Unifying PINNs and FEMs through the notion of Natural Neural Tangent Kernel

ENUMATH 2025

Nilo Schwencke, Roland Maier

TAU Team—INRIA Saclay; A&O—LISN—Paris-Saclay University; CNRS; KIT

September 02, 2025

Context

Problem statement

We aim to solve:

$$\begin{cases} D(u) = f \in \mathrm{L}^2(\Omega \to \mathbb{R}, \mu) & \text{in } \Omega \\ B(u) = g \in \mathrm{L}^2(\partial\Omega \to \mathbb{R}, \sigma) & \text{on } \partial\Omega \end{cases},$$

with D self adjoint. To keep it simple, $D = \Delta$, B trace operator and $g = 0$.

Two notions of solutions

Strong solution : u such that

$$\int_\Omega |\Delta[u] - f|^2 = 0$$

Weak solution : u such that $\forall v \in \mathcal{C}_c^\infty(\Omega)$

$$\int_\Omega \Delta u \, v = \int_\Omega f \, v$$

In which spaces live the solutions ?

$\mathrm{H}_0^1(\Omega)$ space

Note that: $\forall v \in \mathcal{C}_c^\infty(\Omega)$

$$\int_\Omega \Delta u \, v = \int_\Omega \langle \nabla u \, , \, \nabla v \rangle_{\mathbb{R}^{\dim(\Omega)}}$$

Completing $\mathcal{C}_c^\infty(\Omega)$ w.r.t this inner product yields $\mathrm{H}_0^1(\Omega)$, in which weak solutions live.

Lemma (Poincaré)

$O_1 : u \in H_0^1(\Omega) \mapsto u \in L^2(\Omega)$ is continuous.

$\mathrm{H}_0^1(\Omega) \cap \mathrm{H}^2(\Omega)$ space

$\mathrm{H}^2(\Omega)$: max space containing $\mathcal{C}^\infty(\Omega)$ s.t:

- $\mathcal{O}_2 : u \in \mathrm{H}^2(\Omega) \mapsto u \in \mathrm{L}^2(\Omega)$ continuous.
- $\Delta : \mathrm{H}^2(\Omega) \to \mathrm{L}^2(\Omega)$ continuous.

Strong solutions lives in $\mathrm{H}_0^1(\Omega) \cap \mathrm{H}^2(\Omega)$.

Galerkin approach

Choose $(v_k) \in \mathrm{H}^1_0(\Omega)^N$ and approximate a weak solution by solving the linear system: for all $1 \leqslant i, j \leqslant N$

$$\langle v_i \,,\, v_j \rangle_{\mathrm{H}^1_0(\Omega)} = \langle v_i \,,\, f \rangle_{\mathrm{L}^2(\Omega)} \,. \qquad (1)$$

Yields a solution $u_N \in \mathrm{H}_N(\Omega)$:

$$u_N := \sum_{k=1}^{N} \alpha_k v_k,$$

with α a solution to eq. (1) and

$$\mathrm{H}_N(\Omega) := \mathrm{Span}(v_k \;:\; 1 \leqslant k \leqslant N).$$

Lemma

If $u \in H^1_0(\Omega)$ is a weak solution, then $u_N = \Pi_{H_N} u$.

FEMs are a special choice of $(u_k)_{k=1}^N$.

PINNs approach

Choose a Neural Network architecture

$$u : \begin{cases} \mathbb{R}^P & \to & \mathcal{C}^\infty\Omega \\ \theta & \mapsto & u_\theta. \end{cases},$$

and optimize it on the loss:

$$\widehat{\ell}_{D,B}(\boldsymbol{\theta}) := \frac{1}{2S_D} \sum_{i=1}^{S_D} \Big(D[u_{|\boldsymbol{\theta}}](x_i^D) - f(x_i^D) \Big)^2 + \frac{1}{2S_B} \sum_{i=1}^{S_B} \Big(B[u_{|\boldsymbol{\theta}}](x_i^B) - g(x_i^B) \Big)^2,$$

using autodiff to compute D, B (Raissi et al., 2019).

A priori different approaches.

