Tchebycheff's Approximation to Polyomials

Steve Broeker, B.S.
School of Mathematics
Pacifc Lutheran University
Tacoma, Washington

March 5, 1982

Abstract

Tchebycheff was a Russian mathematician. He did most of his work on approximation theory in 1894.

The problem that I wish to address is that I want to approximate a continuous function f defined on an interval [a, b] by a polynomial:

$$P(x) = C_n X^n + C_{n-1} X^{n-1} + \ldots + C_0$$

I want to evaluate this approximation by minimizing expressions of the form:

1. $max\ |f(x) - p(x)|$

   $a \leq x \leq b$

2. $max\ |f(x_i) - p(x_i)|$

   $1 \leq i \leq m$
   $a \leq x_i \leq b$

1 Background

This thesis is submitted in partial fullfillment of the requirements for the degree of Bachelor's of Science in Mathematics at Pacific Lutheran University.

1.1 Introduction

Tchebycheff was a Russian mathematician. He did most of his work on approximation theory in 1894.

The problem that I wish to address is that I want to approximate a continuous function *f* defined on an interval *[a, b]* by a polynomial:

$$P(x) = C_n X^n + C_{n-1} X^{n-1} + \ldots + C_0$$

I want to evaluate this approximation by minimizing expressions of the form:

1. max $|f(x) - p(x)|$

   $a \leq x \leq b$

2. max $|f(x_i) - p(x_i)|$

   $1 \leq i \leq m$
   $a \leq x_i \leq b$

A large part of the work of Tchebycheff involved the very special case when the number of points taken is equal to *n+1*. This topic is called interpolation.

We know that a straight line having the equation $y = ax + b$ can be passed through any two points having distinct abscissas.

Similarly, a parabola $y = ax^2 + bx + c$ can be passed through any three points having distinct abscissas.

I would now like to generalize this concept for *n +1* points. But I first need to do some preliminary work.

2 Vandermonde's Determinant

Theorem:

$$\prod_{0 \leq j < i \leq n}(X_i - X_j) = \begin{vmatrix} 1 & X_0 & X_0^2 & ... & X_0^n \\ 1 & X_1 & X_1^2 & ... & X_1^n \\ . & . & . & ... & . \\ 1 & X_n & X_n^2 & ... & X_n^n \end{vmatrix}$$

Proof by induction:

I will first show that Vandermonde's Determinant (from here on out this determimiant will be referred to as: D_T) holds for *n=1*. For *n=1* we have:

$$\begin{vmatrix} 1 & X_0 \\ 1 & X_1 \end{vmatrix} = (X_1 - X_0)$$

Therefore D_T holds for *n=1*.

Assuming that D_T is true for *n=r*, I must now prove that D_T holds for *n=r+1*. For *n=r+1* we have:

$$\begin{vmatrix} 1 & X_0 & X_0^2 & ... & X_0^r & X_0^{r+1} \\ 1 & X_1 & X_1^2 & ... & X_1^r & X_1^{r+1} \\ . & . & . & ... & . & . \\ 1 & X_{r+1} & X_{r+1}^2 & ... & X_{r+1}^r & X_{r+1}^{r+1} \end{vmatrix}$$

$$= \begin{vmatrix} 1 & X_0 & X_0^2 & ... & X_0^r & X_0^{r+1} - X_0^{r+1} \\ 1 & X_1 & X_1^2 & ... & X_1^r & X_1^{r+1} - X_1^r X_0 \\ . & . & . & ... & . & . \\ 1 & X_{r+1} & X_{r+1}^2 & ... & X_{r+1}^r & X_{r+1}^{r+1} - X_{r+1}^r X_0 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & X_0 & X_0^2 & ... & X_0^r & 0 \\ 1 & X_1 & X_1^2 & ... & X_1^r & X_1^r(X_1 - X_0) \\ . & . & . & ... & . & . \\ 1 & X_{r+1} & X_{r+1}^2 & ... & X_{r+1}^r & X_{r+1}^r(X_{r+1} - X_0) \end{vmatrix}$$

$$= \begin{vmatrix} 1 & X_0 & X_0^2 & ... & X_0^r - X_0^r & 0 \\ 1 & X_1 & X_1^2 & ... & X_1^r - X_1^{r-1} X_0 & X_1^r(X_1 - X_0) \\ . & . & . & ... & . & . \\ 1 & X_{r+1} & X_{r+1}^2 & ... & X_{r+1}^r - X_{r+1}^{r-1} X_0 & X_{r+1}^r(X_{r+1} - X_0) \end{vmatrix}$$

