44. STONE-WEIERSTRASS THEOREM

Here we address an important approximation theorem in the space of continuous functions. An early version of this result is due to K. Weierstrass. M.H. Stone extended and significantly generalized this result to what now bears both of their names.

44.1 Weierstrass approximation theorem.

Many practical tasks and, particularly, calculations on a computer, lead to the following problem: Given a continuous function $f\colon [a,b] \to \mathbb{R}$, is there a way to encode it (say, up to a prescribed error bar ϵ) using a finite collection of numbers. There are many answers to this problem one of which is an approximation by polynomials. An early result on this was proved by K. Weirstrass in 1885:

Theorem 44.1 (Weierstrass approximation theorem) *Let $a < b$ be reals. Then for each $f \in C([a,b],\mathbb{R})$ and each $\epsilon > 0$ there is a polynomial P such that*

$$\sup_{x\in[a,b]} \bigl|f(x) - P(x)\bigr| < \epsilon. \tag{44.1}$$

Equivalently, for each $f \in C([a,b],\mathbb{R})$ there is a sequence $\{P_n\}_{n\in\mathbb{N}}$ of polynomials such that $P_n \to f$ uniformly on $[a,b]$.

Proof. A number of proofs of this result exist some of which require detours to concepts that are quite interesting in their own right but may be found difficult at first (see the textbook). We will give a proof discovered by S. Bernstein in 1912 that is based on probabilistic ideas. This concerns mainly the understanding of why certain choices are made; no particular knowledge of probability is required to follow the steps formally.

First we note that it suffices to prove the claim for continuous functions on $[0,1]$. Indeed, if $f \in C([a,b],\mathbb{R})$ then $g(x) := f(a+(b-a)x)$ defines $g \in C([0,1],\mathbb{R})$. If Q is a polynomial such that $\sup_{x\in[0,1]} |g(x) - Q(x)| < \epsilon$, then $P(x) := Q(\frac{x-a}{b-a})$ is a polynomial such that (44.1) holds.

Let thus $f \in C([0,1],\mathbb{R})$ and define

$$P_n(x) := \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f(k/n). \tag{44.2}$$

This is a polynomial of degree n which has a probabilistic interpretation: Indeed, the expression $\binom{n}{k}x^k(1-x)^{n-k}$ is known to be the probability that n independent coin tosses by a coin with probability of getting heads equal to x result in k heads. Regardless of this interpretation, thanks to the Binomial theorem

$$\forall n \in \mathbb{N}\, \forall a,b \in \mathbb{R}\colon\ (a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}, \tag{44.3}$$

the following identities are true (and will be useful): First,

$$\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = 1, \tag{44.4}$$

which means that $\{\binom{n}{k}x^k(1-x)^{n-k}\}_{k=0}^n$ is indeed a probability mass function. This is obtained by plugging $a := x$ and $b := 1-x$ into (44.2). Taking a derivative of (44.4) with respect to a, multiplying the expression by a then setting $a := x$ and $b := 1-x$ then gives

$$\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}k = nx, \tag{44.5}$$

which quantifies the expected number of heads in n coin tosses. Finally, taking two derivatives with respect to a also gives

$$\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}k^2 = (nx)^2 + nx(1-x), \tag{44.6}$$

which says that the variance of the number of heads in n coin tosses is $nx(1-x)$. The polynomial P_n then has the meaning of the expected value of f of the average number of heads in n coin tosses by a coin with success probability x.

