

ANaGRAM : A Natural Gradient Relative To Adapted Model for efficient PINNs learning

12^{ème} Biennale de la SMAI

Nilo Schwencke, Cyril Furtlehner

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

June 06, 2025



- 1 Physics informed neural networks (PINNs)
 - PINNs in a nutshell
 - Why does it work so bad ?
- 2 Natural Gradient
 - A functional geometry perspective
 - Reproducing Kernel Hilbert Spaces (RKHS) *détour*
 - Kernel and computational perspective

- 3 empirical Natural Gradient
- 4 Application to PINNs
 - PINNs are a quadratic regression problem
 - Natural Gradient and Green's function
- 5 Experiments
 - NTK vs NNTK of PINNs
- 6 Conclusion and Perspectives

Physics informed neural networks (PINNs)

Problem statement

We aim to solve:

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

PINNs key idea

Optimize a neural network $u|_{\theta}$ on the loss:

$$\begin{aligned} \hat{\ell}_{D,B}(\theta) := & \frac{1}{2S_D} \sum_{i=1}^{S_D} \left(D[u|_{\theta}](x_i^D) - f(x_i^D) \right)^2 \\ & + \frac{1}{2S_B} \sum_{i=1}^{S_B} \left(B[u|_{\theta}](x_i^B) - g(x_i^B) \right)^2, \end{aligned}$$

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

Problem

This leads to low accuracy with SGD.

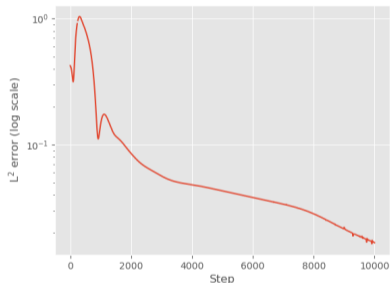


Figure: PINN solution under standard Adam optimization, to Laplace equation in 2D.

Why does it work so bad ?

Intuition from NTK

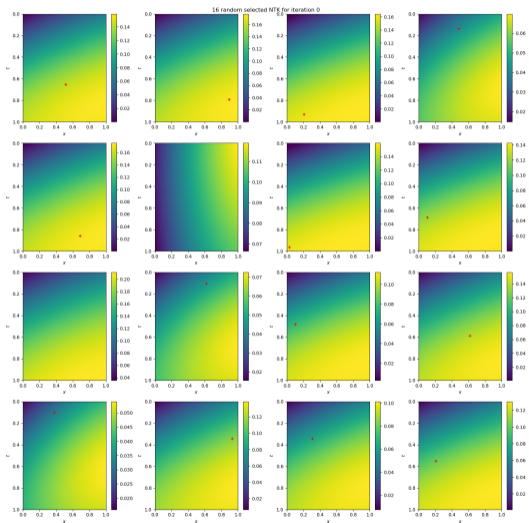


Figure: NTK for Laplace equation.

Intuition from Fourier

Consider the Fourier partial series:

$$S_N : \begin{cases} \mathbb{C}[-N, N] & \rightarrow L^2([0, 1]) \\ (\alpha_k) & \mapsto \sum_{k=-N}^N \alpha_k e^{2i\pi kx} \end{cases}$$

S_N singular values are all 1 (perfect conditioning of the spectrum). **BUT**:

$$\Delta[S_N] \text{ spectrum is } \{4\pi^2 k^2 : 1 \leq k \leq N\}$$

Conclusion: Differential operators strongly impact the spectral condition.

Natural Gradient

Reinterpreting quadratic loss

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

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

In the population limit:

$$\hat{\ell}(\boldsymbol{\theta}) \xrightarrow{S \rightarrow \infty} \mathcal{L}(u_{|\boldsymbol{\theta}}); \quad \mathcal{L}(u) := \frac{1}{2} \|u - f\|_{L^2(\Omega)}^2$$

This yields the Fréchet derivative:

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

and thus the gradient flow:

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

Solution: $u_t = f - e^{-t}(u_0 - f)$.

