DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

A STONE-WEIERSTRASS THEOREM WITHOUT CLOSURE UNDER SUPREMA

R. Preston McAfee
California Institute of Technology
University of Western Ontario

Philip J. Reny
University of Western Ontario

SOCIAL SCIENCE WORKING PAPER 727

April 1990

A STONE-WEIERSTRASS THEOREM WITHOUT CLOSURE UNDER SUPREMA

R. Preston McAfee and Philip J. Reny

ABSTRACT

For a compact metric space X, consider a linear subspace A of $C(X)$ containing the constant functions. One version of the Stone-Weierstrass theorem states that, if A separates points, then the closure of A under both minima and maxima is dense in $C(X)$. Similarly, by the Hahn-Banach theorem, if A separates probability measures, A is dense in $C(X)$. We show that if A separates points from probability measures, then the closure of A under minima is dense in $C(X)$. This theorem has applications in Economic Theory.

A STONE-WEIERSTRASS THEOREM WITHOUT CLOSURE UNDER SUPREMA

R. Preston McAfee and Philip J. Reny*
California Institute of Technology
Pasadena, CA 91125

The classical Stone-Weierstrass theorem states that, if a linear space A of real valued functions defined on a compact metric space X contains the constant functions, is closed under minima and maxima, and separates points, then A is dense in $C(X)$. The purpose of this paper is to provide an alternative structure for sets closed under minima alone, which generates the same result.

The theorem fits between the Stone-Weierstrass theorem and a collorary to the Hahn-Banach theorem. Let X be a compact metric space, with metric ρ, and Δ the set of probability distributions (regular unitary measures) on X. Let δ_x represent the point mass measures:

$$\delta_x(E) = \begin{cases} 1, & \text{if } x \in E \\ 0, & \text{if } x \notin E. \end{cases}$$

For $A \subseteq C(X)$, define the closure under minima and maxima:

$$A_m = \{f : f(x) = \min_{1 \le i \le n} f_i(x), f_i \in A, n \in \mathbb{N}\},$$

$$A_M = \{f : f(x) = \max_{1 \le i \le n} f_i(x), f_i \in A, n \in \mathbb{N}\}.$$

As usual, **1** denotes the constant function one, and $\bar{A}$ the closure of A in supnorm.

Definition 1: *A linear subspace of C(X) containing* **1** *is said to separate points if, for x and y in X,*

$$\int f\, d\delta x = \int f\, d\delta y \quad \text{for } all\ f\ in\ A\ implies\ x = y, \tag{1}$$

and to separate probability distributions if, for $\mu, \nu \in \Delta$,

$$\int f\, d\mu = \int f\, d\nu \quad \text{for } all\ f\ in\ A\ implies\ \mu = \nu,$$

---

* California Institute of Technology and University of Western Ontario. Correspondence to Prof. R.P. McAfee, 228-77 Caltech, Pasadena, CA 91125. The authors gratefully acknowledge the assistance of Charalambos Aliprantis in the preparation of this paper.

and to separate points from probability distributions if, for μ *in* Δ, $x \in X$.

$$\int f \, d\mu = \int f \, d\delta_x, \text{ for } all \ f \ in \ A \ implies \ \mu = \delta_x. \qquad (2)$$

One statement of the Stone-Weierstrass Theorem is

Theorem 2 (Stone-Weierstrass): *If* A *is a linear subspace of* $C(X)$, $1 \in A$ *then* A *separates points if and only if* $\overline{(A_m)_M} = C(\mathbf{X})$.

Condition (1) is equivalent to the more standard definition of separating points, namely that $f(x)=f(y)$ for all $f \in A$ implies $x=y$, and is stated in the somewhat cumbersome way for comparability to two subsequent results. Note that $(A_m)_M$ is a linear space closed under maxima and minima.

A well known corollary[1] to the Hahn-Banach Theorem and Riesz Representation Theorem has a similar flavor to Theorem 2.

Theorem 3 (Corollary to Hahn-Banach): *If* A *is a linear subspace of* $C(X)$, $1 \in A$, *then* A *separates probability distributions if and only if* $\overline{A} = C(X)$.

Thus, one consolidated view of these results is that, if we are given $A \subseteq C(X)$, with $1 \in A$, then A is dense if it separates probability distributions from probability distributions, or if it is closed under minima and maxima and separates points. In the next section, we prove the following intermediate result.

