The Weierstrass Approximation Theorem[1]

The general mathematical ideas include uniform convergence, heat equation on the line, power series, Fourier series, and convolutions, as well as basic probability.

The Theorem

Every continuous function on a bounded interval can be approximated by a polynomial with arbitrarily high precision.

More precisely, for every continuous function $f = f(x)$, $x \in [a, b]$, and for every number $\varepsilon > 0$, there exists a polynomial $P = P(x)$, depending on the function f and the number ε, such that

$$\max_{x\in[a,b]} |f(x) - P(x)| < \varepsilon.$$

Proof number 1 [by Weierstrass himself.]

Step 1. Continue the function f to the whole real line in such a way that the result is still continuous and is equal to zero for $x \notin [a-1, b+1]$. Use the same notation $f = f(x)$ for the result.

Step 2. Let $u = u(t, x)$ be the solution of the heat equation

$$u_t = u_{xx},\ t > 0, x \in \mathbb{R}$$

with initial condition $u(0, x) = f(x)$. In other words,

$$u(t,x) = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{+\infty} f(y) e^{-(x-y)^2/(4t)}\, dy. \tag{1}$$

Step 3. Argue that

$$\lim_{t\to 0+} \max_{x\in[a,b]} |u(t,x) - f(x)| = 0$$

Step 4. Argue that, for every $t > 0$, the function $u = u(t, x)$ is (real) analytic as a function of x.

Step 5. Complete the proof: given $\varepsilon > 0$, use Step 3 to find a $t_0 > 0$ such that

$$\max_{x\in[a,b]} |u(t_0,x) - f(x)| < \frac{\varepsilon}{2}$$

and then approximate $u(t_0, x)$ by the corresponding Taylor polynomial, which is possible by Step 4. □

Proof number 2 [Using trigonometric series]

Step 1. With no loss of generality, assume that $a = 0, b = 1$: this can always be achieved by considering

$$g(t) = f\big(a + t(b-a)\big).$$

Step 2. Compute the Fourier cosine coefficients of $h(t) = f(|\cos t|)$:

$$A_0 = \frac{1}{\pi}\int_0^{\pi} h(t)dt,\ A_k = \frac{2}{\pi}\int_0^{\pi} h(t)\cos(kt)\, dt.$$

Step 3. Argue that if

$$\sigma_{h,n}(t) = A_0 + \sum_{k=1}^{n} \frac{n+1-k}{n+1} A_k \cos(kt)$$

then

$$\lim_{n\to\infty} \max_{t\in[0,2\pi]} |h(t) - \sigma_{h,n}(t)| = 0. \tag{2}$$

Step 4. Argue that, for every $n = 1, 2, \ldots$, there exists a polynomial $T_n = T_n(x)$ such that

$$\cos(nt) = T_n(\cos t). \tag{3}$$

Note that this is true when $n = 1$ ($T_1(x) = x$) and $n = 2$ (because $\cos(2t) = 2\cos^2 t - 1$, so that $T_2(x) = 2x^2 - 1$). Then use `strong` induction and the trig identities to conclude that, because $\cos(nt) = 2\cos t \cos(n-1)t - \cos(n-2)t$, we have $T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x)$ which defines T_n for all n.

[1]Sergey Lototsky, USC; updated on November 22, 2022

Step 5. Complete the proof: keeping in mind that, with suitable numbers $a_{k,n}$,

$$\sigma_{h,n}(t)=\sum_{k=0}^{n}a_{k,n}\cos(kt)=\sum_{k=0}^{n}a_{k,n}T_k(\cos(t)),$$

and therefore

$$\begin{aligned}\max_{t\in[0,2\pi]}|h(t)-\sigma_{h,n}(t)| &= \max_{t\in[0,2\pi]}\left|f(|\cos(t)|)-\sum_{k=0}^{n}a_{k,n}T_k(\cos(t))\right|\\ &= \max_{x\in[0,1]}\left|f(x)-\sum_{k=0}^{n}a_{k,n}T_k(x)\right|,\end{aligned}$$

meaning that, for sufficiently large n, the polynomial $\sum_{k=0}^{n}a_{k,n}T_k(x)$ does the job.

Proof number 3. [Using the Chebyshev inequality/(Weak) Law of Large Numbers]

Step 1. With no loss of generality assume that $a=0, b=1$, so that $x\in[0,1]$.

