

# A Deep Learning Framework for Identifying Weakly Chaotic, Strongly Chaotic, Resonant and Non-resonant Orbits in the Generalized Kicked Rotator

Jingyue Hao<sup>a</sup>, Zhiguo Xu<sup>a</sup>, Jian Zu<sup>b,\*</sup>

<sup>a</sup>School of Mathematics, Jilin University, Changchun 130012, P.R. China

<sup>b</sup>Center for Mathematics and Interdisciplinary Sciences, School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, P.R. China

## Abstract

Identifying the types of orbits is an important topic in the study of chaotic dynamical systems. Beyond the well-known distinctly chaotic and regular motions, we focus on dynamics occurring in regions where regular and chaotic motions coexist and intertwine, which potentially indicating weakly chaotic orbits. This intermediate regime lies between strongly chaotic dynamics, characterized by exponential sensitivity and completely non-chaotic, purely regular behavior. In this paper, we introduce a deep learning framework to identify the types of orbits in the generalized kicked rotator system, which is challenging to study due to its complex and mixed chaotic behaviors. Our deep learning framework can be divided into two steps. First, we propose a novel algorithm that integrates the weighted Birkhoff average, the Lyapunov exponent, and the correlation dimension to identify weakly chaotic orbits. The algorithm categorizes orbits into four types: weakly chaotic, strongly chaotic, and regular orbits (which are further subdivided into resonant and non-resonant orbits), thereby creating a valuable dataset required for deep learning models. Second, we demonstrate that a well-trained 2D-CNN achieves high performance in accurately classifying orbits, largely because it effectively leverages the 2D structural information of the phase space relation. To our knowledge, this is the first paper to identify weakly chaotic orbits using deep learning methods. The method can be easily extend to other models.

**Keywords:** Generalized kicked rotator, Birkhoff averages, Lyapunov exponent, Deep learning, Orbits classification

## 1. Introduction

In the exploration of nonlinear dynamic systems, chaos theory, as one of the major scientific revolutions of the 20th century, has profoundly reshaped our understanding of complexity and predictability in nature. It reveals that even deterministic systems can exhibit extreme sensitivity to initial conditions and long-term behavioral unpredictability—phenomena that are ubiquitous from celestial mechanics to ecological models. Within this framework, identifying different types of orbits becomes a critical endeavor, as it allows researchers to decipher the fundamental transitions between ordered and chaotic motions, thereby uncovering the underlying mechanisms governing system evolution. One illustrative paradigm for studying such dynamics is provided by two-dimensional area-preserving maps. In integrable systems, motion is regular and confined to invariant tori, with each orbit characterized by a specific rotation number. However, this orderly structure is highly fragile under perturbations. According to the Kolmogorov-Arnold-Moser (KAM) theory [1], when a small disturbance is introduced, most invariant tori persist, while others degenerate into isolated periodic orbits, resonant islands, or chaotic regions—giving rise to intricately interwoven dynamical patterns. The coexistence and interaction of these distinct orbital types not only exemplify the universal route from regularity to chaos but

also serve as a cornerstone for understanding stability, controllability, and complexity in physical, biological, and engineering systems.

Lyapunov exponent is a widely recognized method for identifying different dynamical phenomena, as it measures the separation of nearby orbits. As noted by Eckmann and Ruelle [2] and Ott [3], the magnitude of Lyapunov exponents reflects the “strength” of chaotic dynamics. Distinctly positive Lyapunov exponents typically characterize strongly chaotic orbits, while values approaching zero suggest weakly chaotic orbits—a behavior characterized by subexponential divergence. Regions where regular and chaotic dynamics coexist often contain weakly chaotic orbits. These orbits represent an intermediate regime between strongly chaotic dynamics and purely regular dynamics. Macroscopically, weakly chaotic orbits exhibit anomalous dynamical features, such as dynamical aging and anomalous diffusion. Typical examples include maps with indifferent fixed points, polygonal billiards, and Hamiltonian systems featuring sticky islands in phase space [4].

Besides Lyapunov exponent, there are many other methods used to classify dynamical behaviors. Kolmogorov-Sinai entropy (K-S entropy) quantifies the degree of disorder or unpredictability in a dynamical system [5]. A K-S entropy of zero corresponds to a regular system, where the future state can be determined from the initial conditions. In contrast, a positive value indicates that the system exhibits chaotic behavior. Pesin’s theorem establishes the fundamental relationship between K-S entropy and Lyapunov exponent, demonstrating that

\*Corresponding author: zuj100@nenu.edu.cn

under appropriate conditions, K-S entropy equals the sum of all positive Lyapunov exponents [6, 7]. The weighted Birkhoff average (WBA) allows for quick and accurate distinction between regular and chaotic orbits [8, 9]. For chaotic orbits, WBA typically converges slowly due to mixing and sensitivity to initial conditions. In contrast, for regular orbits, the WBA converges significantly faster, often exhibiting super-convergence.

The traditional methods mentioned above typically require explicit knowledge of the system’s governing equations or substantial trajectory data, which limits their applicability in scenarios with only observational data or limited data. Deep Learning (DL) offers a promising solution for classifying motion types from time series data, which enables more comprehensive cartographic studies at reduced computational costs [10, 11, 12, 13]. Furthermore, DL can predict the future behavior of chaotic systems directly from time series data [14, 15]. In [10], Barrio et al. established the efficacy of convolutional neural networks (CNNs) for the binary classification of dynamical system orbits, demonstrating that a properly designed CNN achieves high accuracy in distinguishing regular from chaotic regimes using finite-time orbits substantially shorter than those required by traditional Lyapunov methods. In [13], Lee et al. focused on binary orbit classification, systematically comparing multilayer perceptron (MLP), CNN, and long short-term memory (LSTM) architectures for categorizing time series from the Logistic map and Lorenz system into regular or chaotic categories. In [11], Uzun et al. explored binary system classification, employing a deep learning approach that converts time series into graphic images and uses transfer learning models like SqueezeNet and ResNet to achieve high accuracy in distinguishing between different chaotic systems. The seminal work by Celletti et al. [12] addressed a three-class orbit classification problem, providing a systematic categorization of three fundamental motion types—chaotic, rotational, and librational—in pendulum-like systems and the spin-orbit model, with the InceptionTime CNN architecture exhibiting superior performance.

The standard map in particular has been widely used in diverse scientific fields, such as particle dynamics in accelerators [16], comet dynamics [17], and the autoionization of molecular Rydberg states [18]. Inspired by its extensive applications and theoretical importance in chaos theory, the generalized standard map (also known as generalized kicked rotator) was recently constructed in [19]. This generalized version is considered explanatory and more appropriate for modeling complex systems that cannot be adequately reduced to the original standard map as a first approximation. Given the increased complexity of this generalized kicked rotator (GKR) system, identifying weak chaos and classifying orbits within it using deep learning (DL) has emerged as a significant research endeavor. The GKR system is defined as:

$$\begin{aligned} x_{n+1} &= x_n + y_{n+1}, \\ y_{n+1} &= y_n + K \sum_{j=1}^M \sin(2\pi j x_n), \end{aligned} \quad \text{mod } 1, \quad (1)$$

where  $x_n$  represents the angular position of the particle after

the  $n$ -th iteration, and  $y_n$  the corresponding momentum. The term  $K \sum_{j=1}^M \sin(2\pi j x_n)$  refers to the generalized force, where  $K$  is a positive dimensionless map parameter that controls the extent of nonlinearity of the system,  $M$  is a positive integer. This seemingly simple setup gives rise to rich and complex dynamics, including strong chaos, weak chaos, and invariant tori, making it an ideal testing ground for classifying different types of motion.

In this paper, we propose a novel algorithm for identifying weakly chaotic orbits and classifying orbital dynamics in the GKR system into four distinct categories: strongly chaotic, weakly chaotic, resonant, and non-resonant orbits. A key innovation of our approach lies in the synergistic integration of three complementary methods—the weighted Birkhoff average, Lyapunov exponent, and correlation dimension—which collectively enable a more robust and accurate characterization of weak chaos, a regime often challenging to delineate with single-metric methods. To generate high-quality labeled data for deep learning, we first evolve a set of orbits from given initial conditions over sufficiently long timescales, applying our algorithm to assign a definitive dynamical category to each orbit. These algorithmically generated labels, derived from long-term iterations, are then paired with corresponding shorter finite-length orbital segments to construct the training and validation datasets. This strategy of using extended trajectories for label generation and shorter subsequences as model inputs effectively balances label reliability with data diversity, forming a scalable and practical foundation for subsequent DL-based classification.

In the deep learning component, we employ the 2D-CNN method to fully leverage the informational richness of image-based representations. This image-based approach (unlike direct time-series processing) captures inherent spatial patterns in the trajectory data. By analyzing features in these trajectory images, it effectively identifies orbit types. When trained on a dataset of indexes paired with their corresponding orbits, the model achieves outstanding classification performance, with accuracy exceeding 99%. This significantly outperforms the classification accuracy of models like MLP, InceptionTime CNN, and their Transformer-based hybrids (i.e., MLP-Transformer, InceptionTime CNN-Transformer). Furthermore, the 2D-CNN demonstrates considerable generalization capability, confirming its robustness in handling varied dynamical regimes. To the best of our knowledge, this represents the first study employing a machine learning approach to classify chaotic dynamical systems into four distinct categories—strongly chaotic, weakly chaotic, resonant and non-resonant orbits.

The paper is structured as follows. In Section 2, we detail the algorithms used to identify the weakly chaotic orbits and classify the orbits into the aforementioned four types. In Section 3, we compare the classification performance of several different DL models and demonstrate that the 2D-CNN model achieves exceptional classification results. Finally, in Section 4, we summarize the main findings of our study and discussing their implications.

