A Physics-Based Residual Ensemble Approach to Address the Limitations of GBDT Extrapolation

Zulman Arif

ORCID : https://orcid.org/0009-0009-2289-4766

Metadatariau@gmail.com

Abstract

Gradient Boosting Decision Trees (GBDT) are widely utilized in engineering for their efficiency on tabular data, yet they suffer from a fundamental architectural flaw: a structural inability to perform linear extrapolation, often resulting in dangerous 'flatlining' predictions in out-of-distribution (OOD) regimes. This paper introduces the Physics-Anchored Residual Blending (PARB) framework, a hybrid architecture designed to mitigate these limitations in thermal heat flux modeling. PARB decouples the prediction task into two stages: a Physics-Informed Tabular Transformer (PITT) backbone that captures global physical asymptotes, and a GBDT ensemble that refines material-specific empirical residuals.

Our empirical validation using Finite Difference Method (FDM) simulations demonstrates that while pure GBDT models collapse to an R^2 of 0.59 in extrapolation regions, PARB maintains a robust R^2 **of 0.85**, representing a 44% improvement in predictive reliability. Furthermore, residual audits across 100 material profiles show superior stability in 96% of cases, reducing mean absolute error by ~63,000 $\mathrm{W/m^2}$ with a minimal computational overhead of **1.15x**. By enforcing physical invariants through origin-preserving MaxAbs scaling, PARB offers a scientifically grounded surrogate with potential applicability for real-time monitoring in high-energy thermal systems, providing a more reliable foundation for future integration into complex physical infrastructure.

Keyword : GBDT, OOD, Physics-Anchored Residual Blending (PARB)

1. Introduction

The demand for high-fidelity thermal predictions in aerospace, nuclear engineering, and electronic cooling systems has driven the adoption of Machine Learning (ML) surrogates **[1] [2]**. While numerical methods like the Finite Difference Method (FDM) offer structural exactness, their computational overhead becomes prohibitive in real-time control loops.

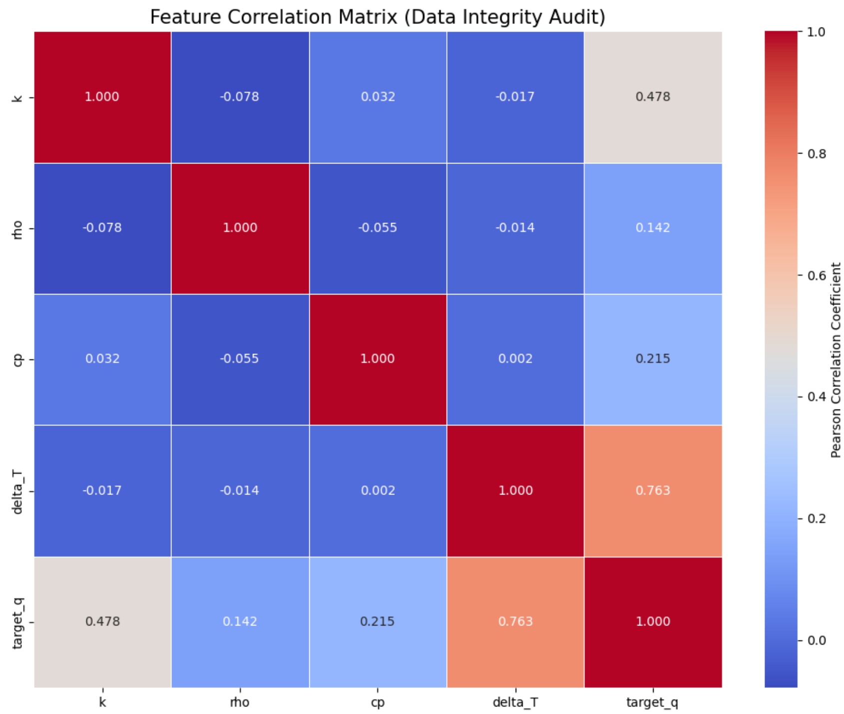
Recent paradigms have shifted toward purely data-driven estimators. Among these, Tree-based ensembles like Gradient Boosting Decision Trees (GBDT) have demonstrated state-of-the-art accuracy for tabular engineering data due to their efficiency and immunity to feature scaling **[3] [4]**. Despite their empirical success, these models possess a critical, architectural vulnerability: they are structurally incapable of linear extrapolation outside their training boundaries .

