Spectral methods in dynamics

Vaughn Climenhaga

January 30, 2013

In the dynamics seminar here at Houston, we're beginning a series of expository talks on statistical properties of dynamical systems. This week's talk was given by Andrew Török and introduces some of the spectral methods for transfer operators that prove useful. This post is basically a set of notes from that talk, aiming to give an informal and accessible introduction to this topic.

1 Observables, invariant measures, and mixing

We consider dynamical systems $T\colon X \to X$ that are expected to exhibit some sort of "chaotic" behaviour. Here X is generally a compact metric space (today we'll mostly consider $X = [0,1]$) and T is continuous, or at least piecewise continuous.

A measurable function $\varphi\colon X \to \mathbb{C}$ is called an "observable". The idea is to consider the time series of functions $\{\varphi \circ T^k\}_{k \geq 0}$, which for "chaotic" systems resembles a sequences of random variables, with $\varphi \circ T^k$ representing the observation φ made at time k. But just how random is this sequence? Is there a sense in which these random variables can really be treated as random, independent, uncorrelated?

This is the central question in studying statistical properties of dynamical systems. There are various results that hold for sequences of independent identically distributed (IID) random variables, and it turns out that these same results hold for many systems of interest.

Let's make this a little more precise. If we flip a coin repeatedly (a fundamental example of an IID sequence), the strong law of large numbers

says that "almost surely" (that is, with probability 1), the fraction of flips that turn up heads will converge to $1/2$. One also has the central limit theorem, which says that with the appropriate averaging, the probability distribution that governs the number of heads that appear will converge to a normal distribution.

The sequence of observations $\varphi \circ T^k$ is certainly not IID – there are strong correlations between φ and $\varphi \circ T^k$ for small values of k. However, there are many examples where this correlation decays as $k \to \infty$. For such examples it is reasonable to ask if probabilistic results such as the strong law of large numbers and the central limit theorem hold.

In order to make probabilistic statements – that is, to treat $(X, \varphi \circ T^k)$ as a probability space with a sequence of random variables – we must first specify what measure we place on X. Let μ be a Borel probability measure on X. The measure μ is said to be *invariant* if $\mu(T^{-1}A) = \mu(A)$ for every Borel $A \subset X$. If we interpret A as the event $x \in A$ (occuring at time 0), then the requirement of invariance amounts to the condition that the probability of the event A is the same at time 0 (that is, $x \in A$) as it is at any later time k (that is, $T^k x \in A$). Thus the sequence of random variables $\varphi \circ T^k$ on (X, μ) is identically distributed, and we only need to worry about the fact that independence fails.

An equivalent definition of invariance is that $\int \varphi \, d\mu = \int \varphi \circ T \, d\mu$ for every $\varphi \in L^1(X, \mu)$. Informally, this means that the expected value of the observables $\varphi \circ T^k$ are all the same – one obtains the same expected value for an observation whether it is made at time 0 (that is, $\int \varphi \, d\mu$) or at some later time k (that is, $\int \varphi \circ T^k \, d\mu$). This duality between definitions in terms of sets and definitions in terms of functions will continue to appear throughout our discussion.

In fact, not much more is needed to obtain the first probabilistic statement, the strong law of large numbers. An invariant measure μ is a ergodic if any (and hence all) of the following three equivalent conditions hold:

1. every invariant set $(A = T^{-1}A)$ has measure 0 or 1;

2. every invariant function $(\varphi = \varphi \circ T)$ is constant μ-a.e.;

3. μ cannot be written as a convex combination of two other invariant measures.

The strong law of large numbers holds for any ergodic measure μ and integrable observable $\varphi \in L^1(X, \mu)$ – this is the Birkhoff ergodic theorem.

What about the central limit theorem? For this we need something more. Consider two events $A, B \subset X$. If the state of the system at time k was completely independent of the state at time 0, then we would have $\mu(A \cap T^{-k}B) = \mu(A)\mu(B)$. Because we consider deterministic systems in which there are short-term correlations, we do not usually have equality in this relation – nevertheless we can ask for it to hold asymptotically, so that

$$\lim_{k \to \infty} \mu(A \cap T^{-k}B) = \mu(A)\mu(B). \tag{1}$$

A measure satisfying (1) is called *mixing*. An important part of establishing the statistical behaviour of a system with respect to this measure is to understand the *rate* of mixing – how quickly do the correlations decay?