The intuition why P_n is a good approximation of f for large n comes from the fact that, by the Weak Law of Large Numbers, this average number tends to x as $n \to \infty$ which means that most of the weight in the sum in (44.7) is on k with $k/n - x$ small. To use this intuition through analytic reasoning, we first note that, by (44.4),

$$f(x) - P_n(x) = \sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}\,\big[f(x) - f(k/n)\big]. \tag{44.7}$$

Next we note that f is uniformly continuous, which means

$$\forall\epsilon > 0\ \exists\delta > 0\ \forall x, y \in [0,1]\colon\ \ |x-y| < \delta\ \Rightarrow\ \big|f(y) - f(x)\big| < \epsilon. \tag{44.8}$$

For $\delta > 0$ related to an $\epsilon > 0$ as in (44.8) we now observe

$$\begin{aligned}\big|f(x) - P_n(x)\big| &\leqslant \sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}\,\big|f(x) - f(k/n)\big| \\ &= \sum_{\substack{k=0,\dots,n\\ |k/n-x|<\delta}}\binom{n}{k}x^k(1-x)^{n-k}\,\big|f(x) - f(k/n)\big| \\ &\qquad + \sum_{\substack{k=0,\dots,n\\ |k/n-x|\geqslant\delta}}\binom{n}{k}x^k(1-x)^{n-k}\,\big|f(x) - f(k/n)\big|.\end{aligned} \tag{44.9}$$

In the first sum we use (44.8) to bound $|f(x) - f(k/n)| < \epsilon$, while in the second sum we simply bound $|f(x) - f(k/n)| \leqslant 2\|f\|$, where $\|f\| := \sup_{x\in[0,1]}|f(x)|$ is the supremum norm of f. In light of (44.4), this gives

$$\big|f(x) - P_n(x)\big| \leqslant \epsilon + 2\|f\| \sum_{\substack{k=0,\dots,n\\ |k/n-x|\geqslant\delta}}\binom{n}{k}x^k(1-x)^{n-k}. \tag{44.10}$$

It remains to bound the sum on the right. Here we note that, for all k that contribute to the sum we have $1 \leqslant (\frac{k-nx}{n\delta})^2$ and so

$$\sum_{\substack{k=0,\dots,n\\|k/n-x|\geqslant\delta}} \binom{n}{k} x^k(1-x)^{n-k} \leqslant \sum_{k=0}^{n} \binom{n}{k} x^k(1-x)^{n-k}\Bigl(\frac{k-nx}{n\delta}\Bigr)^2 \tag{44.11}$$

The right-hand side is now worked out using the identities (44.4–44.6) to equal

$$\frac{1}{n^2\delta^2}\Bigl((nx)^2 + nx(1-x) - 2(nx)^2 + (nx)^2\Bigr) = \frac{1}{n\delta^2}x(1-x) \leqslant \frac{1}{4n\delta^2}. \tag{44.12}$$

Hence we get

$$\bigl|f(x) - P_n(x)\bigr| \leqslant \epsilon + \frac{\|f\|}{2n\delta^2}. \tag{44.13}$$

This bound does not depend on x and so it holds for all $x \in [0,1]$. We thus get

$$\forall n \in \mathbb{N}\colon\ n \geqslant \frac{\|f\|}{2\epsilon\delta^2} \ \Rightarrow\ \sup_{x\in[0,1]} \bigl|f(x) - P_n(x)\bigr| \leqslant 2\epsilon. \tag{44.14}$$

Since $\epsilon > 0$ is arbitrary, this is the desired claim. □

One important consequence of the Weierstrass theorem is:

Lemma 44.2 *Let $f \in C([a,b],\mathbb{R})$ be such that*

$$\forall n \in \mathbb{N}\colon \quad \int_a^b f(x)x^n \mathrm{d}x = 0. \tag{44.15}$$

Then $f = 0$.

