Chapter 10 Weierstrass Approximation Theorem

Not all functions which are infinitely differentiable admit Taylor series expansion. Even if a function has a Taylor series expansion about some point, the convergence of the series may be extremely slow for computation. But we can still approximate a continuous function by a polynomial function or polynomial for short in other ways. Computationally and ideally we would want to be able to obtain the best uniform approximation of a continuous function by a polynomial of a given degree. Usually this is not possible. Hence we seek an effective way to generate a polynomial approximation using a finite set of points on an interval say $[a, b]$, and that the polynomials obtained as the number of points used increases should converge to the given function on $[a, b]$ or the best approximation to the given function on $[a, b]$. Polynomials have been the main stay of numerical approximation, interpolation and geometric modelling and the quadrupling of computing power over the last few years have made polynomial approximation an extremely useful and powerful technique in digital computation. The starting point of finding the best polynomial approximation is the Weierstrass Approximation Theorem. It provides the basic assumption that there is a best uniform approximation of a continuous function by a polynomial of a given degree. It does not tell us how to get it. We shall present this basic result. For the topics on the characterization of best polynomial approximation, the use of trigonometric polynomials and Chebyshev polynomials for the estimate of the error function for the approximation, the intricate relation between the best approximation on an interval and the best approximation on a mesh of points in the interval, the reader is referred to "An Introduction to the Approximation of Functions" by Theodore J. Rivlin.

Classically this well known result of Weierstrass that any continuous function on a closed and bounded interval can be approximated by a polynomial function is phrased as the set of all polynomial functions defined on $[a, b]$ is dense in the *space* of all continuous functions defined on $[a, b]$ with the uniform *norm*. There is a generalization, the Stone-Weierstrass Theorem, to more general subsets of a metric space with the polynomials replaced by an appropriate family of functions and the space is the space of all continuous function on a compact topological space with the sup norm, which is known to be complete.

Theorem 1. Weierstrass Approximation Theorem.
Let I be a closed and bounded interval. Suppose $f : I \to \mathbf{R}$ is a continuous function. Then for each $\varepsilon > 0$, there exists a polynomial function $p_\varepsilon : I \to \mathbf{R}$ such that
$$|f(x) - p_\varepsilon(x)| < \varepsilon \quad \text{for all } x \text{ in } I ,$$
or equivalently sup $\{|f(x) - p_\varepsilon(x)| : x \in I\} < \varepsilon$.

One can give a proof along the lines of (1) that the polynomial functions form a subalgebra that separates points of I, (2) that the closure of this subalgebra is a lattice in $C(I, \mathbf{R})$ the space of all continuous function on I with the sup norm and (3) using the compactness of I, (1) *and* (2) one can find a point on this lattice which is arbitrarily near f. This relies on compactness and argument involving finite subcover of an open cover. This shows the advantage of working in normed linear spaces but is less constructive in nature. We shall present a proof of the theorem due

to Bernstein, using his polynomial sequences. There are many interesting problems that may be solved using Bernstein polynomials.

We shall prove a special case of Theorem 1 when $I = [0, 1]$ first. We now describe the Bernstein polynomials.

Let $f : [0,1] \to \mathbf{R}$ be a function. Then for each integer $n \geq 0$, we define the *Bernstein polynomial* of degree n associated with f to be

$$B_n(f)(x) = \sum_{k=0}^{n} f\left(\tfrac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}$$

Theorem 2.

Suppose $f : [0,1] \to \mathbf{R}$ is a continuous function. Then for each $\varepsilon > 0$, there exists a polynomial function $p : I \to \mathbf{R}$ such that

$$\sup\{|f(x) - p(x)| : x \in I\} < \varepsilon .$$

More specifically, the sequence of Bernstein polynomial $(B_n(f))$ as defined above converges uniformly to f.

Before we proceed with the proof, we shall derive a series of identities which are needed for the proof. The binomial theorem states that for integer $n \geq 0$,

$$(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k} \qquad\text{----------------------- (A)}$$

Hence from (A), for integer $n \geq 1$,

$$(x+y)^{n-1} = \sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-k-1} \qquad\text{----------------------- (1)}$$

Multiply (1) by nx, we obtain

$$nx(x+y)^{n-1} = \sum_{k=0}^{n-1} n\binom{n-1}{k}x^{k+1}y^{n-k-1} \qquad\text{----------------------- (2)}$$

