Metric Space Compactness, Arzelá Ascoli, and Stone-Weierstrass

Ethan Ebbighausen

May 11, 2026

1 Introduction

This document is a supplementary material for my classes (usually either analysis or PDEs). The goal is to view two important theorems about the space of continuous functions, but it takes some slight detours to contextualize the two theorems. In particular, we talk a bit about compactness that leads up to why we have a special theorem about compactness in some function spaces, and a bit about interpolation to guide why we might expect polynomials to be important or dense with respect to continuous functions.

2 Compactness in General Metric Spaces

Compact spaces are nice to work with in analysis in part because they are *sequentially compact*, which means any sequence has an accumulation point within the set. In metric spaces (assuming the Axiom of Countable Choice), this is actually equivalent to being compact.

In $\mathbb{R}^n$, the proof of equivalence is the pair of the Heine-Borel and Bolzano-Weierstrass theorems, which give that both properties are equivalent to a set being closed and compact. In any metric space, a closed set is compact and bounded, but the converse fails.

Example 1: Recall that the support of a function $f : A \to \mathbb{R}$, denoted $\text{supp}(f)$, is the closure of $f^{-1}(\{x \in A : f(x) \neq 0\}$.

Consider a smooth bump ϕ which has $\phi(0) = 1$ and $\text{supp}(\phi) \subset [-1, 1]$. The sequence $\{\phi(x - 2n)\}$ is closed and bounded in $C(\mathbb{R})$, but fails to be sequentially compact since no subsequence has a limit point ($||\phi(x - 2n) - \phi(x - 2m)|| = 1$ if $n \neq m$). Hence, it fails to be compact.

In fact, the closed unit ball fails to be compact in many infinite-dimensional spaces. For any Hilbert space, it is compact iff. the space is finite dimensional (this uses the inner product via the Riesz lemma).

To show this equivalence, we first generate a critereon for compactness in a general metric space. This turns out to be a generalization of being bounded. We call a metric space (X, d) totally bounded if for every $\epsilon > 0$, there exists a finite collection of open balls of radius ϵ which cover X. For example, on $\mathbb{R}$ this is true if and only if the set is bounded. We will see more examples of totally bounded sets in function spaces in the following sections.

Since ϵ-balls form an open cover of any metric space, any compact set is also totally bounded immediately. We may see a converse by generalizing the proof of the Heine-Borel Theorem.

[2.1] A metric space (X, d) is compact if and only if it is closed and totally bounded

Proof:
The forward direction is immediate as above, so let us consider the reverse. Let (X, d) be a metric space which is closed and totally bounded. Let $\{U_\alpha\}$ be an arbitrary open cover of X. Assume for contradiction that X fails to have a finite subcover.

Since X is totally bounded, it has some fixed diameter t. We may cover X by finitely many balls of radius $t/4$. At least one of those balls fails to have a finite subcover as well. Call this ball B_1. We may repeat this same process with the radius $t/8$ to get a ball $B_2 \subset B_1$. Repeating this process renders a sequence of balls B_i of radii $\frac{t}{2^{1+i}}$. Let c_i be the center of ball B_i, and notice that if $j > i$, $c_j \in B_i$ implies $d(c_i, c_j) < \frac{t}{2^{1+i}}$. This sequence is then Cauchy, and converges to some $x \in X$ by the assumption that X is closed.

We must have $x \in U_\alpha$ for some α. Let $\epsilon > 0$ be such that $B_\epsilon(x) \subset U_\alpha$. We may then find N so for $n \geq N$, $c_n \in B_{\epsilon/2}(x)$ and $\frac{t}{2^{1+n}} < \epsilon/2$. This implies $B_n \subset U_\alpha$, contradicting that no finite subcover of this U_α exists.

Q.E.D.

[Where in the above do we use the Axiom of (Countable) Choice?]

Now a sequentially compact metric space is immediately closed, so we need only show the other property to ascertain our goal.