Proof. By additivity of the Riemann integral we get

$$\forall P \text{ polynomial}\colon \quad \int_a^b f(x)P(x)\mathrm{d}x = 0 \tag{44.16}$$

Let $\{P_n\}_{n\in\mathbb{N}}$ be a sequence of polynomials such that $P_n \to f$ uniformly on $[a,b]$. Since f is bounded, we also have the uniform convergence of $fP_n \to f^2$ and, by Theorem 40.1,

$$0 = \int_a^b f(x)P_n(x)\mathrm{d}x \underset{n\to\infty}{\longrightarrow} \int_a^b f(x)^2\mathrm{d}x. \tag{44.17}$$

The claim now follows from the fact that, for continuous f, vanishing of the integral of f^2 implies $f = 0$. □

Here is a consequence of this result for the metric structure of $C([a,b],\mathbb{R})$:

Corollary 44.3 *For each $a < b$, the class of polynomials with rational coefficients is dense in $C([a,b],\mathbb{R})$. In particular, $C([a,b],\mathbb{R})$ is separable.*

Proof. We have already shown that the polynomials are dense in $C([a,b],\mathbb{R})$. To get the claim notice that, by density of $\mathbb{Q}$ in $\mathbb{R}$, every polynomial can be approximated, uniformly on $[a,b]$, by a polynomial with rational coefficients. □

44.2 Stone-Weierstrass theorem.

The metric formulation of the Weierstrass approximation theorem underlies its considerable generalization proved by M.H. Stone in 1937 that deals with continuous real-valued functions on compact (and even locally compact) spaces. The simplest version of this result is as follows:

Theorem 44.4 (Stone-Weierstrass theorem) *For X a compact metric space, let $\mathcal{A} \subseteq C(X, \mathbb{R})$ be a class of functions such that:*

(1) ($\mathcal{A}$ is an algebra)

$$\forall f, g \in \mathcal{A}\ \forall \lambda \in \mathbb{R}\colon\ f + g \in \mathcal{A}\ \wedge\ \lambda f \in \mathcal{A}\ \wedge\ f \cdot g \in \mathcal{A} \tag{44.18}$$

(2) ($\mathcal{A}$ does not vanish at a point)

$$\forall x \in X\, \exists f \in \mathcal{A}\colon\ f(x) \neq 0 \tag{44.19}$$

(3) ($\mathcal{A}$ separates points)

$$\forall x, y \in \mathcal{A}\colon\ x \neq y \Rightarrow\ \exists f \in \mathcal{A}\colon f(x) \neq f(y). \tag{44.20}$$

Then $\mathcal{A}$ is dense in $C(X, \mathbb{R})$.

As the statement perhaps suggest, the proof is largely algebraic. (Indeed, the title of Stone's paper is "Applications of the Theory of Boolean Rings to General Topology.") The motivation for conditions (1-3) is that these are satisfied by $C(X, \mathbb{R})$ itself:

Lemma 44.5 *For any metric space X, the set $C(X, \mathbb{R})$ (or even $C(X, \mathbb{C})$) satisfies (1-3) above.*

Proof. That $C(X, \mathbb{R})$ follows from that fact that $\mathbb{R}$ is an algebra and the operations in (44.18) are taken pointwise. Writing ρ for the metric in X, the function $f(z) := 1 - \rho(x, z)$ is continuous, does not vanish at x and takes different values at x and y. □

Next we observe:

Lemma 44.6 *Let X be a compact metric space. If $\mathcal{A} \subseteq C(X, \mathbb{R})$ obey (1-3) in Theorem 44.4. Then so does its closure $\overline{\mathcal{A}}$ of $\mathcal{A}$ in $C(X, \mathbb{R})$.*

Proof. Let $f, g \in \overline{\mathcal{A}}$ and $\lambda \in \mathbb{R}$. For each $\epsilon > 0$ there exists $f_\epsilon, g_\epsilon \in \mathcal{A}$ such that $\|f - f_\epsilon\| < \epsilon$ and $\|g - g_\epsilon\| < \epsilon$. But then the triangle inequality shows $\|(f+g) - (f_\epsilon + g_\epsilon)\| < 2\epsilon$ and, since this holds for all $\epsilon > 0$, we get $f + g \in \overline{\mathcal{A}}$ as desired. For the remaining operations we note $\|\lambda f - \lambda f_\epsilon\| < |\lambda|\epsilon$ and $\|f \cdot g - f_\epsilon \cdot g_\epsilon\| < \epsilon(\|f\| + \|g_\epsilon\|)$ and observe that $\|g_\epsilon\| \leqslant \epsilon + \|g\|$. This shows that $\overline{A}$ obeys (1). Since (2-3) hold for $\mathcal{A}$ and $\mathcal{A} \subseteq \overline{\mathcal{A}}$, they hold for $\overline{\mathcal{A}}$ as well. □

