7

Bernstein Polynomials

7.1 Introduction

This chapter is concerned with sequences of polynomials named after their creator S. N. Bernstein. Given a function f on $[0, 1]$, we define the Bernstein polynomial

$$B_n(f; x) = \sum_{r=0}^{n} f\left(\frac{r}{n}\right) \binom{n}{r} x^r (1 - x)^{n-r} \tag{7.1}$$

for each positive integer n. Thus there is a sequence of Bernstein polynomials corresponding to each function f. As we will see later in this chapter, if f is continuous on $[0, 1]$, its sequence of Bernstein polynomials converges uniformly to f on $[0, 1]$, thus giving a constructive proof of Weierstrass's Theorem 2.4.1, which we stated in Chapter 2. There are several proofs of this fundamental theorem, beginning with that given by K. Weierstrass [55] in 1885. (See the Notes in E. W. Cheney's text [7]. This contains a large number of historical references in approximation theory.) Bernstein's proof [3] was published in 1912. One might wonder why Bernstein created "new" polynomials for this purpose, instead of using polynomials that were already known to mathematics. Taylor polynomials are not appropriate; for even setting aside questions of convergence, they are applicable only to functions that are infinitely differentiable, and not to all continuous functions. We can also dismiss another obvious candidate, the interpolating polynomials for f constructed at equally spaced points. For the latter sequence of polynomials does *not* converge uniformly to f for *all* $f \in C[0, 1]$, and the same is true of interpolation on any other fixed sequence of abscissas. However, L. Fejér [19] used a method based on Hermite interpolation in a proof published in 1930, which we will discuss in the next section.

Later in this section we will consider how Bernstein discovered his polynomials, for this is not immediately obvious. We will also see that although the convergence of the Bernstein polynomials is slow, they have compensating "shape-preserving" properties. For example, the Bernstein polynomial of a convex function is itself convex.

It is clear from (7.1) that for all $n \geq 1$,

$$B_n(f;0) = f(0) \quad \text{and} \quad B_n(f;1) = f(1), \tag{7.2}$$

so that a Bernstein polynomial for f interpolates f at both endpoints of the interval $[0,1]$.

Example 7.1.1 It follows from the binomial expansion that

$$B_n(1;x) = \sum_{r=0}^{n} \binom{n}{r} x^r (1-x)^{n-r} = (x + (1-x))^n = 1, \tag{7.3}$$

so that the Bernstein polynomial for the constant function 1 is also 1. Since

$$\frac{r}{n}\binom{n}{r} = \binom{n-1}{r-1}$$

for $1 \leq r \leq n$, the Bernstein polynomial for the function x is

$$B_n(x;x) = \sum_{r=0}^{n} \frac{r}{n}\binom{n}{r} x^r (1-x)^{n-r} = x \sum_{r=1}^{n} \binom{n-1}{r-1} x^{r-1}(1-x)^{n-r}.$$

Note that the term corresponding to $r = 0$ in the first of the above two sums is zero. On putting $s = r - 1$ in the second summation, we obtain

$$B_n(x;x) = x \sum_{s=0}^{n-1} \binom{n-1}{s} x^s (1-x)^{n-1-s} = x, \tag{7.4}$$

the last step following from (7.3) with n replaced by $n-1$. Thus the Bernstein polynomial for the function x is also x. ■

We call B_n the *Bernstein operator*; it maps a function f, defined on $[0,1]$, to $B_n f$, where the function $B_n f$ evaluated at x is denoted by $B_n(f;x)$. The Bernstein operator is obviously linear, since it follows from (7.1) that

$$B_n(\lambda f + \mu g) = \lambda\, B_n f + \mu\, B_n g, \tag{7.5}$$

for all functions f and g defined on $[0,1]$, and all real λ and μ. We now require the following definition.

Definition 7.1.1 Let L denote a linear operator that maps a function f defined on $[a, b]$ to a function Lf defined on $[c, d]$. Then L is said to be a *monotone* operator or, equivalently, a *positive* operator if

$$f(x) \geq g(x), \quad x \in [a, b] \quad \Rightarrow \quad (Lf)(x) \geq (Lg)(x), \quad x \in [c, d], \tag{7.6}$$

where we have written $(Lf)(x)$ to denote the value of the function Lf at the point $x \in [a, b]$. ■

We can see from (7.1) that B_n is a monotone operator. It then follows from the monotonicity of B_n and (7.3) that

$$m \leq f(x) \leq M, \quad x \in [0, 1] \quad \Rightarrow \quad m \leq B_n(f; x) \leq M, \quad x \in [0, 1]. \tag{7.7}$$

In particular, if we choose $m = 0$ in (7.7), we obtain

$$f(x) \geq 0, \quad x \in [0, 1] \quad \Rightarrow \quad B_n(f; x) \geq 0, \quad x \in [0, 1]. \tag{7.8}$$

It follows from (7.3), (7.4), and the linear property (7.5) that

$$B_n(ax + b; x) = ax + b, \tag{7.9}$$

for all real a and b. We therefore say that the Bernstein operator *reproduces* linear polynomials. We can deduce from the following result that the Bernstein operator does not reproduce any polynomial of degree greater than one.

Theorem 7.1.1 The Bernstein polynomial may be expressed in the form

$$B_n(f; x) = \sum_{r=0}^{n} \binom{n}{r} \Delta^r f(0)\, x^r, \tag{7.10}$$

where Δ is the forward difference operator, defined in (1.67), with step size $h = 1/n$.

Proof. Beginning with (7.1), and expanding the term $(1 - x)^{n-r}$, we have

$$B_n(f; x) = \sum_{r=0}^{n} f\left(\frac{r}{n}\right) \binom{n}{r} x^r \sum_{s=0}^{n-r} (-1)^s \binom{n-r}{s} x^s.$$

Let us put $t = r + s$. We may write

$$\sum_{r=0}^{n} \sum_{s=0}^{n-r} = \sum_{t=0}^{n} \sum_{r=0}^{t}, \tag{7.11}$$

since both double summations in (7.11) are over all lattice points (r, s) lying in the triangle shown in Figure 7.1. We also have

$$\binom{n}{r}\binom{n-r}{s} = \binom{n}{t}\binom{t}{r},$$

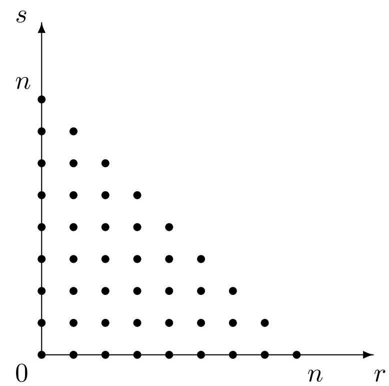

FIGURE 7.1. A triangular array of $\frac{1}{2}(n+1)(n+2)$ lattice points.

and so we may write the double summation as

$$\sum_{t=0}^{n} \binom{n}{t} x^t \sum_{r=0}^{t} (-1)^{t-r} \binom{t}{r} f\left(\frac{r}{n}\right) = \sum_{t=0}^{n} \binom{n}{t} \Delta^t f(0)\, x^t,$$

on using the expansion for a higher-order forward difference, as in Problem 1.3.7. This completes the proof. ■

In (1.80) we saw how differences are related to derivatives, showing that

$$\frac{\Delta^m f(x_0)}{h^m} = f^{(m)}(\xi), \tag{7.12}$$

where $\xi \in (x_0, x_m)$ and $x_m = x_0 + mh$. Let us put $h = 1/n$, $x_0 = 0$, and $f(x) = x^k$, where $n \geq k$. Then we have

$$n^r \Delta^r f(0) = 0 \quad \text{for} \quad r > k$$

and

$$n^k \Delta^k f(0) = f^{(k)}(\xi) = k!\,. \tag{7.13}$$

Thus we see from (7.10) with $f(x) = x^k$ and $n \geq k$ that

$$B_n(x^k; x) = a_0 x^k + a_1 x^{k-1} + \cdots + a_{k-1} x + a_k,$$

say, where $a_0 = 1$ for $k = 0$ and $k = 1$, and

$$a_0 = \binom{n}{k} \frac{k!}{n^k} = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)$$

for $k \geq 2$. Since $a_0 \neq 1$ when $n \geq k \geq 2$, this justifies our above statement that the Bernstein operator does not reproduce any polynomial of degree greater than one.

Example 7.1.2 With $f(x) = x^2$, we have

$$f(0) = 0, \quad \Delta f(0) = f\left(\frac{1}{n}\right) - f(0) = \frac{1}{n^2},$$

and we see from (7.13) that $n^2 \Delta^2 f(0) = 2!$ for $n \geq 2$. Thus it follows from (7.10) that

$$B_n(x^2; x) = \binom{n}{1} \frac{x}{n^2} + \binom{n}{2} \frac{2x^2}{n^2} = \frac{x}{n} + \left(1 - \frac{1}{n}\right) x^2,$$

which may be written in the form

$$B_n(x^2; x) = x^2 + \frac{1}{n} x(1 - x). \tag{7.14}$$

Thus the Bernstein polynomials for x^2 converge uniformly to x^2 like $1/n$, very slowly. We will see from Voronovskaya's Theorem 7.1.10 that this rate of convergence holds for all functions that are twice differentiable. ■

We have already seen in (7.7) that if $f(x)$ is positive on $[0, 1]$, so is $B_n(f; x)$. We now show that if $f(x)$ is monotonically increasing, so is $B_n(f; x)$.

Theorem 7.1.2 The derivative of the Bernstein polynomial $B_{n+1}(f; x)$ may be expressed in the form

$$B'_{n+1}(f; x) = (n + 1) \sum_{r=0}^{n} \Delta f\left(\frac{r}{n + 1}\right) \binom{n}{r} x^r (1 - x)^{n-r} \tag{7.15}$$

for $n \geq 0$, where Δ is applied with step size $h = 1/(n + 1)$. Furthermore, if f is monotonically increasing or monotonically decreasing on $[0, 1]$, so are all its Bernstein polynomials.

Proof. The verification of (7.15) is omitted because it is a special case of (7.16), concerning higher-order derivatives of the Bernstein polynomials, which we prove in the next theorem. To justify the above remark on monotonicity, we note that if f is monotonically increasing, its forward differences are nonnegative. It then follows from (7.15) that $B'_{n+1}(f; x)$ is nonnegative on $[0, 1]$, and so $B_{n+1}(f; x)$ is monotonically increasing. Similarly, we see that if f is monotonically decreasing, so is $B_{n+1}(f; x)$. ■

Theorem 7.1.3 For any integer $k \geq 0$, the kth derivative of $B_{n+k}(f; x)$ may be expressed in terms of kth differences of f as

$$B^{(k)}_{n+k}(f; x) = \frac{(n + k)!}{n!} \sum_{r=0}^{n} \Delta^k f\left(\frac{r}{n + k}\right) \binom{n}{r} x^r (1 - x)^{n-r} \tag{7.16}$$

for all $n \geq 0$, where Δ is applied with step size $h = 1/(n + k)$.

Proof. We write

$$B_{n+k}(f;x) = \sum_{r=0}^{n+k} f\left(\frac{r}{n+k}\right)\binom{n+k}{r} x^r(1-x)^{n+k-r}$$

and differentiate k times, giving

$$B_{n+k}^{(k)}(f;x) = \sum_{r=0}^{n+k} f\left(\frac{r}{n+k}\right)\binom{n+k}{r} p(x), \tag{7.17}$$

where

$$p(x) = \frac{d^k}{dx^k} x^r(1-x)^{n+k-r}.$$

We now use the Leibniz rule (1.83) to differentiate the product of x^r and $(1-x)^{n+k-r}$. First we find that

$$\frac{d^s}{dx^s} x^r = \begin{cases} \dfrac{r!}{(r-s)!} x^{r-s}, & r-s \geq 0, \\[2ex] 0, & r-s < 0, \end{cases}$$

and

$$\frac{d^{k-s}}{dx^{k-s}}(1-x)^{n+k-r} = \begin{cases} (-1)^{k-s}\dfrac{(n+k-r)!}{(n+s-r)!}(1-x)^{n+s-r}, & r-s \leq n, \\[2ex] 0, & r-s > n. \end{cases}$$

Thus the kth derivative of $x^r(1-x)^{n+k-r}$ is

$$p(x) = \sum_s (-1)^{k-s}\binom{k}{s}\frac{r!}{(r-s)!}\frac{(n+k-r)!}{(n+s-r)!} x^{r-s}(1-x)^{n+s-r}, \tag{7.18}$$

where the latter summation is over all s from 0 to k, subject to the constraints $0 \leq r-s \leq n$. We make the substitution $t = r-s$, so that

$$\sum_{r=0}^{n+k}\sum_s = \sum_{t=0}^{n}\sum_{s=0}^{k}. \tag{7.19}$$

A diagram may be helpful here. The double summations in (7.19) are over all lattice points (r,s) lying in the parallelogram depicted in Figure 7.2. The parallelogram is bounded by the lines $s=0$, $s=k$, $t=0$, and $t=n$, where $t = r-s$. We also note that

$$\binom{n+k}{r}\frac{r!}{(r-s)!}\frac{(n+k-r)!}{(n+s-r)!} = \frac{(n+k)!}{n!}\binom{n}{r-s}. \tag{7.20}$$

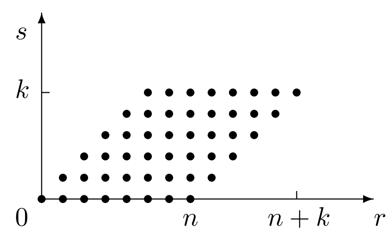

FIGURE 7.2. A parallelogram of $(n+1)(k+1)$ lattice points.

