Weierstrass's theorem - leaving no Stone unturned

B.Sury

Stat-Math Unit
Indian Statistical Institute
Bangalore

Workshop on linear algebra and analysis

University of Hyderabad

September 14-17, 2006

In these talks, we discuss the basic theme of approximating functions by polynomial functions. Although it is exemplified by the classical theorem of Weierstrass, the theme goes much further. Even on the face of it, the advantage of polynomial approximations can be seen from the fact that unlike general continuous functions, it is possible to numerically feed polynomial interpolations of such functions into a computer and the justification that we will be as accurate as we want is provided by the theorems we discuss. In reality, this theme goes deep into subjects like Fourier series and has applications like separability of the space of continuous functions. Marshall Stone's generalisation to compact Hausdorff spaces is natural and important in mathematics. Applications of the Weierstrass approximation theorem abound in mathematics, and we discuss a few of them, including one addressing Gaussian quadrature. Although there are standard reference texts like Rudin's 'Principles of Mathematical Analysis' and Apostol's 'Mathematical Analysis', a reference whose style we have adopted for a lot of this material is Körner's book [K].

The starting point of all our discussions is :

Weierstrass's Theorem (1885) :
If $f : [a, b] \to \mathbf{C}$ is continuous, then for each $\epsilon > 0$, there is a polynomial $P(x)$ such that

$$|f(x) - P(x)| < \epsilon \;\; \forall \;\; x \in [a, b].$$

A topologist would re-phrase this as "the set of polynomials is dense in the space of continuous functions on $[a, b]$ for the metric given by the sup norm." The first question which arises is whether one could not expect a continuous function to be itself expressible at least as a power series if not actually as a polynomial. Unfortunately, even this is too much to expect as the following example shows.

Proposition (C^∞ but not analytic) :
The function $h : \mathbf{R} \to \mathbf{R}$ given by $h(x) = e^{-1/x^2}$ for $x \neq 0$ and $h(0) = 0$ is infinitely differentiable, but there does not exist any $\epsilon > 0$ such that in the interval $|x| < \epsilon$, the $h(x)$ could be expressed as a power series $\sum_{n=0}^{\infty} a_n x^n$.
Proof.
First, let us note that for a function $f(x)$ given on a fixed interval of the form $|x| < \epsilon$ by the convergent power series $\sum_{n=0}^{\infty} a_n x^n$, the coefficients satisfy

$a_n = \frac{f^{(n)}(0)}{n!}$.
For our function $h(x)$, an easy induction shows that for any $x \neq 0$, one has $h^{(n)}(x) = Q_n(1/x)exp(-1/x^2)$ for some polynomial $Q_n(x)$. Thus, $h(x)$ is infinitely differentiable at all $x \neq 0$.
Further, $h(x)$ is infinitely differentiable at $x = 0$ also and $h^{(n)}(0) = 0$ as seen by induction on n and the earlier inductive hypothesis for non-zero points since $exp(t^2)$ diverges faster than any polynomial as $t \to \infty$. Indeed,

$$\frac{h^{(n)}(x) - h^{(n)}(0)}{x} = x^{-1}Q_n(x^{-1})exp(-x^{-2}) \to 0 \quad as \quad x \to 0.$$

Thus, the observation made at the beginning of the proof shows that if $h(x)$ were to be expressible as a power series in any interval of the form $(-\epsilon, \epsilon)$, then its coefficients would all be zero ! Evidently, $h(x)$ is not the zero function in any such interval.

When Weierstrass proved the approximation theorem, he was 70 years old. Twenty years later, another proof was given by the 19-year old Fejer - this is what is charming about mathematics ! It is interesting to learn that in the beginning, Fejer was considered weak in mathematics at school and was required to have special tuition ! Fejer's proof is via Fourier series, and it turns out that *Weierstrass's theorem itself is equivalent to its periodic version.* Towards proving Fejer's theorem which implies Weierstrass's theorem, we recall what Fourier series are.

Let $f : \mathbf{R} \to \mathbf{C}$ be a continuous function which is periodic, of period 2π (equivalently, f can be considered as a function on the unit circle T). One defines the Fourier coefficients of f for any integer r by

$$\hat{f}(r) = \frac{1}{2\pi} \int_0^{2\pi} f(t)exp(-irt)dt.$$

The idea of defining this was clear - it is natural to expect a periodic function to be expressible as a linear combinations of the functions $t \mapsto exp(irt)$ for various r; the coefficient of $exp(irt)$ for a particular r is obtained using the orthogonality property $\int_0^{2\pi} e^{irt}dt = 0$ if $r \neq 0$, of these special functions. The question whether this natural expectation is well-founded is answered by Dirichlet affirmatively for good functions; this is :

Theorem (Dirichlet) :
If f is continuous and has a derivative function, which is continuous and bounded (except possibly at finitely many points), then the sums $S_n(f,t) = \sum_{-n}^{n} \hat{f}(r)exp(irt) \to f(t)$ as $n \to \infty$, at all points t where f is continuous.

