The search for the lost attractor

Mario Pasquato
Département de Physique, Université de Montréal, Montréeal, Québec, Canada
Ciela Institute, Montréal, Québec, Canada
Mila - Québec Artificial Intelligence Institute, Montréal, Québec, Canada
Dipartimento di Fisica e Astronomia, Università di Padova, Padova, Italy
`mario.pasquato@umontreal.ca`

Syphax Haddad
Département de Physique, Université de Montréal, Montreal, Quebec, Canada
Cégep Trois Rivières, Trois Rivières, Quebec, Canada

Pierfrancesco Di Cintio
National Council of Research - Institute of Complex Systems, Sesto Fiorentino, Italy
INAF - Osservatorio Astrofisico di Arcetri, Firenze, Italy
INFN - Sezione di Firenze, Sesto Fiorentino, Italy

Alexandre Adam
Département de Physique, Université de Montréal, Montréal, Québec, Canada
Ciela Institute, Montréal, Québec, Canada
Mila - Québec Artificial Intelligence Institute, Montréal, Québec, Canada

Pablo Lemos
Département de Physique, Université de Montréal, Montréal, Québec, Canada
Ciela Institute, Montréal, Québec, Canada
Mila - Québec Artificial Intelligence Institute, Montréal, Québec, Canada
Center for Computational Astrophysics, Flatiron Institute, NYC, NY, USA

Noé Dia
Département de Physique, Université de Montréal, Montréal, Québec, Canada
Ciela Institute, Montréal, Québec, Canada
Mila - Québec Artificial Intelligence Institute, Montréal, Québec, Canada

Mircea Petrache
Facultad de Matemáticas & Instituto de Ingeneria Matematica y Computacional, Santiago, Chile

Ugo Niccolò Di Carlo
SISSA - Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy
INFN - Sezione di Trieste, Trieste, Italy

Alessandro Alberto Trani
Niels Bohr Institute, Copenhagen, Denmark
Research Center for the Early Universe, School of Science, The University of Tokyo, Tokyo, Japan
Okinawa Institute of Science and Technology, Okinawa, Japan

Laurence Perreault-Levasseur
Département de Physique, Université de Montréal, Montréal, Québec, Canada
Ciela Institute, Montréal, Québec, Canada
Mila - Québec Artificial Intelligence Institute, Montréal, Québec, Canada
Center for Computational Astrophysics, Flatiron Institute, NYC, NY, USA

Yashar Hezaveh
Département de Physique, Université de Montréal, Montréal, Québec, Canada
Ciela Institute, Montréal, Québec, Canada
Mila - Québec Artificial Intelligence Institute, Montréal, Québec, Canada
Center for Computational Astrophysics, Flatiron Institute, NYC, NY, USA

Abstract

N-body systems characterized by r^{-2} attractive forces may display a self-similar collapse known as the gravo-thermal catastrophe or core-collapse. Globular clusters are a real-life example of this in astronomy. In these clusters, collapse is halted by dynamical heating from binary stars. This is known as the binary-burning phase of star cluster evolution. A fraction of Milky Way globular clusters may have already reached this phase. It has been speculated –with guidance from simulations– that macroscopic variables such as central density and velocity dispersion are governed post-core-collapse by an effective, low-dimensional system of ordinary differential equations (ODEs). However, it is hard to distinguish potentially chaotic low-dimensional motion, from high-dimensional stochastic noise. Here we apply three machine learning tools to the time series of relevant macroscopic quantities from state-of-the-art dynamical simulations to constrain the post-collapse dynamics: topological data analysis (TDA) on a lag embedding, Sparse Identification of Nonlinear Dynamics (SINDY), and Tests of Accuracy with Random Points (TARP). Even though TARP suggests that the time series are not purely noise, we do not find a low-dimensional system of equations describing their time evolution.