A functional geometry perspective

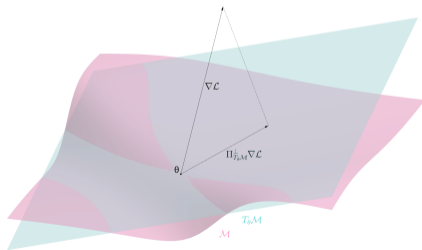
Natural gradient in functional space

The functional space is constrained to:

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

The Natural Gradient is then defined as:

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



Reproducing Kernel Hilbert Spaces (RKHS) *détour*

Definition-Proposition

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

- 1 There exist a function $k : \Omega \times \Omega \rightarrow \mathbb{R}$ such that:
 - $\mathcal{H} = \overline{\text{Span}(k(x, \cdot) : x \in \Omega)}$
 - $\langle k(x, \cdot), k(y, \cdot) \rangle_{\mathcal{H}} = k(x, y)$
- 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 $\mathcal{H} := \overline{\text{Span}(u_i : i \in \mathbb{N})}$, is an RKHS, then its kernel is given by: for all $x, y \in \Omega$

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

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 $\Pi_{\mathcal{H}_0}^{\perp}$ onto \mathcal{H}_0 is given by:

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

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 $\Pi_{T_\theta \mathcal{M}} : L^2(\Omega) \rightarrow L^2(\Omega)$ onto $T_\theta \mathcal{M}$. It is given by the formula:

$$NNTK_\theta(x, y) := \sum_{1 \leq p, q \leq P} (\partial_p u|_\theta(x)) \ G_{\theta pq}^\dagger (\partial_q u|_\theta(y))^t; \quad G_{\theta p, q} := \langle \partial_p u|_\theta, \partial_q u|_\theta \rangle_{L^2(\Omega)}.$$

Corollary

The Natural Gradient update rewrites: $\theta_{t+1} \leftarrow \theta_t - \eta G_{\theta_t}^\dagger \nabla \ell(\theta_t)$; $\ell(\theta) := \mathcal{L}(u|_\theta)$.

Shortcomings

- Computation of the Gram matrix G_{θ_t} is quadratic in the number of parameters.
- Inversion of G_{θ_t} is cubic

We introduce a the empirical Natural Gradient that scales linearly with the number of parameters.

empirical Natural Gradient

(N)NTK in a nutshell

The functional dynamic of (N)GD on the empirical loss $\hat{\ell}$ is described by (Jacot et al., 2018; Rudner et al., 2019):

$$\frac{du_{\theta_t}}{dt}(x) = - \sum_{i=1}^S (N)NTK_{\theta_t}(x, x_i)(u_{|\theta_t}(x_i) - y_i),$$

Key Idea

The empirical dynamics takes place in:

$$\hat{T}_{\theta}\mathcal{M} := \text{Span} (NNTK_{\theta}(x_i, \cdot) : (x_i)_{1 \leq i \leq N}).$$

We can define the empirical Natural Gradient:

$$\theta_{t+1} = \theta_t - \eta du_{|\theta_t}^{\dagger} \left(\Pi_{\hat{T}_{\theta_t}\mathcal{M}}^{\perp} \nabla \mathcal{L}_{|u_{|\theta_t}} \right).$$

Byproduct

Yields an optimal criterion for (x_i) choice:

$$(x_i)^{\star} = \underset{(x_i) \in \Omega^S}{\operatorname{argmin}} \left\| \Pi_{\hat{T}_{\theta, K}^{(x_i)}\mathcal{M}}^{\perp} \nabla \mathcal{L}_{|u_{|\theta_t}} - \nabla \mathcal{L}_{|u_{|\theta_t}} \right\|_{L^2}$$

empirical Natural Gradient

Theorem (ANaGRAM)

Under mild assumptions:

$$du_{|\theta_t}^{\dagger} \left(\Pi_{\hat{T}_{\theta_t}\mathcal{M}}^{\perp} \nabla \mathcal{L}_{|u_{|\theta_t}} \right) = \hat{\phi}_{\theta_t}^{\dagger} \widehat{\nabla} \mathcal{L}_{\theta_t},$$

with: for all $1 \leq p \leq P, 1 \leq i \leq S$

- $\hat{\phi}_{\theta_t i, p} := \partial_p u_{|\theta_t}(x_i)$
- $\widehat{\nabla} \mathcal{L}_{\theta_t i} := \nabla \mathcal{L}_{|u_{|\theta_t}}(x_i)$

Key fact

$\hat{\phi}_{\theta_t}^{\dagger}$ can be computed with a SVD, with complexity $O(\min(PS^2, P^2S))$.

