
Stabilizing PINNs: A regularization scheme for PINN training to avoid unstable fixed points of dynamical systems

Miloš Babić

Christian Doppler Laboratory for Physics-driven Machine Learning in Industrial Applications,
Graz, Austria
Institute of Thermodynamics and Sustainable Propulsion Systems, Graz University of Technology,
Graz, Austria
Know Center Research GmbH, Graz, Austria

Franz M. Rohrhofer

Know Center Research GmbH, Graz, Austria

Bernhard C. Geiger

Signal Processing and Speech Communication Laboratory, Graz University of Technology, Graz, Austria
Know Center Research GmbH, Graz, Austria

Abstract

It was recently shown that the loss function used for training physics-informed neural networks (PINNs) exhibits local minima at solutions corresponding to fixed points of dynamical systems. In the forward setting, where the PINN is trained to solve initial value problems, these local minima can interfere with training and potentially lead to physically incorrect solutions. Building on stability theory, this paper proposes a regularization scheme that penalizes solutions corresponding to unstable fixed points. Experimental results on four dynamical systems, including the Lotka-Volterra model and the van der Pol oscillator, show that our scheme helps avoiding physically incorrect solutions and substantially improves the training success rate of PINNs.

1 Introduction

Physics-informed neural networks (PINNs, [10]) are among the most prominent instantiations of physics-informed machine learning. Capable of including systems of differential equations during training, they have been proposed for solving boundary and initial value problems, for inferring parameters of a differential equation from data and for estimating unobservable scalar or vector fields from measurements. Especially for the forward problem – i.e., for learning the solution to a system of differential equations given boundary and initial conditions – PINNs experience training difficulties for all but the most trivial problem settings.

There is a substantial body of literature that investigates the root causes of and offers remedies for these training difficulties. For example, for stiff problems in which the gradient of the solution function varies strongly across the computational domain, it was shown that reducing the weight of points close to gradient maxima leads to a more well-behaved loss and, hence, successful training [11]. Other studies attribute challenges in PINN training to the use of large computational domains, where techniques such as domain decomposition [5, 8] and sinusoidal feature mappings [15] have

The Third Austrian Symposium on AI, Robotics, and Vision (AIROV26).

