Finite-dimensional sets and finite-dimensional attractors

James C. Robinson

Contents

1

Dynamical systems and global attractors

We will consider attractors for abstract dynamical systems defined on some phase space X (a finite- or infinite-dimensional real Hilbert space in all we do here) with the dynamics given by a semigroup of solution operators $\{S(t)\}_{t\ge 0}$ that satisfy

(i) $S(0) = \mathrm{id}$,
(ii) $S(t)x$ continuous in t and x, and
(iii) $S(t)S(s) = S(t+s)$ for all $t, s \ge 0$.

At other points it will be useful to consider instead a dynamical system that arises from iterating a fixed function $f : X \to X$. Such a system could be derived from a continuous time system if we consider the map $S(T)$ for some fixed $T > 0$.

Such a semigroup can be generated, for example, from a finite-dimensional system of ordinary differential equations,

$$\dot{x} = f(x) \qquad x(0) = x_0 \qquad x \in \mathbb{R}^N,$$

in which case $X = \mathbb{R}^N$ and $S(t)x_0 = x(t; x_0)$.

Or, as a canonical infinite-dimensional example, one can consider the two-dimensional Navier-Stokes equations on a bounded domain $\Omega \subset \mathbb{R}^2$

$$u_t - \Delta u + (u \cdot \nabla)u + \nabla p = f \qquad \nabla \cdot u = 0$$

with an initial condition $u(x,0) = u_0(x)$ and a time-independent forcing function f. In this case if one takes

$$u_0 \in H = \{u : \int_\Omega |u(x)|^2 \, \mathrm{d}x < \infty : \nabla \cdot u = 0\}$$

(i.e. essentially $[L^2(\Omega)]^2$) then there exists a unique solution for all $t \geq 0$ that remains in H.

1.1 Global attractors

In order to prove the existence of a global attractor, one assumes that $S(t)$ is 'dissipative' in some appropriate sense. To do this we make use of the idea of an absorbing set.

A set K is *absorbing* for $S(t)$ if for any bounded subset B of X there exists a time t_B such that

$$S(t)B \subseteq K \qquad \text{for all} \quad t \geq t_B.$$

In a finite-dimensional space, $S(t)$ is dissipative if it has a bounded absorbing set. In an infinite-dimensional space, $S(t)$ is dissipative if

(i) it has a compact absorbing set, or
(ii) it has a bounded absorbing set and $S(t)$ is 'asymptotically compact': for any bounded sequence x_n and and sequence $t_n \to \infty$, $S(t_n)x_n$ has a convergent subsequence.

A dissipative dynamical system has a global attractor $\mathscr{A}$, that is a compact, invariant, attracting set, i.e.

(i) $\mathscr{A}$ is compact;
(ii) $S(t)\mathscr{A} = \mathscr{A}$ for all $t \geq 0$; and
(iii) for any bounded set B,

$$\lim_{t\to\infty} \mathrm{dist}(S(t)B, \mathscr{A}) \to 0 \qquad \text{as} \quad t \to 0,$$

where $\mathrm{dist}(A,B) = \sup_{a\in A} \inf_{b\in B} \|a - b\|$.

For a proof see Hale (1988), Temam (1988), or even Robinson (2001).

There is an alternative, and more analytical characterisation of attractors in terms of complete bounded orbits.

Proposition 1.1 *The global attractor is given by*

$$\mathscr{A} = \{u_0 \in X : \textit{there exists a solution } u(t) \textit{ defined for all } t \in \mathbb{R} \\ \textit{with } u(0) = u_0 \textit{ such that } \|u(t)\| \leq M \ \forall\ t \in \mathbb{R} \textit{ for some } M > 0\}.$$

Proof Suppose that $u_0 \in$ RHS. Then for every $t \geq 0$, $u_0 = S(t)u_{-t}$ for some u_{-t} with $\|u_{-t}\| \leq M$, i.e. $u_{-t} \in B(0, M)$. Then

$$\text{dist}(u_0, \mathscr{A}) = \text{dist}(S(t)u_{-t}, \mathscr{A}) \leq \text{dist}(S(t)B(0, M), \mathscr{A}) \to 0 \qquad \text{as} \quad t \to \infty,$$

and so $u_0 \in \mathscr{A}$. To prove the opposite inclusion, given $u_0 \in A$ it is clear that $\|S(t)u_0\|$ is bounded for all $t \geq 0$, since $S(t)u_0 \in \mathscr{A}$. To extend the solution backwards in time, first find $u(-1) \in \mathscr{A}$ such that $S(1)u(-1) = u_0$ (this is possible since $S(1)\mathscr{A} = \mathscr{A}$), and let $u(-1+t) = S(t)u_{-1}$ for $t \in [0, 1)$. Then continue inductively, choosing $u(-(n+1))$ such that $u(-n) = S(1)u(-(n+1))$ and defining $u(-(n+1)+t) = S(t)u(-(n+1))$ for all $t \in [0, 1)$. The semigroup property ensures that this gives a solution, and since the solution lies within $\mathscr{A}$ for all $t \in \mathbb{R}$ it is bounded. □

1.2 'Box-counting' dimension

Consider a subset X of $\mathbb{R}^n$, and subdivision of $\mathbb{R}^n$ into boxes with sides of length ϵ, i.e. cubes of the form

$$[m_1\epsilon, (m_1 + 1)\epsilon] \times \cdots \times [m_n\epsilon, (m_n + 1)\epsilon].$$

Let $I(X, \epsilon)$ denote the number of these boxes that intersect X. The 'box-counting dimension' is then

$$d_{\text{box}}(X) = \limsup_{\epsilon \to 0} \frac{\log I(X, \epsilon)}{-\log \epsilon}.$$

Now let $N(X, \epsilon)$ denote the minimum number of balls of radius ϵ ('ϵ-balls') required to cover X; in $\mathbb{R}^n$ any ϵ-ball is contained in at most 3^n boxes of side 2ϵ, while each box is contained in a $\sqrt{n}(\epsilon/2)$ ball, from which it follows that

$$I(X, 2\epsilon/\sqrt{n}) \leq N(X, \epsilon) \leq 3^n I(X, 2\epsilon),$$

and hence one could also make the definition

$$d_{\rm box}(X) = \limsup_{\epsilon\to 0} \frac{\log N(X,\epsilon)}{-\log\epsilon}. \tag{1.1}$$

We will adopt this expression, which is also sensible for subsets of infinite-dimensional spaces, as the primary definition of 'box-counting' dimension. Note that the dimension depends on the norm in which we are taking the 'closed balls'.

Definition 1.2 *Let X be a subspace of a Banach space B. The (upper) box-counting dimension of X is*

$$d_{\rm box}(X) = \limsup_{\epsilon\to 0} \frac{\log N(X,\epsilon)}{-\log\epsilon},$$

where $N(X,\epsilon)$ denotes the minimum number of closed balls (in the norm of B) of radius ϵ required to cover X.

(We can also require the balls to have centres in X if we wish.)

The "lim sup" is necessary, as there are simple sets for which the limit as $\epsilon\to 0$ does not exist. Note that for any $d > d_{\rm box}(X)$, there exists an ϵ_0 such that for all $\epsilon < \epsilon_0$,

$$N(X,\epsilon) \le \epsilon^{-d}.$$

1.3 Properties and examples

The following lemma gives some useful properties of this notion of dimension.

Lemma 1.3

(i) $d_{\rm box}(\overline{X}) = d_{\rm box}(X)$*;*
(ii) if $X \subset Y$ *then* $d_{\rm box}(X) \le d_{\rm box}(Y)$*;*
(iii) $d_{\rm box}(X\times) \le d_{\rm box}(X) + d_{\rm box}(Y)$*;*
(iv) if $f : H_1 \to H_2$ *is Hölder continuous with exponent* θ $(0 < \theta \le 1)$*, i.e.*

$$\|f(u) - f(v)\|_2 \le C\|u - v\|_1^{\theta},$$

then $d_{\rm box}(f(X)) \le d_{\rm box}(X)$*; and*
(v) If $X - X = \{x - y :\ x, y \in X\}$ *then* $d_{\rm box}(X - X) \le 2d_{\rm box}(X)$*.*

Proof (i) A cover of X by closed balls also gives a cover over $\overline{X}$; (ii) clear from the definition; (iii) take $d_X > d_{\rm box}(X)$ and $d_Y > d_{\rm box}(Y)$ - choose ϵ_0 small enough that for any $\epsilon < \epsilon_0$ one can cover X by ϵ^{-d_X} balls of radius ϵ and Y by ϵ^{-d_Y} balls of radius ϵ. Then $X \times Y$ can be covered by $\epsilon^{-(d_X+d_Y)}$ balls of radius 2ϵ, from which the result follows; (iv) Suppose that $|f(x) - f(x')| \le C|x - x'|^\theta$. Given $d > d_{\rm box}(X)$, choose ϵ_0 sufficiently small that $d_{\rm box}(X) \le \epsilon^d$ for all $0 < \epsilon < \epsilon_0$. Under f such a covering of X produces a cover of $f(X)$ by sets of diameter no larger than $C\epsilon^\theta$. So

$$N(f(X), C\epsilon^\theta) \le \epsilon^{-d} \qquad \Rightarrow \qquad N(f(X), \delta) \le (\delta/C)^{-d/\theta},$$

and hence $d_{\rm box}(f(X)) \le d_{\rm box}(X)/\theta$. (v) $X - X$ is the image of $X \times X$ under the Lipschitz map $(x, y) \mapsto x - y$. □

For more general theory on the box-counting dimension see Falconer (1990).

We now look at a simple example that will be useful later. Let H be a Hilbert space and $\{e_n\}_{n=1}^\infty$ an orthonormal subset of H. Let $E_\alpha = \{n^{-\alpha}e_n\} \cup \{0\}$ (this set is compact); then $d_{\rm box}(E_\alpha) = 1/\alpha$.

Indeed, for some fixed $\epsilon > 0$ let N be the first n such that $n^{-\alpha} < \epsilon$. Then a single ball of radius ϵ centred at the origin will cover all the points $j^{-\alpha}e_j$ for $j \ge N$. If we cover the first N elements of E_α with balls of radius ϵ this shows that

$$N(E_\alpha, \epsilon) \le \epsilon^{-1/k}.$$

For a lower bound, observe that

$$\|n^{-\alpha}e_n - m^{-\alpha}e_m\|^2 = n^{-2\alpha} + m^{-2\alpha} > 2\epsilon^2$$

if $n, m < N$. So each of the $\{j^{-\alpha}e_j\}_{j=1}^N$ requires a different ball of radius $\epsilon/\sqrt{2}$ to cover it. So

$$N(E_\alpha, \epsilon/\sqrt{2}) \ge \epsilon^{-1/k}.$$

These upper and lower bounds show that $d_{\rm box}(E_\alpha) = 1/\alpha$ as claimed.