The same duality mentioned above happens here as well – mixing can be characterised in terms of functions (instead of sets) as the condition that

$$\lim_{k \to \infty} \int (\varphi \circ T^k)\psi \, d\mu = \left(\int \varphi \, d\mu \right) \left(\int \psi \, d\mu \right) \tag{2}$$

for every $\varphi, \psi \in L^2(X, \mu)$. It turns out that this is a more useful formulation when we wish to compute rates of mixing. The idea is that the convergence in (2) happens exponentially quickly for observables φ, ψ taken from a "reasonably nice" function space, while for more general classes of observables (such as arbitrary elements of L^2), and also for the setwise-defined mixing in (1), the convergence can be arbitrarily slow.

For the time being, then, we can describe our goal as follows. The dynamical system (X, T) may have a great many invariant measures, and by restricting our attention to the ergodic measures, we guarantee that the strong law of large numbers holds thanks to the Birkhoff ergodic theorem. Within the class of ergodic measures, we want to find measures for which the correlations between the observables $\varphi \circ T^k$ decay exponentially, and for which a central limit theorem (and hopefully other statistical laws) can be proved.

So far there has been no mention of anything "spectral", despite the title at the top. The comment above, that we will ultimately need to restrict to some "reasonably nice" function space, suggests how it will enter. To the map $T \colon X \to X$ is associated an induced operator on various function spaces. We will study the spectral properties of this operator, and by choosing an appropriate function space, we will be able to use these properties to deduce various statistical properties.

3

2 Examples

First, though, it is high time for some examples. Both of our chief examples at the present time are defined on the unit interval $X = [0, 1]$. The first example is the doubling map $T \colon x \to 2x \pmod 1$ shown in Figure 1. This map preserves Lebesgue measure and has $T'(x) = 2$ everywhere. (More precisely, almost everywhere, since T is discontinuous at $x = 1/2$.)
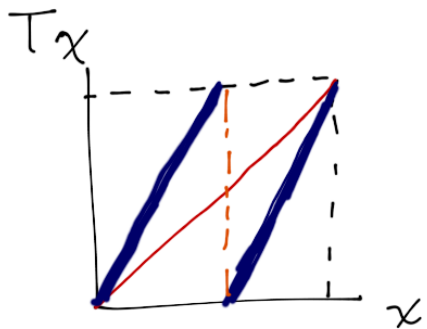

Figure 1: The doubling map $T \colon x \to 2x \pmod 1$.

The second, more general, example, is a piecewise expanding interval map such as the one shown in Figure 2. Here again $X = [0, 1]$, but now $T \colon [0, 1] \to [0, 1]$ comes from a more general class – the interval $[0, 1]$ is partitioned into finitely many intervals, and on each of these the map T is C^2 with $|T'| \geq \lambda > 1$. (For the reader familiar with such maps, note that there is no Markov assumption.)
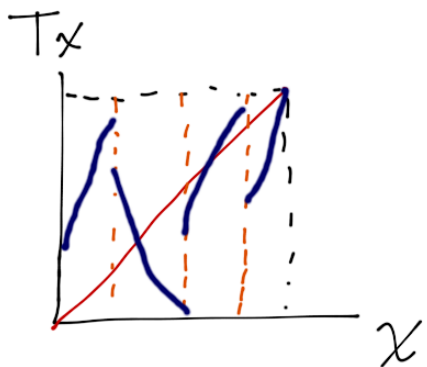

Figure 2: A piecewise expanding interval map.

Lebesgue measure is a natural choice for the first example: it is invariant, it can be shown to be ergodic, and when we speak of "choosing a point at random from the interval", the most natural interpretation is that we mean at random with respect to Lebesgue measure.

What invariant measure should we use for the second example? The map T certainly appears to be a plausible candidate for chaotic behaviour – nearby points are driven apart exponentially fast – but in order to describe the statistical properties of the system we need an invariant measure. In particular, we would like an invariant measure μ that is absolutely continuous with respect to Lebesgue measure, so that a result that is true for μ-a.e. point x will also be true for Lebesgue-a.e. x. So how do we find an absolutely continuous invariant probability measure? (The abbreviation acim or acip is often used.)

Let $\mathcal{M}$ denote the set of absolutely continuous probability measures on $X = [0, 1]$, not necessarily invariant. Elements of $\mathcal{M}$ correspond to functions $\psi \in L^1(X, \text{Leb})$ by the Radon–Nikodym theorem: the association is $d\mu = \psi \, dx$. We want to understand the maps that the dynamics T induces on both the space of measures $\mathcal{M}$ and the space of densities L^1 – invariant measures (densities) will correspond to fixed points of this map.