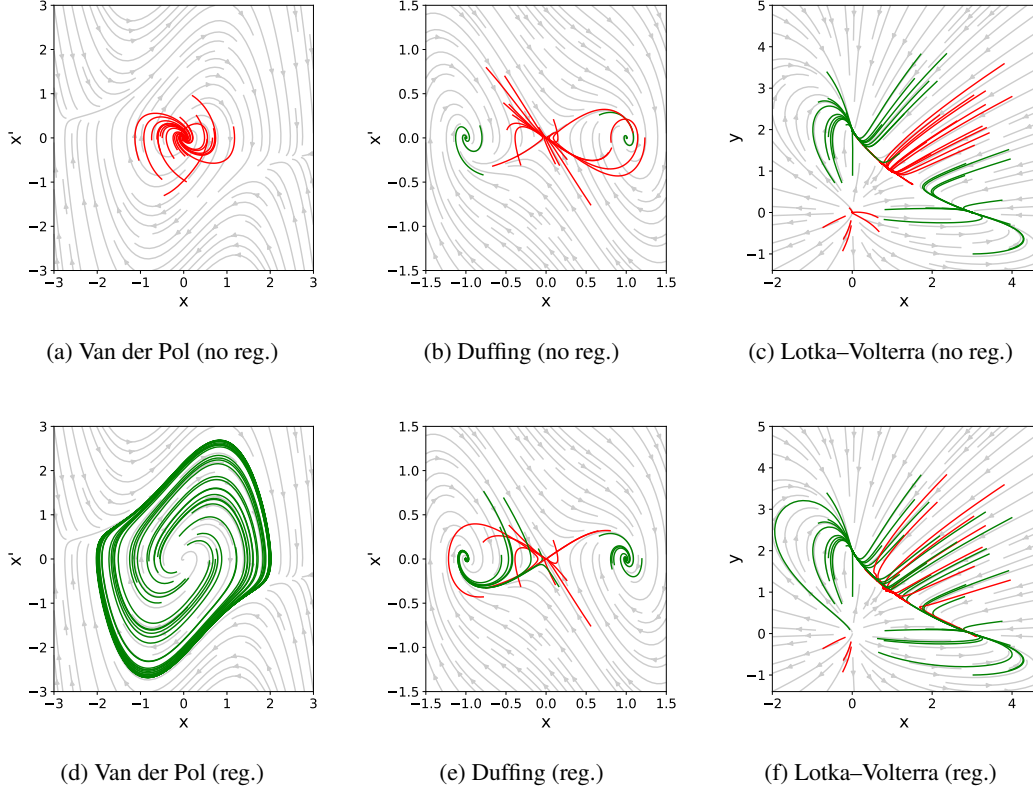


Figure 1: Phase portraits comparing standard PINN training (top row) and regularized training with derivative penalty at unstable fixed points (bottom row) for three dynamical systems: Van der Pol, Duffing, and Lotka–Volterra.

been shown to improve training success. More generally, PINN training is often susceptible to convergence toward the trivial all-zero solution, as standard neural network initializations tend to bias the model toward this outcome. To mitigate this, several strategies have been proposed, including specialized initialization schemes [15], ensemble methods [4], and adaptive collocation point weighting techniques [2, 13].

Interestingly, the all-zero solution is a valid general solution to a large class of differential equations, cf. [15, Prop. 1]. In addition to this all-zero solution, the authors of [12] showed that other, non-zero solutions can become minima of the training loss despite corresponding to unphysical behavior. More specifically, in the context of ordinary differential equations (ODEs), it was shown that solutions at fixed points are always global optima of the physics loss, regardless whether these fixed points are stable or not. While convergence to the all-zero solution can be prevented by some of the remedies mentioned above, convergence to unstable, but non-zero fixed points has not been addressed so far.

In this work, we fill this gap by proposing a regularization scheme that avoids training convergence to unstable fixed points, as illustrated in Fig. 1. More specifically, given a system of ODEs, we linearize the candidate solution around each collocation point and characterize its stability via the eigenvalues of the resulting Jacobian matrix (Section 3). For collocation points for which the candidate solution exhibits unstable behavior, the regularization term is large. Although the approach appears simple, we show in Section 5 that it substantially improves training success for several (low-dimensional) ODEs (Section 4).

2 Background

We focus on autonomous dynamical systems that can be described by ODEs of the general form:

$$x^{(n)} = f(x, x', x'', \dots, x^{(n-1)}), \quad (1)$$

where f is a nonlinear function of time t and the unknown solution function $x(t)$ and where x' , x'' , and $x^{(n)}$ denote the first-, second-, and n th-order derivative with respect to time t . Since any higher-order ODE can be rewritten as a system of first-order ODEs, we consider the multi-dimensional first-order system $x' = f(x)$, where $f = (f_1, f_2, \dots, f_n)$ and $x \in \mathbb{R}^n$. For this system, a *fixed point* $x^* \in \mathbb{R}^n$ satisfies $f(x^*) = 0$. The local stability around x^* can be determined by linearizing the system through Taylor expansion, and then analyzing the resulting Jacobian matrix $J(x^*) = \partial \vec{f} / \partial \vec{x}|_{x^*}$. If all eigenvalues $\{\lambda_1, \dots, \lambda_n\} \subset \mathbb{C}$ of J have negative real parts, the fixed point x^* is asymptotically stable [1]. If at least one eigenvalue has a positive real part, the fixed point is unstable. An unstable fixed point is considered a saddle point if it is stable in some directions and unstable in others.

We use PINNs to approximate the unknown solution $x(t)$ to (1) with a fully connected neural network $x_\theta(t)$, where θ denotes the network's weights and biases. The PINN is trained on both labeled and unlabeled data. Labeled data is used to encode the initial condition (IC) $x(0)$ via the IC loss $\mathcal{L}_{\text{IC}} = \|x_\theta(0) - x(0)\|^2$. Unlabeled data (in the context of PINNs also called *collocation points*) is used to enforce the system dynamics (1) via the physics loss

$$\mathcal{L}_f = \frac{1}{N_{\text{col}}} \sum_{i=1}^{N_{\text{col}}} \|x'_\theta(t_i) - f(x_\theta(t_i))\|^2, \quad (2)$$

where the collocation points $\{t_i\}_{i=1}^{N_{\text{col}}}$ are (randomly) sampled from the temporal domain $0 < t \leq T$, where T is the *simulation time*. Both losses are combined by linear scalarization, i.e. $\mathcal{L} = \mathcal{L}_{\text{IC}} + \mathcal{L}_f$, where an additional regularization loss, central to our study, will be introduced in Section 3.