The hypothesis is rather restrictive; for example, there are continuous functions f for which the sums $S_n(f,0)$ have infinite limsup as observed by Du Bois-Reymond. However, it still does not rule out the possibility of determining a continuous f (possibly not satisfying the hypothesis of Dirichlet's theorem) from its Fourier coefficients $\hat{f}(r)$, $r \in \mathbf{Z}$. This was answered in a surprising manner by the 19-year old Fejer who showed that the sequence $S_n(f,t)$ may not be well-behaved but their averages $\sigma_n = \frac{S_0+S_1+\cdots+S_n}{n+1}$ behave better. His result was :

Theorem (Fejer) :
(i) *If $f : T \to \mathbf{C}$ is Riemann integrable, then at any point t where f is continuous, we have*

$$\sigma_n(f,t) = \frac{1}{n+1}\sum_{k=0}^{n} S_k(f,t) = \sum_{r=-n}^{n} \frac{n+1-|r|}{n+1}\hat{f}(r)exp(irt) \to f(t).$$

(ii) *If $f : T \to \mathbf{C}$ is continuous then*

$$\sigma_n(f,t) = \sum_{-n}^{n} \frac{n+1-|r|}{n+1}\hat{f}(r)exp(irt) \to f(t)$$

uniformly.

Fejer's theorem (ii) above immediately implies the following explicit trigonometric version of Weierstrass's theorem as $\sigma_n(f,-)$ is a trigonometric polynomial for each n. Here, one means by a trigonometric polynomial, a function of the form $\sum_{r=-n}^{n} a_r exp(irt)$. The trigonometric version, in turn, will lead easily to the Weierstrass theorem itself as we shall show :

Weierstrass's theorem - Trigonometric version :
If $f : T \to \mathbf{C}$ be continuous. The, for any $\epsilon > 0$, there exists a trigonometric polynomial P with

$$sup_{t \in T}|f(t) - P(t)| < \epsilon.$$

The class of functions which plays the key role in the proof of Fejer's theorem are the functions $K_n(t) = \sum_{r=-n}^{n} \frac{n+1-|r|}{n+1} exp(irt)$ which occur as 'weights' of the Fourier coefficients $\hat{f}(r)$. Now, the function K_n is called a Fejer kernel function and has the following remarkable properties.

Properties of Fejer's kernel :
We may look at $K_n(t) := \sum_{r=-n}^{n} \frac{n+1-|r|}{n+1} exp(irt)$ for any real t.
(i) $K_n(t) = \frac{1}{n+1}(\frac{Sin((n+1)t/2)}{Sin(t/2)})^2$ for $t \neq 0$.
For $t = 0$, the expession on the right reduces to the limiting value $n+1$ which matches the value $K_n(0)$ clearly.
(ii) $K_n(t) \geq 0$ for all t.
(iii) $K_n \to 0$ uniformly outside $[-\delta, \delta]$ for each positive δ.
(iv) $\frac{1}{2\pi} \int_T K_n(t)dt = 1$.
These properties are easily verified by first principles. If we draw graphs of these functions, we will see that the support (width) gets smaller and smaller as n increases. As the total area of each is 1, these properties are sometimes expressed as asserting that the functions K_n form an approximate identity for the convolution operation.

Before giving the rigorous proof of Fejer's theorem, it is very easy to describe it informally first.
Now

$$\sigma_n(f,t) = \sum_{r=-n}^{n} \frac{n+1-|r|}{n+1} \hat{f}(r) exp(irt)$$

$$= \sum_{r=-n}^{n} \frac{n+1-|r|}{n+1} exp(irt) \frac{1}{2\pi} (\int f(x) exp(-irx)dx)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) K_n(t-x)dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t-x) K_n(x)dx.$$

The idea is that for a positive, small δ, and large n, we have

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f(t-x)K_n(x)dx \approx \frac{1}{2\pi} \int_{-\delta}^{\delta} f(t-x)K_n(x)dx \approx \frac{f(t)}{2\pi} \int_{-\delta}^{\delta} K_n(x)dx.$$

Therefore, we would have $\sigma_n(f,t) \approx f(t)$ for large n. Let us make this rigorous now.

Proof of Fejer's theorem.
(i) We have assumed that f is continuous at a certain point t on the circle.

Being Riemann integrable, f is bounded on the circle; say $|f(x)| \leq M$ for all x. Now, for any $\epsilon > 0$, there is δ depending on t and on ϵ such that $|f(x) - f(t)| \leq \epsilon/2$ whenever $|x - t| < \delta$. By the property (iii) of the Fejer kernels, there is N (depending, of course, on δ and, therefore, on t, ϵ) such that

$$|K_n(x)| \leq \epsilon/4M \quad \forall \quad x \notin [-\delta, \delta] \ , \ n \geq N.$$

Then

$$|\sigma_n(f,t) - f(t)| = |\frac{1}{2\pi}\int_T (f(t-x) - f(t))K_n(x)dx|$$

$$\leq \frac{1}{2\pi}\int_{x\in[-\delta,\delta]} |(f(t-x)-f(t))K_n(x)|dx + \frac{1}{2\pi}\int_{x\notin[-\delta,\delta]} |(f(t-x)-f(t))K_n(x)|dx.$$

Now, in the first integral, one can use the inequality $|f(t-x) - f(t)| \leq \epsilon/2$ for the integrand and use the positivity and property (iv) on K_n being of unit area; this bounds the first integral by $\epsilon/2$. In the second integral, if we use the bound $|f(t-x) - f(t)| \leq 2M$, and the bound $|K_n(x)| \leq \epsilon/4M$, that integral too will be bounded by $\epsilon/2$. This completes the proof of (i) of Fejer's theorem.

Now (ii) follows quite immediately from the proof of (i) by noting that f must be uniformly continuous on the circle and by replacing $\delta(t, \epsilon)$ and $N(t, \epsilon)$ in the proof by constants dependent only on ϵ.

We draw attention, in passing, to a rather interesting consequence of Fejer's theorem :

If f, g are both continuous complex-valued functions on the unit circle, and if $\hat{f}(r) = \hat{g}(r)$ for all integers r, then $f = g$.