Step 2. For sufficiently large N, the polynomial

$$P_{N,f}(x)=\sum_{k=0}^{N}f(k/N)\binom{N}{k}x^k(1-x)^{N-k},\quad \binom{N}{k}=\frac{N!}{k!(N-k)!}, \tag{4}$$

does the job. Indeed, let $X_n(x)$ be Binomial random variable $\mathcal{B}(N,x)$ representing the number of successes in N independent trials, with probability of success in one trial equal to x. Then

$$P_{n,f}(x)=\mathbb{E}f\left(\frac{X_N(x)}{N}\right). \tag{5}$$

Uniform continuity of f implies that, for every $\varepsilon>0$, there is a $\delta>0$ [independent of x] such that

$$\left|f\left(\frac{X_N(x)}{N}\right)-f(x)\right|\le\frac{\varepsilon}{2} \tag{6}$$

as long as

$$\left|\frac{X_N(x)}{N}-x\right|\le\delta. \tag{7}$$

By the Chebyshev inequality,

$$\mathbb{P}\left(\left|\frac{X_N(x)}{N}-x\right|>\delta\right)\le\frac{x(1-x)}{N\delta}\le\frac{1}{4N\delta}, \tag{8}$$

because the mean of $X_N(x)$ is Nx and the variance of $X_N(x)$ is $Nx(1-x)$.

Let $C_f=\max_{0\le x\le 1}|f(x)|$. Then, combining (5)–(8),

$$\begin{aligned}\left|f(x)-P_{N,f}(x)\right| &= \left|f(x)-\mathbb{E}f\left(\frac{X_N(x)}{N}\right)\right|\le\mathbb{E}\left|f(x)-f\left(\frac{X_N(x)}{N}\right)\right|\\ &=\mathbb{E}\left(\left|f(x)-f\left(\frac{X_N(x)}{N}\right)\right|1\left(\left|\frac{X_N(x)}{N}-x\right|\le\delta\right)\right)\\ &+\mathbb{E}\left(\left|f(x)-f\left(\frac{X_N(x)}{N}\right)\right|1\left(\left|\frac{X_N(x)}{N}-x\right|>\delta\right)\right)\\ &\le\frac{\varepsilon}{2}+2C_f\mathbb{P}\left(\left|\frac{X_N(x)}{N}-x\right|>\delta\right)\le\frac{\varepsilon}{2}+\frac{C_f}{2N\delta}.\end{aligned}$$

This completes the proof: given an $\varepsilon>0$, we first find $\delta>0$ according to (6), (7), and then take $N>C_f/(\varepsilon\delta)$ so that $C_f/(2N\delta)<\varepsilon/2$ and

$$\left|f(x)-P_{N,f}(x)\right|<\varepsilon.$$

Comments

(1) The polynomials T_n defined by (3) are called `Chebyshev polynomials` of the first kind.
(2) The polynomials from (4) are called the `Bernstein polynomials`.
(3) A continuous on $[a, b]$ function f is `uniformly` continuous, that is, for every $\varepsilon > 0$, one can find a $\delta > 0$ so that

$$|f(x) - f(y)| < \varepsilon$$

for all $x, y \in [a, b]$ such that $|x - y| < \delta$. *This fact is used in all three proofs.*
(4) The `Stone-Weierstrass` theorem usually refers to an extension of the result to functions defined on sets other than bounded intervals in $\mathbb{R}$ and being approximated by functions other than polynomials.
(5) Recall that the Fourier series of a continuous function can diverge in many points. Accordingly, (2) [known as `Fejér's theorem`] is of independent interest, showing that a continuous function is uniquely determined by its Fourier coefficients, but not necessarily through the Fourier series.

Keeping in mind that one can only compute polynomials, the importance of the result in applications is clear. The result also opens the whole new direction of research: finding the "best" polynomial that does the job. An additional challenge is that the function f to be approximated is often not known. Still, the theorem affirms that polynomial approximation is possible as long as we can prove that the function we are trying to approximate is continuous.

To illustrate the importance of having a general result showing *possibility* of something, consider Hilbert's tenth problem, asking about a procedure for deciding whether a polynomial equation with integer coefficients [known as `Diophantine equations` has integer solutions. Some equations of this type have infinitely many solutions, for example, $x^2 + y^2 = z^2$, where the solutions are Pythagorean triplets. Some equations obviously have no solutions, such as $x^2+y^2+1 = 0$. Turns out, **no such procedure exists in general**, and this was proved by Yuri Matiyasevich in 1970, building on the earlier work by Martin Davis, Hilary Putnam and Julia Robinson.

