WEIERSTRASS PPROXIMATION THEOREM

LING WANG

In this notes, we will give several proofs of the Weierstrass approximation theorem. The original version of this result was established by Karl Weierstrass [W] in 1885 using the Weierstrass transform. Hence, we first show the original proof given by Weierstrass, and next we give a proof using the Bernstein polynomial. Finally, we prove the result by considering the convolution of a sequence of polynomials.

The Weierstrass approximation theorem states precisely as follows.

Theorem 1 (Weierstrass Approximation Theorem)**.** *Let* $f : [a, b] \to \mathbb{R}$ *be a continuous function. Then* f *is on* $[a, b]$ *a uniform limit of polynomials.*

1. First proof of Theorem 1

First proof of Theorem 1*.* We begin by extending f to a bounded uniformly continuous function on $\mathbb{R}$ by defining $f(x) = f(a)(x-a+1)$ on $[a-1, a)$, $f(x) = -f(b)(x-b-1)$ on $(b, b+1]$, and $f(x) = 0$ on $\mathbb{R}\backslash[a-1, b+1]$. In particular, there exists $R > 0$ such that $f(x) = 0$ for $|x| > R$. Hence f is a bounded uniformly continuous function on $\mathbb{R}$. For $h > 0$, we define

$$S_h f(x) = \frac{1}{h\sqrt{\pi}} \int_{-\infty}^{\infty} f(u) e^{-\left(\frac{u-x}{h}\right)^2} \, \mathrm{d}u.$$

Next, we show that $S_h f$ converges uniformly to f as $h \to 0$. Indeed, let $\varepsilon > 0$, then there exists $\delta > 0$ such that $|f(x) - f(y)| < \frac{\varepsilon}{2}$ for all $x, y \in \mathbb{R}$ with $|x - y| < \delta$. Assume $|f(x)| \le M$ on $\mathbb{R}$. Using that $\displaystyle\int_{-\infty}^{\infty} e^{-v^2} \, \mathrm{d}v = \sqrt{\pi}$, one also verifies easily that

$$\frac{1}{h\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{u-x}{h}\right)^2} \, \mathrm{d}u = 1.$$

This implies that we can write

$$f(x) = \frac{1}{h\sqrt{\pi}} \int_{-\infty}^{\infty} f(x) e^{-\left(\frac{u-x}{h}\right)^2} \, \mathrm{d}u.$$

Date: May 7, 2023.

Now let $h_0 > 0$ such that $h_0 < \frac{\varepsilon\delta\sqrt{\pi}}{2M}$, then

$$\begin{aligned}
|S_h f(x) - f(x)| &\leq \frac{1}{h\sqrt{\pi}} \int_{-\infty}^{\infty} |f(u) - f(x)| e^{-\left(\frac{u-x}{h}\right)^2} \, \mathrm{d}u \\
&\leq \frac{\varepsilon}{2} + \frac{1}{h\sqrt{\pi}} \int_{|x-u|\geq\delta} |f(u) - f(x)| e^{-\left(\frac{u-x}{h}\right)^2} \, \mathrm{d}u \\
&\leq \frac{\varepsilon}{2} + \frac{2M}{h\sqrt{\pi}} \int_{|x-u|\geq\delta} e^{-\left(\frac{u-x}{h}\right)^2} \, \mathrm{d}u \\
&= \frac{\varepsilon}{2} + \frac{2M}{\sqrt{\pi}} \int_{|v|\geq\frac{\delta}{h}} e^{-v^2} \, \mathrm{d}v \\
&\leq \frac{\varepsilon}{2} + \frac{2Mh}{\delta\sqrt{\pi}} \int_{|v|\geq\frac{\delta}{h}} |v| e^{-v^2} \, \mathrm{d}v \\
&\leq \frac{\varepsilon}{2} + \frac{4Mh}{\delta\sqrt{\pi}} \int_{0}^{\infty} v e^{-v^2} \, \mathrm{d}v \\
&= \frac{\varepsilon}{2} + \frac{2hM}{\delta\sqrt{\pi}} < \varepsilon
\end{aligned}$$

for all $0 < h \leq h_0$ and all $x \in \mathbb{R}$. Hence, $S_h f$ converges uniformly to f as $h \to 0$.

