

## ON THE OPTIMAL STABILITY OF THE BERNSTEIN BASIS

R. T. FAROUKI AND T. N. T. GOODMAN

ABSTRACT. We show that the Bernstein polynomial basis on a given interval is “optimally stable,” in the sense that no other nonnegative basis yields systematically smaller condition numbers for the values or roots of arbitrary polynomials on that interval. This result follows from a partial ordering of the set of all nonnegative bases that is induced by nonnegative basis transformations. We further show, by means of some low-degree examples, that the Bernstein form is not uniquely optimal in this respect. However, it is the only optimally stable basis whose elements have no roots on the interior of the chosen interval. These ideas are illustrated by comparing the stability properties of the power, Bernstein, and generalized Ball bases.

### 1. INTRODUCTION

To represent a polynomial  $p$  in a digital computer, we store in memory its coefficients  $c_0, \dots, c_n$  in a suitable basis. These coefficients, together with a value  $t$  of the independent variable, serve as input to an evaluation algorithm that furnishes the polynomial value  $p(t)$  as output.

Since most scientific/engineering calculations are performed in floating-point arithmetic, limitations that are imposed by the choice of basis on the expected accuracy of the value  $p(t)$  are clearly of fundamental concern. In theoretical expositions, for example, one usually relies upon the monomial or “power” form  $c_0 + c_1t + \dots + c_nt^n$ , which may be evaluated by Horner’s method. It has been shown, however, that if one is interested only in polynomial values (or roots) on  $t \in [0, 1]$ , the Bernstein form on this interval is systematically more stable than the power form [5], in the following sense:

We imagine the coefficients of both the power and Bernstein forms of  $p$  to be subject to random errors of maximum relative magnitude  $\epsilon$ , and we ask: how do the bounds on the corresponding perturbations  $\delta p(t)$  in the value  $p(t)$  compare? These bounds may be expressed in terms of “condition numbers” (see §2 below) and one finds they are *systematically* smaller — i.e., for every polynomial  $p$  and for each value  $t \in (0, 1)$  — in the Bernstein form.

The enhanced stability of the Bernstein form, as compared to the power form, can be attributed to two simple facts: (i) the power and Bernstein bases are both nonnegative on  $[0, 1]$ ; and (ii) the latter may be transformed into the former by means of a nonnegative matrix. Our purpose here is to explore the implications of these facts in a broader and more fundamental context. Thus, we consider families of nonnegative bases generated by nonnegative transformations and identify their

---

Received by the editor March 2, 1995 and, in revised form, August 28, 1995.  
1991 *Mathematics Subject Classification*. Primary 65G99; Secondary 65D17.

“minimal” members (those not obtainable as nonnegative transformations of any other nonnegative basis).

The Bernstein basis is one example of such “optimally stable” bases, in the sense that there exists no other nonnegative basis that gives systematically smaller condition numbers than it. Although it is not *uniquely* optimal, no other basis in common use enjoys this distinction, and it is uncertain whether other conceivable optimally-stable bases would share the useful properties and algorithms that we associate with the Bernstein form.

We proceed in the remainder of this paper as follows. After reviewing basic ideas concerning polynomial condition numbers in §2, we introduce in §3 a partial ordering of the set of nonnegative bases on a given interval, and identify its minimal elements as optimally stable bases. We show in §4 that the Bernstein basis is one such basis, and we derive results that distinguish it from other possible optimal bases. In §5 and §6 we illustrate the practical consequences of these results in the context of three representations (the power, Bernstein, and generalized Ball forms) commonly used in computer-aided geometric design. Finally, §7 offers some concluding remarks.

## 2. CONDITION NUMBERS

Any set of  $n+1$  linearly independent polynomials,  $\Phi = \{\phi_0(t), \dots, \phi_n(t)\}$ , defines a *basis* for polynomials of degree  $n$  — i.e., we can uniquely express any polynomial  $p(t)$  of degree  $\leq n$  in the form

$$(1) \quad p(t) = \sum_{k=0}^n c_k \phi_k(t)$$

by a suitable choice of coefficients  $c_0, \dots, c_n$ . We shall be concerned here with the *stability* of such representations, i.e., in how sensitive a value or root of  $p$  is to random perturbations of a given maximum relative magnitude  $\epsilon$  in the coefficients  $c_0, \dots, c_n$  corresponding to the basis  $\Phi$ .

A sharp bound on the perturbation  $\delta p(t)$  in (1) may be expressed in terms of a condition number  $C_\Phi(p(t))$  for the *value* of  $p$  in the basis  $\Phi$  as

$$(2) \quad |\delta p(t)| \leq C_\Phi(p(t)) \epsilon \quad \text{where} \quad C_\Phi(p(t)) = \sum_{k=0}^n |c_k \phi_k(t)|.$$

The above bound holds for *arbitrary* — not just infinitesimal — coefficient perturbations  $\epsilon$ . Note that  $C_\Phi(p(t))$  depends as much on the *adopted basis*  $\Phi$  as on the particular polynomial  $p$  under consideration.

Suppose now that  $\tau$  is a simple real root of (1), i.e.,  $p(\tau) = 0 \neq p'(\tau)$ . We can also characterize the sensitivity of  $\tau$  to a perturbation  $\epsilon$  of the coefficients  $c_0, \dots, c_n$  in the basis  $\Phi$  by a condition number. Namely, the displacement  $\delta\tau$  of this root satisfies

$$(3) \quad |\delta\tau| \leq C_\Phi(\tau) \epsilon \quad \text{where} \quad C_\Phi(\tau) = \frac{1}{|p'(\tau)|} \sum_{k=0}^n |c_k \phi_k(\tau)|.$$

The quantity  $C_\Phi(\tau)$  is called the condition number for the *root*  $\tau$  in the basis  $\Phi$  (note that the above bound is strictly valid only in the limit  $\epsilon \rightarrow 0$ ). Further details on these condition number formulations may be found in [5]; for a comprehensive discussion, see [6].

To obtain accurate results when performing floating-point computations with polynomials, it is desirable to employ a representation — i.e., a choice of basis — in which the condition numbers are as small as possible.<sup>1</sup> When comparing condition numbers in *arbitrary* bases  $\Phi = \{\phi_0, \dots, \phi_n\}$  and  $\Psi = \{\psi_0, \dots, \psi_n\}$  one will find, in general, that some polynomials  $p$  and values of  $t$  yield smaller condition numbers in the  $\Phi$  basis, while others are better-conditioned in the  $\Psi$  representation. In other words, no *systematic* inequality between  $C_\Phi(p(t))$  and  $C_\Psi(p(t))$  — valid for all  $p$  and every  $t$  in some domain of interest — can be expected when  $\Phi$  and  $\Psi$  are arbitrarily chosen.