Now, $n\binom{n-1}{k} = n \cdot \dfrac{(n-1)!}{k!(n-1-k)!} = (k+1)\dfrac{n!}{(k+1)!(n-1-k)!} = (k+1)\binom{n}{k+1}$ and so from (2), we have

$$nx(x+y)^{n-1} = \sum_{k=0}^{n-1} n\binom{n-1}{k}x^{k+1}y^{n-k-1} = \sum_{k=0}^{n-1}(k+1)\binom{n}{k+1}x^{k+1}y^{n-k-1}$$
$$= \sum_{k=1}^{n} k\binom{n}{k}x^k y^{n-k} = \sum_{k=0}^{n} k\binom{n}{k}x^k y^{n-k}$$

Evidently, the above equality is true when $n = 0$ and so we have that for any integer $n \geq 0$,

$$nx(x+y)^{n-1} = \sum_{k=0}^{n} k\binom{n}{k}x^k y^{n-k} \qquad\text{--------------------- (B)}$$

From (B) for integer $n \geq 1$,

$$(n-1)x(x+y)^{n-2} = \sum_{k=0}^{n-1} k\binom{n-1}{k}x^k y^{n-k-1} \qquad\text{----------------- (3)}$$

Multiply (3) by nx, we obtain similarly,

$$n(n-1)x^2(x+y)^{n-2} = \sum_{k=0}^{n-1} kn\binom{n-1}{k}x^{k+1}y^{n-k-1} = \sum_{k=0}^{n-1} k(k+1)\binom{n}{k+1}x^{k+1}y^{n-k-1}$$
$$= \sum_{k=1}^{n}(k-1)k\binom{n}{k}x^k y^{n-k} = \sum_{k=0}^{n}(k-1)k\binom{n}{k}x^k y^{n-k} .$$

Plainly, the above equality also holds when $n = 0$, and so we have for any integer $n \geq 0$,

$$n(n-1)x^2(x+y)^{n-2} = \sum_{k=0}^{n}(k-1)k\binom{n}{k}x^k y^{n-k} \qquad\text{------------ (C)}$$

Note that the identities (A), (B) and (C) contain the same factor $\binom{n}{k}x^k y^{n-k}$.

Let $r_k(x) = \binom{n}{k}x^k y^{n-k}$.

Now, taking y to be $1 - x$, so that $x + y = 1$, we let $r_k(x) = \binom{n}{k}x^k y^{n-k} = \binom{n}{k}x^k(1-x)^{n-k}$

. We obtain from (A), $1 = (x+y)^n = \sum_{k=0}^{n} r_k(x)$. That is, for any integer $n \geq 0$,

$$1 = \sum_{k=0}^{n} r_k(x) \qquad \text{-----------------------------} \quad (D)$$

Similarly from (B), we obtain, for any integer $n \geq 0$

$$nx = \sum_{k=0}^{n} k\, r_k(x) \qquad \text{-----------------------------} \quad (E)$$

and from (C) we obtain, for any integer $n \geq 0$

$$n(n-1)x^2 = \sum_{k=0}^{n}(k-1)k\, r_k(x) \qquad \text{--------------------} (F)$$

Then for any integer $n \geq 0$,

$$\sum_{k=0}^{n}(k-nx)^2\, r_k(x) = \sum_{k=0}^{n} n^2 x^2\, r_k(x) - 2nx \sum_{k=0}^{n} k\, r_k(x) + \sum_{k=0}^{n} k^2\, r_k(x)$$

$$= n^2 x^2 \sum_{k=0}^{n} r_k(x) - 2nx \sum_{k=0}^{n} k\, r_k(x) + \sum_{k=0}^{n}[(k-1)k + k]\, r_k(x)$$

$$= n^2 x^2 - 2nx \cdot nx + nx + n(n-1)x^2 = nx(1-x)$$

by (D), (E) and (F).

Therefore, for any integer $n \geq 0$,

$$\sum_{k=0}^{n}(k-nx)^2\, r_k(x) = nx(1-x). \qquad \text{-----------} \quad (G)$$

We now proceeds to the proof of Theorem 2.

Proof of Theorem 2.