3   The transfer operator

To understand the dynamics that T induces on $\mathcal{M}$, first note that the map T induces an action on observables $\varphi \colon X \to \mathbb{C}$ by sending φ to $\varphi \circ T$ – this is called the *Koopman operator* or composition operator and is an important tool in ergodic theory (once we have an invariant measure). So far, however, we have no invariant measure. Note that a measure $\mu \in \mathcal{M}$ given by $d\mu = \psi \, dx$ is invariant if and only if $\int (\varphi \circ T) \cdot \psi \, dx = \int \varphi \cdot \psi \, dx$ for every φ. In particular, μ is invariant if and only if the corresponding density ψ is a fixed point of the operator $\mathcal{P}_T$ defined on $L^1(X, \text{Leb})$ by the condition

$$\int (\varphi \circ T) \cdot \psi \, dx = \int \varphi \cdot (\mathcal{P}_T \psi) \, dx \qquad \text{for all } \varphi \in L^\infty(X, \text{Leb}). \quad (3)$$

The operator $\mathcal{P}_T$ is called the transfer operator (or *Ruelle operator*), and will be the central object of the spectral methods we are describing. We can view $\mathcal{P}_T$ as the action induced by T on the space $\mathcal{M}$ of absolutely continuous measures. Note that in order for $\mathcal{P}_T$ to be a bounded linear operator on L^∞,

5

the map T must be non-singular – that is, it cannot take a set of positive measure to a set of measure 0.

Formally, (3) defines the transfer operator as the (right) adjoint of the Koopman operator. One can verify that (3) determines $\mathcal{P}_T$ uniquely, and that moreover the action of $\mathcal{P}_T$ can be described explicitly by

$$(\mathcal{P}_T \psi)(x) = \sum_{y \in T^{-1}x} \frac{\varphi(y)}{|T'(y)|}. \tag{4}$$

The utility of $\mathcal{P}_T$ for us is two-fold: in the first place, as pointed out above, absolutely continuous invariant measures correspond to fixed points of $\mathcal{P}_T$ – that is, eigenfunctions of $\mathcal{P}_T$ with eigenvalue 1. Furthermore, by iterating (3), the correlation functions on the left-hand side of (2) can be understood in terms of the iterates $\mathcal{P}_T^k \psi$, and these in turn can be understood in terms of spectral properties of $\mathcal{P}_T$ apart from the eigenvalue 1. This is how knowledge of the spectrum of $\mathcal{P}_T$ will lead to information on decay of correlations and other statistical properties: the eigenfunction corresponding to the largest eigenvalue is the density of the absolutely continuous invariant measure, and the presence of a *spectral gap* (defined below) between this eigenvalue and smaller eigenvalues leads to exponential decay of $\mathcal{P}_T^k \psi$ when $\int \psi \, dx = 0$.

4 Decay of correlations for the doubling map

To illustrate what happens, let us return to the example of the doubling map $T \colon x \to 2x \pmod 1$. Here we can write (4) explicitly as

$$(\mathcal{P}_T \psi)(x) = \frac{1}{2} \left[\psi \left(\frac{x}{2} \right) + \psi \left(\frac{1+x}{2} \right) \right]. \tag{5}$$

First we observe that the constant function $\mathbf{1}$ satisfies $\mathcal{P}_T \mathbf{1} = \mathbf{1}$, which corresponds to the fact that Lebesgue measure itself is invariant for T. So how do we carry out the rest of the programme and use the iterates of $\mathcal{P}_T$ to establish a rate of convergence in (2)?

At this point we must confront openly a point that was mentioned in passing earlier. If we consider $\mathcal{P}_T$ as an operator on L^1, then there is not too much we can do. (More on this later, perhaps.) However, if we make a "good" choice of a Banach space on which $\mathcal{P}_T$ acts, then we will be able to get good results. In some sense this is the central challenge of the spectral

method for studying statistical properties of dynamical systems: to find a good Banach space on which the transfer operator acts with a spectral gap. While we will see momentarily that this is not too difficult for the interval maps we are considering here, it turns out to be a much more substantial challenge when we study more general classes of systems.

In the present setting we may consider the subspace $\mathrm{Lip} \subset L^1$ of all Lipschitz continuous functions on $[0,1]$. There is a natural semi-norm

$$|\psi|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|\psi(x) - \psi(y)|}{|x - y|},$$

which fails to be a norm only because it vanishes on all constant functions. We can define a true norm by

$$\|\psi\|_{\mathrm{Lip}} = \|\psi\|_{\infty} + |\psi|_{\mathrm{Lip}},$$

where $\|\psi\|_{\infty} = \sup_x |\psi(x)|$.

