SUPERVISED METRIC REGULARIZATION THROUGH ALTERNATING OPTIMIZATION FOR MULTI-REGIME PHYSICS-INFORMED NEURAL NETWORKS

Enzo Nicolás Spotorno, Josafat Leal Filho, & Antônio Augusto Fröhlich
Department of Informatics and Statistics
Federal University of Santa Catarina
{enzoniko, josafat, guto}@lisha.ufsc.br

ABSTRACT

Standard Physics-Informed Neural Networks (PINNs) often face challenges when modeling parameterized dynamical systems with sharp regime transitions, such as bifurcations. In these scenarios, the continuous mapping from parameters to solutions can result in spectral bias or "mode collapse", where the network averages distinct physical behaviors. We propose a **Topology-Aware PINN (TAPINN)** that aims to mitigate this challenge by structuring the latent space via **Supervised Metric Regularization**. Unlike standard parametric PINNs that map physical parameters directly to solutions, our method conditions the solver on a latent state optimized to reflect the metric-based separation between regimes, showing $\approx 49\%$ lower physics residual (0.082 vs. 0.160). We train this architecture using a phase-based **Alternating Optimization (AO)** schedule to manage gradient conflicts between the metric and physics objectives. Preliminary experiments on the Duffing Oscillator demonstrate that while standard baselines suffer from spectral bias and high-capacity Hypernetworks overfit (memorizing data while violating physics), our approach achieves stable convergence with $\mathbf{2.18}\times$ lower gradient variance than a multi-output Sobolev Error baseline, and significantly fewer parameters than a hypernetwork-based alternative.

1 INTRODUCTION

Physics-Informed Neural Networks (PINNs) have demonstrated significant potential for solving parameterized dynamical systems. However, modeling systems that exhibit qualitative behavioral changes, such as bifurcations from stability to chaos, remains a challenge. The literature suggests that standard Multi-Layer Perceptrons (MLPs) struggle to approximate the discontinuous or non-smooth dependence of the solution on the system parameters due to spectral bias (Rahaman et al., 2019) and the singularity of the Jacobian at bifurcation points (Krishnapriyan et al., 2021). While spectral bias reflects a representational limitation (Rahaman et al., 2019), ill-conditioned Jacobians near bifurcations primarily induce optimization pathologies (Krishnapriyan et al., 2021). Recent works have also explored PINNs for bifurcation detection in nonlinear lattice systems (Shahab et al., 2025). Nevertheless, existing solutions typically involve specialized architectures. **Hyper-PINNs** (Almeida et al., 2021) generate network weights conditioned on parameters to handle these transitions, while aiming to maintain parameter efficiency compared to training separate networks for each parameter. Other approaches, such as Mixture-of-Experts (MoE) (Bischof & Kraus, 2022), explicitly route inputs to different sub-networks. These methods, however, introduce challenges such as routing instability (MoE) or indirect weight generation overhead (HyperPINNs).

In this work, we investigate a single-network architecture that incorporates **Supervised Metric Learning** (Triplet Loss) to regularize the latent space, and we name it **Topology-Aware PINN (TAPINN)**. We hypothesize that by explicitly enforcing a geometric structure in the latent space that mirrors the physical regimes, we can improve the trainability of the solver without the computational overhead of hypernetworks. To stabilize the training of these competing objectives (topology vs. physics), we employ an Alternating Optimization schedule. We choose an LSTM-based encoder to capture temporal dependencies in short observation windows, which is crucial for distinguishing

periodic from chaotic trajectories. Triplet loss is employed to cluster embeddings from trajectories in similar dynamical regimes while separating dissimilar ones, drawing from representation learning principles (Bengio et al., 2013) and manifold assumptions in high-dimensional data (Schroff et al., 2015). Triplets are formed in-batch using known forcing amplitudes F_0 as proxies for regime similarity (anchors/positives share the same F_0, negatives differ; no hard/semi-hard mining); we use Euclidean distance and margin = 0.2. This provides supervision without discrete regime labels. The alternating optimization schedule follows successful block-coordinate strategies in multi-objective PINNs (Wang et al., 2021), ensuring the latent manifold is stabilized before heavy physics enforcement and avoiding severe gradient conflicts. This phased approach is inspired by strategies for mitigating gradient pathologies in multi-objective PINNs (Wang et al., 2021).

