# THE APPLICATION OF TOTAL POSITIVITY TO COMPUTER AIDED CURVE AND SURFACE DESIGN

by

J. M. Lane

R. F. Riesenfeld

UUCS - 79 - 115

# THE APPLICATION OF TOTAL POSITIVITY TO COMPUTER AIDED CURVE AND SURFACE DESIGN*

by

J. M. Lane

R. F. Riesenfeld

Department of Computer Science
University of Utah
Salt Lake City, Utah

July 1977

Key Words and Phrases: Total positivity, variation diminishing approximation, computer aided curve and surface design.

CR Categories: 5.13 (functional approximation), 8.2.

## ABSTRACT

Of particular importance in an interactive curve and surface design system is the interface to the user. The mathematical model employed in the system must be sufficiently flexible for interaction between designer and machine to converge to a satisfactory result. The mathematical theory of Total Positivity is combined with the interactive techniques of Bézier and Riesenfeld in developing new methods of shape representation which retain the valuable variation-diminishing and convex hull properties of Bernstein and B-spline approximation, while providing improvements in the interactive interface to the user. Specifically, extending the Bézier notion of using a polygon to describe a smooth curve, methods of assigning a weight to each vertex which will control the amount of local fit to the polygon or polygonal net are provided. Thus, the designer can cause "cusps" and "flats" easily by manipulating the "tension" at each vertex. Further, the generalization from curves to surfaces can be done with rectilinear data or triangular data. Illustrations are provided from an experimental implementation of the newly constructed models as a demonstration of their feasibility and utility in computer aided curve and surface design.

---

*This research was supported in part by the National Science Foundation (MCS-74-13017-A01) and the Office of Naval Research (N00014-77-C-0157).

# I. INTRODUCTION

Computer Aided Geometric Design is principally concerned with the modelling of physical objects within a digital computer for automated design and manufacture. Potentially, the computer can free the designer from the limitations of traditional drafting techniques while enhancing and accelerating the production process. Working interactively at a graphics terminal or a numerically-controlled drafting machine, a designer can modify or manipulate an existing model or he can design an object ab initio, using trial and error techniques to produce an acceptable design. The computer system can then generate the necessary information for numerically controlled manufacture and production [1].

## Statement of Problem

The two central problems this paper addresses are:

(1) The development of a unifying theory for the construction and analysis of methods of modelling free-form curves and surfaces in a digital computer, and

(2) The application of this theory toward the construction of new techniques for ab initio design.

These problems are part of a study termed Computer Aided Geometric Design (CAGD).

The realization of such a system for curve and surface description imposes certain constraints on the computer-based model. The model must accurately represent a variety of shapes, be amenable to analysis and manipulation, and must take into account the capabilities and limitations of both the designer and the computer. Although it is clear a mathematical model is required, the properties of shape cannot be characterized entirely by the properties of analytic functions and, therefore, new mathematical techniques for synthesizing, storing and retrieving shape information are needed.

To meet these criteria we restrict our attention to vector-valued bounded piecewise analytic functions from a finite dimensional linear space. That is, we have for curves

$$P(t) = \sum_{i \in I} \phi_i(t) P_i, \tag{1.1}$$

and for surfaces

$$P(u,v) = \sum_{i \in I} \phi_i(u,v) P_i, \tag{1.2}$$

where I is some linear ordered, finite set of integers, $P_i \in R^3$, t and (u,v) are elements of some bounded subset of R and $R^2$, respectively, and the $\{\phi_i(t)\}$ and $\{\phi_i(u,v)\}$ are linearly independent sets of bounded piecewise analytic functions.

The set of points $P_i$ clearly constitutes the controlling parameters with respect to the basis $\{\phi_i, i \in I\}$. It remains, then, to find bases which appropriately model the constraints of a CAGD system. Not

only must the control parameters be suitable for interactive communication between designer and computer, but the basis functions must be appropriate for digital representation and computation. At this point we make the distinction between the fitting of a mathematical representation to a predesigned object and the design of an object ab initio, for our choice of basis may depend on which type of design process we wish to model.

In fitting we are concerned with finding a mathematical description of an existing shape or physical object, and in ab initio design the problem is to create a mathematical representation which meets design constraints which may be entirely subjective, aesthetic considerations. Because we know the exact shape a priori, fitting can be done by digitizing points off the object and interpolating the data with an appropriate approximation method. What constitutes a "good" approximation is ultimately up to the designer and interaction with the model may be necessary. Thus there are hard and soft constraints to be met. The hard constraints are the discrete data which must be interpolated, while the soft constraints are manipulated by the designer in determining an acceptable fit. In general, there are few hard constraints in ab initio design and the designer manipulates this increased flexibility to achieve his final design.

While interpolation techniques appear necessary for the fitting problem, this requirement is not present in ab initio design. Of fundamental importance in ab initio design is the interface to the designer. If an interactive design session is to be successful, the designer must be able to manipulate the control points $P_i$ in an iterative process which converges when a satisfactory shape is achieved.

Thus the shape of the curve or surface must respond predictably to the interactive manipulation of the control points. Experience with interpolation techniques indicates that unwanted kinks and oscillations can occur during the design process, making it difficult to predict the response of the curve to manipulation of the data.

## A Review of Bézier and B-spline Techniques

An ab initio system for the design of automobiles which avoids these difficulties is Bézier's Système Unisurf at Renault [1]. In Système Unisurf the control points are associated with the vertices of a polygon for curves and the vertices of a rectilinear net for surfaces. The resulting curve or surface replicates the gross features of the polygon or rectilinear net. More importantly, the design responds predictably to the manipulation of the control points.

In Bézier's method curves are piecewise polynomials where the curve is given on each piece by

$$P(t) = \sum_{i=0}^{m} \phi_i(t) P_i \ , \qquad (1.3)$$

where

$$\phi_i(t) = \binom{m}{i} t^i (1-t)^{m-1} \ ,$$

for $t\varepsilon(0,1)$, $m$ is the degree of the polynomial, and the $P_i$ are the appropriate vertices for that piece. Figure 1.1 shows a typical Bézier curve.

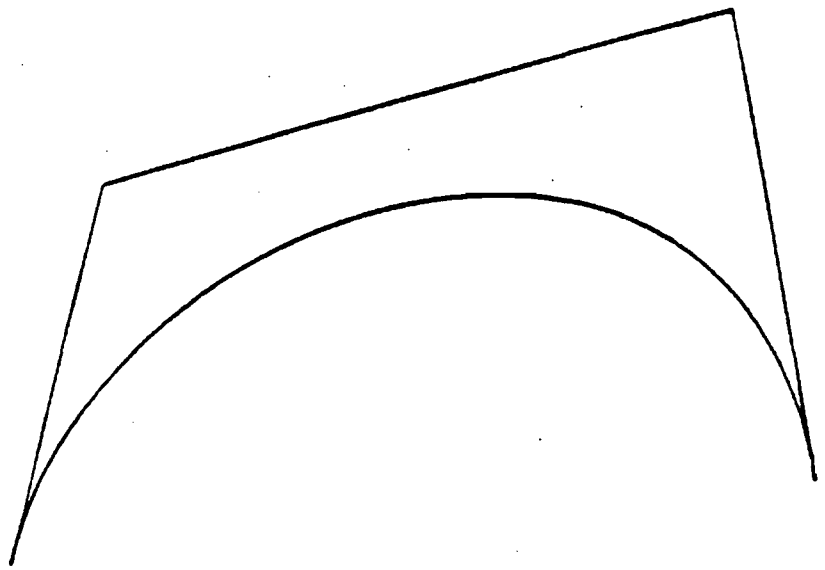

Figure 1.1 Cubic Bézier curve and associated polygon.