## 2. Orbits classification

This section presents an algorithm to identify weakly chaotic orbits and, for the first time, classifies orbits in the GKR system into four distinct categories. For convenient, we take  $\mathbf{x}_n = (x_n, y_n)^T$ , then Eq.(1) can be rewritten as:

$$\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n) = \begin{pmatrix} x_n + y_n + K \sum_{j=1}^M \sin(2\pi j x_n) \\ y_n + K \sum_{j=1}^M \sin(2\pi j x_n) \end{pmatrix} \bmod 1.$$

Clearly,  $\mathbf{f}$  is a map from  $\mathbb{T}^2$  to itself. When  $M = 1$ , the system is called the classical standard map, which was introduced by Chirikov [20]. For  $K = 0$ , the system is completely integrable and every orbit lies on an invariant torus. For  $K \ll 1$ , according to the KAM theory, most of invariant tori are preserved, and resonant “islands” emerge. These islands are surrounded by “chaotic sea”.

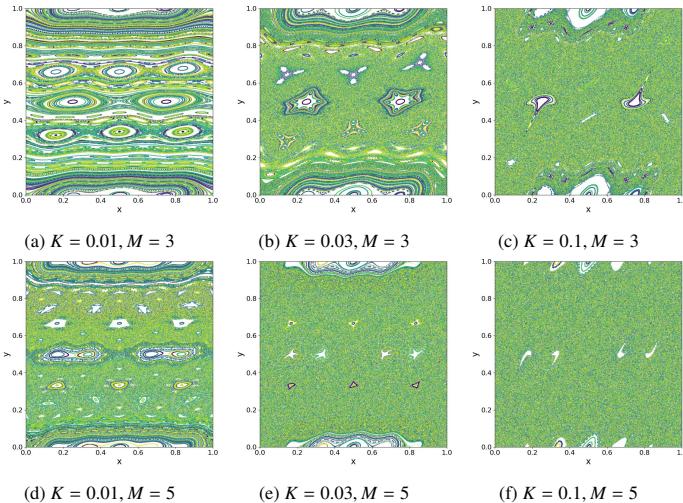


Figure 1: The orbits of the GKR system for  $K = 0.01, 0.03, 0.1$  with  $M = 3$  (Top) and  $M = 5$  (Bottom) from left to right. Here, we take 400 different initial conditions, randomly distributed throughout the phase space and iterate each of these initial conditions through the generalized kicked rotator 1000 steps. Different colors represent orbits with different initial conditions.

Figure 1 illustrates the phase space of the GKR system for varying parameters. It is evident that as the parameters  $K$  and  $M$  increase, the dynamical behavior of the system undergoes significant changes: some invariant tori are preserved, while certain resonant islands gradually sink into the connected chaotic sea, as shown from left to right in both the top and bottom rows. This suggests the possible presence of weakly chaotic orbits during the transition from regularity to chaos.

Figure 1 illustrates the phase space of the GKR system for varying parameters. It is evident that as the parameters  $K$  and  $M$  increase, the dynamical behavior of the system undergoes significant changes: some invariant tori are preserved, while certain resonant islands gradually sink into the connected chaotic sea, as shown from left to right in both the top and bottom rows.

This suggests the possible presence of weakly chaotic orbits during the transition from regularity to chaos.

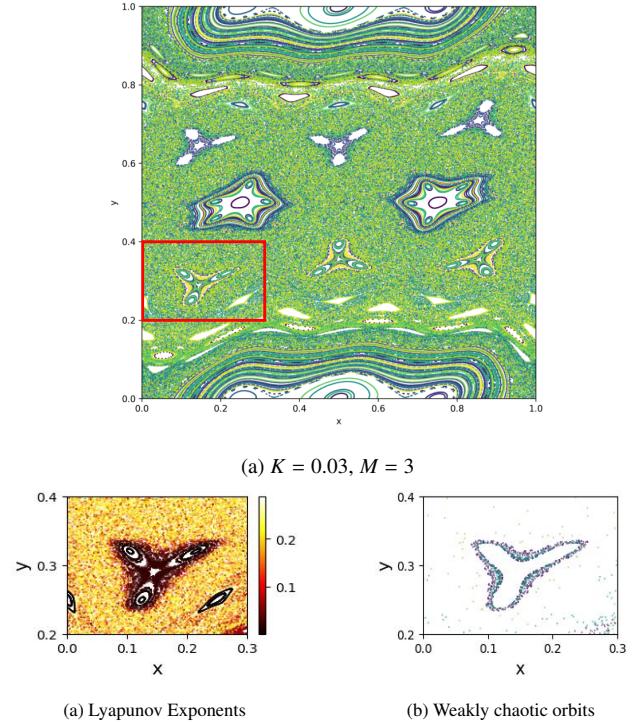


Figure 2: (a) The orbits of the GKR system for  $K = 0.03, M = 3$ . (b) Lyapunov exponent heatmap and (c) weakly chaotic orbits for the region  $[0, 0.3] \times [0.2, 0.4]$  in (a).

As shown in Figure 2, resonant islands coexist with surrounding chaotic orbits within the region  $[0.0, 0.3] \times [0.2, 0.4]$ . The orbits around the edge of the resonant island are candidates for being weakly chaotic orbits. To quantitatively verify this phenomenon, we calculate the Lyapunov exponents (finite-time approximations) within this region for the case  $K = 0.03, M = 3$ . Orbit lying on resonant islands exhibit relatively small Lyapunov exponents (on the order of  $10^{-3}$ ). As orbits move away from these islands, the exponents increase continuously, reaching values around 0.25 in the central chaotic sea, which indicates strong chaos, see Figure 2b. If only regular orbits and the chaotic sea were present, a clear gap in the Lyapunov exponents would be expected. However, as illustrated in Figure 3, the exponents change gradually without a distinct separation. This continuous transition, in the absence of a clear gap, suggests the existence of weakly chaotic orbits forming an intermediate region between the resonant islands and the chaotic sea, as visualized in Figure 2c.

We employ a hierarchical classification framework utilizing three methods sequentially to identify weakly chaotic orbits, as illustrated in Figure 4. First, the weighted Birkhoff average (WBA) method is used to identify an initial set of strongly chaotic orbits (Strongly chaotic I) and regular orbits (Regular I). The remaining orbits are labeled Unclassified I. Second, we compute the Lyapunov exponents (LE) for all orbits. Using the Regular I and Strongly chaotic I groups as references, we establish LE thresholds to classify a portion of the unclassified I

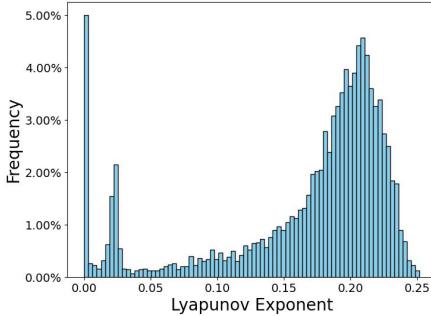


Figure 3: The histogram of Lyapunov exponents for the region  $[0, 0.3] \times [0.2, 0.4]$  with the case  $K = 0.03, M = 3$ .

orbits into strongly chaotic orbits (Strongly chaotic II) and regular orbits (Regular II). The orbits that remain unclassified after this step are labeled Unclassified II. Finally, we perform a correlation dimension (CD) analysis. By examining the CD distributions of the previously classified orbits (Strongly Chaotic I  $\cup$  II and Regular I  $\cup$  II), we establish two thresholds. Based on these thresholds, the Unclassified II orbits are classified into Strongly chaotic III, Regular III, and Weakly chaotic orbits. The detailed process is outlined below.

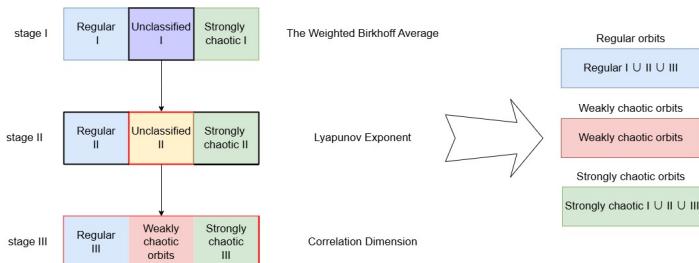


Figure 4: The framework outlines the classification process of different types of orbits in the Eq.(1).

## 2.1. The weighted Birkhoff average

According to [21], the Birkhoff average of a function  $h \in L^1(\mathbb{T}^2, \mathbb{R})$  along the trajectory of the map  $\mathbf{f}$  starting at  $\mathbf{x}_0$  is defined by

$$B_N(h)(\mathbf{x}_0) = \frac{1}{N} \sum_{n=0}^{N-1} h(\mathbf{f}^n(\mathbf{x}_0)), \quad (2)$$

where  $N$  is the number of iterations. Under mild hypotheses, the Birkhoff Ergodic Theorem (see Theorem 4.5.5. in [22]) states that

$$\lim_{N \rightarrow \infty} B_N(h)(\mathbf{x}_0) \triangleq B(h)(\mathbf{x}_0) = \int_{\mathbb{T}^2} h d\mu,$$

where  $\mu$  is an invariant probability measure for the trajectory's closure. Due to edge effects at the two ends of the finite orbit segment, the convergence of (2) may be very slow. In [23], it is shown that even if the orbit lies on a smooth invariant torus with irrational rotation number, the convergence rate of (2) is at best as  $O(N^{-1})$ . By contrast, for chaotic orbits, the convergence

is typically considered to be  $O(N^{-1/2})$ , which is essentially implied by the central limit theorem [24, 25]. Although the convergence rate of the Birkhoff average can distinguish chaotic from regular orbits, the practical difference is often not significant enough for efficient classification.

