Notes on Topology for Functional Analysis

Eric A. Carlen[1]
Rutgers University

January 30, 2013

1 Introduction

What is it that one analyzes in functional analysis? Very often the analysis involves functions defined on a domain in an infinite dimensional vector space with values in the field, $\mathbb{R}$ or $\mathbb{C}$, over which the vectors space is defined. Since the elements of these infinite dimensional vector spaces are often functions themselves; e.g., the vectors space may consist of square integrable functions on $[0,1]$, functions on such vectors spaces are called *functionals*. More generally, a functional is simply a scalar valued function defined on a set of functions on which there is some linear structure.

Many problems in mathematics can be treated effectively by relating them to *variational problems*; i.e., problems of finding maximizers and minimizers of functions – or functionals. The Spectral Theorem is a simple but important example. Many other problems can be treated effectively by some sort of method of successive approximation. The development of both strategies requires a through exploration of the notions of *continuity* and *compactness*, and their interactions with *convexity*, *completeness*, *separability*, and *duality* that we shall develop in this course. We begin by recalling some basic theorems on continuity and compactness.

1.1 Approximation

A basic strategy in functional analysis is *approximation.* In particular, one often tries to approximate "general" elements of some infinite dimensional vector space by "nice" elements from some well understood vector space, possibly even finite dimensional.

For example, consider the vector space $\mathcal{C}([0,1])$ consisting of continuous real valued functions on $[0,1]$. The *Weirstrass Approximation Theorem* says that for any $f \in \mathcal{C}([0,1])$, and any $\epsilon > 0$, there is a polynomial p such that

$$\sup_{0 \le t \le 1} |f(t) - p(t)| \le \epsilon \ .$$

The quantity $\sup_{0\le t\le 1} |f(t) - p(t)|$ is a measure of the *distance* between f and p in $\mathcal{C}([0,1])$. That is, the function d_{uniform} on $\mathcal{C}([0,1]) \times \mathcal{C}([0,1])$ defined by

$$d_{\text{uniform}}(f,g) = \sup_{0 \le t \le 1} |f(t) - g(t)|$$

is a *metric* on $\mathcal{C}([0,1])$, and equipped with this metric, $\mathcal{C}([0,1])$ is a *metric space.*

[1] © 2012 by the author. This article may be reproduced, in its entirety, for non-commercial purposes.

According to the Wierstrass Approximation Theorem, given any $f \in \mathcal{C}([0,1])$, for each $n \in \mathbb{N}$, we can find a polynomial p_n such that $d_{\rm uniform}(f, p_n) \le 1/n$, so that

$$\lim_{n\to\infty} d_{\rm uniform}(f, p_n) = 0 \ ,$$

which we express by writing

$$f = \lim_{n\to\infty} p_n \ ,$$

or by saying that *the sequence* $\{p_n\}_{n\in\mathbb{N}}$ *converges (uniformly) to* f.

Many of the central questions in real analysis have to do with convergence, and especially with two closely related concepts, *continuity* and *compactness*.

A basic theorem from elementary topology says that if K is a closed and bounded subset of $\mathbb{R}^n$, and if f is a continuous real valued function on K, then there exist an $x \in K$ such that

$$f(x) \ge f(y) \qquad \text{for all } y \in K \ .$$

Such an x is called a *maximizer of* f.

In recalling this statement, we did not specify the topology with respect to which "closed" and "continuous" are defined, nor the metric with respect to which "bounded" is defined. There is only one topology on $\mathbb{R}^n$ with respect to which the algebraic operations on it are continuous, which is its usual metric topology, defined in terms of the Euclidean distance. In infinite dimensions, this not the case, and we shall need a variety of topologies for many problems, as we shall explain below. Moreover, not all of these topologies are *metrizable*. That is, we shall deal with topologies that cannot be induced by any metric.

Still, metric topologies are central to our subject, and are extremely convenient to work with, largely because questions of continuity and compactness can both be reduced to questions about sequences in the metric space setting.

1.2 Continuity in metric spaces

A function f from X to Y is continuous if a sufficiently small change in the input results in a small change in the output. This is most conveniently expresses in terms of the *open metric balls* in X:

1.1 DEFINITION (Open ball in metric spaces)**.** Let X be a metric space with metric d. Given a number $r > 0$, and a point $x \in X$, let $B_r(x)$ be defined by

$$B_r(x) = \{\ y \in X \ : \ d(y,x) < r\ \} \ .$$

This set is called the *open ball of radius* r *about* x.

Another piece of notation will be convenient: If X and Y are sets, and $f : X \to Y$, we write $f(A)$ to denote the image of $A \subset X$ under f, and we write $f^{-1}(B)$ to denote the pre-image of $B \subset Y$. We now define continuity in the metric space setting:

1.2 DEFINITION (Continuous functions from one metric space to another)**.** Let (X, d_X) and (Y, d_Y) be two metric spaces. Let f be a function from X to Y. Then f *is continuous at* $x_0 \in X$ in case for every $\epsilon > 0$, there is a $\delta_\epsilon > 0$ such that

$$f(B_{\delta_\epsilon}(x_0)) \subset B_\epsilon(f(x_0)) \ . \tag{1.1}$$

The function f is *continuous* in case it is continuous at each $x_0 \in X$.

1.3 DEFINITION (Convergent sequences in a metric space)**.** Let (X, d_X) be a metric spaces. Then a sequence $\{x_n\}$ in X converges to $x_0 \in X$ if and only if for every $\epsilon > 0$, $B_\epsilon(x_0)$ contains all but finitely many terms of the sequence $\{x_n\}$. We express this by writing $\lim_{n\to\infty} x_n = x_0$.

1.4 THEOREM (Continuity and sequences)**.** *Let (X, d_X) and (Y, d_Y) be two metric spaces. Let f be a function from X to Y. Then f is continuous at $x_0 \in X$ if and only if for every sequence $\{x_k\}$ in X*

$$\lim_{k\to\infty} x_k = x_0 \quad \Rightarrow \quad \lim_{k\to\infty} f(x_k) = f(x_0) \ . \tag{1.2}$$

Proof. Suppose that f is continuous, and that $\lim_{k\to\infty} x_k = x_0$. Pick any $\epsilon > 0$. Since f is continuous, there exists a $\delta > 0$ so that $f(B_\delta(x_0)) \subset B_\epsilon(f(x_0))$. Since $\lim_{k\to\infty} x_k = x_0$, all but finitely many terms of $\{x_k\}$ lie inside $B_\delta(x_0)$. Hence all but finitely many terms of $\{f(x_k)\}$ lie inside $f(B_\delta(x_0)) \subset B_\epsilon(f(x_0))$, and this proves that $\lim_{k\to\infty} f(x_k) = f(x_0)$.

Next suppose that f is not continuous at x_0. Then there exists some $\epsilon > 0$ so that for each positive integer k,

$$f(B_{1/k}(x_0)) \not\subset B_\epsilon(f(x_0)) \ .$$

Define a sequence $\{x_k\}$ by choosing

$$x_k \in B_{1/k}(x_0) \bigcap (f^{-1}(B_\epsilon(f(x_0))))^c \ \neq \emptyset \ .$$

Then $\{x_k\}$ converges to x_0, but $\{f(x_k)\}$ does nor converge to $f(x_0)$. □

There is another characterization of continuity involving the notion of *open sets*, which we now define:

1.5 DEFINITION (Open sets in metric spaces)**.** Let X be a metric space with metric d. A subset U of X is *open* in case *either*:

(1) It is the empty set $\emptyset$, *or else*

(2) For each $x \in U$, there is an $r > 0$, depending on x, such that

$$B_r(x) \subset U \ .$$

It is possible, and as we shall see, useful to characterize the (global) continuity of functions between two metric spaces simply in terms of open sets, without explicit reference to the specific metrics themselves.

1.6 THEOREM (Continuity and open sets)**.** *Let X and Y be metric spaces with metrics d_X and d_Y resepctively. Let f be a function from X to Y. Then f is continuous if and only if for every open set U in Y, $f^{-1}(U)$ is open in X.*

Proof. Suppose that f is continuous, and let U be an open set in Y. If $f^{-1}(U) = \emptyset$, then $f^{-1}(U)$ is open by *(1)*. Otherwise, if $f^{-1}(U) \neq \emptyset$, consider any $x \in f^{-1}(U)$. Then $f(x) \in U$, and since U is open, there exists a $\epsilon > 0$ such that $B_\epsilon(f(x)) \subset U$. Then, since f is continuous at x_0, there is a $\delta > 0$ so that $f(B_\delta(x)) \subset B_\epsilon(f(x)) \subset U$. This means

$$B_\delta(x) \subset f^{-1}(U) \ ,$$

and since x is an arbitrary element of $f^{-1}(U)$, U is open.

