

Picking the Branch Is Not Solving the Problem: What an Imperfect Neural Operator Buys a Physics-Informed Network, Where Its Basin Information Comes From, and the Comparators That Decide Whether the Hybrid Pays

Pranay Mahendrakar

SONYTECH · ORCID 0009-0003-7224-029X

pranaymahendrakar@sonytech.in · mahendrakarpranay@gmail.com · +91 6361723454

Abstract. Physics-informed neural networks (PINNs) can minimise a PDE residual and still land on the wrong solution, while neural operators are cheap to evaluate but lose accuracy outside their training range. A 2026 study joins the two: a crude, extrapolated operator is distilled into a fresh PINN, and the authors conclude that a prior need not be accurate, only able to identify the correct basin from which residual minimisation recovers the solution. In this analysis the author reviews and synthesises published results to show that the principle is sound but old, and that the evidence offered for it leaves open the three questions that decide whether the hybrid is worth building: where the basin information came from, whether the comparison baseline was the one practitioners would use, and whether a PINN is the right refiner once a basin is known. For the bifurcating Allen-Cahn case, the author derives a one-mode estimate of the branch amplitude that agrees with the study's reference peak to within about 3 percent. On that estimate, the seed function hand-placed in the operator's ansatz is closer to the target than the extrapolated teacher. For the lid-driven cavity at Reynolds number 3200, each of the published arms collected here that reaches good accuracy walks a path in Reynolds number, and a curriculum-trained PINN with no operator reports a lower error than the hybrid. Eighteen published results are consolidated in a table, a five-arm comparator procedure is proposed for licensing a basin-selection claim, and the opposite failure is recorded as well: an accurate operator can still start Newton's method outside its basin.

Keywords: physics-informed neural networks; neural operators; basin of attraction; numerical continuation; solution multiplicity; warm start; scientific machine learning; baselines

SONYTECH ORCID: 0009-0003-7224-029X

1. Introduction

A residual is a weak certificate. When a nonlinear PDE has more than one solution, a field that makes the residual small may be the wrong one, and an optimiser that drives the residual toward zero has no reason to prefer the branch the modeller cares about. Classical numerical analysis has lived with this for decades. Newton's method converges quadratically only from inside a basin around a root; continuation methods walk a parameter from an easy problem to a hard one so that each solve starts from a good guess; and deflation modifies the residual so that known roots stop attracting the iteration (Allgower and Georg, 2003; Farrell et al., 2014). Physics-informed neural networks (Raissi et al., 2019; Karniadakis et al., 2021) inherited the problem without inheriting the tools, and a large literature now documents PINNs that converge to trivial, unphysical or merely smooth solutions (Krishnapriyan et al., 2021; Leiteritz and Pfluger, 2021; Daw et al., 2022; Chuang and Barba, 2022; Wang et al., 2023b).

Neural operators sit on the other side of the same trade. They learn a map from parameters to solutions and evaluate it quickly (Lu et al., 2019; Li et al., 2020; Kovachki et al., 2021), and physics-informed variants can be trained from residuals alone (Wang et al., 2021;

Li et al., 2021). Their weakness is extrapolation: outside the training distribution their predictions degrade, sometimes badly (Zhu et al., 2022; Lara Benitez et al., 2023; Serrano et al., 2026).

A September 2026 study by Mousavi, Kadeethum, Bouklas and Goswami proposes to make each method cover the other's weakness (Mousavi et al., 2026; hereafter "the anchor study"). A physics-informed operator is trained on a box of parameters, its spatial basis is frozen, its coefficients are extrapolated to an out-of-distribution target with a low-degree polynomial continuation prior, and the resulting field is distilled into a freshly initialised PINN whose distillation weight decays to zero. On three problems, the viscous Burgers equation, a steady Allen-Cahn equation just above a pitchfork bifurcation, and the lid-driven cavity at Reynolds number $Re = 3200$, the distilled PINN beats a PINN trained from scratch by wide margins. The authors summarise the result in one sentence: "an NO need not accurately predict the solution to be useful; it only needs to identify the correct basin from which PINN optimization can recover it."

This paper takes that sentence seriously as a claim and asks what would have to be true for it to license building the hybrid. The author's answer is that three separate questions are folded into the anchor study's evidence and need to be pulled apart.

The first is provenance. If a crude prior selects the right basin, the basin information has to have come from somewhere. The anchor study reports that it came from the operator. Its own appendices show two other sources: a hand-placed seed function and anti-collapse penalty in the Allen-Cahn case, and eight finite-difference solves plus a four-rung continuation ladder in the cavity case.

The second is the baseline. A from-scratch PINN at the target parameter is the weakest available comparator for a parametric extrapolation problem. The practice the PINN literature itself recommends for high-Reynolds cavity flow is a curriculum in Reynolds number, and the published curriculum results are part of the evidence (Wang et al., 2023a; Wang et al., 2024).

The third is the refiner. Once a basin is known, a PINN is one way to descend into it. A classical solver initialised from the same prior is another, and a growing literature on neural-operator warm starts for Krylov and Newton solvers reports that route working (Zhou et al., 2023; Aghili et al., 2024; Eshaghi et al., 2025; Taghikhani et al., 2025; Wang et al., 2026; Oh et al., 2026).

The paper makes five contributions. First, it splits "wrong-basin" failures three ways, into genuine multiplicity of the PDE, spurious minimisers introduced by the PINN loss, and optimisation pathology, each with a different remedy and a different test (Section 3). Second, it gives a provenance account of the basin information in each of the anchor study's three experiments. This includes a closed-form one-mode estimate of the Allen-Cahn branch that reproduces the study's reference amplitude to about 3 percent and identifies a zero-cost teacher, and a reading of the study's extrapolation stage as a predictor-corrector continuation in a learned basis (Section 4, Figure 1b). Third, it re-reads the cavity result against the published curriculum baselines, one of which reports a lower error without any operator; every published arm collected here that reaches good accuracy at $Re = 3200$ walks a path in Reynolds number (Section 5, Figure 1a). Fourth, it records the mirror image of "accuracy is not required", which is also published: accuracy is not sufficient either, because a basin belongs to a prior and a refiner together, not to the prior alone (Oh et al., 2026; Section 7). Fifth, it consolidates the evidence in a table (Table 1) and a provenance table (Table 2), proposes a five-arm comparator procedure (Algorithm 1) that would license or refuse a basin-selection claim, and lists five studies that would settle what the published record leaves open (Sections 10 and 11).

In this analysis the author reviews and synthesises published results, each credited to the paper that reported it, and derives one closed-form estimate (Section 4), labelled as such.

2. What the anchor study reports

The framework has three parts (Mousavi et al., 2026). Part 1 trains a physics-informed operator with a trunk-branch decomposition, in the DeepONet style (Lu et al., 2019; Wang et al., 2021), over a bounded parameter box, then freezes the trunk and builds a set of trusted coefficient "anchors" at a few base parameters. Part 2 fits a low-degree polynomial through those anchors, extrapolates it to the target parameter, and re-trains a fresh branch head whose output is a bounded correction to the polynomial, with a penalty that keeps the field near the prior. When the target is far away, Part 2 proceeds in stages, appending each converged intermediate solution to the anchor pool. Part 3 hands the extrapolated field to a new multilayer-perceptron PINN that drops the trunk-branch structure and the boundary lift, trains it on the residual plus a distillation term, and drives the distillation weight to zero before training ends. The authors stress that this is not knowledge distillation in the compression sense of Hinton et al. (2015), where the student is capped by an accurate teacher: here the teacher is expected to be wrong, and the student is expected to surpass it.

The three experiments report the following, each over five seeds and as relative L2 error against a reference field, with accuracy defined as one minus that error.

Burgers (one dimension). The operator is trained on viscosity and domain length (ν, L) in $[0.10, 0.20] \times [1, 2]$ and extrapolated to $(0.05, 4)$. Its out-of-distribution error is 0.12. The from-scratch PINN reaches 0.4123 ± 0.1601 ; the distilled PINN reaches 0.0209 ± 0.0043 . The authors describe this problem as having a unique monotone solution manifold, and they report that the baseline's accuracy depends strongly on the seed and occasionally collapses to a poor local solution.

Allen-Cahn (two dimensions, steady). The equation is the Laplacian of u plus $\lambda(u - u^3) = 0$ with zero boundary values on a rectangle of aspect ratio H . Below the first Dirichlet eigenvalue $\lambda_1(H) = \pi^2(1 + 1/H^2)$ only the zero solution exists; above it, a pair of non-trivial branches bifurcates. The operator is trained on (λ, H) in $[35, 80] \times [1, 2]$, comfortably supercritical, and the target is $(12.5, 3)$, about 14 percent above onset, where the authors report a reference peak amplitude of 0.4530. The from-scratch PINN collapses to the zero field in every seed (error 1.00 ± 0.00 , peak 0.0002 ± 0.0002). The extrapolated teacher has the right single-hump shape but overshoots the amplitude. The distilled PINN reaches 0.0037 ± 0.001 .