To accelerate the convergence, we employ the weighted Birkhoff average (WBA) method. Following [21], the WBA is defined as

$$WB_N(h)(\mathbf{x}_0) = \frac{1}{S} \sum_{n=0}^{N-1} \Psi\left(\frac{n}{N}\right) h(\mathbf{f}^n(\mathbf{x}_0)), \quad (3)$$

with the normalization constant

$$S = \sum_{n=0}^{N-1} \Psi\left(\frac{n}{N}\right).$$

$\Psi$  is a  $C^\infty$  weight function

$$\Psi(s) \triangleq \begin{cases} e^{-s(1-s)^{-1}}, & s \in (0, 1), \\ 0, & \text{else} \end{cases} \quad (4)$$

to downplay the influence of the endpoints in a finite time series. This exponential bump function vanishes with infinite smoothness at the endpoints, meaning,  $\lim_{s \rightarrow 0^+} \Psi^{(k)}(s) = \lim_{s \rightarrow 1^-} \Psi^{(k)}(s) = 0$  for all  $k \in \mathbb{N}$ . Consequently, we can define  $\Psi^{(k)}(0) = \Psi^{(k)}(1) = 0$  for all  $k \in \mathbb{N}$ . It is established that

$$\lim_{N \rightarrow \infty} WB_N(h)(\mathbf{x}_0) = B(h)(\mathbf{x}_0).$$

The key advantage of WBA lies in its convergence rate for regular orbits. For chaotic orbits, the convergence remains slow, with  $|WB_N(h)(\mathbf{x}_0) - B(h)(\mathbf{x}_0)| \sim O(N^{-1/2})$ , similar to the standard Birkhoff average [24, 25]. However, for regular orbits, the convergence can be dramatically faster [24]. In particular, if the orbit is conjugate to a rigid rotation with a Diophantine rotation vector  $\omega$  and function  $h$  is  $C^\infty$ , then (3) converges faster than any power, i.e., for any  $k \in \mathbb{N}$ , exists a constant  $C_k$  such that

$$|WB_N(h)(\mathbf{x}_0) - B(h)(\mathbf{x}_0)| < \frac{C_k}{N^k}.$$

Furthermore, if  $h$  is analytic, then the convergence rate is of exponential type  $O(\exp(-N^\zeta))$  for some  $\zeta > 0$  [26].

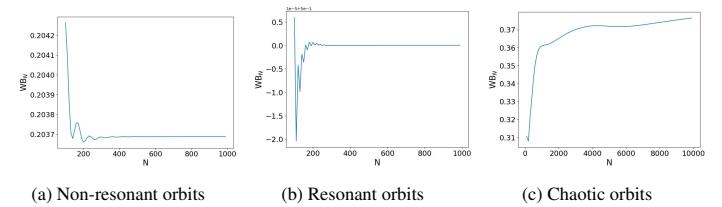


Figure 5: Convergence behavior of the  $WB_N$  series terms

Figure 5 illustrates the terms of the series  $WB_N$  with  $h(x, y) = y$  for three types of orbits in the GKR system. It is evident

that for regular orbits (both non-resonant (a) and resonant orbits (b)),  $WB_N$  converges rapidly, often within  $1 \times 10^3$  iterations. In contrast, for chaotic orbit, no clear convergence is observed even after  $2 \times 10^4$  iterations. This stark contrast demonstrates that analyzing the convergence of WBA provides an effective method for distinguishing chaotic from regular dynamics. However, since visually inspecting the convergence for every orbit is impractical, a quantitative criterion based on this method must be developed.

For a given orbit  $\{\mathbf{f}^n(\mathbf{x}_0) : n \in \mathbb{N}\}$ , we compute the value of the  $WB_N(h)$ , defined in Eq.(3) for two distinct segments of the orbit. The first segment comprises the initial  $N$  iterates,  $\{\mathbf{x}_0, \mathbf{f}(\mathbf{x}_0), \mathbf{f}^2(\mathbf{x}_0), \dots, \mathbf{f}^{N-1}(\mathbf{x}_0)\}$ . The second segment comprises the subsequent  $N$  iterates, starting from the  $N$ -th point:  $\{\mathbf{f}^N(\mathbf{x}_0), \mathbf{f}^{N+1}(\mathbf{x}_0), \dots, \mathbf{f}^{2N-1}(\mathbf{x}_0)\}$ . If the orbit is regular,  $WB_N(h)$  converges rapidly. Consequently, the difference between its value computed over the first  $N$  iterates and the next  $N$  iterates should be negligible. In contrast, for a chaotic orbit, the discrepancy is typically significant. Based on this, we define a discriminating quantity:

$$\Delta_N \triangleq WB_N(h)(\mathbf{x}_0) - WB_N(h)(\mathbf{f}^N(\mathbf{x}_0)). \quad (5)$$

To classify the orbit, we quantify the magnitude of  $\Delta_N$  by defining a measure of agreement between the two WBA estimates:

$$\text{dig}_N = -\log_{10} |\Delta_N|. \quad (6)$$

This metric,  $\text{dig}_N$ , effectively measures the number of significant decimal digits to which the two estimates agree. For chaotic orbits, where  $|\Delta_N| \sim N^{-1/2}$ ,  $\text{dig}_N$  remain relatively small. For regular orbits, both  $WB_N(h)(\mathbf{x})$  and  $WB_N(h)(\mathbf{f}^N(\mathbf{x}))$  converge rapidly to the same limit, causing  $\Delta_N$  to decay quickly to 0; hence,  $\text{dig}_N$  becomes large and increases with  $N$ .

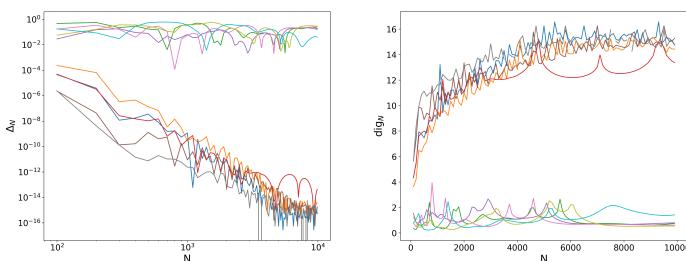


Figure 6: Left panel:  $\Delta_N$  as a function of the number of iterations  $N$  for 10 initial conditions. Right panel:  $\text{dig}_N$  as a function of  $N$  for the corresponding orbits.

Figure 6 plots  $\Delta_N$  against the number of iterations  $N$  for ten orbits in the GKR system. For chaotic orbits,  $\Delta_N$  remains largely unchanged as  $N$  increase, however for resonant or non-resonant orbits,  $\Delta_N$  decays rapidly. We can observe that there is a clear separation between chaotic and regular orbits for  $N \geq 10^4$ . Therefore, we choose  $N = 10^4$  and use  $\text{dig}_{10^4}$  as our classification metric.

Inspired by [27], we establish a threshold to distinguish orbits types. We performed a histogram analysis of  $\text{dig}_{10^4}$  for 5000 different initial conditions randomly selected from  $[0, 1] \times [0, 1]$ ,

as shown in Figure 7. The distribution of  $\text{dig}_{10^4}$  values reveals the concentrated intervals: a low-value region, a high-value region, and a transitional intermediate interval. We can consider the orbits in the low-value region to be strongly chaotic, and those in the high-value region to be regular. In fact, the Lyapunov exponents for orbits in the low-value region are significantly higher than those in the high-value region. Notably, within the transition region, Lyapunov exponents vary continuously. Some orbits show values close to zero (indicating regular orbits), others show high values (indicating strongly chaotic orbits), while another subset displays intermediate values. This phenomenon suggests the possible existence of weakly chaotic orbits, which warrants further investigation. Using *k-means* clustering on the  $\text{dig}_{10^4}$  values, we set thresholds  $C_1$  (left) and  $C_2$  (right) at the centroids of the low-value and high-value clusters. Orbit with  $\text{dig}_{10^4} < C_1$  are classified as strongly chaotic, and those with  $\text{dig}_{10^4} > C_2$  are classified as regular. Orbit falling between  $C_1$  and  $C_2$  (the transitional region) require further analysis.

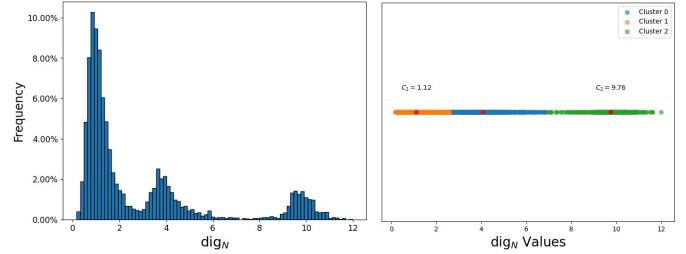


Figure 7: Left panel: Histogram of  $\text{dig}_{10^4}$  values for 5000 randomly selected initial conditions in the system Eq.(1) with  $K = 0.03, M = 3$ . Right panel: The corresponding clustering result of the  $\text{dig}_{10^4}$  data.

As an example, for  $K = 0.03$  and  $M = 3$ , the  $\text{dig}_{10^4}$  distribution (Figure 7) has the low-value and high-value centroids at  $C_1 = 1.12, C_2 = 9.76$ , respectively. We thus classify orbits with  $\text{dig}_{10^4} < 1.12$  as strongly chaotic (Figure 13a) and those with  $\text{dig}_{10^4} > 9.76$  as regular (Figure 13c). Orbit with values between these thresholds are labeled as unclassified (Figure 13b). This procedure can be generalized to any parameters  $K$  and  $M$  using Algorithm 1 to systematically identify strongly chaotic and regular orbits.

## 2.2. Lyapunov exponent

In dynamical systems, the Lyapunov exponent (LE) is a key method for measuring a system's sensitivity to initial conditions, i.e., the exponential rate of divergence (or convergence) between nearby orbits. The Lyapunov exponent of the map  $\mathbf{f}$  at the point  $\mathbf{x}_0$  along the direction  $\mathbf{v}_0$  is defined as:

$$\lim_{N \rightarrow \infty} \frac{\ln \|D\mathbf{f}^N(\mathbf{x}_0)\mathbf{v}_0\|}{N}. \quad (7)$$

In practice, for finite but sufficiently large  $N$ , we compute the finite-time Lyapunov exponent using the formula [24]:

$$\frac{\ln \|D\mathbf{f}^N(\mathbf{x}_0)\mathbf{v}_0\|}{N} = \frac{1}{N} \sum_{n=0}^{N-1} \ln \|\mathbf{v}_n\|, \quad \mathbf{v}_n = \frac{D\mathbf{f}(\mathbf{x}_{n-1})\mathbf{v}_{n-1}}{\|\mathbf{v}_{n-1}\|}. \quad (8)$$

---

**Algorithm 1** Stage I: Orbit Classification by weighted Birkhoff averaging

---

- 1: **Input:** Parameters  $K, M$ , and orbits  $\{(x_n, y_n)\}_{n=0}^{2 \times 10^4}$
- 2: **Output:** Regular I, Strongly chaotic I, and Unclassified I
- 3: **for** each orbit **do**
- 4:     Calculate the  $\text{dig}_{10^4}$  using (6)
- 5: **end for**
- 6: Determine the two thresholds  $C_1$  and  $C_2$  by *k-means* clustering
- 7: **for** each orbit **do**
- 8:     **if**  $\text{dig}_{10^4} < C_1$  **then**  
        Labeled as Strongly chaotic I
- 9:     **else if**  $\text{dig}_{10^4} > C_2$  **then**  
        Labeled as Regular I
- 10:     **else if**  $C_1 \leq \text{dig}_{10^4} \leq C_2$  **then**  
        Labeled as Unclassified I
- 11:     **end if**
- 12: **end for**

---

Here,  $D\mathbf{f}(\mathbf{x})$  denotes the Jacobian matrix of the map  $\mathbf{f}$  evaluated at  $\mathbf{x}$ . The sequence  $\{\mathbf{v}_n\}$  represents the time-evolved perturbation vector. The Lyapunov exponent serves as a fundamental tool for characterizing chaotic behavior, providing crucial insights into the system's long-term evolution and predictability.