As discussed in [12], fixed points of ODEs correspond to global optima of the physics loss, characterized by non-trivial basins of attraction. This arises from the fact that the fixed point condition $f(x^*) = 0$ leads to inherently small physics residuals in the vicinity of fixed points. Specifically, the physics loss (2) becomes small when $f(x_\theta^*) \approx 0$, i.e. when the network output x_θ^* is close to a fixed point x^* . As a consequence, regardless of whether a fixed point is stable or unstable, PINN solutions tend to be attracted to fixed points, often resulting in trajectories that converge to the nearest one (as shown in the Fig. 1). Closely related to convergence issues caused by fixed points is the prominent influence of the trivial zero solution, $x^* = 0$, which is a fixed point in many dynamical systems. For instance, Wong et al. [15] propose learning in sinusoidal space, introducing sinusoidal mappings to initialize PINNs with an appropriate input gradient distribution to overcome trainability issues.

3 Method

To prevent the PINN from converging to unstable fixed points, or at least reduce the likelihood of such convergence, we introduce an additional regularization term into the physics loss, designed to modify the loss landscape such that it penalizes predictions near or at unstable fixed points. The proposed regularization term consists of three factors: i) a factor \mathcal{R}_{LS} that penalizes solutions x_θ for which the Jacobian indicates local instability at a given collocation point t_i ; ii) a factor \mathcal{R}_{SE} that ensures that this regularization is only active at collocation points at which the system has converged to a fixed point, i.e., for which $x'_\theta(t_i) = 0$, and iii) a decaying term C that gradually reduces the influence of regularization.

Stability can be easily determined by first converting higher-order ODEs into systems of first-order ODEs and then computing the eigenvalues of the Jacobian matrix obtained through a Taylor expansion around the fixed point. To penalize a candidate solution x_θ that is unstable at a collocation point t^* , we propose the regularization term

$$\mathcal{R}_{\text{LS}}(t^*) = \sum_{\lambda \in \sigma(J(x_\theta(t^*)))} \max(\text{Re}(\lambda), 0), \quad (3)$$

where $\sigma(J(x_\theta(t^*))) = \{\lambda_1^*, \dots, \lambda_n^*\}$ is the spectrum of the Jacobian matrix evaluated at $x_\theta(t^*)$ and Re is the real part.

Since stability is a statement about fixed points, this regularization is not meaningful unless $x'_\theta(t^*) = 0$. We thus multiply $\mathcal{R}_{\text{LS}}(t^*)$ by a factor $\mathcal{R}_{\text{SE}}(t^*)$ that is obtained by applying a Gaussian kernel to the time derivatives of the ODE:

$$\mathcal{R}_{\text{SE}}(t^*) = \exp(-\|x'_\theta(t^*)\|^2/\varepsilon), \quad (4)$$

where ε is a hyperparameter that adjusts the sensitivity of the regularization around the fixed point. $\mathcal{R}_{\text{SE}}(t^*)$ attains its maximum value of one if and only if all time derivatives vanish simultaneously, i.e., if $x_\theta(t^*)$ is precisely at a fixed point. Candidate solutions converging to unstable fixed points are thus regularized by the term

$$\mathcal{R} = \frac{1}{N} \sum_{i=1}^N [\mathcal{R}_{\text{SE}}(t_i) \times \mathcal{R}_{\text{LS}}(t_i)], \quad (5)$$

where the regularization function is evaluated at the collocation points $\{t_i\}_{i=1}^N$ used to compute the physics loss.

Recognizing the importance of initialization in PINN training, cf. [15], it may suffice to steer away the candidate solution x_θ from unstable fixed points early during training. We thus propose a linear decaying scheme that turns off regularization after a certain percentage $\gamma \in (0, 1)$ of epochs:

$$C = \max \left(C_0 \left(\gamma - \frac{\text{epoch}}{N_{\text{epochs}}} \right), 0 \right), \quad (6)$$

where C_0 denotes the initial regularization coefficient and N_{epochs} the total number of epochs. The regularization scheme is finally used in the total loss function via:

$$\mathcal{L} = \mathcal{L}_{\text{IC}} + \mathcal{L}_f + C \times \mathcal{R}. \quad (7)$$

4 Experimental Setup

In our analysis, we aim to demonstrate the impact of our regularization on the PINN training performance across four distinct dynamical systems, governed by first- or second-order ODEs in one- or two-dimensional settings. These systems include the pitchfork bifurcation, unforced Duffing oscillator, van der Pol oscillator, and Lotka-Volterra model.