$$= \begin{vmatrix} 1 & X_0 & X_0^2 & ... & 0 & 0 \\ 1 & X_1 & X_1^2 & ... & X_1^{r-1}(X_1 - X_0) & X_1^r(X_1 - X_0) \\ . & . & . & ... & . & . \\ 1 & X_{r+1} & X_{r+1}^2 & ... & X_{r+1}^{r-1}(X_{r+1} - X_0) & X_{r+1}^r(X_{r+1} - X_0) \end{vmatrix}$$

Continuing this process we arrive at:

$=$

$$\begin{vmatrix} 1 & 0 & 0 & ... & 0 & 0 \\ 1 & (X_1 - X_0) & X_1(X_1 - X_0) & ... & X_1^{r-1}(X_1 - X_0) & X_1^r(X_1 - X_0) \\ . & . & . & ... & . & . \\ 1 & (X_{r+1} - X_0) & X_{r+1}(X_{r+1} - X_0) & ... & X_{r+1}^{r-1}(X_{r+1} - X_0) & X_{r+1}^r(X_{r+1} - X_0) \end{vmatrix}$$

$=$

$$\begin{vmatrix} (X_1 - X_0) & X_1(X_1 - X_0) & ... & X_1^{r-1}(X_1 - X_0) & X_1^r(X_1 - X_0) \\ (X_2 - X_0) & X_2(X_2 - X_0) & ... & X_2^{r-1}(X_2 - X_0) & X_2^r(X_2 - X_0) \\ . & . & ... & . & . \\ (X_{r+1} - X_0) & X_{r+1}(X_{r+1} - X_0) & ... & X_{r+1}^{r-1}(X_{r+1} - X_0) & X_{r+1}^r(X_{r+1} - X_0) \end{vmatrix}$$

$= (X_1 - X_0)$ multiplied by:

$$\begin{vmatrix} 1 & X_1 & ... & X_1^{r-1} & X_1^r \\ (X_2 - X_0) & X_2(X_2 - X_0) & ... & X_2^{r-1}(X_2 - X_0) & X_2^r(X_2 - X_0) \\ . & . & ... & . & . \\ (X_{r+1} - X_0) & X_{r+1}(X_{r+1} - X_0) & ... & X_{r+1}^{r-1}(X_{r+1} - X_0) & X_{r+1}^r(X_{r+1} - X_0) \end{vmatrix}$$

$= (X_1 - X_0)(X_2 - X_0)$ multiplied by:

$$\begin{vmatrix} 1 & X_1 & ... & X_1^{r-1} & X_1^r \\ 1 & X_2 & ... & X_2^{r-1} & X_2^r \\ . & . & ... & . & . \\ (X_{r+1} - X_0) & X_{r+1}(X_{r+1} - X_0) & ... & X_{r+1}^{r-1}(X_{r+1} - X_0) & X_{r+1}^r(X_{r+1} - X_0) \end{vmatrix}$$

Continuing this process we arrive at:

$$= (X_1 - X_0)(X_2 - X_0)...(X_{r+1} - X_0) \begin{vmatrix} 1 & X_1 & ... & X_1^{r-1} & X_1^r \\ 1 & X_2 & ... & X_2^{r-1} & X_2^r \\ . & . & ... & . & . \\ 1 & X_{r+1} & ... & X_{r+1}^{r-1} & X_{r+1}^r \end{vmatrix}$$

$$= \prod_{1 \leq i \leq r+1} (X_i - X_0) \begin{vmatrix} 1 & X_1 & ... & X_1^{r-1} & X_1^r \\ 1 & X_2 & ... & X_2^{r-1} & X_2^r \\ . & . & ... & . & . \\ 1 & X_{r+1} & ... & X_{r+1}^{r-1} & X_{r+1}^r \end{vmatrix}$$

The matrix on the right is nothing more than Vandermonde's Determinant for r variables. In this case the variables have been numbered from *1* to *r+1* instead of *0* to r. Thus by my assumption that D_T holds for n=r, we have:

$$= \left[\prod_{1 \leq i \leq r+1} (X_i - X_0)\right] \left[\prod_{0 \leq j < i \leq n} (X_i - X_j)\right]$$

$$= \prod_{0 \leq i \leq r+1} (X_i - X_j)$$

Therefore D_T is true for n = *r+1*. Thus (by induction) my proof is complete.