Theorem 4: *If* A *is a linear subspace of* $C(X)$, $1 \in A$, *then* A *separates points from probability distributions if and only if* $\overline{A_m} = \overline{A_M} = C(X)$.

This is an intermediate result in the sense that we can eliminate the closure under maxima in the Stone-Weierstrass theorem if A separates points from probability distributions, and eliminate closure under both maxima and minima if A separates probability distributions from probability distributions. Thus, separating points from probability distributions substitutes for the ability to take maxima in the Stone-Weierstrass theorem.

Consider for example, the set of quadratics on [0,1]:

$$A = \{\alpha_o + \alpha_1 x + \alpha_2 x^2 : (\alpha_o, \alpha_1, \alpha_2) \in \mathbb{R}^3\}.$$

Clearly $\overline{A} = A \neq C[0,1]$. However, A separates points from probability distributions. That is, for if $\mu \in \Delta$, $\mu \neq \delta_y$, then:

$$\int (x-y)^2 \, d\delta_y(x) = 0 < \int (x-y)^2 \, d\mu(x).$$

---

1. See, for example, Friedman (1970), Corollary 4.8.7, p. 153, and note that the norm dual of $C(X)$ is the set of regular signed measures. Since $1 \in A$, $\int f(d\mu^+ - d\mu^-) = 0$ allows $\mu^+, \mu^- \in \Delta$ without loss of generality, where $\mu = \mu^+ - \mu^-$ is the Jordan decomposition of μ.

Therefore, according to Theorem 4, $\overline{A_m} = \overline{A_M} = C[0,1]$. Thus, the present Theorem is not a consequence of the Hahn-Banach Corollary. Similarly, if A is the set of linear functions an $[0,1]$,

$$A = \{\alpha_o + \alpha_1 x : (\alpha_o, \alpha_1) \in \mathbb{R}^2\},$$

A separates points, so $\overline{(A_m)_M} = C[0,1]$. Incidentally, $(A_m)_M$ is the subspace of all piecewise linear functions on $[0,1]$. However, A fails to separate points from probability distributions, and $\overline{A_m}$ is the set of convex functions. This example distinguishes the present Theorem from Stone-Weierstrass.

Problems for which only minima or maxima, but not both, may be taken arise in a natural way in economic theory. Suppose the value v of an object for sale (e.g. an oil lease) is correlated to an observable s (for example, the results of a sample drilling). Let $f(s / v)$ be the density of s, given v. Suppose the potential buyer, but not the seller, knows v. Can the seller on average charge the potential buyer his value v? This reduces to solving the equation

$$v = \int_S z(s) f(s / v) ds ,$$

where $z(s)$ is the price charged when the outcome s arises. Assume s is a draw from a compact metric space S.

If the seller offers the buyer a set $\{z_1, \ldots, z_n\}$ of charges, and lets the buyer choose the charge he likes best (i.e. which minimizes the expected charge) the seller will earn, on average,

$$p(v) = \min_{1 \le i \le n} \int_S z_i(s) f(s / v) ds,$$

assuming the buyer agrees to buy, that is, the minimum expected charge $p(v)$ is less than v. If

$$R = \{\int z(s) f(s / \cdot) ds : z \in C(S)\},$$

then the seller can charge the buyer his value (on average) if the identity is in R_m. Obviously, the seller can get arbitrarily close if $\overline{R_m} = C[0, \bar{v}]$, where values fall in $[0, \bar{v}]$. Note that $\mathbf{1} \in R$ since $f(\cdot / v)$ is a density. This problem, and others like it, are explored in [4] and [5]. We shall return to a special case of this class of problems in the final section.

Proof of Theorem 4

For this section, X is a compact metric space with metric ρ, A is a linear subspace of $C(X)$, and $\mathbf{1} \in A$.

Definition 5: Let $\varepsilon > 0$, $\delta > 0$. *A positive continuous function f is a* ***nearly u-shaped function*** at y *of order* (ε, δ) *iff $f(y) \le \varepsilon$ and* $\rho(x, y) > \delta$ *implies* $f(x) \ge 1$.