Let $\varepsilon > 0$ and M such that $|f(x)| \leq M$ for all x. Then by the above claim, we know that there exists $h_0 > 0$ such that for all $x \in \mathbb{R}$ there is $|S_{h_0} f(x) - f(x)| < \frac{\varepsilon}{2}$. Since $f(u) = 0$ for $|u| > R$, we can write

$$S_{h_0} f(x) = \frac{1}{h_0\sqrt{\pi}} \int_{-R}^{R} f(u) e^{-\left(\frac{u-x}{h_0}\right)^2} \, \mathrm{d}u.$$

On $\left[\frac{-2R}{h_0}, \frac{2R}{h_0}\right]$, the power series of e^{-v^2} converges uniformly, so there exists N such that

$$\left| \frac{1}{h_0\sqrt{\pi}} e^{-\left(\frac{u-x}{h_0}\right)^2} - \frac{1}{h_0\sqrt{\pi}} \sum_{k=0}^{N} \frac{(-1)^k}{k!} \left(\frac{u-x}{h_0}\right)^{2k} \right| < \frac{\varepsilon}{4RM}$$

for all $|x| \leq R$ and all $|u| \leq R$, since in that case $|u - x| \leq 2R$. This implies that

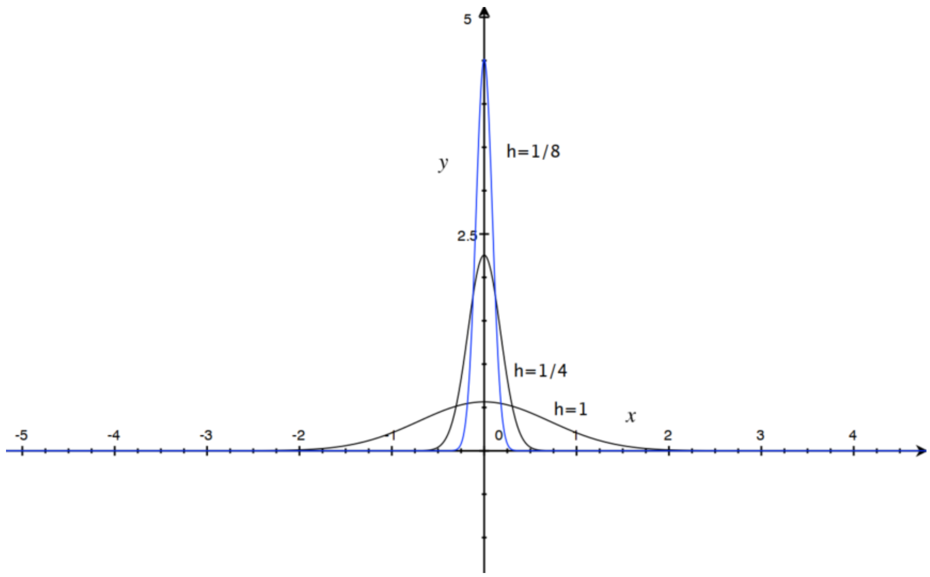
$$\left| S_{h_0} f(x) - \frac{1}{h_0\sqrt{\pi}} \int_{-R}^{R} f(u) \sum_{k=0}^{N} \frac{(-1)^k}{k!} \left(\frac{u-x}{h_0}\right)^{2k} \, \mathrm{d}u \right| < \frac{\varepsilon}{2}$$

for all $|x| \leq R$. If we put

$$P(x) = \frac{1}{h_0\sqrt{\pi}} \int_{-R}^{R} f(u) \sum_{k=0}^{N} \frac{(-1)^k}{k!} \left(\frac{u-x}{h_0}\right)^{2k} \, \mathrm{d}u,$$

then $P(x)$ is a polynomial in x of degree at most $2N$ such that $|S_{h_0} f(x) - P(x)| < \frac{\varepsilon}{2}$ for all $|x| \leq R$. This implies that $|f(x) - P(x)| < \varepsilon$ for all $x \in [a, b]$. □

Remark 2. *The function $S_h f$ is the convolution of f with a Gaussian heat kernel. These heat kernels form an approximate identity. The following figure shows the kernel for the values $h = 1, h = \frac{1}{4}$ and $h = \frac{1}{8}$.*



2. Second proof of Theorem 1

First, we state Chebyshev's inequality used in the following in probabilistic statement.

Lemma 3. *Let X (integrable) be a random variable with finite non-zero variance σ^2 (and thus finite expected value μ). Then for any real number $k > 0$,*

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

***Second proof of Theorem 1**.* It suffices to prove the theorem on $[0, 1]$. Since f is continuous on $[0, 1]$, it is uniformly continuous. This means that for any $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ such that $|f(x) - f(y)| < \varepsilon/2$ for all $x, y \in [0, 1]$ satisfying $|x - y| < \delta_\varepsilon$. Let us fix an $\varepsilon > 0$ and such a corresponding $\delta_\varepsilon > 0$.