Note that for the GKR system (1), the Jacobian matrix

$$D\mathbf{f}(\mathbf{x}) = \begin{bmatrix} 1 + K \sum_{n=1}^M 2\pi n \cos(2\pi n x) & 1 \\ K \sum_{n=1}^M 2\pi n \cos(2\pi n x) & 1 \end{bmatrix}$$

has a determinant  $\det D\mathbf{f}(\mathbf{x}_n) \equiv 1$ , implying area preservation in phase space. In order to compute the Lyapunov exponents, we choose two linearly independent initial perturbation directions, typically  $\mathbf{v}_0^{(1)} = (1, 0)$ ,  $\mathbf{v}_0^{(2)} = (0, 1)$ . Then, using Eq.(8) we obtain two Lyapunov exponents, denoted as  $\lambda_1$  and  $\lambda_2$ . According to [28], the sum of the Lyapunov exponents satisfies

$$\lambda_1 + \lambda_2 = \lim_{N \rightarrow \infty} \frac{1}{N} \ln |\det D\mathbf{f}^N(\mathbf{x})| = 0,$$

where the term  $|\det D\mathbf{f}^N(\mathbf{x})|$  quantifies the cumulative expansion or contraction of an infinitesimal phase space volume over  $N$  iterations. Since  $\lambda_1, \lambda_2$  are equal in magnitude and opposite in sign, we only need to consider either  $|\lambda_1|$  or  $|\lambda_2|$  (denoted hereafter as  $|\lambda|$ ). In general dynamical systems, regular orbits typically correspond to all Lyapunov exponents being non-positive, while chaotic orbits are characterized by at least one positive Lyapunov exponent. Weakly chaotic orbits may exhibit Lyapunov exponents approaching zero. However, since  $f$  is an area-preserving map, both regular and weakly chaotic orbits may yield Lyapunov exponents near zero. To enhance the discriminative power of the visualization, we apply a logarithmic transformation to the Lyapunov exponents, thereby amplifying hierarchical differences in their magnitudes.

We compute  $\log_{10} |\lambda|$  for orbits originating from 5000 random initial conditions in  $[0, 1] \times [0, 1]$ , after  $10^4$  iterations. Fig-

ure 8 shows the corresponding histograms for different parameter sets: the top row for  $M = 3$  with varying  $K$ , and the bottom row for  $M = 5$ . For  $K = 0.01$  and  $M = 3$ , the distribution of  $\log_{10} |\lambda|$  is primarily concentrated around  $-3$ , accompanied by an incipient secondary mound in the  $[-2, -1]$  interval. As  $K$  increases, the peak

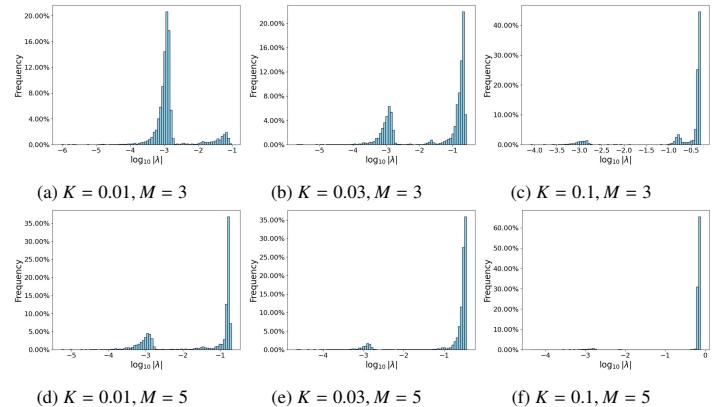


Figure 8: Histograms of  $\log_{10} |\lambda|$  for orbits of Eq.(1) with different  $K, M$ . The initial conditions are 5000 values randomly selected from the  $[0, 1] \times [0, 1]$ .

of  $\log_{10} |\lambda|$  around  $-3$  gradually weakens, while the feature around  $-1$  grows significantly, eventually forming a dominant peak. At  $K = 0.03$ , a lower peak centered near  $-3$  coexists with a dominant peak near  $-0.8$ , with clear separation between the two modes. As  $K$  increases to  $K = 0.1$ , the system becomes predominantly chaotic, with a single prominent peak. A similar evolutionary pattern is observed for  $M = 5$  as  $K$  increases, demonstrating that the transition from regular to chaotic dynamics is consistent for different numbers of harmonics.

Based on the preceding analysis, the value of  $\log_{10} |\lambda|$  exhibit a consistent pattern: those of regular orbits concentrate on the left side of the histogram, while values from strongly chaotic orbits cluster on the right. This clear separation in  $\log_{10} |\lambda|$  between the two previously identified groups provides a basis for further classifying the orbits in the Unclassified I category. To leverage this, we developed Algorithm 2 as an extension of Algorithm 1. The procedure is as follows. We first compute  $\log_{10} |\lambda|$  for the orbits already classified as Regular I and Strongly Chaotic I. From these, we define two thresholds:

- $\text{LE}_{\text{reg}}$ : a value exceeds  $\log_{10} |\lambda|$  for over 95% of the Regular I orbits (i.e.,  $\log_{10} |\lambda| < \text{LE}_{\text{reg}}$ ).
- $\text{LE}_{\text{chaos}}$ : a value that is exceeded by  $\log_{10} |\lambda|$  for over 95% of the Strongly chaotic I orbits (i.e.,  $\log_{10} |\lambda| > \text{LE}_{\text{chaos}}$ ).

These thresholds are then applied to the Unclassified I orbits. Those satisfying  $\log_{10} |\lambda| < \text{LE}_{\text{reg}}$  are labeled as Regular II, as their Lyapunov exponents are consistent with Regular I group. Orbits satisfying  $\log_{10} |\lambda| > \text{LE}_{\text{chaos}}$  are labeled Strongly chaotic II, indicating dynamics similar to Strongly chaotic I.

For the specific case with  $K = 0.03$  and  $M = 3$ , the orbits pre-classified in Algorithm 1 exhibit distinct Lyapunov

---

**Algorithm 2** Stage II: Identify orbits in Unclassified I by Lyapunov exponents

---

```

1: Input: Parameters  $K, M$ , and orbits classified by Stage I
2: Output: Regular II, Strongly chaotic II, and Unclassified
II
3: for Regular I and Strongly chaotic I do
4:   Calculate the  $\log_{10} |\lambda|$ 
5: end for
6: Determine the thresholds  $LE_{\text{reg}}$  and  $LE_{\text{chaos}}$  such that:
   For orbits in Regular I:  $\Pr(LE_{\text{reg}} > \log_{10} |\lambda|) \geq 0.95$ 
   For orbits in Strongly chaotic I:  $\Pr(LE_{\text{chaos}} < \log_{10} |\lambda|) \geq 0.95$ 
7: for Orbit in Unclassified I do
8:   Calculate the  $\log_{10} |\lambda|$ 
9:   if  $\log_{10} |\lambda| < LE_{\text{reg}}$  then
10:    Labeled as Regular II
11:   else if  $\log_{10} |\lambda| > LE_{\text{chaos}}$  then
12:    Labeled as Strongly chaotic II
13:   else if  $LE_{\text{reg}} \leq \log_{10} |\lambda| \leq LE_{\text{chaos}}$  then
14:    Labeled as Unclassified II
15:   end if
16: end for

```

---

exponent distributions. As shown in Figure 9, regular orbits ( $\text{dig}_{10^4} > 9.76$ ) typically have  $\log_{10} |\lambda| < -3.1$ , with over 95% of orbits falling below this value. In contrast, strongly chaotic orbits ( $\text{dig}_{10^4} < 1.12$ ) are characterized by  $\log_{10} |\lambda| > -1.1$ , a threshold exceeded by more than 95% of orbits in this category. Therefore, we conclude that orbits in the ambiguous range  $1.12 \leq \text{dig}_{10^4} \leq 9.76$  that satisfy  $\log_{10} |\lambda| < -3.1$  are classified as regular, while those with  $\log_{10} |\lambda| > -1.1$  are classified as strongly chaotic. As shown in Figure 13, we further classify some orbits from the Unclassified I into strongly chaotic and regular orbits, as shown in Figures 13d and 13f, while a portion remains unclassified, as shown in Figure 13e.

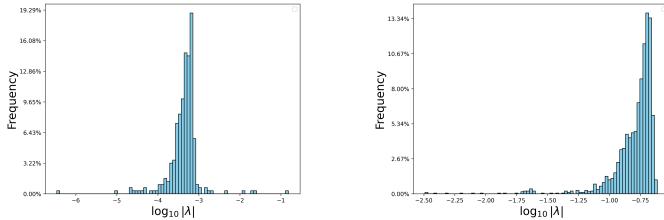


Figure 9: Histograms of  $\log_{10} |\lambda|$  for orbits with  $\text{dig}_{10^4} > 9.76$  (left), and  $\text{dig}_{10^4} < 1.12$  (right).

