Draft version June 4, 2025
Typeset using LaTeX **twocolumn** style in AASTeX631

Quantifying the Evolution of Learned Feature Structure in PINN Latent Space for 2D Burger's Equation via Principal Component Analysis

AstroPilot[1]

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

ABSTRACT

Understanding how Physics-Informed Neural Networks (PINNs) encode complex physical phenomena in their latent spaces is crucial for interpreting their learned representations. This study investigates the statistical structure of the 10-dimensional latent space learned by a PINN for the 2D Burger's equation across 25 viscosity values, a parameter controlling the transition from turbulent-like to diffusive regimes. We applied Principal Component Analysis (PCA) to standardized latent vectors extracted for each viscosity, analyzing the evolution of the eigenvalue spectrum and eigenvector structure. Our analysis quantified how the distribution of variance across latent dimensions changes with viscosity, tracking eigenvalue magnitudes, spectrum concentration (normalized entropy), and effective dimensionality based on variance explained. We also assessed the stability of the dominant principal component directions using cosine similarity. Our results show that as viscosity increases, the variance captured by the leading principal component decreases, and variance becomes more evenly distributed across latent dimensions (increasing spectrum entropy). The PCA-based effective dimensionality exhibits a non-monotonic trend, peaking at intermediate viscosities, which qualitatively aligns with previous intrinsic dimensionality findings. While the primary direction of variation (PC1) shows relative stability across low-to-intermediate viscosities, it undergoes significant rotation at high viscosities, and secondary directions (PC2, PC3) are less stable, particularly when eigenvalues are close. These quantitative findings provide evidence that the PINN adapts its internal latent space structure to the underlying physics. The observed evolution, including changes in variance distribution, non-monotonic complexity, and PC stability, offers insights into how the network implicitly captures physical transitions and potentially reflects principles analogous to coarse-graining as the system simplifies in the diffusion-dominated regime.

Keywords: Multivariate analysis, Principal component analysis, Astronomy data modeling, Computational methods, Algorithms

1. INTRODUCTION

Physics-Informed Neural Networks (PINNs) represent a significant advancement in scientific machine learning, offering a powerful framework for solving partial differential equations (PDEs) and inferring physical laws by embedding governing equations directly into the neural network's loss function. This mesh-free approach provides flexibility for handling complex geometries and high-dimensional problems. Beyond their role as solvers, PINNs also act as sophisticated non-linear function approximators, learning intricate mappings from input parameters, such as spatial and temporal coordinates, to physical quantities like velocity or pressure. A critical, yet often opaque, aspect of these networks is their internal representation space, commonly referred to as the latent space. This is where the network implicitly encodes abstract features of the physical system and its dynamics. Gaining a quantitative understanding of the structure and evolution of this learned latent space is paramount for interpreting what the network has captured about the underlying physics, enhancing trust in its predictions, and enabling reliable use cases like extrapolation and uncertainty quantification.

Interpreting the internal representations learned by deep neural networks remains a formidable challenge, particularly within the context of complex physical systems. The latent space typically results from a high-dimensional, non-linear transformation of the input data, and its dimensions do not generally correspond directly to intuitively understandable physical variables. The core difficulty lies in devising methods to

extract meaningful, quantitative insights into how the network organizes information about the physical state and its evolution within this abstract, learned space. This challenge becomes particularly pronounced when dealing with physical systems governed by parameters that drive transitions between distinct physical regimes. Analyzing how the latent space structure adapts in response to such parametric changes offers a valuable window into the network's internal encoding strategies.

This study focuses on the 2D Burger's equation, a well-known model in fluid dynamics that exhibits rich dynamics influenced by its viscosity parameter, ν. At low viscosities, the equation can develop sharp gradients resembling shocks or turbulent-like structures, while at high viscosities, diffusive effects dominate, leading to smooth profiles. We investigate how a PINN, trained to solve this equation across a range of 25 different viscosity values, encodes this physical transition within its 10-dimensional latent space. Our primary objective is to quantitatively analyze the statistical structure of this latent space for each specific viscosity value independently.

To achieve this, we employ Principal Component Analysis (PCA). For each viscosity setting, we extract a collection of 10-dimensional latent vectors corresponding to various spatial and temporal points computed by the trained PINN. These vectors are then standardized to have zero mean and unit variance for each dimension. We then compute the covariance matrix of the standardized latent vectors, and perform its eigendecomposition. PCA provides a linear dimensionality reduction technique that decomposes the variance of the data into a set of orthogonal components, ordered by the amount of variance they explain. The eigenvalues resulting from the eigendecomposition represent the variance captured by each principal component, and their distribution, known as the eigenvalue spectrum, reveals how the total variance is spread across the learned features. The corresponding eigenvectors define the directions of these principal components within the 10-dimensional latent space.

