Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Physics-Informed Neural Networks for Viscous Flow Past a NACA 0012 Airfoil at Zero Angle of Attack: A Comparative Study of Standard and Curriculum Learning Training Strategies

Dinesh Sai Palli

palli_b231153me@nitc.ac.in

National Institute of Technology Calicut

Sai Saketha Chandra Athkuri

saisakethachandra@nitc.ac.in

National Institute of Technology Calicut

Research Article

Keywords: physics-informed neural networks, curriculum learning, numerical continuation, NACA0012, incompressible Navier–Stokes, Reynolds-number scaling, network interpretability

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11617316/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Physics-Informed Neural Networks for Viscous Flow Past a NACA 0012 Airfoil at Zero Angle of Attack: A Comparative Study of Standard and Curriculum Learning Training Strategies

Dinesh Sai Palli* and Sai Saketha Chandra Athkuri†
Department of Mechanical Engineering, National Institute of Technology Calicut, Kerala 673601, India

Physics-informed neural networks (PINNs) can solve the Navier–Stokes equations without a mesh, yet at moderate Reynolds numbers they often converge to fields with a small training loss and almost no pressure. We study this problem for steady, two-dimensional, incompressible flow past a NACA 0012 airfoil at $\alpha = 0^\circ$ and Re $\in \{100, 500, 1000, 2000, 5000, 10{,}000\}$, comparing two ways of training the same network (eight hidden layers of twenty tanh neurons, velocity–pressure formulation). The standard PINN is trained from a new Glorot initialization at each Reynolds number. The curriculum PINN is trained once along Re = 500 → 1000 → 2000 → 5000 → 10,000, with every stage starting from the weights of the previous one; numerically, this is a natural-parameter continuation in Re. At matched Reynolds numbers the curriculum network finishes with a loss 3.2 to 28.5 times lower and gives a physical pressure coefficient at all five stages, with stagnation peaks between +0.87 and +1.15. The standard network is physical only at Re = 100 and Re = 1000, and above Re = 2000 its C_p is of order 10^{-3}. The curriculum C_p agrees closely with ANSYS Fluent at Re = 500, 1000, 5000 and 10,000. Inside the network, the curriculum PINN keeps more of its capacity in use (54.47% sparsity against 83.28% at Re = 1000), its deepest layer becomes more active at every stage, and its x-momentum residual is lower at every matched Reynolds number. Its neurons also become less correlated with pressure as Re increases: the mean correlation magnitude drops by 83%, from 0.407 to 0.070, while the velocity correlation stays near 0.10. We read this as the network following the thinning of the boundary layer. Taken together, the measurements explain why warm-started training works where independent training does not, rather than only showing that it does.

Keywords: physics-informed neural networks; curriculum learning; numerical continuation; NACA 0012; incompressible Navier–Stokes; Reynolds-number scaling; network interpretability

*B.Tech. Student, Department of Mechanical Engineering, National Institute of Technology Calicut, Kerala 673601, India.
†Assistant Professor, Department of Mechanical Engineering, National Institute of Technology Calicut, Kerala 673601, India.

Nomenclature

x, y	streamwise and transverse coordinates, scaled by chord c
u, v	streamwise and transverse velocity, scaled by U_∞
p	static pressure, scaled by ρU_∞^2
Re	Reynolds number, $U_\infty c/\nu$
U_∞	freestream speed (= 1 in scaled form)
c	chord length (= 1 in scaled form)
ν	kinematic viscosity (= 1/Re in scaled form)
ρ	density (= 1 in scaled form)
α	angle of attack
C_p	pressure coefficient
$\Theta(\boldsymbol{x};\theta)$	feed-forward network with trainable parameters θ
θ	vector of weights and biases
$\mathcal{R}_c, \mathcal{R}_x, \mathcal{R}_y$	continuity, x-momentum and y-momentum residuals
$\mathcal{L}_{\text{total}}, \mathcal{L}_{\text{NS}}, \mathcal{L}_{\text{BC}}$	total, Navier–Stokes residual and boundary-condition losses
w	boundary-condition loss weight
$N_{\text{int}}, N_{\text{BC}}$	numbers of interior collocation and boundary points
N_{e}	number of points used to evaluate the diagnostics
$\mathcal{L}_0^{(s)}$	total loss at the start of curriculum stage s
$h_{\ell j}$	activation of neuron j in hidden layer ℓ
S_ℓ	sparsity of layer ℓ (%)
$\overline{a}_{\ell j}, \overline{a}_\ell$	mean absolute activation of neuron (ℓ, j) and of layer ℓ
$\overline{a}_{\ell,s}$	$\overline{a}_\ell$ at the end of curriculum stage s
$r_{\ell j,q}$	Pearson correlation of neuron (ℓ, j) with quantity q
$\overline{\lvert r\rvert}_q$	mean of $\lvert r_{\ell j,q}\rvert$ over the 160 hidden neurons
RMS$(\cdot)$	root-mean-square operator
Subscript s	curriculum stage index

I. Introduction

Aerodynamic design depends on knowing the flow around a lifting surface. The surface pressure and skin friction set the lift and drag, and the surrounding velocity field governs how the boundary layer grows, whether it separates and how the wake develops. Closed-form solutions of the governing equations exist only for idealized cases, so in practice these quantities are obtained by solving the Navier–Stokes equations numerically with computational fluid dynamics (CFD). In design work a single solution is rarely enough. Shape optimization and sweeps over Reynolds number and angle of attack all require the flow to be recomputed at many operating points.

Mesh-based solvers such as ANSYS Fluent remain the standard tool for this, and for good reason: they are mature, accurate and extensively validated. Their drawback in repeated evaluations is cost. Each new geometry or operating condition needs a new or deformed mesh and a fresh nonlinear solve. As the Reynolds number rises, the boundary layer becomes thinner, roughly in proportion to $\text{Re}^{-1/2}$ [1], so the near-wall grid must be refined further and the solution takes longer to converge. In gradient-based shape optimization the mesh must also follow every change of the surface, and even with efficient adjoint methods [2] each design iteration still needs a complete flow solution. Over hundreds or thousands of evaluations these costs add up quickly. This motivates mesh-free, differentiable models that are cheap to evaluate once trained and can complement conventional CFD.

Physics-informed neural networks (PINNs) [3] are one such approach. A neural network takes the spatial coordinates as input and returns the flow variables. Instead of being fitted to precomputed solutions, it is trained to minimize the residuals of the governing equations at a set of collocation points, together with the mismatch in the boundary conditions. The derivatives in these residuals are computed exactly by automatic differentiation, so no mesh is required and no interior solution data are needed. PINNs have been applied to many forward and inverse problems in fluid mechanics [4, 5, 6], including the reconstruction of velocity and pressure fields from flow visualizations [7], the direct solution of the incompressible Navier–Stokes equations [8] and Reynolds-averaged flow over airfoils [9].

PINNs also have a well-documented weakness. As the Reynolds number increases, a network trained by minimizing the combined loss often converges to a solution that lowers the loss without satisfying the physics [10]. Part of the reason is an imbalance between the gradients of the different loss terms during training [11]. Another is that the pressure

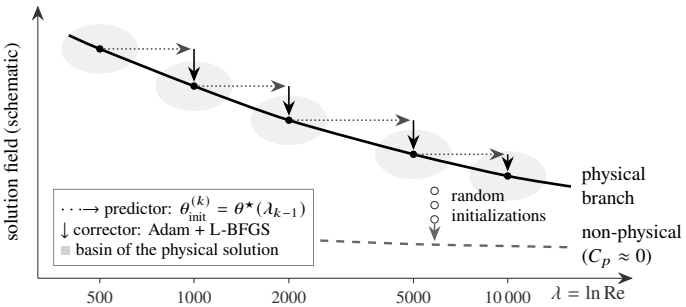


Fig. 1 Curriculum training viewed as natural-parameter continuation in $\lambda = \ln \text{Re}$ (schematic). Each stage starts from the converged solution of the previous one (dotted arrows) and the optimizer corrects it back onto the physical branch. A random initialization at high Reynolds number has no reason to land inside the basin of the physical solution and may instead converge to a non-physical branch.

enters the momentum equations only through its gradient, so the loss is only weakly sensitive to it. For airfoil flow the result can be deceptive: the training loss is small and the velocity field looks reasonable, but the predicted pressure has no stagnation peak at the leading edge. A failure of this kind cannot be seen in the loss curve alone.

Curriculum learning [12] offers a simple remedy. Instead of training directly on the hardest problem, the model is trained on a sequence of problems of increasing difficulty. For a PINN this means training first at a low Reynolds number and then raising Re step by step, starting each stage from the converged weights of the previous one rather than from a random initialization. The idea is the same as numerical continuation [13]: the solution at one Reynolds number serves as the initial guess (the *predictor*) at the next, and the optimizer corrects it (the *corrector*). Because the flow changes smoothly between neighboring Reynolds numbers, each stage begins close to the physical solution, whereas a random initialization at high Re can just as easily fall into a non-physical one. Figure 1 illustrates this, with $\lambda = \ln \text{Re}$ as the continuation parameter.