Conversely, suppose that f has the property that whenever U is open in Y, $f^{-1}(U)$ is open in X. Fix any $x \in X$ and any $\epsilon > 0$. $f^{-1}(B_\epsilon(f(x)))$ is open and contains x. Therefore, there is some $\delta > 0$ such that

$$B_\delta(x) \subset f^{-1}(B_\epsilon(f(x))) \ .$$

Since ϵ is arbitrary, f is continuous at x. Since x is arbitrary, f is continuous. □

1.3 Topological spaces

Since we can characterize continuous functions in terms of open sets, without explicitly mentioning a metric at all, we can "strip away" the metric structure, and deal directly with open sets. This will turn out to be useful.

1.7 DEFINITION (Topological Spaces)**.** Let X be any set, and let $\mathcal{O}$ be any collection of sets in X satisfying:

(1) The empty set $\emptyset$ belongs to $\mathcal{O}$, as does X itself.

(2) The union over any *arbitray* set of sets in $\mathcal{O}$ belongs to $\mathcal{O}$.

(3) The intersection over any *finite* set of sets in $\mathcal{O}$ belongs to $\mathcal{O}$.

In this case, $\mathcal{O}$ is said to be a *topology* on X, and the sets belonging to $\mathcal{O}$ are called *open* sets in X (for the topology in question). A subset A of X is *closed* in case its complement, A^c is open.

The pair $(X, \mathcal{O})$ is said to be a *topological space.*

Note that by De Morgans laws, the intersection of any *arbitrary* set of closed sets in X is itself closed.

The next definitions introduces some more useful terminology

1.8 DEFINITION (Interior, closure and neighborhoods)**.** Let $(X, \mathcal{O})$ be a topological space, and A a subset of X.The *interior* of A, A^o, is the union of all of the open sets contained in A. The *closure of* A, $\overline{A}$, is the intersection of all of the closed sets containing A. Finally for any $x \in X$, the set $\mathcal{N}_x$ of *neighborhoods of* x consists of all sets B such that $x \in B^o$.

1.9 DEFINITION (Hausdorff, normal)**.** A topological space $(X, \mathcal{O})$ is called *Hausdorff* if for any two distinct elements $x, y \in X$, there are disjoint open sets sets U and V with $x \in U$ and $y \in V$.

It is called *normal* if for each $x \in X$, the singleton $\{x\}$ is closed, and moreover, for any two disjoint closed sets A and B, there are disjoint open sets U and V with $A \subset U$ and $B \subset V$.

It is easy to see, and left as an exercise, that if X is any metric space, and $\mathcal{O}$ is the collection of all open sets in X, as defined above in terms of open balls, $\mathcal{O}$ does indeed constitute a topology on X. Thus by Theorem 1.6, the definition of continuity that we make next is consistent with our existing notion of continuity in the metric space setting.

1.10 DEFINITION (Continuous functions between topological spaces)**.** Let $(X, \mathcal{O}_X)$ and $(Y, \mathcal{O}_Y)$ be two topological spaces. A function f from X to Y is *continuous at* $x \in X$ in case for every neighborhood V of $f(x)$, there is a neighborhood U of x such that $f(U) \subset V$.

A function f from X to Y is *continuous* whenver U is open in Y, $f^{-1}(U)$ is open in X.

It is easy to see that $f : X \to Y$ is continuous if and only if it is continuous at each $x \in X$.

We are almost ready to get back to proving theorems. The next theorem we prove concerns one of our core issues: approximation. Given a topological space $(X, \mathcal{O}_X)$, and a subset A of X, what does it mean to "approximate" $x \in X \backslash A$ by elements of A? We shall take it to mean that *every* neighborhood U of x contains points in A:

1.11 DEFINITION (Limit points in a topological space)**.** Let $(X, \mathcal{O}_X)$ be a topological space. If A is any set in X, a point $x \in X$ is a *limit point* of A in case every for every open set U that contains x,

$$A \cap U \neq \emptyset \ .$$

We must be careful to distinguish this notion of limit point of a set from the limit of a sequence. In particular: Let $\{x_k\}$ be a sequence of elements of X. Recall that $\{x_k\}$ *is convergent to* x in case every neighborhood U of x also contains all but finitely many terms in the sequence $\{x_k\}$. On the other hand, x is a limit point of $\{x_k\}$ in case every neighborhood U of x contains *at least one* element of $\{x_k\}$.

In a metric space, there is of course a characterization of limit points in terms of sequences; x is a limit point of A if and only if there is a sequence $\{x_n\}_{n\in\mathbb{N}}$ of elements in A such that $\lim_{n\to\infty} x_{k_n} = x$.

We are now ready for the theorem that justifies the terminology "closed":

1.12 THEOREM (Closed sets and limit points)**.** *Let $(X, \mathcal{O})$ be any topological space. A subset A of X is closed if and only if A contains all of its limit points.*

Proof. Suppose that A is closed, and $x \in A^c$. Since A^c is open, there is an open set U containing x that has an empty intersection with A. Thus, x is not a limit point of A. Since x was an arbitrary point outside A, A must contain all of its limit points.

On the other hand, suppose that A contains all of its limit points. We must show that A is closed, or, what is the same thing, that A^c is open. Consider any point $x \in A^c$. Since it is not a limit point of x, there is an open set U_x containing x that has empty intersection with A. For each $x \in A^c$, chose such a U_x. But then, since for each $x \in A^c$, $x \in U_x \subset A^c$,

$$A^c = \bigcup_{x\in A^c} U_x \ .$$

Thus, by *(2)* in the definition of topological spaces, $A^c = \bigcup_{x\in A^c} U_x$ is open. □

We close this subsection with one more definition:

1.13 DEFINITION (Density)**.** Let $(X, \mathcal{O})$ be a toplogical space. Let $A \subset B \subset X$. Then A *is dense in* B in case the closure of A contains B.

By Theorem 1.12, A is dense in B if and only if every point in B is a limit point in A; i.e, if every point in B can be approximated arbitrarily well by points in A.

1.4 Compactness

1.14 DEFINITION (Compact Sets)**.** Let $(X, \mathcal{O}_X)$ be a topological space. A subset K is called *compact* in X in case for every collection $\mathcal{U}$ of open sets that covers K; i.e.,

$$K \subset \bigcap_{u \in \mathcal{U}} U \ ,$$

there is a finite subset $\mathcal{G}$ of $\mathcal{U}$ that also covers K:

$$K \subset \bigcap_{u \in \mathcal{G}} U \ .$$

$\mathcal{G}$ is called a *finite subcover* of A.

Our first example of a theorem involving compactness is the classical result known as *Dini's Theorem*. In proving this, we shall make use of the following fact: If $\mathcal{U}$ is any set of open subsets of X, then by De Morgan's laws,

$$\left(\bigcup_{U \in \mathcal{U}} U \right)^c = \bigcap_{U \in \mathcal{U}}^{n} U^c \ ,$$

Thus $\mathcal{U}$ is an open cover if and only if $\{ \ U^c \ : \ U \in \mathcal{U} \ \}$ is a set of closed subsets of X with empty intersection.

Therefore, X is compact if and only if whenever $\mathcal{K}$ is a set of closed subsets of X such that

$$\bigcap_{K \in \mathcal{K}} K = \emptyset \ ,$$

there is a finite subset $\{K_1, \ldots, K_n\} \subset \mathcal{K}$ such that

$$\bigcap_{j=1}^{n} K_j = \emptyset \ .$$

This analysis is often summarized by saying that X *is compact if and only if* X *has the "finite intersection property"*.

1.15 THEOREM (Dini's Theorem)**.** *Let X be a compact space, and let $\{f_n\}_{n \in \mathbb{N}}$ be a sequence of real valued functions on X, and suppose that there is a continuous real valued function f on X such that for each $x \in X$, the sequence $\{f_n(x)\}_{n \in \mathbb{N}}$ is monotone non-decreasing,*

$$\lim_{n \to \infty} f_n(x) = f(x) \ .$$

Then

$$\lim_{n \to \infty} f_n = f$$

uniformly.

In other words, pointwise convergence, together with compactness and montonicity, imply uniform convergence. Also note that replacing each f_n by $-f_n$, one converts a monotone non-decreasing sequence into a montone non-increasing sequence, and so the theorem remains true if one replaces "monotone non-decreasing" by "monotone non-increasing".

