Spline functions and total positivity.

M. GASCA *

Dedicado al Profesor Baltasar Rodríguez Salinas, maestro de tantos matemáticos españoles, con admiración y afecto.

Abstract

In this survey we show the close connection between the theory of Spline Functions and that of Total Positivity. In the last section we mention some recent results on totally positive bases which are optimal for shape preserving properties in Computer Aided Geometric Design.

1 Totally positive functions and matrices

Totally positive functions play an important role in Approximation Theory and in many other branches of Mathematics: Analysis, Probability, Combinatorics, Statistics, etc., with applications in Economics, Biology, Computer Aided Geometric Design,... They started to be systematically studied about sixty years ago, when F. R. Gantmacher and M. G. Krein in the former Soviet Union and a little later I. J. Schoenberg in the United States published their pioneering papers on this subject. Anyway, much of the theory was influenced at the beginning by the earlier work of many important mathematicians, as mentioned by Karlin in [19]: Stieltjes, Chebyshev, Bernstein, Haar, Fejér, Schur, Pólya, Fekete,...
In 1968 S. Karlin published his fundamental book Total Positivity [19], where he develops the theory and includes almost all the results appeared until that time. It was published as Volume I but the second volume has been never written until now. The book is so impressive and complete that probably this has caused that no other book on the subject has been published until 1995. However the great quantity of papers devoted to the different applications of the theory encouraged us to promote the organization of an international meeting to summarize them. An International Workshop on Total Positivity and its Applications (IWTPA) took place in Jaca (Spain) in September 1994, meeting together to many of the present researchers in this area. As an important consequence, a new book [17] entitled Total Positivity and its Applications (M. Gasca and C. A. Micchelli as editors) has been just published by Academic Publishers. A detailed history of the first results on Total Positivity can be reconstructed from the historical remarks of [19] and [23] and also from the paper [21] by Allan Pinkus [17].

A real function (or kernel) \( K(x, y) \) of two variables belonging to linearly ordered sets \( X, Y \) is said to be totally positive if for each positive integer and for all

\[
x_1 < x_2 < \cdots < x_m; \quad y_1 < y_2 < \cdots < y_m, \quad x_i \in X, y_j \in Y,
\]

one has

\[
\det(K(x_i, y_j))_{1 \leq i,j \leq m} > 0.
\]

If both \( X, Y \) are finite sets then \( K \) can be considered a finite matrix and then we speak of totally positive (\( TP \)) matrices. Consequently, a \( TP \) matrix is a matrix whose minors are all nonnegative. When the inequalities are strict in the above definitions we speak of strictly totally positive functions or matrices. In the german literature the term total nichtnegativität has been frequently used instead of total positivity. There are many examples of \( TP \) functions, in particular in Statistics and also the Green’s functions associated with many standard boundary-value problems of Sturm-Liouville.

2 The origins of spline functions

Piecewise polynomial functions have been considered for a long time in order to gain flexibility with respect to polynomials, above all when one
is considering large intervais. The aim of maintaining global continuity as much as possible gave rise to the theory of spline functions.

Polynomial splines of degree \( k \) with simple knots are piecewise polynomials of degree \( k \) with global continuity of order \( k - 1 \). The term \textit{spline} to singularize these functions was introduced by I. J. Schoenberg in 1946, due to the thin rods of elastic material traditionally used by designers, which are called splines. An ideal spline of this class passing through several points is in essence represented by a piecewise cubic with global continuity of order 2. More details can be found in many introductory texts on the subject.

Since the study of the involved mathematical problems can be traced until Euler’s time one can say that splines have a very old history. But it was not until Schoenberg that they were systematically studied and frequently used for approximation purposes. Many other authors are mentioned in a historical remark in [23]. However Schoenberg and mathematicians related to him (de Boor, Karlin, Micchelli, Schumaker and many others) have produced, above all from the sixties, most of the important results on splines. Other historical remarks on some aspects of the theory can be found in several papers in [17], for example [20] and [1]. One of the reasons of the great success of splines in Approximation Theory was the discovery of very efficient algorithms to work with them and the simultaneous development of digital computers where those algorithms can be implemented.

Although the original splines had simple knots, repetition of knots was soon associated to lower order of continuity at them, as we present in the next section.

