

MORE ABOUT THE STONE-WEIERSTRASS THEOREM THAN YOU PROBABLY WANT TO KNOW

1. Stone's Contribution.

The really new thing about Stone's approach to the approximation theorem was the approach via lattices of continuous functions, although Lebesgue had noticed the importance of approximating the absolute-value function earlier. There is a segment of the mathematical community formed of people who are as likely to encounter a lattice in their work as an algebra and for whom the lattice version of the theorem is at least as important as the algebra version. Folland takes the point of view that algebras are more important than lattices and relegates the heart of Stone's proof to a lemma (4.49), in which he gets maximum mileage from Stone's argument. Since his refinement can be carried out *a posteriori*, I thought it more important to give the basic proof (of density of a lattice with the two-point interpolation property) first.

I deviated from Folland's presentation of the proof in class, so here's the proof I gave. (Some take suprema first and infima second, as Folland does; I did things in the opposite order; it doesn't really matter.)

Recall that a set L of real-valued functions on a set X is called a **lattice** if it is closed under forming (finite) pointwise suprema and infima, *i.e.*, if whenever $f, g \in L$ the functions $f \vee g$ and $f \wedge g$ defined by

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

belong to L .

Theorem [M. H. Stone]: Let X be a compact Hausdorff space (with at least two points) and $L \subseteq \mathcal{C}(X, \mathbb{R})$ a lattice of continuous functions with the property that for each pair of distinct points $x \neq y$ in X and pair of real numbers a, b there exists $g \in L$ for which $g(x) = a$ and $g(y) = b$. Then for every $f \in \mathcal{C}(X, \mathbb{R})$ and $\epsilon > 0$ there is a function $g \in L$ for which $\|f - g\|_\infty < \epsilon$; *i.e.*, the closure of L in the uniform norm is all of $\mathcal{C}(X, \mathbb{R})$.

Proof. Let $f \in \mathcal{C}(X, \mathbb{R})$ and $\epsilon > 0$ be given. For each pair of points $x, y \in X$ we can by hypothesis find $g_{xy} \in L$ for which $g_{xy}(x) = f(x)$ and $g_{xy}(y) = f(y)$ (even including the possibility that $x = y$). For fixed x , the open set $V_{x,y} = \{z \in X : f(z) - \epsilon < g_{xy}(z)\}$ is an open neighborhood of y , so the family of all such sets covers X and by compactness one can find a finite subfamily $\{V_{x,y_1(x)}, \dots, V_{x,y_n(x)}\}$ that covers X (both the y 's and the number of y 's [and therefore V 's] will depend on x , as we have indicated with the subscripts). Let $g_x = g_{x,y_1(x)} \vee \dots \vee g_{x,y_n(x)} \in L$; then $f(z) - \epsilon < g_x(z)$ holds for all $z \in X$ and $g_x(x) = f(x)$ holds for each $x \in X$. Now play the same game from above: for each $x \in X$ the open set $U_x = \{z \in X : g_x(z) < f(z) + \epsilon\}$ is a neighborhood of x , so the family of all such U_x is an open cover of X from which one can extract a finite subcover $\{U_{x_1}, \dots, U_{x_n}\}$. Let $g = g_{x_1} \wedge \dots \wedge g_{x_n} \in L$; then $f(z) - \epsilon < g(z) < f(z) + \epsilon$ holds for all $z \in X$, *i.e.*, $|g(z) - f(z)| < \epsilon$ holds for all $z \in X$, or $\|g - f\|_\infty < \epsilon$, as desired.

The “two-point interpolation” condition on L can be weakened slightly in an obvious way and the proof will still be valid: simply require that for each pair of distinct points $x \neq y$ in X and pair of real numbers a, b there exist for each $\epsilon > 0$ some $g \in L$ for which $|g(x) - a| < \epsilon$ and $|g(y) - b| < \epsilon$ and the proof will still go through (use $\epsilon/2$ instead of ϵ throughout). This fact is occasionally handy in complex-function-theory contexts. The same proof also adapts to a situation in which the uniform closure of L may not be all of $\mathcal{C}(X, \mathbb{R})$. Suppose that $L \subseteq \mathcal{C}(X, \mathbb{R})$ is a lattice that does not necessarily satisfy a “two-point interpolation” condition, but instead make a hypothesis about its relation to the given $f \in \mathcal{C}(X, \mathbb{R})$; namely, that for each pair of distinct points $x \neq y$ in X there exist for each $\epsilon > 0$ some $g \in L$ for which $|g(x) - f(x)| < \epsilon$ and $|g(y) - f(y)| < \epsilon$. The proof will still go through (using $\epsilon/2$ instead of ϵ throughout), and we shall have proved the following corollary, which is essentially Folland’s **Lemma (4.48)**.

Corollary: Let X be a compact Hausdorff space (with at least two points) and $L \subseteq \mathcal{C}(X, \mathbb{R})$ be a lattice of continuous real-valued functions on X . Then in order that $f \in \mathcal{C}(X, \mathbb{R})$ belong to the closure of L in the uniform-norm topology, it is necessary and sufficient that for each pair of distinct points $x \neq y$ in X there exist for each $\epsilon > 0$ some $g \in L$ for which $|g(x) - f(x)| < \epsilon$ and $|g(y) - f(y)| < \epsilon$.