### 2.3. Correlation dimension

The Correlation dimension (CD) is a fundamental measure in nonlinear time series analysis that quantifies the geometric complexity of a point set in a metric space by examining the statistics of pairwise distances. It serves as a powerful tool for characterizing the underlying structure of dynamical systems, particularly for assessing the effective degrees of freedom in complex systems and for identifying the fractal structure of

strange attractors. The Grassberger-Procaccia (GP) algorithm is the most widely used method for computing the correlation dimension, especially for determining the fractal dimensions of strange attractors from observed time series data.

According to [29][30], for Eq.(1), given an initial point  $\mathbf{x}_0$ , iterating  $N$  times yields an orbit from which the correlation sum  $C(N, r)$  can be computed as:

$$C(N, r) = \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} H(r - \|\mathbf{x}_i - \mathbf{x}_j\|), \quad (9)$$

where  $H(\cdot)$  is the Heaviside step function and  $\|\mathbf{x}_i - \mathbf{x}_j\|$  denotes the Euclidean distance between embedded vectors. As  $r \rightarrow 0$ ,  $C(N, r)$  decreases monotonically to zero. If  $C(N, r)$  follows a power law  $C(N, r) \sim r^D$  in the limits  $N \rightarrow +\infty, r \rightarrow 0$ , then  $D$  is referred to as the correlation dimension estimate of the orbit. This dimension is mathematically expressed as:

$$D = \lim_{r \rightarrow 0} \lim_{N \rightarrow +\infty} \frac{\log C(N, r)}{\log r}. \quad (10)$$

The theoretical limits in Eq.(10) are not attainable in practice with finite data. Therefore, we adopt the following numerical method to estimate  $D$ . We generate orbits of length  $N$  from 30 randomly selected initial conditions. To determine a suitable orbit length  $N$ , we examine the behavior of the correlation integral  $C(N, r)$  as a function of  $N$  for a fixed radius  $r$ , with  $r$  spanning the interval  $[10^{-6}, 10^0]$ . As shown in Figure 10, a similar trend was observed:  $C(N, r)$  plateaus and becomes independent of  $N$  for  $N \geq 10^4$ . Consequently, we set  $N = 10^4$  for further computations. With  $N$  fixed, we then evaluate  $\log C(10^4, r)$  against  $\log r$  over logarithmically spaced radii in the interval  $[10^{-6}, 10^0]$ . This range was chosen to cover the scaling region while avoiding saturation effects and noise, following the methodology of [30, 2]. The correlation dimension  $D$  was estimated from the slope of the linear region in the plot of  $\log C(10^4, r)$  versus  $\log r$ . Figure 10 presents scatter plots of orbits pre-classified as strongly chaotic and regular. The analysis reveals that regular orbit possess a lower correlation dimension than strongly chaotic orbit, aligning with the difference in their underlying dynamical complexity.

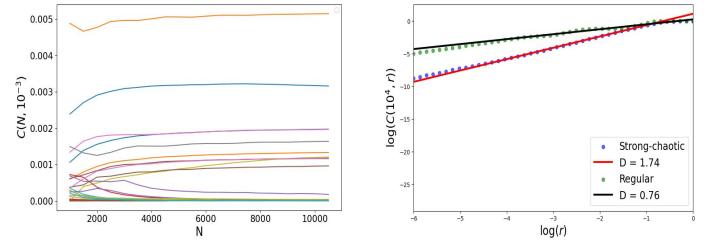


Figure 10: Left panel:  $C(N, 10^{-3})$  as a function of the number of iterations  $N$  for 30 initial conditions. Right panel: Scatter plot of  $\log C(10^4, r)$  against  $\log r$  for orbits that were pre-classified into strongly chaotic and regular types. The black and red lines represent the fitted lines.

To better visualize variations in the correlation dimension across a larger sample, we randomly select 400 initial conditions within  $[0, 1] \times [0, 1]$  for different parameters  $K$  and  $M$ . The

corresponding distributions of  $D$  are visualized using heatmaps. As shown in Figure 11, these heatmaps illustrate the distribution of  $D$  for specific  $K$  and  $M$ , which consistently align with the distinct trajectory patterns previously identified in Figure 1. This finding demonstrates that the correlation dimension  $D$ , as a fundamental measure of set complexity, serves as an effective quantitative descriptor for distinguishing between different types of dynamical behaviors.

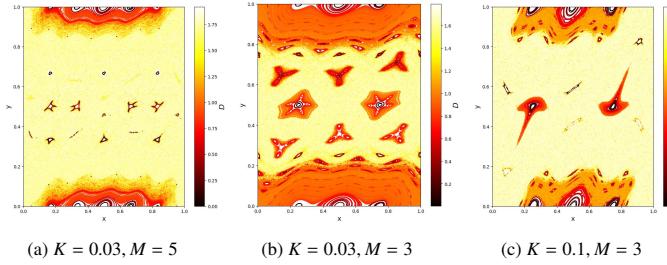


Figure 11: Correlation dimension heatmaps for Eq.(1), computed from 400 orbits with initial conditions randomly sampled from the  $[0, 1] \times [0, 1]$ .

The preceding analysis reveals a consistent pattern in the correlation dimension  $D$ : it is significantly larger for strongly chaotic orbits than for regular ones. This pattern provides a quantitative basis for further classifying orbits in the Unclassified II category. To implement this, we developed Algorithm 3. The procedure begins by computing  $D$  for orbits already categorized as Regular I  $\cup$  II and Strongly Chaotic I  $\cup$  II. Using these results, we define two thresholds:

- $D_{\text{reg}}$ : a value exceeds  $D$  for over 65% of the Regular I  $\cup$  II orbits (i.e.,  $D < D_{\text{reg}}$ ).
- $D_{\text{chaos}}$ : a value that is exceeded by  $D$  for over 65% of the Strongly chaotic I  $\cup$  II orbits (i.e.,  $D > D_{\text{chaos}}$ ).

These thresholds are then applied to the Unclassified II orbits for classification:

- Orbit with  $D > D_{\text{chaos}}$  are labeled as Strongly chaotic III, as their correlation dimensions align with the characteristic high values of the known strongly chaotic orbits.
- Orbit with  $D < D_{\text{reg}}$  are labeled as Regular III, indicating dynamics similar to the known regular orbits.
- Orbit with  $D_{\text{reg}} \leq D \leq D_{\text{chaos}}$  are labeled as Weakly chaotic. This intermediate classification is assigned to orbits whose behavior, as characterized by the synergistic integration of the three complementary methods— $\text{dig}_{10^4}$ ,  $\log_{10} |\lambda|$ , and  $D$ —is neither unequivocally strongly chaotic nor regular.

---

**Algorithm 3 Stage III: Identify orbits in Unclassified II by correlation dimensions**

---

```

1: Input: Parameters  $K, M$ , and orbits classified by Stage II
2: Output: Regular, Strongly chaotic and Weakly chaotic orbits
3: for Regular I  $\cup$  II and Strongly chaotic I  $\cup$  II do
4:   Calculate  $D$ 
5: end for
6: Determine the thresholds  $D_{\text{reg}}$  and  $D_{\text{chaos}}$  such that:
   For orbits in Regular I  $\cup$  II:  $\Pr(D_{\text{reg}} > D) \geq 0.65$ 
   For orbits in Strongly chaotic I  $\cup$  II:  $\Pr(D_{\text{chaos}} < D) \geq 0.65$ 
7: for Orbits in Unclassified II do
8:   Calculate  $D$ 
9:   if  $D < D_{\text{reg}}$  then
10:    Labeled as Regular III
11:   else if  $D > D_{\text{chaos}}$  then
12:    Labeled as Strongly chaotic III
13:   else if  $D_{\text{reg}} \leq D \leq D_{\text{chaos}}$  then
14:    Labeled as Weakly chaotic orbits
15:   end if
16: end for

```

---

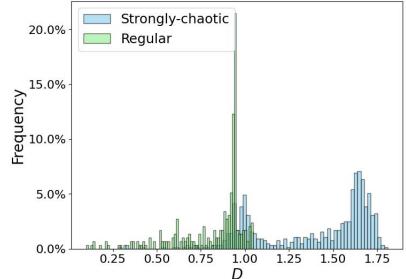


Figure 12: Histogram of the correlation dimension  $D$ . Skyblue and lightgreen bars represent the distributions for Strongly chaotic I  $\cup$  II and Regular I  $\cup$  II, respectively.

For the specific case with  $K = 0.03$  and  $M = 3$ , the orbits pre-classified in Algorithm 1 2 exhibit distinct correlation dimension distributions. As shown in Figure 12, more than 65% of Strongly chaotic I  $\cup$  II orbits have  $D > 1.4$ , while more than 65% of Regular I  $\cup$  II orbits have  $D < 0.93$ . Based on these thresholds, orbits in the Unclassified II set are classified as follows: those with  $D > 1.4$  are classified as Strongly chaotic III, while those with  $D < 0.93$  are classified as Regular III, and the remaining orbits are classified as Weakly chaotic. As visually summarized in Figures 13g, 13h, and 13i, this multi-stage hierarchical strategy successfully segregates all orbits into three distinct categories: Strongly Chaotic, Weakly Chaotic, and Regular.