### 3 Spline functions

Let \( \Delta = \{x_i\}_{1 \leq i \leq r} \) be a set of real points with \( x_1 < x_2 < \cdots < x_r, \) a positive integer and \( \mathcal{M} = (m_1, m_2, \ldots, m_r) \) a vector of integers with \( 1 \leq m_i \leq k + 1 \forall i. \) Consider an interval \([a, b]\), finite or infinite, such that \( x_i \in [a, b] \forall i, \) and denote \( x_0 = a, x_{r+1} = b. \) A piecewise polynomial function \( f \) of degree \( k \) on each of the intervals \([x_i, x_{i+1}], 0 \leq i \leq r-1, [x_r, x_{r+1}] \) and continuity of order \( k - m_i \) at the knot \( x_i \) is called a \textit{spline of degree} \( k \) \textit{with knots} \( x_i \) \textit{of multiplicity} \( m_i, 1 \leq i \leq r. \) Following the notation of [23] the space spanned by these functions will
be denoted by $S(\mathcal{P}_k, \mathcal{M}, \Delta)$.

The role played by the space $\mathcal{P}_k$ can be played by other extended complete Chebyshev spaces, giving rise to the so-called Chebyshevian splines. We restrict our attention to polynomial splines. Moreover, right continuity at the knots could be obviously considered instead of left continuity. Since a spline $s \in S(\mathcal{P}_k, \mathcal{M}, \Delta)$ is determined by $(k+1) \times (r+1)$ coefficients subject to $\sum_{i=1}^{r}(k+1-m_i)$ continuity conditions, it easily follows that the dimension of the space is $k+1 + \sum_{i=1}^{r} m_i$.

For brevity we denote $n = \sum_{i=1}^{r} m_i$ and consequently the dimension is $k+n+1$.

A natural basis for this space is the union of any basis of $\mathcal{P}_k$ with the set of truncated power functions

$$\left\{ (x-x_i)^{k+1-j} \right\}_{1 \leq i \leq r, 1 \leq j \leq m_i},$$

where, as usual,

$$x_i^+ = \begin{cases} x_i, & \text{if } x \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently, a spline $s \in S(\mathcal{P}_k, \mathcal{M}, \Delta)$ can be expressed in the form

$$s(x) = \sum_{i=0}^{k} c_i x^i + \sum_{i=1}^{r} \sum_{j=1}^{m_i} c_{ij} (x-x_i)^{k+1-j}. \quad (1)$$

These bases are not convenient for many purposes. In particular, their functions do not have compact support. On the contrary, many splines have compact support. Furthermore, when we move to the right in the real line, many functions of the basis are not zero, although only $k+1$ truncated powers are needed to represent a polynomial of degree $k$ between two consecutive knots. This means that the truncated power functions are locally linearly dependent.

Looking for splines of minimal support and locally linearly independent, it happens that if $m_i + m_{i+1} + \cdots + m_j > k+1$ there exists nonzero splines $s(x)$ such that

$$s(x) = 0 \text{ for } x < x_i \text{ or } x > x_j.$$

If $m_i + m_{i+1} + \cdots + m_j \leq k+1$, then no such nonzero spline exists. In this sense, splines like these with $m_i + m_{i+1} + \cdots + m_j = k+2$ have
least support. They can be introduced in several different ways, being the use of the divided differences of the Green's function \((x, y)^k_x\), one of the most elegant ways. It requires to form a new partition of points as follows:

\[
y_1 \leq \cdots \leq y_{k+1} \leq a \text{ if } a > -\infty
\]
\[
y_1 \leq \cdots \leq y_{k+1} < x_1 \text{ if } a = -\infty
\]
\[
y_{k+2} = y_{k+3} = \cdots = y_{k+m_1+1} = x_1
\]
\[
y_{k+m_1+2} = y_{k+m_1+3} = \cdots = y_{k+m_1+m_2+1} = x_2
\]

\[
y_{k+n-m_r+2} = y_{k+m_1+3} = \cdots = y_{k+n+1} = x_r
\]
\[
b \leq y_{k+n+2} \leq \cdots \leq y_{n+2k+2} \text{ if } b < \infty
\]
\[
x_r < y_{k+n+2} \leq \cdots \leq y_{n+2k+2} \text{ if } b = \infty.
\]