Lattices of continuous functions in a $\mathcal{C}(X, \mathbb{R})$ frequently arise from a **wedge** that is closed under taking (finite) suprema or infima. A subset $W \subseteq \mathcal{C}(X, \mathbb{R})$ is called a **wedge** if it is closed under addition and under multiplication by nonnegative real numbers. An example is furnished by the **convex functions** on a closed bounded interval $X \subseteq \mathbb{R}$ (or, more generally, on a compact convex set $X \subseteq \mathbb{R}^n$), *i.e.*, the functions that for each pair of points $x_0, x_1 \in X$ satisfy $f((1 - \lambda)x_0 + \lambda x_1) \leq (1 - \lambda)f(x_0) + \lambda f(x_1)$ for all $0 \leq \lambda \leq 1$. It is not difficult to show that these functions are necessarily continuous, and easy to show that the set of all such functions is closed under the operations of addition, multiplication by nonnegative scalars, and forming pointwise suprema ($f \vee g$). If W is a wedge of real-valued functions on X (which can now be any set) then $W - W = \{f - g : f, g \in W\}$ is obviously a vector space of functions, and if W was closed under one of the operations \wedge or \vee then $W - W$ is a lattice. For example, if W was closed under \vee then the fact that the order relation on \mathbb{R} is invariant under translation implies that for f ’s and g ’s in W

$$\begin{aligned} (f_1 - g_1) \vee (f_2 - g_2) + (g_1 + g_2) &= (f_1 + g_2) \vee (f_2 + g_1) \\ (f_1 - g_1) \vee (f_2 - g_2) &= (f_1 + g_2) \vee (f_2 + g_1) - (g_1 + g_2) \in W - W \end{aligned}$$

so $W - W$ is closed under \vee , and since it is a vector space (and thus contains the negative of each of its elements) it must also be closed under \wedge , for $f \wedge g = -[(-f) \vee (-g)]$. Looking at the particular examples of the wedges of convex functions on compact convex subsets X of \mathbb{R} or \mathbb{R}^n , we see that Stone’s theorem implies that every continuous real-valued function on X can be uniformly approximated arbitrarily closely by a difference of two convex functions. This fact has the useful consequence that a countably additive signed measure on (the Borel sets of) X is uniquely determined by its integrals against the convex functions; many uniqueness theorems for measures arise in a similar way.

2. Weierstraß, Bernštejn, et al.

These gentlemen were concerned with approximation by polynomials on an interval. If you know a little something about the Poisson kernel on the unit circle, you will find it easy to see that a continuous function of period 2π can be approximated uniformly by a finite trigonometric series or “trigonometric polynomial” of the form $a_0/2 + \sum_{k=1}^n r^k \{a_k \cos kt + b_k \sin kt\}$, with $r < 1$. The Maclaurin series of the trig functions involved converge uniformly on $[-\pi, \pi]$, so such a trig polynomial can itself be uniformly approximated on $[-\pi, \pi]$ by a polynomial in t . Any continuous function on, say, $[0, 1]$ can be extended to be periodic with period 2π , so such a function can be approximated uniformly on $[0, 1]$ by a polynomial, and there you have a proof of the Weierstraß approximation theorem: true, but rather dull.

A completely different and extremely ingenious approach to establishing the density of the polynomials in $\mathcal{C}([0, 1], \mathbb{R})$ was taken by S. N. Bernštejn. If $f \in \mathcal{C}([0, 1], \mathbb{R})$, its **Bernštejn polynomials** are defined by

$$B_n(f)(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$$

for degrees $n = 0, 1, 2, \dots$. These are “weighted averages” of values of f at n equally spaced points on $[0, 1]$, with the weight at k/n for a given $x \in [0, 1]$ being the probability of k heads out of n flips with a coin whose probability of landing heads is x . It turns out, as we shall see, that $\lim_{n \rightarrow \infty} \|f - B_n(f)\|_\infty = 0$, so the Bernštejn polynomials give *explicit, computable* polynomial approximations to a given continuous function. [I must hasten to add that their convergence to f is *extremely* slow; a hint of this is found below in the fact that the ($\|\cdot\|_\infty$ norm of the) error of approximation for the function $f(x) \equiv x^2$ equals $1/(4n)$.] Their convergence to f is plausible for a number of reasons; here are some, followed by a genuine proof.

It is obvious that the mappings $f \rightarrow B_n(f)$ are linear in f , and they are **positive** or **monotone** in the sense that $f \geq 0$ on $[0, 1]$ implies that $B_n(f) \geq 0$ on $[0, 1]$ (since all the coefficients $f(k/n)$ and all the terms $x^k(1-x)^{n-k}$ of $B_n(f)$ will then be nonnegative on $[0, 1]$); more generally, $f \leq g$ on $[0, 1]$ implies that $B_n(f) \leq B_n(g)$ on $[0, 1]$ for all $n \in \mathbb{N}$. The binomial theorem says

$$[(1-x) + xe^t]^n = \sum_{k=0}^n e^{kt} \binom{n}{k} x^k (1-x)^{n-k},$$

for any choice of t , so for $t = 0$ we have $1 = \sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k}$, no surprise from a probabilistic point of view. Differentiating once or twice with respect to t gives

$$\begin{aligned} [(1-x) + xe^t]^{n-1} nxe^t &= \sum_{k=0}^n ke^{kt} \binom{n}{k} x^k (1-x)^{n-k} \\ [(1-x) + xe^t]^{n-1} nxe^t + [(1-x) + xe^t]^{n-2} n(n-1)(xe^t)^2 &= \sum_{k=0}^n k^2 e^{kt} \binom{n}{k} x^k (1-x)^{n-k} \end{aligned}$$

which for $t = 0$ give respectively

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

$$x^2 + \left(\frac{1}{n}\right)x(1-x) = \sum_{k=0}^n \binom{n}{k}^2 x^k (1-x)^{n-k}$$

respectively. Thus $\lim_{n \rightarrow \infty} \|f - B_n(f)\|_\infty = 0$ for the choices $f(x) \equiv 1$, $f(x) \equiv x$ and $f(x) \equiv x^2$, and it is clear that with sufficient computational industry one could show by the same (or similar) elementary arguments that the same thing happens for $f(x) \equiv x^m$ where $m \in \mathbb{N}$ is arbitrary. Thus if one already believes the Weierstraß approximation theorem, one can prove that the Bernštein polynomials converge uniformly to any given $f \in \mathcal{C}([0, 1], \mathbb{R})$: let $\epsilon > 0$ be given, find a polynomial $p(x)$ for which $|f(x) - p(x)| < \epsilon/3$ for all $x \in [0, 1]$, i.e., $p(x) - \epsilon/3 < f(x) < p(x) + \epsilon/3$ for all $x \in [0, 1]$. Then $[B_n(p)](x) - \epsilon/3 < [B_n(f)](x) < [B_n(p)](x) + \epsilon/3$ holds for all $x \in [0, 1]$, i.e., $\|B_n(f) - B_n(p)\|_\infty < \epsilon/3$, so if n is so large that $\|B_n(p) - p\|_\infty < \epsilon/3$, then also $\|B_n(p) - f\|_\infty < 2\epsilon/3$, so $\|B_n(f) - f\|_\infty < \epsilon$.

