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

Nilo Schwencke [1], Cyril Furtlehner [1]

Abstract: Physics Informed Neural Networks (PINNs) have gained significant interest for solving PDE-driven systems, particularly for data assimilation, but still face many challenges. This paper introduces an approach that enhances the speed and accuracy of PINNs training with a new natural gradient scaling as $\min(P^2S, S^2P)$, where P is the number of parameters and S is the batch size. In addition, it shows an explicit connection with Green's function theory.

Physics Informed Neural Networks

Given a Partial Differential and Boundary Operator:

$$D : \begin{cases} \mathcal{H} & \to & \mathrm{L}^2(\Omega \to \mathbb{R}, \mu) \\ u & \mapsto & D[u] \end{cases}, B : \begin{cases} \mathcal{H} & \to & \mathrm{L}^2(\partial\Omega \to \mathbb{R}, \sigma) \\ u & \mapsto & B[u] \end{cases},$$

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

PINNs key idea: Optimize a neural network $u_{|\theta}$ on the loss

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

Problem: **This leads to low accuracy when using SGD.**

Neural Tangent Kernel (NTK)

Jacot et al. show that for an empirical quadratic loss: $\ell(\theta) := \frac{1}{2}\sum_{i=1}^{S}(u_\theta(x_i) - y_i)^2$,

the functional dynamic of the gradient descent on ℓ can be described by:

$$\frac{\mathrm{d}u_{\theta_t}}{\mathrm{d}t}(x) = -\sum_{i=1}^{S} NTK_{\theta_t}(x, x_i)(u_{|\,\theta_t}(x_i) - y_i),$$

with: $NTK_{\theta_t}(x,y) := \sum_{p=1}^{P}\left(\partial_p u_{|\,\theta_t}(x)\right)(\partial_p u_{|\,\theta_t}(y))^T$.

Natural Gradient

Given the functional quadratic loss: $\mathcal{L} : v \in \mathrm{L}^2(\Omega) \mapsto \frac{1}{2}\,\|v - f\,\|^2_{\mathrm{L}^2(\Omega)}$,

whose gradient is : $\nabla\mathcal{L}_{|\,v} = v - f$, the natural gradient update is given by:

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

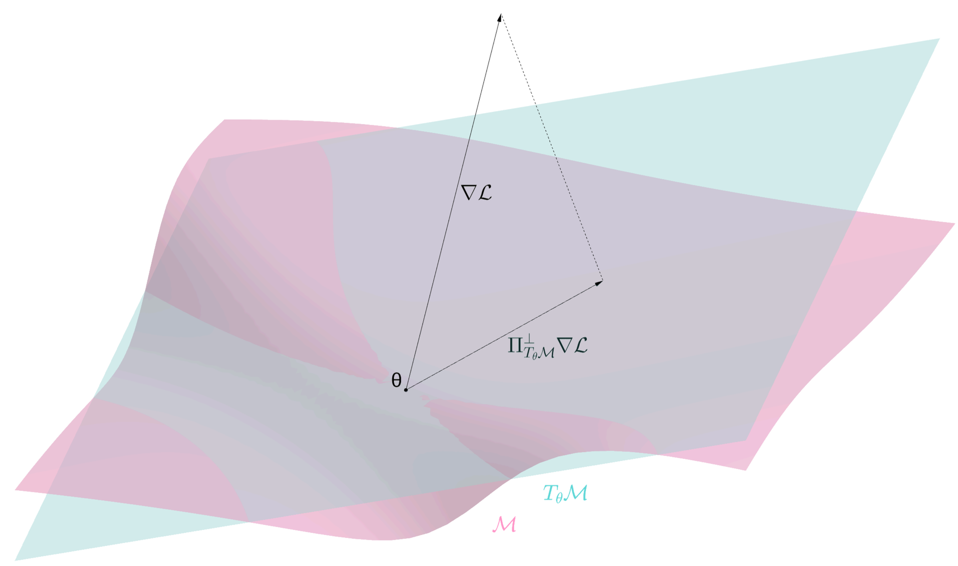


Illustration of Natural Gradient in functional space

Rudner et al. show that functional dynamics then follow the Natural NTK (NNTK):

$$\frac{\mathrm{d}u_{\theta_t}}{\mathrm{d}t}(x) = -\sum_{i=1}^{S} NNTK_{\theta_t}(x, x_i)(u_{|\,\theta_t}(x_i) - y_i), \quad \text{with :}$$

