A CHARACTERIZATION OF BERNSTEIN POLYNOMIALS

H. Berens

Mathematisches Institut
der Universität Erlangen-Nürnberg
Erlangen, W. Germany

R. DeVore

Department of Mathematics
University of South Carolina
Columbia, South Carolina

The Bernstein polynomials have received considerable attention in approximation theory, in part due to their shape preserving properties. On the other hand, their rate of approximation is not as good as that for other methods of approximation. In this note, we want to show that this slow rate of approximation is actually a consequence of their shape preserving, and in a certain sense they have the best rate of approximation among all operators with the same shape preserving properties.

For a function $f \in C[0,1]$, the Bernstein polynomial is defined by

$$B_n(f) := \sum_0^n f(k/n) p_{n,k}; \qquad p_{n,k}(x) := \binom{n}{k} x^k (1-x)^{n-k}.$$

From the differentiation formula [2]

$$(1)\quad \Big[\sum_{0}^{n} a_k\, p_{n,k}\Big]^{(j)} = n(n-1)\dots(n-j+1)\sum_{k=0}^{n-j} \Delta^j a_k\, p_{n-j,k},$$

it follows that B_n preserves convexity of all orders, i.e. $f^{(j)} \geq 0$ on $[0,1]$ implies $[B_n(f)]^{(j)} \geq 0$ on $[0,1]$, $j=0,1,\dots$. Thus B_n is an operator in the class $\mathcal{L}_n$ of all those operators L_n with

$$(2)\quad \begin{array}{ll} \text{i)} & L_n f \in \mathbb{P}_n,\ \text{for all } f \in C[0,1] \\ \text{ii)} & L_n(\ell) = \ell,\ \text{for all } \ell \in \mathbb{P}_1, \\ \text{iii)} & [L_n(f)]^{(j)} \geq 0,\ \text{if } f^{(j)} \geq 0,\ j=0,1,\dots,n. \end{array}$$

Here $\mathbb{P}_k$ is the space of polynomials of degree k.

Each $L_n \in \mathcal{L}_n$ is a positive operator, so its degree of approximation is controlled by $L_n((\cdot-x)^2,x)$. More precisely, if $B = \{f\colon f' \text{ is a.c. and } \|f'\|_\infty = 1\}$ then [1]

$$\sup_{f\in B} \|f-L_n(f)\|_\infty = \frac{1}{2}\|L_n((\cdot-x)^2,x)\|_\infty$$

with $\|\cdot\|_\infty$ the sup norm on $[0,1]$. The main result of this note is:

$$(3)\quad \frac{x(1-x)}{n} = B_n((\cdot-x)^2,x) = \inf_{L_n\in\mathcal{L}_n} L_n((\cdot-x)^2,x),\ 0 \leq x \leq 1.$$

Thus, at least in this sense, B_n has the best rate of approximation among the operators in $\mathcal{L}_n$.

The result (3) can also be formulated in terms of eigenvalues. From (2)iii), it follows that $L_n\colon \mathbb{P}_j \to \mathbb{P}_j, j=0,1,\dots,n$. from which it is easy to conclude that for each j, L_n has an eigenfunction $Q_j \in \mathbb{P}_j$, $Q_j = x^j + \dots$ with corresponding eigenvalue $\lambda_j = \lambda_j(L_n)$. The eigenvalues of B_n are $\lambda_j(B_n) = 1$, $j=0,1$ and $\lambda_j(B_n) = (1-\frac{1}{2})\dots(1-\frac{j-1}{n})$, $j=2,\dots,n$. The inequality (3) is equivalent to the statement

$$(4)\quad \lambda_2(L_n) \leq \lambda_2(B_n),\ L_n \in \mathcal{L}_n,\ n=1,2,\dots .$$

We will show that equality holds in (4) if and only if $L_n = B_n$ and therefore this characterizes the Bernstein polynomials.

We begin with some simple properties of the eigenvalues $\lambda_j(L_n)$ for $L_n \in \mathcal{L}_n$.