Let r be any positive integer such that $r \geq \dfrac{\|f\|_\infty}{\delta_\varepsilon^2 \varepsilon}$. Define the Bernstein polynomials

$$b_{k,r}(x) = \mathrm{P}\left(S_{r,x} = k\right) = \binom{r}{k} x^k (1 - x)^{r-k}$$

where $S_{r,x} \sim \mathrm{Binom}(r, x)$, the binomial distribution. Let $p(x) := \sum_{k=0}^{r} f\left(\frac{k}{r}\right) b_{k,r}(x)$, which is a degree- r polynomial. Then, for any $x \in [0, 1]$,

$$|p(x) - f(x)| = \left| \sum_{k=0}^{r} \left(f\left(\frac{k}{r}\right) - f(x) \right) b_{k,r}(x) \right|$$

$$
\begin{aligned}
&\leq \sum_{|k-rx|<r\delta_\varepsilon} \left| f\left(\frac{k}{r}\right) - f(x) \right| b_{k,r}(x) + \sum_{|k-rx|\geq r\delta_\varepsilon} \left| f\left(\frac{k}{r}\right) - f(x) \right| b_{k,r}(x) \\
&\leq \frac{\varepsilon}{2} + 2\|f\|_\infty \, \mathrm{P}\left(|S_{r,x} - rx| \geq r\delta_\varepsilon\right) \\
&\leq \frac{\varepsilon}{2} + 2\|f\|_\infty \frac{x(1-x)}{r\delta_\varepsilon^2} \qquad \text{(by Lemma 3)} \\
&\leq \frac{\varepsilon}{2} + \frac{\|f\|_\infty}{2r\delta_\varepsilon^2} \\
&\leq \varepsilon
\end{aligned}
$$

where the final inequality uses the assumption on r. □

3. Third proof of Theorem 1

Finally, we give the third proof of Theorem 1. The idea of the proof is to choose a suitable sequence of polynomials $\{Q_n\}$ such that Q_n behaves like a 'Dirac delta function' as n$\to \infty$. Then, the sequence of polynomials $P_n(x) = \displaystyle\int_0^1 Q_n(x-t)f(t)\mathrm{d}t$ converges to $f(x)$ as $n \to \infty$. We will prove this momentarily, but first we need to do the ground work.

Notice that we only need to consider $a = 0$ and $b = 1$, with $f(0) = f(1) = 0$. If we prove this case, then for a general $\tilde{f} \in C([0,1])$, $\exists$ a sequence of polynomials

$$P_n(x) \to \tilde{f}(x) - \tilde{f}(0) - x(\tilde{f}(1) - \tilde{f}(0)) \text{ uniformly.}$$

Hence,

$$\tilde{P}_n(x) = P_n(x) + \tilde{f}(0) + x(\tilde{f}(1) - \tilde{f}(0)) \to \tilde{f}(x) \text{ uniformly.}$$

Lemma 4. *Let* $c_n := \left(\displaystyle\int_{-1}^1 \left(1-x^2\right)^n \, \mathrm{d}x\right)^{-1} > 0$, *and let*

$$Q_n(x) = c_n \left(1-x^2\right)^n.$$

Then,

1. $\forall n,\ \displaystyle\int_{-1}^1 Q_n(x) \, \mathrm{d}x = 1.$
2. $\forall n,\ Q_n(x) \geq 0$ *on* $[-1,1]$, *and*
3. $\forall \delta \in (0,1),\ Q_n \to 0$ *uniformly on* $\delta \leq |x| \leq 1$.

Proof. 2. Immediately clear.

1. $\displaystyle\int_{-1}^1 Q_n(x) \, \mathrm{d}x = c_n \int_{-1}^1 \left(1-x^2\right)^n \, \mathrm{d}x = 1$ by definition of c_n.
3. We first estimate c_n. We have for all $n \in \mathbb{N}$ and $\forall x \in [-1,1]$,

$$\left(1-x^2\right)^n \geq 1 - nx^2.$$

It follows from the calculus

$$g(x) = \left(1 - x^2\right)^n - \left(1 - nx^2\right)$$

satisfies $g(0) = 0$, and

$$g'(x) = n \cdot 2x \left(1 - \left(1 - x^2\right)^{n-1}\right) \geq 0$$

in $[0, 1]$. Thus, $g(x) \geq 0$ by the mean value theorem. Then,

$$\begin{aligned}
\frac{1}{c_n} &= \int_{-1}^{1} \left(1 - x^2\right)^n \, \mathrm{d}x \\
&= 2 \int_{0}^{1} \left(1 - x^2\right)^n \, \mathrm{d}x \\
&> 2 \int_{0}^{1/\sqrt{n}} \left(1 - x^2\right)^n \, \mathrm{d}x \\
&\geq 2 \int_{0}^{1/\sqrt{n}} \left(1 - nx^2\right) \mathrm{d}x \\
&= 2 \left(\frac{1}{\sqrt{n}} - \frac{n}{3} \cdot n^{-3/2}\right) \\
&= \frac{4}{3} \frac{1}{\sqrt{n}} > \frac{1}{\sqrt{n}}.
\end{aligned}$$

Therefore, $c_n < \sqrt{n}$.