3 Interpolation Theorem

Theorem:

There exists a unique polynomial of degree $\leq n$ which assumes prescribed values at $n+1$ distinct points.

Proof:

Let $(x_0, x_1, ..., x_n)$ be the points and $(y_0, y_1, ..., y_n)$ be the prescribed values. We seek a polynomial p such that $p(x_i) = y_i (i = 0, 1, ..., n)$. Since the polynomial is of degree $\leq n$, it may be expressed as:

$$P(X) = \sum_{j=0}^{n} C_j X^j$$

Hence our requirement now reads:

$$P(X_i) = \sum_{j=0}^{n} C_j X_i^j = Y_i (i = 0, 1, ..., n).$$

Written out in matrix from this becomes:

$$\begin{bmatrix} 1 & X_0 & X_0^2 & ... & X_0^n \\ . & . & ... & . & . \\ 1 & X_n & X_n^2 & ... & X_n^n \end{bmatrix} \begin{bmatrix} C_0 \\ . \\ . \\ . \\ C_n \end{bmatrix} = \begin{bmatrix} Y_0 \\ . \\ . \\ . \\ Y_n \end{bmatrix}$$

In this equation the C matrix is unknown while the X and Y matricies are known.

The determinant of the X matrix equals Vandermonde's Determinant and thus has the value:

$$D_T = \prod_{0 \leq j < i \leq n} (X_i - X_j)$$

Since each of the X_i*'s* are distinct, $Det \neq 0$. Thus the matrix has a unique solution and my proof is complete.

4 Interpolation Process

I will now seek to asses the interpolation process as an instrument of approximation. This examination will pertain to our two original expressions:

1. max $|f(x) - p(x)|$

   $a \leq x \leq b$

2. $\max |f(x_i) - p(x_i)|$

$1 \leq i \leq m$

$a \leq x_i \leq b$

The polynomial p of degree $\leq n$ which interpolates to f at $n+1$ points x_i, clearly solves the problem of minimizing the second equation when $m = n+1$.

I will now ask, will the first expression also be small when p is chosen in this way. The answer is certainly not if the behavior of f between the interpolating points is not somehow controlled. It turns out that such control is possible for functions which possess $n+1$ continuous derivatives. Before I address this problem, we need to familiarize ourselves with the Tchebycheff norm, which for a polynomial y defined on the interval $[a,b]$, is:

$$||y||_T = max_{a \leq x \leq b}|y(x)|$$

To show that this is indeed a norm I must show, for polynomials y and x, that:

1. $||y|| > 0$ (unless $y = 0$)

2. $||\lambda y|| = |\lambda| \; ||y||$ (λ is a scalar)

3. $||y + z|| \leq ||y|| + ||z||$

The Tchebycheff norm obviously fits this definition.

5 Theorem 1

Theorem:

If f possesses n continuous derivatives on $[a, b]$. And if p is the polynomial of degree $< n$ which interpolates to f at n nodes x_i in $[a, b]$,

and if $w(x) = \prod (x - x_i)$, then in terms of the Tchebycheff norm:

$$||f - p|| \leq \frac{1}{n!} \; ||f^{(n)}|| \; ||w||$$

Proof:

I will first show to each y in $[a, b]$ there corresponds a $z_y \in [a, b]$ such that:

$$f(y) - p(y) = \frac{1}{n!} \; f^{(n)}(z_y) w(y)$$

This formula is obvious if y is one of the nodes. Otherwise we put:

$$\phi = f - p - \lambda w$$

where λ is chosen to make $\phi(y) = 0$. Namely:

$$\phi(y) = f(y) - p(y) - \lambda w(y)$$

$$0 = f(y) - p(y) - \lambda w(y)$$

$$\lambda = \frac{f(y)-p(y)}{w(y)}$$

It is clear that ϕ vanishes also at the nodes x_i for:

$$\phi(x_i) = f(x_i) - p(x_i) - \lambda x(x_i) = 0$$

Thus ϕ vanishes in at least $n+1$ points of $[a, b]$, the n nodes and the point y. Rolle's theorem states that for $f(x) \in c[a, b]$ and f differentiable at each point of $[a, b]$. If $f(a) = f(b)$ then there is a point $x = \beta$ with $a < \beta < b$ for which $f'(\beta) = 0$.