The algebraic content of the assumption becomes apparent from:

Lemma 44.7 *Suppose $\mathcal{A}$ is a collection of real-valued functions on a set X such that (1-3) in Theorem 44.4 hold. Then*

$$\forall x, y \in X\, \forall f \in \mathbb{R}^X\, \exists h \in \mathcal{A}\colon\ h(x) = f(x)\ \wedge\ h(y) = f(y) \tag{44.21}$$

In short, at any two given points, every function $f\colon X \to \mathbb{R}$ coincides with some $h \in \mathcal{A}$.

Proof. Fix $f \in \mathbb{R}^X$. Supposing first $x = y$, here we use (2) to find $g \in \mathcal{A}$ such that $g(x) \neq 0$ and, using (1), set $h(z) := \frac{f(x)}{g(x)} g(z)$ to get $h \in \mathcal{A}$ with $h(x) = f(x)$. If in turn $x \neq y$ we use (2-3) to find $\alpha, \beta, \gamma \in \mathcal{A}$ such

$$\alpha(x) \neq \alpha(y) \ \wedge \ \beta(x) \neq 0 \ \wedge \ \gamma(y) \neq 0. \tag{44.22}$$

Note that then $z \mapsto \beta(z)[\alpha(z) - \alpha(x)]$ and $z \mapsto \gamma(z)[\alpha(z) - \alpha(y)]$ belong to $\mathcal{A}$. (Since we do not know that constants belong to $\mathcal{A}$, this requires writing each of these as a difference of two functions that are demonstrably in $\mathcal{A}$.) Now use (1) to check that

$$h(z) := \frac{\beta(z)}{\beta(x)} \frac{\alpha(z) - \alpha(y)}{\alpha(x) - \alpha(y)} f(x) + \frac{\gamma(z)}{\gamma(y)} \frac{\alpha(z) - \alpha(x)}{\alpha(y) - \alpha(x)} f(y) \tag{44.23}$$

also lies in $\mathcal{A}$ and coincides with f at x and y. □

The proof of course has an analytic part that lies at the heart of:

Lemma 44.8 *Assuming X to be a compact metric space, suppose $\mathcal{A} \subseteq C(X, \mathbb{R})$ obeys (1-3) in Theorem 44.4. Then*

$$\forall f, g \in \overline{\mathcal{A}}\colon \ \max\{f, g\} \in \overline{\mathcal{A}} \wedge \min\{f, g\} \in \overline{\mathcal{A}} \tag{44.24}$$

where $\max\{f, g\}(x) := \max\{f(x), g(x)\}$ *and* $\min\{f, g\}(x) = \min\{f(x), g(x)\}$.

Proof. For the proof it actually suffices to show that

$$\forall f \in \overline{\mathcal{A}}\colon \ |f| \in \overline{\mathcal{A}} \tag{44.25}$$

Indeed, we then get (44.24) by noting that

$$\max\{f, g\} = \frac{1}{2}\bigl(f + g + |f - g|\bigr) \wedge \max\{f, g\} = \frac{1}{2}\bigl(f + g - |f - g|\bigr) \tag{44.26}$$

along with the fact that $\overline{\mathcal{A}}$ obeys (1). (We are using Lemma 44.6 throughout.)

