Submitted to Bernoulli

Supplement to "On the convergence of PINNs"

NATHAN DOUMÈCHE[1,a], GÉRARD BIAU[1,b] and CLAIRE BOYER[1,c]

[1]Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université,
[a]nathan.doumeche@sorbonne-universite.fr, [b]gerard.biau@sorbonne-universite.fr,
[c]claire.boyer@sorbonne-universite.fr

1. Some reminders of functional analysis on Lipschitz domains

Extension theorems Let $\Omega \subseteq \mathbb{R}^{d_1}$ be an open set and let $K \in \mathbb{N}$ be an order of differentiation. It is not straightforward to extend a function $u \in H^K(\Omega, \mathbb{R}^{d_2})$ to a function $\tilde{u} \in H^K(\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$ such that

$$\tilde{u}|_\Omega = u|_\Omega \quad \text{and} \quad \|\tilde{u}\|_{H^K(\mathbb{R}^{d_1})} \leqslant C_\Omega \|u\|_{H^K(\Omega)},$$

for some constant C_Ω independent of u. This result is known as the extension theorem in Evans (2010, Chapter 5.4) when Ω is a manifold with C^1 boundary. However, the simplest domains in PDEs take the form $]0, L[^3 \times]0, T[$, the boundary of which is not C^1. Fortunately, Stein (1970, Theorem 5 Chapter VI.3.3) provides an extension theorem for bounded Lipschitz domains. We refer the reader to Shvartzman (2010) for a survey on extension theorems.

Example of a non-extendable domain Let the domain $\Omega =]-1, 1[^2 \backslash (\{0\} \times [0, 1[)$ be the square $]-1, 1[^2$ from which the segment $\{0\} \times [0, 1[$ has been removed. Then the function

$$u(x, y) = \begin{cases} 0 & \text{if } x < 0 \text{ or if } y \leqslant 0 \\ \exp(-\frac{1}{y}) & \text{if } x, y > 0, \end{cases}$$

belongs to $C^\infty(\Omega, \mathbb{R})$ but cannot be extended to $\mathbb{R}^2$, since it cannot be continuously extended to the segment $\{0\} \times [0, 1[$. Notice that Ω is not a Lipschitz domain because it lies on both sides of the segment $\{0\} \times [0, 1[$, which belongs to its boundary $\partial\Omega$.

Theorem 1.1 (Sobolev inequalities). *Let $\Omega \subseteq \mathbb{R}^{d_1}$ be a bounded Lipschitz domain and let $m \in \mathbb{N}$. If $m \geqslant d_1/2$, then there exists an operator $\tilde{\Pi} : H^m(\Omega, \mathbb{R}^{d_2}) \to C^0(\Omega, \mathbb{R}^{d_2})$ such that, for any $u \in H^m(\Omega, \mathbb{R}^{d_2})$, $\tilde{\Pi}(u) = u$ almost everywhere. Moreover, there exists a constant $C_\Omega > 0$, depending only on Ω, such that, $\|\tilde{\Pi}(u)\|_{\infty,\Omega} \leqslant C_\Omega \|u\|_{H^m(\Omega)}$.*

Proof. Since Ω is a bounded Lipschitz domain, there exists a radius $r > 0$ such that $\Omega \subseteq B(0, r)$. According to the extension theorem (Stein, 1970, Theorem 5, Chapter VI.3.3), there exists a constant $C_\Omega > 0$, depending only on Ω, such that any $u \in H^m(\Omega, \mathbb{R}^{d_2})$ can be extended to $\tilde{u} \in H^m(B(0, r), \mathbb{R}^{d_2})$, with $\|\tilde{u}\|_{H^m(B(0,r))} \leqslant C_\Omega \|u\|_{H^m(\Omega)}$. Since $m \geqslant d_1/2$, the Sobolev inequalities (e.g., Evans, 2010, Chapter 5.6, Theorem 6) state that there exists a constant $\tilde{C}_\Omega > 0$, depending only on Ω, and a linear embedding $\Pi : H^m(B(0, r), \mathbb{R}^{d_2}) \to C^0(B(0, r), \mathbb{R}^{d_2})$ such that $\|\Pi(\tilde{u})\|_\infty \leqslant \tilde{C}_\Omega \|\tilde{u}\|_{H^m(B(0,r))}$ and $\Pi(\tilde{u}) = \tilde{u}$ in $H^m(B(0, r), \mathbb{R}^{d_2})$. Therefore, $\tilde{\Pi}(u) = \Pi(\tilde{u})|_\Omega$ and $\|\tilde{\Pi}(u)\|_{\infty,\Omega} \leqslant C_\Omega \tilde{C}_\Omega \|u\|_{H^m(\Omega)}$. □

Definition 1.2 (Weak convergence in $L^2(\Omega)$). A sequence $(u_p)_{p \in \mathbb{N}} \in L^2(\Omega)^{\mathbb{N}}$ weakly converges to $u_\infty \in L^2(\Omega)$ if, for any $\phi \in L^2(\Omega)$, $\lim_{p\to\infty} \int_\Omega \phi u_p = \int_\Omega \phi u_\infty$. This convergence is denoted by $u_p \rightharpoonup u_\infty$.

The Cauchy-Schwarz inequality shows that the convergence with respect to the $L^2(\Omega)$ norm implies the weak convergence. However, the converse is not true. For example, the sequence of functions $u_p(x) = \cos(px)$ weakly converges to 0 in $L^2([-\pi,\pi])$, whereas $\|u_p\|_{L^2([-\pi,\pi])} = 1/2$.

Definition 1.3 (Weak convergence in $H^m(\Omega)$). A sequence $(u_p)_{p\in\mathbb{N}} \in H^m(\Omega)^{\mathbb{N}}$ weakly converges to $u_\infty \in H^m(\Omega)$ in $H^m(\Omega)$ if, for all $|\alpha| \leqslant m$, $\partial^\alpha u_p \rightharpoonup \partial^\alpha u_\infty$.

Theorem 1.4 (Rellich-Kondrachov). *Let $\Omega \subseteq \mathbb{R}^{d_1}$ be a bounded Lipschitz domain and let $m \in \mathbb{N}$. Let $(u_p)_{p\in\mathbb{N}} \in H^{m+1}(\Omega,\mathbb{R}^{d_2})$ be a sequence such that $(\|u_p\|_{H^{m+1}(\Omega)})_{p\in\mathbb{N}}$ is bounded. There exists a function $u_\infty \in H^{m+1}(\Omega,\mathbb{R}^{d_2})$ and a subsequence of $(u_p)_{p\in\mathbb{N}}$ that converges to u_∞ both weakly in $H^{m+1}(\Omega,\mathbb{R}^{d_2})$ and with respect to the $H^m(\Omega)$ norm.*

Proof. Let $r > 0$ be such that $\Omega \subseteq B(0,r)$. According to the extension theorem of Stein (1970, Theorem 5, Chapter VI.3.3), there exists a constant $C_r > 0$ such that each u_p can be extended to $\tilde{u}_p \in H^{m+1}(B(0,r),\mathbb{R}^{d_2})$, with $\|\tilde{u}_p\|_{H^{m+1}(B(0,r))} \leqslant C_r\|u_p\|_{H^{m+1}(\Omega)}$. Observing that, for all $|\alpha| \leqslant m$, $\partial^\alpha\tilde{u}_p$ belongs to $H^1(B(0,r),\mathbb{R}^{d_2})$, the Rellich-Kondrachov compactness theorem (Evans, 2010, Theorem 1, Chapter 5.7) ensures that there exists a subsequence of $(\tilde{u}_p)_{p\in\mathbb{N}}$ that converges to an extension of u_∞ with respect to the $H^m(B(0,r))$ norm. Since the subsequence is also bounded, upon passing to another subsequence, it also weakly converges in $H^{m+1}(B(0,r),\mathbb{R}^{d_2})$ to $u_\infty \in H^{m+1}(B(0,r),\mathbb{R}^{d_2})$ (e.g., Evans, 2010, Chapter D.4). Therefore, by considering the restrictions of all the previous functions to Ω, we deduce that there exists a subsequence of $(u_p)_{p\in\mathbb{N}}$ that converges to u_∞ both weakly in $H^{m+1}(\Omega)$ and with respect to the $H^m(\Omega)$ norm. □

2. Some useful lemmas

The nth Bell number B_n (Hardy, 2006) corresponds to the number of partitions of the set $\{1,\dots,n\}$. Bell numbers satisfy the relationship $B_0 = 1$ and

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k. \tag{1}$$

For $K \geqslant 1$ and $u \in C^K(\mathbb{R}^{d_1},\mathbb{R}^{d_2})$, the Kth derivative of u is denoted by $u^{(K)}$.

Lemma 2.1 (Bounding the partial derivatives of a composition of functions). *Let $d_1, d_2 \geqslant 1$, $K \geqslant 0$, $f \in C^K(\mathbb{R}^{d_1},\mathbb{R})$, and $g \in C^K(\mathbb{R},\mathbb{R}^{d_2})$. Then*

$$\|g\circ f\|_{C^K(\mathbb{R}^{d_1})} \leqslant B_K\|g\|_{C^K(\mathbb{R})}(1+\|f\|_{C^K(\mathbb{R}^{d_1})})^K.$$

Proof. Let $K_1 \leqslant K$ and let $\Pi(K_1)$ be the set of all partitions of $\{1,\dots,K_1\}$. According to Hardy (2006, Proposition 1), one has, for all $h \in C^{K_1}(\mathbb{R}^{K_1+d_1},\mathbb{R})$,

$$\partial^{K_1}_{1,2,3,\dots,K_1}(g\circ h) = \sum_{P\in\Pi(K_1)} g^{(|P|)}\circ h \times \prod_{S\in P}\Big[\big(\prod_{j\in S}\partial_j\big)h\Big].$$

Let $\alpha = (\alpha_1,\dots,\alpha_{d_1})$ be a multi-index such that $|\alpha| = K_1$. Setting $\alpha_0 = 0$, $y_j = x_{K_1+j} + (x_{\alpha_1+\cdots+\alpha_{j-1}} + \cdots + x_{\alpha_1+\cdots+\alpha_j-1})$, and letting $h(x_1,\dots,x_{K_1+d_1}) = f(y_1,\dots,y_{d_1})$, we are led to

$$\partial^\alpha(g\circ f) = \sum_{P\in\Pi(K_1)} g^{(|P|)}\circ f \times \prod_{S\in P}\partial^{\alpha(S)} f, \tag{2}$$

where $\alpha(S) = (|\{b \in S, \quad \alpha_1 + \cdots + \alpha_{\ell-1} \leqslant b \leqslant \alpha_1 + \cdots + \alpha_\ell\}|)_{1 \leqslant \ell \leqslant d_1}$. Moreover, by definition of the Bell number, $|\Pi(K_1)| = B_{K_1}$, and, by definition of a partition, $|P| \leqslant K_1$. So,

$$\begin{aligned}\|\partial^\alpha (g \circ f)\|_\infty &\leqslant B_{K_1} \|g\|_{C^{K_1}(\mathbb{R}^{d_1})} \max_{i_1+2i_2+\cdots+K_1 i_{K_1}=K_1} \prod_{j=1}^{K_1} \|f\|^{i_j}_{C^j(\mathbb{R}^{d_1})} \\ &\leqslant B_{K_1} \|g\|_{C^{K_1}(\mathbb{R}^{d_1})} (1 + \|f\|_{C^{K_1}(\mathbb{R}^{d_1})})^{K_1}.\end{aligned}$$

Since this inequality is true for all $K_1 \leqslant K$ and for all $|\alpha| = K_1$, the lemma is proved. □

Lemma 2.2 (Bounding the partial derivatives of a changing of coordinates f). *Let $d_1, d_2 \geqslant 1$, $K \geqslant 0$, $f \in C^K(\mathbb{R}, \mathbb{R})$, and $g \in C^K(\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$. Let $v \in C^K(\mathbb{R}^{d_1}, \mathbb{R}^{d_1})$ be defined by $v(\mathbf{x}) = (f(x_1), \ldots, f(x_{d_1}))$. Then*

$$\|g \circ v\|_{C^K(\mathbb{R}^{d_1})} \leqslant B_K \times \|g\|_{C^K(\mathbb{R}^{d_1})} \times (1 + \|f\|_{C^K(\mathbb{R})})^K.$$

Proof. Let $\alpha = (\alpha_1, \ldots, \alpha_{d_1})$ be a multi-index such that $|\alpha| = K$. For $\mathbf{x} = (x_1, \ldots, x_{d_1})$ and a fixed $i \in \{1, \ldots, d_1\}$, we let $h(t) = g(f(x_1), \ldots, f(x_{i-1}), t, f(x_{i+1}), \ldots, f(x_{d_1}))$. Clearly, $(h \circ f)^{(\alpha_i)}(x_i) = (\partial_i)^{\alpha_i}(g \circ v)(\mathbf{x})$. Thus, according to Lemma 2.1,

$$(h \circ f)^{(\alpha_i)} = \sum_{P_i \in \Pi(\alpha_i)} h^{(|P_i|)} \circ f \times \prod_{S_i \in P_i} f^{(|S_i|)}.$$

Therefore,

$$(\partial_i)^{\alpha_i}(g \circ v)(\mathbf{x}) = \sum_{P_i \in \Pi(\alpha_i)} (\partial_i)^{|P_i|} g \circ v(\mathbf{x}) \prod_{S_i \in P_i} f^{(|S_i|)}(x_i).$$

Letting $i = 1$ and observing that $\partial_j f^{(|S_1|)}(x_1) = 0$ for $j \neq 1$, we see that

$$\partial^\alpha (g \circ v)(\mathbf{x}) = \sum_{P_1 \in \Pi(\alpha_1)} \Big[\prod_{S_1 \in P_1} f^{(|S_1|)}(x_1) \Big] \times (\partial_2)^{\alpha_2} \ldots (\partial_{d_1})^{\alpha_{d_1}} [(\partial_1)^{|P_1|} g \circ v](\mathbf{x}).$$

Repeating the same procedure for $(\partial_1)^{|P_1|} g \circ v, \ldots, (\partial_1)^{|P_1|} \ldots (\partial_{d_1})^{|P_{d_1}|} g \circ v$, we obtain

$$\begin{aligned}\partial^\alpha (g \circ v)(\mathbf{x}) = &\sum_{P_1 \in \Pi(\alpha_1)} \Big[\prod_{S_1 \in P_1} f^{(|S_1|)}(x_1)) \Big] \times \cdots \\ &\cdots \times \sum_{P_{d_1} \in \Pi(\alpha_{d_1})} \Big[\prod_{S_{d_1} \in P_{d_1}} f^{(|S_{d_1}|)}(x_{d_1})) \Big] \times (\partial_1)^{|P_1|} \ldots (\partial_{d_1})^{|P_{d_1}|} g \circ v(\mathbf{x}).\end{aligned}$$

Since $\sum_{S_i \in P_i} |S_i| = \alpha_i$ and $\sum_{i=1}^{d_1} \alpha_i = K$, we conclude that

$$\|\partial^\alpha (g \circ v)\|_\infty \leqslant B_{\alpha_1} \times \cdots \times B_{\alpha_{d_1}} \times \|\partial^\alpha g\|_\infty (1 + \|f\|_{C^K(\mathbb{R})})^K.$$

Using the injective map $\mathcal{M} : \Pi(\alpha_1) \times \cdots \times \Pi(\alpha_{d_1}) \to \Pi(K)$ such that $\mathcal{M}(P_1, \ldots, P_{d_1}) = \cup_{i=1}^{d_1} P_i$, we have $B_{\alpha_1} \times \cdots \times B_{\alpha_{d_1}} \leqslant B_K$. This concludes the proof. □

Lemma 2.3 (Bounding hyperbolic tangent and its derivatives). *For all $K \in \mathbb{N}$, one has*

$$\| \tanh^{(K)} \|_\infty \leqslant 2^{K-1}(K+2)!$$

Proof. The tanh function is a solution of the equation $y' = 1 - y^2$. An elementary induction shows that there exists a sequence of polynomials $(P_K)_{K\in\mathbb{N}}$ such that $\tanh^{(K)} = P_K(\tanh)$, with $P_0(X) = X$ and $P_{K+1}(X) = (1 - X^2) \times P'_K(X)$. Clearly, P_K is a real polynomial of degree $K + 1$, of the form $P_K(X) = a_0^{(K)} + a_1^{(K)} X + \cdots + a_{K+1}^{(K)} X^{K+1}$. One verifies that $a_i^{(K+1)} = (i+1)a_{i+1}^{(K)} - (i-1)a_{i-1}^{(K)}$, with $a_{-1}^{(K)} = a_{K+2}^{(K)} = 0$. The largest coefficient $M(P_K) = \max_{0\leqslant i\leqslant K+1} |a_i^{(K)}|$ of P_K satisfies $M(P_{K+1}) \leqslant 2(K+1) \times M(P_K)$. Thus, since $M(P_1) = 1$, we see that $M(P_K) \leqslant 2^{K-1}K!$. Recalling that $0 \leqslant \tanh \leqslant 1$, we conclude that

$$\| \tanh^{(K)} \|_\infty = \|P_K(\tanh)\|_\infty \leqslant (K+2)M(P_K) \leqslant 2^{K-1}(K+2)!$$

□

In the sequel, for all $\theta \in \mathbb{R}$, we write $\tanh_\theta(x) = \tanh(\theta x)$. We define the sign function such that $\mathrm{sgn}(x) = \mathbf{1}_{x>0} - \mathbf{1}_{x<0}$.

Lemma 2.4 (Characterizing the limit of hyperbolic tangent in Hölder norm). *Let $K \in \mathbb{N}$ and $H \in \mathbb{N}^\star$. Then, for all $\varepsilon > 0$, $\lim_{\theta\to\infty} \| \tanh_\theta^{\circ H} - \mathrm{sgn}\|_{C^K(\mathbb{R}\setminus]-\varepsilon,\varepsilon[)} = 0$.*

Proof. Fix $\varepsilon > 0$. We prove the stronger statement that, for all $m \in \mathbb{N}$, one has

$$\lim_{\theta\to\infty} \theta^m \| \tanh_\theta^{\circ H} - \mathrm{sgn}\|_{C^K(\mathbb{R}\setminus]-\varepsilon,\varepsilon[)} = 0.$$

We start with the case $H = 1$ and then prove the result by induction on H. Observe first, since $\tanh_\theta^{\circ H} - \mathrm{sgn}$ is an odd function, that

$$\| \tanh_\theta^{\circ H} - \mathrm{sgn}\|_{C^K(\mathbb{R}\setminus]-\varepsilon,\varepsilon[)} = \| \tanh_\theta^{\circ H} - \mathrm{sgn}\|_{C^K([\varepsilon,\infty[)}.$$

The case $H = 1$ Assume, to start with, that $K = 0$. For all $x \geqslant \varepsilon$, one has

$$\theta^m |\tanh_\theta(x) - 1| = \frac{2\theta^m}{1+\exp(-2\theta x)} \leqslant \frac{2\theta^m}{1+\exp(-2\theta\varepsilon)}.$$

Therefore, for all $m \in \mathbb{N}$,

$$\theta^m \| \tanh_\theta - \mathrm{sgn}\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[} = \theta^m \| \tanh_\theta - \mathrm{sgn}\|_{\infty,[\varepsilon,\infty[} \leqslant \frac{2\theta^m}{1+\exp(-2\theta\varepsilon)} \xrightarrow{\theta\to\infty} 0.$$

Next, to prove that the result if true for all $K \geqslant 1$, it is enough to show that, for all m,

$$\theta^m \| \tanh_\theta^{(K)} \|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[} \xrightarrow{\theta\to\infty} 0.$$

According to the proof of Lemma 2.3, there exists a sequence of polynomials $(P_K)_{K\in\mathbb{N}}$ such that $\tanh^{(K)} = P_K(\tanh)$ and $P_{K+1}(X) = (1 - X^2) \times P'_K(X)$. Since $\tanh_\theta(x) = \tanh(\theta x)$, one has

$$\begin{aligned}\tanh_\theta^{(K)}(x) &= \theta^K \tanh^{(K)}(\theta x)\\ &= \theta^K (1 - \tanh^2(\theta x)) \times P'_{K-1}(\tanh(\theta x))\\ &= \theta^K (1 - \tanh(\theta x))(1 + \tanh(\theta x)) \times P'_{K-1}(\tanh(\theta x)).\end{aligned}$$

Fix $x \geqslant \varepsilon$. Then, letting $M_K = \|P'_{K-1}\|_{\infty,[-1,1]}$, we are led to

$$|\tanh_\theta^{(K)}(x)| \leqslant 2M_K\theta^K(1-\tanh(\theta x)) \leqslant 4M_K \times \frac{\theta^K}{1+\exp(2\theta x)} \leqslant 4M_K \times \frac{\theta^K}{1+\exp(2\theta\varepsilon)}.$$

This shows that $\theta^m \|\tanh_\theta^{(K)}\|_{\infty,[\varepsilon,\infty[} \leqslant 4M_K \times \frac{\theta^{K+m}}{1+\exp(2\theta\varepsilon)}$. One proves with similar arguments that the same result holds for all $x \leqslant -\varepsilon$. Thus,

$$\theta^m \|\tanh_\theta^{(K)}\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[} \leqslant 4M_K \times \frac{\theta^{K+m}}{1+\exp(2\theta\varepsilon)} \xrightarrow{\theta\to\infty} 0,$$

and the lemma is proved for $H = 1$. **Induction** Assume that that, for all K and all m,

$$\theta^m \|\tanh_\theta^{\circ H} - \mathrm{sgn}\|_{C^K(\mathbb{R}\setminus]-\varepsilon,\varepsilon[)} \xrightarrow{\theta\to\infty} 0. \tag{3}$$

Our objective is to prove that, for all K_2 and all m_2,

$$\theta^{m_2} \|\tanh_\theta^{\circ(H+1)} - \mathrm{sgn}\|_{C^{K_2}(\mathbb{R}\setminus]-\varepsilon,\varepsilon[)} \xrightarrow{\theta\to\infty} 0.$$

If $K_2 = 0$, since, for all $(x,y) \in \mathbb{R}^2$, $|\tanh_\theta(x) - \tanh_\theta(y)| \leqslant \theta|x-y| \times \|\tanh'\|_\infty \leqslant \theta|x-y|$. We deduce that

$$\theta^{m_2} \|\tanh_\theta^{\circ(H+1)} - \tanh_\theta(\mathrm{sgn})\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[} \leqslant \theta^{m_2+1} \|\tanh_\theta^{\circ H} - \mathrm{sgn}\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[}.$$

Therefore, according to (3), we have that $\lim_{\theta\to\infty} \theta^{m_2} \|\tanh_\theta^{\circ(H+1)} - \tanh_\theta(\mathrm{sgn})\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[} = 0$. Since $\tanh_\theta(\mathrm{sgn}) - \mathrm{sgn} = (\tanh(\theta)-1)\mathbf{1}_{x>0} - (\tanh(\theta)-1)\mathbf{1}_{x<0}$, we see that, for all m_2,

$$\lim_{\theta\to\infty} \theta^{m_2} \|\tanh_\theta(\mathrm{sgn}) - \mathrm{sgn}\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[} = 0.$$

Using the triangle inequality, we conclude as desired that, for all m_2,

$$\theta^{m_2} \|\tanh_\theta^{\circ(H+1)} - \mathrm{sgn}\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[} \xrightarrow{\theta\to\infty} 0. \tag{4}$$

Assume now that $K_2 \geqslant 1$. Since $\tanh_\theta^{\circ(H+1)} = \tanh^{\circ H}(\tanh)$, the Faà di Bruno formula (e.g., Comtet, 1974, Chapter 3.4) states that

$$(\tanh_\theta^{\circ(H+1)})^{(K_2)} = \sum_{m_1+2m_2+\cdots+K_2 m_{K_2}=K_2} \frac{K_2!}{\prod_{i=1}^{K_2} m_i! \times i!^{m_i}} \times (\tanh_\theta^{\circ H})^{(m_1+\cdots+m_{K_2})}(\tanh_\theta) \times \prod_{j=1}^{K_2} (\tanh_\theta^{(j)})^{m_j}.$$

Notice that if $|x| \leq \operatorname{arctanh}(1/\sqrt{2})$, $|\tanh(x)| \geqslant \frac{|x|}{2}$ because by calling $f(x) = \tanh(x) - \frac{x}{2}$, $f(0) = 0$ and $f'(x) = (1-\tanh(x)^2) - \frac{1}{2} \geqslant 0$. Therefore, if $|x| \geq \varepsilon$, $|\tanh(\theta x)| \geqslant \min(\frac{1}{\sqrt{2}}, \frac{\theta}{2}\varepsilon) \geqslant \varepsilon$ if $\theta \geqslant 2$ and