Thus since f possesses n continuous derivatives on $[a, b]$, ϕ' vanishes at least once between any two zeros of ϕ and thus vanishes in at least n points.

Also, ϕ'' vanishes at least once between any two zeros of ϕ' and thus vanishes in at least n-1 points. Continuing this argument, we see that $\phi^{(n)}$ has at least one root on the interval $[a, b]$, say at the point zy. By differentiating ϕ with respect to x and remembering that y is a fixed point we have:

$$\phi^{(n)} = f^{(n)} - p^{(n)} - \lambda w^{(n)}$$

And since p is a polynomial of degree $< n$ we have:

$$\phi^{(n)} = f^{(n)} - 0 - \lambda w^{(n)}$$

And since $w(s) = s^n + s^{n-1} + ...$ we have:

$$\phi^{(n)} = f^{(n)} - \lambda n!$$

Thus:

$$f^{(n)}(z_y) = \lambda n!$$

And since:

$$\lambda = \frac{f(y)-p(y)}{w(y)}$$

we have:

$$f^{(n)}(z_y) = \frac{[f(y)-p(y)]n!}{w(y)}$$

$$f(y) - p(y) = \frac{1}{n!}\ f^{(n)}(z_y)w(y)$$

$$|f(y) - p(y)| = |\frac{1}{n!}\ f^{(n)}(z_y)w(y)|$$

Remember, the preceeding is true for all $y \in [a,\ b]$.

Now assume that:

$$max_{a \leq y \leq b}|f(y) - p(y)| \neq max_{a \leq y \leq b}\ |\frac{1}{n!} f^{(n)}(z_y)w(y)|$$

Without loss of generality I can assume that:

$$max_{a \leq y \leq b}|f(y) - p(y)| < max_{a \leq y \leq b}\ |\frac{1}{n!} f^{(n)}(z_y)w(y)|$$

Now for $\alpha \in [a,\ b]$ let:

$$|\frac{1}{n!} f^{(n)}(z_\alpha)w(\alpha)| = max_{a \leq y \leq b}\ |\frac{1}{n!} f^{(n)}(z_y)w(y)|$$

then:

$$> max_{a \leq y \leq b}|f(y) - p(y)|$$

QED

Thus:

$$max_{a \leq y \leq b}|f(y) - p(y)| = max_{a \leq y \leq b}\ |\frac{1}{n!} f^{(n)}(z_y)w(y)|$$

$$||f(y) - p(y)|| = \frac{1}{n!}\ ||f^{(n)}(z_y)w(y)||$$

$$\leq \frac{1}{n!}\ ||f^{(n)}(z_y)||\ ||w(y)||$$

$$\leq \frac{1}{n!}\ ||f^{(n)}||\ ||w(y)||$$

A question that is raised in a natural way by the foregoing theorem is how can we situate the nodes as to optimize the error bound? Since the nodes enter this formula only in the function w, I must attempt to minimize the norm of w.

I will now prove a relationship which will be immediately useful.

6 Theorem 2

Theorem:

$$\sum_{k=0}^{n} A_k cos^k\theta \ = cosn\theta$$

With appropriate coefficients A_k, the leading one, $A_n \ = 2^{n-1}$.

Proof by induction:

I will first show that the equation holds for n = *1*.

$$cos(1 - \theta) \ = cos\theta$$

$$= \textit{0 + (1 x cos } \theta\textit{)}$$

$$= \sum_{k=0}^{1} A_k cos^k\theta$$

The leading coefficient $A_1 = 1 = 2^0 = 2^{1-1} = 2^{n-1}$. Therefore, the equation holds for *n=1*.