This problem may be surmounted by imposing suitable restrictions on the bases we consider. Indeed, such conditions may be independently desirable, on account of useful attributes or algorithms they incur. Specifically, we shall be concerned here only with bases  $\Phi = \{\phi_0, \dots, \phi_n\}$  that are *nonnegative* over an interval  $t \in [a, b]$  of interest, i.e., for  $k = 0, \dots, n$  they satisfy

$$\phi_k(t) \geq 0 \quad \text{for } t \in [a, b].$$

If, in addition, the basis  $\Phi$  forms a *partition of unity* ( $\phi_0(t) + \dots + \phi_n(t) \equiv 1$ ) the representation (1) will exhibit the *convex hull property*:

$$\min_k(c_k) \leq p(t) \leq \max_k(c_k) \quad \text{for } t \in [a, b].$$

In terms of condition numbers, nonnegative bases are of particular interest in the context of the following result:

**Proposition 1.** *Let  $\Phi = \{\phi_0(t), \dots, \phi_n(t)\}$  and  $\Psi = \{\psi_0(t), \dots, \psi_n(t)\}$  be nonnegative bases for degree- $n$  polynomials on  $t \in [a, b]$  such that the latter can be expressed as a nonnegative combination of the former, i.e.,*

$$(4) \quad \psi_j(t) = \sum_{k=0}^n M_{jk} \phi_k(t), \quad j = 0, \dots, n,$$

$$(5) \quad \text{where } M_{jk} \geq 0 \quad \text{for all } 0 \leq j, k \leq n.$$

*Then the condition number for the value of **any** degree- $n$  polynomial  $p(t)$  at **any** point  $t \in [a, b]$  in these bases satisfies the inequality*

$$(6) \quad C_\Phi(p(t)) \leq C_\Psi(p(t)).$$

*Proof.* The result (6) is an immediate consequence of the triangle inequality and the nonnegativity of the matrix elements  $M_{jk}$  and the bases  $\Phi$  and  $\Psi$ . It is a straightforward generalization of Theorem 1 in [5], wherein it was couched specifically in terms of the Bernstein basis  $b_k^n(t) = \binom{n}{k} (1-t)^{n-k} t^k$  for  $k = 0, \dots, n$  on  $[0, 1]$  and the power basis  $\{1, t, \dots, t^n\}$  — note that the latter is given by the nonnegative combination

$$(7) \quad t^j = \sum_{k=j}^n \binom{k}{j} b_k^n(t) \quad \text{for } j = 0, \dots, n$$

of the former. □

<sup>1</sup>It can be shown by *backward error analysis* [12] that the cumulative effect of floating-point arithmetic errors during any computation on given polynomials is equivalent to certain perturbations of their exact coefficients.

The inequality (6) clearly also holds for *root* condition numbers, since these differ from condition numbers for the value only by the magnitude of the derivative at the root (which is independent of the choice of basis).

### 3. PARTIAL ORDERING OF NONNEGATIVE BASES

Let  $\Pi_n$  be the space of all polynomials of degree  $\leq n$  on the interval  $[a, b]$ . We shall denote by  $B_n$  the set of *nonnegative bases* for  $\Pi_n$ .

If  $\Phi = \{\phi_0, \dots, \phi_n\}$  and  $\Psi = \{\psi_0, \dots, \psi_n\}$  are members of  $B_n$ , we write  $\Phi \succsim \Psi$  if there exists a nonnegative  $(n+1) \times (n+1)$  matrix  $\mathbf{M}$  such that

$$(8) \quad \Psi^T = \mathbf{M} \Phi^T.$$

Since the product of two nonnegative matrices is also a nonnegative matrix, the relation  $\succsim$  evidently satisfies the transitivity condition

$$\Psi \succsim \Phi \text{ and } \Phi \succsim \Theta \implies \Psi \succsim \Theta.$$

It can easily be verified that a nonnegative matrix has a nonnegative inverse if and only if it is the product of a permutation matrix and a positive diagonal matrix. Thus, the relations  $\Phi \succsim \Psi$  and  $\Psi \succsim \Phi$  are *simultaneously* satisfied if and only if, under suitable ordering, the elements of  $\Phi$  are constant positive multiples of the elements of  $\Psi$ . In such a case, we write  $\Phi \sim \Psi$ .

We shall also write  $\Phi \prec \Psi$  if  $\Phi \succsim \Psi$  but  $\Phi \not\sim \Psi$ . These relations may be phrased as follows. When  $\Phi \prec \Psi$ , we say “ $\Phi$  precedes  $\Psi$ .” If  $\Phi \sim \Psi$ , we say “ $\Phi$  is similar to  $\Psi$ .” Finally,  $\Phi \succsim \Psi$  is read “ $\Phi$  precedes or is similar to  $\Psi$ .”

The relation  $\succsim$  induces a *partial ordering* among the members of  $B_n$  — we say the ordering is “partial” since there exist pairs of bases  $\Phi$  and  $\Psi$  in  $B_n$  such that neither the matrix  $\mathbf{M}$  defined by (8) nor its inverse is nonnegative; no precedence relation can be established between such bases.

We shall call any nonnegative basis  $\Phi$  a *minimal basis* if there is no basis  $\Psi$  in  $B_n$  that satisfies  $\Psi \prec \Phi$ . Note that, since  $\succsim$  only partially orders  $B_n$ , there may be — modulo similarities — more than one minimal basis. The following theorem demonstrates that minimal bases are “optimally stable” (in the sense of the condition numbers defined in §2):

**Theorem 1.** *Any two bases  $\Phi$  and  $\Psi$  in  $B_n$  satisfy*

$$(9) \quad \Phi \succsim \Psi \iff C_\Phi(p(t)) \leq C_\Psi(p(t)),$$

where the inequality on the right holds for each polynomial  $p \in \Pi_n$  evaluated at every value  $t \in [a, b]$ .

*Proof.* That  $\Phi \succsim \Psi \implies C_\Phi(p(t)) \leq C_\Psi(p(t))$  for each  $p \in \Pi_n$  and every  $t \in [a, b]$  follows directly from Proposition 1 and the definition of  $\succsim$ .

To verify the converse statement, we shall argue by contradiction. Suppose that  $C_\Phi(p(t)) \leq C_\Psi(p(t))$  for each  $p \in \Pi_n$  and any  $t \in [a, b]$ , but the relation  $\Phi \succsim \Psi$  does *not* hold. Then  $\Psi^T = \mathbf{M} \Phi^T$ , where the matrix  $\mathbf{M}$  is not nonnegative, i.e., among its elements  $M_{ij}$  for  $0 \leq i, j \leq n$  we have  $M_{kl} < 0$  for some  $k, l$ . Furthermore, since  $\Phi = \{\phi_0, \dots, \phi_n\}$  and  $\Psi = \{\psi_0, \dots, \psi_n\}$  are nonnegative bases, we clearly must have  $M_{kr} > 0$  for some  $r$ .