Bernštein's 1912 proof that his polynomials converged uniformly to f involved probabilistic considerations explicitly (see below). Others—Bohman and Korovkin—observed, however, that if one wants to make a self-contained proof of the Weierstraß approximation theorem, there is already enough material at hand in the few paragraphs above. The reason is that any given $f \in \mathcal{C}([0, 1])$ is the pointwise infimum on $[0, 1]$ of all the parabolas $p(x) = a(x - x_0)^2 + b$ for which $f(x) \leq p(x)$ holds throughout $[0, 1]$. To see this, let $f \in \mathcal{C}([0, 1], \mathbb{R})$, $x_0 \in [0, 1]$ and $\epsilon > 0$ be given; then there is an open interval $U = (x_0 - \delta, x_0 + \delta)$ in which the inequality $(x - x_0)^2 + [f(x_0) + \epsilon] > f(x)$ holds, and for $a \geq 1$ the inequality $a(x - x_0)^2 + [f(x_0) + \epsilon] > f(x)$ will hold there *a fortiori*. Since $(x - x_0)^2 \geq \delta^2$ outside U and f is bounded (above), it is geometrically obvious (details left to any fanatical reader) that if $a \geq 1$ is chosen sufficiently large, then $a(x - x_0)^2 + [f(x_0) + \epsilon] > f(x)$ will hold throughout $[0, 1]$; so we have exhibited a parabola $p(x)$ that majorizes f everywhere on $[0, 1]$ but for which $p(x) = f(x_0) + \epsilon$, and $\epsilon > 0$ was arbitrary.

We can now change “pointwise” to “uniform” and reduce to the infimum of finitely many parabolas, provided that we tolerate some error; that fact essentially constitutes the proof of the Weierstraß theorem. Let $f \in \mathcal{C}([0, 1], \mathbb{R})$ and $\epsilon > 0$ be given, and for each $y \in [0, 1]$ pick a quadratic function (“parabola”) p_y for which $p_y(y) < f(y) + \epsilon$ but for which $f(x) < p_y(x)$ holds for all $x \in [0, 1]$ (the latter inequality can be made strict without damaging the former by adding a small constant to p_y if necessary). Then y has an open neighborhood U_y in which $f(z) < p_y(z) < f(z) + \epsilon$ holds for all $z \in U_y$. The $\{U_y : y \in [0, 1]\}$ form an open cover of the compact interval $[0, 1]$, so one may select finitely many $\{y_1, \dots, y_r\}$ for which the corresponding U_y cover $[0, 1]$. The function $p = p_{y_1} \wedge \dots \wedge p_{y_r}$ will then satisfy both $p(x) < f(x) + \epsilon$ and $f(x) < p(x)$ for every $x \in [0, 1]$.

A result of the construction given in the preceding paragraph is that for every $1 \leq j \leq r$ the inequalities $f(x) < p(x) \leq p_j(x)$ hold for every $x \in [0, 1]$, while every $x \in [0, 1]$ has an open neighborhood U_j (containing one of the points y_j) such that for every $x \in U_j$

the relation $f(x) < p(x) \leq p_j(x) < f(x) + \epsilon$ holds. The positivity of $g \rightarrow B_n(g)$ and the first of these inequalities imply that $B_n(f) \leq B_n(p) \leq B_n(p_j)$ holds throughout $[0, 1]$, but the second relation implies that $p_j(x) < f(x) + \epsilon$ for $x \in U_j$. Since p_j is a polynomial of degree 2, $B_n(p_j) \rightarrow p_j$ uniformly on $[0, 1]$. Consequently there is an $N_j \in \mathbb{N}$ such that $n \geq N_j$ implies $[B_n(p_j)](x) \leq p_j(x) + \epsilon$ uniformly for $x \in [0, 1]$ and *a fortiori* for $x \in U_j$, so $[B_n(p_j)](x) \leq p_j(x) + \epsilon < f(x) + 2\epsilon$ for $x \in U_j$. Thus if $N_+ = \max\{N_1, \dots, N_r\}$ and $n \geq N_+$ we have $[B_n(f)](x) \leq [B_n(p)](x) \leq [B_n(p_j)](x) < f(x) + 2\epsilon$ at every point of $[0, 1]$. Since $f \in \mathcal{C}([0, 1], \mathbb{R})$ was arbitrary, we can similarly find an N_- such that for $n \geq N_-$ we have $[B_n(-f)](x) < -f(x) + 2\epsilon$, *i.e.*, $f(x) - 2\epsilon < [B_n(f)](x)$ at every point of $[0, 1]$, so $n \geq \max\{N_+, N_-\}$ will give $\|B_n(f) - f\|_\infty < 2\epsilon$, and $\epsilon > 0$ is in our hands.

For a more purely probabilistic look at the Bernšteĭn ideas, see Folland, p. 291 ff. Note, however, that the method of proof we just gave can be abstracted to yield a proof of the following

Theorem [Bohman-Korovkin]: Suppose that $\{K_n\}_{n=1}^\infty$ is a sequence of mappings of $\mathcal{C}([0, 1], \mathbb{R})$ into itself satisfying the following conditions:

- (1) Each K_n is linear (on the vector space $\mathcal{C}([0, 1], \mathbb{R})$) ;
- (2) $f \geq 0 \Rightarrow K_n[f] \geq 0$ for all $n \in \mathbb{N}$;
- (3) $\lim_{n \rightarrow \infty} \|f - K_n[f]\|_\infty = 0$ for $f \equiv 1, x, x^2$.

Then $\lim_{n \rightarrow \infty} \|f - K_n(f)\|_\infty = 0$ for every $f \in \mathcal{C}(X, \mathbb{R})$.

