

Adaptive collocation and residual verification for engineering physics-informed neural networks

Luis Torres

Abstract

Adaptive collocation is often described as placing training points where a physics-informed neural network needs them most. That description hides two distinct operations: estimating a fixed residual objective more efficiently, and deliberately changing the objective to emphasize selected regions. This review derives the distinction using importance sampling, examines the difference between residual magnitude and gradient variance, and develops an analytical hot-spot example to quantify missed-region risk. Applications to wake dynamics, thermal layers, solid mechanics, and interfaces show why sampling must be tied to geometry and engineering quantities of interest. Independent verification, support-preserving sampling, causal time allocation, and complete computational accounting are assembled into a proposed evaluation protocol. The worked probability model connects sampling choices to the risk of missing a critical region.

1. The residual has a measure

Physics-informed neural networks minimize combinations of differential-equation residuals, boundary and initial residuals, and observation errors [1, 2]. A residual average is meaningful only after specifying where the points lie and which measure their average approximates. Uniform spatial sampling estimates a different integral from sampling concentrated near a wall. Uniform samples in a reference coordinate system may not be uniform in physical space when a geometry map has a nonconstant Jacobian.

Let Ω be a physical domain and let $p(x)$ be a normalized target density. For a nonnegative scalar loss $f_\theta(x)$, such as a scaled squared PDE residual, the desired objective is

$$J(\theta) = \int_{\Omega} f_\theta(x) p(x) dx = E_p[f_\theta(X)].$$

If p is uniform, J is a domain-average residual rather than an unnormalized volume integral. The distinction matters when comparing geometries of different volume or cases with different time horizons. Surface losses, volume losses, and observation likelihoods are separate measures; their raw point counts should not silently determine their relative importance.

Residual-adaptive strategies have been compared across multiple PINN problems, and importance-sampling methods have been proposed to improve training efficiency [3, 4]. These studies motivate adaptation but do not supply a universal best density. A method can improve a training objective while worsening an engineering quantity that the objective does not adequately control. Consequently, both the target measure and the engineering validation measure should be declared.

2. Correcting a changed sampling density

Suppose training locations are drawn from a proposal density q with positive support wherever p_{θ} can be nonzero. Importance weighting gives the estimator

$$\begin{aligned} \hat{J} &= (1/N) \sum_{i=1}^N f_{\theta}(X_i) p(X_i)/q(X_i), X_i \sim q. \\ E_q[\hat{J}] &= J. \\ \text{Var}_q(\hat{J}) &= (1/N) [\int f_{\theta}(x)^2 p(x)^2 / q(x) dx - J^2]. \end{aligned}$$

The expectation and variance expressions assume independent draws for the stated fixed θ and density. If one omits p/q , the sample average instead estimates the q -weighted objective. That can be a deliberate and useful choice, such as emphasizing a boundary layer, but it should be described as residual reweighting. It is not merely a lower-variance estimate of the original domain-average residual.

For estimating a fixed nonnegative scalar f , the formal zero-variance density is $q^*(x) = p(x)f(x)/J$ on its nonzero support. Every importance-weighted sample then equals J . This observation is mainly explanatory: evaluating the exact density and normalizing constant may already require the integral being estimated. Practical methods approximate the density using candidate points, coarse partitions, or previously evaluated residuals. Nabian et al. use approximate importance-sampling ideas to avoid evaluating a costly exact sampling distribution at every step [4].

Training requires a gradient, not only a scalar objective. With $g_{\theta}(x) = \nabla_{\theta} f_{\theta}(x)$, minimizing the second moment of the Euclidean norm of the importance-weighted gradient formally gives q proportional to $p\|g_{\theta}\|$, under the same support and integrability conditions. A residual-based density is therefore a proxy for an efficient gradient density, not generally its optimum. Computing per-point parameter-gradient norms can itself be expensive, so the relevant comparison is the improvement per unit total computation.