Surfaces are generated by taking the tensor product of the basis functions with respect to two orthogonal directions, i.e.,

$$P(u,v) = \sum_{i=0}^{m} \binom{m}{i} u^i (1-u)^{m-i} \left( \sum_{j=0}^{n} \binom{n}{j} v^j (1-v)^{n-j} \right) P_{ij} \qquad (1.4)$$

where $(u,v)\epsilon(0,1)x(0,1)$, m,n are the degrees of the polynomial with respect to the u and v directions and the $P_{ij}$ are the vertices of the associated rectilinear network. Figure 1.2 shows a Bézier network and the corresponding surface.

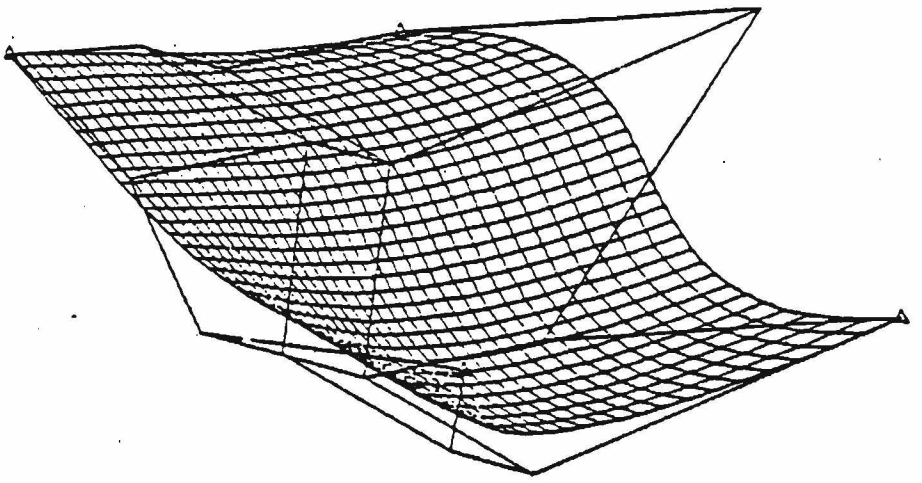

Figure 1.2 Bicubic Bézier surface and associated rectilinear network of points.

Gordon and Riesenfeld [14] have associated the remarkable reproducing power of the Bernstein-Bézier methods with:

(1) the fact that Bernstein approximation is variation-diminishing, and

(2) the fact that the Bernstein basis functions are positive and sum identically to 1, i.e.,

$$\sum_{i=0}^{m} \phi_i(x) \equiv 1$$

and

$$\phi_i(x) > 0$$

for all x, such that

$$x\varepsilon(0,1) \text{ and } i = 0,1,\ldots, m. \qquad (1.5)$$

If we let V(f) denote the number of sign changes of some function f and $V[x_1,x_2, \ldots, x_m]$ denotes the number of sign changes of the indicated sequence, then property (1) is equivalent to

$$V[P] \leq V[P_0,P_1, \ldots, P_m] \tag{1.6}$$

Property (2) is commonly referred to as the convex hull property, since the conditions stated are necessary and sufficient for any curve of the form (1.1) to lie within the convex hull of the coefficients $P_i$.

Note that, in general, the piecewise Bernstein approximation is only $C^0$ continuous. Although it is not difficult to "fix up" $C^1$ continuity, many applications in the automobile and aircraft industries require at least $C^2$ continuity. Riesenfeld [5] has recently proposed vector-valued B-spline approximation as a generalization of Bernstein approximation which retains the variation-diminishing and convex hull properties while improving the continuity class of the curves and surfaces to arbitrary smoothness.

The vector-valued B-spline approximation of degree M-1 to the associated polygon $[P_1,P_2, \ldots, P_n]$ for integral knot spacing is given by

$$P(t) = \sum_{i=1}^{n} \phi_{iM}(t)\, P_i \tag{1.7}$$

where $P_i \epsilon R^3$, $t\epsilon(-\infty,\infty)$, $n \geq M$ and

$$\phi_{iM}(t) = \frac{1}{(M-1)!} \sum_{j=0}^{M} (-1)^j \binom{M}{j} (s-j-i)_+^{M-1} \qquad (1.8)$$

where

$$(s - (j+i))_+^M = \begin{cases} (s-j-i)^M \text{ for } s \geq (i+j) \\ 0 \quad \text{otherwise} \end{cases} \qquad (1.9)$$

As with Bézier's method, the surface equation is given as the tensor product

$$P(u,v) = \sum_{j=1}^{n} \phi_{jN}(u) \left( \sum_{i=1}^{m} \phi_{iM}(v)\, P_{ij} \right), \qquad (u,v) \epsilon R^2$$

$n \geq N$, $m \geq M$ and $\phi_i^M$ represents the ith B-spline basis function of degree (M-1), as in (1.8). An important property of the B-spline basis of degree (M-1) is that each basis function is already a piecewise polynomial of continuity $C^{M-2}$, thus the designer does not, in general, have to explicitly join segments to form the compound curve (surface), although the piecewise nature of the curve (surface) is implicit in the basis. Figures 1.3 and 1.4 show typical B-spline curves, while Figure 1.5 shows a B-spline surface and the associated rectilinear network.

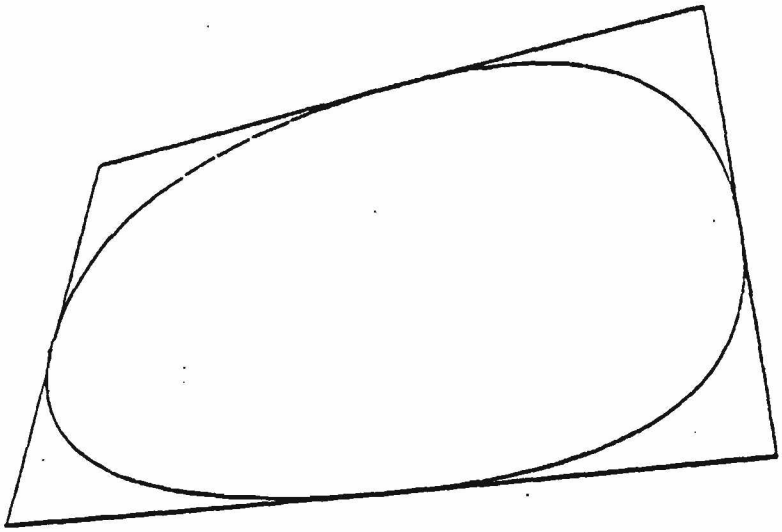

Figure 1.3 Closed quadratic B-spline curve and associated polygon.

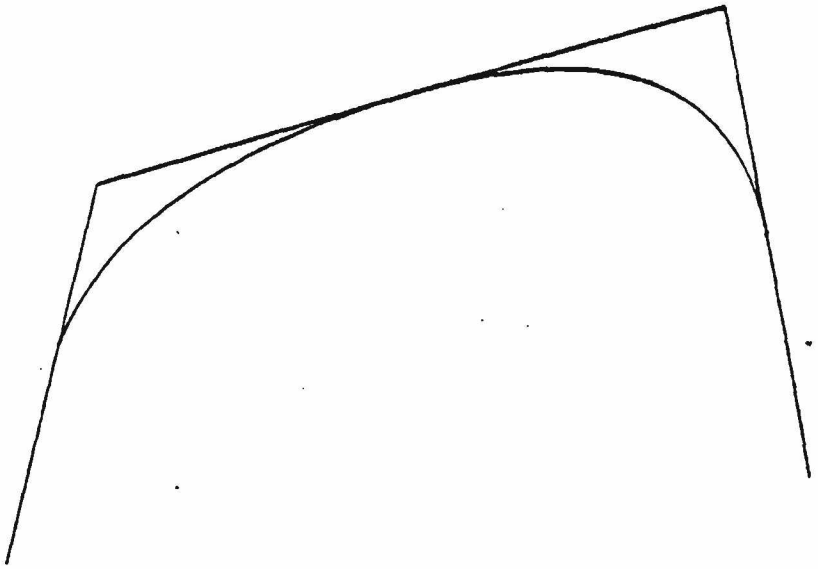

Figure 1.4 Open quadratic B-spline curve and associated polygon.

