-\ell)+j}.$$

From (3.1), we obtain that $c_{kj} = 0$ for all k, j. By injectivity of T, we finally get $a = b$.

We define $Q_\alpha : L^2(D)^m \to L^2(D)^m$ by

$$Q_\alpha := \left(P_{V_0^\perp}P_{V_\alpha^\perp} + (I - P_{V_0^\perp})(I - P_{V_\alpha^\perp})\right)\left(I - (P_{V_0^\perp} - P_{V_\alpha^\perp})^2\right)^{-1/2}.$$

By the same argument as in Section I.4.6 Kato [2013], we can show that Q_α is injective and

$$Q_\alpha P_{V_\alpha^\perp} = P_{V_0^\perp} Q_\alpha,$$

that is, Q_α maps from $\mathrm{Ran}(P_{V_\alpha^\perp})$ to

$$\mathrm{Ran}(P_{V_0^\perp}) \subset \{0\}^{m-\ell} \times L^2(D)^\ell.$$

It follows that

$$\pi_\ell \circ Q_\alpha \circ P_{V_\alpha^\perp} \circ T : L^2(D)^n \to L^2(D)^\ell$$

is injective. □

B.2 Remarks following Lemma 1

Remark 2. *An example that satisfies (3.1) is the neural operator whose L-th layer operator $\mathcal{L}_L$ consists of the integral operator K_L with continuous kernel function k_L, and with continuous activation function σ. Indeed, in this case, we may choose the orthogonal sequence $\{\xi_k\}_{k\in\mathbb{N}}$ in $L^2(D)$ as a discontinuous functions sequence [1] so that $\mathrm{span}\{\xi_k\}_{k\in\mathbb{N}} \cap C(D) = \{0\}$. Then, by $\mathrm{Ran}(\mathcal{L}_L) \subset C(D)^{d_L}$, the assumption (3.1) holds.*

Remark 3. *In the proof of Lemma 1, an operator $B \in \mathcal{L}(L^2(D)^m, L^2(D)^\ell)$,*

$$B = \pi_\ell \circ Q_\alpha \circ P_{V_\alpha^\perp},$$

appears, where $P_{V_\alpha^\perp}$ is the orthogonal projection onto orthogonal complement $V_\alpha^\perp$ of V_α with

$$V_\alpha := \mathrm{span}\left\{\varphi^\alpha_{k,j} \;\middle|\; k \in \mathbb{N}, j \in [m-\ell]\right\} \subset L^2(D)^m,$$

in which $\varphi^\alpha_{k,j}$ is defined for $\alpha \in (0,1)$, $k \in \mathbb{N}$ and $j \in [\ell]$,

$$\varphi^\alpha_{k,j} := \left(0, ..., 0, \underbrace{\sqrt{(1-\alpha)}\varphi_k}_{j-th}, 0, ..., 0, \sqrt{\alpha}\xi_{(k-1)(m-\ell)+j}\right).$$

Here, $\{\varphi_k\}_{k\in\mathbb{N}}$ is an orthonormal basis in $L^2(D)$. Futhermore, $Q_\alpha : L^2(D)^m \to L^2(D)^m$ is defined by

$$Q_\alpha := \left(P_{V_0^\perp}P_{V_\alpha^\perp} + (I - P_{V_0^\perp})(I - P_{V_\alpha^\perp})\right)\left(I - (P_{V_0^\perp} - P_{V_\alpha^\perp})^2\right)^{-1/2},$$

where $P_{V_0^\perp}$ is the orthogonal projection onto orthogonal complement $V_0^\perp$ of V_0 with

$$V_0 := L^2(D)^{m-\ell} \times \{0\}^\ell.$$

The operator Q_α is well-defined for $0 < \alpha < 1/2$ because it holds that

$$\left\| P_{V_\alpha^\perp} - P_{V_0^\perp} \right\|_{\mathrm{op}} < 2\alpha.$$

This construction is given by the combination of "Pairs of projections" discussed in Kato [2013, Section I.4.6] with the idea presented in [Puthawala et al., 2022b, Lemma 29].

[1] e.g., step functions whose supports are disjoint for each sequence.

B.3 Proof of Theorem 1

We begin with

Definition 3. *The set of L-layer neural networks mapping from $\mathbb{R}^d$ to $\mathbb{R}^{d'}$ is*

$$\mathrm{N}_L(\sigma;\mathbb{R}^d,\mathbb{R}^{d'}) := \Big\{ f:\mathbb{R}^d \to \mathbb{R}^{d'} \Big| f(x) = W_L\sigma(\cdots W_1\sigma(W_0x+b_0)+b_1\cdots)+b_L,$$
$$W_\ell \in \mathbb{R}^{d_{\ell+1}\times d_\ell}, b_\ell \in \mathbb{R}^{d_{\ell+1}}, d_\ell \in \mathbb{N}_0(d_0=d,\ d_{L+1}=d'), \ell=0,...,L\Big\},$$

where $\sigma:\mathbb{R}\to\mathbb{R}$ is an element-wise nonlinear activation function. For the class of nonlinear activation functions,

$$\mathrm{A}_0 := \Big\{ \sigma \in C(\mathbb{R}) \Big| \exists n \in \mathbb{N}_0 \text{ s.t. } \mathrm{N}_n(\sigma;\mathbb{R}^d,\mathbb{R}) \text{ is dense in } C(K) \text{ for } \forall K \subset \mathbb{R}^d \text{ compact} \Big\}$$

$$\mathrm{A}_0^L := \Big\{ \sigma \in A_0 \Big| \sigma \text{ is Borel measurable s.t. } \sup_{x\in\mathbb{R}} \frac{|\sigma(x)|}{1+|x|} < \infty \Big\}$$

$$\mathrm{BA} := \Big\{ \sigma \in A_0 \Big| \forall K \subset \mathbb{R}^d \text{ compact}, \forall\epsilon>0, \text{ and } \forall C \geq \mathrm{diam}(K), \exists n \in \mathbb{N}_0,$$
$$\exists f \in \mathrm{N}_n(\sigma;\mathbb{R}^d,\mathbb{R}^d) \text{ s.t. } |f(x)-x| \leq \epsilon,\ \forall x \in K, \text{ and, } |f(x)| \leq C,\ \forall x \in \mathbb{R}^d \Big\}.$$

The set of integral neural operators with L^2-integral kernels is

$$\begin{aligned}
\mathrm{NO}_L(\sigma;D,d_{in},d_{out}) := \Big\{ & G : L^2(D)^{d_{in}} \to L^2(D)^{d_{out}} \Big| \\
& G = K_{L+1}\circ(K_L+b_L)\circ\sigma\cdots\circ(K_2+b_2)\circ\sigma\circ(K_1+b_1)\circ(K_0+b_0), \\
& K_\ell \in \mathcal{L}(L^2(D)^{d_\ell}, L^2(D)^{d_{\ell+1}}),\ K_\ell : f \mapsto \int_D k_\ell(\cdot,y)f(y)dy\Big|_D, \\
& k_\ell \in L^2(D\times D;\mathbb{R}^{d_{\ell+1}\times d_\ell}),\ b_\ell \in L^2(D;\mathbb{R}^{d_{\ell+1}}), \\
& d_\ell \in \mathbb{N},\ d_0=d_{in},\ d_{L+2}=d_{out},\ \ell=0,...,L+2 \Big\}.
\end{aligned} \tag{B.4}$$

Proof. Let $R>0$ such that

$$K \subsetneq B_R(0),$$

where $B_R(0) := \{u \in L^2(D)^{d_{in}} \mid \|u\|_{L^2(D)^{d_{in}}} \leq R\}$. By Theorem 11 of Kovachki et al. [2021b], there exists $L \in \mathbb{N}$ and $\widetilde{G} \in \mathrm{NO}_L(\sigma;D,d_{in},d_{out})$ such that

$$\sup_{a\in K} \Big\| G^+(a) - \widetilde{G}(a) \Big\|_{L^2(D)^{d_{out}}} \leq \frac{\epsilon}{2}, \tag{B.5}$$

and

$$\Big\| \widetilde{G}(a) \Big\|_{L^2(D)^{d_{out}}} \leq 4M, \quad \text{for } a \in L^2(D)^{d_{in}}, \quad \|a\|_{L^2(D)^{d_{in}}} \leq R.$$

We write operator $\widetilde{G}$ by

$$\widetilde{G} = \widetilde{K}_{L+1}\circ(\widetilde{K}_L+\widetilde{b}_L)\circ\sigma\cdots\circ(\widetilde{K}_2+\widetilde{b}_2)\circ\sigma\circ(\widetilde{K}_1+\widetilde{b}_1)\circ(\widetilde{K}_0+\widetilde{b}_0),$$

where

$$\begin{aligned}
& \widetilde{K}_\ell \in \mathcal{L}(L^2(D)^{d_\ell}, L^2(D)^{d_{\ell+1}}),\ \widetilde{K}_\ell : f \mapsto \int_D \widetilde{k}_\ell(\cdot,y)f(y)dy, \\
& \widetilde{k}_\ell \in C(D\times D;\mathbb{R}^{d_{\ell+1}\times d_\ell}),\ \widetilde{b}_\ell \in L^2(D;\mathbb{R}^{d_{\ell+1}}), \\
& d_\ell \in \mathbb{N},\ d_0=d_{in},\ d_{L+2}=d_{out},\ \ell=0,...,L+2.
\end{aligned}$$

We remark that kernel functions $\widetilde{k}_\ell$ are continuous because neural operators defined in Kovachki et al. [2021b] parameterize the integral kernel function by neural networks, thus,

$$\mathrm{Ran}(\widetilde{G}) \subset C(D)^{d_{out}}. \tag{B.6}$$

We define the neural operator $H : L^2(D)^{d_{in}} \to L^2(D)^{d_{in}+d_{out}}$ by

$$H = K_{L+1} \circ (K_L + b_L) \circ \sigma \cdots \circ (K_2 + b_2) \circ \sigma \circ (K_1 + b_1) \circ (K_0 + b_0),$$

where K_ℓ and b_ℓ are defined as follows. First, we choose $K_{inj} \in \mathcal{L}(L^2(D)^{d_{in}}, L^2(D)^{d_{in}})$ as a linear injective integral operator [2].

(i) When $\sigma_1 \in \mathrm{A}_0^L \cap \mathrm{BA}$ is injective,

$$K_0 = \begin{pmatrix} K_{inj} \\ \widetilde{K}_0 \end{pmatrix} \in \mathcal{L}(L^2(D)^{d_{in}}, L^2(D)^{d_{in}+d_1}), \quad b_0 = \begin{pmatrix} O \\ \widetilde{b}_0 \end{pmatrix} \in L^2(D)^{d_{in}+d_1},$$

$$\vdots$$

$$K_\ell = \begin{pmatrix} K_{inj} & O \\ O & \widetilde{K}_\ell \end{pmatrix} \in \mathcal{L}(L^2(D)^{d_{in}+d_\ell}, L^2(D)^{d_{in}+d_{\ell+1}}), \quad b_\ell = \begin{pmatrix} O \\ \widetilde{b}_\ell \end{pmatrix} \in L^2(D)^{d_{in}+d_{\ell+1}},$$

$$(1 \le \ell \le L),$$

$$\vdots$$

$$K_{L+1} = \begin{pmatrix} K_{inj} & O \\ O & \widetilde{K}_{L+1} \end{pmatrix} \in \mathcal{L}(L^2(D)^{d_{in}+d_{L+1}}, L^2(D)^{d_{in}+d_{out}}), \quad b_\ell = \begin{pmatrix} O \\ O \end{pmatrix} \in L^2(D)^{d_{in}+d_{out}}.$$

(ii) When $\sigma_1 = \mathrm{ReLU}$,

$$K_0 = \begin{pmatrix} K_{inj} \\ \widetilde{K}_0 \end{pmatrix} \in \mathcal{L}(L^2(D)^{d_{in}}, L^2(D)^{d_{in}+d_1}), \quad b_0 = \begin{pmatrix} O \\ \widetilde{b}_0 \end{pmatrix} \in L^2(D)^{d_{in}+d_1},$$

$$K_1 = \begin{pmatrix} K_{inj} & O \\ -K_{inj} & O \\ O & \widetilde{K}_1 \end{pmatrix} \in \mathcal{L}(L^2(D)^{d_{in}+d_1}, L^2(D)^{2d_{in}+d_2}),\ b_0 = \begin{pmatrix} O \\ O \\ \widetilde{b}_1 \end{pmatrix} \in L^2(D)^{2d_{in}+d_1},$$

$$\vdots$$

$$K_\ell = \left(\begin{array}{cc:c} K_{inj} & -K_{inj} & O \\ -K_{inj} & K_{inj} & \\ \hdashline \multicolumn{2}{c:}{O} & \widetilde{K}_\ell \end{array}\right) \in \mathcal{L}(L^2(D)^{2d_{in}+d_\ell}, L^2(D)^{2d_{in}+d_{\ell+1}}),$$

$$b_\ell = \begin{pmatrix} O \\ O \\ \widetilde{b}_\ell \end{pmatrix} \in L^2(D)^{2d_{in}+d_{\ell+1}}, \quad (2 \le \ell \le L),$$

$$\vdots$$

$$K_L = \left(\begin{array}{cc:c} K_{inj} & -K_{inj} & O \\ \hdashline \multicolumn{2}{c:}{O} & \widetilde{K}_L \end{array}\right) \in \mathcal{L}(L^2(D)^{2d_{in}+d_L}, L^2(D)^{d_{in}+d_{L+1}}),$$

$$b_L = \begin{pmatrix} O \\ \widetilde{b}_L \end{pmatrix} \in L^2(D)^{d_{in}+d_{L+1}},$$

$$K_{L+1} = \begin{pmatrix} K_{inj} & O \\ O & \widetilde{K}_{L+1} \end{pmatrix} \in \mathcal{L}(L^2(D)^{d_{in}+d_{L+1}}, L^2(D)^{d_{in}+d_{out}}),$$

$$b_{L+1} = \begin{pmatrix} O \\ O \end{pmatrix} \in L^2(D)^{d_{in}+d_{out}}.$$

Then, the operator $H : L^2(D)^{d_{in}} \to L^2(D)^{d_{in}+d_{out}}$ has the form of

$$H := \begin{cases} \begin{pmatrix} K_{inj} \circ K_{inj} \circ \sigma \circ K_{inj} \circ \cdots \circ \sigma \circ K_{inj} \circ K_{inj} \\ \widetilde{G} \end{pmatrix} & \text{in the case of (i).} \\ \begin{pmatrix} K_{inj} \circ \cdots \circ K_{inj} \\ \widetilde{G} \end{pmatrix} & \text{in the case of (ii).} \end{cases}$$

[2]For example, if we choose the integral kernel k_{inj} as $k_{inj}(x, y) = \sum_{k=1}^{\infty} \vec{\varphi}_k(x)\vec{\varphi}_k(y)$, then the integral operator K_{inj} with the kernel k_{inj} is injective where $\{\vec{\varphi}\}_k$ is the orthonormal basis in $L^2(D)^{d_{in}}$.