In our earlier conference paper [14] we introduced this warm-start strategy for viscous flow past the NACA 0012 airfoil and showed, from the pressure distributions, that it remains physical up to Re = 20,000 where independently trained networks do not. That study did not examine *why* the strategy works, or whether the gain could simply come from a longer total training time. The present paper addresses these questions at $\alpha = 0^\circ$ for $\text{Re} \in \{100, 500, 1000, 2000, 5000, 10{,}000\}$, keeping the network, the loss and the per-stage optimizer settings the same for both strategies. Its contributions are:

(i) a quantitative comparison of the training loss and pressure coefficient of the two strategies, identifying the Reynolds number at which independent training stops giving physical solutions, with validation against ANSYS Fluent;

(ii) an analysis of layer-wise activation, sparsity and PDE residuals that traces the curriculum's advantage to measurable differences inside the network; and

(iii) a neuron-level correlation analysis that links the curriculum network's internal encoding to the known Reynolds-number dependence of the flow.

Figure 2 summarizes the workflow.

Section II states the governing equations and boundary conditions, and Sec. III describes the loss, the network, the two training procedures and the diagnostics. The results are presented in Sec. IV, discussed in Sec. V and summarized in Sec. VI.

II. Governing Equations

We consider steady, laminar, incompressible flow in two dimensions and use the primitive-variable, or velocity–pressure (VP), form of the Navier–Stokes equations. Lengths are scaled by the chord, velocities by the freestream speed and pressure by ρU_∞^2, so that $U_\infty = c = \rho = 1$ and $\nu = 1/\text{Re}$. The residuals of continuity and momentum are

$$\mathcal{R}_c \equiv \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0, \tag{1}$$

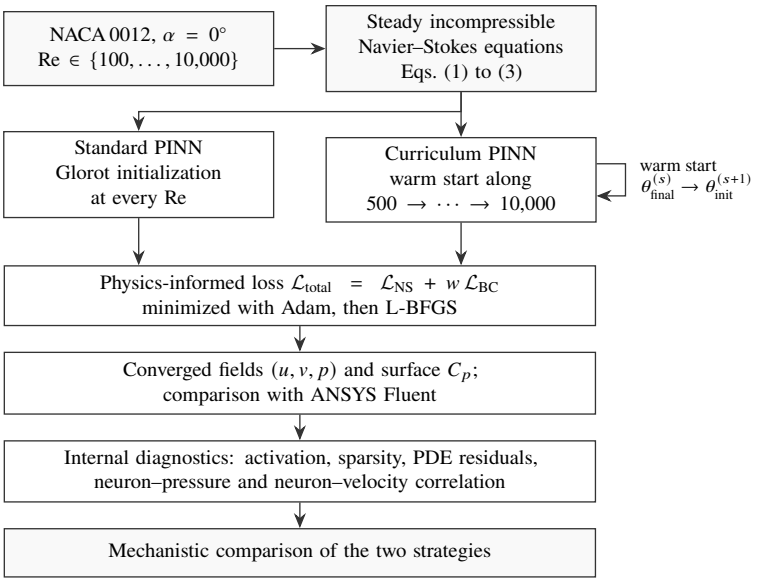


Fig. 2 Workflow of the study. The same boundary-value problem is solved by a standard PINN and by a curriculum PINN over the same Reynolds numbers. The resulting fields are checked against ANSYS Fluent and then compared through the internal diagnostics of Sec. III.D.

$$\mathcal{R}_x \equiv u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + \frac{\partial p}{\partial x} - \frac{1}{\mathrm{Re}}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) = 0, \tag{2}$$

$$\mathcal{R}_y \equiv u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + \frac{\partial p}{\partial y} - \frac{1}{\mathrm{Re}}\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) = 0, \tag{3}$$

where $u(\boldsymbol{x})$ and $v(\boldsymbol{x})$ are the streamwise and transverse velocity, $p(\boldsymbol{x})$ is the static pressure, and $\boldsymbol{x} = (x, y)$. Taking the divergence of the momentum equations and using continuity gives the pressure Poisson relation

$$\nabla^2 p = 2\left(\frac{\partial u}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial v}{\partial x}\right), \tag{4}$$

which shows that the pressure is determined by the velocity gradients. It follows from Eqs. (1) to (3) and is not imposed separately.

The computational domain is the rectangle $x \in [-3, 5]$, $y \in [-1, 1]$ with the airfoil leading edge at the origin and the chord along the x axis. A uniform stream $u = 1$, $v = 0$ is prescribed at the inlet, the gauge pressure $p = 0$ at the outlet, and no-slip $u = v = 0$ on the airfoil surface. The lateral boundaries at $|y| = c$ are close by the standards of external-flow CFD, so some confinement of the pressure field is to be expected; both strategies use the same domain, so the comparison between them is unaffected. The setup is shown in Fig. 3. The section has the standard NACA four-digit thickness distribution with a maximum thickness of 12% of the chord. Because the incidence is zero and the section is symmetric, the exact solution is symmetric about $y = 0$; we did not impose this symmetry on either network.

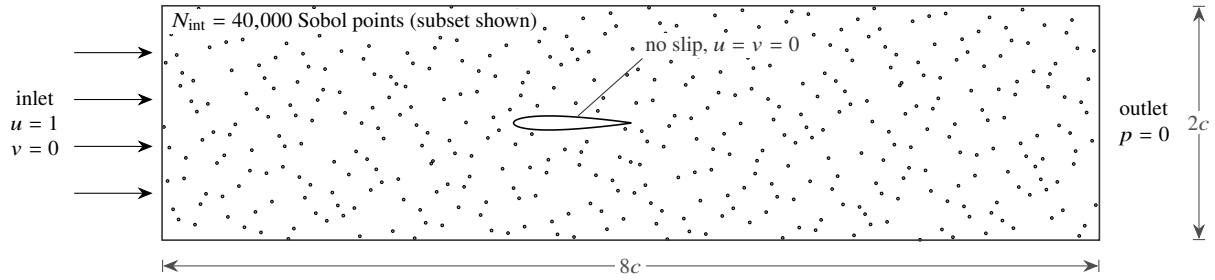
The network $\Theta(\boldsymbol{x}; \theta)$ approximates the triplet $(u, v, p) \approx \Theta(\boldsymbol{x}; \theta)$. Its parameters are found by minimizing the loss defined in Sec. III.A.

III. Methodology

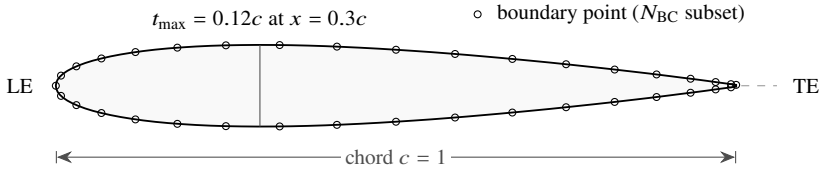
A. Physics-Informed Loss

The total loss is the sum of a residual term and a boundary term,

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{NS}} + w\,\mathcal{L}_{\text{BC}}, \tag{5}$$



(a) Computational domain and boundary conditions.



(b) NACA 0012 section at true aspect ratio, with no-slip boundary points.

Fig. 3 Problem setup. (a) The rectangular domain $x \in [-3, 5]$, $y \in [-1, 1]$ with a uniform inflow at the left, zero gauge pressure at the right and no slip on the airfoil; dots show a subset of the interior collocation points. (b) The section, drawn from the NACA four-digit thickness distribution, with a subset of the boundary points used in $\mathcal{L}_{\text{BC}}$. Point positions are illustrative.

with $w = 1$ in every run reported here. Over N_{int} interior collocation points,

$$\mathcal{L}_{\text{NS}} = \frac{1}{N_{\text{int}}} \sum_{i=1}^{N_{\text{int}}} \Big[\mathcal{R}_c(\boldsymbol{x}_i)^2 + \mathcal{R}_x(\boldsymbol{x}_i)^2 + \mathcal{R}_y(\boldsymbol{x}_i)^2 \Big], \tag{6}$$

and over N_{BC} boundary points,

$$\mathcal{L}_{\text{BC}} = \frac{1}{N_{\text{BC}}} \sum_{j=1}^{N_{\text{BC}}} \Big[\big(\hat{u}_j - u_j^{\text{BC}}\big)^2 + \big(\hat{v}_j - v_j^{\text{BC}}\big)^2 + \big(\hat{p}_j - p_j^{\text{BC}}\big)^2 \Big], \tag{7}$$

where a hat marks a network prediction and the superscript BC a prescribed value. Only the components actually specified at a given boundary point contribute.