A similar analysis can be used to show that the countable set $E_{\log} = \{e_n/\log n\}_{n=2}^\infty \cup \{0\}$ has $d_{\rm box}(E_{\log}) = \infty$.

1.4 Box-counting dimension of attractors

Suppose that A is the attractor of a dynamical system. We now discuss how one can bound the box-counting dimension of A.

In what follows we write $S = S(T)$, the time T map of the dynamical system.

Suppose that given a cover of A by N balls of radius ϵ, we can show that this gives rise to a cover of A by αN balls of radius $\theta\epsilon$ (for some $\theta < 1$), i.e.

$$N(A, \theta\epsilon) \le \alpha N(A, \epsilon). \tag{1.2}$$

Then we can iterate this to deduce that

$$N(A, \theta^k\epsilon) \le \alpha^k N(A, \epsilon)$$

for all $k = 1, 2, \ldots$.

This yields a bound on the dimension of A, since[1]

$$d_{\text{box}}(A) = \limsup_{k\to\infty} \frac{\log N(A, \theta^k\epsilon)}{-\log(\theta^k\epsilon)} = \frac{\log\alpha}{-\log\theta}.$$

How do we go about proving something like (1.2)? The idea is to take an initial covering of A by balls with ϵ small (since we are only concerned with the limit as $\epsilon \to 0$), and map these forward using S. Since A is invariant, a cover of A maps to another cover of $A = SA$. If the balls are small enough then $S(B(x, \epsilon))$ should be well approximated by $S(x) + DS(x)(B(0, \epsilon))$, so if we can (i) understand the image of a ball under a linear map and then (ii) have an efficient way of covering such images, we can hope to derive a bound like (1.2).

1.4.1 The image of a ball under a linear map

Let $L : \mathbb{R}^N \to \mathbb{R}^n$ be a linear map.

Consider the symmetric $N \times N$ matrix $L^T L$. This matrix has a set of orthonormal eigenvectors $\{e_j\}_{j=1}^N$ with corresponding eigenvalues λ_j, i.e. $L^T L e_j = \lambda_j e_j$.

[1] One cannot taking the limit through an arbitrary sequence ϵ_k instead of the limsup as $\epsilon \to 0$; one needs $\epsilon_{k+1} \ge \gamma\epsilon_k$ for some $\gamma > 0$, which of course we have in this case.

We show that each λ_j is non-negative, and that there are at most n non-zero eigenvalues.

First, observe that

$$\lambda_j = (\lambda_j e_j, e_j) = (L^T L e_j, e_j) = (Le_j, Le_j) = \|Le_j\|^2 \geq 0,$$

so that each eigenvalue is non-negative. Write $\lambda_j = \alpha_j^2$.

Next,

$$(Le_i, Le_j) = (L^T L e_i, e_j) = \lambda_j (e_i, e_j) = \lambda_j \delta_{ij}.$$

So the $\{Le_j\}$ are orthogonal in $\mathbb{R}^n$, with $\|Le_j\| = \alpha_j$. Since there can be at most n mutually orthogonal vectors in $\mathbb{R}^n$, it follows that there are at most n non-zero eigenvalues of $L^T L$. Denote these by $\alpha_1^2, \ldots, \alpha_k^2$ (with $k \leq n$).

The non-zero $\alpha_1, \ldots, \alpha_k$ are called the singular values of the matrix L.

If we only want to calculate these singular values, we can just as well work with the matrix LL^T (which is $n \times n$), since if $L^T L e = \lambda e$ then

$$LL^T(Le) = L[L^T Le] = L[\lambda e] = \lambda[Le]$$

and if $LL^T \hat{e} = \lambda \hat{e}$ then

$$L^T[LL^T \hat{e}] = L^T L[L^T \hat{e}] = L^T[\lambda \hat{e}] = \lambda[L^T \hat{e}].$$

Lemma 1.4 *The image of the unit ball in $\mathbb{R}^N$ under a linear map L is an ellipse in $\mathbb{R}^n$ whose semi-axes are Le_j of length α_j.*

Proof (After Section 1.3.1 in Temam, 1988) We have already shown that the $\{Le_j\}$ are orthogonal and that $|Le_j| = \alpha_j$.

Now take some $x \in B_{\mathbb{R}^N}(0,1)$; we can write

$$x = \sum_{j=1}^{k} x_j e_j + y \qquad \sum_{j=1}^{k} |x_j|^2 + |y|^2 \leq 1,$$

where y is orthogonal to the $\{x_j\}$. Then

$$Lx = \sum_{j=1}^{k} x_j (Le_j) = \sum_{j=1}^{\infty} (x_j \alpha_j) \frac{Le_j}{\alpha_j}.$$

This expresses $Lx = \sum_j \xi_j \hat{e}_j$ where the $\{\hat{e}_j\}$ are orthonormal vectors in the directions of Le_j; clearly

$$\sum_j \left(\frac{\xi_j}{\alpha_j}\right)^2 \leq 1,$$

and so the image LB is an ellipse as stated. □

Ellipses can be covered by small balls in an efficient way – for a proof of the following lemma see Chepyzhov & Vishik (2002)

Lemma 1.5 *Let $E \subset \mathbb{R}^n$ be an ellipsoid with semi-axes $\alpha_1 \geq \alpha_2 \geq \cdots$. Then for any $r < \alpha_1$*

$$N(E, \sqrt{2}r) \leq 4^k \frac{\alpha_1 \cdots \alpha_k}{r^k}$$

where k is the largest integer such that $r \leq \alpha_k$.

Putting these ingredients together in an appropriate way (see Chepzhov & Vishik, for example) yields the following theorem:

Theorem 1.6 *Let*

$$\omega_k = \sup_{x \in A} \alpha_1(DS(x)) \cdots \alpha_k(DS(x)).$$

Assume that $\omega_k \leq \bar{\omega}_k$, where $\bar{\omega}_k$ is a concave function of k. Then if $\bar{\omega}_d < 1$, $d_{\text{box}}(A) \leq d$.

Proof Cover A by a collection of balls $\{B(x_j, \epsilon)\}$. Map these forward by S to obtain a new covering of A by ellipses using Lemma 1.4. Cover each ellipse by smaller balls using Lemma 1.5, using the fact that $\omega_d < 1$ to ensure that the number of balls in the new cover can be controlled. The hard part of the proof is that the 'largest integer k such that $r \leq \alpha_k$' that occurs in Lemma 1.5 can vary over A, so one has to be somewhat careful. □

Many examples are known of 'interesting' equations that have finite-dimensional attractors. The canonical 'interesting' example is the two-dimensional Navier-Stokes equations

$$u_t - \Delta u + (u \cdot \nabla)u + \nabla p = f \qquad \nabla \cdot u = 0.$$

It is relatively straightforward to show that if $f \in L^2$ there is a bounded absorbing set in H^1 (i.e. a compact absorbing set in L^2), and that in terms of the Grashof number

$$G = \frac{\|f\|_{L^2}}{\nu^2 \lambda_1}$$

(where λ_1 is the first eigenvalue of the Stokes operator) the dimension of the attractor is bounded by cG in the case of Dirichlet boundary conditions. Some more works shows that one can improve this to $cG^{2/3}(1 + \log G)^{1/3}$ in the periodic case (Constantin & Foias, 1985).

1.5 What does it mean for an attractor to be finite-dimensional?

In the case of the Navier-Stoke equations, one can relate the attractor dimension to the 'smallest lengthscale in the flow', following the heuristic ideas of Landau & Lifshitz (1959).

Essentially their idea is that if there is such a small scale l, then the 'number of degrees of freedom' in the flow should be defined as

$$\frac{|\Omega|}{l^d},$$

where Ω is the d-dimensional domain in which the fluid flows, and $|\Omega|$ denotes the volume of Ω – i.e. the number of 'little boxes of size l' that will fit into Ω.

If we identify 'the number of degrees of freedom of the flow' with the dimension of the global attractor (cf. Doering & Gibbon, 1995), then this suggests that the 'smallest significant length' will be given by

$$l \sim [d_{\text{box}}(A)]^{-1/d}.$$

In the two-dimensional periodic case, this gives an estimate $l \sim G^{-1/3}$ (plus logarithmic corrections) that agrees with the results from the heuristic theory of 2D turbulence due to Kraichnan (the analogue of the 3D Kolmogorov theory).

One could hope for a more dynamical interpretation, however, that is not restricted to fluid dynamics, namely that the existence of a finite-dimensional attractor should imply that the 'asymptotic dynamics are finite dimensional'.

To interpret this more concretely, one might hope for a finite-dimensional system of ODEs in $\mathbb{R}^N$, $\dot{x} = f(x)$ with unique solutions generating a solution mapping $T(t)$, such that

(i) $T(t)$ has a global attractor X;
(ii) there is a homeomorphism $L : A \to X$ such that

$$T(t) = [L \circ S(t) \circ L^{-1}](x).$$

In other words the attractor and its dynamics can be faithfully reproduced (in a one-to-one way) within a finite-dimensional dynamical system.

Such a result appears currently out of reach (for some partial results see Robinson, 1999. The main problem is that if one starts off with a PDE in some infinite-dimensional space H, which we can write here as an abstract ODE $\dot{u} = F(u)$, then the finite-dimensional ODE on $X = LA$ must be given by

$$\dot{x} = f(x) = L(F(L^{-1}(x))).$$

However, uniqueness of solutions requires f to be Lipschitz (or Lipschitz with logarithmic corrections), and we will soon see that one cannot expect L^{-1} to be any better than Hölder continuous in general.

But this potential construction serves to motivate (in part) the abstract embedding theorems that we will now consider, and in particular the attention we pay to properties of the inverse L^{-1} that provides a parametrisation of A in terms of a finite number of parameters.

2

Embedding theorems in finite-dimensional spaces

2.1 Embeddings and parametrisations

We will prove a number of results along the following lines: if X is a compact set with $d_{\rm box}(X) = d$, then for $k > 2d+1$ there are 'many' maps $L : X \to \mathbb{R}^k$ that are one-to-one between X and its image (we will say simply that 'L is one-to-one on X'). Such results (for general finite-dimensional sets) seem to go back to Mañé (1981) who proved such a result for subsets of Banach spaces for which $d_{\rm H}(X - X) < \infty$ (where $d_{\rm H}$ is the Hausdorff dimension).

