The Weierstrass Uniform Approximation Theorem
W. L. Green
Research Horizons – October 11, 2006

In 1885 Karl Weierstrass published the first proof of what is now known as the Weierstrass Uniform Approximation Theorem.

Theorem: Let [a,b] be a compact interval contained in the real line. Then for any continuous function f on [a,b] and any ε > 0, there exists a polynomial p (in one real variable) such that for all x in [a,b] we have $|f(x) - p(x)| \le \varepsilon$.

In stating this theorem today it is common today to employ the uniform norm $\| \; \|_\infty$, given by

$$\|f\|_\infty = \sup\{|f(x)| : x \in [a,b]\},$$

and the conclusion of the theorem then becomes the existence of a polynomial p such that

$$\|f - p\|_\infty \le \varepsilon.$$

Weierstrass's result so caught the imagination of the mathematicians of his day that a veritable Who's Who of the best of them published additional proofs of the result over the next few decades. The list of notables and their proofs includes

Runge	1885
Picard	1891
Volterra	1897
Lebesgue	1898
Mittag-Leffler	1900
Fejér	1900
Lerch	1903
Landau	1908
de la Vallée Poussin	1912
Bernstein	1912
Fejér	1930

(This list, and much of the historical information in this talk, is taken from E. W. Cheney's excellent book *Introduction to Approximation Theory* (AMS Chelsea, 1982, reprinted in 2000)). Note that Fejér gave two different proofs separated by a span of thirty years.

Many of the proofs given above employed clever ideas or special insights from other parts of mathematics, e.g., probability or harmonic analysis, and all modern proofs rely on some moderately deep insight. To see why the theorem requires some subtlety, let us consider a naïve approach.

For any set of n+1 points in the plane, there is a unique polynomial of degree at most n that passes through all of these points. This is the usual *Lagrange interpolating polynomial* for this set of points, and its coefficients can be obtained as the unique solution vector to a certain linear system. By choosing ever larger sets of points on the graph of the function f, say over evenly spaced division points in the interval [a,b], and forming the interpolating polynomial for each of these sets of points, we might hope to find a sequence of polynomials that would converge back to f everywhere on [a,b]. We might even hope that the convergence would be uniform on [a,b], which would prove the Weierstrass Theorem. If the function f is smooth enough, this is actually a reasonable hope. Unfortunately, Bernstein showed in 1912 that if f is the absolute value function on [-1,1], then the sequence of interpolating polynomials (with evenly spaced nodes) converges *only* at the three points -1, 0, and 1 (where it obviously must). In particular, the Lagrange interpolating polynomials not only fail to converge uniformly; they fail to converge at all at almost every point of the interval. Another example, due to Runge, shows that for the derivative of the arctangent, the norm of the difference between f and the interpolating polynomial becomes arbitrarily large as the number of nodes goes to infinity.

Weierstrass proved his theorem by showing that every continuous periodic function on the line can be uniformly approximated by trigonometric polynomials; it then follows from Taylor's Theorem that the function in question can be approximated on any compact subinterval by ordinary polynomials. Lebesgue's proof employs the Taylor series about zero for the function $\sqrt{1-x}$ (which is complex analytic in the open unit disc). This series is monotone on [0,1] and can be shown to converge uniformly on [0,1]; replacing x by $1 - x^2$ then gives a series (but not a Taylor series) that converges uniformly on [-1,1] to the absolute value function. Lebesgue then showed how to use this result on the absolute value function to handle piecewise linear functions and hence arbitrary continuous ones.

Perhaps the most interesting of the proofs that followed Weierstrass's are those of Fejér and Bernstein. Fejér showed that although the partial sums of the Fourier series for a continuous periodic f need not converge back to f, the *Cesaro means* of these partial sums do in fact converge back to f, and quite remarkably do so uniformly. (For periodic functions, this is considerably sharper than the original Weierstrass result.) As before, Taylor's Theorem then finishes the job.

Bernstein, thinking along probabilistic lines, gave an *explicit* sequence of polynomials, based on the values of f at certain points, that converges uniformly on [0,1] to any continuous f. Indeed, the Bernstein polynomials are given by

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

Interestingly, Bernstein's proof proceeds by showing that if the Bernstein polynomials for the three functions f(x) = 1, x, and x^2 converge uniformly on [0,1] to f, then the same is true for any continuous function f on [0,1]. It is easy to see that an affine scaling reduces the general problem of [a,b] to that of [0,1], which again establishes (a slightly sharper form of) the Weierstrass result. Bohman [1952] and Korovkin [1953] noted this aspect of Bernstein's proof, and showed that if B_n is any sequence of linear and monotone operators from C[a,b] into itself, and if $B_n(f)$ converges uniformly to f for each of these three functions f, then $B_n(f)$ converges uniformly to f for all f in C[a,b].

After a few decades passed, some very significant generalizations of the original Weierstrass result appeared. One, due to Müntz [1914], goes as follows. The Weierstrass Theorem is equivalent to the assertion that the span of the set of monomials $\{1, x, x^2, \dots\}$ is uniformly dense in C[0,1]. Which subsets of this collection of monomials also have this property?