As to (44.25), let $f \in \overline{\mathcal{A}}$ and note that f is, being a continuous function on a compact space, bounded. This means $\exists a > 0\colon \operatorname{Ran}(f) \subseteq [a, a]$. By the Weierstrass approximation theorem, there exists a sequence $\{P_n\}_{n \in \mathbb{N}}$ of polynomials in z such that

$$\sup_{z \in [-a, a]} \bigl| |z| - P_n(z) \bigr| \underset{n \to \infty}{\longrightarrow} 0 \tag{44.27}$$

Using $\|\cdot\|$ for the supremum norm, it then follows that also $\| |f| - P_n \circ f \| \to 0$ as $n \to \infty$. But $P_n = P_n(0) + Q_n$ where Q_n is a polynomial without a constant term and $P_n(0) \to 0$ by $z = 0$ case in (44.27). Hence $\| |f| - Q_n \circ f \| \to 0$ as well. Since $Q_n \circ f \in \overline{\mathcal{A}}$ by (1) and $\overline{\mathcal{A}}$ is closed in $C(X, \mathbb{R})$, we conclude that $|f| \in \overline{\mathcal{A}}$. (The reason why we had to remover $P_n(0)$ from P_n is that we do not know whether $\overline{\mathcal{A}}$ contains constants.) □

Remark 44.9 The reference to the Weierstrass approximation theorem may be considered distracting and perhaps even conceptually wrong as Theorem 44.4 is often used to infer Theorem 44.1 as a corollary. And, indeed, this reference can be avoided by proving (44.27) directly. One way is to define $\{P_n\}_{n \in \mathbb{N}}$ recursively by

$$P_0(x) := 0 \ \wedge \ \forall n \in \mathbb{N}\colon P_{n+1}(x) := P_n(x) + \frac{x^2 - P_n(x)^2}{2} \tag{44.28}$$

and noting that $P_n \to |\cdot|$ uniformly on $[-1,1]$. (One then needs to scale the argument by a to get (44.27).) Another is to observe that, for all $|x| < 1$ and $\alpha \in \mathbb{R}$,

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \quad \text{where} \quad \binom{\alpha}{k} := \frac{1}{k!} \prod_{i=0}^{k-1} (\alpha - i). \tag{44.29}$$

(This is so called *Newton's binomial series* although it is nothing more than the Taylor series of $x \mapsto (1+x)^\alpha$ at $x = 0$.) For $\alpha := 1/2$ and $x := z^2 - 1$, the left-hand side becomes $|z|$ while, as $x^k = (-1)^k |1 - z^2|$, the right-hand side is an alternating series that, as it turns out, converges uniformly in $z \in [-1,1]$. The partial sums then give the desired polynomials approximating $z \mapsto |z|$ uniformly.

We are finally ready to give:

Proof of Theorem 44.4. Let $\mathcal{A}$ be as in the statement and let $\overline{\mathcal{A}}$ be its closure in $C(X,\mathbb{R})$. Let $f \in C(X,\mathbb{R})$. By Lemma 44.6, $\overline{\mathcal{A}}$ obeys conditions (1-3) and so for each $x, y \in X$ there exists a function $h_{x,y} \in \overline{\mathcal{A}}$ such that $h_{xy}(x) = f(x)$ and $h_{xy}(y) = f(y)$. (This requires a choice although that seems avoidable thanks to the compactness of X.) Fix $\epsilon > 0$ and $x \in X$. The continuity of f and h_{xy} at y and the fact that $f(y) - h_{xy}(y) = 0$ ensure

$$\forall y \in X\, \exists \delta > 0\, \forall z \in B_X(y,\delta)\colon\; f(z) - h_{xy}(z) > -\epsilon \tag{44.30}$$

and so we may define

$$\delta(y) := \frac{1}{2} \sup\Bigl\{\delta \in (0,1)\colon \forall z \in B_X(y,\delta)\colon\; f(z) - h_{xy}(z) > -\epsilon\Bigr\} \tag{44.31}$$