For our purposes, the key property of the semi-norm $|\cdot|_{\mathrm{Lip}}$ is that it is contracted by the operator $\mathcal{P}_T$.

Proposition 1 *For the doubling map T and any Lipschitz function ψ we have $|\mathcal{P}_T \psi|_{\mathrm{Lip}} \leq \frac{1}{2}|\psi|_{\mathrm{Lip}}$.*

PROOF: Using (5) we have

$$|\mathcal{P}_T \psi|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|(\mathcal{P}_T \psi)(x) - (\mathcal{P}_T \psi)(y)|}{|x - y|}$$
$$= \sup_{x \neq y} \frac{1}{2} \frac{|\psi(x_1) + \psi(x_2) - \psi(y_1) - \psi(y_2)|}{|x - y|},$$

where we write $x_1 = \frac{x}{2}$, $x_2 = \frac{1+x}{2}$, and similarly for y_1, y_2. Now $|x_j - y_j| = \frac{1}{2}|x - y|$ and so

$$|\mathcal{P}_T \psi|_{\mathrm{Lip}} \leq \sup_{x \neq y} \frac{1}{4} \left[\frac{|\psi(x_1) - \psi(y_1)|}{|x_1 - y_1|} + \frac{|\psi(x_2) - \psi(y_2)|}{|x_2 - y_2|} \right] \leq \frac{1}{2}|\psi|_{\mathrm{Lip}}.$$

$\square$

Now we can address the question of decay of correlations in (2). Iterating (3) gives

$$\int (\varphi \circ T^k) \cdot \psi \, dx = \int \varphi \cdot (\mathcal{P}_T^k \psi) \, dx. \tag{6}$$

7

Every $\psi \in \text{Lip}$ can be written as

$$\psi = c_\psi \mathbf{1} + \hat{\psi}, \tag{7}$$

where $c_\psi = \int \psi \, dx$ and $\int \hat{\psi} \, dx = 0$. That is, there is a decomposition

$$\text{Lip} = \mathbb{C}\mathbf{1} \oplus H, \qquad H = \left\{ \psi \in \text{Lip} \mid \int \psi \, dx = 0 \right\}. \tag{8}$$

This decomposition is invariant under the action of $\mathcal{P}_T$: indeed, $\mathbb{C}\mathbf{1}$ is invariant because $\mathcal{P}_T \mathbf{1} = \mathbf{1}$, and to see that H is invariant we use (3) to observe that if $\int \hat{\psi} \, dx = 0$, then

$$\int \mathcal{P}_T \hat{\psi} \, dx = \int (\mathbf{1} \circ T) \hat{\psi} \, dx = \int \hat{\psi} \, dx = 0.$$

It is not hard to show that any $\hat{\psi} \in H$ satisfies $\|\hat{\psi}\|_\infty \leq |\hat{\psi}|_{\text{Lip}}$, because its range $\psi(X) \subset \mathbb{C}$ has diameter $\leq |\hat{\psi}|_{\text{Lip}}$ and contains 0 in its convex hull.

Now we can estimate the rate of decay of the correlation quantity

$$C_k(\varphi, \psi) := \int (\varphi \circ T^k) \cdot \psi \, dx - \left(\int \varphi \, dx \right) \left(\int \psi \, dx \right). \tag{9}$$

Using (6) and the decomposition (7) gives

$$C_k(\varphi, \psi) = \int \varphi \cdot (\mathcal{P}_T^k (c_\psi \mathbf{1} + \hat{\psi})) \, dx - c_\psi \int \varphi \, dx$$

$$= \int \varphi \cdot (\mathcal{P}_T^k \hat{\psi}) \, dx,$$

where the second equality uses the fact that $\mathcal{P}_T^k \mathbf{1} = \mathbf{1}$. We estimate $\mathcal{P}_T^k \hat{\psi}$ using Proposition 1:

$$\|\mathcal{P}_T^k \hat{\psi}\|_\infty \leq |\mathcal{P}_T^k \hat{\psi}|_{\text{Lip}} \leq 2^{-k} |\hat{\psi}|_{\text{Lip}} = 2^{-k} |\psi|_{\text{Lip}},$$

and conclude that for any $\varphi \in L^1$ and $\psi \in \text{Lip}$ we have

$$|C_k(\varphi, \psi)| \leq \|\varphi\|_1 \|\mathcal{P}_T^k \hat{\psi}\|_\infty \leq 2^{-k} \|\varphi\|_1 |\psi|_{\text{Lip}}.$$

This means that for the doubling map, the convergence in (2) happens exponentially quickly provided the observables φ, ψ are sufficiently regular.