By analyzing the evolution of the PCA decomposition across the range of viscosity values, we aim to quantify how the structure of the learned latent representation changes as the underlying physics transitions. Specifically, we track how the eigenvalue spectrum changes, examining metrics such as the magnitudes of the leading eigenvalues, the concentration of variance across components (e.g., using normalized entropy), and the PCA-based effective dimensionality defined by the number of components required to explain a certain percentage of the total variance. Furthermore, we assess the stability of the principal component directions by quantifying the alignment (e.g., using cosine similarity) of the eigenvectors at different viscosities relative to a reference state.

The quantitative changes observed in these PCA metrics across the viscosity range serve as our evidence for how the PINN's latent space adapts to the changing physical regime governed by viscosity. We hypothesize that the statistical structure of the learned features, as revealed by the PCA spectrum and eigenvector stability, will reflect the known physical transitions in the Burger's equation. For instance, a physically more complex regime might be associated with a more distributed variance across latent dimensions or a higher effective dimensionality, while a simpler, diffusion-dominated regime might show variance concentrating along fewer principal components. By analyzing the trends in variance distribution, effective dimensionality, and the stability of principal component directions, we gain insights into how the network implicitly captures the changing dynamics. These findings provide a novel, quantitative perspective on the PINN's learned representations, complementing previous studies on latent space intrinsic dimensionality and offering potential analogies to physical principles like coarse-graining as the system simplifies.

2. METHODS

2.1. *Dataset*

The data analyzed in this study consists of learned latent space representations generated by a Physics-Informed Neural Network (PINN) trained to solve the 2D Burger's equation across a range of viscosity values. The dataset is provided as a single NumPy array with dimensions (N_x, N_t, N_ν, D), where $N_x = 101$ is the number of spatial points, $N_t = 103$ is the number of temporal points, $N_\nu = 25$ is the number of distinct viscosity values, and $D = 13$ is the total number of features per point. The first three features ($D = 0, 1, 2$) correspond to the spatial coordinate (x), time (t), and viscosity (ν), respectively, forming the input mesh information. The remaining ten features ($D = 3$ through 12) constitute the 10-dimensional latent space vector produced by an intermediate layer of the trained PINN for the corresponding (x, t, ν) input.

The 25 unique viscosity values are pre-defined in the dataset and span a range from a minimum value characteristic of turbulent-like behavior to a maximum value where diffusive effects dominate. For each of these 25 viscosity values, the latent space data consists of $N_x \times N_t = 101 \times 103 = 10403$ distinct 10-dimensional vectors, representing the network's internal encoding at

a grid of spatial and temporal points for that specific physical regime.

2.2. *Data Preparation for Per-Viscosity Analysis*

To perform a structured analysis of the latent space evolution, the raw data was processed and organized. The full dataset array was first separated into mesh information and latent space vectors. For each of the 25 viscosity values, the corresponding slice of the latent space data was extracted. This slice, initially of shape (N_x, N_t, D_{latent}), where $D_{latent} = 10$, was then reshaped into a 2D matrix of shape $(N_{samples}, D_{latent})$, where $N_{samples} = N_x \times N_t = 10403$. This resulted in 25 separate matrices, each representing the collection of 10-dimensional latent vectors for a specific viscosity value. Let X_k denote the $N_{samples} \times D_{latent}$ matrix for the k-th viscosity value, where $k \in \{0, 1, \ldots, 24\}$.

2.3. *Principal Component Analysis*

Principal Component Analysis (PCA) was applied independently to the latent space data matrix X_k for each of the 25 viscosity values. PCA is a linear dimensionality reduction technique that identifies orthogonal directions (principal components) in the data that capture the maximum variance.

2.3.1. *Standardization*

Prior to PCA, each $N_{samples} \times D_{latent}$ matrix X_k was standardized. Standardization involves scaling the data such that each of the D_{latent} features (columns) has a mean of zero and a standard deviation of one. This step is crucial because PCA is sensitive to the scale of the features, and standardizing ensures that features with larger variances do not disproportionately influence the principal components. The standardized matrix $\tilde{X}_k$ was computed as:

$$\tilde{X}_k = (X_k - \mu_k)/\sigma_k$$

where μ_k is the vector of means for each of the 10 latent features in X_k, and σ_k is the vector of standard deviations for each feature. Element-wise division is implied.

