

Project Thesis/Report

Contents

Chapter 1

Introduction

Experimenting on biological, physical, and artificial systems in order to generate a more informative dynamical model is a well-established practice in modern science. Traditional methods for modelling physical systems are based on laws of physics that are based either on empirical relationships or on intuition. For systems that evolve with time, physical laws yield mathematical equations that govern how the quantities evolve with time. However our world around is, for the most part, much more complex than that which can be distilled into elegant equations. We do not fully understand many of the more complex systems, nor do they even provide us with a good physical intuition of the underlying principles governing the dynamics. To complicate this - the underlying systems we observe often display sensitive dependence on initial conditions; despite having highly similar initial conditions, differing orbits diverge quickly and to such an extent that it becomes seemingly impossible to retrace their steps back to the original conditions. Even the presence of usually 'negligible' computational noise or measurement error renders long-term point-wise prediction infeasible. There are many difficulties that one encounters while modelling such systems:

- i. One may not have access to the complete states of the systems
- ii. The actual system may be described by functions that behave wildly in the sense their graphs have a wild oscillatory behaviour, i.e., they have a large functional complexity.

Models derived primarily from data can be classified into three categories:

- i. Interpretable models (i.e., they establish relationships between internal physical mechanisms),
- ii. Partially interpretable models capturing some modes of the dynamics,
- iii. Non-Interpretable models, defined as such mainly due to the fact that they are defined on a different phase space that is usually high-dimensional.

Examples in the literature attempting to forecast data from such systems have tried a number of different approaches with differing degrees of success. When we have complete access to states of the system, an ordinary differential equation can be obtained from data (Citations to Sindy and [?, ?])

wherein one could approximate the vector field by a library of functions to obtain interpretable models.

Recent partially interpretable models available in the literature have been based on the Koopman operator (e.g., [?, ?]) to employ observables mapping the data onto a higher-dimensional space. This then makes the dynamics in the higher-dimensional space more amenable for approximation by a linear transformation. Such methods, obviously, do not guarantee an exact reconstructions for nonlinear models, and in practice provide poor long-term consistency for a large class of systems. (Add citation)

The non-interpretable models include the delay embedding and the machine learning algorithms. One could learn a system conjugate to the underlying system by applying the Takens delay embedding [?] when one has “good” observations from the system; this learnt system could then be used for forecast the observed data. Takens delay embedding theorem [?] and its various generalisations (e.g., [?, ?, ?, ?]) establishes the learnability of a system constructed by concatenating sufficiently large previous time-series observations of a dynamical system into a vector (called delay coordinates). This then establishes the existence of a map on the space of delay coordinates equivalent (or topologically conjugate) to the underlying map from which the observed time-series was first obtained. Although topological conjugacy guarantees an alternate representation of the underlying system, the quality of this representation still depends on numerous parameters, making the comprehension of the dynamics rather unreliable (e.g., [?]). One reason for this fragility is that the embedded attractor in the reconstruction space not always an attractor of the map that is learnt in the reconstruction space, despite unquestionably being an invariant set. When the embedded object is not (**definitely/always?**) an attractor (as explained in Chapter 3) it can cause predictions to fail.

Practically, the application of Takens embedding involves learning a map through some technique and consequently one wishes that these would have low functional complexity[?], i.e., functions with fewer oscillatory graphs. Pure machine learning methodology that processes temporal information (like the echo state networks (Citations to echo state networks)) by mapping data onto a higher dimensional space for further processing. Although they perform well on forecasting some dynamical systems, they fail completely on others (Citations here) as there is often no guarantee that the right function was learnt during training.

This thesis deals with the implementation and analysis of non-interpretable models that can guarantee **exact/accurate** reconstruction. The project work concerns the study and implementation of a method (Citation to paper with Adriaan) that incorporates learning a function by mapping the data on to a higher dimensional space using what is called a driven dynamical system (See Chapter 4). With a clear understanding of how the data is mapped onto the higher dimensional space, the method then permits the learning of a dynamical system topologically conjugate to that of the underlying system. Instead of linear regression as in the training of echo state networks, deep learning methods are employed to learn the correct function. With slight modifications to the implementation in the paper (with Adriaan), we show that one can construct accurate non-interpretable models with the ability to reconstruct attractors from more hard-to-forecast systems like the double pendulum. (The forecasting of the time-series from a double pendulum has not been

reported before.) Moreover, we also demonstrate that long-term statistical consistency is preserved.

By solidifying the mathematical underpinnings of our theory, we hope to guarantee the ability to construct models with predictive power ranging from molecular biology to neuroscience. in the near future. (We can modify this sentence when the report is completed.)

The report is organised as follows:

In ¹Chapter ??, we recall the definition of a discrete-time dynamical system, how a discrete-system arises from a flow of an ODE and then proceed to define the inverse limit space and topological conjugacy of autonomous systems. EdN:1

In Section 3, we introduce the problem of forecasting dynamical systems, state the Takens delay embedding theorem and discuss various issues faced while forecasting.