Corollary

There exist P points (\hat{x}_i) such that:

$$\Pi_{\hat{T}_{\theta}\mathcal{M}}^{\perp} \nabla \mathcal{L}_{|u_{|\theta}} = \Pi_{T_{\theta}\mathcal{M}}^{\perp} \nabla \mathcal{L}_{|u_{|\theta}}.$$

Application to PINNs

Key remark

The only difference between the losses:

$$\hat{\ell}_{D,B}(\theta) := \frac{1}{2S_D} \sum_{i=1}^{S_D} \left(D[u|_{\theta}](x_i^D) - f(x_i^D) \right)^2 + \frac{1}{2S_B} \sum_{i=1}^{S_B} \left(B[u|_{\theta}](x_i^B) - g(x_i^B) \right)^2,$$

and $\hat{\ell}(\theta) := \frac{1}{2S} \sum_{i=1}^S (u|_{\theta}(x_i) - f(x_i))^2$ is the use of the operators D and B .

Proposition

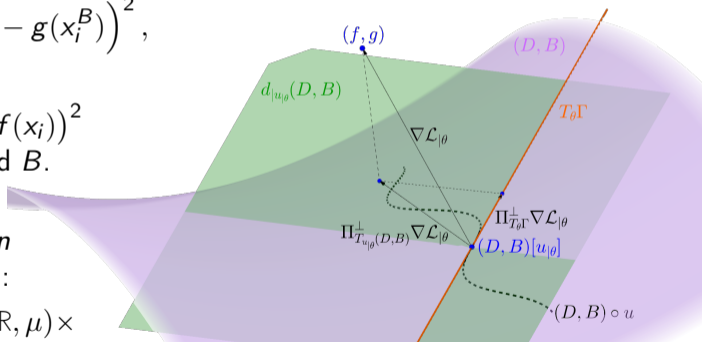
PINNs are a quadratic regression problem with model: $(D, B) \circ u$:

$$\begin{cases} \mathbb{R}^P & \rightarrow \mathcal{H} & \rightarrow L^2(\Omega \rightarrow \mathbb{R}, \mu) \times \\ & & L^2(\partial\Omega \rightarrow \mathbb{R}, \sigma) \\ \theta & \mapsto u|_{\theta} & \mapsto (D[u|_{\theta}], B[u|_{\theta}]) \end{cases}$$

PINNs are a quadratic regression problem

Natural Gradient of PINNs

Figure: Illustration of PINNs Natural Gradient



Definition (Green's function of D)

A Green's function is any kernel function $g : \Omega \times \Omega \rightarrow \mathbb{R}$ such that the operator:

$$R : f \in D[\mathcal{H}] \mapsto \left(x \in \Omega \mapsto \int_{\Omega} g(x, s) f(s) \mu(ds) \right) \in \mathcal{H}$$

verifies the equation: $D \circ R = I_{D[\mathcal{H}]}$

Definition (generalized Green's function of D on $\mathcal{H}_0 \subset \mathcal{H}$)

A generalized Green's function is any kernel function $g : \Omega \times \Omega \rightarrow \mathbb{R}$ such that the operator:

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

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

Theorem

Let $D : \mathcal{H} \rightarrow L^2(\Omega \rightarrow \mathbb{R}, \mu)$ be a linear differential operator and $u : \mathbb{R}^P \rightarrow \mathcal{H}$ a parametric model. Then for all $\theta \in \mathbb{R}^P$, the generalized Green's function of D on $T_\theta \mathcal{M} = \text{Im } du|_\theta$ is given by: for all $x, y \in \Omega$

$$g_{T_\theta \mathcal{M}}(x, y) := \sum_{1 \leq p, q \leq P} \partial_p u|_\theta(x) G_{p,q}^\dagger \partial_q D[u|_\theta](y),$$

with: for all $1 \leq p, q \leq P$

$$G_{pq} := \langle \partial_p D[u|_\theta], \partial_q D[u|_\theta] \rangle_{L^2(\Omega \rightarrow \mathbb{R}, \mu)}.$$

