Chapter 7

Delay Coordinate Embedding

Up to this point, we have known our state space explicitly. But what if we do not know it? How can we then study the dynamics is phase space? A typical case is when our data is a scalar time series (e.g., temperature measured at uniform time intervals in a specific geographic point). Let us consider a continuous time systems with true state \( u \in S \), where \( S \) is our phase space.\(^1\) Then our measured signal is

\[
x_n \triangleq x(u(t_n)) + \eta_n,
\]

where \( \eta_n \) is measurement noise. If we have uniform time sampling \( \Delta t \) (which is a generic case), \( t_n = n \Delta t \). Thus, we usually have a time series

\[
\{x_n\}_{n=1}^N = \{x_1, x_2, x_3, \ldots, x_N\}.
\]

\(^1\)In general, \( S \) is not even known precisely.

---

Example 1: Driven Pendulum

We consider a driven pendulum in Fig. 7.1, described by the following equation:

\[
\ddot{u} + \gamma \dot{u} + \sin u = F \cos \omega t,
\]
or

\[
\begin{align*}
\dot{u} &= v, \\
\dot{v} &= -\gamma v - \sin u + F \cos \theta, \\
\dot{\theta} &= \omega,
\end{align*}
\]  

(7.4)

or

\[
\dot{u} = f(u),
\]

(7.5)

where u = [u, v, \theta]^T \in S^1 \times \mathbb{R}^1 \times S^1. Now let us assume that we measure only the first coordinate u as

\[
x = h(\Pi_1(u)) = cu,
\]

(7.6)

where h : S \to S is the angular position sensor function, \Pi_1 : S \times \mathbb{R} \times S \to S is the projection onto the first coordinate u \in S, and c is the linear sensitivity (c = h'(u*)).

In this example we know S, and the question is if we can reconstruct the dynamics in the phase space and estimate all the relevant system parameters by just having time series of x.

**Example 2: Kicked Elastica**

As an another example we consider an experimental system [8] shown in Fig. 7.2. Here, we have a cantilevered elastic rod with a permanent magnet mounted at its free end. The kicker magnet located near the free end of the rod applies transverse load to the rod tip as it passes over it. The rod displacement can be expressed in terms of its normal modes as

\[
u(s, t) = \sum_{n=1}^{\infty} q_n(t) \varphi_n(s),
\]

(7.7)

where \varphi_n : [0, L] \to \mathbb{R} are normal modes of the system satisfying the boundary conditions, and the q_n(t) are the corresponding time coordinates. The corresponding velocity is \nu = \dot{u}. Therefore,
\( \mathbf{u} = [u, v]^T \in \mathcal{S} \) is the \( \infty \)-dimensional function space. We measure strain \( x(t) \) at \( s = s^* \). The strain \( x \propto \kappa(s^*) = cu''(s^*) \), where \( \kappa \) is the curvature of the rod and the approximation is valid for small \( u \) and \( u' \). The question here would be can we use measure strain time series \( \{x\}_{n=1}^N \) to reconstruct the observed dynamics of the elastic rod. Namely, can we determine the number of the active modes and their contribution to the response at various value of the supply voltage to the kicker magnet? Can we develop a bifurcation diagram while varying supply voltage to the kicker magnet? etc.

**Example 3: Walking Human**

We have seen many examples of the studies on humans kinematics [33, 37] shown in Fig. 7.3, where their gait dynamics are recorded using goniometers (for joint angles) or other instruments (such as video based motion capture). Here, the state of human locomotion \( \mathbf{u} \in \mathcal{S} \) is unknown. We do not even know that is our phase space \( \mathcal{S} \), which might be collection of joint angles, their velocities, plus many unknown variables. We would usually measure some joint angle \( x = c\theta \), where \( \theta \) is the true joint angle, or track motion using video based system, etc.

![Figure 7.3: Examples of studies relating to tracking and identifying human kinematics: elite cyclist on the left [37] and soldier walking on the right [33].](image)

A pertinent question, when we do not know \( \mathcal{S} \): is it reasonable to believe that one even exists? Additionally, can we use the measured kinematic variable to track and predict other physiological processes like muscle fatigue or oxygen consumption?
7.1 Delay Reconstruction

Given the scalar measurement \( \{x_n\}_{n=1}^N \), we can attempt to form a multi-dimensional observable using delay reconstruction. That is, given \( \{x_1, x_2, \ldots\} \), we define:

\[
x_n = [x_n, x_{n-\tau}, \ldots, x_{n-(d-2)\tau}, x_{n-(d-1)\tau}]^T \in \mathbb{R}^{d \times 1},
\]

(7.8)

where \( \tau \) is the lag or delay time \((\tau \Delta t)\) and \( d \) is the embedding dimension. A simple illustration of three-dimensional reconstruction is shown in Fig. 7.4.

