Exploring chaotic time series and phase spaces

de Carvalho Pagliosa, Lucas

DOI:
10.33612/diss.117450127

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
2.1 Initial Considerations

This chapter introduces and describes relevant concepts related to time series (Section 2.2), Dynamical Systems (Section 2.3), and embeddings (Section 2.4). Next, we combine all theory to show how to reconstruct phase spaces from time series in practice (Section 2.5), to finally proceed to important analysis based on phase-states features (Section 2.6).

As noted in Chapter 1, related work encompasses, apart from the fundamental concepts and results related to time series (our main focus), also work in Statistical Learning Theory, Machine Learning, and Information Visualization. Since this second type of related work pertains specifically to the techniques addressing individual research questions, we introduce and describe it only when needed along Chapter 5 to Chapter 9.

2.2 Time Series

A univariate\(^2\) time series \(T_i\) is the sequence of \(n\) observations

\[
T_i = \{x(0), x(1), x(2), \cdots, x(n-1)\}, \ x_k \subset \mathbb{R},
\]  

(2.1)

that models the evolution of some variable \(i\) (e.g., wind speed, relative humidity of the air), representing a feature of some phenomenon of interest (e.g., weather) during an interval of time. In practical scenarios, \(T_i\) is formed after collecting impulses from or solving mathematical equations describing (Butcher, 1996) the phenomenon at sampling rate \(t_s\), which defines the time elapsed between two consecutive observations. Moreover, sampling rates \(t_s\) can be kept constant (Figure 2.1) or change along time, depending on the target application.

Along this manuscript, the subscripted index (such as \(i\) in \(T_i\)) is also appended to the time-series corresponding features, such as

---

1 For some of the topics included in this chapter, different definitions were used in the published articles. Thus, although we have tried our best to standardize the nomenclature in this thesis, the reader may find some divergences when comparing the following chapters with their corresponding articles.

2 In the context of this manuscript, we approach unidimensional time series only. However, the studies developed in this thesis and in the produced articles can be extended to multiple dimensions as performed in (Serrà et al., 2009), without loss of generality.
the number of observations, average, or variance. Further, aggregated indexes will denote variations of the same series. For instance, $T_i$ and $T_j$ represent two different series, while $T_{ij}$, for $j \in [1, s]$, denotes one of the $s$ modifications of $T_i$. In the last case, the features of modified series remain identical to the original by default (i.e., number of observations, average, etc.) unless explicitly stated differently. Such notations become useful when comparing phase spaces (Section 2.3.3) and dealing with surrogate data (Theiler et al., 1992).

In addition to the sampling rate $t_s$, other features such as the length of the series, the initial observation $x(0)$ (especially for chaotic data), and the amount of noise (Graben, 2001) also need to be considered when analyzing a time series. These additional features help quantifying various statistical properties of interest, and are useful to identify when different series might still represent the same phenomenon of interest. Figure 2.2 illustrates the idea on observations from one of the variables of the Lorenz system (Tucker, 1999). As it can be seen by this figure, a time series from the Lorenz system can be represented in different ways. Then, the robustness of some model can be tested against variations of the same series. Nonetheless, we expect the series to be large (at least 1000 observations) and “clean” enough to preserve the nature of the measured variable. According to our point of view, this is not too much to ask for, as no relevant models can really be inferred from too small-noisy datasets. In other words, the series must have sufficient information to unfold the dynamics of the generating rule (see Section 2.3 for details).

Apart from the above, additional notations include the time delay $\tau \in \mathbb{I}^+$, representing the number of observations to be shifted from the current timestamp $t$, such that $x(t \pm \tau) \in T_i$; and the leap time $\rho \in \mathbb{I}^+$, a moment in the future to be forecasted as a single observation.
2.3 Dynamical Systems

A dynamical system \( S^d = \{p_0, \cdots, p_\infty\} \) is a set of \( d \)-dimensional states (also known as points) \( p_t = [p_{t,1}, p_{t,2}, \cdots, p_{t,d}]^3 \) that, driven by a generating (also called governing) rule \( R(\cdot) \), models the behavior of some phenomenon as a function of state trajectories so that

\[
R : S^d \rightarrow S^d,
\]

where \( d \) corresponds to the number of degrees of freedom the system has, i.e., the number variables required to describe \( R(\cdot) \). Further, although \( S^d \) can consist of an infinity of states, in practical terms, a dynamical system is usually represented by the finite set \( S = \{p_0, p_1, \cdots, p_N\} \subset S^d \) of \( N \) states.

2.3.1 Types Of Dynamical Systems

Dynamical systems can be classified in function of their generating rule, as described below.

\footnote{When referring to time-series and phase-spaces attributes, indices will start at 0. For other contexts, such as indices denoting dimensions or position in arrays, we start counting from 1.}
First, the generating rule is classified either as *discrete* or *continuous*, as follows.

**Discrete rules:** Also known as maps, discrete rules are functions of the form $F = \{f_1, f_2, \cdots, f_d\}$ that explicitly relate states based on past values (defining their trajectories) so that $p_{t+1} = F(p_t)$ and

$$S^d = \{F(p_0), F^2(p_0), \cdots, F^\infty(p_0)\}. \quad (2.3)$$

In the above, $F^2(p_t) = F(F(p_t))$, and similarly for higher composition orders.

**Continuous rules:** Also called fluxes, continuous rules are modeled by a set of differential equations

$$p_{t+1} = \partial p_t, \quad (2.4)$$

that describe how $S^d$ varies in the limit. In such scenarios, Equation 2.4 is typically approximated in the form of Equation 2.3 based on discrete methods (Butcher, 1996) solved using the sampling rate $t_s$. Summarizing the above, no matter whether the rule is a discrete map or a continuous flux, $d$ different time series, as described by Equation 2.1, can be generated to represent each dimension of the underlying system.

