Adaptive Sampling in PINNs through the lens of Kernel Theory

CANUM 2026 — Saint-Jacut-de-la-Mer

Nilo Schwencke

joint works with Sheng Wan, Cyriaque Rousselot, Shorouk El-Hassanieh, Alena Shilova and Cyril Furtlehner

LIP – ENS Lyon • INRIA Lyon • CNRS

École polytechnique • University of Bonn • ENS Paris-Saclay • INRIA Saclay – TAU Team

June 04, 2026

Adaptive sampling strategies in PINNs

PINN loss on collocation points

$$\widehat{\mathcal{L}}(\theta) := \frac{1}{S}\sum_{i=1}^{S} |r_\theta(x)|^2 + \text{BC/IC},$$

$$r_\theta(x) := D[u_\theta](x) - f(x).$$

- The training set is a set of **collocation points**.
- Sampling changes the loss landscape seen by the optimizer.
- Adaptive methods often outperform static sampling.

Question

Can these methods be organized by a common principle ? Do they work for natural gradient ?

Examples of adaptive samplers

- **RAR**: add the points with largest residual (Lu et al.2021)

$$x_{\text{new}} = \arg\max_x |r_\theta(x)|$$

- **RAD / RAR-D**: sample according to the residual distribution (Wu et al.2023)

$$p(x) \propto |r_\theta(x)|^\alpha$$

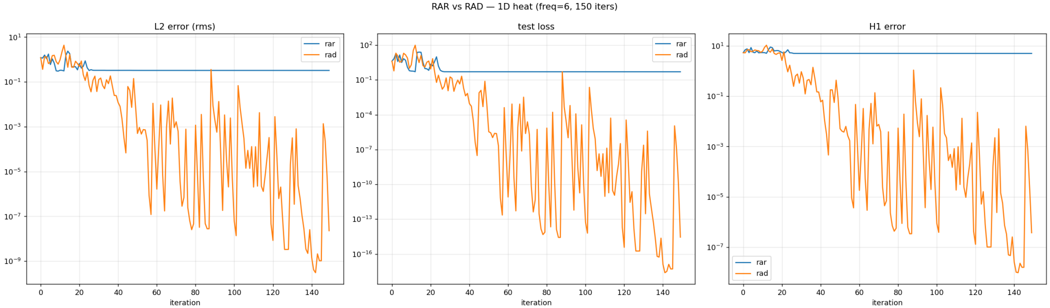
- **R3**: retain high-residual points, resample the others (Daw et al.2023)

$$\mathcal{X} = \mathcal{X}_{\text{retain}} \cup \mathcal{X}_{\text{resample}}$$

- **PINNACLE**: NTK-based convergence-degree sampling (Lau et al.2024)

$$p(x) \propto \widetilde{\Pi}_{T_\theta \mathcal{M}} r_\theta$$

Literature baselines: RAR vs RAD



Same fixed budget. **RAR** (deterministic, largest residual) *over-concentrates* and stalls at ≈ 0.3 — worse than no resampling. **RAD** ($p \propto |r_\theta|^\alpha$) keeps spatial spread and converges to $\approx 3 \cdot 10^{-10}$.

Kernel theory viewpoint

What a kernel encompasses

- The topology:

$$\left\langle \sum_i \alpha_i k(x_i, \cdot) \, , \, \sum_j \beta_j k(y_j, \cdot) \right\rangle_k := \sum_{i,j} \alpha_i \beta_j k(x_i, y_j).$$

- The algebraic information:

$$\mathcal{H}(k) := \overline{\mathrm{Span}\left(k(x, \cdot) \; : \; x \in \Omega\right)}$$

Classical problem

Given a kernel k and a target f, choose points

$$X = \{x_1, \dots, x_S\}$$

to reconstruct f by an interpolant

$$\Pi_X f \in \mathrm{Span}\{k(x_i, \cdot)\}_{i=1}^S.$$

Three type of informations

- The target information (f);
- The representation power of $\mathcal{H}(k)$
- The generalization power of $\mathrm{Span}\{k(x_i, \cdot)\}_{i=1}^S$.

Key point

Kernel interpolation theory already contains **target-dependent sampling** rules.