However, there are proofs of this theorem (and extensions thereof) that take an entirely different approach: see, *e.g.*, H. H. Schaefer, *Banach Lattices and Positive Operators*, Springer-Verlag 1974, Ch. V, p. 308 ff.

Of course the Bernšteĭn proof is set up on $[0, 1]$, but that is merely a normalization: if $g \in \mathcal{C}([a, b], \mathbb{R})$ then $f = (x \rightarrow f((1-x)a + xb)) \in \mathcal{C}([0, 1], \mathbb{R})$, and if $P(x)$ is a polynomial with $\|P - f\|_{[0,1],\infty} < \epsilon$ then $P^*(t) = P((t-a)/(b-a))$ will be a polynomial in t (of the same degree, incidentally) for which $\|P^* - g\|_{[a,b],\infty} < \epsilon$. In particular, given any (small) $\epsilon > 0$ and (possibly large) $b > 0$ we can find a polynomial $p(x)$ for which $|p(x) - |x|| < \epsilon/2$ holds for all $x \in [-b, b]$. If we think of $p(x) = a_0 + \dots + a_n x^n$ then $|a_0| = |p(0)| < \epsilon/2$, so we may replace $p(x)$ by $P(x) = p(x) - p(0)$, a polynomial with zero constant term, and still have $|P(x) - |x|| < \epsilon$ hold for all $x \in [-b, b]$. This gives Folland's **Lemma (4.47)** without the necessity of making the restriction $|x| \leq 1$, though that is not very important.

3. Stone-Weierstraß for Function Algebras.

H. Lebesgue seems to have been among the first to recognize the importance of finding polynomial approximations for the absolute-value function when one tries to find polynomial approximations to continuous functions in general. We have just given one approach to this approximation problem: find a proof for the Weierstraß approximation theorem that doesn't involve finding approximations to $|x|$ as a lemma. Another approach is to

establish the uniform convergence on $[0, 1]$ of the Maclaurin series for $(1 - t)^{1/2}$ to that function, then write $t = (1 - x^2)$ and $|x| = [1 - (1 - x^2)]^{1/2}$ for $x \in [-1, 1]$ and use partial sums of that Maclaurin series (after substituting $t = 1 - x^2$) as uniform approximants (possibly deleting the constant term, which [as we saw above] can be done without serious damage to the approximation).

Why should the Maclaurin series of $\sqrt{1 - t}$ converge uniformly to the function for $t \in [0, 1]$? Various answers can be given. If one can show that the series converges to the function for $|t| < 1$ and that the series of coefficients is absolutely convergent, one will certainly wrap things up: the series will converge uniformly on $[-1, 1]$, and the endpoint values will necessarily be correct because both $\sqrt{1 - t}$ and the sum of its Maclaurin series will be continuous at the endpoints. From the complex-variables standpoint, it is a triviality that for $|z| < 1$

$$f(z) = (1 + z)^a = \sum_{n=0}^{\infty} \binom{a}{n} z^n \quad \text{where} \quad \binom{a}{n} = \frac{a(a-1)\cdots(a-n+1)}{n!}$$

with the series converging uniformly on compact subsets of $|z| < 1$. (Here $(1 + z)^a$ is defined to be $\exp(a \log(1 + z))$ where the branch of the logarithm that is real on \mathbb{R}^+ is taken. We shall be interested only in real a , and there isn't much sense in considering $a < 0$ because $(1 + z)^a$ will become infinite as $z \rightarrow -1^+$ and so the series of coefficients couldn't possibly converge absolutely.) The absolute value of the ratio of the n -th to the $(n - 1)$ -st coefficient of this series is

$$\left| \frac{a(a-1)\cdots(a-n)(a-n+1)}{n!} \frac{(n-1)!}{a(a-1)\cdots(a-n+2)} \right| = \left| \frac{n-a-1}{n} \right|$$

which for large n and real a equals

$$1 - \frac{a+1}{n} .$$

This is one of those situations where the ratio test fails. All is not lost, however, because the ratio of the n -th to the $(n - 1)$ -st coefficient in the series $\sum_{n=1}^{\infty} 1/n^\alpha$ —which is known to converge for $\alpha > 1$ and to diverge otherwise, by comparison with the integral $\int_1^{\infty} 1/x^\alpha dx$ —is

$$\left(\frac{n-1}{n}\right)^\alpha = \left(1 - \frac{1}{n}\right)^\alpha .$$

It does not seem to be well known, but is certainly easy to prove, that if $\sum a_n$ and $\sum b_n$ are two series of positive terms for which the quotient $(a_n/a_{n-1})/(b_n/b_{n-1})$ is ≤ 1 for all sufficiently large n , then $\sum b_n < \infty$ implies $\sum a_n < \infty$. The reason is that if $(a_n/a_{n-1})/(b_n/b_{n-1}) \leq 1$ holds for all $n > N$, then

$$\begin{aligned} a_{N+p} &= (a_{N+p}/a_{N+p-1}) \cdots (a_{N+2}/a_{N+1}) \cdot (a_{N+1}/a_N) \cdot a_N \\ &\leq (b_{N+p}/b_{N+p-1}) \cdots (b_{N+2}/b_{N+1}) \cdot (b_{N+1}/b_N) \cdot a_N = \frac{a_N}{b_N} b_{N+p} \end{aligned}$$

so the terms of $\sum a_n$ are majorized by (a constant multiple of) the corresponding terms of the convergent series $\sum b_n$ for all sufficiently large indices. Now the quotient of the ratios of consecutive terms of $\sum_{n=0}^{\infty} \left| \binom{a}{n} \right|$ and $\sum_{n=1}^{\infty} 1/n^\alpha$ is

$$\frac{1 - (a + 1)/n}{(1 - 1/n)^\alpha} .$$

The limit of this ratio is 1 as $n \rightarrow \infty$, but if we replace $1/n$ by x , take logarithms and take the derivative, we get

$$\begin{aligned} \log \left(\frac{1 - (a + 1)x}{(1 - x)^\alpha} \right) &= \log(1 - (a + 1)x) - \alpha \log(1 - x) \\ \frac{d}{dx} \log \left(\frac{1 - (a + 1)x}{(1 - x)^\alpha} \right) &= -\frac{a + 1}{1 - (a + 1)x} + \alpha \frac{1}{1 - x} . \end{aligned}$$

The limiting value of this derivative as $x \rightarrow 0^+$ is $\alpha - a - 1$, so if $1 < \alpha < a + 1$ the derivative with respect to x will be **negative** and thus the quotient of the ratios will **increase** to its limit 1 as $n \rightarrow \infty$ and $1/n \searrow 0$, a fact which makes the ratios of $\sum_{n=0}^{\infty} \left| \binom{a}{n} \right|$ smaller than

those of $\sum_{n=1}^{\infty} \frac{1}{n^\alpha} < \infty$ and thus will make $\sum_{n=0}^{\infty} \left| \binom{a}{n} \right| < \infty$.

