MATHEMATICS 3103 (Functional Analysis)
YEAR 2009–2010, TERM 2

HANDOUT #5: SPACES OF CONTINUOUS FUNCTIONS

The most important spaces in applications of functional analysis, after the Hilbert spaces, are the spaces $\mathcal{C}(X)$ of bounded continuous functions on a metric space X (equipped with the sup norm). We have already proven the most important fact about $\mathcal{C}(X)$, namely that it is *complete* (hence a Banach space). Here we would like to develop some additional useful properties of the spaces $\mathcal{C}(X)$, especially in the case where X is *compact*.

Separation and extension theorems

The first question one should ask, when introducing a space $\mathcal{F}$ of functions on some set X, is whether there are "enough" functions in the space $\mathcal{F}$. What is "enough"? Well, there are various notions, but at a minimum one would like the functions in $\mathcal{F}$ to *separate points of* X: that is, for each pair $x, y \in X$ with $x \neq y$, there should exist $f \in \mathcal{F}$ such that $f(x) \neq f(y)$.[1] When $\mathcal{F}$ is the space $\mathcal{C}(X)$ of bounded continuous functions on a metric space X, we can prove more: not only do the bounded continuous functions separate *points* of X, they separate *disjoint closed sets* of X. (This of course includes the property of separating points, because every one-point set is closed.)

Theorem 5.1 (Urysohn's lemma for metric spaces) *Let A and B be disjoint nonempty closed sets in a metric space X. Then there exists a function $f \in \mathcal{C}(X)$ such that $f = 1$ on A, $f = 0$ on B, and $0 < f < 1$ on $X \setminus (A \cup B)$.*

PROOF. Let $f(x) = d(x, B)/[d(x, A) + d(x, B)]$. I leave it to you to prove that f is continuous and has the needed properties. □

Remark. If you have studied General Topology, you will know that there are a variety of *separation properties* that a topological space can possess (or not possess): Hausdorff, regular, completely regular, normal, ...[2] The foregoing theorem shows that metric spaces are as well behaved, in the sense of separation properties, as a topological space can possibly be: namely, they are *perfectly normal*, i.e. two disjoint closed sets can be *precisely separated* by a continuous function in the sense given above. A slightly larger class of topological spaces is given by the *normal* spaces: in these, two disjoint closed sets can be separated by a continuous function but not necessarily precisely, i.e. there exists a function $f \in \mathcal{C}(X)$ such that $f = 1$ on A, $f = 0$ on B, and $0 \le f \le 1$ (*non*-strict inequality!) on $X \setminus (A \cup B)$. □

Here is a related and extremely useful property:

---

[1] For instance, we posed a question of this type two weeks ago, for the case when $\mathcal{F}$ is the space X^* of continuous linear functionals on a normed linear space X. Next week we will answer this question affirmatively, using the Hahn–Banach theorem: the continuous linear functionals do separate points of X.

[2] See, for instance, `http://en.wikipedia.org/wiki/Separation_axiom`

Theorem 5.2 (Tietze extension theorem for metric spaces) *Let A be a closed subset of a metric space X, and let $f\colon A \to [a,b] \subset \mathbb{R}$ be a bounded continuous function. Then there exists a continuous function $g\colon X \to [a,b]$ that extends f (i.e. $g \restriction A = f$).*

Since we can always take $a = \inf_{x\in A} f(x)$ and $b = \sup_{x\in A} f(x)$, this result says that we can extend f continuously to all of X without increasing its upper bound or decreasing its lower bound.

PROOF. Without loss of generality we can assume that $[a,b] = [-1,1]$ (why?). Now define subsets $B, C \subseteq A$ by $B = \{x \in A\colon f(x) \le -\frac{1}{3}\}$ and $C = \{x \in A\colon f(x) \ge \frac{1}{3}\}$. By Urysohn's lemma there is a continuous function $f_1\colon X \to \mathbb{R}$ that takes the value $-\frac{1}{3}$ on B and $\frac{1}{3}$ on C and satisfies $|f_1(x)| \le \frac{1}{3}$ for all $x \in X$. By construction $|f(x) - f_1(x)| \le \frac{2}{3}$ for all $x \in A$ (why?).

Now apply the same construction to the function $f - f_1$, dividing the interval $[-\frac{2}{3}, \frac{2}{3}]$ into thirds: that is, let $B_2 = \{x \in A\colon f(x) - f_1(x) \le -\frac{2}{9}\}$ and $C_2 = \{x \in A\colon f(x) - f_1(x) \ge \frac{2}{9}\}$, and use Urysohn's lemma to obtain a continuous function $f_2\colon X \to \mathbb{R}$ that takes the value $-\frac{2}{9}$ on B_2 and $\frac{2}{9}$ on C_2 and satisfies $|f_2(x)| \le \frac{2}{9}$ for all $x \in X$; then $|f(x) - f_1(x) - f_2(x)| \le \frac{4}{9}$ for all $x \in A$ (why?).

Proceeding inductively, we construct a continuous function $f_n\colon X \to \mathbb{R}$ such that $|f_n(x)| \le 2^{n-1}/3^n$ for all $x \in X$ and $|f(x) - \sum_{i=1}^{n} f_i(x)| \le 2^n/3^n$ for all $x \in A$ (why?).

It follows from the bound $\|f_n\|_\infty \le 2^{n-1}/3^n$ that the series $\sum_{n=1}^{\infty} f_n$ is uniformly summable to a continuous (why?) function $g\colon X \to \mathbb{R}$ satisfying $\|g\|_\infty \le 1$ (why?). Moreover, g coincides with f on A (why?). □

Remarks. 1. Note that Urysohn's lemma (but with the *non*-strict inequality $0 \le f \le 1$ on $X \setminus (A \cup B)$) is a *special case* of the Tietze extension theorem (why?).

2. For a different proof of the Tietze extension theorem for metric spaces, see Dieudonné, Section IV.5: he gives an explicit formula, using the distance function d, for an extension function g. But the proof given here (which is the standard proof given in most books) has the advantage that it extends immediately to normal topological spaces, since all it uses is Urysohn's lemma (with non-strict inequality).

3. In Problem 1 of Problem Set #5, I will ask you to prove a slight strengthening of this result, in two directions: firstly, to allow unbounded continuous functions f; and secondly, to show that if f is bounded above (resp. below) but does not actually attain this upper (resp. lower) bound, then g can be chosen so that it does not attain this bound either.

Dini's theorem

The example of the functions $f_n(x) = x^n$ on the space $[0,1]$ shows that the limit of a *pointwise* convergent sequence of continuous functions need not be continuous — not even if the sequence is monotone (e.g. here it is decreasing: $f_1 \ge f_2 \ge f_3 \ge \ldots$) and the space is compact.