PINN Settings. We use fully connected PINNs with four hidden layers, each containing 50 units, and employ the Swish activation function. Training is performed using the Adam optimizer with a learning rate of 0.001 and a decay rate of 1.0. During each training epoch, 1024 collocation points are randomly sampled to compute the physics loss (2), and the total number of epochs is set to $25k$. In the case of first-order systems, we use hard constraints for the initial conditions.

Evaluation Procedure. We train PINNs with and without regularization, for different ICs and simulation times, as convergence to local optima is highly sensitive to these settings, cf. [12]. To account for the inherent randomness in neural network initialization, we train 10 different PINNs for each configuration.

Our primary goal is to investigate a failure mode where the system converges to a local optimum with fundamentally different dynamics. Thus, rather than focusing on the accuracy of $x_\theta(t)$, we evaluate training success rates based on whether training converged to an incorrect local optimum or to the true solution $x_{\text{ref}}(t)$. The solution $x_{\text{ref}}(t)$ is obtained using the Runge–Kutta method; for the pitchfork bifurcation system an analytical solution is known (found in [12]). Training is considered successful if the L_2 relative error from the reference solution remains below 0.15.

Hyperparameters. We set the hyperparameter ε to 0.01, the initial regularization coefficient C_0 to 1.0 and the fraction of epochs γ to 0.5. We study the sensitivity of these hyperparameters in Section 5.1.

Candidate Systems: The **pitchfork bifurcation** system can be used to model population dynamics with negative growth effects at both low and high population levels. It is described by the first-order ODE $x' = x - x^3$ and has one unstable fixed point at $x^* = 0$ and two (asymptotically) stable fixed points at $x^* = 1$ and $x^* = -1$. As shown in [12], the unstable fixed point can cause severe convergence issues for PINNs when the simulation time is long and/or the ICs are close to the unstable point. The **unforced damped Duffing oscillator** is described by the second-order ODE $x'' + x' - x + x^3 = 0$, which gives an unstable saddle point at $(x, x')^* = (0, 0)$ and two (asymptotically) stable fixed points at $(x, x')^* = (\pm 1, 0)$. The **Van der Pol oscillator** [9] is a dynamical system described by the second-order ODE of the form $x'' - (1 - x^2)x' + x = 0$. The system exhibits stable oscillations, referred to in the literature as relaxation oscillations, and possesses a single unstable fixed point at $(x, x')^* = (0, 0)$. The **Lotka–Volterra equations** are a well-known system of first-order ODEs used to describe predator–prey dynamics in biological systems. While the

Table 1: Success rates for unmodified (left number) and regularized (right number) PINN training for different dynamical systems across different simulation times (T) and ICs. Bold numbers indicate the best success rate for the considered setting.

	$T = 11$	$T = 12$	$T = 13$	$T = 14$	$T = 15$
$x(0), x'(0)$	Unforced Duffing Oscillator				
(0.1, 0)	0.5 / 0.4	0.0 / 0.5	0.0 / 0.7	0.0 / 0.5	0.0 / 0.6
(0.2, 0)	0.9 / 0.9	0.1 / 0.6	0.0 / 0.3	0.0 / 0.5	0.0 / 0.7
(0.3, 0)	0.9 / 1.0	0.2 / 0.7	0.0 / 0.3	0.0 / 0.2	0.0 / 0.6
(0.4, 0)	1.0 / 1.0	0.2 / 0.6	0.0 / 0.5	0.0 / 0.5	0.0 / 0.4
(0.5, 0)	0.9 / 1.0	0.7 / 0.8	0.0 / 0.4	0.0 / 0.8	0.0 / 0.7
$x(0), x'(0)$	Van der Pol Oscillator				
(0.1, 0)	0.0 / 1.0	0.0 / 0.2	0.0 / 0.6	0.0 / 0.3	0.0 / 0.1
(0.2, 0)	0.6 / 1.0	0.0 / 0.9	0.0 / 0.6	0.0 / 0.5	0.0 / 0.2
(0.3, 0)	0.9 / 1.0	0.2 / 1.0	0.0 / 1.0	0.0 / 1.0	0.0 / 0.4
(0.4, 0)	1.0 / 1.0	0.8 / 1.0	0.1 / 0.9	0.0 / 1.0	0.0 / 0.6
(0.5, 0)	0.9 / 1.0	0.6 / 1.0	0.0 / 1.0	0.0 / 0.9	0.0 / 0.5
$x(0), y(0)$	Lotka–Volterra Model				
(0.0, 0.1)	0.0 / 1.0	0.0 / 1.0	0.0 / 1.0	0.0 / 1.0	0.0 / 1.0
(0.1, 0.0)	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
(0.0, 0.5)	0.0 / 1.0	0.0 / 1.0	0.0 / 0.9	0.0 / 0.8	0.0 / 0.5
(0.5, 0.0)	0.0 / 1.0	0.0 / 1.0	0.0 / 1.0	0.0 / 0.7	0.0 / 0.6
(1.1, 1.1)	1.0 / 1.0	0.3 / 1.0	0.0 / 1.0	0.0 / 1.0	0.0 / 1.0