![Figure 7.4: Illustration of delay reconstruction of scalar time series into a three-dimensional phase space trajectory using \( \tau \) delay time](image)

Why do we thing that the delay reconstruction might work? Consider

\[
x_n = [x_n, x_{n-\tau}, x_{n-2\tau}]^T \in \mathbb{R}^{d \times 1},
\]

(7.9)

If our time series are from a deterministic system then \( u(t_n - \tau) = F(u(t_n)) \).\(^2\) Thus,

\[
u(t_n - 2\tau) = F(u(t_n - \tau)) = F(F(u(t_n))) = F^2(u(t_n)),
\]

(7.10)

and

\[
x_n = G(u(t_n)) \equiv G(u_n).
\]

(7.11)

\(^2\)Think of \( u(t_n) \) as an initial condition and integrate backwards.
7.1. DELAY RECONSTRUCTION

Or, at least, we might hope to be able to do so. To have any hope of this working, we need to get the dimension \( d \) and delay \( \tau \) right.

In summary, as shown in Fig. 7.5, our deterministic trajectory \( \varphi_t(u_0) : S \to S \) evolves on a manifold \( A \subset S \), with \( \dim A = k \). We are looking for a delay reconstruction \( \Phi : S \to \mathbb{R}^d \), which is a composition of the measurement \( x : S \to \mathbb{R} \) and the embedding procedure \( e : \mathbb{R} \to \mathbb{R}^d \). The delay embedding will map \( u_0 \) through \( x_0 = \Phi(u_0) \) to \( x_0 \), and the embedding will have a trajectory \( \tilde{\varphi}_t(x_0) \).

The big question is: when will this process work properly? We would like all the topological properties to be preserved (i.e., dimension and type of orbit), as well as key dynamical properties like stability, and more generally we would like \( \tilde{\varphi}_t \) to be “like” \( \varphi_t \).

The big answer developed by Whitney, Takes, Sauer, at al. is:

**Embedding Theorem:** Consider an invariant manifold \( A \subset S \) with capacity (i.e., box counting) dimension \( D_F \). Then almost every\(^a\) smooth\(^b\) \( x : S \to \mathbb{R} \) and delay time \( \tau \), the delay map \( \Phi(x, \varphi, \tau) : S \to \mathbb{R}^d \) with \( d > 2D_F \) is an embedding of \( A \) into \( \mathbb{R}^d \) if:

1. There are no periodic orbits with period at \( \tau \) or \( 2\tau \) (in \( A \)).
2. There are only finite number of periodic orbits with period \( n\tau \) (\( 2 \leq n \leq d \)).
3. There are a finite number of equilibria.

\(^a\)i.e., with probability 1.
\(^b\)That is, \( x \in C^1 \)

This is the *fractal delay embedding theorem* by Sauer, Yorke, and Casdagli (1991) \([6]\). To understand this theorem, we need to clarify some terms.

**Definition:** An embedding is a map \( F \) of \( A \) manifold that is a \( C^1 \) *diffeomorphism* (one-to-one, onto, and invertible) for which the Jacobian \( DF \) has full rank everywhere in \( A \).
Schematic of an embedding is shown in Fig. 7.6. Please note that manifold $A$ can be distorted by the embedding $F$.

We still need to understand what the capacity dimension $D_F$ is, which we will address in Dimension Theory chapter. However, right now we can consider a case when $D_F = D$ a “normal” integer dimension. Then the requirement that $d > 2D$ (or $d \geq 2D + 1$) can be understood with the cartoons shown in Fig. 7.7. In Fig. 7.7(a), when $m < D$, we clearly do not have large enough room for the embedding. In Fig. 7.7(b), when $m = D$, we see that the mapping cannot be one-to-one globally (though locally we are ok). In Fig. 7.7(c), when $m = 2D$, we might have an embedding $F$ if we are lucky (top case), but generically we cannot guarantee one-to-one mapping due to expected intersections caused by projection (bottom case). In addition, there are infinite number of “nearby” $F$ that fail too. Finally, in Fig. 7.7(d), when $d = 2D + 1$, we have an embedding that is geometrically ok and preserves the original trajectories dynamical characteristics. Cartoons in Fig. 7.7 are actually describing the Whitney Embedding Theorem [9], which the previously discussed Embedding Theorem generalizes to fractal dimensions.

### 7.1.1 Some Remarks

- In practice, we do not know $D_F$—that is one thing we may be trying to find. So the question is how to estimate $d$?

- The Delay-Embedding theorem says $\tau$ not important (except to avoid periodicities). In practice, this is not true: $\tau$ too small will collapse $A$ onto hyper-diagonal of the embedding space and loose “attractor” in noise. With chaos, if $\tau$ is too large, successive $x_n$ are approximately uncorrelated and the deterministic structure of the attractor may be obscured.
7.1. DELAY RECONSTRUCTION

 Embedding Theorem is only the sufficient condition \((d \geq 2D_F + 1)\), but may not be necessary and we may achieve embeddings at lower dimensions.

