Function spaces and compactness

Vaughn Climenhaga

February 13, 2013

In the last two posts on spectral methods in dynamics, we've used (both explicitly and implicitly) a number of results and a good deal of intuition on function spaces. It seems worth discussing these a little more at length, as a supplement to the weekly seminar posting.

1 Function spaces and extra structure

It is useful to treat real-valued functions (or complex-valued functions, or vector space-valued functions) as elements of a vector space, so that the tools from linear algebra can be applied. Given a set X one may consider the vector space $\mathbb{R}^X$ of all real-valued functions with domain X. If X is finite, say with n elements, then this is just the familiar vector space $\mathbb{R}^n$. The more interesting examples are when X is infinite, and so $\mathbb{R}^X$ is infinite-dimensional. We will focus on the case $X = [0, 1]$, which is reasonably representative.

Generally speaking, the functions $[0, 1] \to \mathbb{R}$ that arise from some application are not entirely arbitrary, but have some degree of regularity – maybe they are continuous, or piecewise continuous, or measurable, or integrable, etc. It turns out that the vector space $\mathbb{R}^{[0,1]}$ is "too large" for many applications, and that it is more suitable to consider a smaller space, whose elements are functions with some extra properties. We will consider some of the ways to do this, paying particular attention to how those choices let us recover certain properties of $\mathbb{R}^n$ that involve extra structure beyond that of the vector space itself:

- *Topology:* We know what it means for a sequence $\vec{x}_k \in \mathbb{R}^n$ to converge to some $\vec{x} \in \mathbb{R}^n$, and we want a similar notion of convergence in a vector space $V \subset \mathbb{R}^{[0,1]}$.

- *Metric and norm:* We want the notion of convergence to come from a metric (distance function) that is compatible with the vector space structure of V – that is, a norm, with respect to which the vector space V becomes a Banach space.

- *Compactness:* A subset of $\mathbb{R}^n$ is compact if every sequence in that subset has a convergent subsequence, and this property is important in many applications and proofs. By the Heine–Borel theorem compactness in $\mathbb{R}^n$ is equivalent to being closed and bounded. How can we determine when a set of functions in V is compact?

2 Continuous functions and Arzelà–Ascoli

The extra structure we seek to place on $V \subset \mathbb{R}^{[0,1]}$ should leverage some of the extra structure that $[0, 1]$ has, beyond simply being an uncountable set. In particular, we may use either the topology of $[0, 1]$ or Lebesgue measure on $[0, 1]$ to define properties of functions $f \colon [0, 1] \to \mathbb{R}$. First we discuss the topological option – later we see what happens when we use the measure-theoretic structure to define the L^p spaces (and others).

The natural space to use is $C(X)$, the space of continuous real-valued functions on $X = [0, 1]$, with the norm $\|f\|_{C^0} = \sup_{x \in [0,1]} |f(x)|$. The space of continuous functions is complete with respect to this norm, and so we have a Banach space. What about compactness? How do we tell if a set $A \subset C(X)$ is compact? Of course A should be closed, but what else do we need? Boundedness is no longer enough: the unit ball in $C(X)$ is not compact, as can be seen by considering the sequence of functions shown in Figure 1.

The solution here is given by the Arzelà–Ascoli theorem: a set $A \subset C(X)$ is pre-compact (has compact closure) if and only if the following conditions are satisfied.

- A is uniformly bounded: $\sup_{f \in A} \sup_{x \in X} |f(x)| < \infty$.

- A is equicontinuous: for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ for every $f \in A$ and $|x - y| < \delta$.

Remark 1 *The proof that these conditions guarantee compactness uses the following strategy, which it is a useful exercise to complete:*
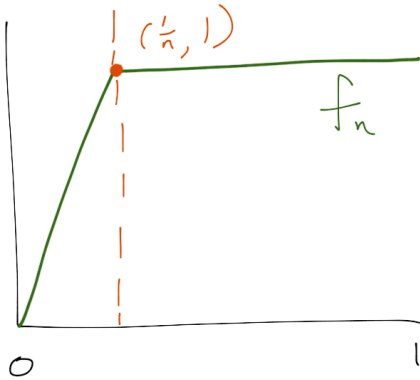
Figure 1: Uniformly bounded but no convergent subsequence.