In particular, the natural gradient of PINNs can be rewritten:

$$\theta_{t+1} \leftarrow \theta_t - \eta du|_{\theta_t}^\dagger \left(x \in \Omega \mapsto \int_{\Omega} g_{T_{\theta_t} \mathcal{M}}(x, y) \nabla \mathcal{L}|_{\theta_t}(y) \mu(dy) \right),$$

Experiments

2D Laplace equation

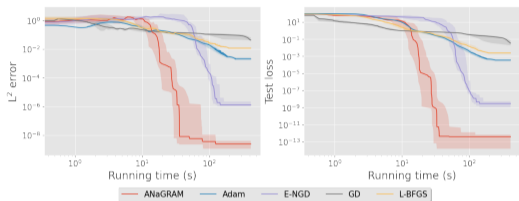


Figure: Performance comparison w.r.t running time for Laplace equation in 2 D:

$$\begin{cases} \Delta u = -2\pi^2 \sin(\pi x_1) \sin(\pi x_2) & \text{in } [0, 1]^2 \\ u = 0 & \text{on } \partial[0, 1]^2 \end{cases}$$

1+1 D Heat equation

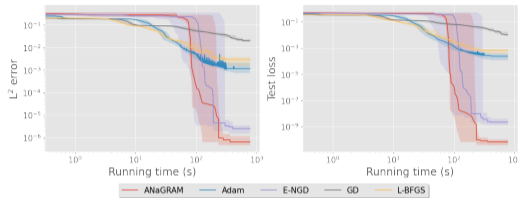


Figure: Performance comparison w.r.t running time for Heat equation in 1+1 D:

$$\begin{cases} \partial_t u - \frac{1}{4} \partial_{xx} u = 0 & \text{in } [0, 1]^2 \\ u = 0 & \text{on } [0, 1] \times \{0, 1\} \\ u = \sin(\pi x) & \text{on } \{0\} \times [0, 1] \end{cases}$$

5 D Laplace equation

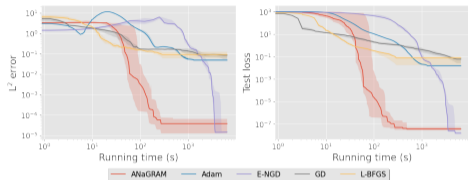


Figure: Performance comparison w.r.t running time for Laplace equation in 5 D:

$$\begin{cases} \Delta u = \pi^2 \sum_{k=1}^5 \sin(\pi x_k) & \text{in } \Omega = [0, 1]^5 \\ u = \sum_{k=1}^5 \sin(\pi x_k) & \text{on } \partial\Omega \end{cases} \quad \begin{cases} \partial_t u - 10^{-3} \partial_{xx} u = 5(u - u^3) & \text{in } \Omega = [0, 1] \times [-1, 1] \\ u = -1 & \text{on } \partial\Omega_b = [0, 1] \times \{-1, 1\} \\ u(0, x) = x^2 \cos(\pi x) & \text{on } \partial\Omega_0 = \{0\} \times [-1, 1] \end{cases}$$

1+1 D Allen-Cahn equation

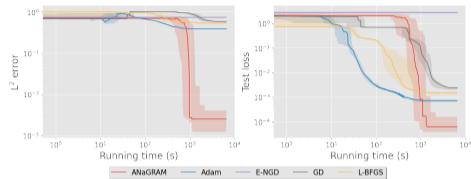
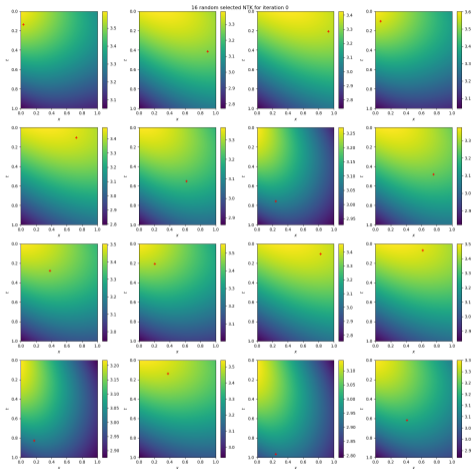
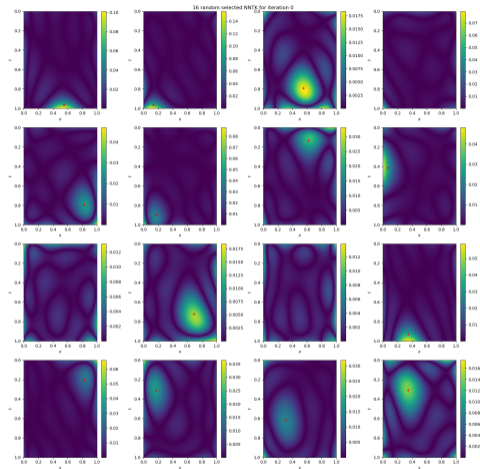