But if we *assume* that the limit is continuous, then under certain conditions we can guarantee that the convergence is uniform:

Theorem 5.3 (Dini's theorem) *Let X be a* compact *metric space. Let (f_n) be a* monotone *(i.e. increasing or decreasing) sequence of real-valued continuous functions that converges pointwise to a* continuous *function g. Then (f_n) converges uniformly to g, i.e. $\|f_n - g\|_\infty \to 0$.*

Proof. Suppose that (f_n) is decreasing, i.e. $f_1 \ge f_2 \ge f_3 \ge \ldots$ (the increasing case can obviously be handled by replacing f_n by $-f_n$ and g by $-g$). Choose $\epsilon > 0$, and let $U_n = \{x \in X\colon f_n(x) - g(x) < \epsilon\}$. Since f_n and g are continuous, U_n is open; and note that $U_1 \subseteq U_2 \subseteq U_3 \subseteq \ldots$ because the sequence (f_n) is decreasing. Furthermore, the sets U_n together *cover* X (why?). So, by compactness of X, there exists a finite subcovering $\{U_{n_1}, \ldots, U_{n_k}\}$. But, setting $N = \max(n_1, \ldots, n_k)$, this means that $U_N = X$ (why?). It follows that for all $n \ge N$ we have $g(x) \le f_n(x) \le f_N(x) < g(x) + \epsilon$ for all $x \in X$. Since this holds for arbitrary $\epsilon > 0$, we have proven that the sequence (f_n) converges uniformly to g. $\Box$

You should convince yourself by simple examples that all three hypotheses of Dini's theorem (compactness of X, monotonicity of convergence, continuity of the limit function) are needed: with any two of the three, a counterexample can be found.

In Problem 3 of Problem Set #5, I will ask you to prove a slight generalization: if (f_n) is a decreasing sequence of *upper semicontinuous* real-valued functions on a compact metric space X that converges pointwise to a *lower semicontinuous* function g, then the convergence is uniform.

The Stone–Weierstrass theorem

One of the most important results in real analysis is the **Weierstrass approximation theorem**[3], which states that any real-valued continuous function f on a *closed bounded* interval $[a, b]$ of the real line can be uniformly approximated by polynomials, i.e. for every $\epsilon > 0$ there exists a polynomial P such that $\|f - P\|_\infty < \epsilon$. Otherwise put, the polynomials are dense in the space $\mathcal{C}[a, b]$ equipped with the sup norm. Not surprisingly, a similar result holds for continuous functions on closed bounded subsets A of $\mathbb{R}^n$: they can be uniformly approximated by multivariate polynomials in the coordinates $x_1, \ldots, x_n$.

The **Stone–Weierstrass theorem**[4] is a vast extension of all these results. To begin with, the domain $[a, b] \subset \mathbb{R}$ or $A \subset \mathbb{R}^n$ is replaced by an arbitrary compact metric space X. But what, in this generality, should replace the polynomials? The answer is obtained, not by trying to find any unique family of functions to replace the polynomials, but rather by asking what is the fundamental property possessed by the polynomials. The key property is that the polynomials form an **algebra**: that is, they are closed not only under the vector-space operations (pointwise addition, and multiplication by scalars) but also under *pointwise multiplication*, i.e. the product of two polynomials is again a polynomial. The Stone–Weierstrass theorem shows that if $\mathcal{A}$ is *any* algebra of continuous functions on a compact metric space X

[3] Proven by the German mathematician Karl Weierstrass (1815–1897) in 1885. Weierstrass was one of the key founders of modern analysis.

[4] Proven by the American mathematician Marshall Stone (1903–1989) in 1937.

that contains the constant functions and separates points of X, then $\mathcal{A}$ is dense in the space $\mathcal{C}(X)$ equipped with the sup norm.

Theorem 5.4 (Stone–Weierstrass theorem) *Let X be a compact metric space, and let $\mathcal{A}$ be any algebra of real-valued continuous functions on X that contains the constant functions and separates points of X. Then $\mathcal{A}$ is dense in $\mathcal{C}(X)$.*

Before beginning the proof, it is useful to make some remarks about algebras of functions. First of all, for any metric space X (not necessarily compact), the space $\mathcal{C}(X)$ of bounded continuous functions on X is an algebra, since the pointwise product of two bounded continuous functions is bounded and continuous. Moreover, we have $\|fg\|_\infty \le \|f\|_\infty \|g\|_\infty$ (why?).[5] It easily follows from this that the bilinear mapping $(f,g) \mapsto fg$ is continuous (why?). And it easily follows from this that if $\mathcal{A}$ is any subalgebra of $\mathcal{C}(X)$ [i.e. any vector subspace of $\mathcal{C}(X)$ that is also closed under pointwise multiplication], then its closure $\overline{\mathcal{A}}$ (in the sup norm) is again a subalgebra.

In the course of the proof of the Stone–Weierstrass theorem, we will actually prove also a variant result that is interesting in its own right: it comes from the fact that the space $\mathcal{C}(X)$ is not only an algebra but is also a *lattice*,[6] namely if f and g are bounded continuous functions then so are the functions $f \vee g$ and $f \wedge g$, where

$$(f \vee g)(x) = \max[f(x), g(x)] \tag{5.1a}$$

$$(f \wedge g)(x) = \min[f(x), g(x)] \tag{5.1b}$$

More generally, a subset $\mathcal{L} \subseteq \mathcal{C}(X)$ is called a **lattice** (of bounded continuous functions on X) if $f, g \in \mathcal{L}$ implies $f \vee g \in \mathcal{L}$ and $f \wedge g \in \mathcal{L}$. A subset $\mathcal{L} \subseteq \mathcal{C}(X)$ is called a **vector lattice** (of bounded continuous functions on X) if it is both a linear subspace and a lattice (i.e. is closed under multiplication by scalars, addition, $\vee$ and $\wedge$).

Theorem 5.5 (Stone–Weierstrass theorem, lattice version) *Let X be a compact metric space, and let $\mathcal{L}$ be any vector lattice of continuous functions on X that contains the constant functions and separates points of X. Then $\mathcal{L}$ is dense in $\mathcal{C}(X)$.*

We will actually prove a bit more than this: namely, we will not require that $f, g \in \mathcal{L}$ imply $f + g \in \mathcal{L}$, but will only require this when g is a constant function; and we will prove a slightly strengthened version of the uniform approximation property, which shows that the approximation can always be from above.

---

[5] A Banach space equipped with a bilinear multiplication satisfying $\|xy\| \le \|x\|\,\|y\|$ is called a **Banach algebra**. The space $\mathcal{C}(X)$ under pointwise multiplication is thus a fundamental example of a *commutative* Banach algebra. On the other hand, the space $\mathcal{B}(X)$ of bounded linear operators on a Banach space X is a fundamental example of a *noncommutative* (when $\dim X > 1$) Banach algebra.

[6] **Warning:** Here I am using the word "lattice" in the sense that this word is used in algebra and functional analysis, namely, as an ordered set in which every pair of elements has a least upper bound and a greatest lower bound. This has nothing whatsoever to do with the meaning of "lattice" in geometry and solid-state physics, namely as (for example) a discrete subset of $\mathbb{R}^n$ that is invariant under translation by some linearly independent vectors $a_1, \ldots, a_n$. It is unfortunate that English uses the same word for both concepts (a little like the different meanings of the word "field" in algebra and in vector calculus). Other languages avoid this confusion: e.g. French denotes the first concept by *treillis* and the second by *réseau*.