Observe that one has

\[
y_{i+k+1} > y_i \quad \forall i.
\]

Now we define the functions

\[
B_i(x) = (y_{i+k+1} - y_i)[y_i, \cdots, y_{i+k+1}](y-x)^k_{+} \quad a \leq x \leq b
\]

or equivalently

\[
B_i(x) = (-1)^k(y_{i+k+1} - y_i)[y_i, \cdots, y_{i+k+1}](x-y)^k_{+},
\]

\(1 \leq i \leq n + k + 1\). Here \([y_i, \cdots, y_{i+k+1}]f(x, y)\) means the divided difference of \(f\) with respect to the variable \(y\) and with arguments \(y_i, \cdots, y_{i+k+1}\).

These functions \(B_i(x)\) (up to a constant factor for different normalizations) are called \(B\)-splines and form a basis of \(S(P_k, M, \Delta)\), with

\[
B_i(x) = 0 \forall x \notin [y_i, y_{i+k+1}]
\]
\[
B_i(x) > 0 \forall x \in (y_i, y_{i+k+1}).
\]

The above ones have been normalized in order to satisfy

\[
\sum_{i=1}^{n+k+1} B_i(x) = 1 \quad \forall x \in [a, b].
\]
$B_i$ is a spline of degree $k$ and knots $y_i, \ldots, y_{i+k+1}$. For each $x$, no more than $k+1$ of the $B_i(x)$'s are different from zero, and the set \{$B_i, B_{i+1}, \ldots, B_j$\}, with $j - 1 \geq k + 1$, is a basis of the space of splines of degree $k$ on $[y_{i+k}, y_{j+1})$ and knots $y_{i+k+1}, \ldots, y_j$.

Nice recurrence relations produce easy computational manipulations of the $B$-splines, which means a great practical advantage with respect to other bases.

It is worth to mention that many of the properties of this type of functions had been studied by the bulgarian mathematician L. Tchakaloff some years before the first papers on splines, but with a different terminology. Moreover, his papers on these questions, written in bulgarian, had a short diffusion. See the paper by B. Bojanov [1] in [17] for more details.

4 B-splines and total positivity

One of the nicest properties of B-splines concerns the sign of collocation matrices for Lagrange interpolation problems.

Let us consider a set of strictly increasing real numbers $y_1 < y_2 < \cdots < y_{n+k+1}$ and let $B_1, B_2, \ldots, B_n$ be the corresponding B-splines defined by (2), with $[y_i, y_{i+k+1}]$ as the support of $B_i$. Let $t_1 < t_2 < \cdots < t_n$ be real numbers and consider the collocation matrix

$$M = \begin{pmatrix} t_1 & t_2 & \cdots & t_n \\ B_1 & B_2 & \cdots & B_n \end{pmatrix} = (B_{j}(t_i))_{1 \leq i, j \leq n}. \quad (3)$$

The theorem that states that the matrix (3) is nonsingular if and only if

$$t_i \in \{x \mid B_i(x) \neq 0\} \quad i = 1, 2, \ldots, n,$$

i.e.

$$y_i < t_i < y_{i+k+1} \quad 1 \leq i \leq n \quad (4)$$

is known as Schoenberg and Whitney although the original Theorem 2 of [22] was presented in a slightly different form, as we shall see. In fact this result is very rich and can be presented in several different forms, some of them giving more information about the sign of the minors of the collocation matrix $M$. 

As we have said above, if \( n \geq k + 2 \) the B-splines \( B_1, B_2, \ldots, B_n \) form a basis of the space of splines of degree \( k \) and knots \( y_k + 2, \ldots, y_n \), i.e. the splines that can be expressed in the form

\[
s(x) = \sum_{j=0}^{k} a_j x^j + \sum_{j=k+2}^{n} c_j (x - y_j)^k.
\]  

(5)

Observe that in this form the Schoenberg-Whitney theorem states that there exists a unique spline (5) with prescribed values at the points \( t_1 < t_2 < \cdots < t_n \) if and only if

\[
t_i < y_{i+k+1} < t_{i+k+1} \quad 1 \leq i \leq n - k - 1.
\]  

(6)

This is precisely the statement of [22, Theorem 2], where B-splines were not used because they had not been introduced yet. The proof is based on Laplace transforms and is quite different to those based on B-splines, which are rather simpler (see [23, Theorem 4.6]).