Natural gradient for PINNs

Reinterpreting quadratic loss

Consider the loss of a classical quadratic regression problem, with batch (x_i):

$$\widehat{\ell}(\boldsymbol{\theta}) := \frac{1}{2S} \sum_{i=1}^{S} \left(u_{|\boldsymbol{\theta}}(x_i) - f(x_i)\right)^2 .$$

In the population limit:

$$\widehat{\ell}(\theta) \overset{S\to\infty}{\longrightarrow} \mathcal{L}(u_{|\boldsymbol{\theta}})\,; \quad \mathcal{L}(u) := \frac{1}{2}\,\|u - f\|^2_{\mathrm{L}^2(\Omega)}$$

This yields the Fréchet derivative:

$$\mathrm{d}\mathcal{L}|_u(h) = \langle \underbrace{u - f}_{\nabla\mathcal{L}_{|u}}, h\rangle_{\mathrm{L}^2(\Omega)},$$

and thus the gradient flow:

$$\begin{cases} u_0 \in \mathrm{L}^2(\Omega) \\ \dot{u}_t = -\nabla\mathcal{L}_{|u_t} = f - u_t \end{cases} .$$

<u>**Solution:**</u> $u_t = f + e^{-t}(u_0 - f)$.

A functional geometry perspective

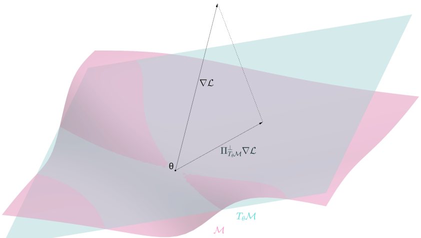
Natural gradient in functional space

The functional space is constrained to:

- $\mathcal{M} := \operatorname{Im} u = \{u_{\boldsymbol{\theta}} \,:\, \boldsymbol{\theta} \in \mathbb{R}^P)$
- $T_{\boldsymbol{\theta}}\mathcal{M} := \operatorname{Im} \mathrm{d}u_{|\boldsymbol{\theta}} = \operatorname{Span}(\partial_p u_{\boldsymbol{\theta}})$

The Natural Gradient is then defined as:

$$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \eta\, \mathrm{d}u^{\dagger}_{|\boldsymbol{\theta}_t}\left(\Pi^{\perp}_{T_{\boldsymbol{\theta}_t}\mathcal{M}} \nabla\mathcal{L}_{|u_{|\boldsymbol{\theta}_t}}\right),$$



Definition-Proposition

An Hilbert space $\mathcal{H}$ of functions $\Omega \to \mathbb{R}$ is a RKHS if and only if the following equivalent conditions are met:

1. *There exist a function $k : \Omega \times \Omega \to \mathbb{R}$ such that:*
   - $\mathcal{H} = \overline{\mathrm{Span}\left(k(x,\cdot) : x \in \Omega\right)}$
   - $\langle k(x,\cdot)\,,\, k(y,\cdot)\rangle_{\mathcal{H}} = k(x,y)$
2. *for all $x \in \Omega$, the evaluation form $e_x : f \in \mathcal{H} \mapsto f(x)$ is continuous.*

Proposition

Any finite dimensional space $\mathcal{H}$ is a RKHS

Proposition

If $\mathcal{H} := \overline{\mathrm{Span}(u_i : i \in \mathbb{N})}$, is an RKHS, then its kernel is given by: for all $x, y \in \Omega$

$$k(x,y) = \sum_{i,j \in \mathbb{N}} u_i(x) G^{\dagger}_{i,j} u_j(y)$$

where $G_{ij} := \langle u_i\,,\, u_j\rangle_{\mathcal{H}}$.

Proposition

If $\mathcal{H} \supset \mathcal{H}_0$ is an RKHS with kernel k, then the orthogonal projection $\Pi^{\perp}_{\mathcal{H}_0}$ onto $\mathcal{H}_0$ is given by:

$$\Pi^{\perp}_{\mathcal{H}_0}(f)(x) = \langle k(x,\cdot)\,,\, f\rangle_{\mathcal{H}}$$

Remark

Applying to $\mathcal{H}_0 = T_\theta \mathcal{M}$, we get natural gradient.