This is immediate from the fact that

$$0 = \sigma_n(f,t) - \sigma_n(g,t) \to f(t) - g(t)$$

as $n \to \infty$.

Fejer's proof of the Weierstrass theorem.

Recall the statement we are trying to prove here :

If $f : [a, b] \to \mathbf{C}$ is continuous, then for each $\epsilon > 0$, there is a polynomial $P(x)$ such that

$$|f(x) - P(x)| < \epsilon \quad \forall \quad x \in [a, b].$$

Note first that the interval $[a,b]$ can be taken to be $[-\pi,\pi]$ without loss of generality. Indeed, replace f by $g(x) = f(a + \frac{(x+\pi)(b-a)}{2\pi})$ and, replace an approximation $Q(x)$ to $g(x)$ by the approximation $P(x) = Q(\frac{2\pi(x-a)}{b-a} - \pi)$. Thus, we assume $[a,b] = [-\pi,\pi]$. Consider the function $F(x)$ whose values in $|x| \leq \pi$ are taken to be $f(|x|)$ and in $|x| > \pi$ so that F has period 2π. This is a continuous function. By Fejer's theorem, there exists $n \geq 1$ and complex numbers $a_{-n}, \cdots, a_{-1}, a_0, a_1, \cdots, a_n$ satisfying

$$|F(t) - \sum_{r=-n}^{n} a_r exp(irt)| < \epsilon/2$$

for all t. But, the series $\sum_{k=0}^{\infty} \frac{(ix)^k}{k!}$ converges uniformly to $exp(ix)$ in any bounded interval $[-M,M]$. Therefore, there exists $m(r)$ corresponding to each $r \in [-n,n]$ so that

$$|\sum_{k=0}^{m(r)} \frac{(irt)^k}{k!} - exp(irt)| \leq \frac{\epsilon}{(4n+2)|a_r|+1}$$

for all t with $|t| \leq 1$. Then, the polynomial $P(t) = \sum_{r=-n}^{n} a_r \sum_{k=0}^{m(r)} \frac{(irt)^k}{k!}$ satisfies, for $t \in [0,1]$,

$$|P(t) - f(t)| = |P(t) - F(t)| < \frac{\epsilon}{2} + \sum_{r=-n}^{n} \frac{\epsilon}{4n+2} = \epsilon.$$

This proves Weierstrass's theorem.

An application to moments :
Here is an application of Weierstrass's theorem which is useful in probability theory where one works with moments.

Theorem (Hausdorff) :
Let $f, g : [a,b] \to \mathbf{C}$ be continuous functions. Then, if equality of the moments

$$\int_a^b x^r f(x)dx = \int_a^b x^r g(x)dx$$

holds for all $r \geq 0$, then $f \equiv g$ on $[a,b]$.
Proof.
Working with $h = f - g$, it suffices to show that if all moments of h vanish,

then $h$must be the zero function. Now, for any polynomial $P(x)$, we have $\int_a^b P(x)h(x)dx = 0$. Let $\{P_n\}$ be a sequence of polynomials converging uniformly on $[a,b]$ to the function $\overline{h(x)}$. Since $\{P_n(x)h(x)\}$ converges uniformly on $[a,b]$ to the function $|h(x)|^2$, we have $\int_a^b |h(x)|^2 = 0$. As the integrand is real and non-negative throughout the interval, it must be zero.

The following example shows that Hausdorff's theorem is invalid if we go to infinite intervals.

Counterexample on $[0,\infty)$.
The moments of the non-zero, real-valued, continuous function

$$h(x) = exp(-x^{1/4})Sin(x^{1/4})$$

on $[0,\infty)$ *are zero.*
To get a positive function as an example, one can look at $g(x) = max(h(x), 0)$.

Lord Kelvin on compasses and tides :
Two practical applications of polynomial approximation emerged from the work of Lord Kelvin. One was the problem of correcting a magnetic compass mounted in a ship (which usually has a lot of iron and steel components). If a true angle of θ to the north is given by the compass as $f(\theta)$ (that is, with with an error $g(\theta) = f(\theta) - \theta$), then it makes sense to approximate $g(\theta)$ for small θ by a trigonometric polynomial of degree 2, say

$$g(\theta) = a_0 + a_1Cos(\theta) + a_2Cos(2\theta) + b_1Sin(\theta) + b_2Sin(2\theta).$$

The point is that by taking a few readings in the port by comparing with known directions θ, one can easily compute the a_i's and the b_i's. Experiments have shown that this approximation is reasonable - the value of the error $g(\theta)$ can usually be determined upto 2 or 3 degrees. Kelvin also designed a compass which can easily be corrected along these lines and was used extensively until the second world war.
The other situation to which Lord Kelvin applied the idea of approximation by trigonometric polynomials is to the prediction of tides. The height $h(t)$ of the tide at time t is known to be a sum of certain periodic functions $h_1(t) + h_2(t) + \cdots + h_N(t)$. For instance, $h_1(t)$ might have, as period, the rotation of the earth with respect to the moon, $h_2(t)$ may have its period to be that of the rotation of the earth with respect to the sun etc. Approximating

the h_i's by trigonometric polynomials, one has an approximation of the form

$$h(t) \approx a_0 + \sum_{r=1}^{N}(a_r Cos(m_r t) + b_r Sin(m_r t)).$$

If one has a record of the value $h(t)$ over a long range $[S, S+T]$ one can compute the coefficients from the easily-proved formulae :

$$\frac{2}{T}\int_S^{S+T} h(t)Cos(m_r t)dt \to a_r,$$

$$\frac{2}{T}\int_S^{S+T} h(t)Sin(m_r t)dt \to b_r$$

as $T \to \infty$. The computations of these integrals can be carried by numerical integration, an area to which polynomial approximation applies, as we will show later. Incidentally, the numbers m_r's are selected from the frequencies of the form $k\lambda$ where $2\pi\lambda^{-1}$ is the period of earth's rotation with respect to the moon etc. Experimentally though, it turns out that one needs to take T large. Very remarkably, Lord Kelvin also built a machine known as a harmonic-analyser to compute the coefficients a_i's and b_i's from the records of measured $h(t)$. This was Government-funded and was used purely to replace brain by brass - hence, it has a claim to be a forerunner of computers which came 20 years later.