#### 2.4. Identification of resonant invariant torus orbits

So far, we have categorized the orbits into three broad types—Strongly chaotic, Weakly chaotic, and Regular—using Algorithms 1, 2, and 3. We now focus on the regular orbits,

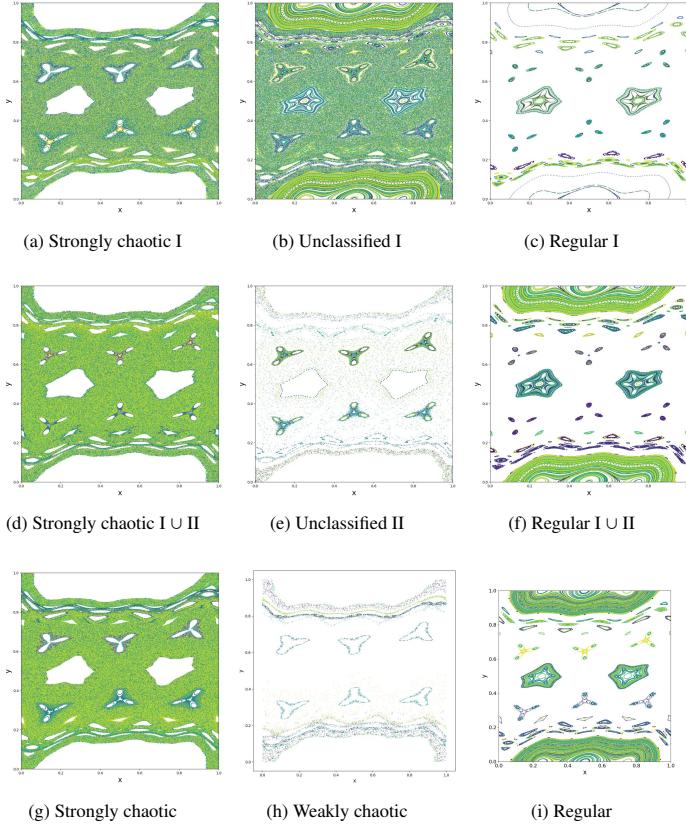


Figure 13: For the case with  $K = 0.03$  and  $M = 3$ , the above algorithms progressively classify the orbits into: Strongly chaotic, Weakly chaotic, and Regular.

which are further subdivided into resonant and non-resonant orbits using Algorithm 4.

Regular orbits consist of two topologically distinct types: resonant orbits and non-resonant orbits. A central quantity for characterizing these orbits is the rotation number  $\omega$ , defined for any initial condition on a regular orbit as:

$$\omega = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} y_{n+1}. \quad (11)$$

This quantity quantifies the average angular rate of motion. If  $\omega$  satisfies a strong irrationality condition, such as the Diophantine condition, the corresponding orbit is non-resonant, and its persistence under small perturbations is assured by the KAM theorem. Conversely, if  $\omega$  is sufficiently well-approximated by rational numbers, the orbit is resonant, a structure typically destroyed by perturbations.

To numerically distinguish between these cases, we employ a method that assesses, with high probability, whether a floating-point computed  $\omega$  approximates a rational or an irrational number [27]. This method defines an indicator,  $\text{dev}_\omega$ , for each computed rotation number. If  $\text{dev}_\omega > s$ , the rotation number is identified as approximating a rational number, indicating that the corresponding orbit is resonant; otherwise, it is considered an irrational approximation, meaning the orbit is non-resonant. In our implementation, the threshold is set to  $s = 0.67$  (see Appendix A for details). Applying Algorithm 4 with parameters

$K = 0.03$  and  $M = 3$  successfully categorizes the orbits within the regular set into resonant and non-resonant types. Thus, we achieve a comprehensive four-fold classification of orbits, establishing a systematic framework that distinguishes strongly chaotic, weakly chaotic, resonant, and non-resonant orbits.

---

**Algorithm 4** Stage IV: Classify regular orbits by rotation number

---

```

1: Input: Regular orbits
2: Output: resonant orbits and non-resonant orbits
3: for orbits in Regular do
4:   Compute  $\text{dev}_\omega$ 
5:   if  $\text{dev}_\omega > 0.67$  then
6:     Labeled as resonant orbit
7:   else if  $\text{dev}_\omega \leq 0.67$  then
8:     Labeled as non-resonant orbit
9:   end if
10: end for
```

---

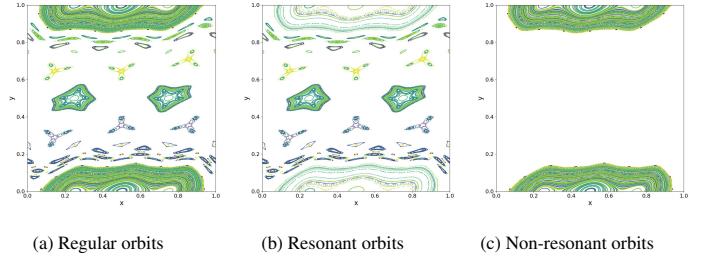


Figure 14: The regular orbits in the case of  $K = 0.03$ ,  $M = 3$  are further classified into resonant orbits and non-resonant orbits.

### 3. Deep Learning classification

This section outlines the generation of high-quality labeled data for deep learning, specifies which deep learning models were selected, explains the training process used for orbit classification, and provides a comparative analysis of their classification performance. MLP is typically employed as the baseline model. InceptionTime CNN architecture has demonstrated strong classification performance in time-series classification tasks, as evidenced by previous related studies. Furthermore, we implement hybrid variants of the MLP and InceptionTime CNN that integrate Transformer architectures, resulting in the MLP-Transformer and InceptionTime CNN-Transformer models. The incorporation of Transformers is motivated by their self-attention mechanism, which excels at capturing complex long-range dependencies inherent in time-series data. Finally, we employ a 2D-CNN model that converts sequential orbital data into image-based representations. This approach enables the model to leverage the inherent spatial patterns in trajectory images for accurate orbit type identification. Among the models evaluated on the same dataset, the 2D-CNN demonstrates the most effective and stable classification performance.

### 3.1. Details about the Data and Model

**Datasets:** To generate high-quality labeled data for deep learning, we first evolve 1000 orbits from given initial conditions over long-term iterations. Each orbit is algorithmically assigned a definitive label: '0' for strongly chaotic orbits, '1' for resonant orbits, '2' for non-resonant orbits, and '3' for weakly chaotic orbits. These reliable labels are then paired with corresponding shorter 900-point orbital segments to construct the dataset, which is partitioned into training (64%), validation (16%), and test (20%) sets. This strategy leverages long-term evolution for robust labeling and uses shorter segments as model inputs, effectively balancing label reliability with data diversity to establish a scalable and practical foundation for deep learning-based classification.

**MLP and MLP-Transformer:** In classification tasks, MLP is typically employed as a benchmark model, trained on our specialized dataset derived from orbital trajectories. Each sample is constructed by concatenating two feature sets: a down-sampled time series, created by selecting every 10th point from the original 900-length orbit and then flattened; and a vector of statistical descriptors, including the mean, minimum, maximum, variance, correlation coefficient, and average density. Our MLP architecture comprises an input layer, four fully-connected hidden layers (each containing 500 neurons), and an output layer with four neurons corresponding to the classification categories. Hidden layers utilize the ReLU activation function, batch normalization and dropout regularization ( $p = 0.1$ ). The output layer employs a softmax function to generate class probabilities. The training process employs the AdamW optimizer with a weight decay coefficient of  $10^{-5}$  and gradient clipping. The learning rate is scheduled using the OneCycle policy with a peak learning rate of 0.01. The loss function is cross-entropy, and the model is trained for 500 epochs. During training, we apply updates in batches of 64. The model achieving the highest validation accuracy is saved for final evaluation on the test set. Building upon this MLP baseline, we design an MLP-Transformer hybrid architecture that processes the input through dual parallel pathways. In the Transformer pathway, the input is reshaped into a sequence and augmented with sinusoidal positional encoding, then processed by a 3-layer encoder with 8-headed self-attention, followed by adaptive average pooling, while the MLP pathway consists of two 256-neuron layers. Both pathways employ batch normalization, ReLU activation, and dropout. Their respective representations are concatenated and processed through three fusion layers (512, 256, and 4 neurons) utilizing identical regularization techniques. This hybrid model maintains the same optimization strategies and training protocols as the MLP baseline.

**InceptionTime CNN and InceptionTime CNN-Transformer:** Following the approach in [12], which applies the InceptionTime CNN architecture to classify regular and chaotic motions in the forced pendulum and the spin-orbit systems, our work employs a standard InceptionTime CNN model with six Inception modules. The network input is a two-channel time series representing the temporal evolution of the x- and y-coordinates over 900 iterations. Each Inception module begins with a bottleneck layer, followed by parallel

branches: three convolutional layers with kernel sizes of 39, 19, and 9, and one max-pooling branch that is subsequently processed by  $1 \times 1$  convolution. The outputs of these branches are concatenated, processed through batch normalization, and activated via the ReLU function. The final feature representation is subjected to global average pooling before being passed to a fully connected layer for classification. The model is trained using the cross-entropy loss function and the Adam optimizer with a learning rate of 0.0008. Training proceeds for 100 epochs with a batch size of 32, and the model weights yielding the highest validation accuracy are retained for final evaluation on the test dataset. To enhance the capture of long-range temporal dependencies and refine feature representation, we further introduce an augmented architecture—denoted as InceptionTime CNN-Transformer—by integrating a Transformer encoder after the Inception modules. The feature maps extracted by the six Inception modules are reshaped and augmented with sinusoidal positional encodings, which inject order information using sine and cosine functions of varying frequencies. The resulting sequence is processed by a two-layer Transformer encoder employing a four-head self-attention mechanism. The output is then pooled globally and passed to the classification layer. The training protocol for this InceptionTime CNN-Transformer hybrid model remains identical to that of the standard InceptionTime CNN described above

**2D-CNN:** We design a CNN architecture employing two-dimensional convolutional layers, hereinafter referred to as the 2D-CNN model. The network accepts as input a visual encoding of orbits: each trajectory, with a length of 900 points sampled from the phase space  $[0, 1] \times [0, 1]$ , is encoded as a  $64 \times 64$  pixel scatter plot, where the orbit is uniquely colored according to its pre-assigned class label. This transformation allows the model to leverage spatial feature hierarchies inherent in the visual representation rather than raw coordinate sequences. The model is structured into a feature extraction module followed by a classification module. The feature extractor consists of three convolutional blocks. Each block contains a convolutional layer with a  $3 \times 3$  kernel, stride 1, followed by a ReLU activation function and a  $2 \times 2$  max-pooling operation. The number of channels increases across the blocks—32 in the first, 64 in the second, and 128 in the third. The resulting feature maps are then flattened into a one-dimensional vector. The classification module comprises two fully connected layers, each with 512 neurons and followed by ReLU activation and dropout regularization with a rate of 0.5. The final output layer produces logits for the four classes, which are passed through a softmax function to generate class probabilities. The model is trained using the Adam optimizer with a learning rate of 0.0008 and the cross-entropy loss function. Training employs mini-batches of 32 for up to 100 epochs, with early stopping implemented at a patience of 5. The model achieving the highest validation accuracy is saved and evaluated on the test dataset to assess classification performance.