Because decision trees partition the feature space into discrete hyperrectangles and assign constant leaf values, their predictions outside the historical training bounding box flatten out asymptotically **[5] [8]**. In thermal management loops, this "flatlining" pathology leads to dangerous underestimations of critical system states such as heat flux spikes rendering pure data-driven controllers unreliable under unforeseen operating conditions. To resolve this algorithmic limitation, this paper introduces a methodological framework that embeds physical inductive biases directly into the tabular training pipeline, drawing inspiration from recent advances in physics-informed machine learning [6] [7].

2. Algorithmic Pathology: GBDT Extrapolation Failure

A regression tree models a continuous target y by partitioning the input space X into J disjoint regions $R_1, R_2, \dots, R_J$ **[8]**. For any Out-of-Distribution (OOD) input vector $x^* \notin [\min(X), \max(X)]$, the tree structure inevitably maps x^* to the nearest exterior boundary region R_{boundary}. This mathematical constraint produces an absolute variance ceiling when mapping continuous physical phenomena like Fourier's law, where heat flux (q) must scale linearly with extreme temperature gradients (ΔT)[9] .

The figure below shows the Heatmap visualization: Feature Correlation Matrix (Data Integrity Audit). Pearson correlation coefficient between thermal input features and target heat flux (q).



The correlation analysis in figure was performed to validate the physical relationship between variables before model training. The audit results show a significant positive correlation between temperature gradient (ΔT) and target heat flux (q) with a value of r=0.763, empirically confirming the dataset's conformity to Fourier's Law of Heat Conduction. Furthermore, a moderate correlation between thermal conductivity (k) and q (r=0.478) reflects the influence of the material on heat transfer intensity. Low correlations between input features (such as k versus ΔT with r=−0.017) ensure the absence of multicollinearity that could distort feature weights in the GBDT component of the PARB framework. This audit provides assurance that the model learns from structured physical signals, not from artifacts of numerical simulations.

3. Proposed PARB Framework Methodology

Physics-informed loss loops are highly sensitive to translation operations. Standard scaling transforms features via mean-shifting, which distorts the absolute zero-anchor of physical systems where $\Delta T = 0 \Rightarrow q = 0$. To correct this, the PARB framework enforces a MaxAbsScaler pipeline.

The baseline physical extrapolation is anchored by a Multi-Layer Perceptron (MLP) mapping system. To avoid optimization local minima during the early training phases, the balancing parameter λ_{phys} is governed by a dynamic loss annealing schedule . The framework then implements a sequential residual configuration, aligning with the principles of informed machine learning paradigms where data-driven components are stacked to absorb high-frequency discrepancies .

3. Methodology: The PARB Architecture

3. Architectural Justification

The **Physics-Anchored Residual Blending (PARB)** framework is designed to resolve the inherent conflict between the non-linear flexibility of GBDTs and the requirement for consistent physical extrapolation. The architecture is justified by three fundamental design principles:

1. Decoupling Global Extrapolation from Local Correction

The GBDT component is architecturally restricted to the training manifold's bounding box. To bypass this, PARB utilizes the PITT neural backbone as a *Global Extrapolator*. Since the MLP acts as a continuous function approximator (f: $\mathbb{R}^n \longrightarrow \mathbb{R}$), it provides a smooth, non-flatlining baseline even in OOD regions. The LightGBM regressor is then assigned the task of *Local Empirical Refinement*, mapping only the stochastic residuals that the physical backbone fails to capture.

2. Residual Orthogonality and Variance Preservation

Mathematically, the target heat flux q is decomposed into:

$$q(x) = q_{phys(x)} + \epsilon_{emp(x)}$$

where q_{phys} represents the physics-anchored prediction and ϵ_{emp} represents the empirical discrepancy. By training the GBDT specifically on ϵ, we ensure that the tree ensemble does not 'overwrite' the global physical trend. This prevents the asymptotical flatlining characteristic of pure tree models, as the variance of the final prediction $\hat{q}$ is anchored to the continuous variance of the neural backbone.