1. *Given any sequence $f_n \in A$, use uniform boundedness and a diagonalisation argument to find a subsequence that converges at every rational number. (Or on some other countable dense set.)*

2. *Use equicontinuity to guarantee that $\{f_{n_k}(x)\}_{k \geq 1}$ is Cauchy for every $x \in [0,1]$, and hence converges.*

In particular, one can consider the subspace $C^\alpha(X) \subset C(X)$ of Hölder continuous functions with exponent $\alpha \in (0,1)$ – this is a Banach space with norm

$$\|f\|_{C^\alpha} = \|f\|_{C^0} + |f|_\alpha, \qquad |f|_\alpha = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|^\alpha}.$$

When $\alpha = 1$ this is the space of Lipschitz functions. If $A \subset C^\alpha(X)$ is uniformly bounded *in the C^α norm*, then it is uniformly bounded in the C^0 norm and equicontinuous, and hence it is pre-compact *in the C^0 norm*.

It is important to note here the structure of the last statement – we have two norms, $\|\cdot\|_{C^\alpha}$ and $\|\cdot\|_{C^0}$, such that **uniform boundedness in one norm implies pre-compactness in the other**. This is the closest that we can come to an infinite dimensional analogue of Heine–Borel: as a consequence of Riesz's lemma, every infinite-dimensional Banach space has a uniformly bounded sequence with no convergent subsequence.

In our study of spectral methods in dynamics, an important step is always to find two norms with this relationship: uniform boundedness in one implies pre-compactness in the other. We remark that the Arzelà–Ascoli theorem actually gives just a little bit more than this: given a sequence $f_n \in C(X)$

that is uniformly bounded in the C^α norm, pre-compactness only guarantees the existence of a limit point $f_{n_k} \xrightarrow{C^0} f \in C^0$, but in fact the limit point f is in C^α as well, because any modulus of continuity for the sequence f_n is also a modulus of continuity for any limit point.

Another important family of function spaces, which leverages not only the topological but also the differentiable structure of the unit interval, are the spaces C^r, defined inductively as

$$C^{r+1} = \{f \colon [0,1] \to \mathbb{R} \mid f \text{ is differentiable and } f' \in C^r\}.$$

Here r need not be an integer (the base case for the induction is $0 \leq r < 1$), so for example, for $0 < \alpha < 1$, $C^{1+\alpha}$ is the space of differentiable functions whose derivatives are Hölder continuous with exponent α. The space C^r becomes a Banach space when endowed with the norm inductively given by

$$\|f\|_{C^{r+1}} = \|f\|_{C^0} + \|f'\|_{C^r}.$$

For example, on C^1 the appropriate norm is

$$\|f\|_{C^1} = \|f\|_{C^0} + \|f'\|_{C^0}. \tag{1}$$

The relationship discussed above between uniform boundedness in one norm and pre-compactness in another can be stated quite generally for this family of norms: **uniform boundedness in the C^r norm implies pre-compactness in the C^s norm for any $0 \leq s < r$.** This relationship is often expressed by saying that "C^r is compactly embedded in C^s for $r > s$".

3   Lp spaces

In terms of the measure-theoretic structure of the unit interval, the most important function spaces are the L^p spaces

$$
\begin{aligned}
L^p &= L^p([0,1], dx) \\
&= \bigg\{ f \colon [0,1] \to \mathbb{R} \,\Big|\, f \text{ is measurable and} \\
&\qquad\qquad \|f\|_p := \left(\int_{[0,1]} |f(x)|^p \, dx \right)^{\frac{1}{p}} < \infty \bigg\},
\end{aligned}
$$

where $1 \le p < \infty$, and

$$L^\infty = \{f\colon [0,1] \to \mathbb{R} \mid f \text{ is measurable and } \|f\|_\infty < \infty\},$$

where $\|f\|_\infty = \sup\{L \ge 0 \mid \{x \in [0,1] \mid |f(x)| > L\} \text{ has positive Lebesgue measure}\}$ is the essential supremum of f.