An immediate consequence of such a result is that the inverse mapping from $L(X)$ back to X is continuous:

Lemma 2.1 *If X is compact and $f : X \to Y$ is continuous and one-to-one, then $f^{-1} : f(X) \to X$ is continuous.*

Proof If not there exists an $\epsilon > 0$, a $y \in f(X)$ and a sequence $\{y_n\} \in f(X)$ such that

$$y_n \to y \qquad \text{but} \qquad |f^{-1}(y_n) - f^{-1}(y)| > \epsilon. \tag{2.1}$$

However, $f^{-1}(y_n)$ is a sequence in the compact set, so it has a subsequence (which we relabel) such that $f^{-1}(y_n) \to x \in X$. Since f is continuous, it follows that $y_n \to f(x)$, so that $y = f(x)$. But then $x = f^{-1}(y)$, which contradicts the second part of (2.1). $\square$

If there is a continuous embedding of a compact X into $\mathbb{R}^n$, then the inverse of such a map gives a parametrisation of X, i.e. one can specify a point on X using only n coordinates.

In fact we will take some care to get more particular information on the continuity of L^{-1}, in most cases being able to show that it is Hölder continuous, i.e. that for some $0 < \theta < 1$,

$$|L^{-1}(x) - L^{-1}(y)| \leq C|x - y|^{\theta}.$$

The material here is essentially a combination of arguments in the papers by Sauer, Yorke, & Casdagli (1991), and that of Hunt & Kalsohin (1999).

We will initially consider embeddings by means of general linear maps $L : \mathbb{R}^N \to \mathbb{R}^k$, and aim to show that 'almost every' such map is a 'nice' embedding (i.e. an embedding with a Hölder continuous inverse).

Let $\{e_\alpha\}_{\alpha=1}^N$ and $\{\hat{e}_j\}_{j=1}^k$ be orthonormal bases for $\mathbb{R}^N$ and $\mathbb{R}^k$ respectively. We can represent such a linear map as an $N \times k$ matrix

$$L = \begin{pmatrix} l_1 & l_2 & \cdots & \cdots & l_N \end{pmatrix}, \tag{2.2}$$

where each l_α is an element of $\mathbb{R}^k$ that represents $L(e_\alpha)$. We can write $l_\alpha = \sum_j l_{\alpha,j}\hat{e}_j$.

We consider the set Q of all linear maps that are given in the form (2.2), where

$$\sum_{\alpha=1}^{N} |l_\alpha|^2 \leq 1.$$

This is equivalent to taking $\{l_{\alpha,j}\}$ in the unit ball in $\mathbb{R}^{Nk}$. We equip Q with a probability measure μ, equal to Lebesgue measure on this ball, normalised so that the total measure of Q is equal to one.

Note that for any $L \in Q$, $|Lx| \leq |x|$ for every $x \in \mathbb{R}^N$ (and since L is linear, $|Lx - Ly| \leq |x - y|$, i.e. L is Lipschitz with Lipschitz constant at most 1).

A key part of the proof is an estimate of 'how many' elements of Q map some given x to elements with small norm. We will derive this as a corollary of the following result. (Recall that the singular values of M are the square roots of the eigenvalues of M^TM (or equivalently of MM^T).)

Lemma 2.2 *(After Lemma 4.2 of Sauer et al.) Let* $M : \mathbb{R}^m \to \mathbb{R}^n$ *be a* *linear map. For a positive integer r ($1 \le r \le n$), let $\alpha_r > 0$ be the rth largest singular value of M. Then*

$$\frac{\mathrm{Vol}\,\{x \in B_m(\rho) : \ |F(x)| < \delta\}}{\mathrm{Vol}(B_m(\rho))} \le m \left(\frac{\delta}{\alpha_r \rho}\right)^r . \tag{2.3}$$

We denote by $\Omega_n = \pi^{n/2}/\Gamma(1+\frac{n}{2})$ the volume of the unit ball in $\mathbb{R}^n$.

Proof The image of $B_m(\rho)$ under M, $MB_m(\rho)$, is an ellipse, whose semiaxes are $\{\rho\alpha_i\}$, where the α_i are the singular values of M (cf. Lemma 1.4). So

$$\begin{aligned}
\mathrm{Vol}\{x \in B_m(\rho) : \ |Mx| < \delta\} &= \Omega_n \left[\prod_{j=1}^{n} \min\left(\frac{\delta}{\alpha_j}, \rho\right)\right] \Omega_{m-n}\rho^{m-n} \\
&\le \Omega_n \Omega_{m-n} \left[\prod_{j=1}^{r} \frac{\delta}{\alpha_j}\right] \left[\prod_{j=r+1}^{n} \rho\right] \rho^{m-n} \\
&= \Omega_n \Omega_{m-n} \left(\frac{\delta}{\alpha_r}\right)^r \rho^{m-r}
\end{aligned}$$

for any $r \in \{1, \ldots, n\}$. Since $\mathrm{Vol}\, B_m(\rho) = \Omega_m \rho^m$, the LHS of (2.3) is bounded by

$$\begin{aligned}
\mathrm{LHS} &\le \frac{\Omega_n \Omega_{m-n}}{\Omega_m} \left(\frac{\delta}{\alpha_r \rho}\right)^r \\
&= \frac{\Gamma(1+\frac{n}{2})\Gamma(1+\frac{m-n}{2})}{\Gamma(1+\frac{m}{2})} \left(\frac{\delta}{\alpha_r \rho}\right)^r \\
&= \frac{\Gamma(1+\frac{n}{2})\Gamma(1+\frac{m-n}{2})}{\Gamma(2+\frac{m}{2})} (1+\tfrac{m}{2}) \left(\frac{\delta}{\alpha_r \rho}\right)^r \\
&= B(1+\tfrac{n}{2}, 1+\tfrac{m-n}{2})(1+\tfrac{m}{2}) \left(\frac{\delta}{\alpha_r \rho}\right)^r \\
&\le (1+\tfrac{m}{2}) \left(\frac{\delta}{\alpha_r \rho}\right)^r \\
&\le m \left(\frac{\delta}{\alpha_r \rho}\right)^r .
\end{aligned}$$

Here $B(a, b)$ is the beta function, which is equal to $\Gamma(a)\Gamma(b)/\Gamma(a+b)$ and is given by the integral $\int_0^1 t^{a-1}(1-t)^{b-1}\,\mathrm{d}t$; clearly $B(a, b) \le 1$ for $a, b \ge 1$. □

We now deduce the key estimate for the embedding result as a corollary of this.

Corollary 2.3 *Let Q be the set of linear maps $L : \mathbb{R}^N \to \mathbb{R}^k$ described above, equipped with the measure μ. Then for any $x \in \mathbb{R}^N$ and any $\epsilon > 0$,*

$$\mu\{L \in Q: \ |Lx| < \epsilon\} \le Nk \left(\frac{\epsilon}{|x|}\right)^k.$$

Proof Let $\{e_\alpha\}_{\alpha=1}^N$ be a basis for $\mathbb{R}^N$, and $\{\hat{e}_j\}_{j=1}^k$ be a basis for $\mathbb{R}^k$. We can write any $L \in \mathscr{L}(\mathbb{R}^N, \mathbb{R}^k)$ in the form

$$L = \sum_{\alpha,j} c_{\alpha,j} L_{\alpha,j},$$

where $L_{\alpha,j}$ is the linear map from $\mathbb{R}^N$ into $\mathbb{R}^k$ given by

$$L_{\alpha,j}(z) = (z, e_\alpha)\hat{e}_j \qquad \alpha = 1, \ldots, N; \ j = 1, \ldots, k$$

(i.e. $L_{\alpha,j}$ is the linear map that sends e_α to $\hat{e}_j$).

Then for any $z \in \mathbb{R}^N$,

$$Lz = \sum_{\alpha,j} c_{\alpha,j}(z, e_\alpha)\hat{e}_j = M_z \underline{c},$$

$\underline{c} \in \mathbb{R}^{Nk}$ with components $c_{\alpha,j}$, and where M_z is a transformation from $\mathbb{R}^{Nk}$ into $\mathbb{R}^k$, with components

$$[M_z]_{i,\{\alpha,j\}} = (z, e_\alpha)\delta_{ij} \qquad \alpha = 1, \ldots, N; \ i, j = 1, \ldots, k.$$

In order to apply Lemma 2.2 we need to find the singular values of M_z. We calculate these by considering $M_z M_z^T$ rather than $M_z^T M_z$, since

$$\begin{aligned}[M_z M_z^T]_{r,s} &= \sum_{\alpha,j} [M_z]_{r,\{\alpha,j\}} [M_z]_{s,\{\alpha,j\}} \\ &= (z, e_\alpha)\delta_{sj}(z, e_\alpha)\delta_{rj} = |z|^2 \delta_{rs} :\end{aligned}$$

MM^T has k non-zero singular values, all of which are $|z|$. Using this the result follows at once from Lemma 2.2. □

2.2 An 'abstract' embedding theorem for subsets of $\mathbb{R}^N$

Theorem 2.4 *Let X be a compact subset of $\mathbb{R}^N$. If $k > 2d_{\text{box}}(X)$ then* *given α with*

$$0 < \alpha < 1 - \frac{2d}{k},$$

for almost every linear map $L : \mathbb{R}^N \to \mathbb{R}^k$ there exists a $C = C_L$ such that

$$|Lx - Ly| \geq C|x - y|^{1/\alpha}; \tag{2.4}$$

in particular L is one-to-one on X.

Proof Take $d > d_{\text{box}}(X)$; let $Z_n = \{z \in X - X : \ |z| \geq 2^{-n}\}$, and set

$$Q_n = \{L \in Q : \ |Lz| \leq 2^{-n/\alpha} \text{ for some } z \in Z_n\}.$$

This Q_n is essentially the set of ‘bad’ linear maps for which (2.4) does not hold for some (x, y) with $x - y \in Z_n$.

Cover Z_n with balls $B(z_j, 2^{-n/\alpha})$; we need no more than $2^{2nd/\alpha}$ of these. Let $Y_j = Z_n \cap B(z_j, 2^{-n/\alpha})$ and choose some $z_0 \in Y_j$. Then

$$|Lz_j| > 2 \cdot 2^{-n/\alpha} \qquad \Rightarrow \qquad |Lz| > 2^{-n/\alpha} \quad \forall\ z \in Y_j. \tag{2.5}$$

So if things go wrong in Y_j, we must have

$$|Lz_j| \leq 2 \cdot 2^{-n/\alpha}.$$

Since $|z_j| \geq 2^{-n}$, the measure of $L \in Q$ for which things go wrong in Y_j is bounded by

$$Nk \left(\frac{2 \cdot 2^{-n/\alpha}}{2^{-n}} \right)^k.$$

Since it requires no more than $2^{2nd/\alpha}$ balls to cover Z_n, the total measure of maps for which things fail in Z_n (the set Q_n) is bounded by

$$\mu(Q_n) \leq Nk 2^k 2^{nk} 2^{-n(k-2d)/\alpha} = C_k 2^{[k-(k-2d)/\alpha]n}.$$

We now ensure that $\sum \mu(Q_n) < \infty$: do this this we require the exponent in the power of 2 to be negative,

$$k - \frac{(k - 2d)}{\alpha} < 0.$$

This means we must take $k > 2d$, and then $\alpha < 1-(2d/k)$ as in the statement of the theorem.