In Section 4, a driven dynamical system is defined and discuss the properties of a specific class of driven dynamical systems that we make use of in this thesis/report/paper.

Finally in Section 5 we show the implementation of these forecasting methods, and conclusions are provided in Section 6.

¹EdNOTE: M: Please label Sections as Chapters as this is apt for a thesis; for the paper that we will write soon, it would be in Sections. Please note the labelling in the tex file, and you may like it to be cross references like this

Chapter 2

Discrete-time Dynamical Systems

In this chapter, we provide a brief description of what a discrete-time dynamical system is and what it means for it to exhibit chaos. We refer to [?, ?] for more details.

At its most elementary level, a dynamical system is just something that evolves deterministically through time. In the context of this thesis, deterministic refers to the fact that a system evolves according to specified rules rather than based on random events. Dynamical systems arise in a variety of situations. A continuous-time dynamical system describes the states for all values of the time. Specifically, if the motion of a pendulum in which the quantities such as the angular position and angular momentum are known at all times, then it is a continuous-time dynamical system. The equations of the dynamical system can take the form of one or more ordinary differential equations that determine the relevant quantities at any future time if we know the initial location and momentum. In ecology, discrete-time dynamical systems are widely used to model population growth. The model in this case is a function calculating the following generation's population given the population of the previous generation. If we know the starting population, we may once again calculate the population at any time in the future.

Formally, a function $T : U \rightarrow U$, where U is some set is a *discrete-time dynamical system* and its iterates $\{u, Tu, T^2u, \dots\}$, where T^n denotes the n -fold composition of T with itself, describe the evolution of an initial condition $u \in U$ (Note that we drop the brackets and denote $T(u)$ by Tu frequently for simplicity in notation).

Continuous-time dynamical systems modelled using ordinary differential equations can give rise to discrete-time dynamical systems. To see this, consider a differential equation $\dot{x} = f(x)$, $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $n \in \mathbb{N}$ given to have a unique solution passing through each point $x \in \mathbb{R}^n$, the flow of the equation is defined to be a mapping $\varphi : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$ that satisfies the following properties ² EdN:2

- i. $x(t)$ is a solution of the differential equation,

²EdNOTE: M: Please change this or revert back to how it was defined earlier. We are describing φ and in the first two points, it does not concern φ

- ii. $x(0) = x_0$, and
- iii. $\varphi(x_0, t) = x(t)$.

By fixing $t = K \in (0, \infty)$, we can define the *time- K map* as $T(x) := \varphi(x_0, K)$, and it is easily verified that $T \circ T(x_0) = \varphi(x_0, 2K)$. In general the m^{th} iterate of x_0 under T would be the value of the solution of the ODE evaluated at time mK with the initial condition x_0 . Thus ordinary differential equations give rise to a discrete-time dynamical system by sampling the value of x at time intervals K units apart.

A numerical discretization of a differential equation can also give rise to a discrete-time dynamical system. For instance, Euler's method approximates $\dot{x}(t)$ by $(x(t+h) - x(t))/h$; if h is fixed throughout, the solution of a differential equation $\dot{x} = f(x)$ can be approximated at the time instant $t + (m+1)h$ by iterating the equation

$$x(t + (m+1)h) = x(t + mh) + hf(x(t + mh)).$$

Adopting more succinct notation by replacing $x(t + mh)$ with u_m , we rewrite the above equation as

$$u_{m+1} = u_m + hf(u_m)$$

or, in even simpler terms, as the discrete-time dynamical system with map $T(u) = u + hf(u)$, where T , u and f are understood to be as above.

Of course, discrete-time dynamical systems need not always arise through a differential equation, and once again we may consider the field of ecology where discrete dynamical systems are often directly derived or assumed. There is a school of thought that advocates discrete-time dynamical systems to be more natural for modelling real-world observations than differential equations and we refer the interested reader to [?]. **too philosophical, or actually good?** ³

EdN:3

2.0.1 Invariant Sets

A core concept in the study of dynamical systems is that of invariance. Given a discrete-time dynamical system $T : U \rightarrow U$, a subset $A \subset U$ is said to be an *invariant set* if $T(A) = A$. We also define the *orbit of T* to be a sequence $\bar{u} = \{u_n\}_{n \in \mathbb{Z}}$ obeying the update equation, $u_{n+1} = Tu_n$, $n \in \mathbb{Z}$.