I will now assume that the formula is true for *n=r* and the leading coefficient $A_r = 2^{r-1}$. I need to show that the relationship is true for *n=r+1*. Thus:

$$cos(r + 1)\theta \ = cos(r\theta \ + \theta)$$

$$= cosr\theta cos\theta \text{ - } sinr\theta sin\theta$$

Since:

$$= cos(A \pm B) = cosAcosB \pm sinAsinB$$

We have:

$$= 2cosr\theta cos\theta \text{ - } cosr\theta cos\theta \text{ - } sinr\theta sin\theta$$

$$= 2cosr\theta cos\theta \text{ - } cos(r \text{ - } 1)\theta$$

$$= 2cos\theta \ \sum_{k=0}^{r} A_k cos^k\theta - \sum_{k=0}^{r-1} B_k cos^k\theta$$

$$= \sum_{k=0}^{r} 2A_k cos^{k+1}\theta - \sum_{k=0}^{r-1} B_k cos^k\theta$$

$$= 2A_k cos^{r+1}\theta + 2A_{k+1} cos^r\theta + \sum_{k=0}^{r-1}(2A_k - B_k)cos^k\theta$$

$$= \sum_{k=0}^{r+1} C_k cos^k\theta$$

Where (C_{r+1} = $2A_r$), (C_r = $2A_{r-1}$), and (C_i = $2A_i$ - B_i) for ($0 \leq i \leq r-1$).
The relationship thus holds for *n=r+1*. Therefore (by induction), my proof is complete.

7 Theorem 3

Theorem:

The norm of:

$$w(x) = \prod_{i=1}^{n}(X - X_i)$$

is minimized on *[-1, 1]* when:

$$x_i = cos[(2i-1)\frac{\Pi}{2n}]$$

Proof:

Letting:

$$T_n(x) = \sum_{k=0}^{n} A_k X^k$$

We have:

$$T_n(cos\theta) = cosn\theta$$

To obtain the roots of T_n we set:

$$T_n(cos\theta) = 0$$

We thus have:

$$T_n(cos\theta) = cosn\theta = 0$$

$$n\theta = arccos0$$

$$n\theta = \frac{(2i-1)\Pi}{2} \quad (i = 1,2, \ldots)$$

$$\theta = \frac{(2i-1)\Pi}{2n}$$

$$cos\theta = cos[(2i-1)\frac{\Pi}{2n}]$$

Thus the roots of T_n are the x_i given above. The polynomial: $U = 2^{1-n}T_n$, is a multiple of *W* since *U* and *W* have the same zeros. The maximum of $|U(x)|$ on *[-1, 1]* occurs then at the points:

$$y_i = cosi\frac{\Pi}{n}$$

Since:

$$T_n(y_i) = cosni\frac{\Pi}{n} = cosi\Pi = (-1)^i$$

Now if possible, let *V* be another polynomial of the same form as *W*, for which: $||V|| < ||U||$. Then:

$$V(y_0) < ||U|| = U(y_0)$$

Thus:

$$V(y_0) < U(y_0)$$

and:

$$V(y_0) < ||U|| = |U(y_1)| = |-1|$$

Now:

$$V(y_1) > \text{-}1,$$

Since if:

$$V(y_1) \leq \text{-}1,$$

Then:

$$V(y_1) \geq |-1| = ||U||$$

Thus:

$$V(y_1) > -1 = U(y_1)$$

Continuing this process we see that *(U - V)* must vanish at least once in each interval *[* (y_1, y_0), (y_2, y_1), ... *]* for a total of n times. But this is not possible since V and W have degree n and a leading coefficient of *1*. Their difference is therefore of degree $\leq n$.

Thus: $||V|| \geq ||U||$, and W is thus minimized on *[-1, 1]* when:

$$x_i = cos[(2i - 1)\frac{\Pi}{2n}]$$

I now need to expand this result to apply for the general interval *[a, b]* to *[-1, 1]* where the following serves our purpose:

$$x = \frac{a-2y+b}{a-b}$$

Our transformation is continuous except when: $a = b$, which is an interval of one point and thus not permitted. Solving for y we have:

$$x = \frac{a-2y+b}{a-b}$$

$$(a - b)x = (a + b) - 2y$$

$$2y = (a + b) + (b - a)x$$

$$y = \frac{1}{2} [(a + b) + (b - a)x] \quad \text{from } [-1, 1] \rightarrow [a, b] \; .$$