Lid-driven cavity (two dimensions, steady). The operator is trained on Re in $[100, 800]$ at unit aspect ratio, with eight finite-difference solves supplying the anchors, and extrapolated through a four-rung ladder 1400, 2000, 2500, 3200. The operator's in-distribution error is already 0.2586 ± 0.1294 and its error at the

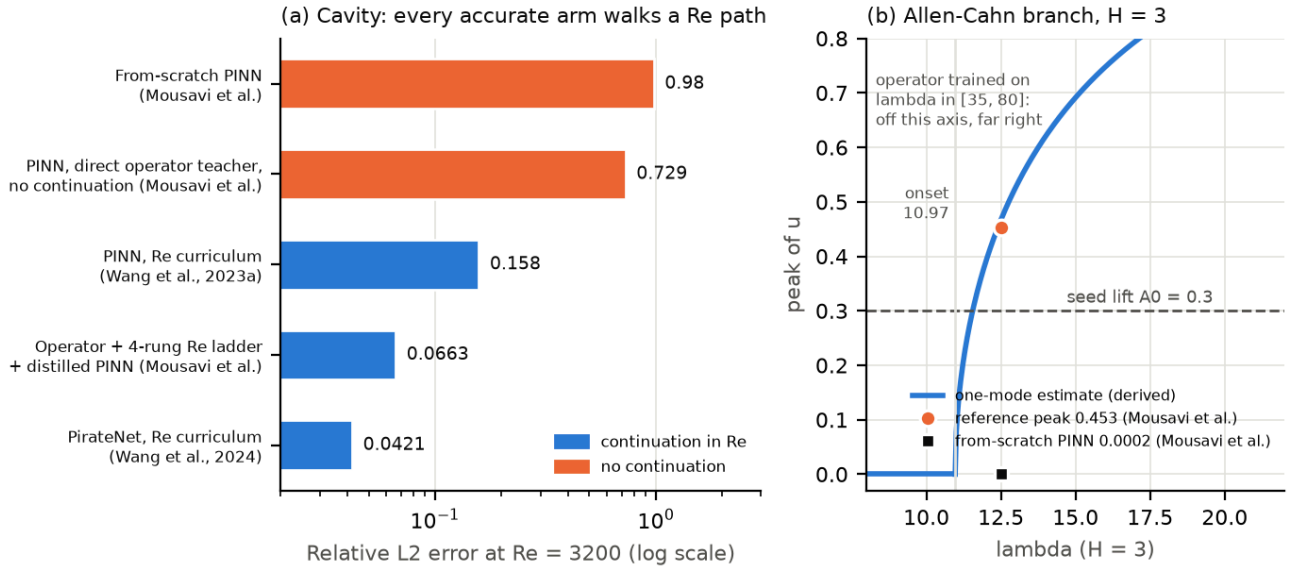


Figure 1. (a) Published relative L2 errors for the lid-driven cavity at $Re = 3200$: from-scratch PINN, PINN distilled from the operator without continuation, and PINN distilled from the operator after a four-rung Reynolds ladder (Mousavi et al., 2026); Reynolds-curriculum PINN (Wang et al., 2023a); PirateNet with a Reynolds curriculum (Wang et al., 2024). Lid regularisations and references differ between sources (Section 5). (b) Steady Allen-Cahn branch amplitude at $H = 3$: the curve is the author's one-mode Galerkin estimate $a^2 = (16/9)(\lambda - \lambda_{1,1})/\lambda$, a derivation and not a measurement; the reference peak 0.4530 and the from-scratch PINN peak 0.0002 at $\lambda = 12.5$ are from Mousavi et al. (2026), as is the seed amplitude 0.3.

target is 0.53. The from-scratch PINN reaches 0.9800 ± 0.0248 , the distilled PINN 0.0663 ± 0.0099 . In an ablation that skips Part 2, the teacher is the Part 1 operator evaluated directly at $Re = 3200$, and the student distilled from it reaches 27.09 ± 13.49 percent accuracy, an error of about 0.73; the study does not report the direct teacher's own error. The final PDE residuals of the three arms are $6.3e-6$ (full framework), $1.3e-5$ (no Part 2) and $2.9e-5$ (baseline), and the authors read the similarity of these residuals, set against very different fields, as the signature of a basin-selection failure rather than an optimisation failure.

The anchor study also compares its cavity student with PirateNet (Wang et al., 2024), the residual-adaptive architecture it names as the state of the art for a single cavity at $Re = 3200$: 0.0663 against 0.0421 in relative L2 error, 297 thousand against roughly 1.32 million trainable parameters, 250 thousand against 630 thousand optimisation steps, and $6.5e9$ against $2.6e9$ point evaluations. It notes that Parts 1 to 3 together take 546 thousand steps, and that Part 1 is a one-time cost reusable across targets.

Two small fidelity notes. First, the anchor study's Table 2 gives the Allen-Cahn teacher's out-of-distribution error as 0.78, which under its own definition is 22 percent accuracy, while its text twice describes the same teacher as 32 percent accurate; this paper uses the table value and flags the discrepancy. Second, the cavity figure of 0.0663 appears both in Table 2, against the per-problem reference field, and in Table 3, labelled as error against Ghia et al. (1982); the study does not say whether these are the same computation. Neither note changes the

argument below.

3. Three ways to land in the wrong basin

"Basin" is used in the anchor study, and in much of the literature it draws on, for at least three different situations. They share a symptom, a trained network whose field is wrong though its residual is small, but they have different causes, different remedies, and different consequences for what a prior can contribute.

3.1 Genuine multiplicity of the PDE

Some boundary-value problems have several solutions for the same data. The steady Allen-Cahn equation in the anchor study is the textbook case: above $\lambda_{1,1}$ the zero field and a pair of signed branches all satisfy the equation exactly. Bifurcation from a simple eigenvalue is classical (Crandall and Rabinowitz, 1971), and the existence and stability of stationary solutions for this one-dimensional reaction-diffusion family was worked out by Chafee and Infante (1974). Multiplicity of this kind is a property of the continuous problem. No change to the loss, the sampler or the optimiser removes it, because every branch has zero residual. A method that returns one solution has to be told which one, and the telling is information that comes from outside the residual.

This is the situation in which the anchor study's principle is cleanest, and the Allen-Cahn experiment is its cleanest demonstration. It is also the situation that classical numerics handles with the most mature tools. Deflation modifies the residual so that already-found solutions are no longer roots, and Farrell et al. (2014) report that

Newton's method with deflation, started from a zero initial guess each time, finds three steady Allen-Cahn solutions on their benchmark, one unstable and two stable. The PINN literature has ported the idea: structure-probing deflation for networks (Gu et al., 2020), homotopy-continuation PINNs for inverse problems with several solutions (Zheng et al., 2023), Newton-informed neural operators that learn several solutions at once (Hao et al., 2024), and Deflation-PINNs (Disaro et al., 2026), whose authors report two results: all six stable states of a Landau-de Gennes liquid-crystal problem on a square, recovered in one unsupervised run and then refined by a gradient-flow and Newton solve that the network initialises; and all three solutions, 0 , $+u^*$ and $-u^*$, of a steady Allen-Cahn problem on the unit disk at $\lambda = 6$, just above the first Dirichlet eigenvalue of about 5.783 , recovered in one run with no classical refinement. The Allen-Cahn result is the closest published analogue to the anchor study's near-onset pitchfork. Zou, Wang and Karniadakis (2025) report an even simpler route: training 50 randomly initialised PINNs on the steady Allen-Cahn problem with plus-and-minus-one boundary values that Farrell et al. also use, at a different interface parameter, recovers three distinct solutions, and five independently initialised PINNs on a cavity driven by a sinusoidal lid produce three solution types, one stable symmetric, one unstable asymmetric and one stable asymmetric, which the authors check by using them to initialise finite-element and spectral-element solvers.

The point for the anchor study is not that these methods beat it; none was run on its benchmark. The point is that when multiplicity is genuine, finding the branches and choosing among them are separate tasks. Ensembles and deflation enumerate branches. Choosing the physically relevant one needs a criterion: stability, symmetry, continuity from a known state, or agreement with an observation. The anchor study supplies that criterion implicitly, through continuity of the solution branch from the training box, and Section 4 asks how much of it the operator actually carries.

3.2 Spurious minimisers introduced by the PINN loss

Other "wrong basins" are not solutions of the PDE at all. They are minimisers of the discretised, penalised loss that a PINN actually optimises. Leiteritz and Pfluger (2021) show that with too few collocation points a PINN can learn a trivial function that satisfies the physics penalty at the sampled points by construction. Daw et al. (2022) argue that PINN training depends on solution information propagating inward from boundary and initial points, and that poor sampling lets the network settle on a trivial solution with a highly imbalanced residual field. Basir and Senocak (2022) attribute poor convergence on elliptic problems to incomparable scales between the boundary and domain terms of the composite loss.

For a spurious minimiser, the remedy is to fix the problem being solved: change the sampling, rebalance or harden the boundary terms, regularise the singularity, add data. A good prior can steer around a spurious minimiser, but it cannot certify that the end point solves the PDE, because the end point is judged by the same loss that admitted the spurious minimiser. Telling 3.1 from 3.2 requires a check outside the loss, such as handing the converged field to an independent solver and seeing whether it stays put (Zou et al., 2025, do exactly this).

3.3 Optimisation pathology that looks like a basin

The third case has no second solution at all. The PINN loss is ill-conditioned and the optimiser stalls, or reaches a low-frequency approximation and stops improving. Krishnapriyan et al. (2021) show that the soft PDE regulariser can make the problem more ill-conditioned, that the failures are not a matter of network expressivity, and that curriculum regularisation and sequence-to-sequence training reduce error by one to two orders of magnitude. Rathore et al. (2024) connect ill-conditioned differential operators to ill-conditioning of the PINN loss and report that Adam followed by L-BFGS, and a Nystrom-preconditioned Newton-CG method, improve training substantially. Wang et al. (2020a) trace failures to imbalanced back-propagated gradients between loss terms; Wang et al. (2020b) give a neural-tangent-kernel account in which the loss components converge at very different rates; Wang et al. (2022) attribute failures on dynamical problems to violation of temporal causality; spectral bias, the tendency of networks to fit low frequencies first (Rahaman et al., 2018), is mitigated by Fourier feature embeddings (Tancik et al., 2020), which the anchor study itself uses. Remedies of this family are surveyed by Cuomo et al. (2022) and consolidated into a training recipe by Wang et al. (2023a).