2 METHODOLOGY

Below, we formalize the problem setting, describe the proposed encoder-generator architecture, and detail the loss functions and phased training procedure. We focus on a parameterized dynamical system governed by the residual $\mathcal{N}[\mathbf{x}(t); \lambda] = 0$. The objective is to reconstruct the full solution trajectory $\hat{\mathbf{x}}(t)$ given a short observation window $\mathbf{x}_{obs}$ associated with a parameter λ. Unlike standard Initial Value Problem (IVP) solvers that propagate forward from a single state $\mathbf{x}(0)$, our data-assimilation approach requires the encoder to process an observation window of the first 100 timesteps (10% of the trajectory) to accurately infer the underlying dynamical regime before reconstruction. We specifically address the regime where $\nabla_\lambda \mathbf{x}$ may be ill-conditioned (e.g., near a bifurcation), making direct regression from λ difficult for standard optimizers. Here, λ is scalar (forcing amplitude F_0). We assume piecewise smoothness away from bifurcations. Unlike parametric baselines, λ is used only indirectly via observed trajectories $\mathbf{x}_{obs}$.

TAPINN consists of an LSTM-based Encoder (E) and a PINN Generator (G). We use the term "Topology-Aware" to indicate that the method explicitly structures the metric geometry of the latent space to mirror separation between dynamical regimes (rather than computing strict topological invariants). The **Encoder:** $z = E(\mathbf{x}_{obs})$. Maps the input window to a latent vector z. The **Generator:** $\hat{\mathbf{x}}(t) = G(t, z)$, implemented as a 4-layer MLP with 32 hidden units and $\tanh$ activations. Unlike parametric baselines that require the known parameter λ, our method infers regime information solely from the short observation window $\mathbf{x}_{\text{obs}}$ (first 100 timesteps), targeting realistic data-assimilation settings where λ is unknown. LSTM is selected for its effectiveness in capturing long-term dependencies in dynamical trajectories; simpler alternatives (e.g., 1D CNNs) were explored but showed inferior regime discrimination in preliminary tests, therefore we left deeper experimentation for future work. We define a composite loss function $\mathcal{L}_{total} = \mathcal{L}_{data} + \alpha\mathcal{L}_{physics} + \beta\mathcal{L}_{metric}$. α and β are tuned via grid search (α=1.0, β=0.1 in our experiments); adaptive strategies remain future work (Wang et al., 2021).

To mitigate gradient conflicts observed when optimizing these terms simultaneously (Wang et al., 2021), we utilize a block-coordinate descent strategy. **Phase I (Metric Alignment)** optimizes the Encoder using a Triplet Loss $\mathcal{L}_{metric} = \max(0, d(z_a, z_p) - d(z_a, z_n) + m)$. This step organizes the latent space such that inputs from the same physical regime are clustered together. Intuitively, this geometric structuring forces the latent space to act as a linearized surrogate parameter manifold. By artificially grouping trajectories with similar forcing amplitudes and separating disparate ones, we reduce the effective condition number of the parameter-to-solution mapping, preventing the generator from collapsing into a spectral average of distinct periodic and chaotic regimes. This creates a more linearly separable representation, supporting manifold learning techniques like t-SNE for visualization (Maaten & Hinton, 2008). **Phase II (Physics Reconstruction)** optimizes the Generator on $\mathcal{L}_{physics}$ while freezing the Encoder, providing the solver with a stable conditioning variable z.

Interleaved Joint Tuning: We train over 30 epochs as follows: Phase I (5 epochs, encoder-only on $\mathcal{L}_{metric}$), Phase II (20 epochs, generator-only on $\mathcal{L}_{physics} + \mathcal{L}_{data}$ with encoder frozen), then interleaved joint updates on $\mathcal{L}_{total}$ every $k = 5$ batches ($\approx 20\%$ of steps). This stabilizes the latent manifold early while keeping the encoder adaptable, preventing generator overfitting to a stale representation. By encouraging a latent space that linearizes the parameter-to-solution manifold, the approach may implicitly reduce the effective condition number of the Jacobian seen by the generator, potentially alleviating optimization difficulties near bifurcations identified in prior work (Krishnapriyan et al., 2021).

Table 1: **Performance Comparison.** We report Physics Residual (mean squared ODE residual $\mathcal{L}_{physics} = \frac{1}{N_c}\sum \|\mathcal{N}[\hat{\mathbf{x}}(t_i);\lambda]\|_2^2$ over $N_c = 10^4$ uniformly sampled collocation points $t_i \in [0, T]$; no additional weighting), Model Complexity (Exact Parameters), and Test MSE. Note that while HyperPINN achieves low Data MSE, its high Physics Residual indicates overfitting. Our method achieves the best physics compliance with $5\times$ fewer parameters than HyperPINN. All other models have a similar parameter count for fairness. Since the Parametric baseline doesn't use the encoder we matched its parameter count by increasing the MLP capacity.