1 Introduction

Point particles interacting through inverse square gravitational forces are ubiquitous in astronomy: from galaxies in galaxy clusters, to stars in galaxies and star clusters, down in scale to planetesimals and even dust [16]. While the general gravitational N-body problem is approached numerically either by brute forcing the pairwise interactions [direct N-body; 1] or by approximated schemes (multi-particle collision, MPC [8], Monte-Carlo [9]), it is expected that in specific regimes there may emerge tractable effective equations driving the evolution of relevant macroscopic quantities. It has been suggested that this is the case of post core-collapse oscillations in star clusters [2, 3]. E.g. [2] claim to find a three-dimensional attractor that qualitatively resembles the Rössler attractor [see 12].

Paper contribution: For the first time, we apply three independent machine learning tools to analyze state-of-the-art simulations of star cluster core collapse. We do not find clear evidence of effective low dimensional dynamics. Topological data analysis [TDA; 17] when applied to a lag embedding of the core density and velocity dispersion time series produces a persistence diagram that appears indistinguishable from noise; similarly, Sparse Identification of Nonlinear Dynamical Systems [SINDY; 4] fails to yield an equation reproducing the observed time series. Finally, Tests of Accuracy with Random Points [TARP; 11] find that samples from our original time series do not lie in-distribution with respect to random reshufflings. The post core-collapse evolution is thus not purely independent identically distributed noise, but this is not sufficient to conclude that it is the result of low-dimensional dynamics.

2 Simulations

We have performed a hybrid MPC-particle-mesh simulation with $N = 10^5$ particles where the collective gravitational force is evaluated with a standard particle in cell method on a $32 \times 16 \times 128$ spherical grid. The Poisson equation is solved with a finite difference scheme. Collisions are resolved with the MPC stochastic scheme [8] on the same grid, conditioned every time step Δt with the cell-based interaction probability $p_i = \text{Erf}\left(\beta \Delta t \nu_c\right)$, where $\nu_c = 8\pi G^2 \bar{m}_i^2 n_i \log \Lambda / \sigma_i^3$ is the typical collision frequency in cell i and β a dimensionless constant proportional to the total number of cells. Initial positions and velocities are sampled from an isotropic Plummer model with scale radius r_s and mass scale M. The equations of motion for the N particles are propagated between each collision step with a standard second order leapfrog scheme with a fixed timestep. All radii are expressed in units of the scale radius r_s, while the time unit is given as a function of the total mass M and the gravitational constant G as $t_{\text{dyn}} \equiv \sqrt{r_s^3 / GM}$. We fix $\Delta t = t_{\text{dyn}}/100$ and run until $2 \times 10^5 t_{\text{dyn}}$. As
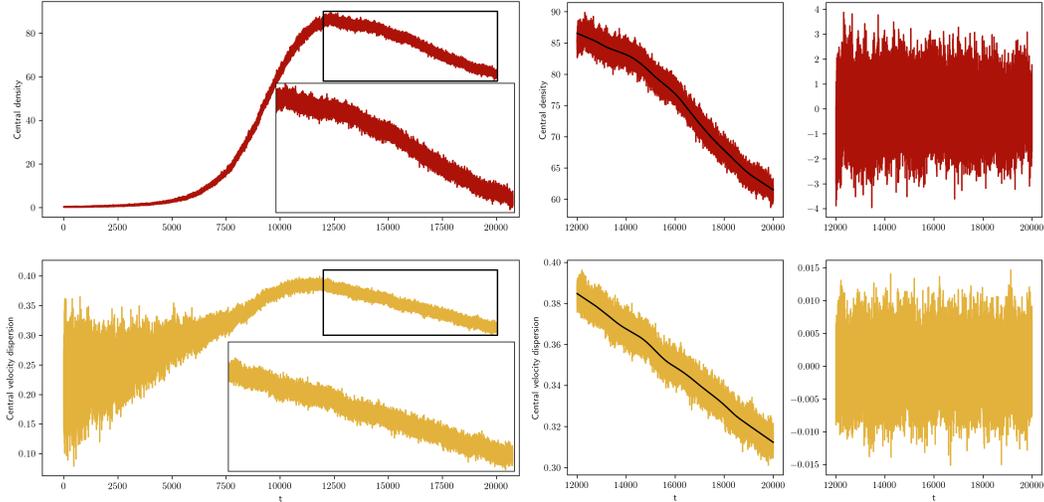
Figure 1: **Left panel:** Evolution of density (top, red) and velocity dispersion (bottom, yellow) in our simulation. Zoom in on the post-core-collapse part of the time series (inset panels). **Right panel:** Trend fitting (left, black solid lines) and detrended time series (right).

a function of time, we save the values of the central density and velocity dispersion (both evaluated within the time-dependent radius enclosing the 5% of M), as shown in Fig. 1, left panel.