The set of nearly u-shaped functions at y of order (ε, δ) is denoted $U(y, \varepsilon, \delta)$.
We shall make use of three obvious properties of the sets $U(y, \varepsilon, \delta)$.

$$0<\varepsilon \le \varepsilon_o, 0<\delta \le \delta_o \Rightarrow U(y,\varepsilon,\delta) \subseteq U(y,\varepsilon_o,\delta_o), \tag{2}$$

$$\text{each } U(y,\varepsilon,\delta) \text{ is convex, and} \tag{3}$$

$$\text{each } U(y,\varepsilon,\delta) \text{ has nonempty interior.} \tag{4}$$

The last fact follows from the observation that the $\varepsilon / 4$ ball around $\varepsilon / 2 + \rho(\cdot, y) / \delta$ is contained in $U(y, \varepsilon, \delta)$. The following lemma shows that $\overline{A_m} = C(X)$ if and only if A contains u-shaped functions at every $x \in X$ of all orders (ε, δ). This lemma is critical to the proof of the theorem.
Lemma 6: *Suppose $A \subseteq C(X)$ is a linear subspace, $\mathbf{1} \in A$. Then $\overline{A_m} = C(X)$ if and only if for all y in X and all $\varepsilon,\delta>0$, $U(y,\varepsilon,\delta) \cap A \neq \emptyset$.*

Proof: ($\Rightarrow$) Fix $y \in X$, $\varepsilon > 0$, and $\delta > 0$. Since $\overline{A_m} = C(X)$, there are $f_i, \ldots, f_n \in A$ so that

$$\left| \min_{1 \le i \le n} f_i(x) - (\varepsilon / 2 + \rho(x,y) / \delta) \right| < \varepsilon / 2, \ \forall x \in X.$$

Thus, there exists $j \in \{1, \ldots, n\}$ with

$$\left| f_j(y) - \varepsilon / 2 \right| < \varepsilon / 2.$$

From

$$f_j(x) \ge \min_{1 \le i \le n} f_i(x) \ge \rho(x,y) / \delta,$$

we easily infer $f_j \in U(y,\varepsilon,\delta)$.
($\Leftarrow$) Fix $f \in C(X)$, and $\varepsilon > 0$. Define

$$\alpha = \max_{x \in X} f(x) - \min_{x \in X} f(x).$$

If $\alpha = 0$, we're done, since $\mathbf{1} \in A$. So suppose $\alpha > 0$. Since f is continuous, there is a $\beta>0$ so that

$$\rho(x,y) < \beta \Rightarrow \left| f(y) - f(x) \right| < \varepsilon / 2.$$

For each $y \in X$, choose $g \in A \cap U(y, \varepsilon / 3\alpha, \beta)$, and define

$$h = \alpha g + (f(y) + \varepsilon / 2)\mathbf{1} \in A.$$

Note

$$| h(y) - f(y) | = \alpha g(y) + \varepsilon / 2 \le \alpha(\varepsilon / 3\alpha) + \varepsilon / 2 < \varepsilon.$$

For $\rho(x, y) < \beta$,

$$h(x) - f(x) = \alpha g(x) + f(y) - f(x) + \varepsilon / 2 \ge f(y) - f(x) + \varepsilon / 2 \ge 0.$$

For $\rho(x, y) \ge \beta$, we have

$$h(x) - f(x) = \alpha g(x) + f(y) - f(x) + \varepsilon / 2 \ge \alpha + f(y) - f(x) + \varepsilon/2 \ge 0,$$

by the definition of α.

Thus $h(x) \ge f(x)$ and $h(y) < f(y) + \varepsilon$. Now define the set (recall that h depends on y):

$$S(y) = \{x : h(x) < f(x) + \varepsilon\}.$$

Clearly, $\{S(y) : y \in X\}$ forms an open cover of X, since $y \in S(y)$. Because X is compact, there is a finite subcover $S(x_1), \ldots, S(x_n)$, with associated functions $h_1, \cdots, h_n$.
By construction, $0 \le \min_{1 \le i \le n} h_i(x) - f(x) < \varepsilon$ for all $x \in X$, and thus $f \in \overline{A_m}$ as desired. QED

Remark: The nearly u-shaped functions permit approximation from above, in the sense that the lower envelope, produced by minima, approximates any function. This occurs because u-shaped functions take minima near a chosen point y, and then rise sufficiently rapidly away from y.