Figure: Performance comparison w.r.t running time for Allen-Cahn equation in 1+1 D:

NTK vs NNTK of PINNs



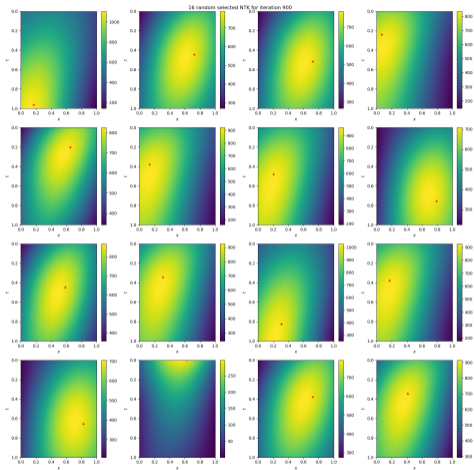
(a) NTK



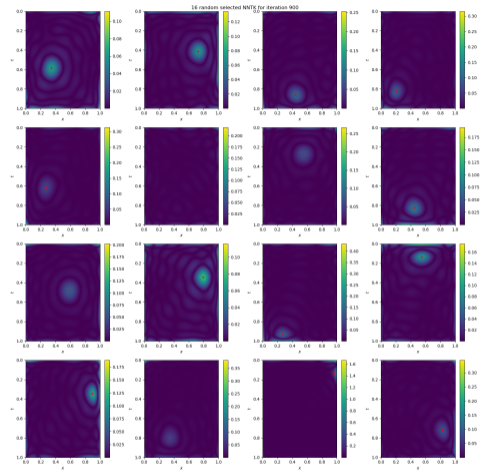
(b) NNTK

Figure: Comparison of NTK and NNTK at **initialization** for Heat equation in 1+1D

NTK vs NNTK of PINNs



(a) NTK



(b) NNTK

Figure: Comparison of NTK and NNTK at the **end of optimization** for Heat equation in 1+1D

Conclusion and Perspectives

Conclusions

- Anagram gives a theoretically founded simplification to any natural-gradient algorithm lowering the complexity from $O(P^3)$ to $O(\min(PN^2, P^2N))$, which is above stochastic gradient descent only by a factor $\min(P, N)$.
- In the case of PINNs, we prove that natural gradient correspond to an optimal linear update following the Green's function.
- Empirical results are improved by several orders of magnitude.
- The SVD cut-off factor appears to be a pivotal hyper-parameter of the algorithm.

Perspectives

- Design of an optimal collocation points procedure, coupled with SVD cut-off factor adaptation strategy.
- Establish theoretical connections with classical algorithms, such as FEMs, FDMs, *etc.*
- Include data assimilation in this theoretical setting, and understand its regularizing effect.
- Include common optimization techniques (e.g. Momentum)
- Extend it to Operator learning
- Application to HJB equation and control

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. FURTLER (2025): “ANaGRAM: A Natural Gradient Relative to Adapted Model for Efficient PINNs Learning,” in *The Thirteenth International Conference on Learning Representations*.