Now taking  $p = \psi_k$  and choosing a value  $t \in [a, b]$  such that both  $\phi_l(t) > 0$  and  $\phi_r(t) > 0$ , we observe that

$$C_\Phi(p(t)) = \sum_{j=0}^n |M_{kj}| \phi_j(t) > \left| \sum_{j=0}^n M_{kj} \phi_j(t) \right| = |\psi_k(t)| = C_\Psi(p(t)),$$

where we make use of the triangle inequality and the fact that  $M_{kl}$  and  $M_{kr}$  are of opposite sign. Since the above contradicts our initial supposition, we must conclude that it is false.  $\square$

Note that, in Theorem 1,  $\Phi \sim \Psi \Rightarrow C_\Phi(p(t)) \equiv C_\Psi(p(t))$  in equation (9) for each  $p \in \Pi_n$  and every  $t \in [a, b]$ .

#### 4. OPTIMAL STABILITY OF THE BERNSTEIN BASIS

Consider now the Bernstein basis  $\mathcal{B}$  defined on  $t \in [a, b]$  by

$$(10) \quad b_j^n(t) = \binom{n}{j} \frac{(b-t)^{n-j}(t-a)^j}{(b-a)^n} \quad \text{for } j = 0, \dots, n.$$

We will demonstrate that  $\mathcal{B}$  has optimal stability in  $B_n$ .

**Theorem 2.** *Suppose that  $\Psi = \{\psi_0, \dots, \psi_n\}$  and  $\tilde{\Psi} = \{\tilde{\psi}_0, \dots, \tilde{\psi}_n\}$  are bases in  $B_n$  satisfying*

$$\begin{aligned} \psi_j^{(i)}(a) &= 0 \quad \text{for } i = 0, \dots, j-1 \text{ and } j = 1, \dots, n, \\ \tilde{\psi}_j^{(i)}(b) &= 0 \quad \text{for } i = 0, \dots, n-j-1 \text{ and } j = 0, \dots, n-1. \end{aligned}$$

*Then, if  $\Phi \in B_n$  satisfies both  $\Phi \preceq \Psi$  and  $\Phi \preceq \tilde{\Psi}$ , we have  $\Phi \sim \mathcal{B}$ .*

*Proof.* Let  $M_{ij}$  for  $0 \leq i, j \leq n$  be the elements of the nonnegative matrix  $\mathbf{M}$  such that

$$(11) \quad \tilde{\Psi}^T = \mathbf{M} \Phi^T.$$

We will prove by induction that the basis  $\Phi$  may be ordered such that the matrix  $\mathbf{M}$  is lower triangular. Suppose that, for a suitable ordering of  $\Phi$ , there is a row  $k$  such that

$$M_{ij} = 0 \quad \text{for } j \geq i + 1 \text{ and } i = 0, \dots, k - 1.$$

This hypothesis is vacuous if  $k = 0$ . Note that it is unaffected by reordering  $\phi_k, \dots, \phi_n$ . Note also that, for  $\tilde{\psi}_0, \dots, \tilde{\psi}_{k-1}$  to be linearly independent under this hypothesis, we must have  $M_{ii} > 0$  for  $i = 0, \dots, k - 1$ . Now for  $k \leq n - 1$ ,

$$0 = \tilde{\psi}_k(b) = \sum_{j=0}^n M_{kj} \phi_j(b),$$

and since  $M_{kj}$  and  $\phi_j(b)$  are nonnegative, we must have  $M_{kj} \phi_j(b) = 0$  for  $j = 0, \dots, n$ . If we were to suppose that  $M_{kj} > 0$  for  $j = k, \dots, n$ , then  $\phi_j(b) = 0$  for  $j = k, \dots, n$ . However, since  $\tilde{\psi}_j(b) = 0$  for  $j = 0, \dots, k - 1$ , we may infer from our hypothesis that  $\phi_j(b) = 0$  for  $j = 0, \dots, k - 1$  also. But the values  $\phi_0(b), \dots, \phi_n(b)$  cannot all vanish if  $\Phi$  is a basis for  $\Pi_n$ . Our supposition that  $M_{jk} > 0$  for  $j = k, \dots, n$  must therefore be false. Thus, by reordering  $\phi_k, \dots, \phi_n$  if necessary, we may assume that  $M_{kn} = 0$ .

If  $k = n - 1$ , the inductive proof is complete. Otherwise, for  $k \leq n - 2$ , we note that

$$(12) \quad 0 = \tilde{\psi}'_k(b) = \sum_{j=0}^{n-1} M_{kj} \phi'_j(b),$$

and, again,  $\sum_{j=0}^{n-1} M_{kj} \phi_j(b) = 0$ . Now supposing that  $M_{kj} > 0$  for  $j = k, \dots, n - 1$  we have, as before,  $\phi_j(b) = 0$  for  $j = k, \dots, n - 1$ . Since the basis  $\Phi$  is nonnegative on  $[a, b]$  this implies that  $\phi'_j(b) \leq 0$  for  $j = k, \dots, n - 1$ . Now  $\tilde{\psi}_j(b) = \tilde{\psi}'_j(b) = 0$  for  $j = 0, \dots, k - 1$ , and by our hypothesis we also have  $\phi_j(b) = \phi'_j(b) = 0$  for  $j = 0, \dots, k - 1$ . To satisfy equation (12) under these conditions, we must have  $\phi'_j(b) = 0$  for  $j = k, \dots, n - 1$ . Hence,  $\phi_j(b) = \phi'_j(b) = 0$  for  $j = 0, \dots, n - 1$ , i.e., the  $n$  polynomials  $\phi_0, \dots, \phi_{n-1}$  lie in the  $(n - 1)$ -dimensional space  $\{p \in \Pi_n : p(b) = p'(b) = 0\}$ , which contradicts their linear independence. The supposition that  $M_{kj} > 0$  for  $j = k, \dots, n - 1$  must therefore be false, and on reordering  $\phi_k, \dots, \phi_{n-1}$  if necessary we may assume that  $M_{k,n-1} = 0$ .

Continuing in this manner, one can show that  $M_{kj} = 0$  for  $j = k + 1, \dots, n$ , and the inductive proof is complete. So we can henceforth assume that (11) holds, where  $\mathbf{M}$  is a nonnegative, lower triangular matrix.