Theorem 5.6 (Stone–Weierstrass theorem, slightly strengthened lattice version) *Let X be a compact metric space, and let $\mathcal{L}$ be any nonempty lattice of continuous functions on X that separates points of X and is closed under multiplication by scalars and addition of constants (in short, $f \in \mathcal{L}$ and $\alpha, \beta \in \mathbb{R}$ imply $\alpha f + \beta \in \mathcal{L}$).*[7] *Then for any $f \in \mathcal{C}(X)$ and any $\epsilon > 0$, there exists $g \in \mathcal{L}$ such that $f \le g \le f + \epsilon$.*

We will prove this using a sequence of lemmas. The first lemma is a generalization of Dini's theorem:

Lemma 5.7 (Dini's theorem for lattices) *Let $\mathcal{L}$ be a lattice of continuous real-valued functions on a compact metric space X, and suppose that the function g defined by*

$$g(x) \;=\; \inf_{f \in \mathcal{L}} f(x) \tag{5.2}$$

is continuous. Then, for each $\epsilon > 0$, there exists $h \in \mathcal{L}$ such that $g \le h \le g + \epsilon$.

Do you see how this contains Dini's theorem as a special case? Just take $\mathcal{L} = \{f_1, f_2, \ldots\}$.

Proof of Lemma 5.7. The proof is a minor modification of the proof of Dini's theorem. Choose $\epsilon > 0$, and for each $f \in \mathcal{L}$ let $U_f = \{x \in X\colon f(x) - g(x) < \epsilon\}$. Since f and g are continuous, U_f is open. Furthermore, the sets U_f together *cover* X (why?). So, by compactness of X, there exists a finite subcovering $\{U_{f_1}, \ldots, U_{f_k}\}$. Then $h = f_1 \wedge f_2 \wedge \ldots \wedge f_k$ does what is needed (why?). □

Remark. Note that this proof did not really use that $\mathcal{L}$ is a lattice, but only that it is a **$\wedge$-semilattice**, i.e. $f, g \in \mathcal{L}$ implies $f \wedge g \in \mathcal{L}$.

The second lemma is almost trivial but is worth stating explicitly:

Lemma 5.8 *Let $\mathcal{F}$ be any family of real-valued functions on X that separates points of X and is closed under multiplication by scalars and addition of constants. Then for every pair of distinct points $x, y \in X$ and every pair of real numbers a, b, there exists $f \in \mathcal{F}$ such that $f(x) = a$ and $f(y) = b$.*

Proof. Since $\mathcal{F}$ separates points of X, we can find $g \in \mathcal{F}$ such that $g(x) \neq g(y)$. Then

$$f \;=\; \frac{a-b}{g(x)-g(y)}\, g \;+\; \frac{bg(x) - ag(y)}{g(x)-g(y)} \tag{5.3}$$

has all the needed properties (why?). □

And one last lemma:

[7] Note that this implies in particular that $\mathcal{L}$ contains the constant functions, by taking $\alpha = 0$ and observing that $\mathcal{L}$ is nonempty (why do I need the latter?).

By the way, why did I require *explicitly* that $\mathcal{L}$ be nonempty? If X has at least two points, then the hypothesis that $\mathcal{L}$ separates points of X implies that $\mathcal{L}$ has to be nonempty. But if X has only one point, then $\mathcal{L} = \varnothing$ satisfies all the hypotheses of the theorem but is not dense in $\mathcal{C}(X)$! Sometimes it is a pain to have to worry about all these trivial degenerate cases …

Lemma 5.9 *Let X be a compact metric space, and let $\mathcal{L}$ be a nonempty lattice of continuous functions on X that separates points of X and is closed under multiplication by scalars and addition of constants. Then for any closed subset $B \subset X$, any point $p \in X \setminus B$, and any real numbers a, b, there exists a function $g \in \mathcal{L}$ such that $g \geq a$, $g(p) = a$ and $g > b$ on B.*

PROOF. By Lemma 5.8 we can choose, for each $x \in B$, a function $g_x \in \mathcal{L}$ such that $g_x(p) = a$ and $g_x(x) = b+1$. Now let $U_x = \{y \in X\colon g_x(y) > b\}$. Then the sets $\{U_x\}_{x \in B}$ cover B (why?); and since B is compact (why?), we can choose a finite subcovering $\{U_{x_1}, \ldots, U_{x_n}\}$. Then $g = g_{x_1} \vee \ldots \vee g_{x_n} \vee a$ has all the needed properties (why?). $\Box$

Remark. It *appears* this proof did not really use that $\mathcal{L}$ is a lattice, but only that it is a $\vee$**-semilattice**, i.e. $f, g \in \mathcal{L}$ implies $f \vee g \in \mathcal{L}$. But since $\mathcal{L}$ is closed under multiplication by scalars (including the scalar -1) and $f \wedge g = -(-f \vee -g)$, it turns out that $\mathcal{L}$ has to be a lattice after all. And we *did* use the closure under multiplication by scalars (including negative scalars) in the proof of Lemma 5.8.

We are now ready to prove the lattice version of the Stone–Weierstrass theorem:

PROOF OF THE SLIGHTLY STRENGTHENED LATTICE VERSION OF THE STONE–WEIERSTRASS THEOREM. Given $f \in \mathcal{C}(X)$, let us define $\mathcal{L}_{\geq f} = \{g \in \mathcal{L}\colon g \geq f\}$. Note that $\mathcal{L}_{\geq f}$ is also a lattice (why?). I claim that

$$f(p) \;=\; \inf_{g \in \mathcal{L}_{\geq f}} g(p) \tag{5.4}$$

for all $p \in X$. If we can prove this, we are done, because the desired conclusion then follows immediately from Lemma 5.7.

To prove (5.4), fix any $p \in X$ and choose any $\delta > 0$. Since f is continuous, the set

$$B \;=\; \{x \in X\colon f(x) \geq f(p) + \delta\} \tag{5.5}$$

is closed. Since X is compact, f is bounded on X, say by M. By Lemma 5.9 with $a = f(p) + \delta$ and $b = M$, we can find $g \in \mathcal{L}$ such that $g \geq f(p) + \delta$, $g(p) = f(p) + \delta$ and $g > M$ on B. Then $g > M \geq f$ on B and $g \geq f(p) + \delta > f$ on $X \setminus B$ (why?), hence $g > f$ everywhere. It follows that $g \in \mathcal{L}_{\geq f}$. On the other hand, we have $g(p) \leq f(p) + \delta$ (actually equality, but we don't need this). Since $\delta > 0$ was arbitrary, we have proven (5.4). $\Box$

To prove the standard (i.e. algebra) version of the Stone–Weierstrass theorem, we need a lemma that allows us to approximate lattices by algebras, which comes down to approximating the absolute-value function by polynomials:

Lemma 5.10 *Given $\epsilon > 0$, there is a real polynomial P_ϵ in one variable such that*

$$\sup_{-1 \leq s \leq 1} \Big| |s| - P_\epsilon(s) \Big| \;<\; \epsilon\,. \tag{5.6}$$

This is of course a special case of the Weierstrass approximation theorem for the interval $[-1, 1]$, but let us give two direct elementary proofs:

First proof of Lemma 5.10. Let $\sum_{n=0}^{\infty} a_n t^n$ be the binomial series for $(1-t)^{1/2}$, namely $a_0 = 1$, $a_1 = -\frac{1}{2}$ and

$$a_n \;=\; -\,\frac{(2n-3)!!}{2^n n!} \tag{5.7}$$

[recall that $k!! = k(k-2)(k-4)\cdots 1$ for k odd]. It is a well-known fact that this series converges uniformly to $(1-t)^{1/2}$ for all $t \in [0,1]$ (and indeed for all complex t with $|t| \le 1$).[8] Hence, given $\epsilon > 0$, we can choose N so that

$$\sup_{0 \le t \le 1} |(1-t)^{1/2} - Q_N(t)| \;<\; \epsilon \tag{5.8}$$

where $Q_N(t) = \sum_{n=0}^{N} a_n t^n$. Then $P_\epsilon(s) = Q_N(1-s^2)$ has all the needed properties (why?). □

Second proof of Lemma 5.10. We shall construct a sequence $(p_n)_{n=1}^{\infty}$ of real polynomials which in the interval $[-1,1]$ satisfy $0 = p_1 \le p_2 \le \ldots$ and which converge uniformly to $|s|$. We define the p_n inductively, taking $p_1 = 0$ and then

$$p_{n+1}(s) \;=\; p_n(s) \,+\, \tfrac{1}{2}[s^2 - p_n(s)^2]\,. \tag{5.9}$$

Let us prove by induction that $p_{n+1}(s) \ge p_n(s)$ and $p_n(s) \le |s|$ and for $s \in [-1,1]$. From the defining recursion (5.9) we see that the second result (for any given n) implies the first (for the same n). On the other hand,

$$\begin{aligned} |s| - p_{n+1}(s) &\;=\; |s| - p_n(s) \,-\, \tfrac{1}{2}[s^2 - p_n(s)^2] \\ &\;=\; [|s| - p_n(s)]\left(1 \,-\, \tfrac{1}{2}[|s| + p_n(s)]\right) \end{aligned} \tag{5.10}$$

and from $p_n(s) \le |s|$ we deduce $\frac{1}{2}[|s| + p_n(s)] \le |s| \le 1$, so $p_n(s) \le |s|$ implies $p_{n+1}(s) \le |s|$. Thus, for each $s \in [-1,1]$ the sequence $(p_n(s))$ is increasing and bounded above (by $|s|$), hence converges to a limit $P(s)$. But passing to the limit $n \to \infty$ in the recursion (5.9) we see that $s^2 - P(s)^2 = 0$; and since $P(s) \ge 0$ this means that $P(s) = |s|$. Since the absolute-value function is continuous and the sequence (p_n) is increasing, Dini's theorem shows that the

[8] For a proof, see e.g. `http://en.wikipedia.org/wiki/Binomial_series` For what it's worth, the uniform convergence for $t \in [0,1]$ follows from the pointwise convergence together with Dini's theorem, since $a_n < 0$ for all $n \ge 1$.

convergence is uniform.[9] $\square$

Corollary 5.11 *Let X be any metric space (not necessarily compact) and let $\mathcal{A}$ be any* closed *subalgebra of $\mathcal{C}(X)$ that contains the constant functions. Then $\mathcal{A}$ is a vector lattice.*

PROOF. First consider any $f \in \mathcal{A}$ with $\|f\|_\infty \le 1$. Then, given $\epsilon > 0$ we have $\| \, |f| - P_\epsilon(f) \, \|_\infty \le \epsilon$ where P_ϵ is the polynomial given by Lemma 5.10 (why? why do we have only non-strict inequality?). Since $\mathcal{A}$ is an algebra containing the constant functions, we have $P_\epsilon(f) \in \mathcal{A}$ (why? where did we use the fact that $\mathcal{A}$ contains the constant functions?). Since this holds for every $\epsilon > 0$, and $\mathcal{A}$ is closed, we conclude that $|f| \in \mathcal{A}$.

Now consider any nonzero $f \in \mathcal{A}$. Then $f/\|f\|_\infty$ belongs to $\mathcal{A}$ (why?) and has sup norm 1, so $|f|/\|f\|_\infty$ belongs to $\mathcal{A}$ by what has just been said, hence $|f| \in \mathcal{A}$ as well (why?). Thus $\mathcal{A}$ contains the absolute value of each function that is in $\mathcal{A}$. But since

$$f \vee g \;=\; \tfrac{1}{2}(f+g) \,+\, \tfrac{1}{2}|f-g| \tag{5.11a}$$

$$f \wedge g \;=\; \tfrac{1}{2}(f+g) \,-\, \tfrac{1}{2}|f-g| \tag{5.11b}$$

we conclude that $\mathcal{A}$ is a lattice, hence a vector lattice. $\square$

---

[9] The uniform convergence can also be proven by an explicit bound. Since $p_n(s) \ge 0$, we have from (5.10)

$$|s| - p_{n+1}(s) \;\le\; (1 - \tfrac{1}{2}|s|)\,[|s| - p_n(s)]$$

and hence

$$|s| - p_n(s) \;\le\; (1 - \tfrac{1}{2}|s|)^n\,[|s| - p_0(s)] \;=\; |s|\,(1 - \tfrac{1}{2}|s|)^n \;.$$

But simple calculus shows that for $n \ge 1$ we have

$$\sup_{-1 \le s \le 1} |s|\,(1 - \tfrac{1}{2}|s|)^n \;=\; \frac{2\,n^n}{(n+1)^{n+1}} \;\le\; \frac{2}{n+1}$$

[achieved at $|s| = 2/(n+1)$], which tends to zero as $n \to \infty$.

Note, by the way, that this bound gives (up to a constant factor) the correct rate of convergence, because we also have from (5.10) [using $p_n(s) \le |s|$]

$$|s| - p_n(s) \;\ge\; |s|\,(1 - |s|)^n$$

and

$$\sup_{-1 \le s \le 1} |s|\,(1 - |s|)^n \;=\; \frac{n^n}{(n+1)^{n+1}} \;\ge\; \frac{1}{e(n+1)}$$

(why?) [achieved at $|s| = 1/(n+1)$].

It is relevant to note that the slowest convergence of $p_n(s)$ to $|s|$ is obtained near the point $s = 0$ where the function $|s|$ is nonanalytic. This is not an accident. Indeed, the Russian mathematician Sergei Natanovich Bernstein (1880–1968) proved in 1912 that a function f on $[-1, 1]$ can be approximated by a polynomial of degree n with an error that decreases *exponentially* in n *if and only if* f can be extended to an *analytic* function in some complex neighborhood of $[-1, 1]$. Bernstein furthermore proved that it is not possible to approximate $|s|$ in $[-1, 1]$ by a polynomial of degree n with an approximation better than order $1/n$. These results belong to the area of analysis known as *approximation theory*. A nice introduction to approximation theory can be found in Allan Pinkus, Negative theorems in approximation theory, *Amer. Math. Monthly* **110**, 900–911 (2003).