Separately, $S^d$ can be either *deterministic* or *stochastic*, based on the nature of the generating rule, as follows.

**Non-deterministic (stochastic) dynamical systems** are used to model unknown influences by means of random or conditional parameters. An example of such system is the two-dimensional random walk

$$p_{t+1} = \sum_{t=0}^{n} Z(p_t), \quad (2.5)$$

where $Z(p_t)$ is a Markov chain (Meyn and Tweedie, 2009) over each state based on the probability density function $P([p_0, \cdots, p_t])$. Among other applications, the random walk, depicted in Figure 2.3(a), is a simplified model that mimics the Brownian motion used in physics to represent the random motion of fluid molecules (Einstein, 1956). Random walks are also present in several other disciplines such as economics, chemistry, computing, and biology. In this scenario, when one analyzes each dimension of Equation 2.5 as a time series, the most common approach is to use statistical tools (Box and Jenkins, 2015) to support modeling.
and prediction.

**Deterministic** dynamical systems, on the other hand, have a well-defined generating rule $R(\cdot)$ that produces a single and unique state in the future, given a starting moment. Nonetheless, deterministic systems may present chaotic behaviors. A well-known example is the Lorenz system, designed to model atmospheric data to support weather forecasting (Tucker, 1999) in the form

$$
\begin{bmatrix}
\partial \\
\sigma(y-x) \\
x(\rho-z) - y \\
x y - \beta z
\end{bmatrix}.
$$

In this context, parameters $\sigma, \beta, \rho$ are adjusted to simulate different environmental conditions. A chaotic system is typically observed using $\rho = 28, \sigma = 10$ and $\beta = 8/3$. In this case, approaches like non-linear regression (Bates and Watts, 1988) to forecast observations may lead to poor results when applied directly over dimensions (i.e., time series), as small disturbances tend to evolve to completely different trajectories. Phase-space methods aim to improve modeling by considering phase states and their orbits instead (see Chapter 4).

Figure 2.3: Example of dynamical systems. (a) Stochastic random walk. (b) The Lorenz system was created using $p_0 = \{-13, -14, 47\}$, $t_s = 0.01$, $n = 5001$, $\sigma = 10$, $\beta = 8/3$, and $\rho = 28$. Parameters $\sigma, \beta, \rho$ were set with values known to produce a chaotic behavior.

Lastly, it is worth to say that although systems usually present a mixture of both deterministic and stochastic observations, this thesis focused on exploring predominantly deterministic time series\footnote{We add artificial noise in most of our experiments when dealing with deterministic generating rules.} due to the typical chaotic/cyclical behavior of natural phenomena (Andrews and Herzberg, 1985).
2.3.2 Orbits And Attractors

Let $F$ be a function that represents either a map or a flux after solving the describing differential equations (Equation 2.4). Given a state $p_t \in S \subset S^d$, its $k$-trajectory or $k$-orbit is the set of states $\{p_t, F(p_t), F^2(p_t), \ldots, F^k(p_t)\}$ that defines the temporal evolution of $p_t$ to $p_{t+k}$. A state $p_t$ is called fixed if $F(p_t) = p_t$, and $k$-periodic when $F^k(p_t) = p_t$. A fixed state is also stable or unstable if its nearest states are attracted or repelled to it during the course of their orbits, respectively. Moreover, due to the required notion of distance to measure nearest neighbors, states are assumed to lie in some metric space, such as the Euclidean space $\mathbb{E}^d$, which implies $S^d \subseteq \mathbb{E}^d$. Thus, the state $p_t$ is a neighbor of $p_t$ if it lies in the interior of the open ball $B(p_t, \varepsilon)$, centered in $p_t$ and with radius $\varepsilon$, in form

$$B(p_t, \varepsilon) = \{p_{t'} \in \mathbb{E}^d : \|p_t - p_{t'}\|_2 < \varepsilon\}, \quad (2.7)$$

where $\|\cdot\|_2$ is the Euclidean norm. In this context, if $\lim_{k \to \infty} F^k(p_t) = p_t$, then $p_t$ is a sink or an attractor. On the other hand, if states of the image $F(B(p_t, \varepsilon))$ become more distant from $p_t$ than when they were in $B(p_t, \varepsilon)$, i.e., they are repelled from $p_t$ along their orbits, then such point is called a source state. The basin of attraction is the region formed by the smallest, but sufficient radius $\varepsilon$, such as neighbors of $p_t$ are attracted to it. Moreover, fixed points can behave differently across dimensions, such that saddles may be formed (for $S^{d>1}$), as illustrated in Figure 2.4.

\[\text{Figure 2.4: Different types of attractors. From left to right: } p_t \text{ is (a) an attractor point or sink; (b) a repelling point or source; or (c) a saddle point. Adapted from Alligood et al. (1996).}\]

Based on the above concepts, one may realize that it is not uncommon to find multiple and different types of attractors that, together, define the dynamics of a system. Such orbits sometimes evolve into nonlinear trajectories that are useful to visualize (for low dimensions) and measure important features on the space (Section 2.6). For instance, two-dimensional attractors can be depicted
by cobweb plots, while isolated circuits are called limit circles (Al-ligood et al., 1996). Moreover, periodicities of high-dimensional systems may form \(d\)-dimensional tori. On the other hand, more complex structures like fractals (Section 2.6.1) and manifolds (Section 2.4) are known as strange attractors (Mandelbrot, 1977; Alligood et al., 1996; Lee, 2003), as it is the case of the famous Lorenz system (Equation 2.6). In the latter case, any initial point \(p_0\) will, eventually, converge and be bounded to the trajectories of the attractor, as illustrated in Figure 2.5.