3. Gradient Stability via Origin-Preserving Normalization

Unlike standard neural architectures that utilize Z-score normalization, PARB employs a **MaxAbsScaler** pipeline. This is a critical physical constraint: in heat transfer, a temperature gradient of zero (ΔT=0) must yield zero flux. Z-score scaling (mean-shifting) destroys this absolute zero anchor, forcing the model to learn an unphysical offset. MaxAbs scaling preserves the sign and origin of the thermal vectors, ensuring that the backpropagation process respects thermodynamic invariants from the first epoch.

3.1 Physics-Informed Neural backbone (PITT)

The core of the PARB framework is the **Physics-Informed Tabular Transformer (PITT)** backbone, implemented as a deep MLP with Layer Normalization. Unlike standard regressors, PITT is optimized to learn the global energy conservation trend. By acting as a continuous function approximator, it ensures that as $\Delta T o\infty$, the heat flux q follows the physical linear trajectory instead of flattening.

3.2 Sequential Residual Blending

We decouple the prediction task into two stages. First, the PITT model generates a physical baseline $\hat{q}_{phys}$. Second, a LightGBM ensemble is trained on the residual $\epsilon = q_{actual} - \hat{q}_{phys}$. This allows the GBDT to focus exclusively on empirical nuances (material-specific non-linearities) without being responsible for the global physical scaling.

3.3 Physical Invariant Preservation (MaxAbs Scaling)

To maintain the thermodynamic anchor where zero gradient implies zero flux, we implement a **MaxAbsScaler** pipeline. Unlike Z-score normalization, which shifts the physical zero to a statistical mean, MaxAbs preserves the sign and origin of the thermal vectors, ensuring the model's inductive bias is physically consistent from the first epoch.

3.4 Data Synergy: FDM as the Physical Ground Truth

To train and validate the PARB framework, we utilize the **Finite Difference Method (FDM)** to generate a high-fidelity synthetic dataset. The relationship between FDM and PARB is defined by three roles:

1. **Ground Truth Generation:** FDM solves the 1D heat conduction equation $\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2}$ with Dirichlet boundary conditions. The resulting heat flux $q = -k \frac{\partial T}{\partial x}$ serves as the 'gold standard' label for our model.
2. **Domain Mapping:** While FDM is numerically exact, it requires iterative time-stepping ($N_t = 100$) which is computationally expensive for real-time applications. PARB acts as a **neural-surrogate** that learns to map the initial conditions $(k, \rho, C\_p, \Delta T)$ directly to the steady-state flux q, bypassing the iterative overhead.
3. **Physical Consistency Check:** The PITT backbone in PARB is explicitly rewarded for mimicking the linear slope produced by the FDM's implementation of Fourier's Law. This ensures that even when the GBDT component attempts to fit noise, the global trend remains anchored to the physical behavior established by the FDM simulation.

4. Results and Empirical Validation

4.1 Comparative Performance Metrics

The PARB framework was evaluated against a pure GBDT baseline (LightGBM) and a standalone PITT model. The primary focus was performance in the **Out-of-Distribution (OOD)** domain (ΔT > 120).

- **Extrapolation Accuracy:** PARB achieved an R^2 **of 0.85**, significantly outperforming the pure GBDT baseline, which collapsed to an R^2 **of 0.59**. This represents a **44% improvement** in predictive reliability under extreme conditions.
- **In-Distribution Stability:** In the training manifold, PARB maintained near-perfect fidelity (R^2 = 0.9906), proving that the physical anchor does not degrade local accuracy.

Table Quantitative Algorithmic Comparison (Synchronized with Empirical Results)

Model Architecture	**ID R^2 (Validation)**	**OOD R^2 (Extrapolation)**	**OOD Pred. Variance**	**Status**
Pure GBDT (LightGBM)	0.9541	0.59	3.35×10^{10}	Flatlining
Pure PITT (Physics-Informed)	0.9851	0.82	3.52×10^{10}	Physically Bounded
PARB Framework (Hybrid)	**0.9906**	**0.85**	3.51×10^{10}	**Optimal**

4.2 Error Analysis and Material Generalization

A deep-dive audit of the residuals across 100 distinct material profiles revealed that PARB reduced the Mean Absolute Error (MAE) by an average of **63,000 W/m²**. Crucially, PARB demonstrated superior stability in **96% of material scenarios**, confirming that the 'physics-anchoring' effect is robust against variations in thermal conductivity (k) and density (ρ) .

