Weierstrass's Theorem – Leaving no 'Stone' Unturned

B Sury

In this article, 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 – to Gaussian quadrature for instance.

B Sury was associated with *Resonance* during 1999–2005. In this article, he celebrates a wonderful theorem which he introduces as:

An ancient important question it was,
to approximate by polynomials without loss.
For the functions in C[0,1]
it was beautifully done
by the great Karl Weierstrass!

1. Weierstrass's Classical Theorem

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

Keywords
Weierstrass approximation, Fejer's theorem, Fejer kernel, polynomial approximation, Hausdorff theorem, moments, Bernstein polynomial, Gaussian quadrature, Legendre polynomial, Fourier series.

The question arises: Can one expect a continuous function to be expressible as a power series if not a polynomial? But even this is too much to expect.

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* $h(x) = \mathrm{e}^{-1/x^2}$ *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,

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

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.

2. Fejer Kernel for a Trigonometric Version

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.

Weierstrass's theorem is equivalent to its periodic version.

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

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

The idea of defining this was clear – it is natural to expect a periodic function to be expressible as a linear combination 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 $\int_0^{2\pi} \mathrm{e}^{irt} \mathrm{d}t = 0$ 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:

It is natural to expect a periodic function to be expressible as a linear combination of the functions $t \mapsto \exp(irt)$.

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

This does not rule out the possibility of determining a continuous function f from its Fourier coefficients.

Fejer's theorem implies the explicit trigonometric version of Weierstrass's theorem.

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*

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

(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. Then, for any $\epsilon > 0$, there exists a trigonometric polynomial P with

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

The trignometric version in turn leads to the Weierstrass theorem.

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.

If we draw graphs of these functions we see that the support (width) gets smaller as *n* increases.

***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) $\frac{1}{2\pi} \int_T K_n(t) \mathrm{d}t = 1$.

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

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

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

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

The functions K_n form an approximate identity for the convolution operation.

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

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

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

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

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

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 $n \to \infty$.

3. 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

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

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.

4. 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*

$$\int_a^b x^r f(x) \mathrm{d}x = \int_a^b x^r g(x) \mathrm{d}x$$

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 $\int_a^b P(x)h(x)\mathrm{d}x = 0$. 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.

Hausdorff's theorem is invalid if we go to infinite intervals.

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 $g(x) = \max(h(x), 0)$.

5. 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 $M = \text{Sup}\{|f(x)| : x \in [0, 1]\}$. 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

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

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,

$$\begin{aligned} B_n(x;(x-x_0)^2) &= x^2 + \frac{1}{n}(x - x^2) - 2x_0 x + x_0^2 \\ &= (x-x_0)^2 + \frac{1}{n}(x-x^2). \end{aligned}$$

Feeding this in 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 point-wise infimum of parabolae as above. Going from 'point-wise' to 'uniform' amounts to reducing to the case of infimum of finitely many parabolae; this is accomplished if we tolerate some error.

Two practical applications of polynomial approximation emerged from the work of Lord Kelvin.

6. 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

$$g(\theta) = a_0 + a_1\cos(\theta) + a_2\cos(2\theta) + b_1\sin(\theta) + b_2\sin(2\theta).$$

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 up to 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

Lord Kelvin applied the idea of approximation by trigonometric polynomials to the prediction of tides.

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

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:

$$\frac{2}{T}\int_S^{S+T} h(t)\cos(m_r t)\mathrm{d}t \to a_r,$$

$$\frac{2}{T}\int_S^{S+T} h(t)\sin(m_r t)\mathrm{d}t \to b_r$$

as $T \to \infty$. The computations of these integrals can be carried out 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.

The computation of these integrals can be done by numerical integration.

Lord Kelvin also built a machine known as a harmonic analyser.

7. Remarks on Gaussian Quadrature

The method of approximating an integral by the interpolating polynomial at some points was carefully considered by Gauss. He showed that this approximation is exact for polynomials of degree $< 2n$ if the n points are the zeroes of the so-called Legendre polynomial $P_n(x) = \frac{1}{2^n n!}\frac{\mathrm{d}^n}{\mathrm{d}x^n}(x^2-1)^n$. 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 $\leq 2n-1$,

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

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 $\sum_{r=1}^{n} a_r = 2$.

The method of approximating an integral by interpolating polynomials was studied by Gauss.

On an interval of lenth 4, the only continuous functions which are uniformly approximable by integral polynomials are those polynomials themselves.

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. An application of the Weierstrass approximation theorem is the following beautiful theorem due to Stieltjes:

Let f be any continuous function on $[-1,1]$. Then $G_n(f) \to \int_{-1}^{1} f(x)\mathrm{d}x$ as $n \to \infty$.

8. What about Integral Polynomials?

If we look at polynomials with integer coefficients, a consequence of the very special properties of Chebychev polynomials is that, 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 intervals of smaller length. In fact an easy consequence of Bernstein's proof is:

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!

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.

The interested reader is referred to the proofs of these assertions and their generalizations due to Le Baron Ferguson in $[F]$.

Acknowledgements

It is a pleasure to acknowledge the feedback from Shailesh Shirali on this article. Initially, I had written a two-part article with the second part containing some advanced material. On the referee's suggestion to make it more UG-friendly, I have more or less removed the second part completely. I thank him for this suggestion. On some future occasion, I hope to discuss in detail the question of

differentiability of Fourier series, the Stone–Weierstrass theorem for compact Hausdorff spaces and also the interesting application to integer polynomials alluded to above. Although there are many textbooks discussing different aspects of the subject-matter above, I suggest the curious student to the friendly text by Körner which we have also followed to a large extent.

Suggested Reading

[1] **T M Apostol, *Mathematical Analysis*, Addison-Wesley Publishing Company, 1981.**

[2] **Le Baron O Ferguson, What can be approximated by polynomials with integer coefficients, *American Math. Monthly*, pp.403–414, May 2006.**

[3] **T Korner, *Fourier analysis*, Cambridge University Press, 1989.**

[4] **W Rudin, *Principles of Mathematical Analysis*, McGraw-Hill Inc., 1976.**

Address for Correspondence
B Sury
Stat-Math Unit
Indian Statistical Institute
8th Mile Mysore Road
Bangalore 560 059m India.
Email: sury@isibang.ac.in