Since f is continuous and [0, 1] is compact, by Theorem 7 Chapter 3 the image f ([0,1]) is compact and so by the Heine-Borel Theorem (*see Theorem 43 Chapter 2*) f ([0, 1]) is closed and bounded. Therefore, there exists a real number $M > 0$ such that
$$|f(x)| \leq M \text{ for all } x \text{ in } [0,1].$$

Given $\varepsilon > 0$, since f is uniformly continuous on [0, 1] (see Theorem 29 Chapter 3), there exists $\delta > 0$ such that for all x, y in [0, 1],
$$|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon/2. \qquad \text{------------} \quad (1)$$

We now estimate how close the Bernstein polynomial $B_n(f)$ is to f for integer $n \geq 1$.

$$|f(x) - B_n(f)(x)| = \left| f(x) - \sum_{k=0}^{n} f(\tfrac{k}{n})\binom{n}{k}x^k(1-x)^{n-k} \right|$$

$$= \left| f(x) - \sum_{k=0}^{n} f(\tfrac{k}{n}) r_k(x) \right|$$

$$= \left| f(x) \sum_{k=0}^{n} r_k(x) - \sum_{k=0}^{n} f(\tfrac{k}{n}) r_k(x) \right| \qquad \text{using identity (D)}$$

$$= \left| \sum_{k=0}^{n}\left(f(x) - f(\tfrac{k}{n}) \right) r_k(x) \right|. \qquad \text{------------------} \quad (2)$$

We next examine the sum on the right of (2) according to whether $|x - \tfrac{k}{n}| < \delta$ or $|x - \tfrac{k}{n}| \geq \delta$, where δ is given in (1).

If $|x - \tfrac{k}{n}| < \delta$, then by (1),

$$\left| f(x) - f(\tfrac{k}{n}) \right| < \tfrac{\varepsilon}{2}. \qquad \text{------------------ (3)}$$

Suppose now $|x - \tfrac{k}{n}| \geq \delta$. Then $|nx - k| \geq n\delta$. Hence

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

$$\leq 2M \frac{(nx - k)^2}{n^2 \delta^2} \qquad \text{-------------------- (4)}$$

because $\dfrac{|nx - k|}{n\delta} \geq 1$.

Therefore, for any x in [0,1] and for $0 \leq k \leq n$.

$$\left| f(x) - f(\tfrac{k}{n}) \right| \leq \tfrac{\varepsilon}{2} + \frac{2M}{\delta^2}\left(x - \tfrac{k}{n}\right)^2. \qquad \text{-------------- (5)}$$

(We add the term $\tfrac{\varepsilon}{2}$ so that we can combine (3) and (4) in one inequality for simplicity.)

Using (2) and the fact that $r_k(x) \geq 0$ for all x in [0,1] and $0 \leq k \leq n$, we get

$$|f(x) - B_n(f)(x)| = \left| \sum_{k=0}^{n} \left(f(x) - f(\tfrac{k}{n}) \right) r_k(x) \right|$$

$$\leq \sum_{k=0}^{n} \left| \left(f(x) - f(\tfrac{k}{n}) \right) \right| r_k(x)$$

$$\leq \sum_{k=0}^{n} \left(\tfrac{\varepsilon}{2} + \frac{2M}{\delta^2}\left(x - \tfrac{k}{n}\right)^2 \right) r_k(x) \qquad \text{by inequality (5)}$$

$$= \tfrac{\varepsilon}{2} \sum_{k=0}^{n} r_k(x) + \frac{2M}{\delta^2} \sum_{k=0}^{n} \left(x - \tfrac{k}{n}\right)^2 r_k(x)$$

$$= \tfrac{\varepsilon}{2} + \frac{2M}{\delta^2 n^2} \sum_{k=0}^{n} (nx - k)^2 r_k(x) \qquad \text{by identity (D)}$$

$$= \tfrac{\varepsilon}{2} + \frac{2M}{\delta^2 n^2} nx(1 - x) \qquad \text{by identity (G)}$$

$$= \tfrac{\varepsilon}{2} + \frac{2M}{\delta^2 n} x(1 - x)$$

$$< \tfrac{\varepsilon}{2} + \frac{2M}{\delta^2 n}$$

because $x(1-x) < 1$ for x in [0,1].

Hence, for any x in [0,1] and any $n \geq 1$,

$$|f(x) - B_n(f)(x)| < \tfrac{\varepsilon}{2} + \frac{2M}{\delta^2 n} \qquad \text{-------------- (6)}$$

Since $\dfrac{2M}{\delta^2 n} \to 0,$ there exists a positive integer N such that

$$n \geq N \Rightarrow \frac{2M}{\delta^2 n} < \tfrac{\varepsilon}{4}.$$

It then follows from (6) that

$$n \geq N \Rightarrow |f(x) - B_n(f)(x)| < \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{4} = \tfrac{3\varepsilon}{4} \;\; \text{for all } x \text{ in } [0,1].$$

Hence,

$$n \geq N \Rightarrow \sup\{ |f(x) - B_n(f)(x)| : x \in [0, 1]\} \leq \tfrac{3\varepsilon}{4} < \varepsilon.$$

This shows that $B_n(f) \to f$ uniformly on [0,1]. We may take the polynomial function p to be $B_N(f)$ and $\sup\{ |f(x) - p(x)| : x \in [0, 1]\} < \varepsilon$. This completes the proof of Theorem 2.