The proof of the Stone–Weierstrass theorem is now a virtual triviality:

Proof of the Stone–Weierstrass theorem. As previously observed, the closure $\overline{\mathcal{A}}$ (in the sup norm) is again a subalgebra; and by Corollary 5.11 it is a lattice. It then follows from the lattice version of the Stone–Weierstrass theorem that $\overline{\mathcal{A}}$ is dense in $\mathcal{C}(X)$; but since $\overline{\mathcal{A}}$ is closed, this means that $\overline{\mathcal{A}} = \mathcal{C}(X)$. $\square$

Thus far we have been considering the space $\mathcal{C}(X)$ of *real-valued* continuous functions on X. What about the space $\mathcal{C}_{\mathbb{C}}(X)$ of *complex-valued* continuous functions on X? Here it is *false* that an algebra that separates points and contains the constant functions must be dense in $\mathcal{C}_{\mathbb{C}}(X)$: for instance, if X is the closed unit disc in the complex plane and $\mathcal{A}$ is the algebra of polynomials in the complex variable z, then all the elements of $\overline{\mathcal{A}}$ will be *analytic* functions when restricted to the open unit disc; so $\overline{\mathcal{A}}$ will certainly not contain, for instance, the function $f(z) = \overline{z}$. Instead we have the following weaker result:

Theorem 5.12 (Stone–Weierstrass theorem, complex version) *Let X be a compact metric space, and let $\mathcal{A}$ be any algebra of complex-valued continuous functions on X that contains the constant functions, separates points of X, and is invariant under complex conjugation (i.e. $f \in \mathcal{A}$ implies $\overline{f} \in \mathcal{A}$). Then $\mathcal{A}$ is dense in $\mathcal{C}_{\mathbb{C}}(X)$.*

Proof. The hypothesis implies that, for every $f \in \mathcal{A}$, the functions $\operatorname{Re} f = (f + \overline{f})/2$ and $\operatorname{Im} f = (f - \overline{f})/2i$ also belong to $\mathcal{A}$. Therefore, if $\mathcal{A}_0$ is the (real) subalgebra of $\mathcal{A}$ consisting of real-valued functions, we can conclude that $\mathcal{A}_0$ separates points of X (why?) and contains the (real) constant functions. So the Stone–Weierstrass theorem implies that $\mathcal{A}_0$ is dense in $\mathcal{C}_{\mathbb{R}}(X)$. But since $\mathcal{A} = \mathcal{A}_0 + i\mathcal{A}_0$ (why?), we conclude that $\mathcal{A}$ is dense in $\mathcal{C}_{\mathbb{C}}(X) = \mathcal{C}_{\mathbb{R}}(X) + i\mathcal{C}_{\mathbb{R}}(X)$. $\square$

Let us now give some classic applications of the Stone–Weierstrass theorem. The first one is the original Weierstrass approximation theorem for compact subsets of $\mathbb{R}^n$:

Theorem 5.13 (Weierstrass approximation theorem for compact subsets of $\mathbb{R}^n$) *Any real-valued continuous function on a closed bounded subset $X \subset \mathbb{R}^n$ is the limit of a sequence of polynomials that converges uniformly on X.*

Take now for X the unit circle in $\mathbb{R}^2$, parametrized by the angle θ, so that the continuous functions on X can be identified with the continuous functions on $\mathbb{R}$ that are periodic of period 2π, or equivalently with the continuous functions on $[-\pi, \pi]$ that satisfy $f(-\pi) = f(\pi)$. And take for $\mathcal{A}$ the (complex) algebra generated by the functions 1, $e^{i\theta}$ and $e^{-i\theta}$: that is, the elements of $\mathcal{A}$ are the *trigonometric polynomials* $\sum_{n=-N}^{N} a_n e^{in\theta}$ with coefficients $a_n \in \mathbb{C}$. Then the complex version of the Stone–Weierstrass theorem gives:

Theorem 5.14 (Weierstrass approximation theorem for trigonometric polynomials) *Any complex-valued continuous function on $\mathbb{R}$ that is periodic of period 2π is the limit of a sequence of trigonometric polynomials that converges uniformly on $\mathbb{R}$.*

Alternatively, we could have applied the real version of the Stone–Weierstrass theorem to the (real) algebra $\mathcal{A}_0$ generated by the functions 1, $\cos\theta$ and $\sin\theta$: standard trig identities show that the elements of $\mathcal{A}_0$ are precisely the *real trigonometric polynomials* $\sum_{n=0}^{N} a_n \cos(n\theta) + \sum_{n=1}^{N} b_n \sin(n\theta)$ with coefficients $a_n, b_n \in \mathbb{R}$.

Note also that the *only* continuous functions on $\mathbb{R}$ that can be uniformly approximated (or even pointwise approximated) by trigonometric polynomials are those that are periodic of period 2π, because the trigonometric polynomials are of course periodic and the periodicity survives uniform (or even pointwise) limits. On the other hand, in the L^2 norm the situation is very different:

Theorem 5.15 *In the space $\mathcal{C}[-\pi,\pi]$ equipped with the L^2 norm, the trigonometric polynomials are dense. (This is an incomplete inner-product space, but it follows immediately that the trigonometric polynomials are also dense in its completion, which is the space $L^2[-\pi,\pi]$.)*

Proof. The Weierstrass approximation theorem for trigonometric polynomials tells us that the trigonometric polynomials are dense, in the sup norm, in the linear subspace $\mathcal{C}_{\rm per}[-\pi,\pi] \subsetneq \mathcal{C}[-\pi,\pi]$ defined by

$$\mathcal{C}_{\rm per}[-\pi,\pi] \;=\; \{f \in \mathcal{C}[-\pi,\pi]\colon\, f(-\pi) = f(\pi)\}\;. \tag{5.12}$$

Since the L^2 norm is weaker than the sup norm, it follows that the trigonometric polynomials are also dense, in the L^2 norm, in $\mathcal{C}_{\rm per}[-\pi,\pi]$. So it suffices to show that $\mathcal{C}_{\rm per}[-\pi,\pi]$ is dense in $\mathcal{C}[-\pi,\pi]$ in the L^2 norm. To do this, consider any $f \in \mathcal{C}[-\pi,\pi]$, and define $g_n \in \mathcal{C}_{\rm per}[-\pi,\pi]$ by redefining f in a small neighborhood $[\pi - 1/n, \pi]$ so as to make it periodic (and still continuous):

$$g_n(x) \;=\; \begin{cases} f(x) & \text{for } -\pi \le x \le \pi - 1/n \\ f(\pi-1/n) + n[f(-\pi) - f(\pi-1/n)](x-\pi+1/n) & \text{for } \pi - 1/n \le x \le \pi \end{cases}$$

Then g_n coincides with f on $[-\pi, \pi-1/n]$ and differs from f by at most $2\|f\|_\infty$ on $[\pi-1/n,\pi]$, so we have

$$\|f-g_n\|_2^2 \;=\; \int_{-\pi}^{\pi} |f(x)-g_n(x)|^2\,dx \;\le\; (1/n)\,4\|f\|_\infty^2\,,$$

which tends to zero as $n \to \infty$. Thus $g_n \to f$ in L^2 norm, which proves that $\mathcal{C}_{\rm per}[-\pi,\pi]$ is dense in $\mathcal{C}[-\pi,\pi]$ in the L^2 norm. $\square$

In the first week of this course we posed (but did not solve) the question of the separability of the spaces $\mathcal{C}(X)$ [equipped as usual with the sup norm]. We can now give the answer. When X is a closed bounded interval $[a,b]$ of the real line, or more generally a closed bounded subset of $\mathbb{R}^n$, then the separability of $\mathcal{C}(X)$ is a consequence of the Weierstrass approximation theorem: indeed, the polynomials with rational coefficients form a countable (why?) dense (why?) set in $\mathcal{C}(X)$. For a general compact metric space X, we can use the Stone–Weierstrass theorem to prove the separability of $\mathcal{C}(X)$:

Theorem 5.16 *If X is a compact metric space, the space $\mathcal{C}(X)$ is separable.*

Proof. Recall from Problem 8 of Problem Set #1 that a collection $\mathcal{G}$ of nonempty open sets of X is called a **basis** for the open sets of X if every nonempty open set of X is the union of some subcollection of the collection $\mathcal{G}$; and recall the theorem that a metric space X is separable if and only if there exists a countable basis for the open sets of X. Since X is compact, it is separable, so let (U_n) be a countable basis for the open sets of X, and define $f_n(x) = d(x, X \setminus U_n)$. The functions f_n are continuous (why?) and bounded (why?). The monomials $f_1^{\alpha_1} \cdots f_n^{\alpha_n}$ with $n \in \mathbb{N}$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{N}$ form a countable family $\mathcal{F}$ (why?), and the vector space $\mathcal{A}$ generated by the family $\mathcal{F}$ is the subalgebra of $\mathcal{C}(X)$ generated by the (f_n). If we prove that $\mathcal{A}$ is dense in $\mathcal{C}(X)$, we will be done, because this implies that $\mathcal{F}$ is total in $\mathcal{C}(X)$, and hence by Problem 8(a) of Problem Set #3 that $\mathcal{C}(X)$ is separable. Now $\mathcal{A}$ obviously contains the constant functions, so we need only check that $\mathcal{A}$ separates points of X. But if $x \neq y$, then there is a U_n such that $x \in U_n$ and $y \notin U_n$ (why?), hence $g_n(x) \neq 0$ (why?) and $g_n(y) = 0$. $\square$

What if X is noncompact? In Proposition 1.18 we proved that $\ell^\infty = \mathcal{C}(\mathbb{N})$ is nonseparable; and in Problem 5 of Problem Set #1 you used a similar technique to show that $\mathcal{C}(\mathbb{R})$ is nonseparable. In Problem 2(b) of Problem Set #5 you are asked to extend this method to show that $\mathcal{C}(X)$ is nonseparable *whenever* X is noncompact.

Compactness in $\mathcal{C}(X)$ and the Arzelà–Ascoli theorem

In Handout #2 and Problems 3 and 4 of Problem Set #2, we discussed some criteria for compactness in the sequence spaces ℓ^p and c_0: more precisely, we found sufficient conditions for compactness in ℓ^1 and ℓ^2 (the obvious analogues would work in ℓ^p for any $p < \infty$) and a necessary and sufficient condition for compactness in c_0. Now we would like to discuss criteria for compactness in the space $\mathcal{C}(X)$ [equipped as usual with the sup norm] when X is a *compact* metric space.

We begin by discussing the key new notion, which is the **equicontinuity** of a family of functions. First recall that if (X, d_X) and (Y, d_Y) are two metric spaces, then a mapping $f \colon X \to Y$ is called

- **continuous at x_0** if for each $\epsilon > 0$, there exists $\delta > 0$ (depending of course on ϵ) such that, for all $x' \in X$, $d_X(x_0, x') < \delta$ implies $d_Y(f(x_0), f(x')) < \epsilon$;
- **continuous** if for each $\epsilon > 0$ and each $x \in X$, there exists $\delta > 0$ (depending on ϵ and x) such that, for all $x' \in X$, $d_X(x, x') < \delta$ implies $d_Y(f(x), f(x')) < \epsilon$;
- **uniformly continuous** if for each $\epsilon > 0$, there exists $\delta > 0$ (depending on ϵ) such that, for all $x, x' \in X$, $d_X(x, x') < \delta$ implies $d_Y(f(x), f(x')) < \epsilon$.

Now suppose that we have a *family* $\mathcal{F}$ of mappings from X to Y. If each function of the family is continuous at x_0 (or continuous, or uniformly continuous), then for each $f \in \mathcal{F}$ we can say that "for each $\epsilon > 0$ there exists $\delta > 0$ such that ..." but the δ might of course depend on the specific function $f \in \mathcal{F}$ being considered. The family is called *equicontinuous* if the δ can be chosen to be uniform for all $f \in \mathcal{F}$. More precisely, a family $\mathcal{F}$ of mappings from X to Y is called

- **equicontinuous at x_0** if for each $\epsilon > 0$, there exists $\delta > 0$ (depending on ϵ) such that, for all $x' \in X$ and all $f \in \mathcal{F}$, $d_X(x_0, x') < \delta$ implies $d_Y(f(x_0), f(x')) < \epsilon$;

- **equicontinuous** if for each $\epsilon > 0$ and each $x \in X$, there exists $\delta > 0$ (depending on ϵ and x) such that, for all $x' \in X$ and all $f \in \mathcal{F}$, $d_X(x, x') < \delta$ implies $d_Y(f(x), f(x')) < \epsilon$;

- **uniformly equicontinuous**[10] if for each $\epsilon > 0$, there exists $\delta > 0$ (depending on ϵ) such that, for all $x, x' \in X$ and all $f \in \mathcal{F}$, $d_X(x, x') < \delta$ implies $d_Y(f(x), f(x')) < \epsilon$.

Thus, equicontinuity at x_0 for a family $\mathcal{F}$ implies in particular that each of the functions in $\mathcal{F}$ is continuous at x_0 (and likewise for the other two notions), but it is much stronger.

Of course, if $\mathcal{F}$ is a *finite* set of functions, each of which is continuous at x_0, then $\mathcal{F}$ is equicontinuous at x_0 (and likewise for the other two notions). More generally, if $\mathcal{F}_1, \ldots, \mathcal{F}_k$ is a finite collection of families, each of which is equicontinuous at x_0, then their union $\bigcup\limits_{i=1}^{k} \mathcal{F}_i$ is also equicontinuous at x_0 (and likewise for the other two notions).

Example 1. The family of functions $(f_n)_{n=0}^{\infty}$ on $[0, 1]$ defined by $f_n(x) = x^n$ is *not* equicontinuous at $x_0 = 1$ (why?).

Example 2. Let X and Y be an arbitrary pair of metric spaces. Fix $\alpha > 0$ and $M < \infty$, and consider the family $\mathcal{F}_{\alpha,M}$ of functions $f\colon X \to Y$ defined by

$$\mathcal{F}_{\alpha,M} \;=\; \{f\colon d_Y(f(x), f(x')) \;\le\; M\, d_X(x, x')^{\alpha} \text{ for all } x, x' \in X\} \,. \tag{5.13}$$

Then $\mathcal{F}_{\alpha,M}$ is uniformly equicontinuous (why?).

These are the **Hölder-continuous functions of order** α, with Hölder seminorm at most M. If $\alpha = 1$ they are the **Lipschitz-continuous functions** with Lipschitz seminorm at most M. □

We can now state the fundamental criterion concerning compactness in $\mathcal{C}(X)$ where X is a *compact* metric space:

Theorem 5.17 (Arzelà–Ascoli theorem)[11] *Let X be a compact metric space. Then:*

(a) A subset $\mathcal{F} \subset \mathcal{C}(X)$ is compact if and only if it is closed, bounded and equicontinuous.