[2.2] A sequentially compact metric space is totally bounded

Proof:

By contradiction, assume that the sequentially compact metric space (X, d) is not totally bounded. Then, given any $\epsilon > 0$, we may (by the Axiom of choice) pick some sequence $\{x_n\}$ such that $B_\epsilon(x_n) \cap B_\epsilon(x_m) = \emptyset$ for $n \neq m$. Thus, no subsequence of this sequence may converge, contradicting that the space is sequentially compact.

Q.E.D.

3 The Case of Continuous Functions

From Example 1 above, we notice that a closed and bounded set in $C(\mathbb{R})$ is not compact. One simplification to understand what total boundedness looks like here may be to reduce to considering $C([a,b])$. We try to generate a similar example.

Example 2:

Let $\epsilon = b - a$. Let $x_n = a + \sum_{i=1}^{n} \epsilon 2^{-n}$. Then, $d(x_n, x_{n+1}) = \frac{\epsilon}{2^{n+1}}$. Let ϕ be a smooth bump so $\phi(0) = 1$ and $\text{supp}(\phi) \subset [-1, 1]$. Consider the sequence $\phi_n(x) = \phi(2^{n+2}x - x_n)$. Then, the ϕ_n are all supported in $[a, b]$ and $||\phi_n - \phi_m|| = 1$ if $n \neq m$.

Since our domain is a compact space, the centers of these bumps must converge to some limit point (endpoint b in the example). Therefore, instead of acting the same, we force the ϕ_n to be sharper and sharper bumps to keep the same disjoint support property. It might not be clear that this is true in every case just yet, but it turns out that this issue is exactly what we want to target. Let us formalize the "sharpness".

A family of functions $\mathcal{F} \subset C(X)$ is said to be equicontinuous if for all $x \in X$ and all $\epsilon > 0$, there exists some $\delta > 0$ such that if $d(x, y) < \delta$, then for any $f \in \mathcal{F}$, $d(f(x), f(y)) < \epsilon$. In other words, the modulus of continuity is the same across the family. The family is said to be uniformly equicontinuous if the equivalent idea holds with uniform continuity: for all $\epsilon > 0$ there exists $\delta > 0$ such that $d(x, y) < \epsilon$ implies $d(f(x), f(y)) < \delta$ for all $x, y \in X$ and $f \in \mathcal{F}$.

Example 3:

The family $\{\alpha x\}_{\alpha \in [0,1]}$ is uniformly equicontinuous on $\mathbb{R}$ (in fact, it is equiLipschitz!). The family $\{\sin(\alpha x)\}_{\alpha \in \mathbb{R}}$ is not equicontinuous on any closed interval. The family $\{\alpha x^2\}_{\alpha \in [0,1]}$ is equicontinuous but not uniformly equicontinuous on $\mathbb{R}$.

[3.1] Arzelá-Ascoli Theorem: Consider a sequence $\{f_n\}$ of real valued functions on the closed and bounded interval $[a, b] \subset \mathbb{R}$. If $\{f_n\}$ is uniformly equicontinuous and uniformly bounded, then there exists a subsequence $\{f_{n_k}\}$ which converges uniformly to some $f \in C([a, b])$ [1,2]. We will follow the proof presented in [4]

Proof:

The proof of this theorem is a premier example of what is now known as a diagonalization argument. We will create a sequence of nested subsequences with some property improving as we go down the ladder. Taking the diagonal elements gives a subsequence embodying the property we desire.

In this case, our goal is to create subsequences that converge at more and more points of $[a, b]$. Consider $\mathbb{Q} \cap [a, b] = \{x_1, x_2, ...\}$. Let C be the uniform bound of the sequence. Since $f_n(x_1) \in [-C, C]$, we know that there is a subsequence n_k so $\{f_{n_k}(x_1)\}$ is convergent. We set $f_{1,k} = f_{n_k}$. We may similarly find some a subsequence so $\{f_{1,k_m}(x_2)\}$ is convergent, and we label this as $f_{2,m} = f_{1,k_m}$. Repeating indefinitely gives a collection $\{f_{n,k}\}$.