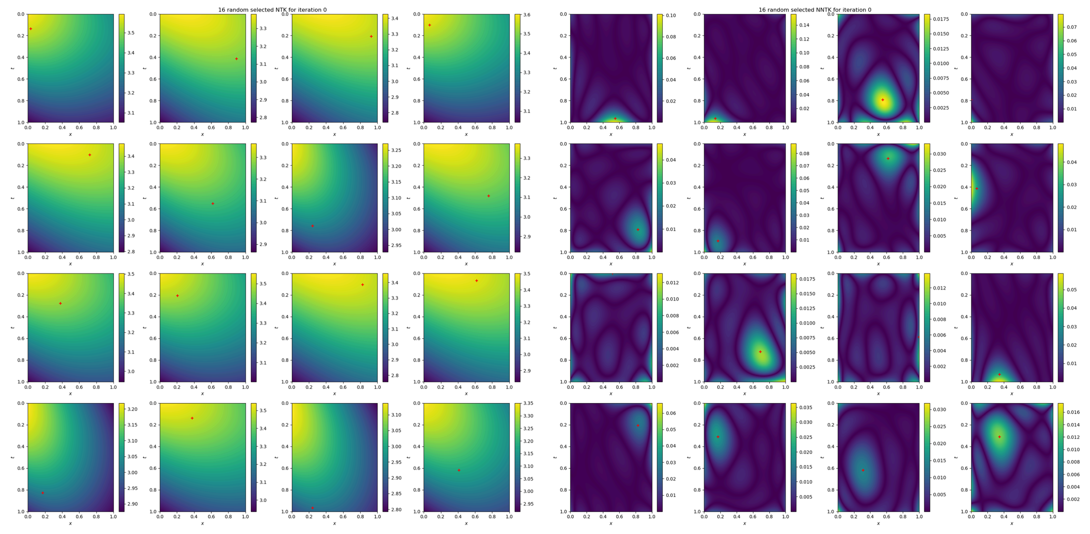
$$NNTK_{\boldsymbol{\theta}}(x,y) := \sum_{1\leq p,q\leq P}\left(\partial_p u_{|\,\boldsymbol{\theta}}(x)\right)G^{\dagger}_{\boldsymbol{\theta}_{pq}}\left(\partial_q u_{|\,\boldsymbol{\theta}}(y)\right)^t; \quad G_{\boldsymbol{\theta}_{\boldsymbol{p},\boldsymbol{q}}} := \left\langle \partial_p u_{|\,\boldsymbol{\theta}}\,,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

.