In summary, if the conditions of the Embedding Theorem are satisfied, we have the commutator diagram shown in Fig. 7.8, where \(x\) is the measurement function and \(e\) is the delay-reconstruction procedure. Thus, the true dynamics can be expressed as:

\[
\mathbf{u}(t) = \varphi_t(\mathbf{u}_0)
= \Phi^{-1} \circ \hat{\varphi}_t \circ \Phi(\mathbf{u}_0),
\]

that is

\[
\varphi_t = \Phi^{-1} \circ \hat{\varphi}_t \circ \Phi.
\]

Therefore, in this sense, the dynamics in \(\mathbb{R}^d\) and \(\mathcal{A}\) are equivalent. Further, if we study, estimate, model \(\hat{\varphi}_t\) on \(\mathbb{R}^d\), we know important properties of \(\varphi_t\) on \(\mathcal{A}\).

![Figure 7.7: Cartoon illustration of the Whitney Embedding Theorem](image)

Figure 7.7: Cartoon illustration of the Whitney Embedding Theorem

- Embedding Theorem is only the sufficient condition \((d \geq 2D_F + 1)\), but may not be necessary and we may achieve embeddings at lower dimensions.

In summary, if the conditions of the Embedding Theorem are satisfied, we have the commutator diagram shown in Fig. 7.8, where \(x\) is the measurement function and \(e\) is the delay-reconstruction procedure. Thus, the true dynamics can be expressed as:

\[
\mathbf{u}(t) = \varphi_t(\mathbf{u}_0)
= \Phi^{-1} \circ \hat{\varphi}_t \circ \Phi(\mathbf{u}_0),
\]

that is

\[
\varphi_t = \Phi^{-1} \circ \hat{\varphi}_t \circ \Phi.
\]

Therefore, in this sense, the dynamics in \(\mathbb{R}^d\) and \(\mathcal{A}\) are equivalent. Further, if we study, estimate, model \(\hat{\varphi}_t\) on \(\mathbb{R}^d\), we know important properties of \(\varphi_t\) on \(\mathcal{A}\).

![Figure 7.8: Commutator Diagram for the Delay Embedding](image)

Figure 7.8: Commutator Diagram for the Delay Embedding
7.2 Phase Space Reconstruction

Two basic parameters need to be determined for delay coordinate embedding: delay time $\tau$ and the embedding dimension $d$. The delay time should be large enough that each coordinate of the reconstructed phase point is distinct and not redundant. Taking a too small delay results in delay vectors close to the hyper-diagonal of the reconstructed phase space and any variations transverse to it are not well defined. At the same time, for a chaotic response, the delay should small enough that coordinates are not statistically independent of each other and the attractor geometry is not complicated or obscured. For example, the ideal delay for a harmonic signal would be a quarter of its period. This suggests that the first zero crossing of the auto-covariance is a good choice for the delay (e.g., for a harmonic signal this happens at quarter of the period as shown in Fig. 7.9), but for chaotic time series this would generally overestimate the needed delay since it cannot account for nonlinear correlations.

Average mutual information (AMI), being a measure of nonlinear correlations, in the data provides another alternative. Its first minimum, when identifiable is usually a good estimate of optimal delay as it identifies a delay at which a new coordinate provides maximal new information compared to its delayed version while still being non-linearly correlated to it. However, AMI estimates are very sensitive to noise, which usually makes it impossible to identify the first minimum in AMI data even if it exists. Another alternative is the use of characteristic length spectrum (CLS), which is highly robust to noise. The simplest of all methods is to just look at the reconstructed phase portrait and, when possible, identify the delay that provide a good cover of space without causing complex geometry.

The selection of the embedding dimension can be guided by the embedding theorem, which stipulated that any dimension higher than twice the capacity dimension of the attractor provides a good embedding. However, we do not usually know system’s capacity dimension a priori, and need to
Figure 7.10: Basic idea of false nearest neighbors is illustrated by a three-dimensional closed curve (blue line) and its two-dimensional projection (red line). The thin dotted lines show two-dimensional projections onto other planes. The points that are false neighbors in the projection \((A'\text{ and } B')\) are far apart in higher dimensions \((A \text{ and } B)\), while the true neighbors \((C' \text{ and } D')\) remain close neighbors \((C \text{ and } D)\).

estimate it using the measured data and the corresponding reconstructed trajectory. In addition, this is only a sufficient condition and there may as well be perfectly good lower-dimensional embeddings. The necessary condition on the minimally needed embedding dimension is usually provided by the false nearest neighbor (FNN) method.

The idea behind the FNN method is simple (see Fig. 7.10): if two points in \(d\) dimensions are true nearest neighbors to each other, then adding \((d + 1)\)-th coordinate should not separate them from each other. If, however, they were nearest neighbors only due to the projection from a higher dimensional structure they will separate from each other in \((d + 1)\)-dimensional space and thus would constitute false neighbors in \(d\) dimensions. The minimum embedding dimension is identified by by the smallest dimension having zero fraction of the FNNs.
Problems

Problem 7.1

Consider the Hénon map [20],

\[ x_{n+1} = a - x_n^2 + by_n, \quad y_{n+1} = x_n, \]

which yields chaotic solutions for \( a = 1.4 \) and \( b = 0.3 \).

1. Using a typical sequence of \( x_n \), create different two dimensional phase portraits using delay times \( \tau = 1, 2, \ldots \).

2. Which picture gives the clearest information about the original system? Why?

3. Rewrite the map in delay coordinates with unit delay and interpret the results.