B. Network Architecture

Both strategies use the same fully connected network,

$$\Theta(\boldsymbol{x}; \theta) = \mathcal{O} \circ \mathcal{H}_8 \circ \cdots \circ \mathcal{H}_1(\boldsymbol{x}), \qquad \Theta : \mathbb{R}^2 \to \mathbb{R}^3, \tag{8}$$

in which each hidden map $\mathcal{H}_\ell : \mathbb{R}^{20} \to \mathbb{R}^{20}$ (with $\mathbb{R}^2 \to \mathbb{R}^{20}$ for $\ell = 1$) is $\mathcal{H}_\ell(\cdot) = \sigma\big(\boldsymbol{W}_\ell \cdot + \boldsymbol{b}_\ell\big)$ and the output map $\mathcal{O}$ is affine. The activation is

$$\sigma(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}, \tag{9}$$

which is bounded in $(-1, 1)$ and infinitely differentiable. Smoothness matters here because the momentum residuals contain second derivatives of the outputs, which are obtained by automatic differentiation through the network. Weights are drawn from the Glorot-normal distribution [15]. The parameter count is

$$\underbrace{2 \times 20}_{\boldsymbol{W}_1} + \underbrace{7 \times 20 \times 20}_{\boldsymbol{W}_{2\text{–}8}} + \underbrace{20 \times 3}_{\boldsymbol{W}_{\mathcal{O}}} + \underbrace{8 \times 20 + 3}_{\text{biases}} = 3{,}063. \tag{10}$$

Figure 4 shows the network together with the differentiation step that produces the residuals.

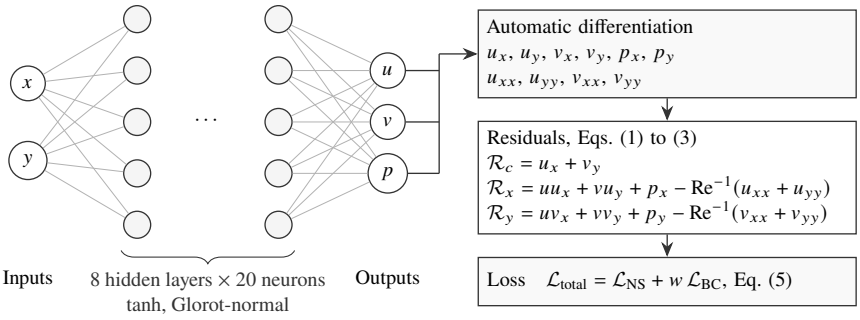


Fig. 4 Network and residual evaluation. The inputs are the coordinates (x, y) and the outputs are u, v and p. All derivatives needed in Eqs. (1) to (3) are obtained by automatic differentiation of the network outputs.

Algorithm 1 Curriculum (warm-start) training of a PINN along a Reynolds-number schedule

Require: schedule $\mathrm{Re}^{(1)} < \mathrm{Re}^{(2)} < \cdots < \mathrm{Re}^{(K)}$
Require: network $\Theta(\boldsymbol{x}; \theta)$
draw $\theta^{(0)} \sim \mathcal{N}_{\text{Glorot}}$
for $s = 1$ **to** K **do**
  $\mathrm{Re} \leftarrow \mathrm{Re}^{(s)}$
  if $s > 1$ **then**
    $\theta^{(s)}_{\text{init}} \leftarrow \theta^{(s-1)}_{\text{final}}$ {warm start (predictor)}
  else
    $\theta^{(s)}_{\text{init}} \leftarrow \theta^{(0)}$
  end if
  $\theta^{(s)}_{\text{final}} \leftarrow \arg\min_\theta \ \mathcal{L}_{\text{total}}(\theta; \mathrm{Re}^{(s)})$ from $\theta^{(s)}_{\text{init}}$ {Adam, then L-BFGS (corrector)}
end for
return $\{\theta^{(s)}_{\text{final}}\}_{s=1}^{K}$

C. Training Strategies

The *standard* PINN is trained separately at each of the six Reynolds numbers, each time from a new Glorot-normal draw. The *curriculum* PINN is initialized once, trained at Re = 500, and then passed through the remaining four stages without reinitialization, as set out in Algorithm 1. In the language of Sec. I, the line marked *warm start* is the predictor and the call to the optimizer is the corrector. Every stage of either strategy uses Adam followed by L-BFGS. Table 1 lists the remaining settings. Architecture, loss, domain, boundary conditions and collocation points are identical for the two strategies.

The per-stage optimizer budgets are comparable, but they are not identical, and by construction the curriculum network has seen more total iterations by the time it reaches a given Reynolds number. We return to this point in Sec. V.C.

D. Diagnostics

A low loss does not by itself certify a physical solution, so each trained network is examined with the following measures. All are evaluated on a set of N_{e} points $\{\boldsymbol{x}_i\}_{i=1}^{N_{\text{e}}}$ in the domain. We write $h_{\ell j}(\boldsymbol{x})$ for the activation of neuron j in hidden layer ℓ, and $W_\ell = 20$ for the layer width.

RMS residuals. For each equation,

$$\mathrm{RMS}(\mathcal{R}_k) = \left(\frac{1}{N_{\text{e}}} \sum_{i=1}^{N_{\text{e}}} \mathcal{R}_k(\boldsymbol{x}_i)^2 \right)^{1/2}, \qquad k \in \{c, x, y\}. \tag{11}$$

Table 1 Computational settings for the standard and curriculum PINNs at $\alpha = 0°$.

Parameter	Standard PINN	Curriculum PINN
Architecture	8×20, tanh	8×20, tanh
Formulation	VP: $(x, y) \to (u, v, p)$	VP: $(x, y) \to (u, v, p)$
Parameters $\lvert\theta\rvert$	3 063	3 063
Initialization	Glorot normal (each Re)	Glorot (Stage 1 only)
Warm start	None	Weights from previous stage
Re schedule	{100, 500, 1k, 2k, 5k, 10k}	$500 \to 1\text{k} \to 2\text{k} \to 5\text{k} \to 10\text{k}$
Adam epochs	40 000 ($\eta = 10^{-3}$)	40 000 (η decreasing)
L-BFGS epochs	~15 000	15 000–30 000
BC weight w	1	1
Domain	$x \in [-3, 5]$, $y \in [-1, 1]$	$x \in [-3, 5]$, $y \in [-1, 1]$
Inlet	$u = 1$, $v = 0$	$u = 1$, $v = 0$
Outlet	$p = 0$	$p = 0$
Airfoil	$u = v = 0$ (no slip)	$u = v = 0$ (no slip)
Interior points	40 000 (Sobol)	40 000 (Sobol)
Boundary points	1 000	1 000
Hardware	A100-SXM4-40 GB	A100-SXM4-40 GB
Wall time per stage	~1.3 h	~1.3 h

Sparsity. The sparsity of layer ℓ is the percentage of its activations whose magnitude is below $\varepsilon = 0.01$,

$$S_\ell = \frac{100}{N_\mathrm{e}\, W_\ell} \sum_{i=1}^{N_\mathrm{e}} \sum_{j=1}^{W_\ell} \mathbf{1}\big[\, |h_{\ell j}(\boldsymbol{x}_i)| < \varepsilon \,\big], \tag{12}$$

where $\mathbf{1}[\cdot]$ is the indicator function. The network sparsity $\overline{S}$ is the average of S_ℓ over the eight hidden layers, in percent. A high $\overline{S}$ means that much of the network is effectively idle.

Mean activation and neuron status. The mean absolute activation of neuron j in layer ℓ, and its average over the layer, are

$$\overline{a}_{\ell j} = \frac{1}{N_\mathrm{e}} \sum_{i=1}^{N_\mathrm{e}} |h_{\ell j}(\boldsymbol{x}_i)|, \qquad \overline{a}_\ell = \frac{1}{W_\ell} \sum_{j=1}^{W_\ell} \overline{a}_{\ell j}. \tag{13}$$

For the curriculum network we write $\overline{a}_{\ell,s}$ for the value of $\overline{a}_\ell$ at the end of stage s. Each neuron is classified by $\overline{a}_{\ell j}$ as

$$\text{dead if } \overline{a}_{\ell j} < 0.001, \qquad \text{saturated if } \overline{a}_{\ell j} > 0.99, \qquad \text{active otherwise.} \tag{14}$$

Neuron–physics correlation. Let $\boldsymbol{h}_{\ell j}$ and $\boldsymbol{q}$ be the vectors of values of $h_{\ell j}$ and of a predicted quantity q (such as $\hat{p}$ or $\hat{u}$) at the N_e points, and let $\tilde{\boldsymbol{h}}_{\ell j}$ and $\tilde{\boldsymbol{q}}$ be the same vectors with their means subtracted. The Pearson correlation and its average magnitude over the 160 hidden neurons are

$$r_{\ell j,q} = \frac{\tilde{\boldsymbol{h}}_{\ell j} \cdot \tilde{\boldsymbol{q}}}{\|\tilde{\boldsymbol{h}}_{\ell j}\|\, \|\tilde{\boldsymbol{q}}\|}, \qquad \overline{|r|}_q = \frac{1}{160} \sum_{\ell=1}^{8} \sum_{j=1}^{20} |r_{\ell j,q}|. \tag{15}$$