It then follows from (7.17), (7.18), (7.19), and (7.20) that the kth derivative of $B_{n+k}(f;x)$ is

$$\frac{(n+k)!}{n!}\sum_{t=0}^{n}\sum_{s=0}^{k}(-1)^{k-s}\binom{k}{s} f\left(\frac{t+s}{n+k}\right)\binom{n}{t} x^t(1-x)^{n-t}.$$

Finally, we note from Problem 1.3.7 that

$$\sum_{s=0}^{k}(-1)^{k-s}\binom{k}{s} f\left(\frac{t+s}{n+k}\right) = \Delta^k f\left(\frac{t}{n+k}\right),$$

where the operator Δ is applied with step size $h = 1/(n+k)$. This completes the proof. ■

By using the connection between differences and derivatives, we can deduce the following valuable result from Theorem 7.1.3.

Theorem 7.1.4 If $f \in C^k[0,1]$, for some $k \geq 0$, then

$$m \leq f^{(k)}(x) \leq M, \quad x \in [0,1] \quad \Rightarrow \quad c_k m \leq B_n^{(k)}(f;x) \leq c_k M, \quad x \in [0,1],$$

for all $n \geq k$, where $c_0 = c_1 = 1$ and

$$c_k = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right), \quad 2 \leq k \leq n.$$

Proof. We have already seen in (7.7) that this result holds for $k = 0$. For $k \geq 1$ we begin with (7.16) and replace n by $n-k$. Then, using (7.12) with $h = 1/n$, we write

$$\Delta^k f\left(\frac{r}{n}\right) = \frac{f^{(k)}(\xi_r)}{n^k}, \tag{7.21}$$

where $r/n < \xi_r < (r+k)/n$. Thus

$$B_n^{(k)}(f;x) = \sum_{r=0}^{n-k} c_k f^{(k)}(\xi_r) x^r (1-x)^{n-k-r},$$

and the theorem follows easily from the latter equation. One consequence of this result is that if $f^{(k)}(x)$ is of fixed sign on $[0,1]$, then $B_n^{(k)}(f;x)$ also has this sign on $[0,1]$. For example, if $f''(x)$ exists and is nonnegative on $[0,1]$, so that f is convex, then $B_n''(f;x)$ is also nonnegative and $B_n(f;x)$ is convex. ■

Bernstein's discovery of his polynomials was based on an ingenious probabilistic argument. Suppose we have an event that can be repeated and has only two possible outcomes, A and B. One of the simplest examples is the tossing of an unbiased coin, where the two possible outcomes, heads and tails, both occur with probability 0.5. More generally, consider an event where the outcome A happens with probability $x \in [0,1]$, and thus the outcome B happens with probability $1-x$. Then the probability of A happening precisely r times followed by B happening $n-r$ times is $x^r(1-x)^{n-r}$. Since there are $\left(\begin{array}{c} n \\ r \end{array}\right)$ ways of choosing the order of r outcomes out of n, the probability of obtaining r outcomes A and $n-r$ outcomes B in *any* order is given by

$$p_{n,r}(x) = \left(\begin{array}{c} n \\ r \end{array}\right) x^r(1-x)^{n-r}.$$

Thus we have

$$p_{n,r}(x) = \left(\frac{n-r+1}{r}\right)\left(\frac{x}{1-x}\right) p_{n,r-1}(x),$$

and it follows that

$$p_{n,r}(x) > p_{n,r-1}(x) \quad \text{if and only if} \quad r < (n+1)x.$$

We deduce that $p_{n,r}(x)$, regarded as a function of r, with x and n fixed, has a peak when $r = r_x \approx nx$, for large n, and is monotonically increasing for $r < r_x$ and monotonically decreasing for $r > r_x$. We already know that

$$\sum_{r=0}^{n} p_{n,r}(x) = B_n(1;x) = 1,$$

and in the sum

$$\sum_{r=0}^{n} p_{n,r}(x) f\left(\frac{r}{n}\right) = B_n(f;x),$$

where $f \in C[0,1]$ and n is large, the contributions to the sum from values of r sufficiently remote from r_x will be negligible, and the significant part of the sum will come from values of r close to r_x. Thus, for n large,

$$B_n(f;x) \approx f\left(\frac{r_x}{n}\right) \approx f(x),$$

and so $B_n(f;x) \to f(x)$ as $n \to \infty$. While this is by no means a rigorous argument, and is thus *not* a proof, it gives some insight into how Bernstein was motivated in his search for a proof of the Weierstrass theorem.

Example 7.1.3 To illustrate Bernstein's argument concerning the polynomials $p_{n,r}$, let us evaluate these polynomials when $n = 8$ and $x = 0.4$. The resulting values of $p_{n,r}(x)$ are given in the following table:

r	0	1	2	3	4	5	6	7	8
$p_{n,r}(x)$	0.02	0.09	0.21	0.28	0.23	0.12	0.04	0.01	0.00

In this case, the largest value of $p_{n,r}$ is attained for $r = r_x = 3$, consistent with our above analysis, which shows that $r_x \approx nx = 3.2$. ■

Theorem 7.1.5 Given a function $f \in C[0,1]$ and any $\epsilon > 0$, there exists an integer N such that

$$|f(x) - B_n(f;x)| < \epsilon, \quad 0 \leq x \leq 1,$$

for all $n \geq N$.

Proof. In other words, the above statement says that the Bernstein polynomials for a function f that is continuous on $[0,1]$ converge uniformly to f on $[0,1]$. The following proof is motivated by the plausible argument that we gave above.

We begin with the identity

$$\left(\frac{r}{n} - x\right)^2 = \left(\frac{r}{n}\right)^2 - 2\left(\frac{r}{n}\right)x + x^2,$$

multiply each term by $\binom{n}{r} x^r(1-x)^{n-r}$, and sum from $r = 0$ to n, to give

$$\sum_{r=0}^{n} \left(\frac{r}{n} - x\right)^2 \binom{n}{r} x^r(1-x)^{n-r} = B_n(x^2;x) - 2xB_n(x;x) + x^2B_n(1;x).$$

It then follows from (7.3), (7.4), and (7.14) that

$$\sum_{r=0}^{n} \left(\frac{r}{n} - x\right)^2 \binom{n}{r} x^r(1-x)^{n-r} = \frac{1}{n}x(1-x). \tag{7.22}$$

For any fixed $x \in [0,1]$, let us estimate the sum of the polynomials $p_{n,r}(x)$ over all values of r for which r/n is not close to x. To make this notion precise, we choose a number $\delta > 0$ and let S_δ denote the set of all values of r satisfying $\left|\frac{r}{n} - x\right| \geq \delta$. We now consider the sum of the polynomials $p_{n,r}(x)$ over all $r \in S_\delta$. Note that $\left|\frac{r}{n} - x\right| \geq \delta$ implies that

$$\frac{1}{\delta^2}\left(\frac{r}{n} - x\right)^2 \geq 1. \tag{7.23}$$

Then, using (7.23), we have

$$\sum_{r\in S_\delta}\binom{n}{r}x^r(1-x)^{n-r}\le\frac{1}{\delta^2}\sum_{r\in S_\delta}\left(\frac{r}{n}-x\right)^2\binom{n}{r}x^r(1-x)^{n-r}.$$

The latter sum is not greater that the sum of the same expression over *all* r, and using (7.22), we have

$$\frac{1}{\delta^2}\sum_{r=0}^{n}\left(\frac{r}{n}-x\right)^2\binom{n}{r}x^r(1-x)^{n-r}=\frac{x(1-x)}{n\delta^2}.$$

Since $0\le x(1-x)\le\frac{1}{4}$ on $[0,1]$, we have

$$\sum_{r\in S_\delta}\binom{n}{r}x^r(1-x)^{n-r}\le\frac{1}{4n\delta^2}. \tag{7.24}$$

Let us write

$$\sum_{r=0}^{n}=\sum_{r\in S_\delta}+\sum_{r\notin S_\delta},$$

where the latter sum is therefore over all r such that $|\frac{r}{n}-x|<\delta$. Having split the summation into these two parts, which depend on a choice of δ that we still have to make, we are now ready to estimate the difference between $f(x)$ and its Bernstein polynomial. Using (7.3), we have

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

and hence

$$\begin{aligned}f(x)-B_n(f;x)=&\sum_{r\in S_\delta}\left(f(x)-f\left(\frac{r}{n}\right)\right)\binom{n}{r}x^r(1-x)^{n-r}\\&+\sum_{r\notin S_\delta}\left(f(x)-f\left(\frac{r}{n}\right)\right)\binom{n}{r}x^r(1-x)^{n-r}.\end{aligned}$$

We thus obtain the inequality

$$\begin{aligned}|f(x)-B_n(f;x)|\le&\sum_{r\in S_\delta}\left|f(x)-f\left(\frac{r}{n}\right)\right|\binom{n}{r}x^r(1-x)^{n-r}\\&+\sum_{r\notin S_\delta}\left|f(x)-f\left(\frac{r}{n}\right)\right|\binom{n}{r}x^r(1-x)^{n-r}.\end{aligned}$$

Since $f \in C[0,1]$, it is bounded on $[0,1]$, and we have $|f(x)| \leq M$, for some $M > 0$. We can therefore write

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

for all r and all $x \in [0,1]$, and so

$$\sum_{r \in S_\delta} \left| f(x) - f\left(\frac{r}{n}\right) \right| \binom{n}{r} x^r (1-x)^{n-r} \leq 2M \sum_{r \in S_\delta} \binom{n}{r} x^r (1-x)^{n-r}.$$

On using (7.24) we obtain

$$\sum_{r \in S_\delta} \left| f(x) - f\left(\frac{r}{n}\right) \right| \binom{n}{r} x^r (1-x)^{n-r} \leq \frac{M}{2n\delta^2}. \tag{7.25}$$

Since f is continuous, it is also uniformly continuous, on $[0,1]$. Thus, corresponding to any choice of $\epsilon > 0$ there is a number $\delta > 0$, depending on ϵ and f, such that

$$|x - x'| < \delta \ \Rightarrow \ |f(x) - f(x')| < \frac{\epsilon}{2},$$

for all $x, x' \in [0,1]$. Thus, for the sum over $r \notin S_\delta$, we have

$$\begin{aligned} \sum_{r \notin S_\delta} \left| f(x) - f\left(\frac{r}{n}\right) \right| \binom{n}{r} x^r (1-x)^{n-r} &< \frac{\epsilon}{2} \sum_{r \notin S_\delta} \binom{n}{r} x^r (1-x)^{n-r} \\ &< \frac{\epsilon}{2} \sum_{r=0}^{n} \binom{n}{r} x^r (1-x)^{n-r}, \end{aligned}$$

and hence, again using (7.3), we find that

$$\sum_{r \notin S_\delta} \left| f(x) - f\left(\frac{r}{n}\right) \right| \binom{n}{r} x^r (1-x)^{n-r} < \frac{\epsilon}{2}. \tag{7.26}$$

On combining (7.25) and (7.26), we obtain

$$|f(x) - B_n(f;x)| < \frac{M}{2n\delta^2} + \frac{\epsilon}{2}.$$

It follows from the line above that if we choose $N > M/(\epsilon\delta^2)$, then

$$|f(x) - B_n(f;x)| < \epsilon$$

for all $n \geq N$, and this completes the proof. ■

Using the methods employed in the above proof, we can show, with a little greater generality, that if f is merely bounded on $[0,1]$, the sequence $(B_n(f;x))_{n=1}^{\infty}$ converges to $f(x)$ at any point x where f is continuous. We will now discuss some further properties of the Bernstein polynomials.

Theorem 7.1.6 If $f \in C^k[0,1]$, for some integer $k \geq 0$, then $B_n^{(k)}(f;x)$ converges uniformly to $f^{(k)}(x)$ on $[0,1]$.

Proof. We know from Theorem 7.1.5 that the above result holds for $k = 0$. For $k \geq 1$ we begin with the expression for $B_{n+k}^{(k)}(f;x)$ given in (7.16), and write

$$\Delta^k f\left(\frac{r}{n+k}\right) = \frac{f^{(k)}(\xi_r)}{(n+k)^k},$$

where $r/(n+k) < \xi_r < (r+k)/(n+k)$, as we did similarly in (7.21). We then approximate $f^{(k)}(\xi_r)$, writing

$$f^{(k)}(\xi_r) = f^{(k)}\left(\frac{r}{n}\right) + \left(f^{(k)}(\xi_r) - f^{(k)}\left(\frac{r}{n}\right)\right).$$

We thus obtain

$$\frac{n!(n+k)^k}{(n+k)!} B_{n+k}^{(k)}(f;x) = S_1(x) + S_2(x), \tag{7.27}$$

say, where

$$S_1(x) = \sum_{r=0}^{n} f^{(k)}\left(\frac{r}{n}\right)\binom{n}{r} x^r(1-x)^{n-r}$$

and

$$S_2(x) = \sum_{r=0}^{n} \left(f^{(k)}(\xi_r) - f^{(k)}\left(\frac{r}{n}\right)\right)\binom{n}{r} x^r(1-x)^{n-r}.$$

In $S_2(x)$, we can make $|\xi_r - \frac{r}{n}| < \delta$ for all r, for any choice of $\delta > 0$, by taking n sufficiently large. Also, given any $\epsilon > 0$, we can choose a positive value of δ such that

$$\left|f^{(k)}(\xi_r) - f^{(k)}\left(\frac{r}{n}\right)\right| < \epsilon,$$

for all r, because of the uniform continuity of $f^{(k)}$. Thus $S_2(x) \to 0$ uniformly on $[0,1]$ as $n \to \infty$. We can easily verify that

$$\frac{n!(n+k)^k}{(n+k)!} \to 1 \quad \text{as} \quad n \to \infty,$$

and we see from Theorem 7.1.5 with $f^{(k)}$ in place of f that $S_1(x)$ converges uniformly to $f^{(k)}(x)$ on $[0,1]$. This completes the proof. ■

As we have just seen, not only does the Bernstein polynomial for f converge to f, but derivatives converge to derivatives. This is a most remarkable property. In contrast, consider again the sequence of interpolating polynomials (p_n^*) for e^x that appear in Example 2.4.4. Although this sequence of polynomials converges uniformly to e^x on $[-1,1]$, this does *not* hold for their derivatives, because of the oscillatory nature of the error of interpolation.