Method	**Physics Res.** (↓)	**Params**	**Data MSE** (↓)
Parametric Baseline	0.160	8,577	0.392
Multi-Output	0.192	8,069	0.426
HyperPINN	0.158	39,169	**0.281**
Ours (AO)	**0.082**	8,003	0.425

3 Experiments and Results

We evaluate the method on the **Duffing Oscillator**, a canonical system exhibiting chaos. The system is governed by $\ddot{x}+\delta\dot{x}+\alpha x+\beta x^3 = F_0\cos(\omega t)$, with standard parameters $\delta = 0.3$, $\alpha = -1$, $\beta = 1$, $\omega = 1$. Varying the forcing amplitude $F_0 \in [0.3, 0.8]$ induces transitions from periodic to chaotic regimes. Such multi-scale dynamical systems are known to induce severe optimization pathologies in PINNs, as characterized by (Krishnapriyan et al., 2021). While the Duffing system exhibits clear and relatively sharp regime transitions, subtler or higher-codimension bifurcations in systems such as the Lorenz attractor or reaction-diffusion equations (e.g., Allen-Cahn), broader validation is left for future work. Simulating regimes from Period-1 to Chaos by varying the forcing amplitude $F_0 \in [0.3, 0.8]$, we generated a dataset of 500 trajectories for each regime ($F_0 \in \{0.3, 0.5, 0.8\}$) using a 4th-order Runge-Kutta solver with $dt = 0.01$. The observation window consists of the first 100 timesteps (10% of the trajectory). Models were trained for 30 epochs using the Adam optimizer with a learning rate of 10^{-3}.

We compare our approach against three baselines. **(I) Parametric Baseline:** A standard PINN, a 4-layer MLP with 64 hidden-units each, that receives the explicit parameter λ as input ($Input : t, \lambda$). Unlike this and the HyperPINN baseline (both conditioned on known λ), TAPINN infers regime from partial observations $\mathbf{x}_{obs}$, targeting data-assimilation settings where λ is unknown. **(II) HyperPINN Baseline:** A hypernetwork architecture where weights are predicted as functions of λ. This represents a high-capacity alternative for multi-regime modeling, designed for parameter efficiency (Almeida et al., 2021). **(III) Multi-Output Baseline:** Identical to our architecture (LSTM Encoder + Generator) but trained via **standard joint optimization** with a Sobolev (H^1) loss that enforces constraints on both $\mathbf{x}$ and $\dot{\mathbf{x}}$ (Czarnecki et al., 2017). This isolates the combined effect of our Metric Regularization and Alternating Optimization against standard derivative-constrained training.

In our experiments, the Multi-Output baseline exhibited severe numerical instability. While Sobolev training is effective for smooth functions, the inclusion of gradient constraints near the bifurcation point resulted in gradient norms $\mathbf{2.14}\times$ higher on average compared to our method, with a variance $\mathbf{2.18}\times$ larger (see Figure 1a). This suggests that enforcing high-order constraints on the encoder without explicit metric regularization leads to an unstructured latent space, exacerbating optimization pathologies near bifurcations. Moreover, we include HyperPINNs as a high-capacity benchmark. As shown in Table 1, HyperPINN suffers from a "memorization" pathology: it achieves the lowest Data MSE (0.281) but a high Physics Residual (0.158), indicating it overfits trajectory points without satisfying the governing ODE. One might question whether TAPINN's lower physics residual is merely a regularization artifact of its smaller capacity compared to HyperPINN (8,003 vs. 39,169 parameters). However, our ablation demonstrates otherwise. Both the Parametric and Multi-Output baselines operate at the same ≈ 8k capacity but yield physics residuals approximately $2\times$ higher than our approach (0.160 and 0.192 vs. 0.082). This confirms that the topology-aware structuring, rather than merely restricted capacity, is the primary driver of the improved physical compliance.