Three kernel quantities to keep apart

Residual

$$r_X(x) = f(x) - \Pi_X f(x)$$

target error

Power function

$$P_X(x) = \left\| k(\cdot, x) - \Pi_X k(\cdot, x) \right\|_{\mathcal{H}_k}$$

missing direction

Leverage scores

$$\ell_k(x) = \left\| \Pi_X k(\cdot, x) \right\|_{\mathcal{H}_k}$$

correlation to the training data

Interpretation

Residual tells where we are wrong; power tells what is new; leverage tells the dependence to training data.

Classical target-dependent greedy rules

Sampling rules

$$f\text{-greedy:} \quad x_{n+1} = \operatorname*{argmax}_x |r_X(x)|,$$

$$P\text{-greedy:} \quad x_{n+1} = \operatorname*{argmax}_x P_X(x),$$

$$fP\text{-greedy:} \quad x_{n+1} = \operatorname*{argmax}_x |r_X(x)| P_X(x),$$

$$fP\text{-greedy:} \quad x_{n+1} = \operatorname*{argmax}_x \frac{|r_X(x)|}{P_X(x)}.$$

One-parameter family (Wenzel et al.2023)

$$x_{n+1} = \operatorname*{argmax}_x \ |r_n(x)|^{\beta} \, P_n(x)^{\,1-\beta}, \quad \beta \in [0, \infty]$$

$\beta = 0 :$ P-greedy $\quad \frac{1}{2} :$ $f{\cdot}P \quad 1 :$ $f \quad \infty :$ f/P

Meaning

- f-greedy: target-driven refinement.
- P-greedy: geometric exploration.
- fP,f/P-greedy: target, geometry balance.

What theory guarantees

- P-greedy: near-optimal (Sobolev), points (Santin and Haasdonk2017).
- target-dependence speeds the rate by $n^{-\beta/2}$ (Wenzel et al.2023).
- f-greedy is **optimal** (best n-term) in Sobolev RKHS (Santin et al.2024).

ANaGRAM as an empirical-kernel solver

ANaGRAM makes the kernel explicit

Empirical residual features

$$\widehat{\Phi}_\theta := \left(\partial_p \left[D(u_\theta)(x_i)\right]\right)_{1\leq p\leq P, 1\leq i\leq S}.$$

$$\widehat{\Phi}_\theta = U\Lambda V^\top.$$

Empirical NTK

$$K_\theta = \widehat{\Phi}_\theta^\top \widehat{\Phi}_\theta.$$

The natural-gradient step solves

$$\widehat{\Phi}_\theta^\top \widehat{\Phi}_\theta \alpha = \widehat{r_\theta}; \quad \widehat{r}_{\theta_i} := r_\theta(x_i).$$

Consequence

One ANaGRAM step is a **kernel regression of the PDE residual** in the empirical NTK.

Truncated natural-gradient step

$$\Delta_k = U_k \Lambda_k^{-1} V_k^\top \widehat{r}_\theta.$$

Independent reconstruction criterion

On a fresh extrapolation set Z:

$$\mathrm{RCE}(k)^2 = \left\| r_Z - \Phi_Z^\top \Delta_k \right\|_{L^2(Z)}^2, \qquad k_{\mathrm{opt}} = \operatorname*{argmin}_k \mathrm{RCE}(k).$$

Role of k_{opt}

The same spectral regularization is used for all sampling strategies; only the collocation set changes.

New sampling strategies ideas

leverage-RAR: geometry-aware residual refinement

Rule (per operator block, each step)

With $\widehat{\Phi}_\theta = U \Lambda V^\top$ and leverage

$$h_i = \sum_{k \le k_{\mathrm{opt}}} V_{k,i}^2 = (V_{k_{\mathrm{opt}}} V_{k_{\mathrm{opt}}}^\top)_{ii} \in [0,1]$$

- **drop** x_i if $\sqrt{h_i\, S / k_{\mathrm{opt}}} < u, \;\; u \sim \mathcal{U}(0,1)$ ($S = \#\text{pts}$);
- **insert** the candidates of largest residual $|r_\theta(y)|$ (RAR).

Kernel reading

- h_i : how much is the point x_i important for approximation ?
- low leverage $\Rightarrow$ redundant $\Rightarrow$ safe to drop.

$\simeq$ leverage sampling $\times$ RAR

- prune points the kernel *over-represents* (geometry / actionability);
- add points where the PDE is *most violated* (target / residual).