and observe that $f(z) - h_{xy}(z) > -\epsilon$ for $z \in B_X(y, \delta(y))$. The balls $\{B_X(y,\delta(y))\colon y \in X\}$ form an open cover of X and so there exists $n \in \mathbb{N}$ and $y_0, \dots, y_n \in X$ such that $X = \bigcup_{i=0}^n B_X(y_i, \delta(y_i))$. Thanks to Lemmas 44.8, the function $g_x\colon X \to \mathbb{R}$ defined by

$$g_x(z) := \max\{h_{xy_0}, \dots, h_{xy_n}\} \tag{44.32}$$

then obeys

$$\forall x \in X\colon\;\; g_x \in \overline{\mathcal{A}} \,\wedge\, g_x(x) = x \,\wedge\, \forall z \in X\colon g_x(z) > f(z) - \epsilon \tag{44.33}$$

where the last part follows from $g_x(z) \geqslant h_{xy_i}(z)$ whenever $z \in B_X(y_i, \delta(y_i))$.

We now repeat the same argument with the functions $\{g_x\colon x \in X\}$. Indeed, for all $x \in X$, the function $f - g_x$ vanishes at x and is continuous there. Hence

$$\delta'(x) := \frac{1}{2} \sup\Bigl\{\delta \in (0,1)\colon \forall z \in B_X(x,\delta)\colon\; f(z) - g_z(z) < \epsilon\Bigr\} \tag{44.34}$$

lies in $(0,1)$ and obeys $f(z) - g_x(z) < \epsilon$ whenever $z \in B_X(x, \delta'(x))$. Since the open balls $\{B_X(x,\delta'(x))\colon x \in X\}$ cover X which is compact, there exist $m \in \mathbb{N}$ and $x_0, \dots, x_m \in X$ such that $X = \bigcup_{i=0}^m B_X(x_i, \delta'(x_i))$. Define $g\colon X \to \mathbb{R}$ by

$$g(z) := \min\{g_{x_0}, \dots, g_{x_m}\} \tag{44.35}$$

Lemmas 44.8 then shows

$$g \in \overline{\mathcal{A}} \,\wedge\, \forall z \in X\colon\; -\epsilon < g(z) - f(z) < \epsilon \tag{44.36}$$

where the left inequality follows from the inequality in (44.33) while the right inequality uses that $g_z(z) < f(z) - \epsilon$ whenever $z \in B_X(x, \delta'(x))$. We thus have $\|f - g\| \leqslant \epsilon$. As this holds for all $\epsilon > 0$ and $\overline{\mathcal{A}}$ is closed, we conclude $f \in \overline{\mathcal{A}}$ as desired. $\square$

44.3 Applications and extensions.

The above approximation theorems have some standard applications that we will discuss next. First, Theorem 44.4 indeed subsumes Theorem 44.1 because the polynomials on any bounded intereval $[a,b]$ obey condition (1-3) there. A typical situation where Theorem 44.4 provides a considerable benefit is the subject of:

Lemma 44.10 *Let X and Y be compact metric spaces and let $X \times Y$ be endowed with a product metric that makes it compact. Then*

$$\left\{ \sum_{i,j=1}^{n} f_i \otimes g_j \colon f_0,\dots,f_n \in C(X,\mathbb{R}) \,\wedge\, g_0,\dots,g_n \in C(Y,\mathbb{R}) \right\} \tag{44.37}$$

where $f \otimes g(x,y) := f(x)g(y)$, is dense in $C(X \times Y, \mathbb{R})$.