Proof of Theorem 1.

Suppose $I = [a, b]$ a closed and bounded interval and $f : I \to \mathbf{R}$ is a continuous function. Let g : $[0,1] \to [a, b]$ be the bijective linear map defined by $g(t) = a + t(b-a)$ for t in [0,1]. Then g is continuous, $g(0) = a$ and $g(1) = b$. Since f is continuous, the

composite $f \circ g : [0,1] \to \mathbf{R}$ is also continuous. Hence by Theorem 2, for any $\varepsilon > 0$, there exists a positive integer N such that for any integer $n \geq N$, the Bernstein polynomial $B_n (f \circ g)$ satisfies

$$|f \circ g(x) - B_n(f \circ g)(x)| < \varepsilon \quad \text{for all } x \text{ in } [0, 1]. \quad \text{----------------- (1)}.$$

Now g is a continuous injective map and so g has a continuous inverse function. (see Theorem 23 Chapter 3). Indeed the inverse function $g^{-1} : [a, b] \to [0, 1]$ is given by

$$g^{-1}(t) = \frac{t-a}{b-a} \quad \text{for } t \text{ in } [a, b] .$$

Thus by (1) for all t in $[a, b]$,

$$|f(t) - B_N(f \circ g)(g^{-1}(t))| < \varepsilon.$$

Hence

$$\left| f(t) - B_N(f \circ g)(\frac{t-a}{b-a}) \right| < \varepsilon \quad \text{for all } t \text{ in } [a, b].$$

Since $B_N (f \circ g)$ is a polynomial function, $p_\varepsilon(t) = B_N(f \circ g)(\frac{t-a}{b-a})$ is a polynomial function in t and $|f(t) - p_\varepsilon(t)| < \varepsilon$ for all t in I,

If we let $q_n(t) = B_n(f \circ g)(\frac{t-a}{b-a})$, then

$$q_n(t) = B_n(f \circ g)(\frac{t-a}{b-a}) = \sum_{k=0}^{n} f \circ g(\frac{k}{n})\binom{n}{k}(\frac{t-a}{b-a})^k (1 - (\frac{t-a}{b-a}))^{n-k}$$

$$= \sum_{k=0}^{n} f \circ g(\frac{k}{n})\binom{n}{k}\left(\frac{t-a}{b-a}\right)^k \left(\frac{b-t}{b-a}\right)^{n-k}$$

$$= \sum_{k=0}^{n} f\left(a + \frac{k}{n}(b-a)\right)\binom{n}{k}\left(\frac{t-a}{b-a}\right)^k \left(\frac{b-t}{b-a}\right)^{n-k} .$$

It follows from (1) that $q_n \to f$ uniformly on $[a, b]$.

Remark.

1. The Weierstrass Approximation Theorem uses the uniform continuity of f which depends on the compactness of the interval $[a, b]$. In general, any continuous function defined on a non-compact interval (or domain) need not be uniformly continuous. Since the only compact intervals of $\mathbf{R}$ are the closed and bounded interval by the Heine-Borel Theorem, we cannot hope to extend the approximation result to any continuous function to the whole of $\mathbf{R}$. However, in practice, we use the approximation theorem in a closed and bounded interval containing the region of interest.

2. A more general result is the Stone-Weierstrass Theorem for compact metric spaces. There is still the more general Theorem in the space of all continuous real valued function on a Hausdorff topological space with the compact open topology. For the metric space version, assume A is a compact metric space, let $C(A, \mathbf{R})$ be the set of all continuous real valued functions on A with the sup metric. Let $\mathscr{B} \subseteq C(A, \mathbf{R})$ be such that it is an algebra, i.e., $f, g \in \mathscr{B}$ and $\alpha \in \mathbf{R}$ implies that, $f + g$, $f g$, and $\alpha f \in \mathscr{B}$. If $\mathscr{B}$ contains 1, the constant function and is separating, i.e., for all a, b in A , $a \neq b$ implies that there exists a function f in $\mathscr{B}$ such that $f(a) \neq f$ (b), then $\mathscr{B}$ is dense in $C(A, \mathbf{R})$, i.e., the closure of $\mathscr{B}$ in $C(A, \mathbf{R})$ is $C(A, \mathbf{R})$. This gives an approximation on more abstract compact metric spaces other than the closed and bounded interval in $\mathbf{R}$. Note that when A is compact, the space $C(A, \mathbf{R})$ with the compact open topology and that with the sup metric coincide.