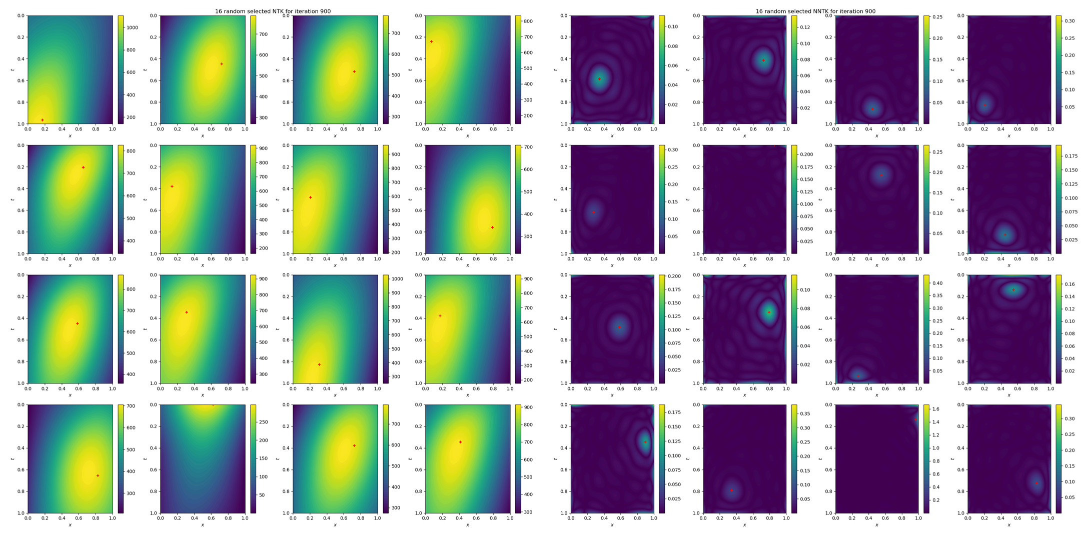
Lemma $\Pi^{\perp}_{T_{\boldsymbol{\theta}_t}\mathcal{M}}$ is an integral operator whose kernel is $NNTK_{\theta_t}$.

In particular the natural gradient update rewrites:

$$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \eta\, G^{\dagger}_{\boldsymbol{\theta}_t}\nabla_{\boldsymbol{\theta}_t}\mathcal{L}(u_{|\,\boldsymbol{\theta}_t})$$



Comparison of NTK and NNTK at initialization for Heat equation in (1+1) D.
Reading: the red cross being x_i, the plot represents the function $(N)NTK_{\boldsymbol{\theta}_0}(x_i, \cdot)$



Comparison of final NTK and NNTK for Heat equation in (1+1) D.
Reading: the red cross being x_i, the plot represents the function $(N)NTK_{\boldsymbol{\theta}_{\mathrm{end}}}(x_i, \cdot)$

Empirical Natural Gradient

We introduce the **empirical Natural Gradient** as the projection of $\nabla\mathcal{L}_{|\,u_{|\,\boldsymbol{\theta}_t}}$ onto the **empirical Tangent Space**:

$$\hat{T}^{NNTK}_{\boldsymbol{\theta},(x_i)}\mathcal{M} := \mathrm{Span}(NNTK_{\boldsymbol{\theta}}(\cdot, x_i) : (x_i)_{1\leq i\leq S}) \subset T_{\boldsymbol{\theta}}\mathcal{M},$$

where $(x_i)_{1\leq i\leq S}$ are given collocation points.

Theorem empirical Natural Gradient is accurately approximated by:

$$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \eta\,\hat{\phi}^{\dagger}_{\boldsymbol{\theta}}\,\widehat{\nabla\mathcal{L}}_{|\,u_{|\,\boldsymbol{\theta}}}; \quad \hat{\phi}_{\boldsymbol{\theta}_{i,p}} := \partial_p u_{|\,\boldsymbol{\theta}}(x_i)\,; \quad \widehat{\nabla\mathcal{L}}_{|\,u_{|\,\boldsymbol{\theta}_i}} := \nabla\mathcal{L}_{|\,u_{|\,\boldsymbol{\theta}}}(x_i),$$

resulting in $O(\min(P^2S, S^2P))$ computational complexity.