Illustration of natural gradient dynamics

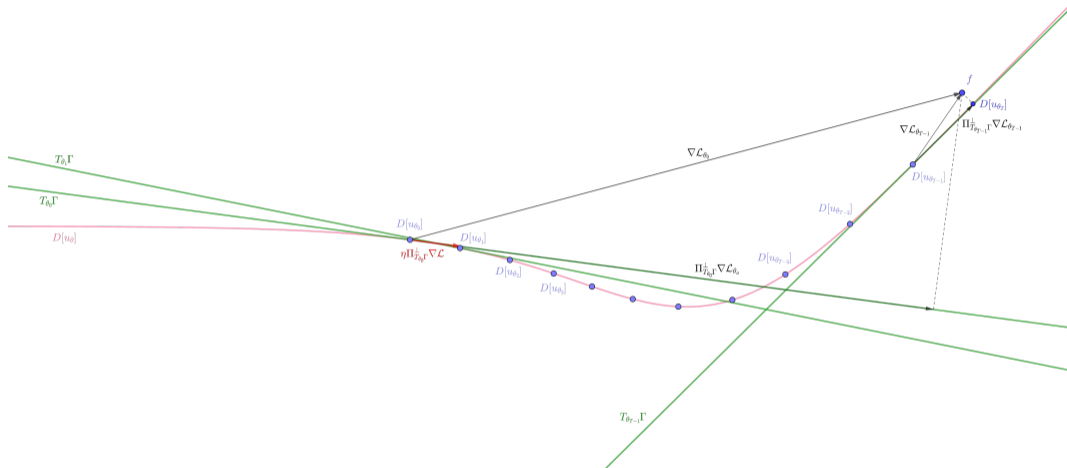


Figure: Illustration of PINNs learning process under natural gradient, as successive applications of Green's function

Remark

$$(x_i)^* = \operatorname{argmin}_{(x_i) \in \Omega^S} \left\| \Pi_{\hat{T}_{\theta, K}^{\perp}(x_i) \mathcal{M}} \nabla \mathcal{L}|_{u|_{\theta_t}} - \nabla \mathcal{L}|_{u|_{\theta_t}} \right\|_{L^2} = \operatorname{argmin}_{(x_i) \in \Omega^S} \inf_{\alpha \in \mathbb{R}^S} \left\| \sum_{i=1}^S \alpha_i K_{\theta}(x_i, \cdot) - \nabla \mathcal{L}|_{u|_{\theta_t}} \right\|_{L^2}^2$$

Consequence

$(x_i)^*$ can be “learned” by the minimization through natural gradient descent of

$$u : \begin{cases} \Omega^S \times \mathbb{R}^S & \rightarrow L^2(\Omega \rightarrow \mathbb{R}, \mu) \\ ((x_i), \alpha) & \mapsto \sum_{i=1}^S \alpha_i K_{\theta}(x_i, \cdot) \end{cases}$$

Even better: closed form formulas exist !

Proposition

- $\langle \partial_{\alpha_i} u_{\theta}, \partial_{\alpha_j} u_{\theta} \rangle = K_{\theta}(x_i, x_j)$
- $\langle \partial_{x_i} u_{\theta}, \partial_{\alpha_j} u_{\theta} \rangle = \alpha_j \partial_1 K_{\theta}(x_i, x_j)$
- $\langle \partial_{x_i} u_{\theta}, \partial_{x_j} u_{\theta} \rangle = \alpha_i \partial_2 \partial_1 K_{\theta}(x_i, x_j) \alpha_j$
- $\langle \partial_{\alpha_i} u_{\theta}, \nabla \mathcal{L} \rangle = \Pi_{T_{\theta} \mathcal{M}} \nabla \mathcal{L}(x_i) \simeq \nabla \mathcal{L}(x_i)$
- $\langle \partial_{x_i} u_{\theta}, \nabla \mathcal{L} \rangle = \alpha_i \Pi_{T_{\theta} \mathcal{M}} \nabla \mathcal{L}'(x_i) \simeq \alpha_i \nabla \mathcal{L}'(x_i)$