3 Methods

3.1 Topological data analysis on lag embedding

TDA is the study of the topological properties of point clouds derived from data. The primary tool within TDA is persistent homology, which provides a multi-scale description of the homological features of a dataset such as connected components (H_0), loops (H_1), and voids (H_2). The main output of persistent homology is the barcode or persistence diagram (PD), visualizing the birth and death of topological features as one varies a scale parameter. See [6, 13, 5, 7] for comprehensive discussion.

We applied persistent homology to a lag embedding of the time series of velocity dispersion and of central density generated from our simulations. To achieve this, we deterended our time series by subtracting a polynomial fit (see Fig. 1, right panel), calculated the autocorrelation and sampled at regular intervals such that the autocorrelation signal became insignificant. This was the basis for computing a four-dimensional lag embedding. Our lag embeddings reconstruct a higher-dimensional phase space from a single time series, effectively transforming a one-dimensional series into a point cloud. This is justified by Takens' theorem, asserting that, under certain conditions, the dynamics of a system can be reconstructed using the time-delayed versions of even a single observable [14].

By applying TDA to this lag-embedded space, we can discern topological features that might correspond to dynamical structures or patterns in the original time series. For instance, loops revealed by persistent homology might indicate periodic orbits or recurrent dynamics. We used the *ripser* library [15] to compute persistent homology features on our timeseries.

3.2 Sparse Identification of Nonlinear Dynamics (SINDY)

SINDY is an algorithmic approach designed to derive governing equations of dynamical systems directly from observational data [4]. The central idea is to compute derivatives of the observed state variables and then express these derivatives as a sparse linear combination of a pre-defined library of functions. Here we relied on a polynomial library up to third degree, with interactions. We filtered our time series by removing high-frequency Fourier modes and took first order numerical derivatives of velocity dispersion and of central density generated from our simulations. We thus used SINDY
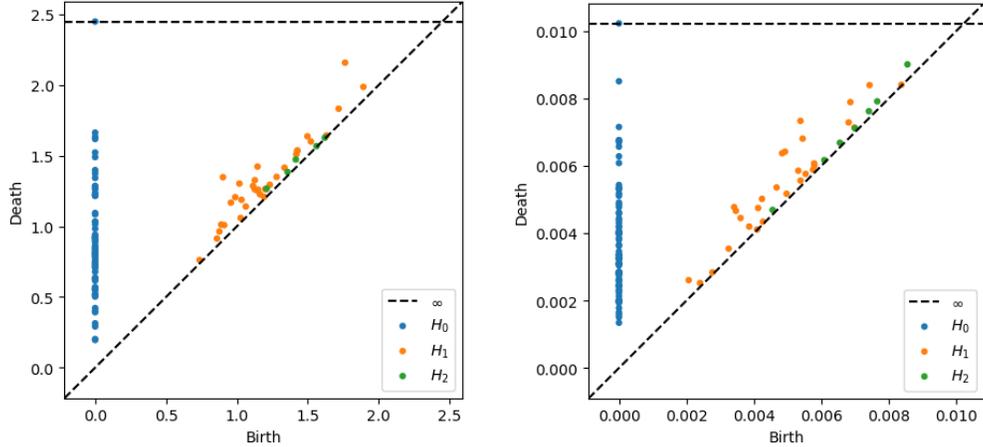
Figure 2: Persistence diagrams of central density (left) and velocity dispersion (right). Blue dots (H_0) represent connected components, essentially subclusters of points in the point cloud. Orange dots (H_1) represent rings, the kind of feature we expect if the evolution in lag embedding space is (quasi) periodic, e.g. approaching a limit cycle. Green dots (H_2) represent voids, which may correspond to a more complex but still low-dimensional dynamics on a 2D surface.

with LASSO regularization and MSE loss to look for second-order ODEs, the minimum order that allows for oscillatory behavior. At the implementation level, we relied on the *pysindy* library [10].