PINNs are a quadratic regression

Applying our algorithm to PINNs only requires to notice that PINNs are a quadratic regression problem, using the model:

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

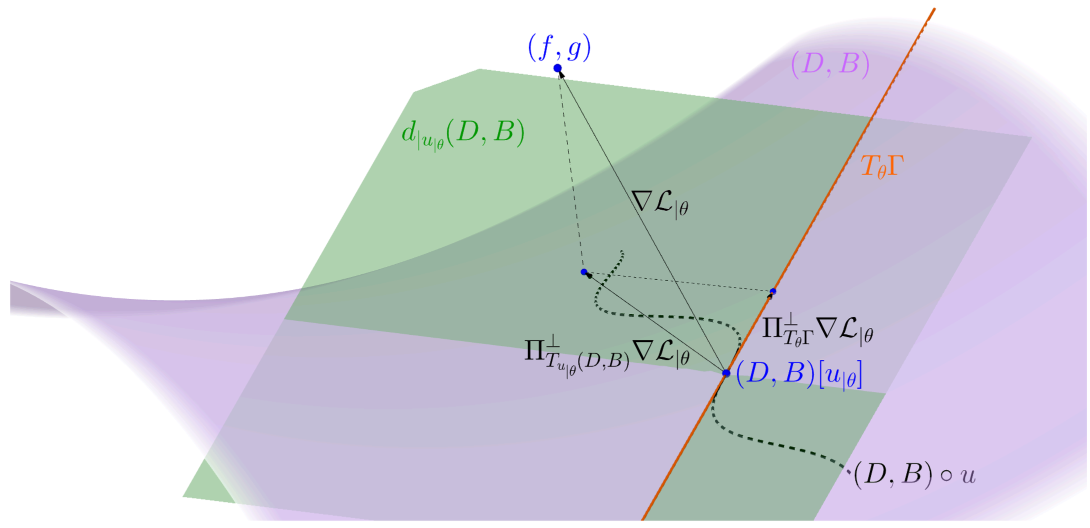


Illustration of Natural Gradient of PINNs

Connection to Green's function theory

Definition A generalized Green's function is any kernel function $g : \Omega\times\Omega \to \mathbb{R}$ 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: $D\circ R = \Pi^{\perp}_{D[\mathcal{H}_0]}$

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

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

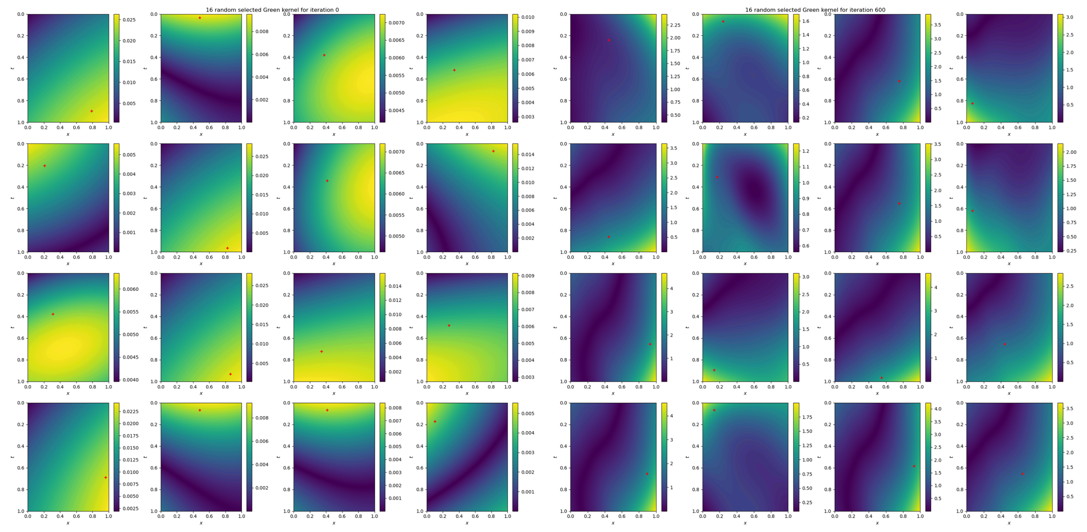
with for all $1\leq p,q\leq P$

$$G_{pq} := \left\langle \partial_p D[u_{|\,\boldsymbol{\theta}}]\,, \partial_q D[u_{|\,\boldsymbol{\theta}}]\right\rangle_{\mathrm{L}^2(\Omega\to\mathbb{R},\mu)}.$$