A good initial condition helps here too, because it shortens the path the optimiser must travel through a badly conditioned landscape. But help of this kind is acceleration, not basin selection, and it can be bought in other ways: a better optimiser, a better architecture, a curriculum.

3.4 Why the split matters for the anchor study

The anchor study's three problems fall into different cells. Allen-Cahn is 3.1: the zero field is an exact solution. Burgers, which the study itself describes as having a unique monotone solution manifold, is 3.3: the baseline's seed-dependent accuracy and occasional collapse to "a poor local solution" are optimisation failures, and the distilled PINN's gain there is acceleration and variance reduction, not branch choice. The cavity is the hard case, and the study's evidence does not place it. Wang et al. (2023b) study the cavity directly at $Re = 2000, 3000$ and 5000 , with the lid also smoothed at the corners, by a hyperbolic-cosine profile rather than the anchor study's smotherstep, and report

that vanilla PINNs return two classes of solution: one agrees with direct numerical simulation, and the other, in their words, "is an unstable solution to the Navier-Stokes equations and not physically realizable". They attribute this multiplicity to singularities and unbounded vorticity, and they report that an entropy-viscosity term, or a single labelled measurement point, steers training to the DNS-like answer, with velocity error under 1 percent from one data point at $Re = 2000$. Read literally, their description puts the cavity in cell 3.1, a genuine but unstable second steady state, with stability as the selection criterion. Stability is also what time-marching and relaxation solvers select by construction, which is consistent with the classical benchmarks reporting one steady solution per Reynolds number up to $Re = 10,000$ (Marchi et al., 2021) and $Re = 21,000$ (Erturk et al., 2005). Read through their attribution to singular behaviour, the second class could instead be partly an artefact of the penalised, sampled loss, which is cell 3.2. The anchor study's baseline could also simply have stalled (cell 3.3). The near-equality of residuals ($2.9e-5$ against $6.3e-6$, a factor of about 4.6) does not separate the three readings. A soft-constrained PINN can reach a small interior residual while violating the boundary data, and the study's ablation table reports PDE residuals but not boundary losses.

None of this undermines the anchor study's measured gains. It changes what they are gains in. Only one of the three experiments, Allen-Cahn, is unambiguously a test of basin selection in the sense the headline sentence uses; Burgers is not one, and the cavity may or may not be.

4. Where the basin information comes from

If a prior that is 78 percent wrong in L2 can still select the right basin, the information that does the selecting is coarse: a sign, a shape, a topology. The question is whether the operator, as opposed to the modeller or the external solver, is what supplies it. Table 2 sets out the answer for each experiment; this section works through the two that matter.

4.1 Allen-Cahn: the branch is in the ansatz, and in closed form

The anchor study's Appendix A.3 describes the operator's ansatz for this problem. The field is written as a fixed seed lift, $A_0 \sin(\pi x) \sin(\pi y)$ with $A_0 = 0.3$ on the reference square, plus a boundary multiplier times the trunk-branch expansion. The authors note that the seed is an eigenfunction of the Laplacian with eigenvalue exactly $\lambda_1(H)$, so that it points along the principal eigendirection from which the positive branch bifurcates, and that it "explicitly breaks the $u = 0$ degeneracy at initialization". Part 1 adds, in their Appendix A.6, a soft amplitude floor that penalises fields collapsing toward zero. The operator's base box is restricted to λ in $[35, 80]$, far from onset, in what the authors call the

comfortably supercritical regime, and the positive seed biases the ansatz toward the positive member of the branch pair there, as the authors themselves note. The Part 3 student, by contrast, drops the boundary lift, so the from-scratch baseline sees neither the seed nor the floor.

The comparison in Section 4.2 of the anchor study is therefore between a PINN with no branch information and a PINN distilled from an operator whose branch information was hand-placed by the modellers. That is a fair test of whether the information survives two extrapolation stages and a change of architecture. It is not a test of whether the operator learned it.

How much information is needed? Near a simple bifurcation, the non-trivial branch is, to leading order, a multiple of the principal eigenfunction ϕ_1 , with an amplitude that grows like the square root of the distance from onset. The anchor study states this scaling in its Section 3.3 and attributes the teacher's overshoot to the polynomial prior's inability to represent it. The amplitude can be estimated by projecting the equation onto ϕ_1 . The derivation below is the author's and is not a measurement.

Write $u = a \phi_1$ with $\phi_1 = \sin(\pi x) \sin(\pi y / H)$ on $[0, 1] \times [0, H]$. Multiplying the equation by ϕ_1 and integrating gives $(\lambda - \lambda_1) = \lambda a^3$. On this rectangle $= H/4$ and $= 9H/64$, so

$$a^2 = (16/9) (\lambda - \lambda_1) / \lambda.$$

At $H = 3$, $\lambda_1 = \pi^2 (10/9) = 10.966$, and at $\lambda = 12.5$ this gives $a = 0.467$. The anchor study reports a reference peak of 0.4530 from a finite-difference Newton solve. The one-mode estimate is 3.1 percent high, a gap plausibly due to the higher modes the projection neglects. Figure 1b plots the estimate against λ with the study's reference value, the baseline's collapsed amplitude, and the seed amplitude marked.

Two consequences follow, and both are inferences rather than measurements. First, at this target the branch's amplitude is available to within a few percent from a one-line formula that needs neither an operator nor a reference solve. Second, under the one-mode approximation the bare seed, $0.3 \phi_1$, sits at a relative L2 distance of about $(0.467 - 0.3)/0.467 = 0.36$ from the branch. The extrapolated teacher, by the anchor study's Table 2, sits at 0.78. On that approximation, the cheapest teacher available, the seed the operator was built on, is closer in L2 to the branch than the operator it was used to build; whether that makes it a better initial field is what Study 1 would test. Whether a PINN distilled from the seed alone would also recover the branch in every seed has not been tested; the anchor study does not report that arm. It is the single most informative control the Allen-Cahn experiment lacks.

The caveat is real, and it cuts in the anchor study's favour for other problems. Allen-Cahn on a rectangle is unusually transparent: one simple eigenvalue, one branch direction, a closed-form critical parameter. In a problem with a degenerate eigenvalue, a secondary bifurcation, or a branch that has travelled far from onset, no one-mode formula exists, and a learned operator that carries branch identity across a parameter range could supply something no hand calculation does. The anchor study's evidence does not reach such a problem. The Allen-Cahn experiment shows that branch identity survives extrapolation and distillation; it does not show that the operator was the cheapest or the only source of it.

4.2 Cavity: eight solves and a ladder

For the cavity the anchor study is explicit that external data enter: Part 1's anchors are eight finite-difference reference fields at Re in the base box, reprojected onto the frozen basis, and these are "the only external data the main cavity pipeline ever consumes". The reference used to score the result is a finite-difference streamfunction-vorticity solver on a uniform 129-point grid, iterated to a steady state at the target Reynolds number with the same regularised lid (the study's Appendix D). Part 2 then advances through 1400, 2000, 2500 and 3200, fitting the polynomial prior afresh at each rung after appending the previous rung's converged field to the anchor pool.

Two things about this matter. The first is that the operator by itself is a poor model of the flow even inside its box: an in-distribution error of 0.2586 ± 0.1294 is far from the 0.0015 it achieves for Allen-Cahn. Whatever the teacher contributes at $Re = 3200$ is, on the author's reading, coarse structural information: a primary vortex in roughly the right place, and corner vortices. That reading is consistent with the anchor study's description of the failed ablation as having a mislocated primary vortex and no lower corner vortices.

The second is that the ablation isolates the ladder, not the operator. With the operator held fixed and Part 2 removed, the student's accuracy drops from 93 percent to 27 percent. The component whose removal drops the distilled student from 93 to 27 percent accuracy is the staged continuation in Reynolds number. With it, the teacher handed to the student is 53 percent wrong; the error of the direct teacher used without it is not reported. That is one of the oldest tools in the numerical treatment of nonlinear problems. Continuation in a physical parameter is how classical solvers reach hard nonlinear regimes (Allgower and Georg, 2003); Oh et al. (2026) describe incremental continuation as the conventional stabiliser for Newton when the basin is small; and, as Section 5 shows, continuation in Reynolds number is also how the published PINN baselines reach $Re = 3200$.

4.3 Part 2 is a predictor-corrector continuation in a learned basis

It helps to describe Part 2 in the vocabulary of numerical continuation, because once described that way its role in the results is easier to see. Classical predictor-corrector continuation advances along a solution curve in two moves (Allgower and Georg, 2003). A predictor extrapolates the known part of the curve to the next parameter value, and a corrector, usually Newton's method, pulls the prediction back onto the curve. The step size is kept small enough for the prediction to land inside the corrector's basin, and each converged point is added to the history used by the next predictor.

