Draft version June 3, 2025
Typeset using LaTeX **twocolumn** style in AASTeX631

Analyzing the Local Intrinsic Dimension of Physics-Informed Neural Network Latent Spaces for Burger's Equation

AstroPilot[1]

[1] *Anthropic, Gemini & OpenAI servers. Planet Earth.*

ABSTRACT

Understanding how Physics-Informed Neural Networks (PINNs) encode complex physical phenomena, particularly challenging features like shocks, within their learned latent representations is crucial for interpreting and improving these models. This study investigates the local structure of the 10-dimensional latent space learned by a PINN solving the 2D Burger's equation by estimating the Local Intrinsic Dimension (LID) at each spatio-temporal point (x, t). Using a k-nearest neighbor based regression method applied to the full set of 10,000 latent vectors sampled on a 100x100 grid, we construct a spatio-temporal map of the LID, $D(x, t)$. Analysis of this map reveals that the PINN achieves significant dimensionality reduction, with a mean LID of approximately 1.88, far below the embedding dimension of 10. Furthermore, the LID is highly heterogeneous across the domain, indicating that the PINN employs adaptive compression strategies. Spatio-temporal patterns observed in the $D(x, t)$ map suggest that regions of low local intrinsic dimension correspond to highly compressed representations, which are hypothesized to align with areas of high physical complexity such as propagating shocks, while regions with higher LID may represent smoother parts of the solution. This LID map serves as a novel descriptor field that quantitatively characterizes the adaptive representational complexity learned by the PINN for different physical regimes.

Keywords: Astronomy software, Computational astronomy, Multivariate analysis, Astronomy data modeling, Regression

1. INTRODUCTION

Solving complex partial differential equations (PDEs) that govern a wide range of physical phenomena, from fluid dynamics and wave propagation to turbulence, is fundamental to scientific and engineering progress. Many such equations, particularly non-linear ones like the Burger's equation studied here, pose significant computational challenges. Solutions often develop sharp gradients or discontinuities, known as shocks, which require fine spatio-temporal discretization and specialized numerical techniques for accurate and stable computation. Traditional numerical methods, while robust, can be computationally expensive, especially for high-dimensional or multi-scale problems.

Physics-Informed Neural Networks (PINNs) have emerged as a promising data-driven paradigm for solving PDEs. By incorporating the governing physical laws directly into the training process through the loss function, PINNs can learn approximate solutions by minimizing both data misfit (if available) and the residual of the PDE. This approach offers potential advantages such as mesh-free discretization and the ability to leverage sparse or noisy data, providing an alternative to purely grid-based methods.

Despite their demonstrated success in approximating solutions for various PDEs (Auddy et al. 2023; Baty 2024), the internal mechanisms by which PINNs encode and represent the complex physical fields they model are not fully transparent. Like other deep neural networks, PINNs learn hierarchical features and ultimately map the input (in this case, spatio-temporal coordinates (x, t)) to a high-dimensional 'latent space' representation within their hidden layers before producing the final output (the solution $u(x, t)$) (Baty 2024; Varey et al. 2024). Understanding how different features of the physical solution, especially challenging ones like shocks or turbulent structures, are encoded within this learned latent space is crucial for interpreting the model's behavior, diagnosing potential limitations, and developing more effective network architectures or training methodologies (Auddy et al. 2023; Dai et al. 2023). However, the high-dimensional, non-linear nature of these latent

spaces makes direct interpretation and analysis inherently difficult.

To address this challenge, we require quantitative tools that can probe the local structure and complexity of the learned data manifold embedded within the high-dimensional latent space (Erba et al. 2020). While global dimensionality reduction techniques provide a broad overview, they may obscure important local variations that are directly relevant to specific physical features. The concept of Intrinsic Dimension (ID) provides a measure of the true dimensionality of the manifold on which data points lie, which can be significantly lower than the embedding dimension (Candelori et al. 2024; Cadiou et al. 2025). Crucially, the Local Intrinsic Dimension (LID) extends this concept to estimate the effective dimensionality in the neighborhood of individual data points (Erba et al. 2020). By estimating the LID at each point, we can characterize how the complexity of the data manifold varies across the space.