2.3.2. *Covariance Matrix Calculation and Eigendecomposition*

For each standardized matrix $\tilde{X}_k$, the $D_{latent} \times D_{latent}$ covariance matrix C_k was calculated. The covariance matrix captures the variance of each feature and the covariance between pairs of features. Given $\tilde{X}_k$ with $N_{samples}$ rows (observations) and D_{latent} columns (features), the covariance matrix is given by:

$$C_k = \frac{1}{N_{samples} - 1} \tilde{X}_k^T \tilde{X}_k$$

Eigendecomposition was then performed on the covariance matrix C_k. This decomposition yields a set of D_{latent} eigenvalues $\lambda_k = \{\lambda_{k,1}, \lambda_{k,2}, \ldots, \lambda_{k,10}\}$ and a corresponding set of D_{latent} eigenvectors $V_k = \{v_{k,1}, v_{k,2}, \ldots, v_{k,10}\}$. The eigenvectors represent the directions of the principal components in the 10-dimensional latent space, and the eigenvalues represent the amount of variance in the data that lies along each corresponding eigenvector direction. The eigenvalues and their corresponding eigenvectors were sorted in descending order of eigenvalue magnitude, such that $\lambda_{k,1} \geq \lambda_{k,2} \geq \cdots \geq \lambda_{k,10}$. The sorted eigenvalues and eigenvectors for each viscosity were stored for subsequent analysis.

2.4. *Analysis of Eigenvalue Spectrum Evolution*

The sorted eigenvalues obtained from the PCA of each per-viscosity latent space were analyzed to quantify how the distribution of variance across the latent dimensions changes with viscosity.

2.4.1. *Variance Explained*

For each viscosity k, the proportion of total variance explained by the j-th principal component was calculated as $\text{variance explained}_j = \lambda_{k,j} / \sum_{i=1}^{10} \lambda_{k,i}$. The cumulative variance explained by the first p principal components was calculated as $\sum_{j=1}^{p} \text{variance explained}_j$. The magnitudes of the leading eigenvalues $(\lambda_{k,1}, \lambda_{k,2}, \lambda_{k,3})$ and the cumulative variance explained were tracked as functions of the viscosity value ν_k.

2.4.2. *Effective Dimensionality*

A PCA-based effective dimensionality was determined for each viscosity. This metric is defined as the minimum number of principal components required to explain a certain cumulative percentage of the total variance. Common thresholds (e.g., 90%, 95%, 99%) were used to observe how the effective dimensionality, according to this criterion, changes across the viscosity range.

2.4.3. *Spectrum Concentration*

To quantify the concentration or flatness of the eigenvalue spectrum, the normalized Shannon entropy of the variance explained ratios was calculated. Let $p_{k,j} = \lambda_{k,j} / \sum_{i=1}^{10} \lambda_{k,i}$ be the proportion of variance explained by the j-th principal component for viscosity k. The Shannon entropy H_k is given by $H_k = -\sum_{j=1}^{10} p_{k,j} \log_2(p_{k,j})$. This entropy was then normalized by the maximum possible entropy for a 10-dimensional space, which occurs when variance is evenly distributed ($\lambda_{k,j} = \text{constant for all } j$), $H_{max} = \log_2(10)$.

The normalized entropy $H_{norm,k} = H_k/H_{max}$ provides a measure of spectrum concentration, ranging from 0 (variance concentrated in one dimension) to 1 (variance equally distributed across all dimensions). The evolution of $H_{norm,k}$ with viscosity was analyzed.

2.5. *Analysis of Principal Component Stability*

The stability of the directions of the principal components across different viscosity values was assessed by quantifying the alignment of the corresponding eigenvectors.

2.5.1. *Reference Eigenvectors*

The set of sorted eigenvectors obtained for the lowest viscosity value, ν_{min}, was chosen as the reference set, denoted as $V_{ref} = \{v_{min,1}, v_{min,2}, \ldots, v_{min,10}\}$.

2.5.2. *Cosine Similarity*

For each viscosity ν_k ($k > 0$) and its set of sorted eigenvectors $V_k = \{v_{k,1}, v_{k,2}, \ldots, v_{k,10}\}$, the similarity between the j-th eigenvector $v_{k,j}$ and the corresponding j-th reference eigenvector $v_{min,j}$ was measured using the absolute cosine similarity:

$$S_j(\nu_k) = |v_{k,j} \cdot v_{min,j}|$$

The absolute value is taken because eigenvectors represent directions, and a vector v and $-v$ represent the same line. The cosine similarity ranges from 0 (orthogonal directions) to 1 (perfectly aligned directions). This metric was computed for the first few dominant principal components (e.g., $j = 1, 2, 3$) across all viscosity values to analyze their directional stability relative to the state at minimum viscosity.