Part 2 of the anchor study does each of these things. The low-degree polynomial fitted through the anchor coefficients is the predictor: it extrapolates the branch in the coefficients of a fixed basis. The retrained branch head, constrained to a bounded correction of the polynomial and penalised in the field metric for departing from it, is the corrector. It minimises the residual in the frozen basis, which plays the role of a reduced-order discretisation. The staged ladder, with each converged rung appended to the anchor pool and the polynomial refitted, is step-size control with a growing history. The bounded-correction cap is a trust region. The novelty lies in the basis, which is learned rather than chosen, and in the corrector, which is a gradient-trained network head rather than Newton's method. The continuation logic is classical.

Seen this way, the ablation of Section 4.2 is unsurprising. Removing Part 2 removes the predictor-corrector march and evaluates the learned map four times beyond its training range in a single step. That this fails is consistent with what the extrapolation literature reports. Zhu et al. (2022) report that purely data-driven deep operators are usually limited to interpolation and can fail with large errors under extrapolation, and propose fine-tuning with the PDE loss or with sparse new observations as remedies. Lara Benitez et al. (2023) report that standard operators do not perform well on every physics problem and study bounds on out-of-distribution risk. Serrano et al. (2026) report that operators often struggle when test inputs lie outside the training distribution. Transfer approaches that work under conditional shift use labelled target data (Goswami et al., 2022). The anchor study's own framing, an "imperfect ancestor", undersells the result: by the time the teacher reaches $Re = 3200$ it is no longer the ancestor's prediction. It is the end point of a four-step continuation computed in the ancestor's basis.

This matters for what the principle claims. "An NO need not accurately predict the solution" is true of the Part 1 operator, whose direct prediction at the target does not work. What does work is a continuation that uses the operator's basis as a reduced space. That is a useful idea in its own right, a learned reduced basis for

continuation, and it would be tested by comparing it with continuation in a classical reduced basis such as proper orthogonal decomposition modes of the eight anchor solves. The anchor study does not report that comparison either.

4.4 Burgers: no basin to select

The Burgers experiment has a unique solution and an analytical reference via the Cole-Hopf transform. The distilled PINN's improvement there, from 0.41 to 0.021 in error, with the standard deviation across seeds shrinking from 0.16 to 0.004, is large and useful. It is an acceleration-and-reliability result of the kind reported for meta-learned and transferred PINN initialisations on parameterised problems (Penwarden et al., 2021; Desai et al., 2021) and for learned initialisation more broadly (Anurag et al., 2026). It is not evidence about basin selection, and the anchor study is careful to present it as establishing "the fundamental transfer mechanism" rather than as a multiplicity case.

5. The baseline that decides the cavity result

McGreiv and Hakim (2024) reviewed the machine-learning-for-PDE literature and report that 79 percent (60 of 76) of articles claiming to outperform a standard numerical method on a fluid-related PDE compared against a weak baseline. They propose two rules: compare at equal accuracy or equal runtime, and compare against an efficient numerical method. Their review concerns ML solvers against numerical methods, not ML methods against each other. The logic transfers, though: a claim that a component helps is only as strong as the best alternative it was compared with.

For the cavity at $Re = 3200$, the anchor study's comparator for the distillation effect is a PINN trained from scratch at the target. The PINN literature does not recommend that procedure. Wang et al. (2023a) report that, in their experience, PINNs trained directly at high Reynolds number are unstable and liable to converge to erroneous solutions, that PINNs struggle beyond $Re = 1000$, and that a curriculum through $Re = 100, 400, 1000, 3200$ with $5e4, 5e4, 1e5$ and $5e5$ iterations reaches a relative L2 error of $1.58e-1$ at $Re = 3200$. PirateNets (Wang et al., 2024) uses a curriculum through $100, 400, 1000, 1600, 3200$ with $1e4, 2e4, 5e4, 5e4$ and $5e5$ iterations, which sums to the 630 thousand steps the anchor study cites, and reports $4.21e-2$. The anchor study's own methodology section notes that parameter continuation is common in PINN curricula and cites Krishnapriyan et al. (2021) for it. Jangir et al. (2026) report that physics-only PINNs often lose accuracy in convection-dominated high-Reynolds-number flows, and, for a parameterised cavity PINN trained over Re 500 to 1000, that about 5 percent supervised CFD data suffices for accurate prediction. Takao and Ii (2025) report that, in an inverse setting with interior reference data, initialising from a pretrained cavity PINN speeds

convergence relative to random initialisation, more so when the pretraining geometry is similar.

Figure 1a puts the published errors on one axis. The arms that fail, the from-scratch PINN (0.98) and the no-continuation ablation (about 0.73), share one feature: neither walks a path in Reynolds number. The arms that succeed, the expert's-guide curriculum (0.158), the anchor framework (0.0663) and PirateNet (0.0421), all do. Wang et al. (2023b) add results at nearby Reynolds numbers that qualify the pattern. With no labelled data, their two-network entropy-viscosity PINN at $Re = 3000$ reduces the velocity error "to less than 3%"; it anneals the artificial-viscosity coefficient from 0.05 to 0.002 across training stages, which is a continuation in artificial viscosity rather than in Reynolds number, a homotopy of the same family as a Reynolds path. Transfer learning along a Reynolds path, initialising a $Re = 3000$ network from one trained at $Re = 2000$, does worse, with a u-velocity error of 10.26 percent. And at $Re = 2000$ a single labelled point, with no path at all, brings the velocity error under 1 percent (Table 1). The operator-distillation arm sits between the two curriculum PINNs in error, uses fewer steps and parameters than PirateNet, and uses about 2.5 times as many point evaluations.

Three cautions apply to reading Figure 1a as a ranking. The lid regularisations differ: the anchor study uses a smootherstep margin at each corner, PirateNet a hyperbolic-cosine profile with $C_0 = 50$, and the expert's guide reports a uniform lid. The anchor study states that its anchor solves are checked against the Marchi et al. (2021) benchmark with "the few-percent offset expected from the regularized-versus-singular lid" accounted for. An offset of a few percent is the same order as the gap between the anchor framework and PirateNet. Second, the comparisons use different networks, optimisers and budgets. Third, PirateNet's architecture is itself a remedy for the optimisation pathologies of Section 3.3, so it confounds architecture with curriculum.

What the figure does support is narrower and, for the anchor study's headline, more important. None of the results collected here shows an operator teacher rescuing a cavity PINN that a Reynolds-number curriculum would not also have rescued. The relevant comparator, the same Part 3 student trained through the same four-rung ladder 1400, 2000, 2500, 3200 with no teacher at all, is not reported. Until it is, the cavity experiment shows that continuation plus a teacher works, that continuation without a teacher works in other hands, and that a teacher without continuation does not. It does not isolate a contribution of the teacher.

The anchor study's efficiency comparison, 250 thousand Part 3 steps against PirateNet's 630 thousand, needs a similar caveat, taken up in Section 6.3.

6. Is a PINN the right refiner?

Suppose the basin has been selected. The anchor study's design then hands the field to a PINN to descend into it. The obvious alternative is to hand it to the solver that produced the anchors and the scoring reference.

6.1 The warm-start literature

A substantial literature uses neural-operator predictions as initial guesses for classical solvers and reports the gains in iteration counts and wall-clock time rather than in PINN accuracy. Zhou et al. (2023) use an equivariant neural operator to map low-fidelity to high-fidelity fields and warm-start steady-state flow solvers, reporting at least two-fold faster convergence without loss of accuracy across low-Reynolds cylinder flow, high-Reynolds airfoil flow and a wing. Aghili et al. (2024) train a Fourier neural operator with data and residual terms to produce initial guesses for Newton's method on nonlinear elliptic problems discretised by finite differences, and report large gains over naive initial guesses, particularly for strongly nonlinear and anisotropic cases. Eshaghi et al. (2025) report reductions of up to 90 percent in computational time for Krylov solvers warm-started by neural operators, while keeping the solvers' convergence guarantees. Taghikhani et al. (2025) refine a physics-informed neural field by Newton iterations in nonlinear finite-element solid mechanics. Wang et al. (2026) pretrain a physics-informed neural operator on unstructured point clouds and use its predictions, with relative errors around 1 percent, to warm-start finite-element solves, reporting speedups of up to an order of magnitude over zero initial guesses. In the linear setting, Zhang et al. (2022) interleave a DeepONet with relaxation sweeps so that the network handles the low-frequency error modes that relaxation handles poorly. Markidis (2021) reached the same division of labour from the PINN side: PINN linear solvers resolve low frequencies quickly and high frequencies slowly, and integrating them with traditional solvers produced solvers competitive with PETSc conjugate gradient. Learned components inside classical solvers, rather than in front of them, are reported to preserve convergence guarantees (Hsieh et al., 2019) and to correct discretisation error (Um et al., 2020; Kochkov et al., 2021).

These results share the anchor study's premise, that a cheap approximate field is valuable as a starting point, and differ from it in where the final accuracy comes from. In the warm-start design the refiner is a discretisation with known convergence order and a stopping criterion on its own residual. In the anchor design the refiner is a PINN whose final accuracy is set by its optimiser, architecture and collocation.

6.2 What a PINN refiner costs

The comparative evidence on PINNs as forward solvers is not favourable. Grossmann et al. (2023) compare PINNs with the finite element method on Poisson problems in one to three dimensions, a one-dimensional Allen-Cahn problem and semilinear Schrodinger problems, and report that PINNs did not beat FEM in solution time or accuracy in their study, although they were sometimes faster to evaluate once trained. Chuang and Barba (2022) report that about 32 hours of PINN training were needed to match the accuracy of a 16×16 finite-difference simulation of the Taylor-Green vortex that took under 20 seconds. Wight and Zhao (2020) report that applying a PINN directly to phase-field equations often fails to give accurate solutions and needs adaptive sampling in space and time.