Let $\delta > 0$. We claim that $\lim_{n\to\infty} \sqrt{n} \left(1 - \delta^2\right)^n = 0$. Indeed,

$$\begin{aligned}
\lim_{n\to\infty} \left(\sqrt{n} \left(1 - \delta^2\right)^n\right)^{1/n} &= \lim_{n\to\infty} \left(n^{1/n}\right)^{1/2} \left(1 - \delta^2\right) \\
&= 1 - \delta^2 < 1.
\end{aligned}$$

Therefore,

$$\lim_{n\to\infty} \sqrt{n} \left(1 - \delta^2\right)^n = 0.$$

Let $\varepsilon > 0$, and choose $M \in \mathbb{N}$ such that for all $n \geq M$,

$$\sqrt{n} \left(1 - \delta^2\right)^n < \varepsilon.$$

Then, $\forall\, n \geq M$ and $\forall\, x : \delta \leq |x| \leq 1$,

$$\left|c_n \left(1 - x^2\right)^n\right| < \sqrt{n} \left(1 - x^2\right)^n \leq \sqrt{n} \left(1 - \delta^2\right)^n < \varepsilon,$$

which means $Q_n \to 0$ uniformly on $\delta \leq |x| \leq 1$. □

We now prove the Weierstrass Approximation Theorem.

Third proof of Theorem 1. Suppose $f \in C([0,1]), f(0) = f(1) = 0$. We extend f to an element of $C(\mathbb{R})$ by setting $f(x) = 0$ for all $x \notin [0,1]$. We furthermore define

$$P_n(x) = \int_0^1 f(t)Q_n(t-x)\mathrm{d}t.$$

Note that $P_n(x)$ is in fact a polynomial. Furthermore, observe that for $x \in [0,1]$

$$\begin{aligned} P_n(x) &= \int_0^1 f(t)Q_n(t-x)\mathrm{d}t \\ &= \int_{-x}^{1-x} f(x+t)Q_n(t)\mathrm{d}t \\ &= \int_{-1}^{1} f(x+t)Q_n(t)\mathrm{d}t. \end{aligned}$$

The second equality is true by a change of variable, and the last equality is true as $f(x+t) = 0$ for $t \notin [-x, 1-x]$.

We now prove $P_n \to f$ uniformly on $[0,1]$. Let $\varepsilon > 0$. Since f is uniformly continuous on $[0,1], \exists \delta > 0$ such that $\forall |x-y| \le \delta, |f(x) - f(y)| < \frac{\varepsilon}{2}$. Let $C = \sup\{f(x) \mid x \in [0,1]\}$. Choose $M \in \mathbb{N}$ such that $\forall n \ge M$,

$$\sqrt{n}\left(1-\delta^2\right)^n < \frac{\varepsilon}{8C}.$$

Thus, $\forall n \ge M, \forall x \in [0,1]$, by Lemma 4,

$$\begin{aligned} |P_n(x) - f(x)| &= \left| \int_{-1}^{1} (f(x-t) - f(x))Q_n(t)\mathrm{d}t \right| \\ &\le \int_{-1}^{1} |f(x-t) - f(x)|Q_n(t)\mathrm{d}t \\ &\le \int_{|t|\le\delta} |f(x-t) - f(x)|Q_n(t)\mathrm{d}t + \int_{\delta\le|t|\le 1} |f(x-t) - f(x)|Q_n(t)\mathrm{d}t \\ &\le \frac{\varepsilon}{2}\int_{|t|\le\delta} Q_n(t)\mathrm{d}t + \sqrt{n}\left(1-\delta^2\right)^n \int_{\delta\le|t|\le 1} 2C\ \mathrm{d}t \\ &< \frac{\varepsilon}{2} + 4C\sqrt{n}\left(1-\delta^2\right)^n < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

□

References

[W] Weierstrass, K., Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen. *Verl. d. Kgl. Akad. d. Wiss. Berlin* **2** (1885) 633-639.

School of Mathematical Sciences, Peking University, Beijing 100871, China.
Email address: `lingwang@stu.pku.edu.cn`