3. RESULTS

This study quantitatively investigates the structure and evolution of the 10-dimensional latent space learned by a Physics-Informed Neural Network (PINN) for the 2D Burger's equation. Employing Principal Component Analysis (PCA) on standardized latent vectors extracted for 25 distinct viscosity values (ν), ranging from 0.01 to 1.0, we analyze how the learned representation adapts to changes in the governing physical parameter. For each viscosity, the analysis is based on 10,403 latent space samples, corresponding to a grid of spatial and temporal points. Prior to PCA, the latent features were standardized per viscosity to ensure that the analysis is not biased by differences in feature scales or means, as motivated by observed variations in pre-standardization statistics across features and viscosities.

3.1. *Evolution of the eigenvalue spectrum with viscosity*

The PCA applied to the 10-dimensional standardized latent space for each viscosity ν yields a spectrum of 10 eigenvalues, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{10}$, representing the variance captured by each corresponding principal component (PC). The evolution of these eigenvalues with viscosity provides insight into how the distribution of variance across the latent dimensions changes, as shown in Figure 1.

As depicted in Figure 1, the magnitudes of the three largest eigenvalues ($\lambda_1, \lambda_2, \lambda_3$) exhibit distinct trends as viscosity increases. The dominant eigenvalue, λ_1, which represents the direction of maximum variance in the latent space, shows a clear decreasing trend with increasing viscosity. For instance, at the lowest viscosity ($\nu = 0.01$), $\lambda_1 \approx 7.10$. This value drops to ≈ 5.50 at $\nu = 0.1$ and further to ≈ 3.93 at the highest viscosity ($\nu = 1.0$). This decrease in the magnitude of the leading eigenvalue indicates that the primary mode of variation in the learned latent representation becomes less pronounced as the system transitions towards the diffusion-dominated regime.

In contrast, the second (λ_2) and third (λ_3) eigenvalues display more complex, non-monotonic behavior (Figure 1). λ_2 initially increases from ≈ 1.58 at $\nu = 0.01$, reaching a peak around intermediate viscosities ($\nu \approx 0.08 - 0.15$), with a value of ≈ 2.44 at $\nu = 0.1$. Beyond this peak, λ_2 decreases but remains relatively large, ending at ≈ 3.25 at $\nu = 1.0$. Notably, at higher viscosities ($\nu \geq 0.2$), λ_2 becomes comparable to, and in some cases even larger than, the λ_1 values observed at the lowest viscosities. The ratio λ_2/λ_1 increases significantly with viscosity, from ≈ 0.22 at $\nu = 0.01$ to ≈ 0.50 at $\nu = 0.1$ and ≈ 0.83 at $\nu = 1.0$. Similarly, λ_3 increases from ≈ 1.06 at $\nu = 0.01$ to a peak (e.g., ≈ 1.61 at $\nu = 0.1$), and then decreases, ending at ≈ 2.18 at $\nu = 1.0$. The ratio λ_3/λ_1 also increases substantially, from ≈ 0.15 at $\nu = 0.01$ to ≈ 0.29 at $\nu = 0.1$ and ≈ 0.55 at $\nu = 1.0$.

This shift in eigenvalue magnitudes signifies a fundamental change in how variance is distributed across the latent dimensions. At low viscosity, variance is highly concentrated along the first principal component, with PC1 explaining approximately 70.9% of the total variance. The first three PCs combined explain 97.3%. As viscosity increases, the dominance of PC1 diminishes (55.0% at $\nu = 0.1$, 39.3% at $\nu = 1.0$), and the contributions from PC2 and PC3 become increasingly significant. At $\nu = 1.0$, PC2 explains 32.5% and PC3 explains 21.8