### 3.2. Deep Learning classification results

Under the experimental configurations detailed for the aforementioned five models, we evaluate the classification performance of each model with varying parameters  $K$  and  $M$ —consistent with the parameters outlined in Figure 1—using identical training and testing datasets. The results, summarized in Table 1, reveal that the 2D-CNN model demonstrates exceptional performance, consistently achieving classification accuracy exceeding 99%. In contrast, the other four models exhibit significant performance fluctuations and failed to reach the same level of high, stable accuracy. The substantial performance gap suggests that the 2D-CNN is inherently well-suited for capturing spatially local patterns, which is particularly advantageous for this classification task.

Models \ Parameters	K=0.01,M=3	K=0.03,M=3	K=0.1,M=3
<b>MLP</b>	80.50% $\pm$ 0.50%	82.50% $\pm$ 1.50%	90.50% $\pm$ 0.50%
<b>MLP-Transformer</b>	88.00% $\pm$ 0.50%	84.00% $\pm$ 0.50%	94.00% $\pm$ 0.50%
<b>InceptionTime-CNN</b>	91.00% $\pm$ 0.50%	90.00% $\pm$ 0.30%	92.25% $\pm$ 0.25%
<b>InceptionTime-CNN-Transformer</b>	92.00% $\pm$ 0.50%	91.00% $\pm$ 0.25%	94.75% $\pm$ 0.25%
<b>2D-CNN</b>	<b>99.50%<math>\pm</math>0.25%</b>	<b>99.45%<math>\pm</math>0.45%</b>	<b>99.50%<math>\pm</math>0.25%</b>

Models \ Parameters	K=0.01,M=5	K=0.03,M=5	K=0.1,M=5
<b>MLP</b>	84.50% $\pm$ 0.50%	91.00% $\pm$ 1.25%	97.00% $\pm$ 1.00%
<b>MLP-Transformer</b>	86.25% $\pm$ 0.50%	94.00% $\pm$ 0.75%	97.50% $\pm$ 0.50%
<b>InceptionTime-CNN</b>	91.50% $\pm$ 0.25%	94.00% $\pm$ 0.50%	97.25% $\pm$ 0.25%
<b>InceptionTime-CNN-Transformer</b>	92.00% $\pm$ 0.75%	96.50% $\pm$ 0.75%	97.50% $\pm$ 0.50%
<b>2D-CNN</b>	<b>99.50%<math>\pm</math>0.50%</b>	<b>99.50%<math>\pm</math>0.50%</b>	<b>99.38%<math>\pm</math>0.31%</b>

Table 1: Comparison of classification accuracy across the five DL models with different parameters.

To further validate the predictive reliability of the 2D-CNN model, we visualize trajectory diagrams for the four orbit types using the model’s predicted labels, as shown in Figure 15. These predicted trajectories exhibit near-perfect agreement with the trajectories displayed in Figures 13g, 13h, 14b, and 14c. This close visual correspondence demonstrates that the 2D-CNN model accurately captures the underlying dynamical structures, affirming its high fidelity in identifying the essential features of different orbital types.

Although the dataset is largely reliable, it contains a small proportion of mislabeled orbits. The 2D-CNN model successfully reclassifies these erroneous cases into their appropriate dynamical categories. Table 2 provides a comparative summary of orbit-type distributions and misclassification counts for the numerical algorithm and the deep learning model. For instance, the numerical algorithm initially identified 137 orbits; however, validation using trajectory plots generated with longer iterations confirmed that 2 of these were misclassified. In contrast, the 2D-CNN model classified 133 orbits as resonant, with only one misclassification. To illustrate the corrective capability of the 2D-CNN, Figure 16a displays specific instances where orbits initially misclassified by the numerical algorithm were correctly identified by the deep learning model. For example, Orbit **b** was originally labeled as a resonant orbit by the

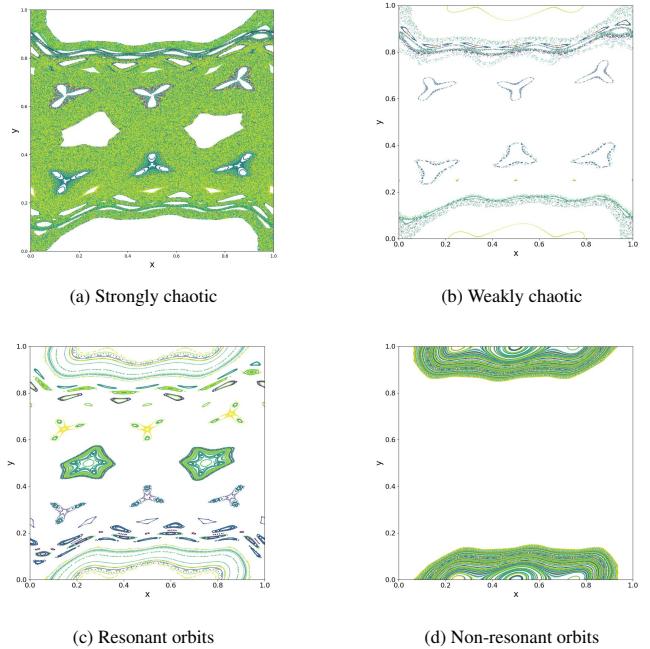


Figure 15: Trajectory diagrams of the four orbit types classified by the 2D-CNN model.

numerical algorithm but was accurately reclassified as a non-resonant orbit by the 2D-CNN. In contrast, Figure 16b presents cases where the 2D-CNN misclassified orbits that were originally correct in the algorithmic labeling. For instance, Orbit **e**, correctly categorized as a resonant orbit by the numerical algorithm, was erroneously classified as a weakly chaotic orbit by the 2D-CNN. This phenomenon underscores that the deep learning model can capture fundamental patterns even when reference labels are wrong. Nevertheless, high-quality labeled data remains indispensable for reliably distinguishing challenging borderline cases, which are particularly susceptible to misclassification.

	Numerical algorithm	Classified incorrectly	2D-CNN	Classified incorrectly
<b>Class ‘0’</b>	669	0	671	2
<b>Class ‘1’</b>	137	2	133	1
<b>Class ‘2’</b>	177	1	177	0
<b>Class ‘3’</b>	17	0	20	3
<b>Total</b>	1000	3	1000	6

Table 2: A comparative overview of the orbit counts for each dynamical type, obtained through the numerical algorithm and predicted by the trained 2D-CNN.

To systematically evaluate the classification performance of our 2D-CNN model across a continuous parameter space rather than at isolated points, we expanded our analysis to the intervals  $K \in [0.005, 0.3]$  and  $M \in [2, 6]$ . For each integer value of  $M$  in this range, ten values of  $K$  were randomly sampled, resulting in a total of 50 distinct parameter pairs. For every resulting parameter pair  $(K, M)$ , we first generated the dynamical labels using

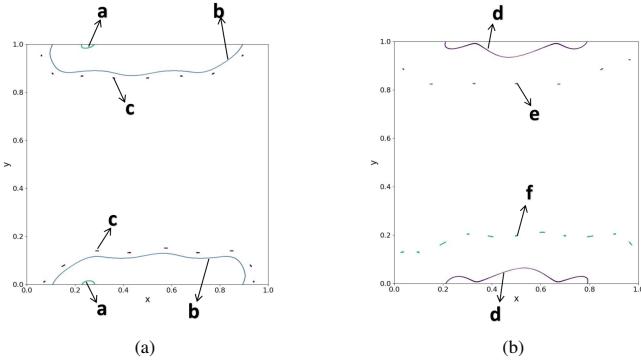


Figure 16: (a)The numerical algorithm-based method misclassified the orbit, whereas the 2D-CNN accurately determined its type. (b)The numerical algorithm-based method correctly classified the orbit, whereas the 2D-CNN misidentified the orbit type.

our numerical algorithm. These labels were then paired with the corresponding 900-length orbital segments to form the individual data subsets. Finally, all subsets from the different parameter pairs were combined into a single dataset for deep learning. To assess the model’s performance on unseen parameter combinations within the same operational domain, we selected six distinct pairs from these intervals and evaluated each in a separate test run, ensuring no overlap with the training pairs. As illustrated in Figure 17, the model achieved exceptional classification accuracy, surpassing 99%. Moreover, as shown in Figure 18, it maintained high performance when evaluated on parameter pairs outside the training intervals (specifically,  $K = 0.03, M = 10$ ;  $K = 0.5, M = 3$ ; and  $K = 0.001, M = 3$ ), confirming its robust generalization capability. These results indicate that the model captures underlying dynamical patterns consistent across the parameter space, rather than overfitting to specific training examples.

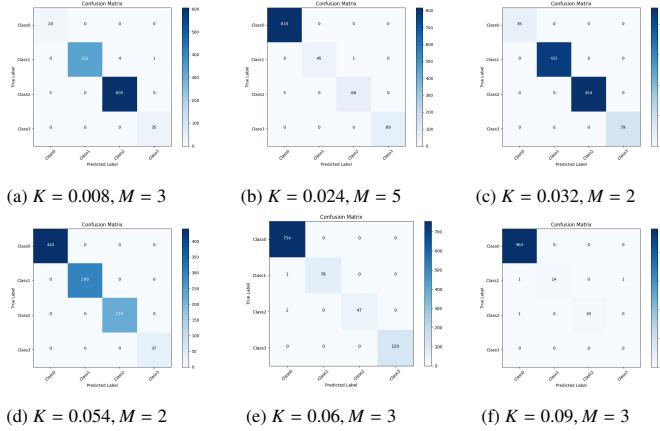


Figure 17: Classification performance of the 2D-CNN model evaluated under six distinct parameter pairs.