Lemma. <u>For any</u> $L_n \in \mathcal{L}_n$ <u>we have</u>

$$\lambda_0(L_n) = \lambda_1(L_n) = 1 \geq \lambda_2(L_n) \geq \ldots \geq \lambda_n(L_n) \geq 0$$

<u>Proof</u>. Since L_n preserves linear functions, $\lambda_0(L_n) = \lambda_1(L_n) = 1$. For the function $e_j(x) = x^j/j!$, we have $L_n(e_j) = \lambda_j e_j + a_{j-1,j}\, e_{j-1} + \ldots$. Since $e_j^{(i)} \geq 0$, $i=0,1,\ldots,n$, the polynomial $L_n(e_j)$ must be completely monotonic on $[0,1]$ (property (2)iii) and so all the coefficients of $L_n(e_j)$ are non-negative. In particular

$$\lambda_j(L_n) = [L_n(e_j)]^{(j)} \geq 0.$$

In addition, $e_{j-1} - e_j$ has a non-negative $(j-1)^{\underline{st}}$ derivative and so (2)iii) also implies that

$$0 \leq [L_n(e_{j-1}-e_j)]^{(j-1)} = \lambda_{j-1} - \lambda_j x - a_{j-1,j},\quad 0 \leq x \leq 1.$$

When $x=1$, we find $\lambda_{j-1} - \lambda_j \geq a_{j-1,j} \geq 0$, as desired.

Theorem. For any $L_n \in \mathcal{L}_n$.

(5) $\lambda_2(L_n) \leq \lambda_2(B_n) = 1 - \frac{1}{n}$

<u>or equivalently</u>

(6) $\frac{x(1-x)}{n} = B_n((\cdot-x)^2, x) \leq L_n((\cdot-x)^2, x)$

<u>Furthermore, equality can hold in</u> (5) <u>or</u> (6) <u>if and only if</u> $L_n = B_n$.

<u>Proof</u>. The operator L_n can be represented as

$$L_n(f,x) = \sum_{k=0}^{n} a_k(f)\, p_{n,k}(x);\qquad a_k(f) = \int_0^1 f\, d\alpha_k$$

with $d\alpha_k$ a Borel measure. Here we are using the fact that

$(p_{n,k})_0^n$ is a basis for $\mathbb{P}_n$. Now, L_n preserves linear functions and so

$$\int_0^1 d\alpha_k = 1; \quad \int_0^1 t\, d\alpha_k = k/n, \; k=0,1,\dots,n,$$

where we used the fact that $1 = \sum_0^n p_{n,k}$; $x = \sum_0^n k/n\, p_{n,k}(x)$, uniquely. The two measures $d\alpha_0$ and $d\alpha_n$ are easily seen to be positive. Indeed if $f \geq 0$ then $a_0(f) = L_n(f,0) \geq 0$, similarly for $d\alpha_n$. In addition $0 \leq L_n(t^2,0) \leq L_n(t,0) = 0$, because $t^2 \leq t$ on $[0,1]$. This means that $\int_0^1 t^2\, d\alpha_0 = 0$. Since $d\alpha_0$ is a positive measure we must have $d\alpha_0 = d\rho_0$ with $d\rho_t$ denoting the Dirac measure with unit mass at t. In a similar way, we find $d\alpha_n = d\rho_1$.

We now want to prove that $d\alpha_1$ and $d\alpha_{n-1}$ are positive measures. It is enough to do this for $d\alpha_{n-1}$ since a similar argument (or symmetry) handles the case $d\alpha_1$. If f is any nonnegative function on $[0,1]$ with $f(1) = 0$ then

$$0 \leq \lim_{x\to 1} \frac{L_n(f,x)}{p_{n,n-1}(x)} = \lim_{x\to 1} \sum_0^{n-1} a_k(f) \frac{p_{n,k}(x)}{p_{n,n-1}(x)} = a_{n-1}(f)$$