Theorem 4: *Suppose A is a linear subspace of $C(X)$, where X is a compact metric space, and $1 \in A$. Then $\overline{A_m} = C(X)$ if and only if A separates points from probability distributions.*

Proof: ($\Rightarrow$) Suppose $\mu \in \Delta, \mu \ne \delta_y$. Then there exists $\delta > 0$ so that

$$\int_{N_\delta(y)} d\mu(x) < 1,$$

where $N_\delta(y) = \{x \,/\, \rho(x, y) < \delta\}$. Let $\varepsilon < 1 - \int_{N_\delta(y)} d\mu(x)$, and choose $f \in U(y, \varepsilon, \delta) \cap A$. Such a function exists by Lemma 6. Then

$$\int f \, d\mu = \int_{N_\delta(y)} f d\mu + \int_{X \setminus N_\delta(y)} f d\mu \ge \int_{X \setminus N_\delta(y)} f d\mu \ge \int_{X \setminus N_\delta(y)} d\mu$$

$$= 1 - \int_{N_\delta(y)} d\mu > \varepsilon \ge f(y),$$

and so $f \in A$ and f separates y from μ, as desired.
($\Leftarrow$) Suppose by way of contradiction that $\overline{A_m} \neq C(X)$. By Lemma 6, there exists y, ε_o>0, and δ_o>0 so that $U(y, \varepsilon_o, \delta_o) \cap A = \emptyset$. Since A is linear, and hence convex and $U(y, \varepsilon_o, \delta_o)$ is convex, with nonempty interior, there is a separating functional [2]. Thus, there is a signed measure μ, and a constant c with :

$$\text{for all } g \in A \text{ and all } f \in U(y, \varepsilon_o, \delta_o), \text{ we have } \int g\,d\mu \leq c \leq \int f\,d\mu. \tag{6}$$

Since A is a linear space, we can assume c=0 and $\int g\,d\mu$=0 for all g in A.

Let $\mu = \mu^+ - \mu^-$ be the Jordan Decomposition of μ (see [6, pp. 235-6]), with associated sets S^+ and S^- satisfying:

$$S^+ \cap S^- = \emptyset \text{ and } \int_{(S^+)^c} d\mu^- = \int_{(S^-)^c} d\mu^+ = 0.$$

Since $\mathbf{1} \in A$, $\int d\,\mu^+ = \int d\mu^-$.

Since both μ^+, and μ^- are finite, we may take $\mu^+, \mu^- \in \Delta$ without loss of generality, by rescaling. Neither μ^+ nor μ^- can be δ_y, for if either is equal to δ_y, (5) contradicts (6). Since μ^- is regular (see [1, Theorem 1.1, p.7]), there is a closed set $\Psi \subseteq S^-$ and $0 < \delta \leq \delta_o$ so that $\Psi \cap N_\delta(y) = \emptyset$ and $\mu^-(\Psi) > 0$. Choose $K > 1 / \mu^-(\Psi) \geq 1$, and define

$$f(x) = \begin{cases} 0, & \text{if } x \in N_\delta(y) \\ K, & \text{if } x \in \Psi \\ 1, & \text{if } x \notin \Psi \cup N_\delta(y) \end{cases}$$

and observe, since $\Psi \cap S^+ = \emptyset$, that

$$\int f\,d\mu = \int f\,d\mu^+ - \int f\,d\mu^- \leq 1 - K\int_\Psi d\mu^- < 0. \tag{7}$$

By [1, Theorem 1.2, p. 8][3] there is a sequence $\{f_n\} \subseteq C(X)$ satisfying

(a) $f_n(x) \geq 1$ for all $x \notin N_\delta(x_o)$;

(b) $0 \leq f_n(x) \leq K$ for all $x \in X$;

(c) $f_n(y) = 0$; and

2. See [2, Theorem 8 in Part I, p. 417]. It is important to note that μ is a regular measure. See [2, Theorem 2, in Part I, p.262].
3. This is a straightforward application of the Tietze Extension Theorem.

(d) $f_n(x) \to f(x)$ for all $x \in X$.

By (a)-(c), $f_n \in U(y, \varepsilon_o, \delta) \subseteq U(y, \varepsilon_o, \delta_o)$. By (b), (d), and (7):

$$\int f_n \, d\mu \to \int f \, d\mu < 0,$$

which contradicts (6) and $c=0$. This completes the proof. Q.E.D.

Remark: $\overline{A_m}$ may be replaced by $\overline{A_M}$ in the statement of the theorem, by noting $A_M = -((-A)_m) = -A_m$, since A is linear. In addition, if A is a convex cone and both **1** and **–1** are in A, lemma 6 and theorem 4 continue to hold, with trivial modifications of the proof.