Two examples of invariant sets include a fixed point where $Tu = u$ for $u \in U$, and a periodic orbit, i.e., a set of iterates $\{u, Tu, T^2u, \dots, T^pu\}$, where $T^{p+1}u = u$ for some p . The entire space U could be also be invariant. Consider for example the space $U = [0, 1]$, where $Tu = 4u(1 - u)$ and then U is invariant. (*as every $u \in U$ can be written as $u = 4x(1 - x)$ for some $x \in U$*)

³EdNOTE: M: It is good! I have moved this paragraph

We may learn a great deal about the iterates of a dynamical system by considering the types of invariant sets of a discrete-time dynamical system. For example, if $U = [0, 1]$ has map $Tu = u/2$, then the only invariant set is $\{0\}$, and all orbits approach this invariant set as time flows in the forward direction. Indeed, if a some non-zero invariant set (call it B) exists, then there is some $r \in (0, 1] \cap B$. But $r \notin T(B)$ since any orbit with initial value r will be a decreasing sequence. Moreover, every orbit of T will be decreasing and therefore approach the value 0 as $n \rightarrow \infty$.

One may ask if every orbit approaches an invariant set? In general the answer is no, since for the dynamical system $T : \mathbb{R} \rightarrow \mathbb{R}$ defined by $Tu = 2u$, any orbit that does not intersect the invariant set $\{0\}$, will not approach any invariant set.

However, when the space U is compact, all orbits approach an invariant set. This is since the set of limit points of the orbit can be shown to be invariant [?].⁴ So, when U is compact, the ω -limit set $\omega(u; T)$ of a point u defined to be the collection of limit points of the sequence $\{x, Tu, T^2u, \dots\}$ is nonempty, and $\omega(u; T)$ is invariant. For instance, for the map $Tu = u^2$ defined on $[0, 1]$ all orbits lie in the invariant set $\{1\}$ or else would approach the invariant set $\{0\}$.⁵

We discuss various properties of invariant sets: Invariant sets can be attracting or repelling depending on how orbits in its vicinity behave. Recall the examples $Tu = u/2$ and $Tu = 2u$ defined on \mathbb{R} ($\{0\}$ “attracts” orbits) or the example, $Tu = u^2$ defined on $[0, 1]$ ($\{1\}$ “repels” orbits). We are interested in attractive invariant sets since they capture the long-term dynamics as time increases, and in particular the invariant sets that are called attractors whose definition we recall next.

Definition 1. Let $T : U \rightarrow U$, where U is a metric space with metric d . A compact subset $A \subset U$ is said to be an attractor if it satisfies the three conditions:

1. A is invariant.
2. A is asymptotically stable, i.e., for every $\epsilon > 0$ and for all u so that $d(u, A) < \epsilon$, we have $d(T^n u, A) \rightarrow 0$ as $n \rightarrow \infty$.
3. A has Lyapunov stability, i.e., for every $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ so that $d(u, A) < \delta$ implies $d(T^n u, A) < \epsilon$ for all $n \geq 0$.

Indeed in our previous examples, it can be verified that the system $Tu = u/2$ defined on \mathbb{R} and $Tu = u^2$ defined on $[0, 1]$ the singleton set $\{0\}$ is an attractor. For the system, $Tu = 1 - |2u - 1|$ on $[0, 1]$, one may easily verify that the only attractor is the entire space $[0, 1]$. This is since between any two points $u < v$ in $[0, 1]$, and for any a, b so that $u \leq a < b \leq v$, we can find an n so that $T^n(a, b) = [0, 1]$ (as would be explained later in this chapter).

⁴EDNOTE: M: I have edited here

⁵EDNOTE: B: not relevant other than being an additional example perhaps? M: Have added another simple example

⁶EDNOTE: For let B be an attractor of U . From above, we know that B is itself invariant, is asymptotically stable and has Lyapunov stability. Now suppose there is some attractor $C \subsetneq B$, so there is a $b \in B$ which is not in C M: Have changed the example

A dynamical system can have several attractors, and may also be contained in another attractor. For the example, $Tu = u^2$ on $[0, 1]$, both $\{0\}$ and $[0, 1]$ are attractors. It is a known result that for dynamical systems on a compact space there always exists an attractor.

The dynamics restricted to an invariant set can be complicated. For instance an invariant set could be just a single point or it could have an infinite set. If the invariant set is infinite then there is a possibility of complicated dynamics, and a particular well-studied phenomenon of such complexity gives rise to what is called chaotic behaviour, and has been studied in detail over the last fifty years.

2.0.2 Chaos

EdN:7 ⁷

In the 1960s, a number of mathematicians and mathematically interested scientists independently discovered chaos in the mathematical sense. The meteorologist Edward Lorenz may have been the first to explain this phenomenon in his 1963 paper [?]. The notions of invariance, attractivity and chaos may also be described for continuous systems, and Lorenz's system comprised of a system of differential equations. The narrative of Lorenz's discovery of chaos, as well as the history of other forerunners in this subject, are fascinating. For those who are interested, we highly recommend James Gleick's book *Chaos: The Making of a New Science* [?]. It clearly illustrates these experiences, explains why chaos was such a startling and crucial mathematical and scientific discovery, and describes the underlying mathematical notions for non-specialists.