Since $\sum \mu(Q_n) < \infty$, it follows that[1] almost every $L \in Q$ is contained in only a finite number of the Q_n. For such an L, there exists an n_0 such that $L \notin Q_n$ for all $n \geq n_0$, i.e. such that

$$|x-y| \geq 2^{-n} \quad \Rightarrow \quad |L(x-y)| \geq 2^{-n/\alpha} \qquad \text{for all} \quad n \geq n_0.$$

Then if $2^{-n} \leq |x-y| < 2^{-(n-1)}$ with $n \geq n_0$ we have

$$|L(x-y)| \geq 2^{-n/\alpha} \geq 2^{-1/\alpha}|x-y|^{1/\alpha},$$

while if $|x-y| \geq 2^{-n_0}$, since X is bounded we must have $|x-y| \leq M$ for some $M > 0$, so that

$$|L(x-y)| \geq 2^{-n_0/\alpha} = [2^{-n_0/\alpha}/M^{1/\alpha}]M^{1/\alpha} \geq [M2^{n_0}]^{-1/\alpha}|x-y|^{1/\alpha},$$

and (2.4) holds for $C_L = \min([M2^{n_0}]^{-1/\alpha}, 2^{-1/\alpha})$. □

2.3 The Takens time-delay embedding theorem in $\mathbb{R}^N$

We now prove a version of a more dynamical embedding result – the Takens time-delay embedding theorem. The idea here is to replace a linear map by a single scalar observation $h : X \to \mathbb{R}$, but repeated (at 'equal time intervals') a number of times. To make the statement and proof simpler, we take $g = S(T)$, and consider the k-fold observation mapping

$$x \mapsto [h(x), h(g(x)), \ldots, h(g^{k-1}(x))].$$

In the proof we will need the following lemma. We omit the proof, which follows very similar lines to the argument used to prove Theorem 2.4

Lemma 2.5 *(Sauer et al., Lemma 4.3) Let S be a compact subset of $\mathbb{R}^{2k}$, and let and $G_0, G_1, \ldots, G_t$ be Lipschitz maps from S to $\mathbb{R}^n$. Assume that for each $z \in S$, the $n \times t$ matrix*

$$M_z = \begin{pmatrix} G_1(z) & \cdots & G_t(z) \end{pmatrix}$$

[1] This is the Borel-Cantelli Lemma. To prove this, consider $\mathcal{Q} = \cap_{j=1}^{\infty} \cup_{n \geq j} Q_n$, which consists precisely of those L that lie in infinitely many of the Q_ns. To see that $\mu(\mathcal{Q}) = 0$, observe that for every j,

$$\mu(\mathcal{Q}) \leq \mu\left(\cup_{n \geq j} Q_n\right) \leq \sum_{n \geq j} \mu(Q_n),$$

and that the right-hand side tends to zero as $j \to \infty$ since $\sum \mu(Q_n) < \infty$.

has rank at least r and that $d_{\rm box}(\overline{S}) < r$. Take

$$0 < \gamma < 1 - \frac{d_{\rm box}(\overline{S})}{r}.$$

Then for almost every $\alpha \in \mathbb{R}^t$ there exists a constant $C = C_\alpha$ such that

$$|G_\alpha(z)| \geq C_\alpha |z|^{1/\gamma} \qquad \text{for all} \quad z \in S,$$

where

$$G_\alpha = G_0 + \sum_{j=1}^{t} \alpha_j G_j.$$

We can now prove a version of Sauer et al.'s Theorem 4.13, that gives Hölder continuity of the parametrisation along the lines of Theorem 4.1 in Hunt & Kaloshin (the statement of the theorem there is incorrect, since linear maps will not suffice).

After Theorem 2.4, our intuition should be that if we have k 'independent' measurements and $k > 2d_{\rm box}(X)$, this should provide an embedding. However, our measurements in the time delay case arise from iterates of some $x \in X$. If $x, \ldots, g^{k-1}(x)$ are distinct then the condition $k > 2d_{\rm box}(X)$ should suffice; but note that if x lies on a periodic orbit of period p, $\{x, \ldots, g^{k-1}(x)\}$ contains only p distinct points, so we will also have to limit the size of the sets X_p of p-periodic points, requiring $d_{\rm box}(X_p) < p/2$ for $p = 1, \ldots, k$. These conditions are sufficient to prove a result that closely parallels Theorem 2.4.

Theorem 2.6 *Let X be a compact subset of $\mathbb{R}^N$ with $d_{\rm box}(X) = d$, and $g : X \to X$ a Lipschitz map. Assume that*

(i) $k > 2d$, and

(ii) the set X_p of p-periodic points of g (i.e. $x \in X$ such that $g^p(x) = x$) satisfies $2d_{\rm box}(X_p) < p$ for all $p = 1, \ldots, k$.

Choose γ with

$$0 < \gamma < 1 - \max\left(\max_{1 \leq p \leq k} \frac{2d_{\rm box}(X_p)}{p}, \frac{2d_{\rm box}(X)}{k} \right).$$

Let $h_1, \ldots, h_m$ be a basis for the polynomials in N variables of degree at most $2k$, and given any θ-Hölder function $h_0 : \mathbb{R}^N \to \mathbb{R}$ define

$$h_\alpha = h_0 + \sum_{j=1}^{m} \alpha_j h_j.$$

Then for almost every $\alpha \in \mathbb{R}^m$ the k-fold observation map $F_k : X \to \mathbb{R}^k$ defined by

$$F_k[h_\alpha, g](x) = \Big(h_\alpha(x), h_\alpha(g(x)), \ldots, h_\alpha(g^{k-1}(x))\Big)^T \tag{2.6}$$

satisfies

$$|F_k[h_\alpha, g](x) - F_k[h_\alpha, g](y)| \geq C_\alpha |x - y|^{1/\gamma} \tag{2.7}$$

for some $C_\alpha > 0$; in particular $F_k[h_\alpha, g]$ is one-to-one on X for almost every $\alpha \in \mathbb{R}^m$.

Proof For $i = 0, 1, \ldots, m$ define

$$F_i(x) = \begin{pmatrix} h_i(x) \\ h_i(g(x)) \\ \vdots \\ h_i(g^{k-1}(x)) \end{pmatrix},$$

so that by definition

$$F(h_\alpha, g) = F_0 + \sum_{j=1}^{m} \alpha_j F_j.$$

In order to apply Lemma 2.5 we need to check, for each $x \neq y$, the rank of the matrix

$$\begin{aligned} M_{(x,y)} &= \begin{pmatrix} F_1(x) - F_1(y) & \cdots & F_m(x) - F_m(y) \end{pmatrix} \\ &= \begin{pmatrix} h_1(x) - h_1(y) & \cdots & h_m(x) - h_m(y) \\ \vdots & \ddots & \vdots \\ h_1(g^{k-1}(x)) - h_1(g^{k-1}(y)) & \cdots & h_m(g^{k-1}(x)) - h_m(g^{k-1}(y)) \end{pmatrix}. \end{aligned}$$

In order to analyse this, it is helpful to write it in the form $M = JH$, where

$$H_{(x,y)} = \begin{pmatrix} h_1(z_1) & \cdots & h_m(z_1) \\ \vdots & \ddots & \vdots \\ h_1(z_q) & \cdots & h_m(z_q) \end{pmatrix},$$

with all of the $z_1, \ldots, z_q$ distinct (we have $q \leq 2k$), and where $J_{(x,y)}$ is a $k \times q$ matrix each of whose rows consists of zeros except for one 1 and one -1. Given any $\xi \in \mathbb{R}^q$, we can find a set of coefficients $\{\alpha_j\}_{j=1}^m$ such that $\sum_j \alpha_j h_j(z_l) = \xi_l$, i.e. such that $H_{(x,y)}\underline{\alpha} = \xi$. This implies that the rank of H is q; since $J : \mathbb{R}^q \to \mathbb{R}^k$, we only need to check the rank of J.

We split the set $X \times X \setminus \{0\}$ into three disjoint sets of pairs (x, y), and show that that $F_\alpha(x) \neq F_\alpha(y)$ for almost every α on each of these sets. It then follows that $F_\alpha(x) \neq F_\alpha(y)$ for almost every α, for any $x \neq y$, $x, y \in X$.

Case 1: x and y are not both periodic of period $\leq k$.

In this case wlog $\{x, g(x), \ldots, g^{k-1}(x)\}$ consists of k discrete points and $\{y, \ldots, g^{k-1}(y)\}$ consists of $r \geq 1$ points distinct from the iterates of x. So

$$J_{(x,y)} = \left(\begin{array}{ccccc|ccccc} 1 & 0 & \cdots & \cdots & 0 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 & 0 & -1 & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & \cdots & 0 & 0 & 0 & \vdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & -1 & \cdots & 0 \end{array}\right),$$

where the left block is $k \times k$ and the right block is $r \times k$ (with the top $r \times r$ entries being minus the identity).

It follows that the rank of $J : \mathbb{R}^{k+r} \to \mathbb{R}^k$ is k, and so is the rank of $F = JH$. Since the set of pairs $x \neq y$ has box-counting dimension at most $2d$, and we have just shown that rank $M_{(x,y)} = k > 2d$ by assumption, the conditions of Lemma 2.5 are met for this choice of (x, y), i.e. for all such (x, y), $F_\alpha(x) \neq F_\alpha(y)$ for almost every α.

Case 2: x and y lie in distinct periodic orbits of period $\leq k$.

Suppose that p and q are the minimal integers such that $g^p(x) = x$ and $g^q(y) = y$, wlog $1 \leq q \leq p \leq k$. Then J has rank at least p (its top left $p \times p$ entries are the $p \times p$ identity matrix). So rank $M_{(x,y)} \geq p$ in this case, while by assumption the set of pairs of periodic points of period $\leq p$ has dimension $< p$. So once more we can apply Lemma 2.5.

Case 3: x and y lie on the same periodic orbit of period $\leq k$.