The anchor study's own pipeline contains a candidate refiner that is likely to be cheaper; Study 3 in Section 11 would check. Its cavity reference is a finite-difference solver on a 129-point uniform grid, iterated to steady state at $Re = 3200$ with the regularised lid; its Allen-Cahn reference is a Newton solve on a five-point stencil. Both solvers produce the very fields against which the PINN is scored. For Allen-Cahn, the five-point Newton solve plausibly needs little more than an initial guess on the right side of the zero solution, which the seed provides; the anchor study does not say how that solve was initialised. For the cavity, the reference solver reached $Re = 3200$ on its own, by the anchor study's account; whether it did so from a cold start or by continuation is not stated.

This does not make the hybrid pointless. A PINN refiner earns its place where a mesh-based solver is awkward: irregular or high-dimensional domains, inverse problems, problems where the solution is wanted as a smooth function rather than a grid, or settings where data and physics must be combined (Karniadakis et al., 2021; Cuomo et al., 2022). None of the anchor study's three benchmarks is of that kind; all three are low-dimensional forward problems on rectangles with standard reference solvers. On those benchmarks, the second rule of McGreivy and Hakim, compare against an efficient numerical method, would ask whether the operator-plus-PINN pipeline is faster than the operator-plus-solver pipeline at equal accuracy, and the published evidence does not answer it.

6.3 When the amortisation argument holds

The anchor study's strongest economic argument is amortisation: Part 1 is paid once and reused across targets. The argument is sound in form, and the warm-start papers above make it too. Zhou et al. (2023) report overall time savings across multiple simulations even after counting model development. Eshaghi et al. (2025) motivate theirs by many-query, real-time and design tasks. The question is what the amortised pipeline should be compared with.

For a single target, the comparison is with a single-target method. The anchor study pays for Part 1 (operator training plus eight reference solves), Part 2 (four continuation rungs) and Part 3 (a 250-thousand-step PINN), against a curriculum PINN or a continuation solve that pays only for the path. For many targets, the comparison is with methods that are themselves amortised: a physics-informed operator trained across the whole range of interest (Wang et al., 2021; Li et al., 2021), a parameterised PINN that takes Reynolds number as an input (Jangir et al., 2026, report one), or a meta-learned PINN initialisation (Penwarden et al., 2021). Against those, the anchor's distinctive claim is extrapolation: its targets lie outside the training box, where an amortised operator alone would fail. That claim is real, but it is exactly the case in which Section 4.3 locates the work in a staged continuation rather than in the operator. On this reading, a many-target user who wants extrapolation pays one continuation per target either way.

None of this makes amortisation wrong. It means the anchor study's single-target comparison with PirateNet, 250 thousand Part 3 steps against 630 thousand, compares one stage of an amortised pipeline with the whole of an unamortised one. The end-to-end figure the study also gives, 546 thousand steps plus eight solves, is the like-for-like number for a single target.

7. The mirror image: accuracy is not sufficient either

The anchor study's headline is that accuracy is not necessary. Oh et al. (2026) report the converse, from the warm-start side. They note that neural operators are increasingly used to warm-start Newton solvers "on the premise that a low test error places the initial guess inside the basin of attraction", and that this premise is unreliable. In a nearly incompressible hyperelasticity problem, an operator trained to relative L2 error of order $1e-3$ produced initial states whose discrete Jacobian was indefinite. Mean-squared training controls the error on average and leaves localised pointwise violations of the physics, in their case a predicted volume change dispersed away from one. A short label-free fine-tuning phase against the discrete energy restored a positive-definite Jacobian, and with an inexact outer loop the warm-started Newton method converged across the full loading range, at up to 5.4 times the wall-clock speed of incremental continuation on a 3D problem with 6.4 million degrees of freedom. They report the same effect on the anchor study's own benchmark family: for a lid-driven cavity at $Re = 3000$, a warm-started Newton method fails from an operator trained only to a tolerance of $5e-2$ and converges from one trained to $1e-2$.

Put the two results side by side. A teacher 78 percent wrong in L2 was inside the basin of a gradient-trained PINN with a decaying distillation term. A teacher wrong by something of order 0.1 percent in L2 was outside the

basin of a Newton-CG solver. Both are consistent with a single statement: a basin is a property of the prior and the refiner together, not of the prior alone. L2 error is a summary statistic of the prior and carries no information about the refiner. The author's conjecture about why the two refiners differ is that a gradient-based PINN with a soft, decaying pull toward the teacher has a wide and forgiving basin, because the teacher only biases the early trajectory before the residual takes over. A Newton method with a Krylov inner solve, on Oh et al.'s account, has a basin bounded by the local spectrum of the Jacobian, which a small localised error can break.

For practice, the consequence is that "the operator need not be accurate" should not travel without the refiner attached. A reader who builds an operator-to-Newton pipeline on the anchor study's principle, and accepts a crude operator on its authority, risks the failure Oh et al. describe. The right currency for a prior is basin membership for the refiner in hand, and the right test is to run that refiner from that prior, not to measure the prior's distance from a reference.

The same point bears on why the no-Part-2 ablation fails while the full framework succeeds. The continued teacher is 53 percent wrong in L2. The direct teacher's own error is not reported, so the published record gives no L2 comparison between the two teachers; what it gives is the outcome, students that reach 93 and 27 percent accuracy. The author's conjecture is that something in the structure of the continued teacher, such as the placement of the primary vortex, which the anchor study reports mislocated in the ablated student, puts it on the right side of the PINN's basin boundary. If so, that structure is what continuation supplies, which is why Section 4.2 attributes the effect to the ladder.

8. What "refines beyond the operator's accuracy" can mean

The anchor study's abstract says the residual refines the solution "beyond the operator's accuracy". That is literally true in all three experiments: 0.0209 against 0.12, 0.0037 against 0.78, 0.0663 against 0.53. Two qualifications put it in proportion.

First, the ceiling is the PINN's, not the operator's. Wang and Lai (2023) report that networks in practice struggle to push prediction errors below about $1e-5$ even with large size and long training, and that multi-stage training, in which each stage fits the previous stage's residue, reaches near machine precision for regression and PINNs. Muller and Zeinhofer (2023) report that energy natural gradient descent gives errors several orders of magnitude smaller than standard optimisers. Rathore et al. (2024) report large gains from second-order methods. A distilled PINN inherits the floor of its own optimiser and architecture, whatever the teacher. The anchor study's cavity error of $6.63e-2$ sits more than three orders of magnitude above the $1e-5$ level at which

Wang and Lai say networks typically stall. The author's reading, an inference and not a measurement, is that it is set by the high-Reynolds flow and the training recipe rather than by the teacher.

Second, for the cavity the reported error is against a reference with its own approximations. The 129-point reference grid at $Re = 3200$ is coarse compared with the 601-point grid Erturk et al. (2005) used to compute steady solutions up to $Re = 21,000$ with residuals below $1e-10$, or the 8192-point multigrid with Richardson extrapolation that Marchi et al. (2021) used up to $Re = 10,000$. The anchor study lists both, alongside the spectral benchmark of Botella and Peyret (1998), as the standard yardsticks. The few-percent regularisation offset the anchor study mentions is the same order as the reported error. None of this means the PINN is inaccurate; it means that differences of a few percent between cavity results computed with different lids and references, including the 0.0663 against 0.0421 comparison with PirateNet, are within the uncertainty the setup itself introduces.

9. The strongest case for the anchor study's principle

The argument so far has pressed on the evidence, so the case for the principle should be stated at full strength.

First, the Allen-Cahn experiment is a clean demonstration of something non-trivial. Branch identity, placed in the operator's ansatz at λ between 35 and 80 and H between 1 and 2, survives a two-stage polynomial extrapolation to $\lambda = 12.5$ and $H = 3$, survives transfer into a network with a different architecture and no boundary lift, and survives the decay of the distillation weight to zero. In five of five seeds the student ends on the branch, and the from-scratch student ends at zero in five of five. That the information was hand-placed does not make its preservation trivial. Distillation into a fresh network could have lost the sign; the decaying weight could have released the student back toward the zero field; the bounded-correction cap could have prevented convergence. None of these happened.

Second, the parametric setting is where hand-placed information runs out. In a real design loop the modeller does not have a one-mode formula for every target, and the operator's value is that it carries branch information learned over a range to targets nobody has solved. The anchor study's frame, a neural operator as a reusable carrier of solution-branch information, is a reasonable bet for that setting, even if the three benchmarks do not exercise it.

Third, the initialisation literature supports the general direction. LIGO-PINN reports that learned initialisation alleviates catastrophic PINN failures, including convergence to trivial solutions, and argues that the role of initial weights has been under-investigated (Anurag et

al., 2026). Zou et al. (2025) find that initialisation shapes which solution a PINN finds. Takao and Ii (2025), in an inverse setting with interior reference data, and Penwarden et al. (2021) report faster convergence from transferred or meta-learned initialisations, and Desai et al. (2021) report one-shot transfer for linear problems. Zhu et al. (2022) report that fine-tuning a pretrained DeepONet with the PDE loss is one of several reliable routes to extrapolation, and Subramanian et al. (2023) report that pre-trained scientific models fine-tune to out-of-distribution physics with far fewer examples than training from scratch. The anchor study's distillation-with-decay is a reasonable member of this family, and its persistent but decaying teacher is a design choice that the warm-start papers cited here do not make.