EdN:8

There are several intricately different definitions of chaos available in the literature, with each of them indicating some aspect of complexity ⁸ In practice, it is not possible to verify which definition or notion does the underlying dynamical system satisfies when only data is observed from it. We do, however, specifically recall the definition of chaos in the sense of Devaney [?, ?] so as to understand some nuances behind the complexity. Devaney's definition of chaos has three requirements, with the first being the notion of sensitive dependence on initial conditions.

Definition 2. A dynamical system $T : U \rightarrow U$ is said to have sensitive dependence on initial conditions (SDIC) if there exists a $\delta > 0$ such that for every $u \in U$ and in every neighborhood of $u \in U$ there exists a $v \in U$ and an integer $N := N(u, v) \in \mathbb{Z}$ such that $d(T^N u, T^N v) > \delta$.

It is important to acquire a sense of this definition as it can easily be misunderstood. It is common misconception to interpret SDIC as two close points (u and v) that eventually become separated by a distance δ under iteration by T . But this is not true. To fully comprehend all the subtleties of the concept, one needs to discuss it more thoroughly by considering each sentence with care and

⁷EDNOTE: should one add paragraphs like these? M: this is a thesis, and not a paper; the instructions sent to the external examiner regarding the thesis evaluation says that the student suggest strongly that it is good to write

⁸EDNOTE: B: what are we saying here? M: Devaney's definition is not the only one, there are about half a dozen definitions as well

attention, whereafter one may examine how they fit together to convey the concept of sensitive dependence.⁹

EdN:9

To this end, we make 3 remarks:

1. First, the $\delta > 0$ in the definition of SDIC is independent of u .
2. Second, in every neighborhood of u , we may not necessarily find all points v in the neighborhood distinct from u that would separate from the forward iterates of u .
3. Finally, N depends upon u and v chosen, and their iterates may not separate for ever (i.e., for all $n > N$) and we allow their iterates to get arbitrarily close in the future.

10

EdN:10

The second part in Devaney's definition of chaos concerns topological transitivity.

Definition 3. A dynamical system $F : U \rightarrow U$ is topologically transitive if for any pair of nonempty open sets E_1 and E_2 there exists a $n \in \mathbb{N}$ such that $T^n(E_1) \cap E_2 \neq \emptyset$.

Topologically transitivity implies that the iterates of an open set of initial conditions gets mixed up with other open sets. On a compact metric space, it can be shown that topological transitivity also implies the existence of a point whose forward iterates are dense [?]; or in other words, the orbit going through this point will be dense in the compact metric space. In fact, it is topologically more likely that the choice of an arbitrary point will be one whose iterates are almost dense.

Definition 4. A subset of a topological space U is said to be a meager set if it can be written as a countable union of sets with empty interior. The complement of the a meager set is set to be a residual set.

To be more precise (see [?]), for a discrete-time dynamical system on a compact space, the set of points with dense iterates are residual, and they are typical or likely to be observed in practice. In this thesis, we assume that when data is observed from a topologically transitive system, we assume that the data arises from a dense orbit.

We may now define the notion of Chaos as formulated by Devaney[?].

Definition 5. A dynamical system $T : U \rightarrow U$ is said to exhibit chaos in the sense of Devaney if it satisfies the three properties:

1. T has SDIC.

⁹EDNOTE: \leftarrow is this talking too simply perhaps? M: No! it is not so

¹⁰EDNOTE: M: It is probably best to see a few numerical simulations to understand SDIC. (Please use the logistic map and perhaps the trajectories of the double pendulum data).