Now we recall that the zeros of $T_n(x)$ are:

$$x_i = cos[(2i - 1)\frac{\Pi}{2n}]$$

And thus the corresponding interpolation points in *[a, b]* are:

$$y_i = \frac{1}{2} [(b - a)cos[(2i - 1)\frac{\Pi}{2n}] + (a + b)]$$

Therefore, for a given function f of degree $\leq n$, which posseses n continuous derivatives on *[a, b]*, we can construct an approximation p such that the mean of *(f - p)* is give by:

$$||f - p|| \leq \frac{1}{n!} ||f^{(n)}|| \; ||w||$$

where the norm of $w(y) = \prod_{i=1}^{n} (y - y_i)$ is minimized on *[1, b]*, when:

$$y_i = \frac{1}{2} [(b - a)cos[(2i - 1)\frac{\Pi}{2n}] + (a + b)]$$

For the maximum derivation of $\prod_{i=1}^{n} (y - y_i)$ from zero in *[1, b]*, we have:

$$max_{a \leq y \leq b} \prod_{i=1}^{n} |y - y_i|$$

Now:

$$y - y_i = \frac{1}{2} \, [(b - a)x + (a + b)] - \frac{1}{2} \, [(b - a)x_i + (a + b)]$$

$$= \frac{1}{2} \, [(b - a)x + (a + b) - (b - a)x_i + (a + b)]$$

$$= \frac{1}{2} \, [(b - a)(x - x_i)]$$

$$= \frac{(b-a)(x-x_i)}{2}$$

Thus:

$$max_{a \leq y \leq b} \prod_{i=1}^{n} |y - y_i| = \left[\frac{(b-a)}{2}\right]^n \prod_{i=1}^{n} (x - x_i)$$

Remembering that: $W(x) = x^{1-n} T_n$
and that the maximum value of: $T_n(x) = 1$.
We have:

$$= \left[\frac{(b-a)}{2}\right]^n 2^{1-n}$$

$$= \left[\frac{(b-a)}{2}\right]^n \frac{1}{2^{n-1}}$$

$$= \frac{(b-a)^n}{2^{2n-1}}$$

Clearly Tchebycheff's polynomial has its limitations as an approximation. This would tend to show us that the use of the interpolation process as a method of approximation would be inadequate for many applications. I will not present a theorem which will introduce an approach which gives much better results.

8 Weierstrass Approximation Theorem

Theorem:

Let f be a continuous function defined on $[a, b]$. To each $\epsilon > 0$, there corresponds a polynomial p such that $||f - p|| < \epsilon$. Thus: $|f - p| < \epsilon$ for all $x \, \epsilon \, [a, b]$.

Proof (by Bernstein):

Bernstein constructed, for a given $f \in c[1, 1])$, a sequence of polynomials (now called Bernstein polynomials) $B_n f$ by means of the formula:

$$(B_n f)(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\left(\begin{array}{c} n \\ k \end{array}\right)x^k(1-x)^{n-k}$$

Where: $\left(\begin{array}{c} n \\ k \end{array}\right)$ is the binomial coefficient: $\frac{n!}{(n-k)!k!}$.

For $f,g \in c[0, 1]$ we have:

$$B_n(f+g) = \sum_{k=0}^{n} (f+g)\left(\frac{k}{n}\right)\left(\begin{array}{c} n \\ k \end{array}\right)x^k(1-x)^{n-k}$$

$$= \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\left(\begin{array}{c} n \\ k \end{array}\right)x^k(1-x)^{n-k} + \sum_{k=0}^{n} g\left(\frac{k}{n}\right)\left(\begin{array}{c} n \\ k \end{array}\right)x^k(1-x)^{n-k}$$

$$= B_n f + B_n g$$

And for α a scalar we have:

$$B_n(\alpha f) = \sum_{k=0}^{n} \alpha f\left(\frac{k}{n}\right)\left(\begin{array}{c} n \\ k \end{array}\right)x^k(1-x)^{n-k}$$

$$= \alpha \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\left(\begin{array}{c} n \\ k \end{array}\right)x^k(1-x)^{n-k}$$

$$= \alpha(B_n f)$$

B_n is thus a linear operator.

By definition, for B_n to be a monotone operator, for:

$$f,g \in c[0, 1]$$

$$f \geq g \rightarrow B_n f \geq B_n g$$

Now since $n \geq k$ and $x \in [0, 1]$:

$$\frac{n!}{(n-k)!k!}\, x^k(1-x)^{n-k} \geq 0$$

Thus if $f \geq g$ then:

$$f\left(\frac{k}{n}\right)\left(\begin{array}{c} n \\ k \end{array}\right)x^k(1-x)^{n-k} \geq g\left(\frac{k}{n}\right)\left(\begin{array}{c} n \\ k \end{array}\right)x^k(1-x)^{n-k}$$

And therefore:

$$\sum_{k=0}^{n} f\left(\frac{k}{n}\right)\left(\begin{array}{c} n \\ k \end{array}\right)x^k(1-x)^{n-k} \geq \sum_{k=0}^{n} g\left(\frac{k}{n}\right)\left(\begin{array}{c} n \\ k \end{array}\right)x^k(1-x)^{n-k}$$

Thus $B_n f \geq B_n g$ and therefore $B_n f$ is a monotone operator. I will now prove a theorem essential to the completion of this proof.