Pressure coefficient and physicality criterion. On the airfoil surface,

$$C_p = \frac{p - p_\infty}{\frac{1}{2}\rho\, U_\infty^2} = 2p, \tag{16}$$

since $U_\infty = \rho = 1$ and the reference pressure is the outlet value $p_\infty = 0$. In the absence of a reference solution during training, we call a predicted distribution *physical* if and only if

$$C_p^{\max} > 0.5, \tag{17}$$

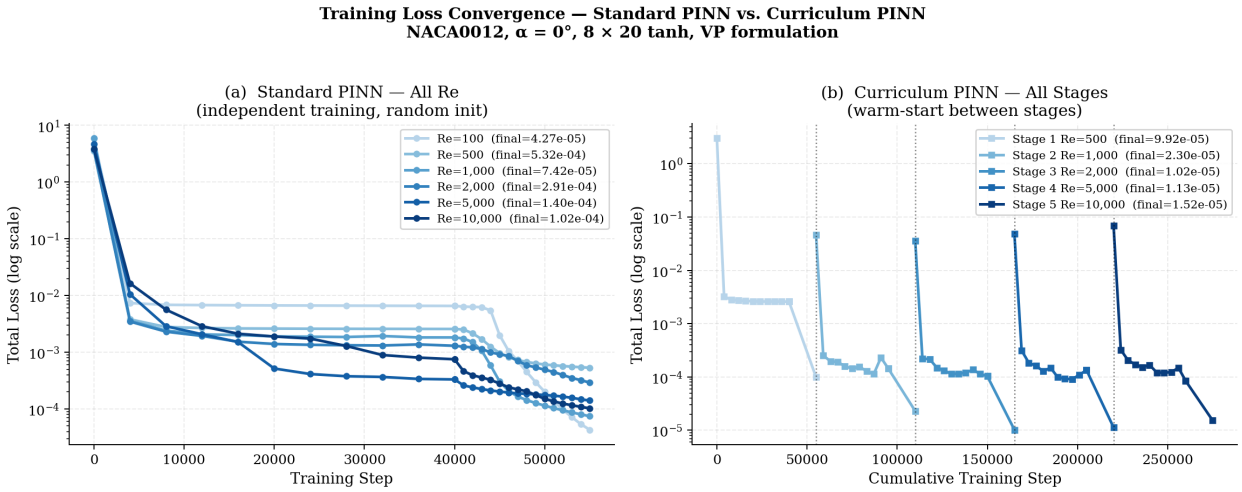


Fig. 5 Total loss $\mathcal{L}_{\text{total}}$ (log scale) against training step. (a) Standard PINN at all six Reynolds numbers, each started from a Glorot initialization. (b) Curriculum PINN over its five stages; the jumps mark the transfer of weights to the next Reynolds number.

Table 2 Final loss, improvement ratio and physicality of C_p (criterion of Eq. (17)) at $\alpha = 0°$.

Re	$\mathcal{L}_{\text{total}}^{\text{std}}$	$\mathcal{L}_{\text{total}}^{\text{curr}}$	Ratio	Physical C_p, std	Physical C_p, curr
100	4.27×10^{-5}	—	—	Yes	—
500	5.32×10^{-4}	9.92×10^{-5}	5.4×	No	Yes
1000	7.42×10^{-5}	2.30×10^{-5}	3.2×	Yes	Yes
2000	2.91×10^{-4}	1.02×10^{-5}	28.5×	No	Yes
5000	1.40×10^{-4}	1.13×10^{-5}	12.3×	No	Yes
10 000	1.02×10^{-4}	1.52×10^{-5}	6.7×	No	Yes

that is, if a stagnation peak is present. A distribution that is negative everywhere, close to zero everywhere, or marked by abrupt non-monotone jumps is classed as *non-physical* and taken to signal convergence to a spurious stationary point. The threshold is deliberately loose; the ANSYS Fluent comparison in Sec. IV.B.1 is the stricter test.

IV. Results

A. Training Loss

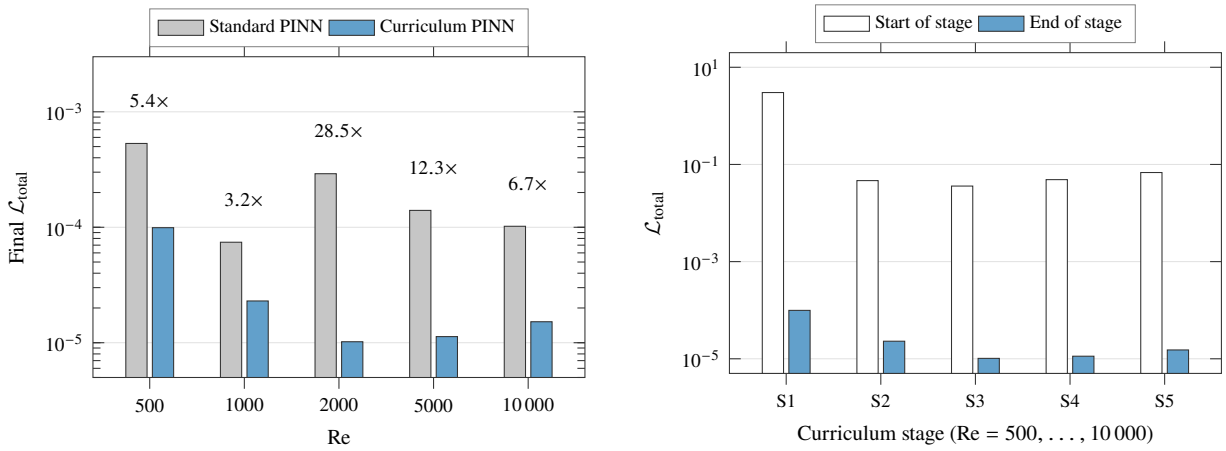
Figure 5 shows how the loss evolves. All runs converge in the usual numerical sense, falling by four to five orders of magnitude before they flatten out, so the loss curves alone do not tell the two strategies apart. The difference lies in where each run ends up.

Table 2 gives the final values. Between Re = 500 and Re = 10,000 the curriculum network always finishes lower, by a factor of 3.2 to 28.5. The largest gap, at Re = 2000, comes just after the last Reynolds number at which the standard network is still physical (Fig. 6a). That is where we would expect the starting point to matter most, since the standard run has lost the physical basin while the curriculum run begins the stage inside it. The smallest gap, at Re = 1000, is at the only Reynolds number above 100 where both networks are physical.

The effect of the warm start is clearest at the start of each stage. Let $\mathcal{L}_0^{(s)}$ be the loss just after the weights are transferred, before any update at the new Reynolds number. Increasing Re does push the loss up, by a factor of 469 to 6,014 relative to the end of the previous stage (Table 3), but every warm-started stage still begins well below the value $\mathcal{L}_0^{(1)} = 3.02$ that follows the random initialization of Stage 1. The ratio $\mathcal{L}_0^{(1)}/\mathcal{L}_0^{(s)}$ lies between 44 and 84 for Stages 2–5, and is 64 on average. In continuation terms, the predictor starts the optimizer about two orders of magnitude closer to a solution than a random draw does (Fig. 6b).

Table 3 Loss at the start of each curriculum stage, after weight transfer, compared with the final loss of the preceding stage.

Stage	Re	$\mathcal{L}_0^{(s)}$	Initial / previous final
1	500	3.02	— (random)
2	1000	4.65×10^{-2}	469×
3	2000	3.59×10^{-2}	1,565×
4	5000	4.87×10^{-2}	4,765×
5	10 000	6.82×10^{-2}	6,014×



(a) Final loss; labels give the ratio std/curr.

(b) Warm-start effect: initial and final loss per stage.

Fig. 6 Training loss at matched Reynolds numbers (a) and at the start and end of each curriculum stage (b). Data from Tables 2 and 3.