The solution of the heat equation *might not be a real analytic function of* t even if the initial condition $\varphi(x) = u(0, x)$ is a real analytic function of x. Indeed, writing

$$u(t, x) = \sum_{k=0}^{\infty} u_k(x)t^k, \tag{9}$$

we conclude from the equation that

$$\sum_{k=0}^{\infty} \big((k + 1)u_{k+1}(x) - u_k''(x)\big)t^k = 0$$

or

$$(k + 1)u_{k+1}(x) = u_k''(x),$$

or

$$u_k(x) = \frac{\varphi^{(2k)}(x)}{k!},$$

where $\varphi^{(n)}$ is the n-th derivative of φ. For example, take

$$\varphi(x) = \frac{1}{1 - x} = \sum_{k=0}^{\infty} x^k, \ |x| < 1.$$

Then $\varphi^{(2k)}(0) = (2k)!$ so that, for $x = 0$, the right-hand side of (9) diverges for all $t > 0$. Together with non-uniqueness of solution, we see that the heat equation does not obey the Cauchy-Kovalevskaya theorem. The reason is that the set $\{(t, x) : t = 0\}$ is a characteristic surface for the equation, which is not allowed by the Cauchy-Kovalevskaya theorem. Keeping in mind that the heat equation also violates certain laws of physics [e.g. in connection with infinite propagation speed], it is truly remarkable that the equation is still an interesting and useful object to study in both mathematics and physics.

An infinitely differentiable function of a real variable $f = f(x)$ can fail to be real analytic in two main ways:

(1) The Taylor series of f at x_0 converges, but not to $f(x)$ when $x \neq x_0$. The main example is

$$f(x) = \begin{cases} e^{-1/x^2}, & x \neq 0, \\ 0, & x = 0, \end{cases}$$

with $x_0 = 0$: the Taylor series of f at zero is identically equal to zero.

(2) The Taylor series of f at x_0 diverges for all $x \neq x_0$. We already know that such functions can be solutions of differential equations, both ordinary and with partial derivatives. Another standard example is

$$f(x) = \frac{1}{\sqrt{\pi}} \int_0^{+\infty} \cos(xy) e^{-(\ln y)^2}\, dy. \tag{10}$$

You can (easily) verify that the integral converges for all x and, for $k = 1, 2, \ldots,$

$$f^{(n)}(0) = \begin{cases} (-1)^k e^{(2k+1)^2/4}, & n = 2k, \\ 0, & n = 2k-1, \end{cases}$$

so the Taylor series for f at $x_0 = 0$ has zero radius of convergence.

A different kind of polynomial approximation. Let $w = w(x)$, $x \in \mathbb{R}$, be a non-negative function, also called `weight function`, such that $s_n(w) = \int_{\mathbb{R}} x^n w(x)\, dx < \infty$ for all $n = 0, 1, 2, \ldots$. Define the space $H_w = \{f = f(x) : \int_{\mathbb{R}} |f(x)|^2 w(x)\, dx < \infty\}$, and, for $f, g \in H_w$, introduce the notations $(f, g)_w = \int_{\mathbb{R}} f(x)g(x)w(x)\, dx$, $\|f\|_w = \sqrt{(f, f)_w}$. Then all polynomials belong to H_w, and the following question makes sense: *is it true that, for every $f \in H_w$ and $\varepsilon > 0$, there exists a polynomial $P = P_{\varepsilon,f}(x)$ such that $\|f - P_{\varepsilon,f}\|_w < \varepsilon$?*

IF the Fourier transform of the function w is analytic at the origin, THEN the answer to the above question is YES.

IN GENERAL, the answer is NO, and the standard example, $w(x) = e^{-(\ln x)^2} 1(x > 0)$, is known as the `Stieltjes-Wigert weight function`. By direct computations, if $g(x) = \sin(2\pi \ln x)$, then, with $a = (n+1)/2$ and using $\sin(2\pi(x + a)x) = (-1)^{n+1} \sin(2\pi x)$,

$$\int_0^{\infty} x^n \sin(2\pi \ln x)\, e^{-(\ln x)^2}\, dx = \int_{\mathbb{R}} \sin(2\pi y) e^{ny - y^2 + y}\, dy = e^{a^2} \int_{\mathbb{R}} \sin(2\pi x) e^{-(x-a)^2}\, dx$$
$$= (-1)^{n+1} e^{a^2} \int_{\mathbb{R}} \sin(2\pi x) e^{-x^2}\, dx = 0$$

for every $n = 0, 1, 2, \ldots$. As a result, for every polynomial P, we have $(g, P)_w = 0$ and therefore

$$\|g - P\|_w^2 = \|g\|_w^2 + \|P\|_w^2 \geq \|g\|_w^2.$$

Note that the function from (10) is, up to a constant, the Fourier cosine transform of the Stieltjes-Wigert weight function, and is NOT analytic at the origin.

A closely related question is the `problem of moments`: given two (continuous) weight functions w and h, with $s_n(w) = s_n(h) < \infty$ for all $n = 0, 1, 2, \ldots$, is it true that $w(x) = h(x)$? In probabilistic language, we are asking whether two absolutely continuous random variables with equal moments have the same distribution. Once again, the Stieltjes-Wigert weight function shows that, in general, the answer is NO.