(b) A subset $\mathcal{F} \subset \mathcal{C}(X)$ is relatively compact (i.e. its closure is compact) if and only if it is bounded and equicontinuous.

---

[10] A better term would probably be **equi-(uniformly continuous)**, but this is unwieldy both to say and to write.

[11] The sufficient condition for compactness was proven in 1883 by the Italian mathematician Giulio Ascoli (1843–1896). The necessary condition was proven in 1895 by another Italian, Cesare Arzelà (1874–1912).

Note that here boundedness of a set $\mathcal{F} \subset \mathcal{C}(X)$ means boundedness in sup norm, i.e. there exists $M < \infty$ such that $\|f\|_\infty \le M$ for all $f \in \mathcal{F}$, or equivalently $|f(x)| \le M$ for all $f \in \mathcal{F}$ and all $x \in X$. This is sometimes called "uniform boundedness", where "uniform" means both "uniform in x" and "uniform in f".

As we shall see, the formulations (a) and (b) of the Arzelà–Ascoli theorem are almost trivially interderivable. Some books prefer to state one or the other; I prefer to state both, as both are useful.

We shall prove the Arzelà–Ascoli theorem by a sequence of lemmas.

First, if $\mathcal{F}$ is a family of mappings from metric space X to a metric space Y, let us denote by $\overline{\mathcal{F}}^{(\mathrm{ptwise})}$ the "closure" of $\mathcal{F}$ with respect to *pointwise* convergence, i.e. the set of all functions $f\colon X \to Y$ for which there exists a sequence $(f_n)_{n=1}^\infty$ in $\mathcal{F}$ such that $f(x) = \lim\limits_{n\to\infty} f_n(x)$ for all $x \in X$.[12] We then have:

Lemma 5.18 *Let X and Y be metric spaces, and let $\mathcal{F}$ be a family of mappings from X to Y that is equicontinuous at x_0 (or equicontinuous, or uniformly equicontinuous). Then $\overline{\mathcal{F}}^{(\mathrm{ptwise})}$ is likewise equicontinuous at x_0 (or equicontinuous, or uniformly equicontinuous). In particular, each $f \in \overline{\mathcal{F}}^{(\mathrm{ptwise})}$ is continuous at x_0 (or continuous, or uniformly continuous).*

Proof. Let us prove the version for equicontinuity at x_0 (the others are similar). Let $(f_n)_{n=1}^\infty$ be a sequence in $\mathcal{F}$ converging pointwise to a function $f \in \overline{\mathcal{F}}^{(\mathrm{ptwise})}$. By hypothesis, for each $\epsilon > 0$ there exists $\delta > 0$ (depending on ϵ) such that, for all $x' \in X$ and all n, $d_X(x_0, x') < \delta$ implies $d_Y(f_n(x_0), f_n(x')) < \epsilon$. But now we can take the limit $n \to \infty$ in this latter inequality to conclude that $d_Y(f(x_0), f(x')) \le \epsilon$. This implies that every $f \in \overline{\mathcal{F}}^{(\mathrm{ptwise})}$ satisfies the continuity condition with the *same* choice of $\delta = \delta(\epsilon)$ that works for $\mathcal{F}$ [except that we now have non-strict instead of strict inequality, but that is no harm]. This proves that the family $\overline{\mathcal{F}}^{(\mathrm{ptwise})}$ is equicontinuous at x_0. □

From this lemma we can already see that forms (a) and (b) of the Arzelà–Ascoli theorem are trivially interderivable. Indeed, if a subset $\mathcal{F}$ is bounded (in sup norm), then so is its (sup-norm) closure $\overline{\mathcal{F}}$; and if a subset $\mathcal{F}$ is equicontinuous, then Lemma 5.18 shows that its sup-norm closure $\overline{\mathcal{F}}$ is equicontinuous as well (since even the potentially larger set $\overline{\mathcal{F}}^{(\mathrm{ptwise})}$ is equicontinuous). You should now complete the proof that (a) implies (b), and vice versa.

Lemma 5.19 *Let X and Y be metric spaces, with Y complete, and let (f_n) be an equicontinuous sequence of mappings from X to Y that is pointwise convergent on a* dense *subset $D \subset X$. Then (f_n) is pointwise convergent on all of X, and the limit function is continuous.*

Proof. Consider a point $x \in X$. Then, given $\epsilon > 0$ there exists $\delta > 0$ (depending on x and ϵ) such that

$$d_Y(f_n(x), f_n(x')) < \epsilon \quad \text{for all } n \text{ whenever } d_X(x, x') < \delta\,. \tag{5.14}$$

[12] This is not a standard notation, but I think it is useful here.

Since D is dense in X, there exists $x' \in D$ such that $d_X(x, x') < \delta$. Now, by hypothesis the sequence $(f_n(x'))$ is convergent in Y, hence a Cauchy sequence, i.e. there exists an integer N such that

$$d_Y(f_m(x'), f_n(x')) < \epsilon \quad \text{whenever } m, n \geq N\ . \tag{5.15}$$

It follows by the triangle inequality that

$$\begin{aligned} d_Y(f_m(x), f_n(x)) &\leq d_Y(f_m(x), f_m(x')) + d_Y(f_m(x'), f_n(x')) + d_Y(f_n(x'), f_n(x)) \\ &< 3\epsilon \quad \text{whenever } m, n \geq N\ , \end{aligned} \tag{5.16}$$

which shows that $(f_n(x))$ is a Cauchy sequence in Y. And since Y is complete, this means that $(f_n(x))$ is convergent.

The continuity of the limit function was already proven in Lemma 5.18. □

That was a typical "3ϵ argument". Here is another:

Lemma 5.20 *Let X and Y be metric spaces, with X* compact*, and let (f_n) be an equicontinuous sequence of mappings from X to Y that is pointwise convergent to a function f. Then (f_n) is* uniformly *convergent to f.*

PROOF. Choose $\epsilon > 0$. By equicontinuity, for each $x \in X$ there exists an open set $U_x \ni x$ such that $d_Y(f_n(x), f_n(x')) < \epsilon$ for all n and all $x' \in U_x$. From this it also follows that $d_Y(f(x), f(x')) \leq \epsilon$ for all $x' \in U_x$ (why?).