Application to differentiability of Fourier series :
Another direct consequence of polynomial approximation is that under some conditions, the Fourier series can be differentiated term by term. This is :

Theorem.
Let $f : T \to \mathbf{C}$ be continuous. If the series $\sum_{n\in\mathbf{Z}} |n\hat{f}(n)|$ converges, then f is a C^1-function, and the series $\sum_{r=-n}^{n} ir\hat{f}(r)exp(irt) \to f'(t)$ uniformly as $n \to \infty$.
Proof.
Consider the sequence of functions $f_n(t) = S_n(f,t) := \sum_{r=-n}^{n}\hat{f}(r)exp(irt)$ on the circle. Since $|\hat{f}(r)| \leq |r||\hat{f}(r)|$ for $r \neq 0$, the comparison test shows that $\sum_{r=-n}^{n}|\hat{f}(r)|$ converges as $n \to \infty$. We claim that $f_n(t) \to f(t)$ uniformly on the circle. Given any $\epsilon > 0$, there is n_0 depending on it such that for all $m \geq n \geq n_0$ and all t on the circle, one has

$$|\sum_{n\leq|r|\leq m}\hat{f}(r)exp(irt)| \leq \sum_{n\leq|r|\leq m}|\hat{f}(r)exp(irt)| = \sum_{n\leq|r|\leq m}|\hat{f}(r)| < \epsilon.$$

Thus, by Weierstrass's M-test, $f_n(t)$ does converge uniformly to some function $g(t)$. Since $f_n(t)$'s are continuous, so is $g(t)$.
As $\sum_{r=-n}^{n} \hat{f}(r)exp(irt) \to g(t)$ uniformly, for any $k \in \mathbf{Z}$, we get

$$\sum_{r=-n}^{n} \hat{f}(r)exp(i(r-k)t) \to g(t)exp(-ikt)$$

uniformly. Hence,

$$\hat{f}(k) = \sum_{r=-n}^{n} \hat{f}(r)\frac{1}{2\pi}\int exp(i(r-k)t)dt \to \frac{1}{2\pi}\int g(t)exp(-ikt)dt = \hat{g}(k)$$

as $n \to \infty$. Thus, $g = f$ since a continuous function with all Fourier coefficients zero is the zero function.
Note that the above analysis applies to the functions

$$f_n'(t) = \sum_{r=-n}^{n} ir\hat{f}(r)exp(irt)$$

because of the hypothesis that $\sum_{r=-n}^{n} |r||\hat{f}(r)|$ converges as $n \to \infty$. We conclude that $f_n'(t)$ converge uniformly to a continuous function $h(t)$.
Finally, using the following standard fact for our sequence f_n, the result follows :
if $\{g_n\}_n$ are C^1-functions from the circle such that $g_n \to g$ and $g_n' \to h$ uniformly as $n \to \infty$, then g is in C^1 and $g' = h$.

How bizarre are continuous functions ?
In contrast with the previous theorem where some conditions forced differentiability, there are continuous functions which are nowhere differentiable, as Weierstrass showed. One can imagine the effect this must have had those days when everyone thought a continuous function could fail to be differentiable only at a very few points. Weierstrass's example is :

Example :
The series $\sum_{r=0}^{n} \frac{Sin((r!)^2t)}{r!}$ converges uniformly on T as $n \to \infty$ to a function $h(t)$ which is continuous and nowhere-differentiable.

This example was a fore-runner of various pathological examples like Peano's space-filling curve (a continuous, surjective function from $\mathbf{R}$ to $\mathbf{R}^2$). Lest we

think that these are bizarre and useless, it should be borne in mind that these ideas were the ones which led to the study of random functions and Brownian motion which have so many practical applications.
We mentioned earlier that an example due to Du Bois-Reymond in 1876 shattered the then-prevalent hope that the Fourier series of any continuous function must converge at every point. Stalwarts like Dirichlet, Riemann, Weierstrass and Dedekind had earlier believed this would indeed be true. Later, in 1926 Kolmogorov produced examples of a Lebesgue integrable (but not even Riemann integrable, let alone continuous) functions whose Fourier series diverged at every point. After that, opinion began to swing towards the belief that there could be continuous functions whose Fourier series diverged everywhere ! Finally, in 1964, Carleson proved the very surprising result that the Fourier series of a continuous function (indeed, of an L^2-function) must converge at all points except for a set of measure zero !