On comparing the complexity of the proofs of Theorems 7.1.5 and 7.1.6, it may seem surprising that the *additional* work required to complete the proof of Theorem 7.1.6 for $k \geq 1$ is so little compared to that needed to prove Theorem 7.1.5.

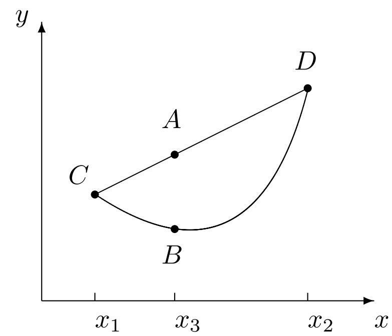

FIGURE 7.3. A and B are the points on the chord CD and on the graph of the convex function $y = f(x)$, respectively, with abscissa $x_3 = \lambda x_1 + (1 - \lambda)x_2$.

We now state results concerning the Bernstein polynomials for a convex function f. First we define convexity and show its connection with second-order divided differences.

Definition 7.1.2 A function f is said to be convex on $[a, b]$ if for any $x_1, x_2 \in [a, b]$,

$$\lambda f(x_1) + (1 - \lambda)f(x_2) \geq f(\lambda x_1 + (1 - \lambda)x_2) \tag{7.28}$$

for any $\lambda \in [0, 1]$. Geometrically, this is just saying that a chord connecting any two points on the convex curve $y = f(x)$ is never below the curve. This is illustrated in Figure 7.3, where CD is such a chord, and the points A and B have y-coordinates $\lambda f(x_1) + (1 - \lambda)f(x_2)$ and $f(\lambda x_1 + (1 - \lambda)x_2)$, respectively. ■

If f is twice differentiable, f being convex is equivalent to f'' being nonnegative. Of course, functions can be convex without being differentiable. For example, we can have a convex polygonal arc.

Theorem 7.1.7 A function f is convex on $[a, b]$ if and only if all second-order divided differences of f are nonnegative.

Proof. Since a divided difference is unchanged if we alter the order of its arguments, as we see from the symmetric form (1.21), it suffices to consider the divided difference $f[x_0, x_1, x_2]$ where $a \leq x_0 < x_1 < x_2 \leq b$. Then we obtain from the recurrence relation (1.22) that

$$f[x_0, x_1, x_2] \geq 0 \quad \Leftrightarrow \quad f[x_1, x_2] \geq f[x_0, x_1]. \tag{7.29}$$

On multiplying the last inequality throughout by $(x_2 - x_1)(x_1 - x_0)$, which is positive, we find that both inequalities in (7.29) are equivalent to

$$(x_1 - x_0)(f(x_2) - f(x_1)) \geq (x_2 - x_1)(f(x_1) - f(x_0)),$$

which is equivalent to

$$(x_1 - x_0)f(x_2) + (x_2 - x_1)f(x_0) \geq (x_2 - x_0)f(x_1). \tag{7.30}$$

If we now divide throughout by $x_2 - x_0$ and write $\lambda = (x_2 - x_1)/(x_2 - x_0)$, we see that $x_1 = \lambda x_0 + (1 - \lambda)x_2$, and it follows from (7.30) that

$$\lambda f(x_0) + (1 - \lambda)f(x_2) \geq f(\lambda x_0 + (1 - \lambda)x_2),$$

thus completing the proof. ■

The proofs of the following two theorems are held over until Section 7.3, where we will state and prove generalizations of both results.

Theorem 7.1.8 If $f(x)$ is convex on $[0, 1]$, then

$$B_n(f; x) \geq f(x), \quad 0 \leq x \leq 1, \tag{7.31}$$

for all $n \geq 1$. ■

Theorem 7.1.9 If $f(x)$ is convex on $[0, 1]$,

$$B_{n-1}(f; x) \geq B_n(f; x), \quad 0 \leq x \leq 1, \tag{7.32}$$

for all $n \geq 2$. The Bernstein polynomials are equal at $x = 0$ and $x = 1$, since they interpolate f at these points. If $f \in C[0, 1]$, the inequality in (7.32) is *strict* for $0 < x < 1$, for a given value of n, unless f is linear in each of the intervals $\left[\frac{r-1}{n-1}, \frac{r}{n-1}\right]$, for $1 \leq r \leq n - 1$, when we have simply $B_{n-1}(f; x) = B_n(f; x)$. ■

Note that we have from Theorem 7.1.4 with $k = 2$ that if $f''(x) \geq 0$, and thus f is convex on $[0, 1]$, then $B_n(f; x)$ is also convex on $[0, 1]$. In Section 7.3 we will establish the stronger result that $B_n(f; x)$ is convex on $[0, 1]$, provided that f is convex on $[0, 1]$.

We conclude this section by stating two theorems concerned with estimating the error $f(x) - B_n(f; x)$. The first of these is the theorem due to Elizaveta V. Voronovskaya (1898–1972), which gives an asymptotic error term for the Bernstein polynomials for functions that are twice differentiable.

Theorem 7.1.10 Let $f(x)$ be bounded on $[0, 1]$. Then, for any $x \in [0, 1]$ at which $f''(x)$ exists,

$$\lim_{n \to \infty} n(B_n(f; x) - f(x)) = \frac{1}{2}x(1 - x)f''(x). \quad ■ \tag{7.33}$$

See Davis [10] for a proof of Voronovskaya's theorem.

Finally, there is the following result that gives an upper bound for the error $f(x) - B_n(f; x)$ in terms of the modulus of continuity, which we defined in Section 2.6.

Theorem 7.1.11 If f is bounded on $[0, 1]$, then

$$\|f - B_n f\| \leq \frac{3}{2}\,\omega\left(\frac{1}{\sqrt{n}}\right), \tag{7.34}$$

where $\|\cdot\|$ denotes the maximum norm on $[0, 1]$. ■

See Rivlin [48] for a proof of this theorem.

Example 7.1.4 Consider the Bernstein polynomial for $f(x) = |x - \frac{1}{2}|$,

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

The difference between $B_n(f; x)$ and $f(x)$ at $x = \frac{1}{2}$ is

$$\frac{1}{2^n} \sum_{r=0}^{n} \left|\frac{r}{n} - \frac{1}{2}\right| \binom{n}{r} = e_n,$$

say. Let us now choose n to be even. We note that

$$\left(\frac{1}{2} - \frac{r}{n}\right)\binom{n}{r} = \left(\frac{n-r}{n} - \frac{1}{2}\right)\binom{n}{n-r}$$

for all r, and that the quantities on each side of the above equation are zero when $r = n/2$. It follows that

$$2^n e_n = \sum_{r=0}^{n} \left|\frac{r}{n} - \frac{1}{2}\right| \binom{n}{r} = 2 \sum_{r=0}^{n/2} \left(\frac{1}{2} - \frac{r}{n}\right)\binom{n}{r}. \tag{7.35}$$

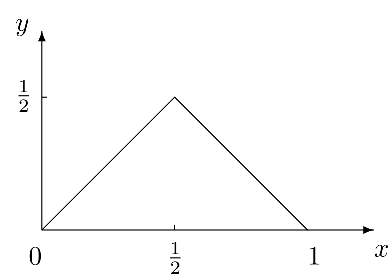

FIGURE 7.4. The function $f(x) = |x - \frac{1}{2}|$ on $[0, 1]$.

Let us split the last summation into two. We obtain

$$\sum_{r=0}^{n/2} \binom{n}{r} = \frac{1}{2}\left(\binom{n}{n/2} + \sum_{r=0}^{n} \binom{n}{r}\right) = \frac{1}{2}\binom{n}{n/2} + 2^{n-1},$$

and since

$$\frac{r}{n}\binom{n}{r} = \binom{n-1}{r-1}, \qquad r \geq 1,$$

we find that

$$2\sum_{r=0}^{n/2} \frac{r}{n}\binom{n}{r} = 2\sum_{r=1}^{n/2}\binom{n-1}{r-1} = \sum_{r=1}^{n}\binom{n-1}{r-1} = 2^{n-1}.$$

It then follows from (7.35) that

$$e_n = \frac{1}{2^{n+1}}\binom{n}{n/2} \sim \frac{1}{\sqrt{2\pi n}}$$

for n large. The last step follows on using Stirling's formula for estimating $n!$ (see Problem 2.1.12). This shows that $\|f - B_n f\| \to 0$ at least as slowly as $1/\sqrt{n}$ for the function $f(x) = |x - \frac{1}{2}|$, where $\|\cdot\|$ denotes the maximum norm on $[0, 1]$. ■

Problem 7.1.1 Show that

$$B_n(x^3; x) = x^3 + \frac{1}{n^2}x(1-x)(1+(3n-2)x),$$

for all $n \geq 3$.

Problem 7.1.2 Show that

$$B_n(e^{\alpha x}; x) = (xe^{\alpha/n} + (1-x))^n$$

for all integers $n \geq 1$ and all real α.

Problem 7.1.3 Deduce from Definition 7.1.2, using induction on n, that a function f is convex on $[a, b]$ if and only if

$$\sum_{r=0}^{n} \lambda_r f(x_r) \geq f\left(\sum_{r=0}^{n} \lambda_r x_r\right),$$

for all $n \geq 0$, for all $x_r \in [a, b]$, and for all $\lambda_r \geq 0$ such that

$$\lambda_0 + \lambda_1 + \cdots + \lambda_n = 1.$$

Problem 7.1.4 Find a function f and a real number λ such that f is a polynomial of degree two and $B_n f = \lambda f$. Also find a function f and a real number λ such that f is a polynomial of degree three and $B_n f = \lambda f$.

Problem 7.1.5 Verify Voronovskaya's Theorem 7.1.10 directly for the two functions x^2 and x^3.

7.2 The Monotone Operator Theorem

In the 1950s, H. Bohman [5] and P. P. Korovkin [31] obtained an amazing generalization of Bernstein's Theorem 7.1.5. They found that as far as convergence is concerned, the crucial properties of the Bernstein operator B_n are that $B_n f \to f$ uniformly on $[0,1]$ for $f = 1$, x, and x^2, and that B_n is a monotone linear operator. (See Definition 7.1.1.) We now state the Bohman–Korovkin theorem, followed by a proof based on that given by Cheney [7].

Theorem 7.2.1 Let (L_n) denote a sequence of monotone linear operators that map a function $f \in C[a,b]$ to a function $L_n f \in C[a,b]$, and let $L_n f \to f$ uniformly on $[a,b]$ for $f = 1$, x, and x^2. Then $L_n f \to f$ uniformly on $[a,b]$ for all $f \in C[a,b]$.

Proof. Let us define $\phi_t(x) = (t-x)^2$, and consider $(L_n\phi_t)(t)$. Thus we apply the linear operator L_n to ϕ_t, regarded as a function of x, and then evaluate $L_n\phi_t$ (which is also a function of x) at $x = t$. Since L_n is linear, we obtain

$$(L_n\phi_t)(t) = t^2(L_n g_0)(t) - 2t(L_n g_1)(t) + (L_n g_2)(t),$$

where

$$g_0(x) = 1, \qquad g_1(x) = x, \qquad g_2(x) = x^2.$$

Hence

$$(L_n\phi_t)(t) = t^2[(L_n g_0)(t) - 1] - 2t[(L_n g_1)(t) - t] + [(L_n g_2)(t) - t^2].$$

On writing $\|\cdot\|$ to denote the maximum norm on $[a,b]$, we deduce that

$$\|L_n\phi_t\| \leq M^2\|L_n g_0 - g_0\| + 2M\|L_n g_1 - g_1\| + \|L_n g_2 - g_2\|,$$

where $M = \max(|a|, |b|)$. Since for $i = 0$, 1, and 2, each term $\|L_n g_i - g_i\|$ may be made as small as we please, by taking n sufficiently large, it follows that

$$(L_n\phi_t)(t) \to 0 \quad \text{as} \quad n \to \infty, \tag{7.36}$$

uniformly in t.

Now let f be any function in $C[a,b]$. Given any $\epsilon > 0$, it follows from the uniform continuity of f that there exists a $\delta > 0$ such that for all $t, x \in [a,b]$,

$$|t - x| < \delta \Rightarrow |f(t) - f(x)| < \epsilon. \tag{7.37}$$

On the other hand, if $|t - x| \geq \delta$, we have

$$|f(t) - f(x)| \leq 2\|f\| \leq 2\|f\|\frac{(t-x)^2}{\delta^2} = \alpha\phi_t(x), \tag{7.38}$$

say, where $\alpha = 2\|f\|/\delta^2 > 0$. Note that $\phi_t(x) \geq 0$. Then, from (7.37) and (7.38), we see that for *all* $t, x \in [a, b]$,

$$|f(t) - f(x)| \leq \epsilon + \alpha\phi_t(x),$$

and so

$$-\epsilon - \alpha\phi_t(x) \leq f(t) - f(x) \leq \epsilon + \alpha\phi_t(x). \tag{7.39}$$

At this stage we make use of the monotonicity of the linear operator L_n. We apply L_n to each term in (7.39), regarded as a function of x, and evaluate each resulting function of x at the point t, to give

$$\begin{aligned}-\epsilon(L_n g_0)(t) - \alpha(L_n\phi_t)(t) &\leq f(t)(L_n g_0)(t) - (L_n f)(t)\\ &\leq \epsilon(L_n g_0)(t) + \alpha(L_n\phi_t)(t).\end{aligned}$$

Observe that $(L_n\phi_t)(t) \geq 0$, since L_n is monotone and $\phi_t(x) \geq 0$. Thus we obtain the inequality

$$|f(t)(L_n g_0)(t) - (L_n f)(t)| \leq \epsilon\|L_n g_0\| + \alpha(L_n\phi_t)(t). \tag{7.40}$$

If we now write $L_n g_0 = 1 + (L_n g_0 - g_0)$, we obtain

$$\|L_n g_0\| \leq 1 + \|L_n g_0 - g_0\|,$$

and so derive the inequality

$$|f(t)(L_n g_0)(t) - (L_n f)(t)| \leq \epsilon(1 + \|L_n g_0 - g_0\|) + \alpha(L_n\phi_t)(t). \tag{7.41}$$

We now write

$$f(t) - (L_n f)(t) = [f(t)(L_n g_0)(t) - (L_n f)(t)] + [f(t) - f(t)(L_n g_0)(t)],$$

and hence obtain the inequality

$$\begin{aligned}|f(t) - (L_n f)(t)| \leq\, &|f(t)(L_n g_0)(t) - (L_n f)(t)|\\ &+|f(t) - f(t)(L_n g_0)(t)|.\end{aligned} \tag{7.42}$$

In (7.41) we have already obtained an upper bound for the first term on the right of (7.42), and the second term satisfies the inequality

$$|f(t) - f(t)(L_n g_0)(t)| \leq \|f\|\;\|L_n g_0 - g_0\|. \tag{7.43}$$

Then, on substituting the two inequalities (7.41) and (7.43) into (7.42), we find that

$$|f(t) - (L_n f)(t)| \leq \epsilon + (\|f\| + \epsilon)\,\|L_n g_0 - g_0\| + \alpha(L_n\phi_t)(t). \tag{7.44}$$

On the right side of the above inequality we have ϵ plus two nonnegative quantities, each of which can be made less than ϵ for all n greater than some sufficiently large number N, and so

$$|f(t) - (L_n f)(t)| < 3\epsilon,$$

uniformly in t, for all $n > N$. This completes the proof. ■

Remark If we go through the above proof again, we can see that the following is a valid alternative version of the statement of Theorem 7.2.1. We will find this helpful when we discuss the Hermite–Fejér operator.