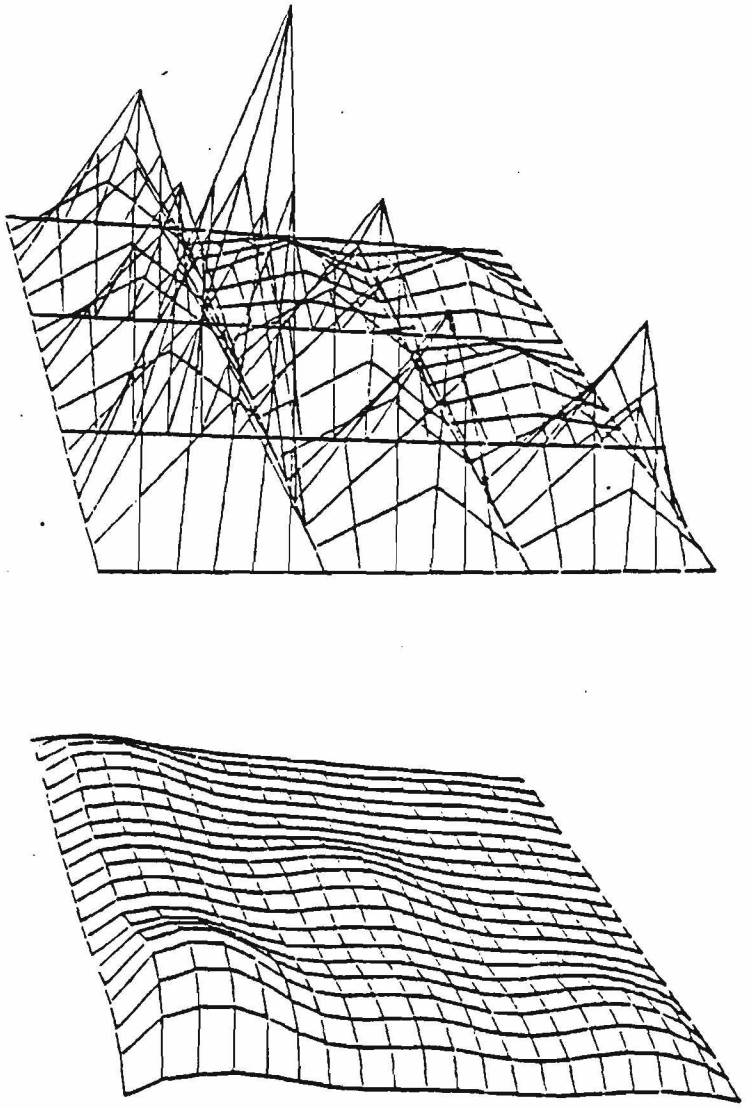

Figure 1.5 Bicubic B-spline surface and associated rectilinear network of points.

We noted earlier that interpolation methods seem ill-suited for ab initio design in that it may be difficult to control interactively the occurrence of kinks and oscillations in the design, while it appears the variation-diminishing methods of Bézier and Riesenfeld perform well. The mathematical statement of these observations is captured in the following exclusion theorem due to Schoenberg [8].

Theorem 1.1. Let $L(f)$ be a linear transformation defined for all continuous functions $f(x)$ on $[0,1]$, $L(f) \neq f$ for some $f$. Then if $L(a + bx + cx^2) = a + bx + cx^2$, for all $a, b, c \in R$ (reals), then $L$ cannot be variation-diminishing.

From the point of view of ab initio design then, techniques of approximation which have a high degree of polynomial "reproductive power," sometimes called polynomial precisions, cannot be variation-diminishing, and thus there is no guarantee that you can "control" oscillations during manipulation of the vertices of the defining polygonal network. Since most polynomial and polynomial spline techniques of interpolation are precise for more than linear polynomials, i.e., they reproduce polynomials of degree two and higher, they cannot be variation-diminishing. Thus Schoenberg's theorem establishes a firm foundation for approach to CAGD. If the problem is to fit hard constraints, then interpolation methods would seem superior to variation-diminishing methods, while variation-diminishing methods appear superior for ab initio design.

Although vector-valued B-spline approximation offers an attractive generalization of Bézier's methods for curve and surface design, both techniques possess inherent weaknesses. The extension from curves to surfaces requires a rectilinear network of points. Thus a designer is

restricted to rectilinear data as control points. There is no simple way of adding and deleting points within this restrictive topology. In general, the designer must delete or add a whole row and column of points to maintain the rectilinear topology of the data. This is clearly unacceptable for it would generate global changes to the surface. Further, Bernstein and high degree (continuity) B-spline approximation techniques have low "reproductive" power in the sense that the curves are poor approximations to the shape of the defining polygon [5]. The ability to control interactively the amount of local fit to the polygon while maintaining a high continuity class would be an important feature of an ab initio design system.

This paper is an investigation into the development of mathematical models which retain the valuable variation-diminishing and convex hull properties of Bernstein and B-spline approximation, while providing improvements in the interface to the user for ab initio design. A family of prototypes which includes the Bernstein and Bézier method is presented. Specifically, methods of assigning weights to each vertex which will control the amount of local fit to the polygon or polygonal net are provided. Thus the designer can easily cause "pseudo-cusps" and "pseudo-flats" by manipulating the "weights" at each vertex. Further, the generalization from curves to surfaces can be done with rectilinear data or triangular data. Therefore, the difficulty involved with the addition and deletion of points in a rectilinear network can be avoided. The feasibility and utility of the newly constructed models for interactive design are demonstrated on an experimental system for curve and system design.

A unifying theory, based on total positivity [3], is developed here for the construction of such models. Total positivity is a concept that has played an important role in various mathematical sciences. The application of the theory of total positivity to CAGD has resulted not only in new techniques for synthesizing shape operators for CAGD but also has provided general methods for analyzing the Bernstein-Bézier and B-spline methods previously discussed. Section II provides a general discussion of the representative and manipulation rules for totally positive functions and their interrelationships. In Section III we apply the theory of total positivity to the analysis of some existing methods of ab initio design and then proceed to the construction of new shape operators for curve and surface design. In Section IV we discuss the utility of these methods, as demonstrated on the curve and surface design system.

## II. TOTAL POSITIVITY--A REVIEW

### Total Positivity and the Variation-Diminishing Property

In Section I we established the usefulness of variation-diminishing methods for ab initio design. Due to the importance of Schoenberg's exclusion theorem (1.1) in this connection, we restate and prove it here.

Theorem 2.1 [8]. Let $T(f)$ be a linear transformation defined for all $f(x)$ continuous on an interval $[a,b]$, where $g(x) = T(f)$ is itself a continuous function in $[a,b]$, with the following properties:

1. $T(c + dx) = c + dx$, for all $c, d \in$ reals.
2. $V[T(f)] \leq V[f]$.
3. There exists $f$ such that $T(f) \neq f$, i.e., $T$ is not the identity transformation.

Then there exist real numbers $\alpha, \beta, \gamma$ such that

$$T(\alpha + \beta x + \gamma x^2) \neq \alpha + \beta x + \gamma x^2,$$

for $x \in [a,b]$.