2. T is topologically transitive.
3. The set of periodic points of T are dense in U .

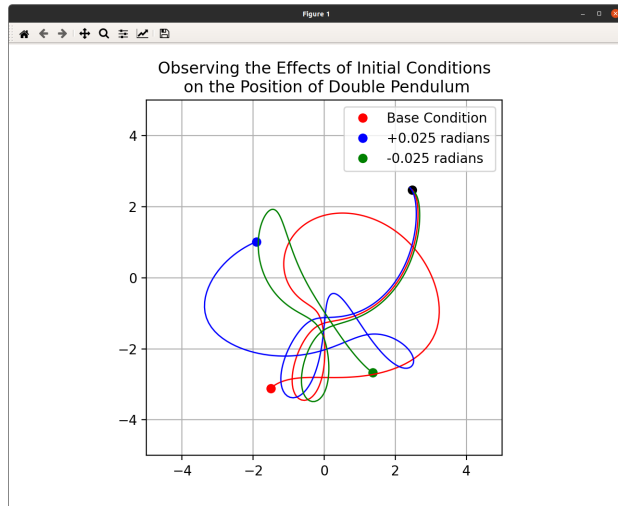


Fig. 1. Starting to diverge

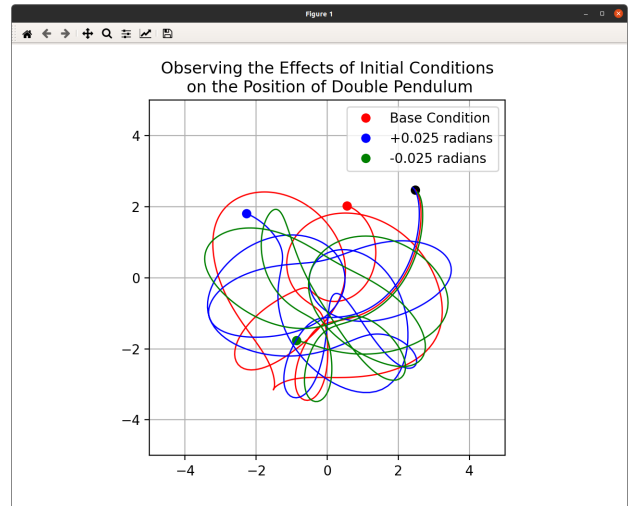
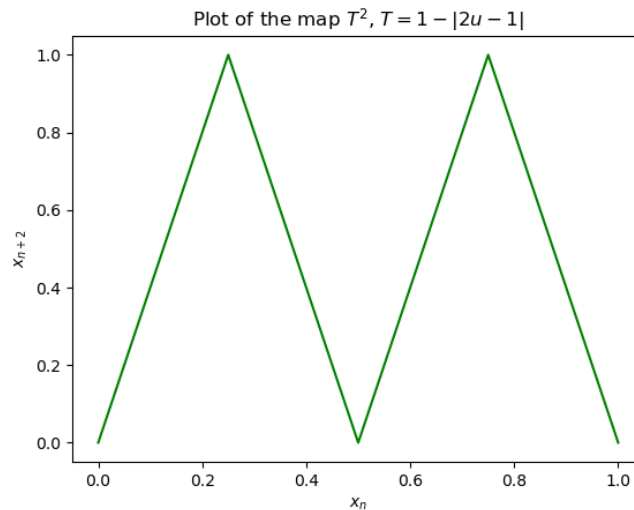


Fig. 2. After more time-steps

EdN:11 11 12

EdN:12

Example 1. The standard tent map $Tu = 1 - |2u - 1|$ defined on $[0, 1]$ is a well known example of a dynamical system satisfying these three properties.



Let us now consider why this map in the above example exhibits chaos in the sense of Devaney. The graph of the map T is piecewise linear with two straight lines, one connecting the points $(0, 0)$

¹¹EDNOTE: M: Please put this picture in perspective and refer after the definition of SDIC. You can also say that you pick two positions that are slightly different but with the same angular velocity. Also in the text, say that you describe the DP in a later chapter.

¹²EDNOTE: M: I realize that need figures in the eps format to have the commutativity diagrams compile as well.

and $(1/2, 1)$ and the other connecting $(1/2, 1)$ with $(1, 0)$. They form a so-called tent with base centered at $u = 1/2$. The graph of the map T^2 comprises two symmetric tents with their base centered at $1/4$ and $3/4$.

In general, the graph of T^k contains 2^{k-1} tents. Given any point u and an open neighborhood $W \subset [0, 1]$ of u , we can find $k \in \mathbb{Z}$ large enough so that $T^k(W) = [0, 1]$ because we can accommodate a tent whose base is contained in W .¹³ This implies the existence of two points w_1 and w_2 in W so that $d(T^k w_1, T^k w_2) = 1$, and by the triangle inequality at least one of the inequalities $d(T^k u, T^k w_1) > \delta$ or $d(T^k u, T^k w_2) > \delta$ hold when $\delta \in (0, 1/2)$. So T has sensitive dependence on initial conditions.¹⁴

EdN:14

Next, let W_1 and W_2 be two nonempty open sets. Given any open set W_1 , we can find a $k \in \mathbb{N}$ large enough so that $T^k(W_1) = [0, 1]$ because we can accommodate a tent with base contained in W . So $T^k(W_1) \cap W_2 \neq \emptyset$, and thus T has topological transitivity.

Finally, for every open interval (a, b) , the graph of T^k intersects the graph of the identity map on $[0, 1]$. This follows from the fact established above that there exists a $k \in \mathbb{N}$ such that $T^k(a, b) = [0, 1]$. If the intersection point has coordinates of the form (p, p) , the point p is then a fixed point of T^k and therefore also a periodic point of T . Since we have found a periodic point in the interval W , the set of periodic points are dense in T .¹⁵

EdN:15

2.1 Conjugacy

We now turn our attention to the subject of conjugacy that describes when two dynamical systems are equivalent dynamically.

To show topological similarity or sameness between two metric or topological spaces, one needs to establish a homeomorphism between the two spaces. However, in the study of dynamical systems defined on two spaces, establishing a homeomorphism does not indicate that the systems are (physically) related in any way. For instance, the maps $Tu = u^2$ and $Tu = 1 - |2u - 1|$ defined on $[0, 1]$ have totally different behaviour.¹⁶ To find systems that are dynamically similar, one rather needs to establish a kind of dynamical equivalence which we illustrate in a simple commutativity diagram below.