Conclusion

When the metric space X is an interval [a, b] of the real line, the Stone-Weierstrass theorem has an appealing corollary, namely that if **1** and a strictly increasing function are in A, then $\overline{(A_m)_M} = C[a,b]$. There is an analogous corollary for the present theorem.

Corollary 7 : *Suppose A is a linear subspace of* $C[a,b]$ containing **1** and two functions f and g satisfying:

$$f \textit{ is strictly increasing}, \tag{8}$$

$$\text{and } \frac{g(x)-g(y)}{f(x)-f(y)} \textit{ is strictly increasing in } x \neq y, \text{ for all } y.^4 \tag{9}$$

Then $\overline{A_m} = C[a,b]$.

Proof: Observe that, if $x < y < z$, then

$$\frac{g(x)-g(y)}{f(x)-f(y)} \leq \frac{g(z)-g(y)}{f(z)-f(y)}.$$

Therefore, there is a function (not necessarily continuous) α so that

$$\lim_{x \to y^-} \frac{g(x)-g(y)}{f(x)-f(y)} \leq \alpha(y) \leq \lim_{x \to y^+} \frac{g(x)-g(y)}{f(x)-f(y)}.$$

Moreover, α is strictly increasing, for if $x < y < z$:

$$\alpha(x) < \frac{g(y)-g(x)}{f(y)-f(x)} < \frac{g(y)-g(z)}{f(y)-f(z)} < \alpha(z).$$

Consider $\beta_x(y) = g(y) - \alpha(x) f(y)$, and note β_x is in A, and satisfies

4. The function $(g(x)-g(y))/(f(x)-f(y))$ necessarily has left and right limits as $x \to y$, if (9) holds, for $a<y<b$.

$$\beta_x(y) - \beta_x(x) = (f(y) - f(x))\left[\frac{g(y) - g(x)}{f(y) - f(x)} - \alpha(x)\right] \geq 0,$$

with equality if and only if $y = x$. Thus, if $\nu \neq \delta_x$,

$$\int_a^b \beta_x(y) d\nu(y) > \int_a^b \beta_x(x) d\nu(y) = \beta_x(x).$$

Consequently, (5) is satisfied Q.E.D.

Remark: If f and g are twice differentiable, (8) and (9) reduce to $f' > 0$ and $(g'/f')' > 0$, which are easy to check, as we illustrate in the following example.

Example 7: Suppose a random variable s has cumulative distribution function s^{ν} for $s \in [0,1]$. An economic agent who knows v is to be offered a menu $\{z_i\}$ of payments. This agent chooses the charge with the least expected value:

$$p(v) = \min_{1 \leq i \leq n} \int_o^1 z_i(s) v s^{v-1} ds.$$

Is the set of such charges dense in $C[0,1]$? That is, if the agent's value of an object for sale is $\pi(v)$, is there a menu $\{z_i(s)\}$ that approximately charges the agent his value?

The answer is yes. Consider

$$A = \{f : f(v) = \int_o^1 z(s) v s^{v-1} ds, z \in C[0,1]\}.$$

Note that A contains **1** (using $z = 1$), $f(v) = v / (v+1)$ (for $z(s) = s$)), and $g(v) = v / (v+2)$ (for $z(s) = s^2$). It is easily verified that f and g satisfy the hypotheses of the corollary, so $\overline{A_m} = C[0,1]$.

References

[1] P. Billingsley, *Convergence of Probability Measures*, (New York: John Wiley and Sons), 1968.

[2] N. Dunford and J.T. Schwartz, *Linear Operators, Part I: General Theory*, (New York: John Wiley and Sons), 1988.

[3] A. Friedman, *Foundations of Modern Analysis*, *(New York: Dover Publications); 1970.*

[4] R.P. McAfee, J. McMillan, and P. Reny, "Extracting the Surplus in Common Value Auctions," *Econometrica*, 57, No. 6, November 1989: 1451-1459.

[5] R.P. McAfee and P. Reny, "Correlated Information and Mechanism Design," (University of Western Ontario); forthcoming, *Econometrica*, 1990.

[6] H.L. Royden, *Real Analysis*, Second Edition, (New York: MacMillan Publishing Co.); 1968.