$\varepsilon \geqslant \frac{1}{\sqrt{2}}$. This is why for $\theta \geqslant 2$ and $\varepsilon \leqslant 1$,

$$\|(\tanh_\theta^{\circ H})^{(m_1+\cdots+m_{K_2})}(\tanh_\theta)\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[} \leqslant \|(\tanh_\theta^{\circ H})^{(m_1+\cdots+m_{K_2})}\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[}.$$

Therefore, from the triangular inequality on $\|\cdot\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[}$,

$$\|(\tanh_\theta^{\circ(H+1)})^{(K_2)}\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[} \leqslant \sum_{m_1+2m_2+\cdots+K_2m_{K_2}=K_2} \frac{K_2!}{\prod_{i=1}^{K_2} m_i! \times i!^{m_i}} \times \|(\tanh_\theta^{\circ H})^{(m_1+\cdots+m_{K_2})}\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[} \prod_{j=1}^{K_2} \|\tanh_\theta^{(j)}\|^{m_j}_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[}.$$

According to the induction hypothesis (3), one has, for all $K \geqslant 1$ and all $m \in \mathbb{N}$,

$$\lim_{\theta\to\infty} \theta^m \|(\tanh_\theta^{\circ H})^{(K)}\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[} = 0.$$

We deduce from the above that for all $K_2 \geqslant 1$ and all m_2,

$$\theta^{m_2}\|(\tanh_\theta^{\circ(H+1)})^{(K_2)}\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[} \xrightarrow{\theta\to\infty} 0. \tag{5}$$

Combining (4) and (5), it comes that $\lim_{\theta\to\infty}\theta^{m_2}\|\tanh_\theta^{\circ(H+1)} - \mathrm{sgn}\|_{C^{K_2}(\mathbb{R}\setminus]-\varepsilon,\varepsilon[)} = 0$. □

Corollary 2.5 (Bounding hyperbolic tangent compositions and their derivatives). *Let $K \in \mathbb{N}$ and $H \in \mathbb{N}^\star$. Then, for or all $\theta \in \mathbb{R}$, $\|(\tanh_\theta^{\circ H})^{(K)}\|_\infty < \infty$.*

Proof. An induction as the one of Lemma 2.4 shows that $\|(\tanh_\theta^{\circ H})^{(K)}\|_{\infty,\mathbb{R}\setminus]-\varepsilon,\varepsilon[} < \infty$. In addition, since $\tanh_\theta^{\circ H} \in C^\infty(\mathbb{R},\mathbb{R})$, $\|(\tanh_\theta^{\circ H})^{(K)}\|_{\infty,[-\varepsilon,\varepsilon]} < \infty$. □

When $d_1 = d_2 = 1$, the observations $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n) \in \mathbb{R}^2$ can be reordered as $(\mathbf{X}_{(1)}, Y_{(1)}), \ldots, (\mathbf{X}_{(n)}, Y_{(n)})$ according to increasing values of the $\mathbf{X}_i$, that is, $\mathbf{X}_{(1)} \leqslant \cdots \leqslant \mathbf{X}_{(n)}$. Moreover, we let $\mathcal{G}(n, n_r) = \{(\mathbf{X}_i, Y_i), 1 \leqslant i \leqslant n\} \cup \{\mathbf{X}_j^{(r)}, 1 \leqslant j \leqslant n_r\}$, and denote by $\delta(n, n_r)$ the minimum distance between two distinct points in $\mathcal{G}(n, n_r)$, i.e.,

$$\delta(n, n_r) = \min_{\substack{z_1, z_2 \in \mathcal{G}(n,n_r)\\ z_1 \neq z_2}} |z_1 - z_2|. \tag{6}$$

Lemma 2.6 (Exact estimation with hyperbolic tangent). *Assume that $d_1 = d_2 = 1$, and let $H \geqslant 1$. Let the neural network $u_\theta \in \mathrm{NN}_H(n-1)$ be defined by*

$$u_\theta(x) = Y_{(1)} + \sum_{i=1}^{n-1} \frac{Y_{(i+1)} - Y_{(i)}}{2}\left[\tanh_\theta^{\circ H}\left(x - \mathbf{X}_{(i)} - \frac{\delta(n, n_r)}{2}\right) + 1\right].$$

Then, for all $1 \leqslant i \leqslant n$,

$$\lim_{\theta\to\infty} u_\theta(\mathbf{X}_i) = Y_i.$$

Moreover, for all order $K \in \mathbb{N}^\star$ of differentiation and all $1 \leqslant j \leqslant n_r$,

$$\lim_{\theta\to\infty} u_\theta^{(K)}(\mathbf{X}_j^{(r)}) = 0.$$

Proof. Applying Lemma 2.4 with $\varepsilon = \delta(n,n_r)/4$ and letting

$$G = \mathbb{R}\backslash \cup_{i=1}^{n}]\mathbf{X}_{(i)} + \frac{1}{4}\delta(n,n_r), \mathbf{X}_{(i)} + \frac{3}{4}\delta(n,n_r)[,$$

one has, for all K, $\lim_{\theta\to\infty} \|u_\theta - u_\infty\|_{C^K(G)} = 0$, where

$$u_\infty(x) = Y_{(1)} + \sum_{i=1}^{n-1} \big[Y_{(i+1)} - Y_{(i)}\big] \times \mathbf{1}_{x > \mathbf{X}_{(i)} + \frac{\delta(n,n_r)}{2}}.$$

Clearly, for all $1 \leqslant i \leqslant n$, $u_\infty(\mathbf{X}_i) = Y_i$. Since $u'_\infty(x) = 0$ for all $x \in G$, and since $\mathbf{X}_j^{(r)} \in G$ for all $1 \leqslant j \leqslant n_r$, we deduce that $u_\infty^{(K)}(\mathbf{X}_j^{(r)}) = 0$. This concludes the proof. □

Definition 2.7 (Overfitting gap). For any $n, n_e, n_r \in \mathbb{N}^\star$ and $\lambda_{(\text{ridge})} \geqslant 0$, the overfitting gap operator OG_{n,n_e,n_r} is defined, for all $u \in C^\infty(\bar{\Omega}, \mathbb{R}^{d_2})$, by

$$\text{OG}_{n,n_e,n_r}(u) = |R^{(\text{ridge})}_{n,n_e,n_r}(u) - \mathscr{R}_n(u)|.$$

Lemma 2.8 (Monitoring the overfitting gap). *Let $\varepsilon > 0$, $\lambda_{(\text{ridge})} \geqslant 0$, $H \geqslant 2$, and $D \in \mathbb{N}^\star$. Let $n, n_e, n_r \in \mathbb{N}^\star$. Let $\hat{\theta} \in \Theta_{H,D}$ be a parameter such that (i) $R^{(\text{ridge})}_{n,n_e,n_r}(u_{\hat{\theta}}) \leqslant \inf_{u\in \text{NN}_H(D)} R^{(\text{ridge})}_{n,n_e,n_r}(u) + \varepsilon$ and (ii) $\text{OG}_{n,n_e,n_r}(u_{\hat{\theta}}) \leqslant \varepsilon$. Then*

$$\mathscr{R}_n(u_{\hat{\theta}}) \leqslant \inf_{u\in \text{NN}_H(D)} \mathscr{R}_n(u) + 2\varepsilon + o_{n_e,n_r\to\infty}(1).$$

Proof. On the one hand, since $\mathscr{R}_n \leqslant R^{(\text{ridge})}_{n,n_e,n_r} + \text{OG}_{n,n_e,n_r}$, assumptions (i) and (ii) imply that $\mathscr{R}_n(u_{\hat{\theta}}) \leqslant \inf_{u\in \text{NN}_H(D)} R^{(\text{ridge})}_{n,n_e,n_r}(u) + 2\varepsilon$. On the other hand, $R^{(\text{ridge})}_{n,n_e,n_r} - \text{OG}_{n,n_e,n_r} \leqslant \mathscr{R}_n$. The proof of Theorem 4.6 reveals that there exists a sequence $(\theta(n_e,n_r))_{n_e,n_r\in\mathbb{N}} \in \Theta_{H,D}^{\mathbb{N}}$ such that $\lim_{n_e,n_r\to\infty} \text{OG}_{n,n_e,n_r}(u_{\theta(n_e,n_r)}) = 0$ and $\lim_{n_e,n_r\to\infty} \mathscr{R}_n(u_{\theta(n_e,n_r)}) = \inf_{u\in\text{NN}_H(D)} \mathscr{R}_n(u)$. Thus, $\inf_{u\in\text{NN}_H(D)} R^{(\text{ridge})}_{n,n_e,n_r}(u) \leqslant \inf_{\text{NN}_H(D)} \mathscr{R}_n(u) + o_{n_e,n_r\to\infty}(1)$. We deduce that

$$\mathscr{R}_n(u_{\hat{\theta}}) \leqslant \inf_{u\in \text{NN}_H(D)} \mathscr{R}_n(u) + 2\varepsilon + o_{n_e,n_r\to\infty}(1).$$

□

Lemma 2.9 (Minimizing sequence of the theoretical risk.). *Let $H, D \in \mathbb{N}^\star$. Define the sequence $(v_p)_{p\in\mathbb{N}} \in \text{NN}_H(D)^{\mathbb{N}}$ of neural networks by $v_p(\mathbf{x}) = \tanh_p \circ \tanh^{\circ(H-1)}(\mathbf{x})$. Then, for any $\lambda_e > 0$,*

$$\lim_{p\to\infty} \lambda_e(1 - v_p(1))^2 + \frac{1}{2}\int_{-1}^{1} \mathbf{x}^2 (v'_p)^2(\mathbf{x}) d\mathbf{x} = 0.$$

Proof. $\tanh^{\circ(H-1)}$ is an increasing C^∞ function such that $\tanh^{\circ(H-1)}(0) = 0$. Therefore, Lemma 2.4 shows that $\lim_{p\to\infty} v_p(1) = 1$, so that $\lim_{p\to\infty} \lambda_e(1 - v_p(1))^2 = 0$. This shows the convergence of the left-hand term of the lemma.

To bound the right-hand term, we have, according to the chain rule,

$$|v'_p(\mathbf{x})| \leqslant p \|\tanh^{\circ(H-1)}\|_{C^1(\mathbb{R})} |\tanh'(p \tanh^{\circ(H-1)}(\mathbf{x}))|,$$

with $\|\tanh^{\circ(H-1)}\|_{C^1(\mathbb{R})} < \infty$ by Corollary 2.5. Thus,

$$\int_{-1}^{1} \mathbf{x}^2 (v_p')^2(\mathbf{x}) d\mathbf{x} \leqslant \|\tanh^{\circ(H-1)}\|_{C^1(\mathbb{R})}^2 \int_{-1}^{1} p^2 \mathbf{x}^2 (\tanh'(p \tanh^{\circ(H-1)}(\mathbf{x})))^2 d\mathbf{x}.$$

Notice that $\mathbf{x}^2(\tanh'(p\tanh^{\circ(H-1)}(\mathbf{x})))^2$ is an even function, so that

$$\int_{-1}^{1} \mathbf{x}^2 (v_p')^2(\mathbf{x}) d\mathbf{x} \leqslant 2\|\tanh^{\circ(H-1)}\|_{C^1(\mathbb{R})}^2 \int_{0}^{1} p^2 \mathbf{x}^2 (\tanh'(p \tanh^{\circ(H-1)}(\mathbf{x})))^2 d\mathbf{x}.$$

Remark that $(\tanh')^2(\mathbf{x}) = (1-\tanh(\mathbf{x}))^2(1+\tanh(\mathbf{x}))^2 \leqslant 16\exp(-2\mathbf{x})$, so that

$$\int_{-1}^{1} \mathbf{x}^2 (v_p')^2(\mathbf{x}) d\mathbf{x} \leqslant 32\|\tanh^{\circ(H-1)}\|_{C^1(\mathbb{R})}^2 \int_{0}^{1} p^2 \mathbf{x}^2 \exp(-2p \tanh^{\circ(H-1)}(\mathbf{x})) d\mathbf{x}.$$

If $H = 1$, then the change of variable $\bar{\mathbf{x}} = p\mathbf{x}$ states that

$$\int_0^1 p^2 \mathbf{x}^2 \exp(-2p\mathbf{x}) d\mathbf{x} \leqslant p^{-1} \int_0^\infty \bar{\mathbf{x}}^2 \exp(-2\bar{\mathbf{x}}) d\bar{\mathbf{x}} \xrightarrow{p\to\infty} 0$$

and the lemma is proved.

If $H \geqslant 2$, notice that $\tanh(\mathbf{x}) \geqslant \mathbf{x}\mathbf{1}_{\mathbf{x}\leqslant 1}/2 + \mathbf{1}_{\mathbf{x}\geqslant 1}/2$ for all $\mathbf{x} \geqslant 0$, and therefore we have that $\tanh^{\circ(H-1)}(\mathbf{x}) \geqslant \mathbf{x}\mathbf{1}_{\mathbf{x}\leqslant 2^{H-1}}/2^H + \mathbf{1}_{\mathbf{x}\geqslant 2^{H-1}}/2^H$. Thus, using the change of variable $\bar{\mathbf{x}} = p\mathbf{x}$,

$$\begin{aligned}\int_0^1 p^2 \mathbf{x}^2 \exp(-2p \tanh^{\circ(H-1)}(\mathbf{x})) d\mathbf{x} &\leqslant \int_0^1 p^2 \mathbf{x}^2 \exp(-2^{H-1} p\mathbf{x}) d\mathbf{x} \\ &\leqslant p^{-1} \int_0^\infty \bar{\mathbf{x}}^2 \exp(-2^{H-1}\bar{\mathbf{x}}) d\bar{\mathbf{x}}.\end{aligned}$$

Since this upper bound vanishes as $p \to \infty$, this conclude s the proof when $H \geqslant 2$.

□

Definition 2.10 (Weak lower semi-continuity). A fonction $I : H^m(\Omega) \to \mathbb{R}$ is weakly lower semi-continuous on $H^m(\Omega)$ if, for any sequence $(u_p)_{p\in\mathbb{N}} \in H^m(\Omega)^{\mathbb{N}}$ that weakly converges to $u_\infty \in H^m(\Omega)$ in $H^m(\Omega)$, one has $I(u_\infty) \leqslant \liminf_{p\to\infty} I(u_p)$.

The following technical lemma will be useful for the proof of Proposition 5.6.

Lemma 2.11 (Weak lower semi-continuity with convex Lagrangians). *Let the Lagrangian* $L \in C^\infty(\mathbb{R}^{\binom{d_1+m}{m}d_2} \times \cdots \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_1}, \mathbb{R})$ *be such that, for any* $x^{(m)}, \dots, x^{(0)}$*, and* z*, the function* $x^{(m+1)} \mapsto L(x^{(m+1)}, \dots, x^{(0)}, z)$ *is convex and nonnegative.*

Then the function $I : u \mapsto \int_\Omega L((\partial^{m+1}_{i_1,\dots,i_{m+1}} u(\mathbf{x}))_{1\leqslant i_1,\dots,i_{m+1}\leqslant d_1}, \dots, u(\mathbf{x}), \mathbf{x}) d\mathbf{x}$ *is lower-semi continuous for the weak topology on* $H^{m+1}(\Omega, \mathbb{R}^{d_2})$.

Proof. This results generalizes Evans (2010, Theorem 1, Chapter 8.2), which treats the case $m = 0$. Let $(u_p)_{p\in\mathbb{N}} \in H^{m+1}(\Omega, \mathbb{R}^{d_2})^{\mathbb{N}}$ be a sequence that weakly converges to $u_\infty \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ in

$H^{m+1}(\Omega, \mathbb{R}^{d_2})$. Our goal is to prove that $I(u_\infty) \leqslant \liminf_{p\to\infty} I(u_p)$. Upon passing to a subsequence, we can suppose that $\lim_{p\to\infty} I(u_p) = \liminf_{p\to\infty} I(u_p)$.

As a first step, we strengthen the convergence of $(u_p)_{p\in\mathbb{N}}$ by showing that for any $\varepsilon > 0$, there exists a subset E_ε of Ω such that $|\Omega\backslash E_\varepsilon| \leqslant \varepsilon$ (the notation $|\cdot|$ stands for the Lebesgue measure), and such that there exists a subsequence that uniformly converges on E_ε, as well as its derivatives. Recalling that a weakly convergent sequence is bounded (e.g., Evans, 2010, Chapter D.4), one has $\sup_{p\in\mathbb{N}} \|u_p\|_{H^{m+1}(\Omega)} < \infty$. Theorem 1.4 ensures that a subsequence of $(u_p)_{p\in\mathbb{N}}$ converges to, say, $u_\infty \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ with respect to the $H^m(\Omega)$ norm. Upon passing again to another subsequence, we conclude that for all $|\alpha| \leqslant m$ and for almost every x in Ω, $\lim_{p\to\infty} \partial^\alpha u_p(x) = \partial^\alpha u_\infty(x)$ (see, e.g. Brezis, 2010, Theorem 4.9). Finally, by Egorov's theorem (Evans, 2010, Chapter E.2), for any $\varepsilon > 0$, there exists a measurable set E_ε such that $|\Omega\backslash E_\varepsilon| \leqslant \varepsilon$ and such that, for all $|\alpha| \leqslant m$, $\lim_{p\to\infty} \|\partial^\alpha u_p - \partial^\alpha u_\infty\|_{L^\infty(E_\varepsilon)} = 0$.

Our next goal is to bound the function L. Let $F_\varepsilon = \{x \in \Omega, \sum_{|\alpha|\leqslant m+1} |\partial^\alpha u_\infty(x)| \leqslant \varepsilon^{-1}\}$ and $G_\varepsilon = E_\varepsilon \cap F_\varepsilon$. Observe that $\lim_{\varepsilon\to 0} |\Omega\backslash G_\varepsilon| = 0$. Since, for all $|\alpha| \leqslant m+1$, $\|\partial^\alpha u_\infty\|_{\infty, G_\varepsilon} < \infty$, and since $\lim_{p\to\infty} \|\partial^\alpha u_p - \partial^\alpha u_\infty\|_{L^\infty(G_\varepsilon)} = 0$, then, for all p large enough, $(\|\partial^\alpha u_p\|_{L^\infty(G_\varepsilon)})_{p\in\mathbb{N}}$ is bounded. For now, for the ease of notation, we denote $((\partial^{m+1}_{i_1,\ldots,i_{m+1}} u(z))_{1\leqslant i_1,\ldots,i_{m+1}\leqslant d_1}, \ldots, u(z), z)$ by $(D^{m+1}u(z), \ldots, u(z), z)$. Therefore, since the Lagrangian L is smooth and Ω is bounded, for all p large enough, $(\|L(D^{m+1}u_p(\cdot), \ldots, Du_p(\cdot), u_p(\cdot), \cdot)\|_{L^\infty(G_\varepsilon)})_{p\in\mathbb{N}}$ is bounded as well.

To conclude the proof, we take advantage of the convexity of the Lagrangian L. Let J_{m+1} be the Jacobian matrix of L along the vector $x^{(m+1)}$. The convexity of L implies

$$\begin{aligned} &L(D^{m+1}u_p(z), \ldots, u_p(z), z) \\ &\quad\geqslant L(D^{m+1}u_\infty(z), D^m u_p(z) \ldots, u_p(z), z) \\ &\qquad + J_{m+1}(D^{m+1}u_\infty(z), D^m u_p(z) \ldots, u_p(z), z) \times (D^{m+1}u_p(z) - D^{m+1}u_\infty(z)). \end{aligned}$$

Using the fact that $L \geqslant 0$ and that $I(u_p) \geqslant \int_{G_\varepsilon} L(D^{m+1}u_p(z), \ldots, u_p(z), z)dz$, we obtain

$$\begin{aligned} I(u_p) \geqslant \int_{G_\varepsilon} &L(D^{m+1}u_\infty(z), D^m u_p(z), \ldots, u_p(z), z) \\ &+ J_{m+1}(D^{m+1}u_\infty(z), D^m u_p(z), \ldots, u_p(z), z) \times (D^{m+1}u_p(z) - D^{m+1}u_\infty(z))dz. \end{aligned}$$

Since $(\|L(D^{m+1}u_p(\cdot), \ldots, Du_p(\cdot), u_p(\cdot), \cdot)\|_{L^\infty(G_\varepsilon)})_{p\in\mathbb{N}}$ is bounded for p large enough, and since, for all $|\alpha| \leqslant m$, $\lim_{p\to\infty} \|\partial^\alpha u_p - \partial^\alpha u_\infty\|_{L^\infty(G_\varepsilon)} = 0$, the dominated convergence theorem ensures that

$$\lim_{p\to\infty} \int_{G_\varepsilon} L(D^{m+1}u_\infty(z), D^m u_p(z), \ldots, u_p(z), z)dz = \int_{G_\varepsilon} L(D^{m+1}u_\infty(z), \ldots, u_\infty(z), z)dz.$$

Since (i) L is smooth (and therefore Lipschitz on bounded domains), (ii) for all p large enough, $(\|\partial^\alpha u_p\|_{L^\infty(G_\varepsilon)})_{p\in\mathbb{N}}$ is bounded, and (iii) for all $|\alpha| \leqslant m$, $\lim_p \|\partial^\alpha u_p - \partial^\alpha u_\infty\|_{L^\infty(G_\varepsilon)} = 0$, we have that $\lim_{p\to\infty} \|J_{m+1}(D^{m+1}u_\infty(\cdot), D^m u_p(\cdot), \ldots, u_p(\cdot), \cdot) - J_{m+1}(D^{m+1}u_\infty(\cdot), \ldots, u_\infty(\cdot), \cdot)\|_{L^\infty(G_\varepsilon)} = 0$. Therefore, since $D^{m+1}u_p \rightharpoonup D^{m+1}u_\infty$,

$$\lim_{p\to\infty} \int_{G_\varepsilon} J_{m+1}(D^{m+1}u_\infty(z), D^m u_p(z), \ldots, u_p(z), z) \times (D^{m+1}u_p(z) - D^{m+1}u_\infty(z))dz = 0.$$

Hence, $\lim_{p\to\infty} I(u_p) \geqslant \int_{G_\varepsilon} L(D^{m+1}u_\infty(z), \ldots, u_\infty(z), z)dz$. Finally, applying the monotone convergence theorem with $\varepsilon \to 0$ shows that $\lim_{p\to\infty} I(u_p) \geqslant I(u_\infty)$, which is the desired result. □

Lemma 2.12 (Measurability of $\hat{u}_n$). *Let* $\hat{u}_n = \arg\min_{u \in H^{m+1}(\Omega,\mathbb{R}^{d_2})} \mathscr{R}_n^{(\mathrm{reg})}(u)$*, where, for all* $u \in H^{m+1}(\Omega,\mathbb{R}^{d_2})$*,*

$$\mathscr{R}_n^{(\mathrm{reg})}(u) = \frac{\lambda_d}{n}\sum_{i=1}^n \|\tilde{\Pi}(u)(\mathbf{X}_i) - Y_i\|_2^2 + \lambda_e \mathbb{E}\|\tilde{\Pi}(u)(\mathbf{X}^{(e)}) - h(\mathbf{X}^{(e)})\|_2^2$$
$$+ \frac{1}{|\Omega|}\sum_{k=1}^M \|\mathscr{F}_k(u,\cdot)\|_{L^2(\Omega)} + \lambda_t \|u\|^2_{H^{m+1}(\Omega)}.$$

Then $\hat{u}_n$ *is a random variable.*

Proof. Recall that

$$\mathscr{R}_n^{(\mathrm{reg})}(u) = \mathcal{A}_n(u,u) - 2\mathcal{B}_n(u) + \frac{\lambda_d}{n}\sum_{i=1}^n \|Y_i\|^2 + \lambda_e \mathbb{E}\|h(\mathbf{X}^{(e)})\|_2^2 + \frac{1}{|\Omega|}\sum_{k=1}^M \int_\Omega B_k(\mathbf{x})^2 d\mathbf{x}.$$

Throughout we use the notation $\mathcal{A}_{(\mathbf{x},e)}(u,u)$ instead of $\mathcal{A}_n(u,u)$, to make the dependence of $\mathcal{A}_n$ in the random variables $\mathbf{x} = (\mathbf{X}_1,\ldots,\mathbf{X}_n)$ and $e = (\varepsilon_1,\ldots,\varepsilon_n)$ more explicit. We do the same with $\mathcal{B}_n$. For a given a normed space $(F,\|\cdot\|)$, we let $\mathscr{B}(F,\|\cdot\|)$ be the Borel σ-algebra on F induced by the norm $\|\cdot\|$.