Proof. Fix $\epsilon > 0$. Define the sets K_ℓ, $\ell \in \mathbb{N}$, by

$$K_\ell := \{\ x \in X\ :\ f(x) - f_\ell(x) \geq \epsilon\ \} = \{\ x \in X\ :\ |f(x) - f_\ell(x)| \geq \epsilon\ \}\ .$$

Since f and f_ℓ are continuous, K_ℓ is closed. Also, since for each x, $\lim_{\ell\to\infty} f_\ell(x) = f(x)$,

$$\bigcap_{k=1}^{\infty} K_\ell = \emptyset.$$

Then, by the compactness of X, there is some $n \in \mathbb{N}$ such that

$$\bigcap_{\ell=1}^{n} K_\ell = \emptyset.$$

Hence, for all $\ell > n$, and all x, $|f_\ell(x) - f(x)| < \epsilon$. Since the sequence is monotone non-decreasing, it then follows that

$$m \geq \ell \qquad \Rightarrow \qquad |f_m(x) - f(x)| < \epsilon\ ,$$

and so $d(f_m, f) < \epsilon$ for all $m \geq \ell$. □

Next, we turn to one of the main theorems on compactness.

1.16 THEOREM (Compactness, Continuity, and Minima)**.** *Let $(X, \mathcal{O})$ be any topological space, and let K be a compact subset of X. Let f be a functions from X to $\mathbb{R}$ that is continuous when $\mathbb{R}$ is equipped with its usual metric topology. Then there exosts and $x \in K$ so that*

$$f(x) \geq f(y) \qquad \text{for all} \qquad y \in K\ . \tag{1.3}$$

Proof. Consider the open sets $(-n, \infty)$ in $\mathbb{R}$, since f is continuous,

$$\mathcal{U} = \{\ f^{-1}((-n, \infty))\ \ : n \in \mathbb{N}\ \}$$

is an open cover of X, and hence K. Since K is compact, there exists a finite subcover. But since the sets in our cover are *nested*; i.e., since for $n > m > 0$, $f^{-1}((-m, \infty)) \subset f^{-1}((-n, \infty))$, a single set in our cover suffices; i.e., there is an n with $K \subset f^{-1}((-n, \infty))$. In particular, f is bounded from below on K.

Now let a be the greatest lower bound of the numbers $f(y)$ for $y \in K$. We claim that there exists an $x \in K$ with $f(x) = a$. If so, then plainly (1.3) is true.

To prove this, let us suppose that there is no such x. Then

$$\mathcal{U} = \{\ f^{-1}((a + 1/n, \infty))\ \ : n \in \mathbb{N}\ \}$$

is an open cover of K. This means that there is a finite subcover, and again, since the sets in the open cover are nested, a single one of them, say $f^{-1}((a+1/n, \infty))$, covers K. But this would mean that $f(y) \geq a + 1/n$ for each y in k, which is not possible since a is the greatest lower bound. □

Any point x for which (1.3) is true is called a *maximizer of f on X*. Likewise, any point x for which

$$f(x) \le f(y) \qquad \text{for all} \qquad y \in K \ . \tag{1.4}$$

is called a *minimizer of f on X*.

There are several important things to notice from this proof. First, if f is continuous, so is $-f$, and a minimizer of $-f$ is a maximizer of f. Hence, the theorem implies the existence of both minimizers and maximizers for continuous functions on compact sets.

Now suppose we have a real valued function f defined on a set K, and we want to know if f has a minimizer in K. If we can find a toplogy on K that makes f continuous, and makes K compact, then we can apply the previous theorem.

However, the demands of continuity and compactness pull in opposite directions when we look for our topology: The topology has to have sufficiently many open sets in it for f to be continuous, since we need $f^{-1}(U)$ to be open for every open set U in $\mathbb{R}$. On the other hand, the more open sets we include in our topology, the more open covers we have to consider when showing that every open cover has a finite subcover.

Very often, one is stuck between a rock and a hard place, and there is *no* topology that both makes f continuous, and K compact. Indeed, there are many very nice functions f on $\mathbb{R}$ – such as the *identity function*, $f(x) = x$ – that simply do not have minimizers or maximizers. While $\mathbb{R}$ is compact under the trivial topology $\mathcal{O} = \{\emptyset, \mathbb{R}\}$, and while the identity function is continuous under the usual metric topology on $\mathbb{R}$, the fact that the identity function does not have either a maximizer or a minimizer shows that there is *no* topology $\mathcal{O}$ on $\mathbb{R}$ under which $\mathbb{R}$ is compact and the identity function is continuous from $(\mathbb{R}, \mathcal{O})$ to $\mathbb{R}$ equipped with its usual metric topology.

A situation that is frequently encountered in applications is that a function f on X does have, say, a minimizer, but not a maximizer. Also in this situation, it is impossible to find a topology for which f is continuous and X is compact, since them both minima and maxima would exist.

However, if we are only looking for minima, it is worth noticing that in our proof of Theorem 1.16, we did not use the full strength of the continuity hypothesis. The same proof yields the same conclusion if we assume only the property that $f^{-1}((t,\infty))$ is open for each t in $\mathbb{R}$.

1.17 DEFINITION (Upper and lower semicontinuous function)**.** Let $(X, \mathcal{O}_X)$ be a topological space. A function f from X to $\mathbb{R}$ is called *lower semicintinuous* in case for all t in $\mathbb{R}$, $f^{-1}((t,\infty))$ is open. It is called *upper semicintinuous* in case for all t in $\mathbb{R}$, $f^{-1}((-\infty,t))$ is open.

Thus, we can prove existence of minimizers for f on X by finding a topology that makes f lower semicontinuous, and K compact. This turns out to be a very useful strategy, as we shall see.

Still, to use either Theorem 1.15 or Theorem 1.16, we need criteria for compactness. How can we tell if a set X is compact? In metric spaces, we can reduce this to a question about sequences. We first make three definitions, two for metric spaces, and one for topological spaces.

1.18 DEFINITION (Total boundedness)**.** A set K in a metric space (X, d) is *totally bounded* in case for every $\epsilon > 0$, there exists a finite set $\{x_1, \ldots, x_n\} \subset K$ such that

$$K \subset \bigcup_{j=1}^{n} B_\epsilon(x_j) \ ;$$

i.e., for each $\epsilon > 0$, K has a finite cover by open balls of radius ϵ.

1.19 DEFINITION (Cauchy sequences and completeness)**.** Let (X,d) be a metric space. A sequence $\{x_k\}_{k\in\mathbb{N}}$ is a *Cauchy sequence* in case for each $\epsilon > 0$, there is some m so that $B_\epsilon(x_m)$ contains all but finitely many terms in the sequence.

A set $K \subset D$ is *complete* in case whenever $\{x_k\}_{k\in\mathbb{N}}$ is Cauchy, then $\{x_k\}_{k\in\mathbb{N}}$ converges to an element $x_0 \in K$; i.e., $\lim_{k\to\infty} x_k = x_0$.

1.20 DEFINITION (totoal boundedness)**.** A set K in a metric space (X,d) is *totally bounded* in case for every $\epsilon > 0$, there exists a finite set $\{x_1,\dots,x_n\} \subset K$ such that

$$K \subset \bigcup_{j=1}^n B_\epsilon(x_j) \ ;$$

i.e., for each $\epsilon > 0$, K has a finite cover by open balls of radius ϵ.

1.21 DEFINITION (Sequential compactness)**.** A set K in a topological space $(X,\mathcal{O})$ is *sequentially compact* in case every sequence $\{x_n\}_{n\in\mathbb{N}}$ in K has a convergent subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ that converges to an elements of K.

1.22 THEOREM (Compactness in a Metric space)**.** *Let (X,d) be any metric space, and let K be any subset of X. Then the following are equivalent:*

(1) K is sequentially compact.

(2) K is totally bounded and complete.

(3) K is compact.

Proof of Theorem 1.21. We first show that *(3)* implies *(2)*. Suppose K is compact. For each $\epsilon > 0$, $\{\ B_\epsilon(x)\ :\ x \in K\ \}$ is an open cover of K. Since K is compact, there exist a finite set $\{x_1,\dots,x_n\} \subset K$ such that $K \subset \bigcup_{j=1}^n B_\epsilon(x_j)$. Thus, K is totally bounded.

