Quantifying the Complexity of Dynamical Systems to Improve ML Benchmarking

Victor Huynh, Vincent Thai

Brains in Silicon, Boahen Lab, Advised by Max Kanwal

Motivation

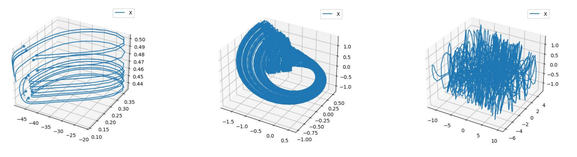
Dynamical systems* are everywhere!
* a system that evolves over time according to a specific set of rules

- Hard to predict, so we use machine learning (ML)
- Some ML algorithms work better for systems of certain complexity than others
- Helpful to characterize a **system's complexity**** in order to choose the best algorithm for the job
  ** difficulty of describing a system

Goal:
- Establish **mathematical** definition of complexity
- Find **parameters** that influence complexity

Complexity

LOW **MEDIUM** **HIGH**



Roadmap

1. Literature review to identify **mathematical measures** currently used to quantify complexity
2. Translate those math **equations into code**
3. **Numerical analysis** on equation parameters to investigate what variables affect complexity

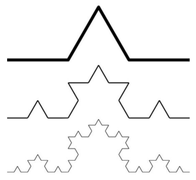
Acknowledgements

1. William Gilpin. "Chaos as an interpretable benchmark for forecasting and data-driven modelling" Advances in Neural Information Processing Systems (NeurIPS) 2021 https://arxiv.org/abs/2110.05266
2. Bhamidipaty, L. M., Bruzzese, T., Tran, C., Mrad, R. R., & Kanwal, M. (2023, November 2). Dynadojo: An extensible benchmarking platform for scalable... OpenReview. https://openreview.net/forum?id=pTSNoBTk8E

1. Mathematical Measures

Dimension: Correlation dimension, Kaplan Yorke dimension
Entropy: Multiscale entropy, Pesin entropy
Chaos: Maximum Lyapunov, Lyapunov Spectrum

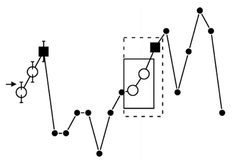
Dimension
characterizes the amount of fractal space occupied by the system



$$C(\epsilon) \equiv \lim_{N\to\infty} \frac{1}{N^2} \sum_{i,j=1}^{\infty} H\left(\epsilon - \left|x_i - x_j\right|\right)$$

(Correlation Dimension)

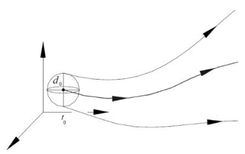
Entropy
quantifies informational predictability across the system



$$\mathrm{ApEn} \approx \Sigma - \log\,(1+A_i)/(1+B_i)$$

(Approximate Entropy)

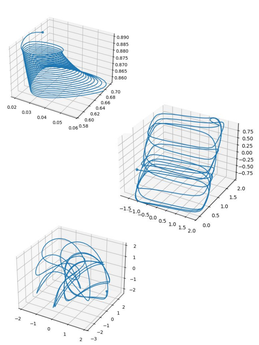
Chaos
measures the sensitivity to perturbations in initial condition



$$|\delta \mathbf{Z}(t)| \approx e^{\lambda t} |\delta \mathbf{Z}_0|$$

(Lyapunov Exponent)

Correlation Matrix



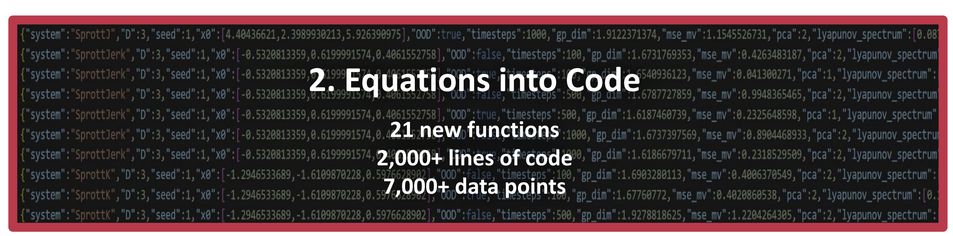
These measures uniquely capture three aspects of complexity

A) contextual
B) temporal
C) spatial

Importance of having various measures to be comprehensive

2. Equations into Code

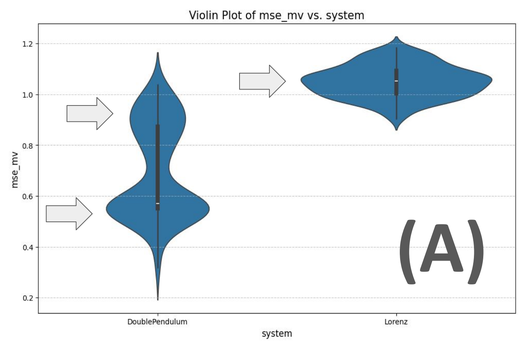
21 new functions
2,000+ lines of code
7,000+ data points

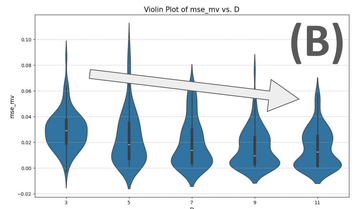


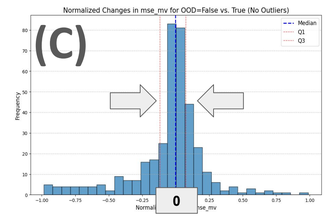
3. Numerical Analysis

Used Gilpin's database of **131 dynamical systems**
- Generated **trajectories from** these **systems** and applied complexity measures

Swept across **four parameters** to see how they affected measure values
- Embed dimension, init. condition, traj. length, in/out of distribution







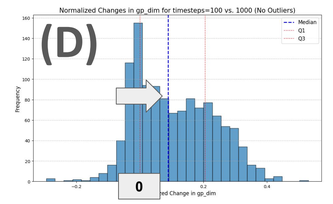
Graphs

(A) Distribution of complexities across 100 init condit for two systems

(B) Distribution of complexities across 20 init condit for different embedding dimensions

(C) Histogram of the change in complexities for ID vs OOD init condit for all systems

(D) Histogram of the change in complexities for trajectories of 100vs1000 for all systems



Key Takeaways

1. System complexities are only weakly correlated with embed dim **(B)** and in/out of distribution initial conditions **(C)**
2. Complexity saturates around 1000 timesteps, slight increase overall **(D)**
3. For certain systems with multi-modality, such as Double Pendulum, initial condition do significantly impact complexity! **(A)**
4. The correlation matrix captures three distinguishable aspects of complexity

Next Steps:

- **Sweep system coefficients** to see if they have any effect on complexity
- Implement **more complexity measures** to see if they reveal new aspect of complexity we haven't seen yet
- Implement **more systems** to test complexity measures on, beyond those defined by equations