Our goal is to prove that the function

$$\hat{u}_n : (\Omega^n \times \mathbb{R}^{nd_2}, \mathscr{B}(\Omega^n \times \mathbb{R}^{nd_2}, \|\cdot\|_2)) \to (H^{m+1}(\Omega,\mathbb{R}^{d_2}), \mathscr{B}(H^{m+1}(\Omega,\mathbb{R}^{d_2}), \|\cdot\|_{H^{m+1}(\Omega)}))$$
$$(\mathbf{x},e) \mapsto \underset{u \in H^{m+1}(\Omega,\mathbb{R}^{d_2})}{\arg\min} \mathcal{A}_{(\mathbf{x},e)}(u,u) - 2\mathcal{B}_{(\mathbf{x},e)}(u)$$

is measurable. Recall that $H^{m+1}(\Omega,\mathbb{R}^{d_2})$ is a Banach space separable with respect to its norm $\|\cdot\|_{H^{m+1}(\Omega)}$. Let $(v_q)_{q\in\mathbb{N}} \in H^{m+1}(\Omega,\mathbb{R}^{d_2})^{\mathbb{N}}$ be a sequence dense in $H^{m+1}(\Omega,\mathbb{R}^{d_2})$. Note that, for any $\mathbf{x} \in \Omega^n$ and any $e \in \mathbb{R}^{nd_2}$, one has $\min_{u\in H^{m+1}(\Omega,\mathbb{R}^{d_2})} \mathcal{A}_{(\mathbf{x},e)}(u,u) - 2\mathcal{B}_{(\mathbf{x},e)}(u) = \inf_{q\in\mathbb{N}} \mathcal{A}_{(\mathbf{x},e)}(v_q,v_q) - 2\mathcal{B}_{(\mathbf{x},e)}(v_q)$. This identity is a consequence of the fact that the function $u \mapsto \mathcal{A}_{(\mathbf{x},e)}(u,u) - 2\mathcal{B}_{(\mathbf{x},e)}(u)$ is continuous for the $H^{m+1}(\Omega)$ norm, as shown in the proof of Proposition 5.5). Moreover, according to this proof, each function $F_q(\mathbf{x},e) := \mathcal{A}_{(\mathbf{x},e)}(u_q,u_q) - 2\mathcal{B}_{(\mathbf{x},e)}(u_q)$ is a composition of continuous functions, and is therefore measurable. Thus, the function

$$G(\mathbf{x},e) := \min_{u\in H^{m+1}(\Omega,\mathbb{R}^{d_2})} \mathcal{A}_{(\mathbf{x},e)}(u,u) - 2\mathcal{B}_{(\mathbf{x},e)}(u) = \inf_{q\in\mathbb{N}} \mathcal{A}_{(\mathbf{x},e)}(u_q,u_q) - 2\mathcal{B}_{(\mathbf{x},e)}(u_q)$$

is measurable.

Next, since Ω, $\mathbb{R}$, and $H^{m+1}(\Omega,\mathbb{R}^{d_2})$ are separable, we know that the σ-algebras $\mathscr{B}(\Omega^n \times \mathbb{R}^{nd_2} \times H^{m+1}(\Omega,\mathbb{R}^{d_2}), \|\cdot\|_\otimes)$ and $\mathscr{B}(\Omega^n \times \mathbb{R}^{nd_2}, \|\cdot\|_2) \otimes \mathscr{B}(H^{m+1}(\Omega,\mathbb{R}^{d_2}), \|\cdot\|_{H^{m+1}(\Omega)})$ are identical, where $\|(\mathbf{x},e,u)\|_\otimes = \|(\mathbf{x},e)\|_2 + \|u\|_{H^{m+1}(\Omega)}$ (see, e.g. Rogers and Williams, 2000, Chapter II.13, E13.11c). This implies that the coordinate projections $\Pi_{\mathbf{x},e}$ and Π_u—defined for $(\mathbf{x},e) \in \Omega^n \times \mathbb{R}^{nd_2}$ and $u \in H^{m+1}(\Omega,\mathbb{R}^{d_2})$ by $\Pi_{\mathbf{x},e}(\mathbf{x},e,u) = (\mathbf{x},e)$ and $\Pi_u(\mathbf{x},e,u) = u$—are $\|\cdot\|_\otimes$ measurable. It is easy to check that, for any $(\mathbf{x},e) \in \Omega^n \times \mathbb{R}^{nd_2}$ and $u \in H^{m+1}(\Omega,\mathbb{R}^{d_2})$, if $\lim_{p\to\infty} \|(\mathbf{x}_p,e_p,u_p) - (\mathbf{x},e,u)\|_\otimes = 0$, then $\lim_{p\to\infty} \|\tilde{\Pi}(u_p) - \tilde{\Pi}(u)\|_{\infty,\Omega} = 0$ and, since $\tilde{\Pi}(u) \in C^0(\Omega,\mathbb{R}^{d_2})$, $\lim_{p\to\infty} \mathcal{A}_{\mathbf{x}_p,e_p}(u_p,u_p) - 2\mathcal{B}_{\mathbf{x}_p,e_p}(u_p) = \mathcal{A}_{\mathbf{x},e}(u,u) - 2\mathcal{B}_{\mathbf{x},e}(u)$. This proves that $I : (\Omega^n \times \mathbb{R}^{nd_2} \times H^{m+1}(\Omega,\mathbb{R}^{d_2}), \mathscr{B}(\Omega^n \times \mathbb{R}^{nd_2} \times H^{m+1}(\Omega,\mathbb{R}^{d_2}), \|\cdot\|_\otimes)) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ defined by

$$I(\mathbf{x},e,u) = \mathcal{A}_{(\mathbf{x},e)}(u,u) - 2\mathcal{B}_{(\mathbf{x},e)}(u)$$

is continuous with respect to $\|\cdot\|_\otimes$ and therefore measurable. According to the above, the function

$$\tilde{I}(\mathbf{x}, e, u) = I(\mathbf{x}, e, u) - G \circ \Pi_{x,e}(\mathbf{x}, e, u)$$

is also measurable. Observe that, by definition, $\hat{u}_n = J \circ (\mathbf{X}_1, \dots, \mathbf{X}_n, \varepsilon_1, \dots, \varepsilon_n)$, where $J(\mathbf{x}, e) = \Pi_u(\tilde{I}^{-1}(\{0\}) \cap (\{(\mathbf{x}, e)\} \times H^{m+1}(\Omega, \mathbb{R}^{d_2})))$. For any measurable set $S \in \mathscr{B}(H^{m+1}(\Omega, \mathbb{R}^{d_2}, \|\cdot\|_{H^{m+1}(\Omega)})$, $J^{-1}(S) = \Pi_{x,e}(\tilde{I}^{-1}(\{0\}) \cap (\Omega^n \times \mathbb{R}^{nd_2} \times S)) \in \mathscr{B}(\Omega^n \times \mathbb{R}^{nd_2})$. (Notice that $J^{-1}(S)$ is the collection of all pairs $(\mathbf{x}, e) \in \Omega^n \times \mathbb{R}^{nd_2}$ satisfying $\arg\min_{u \in H^{m+1}(\Omega, \mathbb{R}^{d_2})} \mathcal{A}_{(\mathbf{x},e)}(u, u) - 2\mathcal{B}_{(\mathbf{x},e)}(u) \in S$.) To see this, jut note that for any set $\tilde{S} \in \mathscr{B}(\Omega^n \times \mathbb{R}^{nd_2}, \|\cdot\|_2) \otimes \mathscr{B}(H^{m+1}(\Omega, \mathbb{R}^{d_2}), \|\cdot\|_{H^{m+1}(\Omega, \mathbb{R}^{d_2})})$, one has $\Pi_{x,e}(\tilde{S}) \in \mathscr{B}(\Omega^n \times \mathbb{R}^{nd_2}, \|\cdot\|_2)$ (see, e.g. Rogers and Williams, 2000, Lemma 11.4, Chapter II). We conclude that the function J is measurable and so is $\hat{u}_n$. □

Let $B(1, \|\cdot\|_{H^{m+1}(\Omega)}) = \{u \in H^{m+1}(\Omega, \mathbb{R}^{d_2}), \quad \|u\|_{H^{m+1}(\Omega)} \leqslant 1\}$ be the ball of radius r centered at 0. Let $N(B(1, \|\cdot\|_{H^{m+1}(\Omega)})), \|\cdot\|_{H^{m+1}(\Omega)}, r)$ be the minimum number of balls of radius r according to the norm $\|\cdot\|_{H^{m+1}(\Omega)}$ needed to cover the space $B(1, \|\cdot\|_{H^{m+1}(\Omega)})$.

Lemma 2.13 (Entropy of $H^{m+1}(\Omega, \mathbb{R}^{d_2})$). *Let $\Omega \subseteq \mathbb{R}^{d_1}$ be a Lipschitz domain. For $m \geqslant 1$, one has*

$$\log N(B(1, \|\cdot\|_{H^{m+1}(\Omega)}), \|\cdot\|_{H^{m+1}(\Omega)}, r) = \underset{r \to 0}{O}(r^{-d_1/(m+1)}).$$

Proof. According to the extension theorem (Stein, 1970, Theorem 5, Chapter VI.3.3), there exists a constant $C_\Omega > 0$, depending only on Ω, such that any $u \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ can be extended to $\tilde{u} \in H^{m+1}(\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$, with $\|\tilde{u}\|_{H^{m+1}(\mathbb{R}^{d_1})} \leqslant C_\Omega \|u\|_{H^{m+1}(\Omega)}$. Let $r > 0$ be such that $\Omega \subseteq B(r, \|\cdot\|_2)$ and let $\phi \in C^\infty(\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$ be such that

$$\phi(\mathbf{x}) = \begin{cases} 1 \text{ for } \mathbf{x} \in \Omega \\ 0 \text{ for } \mathbf{x} \in \mathbb{R}^{d_1}, |x| \geqslant r. \end{cases}$$

Then, for any $u \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$, (i) $\phi\tilde{u} \in H^{m+1}(\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$, (ii) $\phi\tilde{u}|_\Omega = u$, and (iii) there exists a constant $\tilde{C}_\Omega > 0$ such that $\|\phi\tilde{u}\|_{H^{m+1}(\mathbb{R}^{d_1})} \leqslant \tilde{C}_\Omega \|u\|_{H^{m+1}(\Omega)}$. The lemma follows from Nickl and Pötscher (2007, Corollary 4). □

Lemma 2.14 (Empirical process L^2). *Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be i.i.d. random variables, with common distribution $\mu_\mathbf{X}$ on Ω. Then there exists a constant $C_\Omega > 0$, depending only on Ω, such that*

$$\mathbb{E}\Big(\sup_{\|u\|_{H^{m+1}(\Omega)} \leqslant 1} \mathbb{E}\|\tilde{\Pi}(u)(\mathbf{X})\|_2^2 - \frac{1}{n}\sum_{i=1}^n \|\tilde{\Pi}(u)(\mathbf{X}_i)\|_2^2\Big) \leqslant \frac{d_2^{1/2} C_\Omega}{n^{1/2}},$$

and

$$\mathbb{E}\Big(\Big(\sup_{\|u\|_{H^{m+1}(\Omega)} \leqslant 1} \mathbb{E}\|\tilde{\Pi}(u)(\mathbf{X})\|_2^2 - \frac{1}{n}\sum_{i=1}^n \|\tilde{\Pi}(u)(\mathbf{X}_i)\|_2^2\Big)^2\Big) \leqslant \frac{d_2 C_\Omega}{n},$$

where $\tilde{\Pi}$ is the Sobolev embedding (see Theorem 1.1).

Proof. For any $u \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$, let

$$Z_{n,u} = \mathbb{E}\|\tilde{\Pi}(u)(\mathbf{X}_i)\|_2^2 - \frac{1}{n}\sum_{j=1}^n \|\tilde{\Pi}(u)(\mathbf{X}_i)\|_2^2 \quad \text{and} \quad Z_n = \sup_{\|u\|_{H^{m+1}(\Omega)} \leqslant 1} Z_{n,u}.$$

For any $u, v \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ such that $\|u\|_{H^{m+1}(\Omega)} \leqslant 1$ and $\|v\|_{H^{m+1}(\Omega)} \leqslant 1$, we have

$$\begin{aligned}
&\Big|\frac{1}{n}(\|\tilde{\Pi}(u)(\mathbf{X}_i)\|_2^2 - \mathbb{E}\|\tilde{\Pi}(u)(\mathbf{X}_i)\|_2^2) - \frac{1}{n}(\|\tilde{\Pi}(v)(\mathbf{X}_i)\|_2^2 - \mathbb{E}\|\tilde{\Pi}(v)(\mathbf{X}_i)\|_2^2)\Big| \\
&\quad \leqslant \frac{2}{n}(\|\tilde{\Pi}(u-v)(\mathbf{X}_i)\|_2 + \mathbb{E}\|\tilde{\Pi}(u-v)(\mathbf{X}_i)\|_2) \\
&\quad \leqslant \frac{4C_\Omega}{n}\sqrt{d_2}\|u-v\|_{H^{m+1}(\Omega)} \qquad \text{(by applying Theorem 1.1).}
\end{aligned}$$

Therefore, applying Hoeffding's, Azuma's and Dudley's theorem similarly as in the proof of Theorem 5.2 shows that

$$\mathbb{E}(Z_n) \leqslant 24 C_\Omega d_2^{1/2} n^{-1} \int_0^\infty [\log N(B(1, \|\cdot\|_{H^{m+1}(\Omega)}), \|\cdot\|_{H^{m+1}(\Omega)}, r)]^{1/2} dr.$$

Lemma 2.13 shows that there exists a constant C'_Ω, depending only on Ω, such that $\mathbb{E}(Z_n) \leqslant C'_\Omega d_2^{1/2} n^{-1/2}$. Applying McDiarmid's inequality as in the proof of Theorem 5.2 shows that $\mathrm{Var}(Z_n) \leqslant 16 C_\Omega^2 d_2 n^{-1}$. Finally, since $\mathbb{E}(Z_n^2) \leqslant \mathrm{Var}(Z_n) + \mathbb{E}(Z_n)^2$, we deduce that

$$\mathbb{E}(Z_n^2) \leqslant \frac{d_2}{n}\big((C'_\Omega)^2 + 16 C_\Omega^2\big).$$

□

Lemma 2.15 (Empirical process). *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n, \varepsilon_1, \ldots, \varepsilon_n$ be independent random variables, such that $\mathbf{X}_i$ is distributed along $\mu_{\mathbf{X}}$ and ε_i is distributed along μ_ε, such that $\mathbb{E}(\varepsilon) = 0$. Then there exists a constant $C_\Omega > 0$, depending only on Ω, such that*

$$\mathbb{E}\Big(\Big(\sup_{\|u\|_{H^{m+1}(\Omega)} \leqslant 1} \frac{1}{n}\sum_{j=1}^n \langle \tilde{\Pi}(u)(\mathbf{X}_j) - \mathbb{E}(\tilde{\Pi}(u)(\mathbf{X})), \varepsilon_j\rangle\Big)^2\Big) \leqslant \frac{d_2 \mathbb{E}\|\varepsilon\|_2^2}{n} C_\Omega,$$

where $\tilde{\Pi}$ is the Sobolev embedding.

Proof. First note, since $H^{m+1}(\Omega, \mathbb{R}^{d_2})$ is separable and since, for all $u \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$, the function $(\mathbf{x}_1, \ldots, \mathbf{x}_n, e_1, \ldots, e_n) \mapsto \frac{1}{n}\sum_{j=1}^n \langle \tilde{\Pi}(u)(\mathbf{x}_j) - \mathbb{E}(\tilde{\Pi}(u)(\mathbf{X})), e_j\rangle$ is continuous, that the quantity $Z = \sup_{\|u\|_{H^{m+1}(\Omega)} \leqslant 1} \frac{1}{n}\sum_{j=1}^n \langle \tilde{\Pi}(u)(\mathbf{X}_j) - \mathbb{E}(\tilde{\Pi}(u)(\mathbf{X})), \varepsilon_j\rangle$ is a random variable. Moreover, $|Z| \leqslant 2C_\Omega\sqrt{d_2}\sum_{j=1}^n \|\varepsilon_j\|_2/n$, where C_Ω is the constant of Theorem 1.1. Thus, $\mathbb{E}(Z^2) < \infty$.

Define, for any $u \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$,

$$Z_{n,u} = \frac{1}{n}\sum_{j=1}^n \langle \tilde{\Pi}(u)(\mathbf{X}_j) - \mathbb{E}(\tilde{\Pi}(u)(\mathbf{X})), \varepsilon_j\rangle \quad \text{and} \quad Z_n = \sup_{\|u\|_{H^{m+1}(\Omega)} \leqslant 1} Z_{n,u}.$$

For any $u, v \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$, we have

$$\begin{aligned}
&\Big|\frac{1}{n}\langle \tilde{\Pi}(u)(\mathbf{X}_i) - \mathbb{E}(\tilde{\Pi}(u)(\mathbf{X})), \varepsilon_i\rangle - \frac{1}{n}\langle \tilde{\Pi}(v)(\mathbf{X}_i) - \mathbb{E}(\tilde{\Pi}(u)(\mathbf{X})), \varepsilon_i\rangle\Big| \\
&\quad = \frac{1}{n}|\langle \tilde{\Pi}(u-v)(\mathbf{X}_i) - \mathbb{E}(\tilde{\Pi}(u-v)(\mathbf{X})), \varepsilon_i\rangle|
\end{aligned}$$

$$\leqslant \frac{2C_\Omega}{n}\sqrt{d_2}\|u - v\|_{H^{m+1}(\Omega)}\|\varepsilon_i\|_2 \qquad \text{(by applying Theorem 1.1).}$$

Using that ε is independent of $\mathbf{X}$, so that the conditional expectation of Z_n is indeed a real expectation with $\varepsilon_1, \ldots, \varepsilon_n$ fixed, we can apply Hoeffding's, Azuma's and Dudley's theorem similarly as in the proof of Theorem 5.2 to show that

$$\mathbb{E}(Z_n \mid \varepsilon_1, \ldots, \varepsilon_n) \leqslant \frac{24C_\Omega}{n}\sqrt{d_2}\Big(\sum_{i=1}^n \|\varepsilon_i\|_2^2\Big)^{1/2} \times \int_0^\infty [\log N(B(1, \|\cdot\|_{H^{m+1}(\Omega)}), \|\cdot\|_{H^{m+1}(\Omega)}, r)]^{1/2} dr.$$

Hence, according to Lemma 2.13, there exists a constant $C'_\Omega > 0$, depending only on Ω, such that $\mathbb{E}(Z_n \mid \varepsilon_1, \ldots, \varepsilon_n) \leqslant C'_\Omega n^{-1}\sqrt{d_2}\Big(\sum_{i=1}^n \|\varepsilon_i\|_2^2\Big)^{1/2}$. We deduce that

$$\mathbb{E}(Z_n) \leqslant C'_\Omega\sqrt{d_2}\frac{(\mathbb{E}\|\varepsilon\|_2^2)^{1/2}}{n^{1/2}},$$

and

$$\mathrm{Var}(\mathbb{E}(Z_n \mid \varepsilon_1, \ldots, \varepsilon_n)) \leqslant \mathbb{E}(\mathbb{E}(Z_n \mid \varepsilon_1, \ldots, \varepsilon_n)^2) \leqslant (C'_\Omega)^2 d_2 \frac{\mathbb{E}\|\varepsilon\|_2^2}{n}.$$

Applying McDiarmid's inequality as in the proof of Theorem 5.2 shows that

$$\mathrm{Var}(Z_n \mid \varepsilon_1, \ldots, \varepsilon_n) \leqslant 16C_\Omega^2 d_2 \frac{1}{n^2}\sum_{i=1}^n \|\varepsilon_i\|_2^2.$$

The law of the total variance ensures that

$$\begin{aligned}\mathrm{Var}(Z_n) &= \mathrm{Var}(\mathbb{E}(Z_n \mid \varepsilon_1, \ldots, \varepsilon_n)) + \mathbb{E}(\mathrm{Var}(Z_n \mid \varepsilon_1, \ldots, \varepsilon_n)) \\ &\leqslant \frac{d_2\mathbb{E}\|\varepsilon\|_2^2}{n}\big((C'_\Omega)^2 + 16C_\Omega^2\big).\end{aligned}$$

Since $\mathbb{E}(Z_n^2) \leqslant \mathrm{Var}(Z_n) + \mathbb{E}(Z_n)^2$, we deduce that

$$\mathbb{E}(Z_n^2) \leqslant \frac{d_2\mathbb{E}\|\varepsilon\|_2^2}{n}\big(2(C'_\Omega)^2 + 16C_\Omega^2\big).$$

□

3. Proofs of Proposition 2.3

De Ryck, Lanthaler and Mishra (2021, Theorem 5.1) ensures that NN_2 is dense in $(C^\infty([0,1]^{d_1}, \mathbb{R}), \|\cdot\|_{C^K([0,1]^{d_1})})$ for all $d_1 \geqslant 1$ and $K \in \mathbb{N}$. Note that the authors state the result for Hölder spaces $(W^{K+1,\infty}([0,1]^{d_1}), \|\cdot\|_{W^{K,\infty}(]0,1[^{d_1})})$ (see Evans, 2010, for a definition). Clearly, $C^\infty([0,1]^{d_1}) \subseteq W^{K+1,\infty}([0,1]^{d_1})$ and the norms $\|\cdot\|_{C^K}$ and $\|\cdot\|_{W^{K,\infty}}$ coincide on $C^\infty([0,1]^{d_1})$.

Our proof generalizes this result to any bounded Lipschitz domain Ω, to any number $H \geqslant 2$ of layers, and to any output dimension d_2. We stress that for any $U \subseteq \mathbb{R}^{d_1}$, the set $\mathrm{NN}_2 \subseteq C^\infty(\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$ can of course be seen as a subset of $C^\infty(U, \mathbb{R}^{d_2})$.

Generalization to any bounded Lipschitz domain Ω In this and the next paragraph, $d_2 = 1$. Our objective is to prove that NN_2 is dense in $(C^\infty(\bar{\Omega}, \mathbb{R}), \|\cdot\|_{C^K(\Omega)})$. Let $f \in C^\infty(\bar{\Omega}, \mathbb{R})$. Since Ω is bounded, there exists an affine transformation $\tau : x \mapsto A_\tau x + b_\tau$, with $A_\tau \in \mathbb{R}^\star$ and $b_\tau \in \mathbb{R}^{d_1}$, such that $\tau(\Omega) \subseteq [0,1]^d$. Set $\hat{f} = f(\tau^{-1})$. According to the extension theorem for Lipschitz domains of Stein (1970, Theorem 5 Chapter VI.3.3), the function $\hat{f}$ can be extended to a function $\tilde{f} \in W^{K,\infty}([0,1]^{d_1})$ such that $\tilde{f}|_{\tau(\Omega)} = \hat{f}|_{\tau(\Omega)}$. Fix $\epsilon > 0$. According to De Ryck, Lanthaler and Mishra (2021, Theorem 5.1), there exists $u_\theta \in \mathrm{NN}_2$ such that $\|u_\theta - \tilde{f}\|_{W^{K,\infty}([0,1]^d)} \leqslant \epsilon$. Since $\tilde{f}$ is an extension of $\hat{f}$, $\tilde{f}|_{\tau(\Omega)} \in C^\infty(\bar{\Omega})$ and one also has $\|u_\theta - \hat{f}\|_{C^K(\tau(\Omega))} \leqslant \epsilon$.

Now, let $m \in \mathbb{N}$ and let α be a multi-index such that $\sum_{i=1}^{d_1} \alpha_i = m$. Then, clearly, $\partial^\alpha(\hat{f}(\tau)) = A_\tau^m \times \partial^\alpha \hat{f}(\tau)$. Therefore, $\|u_\theta(\tau) - \hat{f}(\tau)\|_{C^K(\Omega)} \leqslant \epsilon \times \max(1, A_\tau^K)$, that is

$$\|u_\theta(\tau) - f\|_{C^K(\Omega)} \leqslant \epsilon \times \max(1, A_\tau^K).$$

But, since τ is affine, $u_\theta(\tau)$ belongs to NN_2. This is the desires result. **Generalization to any number** $H \geqslant 2$ **of layers** We show in this paragraph that NN_H is dense in $(C^\infty(\bar{\Omega}, \mathbb{R}), \|\cdot\|_{C^K(\Omega)})$ for all $H \geqslant 2$. The case $H = 2$ has been treated above and it is therefore assumed that $H \geqslant 3$.