Next, let $\{x_k\}_{k\in\mathbb{N}}$ be a Cauchy sequence in K. Since $\{x_n\}$ is Cauchy, for each m we can fins a ball $B_{1/m}(z_m)$ that contains all but finitely many of the terms in the sequence. For each $m \in \mathbb{N}$, define

$$C_m := \left(\bigcup_{d(y,z_m)>2/m} B_{1/m}(y) \right)^c ,$$

and note that C_m is closed. If $z \in B_{1/m}(z_m)$, and $d(y,z_m) > 2/m$, then

$$d(z,y) \geq d(z_m,y) - d(z_m,z) \geq \frac{2}{m} - \frac{1}{m} = \frac{1}{m} \ .$$

Thus, $z \notin B_{1/m}(y)$. This is true for any y with $d(y,z_m) > 2/m$, $z \in C_m$; i.e, $B_{1/m}(z_m) \subset C_m$.

Next, since obviously $z \in B_{1/m}(z)$, if $d(z,z_m) > 2/m$, $z \in \bigcup_{d(y,z_m)>2/m} B_{1/m}(y)$. That is, $(B_{2/m}(z_m))^c \subset C_m^c$. Thus, $C_m \subset B_{2/m}(z_m)$. Altogether, we have:

$$B_{1/m}(z_m) \subset C_m \subset B_{2/m}(z_m) \ . \tag{1.5}$$

Then, for any finite n, $\displaystyle\bigcap_{m>1}^n C_m \neq \emptyset$ since each C_m contains all but finitely many terms in $\{x_n\}_{n\in\mathbb{N}}$.

Therefore, by compactness, and the finite intersection property in particular, $\displaystyle\bigcap_{m>1}^\infty C_m \neq \emptyset$. Chose

x_0 in this infinite intersection. By (1.5), $\lim_{n\to\infty} x_n = x_0$, and $x_0 \in K$. This shows that every Cauchy sequence in K converges to a point in K, and thus, K is complete.

We now show that *(2)* implies *(1)*. Let $\{x_n\}_{n\in\mathbb{N}}$ be any sequence in K. For each $m \in \mathbb{N}$, choose a finite $1/m$ cover of K by open ball of radius $1/m$. By the pigeonhole principle at least one of these contains an infinite subsequence. Using Cantor's diagonal construction as above, only more simply, we construct a subsequence $\{x_m^{(m)}\}_{m\in\mathbb{N}}$ such that for each $r > 0$, all but finitely many terms lie in a ball of radius r. Thus, $\{x_m^{(m)}\}_{m\in\mathbb{N}}$ is a Cauchy sequence, and since K is complete, it converges to some $x_0 \in K$. This shows that K is sequentially complete.

Finally, we show that *(1)* implies *(3)* This part of the proof is more complicated, and we proceed in four steps.

Step 1: K is bounded: We first show that K is bonded, which means that

$$\sup_{x,y\in K} d(x,y) < \infty \ .$$

This supremum is called the *diameter* of K.

To see that the diameter is finite, suppose that it is not. Under this hypothesis, we construct a sequence $\{x_n\}_{n\in\mathbb{N}}$ as follows. First, fix any $x \in X$. Now for each $n \in N$, choose some $x_n \in K\backslash B_n(x)$. The set $K\backslash B_n(x)$ is not empty when the diameter of K is infinite.

Then, by hypothesis, there is a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ and some $y \in K$ such that

$$\lim_{k\to\infty} x_{n_k} = y \ . \tag{1.6}$$

Then by the triangle inequality, we would have

$$d(x, x_{n_k}) \leq d(x,y) + d(y, x_{n_k}) \ .$$

But this cannot be: By construction, $d(x, x_{n_k}) > n_k$, while $d(x,y)$ is some fixed, finite number, and for all sufficiently large k, $d(y, x_{n_k}) \leq 1$, by (1.6). This contradiction shows that K must be bounded.

Step 2: K contains a dense sequence: We next show that there is a sequence $\{x_n\}_{n\in\mathbb{N}}$ that is *dense* in K; i.e., that for every $\epsilon > 0$, and every $x \in K$, there is some n such that $d(x_n, x) < \epsilon$.

In other words, the sequence $\{x_n\}_{n\in\mathbb{N}}$ passes arbitrarily close to every point in K. Here is how to construct it:

Pick the first term x_1 arbitrarily. We then define the rest of the sequence recursively as follows: Suppose that $\{x_1, \ldots, x_k\}$ have been chosen. For each $y \in K$, define

$$d_k(y) := \min_{1\leq j\leq k} \{d(y, x_j)\} \ .$$

This is, by definition, the distance from y to the set $\{x_1, \ldots, x_k\} \subset K$, and of course, this is no greater than the diameter of K, which is finite by the first step.

Therefore, d_k, defined by

$$d_k := \sup_{y\in K} d_k(y)$$

is no greater than the diameter of K.

Armed with this knowledge, we are ready to choose x_{k+1}: We choose x_{k+1} to be any element of K with

$$d_k(x_{k+1}) \geq \frac{1}{2} d_k \ .$$

We now claim that $\lim_{k\to\infty} d_k = 0$. It should be clear that $\{x_n\}_{n\in\mathbb{N}}$ is dense if and only if this is the case. So, to complete Step 2, we need to prove that $\lim_{k\to\infty} d_k = 0$.

Towards this end, the first thing to observe is that $\{d_k\}_{k\in\mathbb{N}}$ is a monotone decreasing sequence, bounded below by zero: Indeed, for any sets $A \subset B \subset K$, the distance from y to B is no greater than the distance from y to A. Therefore, we only have to show that *some subsequence* of $\{d_k\}_{k\in\mathbb{N}}$ converges to zero.

To do this, let $\{x_{k_n}\}_{n\in\mathbb{N}}$ be a convergent subsequence of $\{x_k\}_{k\in\mathbb{N}}$, and let y be the limit; i.e.,

$$\lim_{k\to\infty} x_{k_n} = y \ .$$

Then by the triangle inequality $d(x_{k_n}, x_{k_{n+1}}) \leq d(x_{k_n}, y) + d(y, x_{k_{n+1}})$, and since

$$\lim_{n\to\infty} d(x_{k_n}, y) = \lim_{n\to\infty} d(y, x_{k_{n+1}}) = 0 \ ,$$

we have

$$\lim_{k\to\infty} d(x_{k_n}, x_{k_{n+1}}) = 0 \ .$$

But since

$$x_{k_n} \in \{x_1, \dots, x_{k_{n+1}-1}\} \ ,$$

$$d(x_{k_n}, x_{k_{n+1}}) \geq d_{k_{n+1}-1}(x_{k_{n+1}}) \geq \frac{1}{2} d_{k_{n+1}-1} \ .$$

Therefore,

$$\lim_{n\to\infty} d_{k_{n+1}-1} = 0 \ ,$$

and then, since the entire sequence is monotone decreasing, $\lim_{k\to\infty} d_k = 0$. Hence, the sequence we have constructed is dense.

Step 3: Given any open cover of K, there exists a countable subcover. To prove this, consider any open cover $\mathcal{G}$ of K. Consider the set of open balls $B_r(x_k)$ where $r > 0$ is rational, and $\{x_k\}_{k\in\mathbb{N}}$ is the dense sequence that we have constructed in Step 2. This set of balls is countable since a countable union of countable sets is countable.

The countable subcover is constructed as follows: For each rational $r > 0$ and each $k \in \mathbb{N}$, choose $U_{r,k}$ to be *some* open set in $\mathcal{G}$ that contains $B_r(x_k)$ if there is such a set, and otherwise, do not define $U_{r,k}$. Let $\mathcal{U}$ be the set of open sets defined in this way; clearly $\mathcal{U}$ is countable by construction.

We now claim that $\mathcal{U}$ is an open cover of K. Clearly the sets in $\mathcal{U}$ are open. To see that they cover, pick any $x \in K$. Since $\mathcal{G}$ is an open cover of K, $x \in V$ for some $V \in \mathcal{G}$. Then, since V is open, for some rational $r > 0$, $B_{2r}(x) \subset V$.

Then, since $\{x_k\}_{k\in\mathbb{N}}$ is dense, there is some k with $x_k \in B_r(x)$. But then $x \in B_r(x_k)$ and

$$B_r(x_k) \subset B_{2r}(x) \subset V \ ,$$

(where the first containment holds by the triangle inequality). This shows that for the pair (r,k), there is *some* $V \in \mathcal{G}$ containing $B_r(x_k)$. Therefore, by construction, $U_{r,k} \in \mathcal{U}$ contains $B_r(x_k)$, and hence $x \in U_{r,k}$. Since x is an arbitrary element of K, $\mathcal{U}$ covers K.

Step 4: Some finite subcover of the countable cover is a cover. Order the sets in our countable cover $\mathcal{U}$ into a sequence of open sets $\{U_k\}_{k\in\mathbb{N}}$ that covers K.