3.2. *Effective dimensionality from PCA*

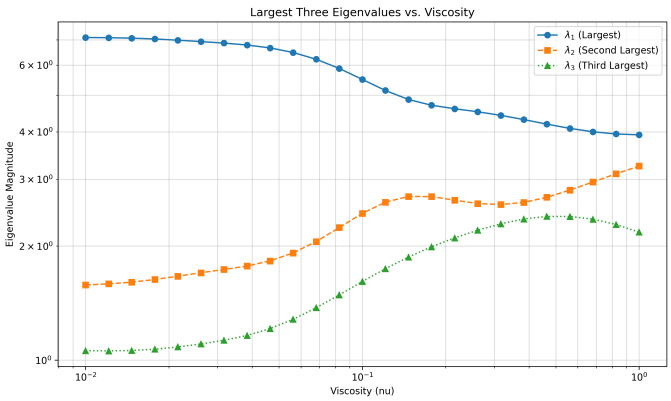

Figure 1. Evolution of the three largest eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$ from Principal Component Analysis of the latent space as a function of viscosity (ν). The magnitude of the largest eigenvalue (λ_1) decreases significantly with increasing viscosity, while the second (λ_2) and third (λ_3) eigenvalues generally increase. This shift in eigenvalue magnitudes indicates that the variance in the latent space becomes less concentrated in the dominant principal component and more distributed across secondary components as viscosity increases.

The number of principal components required to explain a specific cumulative percentage of the total variance serves as a PCA-based measure of effective dimensionality within the 10-dimensional latent space. Analyzing this metric for thresholds of 90%, 95%, and 99% variance explained reveals how the "size" of the subspace needed to capture most of the latent variability changes with viscosity, as shown in Figure 2.

For the 90% variance threshold, as seen in Figure 2, the number of required PCs generally remains low (2-3 PCs) across most of the viscosity range, increasing slightly to 3-4 PCs only at the highest viscosities. For the 95% variance threshold, 3 PCs are typically sufficient at low to intermediate viscosities. This number increases to 4 PCs for viscosities $\nu \geq 0.26$ and remains at 4 for $\nu = 1.0$.

The trend for the 99% variance threshold shows a more pronounced and interesting behavior (Figure 2). At the lowest viscosity ($\nu = 0.01$), 4 PCs are needed to capture 99% of the variance (as 3 PCs capture 97.3%). As viscosity increases, the number of required PCs rises, peaking in the intermediate viscosity range (around $\nu \approx 0.04 - 0.2$) at 5-6 PCs. Beyond this intermediate range, the number of PCs needed for 99% variance slightly decreases or stabilizes at 4-5 PCs for higher viscosities. For example, at $\nu = 0.1$, where 3 PCs capture 95.5% variance, more PCs are needed for 99%. At $\nu = 1.0$, where 4 PCs capture approximately 98.3% variance (93.5% by PC1-3 + 4.8% by PC4), 4 or 5 PCs are likely needed for 99%.

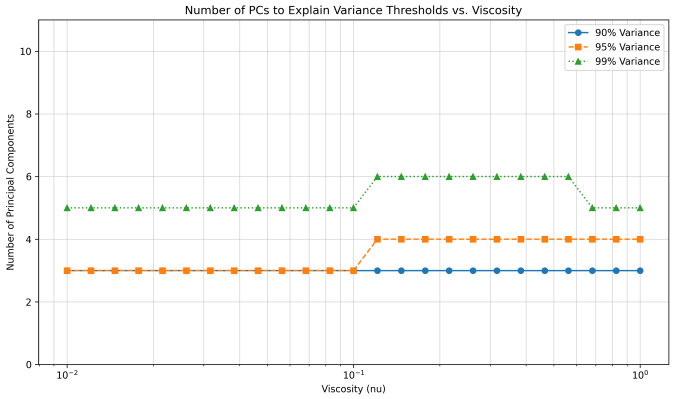

Figure 2. Number of principal components (PCs) required to explain 90%, 95%, and 99% of the total variance in the 10-dimensional latent space of a Physics-Informed Neural Network for the 2D Burger's equation, as a function of viscosity (ν). The number of PCs needed for 99% variance exhibits a non-monotonic trend, increasing at intermediate viscosities before decreasing at higher values, suggesting that the latent space representation is most complex in the intermediate regime.

This non-monotonic trend observed in the number of PCs required to explain 99% of the variance, with a peak in the intermediate viscosity regime (Figure 2), qualitatively aligns with previous findings on the intrinsic dimensionality (ID) of the latent space using methods like TwoNN. While the absolute values of PCA-based effective dimensionality (maximum around 6 PCs) are significantly lower than the TwoNN ID estimates (maximum around 40), which is expected as PCA measures linear dimensionality within the 10D ambient space while ID estimates the dimensionality of the potentially nonlinear manifold, the congruence in the non-monotonic pattern suggests that both metrics are capturing a similar underlying change in the complexity of the learned representation as the physics evolves. The peak at intermediate viscosities indicates that this regime requires the PINN to utilize a larger effective subspace within its 10-dimensional latent space to encode the physical state.