EdN:16

$$\begin{array}{ccc} U & & U \\ & \searrow & \swarrow \\ & V & V. \end{array} \quad (2.1)$$

¹³EDNOTE: B: This last phrase doesn't make sense to me. M: The number of tents grow with k and hence this is obvious; I explained in one of our meetings. Does it make sense?

¹⁴EDNOTE: M: It is advised not to introduced newlines in a thesis. For the presentation, yes

¹⁵EDNOTE: Expand further on reason for density. M: fixed point of T^k is a periodic point, and we have found in an arbitrary interval.

¹⁶EDNOTE: "related" - in what sense am I saying it here? Physically? M: Dynamically

To understand the diagram we note that if we travel “right and then down,” the diagram instructs us to use T (top arrow right) first, followed by ϕ . (right arrow downwards). Consequently the pathway amounts to finding $\phi(T(u))$. If we proceed “down and then right”, the diagram instructs us to apply ϕ (left arrow downwards) first, and then apply S (bottom arrow right). This then amounts to finding $S(\phi(u))$. When the relationships in this diagram hold, we say $\phi(T(u)) = S(\phi(u))$, and we formally denote it as $\phi \circ T = S \circ \phi$.

Definition 6. Consider the dynamical systems $T : U \rightarrow U$, $S : V \rightarrow V$ and suppose the relationship $\phi \circ T = S \circ \phi$ holds. If ϕ is a homeomorphism, then S is said to be conjugate to T and ϕ is called a conjugacy. If we relax the criterion on ϕ and merely require ϕ to be continuous and surjective, ϕ is then a semi-conjugacy between T and S ; S is said to be semi-conjugate to T .

EdN:17 17

When S is conjugate to T , the dynamics of the two systems are in some way “dynamically equivalent” and specifically they are in one-to-one correspondence with one another. However when S is semi-conjugate to T with ϕ a many-to-one mapping, the dynamics on V provides a coarse-grained description of the dynamics on U [?]. When S is semi-conjugate to T , it is also common to call S a factor of T , or conversely that T is an extension of S . In essence, an extension (e.g., [?]) is a larger system capturing all of the important dynamics of its factor.

It is a very hard or nearly an impossible task to establish the existence of a conjugacy or a semi-conjugacy ϕ between two systems [?]. However, one can verify that a function ϕ satisfies the commutativity diagram ???. For example $\phi(x) = \sin(\pi x/2)^2$ is a conjugacy between two systems $Tu = 1 - |2u - 1|$ and $Sv = 4v(1 - v)$ defined on $U = [0, 1]$.

In establishing that two systems are conjugate (semi-conjugate) to one another, we have also shown that one may choose to work with one systems as opposed to the second and still be guaranteed to obtain information on the latter. This is incredibly useful while forecasting the future evolution of dynamical systems.

¹⁷EDNOTE: M: Saying ϕ is then a semi-conjugacy between T and S could be ambiguous as it does not specify which system is on the top in the commutativity. So please edit accordingly

Chapter 3

The Learning Problem

Introduction...

Consider a relatively simple learning problem:

Given (u_0, u_1, \dots, u_m) for $m \in \mathbb{Z}$, a finite segment of an orbit of the map T where T is defined by the update equation $u_{n+1} = Tu_n$, forecast the values u_{m+1}, u_{m+2} where the map T is unknown, given that u_m .

A practical example of this would be - given the sequential coordinates of an object moving in space, predict the future positions of that object. However we are very rarely, if at all, presented with a problem where the entire state information is available to us in the form u_n at some time-step $n \in \mathbb{N}$. Consider then a more involved learning problem:

Suppose we only have the observations $\theta(u_0), \theta(u_1), \dots, \theta(u_m)$ of the true system states u in an unknown dynamical system (U, T) and we wish to predict the values $\theta(u_{m+1}), \theta(u_{m+2})$ and so on. First we recall the method of Takens delay embedding.

In this method we consider a discrete-time dynamical system defined on a nice space (a smooth manifold) that can be obtained as time- K map of a continuous time-dynamical system¹⁸. We make observations from such a system, i.e., we observe the evolution of $\theta(w_0)$, i.e., we observe a finite set of values of $u_n := \theta(w_n) = \theta(Tw_{n-1})$. The sequence $\{u_n\}$ represents a scalar time-series and intuitively θ may be thought to represent a probe inserted into a larger system and itself measuring/extracting only a small part of the greater system state at time $t = n$. Consider for example a thermometer erected to measure the ambient temperature in a local village. This measurement function, the thermometer is capturing a single aspect, the temperature, of a grander dynamical system entailing the present weather of the surrounding area, but even more than that, it is measuring an incredibly

EdN:18

¹⁸EDNOTE: M: edited here

small part of the global weather system.