Proof: Assume the converse, that $g(x) = T(f)$ *does* preserve quadratic polynomials, i.e., $T(x^2) = x^2$. Let $a < x_0 < b$. We will show that $g(x_0) = f(x_0)$, which implies $T$ must be the identity transformation, in contradiction of (3) above.

Assume, for instance, that

$$g(x_0) > f(x_0)$$

(A similar argument can be made for $g(x_0) < f(x_0)$). Since

$$T(f(x) - a - bx - cx^2) = g(x) - a - bx - cx^2$$

by the variation-diminishing property we have

$$V[g(x) - a - bx - cx^2] \leq V[f(x) - a - bx - cx^2]. \tag{2.1}$$

That is, the graph of $g(x)$ crosses the graph of any quadratic polynomial no more often that the graph of $f(x)$ does. But for $\alpha$ such that

$$f(x_0) < \alpha < g(x_0) ,$$

the parabola $y = \alpha + \beta(x - x_0)^2$ enjoys, for sufficiently large positive $\beta$, the following properties:

(i) It is crossed by $y = g(x)$ at least twice.

(ii) It is not crossed by $y = f(x)$.

(Consider the zeros of $g(x) - (\alpha + \beta(x - x_0)^2$ and $f(x) - (\alpha + \beta(x - x_0)^2$.) These conclusions contradict the variation-diminishing property (2.1) and the theorem is established.

We have already alluded to the strong interrelationship between variation-diminishing methods and totally positive functions. This relationship is established in the following theorem due to Karlin [3].

Theorem 2.2 [3]. Let $K(x,y)$ be a function of two variables $x \in X$ and $y \in Y$, where $X$ and $Y$ are linearly ordered subsets of the real line and consider the transformation

$$g(x) = T(f)(x) = \int_Y K(x,y)\, f(y)\, dy, \tag{2.2}$$

where f and K are bounded and Riemann integrable on Y. When Y consists of a discrete set, we interpret the integral as a sum.

If K is totally positive of order r, then

$$V[g] \leq V[f] \quad \text{provided that} \quad V[f] \leq r-1. \tag{2.3}$$

Note that equation (1.1) is a vector-valued form of (2.2) where Y is a discrete set. The proof of Theorem 2.2 follows as a corollary to the statement that, for a totally positive matrix A and any vector $\bar{x}$ (of proper length), $A\bar{x}$ has no more sign changes than does $\bar{x}$. Before we can prove either of these assertions a general discussion of the representation and manipulation rules for totally positive functions is necessary, after which the proof of theorem 2.2 will be resumed.

Total Positivity

Definition 2.1. A real function K(x,y) of two variables ranging over linearly ordered sets X and Y, respectively is said to be totally positive of order r (abbr. $TP_r$) if for all sequences $x_1 < x_2 < \ldots < x_m$, $y_1 < y_2 < \ldots < y_m$, $x_i \in X$, $y_i \in Y$, $1 \leq m \leq r$ we have the inequalities

$$K\begin{pmatrix} x_1, x_2, \ldots, x_m \\ y_1, y_2, \ldots, y_m \end{pmatrix} = \begin{vmatrix} K(x_1,y_1), \ldots, K(x_1, y_m) \\ \vdots \qquad\qquad \vdots \\ K(x_m,y_1), \ldots, K(x_m, y_m) \end{vmatrix} \geq 0. \tag{2.4}$$

If strict inequality holds in 2.4, we say K(x,y) is strictly totally positive of order r (abbreviated $STP_r$). The subscript is normally omitted when a function is (strictly) totally positive of all orders.

Note that total positivity implies positivity. As examples of totally positive functions, we have:

(a) $e^{xy}$, $X \times Y \equiv R^2$ is STP

(b) $K_m(x,i) = \binom{m}{i} x^i (1-x)^{m-i}$ is STP of order m+1 on $i \in I = \{0,1, \ldots, m\}$ and $x \in (0,1)$.

(c) The square n x n matrix $A = [a(i,j) = \delta_{ij}]$ is TP.

The verification of (c) is trivial, while (b) will be shown to follow immediately from (a). Note that the function K(x,i) in (b) represents the Bernstein basis of degree m-1, and the proof of (b) in conjunction with Theorem 2.2 would be sufficient to show that Bernstein-Bézier methods are variation-diminishing. We provide the proofs for (a) and (b) below.

Theorem 2.3. $e^{xy}$ is STP on $X \times Y = R^2$.

Proof: We first show that

$$e\begin{pmatrix} x_1, \ldots, x_n \\ y_1, \ldots, y_n \end{pmatrix} = \begin{vmatrix} e^{x_1 y_1} & \cdots & e^{x_1 y_n} \\ \vdots & & \vdots \\ e^{x_n y_1} & \cdots & e^{x_n y_n} \end{vmatrix} \neq 0$$

for $x_1, x_2, \ldots x_n$ distinct in X and $y_1, y_2, \ldots, y_n$ distinct in Y. This fact follows immediately if we can show that any exponential polynomial of the form

$$\sum_{i=1}^{n} a_i e^{y_i x} \tag{2.5}$$

where

$$\sum_{i=1}^{n} a_i^2 > 0$$

has at most n-1 zeros.

Denoting the number of distinct zeros of a function by $Z(f)$, we prove (2.5) as a lemma.

Lemma 2.1. Let $y_i$, $i=1, \ldots, n$, be a set of distinct real numbers. Then

$$Z\left(\sum_{i=1}^{n} a_i e^{y_i x}\right) \leq n-1$$

where

$$\sum_{i=1}^{n} a_i^2 > 0 .$$

Proof: (by mathematical induction) Let $n = 1$. Then clearly $a_i e^{y_i x}$ has no zeros.

Now assume the hypothesis holds for $n = k-1$, i.e.,

$$Z\left(\sum_{i=1}^{k-1} a_i e^{y_i x}\right) \leq k-2$$

for all $a_i \varepsilon R$ such that

$$\sum_{i=1}^{k-1} a_i^2 > 0 . \tag{2.6}$$

We must show that

$$Z\left(\sum_{i=1}^{k} a_i e^{y_i x}\right) \leq k-1$$

for all $a_i \varepsilon R$ such that

$$\sum_{i=1}^{k} a_i^2 > 0. \tag{2.7}$$

Assume without loss of generality $a_1 \neq 0$. Since $e^{-y_1 x}$ has no zeros, we have

$$Z\left(e^{-y_1 x} \sum_{i=1}^{k} a_i e^{y_i x}\right) = Z\left(\sum_{i=1}^{k} a_i e^{y_i x}\right) \tag{2.8}$$

Differentiating $e^{-y_1 x} \sum_{i=1}^{k} a_i e^{y_i x}$ with respect to x, we get

$$\sum_{i=2}^{k} a_i' e^{y_i' x}$$

where

$$y_i' = y_i - y_1$$

and

$$a_i' = a_i(y_i - y_1) \tag{2.9}$$

which, by our induction assumption, has at most k-2 zeros. It follows from Rolle's theorem that a function can have at most one more distinct zero than its derivative, therefore,

$$Z\left(\sum_{i=1}^{k} a_i e^{y_i x}\right) \leq k-1 \tag{2.10}$$

Now we show that for $x_1 < x_2 < \ldots < x_n$, $x_i \varepsilon X$ and $y_1 < y_2 < \ldots < y_n$, $y_i \varepsilon Y$, the determinant

$$e\begin{pmatrix} x_1, \ldots, x_n \\ y_1, \ldots, y_n \end{pmatrix}$$

cannot achieve both negative and positive values.