3.3 Tests of Accuracy with Random Points

TARP is a necessary and sufficient coverage test to assess whether a set of points has been drawn from a given distribution for which a generative model is available [11]. TARP has been originally designed to test the accuracy of posterior estimators by using random points to assess coverage probabilities, but here we are applying it essentially as an anomaly detection technique, for the task of distinguishing our original time series from random reshuffled versions of it. For our purposes, the main output of TARP is the expected coverage probability as a function of credible intervals. When plotted, this reveals that a sample is out of distribution if it deviates from the identity line. Unlike the previous two approaches, finding such a deviation through TARP does not prove, per se, that the dynamics producing our time series in low dimensional.

4 Results

4.1 Topological data analysis on lag embedding

The PDs of the central density and velocity dispersion time series are shown in Fig. 2. In a PD, stable topological structure is represented by points far from the diagonal. Qualitatively, these do not appear to be present in Fig. 2 for rings and voids, suggesting that these topological features picked up by persistent homology are just noise. Quantitatively, we tested this by computing the persistent entropy statistic on our original time series and on 100 random re-shufflings thereof. Persistent entropy is the Shannon entropy of the lifetime of features in a persistence diagram. The persistent entropy for the original time series typically falls within the distribution of the reshuffled time series, both for density and velocity dispersion. This is not the case for a time series obtained from the Rössler attractor, which happens to have systematically lower persistent entropy than its reshufflings.

4.2 Sparse Identification of Nonlinear Dynamics (SINDY)

We conducted a thorough analysis of the ODEs proposed by SINDY while varying the strength of regularization, which results in different levels of sparsity in the ODE coefficients. Neither a grid search nor a meticulous manual exploration resulted in equations able to reproduce our time series in the long term. In Fig. 3 (left panel) we show the solution of one such equation over time. While it

4
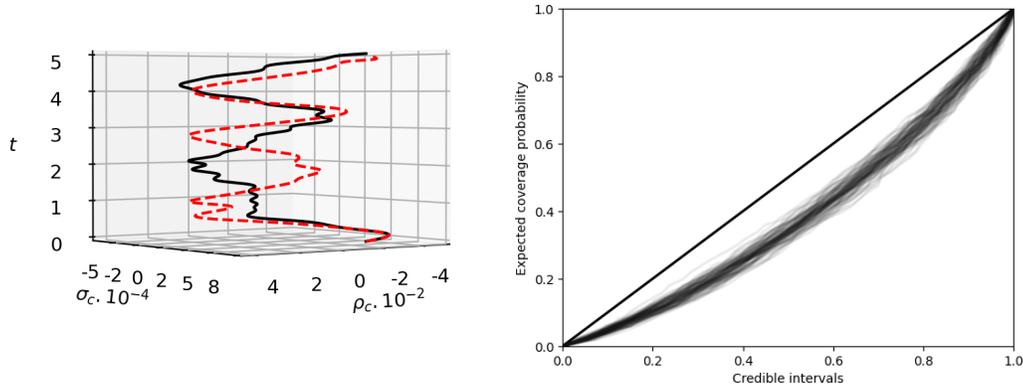
Figure 3: **Left panel**: solution of a hand-picked ODE found by SINDY (red dashed line) versus the smoothed time series it was learned from (solid black line). Up until $t \approx 0.5$ it follows the data closely, then it deviates catastrophically even though it still shows some oscillatory behavior. Times are in arbitrary units. **Right panel:** TARP coverage diagram. The 100 semitransparent solid curves correspond each to a run of TARP. They all present bias, systematically deviating downwards with respect to the diagonal.

initially reproduces the time series it is learned from, it eventually departs dramatically from it. This was the case for all systems of ODEs we found.