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

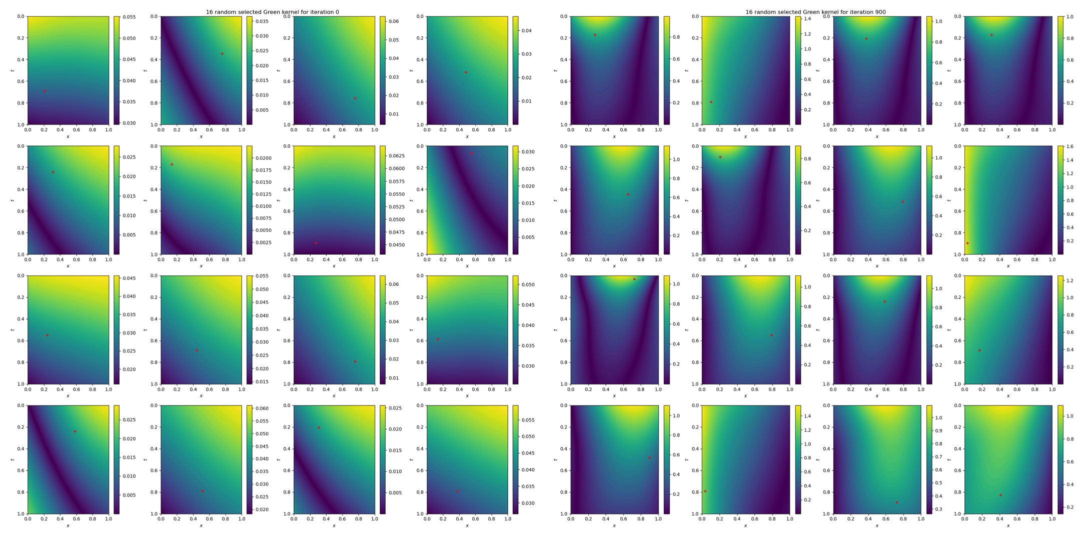
$$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \eta\,\mathrm{d}u^{\dagger}_{|\,\boldsymbol{\theta}_t}\left(x\in\Omega \mapsto \int_\Omega g_{T_{\boldsymbol{\theta}_t}\mathcal{M}}(x,y)\nabla\mathcal{L}_{|\,\boldsymbol{\theta}_t}(y)\mu(\mathrm{d}y)\right),$$

Interpretation: **Following the natural gradient for PINNs boils down to infinitesimally moving in the direction of the solution to the linearized equation.**



(a) Initialization (b) Final

Generalized Green's functions for Laplace equation in 2 D.
Reading: the red cross being x_i, the plot represents the function $g_{T_{\boldsymbol{\theta}_0}\mathcal{M}}(x_i, \cdot)$



(a) Initialization (b) Final

Generalized Green's functions for Heat equation in (1+1) D
Reading: the red cross being x_i, the plot represents the function $g_{T_{\boldsymbol{\theta}_{\mathrm{end}}}\mathcal{M}}(x_i, \cdot)$