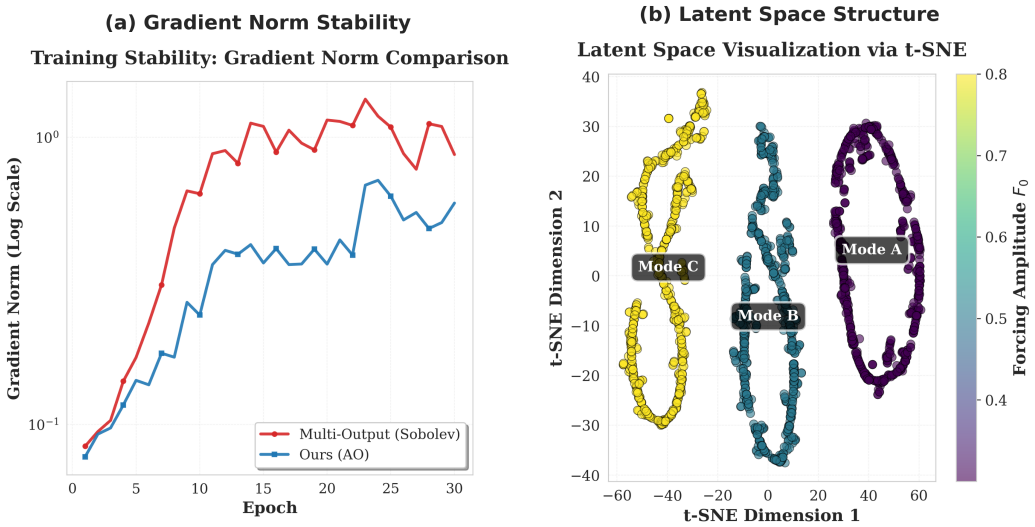

Figure 1: Qualitative analysis of the Topology-Aware training process: (a) Gradient norms during training. The Multi-Output baseline (orange) shows significant spikes near regime transitions compared to ours (blue). (b) t-SNE visualization of the learned embeddings z, colored by regime (F_0). Distinct clusters emerge without discrete supervision.

In contrast, our Topology-Aware approach achieves the lowest physics error (0.082) while using substantially fewer parameters ($8,003$ vs $39,169$), demonstrating that structuring the latent space can be more parameter-efficient than brute-force weight generation. Beyond visual clustering, we quantify the structure of the learned latent space z by training a linear probe to regress the physical parameter F_0 from z. Our method achieves a **Prognostics MSE** of 3.5×10^{-4}, confirming that the encoder learns a highly structured, linearizable representation of the chaotic regimes (see Figure 1b), which is not explicitly enforced by the physics loss alone. While this initial proof-of-concept focuses on characterizing the optimization stability and the resulting latent space structure, a rigorous comparison of the reconstructed trajectories against ground truth in physical coordinates remains a priority for extended validation. Finally, to isolate the contribution of our phased training strategy versus the metric loss itself, we compare our approach against a version trained with standard joint optimization (summing Metric, Physics, and Data losses). This "Joint Training" configuration yielded a significantly higher physics residual (≈ 0.158), performing nearly identically to the standard baseline (0.160). This failure confirms that metric regularization alone is insufficient; the Alternating Optimization schedule is critical to resolve gradient conflicts, allowing the latent topology to stabilize before being subjected to the competing gradients of the physics constraints.

4 Conclusion

We presented an exploratory proof-of-concept Topology-Aware PINN framework that leverages Supervised Metric Regularization and Alternating Optimization to effectively address gradient conflicts and regime transitions in multi-regime dynamical systems. The encouraging preliminary results on the Duffing oscillator, demonstrating substantially improved physics compliance, training stability, and well-structured latent representations, highlight the promise of explicitly organizing the latent space. These findings naturally open several exciting avenues for future work, including theoretical analyses of Jacobian conditioning and spectral properties, rigorous statistical validation across random seeds, sensitivity studies on observation-window length, and broader empirical evaluation on PDE systems, continuous parameter regimes, and noisy data. In addition, comparisons with capacity-matched variants of high-parameter architectures such as HyperPINNs will provide deeper insight into the benefits of metric regularization. Lastly, future work will also focus on extensive benchmarking of TAPINN against domain decomposition methods (Jagtap & Karniadakis, 2020) and operator learning frameworks (Li et al., 2021). We believe this lightweight approach offers a practical and efficient pathway toward more robust physics-informed modeling of complex dynamical systems and will be extended in subsequent research.

ACKNOWLEDGMENTS

This work was partially funded by Fundação de Apoio da UFMG (Fundep), through Linha VI – Conectividade Veicular, a prioritary program from Mover (Mobilidade Verde e Inovação), project AutoDL (29271.03.01/2023.04-00) and Auto5G (29271.02.01/2022.01-00).

REFERENCES

Diogo Almeida et al. HyperPINNs: Learning parameterized differential equations with physics-informed hypernetworks. *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.

Rafael Bischof and Michael A Kraus. Mixture-of-experts-ensemble meta-learning for physics-informed neural networks. In *Proceedings of 33. forum bauinformatik*, 2022.

Wojciech M Czarnecki et al. Sobolev training for neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Ameya D Jagtap and George Em Karniadakis. Extended physics-informed neural networks (XPINNs). *Communications in Computational Physics*, 2020.

Aditi Krishnapriyan et al. Characterizing possible failure modes in physics-informed neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Zongyi Li et al. Fourier neural operator for parametric partial differential equations. *International Conference on Learning Representations (ICLR)*, 2021.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

Nasim Rahaman et al. On the spectral bias of neural networks. *International Conference on Machine Learning (ICML)*, 2019.

Florian Schroff et al. FaceNet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

Muhammad Luthfi Shahab, Fidya Almira Suheri, Rudy Kusdiantara, and Hadi Susanto. Physics-informed neural networks for high-dimensional solutions and snaking bifurcations in nonlinear lattices. *Physica D: Nonlinear Phenomena*, pp. 134836, 2025.

Sifan Wang et al. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 2021.