9 Theorem on Monotone Operators

Theorem:

For a sequence of monotone linear operators L_n on $C[1, b]$, the following conditions are equivalent:

1. $L_n f \to f$ for all $f \in c[a, b]$
2. $L_n f \to f$ for the three functions $f(x) = 1, x, x^2$
3. $L_n f \to 1$ and $(L_n \phi_t)(t) \to 0$ in t where $\phi_t(x) = (t - x)^2$

Proof:

$(1 \to 2)$ is trivial.

$(2 \to 3)$

Define: $f_i(x) = x^i$.

Now:

$$\phi_t(x) = (t - x)^2$$

$$= t^2 - 2tx + x^2$$

$$\phi_t = t^2 f_0 - 2t f_1 + f_2$$

$$L_n \phi_t = t^2 L_n f_0 - 2t L_n f_1 + L_n f_2$$

$$(L_n \phi_t)(t) = t^2 (L_n f_0)(t) - t^2 - 2t(L_n f_1)(t) + 2t^2 + (L_n f_2)(t) - t^2$$

$$= t^2[(L_n f_0)(t) - 1] - 2t[(L_n f_1)(t) - t] + [(L_n f_2)(t) - t^2]$$

$$\leq t^2 ||L_n f_0 - 1|| - |2t| ||L_n f_1 - t|| + ||L_n f_2 - t^2||$$

$$\to t^2 ||1 - 1|| - |2t| ||t - t|| + ||t^2 - t^2|| = 0$$

$(3 \to 1)$

We begin by selecting σ such that:

$$|x - y| < \sigma \to |f(x) - f(y)| < \epsilon \ (\sigma > 0, \epsilon > 0)$$

Now set $\alpha = 2||f||\sigma^{-2}$ and let t be an arbitrary but fixed point of $[1, b]$. If $|t - x| < \sigma$, then $|f(t) - f(x)| < \epsilon$. Whereas if $|t - x| \geq \sigma$, then:

$$|f(t) - f(x)| \leq |f(t)| + |f(x)|$$

$$\leq 2||f||$$

$$\leq 2||f|| \ \frac{(t-x)^2}{\sigma^2}$$

$$= \alpha\sigma_t(x)$$

Thus for all x, the following inequality is statisfied:

$$-\epsilon - \alpha\sigma_t(x) \leq f(t) - f(x) \leq \epsilon + \alpha\sigma_t(x)$$

Let $f_0(x) = 1$. Then we have:

$$-\epsilon f_0 - \alpha\sigma_t \leq f(t) f_0 - f \leq \epsilon f_0 + \alpha\sigma_t$$

By the linearity and monotonicity of L_n we have:

$$-\epsilon(L_n f_0)(t) - \alpha(L_n\sigma_t) \leq f(t)(L_n f_0)(t) - (L_n f)(t) \leq \epsilon(L_n f_0)(t) + \alpha(L_n\sigma_t)(t)$$

This yields:

$$|f(t)(L_n f_0)(t) - (L_n f_t)(t)| \leq \epsilon(L_n f_0)(t) + \alpha(L_n\sigma_t)(t) \leq \epsilon||L_n f_0|| + \alpha(L_n\sigma_t)(t)$$

Since $L_n f_0 \to f_0$ and $(L_n\sigma_t)(t) \to 0$ we have that the above expression goes to $|f(t) - (L_n f)(t)| < \epsilon$.

10 Proof of the Weierstrass Theorem

I am first going to prove the theorem for the interval *[0, 1]* and then extend it to a given interval *[a, b]*. I will show that for any $f \in c[0, 1]$. The Bernstein polynomials $B_n f$ converge to *f*. the linearity and monotonicity of B_n have already been mentioned. By the theorem on monotone operators it will suffice to show that $B_n f \to f$ for $f(x) = 1, x,$ *and* x^2.