The roles of knots and interpolation points can be interchanged, as it is seen in [9]. Let us write the system defined by the Lagrange interpolation problem considered above for splines (5):

\[
\sum_{i=0}^{k} a_i t_j^i + \sum_{i=k+2}^{n} c_i (t_j - y_i)^k = z_j \quad 1 \leq j \leq n.
\]  

(7)

where \( z_j \) is the prescribed value of \( s(x) \) at \( t_j \). For \( 1 \leq i \leq n - k + 1 \) let \( M_i(x) \) be the B-spline of degree \( k \) with knots \( t_i, \ldots, t_{i+k+1} \):

\[
M_i(x) = M(x \mid t_i, \ldots, t_{i+k+1}) = [t_i, \ldots, t_{i+k+1}]^k (y - x)^k,
\]

where, as said above, the divided difference is taken on the argument \( y \). Observe that in this case we have not normalized the B-spline with the factor \((t_{i+k+1} - t_i)\).

Recall that

\[
[t_i, \ldots, t_{i+k+1}]f = \sum_{j=1}^{i+k+1} \frac{f(t_j)}{\prod_{\substack{r=1 \atop r \neq j}}^{i+k+1} (t_r - t_j)}.
\]
and that this is zero when \( f \) is a polynomial of degree not greater than \( k \).

We can replace in the system (7) the \( r \)th equation \((r = n, n-1, \ldots, k+2)\) by an adequate linear combination of it and the \( k+1 \) precedent ones in order to get zeros in the first \( k+1 \) coefficients: observe that for \( r = n, n-1, \ldots, k+2 \) one has

\[
a_h \sum_{j=r-k}^{r} \frac{t_j^h}{r} = a_h [t_{r-k-1}, \ldots, t_r] t^h = 0, \quad 0 \leq h \leq k,
\]

\[
c_h \sum_{j=r-k}^{r} \frac{(t_j - y_h)^k}{r} = \prod_{i=r-k+1}^{r} (t_i - t_j)
\]

\[
c_h [t_{r-k-1}, \ldots, t_r] (t_j - y_h)^k = c_h M_{r-k-1}(y_h), \quad k+2 \leq h \leq n,
\]

where the divided difference is taken on the argument denoted by the dot.

The new system can be written

\[
\sum_{i=0}^{k} a_i t^i_j + \sum_{i=k+2}^{n} c_j (t_j - y_i)^k = z_j, \quad 1 \leq j \leq k + 1
\]

\[
\sum_{i=k+2}^{n} c_i M_j(y_i) = \hat{z}_j, \quad 1 \leq j \leq n - k - 1
\]

and it has a solution for all data \( z_1, z_2, \ldots, z_{k+1}, \hat{z}_1, \hat{z}_2, \ldots, \hat{z}_{n-k-1} \), that is all \( z_1, z_2, \ldots, z_n \), if and only if

\[
\det M_j(y_j)_{1 \leq i \leq n-k-1; 1 \leq j \leq n} \neq 0.
\]

In other words, the interpolation problem (7) has a unique solution if and only if the interpolation problem determined by the space spanned by \( M_1, \ldots, M_{n-k-1} \) and the interpolation points \( y_{k+2}, y_{k+3}, \ldots, y_n \) has a unique solution. According to (3) (4) that happens if and only if

\[
y_j \in \{ x \mid M_{j-k-1}(x) \neq 0 \} \quad j = k + 2, k + 3, \ldots, n,
\]

i.e.

\[
t_j - k - 1 < y_j < t_j - k + 2 \leq j \leq n.
\]
Observe that (4) implies (10) and that (10) implies (4) except for the conditions for \(y_1, y_2, \ldots, y_{k+1}\) that do not appear in the last version of the problem, and remark that the roles of knots and interpolation points have been interchanged in both results.

Let us observe too that the matrix

\[
M_1(\eta_j)_{1 \leq i \leq n-k-1; 1 \leq j \leq n-k+2}
\]

of the coefficients of the last \(n-k-1\) equations of (8) is the Schur complement of the Vandermonde matrix \(\left(\eta_j^i\right)_{0 \leq i \leq k; 1 \leq j \leq k+1}\) in the coefficients matrix of the whole system (7).