For the case of (ii), we have used the fact

$$\begin{pmatrix} I & -I \end{pmatrix} \circ \mathrm{ReLU} \circ \begin{pmatrix} I \\ -I \end{pmatrix} = I.$$

Thus, in both cases, H is injective.

In the case of (i), as $\sigma \in A_0^L$, we obtain the estimate

$$\|\sigma(f)\|_{L^2(D)^{d_{in}}} \leq \sqrt{2|D|d_{in}} C_0 + \|f\|_{L^2(D)^{d_{in}}} \ , \ f \in L^2(D)^{d_{in}},$$

where

$$C_0 := \sup_{x \in \mathbb{R}} \frac{|\sigma(x)|}{1+|x|} < \infty.$$

Then we evaluate for $a \in K (\subset B_R(0))$,

$$\begin{aligned} &\|H(a)\|_{L^2(D)^{d_{in}+d_{out}}} \\ &\leq \left\|\widetilde{G}(a)\right\|_{L^2(D)^{d_{out}}} + \|K_{inj} \circ K_{inj} \circ \sigma \circ K_{inj} \circ \cdots \circ \sigma \circ K_{inj} \circ K_{inj}(a)\|_{L^2(D)^{d_{in}}} \\ &\leq 4M + \sqrt{2|D|d_{in}} C_0 \sum_{\ell=1}^{L} \|K_{inj}\|_{\mathrm{op}}^{\ell+1} + \|K_{inj}\|_{\mathrm{op}}^{L+2} R =: C_H. \end{aligned} \tag{B.7}$$

In the case of (ii), we find the estimate, for $a \in K$,

$$\|H(a)\|_{L^2(D)^{d_{in}+d_{out}}} \leq 4M + \|K_{inj}\|_{\mathrm{op}}^{L+2} R < C_H. \tag{B.8}$$

From (B.6) (especially, $\mathrm{Ran}(\pi_1 H) \subset C(D)$) and Remark 2, we can choose an orthogonal sequence $\{\xi_k\}_{k \in \mathbb{N}}$ in $L^2(D)$ such that (3.1) holds. By applying Lemma 1, as $T = H$, $n = d_{in}$, $m = d_{in} + d_{out}$, $\ell = d_{out}$, we find that

$$G := \underbrace{\pi_{d_{out}} \circ Q_\alpha \circ P_{V_\alpha^\perp}}_{=:B} \circ H : L^2(D)^{d_{in}} \to L^2(D)^{d_{out}},$$

is injective. Here, $P_{V_\alpha^\perp}$ and Q_α are defined as in Remark 3; we choose $0 < \alpha << 1$ such that

$$\left\|P_{V_\alpha^\perp} - P_{V_0^\perp}\right\|_{\mathrm{op}} < \min\left(\frac{\epsilon}{10 C_H}, 1\right) =: \epsilon_0,$$

where $P_{V_0^\perp}$ is the orthogonal projection onto

$$V_0^\perp := \{0\}^{d_{in}} \times L^2(D)^{d_{out}}.$$

By the same argument as in the proof of Theorem 15 in Puthawala et al. [2022a], we can show that

$$\|I - Q_\alpha\|_{\mathrm{op}} \leq 4\epsilon_0.$$

Furthermore, since B is a linear operator, $B \circ K_{L+1}$ is also a linear operator with integral kernel $(Bk_{L+1}(\cdot, y))(x)$, where $k_{L+1}(x,y)$ is the kernel of K_{L+1}. This implies that

$$G \in \mathrm{NO}_L(\sigma; D, d_{in}, d_{out}).$$

We get, for $a \in K$,

$$\left\|G^+(a) - G(a)\right\|_{L^2(D)^{d_{out}}} \leq \underbrace{\left\|G^+(a) - \widetilde{G}(a)\right\|_{L^2(D)^{d_{out}}}}_{\text{(B.5)} \leq \frac{\epsilon}{2}} + \left\|\widetilde{G}(a) - G(a)\right\|_{L^2(D)^{d_{out}}}. \tag{B.9}$$

Using (B.7) and (B.8), we then obtain

$$\begin{aligned} &\left\|\widetilde{G}(a) - G(a)\right\|_{L^2(D)^{d_{out}}} = \left\|\pi_{d_{out}} \circ H(a) - \pi_{d_{out}} \circ Q_\alpha \circ P_{V_\alpha^\perp} \circ H(a)\right\|_{L^2(D)^{d_{out}}} \\ &\leq \left\|\pi_{d_{out}} \circ (P_{V_0^\perp} - P_{V_\alpha^\perp} + P_{V_\alpha^\perp}) \circ H(a) - \pi_{d_{out}} \circ Q_\alpha \circ P_{V_\alpha^\perp} \circ H(a)\right\|_{L^2(D)^{d_{out}}} \\ &\leq \left\|\pi_{d_{out}} \circ (P_{V_0^\perp} - P_{V_\alpha^\perp}) \circ H(a)\right\|_{L^2(D)^{d_{out}}} + \left\|\pi_{d_{out}} \circ (I - Q_\alpha) \circ P_{V_\alpha^\perp} \circ H(a)\right\|_{L^2(D)^{d_{out}}} \\ &\leq 5\epsilon_0 \|H(a)\|_{L^2(D)^{d_{in}+d_{out}}} \leq \frac{\epsilon}{2}. \end{aligned} \tag{B.10}$$

Combining (B.9) and (B.10), we conclude that

$$\sup_{a\in K}\left\|G^{+}(a)-G(a)\right\|_{L^2(D)^{d_{out}}}\leq \frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon.$$

□

B.4 Remark following Theorem 1

Remark 4. *We make the following observations using Theorem 1:*

(i) ReLU and Leaky ReLU functions belong to $\mathrm{A}_0^L \cap \mathrm{BA}$ due to the fact that $\{\sigma \in C(\mathbb{R}) \mid \sigma$ is not a polynomial$\} \subseteq A_0$ (see Pinkus [1999]), and both the ReLU and Leaky ReLU functions belong to BA (see Lemma C.2 in Lanthaler et al. [2022]). We note that Lemma C.2 in Lanthaler et al. [2022] solely established the case for ReLU. However, it holds true for Leaky ReLU as well since the proof relies on the fact that the function $x \mapsto \min(\max(x,R),R)$ can be exactly represented by a two-layer ReLU neural network, and a two-layer Leaky ReLU neural network can also represent this function. Consequently, Leaky ReLU is one of example that satisfies (ii) in Theorem 1.

(ii) We emphasize that our infinite-dimensional result, Theorem 1, deviates from the finite-dimensional result. Puthawala et al. [2022a, Theorem 15] assumes that $2d_{in}+1 \leq d_{out}$ due to the use of Whitney's theorem. In contrast, Theorem 1 does not assume any conditions on d_{in} and d_{out}, that is, we are able to avoid invoking Whitney's theorem by employing Lemma 1.

(iii) We provide examples that injective universality does not hold when $L^2(D)^{d_{in}}$ and $L^2(D)^{d_{out}}$ are replaced by $\mathbb{R}^{d_{in}}$ and $\mathbb{R}^{d_{out}}$: Consider the case where $d_{in}=d_{out}=1$ and $G^+:\mathbb{R}\to\mathbb{R}$ is defined as $G^+(x)=\sin(x)$. We can not approximate $G^+:\mathbb{R}\to\mathbb{R}$ by an injective function $G:\mathbb{R}\to\mathbb{R}$ in the set $K=[0,2\pi]$ in the L^∞-norm. According to the topological degree theory (see Cho and Chen [2006, Theorem 1.2.6(iii)]), any continuous function $G:\mathbb{R}\to\mathbb{R}$ which satisfies $\|G-G^+\|_{C([0,2\pi])}<\varepsilon$ satisfies the equation on both intervals $I_1=[0,\pi]$, $I_2=[\pi,2\pi]$ $deg(G,I_j,s)=deg(G^+,I_j,s)=1$ for all $s\in[-1+\varepsilon,1-\varepsilon]$, $j=1,2$. This implies that $G:I_j\to\mathbb{R}$ obtains the value $s\in[-1+\varepsilon,1-\varepsilon]$ at least once. Hence, G obtains the values $s\in[-1+\varepsilon,1-\varepsilon]$ at least two times on the interval $[0,2\pi]$ and is it thus not injective. It is worth noting that the degree theory exhibits significant differences between the infinite-dimensional and finite-dimensional cases [Cho and Chen, 2006]).

C Details of Section 3.3

C.1 Finite rank approximation

We consider linear integral operators K_ℓ with L^2 kernels $k_\ell(x,y)$. Let $\{\varphi_k\}_{k\in\mathbb{N}}$ be an orthonormal basis in $L^2(D)$. Since $\{\varphi_k(y)\varphi_p(x)\}_{k,p\in\mathbb{N}}$ is an orthonormal basis of $L^2(D\times D)$, integral kernels $k_\ell \in L^2(D\times D;\mathbb{R}^{d_{\ell+1}\times d_\ell})$ in integral operators $K_\ell\in\mathcal{L}(L^2(D)^{d_\ell},L^2(D)^{d_{\ell+1}})$ has the expansion

$$k_\ell(x,y)=\sum_{k,p\in\mathbb{N}} C_{k,p}^{(\ell)}\varphi_k(y)\varphi_p(x),$$

then integral operators $K_\ell\in\mathcal{L}(L^2(D)^{d_\ell},L^2(D)^{d_{\ell+1}})$ take the form

$$K_\ell u(x)=\sum_{k,p\in\mathbb{N}} C_{k,p}^{(\ell)}(u,\varphi_k)\varphi_p(x),\ u\in L^2(D)^{d_\ell},$$

where $C_{k,p}^{(\ell)}\in\mathbb{R}^{d_{\ell+1}\times d_\ell}$ whose (i,j)-th component $c_{k,p,ij}^{(\ell)}$ is given by

$$c_{k,p,ij}^{(\ell)}=(k_{\ell,ij},\varphi_k\varphi_p)_{L^2(D\times D)}.$$

Here, we write $(u,\varphi_k)\in\mathbb{R}^{d_\ell}$,

$$(u,\varphi_k)=\left((u_1,\varphi_k)_{L^2(D)},...,(u_{d_\ell},\varphi_k)_{L^2(D)}\right).$$

We define $K_{\ell,N} \in \mathcal{L}(L^2(D)^{d_\ell}, L^2(D)^{d_{\ell+1}})$ as the truncated expansion of K_ℓ by N finite sum, that is,

$$K_{\ell,N}u(x) := \sum_{k,p \leq N} C^{(\ell)}_{k,p}(u, \varphi_k)\varphi_p(x).$$

Then $K_{\ell,N} \in \mathcal{L}(L^2(D)^{d_\ell}, L^2(D)^{d_{\ell+1}})$ is a finite rank operator with rank N. Furthermore, we have

$$\|K_\ell - K_{\ell,N}\|_{\mathrm{op}} \leq \|K_\ell - K_{\ell,N}\|_{\mathrm{HS}} = \left(\sum_{k,p \geq N} \sum_{i,j} |c^{(\ell)}_{k,p,ij}|^2 \right)^{1/2}, \tag{C.1}$$

which implies that as $N \to \infty$,

$$\|K_\ell - K_{\ell,N}\|_{\mathrm{op}} \to 0.$$

C.2 Layerwise injectivity

We first revisit layerwise injectivity and bijectivity in the case of the finite rank approximation. Let $K_N : L^2(D)^n \to L^2(D)^m$ be a finite rank operator defined by

$$K_N u(x) := \sum_{k,p \leq N} C_{k,p}(u, \varphi_k)\varphi_p(x), \; u \in L^2(D)^n,$$

where $C_{k,p} \in \mathbb{R}^{m \times n}$ and $(u, \varphi_p) \in \mathbb{R}^n$ is given by

$$(u, \varphi_p) = \left((u_1, \varphi_p)_{L^2(D)}, ..., (u_n, \varphi_p)_{L^2(D)} \right).$$

Let $b_N \in L^2(D)^n$ be defined by

$$b_N(x) := \sum_{p \leq N} b_p \varphi_p(x),$$

in which $b_p \in \mathbb{R}^m$. As analogues of Propositions 1 and 2, we obtain the following characterization.

Proposition 5. *(i) The operator*

$$\mathrm{ReLU} \circ (K_N + b_N) : (\mathrm{span}\{\varphi_k\}_{k \leq N})^n \to L^2(D)^m,$$

is injective if and only if for every $v \in (\mathrm{span}\{\varphi_k\}_{k \leq N})^n$,

$$\left\{ u \in L^2(D)^n \;\middle|\; \vec{u}_N \in \mathrm{Ker}(C_{S,N}) \right\} \cap X(v, K_N + b_N) \cap (\mathrm{span}\{\varphi_k\}_{k \leq N})^n = \{0\}.$$

where $S(v, K_N + b_N) \subset [m]$ and $X(v, K_N + b_N)$ are defined in Definition 2, and

$$\vec{u}_N := ((u, \varphi_p))_{p \leq N} \in \mathbb{R}^{Nn}, \;\; C_{S,N} := \left(C_{k,q}\big|_{S(v, K_N + b_N)} \right)_{k,q \in [N]} \in \mathbb{R}^{N|S(v,K_N+b_N)| \times Nn}. \tag{C.2}$$

(ii) Let σ be injective. Then the operator

$$\sigma \circ (K_N + b_N) : (\mathrm{span}\{\varphi_k\}_{k \leq N})^n \to L^2(D)^m,$$

is injective if and only if C_N is injective, where

$$C_N := (C_{k,q})_{k,q \in [N]} \in \mathbb{R}^{Nm \times Nn}. \tag{C.3}$$

Proof. The above statements follow from Propositions 1 and 2 by observing that $u \in \mathrm{Ker}\,(K_N)$ is equivalent to (cf. (C.2) and (C.3))

$$\sum_{k,p \leq N} C_{k,p}(u, \varphi_k)\varphi_p = 0, \iff C_N \vec{u}_N = 0.$$

□

C.3 Global injectivity

We revisit global injectivity in the case of finite rank approximation. As an analogue of Lemma 1, we have the following

Lemma 2. *Let $N, N' \in \mathbb{N}$ and $n, m, \ell \in \mathbb{N}$ with $N'm > N'\ell \geq 2Nn + 1$, and let $T : L^2(D)^n \to L^2(D)^m$ be a finite rank operator with N' rank, that is,*

$$\mathrm{Ran}(T) \subset (\mathrm{span}\{\varphi_k\}_{k \leq N'})^m, \tag{C.4}$$

and Lipschitz continuous, and

$$T : (\mathrm{span}\{\varphi_k\}_{k \leq N})^n \to L^2(D)^m,$$

is injective. Then, there exists a finite rank operator $B \in \mathcal{L}(L^2(D)^m, L^2(D)^\ell)$ with rank N' such that

$$B \circ T : (\mathrm{span}\{\varphi_k\}_{k \leq N})^n \to (\mathrm{span}\{\varphi_k\}_{k \leq N'})^\ell,$$

is injective.