We will now give an important application of Weierstrass's approximation theorem called the Riemann-Lebesgue lemma. To motivate it, recall the convolution of two functions $f, g : \mathbf{R} \to \mathbf{C}$ defined by

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x-y)g(y)dy$$

is a 'smoothing' operator. For instance, if f is continuous and bounded, and g is Riemann integrable on each finite $[a, b]$ and satisfies $\int_{-\infty}^{\infty} |g(x)|dx < \infty$, then $f * g$ is continuous and bounded. One has similar (and somewhat better) results for the convolution of functions on the circle. For two continuous functions on the circle, the n-th Fourier coefficient of their convolution is the product of the corresponding Fourier coefficients. In fact, the set of all continuous functions forms a commutative ring under the operations of pointwise addition and of convolution. That it does not have an identity is a consequence of the :

Riemann-Lebesgue lemma :
Let $f : T \to \mathbf{C}$ be continuous. Then, $\hat{f}(n) \to 0$ as $|n| \to \infty$.

Before proving this lemma, let us point out :

Corollary ($C(S^1)$ has no identity) :
There is no continuous function g on the circle so that $g * f = f$ for every continuous f.

For, if there were, then one would have $\hat{g}(n)\hat{f}(n) = \hat{f}(n)$ for all n. For the function $f(t) = exp(int)$, we would have $\hat{g}(n) = 1$ for all n, which would make it rather difficult (!) for $\hat{g}(n)$ to converge to zero as $|n|$ gets arbitrarily large.

Proof of Riemann-Lebesgue lemma.
For any $\epsilon > 0$, using Fejer's trigonometric version of the Weierstrass theorem, we have a trigonometric polynomial $P(t) = \sum_{r=-n}^{n} a_r exp(irt)$ so that

$$Sup_{t \in T}|f(t) - P(t)| < \epsilon.$$

For $|N| > n$, $\hat{P}(N) = 0$ and so

$$|\hat{f}(N)| = |(\widehat{f - P})(N)|$$

$$= \frac{1}{2\pi}|\int (f-P)(t)exp(iNt)dt| \leq \frac{1}{2\pi}\int |(f-P)(t)|dt \leq \epsilon.$$

Lemma (Mean square convergence) :
Let $f : T \to \mathbf{C}$ be continuous, and as before,

$$(S_n f)(t) = \sum_{-n}^{n} \hat{f}(r)exp(irt).$$

Then,

$$||f - S_n f||_2 \to 0 \quad as \quad n \to \infty.$$

Of course, we note that mean square convergence does not imply even pointwise convergence; there are examples.
Proof.
By the trigonometric version of the Weierstrass approximation theorem, given $\epsilon > 0$, there is a polynomial $P_n(t) = \sum_{r=-n}^{n} a_r exp(irt)$ such that $|P_n(t) - f(t)| < \epsilon$. Therefore,

$$||P_n - f||_2^2 = \frac{1}{2\pi}\int |P_n(t) - f(t)|^2 dt \leq \epsilon^2.$$

If $N > n$, then $S_N(P_n) = P_n$; so

$$||f - S_N(f)||_2 \leq ||f - P_n||_2 + ||P_n - S_N(P_n)||_2 + ||S_N f - S_N P_n||_2$$

$$\leq ||f - P_n||_2 + ||S_N(f - P_n)||_2.$$

But, by using how minimisation problem is solved in an inner product space (discussed in our linear algebra lectures), it follows that $||S_N g||_2 \leq ||g||_2$ for all g. Thus, we have

$$||f - S_N f||_2 \leq 2\epsilon$$

which proves the lemma.

Application to Gaussian quadrature :
We discuss an interesting application to quadrature. The method of approximating an integral by the interpolating polynomial at some points was viewed carefully by Gauss. He showed that this approximation is exact for polynomials of degree $< 2n$ if the n points are the zeroes of the Legendre polynomial $P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n$. More precisely, he proved that if $x_1, \cdots, x_n$ are the zeroes of $P_n(x)$ then there exist some $a_1, \cdots, a_n$ so that, for any polynomial $P(x)$ of degree $\leq 2n-1$,

$$\int_{-1}^{1} P(x)dx = \sum_{r=1}^{n} a_r P(x_r).$$

Note that each $a_k > 0$ as seen by applying Gauss's theorem to the polynomial $P(x) = \prod_{i \neq k}(x - x_i)^2$ of degree $2n-2$ and observing that the right side becomes $a_k P(x_k)$. Also, applying the Gauss theorem to the constant polynomial $P \equiv 1$, we have $\sum_{r=1}^{n} a_r = 2$.
For any function f, write $G_n(f) = \sum_{r=1}^{n} a_r f(x_r)$ with x_r, a_r as in Gauss's result. The following beautiful theorem was proved by Stieltjes :

Theorem (Stieltjes) :
Let f be any continuous function on $[-1,1]$. Let $\mathcal{P}_d$ denote the space of polynomials of degree $\leq d$. Then,
(a) $|G_n(f) - \int_{-1}^{1} f(x)dx| \leq 4 \inf\{sup_{t\in[-1,1]}|f(t) - P(t)| : P \in \mathcal{P}_{2n-1}\}$.
(b) $G_n(f) \to \int_{-1}^{1} f(x)dx$ *as* $n \to \infty$.
Proof.
(a) Let $P \in \mathcal{P}_{2n-1}$. By Gauss's theorem, $G_n(P) = \int_{-1}^{1} P(x)dx$. Therefore,

$$|G_n(f) - \int_{-1}^{1} f(x)dx| \leq |G_n(f) - G_n(P)| + |\int_{-1}^{1} f(x)dx - \int_{-1}^{1} P(x)dx|$$

$$\leq \sum_{r=1}^{n} a_r|(f(x_r) - P(x_r))| + |\int_{-1}^{1} |f(x) - P(x)|dx \leq 4 \; sup_{t\in[-1,1]}|f(t) - P(t)|.$$

This proves (a).
(b) Weierstrass's approximation theorem shows

$$\inf\{sup_{t\in[-1,1]}|f(t) - P(t)| : P \in \mathcal{P}_{2n-1}\} \to 0$$

as $n \to \infty$; this proves (b) immediately.