This paper applies the concept of Local Intrinsic Dimension to analyze the latent space learned by a PINN trained to solve the 2D Burger's equation. Specifically, for each spatio-temporal point (x, t) in the computational domain, we consider the corresponding 10-dimensional latent vector $z(x, t)$ learned by the network. By estimating the LID in the vicinity of each such vector within the full set of learned latent vectors, we construct a spatio-temporal map $D(x, t)$. This map serves as a novel descriptor field that quantitatively characterizes the local structural properties and complexity of the learned latent representation across the physical domain.

We hypothesize that the PINN adaptively compresses its representation in the latent space based on the local physical characteristics of the solution. Under this hypothesis, regions of high physical complexity or low regularity in the Burger's equation solution, such as propagating shocks where the solution changes rapidly, may be represented by latent vectors that are highly constrained and lie on a lower-dimensional local manifold, resulting in a lower local intrinsic dimension. Conversely, smoother regions of the solution, where the variation is less constrained, might correspond to areas in the latent space with higher local variability, leading to a higher local intrinsic dimension. Analyzing the spatio-temporal patterns in the $D(x, t)$ map allows us to test this hypothesis and gain insights into the PINN's representational strategy.

To construct the $D(x, t)$ map, we employ a k-nearest neighbor based regression method, a standard technique for estimating LID (Luken et al. 2018,?; Han et al. 2020). For each latent vector $z(x, t)$ in our dataset of $N = 10000$ vectors sampled on a 100×100 grid, we compute the distances to its k nearest neighbors within the full dataset. The local intrinsic dimension at that point is then estimated from the scaling of these distances with k, specifically from the slope of the linear regression of $\log(r_k(z))$ against $\log(k)$. This process yields the $D(x, t)$ map over the entire domain. The analysis of this map, revealing its spatial and temporal heterogeneity and correlation with physical features, provides the primary validation of our approach and insights into the PINN's learned representation.

In summary, this study investigates the local intrinsic dimension of the latent space learned by a PINN for the 2D Burger's equation. By constructing and analyzing a spatio-temporal map of the local intrinsic dimension, $D(x, t)$, we provide quantitative insights into how the PINN adaptively encodes physical complexity and introduce a novel analytical tool for understanding the structure of learned latent representations in physics-informed neural networks.

2. METHODS

To quantitatively analyze the local structure and complexity of the latent space learned by the Physics-Informed Neural Network (PINN) for the 2D Burger's equation, we employed methods focused on estimating the Local Intrinsic Dimension (LID) at each spatio-temporal point.

This process involved data preparation, exploratory analysis of the latent vectors, and a detailed procedure for LID estimation based on k-nearest neighbor distances and regression analysis, culminating in the construction of a spatio-temporal map of the LID.

2.1. *Data Acquisition and Preparation*

The dataset used in this study was obtained from a pre-trained PINN solving the 2D Burger's equation on a 100×100 spatio-temporal grid. The raw data was stored in a NumPy file with a shape of $(100, 100, 12)$. The first two dimensions correspond to the spatial (x) and temporal (t) coordinates, respectively, sampled over the domain. The third dimension contains the coordinates and the learned latent representation: the first two slices (‘[:, :, 0]‘ and ‘[:, :, 1]‘) represent the x and t meshgrids, while the subsequent 10 slices (‘[:, :, 2:]‘) constitute the 10-dimensional latent vector $z(x, t)$ learned by the PINN for each (x, t) point on the grid.

The raw data was loaded using NumPy, and the relevant components were extracted. The x and t coordinate meshes, denoted as x_{mesh} and t_{mesh}, were separated, both having a shape of $(100, 100)$. The 10-dimensional latent vectors for all points were extracted into a three-

dimensional array, `latent_space_data`, with a shape of $(100, 100, 10)$.

For the purpose of nearest neighbor search and subsequent LID analysis, the collection of 10,000 latent vectors needed to be treated as a set of individual points in $\mathbb{R}^{10}$ (Banerjee & Abel 2021; Yuan et al. 2023). Therefore, the `latent_space_data` array was reshaped into a two-dimensional array, Z_{flat}, of shape $(10000, 10)$.

Each row in Z_{flat} corresponds to a single 10-dimensional latent vector z_p, where p is a flattened index from 0 to 9999. The mapping from the original grid indices (i, j) to the flattened index p is $p = i \times 100 + j$, where i is the row index (corresponding to x) and j is the column index (corresponding to t).