Theorem: Let $0 < n_1 < n_2 < \dots$ be an increasing sequence of positive integers. Then the span of $\{1, x^{n_1}, x^{n_2}, ...\}$ is uniformly dense in C[0,1] if and only if $\sum_i \frac{1}{n_i}$ diverges.

Note that Müntz's Theorem is specific to [0,1]. (Consider for example the even functions in C[-1,1], which are the closed linear span of the even powers of x.) To prove Müntz's Theorem it actually suffices to show that a corresponding result holds for the space $L^2[0,1]$. In $L^2[0,1]$, we can compute the distance from a monomial x^m to the space spanned by any finite subset of $\{1, x^{n_1}, x^{n_2}, ...\}$. The proof then shows that if x^m is not in the set $\{1, x^{n_1}, x^{n_2}, ...\}$, then this distance goes to zero if and only if $\sum_i \frac{1}{n_i}$ diverges. In particular, the proof shows that if the span of $\{1, x^{n_1}, x^{n_2}, ...\}$ is not dense, then the only monomials that are approximable by this span are the elements of the set $\{1, x^{n_1}, x^{n_2}, ...\}$ itself.

In 1939 Marshall Stone published a generalization in a different direction. Stone's result applies to the space C(X) for any compact Hausdorff space, and the key to the result is finding a generalization of the polynomials. Stone pointed out that the space C[a,b] is an algebra, and that the polynomials form a subalgebra containing the constant functions.

Theorem: Let X be a compact Hausdorff space, and let A be a subalgebra of C(X) that contains the constant function 1 and separates the points of X. (That is, for every pair x and y of distinct point of X there exists a function f in A with $f(x) \neq f(y)$.) Then the algebra A is uniformly dense in C(X).

There is also a version of the Stone-Weierstrass Theorem for the algebra of all continuous complex-valued function on X; in this case it is necessary to assume that the subalgebra is closed under complex conjugation.

The proof of Stone's version of the theorem brings us back again to the absolute value function. Using the Weierstrass version, we know that on the (compact) range of any function f in C(X), the absolute value function is uniformly approximable by polynomials. The composition of these polynomials with f then approximates the absolute value of f uniformly on X. In turn, this means that if f lies in the norm closure of A, then so does its absolute value. By standard arguments it then follows that a uniformly closed subalgebra of C(X) is closed under the lattice operations (taking the maximum or minimum of two functions pointwise). Stone then showed that a uniformly closed lattice that separates the point of X is all of C(X). In particular, we see that by investigating the approximation of the absolute value function, Bernstein and Lebesgue had gone directly to the central core of the issue, which reappears in Stone's investigations as the need for the lattice property.

We conclude with one last generalization of Weierstrass's original result, and also of Stone's result. The algebras C(X) are known to be, up to isometric *-isomorphism, the most general commutative C*-algebras with identity. What about the non-commutative C*-algebras? It turns out that there is a version (actually more than one) of the Stone-Weierstrass Theorem for these as well. We need a little explanation.

A C*-algebra is, up to isometric *-isomorphism, a norm-closed adjoint-stable subalgebra of the algebra B(H) of all bounded linear operators on a Hilbert space H. A *state* on a C*-algebra is a positive linear functional of norm one. The states on the complex version of C(X) are precisely the Radon probability measures on X, and the points of X, thought of as evaluation measures, are the extreme points of the compact convex set of all states of C(X). Thus the Stone's version of the Weierstrass Theorem says this: if A is a norm-closed *-subalgebra of C(X) that contains the identity of A, and if A separates the extremal states of C(X), then A is all of C(X).

Theorem: Let B be any Type I C*-algebra with identity I. Let A be any C*-subalgebra of B such that A contains I, and suppose that for any two extremal states of B there is an element of A on which they differ. Then B is all of A.

This is a satisfactory generalization of Stone's result, but not entirely so since it applies only to Type I algebras. (A C*-algebra is of Type I if and only if whenever it is represented on a Hilbert space, the representation can be described entirely in terms of tensor products and direct sums of commutative algebras and algebras of the form B(H). These are the unsurprising ones. All commutative C*-algebras are of Type I.) There is another generalization that applies to all C*-algebras with identity, and it goes like this.

Theorem: Let B be a C*-algebra with identity I. Let A be any C*-subalgebra of B such that A contains I, and suppose that for any two distinct points in the weak*-closure of the extremal states of B there is an element of A on which they differ. Then B is all of A.

When A is C(X), and hence the states of A are the Radon probability measures on X, it is not hard to check that the extremal states of C(X) are weak*-closed in the space of states. Thus in the case of C(X), both of these generalizations reduce to Stone's theorem. Unfortunately, in the case of some very ordinary and commonly encountered non-commutative C*-algebras e.g., Glimm's UHF-algebras), the extremal states are *weak*-dense* in the space of states on A. Thus it would be very useful to solve the following open problem.

Open problem: Suppose B is a C*-algebra with identity I. Suppose A is a C*-subalgebra of B such that A contains I, and such that for any two distinct extremal states of B there is an element of A on which they differ. Must A be all of B?

For a clear discussion of the Stone-Weierstrass Theorems for non-commutative C*-algebras, see J Dixmier's *C*-algebras and their Representations* (North Holland, 1977), translated from the French (Gauthier-Villars, 1969).

WLG, 10/11/06