Suppose that for each n, it is *not* the case that

$$K \subset \bigcup_{k=1}^{n} U_k \ . \tag{1.7}$$

Then we can construct a sequence $\{x_n\}_{n\in\mathbb{N}}$ by choosing

$$x_n \in K \backslash \left(\bigcup_{k=1}^{n} U_k \right) \ .$$

Let $\{x_{n_j}\}_{j\in\mathbb{N}}$ be a convergent subsequence with $\lim_{j\to\infty} x_{n_j} = y \in K$. Then, since $\mathcal{U}$ is an open cover of K, there is some U_k with $y \in U_k$. But then all but finitely many terms of the sequence $\{x_{n_j}\}_{j\in\mathbb{N}}$ lie in U_k, and so the whole sequence lies in some finite union of the sets in $\mathcal{U}$. This is a contradiction, and so (1.7) is true for some $n \in \mathbb{N}$. □

In the broader setting of topological spaces, there is no relation between compactness and sequential compactness. There are topological spaces that are compact, but not sequentially compact, and there are sequentially compact spaces that are not compact: This theorem *does not extend* to the general setting of topological spaces; it is important that the topology be a metric topology. Likewise, in the general topological setting, it is not true that a function f is continuous if and only if it takes convergent sequences to convergent sequences.

The notion of compactness as we have defined it in terms of open covers is a 20th century notion. In the 19th century, mathematicians thought about compactness issues in terms of sequential compactness.

1.5 The Arzelà-Ascoli Theorem

Let X be a compact topological space. Let $\mathcal{C}(X,\mathbb{R})$ denote the set of continuous real valued functions on X. If $f, g \in \mathcal{C}(X,\mathbb{R})$, then $x \mapsto |f(x) - g(x)|$ also belongs to $\mathcal{C}(X,\mathbb{R})$, Therefore, by Theorem 1.16, there exists an $x_0 \in X$ such that

$$|f(x_0) - g(x_0)| \geq |f(x) - g(x)| \ .$$

Define $d(f,g)$ to be this maximum value of $|f(x) - g(x)|$. It is obvious that $d(f,g) = d(g,f)$ and that $d(f,g) = 0$ if and only if $f = g$. Next, for f, g and h in $\mathcal{C}(X,\mathbb{R})$,

$$\begin{aligned} d(f,g) = |f(x_0) - g(x_0)| &\leq |f(x_0) - h(x_0)| + |h(x_0) - g(x_0)| \\ &\leq \sup_{x\in X} |f(x) - h(x)| + \sup_{x\in X} |h(x) - g(x)| \\ &= d(f,h) + d(h,g) \ . \end{aligned}$$

Thus, d is a metric on $\mathcal{C}(X,\mathbb{R})$, known as the *uniform metric*. For brevity of notation, we shall write $\mathcal{C}(X,\mathbb{R})$ to denote the metric space $(\mathcal{C}(X,\mathbb{R}), d)$.

1.23 THEOREM (Completeness of $\mathcal{C}(X,\mathbb{R})$)**.** *$\mathcal{C}(X,\mathbb{R})$ is a complete metric space.*

Proof. Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $(\mathcal{C}(X,\mathbb{R}),d)$. For each $x \in X$,

$$|f_n(x) - f_m(x)| \le d(f_n, f_m) \ ,$$

and so $\{f_n(x)\}_{n\in\mathbb{N}}$ is also a Cauchy sequence. Since $\mathbb{R}$ is complete, $\{f_n(x)\}_{n\in\mathbb{N}}$ is convergent. Define $f(x)$ to be the limit of this sequence. (Since $\mathbb{R}$ is Hausdorff, the limit is unique, and so $f(x)$ is well-defined.)

Now, fix $\epsilon > 0$, Since $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence, there exists an N_ϵ such that

$$n, m \ge N_\epsilon \quad \Rightarrow \quad d(f_n, f_m) \le \frac{\epsilon}{2} \ .$$

For each x,

$$|f_n(x) - f(x)| \le |f_n(x) - f_m(x)| + |f_m(x) - f(x)| \ .$$

Now choose m so large that $m \ge N_\epsilon$ and $|f_m(x) - f(x)| < \epsilon/2$. Thus, for all all $x \in X$,

$$n \ge N_\epsilon \quad \Rightarrow \quad |f_n(x) - f(x)| < \epsilon \ .$$

That is, since N_ϵ does not dependent on x, f is the uniform limit of the sequence $\{f_n\}$.

We now claim that $f \in \mathcal{C}(X,\mathbb{R})$. For all $x, y \in X$,

$$|f(x) - f(y)| \le |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \ .$$

Then, for $n > N_{\epsilon/3}$

$$|f(x) - f(y)| \le \frac{2}{3}\epsilon + |f_n(x) - f_n(y)| \ .$$

Since f_n is continuous, there is a $\delta_\epsilon > 0$ so that

$$d(y,x) < \delta_\epsilon \quad \Rightarrow \quad |f_n(x) - f_n(y)| < \frac{\epsilon}{3} \ .$$

Thus,

$$d(y,x) < \delta_\epsilon \quad \Rightarrow \quad |f(x) - f(y)| < \epsilon \ .$$

Thus f is continuous a x, and since $x \in X$ is arbitrary, $f \in \mathcal{C}(X,\mathbb{R})$, and have already seen that $\lim_{n\to\infty} d(f_n, f) = 0$. We have shown that $\mathcal{C}(X,\mathbb{R})$ is complete. □

We shall now characterize the compact sets in $\mathcal{C}(X,\mathbb{R})$.

1.24 DEFINITION (Equicontinuous, pointwise bounded)**.** Let $\mathcal{F} \subset \mathcal{C}(X,\mathbb{R})$. Then $\mathcal{F}$ is *equicontinuous* in case for each $\epsilon > 0$ and each $x \in X$, there is a neighborhood U_ϵ of x such that

$$y \in U_\epsilon \Rightarrow |f(y) - f(x)| < \epsilon \qquad \text{for all } f \in \mathcal{F} \ .$$

Also, $\mathcal{F}$ is *pointwise bounded* in case for each $x \in X$, $\{f(x) \ : f \in \mathcal{F}\}$ is a bounded subset of $\mathbb{R}$.

1.25 THEOREM (Arzelà-Ascoli)**.** *Let X be a compact topological space space, and let $\mathcal{F}$ be an equicontinuous and pointwise bounded subset of $\mathcal{C}(X,\mathbb{R})$. Then the closure of $\mathcal{F}$ is compact if and only if $\mathcal{F}$ is pointwise bounded and equicontinuous.*

Proof. Suppose that $\mathcal{F}$ is pointwise bounded and equicontinuous. Since $\mathcal{C}(X, \mathbb{R})$, the closure of $\mathcal{F}$ is complete, and hence it suffices to show that $\mathcal{F}$ is totally bounded.

Fix $\epsilon > 0$. Since $\mathcal{F}$ is equicontinous, at each $x \in X$, there is an open set U_x such that $|f(y) - f(x)| < \epsilon/3$ for all $y \in U_x$. Since X is compact, and since $\{U_x \ : \ x \in X\}$ is an open cover of X, there exists a finite subcover $\{U_{x_1}, \dots, U_{x_n}\}$.

We now claim that for this set $\{x_1, \dots, x_m\}$,

$$\max\{|f(x_j) - g(x_j)|\} < \frac{\epsilon}{3} \qquad \Rightarrow \qquad d(f,g) < \epsilon \ .$$

To see this, note that every x belongs to some U_{x_j}, and hence

$$|f(x) - g(x)| \leq |f(x) - f(x_j)|| + |f(x_j) - g(x_j)| + |g(x_j) - g(x)| < \epsilon \ .$$

Now define a map Φ from $\mathcal{F}$ to $\mathbb{R}^n$ by

$$\Phi(f) = (f(x_1), \dots, f(x_n)) \ .$$

Now, since $\mathcal{F}$ is pointwise bounded, $\Phi(\mathcal{F})$ is a bounded, and hence totally bounded subset of $\mathbb{R}^n$. Conider any open cover of $\Phi(\mathcal{F})$ by balls of radius $\epsilon/3$. Then the inverse images of these balls are an open cover of $\mathcal{F}$ ball by sets of diameter no more than ϵ. Replacing them with the balls of radius ϵ about any point in them, we have our finite open cover by balls of radius ϵ. sine $\epsilon > 0$is arbitrary, $\mathcal{F}$ is totally bounded.