The Schoenberg and Whitney theorem holds also for splines with multiple knots. See [23] for more details and historical remarks.

Anyway, there are more interesting properties involved in the collocation matrix (3). In fact, it is a totally positive matrix. The total positivity was previewed by Schoenberg and proved first by Karlin [19] even for Chebyshevian splines (see Section 10 of [19] for historical remarks). De Boor [21] proved it again in a different and simpler way. The total positivity of the matrix means the nonnegativity of all minors of the matrix, but in general it is difficult for TP matrices to know, without calculating it, if a minor of the matrix will be positive or zero. De Boor, in the same paper, proved a nice property of the collocation matrix of the B-splines: any minor of it is positive if and only if the diagonal entries of the minor are all positive, and consequently the minor is zero if and only if there is at least one zero in the diagonal of the minor. Totally positive matrices having this property form an intermediate class between totally positive and strictly totally positive matrices, and have been called by us in [9] almost strictly totally positive matrices.

Dealing with splines, the technique of knot insertion is very useful. Let \(\tau\) and \(\mu\) be two strictly increasing sequences of knots such that \(\tau\) is a subsequence of \(\mu\). Then the space of splines of degree \(k\) and knots \(\tau\) is a subspace of that of the same degree and knots \(\mu\). Let \(b_\tau\) and \(b_\mu\) the row vectors of the corresponding B-spline bases. Then one has

\[
b_\tau = b_\mu A_{\tau\mu}.
\]

The \((i,j)\) entry of \(A_{\tau\mu}\) is usually denoted \(\alpha_j(i)\) and considered as a function of \(i\) is called a discrete B-spline. The matrix \(A_{\tau\mu}\) is totally
positive and is called the collocation matrix of the discrete B-splines. Even, in this case, the matrix is almost strictly totally positive in the above terminology. See [20] for a brief summary and references of the properties of discrete B-splines.

5 Variation diminution

Let us denote $S^-(x_1, x_2, \cdots, x_m)$ the number of sign changes of the sequence of real numbers $x_1, x_2, \cdots, x_m$ discarding zero terms and let $f$ be a function defined in an ordered real set. Denote

$$S^- f = \sup S^-(f(t_1), f(t_2), \cdots, f(t_m))$$

where the supremum is taken over all $t_1, t_2, \cdots, t_m, (t_i \in I)$ with $m$ arbitrary but finite.

Analogously, let $S^+(x_1, x_2, \cdots, x_m)$ be the maximum number of sign changes of the sequence $x_1, x_2, \cdots, x_m$, the zero terms being permitted to take on arbitrary signs. Denote

$$S^+ f = \sup S^+(f(t_1), f(t_2), \cdots, f(t_m)).$$

An important property of totally positive matrices (as a special case of sign-regular matrices, see [19]) is that if $A$ is a nonsingular totally positive matrix of order $n$, then

$$S^-(Ax) \leq S^-(x) \quad \forall x \in \mathbb{R}^n.$$

$$S^+(Ax) \leq S^-(x) \quad \forall x \in \mathbb{R}^n.$$

If $A$ is strictly totally positive, then, for all $x \in \mathbb{R}^n \setminus \{0\}$,

$$S^+(Ax) \leq S^-(x).$$

This means that $A$ has the variation-diminishing property. Roughly speaking, the number of sign changes of the transformed of a sequence is not greater than that of the original sequence. This property has important applications in Computer Aided Geometric Design.

A system of nonnegative functions $(u_i)_{1 \leq i \leq n}$ is said to be totally positive on $I$ if the collocation matrix

$$M\left( \begin{array}{cccc} t_1 & t_2 & \cdots & t_n \\ u_1 & u_2 & \cdots & u_n \end{array} \right) = (u_j(t_i))_{1 \leq i, j \leq n} t_1 < t_2 < \cdots < t_n, t_i \in I$$

satisfies the following conditions:

$$S^-(M) \leq S^-(x) \quad \forall x \in \mathbb{R}^n.$$

$$S^+(M) \leq S^-(x) \quad \forall x \in \mathbb{R}^n.$$
Spline functions and total positivity

is always totally positive. The system is normalized totally positive (NTP) if it is totally positive and \( \sum_{i=1}^{n} u_i = 1 \). For such a system and \( P_1, \ldots, P_n \in \mathbb{R}^k \), we may define a curve \( \gamma(t) \) by

\[
\gamma(t) = \sum_{i=1}^{n} u_i(t) P_i.
\]

(11)

The points \( P_1, \ldots, P_n \) are called control points, because we expect to modify the shape of the curve by changing adequately these points. The polygon with vertices \( P_1, \ldots, P_n \) is called control polygon of \( \gamma \).