So good, the series of coefficients converges absolutely and thus the power series converges absolutely and uniformly on $[-1, 1]$; is there a “real-variable reason” that it converges to $(1 + x)^a$ for $x \in (-1, 1)$? One such reason is furnished by the integral form of the remainder in the Taylor series: if $f(x)$ has $N + 1$ continuous derivatives in an interval containing zero, then N successive integrations by parts give

$$f(x) = \sum_{n=0}^N \frac{f^{(n)}(0)}{n!} x^n + \frac{1}{N!} \int_0^x (x - t)^N f^{(N+1)}(t) dt . \quad (*)$$

For $f(x) = (1 + x)^a$, one has $f^{(N+1)}(t) = a(a - 1) \cdots (a - N) \cdot (1 + x)^{a-N-1}$

The limit of this quotient is 1 as $n \rightarrow \infty$, but if we replace $1/n$ by x , take logarithms and take the derivative, we get

$$\begin{aligned} \log \left(\frac{1 - (a + 1)x}{(1 - x)^\alpha} \right) &= \log(1 - (a + 1)x) - \alpha \log(1 - x) \\ \frac{d}{dx} \log \left(\frac{1 - (a + 1)x}{(1 - x)^\alpha} \right) &= -\frac{a + 1}{1 - (a + 1)x} + \alpha \frac{1}{1 - x} . \end{aligned}$$

The limiting value of this derivative as $x \searrow 0$ is $\alpha - a - 1$, so if $1 < \alpha < a + 1$ the derivative with respect to x will be **negative** near $x = 0^+$ and thus the quotient of the ratios will

eventually **increase** to its limit 1 as $n \rightarrow \infty$ and $1/n \searrow 0$, a fact which makes the ratios of $\sum_{n=0}^{\infty} \left| \binom{a}{n} \right|$ smaller than those of $\sum_{n=1}^{\infty} \frac{1}{n^\alpha} < \infty$ and thus will make $\sum_{n=0}^{\infty} \left| \binom{a}{n} \right| < \infty$.

So good, the series of coefficients converges absolutely and thus the power series converges absolutely and uniformly on $[-1, 1]$; is there a “real-variable reason” that it converges to $(1+x)^a$ for $x \in (-1, 1)$? One such reason is furnished by the integral form of the remainder in the Taylor series: if $f(x)$ has $N+1$ continuous derivatives in an interval containing zero, then N successive integrations by parts give

$$f(x) = \sum_{n=0}^N \frac{f^{(n)}(0)}{n!} x^n + \frac{1}{N!} \int_0^x (x-t)^N f^{(N+1)}(t) dt. \quad (\$)$$

For $f(x) = (1+x)^a$, one has

$$f^{(N+1)}(t) = a(a-1)\cdots(a-N) \cdot (1+t)^{a-N-1} = (N+1)! \binom{a}{N+1} (1+t)^{a-N-1}$$

so the integral form of the remainder term for the N -th-degree Maclaurin polynomial is

$$\begin{aligned} & \frac{1}{N!} \int_0^x (x-t)^N (N+1)! \binom{a}{N+1} (1+t)^{a-N-1} dt = \\ & (N+1) \binom{a}{N+1} \int_0^x (1+t)^{a-1} \left[\frac{x-t}{1+t} \right]^N dt. \end{aligned} \quad (\$\$)$$

The absolute value of the expression in brackets in the integral in ($\$\$$) is bounded by 1 for $-1 \leq t \leq 1$. This is obvious for $0 \leq t \leq x \leq 1$ and almost as obvious for $-1 \leq x \leq t \leq 0$, since in the latter case we have

$$\left| \frac{x-t}{1+t} \right| = \frac{t-x}{t-(-1)} \leq 1.$$

Since $a > 1$, the function $(1+t)^{a-1}$ is integrable on $[-1, 1]$, so we can bound the remainder term by

$$\begin{aligned} & \left| (N+1) \binom{a}{N+1} \int_0^x (1+t)^{a-1} \left[\frac{x-t}{1+t} \right]^N dt \right| \leq \\ & (N+1) \left| \binom{a}{N+1} \right| \left| \int_0^x (1+t)^{a-1} dt \right| \leq \\ & (N+1) \left| \binom{a}{N+1} \right| \max \left\{ \frac{2^a - 1}{a}, \frac{1}{a} \right\}. \end{aligned}$$

We saw on pp. 6–7 above that for sufficiently large n there is a constant M for which $\left| \binom{a}{n} \right| \leq M \cdot (1/n^\alpha)$, provided that $1 < \alpha < a+1$. Consequently $(N+1) \left| \binom{a}{N+1} \right| \leq M \cdot (1/(N+1)^{\alpha-1}) \rightarrow 0$ as $N \rightarrow \infty$, so for exponents $a > 0$ the remainder term for the

binomial series of $(1+x)^a$ converges to zero uniformly for $x \in [-1, 1]$. Thus the series not only converges absolutely uniformly on the interval, but converges to the correct value.