system’s behavior can be complex depending on its specific formulation, in our work we consider a generalized version of the Lotka-Volterra equations:

$$x' = x(3 - x - 2y) \quad y' = y(2 - x - y) \quad (8)$$

This system exhibits two (asymptotically) stable fixed points at $(x, y)^* = (0, 2)$ and $(x, y)^* = (3, 0)$, and two unstable fixed points at $(x, y)^* = (0, 0)$ and $(x, y)^* = (1, 1)$, with the latter being a saddle point.

5 Results

Training Success Rates. Table 1 shows the quantitative results of our experiments for unmodified PINN training and for our proposed regularization scheme (7) with default hyperparameters as reported in Section 4. We deliberately selected ICs and simulation times such that unmodified PINN training exhibits convergence problems (e.g., the simulation times were chosen longer for the pitchfork bifurcation system than in [12]). Furthermore, we did not apply training modifications proposed previously, such as loss weighting, adaptive collocation point sampling, etc., as they are orthogonal to our regularization scheme. In all of the unsuccessful training runs, training converged to the local optimum corresponding to the unstable fixed point of the dynamical system, confirming the results of [12] that unstable fixed points play a major role in PINN training. As the table shows, regularization substantially improves success rates for all considered systems and problem parameterizations, validating the effectiveness of our proposal even for longer simulation times.

Unforced Duffing and van der Pol Oscillators. For the unforced Duffing and the van der Pol oscillator we present an additional evaluation. Specifically, we fix the simulation time at $T = 12.5$ and $T = 15$ respectively and randomly sample 20 ICs from a Gaussian distribution centered at the unstable fixed point $(x, x')^* = (0, 0)$, with a covariance matrix $\Sigma = 0.25\mathbb{I}$. For the van der Pol system, we increased the number of epochs to $50k$. The simulated trajectories in the phase portrait are shown in Fig. 1. Consistent with the results in Table 1, the unmodified PINN frequently fails by converging to the unstable fixed point, whereas the regularized PINN exhibits much better performance. Indeed, the

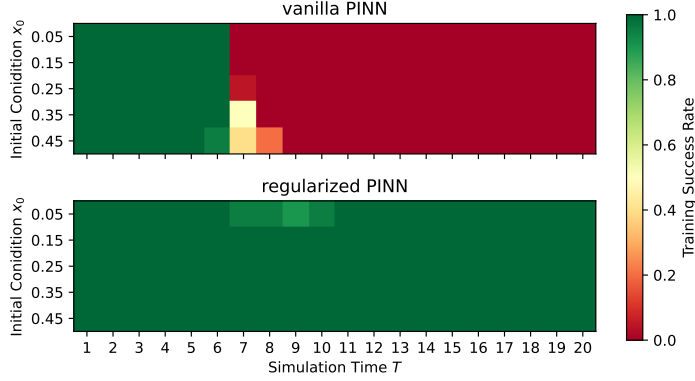


Figure 2: Success rates of unmodified (top) and regularized (bottom) training runs over different initial conditions x_0 and simulation times T .

training success rates are up from 15% to 70% for the Duffing and from 0% to 100% for the van der Pol oscillator.

Pitchfork Bifurcation. For the pitchfork bifurcation system, experiments were performed over a broader range of simulation times, $T \in \{1, \dots, 20\}$, in order to better illustrate the combined effect of simulation time and initial condition on the training results. The results are visualized as a heatmap in Fig. 2. From the figure, we can observe that while the unmodified PINN training begins to fail at approximately $T = 7$, the regularized PINN continues to achieve successful training for all of the tested simulation times. Only for simulation times in the range $7 \geq T \geq 10$ with an IC of $x_0 = 0.05$, the regularized PINN exhibits a slight performance degradation, which we believe to be an artifact of the specific experimental settings and potentially resolvable through more optimized hyperparameter tuning.