We shall now prove by induction that

$$(13) \quad \phi_j(t) = c_j(b - t)^{n-j}(t - a)^j$$

for nonzero constants  $c_0, \dots, c_n$ . As our inductive hypothesis we suppose that, for some  $k$  between 0 and  $n$ , expression (13) holds for  $j = n - k + 1, \dots, n$ . This hypothesis is vacuous for  $k = 0$ .

Let  $\Psi^T = \mathbf{N} \Phi^T$  for some nonnegative matrix  $\mathbf{N}$  having elements  $N_{ij}$ ,  $0 \leq i, j \leq n$ . First, suppose that  $k \leq n - 1$ . Then

$$0 = \psi_{n-k}(a) = \sum_{j=0}^n N_{n-k,j} \phi_j(a),$$

and hence  $N_{n-k,j} \phi_j(a) = 0$  for  $j = 0, \dots, n$ . If we suppose that  $N_{n-k,j} > 0$  for  $j = 0, \dots, n - k$ , then  $\phi_j(a) = 0$  for  $j = 0, \dots, n - k$ . But, by our hypothesis,  $\phi_j$  has the form (13) for  $j = n - k + 1, \dots, n$ , and so  $\phi_j(a) = 0$  for  $j = n - k + 1, \dots, n$  also. However, the values  $\phi_0(a), \dots, \phi_n(a)$  cannot all vanish if  $\Phi$  is a basis for  $\Pi_n$ , so we conclude that the supposition  $N_{n-k,j} > 0$  for  $j = 0, \dots, n - k$  must be false. Hence, we must have  $N_{n-k,l} = 0$  for some  $l$ , where  $0 \leq l \leq n - k$ .

Next suppose that  $k \leq n - 2$ . Then we have

$$(14) \quad 0 = \psi'_{n-k}(a) = \sum_{j=0}^n N_{n-k,j} \phi'_j(a),$$

and similarly  $\sum_{j=0}^n N_{n-k,j} \phi_j(a) = 0$ . We suppose that  $N_{n-k,j} > 0$  for  $j = 0, \dots, n - k$  ( $j \neq l$ ). Then, as before,  $\phi_j(a) = 0$  for  $j = 0, \dots, n - k$  ( $j \neq l$ ). Since the basis  $\Phi$  is nonnegative on  $[a, b]$  this implies that  $\phi'_j(a) \geq 0$  for  $j = 0, \dots, n - k$  ( $j \neq l$ ). Now by our hypothesis,  $\phi_j$  has the form (13) for  $j = n - k + 1, \dots, n$ , and so  $\phi_j(a) = \phi'_j(a) = 0$  for  $j = 0, \dots, n - k$  ( $j \neq l$ ). In order to satisfy equation (14) under these conditions, we must have  $\phi'_j(a) = 0$  for  $j = 0, \dots, n - k$  ( $j \neq l$ ). This implies that the set  $\{\phi_j : 0 \leq j \leq n, j \neq l\}$  of  $n$  polynomials lies in the  $(n - 1)$ -dimensional space  $\{p \in \Pi_n : p(a) = p'(a) = 0\}$ , which contradicts their linear independence. Thus, the supposition that  $N_{n-k,j} > 0$  for  $j = 0, \dots, n - k$

( $j \neq l$ ) is false, and we infer that  $N_{n-k,m} = 0$  for some  $m$ , where  $0 \leq m \leq n - k$  and  $m \neq l$ .

Continuing in this manner, one can show that only one of the elements  $N_{n-k,j}$  for  $j = 0, \dots, n - k$  is nonzero. Thus, for some  $r$  between 0 and  $n - k$ , we make use of (13) to deduce that

$$(15) \quad \psi_{n-k} = c_r \phi_r + \sum_{j=n-k+1}^n c_j (b-t)^{n-j} (t-a)^j$$

for some nonnegative constants  $c_r$  and  $c_{n-k+1}, \dots, c_n$ . Now since  $\tilde{\psi}_r^{(i)}(b) = 0$  for  $i = 0, \dots, n - r - 1$ , the relation (11) — where  $\mathbf{M}$  is lower triangular — shows that

$$\phi_r^{(i)}(b) = 0 \quad \text{for } i = 0, \dots, n - r - 1.$$

On the other hand, equation (15) and the requirement that  $\psi_{n-k}^{(i)}(a) = 0$  for  $i = 0, \dots, n - k - 1$  indicate that

$$\phi_r^{(i)}(a) = 0 \quad \text{for } i = 0, \dots, n - k - 1.$$

From the above conditions we see that, for  $\phi_r$  to not vanish identically, we must have  $r = n - k$ . Hence we deduce that  $\phi_{n-k}(t)$  is a constant positive multiple of  $(b - t)^k (t - a)^{n-k}$ .

Hence (13) is established for  $j = n - k$ , and the inductive step is complete. Thus, the form (13) holds for  $j = 0, \dots, n$  and we have  $\Phi \sim \mathcal{B}$ . □

**Corollary 1.** *The basis  $\mathcal{B}$  is minimal.*

*Proof.* Suppose  $\Phi \in B_n$  is such that  $\Phi \lesssim \mathcal{B}$ . Since the basis  $\mathcal{B}$  satisfies the conditions on both  $\Psi$  and  $\tilde{\Psi}$  stipulated in Theorem 2, this theorem implies that  $\Phi \sim \mathcal{B}$ . Thus, there is no basis  $\Phi$  in  $B_n$  such that  $\Phi \prec \mathcal{B}$ . □

Corollary 1 establishes the optimal stability of the Bernstein basis. We note also the following corollaries to Theorem 2:

**Corollary 2.** *If  $\Phi \in B_n$  satisfies both  $\Phi \lesssim \{1, t - a, \dots, (t - a)^n\}$  and  $\Phi \lesssim \{1, b - t, \dots, (b - t)^n\}$ , then  $\Phi \sim \mathcal{B}$ .*

**Corollary 3.** *Suppose that  $\Phi \in B_n$  satisfies  $\Phi \lesssim \{1, t - a, \dots, (t - a)^n\}$  and  $\Phi$  is symmetric, i.e.,  $\Phi(t) \sim \Phi(a + b - t)$ . Then  $\Phi \sim \mathcal{B}$ .*

*Remark 1.* The Bernstein basis is *not* the only minimal basis. Consider, for example,

$$\Psi = \{(1 - t)^2, (2t - 1)^2, t^2\}$$

as a basis for  $\Pi_2$  on  $t \in [0, 1]$ . Suppose that  $\Phi = \{\phi_0, \phi_1, \phi_2\} \in B_2$  satisfies  $\Phi \lesssim \Psi$ . Then  $(1 - t)^2 = a\phi_0 + b\phi_1 + c\phi_2$  with  $a, b, c \geq 0$ . Since

$$a\phi_0^{(i)}(1) + b\phi_1^{(i)}(1) + c\phi_2^{(i)}(1) = 0 \quad \text{for } i = 0, 1$$

the argument used in the proof of Theorem 2 shows that only one of  $a, b, c$  can be nonzero. So we may assume  $\phi_0(t) = (1 - t)^2$ , and similar arguments lead us to conclude that  $\phi_2(t) = t^2$ . Hence  $(2t - 1)^2 = a(1 - t)^2 + b\phi_1(t) + ct^2$  with  $a, b, c \geq 0$ . Setting  $t = \frac{1}{2}$  gives  $a = c = 0$ , and hence we have  $\Phi \sim \Psi$ .

