Title

B. Nortier

Abstract

Abstract goes here.

1 Introduction

Traditional methods for modelling physical systems involves deriving governing equations by considering the world and discover some fundamental relationships between encountered quantities and phenomena. These fundamental relationships are then expressed mathematically and used as the foundations to derive more complicated relationships. Consider the example of Newton's law F=ma However our world around is, for the most part, much more complex than that which can be distilled into elegant equations. Many of the more complex systems we don't fully understand, nor even have good physical intuition of. To complicate this – many of the physical phenomena we observe in nature are chaotic. Mathematically, this would imply what we call sensitive dependence on initial conditions. Despite having highly similar initial conditions, differing orbits will diverge quickly and by such a large degree that it becomes seemingly impossible to retrace them steps back to the origin conditions.The presence of the smallest of computational noise or measurement error makes long-term pointwise prediction very difficult to minimise. We may however exploit the environment of data-overload we currently find ourselves in. This is one where an incredible amount of information is being gathered and generated on a daily basis. Alongside this we have also witnessed the advent of greater computing power and storage. This project work focusees on constructing governing models for real-world systems in some sense 'blindly', by making use of the abundance of information we have at our disposal. An advantage of this approach is the fact that we need only make a very few starting assumptions about the system we choose to investigate, allowing the methodology it to be greatly generalised. In establishing topological conjugacy between underlying systems we can then, amongst others, establish long-term accurate results.

By solidifying the mathematical underpinnings of our theory, we may guarantee having the ability to construct models with predictive power ranging from molecular biology to neuroscience. The goal of the methodology described here is to be able to predict the dynamics into the future by constructing a model that will accurately mimic the dynamics of a system and hence also provide a framework for future evolution of a dynamical system.

Additional to Intro?
Examples in literature attempting to forecast data from such systems abound: One could model the unobserved states as stochastic quantities. Or one could map the data to a higher-dimensional space by means of delayed-time map of observations and then hope to see some of the the unobserved characteristics of the data revealed.

2 Possible addition or Post-introduction:

Expand on non-autonomous systems, their unique challenges and chaotic dynamics?

Long-term consistency while modeling **chaotic** dynamical systems is an issue. When systems have sensitive dependence on initial conditions, negligible numerical errors multiply within a very short time gap to the point that it becomes unfeasible to track specific orbits. We'll focus on case where attractor is present.

3 Setup and Learning Problem

Consider a relatively simple learning problem:

Given (u_0, u_1, ..., u_m), forecast um+1, um+2, ..., given that u_n is defined by relation $u_{n+1} = Tu_n$.

However, T is unknown.

More involved learning problem: We have θ(u0), θ(u1), ..., θ(um) in an unknown dynamical system (U,T). Forecast θ(um+1), θ(um+2)

Now consider a more refined scenario:

1. a continuous time system with true state u in U (input space)

2. Measured signal = x_n (where: x_n homeo x(u(tn)) + η_n <− noise

We have a scalar time-series.
We construct multi-dimensional observable
How?

1. Delay-coordinates: Given x1,x2,...: define x_n = [x_n,x_n-τ,..., x_n-(d-2)τ,x_n-(d-1)τ].T $\in \mathbb{R}^d$

2. τ=lag

3. d=embedding dimension (park those two for now)

Given: $x' = f(x)$ where $x(t) \in \mathbb{R}$ (x(t) are the states, f is assumed smooth)
Define flow: Φ(x(t0),M)=x(t0+M), Φ:U×R → U
we can relate the two:
d/dtΦ (x(t0),t) = d/dt x(t0+M) = f(x(t0+M)) =f(Φ(x(t0),t)
Now define time-map ΦM : x(t0+M)= ΦT(x(t0))
i.e. x(t0+kM) = Φk
M(x(t0))= ΦM∘ ΦM ∘...∘ ΦM(x(t0) = k-fold composition

BUT: we only get to see a 1D time series θ(x(t)), where θ: R
N→R is a smooth observation/measurement function.

Can information about the system state x(t) be retained in this time series data θ(x(t))?

If we know f, then easily. But we don't.

However: Takens' theorem gives some hope by saying yes!

3.1 Some side-questions

Let's talk about the fact that the manifold U is an attractor etc? Takens only holds for attractors, so that is then why we immediately "restrict" U by only considering the subset containing the attractor right? (ie. Only consider the space occupied by attractor. This is then why we demand that U is compact and T surjective? What about density of attractors? (1)

...

4 Takens Theorem

1. What's that?
   (a) Takens defines delay-coord-map Fθ
      i. Fθ : [θ(T-2du), ... , θ(T-1u), θ(u)] → [θ(T-2d+1u), ... , θ(T-1u), θ(Tu)]
      ii. Where T is flow T$u_{n-1} = u_n$ (previously our flow was Φ)
      iii. θ is measurement function,
   (b) When confining DS to manifold U (in which system is contained), Takens showed that, if certain smoothness conditions are satisfied on T and θ then delay-coord map Fθ is embedding of U onto the reconstruction space for *almost every choice(2)* of measurement function θ (the observable)
   (c) Or alternatively: Takens' embedding theorem guarantees that almost all dynamical systems can be reconstructed from just one noiseless observation sequence.
   (d) i.e. for great number of possible observation functions θ, Fθ preserves the topology of U
   (e) *See Graph 1*

i. If we are in a state u, and then evolve towards state Tu before embedding into R(2m+1), then it is the same as embedding into R(2m+1) and then evolving through the map Fθ

(f) Thus: information about U can be retained in the time series' (observation) output. By preserving the topology of the manifold U in the reconstruction space X, topological invariants of manifold are preserved, of which dimensionality is one (again considered later important later. (3)

2. Problems with Takens

(a) Even if we can find Fθ, our approximation of Fθ is a map from a larger set R2m+1 containing the embedded attractor. No theoretical guarantees that Fθ will keep this attractor the same. (should have defined attractor earlier?)