Consider the $\{f_{n,n}\}_{n\in\mathbb{N}}$. By construction, at any rational point in $[a, b]$, $f_{n,n}$ converges. The density of the rationals now implies that $f_{n,n}(x)$ converges for any x. Indeed, pick some $\epsilon > 0$. We may find some δ so that for all n, $d(x, y) < \delta$ implies $d(f_n(x), f_n(y)) < \delta/3$. Pick some rational y within this δ-ball of x, and there exists some N so for all $n, m \geq N$

$$\begin{aligned} d(f_{n,n}(x), f_{m,m}(x)) &\leq d(f_{n,n}(x), f_{n,n}(y)) + d(f_{n,n}(y), f_{m,m}(y)) + d(f_{m,m}(y), f_{m,m}(x)) \\ &\leq \epsilon \end{aligned}$$

and the sequence is pointwise Cauchy. Consider N. By the uniform equicontinuity, this same N works for all $x \in B_\delta(y)$. Therefore, we consider $B_\delta(y)$ for each rational $y \in [a, b]$, and call the associated number $N(y)$. This generates an open cover of a compact set, so we have some finite subcover with centers $y_1, ..., y_m$. Set $N = \max\{N(y_1), ..., N(y_m)\}$, and this N now works for all $x \in [a, b]$. In other words, for $n, m \geq N$ and any x,

$$\begin{aligned} d(f_{n,n}(x), f_{m,m}(x)) &\leq d(f_{n,n}(x), f_{n,n}(y_j)) + d(f_{n,n}(y_j), f_{m,m}(y_j)) + d(f_{m,m}(y_j), f_{m,m}(x)) \\ &\leq \epsilon \end{aligned}$$

where y_j is one of the centers. We have now shown that the sequence is uniformly Cauchy, and so it has a uniform limit.

Q.E.D.

We may immediately upgrade the Arzelá-Ascoli theorem to a family of functions. If a family $\mathcal{F}$ is closed, uniformly equicontinuous and uniformly bounded, the latter two properties hold on any sequence in $\mathcal{F}$. The sequence must then have a convergent subsequence, and the limit must be in $\mathcal{F}$ since it is closed. Therefore, in the metric space $C([a, b])$, $\mathcal{F}$ is sequentially compact. As we know from the previous section, $\mathcal{F}$ is compact. This may be further extended to the algebra of continuous functions on a locally compact Hausdorff space, as opposed to a compact subset of the real line.

Example 3:

The sequence in Example 1 is uniformly bounded and uniformly equicontinuous, but not precompact. Hence, the boundedness assumption for the domain is necessary.

However, if we consider a uniformly continuous function f on an open interval (a, b), then f may be extended to the closed interval $[a, b]$ with the same modulus of continuity. To see this, note that if it were not the case, we must have a sequence $\{x_n\}$ converging to one of the endpoints such that $f(x_n)$ fails to converge, i.e. there exists some $\epsilon > 0$ such given any n, there is $m > n$ with $d(f(x_n), f(x_m)) > \epsilon$. This violates uniform continuity. For this same ϵ, we may pick $\delta > 0$ so $d(x, y) < \delta$ gives $d(f(x), f(y)) < \epsilon$. Considering that all x_n are in a δ-neighborhood of the endpoint for large n, we have a direct contradiction.

We may use this fact to extend a uniformly equicontinuous sequence $\{f_n\}$ on (a, b) to a uniformly equicontinuous family on $[a, b]$ (though the modulus of continuity may change slightly).

4 Functional Approximation

In many cases in analysis, we prove some result by showing it holds on a dense set of points. For example, in the Arzelá-Ascoli Theorem above, we proved convergence on the dense set of rational points to show it held at every point. This often allows us to reduce to something technically easier. On the other hand, when doing numerical calculations, we often don't have a great form to look at an abstract function and only have a finite number of computed function values. We want to represent the function by something easy to compute without introducing too much error.