Thus $d\alpha_{n-1} = d\mu_{n-1} + c\,d\rho_1$ with $d\mu_{n-1} \geq 0$ and supported on $[0,1)$ and c a constant. We want to see that $c \geq 0$. To this end, consider the function $f_\varepsilon(x) \equiv (x-\varepsilon)_+^n$, $0 < \varepsilon < 1$. Since $f_\varepsilon^{(k)} \geq 0$, $k=0,1,\dots,n$, we have from (1) and (2)iii)

$$\Delta^i a_0(f_\varepsilon) = \lim_{x\to 0} [L_n(f_\varepsilon)]^{(i)}(x) \geq 0, \; i=0,1,\dots,n.$$

Here $\Delta^i a_k = \sum_{\nu=0}^{i} \binom{i}{\nu}(-1)^{\nu+i} a_{k+\nu}$

It follows easily that $(a_i(f_\varepsilon))$ is a monotone sequence:

$$(7) \quad 0 = a_0(f_\varepsilon) \leq \dots \leq a_{n-1}(f_\varepsilon) \leq a_n(f_\varepsilon) = (1-\varepsilon)_+^n .$$

In particular $0 \leq a_{n-1}(f_\varepsilon) = \int_\varepsilon^1 f_\varepsilon\, d\mu_{n-1} + c(1-\varepsilon)^n$. Dividing by $(1-\varepsilon)^n$ and letting $\varepsilon \to 1$ easily establishes that $c \geq 0$, as desired. Note $(1-\varepsilon)^{-n} f_\varepsilon \leq 1$ for all $0 < \varepsilon < 1$.

Now we can estimate the eigenvalue $\lambda_2(L_n)$. Let ℓ be the linear function which satisfies $\ell(\frac{n-1}{n}) = (\frac{n-1}{n})^2$, $\ell'(\frac{n-1}{n}) = 2(\frac{n-1}{n})$. Then $\ell(t) \leq t^2$ for all $0 \leq t \leq 1$ and so

$$(\frac{n-1}{n})^2 = a_{n-1}(\ell) \leq a_{n-1}(t^2).$$

But

$$L_n(t^2,x) = \sum_{i=0}^{2} c_i x^i = \sum_0^n Q(k/n)\, p_{n,k}(x)$$

for some quadratic Q. Here we use the fact that $x^2 = \sum_0^n Q_2(k/n) p_{n,k}(x)$ for some quadratic Q_2. Now we know that $Q(0) = 0$, $Q(1) = 1$, $Q(\frac{n-1}{n}) \geq (\frac{n-1}{n})^2$. Thus $Q(t) \geq t^2$ for all $0 \leq t \leq 1$ and therefore

$$L_n(t^2,x) = \sum_0^n Q(k/n)\, p_{n,k}(x) \geq \sum_0^n (k/n)^2 p_{n,k}(x) = B_n(t^2,x).$$

Since both L_n and B_n preserve linear functions

(9) $\quad L_n((\cdot-x)^2,x) \geq B_n((\cdot-x)^2,x) = \dfrac{x(1-x)}{n}, \quad 0 \leq x \leq 1$

which establishes (6).

To prove (5), note that

$$L_n((t-x)^2,x) = L_n(t^2,x) - 2xL_n(t,x) + x^2 = (\lambda_2(L_n)-1)x^2+ax+b$$

Comparing terms with (8) shows that

$$[1-\lambda_2(L_n)]\; x\; (1-x) \geq \frac{x(1-x)}{n} \quad 0 \leq x \leq 1$$

from which we get $\lambda_2(L_n) \leq 1 - \frac{1}{n} = \lambda_2(B_n)$ and so we have proved (5) and (6).

We now want to discuss when equality can hold in (5) or (6). We will show that equality in (5) or (6) implies $L_n = B_n$. This will be done by showing that $d\alpha_k = d\rho_{k/n}$, $k=0,1,\ldots,n$.

For this purpose, it is enough to show that the measures $d\alpha_k$ are non-negative since by (2)ii) we already know that

(10) $$\int_0^1 t^i \, d\alpha_k = \left(\frac{k}{n}\right)^i$$

$i = 0,1$, and equality in (5) means that this holds for $i = 2$ as well. It is well known that any positive measure satisfying (10) for $i = 0,1,2$ must equal $d\rho_{k/n}$.