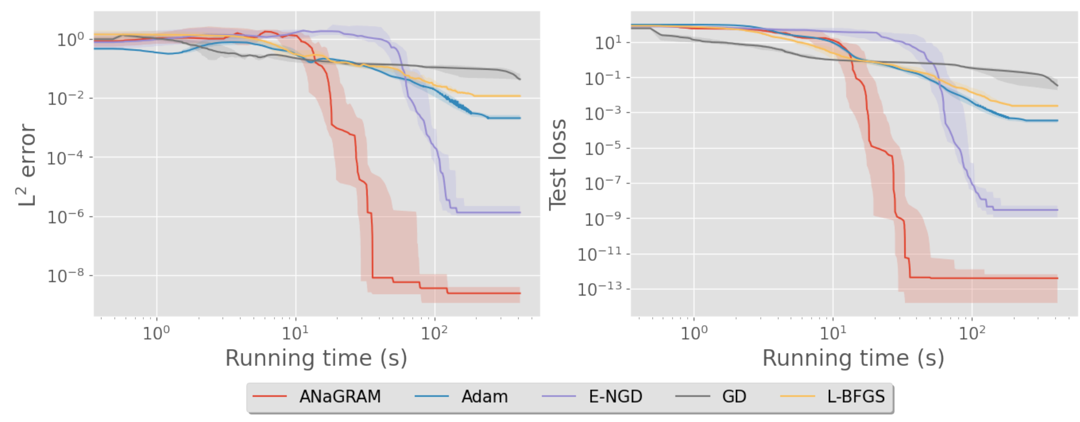
Collocation points selection criterion

One of the pleasant byproducts of the empirical natural gradient formulation is that it gives a straightforward criterion for collocation points selection:

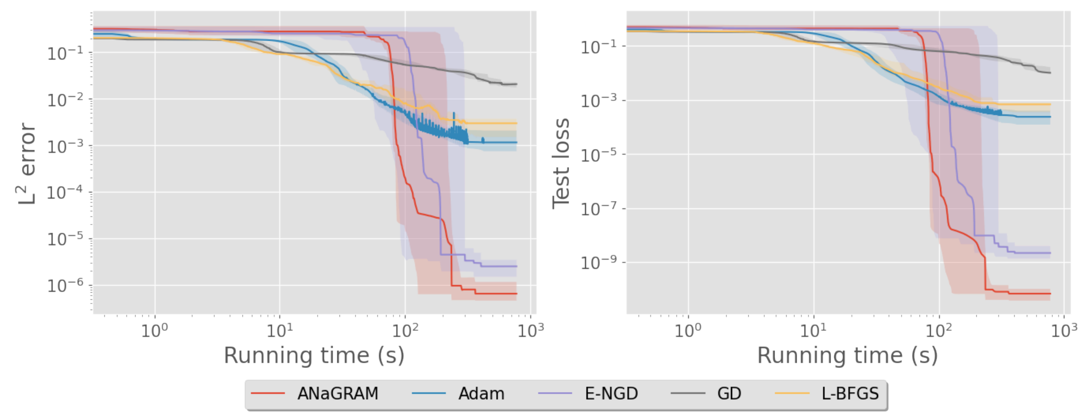
$$(x_i^*) := \operatorname*{argmin}_{(x_i)\in\Omega^S}\left\|\Pi^{\perp}_{\mathrm{Span}(NNTK_{\boldsymbol{\theta}}(x_i,\cdot):1\leq i\leq S)}\left(\nabla\mathcal{L}_{|\,u_{|\,\theta}}\right) - \nabla\mathcal{L}_{|\,u_{|\,\theta}}\right\|^2_{\mathcal{H}}.$$

Making the most of this criterion is currently the subject of an in-depth study.