Lemma 2.2. Given $x_1 < x_2 < \ldots < x_n$, $x_i \varepsilon X$ and $y_1 < y_2 < \ldots < y_n$

$y_i \varepsilon Y$, where $X \times Y = R^2$, we have

$$e\begin{pmatrix} x_1, \ldots, x_n \\ y_1, \ldots, y_n \end{pmatrix} = \begin{vmatrix} e^{x_1 y_1} & \ldots & e^{x_1 y_n} \\ \vdots & & \vdots \\ e^{x_n y_1} & \ldots & e^{x_n y_n} \end{vmatrix} \tag{2.11}$$

is of one strict sign.

Proof: This will be proved by contradiction. Fixing $y_1 < y_2 < \ldots < y_n$, $y_i \varepsilon Y$, let $x_1 < x_2 < \ldots < x_n$, and $z_1 < z_2 < \ldots < z_n$, $x_i, z_i$ in $X$ be such that

$$e\begin{pmatrix} x_1, \ldots, x_n \\ y_1, \ldots, y_n \end{pmatrix} > 0 > e\begin{pmatrix} z_1, \ldots, z_n \\ y_1, \ldots, y_n \end{pmatrix} \tag{2.12}$$

Since $e^x$ is a continuous real function, we have for some $\lambda \varepsilon (0,1)$ the determinant

$$D(\lambda) = e\begin{pmatrix} y_1, & \ldots & , y_n \\ \lambda x_1 + (1-\lambda)\, z_1, & \ldots, & \lambda x_n + (1-\lambda)\, z_n \end{pmatrix} \equiv 0$$

By (2.5) this is impossible for distinct $x_i$ and $z_i$. Therefore, there exist $i,j$, $i \neq j$, such that

$$\lambda x_i + (1-\lambda)\, z_i = \lambda x_j + (1-\lambda)\, z_j$$

which implies

$$0 = \frac{\lambda}{1-\lambda} = \frac{z_j - z_i}{x_i - x_j} \quad ,$$

contradicting the strictly increasing nature of the $x_i$ and $z_i$.

Now all that remains to complete the proof of Theorem 2.3 is to exhibit a set of increasing $x_i$ and increasing $z_i$, for which the determinant in (2.11) is positive. If we let $x_i = i$, $i = 0, 1, \ldots, n$, and $y_i = \ln z_i$ where $z_i$ is a strictly increasing sequence of positive reals numbers, then (2.11) specializes to

$$\begin{vmatrix} e^{x_0 y_0} & e^{x_0 y_n} \\ \\ e^{x_n y_0} & e^{x_n y_n} \end{vmatrix} = \begin{vmatrix} 1 & z_0^1 & \cdots & z_0^n \\ \vdots & \vdots & & \vdots \\ 1 & z_n^1 & & z_n^n \end{vmatrix}$$

$$= \prod_{i>j} (z_i - z_j) > 0, \tag{2.13}$$

since the determinant on the right hand side is the well-known Vandemonde determinant [15].

In order to prove the total positivity of the Bernstein basis, we need the following two composition rules for totally positive functions.

Theorem 2.4 [3]. Let $K(x,y)$ be $TP_r(STP_r)$ on X and Y and let $\phi(x)$ and $\psi(y)$ be nonnegative (positive) on X and Y, respectively. Then

$$L(x,y) = \phi(x)\, \psi(y)\, K(x,y) \tag{2.14}$$

is $TP_r(STP_r)$.

Proof: The conclusion (2.14) follows immediately from the fact that the determinant is a multilinear function of its rows or columns.

Theorem 2.5 [3]. Let $K(x,y)$ be $TP_r(STP_r)$ on X and Y and let $\phi(u)$ and $\psi(v)$, $u \in U$, $v \in V$ be strictly increasing functions mapping U and V into

X and Y, respectively. Then

$$L(u,v) = K[\phi(u), \psi(v)] \tag{2.15}$$

is $TP_r$($STP_r$), $u\in U$, $v\in V$.

Proof: The result (2.15) follows immediately from the obvious fact that $\phi$ and $\psi$ map increasing sequences into increasing sequences.

Theorem 2.6. The Bernstein basis $K_m(i,x) = \binom{m}{i} x^i(1-x)^{m-i}$ $i=0,1, \ldots, m$, $x\in(0,1)$ is $TP_{m+1}$ on $\{0,1, \ldots, m\}$ and $(0,1)$.

Proof: Rewriting $K_m(i,x) = \binom{m}{i} x^i(1-x)^{n-i}$ as

$$K_m(i,x) = \binom{m}{i} (1-x)^m e^{i\ln(\frac{x}{1-x})} , \tag{2.16}$$

the conclusion follows directly from Theorems 2.3, 2.4 and 2.5.

We will find the following composition formula for totally positive functions of fundamental importance in the development of shape operators for CAGD.

Theorem 2.7 [3]. Let $K(x,z)$ and $L(z,y)$ $x\in X$, $y\in Y$ and $z\in Z$ be Riemann integrable functions of z, where X, Y and Z are linearly ordered sets of the real line, and define

$$M(x,y) = \int_Z K(x,z)\, L(z,y)\, dz,$$

where again we interpret the integral as a sum when Z is discrete. Then, if $K(x,z)$ is $TP_r$ on $X \times Z$ and $L(z,y)$ is $TP_s$ on $Z \times Y$, then $M(x,y)$ is $TP_t$ on $X \times Y$, where $t = \min(r,s)$.

Proof: The conclusion follows easily from the generalized Cauchy - Binet formula for the composition of determinants given by

$$M\begin{pmatrix} x_1, x_2, \ldots, x_m \\ y_1, y_2, \ldots, y_m \end{pmatrix} =$$

$$\int \cdots \int_{a \le z_1 < \ldots < z_m \le b} K\begin{pmatrix} x_1, \ldots, x_m \\ z_1, \ldots, z_m \end{pmatrix} L\begin{pmatrix} z_1, \ldots, z_m \\ y_1, \ldots, y_m \end{pmatrix} dz_1, \ldots, dz_m \tag{2.17}$$

where M, K and L are defined as above and we assume Z = [a,b] for simplicity. Formula (2.17) is a direct extension of the Cauchy-Binet formula [17] for matrices, which we restate below.

Theorem 2.8 [17]. Let A, B and C denote matrices of real numbers or orders n x m, n x k, and k x m, respectively. If A = BC, then

$$A\begin{pmatrix} i_1, i_2, \ldots, i_p \\ j_1, j_2, \ldots, j_p \end{pmatrix} = \sum_{1 \le \alpha_1 < \ldots < \alpha_n \le n} B\begin{pmatrix} i_1, i_2, \ldots, i_p \\ \alpha_1, \alpha_2, \ldots, \alpha_p \end{pmatrix} C\begin{pmatrix} \alpha_1, \alpha_2, \ldots, \alpha_p \\ j_1, j_2, \ldots, j_p \end{pmatrix} \tag{2.18}$$

where the lefthand term stands for the minor of A involving rows $i_1, i_2, \ldots, i_p$ and $j_1, j_2, \ldots, j_p$, respectively.