Proof. From (C.4), $T : L^2(D)^n \to L^2(D)^m$ has the form of

$$T(a) = \sum_{k \leq N'} (T(a), \varphi_k)\varphi_k,$$

where $(T(a), \varphi_k) \in \mathbb{R}^m$. We define $\mathbf{T} : \mathbb{R}^{Nn} \to \mathbb{R}^{N'm}$ by

$$\mathbf{T}(\mathbf{a}) := ((T(\mathbf{a}), \varphi_k))_{k \in [N']} \in \mathbb{R}^{N'm}, \ \mathbf{a} \in \mathbb{R}^{Nn},$$

where $T(\mathbf{a}) \in L^2(D)^m$ is defined by

$$T(\mathbf{a}) := T\left(\sum_{k \leq N} a_k \varphi_k\right) \in L^2(D)^m,$$

in which $a_k \in \mathbb{R}^n$, $\mathbf{a} = (a_1, ..., a_N) \in \mathbb{R}^{Nn}$.

Since $T : L^2(D)^n \to L^2(D)^m$ is Lipschitz continuous, $\mathbf{T} : \mathbb{R}^{Nn} \to \mathbb{R}^{N'm}$ is also Lipschitz continuous. As $N'm > N'\ell \geq 2Nn + 1$, we can apply Lemma 29 from Puthawala et al. [2022a] with $D = N'm$, $m = N'\ell$, $n = Nn$. According to this lemma, there exists a $N'\ell$-dimensional linear subspace $\mathbf{V}^\perp$ in $\mathbb{R}^{N'm}$ such that

$$\left\| P_{\mathbf{V}^\perp} - P_{\mathbf{V}_0^\perp} \right\|_{\mathrm{op}} < 1,$$

and

$$P_{\mathbf{V}^\perp} \circ \mathbf{T} : \mathbb{R}^{Nn} \to \mathbb{R}^{N'm},$$

is injective, where $\mathbf{V}_0^\perp = \{0\}^{N'(m-\ell)} \times \mathbb{R}^{N'\ell}$. Furthermore, in the proof of Theorem 15 of Puthawala et al. [2022a], denoting

$$\mathbf{B} := \pi_{N'\ell} \circ \mathbf{Q} \circ P_{\mathbf{V}^\perp} \in \mathbb{R}^{N'\ell \times N'm},$$

we are able to show that

$$\mathbf{B} \circ \mathbf{T} : \mathbb{R}^{Nn} \to \mathbb{R}^{N'\ell}$$

is injective. Here, $\pi_{N'\ell} : \mathbb{R}^{N'm} \to \mathbb{R}^{N'\ell}$

$$\pi_{N'\ell}(a, b) := b, \quad (a, b) \in \mathbb{R}^{N'(m-\ell)} \times \mathbb{R}^{N'\ell},$$

where $\mathbf{Q} : \mathbb{R}^{N'm} \to \mathbb{R}^{N'm}$ is defined by

$$\mathbf{Q} := \left(P_{\mathbf{V}_0^\perp} P_{\mathbf{V}^\perp} + (I - P_{\mathbf{V}_0^\perp})(I - P_{\mathbf{V}^\perp})\right)\left(I - (P_{\mathbf{V}_0^\perp} - P_{\mathbf{V}^\perp})^2\right)^{-1/2}.$$

We define $B : L^2(D)^m \to L^2(D)^\ell$ by

$$Bu = \sum_{k,p \leq N'} \mathbf{B}_{k,p}(u, \varphi_k)\varphi_p,$$

where $\mathbf{B}_{k,p} \in \mathbb{R}^{\ell \times m}$, $\mathbf{B} = (\mathbf{B}_{k,p})_{k,p \in [N']}$. Then $B : L^2(D)^m \to L^2(D)^\ell$ is a linear finite rank operator with N' rank, and

$$B \circ T : L^2(D)^n \to L^2(D)^\ell$$

is injective because, by the construction, it is equivalent to

$$\mathbf{B} \circ \mathbf{T} : \mathbb{R}^{Nn} \to \mathbb{R}^{N'\ell},$$

is injective. □

C.4 Proof of Theorem 2

Definition 4. *We define the set of integral neural operators with N rank by*

$$\begin{aligned}
&\mathrm{NO}_{L,N}(\sigma; D, d_{in}, d_{out}) := \Big\{ G_N : L^2(D)^{d_{in}} \to L^2(D)^{d_{out}} \Big| \\
&G_N = K_{L+1,N} \circ (K_{L,N} + b_{L,N}) \circ \sigma \cdots \circ (K_{2,N} + b_{2,N}) \circ \sigma \circ (K_{1,N} + b_{1,N}) \circ (K_{0,N} + b_{0,N}), \\
&K_{\ell,N} \in \mathcal{L}(L^2(D)^{d_\ell}, L^2(D)^{d_{\ell+1}}),\ K_{\ell,N} : f \mapsto \sum_{k,p \leq N} C_{k,p}^{(\ell)} (f, \varphi_k) \varphi_p, \\
&b_{\ell,N} \in L^2(D; \mathbb{R}^{d_{\ell+1}}),\ b_{\ell,N} = \sum_{p \leq N} b_p^{(\ell)} \varphi_m \\
&C_{k,p}^{(\ell)} \in \mathbb{R}^{d_{\ell+1} \times d_\ell},\ b_p^{(\ell)} \in \mathbb{R}^{d_{\ell+1}},\ k, p \leq N, \\
&d_\ell \in \mathbb{N},\ d_0 = d_{in},\ d_{L+2} = d_{out},\ \ell = 0, ..., L+2 \Big\}.
\end{aligned} \tag{C.5}$$

Proof. Let $R > 0$ such that

$$K \subsetneq B_R(0),$$

where $B_R(0) := \{u \in L^2(D)^{d_{in}} \mid \|u\|_{L^2(D)^{d_{in}}} \leq R\}$. As ReLU and Leaky ReLU function belongs to $\mathrm{A}_0^L \cap \mathrm{BA}$, by Theorem 11 of Kovachki et al. [2021b], there exists $L \in \mathbb{N}$ and $\widetilde{G} \in \mathrm{NO}_L(\sigma; D, d_{in}, d_{out})$ such that

$$\sup_{a \in K} \left\| G^+(a) - \widetilde{G}(a) \right\|_{L^2(D)^{d_{out}}} \leq \frac{\epsilon}{3}. \tag{C.6}$$

and

$$\left\| \widetilde{G}(a) \right\|_{L^2(D)^{d_{out}}} \leq 4M, \quad \text{for } a \in L^2(D)^{d_{in}}, \quad \|a\|_{L^2(D)^{d_{in}}} \leq R.$$

We write operator $\widetilde{G}$ by

$$\widetilde{G} = \widetilde{K}_{L+1} \circ (\widetilde{K}_L + \widetilde{b}_L) \circ \sigma \cdots \circ (\widetilde{K}_2 + \widetilde{b}_2) \circ \sigma \circ (\widetilde{K}_1 + \widetilde{b}_1) \circ (\widetilde{K}_0 + \widetilde{b}_0),$$

where

$$\begin{aligned}
&\widetilde{K}_\ell \in \mathcal{L}(L^2(D)^{d_\ell}, L^2(D)^{d_{\ell+1}}),\ \widetilde{K}_\ell : f \mapsto \int_D \widetilde{k}_\ell(\cdot, y) f(y) dy, \\
&\widetilde{k}_\ell \in L^2(D \times D; \mathbb{R}^{d_{\ell+1} \times d_\ell}),\ \widetilde{b}_\ell \in L^2(D; \mathbb{R}^{d_{\ell+1}}), \\
&d_\ell \in \mathbb{N},\ d_0 = d_{in},\ d_{L+2} = d_{out},\ \ell = 0, ..., L+2.
\end{aligned}$$

We set $\widetilde{G}_{N'} \in \mathrm{NO}_{L,N'}(\sigma; D, d_{in}, d_{out})$ such that

$$\widetilde{G}_{N'} = \widetilde{K}_{L+1,N'} \circ (\widetilde{K}_{L,N'} + \widetilde{b}_{L,N'}) \circ \sigma \cdots \circ (\widetilde{K}_{2,N'} + \widetilde{b}_{2,N'}) \circ \sigma \circ (\widetilde{K}_{1,N'} + \widetilde{b}_{1,N'}) \circ (\widetilde{K}_{0,N'} + \widetilde{b}_{0,N'}),$$

where $\widetilde{K}_{\ell,N'} : L^2(D)^{d_\ell} \to L^2(D)^{d_{\ell+1}}$ is defined by

$$\widetilde{K}_{\ell,N'} u(x) = \sum_{k,p \leq N'} C_{k,p}^{(\ell)} (u, \varphi_k) \varphi_p(x),$$

where $C^{(\ell)}_{k,p} \in \mathbb{R}^{d_{\ell+1}\times d_\ell}$ whose (i,j)-th component $c^{(\ell)}_{k,p,ij}$ is given by

$$c^{(\ell)}_{k,p,ij} = (\widetilde{k}_{\ell,ij}, \varphi_k\varphi_p)_{L^2(D\times D)}.$$

Since

$$\left\|\widetilde{K}_\ell - \widetilde{K}_{\ell,N'}\right\|^2_{\mathrm{op}} \le \left\|\widetilde{K}_\ell - \widetilde{K}_{\ell,N'}\right\|^2_{\mathrm{HS}} = \sum_{k,p\ge N'+1}\sum_{i,j}|c^{(\ell)}_{k,p,ij}|^2 \to 0 \text{ as } N' \to \infty,$$

there is a large $N' \in \mathbb{N}$ such that

$$\sup_{a\in K}\left\|\widetilde{G}(a) - \widetilde{G}_{N'}(a)\right\|_{L^2(D)^{d_{out}}} \le \frac{\epsilon}{3}. \tag{C.7}$$

Then, we have

$$\begin{aligned}\sup_{a\in K}\left\|\widetilde{G}_{N'}(a)\right\|_{L^2(D)^{d_{out}}} &\le \sup_{a\in K}\left\|\widetilde{G}_{N'}(a) - \widetilde{G}(a)\right\|_{L^2(D)^{d_{out}}} + \sup_{a\in K}\left\|\widetilde{G}(a)\right\|_{L^2(D)^{d_{out}}} \\ &\le 1 + 4M.\end{aligned}$$

We define the operator $H_{N'} : L^2(D)^{d_{in}} \to L^2(D)^{d_{in}+d_{out}}$ by

$$H_{N'}(a) = \begin{pmatrix} H_{N'}(a)_1 \\ H_{N'}(a)_2 \end{pmatrix} := \begin{pmatrix} K_{inj,N}\circ\cdots\circ K_{inj,N}(a) \\ \widetilde{G}_{N'}(a) \end{pmatrix},$$

where $K_{inj,N} : L^2(D)^{d_{in}} \to L^2(D)^{d_{in}}$ is defined by

$$K_{inj,N}u = \sum_{k\le N}(u,\varphi_k)\varphi_k.$$

As $K_{inj,N}\,(\mathrm{span}\{\varphi_k\}_{k\le N})^{d_{in}} \to L^2(D)^{d_{in}}$ is injective,

$$H_{N'} : (\mathrm{span}\{\varphi_k\}_{k\le N})^{d_{in}} \to (\mathrm{span}\{\varphi_k\}_{k\le N})^{d_{in}} \times (\mathrm{span}\{\varphi_k\}_{k\le N'})^{d_{out}},$$

is injective. Furthermore, by the same argument (ii) (construction of H) in the proof of Theorem 1,

$$H_{N'} \in NO_{L,N'}(\sigma; D, d_{in}, d_{out}),$$

because both of two-layer ReLU and Leaky ReLU neural networks can represent the identity map. Note that above $K_{inj,N}$ is an orthogonal projection, so that $K_{inj,N}\circ\cdots\circ K_{inj,N} = K_{inj,N}$. However, we write above $H_{N'}(a)_1$ as $K_{inj,N}\circ\cdots\circ K_{inj,N}(a)$ so that it can be considered as combination of $(L+2)$ layers of neural networks.

We estimate that for $a \in L^2(D)^{d_{in}}$, $\|a\|_{L^2(D)^{d_{in}}} \le R$,

$$\|H_{N'}(a)\|_{L^2(D)^{d_{in}+d_{out}}} \le 1 + 4M + \|K_{inj}\|^{L+2}_{\mathrm{op}}\,R =: C_H.$$