First results for collocation learning in Fourier space

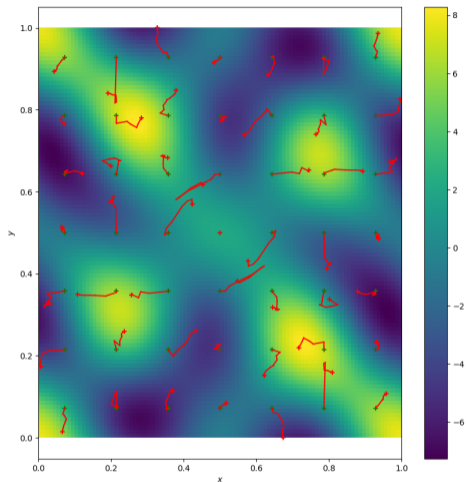


Figure: Points learning dynamic

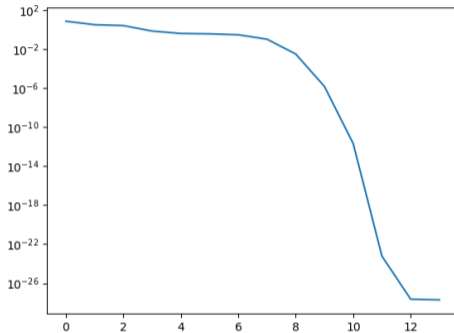


Figure: $\left\| \Pi_{\hat{\mathcal{T}}_{\theta, K}^{\perp}(x_i)} \nabla \mathcal{L}|_{u|_{\theta_t}} - \nabla \mathcal{L}|_{u|_{\theta_t}} \right\|_{L^2}$ wrt (x_i)
learning steps

In the population limit, the natural gradient of PINNs is the update:

$$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \eta \text{d}((D, B) \circ u)_{|\boldsymbol{\theta}_t}^\dagger \left(\Pi_{T_{\boldsymbol{\theta}_t} \Gamma}^\perp \nabla \mathcal{L}_{|u_{|\boldsymbol{\theta}_t}} \right)$$

Corollary

The kernel of $\Pi_{T_{\boldsymbol{\theta}} \Gamma}$ is: for all $x, y \in (\Omega \times \partial\Omega)^2$

$$\begin{aligned} \text{NNTK}_{\boldsymbol{\theta}}(x, y) &= \sum_{1 \leq p, q \leq P} \partial_p(D, B)[u_{|\boldsymbol{\theta}}](x) G_{\boldsymbol{\theta}_{p,q}}^\dagger \partial_q(D, B)[u_{|\boldsymbol{\theta}}](y) \\ &= \sum_{1 \leq p, q \leq P} (\partial_p D[u_{|\boldsymbol{\theta}}](x_1), \partial_p B[u_{|\boldsymbol{\theta}}](x_2)) G_{\boldsymbol{\theta}_{p,q}}^\dagger (\partial_q D[u_{|\boldsymbol{\theta}}](y_1), \partial_q B[u_{|\boldsymbol{\theta}}](y_2)), \end{aligned}$$

where for all $1 \leq p, q \leq P$

$$\begin{aligned} G_{\boldsymbol{\theta}_{p,q}} &:= \langle \partial_p(D, B)[u_{|\boldsymbol{\theta}}], \partial_q(D, B)[u_{|\boldsymbol{\theta}}] \rangle_{L^2(\Omega \rightarrow \mathbb{R}, \mu) \times L^2(\partial\Omega \rightarrow \mathbb{R}, \sigma)} \\ &= \langle \partial_p D[u_{|\boldsymbol{\theta}}], \partial_q D[u_{|\boldsymbol{\theta}}] \rangle_{L^2(\Omega \rightarrow \mathbb{R}, \mu)} + \langle \partial_p B[u_{|\boldsymbol{\theta}}], \partial_q B[u_{|\boldsymbol{\theta}}] \rangle_{L^2(\partial\Omega \rightarrow \mathbb{R}, \sigma)}. \end{aligned}$$

Corollary

The kernel of $\Pi_{\hat{T}_{\theta}\Gamma}$ is: for all $x, y \in (\Omega \times \partial\Omega)^2$

$$\hat{k}(x, y) = \sum_{1 \leq i, j \leq S} NNTK_{\theta}(x, x_i) \hat{G}_{\theta}^{\dagger} NNTK_{\theta}(x_j, y), \text{ where}$$

$$G_{\theta_{i,j}} := \langle NNTK_{\theta}(\cdot, x_i), NNTK_{\theta}(x_j, \cdot) \rangle_{L^2(\Omega \rightarrow \mathbb{R}, \mu) \times L^2(\partial\Omega \rightarrow \mathbb{R}, \sigma)} = NNTK_{\theta}(x_i, x_j)$$

Theorem (ANaGRAM for PINNs)