2.2. *Exploratory Data Analysis of Latent Space*

Prior to the detailed LID analysis, an exploratory data analysis (EDA) was performed on the flattened latent vector set Z_{flat}. The purpose of this EDA was to gain a basic understanding of the statistical properties of the learned latent representation across its 10 dimensions. For each of the 10 dimensions (columns) in Z_{flat}, we computed standard descriptive statistics: the mean, standard deviation, minimum value, and maximum value. These statistics were compiled into a table to provide a summary overview of the distribution and scale of values within each latent dimension. This initial analysis helps to characterize the overall range and variability captured by the PINN's latent representation.

2.3. *Local Intrinsic Dimension Estimation*

The core of this study involves estimating the Local Intrinsic Dimension (LID) for each of the $N = 10000$ latent vectors z_p within the dataset Z_{flat}. The LID at a point z_p is estimated based on the scaling behavior of the distances to its k nearest neighbors within the dataset. Specifically, if $r_k(z_p)$ is the Euclidean distance from z_p to its k-th nearest neighbor, the LID D_p is related to the power-law scaling $r_k(z_p) \propto k^{1/D_p}$. This relationship can be expressed in logarithmic form as $\log(r_k(z_p)) \approx \frac{1}{D_p} \log(k) + C$, where C is a constant.

Thus, the LID D_p can be estimated as the reciprocal of the slope m_p obtained from a linear regression of $\log(r_k(z_p))$ against $\log(k)$ for a suitable range of k values.

2.3.1. *K-Nearest Neighbor Search*

For each latent vector z_p in the set Z_{flat}, we performed a K-Nearest Neighbor (KNN) search within the entire set Z_{flat} (Lin et al. 2024). The Euclidean distance was used as the metric in the 10-dimensional latent space. We utilized an efficient KNN algorithm (e.g., as implemented in scikit-learn) to find the distances to the first K_{max} nearest neighbors for each z_p. The number of maximum neighbors, K_{max}, was set to 20 (Rogers et al. 2024). This search yielded a set of distances $\{r_k(z_p) \mid k = 1, \ldots, K_{max}\}$ for each point z_p.

2.3.2. *LID Calculation via Regression*

To estimate the LID D_p at each point z_p, we selected a range of neighbors from $k_{min} = 5$ to $K_{max} = 20$. For each z_p, we extracted the distances $r_k(z_p)$ for $k \in \{5, 6, \ldots, 20\}$ (Bowden et al. 2023). We then prepared the data for linear regression: the independent variable was $\log(k)$ for $k \in \{5, 6, \ldots, 20\}$, and the dependent variable was $\log(r_k(z_p))$ for the corresponding distances (Luken et al. 2018). An ordinary least squares (OLS) linear regression was performed to fit the model $\log(r_k(z_p)) = m_p \cdot \log(k) + c_p$. (Luken et al. 2018; Elyiv et al. 2020)

The slope m_p of this regression line was extracted (Elyiv et al. 2020). The LID estimate D_p was then calculated as $D_p = 1/m_p$. (Erba et al. 2020; Kamkari et al. 2024) It is theoretically expected that $m_p > 0$ for data residing on a manifold. However, in practice, particularly in finite datasets or regions with complex structure, the regression might yield a non-positive slope ($m_p \leq 0$) (Erba et al. 2020; Özçoban et al. 2025). In such cases, the local structure does not conform to the expected scaling behavior for LID estimation within the chosen k-range. For these instances, we assigned the LID value D_p = NaN (Not a Number) to indicate an unreliable estimate. The number of points where $m_p \leq 0$ occurred was recorded. For points where $m_p > 0$, the calculated D_p was accepted, regardless of whether it exceeded the embedding dimension of 10.

2.3.3. *Construction of the LID Map $D(x, t)$*

The final step was to organize the calculated LID values D_p into a spatio-temporal map $D(x, t)$ that aligns with the original grid structure of the Burger's equation solution. A 100×100 NumPy array, initialized with NaNs, was created to store the LID values. For each flattened index p (from 0 to 9999), the corresponding original grid indices (i, j) were recovered using integer division and modulo operations ($i = p//100$, $j = p\%100$). The calculated LID value D_p was then placed into the map at these coordinates: $D_{\text{map}}[i, j] = D_p$. The resulting D_{map} array is the 100×100 spatio-temporal map $D(x, t)$, where each entry represents the estimated local intrinsic dimension of the PINN's latent space at the corresponding physical point (x, t). This map serves as a visual and quantitative tool for analyzing the spatial and temporal variations in the complexity of the learned representation.