Suppose that p and q are the minimal integers such that $g^p(x) = x$ and $g^q(x) = y$, with $1 \leq q < p \leq k$. As an illustrative example, if $p = 7$ and $q = 4$, then J is of the form

$$J_{(x,y)} = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{pmatrix}.$$

The rank of such a matrix is at least $p/2$, and hence rank $M_{(x,y)} \geq p/2$ in this case. The box-counting dimension of the set of pairs that lie on the same periodic orbit if bounded by $d_{\text{box}}(A_p)$, since any such (x, y) is contained in the image of A_p under one of the Lipschitz mappings $x \mapsto (x, g^j(x))$ for some $j = 1, \ldots, k$. □

2.4 Periodic orbits and the Lipschitz constant for ODEs

If we recast Theorem 2.6 in terms of a continuous flow generated by a Lipschitz ODE (so $g = S(T)$ for some $T > 0$ and $\theta = 1$) then observe that there can be no embedding result if there are periodic orbits of period T or even $2T$. This follows from the statement of the theorem (since in this case the dimension of X_1 and X_2 is at least one, and the condition $\dim(X_p) < p/2$ cannot be satisfied for $p = 1, 2$), but this is not simply an artefact of the proof, as the following arguments from Sauer et al.'s paper show.

Indeed, if there is a periodic orbit Γ of period T this is a topological circle. But under the time-delay mapping, any point on Γ maps onto a line in $\mathbb{R}^k$. One cannot map a circle onto a line using any continuous one-to-one mapping, so the theorem must fail in this case.

Such an embedding is also impossible if there is a periodic orbit Γ of period 2. Consider the map $x \mapsto d(x) = h(g(x)) - h(x)$. Then either $d(x) \equiv 0$ on Γ, or there is some $x_0 \in \Gamma$ such that $d(x_0) \neq 0$. In the latter case,

$$d(g(x_0)) = h(g^2(x_0)) - h(g(x_0)) = h(x_0) - h(g(x_0)) = -d(x_0),$$

so that there must be some $x^* \in \Gamma$ with $d(x^*) = 0$. But this implies that $h(x^*) = h(g(x^*))$, so that x^* and $g(x^*)$ (which are distinct points) are mapped to the same point in $\mathbb{R}^k$ whatever the value of k.

So it is useful to have a result that guarantees the non-existence of periodic orbits with certain periods.

We give a simple proof here of such a result for ODEs due to Robinson & Vidal López (2006), following ideas in Kukavica (1994).

Theorem 2.7 *Any periodic orbit of the equation $\dot{x} = f(x)$, where f has Lipschitz constant L, has period $T \geq 1/L$.*

Yorke (1969) showed that the period is in fact bounded below by $2\pi/L$ (and this result is sharp in general).

Proof Fix $\tau > 0$ and set $v(t) = x(t) - x(t - \tau)$. Then

$$v(t) - v(s) = \int_s^t \dot{v}(r)\,\mathrm{d}r.$$

Integrating both sides with respect to s from 0 to T gives

$$Tv(t) = \int_0^T \left(\int_s^t \dot{v}(r)\,\mathrm{d}r \right) \mathrm{d}s$$

and so

$$T|v(t)| \leq \int_0^T \int_0^T |\dot{v}(r)|\,\mathrm{d}r\,\mathrm{d}s \leq T \int_0^T |\dot{v}(r)|\,\mathrm{d}r,$$

i.e.

$$\begin{aligned} |x(t) - x(t - \tau)| \leq \int_0^T |\dot{v}(s)|\,\mathrm{d}s &= \int_0^T |f(x(s)) - f(x(s - \tau))|\,\mathrm{d}s \\ &\leq L \int_0^T |x(s) - x(s - \tau)|\,\mathrm{d}s. \end{aligned}$$

Therefore

$$\int_0^T |x(t) - x(t - \tau)|\,\mathrm{d}t \leq LT \int_0^T |x(s) - x(s - \tau)|\,\mathrm{d}s,$$

and it follows that if $LT < 1$ then

$$\int_0^T |x(t) - x(t - \tau)|\,\mathrm{d}t = 0.$$

Thus $x(t) = x(t - \tau)$ for all $\tau > 0$, i.e. $x(t)$ is constant. □

3

Embeddings in infinite-dimensional spaces

We now want to follow a similar programme for subsets of infinite-dimensional spaces – for simplicity we treat the case of subsets of a (real separable) Hilbert space H (one can also treat the Banach space case with a little more effort). The arguments are similar to the finite-dimensional case, but involve some additional ideas. This proof comes essentially from Hunt & Kaloshin (1999).

3.1 Prevalence

We first have to understand what we might mean by 'almost every' linear map from H into $\mathbb{R}^k$. Although we could think of such a map as a countable collection of elements of $\mathbb{R}^k$ (giving the image of each element of some orthonormal basis of H), we cannot put a uniform probability measure on 'the unit ball in $\mathbb{R}^{\infty k}$' (whatever that might mean).

Instead, we adopt the notion of 'prevalence' introduced by[1] Hunt, Sauer, & Yorke (1992). In fact we have already seen this idea in action, in the finite-dimensional version of the Takens theorem: there we proved that given any initial Lipschitz function $h_0 : \mathbb{R}^N \to \mathbb{R}$, the perturbed function

$$h_0 + \sum_{j=1}^{N} \alpha_j h_j$$

[1] Strictly speaking the idea was introduced much earlier by Christensen (1974) in work on the differentiability of Lipschitz mappings. Hunt et al. 'rediscovered' the definition, and coined the suggestive term 'prevalence'.

makes the k-fold observation map one-to-one for almost every $\underline{\alpha} \in \mathbb{R}^N$.

We can make this idea more formal with the following definition.

Definition 3.1 *Let V be a normed space, and S a subset of V. Then S is* prevalent *if there exists a compact subset Q of V ('the probe set') equipped with a probability measure μ, such that for any $v \in V$,*

$$v + q \in S \qquad \text{for } \mu - \text{almost every } q \in Q.$$

(In the finite-dimensional Takens theorem, Q was the set of all polynomials in N variables of degree $\leq 2k$.)

Lemma 3.2 *If S is a prevalent subset of V then S is dense in V.*

Proof Given any $v \in V$ and $\epsilon > 0$, cover Q (which is compact) by a finite number of balls of radius ϵ. At least one these balls, say $B(x, \epsilon)$, must have positive μ-measure. It follows that $(v - x) + B(x, \epsilon) = B(v, \epsilon)$ must contain a point of S. □

It is also the case that just as a countable intersection of sets of full measure has full measure, so does a countable intersection of prevalent sets (for a proof see Hunt et al.).

3.1.1 A nice probe set

To construct an appropriate probe set in this infinite-dimensional case, we start with a sequence of finite-dimensional (d_j-dimensional) subspaces $\{V_j\}_{j=1}^{\infty}$ of H.

For any $u \in H$, denote by u^* the element of H^* given by $u^*(x) = (u, x)$. Then we define Q to be those linear maps from H into $\mathbb{R}^k$ given by

$$Q = \{L = (L_1, \ldots, L_k) : \; L_n = \sum_{j=1}^{\infty} j^{-2} \phi_{nj}^*, \quad \phi_{nj} \in V_j \quad \text{with } \|\phi_{nj}\| \leq 1\}$$

This looks unpleasant, but one can show that Q is a compact subset of $\mathscr{L}(H, \mathbb{R}^k)$ [the space of all bounded linear maps from H into $\mathbb{R}^k$], and it is clear that any $L \in Q$ can somehow be 'approximated' by a map from some $\mathbb{R}^N$ (more precisely $\mathbb{R}^{d_j}$ for some j) into $\mathbb{R}^k$.

One can define a measure on Q as the product of the probability measures obtained by choosing each ϕ_{nj} from a uniform distribution on the unit ball in $\mathbb{R}^{d_j}$ (since this is isometric to the unit ball in V_j).

We then have the following estimate, which (as before) is really the key element of the whole proof.

Lemma 3.3 *Given any $z \in H$, for any $f \in \mathscr{L}(H, \mathbb{R}^k)$, for any $j \in \mathbb{N}$,*

$$\mu\{L \in Q : \ |(f + L)z| < \epsilon\} \leq \left(\frac{\epsilon j^{-2} d_j^{1/2}}{\|P_j z\|} \right)^k,$$

where P_j denotes the orthogonal projection onto V_j.

Essentially this says that we can bound the LHS by looking at 'the finite-dimensional component from V_j'. Note that the dependence on the dimension of the larger space (here $d_j^{k/2}$ when considering the d_j-dimensional space V_j) is significantly worse than before (it was Nk when considering maps from $\mathbb{R}^N$ into $\mathbb{R}^k$). This arises from the somewhat more complicated definition of the space Q.

Recall that when applying the equivalent bound in the finite-dimensional case ($\mu \leq C_k(\epsilon/\|z\|)^k$) we looked at sets of $z \in X - X$ with $\|z\| \geq \epsilon$. We will do something similar in the infinite-dimensional case, but now we have to replace $\|z\|$ in the denominator by $\|P_j z\|$. So we will need to ensure that if we have $\|z\| \geq \epsilon$, we still have $\|P_j z\| \sim \epsilon$. This is possible if, for example, we know that every point of X lies within $\epsilon/3$ of $P_j H$, for then

$$\|P_j z\| = \|P_j(x-y)\| \geq \|x-y\| - \|x - P_j x\| - \|y - P_j y\| \geq \epsilon - \epsilon/3 - \epsilon/3 = \epsilon/3.$$

We will therefore choose our spaces V_j in such a way that they approximate X 'nicely enough'.

3.2 The thickness exponent

To quantify how nicely X can be approximated by linear subsapces, Hunt & Kaloshin introduced the 'thickness exponent'.The 'thickness exponent' (or simply 'thickness').

Definition 3.4 *Let X be a subset of a Banach space $\mathscr{B}$. The thickness exponent of X in $\mathscr{B}$, $\tau(X;\mathscr{B})$ is given by*

$$\tau(X;\mathscr{B}) = \limsup_{\epsilon\to 0} \frac{\log D(X,\epsilon)}{-\log\epsilon}, \tag{3.1}$$

where $D(X,\epsilon)$ is the dimension of the smallest linear subspace V of $\mathscr{B}$ such that

$$\mathrm{dist}_{\mathscr{B}}(X,V) \le \epsilon,$$

i.e. every point in X lies within ϵ of V (in the norm of $\mathscr{B}$).

Note that $\tau(X) \le d_{\mathrm{box}}(X)$: if one covers X by $N(X,\epsilon)$ balls of radius ϵ, clearly the linear subspace spanned by the centres of these balls (whose dimension is no larger than $N(X,\epsilon)$) approximates X to within ϵ.