The optimal stability of the Bernstein basis is closely related to another optimal attribute, recently identified by Carnicer and Peña [2] — namely, its optimal *shape-preserving properties*. They show that any normalized totally positive basis may be obtained from the Bernstein basis by multiplication with a totally positive transformation matrix (a *totally positive* matrix is one whose minors are all nonnegative; a basis  $\phi_0, \dots, \phi_n$  is totally positive on the interval  $t \in [a, b]$  if the collocation matrix with elements  $M_{jk} = \phi_j(t_k)$ ,  $0 \leq j, k \leq n$ , is totally positive for any sequence  $t_0, \dots, t_n \in [a, b]$ ).

Carnicer and Peña have subsequently considered nonnegative bases, and from Theorem 3.9 in their recent paper [3] one can independently infer that the Bernstein basis is minimal with respect to the partial order  $\preceq$  among the set of all nonnegative bases, and is therefore optimally stable.

We now give some results that serve to differentiate the Bernstein basis from other possible optimally-stable bases. If  $\phi(t)$  and  $\psi(t)$  are polynomials on  $t \in [a, b]$  we shall write  $Z(\phi) \subseteq Z(\psi)$  if the zeros of  $\psi$  (with multiplicity) include those of  $\phi$ , i.e., if  $\phi^{(j)}(\tau) = 0$  for  $j = 0, \dots, k$  for any  $\tau \in [a, b]$  and  $k \geq 0$ , then  $\psi^{(j)}(\tau) = 0$  for  $j = 0, \dots, k$ .

**Proposition 2.** *Suppose  $\Phi \in B_n$  includes two elements  $\phi_i$  and  $\phi_j$  ( $i \neq j$ ) such that  $Z(\phi_i) \subseteq Z(\phi_j)$ . Then  $\Phi$  is not minimal.*

*Proof.* Let  $\tau_1, \dots, \tau_r$  denote all the zeros of  $\phi_i$ , with multiplicity, on  $[a, b]$ . Then  $\phi_i(t) = |(t - \tau_1) \cdots (t - \tau_r)| p_i(t)$ , where  $p_i(t) > 0$  for  $t \in [a, b]$ . So  $\phi_j(t) = |(t - \tau_1) \cdots (t - \tau_r)| p_j(t)$ , where  $p_j(t) \geq 0$  for  $t \in [a, b]$ . Choose  $c > 0$  so that  $c p_j(t) < p_i(t)$  for  $t \in [a, b]$ . Then  $\phi_i(t) - c \phi_j(t) \geq 0$  for  $t \in [a, b]$ . Now define  $\Psi = \{\psi_0, \dots, \psi_n\}$  by  $\psi_k = \phi_k$  if  $k \neq i$ , and  $\psi_i = \phi_i - c \phi_j$ . Then  $\phi_k = \psi_k$  for  $k \neq i$ , and  $\phi_i = \psi_i + c \psi_j$ . Hence  $\Psi \prec \Phi$ , so  $\Phi$  is not minimal.  $\square$

**Theorem 3.** *The Bernstein basis (10) is the only minimal basis with no zeros on  $t \in (a, b)$ .*

*Proof.* Let  $\Phi = \{\phi_0, \dots, \phi_n\}$  be a minimal basis with no zeros on  $(a, b)$ . Suppose that  $\phi_i$  has zeros of multiplicity  $\ell_i$  and  $r_i$  at  $t = a$  and  $t = b$  for  $i = 0, \dots, n$ . Suppose further that  $\ell_i = \ell_j$  for some  $i \neq j$ . If  $r_i \leq r_j$ , then  $Z(\phi_i) \subseteq Z(\phi_j)$ , while if  $r_j \leq r_i$ , then  $Z(\phi_j) \subseteq Z(\phi_i)$ . By Proposition 2, this contradicts  $\Phi$  being minimal. So  $\ell_0, \dots, \ell_n$  must be distinct, and we may order the basis such that  $\ell_i = i$  for  $i = 0, \dots, n$ . If, for some  $i < j$ , we have  $r_i \leq r_j$ , then  $Z(\phi_i) \subseteq Z(\phi_j)$ , which again contradicts the minimality of  $\Phi$ . So  $i < j$  must imply that  $r_i > r_j$ , and hence  $r_i = n - i$  for  $i = 0, \dots, n$ . Thus, we have  $\phi_i(t) = c_i(b - t)^{n-i}(t - a)^i$  where  $c_i > 0$  for  $i = 0, \dots, n$ .  $\square$

A closely related result is Theorem 7 of [5], which states that if the basis  $\Phi = \{\phi_0, \dots, \phi_n\}$  is such that every root — real or complex — of each basis function lies outside the open disk  $D$  in the complex plane having the real interval  $t \in (a, b)$  as diameter, then  $\mathcal{B} \prec \Phi$ .

*Remark 2.* For any value  $\frac{1}{2} < a < 1$ , the following form a minimal basis for  $\Pi_2$  on  $t \in [0, 1]$  that is a partition of unity, yet is distinct from the Bernstein basis:

$$\phi_0(t) = \frac{2t(1-t)}{a}, \quad \phi_1(t) = \frac{(2a-1)t^2}{a^2}, \quad \phi_2(t) = \frac{(t-a)^2}{a^2}.$$

The minimality of this basis can be verified by methods similar to those used in Remark 1.

5. POWER, BERNSTEIN, AND BALL BASES

We now illustrate the above ideas in the context of some parametric curve and surface representations commonly used in computer-aided geometric design. The properties of the Bernstein-Bézier form

$$(16) \quad \mathbf{r}(t) = \sum_{k=0}^n \mathbf{p}_k b_k^n(t), \quad b_k^n(t) = \binom{n}{k} (1-t)^{n-k} t^k$$

of a degree- $n$  curve defined on  $t \in [0, 1]$  are well known [4]. Note that  $n$  is only an *upper bound* on the true degree of the curve. It is not possible, by mere inspection of the Bézier control points  $\mathbf{p}_0, \dots, \mathbf{p}_n$ , to ascertain whether  $\mathbf{r}(t)$  is, in fact, a lower-degree curve “masquerading” as a degree- $n$  curve (the condition for this is that the  $n$ th forward difference  $\Delta^n \mathbf{p}_0$  of the control points should vanish).