Let (L_n) denote a sequence of monotone linear operators that map functions $f \in C[a,b]$ to functions $L_n f \in C[a,b]$. Then if $L_n g_0 \to g_0$ uniformly on $[0,1]$, and $(L_n \phi_t)(t) \to 0$ uniformly in t on $[0,1]$, where g_0 and ϕ_t are defined in the proof of Theorem 7.2.1, it follows that $L_n f \to f$ uniformly on $[0,1]$ for all $f \in C[a,b]$. ■

Example 7.2.1 We see from Examples 7.1.1 and 7.1.2 that

$$B_n(1;x) = 1, \quad B_n(x;x) = x, \quad \text{and} \quad B_n(x^2;x) = x^2 + \frac{1}{n}x(1-x).$$

Thus $B_n(f;x)$ converges uniformly to $f(x)$ on $[0,1]$ for $f(x) = 1$, x, and x^2, and since the Bernstein operator B_n is also linear and monotone, it follows from the Bohman–Korovkin Theorem 7.2.1 that $B_n(f;x)$ converges uniformly to $f(x)$ on $[0,1]$ for all $f \in C[0,1]$, as we already found in Bernstein's Theorem 7.1.5. ■

We now recall the Hermite interpolating polynomial p_{2n+1}, defined by (1.38). If we write

$$q_{2n+1}(x) = \sum_{i=0}^{n} [a_i u_i(x) + b_i v_i(x)], \tag{7.45}$$

where u_i and v_i are defined in (1.39) and (1.40), then

$$q_{2n+1}(x_i) = a_i, \quad q'_{2n+1}(x_i) = b_i, \quad 0 \le i \le n. \tag{7.46}$$

If we now choose

$$a_i = f(x_i), \quad b_i = 0, \quad 0 \le i \le n, \tag{7.47}$$

where the x_i are the zeros of the Chebyshev polynomial T_{n+1} and f is a given function defined on $[-1,1]$, it follows that

$$q_{2n+1}(x) = \sum_{i=0}^{n} f(x_i) u_i(x) = (L_n f)(x), \tag{7.48}$$

say, where u_i is given by (2.103), and so

$$(L_n f)(x) = \left(\frac{T_{n+1}(x)}{n+1} \right)^2 \sum_{i=0}^{n} f(x_i) \frac{1 - x_i x}{(x - x_i)^2}. \tag{7.49}$$

We call L_n the Hermite–Fejér operator. It is clear that L_n is a linear operator, and since

$$0 \leq 1 - x_i x \leq 2 \quad \text{for} \quad -1 \leq x \leq 1, \tag{7.50}$$

for all i, we see that L_n is monotone. We also note that L_n reproduces the function 1, since the derivative of 1 is zero and $(L_n 1)(x)$ interpolates 1 on the Chebyshev zeros. It is also obvious that L_n does *not* reproduce the functions x and x^2, since their first derivatives are not zero on all the Chebyshev zeros. Let us apply L_n to the function $\phi_t(x) = (t-x)^2$. We obtain

$$(L_n\phi_t)(t) = \left(\frac{T_{n+1}(t)}{n+1}\right)^2 \sum_{i=0}^{n} (1 - x_i t),$$

and it follows from (7.50) that

$$|(L_n\phi_t)(t)| \leq \frac{2}{n+1},$$

so that $(L_n\phi_t)(t) \to 0$ uniformly in t on $[-1,1]$. Thus, in view of the alternative statement of Theorem 7.2.1, given in the remark following the proof of the theorem, we deduce the following result as a special case of the Bohman–Korovkin Theorem 7.2.1.

Theorem 7.2.2 Let (L_n) denote the sequence of Hermite–Fejér operators, defined by (7.49). Then $L_n f \to f$ uniformly for all $f \in C[-1,1]$. ■

Theorem 7.2.2, like Bernstein's Theorem 7.1.5, gives a constructive proof of the Weierstrass theorem. A direct proof of Theorem 7.2.2, which does not explicitly use the Bohman–Korovkin theorem, is given in Davis [10]. We will give another application of the Bohman–Korovkin theorem in the next section.

We can show (see Problem 7.2.1) that the only linear monotone operator that reproduces 1, x, and x^2, and thus all quadratic polynomials, is the identity operator. This puts into perspective the behaviour of the Bernstein operator, which reproduces linear polynomials, but does not reproduce x^2, and the Hermite–Fejér operator, which does not reproduce x or x^2.

Problem 7.2.1 Let L denote a linear monotone operator acting on functions $f \in C[a,b]$ that reproduces 1, x, and x^2. Show that $(L_n\phi_t)(t) = 0$, where $\phi_t(x) = (t-x)^2$. By working through the proof of Theorem 7.2.1 show that for a given $f \in C[a,b]$, (7.40) yields

$$|f(t) - (Lf)(t)| \leq \epsilon$$

for all $t \in [a,b]$ and any given $\epsilon > 0$. Deduce that $Lf = f$ for all $f \in C[a,b]$, and thus L is the identity operator.

Problem 7.2.2 Deduce from Theorem 7.2.2 that

$$\lim_{n\to\infty}\left(\frac{T_{n+1}(x)}{n+1}\right)^2\sum_{i=0}^{n}\frac{1-x_i x}{(x-x_i)^2}=1,$$

where the x_i denote the zeros of T_{n+1}.

7.3 On the q-Integers

In view of the many interesting properties of the Bernstein polynomials, it is not surprising that several generalizations have been proposed. In this section we discuss a generalization based on the q-integers, which are defined in Section 1.5. Let us write

$$B_n(f;x)=\sum_{r=0}^{n} f_r \left[\begin{array}{c} n \\ r \end{array}\right] x^r \prod_{s=0}^{n-r-1}(1-q^s x) \tag{7.51}$$

for each positive integer n, where f_r denotes the value of the function f at $x=[r]/[n]$, the quotient of the q-integers $[r]$ and $[n]$, and $\left[\begin{array}{c} n \\ r \end{array}\right]$ denotes a q-binomial coefficient, defined in (1.116). Note that an empty product in (7.51) denotes 1. When we put $q=1$ in (7.51), we obtain the classical Bernstein polynomial, defined by (7.1), and in this section we consistently write $B_n(f;x)$ to mean the *generalized* Bernstein polynomial, defined by (7.51). In Section 7.5, whenever we need to emphasize the dependence of the generalized Bernstein polynomial on the parameter q, we will write $B_n^q(f;x)$ in place of $B_n(f;x)$.

We see immediately from (7.51) that

$$B_n(f;0)=f(0) \quad \text{and} \quad B_n(f;1)=f(1), \tag{7.52}$$

giving interpolation at the endpoints, as we have for the classical Bernstein polynomials. It is shown in Section 8.2 that every q-binomial coefficient is a polynomial in q (called a Gaussian polynomial) with coefficients that are all positive integers. It is thus clear that B_n, defined by (7.51), is a linear operator, and with $0<q<1$, it is a monotone operator that maps functions defined on $[0,1]$ to P_n. The following theorem involves q-differences, which are defined in Section 1.5. This result yields Theorem 7.1.1 when $q=1$.

Theorem 7.3.1 The generalized Bernstein polynomial may be expressed in the form

$$B_n(f;x)=\sum_{r=0}^{n}\left[\begin{array}{c} n \\ r \end{array}\right]\Delta_q^r f_0\, x^r, \tag{7.53}$$

where

$$\Delta_q^r f_j=\Delta_q^{r-1}f_{j+1}-q^{r-1}\Delta_q^{r-1}f_j, \quad r\geq 1,$$

with $\Delta_q^0 f_j=f_j=f([j]/[n])$.

Proof. Here we require the identity,

$$\prod_{s=0}^{n-r-1}(1-q^s x)=\sum_{s=0}^{n-r}(-1)^s q^{s(s-1)/2}\left[\begin{array}{c} n-r \\ s \end{array}\right] x^s, \qquad (7.54)$$

which is equivalent to (8.12) and reduces to a binomial expansion when we put $q=1$. Beginning with (7.51), and expanding the term consisting of the product of the factors $(1-q^s x)$, we obtain

$$B_n(f;x)=\sum_{r=0}^{n} f_r \left[\begin{array}{c} n \\ r \end{array}\right] x^r \sum_{s=0}^{n-r}(-1)^s q^{s(s-1)/2}\left[\begin{array}{c} n-r \\ s \end{array}\right] x^s.$$

Let us put $t=r+s$. Then, since

$$\left[\begin{array}{c} n \\ r \end{array}\right]\left[\begin{array}{c} n-r \\ s \end{array}\right]=\left[\begin{array}{c} n \\ t \end{array}\right]\left[\begin{array}{c} t \\ r \end{array}\right],$$

we may write the latter double sum as

$$\sum_{t=0}^{n}\left[\begin{array}{c} n \\ t \end{array}\right] x^t \sum_{r=0}^{t}(-1)^{t-r} q^{(t-r)(t-r-1)/2}\left[\begin{array}{c} t \\ r \end{array}\right] f_r=\sum_{t=0}^{n}\left[\begin{array}{c} n \\ t \end{array}\right]\Delta_q^t f_0\, x^t,$$

on using the expansion for a higher-order q-difference, as in (1.121). This completes the proof. ■

We see from (1.33) and (1.113) that

$$\frac{\Delta_q^k f(x_0)}{q^{k(k-1)/2}\,[k]!}=f[x_0,x_1,\ldots,x_k]=\frac{f^{(k)}(\xi)}{k!},$$

where $x_j=[j]$ and $\xi\in(x_0,x_k)$. Thus q-differences of the monomial x^k of order greater than k are zero, and we see from Theorem 7.3.1 that for all $n\geq k$, $B_n(x^k;x)$ is a polynomial of degree k.

We deduce from Theorem 7.3.1 that

$$B_n(1;x)=1. \qquad (7.55)$$

For $f(x)=x$ we have $\Delta_q^0 f_0=f_0=0$ and $\Delta_q^1 f_0=f_1-f_0=1/[n]$, and it follows from Theorem 7.3.1 that

$$B_n(x;x)=x. \qquad (7.56)$$

For $f(x)=x^2$ we have $\Delta_q^0 f_0=f_0=0$, $\Delta_q^1 f_0=f_1-f_0=1/[n]^2$, and

$$\Delta_q^2 f_0=f_2-(1+q)f_1+qf_0=\left(\frac{[2]}{[n]}\right)^2-(1+q)\left(\frac{[1]}{[n]}\right)^2.$$

We then find from Theorem 7.3.1 that

$$B_n(x^2;x) = x^2 + \frac{x(1-x)}{[n]}. \tag{7.57}$$

The above expressions for $B_n(1;x)$, $B_n(x;x)$, and $B_n(x^2;x)$ generalize their counterparts given earlier for the case $q=1$ and, with the help of Theorem 7.2.1, lead us to the following theorem on the convergence of the generalized Bernstein polynomials.

Theorem 7.3.2 Let (q_n) denote a sequence such that $0 < q_n < 1$ and $q_n \to 1$ as $n \to \infty$. Then, for any $f \in C[0,1]$, $B_n(f;x)$ converges uniformly to $f(x)$ on $[0,1]$, where $B_n(f;x)$ is defined by (7.51) with $q = q_n$.

Proof. We saw above from (7.55) and (7.56) that $B_n(f;x) = f(x)$ for $f(x) = 1$ and $f(x) = x$, and since $q_n \to 1$ as $n \to \infty$, we see from (7.57) that $B_n(f;x)$ converges uniformly to $f(x)$ for $f(x) = x^2$. Also, since $0 < q_n < 1$, it follows that B_n is a monotone operator, and the proof is completed by applying the Bohman–Korovkin Theorem 7.2.1. ■

We now state and prove the following generalizations of Theorems 7.1.8 and 7.1.9.

Theorem 7.3.3 If $f(x)$ is convex on $[0,1]$, then

$$B_n(f;x) \geq f(x), \quad 0 \leq x \leq 1, \tag{7.58}$$

for all $n \geq 1$ and for $0 < q \leq 1$.