Let $f \in C^\infty(\bar{\Omega}, \mathbb{R})$. Introduce the function v defined by

$$v(x_1, \ldots, x_{d_1}) = (\tanh^{\circ(H-2)}(x_1), \ldots, \tanh^{\circ(H-2)}(x_{d_1})),$$

where $\tanh^{\circ(H-2)}$ stands for the tanh function composed $(H-2)$ times with itself. For all $u_\theta \in \mathrm{NN}_2$, $u_\theta(v) \in \mathrm{NN}_H$ is a neural network such that the first weights matrices $(W_\ell)_{1 \leqslant \ell \leqslant H-2}$ are identity matrices and the first offsets $(b_\ell)_{1 \leqslant \ell \leqslant H-2}$ are equal to zero. Since tanh is an increasing C^∞ function, v is a C^∞ diffeomorphism. Therefore, $v(\Omega)$ is a bounded Lipschitz domain and $f(v^{-1}) \in C^\infty(v(\Omega), \mathbb{R})$. Lemma 2.2 shows that $f(v^{-1}) \in C^\infty(\bar{v}(\Omega), \mathbb{R})$, where $\bar{v}(\Omega)$ is the closure of $v(\Omega)$. According to the previous paragraph, there exists a sequence $(\theta_m)_{m \in \mathbb{N}}$ of parameters such that $u_{\theta_m} \in \mathrm{NN}_2$ and

$$\lim_{m \to \infty} \|u_{\theta_m} - f(v^{-1})\|_{C^K(v(\Omega))} = 0.$$

Thus, u_{θ_m} approximates $f(v^{-1})$, and we would like $u_{\theta_m}(v)$ to approximate f. From Lemma 2.2,

$$\|u_{\theta_m}(v) - f\|_{C^K(\Omega)} \leqslant B_K \times \|u_{\theta_m} - f \circ v^{-1}\|_{C^K(\Omega)} \times (1 + \|\tanh^{\circ H-2}\|_{C^K(\mathbb{R})})^K,$$

while Corollary 2.5 asserts that $\|\tanh^{\circ H-2}\|_{C^K(\mathbb{R})} < \infty$. Therefore, we deduce that $\lim_{m\to\infty} \|u_{\theta_m}(v) - f\|_{C^K(\Omega)} = 0$ with $u_{\theta_m}(v) \in \mathrm{NN}_H$, which proves the lemma for $H \geqslant 2$.

Generalization to all output dimension d_2 We have shown so far that for all $H \geqslant 2$, NN_H is dense in $(C^\infty(\bar{\Omega}, \mathbb{R}), \|\cdot\|_{C^K(\Omega)})$. It remains to establish that NN_H is dense in $(C^\infty(\bar{\Omega}, \mathbb{R}^{d_2}), \|\cdot\|_{C^K(\Omega)})$ for any output dimension d_2.

Let $f = (f_1, \ldots, f_{d_2}) \in C^\infty(\Omega, \mathbb{R}^{d_2})$. For all $1 \leqslant i \leqslant d_2$, let $(\theta_m^{(i)})_{m \in \mathbb{N}} \in (\mathrm{NN}_H)^{\mathbb{N}}$ be a sequence of neural networks such that $\lim_{m\to\infty} \|u_{\theta_m^{(i)}} - f_i\|_{C^K(\Omega)} = 0$. Denote by $u_{\theta_m} = (u_{\theta_m^{(1)}}, \ldots, u_{\theta_m^{(d_2)}})$ the stacking of these sequences. For all $m \in \mathbb{N}$, $u_{\theta_m} \in \mathrm{NN}_H$ and $\lim_{m\to\infty} \|u_{\theta_m} - f\|_{C^K(\Omega)} = 0$. Therefore, NN_H is dense in $(C^\infty(\bar{\Omega}, \mathbb{R}), \|\cdot\|_{C^K(\Omega)})$.

4. Proofs of Section 3

4.1. Proof of Proposition 3.1

Consider $u_{\hat{\theta}(p,n_r,D)} \in \mathrm{NN}_H(D)$, the neural network defined by

$$u_{\hat{\theta}(p,n_r,D)}(\mathbf{x}) = Y_{(1)} + \sum_{i=1}^{n-1} \frac{Y_{(i+1)} - Y_{(i)}}{2} \Big[\tanh_p^{\circ H} \Big(\mathbf{x} - \mathbf{X}_{(i)} - \frac{\delta(n,n_r)}{2} \Big) + 1 \Big],$$

where $\delta(n,n_r)$ is defined in (6) and where the observations have been reordered according to increasing values of the $\mathbf{X}_{(i)}$. According to Lemma 2.6, one has, for all $1 \leqslant i \leqslant n$, $\lim_{p\to\infty} u_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_i) = Y_i$. Moreover, for all order $K \geqslant 1$ of differentiation and all $1 \leqslant j \leqslant n_r$, $\lim_{p\to\infty} u^{(K)}_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_j^{(r)}) = 0$. Recalling that $\mathscr{F}(u,\mathbf{x}) = mu''(\mathbf{x}) + \gamma u'(\mathbf{x})$, we have $\|\mathscr{F}(u,\mathbf{x})\|_2 \leqslant m\|u''(\mathbf{x})\|_2 + \gamma\|u'(\mathbf{x})\|_2$. We therefore conclude that $\lim_{p\to\infty} R_{n,n_r}(u_{\hat{\theta}(p,n_r,D)}) = 0$, which is the first statement of the proposition.

Next, using the Cauchy-Schwarz inequality, we have that, for any function $f \in C^2(\mathbb{R})$ and any $\varepsilon > 0$,

$$2\varepsilon \int_{-\varepsilon}^{\varepsilon} (mf'' + \gamma f')^2 \geqslant \Big(\int_{-\varepsilon}^{\varepsilon} mf'' + \gamma f' \Big)^2 = \big[m(f'(\varepsilon) - f'(-\varepsilon)) + \gamma (f(\varepsilon) - f(-\varepsilon)) \big]^2.$$

Thus,

$$\begin{aligned}
&\mathscr{R}_n(u_{\hat{\theta}(p,n_r,D)}) \\
&\quad \geqslant \frac{1}{T} \int_{[0,T]} \mathscr{F}(u_{\hat{\theta}(p,n_r,D)}, \mathbf{x})^2 d\mathbf{x} \\
&\quad \geqslant \frac{1}{T} \sum_{i=1}^{n} \int_{\mathbf{X}_{(i)}+\delta(n,n_r)/2-\varepsilon}^{\mathbf{X}_{(i)}+\delta(n,n_r)/2+\varepsilon} \mathscr{F}(u_{\hat{\theta}(p,n_r,D)}, \mathbf{x})^2 d\mathbf{x} \\
&\quad \geqslant \frac{1}{T} \sum_{i=1}^{n} \frac{1}{2\varepsilon} \big[m(u'_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_{(i)} + \delta(n,n_r)/2 + \varepsilon) - u'_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_{(i)} + \delta(n,n_r)/2 - \varepsilon)) \\
&\qquad\quad + \gamma (u_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_{(i)} + \delta(n,n_r)/2 + \varepsilon) - u_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_{(i)} + \delta(n,n_r)/2 - \varepsilon)) \big]^2.
\end{aligned}$$

Observe that, as soon as $\delta(n,n_r)/4 > \varepsilon$, one has, for all $1 \leqslant i \leqslant n-1$,

$$\lim_{p\to\infty} u_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_{(i)} + \delta(n,n_r)/2 + \varepsilon) - u_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_{(i)} + \delta(n,n_r)/2 - \varepsilon) = Y_{(i+1)} - Y_{(i)},$$

and, for all $1 \leqslant i \leqslant n-1$,

$$\lim_{p\to\infty} u'_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_{(i)} + \delta(n,n_r)/2 + \varepsilon) - u'_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_{(i)} + \delta(n,n_r)/2 - \varepsilon) = 0.$$

Hence, for any $0 < \varepsilon < \delta(n,n_r)/4$,

$$\begin{aligned}
&\sum_{i=1}^{n} \frac{1}{2\varepsilon} \big[m(u'_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_{(i)} + \delta(n,n_r)/2 - \varepsilon) - u'_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_{(i)} + \delta(n,n_r)/2 - \varepsilon)) \\
&\qquad + \gamma (u_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_{(i)} + \delta(n,n_r)/2 - \varepsilon) - u_{\hat{\theta}(p,n_r,D)}(\mathbf{X}_{(i)} + \delta(n,n_r)/2 - \varepsilon)) \big]^2
\end{aligned}$$

$$\xrightarrow[p\to\infty]{} \gamma \times \frac{\sum_{i=1}^{n-1}(Y_{(i+1)}-Y_{(i)})^2}{2\varepsilon}.$$

We have just proved that, for any $0<\varepsilon<\delta(n,n_r)/4$, there exists $P\in\mathbb{N}$ such that, for all $p\geqslant P$,

$$\mathcal{R}_n(u_{\hat{\theta}(p,n_r,D)}) \geqslant \gamma \times \frac{\sum_{i=1}^{n-1}(Y_{(i+1)}-Y_{(i)})^2}{2\varepsilon T}.$$

We conclude as desired that $\lim_{p\to\infty}\mathcal{R}_n(u_{\hat{\theta}(p,n_r,D)})=\infty$, since we suppose that there exists two observations $Y_{(i)}\neq Y_{(j)}$.

4.2. Proof of Proposition 3.2

Let $u_{\hat{\theta}(p,n_e,n_r,D)}\in \mathrm{NN}_H(4)$ be the neural network defined by

$$\begin{aligned} u_{\hat{\theta}(p,n_e,n_r,D)}(x,t) = &\tanh^{\circ H}(x+0.5+pt)-\tanh^{\circ H}(x-0.5+pt)\\ &+\tanh^{\circ H}(0.5+pt)-\tanh^{\circ H}(1.5+pt).\end{aligned}$$

Clearly, for any $p\in\mathbb{N}$, $u_{\hat{\theta}(p,n_e,n_r,D)}$ satisfies the initial condition

$$u_{\hat{\theta}(p,n_e,n_r,D)}(x,0)=\tanh^{\circ H}(x+0.5)-\tanh^{\circ H}(x-0.5)+\tanh^{\circ H}(0.5)-\tanh^{\circ H}(1.5).$$

We are going to prove in the next paragraphs that the derivatives of $u_{\hat{\theta}(p,n_e,n_r,D)}$ vanish as $p\to\infty$, starting with the temporal derivative and continuing with the spatial ones. According to Lemma 2.4, for all $\varepsilon>0$ and all $x\in[-1,1]$, $\lim_{p\to\infty}\|u_{\hat{\theta}(p,n_e,n_r,D)}(x,\cdot)\|_{C^2([\varepsilon,T])}=0$. Therefore, for any $\mathbf{X}_i^{(e)}\in\{-1,1\}\times[0,T]$, $\lim_{p\to\infty}\|u_{\hat{\theta}(p,n_e,n_r,D)}(\mathbf{X}_i^{(e)})\|_2=0$ and, for any $\mathbf{X}_j^{(r)}\in\Omega$, $\lim_{p\to\infty}\|\partial_t u_{\hat{\theta}(p,n_e,n_r,D)}(\mathbf{X}_j^{(r)})\|_2=0$ (since $\mathbf{X}_j^{(r)}\notin\partial\Omega$).

Letting $v(x,t)=\tanh^{\circ H}(x+0.5+pt)-\tanh^{\circ H}(x-0.5+pt)$, it comes that $\partial^2_{x,x}u_{\hat{\theta}(p,n_e,n_r,D)}=p^{-2}\partial^2_{t,t}v$. Thus, invoking again Lemma 2.4, for all $\varepsilon>0$, and all $x\in[-1,1]$,

$$\lim_{p\to\infty}p^{-2}\|\partial^2_{t,t}v(x,\cdot)\|_{\infty,[\varepsilon,T]}=\lim_{p\to\infty}\|\partial^2_{x,x}u_{\hat{\theta}(p,n_e,n_r,D)}(x,\cdot)\|_{\infty,[\varepsilon,T]}=0.$$

Therefore, for any $\mathbf{X}_j^{(r)}\in\Omega$, one has $\lim_{p\to\infty}\|\partial^2_{x,x}u_{\hat{\theta}(p,n_e,n_r,D)}(\mathbf{X}_j^{(r)})\|_2=0$ and, in turn, one has $\lim_{p\to\infty}\|\mathscr{F}(u_{\hat{\theta}(p,n_e,n_r,D)},\mathbf{X}_j^{(r)})\|_2=0$. Thus, for all $n_e,n_r\geqslant 0$, $\lim_{p\to\infty}R_{n_e,n_r}(u_{\hat{\theta}(p,n_e,n_r,D)})=0$.

Next, observe that $\mathscr{R}(u_{\hat{\theta}(p,n_e,n_r,D)})\geqslant\int_{[-1,1]\times[0,T]}(\partial_t u_{\hat{\theta}(p,n_e,n_r,D)}-\partial^2_{x,x}u_{\hat{\theta}(p,n_e,n_r,D)})^2$. By the Cauchy-Schwarz inequality, for any $\delta>0$,

$$\begin{aligned}&\int_{[-1,1]\times[0,T]}(\partial_t u_{\hat{\theta}(p,n_e,n_r,D)}-\partial^2_{x,x}u_{\hat{\theta}(p,n_e,n_r,D)})^2\\ &\geqslant \delta^{-1}\int_{x=-1}^{1}\Big(\int_{t=0}^{\delta}\partial_t u_{\hat{\theta}(p,n_e,n_r,D)}(x,t)-\partial^2_{x,x}u_{\hat{\theta}(p,n_e,n_r,D)}(x,t)\Big)^2dx\\ &\geqslant \delta^{-1}\int_{x=-1}^{1}\Big(u_{\hat{\theta}(p,n_e,n_r,D)}(x,\delta)-u_{\hat{\theta}(p,n_e,n_r,D)}(x,0)-\int_{t=0}^{\delta}\partial^2_{x,x}u_{\hat{\theta}(p,n_e,n_r,D)}(x,t)dt\Big)^2dx.\end{aligned}$$

Invoking again Lemma 2.4, we know that $\lim_{p\to\infty} \|u_{\hat\theta(p,n_e,n_r,D)}(\cdot,\delta)\|_{[-1,1],\infty} = 0$. Moreover, for all $t > 0$ and all $-1 \leqslant x \leqslant 1$, $\lim_{p\to\infty} \partial^2_{x,x} u_{\hat\theta(p,n_e,n_r,D)}(x,t) = 0$. Besides, by Corollary 2.5, $\|\partial^2_{x,x} u_{\hat\theta(p,n_e,n_r,D)}\|_{\infty,[0,1]\times[-1,1]} \leqslant 2\|\tanh^{\circ H}\|_{C^2(\mathbb{R})} < \infty$. Thus, by the dominated convergence theorem, for any $\delta > 0$ and all p large enough,

$$\mathscr{R}(u_{\hat\theta(p,n_e,n_r,D)}) \geqslant \frac{1}{2\delta}\int_{x=-1}^{1} \big(u_{\hat\theta(p,n_e,n_r,D)}(x,0)\big)^2 dx.$$

Noticing that $u_{\hat\theta(p,n_e,n_r,D)}(x,0)$ corresponds to the initial condition, that does not depends on p, we conclude that $\lim_{p\to\infty} \mathscr{R}(u_{\hat\theta(p,n_e,n_r,D)}) = \infty$.

5. Proofs of Section 4

5.1. Proof of Proposition 4.2

Recall that each neural network $u_\theta \in \mathrm{NN}_H(D)$ is written as $u_\theta = \mathcal{A}_{H+1} \circ (\tanh\circ\mathcal{A}_H) \circ \cdots \circ (\tanh\circ\mathcal{A}_1)$, where each $\mathcal{A}_k : \mathbb{R}^{L_{k-1}} \to \mathbb{R}^{L_k}$ is an affine function of the form $\mathcal{A}_k(x) = W_k x + b_k$, with W_k a $(L_{k-1}\times L_k)$-matrix, $b_k \in \mathbb{R}^{L_k}$ a vector, $L_0 = d_1$, $L_1 = \cdots = L_H = D$, $L_{H+1} = d_2$, and $\theta = (W_1, b_1, \ldots, W_{H+1}, b_{H+1}) \in \mathbb{R}^{\sum_{i=0}^{H}(L_i+1)\times L_i}$. For each $i \in \{1,\ldots,d_1\}$, we let π_i be the projection operator on the ith coordinate, defined by $\pi_i(x_1,\ldots,x_{d_1}) = x_i$. Similarly, for a matrix $W = (W_{i,j})_{1\leqslant i\leqslant d_2, 1\leqslant j\leqslant d_1}$, we let $\pi_{i,j}(W) = W_{i,j}$ and $\|W\|_\infty = \max_{1\leqslant i\leqslant d_2, 1\leqslant j\leqslant d_1} |W_{i,j}|$. Note that $\|W_k\mathbf{x}\|_\infty \leqslant L_{k-1}\|W_k\|_\infty\|\mathbf{x}\|_\infty$. Clearly, $\max_{1\leqslant k\leqslant H+1}(\|W_k\|_\infty, \|b_k\|_\infty) \leqslant \|\theta\|_\infty \leqslant \|\theta\|_2$. Finally, we recursively define the constants $C_{K,H}$ for all $K \geqslant 0$ and all $H \geqslant 1$ by $C_{0,H} = 1$, $C_{K,1} = 2^{K-1}\times(K+2)!$, and

$$C_{K,H+1} = B_K 2^{K-1}(K+2)! \max_{\substack{i_1,\ldots,i_K\in\mathbb{N}\\ i_1+2i_2+\cdots+Ki_K=K}} \prod_{1\leqslant \ell\leqslant K} C_{\ell,H}, \tag{7}$$

where B_K is the Kth Bell number, defined in (1).

We prove the proposition by induction on H, starting with the case $H = 1$. Clearly, for $H = 1$, one has

$$\|u_\theta\|_\infty \leqslant \|W_2 \times \tanh\circ\mathcal{A}_1\|_\infty + \|b_2\|_\infty \leqslant \|W_2\|_\infty D + \|b_2\|_\infty \leqslant (D+1)\|\theta\|_2. \tag{8}$$

Next, for any multi-index $\alpha = (\alpha_1,\ldots,\alpha_{d_1})$ such that $|\alpha| \geqslant 1$,

$$\partial^\alpha u_\theta(\mathbf{x}) = W_2 \begin{pmatrix} \pi_{1,1}(W_1)^{\alpha_1}\times\cdots\times\pi_{1,d_1}(W_1)^{\alpha_{d_1}}\times\tanh^{(|\alpha|)}(\pi_1(\mathcal{A}_1(\mathbf{x}))) \\ \vdots \\ \pi_{1,d_1}(W_1)^{\alpha_1}\times\cdots\times\pi_{d_1,d_1}(W_1)^{\alpha_{d_1}}\times\tanh^{(|\alpha|)}(\pi_{d_1}(\mathcal{A}_1(\mathbf{x}))) \end{pmatrix}. \tag{9}$$

Upon noting that $|\pi_{1,d_1}(W_1)| \leqslant \|\theta\|_\infty$, we see that

$$\|\partial^\alpha u_\theta\|_\infty \leqslant D\|W_2\|_\infty\|\theta\|_2^{|\alpha|}\|\tanh^{(|\alpha|)}\|_\infty \leqslant D\|\theta\|_2^{1+|\alpha|}\|\tanh^{(|\alpha|)}\|_\infty. \tag{10}$$

Therefore, combining (8) and (10), for any $K \geqslant 1$, $\|u_\theta\|_{C^K(\mathbb{R}^{d_1})} \leqslant (D+1)\max_{k\leq K}\|\tanh^{(k)}\|_\infty(1+\|\theta\|_2)^K\|\theta\|_2$. Applying Lemma 2.3, we conclude that, for all $u \in \mathrm{NN}_1(D)$ and for all $K \geqslant 0$,

$$\|u_\theta\|_{C^K(\mathbb{R}^{d_1})} \leqslant C_{K,1}(D+1)(1+\|\theta\|_2)^K\|\theta\|_2.$$

Induction Assume that for a given $H \geqslant 1$, one has, for any neural network $u_\theta \in \mathrm{NN}_H(D)$ and any $K \geqslant 0$,

$$\|u_\theta\|_{C^K(\mathbb{R}^{d_1})} \leqslant C_{K,H}(D+1)^{1+KH}(1+\|\theta\|_2)^{KH}\|\theta\|_2. \tag{11}$$

Our objective is to show that for any $u_\theta \in \mathrm{NN}_{H+1}(D)$ and any $K \geqslant 0$,

$$\|u_\theta\|_{C^K(\mathbb{R}^{d_1})} \leqslant C_{K,H+1}(D+1)^{1+K(H+1)}(1+\|\theta\|_2)^{K(H+1)}\|\theta\|_2.$$

For such a u_θ, we have, by definition, $u_\theta = \mathcal{A}_{H+2} \circ \tanh \circ v_\theta$, where $v_\theta \in \mathrm{NN}_H(D)$ (by a slight abuse of notation, the parameter of v_θ is in fact $\theta' = (W_1, b_1, \ldots, W_{H+1}, b_{H+1})$ while $\theta = (W_1, b_1, \ldots, W_{H+2}, b_{H+2})$, so $\|\theta'\|_2 \leqslant \|\theta\|_2$ and $\|\theta'\|_\infty \leqslant \|\theta\|_\infty$). Consequently,

$$\|u_\theta\|_\infty \leqslant \|W_{H+2}\|_\infty D + \|b_{H+2}\|_\infty \leqslant (D+1)\|\theta\|_2. \tag{12}$$

In addition, for any multi-index $\alpha = (\alpha_1, \ldots, \alpha_{d_1})$ such that $|\alpha| \geqslant 1$,

$$\partial^\alpha u_\theta(\mathbf{x}) = W_{H+2}\begin{pmatrix} \partial^\alpha(\tanh \circ \pi_1 \circ v_\theta(\mathbf{x})) \\ \vdots \\ \partial^\alpha(\tanh \circ \pi_D \circ v_\theta(\mathbf{x})) \end{pmatrix}.$$

Thus, $\|\partial^\alpha u_\theta\|_\infty \leqslant D\|W_{H+2}\|_\infty \max_{j \leqslant D} \|\tanh \circ \pi_j \circ v_\theta\|_{C^K(\mathbb{R}^{d_1})}$. Invoking identity (2), one has

$$\|\tanh \circ \pi_j \circ v\|_{C^K(\mathbb{R}^{d_1})} \leqslant B_K \|\tanh\|_{C^K(\mathbb{R})} \max_{i_1+2i_2+\cdots+Ki_K=K} \prod_{1 \leqslant \ell \leqslant K} \|\pi_j \circ v_\theta\|^{i_\ell}_{C^\ell(\mathbb{R}^{d_1})}.$$

Observing that $\pi_j \circ v_\theta$ belongs to $\mathrm{NN}_H(D)$, Lemma 2.3 and inequality (11) show that

$$\|\tanh \circ \pi_j \circ v_\theta\|_{C^\ell(\mathbb{R}^{d_1})} \leqslant C_{\ell,H+1}(D+1)^{1+\ell H}(1+\|\theta\|_2)^{1+\ell H}\|\theta\|_2.$$

Therefore, $\|\partial^\alpha u_\theta\|_\infty \leqslant C_{K,H+1}(D+1)^{1+KH}(1+\|\theta\|_2)^{K(H+1)}\|\theta\|_2$, which concludes the induction.

To complete the proof, it remains to show that the exponent of $\|\theta\|_2$ is optimal. To this aim, we let $d_1 = d_2 = 1$, $D = 1$. For each $H \geqslant 1$, we consider the sequence $(\theta_m^{(H)})_{m\in\mathbb{N}}$ defined by $\theta_m^{(H)} = (W_1^{(m)}, b_1^{(m)}, \ldots, W_{H+1}^{(m)}, b_{H+1}^{(m)})$, with $W_i^m = m$ and $b_i^m = 0$. Then, for all $\theta = (W_1, b_1, \ldots, W_{H+1}, b_{H+1}) \in \Theta_{H,1}$, the associated neural network's derivatives satisfy

$$\|u_\theta^{(k)}\|_\infty = \|(\tanh^{\circ H})^{(K)}\|_\infty |W_{H+1}| \prod_{i=1}^{H} |W_i|^K.$$

Next, since $\|\theta_m^{(H)}\|_2 = m\sqrt{H+1}$, we have

$$\|u_{\theta_m^{(H)}}\|_{C^K(\mathbb{R}^{d_1})} \geqslant \left\|u^{(K)}_{\theta_m^{(H)}}\right\|_\infty \geqslant \left\|(\tanh^{\circ H})^{(K)}\right\|_\infty m^{1+HK} \geqslant \bar{C}(H,K)\|\theta_m^{(H)}\|_2^{1+HK},$$

where $\bar{C}(H,K) = (H+1)^{-(1+HK)/2}\|(\tanh^{\circ H})^{(K)}\|_\infty$. Since $\lim_{m\to\infty} \|\theta_m^{(H)}\|_2 = \infty$, we conclude that the bound of inequality (11) is tight.