By compactness of X, there is a finite subcollection $\{U_{x_1}, \ldots, U_{x_k}\}$ that covers X. Now choose N large enough so that $d_Y(f_n(x_i), f(x_i)) < \epsilon$ for all $n \geq N$ and $1 \leq i \leq k$ (why is this possible?). Then for every $x' \in X$ there exists an index i ($1 \leq i \leq k$) such that $x' \in U_{x_i}$, and we have

$$\begin{aligned} d_Y(f_n(x'), f(x')) &\leq d_Y(f_n(x'), f_n(x_i)) + d_Y(f_n(x_i), f(x_i)) + d_Y(f(x_i), f(x')) \\ &< 3\epsilon \quad \text{whenever } n \geq N\ . \end{aligned} \tag{5.17}$$

This shows that (f_n) converges uniformly to f. □

Lemma 5.21 (diagonal argument) *Let (f_n) be a sequence of mappings from a* countable *set S into a metric space Y, with the property that for each $x \in S$, every subsequence of $(f_n(x))$ has a convergent sub-subsequence. [This would hold, in particular, if the set $\{f_n(x)\}_{n \geq 1}$ is relatively compact in Y for each $x \in S$.] Then there exists a subsequence (f_{n_i}) that converges (pointwise) for all $x \in S$.*

PROOF. We discussed the "diagonal trick" in the solutions to Problem 2(b) of Problem Set #2, but let us review it. The set S is either finite or countably infinite, so we enumerate it as $S = \{x_1, x_2, \ldots\}$. We now extract from the sequence (f_n) a subsequence (f_{n_i}) [where $n_1 < n_2 < \ldots$] that converges at x_1. We then extract from the subsequence (f_{n_i}) a sub-subsequence that converges at x_2, and so forth. In general, we will introduce a two-dimensional array $(f_{i,j})$ of functions,

$$\begin{array}{ccccc} f_{0,1} & f_{0,2} & f_{0,3} & f_{0,4} & \cdots \\ f_{1,1} & f_{1,2} & f_{1,3} & f_{1,4} & \cdots \\ f_{2,1} & f_{2,2} & f_{2,3} & f_{2,4} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

where the zeroth row is the original sequence (i.e. $f_{0,j} = f_j$), the first row is the first chosen subsequence (i.e. $f_{1,j} = f_{n_j}$), the second row is the chosen sub-subsequence, and so forth. In general the ith row is a subsequence of the $(i-1)$st row, chosen so that the sequence converges when evaluated at x_i. Since the ith row is in fact a subsequence of *all* the preceding rows (why?), we conclude that in the ith row the sequence converges when evaluated at $x_1, \ldots, x_i$. If S is finite (say, $S = \{x_1, x_2, \ldots, x_N\}$), then the last ($N$th) row of this array gives the desired subsequence. If S is countably infinite, then we build an infinite array, and the desired subsequence is given by the *diagonal* sequence $f_{j,j}$. (Why does this work?) □

You should make sure you understand the logic behind the diagonal argument, as it is frequently used in analysis.

We are now ready to prove the Arzelà–Ascoli theorem. We start with the "main" direction of the theorem, proving for simplicity version (b), i.e. boundedness plus equicontinuity implies relative compactness.

Proof of boundedness plus equicontinuity $\Longrightarrow$ relative compactness. Since X is compact, it is separable, so choose a countable dense set $D \subset X$. Now consider any sequence (f_n) in $\mathcal{F}$. Since $\mathcal{F}$ is bounded in sup norm, the set $\{f_n(x)\}$ is certainly bounded for each $x \in D$, hence relatively compact in $\mathbb{R}$. So by Lemma 5.21 there exists a subsequence (f_{n_i}) that converges (pointwise) for all $x \in D$. Then by Lemma 5.19 the sequence (f_{n_i}) is pointwise convergent on the whole space X, to a limit function f that is continuous. And by Lemma 5.20 the convergence of (f_{n_i}) to f is uniform, i.e. $f_{n_i} \to f$ in $\mathcal{C}(X)$. □

Remark. It is easy to see that the requirement of sup-norm boundedness of the family $\mathcal{F}$ can be weakened to pointwise boundedness (i.e. the set $\{f(x)\colon f \in \mathcal{F}\}$ is bounded in $\mathbb{R}$ for each $x \in X$) or even pointwise boundedness on a dense subset $D \subset X$. Indeed, that is all that this proof uses. (Of course, sup-norm boundedness then *follows* by invoking the converse direction of the Arzelà–Ascoli theorem.) □

Now we prove the "converse" direction of the Arzelà–Ascoli theorem, once again in version (b), i.e. relative compactness implies boundedness plus equicontinuity. Indeed, with no extra work we will show uniform equicontinuity.

Proof of relative compactness $\Longrightarrow$ boundedness plus uniform equicontinuity. Since $\overline{\mathcal{F}}$ is compact, it is certainly bounded, hence $\mathcal{F} \subseteq \overline{\mathcal{F}}$ is bounded as well. Moreover, $\overline{\mathcal{F}}$ is totally bounded, i.e. for each $\epsilon > 0$ we can choose a ϵ-net $\{f_1, \ldots, f_m\} \subseteq \overline{\mathcal{F}}$. Since each f_i is uniformly continuous (why?) and the family is finite, there exists $\delta > 0$ such that

$$|f_i(x) - f_i(x')| \;<\; \epsilon \quad \text{for all } i \text{ whenever } d_X(x, x') < \delta \tag{5.18}$$

(how does this use the finiteness of the family?). Then, by definition of ϵ-net, for any $f \in \overline{\mathcal{F}}$ there exists an index i such that $\|f - f_i\|_\infty < \epsilon$. We then have

$$\begin{aligned} |f(x) - f(x')| &\le |f(x) - f_i(x)| + |f_i(x) - f_i(x')| + |f_i(x') - f(x')| \\ &< 3\epsilon \quad \text{whenever } d_X(x, x') < \delta \, . \end{aligned} \tag{5.19}$$

Since this holds for all $f \in \overline{\mathcal{F}}$, we have proven that the family $\overline{\mathcal{F}}$ is uniformly equicontinuous. Hence so is $\mathcal{F}$. $\square$

Remark. In Problem 6 of Problem Set #5 you will prove that equicontinuity and uniform equicontinuity are in fact *equivalent* properties whenever X is compact.

Example 2, revisited. Let X be a *compact* metric space and let $Y = \mathbb{R}$. Fix $\alpha > 0$ and $M < \infty$. Then the family

$$\mathcal{F}_{\alpha,M} = \{f \colon |f(x) - f(x')| \le M d_X(x, x')^\alpha \text{ for all } x, x' \in X\} \subseteq \mathcal{C}(X) \tag{5.20}$$

is uniformly equicontinuous. This family is not bounded (why?), but the subfamily

$$\mathcal{G}_{\alpha,M} = \{f \colon |f(x)| \le M \text{ and } |f(x) - f(x')| \le M \, d_X(x, x')^\alpha \text{ for all } x, x' \in X\} \tag{5.21}$$

is bounded (and of course still uniformly equicontinuous). Moreover, the family $\mathcal{G}_{\alpha,M}$ is closed as well (why?). It follows that $\mathcal{G}_{\alpha,M}$ is a compact subset of $\mathcal{C}(X)$.

In particular, suppose that X is a closed interval $[a, b]$ of the real line, and consider the family of functions that are continuously differentiable on $[a, b]$ with both f and f' bounded in absolute value by M:

$$\mathcal{D}_M = \{f \in \mathcal{C}^1[a, b] \colon |f(x)| \le M \text{ and } |f'(x)| \le M \text{ for all } x \in X\} \, . \tag{5.22}$$

Then we have $\mathcal{D}_M \subset \mathcal{G}_{1,M}$ (why?), so $\mathcal{D}_M$ is a relatively compact subset of $\mathcal{C}(X)$. [With a little extra work it can be shown that the closure of $\mathcal{D}_M$ is precisely $\mathcal{G}_{1,M}$.]

In Problem 8 of Problem Set #5 you will apply this type of argument to prove some very important theorems in *complex* analysis.