![Figure 2.5: Example of different orbits of the Lorenz system using 20 random initial points \(p_0\). As it is noticed, all trajectories eventually converge to the attractor, never leaving afterwards. In order to simplify the visualization, only a two-dimensional system is shown.](image)

Despite different types of generating rules can lead to previously mentioned attractors, strange attractors are typically found in chaotic systems due to their sensitiveness to the initial conditions. In such scenarios, two almost identical states \(\|p_{t'} - p_t\| \leq \varepsilon \to 0^+\) tend to evolve to completely different orbits (even though remaining restricted to the form of the attractor) as time elapses, eventually getting close to each other again after a certain number of iterations. Such factor makes those systems especially hard to predict, as minimum errors/fluctuations in data sampling and modeling (even in the limited capacity of float number representation) may be enough to change orbit trajectories.

### 2.3.3 Phase Space

A deterministic generating rule \(R(\cdot)\) maps the state \(p_t \in S \subset S^d\) to, ideally, a single state \(p_{t+1}\) in the future. Conversely, unpredictable factors such as random processes can disrupt such mapping for
stochastic systems. Therefore, for deterministic data, the analysis of this rule offers, as main advantage, a more consistent approach to: i) identify patterns, cycles and trends; ii) forecast observations; and iii) correlate systems. The quality of such analyses directly depends on the number of states in \( S \) and how they are rearranged in the space. For the case when \( S \) is characterized by a sufficient set of states representing all possible dynamics of \( S^d \), then \( S \) is called the \textit{phase space} of \( S^d \). When such a phase space is extracted from the time series, the variable \textit{time} has no longer influence on the system \cite{Pagliosa2017}. Therefore, the phase space can be used to interpret how the analyzed phenomenon behaves along \textit{any given period of time}, thereby considerably simplifying the analysis and, in particular, the prediction. Note that if \( S \) is a phase space, then it may be represented by a finite set of states with potentially lower dimension than \( d \), as the dynamics of the system may converge to its attractor.

Although Figure 2.1(b) already exemplifies the phase space of the Lorenz system, let us reinforce this concept using another, simpler, example given by the nonlinear motion of the pendulum

\[
\frac{\partial^2 \theta}{\partial t^2} + \omega \sin \theta = 0, \tag{2.8}
\]

where \( \theta \) defines the angle between the pendulum rod and the vertical line, \( \omega = \frac{g}{L} \) is the motion frequency, \( g \) is the gravitational acceleration, and \( L \) gives the pendulum length. Such a system can be expressed in terms of the angle \( x = \theta \) and the angular velocity \( y = \frac{\partial \theta}{\partial t} \), as the other parameters remain constant. Thus, one can use these two variables to reconstruct the phase space according to the relation

\[
\frac{\partial}{\partial t} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ -\omega \sin x \end{bmatrix}, \tag{2.9}
\]

which represents all possible combinations between angle and angular velocity that a pendulum may have.

Despite the visualization of phase states (in the form of two/three-dimensional trajectories) is usually enough to analyze patterns and behaviors for well-defined structures, a vector field (the gradients of \textit{Equation 2.9}) represents a more intuitive visual depiction to track the \textit{dynamics} of the system when analyzing dense spaces, as illustrated in Figure 2.6. From this figure, it is observed the existence of sink and source states that can be useful to predict cycles and future observations \cite{Boccaletti2008}. Additionally, more refined approaches such as feature detection methods \cite{Post2003}, geometric methods \cite{McLoughlin2018}.
et al., 2010), and texture-based methods (Laramee et al., 2004) can be applied to highlight deeper insights.

![Vectors](image)

Figure 2.6: Vector field of the phase space of the simple pendulum, given by Equation 2.9. Small dashes represent the gradients of states. For simplicity, the arrows were removed, but the overall orientation of states is depicted by solid lines. Using this picture, one can interpret, for instance, that with greater velocities the pendulum has enough energy to rotate, while preserving an oscillating pattern at lower speeds, eventually converging to an equilibrium point when its velocity reaches zero.

Finally, it is worth to say that phase-space analysis is also possible when the generating rule is unknown, as shown in more detail in Section 2.5.

2.4 IMMERSION AND EMBEDDING

This section presents the basic concepts regarding topology and diffeomorphism, which later inspired Takens to propose his embedding theorem to reconstruct the phase space from univariate time series (Takens, 1981).

**Topological space:** A topological space is a set of open sets that, following axioms based on set theory, characterizes one of the most general structures of mathematical spaces. As an open set is an abstract concept that generalizes the notion of an open interval in \( \mathbb{R} \), a topological space is mainly defined by a set of points and their relation with their neighborhoods. Thus, other spaces such as the metric and normed spaces are specializations of the topological space, since additional constraints and measures are defined (Mendelson, 1975). Further, more complex spaces where the concept of distance and direction may be needed in order to perform deeper analyses. Normally, this space is the Euclidean
space $\mathbb{E}^d$ (or equivalently the real space $\mathbb{R}^d$), where the notion of neighborhood is given as in Equation 2.7.