Next, suppose that the closure of $\mathcal{F}$ is compact. Then $\mathcal{F}$ is totally bounded, so that for each $\epsilon > 0$, there is a finite set $\{f_1, \dots, f_n\} \subset \mathcal{F}$ such that

$$\mathcal{F} \subset \bigcup_{j=1}^{n} B_{\epsilon/4}(f_j) \ .$$

For any x, and choosing, say, $\epsilon = 4$, we have

$$\{f(x) \ : f \in \mathcal{F} \ \} \subset \bigcup_{j=1}^{n} (f_j(x) - 1, f_j(x) + 1) \ .$$

Thus, $\mathcal{F}$ is point wise bounded.

Next, for each $x \in X$, define $U_x \subset X$ by

$$U_x = \bigcap_{j=1}^{n} \{ \ y \in X \ : \ |f_j(y) - f_j(x)| < \epsilon/4\} \ .$$

Evidently, U_x is a finite intersection of open sets, and therefore open. Also, $x \in U_x$. Thus

$$\bigcup_{x \in X} U_x = X \ ,$$

and then since X is compact, there is a finite set $\{x_1, \dots, x_m\} \subset X$ such that

$$\bigcup_{k=1}^{m} U_{x_k} = X \ .$$

Now, fix any $x \in X$. By the above, $x \in U_{x_k}$ for some $k \in \{1, \ldots, m\}$. For any $f \in \mathcal{F}$, there is some $j \in \{1, \ldots, n\}$ such that $d(f, f_j) < \epsilon/4$. But then for any $y \in U_{x_k}$,

$$\begin{aligned} |f(y) - f(x)| &\leq |f(y) - f_j(y)| + |f_j(y) - f_j(x)| + |f_j(x) - f(x)| \\ &\leq \frac{\epsilon}{4} + |f_j(y) - f_j(x)| + \frac{\epsilon}{4} \end{aligned} \tag{1.8}$$

where the first inequality is the triangle inequality, the second uses the fact that $d(f, f_j) < \epsilon/4$.

Then, once more by the triangle inequality, and then the definition of U_{x_k},

$$|f_j(y) - f_j(x)| \leq |f_j(y) - f_j(x_k)| + |f_j(x_k) - f_j(x)| \leq \frac{\epsilon}{2} \ .$$

Altogether, we have $|f(x) - f(y)| < \epsilon$ whenever $y \in U_{x_k}$, and $f \in \mathcal{F}$, so that U_{x_k} is the required neighborhood of x. Hence, $\mathcal{F}$ is equicontinuous. □

1.6 The Stone-Wierstrass Theorem

The Stone-Wierstrass Theorem is an approximation theorem that generalizes the classical Wierstrass Approximation Theorem that we discussed at the beginning of these notes.

We begin with two definitions. Let X be a compact topological space, and let $\mathcal{C}(X, \mathbb{R})$ be the space of continuous real valued functions on X equipped with the uniform metric. A subset $\mathcal{A}$ of $\mathcal{C}(X, \mathbb{R})$ is an *algebra* in case $\mathcal{A}$ is a vector subspace over $\mathbb{R}$ of $\mathcal{C}(X, \mathbb{R})$ equipped with its usual rules of addition and scalar multiplication, and if, moreover, for every $f, g \in \mathcal{A}$, the pointwise product fg also belongs to $\mathcal{A}$.

A subset $\mathcal{A}$ of $\mathcal{C}(X, \mathbb{R})$ is *separating* in case for pair of distinct points x, y in X, there is an $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.

Notice that if X is not Hausdorff, not even $\mathcal{C}(X, \mathbb{R})$ the space of *all* continuous real valued functions on X is separating. Indeed, if X is not Hausdorff, there exist two distinct points x and y in X such that every neighborhood U of x contains y. But then for any continuous function f, $f(x) = f(y)$. Hence throughout this subsection, we shall only be concerned with Hausdorff topological spaces.

The primary example of a separating algebra to keep in mind is $X = [0, 1]$, with $\mathcal{A}$ being the algebra of all polynomials in the real variable $x \in [0, 1]$. To see that this algebra is separating, consider the polynomial $p(x) = x$. Then for $x_0 \neq x_1$ in X, $p(x_0) \neq p(x_1)$.

1.26 THEOREM (Stone-Wierstrass)**.** *Let X be a compact topological space, and let $\mathcal{A}$ be a subset of $\mathcal{C}(X, \mathbb{R})$ that is a separating algebra. Let $\mathcal{B}$ be the uniform closure of $\mathcal{A}$. Then either $\mathcal{B} = \mathcal{C}(X, \mathbb{R})$, or else $\mathcal{B}$ consisits of all continuous functions on X that vanish at some fixed point x_0. In particular, if $\mathcal{A}$ contains the constant functions, $\mathcal{B} = \mathcal{C}(X, \mathbb{R})$.*

We will prove Theorem 1.26 as a consequence of two lemmas, and shall make use of the *partial order* in $\mathcal{C}(X, \mathbb{R})$: If $f, g \in \mathcal{C}(X, \mathbb{R})$, we write $f \leq g$ in case $f(x) \leq g(x)$ for all $x \in X$. With this partial order, $\mathcal{C}(X, \mathbb{R})$ is a *lattice*: Given any $f, g \in \mathcal{C}(X, \mathbb{R})$ there is a unique function $g \wedge f \in \mathcal{C}(X, \mathbb{R})$ such that $g \wedge f \leq f, g$, and such that $h \leq g \wedge f$ whenever $h \leq f, g$. Of course, $g \wedge f$ is defined by

$$g \wedge f(x) = \min\{ \ f(x) \ , \ g(x) \ \} \ ,$$

which is continuous.

Likewise, given any $f, g \in \mathcal{C}(X, \mathbb{R})$ there is a unique function $g \vee f \in \mathcal{C}(X, \mathbb{R})$ such that $f, g \leq g \vee f$, and such that $g \vee f \leq h$ whenever $f, g \leq h$. Of course, $g \wedge f$ is defined by

$$g \vee f(x) = \max\{\ f(x)\ ,\ g(x)\ \}\ ,$$

which is continuous.

A subset $\mathcal{F}$ of $\mathcal{C}(X, \mathbb{R})$ is itself a lattice if whenever $f, g \in \mathcal{F}$, then both $f \wedge g$ and $f \vee g$ belong to $\mathcal{F}$. Then observing that

$$f \wedge g = \frac{1}{2}(f + g - |f - g|) \qquad \text{and} \qquad f \vee g = \frac{1}{2}(f + g + |f - g|)\ , \tag{1.9}$$

we see that a subset $\mathcal{F}$ of $\mathcal{C}(X, \mathbb{R})$ is itself a lattice if whenever $f \in \mathcal{F}$, then $|f| \in \mathcal{F}$.

1.27 LEMMA (Limit point criterion for lattices in $\mathcal{C}(X, \mathbb{R})$)**.** *Let X be a compact Hausdorff space. Let $\mathcal{F} \subset \mathcal{C}(X, \mathbb{R})$ be a lattice.*

If f is any element of $\mathcal{C}(X, \mathbb{R})$ with the property that for every $x, y \in X$, there exists a function $f_{x,y} \in \mathcal{F}$ for which

$$f_{x,y}(x) = f(x) \qquad \text{and} \qquad f_{x,y}(y) = f(y)\ . \tag{1.10}$$

Then f is a limit point of $\mathcal{F}$; i.e., it belongs to the closure of $\mathcal{F}$.

Proof. Fix any $f \in \mathcal{C}(X, \mathbb{R})$ with the property every $x, y \in X$, there exists a function $f_{x,y} \in \mathcal{F}$ such that (1.10) is satisfied. Fix any $\epsilon > 0$. We must show that there exists some $g \in \mathcal{F}$ with $|g(x) - f(x)| < \epsilon$ for all $x \in X$.

First, for each $(x, y) \in X \times X$, make some choice of $f_{x,y}$, and define the open set $U_{x,y} \subset X$ by

$$U_{x,y} = \{\ z\ :\ f_{x,y}(z) < f(z) + \epsilon\ \}\ .$$

Evidently, $x, y \in U_{x,y}$. Therefore

$$X = \bigcup_{x \in X} U_{x,y}\ ,$$

and then, since X is compact, there exists a finite set $\{x_1, \ldots, x_n\} \subset X$ such that

$$X = \bigcup_{j=1}^{n} U_{x_j,y}\ .$$

Now define the function f_y by

$$f_y = f_{x_1,y} \wedge f_{x_2,y} \wedge \cdots \wedge f_{x_n,y}\ .$$

Since $\mathcal{F}$ is a lattice, $f_y \in \mathcal{F}$, and

$$f_y \leq f + \epsilon$$

in the lattice order; i.e., everywhere on X.