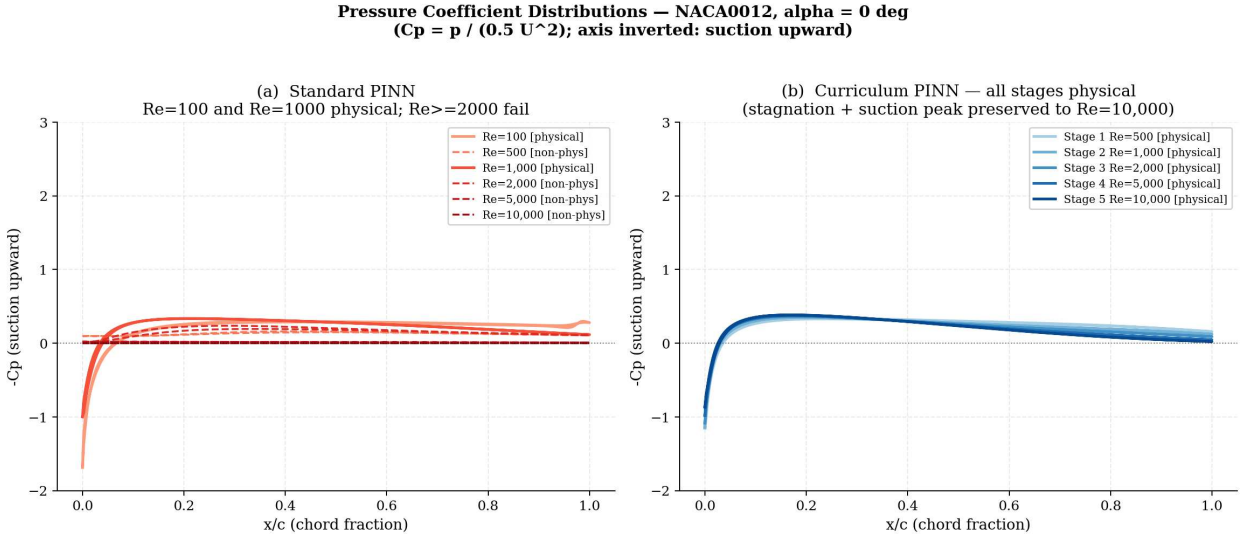


Fig. 7 Surface pressure coefficient on the NACA 0012 at $\alpha = 0°$, plotted as $-C_p$ following aerodynamic convention. (a) Standard PINN: only Re = 100 and Re = 1000 show a stagnation peak. (b) Curriculum PINN: all five stages are physical, with $C_p^{\max} \in [+0.87, +1.15]$.

B. Pressure Coefficient and Flow Fields

The surface pressure is where the two strategies part company physically, and not only numerically. Figure 7 compares the C_p distributions over the full Reynolds-number range, and Fig. 8 marks which runs pass the physicality test.

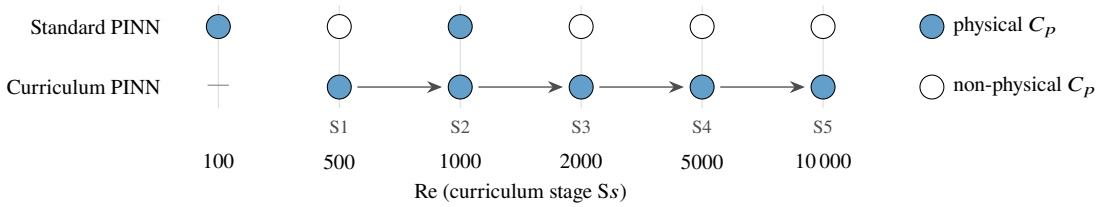


Fig. 8 Physicality of the predicted pressure coefficient (criterion of Eq. (17)) for each run. The curriculum network starts at Re = 500 and stays physical through Stage 5, each stage warm-started from the previous one (arrows); the standard network is physical only at Re = 100 and Re = 1000.

The standard PINN produces a physical distribution only at Re = 100 ($C_p^{\max} = +1.687$) and Re = 1000 ($C_p^{\max} = +0.999$). Its result at Re = 500 is the most telling failure, with C_p between -0.160 and -0.096 over the entire chord. A pressure coefficient that is negative everywhere is impossible when the flow stagnates at the leading edge, so here the network has satisfied the loss without describing the flow at all. From Re = 2000 upward, C_p drops to $\mathcal{O}(10^{-3})$ and the pressure field no longer carries any aerodynamic information. This is the weak coupling of pressure described in Sec. I.

Every curriculum stage, in contrast, shows a stagnation peak, a suction region and recovery towards the trailing edge. The peak falls as the Reynolds number increases, from $C_p^{\max} = 1.154$ at Re = 500 and 0.999 at Re = 1000 to 0.868 at Re = 10,000. That trend is expected. At low Re the thick viscous layer near the nose contributes to the normal stress and lifts the stagnation pressure above its inviscid value of one; as Re grows, this contribution fades and the stagnation region becomes narrower.

1. Comparison with ANSYS Fluent

The criterion of Eq. (17) checks only that a stagnation peak exists, so we also compared the curriculum solutions with ANSYS Fluent at Re = 500, 1000, 5000 and 10,000 (Fig. 9). At all four, the PINN captures the stagnation peak, the suction plateau and the recovery over the rear of the chord, and the two curves stay close over most of the section. The only visible differences are in the mid-chord suction region at Re = 10,000. We attribute them to the limited capacity of an 8×20 network once the boundary layer is only a few hundredths of a chord thick.

2. Velocity and Pressure Fields

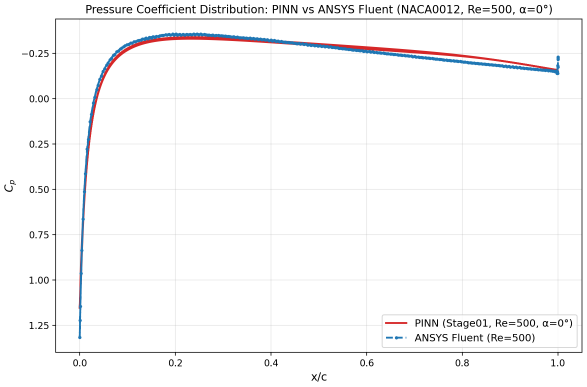
Figures 10 to 12 show the complete fields. At Re = 1000 the standard PINN resolves an attached boundary layer and a wake. At Re = 10,000 it still produces a thin wake that looks plausible in u, yet its C_p lies between -0.0004 and $+0.0013$, and nothing in the velocity contours would warn a reader of this. The curriculum network at the same Reynolds number gives a thin attached boundary layer and a pressure field with $C_p^{\max} = +0.868$; this is the highest Reynolds number in the study with a physical solution.

C. Internal Representations

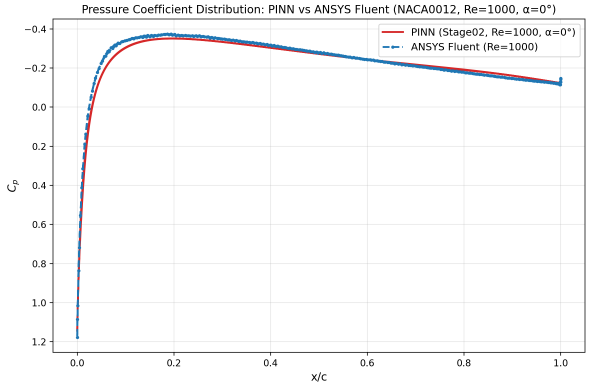
The loss and the surface pressure establish that the curriculum network does better. The diagnostics of Sec. III.D let us look at how it differs internally.

1. Activation of the Deepest Layer

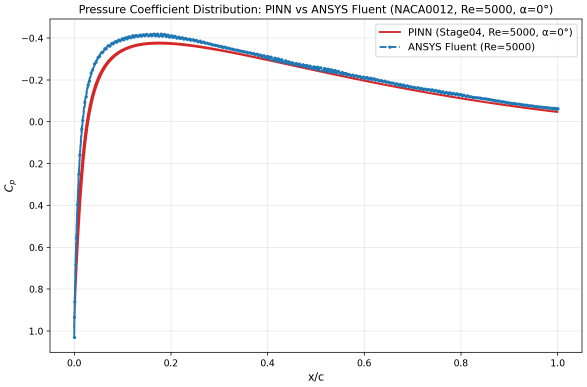
Figure 13(a) shows the mean absolute activation of each layer of the curriculum network at each stage. Every layer becomes more active as training moves along the schedule, the middle layers most of all. For comparing the two strategies the deepest layer is the most useful (Fig. 13b). In the standard network the Layer 8 activation has no clear trend: it is largest at Re = 500 ($\overline{a}_8 = 0.572$), drops to 0.398 at Re = 5000 and climbs again at Re = 10,000, much as one would expect if the independent runs settle into different kinds of minimum. In the curriculum network it rises at every stage, from $\overline{a}_{8,1} = 0.554$ to $\overline{a}_{8,5} = 0.634$, and the ratio $\overline{a}_{8,s}/\overline{a}_{1,s}$ goes from 0.871 to 0.933. The deepest layer is gradually brought up to the activity of the first, which suggests that gradients continue to reach the full depth of the network as the problem gets harder.



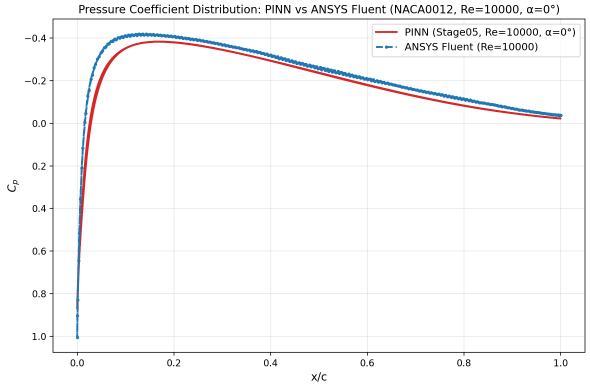
(a) Re = 500 (Stage 1).



(b) Re = 1000 (Stage 2).



(c) Re = 5000 (Stage 4).



(d) Re = 10,000 (Stage 5).