Friz & Robinson (1999) showed that if U is a sufficiently regular bounded domain in $\mathbb{R}^n$, then if X is a subset of $L^2(U)$ that consists of functions that are uniformly bounded in the Sobolev space $H^s(U)$, it follows that $\tau(X) \le n/s$. [In particular this shows that the attractors of PDEs that are 'smooth' (bounded in H^s for all s) have thickness exponent zero.]

Lemma 3.5 *Let $U \subset \mathbb{R}^m$ be a smooth bounded domain. Let X be a compact subset of $[L^2(U)]^n$ such that*

$$\sup_{u\in X} \|u\|_{[H^s(U)]^n} < +\infty$$

for some $s \ge 1$. Then $\tau(X) \le m/s$.

Proof We give a slightly incorrect proof which ignores the issue of boundary conditions, but contains the essential ideas (this is the proof you'll find in Friz & Robinson's paper; the boundary conditions can be fixed by considering extensions of our original functions to larger domains such that the extended functions have compact support, see Robinson, 2008).

A proof for the case $n = 1$ is sufficient; if $n > 1$ the argument can be applied to each component of the functions in X.

Let A be the Laplacian operator on U, with Dirichlet boundary conditions ($u = 0$ on ∂U). The Laplacian on such a domain has a sequence $\{w_j\}$ of eigenfunctions with corresponding eigenvalues λ_j ($Aw_j = \lambda_j w_j$) which, if ordered so that $\lambda_{j+1} \geq \lambda_j$, satisfy $\lambda_j \sim j^{2/m}$ (see Davies (1995) for example).

Now, if $u \in H^s$ then $u \in D(A^{s/2})$ [in fact this also requires some conditions at the boundary] with

$$\|A^{s/2}u\| \leq C_s \|u\|_{H^s}.$$

Now consider the projection P_k of u onto the space spanned by the first k eigenfunctions of A,

$$P_k u = \sum_{j=1}^{k} (u, w_j) w_j,$$

and its orthogonal complement $Q_k = I - P_k$. Then

$$\begin{aligned} \|u - P_k u\| &\leq \|Q_k u\| \\ &= \|Q_k A^{-s/2} A^{s/2} u\| \\ &= \|Q_k A^{-s/2}\|_{\text{op}} \|A^{s/2} u\| \\ &\leq \lambda_{k+1}^{-s/2} \|u\|_{H^s} \\ &\leq C k^{-s/m}, \end{aligned} \tag{3.2}$$

for some constant C, and so $\tau(X) \leq m/s$. □

3.3 Abstract embedding for finite-dimensional subsets of H

The following is Hunt & Kaloshin's embedding theorem. The proof closely follows the finite-dimensional one, choosing V_j to be spaces that approximate X to within $2^{-j/\alpha}/3$.

Theorem 3.6 *Let X be a compact subset of a real Hilbert space H, with $d_{\text{box}}(X) = d < \infty$ and $\tau(X) = \tau$. Then for any integer $k > 2d$ and any α with*

$$0 < \alpha < \frac{k - 2d}{k(1 + (\tau/2))} \tag{3.3}$$

there exists a prevalent set of bounded linear maps $L : \mathscr{B} \to \mathbb{R}^k$ such that

$$|Lx - Ly| \geq C_L \|x - y\|^{\alpha} \qquad \text{for all} \qquad x, y \in X. \tag{3.4}$$

In particular, L is injective on X.

How good is the Hölder exponent? If one lets $k \to \infty$ then it approaches

$$\frac{1}{1+(\tau/2)};$$

it can only be made arbitrarily close to 1 if the thickness of X is zero.

But how good is the bound obtained in the theorem? A simple collection of examples – the sets $E_\alpha = \{n^{-\alpha}e_n\}_{n=1}^\infty$ considered in Section 1.3, with $\dim(E_\alpha) = 1/\alpha$ – suffice to show that the bound is 'asymptotically sharp' (as $k \to \infty$).

We will require two simple lemmas, the first of which we state without proof.

Lemma 3.7 *Suppose that $L : H \to \mathbb{R}^k$ is onto and linear. Then $U = (\ker L)^\perp$ has dimension k, and L can be decomposed uniquely as MP, where P is the orthogonal projection onto U and $M : U \to \mathbb{R}^k$ is an invertible linear map.*

Lemma 3.8 *Let P be any orthogonal projection in H, and $\{e_j\}_{j=1}^\infty$ any orthonormal subset of H. Then*

$$\operatorname{rank} P \geq \sum_{j=1}^\infty \|Pe_j\|^2,$$

with equality guaranteed if $\{e_j\}_{j=1}^\infty$ is a basis for H.

Proof Suppose that P has rank k. Then there exists an orthonormal basis $\{u_1, \ldots, u_k\}$ for PH, so that for any $x \in H$,

$$Px = \sum_j (x, u_j)u_j.$$

In particular, $Pe_i = \sum_j (e_i, u_j)u_j$, so that

$$\|Pe_i\|^2 = (Pe_i, Pe_i) = (Pe_i, e_i) = \sum_{j=1}^k (e_i, u_j)(u_j, e_i) = \sum_{j=1}^k |(e_i, u_j)|^2.$$

It follows that

$$\sum_{i=1}^\infty \|Pe_i\|^2 = \sum_{i=1}^\infty \sum_{j=1}^k |(e_i, u_j)|^2 = \sum_{j=1}^k \sum_{i=1}^\infty |(e_i, u_j)|^2 \leq \sum_{j=1}^k \|u_j\|^2 = k,$$

with equality if the $\{e_i\}$ form a basis for H. □

So now suppose that there exists an orthogonal projection P that is injective on E_α and such that P^{-1} has Hölder exponent θ, $\|P(x-y)\| \geq C\|x-y\|^{1/\theta}$. Then in particular (since $0 \in E_\alpha$)

$$\|P(n^{-\alpha}e_n)\| \geq C\|n^{-\alpha}e_n\|^{1/\theta} \qquad \Rightarrow \|Pe_n\| \geq Cn^{[1-(1/\theta)]\alpha}.$$

It follows from Lemma 3.8 that

$$\operatorname{rank} P \geq \sum_{n=1}^{\infty} \|Pe_n\|^2 \geq C^2 \sum_{n=1}^{\infty} n^{2(1-(1/\theta))\alpha}.$$

If the rank of P is finite then the power of n must be less than -1, i.e. we must have

$$\theta < \frac{1}{1+(1/2\alpha)} = \frac{1}{1+(\dim(E_\alpha)/2)}. \tag{3.5}$$

This is not quite the bound we are after. However, recall that $\tau(E_\alpha) \leq \dim(E_\alpha)$. If we had $\tau(E_\alpha) < \dim(E_\alpha)$ then we could use the embedding theorem for k sufficiently large to improve on the bound in (3.5). But we cannot do this, and so in fact we must have $\tau(E_\alpha) = \dim(E_\alpha)$, showing that the best possible bound is indeed

$$\frac{1}{1+(\tau(E_\alpha)/2)}.$$

This bound cannot be improved if we insist on using the thickness exponent; but if we define a different quantity that measures how well X is approximate by finite-dimensional sets (rather than subspaces) we can do better.

3.4 Lipschitz deviation

A first step towards generalising the thickness exponent, the m-Lipschitz deviation was introduced by Olson & Robinson (2005). Denote by $\delta_m(X,\epsilon)$ the smallest dimension of a linear subspace U such that

$$\operatorname{dist}(X, G_U[\phi]) < \epsilon$$

for some m-Lipschitz function $\phi : U \to U^\perp$,

$$\|\phi(u) - \phi(v)\| \leq m\|u-v\| \qquad \text{for all} \qquad u, v \in U,$$

where $U^\perp$ is the orthogonal complement of U in H. We will write $G_U[\phi]$ for the graph of ϕ over U:

$$G_U[\phi] = \{\, u + \phi(u) : \ u \in U \,\}.$$

The m-Lipschitz deviation is given by

$$\mathrm{dev}_m(X) = \limsup_{\epsilon \to 0} \frac{\log \delta_m(X, \epsilon)}{-\log \epsilon}.$$

Pinto de Moura & Robinson (2009) defined the Lipschitz deviation of X to be

$$\mathrm{dev}(X) = \lim_{m \to \infty} \mathrm{dev}_m(X);$$

since $\mathrm{dev}_m(X)$ is non-increasing in m the limit clearly exists provided that $\mathrm{dev}_m(X)$ is finite for some $m > 0$.

Note that $\mathrm{dev}(X)$ is bounded above by $\tau(X)$, since $\mathrm{dev}_m(X) \leq \tau(X)$ for every $m > 0$ (one can always approximate by the graph of the zero function, which is m-Lipschitz). In fact this inequality can be strict.

One can replace the thickness by the Lipschitz deviation in Theorem 3.6. The key point in using the thickness exponent is that if $x, y \in X$ with $\|x - y\| \geq \epsilon$, and we choose a V such that $\mathrm{dist}(X, V) \leq \epsilon/3$, then $\|P(x-y)\| \geq \epsilon$, i.e. that $\|P(x - y)\|$ is comparable with $\|x - y\|$.

We can do something similar with graphs of Lipschitz functions, and this is why the proof still works:

Lemma 3.9 *Suppose that $\phi : U \to U^\perp$ is m-Lipschitz. Then*

$$|Qx - \phi(Px)| \leq \sqrt{2}\, m\, \mathrm{dist}(x, G_U[\phi]), \tag{3.6}$$

where P is the orthogonal projection onto U and $Q = I - P$.

Proof We have

$$\mathrm{dist}(x, G_U[\phi])^2 = \inf_{x \in U} \ |Px - u|^2 + |Qx - \phi(u)|^2.$$

But for any $x \in U$,

$$\begin{aligned}|Qx - \phi(Px)|^2 &= |Qx - \phi(u) + \phi(u) - \phi(Px)|^2 \\ &\leq 2|Qx - \phi(u)|^2 + 2|\phi(u) - \phi(Px)|^2 \\ &\leq 2|Qx - \phi(u)|^2 + 2m^2|u - Px|^2 \\ &\leq 2m^2(|Qx - \phi(u)|^2 + |u - Px|^2).\end{aligned}$$

Since this holds for all $x \in U$ we have (3.6). □

The advantage of this definition is that $\text{dev}(A) = 0$ for the attractors of a large class of partial differential equations.

There is an extensive theory of approximate inertial manifolds, introduced by Foias, Manley, & Temam (1998) and developed subsequently by a number of authors. Important here is that it is possible to prove the existence of families of "approximate inertial manifolds of exponential order" (Debussche & Temam, 1994; Rosa, 1995; Pinto de Moura & Robinson, 2009): these are precisely a collection of 1-Lipschitz functions ϕ_n from the finite-dimensional spaces H_n (in fact spanned by a finite collection of eigenfunctions of the linear term occurring in the equation), such that

$$\phi_n : H_n \to H_n^\perp \qquad |\phi(x) - \phi(y)| \leq |x - y|$$

and

$$\text{dist}(A, G_{H_n}[\phi_n]) \leq C\text{e}^{-\lambda_n^\alpha},$$

for some $\alpha > 0$, where n is the n^{th} eigenvalue, usually $\lambda_n \sim n^s$ for some s. Using this in the definition of Lipschitz deviation shows that $\text{dev}(A) = 0$.