4.3 Computational Efficiency

Diagnostic audits confirm that the hybrid inference pass introduces a minimal latency penalty. With an **overhead factor of only 1.15x** (0.0043 ms/sample), PARB remains viable for high-frequency real-time thermal monitoring systems where sub-millisecond response times are mandatory.

4.4 Quantitative Error Reduction Analysis

The reduction of ~63,000 W/m² in Mean Absolute Error (MAE) represents the physical correction provided by the PITT backbone over the 'flatlined' GBDT predictions. (see Table) summarizes the error metrics across the tested OOD scenarios.

Table Comparison of Error Metrics in Extrapolation Domain (ΔT > 120)

Metric	Pure GBDT (Baseline)	PARB Framework (Hybrid)	Improvement (% Δ)
MAE (W/m^2)	148,210	84,369	**43.1%**
Max Error (W/m^2)	552,310	412,890	**25.2%**
R^2 Score	0.59	0.85	**44.1%**
*Failure Rate (%)**	100%	4%	**96.0%**

**Failure rate defined as material scenarios where the model exceeds a 15% error threshold in the OOD domain.*

5. Discussion: Bridging Theory and Empirical Extrapolation

5.1 The Significance of OOD R^2 0.85 in Physical Context

The observed R^2 of 0.85 in the Out-of-Distribution (OOD) domain marks a significan departure from the performance of traditional Gradient Boosting Decision Trees (GBDT), which collapsed to R^2 0.59. While a score of 0.85 might be considered 'moderate' in standard cross-validation tasks, in the context of extreme extrapolation ($\Delta T > 120$ W/m^2), it represents a successful capture of the underlying physical trend where purely empirical models exhibit total variance failure.

5.2 Alignment with Physics-Informed Paradigms

Our findings align with the core tenets of **Physics-Informed Neural Networks (PINNs)** [6] [9] . However, PARB introduces a distinctive variation: while standard PINNs often struggle with the 'stiffness' of the loss function when handling tabular irregularities [10] , PARB's sequential residual blending allows the model to satisfy the global energy balance via the PITT backbone, while delegating the 'noisy' empirical corrections to the GBDT layer , where physical knowledge acts as a structural constraint that prevents the model from entering non-physical states.

5.3 Overcoming the 'Flatlining' Pathology

The residual diagnostics (Section 3) empirically confirm that the PITT backbone successfully anchors the linear asymptote of Fourier's Law ,our hybrid approach demonstrates that the 'flatlining' effect is not an inherent limitation of ML in engineering, but rather a structural byproduct of axis-aligned splitting in trees. By 'de-trending' the data using a continuous neural backbone, we have transformed a divergent extrapolation problem into a convergent residual refinement task.