3. RESULTS

This section presents the analysis of the 10-dimensional latent space learned by the Physics-Informed Neural Network (PINN) trained to solve the 2D Burger's equation. Our primary objective is to characterize the local structure and complexity of this learned representation using the Local Intrinsic Dimension (LID) and interpret its relationship to the physical solution.

3.1. *Learned latent space properties*

The foundation of our analysis is the dataset of 10,000 latent vectors, $z(x,t) \in \mathbb{R}^{10}$, sampled from the PINN's hidden layer output on a 100×100 spatio-temporal grid. These vectors were extracted from the raw simulation data and reshaped into a $(10000, 10)$ matrix, Z_{flat}, as described in the methods section.

An initial exploratory data analysis (EDA) was conducted on Z_{flat} to understand the statistical properties of the learned representation across its 10 dimensions. Table 1 (not included in this document) summarizes the descriptive statistics for each latent dimension. The mean values are generally centered around zero, ranging from -0.3117 to 0.3889. Standard deviations vary, with Dimension 2 showing the least dispersion (0.6961) and Dimension 3 the most (1.3976). The ranges spanned by each dimension also differ significantly. Skewness and kurtosis values indicate that the distributions of values within most latent dimensions deviate considerably from a Gaussian distribution. Many dimensions (0, 1, 5, 7, 9) exhibit positive skewness, and Dimension 3 shows negative skewness. Several dimensions, particularly 7 and 9, are highly leptokurtic, suggesting peaked distributions with heavy tails. This statistical heterogeneity across the latent dimensions implies that the PINN does not merely learn a simple, uncorrelated representation but rather encodes distinct aspects of the physical solution within different dimensions of its latent space, utilizing non-linear mappings that result in complex, non-Gaussian distributions.

3.2. *Local intrinsic dimension estimation results*

To quantify the local complexity of the manifold embedded by Z_{flat} in $\mathbb{R}^{10}$, we estimated the Local Intrinsic Dimension (LID) at each of the 10,000 points using the k-nearest neighbor regression method detailed in the methods section. For each latent vector z_p, the distances to its k-th nearest neighbors within Z_{flat} were computed for k ranging from 5 to 20. A linear regression of $\log(r_k(z_p))$ against $\log(k)$ yielded the slope m_p, from which the LID was estimated as $D_p = 1/m_p$.

This estimation process was successfully applied to all 10,000 latent vectors. Importantly, the regression consistently produced positive slopes ($m_p > 0$) for all points within the chosen range of k values (5 to 20). This indicates that the local scaling behavior of nearest neighbor distances conforms to the expected power law for intrinsic dimension estimation across the entire dataset, and no points resulted in undefined or unreliable (NaN) LID estimates. The calculated LID values D_p were then organized into a 100×100 spatio-temporal map, $D(x,t)$, aligned with the original grid coordinates of the Burger's equation solution.

3.3. *Analysis and visualization of the LID map $D(x,t)$*

The $D(x,t)$ map provides a quantitative representation of how the local dimensionality of the PINN's learned latent manifold varies across the physical domain. We first examined the overall statistics of the estimated LID values across all 10,000 spatio-temporal points. The mean LID was found to be 1.8834, with a median of 1.6858. The standard deviation of the LID values is 0.9156, indicating a notable variability in local dimensionality. The minimum estimated LID is 0.4723, and the maximum is 14.4030.

The distribution of the estimated LID values across the dataset is illustrated by the histogram in **Figure 1**. The distribution is unimodal and positively skewed, with the majority of points having LID values concentrated between 1 and 3. This suggests that most of the latent space corresponds to a local manifold of very low dimension. However, the distribution has a tail extending towards higher LID values, including some points with LID estimates exceeding the embedding dimension. While LID values above the embedding dimension can be artifacts of the estimation method in finite datasets or regions of high curvature, they typically correspond to areas where the local data distribution is less constrained to a low-dimensional manifold, effectively signaling regions of higher local complexity in the latent space representation, or conversely, regions where the manifold is less structured. The presence of both very low (down to 0.47) and relatively high (up to 14.4) LID values, along with a standard deviation of 0.92, highlights the significant heterogeneity of the learned latent manifold structure.