When q is updated from the current model, one can condition on the chosen density and treat it as fixed while computing the importance-weighted gradient. Automatically differentiating only the denominator q_{θ} in the sampled expression introduces extra terms and does not give the intended gradient estimator. A treatment that differentiates a parameter-dependent sampling distribution requires the corresponding distributional terms. A practical implementation should state whether the sampler and weights are detached from automatic differentiation.

3. A hot-spot example with exact probabilities

Consider a unit-volume domain with a region H occupying one percent of its volume. Suppose the squared residual f is 100 inside H and 1 elsewhere. Under the uniform target density, the exact mean is $0.01 \times 100 + 0.99 \times 1 = 1.99$. The second moment is 100.99 and the variance of a single uniform sample is $100.99 - 1.99^2 = 97.0299$. An average of N independent uniform samples has variance $97.0299/N$.

The probability of drawing no point in H is 0.99^N . It is approximately 0.366 for $N = 100$ and 0.00657 for $N = 500$. Thus a training or verification set with one hundred independent uniform points has a substantial chance of missing the entire high-residual region. These probabilities describe the specified toy field. Their interpretation depends on the specified residual field, region size, and independent uniform sampling assumptions.

Now choose a proposal that places half its probability in H and half outside. Its densities are 50 inside H and $0.5/0.99$ outside. The importance-weighted sample values become 2 inside and 1.98 outside. Their mean is still 1.99, and their single-sample variance is 0.0001. Without importance weights, the expected sampled loss is instead 50.5. The dramatic numerical difference illustrates both the potential value of informed sampling and the need to identify which objective is being estimated.

Sampling and estimator	Expected reported loss	Single-sample variance
Uniform sampling, ordinary average	1.99	97.0299
Half the probability in H , corrected average	1.99	0.0001
Half the probability in H , uncorrected average	50.5	2450.25

The example assumes that H is already known. Discovering it is part of the real computational problem. A proposal fitted from a sparse candidate set may never discover a thin region that receives no candidates. High residuals at observed points guide exploitation; continued exploration is needed to reveal unobserved errors. Moreover, reducing the variance of this scalar residual estimate does not prove an equal improvement in the gradient estimator or the final engineering field.

4. Preserve exploration and account for geometry

A support-preserving proposal can mix an adaptive density q_a with the target density: $q = (1 - \varepsilon)q_a + \varepsilon p$, where $0 < \varepsilon \leq 1$. Then $p/q \leq 1/\varepsilon$ wherever p is positive. The mixture guarantees exploration under p and bounds the importance ratio. It does not guarantee a finite variance when f itself lacks a finite second moment, nor does it guarantee that a practical finite sample resolves every layer. The exploration fraction is a numerical design parameter that should be reported.

Separate interior, boundary, initial, and interface samplers. Each should have its own normalized target measure and physical scaling. Increasing boundary points should reduce uncertainty in the boundary estimate, rather than accidentally multiplying the physical importance of the boundary term through an unnormalized sum. In curved or moving geometries, use the correct volume or surface Jacobian when sampling in reference coordinates. A uniform distribution in a parameter coordinate is not necessarily a uniform distribution along a physical boundary.

For material interfaces, sample both sides as well as the interface itself. A globally smooth network can obscure a legitimate derivative jump, and an interior residual evaluated away from the interface

cannot verify a transmission condition. Finite-basis and domain-decomposition approaches offer ways to localize approximation, while variational and hp-variational PINNs introduce test functions and quadrature with their own resolution requirements [5, 6, 7]. Adaptation should respect these formulations rather than treating all points as interchangeable residual samples.

Exact boundary ansatzes and constrained optimization also change the sampling task [8, 9]. An algebraically satisfied Dirichlet trace may not require a boundary penalty, but derivatives near that boundary and other physical boundary conditions still need verification. An augmented-Lagrangian condition is not necessarily exact at finite optimization tolerance. Sampling requirements follow the actual residual system, not the label hard constraint.