Fig. 9 Curriculum PINN (solid red) against ANSYS Fluent (dashed blue) at four Reynolds numbers, plotted with $-C_p$ upward. The two agree closely at every stage.

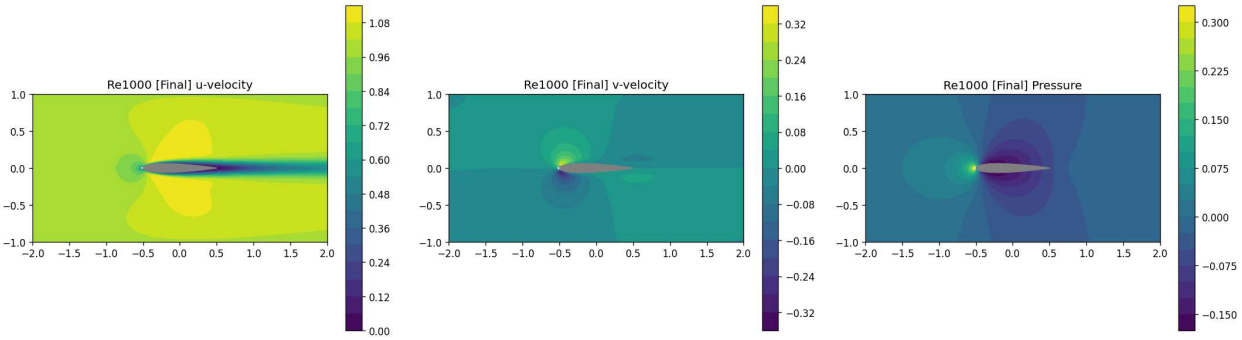


Fig. 10 Standard PINN at Re = 1000, $\alpha = 0°$: u, v and p from left to right. The boundary layer and wake are resolved; this is one of only two Reynolds numbers at which the standard network is physical.

2. Sparsity

Network sparsity is listed in Table 4 and plotted in Fig. 14(a). The clearest contrast is at Re = 1000, where the standard network is 83.28% sparse, so that fewer than 17% of its activations exceed ε. The curriculum network at the same Reynolds number is 54.47% sparse, a difference of 28.81 percentage points. From Re = 1000 upward the curriculum network is always the less sparse of the two. The one exception is Re = 500, where the curriculum run is still in its first, randomly initialized stage and the two strategies have not yet diverged. There the standard network is 5.44 points less sparse, but it is also non-physical, so sparsity alone is clearly not a measure of solution quality.

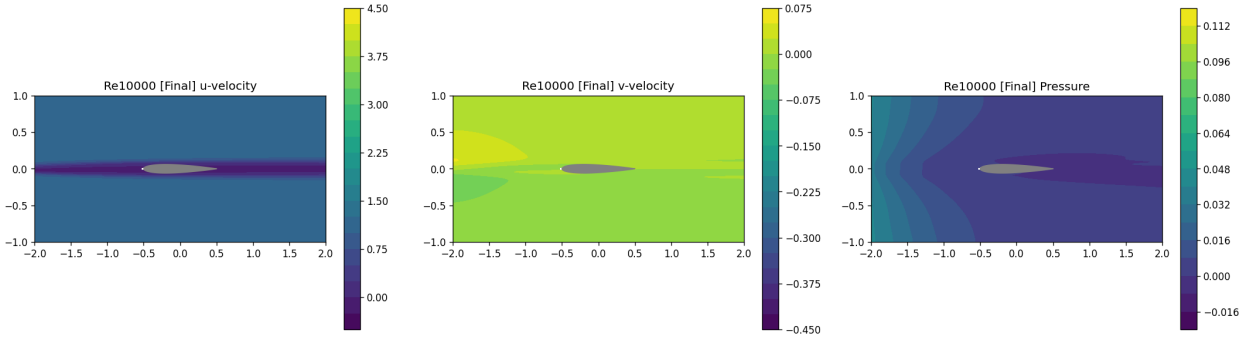


Fig. 11 Standard PINN at Re = 10,000, $\alpha = 0°$. A thin wake is visible, yet $C_p \in [-0.0004, +0.0013]$. A plausible velocity field does not imply a correct pressure field.

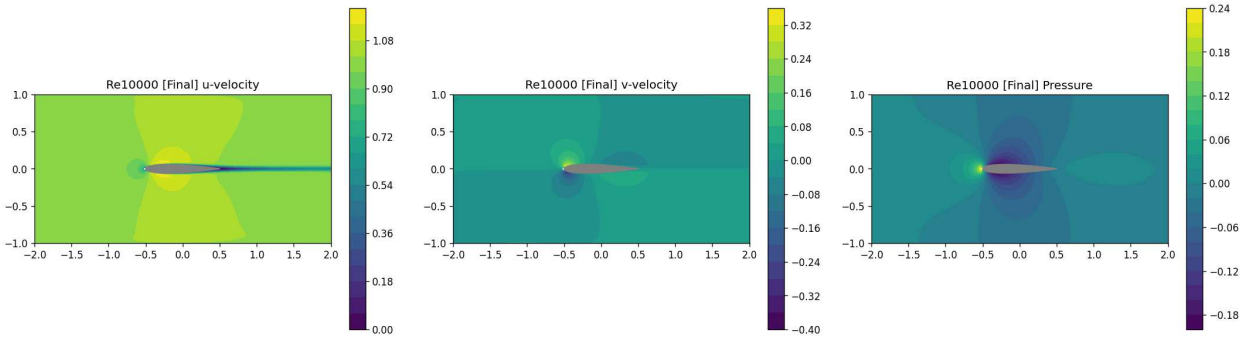
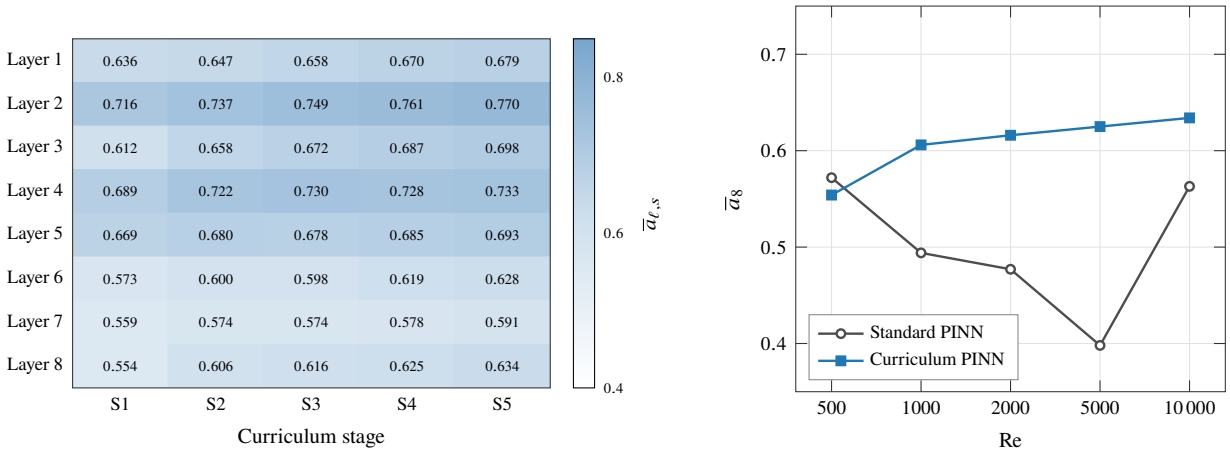


Fig. 12 Curriculum PINN, Stage 5, at Re = 10,000, $\alpha = 0°$. Both the velocity field, with its thin attached boundary layer, and the pressure field ($C_p^{\max} = +0.868$) are physical.



(a) Curriculum PINN: $\overline{a}_{\ell,s}$ by layer and stage.

(b) Layer 8 activation against Re.

Fig. 13 Layer-wise activation of the curriculum PINN (a) and comparison of the deepest layer between the two strategies (b).

3. PDE Residuals

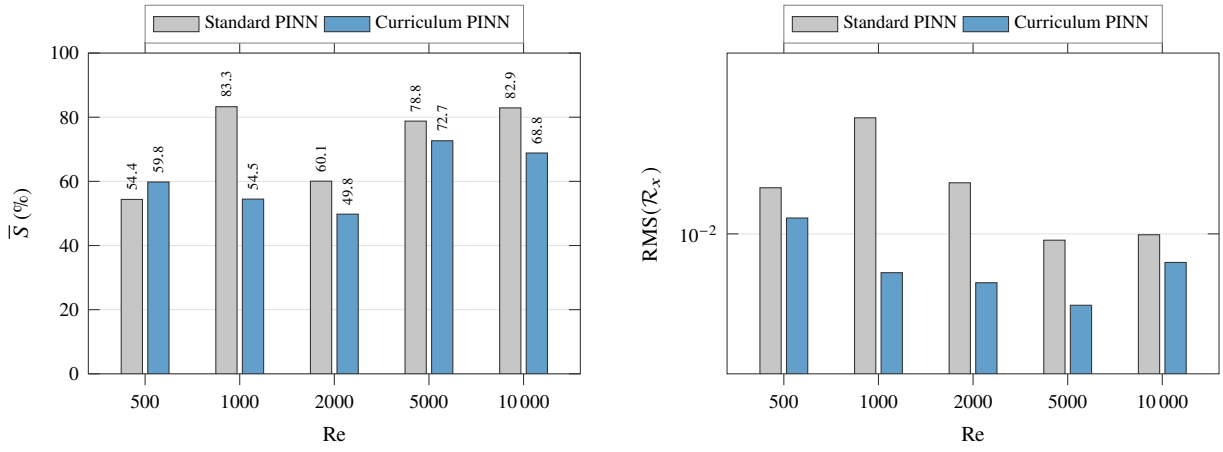
The RMS residuals in Table 5 point to a trade-off. The curriculum network has the smaller x-momentum residual at every matched Reynolds number (Fig. 14b) and the smaller y-momentum residual everywhere except Re = 10,000. Its continuity residual, on the other hand, is larger at four of the five Reynolds numbers, with Re = 1000 the exception. Warm-started training seems to favour momentum balance over strict pointwise mass conservation, and the end result is still a physically correct pressure distribution. Section V.B returns to this point.