Lotka-Volterra Model. For the Lotka–Volterra model, we uniformly sample 50 ICs from the interval $(-1, 4)^2$ along both x and y axis, and use a fixed simulation time of $T = 12.5$. The phase portrait for this system is presented in Fig. 1. While the proposed regularization is not as effective as in the two second-order ODEs, it still yields a notable improvement in training success rate from 62% to 78%.

5.1 Hyperparameter Sensitivity Analysis

We analyze the sensitivity of our regularization scheme w.r.t. the three hyperparameters C_0 , ε and λ . Rather than varying all three hyperparameters simultaneously, which would be computationally expensive, in this experiment we keep two parameters fixed and vary the third.

We conduct the experiment on the Duffing oscillator using the IC $(x_0, x'_0) = (0.01, 0.0)$ and simulation time of $T = 15$. While we believe that the optimal hyperparameters still depend on the considered system, its ICs, and the required simulation time, the results presented in Table 2 suggests that the performance of our regularization scheme depends only mildly on even large variations of the hyperparameters, and always improves upon unmodified PINN training (which in this case has a success rate of 0% for the considered 20 runs).

5.2 Ablation Study

Penalizing candidate solutions that converge to static equilibria via (4) is, when combined with the coefficient decay (6), a valid regularization scheme on its own – at least for systems that do not have (asymptotically) stable fixed points. In this section, we thus investigate the performance of PINN training according to (7), both with and without including the local stability term (3) in the regularization loss.

Table 2: Hyperparameter sensitivity analysis for the proposed regularization scheme. Success rates are shown for 20 training runs for the Duffing oscillator with IC $(x_0, x'_0) = (0.01, 0.0)$ and simulation time $T = 15$. Unmodified PINN training fails for this setting.

ϵ	γ	C	0.001	0.01	0.1	1.0	10.0	100.0	1000.0		
0.01	0.5	Success rate	0	0.3	0.4	0.35	0.6	0.6	0.6		
C	γ	ϵ	0.0001	0.001	0.01	0.1	1.0	10.0	100		
1.0	0.5	Success rate	0.5	0.4	0.45	0.45	0.35	0.6	0.4		
ϵ	C	γ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.01	1.0	Success rate	0.5	0.7	0.45	0.4	0.7	0.75	0.6	0.6	0.4

To this effect, we compare the two schemes on the Duffing oscillator, as its physical trajectories converges to one of the asymptotically stable fixed points, a property for which taking stability into account seems as the most sensible approach. We also compared the two schemes for the Van der Pol oscillator, a system for which the regularization scheme shown in (4) is, in theory, a valid scheme on its own as the system does not exhibit any stable fixed point. We perform the experiment on a single simulation time and IC, but optimized the PINN performance over all three hyperparameters in a grid search for each regularization scheme on each system separately. We chose a simulation time of $T = 15$ for both systems and ICs $(x_0, x'_0) = (0.01, 0.0)$ for the Duffing oscillator and $(x_0, x'_0) = (0.1, 0.0)$ for the Van der Pol oscillator. These settings were deliberately chosen to avoid perfect success rates for either approach, thereby creating challenging conditions that highlight meaningful differences between the two methods. Table 3 reports the highest success rates achieved during hyperparameter optimization. As shown, the best performance achieved by the combined loss function (7) is either comparable to (in the case of the van der Pol oscillator) or superior to (in the case of the unforced Duffing oscillator) that achieved by the static equilibrium regularization (4) alone. (There may not be an ordering between the two regularization schemes for fixed sets of hyperparameters, however.) Thus, we conclude that regularizing against unstable fixed points is necessary for systems that have asymptotically stable fixed points, and does not deteriorate performance for systems that do not.

Table 3: Comparing regularization with and without taking local stability \mathcal{L}_{LS} into account. The success rate computed over 20 training runs, maximized over all hyperparameter combinations, is shown.