Proof (idea). With the help of Lemma 44.5, we readily check that the collection (44.37) satisfies (1-3) of Theorem 44.4. □

Theorem 44.4 also allows us to work with other classes of approximating functions than polynomials. A standard example is:

Lemma 44.11 *The linear vector space*

$$\left\{ x \mapsto \sum_{k=0}^{n} \Big(a_k \sin(2\pi k x) + b_k \cos(2\pi k x) \Big) \colon n \in \mathbb{N} \,\wedge\, a_0,\dots,a_n,b_0,\dots,b_n \in \mathbb{R} \right\} \tag{44.38}$$

is dense in

$$\Big\{ f \in C\big([0,1],\mathbb{R}\big) \colon f(0) = f(1) \Big\}. \tag{44.39}$$

Proof (idea). We have to show that the set (44.38) satisfies (1-3) of Theorem 44.4. In order to show that this is an algebra, we have to show that the product of any two sines or cosines appearing in (44.38) can be written as a single sine or cosine. This is based on the identities of the form $\sin(\alpha)\cos(\beta) = \frac{1}{2}[\sin(\alpha+\beta) + \sin(\alpha-\beta)]$ etc.

Since $\sin(x)^2 + \cos(x)^2 = 1$, the collection (44.38) does not vanish at a point. A minor hurdle comes in the verification of (3). It is easy to check that if $\sin(2\pi x) = \sin(2\pi y)$ and $\cos(2\pi x) = \cos(2\pi y)$, then $y - x \in 2\pi\mathbb{Z}$. So the functions separate all pairs of points except for $x = 0$ and $y = 1$. And, no surprise, all of the functions in (44.38) take equal values at the endpoints of $[0,1]$. In order to apply Theorem 44.4, we thus have to take $X := [0,1)$ with $\rho(x,y) := \min\{|x-y|, |1-x+y|\}$. (This corresponds to wrapping X into a circle of unit circumference; the ρ-distance is then the shorter of the two arcs connecting x and y.) Then (X,ρ) is compact and (44.38) separates points. □

The trick used in the proof of Lemma 44.11 shows that the separation of points is not a very serious restriction: If all functions evaluate to the same value at two distinct points, we can identify these two points into one and still preserve compactness and get continuous functions with the required attributes satisfies. We can even handle any finite number of pairs of points this way (an infinite number of pairs might run into issues with compactness).

The condition that the functions do not all vanish at a point is harder to overcome and Theorem 44.4 is sometimes stated without it but with the conclusion modified to "either $\mathcal{A}$ is dense in $C(X,\mathbb{R})$ or there exists $x \in X$ such that $\forall f \in \mathcal{A}\colon f(x) = 0$" which, in the second alternative, is hardly illuminating. However, an easy modification gives:

Theorem 44.12 *Let X be a compact metric space and let $x_0 \in X$. Suppose*

$$\mathcal{A} \subseteq \{f \in C(X,\mathbb{R})\colon f(x_0) = 0\} \tag{44.40}$$

obeys (1) and (3) of Theorem 44.4 as stated and obeys (2) for all $x \neq x_0$. Then

$$\mathcal{A} \text{ is dense in } \{f \in C(X,\mathbb{R})\colon f(x_0) = 0\} \tag{44.41}$$

Proof. Let

$$\mathcal{A}' := \{c1 + f\colon c \in \mathbb{R} \wedge f \in \mathcal{A}\} \tag{44.42}$$

Since function 1 is a unity under pointwise multiplication, $\mathcal{A}'$ is an algebra; i.e., (1) holds for $\mathcal{A}'$. As $\mathcal{A} \subseteq \mathcal{A}'$, so does (3). As to (2), once $c \neq 0$ we have $(c1 + f)(x_0) \neq 0$ and so $\mathcal{A}'$ also obeys (2) on all of X. By Theorem 44.4, $\mathcal{A}'$ is dense in $C(X,\mathbb{R})$. But this means that for each $f \in C(X,\mathbb{R})$ with $f(x_0) = 0$ and each $\epsilon > 0$ there is $c \in \mathbb{R}$ and $h \in \mathcal{A}$ such that $\|f - (c1 + h)\| < \epsilon$. Invoking this at at x_0 where $f(x_0) = 0$ and $h(x_0) = 0$ shows $|c| < \epsilon$. It follows that $\|f - h\| < 2\epsilon$, proving the claim. □