**Manifolds:** The correspondence between topological and Euclidean spaces may be given through manifolds. More precisely, a $d$-manifold $\mathcal{M}$ is a topological space such that each of its open sets $p \in \mathcal{M}$ can be mapped to a point $p \in \mathbb{E}^d$ and vice-versa, i.e., $p \approx p$, without losing any topological property (neighborhood relationships). Under those circumstances, a $d$-manifold is said to be locally homeomorphic to $\mathbb{E}^d$ (Mendelson, 1975; Lee, 2003; LaValle, 2006). By this definition, examples of unidimensional manifolds consist of open intervals in $\mathbb{E}$ and circles in $\mathbb{E}^2$, respectively, while surfaces such as planes, spheres, and tori in $\mathbb{E}^3$ are representations of two-dimensional manifolds.

**Differentiable manifolds:** A differentiable manifold is a manifold locally defined by a set of $C^k$ differential equations that provide additional information to the abstract topological space $\mathcal{M}$. With these functions, one can unambiguously define directional derivatives and tangent spaces to perform infinitesimal calculus and deform manifolds. Some of those deformations, which are in the form $F : \mathcal{M} \rightarrow \mathcal{N}$, receive special attention depending on the properties they preserve. For instance, if $T_p\mathcal{M}$ is the tangent space (Lee, 2003) on the point $p$ in the manifold $\mathcal{M}$, then an immersion is a function whose derivative $\partial_p F$ (partial of $F$ with respect to $p$) is everywhere injective

$$\partial_p F : T_p\mathcal{M} \rightarrow T_F(p)\mathcal{N}, \quad (2.10)$$

which guarantees the resulting image ($\mathcal{N}$) has well-defined derivatives in all its domain ($\mathcal{M}$). However, the image of an immersion is not necessarily a manifold. Figure 2.7 illustrates the possible scenarios involving an immersion. An embedding, on the other hand, is a transformation that, besides being an immersion, is an injective function itself that also creates a diffeomorphism\(^5\) between $\mathcal{M}$ and $\mathcal{N}$. Therefore, in contrast to immersions, the image of an embedding is always a manifold, as illustrated in Figure 2.8. Moreover, if the manifold is compact\(^6\) (as is the case of most attractors), then every injective immersion is an embedding.

The motivation behind these deformations is to transform the topological properties of a manifold into a more intuitive, and easier to process, representation such as a surface in the Euclidean space. One of the most famous examples to illustrate this concept

---

5 A diffeomorphism is an invertible function that maps one differentiable manifold to another, such that both the function and its inverse are smooth.

6 A manifold is compact if it is finite and has limit points.
Figure 2.7: Example of immersions. (a) The eight-shaped closed curve is an immersion of the open set \( (-\frac{\pi}{2}, \frac{3\pi}{2}) \) into \( \mathbb{E}^2 \). (b) The cuspidal cubic (middle) is not an immersion, as the partial of \( f(t) \) is not injective in 0. (c) The nodal cubic (bottom) is an immersion: \( f'(t) = (2t, 3t^2 - 1) = (0, 0) \) has no solution in \( t \). All images are non manifolds. Adapted from Tu (2010).

is the immersion of the Klein bottle, a 2-manifold whose topology can be described by an identification (LaValle, 2006) in the form of a square. In this representation, points near edges should remain together so that the orientation of similar arrows match. Thus, despite the topological space has enough information to describe how points behave, the immersion of the Klein bottle into \( \mathbb{E}^3 \) (Figure 2.9) turns it much easier to understand and study, even if the resulted image is not a manifold. Similarly, one can embed the Klein bottle into \( \mathbb{E}^4 \) to remove the observed self-intersections.

Finding a space in which a manifold can be embedded is not a trivial task in most cases. In this context, Whitney (1936) proposed a theorem saying that \( \mathbb{E}^{2d+1} \) is a sufficient space to embed a \( d \)-manifold, since no two points from a \( d \)-dimensional manifold could be mapped to the same point in the \( (2d + 1) \)-dimensional space (Ghomi and Greene, 2011; Forstnerič, 2011). It is worth to re-
Figure 2.8: The function is not an embedding in $\mathbb{R}^2$, but it is in $\mathbb{R}^3$. In this example, $t \in (\frac{-\pi}{2}, \frac{3\pi}{2})$. Adapted from Tu (2010).

Figure 2.9: Identification of the topological space (left) and the resulting image of the immersion of the Klein bottle into $\mathbb{E}^3$ (right). Points in this topology should be close to each other such that the orientation of similar arrows are equal.

Inforce, however, that this theorem elaborates a sufficient, but not necessary condition, such that lower dimensions may be enough to embed a manifold, as it is the case of the Klein bottle. According to Whitney’s theorem, such 2-dimensional manifold can be embedded in $\mathbb{E}^5$, but $\mathbb{E}^4$ is already enough.

Extending this study, Takens (1981) proposed his own embedding theorem, described next. Let $\mathcal{M}$ be a compact manifold of dimension $d^7$. For pairs $(\varphi, y)$, where $\varphi : \mathcal{M} \to \mathcal{M}$ is a diffeomorphism and $y : \mathcal{M} \to \mathbb{R}$ is a smooth function, it is a generic property that the map $\Phi : \mathcal{M} \to \mathbb{E}^{2d+1}$, in the form

$$\Phi_{(\varphi, y)}(p) = (y(p), y \circ \varphi(p), \cdots, y \circ \varphi^{2d}(p)),$$

is an embedding. In other words, the main contribution of Takens’ theorem was to show that a single quantity of the manifold $\mathcal{M}$ is enough to embed it in $\mathbb{E}^{2d+1}$. However, like Whitney, Takens’ theorem elaborates a sufficient, but not a necessary condition. As matter of fact, the space $\mathbb{E}^{2d+1}$ is usually an overestimation, and

---

7 The dimension here is the Euclidean space in which the manifold lies, not the dimension it is homeomorphic to.
finding a lower, simpler embedding dimension, from now on referred to as \( m \), is desirable in order to decrease the computational costs involved in modeling and prediction, especially when dealing with large volumes of continuously collected data, also known as data streams (Muthukrishnan, 2005). For instance, a sufficient space to embed the 2-manifold, 3-dimensional Lorenz system, according to Takens is \( \mathbb{E}^7 \), but \( \mathbb{E}^{m=3} \) is already enough to unfold the Lorenz attractor dynamics.