The same argument as before shows that the estimate of the Hölder exponent in terms of the Lipschitz deviation cannot be improved (for the sets E_α we have $\text{dev}(E_\alpha) = \tau(E_\alpha) = d_{\text{box}}(E_\alpha)$).

3.5 The Takens time-delay embedding for subsets of H

We now combine the infinite-dimensional abstract theorem that we have just proved with a (generalised) version of the finite-dimensional Takens theorem, to prove an infinite-dimensional version.

The idea is essentially straightforward. Suppose that we have a Lipschitz map $\Phi : H \to H$ that has a finite-dimensional invariant set A. We suppose too for simplicity (but this can be relaxed) that $\mathrm{dev}(A) = 0$.

Theorem 3.6 guarantees that we can find a linear map $L : H \to \mathbb{R}^k$ that has an inverse that is Hölder continuous, with Hölder exponent θ as close to one as we like, provided we take k sufficiently large.

We now consider the set $X = LA$, and define a dynamical system on X by iterating the map $g = L \circ \Phi \circ L^{-1}$. Since L and Φ are Lipschitz but L^{-1} is only Hölder continuous, g is only Hölder continuous (with the same exponent θ as L^{-1}). However, unlike a general Hölder continuous map, whose Hölder exponent decreases under the operation of composition, the particular form of g means that

$$g^j = L \circ \Phi^j \circ L^{-1},$$

and so every iterate of g has the same Hölder exponent. This very particular form of g allows us to apply the following Hölder generalisation of the Takens result of Sauer et al. We make a very strong assumption on the non-existence of periodic orbits. Again this can be relaxed at the expense of a less clean result.

Theorem 3.10 *Let X be a compact subset of $\mathbb{R}^N$ with $d_{\mathrm{box}}(X) = d$, and $g : X \to X$ a map such that g^r is a θ-Hölder function for any $r \in \mathbb{N}$. Let $k > 2d/\theta$ ($k \in \mathbb{N}$) and assume that the set X_p of p-periodic points of g is empty for all $p = 1, \ldots, k$.*

Let $h_1, \ldots, h_m$ be a basis for the polynomials in N variables of degree at most $2k$, and given any θ-Hölder function $h_0 : \mathbb{R}^N \to \mathbb{R}$ define

$$h_\alpha = h_0 + \sum_{j=1}^{m} \alpha_j h_j.$$

Then for any γ with

$$0 < \gamma < 1 - \frac{2d}{k\theta},$$

the k-fold observation map $F_k : X \to \mathbb{R}^N$ defined by

$$F_k[h_\alpha, g](x) = \Big(h_\alpha(x), h_\alpha(g(x)), \ldots, h_\alpha(g^{k-1}(x))\Big)^T \tag{3.7}$$

satisfies

$$|F_k[h_\alpha, g](x) - F_l[h_\alpha, g](y)| \geq C_\alpha |x - y|^{1/\gamma}$$

for almost every $\alpha \in \mathbb{R}^m$; in particular $F_k[h_\alpha, g]$ is one-to-one on X.

We can now essentially combine this result with Theorem 3.6 to prove the following. For the details (but without the Hölder continuity) see Robinson (2005).

Theorem 3.11 *Let $\mathscr{A}$ be a compact subset of a Hilbert space H with upper box-counting dimension $d_{\rm box}(\mathscr{A}) = d$ and Lipschitz deviation zero. Choose an integer $k > (2+\tau)d$, and suppose further that $\mathscr{A}$ is an invariant set for a Lipschitz map $\Phi : H \to H$ and that there are no p-periodic points of Φ for $p = 1, \ldots, k$. Then for any α with*

$$0 < \alpha < 1 - \frac{k}{2d}$$

a prevalent set of Lipschitz maps $f : H \to \mathbb{R}$ make the k-fold observation map $D_k[f,\Phi] : H \to \mathbb{R}^k$ defined by

$$D_k[f,\Phi](u) = \Big(f(u), f(\Phi(u)), \ldots, f(\Phi^{k-1}(u))\Big) \tag{3.8}$$

satisfy

$$|D_k[f,\Phi](u) - D_k[f,\Phi](v)| \geq C_f \|u - v\|^{1/\alpha};$$

in particular $D_k[f,\phi]$ is one-to-one on $\mathscr{A}$.

For a result that limits the periods of periodic orbits for a class of infinite-dimensional dynamical systems arising for semilinear PDEs, see Robinson & Vidal López (2006).

3.6 Parametrisation by point values

Finally we will give a sketch proof of a result due to Friz & Robinson (2001) [and subsequently developed in Friz, Kukavica, & Robinson (2001) and Kukavica & Robinson (2004)] which gives a parametrisation of an attractor that consists of analytic functions in terms of the values of the functions at a sufficient number of points in the domain.

We will sketch a proof of the following result for attractors consisting of 2-component periodic functions on $\mathbb{R}^2$ (thinking of the 2D Navier-Stokes equations as the prime application), although a more general result is possible for d-component analytic functions on a bounded (or periodic) domain in $\mathbb{R}^m$ (Kukavica & Robinson, 2004). Although the main application of this result is to the attractors of various dissipative partial differential equations, note that the theorem in fact treats only a collection of functions with particular properties.

Essentially, the theorem says if the attractor consists of real analytic functions and has box-counting dimension d, then different elements of the attractor can be distinguished by comparing a finite number of their point values, with the number of observations required, k, comparable to the attractor dimension ($k \geq 16d + 1$): if u and v are elements of the attractor and

$$u(x_j) = v(x_j) \qquad j = 1, \ldots, k$$

then we must have $u(x) = v(x)$ throughout Ω. [For a similar result in the context of purely analytic systems see Sontag (2002).]

We write $Q = [0, 2\pi]^2$.

Theorem 3.12 *Let $\mathscr{A}$ be a compact subset of $L^2(Q, \mathbb{R}^2)$ with finite dimension $d_{\text{box}}(\mathscr{A})$ that consists of real analytic functions*[1]*. Then for $k \geq 16d_f(\mathscr{A}) + 1$ almost every set $\mathbf{x} = (x_1, \ldots, x_k)$ of k points in Ω makes the map $E_{\mathbf{x}}$, defined by*

$$E_{\mathbf{x}}[u] = (u(x_1), \ldots, u(x_k))$$

one-to-one between X and its image.

Furthermore the point values of u at $(x_1, \ldots, x_k)$ parametrize $\mathscr{A}$: the map $E_{\mathbf{x}}^{-1} : E_{\mathbf{x}}[\mathscr{A}] \to \mathscr{A}$ is continuous from $\mathbb{R}^{2k}$ into $L^2(Q, \mathbb{R}^2)$ and into $C^r(Q, \mathbb{R}^2)$ for every $r \in \mathbb{N}$.

"Almost every" is with respect to Lebesgue measure on Ω^k.

Before we begin to sketch the proof, we give an idea of why the condition of analyticity is required. Suppose that we have chosen $(x_1, \ldots, x_k)$, and that these are in fact a 'bad' set of points. This means that there are $u, v \in \mathscr{A}$ with $u \neq v$ such that

$$u(x_j) = v(x_j) \qquad \text{for every} \quad j = 1, \ldots, k.$$

Looking instead at the set of differences $X = \mathscr{A} - \mathscr{A}$, our chosen points are

[1] The most general result (whose proof is in some ways simpler) is in fact valid for sets $\mathscr{A}$ of functions defined on some bounded set Ω for which every non-zero difference $w = u - v$ has finite order of vanishing: this means that for every such w and every $x \in \Omega$ there exists a multi-index α such that $\partial^\alpha w(x) \neq 0$.

'bad' if there exists some non-zero $w \in X$ such that

$$w(x_j) = 0 \qquad \text{for every} \quad j = 1, \dots, k.$$

In other words, a collection of points is 'bad' if there is a non-zero element of X that is simultaneously zero at all these points. We use the analyticity of w to limit the size of its set of zeros.

We will need a more quantitative description of the analyticity of functions in $\mathscr{A}$, for which we make use of the notion of Gevrey classes (which were used to discuss the analyticity of solutions of the Navier-Stokes equations by Foias & Temam, 1989). These are only comfortably described in the periodic case, where we can use Fourier series.

Denote by A the negative Laplacian on Q with periodic boundary conditions and zero average. If we write

$$[L^2(Q)]^2 = \{u = \sum_{j \in \mathbb{Z}} u_j \mathrm{e}^{\mathrm{i} j \cdot x} : \ \sum_j |u_j|^2 < \infty\}$$

with $u_j \in \mathbb{C}^2$ and $u_{-j} = \overline{u_j}$, then since $A\mathrm{e}^{\mathrm{i} j \cdot x} = |j|^2 \mathrm{e}^{\mathrm{i} j \cdot x}$, we have

$$\|Au\|^2 = \sum_j |j|^4 |u_j|^2,$$

and more generally we can define

$$A^k u = \sum_j |j|^k u_j \mathrm{e}^{\mathrm{i} j \cdot x}$$

with $D(A^k u)$ (those u with $A^k u \in L^2$) given by

$$D(A^k u) = \{u = \sum_{j \in \mathbb{Z}} u_j \mathrm{e}^{\mathrm{i} j \cdot x} : \ \sum_j |j|^{2k} |u_j|^2 < \infty\}.$$

For a given $\tau > 0$ we define the Gevrey class G_τ to be $D(\mathrm{e}^{\tau A^{1/2}})$, i.e.

$$\{u = \sum_{j \in \mathbb{Z}} u_j \mathrm{e}^{\mathrm{i} j \cdot x} : \ \sum_j \mathrm{e}^{2\tau |j|} |u_j|^2 < \infty\},$$

and set

$$\|u\|_\tau^2 = \sum_j \mathrm{e}^{2\tau |j|} |u_j|^2.$$

One can show that any real analytic function on Q must be contained in one of the classes G_τ, for some $\tau > 0$ (see John, 1991); conversely, any

$u \in G_\tau$ is real analytic and can be extended to an analytic function in the region

$$S = \{z : \operatorname{Re} z \in Q, \ |\operatorname{Im} z| \le \tau/8\}$$

with

$$\sup_{z \in S} |u(z)| \le \beta \|u\|_\tau$$

for some $\beta > 0$. For a proof of this – which amounts to showing that the Taylor expansion of u converges, since $u \in G_\tau$ implies that $\|\partial^\alpha u\|_\infty \le C\|u\|_\tau (2/\tau)^k k!$ (where $k = |\alpha|$), and then using this to obtain an estimate on $|u(z)|$ – see Friz & Robinson (2001).

