

Analysis I: Weierstrass Approximation

Mark Reeder Boston College

Nov 12 2016

I. Let $C([a, b])$ be the set of continuous functions $f : [a, b] \rightarrow \mathbb{R}$. Inside $C([a, b])$ we have the *polynomial functions*. These have the form $f(x) = c_0 + c_1x + \cdots + c_nx^n$, where the $c_i \in \mathbb{R}$ are constants and n is a non-negative integer. Polynomial functions are the only functions that can be written down completely. In this sense, polynomials are to continuous functions as rational numbers are to real numbers.

Informally, the Weierstrass Approximation Theorem (WAT) asserts that any continuous function on $[a, b]$ may be approximated uniformly well by a polynomial function. It is one of the most important results in Analysis.

To state the WAT precisely we recall first that $C([a, b])$ is a metric space, with distance function

$$d(f, g) = \max_{x \in [a, b]} |f(x) - g(x)|.$$

We recall as well that a subset S of a metric space M is *dense* if for all $f \in M$ there exists $\epsilon > 0$ such that $M_\epsilon(f) \cap S \neq \emptyset$. This is a precise way of saying that any point in M is arbitrarily well-approximated (according to the metric on M) by an element of S . We can now state the theorem.

Weierstrass Approximation Theorem

The set of polynomial functions on $[a, b]$ is dense in $C([a, b])$.

We can also state the WAT in this way: given a continuous function $f : [a, b] \rightarrow \mathbb{R}$ and a real number $\epsilon > 0$, there exists a polynomial function $p : [a, b] \rightarrow \mathbb{R}$ such that

$$|f(x) - p(x)| < \epsilon \quad \text{for all } x \in [a, b].$$

If we can prove the WAT for the interval $[0, 1]$, then via the bijection $[a, b] \rightarrow [0, 1]$ sending

$$x \mapsto \frac{x - a}{b - a},$$

we will have proved the WAT on $[a, b]$. So from now on we assume $[a, b] = [0, 1]$.

II. Weierstrass proved his theorem in 1885. His proof was greatly simplified by Sergei Bernstein in 1912, using the now-famous **Bernstein Polynomials**

$$b_{kn}(x) = \binom{n}{k} x^k (1-x)^{n-k}.$$

Here n is a nonnegative integer and $k = 0, 1, 2, \dots, n$. Each $b_{kn} \geq 0$ on $[0, 1]$, and is > 0 on $(0, 1)$. Except for b_{00} the function b_{kn} has a unique maximum at k/n . The factor $\binom{n}{k}$ makes the area under the graph of b_{kn} equal to $1/(n+1)$, independent of k .

Bernstein arrived at his polynomials via probability theory. They are now also widely used in computer graphics (Bezier Curves).

We will be working with linear combinations of Bernstein polynomials for a fixed n :

$$\sum_{k=0}^n c_k b_{kn}(x),$$

where c_0, c_1, \dots, c_n are constants. For this we need the identities

$$\sum_{k=0}^n b_{kn}(x) = 1, \tag{1}$$

$$\sum_{k=0}^n k b_{kn}(x) = nx, \tag{2}$$

$$\sum_{k=0}^n k^2 b_{kn}(x) = nx(nx - x + 1). \tag{3}$$

To prove these identities, start with the binomial expansion

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}. \tag{4}$$

Note that b_{kn} is obtained from $\binom{n}{k} x^k y^{n-k}$ by setting $y = 1 - x$. Setting $y = 1 - x$ in (4) then gives (1). Next, differentiate both sides of (4) with respect to x and then multiply by x . We get

$$x \cdot n(x + y)^{n-1} = \sum_{k=0}^n \binom{n}{k} k x^k y^{n-k}. \tag{5}$$

Setting $y = 1 - x$ in (5) gives (2). Do it again: differentiate both sides of (5) with respect to x and then multiply by x . We get (after using the product rule and simplifying)

$$nx(nx + y)^{n-2} = \sum_{k=0}^n \binom{n}{k} k^2 x^k y^{n-k}. \tag{6}$$

Setting $y = 1 - x$ in (6) gives (3).