3.3. *Spectrum concentration: Normalized Shannon entropy*

To quantify how evenly the variance is distributed across the 10 principal components, we calculated the normalized Shannon entropy of the eigenvalue spectrum for each viscosity. A lower normalized entropy implies that variance is concentrated in a few leading PCs (a peaked spectrum), while a higher value indicates a more even distribution across dimensions (a flatter spectrum).

As shown in Figure 3, the normalized entropy of the eigenvalue spectrum generally increases with viscosity. At the lowest viscosity ($\nu = 0.01$), the normalized entropy is approximately 0.392, reflecting the strong dominance of PC1 as seen in Figure 1. As viscosity increases, the entropy rises, reaching approximately 0.504 at $\nu = 0.1$. The entropy continues to increase, peaking at intermediate-to-high viscosities (around $\nu \approx 0.3-0.7$) with values between 0.56 and 0.58. At the highest viscosity ($\nu = 1.0$), the normalized entropy is approximately 0.564, showing a slight decrease or plateau compared to the peak.

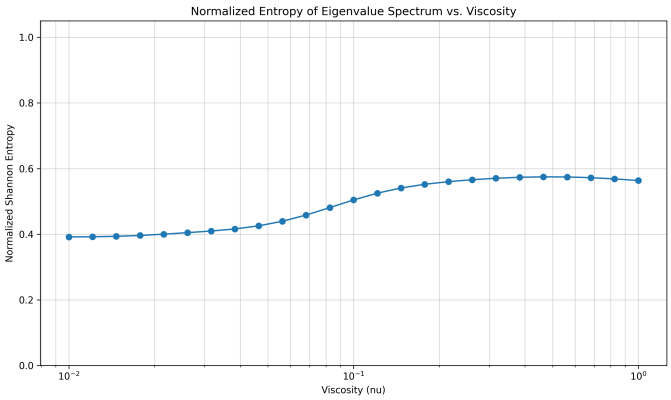

Figure 3. Normalized Shannon entropy of the principal component analysis eigenvalue spectrum of the latent space as a function of viscosity (ν). The entropy quantifies the evenness of variance distribution across principal components, with increasing entropy indicating variance becomes more distributed with increasing viscosity. This suggests the latent space representation shifts from a structure dominated by a few components at low ν to one with variance spread across more components at higher ν.

The overall increasing trend in entropy from low to high viscosity (Figure 3) suggests that, within the fixed 10-dimensional latent space, the variance becomes less concentrated in the leading principal components and is spread more broadly across the available dimensions. This implies that the network's representation becomes less "sparse" in terms of variance distribution as viscosity increases. This finding, indicating a flatter spectrum at higher viscosities, might seem counterintuitive if one expects a simpler, diffusion-dominated regime to result in a representation concentrated in fewer dimensions (lower entropy). However, this could be interpreted as the network re-purposing its fixed latent dimensions; even if the complexity of the underlying physical state decreases, the network might distribute the information more evenly across its features rather than collapsing it onto fewer dimensions. The slight decrease in entropy at the very highest viscosities, along with the slight decrease in PCs for 99% variance in this regime (Figures 3 and 2), provides some evidence that a form of simplification is eventually reflected in the PCA metrics.

3.4. *Stability of principal components (Eigenvectors)*

The stability of the directions of the principal components across different viscosity values was assessed by calculating the absolute cosine similarity between the eigenvectors at each viscosity and a reference set of eigenvectors obtained at the lowest viscosity ($\nu = 0.01$). A similarity score close to 1 indicates high alignment, while a score close to 0 indicates orthogonality (significant rotation). These similarities are plotted in Figure 4.

As shown in Figure 4, the direction of the first principal component (PC1) demonstrates relatively high stability across a wide range of low to intermediate viscosities. The similarity score is 1.0 at $\nu = 0.01$ (by definition) and remains high (e.g., ≈ 0.985 at $\nu = 0.1$). However, as viscosity increases into the higher regime, the stability of PC1 decreases, with the similarity dropping to approximately 0.679 at $\nu = 1.0$. This indicates that while the primary direction of variation learned at low viscosity maintains some relevance, it undergoes a substantial rotation in the high-viscosity regime, reflecting a qualitative change in the most dominant feature captured by the network.