Note that if $\mathcal{A}$ vanishes at more than one point, it also fails to separate points. So, in this case we first need to identify all the points where all the functions in $\mathcal{A}$ vanishe into one and then apply the above theorem. Using these ideas we can show:

Lemma 44.13 *The linear vector space*

$$\left\{x \mapsto \sum_{k=1}^{n} a_k \sin(\pi k x)\colon n \in \mathbb{N} \smallsetminus \{0\} \wedge a_1, \dots, a_n \in \mathbb{R}\right\} \tag{44.43}$$

is dense in

$$\left\{f \in C\big([0,1],\mathbb{R}\big)\colon f(0) = 0 = f(1)\right\}. \tag{44.44}$$

We leave a proof of this lemma to homework. The key idea is to show that, for all $m, n \in \mathbb{N}$ such that $m - n$ is even there exist $\{c_k\}_{k\in\mathbb{N}}$ with $\sum_{k=1}^{\infty} |c_k| < \infty$ such that

$$\forall x \in [0,1]\colon\quad \cos(\pi n x) - \cos(\pi m x) = \sum_{k=1}^{\infty} c_k \sin(\pi k x) \tag{44.45}$$

which is needed to conclude that the closure of (44.43) in (44.44) is an algebra.

Notably, there is even a version of Stone-Weierstrass theorem that does not require X to be compact but rather only locally compact. One then has to curb the space of continuous functions to only those that "vanish at infinity." We refer the reader to standard analysis textbooks where this is treated in detail.

Another interesting extension is that to the complex valued functions. Unlike many other theorems in analysis, this extension is not automatic and, in fact, as shown in a homework exercise, the result is FALSE under the conditions (1-3) alone. This is not surprising as ordering of the reals has been used heavily in the proof. A version that works for complex-valued functions is as follows:

Theorem 44.14 (complex-valued Stone-Weierstrass theorem) *For X a compact metric space, let $\mathcal{A} \subseteq C(X,\mathbb{C})$ be a class of functions such that:*

(1) ($\mathcal{A}$ is an algebra)

$$\forall f,g \in \mathcal{A}\ \forall \lambda \in \mathbb{C}\colon\ f+g \in \mathcal{A}\ \wedge\ \lambda f \in \mathcal{A}\ \wedge\ f\cdot g \in \mathcal{A} \tag{44.46}$$

(2) ($\mathcal{A}$ does not vanish at a point)

$$\forall x \in X\, \exists f \in \mathcal{A}\colon\ f(x) \neq 0 \tag{44.47}$$

(3) ($\mathcal{A}$ separates points)

$$\forall x,y \in \mathcal{A}\colon\ x \neq y \Rightarrow\ \exists f \in \mathcal{A}\colon f(x) \neq f(y). \tag{44.48}$$

(4) ($\mathcal{A}$ is self-adjoint)

$$\forall f \in \mathcal{A}\colon \bar{f} \in \mathcal{A} \tag{44.49}$$

where $\bar{f}$ is the complex conjugate of f.

Then $\mathcal{A}$ is dense in $C(X,\mathbb{C})$.

Proof. Let $\mathcal{A}' := \mathcal{A} \cap C(X,\mathbb{R})$ be the real valued functions in $\mathcal{A}$. Since $f,\bar{f} \in \mathcal{A}$ implies $f+\bar{f}, \mathrm{i}(f-\bar{f}) \in \mathcal{A}'$, we readily check that $\mathcal{A}'$ satisfies (1-3) of Theorem 44.4. Hence $\mathcal{A}'$ is dense in $C(X,\mathbb{R})$. But each $f \in C(X,\mathbb{R})$ admits a representation $f = g + \mathrm{i}h$ for some $g,h \in C(X,\mathbb{R})$ and so $\mathcal{A} = \{g+\mathrm{i}h\colon g,h \in \mathcal{A}'\}$ is thus dense in $C(X,\mathbb{C})$. □

Applications of the complex-valued Stone-Weierstrass theorem will be given when we discuss Fourier series.