Proof. For each $x \in [0,1]$, let us define

$$x_r = \frac{[r]}{[n]} \quad \text{and} \quad \lambda_r = \left[\begin{array}{c} n \\ r \end{array} \right] x^r \prod_{s=0}^{n-r+1} (1 - q^s x), \quad 0 \leq r \leq n.$$

We see that $\lambda_r \geq 0$ when $0 < q \leq 1$ and $x \in [0,1]$, and note from (7.55) and (7.56), respectively, that

$$\lambda_0 + \lambda_1 + \cdots + \lambda_n = 1$$

and

$$\lambda_0 x_0 + \lambda_1 x_1 + \cdots + \lambda_n x_n = x.$$

Then we obtain from the result in Problem 7.1.3 that if f is convex on $[0,1]$,

$$B_n(f;x) = \sum_{r=0}^{n} \lambda_r f(x_r) \geq f\left(\sum_{r=0}^{n} \lambda_r x_r\right) = f(x),$$

and this completes the proof. ■

Theorem 7.3.4 If $f(x)$ is convex on $[0,1]$,

$$B_{n-1}(f;x) \geq B_n(f;x), \quad 0 \leq x \leq 1, \tag{7.59}$$

for all $n \geq 2$, where $B_{n-1}(f;x)$ and $B_n(f;x)$ are evaluated using the *same* value of the parameter q. The Bernstein polynomials are equal at $x = 0$ and $x = 1$, since they interpolate f at these points. If $f \in C[0,1]$, the inequality in (7.59) is *strict* for $0 < x < 1$ unless, for a given value of n, the function f is linear in each of the intervals $\left[\frac{[r-1]}{[n-1]}, \frac{[r]}{[n-1]}\right]$, for $1 \leq r \leq n-1$, when we have simply $B_{n-1}(f;x) = B_n(f;x)$.

Proof. In the proof given by Davis [10] for the special case of this theorem when $q = 1$, the difference between two consecutive Bernstein polynomials is expressed in terms of powers of $x/(1-x)$. This is not appropriate for $q \neq 1$, and our proof follows that given by Oruç and Phillips [40]. For $0 < q < 1$, let us write

$$\begin{aligned}
&(B_{n-1}(f;x) - B_n(f;x)) \prod_{s=0}^{n-1} (1-q^s x)^{-1} \\
&\quad = \sum_{r=0}^{n-1} f\left(\frac{[r]}{[n-1]}\right) \left[\begin{array}{c} n-1 \\ r \end{array}\right] x^r \prod_{s=n-r-1}^{n-1} (1-q^s x)^{-1} \\
&\quad\quad - \sum_{r=0}^{n} f\left(\frac{[r]}{[n]}\right) \left[\begin{array}{c} n \\ r \end{array}\right] x^r \prod_{s=n-r}^{n-1} (1-q^s x)^{-1}.
\end{aligned}$$

We now split the first of the above summations into two, writing

$$x^r \prod_{s=n-r-1}^{n-1} (1-q^s x)^{-1} = \psi_r(x) + q^{n-r-1} \psi_{r+1}(x),$$

say, where

$$\psi_r(x) = x^r \prod_{s=n-r}^{n-1} (1-q^s x)^{-1}. \tag{7.60}$$

On combining the resulting three summations, the terms in $\psi_0(x)$ and $\psi_n(x)$ cancel, and we obtain

$$(B_{n-1}(f;x) - B_n(f;x)) \prod_{s=0}^{n-1} (1-q^s x)^{-1} = \sum_{r=1}^{n-1} \left[\begin{array}{c} n \\ r \end{array}\right] a_r \psi_r(x), \tag{7.61}$$

where

$$a_r = \frac{[n-r]}{[n]} f\left(\frac{[r]}{[n-1]}\right) + q^{n-r} \frac{[r]}{[n]} f\left(\frac{[r-1]}{[n-1]}\right) - f\left(\frac{[r]}{[n]}\right). \tag{7.62}$$

It is clear from (7.60) that each $\psi_r(x)$ is nonnegative on $[0,1]$ for $0 \le q \le 1$, and thus from (7.61), it will suffice to show that each a_r is nonnegative. Let us write

$$\lambda = \frac{[n-r]}{[n]}, \quad x_1 = \frac{[r]}{[n-1]}, \quad \text{and} \quad x_2 = \frac{[r-1]}{[n-1]}.$$

It then follows that

$$1 - \lambda = q^{n-r}\frac{[r]}{[n]} \quad \text{and} \quad \lambda x_1 + (1-\lambda)x_2 = \frac{[r]}{[n]},$$

and we see immediately, on comparing (7.62) and (7.28), that

$$a_r = \lambda f(x_1) + (1-\lambda)f(x_2) - f(\lambda x_1 + (1-\lambda)x_2) \ge 0,$$

and so $B_{n-1}(f;x) \ge B_n(f;x)$. We obviously have equality at $x = 0$ and $x = 1$, since all Bernstein polynomials interpolate f at these endpoints. The inequality will be strict for $0 < x < 1$ unless every a_r is zero; this can occur only when f is linear in each of the intervals between consecutive points $[r]/[n-1]$, $0 \le r \le n-1$, when we have $B_{n-1}(f;x) = B_n(f;x)$ for $0 < x < 1$. This completes the proof. ■

We now give an algorithm, first published in 1996 (see Phillips [42]), for evaluating the generalized Bernstein polynomials. When $q = 1$ it reduces to the well-known de Casteljau algorithm (see Hoschek and Lasser [26]) for evaluating the classical Bernstein polynomials.

Algorithm 7.3.1 This algorithm begins with the value of q and the values of f at the $n+1$ points $[r]/[n]$, $0 \le r \le n$, and computes $B_n(f;x) = f_0^{[n]}$, which is the final number generated by the algorithm.

input: q; $f([0]/[n]), f([1]/[n]), \ldots, f([n]/[n])$
 for $r = 0$ **to** n
  $f_r^{[0]} := f([r]/[n])$
 next r
 for $m = 1$ **to** n
  for $r = 0$ **to** $n - m$
   $f_r^{[m]} := (q^r - q^{m-1}x)f_r^{[m-1]} + x f_{r+1}^{[m-1]}$
  next r
 next m
output: $f_0^{[n]} = B_n(f;x)$ ■

The following theorem justifies the above algorithm.

Theorem 7.3.5 For $0 \le m \le n$ and $0 \le r \le n - m$, we have

$$f_r^{[m]} = \sum_{t=0}^{m} f_{r+t} \left[\begin{array}{c} m \\ t \end{array} \right] x^t \prod_{s=0}^{m-t-1} (q^r - q^s x), \tag{7.63}$$

and, in particular,

$$f_0^{[n]} = B_n(f;x). \tag{7.64}$$

Proof. We use induction on m. From the initial conditions in the algorithm, $f_r^{[0]} := f([r]/[n]) = f_r,\ \ 0 \le r \le n$, it is clear that (7.63) holds for $m = 0$ and $0 \le r \le n$. Note that an empty product in (7.63) denotes 1. Let us assume that (7.63) holds for some m such that $0 \le m < n$, and for all r such that $0 \le r \le n - m$. Then, for $0 \le r \le n - m - 1$, it follows from the algorithm that

$$f_r^{[m+1]} = (q^r - q^m x) f_r^{[m]} + x f_{r+1}^{[m]},$$

and using (7.63), we obtain

$$\begin{aligned} f_r^{[m+1]} = {} & (q^r - q^m x) \sum_{t=0}^{m} f_{r+t} \begin{bmatrix} m \\ t \end{bmatrix} x^t \prod_{s=0}^{m-t-1} (q^r - q^s x) \\ & + x \sum_{t=0}^{m} f_{r+t+1} \begin{bmatrix} m \\ t \end{bmatrix} x^t \prod_{s=0}^{m-t-1} (q^{r+1} - q^s x). \end{aligned}$$

The coefficient of f_r on the right of the latter equation is

$$(q^r - q^m x) \prod_{s=0}^{m-1} (q^r - q^s x) = \prod_{s=0}^{m} (q^r - q^s x),$$

and the coefficient of f_{r+m+1} is x^{m+1}. For $1 \le t \le m$, we find that the coefficient of f_{r+t} is

$$\begin{aligned} & (q^r - q^m x) \begin{bmatrix} m \\ t \end{bmatrix} x^t \prod_{s=0}^{m-t-1} (q^r - q^s x) + \begin{bmatrix} m \\ t-1 \end{bmatrix} x^t \prod_{s=0}^{m-t} (q^{r+1} - q^s x) \\ & = a_t x^t \prod_{s=0}^{m-t-1} (q^r - q^s x), \end{aligned}$$

say. We see that

$$a_t = (q^r - q^m x) \begin{bmatrix} m \\ t \end{bmatrix} + q^{m-t}(q^{r+1} - x) \begin{bmatrix} m \\ t-1 \end{bmatrix}$$

and hence

$$a_t = q^r \left(\begin{bmatrix} m \\ t \end{bmatrix} + q^{m+1-t} \begin{bmatrix} m \\ t-1 \end{bmatrix} \right) - q^{m-t} x \left(q^t \begin{bmatrix} m \\ t \end{bmatrix} + \begin{bmatrix} m \\ t-1 \end{bmatrix} \right).$$

It is easily verified (see (8.7) and (8.8)) that

$$\begin{bmatrix} m \\ t \end{bmatrix} + q^{m+1-t} \begin{bmatrix} m \\ t-1 \end{bmatrix} = q^t \begin{bmatrix} m \\ t \end{bmatrix} + \begin{bmatrix} m \\ t-1 \end{bmatrix} = \begin{bmatrix} m+1 \\ t \end{bmatrix}$$

and thus

$$a_t = (q^r - q^{m-t}x) \left[\begin{array}{c} m+1 \\ t \end{array} \right].$$

Hence the coefficient of f_{r+t}, for $1 \le t \le m$, in the above expression for $f_r^{[m+1]}$ is

$$\left[\begin{array}{c} m+1 \\ t \end{array} \right] x^t \prod_{s=0}^{m-t} (q^r - q^s x),$$

and we note that this also holds for $t = 0$ and $t = m+1$. Thus we obtain

$$f_r^{[m+1]} = \sum_{t=0}^{m+1} f_{r+t} \left[\begin{array}{c} m+1 \\ t \end{array} \right] x^t \prod_{s=0}^{m-t} (q^r - q^s x),$$

and this completes the proof by induction. ■

The above algorithm for evaluating $B_n(f;x)$ is not unlike Algorithm 1.1.1 (Neville–Aitken). In the latter algorithm, each quantity that is computed is, like the final result, an interpolating polynomial on certain abscissas. Similarly, in Algorithm 7.3.1, as we see in (7.63), each intermediate number $f_r^{[m]}$ has a form that resembles that of the final number $f_0^{[n]} = B_n(f;x)$. We now show that each $f_r^{[m]}$ can also be expressed simply in terms of q-differences, as we have for $B_n(f;x)$ in (7.53).

Theorem 7.3.6 For $0 \le m \le n$ and $0 \le r \le n-m$, we have

$$f_r^{[m]} = \sum_{s=0}^{m} q^{(m-s)r} \left[\begin{array}{c} m \\ s \end{array} \right] \Delta_q^s f_r \, x^s. \tag{7.65}$$

Proof. We may verify (7.65) by induction on m, using the recurrence relation in Algorithm 7.3.1. Alternatively, we can derive (7.65) from (7.63) as follows. First we write

$$\prod_{s=0}^{m-t-1} (q^r - q^s x) = q^{(m-t)r} \prod_{s=0}^{m-t-1} (1 - q^s y),$$

where $y = x/q^r$, and we find with the aid of (7.54) that

$$\prod_{s=0}^{m-t-1} (q^r - q^s x) = \sum_{j=0}^{m-t} (-1)^j q^{j(j-1)/2 + (m-t-j)r} \left[\begin{array}{c} m-t \\ j \end{array} \right] x^j.$$

On substituting this into (7.63), we obtain

$$f_r^{[m]} = \sum_{t=0}^{m} f_{r+t} \left[\begin{array}{c} m \\ t \end{array} \right] x^t \sum_{j=0}^{m-t} (-1)^j q^{j(j-1)/2 + (m-t-j)r} \left[\begin{array}{c} m-t \\ j \end{array} \right] x^j.$$

If we now let $s = j + t$, we may rewrite the above double summation as

$$\sum_{s=0}^{m} q^{(m-s)r} \left[\begin{array}{c} m \\ s \end{array} \right] x^s \sum_{j=0}^{s} (-1)^j q^{j(j-1)/2} \left[\begin{array}{c} s \\ j \end{array} \right] f_{r+s-j},$$

which, in view of (1.121), gives

$$f_r^{[m]} = \sum_{s=0}^{m} q^{(m-s)r} \left[\begin{array}{c} m \\ s \end{array} \right] \Delta_q^s f_r \, x^s,$$

and this completes the proof. ■

Problem 7.3.1 Verify (7.65) directly by induction on m, using the recurrence relation in Algorithm 7.3.1.

Problem 7.3.2 Work through Algorithm 7.3.1 for the case $n = 2$, and so verify directly that $f_0^{[2]} = B_2(f; x)$.

7.4 Total Positivity

We begin this section by defining a totally positive matrix, and discuss the nature of linear transformations when the matrix is totally positive. We will apply these ideas in Section 7.5 to justify further properties of the Bernstein polynomials concerned with shape, such us convexity.

Definition 7.4.1 A real matrix $\mathbf{A}$ is called *totally positive* if all its minors are nonnegative, that is,

$$\mathbf{A} \left(\begin{array}{c} i_1, i_2, \ldots, i_k \\ j_1, j_2, \ldots, j_k \end{array} \right) = \det \left[\begin{array}{ccc} a_{i_1,j_1} & \cdots & a_{i_1,j_k} \\ \vdots & & \vdots \\ a_{i_k,j_1} & \cdots & a_{i_k,j_k} \end{array} \right] \geq 0, \tag{7.66}$$

for all $i_1 < i_2 < \cdots < i_k$ and all $j_1 < j_2 < \cdots < j_k$. We say that $\mathbf{A}$ is *strictly totally positive* if all its minors are *positive*, so that $\geq$ is replaced by $>$ in (7.66). ■

It follows, on putting $k = 1$ in (7.66), that a *necessary* condition for a matrix to be totally positive is that all its elements are nonnegative.