(b) Takens only holds for noiseless observations.

i. Assume F was learnt but due to measurement error/noise we get V, not an attractor, then will tend somewhere else. This problem will later be removed by some properties of g (USP/UAP)

(c) Takens only holds if T is invertible.

(d) Functional complexity and stability of embedding Fθ depend on choice of observable θ

Now: Although we cannot necessarily find a map T:U $\rightarrow$ U, there does exist a map from Φ: Φ2d,f(um) $\rightarrow$ Φ2d,f(um+1) whenever U is a manifold of dimension less than d (should we say more about d?)

Graph 2 (In sketch: W==U)

We're one step closer to recovering information about the system states

5 Problem

We're trying to approximate a map that might not even exist.

Goal: predict dynamics into future and construct model from data which mimics dynamics.

6 So how do we guarantee existence?

1. Assumptions

   (a) assume a compact, discrete time, dynamical system (U,T) (implicitly T cts)
      i. T must be surjective. (This is important, we return to it later)
   (b) A driven dynamical system g, who takes input in U. (implicitly g is cts)
      i. g will be randomly initiated neural net <– wil explain why later
      ii. g(u_n-1, x_n-1) = x_n
   (c) Bi-infinite sequence u <– physically: running for long time

2. We can easily show existence of the function g:
   (a) g = identity on X
   But then: g ignores the input u and any constant bi-infinite sequence would be a solution to input u.
   (b) However this gives trivial soln, can we find something better?
   (c) Define: reachable soln space XU, entire solutions, YT
      i. Give own example of entire sol'n
      ii. *Graph 3*

3. Choosing g
   (a) Choosing g st "not quench temporal structure in u"
   (b) (4)
   (c) g must be cts, invertible (for SI-invertibility)
   (d) Specific w.r.t tanh:
   Describe set of all solutions of g for given input u

7 Can we show uniqueness of an entire solution?

Graph 4

1. UAP: regardless of starting position, all trajectories converge to a single trajectory. This trajectory is the unique solution sequence x to the input sequence u from the previous slide.

2. USP: *insert more info here*

   (a) The USP - we can start the system anywhere and in time it will forget
   (b) No new complexity
   (c) USP ≡ g is topological contraction ≡ UAP (we'll show this)
      i. 1. Define topological contraction

3. but infinite sequences are difficult to manage
   (a) i. If both SI-invertibility and USP hold for g, then our causal mapping becomes much simpler: we can drop the infinite sequence to focus only on the most recent two terms. (Call it H2)
   AND H2 becomes embedding

8 Establishing Embedding and Conjugacy

1. Define: Embedding, Causal Embedding, Topological Conjugacy
   (a) Explain conjugacy more perhaps?
2. *Graph 5*
3. SI-invertibility, guaranteeing us the function G_T on X×X. <— theoretical basis to try approximate!
4. If g is SI-invertible and u_n in U is an orbit of T then:
   (a) G_T exists
   (b) if g also has USP then (YT, G_T) top.semi-conjug to ($\hat{U}$, $\hat{T}$)
   (c) if T:U → U is homeo then (YT, G_T) is top.conjug to ($\hat{U}$, $\hat{T}$) and top.conjug to (U,T) so g is causal embedding
   (d) WE can learn single-delay dynamics
      i. forecast future values of G_T
      ii. forecast future values of u_n
5. Question: why can one learn single-delay lag dynamics of the driven states through G_T with **enough** data?

9 Why G_T?

Why do we do this?

1. Essentially embed attractor into higher dimensional space X×X.
2. Hence: more 'dimensional room' for the dynamics to move, and we might hope that the dynamics of G_T is in some sense simpler than that of T.
   OR
   Embed into state space with much higher dimension s.t. unobserved features of sequential iput data (u_n) can emerge in (x_n)

10 Progression to Γma instead of G_T

1. Define Γ:(x_n-1, x_n) $\rightarrow u_n$
   (a) Existence? Γ exists when G exists
2. New conjugacy
   Graph 6
3. Why not G_T?
   (a) dimension on which G_T is defined can be high (wasn't it the goal to increase dimension?)
   (b) If we know Γ then we can drive the system autonomously.
   And then we know G_T anyway
   (c) Also: numerical errors could accumulate when working
   Graph 7

11 Programmatically

1. Implementation
   (a) Explain the choice of g
   (b) Explain DC, SDD, MDD, AMDD
   (c) Sparse matrices
   (d) Explain why RNN is chosen (above, for example, FNN)

2. Some Examples
   (a) Double Pendulum successes
   (b) Data dying down (ie. Not surjective and make more remarks here on surjectivity as promised in the beginning)
   (c) Attractors from previous article such as Lorenz, Henon
   (d) Other attractors I've experimented with.

12 Comparison with/Mention of previous work

1. Mention previous work, such as SyndiPi, and that it doesn't work
2. Data-driven approaches alongside machine-learning algorithms have outperformed the classic delay-coordinate embedding as formulated by Takens in the prediction of chaotic dynamical systems' evolution.

13 Further Work

MD

14 More things to Add

Unsure here whether it's important

1. Defining attractors
2. Instability due to perturbation and how prevented
3. Defining a process?
4. *Graph 8*
5. State contractions and their importance
6. More actual theorems
7. Worry:

(a) So far I have very little "mathematical" theorems or proofs. Is this a problem
This will, however, probably be added when expanding on attractors and USP

8. Should I show/state that USP == UAP (equivalent)?