Here, we repeat an argument similar to the one in the proof of Lemma 2: $H_{N'} : L^2(D)^{d_{in}} \to L^2(D)^{d_{in}+d_{out}}$ has the form of

$$H_{N'}(a) = \left(\sum_{k\le N}(H_{N'}(a)_1,\varphi_k)\varphi_k,\ \sum_{k\le N'}(H_{N'}(a)_2,\varphi_k)\varphi_k\right).$$

where $(H_{N'}(a)_1,\varphi_k) \in \mathbb{R}^{d_{in}}$, $(H_{N'}(a)_2,\varphi_k) \in \mathbb{R}^{d_{out}}$. We define $\mathbf{H}_{N'} : \mathbb{R}^{Nd_{in}} \to \mathbb{R}^{Nd_{in}+N'd_{out}}$ by

$$\mathbf{H}_{N'}(\mathbf{a}) := \left[((H_{N'}(\mathbf{a})_1,\varphi_k))_{k\in[N]},\ ((H_{N'}(\mathbf{a})_2,\varphi_k))_{k\in[N']}\right] \in \mathbb{R}^{Nd_{in}+N'd_{out}},\ \mathbf{a} \in \mathbb{R}^{Nd_{in}},$$

where $H_{N'}(\mathbf{a}) = (H_{N'}(\mathbf{a})_1, H_{N'}(\mathbf{a})_2) \in L^2(D)^{d_{in}+d_{out}}$ is defined by

$$H_{N'}(\mathbf{a})_1 := H_{N'}\left(\sum_{k\le N}a_k\varphi_k\right)_1 \in L^2(D)^{d_{in}},$$

$$H_{N'}(\mathbf{a})_2 := H_{N'}\left(\sum_{k\le N'} a_k\varphi_k\right)_2 \in L^2(D)^{d_{out}},$$

where $a_k \in \mathbb{R}^{d_{in}}$, $\mathbf{a} = (a_1, ..., a_N) \in \mathbb{R}^{Nd_{in}}$. Since $H_{N'} : L^2(D)^{d_{in}} \to L^2(D)^{d_{in}+d_{out}}$ is Lipschitz continuous, $\mathbf{H}_{N'} : \mathbb{R}^{Nd_{in}} \to \mathbb{R}^{N'd_{out}}$ is also Lipschitz continuous. As

$$Nd_{in} + N'd_{out} > N'd_{out} \ge 2Nd_{in} + 1,$$

we can apply Lemma 29 of Puthawala et al. [2022a] with $D = Nd_{in} + N'd_{out}$, $m = N'd_{out}$, $n = Nd_{in}$. According to this lemma, there exists a $N'd_{out}$-dimensional linear subspace $\mathbf{V}^{\perp}$ in $\mathbb{R}^{Nd_{in}+N'd_{out}}$ such that

$$\left\|P_{\mathbf{V}^{\perp}} - P_{\mathbf{V}_0^{\perp}}\right\|_{op} < \min\left(\frac{\epsilon}{15C_{H_N}},\, 1\right) =: \epsilon_0$$

and

$$P_{\mathbf{V}^{\perp}} \circ \mathbf{H}_{N'} : \mathbb{R}^{Nd_{in}} \to \mathbb{R}^{Nd_{in}+N'd_{out}},$$

is injective, where $\mathbf{V}_0^{\perp} = \{0\}^{Nd_{in}} \times \mathbb{R}^{N'd_{out}}$. Furthermore, in the proof of Theorem 15 of Puthawala et al. [2022a], denoting by

$$\mathbf{B} := \pi_{N'd_{out}} \circ \mathbf{Q} \circ P_{\mathbf{V}^{\perp}},$$

we can show that

$$\mathbf{B} \circ \mathbf{H}_{N'} : \mathbb{R}^{Nd_{in}} \to \mathbb{R}^{N'd_{out}},$$

is injective, where $\pi_{N'd_{out}} : \mathbb{R}^{Nd_{in}+N'd_{out}} \to \mathbb{R}^{N'd_{out}}$

$$\pi_{N'd_{out}}(a, b) := b, \quad (a, b) \in \mathbb{R}^{Nd_{in}} \times \mathbb{R}^{N'd_{out}},$$

and $\mathbf{Q} : \mathbb{R}^{Nd_{in}+N'd_{out}} \to \mathbb{R}^{Nd_{in}+N'd_{out}}$ is defined by

$$\mathbf{Q} := \left(P_{\mathbf{V}_0^{\perp}} P_{\mathbf{V}^{\perp}} + (I - P_{\mathbf{V}_0^{\perp}})(I - P_{\mathbf{V}^{\perp}})\right)\left(I - (P_{\mathbf{V}_0^{\perp}} - P_{\mathbf{V}^{\perp}})^2\right)^{-1/2}.$$

By the same argument in proof of Theorem 15 in Puthawala et al. [2022a], we can show that

$$\|I - \mathbf{Q}\|_{\mathrm{op}} \le 4\epsilon_0.$$

We define $B : L^2(D)^{d_{in}+d_{out}} \to L^2(D)^{d_{out}}$

$$Bu = \sum_{k,p\le N'} \mathbf{B}_{k,p}(u, \varphi_k)\varphi_p,$$

$\mathbf{B}_{k,p} \in \mathbb{R}^{d_{out}\times(d_{in}+d_{out})}$, $\mathbf{B} = (\mathbf{B}_{k,p})_{k,p\in[N']}$, then $B : L^2(D)^{d_{in}+d_{out}} \to L^2(D)^{d_{out}}$ is a linear finite rank operator with N' rank. Then,

$$G_{N'} := B \circ H_{N'} : L^2(D)^{d_{in}} \to L^2(D)^{d_{out}},$$

is injective because by the construction, it is equivalent to

$$\mathbf{B} \circ \mathbf{H}_{N'} : \mathbb{R}^{Nd_{in}} \to \mathbb{R}^{N'd_{out}},$$

is injective. Furthermore, we have

$$G_{N'} \in NO_{L,N'}(\sigma; D, d_{in}, d_{out}).$$

Indeed, $H_{N'} \in NO_{L,N'}(\sigma; D, d_{in}, d_{out})$, B is the linear finite rank operator with N' rank, and multiplication of two linear finite rank operators with N' rank is also a linear finite rank operator with N' rank.

Finally, we estimate for $a \in K$,

$$\begin{aligned}
&\left\|G^+(a) - G_{N'}(a)\right\|_{L^2(D)^{d_{out}}} \\
&= \underbrace{\left\|G^+(a) - \widetilde{G}(a)\right\|_{L^2(D)^{d_{out}}}}_{(C.6)\le\frac{\epsilon}{3}} + \underbrace{\left\|\widetilde{G}(a) - \widetilde{G}_{N'}(a)\right\|_{L^2(D)^{d_{out}}}}_{(C.7)\le\frac{\epsilon}{3}} + \left\|\widetilde{G}_{N'}(a) - G_{N'}(a)\right\|_{L^2(D)^{d_{out}}}.
\end{aligned} \tag{C.8}$$

Using notation $(a,\varphi_k)\in\mathbb{R}^{d_{in}}$, and $\mathbf{a}=((a,\varphi_k))_{k\in[N]}\in\mathbb{R}^{Nd_{in}}$, we further estimate for $a\in K$,

$$
\begin{aligned}
&\left\|\widetilde{G}_{N'}(a)-G_{N'}(a)\right\|_{L^2(Q)^{d_{out}}} = \|\pi_{d_{out}}H_{N'}(a)-B\circ H_{N'}(a)\|_{L^2(Q)^{d_{out}}}\\
&= \|\pi_{N'd_{out}}\mathbf{H}_{N'}(\mathbf{a})-\mathbf{B}\circ\mathbf{H}_{N'}(\mathbf{a})\|_2\\
&= \|\pi_{N'd_{out}}\circ\mathbf{H}_{N'}(\mathbf{a})-\pi_{N'd_{out}}\circ\mathbf{Q}\circ P_{\mathbf{V}^\perp}\circ\mathbf{H}_{N'}(\mathbf{a})\|_2\\
&\le \left\|\pi_{N'd_{out}}\circ(P_{\mathbf{V}_0^\perp}-P_{\mathbf{V}^\perp}+P_{\mathbf{V}^\perp})\circ\mathbf{H}_{N'}(\mathbf{a})-\pi_{N'd_{out}}\circ\mathbf{Q}\circ P_{\mathbf{V}^\perp}\circ\mathbf{H}_{N'}(\mathbf{a})\right\|_2\\
&\le \left\|\pi_{N'd_{out}}\circ(P_{\mathbf{V}_0^\perp}-P_{\mathbf{V}^\perp})\circ\mathbf{H}_{N'}(\mathbf{a})\right\|_2+\|\pi_{N'd_{out}}\circ(I-\mathbf{Q})\circ P_{\mathbf{V}^\perp}\circ\mathbf{H}_{N'}(\mathbf{a})\|_2\\
&\le 5\epsilon_0 \underbrace{\|\mathbf{H}_{N'}(\mathbf{a})\|_2}_{=\|H_{N'}(a)\|_{L^2(D)^{d_{out}}}<C_H} \le \frac{\epsilon}{3},
\end{aligned}
\tag{C.9}
$$

where $\|\cdot\|_2$ is the Euclidean norm. Combining (C.8) and (C.9), we conclude that

$$\sup_{a\in K}\left\|G^+(a)-G_{N'}(a)\right\|_{L^2(D)^{d_{out}}}\le\frac{\epsilon}{3}+\frac{\epsilon}{3}+\frac{\epsilon}{3}=\epsilon.$$

□

D Details of Section 4.1

D.1 Proof of Proposition 3

Proof. Since W is bijective, and σ is surjective, it is enough to show that $u\mapsto Wu+K(u)$ is surjective. We observe that for $z\in L^2(D)^n$,

$$Wu+K(u)=z,$$

is equivalent to

$$H_z(u):=-W^{-1}K(u)+W^{-1}z=u.$$

We will show that $H_z:L^2(D)^n\to L^2(D)^n$ has a fixed point for each $z\in L^2(D)^n$. By the Leray-Schauder theorem, see Gilbarg and Trudinger [2001, Theorem 11.3], $H:L^2(D)\to L^2(D)$ has a fixed point if the union $\bigcup_{0<\lambda\le1}V_\lambda$ is bounded, where the sets

$$
\begin{aligned}
V_\lambda &:= \{u\in L^2(D):\ u=\lambda H_z(u)\}\\
&= \{u\in L^2(D):\ \lambda^{-1}u=H_z(u)\}\\
&= \{u\in L^2(D):\ -\lambda^{-1}u=W^{-1}K(u)-W^{-1}z\},
\end{aligned}
$$

are parametrized by $0<\lambda\le1$.

As the map $u\mapsto\alpha u+W^{-1}K(u)$ is coercive, there is an $r>0$ such that for $\|u\|_{L^2(D)^n}>r$,

$$\frac{\left\langle\alpha u+W^{-1}K(u),u\right\rangle_{L^2(D)^n}}{\|u\|_{L^2(D)^n}}\ge\|W^{-1}z\|_{L^2(D)^n}.$$

Thus, we have that for $\|u\|_{L^2(D)^n}>r$

$$
\begin{aligned}
&\frac{\left\langle W^{-1}K(u)-W^{-1}z,u\right\rangle_{L^2(D)^n}}{\|u\|^2_{L^2(D)^n}}\\
&\ge\frac{\left\langle\alpha u+W^{-1}K(u),u\right\rangle_{L^2(D)^n}-\left\langle\alpha u+W^{-1}z,u\right\rangle_{L^2(D)^n}}{\|u\|^2_{L^2(D)^n}}\\
&\ge\frac{\|W^{-1}z\|_{L^2(D)^n}}{\|u\|_{L^2(D)^n}}-\frac{\left\langle W^{-1}z,u\right\rangle_{L^2(D)^n}}{\|u\|^2_{L^2(D)^n}}-\alpha\ge-\alpha>-1,
\end{aligned}
$$

and, hence, for all $\|u\|_{L^2(D)} > r_0$ and $\lambda \in (0,1]$ we have $u \notin V_\lambda$. Thus

$$\bigcup_{\lambda \in (0,1]} V_\lambda \subset B(0, r_0).$$

Again, by the Leray-Schauder theorem (see Gilbarg and Trudinger [2001, Theorem 11.3]), H_z has a fixed point. □

D.2 Examples for Proposition 3

Example 2. *We consider the case where $n = 1$ and $D \subset \mathbb{R}^d$ is a bounded interval. We consider the non-linear integral operator,*

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

and $k(x, y, t)$ is bounded, that is, there is $C_K > 0$ such that

$$|k(x, y, t)| \le C_K, \ x, y \in D, \ t \in \mathbb{R}.$$

If $\left\|W^{-1}\right\|_{\mathrm{op}}$ is small enough such that

$$1 > \left\|W^{-1}\right\|_{\mathrm{op}} C_K |D|,$$

then, for $\alpha \in \left(\left\|W^{-1}\right\|_{\mathrm{op}} C_K |D|, 1\right)$, $u \mapsto \alpha u + W^{-1}K(u)$ is coercive. Indeed, we have for $u \in L^2(D)$,

$$\begin{aligned}
&\frac{\left\langle \alpha u + W^{-1}K(u), u \right\rangle_{L^2(D)}}{\|u\|_{L^2(D)}} \\
&\ge \alpha \|u\|_{L^2(D)} - \left\|W^{-1}\right\|_{\mathrm{op}} \|K(u)\|_{L^2(D)} \ge \underbrace{\left(\alpha - \left\|W^{-1}\right\|_{\mathrm{op}} C_K |D|\right)}_{>0} \|u\|_{L^2(D)}.
\end{aligned}$$

For example, we can consider a kernel

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

where $\sigma_s : \mathbb{R} \to \mathbb{R}$ is the sigmoid function defined by

$$\sigma_s(t) = \frac{1}{1 + e^{-t}}.$$

There are functions $a, b, c \in C(\overline{D} \times \overline{D})$ such that

$$\sum_{j=1}^{J} \|c_j\|_{L^\infty(D \times D)} < \left\|W^{-1}\right\|_{\mathrm{op}}^{-1} |D|^{-1}.$$

Example 3. *Again, we consider the case where $n = 1$ and $D \subset \mathbb{R}^d$ is a bounded set. We assume that $W \in C^1(\overline{D})$ satisfies $0 < c_1 \le W(x) \le c_2$. For simplicity, we assume that $|D| = 1$. We consider the non-linear integral operator*

$$K(u)(x) := \int_D k(x, y, u(x))u(y)dy, \ x \in D, \tag{D.1}$$

where

$$k(x, y, t) = \sum_{j=1}^{J} c_j(x, y)\sigma_{wire}(a_j(x, y)t + b_j(x, y)), \tag{D.2}$$

in which $\sigma_{wire} : \mathbb{R} \to \mathbb{R}$ is the wavelet function defined by

$$\sigma_{wire}(t) = \mathit{Im}\,(e^{i\omega t} e^{-t^2}),$$

and $a_j, b_j, c_j \in C(\overline{D} \times \overline{D})$ are such that the $a_j(x, y)$ are nowhere vanishing functions, that is, $a_j(x, y) \neq 0$ for all $x, y \in \overline{D} \times \overline{D}$.

The next lemma holds for any activation function with exponential decay, including the activation function σ_{wire} and settles the key condition for Proposition 3 to hold.