5.2. Lipschitz dependence of the Hölder norm in the NN parameters

Proposition 5.1 (Lipschitz dependence of the Hölder norm in the NN parameters). *Consider the class* $\mathrm{NN}_H(D) = \{u_\theta, \theta \in \Theta_{H,D}\}$. *Let* $K \in \mathbb{N}$. *Then there exists a constant* $\tilde{C}_{K,H} > 0$, *depending only on* K *and* H, *such that, for all* $\theta, \theta' \in \Theta_{H,D}$,

$$\|u_\theta - u_{\theta'}\|_{C^K(\Omega)} \leqslant \tilde{C}_{K,H}(1 + d_1 M(\Omega))(D+1)^{H+KH^2}(1 + \|\theta\|_2)^{H+KH^2}\|\theta - \theta'\|_2,$$

where $M(\Omega) = \sup_{\mathbf{x}\in\Omega} \|\mathbf{x}\|_\infty$.

Proof. We recursively define the constants $\tilde{C}_{K,H}$ for all $K \geqslant 0$ and all $H \geqslant 1$ by $\tilde{C}_{K,1} = (K+2)2^{2K-1}(K+2)!(K+3)!$, and

$$\tilde{C}_{K,H+1} = C_{K,H+1}[1 + (K+1)B_K 2^{2K-1}(K+3)!(K+2)!\tilde{C}_{K,H}].$$

Recall that π_i is the projection operator on the ith coordinate, defined by $\pi_i(x_1, \ldots, x_{d_1}) = x_i$. Before embarking on the proof, observe that by identity (2), we have, for all $u_1, u_2 \in C^K(\Omega, \mathbb{R}^D)$, for all $1 \leqslant i \leqslant D$,

$$\begin{aligned}\partial^\alpha(\tanh\circ\pi_i \circ u_1 - \tanh\circ\pi_i \circ u_2) = \sum_{P\in\Pi(K)} &[\tanh^{(|P|)} \circ\pi_i \circ u_1] \prod_{S\in P} \partial^{\alpha(S)}(\pi_i \circ u_1) \\ &- [\tanh^{(|P|)} \circ\pi_i \circ u_2] \prod_{S\in P} \partial^{\alpha(S)}(\pi_i \circ u_2).\end{aligned}$$

In addition, for two sequences $(a_i)_{1\leqslant i\leqslant n}$ and $(b_i)_{1\leqslant i\leqslant n}$,

$$\prod_{i=1}^n a_i - \prod_{i=1}^n b_i = \sum_{i=1}^n (a_i - b_i)\Big(\prod_{j=i+1}^n a_j\Big)\Big(\prod_{j=1}^{i-1} b_j\Big) \leqslant n \max_{1\leqslant i\leqslant n}\{|a_i - b_i|\} \prod_{i=1}^n \max(|a_i|, |b_i|). \tag{13}$$

Observe that for any $1 \leqslant i \leqslant d_2$ and $P \in \Pi(K)$, the term $[\tanh^{(|P|)} \circ\pi_i \circ u_1] \prod_{S\in P} \partial^{\alpha(S)}(\pi_i \circ u_1) - [\tanh^{(|P|)} \circ\pi_i \circ u_2] \prod_{S\in P} \partial^{\alpha(S)}(\pi_i \circ u_2)$ is the difference of two products of $|P|+1$ terms to which we can apply (13). So,

$$\begin{aligned}&\Big\|[\tanh^{(|\pi|)} \circ\pi_i \circ u_1] \prod_{S\in P} \partial^{\alpha(S)}(\pi_i \circ u_1) - [\tanh^{(|\pi|)} \circ\pi_i \circ u_2] \prod_{S\in \pi} \partial^{\alpha(S)}(\pi_i \circ u_2)\Big\|_{\infty,\Omega} \\ &\quad\leqslant (|P|+1)\big(\|\tanh^{(|P|)}\|_{\mathrm{Lip}}\|u_1 - u_2\|_{\infty,\Omega} + \|u_1 - u_2\|_{C^K(\Omega)}\big) \\ &\quad\quad\times \|\tanh^{(|P|)}\|_\infty \prod_{S\in P} \max(\|\partial^{\alpha(S)}u_1\|_{\infty,\Omega}, \|\partial^{\alpha(S)}u_2\|_{\infty,\Omega}).\end{aligned} \tag{14}$$

Notice finally that $\|\tanh^{(|P|)}\|_{\mathrm{Lip}} = \|\tanh^{(|P|+1)}\|_\infty$.

With the preliminary results out of the way, we are now equipped to prove the statement of the proposition, by induction on H. Assume first that $H = 1$. We start by examining the case $K = 0$ and then generalize to all $K \geqslant 1$. Let $u_\theta = \mathcal{A}_2 \circ \tanh\circ\mathcal{A}_1$ and $u_{\theta'} = \mathcal{A}'_2 \circ \tanh\circ\mathcal{A}'_1$. Notice that

$$\|\mathcal{A}_1 - \mathcal{A}'_1\|_{\infty,\Omega} \leqslant \|b_1 - b'_1\|_\infty + d_1 M(\Omega)\|W_1 - W'_1\|_\infty \leqslant \|\theta - \theta'\|_2(1 + d_1 M(\Omega)),$$

where $M(\Omega) = \max_{\mathbf{x}\in\Omega} \|\mathbf{x}\|_\infty$. Since $\|\tanh\|_{\text{Lip}} = 1$, we deduce that $\|\tanh\circ\mathcal{A}_1 - \tanh\circ\mathcal{A}'_1\|_\infty \leqslant \|\theta - \theta'\|_2(1 + d_1 M(\Omega))$. Similarly, $\|\mathcal{A}_2 - \mathcal{A}'_2\|_{\infty,B(1,\|\cdot\|_\infty)} \leqslant \|\theta - \theta'\|_2(1 + D)$. Next,

$$\begin{aligned}\|u_\theta - u_{\theta'}\|_{\infty,\Omega} &\leqslant \|(\mathcal{A}_2 - \mathcal{A}'_2)\circ\tanh\circ\mathcal{A}_1\|_{\infty,\Omega} + \|\mathcal{A}'_2\circ\tanh\circ\mathcal{A}_1 - \mathcal{A}'_2\circ\tanh\circ\mathcal{A}'_1)\|_{\infty,\Omega}\\ &\leqslant \|\mathcal{A}_2 - \mathcal{A}'_2\|_{\infty,B(1,\|\cdot\|_\infty)} + D\|W'_2\|_\infty\|\tanh\circ\mathcal{A}_1 - \tanh\circ\mathcal{A}'_1\|_{\infty,\Omega}\\ &\leqslant \|\theta - \theta'\|_2(1 + D + D\|\theta'\|_2(1 + d_1 M(\Omega)))\\ &\leqslant \tilde{C}_{0,1}(1 + d_1 M(\Omega))(D+1)(1 + \max(\|\theta\|_2, \|\theta'\|_2))\|\theta - \theta'\|_2.\end{aligned}$$

This shows the result for $H = 1$ and $K = 0$. Assume now that $K \geqslant 1$, and let α be a multi-index such that $|\alpha| = K$. Observe that

$$\begin{aligned}\|\partial^\alpha(u_\theta - u_{\theta'})\|_{\infty,\Omega} \leqslant{}& \|(W_2 - W'_2)\partial^\alpha(\tanh\circ\mathcal{A}_1)\|_{\infty,\Omega}\\ &+ \|W'_2\partial^\alpha(\tanh\circ\mathcal{A}_1 - \tanh\circ\mathcal{A}'_1)\|_{\infty,\Omega}.\end{aligned} \tag{15}$$

By Lemma 2.3 and an argument similar to the inequality (9), we have

$$\begin{aligned}\|(W_2 - W'_2)\partial^\alpha(\tanh\circ\mathcal{A}_1)\|_{\infty,\Omega} &\leqslant (D+1)\|\theta - \theta'\|_2\|\theta\|_2^K\|\tanh\|_{C^K(\mathbb{R})}\\ &\leqslant 2^{K-1}(K+2)!(D+1)\|\theta - \theta'\|_2\|\theta\|_2^K.\end{aligned} \tag{16}$$

In order to bound the second term on the right-hand side of (15), we use inequality (14) with $u_1 = \mathcal{A}_1$ and $u_2 = \mathcal{A}'_1$. In this case, the only non-zero term on the right-hand side of (14) corresponds to the partition $\pi = \{\{1\},\{2\},\dots,\{K\}\}$. Recall that $\|\mathcal{A}_1 - \mathcal{A}'_1\|_{\infty,\Omega} \leqslant \|\theta - \theta'\|_2(1 + d_1 M(\Omega))$, and note that whenever $|\alpha| = 1$, $\|\partial^\alpha(\mathcal{A}_1 - \mathcal{A}'_1)\|_{\infty,\Omega} \leqslant \|\theta - \theta'\|_2$. Therefore, $\|\mathcal{A}_1 - \mathcal{A}'_1\|_{C^K(\Omega)} = \|\mathcal{A}_1 - \mathcal{A}'_1\|_{C^1(\Omega)} \leqslant \|\theta - \theta'\|_2(1 + d_1 M(\Omega))$. Observe that $\prod_{B\in\{\{1\},\{2\},\dots,\{K\}\}} \max(\|\partial^{\alpha(B)}\mathcal{A}_1\|_{\infty,\Omega}, \|\partial^{\alpha(B)}\mathcal{A}'_1\|_{\infty,\Omega}) \leqslant \max(\|\theta\|_2, \|\theta'\|_2)^K$. Thus, putting all the pieces together, we are led to

$$\begin{aligned}&\|\partial^\alpha(\tanh\circ\mathcal{A}_1 - \tanh\circ\mathcal{A}'_1)\|_{\infty,\Omega}\\ &\quad\leqslant (K+1)\|\tanh^{(K+1)}\|_\infty\|\theta - \theta'\|_2(1 + d_1 M(\Omega))\|\tanh^{(K)}\|_\infty\max(\|\theta\|_2, \|\theta'\|_2)^K.\end{aligned}$$

Now, by Lemma 2.3, $\|\tanh^{(K)}\|_\infty \leqslant 2^{K-1}(K+2)!$ So,

$$\begin{aligned}&\|\partial^\alpha(\tanh\circ\mathcal{A}_1 - \tanh\circ\mathcal{A}'_1)\|_{\infty,\Omega}\\ &\quad\leqslant (K+1)2^{2K-1}(K+2)!(K+3)!\|\theta - \theta'\|_2(1 + d_1 M(\Omega))\max(\|\theta\|_2, \|\theta'\|_2)^K.\end{aligned} \tag{17}$$

Combining inequalities (15), (16), and (17), we conclude that

$$\|\partial^\alpha(u_\theta - u_{\theta'})\|_{\infty,\Omega} \leqslant \tilde{C}_{K,1}(1 + d_1 M(\Omega))(D+1)(1 + \max(\|\theta\|_2, \|\theta'\|_2))^{K+1}\|\theta - \theta'\|_2,$$

so that $\|u_\theta - u_{\theta'}\|_{C^K(\Omega)} \leqslant \tilde{C}_{K,1}(1 + d_1 M(\Omega))(D+1)(1 + \max(\|\theta\|_2, \|\theta'\|_2))^{K+1}\|\theta - \theta'\|_2$.

Induction Fix $H \geqslant 1$, and assume that for all $u_\theta, u_{\theta'} \in \text{NN}_H(D)$ and all $K \geqslant 0$,

$$\begin{aligned}&\|u_\theta - u_{\theta'}\|_{C^K(\Omega)}\\ &\quad\leqslant \tilde{C}_{K,H}(1 + d_1 M(\Omega))(D+1)^{H+KH^2}(1 + \max(\|\theta\|_2, \|\theta'\|_2))^{H+KH^2}\|\theta - \theta'\|_2.\end{aligned} \tag{18}$$

Let $u_\theta, u_{\theta'} \in \mathrm{NN}_{H+1}(D)$. Observe that $u_\theta = \mathcal{A}_{H+2} \circ \tanh \circ v_\theta$ and $u_{\theta'} = \mathcal{A}'_{H+2} \circ \tanh \circ v_{\theta'}$, where $v_\theta, v_{\theta'} \in \mathrm{NN}_H(D)$. Moreover,

$$\begin{aligned}
&\|\partial^\alpha(u_\theta - u_{\theta'})\|_{\infty,\Omega} \\
&\quad \leqslant \|(W_{H+2} - W'_{H+2})\partial^\alpha(\tanh \circ v_\theta)\|_{\infty,\Omega} + \|W'_{H+2}\partial^\alpha(\tanh \circ v_\theta - \tanh \circ v_{\theta'})\|_{\infty,\Omega} \\
&\quad \leqslant D(\|\theta - \theta'\|_2 \times \|\partial^\alpha(\tanh \circ v_\theta)\|_{\infty,\Omega} + \|\theta'\|_2 \times \|\partial^\alpha(\tanh \circ v_\theta - \tanh \circ v_{\theta'})\|_{\infty,\Omega}). \qquad (19)
\end{aligned}$$

Since $\tanh \circ v_\theta \in \mathrm{NN}_{H+1}(D)$, we have, by Proposition 4.2,

$$\|\partial^\alpha(\tanh \circ v_\theta)\|_{\infty,\Omega} \leqslant C_{K,H+1}(D+1)^{1+K(H+1)}(1+\|\theta\|_2)^{K(H+1)}\|\theta\|_2. \qquad (20)$$

Moreover, using (14), Lemma 2.3, and the definition of $C_{K,H+1}$ in (7), we have

$$\begin{aligned}
&\|\partial^\alpha(\tanh \circ v_\theta - \tanh \circ v_{\theta'})\|_{\infty,\Omega} \\
&\quad \leqslant B_K(K+1)\|\tanh^{(K+1)}\|_\infty \|v_\theta - v_{\theta'}\|_{C^K(\Omega)}\|\tanh^{(K)}\|_\infty \\
&\qquad \times C_{K,H+1}(D+1)^{KH}(1+\max(\|\theta\|_2, \|\theta'\|_2))^{KH} \\
&\quad \leqslant 2^{2K-1}(K+3)!(K+2)!B_K(K+1)\|v_\theta - v_{\theta'}\|_{C^K(\Omega)} \\
&\qquad \times C_{K,H+1}(D+1)^{KH}(1+\max(\|\theta\|_2, \|\theta'\|_2))^{KH}. \qquad (21)
\end{aligned}$$

The term $\|v_\theta - v_{\theta'}\|_{C^K(\Omega)}$ in (21) can be upper bounded using the induction assumption (18). Thus, combining (19), (20), and (21), we conclude as desired that for all $u_\theta, u_{\theta'} \in \mathrm{NN}_{H+1}(D)$ and all $K \in \mathbb{N}$,

$$\begin{aligned}
\|u_\theta - u_{\theta'}\|_{C^K(\Omega)} &\leqslant \tilde{C}_{K,H+1}(1 + d_1 M(\Omega))(D+1)^{(H+1)+K(H+1)^2} \\
&\qquad \times (1+\max(\|\theta\|_2, \|\theta'\|_2))^{(H+1)+K(H+1)^2}\|\theta - \theta'\|_2.
\end{aligned}$$

□

5.3. Uniform approximation of integrals

Throughout this section, the parameters $H, D \in \mathbb{N}^\star$ are held fixed, as well as the neural architecture $\mathrm{NN}_H(D)$ parameterized by $\Theta_{H,D}$. We let d be a metric in $\Theta_{H,D}$, and denote by $B(r,d)$ the closed ball in $\Theta_{H,D}$ centered at 0 and of radius r according to the metric d, that is, $B(r,d) = \{\theta \in \Theta_{H,D},\ d(0,\theta) \leqslant r\}$.

Theorem 5.2 (Uniform approximation of integrals). *Let $\Omega \subseteq \mathbb{R}^{d_1}$ be a bounded Lipschitz domain, let $\alpha_1 > 0$, and let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a sequence of i.i.d. random variables in $\bar\Omega$, with distribution μ_X. Let $f : C^\infty(\bar\Omega, \mathbb{R}^{d_2}) \times \bar\Omega \to \mathbb{R}^{d_2}$ be an operator, and assume that the following two requirements are satisfied:*

(i) there exist $C_1 > 0$ and $\beta_1 \in [0, 1/2[$ such that, for all $n \geqslant 1$ and all $\theta, \theta' \in B(n^{\alpha_1}, \|.\|_2)$,

$$\|f(u_\theta, \cdot) - f(u_{\theta'}, \cdot)\|_{\infty,\bar\Omega} \leqslant C_1 n^{\beta_1}\|\theta - \theta'\|_2; \qquad (22)$$

(ii) there exist $C_2 > 0$ and $\beta_2 \in [0, 1/2[$ satisfying $\beta_2 > \alpha_1 + \beta_1$ such that, for all $n \geqslant 1$ and all $\theta \in B(n^{\alpha_1}, \|.\|_2)$,

$$\|f(u_\theta, \cdot)\|_{\infty,\bar\Omega} \leqslant C_2 n^{\beta_2}. \qquad (23)$$

Then, almost surely, there exists $N \in \mathbb{N}^\star$ such that, for all $n \geqslant N$,

$$\sup_{\theta \in B(n^{\alpha_1}, \|.\|_2)} \Big\| \frac{1}{n} \sum_{i=1}^{n} f(u_\theta, \mathbf{X}_i) - \int_{\bar{\Omega}} f(u_\theta, \cdot) d\mu_X \Big\|_2 \leqslant \log^2(n) n^{\beta_2 - 1/2}.$$

(Notice that the rank N is random.)

Proof. Let us start the proof by considering the case $d_2 = 1$. For a given $\theta \in B(n^{\alpha_1}, \|\cdot\|_2)$, we let

$$Z_{n,\theta} = \frac{1}{n} \sum_{i=1}^{n} f(u_\theta, \mathbf{X}_i) - \int_{\bar{\Omega}} f(u_\theta, \cdot) d\mu_X.$$

We are interested in bounding the random variable

$$Z_n = \sup_{\theta \in B(n^{\alpha_1}, \|\cdot\|_2)} |Z_{n,\theta}| = \sup_{\theta \in B(n^{\alpha_1}, \|\cdot\|_2)} Z_{n,\theta}.$$

Note that there is no need of absolute value in the rightmost term since, for any $\theta = (W_1, b_1, \ldots, W_{H+1}, b_{H+1}) \in B(n^{\alpha_1}, \|\cdot\|_2)$, it is clear that $\theta' = (W_1, b_1, \ldots, W_H, b_H, -W_{H+1}, -b_{H+1}) \in B(n^{\alpha_1}, \|\cdot\|_2)$ and $u_{\theta'} = -u_\theta$. Let $M(\Omega) = \max_{\mathbf{x} \in \bar{\Omega}} \|x\|_2$. Using inequality (22), we have, for any $\theta, \theta' \in B(n^{\alpha_1}, \|\cdot\|_2)$,

$$\Big| \frac{1}{n} \Big(f(u_\theta, \mathbf{X}_i) - \int_{\bar{\Omega}} f(u_\theta, \cdot) d\mu_X \Big) - \frac{1}{n} \Big(f(u'_\theta, \mathbf{X}_i) - \int_{\bar{\Omega}} f(u'_\theta, \cdot) d\mu_X \Big) \Big| \leqslant 2C_1 n^{\beta_1 - 1} \|\theta - \theta'\|_2.$$

According to Hoeffding's theorem (van Handel, 2016, Lemma 3.6), the random variable $n^{-1}(f(u_\theta, \mathbf{X}_i) - \int_{\bar{\Omega}} f(u_\theta, \cdot) d\mu_X) - n^{-1}(f(u'_\theta, \mathbf{X}_i) - \int_{\bar{\Omega}} f(u'_\theta, \cdot) d\mu_X)$ is subgaussian with parameter $4C_1^2 n^{2\beta_1 - 2} \|\theta - \theta'\|_2^2$. Invoking Azuma's theorem (van Handel, 2016, Lemma 3.7), we deduce that $Z_{n,\theta} - Z_{n,\theta'}$, is also subgaussian, with parameter $4C_1^2 n^{2\beta_1 - 1} \|\theta - \theta'\|_2^2$. Since $\mathbb{E}(Z_{n,\theta}) = 0$, we conclude that for all $n \geqslant 1$, $(Z_{n,\theta})_{\theta \in B(n^{\alpha_1}, \|\cdot\|_2)}$ is a subgaussian process on $B(n^{\alpha_1}, \|\cdot\|_2)$ for the metric $d(\theta, \theta') = 2C_1 n^{\beta_1 - 1/2} \|\theta - \theta'\|_2$. Moreover, since $\theta \mapsto Z_{n,\theta}$ is continuous for the topology induced by the metric d, $(Z_{n,\theta})_{\theta \in B(n^{\alpha_1}, \|\cdot\|_2)}$ is separable (van Handel, 2016, Remark 5.23). Thus, by Dudley's theorem (van Handel, 2016, Corollary 5.25)

$$\mathbb{E}(Z_n) \leqslant 12 \int_0^\infty [\log N(B(n^{\alpha_1}, \|\cdot\|_2), d, r)]^{1/2} dr,$$

where $N(B(n^{\alpha_1}, \|\cdot\|_2), d, r)$ is the minimum number of balls of radius r according to the metric d needed to cover the space $B(n^{\alpha_1}, \|\cdot\|_2)$. Clearly, $N(B(n^{\alpha_1}, \|\cdot\|_2),\ d,\ r) = N(B(n^{\alpha_1}, \|\cdot\|_2),\ \|\cdot\|_2,\ n^{1/2 - \beta_1} r/(2C_1))$. Thus,

$$\mathbb{E}(Z_n) \leqslant 24 C_1 n^{\beta_1 - 1/2} \int_0^\infty [\log N(B(n^{\alpha_1}, \|\cdot\|_2), \|\cdot\|_2, r)]^{1/2} dr$$

and, in turn,

$$\mathbb{E}(Z_n) \leqslant 24 C_1 n^{\alpha_1 + \beta_1 - 1/2} \int_0^\infty [\log N(B(1, \|\cdot\|_2),\ \|\cdot\|_2, r)]^{1/2} dr.$$

Upon noting that $N(B(1, \|\cdot\|_2),\ \|\cdot\|_2, r) = 1$ for $r \geqslant 1$, we are led to

$$\mathbb{E}(Z_n) \leqslant 24 C_1 n^{\alpha_1 + \beta_1 - 1/2} \int_0^1 [\log N(B(1, \|\cdot\|_2),\ \|\cdot\|_2, r)]^{1/2} dr.$$

Since $\Theta_{H,D} = \mathbb{R}^{(d_1+1)D+(H-1)D(D+1)+(D+1)d_2}$, according to van Handel (2016, Lemma 5.13), one has

$$\log N(B(1, \|\cdot\|_2), \ \|\cdot\|_2, r) \leqslant [(d_1+1)D + (H-1)D(D+1) + (D+1)d_2] \log(3/r).$$

Notice that $\int_0^1 \log(3/r)^{1/2} dr \leqslant 3/2$. Therefore,

$$\mathbb{E}(Z_n) \leqslant 36 C_1 [(d_1+1)D + (H-1)D(D+1) + (D+1)d_2]^{1/2} n^{\alpha_1+\beta_1-1/2}. \tag{24}$$

Next, observe that, by definition of $Z_n = Z_n(\mathbf{X}_1, \ldots, \mathbf{X}_n)$,

$$\begin{aligned}
&\sup_{\mathbf{x}_i \in \mathbb{R}^{d_1}} Z_n(\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}, \mathbf{x}_i, \mathbf{X}_{i+1}, \ldots, \mathbf{X}_n) - \inf_{\mathbf{x}_i \in \mathbb{R}^{d_1}} Z_n(\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}, \mathbf{x}_i, \mathbf{X}_{i+1}, \ldots, \mathbf{X}_n) \\
&\leqslant 2n^{-1} \sup_{\theta \in B(n^{\alpha_1}, \|\cdot\|_2)} \Big\| f(u_\theta, \mathbf{X}_i) - \int_{\bar{\Omega}} f(u_\theta, \cdot) d\mu_X \Big\|_2 \\
&\leqslant 4n^{-1} \sup_{\theta \in B(n^{\alpha_1}, \|\cdot\|_2)} \|f(u_\theta, \cdot)\|_\infty.
\end{aligned}$$

Using inequality (23), McDiarmid's inequality (van Handel, 2016, Theorem 3.11) ensures that Z_n is subgaussian with parameter $4C_2^2 n^{2\beta_2-1}$. In particular, for all $t_n \geqslant 0$, $\mathbb{P}(|Z_n - \mathbb{E}(Z_n)| \geqslant t_n) \leqslant 2\exp(-n^{1-2\beta_2} t_n^2/(8C_2^2))$, which is summable with $t_n = C_3 n^{\beta_2 - 1/2} \log^2(n)$, where C_3 is any positive constant. Thus, recalling that $\beta_2 > \alpha_1 + \beta_1$, the Borel-Cantelli lemma and (24) ensure that, almost surely, for all n large enough, $0 \leqslant Z_n \leqslant 2C_3 n^{\beta_2-1/2} \log^2(n)$. Taking $C_3 = 1/2$ yields the desired result.