In fact, this definition cheats a little bit, because elements of an L^p space are actually *equivalence classes* of functions, where two functions are equivalent if they agree on a set of full Lebesgue measure. This throws a small technical monkey wrench into many arguments involving L^p spaces, since strictly speaking an expression like $f(x)$ for $f \in L^p$ has no meaning unless it is inside an integral sign. One way to avoid these technicalities is to emphasise the role of elements of L^p not necessary as functions, but rather as linear functionals.

Recall that if $\mathcal{B}$ is a Banach space, then $\mathcal{B}^*$ is the dual space of continuous linear functionals $\mathcal{B} \to \mathbb{R}$. The L^p spaces have the property that

$$(L^p)^* = L^q \text{ for } 1 < p, q < \infty \text{ such that } \frac{1}{p} + \frac{1}{q} = 1,$$

where $f \in L^p$ defines a linear functional on L^q by

$$g \mapsto \int f \cdot g \, dx \text{ for } g \in L^q. \tag{2}$$

Thus instead of thinking of a function $f \in L^p$, we may think of the associated functional in (2), which is obtained by integrating the function f against *test functions* from a suitable space. In this case the space of test functions is taken to be L^q, but there are many other examples we could consider – eventually this leads to the idea of considering distributions in place of functions, but we will not go this far here.

Remark 2 *Before moving on, we note that* $(L^1)^* = L^\infty$, *but* $(L^\infty)^*$ *is a larger space than* L^1.

4   Weak derivatives

An important use of this alternate viewpoint – functions as continuous linear functionals – is to define the weak derivative of a function. If $f\colon [0,1] \to \mathbb{R}$ is

differentiable, then for any differentiable $g \colon [0,1] \to \mathbb{R}$ with $g(0) = g(1) = 0$, integration by parts gives

$$\int f' \cdot g \, dx = - \int f \cdot g' \, dx. \tag{3}$$

Equation (3) characterises the derivative f', which motivates the following definition: $h \in L^1$ is the weak derivative of $f \in L^1$ if

$$\int h \cdot \varphi \, dx = - \int f \cdot \varphi' \, dx \text{ for all } \varphi \in \mathcal{G}, \tag{4}$$

where the space of test functions is $\mathcal{G} = \{\varphi \in C^1([0,1], \mathbb{R}) \mid \varphi(0) = \varphi(1) = 0\}$. Write $h = Df$ in this case.

Example 1 *The absolute value function $f(x) = |x|$ has as its derivative the step function $Df(x) = -1(x < 0), 1(x > 0)$. Note that the value of $Df(0)$ is not uniquely defined because Df is considered as an element of L^1.*

Writing $g(x) = Df(x)$ for the step function just described, we see that g does not have a weak derivative in L^1. Indeed, this is true for any function with a jump discontinuity.

Using mollifiers one can show that any L^1 function f can be L^1 approximated by (infinitely) differentiable functions f_ϵ such that f'_ϵ approximates Df in L^1. This can be used to show that the usual product rule for derivatives holds for weak derivatives as well: $D(fg) = (Df) \cdot g + f \cdot (Dg)$, as long as f and g both have weak derivatives. The space of L^1 functions with a weak derivative in L^1 is denoted $W^{1,1}$ and is an important example of a Sobolev space. Here the norm is

$$\|f\|_{W^{1,1}} = \|f\|_{L^1} + \|Df\|_{L^1},$$

which can be viewed as an analogue of the definition of the C^1 norm in (1). Moreover, just as the C^1 unit ball is C^0 compact, so also the $W^{1,1}$ unit ball is L^1 compact, as we will see.

5 Kolmogorov–Riesz compactness theorem

In understanding compactness for subsets of function spaces, it is useful to recall that the Heine–Borel theorem can be generalised to arbitrary complete

metric spaces as follows: a set is compact if and only if it is closed and totally bounded. In particular, for Banach spaces, pre-compactness is equivalent to being totally bounded.

The Arzelà–Ascoli theorem gives a necessary and sufficient condition for a set in C^0 to be totally bounded (and hence pre-compact). A similar result in the L^p spaces is the Kolmogorov–Riesz compactness theorem – an expository account of this theorem and its relationship to the Arzelà–Ascoli theorem is given in a recent paper by H. Hanche–Olsen and H. Holden, The Kolmogorov–Riesz compactness theorem (*Expo. Math.* **28** (2010), 385–394).