We now show that all the measures $d\alpha_0,\ldots,d\alpha_n$ are positive. Suppose this is not the case and let μ be the largest integer $\le n$ such that $d\alpha_\mu$ is not positive. Then $\mu \le n-2$ because we have shown earlier that $d\alpha_{n-1}$ and $d\alpha_n$ are positive. We then know that $d\alpha_j = d\rho_{j/n}$, $j = \mu + 1,\ldots,n$. If $f \ge 0$ vanishes at $\frac{j}{n}$, $j = \mu+1,\ldots,n$, then $a_j(f) = 0$ so that

$$a_\mu(f) = \lim_{x\to 1} \sum_{k=0}^{\mu} a_k(f) \frac{p_{n,k}(x)}{p_{n,\mu}(x)} = \lim_{x\to 1} \frac{L_n(f,x)}{p_{n,\mu}(x)} \ge 0$$

Hence $d\alpha_\mu = d\beta + d\gamma$ with $d\beta \ge 0$ and $d\gamma = \sum_{\mu+1}^{n} c_j \, d\rho_{j/n}$.

We will show that $c_j \ge 0$, $j = \mu+1,\ldots,n$ which will complete the proof. First we want to see that $c_j = 0$, $j = \mu+2,\ldots,n$. Assume this were not the case and let ν be the largest integer $\nu \ge \mu+2$ for which $c_\nu \ne 0$.

Consider the function $f_\varepsilon(x) = (x-\varepsilon)_+^n$ which we used earlier. We have

$$0 = a_0(f_\varepsilon) \le \ldots \le a_\mu(f_\varepsilon) \le f_\varepsilon\left(\frac{\mu+1}{n}\right) \le \ldots \le f_\varepsilon(1).$$

Now, when $\nu \ne n$, take $\varepsilon > \frac{\nu}{n}$ so that $f_\varepsilon\left(\frac{\mu+1}{n}\right) = 0$ and hence $a_\mu(f_\varepsilon) = 0$, thus

$$\int_0^1 f_\varepsilon \, d\alpha_\mu = 0$$

Since $d\alpha_\mu \ge 0$ on $(\frac{\nu}{n},1]$, we have (taking $\varepsilon \sim \frac{\nu}{n}$) $d\alpha_\mu \equiv 0$ on $(\frac{\nu}{n},1]$.

This holds when $\nu = n$ as well. Now, take $\frac{\mu+1}{n} < \varepsilon < \frac{\nu}{n}$. Then

$$(11) \qquad 0 \le a_\mu(f_\varepsilon) = \int_\varepsilon^{\frac{\nu}{n}} f_\varepsilon d\beta + c_\nu f_\varepsilon(\tfrac{\nu}{n}) \le f_\varepsilon(\tfrac{\mu+1}{n}) = 0$$

Dividing by $f_\varepsilon(\frac{\nu}{n})$ and letting $\varepsilon \to \frac{\nu}{n}$ shows that $c_\nu = 0$ since the integral term will tend to 0. This gives a contradiction and shows that each $c_j = 0$, $j = \mu+2,\ldots,n$.

The same argument that arrived at the lefthand inequality in (11) applies to the case $\nu = \mu+1$. Dividing by $f_\varepsilon(\frac{\nu}{n})$ and taking a limit in this case shows that $c_{\mu+1} \ge 0$. Thus, we have shown that $c_j \ge 0$, $j = \mu+1,\ldots,n$ and as observed above this proved that $d\alpha_\mu \ge 0$ as desired.

REFERENCES

[1] R. DeVore, Approximation of Continuous Functions by Positive Linear Operators, Springer Lecture Notes in Mathematics, #293, Berlin, 1972, 289 pp.

[2] G. G. Lorentz, Bernstein Polynomials, Toronto U. Press, Toronto, 1953, 130 pp.

The first author acknowledges National Science Foundation support in Grant MCS 77-22982.