Proof: The conclusion will follow if we can establish the corresponding result for any square matrix which is the product of two rectangular matrices. So, suppose that a square matrix $C = \|c_{ij}\|_1^m$ is the product of two rectangular matrices $A = \|a_{ik}\|$ and $B = \|b_{kj}\|$ of dimension m x n and n x m, respectively. That is,

$$c_{ij} = \sum_{\alpha=1}^{n} a_{i\alpha} b_{\alpha j} \qquad (i,j = 1,2, \ldots, m). \tag{2.19}$$

By (2.19) the determinant of C can be represented in the form

$$\begin{vmatrix} c_{11} & \cdots & c_{1m} \\ \cdots & \cdots & \cdots \\ c_{m1} & \cdots & c_{mm} \end{vmatrix} = \begin{vmatrix} \sum_{\alpha_1=1}^{n} a_{1\alpha_1} b_{\alpha_1 1} & \cdots & \sum_{\alpha_m=1}^{n} a_{1\alpha_m} b_{\alpha_m m} \\ \cdots & \cdots & \cdots \\ \sum_{\alpha_1=1}^{n} a_{m\alpha_1} b_{\alpha_1 1} & \cdots & \sum_{\alpha_m=1}^{n} a_{m\alpha_m} b_{\alpha_m m} \end{vmatrix}$$

$$= \sum_{\alpha_1, \ldots, \alpha_m=1}^{n} \begin{vmatrix} a_{1\alpha_1} b_{\alpha_1 1} & \cdots & a_{1\alpha_m} b_{\alpha_m m} \\ \cdots & \cdots & \cdots \\ a_{m\alpha_1} b_{\alpha_1 1} & \cdots & a_{m\alpha_m} b_{\alpha_m m} \end{vmatrix}$$

$$= \sum_{\alpha_1, \ldots, \alpha_m=1}^{n} A\begin{pmatrix} 1 & 2 & \ldots & m \\ \alpha_1 & \alpha_2 & \ldots & \alpha_m \end{pmatrix} b_{\alpha_1 1} b_{\alpha_2 2} \cdots b_{\alpha_m m} \tag{2.20}$$

If $m > n$, then among the numbers $\alpha_1, \alpha_2, \ldots, \alpha_m$ there are always at least two that are equal, so that every summand on the right-hand side of (2.20) is zero. Hence in this case $C = 0$.

Now let $m \leq n$. Then in the sum on the right-hand side of (2.20) all those summands will be zero in which at least two of the subscripts $\alpha_1, \alpha_2, \ldots, \alpha_m$ are equal. All the remaining summands of (2.20) can

be split into groups of m! terms each by combining into one group those summands that differ from each other only in the order of the subscripts $\alpha_1, \alpha_2, \ldots, \alpha_m$ (so that within each such group the subscripts $\alpha_1, \alpha_2, \ldots, \alpha_m$ have one and the same set of values). Now within one such group the sum of the corresponding terms is

$$\Sigma\, \varepsilon(\alpha_1, \alpha_2, \ldots, \alpha_m)\, A\begin{pmatrix} 1 & 2 & \ldots & m \\ k_1 & k_2 & \ldots & k_m \end{pmatrix} b_{\alpha_1} \quad \ldots \quad b_{\alpha_m m} =$$

$$= A\begin{pmatrix} 1 & 2 & \ldots & m \\ k_1 & k_2 & \ldots & k_m \end{pmatrix} \Sigma\, \varepsilon(\alpha_1, \alpha_2, \ldots, \alpha_m)\, b_{\alpha_1 1} \quad \ldots \quad b_{\alpha_m m}$$

$$= A\begin{pmatrix} 1 & 2 & \ldots & m \\ k_1 & k_2 & \ldots & k_m \end{pmatrix} B\begin{pmatrix} k_1 & k_2 & \ldots & k_m \\ 1 & 2 & \ldots & m \end{pmatrix},$$

where $k_1 < k_2 < \ldots < k_m$ is the normal order of the subscripts $\alpha_1$, $\alpha_2, \ldots, \alpha_m$ and $\varepsilon(\alpha_1, \alpha_2, \ldots, \alpha_m) = (-1)^N$, where N is the number of transpositions of the indices needed to put the permutation $\alpha_1, \alpha_2, \ldots \alpha_m$ into normal order.

Hence (2.19) implies (2.18).

Variation-Diminishing Transformations

We are now in a position to prove Theorem 2.2. As noted earlier, Theorem 2.2 is a direct consequence of the corresponding theorem for matrices, which we establish as Theorem 2.9 below. The proof given here is essentially that given by Schoenberg [16,21]. In the follow-

ing definitions and theorems, sequences and matrices will often be represented with the functional notation x(i), a(i,j), emphasizing the fact that we are dealing with functions over discrete sets.

Definition 2.2. Given a function f(t) defined in I, an ordered set of the real line. We define the variation of f over I as

$$V(f) = \sup V[f(t_1), f(t_2), \ldots f(t_m)] ,$$

where the supremum is extended over all sets $t_1 < t_2 \ldots < t_m (t_i \varepsilon I)$, m arbitrary but finite and V[x(1), x(2), ..., x(m)] is the number of sign changes of the indicated sequence, zero terms discarded. By convention V[0,0, ..., 0] = -1.

Definition 2.3. Given a sequence X = [x(1), x(2), ..., x(n)] of real numbers, we define N[x(1), x(2), ..., x(n)] as the number of x(i)'s which are nonzero.

Definition 2.4. A real m x n matrix A = [a(i,j)] is said to be variation-diminishing if

$$y(i) = \sum_{j=1}^{n} a(i,j)\, x(j), \qquad (i = 1,2, \ldots, m)$$

implies that

$$V[y(1), y(2), \ldots, y(m)] \leq V[x(1), \ldots, x(n)] .$$

Lemma 2.1. Consider the following matrix operations:

(i) Multiplication of a row or column by a non-negative constant;

(ii) Addition of a row or column to an adjacent row or column;

(iii) Omission of a row or column.

These operations when applied to a TP matrix yield a TP matrix.

Proof: (iii) is trivial, while (i) follows from Theorem 2.4. (ii) is evident since any minor in the new matrix would be the sum of minors in the old.

Lemma 2.2. Let $X = [x(i), i = 1,2,\ldots,m]$ be a sequence of real numbers. If for all

$$i_0 < i_1 < \ldots < i_p, \quad p < m,$$

we have

$$V[x(i_0), \ldots, x(i_p)] \leq p - 1, \tag{2.21}$$

then

$$V[x(1), \ldots, x(m)] \leq p - 1 .$$

Proof: Assume the converse, that there exists

$$i_0 < i_{1'} < \ldots < i_{p+1}$$

such that

$$V[x(i_0), \ldots, x(i_{p+1})] > p - 1$$

then

$$V[x(i_0), \ldots, x(i_{p+1})] = p, \tag{2.22}$$

since all other cases immediately contradict (2.21). But to preserve (2.21) for all subsequences of $[x(i_0), \ldots, x(i_{p+1})]$ the sequence must alternate in sign. However, we than have $V[x(i_0), \ldots, x(i_{p+1})] \geq p+1$ a contradiction to (2.22). From here a simple induction argument on the length of the subsequence yields the desired result.

Lemma 2.3. Let A be a TP, real m x n matrix of rank p and define

$$y(i) = \sum_{j=1}^{n} a(i,j)\, x(j), \qquad i = 1,2, \ldots, m$$

where $X = [x(j),\ j = 1,2, \ldots, n]$ is some real sequence. Let A' be the submatrix of A consisting of the rows $i_0, i_1, \ldots, i_p$ and define

$$\alpha(K) = A\begin{pmatrix} i_0, & \ldots, & i_{k-1}, & i_{k+1}, & \ldots, & i_p \\ j_1, & & \ldots & & & j_p \end{pmatrix}$$

for some selection of columns $j_1, j_2, \ldots, j_p$. Then

$$\sum_{k=0}^{p} (-1)^k\, \alpha(k)\, y(i_k) = 0 .$$

Proof: Since

$$y(i_k) = \sum_{j=1}^{n} a(i_k,j)\, x(j) ,$$

we have

$$\sum_{k=0}^{p} (-1)^k\, \alpha(k)\, y(i_k)$$

$$= \sum_{k=0}^{p} (-1)^k\, \alpha(k) \sum_{j=1}^{n} a(i_k,j)\, x(j)$$

$$= \sum_{j=1}^{n} (-1)^j\, x(j) \left( \sum_{k=0}^{p} (-1)^{k+j}\, \alpha(k)\, a(i_k,j) \right) . \qquad (2.23)$$

But for j not a member of $[j_1, j_2, \ldots, j_p]$ the inner sum is the

expansion of some minor of A of order p+1 and thus is zero. Similarly, for j a member of $[j_1, j_2, \ldots, j_p]$, the inner sum corresponds to the expansion of the determinant of some matrix which possesses two identical columns and again the sum is zero. Therefore, (2.23) must be identically zero.