Theorem 7.4.1 A real matrix $\mathbf{A} = (a_{ij})$ is totally positive if

$$\mathbf{A} \left(\begin{array}{c} i, i+1, \ldots, i+k \\ j, j+1, \ldots, j+k \end{array} \right) \geq 0 \quad \text{for all } i, j, \text{ and } k. \tag{7.67}$$

Similarly, the matrix $\mathbf{A}$ is strictly totally positive if the minors given in (7.67), which are formed from consecutive rows and columns, are all positive. ■

For a proof, see Karlin [28]. In view of Theorem 7.4.1, we can determine whether $\mathbf{A}$ is totally positive or strictly totally positive by testing the positivity of only those minors that are formed from consecutive rows and columns, rather than having to examine *all* minors.

Example 7.4.1 Let us consider the Vandermonde matrix

$$\mathbf{V} = \mathbf{V}(x_0, \ldots, x_n) = \begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix}. \tag{7.68}$$

As we showed in Chapter 1 (see Problem 1.1.1),

$$\det \mathbf{V}(x_0, \ldots, x_n) = \prod_{i>j} (x_i - x_j). \tag{7.69}$$

Let $0 < x_0 < x_1 < \cdots < x_n$. Then we see from (7.69) that $\det \mathbf{V} > 0$, and we now prove that $\mathbf{V}$ is strictly totally positive. Using Theorem 7.4.1, it is sufficient to show that the minors

$$\det \begin{bmatrix} x_i^j & x_i^{j+1} & \cdots & x_i^{j+k} \\ x_{i+1}^j & x_{i+1}^{j+1} & \cdots & x_{i+1}^{j+k} \\ \vdots & \vdots & & \vdots \\ x_{i+k}^j & x_{i+k}^{j+1} & \cdots & x_{i+k}^{j+k} \end{bmatrix}$$

are positive for all nonnegative i, j, k such that $i+k, j+k \le n$. On removing common factors from its rows, the above determinant may be expressed as

$$(x_i \cdots x_{i+k})^j \det \mathbf{V}(x_i, \ldots, x_{i+k}) > 0,$$

since

$$\det \mathbf{V}(x_i, \ldots, x_{i+k}) = \prod_{i \le r < s \le i+k} (x_s - x_r) > 0.$$

This completes the proof that $\mathbf{V}$ is strictly totally positive. ■

If $\mathbf{A} = \mathbf{BC}$, where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ denote matrices of orders $m \times n, m \times k$, and $k \times n$, respectively, then

$$\mathbf{A}\begin{pmatrix} i_1, \ldots, i_p \\ j_1, \ldots, j_p \end{pmatrix} = \sum_{\beta_1 < \cdots < \beta_p} \!\!\cdots\!\! \sum \mathbf{B}\begin{pmatrix} i_1, \ldots, i_p \\ \beta_1, \ldots, \beta_p \end{pmatrix} \mathbf{C}\begin{pmatrix} \beta_1, \ldots, \beta_p \\ j_1, \cdots, j_p \end{pmatrix}. \tag{7.70}$$

This is known as the Cauchy–Binet determinant identity. It follows immediately from this most useful identity that the product of totally positive matrices is a totally positive matrix, and the product of strictly totally positive matrices is a strictly totally positive matrix.

Definition 7.4.2 Let $\mathbf{v}$ denote the sequence (v_i), which may be finite or infinite. Then we denote by $S^-(\mathbf{v})$ the number of strict sign changes in the sequence $\mathbf{v}$. ■

For example, $S^-(1,-2,3,-4,5,-6) = 5$, $S^-(1,0,0,1,-1) = 1$, and $S^-(1,-1,1,-1,1,-1,\ldots) = \infty$, where the last sequence alternates $+1$ and -1 indefinitely. It is clear that inserting zeros into a sequence, or deleting zeros from a sequence, does not alter the number of changes of sign. Also, deleting terms of a sequence does not increase the number of changes of sign. We use the same notation to denote sign changes in a function.

Definition 7.4.3 Let

$$v_i = \sum_{k=0}^{n} a_{ik} u_k, \quad i = 0, 1, \ldots, m,$$

where the u_k and the a_{ik}, and thus the v_i, are all real. This linear transformation is said to be *variation-diminishing* if

$$S^-(\mathbf{v}) \leq S^-(\mathbf{u}). \qquad ■$$

Definition 7.4.4 A matrix $\mathbf{A}$, which may be finite or infinite, is said to be *m-banded* if there exists an integer l such that $a_{ij} \neq 0$ implies that $l \leq j - i \leq l + m$. ■

This is equivalent to saying that all the nonzero elements of $\mathbf{A}$ lie on $m+1$ diagonals. We will say that a matrix $\mathbf{A}$ is *banded* if it is m-banded for some m. Note that every finite matrix is banded. We have already met 1-banded and 2-banded (tridiagonal) matrices in Chapter 6. In this section we will be particularly interested in 1-banded matrices, also called *bidiagonal* matrices, because of Theorem 7.4.3 below.

We now come to the first of the main results of this section.

Theorem 7.4.2 If $\mathbf{T}$ is a totally positive banded matrix and $\mathbf{v}$ is any vector for which $\mathbf{Tv}$ is defined, then

$$S^-(\mathbf{Tv}) \leq S^-(\mathbf{v}). \qquad ■$$

For a proof of this theorem see Goodman [22].

When we first encounter it, the question of whether a linear transformation is variation-diminishing may not seem very interesting. However, building on the concept of a variation-diminishing linear transformation, we will see in Section 7.5 that the number of sign changes in a function f defined on $[0,1]$ is not increased if we apply a Bernstein operator, and we say that Bernstein operators are shape-preserving. This property does not always hold, for example, for interpolating operators.

Example 7.4.2 Let $\mathbf{v}$ denote an infinite real sequence for which $S^-(\mathbf{v})$ is finite. Consider the sequence $\mathbf{w} = (w_i)_{i=0}^\infty$ defined by

$$w_i = v_i + v_{i-1} \quad \text{for} \quad i \geq 1, \quad \text{and} \quad w_0 = v_0.$$

Then $\mathbf{w} = \mathbf{Tv}$, where

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

Let us consider the minors of $\mathbf{T}$ constructed from consecutive rows and columns. Any such minor whose leading (top left) element is 0 has either a whole row or a whole column of zeros, and so the minor is zero. It is also not hard to see that any minor constructed from consecutive rows and columns whose leading element is 1 has itself the value 1. Thus, by Theorem 7.4.1, the matrix $\mathbf{T}$ is totally positive, and so we deduce from Theorem 7.4.2 that $S^-(\mathbf{w}) = S^-(\mathbf{Tv}) \leq S^-(\mathbf{v})$. ■

Theorem 7.4.3 A finite matrix is totally positive if and only if it is a product of 1-banded matrices with nonnegative elements. ■

For a proof of this theorem, see de Boor and Pinkus [13]. An immediate consequence of Theorem 7.4.3 is that the product of totally positive matrices is totally positive, as we have already deduced above from the Cauchy–Binet identity.

Example 7.4.3 To illustrate Theorem 7.4.3, consider the 1-banded factorization

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

The four matrices in the above product are indeed all 1-banded matrices with nonnegative elements, and their product is totally positive. ■

We now state a theorem, and give a related example, concerning the factorization of a matrix into the product of lower and upper triangular matrices.

Theorem 7.4.4 A matrix $\mathbf{A}$ is strictly totally positive if and only if it can be expressed in the form $\mathbf{A} = \mathbf{LU}$ where $\mathbf{L}$ is a lower triangular matrix, $\mathbf{U}$ is an upper triangular matrix, and both $\mathbf{L}$ and $\mathbf{U}$ are totally positive matrices. ■

For a proof, see Cryer [9].

Example 7.4.4 To illustrate Theorem 7.4.4, we continue Example 7.4.3, in which the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$$

is expressed as a product of four 1-banded matrices. If we multiply the first two of these 1-banded matrices, and also multiply the third and fourth, we obtain the **LU** factorization

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 2 \end{bmatrix} = \mathbf{LU},$$

and it is easy to verify that **L** and **U** are both totally positive. ■

The matrix **A** in Example 7.4.4 is the 3×3 Vandermonde matrix $\mathbf{V}(1,2,3)$. In Section 1.2 we gave (see Theorem 1.2.3) the **LU** factorization of the general Vandermonde matrix.

Example 7.4.5 Let

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Then we can easily verify that **A** is totally positive, and it is obviously not strictly totally positive. We give two different **LU** factorizations of **A**:

$$\mathbf{A} = \mathbf{LU} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where **L** is totally positive but **U** is not, and

$$\mathbf{A} = \mathbf{LU} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where both **L** and **U** are totally positive. This example shows that we cannot replace "strictly totally positive" by "totally positive" in the statement of Theorem 7.4.4. ■

Definition 7.4.5 For a real-valued function f on an interval I, we define $S^-(f)$ to be the number of sign changes of f, that is,

$$S^-(f) = \sup S^-(f(x_0), \ldots, f(x_m)),$$

where the supremum is taken over all increasing sequences $(x_0, \ldots, x_m)$ in I, for all m.

Definition 7.4.6 We say that a sequence $(\phi_0, \ldots, \phi_n)$ of real-valued functions on an interval I is totally positive if for any points $0 < x_0 < \cdots < x_n$ in I, the collocation matrix $(\phi_j(x_i))_{i,j=0}^n$ is totally positive. ■

Theorem 7.4.5 Let $\psi_i(x) = \omega(x)\phi_i(x)$, for $0 \leq i \leq n$. Then, if $\omega(x) \geq 0$ on I and the sequence of functions $(\phi_0, \ldots, \phi_n)$ is totally positive on I, the sequence $(\psi_0, \ldots, \psi_n)$ is also totally positive on I.

Proof. This follows easily from the definitions. ■

Theorem 7.4.6 If $(\phi_0, \ldots, \phi_n)$ is totally positive on I, then for any numbers $a_0, \ldots, a_n$,

$$S^-(a_0\phi_n + \cdots + a_n\phi_n) \leq S^-(a_0, \ldots, a_n). \quad ■$$

For a proof of this theorem see Goodman [22].

Definition 7.4.7 Let L denote a linear operator that maps each function f defined on an interval $[0, 1]$ onto Lf defined on $[0, 1]$. Then we say that L is variation-diminishing if

$$S^-(Lf) \leq S^-(f). \quad ■$$

Problem 7.4.1 Show that an $n \times n$ matrix has $(2^n - 1)^2$ minors, of which $\frac{1}{4}n^2(n + 1)^2$ are formed from consecutive rows and columns. How many minors are there in these two categories for an $m \times n$ matrix?

Problem 7.4.2 Show that the matrix

$$\begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 \\ 0 & a_0 & a_1 & a_2 & a_3 \\ 0 & 0 & a_0 & a_1 & a_2 \\ 0 & 0 & 0 & a_0 & a_1 \\ 0 & 0 & 0 & 0 & a_0 \end{bmatrix}$$

is totally positive, where $a_i \geq 0$ for all i, and $a_i^2 - a_{i-1}a_{i+1} \geq 0$ for $1 \leq i \leq 3$.

Problem 7.4.3 Let $\mathbf{v}$ denote an infinite real sequence for which $S^-(\mathbf{v})$ is finite. Consider the sequence $\mathbf{w}$ defined by

$$w_r = \sum_{s=0}^{r} v_s \quad \text{for} \quad r \geq 0.$$

Show that $S^-(\mathbf{w}) \leq S^-(\mathbf{v})$.

Problem 7.4.4 Repeat Problem 7.4.3 with the sequence $\mathbf{w}$ defined by

$$w_r = \sum_{s=0}^{r} \binom{r}{s} v_s \quad \text{for} \quad r \geq 0,$$

again showing that $S^-(\mathbf{w}) \leq S^-(\mathbf{v})$.

Problem 7.4.5 Show that the matrix

$$\begin{bmatrix} (x_0+t_0)^{-1} & (x_0+t_1)^{-1} & 1 & x_0 \\ (x_1+t_0)^{-1} & (x_1+t_1)^{-1} & 1 & x_1 \\ (x_2+t_0)^{-1} & (x_2+t_1)^{-1} & 1 & x_2 \\ (x_3+t_0)^{-1} & (x_3+t_1)^{-1} & 1 & x_3 \end{bmatrix}$$

is totally positive if $0 < x_0 < x_1 < x_2 < x_3$ and $0 < t_0 < t_1$.

7.5 Further Results

This section is based on the work of Goodman, Oruç and Phillips [23]. We will use the theory of total positivity, developed in the last section, to justify shape-preserving properties of the generalized Bernstein polynomials. We will also show that if a function f is convex on $[0,1]$, then for each x in $[0,1]$ the generalized Bernstein polynomial $B_n(f;x)$ approaches $f(x)$ monotonically from above as the parameter q increases, for $0 < q \leq 1$.

In the last section we saw that for $0 < x_0 < x_1 < \cdots < x_n$, the Vandermonde matrix $\mathbf{V}(x_0,\ldots,x_n)$ is strictly totally positive. It then follows from Definition 7.4.6 that the sequence of monomials $(x^i)_{i=0}^n$ is totally positive on any interval $[0,\infty)$. We now make the change of variable $t = x/(1-x)$, and note that t is an increasing function of x. Thus, if $t_i = x_i/(1-x_i)$, and we now let $0 < x_0 < x_1 < \cdots < x_n < 1$, it follows that $0 < t_0 < t_1 < \cdots < t_n$.

Since the Vandermonde matrix $\mathbf{V}(t_0,\ldots,t_n)$ is strictly totally positive, it follows that the sequence of functions

$$\left(1, \frac{x}{1-x}, \frac{x^2}{(1-x)^2}, \ldots, \frac{x^n}{(1-x)^n}\right)$$

is totally positive on $[0,1]$. We also see from Theorem 7.4.5 that the sequence of functions

$$\left((1-x)^n, x(1-x)^{n-1}, x^2(1-x)^{n-2}, \ldots, x^{n-1}(1-x), x^n\right) \tag{7.71}$$

is totally positive on $[0,1]$. Since the $n+1$ functions in the sequence (7.71) are a basis for P_n, the subspace of polynomials of degree at most n, they are a basis for all the classical Bernstein polynomials of degree n, defined by (7.1), and we can immediately deduce the following powerful result from Theorem 7.4.6.