Lemma 3. *Assume that $|D| = 1$ and the activation function $\sigma : \mathbb{R} \to \mathbb{R}$ is continuous. Assume that there exists $M_1, m_0 > 0$ such that*

$$|\sigma(t)| \leq M_1 e^{-m_0|t|}, \quad t \in \mathbb{R}.$$

Let $a_j, b_j, c_j \in C(\overline{D} \times \overline{D})$ be such that $a_j(x, y)$ are nowhere vanishing functions. Moreover, let $K : L^2(D) \to L^2(D)$ be a non-linear integral operator given in (D.1) *with a kernel satisfying* (D.2), *$\alpha > 0$ and $0 < c_0 \leq W(x) \leq c_1$. Then function $F : L^2(D) \to L^2(D)$, $F(u) = \alpha u + W^{-1}K(u)$ is coercive.*

Proof. As $\overline{D}$ is compact, there is $a_0 > 0$ such that for all $j = 1, 2, \ldots, J$ we have $|a_j(x, y)| \geq a_0$ a.e. and $|b_j(x, y)| \leq b_0$ a.e. We point out that $|\sigma(t)| \leq M_1$. Next, let $\varepsilon > 0$ be such that

$$(\sum_{j=1}^{J} \|W^{-1}c_j\|_{L^\infty(D\times D)})M_1\varepsilon < \frac{\alpha}{4}, \tag{D.3}$$

$\lambda > 0$, and $u \in L^2(D)$. We define the sets

$$D_1(\lambda) = \{x \in D : \ |u(x)| \geq \varepsilon\lambda\},$$
$$D_2(\lambda) = \{x \in D : \ |u(x)| < \varepsilon\lambda\}.$$

Then, for $x \in D_2(\lambda)$,

$$\sum_{j=1}^{J} \|W^{-1}c_j\|_{L^\infty(D\times D)} |\sigma(a_j(x, y)u(x) + b_j(x, y))u(x)|$$
$$\leq \sum_{j=1}^{J} \|W^{-1}c_j\|_{L^\infty(D\times D)} M_1 \epsilon\lambda \underset{(D.3)}{\leq} \frac{\alpha}{4}\lambda.$$

After ε is chosen as in the above, we choose $\lambda_0 \geq \max(1, b_0/(a_0\varepsilon))$ to be sufficiently large so that for all $|t| \geq \varepsilon\lambda_0$ it holds that

$$\left(\sum_{j=1}^{J} \|W^{-1}c_j\|_{L^\infty(D\times D)}\right) M_1 \mathrm{exp}(-m_0|a_0 t - b_0|)t < \frac{\alpha}{4}.$$

Here, we observe that, as $\lambda_0 \geq b_0/(a_0\varepsilon)$, we have that for all $|t| \geq \varepsilon\lambda_0$, $a_0|t| - b_0 > 0$. Then, when $\lambda \geq \lambda_0$, we have for $x \in D_1(\lambda)$,

$$\left(\sum_{j=1}^{J} \|W^{-1}c_j\|_{L^\infty(D\times D)}\right) \left|\sigma\Big(a_j(x, y)u(x) + b_j(x, y)\Big)u(x)\right| \leq \frac{\alpha}{4}.$$

When $u \in L^2(D)$ has the norm $\|u\|_{L^2(D)} = \lambda \geq \lambda_0 \geq 1$, we have

$$\begin{aligned}
& \left| \int_D \int_D W(x)^{-1} k(x, y, u(x)) u(x) u(y) dx dy \right| \\
\leq \ & \int_D \left(\int_{D_1} (\sum_{j=1}^{J} \|W^{-1} c_j\|_{L^\infty(D \times D)}) M_1 \mathrm{exp}\Big(-m_0 |a_0 |u(x)| - b_0| \Big) |u(x)| dx \right) |u(y)| dy \\
& + \int_D \left(\int_{D_2} \sum_{j=1}^{J} \|W^{-1} c_j\|_{L^\infty(D \times D)} |\sigma(a_j(x, y) u(x) + b_j(x, y))| |u(x)| dx \right) |u(y)| dy \\
\leq \ & \frac{\alpha}{4} \|u\|_{L^2(D)} + \frac{\alpha}{4} \lambda \|u\|_{L^2(D)} \\
\leq \ & \frac{\alpha}{2} \|u\|_{L^2(D)}^2.
\end{aligned}$$

Hence,

$$\frac{\langle \alpha u + W^{-1} K(u), u \rangle_{L^2(D)}}{\|u\|_{L^2(D)}} \geq \frac{\alpha}{2} \|u\|_{L^2(D)},$$

and the function $u \to \alpha u + W^{-1} K(u)$ is coercive. □

D.3 Proof of Proposition 4

Proof. (Injectivity) Assume that

$$\sigma(W u_1 + K(u_1) + b) = \sigma(W u_2 + K(u_2) + b).$$

where $u_1, u_2 \in L^2(D)^n$. Since σ is injective and $W : L^2(D)^n \to L^2(D)^n$ is bounded linear bijective, we have

$$u_1 + W^{-1} K(u_1) = u_2 + W^{-1} K(u_2) =: z.$$

Since the mapping $u \mapsto z - W^{-1} K(u)$ is contraction (because $W^{-1} K$ is contraction), by the Banach fixed-point theorem, the mapping $u \mapsto z - W^{-1} K(u)$ admit a unique fixed-point in $L^2(D)^n$, which implies that $u_1 = u_2$.

(Surjectivity) Since σ is surjective, it is enough to show that $u \mapsto Wu + K(u) + b$ is surjective. Let $z \in L^2(D)^n$. Since the mapping $u \mapsto W^{-1} z - W^{-1} b - W^{-1} K(u)$ is contraction, by Banach fixed-point theorem, there is $u^* \in L^2(D)^n$ such that

$$u^* = W^{-1} z - W^{-1} b - W^{-1} K(u^*) \iff W u^* + K(u^*) + b = z.$$

□

D.4 Examples for Proposition 4

Example 4. *We consider the case of $n = 1$, and $D \subset [0, \ell]^d$. We consider Volterra operators*

$$K(u)(x) = \int_D k(x, y, u(x), u(y)) u(y) dy,$$

where $x = (x_1, \ldots, x_d)$ and $y = (y_1, \ldots, y_d)$. We recall that K is a Volterra operator if

$$k(x, y, t, s) \neq 0 \implies y_j \leq x_j \quad \text{for all } j = 1, 2, \ldots, d. \tag{D.4}$$

In particular, when $D = (a, b) \subset \mathbb{R}$ is an interval, the Volterra operators are of the form

$$K(u)(x) = \int_a^x k(x, y, u(x), u(y)) u(y) dy,$$

and if x is considered as a time variable, the Volterra operators are causal in the sense that the value of $K(u)(x)$ at the time x depends only on $u(y)$ at the times $y \leq x$.

Assume that $k(x,y,t,s) \in C(\overline{D}\times\overline{D}\times\mathbb{R}\times\mathbb{R})$ *is bounded and uniformly Lipschitz smooth in the* t *and* s *variables, that is,* $k \in C(\overline{D}\times\overline{D}; C^{0,1}(\mathbb{R}\times\mathbb{R}))$.

Next, we consider the non-linear operator $F : L^2(D) \to L^2(D)$,

$$F(u) = u + K(u). \tag{D.5}$$

Assume that $u, w \in L^2(D)$ *are such that* $u + K(u) = w + K(w)$, *so that* $w - u = K(u) - K(w)$. *Next, we will show that then* $u = w$. *We denote and* $D(z_1) = D \cap ([0, z_1] \times [0,\ell]^{d-1})$ *and*

$$\|k\|_{C(\overline{D}\times\overline{D};C^{0,1}(\mathbb{R}\times\mathbb{R}))} := \sup_{x,y\in D} \|k(x,y,\cdot,\cdot)\|_{C^{0,1}(\mathbb{R}\times\mathbb{R})},$$
$$\|k\|_{L^\infty(D\times D\times\mathbb{R}\times\mathbb{R})} := \sup_{x,y\in D, s,t\in\mathbb{R}} |k(x,y,s,t)|.$$

Then for $x \in D(z_1)$ *the Volterra property of the kernel implies that*

$$\begin{aligned}
|u(x) - w(x)| &\le \int_D |k(x,y,u(x),u(y))u(y) - k(x,y,w(x),w(y))w(y)|dy \\
&\le \int_{D(z_1)} |k(x,y,u(x),u(y))u(y) - k(x,y,w(x),u(y))u(y)|dy \\
&\quad + \int_{D(z_1)} |k(x,y,w(x),u(y))u(y) - k(x,y,w(x),w(y))u(y)|dy \\
&\quad + \int_{D(z_1)} |k(x,y,w(x),w(y))u(y) - k(x,y,w(x),w(y))w(y)|dy \\
&\le 2\|k\|_{C(\overline{D}\times\overline{D};C^{0,1}(\mathbb{R}\times\mathbb{R}))}\|u-w\|_{L^2(D(z_1))}\|u\|_{L^2(D(z_1))} \\
&\qquad + \|k\|_{L^\infty(D\times D\times\mathbb{R}\times\mathbb{R})}\|u-w\|_{L^2(D(z_1))}\sqrt{|D(z_1)|},
\end{aligned}$$

so that for all $0 < z_1 < \ell$,

$$\begin{aligned}
&\|u-w\|^2_{L^2(D(z_1))} \\
&= \int_0^{z_1}\left(\int_0^\ell \cdots \int_0^\ell \mathbf{1}_{D(x)}|u(x)-w(x)|^2 dx_d dx_{d-1}\dots dx_2\right)dx_1 \\
&\le z_1\ell^{d-1}\Bigg(2\|k\|_{C(\overline{D}\times\overline{D};C^{0,1}(\mathbb{R}\times\mathbb{R}))}\|u-w\|_{L^2(D(z_1))}\|u\|_{L^2(D(z_1))} \\
&\qquad\qquad + \|k\|_{L^\infty(D\times D\times\mathbb{R}\times\mathbb{R})}\|u-w\|_{L^2(D(z_1))}\sqrt{|D(z_1)|}\Bigg)^2 \\
&\le z_1\ell^{d-1}\left(\|k\|_{C(\overline{D}\times\overline{D};C^{0,1}(\mathbb{R}\times\mathbb{R}))}\|u\|_{L^2(D)} + \|k\|_{L\infty(D\times D\times\mathbb{R}\times\mathbb{R})}\sqrt{|D|}\right)^2\|u-w\|^2_{L^2(D(z_1))}.
\end{aligned}$$

Thus, when z_1 *is so small that*

$$z_1\ell^{d-1}\left(\|k\|_{C(\overline{D}\times\overline{D};C^{0,1}(\mathbb{R}^n))}\|u\|_{L^2(D)} + \|k\|_{L\infty(D\times D\times\mathbb{R}\times\mathbb{R})}\sqrt{|D|}\right)^2 < 1,$$

we find that $\|u-w\|_{L^2(D(z_1))} = 0$, *that is,* $u(x) - w(x) = 0$ *for* $x \in D(z_1)$. *Using the same arguments as above, we see for all* $k \in \mathbb{N}$ *that that if* $u = w$ *in* $D(kz_1)$ *then* $u = w$ *in* $D((k+1)z_1)$. *Using induction, we see that* $u = w$ *in* D. *Hence, the operator* $u \mapsto F(u)$ *is injective in* $L^2(D)$.

Example 5. *We consider derivatives of Volterra operators in the domain* $D \subset [0,\ell]^d$. *Let* $K : L^2(D) \to L^2(D)$ *be a non-linear operator*

$$K(u) = \int_D k(x,y,u(y))u(y)dy, \tag{D.6}$$

where $k(x,y,t)$ *satisfies* (D.4), *is bounded, and* $k \in C(\overline{D}\times\overline{D}; C^{0,1}(\mathbb{R}\times\mathbb{R}))$. *Let* $F_1 : L^2(D) \to L^2(D)$ *be*

$$F_1(u) = u + K(u). \tag{D.7}$$

Then the Fréchet derivative of K at $u_0 \in L^2(D)$ to the direction $w \in L^2(D)$ is

$$DF_1|_{u_0}(w) = w(x) + \int_D k_1(x, y, u_0(y))w(y)dy, \tag{D.8}$$

where

$$k_1(x, y, u_0(y)) = u_0(y)\frac{\partial}{\partial t}k(x, y, t)\Big|_{t=u_0(x)} + k(x, y, u_0(y)), \tag{D.9}$$

is a Volterra opertor satisfying

$$k_1(x, y, t) \neq 0 \implies y_j \leq x_j \quad \text{for all } j = 1, 2, \ldots, d. \tag{D.10}$$

As seen in Example 4, the operator $DF_1|_{u_0} : L^2(D) \to L^2(D)$ is injective.

E Details of Section 4.2

In this appendix, we prove Theorem 3. We recall that in that theorem, we consider the case when $n = 1$, $D \subset \mathbb{R}$ is a bounded interval, and the operator F_1 is of the form

$$F_1(u)(x) = W(x)u(x) + \int_D k(x, y, u(y))u(y)dy,$$

where $W \in C^1(\overline{D})$ satisfies $0 < c_1 \leq W(x) \leq c_2$, the function $(x, y, s) \mapsto k(x, y, s)$ is in $C^3(\overline{D} \times \overline{D} \times \mathbb{R})$, and that in $\overline{D} \times \overline{D} \times \mathbb{R}$ its three derivatives and the derivatives of W are all uniformly bounded by c_0, that is,

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

We recall that the identical embedding $H^1(D) \to L^\infty(D)$ is bounded and compact by Sobolev's embedding theorem.

As we will consider kernels $k(x, y, u_0(y))$, we will consider the non-linear operator F_1 mainly as an operator in a Sobolev space $H^1(D)$.

The Frechet derivative of F_1 at u_0 to direction w, denoted by $A_{u_0}w = DF_1|_{u_0}(w)$ is given by

$$A_{u_0}w = W(x)w(x) + \int_D k(x, y, u_0(y))w(y)dy + \int_D u_0(y)\, \frac{\partial k}{\partial u}(x, y, u_0(y))w(y)dy. \tag{E.2}$$

The condition (E.1) implies that

$$F_1 : H^1(D) \to H^1(D), \tag{E.3}$$

is a locally Lipsichitz smooth function and that the operator

$$A_{u_0} : H^1(D) \to H^1(D),$$

given in (E.2), is defined for all $u_0 \in C(\overline{D})$ as a bounded linear operator.

When $\mathcal{X}$ is a Banach space, we let $B_{\mathcal{X}}(0, R) = \{v \in \mathcal{X} : \|v\|_{\mathcal{X}} < R\}$ and $\overline{B}_{\mathcal{X}}(0, R) = \{v \in \mathcal{X} : \|v\|_{\mathcal{X}} \leq R\}$ be the open and closed balls in $\mathcal{X}$, respectively.

We consider the Hölder spaces $C^{n,\alpha}(\overline{D})$ and their image in (leaky) ReLU-type functions. Let $a \geq 0$ and $\sigma_a(s) = \mathrm{ReLU}(s) - a\,\mathrm{ReLU}(-s)$. We will consider the image of the closed ball of $C^{1,\alpha}(\overline{D})$ in the map σ_a, that is $\sigma_a(\overline{B}_{C^{1,\alpha}(\overline{D})}(0, R)) := \{\sigma_a \circ g \in C(\overline{D}) : \|g\|_{C^{1,\alpha}(\overline{D})} \leq R\}$.

We will below assume that for all $u_0 \in C(\overline{D})$ the integral operator satisfies

$$A_{u_0} : H^1(D) \to H^1(D) \text{ is an injective operator.} \tag{E.4}$$

This condition is valid when $K(u)$ is a Volterra operator, see Examples 4 and 5. As the integral operators A_{u_0} are Fredholm operators having index zero. This implies that the operators (E.4) are bijective.

The inverse operator $A_{u_0}^{-1} : H^1(D) \to H^1(D)$ can be written as

$$A_{u_0}^{-1}v(x) = \widetilde{W}(x)v(x) - \int_D \widetilde{k}_{u_0}(x,y)v(y)dy, \tag{E.5}$$

where $\widetilde{k}_{u_0}, \partial_x \widetilde{k}_{u_0} \in C(\overline{D} \times \overline{D})$ and $\widetilde{W} \in C^1(\overline{D})$.