Bernstein's constructive proof :
A constructive proof of Weierstrass's theorem was given by Sergei Bernstein in 1911. His result is :

Theorem (Bernstein) :
Let $f : [0,1] \to \mathbf{R}$ be continuous. For each natural number n, consider the corresponding Bernstein polynomial of f given by

$$B_n(x; f) = \sum_{r=0}^{n} \binom{n}{r} x^r (1-x)^{n-r} f(r/n).$$

Then, the sequence $B_n(x; f)$ converges uniformly on $[0,1]$ to $f(x)$.

Heuristic idea of proof :
Before proceeding to prove the theorem rigorously, we stop for a moment to reflect on the statement. We have placed the weights $\binom{n}{r}x^r(1-x)^{n-r}$ at the points r/n to the values of f and these weights add up to 1. One could imagine this as follows. Consider a dartboard of unit area and for any fixed x between 0 and 1, consider a region of area x coloured black. If n darts are thrown at the board at random, and if r of them land in the black region, a reward of $f(r/n)$ rupees is given. What would the average winnings be as the number n of throws increases ? Since x^r is the probability of r darts landing in the black region, and $(1-x)^{n-r}$ is the probability that the other $n-r$ darts landing outside the black region and $\binom{n}{r}$ is the number of ways of choosing r darts from the n thrown, the probability of getting exactly r darts in the black region is the product of these three numbers. The expectation (or average winnings) is precisely the number $B_n(x; f)$. As the number n of trials increases, it is more and more probable that the proportion r/n of darts landing in the black region gets closer and closer to the whole area x and thus, the expectation gets closer and closer to $f(x)$.

Proof of Bernstein's theorem.
A crucial property of Bernstein polynomials is $B_n(x; C) = C$ for a constant

polynomial C; indeed,

$$B_n(x;C) = CB_n(x;1) = C\sum_{r=0}^{n}\binom{n}{r}x^r(1-x)^{n-r} = C.$$

The other useful property of the Bernstein polynomials which is clear from their definition is :
$B_n(x;g) \geq B_n(x;h)$ if $g \geq h$; in particular, $B_n(x;g) \geq 0$ for a positive function g.
Let $\epsilon > 0$ be given. Now, on the compact interval $[0,1]$, f is automatically uniformly continuous; so $\exists\delta > 0$ such that, for all $x, y \in [0,1]$,

$$|x-y| \leq \delta \Rightarrow |f(x) - f(y)| \leq \epsilon.$$

Put $M = Sup\{|f(x)| : x \in [0,1]\}$. Fix $x_0 \in [0,1]$. We will try to bound the function $f - f(x_0)$ above by a positive, nonconstant function. If $x \in [x_0 - \delta, x_0 + \delta]$, then this is easy as $|f(x) - f(x_0)| \leq \epsilon$. If $|x - x_0| > \delta$, then we simply look at the bound

$$|f(x) - f(x_0)| \leq 2M < 2M(\frac{x-x_0}{\delta})^2.$$

Hence, we have for every x, the bound $|f(x) - f(x_0)| \leq 2M(\frac{x-x_0}{\delta})^2 + \epsilon$.
Using the remark about $B_n(x;g)$ being monotonic in g, we then have

$$|B_n(x; f - f(x_0))| \leq B_n(x; 2M(\frac{x-x_0}{\delta})^2 + \epsilon) = \frac{2M}{\delta^2}B_n(x;(x-x_0)^2) + \epsilon.$$

Hence,

$$|B_n(x;f) - f(x_0)| = |B_n(x; f - f(x_0))| \leq \frac{2M}{\delta^2}B_n(x;(x-x_0)^2) + \epsilon.$$

Now,

$$B_n(x;(x-x_0)^2) = x^2 + \frac{1}{n}(x-x^2) - 2x_0x + x_0^2 = (x-x_0)^2 + \frac{1}{n}(x-x^2).$$

Feeding this is the previous inequality, we have

$$|B_n(x;f) - f(x_0)| \leq \frac{2M}{\delta^2}(x-x_0)^2 + \frac{2M}{\delta^2 n}(x-x^2) + \epsilon.$$

Putting $x = x_0$ and observing that the maximum value of $x_0 - x_0^2$ is $1/4$, we get

$$|B_n(x_0; f) - f(x_0)| \leq \frac{M}{2\delta^2 n} + \epsilon.$$

For large n, one has the first term $\frac{M}{2\delta^2 n} < \epsilon$ so that

$$|B_n(x_0; f) - f(x_0)| \leq 2\epsilon.$$

This proves the theorem.

Reflecting on Bernstein's proof :
It is clear that the idea of the proof is geometric and very simple. For a continuous $f \in C([0,1])$, and $x_0 \in [0,1]$, and $\epsilon > 0$, continuity of f implies there exists an interval $(x_0 - \delta, x_0 + \delta)$ in which $(x - x_0)^2 + (f(x_0) + \epsilon) > f(x)$. Thus, we will have a parabola of the form $p(x) = a(x - x_0)^2 + (f(x_0) + \epsilon)$ which majorizes f in $[0,1]$ but has the value $f(x_0) + \epsilon$ at x_0. Therefore, $f(x)$ is the pointwise infimum of parabolas as above. Going from 'pointwise' to 'uniform' amounts to reducing to the case of infimum of finitely many parabolas; this is accomplished if we tolerate some error.