The second (PC2) and third (PC3) principal components show less consistent stability (Figure 4). PC2 initially shows good stability (similarity $\approx$ 0.966 at $\nu = 0.1$), but its direction changes more significantly than PC1 at higher viscosities. The similarity for PC2 drops below 0.7 in the intermediate-to-high viscosity range ($\nu \in [0.261, 0.5623]$) and is approximately 0.634 at $\nu = 1.0$. PC3 also starts with high similarity (≈ 0.975 at $\nu = 0.1$) and its stability decreases with increasing viscosity, reaching ≈ 0.887 at $\nu = 1.0$. It appears more stable than PC2 at the highest viscosities, though its stability is lower than PC1's initial stability. Notably, in the range where PC2 stability drops below 0.7, the eigenvalues λ_2 and λ_3 become relatively close (e.g., at $\nu = 0.5623$, $\lambda_2/\lambda_3 \approx 1.17$), as seen in Figure 1. This proximity of eigenvalues suggests that the directions of PC2 and PC3 may be less uniquely defined and more susceptible to mixing or rotation, impacting their individual stability measures. While individual eigenvector directions might rotate, the subspace spanned by these components might be more stable.

The varying stability of the eigenvectors, as shown in Figure 4, suggests that the features learned by the PINN are not static but adapt their orientation in the latent

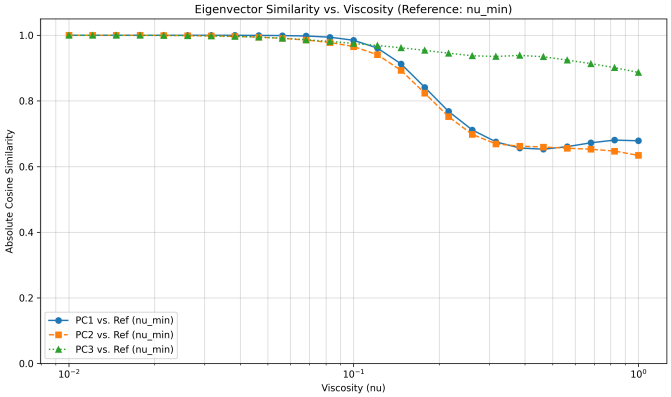

Figure 4. Absolute cosine similarity of the top three principal component (PC) eigenvectors of the latent space, calculated with respect to the eigenvectors at the lowest viscosity ($\nu = 0.01$), as a function of viscosity (ν). PC1 shows high stability across a wide range of viscosities, while PC2 and PC3 exhibit greater changes in direction at higher viscosities. This indicates that the most dominant learned feature direction is relatively robust to changes in viscosity, while secondary features undergo more significant reorientation.

space as the underlying physics changes. The relatively stable core represented by PC1 at lower viscosities might capture fundamental aspects of the solution structure that persist across different regimes, while the rotation at high ν signifies a necessary reorientation. The lower stability of secondary components (PC2, PC3) suggests that the finer details or less dominant modes of variation are more fluid in their representation.

3.5. *Interpretation and synthesis*

The PCA analysis provides a quantitative framework for understanding how the PINN's latent space structure evolves with viscosity, complementing and extending insights from intrinsic dimensionality (ID) analysis. The decrease in the magnitude of the leading eigenvalue (λ_1) with increasing viscosity (Figure 1) is consistent with the physical intuition that the Burger's equation solution simplifies in the diffusion-dominated regime, with smoother profiles requiring less variance to describe.

The trend in PCA-based effective dimensionality, particularly for the 99% variance threshold, which exhibits a non-monotonic behavior with a peak at intermediate viscosities (Figure 2), strongly echoes the non-monotonic trend observed previously for the latent space's intrinsic dimensionality. This alignment supports the conclusion that the intermediate viscosity regime represents the most complex state for the PINN's learned representation, likely corresponding to the physical regime where the interplay between advection and diffusion is most intricate.

The evolution of the normalized Shannon entropy of the eigenvalue spectrum (Figure 3) presents a more nuanced picture. The general increase in entropy with viscosity indicates that variance becomes more evenly distributed across the 10 latent dimensions. While a simpler physical system might intuitively suggest a representation concentrated in fewer dimensions (lower entropy), the PINN operates within a fixed 10D latent space. The increased entropy could reflect how the network utilizes these dimensions, perhaps distributing information differently even if the underlying manifold dimension decreases. The slight decrease or plateau in entropy at the very highest viscosities, along with the slight decrease in PCs for 99% variance in this regime (Figures 3 and 2), provides some evidence that a form of simplification is eventually reflected in the PCA spectrum.