2.5 Reconstructing Phase Spaces

The previous sections have introduced the mathematical support on dynamical systems and immersions. Next, this section combines those concepts and describes how a time series can be embedded in practice.

As previously discussed, the process of finding a finite set \( S \subset S^d \) resembling the dynamics of \( S^d \) is known as unfolding or reconstructing the phase space of \( S^d \). If one knows \( R(\cdot) \), the reconstruction becomes quite straightforward after generating enough states of the respective map or discretized flux. However, this process becomes more difficult when the generating rule is unknown, as it is the case of real-world data sampled from some arbitrary time-dependent phenomenon. Additionally, an even more problematic issue is the lack of information on the data; humans tend to model phenomena in terms of the variables they observe and know, which usually tend to be an insufficient and inaccurate representation of the underlying phenomenon. Separately, data measurements may in practice be corrupted or have missing values, forcing the analyst to disregard them. Summarizing, one may face several scenarios in which only a small number of dimensions is available for analysis. In the limit, we consider the case where just a single dimension \( i \in [1, d] \) of \( S^d \) is considered.

A dynamical system \( S^d \), especially when modeling natural phenomena, usually presents recurrent patterns and observations. In addition, it is expected that variables composing such a system do not only impact themselves, but directly or indirectly affect other variables along time. Such correlation can indeed be noticed in the Lorenz system (Equation 2.6) and in the simple pendulum map (Equation 2.9). Further, if one represents the \( i \)th \( (i \in [1, d]) \) component of all phase states as the time series \( T_i \), it is reasonable to expect that such observations have, even that implicitly, informa-

\[ \text{There exist methods that analyze the impact of using more than one time series to reconstruct the phase space (Cao et al., 1998). However, this matter is out of the scope of this thesis.} \]
tion related to other variables of $R(\cdot)^9$. In order to take advantage of this relation, one can rely on Takens’ embedding theorem (Takens, 1981) to embed a $d$-dimensional manifold $M$ into $\mathbb{E}^{2d+1}$ according to Equation 2.11, where $y(\cdot)$ is interpreted as a direct map to access the observations of $T_i$. Thus, according to Takens, a time series $T_i$ can be embedded into a space that is diffeomorphic to $S^d$ or, more precisely, to its phase space $S$. In this situation, the phase space will be represented by the $N_i \times (2d + 1)$ trajectory matrix, denoted from now on to as $\Phi_i$, in form

$$
\Phi_i = \begin{bmatrix}
y(p_0) & y \circ \varphi(p_0) & y \circ \varphi^2(p_0) & \cdots & y \circ \varphi^{2d}(p_0) \\
y(p_1) & y \circ \varphi(p_1) & y \circ \varphi^2(p_1) & \cdots & y \circ \varphi^{2d}(p_1) \\
y(p_2) & y \circ \varphi(p_2) & y \circ \varphi^2(p_2) & \cdots & y \circ \varphi^{2d}(p_2) \\
y(p_3) & y \circ \varphi(p_3) & y \circ \varphi^2(p_3) & \cdots & y \circ \varphi^{2d}(p_3) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
y(p_{N_i-1}) & y \circ \varphi(p_{N_i-1}) & y \circ \varphi^2(p_{N_i-1}) & \cdots & y \circ \varphi^{2d}(p_{N_i-1})
\end{bmatrix},
\tag{2.12}
$$

so that, mathematically

$$
T_i \rightarrow \Phi_i \approx S \subset S^d. \tag{2.13}
$$

In this context, Takens also proposed a convenient diffeomorphic function $\varphi$ in the form

$$
\varphi^\tau(p) : p \rightarrow \tau p, \quad \tau \in \mathbb{I}^+, \tag{2.14}
$$

later commonly known as the method of delays due to its time displacement characteristics. Assuming the manifold is discretized as a non-uniform grid (see Figure 2.10), such a direction could be, for instance, the dimension $i$, so that $\varphi$ shifts the point $p \in M$ $\tau$ units to the right. Finally, the function $y(\cdot)$ merely maps the component $p_{t,i}$ from the phase state $p_t$ to $x(t) \in T_i$, as illustrated in Figure 2.11.