For both of these reasons, we may try to view a variety of functions by approximating them by polynomials. Recall from linear algebra that if we have points (x_i, y_i) for $i = 0, 1, 2, ..., n$, these exists a unique degree-n polynomial passing through these points given by

$$p(x) = \sum_{i=0}^{n} y_i l_i(x) \text{ for } \quad l_i(x) = \prod_{j \neq i} \frac{x - x_j}{x_i - x_j}$$

This is called the Lagrange form of the interpolant, where the interpolant is the polynomial passing through our desired points. If we have $y_i = f(x_i)$, then this polynomial interpolates the function f. How do we know whether this polynomial is close to the function, however? If we assume f is quite regular, in fact that $f \in C^{n+1}$, then we may see that, for some ξ_x depending on x,

$$f(x) - p(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi_x) \prod_{i=0}^{n} x - x_i$$

To prove this fact, we denote $w(t) = \prod_{i=0}^{n} t - x_i$ and construct $\phi(t) = f(t) - p(t) - \frac{f(x)-p(x)}{w(x)} w(t)$. Then, $\phi(t)$ has $n+1$ zeroes in an interval containing all of the x_i, and so by Rolle's theorem we have $\phi^{(n+1)}(\xi) = 0$ for some point in the same domain. Rearranging the derivative gives the above equation [3].

We might expect that for continuous functions, we should always see that the interpolants uniformly approximate the function as we take many, many nodes. However, this is not true. These are several examples of nonconvergence, but an easy one is due to Runge from 1901, where he investigated $(x^2+1)^{-1}$ on the interval $[-5, 5]$ with evenly spaced nodes and saw that the interpolants fail to converge [Runge : https://archive.org/details/zeitschriftfrma12runggoog]. Roughly speaking, this is because of the complex plane. The function $(x^2+1)^{-1}$ has poles at $\pm i$, and so it is discontinuous here. Polynomials, of course, are continuous and entire, and so they cannot uniformly converge to a discontinuous function in this ball in the complex plane. Depending on our choice of nodes, that means we may also fail to converge on real intervals.

This may influence us to look into which functions are real analytic (may be represented by a power series converging everywhere). Indeed, in the study of such functions, Weierstrass was able to prove the following approximation theorem.

[4.1] Weierstrass Approximation Theorem: Suppose f is a continuous function on the compact interval $[a, b]$. For every $\epsilon > 0$, there exists a polynomial p such that $||f - p||_{C([a,b])} < \epsilon$.

There are several constructive proofs for this theorem, including a proof using Bernstein Polynomials (originally given by Bernstein) and another using convolution with a family of polynomial kernels (given by Rudin). If anyone who speaks German is willing to take a look, I would love to know what the original proof of Weierstrass entails.

Going back to interpolation momentarily, one may combine the above theorem with the Chebyshev alternation theorem to show that, given a continuous function, there exists some system of nodes so interpolation converges to the function.

What makes the polynomials special so that they can do this? Could we do this with trigonometric functions or exponentials First, notice that $C([a, b])$ with the supremum norm is a Banach algebra (a complete normed vector space closed under multiplication and with some other algebraic structure). The set of polynomials are also a vector subspace closed under multiplication, has the 1 function, and has the important property that they can separate points: given $x \neq y$, we may find p a function so $p(x) \neq p(y)$. In fact, this is enough for our approximation!

[4.2] Stone-Weierstrass Theorem: Suppose X is a compact Hausdorff space and A is a vector subspace of $C(X, \mathbb{R})$ which is closed under multiplication and contains a nonzero constant function (hence contains the 1 function). Then, A is dense in $C(X, \mathbb{R})$ if and only if it separates points.