5. Residual magnitude versus engineering importance

A region with the largest residual need not dominate the error in the quantity the engineer cares about. For a linear problem $Lu = f$ with appropriate homogeneous error boundary conditions, let u_h be an approximation, $e = u_h - u$, and $r = Lu_h - f$. For a linear quantity $Q(e) = \langle c, e \rangle$, an adjoint z satisfying $L^* z = c$ gives, when the integrations by parts and boundary conditions are valid,

$$Q(e) = \langle c, e \rangle = \langle z, Le \rangle = \langle z, r \rangle.$$

This identity explains why a residual weighted by adjoint influence can be more relevant than residual magnitude alone. A small residual in a region with large influence on a sensor output can matter more than a larger residual elsewhere. For nonlinear problems the corresponding argument requires linearization and remainder terms. Unknown or unsatisfied boundary conditions add boundary contributions. Certification with a learned adjoint requires controlling its approximation error as well.

In flow engineering, quantities such as pressure drop, lift, or wake frequency depend on different spatial and temporal structures. Frequency-compensated PINNs illustrate the importance of representing periodic flow behavior across geometry and operating conditions [10]. Fourier features, sinusoidal representations, and analyses of multiscale eigenvectors address approximation capacity [11, 12, 13]. Adaptive collocation addresses where the equations are tested. One cannot compensate indefinitely for inadequate frequency representation simply by adding points, or for an unresolved residual region simply by adding frequencies.

Mixed-variable flow models, hidden-fluid reconstruction, and vortex-induced-vibration studies further illustrate that outputs and observations alter the relevant residual system [14, 15, 16]. In heat transfer and additive manufacturing, moving heat sources and thin thermal layers motivate spatial and temporal refinement [17, 18]. In solid mechanics, stress concentrations and constitutive consistency require checks beyond a displacement residual [19, 20, 21]. These examples justify application-specific refinement criteria; they do not establish that residual-only adaptation captures every important engineering error.

6. Time, optimization, and measurement points

Time-dependent problems introduce a causality issue. An optimizer may reduce late-time residuals while the early-time state remains inaccurate, creating an apparently favorable global average. Causal training strategies explicitly account for temporal ordering [22]. Adaptive time sampling should record whether early intervals must meet a criterion before later ones receive substantial weight. A sampling density that shifts toward easy intervals can otherwise reduce the displayed loss without improving the trajectory where it matters.

The residual components themselves may be imbalanced. Gradient pathologies, neural-tangent-kernel analyses, and documented PINN failures show that optimization can favor some terms over others [23, 24, 25]. Self-adaptive and relative-progress weighting methods change term emphasis [26, 27]. Gradient-enhanced PINNs add residual derivative information [28]. These are separate interventions from spatial sampling. A study that changes weights, architecture, residual order, and collocation simultaneously cannot isolate the benefit of its sampler.

Physical collocation points must also be distinguished from measurement locations. Evaluating a PDE residual at another point is computational work using an assumed model. Acquiring a sensor reading supplies information from the physical system. Sparse-sensor design addresses observation placement and information content [29]. Increasing collocation density cannot eliminate an inverse ambiguity that requires a new measurement, and uncertainty in boundary data cannot be cured merely by enforcing those uncertain values at more points.

7. Independent verification and computational accounting

Use an independent verification set whose design is fixed before comparing samplers. It should include a target-measure sample, structured checks near boundaries and interfaces, and task-specific evaluations. The same adaptive points used to minimize a residual are an optimistic place to estimate final error. If an independent audit repeatedly guides model revisions, retain another final audit set or disclose that the procedure has become iterative development.

Conditional convergence results for certain linear PDE classes explain why residual control can support solution accuracy under appropriate assumptions [30]. Inverse-problem stability estimates make additional assumptions on observations and the PDE [31]. Neither result implies that a finite, adaptively selected training set certifies the continuous residual norm. Verification requires sufficient quadrature, support coverage, and an understanding of the stability relationship for the specific problem.