Under mild assumptions, the empirical natural gradient update:

$$\theta_{t+1} \leftarrow \theta_t - \eta d((D, B) \circ u)_{|\theta_t}^{\dagger} \left(\Pi_{\hat{T}_{\theta_t}\Gamma}^{\perp} \nabla \mathcal{L}|_{u|\theta_t} \right),$$

does not require to estimate a Gram matrix. More precisely, we have:

$$d((D, B) \circ u)_{|\theta_t}^{\dagger} \left(\Pi_{\hat{T}_{\theta_t}\Gamma}^{\perp} \nabla \mathcal{L}|_{u|\theta_t} \right) = \hat{\phi}_{\theta_t}^{\dagger} \widehat{\nabla \mathcal{L}}_{\theta_t},$$

where: for all $1 \leq p \leq P, 1 \leq i \leq S$

- $\hat{\phi}_{\theta_{ti}, p} := (\partial_p D[u|_{\theta_t}](x_{i1}), \partial_p B[u|_{\theta_t}](x_{i2}))$
- $\widehat{\nabla \mathcal{L}}_{\theta_{ti}} := \nabla \mathcal{L}|_{u|\theta_t}(x_i)$

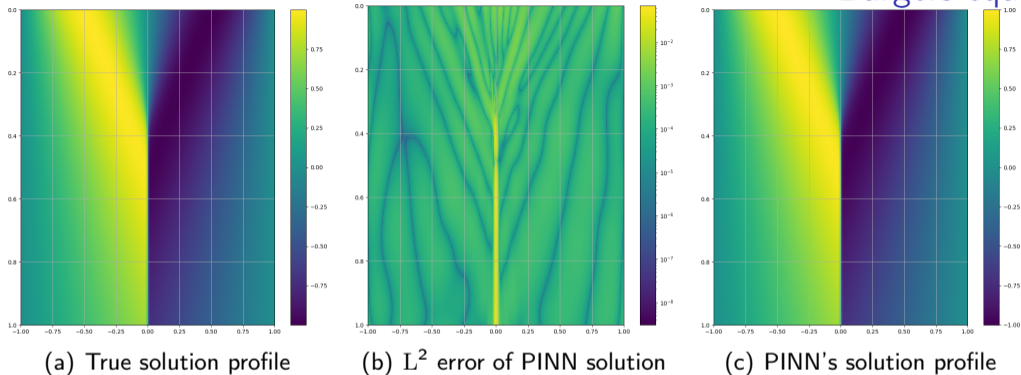


Figure: PINN's solution under Adam (15k steps) + L-BFGS (15k) steps for Burgers equation in 1+1 D:

$$\begin{cases} \partial_t u + u \partial_x u = \nu \partial_{xx} u & \text{in } [0, 1] \times [-1, 1] \\ u = 0 & \text{on } [0, 1] \times \{-1, 1\} \\ u = -\sin(\pi x) & \text{on } \{0\} \times [-1, 1] \end{cases}$$

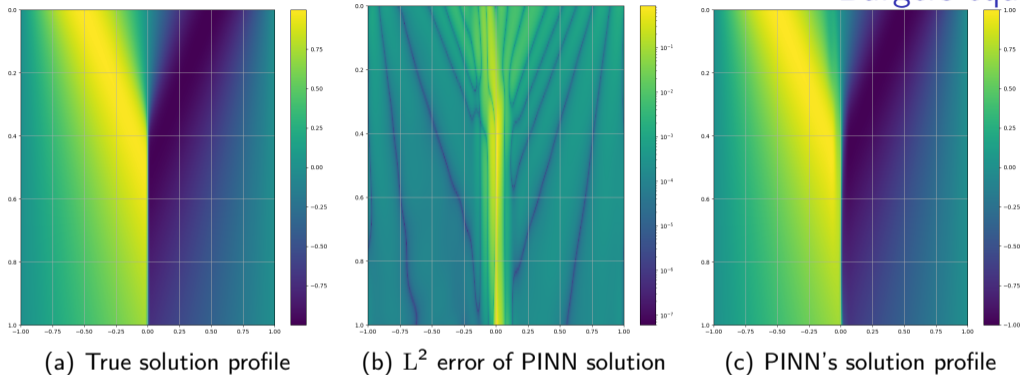


Figure: PINN's solution under Anagram (500 steps) for Burgers equation in 1+1 D:

$$\begin{cases} \partial_t u + u \partial_x u = \nu \partial_{xx} u & \text{in } [0, 1] \times [-1, 1] \\ u = 0 & \text{on } [0, 1] \times \{-1, 1\} \\ u = -\sin(\pi x) & \text{on } \{0\} \times [-1, 1] \end{cases}$$