Definition-Proposition (Schwencke and Furtlehner (2025))

*The **Natural Neural Tangent Kernel (NNTK)** is the kernel of the projection* $\Pi_{T_{\boldsymbol{\theta}}\mathcal{M}} : L^2(\Omega) \to L^2(\Omega)$ *onto* $T_{\boldsymbol{\theta}}\mathcal{M}$. *It is given by the formula:*

$$NNTK_{\boldsymbol{\theta}}(x,y) := \sum_{1\leqslant p,q\leqslant P} \partial_p u_{|\boldsymbol{\theta}}(x) \quad G^{\dagger}_{\boldsymbol{\theta}\,pq} \partial_q u_{|\boldsymbol{\theta}}(y); \quad G_{\boldsymbol{\theta}\,p,q} \quad := \left\langle \partial_p u_{|\boldsymbol{\theta}} \,,\, \partial_q u_{|\boldsymbol{\theta}} \right\rangle_{L^2(\Omega)} .$$

Corollary

The Natural Gradient update rewrites: $\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \eta\, G^{\dagger}_{\boldsymbol{\theta}_t} \nabla \ell(\boldsymbol{\theta}_t); \quad \ell(\boldsymbol{\theta}) := \mathcal{L}(u_{|\boldsymbol{\theta}})$.

Application to PINNs

Those definitions extend to PINNs by replacing the model u with:

$$(D,B)\circ u : \begin{cases} \mathbb{R}^P & \to & \mathcal{C}^\infty(\Omega) & \to & \mathrm{L}^2(\Omega \to \mathbb{R}) \times \mathrm{L}^2(\partial\Omega \to \mathbb{R}) \\ \theta & \mapsto & u_\theta & \mapsto & (D[u_\theta], B[u_\theta]). \end{cases} ,$$

Remark

If we assume $\operatorname{Im} u \subset \mathrm{H}^1_0(\Omega)$, then $B[u_\theta] = 0$ and this simplifies to considering $\Delta \circ u : \mathbb{R}^P \to \mathrm{L}^2(\Omega \to \mathbb{R})$. Then in particular $\operatorname{Im} \Delta \circ u \subset \mathrm{H}^1_0(\Omega) \cap \mathrm{H}^2(\Omega)$

Let us introduce the smooth tangent space $\mathcal{TM}_\theta := \operatorname{Im} \mathrm{d}u_\theta$ and the smooth NNTK :

$$\mathcal{NNTK}_{\boldsymbol{\theta}}(x,y) := \sum_{1 \leqslant p,q \leqslant P} \partial_p u_{|\boldsymbol{\theta}}(x) \quad G^{\dagger}_{\boldsymbol{\theta}\,pq} \partial_q u_{|\boldsymbol{\theta}}(y);\, G_{\boldsymbol{\theta}\,p,q} \quad := \left\langle \partial_p \Delta \circ u_{|\boldsymbol{\theta}} \,,\, \partial_q \Delta \circ u_{|\boldsymbol{\theta}} \right\rangle_{\mathrm{L}^2(\Omega)} .$$

Lemma (Poincaré)

The bilinear form: $\langle u \,,\, v \rangle_\Delta := \int_\Omega \Delta[u] \Delta[v]$ *is an inner product on* $H^1_0(\Omega) \cap H^2(\Omega)$*.*

In particular if we endow $\mathcal{TM}_\theta$ with $\langle \cdot \,,\, \cdot \rangle_\Delta$, $\mathcal{NNTK}$ is a reproducing kernel on $\mathcal{TM}$. Furthemore for all $x, y \in \Omega$

$$\begin{aligned} NNTK(x,y) &= \langle NNTK(x,\cdot) \,,\, NNTK(\cdot,y) \rangle_{\mathrm{L}^2(\Omega)} \\ &= \langle \Delta\left[\mathcal{NNTK}(x,\cdot)\right] \,,\, \Delta\left[\mathcal{NNTK}(\cdot,y)\right] \rangle_\Delta = \Delta\left[\Delta\left[\mathcal{NNTK}(x,\cdot)\right](\cdot,y)\right] \end{aligned}$$

In particular Δ is then an isometry, thus $\Delta^{-1} = \Delta^*$.