Count the cost of candidate generation, residual evaluation, density normalization, resampling, differentiation, and actual optimization. A method that reduces gradient steps may still take longer if every update scores a large candidate pool. Report accuracy against elapsed time as well as against training iterations, and include memory use when candidate arrays or derivative graphs become large. Conventional solver comparisons remain relevant because the best adaptive PINN may still be less efficient than a mesh-based solve for the intended task [32].

Benchmark collections and common implementations support reproducibility [33, 34, 35]. They should be supplemented by engineering stress cases, such as a thin layer, a localized source, a geometry change, or a long temporal horizon. Published training guidance provides a useful baseline for optimizer and scaling choices [36]. The sampler comparison should hold those choices fixed initially, then examine interactions through explicitly described ablations.

Item to report	Why it is needed
Target measure and all proposal densities	Distinguishes objective reweighting from corrected estimation
Exploration fraction and candidate-pool policy	Reveals coverage and missed-region risk
Resampling frequency and differentiation treatment	Makes the actual gradient estimator reproducible
Separate volume, surface, and time budgets	Prevents hidden changes in physical term weighting
Independent residual and engineering checks	Separates training success from physical accuracy
Total time and failed runs	Exposes adaptation overhead and instability

8. A controlled sampler experiment for a thermal layer

A useful proposed comparison is transient conduction with a localized heat source and a resolved conventional reference solution. Keep the network, dimensionless residual, initial condition, boundary treatment, optimizer, and total computational budget fixed. Compare uniform collocation, an uncorrected residual-focused sampler, and an importance-corrected sampler. The first and third estimate the same target objective; the second intentionally emphasizes a different one. Report this distinction before comparing final losses.

Measure temperature error over the complete domain, normal heat-flux error on the relevant surface, energy-balance error, and maximum-temperature error. Record the probability mass assigned to the source region and the fraction of the domain receiving few candidates. Evaluate all three models on identical independent quadrature and time slices. If the adaptive model reduces hot-spot error but increases background error, the result may still be useful for the intended decision, but the tradeoff should be visible.

A stale proposal can remain an unbiased importance proposal for the current residual as long as the actual proposal density used to generate the samples is known and the correct ratio is retained. Its efficiency may deteriorate as the field changes. In contrast, reusing samples and assigning them weights from a newly fitted density that did not generate them generally invalidates the simple

estimator. Store proposal metadata with batches, or use a sampling scheme whose weighting rule explicitly accounts for sample reuse.

For adaptively selected densities, conditional unbiasedness of each fresh weighted batch does not mean that every retrospective statistic is unbiased. Choosing the best checkpoint using noisy audit estimates favors checkpoints with favorable estimation noise. Stopping after an unusually small residual estimate produces a related selection effect. Separate training feedback from the final audit and retain uncertainty estimates for Monte Carlo verification when they are material to the acceptance decision.

This controlled experiment would distinguish improved numerical integration, changed physical emphasis, and improved optimization. Adding a separate architecture ablation could then test whether the layer requires a richer representation. The sequence avoids attributing every improvement to sampling when several mechanisms have changed simultaneously, and it supplies a reproducible basis for extending the comparison to wakes, stress concentrations, and moving-source manufacturing problems.

9. Conclusions

Adaptive collocation has two legitimate purposes: improving the estimation of a specified residual objective, and changing emphasis toward important regions or times. Importance weighting distinguishes the first from the second. Residual-based sampling can be effective, but the density that estimates a scalar residual efficiently is not generally the density that optimizes parameter gradients or engineering quantities most efficiently.

The analytical hot-spot example shows why exploration and independent verification matter even when average training losses appear small. For engineering use, a sampler should preserve coverage, respect physical measures and interfaces, and be evaluated with complete computational cost. Its success should ultimately be judged by independently verified quantities of interest rather than by a favorable loss trace alone.