For a normalized totally positive system the curve (11) lies in the convex hull of the control polygon and one has an interesting shape preserving property, which is very convenient for design purposes and we call endpoint interpolation property: the initial and final endpoints of the curve and the initial and final endpoints (respectively) of the control polygon coincide. Moreover, other variation-diminishing properties of such systems imply that the monotonicity or convexity of the control polygon are inherited by the curve, and the length, angular variation and number of inflections of the curve are respectively bounded by those of the control polygon.

Totally positive systems are also called ordered complete Chebyshev systems and from the precedent section we deduce that the B-spline basis is one of these systems. As any other of them, B-splines satisfy the following variation diminishing property: for any nontrivial vector \( (c_1, c_2, \ldots, c_n) \)

\[
S^- \left( \sum_{i=1}^{n} c_i B_i(x) \right) \leq S^- (c_1, c_2, \ldots, c_n)
\]

where \( S^- (\sum_{i=1}^{n} c_i B_i(x)) \) is taken in the whole real line (see [23], Theorem 4.76).

Bernstein polynomials

\[
B_{k,i}(x) = \binom{k}{i} (1 - x)^{k-i} x^i, \quad 0 \leq i \leq k
\]

form another normalized totally positive system in \([0,1]\). In fact, they can be introduced as the B-spline basis of degree \( k \) when we take as knots \( t_1 = t_2 = \cdots = t_{k+1} = 0, t_{k+2} = t_{k+3} = \cdots = t_{2k+2} = 1 \), that is all the knots are placed at the endpoints of the interval \([0,1]\).
6 Totally positive bases and optimal bases

We have seen above that the B-spline collocation matrix for different points is almost strictly totally positive. Analogously to the definition of TP systems given in the precedent section we can define almost strictly totally positive systems of functions. As it has been seen in [5], for totally positive bases $B$ of continuous functions the following properties are equivalent:

(i) $B$ is almost strictly totally positive.

(ii) A Schoenberg-Whitney theorem holds for $B$.

(iii) The functions in $B$ are locally linearly independent.

In [7] Carnicer and Peña have given an interesting survey of recent results on optimal bases for Computer Aided Geometric Design. In finite dimensional spaces with totally positive bases there exists a basis (unique up to a positive constant factor for each function) called B-basis such that any totally positive basis of the space can be generated from the B-basis by means of a totally positive matrix. These bases are optimal in several senses. They have optimal shape preserving properties in the sense that the control polygon of a curve constructed with the B-basis imitates better the shape of the curve. These bases are also least supported and have other numerical advantages. Examples of them are again, among others, the Bernstein basis and the B-splines basis.

For many of these last properties it has been very important a better knowledge of TP matrices which has been the result of a series of papers by M. Gasca and J. M. Peña in the last years. In those papers the basis tool has been the systematic use of the so called Neville elimination, that is a matricial elimination procedure which consists of producing zeros in the $k$th column of the matrix in the rows $n, n - 1, \ldots, k + 1$ adding to each row an adequate multiple of the previous one, instead of adding a multiple of the $k$th row as in Gauss elimination. This kind of elimination has proved to be very useful to deal with totally positive matrices. The main point for this usefulness is that Neville elimination is performed by bidiagonal elementary matrices with unit diagonal, which are totally positive if the nonzero elements are nonnegative. On the contrary, in general, the elementary matrices which are used in Gauss elimination are not totally positive.
See for example [10-16] for more details and references.

References


Spline functions and total positivity


Departamento de Matemática Aplicada
Facultad de Ciencias
Universidad de Zaragoza
50009 Zaragoza
España

Recibido: 23 de Mayo de 1996