Definition (Solution in the least-square sense)

A function $u \in \mathcal{H}_0 \subset \mathcal{H}$ is a solution in the least square sense if

$$u \in \underset{v \in \mathcal{H}_0}{\operatorname{argmin}} \|D[u] - f\|_{\mathrm{L}^2(\Omega)}$$

Definition (Green's function of D on $\mathcal{H}_0 \subset \mathcal{H}$ in the least-square sense)

A Green's function in LS sense is any $g : \Omega \times \Omega \to \mathbb{R}$ such that the operator:

$$R : f \in \mathrm{L}^2(\Omega \to \mathbb{R}, \mu) \mapsto \left(x \in \Omega \mapsto \int_\Omega g(x,s) f(s) \mu(\mathrm{d}s) \right) \in \mathcal{H}$$

verifies the equation: $D \circ R = \Pi^{\perp}_{D[\mathcal{H}_0]}$

Proposition

$\Delta \left[\mathcal{NNTK}(x,\cdot)\right]$ *is a Green function in LS sense on* $T\mathcal{M}_\theta$*. In particular:*

and a solution (in the least-square sense) is obtained in $\mathcal{TM}$ by:

$$\langle \Delta \left[\mathcal{NNTK}(x,\cdot)\right], f \rangle_{\mathrm{L}^2(\Omega)} = \langle \mathcal{NNTK}(x,\cdot), \Delta^*[f] \rangle_\Delta = \Delta^{-1}[f](x)$$

Fix $(v_k)_{k=1}^N$ and consider the synthesis operator:

$$\mathcal{T} : \begin{cases} \mathbb{R}^P & \rightarrow & \mathrm{H}_0^1(\Omega) \\ \theta & \mapsto & \sum_{k=1}^N \theta_k v_k . \end{cases} ,$$

Then we can rewrite the Galerkin scheme as:

$$\boldsymbol{\theta}_{t+1} \leftarrow 0 - 1 \times G_{\boldsymbol{\theta}_t}^\dagger \underbrace{\langle \partial_p \mathcal{T}_\theta \,,\, f \rangle_{\mathrm{L}^2(\Omega)}}_{\text{No } \nabla} ; \quad G_{\theta\, p,q} := \langle \partial_p \nabla \circ \mathcal{T}_\theta \,,\, \partial_q \nabla \circ \mathcal{T}_\theta \rangle_{\mathrm{L}^2(\Omega)} = \langle \partial_p \mathcal{T}_\theta \,,\, \partial_q \mathcal{T}_\theta \rangle_{\mathrm{H}_0^1}$$

The only difference is that ∇ is missing on the RHS.

But we can “make it appear” by:

$$\langle \partial_p \mathcal{T}_\theta \,,\, f \rangle_{\mathrm{L}^2(\Omega)} = \langle O_1 \left[\partial_p \mathcal{T}_\theta \right] \,,\, f \rangle_{\mathrm{L}^2(\Omega)} = \langle \partial_p \mathcal{T}_\theta \,,\, O_1^* f \rangle_{\mathrm{H}_0^1} = \langle \nabla \partial_p \mathcal{T}_\theta \,,\, \nabla O_1^* f \rangle_{\mathrm{L}^2(\Omega)} .$$

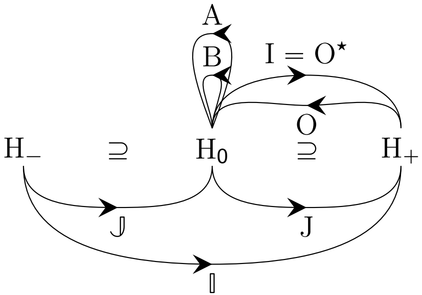


Figure: Schematic diagram of a Hilbert Rigging (a.k.a Gelfand triple).

O embedding operator of H_+ into H_0

$\mathrm{I} := \mathrm{O}^\star : \mathrm{H}_0 \to \mathrm{H}_+$

$\mathrm{A} := \mathrm{OI} : \mathrm{H}_0 \to \mathrm{H}_0 \quad (\text{note that } A^\star = A)$