Fourth, the decaying-teacher design addresses a plausible weakness of pure warm starts: a single initialisation may be forgotten early in a badly conditioned optimisation. Keeping the teacher in the loss for a fraction of training, and gating it where the student's residual is large, as the anchor study does for Burgers, plausibly keeps the trajectory in the basin longer than a single initialisation would. The anchor study does not report the comparison between warm start only and decaying distillation, so this advantage is plausible rather than shown.

None of these points answers the three questions of Sections 4 to 6. They establish that the questions are worth answering.

10. Consolidated evidence and a comparator procedure

Table 1 collects the published results the argument rests on, one row per result, with the source named on the row. Values are as reported by each source; "rel. L2" is relative L2 error.

Read down the "prior" column and a pattern appears. Every successful route in Table 1 to the right solution of a multiplicity or high-Reynolds problem uses one of five kinds of information: continuity along a parameter path (curricula, the anchor study's ladder, classical continuation), a symmetry-breaking or branch-identifying seed (the anchor study's sine seed, structure probing), a single observation (Wang et al., 2023b), a regularisation of the problem itself (their entropy viscosity), or explicit enumeration with a criterion (deflation, ensembles). The neural operator in the anchor study is a container for the first two. That is a legitimate engineering role. It is not evidence that the operator is the source.

Table 2 records, for each of the anchor study's experiments, where the basin information came from and which comparator is missing.

Algorithm 1 states the comparator procedure the author proposes for licensing a claim that a learned prior selects a basin. Each step names the arm whose result would

Table 1. Published results bearing on whether a crude prior selects a basin, and what else would select it.

Source	Problem	Prior or initialiser	Refiner	Reported result
Mousavi et al. (2026)	Burgers 1D, (0.05, 4)	extrapolated operator, err 0.12	distilled PINN	rel. L2 0.0209 vs 0.4123 from scratch
Mousavi et al. (2026)	Allen-Cahn 2D, lambda 12.5, H 3	operator with sin-sin seed, err 0.78	distilled PINN	0.0037 vs 1.00 (collapse to zero)
Mousavi et al. (2026)	cavity Re 3200	operator + 4-rung Re ladder, err 0.53	distilled PINN	0.0663 vs 0.9800 from scratch
Mousavi et al. (2026)	cavity Re 3200	operator, no ladder	distilled PINN	27.09% accuracy (err about 0.73)
Wang et al. (2023a)	cavity Re 3200	Re curriculum 100-400-1000-3200	PINN	rel. L2 1.58e-1
Wang et al. (2024)	cavity Re 3200	Re curriculum to 3200, 630K steps	PirateNet	rel. L2 4.21e-2
Wang et al. (2023b)	cavity Re 2000	one labelled point	PINN	unique DNS-like solution, velocity err < 1%
Zou et al. (2025)	Allen-Cahn 2D, +/-1 boundary data	50 random initialisations	PINN, then FEM	three distinct solutions found
Farrell et al. (2014)	Allen-Cahn 2D	zero guess + deflation	Newton	one unstable, two stable solutions
Disaro et al. (2026)	Landau-de Gennes, square	deflation loss	PINN, then gradient flow + Newton	all six stable states, one run
Disaro et al. (2026)	Allen-Cahn, unit disk, lambda 6	deflation loss	PINN, no classical refinement	all three solutions, one run
Oh et al. (2026)	hyperelasticity 2D/3D	operator, rel. L2 O(1e-3)	Newton-Krylov	indefinite Jacobian; fixed by label-free tuning, up to 5.4x
Eshaghi et al. (2025)	PDE benchmarks, Krylov solves	neural operator	CG, GMRES	up to 90% less compute time
Zhou et al. (2023)	steady flows: cylinders, airfoils, wing	neural operator	steady solver	at least 2x faster convergence
Wang et al. (2026)	elasticity, hyperelasticity	PI neural operator, err about 1%	FEM	up to 10x speedup vs zero guess
Krishnapriyan et al. (2021)	convection, reaction	curriculum regularisation	PINN	1-2 orders lower error
Chuang and Barba (2022)	Taylor-Green, cylinder	none	PINN vs FD	32 h vs < 20 s; no vortex shedding
McGreivy and Hakim (2024)	literature review	n/a	n/a	60 of 76 claims vs weak baselines

Table 2. Provenance of basin information in the three experiments of Mousavi et al. (2026), from the study's own appendices.

Experiment	Multiplicity in PDE?	Hand-supplied branch information	External solves	Parameter path	Missing comparator
Burgers	no (unique, per the study)	none needed	none (Cole-Hopf reference)	single shot	meta-learned or transferred PINN init
Allen-Cahn	yes (zero field exact)	sin-sin seed lift $A_0 = 0.3$; amplitude floor	Gauss-Newton anchor polish, no reference	two rungs, 18 then 12.5	PINN distilled from seed alone
Cavity	not established; see Section 3.4	none reported	8 FD solves in base box	four rungs to 3200	same PINN, same ladder, no teacher

decide it.

The procedure is deliberately ordered so that the cheapest alternatives are tried first. Arm A1 would settle the cavity question of Section 5, arm A2 the Allen-Cahn question of Section 4.1, and arm A4 the refiner question of Section 6. Step S1, made with an independent solver check, would settle which cell of Section 3 the cold-start failure belongs to. None of the five arms requires new methodology; all of them use components the anchor study already built.

11. Studies that would settle the open part

Five studies, each using components that already exist, would settle what the published record leaves open.

Study 1: the seed-only teacher. On the anchor study's Allen-Cahn target, distil the Part 3 PINN from A_0 ϕ_1 alone, and separately from the one-mode estimate a ϕ_1 with $a = 0.467$, using the same schedule and seeds. If either recovers the branch in five of five seeds at comparable accuracy, the experiment would show that the operator preserves branch information but adds nothing a closed form does not. If both fail where the operator teacher succeeds, the operator's higher-mode content is doing work, and the anchor study's principle gains real support.

Algorithm 1. Licensing a claim that a learned prior selects the basin

```

Input: target problem  $P^*$ , learned prior  $T$ , refiner  $R$ , budget  $B$ 
S1  classify the failure of  $R$  from a cold start on  $P^*$ :
      (a) multiplicity of the PDE (Sec 3.1):  $\geq 2$  exact solutions,
          shown by theory, deflation, or ensemble + solver check
      (b) spurious loss minimiser (Sec 3.2): cold-start end point
          fails an independent solver check or the boundary data
      (c) optimisation stall (Sec 3.3): residual not small at  $B$ 
      if (c) only: the claim is acceleration, not basin selection; stop
S2  arm A0:  $R$  from a cold start (lower bound)
S3  arm A1:  $R$  along the parameter path  $T$  was given,
          with no  $T$  (continuation)
S4  arm A2:  $R$  from the cheapest hand-made prior that carries the
          same branch information as  $T$  (seed, symmetry, one
          observation) (cheap prior)
S5  arm A3:  $R$  from  $T$  (the claim)
S6  arm A4: an efficient classical solver from  $T$  and from the A2
          prior, run to the accuracy of A3 (refiner check)
S7  license " $T$  selects the basin for  $R$ " only if
      A3 succeeds where A0 fails, A3 beats A1 and A2 at equal budget,
      and the end point of A3 passes an independent solver check
      license "the hybrid pays" only if A3 also beats A4 at equal
      accuracy, counting the cost of training  $T$  and of every
      external solve that built  $T$ 

```

Study 2: the teacherless ladder. On the cavity, train the same Part 3 primitive-variable PINN through 1400, 2000, 2500, 3200 with the same total step budget and no distillation term. The reported 93-against-27 split in the ablation says the ladder matters. This study would say whether the teacher matters once the ladder is present.

Study 3: the classical refiner. For each target, start the anchor study's own reference solver from the Part 2 field and from a cold start, and report iterations and wall-clock time to the reference tolerance. Compare with the Part 3 PINN's wall-clock time to its reported accuracy. This is McGreivy and Hakim's second rule applied to the hybrid's own components.

Study 4: an independent check of the wrong basin. Report the boundary-condition losses of the from-scratch and no-ladder cavity students, and hand their converged fields to two classical solvers: a Newton solve of the discretised steady equations, and a pseudo-time-marching solve. If Newton converges to a root near the PINN field, that field is close to a genuine steady solution; if time-marching then leaves it for the reference, the solution is unstable, and the cavity belongs to cell 3.1 with stability as the criterion, as Wang et al. (2023b)'s description suggests. If Newton moves to the reference or fails to converge, the cold-start failure was a spurious minimiser or a stall (Sections 3.2 and 3.3), and "basin selection" should be replaced by "loss-landscape rescue" for this experiment. Either answer is informative, and the two solvers are already in the anchor study's pipeline in some form.

Study 5: a problem where hand calculation fails. Test the framework on a multiplicity problem without a one-mode formula: a secondary bifurcation, a degenerate eigenvalue, or a branch far from onset, with the target chosen so that the branch's identity at the target cannot be read off the training box. This is the setting the

parametric argument of Section 9 needs, and none of the three benchmarks provides it. Deflation-based enumeration (Farrell et al., 2014; Disaro et al., 2026) gives the ground truth for which branches exist.

12. What is not known

Whether an operator teacher contributes anything to the Allen-Cahn result beyond the seed it was built on is not known. The one-mode estimate makes it plausible that it does not, but the seed-only arm has not been run.