The analysis of eigenvector stability (Figure 4) reveals that the primary direction of variation (PC1) is relatively robust across low-to-intermediate viscosities but undergoes significant rotation at high viscosities. This suggests that the most prominent learned feature adapts its orientation in the latent space as the physics transitions. The lower stability of secondary components (PC2, PC3), especially when eigenvalues are close (Figure 1), highlights the dynamic nature of the latent space structure beyond the dominant mode.

Collectively, these PCA results provide quantitative evidence that the PINN adapts the structure of its learned latent space in response to changes in the physical parameter ν. The observed trends in variance distribution (Figure 1), effective dimensionality (Figure 2), spectrum concentration (Figure 3), and eigenvector stability (Figure 4) are consistent with the network encoding the physical transitions of the 2D Burger's equation. The decrease in the dominance of the leading PC and the eventual slight decrease in effective dimensionality at high viscosity align with the concept of physical simplification. The non-monotonic complexity profile, captured by both PCA effective dimensionality and previous ID estimates, points to the intermediate viscosity regime as the most representationally complex state. These findings suggest that the PINN's latent space organization reflects principles analogous to coarse-graining, where the network implicitly learns representations that become simpler as the physical system simplifies in the diffusion-dominated limit, albeit with a complex reorganization of variance and feature directions within the fixed latent dimensions.

4. CONCLUSIONS

Interpreting the internal representations learned by deep neural networks, particularly Physics-Informed Neural Networks (PINNs) solving complex physical systems, remains a significant challenge. This study addressed this by quantitatively investigating the structure and evolution of the 10-dimensional latent space learned by a PINN for the 2D Burger's equation across a range of 25 viscosity values, a parameter known to drive transitions between turbulent-like and diffusion-dominated regimes.

Our approach involved applying Principal Component Analysis (PCA) independently to the standardized latent vectors extracted for each viscosity value. This allowed us to analyze how the statistical distribution of variance across the latent dimensions, represented by the eigenvalue spectrum, and the orientation of the principal directions, represented by the eigenvectors, changed as the underlying physics evolved. We quantified these changes through metrics such as eigenvalue magnitudes, PCA-based effective dimensionality (variance explained), normalized Shannon entropy of the spectrum (spectrum concentration), and the absolute cosine similarity of eigenvectors relative to a reference state.

The results provide quantitative evidence for the PINN's adaptation of its latent space structure to the physical regime. We observed a consistent decrease in the magnitude of the dominant eigenvalue (λ_1) with increasing viscosity, indicating that the primary mode of variation in the latent space becomes less pronounced as the system simplifies. Concurrently, the relative contributions of the second and third principal components increased, suggesting a more distributed variance across the leading dimensions at higher viscosities.

The PCA-based effective dimensionality, particularly when considering the number of components needed to explain 99% of the variance, showed a non-monotonic trend, peaking in the intermediate viscosity regime. This finding qualitatively aligns with previous analyses using intrinsic dimensionality methods and supports the conclusion that the intermediate regime, where advection and diffusion effects are balanced, represents the most complex state for the network's learned representation within the 10-dimensional space.

The normalized Shannon entropy of the eigenvalue spectrum generally increased with viscosity, indicating that the variance became more evenly distributed across the 10 latent dimensions. While potentially counterintuitive for a simpler physical system, this suggests that the network utilizes its fixed-dimensional latent space differently in the diffusion-dominated regime, distributing information more broadly rather than collapsing it onto fewer dimensions. A slight decrease or plateau in entropy at the highest viscosities hints at some simplification eventually being reflected in the spectrum flatness.

Analysis of eigenvector stability revealed that the primary direction of variation (PC1) maintains relative stability across low-to-intermediate viscosities but undergoes a significant rotation in the high-viscosity regime. This indicates that the most prominent learned feature reorients itself as the physical dynamics change fundamentally. Secondary components (PC2, PC3) exhibited less stability, particularly when their corresponding eigenvalues were close, highlighting the dynamic nature of the latent space organization beyond the dominant mode.

From these results, we learn that the PINN's latent space is not a static encoding but dynamically adapts its statistical structure in response to changes in governing physical parameters. The evolution of variance distribution, effective dimensionality, and eigenvector directions provides a quantitative window into how the network implicitly captures physical transitions. The observed non-monotonic complexity profile mirrors known aspects of the Burger's equation dynamics. The findings suggest that the network's learned representation for simpler, diffusion-dominated regimes involves a different organization of variance and feature directions within the fixed latent dimensions compared to more complex regimes, potentially reflecting principles analogous to coarse-graining where the representation simplifies as the physical system does. This study provides a quantitative framework for interpreting the learned representations of PINNs and demonstrates their capacity to encode physical complexity in their internal structure.