Theorem 7.5.1 Let $B_n(f;x)$ denote the classical Bernstein polynomial of degree n for the function f. Then

$$S^-(B_n f) \le S^-(f) \tag{7.72}$$

for all f defined on $[0,1]$, and thus the classical Bernstein operator B_n is variation-diminishing.

Proof. Using Theorem 7.4.6, we have

$$S^-(B_n f) \le S^-(f_0, f_1, \ldots, f_n) \le S^-(f),$$

where $f_r = f(r/n)$. ■

For each q such that $0 < q \le 1$, and each $n \ge 1$, we now define

$$P^q_{n,j}(x) = x^j \prod_{s=0}^{n-j-1} (1 - q^s x), \quad 0 \le x \le 1, \tag{7.73}$$

for $0 \le j \le n$. These functions are a basis for P_n, and are thus a basis for all the generalized Bernstein polynomials of degree n, defined by (7.51). We have already seen above that $(P^1_{n,0}, P^1_{n,1}, \ldots, P^1_{n,n})$ is totally positive on $[0,1]$, and we will show below that the same holds for $(P^q_{n,0}, P^q_{n,1}, \ldots, P^q_{n,n})$, for any q such that $0 < q \le 1$.

Since the functions defined in (7.73) are a basis for P_n, it follows that for any choice of q and r satisfying $0 < q, r \le 1$, there exists a nonsingular matrix $\mathbf{T}^{n,q,r}$ such that

$$\begin{bmatrix} P^q_{n,0}(x) \\ \vdots \\ P^q_{n,n}(x) \end{bmatrix} = \mathbf{T}^{n,q,r} \begin{bmatrix} P^r_{n,0}(x) \\ \vdots \\ P^r_{n,n}(x) \end{bmatrix}. \tag{7.74}$$

Theorem 7.5.2 For $0 < q \le r \le 1$ all elements of the matrix $\mathbf{T}^{n,q,r}$ are nonnegative.

Proof. We use induction on n. Since $\mathbf{T}^{1,q,r}$ is the 2×2 identity matrix, its elements are obviously nonnegative. Let us assume that the elements of $\mathbf{T}^{n,q,r}$ are all nonnegative for some $n \ge 1$. Then, since

$$P^q_{n+1,j+1}(x) = x P^q_{n,j}(x), \quad 0 \le j \le n, \tag{7.75}$$

for all q such that $0 < q \le 1$, we see from (7.75) and (7.74) that

$$\begin{bmatrix} P^q_{n+1,1}(x) \\ \vdots \\ P^q_{n+1,n+1}(x) \end{bmatrix} = \mathbf{T}^{n,q,r} \begin{bmatrix} P^r_{n+1,1}(x) \\ \vdots \\ P^r_{n+1,n+1}(x) \end{bmatrix}. \tag{7.76}$$

Also, we have

$$P^q_{n+1,0}(x) = (1 - q^n x)P^q_{n,0}(x) = (1 - q^n x)\sum_{j=0}^{n} t^{n,q,r}_{0,j} P^r_{n,j}(x), \qquad (7.77)$$

where $(t^{n,q,r}_{0,0}, t^{n,q,r}_{0,1}, \ldots, t^{n,q,r}_{0,n})$ denotes the first row of the matrix $\mathbf{T}^{n,q,r}$. If we now substitute

$$(1 - q^n x)P^r_{n,j}(x) = P^r_{n+1,j}(x) + (r^{n-j} - q^n)P^r_{n+1,j+1}(x)$$

in the right side of (7.77) and simplify, we obtain

$$\begin{aligned} P^q_{n+1,0}(x) &= t^{n,q,r}_{0,0} P^r_{n+1,0}(x) + (1 - q^n)t^{n,q,r}_{0,n} P^r_{n+1,n+1}(x) \\ &\quad + \sum_{j=1}^{n} \left((r^{n+1-j} - q^n)t^{n,q,r}_{0,j-1} + t^{n,q,r}_{0,j}\right) P^r_{n+1,j}(x). \end{aligned} \qquad (7.78)$$

Then, on combining (7.76) and (7.78), we find that

$$\begin{bmatrix} P^q_{n+1,0}(x) \\ P^q_{n+1,1}(x) \\ \vdots \\ P^q_{n+1,n+1}(x) \end{bmatrix} = \begin{bmatrix} t^{n,q,r}_{0,0} & \mathbf{v}^T_{n+1} \\ & \\ \mathbf{0} & \mathbf{T}^{n,q,r} \end{bmatrix} \begin{bmatrix} P^r_{n+1,0}(x) \\ P^r_{n+1,1}(x) \\ \vdots \\ P^r_{n+1,n+1}(x) \end{bmatrix}, \qquad (7.79)$$

so that

$$\mathbf{T}^{n+1,q,r} = \begin{bmatrix} t^{n,q,r}_{0,0} & \mathbf{v}^T_{n+1} \\ & \\ \mathbf{0} & \mathbf{T}^{n,q,r} \end{bmatrix}. \qquad (7.80)$$

In the block matrix on the right side of (7.80) $\mathbf{0}$ denotes the zero vector with $n+1$ elements, and $\mathbf{v}^T_{n+1}$ is the row vector whose elements are the coefficients of $P^r_{n+1,1}(x), \ldots, P^r_{n+1,n+1}(x)$, given by (7.78). On substituting $x = 0$ in (7.80), it is clear that $t^{n,q,r}_{0,0} = 1$. We can deduce from (7.78) that if $0 < q \le r \le 1$, the elements of $\mathbf{v}^T_{n+1}$ are nonnegative, and this completes the proof by induction. ■

It follows from (7.80) and the definition of $\mathbf{v}^T_{n+1}$ that

$$t^{n+1,q,r}_{0,0} = t^{n,q,r}_{0,0} \qquad (7.81)$$

and

$$t^{n+1,q,r}_{0,j} = (r^{n+1-j} - q^n)t^{n,q,r}_{0,j-1} + t^{n,q,r}_{0,j}, \quad 1 \le j \le n. \qquad (7.82)$$

We will require this recurrence relation, which expresses the elements in the first row of $\mathbf{T}^{n+1,q,r}$ in terms of those in the first row of $\mathbf{T}^{n,q,r}$, in the proof of our next theorem. This shows that the transformation matrix $\mathbf{T}^{n,q,r}$ can be factorized as a product of 1-banded matrices. First we require the following lemma.

Lemma 7.5.1 For $m \geq 1$ and any real r and a, let $\mathbf{A}(m,a)$ denote the $m \times (m+1)$ matrix

$$\begin{bmatrix} 1 & r^m - a & & & & \\ & 1 & r^{m-1} - a & & & \\ & & \cdot & \cdot & & \\ & & & \cdot & \cdot & \\ & & & & 1 & r - a \end{bmatrix}.$$

Then

$$\mathbf{A}(m,a)\mathbf{A}(m+1,b) = \mathbf{A}(m,b)\mathbf{A}(m+1,a). \tag{7.83}$$

Proof. For $i = 0, \ldots, m-1$ the ith row of each side of (7.83) is

$$\left[0, \ldots, 0, 1, r^{m+1-i} + r^{m-i} - a - b, (r^{m-i} - a)(r^{m-i} - b), 0, \ldots, 0\right]. \quad \blacksquare$$

For $1 \leq j \leq n-1$, let $\mathbf{B}_j^{(n)}$ denote the 1-banded $(n+1) \times (n+1)$ matrix that has units on the main diagonal, and has the elements

$$r^j - q^{n-j}, r^{j-1} - q^{n-j}, \ldots, r - q^{n-j}, 0, \ldots, 0$$

on the diagonal above the main diagonal, where there are $n-j$ zeros at the lower end of that diagonal. Thus, for example,

$$\mathbf{B}_2^{(n)} = \begin{bmatrix} 1 & r^2 - q^{n-2} & & & & \\ & 1 & r - q^{n-2} & & & \\ & & 1 & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & 1 \end{bmatrix}.$$

The matrix $\mathbf{B}_j^{(n+1)}$ can be expressed in a block form involving the matrix $\mathbf{B}_j^{(n)}$. We can verify that

$$\mathbf{B}_1^{(n+1)} = \begin{bmatrix} 1 & \mathbf{c}_0^T \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{7.84}$$

and

$$\mathbf{B}_{j+1}^{(n+1)} = \begin{bmatrix} 1 & \mathbf{c}_j^T \\ \mathbf{0} & \mathbf{B}_j^{(n)} \end{bmatrix} \tag{7.85}$$

for $1 \leq j \leq n-1$, where each $\mathbf{c}_j^T$ is a row vector, $\mathbf{0}$ denotes the zero vector, and $\mathbf{I}$ is the unit matrix of order $n+1$.

Theorem 7.5.3 For $n \geq 2$ and any q, r, we have

$$\mathbf{T}^{n,q,r} = \mathbf{B}_1^{(n)}\mathbf{B}_2^{(n)} \cdots \mathbf{B}_{n-1}^{(n)}, \tag{7.86}$$

where $\mathbf{B}_j^{(n)}$ is the 1-banded matrix defined above.

Proof. We use induction on n. For all $n \geq 2$ let

$$\mathbf{S}^{n,q,r} = \mathbf{B}_1^{(n)}\mathbf{B}_2^{(n)} \cdots \mathbf{B}_{n-1}^{(n)}. \tag{7.87}$$

It is easily verified that

$$\mathbf{T}^{2,q,r} = \mathbf{S}^{2,q,r} = \mathbf{B}_1^{(2)} = \begin{bmatrix} 1 & r-q & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let us assume that for some $n \geq 2$, $\mathbf{T}^{n,q,r} = \mathbf{S}^{n,q,r}$. It follows from (7.84) and (7.85) that

$$\mathbf{S}^{n+1,q,r} = \begin{bmatrix} 1 & \mathbf{c}_0^T \\ & \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{c}_1^T \\ & \\ \mathbf{0} & \mathbf{B}_1^{(n)} \end{bmatrix} \cdots \begin{bmatrix} 1 & \mathbf{c}_{n-1}^T \\ & \\ \mathbf{0} & \mathbf{B}_{n-1}^{(n)} \end{bmatrix}. \tag{7.88}$$

If we carry out the multiplication of the n block matrices on the right of (7.88), then, using (7.86), we see that

$$\mathbf{S}^{n+1,q,r} = \begin{bmatrix} 1 & \mathbf{d}^T \\ & \\ \mathbf{0} & \mathbf{T}^{n,q,r} \end{bmatrix},$$

where $\mathbf{d}^T$ is a row vector. Thus it remains only to verify that the first rows of $\mathbf{T}^{n+1,q,r}$ and $\mathbf{S}^{n+1,q,r}$ are equal. Let us denote the first row of $\mathbf{S}^{n,q,r}$ by

$$\left[s_{0,0}^{n,q,r}, s_{0,1}^{n,q,r}, \ldots, s_{0,n}^{n,q,r} \right].$$

We will show that $s_{0,n}^{n,q,r} = 0$. Let us examine the product of the $n-1$ matrices on the right of (7.87). We can show by induction on j that for $1 \leq j \leq n-1$, the product $\mathbf{B}_1^{(n)}\mathbf{B}_2^{(n)} \cdots \mathbf{B}_j^{(n)}$ is j-banded, where the nonzero elements are on the main diagonal and the j diagonals above the main diagonal. (See Problem 7.5.2.) Thus $\mathbf{S}^{n,q,r}$ is $(n-1)$-banded, and so the last element in its first row, $s_{0,n}^{n,q,r}$, is zero.

Now let us write the matrix $\mathbf{B}_j^{(n+1)}$ in the block form

$$\mathbf{B}_j^{(n+1)} = \begin{bmatrix} \mathbf{A}(j, q^{n+1-j}) & \mathbf{O} \\ & \\ \mathbf{C}_j & \mathbf{D}_j \end{bmatrix},$$

where $\mathbf{A}(j, q^{n+1-j})$ is the $j \times (j+1)$ matrix defined in Lemma 7.5.1, $\mathbf{O}$ is the $j \times (n+1-j)$ zero matrix, $\mathbf{C}_j$ is $(n+2-j) \times (j+1)$, and $\mathbf{D}_j$ is $(n+2-j) \times (n+1-j)$. Thus

$$\mathbf{B}_1^{(n+1)}\mathbf{B}_2^{(n+1)} \cdots \mathbf{B}_j^{(n+1)} = \begin{bmatrix} \mathbf{A}(1, q^n) \cdots \mathbf{A}(j, q^{n+1-j}) & \mathbf{0}^T \\ \mathbf{F}_j & \mathbf{G}_j \end{bmatrix}, \quad (7.89)$$

where $\mathbf{A}(1, q^n) \cdots \mathbf{A}(j, q^{n+1-j})$ is $1 \times (j+1)$ and $\mathbf{0}^T$ is the zero vector with $n+1-j$ elements. In particular, on putting $j = n$ in (7.89), we see from (7.87) that the first row of $\mathbf{S}^{n+1,q,r}$ is

$$\left[\mathbf{A}(1, q^n)\mathbf{A}(2, q^{n-1}) \cdots \mathbf{A}(n-1, q^2)\mathbf{A}(n, q), 0\right] = \left[\mathbf{w}^T, 0\right], \quad (7.90)$$

say, where $\mathbf{w}^T$ is a row vector with $n+1$ elements. (We note in passing that this confirms our earlier observation that the last element of the first row of $\mathbf{S}^{n+1,q,r}$ is zero.) In view of Lemma 7.5.1, we may permute the quantities $q^n, q^{n-1}, \ldots, q$ in (7.90), leaving $\mathbf{w}^T$ unchanged. In particular, we may write

$$\mathbf{w}^T = \mathbf{A}(1, q^{n-1})\mathbf{A}(2, q^{n-2}) \cdots \mathbf{A}(n-1, q)\mathbf{A}(n, q^n). \quad (7.91)$$

Now, by comparison with (7.90), the product of the first $n-1$ matrices in (7.91) is the row vector containing the first n elements in the first row of $\mathbf{S}^{n,q,r}$, and thus

$$\mathbf{w}^T = [s_{0,0}^{n,q,r}, \ldots, s_{0,n-1}^{n,q,r}] \begin{bmatrix} 1 & r^n - q^n & & \\ & \ddots & \ddots & \\ & & 1 & r - q^n \end{bmatrix}$$

$$= [t_{0,0}^{n,q,r}, \ldots, t_{0,n-1}^{n,q,r}] \begin{bmatrix} 1 & r^n - q^n & & \\ & \ddots & \ddots & \\ & & 1 & r - q^n \end{bmatrix}.$$

This gives

$$s_{0,0}^{n+1,q,r} = t_{0,0}^{n,q,r}$$

and

$$s_{0,j}^{n+1,q,r} = (r^{n+1-j} - q^n)t_{0,j-1}^{n,q,r} + t_{0,j}^{n,q,r}, \quad 1 \le j \le n,$$

where we note that $t_{0,n}^{n,q,r} = 0$. It then follows from (7.81) and (7.82) that

$$s_{0,j}^{n+1,q,r} = t_{0,j}^{n+1,q,r}, \quad 0 \le j \le n,$$

and since $s_{0,n+1}^{n+1,q,r} = 0 = t_{0,n+1}^{n+1,q,r}$, (7.86) holds for $n+1$. This completes the proof. ■

If $0 < q \le r^{n-1} \le 1$, all elements in the 1-banded matrices $\mathbf{B}_j^{(n)}$ on the right of (7.86) are nonnegative. Then, from Theorem 7.4.3, we immediately have the following result concerning the total positivity of $\mathbf{T}^{n,q,r}$.