Furthermore, since $f_{x_j,y}(y) = f(y)$ for each j, $f_y(y) = f(y)$. Therefore, defining the open set V_y by

$$V_y := \{\ z \in X\ :\ f(z) - \epsilon < f_y(z)\ \}\ ,$$

we have $y \in V_y$, and hence

$$X = \bigcup_{y \in X} V_y \ ,$$

hen, since X is compact, there exists a finite set $\{y_1, \ldots, y_m\} \subset X$ such that

$$X = \bigcup_{k=1}^{m} V_{y_k} \ .$$

Now define g by

$$g = f_{y_1} \vee f_{y_2} \vee \ldots, \vee f_{y_m} \ .$$

Then since $\mathcal{F}$ is a lattice, $g \in \mathcal{F}$, and by construction,

$$f - \epsilon \le g \le f + \epsilon \ ,$$

which means that $|f(x) - g(x)| < \epsilon$ for all $x \in X$. □

1.28 LEMMA (A closed algebra in $\mathcal{C}(X, \mathbb{R})$ is a lattice)**.** *Let X be a compact Hausdorff space. Let $\mathcal{B}$ be a closed subset of $\mathcal{C}(X, \mathbb{R})$ that is also a subalgebra of $\mathcal{C}(X, \mathbb{R})$. Then $\mathcal{B}$ is a lattice.*

Proof. By the remarks we have made concerning (1.9), it suffices to show that for all $f \in \mathcal{B}$, $|f| \in \mathcal{B}$. Since X is compact and f is continuous, f is bounded above and below, and hence there is a finite positive number c such that $|cf| \le 1$. Then since $|cf| = c|f|$, we may freely suppose that $|f| \le 1$.

Therefore, fix any $f \in \mathcal{B}$ with $|f| \le 1$, We shall complete the proof by showing that there exists a sequence of polynomials $\{p_n\}_{n \in \mathbb{N}}$ so that

$$|f| = \lim_{n \to \infty} p_n(f^2) \tag{1.11}$$

in the uniform topology. Since $\mathcal{B}$ is an algebra, $p_n(f^2) \in \mathcal{B}$ for each n, and then since $\mathcal{B}$ is closed, $|f| \in \mathcal{B}$.

For any number $a \in [0, 1]$, we define a sequence $\{b_n\}_{n \in \mathbb{N}}$ recursively as follows: We set $b_1 = 0$ and then for al $n \in \mathbb{N}$,

$$b_{n+1} = b_n + \frac{a - b_n^2}{2} \ .$$

Notice that

$$b_1 = 0 \ , \quad b_2 = \frac{a}{2} \ , \quad b_3 = a - \frac{a^2}{8} \ ,$$

and so forth. It is easy to see by induction that for each n, there is a polynomial p_n, independent of the value of a, so that such that $b_n = p_n(a)$.

We claim that $\sqrt{a} = \lim_{n \to \infty} b_n$. This will give us a sequence of polynomials $\{p_n\}_{n \in \mathbb{N}}$ such that for each $a \in [0, 1]$,

$$\sqrt{a} = \lim_{n \to \infty} p_n(a) \ ,$$

and therefore, such that

$$|f(x)| = \lim_{n \to \infty} p_n(f^2(x))$$

for all x in X. Then, since X is compact, Dini's Theorem implies that (1.11) is true with uniform convergence.

Hence, we need only verify the claim that $\sqrt{a} = \lim_{n\to\infty} b_n$. To do this, note that

$$\sqrt{a} - b_{n+1} = \sqrt{a} - b_n - \frac{(\sqrt{a} - b_n)(\sqrt{a} + b_n)}{2} = (\sqrt{a} - b_n)\left(1 - \frac{\sqrt{a} + b_n}{2}\right) .$$

Since $a \le 1$, as long as $b_n \le \sqrt{a}$, the right hand side is non-negative, and therefore $b_{n+1} \le \sqrt{a}$. Since $b_1 \le \sqrt{a}$, it follows that $\sqrt{a}$ is an upper bound for the sequence $\{b_n\}_{n\in\mathbb{N}}$.

Now, knowing that $b_n^2 \le a$ for all n, it is clear from the definition that $\{b_n\}_{n\in\mathbb{N}}$ is a monotone non-decreasing sequence. Therefore the limit $b = \lim_{n\to\infty} b_n$ exists and satisfies

$$b = b + \frac{a - b^2}{2} .$$

This means that $b^2 = a$, and since $b \ge 0$, $b = \sqrt{a}$. □

Proof of Theorem 1.26: Fix $x \ne y$ in X, and consider the linear transformation from $\mathcal{A}$ to $\mathbb{R}^2$ given by

$$f \mapsto (f(x), f(y)) .$$

The range of this linear transformation is a subspace S of $\mathbb{R}^2$.

Since $\mathcal{A}$ separates, there can be at most one point $x_0 \in X$ for which $g(x_0) = 0$ for all $g \in \mathcal{A}$.

Let us first assume first that neither x nor y is such a point. Since $\mathcal{A}$ is an algebra, and a vector space in particular, if g is in $\mathcal{A}$ so is very multiple of g. By assumption, there is some $g \in \mathcal{A}$ such that $g(x) \ne 0$, and by choosing an appropriate multiple, we may arrange that $g(x) = 1$.

Thus, S contains a vector of the form $(1, a)$. (with Since $\mathcal{A}$ separates, we can choose $g \in \mathcal{A}$ so that $a \ne 1$.

Now there are two cases to consider. If also $a \ne 0$, then the two vectors $(1, a)$ and $(1, a^2)$ are linearly independent, and $(1, a^2)$ also belongs to S since $\mathcal{A}$ is an algebra (so that $g^2 \in \mathcal{A}$). On the other hand if $a = 0$ then S contains the vector $(1, 0)$, and, since there is some other g with $g(y) = 1$, there is some $b \in \mathbb{R}$ such that $(b, 1)\ inS$. Hence in this case, S contains the two vectors $(1, 0)$ and $(b, 1)$ which are linearly independent. Either way, $S = \mathbb{R}^2$, and so we have proved that as long as $g(x) \ne 0$ and $h(y)\ ne0$ for *some* $g, h \in \mathcal{A}$, then S is all of $\mathbb{R}^2$.

This has the consequence that for *any* $f \in \mathcal{C}(X, \mathbb{R})$, we can find a function $f_{x,y} \in \mathcal{A}$ for which

$$(f(x), f(y)) = (f_{x,y}(x), f_{x,y}(y)) . \tag{1.12}$$

Now we have two cases once more: Suppose first that there is no point $x_0 \in X$ with $f(x_0) = 0$ for all $f \in \mathcal{A}$. Then the above argument applies for *all* x and y in X and all $f \in \mathcal{C}(X, \mathbb{R})$, we can find $f_{x,y} \in \mathcal{A}$ such that (1.12) is true. Moreover, by Lemma 1.28, $\mathcal{B}$ is a lattice. Therefore, by Lemma 1.27, f is a linit point of $\mathcal{B}$, and since $\mathcal{B}$ is closed, $f \in \mathcal{B}$. Since f is an arbitrary element of $\mathcal{C}(X, \mathbb{R})$, we see that in this case, $\mathcal{B} = \mathcal{C}(X, \mathbb{R})$.

The remaining case to consider is that in which there is one point x_0 such that $g(x_0) = 0$ for all $g \in \mathcal{A}$, and hence$\mathcal{B}$, so that $\mathcal{B}$ is certainly contained in the closed subset of $\mathcal{C}(X, \mathbb{R})$ consisting of continuous functions f on X such that $f(x_0) = 0$.

Let f be any such function. The argument made above show that as long as neither x nor y equals x_0, then there is some $g \in \mathcal{A}$, and hence $\mathcal{B}$, for which (1.12) is true. Now suppose that $x = x_0$, and $y \neq x_0$. Then we trivially have

$$f(x_0) = g(x_0) = 0$$

for all $g \in \mathcal{B}$. And since $\mathcal{A}$ separates, and is a vector space, we can choose g so that $f(y) = g(y)$. Therefore, for any $f \in \mathcal{C}(X, \mathbb{R})$ with $f(x_0) = 0$, no matter how x and y are chosen, we can can find $g_{x,y} \in \mathcal{B}$ so that (1.12) is true.

Then the argument made above shows that every $f \in \mathcal{C}(X, \mathbb{R})$ with $f(x_0) = 0$ is a limit point of $\mathcal{B}$, and hence belongs to $\mathcal{B}$. Therefore, in this second case, $\mathcal{B}$ is the subset of $\mathcal{C}(X, \mathbb{R})$ consisting of functions f with $f(x_0) = 0$. □

In our proof of Theorem 1.26, we made use of the fact that our functions f were real valued, and not complex valued: The real numbers are ordered, while the complex numbers are not, and the order on the complex number played a crucial role in the proof through our use of Lemma 1.27.