Proof (Very sketchy)

Rather than consider A itself, consider the set of differences $X = A - A$. Since each element of X is analytic, it lies in some G_τ. Consider the collection $\{X_n\}$ of sets

$$X_n = X \cap G_{1/n}.$$

Then $X = \cup_{n\in\mathbb{N}} X_n$.

Now fix $n \in \mathbb{N}$: we will show that for almost every collection of k points, the map $u \mapsto (u(x_1), \ldots, u(x_k))$ is non-zero on X_n.

Now, since X_n is bounded in $G_{1/n}$, it is possible to show (using an argument very similar to that used in Lemma 3.5) that the thickness of X_n measured in $G_{\varphi/n}$ is zero for any $0 < \varphi < 1$. By showing that for a set bounded in $G_{1/n}$ the identity map from L^2 into $G_{\varphi/n}$ is Hölder continuous (with Hölder exponent no less than $1 - \varphi$) one can also prove that the box-counting dimension of X_n measured in $G_{\varphi/n}$ is bounded according to

$$d_{\text{box}}(X_n; G_{\varphi/n}) \le \frac{2 d_{\text{box}}(A)}{1 - \varphi}.$$

The embedding theorem for subsets of Hilbert spaces now guarantees that there exists a θ-Hölder parametrisation of X_n using any $N > 4d_{\text{box}}(A)/(1 - \varphi)$ parameters, i.e. there exists a map $w : \mathbb{R}^N \to G_{\varphi/n}$ such that $w(x;\epsilon)$ ranges over all of X_n, and the mapping w is analytic in x and Hölder continuous in ϵ (as a map from $\mathbb{R}^N$ into $G_{\varphi/n}$) with Hölder exponent

$$\theta < 1 - \frac{4 d_{\text{box}}(A)/(1 - \varphi)}{N}.$$

That w is Hölder continuous into $G_{\varphi/n}$ implies that all the derivatives of w depend on ϵ in a θ-Hölder way. Because of this we can obtain a controlled description of the zero set of w: it is contained in a countable collection of manifolds given in the form

$$(x_1, x_2(x_1, \epsilon); \epsilon) \qquad \text{or} \qquad (x_1(x_2, \epsilon); \epsilon),$$

where $x_2(\cdot,\cdot)$ [or $x_1(\cdot,\cdot)$] is a θ-Hölder function of its $N+1$ arguments.

Collections of k simultaneous zeros are therefore contained in a countable collection of sets of the form

$$(x_1, y_1(x_1, \epsilon)) \times (y_2(x_2, \epsilon), x_2) \times \cdots \times (x_k, y_k(x_k, \epsilon)) \times \epsilon,$$

i.e. sets of the form

$$(X, Y(X, \epsilon); \epsilon),$$

where $Y = (y_1, \ldots, y_k)$ is a θ-Hölder function of the $N+k$ arguments $X = (x_1, \ldots, x_k)$ and ϵ.

It follows from a simple lemma on the Hausdorff dimension[1] (if $f : \mathbb{R}^N \to \mathbb{R}^m$ is θ-Hölder then the subset $\{(x, f(x)) : x \in \mathbb{R}^N\}$ of $\mathbb{R}^{N+m}$ has Hausdorff dimension at most $N + m(1-\theta)$; see Friz et al., 2001) that

$$d_{\mathrm{H}}(\text{one such product in } Q^k \times \mathbb{R}^N) \leq N + k + k(1-\theta).$$

Now, the Hausdorff dimension is stable under countable unions, so that

$$d_{\mathrm{H}}(\text{all bad products in } Q^k \times \mathbb{R}^N) \leq N + k(2-\theta);$$

and this dimension does not increase under projections (which are Lipschitz), so that

$$d_{\mathrm{H}}(\text{all bad collections in } Q^k) \leq N + k(2-\theta).$$

Since Hausdorff measure is proportional to Lebesgue measure, it follows that if

$$d_{\mathrm{H}}(\text{bad choices}) < d_{\mathrm{H}}(Q^k) = 2k.$$

It follows that it is sufficient to take $N < \theta k$ to ensure that almost every choice of k points is 'good'. By choosing φ very close to one we can take θ as close as we like to

$$1 - \frac{4 d_{\mathrm{box}}(A)}{N},$$

[1] The Hausdorff dimension is based on the definition of the Hausdorff measure $\mathscr{H}^s$, a generalisation of Lebesgue measure to non-integral values of s (for $s \in \mathbb{N}$ the s-dimensional Hausdorff measure is proportional to the s-dimensional Lebesgue measure). One defines $d_{\mathrm{H}}(X) = \inf\{s : \mathscr{H}^s(X) = 0\}$. One always has $d_{\mathrm{H}}(X) \leq d_{\mathrm{box}}(X)$, and unlike the box-counting dimension, the Hausdorff dimension is stable under countable unions (i.e. if $d_{\mathrm{H}}(X_j) \leq d$ then $d_{\mathrm{H}}(\cup_j X_j) \leq d$).

and so we require

$$k > \frac{N}{1 - \frac{4d_{\text{box}}(A)}{N}}.$$

The right-hand side is minimised if we take $N = 8d_{\text{box}}(A)$, and yields $k > 16d_{\text{box}}(A)$ as in the statement of the theorem. □

Note that the spatial nodes could be anywhere (there just have to be enough of them). But if we decided to divide the space into equal boxes and place one node in each box, then the side of the box would be $l \sim d_{\text{box}}(A)^{-1/2}$. This gives a rigorous derivation of the Landau–Lifshitz 'minimum lengthscale'.

References

Chepyzhov V V & Vishik M I (2002) *Attractors for Equations of Mathematical Physics* (Providence, AMS Colloquium Publications vol. 49, A.M.S.).

Christensen J P R (1973) Measure theoretical zero sets in infinite-dimensional spaces and applications to differntiability7 of Lipschitz mappings. *Publ. Dép. Math. (Lyon)* **10**, 29–39.

Constantin P & Foias C (1985) Global Lyapunov exponents, Kaplan-Yorke formulas and the dimension of the attractor for 2D Navier-Stokes equation. *Commun. Pure. Appl. Math.*, 38: 1–27

Debussche A & Temam R (1994) Convergent families of approximate inertial manifolds, *J. Math. Pures Appl.* **73** 489–522.

Doering C R & Gibbon J D (1995) *Applied Analysis of the Navier–Stokes equations* (Cambridge University Press, Cambridge).

Falconer K J (1990) *Fractal Geometry* (Wiley, Chichester).

Foias C, Manley O M, & Temam R (1988) Modelling of the interaction of small and large eddies in two-dimensional turbulent flows. *Math Modell. Numer. Anal.* **22**, 93–114.

Foias C & Temam R (1989) Gevrey class regularity for the solutions of the Navier-Stokes equations. *J. Funct. Anal.* **87**, 359–369.

Friz P K & Robinson J C (1999) Smooth attractors have zero "thickness". *J. Math. Anal. Appl.* **240** 37–46.

Friz P K & Robinson J C (2001) Parametrising the attractor of the two-dimensional Navier-Stokes equations with a finite number of nodal values. *Physica D* **148** 201–220.

Friz P K, Kukavica I, & Robinson J C (2001) Nodal parametrisation of analytic attractors. *Disc. Cont. Dyn. Sys.* **7** 643-657.

Hale J K (1988) *Asymptotic Behavior of Dissipative Systems*, Mathematical Surveys and Monographs Number 25 (American Mathematical Society, Providence, RI).

Hunt B R, Sauer T, & Yorke J A (1992) *Prevalence: a translation-invariant almost every for infinite dimensional spaces.* Bull. Amer. Math. Soc. **27** 217–238; (1993) Prevalence: an addendum. *Bull. Amer. Math. Soc.* **28** 306–307.

Hunt B R & Kaloshin V Y (1999) Regularity of embeddings of infinite-dimensional fractal sets into finite-dimensional spaces. *Nonlinearity* **12** 1263–1275.

John F (1991) *Partial differential equations* (Reprint of the fourth edition; Springer-Verlag, New York).

Kukavica I & Robinson J C (2004) Distinguishing smooth functions by a finite number of point values, and a version of the Takens embedding theorem. *Physica D* **196**, 45-66.

Kukavica I (1994) An absence of a certain class of periodic solutions in the Navier-Stokes equations. *J. Dynam. Differential Equations* **6**, 175–183.

Landau, L D & Lifshitz E M (1959) *Fluid Mechanics* (Course of Theoretical Physics, vol. 6., Pergamon).

Mañé R (1981) On the dimension of the compact invariant sets of certain nonlinear maps, *Springer Lecture Notes in Math.* **898**, Springer, New York, 230-242.

Olson E J and Robinson J C (2005) Almost bi-Lipschitz embeddings and almost homogeneous sets, *Trans. Amer. Math. Soc.*, to appear.

Pinto de Moura E & Robinson J C (2009) Global attractors with zero Lipschitz deviation, submitted to *Nonlinearity.*

Robinson, J C (1999) Global attractors: topology and finite-dimensional dynamics. *J. Dynam. Diff. Eq.* **11**, 557–581.

Robinson J C (2001) *Infinite-dimensional dynamical systems* (Cambridge University Press, Cambridge).

Robinson J C (2005) A topological delay embedding theorem for infinite-dimensional dynamical systems, *Nonlinearity* **18** 2135–2143.

Robinson J C (2007) Parametrization of global attractors, experimental observations, and turbulence, *J. Fluid Mech.* **578** 495–507.

Robinson, J C (2008) A topological time-delay embedding theorem for infinite-dimensional cocycle dynamical systems. *Discrete Contin. Dyn. Syst. Ser. B* **9**, 731–741.

Robinson J C & Vidal López, A (2006) Minimal periods of semilinear evolution equations with Lipschitz nonlinearity. *J. Diff. Eq.* **220**, 396–406

Rosa R (1995) Approximate inertial manifolds of exponential order, *Discrete Contin. Dynam. Systems* **1** 421–448.

Sauer T, Yorke J A, & Casdagli M (1991) Embedology *J. Stat. Phys.* **71** 529-547

Temam R (1988) *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, Springer Applied Mathematical Sciences Volume 68 (Springer Verlag, Berlin).

Yorke J A (1969) Periods of periodic solutions and the Lipschitz constant. *Proc. Am. Math. Soc.* **22**, 509–512.