We will consider the inverse function of the map F_1 in a set $\mathcal{Y} \subset \sigma_a(\overline{B}_{C^{1,\alpha}(\overline{D})}(0,R))$ that is a compact subset of the Sobolev space $H^1(D)$. To this end, we will cover the set $\mathcal{Y}$ with small balls $B_{H^1(D)}(g_j, \varepsilon_0)$, $j = 1, 2, \dots, J$ of $H^1(D)$, centered at $g_j = F_1(v_j)$, where $v_j \in H^1(D)$. We will show that when $g \in B_{H^1(D)}(g_j, 2\varepsilon_0)$, that is, g is $2\varepsilon_1$-close to the function g_j in $H^1(D)$, the inverse map of F_1 can be written as a limit $(F_1^{-1}(g), g) = \lim_{m\to\infty} \mathcal{H}_j^{\circ m}(v_j, g)$ in $H^1(D)^2$, where

$$\mathcal{H}_j \begin{pmatrix} u \\ g \end{pmatrix} = \begin{pmatrix} u - A_{v_j}^{-1}(F_1(u) - F_1(v_j)) + A_{v_j}^{-1}(g - g_j) \\ g \end{pmatrix}.$$

That is, near g_j we can approximate F_1^{-1} as a composition $\mathcal{H}_j^{\circ m}$ of $2m$ layers of neural operators.

To glue the local inverse maps together, we use a partition of unity in the function space $\mathcal{Y}$ given by integral neural operators

$$\Phi_{\vec{i}}(v,w) = \pi_1 \circ \phi_{\vec{i},1} \circ \phi_{\vec{i},2} \circ \cdots \circ \phi_{\vec{i},\ell_0}(v,w), \quad \text{where} \quad \phi_{\vec{i},\ell}(v,w) = (F_{y_\ell, s(\vec{i},\ell), \epsilon_1}(v,w), w),$$

$\pi_1(v,w) = v$ maps a pair (v,w) to the first function v, and $\vec{i}$ belongs to a finite index set $\mathcal{I} \subset \mathbb{Z}^{\ell_0}$, $\epsilon_1 > 0$ and $y_\ell \in D$ $(\ell = 1, ..., \ell_0)$, where $s(\vec{i},\ell) := i_\ell \epsilon_1$. Here, $F_{z,s,h}(v,w)$ are integral neural operators with distributional kernels

$$F_{z,s,h}(v,w)(x) = \int_D k_{z,s,h}(x,y,v(x),w(y))dy,$$

where $k_{z,s,h}(x,y,v(x),w(y)) = v(x)\mathbf{1}_{[s-\frac{1}{2}h, s+\frac{1}{2}h)}(w(y))\delta(y-z)$, $\mathbf{1}_A$ is the indicator function of a set A and $y \mapsto \delta(y-z)$ is the Dirac delta distribution at the point $z \in D$. Using these, we can write the inverse of F_1 at $g \in \mathcal{Y}$ as

$$F_1^{-1}(g) = \lim_{m\to\infty} \sum_{\vec{i}\in\mathcal{I}} \Phi_{\vec{i}} \mathcal{H}_{j(\vec{i})}^{\circ m} \begin{pmatrix} v_{j(\vec{i})} \\ g \end{pmatrix}, \tag{E.6}$$

where $j(\vec{i}) \in \{1, 2, \dots, J\}$ are suitably chosen and the limit is taken in the norm topology of $H^1(D)$. This result is summarized by the following theorem, a modified version of Theorem 3 where the inverse operator F_1^{-1} in (E.6) have refined the partition of unity $\Phi_{\vec{i}}$ so that we use indexes $\vec{i} \in \mathcal{I} \subset \mathbb{Z}^{\ell_0}$ instead of $j \in \{1, \dots, J\}$.

Theorem 4. *Assume that F_1 satisfies the above assumptions* (E.1) *and* (E.4) *and that $F_1 : H^1(D) \to H^1(D)$ is a bijection. Let $\mathcal{Y} \subset \sigma_a(\overline{B}_{C^{1,\alpha}(\overline{D})}(0,R))$ be a compact subset the Sobolev space $H^1(D)$, where $\alpha > 0$ and $a \geq 0$. Then the inverse of $F_1 : H^1(D) \to H^1(D)$ in $\mathcal{Y}$ can written as a limit* (E.6) *that is, as a limit of integral neural operators.*

Observe that Theorem 4 includes the case where $a = 1$, in which case $\sigma_a = Id$ and $\mathcal{Y} \subset \sigma_a(\overline{B}_{C^{1,\alpha}(\overline{D})}(0,R)) = \overline{B}_{C^{1,\alpha}(\overline{D})}(0,R))$. We note that when σ_a is a leaky ReLU-function with parameter $a > 0$, Theorem 4 can be applied to compute the inverse of $\sigma_a \circ F_1$ given by $F_1^{-1} \circ \sigma_a^{-1}$, where $\sigma_a^{-1} = \sigma_{1/a}$. Note that the assumption that $\mathcal{Y} \subset \sigma_a(\overline{B}_{C^{1,\alpha}(\overline{D})}(0,R))$ makes it possible to apply Theorem 4 in the case when one trains deep neural networks having layers $\sigma_a \circ F_1$ and the parameter a of the leaky ReLU-function is a free parameter which is also trained.

Proof. As the operator F_1 can be multiplied by function $W(x)^{-1}$, it is sufficient to consider the case when $W(x) = 1$.

Below, we use the fact that, because $D \subset \mathbb{R}$, Sobolev's embedding theorem yields that the embedding $H^1(D) \to C(\overline{D})$ is bounded and there is $C_S > 0$ such that

$$\|u\|_{C(\overline{D})} \leq C_S \|u\|_{H^1(D)}. \tag{E.7}$$

For clarity, we denote the norm of $C(\overline{D})$ by $\|u\|_{L^\infty(D)}$.

Next we consider the Frechet derivatives of F_1. We recall that the 1st Frechet derivative of F_1 at u_0 is the operator A_{u_0}. The 2nd Frechet derivative of F_1 at u_0 to directions w_1 and w_2 is

$$\begin{aligned} D^2F_1|_{u_0}(w_1,w_2) &= \int_D 2\frac{\partial k}{\partial u}(x,y,u_0(y))w_1(y)w_2(y)dy + \int_D u_0(y)\,\frac{\partial k^2}{\partial u^2}(x,y,u_0(y))w_1(y)w_2(y)dy \\ &= \int_D p(x,y)w_1(y)w_2(y)dy, \end{aligned}$$

where

$$p(x,y) = 2\frac{\partial k}{\partial u}(x,y,u_0(y)) + u_0(y)\frac{\partial k^2}{\partial u^2}(x,y,u_0(y)), \tag{E.8}$$

and

$$\frac{\partial}{\partial x}p(x,y) = 2\frac{\partial^2 k}{\partial u \partial x}(x,y,u_0(y)) + u_0(y)\frac{\partial k^3}{\partial u^2 \partial x}(x,y,u_0(y)). \tag{E.9}$$

Thus,

$$\|D^2F_1|_{u_0}(w_1,w_2)\|_{H^1(D)} \le 3|D|^{1/2}\|k\|_{C^3(D\times D\times\mathbb{R})}(1+\|u_0\|_{L^\infty(D)})\|w_1\|_{L^\infty(D)}\|w_2\|_{L^\infty(D)}. \tag{E.10}$$

When we freeze the function u in kernel k to be u_0, we denote

$$K_{u_0}v(x) = \int_D k(x,y,u_0(y))v(y)dy.$$

Lemma 4. *For $u_0, u_1 \in C(\overline{D})$ we have*

$$\|K_{u_1} - K_{u_0}\|_{L^2(D)\to H^1(D)} \le \|k\|_{C^2(D\times D\times\mathbb{R})}|D|\|u_1-u_0\|_{L^\infty(D)}.$$

and

$$\|A_{u_1} - A_{u_0}\|_{L^2(D)\to H^1(D)} \le 2\|k\|_{C^2(D\times D\times\mathbb{R})}|D|(1+\|u_0\|_{L^\infty(D)})\|u_1-u_0\|_{L^\infty(D)}. \tag{E.11}$$

Proof. Denote

$$\begin{aligned} M_{u_0}v(x) &= \int_D u_0(y)\,\frac{\partial k}{\partial u}(x,y,u_0(y))v(y)dy, \\ N_{u_1,u_2}v(x) &= \int_D u_1(y)\,\frac{\partial k}{\partial u}(x,y,u_2(y))v(y)dy. \end{aligned}$$

We have

$$M_{u_2}v - M_{u_1}v = (N_{u_2,u_2}v - N_{u_2,u_1}v) + (N_{u_2,u_1}v - N_{u_1,u_1}v).$$

By Schur's test for continuity of integral operators,

$$\begin{aligned} \|K_{u_0}\|_{L^2(D)\to L^2(D)} &\le \left(\sup_{x\in D}\int_D |k(x,y,u_0(y))|dy\right)^{1/2}\left(\sup_{y\in D}\int_D |k(x,y,u_0(y))|dx\right)^{1/2} \\ &\le \|k\|_{C^0(D\times D\times\mathbb{R})}, \end{aligned}$$

and

$$\begin{aligned} &\|M_{u_0}\|_{L^2(D)\to L^2(D)} \\ \le\ & \left(\sup_{x\in D}\int_D |u_0(y)\,\frac{\partial k}{\partial u}(x,y,u_0(y))|dy\right)^{1/2}\left(\sup_{y\in D}\int_D |u_0(y)\,\frac{\partial k}{\partial u}(x,y,u_0(y))|dx\right)^{1/2} \\ \le\ & \|k\|_{C^1(D\times D\times\mathbb{R})}\|u\|_{C(\overline{D})}, \end{aligned}$$

and

$$\begin{aligned}
&\|K_{u_2} - K_{u_1}\|_{L^2(D)\to L^2(D)} \\
\leq\ & \left(\sup_{x\in D}\int_D |k(x,y,u_2(y)) - k(x,y,u_1(y))|dy\right)^{1/2} \\
&\qquad \times\left(\sup_{y\in D}\int_D |k(x,y,u_2(y)) - k(x,y,u_1(y))|dx\right)^{1/2} \\
\leq\ & \left(\sup_{x\in D}\int_D \|k\|_{C^1(D\times D\times\mathbb{R})}|u_2(y) - u_1(y))|dy\right)^{1/2} \\
&\qquad \times\left(\sup_{y\in D}\int_D \|k\|_{C^1(D\times D\times\mathbb{R})}|u_2(y) - u_1(y))|dx\right)^{1/2} \\
\leq\ & \|k\|_{C^1(D\times D\times\mathbb{R})}\left(\sup_{x\in D}\int_D |u_2(y) - u_1(y))|dy\right)^{1/2}\left(\sup_{y\in D}\int_D |u_2(y) - u_1(y))|dx\right)^{1/2} \\
\leq\ & \|k\|_{C^1(D\times D\times\mathbb{R})}\left(|D|^{1/2}\|u_2 - u_1\|_{L^2(D)}\right)^{1/2}\left(|D|\sup_{y\in D}|u_2(y) - u_1(y))|\right)^{1/2} \\
\leq\ & \|k\|_{C^1(D\times D\times\mathbb{R})}|D|^{3/4}\|u_2 - u_1\|_{L^2(D)}^{1/2}\|u_2 - u_1\|_{L^\infty(D)}^{1/2} \\
\leq\ & \|k\|_{C^1(D\times D\times\mathbb{R})}|D|\|u_2 - u_1\|_{L^\infty(D)},
\end{aligned}$$

and

$$\begin{aligned}
&\|N_{u_2,u_2} - N_{u_2,u_1}\|_{L^2(D)\to L^2(D)} \\
\leq\ & \left(\sup_{x\in D}\int_D |u_2(y)k(x,y,u_2(y)) - u_2(y)k(x,y,u_1(y))|dy\right)^{1/2} \\
&\qquad \times\left(\sup_{y\in D}\int_D |u_2(y)k(x,y,u_2(y)) - u_2(y)k(x,y,u_1(y))|dx\right)^{1/2} \\
\leq\ & \|k\|_{C^1(D\times D\times\mathbb{R})}|D|^{3/4}\|u_2\|_{C^0(D)}\|u_2 - u_1\|_{L^2(D)}^{1/2}\|u_2 - u_1\|_{L^\infty(D)}^{1/2} \\
\leq\ & \|k\|_{C^1(D\times D\times\mathbb{R})}|D|\cdot\|u_2\|_{C^0(D)}\|u_2 - u_1\|_{L^\infty(D)},
\end{aligned}$$

and

$$\begin{aligned}
&\|N_{u_2,u_1} - N_{u_1,u_1}\|_{L^2(D)\to L^2(D)} \\
\leq\ & \left(\sup_{x\in D}\int_D |(u_2(y) - u_1(y))k(x,y,u_1(y))|dy\right)^{1/2} \\
&\qquad \times\left(\sup_{y\in D}\int_D |(u_2(y) - u_1(y))k(x,y,u_1(y))|dx\right)^{1/2} \\
\leq\ & \|k\|_{C^0(D\times D\times\mathbb{R})}|D|\cdot\|u_2 - u_1\|_{L^\infty(D)},
\end{aligned}$$

so that

$$\begin{aligned}
&\|M_{u_2} - M_{u_1}\|_{L^2(D)\to L^2(D)} \\
\leq\ & \|k\|_{C^1(D\times D\times\mathbb{R})}|D|(1 + \|u_2\|_{C^0(D)})\|u_2 - u_1\|_{L^\infty(D)}.
\end{aligned}$$

Also, when $D_x v = \frac{dv}{dx}$,

$$\begin{aligned}
&\|D_x \circ K_{u_0}\|_{L^2(D)\to L^2(D)} \\
\leq\ & \left(\sup_{x\in D}\int_D |D_x k(x,y,u_0(y))|dy\right)^{1/2}\left(\sup_{y\in D}\int_D |D_x k(x,y,u_0(y))|dx\right)^{1/2} \\
\leq\ & \|k\|_{C^1(D\times D\times\mathbb{R})},
\end{aligned}$$

and

$$
\begin{aligned}
&\|D_x \circ K_{u_1} - D_x \circ K_{u_0}\|_{L^2(D)\to L^2(D)} \\
\leq\ & \left(\sup_{x\in D}\int_D |D_x k(x,y,u_1(y)) - D_x k(x,y,u_0(y))|dy\right)^{1/2} \\
&\qquad \times\left(\sup_{y\in D}\int_D |D_x k(x,y,u_1(y)) - D_x k(x,y,u_0(y))|dx\right)^{1/2} \\
\leq\ & \left(\sup_{x\in D}\int_D \|k\|_{C^2(D\times D\times\mathbb{R})}|u_1(y)-u_0(y))|dy\right)^{1/2} \\
&\qquad \times\left(\sup_{y\in D}\int_D \|k\|_{C^2(D\times D\times\mathbb{R})}|u_1(y)-u_0(y))|dx\right)^{1/2} \\
\leq\ & \|k\|_{C^2(D\times D\times\mathbb{R})}\left(\sup_{x\in D}\int_D |u_1(y)-u_0(y))|dy\right)^{1/2}\left(\sup_{y\in D}\int_D |u_1(y)-u_0(y))|dx\right)^{1/2} \\
\leq\ & \|k\|_{C^2(D\times D\times\mathbb{R})}\left(|D|^{1/2}\|u_1-u_0\|_{L^2(D)}\right)^{1/2}\left(|D|\sup_{y\in D}|u_1(y)-u_0(y))|\right)^{1/2} \\
\leq\ & \|k\|_{C^2(D\times D\times\mathbb{R})}|D|^{3/4}\|u_1-u_0\|_{L^2(D)}^{1/2}\|u_1-u_0\|_{L^\infty(D)}^{1/2} \\
\leq\ & \|k\|_{C^2(D\times D\times\mathbb{R})}|D|\|u_1-u_0\|_{L^\infty(D)}.
\end{aligned}
$$