For the particular case $a = 1/2$ of the binomial series—the only one we need in order to show that uniformly-closed algebras of functions on a compact Hausdorff space contain the absolute values of their elements—there are some *ad hoc* methods available to get uniform convergence (to the correct values) of the Maclaurin series for $\sqrt{1-t}$ on the interval $0 \leq t \leq 1$; Folland gives one of these in **Lemma (4.47)** and exercise 66. A nice way to do this that is not quite so *ad hoc* is to observe that the function $f(t) = 2^{1/2} - (1-t)^{1/2}$ has the property that it and all its derivatives are nonnegative on the interval $(-1, 1)$. A function $f \in \mathcal{C}^\infty(a, b)$ is said to be **completely monotone** in a subinterval of (a, b) when it and all its derivatives are nonnegative in the subinterval, and Bernšteĭn has another theorem:

Theorem [S. N. Bernšteĭn]: Suppose $f \in \mathcal{C}^\infty(a, b)$ is completely monotone on $[c, b)$, where $a < c < b$. Then (1): for every $x_0 \in [c, b)$ the Taylor series for f with base point x_0 converges to $f(x)$ for every $x_0 \leq x < b$, uniformly for any $\delta > 0$ with $x_0 \leq x \leq b - \delta < b$. (2): If f is completely monotone on the whole interval (a, b) , then for every $x_2 \in (a, b)$ and $\delta > 0$ with $x_2 + \delta < b$, the Taylor series for f with base point x_2 converges to $f(x)$ uniformly for x in $[x_2 - \delta, x_2 + \delta] \cap (a, b)$. (3): If $\lim_{x \nearrow b} f(x) = S < \infty$ exists, then the Taylor series for f with base point x_0 converges to S for $x = b$, and if f is completely monotone on the whole interval (a, b) then the Taylor series for f with base point x_0 converges to $f(x)$ absolutely and uniformly for x in $[x_0 - (b - x_0), b] \cap (a, b)$.

Proof. For any x with $x_0 \leq x < b$ one can write the familiar finite Taylor expansion with remainder

$$f(x) = \sum_{n=0}^N \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n + \frac{f^{(N+1)}(\eta_{N,x})}{(N+1)!} (x - x_0)^{N+1}$$

(where $\eta_{N,x} \in (x_0, x)$ depends on x and N). Since everything in sight is nonnegative, we have

$$f(x) \geq \sum_{n=0}^N \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n,$$

and since for $x \geq x_0$ the corresponding formal Taylor series is a series of nonnegative terms, the formal Taylor series for $f(x)$, namely

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n, \tag{*}$$

is a series whose partial sums are bounded above by $f(x)$ and consequently is a convergent series. We do not yet know the value to which (*) converges. However, if $x - x_0 < (1/2)(b - x_0)$ then the midpoint $x_1 = (x_0 + b)/2$ of the interval from x_0 to b will lie to the right of x , while the point $\xi = x_1 + (x - x_0)$ —as far to the right of x_1 as x is to the

right of x_0 —will still lie to the left of b (verification trivial). It then follows from the same argument that established the convergence of (*) that the series

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_1)}{n!} (\xi - x_1)^n \quad (**)$$

converges and thus that the n -th term of (**) tends to zero. Since all the $f^{(n)}$'s are nondecreasing functions (their derivatives are nonnegative), we have by our choice of $x_1 > x > \eta_{N,x}$ and of ξ

$$0 \leq \frac{f^{(N+1)}(\eta_{N,x})}{(N+1)!} (x - x_0)^{N+1} \leq \frac{f^{(N+1)}(x_1)}{(N+1)!} (x - x_0)^{N+1} = \frac{f^{(N+1)}(x_1)}{(N+1)!} (\xi - x_1)^{N+1} \rightarrow 0$$

as $N \rightarrow \infty$, so the remainder term for the partial sums of (*) tends to zero and thus (*) actually converges to $f(x)$, **provided that x is closer to x_0 than half the distance from x_0 to b .**

We now know that there are intervals $[x_0, d] \subseteq [x_0, b]$ in which (*) converges to f . Let $e \leq b$ be the supremum of all such d , and suppose it were true that $x_0 < e < b$. We could then find an x_2 with $x_0 < x_2 < e$ and such that $e - x_2 < (1/4)(b - x_2)$, and what we have already proved implies that

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_2)}{k!} (x - x_2)^k \quad (***)$$

holds for $x - x_2 < (1/2)(b - x_2)$ and thus in particular holds for all x in an open interval whose right endpoint lies to the right of e . Replacing $x - x_0$ by $(x - x_2) + (x_2 - x_0)$ in (*) for $x_2 < x$ gives

$$\begin{aligned} & \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n = \\ & \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} [(x - x_2) + (x_2 - x_0)]^n = \\ \text{[binomial theorem]} & \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} (x - x_2)^k (x_2 - x_0)^{n-k} = \\ & \sum_{k=0}^{\infty} \frac{1}{k!} \left\{ \sum_{n=k}^{\infty} \frac{f^{(n)}(x_0)}{(n-k)!} (x_2 - x_0)^{n-k} \right\} (x - x_2)^k = \\ \text{[} n = j + k \text{]} & \sum_{k=0}^{\infty} \frac{1}{k!} \left\{ \sum_{j=0}^{\infty} \frac{f^{(j+k)}(x_0)}{j!} (x_2 - x_0)^j \right\} (x - x_2)^k = \\ \text{(***)} & \sum_{k=0}^{\infty} \frac{f^{(k)}(x_2)}{k!} (x - x_2)^k \end{aligned}$$

where the passage from the third to the fourth displayed line is justified because a doubly-indexed series of nonnegative terms can be grouped and summed in any order without changing the sum, and the passage from the fifth to the sixth line is justified because the series (*) is known to converge to $f(x)$ for $x_0 < x < e$, and can thus be repeatedly differentiated term-by-term to yield the values of the derivatives of f at x_2 (the expression enclosed by large braces is the term-by-term k -th derivative of (*), evaluated at $x = x_2$). The r. h. sides of (***) and (****) are identical, but the former series is known to converge to $f(x)$ for all x in an open interval whose right endpoint lies to the right of e , while the latter is equal to the sum of the series (*). We thus see that the series (*) converges to $f(x)$ for some values of x lying to the right of e , and that contradicts the definition of e ; so $e = b$ must hold and we have established assertion (1) of the theorem.