Unforced Duffing Oscillator ($T = 15, x_0 = 0.01$)		Van der Pol Oscillator ($T = 15, x_0 = 0.1$)	
Regularization	Success	Regularization	Success
–	0	–	0
\mathcal{R}_{SE}	0.6	\mathcal{R}_{SE}	0.7
$\mathcal{R}_{SE} \times \mathcal{R}_{LS}$	0.7	$\mathcal{R}_{SE} \times \mathcal{R}_{LS}$	0.7

6 Discussion, Limitation & Outlook

Training Difficulties due to Fixed Points. In [12], it was shown that fixed points—regardless of their local stability—correspond to local minima in the physics loss function, which can lead to severe convergence issues in PINNs, particularly for long simulation times and/or when ICs are close to these fixed points. Closely related are common training difficulties in PINNs caused by the zero solution $x^* = 0$, which is a trivial solution and fixed point for many dynamical systems. This solution is especially attractive due to common network initialization schemes that tend to bias the model toward it. Our experiments on the Lotka–Volterra model demonstrated that fixed points may cause convergence issues not only when the IC is close to them, but also when trajectories pass near them—even if the IC is set far away. This specific scenario is illustrated in the case of

Lotka-Volterra system in Fig. 1, where simulated trajectories are drawn toward the unstable fixed point at $(x, y)^* = (1, 1)$, despite originating from regions that are initially far away.

Overall Performance. Our experiments consistently highlight the prominent role of fixed points in training PINNs, demonstrating that the vanilla, unmodified PINN framework struggles to simulate dynamical systems in which fixed points interfere—either because trajectories originate near them or pass close to them during the simulation. With our proposed regularization scheme, introduced in Section 3, we specifically aim to mitigate these training difficulties by leveraging both the fixed point condition, which states that $f(x^*) = 0$, and the local stability information provided by the eigenvalues $\{\lambda_i\}$ of the Jacobian $J(x^*)$. Our results on various dynamical systems demonstrate that incorporating the regularization scheme into the overall loss function substantially improves training success rates. This improvement was shown quantitatively in Table 1 and qualitatively in Figures 1 and 2.

Limitations due to Saddle Points. Although the training success rate improved across many tested settings, our results also indicate that the proposed regularization scheme does not fully eliminate training difficulties. Manual inspection of the training process (not shown in the manuscript) revealed that many unsuccessful training outcomes—despite the proposed regularization being active—can be traced to the presence of saddle points, which constitute a specific class of unstable fixed points. Saddle points are characterized by the coexistence of stable and unstable manifolds: along certain directions (corresponding to negative eigenvalues of the Jacobian), trajectories are attracted toward the fixed point, while along others (positive eigenvalues), they are repelled (cf. Fig 1). This mixed stability behavior leads to a critical sensitivity to ICs. In particular, the stable manifold of a saddle point effectively acts as a *separatrix*, dividing the phase space into regions of qualitatively different dynamics. If the candidate solution $x_\theta(t)$ at network initialization lies on the “wrong side” of this separatrix, the regularization term—which is designed to repel trajectories from fixed points—may inadvertently drive the solution further into the repelling region. Once the regularization decays or is deactivated, the dynamics governed solely by the physical loss may guide the trajectory back toward the saddle point, ultimately resulting in convergence failure. In contrast, if the initial candidate solution lies on the “correct side” of the stable manifold, the regularization successfully steers the trajectory away from the saddle point, allowing it to follow the correct streamlines toward the physically meaningful solution. This limitation highlights the importance of network initialization in the presence of saddle-type fixed points, where the delicate geometry of the phase space can significantly affect training outcomes.

Outlook. Recognizing this limitation of the proposed regularization scheme motivates the discussion of combining it with existing techniques from the literature. For instance, sequence-to-sequence learning [6] or causality-respecting methods [14, 7, 3] can be naturally integrated with our approach to progressively extend the region of the simulation. These methods allow the training process to begin in a narrow temporal window around the IC and gradually expand toward the full simulation domain as training progresses. This approach could potentially circumvent the limitations introduced by saddle points, as their separatrices would not significantly influence the optimization during the initial training phase.

Finally, we would like to comment on the role of the static equilibrium regularization term (4) within the overall regularization framework. This component is designed to prevent PINN training from becoming trapped in trivial solutions that remain constant over time. Although its effectiveness was demonstrated only for systems governed by ODEs, we believe that it is also applicable to dynamical systems described by partial differential equations. Given that PINN training has also been shown to be adversely affected by steady-state solutions in PDE-based systems [12], the static equilibrium regularization may help push the solution away from such steady states during the early stages of training, thereby facilitating convergence toward transient, physically meaningful dynamics.