The generalization to the case $d_2 \geqslant 2$ is easy. Just note, letting $f = (f_1, \ldots, f_{d_2})$, that

$$\begin{aligned}
&\sup_{\theta \in B(n^{\alpha_1}, \|\cdot\|_2)} \Big\| \frac{1}{n} \sum_{i=1}^n f(u_\theta, \mathbf{X}_i) - \int_{\bar{\Omega}} f(u_\theta, \cdot) d\mu_X \Big\|_2 \\
&\leqslant \sqrt{d_2} \max_{1 \leqslant j \leqslant d_2} \sup_{\theta \in B(n^{\alpha_1}, \|\cdot\|_2)} \Big\| \frac{1}{n} \sum_{i=1}^n f_j(u_\theta, \mathbf{X}_i) - \int_{\bar{\Omega}} f_j(u_\theta, \cdot) d\mu_X \Big\|_2.
\end{aligned}$$

Taking $C_3 = d_2^{-1/2}/2$ as above leads to the result. □

Proposition 5.3 (Condition function). *Let Ω be a bounded Lipschitz domain, let E be a closed subset of $\partial\Omega$, and let $h \in \mathrm{Lip}(E, \mathbb{R}^{d_2})$. Then the operator $\mathscr{H}(u, \mathbf{x}) = \mathbf{1}_{\mathbf{x} \in E} \|u(\mathbf{x}) - h(\mathbf{x})\|^2$ satisfies inequalities* (22) *and* (23) *with $\alpha_1 < (3+H)^{-1}/2$, $\beta_1 = (1+H)\alpha_1$, and $1/2 > \beta_2 \geqslant (3+H)\alpha_1$.*

Proof. First note, since $\mathrm{Lip}(E, \mathbb{R}^{d_2}) \subseteq C^0(E, \mathbb{R}^{d_2})$, that $\|h\|_\infty < \infty$. Observe also that for any $v, w \in \mathbb{R}^{d_2}$, $|\|v\|_2^2 - \|w\|_2^2| = |\langle v+w, v-w \rangle| \leqslant \|v+w\|_2 \|v-w\|_2 \leqslant d_2 \|v+w\|_\infty \|v-w\|_\infty$, where $\langle \cdot, \cdot \rangle$ denotes the canonical scalar product. Thus, we obtain, for all $\theta, \theta' \in B(n^{\alpha_1}, \|\cdot\|_2)$ and all $\mathbf{x} \in E$,

$$\begin{aligned}
|\mathscr{H}(u_\theta, \mathbf{x}) - \mathscr{H}(u_{\theta'}, \mathbf{x})| &\leqslant (\|u_\theta(\mathbf{x})\|_2 + \|u_{\theta'}(\mathbf{x})\|_2 + 2\|h(\mathbf{x})\|_2) \|u_\theta(\mathbf{x}) - u_{\theta'}(\mathbf{x})\|_2 \\
&\leqslant d_2 (\|u_\theta\|_{\infty,\bar{\Omega}} + \|u_{\theta'}\|_{\infty,\bar{\Omega}} + 2\|h\|_\infty) \|u_\theta - u_{\theta'}\|_{\infty,\bar{\Omega}} \\
&\leqslant d_2 (2(D+1)n^{\alpha_1} + 2\|h\|_\infty) \|u_\theta - u_{\theta'}\|_{\infty,\bar{\Omega}} \quad \text{(by inequality (12))} \\
&\leqslant 2d_2((D+1)n^{\alpha_1} + \|h\|_\infty) \tilde{C}_{0,H} (1 + d_1 M(\Omega)) \\
&\qquad \times (D+1)^H (1+n^{\alpha_1})^H \|\theta - \theta'\|_2 \quad \text{(by Proposition 5.1)} \\
&\leqslant C_1 n^{\beta_1} \|\theta - \theta'\|_2,
\end{aligned}$$

where $\beta_1 = (1+H)\alpha_1$ and $C_1 = 2^{H+1} d_2(D+1+\|h\|_\infty)\tilde{C}_{0,H}(1+d_1 M(\Omega))(D+1)^H$.

Next, using (12) once again, for all $\theta \in B(n^{\alpha_1}, \|.\|_2)$, $\|\mathscr{H}(u_\theta,\cdot)\|_{\infty,\bar{\Omega}} \leqslant d_2(\|u_\theta\|_{\infty,\bar{\Omega}} + \|h\|_\infty)^2 \leqslant d_2((D+1)n^{\alpha_1} + \|h\|_\infty)^2 \leqslant C_2 n^{2\alpha_1}$. Recall that for inequality (23), β_2 must satisfy $\alpha_1 + \beta_1 < \beta_2 < 1/2$. This is true for $\beta_2 = (3+H)\alpha_1$, which completes the proof. □

Proposition 5.4 (Polynomial operator). *Let Ω be a bounded Lipschitz domain, and let $\mathscr{F} \in \mathscr{P}_{\mathrm{op}}$. Then the operator $\mathbf{1}_{\mathbf{x}\in\Omega}\mathscr{F}(u_\theta, \mathbf{x})^2$ satisfies inequalities (22) and (23) with $\alpha_1 < [2 + H(1 + (2 + H)\deg(\mathscr{F}))]^{-1}/2$, $\beta_1 = H(1 + (2 + H)\deg(\mathscr{F}))\alpha_1$, and $1/2 > \beta_2 \geqslant [2 + H(1 + (2 + H)\deg(\mathscr{F}))]\alpha_1$.*

Proof. Let $\mathscr{F} \in \mathscr{P}_{\mathrm{op}}$. By definition, there exist a degree $s \geqslant 1$, a polynomial $P \in C^\infty(\mathbb{R}^{d_1}, \mathbb{R})[Z_{1,1}, \ldots, Z_{d_2,s}]$, and a sequence $(\alpha_{i,j})_{1\leqslant i\leqslant d_2, 1\leqslant j\leqslant s}$ of multi-indexes such that, for any $u \in C^\infty(\bar{\Omega}, \mathbb{R}^{d_2})$, $\mathscr{F}(u,\cdot) = P((\partial^{\alpha_{i,j}} u_i)_{1\leqslant i\leqslant d_2, 1\leqslant j\leqslant s})$. Namely, there exists $N(P) \in \mathbb{N}^\star$, exponents $I(i,j,k) \in \mathbb{N}$, and functions $\phi_1, \ldots, \phi_{N(P)} \in C^\infty(\bar{\Omega}, \mathbb{R})$, such that $P(Z_{1,1}, \ldots, Z_{d_2,s}) = \sum_{k=1}^{N(P)} \phi_k \times \prod_{i=1}^{d_2} \prod_{j=1}^{s} Z_{i,j}^{I(i,j,k)}$. Recall, by Definition 4.5, that $\deg(\mathscr{F}) = \max_k \sum_{i=1}^{d_2} \sum_{j=1}^{s} (1 + |\alpha_{i,j}|) I(i,j,k)$.

Now, according to Proposition 4.2, there exists a positive constant $C_{\deg(\mathscr{F}),H}$ such that

$$
\begin{aligned}
&\|\mathscr{F}(u_\theta,\cdot)^2\|_{\infty,\bar{\Omega}} \\
&\leqslant \Big[\sum_{k=1}^{N(P)} \|\phi_k\|_{\infty,\bar{\Omega}} \prod_{i=1}^{d_2}\prod_{j=1}^{s} \|\partial^{\alpha_{i,j}} u_\theta\|_{\infty,\bar{\Omega}}^{I(i,j,k)}\Big]^2 \\
&\leqslant N^2(P)\big[\max_{1\leqslant k\leqslant N(P)} \|\phi_k\|_{\infty,\bar{\Omega}}\big]^2 C^2_{\deg(\mathscr{F}),H} (D+1)^{2H\deg(\mathscr{F})} (1 + \|\theta\|_2)^{2H\deg(\mathscr{F})}.
\end{aligned}
$$

Thus, for any $\theta \in B(n^{\alpha_1}, \|\cdot\|_2)$, $\|\mathscr{F}(u_\theta,\cdot)^2\|_{\infty,\bar{\Omega}} \leqslant C_2 n^{\beta_2}$, where

$$
C_2 = 2^{2H\deg(\mathscr{F})} N^2(P)\big[\max_{1\leqslant k\leqslant N(P)} \|\phi_k\|_{\infty,\bar{\Omega}}\big]^2 C^2_{\deg(\mathscr{F}),H} (D+1)^{2H\deg(\mathscr{F})},
$$

and for any $\beta_2 \geqslant 2H\deg(\mathscr{F})\alpha_1$.

Next, observe that, any u and v, $||u|^2 - |v|^2| = |(u+v)(u-v)| \leqslant |u+v||u-v|$. Therefore,

$$
\begin{aligned}
|\mathscr{F}(u_\theta, \mathbf{x})^2 - \mathscr{F}(u_{\theta'}, \mathbf{x})^2| &\leqslant \big(|\mathscr{F}(u_\theta,\mathbf{x})| + |\mathscr{F}(u_{\theta'},\mathbf{x})|\big)|\mathscr{F}(u_\theta,\mathbf{x}) - \mathscr{F}(u_{\theta'},\mathbf{x})| \\
&\leqslant 2C_2^{1/2} n^{H\deg(\mathscr{F})\alpha_1} |\mathscr{F}(u_\theta,\mathbf{x}) - \mathscr{F}(u_{\theta'},\mathbf{x})|.
\end{aligned}
$$

Using inequality (13) (remark that the product $\prod_{i=1}^{d_2}\prod_{j=1}^{s} Z_{i,j}^{I(i,j,k)}$ has less than $\deg(\mathscr{F})$ terms different from 1), it is easy to see that

$$
\begin{aligned}
|\mathscr{F}(u_\theta,\mathbf{x}) - \mathscr{F}(u_{\theta'},\mathbf{x})| \leqslant{}& N(P)\big[\max_{1\leqslant k\leqslant N(P)} \|\phi_k\|_{\infty,\bar{\Omega}}\big] \deg(\mathscr{F}) \|u_\theta - u_{\theta'}\|_{C^{\deg(\mathscr{F})}(\Omega)} \\
&\times \max_{1\leqslant k\leqslant N(P)} \prod_{i,j} \max(\|u_\theta\|_{C^{|\alpha_{i,j}|}(\Omega)}, \|u_{\theta'}\|_{C^{|\alpha_{i,j}|}(\Omega)})^{I(i,j,k)}.
\end{aligned}
$$

From Proposition 4.2, we deduce that

$$
\begin{aligned}
&\max_{1\leqslant k\leqslant N(P)} \prod_{i,j} \max(\|u_\theta\|_{C^{|\alpha_{i,j}|}(\Omega)}, \|u_{\theta'}\|_{C^{|\alpha_{i,j}|}(\Omega)})^{I(i,j,k)} \\
&\leqslant C_{\deg(\mathscr{F}),H} (D+1)^{H\deg(\mathscr{F})} (1 + \max(\|\theta\|_2, \|\theta'\|_2))^{H\deg(\mathscr{F})}.
\end{aligned}
$$

Combining the last two inequalities with Proposition 5.1 gives that

$$\begin{aligned}
&|\mathscr{F}(u_\theta, \mathbf{x}) - \mathscr{F}(u_{\theta'}, \mathbf{x})| \\
&\quad\leqslant N(P)\big[\max_{1\leqslant k\leqslant N(P)} \|\phi_k\|_{\infty,\bar{\Omega}}\big] \deg(\mathscr{F}) \tilde{C}_{\deg(\mathscr{F}),H}(1+d_1 M(\Omega))\|\theta-\theta'\|_2 \\
&\qquad\times C_{\deg(\mathscr{F}),H}(D+1)^{H(1+(1+H)\deg(\mathscr{F}))}(1+\max(\|\theta\|_2, \|\theta'\|_2))^{H(1+(1+H)\deg(\mathscr{F}))}.
\end{aligned}$$

Hence, for all $\theta, \theta' \in B(n^{\alpha_1}, \|\cdot\|_2)$, $|\mathscr{F}(u_\theta, \mathbf{x})^2 - \mathscr{F}(u_{\theta'}, \mathbf{x})^2| \leqslant C_1 n^{\beta_1}\|\theta-\theta'\|_2$, where

$$\begin{aligned}
C_1 = 2C_2^{1/2} N(P)\big[\max_{1\leqslant k\leqslant N(P)} \|\phi_k\|_{\infty,\bar{\Omega}}\big]\big] \deg(\mathscr{F}) \tilde{C}_{\deg(\mathscr{F}),H}(1+d_1 M(\Omega)) \\
\times C_{\deg(\mathscr{F}),H}(D+1)^{H(1+(1+H)\deg(\mathscr{F}))} 2^{H(1+(1+H)\deg(\mathscr{F}))}
\end{aligned}$$

and $\beta_1 = H(1+(2+H)\deg(\mathscr{F}))\alpha_1$.

Recall that for inequality (23), β_2 must satisfy $\alpha_1 + \beta_1 < \beta_2 < 1/2$. This is true for $\beta_2 = [2 + H(1 + (2+H)\deg(\mathscr{F}))]\alpha_1$ and $\alpha_1 < [2 + H(1 + (2+H)\deg(\mathscr{F}))]^{-1}/2$. □

5.4. Proof of Theorem 4.6

Let $u_0 = 0 \in \mathrm{NN}_H(D)$ be the neural network with parameter $\theta = (0, \ldots, 0)$. Obviously, $R^{(\mathrm{ridge})}_{n,n_e,n_r}(u_0) = R_{n,n_e,n_r}(u_0)$. Also,

$$R_{n,n_e,n_r}(u_0) \leqslant \frac{\lambda_d}{n}\sum_{i=1}^{n}\|Y_i\|_2^2 + \lambda_e\|h\|_\infty + \frac{1}{n_r}\sum_{k=1}^{M}\sum_{\ell=1}^{n_r}\|\mathscr{F}_k(0, \mathbf{X}_\ell^{(r)})\|_2^2.$$

Since each $\mathscr{F}_k$ is a polynomial operator (see Definition 4.4), it takes the form

$$\mathscr{F}_k(u, \mathbf{x}) = \sum_{\ell=1}^{N(P_k)} \phi_{\ell,k} \prod_{i=1}^{d_2}\prod_{j=1}^{s_k} (\partial^{\alpha_{i,j,k}} u_i(\mathbf{x}))^{I_k(i,j,\ell)}.$$

Therefore,

$$\begin{aligned}
R_{n,n_e,n_r}(u_0) &\leqslant \frac{\lambda_d}{n}\sum_{i=1}^{n}\|Y_i\|_2^2 + \lambda_e\|h\|_\infty + \sum_{k=1}^{M}\sum_{\ell=1}^{N(P_k)}\|\phi_{\ell,k}\|_{\infty,\bar{\Omega}} \\
&:= I, \qquad (25)
\end{aligned}$$

where I does not depend on $\lambda_{(\mathrm{ridge})}$, n_e, and n_r.

Let $(\hat{\theta}^{(\mathrm{ridge})}(p, n_e, n_r, D))_{p\in\mathbb{N}}$ be any minimizing sequence of the empirical risk of the ridge PINN, i.e., $\lim_{p\to\infty} R^{(\mathrm{ridge})}_{n,n_e,n_r}(u_{\hat{\theta}^{(\mathrm{ridge})}(p,n_e,n_r,D)}) = \inf_{\theta\in\Theta_{H,D}} R^{(\mathrm{ridge})}_{n,n_e,n_r}(u_\theta)$. In the rest of the proof, we let $n_{r,e} = \min(n_r, n_e)$. We will make use of the following three sets: $\mathcal{E}_1(n_{r,e}) = \{\theta \in \Theta_{H,D},\ \|\theta\|_2 \geq n_{r,e}^{\kappa}\}$, $\mathcal{E}_2(n_{r,e}) = \{\theta \in \Theta_{H,D},\ n_{r,e}^{\kappa/4} \leq \|\theta\|_2 \leq n_{r,e}^{\kappa}\}$, and $\mathcal{E}_3(n_{r,e}) = \{\theta \in \Theta_{H,D},\ \|\theta\|_2 \leq n_{r,e}^{\kappa/4}\}$. Clearly, $\Theta_{H,D} = \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3$. The proof relies on the argument that almost surely, given any n_r and n_e, for all p large enough, $\hat{\theta}^{(\mathrm{ridge})}(p, n_e, n_r, D) \in \mathcal{E}_2 \cup \mathcal{E}_3$. Moreover, on $\mathcal{E}_2 \cup \mathcal{E}_3$, the empirical risk function $R^{(\mathrm{ridge})}_{n,n_e,n_r}$ is close to the theoretical risk $\mathscr{R}_n$, when $n_{r,e}$ is large enough. For clarity, the proof is divided into four steps.

Step 1 We start by observing that, for any $\theta \in \mathcal{E}_1(n_{r,e})$, $R^{(\text{ridge})}_{n,n_e,n_r}(\theta) \geqslant \lambda_{(\text{ridge})}\|\theta\|_2^2 \geqslant n^{\kappa}_{r,e}$. Therefore, according to (25), once $n_{r,e} \geq (I+1)^{1/\kappa}$,

$$\inf_{\theta\in\mathcal{E}_3(n_{r,e})} R^{(\text{ridge})}_{n,n_e,n_r}(u_\theta) + 1 \leqslant R^{(\text{ridge})}_{n,n_e,n_r}(u_0) + 1 \leqslant \inf_{\theta\in\mathcal{E}_1(n_{r,e})} R^{(\text{ridge})}_{n,n_e,n_r}(u_\theta).$$

This shows that, for all $n_{r,e}$ large enough and for all p large enough, $\hat{\theta}^{(\text{ridge})}(p, n_e, n_r, D) \notin \mathcal{E}_1(n_{r,e})$.
Step 2 Applying Proposition 5.3 and Proposition 5.4 with $\alpha_1 = \kappa$ and $\beta_2 = (2 + H(1 + (2 + H)\max_k \deg(\mathscr{F}_k)))\alpha_1$, and then Theorem 5.2, we know that, almost surely, there exists $N \in \mathbb{N}^\star$ such that, for all $n_{r,e} \geqslant N$,

$$\begin{aligned}\sup_{\theta\in\mathcal{E}_2(n_{r,e})\cup\mathcal{E}_3(n_{r,e})} &\Big|\frac{1}{n_e}\sum_{j=1}^{n_e} \|u_\theta(\mathbf{X}^{(e)}_j) - h(\mathbf{X}^{(e)}_j)\|_2^2 - \mathbb{E}\|u_\theta(\mathbf{X}^{(e)}) - h(\mathbf{X}^{(e)})\|_2^2\Big| \\ &\leqslant \log^2(n_{r,e}) n_{r,e}^{\beta_2 - 1/2} \qquad (26)\end{aligned}$$

and, for each $1 \leqslant k \leqslant M$,

$$\sup_{\theta\in\mathcal{E}_2(n_{r,e})\cup\mathcal{E}_3(n_{r,e})} \Big|\frac{1}{n_r}\sum_{\ell=1}^{n_r} \mathscr{F}_k(u_\theta, \mathbf{X}^{(r)}_\ell)^2 - \frac{1}{|\Omega|}\int_\Omega \mathscr{F}_k(u_\theta, \mathbf{x})^2 d\mathbf{x}\Big| \leqslant \log^2(n_{r,e}) n_{r,e}^{\beta_2 - 1/2}. \qquad (27)$$

Thus, almost surely, for all $n_{r,e}$ large enough and for all $\theta \in \mathcal{E}_2(n_{r,e})$,

$$R^{(\text{ridge})}_{n,n_e,n_r}(u_\theta) \geqslant \mathscr{R}_n(u_\theta) + \lambda_{(\text{ridge})}\|\theta\|_2^2 - (M+1)\log^2(n_{r,e}) n_{r,e}^{\beta_2 - 1/2}.$$

But, for all $\theta \in \mathcal{E}_2(n_{r,e})$, $\lambda_{(\text{ridge})}\|\theta\|_2^2 \geqslant n_{e,r}^{-\kappa/2}$. Upon noting that $-\kappa/2 > \beta_2 - 1/2$, we conclude that, almost surely, for all $n_{r,e}$ large enough and for all $\theta \in \mathcal{E}_2(n_{r,e})$, $R^{(\text{ridge})}_{n,n_e,n_r}(u_\theta) \geqslant \mathscr{R}_n(u_\theta)$.
Step 3 Clearly, for all $\theta \in \mathcal{E}_3(n_{r,e})$, $\lambda_{(\text{ridge})}\|\theta\|_2^2 \leqslant n_{e,r}^{-\kappa/2}$. Using inequalities (26) and (27), we deduce that, almost surely, for all $n_{r,e}$ large enough and for all $\theta \in \mathcal{E}_3(n_{r,e})$, $|R^{(\text{ridge})}_{n,n_e,n_r}(u_\theta) - \mathscr{R}_n(u_\theta)| \leqslant (M+2)\log^2(n_{r,e}) n_{r,e}^{-\kappa/2}$.
Step 4 Fix $\varepsilon > 0$. Let $(\theta_p)_{p\in\mathbb{N}}$ be any minimizing sequence of the theoretical risk function $\mathscr{R}_n$, that is, $\lim_{p\to\infty}\mathscr{R}_n(u_{\theta_p}) = \inf_{\theta\in\Theta_{H,D}}\mathscr{R}_n(u_\theta)$. Thus, by definition, there exists some $P_\varepsilon \in \mathbb{N}$ such that $|\mathscr{R}_n(u_{\theta_{P_\varepsilon}}) - \inf_{\theta\in\Theta_{H,D}}\mathscr{R}_n(u_\theta)| \leqslant \varepsilon$.

For fixed $n_{r,e}$, according to Step 1, we have, for all p large enough, $\hat{\theta}^{(\text{ridge})}(p, n_e, n_r, D) \in \mathcal{E}_2(n_{r,e}) \cup \mathcal{E}_3(n_{r,e})$. So, according to Step 2 and Step 3,

$$\mathscr{R}_n(u_{\hat{\theta}^{(\text{ridge})}(p,n_e,n_r,D)}) \leqslant R^{(\text{ridge})}_{n,n_e,n_r}(u_{\hat{\theta}^{(\text{ridge})}(p,n_e,n_r,D)}) + (M+2)\log^2(n_{r,e}) n_{r,e}^{-\kappa/2}.$$

Now, by definition of the minimizing sequence $(\hat{\theta}^{(\text{ridge})}(p, n_e, n_r, D))_{p\in\mathbb{N}}$, for all p large enough, $R^{(\text{ridge})}_{n,n_e,n_r}(u_{\hat{\theta}^{(\text{ridge})}(p,n_e,n_r,D)}) \leqslant \inf_{\theta\in\Theta_{H,D}} R^{(\text{ridge})}_{n,n_e,n_r}(u_\theta) + \varepsilon$. Also, according to Step 3,

$$\begin{aligned}\inf_{\theta\in\mathcal{E}_2(n_{r,e})\cup\mathcal{E}_3(n_{r,e})} R^{(\text{ridge})}_{n,n_e,n_r}(u_\theta) &\leqslant \inf_{\theta\in\mathcal{E}_3(n_{r,e})} R^{(\text{ridge})}_{n,n_e,n_r}(u_\theta) \\ &\leqslant \inf_{\theta\in\mathcal{E}_3(n_{r,e})} \mathscr{R}_n(u_\theta) + (M+2)\log^2(n_{r,e}) n_{r,e}^{-\kappa/2}.\end{aligned}$$

Observe that, for all $n_{r,e}$ large enough, $\theta_{P_\varepsilon} \in \mathcal{E}_3(n_{r,e})$. Therefore, $\inf_{\theta\in\mathcal{E}_3(n_{r,e})} \mathscr{R}_n(u_\theta) \leqslant \mathscr{R}_n(u_{\theta_{P_\varepsilon}})$. Combining the previous inequalities, we conclude that, almost surely, for all $n_{r,e}$ large enough and for all p large enough,

$$\mathscr{R}_n(u_{\hat{\theta}^{(\mathrm{ridge})}(p,n_e,n_r,D)}) \leqslant \inf_{\theta\in\Theta_{H,D}} \mathscr{R}_n(u_\theta) + 3\varepsilon.$$

Since ε is arbitrary, almost surely, $\lim_{n_e,n_r\to\infty} \lim_{p\to\infty} \mathscr{R}_n(u_{\hat{\theta}^{(\mathrm{ridge})}(p,n_e,n_r,D)}) = \inf_{\theta\in\Theta_{H,D}} \mathscr{R}_n(u_\theta)$.