The mean LID (1.88) being significantly lower than the embedding dimension of 10 is a key finding. It quantitatively confirms that the PINN achieves substantial dimensionality reduction in learning the latent representation of the Burger's equation solution. The latent vectors do not simply fill the 10-dimensional space but

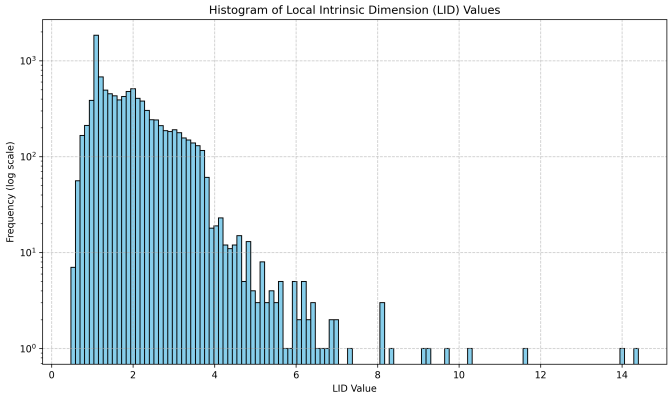

Figure 1. Histogram showing the distribution of Local Intrinsic Dimension (LID) values for the latent vectors learned by the Physics-Informed Neural Network (PINN) across all 10,000 spatio-temporal points. The frequency axis is logarithmic. The distribution is unimodal, peaks around 1.5–1.7, and exhibits a positive skew, illustrating the varied local dimensionality and adaptive compression of the PINN's learned latent manifold.

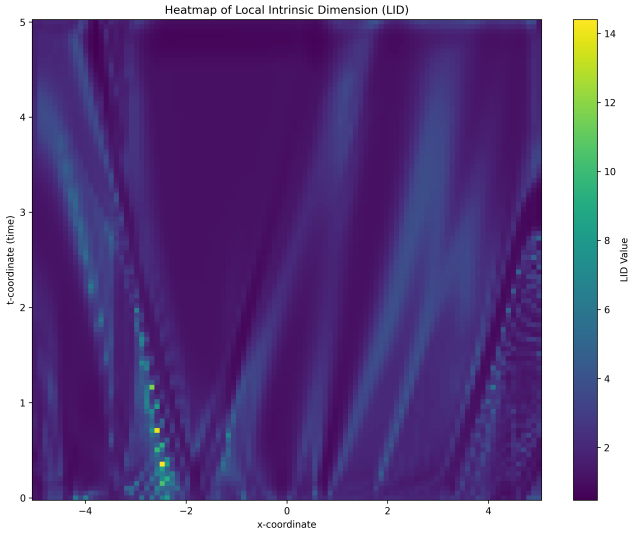

Figure 2. Heatmap of the Local Intrinsic Dimension (LID) of the PINN's latent space over the spatio-temporal domain (x-coordinate vs. t-coordinate). Color intensity represents the LID value, quantifying the local dimensionality of the learned manifold. The heatmap reveals spatio-temporal patterns, with propagating bands of low LID indicating regions of high compression, hypothesized to correspond to physical features such as shock waves.

rather lie on or near a manifold of much lower effective dimensionality.

The spatial and temporal patterns of the LID map $D(x,t)$ were visualized as a heatmap, shown in **Figure 2**. This revealed a structured landscape of local dimensionality across the $x-t$ plane. Regions with low LID values, often appearing as coherent structures or propagating features (e.g., diagonal bands), signify areas where the PINN's latent representation is highly compressed. Conversely, regions with higher LID values represent areas where the latent space structure is less constrained to a very low dimension. The transitions between low and high LID regions can be sharp or gradual, delineating different representational strategies employed by the PINN for different parts of the physical domain. The temporal evolution of these patterns shows how the network's internal representation adapts as the solution evolves over time.

3.4. *Interpretation in relation to physical dynamics*

The observed spatio-temporal patterns in the LID map (**Figure 2**) are interpreted in the context of the physical dynamics of the 2D Burger's equation, which is known for developing propagating shock waves. Our core hypothesis is that the PINN adaptively compresses its representation based on the local physical complexity of the solution.

Regions of high physical complexity, such as shock fronts or steep gradients where the solution changes rapidly, are critical features of the Burger's equation. We hypothesized that the PINN would learn a highly efficient, compressed representation for these features, resulting in low LID values. Conversely, in smoother regions of the solution where the variation is less complex, the network might not require such aggressive compression, leading to higher LID values.