Given any function $f \in C([0, 1])$, we define the **Bernstein polynomial of f** by

$$f_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) b_{kn}(x).$$

Thus f is sampled at the equally spaced points $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n} = 1$, and the coefficient of b_{kn} is the value of f at the point $\frac{k}{n}$ where b_{kn} is maximal.

We will show that $f_n \rightarrow f$ in the metric space $C([0, 1])$. Thus, the WAT will be proved by showing that f is approximated by its Bernstein polynomials.

The actual value of $f(x)$ can be written using (1), as

$$f(x) = f(x) \cdot 1 = f(x) \cdot \sum_{k=0}^n b_{kn}(x) = \sum_{k=0}^n f(x) b_{kn}(x),$$

so that

$$f(x) - f_n(x) = \sum_{k=0}^n \left[f(x) - f\left(\frac{k}{n}\right) \right] b_{kn}(x).$$

Hence by the triangle inequality,

$$|f(x) - f_n(x)| \leq \sum_{k=0}^n \left| f(x) - f\left(\frac{k}{n}\right) \right| b_{kn}(x). \quad (7)$$

Let $\epsilon > 0$ be given. Since $[0, 1]$ is compact, f is uniformly continuous on $[0, 1]$ so there exists δ such that for all $t, x \in [0, 1]$, if $|t - x| < \delta$ then $|f(t) - f(x)| < \epsilon/2$. These numbers ϵ and δ are fixed from now on.

In (7) the point x will be within δ of some sample points $\frac{k}{n}$ but not others. So we set

$$K_1 = \{k : |x - \frac{k}{n}| < \delta\}, \quad K_2 = \{k : |x - \frac{k}{n}| \geq \delta\}.$$

Now

$$\sum_{k \in K_1} \left| f(x) - f\left(\frac{k}{n}\right) \right| b_{kn}(x) < \frac{\epsilon}{2} \sum_{k \in K_1} b_{kn}(x) \leq \frac{\epsilon}{2} \sum_{k=0}^n b_{kn}(x) = \frac{\epsilon}{2},$$

using (1), and the fact that all $b_{kn}(x)$ are non-negative.

In the sum over K_2 we do not know that $|f(x) - f(\frac{k}{n})|$ is small, but at least we know it is bounded. That is, by continuity of f there is a constant M such that $|f| \leq M$ on $[0, 1]$, so

$$|f(x) - f\left(\frac{k}{n}\right)| \leq 2M.$$

If $k \in K_2$ we have $|nx - k| \geq n\delta$ so that $(nx - k)^2 \geq (n\delta)^2$. Hence

$$\begin{aligned}
\sum_{k \in K_2} |f(x) - f\left(\frac{k}{n}\right)| b_{kn}(x) &\leq 2M \sum_{k \in K_1} b_{kn}(x) \\
&\leq \frac{2M}{(n\delta)^2} \sum_{k \in K_1} (nx - k)^2 b_{kn}(x) \\
&\leq \frac{2M}{(n\delta)^2} \sum_{k=1}^n (nx - k)^2 b_{kn}(x) \\
&= \frac{2M}{(n\delta)^2} \sum_{k=1}^n (n^2 x^2 - 2n x k + k^2) b_{kn}(x) \\
&= \frac{2M}{(n\delta)^2} [n^2 x^2 - 2n x (nx) + nx(nx - x + 1)] \\
&= \frac{2M}{n\delta^2} \cdot x(1 - x),
\end{aligned}$$

On the next-to-last line we used (1)-(3). Finally, $x(1 - x) \leq 1/4$ on $[0, 1]$ so we have shown that

$$\sum_{k \in K_2} |f(x) - f\left(\frac{k}{n}\right)| b_{kn}(x) < \frac{M}{2n\delta^2}. \quad (8)$$

This last will be $< \epsilon/2$ as long as $n \geq M/(\epsilon\delta^2)$. Thus we are saved by the miraculous cancellation of the terms with n^2 , which allows n to survive in the denominator of (8).

Putting the estimates for the sums over K_1 and K_2 together, we have shown, for all $x \in [0, 1]$, that

$$n \geq \frac{M}{\epsilon\delta^2} \quad \Rightarrow \quad |f(x) - f_n(x)| < \epsilon.$$

Thus, $f_n \rightarrow f$ in $C([0, 1])$, as was to be shown, and the WAT is proved.