5.5. Proof of Theorem 4.7

The result is a direct consequence of Theorem 4.6, Proposition 2.3 and of the continuity of $\mathscr{R}_n$ with respect to the $C^K(\Omega)$ norm.

6. Proofs of Section 5

6.1. Proof of Proposition 5.5

Since the functions in $H^{m+1}(\Omega, \mathbb{R}^{d_2})$ are only defined almost everywhere, we first have to give a meaning to the pointwise evaluations $u(\mathbf{X}_i)$ when $u \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$. Since Ω is a bounded Lipschitz domain and $(m+1) > d_1/2$, we can use the Sobolev embedding of Theorem 1.1. Clearly, $\tilde{\Pi}$ is linear and $\|\tilde{\Pi}(u)\|_\infty \leqslant C_\Omega \|u\|_{H^{m+1}(\Omega)}$. The natural choice to evaluate $u \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ at the point $\mathbf{X}_i$ is therefore to evaluate its unique continuous modification $\tilde{\Pi}(u)$ at $\mathbf{X}_i$.

By assumption, $\mathscr{F}_k(u,\cdot) = \mathscr{F}_k^{(\mathrm{lin})}(u,\cdot) + B_k$, where $\mathscr{F}_k^{(\mathrm{lin})}(u,\cdot) = \sum_{|\alpha|\leqslant K} \langle A_{k,\alpha}, \partial^\alpha u\rangle$ and $A_{k,\alpha} \in C^\infty(\bar{\Omega}, \mathbb{R}^{d_1})$. Next, consider the symmetric bilinear form, defined for all $u, v \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ by

$$\begin{aligned}\mathcal{A}_n(u,v) = \frac{\lambda_d}{n}\sum_{i=1}^n \langle \tilde{\Pi}(u)(\mathbf{X}_i), \tilde{\Pi}(v)(\mathbf{X}_i)\rangle + \lambda_e \mathbb{E}\langle \tilde{\Pi}(u)(\mathbf{X}^{(e)}), \tilde{\Pi}(v)(\mathbf{X}^{(e)})\rangle \\ + \frac{1}{|\Omega|}\sum_{k=1}^M \int_\Omega \mathscr{F}_k^{(\mathrm{lin})}(u,\mathbf{x})\mathscr{F}_k^{(\mathrm{lin})}(v,\mathbf{x})d\mathbf{x} + \frac{\lambda_t}{|\Omega|}\sum_{|\alpha|\leqslant m+1}\int_\Omega \langle \partial^\alpha u(\mathbf{x}), \partial^\alpha v(\mathbf{x})\rangle d\mathbf{x},\end{aligned}$$

along with the linear form defined for all $u \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ by

$$\begin{aligned}\mathcal{B}_n(u) = \frac{\lambda_d}{n}\sum_{i=1}^n \langle Y_i, \tilde{\Pi}(u)(\mathbf{X}_i)\rangle + \lambda_e \mathbb{E}\langle \tilde{\Pi}(u)(\mathbf{X}^{(e)}), h(\mathbf{X}^{(e)})\rangle \\ - \frac{1}{|\Omega|}\sum_{k=1}^M \int_\Omega B_k(\mathbf{x})\mathscr{F}_k^{(\mathrm{lin})}(v,\mathbf{x})d\mathbf{x}.\end{aligned}$$

Observe that

$$\mathcal{A}_n(u,u) - 2\mathcal{B}_n(u) = \mathscr{R}_n^{(\mathrm{reg})}(u) - \frac{\lambda_d}{n}\sum_{i=1}^n \|Y_i\|_2^2 - \lambda_e \mathbb{E}\|h(\mathbf{X}^{(e)})\|_2^2 - \frac{1}{|\Omega|}\sum_{k=1}^M \int_\Omega B_k(\mathbf{x})^2 d\mathbf{x}.$$

In addition, $\mathcal{A}_n(u,u) \geqslant \lambda_t \|u\|^2_{H^{m+1}(\Omega)}$, where $\lambda_t > 0$, so that $\mathcal{A}_n$ is coercive on the normed space $(H^{m+1}(\Omega), \|\cdot\|_{H^{m+1}(\Omega)})$. Since $(m+1) > \max(d_1/2, K)$, one has that

$$|\mathcal{A}_n(u,v)| \leqslant ((\lambda_d + \lambda_e)C_\Omega^2 + \sum_{1\leqslant k\leqslant M} (\sum_{|\alpha|\leqslant K} \|A_{k,\alpha}\|_{\infty,\Omega})^2 + \lambda_t)\|u\|_{H^{m+1}(\Omega)}\|v\|_{H^{m+1}(\Omega)},$$

and

$$|\mathcal{B}_n(u)| \leqslant C_\Omega\Big(\frac{\lambda_d}{n}\sum_{i=1}^n \|Y_i\|_2 + \lambda_e\|h\|_\infty + \sum_{k=1}^M (\|B_k\|_{\infty,\Omega} \sum_{|\alpha|\leqslant K} \|A_{k,\alpha}\|_{\infty,\Omega})\Big)\|u\|_{H^{m+1}(\Omega)}.$$

This shows that the operators $\mathcal{A}_n$ and $\mathcal{B}_n$ are continuous. Therefore, by the Lax-Milgram theorem (e.g., Brezis, 2010, Corollary 5.8), there exists a unique $\hat{u} \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ such that $\mathcal{A}_n(\hat{u},\hat{u}) - 2\mathcal{B}_n(\hat{u}) = \min_{u\in H^{m+1}(\Omega,\mathbb{R}^{d_2})} \mathcal{A}_n(u,u) - 2\mathcal{B}_n(u)$. This directly implies that $\hat{u}$ is the unique minimizer of $\mathscr{R}_n^{(\mathrm{reg})}$ over $H^{m+1}(\Omega, \mathbb{R}^{d_2})$. Furthermore, the Lax-Milgram theorem also states that $\hat{u}$ is the unique element of $H^{m+1}(\Omega,\mathbb{R}^{d_2})$ such that, for all $v \in H^{m+1}(\Omega,\mathbb{R}^{d_2})$, $\mathcal{A}_n(\hat{u}, v) = \mathcal{B}_n(v)$. This concludes the proof of the proposition.

6.2. Proof of Proposition 5.6

Let $\hat{u}_n$ be the unique minimizer of the regularized theoretical risk $\mathscr{R}_n^{(\mathrm{reg})}$ over $H^{m+1}(\Omega,\mathbb{R}^{d_2})$ given by Proposition 5.5. Notice that

$$\inf_{u\in C^\infty(\bar{\Omega},\mathbb{R}^{d_2})} \mathscr{R}_n^{(\mathrm{reg})}(u) = \inf_{u\in H^{m+1}(\Omega,\mathbb{R}^{d_2})} \mathscr{R}_n^{(\mathrm{reg})}(u) = \mathscr{R}_n(\hat{u}_n).$$

The first equality is a consequence of the density of $C^\infty(\bar{\Omega},\mathbb{R}^{d_2})$ in $H^{m+1}(\Omega,\mathbb{R}^{d_2})$, together with the continuity of the function $\mathscr{R}_n^{(\mathrm{reg})} : H^{m+1}(\Omega,\mathbb{R}^{d_2}) \to \mathbb{R}$ with respect to the $H^{m+1}(\Omega)$ norm (see the proof of Proposition 5.5). The density argument follows from the extension theorem of Stein (1970, Chapter VI.3.3, Theorem 5) and from Evans (2010, Chapter 5.3, Theorem 3).

Our goal is to show that the regularized theoretical risk satisfies some form of continuity, so that we can connect $\mathscr{R}^{(\mathrm{reg})}(u_p)$ and $\mathscr{R}^{(\mathrm{reg})}(\hat{u}_n)$. Recall that, by assumption, $\mathscr{F}_k(u,\cdot) = \mathscr{F}_k^{(\mathrm{lin})}(u,\cdot) + B_k$, where $\mathscr{F}_k^{(\mathrm{lin})}(u,\cdot) = \sum_{|\alpha|\leqslant K}\langle A_{k,\alpha}(\cdot), \partial^\alpha u(\cdot)\rangle$ and $A_{k,\alpha} \in C^\infty(\bar{\Omega},\mathbb{R}^{d_1})$. Observe that

$$\mathscr{R}_n^{(\mathrm{reg})}(u) = F(u) + \frac{1}{|\Omega|}I(u), \tag{28}$$

where

$$F(u) = \frac{\lambda_d}{n}\sum_{i=1}^n \|\tilde{\Pi}(u)(\mathbf{X}_i) - Y_i\|_2^2 + \lambda_e\mathbb{E}\|\tilde{\Pi}(u)(\mathbf{X}^{(e)}) - h(\mathbf{X}^{(e)})\|_2^2,$$

$$I(u) = \int_\Omega L((\partial^{m+1}_{i_1,\ldots,i_{m+1}} u(\mathbf{x}))_{1\leqslant i_1,\ldots,i_{m+1}\leqslant d_1}, \ldots, u(\mathbf{x}), \mathbf{x})d\mathbf{x},$$

and where the function L satisfies

$$L(x^{(m+1)}, \ldots, x^{(0)}, z) = \sum_{k=1}^M \Big(B_k(z) + \sum_{|\alpha|\leqslant K}\langle A_{k,\alpha}(z), x_\alpha^{(|\alpha|)}\rangle\Big)^2 + \lambda_t \sum_{j=0}^{m+1}\|x^{(j)}\|_2^2.$$

(The term $x^{(j)} \in \mathbb{R}^{\binom{d_1+j-1}{j-1}d_2}$ corresponds to the to the concatenation of all the partial derivatives of order j, i.e., to the term $(\partial^j_{i_1,\dots,i_j} u(\mathbf{x}))_{1\leqslant i_1,\dots,i_j\leqslant d_1}$.) Clearly, $L \geqslant 0$ and, since $(m+1) > K$, the Lagrangian L is convex in $x^{(m+1)}$. Therefore, according to Lemma 2.11, the function I is weakly lower-semi continuous on $H^{m+1}(\Omega, \mathbb{R}^{d_2})$.

Now, let us proceed by contradiction and assume that there is a sequence $(u_p)_{p\in\mathbb{N}}$ of functions such that (i) $u_p \in C^\infty(\bar{\Omega}, \mathbb{R}^{d_2})$, (ii) $\lim_{p\to\infty} \mathscr{R}_n^{(\mathrm{reg})}(u_p) = \mathscr{R}_n^{(\mathrm{reg})}(\hat{u}_n)$, and (iii) $(u_p)_{p\in\mathbb{N}}$ does not converge to $\hat{u}_n$ with respect to the $H^m(\Omega)$ norm. Therefore, upon passing to a subsequence, there exists $\varepsilon > 0$ such that, for all $p \geqslant 0$, $\|u_p - \hat{u}_n\|_{H^m(\Omega)} \geqslant \varepsilon$.

Since $\mathscr{R}_n^{(\mathrm{reg})}(u_p) \geqslant \lambda_t \|u_p\|_{H^{m+1}(\Omega)}$, $\lambda_t > 0$, and $(u_p)_{p\in\mathbb{N}}$ is a minimizing sequence, $(u_p)_{p\in\mathbb{N}}$ is bounded in $H^{m+1}(\Omega, \mathbb{R}^{d_2})$. Therefore, Theorem 1.4 states that passing to a subsequence, $(u_p)_{p\in\mathbb{N}}$ converges to a limit, say u_∞, both weakly in $H^{m+1}(\Omega, \mathbb{R}^{d_2})$ and with respect to the $H^m(\Omega)$ norm. Then, since I is weakly lower-semi continuous on $H^{m+1}(\Omega, \mathbb{R}^{d_2})$, we deduce that

$$\lim_{p\to\infty} I(u_p) \geqslant I(u_\infty). \tag{29}$$

Recalling the definition of $\tilde{\Pi}$ in Theorem 1.1, we know that there exists a constant $C_\Omega > 0$ such that $\|u_p - \tilde{\Pi}(u_\infty)\|_{\infty,\Omega} = \|\tilde{\Pi}(u_p - u_\infty)\|_{\infty,\Omega} \leqslant C_\Omega \|u_p - u_\infty\|_{H^m(\Omega)}$. We deduce that $\lim_{p\to\infty} F(u_p) = F(u_\infty)$. Therefore, combining this result with (28) and (29), we deduce that $\lim_{p\to\infty} \mathscr{R}_n^{(\mathrm{reg})}(u_p) \geqslant \mathscr{R}_n^{(\mathrm{reg})}(u_\infty)$. However, recalling that $\lim_{p\to\infty} \mathscr{R}_n^{(\mathrm{reg})}(u_p) = \mathscr{R}_n^{(\mathrm{reg})}(\hat{u}_n)$ and that $\hat{u}_n$ is the unique minimizer of $\mathscr{R}_n^{(\mathrm{reg})}$ over $H^{m+1}(\Omega, \mathbb{R}^{d_2})$, we conclude that $u_\infty = \hat{u}_n$.

We just proved that there exists a subsequence of $(u_p)_{p\in\mathbb{N}}$ which converges to $\hat{u}_n$ with respect to the $H^m(\Omega)$ norm. This contradicts the assumption $\|u_p - \hat{u}_n\|_{H^m(\Omega)} \geqslant \varepsilon$ for all $p \geqslant 0$.

6.3. Proof of Theorem 5.7

The result is an immediate consequence of Theorem 4.7, Propositions 5.5, and Proposition 5.6.

6.4. Proof of Theorem 5.8

Throughout the proof, since no data are involved, we denote the regularized theoretical risk by $\mathscr{R}^{(\mathrm{reg})}$ instead of $\mathscr{R}_n^{(\mathrm{reg})}$. Also, to make the dependence in the hyperparameter λ_t transparent, we denote by $u(\lambda_t)$ the unique minimizer of $\mathscr{R}^{(\mathrm{reg})}$ instead of $\hat{u}_n$.

We proceed by contradiction and assume that $\lim_{\lambda_t\to 0} \|u(\lambda_t) - u^\star\|_{H^m(\Omega)} \neq 0$. If this is true, then, upon passing to a subsequence $(\lambda_{t,p})_{p\in\mathbb{N}}$ such that $\lim_{p\to\infty} \lambda_{t,p} = 0$, there exists $\varepsilon > 0$ such that, for all $p \geqslant 0$, $\|u(\lambda_{t,p}) - u^\star\|_{H^m(\Omega)} \geqslant \varepsilon$.

Notice that $\|u(\lambda_{t,p})\|_{H^{m+1}(\Omega)} \leqslant \mathscr{R}^{(\mathrm{reg})}(u^\star)/\lambda_{t,p} = \|u^\star\|_{H^{m+1}(\Omega)}$. Theorem 1.4 proves that upon passing to a subsequence, $(u(\lambda_{t,p}))_{p\in\mathbb{N}}$ converges with respect to the $H^m(\Omega)$ norm to a function $u_\infty \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$. Since $m \geqslant K$, the theoretical risk $\mathscr{R}$ is continuous with respect to the $H^m(\Omega)$ norm and we have that $\mathscr{R}(u_\infty) = \lim_{p\to\infty} \mathscr{R}(u(\lambda_{t,p}))$. Moreover, by definition of $u(\lambda_{t,p})$ and since $\mathscr{R}(u^\star) = 0$, we have that $\mathscr{R}(u(\lambda_{t,p})) + \lambda_{t,p}\|u(\lambda_{t,p})\|_{H^{m+1}(\Omega)} \leqslant \lambda_{t,p}\|u^\star\|_{H^{m+1}(\Omega)}$. Therefore, $\mathscr{R}(u_\infty) = 0$ and $u_\infty = u^\star$. This contradicts the assumption that for all $p \geqslant 0$, $\|u(\lambda_{t,p}) - u^\star\|_{H^m(\Omega)} \geqslant \varepsilon$.

6.5. Proof of Proposition 5.11

We prove the proposition in several steps. In the sequel, given a measure μ on Ω and a function $u \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$, we let $\|u\|_{L^2(\mu)}^2 = \int_\Omega \|\tilde{\Pi}(u)(\mathbf{x})\|_2^2 d\mu(\mathbf{x})$, where, as usual, $\tilde{\Pi}(u)$ is the unique continuous function such that $\tilde{\Pi}(u) = u$ almost everywhere.

Step 1: Decomposing the problem into two simpler ones Following the framework of Arnone et al. (2022), the core idea is to decompose the problem into two simpler ones thanks to the linearity in $\hat{u}_n$ and in Y_i of the identity

$$\forall v \in H^{m+1}(\Omega, \mathbb{R}^{d_2}), \quad \mathscr{A}_n(\hat{u}_n, v) = \mathscr{B}_n(v)$$

of Proposition 5.5. Thus, recalling that $Y_i = u^\star(\mathbf{X}_i) + \varepsilon_i$, we let

$$\mathscr{B}_n^\star(v) = \frac{\lambda_d}{n}\sum_{i=1}^n \langle u^\star(\mathbf{X}_i), \tilde{\Pi}(v)(\mathbf{X}_i)\rangle + \lambda_e \mathbb{E}\langle \tilde{\Pi}(v)(\mathbf{X}^{(e)}), h(\mathbf{X}^{(e)})\rangle$$
$$- \frac{1}{|\Omega|}\sum_{k=1}^M \int_\Omega B_k(\mathbf{x}) \mathscr{F}_k^{(\mathrm{lin})}(v, \mathbf{x}) d\mathbf{x}$$

and

$$\mathscr{B}_n^{(\mathrm{noise})}(v) = \frac{\lambda_d}{n}\sum_{i=1}^n \langle \varepsilon_i, \tilde{\Pi}(v)(\mathbf{X}_i)\rangle.$$

Clearly, $\mathscr{B}_n = \mathscr{B}_n^\star + \mathscr{B}_n^{(\mathrm{noise})}$. Using Proposition 5.5 with Y_i instead of ε_i, and setting $\lambda_e = 0$, we see that there exists a unique $\hat{u}_n^{(\mathrm{noise})} \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ such that, for all $v \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$, $\mathscr{A}_n(\hat{u}_n^{(\mathrm{noise})}, v) = \mathscr{B}_n^{(\mathrm{noise})}(v)$. Furthermore, $\hat{u}_n^{(\mathrm{noise})}$ is the unique minimizer over $H^{m+1}(\Omega, \mathbb{R}^{d_2})$ of

$$\mathscr{R}_n^{(\mathrm{noise})}(u) = \frac{\lambda_d}{n}\sum_{i=1}^n \|\tilde{\Pi}(u)(\mathbf{X}_i) - \varepsilon_i\|_2^2 + \lambda_e \mathbb{E}\|u(\mathbf{X}^{(e)})\|_2^2 + \frac{1}{|\Omega|}\sum_{k=1}^M \int_\Omega \mathscr{F}_k^{(\mathrm{lin})}(u, \mathbf{x})^2 d\mathbf{x}$$
$$+ \lambda_t \|u\|_{H^{m+1}(\Omega)}^2.$$

Similarly, Proposition 5.5 shows that there exists a unique $\hat{u}_n^\star \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ such that, for all $v \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$, $\mathscr{A}_n(\hat{u}_n^\star, v) = \mathscr{B}_n^\star(v)$, and $\hat{u}_n^\star$ is the unique minimizer over $H^{m+1}(\Omega, \mathbb{R}^{d_2})$ of

$$\mathscr{R}_n^\star(u) = \frac{\lambda_d}{n}\sum_{i=1}^n \|\tilde{\Pi}(u - u^\star)(\mathbf{X}_i)\|_2^2 + \lambda_e \mathbb{E}\|\tilde{\Pi}(u)(\mathbf{X}^{(e)}) - h(\mathbf{X}^{(e)})\|_2^2$$
$$+ \frac{1}{|\Omega|}\sum_{k=1}^M \int_\Omega \mathscr{F}_k(u, \mathbf{x})^2 d\mathbf{x} + \lambda_t \|u\|_{H^{m+1}(\Omega)}^2.$$

By the bilinearity of $\mathscr{A}_n$, one has, for all $v \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$, $\mathscr{A}_n(\hat{u}_n^\star + \hat{u}_n^{(\mathrm{noise})}, v) = \mathscr{B}_n(v)$. However, according to Proposition 5.5, $\hat{u}_n$ is the unique element of $H^{m+1}(\Omega, \mathbb{R}^{d_2})$ satisfying this property. Therefore, $\hat{u}_n = \hat{u}_n^\star + \hat{u}_n^{(\mathrm{noise})}$. **Step 2: Some properties of the minimizers** According to Lemma 2.12, $\hat{u}_n$, $\hat{u}_n^\star$, and $\hat{u}_n^{(\mathrm{noise})}$ are random variables. Our goal in this paragraph is to prove that $\mathbb{E}\|\hat{u}_n\|_{H^{m+1}(\Omega)}^2$,

$\mathbb{E}\|\hat{u}_n^\star\|_{H^{m+1}(\Omega)}^2$, and $\mathbb{E}\|\hat{u}_n^{(\mathrm{noise})}\|_{H^{m+1}(\Omega)}^2$ are finite, so that we can safely use conditional expectations on $\hat{u}_n$, $\hat{u}_n^\star$, and $\hat{u}_n^{(\mathrm{noise})}$. Recall that, since $\lambda_t\|\hat{u}_n\|_{H^{m+1}(\Omega)}^2 \leqslant \mathscr{R}_n^{(\mathrm{reg})}(\hat{u}_n) \leqslant \mathscr{R}_n^{(\mathrm{reg})}(0)$, and since $\mathscr{F}_k^{(\mathrm{lin})}(0,\cdot) = 0$,

$$\lambda_t\|\hat{u}_n\|_{H^{m+1}(\Omega)}^2 \leqslant \frac{\lambda_d}{n}\sum_{i=1}^n \|Y_i\|_2^2 + \lambda_e\mathbb{E}\|h(\mathbf{X}^{(e)})\|_2^2 + \frac{1}{|\Omega|}\sum_{k=1}^M \int_\Omega B_k(\mathbf{x})^2 d\mathbf{x}.$$

Hence,

$$\mathbb{E}\|\hat{u}_n\|_{H^{m+1}(\Omega)}^2 \leqslant \lambda_t^{-1}\Big(\lambda_d\mathbb{E}\|u^\star(\mathbf{X}) + \varepsilon\|_2^2 + \lambda_e\mathbb{E}\|h(\mathbf{X}^{(e)})\|_2^2 + \frac{1}{|\Omega|}\sum_{k=1}^M \int_\Omega B_k(\mathbf{x})^2 d\mathbf{x}\Big).$$

Similarly,

$$\mathbb{E}\|\hat{u}_n^\star\|_{H^{m+1}(\Omega)}^2 \leqslant \lambda_t^{-1}\Big(\lambda_d\mathbb{E}\|u^\star(\mathbf{X})\|_2^2 + \lambda_e\mathbb{E}\|h(\mathbf{X}^{(e)})\|_2^2 + \frac{1}{|\Omega|}\sum_{k=1}^M \int_\Omega B_k(\mathbf{x})^2 d\mathbf{x}\Big),$$

and $\mathbb{E}\|\hat{u}_n^{(\mathrm{noise})}\|_{H^{m+1}(\Omega)}^2 \leqslant \lambda_t^{-1}\lambda_d\mathbb{E}\|\varepsilon\|_2^2$.

Step 3: Bias-variance decomposition In this paragraph, we use the notation $\mathcal{A}_{(\mathbf{x},e)}(u,u)$ instead of $\mathcal{A}_n(u,u)$, to make the dependence of $\mathcal{A}_n$ in the random variables $\mathbf{x} = (\mathbf{X}_1,\dots,\mathbf{X}_n)$ and $e = (\varepsilon_1,\dots,\varepsilon_n)$ more explicit. We do the same with $\mathcal{B}_n$ and $\hat{u}_n^{(\mathrm{noise})}$. Observe that, for any $(\mathbf{x},e) \in \Omega^n \times \mathbb{R}^{nd_2}$ and for any $u \in H^{m+1}(\Omega,\mathbb{R}^{d_2})$, one has

$$\mathcal{A}_{(\mathbf{x},-e)}(u,u) - 2\mathcal{B}_{(\mathbf{x},e)}^{(\mathrm{noise})}(u) = \mathcal{A}_{(\mathbf{x},e)}(-u,-u) - 2\mathcal{B}_{(\mathbf{x},-e)}^{(\mathrm{noise})}(-u).$$

Therefore, $\hat{u}_{(\mathbf{x},e)}^{(\mathrm{noise})} = -\hat{u}_{(\mathbf{x},-e)}^{(\mathrm{noise})}$.