Theorem 2.9 [21]. If a real m x n matrix A = [a(i,j)] is TP, then it is variation-diminishing.

Proof: We first assert that it is sufficient to demonstrate this under the assumption that

$$V[x(1),\ldots,x(n)] = N[x(1),\ldots,x(n)] - 1. \qquad (2.24)$$

For if any x(k) = 0 we compress X and A by

$$x'(j) = \begin{cases} x(j) & j < k \\ x(j+1) & j \geq k \end{cases} \qquad (j = 1, \ldots, n-1)$$

$$a'(i,j) = \begin{cases} a(i,j) & j < k \\ a(i,j+1) & j \geq k \end{cases} \qquad (j = 1, \ldots, n-1)$$

We then have y(i) defined by

$$y(i) = \sum_{j=1}^{n-1} a'(i,j)\, x'(j) .$$

The matrix [a'(i,j)] is TP by Lemma 2.1 and if we can show

$$V[y(1), \ldots, y(m)] \leq V[x'(1), \ldots, x'(n-1)],$$

then since

$$V[x'(1), \ldots, x'(n-1)] = V[x(1), \ldots, x(n)]$$

it will follow that

$$V[y(1), \ldots, y(m)] \leq V[x(1), \ldots, x(n)] .$$

So we can assume no $x_k$ is zero.

Next, we can assume x(k) and x(k+1) are of opposite sign. If this is not true, then there exists $\lambda > 0$ such that $x(k+1) = \lambda x(k)$. If we compress X and A by

$$x'(j) = \begin{cases} x(j) & j < k \\ x(j+1) & j \geq k \end{cases} \qquad (j = 1, \ldots, n-1)$$

$$a'(i,j) = \begin{cases} a(i,j) & j < k \\ a(i,k) + \lambda a(i,k+1) & j = k \\ a(i,j+1) & j > k \end{cases} \qquad (j = 1, \ldots, n-1) \tag{2.25}$$

then we have

$$y(i) = \sum_{j=1}^{n-1} a'(i,j)\, x'(j) .$$

The matrix [a'(i,j)] is TP by Lemma 2.1 and if we can show that

$$V[y(1), \ldots, y(m)] \leq V[x'(1), \ldots, x'(n-1)],$$

then since

$$V[x'(1), \ldots, x'(n-1)] = V[x(1), \ldots, x(n)],$$

it will follow that

$$V[y(1), \ldots, y(m)] \leq V[x(1), \ldots, x(n)]$$

We have shown that we may assume (2.24). We complete the proof then with the following lemma.

Lemma 2.4. [21]. If a real matrix $A = [a(i,j)]$, $(i=1, \ldots, m, j=1, \ldots, n)$ is TP, then

$$V[y(1), \ldots, y(m)] \leq N[x(1), \ldots, x(n)] - 1$$

From above we may assume no $x_k$ is zero. That is, we may assume $N[x(1), \ldots, x(n)] = n$. Since $r(A)$ (rank of A) $\leq n$, we would be done if we can show that

$$V[y(1), \ldots, y(m)] \leq r(A) - 1,$$

We prove this result, in turn, by the following lemma.

Lemma 2.5. [21]. If a real $m \times n$ matrix $A = [a(i,j)]$ is TP, then

$$V[y(1), \ldots, y(m)] \leq r(A) - 1$$

Proof: We proceed by induction on $r(A)$. The result is clearly true if $r(A) = 0$ or 1. Suppose that it has been established for $r(A) = 0, 1, \ldots, p-1$. We must show that it is true for $p$.

We may suppose that $m > p$, since if $m = p$, then $V[y(1),\ldots,y(m)] \leq p-1$ trivially. By Lemma 2.2 it is enough to show that, if $1 \leq i_0 < i, < \ldots < i_p \leq m$, then $V[y(i_0), \ldots, y(i_p)] \leq p-1$. If $A'$ is the submatrix of $A$, consisting of the rows $i_0, i_1, \ldots, i_p$, then $A'$ is TP. There are two cases:

$r(A') < p$ and $r(A') = p$. If $r(A') < p$, then

$$V[y(i_0),\ldots,y(i_p)] \leq r(A') - 1 < p - 1$$

by our induction assumption.

So we assume $r(A') = p$. Let $j_1,\ldots,j_p$ be a selection of columns of $A'$ and set

$$\alpha(k) = A\begin{pmatrix} i_0,\ldots,i_{k-1}, i_{k+1},\ldots,i_p \\ j_1,\ldots,j_p \end{pmatrix}$$

for $0 \leq k \leq p$. Since $A'$ is TP, no two $\alpha(k)$'s are of opposite sign and because $r(A') = p$, we can so choose $j_1,\ldots,j_p$ such that not all $\alpha(k)$'s are zero. We have from Lemma 2.3

$$\alpha(0)\, y(i_0) - \alpha(1)\, y(i_1) + \ldots + (-1)^p \alpha(p)\, y(i_p) = 0 \tag{2.26}$$

It is easily seen that $V[y(i_0), y(i_1),\ldots,y(i_p)] = p$ is not compatible with (2.26) so that $V[y(i_0),\ldots,y(i_p)] \leq p-1$. Thus we have proved Lemma 2.5, thereby proving Lemma 2.4, thereby proving the theorem.

Theorem 2.2 is established in the following three corollaries of Theorem 2.9.

Corollary 2.9.1. Let $[y(1),\ldots,y(m)]$, $[x(1),\ldots,x(n)]$ and $A$ be given as in (2.20) where now we assume

$$V[x(1),\ldots,x(n)] \leq r-1, \quad r \leq n-1,$$

and A is $TP_r$. Then

$$V[y(1),\ldots,y(m)] \leq V[x(1),\ldots,x(n)].$$

Proof: Using the techniques given in (2.24) and (2.25) we can compress A and $[x(1),\ldots,x(n)]$ such that

$$y(i) = \sum_{j=1}^{N} a'(i,j)\, x'(j)$$

where $V[x'(1),\ldots,x'(N)] = N-1 = V[x(1),\ldots,x(n)]$ and $A' = [a'(i,j)]$ is a TP mxr matrix. The conclusion now follows from Theorem 2.9.

Corollary 2.9.2. Let $f(x) = \sum_{i=1}^{n} \phi(i,x)\, a_i$. If $\{\phi(i,x),\ i=1,\ldots,n\}$ is TP on $x = [a,b]$ and $I = \{1,2,\ldots,n\}$, then $\{\phi(i,x),\ i=1,\ldots,n\}$ is variation-diminishing.

Proof: We must show

$$V(f(x)) \leq V[a_1,\ldots,a_n] .$$

That is, for any finite sequence of points $x_j$, $j=1,\ldots,m$ we must show

$$V[f(x_1),\ldots f(x_m)] \leq V[a_1,\ldots,a_n] \quad ,$$

where

$$f(x_j) = \sum_{i=1}^{n} \phi(i,x_j)\, a_i$$

for j=1,...,m. But this is exactly the result of the above theorem and we're done.