Table 4 Network sparsity $\overline{S}$ (%) for the two strategies. Lower values indicate greater use of the network's capacity.

Re	$\overline{S}^{\text{std}}$ (%)	$\overline{S}^{\text{curr}}$ (%)	Δ (pp)	Lower $\overline{S}$
500	54.37	59.81	−5.44	Standard
1000	83.28	54.47	+28.81	Curriculum
2000	60.06	49.78	+10.28	Curriculum
5000	78.78	72.66	+6.12	Curriculum
10 000	82.91	68.84	+14.07	Curriculum

Table 5 RMS PDE residuals of the standard (Std) and curriculum (Cur) PINNs at $\alpha = 0°$.

Re	RMS($\mathcal{R}_c$)		RMS($\mathcal{R}_x$)		RMS($\mathcal{R}_y$)	
	Std	Cur	Std	Cur	Std	Cur
500	9.1×10^{-3}	2.9×10^{-2}	1.7×10^{-2}	1.2×10^{-2}	1.2×10^{-2}	9.3×10^{-3}
1000	4.2×10^{-2}	1.3×10^{-2}	3.8×10^{-2}	6.4×10^{-3}	3.6×10^{-2}	6.3×10^{-3}
2000	6.5×10^{-3}	1.7×10^{-2}	1.8×10^{-2}	5.7×10^{-3}	1.1×10^{-2}	7.7×10^{-3}
5000	4.5×10^{-3}	1.8×10^{-2}	9.3×10^{-3}	4.4×10^{-3}	6.1×10^{-3}	4.2×10^{-3}
10 000	9.0×10^{-3}	1.6×10^{-2}	9.9×10^{-3}	7.2×10^{-3}	5.3×10^{-3}	6.4×10^{-3}



(a) Network sparsity, Eq. (12).

(b) RMS x-momentum residual (log scale).

Fig. 14 Network sparsity (a) and x-momentum residual (b) of the two strategies. Data from Tables 4 and 5.

4. Decoupling of Pressure Encoding

The sharpest difference between the networks appears at the level of individual neurons (Eq. (15), Fig. 15). In the curriculum network, the mean correlation magnitude between activations and predicted pressure decreases steadily along the schedule, from $\overline{|r|}_p = 0.407$ at Stage 1 (Re = 500) to 0.070 at Stage 5 (Re = 10,000), a drop of 83%. The velocity correlation barely moves over the same stages and stays at $\overline{|r|}_u \approx 0.10$.

There is a simple physical explanation for this. At Re = 500 the viscous region around the airfoil is thick, and the pressure field, governed by the elliptic relation of Eq. (4), is the dominant structure in the solution; many neurons follow it. As Re increases, the viscous effects are confined to a layer of thickness $\mathcal{O}(\text{Re}^{-1/2})$, the outer pressure field approaches the smooth inviscid distribution, and the near-wall velocity gradients become the features that are hardest to represent and so occupy the most neurons. The curriculum network tracks this change, its neurons moving gradually away from the pressure towards the geometry and the boundary layer.

Nothing similar happens in the standard network. Its pressure correlation goes up, down and up again, and stays near 0.32 for Re $\geq$ 5000. Because its pressure fields at those Reynolds numbers are close to zero, the network is in effect encoding the wrong pressure. None of this can be seen in loss curves or C_p plots; it only appears when the neurons themselves are examined. It also shows that a PINN trained along a continuation path can develop an internal representation that responds to the Reynolds number in a physically sensible way.

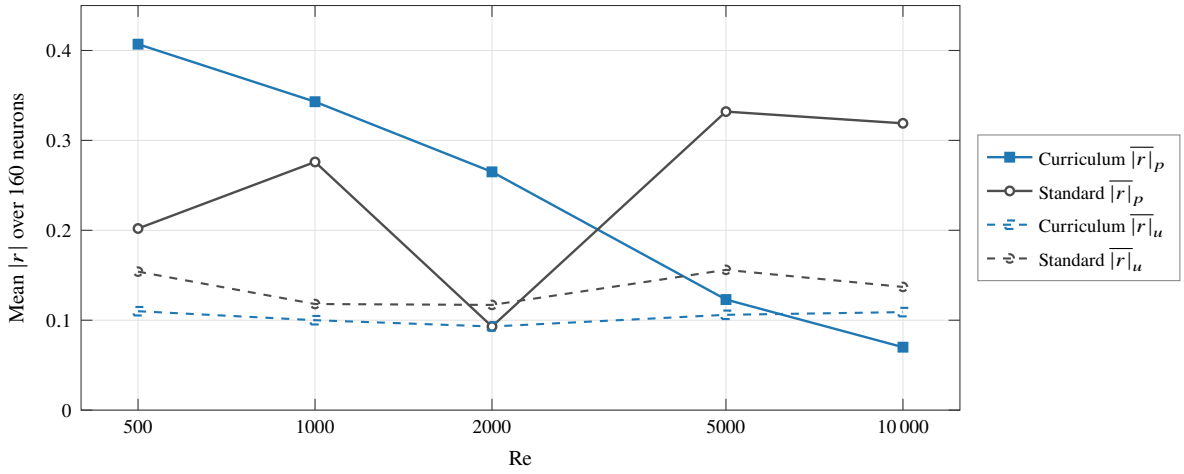
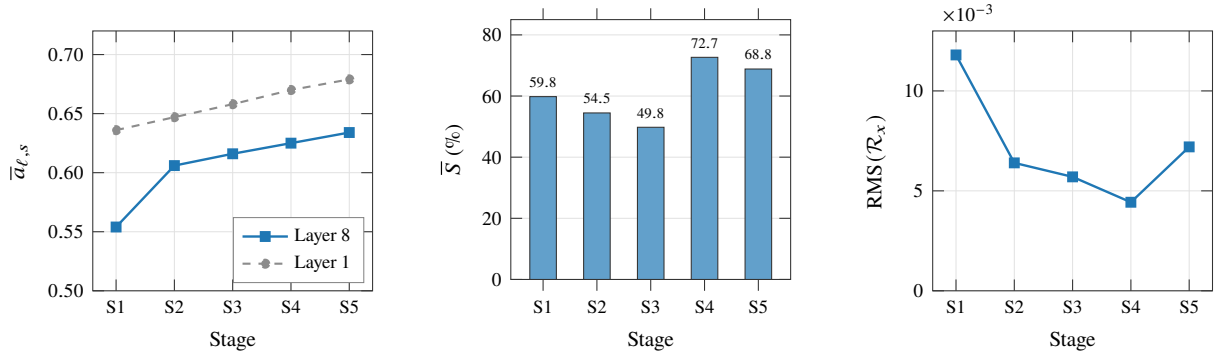


Fig. 15 Mean absolute Pearson correlation of neuron activations with predicted pressure, $\overline{|r|}_p$ (solid), and streamwise velocity, $\overline{|r|}_u$ (dashed). The curriculum network's pressure correlation falls monotonically from 0.407 to 0.070; the standard network shows no systematic trend.



(a) Deep and shallow activation.

(b) Sparsity.

(c) x-momentum residual.

Fig. 16 Stage-to-stage evolution of the curriculum PINN: activation of Layers 8 and 1 (a), network sparsity (b) and RMS x-momentum residual (c).

5. Neuron Status and Stage-to-Stage Evolution

At Stage 5 (Re = 10,000), none of the 160 hidden neurons of the curriculum network is dead ($\overline{a}_{\ell j} < 0.001$) or saturated ($\overline{a}_{\ell j} > 0.99$); all of them are active by the definition in Eq. (14). This is compatible with the network-averaged sparsity of 68.84% at the same stage. Many individual activations are small, but no neuron is inactive over the whole domain.