References

- [1] Maziar Raissi, Paris Perdikaris, George Em Karniadakis. 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 (2019). DOI.
- [2] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, et al.. Physics-informed machine learning. *Nature Reviews Physics* 3, 422-440 (2021). DOI.
- [3] Chenxi Wu, Min Zhu, Qinyang Tan, et al.. A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks. arXiv:2207.10289 (2022). arXiv.
- [4] Mohammad Amin Nabian, Rini Jasmine Gladstone, Hadi Meidani. Efficient training of physics-informed neural networks via importance sampling. arXiv:2104.12325 (2021). arXiv.
- [5] Ben Moseley, Andrew Markham, Tarje Nissen-Meyer. Finite Basis Physics-Informed Neural Networks (FBPINNs): a scalable domain decomposition approach for solving differential equations. arXiv:2107.07871 (2021). arXiv.
- [6] E. Kharazmi, Z. Zhang, G. E. Karniadakis. Variational Physics-Informed Neural Networks For Solving Partial Differential Equations. arXiv:1912.00873 (2019). arXiv.

- [7] Ehsan Kharazmi, Zhongqiang Zhang, George Em Karniadakis. hp-VPINNs: Variational Physics-Informed Neural Networks With Domain Decomposition. arXiv:2003.05385 (2020). arXiv.
- [8] N. Sukumar, Ankit Srivastava. Exact imposition of boundary conditions with distance functions in physics-informed deep neural networks. arXiv:2104.08426 (2021). arXiv.
- [9] Lu Lu, Raphael Pestourie, Wenjie Yao, et al.. Physics-informed neural networks with hard constraints for inverse design. arXiv:2102.04626 (2021). arXiv.
- [10] Tongtao Zhang, Biswadip Dey, Pratik Kakkar, Arindam Dasgupta, Amit Chakraborty. Frequency-compensated PINNs for Fluid-dynamic Design Problems. ML4Eng Workshop, NeurIPS 2020. arXiv:2011.01456 (2020). arXiv.
- [11] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, et al.. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. arXiv:2006.10739 (2020). arXiv.
- [12] Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, et al.. Implicit Neural Representations with Periodic Activation Functions. arXiv:2006.09661 (2020). arXiv.
- [13] Sifan Wang, Hanwen Wang, Paris Perdikaris. On the eigenvector bias of Fourier feature networks: From regression to solving multi-scale PDEs with physics-informed neural networks. arXiv:2012.10047 (2020). arXiv.
- [14] Chengping Rao, Hao Sun, Yang Liu. Physics-informed deep learning for incompressible laminar flows. arXiv:2002.10558 (2020). arXiv.
- [15] Maziar Raissi, Alireza Yazdani, George Em Karniadakis. Hidden Fluid Mechanics: A Navier-Stokes Informed Deep Learning Framework for Assimilating Flow Visualization Data. arXiv:1808.04327 (2018). arXiv.
- [16] Maziar Raissi, Zhicheng Wang, Michael S. Triantafyllou, George Em Karniadakis. Deep Learning of Vortex Induced Vibrations. arXiv:1808.08952 (2018). arXiv.
- [17] Shuheng Liao, Tianju Xue, Jihoon Jeong, et al.. Hybrid thermal modeling of additive manufacturing processes using physics-informed neural networks for temperature prediction and parameter identification. arXiv:2206.07756 (2022). arXiv.
- [18] Qiming Zhu, Zeliang Liu, Jinhui Yan. Machine learning for metal additive manufacturing: Predicting temperature and melt pool fluid dynamics using physics-informed neural networks. arXiv:2008.13547 (2020). arXiv.
- [19] Ehsan Haghighat, Maziar Raissi, Adrian Moure, et al.. A deep learning framework for solution and discovery in solid mechanics. arXiv:2003.02751 (2020). arXiv.
- [20] R. Sharma, Y. B. Guo. Thermal-Mechanical Physics Informed Deep Learning For Fast Prediction of Thermal Stress Evolution in Laser Metal Deposition. arXiv:2412.18786 (2024). arXiv.
- [21] Filippo Masi, Ioannis Stefanou, Paolo Vannucci, Victor Maffi-Berthier. Thermodynamics-based Artificial Neural Networks for constitutive modeling. arXiv:2005.12183 (2020). arXiv.
- [22] Sifan Wang, Shyam Sankaran, Paris Perdikaris. Respecting causality is all you need for training physics-informed neural networks. arXiv:2203.07404 (2022). arXiv.
- [23] Sifan Wang, Yujun Teng, Paris Perdikaris. Understanding and mitigating gradient pathologies in physics-informed neural networks. arXiv:2001.04536 (2020). arXiv.
- [24] Sifan Wang, Xinling Yu, Paris Perdikaris. When and why PINNs fail to train: A neural tangent kernel perspective. arXiv:2007.14527 (2020). arXiv.
- [25] Aditi S. Krishnapriyan, Amir Gholami, Shandian Zhe, et al.. Characterizing possible failure modes in physics-informed neural networks. arXiv:2109.01050 (2021). arXiv.
- [26] Levi McClenny, Ulisses Braga-Neto. Self-Adaptive Physics-Informed Neural Networks using a Soft Attention Mechanism. arXiv:2009.04544 (2020). arXiv.
- [27] Rafael Bischof, Michael Kraus. Multi-Objective Loss Balancing for Physics-Informed Deep Learning. arXiv:2110.09813 (2021). arXiv.
- [28] Jeremy Yu, Lu Lu, Xuhui Meng, George Em Karniadakis. Gradient-enhanced physics-informed neural networks for forward and inverse PDE problems. arXiv:2111.02801 (2021). arXiv.