Thus,

$$\|K_{u_0}\|_{L^2(D)\to H^1(D)} \quad\leq\quad \|k\|_{C^1(D\times D\times\mathbb{R})},$$

and

$$\|M_{u_0}\|_{L^2(D)\to H^1(D)} \quad\leq\quad \|u_0\|_{C^0(D)}\|k\|_{C^1(D\times D\times\mathbb{R})},$$

and

$$\|K_{u_1} - K_{u_0}\|_{L^2(D)\to H^1(D)} \quad\leq\quad \|k\|_{C^2(D\times D\times\mathbb{R})}|D|\|u_1-u_0\|_{L^\infty(D)}.$$

Similarly,

$$\|M_{u_1} - M_{u_0}\|_{L^2(D)\to H^1(D)} \quad\leq\quad \|k\|_{C^2(D\times D\times\mathbb{R})}|D|(1+\|u_2\|_{C^0(D)})\|u_1-u_0\|_{L^\infty(D)}.$$

As $A_{u_1} = K_{u_1} + M_{u_1}$, the claim follows. □

As the embedding $H^1(D)\to C(\overline{D})$ is bounded and has norm C_S, Lemma 4 implies that for all $R>0$ there is

$$C_L(R) = 2\|k\|_{C^2(D\times D\times\mathbb{R})}|D|(1+C_S R),$$

such that the map,

$$u_0 \mapsto DF_1|_{u_0}, \quad u_0 \in \overline{B}_{H^1}(0,R), \tag{E.12}$$

is a Lipschitz map $\overline{B}_{H^1}(0,R)\to\mathcal{L}(H^1(D),H^1(D))$ with Lipschitz constant $C_L(R)$, that is,

$$\|DF_1|_{u_1} - DF_1|_{u_2}\|_{H^1(D)\to H^1(D)} \leq C_L(R)\|u_1-u_2\|_{H^1(D)}. \tag{E.13}$$

As $u_0\mapsto A_{u_0} = DF_1|_{u_0}$ is continuous, the inverse $A_{u_0}^{-1}: H^1(D)\to H^1(D)$ exists for all $u_0\in C(\overline{D})$, and the embedding $H^1(D)\to C(\overline{D})$ is compact, we have that for all $R>0$ there is $C_B(R)>0$ such that

$$\|A_{u_0}^{-1}\|_{H^1(D)\to H^1(D)} \leq C_B(R), \quad \text{for all } u_0\in\overline{B}_{H^1}(0,R). \tag{E.14}$$

Let $R_1, R_2 > 0$ be such that $\mathcal{Y}\subset\overline{B}_{H^1}(0,R_1)$ and $X = F_1^{-1}(\mathcal{Y})\subset\overline{B}_{H^1}(0,R_2)$. Below, we denote $C_L = C_L(2R_2)$ and $C_B = C_B(R_2)$.

Next we consider inverse of F_1 in $\mathcal{Y}$. To this end, let us consider $\varepsilon_0 > 0$, which we choose later to be small enough. As $\mathcal{Y} \subset \overline{B}_{H^1}(0,R)$ is compact there are a finite number of elements $g_j = F_1(v_j) \in \mathcal{Y}$, where $v_j \in X$, $j = 1, 2, \ldots, J$ such that

$$\mathcal{Y} \subset \bigcup_{j=1}^{J} B_{H^1(D)}(g_j, \varepsilon_0).$$

We observe that for $u_0, u_1 \in X$,

$$A_{u_1}^{-1} - A_{u_0}^{-1} = A_{u_1}^{-1}(A_{u_1} - A_{u_0})A_{u_0}^{-1},$$

and hence the Lipschitz constant of $A_{\cdot}^{-1} : u \mapsto A_u^{-1}$, $X \to \mathcal{L}(H^1(D), H^1(D))$ satisfies

$$Lip(A_{\cdot}^{-1}) \leq C_A = C_B^2 C_L, \tag{E.15}$$

see (E.11).

Let us consider a fixed j and $g_j \in \mathcal{Y}$. When g satisfies

$$\|g - g_j\|_{H^1(D)} < 2\varepsilon_0, \tag{E.16}$$

the equation

$$F_1(u) = g, \quad u \in X,$$

is equivalent to the fixed point equation

$$u = u - A_{v_j}^{-1}(F_1(u) - F_1(v_j)) + A_{v_j}^{-1}(g - g_j),$$

that is equivalent to the fixed point equation

$$H_j(u) = u,$$

for the function $H_j : H^1(D) \to H^1(D)$,

$$H_j(u) = u - A_{v_j}^{-1}(F_1(u) - F_1(v_j)) + A_{v_j}^{-1}(g - g_j).$$

Note that H_j depends on g, and thus we later denote $H_j = H_j^g$. We observe that

$$H_j(v_j) = v_j + A_{v_j}^{-1}(g - g_j). \tag{E.17}$$

Let $u, v \in \overline{B}_{H^1}(0, 2R_2)$. We have

$$F_1(u) = F_1(v) + A_v(u - v) + B_v(u - v), \quad \|B_v(u - v)\| \leq C_0\|u - v\|^2,$$

where, see (E.10),

$$C_0 = 3|D|^{1/2}\|k\|_{C^3(D\times D\times\mathbb{R})}(1 + 2C_S R_2)C_S^2,$$

so that for $u_1, u_2 \in \overline{B}_{H^1}(0, 2R_2)$,

$$\begin{aligned} & u_1 - u_2 - A_{v_j}^{-1}(F_1(u_1) - F_1(u_2)) \\ = \quad & u_1 - u_2 - A_{u_2}^{-1}(F_1(u_1) - F_1(u_2)) - (A_{u_2}^{-1} - A_{v_j}^{-1})(F_1(u_1) - F_1(u_2)), \end{aligned}$$

and

$$\begin{aligned} & \|u_1 - u_2 - A_{u_2}^{-1}(F_1(u_1) - F_1(u_2))\|_{H^1(D)} \\ = \quad & \|A_{u_2}^{-1}(B_{u_2}(u_1 - u_2))\|_{H^1(D)} \\ \leq \quad & \|A_{u_2}^{-1}\|_{H^1(D)\to H^1(D)}\|B_{u_2}(u_1 - u_2)\|_{H^1(D)} \\ \leq \quad & \|A_{u_2}^{-1}\|_{H^1(D)\to H^1(D)}C_0\|u_1 - u_2\|_{H^1(D)}^2, \\ \leq \quad & C_B C_0\|u_1 - u_2\|_{H^1(D)}^2, \end{aligned}$$

and

$$\begin{aligned}
& \|(A_{u_2}^{-1} - A_{v_j}^{-1})(F_1(u_1) - F_1(u_2))\|_{H^1(D)} \\
\le\ & \|A_{u_2}^{-1} - A_{v_j}^{-1}\|_{H^1(D)\to H^1(D)} \|F_1(u_1) - F_1(u_2)\|_{H^1(D)} \\
\le\ & Lip_{\overline{B}_{H^1}(0,2R_2)\to H^1(D)}(A_{\cdot}^{-1})\|u_2 - v_j\| Lip_{\overline{B}_{H^1}(0,2R_2)\to H^1(D)}(F_1)\|u_2 - u_1\|_{H^1(D)} \\
\le\ & C_A\|u_2 - v_j\|(C_B + 4C_0R_2)\|u_2 - u_1\|_{H^1(D)},
\end{aligned}$$

see (E.2), and hence, when $\|u - v_j\| \le r \le R_2$,

$$\begin{aligned}
& \|H_j(u_1) - H_j(u_2)\|_{H^1(D)} \\
\le\ & \|u_1 - u_2 - A_{v_j}^{-1}(F_1(u_1) - F_1(u_2))\|_{H^1(D)} \\
\le\ & \|u_1 - u_2 - A_{u_2}^{-1}(F_1(u_1) - F_1(u_2))\|_{H^1(D)} + \|(A_{u_2}^{-1} - A_{v_j}^{-1})(F_1(u_1) - F_1(u_2))\|_{H^1(D)} \\
\le\ & \bigg(C_BC_0(\|u_1 - v_j\|_{H^1(D)} + \|u_2 - v_j\|_{H^1(D)}) + C_A(C_B + 4C_0R_2)\|u_2 - v_j\|\bigg) \cdot \|u_2 - u_1\|_{H^1(D)} \\
\le\ & C_H r\|u_2 - u_1\|_{H^1(D)},
\end{aligned}$$

where

$$C_H = 2C_BC_0 + C_A(C_B + 4C_0R_2).$$

We now choose

$$r = \min(\frac{1}{2C_H}, R_2).$$

We consider

$$\varepsilon_0 \le \frac{1}{8C_B}\frac{1}{2C_H}.$$

Then, we have

$$r \ge 2C_B\varepsilon_0/(1 - C_H r).$$

Then, we have that $\mathrm{Lip}_{\overline{B}_{H^1}(0,2R_2)\to H^1(D)}(H_j) \le a = C_H r < \frac{1}{2}$, and

$$r \ge \|A_{v_j}^{-1}\|_{H^1(D)\to H^1(D)}\|g - g_j\|_{H^1(D)}/(1 - a),$$

and for all $u \in \overline{B}_{H^1}(0, R_2)$ such that $\|u - v_j\| \le r$, we have $\|A_{v_j}^{-1}(g - g_j)\|_{H^1(D)} \le (1 - a)r$.
Then,

$$\begin{aligned}
\|H_j(u) - v_j\|_{H^1(D)} &\le \|H_j(u) - H_j(v_j)\|_{H^1(D)} + \|H_j(v_j) - v_j\|_{H^1(D)} \\
&\le a\|u - v_j\|_{H^1(D)} + \|v_j + A_{v_j}^{-1}(g - g_j) - v_j\|_{H^1(D)} \\
&\le ar + \|A_{v_j}^{-1}(g - g_j)\|_{H^1(D)} \le r,
\end{aligned}$$

that is, H_j maps $\overline{B}_{H^1(D)}(v_j, r)$ to itself. By Banach fixed point theorem, $H_j : \overline{B}_{H^1(D)}(v_j, r) \to \overline{B}_{H^1(D)}(v_j, r)$ has a fixed point.

Let us denote

$$\mathcal{H}_j \begin{pmatrix} u \\ g \end{pmatrix} = \begin{pmatrix} H_j^g(u) \\ g \end{pmatrix} = \begin{pmatrix} u - A_{v_j}^{-1}(F_1(u) - F_1(v_j)) + A_{v_j}^{-1}(g - g_j) \\ g \end{pmatrix}.$$

By the above, when we choose ε_0 to have a value

$$\varepsilon_0 < \frac{1}{8C_B}\frac{1}{2C_H},$$

the map F_1 has a right inverse map $\mathcal{R}_j$ in $B_{H^1}(g_j, 2\varepsilon_0)$, that is,

$$F_1(\mathcal{R}_j(g)) = g, \quad \text{for } g \in B_{H^1}(g_j, 2\varepsilon_0), \tag{E.18}$$

and by Banach fixed point theorem it is given by the limit

$$\mathcal{R}_j(g) = \lim_{m\to\infty} w_{j,m}, \quad g \in B_{H^1}(g_j, 2\varepsilon_0), \tag{E.19}$$

in $H^1(D)$, where

$$w_{j,0} = v_j, \tag{E.20}$$
$$w_{j,m+1} = H_j^g(w_{j,m}). \tag{E.21}$$

We can write for $g \in B_{H^1}(g_j, 2\varepsilon_0)$,

$$\begin{pmatrix} \mathcal{R}_j(g) \\ g \end{pmatrix} = \lim_{m\to\infty} \mathcal{H}_j^{\circ m} \begin{pmatrix} v_j \\ g \end{pmatrix},$$

where the limit takes space in $H^1(D)^2$ and

$$\mathcal{H}_j^{\circ m} = \mathcal{H}_j \circ \mathcal{H}_j \circ \cdots \circ \mathcal{H}_j, \tag{E.22}$$

is the composition of m operators $\mathcal{H}_j$. This implies that $\mathcal{R}_j$ can be written as a limit of finite iterations of neural operators H_j (we will consider how the operator $A_{v_j}^{-1}$ can be written as a neural operator below).