To address this problem, the four functions

$$\beta_0^3(t) = (1-t)^2, \quad \beta_1^3(t) = 2(1-t)^2 t, \quad \beta_2^3(t) = 2(1-t)t^2, \quad \beta_3^3(t) = t^2$$

have been proposed [1] as a basis for cubic polynomials. This basis has the advantage that, when a cubic is described in terms of “Ball control points”  $\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$  as

$$\mathbf{r}(t) = \sum_{j=0}^3 \mathbf{b}_j \beta_j^3(t),$$

a coincidence of the two interior points ( $\mathbf{b}_1 = \mathbf{b}_2$ ) is sufficient and necessary for  $\mathbf{r}(t)$  to degenerate to a *quadratic* curve, i.e., a parabola segment.

A generalization of the cubic Ball basis that retains the “degree-reduction property” has been given in [9]. For polynomials of *odd* degree

$$n = 2m + 1,$$

the generalized Ball basis on  $t \in [0, 1]$  is defined by

$$(17) \quad \beta_j^n(t) = \binom{m+j}{j} (1-t)^{m+1} t^j = \beta_{n-j}^n(1-t)$$

for  $j = 0, \dots, m$ . A degree- $n$  curve with control points  $\mathbf{b}_0, \dots, \mathbf{b}_n$  in the basis (17) exhibits the *convex-hull* and *variation-diminishing* properties in common with Bézier curves [7, 8]. Furthermore, a coincidence

$$\mathbf{b}_{m-r+1} = \mathbf{b}_{m-r+2} = \dots = \mathbf{b}_{m+r-1} = \mathbf{b}_{m+r}$$

of the  $2r$  “middle” control points induces an  $r$ -fold degree reduction, i.e., the curve is then of true degree  $n - r$  rather than  $n$ . The Ball form also offers a recursive algorithm [9] for evaluating/subdividing curves that is actually somewhat more efficient than the de Casteljau algorithm for Bézier curves.

For curves of *even* degree,  $n = 2m$ , the Ball basis is defined by (17) for  $j = 0, \dots, m - 1$  while the “central” basis function is given by

$$\beta_m^n(t) = \binom{2m}{m} (1-t)^m t^m.$$

The methods outlined above can be readily modified to accommodate even-degree curves; see [7, 8, 9, 10] for details.

While the Ball and Bézier representations share many useful features, we will show that the latter is superior in one important respect — it is a “better-conditioned” or more *stable* representation. Since both bases are nonnegative, we

see from Theorem 1 that this is equivalent to the Ball basis being expressible as a nonnegative combination of the Bernstein basis.

Furthermore, whereas the Bernstein form is systematically more stable than the power form, the Ball representation does *not* inherit this property. Indeed, since the Ball basis is symmetric, Corollary 3 indicates that it cannot be systematically more stable than the power basis, i.e., the power basis is *not* a nonnegative combination of the Ball basis. We now proceed to give explicit representations for the Ball basis in terms of the Bernstein basis and for the power basis in terms of the Ball basis.<sup>2</sup>

**5.1. The Ball basis in terms of the Bernstein basis.** To express the  $j$ th Ball basis function of (odd) degree  $n = 2m + 1$  in terms of the degree- $n$  Bernstein basis, i.e., to determine the matrix elements  $\Lambda_{jk}$  in

$$\beta_j^n(t) = \sum_{k=0}^n \Lambda_{jk} b_k^n(t),$$

we multiply (17) by the binomial expansion of  $1 = [(1 - t) + t]^{m-j}$ :

$$\begin{aligned} \beta_j^n(t) &= \binom{m+j}{j} (1-t)^{m+1} t^j \sum_{i=0}^{m-j} \binom{m-j}{i} (1-t)^{m-j-i} t^i \\ &= \binom{m+j}{j} \sum_{i=0}^{m-j} \binom{m-j}{i} (1-t)^{2m+1-(i+j)} t^{i+j}. \end{aligned}$$

Setting  $k = i + j$  in the above then gives

$$(18) \quad \beta_j^n(t) = \sum_{k=j}^m \frac{\binom{m+j}{j} \binom{m-j}{k-j}}{\binom{n}{k}} b_k^n(t) \quad \text{for } j = 0, \dots, m.$$

This defines rows  $j = 0, \dots, m$  of the matrix  $\{\Lambda_{jk}\}_{0 \leq j, k \leq n}$ . Analogously, one finds that rows  $j = m + 1, \dots, n$  are determined by the expression

$$(19) \quad \beta_{n-j}^n(t) = \sum_{k=m+1}^{n-j} \frac{\binom{m+j}{j} \binom{m-j}{k-m-1}}{\binom{n}{k}} b_k^n(t) \quad \text{for } j = 0, \dots, m.$$

Hence from (18) and (19) we have for  $j = 0, \dots, m$  a complete description of the basis-conversion matrix as follows:

$$\Lambda_{jk} = \begin{cases} 0 & \text{when } 0 \leq k < j, \\ \frac{\binom{m+j}{j} \binom{m-j}{k-j}}{\binom{n}{k}} & \text{when } j \leq k \leq m, \\ 0 & \text{when } m < k \leq n, \end{cases}$$

$$\Lambda_{n-j,k} = \begin{cases} 0 & \text{when } 0 \leq k < m + 1, \\ \frac{\binom{m+j}{j} \binom{m-j}{k-m-1}}{\binom{n}{k}} & \text{when } m + 1 \leq k \leq n - j, \\ 0 & \text{when } n - j < k \leq n. \end{cases}$$

<sup>2</sup>We use the unit interval for simplicity; the arguments generalize readily to an arbitrary interval  $t \in [a, b]$  if we take  $1, t - a, \dots, (t - a)^n$  or  $1, b - t, \dots, (b - t)^n$  for the power basis.

The generalized Ball basis of odd degree  $n = 2m + 1$  is thus evidently a nonnegative combination<sup>3</sup> of the Bernstein basis of the same degree. By similar arguments, it can be verified that this result also holds for even-degree Ball and Bernstein bases — see, for example, [10].