Therefore, given the time series $T_i = \{x(0), \cdots, x(n_i - 1)\}$, the phase space $\Phi_i$ with $N_i$ states can be reconstructed according to Equation 2.12. More precisely, the method of delays reconstructs each phase space as

$$
\phi_i(t) = [x(t), x(t + \tau), x(t + 2\tau), \cdots, x(t + 2d\tau)], \tag{2.15}
$$

where $\tau$ is the time delay, as defined in Section 2.2. Moreover, it is worth to note that, because $\Phi_i \approx S$, i.e., the reconstructed

---

9 As consequence, it is worth to mention that the quality of the reconstructed phase space depends on the amount of influence $T_i$ brings about other variables.
phase space is diffeomorphic to $S$, it is known that Equation 2.15 locally preserves neighboring properties of phase states. According to our notation, this relation is expressed as

$$\Phi_i(t) \approx p_t. \quad (2.16)$$

Other strategies to reconstruct the phase space are either modifications of Equation 2.15 (Broomhead and King, 1986) or based on different diffeomorphism functions. For instance, Packard et al. (1980) proposed the method of derivative coordinates, where each row in Equation 2.12 is defined as

$$\varphi^\tau(p): p \to \frac{\partial^\tau F(p)}{\partial \tau}, \quad \tau \in \mathbb{I}^+. \quad (2.17)$$

This method creates phase states similarly to the method of delays but uses infinitesimal time delays, which is impractical when the generating rule is unknown. Nevertheless, one can assume the time series is structured as $T_i = \{x(-h_i), \cdots, x(0), \cdots, x(h_i)\}$, where $h_i = (n_i - 1)/2$, and approximate Equation 2.17 by finite differences (Canuto and Tabacco, 2008) as

$$\phi_i(t) = \begin{bmatrix} x(t), \frac{x(t + \tau) - x(t)}{\tau}, \frac{x(t + \tau) - 2x(t) + x(t - \tau)}{\tau^2}, & \cdots \end{bmatrix}.$$

$$\quad (2.18)$$

Figure 2.12 illustrates both methods for the Lorenz system. While there is no clear evidence about which of these methods is the most appropriate, Ravindra and Hagedorn (1998) elaborate
that the method of delays produces better results when analyzing nonlinear time series. In fact, the method of delays is the most used approach in the literature (Stark, 1999; Yap and Rozell, 2011; Yap et al., 2014). It is worth to say that the method of delays, as originally proposed (Equation 2.15), assumes an uniform $\tau$. Nonetheless, there are articles that investigate the usage of multiple time delays (Breedon and Packard, 1992; Manabe and Chakraborty, 2007).

![Diagram](image)

**Figure 2.11:** Diffeomorphism according to the method of delays. The method of delays allows the reconstruction of the phase space using a single dimension $i$, represented as the time series $T_i$. In the proposed coordinate system, $i$ represents the first dimension/component.

\[
\begin{align*}
P_t &= [p_{t,1}, p_{t,2}, p_{t,3}] \\
\varphi(p_t) &= [p_{t+\tau,1}, p_{t+\tau,2}, p_{t+\tau,3}] \\
\varphi^2(p_t) &= [p_{t+2\tau,1}, p_{t+2\tau,2}, p_{t+2\tau,3}] \\
y \circ p_t &= x(t) \\
y \circ \varphi(p_t) &= x(t + \tau) \\
y \circ \varphi^2(p_t) &= x(t + 2\tau)
\end{align*}
\]

**Figure 2.12:** Despite the different results, the reconstructed phase space using either the method of delays (a) and derivatives (b) preserve topological properties and, most importantly, the dynamics of the original Lorenz system (Figure 2.3(b)). The reconstruction was performed using $m = 3$ and $\tau = 8$ in both methods. For simplicity, only the first two dimensions are visualized.
2.6 PHASE SPACE FEATURES

Based on reconstructed phase space, several methods were proposed to identify and predict chaotic time series (Farmer and Sidorowich, 1987; Andrievskii and Fradkov, 2003; Boccaletti and Bragard, 2008; de Mello and Yang, 2009). Next, concepts such as correlation dimensions and Lyapunov exponents, important to these two types of analysis, are described.

2.6.1 Fractal Dimension

Following the Gestalt principles (Chang et al., 2007), a geometrical object (shape) can be described in term of its patterns and how these are arranged in space. Further, a pattern can be defined as a feature that repetitively occurs at ε spatial units of measure. For instance, patterns can be defined in function of length, area, or volume, in one, two, and three-dimensional spaces, respectively. Therefore, if a D-dimensional object presents N patterns, as illustrated in Figure 2.13, one can notice the relation of proportionality (∝)

\[ N \propto \varepsilon^D \Rightarrow D \approx \log N / \log \varepsilon. \] (2.19)

With the above, a fractal can be defined as an object whose patterns are given in function of the object itself (Mandelbrot, 1977). If patterns are perfect replicas occurring at every scale ε, the fractal follows a self-similar pattern, as it is the case of the Koch snowflake (Figure 2.14). Differently from other shapes, fractals usually do not have a uniform relation between ε and N, so that D, also called the fractal dimension, is often a real number. Despite not being unique, this quantity turns to be an important space descriptor, as it abstracts the complexity of a shape.

Use for dynamical systems: In the scope of dynamical systems, one can compute the fractal dimension D based on the orbits of the
phase states to quantify the structure of $S^d$. Among other applications, the fractal dimension becomes useful to distinguish deterministic/chaotic from stochastic systems, as explained next. Suppose we have a dynamical system with a $d'$-dimensional attractor. The fractal dimension of such an attractor is less or equal to $d'$, i.e., $D \leq d'$ (due to the finite data, noise, and no self-similar patterns). If the attractor is embedded or lies in some higher dimension, the fractal dimension stays the same. Therefore, the fractal dimension of a reconstructed phase space is \textit{invariant to the embedding dimension $m$}. If the generating rule is guided by a stochastic process, e.g., a Normal distribution, the space has greater probability to be equally filled by states, leading the fractal dimension to always be equal to the embedding dimension. In summary, the following relation is noticed:

1. if the fractal dimension $D$ does not vary with the embedding dimension $m$, then $m$ is greater than the dimension of the attractor, and $T_i$ has more probability to be deterministic (chaotic or not);

2. if the fractal dimension $D$ is equal to the embedding dimension $m$ for different values of $m$, either the space does not unfold the dynamics of the attractor or $T_i$ is stochastic (non-deterministic).