For (2), suppose that f is completely monotone on the whole interval (a, b) and that $x_2 \in (a, b)$ and $\delta > 0$ with $x_2 + \delta < b$ are given. The series

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_2)}{n!} (x - x_2)^n$$

converges for $x = x_2 + \delta$ and therefore (since its coefficients are nonnegative) it converges uniformly for $x \in [x_2 - \delta, x_2 + \delta]$; the only question is whether it converges to $f(x)$ at points on the left of x_2 . Given x with $x_2 - \delta < x < x_2$, fix any $x_0 \in (a, b)$ with $x_2 - \delta < x_0 < x$; then $|x - x_2| + |x_2 - x_0| = x_2 - x + x_2 - x_0 < (x_2 + \delta) - x_0$. Since $x_2 + \delta \in (a, b)$, we know that the series

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} ((x_2 + \delta) - x_0)^n$$

converges; its terms majorize those of

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} [|x - x_2| + |x_2 - x_0|]^n$$

(all the coefficients are nonnegative), and therefore this series also converges. The (familiar-looking) manipulation

$$\begin{aligned} & \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} [|x - x_2| + |x_2 - x_0|]^n = \\ \text{[binomial theorem]} & \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} \sum_{k=0}^n \frac{n!}{k! (n-k)!} |x - x_2|^k |x_2 - x_0|^{n-k} = \\ & \sum_{k=0}^{\infty} \frac{1}{k!} \left\{ \sum_{n=k}^{\infty} \frac{f^{(n)}(x_0)}{(n-k)!} |x_2 - x_0|^{n-k} \right\} |x - x_2|^k = \\ \text{[} n = j + k \text{]} & \sum_{k=0}^{\infty} \frac{1}{k!} \left\{ \sum_{j=0}^{\infty} \frac{f^{(j+k)}(x_0)}{j!} |x_2 - x_0|^j \right\} |x - x_2|^k \end{aligned}$$

is therefore justified, because these are (doubly indexed) series of nonnegative terms; but then the manipulation is justified without the absolute-value signs, because all the series involved are absolutely convergent. Consequently we again have

$$\begin{aligned}
 f(x) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n = \\
 &\quad \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} [(x - x_2) + (x_2 - x_0)]^n = \\
 \text{[binomial theorem]} \quad &\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} (x - x_2)^k (x_2 - x_0)^{n-k} = \\
 &\sum_{k=0}^{\infty} \frac{1}{k!} \left\{ \sum_{n=k}^{\infty} \frac{f^{(n)}(x_0)}{(n-k)!} (x_2 - x_0)^{n-k} \right\} (x - x_2)^k = \\
 \text{[} n = j + k \text{]} \quad &\sum_{k=0}^{\infty} \frac{1}{k!} \left\{ \sum_{j=0}^{\infty} \frac{f^{(j+k)}(x_0)}{j!} (x_2 - x_0)^j \right\} (x - x_2)^k = \\
 &\sum_{k=0}^{\infty} \frac{f^{(k)}(x_2)}{k!} (x - x_2)^k
 \end{aligned}$$

but this relation now shows that the series with base point x_2 converges to $f(x)$ even for $x_2 - \delta < x < x_2$. (The case $x = x_2 - \delta$ follows by letting $x \searrow x_2 - \delta$.)

Finally, if (3): If $\lim_{x \nearrow b} f(x) = S < \infty$ exists, then the fact that f is a nondecreasing function implies that the partial sums of (*) are majorized by S even for $x = b$, so (*) converges absolutely and uniformly for x in $[x_0 - (b - x_0), b] \cap (a, b]$; since both f and the sum of (*) have extensions to b that are continuous from the left and these agree near b , (*) must also converge to S for $x = b$.

It now appears as a special case of this theorem that because the function $f(t) = 2^{1/2} - (1-t)^{1/2}$ has the property that it and all its derivatives are nonnegative on the interval $(-1, 1)$ and $\lim_{t \nearrow 1} f(t) = 2^{1/2}$, its Maclaurin series converges uniformly and absolutely to $f(t)$ on $[-1, 1]$. The same is therefore true of the Maclaurin series of $\sqrt{1-t}$ —and Bernštein's theorem tells us this without our having to compute a single coefficient.

4. After that long digression—

Let us wind up our examination of the proof of the algebra version of Stone-Weierstraß. Any of the methods of §3 above will produce for any $\epsilon > 0$ and symmetric interval $[-b, b] \subseteq \mathbb{R}$ a polynomial $P(x)$ with constant term zero for which $\left| |x| - P(x) \right| < \epsilon$ uniformly on $[-b, b]$. It follows that given compact Hausdorff X and an algebra $\mathcal{A} \subseteq \mathcal{C}(X, \mathbb{R})$, one can produce for any $\epsilon > 0$ and $f \in \mathcal{A}$ an element $P(f) \in \mathcal{A}$ with $\| |f| - P(f) \|_{\infty} < \epsilon$, and thus that the

uniform closure $\overline{\mathcal{A}}$ contains the absolute value of any of its elements. It is obvious that the relations

$$\begin{aligned}\|(f \pm g) - (f_0 \pm g_0)\|_\infty &\leq \|f - f_0\|_\infty + \|g - g_0\|_\infty \\ \|(fg) - (f_0g_0)\|_\infty &\leq \|f\|_\infty \|g - g_0\|_\infty + \|f - f_0\|_\infty \|g_0\|_\infty\end{aligned}$$

imply that $\overline{\mathcal{A}}$ is an algebra if \mathcal{A} is; but an algebra that contains the absolute value of each of its elements is a lattice, because of the relations

$$f \vee g = \frac{f+g}{2} + \left| \frac{f-g}{2} \right|, \quad f \wedge g = \frac{f+g}{2} - \left| \frac{f-g}{2} \right|$$

showing that the lattice operations can be expressed in terms of the absolute-value function. These relations are geometrically obvious: the greater of two numbers equals their average plus half the distance between them, while the smaller equals their average minus half the distance between them. They can also be proved without considering pointwise values by remembering that $|\varphi| = \varphi \vee (-\varphi)$: this relation makes $|f-g| = (f-g) \vee (g-f)$ and then the translation-invariance of the order makes $(f+g) + |f-g| = (f+g) + [(f-g) \vee (g-f)] = (2f) \vee (2g)$ (now divide by 2 and use the invariance of order under multiplication by positive numbers). The relation one needs for infima then follows from $f \wedge g = -[(-f) \vee (-g)]$. So

Theorem [Stone-Weierstraß Theorem for subalgebras]: If X is a compact Hausdorff space and $\mathcal{A} \subseteq \mathcal{C}(X, \mathbb{R})$ is an algebra with the two-point interpolation property, then $\overline{\mathcal{A}}$ is a closed dense sublattice of $\mathcal{C}(X, \mathbb{R})$ and therefore $\overline{\mathcal{A}} = \mathcal{C}(X, \mathbb{R})$.