**5.2. The power basis in terms of the Ball basis.** The most convenient means of expressing the power basis  $1, t, \dots, t^n$  in terms of the generalized Ball basis of odd degree  $n = 2m + 1$  is to follow the approach used in [9]. We note that the polynomial  $T_n[f(t)]$  of degree  $n = 2m + 1$  that interpolates the values and first  $m$  derivatives

$$f(0), f'(0), f''(0), \dots, f^{(m)}(0) \quad \text{and} \quad f(1), f'(1), f''(1), \dots, f^{(m)}(1)$$

of a function  $f(t)$  at  $t = 0$  and  $t = 1$  can be expressed as

$$\begin{aligned} T_n[f(t)] = & \sum_{k=0}^m (1-t)^{m+1} t^k \frac{1}{k!} \frac{d^k}{dt^k} \frac{f(t)}{(1-t)^{m+1}} \Big|_{t=0} \\ & + \sum_{k=0}^m (1-t)^k t^{m+1} (-1)^k \frac{1}{k!} \frac{d^k}{dt^k} \frac{f(t)}{t^{m+1}} \Big|_{t=1}. \end{aligned}$$

Since  $T_n[t^j] \equiv t^j$  for  $j = 0, \dots, n$ , the matrix coefficients  $\Gamma_{jk}$  that express the first  $n$  powers of  $t$  in terms of the Ball basis of degree  $n$ ,

$$t^j = \sum_{k=0}^n \Gamma_{jk} \beta_k^n(t),$$

are evidently given by

$$\begin{aligned} \Gamma_{jk} &= \frac{m!}{(m+k)!} \frac{d^k}{dt^k} \frac{t^j}{(1-t)^{m+1}} \Big|_{t=0}, \\ \Gamma_{j,n-k} &= (-1)^k \frac{m!}{(m+k)!} \frac{d^k}{dt^k} \frac{t^j}{t^{m+1}} \Big|_{t=1} \end{aligned}$$

for  $j = 0, \dots, n$  and  $k = 0, \dots, m$ . Further, by use of the Leibniz rule

$$\frac{d^r}{dt^r} fg = \sum_{i=0}^r \binom{r}{i} \frac{d^{r-i}}{dt^{r-i}} f \frac{d^i}{dt^i} g$$

for the derivatives of the product of functions  $f$  and  $g$ , it can be verified by straightforward but tedious calculations that these expressions reduce to

$$\Gamma_{jk} = \begin{cases} 0 & \text{if } k < j, \\ \frac{(k-j+1) \cdots k}{(m+k-j+1) \cdots (m+k)} & \text{if } k \geq j, \end{cases}$$

$$\Gamma_{j,n-k} = (-1)^k \frac{(j-m-k) \cdots (j-m-1)}{(m+1) \cdots (m+k)}$$

for  $j = 0, \dots, n$  and  $k = 0, \dots, m$ .

From the above it is evident that, while the first  $m + 1$  columns of the matrix  $\{\Gamma_{jk}\}_{0 \leq j, k \leq n}$  have only nonnegative entries, those in the last  $m + 1$  columns are

<sup>3</sup>In fact, the matrix that transforms the Bernstein basis into the Ball basis is not only nonnegative, but *totally positive* — i.e., all of its minors are nonnegative [8].

not all nonnegative — in particular, those elements for which  $j > m + k$  and  $k$  is odd are negative. This can be verified by enumerating a few low-degree examples:

$$\begin{aligned} \begin{bmatrix} 1 \\ t \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0^1(t) \\ \beta_1^1(t) \end{bmatrix}, \\ \begin{bmatrix} 1 \\ t \\ t^2 \\ t^3 \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1/2 & 1/2 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1/2 & 1 \end{bmatrix} \begin{bmatrix} \beta_0^3(t) \\ \beta_1^3(t) \\ \beta_2^3(t) \\ \beta_3^3(t) \end{bmatrix}, \\ \begin{bmatrix} 1 \\ t \\ t^2 \\ t^3 \\ t^4 \\ t^5 \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1/3 & 1/2 & 1/2 & 2/3 & 1 \\ 0 & 0 & 1/6 & 1/6 & 1/3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1/3 & 1 \\ 0 & 0 & 0 & 1/6 & -2/3 & 1 \end{bmatrix} \begin{bmatrix} \beta_0^5(t) \\ \beta_1^5(t) \\ \beta_2^5(t) \\ \beta_3^5(t) \\ \beta_4^5(t) \\ \beta_5^5(t) \end{bmatrix}. \end{aligned}$$

Similar results are obtained for even-degree bases.

### 6. THE WILKINSON POLYNOMIALS

Two polynomials, first studied [11] by J. H. Wilkinson in 1959, have become *de facto* “benchmarks” for assessing the stability of various representations for univariate polynomials. Both are of degree 20 and have 20 distinct real roots: in the first case the roots are equidistant, while in the second the spacing of the roots corresponds to a geometric progression. The stability of the power and Bernstein forms of these polynomials was assessed in [5]; here we extend this comparison to include the generalized Ball representation.

**6.1. Equidistant roots.** The Wilkinson polynomial with equally spaced roots

$$P(t) = \prod_{k=1}^n (t - k/n), \quad n = 20,$$

is notoriously ill-conditioned. Figure 1 shows the computed root condition numbers for this polynomial in the power, Bernstein, and Ball bases.

While the power form is extremely unstable, the Bernstein and Ball representations both offer a significant improvement. The Ball form is, in fact, only marginally worse than the Bernstein form in this case, with condition numbers no more than about one order of magnitude greater than the latter. This would typically result in the loss of about one decimal digit of accuracy in floating-point root calculations.

**6.2. Geometric sequence of roots.** For the Wilkinson polynomial with geometrically spaced roots,

$$P(t) = \prod_{k=1}^n (t - 2/2^k), \quad n = 20,$$

it can be seen in Figure 2 that the situation is dramatically different.

While the power and Bernstein forms are exceedingly well-conditioned — the latter being slightly better than the former — the Ball representation is seen to be extremely unstable. Indeed, random fractional perturbations of order  $10^{-10}$  in the Ball coefficients of this polynomial can lead to displacements of order unity for those roots in the range  $0.001 \lesssim t \lesssim 0.05$ .