Why it is robust

Near convergence k_{opt} shrinks and leverages saturate $\Rightarrow$ few swaps: it *self-throttles* (no RAR-style over-concentration);

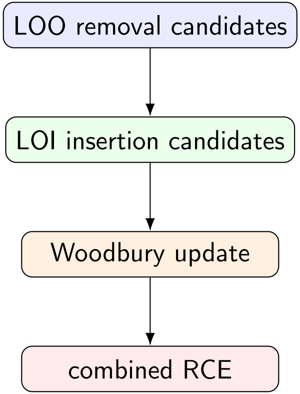
Greedy Woodbury exchange

Problem with marginal scores

- A point can be good to add alone.
- Another point can be harmless to remove alone.
- But several exchanges interact through the update.

Exchange score

$$S(i, y) = \|r_Z\|^2 - \|r_Z - \Phi_Z^\top (d_i + d_y)\|^2.$$



Idea

Accept exchanges one by one, and stop when the combined extrapolation RCE stops decreasing.

First results

Benchmark

- 1D heat equation.
- Frequency 6.
- 150 iterations.
- GPU implementation.
- Same ANaGRAM update for all methods.

Metric

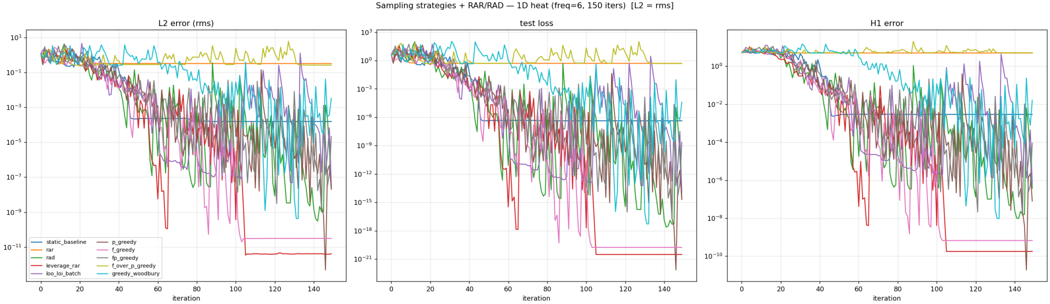
$$\|u_\theta - u^\star\|_{L^2}$$

- best: lowest error reached;
- final: error at iteration 150.

Important convention

Because of banded updates, the best error is often more informative than the final error.

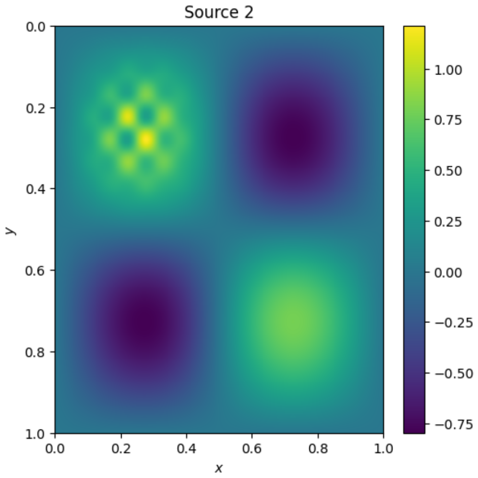
Error trajectories across strategies



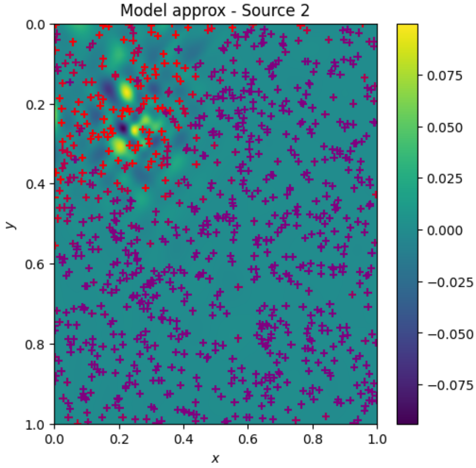
L^2 is the rms error over the eval grid. RAR, RAD,leverage-RAR and f-greedy descend and *stay*; P/fP/Woodbury dive then oscillate up; f/P stays high.

Colegram : *Collocation points* selection

Leverage scores after training



(a) Target function



(b) Leverage score based clustering

Remark

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

Consequence

$(x_i)^\star$ can be "learned" by the minimization through natural gradient descent of

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

Even better: closed form formulas exist !