Whether an operator teacher contributes anything to the cavity result beyond the Reynolds-number ladder is not known. No result the author found trains the same student through the same ladder without a teacher.

Whether the failed PINN arms on the regularised cavity at $Re = 3200$ sit near a genuine, unstable second steady solution or at spurious minimisers of the penalised loss is not known. The anchor study's residual comparison does not distinguish the cases, and boundary losses were not reported.

Whether the operator-plus-PINN pipeline is cheaper than the operator-plus-classical-solver pipeline at equal accuracy, on any of the three benchmarks, is not known.

How wide the basin of a decaying-distillation PINN is, measured as the largest teacher error from which the branch is still recovered, is not known for any problem. The anchor study shows one point inside it per experiment.

Whether the robustness to a crude teacher reported for gradient-trained PINNs carries over to Newton refiners is known to fail in at least one case (Oh et al., 2026) and is not known in general.

13. Limitations

This analysis rests heavily on one preprint, posted a month before this paper; its numbers are taken from its text and tables. Of the two internal inconsistencies noted in Section 2, the first is resolved by preferring the table and the second is flagged, not resolved. The one-mode amplitude estimate is a leading-order approximation; its agreement with the anchor study's reference to about 3 percent supports its use here, but the 0.36 relative distance attributed to the seed is computed against the one-mode branch, not against the reference field, and the true distance could differ by a few hundredths. The comparison in Figure 1a juxtaposes results computed with different lid regularisations, networks, optimisers, references and budgets, and is presented as an illustration of which arms walk a parameter path, not as a ranking. The warm-start results in Section 6.1 come from different problems, mostly elliptic, solid-mechanics and steady-flow problems, and are evidence that the classical-refiner route exists, not that it would win on the anchor study's benchmarks. Several sources are 2025 and 2026 preprints. Finally, Section 3.4 leaves the cavity's cold-start failure unplaced between a genuine unstable steady state, a spurious minimiser of the penalised loss, and an optimisation stall. It rests on Wang et al. (2023b)'s description of the second PINN solution class and on the benchmark literature's practice of reporting one steady solution per Reynolds number. The author does not prefer one reading, and Study 4 would settle it.

14. Conclusion

The anchor study's principle, that a crude prior can be useful because it only needs to pick the basin, is correct and has a long history. It is why Newton's method is so often run along a continuation path, and why deflation exists. What the study's evidence does not establish is that the neural operator is where the basin information came from, that a PINN trained from scratch is the baseline that decides the question, or that a PINN is the refiner one should use once the basin is known. In the Allen-Cahn experiment, the branch information was placed in the operator's ansatz by its builders, and a one-line formula reproduces the target branch's amplitude to about 3 percent. In the cavity experiment, the component whose removal breaks the result is a staged continuation in Reynolds number, and continuation without an operator reaches comparable published accuracy. On the two benchmarks without a closed-form reference, a standard solver was already in the pipeline.

The opposite failure is published as well. A prior whose error is of order a tenth of a percent can still start Newton's method outside its basin. Taken together, the two results say that "need not be accurate" and "need not be close" are properties of a prior paired with a refiner, and that L2 error is the wrong currency for either. The five comparator arms of Algorithm 1 are cheap, use only components the anchor study already built, and would

turn its principle from a plausible principle into a licensed claim, or show where it stops.

References

- Aghili, J., Franck, E., Hild, R., Michel-Dansac, V., & Vigon, V. (2024). Accelerating the convergence of Newton's method for nonlinear elliptic PDEs using Fourier neural operators. *Communications in Nonlinear Science and Numerical Simulation*. arXiv:2403.03021. doi:10.48550/arXiv.2403.03021.
- Allgower, E. L., & Georg, K. (2003). *Introduction to Numerical Continuation Methods*. Society for Industrial and Applied Mathematics. doi:10.1137/1.9780898719154.
- Anurag, N., Adhikari, S., Kapoor, T., & Muralidhar, N. (2026). LIGO-PINN: Learned Initialization via Gated Optimization to Alleviate Convergence Failures in Physics Informed Neural Networks. arXiv:2607.14233. doi:10.48550/arXiv.2607.14233. [PREPRINT]
- Basir, S., & Senocak, I. (2022). Critical Investigation of Failure Modes in Physics-informed Neural Networks. arXiv:2206.09961. doi:10.48550/arXiv.2206.09961. [PREPRINT]
- Botella, O., & Peyret, R. (1998). Benchmark spectral results on the lid-driven cavity flow. *Computers and Fluids*, 27(4), 421-433. doi:10.1016/S0045-7930(98)00002-4.
- Chafee, N., & Infante, E. F. (1974). A bifurcation problem for a nonlinear partial differential equation of parabolic type. *Applicable Analysis*, 4(1), 17-37. doi:10.1080/00036817408839081.
- Chuang, P.-Y., & Barba, L. A. (2022). Experience report of physics-informed neural networks in fluid simulations: pitfalls and frustration. arXiv:2205.14249. doi:10.48550/arXiv.2205.14249. [PREPRINT]
- Crandall, M. G., & Rabinowitz, P. H. (1971). Bifurcation from simple eigenvalues. *Journal of Functional Analysis*, 8(2), 321-340. doi:10.1016/0022-1236(71)90015-2.
- Cuomo, S., Schiano di Cola, V., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific Machine Learning through Physics-Informed Neural Networks: Where we are and What's next. arXiv:2201.05624. doi:10.48550/arXiv.2201.05624.
- Daw, A., Bu, J., Wang, S., Perdikaris, P., & Karpatne, A. (2022). Mitigating Propagation Failures in Physics-informed Neural Networks using Retain-Resample-Release (R3) Sampling. arXiv:2207.02338. doi:10.48550/arXiv.2207.02338.
- Desai, S., Mattheakis, M., Joy, H., Protopapas, P., & Roberts, S. (2021). One-Shot Transfer Learning of Physics-Informed Neural Networks. arXiv:2110.11286. doi:10.48550/arXiv.2110.11286. [PREPRINT]
- Disaro, S., Maity, R. R., & Bacho, A. (2026). Deflation-PINNs: Learning Multiple Solutions for PDEs and Landau-de Gennes. arXiv:2603.27936. doi:10.48550/arXiv.2603.27936. [PREPRINT]
- Erturk, E., Corke, T. C., & Gokcol, C. (2005). Numerical solutions of 2-D steady incompressible driven cavity flow at high Reynolds numbers. *International Journal for Numerical Methods in Fluids*, 48(7), 747-774. doi:10.1002/fld.953.
- Eshaghi, M. S., Anitescu, C., Valizadeh, N., Wang, Y., Zhuang, X., & Rabczuk, T. (2025). Nows: Neural Operator Warm Starts for Accelerating Iterative Solvers. arXiv:2511.02481. doi:10.48550/arXiv.2511.02481.
- Farrell, P. E., Birkisson, A., & Funke, S. W. (2014). Deflation techniques for finding distinct solutions of nonlinear partial differential equations. arXiv:1410.5620. doi:10.48550/arXiv.1410.5620.
- Ghia, U., Ghia, K. N., & Shin, C. T. (1982). High-Re solutions for incompressible flow using the Navier-Stokes equations and a multigrid method. *Journal of Computational Physics*, 48(3), 387-411. doi:10.1016/0021-9991(82)90058-4.
- Goswami, S., Kontolati, K., Shields, M. D., & Karniadakis, G. E. (2022). Deep transfer operator learning for partial differential equations under conditional shift. *Nature Machine Intelligence*.