Example of an application of the Weierstrass Theorem.

Here is an application of the Weierstrass Approximation Theorem.

Suppose f is a continuous function defined on [0, 1]. Suppose f satisfy that $\int_0^1 f(x)x^n dx = 0$ for all integer $n \geq 0$. Then f is identically the 0 constant function. By Theorem 2, the sequence of Bernstein polynomial ($B_n(f)$) converges uniformly to f. By the supposition $\int_0^1 f(x)B_n(f)(x)dx = 0$. Since f is continuous and so f is bounded on [0, 1], it follows that $f B_n(f)$ converges uniformly to f^2. Therefore, either by Theorem 7 Chapter 8 or Theorem 1 Chapter 9,

$$\int_0^1 f^2(x)dx = \lim_{n \to \infty} \int_0^1 f(x)B_n(f)(x)dx = 0.$$

Therefore, since $f^2 \geq 0$, $\int_0^1 f^2(x)dx = 0$ implies that f^2 is identically the zero function on [0, 1]. It follows that f is identically zero.

Exercises 3

1. Prove that if f is increasing on [0, 1], then the Bernstein polynomial $B_n(f)$ is also increasing on [0, 1].

2. Show that if f is continuous on [0, 1], then there is a sequence of polynomial functions $p_n(x)$ such that $f(x) = \sum_{n=1}^{\infty} p_n(x)$. [Hint: use Weierstrass Approximation Theorem.]

3. Suppose f is continuous on [0, 1]. Prove that if $\int_0^1 f(x)x^n dx = 0$ for $n = 0, 1, 2,$ … , then f is identically 0.

4. Prove that the function $f : (a, b) \to \mathbf{R}$ defined by $f(x) = 1/(b - x)$ cannot be approximated uniformly by polynomials on (a, b).

5. Prove that the function $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) = e^x$ cannot be approximated uniformly by polynomials on $\mathbf{R}$.

6. Suppose $f : [0, 1] \to \mathbf{R}$ is bounded. Modify the proof of Theorem 2 to show that if $f : [0, 1] \to \mathbf{R}$ is continuous at x_0 in [0, 1], then $B_n(f)(x_0) \to f(x_0)$. Hence, or otherwise, show that if f is continuous on a closed interval A in [0, 1], then $B_n(f)(x)$ converges to f uniformly on A.

   [Hint show that if f is continuous at x_0, then given any $\varepsilon > 0$ there is a neighbourhood B of x_0 in [0, 1], i.e. the intersection of an open interval containing x_0 with [0, 1] and an integer N such that for all integer $n \geq N$ and for all x in B, $|B_n(f)(x) - f(x)| < \varepsilon$. If A is compact and f is continuous at each point of A, obtain a covering of A by such neighbourhoods and use compactness to deduce uniform convergence on A.]

7. Suppose f is continuous on [0, 1]. Prove that if $\int_0^1 f(x)(x^{n+1}(1 - x)^2)'' dx = 0$ for all integers $n \geq 1$, then f is a linear function.

8. Suppose f is continuous on $[0, 1]$ and $\int_0^1 f(x)g''(x)dx = 0$ for any function g : $[0, 1] \to \mathbf{R}$ which is continuously twice differentiable on $[0,1]$ and equal to zero in neighbourhoods of the points 0 and 1. Prove that f is a linear function.

9. Suppose $f:[a, b] \to \mathbf{R}$ is continuous, prove that
$$\lim_{x \to \infty} \int_a^b f(t)\sin(xt)dt = 0.$$

10. Without using integration, prove that any continuous function on the closed and bounded interval $[a, b]$ has an anti-derivative on $[a, b]$.

11. Suppose $f:[0, 1] \to \mathbf{R}$ is continuous. Prove that if $f:[0, 1] \to \mathbf{R}$ is differentiable at x_0 in $[0, 1]$, then $(B_n(f))'(x_0) \to f'(x_0)$.

12. Suppose $f:[0, 1] \to \mathbf{R}$ is continuous. Prove that if f is continuously differentiable on $[0, 1]$, then $(B_n(f))'(x) \to f'(x)$ uniformly on $[0,1]$ as $n \to \infty$. Moreover, if f is C^r, i.e., r-times continuously differentiable on $[0, 1]$, then $(B_n(f))^{(r)}(x) \to f^{(r)}(x)$ uniformly on $[0,1]$ as $n \to \infty$.