Proposition

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

First results: collocation learning in Fourier space

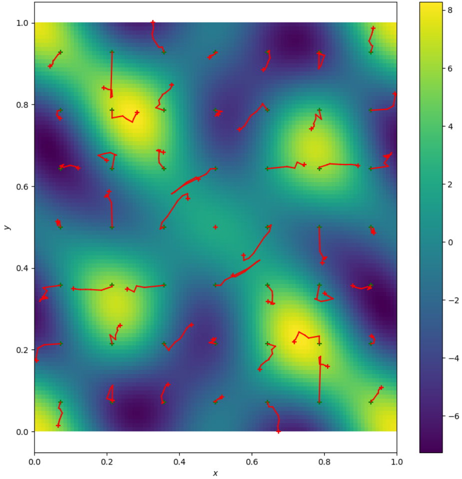


Figure: Points learning dynamic

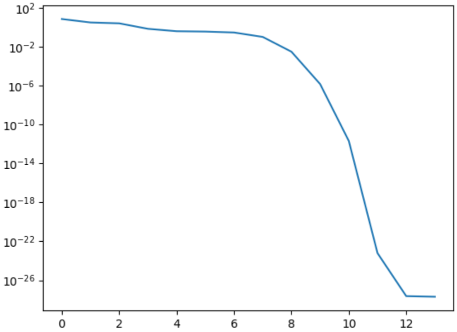


Figure: $\left\| \Pi^{\perp}_{\widehat{T}^{(x_i)}_{\theta,K}\mathcal{M}} \nabla \mathcal{L}_{|u_{\theta_t}} - \nabla \mathcal{L}_{|u_{\theta_t}} \right\|_{L^2}$ wrt (x_i) learning steps

Takeaways

1. **Adaptive PINN sampling = target-dependent point selection in the empirical NTK** — RAR, RAD, R3 and NTK-sampling become special cases of one kernel principle.
2. **Predictive**: on 1D heat, resampling moves the error by ~ 8 orders; residual + geometry samplers (**leverage-RAR**, f-greedy) win, while pure residual (RAR) *over-concentrates and fails*; exploration is the cure.
3. **Algorithmic**: the view yields new samplers (greedy Woodbury exchange; Pareto over residual–power–leverage); stabilizing them and going beyond heat are next.
4. **Colegram**: should be combined with resampling.

Thank you for your attention ! Questions welcome.

ScimBa

DAW, A., J. BU, S. WANG, P. PERDIKARIS, AND A. KARPATNE (2023): "Mitigating Propagation Failures in Physics-informed Neural Networks Using Retain-Resample-Release (R3) Sampling," Comment: 39 pages, 53 figures, 6 tables.

LAU, G. K. R., A. HEMACHANDRA, S.-K. NG, AND B. K. H. LOW (2024): "PINNACLE: PINN Adaptive ColLocation and Experimental Points Selection," in *The Twelfth International Conference on Learning Representations*.

LU, L., X. MENG, Z. MAO, AND G. E. KARNIADAKIS (2021): "DeepXDE: A Deep Learning Library for Solving Differential Equations," *SIAM Rev.*, 63, 208–228.

SANTIN, G. AND B. HAASDONK (2017): "Convergence Rate of the Data-Independent P-Greedy Algorithm in Kernel-Based Approximation," *Dolomites Research Notes on Approximation*, 10, 68–78.

SANTIN, G., T. WENZEL, AND B. HAASDONK (2024): "On the Optimality of Target-Data-Dependent Kernel Greedy Interpolation in Sobolev Reproducing Kernel Hilbert Spaces," *SIAM J. Numer. Anal.*, 62, 2249–2275.

WENZEL, T., G. SANTIN, AND B. HAASDONK (2023): "Analysis of Target Data-Dependent Greedy Kernel Algorithms: Convergence Rates for f-, $$f \cdot P$$- and f/P-Greedy," *Constr Approx*, 57, 45–74.

WU, C., M. ZHU, Q. TAN, Y. KARTHA, AND L. LU (2023): "A Comprehensive Study of Non-Adaptive and Residual-Based Adaptive Sampling for Physics-Informed Neural Networks," *Computer Methods in Applied Mechanics and Engineering*, 403, 115671.