Before we start with the generalization due to Marshall Stone of Weierstrass's theorem, we briefly discuss the very interesting question of approximation by integral polynomials. These results are due to Le Baron Ferguson and the proofs are very easy (see [F]) but we do not give them here.

Chebychev is special :
It is a well-known fact that among all *monic* polynomials of a given degree n on $[-1,1]$ the unique polynomial which has the least sup norm is the Chebychev polynomial $T_n(x) = \frac{1}{2^{n-1}} Cos(n\ Cos^{-1}(x))$. For a general interval $[a,b]$, the corresponding polynomial is

$$P_n(x) = 2((b-a)/4)^n Cos(n\ Cos^{-1}[(2x - b - a)(b - a)]).$$

Note that it has sup norm at least 2 if $b - a \geq 4$. Therefore, as any nonconstant integral polynomial is the product of its top coefficient (an integer) and a monic polynomial, we have the first observation :
On an interval $[a,b]$ with $b - a \geq 4$, each nonconstant integer polynomial has sup norm ≥ 2.
Using this, we see that if f is any continuous function on an interval $[a,b]$ of

length ≥ 4 and if it is uniformly a limit of integral polynomials $P_n(x)$, then the Cauchy property implies that $P_n(x)$ is eventually a constant sequence and is, therefore, equal to $f(x)$. In other words, :
Theorem (uninteresting approximation in length ≥ 4) :
On an interval of length at least 4, the only continuous functions approximable uniformly by integral polynomials are those polynomials themselves.

On the other hand, the problem becomes more interesting for intevals of smaller length. In fact the following theorem is a quite easy consequence of Bernstein's proof above :

Theorem (intervals of length 1) :
A continuous function f on $[0,1]$ is a uniform limit of integral polynomials if and only if $f(0), f(1)$ are integers. In particular, $Sin(x)$ is not, while $Sin(\pi x)$ is a limit !
This is obviously carried over to any interval of the form $[n, n+1]$.

The problem becomes even more interesting for the interval $[-1,1]$ when the condition for approximability of f turns out to be : $f(-1), f(0), f(1)$ are integers and $f(-1)+f(1)$ is even. With a careful analysis, Le Baron Ferguson finally proves the following result. First, for any subset I of $\mathbf{R}$, he forms the set $J(I)$ consisting of all algebraic integers in I all of whose conjugates lie in I. For example, $J([-1,1]) = \{-1,0,1\}$ and $J([-\sqrt{2},\sqrt{2}]) = \{\pm\sqrt{2},\pm 1, 0\}$. He proves the pretty theorem :

Theorem (integral polynomial approximation in length < 4):
Let f be a continuous function on an interval I of length < 4. Then, f is uniformly approximable by integral polynomials if and only if the interpolating polynomial for f on $J(I)$ is integral.

Stone's turn :

Weierstrass's theorem was generalized to compact Hausdorff topological spaces by Marshall Stone in 1937. For instance, if one has a closed disc in the plane, then the question as to whether continuous functions on it can be approximated (uniformly) by polynomial functions is answered affirmatively by such a generalization. Stone published another proof in 1948 - this appeared in Mathematics Magazine, a journal of undergraduate education ! This generalization to subalgebras of the algebra of continuous functions on a compact, Hausdorff space has come to be known as the Stone-Weierstrass theorem; as a matter of fact, the version for algebras follows from a version for function lattices. Ultimately, the proof depends on the Taylor series of the function $\sqrt{1-t}$; more precisely, on the fact that $\sum_{n=0}^{\infty}|\binom{1/2}{n}| < \infty$. First, we define the spaces occurring in the Stone-Weierstrass theorem.

Definitions.
Let K be a *compact metric space*. The set $C(K)$ of continuous functions $f : K \to \mathbf{R}$ forms a *Banach algebra* under the norm $||f||_{\infty} := Sup_{t\in K}|f(t)|$. A subset A of $C(K)$ is said to be a *subalgebra with unity* if it is a real subspace of $C(K)$, contains the constant function 1 and the product of any two of its elements. Note that such an A contains all constant functions. One says that A *separates points* if, for any two distinct points $x \neq y$ in K, one can find a function $f \in A$ so that $f(x) \neq f(y)$. It is clear from the definition that if A is a subalgebra of $C(K)$ containing unity and separates points of K, then its closure in $C(K)$ also has all these properties. In addition, it is closed.
Note that for any compact interval $[a, b]$ in $\mathbf{R}$, the set of all real polynomial functions on $[a, b]$ is a subalgebra with unity which separates points of $[a, b]$. It is clearly not closed.

Stone-Weierstrass theorem.
Let K be a compact metric space, and let $A \subseteq C(K)$ be a closed subalgebra with unity which separates points of K. Then, $A = C(K)$.

It is enough to take K to be a compact topological space and the whole proof goes through. However, we will see that the exact analogue of the Stone-Weierstrass theorem is not true over complex numbers but, will hold with an additional assumption; in fact, the above Stone-Weierstrass theorem implies the correct complex version.