Now applying the binomial theorem, which states that:

$$\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = (a+b)^n$$

We have:

$$(B_n 1)(x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = [x + (1-x)]^n = 1$$

And for the function $f(x) = x$:

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

$$= \sum_{k=1}^{n} \frac{k}{n} \binom{n}{k} x^k (1-x)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{kn!}{n(n-k)!k!} x^k (1-x)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{(n-1)!}{[n-1-(k-1)]!(k-1)!} x^k (1-x)^{n-k}$$

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

$$= x\, [x + (1-x)]^{n-1} = x$$

For the function $f(x) = x^2$ we have:

$$(B_n f)(x) = \sum_{k=0}^{n} \left(\frac{k}{n}\right)^2 \binom{n}{k} x^k (1-x)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{k}{n} \frac{(n-1)!}{[n-1-(k-1)]!(k-1)!} x^k (1-x)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{k}{n} \binom{n-1}{k-1} x^k (1-x)^{n-k}$$

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

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

$$\sum_{k=1}^{n} \frac{1}{n} \binom{n-1}{k-1} x^k (1-x)^{n-k}$$

$$= \frac{(n-1)}{n} \sum_{k=2}^{n} \left(\frac{k-1}{n-1}\right) \left(\frac{(n-1)!}{[n-1-(k-1)]!(k-1)!}\right) x^k (1-x)^{n-k} + \frac{1}{n}$$

$$\sum_{k=1}^{n} \frac{1}{n} \binom{n-1}{k-1} x^k (1-x)^{n-k}$$

$$= \frac{(n-1)}{n} \sum_{k=2}^{n} \left(\frac{(n-2)!}{[n-2-(k-2)]!(k-2)!}\right) x^k (1-x)^{n-k} + \frac{1}{n}$$

$$\sum_{k=1}^{n} \frac{1}{n} \binom{n-1}{k-1} x^k (1-x)^{n-k}$$

$$= \frac{(n-1)}{n} \sum_{k=2}^{n} \binom{n-2}{k-2} x^k (1-x)^{n-k} + \frac{1}{n}$$

$$\sum_{k=1}^{n} \binom{n-1}{k-1} x^k (1-x)^{n-k}$$

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

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

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

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

$$= \frac{(n-1)x^2}{n} [x + (1-x)]^{n-2} + \frac{x}{n} [x + (1-x)]^{n-1}$$

$$= \frac{(n-1)x^2}{n} + \frac{x}{n} \rightarrow x^2$$

I will extend this theorem to apply to an arbitrary interval *[a, b]*. That is, for a function $f \in c[a, b]$.

Cleary; $g(x) = a + x(b - a)$ is continuous on *[0, 1]*

and on *[0, 1]*, $g(x) = [a, b]$. Thus for $\sigma(x) = fg$, $\sigma(x)$ is continuous on *[0, 1]*.

Therefore, the Bernstein polynomials converge to $\sigma(x)$ on *[0, 1]*. But:

$$\sigma(x) = f[a = x(b-a)] \ (x \in [0,\ 1])$$

$$= f(y) \ (x \in [a,\ b])$$

Thus, the Bernstein polynomials converge to f on $[a,\ b]$.

This theorem would serve to indicate that a polynomial constructed from a sequence of polynomials would satisfy our requirements for an approximation to $f \in c[a,\ b]$.

11 Acknowledgements

I would like to thank Dr. Gary Peterson, Dr. Bryan Dorner, and Paul Liebelt for the help they have given me inresearching this subject.

12 Bibliography

1. E. W. Cheney, "Introduction to Approximation Theory", McGraw-Hill, New York, New York, 1966.

2. Philip J. Davis, "Interpolation and Approximation", Dover, New York, New York, 1975.

3. Eugene Isaacson and Herbert Bishop Keller, "Analysis of Numerical Methods" John Wiley and Sons, New York, New York, 1966.

4. Shan S. Kuo, "Numerical Methods and Computers", Addison-Wesley, Reading, Massachusetts, 1965.

5. Cornelius Lanczos, "Applied Analysis", Prentice Hall, Englewood Cliffs, New Jersey.

6. Günter Meinardus, translated by Larry L. Shumaker, "Approximation of Functions: Theory and Numerical Methods", Springer-Verlag, New York, New York, 1967