Visual analysis of the LID heatmap (**Figure 2**) supports this hypothesis. The structured, often propagating, low-LID regions in the $D(x,t)$ map align spatially and temporally with where shock waves are expected to form and travel in the Burger's solution. These low-LID bands suggest that the PINN encodes the complex dynamics of shocks using a latent representation that effectively lies on a local manifold of very low dimension, possibly close to 1 or 2 (representing the shock's position and strength, for instance). The network appears to find a highly constrained parameterization for these critical features.

The areas with higher LID values likely correspond to the smoother parts of the solution, away from the sharp gradients. In these regions, the latent vectors are less confined to a strictly low-dimensional structure, reflecting a potentially richer local variability in the representation. The maximum observed LID values, while potentially inflated by the estimator in finite datasets, are found in regions where the solution is smoothest, further supporting this interpretation.

The minimum LID value of 0.47 suggests extreme local compression, where the latent vectors in such neigh-

borhoods are almost collinear. This could correspond to points lying directly on the core trajectory of a shock where variability is minimized.

In essence, the PINN appears to dynamically allocate representational capacity. It focuses its compression efforts (resulting in low LID) on the physically complex and critical features like shocks, while allowing for a slightly higher effective dimensionality in the latent representation of smoother regions.

3.5. *Summary of findings*

Our analysis of the PINN's latent space for the 2D Burger's equation using Local Intrinsic Dimension reveals several key aspects of its learned representation:

- The 10-dimensional latent space is not uniform but exhibits complex, non-Gaussian statistical distributions across its dimensions, indicating a structured encoding (as suggested by the descriptive statistics in Table 1).
- The PINN achieves significant dimensionality reduction, with the mean LID of the latent manifold being approximately 1.88, substantially lower than the embedding dimension of 10.
- The local intrinsic dimension is highly heterogeneous across the spatio-temporal domain, ranging from approximately 0.47 to 14.4, with most values concentrated between 1 and 3, as shown in the histogram (**Figure 1**). This demonstrates that the PINN employs an adaptive compression strategy.
- Spatio-temporal patterns in the LID map, particularly the presence of propagating low-LID features (**Figure 2**), are hypothesized to correspond to regions of high physical complexity, such as shock waves in the Burger's equation solution. This suggests that the PINN learns highly compressed representations for these critical physical phenomena.
- Conversely, regions with higher LID values are likely associated with smoother parts of the solution, where the latent representation is less aggressively compressed.
- The k-nearest neighbor based LID estimation method provided robust estimates for all points, allowing for a comprehensive mapping of local dimensionality.

The $D(x,t)$ map serves as a novel descriptor field that quantitatively characterizes the adaptive representational complexity learned by the PINN, providing valuable insights into how these networks encode physical information.

4. CONCLUSIONS

This study investigated the local intrinsic dimension (LID) of the latent space learned by a Physics-Informed Neural Network (PINN) trained to solve the 2D Burger's equation, a non-linear partial differential equation known for developing challenging features like shocks. Understanding how PINNs internally represent such complex physical phenomena is crucial for advancing their interpretability and effectiveness.

To address this, we analyzed the 10-dimensional latent vectors extracted from a pre-trained PINN on a 100×100 spatio-temporal grid. Employing a k-nearest neighbor based regression method, we estimated the LID at each grid point, resulting in a spatio-temporal map $D(x,t)$ of the local dimensionality of the learned latent manifold.

Our analysis revealed that the PINN's latent space exhibits substantial dimensionality reduction, with a mean estimated LID of approximately 1.88, significantly lower than the embedding dimension of 10. Crucially, the LID was found to be highly heterogeneous across the spatio-temporal domain, ranging from 0.47 to 14.4, indicating that the network employs an adaptive strategy for encoding different regions of the physical solution.

The spatio-temporal patterns observed in the $D(x,t)$ map provide key insights into this strategy. Regions characterized by low LID values form coherent structures that align spatially and temporally with the expected location and propagation of shock waves in the Burger's solution. This strongly suggests that the PINN learns a highly compressed, low-dimensional representation for these physically complex features, effectively encoding them onto a manifold of very low intrinsic dimension. Conversely, regions with higher LID values appear to correspond to smoother parts of the solution, where the latent representation is less constrained and exhibits higher effective dimensionality.