Proof:

The proof we will present was given by M.H.Stone [5] in 1948 (He actually first proved the result in 1937, but gave this simpler proof later), that uses a clever idea that the maximum and minimum of finitely many functions in this subalgebra A is also in the closure of the subalgebra. This came from the idea that the Weierstrass approximation theorem works due to the fact that we can approximate continuous functions by piecewise linear functions, and these may be approximated by polynomials. Let us prove the maximum and minimum fact before proceeeding.

First, we reduce to noting that A is dense if and only if $\overline{A} = C(X, \mathbb{R})$, so we consider the closed, unital subalgebra $\overline{A}$. Let $f \in \overline{A}$ have $f \geq 0$. Since f is bounded, we may assume $0 \leq f \leq 1$. Let $f(t) = 1 - g(t)$, and recall that the Taylor series $\sqrt{1-t} = \sum_{n=0}^{\infty} c_n t^n$ converges uniformly on $[0, 1]$. Thus,

$$\sqrt{f(t)} = \sqrt{1 - g(t)} = \sum_{n=0}^{\infty} c_n (g(t))^n$$

As the finite approximations of the Taylor series are in $\overline{A}$ and uniformly approximate $\sqrt{f}$, then $\sqrt{f} \in \overline{A}$ as well. This is helpful because it allows us to take $|f| = \sqrt{f^2} \in \overline{A}$ for any $f \in \overline{A}$, in which case

$$\max\{f, g\} = \frac{1}{2}((f+g) + |f-g|) \quad \min\{f, g\} = \frac{1}{2}(f + g - |f-g|)$$

We now make a second reduction. We show that for any $f \in X$, $f \in \overline{A}$ if and only if for any $x, y \in X$ and for every $\epsilon > 0$ there exists $g \in \overline{A}$ so $|f(x) - g(x)| < \epsilon$ and $|f(y) - g(y)| < \epsilon$.

To see that this is gives the above statement, notice that given an algebra A containing a nonzero constant function that separates points, we may take any x and y, with p a function separating them. We then may generate $g_1 = p - p(x) * 1$ and $g_2 = p - p(y) * 1$ two functions in A. Hence, $f(x)g_1 + f(y)g_2$ satisfies the approximation property of the claim.

We now prove the claim. Again, we prove that the condition implies f is in the closure (because the other directions is trivial). Given x, y, ϵ, let f_{xy} be the function satisfying the condition. Let $G_y = \{z : f(z) - f_{xy}(z) < \epsilon\}$. Notice $x, y \in G_y$, and so a union of sets of the form G_y cover X. Since X is compact, there is a finite subcover G_{y_i} for $i = 1, 2, ..., n$. Set $g_x = \max\{f_{xy_1}, ..., f_{xy_n}\}$. The cover property shows that $g_x(z) > f(z) - \epsilon$ for all $z \in X$. We also know that $g_x(x) < f(x) + \epsilon$ since this condition holds for all of the f_{x,y_i}. Let $H_x = \{z : g(z) < f(z) + \epsilon\}$, which contains x. Again, we may get a cover of X by such sets, take a finite subcover, and now set $g = \min\{g_{x_1}, ..., g_{x_m}\}$. This implies $f(z) - \epsilon < g(z) < f(z) + \epsilon$ for all z. We have now approximated f to tolerance ϵ by some function in $\overline{A}$.

Q.E.D.

References

[1] Cesare Arzelá. Un'osservazione intorno alle serie di funzioni. *Rend. Dell' Accad. R. Delle Sci. dell'Istituto di Bologna*, pages 142–159, 1883.

[2] G.Ascoli. Le curve limite di una varietá data di curve. *R. Accademia dei Lincei*, 18.

[3] David R. Kincaid and E. Ward Cheney. *Numerical Analysis: Mathematics of Scientific Computing.* Brooks/Cole, Pacific Grove, Calif., 1991.

[4] W. Rudin. *Principles of Mathematical Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1953.

[5] M. H. Stone. The generalized weierstrass approximation theorem. *Mathematics Magazine*, 21(4):167–184, 1948.