Folland gives refinements of this in his (4.45), in which the question of closing off \mathcal{A} is finessed by assuming that \mathcal{A} is closed and \mathcal{A} is merely required to **separate points of X** in the obvious sense. An algebra \mathcal{A} that separates points and contains the constant functions automatically has the two-point interpolation property, because if $x \neq y$ in X , $f(x) \neq f(y)$, and b_1 and b_2 are given, then the system of equations in the unknowns a_1 and a_2 $a_1 \cdot 1 + a_2 f(x) = b_1$, $a_1 \cdot 1 + a_2 f(y) = b_2$ has determinant $\det \begin{pmatrix} 1 & f(x) \\ 1 & f(y) \end{pmatrix} = f(y) - f(x) \neq 0$ and thus possesses a solution. The case $1 \in \mathcal{A}$ is the most frequently met case.

The power of the theorem is enormous: it obviously implies the classical Weierstraß theorem, not merely for single-variable polynomials on a closed interval of the real line but also for polynomials in the n coördinates on a compact set $X \subseteq \mathbb{R}^n$. Other applications may be found in Folland. The reader is invited to supply a proof of the Tietze extension theorem for closed subsets of (“large”) compact parallelepipeds in \mathbb{R}^n along the following lines: Let $X \subseteq Z = ([a_1, b_1] \times \cdots \times [a_n, b_n]) \subseteq \mathbb{R}^n$ be a compact subset of a closed parallelepiped Z . Then $\mathcal{C}(Z, \mathbb{R})|_X \subseteq \mathcal{C}(X, \mathbb{R})$ is a subalgebra (and sublattice) with the two-point interpolation property and therefore dense in $\mathcal{C}(X, \mathbb{R})$ by the Stone-Weierstraß theorem; it follows (**why?**) that it is all of $\mathcal{C}(X, \mathbb{R})$, and moreover that every element of $\mathcal{C}(X, \mathbb{R})$ can be extended to an element of $\mathcal{C}(Z, \mathbb{R})$ with the same uniform norm (and indeed the same maximum and minimum values).

Suppose the scalars are \mathbb{C} rather than \mathbb{R} . Then “lattice” makes no sense, but one can still talk about subalgebras $\mathcal{A} \subseteq \mathcal{C}(X, \mathbb{C}) = \mathcal{C}(X)$; is a subalgebra \mathcal{A} that separates

points of X necessarily dense in $\mathcal{C}(X)$? Folland gives the standard counterexample on p. 135. Another way to look at that example requires complex function theory, but no integration: with $\mathbb{D} = \{z \in \mathbb{C} : |z| \leq 1\}$ and $\mathbb{T} = \partial\mathbb{D}$ as usual, consider the function $\bar{z} \in \mathcal{C}(\mathbb{T})$ as being the same function as $1/z$ (which it is). Then there is no hope of approximating it by a polynomial $p(z)$ in the complex variable z , because by the maximum modulus theorem (multiply by z on \mathbb{T} to get the first equality) one has $\sup_{z \in \mathbb{T}} |1/z - p(z)| = \sup_{z \in \mathbb{T}} |1 - z \cdot p(z)| = \sup_{z \in \mathbb{D}} |1 - z \cdot p(z)| \geq |1 - 0 \cdot p(0)| = 1$. Indeed, it is not difficult to show that the uniform closure in $\mathcal{C}(\mathbb{T})$ of the polynomials in z consists exactly of those functions that have a (unique [again by the maximum modulus theorem]) continuous extension to \mathbb{D} that is holomorphic at every point of \mathbb{D}° .

Well, there is a quick fix if a simple-minded one: require of $\mathcal{A} \subseteq \mathcal{C}(X)$ that it contain with each function f its complex conjugate function. This happens if and only if \mathcal{A} contains with each function f the real and imaginary parts of f . When that happens $\mathcal{A} \cap \mathcal{C}(X, \mathbb{R})$ is closed under multiplication (why?), and it separates points of X if and only if \mathcal{A} does (if $f(x) \neq f(y)$ then the same must be true of the real or the imaginary part of f , or of both). \mathcal{A} is now seen to be the set of all functions $f = g + ih$ with $g, h \in \mathcal{A} \cap \mathcal{C}(X, \mathbb{R})$, and if that real algebra is dense in $\mathcal{C}(X, \mathbb{R})$ then $\mathcal{A} \subseteq \mathcal{C}(X)$ will be dense. For example, we can get a dense subalgebra of $\mathcal{C}(\mathbb{T})$ by taking polynomials in z and \bar{z} or (equivalently) polynomials in positive *and negative* powers of z . It is a theorem, incidentally, that if $A(\mathbb{D}) \subseteq \mathcal{C}(\mathbb{T})$ is the closure of the polynomials in z (or equivalently the subalgebra consisting of restrictions of functions continuous on \mathbb{D} and holomorphic on \mathbb{D}°), and if \mathcal{A} is a $\|\cdot\|_\infty$ -closed subalgebra of $\mathcal{C}(\mathbb{T})$ with $A(\mathbb{D}) \subseteq \mathcal{A} \subseteq \mathcal{C}(\mathbb{T})$, then either $\mathcal{A} = A(\mathbb{D})$ or $\mathcal{A} = \mathcal{C}(\mathbb{T})$: the algebra $A(\mathbb{D})$ is a **maximal** proper closed subalgebra. (This is the **Wermer maximality theorem**.)