Unfortunately, the fractal dimension $D$ cannot be exactly computed in practice. Because fractals may change patterns as $\varepsilon \to 0$, one would need an infinity amount of data ($N \to \infty$) to find the true value of $D$. Thus, several approaches were proposed to \textit{estimate} the fractal dimension of an object. These include the box-counting dimension $D_0$; information dimension $D_1$; \textit{correlation dimension} $D_2$ and $D_L$ \textit{Lyapunov dimension} $D_L$ (Clark, 1990; Theiler, 1990; Ding et al., 1993; Alligood et al., 1996).

While all these dimensions have relative advantages, the correlation dimension $D_2$ is one of the most used methods in the literature (Kantz and Schreiber, 2004). Nonetheless, there is evidence
2.6 Phases Space Features

that \( D_2 \) and \( D_L \) are numerically close to each other and less prone to be affected by noise for small datasets (Otani and Jones, 1997). As consequence, we next detail the \( D_2 \) method, followed by the \( D_L \) method (Section 2.6.3).

2.6.2 Correlation Dimension

The correlation dimension of the system \( S^d \) is based on the correlation integral

\[
\hat{C}(\varepsilon) = \int \int \theta(\varepsilon - \|p_t - p_{t'}\|_2)dH(p_t)dH(p_{t'}),
\]

where \( \varepsilon \) is an open-ball radius, \( \|\cdot\|_2 \) is the Euclidean norm, and \( H(\cdot) \) is the invariant distribution of \( S^d \). As we have \( \Phi_i \approx S \in S^d \), the correlation sum, proposed by Grassberger and Procaccia (1983), aims to estimate the correlation integral for the discrete set of \( N \) phase states, embedded using parameters \( (m, \tau) \), as

\[
C(m, \varepsilon) = \sum_{t<t'}^{N-1} \theta(\varepsilon - \|p_t - p_{t'}\|_2) \frac{2}{N \ast (N-1)},
\]

where \( \theta \) is the Heaviside step function

\[
\theta(x) = \begin{cases} 
0, & \text{if } x < 0, \\
1, & \text{otherwise}. 
\end{cases}
\]

The correlation dimension \( D_2 \) is then estimated from the fractal integral (Equation 2.19) as

\[
\lim_{\varepsilon \to 0} \lim_{N \to \infty} C(m, \varepsilon) \approx \varepsilon^{D_2} \Rightarrow D_2 \approx \lim_{\varepsilon \to 0} \lim_{N \to \infty} \frac{\log C(m, \varepsilon)}{\log \varepsilon}. \quad (2.23)
\]

However, as the phase space is represented by a finite set of states, one cannot infinitely decrease the radius \( \varepsilon \), also referred to as scaling factor in this case. Thus, the most common approach is to assume the fractal to have self-similar patterns and extrapolate it as \( \varepsilon \to 0 \). Among the alternatives, this estimation is obtained by computing the slope of the regression line that best fits the points in the plot \( \log C(m, \varepsilon) \) versus \( \log \varepsilon \), as shown in Figure 2.15. In addition, as the correlation dimension should not vary for chaotic attractors (as it is the case for the Lorenz system) when \( m \geq D \), a better estimate can be achieved by taking the average slope for multiple embeddings. For instance, this approach estimates \( D_2 = 2.04 \) for the Lorenz system, where the true fractal dimension is known to be 2.05 (Grassberger and Procaccia, 1983). Despite the good result for this particular case, estimating the correlation dimension...
is not trivial for general systems, as the algorithm is sensitive to several parameters including the number of observations $N$, the scaling factor $\varepsilon$, and the interval from where the slope is computed.

![Figure 2.15: The correlation dimension of the Lorenz system ($\tau = 8$) can be estimated as the average slope of the log-log plot between the correlation sum versus the scaling factor $\varepsilon$, over different embedding dimensions (from top to bottom: $m = [3, 6]$). The interval used to compute the slope is bounded by bolder lines.](image)

### 2.6.3 Lyapunov Exponents

Let $J_t$ be the $d \times d$-Jacobian matrix in the form

$$J_t = \begin{bmatrix}
\frac{\partial f_1}{\partial p_{t,1}} & \cdots & \frac{\partial f_1}{\partial p_{t,d}} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_d}{\partial p_{t,1}} & \cdots & \frac{\partial f_d}{\partial p_{t,d}}
\end{bmatrix}, \tag{2.24}
$$

where $F = \{f_1, f_2, \cdots, f_d\}$ is a $d$-dimensional function applied to the point $p_t$. Among several applications, the Jacobian matrix can be used as a transformation to linearly approximate states through Taylor expansion (Canuto and Tabacco, 2008) as

$$F(p_t + h) \approx F(p_t) + J_t h, \tag{2.25}$$

where $h \to 0$ is an infinitesimal displacement vector. Moreover, if $(\lambda_i, v_i)$, $i \in [1, d]$ are eigenpairs of the Jacobian $J_t$, then the eigenvalues $\lambda_i$ are the coefficients of the linear combination $J_t = \lambda_1 v_1 + \cdots + \lambda_2 v_2 + \cdots + \lambda_d v_d$ that quantifies the rate of variation $p_t$ has in each dimension. In other words, the Jacobian can be used as an approximation of the derivatives of $F$.