However, the true state u of a system is seldom, if ever fully known. In almost all cases we can merely insert probes into a system to obtain partial information by measurements taken. Moreover, EdN:19 the process of taking a measurement itself introduces 2 additional difficulties: ¹⁹

- i. A series of measurements over a specific time-interval is inherently a discretisation procedure of the underlying continuous-time dynamical system. (Hence why we restricted our attention to discrete-time systems in the preceding section)
- ii. A probe will never be fully accurate and so the act of measurement introduces a certain measure of numerical noise/inaccuracy.

From here we construct a multidimensional observable using the method of stacking previous observations, i.e., we create delay-coordinate map defined by $\Phi_{k,\theta}(w) := (\theta(T^{-k}w), \dots, \theta(T^{-1}w), \theta(w))$. The essence of Takens theorem is that When k is sufficiently large, we can define a dynamical system on the \mathbb{R}^{k+1} that whose states are $\Phi_{k,\theta}(w)$, $\Phi_{k,\theta}(Tw)$, $\Phi_{k,\theta}(T^2w)$ and this dynamical system is topologically conjugate to the unknown underlying system (W, T) . We recall the Takens delay embedding theorem next.

EdN:20 ²⁰

3.0.1 Taken Embedding Theorem

We define first the concepts of homeomorphism and embedding:

Definition 7. A homeomorphism is a function $f : Z \rightarrow Y$ between two topological spaces Z and Y that is continuous, bijective and has a continuous inverse.

Definition 8. Consider a homeomorphism $f : Z \rightarrow Y$ for $Y \in X$. Z is said to be embedded in X by f .

Takens Theorem states a result that establishes a relationship between the observed and underlying dynamical systems by showing that the concatenation of a sufficiently large number of previous observations into a vector will, under certain conditions, generate a map between the vectors from EdN:21 the respective systems. ²¹ We formulate the theorem from [?].

Theorem 1 (Takens Embedding Theorem (adopted from [?]). Let W be a compact manifold of dimension m , and $d \geq m$ so that $2d$ is an integer. It is a generic property for the pair (T, θ) , where $T : W \rightarrow W$ is a smooth diffeomorphism, and $\theta : W \rightarrow \mathbb{R}$ a smooth function, the map

¹⁹EDNOTE: M: Why is discretization a difficulty? Maybe you replace difficulty by saying that we need consider two aspects

²⁰EDNOTE: M: I have omitted many statements here

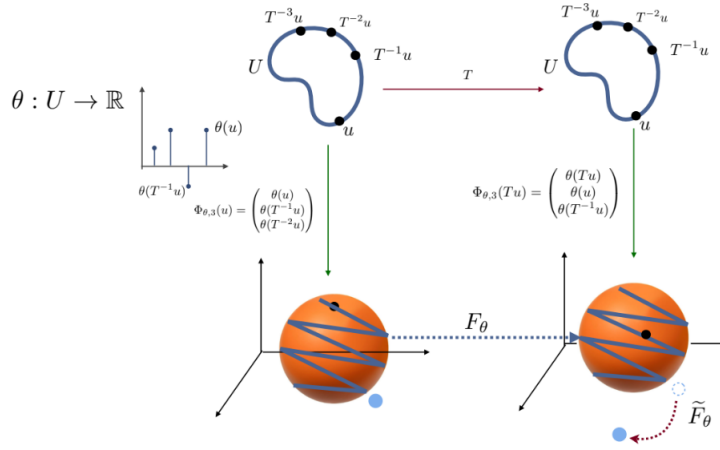
²¹EDNOTE: B: Check wording. M: Edited the statement of the theorem

$\Phi_{2d,\theta} : W \rightarrow \mathbb{R}^{2d+1}$ defined on W by $\Phi_{2d,\theta}(w) := (\theta(T^{-2d}w), \dots, \theta(T^{-1}w), \theta(w))$ is a diffeomorphic embedding; by ‘smooth’ we mean at least C^2 . Consequently, there exists a map $F_\theta : \Phi_{2d,\theta}(W) \rightarrow \Phi_{2d,\theta}(W)$ defined by

$$F_\theta : (\theta(T^{-2d}w), \dots, \theta(T^{-1}w), \theta(w)) \mapsto (\theta(T^{-2d+1}w), \dots, \theta(w), \theta(Tw))$$

so that (W, T) is topologically conjugate to $(\Phi_{2d,\theta}(W), F_\theta)$.

By generic we mean a residual set on a certain topology on an appropriate spaces of functions (that we do not describe here). By C^2 , we mean a function that is twice differentiable, and the second derivative is continuous. (In our scenario, the input space U is considered the attractor.)



22

EdN:22

If we can learn the map F_θ , from sufficient number of data points of $\{\phi_{2d,\theta}(w_n)\}$, then we also know how to forecast how $\theta(w_n)$.