Since, by assumption, ε has the same law as $-\varepsilon$, this implies $\mathbb{E}(\hat{u}_n^{(\mathrm{noise})} \mid \mathbf{X}_1,\dots,\mathbf{X}_n) = 0$, and so $\mathbb{E}(\hat{u}_n^{(\mathrm{noise})}) = 0$. Moreover, since $\hat{u}_n^\star$ is a measurable function of $\mathbf{X}_1,\dots,\mathbf{X}_n$, we have $\mathbb{E}(\hat{u}_n^\star \mid \mathbf{X}_1,\dots,\mathbf{X}_n) = \hat{u}_n^\star$. Recalling (Step 1) that $\hat{u}_n = \hat{u}_n^\star + \hat{u}_n^{(\mathrm{noise})}$, we deduce the following bias-variance decomposition:

$$\mathbb{E}\|\hat{u}_n - u^\star\|_{L^2(\mu_{\mathbf{X}})}^2 = \mathbb{E}\|\hat{u}_n^\star - u^\star\|_{L^2(\mu_{\mathbf{X}})}^2 + \mathbb{E}\|\hat{u}_n^{(\mathrm{noise})}\|_{L^2(\mu_{\mathbf{X}})}^2. \quad (30)$$

Step 4: Bounding the bias Recall that $\hat{u}_n^\star$ minimizes $\mathscr{R}_n^\star$ over $H^{m+1}(\Omega,\mathbb{R}^{d_2})$, so that $\mathscr{R}_n^\star(u^\star) \geqslant \mathscr{R}_n^\star(\hat{u}_n^\star)$. Therefore, $\mathrm{PI}(u^\star) + \lambda_t\|u^\star\|_{H^{m+1}(\Omega)}^2 \geqslant \frac{\lambda_d}{n}\sum_{i=1}^n \|\tilde{\Pi}(\hat{u}_n^\star - u^\star)(\mathbf{X}_i)\|_2^2$. We deduce that

$$\begin{aligned}
&\frac{1}{\lambda_d}\big(\mathrm{PI}(u^\star) + \lambda_t\|u^\star\|_{H^{m+1}(\Omega)}^2\big) \\
&\geqslant \frac{\|\hat{u}_n^\star - u^\star\|_{H^{m+1}(\Omega)}^2}{n}\sum_{i=1}^n \Big\|\tilde{\Pi}\Big(\frac{\hat{u}_n^\star - u^\star}{\|\hat{u}_n^\star - u^\star\|_{H^{m+1}(\Omega)}}\Big)(\mathbf{X}_i)\Big\|_2^2 \\
&\geqslant \|\hat{u}_n^\star - u^\star\|_{L^2(\mu_{\mathbf{X}})}^2 \\
&\quad - \|\hat{u}_n^\star - u^\star\|_{H^{m+1}(\Omega)}^2 \sup_{\|u\|_{H^{m+1}(\Omega)}\leqslant 1}\Big(\mathbb{E}\|\tilde{\Pi}(u)(\mathbf{X})\|_2^2 - \frac{1}{n}\sum_{i=1}^n \|\tilde{\Pi}(u)(\mathbf{X}_i)\|_2^2\Big)
\end{aligned}$$

$$\geqslant \|\hat{u}_n^\star - u^\star\|_{L^2(\mu_{\mathbf{X}})}^2$$

$$- 2\big(\|\hat{u}_n^\star\|_{H^{m+1}(\Omega)}^2 + \|u^\star\|_{H^{m+1}(\Omega)}^2\big) \sup_{\|u\|_{H^{m+1}(\Omega)}\leqslant 1} \Big(\mathbb{E}\|\tilde{\Pi}(u)(\mathbf{X})\|_2^2 - \frac{1}{n}\sum_{i=1}^n \|\tilde{\Pi}(u)(\mathbf{X}_i)\|_2^2\Big).$$

Moreover, $\mathrm{PI}(u^\star) + \lambda_t\|u^\star\|_{H^{m+1}(\Omega)}^2 \geqslant \lambda_t\|\hat{u}_n^\star\|_{H^{m+1}(\Omega)}^2$. Taking expectations, we conclude by Lemma 2.14 that there exists a constant C'_Ω, depending only on Ω, such that

$$\mathbb{E}\|\hat{u}_n^\star - u^\star\|_{L^2(\mu_{\mathbf{X}})}^2 \leqslant \frac{1}{\lambda_d}\big(\mathrm{PI}(u^\star) + \lambda_t\|u^\star\|_{H^{m+1}(\Omega)}^2\big) + \frac{C'_\Omega d_2^{1/2}}{n^{1/2}}\Big(2\|u^\star\|_{H^{m+1}(\Omega)}^2 + \frac{\mathrm{PI}(u^\star)}{\lambda_t}\Big).$$

Step 5: Bounding the variance Since $\hat{u}_n^{(\mathrm{noise})}$ minimizes $\mathscr{R}_n^{(\mathrm{noise})}$ over $H^{m+1}(\Omega, \mathbb{R}^{d_2})$, we have $\mathscr{R}_n^{(\mathrm{noise})}(0) \geqslant \mathscr{R}_n^{(\mathrm{noise})}(\hat{u}_n^{(\mathrm{noise})})$. So,

$$\frac{\lambda_d}{n}\sum_{i=1}^n \|\varepsilon_i\|_2^2 \geqslant \frac{\lambda_d}{n}\sum_{i=1}^n \|\tilde{\Pi}(\hat{u}_n^{(\mathrm{noise})})(\mathbf{X}_i) - \varepsilon_i\|_2^2.$$

Observing that $\|\tilde{\Pi}(\hat{u}_n^{(\mathrm{noise})})(\mathbf{X}_i) - \varepsilon_i\|_2^2 = \|\tilde{\Pi}(\hat{u}_n^{(\mathrm{noise})})(\mathbf{X}_i)\|_2^2 - 2\langle\tilde{\Pi}(\hat{u}_n^{(\mathrm{noise})})(\mathbf{X}_i), \varepsilon_i\rangle + \|\varepsilon_i\|_2^2$, we deduce that

$$\frac{2}{n}\sum_{i=1}^n \langle\tilde{\Pi}(\hat{u}_n^{(\mathrm{noise})})(\mathbf{X}_i), \varepsilon_i\rangle \geqslant \frac{1}{n}\sum_{i=1}^n \|\tilde{\Pi}(\hat{u}_n^{(\mathrm{noise})})(\mathbf{X}_i)\|_2^2,$$

and

$$\Big\langle \int_\Omega \tilde{\Pi}(\hat{u}_n^{(\mathrm{noise})})d\mu_{\mathbf{X}}, \frac{2}{n}\sum_{i=1}^n \varepsilon_i\Big\rangle + \frac{2}{n}\sum_{i=1}^n \Big\langle \tilde{\Pi}(\hat{u}_n^{(\mathrm{noise})})(\mathbf{X}_i) - \int_\Omega \tilde{\Pi}(\hat{u}_n^{(\mathrm{noise})})d\mu_{\mathbf{X}}, \varepsilon_i\Big\rangle$$

$$\geqslant \frac{1}{n}\sum_{i=1}^n \|\tilde{\Pi}(\hat{u}_n^{(\mathrm{noise})})(\mathbf{X}_i)\|_2^2.$$

Therefore,

$$\begin{aligned}\|\hat{u}_n^{(\mathrm{noise})}\|_{L^2(\mu_{\mathbf{X}})}^2 &\leqslant \Big\langle \int_\Omega \tilde{\Pi}(\hat{u}_n^{(\mathrm{noise})})d\mu_{\mathbf{X}}, \frac{2}{n}\sum_{i=1}^n \varepsilon_i\Big\rangle \\ &+ \|\hat{u}_n^{(\mathrm{noise})}\|_{H^{m+1}(\Omega)} \sup_{\|u\|_{H^{m+1}(\Omega)}\leqslant 1} \frac{1}{n}\sum_{j=1}^n \langle\tilde{\Pi}(u)(\mathbf{X}_j) - \mathbb{E}(\tilde{\Pi}(u)(\mathbf{X})), \varepsilon_j\rangle \\ &+ \|\hat{u}_n^{(\mathrm{noise})}\|_{H^{m+1}(\Omega)}^2 \sup_{\|u\|_{H^{m+1}(\Omega)}\leqslant 1} \Big(\mathbb{E}\|\tilde{\Pi}(u)(\mathbf{X}_i)\|_2^2 - \frac{1}{n}\sum_{i=1}^n \|\tilde{\Pi}(u)(\mathbf{X}_i)\|_2^2\Big) \\ &:= A + B + C.\end{aligned}$$

According to the Cauchy-Schwarz inequality,

$$\mathbb{E}(A) \leqslant \Big(\mathbb{E}\Big\|\int_\Omega \tilde{\Pi}(\hat{u}_n^{(\mathrm{noise})})d\mu_{\mathbf{X}}\Big\|_2^2\Big)^{1/2} \times \frac{2(\mathbb{E}\|\varepsilon\|_2^2)^{1/2}}{n^{1/2}},$$

and so, by Jensen's inequality,

$$\mathbb{E}(A) \leqslant \big(\mathbb{E}\|\hat{u}_n^{(\text{noise})}\|_{L^2(\mu_{\mathbf{X}})}^2\big)^{1/2} \times \frac{2(\mathbb{E}\|\varepsilon\|_2^2)^{1/2}}{n^{1/2}}.$$

The inequality $\mathscr{R}_n^{(\text{noise})}(0) \geqslant \mathscr{R}_n^{(\text{noise})}(\hat{u}_n^{(\text{noise})})$ also implies that

$$\frac{\lambda_d}{n}\sum_{i=1}^{n}\|\varepsilon_i\|_2^2 \geqslant \frac{\lambda_d}{n}\sum_{i=1}^{n}\|\tilde{\Pi}(\hat{u}_n^{(\text{noise})})(\mathbf{X}_i) - \varepsilon_i\|_2^2 + \lambda_t\|\hat{u}_n^{(\text{noise})}\|_{H^{m+1}(\Omega)}^2.$$

Therefore,

$$\frac{\lambda_d}{n\lambda_t}\sum_{i=1}^{n} 2\langle\tilde{\Pi}(\hat{u}_n^{(\text{noise})})(\mathbf{X}_i), \varepsilon_i\rangle \geqslant \|\hat{u}_n^{(\text{noise})}\|_{H^{m+1}(\Omega)}^2,$$

and

$$\frac{\lambda_d}{\lambda_t}\sup_{\|u\|_{H^{m+1}(\Omega)}\leqslant 1}\frac{1}{n}\sum_{j=1}^{n}\langle\tilde{\Pi}(u)(\mathbf{X}_j), \varepsilon_j\rangle \geqslant \|\hat{u}_n^{(\text{noise})}\|_{H^{m+1}(\Omega)}.$$

By Theorem 1.1, if $\|u\|_{H^{m+1}(\Omega)} \leqslant 1$, then $\langle\mathbb{E}(\tilde{\Pi}(u)(\mathbf{X})), \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\rangle \leqslant \frac{C_\Omega d_2^{1/2}}{n}\|\sum_{i=1}^{n}\varepsilon_i\|_2$. Thus,

$$\begin{aligned}&\|\hat{u}_n^{(\text{noise})}\|_{H^{m+1}(\Omega)}\\ &\leqslant \frac{\lambda_d}{\lambda_t}\Big(\frac{C_\Omega d_2^{1/2}}{n}\|\sum_{i=1}^{n}\varepsilon_i\|_2 + \sup_{\|u\|_{H^{m+1}(\Omega)}\leqslant 1}\frac{1}{n}\sum_{j=1}^{n}\langle\tilde{\Pi}(u)(\mathbf{X}_j) - \mathbb{E}(\tilde{\Pi}(u)(\mathbf{X})), \varepsilon_j\rangle\Big).\end{aligned}$$

Using Lemma 2.15 together with the fact that, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}$, $(\mathbf{x}+\mathbf{y})^2 \leqslant 2(\mathbf{x}^2+\mathbf{y}^2)$,

$$\mathbb{E}\|\hat{u}_n^{(\text{noise})}\|_{H^{m+1}(\Omega)}^2 \leqslant \frac{4\lambda_d^2}{n\lambda_t^2}C_\Omega^2 d_2\mathbb{E}\|\varepsilon\|_2^2.$$

Similarly, observing that for all random variables $X, Y \in \mathbb{R}$, $\mathbb{E}(XY)^2 \leqslant \mathbb{E}(X^2)\mathbb{E}(Y^2)$,

$$\mathbb{E}(B) \leqslant \frac{4\lambda_d}{n\lambda_t}C_\Omega^2 d_2\mathbb{E}\|\varepsilon\|_2^2.$$

Moreover, by Lemma 2.14 and the inequality $\mathbb{E}(XYZ)^2 \leqslant \mathbb{E}(X^2)\mathbb{E}(Y^2)\mathbb{E}(Z^2)$,

$$\mathbb{E}(C) \leqslant \frac{\lambda_d^2}{n^{3/2}\lambda_t^2}C_\Omega^2 d_2^{3/2}\mathbb{E}\|\varepsilon\|_2^2.$$

Therefore, we conclude that there exists a constant $C_\Omega > 0$, depending only on Ω, such that

$$\begin{aligned}\mathbb{E}\|\hat{u}_n^{(\text{noise})}\|_{L^2(\mu_{\mathbf{X}})}^2 &\leqslant \big(\mathbb{E}\|\hat{u}_n^{(\text{noise})}\|_{L^2(\mu_{\mathbf{X}})}^2\big)^{1/2}\frac{2(\mathbb{E}\|\varepsilon\|_2^2)^{1/2}}{n^{1/2}}\\ &\quad + \frac{4\lambda_d}{n\lambda_t}C_\Omega^2 d_2\mathbb{E}\|\varepsilon\|_2^2 + \frac{\lambda_d^2}{n^{3/2}\lambda_t^2}C_\Omega^2 d_2^{3/2}\mathbb{E}\|\varepsilon\|_2^2.\end{aligned}$$

Hence, using elementary algebra,

$$\big(\mathbb{E}\|\hat{u}_n^{(\text{noise})}\|^2_{L^2(\mu_{\mathbf{X}})}\big)^{1/2} \leqslant \frac{(\mathbb{E}\|\varepsilon\|_2^2)^{1/2}}{n^{1/2}}\Big(2 + 2C_\Omega d_2^{3/4}\Big(\frac{\lambda_d^{1/2}}{\lambda_t^{1/2}} + \frac{\lambda_d}{\lambda_t n^{1/4}}\Big)\Big)$$

and

$$\mathbb{E}\|\hat{u}_n^{(\text{noise})}\|^2_{L^2(\mu_{\mathbf{X}})} \leqslant \frac{8\mathbb{E}\|\varepsilon\|_2^2}{n}\Big(1 + C_\Omega d_2^{3/2}\Big(\frac{\lambda_d}{\lambda_t} + \frac{\lambda_d^2}{\lambda_t^2 n^{1/2}}\Big)\Big).$$

Step 6: Putting everything together Combining Steps 3, 4, and 5, we conclude that

$$\mathbb{E}\|\hat{u}_n - u^\star\|^2_{L^2(\mu_{\mathbf{X}})} \leqslant \frac{1}{\lambda_d}\big(\mathrm{PI}(u^\star) + \lambda_t\|u^\star\|^2_{H^{m+1}(\Omega)}\big) + \frac{C'_\Omega d_2^{1/2}}{n^{1/2}}\Big(2\|u^\star\|^2_{H^{m+1}(\Omega)} + \frac{\mathrm{PI}(u^\star)}{\lambda_t}\Big) + \frac{8\mathbb{E}\|\varepsilon\|_2^2}{n}\Big(1 + C_\Omega d_2^{3/2}\Big(\frac{\lambda_d}{\lambda_t} + \frac{\lambda_d^2}{\lambda_t^2 n^{1/2}}\Big)\Big).$$

6.6. Proof of Proposition 5.12

By definition, $\hat{u}_n$ minimizes $\mathscr{R}_n^{(\text{reg})}$ over $H^{m+1}(\Omega, \mathbb{R}^{d_2})$. So, $\mathscr{R}_n^{(\text{reg})}(u^\star) \geqslant \mathscr{R}_n^{(\text{reg})}(\hat{u}_n)$. Moreover, since

$$\|\tilde{\Pi}(\hat{u}_n)(\mathbf{X}_i) - Y_i\|_2^2 = \|\tilde{\Pi}(\hat{u}_n - u^\star)(\mathbf{X}_i)\|_2^2 - 2\langle\tilde{\Pi}(\hat{u}_n - u^\star)(\mathbf{X}_i), \varepsilon_i\rangle + \|\varepsilon_i\|_2^2,$$

one has

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n\|\tilde{\Pi}(\hat{u}_n)(\mathbf{X}_i) - Y_i\|_2^2\\ &\quad\geqslant -2\|\hat{u}_n - u^\star\|_{H^{m+1}(\Omega)} \times \sup_{\|u\|_{H^{m+1}(\Omega)}\leqslant 1}\frac{1}{n}\sum_{j=1}^n\langle\tilde{\Pi}(u)(\mathbf{X}_j) - \mathbb{E}(\tilde{\Pi}(u)(\mathbf{X})), \varepsilon_j\rangle\\ &\qquad - 2\Big\langle\int_\Omega\tilde{\Pi}(\hat{u}_n - u^\star)d\mu_{\mathbf{X}}, \frac{1}{n}\sum_{i=1}^n\varepsilon_i\Big\rangle + \frac{1}{n}\sum_{i=1}^n\|\varepsilon_i\|_2^2.\end{aligned}$$

Thus,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n\|\tilde{\Pi}(\hat{u}_n)(\mathbf{X}_i) - Y_i\|_2^2\\ &\quad\geqslant -2(\|\hat{u}_n\|_{H^{m+1}(\Omega)} + \|u^\star\|_{H^{m+1}(\Omega)})\sup_{\|u\|_{H^{m+1}(\Omega)}\leqslant 1}\frac{1}{n}\sum_{j=1}^n\langle\tilde{\Pi}(u)(\mathbf{X}_j) - \mathbb{E}(\tilde{\Pi}(u)(\mathbf{X})), \varepsilon_j\rangle\\ &\qquad - 2\Big\langle\int_\Omega\tilde{\Pi}(\hat{u}_n - u^\star)d\mu_{\mathbf{X}}, \frac{1}{n}\sum_{i=1}^n\varepsilon_i\Big\rangle + \frac{1}{n}\sum_{i=1}^n\|\varepsilon_i\|_2^2. \qquad (31)\end{aligned}$$

Recall from Steps 4 and 5 of the proof of Theorem 5.11 that

$$\mathbb{E}\|\hat{u}_n\|^2_{H^{m+1}(\Omega)} \leqslant 2\mathbb{E}\|\hat{u}_n^\star\|^2_{H^{m+1}(\Omega)} + 2\mathbb{E}\|\hat{u}_n^{(\text{noise})}\|^2_{H^{m+1}(\Omega)}$$

$$\leqslant 2\Big(\frac{\mathrm{PI}(u^\star)}{\lambda_t} + \|u^\star\|^2_{H^{m+1}(\Omega)}\Big) + \frac{8\lambda_d^2}{n\lambda_t^2} C_\Omega^2 d_2 \mathbb{E}\|\varepsilon\|_2^2$$

Therefore, Lemma 2.15 and the inequality $\mathbb{E}(XY)^2 \leqslant \mathbb{E}(X)^2\mathbb{E}(Y)^2$ show that

$$\mathbb{E}\Big(\|\hat{u}_n\|_{H^{m+1}(\Omega)} \sup_{\|u\|_{H^{m+1}(\Omega)}\leqslant 1} \frac{1}{n}\sum_{j=1}^n \langle \tilde{\Pi}(u)(\mathbf{X}_j) - \mathbb{E}(\tilde{\Pi}(u)(\mathbf{X})), \varepsilon_j\rangle\Big) = \underset{n\to\infty}{O}\Big(\frac{\lambda_d}{n\lambda_t}\Big).$$

By Theorem 5.11,

$$\mathbb{E}\Big|\Big\langle \int_\Omega \tilde{\Pi}(\hat{u}_n - u^\star) d\mu_{\mathbf{X}}, \frac{1}{n}\sum_{i=1}^n \varepsilon_i\Big\rangle\Big| \leqslant \big(\mathbb{E}\|u^\star - \hat{u}_n\|^2_{L^2(\mu_{\mathbf{X}})}\big)^{1/2}\frac{\mathbb{E}\|\varepsilon\|_2^2}{n^{1/2}} = \underset{n\to\infty}{O}\Big(\frac{\lambda_d}{n^2\lambda_t}\Big)^{1/2}.$$

Combining these three results with (31), we conclude that

$$\mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^n \|\tilde{\Pi}(\hat{u}_n)(\mathbf{X}_i) - Y_i\|_2^2\Big) \geqslant \mathbb{E}\|\varepsilon\|_2^2 + \underset{n\to\infty}{O}\Big(\frac{\lambda_d}{n\lambda_t}\Big).$$

Therefore, since $\lim_{n\to\infty} \frac{\lambda_d^2}{n\lambda_t} = 0$ and since $\mathscr{R}_n^{(\mathrm{reg})}(\hat{u}_n) = \frac{\lambda_d}{n}\sum_{i=1}^n \|\tilde{\Pi}(\hat{u}_n)(\mathbf{X}_i) - Y_i\|_2^2 + \mathrm{PI}(\hat{u}_n) + \lambda_t\|\hat{u}_n\|^2_{H^{m+1}(\Omega)}$,

$$\mathbb{E}\big(\mathscr{R}_n^{(\mathrm{reg})}(\hat{u}_n)\big) \geqslant \lambda_d\mathbb{E}\|\varepsilon\|_2^2 + \mathbb{E}(\mathrm{PI}(\hat{u}_n)) + \underset{n\to\infty}{o}(1).$$

Similarly, almost everywhere,

$$\frac{1}{n}\sum_{i=1}^n \|\tilde{\Pi}(\hat{u}^\star)(\mathbf{X}_i) - Y_i\|_2^2 = \frac{1}{n}\sum_{i=1}^n \|\varepsilon_i\|_2^2.$$

Hence,

$$\mathbb{E}\big(\mathscr{R}_n^{(\mathrm{reg})}(u^\star)\big) = \lambda_d\mathbb{E}\|\varepsilon\|_2^2 + \mathrm{PI}(u^\star) + \lambda_t\|u^\star\|^2_{H^{m+1}(\Omega)}.$$

Since $\mathbb{E}(\mathscr{R}_n^{(\mathrm{reg})}(\hat{u}_n)) \leqslant \mathbb{E}(\mathscr{R}_n^{(\mathrm{reg})}(u^\star))$ and since $\lambda_t \to 0$, we are led to

$$\mathbb{E}(\mathrm{PI}(\hat{u}_n)) \leqslant \mathrm{PI}(u^\star) + \underset{n\to\infty}{o}(1),$$

which is the desired result.

References

ARNONE, E., KNEIP, A., NOBILE, F. and SANGALLI, L. M. (2022). Some first results on the consistency of spatial regression with partial differential equation regularization. *Stat. Sinica* **32** 209–238. https://doi.org/10.5705/ss.202019.0346

BREZIS, H. (2010). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, New York. https://doi.org/10.1007/978-0-387-70914-7

COMTET, L. (1974). *Advanced Combinatorics : The Art of Finite and Infinite Expansions*. Springer, Dordrecht. https://doi.org/10.1007/978-94-010-2196-8

EVANS, L. C. (2010). *Partial Differential Equations*, 2nd ed. *Graduate Studies in Mathematics* **19**. American Mathematical Society, Providence. https://doi.org/10.1090/gsm/019

HARDY, M. (2006). Combinatorics of partial derivatives. *Electron. J. Comb.* **13** R1. https://doi.org/10.48550/arXiv.math/0601149

NICKL, R. and PÖTSCHER, B. M. (2007). Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. *J. Theor. Probab.* **20** 177–199. https://doi.org/10.1007/s10959-007-0058-1

ROGERS, L. C. G. and WILLIAMS, D. (2000). *Diffusions, Markov processes and Martingales* **1, Foundations**, 2nd ed. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511805141

DE RYCK, T., LANTHALER, S. and MISHRA, S. (2021). On the approximation of functions by tanh neural networks. *Neural Netw.* **143** 732–750. https://doi.org/10.1016/j.neunet.2021.08.015

SHVARTZMAN, P. (2010). On Sobolev extension domains in $\mathbb{R}^n$. *J. Funct. Anal.* **258** 2205–2245. https://doi.org/10.1016/j.jfa.2010.01.002

STEIN, E. M. (1970). *Singular Integrals and Differentiability Properties of Functions*. *Princeton Mathematical Series* **30**. Princeton University Press, Princeton. https://doi.org/10.1515/9781400883882

VAN HANDEL, R. (2016). *Probability in High Dimension*. APC 550 Lecture Notes, Princeton University. https://doi.org/10.21236/ADA623999