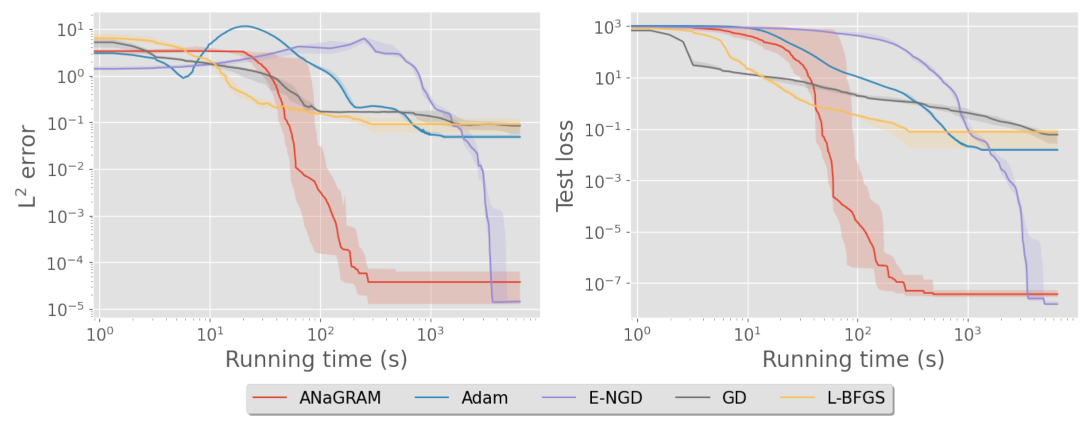
Results



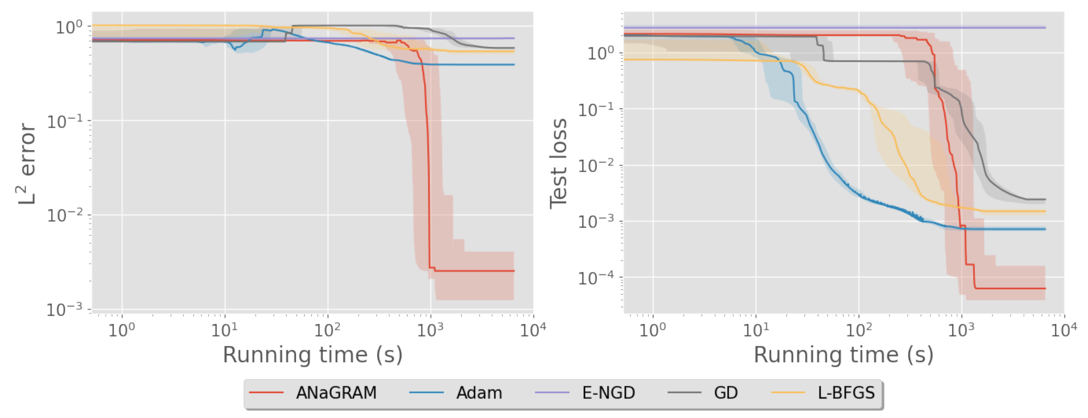
Median L^2 errors and Test losses for Laplace equation in 2 D



Median L^2 errors and Test losses for Heat equation in (1+1) D



Median L^2 errors and Test losses for Laplace equation in 5 D



Median L^2 errors and Test losses for Allen-Cahn equation in (1+1) D

References

- Müller, J., and Zeinhofer, M. (2024). Position: Optimization in SciML Should Employ the Function Space Geometry. In Forty-First International Conference on Machine Learning.
- Müller, Johannes, and Marius Zeinhofer. "Achieving high accuracy with PINNs via energy natural gradient descent." International Conference on Machine Learning. PMLR, 2023.
- Rudner, Tim GJ, et al. "The natural neural tangent kernel: Neural network training dynamics under natural gradient descent." 4th workshop on Bayesian Deep Learning (NeurIPS 2019). 2019.
- Bai, Qinxun, Steven Rosenberg, and Wei Xu. "A geometric understanding of natural gradient." arXiv preprint arXiv:2202.06232 (2022).
- Cuomo, Salvatore, et al. "Scientific machine learning through physics–informed neural networks: Where we are and what's next." Journal of Scientific Computing 92.3 (2022): 88.
- Raissi, M., Perdikaris, P., and Karniadakis, G. E. (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.
- Jacot, Arthur, Franck Gabriel, and Clément Hongler. "Neural tangent kernel: Convergence and generalization in neural networks." Advances in neural information processing systems 31 (2018).
- Wang, Sifan, Xinling Yu, and Paris Perdikaris. "When and why PINNs fail to train: A neural tangent kernel perspective." Journal of Computational Physics 449 (2022): 110768.
- Paulsen, Vern I., and Mrinal Raghupathi. An introduction to the theory of reproducing kernel Hilbert spaces. Vol. 152. Cambridge university press, 2016.

[1] Université Paris-Saclay, LISN-A&O, INRIA-TAU. This work has been supported by the french ANR grant Scalp (ANR-24-CE23-1320). Contact : nilo.schwencke@protonmail.com