4.3   Tests of Accuracy with Random Points

Fig. 3 (right panel) shows 100 runs of TARP, none of which overlaps the diagonal line. Our original time series is thus biased (out-of-distribution) with respect to its reshufflings, suggesting that reshuffling erases some temporal dependence. In other words, our time series is not a sequence of independent identically distributed random variables. This can be due to various reasons, including poor trend removal in the pre-processing phase; it does not *per se* indicate that our time series are generated by a low dimensional dynamics. It does justify further investigation into the matter.

5   Conclusions

We applied three machine learning tools that are new to astronomy, especially in the context of star cluster dynamics, to the problem of characterizing the dynamical evolution of a simulated N-body system in the post core-collapse phase. TDA on a lag embedding of the central density and velocity dispersion time series, as well as SINDY, did not provide us with evidence that the dynamical evolution is driven by a low-dimensional system of ODEs, despite the expectations of [2]. On the other hand, TARP revealed that the time series we analyzed are out-of-distribution with respect to randomly reshuffled time series. This suggests that there is some structure in the (apparent) noise. The search for the lost attractor is not over yet.

Acknowledgments and Disclosure of Funding

References

[1] Sverre J. Aarseth. *Gravitational N-Body Simulations*. Cambridge university press, 2003.

[2] Joseph L. Breeden and Haldan N. Cohn. Chaos in Core Oscillations of Globular Clusters. *Astrophysical Journal*, 448:672, August 1995.

[3] Philip G. Breen and Douglas C. Heggie. Dynamical evolution of black hole subsystems in idealized star clusters. *Monthly Notices of the Royal Astronomical Society*, 432(4):2779–2797, July 2013.

[4] Steven Brunton, Joshua Proctor, and Nathan Kutz. Sparse identification of nonlinear dynamics (sindy). In *APS Division of Fluid Dynamics Meeting Abstracts*, pages L8–005, 2016.

[5] Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve Oudot. *The structure and stability of persistence modules*, volume 10. Springer, 2016.

[6] Frédéric Chazal and Bertrand Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence*, 4:667963, 2021.

[7] Tamal Krishna Dey and Yusu Wang. *Computational topology for data analysis*. Cambridge University Press, 2022.

[8] Pierfrancesco Di Cintio, Mario Pasquato, Hyunwoo Kim, and Suk-Jin Yoon. Introducing a new multi-particle collision method for the evolution of dense stellar systems. Crash-test N-body simulations. A&A, 649:A24, May 2021.

[9] M. Giersz and R. Spurzem. A stochastic Monte Carlo approach to model real star cluster evolution - II. Self-consistent models and primordial binaries. *Monthly Notices of the Royal Astronomical Society*, 317(3):581–606, September 2000.

[10] Alan Kaptanoglu, Brian de Silva, Urban Fasel, Kadierdan Kaheman, Andy Goldschmidt, Jared Callaham, Charles Delahunt, Zachary Nicolaou, Kathleen Champion, Jean-Christophe Loiseau, J. Kutz, and Steven Brunton. PySINDy: A comprehensive Python package for robust sparse system identification. *The Journal of Open Source Software*, 7(69):3994, January 2022.

[11] Pablo Lemos, Adam Coogan, Yashar Hezaveh, and Laurence Perreault-Levasseur. Sampling-Based Accuracy Testing of Posterior Estimators for General Inference. *arXiv e-prints*, page arXiv:2302.03026, February 2023.

[12] C. Letellier and O. E. Rossler. Rossler attractor. *Scholarpedia*, 1(10):1721, 2006. revision #91731.

[13] Elizabeth Munch. A user's guide to topological data analysis. *Journal of Learning Analytics*, 4(2):47–61, 2017.

[14] Floris Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980: proceedings of a symposium held at the University of Warwick 1979/80*, pages 366–381. Springer, 2006.

[15] Christopher Tralie, Nathaniel Saul, and Rann Bar-On. Ripser.py: A lean persistent homology library for python. *The Journal of Open Source Software*, 3(29):925, Sep 2018.

[16] M. Trenti and P. Hut. N-body simulations (gravitational). *Scholarpedia*, 3(5):3930, 2008. revision #91544.

[17] Larry Wasserman. Topological data analysis. *Annual Review of Statistics and Its Application*, 5:501–532, 2018.