3.0.2 Practical Issues of Takens theorem

Despite the fact that the Takens' Embedding Theorem is a powerful result and provides compelling reason to believe that it is possible to accurately *B: right word? M: Yes, I have edited the sentence* reconstruct a system that is conjugate to the underlying system, it does present some serious practical limitations:

1. Even if we can indeed find F_θ our *approximation* of F_θ is a map from a larger set \mathbb{R}^{2d+1} containing the embedded attractor. There are, however, no theoretical guarantees that F_θ will retain $V = \Phi_{2d,\theta}(W)$ as an attractor although W could be an attractor. ²³

EdN:23

²²EDNOTE: M: Please replace this figure with the one with in the Supplement in Adriaan's paper where w appears instead of u and this would be consistent with the statement in the theorem. Here, and throughout if you reproduce a figure, it must be stated where it was reproduced from.

²³EDNOTE: B: We haven't properly said that U is an attractor, or have we. M: I have rephrased

2. Takens theorem is stated for noiseless observations. Due to noise ϵ_n , the delay vector $\Phi_{2d,\theta}(w_n) + \epsilon_n$, may lie outside V , and further, due to the chaotic nature of the underlying system (i.e. the fact that it has SIDC), the evolution of $\Phi_{2d,\theta}(w_n) + \epsilon_n$ under the map F_θ may move completely out of V . This problem can be overcome by using a driven dynamical system with some properties, and we discuss this in Chapter *.
3. It is an unfortunate case that the functional complexity (i.e., the relative number of oscillatory graphs [?]) as well as stability of the embedding F_θ (cite?) depends on the choice of observable θ . This means that we may even learn a map that complicates our learning task and in so doing adds obstacles to the road to learning a conjugate map with predictive power. ²⁴

EdN:24

Another limitation owes itself to the requirement that T be a surjective map. This is not always the case in practical examples and we may consider the case of a damped double pendulum ²⁵

EdN:25

Continue writing here

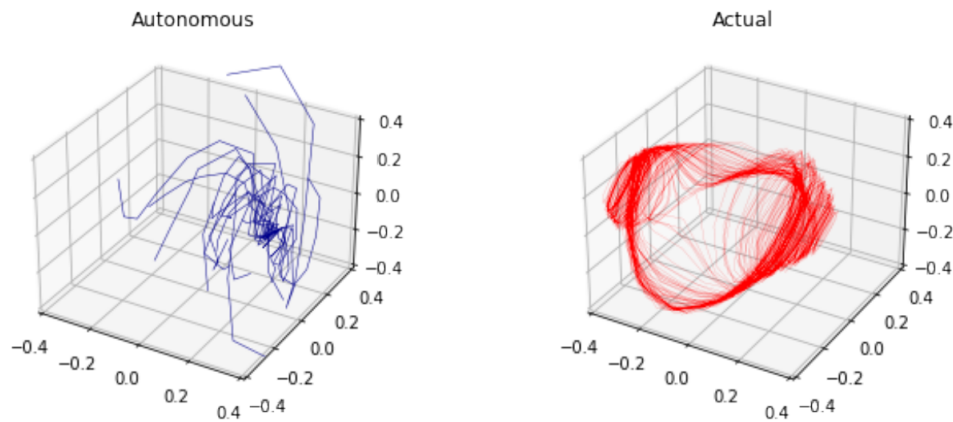


Fig. 1. Illustration of a system having failed to learn the map

Explanation continued here Traalala lorem ipsum et dolore faria maria didi klapogaani koro di tolere mi strem srem krem lem lemmy bemmy bem te dem dem

Continue explaining here

²⁴EdNOTE: Am I talking nonsense here? M: Please avoid this point and delete it

²⁵EdNOTE: M: Please remove the paragraph as well. We do not overcome this problem. You can add this in the next Chapter where you need T to be surjective or else you could include this at the begining of this chapter where you explain the forecasting problem.

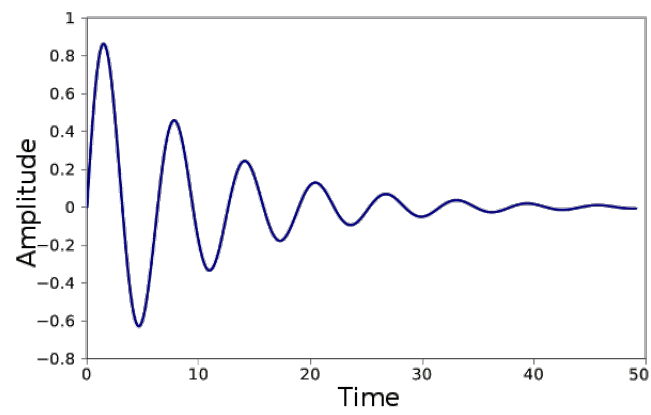


Fig. 2. A damped pendulum