**Use for dynamical systems:** In the scope of dynamical systems, $F$ represents the map or discretized flux of the generating rule
\textbf{2.6 Phase Space Features}

$R(\cdot)$ describing the dynamics of $S^d$, and $p_t$ is a phase state on that system. In that sense, if $p_t$ and another state are separated by $h$ units of measurement after $k$ iterations of their orbits, so that $\delta(k) = \|\delta(k)_1, \cdots, \delta(k)_d\|_2 = \|h\|_2$ is the norm of such distance (Figure 2.16), the divergent rate for each dimension $i \in [1, d]$ can be described by the corresponding eigenvalue $\lambda_i$ of $J^k_t$, defined from the chain rule (Alligood et al., 1996) as

$$J^k_t = J_{t+k-1} J_{t+k-2} \cdots J_t.$$  \hspace{1cm} (2.26)

For the case of chaotic time series, however, close trajectories are known to exponentially diverge from each other after $k$ iterations (Figure 2.16), such that

$$\lim_{k \to \infty} \delta(k)_i \approx \delta(0)_i \exp(\lambda_i k),$$  \hspace{1cm} (2.27)

is the divergence rate for dimension $i$. As consequence, the average variation between states along the dimension $i$, per iteration, is roughly approximated as $\lambda_i^{1/k}$.

![Figure 2.16: Divergence of initially close orbits in chaotic systems. State trajectories tend to diverge exponentially in chaotic systems, so that two states, initially close, evolve to completely different orbits after $k$ iterations. In this example, $k = 3$.](image)

The relation among chaotic trajectories has motivated the definition of the $i$th local Lyapunov exponent as

$$L_i = \lim_{k \to \infty} \frac{1}{k} \log \lambda_i,$$  \hspace{1cm} (2.28)

in attempt to quantify the amount (and in which direction) the dynamical system is varying. Kaplan and Yorke (1979) proposed to rank local Lyapunov exponents, in the form of $L_1 \geq L_2 \geq \cdots \geq L_d$, and use them to estimate the Lyapunov dimension $D_L$ as

$$D_L = j + \sum_{i=1}^j \frac{\lambda_i}{\lambda_j + 1},$$  \hspace{1cm} (2.29)

where $j$ is the largest integer so that $\lambda_1 + \lambda_2 + \cdots + \lambda_j \geq 0$. 

31
Besides estimating the fractal dimension, Lyapunov exponents are used to measure a system sensitivity to initial conditions, i.e., chaos (Alligood et al., 1996). More precisely, if one computes the (global) Lyapunov exponent as the most representative dispersion of \( J_t^k \) (some authors use the trace of \( J_t^k \), i.e., \( \sum_{i=1}^d \lambda_i \), instead), we define

\[
\lambda = \max(L_i),
\]

and then the system either tends to converge to attractor points \( (\lambda < 0) \), be conservative \( (\lambda = 0) \), or present unstable and chaotic behavior \( (\lambda > 0) \) after \( k \) iterations. In this context, \( \lambda \) is also used to define the prediction horizon, i.e., the maximum number of future observations one can forecast within a reliable margin of confidence, as

\[
H = -\frac{\log E}{\lambda},
\]

given a training error \( E \). To exemplify this concept, assume that an algorithm models a time series achieving training error \( E = 0.001 \) for a \( \lambda = 0.692 \) system. Despite the small training error, by Equation 2.31, only \( H = -\frac{\log 0.001}{0.692} = 9.98 \approx 10 \) observations can be predicted with enough confidence. Without knowing that, one may wrongly assumed that the algorithm overfitted, being incapable of generalization (Vapnik, 1998; Luxburg and Schölkopf, 2011). Actually, the algorithm has likely learned the data well, but its forecasting capabilities are limited by the chaotic nature of the analyzed system. Thus, small divergences in the representation of initial conditions lead orbits to exponentially diverge after \( k \) iterations, especially if one uses recursive forecasting to respect the butterfly effect (Brock et al., 1992).

Unfortunately, the Lyapunov dimension is unfeasible to compute in practical scenarios when generating rules are unknown and/or when the Jacobian matrix is not computable. In order to overcome this issue, one needs to first reconstruct the phase space from a time series \( T_i \in [1, d] \) and compute Equation 2.27 for several reference states. Similarly to the correlation dimension, one can extrapolate \( k \to \infty \) when dealing with finite datasets by taking the slope of a linear region on the \( S_i(k) \) versus \( k \) plot, where \( S(k) = \lambda k \).

### 2.7 Final Considerations

This chapter introduced the main concepts associated with this thesis, starting with the definition of time series to later cover important dynamical-systems characteristics such as trajectories, orbits, and attractors. As highlighted, a fundamental concept and
tool for the analysis of such systems is the phase space, for which several reconstruction methods have been outlined. By unfolding the dynamical behavior in such space, one can reveal important features about the underlying phenomenon and, therefore, reach a better understanding of the series and its properties. In addition, due to the map between phase states and time-series observations, one can take advantage of modeling phase-space trajectories to forecast chaotic time series.

However, characterizing dynamical systems by analyzing their phase-space representations is challenging. First and foremost, it is not evident how to reconstruct a phase space in practice, given a sampled time-dependent phenomenon. Secondly, this reconstruction is subject to various parameters and quality metrics, and it is not evident how to compute, or even define, which is the optimal reconstruction (apart from relatively simple dynamical systems having well-known generating rules). Therefore, the concepts described in this chapter were important to either understand the related work and ii) to formulate the proposed research questions in Chapter 1.

Based on the concepts introduced here, we next exemplify the challenges involved in interpreting dynamical systems by introducing more complex examples (Chapter 3). Further on, we describe techniques for estimating embedding parameters in Chapter 4.