#### 4. Conclusion

Our study presents a novel framework for the granular classification of orbital dynamics within the GKR system, success-

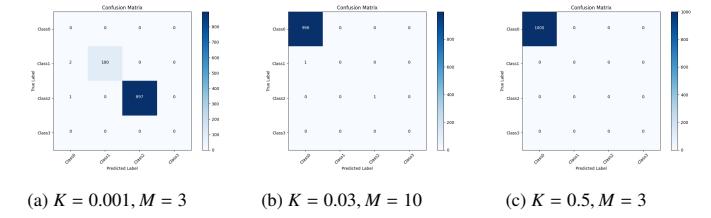


Figure 18: Generalization performance of the 2D-CNN on parameters outside the training domain.

fully distinguishing four distinct categories: strongly chaotic, weakly chaotic, resonant, and non-resonant orbits. A key innovation lies in the synergistic integration of three complementary methods—the weighted Birkhoff average, Lyapunov exponent, and correlation dimension—which enables robust identification, particularly of the elusive weak chaos regime where single-metric approaches often fail. This foundational work facilitated the creation of a high-quality, algorithmically-labeled dataset pairing long-term integration labels with shorter trajectory segments, effectively balancing classification accuracy with practical applicability for deep learning.

Subsequently, we demonstrated that a well-trained 2D-CNN architecture excels in this classification task, significantly outperforming models including MLP, InceptionTime CNN, and their hybrid Transformer variants (MLP-Transformer and InceptionTime CNN-Transformer). The superior performance is attributed to the network’s ability to capture the inherent spatial patterns in the trajectory data. Moreover, the 2D-CNN exhibited a corrective capability, refining the initial algorithmic labels by rectifying some misclassifications, which underscores its ability to capture underlying dynamical principles beyond the training labels. Additionally, the model demonstrates considerable generalization capability across varied dynamical regimes, confirming its robustness. To our knowledge, this work represents the first successful application of machine learning to classify a dynamical system into four distinct orbital categories, providing an intuitive visualization that enhances the understanding of weakly chaotic orbits.

#### 5. Acknowledge

This work was supported by the National Natural Science Foundation of China (Grant No. 12271400 to Z. Xu) and the Science and Technology Development Plan Project of Jilin Province (Grant No. 20250102016JC to J. Zu).

#### Appendix A. Numerical identification of rational numbers

Given a rotation number  $\omega \in [0, 1]$ , and an interval  $I_\delta(\omega)$  for a small  $\delta$ :

$$I_\delta(\omega) \equiv (\omega - \delta, \omega + \delta). \quad \delta = 10^{-tol}$$

We denote the smallest denominator for a rational in an interval  $I_\delta(\omega)$  by:

$$q_{\min}(I_\delta) \equiv \min \left\{ q \in \mathbb{N} : \frac{p}{q} \in I_\delta, p \in \mathbb{Z} \right\}.$$

[27, Algorithm 2] provides a numerical method for finding  $q_{\min}$ . Applying this algorithm to  $10^4$  randomly chosen floating-point numbers in  $(0, 1)$  with uniform distribution, we compute  $q_{\min}$  and plot the histogram of  $\log_{10}(q_{\min})$  (see Figure A.19a).

We know that if there exists a rational number  $p/q \in I_\delta(\omega)$  with a sufficiently small denominator  $q$ , then  $\omega$  is well-approximated by this rational. Conversely, if all such rationals within  $I_\delta(\omega)$  have large denominators, we would typically expect  $\omega$  to approximate an irrational number. However, if the denominator  $q$  is excessively large, it may indicate that  $\omega$  is more likely an approximation of a rational number that narrowly missed being within the interval. To quantify whether  $q$  is small, large, or excessively large, [27] introduces a quantitative indicator  $\text{dev}_\omega$ :

$$\text{dev}_\omega = |\log_{10}(q_{\min}(I_\delta(\omega))) - \text{tol}/2|$$

We classify rotation numbers based on this indicator, if  $\text{dev}_\omega > s$ , then denominator  $q$  is either small or excessively large, and we consider the rotation number to more closely approximate a rational. Otherwise,  $\omega$  is deemed closer to an irrational. In this paper, we set  $\delta = 10^{-6}$  and  $s = 0.67$ , which implies 95.65% of the randomly sampled values,  $q_{\min}$  fall within the interval  $[10^{3-s}, 10^{3+s}]$ , as shown in Figure A.19b.

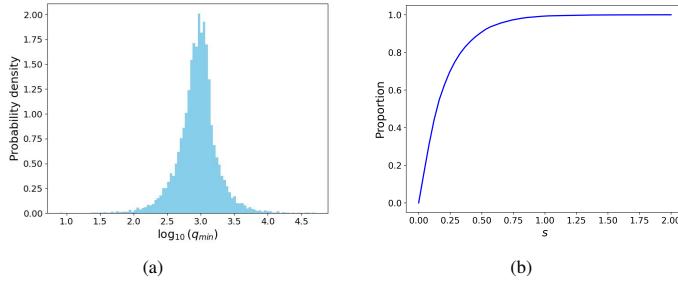


Figure A.19: (a) Probability density of  $\log_{10}(q_{\min})$  with  $\delta = 10^{-6}$  for  $10^4$  randomly chosen floating-point numbers in  $[0, 1]$ . (b) A graph of the probability that  $q_{\min} \in [10^{3-s}, 10^{3+s}]$ .

## References

- [1] R. De la Llave, A tutorial on KAM theory, American Mathematical Society, 2001.
- [2] J. P. Eckmann, D. Ruelle, Ergodic theory of chaos and strange attractors, *Reviews of Modern Physics* 57 (1985) 617.
- [3] E. Ott, Chaos in dynamical systems, Cambridge University Press, 2002.
- [4] R. Klages, Weak chaos, infinite ergodic theory, and anomalous dynamics, in: From Hamiltonian Chaos to Complex Systems: A Nonlinear Physics Approach, Springer, 2013, pp. 3–42.
- [5] Y. Sinai, Kolmogorov-sinai entropy, *Scholarpedia* 4 (2009) 2034.
- [6] J. R. Dorfman, An introduction to chaos in nonequilibrium statistical mechanics, 14, Cambridge University Press, 1999.
- [7] Y. B. Pesin, Characteristic exponents and smooth ergodic theory, *Russian Mathematical Surveys* 32 (1977) 55.
- [8] S. Das, C. B. Dock, Y. Saiki, M. Salgado-Flores, E. Sander, J. Wu, J. A. Yorke, Measuring quasiperiodicity, *Europhysics Letters* 114 (2016) 40005.
- [9] J. D. Meiss, E. Sander, Resonance and weak chaos in quasiperiodically-forced circle maps, *Communications in Nonlinear Science and Numerical Simulation* 142 (2025) 108562.
- [10] R. Barrio, Á. Lozano, A. Mayora-Cebollero, C. Mayora-Cebollero, A. Miguel, A. Ortega, S. Serrano, R. Vigara, Deep learning for chaos detection, *Chaos: An Interdisciplinary Journal of Nonlinear Science* 33 (2023).
- [11] S. Uzun, S. Kaçar, B. Arıcıoğlu, Deep learning based classification of time series of chaotic systems over graphic images, *Multimedia Tools and Applications* 83 (2024) 8413–8437.
- [12] A. Celletti, C. Gales, V. Rodriguez-Fernandez, M. Vasile, Classification of regular and chaotic motions in hamiltonian systems with deep learning, *Scientific Reports* 12 (2022) 1890.
- [13] W. S. Lee, S. Flach, Deep learning of chaos classification, *Machine Learning: Science and Technology* 1 (2020) 045019.
- [14] J. Pathak, B. Hunt, M. Girvan, Z. Lu, E. Ott, Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach, *Physical Review Letters* 120 (2018) 024102.
- [15] X. Zhang, A. Chen, H.-K. Zhang, J. Zu, Deep learning for chaotic system via euler difference neural network, *Communications in Mathematical Sciences* 22 (2024) 2253–2270.
- [16] F. M. Izraelev, Nearly linear mappings and their applications, *Physica D: Nonlinear Phenomena* 1 (1980) 243–266.
- [17] T. Y. Petrosky, Chaos and cometary clouds in the solar system, *Physics Letters A* 117 (1986) 328–332.
- [18] F. Benvenuto, G. Casati, D. L. Shepelyansky, Chaotic autoionization of molecular Rydberg states, *Physical Review Letters* 72 (1994) 1818.
- [19] K. Cetin, U. Tirnakli, B. M. Boghosian, A generalization of the standard map and its statistical characterization, *Scientific Reports* 12 (2022) 8575.

- [20] B. V. Chirikov, A universal instability of many-dimensional oscillator systems, *Physics Reports* 52 (1979) 263–379.
- [21] S. Das, Y. Saiki, E. Sander, J. A. Yorke, Quasiperiodicity: rotation numbers, *The Foundations of Chaos Revisited: From Poincaré to Recent Advancements* (2016) 103–118.
- [22] M. Brin, G. Stuck, *Introduction to dynamical systems*, Cambridge University Press, 2002.
- [23] A. G. Kachurovskii, The rate of convergence in ergodic theorems, *Russian Mathematical Surveys* 51 (1996) 653.
- [24] S. Das, Y. Saiki, E. Sander, J. A. Yorke, Quantitative quasiperiodicity, *Nonlinearity* 30 (2017) 4111.
- [25] Z. Levnajić, I. Mezić, Ergodic theory and visualization. I. mesochronic plots for visualization of ergodic partition and invariant sets, *Chaos: An Interdisciplinary Journal of Nonlinear Science* 20 (2010).
- [26] Z. Tong, Y. Li, Exponential convergence of the weighted birkhoff average, *Journal de Mathématiques Pures et Appliquées* 188 (2024) 470–492.
- [27] E. Sander, J. D. Meiss, Birkhoff averages and rotational invariant circles for area-preserving maps, *Physica D: Nonlinear Phenomena* 411 (2020) 132569.
- [28] L. Barreira, Y. B. Pesin, *Lyapunov exponents and smooth ergodic theory*, American Mathematical Society., 2002.
- [29] P. Grassberger, I. Procaccia, Measuring the strangeness of strange attractors, *Physica D: Nonlinear Phenomena* 9 (1983) 189–208.
- [30] M. Ding, C. Grebogi, E. Ott, T. Sauer, J. A. Yorke, Estimating correlation dimension from a chaotic time series: when does plateau onset occur?, *Physica D: Nonlinear Phenomena* 69 (1993) 404–424.