Theorem 7.5.4 If $0 < q \leq r^{n-1} \leq 1$, the transformation matrix $\mathbf{T}^{n,q,r}$ is totally positive. ■

Theorem 7.5.4 has the following important consequence for the generalized Bernstein polynomials.

Theorem 7.5.5 For $0 < q \leq 1$, the set of functions $(P^q_{n,0}, \ldots, P^q_{n,n})$, which are a basis for all generalized Bernstein polynomials of degree n, is totally positive on $[0, 1]$.

Proof. Let $\mathbf{A}^q_n$ denote the collocation matrix $(P^q_{n,j}(x_i))^n_{i,j=0}$, where we have $0 \leq x_0 < \cdots < x_n \leq 1$. Then we see from (7.74) that

$$\mathbf{A}^q_n = \mathbf{T}^{n,1,q}\mathbf{A}^1_n. \tag{7.92}$$

For every q such that $0 < q \leq 1$, $\mathbf{A}^q_n$ is the product of two totally positive matrices, and so is itself totally positive. It then follows from Definition 7.4.6 that $(P^q_{n,0}, \ldots, P^q_{n,n})$ is totally positive on $[0, 1]$. ■

Let p denote any polynomial in P_n, and let q, r denote any real numbers such that $0 < q, r \leq 1$. Since $(P^q_{n,0}, \ldots, P^q_{n,n})$ and $(P^r_{n,0}, \ldots, P^r_{n,n})$ are both bases for P_n, there exist real numbers $a^q_0, \ldots, a^q_n$ and $a^r_0, \ldots, a^r_n$ such that

$$p(x) = a^q_0 P^q_{n,0}(x) + \cdots + a^q_n P^q_{n,n}(x) = a^r_0 P^r_{n,0}(x) + \cdots + a^r_n P^r_{n,n}(x), \tag{7.93}$$

and we can deduce from (7.74) that

$$[a^q_0, a^q_1, \ldots, a^q_n]\mathbf{T}^{n,q,r} = [a^r_0, a^r_1, \ldots, a^r_n]. \tag{7.94}$$

If $0 < q \leq r^{n-1}$, the matrix $\mathbf{T}^{n,q,r}$ is totally positive and (see Problem 7.5.1) so is its transpose. In particular, the matrix $\mathbf{T}^{n,r,1}$ is totally positive for all r such that $0 < r \leq 1$. Thus we see from (7.94) and Theorem 7.4.2 that

$$S^-\left(a^1_0, \ldots, a^1_n\right) \leq S^-\left(a^r_0, \ldots, a^r_n\right) \leq S^-\left(a^q_0, \ldots, a^q_n\right),$$

where

$$p(x) = a^1_0 P^1_{n,0}(x) + \cdots + a^1_n P^1_{n,n}(x). \tag{7.95}$$

Since $(P^1_{n,0}, \ldots, P^1_{n,n})$ is totally positive, it follows from Theorem 7.4.6 that for $0 < q \leq r^{n-1} \leq 1$ and with p defined by (7.93) and (7.95),

$$S^-(p) \leq S^-\left(a^1_0, \ldots, a^1_n\right) \leq S^-\left(a^r_0, \ldots, a^r_n\right) \leq S^-\left(a^q_0, \ldots, a^q_n\right). \tag{7.96}$$

We can now state a generalization of Theorem 7.5.1.

Theorem 7.5.6 Let $B^q_n(f; x)$ denote the generalized Bernstein polynomial that we denoted by $B_n(f; x)$ in (7.51). Then

$$S^-(B^q_n f) \leq S^-(f) \tag{7.97}$$

on $[0, 1]$, and thus the operator B^q_n is variation-diminishing.

Proof. Let us choose

$$p(x) = B_n^q(f;x) = a_0^q P_{n,0}^q(x) + \cdots + a_n^q P_{n,n}^q(x)$$

in (7.97). We have already noted that the q-binomial coefficient $\begin{bmatrix} n \\ r \end{bmatrix}$ is a polynomial in q with positive integer coefficients, and so is positive if $q > 0$. Thus, for $q > 0$,

$$S^-(B_n^q f) \leq S^-(f_0, f_1, \ldots, f_n) \leq S^-(f),$$

where $f_r = f([r]/[n])$. ■

Let p denote any linear polynomial; that is, $p \in P_1$. Then, since B_n^q reproduces linear polynomials, we may deduce the following result from Theorem 7.5.6.

Theorem 7.5.7 For any function f and any linear polynomial p, we have

$$S^-(B_n^q f - p) \leq S^-(B_n^q(f-p)) \leq S^-(f-p), \tag{7.98}$$

for $0 < q \leq 1$. ■

The next two theorems readily follow from Theorem 7.5.7.

Theorem 7.5.8 Let f be monotonically increasing (decreasing) on $[0,1]$. Then the generalized Bernstein polynomial $B_n^q f$ is also monotonically increasing (decreasing) on $[0,1]$, for $0 < q \leq 1$.

Proof. We have already proved this in Theorem 7.1.2 when $q = 1$. Let us replace p in (7.98) by the constant c. Then, if f is monotonically increasing on $[0,1]$,

$$S^-(B_n^q f - c) \leq S^-(f - c) \leq 1$$

for all choices of constant c, and thus $B_n^q f$ is monotonically increasing or decreasing. Since

$$B_n^q(f;0) = f(0) \leq f(1) = B_n^q(f;1),$$

$B_n^q f$ must be monotonically *increasing*. On the other hand, if f is monotonically *decreasing*, we may replace f by $-f$, and repeat the above argument, concluding that $B_n^q f$ is monotonically decreasing. ■

Theorem 7.5.9 If f is convex on $[0,1]$, then $B_n^q f$ is also convex on $[0,1]$, for $0 < q \leq 1$.

Proof. Let p denote any linear polynomial. Then if f is convex, the graph of p can intersect that of f at no more than two points, and thus $S^-(f-p) \leq 2$. It follows from (7.98) that for any q such that $0 < q \leq 1$,

$$S^-(B_n^q f - p) = S^-(B_n^q(f-p)) \leq S^-(f-p) \leq 2. \tag{7.99}$$

Suppose the graph of p intersects that of $B_n^q f$ at a and b. Then we have $p(a) = B_n^q(f;a)$ and $p(b) = B_n^q(f;b)$, where $0 < a < b < 1$, and we see from (7.99) that $B_n^q f - p$ cannot change sign in (a,b). As we vary a and b, a continuity argument shows that the sign of $B_n^q f - p$ on (a,b) is the same for all a and b, $0 < a < b < 1$. From the convexity of f we see that in the limiting case where $a = 0$ and $b = 1$, $0 \leq p(x) - f(x)$ on $[0,1]$, so that

$$0 \leq B_n^q(p-f;x) = p(x) - B_n^q(f;x), \quad 0 \leq x \leq 1,$$

and thus B_n^q is convex. ■

We conclude this section by proving that if f is convex, the generalized Bernstein polynomials $B_n^q f$, for n fixed, are monotonic in q.

Theorem 7.5.10 For $0 < q \leq r \leq 1$ and for f convex on $[0,1]$, we have

$$f(x) \leq B_n^r(f;x) \leq B_n^q(f;x), \quad 0 \leq x \leq 1. \tag{7.100}$$

Proof. It remains only to establish the second inequality in (7.100), since the first inequality has already been proved in Theorem 7.3.3. Let us write

$$\zeta_{n,j}^q = \frac{[j]}{[n]} \quad \text{and} \quad a_{n,j}^q = \left[\begin{array}{c} n \\ j \end{array} \right].$$

Then, for any function g on [0,1],

$$B_n^q(g;x) = \sum_{j=0}^{n} g(\zeta_{n,j}^q) a_{n,j}^q P_{n,j}^q(x) = \sum_{j=0}^{n} \sum_{k=0}^{n} g(\zeta_{n,j}^q) a_{n,j}^q t_{j,k}^{n,q,r} P_{n,k}^r(x),$$

and thus

$$B_n^q(g;x) = \sum_{k=0}^{n} P_{n,k}^r(x) \sum_{j=0}^{n} t_{j,k}^{n,q,r} g(\zeta_{n,j}^q) a_{n,j}^q. \tag{7.101}$$

With $g(x) = 1$, this gives

$$1 = \sum_{j=0}^{n} a_{n,j}^q P_{n,j}^q(x) = \sum_{k=0}^{n} P_{n,k}^r(x) \sum_{j=0}^{n} t_{j,k}^{n,q,r} a_{n,j}^q$$

and hence

$$\sum_{j=0}^{n} t_{j,k}^{n,q,r} a_{n,j}^q = a_{n,k}^r, \quad 0 \leq k \leq n. \tag{7.102}$$

On putting $g(x) = x$ in (7.101), we obtain

$$x = \sum_{j=0}^{n} \zeta_{n,j}^q a_{n,j}^q P_{n,j}^q(x) = \sum_{k=0}^{n} P_{n,k}^r(x) \sum_{j=0}^{n} t_{j,k}^{n,q,r} \zeta_{n,j}^q a_{n,j}^q.$$

Since

$$\sum_{j=0}^{n} \zeta_{n,j}^{r} a_{n,j}^{r} P_{n,j}^{r}(x) = x,$$

we have

$$\sum_{j=0}^{n} t_{j,k}^{n,q,r} \zeta_{n,j}^{q} a_{n,j}^{q} = \zeta_{n,k}^{r} a_{n,k}^{r}, \quad 0 \leq k \leq n. \tag{7.103}$$

Let us write

$$\lambda_j = \frac{t_{j,k}^{n,q,r} a_{n,j}^{q}}{a_{n,k}^{r}},$$

and we see from (7.102) and (7.103), respectively, that

$$\sum_{j=0}^{n} \lambda_j = 1 \quad \text{and} \quad \zeta_{n,k}^{r} = \sum_{j=0}^{n} \lambda_j \zeta_{n,j}^{q}.$$

It then follows from Problem 7.1.3 that if f is convex,

$$f(\zeta_{n,k}^{r}) = f\left(\sum_{j=0}^{n} \lambda_j \zeta_{n,j}^{q} \right) \leq \sum_{j=0}^{n} \lambda_j f(\zeta_{n,j}^{q}),$$

which gives

$$f(\zeta_{n,k}^{r}) \leq \sum_{j=0}^{n} (a_{n,k}^{r})^{-1} t_{j,k}^{n,q,r} a_{n,j}^{q} f(\zeta_{n,j}^{q}). \tag{7.104}$$

On substituting

$$P_{n,j}^{q}(x) = \sum_{k=0}^{n} t_{j,k}^{n,q,r} P_{n,k}^{r}(x),$$

obtained from (7.74), into

$$B_n^q(f;x) = \sum_{j=0}^{n} f(\zeta_{n,j}^{q}) a_{n,j}^{q} P_{n,j}^{q}(x),$$

we find that

$$B_n^q(f;x) = \sum_{k=0}^{n} a_{n,k}^{r} P_{n,k}^{r}(x) \sum_{j=0}^{n} (a_{n,k}^{r})^{-1} t_{j,k}^{n,q,r} f(\zeta_{n,j}^{q}) a_{n,j}^{q}.$$

It then follows from (7.104) that

$$B_n^q(f;x) \geq \sum_{k=0}^{n} a_{n,k}^{r} P_{n,k}^{r}(x) f(\zeta_{n,k}^{r}) = B_n^r(f;x),$$

and this completes the proof. ■

Problem 7.5.1 Given that

$$\det \mathbf{A} = \det \mathbf{A}^T,$$

for any square matrix $\mathbf{A}$, deduce from Definition 7.4.1 that if the matrix $\mathbf{A}$ is totally positive, so also is $\mathbf{A}^T$.

Problem 7.5.2 Let $\mathbf{A}_1, \mathbf{A}_2, \ldots$ denote $m \times m$ matrices that are 1-banded, and whose nonzero elements are on the main diagonal and the diagonal above the main diagonal. Show by induction on j that for $1 \leq j \leq m-1$, the product $\mathbf{A}_1\mathbf{A}_1 \cdots \mathbf{A}_j$ is a j-banded matrix.