- [29] Krithika Manohar, Bingni W. Brunton, J. Nathan Kutz, Steven L. Brunton. Data-Driven Sparse Sensor Placement for Reconstruction. arXiv:1701.07569 (2017). arXiv.
- [30] Yeonjong Shin, Jerome Darbon, George Em Karniadakis. On the convergence of physics informed neural networks for linear second-order elliptic and parabolic type PDEs. arXiv:2004.01806 (2020). arXiv.
- [31] Siddhartha Mishra, Roberto Molinaro. Estimates on the generalization error of Physics Informed Neural Networks (PINNs) for approximating a class of inverse problems for PDEs. arXiv:2007.01138 (2020). arXiv.
- [32] Tamara G. Grossmann, Urszula Julia Komorowska, Jonas Latz, Carola-Bibiane Schönlieb. Can Physics-Informed Neural Networks beat the Finite Element Method?. arXiv:2302.04107 (2023). arXiv.
- [33] Zhongkai Hao, Jiachen Yao, Chang Su, et al.. PINNacle: A Comprehensive Benchmark of Physics-Informed Neural Networks for Solving PDEs. arXiv:2306.08827 (2023). arXiv.
- [34] Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, et al.. PDEBENCH: An Extensive Benchmark for Scientific Machine Learning. arXiv:2210.07182 (2022). arXiv.
- [35] Lu Lu, Xuhui Meng, Zhiping Mao, George E. Karniadakis. DeepXDE: A deep learning library for solving differential equations. arXiv:1907.04502 (2019). arXiv.
- [36] Sifan Wang, Shyam Sankaran, Hanwen Wang, Paris Perdikaris. An Expert's Guide to Training Physics-informed Neural Networks. arXiv:2308.08468 (2023). arXiv.
- [37] Rajat Arora, Pratik Kakkar, Amit Chakraborty, Biswadip Dey. Physics-informed neural networks for modeling rate- and temperature-dependent plasticity. Machine Learning and the Physical Sciences Workshop, NeurIPS 2022. arXiv:2201.08363v3 (2022). arXiv.