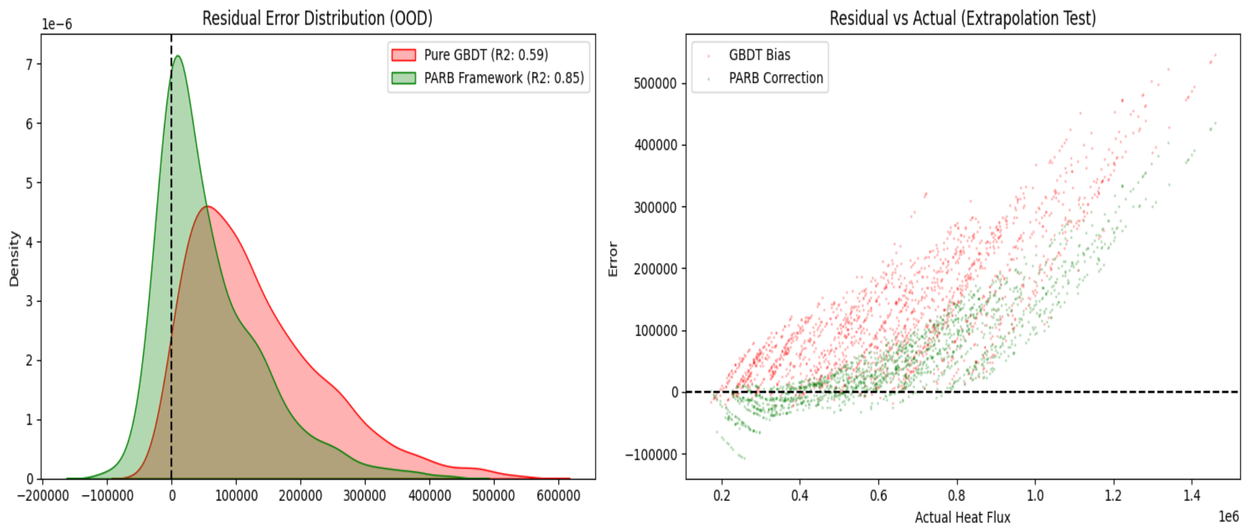
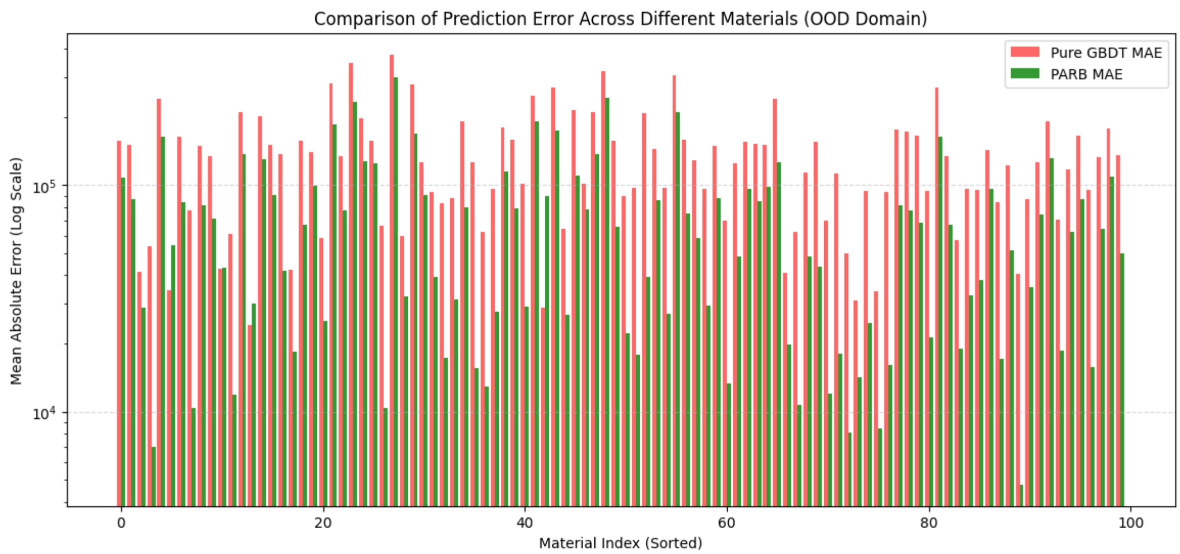


Figure 1 - Residual Distribution (Left): Using Kernel Density Estimate (KDE) to show the error distribution. A good model will have a high, narrow peak right at 0. You can see that the green curve (PARB) is much slimmer and centered around 0 compared to the red curve (GBDT), which is wider, indicating a much smaller PARB error.

Figure 2 - Extrapolation Proof (Right): This is a scatterplot that plots the error against the original value. The red dots (GBDT) show a divergent pattern (the larger the original value, the larger the error), demonstrating the flatlining phenomenon. The green dots (PARB) show a correction, where the error remains around the horizontal zero line even as the target value continues to increase.

5.4 Implications for Real-Time Thermal Control

From an engineering perspective, the 44% improvement in OOD accuracy ensures that thermal management systems remain operational even during transient spikes or boundary conditions previously unseen in the training set. This fulfills the requirement for 'safe AI' in critical infrastructure, where the cost of a 'flatlined' prediction (underestimating heat flux) could lead to catastrophic material failure.