Lemma (on Taylor series of $\sqrt{1-t}$) :
The Taylor series of $f(t) = (1-t)^{1/2}$ at 0 converges absolutely and uniformly to $f(t)$ on $[-1, 1]$.
Proof.
Formally, one would write $\sqrt{1-t} = 1 - \sum_{n=1}^{\infty} a_n t^n$ where

$$a_n = (-1)^{n-1}\binom{1/2}{n} = \frac{(-1)^{n-1}}{n!}\prod_{k=0}^{n-1}(1/2 - k) = \frac{(2n-2)!}{2^{2n-1}n!(n-1)!}.$$

Since $a_n \geq 0$ and the ratio $\frac{a_{n+1}}{a_n}$ converges to 1, the ratio test clearly tells us that the series converges pointwise for each $t \in (-1,1)$. Checking similarly that the derivative of the series converges pointwise, simple manipulations and an application of Stirling's formula $lim_{n\to\infty}\frac{n!}{\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}} = 1$, tells us that the series converges uniformly on $[-1,1]$ to $f(t)$.

A subset L of $C(K)$ is said to be a *lattice* if, for all $f,g \in L$ the functions $(f \vee g)(x) := max(f(x),g(x))$ and $(f \wedge g)(x) := min(f(x),g(x))$ are in L as well.

Proposition (on subalgebras being lattices) :
Let $A \subseteq C(K)$ be a closed subalgebra with unity. Then,
(a) *if $f \in A$, and $f \geq 0$, then $\sqrt{f} \in A$,*
(b) *if $f \in A$ then $|f| \in A$,*
(c) *A is a lattice.*
Proof.
(a) follows from using $F = \frac{f}{||f||_\infty}$ and applying the previous lemma to $\sqrt{1-(1-F)}$. We omit details.
(b) Noting $|f| = \sqrt{f^2}$, this follows from (a).
(c) As $f \vee g = \frac{1}{2}(f+g+|f-g|)$ and $f \wedge g = \frac{1}{2}(f+g-|f-g|)$, the result follows from (b).

Proof of Stone-Weierstrass theorem.
Let $\epsilon > 0$ let $f \in C(K)$. For fixed points $s \neq t$ in K, there is some $h \in A$ so that $h(s) \neq h(t)$. Now, for any real u,v look at the function

$$H_{u,v}(x) := v + (u-v)\frac{h(x)-h(t)}{h(s)-h(t)}$$

on K. Clearly, it is in A and $H_{u,v}(s) = u, H_{u,v}(t) = v$.
Let us take $u = f(s), v = f(t)$ and call the corresponding $H_{u,v}$ as $f_{s,t}$.
Thus, $f_{s,t} \in A$ and $f_{s,t}(s) = f(s), f_{s,t}(t) = f(t)$. We keep s fixed and vary t.
Then the sets

$$U_t := \{v \in K : f_{s,t}(v) < f(v) + \epsilon\}$$

are open and contain t. Thus, $K = \cup_t U_t$ and using compactness of K, there are elements $t_1, \cdots, t_n \in K$ such that

$$K = \cup_{i=1}^n U_{t_i}.$$

Take $h_s := min(f_{s,t_i}; i \leq n)$. The proposition tells us that $h_s \in A, h_s(s) = f(s)$ and $h_s < f + \epsilon$. If we define

$$V_s := \{v \in K : h_s(v) > f(v) - \epsilon\},$$

then V_s is open and $K \subset \cup_{s \in K} V_s$. By compactness, there are $s_1, \cdots, s_m \in K$ so that

$$K \subset \cup_{i=1}^m V_{s_i}.$$

Putting $g = max(h_{s_i}; i \leq m)$, we have $g \in A$ and

$$f - \epsilon < g < f + \epsilon.$$

This is just the statement $||f - g||_\infty < \epsilon$. Hence A is dense in $C(K)$. But, it is assumed to be closed; hence $A = C(K)$, which completes the proof.

It is quite easy to deduce the Tietze extension theorem for compact metric spaces from the above theorem. Tietze's theorem asserts that *if Y is a closed subset of a compact metric space X, then every continuous function on Y can be extended to a continuous function on the whole of X while preserving the uniform norm.*

Counterexample over C :
That the exact analogue of the real Stone-Weierstrass theorem fails for complex numbers can be seen by using a little bit of complex analysis. One argument using complex integration goes as follows. We claim that for $\epsilon \in (0, 1)$, there is no complex polynomial $P(z)$ such that $|\bar{z} - P(z)| < \epsilon$ for all z on the unit circle $T : |z| = 1$. Indeed, assuming the existence of such a P, we have $\int_T zP(z)dz = 0$. But, then

$$1 = \int_T |z|^2 dz = \int_T (z(\bar{z} - P(z)) + zP(z)) = 0$$

a contradiction.
Here is a different argument which avoids integration but uses the maximum modulus principle on the disc $D = \{z : |z| \leq 1\}$. On T, if $\bar{z} = 1/z$ were approximated by a polynomial $P(z)$, then by the maximum modulus principle, we would have

$$sup_T |1/z - P(z)| = sup_T |1 - zP(z)| = sup_D |1 - zP(z)| \geq 1$$

which rules out any possibility of the left hand side going to zero.
This example shows that the complex polynomials on T are not dense in $C(T, \mathbf{C})$ although they separate points. Thus, the following version is the correct complex analogue of Stone-Weierstrass theorem and can be deduced trivially (by looking at $(f+\bar{f})/2, (f-\bar{f})/2i$ for $f \in A$) :

Complex Stone-Weierstrass theorem :
Let K be a compact metric space, and let $A \subseteq C(K, \mathbf{C})$ be a closed subalgebra which contains constants and the complex conjugates of all its elements and separates points. Then $A = C(K, \mathbf{C})$.

References :

[K] : T.Körner, *Fourier analysis*, Cambridge University Press 1989.
[F] : Le Baron O.Ferguson, *What can be approximated by polynomials with integer coefficients*, American Math. Monthly, May 2006, P.403-414.