This is not simply an artifact of the proof: If in the statement of the theorem we replace $\mathcal{C}(X, \mathbb{R})$ by, $\mathcal{C}(X, \mathbb{C})$, the space of continuous complex valued functions on X, the statement becomes false.

To see this, take X to be the closed unit disc in the complex plane $\mathbb{C}$. Take $\mathcal{A}$ to be the algebra of all complex polynomials in the complex variable z, which clearly separates. Polynomials in z are analytic, and uniform limits of analytic functions are analytic, and so the closure of $\mathcal{A}$ consists of functions that are analytic in the interior of the the unit disc. Obviously, not every continuous function of the closed unit disc is analytic in the interior of the disc; $f(z) = z^*$, the complex conjugate of z, is an example. Hence, the uniform closure of $\mathcal{A}$ is not the full set of continuous complex valued functions on the closed unit disc.

However, under one simple additional condition on the algebra $\mathcal{A}$, one can reduce the complex valued case to the real case.

A (complex) subalgebra $\mathcal{A}$ of the algebra of complex valued function on a compact Hausdorf space is called a $*$-*algebra* in case it is closed under complex conjugation. That is, whenver $f \in \mathcal{A}$, then $f^* \in \mathcal{A}$, where f^* is the function defined by $f^*(x) = (f(x))^*$ for all $x \in X$.

In this case, for every $f \in \mathcal{A}$, the real and imaginary parts of f both belong to $\mathcal{A}$. It is also easy to see that when $\mathcal{A}$ separates, so does the real algebra consisting of the real and imaginary parts of functions in $\mathcal{A}$. Applying the Stone-Wierstrass Theorem to this algebra, one can separately approximate, in the uniform metric, the real and imaginary parts of any continuous complex valued function on X by functions in $\mathcal{A}$.

In summary, we have:

1.29 THEOREM (Complex Stone-Wierstrass). *Let X be a compact topological space, and let $\mathcal{A}$ be a subset of $\mathcal{C}(X, \mathbb{C})$ that is a separating $*$-algebra. Let $\mathcal{B}$ be the uniform closure of $\mathcal{A}$. Then either $\mathcal{B} = \mathcal{C}(X, \mathbb{C})$, or else $\mathcal{B}$ consisits of all continuous functions on X that vanish at some fixed point x_0. In particular, if $\mathcal{A}$ contains the constant functions, $\mathcal{B} = \mathcal{C}(X, \mathbb{R})$.*

Here is one important application of Theorem 1.29: Let X be the unit circle in $\mathbb{C}$, with its usual topology. Let $\mathcal{A} \subset \mathcal{C}(x, \mathbb{C})$ be the set consisting of functions f of the form

$$f(z) = \sum_{j=-n}^{n} a_j z^n$$

for some $n \in \mathbb{N}$, and some numbers $a_{-n}, \ldots, a_n$ in $\mathbb{C}$. (Each element of X is a complex number z, and z^n denotes the nth power of z.) The elements of $\mathcal{A}$ are called *complex trigonomentric polynomials*

It is easy to see that $\mathcal{A}$ is a $*$-algebra, and that $\mathcal{A}$ separates. Hence, by Theorem 1.29, $\mathcal{A}$ is dense in $\mathcal{C}(X, \mathbb{C})$. This proves:

1.30 THEOREM (Density of Complex Trigonometric Polynomials)**.** *Let X be the unit circle in $\mathbb{C}$, with its usual topology. Then the set of complex trigonometric polynomials is dense in $\mathcal{C}(X, \mathbb{C})$, with respect to the uniform topology.*

2 Tychonoff's Theorem

Let X be a set. The *Cartesian product of X with itself*, $X \times X$, is the set of all ordered pairs (x_1, x_2) of elements of X. Of course (x_1, x_2) is the graph of a unique function $f : \{1, 2\} \to X$, namely the one with $f(1) = x_1$ and $f(2) = x_2$. (One can accomodate Cartesian products of two different sets Y and Z in this framework by considering $X = Y \cup Z$ and restricting attention to functions f such that $f(1) \in Y$ and $f(2) \in Z$. No real generality is lost in taking the sets to be the same, and the notation is much simpler, so that is how we shall proceed.)

More generally, given any set non-empty S, the Cartesian product of X indexed by S, denotes X^S, is the set of all functions from S to X. For example, $X^{\mathbb{N}}$ is the set of all infinite sequences $\{x_n\}_{n \in \mathbb{N}}$ of elements of X.

On any Cartesian product, there is a natural family of functions with values in X, namely the *coordinate functions*: For each $s \in S$, define

$$\varphi_s : X^S \to X$$

by

$$\varphi(f) = f(s) \ .$$

That is, one simply evaluates the function $f \in X^S$ at s.

Note that when $S = \{1, 2\}$, $\varphi_j((x_1, x_2)) = x_j$, which is why the φ_s are called coordinate functions.

Now suppose that X is a topological space. Is there a nice topology on X^S for which all of the coordinate functions are continuous? Of course, there are plenty of topologies on X^S for which all of the coordinate functions are continuous, but they are not necessarily very nice: For example, if we equip X^S with the power set topology; i.e., the topology consisting of *all* subsets of X^S, then *every* function on X^S is continuous.

However, with such a topology, as long as X has infinitely many elements, X^S will not be compact. On the other hand, if we take $X = [0, 1]$ and $S = \{1, 2\}$, then we can identify X^S with the closed unit square in $\mathbb{R}^2$. If we equip this with its usual metric toplogy, then both coordinate functions are clearly continuous, and the Cartesian product itself is compact.

Tychonoff's theorem says that there is a natural topology on any Cartesian product that is "nice" in the sense that under this topology we have both:

(1) *If X is compact, then so is X^S, no matter what non-empty set S is.*

(2) *For each $s \in S$, the coordinate function φ_s is continuous*

To find such a topology $\mathcal{O}$, we must include enough open sets to make each ϕ_s continuous, and should avoid introducing too many more than that. The more open sets we include, the more open covers there are to consider, and if we produce to many of these, some might well lack finite subcovers.

The way to proceed turns on the following very simple observation: Suppose that Y is any set, and $\mathcal{A}$ is any set of subsets of Y. Let $\mathcal{U}$ and $\mathcal{V}$ be two topologies on Y such that $\mathcal{A} \subset \mathcal{U}$ and $\mathcal{A} \subset \mathcal{V}$. Then $\mathcal{U} \cap \mathcal{V}$ is also a topology – it is very easy to see that is satisfies the three requirements in the definition, and clearly, it contains $\mathcal{A}$.

More generally, consider the set of *all* topologies $\mathcal{U}$ on Y that contain $\mathcal{A}$. The intersection of all of the topologies is again a topology that contains $\mathcal{A}$. By construction, it is the *smallest* topology on Y that contains $\mathcal{A}$ in the sense that any other topology that contains $\mathcal{A}$ also contains every set in this one. One often refers to this topology as the *topology of* Y *generated by* $\mathcal{A}$.

2.1 DEFINITION (Product topology)**.** Let $(X, \mathcal{O})$ be a topological space, and S a nonempty set. The *product toplogy* on a Cartesian product X^S the smallest topology on X^S containing all of the sets of the form

$$\phi_s^{-1}(U) \ , \qquad s \in S \ , \quad U \in \mathcal{O} \ .$$

By construction, each of the coordinate functions is continuous when X^S is equipped with the product Topology. Moreover:

2.2 THEOREM (Tychonoff's Theorem)**.** *Let X be a compact topological space, and S any non-empty set. Then X^S, equipped with the product topology, is compact.*

The special case of this theorem in which X is a compact metric space and S is countable (or finite) is fairly easy to prove using the theorems presented so far in these notes. This is developed in the exercises that follow. The general case involves either the theory of "nets" or the theory of "filters", and this would be a digression, since we shall not invoke the general case in this book, nor shall we have any other occasion to use the theory of nest of filters. Furthermore, the proof of the general case involves the axiom of choice in a much more subtle way than does the spacial case. This is not a problem, but discussion of these subtleties would take us far afield. (The axiom of choice enters the subject, even in the special case, in an essential way: It is the axion of choice whihc assures us that X^S is non-empty: one can always choose, for each $x \in s$, some $x(s) \in X$.)

It is well worth knowing the general case nonetheless. It shows that advantage of the 20th century notion of *compactness*, as defined above, in terms of open covers, and the 19th century notion of sequential compactness