Figure 16 brings together three trends over the stages. The Layer 8 activation increases each time, from 0.554 to 0.634, while Layer 1 changes much less. Sparsity is not monotone: it goes from 59.8% to 54.5%, 49.8%, 72.7% and 68.8%, and the jump at Stage 4 coincides with the step from Re = 2000 to 5000, the largest in $\ln \mathrm{Re}$. The x-momentum residual decreases from 1.18×10^{-2} at Stage 1 to 4.43×10^{-3} at Stage 4 and then rises a little at Stage 5.

D. Consolidated Diagnostics

Tables 6 and 7 collect the main diagnostics of each network for reference.

Table 6 Standard PINN: summary of diagnostics at $\alpha = 0°$.

Re	$\mathcal{L}_{\text{total}}$	$\overline{S}$ (%)	RMS($\mathcal{R}_c$)	RMS($\mathcal{R}_x$)	C_p
100	4.27×10^{-5}	64.81	8.1×10^{-3}	4.6×10^{-2}	Physical
500	5.32×10^{-4}	54.37	9.1×10^{-3}	1.7×10^{-2}	Non-physical
1000	7.42×10^{-5}	83.28	4.2×10^{-2}	3.8×10^{-2}	Physical
2000	2.91×10^{-4}	60.06	6.5×10^{-3}	1.8×10^{-2}	Non-physical
5000	1.40×10^{-4}	78.78	4.5×10^{-3}	9.3×10^{-3}	Non-physical
10 000	1.02×10^{-4}	82.91	9.0×10^{-3}	9.9×10^{-3}	Non-physical

Table 7 Curriculum PINN: summary of diagnostics at $\alpha = 0°$.

Stage	Re	$\mathcal{L}_{\text{total}}$	$\overline{S}$ (%)	RMS($\mathcal{R}_c$)	RMS($\mathcal{R}_x$)	C_p
1	500	9.92×10^{-5}	59.81	2.9×10^{-2}	1.2×10^{-2}	Physical
2	1000	2.30×10^{-5}	54.47	1.3×10^{-2}	6.4×10^{-3}	Physical
3	2000	1.02×10^{-5}	49.78	1.7×10^{-2}	5.7×10^{-3}	Physical
4	5000	1.13×10^{-5}	72.66	1.8×10^{-2}	4.4×10^{-3}	Physical
5	10 000	1.52×10^{-5}	68.84	1.6×10^{-2}	7.2×10^{-3}	Physical

V. Discussion

A. Why the Curriculum Works

These results are consistent with the continuation picture of Sec. I. At moderate and high Reynolds number the PINN loss has several basins, and some of them hold fields with no stagnation recovery and almost no pressure. A random initialization has no reason to prefer one basin over another, and the standard network landed in a non-physical one at four of the five Reynolds numbers above 100. The warm start shifts the odds towards the physical basin, because each curriculum stage begins with a field that already has the right structure: a stagnation point, an attached boundary layer and a wake. Between neighbouring Reynolds numbers of the schedule these features change only in degree. The boundary layer gets thinner, the stagnation region narrower and the pressure away from the wall more uniform. The optimizer has to adjust the solution rather than find it from scratch.

Three separate observations support this interpretation.

(1) *Depth of the minimum.* The curriculum network reaches a final loss 3.2 to 28.5 times lower (Table 2), and each stage starts on average about 64 times closer, in loss, than a random draw (Table 3).

(2) *Use of depth.* The Layer 8 activation rises at every stage (Fig. 13b). In the standard network the same quantity drops to 0.398 at Re = 5000, which suggests that the deep layers contribute little to the solution there.

(3) *Physical encoding.* The neuron–pressure correlation of the curriculum network decreases with Reynolds number in the same way that the relative importance of pressure in the flow does (Fig. 15). The standard network shows no such dependence.

B. Momentum versus Continuity

The curriculum network balances momentum better and conserves mass less well (Table 5). A plausible reason is that each warm-started stage begins near a field that already satisfies momentum. Raising Re mainly changes the viscous term in Eqs. (2) and (3) and leaves the continuity residual of the transferred field as it was, so the optimizer works mostly on momentum, where the loss has gone up, and reduces the inherited continuity error more slowly. For the surface pressure this compromise does no harm, as the Fluent comparison shows.

C. Effective Training Budget

One could argue that the curriculum network simply trained for longer: by Re = 10,000 it has been through five stages, while each standard run had only one. We do not think budget explains the difference, for two reasons. The standard loss curves (Fig. 5) spend long stretches on plateaus during the Adam phase and fall only slowly late in the L-BFGS phase, so continuing from the same point would probably change little. More importantly, the standard network fails qualitatively, producing a pressure field with no stagnation peak. Such a field lies in a different basin, and running a

local optimizer for longer does not move a solution from one basin to another. A standard run with the same total budget as the curriculum would settle the question directly; the conclusion here rests on the two arguments above.

VI. Conclusions

We compared standard and curriculum-trained PINNs for steady incompressible flow past a NACA 0012 airfoil at zero incidence, keeping the architecture, the loss and the per-stage optimizer settings the same. Our main findings are the following.

(1) The curriculum PINN gives a physically correct pressure coefficient at every Reynolds number of its schedule, $\mathrm{Re} \in \{500, 1000, 2000, 5000, 10{,}000\}$. For surface pressure prediction this raises the useful range of this network from Re = 1000 to Re = 10,000. The curriculum C_p agrees closely with ANSYS Fluent at Re = 500, 1000, 5000 and 10,000.

(2) At matched Reynolds numbers its final loss is 3.2 to 28.5 times lower. The largest gain occurs at Re = 2000, just past the point where independent training stops yielding physical solutions.

(3) The curriculum network keeps more of its capacity in use, its deepest layer becomes more active at every stage, and its x-momentum residual is lower at every matched Reynolds number. Its continuity residual is higher at most of them, a trade-off that does not affect the surface pressure.

(4) The correlation between its neurons and the predicted pressure falls by 83% across the schedule while the velocity correlation stays near 0.10. This mirrors the shift from a pressure-dominated to a boundary-layer-dominated flow, and shows that the network's internal representation responds to the Reynolds number in a physically interpretable way.

Treating warm-started training as natural-parameter continuation in $\ln \mathrm{Re}$ accounts for all of these observations. The study also shows that internal diagnostics can expose non-physical PINN solutions that the loss curve hides, and we would argue that they should be reported alongside the loss whenever a trained PINN is assessed.

Data Availability

The trained network weights, training logs and ANSYS Fluent surface-pressure data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] H. Schlichting and K. Gersten, *Boundary-Layer Theory*, 9th ed. Berlin: Springer, 2017.

[2] A. Jameson, "Aerodynamic design via control theory," *J. Sci. Comput.*, vol. 3, no. 3, pp. 233–260, 1988.

[3] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.

[4] S. Cai, Z. Mao, Z. Wang, M. Yin, and G. E. Karniadakis, "Physics-informed neural networks (PINNs) for fluid mechanics: A review," *Acta Mechanica Sinica*, vol. 37, no. 12, pp. 1727–1738, 2021.

[5] L. Sun, H. Gao, S. Pan, and J.-X. Wang, "Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data," *Comput. Methods Appl. Mech. Eng.*, vol. 361, p. 112732, 2020.

[6] L. Lu, X. Meng, Z. Mao, and G. E. Karniadakis, "DeepXDE: A deep learning library for solving differential equations," *SIAM Review*, vol. 63, no. 1, pp. 208–228, 2021.

[7] M. Raissi, A. Yazdani, and G. E. Karniadakis, "Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations," *Science*, vol. 367, no. 6481, pp. 1026–1030, 2020.

[8] X. Jin, S. Cai, H. Li, and G. E. Karniadakis, "NSFnets (Navier–Stokes flow nets): Physics-informed neural networks for the incompressible Navier–Stokes equations," *J. Comput. Phys.*, vol. 426, p. 109951, 2021.

[9] H. Eivazi, M. Tahani, P. Schlatter, and R. Vinuesa, "Physics-informed neural networks for solving Reynolds-averaged Navier–Stokes equations," *Phys. Fluids*, vol. 34, no. 7, p. 075117, 2022.

[10] A. S. Krishnapriyan, A. Gholami, S. Zhe, R. Kirby, and M. W. Mahoney, "Characterizing possible failure modes in physics-informed neural networks," in *Advances in Neural Information Processing Systems*, vol. 34, pp. 26548–26560, 2021.

[11] S. Wang, Y. Teng, and P. Perdikaris, "Understanding and mitigating gradient flow pathologies in physics-informed neural networks," *SIAM J. Sci. Comput.*, vol. 43, no. 5, pp. A3055–A3081, 2021.

[12] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proc. 26th Int. Conf. Machine Learning*, pp. 41–48, 2009.

[13] E. L. Allgower and K. Georg, *Introduction to Numerical Continuation Methods*, Classics in Applied Mathematics, vol. 45. Philadelphia: SIAM, 2003.

[14] D. S. Palli and S. S. C. Athkuri, "Physics-informed neural networks with curriculum learning for viscous flow past a NACA 0012 airfoil," in *AIAA SciTech Forum*, 2026.

[15] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Stat.*, pp. 249–256, 2010.