- arXiv:2204.09810. doi:10.48550/arXiv.2204.09810.
- Grossmann, T. G., Komorowska, U. J., Latz, J., & Schonlieb, C.-B. (2023). Can Physics-Informed Neural Networks beat the Finite Element Method? arXiv:2302.04107. doi:10.48550/arXiv.2302.04107.
- Gu, Y., Wang, C., & Yang, H. (2020). Structure Probing Neural Network Deflation. *Journal of Computational Physics*. arXiv:2007.03609. doi:10.48550/arXiv.2007.03609.
- Hao, W., Liu, X., & Yang, Y. (2024). Newton Informed Neural Operator for Computing Multiple Solutions of Nonlinear Partial Differential Equations. arXiv:2405.14096. doi:10.48550/arXiv.2405.14096. [PREPRINT]
- Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the Knowledge in a Neural Network. arXiv:1503.02531. doi:10.48550/arXiv.1503.02531. [PREPRINT]
- Hsieh, J.-T., Zhao, S., Eismann, S., Mirabella, L., & Ermon, S. (2019). Learning Neural PDE Solvers with Convergence Guarantees. arXiv:1906.01200. doi:10.48550/arXiv.1906.01200.
- Jangir, A., Clements, R., Goyal, R., & Tabor, G. (2026). Sparse-Supervised Hybrid Parameterized Physics-Informed Neural Networks for Incompressible Flows Across Reynolds Numbers. *Physics of Fluids*. arXiv:2602.04670. doi:10.48550/arXiv.2602.04670.
- Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3(6), 422-440. doi:10.1038/s42254-021-00314-5.
- Kochkov, D., Smith, J. A., Alieva, A., Wang, Q., Brenner, M. P., & Hoyer, S. (2021). Machine learning accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*. arXiv:2102.01010. doi:10.48550/arXiv.2102.01010.
- Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2021). Neural Operator: Learning Maps Between Function Spaces. *Journal of Machine Learning Research*. arXiv:2108.08481. doi:10.48550/arXiv.2108.08481.
- Krishnapriyan, A. S., Gholami, A., Zhe, S., Kirby, R. M., & Mahoney, M. W. (2021). Characterizing possible failure modes in physics-informed neural networks. *NeurIPS 2021*. arXiv:2109.01050. doi:10.48550/arXiv.2109.01050.
- Lara Benitez, J. A., Furuya, T., Faucher, F., Kratsios, A., Tricoche, X., & de Hoop, M. V. (2023). Out-of-distribution risk bounds for neural operators with applications to the Helmholtz equation. *Journal of Computational Physics*. arXiv:2301.11509. doi:10.48550/arXiv.2301.11509.
- Leiteritz, R., & Pfluger, D. (2021). How to Avoid Trivial Solutions in Physics-Informed Neural Networks. arXiv:2112.05620. doi:10.48550/arXiv.2112.05620. [PREPRINT]
- Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). Fourier Neural Operator for Parametric Partial Differential Equations. arXiv:2010.08895. doi:10.48550/arXiv.2010.08895.
- Li, Z., Zheng, H., Kovachki, N., Jin, D., Chen, H., Liu, B., Azizzadenesheli, K., & Anandkumar, A. (2021). Physics-Informed Neural Operator for Learning Partial Differential Equations. arXiv:2111.03794. doi:10.48550/arXiv.2111.03794.
- Lu, L., Jin, P., & Karniadakis, G. E. (2019). DeepONet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *Nature Machine Intelligence*. arXiv:1910.03193. doi:10.48550/arXiv.1910.03193.
- Marchi, C. H., Santiago, C. D., & de Carvalho Jr., C. A. R. (2021). Lid-Driven Square Cavity Flow: A Benchmark Solution With an 8192 x 8192 Grid. *Journal of Verification, Validation and Uncertainty Quantification*, 6(4). doi:10.1115/1.4052149.
- Markidis, S. (2021). The Old and the New: Can Physics-Informed Deep-Learning Replace Traditional Linear Solvers? arXiv:2103.09655. doi:10.48550/arXiv.2103.09655.
- McGreivy, N., & Hakim, A. (2024). Weak baselines and reporting biases lead to overoptimism in machine learning for fluid-related partial differential equations. *Nature Machine Intelligence*. arXiv:2407.07218. doi:10.48550/arXiv.2407.07218.
- Mousavi, S. M., Kadeethum, T., Bouklas, N., & Goswami, S. (2026). Learning Physics from an Imperfect Ancestor. arXiv:2609.24947. doi:10.48550/arXiv.2609.24947. [PREPRINT]
- Muller, J., & Zeinhofer, M. (2023). Achieving High Accuracy with PINNs via Energy Natural Gradients. arXiv:2302.13163. doi:10.48550/arXiv.2302.13163.
- Oh, J., Lee, Y., Darbon, J., & Karniadakis, G. E. (2026). Spectrally Safe Neural Operator Warm-Starts for Large-Scale Newton Solvers. arXiv:2606.21828. doi:10.48550/arXiv.2606.21828. [PREPRINT]
- Penwarden, M., Zhe, S., Narayan, A., & Kirby, R. M. (2021). A Metalearning Approach for Physics-Informed Neural Networks (PINNs): Application to Parameterized PDEs. *Journal of Computational Physics*. arXiv:2110.13361. doi:10.48550/arXiv.2110.13361.
- Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F. A., Bengio, Y., & Courville, A. (2018). On the Spectral Bias of Neural Networks. *ICML 2019*. arXiv:1806.08734. doi:10.48550/arXiv.1806.08734.
- Raissi, M., Perdikaris, P., & 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. doi:10.1016/j.jcp.2018.10.045.
- Rathore, P., Lei, W., Frangella, Z., Lu, L., & Udell, M. (2024). Challenges in Training PINNs: A Loss Landscape Perspective. arXiv:2402.01868. doi:10.48550/arXiv.2402.01868.
- Serrano, L., Han, J., Oyallon, E., Ho, S., & Morel, R. (2026). Test-time Generalization for Physics through Neural Operator Splitting. arXiv:2602.00884. doi:10.48550/arXiv.2602.00884. [PREPRINT]
- Subramanian, S., Harrington, P., Keutzer, K., Bhimji, W., Morozov, D., Mahoney, M., & Gholami, A. (2023). Towards Foundation Models for Scientific Machine Learning: Characterizing Scaling and Transfer Behavior. *NeurIPS 2023*. arXiv:2306.00258. doi:10.48550/arXiv.2306.00258.
- Taghikhani, K., Yamazaki, Y., Varghese, J. P., Apel, M., Najian Asl, R., & Rezaei, S. (2025). Neural-Initialized Newton: Accelerating Nonlinear Finite Elements via Operator Learning. arXiv:2511.06802. doi:10.48550/arXiv.2511.06802. [PREPRINT]
- Takao, R., & Ii, S. (2025). Fine-tuning physics-informed neural networks for cavity flows using coordinate transformation. arXiv:2508.01122. doi:10.48550/arXiv.2508.01122. [PREPRINT]
- Tancik, M., Srinivasan, P. P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J. T., & Ng, R. (2020). Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. *NeurIPS 2020*. arXiv:2006.10739. doi:10.48550/arXiv.2006.10739.
- Um, K., Brand, R., Fei, Y., Holl, P., & Thurey, N. (2020). Solver-in-the-Loop: Learning from Differentiable Physics to Interact with Iterative PDE-Solvers. *NeurIPS 2020*. arXiv:2007.00016. doi:10.48550/arXiv.2007.00016.
- Wang, S., Teng, Y., & Perdikaris, P. (2020a). Understanding and mitigating gradient pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*. arXiv:2001.04536. doi:10.48550/arXiv.2001.04536.
- Wang, S., Yu, X., & Perdikaris, P. (2020b). When and why PINNs fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*. arXiv:2007.14527. doi:10.48550/arXiv.2007.14527.
- Wang, S., Wang, H., & Perdikaris, P. (2021). Learning the solution operator of parametric partial differential equations with physics-informed DeepOnets. *Science Advances*. arXiv:2103.10974. doi:10.48550/arXiv.2103.10974.

- Wang, S., Sankaran, S., & Perdikaris, P. (2022). Respecting causality is all you need for training physics-informed neural networks. arXiv:2203.07404. doi:10.48550/arXiv.2203.07404.
- Wang, S., Sankaran, S., Wang, H., & Perdikaris, P. (2023a). An Expert's Guide to Training Physics-informed Neural Networks. arXiv:2308.08468. doi:10.48550/arXiv.2308.08468. [PREPRINT]
- Wang, Z., Meng, X., Jiang, X., Xiang, H., & Karniadakis, G. E. (2023b). Solution multiplicity and effects of data and eddy viscosity on Navier-Stokes solutions inferred by physics-informed neural networks. arXiv:2309.06010. doi:10.48550/arXiv.2309.06010. [PREPRINT]
- Wang, S., Li, B., Chen, Y., & Perdikaris, P. (2024). PirateNets: Physics-informed Deep Learning with Residual Adaptive Networks. Journal of Machine Learning Research. arXiv:2402.00326. doi:10.48550/arXiv.2402.00326.
- Wang, Y., & Lai, C.-Y. (2023). Multi-stage Neural Networks: Function Approximator of Machine Precision. arXiv:2307.08934. doi:10.48550/arXiv.2307.08934.
- Wang, Y., Hao, Z., Eshaghi, M. S., Anitescu, C., Zhuang, X., Rabczuk, T., & Liu, Y. (2026). Pretrain Finite Element Method: A Pretraining and Warm-start Framework for PDEs via Physics-Informed Neural Operators. arXiv:2601.03086. doi:10.48550/arXiv.2601.03086. [PREPRINT]
- Wight, C. L., & Zhao, J. (2020). Solving Allen-Cahn and Cahn-Hilliard Equations using the Adaptive Physics Informed Neural Networks. arXiv:2007.04542. doi:10.48550/arXiv.2007.04542.
- Zhang, E., Kahana, A., Kopanicakova, A., Turkel, E., Ranade, R., Pathak, J., & Karniadakis, G. E. (2022). Blending Neural Operators and Relaxation Methods in PDE Numerical Solvers. arXiv:2208.13273. doi:10.48550/arXiv.2208.13273.
- Zheng, H., Huang, Y., Huang, Z., Hao, W., & Lin, G. (2023). HomPINNs: homotopy physics-informed neural networks for solving the inverse problems of nonlinear differential equations with multiple solutions. Journal of Computational Physics. arXiv:2304.02811. doi:10.48550/arXiv.2304.02811.
- Zhou, X.-H., Han, J., Zafar, M. I., Wolf, E. M., Schrock, C. R., Roy, C. J., & Xiao, H. (2023). Neural operator-based super-fidelity: A warm-start approach for accelerating steady-state simulations. arXiv:2312.11842. doi:10.48550/arXiv.2312.11842.
- Zhu, M., Zhang, H., Jiao, A., Karniadakis, G. E., & Lu, L. (2022). Reliable extrapolation of deep neural operators informed by physics or sparse observations. Computer Methods in Applied Mechanics and Engineering. arXiv:2212.06347. doi:10.48550/arXiv.2212.06347.
- Zou, Z., Wang, Z., & Karniadakis, G. E. (2025). Learning and discovering multiple solutions using physics-informed neural networks with random initialization and deep ensemble. arXiv:2503.06320. doi:10.48550/arXiv.2503.06320. [PREPRINT]