In our setting (where we consider L^p spaces with respect to a *finite* measure), the Kolmogorov–Riesz theorem can be stated as follows: a set $\mathcal{F} \subset L^p$ is totally bounded (in the L^p norm) if and only if

1. $\mathcal{F}$ is bounded, and

2. for every $\varepsilon > 0$ there is $\delta > 0$ such that $\|f \circ T_\gamma - f\|_p < \varepsilon$ for every $f \in \mathcal{F}$ and $|\gamma| < \delta$, where $T_\gamma \colon x \mapsto x + \gamma$.

In other words, to go from bounded to totally bounded one needs the added condition that small changes to the argument result in (uniformly) small changes in the function, with respect to the L^p norm.

Roughly speaking the idea is that if a set can be "approximately embedded" into a totally bounded set, then it must itself be totally bounded – this is Lemma 1 in the paper referred to above. Then the condition on $f \circ T_\rho - f$ for $f \in \mathcal{F}$ allows the set $\mathcal{F}$ to be "approximately embedded" into a bounded set in $\mathbb{R}^n$ by averaging f over small neighbourhoods in its domain. This is of course a very rough description and one should read the paper for the complete proof and precise formulation of what it means to be "approximately embedded".

6   Bounded variation and Helly's theorem

One can use the Kolmogorov–Riesz theorem to show that $W^{1,1}$ is compactly embedded in L^1. (This is a special case of the Rellich–Kondrachov theorem.) However, since functions with jump discontinuities are not in $W^{1,1}$, we want to use a bigger function space in order to study spectral properties of the transfer operator.

The definition of weak derivative can be generalised if one is willing to allow Df to live somewhere besides L^1. Recall that we want Df to satisfy

$$\int (Df) \cdot \varphi \, dx = - \int f \cdot \varphi' \, dx$$

for every test function $\varphi \in \mathcal{G}$, the space of C^1 functions on the interval that vanish at the endpoints. The left-hand side defines a linear functional $\mathcal{G} \to \mathbb{R}$, and given any $f \in L^1$ we may define Df as such a linear functional by setting

$$(Df)(\varphi) = - \int f \cdot \varphi' \, dx.$$

If $f \notin W^{1,1}$, this functional is not given by integration against an L^1 function, but now the definition makes sense for any $f \in L^1$. Moreover, the space of linear functionals on $\mathcal{G}$ carries a natural norm: the norm of $\ell \colon \mathcal{G} \to \mathbb{R}$ is

$$\|\ell\|_{\mathcal{G}^*} = \sup\{|\ell(\varphi)| \mid \varphi \in \mathcal{G}, \|\varphi\|_{C^0} \leq 1\}.$$

A functional ℓ is continuous if and only if $\|\ell\| < \infty$. Recalling our discussion of bounded variation functions in an earlier post, we see that $\|Df\|_{\mathcal{G}^*} = |f|_{BV}$, and so

$$BV = \{f \in L^1 \mid \|Df\|_{\mathcal{G}^*} < \infty\}.$$

The BV norm can be written as $\|f\|_{BV} = \|f\|_{L^1} + \|Df\|_{\mathcal{G}^*}$. Note that BV is exactly the set of functions $f \in L^1$ for which Df is a *continuous* linear functional on $\mathcal{G}$.

Helly's selection theorem states that BV is compactly embedded in L^1. (This is not to be confused with Helly's theorem in geometry.) This is a consequence of the Kolmogorov–Riesz compactness theorem, because a relatively straightforward computation shows that

$$\|f \circ T_\gamma - f\|_{L^1} \leq |f|_{BV} |\gamma|.$$

(See Lemma 11 and Theorem 12 in the paper of Hanche–Olsen and Holden referenced above.) We remark that one can also give a direct proof following the hint given in Footnote 8 of Keller and Liverani's A spectral gap for a one-dimensional lattice of coupled piecewise expanding interval maps: given $f \in BV$, let f_n be the step function that is constant on each dyadic interval $[k, k+1]/2^n$, with value equal to the average of f on that interval. Then the functions f_n approach f in L^1, and the problem reduces to finding a suitable subsequence of step functions.