As $\mathcal{Y} \subset \sigma_a(\overline{B}_{C^{1,\alpha}(\overline{D})}(0,R))$, there are finite number of points $y_\ell \in D$, $\ell = 1, 2, \ldots, \ell_0$ and $\varepsilon_1 > 0$ such that the sets

$$Z(i_1, i_2, \ldots, i_{\ell_0}) = \{g \in Y :\ (i_\ell - \frac{1}{2})\varepsilon_1 \le g(y_\ell) < (i_\ell + \frac{1}{2})\varepsilon_1, \text{ for all } \ell\},$$

where $i_1, i_2, \ldots, i_{\ell_0} \in \mathbb{Z}$, satisfy the condition

$$\text{If } (Z(i_1, i_2, \ldots, i_{\ell_0}) \cap \mathcal{Y}) \cap B_{H^1(D)}(g_j, \varepsilon_0) \neq \emptyset \text{ then } Z(i_1, i_2, \ldots, i_{\ell_0}) \cap \mathcal{Y} \subset B_{H^1(D)}(g_j, 2\varepsilon_0). \tag{E.23}$$

To show (E.23), we will below use the mean value theorem for function $g = \sigma_a \circ v \in \mathcal{Y}$, where $v \in C^{1,\alpha}(\overline{D})$. First, let us consider the case when the parameter a of the leaky ReLU function σ_a is strictly positive. Without loss of generality, we can assume that $D = [0,1]$ and $y_\ell = h\ell$, where $h = 1/\ell_0$ and $\ell = 0, 1, \ldots, \ell_0$. We consider $g \in \mathcal{Y} \cap Z(i_1, i_2, \ldots, i_{\ell_0}) \subset \sigma_a(\overline{B}_{C^{1,\alpha}(\overline{D})}(0,R))$ of the form $g = \sigma_a \circ v$. As a is non-zero, the inequality $(i_\ell - \frac{1}{2})\varepsilon_1 \le g(y_\ell) < (i_\ell + \frac{1}{2})\varepsilon_1$ is equivalent to $\sigma_{1/a}((i_\ell - \frac{1}{2})\varepsilon_1) \le v(y_\ell) < \sigma_{1/a}((i_\ell + \frac{1}{2})\varepsilon_1)$, and thus

$$\sigma_{1/a}(i_\ell \varepsilon_1) - A\varepsilon_1 \le v(y_\ell) < \sigma_{1/a}(i_\ell \varepsilon_1) + A\varepsilon_1, \tag{E.24}$$

where $A = \max(1, a, 1/a)$, that is, for $g = \sigma_a(v) \in Z(i_1, i_2, \ldots, i_{\ell_0})$ the values $v(y_\ell)$ are known within small errors. By applying mean value theorem on the interval $[(\ell_1 - 1)h, \ell_1 h]$ for function v we see that there is $x' \in [(\ell_1 - 1)h, \ell_1 h]$ such that

$$\frac{dv}{dx}(x') = \frac{v(\ell_1 h) - v((\ell_1 - 1)h)}{h},$$

and thus by (E.24),

$$|\frac{dv}{dx}(x') - d_{\ell,\vec{i}}| \le 2A\frac{\varepsilon_1}{h}, \tag{E.25}$$

where

$$d_{\ell,\vec{i}} = \frac{1}{h}(\sigma_{1/a}(i_\ell \varepsilon_1) - \sigma_{1/a}((i_\ell - 1)\varepsilon_1)), \tag{E.26}$$

Observe that these estimates are useful when ε_1 is much smaller that h. As $g = \sigma_a \circ v \in \mathcal{Y} \subset \sigma_a(\overline{B}_{C^{1,\alpha}(\overline{D})}(0,R))$, we have $v \in \overline{B}_{C^{1,\alpha}(\overline{D})}(0,R)$, so that $\frac{dv}{dx} \in \overline{B}_{C^{0,\alpha}(\overline{D})}(0,R)$ satisfies (E.25) implies that

$$|\frac{dv}{dx}(x) - d_{\ell,\vec{i}}| \le 2A\frac{\varepsilon_1}{h} + Rh^\alpha, \quad \text{for all } x \in [(\ell_1 - 1)h, \ell_1 h]. \tag{E.27}$$

Moreover, (E.24) and $v \in \overline{B}_{C^{1,\alpha}(\overline{D})}(0,R)$ imply

$$|v(x) - \sigma_{1/a}(i_\ell \varepsilon_1)| < A\varepsilon_1 + Rh, \tag{E.28}$$

for all $x \in [(\ell_1 - 1)h, \ell_1 h]$.

Let $\varepsilon_2 = \varepsilon_2/A$. When we first choose ℓ_0 to be large enough (so that $h = 1/\ell_0$ is small) and then ε_1 to be small enough, we may assume that

$$\max(2A\frac{\varepsilon_1}{h} + Rh^\alpha, A\varepsilon_1 + Rh) < \frac{1}{8}\varepsilon_2. \tag{E.29}$$

Then for any two functions $g, g' \in \mathcal{Y} \cap Z(i_1, i_2, \dots, i_{\ell_0}) \subset \sigma_a(\overline{B}_{C^{1,\alpha}(\overline{D})}(0,R))$ of the form $g = \sigma_a \circ v, g' = \sigma_a \circ v'$ the inequalities (E.27) and (E.28) imply

$$\begin{aligned} |\frac{dv}{dx}(x) - \frac{dv'}{dx}(x)| &< \frac{1}{4}\varepsilon_2, \\ |v(x) - v'(x)| &< \frac{1}{4}\varepsilon_2, \end{aligned} \tag{E.30}$$

for all $x \in D$. As $v, v' \in \overline{B}_{C^{1,\alpha}(\overline{D})}(0,R)$, this implies

$$\|v - v'\|_{C^1(\overline{D})} < \frac{1}{2}\varepsilon_2,$$

As the embedding $C^1(\overline{D}) \to H^1(D)$ is continuous and has norm less than 2 on the interval $D = [0,1]$, we see that

$$\|v - v'\|_{H^1(\overline{D})} < \varepsilon_2,$$

and thus

$$\|g - g'\|_{H^1(\overline{D})} < A\varepsilon_2 = \varepsilon_0.$$

Hence, the property (E.23) follows.

We next consider the case when the parameter a of the leaky relu function σ_a is zero. Again, we assume that $D = [0,1]$ and $y_\ell = h\ell$, where $h = 1/\ell_0$ and $\ell = 0, 1, \dots, \ell_0$. We consider $g \in \mathcal{Y} \cap Z(i_1, i_2, \dots, i_{\ell_0}) \subset \sigma_a(\overline{B}_{C^{1,\alpha}(\overline{D})}(0,R))$ of the form $g = \sigma_a(v)$ and an interval $[\ell_1 h, (\ell_1+1)h] \subset D$, where $1 \le \ell_1 \le \ell_0 - 2$. We will consider four cases. First, if g does not obtain the value zero on the interval $[\ell_1 h, (\ell_1+1)h]$ the mean value theorem implies that there is $x' \in [\ell_1 h, (\ell_1+1)h]$ such that $\frac{dg}{dx}(x') = \frac{dv}{dx}(x')$ is equal to $d = (g(\ell_1 h) - g([(\ell_1 - 1)h))/h$. Second, if g does not obtain the value zero on either of the intervals $[(\ell_1 - 1)h, \ell_1 h]$ or $[(\ell_1+1)h, (\ell_1+2)h]$, we can use the mean value theorem to estimate the derivatives of g and v at some point of these intervals similarly to the first case. Third, if g does not vanish identically on the interval $[\ell_1 h, (\ell_1+1)h]$ but it obtains the value zero on the both intervals $[(\ell_1 - 1)h, \ell_1 h]$ and $[(\ell_1+1)h, (\ell_1+2)h]$, the function v has two zeros on the interval $[(\ell_1 - 1)h, (\ell_1+2)h]$ and the mean value theorem implies that there is $x' \in [(\ell_1 - 1)h, (\ell_1+2)h]$ such that $\frac{dv}{dx}(x') = 0$. Fourth, if none of the above cases are valid, g vanishes identically on the interval $[\ell_1 h, (\ell_1+1)h]$. In all these cases the fact that $\|v\|_{C^{1,\alpha}(\overline{D})} \le R$ implies that the derivative of g can be estimated on the whole interval $[\ell_1 h, (\ell_1+1)h]$ within a small error. Using these observations we see for any $\varepsilon_2, \varepsilon_3 > 0$ that if $y_\ell \in D = [d_1, d_2] \subset \mathbb{R}$, $\ell = 1, 2, \dots, \ell_0$ are a sufficiently dense grid in D and ε_1 to be small enough, then the derivatives of any two functions $g, g' \in \mathcal{Y} \cap Z(i_1, i_2, \dots, i_{\ell_0}) \subset \sigma_a(\overline{B}_{C^{1,\alpha}(\overline{D})}(0,R))$ of the form $g = \sigma_a(v), g' = \sigma_a(v')$ satisfly $\|g - g'\|_{H^1([d_1+\varepsilon_3, d_2-\varepsilon_3])} < \varepsilon_2$. As the embedding $C^1([d_1+\varepsilon_3, d_2-\varepsilon_3]) \to H^1([d_1+\varepsilon_3, d_2-\varepsilon_3])$ is continuous,

$$\begin{aligned} \|\sigma_a(v)\|_{H^1([d_1, d_1+\varepsilon_3])} &\le c_a \|v\|_{C^{1,\alpha}(\overline{D})} \sqrt{\varepsilon_3}, \\ \|\sigma_a(v)\|_{H^1([d_2-\varepsilon_3, d_2])} &\le c_a \|v\|_{C^{1,\alpha}(\overline{D})} \sqrt{\varepsilon_3}, \end{aligned}$$

and ε_2 and ε_3 can be chosen to be arbitrarily small, we see that the property (E.23) follows. Thus the property (E.23) is shown in all cases.

By our assumptions $\mathcal{Y} \subset \sigma_a(B_{C^{1,\alpha}(\overline{D})}(0,R))$ and hence $g \in \mathcal{Y}$ implies that $\|g\|_{C(\overline{D})} \le AR$. This implies that $\mathcal{Y} \cap Z(i_1, i_2, \dots, i_{\ell_0})$ is empty if there is ℓ such that $|i_\ell| > 2AR/\varepsilon_1 + 1$. Thus, there is a finite set $\mathcal{I} \subset \mathbb{Z}^{\ell_0}$ such that

$$\mathcal{Y} \subset \bigcup_{\vec{i} \in \mathcal{I}} Z(\vec{i}), \tag{E.31}$$

$$Z(\vec{i}) \cap \mathcal{Y} \neq \emptyset, \quad \text{for all } \vec{i} \in \mathcal{I}, \tag{E.32}$$

where we use notation $\vec{i} = (i_1, i_2, \dots, i_{\ell_0}) \in \mathbb{Z}^{\ell_0}$. On the other hand, we have chosen $g_j \in \mathcal{Y}$ such that $B_{H^1(D)}(g_j, \varepsilon_0)$, $j = 1, \dots, J$ cover $\mathcal{Y}$. This implies that for all $\vec{i} \in \mathcal{I}$ there is $j = j(\vec{i}) \in \{1, 2, \dots, j\}$ such that there exists $g \in Z(\vec{i}) \cap B_{H^1(D)}(g_j, \varepsilon_0)$. By (E.23), this implies that

$$Z(\vec{i}) \subset B_{H^1(D)}(g_{j(\vec{i})}, 2\varepsilon_0). \tag{E.33}$$

Thus, we see that $Z(\vec{i})$, $\vec{i} \in \mathcal{I}$ is a disjoint covering of $\mathcal{Y}$, and by (E.33), in each set $Z(\vec{i}) \cap \mathcal{Y}$, $\vec{i} \in \mathcal{I}$ the map $g \to \mathcal{R}_j(g)$ we have constructed a right inverse of the map F_1.

Below, we denote $s(\vec{i}, \ell) = i_\ell \varepsilon_1$. Next we construct a partition of unity in $\mathcal{Y}$ using maps

$$F_{z,s,h}(v,w)(x) = \int_D k_{z,s,h}(x, y, v(x), w(y))dy,$$

where

$$k_{z,s,h}(x, y, v(x), w(y)) = v(x)\mathbf{1}_{[s-\frac{1}{2}h, s+\frac{1}{2}h)}(w(y))\delta(y - z).$$

Then,

$$F_{z,s,h}(v,w)(x) = \begin{cases} v(x), & \text{if } -\frac{1}{2}h \le w(z) - s < \frac{1}{2}h, \\ 0, & \text{otherwise.} \end{cases}$$

Next, for all $\vec{i} \in \mathcal{I}$ we define the operator $\Phi_{\vec{i}} : H^1(D) \times \mathcal{Y} \to H^1(D)$,

$$\Phi_{\vec{i}}(v,w) = \pi_1 \circ \phi_{\vec{i},1} \circ \phi_{\vec{i},2} \circ \cdots \circ \phi_{\vec{i},\ell_0}(v,w),$$

where $\phi_{\vec{i},\ell} : H^1(D) \times \mathcal{Y} \to H^1(D) \times \mathcal{Y}$ are the maps

$$\phi_{\vec{i},\ell}(v,w) = (F_{y_\ell, s(\vec{i},\ell), \varepsilon_1}(v,w), w),$$

and $\pi_1(v,w) = v$ maps a pair (v,w) to the first function v. It satisfies

$$\Phi_{\vec{i}}(v,w) = \begin{cases} v, & \text{if } -\frac{1}{2}\varepsilon_1 \le w(y_\ell) - s(\vec{i},\ell) < \frac{1}{2}\varepsilon_1 \text{ for all } \ell, \\ 0, & \text{otherwise.} \end{cases}$$

Observe that here $s(\vec{i},\ell) = i_\ell \varepsilon_1$ is close to the value $g_{j(\vec{i})}(y_\ell)$. Now we can write for $g \in Y$

$$F_1^{-1}(g) = \sum_{\vec{i} \in \mathcal{I}} \Phi_{\vec{i}}(\mathcal{R}_{j(\vec{i})}(g), g),$$

with suitably chosen $j(\vec{i}) \in \{1, 2, \dots, J\}$.

Let us finally consider $A_{u_0}^{-1}$ where $u_0 \in C(\overline{D})$. Let us denote

$$\widetilde{K}_{u_0} w = \int_D u_0(y) \frac{\partial k}{\partial u}(x, y, u_0(y)) w(y) dy,$$

and $J_{u_0} = K_{u_0} + \widetilde{K}_{u_0}$ be the integral operator with kernel

$$j_{u_0}(x,y) = k(x, y, u_0(y)) + u_0(y) \frac{\partial k}{\partial u}(x, y, u_0(y)).$$

We have

$$(I + J_{u_0})^{-1} = I - J_{u_0} + J_{u_0}(I + J_{u_0})^{-1} J_{u_0},$$

so that when we write the linear bounded operator

$$A_{u_0}^{-1} = (I + J_{u_0})^{-1} : H^1(D) \to H^1(D),$$

as an integral operator

$$(I + J_{u_0})^{-1} v(x) = v + \int_D m_{u_0}(x,y) v(y) dy,$$

we have

$$\begin{aligned}
& (I + J_{u_0})^{-1} v(x) \\
= \;& v(x) - J_{u_0} v(x) \\
& + \int_D \left(\int_D \left\{ j_{u_0}(x,y') j_{u_0}(y,y') + \left(\int_D j_{u_0}(x,y') m_{u_0}(y',x') j_{u_0}(x',y) dx' \right) \right\} dy' \right) v(y) dy \\
= \;& v(x) - \int_D \widetilde{j}_{u_0}(x,y) v(y) dy,
\end{aligned}$$

where

$$\widetilde{j}_{u_0}(x,y) = -j_{u_0}(x,y) + \int_D (j_{u_0}(x,y') j_{u_0}(y,y') dy' + \int_D \int_D j_{u_0}(x,y') m_{u_0}(y',x') j_{u_0}(x',y) dx' dy'.$$

This implies that the operator $A_{u_0}^{-1} = (I + J_{u_0})^{-1}$ is a neural operator, too. Observe that $\widetilde{j}_{u_0}(x,y), \partial_x \widetilde{j}_{u_0}(x,y) \in C(\overline{D} \times \overline{D})$.

This proves Theorem 3. $\square$