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 ([**?, ?**] 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. **(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 explicitly known to be 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 reconstruction. The project work concerns the study and implementation of a method ([**?**]) 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 ([**?**]), 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 more intricate details of proofs are referred to where relevant throughout. This report is organised as follows:

In Chapter 2, 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.

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

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

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 $T : U \to U$, where U is some set is a *discrete-time dynamical system* and its iterates $\{u, Tu, T^2u, \ldots\}$, 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 frequently drop the brackets and denote $T(u)$ by Tu so as to simplify notations).

Continuous-time dynamical systems modelled using ordinary differential equations can give rise to discrete-time dynamical systems. To see this, consider a differential equation $\dot{x} = f(x)$, $f : \mathbb{R}^n \to \mathbb{R}^n$, $n \in \mathbb{N}$ given to have a unique solution passing through each point $x \in \mathbb{R}^n$, call it $x(t)$ where $x(0) = x_0$.

Definition 1 (**Flow of an Equation**). The flow of the equation $\dot{x} = f(x)$ is defined to be a mapping $\varphi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ where $\varphi(x_0, t) = x(t)$ for the solution $x(t)$ with $x(0) = x_0$

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 mth 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, i.e. $\varphi(x_0, mK)$. Thus ordinary differential equations give rise to a discrete-time dynamical system by sampling the value of the solution $x(t)$ 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

$$x(t + (m + 1)h) = x(t + mh) + hf(x(t + mh)) \tag{2.1}$$

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

$$u_{m+1} = u_m + hf(u_m) \tag{2.2}$$

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. **(Citation)** 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 [**?**].

2.1 Invariant Sets

2.2 Chaos

2.3 Conjugacy

Chapter 3

The Learning Problem

In this chapter we acquaint ourselves with the question of forecasting dynamical systems with unknown underlying structures, state and discuss the shortcomings of the Takens delay embedding theorem and discuss various issues faced while forecasting.

Consider a relatively simple learning problem: Given the sequence $(u_0, u_1, \ldots, u_m)$ 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+1} where the map T is unknown, given that u_m.

A practical example of this would be the subsequent scenario: given the time-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 $\theta(u_0), \theta(u_1), \ldots, \theta(u_m)$ 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 forth.

First we establish the method of Takens delay embedding. In this method, one considers 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 system, concepts which will be formalised below. We make observations from such a system, i.e., we observe the evolution of $\theta(w_0)$ (where w_0 represents the initial state) 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 which is itself only measuring/extracting 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 only a single aspect, the temperature, of a much grander dynamical system entailing the present weather of the surrounding area. Even more than that though- it is measuring a miniscule 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 at most insert probes into a system to obtain partial information by means of the measurements taken. Moreover, the process of taking a measurement itself introduces 2 additional aspects which complicate the problem:

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 $\Phi_{k,\theta}(w) := (\theta(T^{-k}w)\ldots,\theta(T^{-1}w),\theta(w))$. The essence of Takens theorem relates to the fact that when k is sufficiently large, we can define a dynamical system on the space $\mathbb{R}^{k+1}$ whose states are $\Phi_{k,\theta}(w)$, $\Phi_{k,\theta}(Tw)$, $\Phi_{k,\theta}(T^2w)$, etc. and this dynamical system is topologically conjugate to the unknown underlying system (W,T). We recall the Takens delay embedding theorem next.

3.1 Takens Embedding Theorem