7 Conclusion

In this work, we introduced a regularization scheme for avoiding unstable fixed points when training PINNs on dynamical systems governed by ODEs. This approach proves especially useful in challenging scenarios, such as those involving long simulation times or initial conditions near unstable fixed points, where training would otherwise be difficult or infeasible. While this method does not entirely

eliminate convergence to local optima, it can be effectively combined with other strategies designed to mitigate such failure modes.

Although ODEs remain relevant in many fields, including control theory, systems biology, and mechanical engineering, the extension of this approach to partial differential equations (PDEs) is a compelling direction for future work. This is not a trivial step, as analyzing the stability of steady-state solutions in PDEs is significantly more complex. Nonetheless, our results demonstrate that explicitly addressing unstable fixed points can substantially improve PINN training, providing a promising foundation for broader applications.

References

- [1] W.E. Boyce and R.C. DiPrima. *Elementary differential equations and boundary value problems*. 8th. Wiley New York, 2004.
- [2] Arka Daw et al. “Mitigating Propagation Failures in Physics-informed Neural Networks using Retain-Resample-Release (R3) Sampling”. In: *Proc. Int. Conf. on Machine Learning (ICML)*. Honolulu, Hawaii, USA, July 2023, pp. 7264–7302.
- [3] Jia Guo, Haifeng Wang, and Chenping Hou. “A novel adaptive causal sampling method for physics-informed neural networks”. In: *arXiv preprint arXiv:2210.12914* (2022).
- [4] Katsiaryna Haitsiukevich and Alexander Ilin. “Improved training of physics-informed neural networks with model ensembles”. In: *Proc. Int. Joint Conf. on Neural Networks (IJCNN)*. 2023, pp. 1–8.
- [5] Ameya D Jagtap and George Em Karniadakis. “Extended physics-informed neural networks (XPINNs): A generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations”. In: *Communications in Computational Physics* 28.5 (2020).
- [6] Aditi Krishnapriyan et al. “Characterizing possible failure modes in physics-informed neural networks”. In: *Advances in neural information processing systems* 34 (2021), pp. 26548–26560.
- [7] Rambod Mojtani, Maciej Balajewicz, and Pedram Hassanzadeh. “Lagrangian PINNs: A causality-conforming solution to failure modes of physics-informed neural networks”. In: *arXiv preprint arXiv:2205.02902* (2022).
- [8] Ben Moseley, Andrew Markham, and Tarje Nissen-Meyer. “Finite basis physics-informed neural networks (FBPINNs): a scalable domain decomposition approach for solving differential equations”. In: *Advances in Computational Mathematics* 49.4 (2023), p. 62.
- [9] Balth. van der Pol Jun. “LXXXVIII. On “relaxation-oscillations””. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11 (1926), pp. 978–992.
- [10] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations”. In: *Journal of Computational Physics* 378 (2019), pp. 686–707.
- [11] Franz M. Rohrhofer et al. “Approximating Families of Sharp Solutions to Fisher’s Equation with Physics-Informed Neural Networks”. In: *Computer Physics Communications* 307 (2025). open-access: [arXiv:2402.08313](https://arxiv.org/abs/2402.08313) [cs.LG], p. 109422. DOI: [10.1016/j.cpc.2024.109422](https://doi.org/10.1016/j.cpc.2024.109422).
- [12] Franz M. Rohrhofer et al. “On the Role of Fixed Points of Dynamical Systems in Training Physics-Informed Neural Networks”. In: *Trans. Machine Learning Research* 1 (2023). Open-access.
- [13] Sifan Wang, Shyam Sankaran, and Paris Perdikaris. “Respecting causality for training physics-informed neural networks”. In: *Computer Methods in Applied Mechanics and Engineering* 421 (2024), p. 116813. DOI: [10.1016/j.cma.2024.116813](https://doi.org/10.1016/j.cma.2024.116813).
- [14] Sifan Wang, Shyam Sankaran, and Paris Perdikaris. “Respecting causality for training physics-informed neural networks”. In: *Computer Methods in Applied Mechanics and Engineering* 421 (2024), p. 116813.
- [15] Jian Cheng Wong et al. “Learning in Sinusoidal Spaces With Physics-Informed Neural Networks”. In: *IEEE Transactions on Artificial Intelligence* 5.3 (2024), pp. 985–1000. DOI: [10.1109/TAI.2022.3192362](https://doi.org/10.1109/TAI.2022.3192362).