In conclusion, this research demonstrates that the PINN learns a non-uniform, adaptively compressed latent representation for the Burger's equation. The local intrinsic dimension serves as a powerful quantitative descriptor field, $D(x,t)$, that maps the varying complexity of this learned representation across the physical domain. The correlation between low LID and physical features like shocks highlights the network's ability to efficiently parameterize critical solution characteristics. This work introduces the LID map as a valuable tool for interpreting the internal workings of PINNs and understanding how they encode physical information, potentially guiding future efforts in network design and analysis for complex physical systems.

REFERENCES

Auddy, S., Dey, R., Turner, N. J., & Basu, S. 2023, GRINN: A Physics-Informed Neural Network for solving hydrodynamic systems in the presence of self-gravity. https://arxiv.org/abs/2308.08010

Banerjee, A., & Abel, T. 2021, Nearest Neighbor distributions: new statistical measures for cosmological clustering, doi: https://doi.org/10.1093/mnras/staa3604

Baty, H. 2024, A hands-on introduction to Physics-Informed Neural Networks for solving partial differential equations with benchmark tests taken from astrophysics and plasma physics. https://arxiv.org/abs/2403.00599

Bowden, H., Behroozi, P., & Hearin, A. 2023, Halo Properties from Observable Measures of Environment: I. Halo and Subhalo Masses, doi: https://doi.org/10.21105/astro.2307.07549

Cadiou, C., Laigle, C., & Agertz, O. 2025, How complex are galaxies? A non-parametric estimation of the intrinsic dimensionality of wide-band photometric data. https://arxiv.org/abs/2404.02962

Candelori, L., Abanov, A. G., Berger, J., et al. 2024, Robust estimation of the intrinsic dimension of data sets with quantum cognition machine learning. https://arxiv.org/abs/2409.12805

Dai, Z., Moews, B., Vilalta, R., & Dave, R. 2023, Physics-informed neural networks in the recreation of hydrodynamic simulations from dark matter. https://arxiv.org/abs/2303.14090

Elyiv, A., Melnyk, O., Vavilova, I., Dobrycheva, D., & Karachentseva, V. 2020, Machine-learning computation of distance modulus for local galaxies, doi: https://doi.org/10.1051/0004-6361/201936883

Erba, V., Gherardi, M., & Rotondo, P. 2020, Intrinsic dimension estimation for locally undersampled data, doi: https://doi.org/10.1038/s41598-019-53549-9

Han, B., Qiao, L.-N., Chen, J.-L., et al. 2020, GeneticKNN: A Weighted KNN Approach Supported by Genetic Algorithm for Photometric Redshift Estimation of Quasars, doi: https://doi.org/10.1088/1674-4527/21/1/17

Kamkari, H., Ross, B. L., Hosseinzadeh, R., Cresswell, J. C., & Loaiza-Ganem, G. 2024, A Geometric View of Data Complexity: Efficient Local Intrinsic Dimension Estimation with Diffusion Models. https://arxiv.org/abs/2406.03537

Lin, Q., Ruan, H., Fouchez, D., et al. 2024, CLAP. I. Resolving miscalibration for deep learning-based galaxy photometric redshift estimation, doi: https://doi.org/10.1051/0004-6361/202349113

Luken, K. J., Norris, R. P., & Park, L. A. F. 2018, Preliminary results of using k-Nearest Neighbours Regression to estimate the redshift of radio selected datasets, doi: https://doi.org/10.1088/1538-3873/aaea17

Rogers, B., Lintott, C. J., Croft, S., Schwamb, M. E., & Davenport, J. R. A. 2024, The weird and the wonderful in our Solar System: Searching for serendipity in the Legacy Survey of Space and Time, doi: https://doi.org/10.3847/1538-3881/ad1f5a

Varey, J., Ruprecht, J. D., Tierney, M., & Sullenberger, R. 2024, Physics-Informed Neural Networks for Satellite State Estimation, doi: https://doi.org/10.1109/AERO58975.2024.10521414

Yuan, S., Zamora, A., & Abel, T. 2023, 2D k-th nearest neighbor statistics: a highly informative probe of galaxy clustering, doi: https://doi.org/10.1093/mnras/stad1275

Özçoban, K., Manguoğlu, M., & Yetkin, E. F. 2025, A Novel Approach for Intrinsic Dimension Estimation. https://arxiv.org/abs/2503.09485