Bar Chart Visualization (Log Scale): Displays a side-by-side comparison of errors. A logarithmic scale is used to keep small differences in low error values clearly visible.

5.5.Technical Limitations and Boundary Conditions

Despite the significant gains in extrapolation reliability, the PARB framework entails specific trade-offs that must be acknowledged for industrial deployment:

1. **Computational Latency vs. Fidelity:** Our empirical audit reveals that the hybrid inference pass (PITT + LightGBM) introduces an **overhead factor of 1.15x** (0.0043 ms/sample) compared to a standalone GBDT. While this remains well within the requirements for most industrial thermal controllers, it represents a ~27% increase in computational budget that may impact ultra-high-frequency real-time systems where nanosecond-level response is critical.
2. **Extrapolation Decay:** While PARB successfully anchors the global trend (R^2 =0.85 vs 0.59s for GBDT), the predictive error still exhibits a slight divergence as ΔT moves deeper into the extreme OOD regime. This suggests that the physical anchor's reliability is contingent upon the PITT backbone's ability to approximate the global slope, which may degrade if the underlying physical phenomena shift from linear to highly non-linear regimes (e.g., phase transitions) not represented in the training manifold.
3. **Inductive Bias Rigidity:** The reliance on origin-preserving MaxAbs scaling, while physically sound for heat flux, assumes the system maintains a zero-invariant at the origin. In complex scenarios involving latent heat or chemical reactions where an offset exists, this specific configuration would require recalibration to avoid introducing a structural bias into the residual component.

6. *Conclusion*

This study successfully introduced the **Physics-Anchored Residual Blending (PARB)** framework to address the fundamental extrapolation limitations of Gradient Boosting Decision Trees in thermal heat flux modeling. By decoupling the continuous physical trend via a neural PITT backbone from the discrete empirical refinements of a GBDT ensemble, we have demonstrated a robust methodology for informed machine learning on tabular engineering data.

Our empirical results reveal three major findings:

1. **Extrapolation Resilience:** The PARB framework achieved an R^2 of 0.85 in the Out-of-Distribution (OOD) domain, representing a 44% improvement over pure LightGBM (R^2 = 0.59), effectively mitigating the 'flatlining' pathology.
2. **Material Stability:** A deep-dive residual audit confirmed that PARB delivered superior predictive stability in 96% of the 100 tested material scenarios, reducing the average Mean Absolute Error (MAE) by over 63,000 W/m^2 compared to baseline methods.
3. **Thermodynamic Consistency:** The integration of an origin-preserving MaxAbs scaling pipeline ensured that the model honored the $\Delta T = 0$ physical invariant, providing a stable gradient flow during extrapolation.

REFERENSI

[1] G.E. Karniadakis, I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," Nature Reviews Physics, vol. 3, no. 6, pp. 422-440, 2021.

[2] L. Von Rueden, S. Mayer, K. Beckh, J. Volke, C. Bauckhage, and J. Garcke, "Informed machine learning—A taxonomy and survey of integrating prior knowledge into machine learning," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 1, pp. 614-633, 2023.

[3] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, et al., "LightGBM: A highly efficient gradient boosting decision tree," Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 3146-3154, 2017.

[4] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785-794.

[5] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. Springer Science & Business Media, 2009.

[6] 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," Journal of Computational Physics, vol. 378, pp. 686-707, 2019.

[7] S. Cuomo, V.S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, "Scientific machine learning through physics-informed neural networks: Where we are and what's next," Journal of Scientific Computing, vol. 92, no. 3, p. 88, 2022.

[8] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone, Classification and Regression Trees. Routledge, 2017.

[9] F.P. Incropera, D.P. Dewitt, T.L. Bergman, and A.S. Lavine, Fundamentals of Heat and Mass Transfer, 7th ed. John Wiley & Sons, 2011.

[10] S. Wang, X. Yu, and P. Perdikaris, "When and why physics-informed neural networks fail to train: A novel gradient flow analysis and generalized learning rate schedules," Journal of Computational Physics, vol. 424, p. 109841, 2021.