5 Spectral properties

The above results for the doubling map are a particular case of the spectral method, although we did not yet describe explicitly the role of the spectrum of $\mathcal{P}_T$. Recall that the spectrum of the operator $\mathcal{P}_T\colon \mathrm{Lip} \to \mathrm{Lip}$ is the set

$$\sigma(\mathcal{P}_T) = \{\lambda \in \mathbb{C} \mid \mathcal{P}_T - \lambda I \text{ is not an invertible operator on } \mathrm{Lip}\},$$

which contains (but is not necessarily equal to) the set of eigenvalues of $\mathcal{P}_T$ (the *point spectrum*). We emphasise that this is a very general definition, valid for any bounded linear operator on any Banach space, not just $\mathcal{P}_T$ acting on Lip. A basic fact in functional analysis is that the spectrum is always compact and non-empty.

In the example above, the constant function $\mathbf{1}$ is an eigenfunction with eigenvalue 1, and using this invariant decomposition $\mathrm{Lip} = \mathbb{C}\mathbf{1} \oplus H$ from (8), we have $\sigma(\mathcal{P}_T) = \{1\} \cup \sigma(\mathcal{P}_T|_H)$. That is, apart from the eigenvalue at 1, the spectrum of $\mathcal{P}_T$ is determined by its action on the subspace H.

Recall from functional analysis that if we write $\rho(\mathcal{P}_T) = \sup\{|\lambda| \mid \lambda \in \sigma(\mathcal{P}_T)\}$ for the *spectral radius* of $\mathcal{P}_T$, we have

$$\rho(\mathcal{P}_T) = \lim_{n \to \infty} \|\mathcal{P}_T^n\|^{1/n} \leq \|\mathcal{P}_T\|. \tag{10}$$

To determine the spectrum of $\mathcal{P}_T|_H$ we can use either the Lipschitz norm $\|\cdot\|_{\mathrm{Lip}}$ or the semi-norm $|\cdot|_{\mathrm{Lip}}$, because on the subspace H the semi-norm becomes a norm and the two are equivalent:

$$|\hat\psi|_{\mathrm{Lip}} \leq \|\hat\psi\|_{\mathrm{Lip}} = \|\hat\psi\|_\infty + |\hat\psi|_{\mathrm{Lip}} \leq 2|\hat\psi|_{\mathrm{Lip}}.$$

(This fails outside of H, where to apply (10) we would need to use $\|\cdot\|_{\mathrm{Lip}}$.) From Proposition 1 and (10) we see that $\rho(\mathcal{P}_T|_H) \leq \frac{1}{2}$. Thus the spectrum of $\mathcal{P}_T$ has the structure shown in Figure 3: there is a single eigenvalue at 1, and the rest of the spectrum is contained in the disc with centre 0 and radius $1/2$.

Going beyond the doubling map to such examples as the piecewise expanding interval maps discussed above, the goal is to carry out a similar procedure by finding a suitable Banach space $\mathcal{B}$ of functions on which the transfer operator acts with a *spectral gap*: that is, where there is a single eigenvalue (or at most finitely many) lying on the unit circle, and the rest of $\sigma(\mathcal{P}_T)$ is contained in a disc of radius $\rho < 1$. Then one is able to draw the following conclusions.
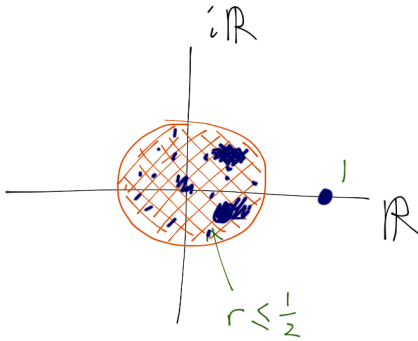
9

Figure 3: A gap in the spectrum of $\mathcal{P}_T$.

1. The eigenfunction(s) corresponding to the eigenvalue 1 are the densities for the absolutely continuous invariant measures.

2. Given any $r \in (\rho, 1)$, there is a constant C_r such that $\|\mathcal{P}_T^k\|_{\mathcal{B}} \leq C_r r^k$, and so the correlations $C_k(\varphi, \psi)$ decay like r^k when the observables φ and ψ are chosen from suitable function spaces.

Eventually it is also interesting to consider a more general class of transfer operators associated to *potential functions* for which the largest eigenvalue may not be 1, but first we will spend some time (in the next few posts) developing the general theory of how one can prove the existence of a spectral gap for piecewise expanding interval maps.