$\mathrm{B} := \sqrt{A} : \mathrm{H}_0 \to \mathrm{H}_0$

$\mathrm{J} := \mathrm{O}^{-1}\mathrm{B} : \mathrm{H}_0 \to \mathrm{H}_+$

$\mathbb{J} := \overline{\mathrm{B}} : \mathrm{H}_- \to \mathrm{H}_0$

$\mathbb{I} := \overline{\mathrm{I}} : \mathrm{H}_- \to \mathrm{H}_+$

Hilbert Riggings framework

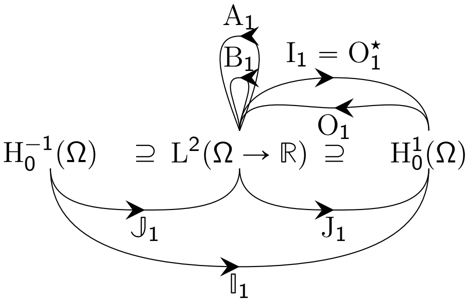


Figure: Schematic diagram of a Hilbert Rigging for triple $H_0^{-1}(\Omega) \supseteq L^2(\Omega \to \mathbb{R}) \supseteq H_0^1(\Omega)$. The lower part is a commutative diagram.

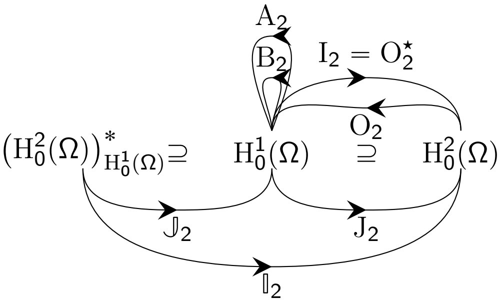


Figure: Schematic diagram of a Hilbert Rigging for $L^2(\Omega \to \mathbb{R}) \supseteq H_0^1(\Omega) \supseteq H_0^1(\Omega) \cap H^2(\Omega)$. The lower part is a commutative diagram.

Hilbert Riggings framework

$$
\begin{tikzcd}
\mathrm{H}_0^{-1}(\Omega) \arrow[rr, bend right, "\mathbb{J}_1"'] \arrow[rrrr, bend right=50, "\mathbb{I}_1"'] & \supseteq & \mathrm{L}^2(\Omega \to \mathbb{R}) \arrow[loop above, "\mathrm{B}_1"] \arrow[loop above, looseness=12, "\mathrm{A}_1"] \arrow[rr, bend left, "\mathrm{I}_1 = \mathrm{O}_1^\star"] \arrow[rr, bend right, "\mathrm{J}_1 = \mathbb{J}_2"'] \arrow[rrrr, bend right=50, "\mathbb{I}_2 = \Delta^{-1}"'] & \supseteq & \mathrm{H}_0^1(\Omega) \arrow[ll, bend left=10, "\mathrm{O}_1"] \arrow[loop above, "\mathrm{B}_2"] \arrow[loop above, looseness=12, "\mathrm{A}_2"] \arrow[rr, bend left, "\mathrm{I}_2 = \mathrm{O}_2^\star"] \arrow[rr, bend right, "\mathrm{J}_2"'] & \supseteq & \mathrm{H}_0^2(\Omega) \arrow[ll, bend left=10, "\mathrm{O}_2"]
\end{tikzcd}
$$

Figure: Strong and Weak Hilbert Riggings in Sobolev spaces. The lower part is a commutative diagram.

In particular $I_2 = \mathbb{I}_2 O_1$. Thus $O_2 = I_1 \mathbb{I}_2^*$ *i.e.* $O_2 \mathbb{I}_2 = I_1$.

This is the classical result stating that a weak solution is a strong solution if $f \in \mathrm{L}^2(\Omega)$ and u is in $\mathrm{H}_0^1(\Omega) \cap \mathrm{H}^2(\Omega)$.

Thank you for your attention !

Jacot, A., F. Gabriel, and C. Hongler (2018): "Neural Tangent Kernel: Convergence and Generalization in Neural Networks," *Advances in neural information processing systems*, 31.

Raissi, M., P. Perdikaris, and G. Karniadakis (2019): "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations," *Journal of Computational Physics*, 378, 686–707.

Rudner, T. G., F. Wenzel, Y. W. Teh, and Y. Gal (2019): "The Natural Neural Tangent Kernel: Neural Network Training Dynamics under Natural Gradient Descent," in *4th Workshop on Bayesian Deep Learning (NeurIPS 2019)*.

Schwencke, N. and C. Furtlehner (2025): "ANaGRAM: A Natural Gradient Relative to Adapted Model for Efficient PINNs Learning," in *The Thirteenth International Conference on Learning Representations*.