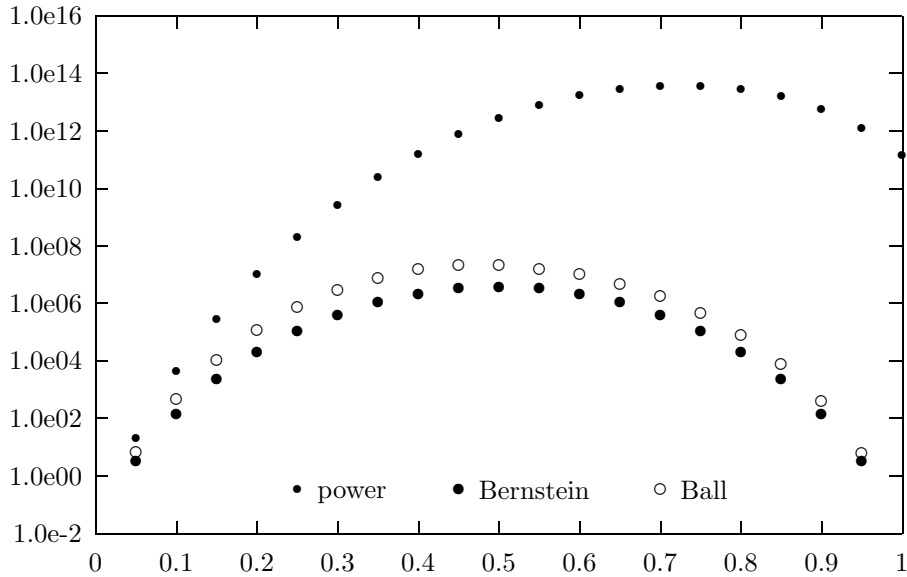


FIGURE 1. Root condition numbers for the first Wilkinson polynomial

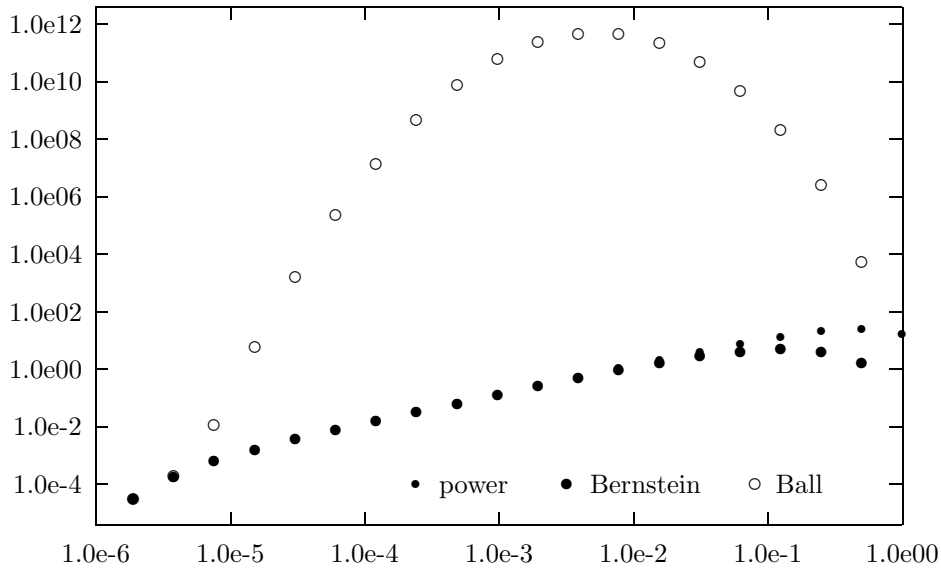


FIGURE 2. Root condition numbers for the second Wilkinson polynomial

With standard double-precision arithmetic (corresponding to a resolution of about 15 decimal digits), the power and Bernstein forms would allow all roots to be computed to 14 or 15 accurate digits, whereas the Ball form would yield only about 4 or 5 accurate digits for the most ill-conditioned roots.

This example offers a convincing illustration of the fact that the Ball form is not only systematically less stable than the Bernstein form — it may also be much

less stable than the power form. Incidentally, it is interesting to note that the Chebyshev basis on  $t \in [0, 1]$  also gives a very unstable representation of this polynomial; see Example 4.2' in [6]. Some of the root condition numbers are as large as  $10^{55}$ ! (that's an exclamation mark, not a factorial —  $10^{55}$  is surely a sufficiently impressive number in its own right).

## 7. CONCLUDING REMARKS

The intrinsic stability of the adopted representation scheme is an important issue that can profoundly influence the accuracy and reliability of various calculations on parametric curves and surfaces. Since such calculations are usually performed in floating-point arithmetic, it is proper that new schemes be judged as much upon this basis as on purely geometric considerations. We have shown that, in addition to its attractive geometric attributes, the Bernstein–Bézier form is “optimally stable” — it is *impossible* to formulate alternate representations, based on nonnegative basis functions, that yield systematically smaller condition numbers. The power and generalized Ball bases, for example, are systematically worse-conditioned than the Bernstein basis (often by very large factors, as exemplified by the Wilkinson examples).

## REFERENCES

1. A. A. Ball, *CONSURF part one: Introduction to conic lofting tile*, Comput. Aided Design **6** (1974), 243–249.
2. J. M. Carnicer and J. M. Peña, *Shape preserving representations and optimality of the Bernstein basis*, Adv. Comp. Math. **1** (1993), 173–196. MR **94i**:65138
3. J. M. Carnicer and J. M. Peña, *Total positivity and optimal bases*, in Total Positivity and its Applications (M. Gasca and C. A. Micchelli, eds.), Kluwer Academic Publishers, Dordrecht, 1996, pp. 133–155.
4. G. Farin, *Curves and Surfaces for Computer Aided Geometric Design*, Academic Press, Boston, 1993. MR **93k**:65016
5. R. T. Farouki and V. T. Rajan, *On the numerical condition of polynomials in Bernstein form*, Comput. Aided Geom. Design **4** (1987), 191–216. MR **89a**:65028
6. W. Gautschi, *Questions of numerical condition related to polynomials*, in Studies in Numerical Analysis, MAA Studies in Mathematics **24**, (G. H. Golub, ed.) 1984, 140–177. CMP 20:07
7. T. N. T. Goodman and H. B. Said, *Properties of generalized Ball curves and surfaces*, Comput. Aided Design **23** (1991), 554–560.
8. T. N. T. Goodman and H. B. Said, *Shape preserving properties of the generalized Ball basis*, Comput. Aided Geom. Design **8** (1991), 115–121. MR **92c**:65181
9. H. B. Said, *A generalized Ball curve and its recursive algorithm*, ACM Trans. Graphics **8** (1989), 360–371.
10. T. V. To, *Polar Form Approach to Geometric Modeling*, Dissertation No. 92, Asian Institute of Technology, Bangkok, Thailand, 1992.
11. J. H. Wilkinson, *The evaluation of the zeros of ill-conditioned polynomials. Parts I & II*, Numer. Math. **1** (1959), 150–166. MR **22**:321
12. J. H. Wilkinson, *Rounding Errors in Algebraic Processes*, Dover (reprint), New York, 1963. CMP 94:14

DEPARTMENT OF MECHANICAL ENGINEERING & APPLIED MECHANICS, UNIVERSITY OF MICHIGAN, ANN ARBOR, MICHIGAN 48109  
*E-mail address:* farouki@engin.umich.edu

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, UNIVERSITY OF DUNDEE, DUNDEE DD1 4HN, SCOTLAND  
*E-mail address:* tgoodman@mcs.dundee.ac.uk