Corollary 2.9.3. Let $\phi(x,y)$ and $g(y)$ be bounded and continuous functions on [a,b] x [c,d] and [c,d], respectively. Let

$$f(x) = \int_c^d \phi(x,y)\, g(y)\, dy \quad .$$

Then, if $\phi(x,y)$ is TP, we have $V[f(x)] \leq V[g(y)]$.

Proof: We must show $V[f(x_1),\ldots,f(x_m)] \leq V[g(y_1),\ldots,g(y_n)]$ $\forall$ $x_i \in [a,b]$, $y_j \in [c,d]$, m and n fixed but arbitrary.

We approximate $f(x_j)$ arbitrarily closely by the Riemann sum

$$f_1(x_i) = h \sum_{j=1}^{n} \phi(x_i,y_j)\, g(y_j),$$

such that $\operatorname{sgn}(f_1(x_i)) = \operatorname{sgn}(f(x_i))$ for i=1,...,m. The conclusion now follows from Corollary 2.9.1.

Corollary 2.9.3 [3]. If in Theorem 2.9 we have, in fact ,

$$V[y(1),\ldots,y(m)] = V[x(1),\ldots,x(n)],$$

for A STP, then {y(i), i=1, i=1,...,m} and {x(j), j-1,...,n} exhibit the same arrangement of signs.

Proof: From the above arguments we know it is enough to assume $V[x(1),\ldots,x(n)] = n-1$. Under this assumption we need only show that the first component of [x(1),...,x(n)] has the same sign as the first nonzero component of [y(1),...,y(m)]. Choose $i_1 < i_2 < \ldots < i_n$ such that $y_{i_v}\, y_{i_{v+1}} < 0$. Since $y_{i_v} = \sum_{j=1}^{n} a_{i_v j}\, x_j$ (v = 1,2,...,n),

we have

$$x_1 = \frac{\begin{vmatrix} y_{i_1} & a_{i_1 2} & \cdots & a_{i_1 n} \\ y_{1_2} & a_{1_2 2} & \cdots & a_{i_2 n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ y_{i_n} & a_{i_n 2} & \cdots & a_{i_n n} \end{vmatrix}}{A\begin{pmatrix} i_1, i_2, \ldots i_n \\ 1,\ 2,\ \ldots\ n \end{pmatrix}}$$

$$= \frac{\sum_{v=1}^{n} (-1)^{v+1} A\begin{pmatrix} i_1, i_2, \ldots, i_{v-1}, i_{v+1}, \ldots, i_n \\ 2,3,\ldots \quad \ldots,n \end{pmatrix}}{A\begin{pmatrix} i_1, i_2, \ldots, i_n \\ 1,2,\ldots,n \end{pmatrix}}$$

Now the minors of A are strictly positive and the values $(-1)^{v+1} y_{i_v}$ $(v = 1,2,\ldots,m)$ maintain the same sign. Thus $\operatorname{sgn}(x_1) = \operatorname{sgn}(y_{i_1})$.

Theorem 2.2 combined with Theorem 2.6 provide enough power to conclude that scalar-valued Bernstein approximation is variation-diminishing. However, as we established in Section I, we are interested in vector-valued approximation and transformation methods, thus we must provide some extension of the scalar-valued theory to vector-valued mathematics. This is provided in the following definition and theorem due to W. W. Meyer [5].

Definition 2.5 [5]. A vector-valued approximation or transformation method is variation-diminishing if it is variation-diminishing as a scalar-valued method.

Theorem 2.10 [5]. If a vector-valued transformation is invariant under euclidean transformation, then no (hyper) plane is pierced more often by a vector-valued transformation than by the primitive curve itself.

Proof: Surely the theorem is true for any principal (hyper) plane $x(i) = 0$ because it is variation-diminishing in the particular coordinate function $x_i(t)$. But any (hyper) plane can be designated as a principal one by a euclidean transformation. Hence the result is true for all (hyper) planes.

It is not difficult to show that transformations of the form (1.1) are invariant under euclidean transformation if and only if

$$\sum_i \phi(i,x) = 1 \quad . \tag{2.27}$$

This fact does not form a formidable obstacle to us, since, given a TP basis $\phi(i,x)$, by Theorem 2.4 we can construct another TP basis

$$\phi'(i,x) = \phi(i,x) / \sum_j \phi(j,x) \tag{2.28}$$

such that

$$\sum_i \phi'(i,x) = 1 \tag{2.29}$$

We have established the useful fact that the Bézier method is variation-diminishing. In Section III we continue with the analysis

of other mathematical schemes in use or proposed for use in *ab initio* design. We then apply the theory of total positivity to the construction of a whole family of new models for CAGD.

## III. APPLICATION OF TOTAL POSITIVITY TO CAGD

Through the application of the theory of total positivity we have shown in Section II that Bernstein-Bézier methods are variation-diminishing. From Section I we know that widely used techniques, such as polynomial and spline interpolation, are not variation-diminishing. In this chapter we continue this type of analysis by applying the techniques developed in Section II to other methods in use, or proposed for use, in ab initio design.

### Piecewise Bernstein-Bézier Methods

In general, it is vector-valued piecewise Bernstein approximation, not Bernstein approximation, which is used in CAGD. As noted in Section I, shape is essentially non-analytic in nature, therefore, it is not surprising that we often find, when working interactively, that a curve segment is not sufficiently flexible to adopt a desired shape. We may then either increase the order of approximation (and thus that of the polygon) or the segment may be split into two or more segments. (see Figure 3.1). Curve splitting is simple mathematically and has the advantage computationally of retaining a reasonable (polynomial) order of approximation. Since the resulting curve is a piecewise Bernstein approximation, the question arises whether this basis is variation-diminishing as well. We establish this fact in Theorem 3.1, but first

we generalize the definition of the Bernstein operator to the range [a,b].

Definition 3.1 [15]. The Bernstein approximation of degree n to the polygon $P = [P_0,\ldots,P_n]$ on the interval [a,b] is given by

$$B_n[P;a,b] = \frac{1}{(b-a)^n} \sum_{k=0}^{n} P_i \binom{n}{k} (x-a)^k (b-x)^{n-k} \qquad (3.1)$$

Theorem 3.1. Let $B_n[P; 0,1]$ and $B_m[Q;1,2]$ be Bernstein approximations of degrees n and m to the polygons $P = [P_0,\ldots,P_n]$ on the (0,1) and $Q = [P_n,\ldots,P_{n+m}]$ on (1,2), respectively (see Figure 3.1). Then

$$V[B_n[P] + B_m[Q]] \leq V[P_0,\ldots,P_{n+m}] . \qquad (3.2)$$

Proof: (3.2) is easily seen to be true if the sign of $P_n$ equals the sign of $P_{n-1}$ or $P_{n+1}$, since then $V[P] + V[Q] = V[P_0,\ldots,P_{n+m}]$. Assume this is not the case, i.e., $P_n = 0$ and $P_{n-1}$ and $P_{n+1}$ are of opposite sign. Again (3.2) is true if either

$$V[B_n[P; 0,1]] < V[P] \qquad (3.3)$$

or

$$V[B_m[Q;1,2]]] < V[Q] , \qquad (3.4)$$

since $\max(V[P] + V[Q]) = V[P_0,\ldots,P_{n+m}] + 1$. So assume this is not the case, i.e., $V[B_n[P; 0,1]] = V[P]$ and $V[B_m[Q; 1,2]] = V[Q]$. But then by Corollary 2.9.3 $B_n[P; 0,1]$ and $B_m[Q; 1,2]$ must have the same arrangement of sign changes as P and Q, respectively. Thus (3.2) must hold in this case as well.

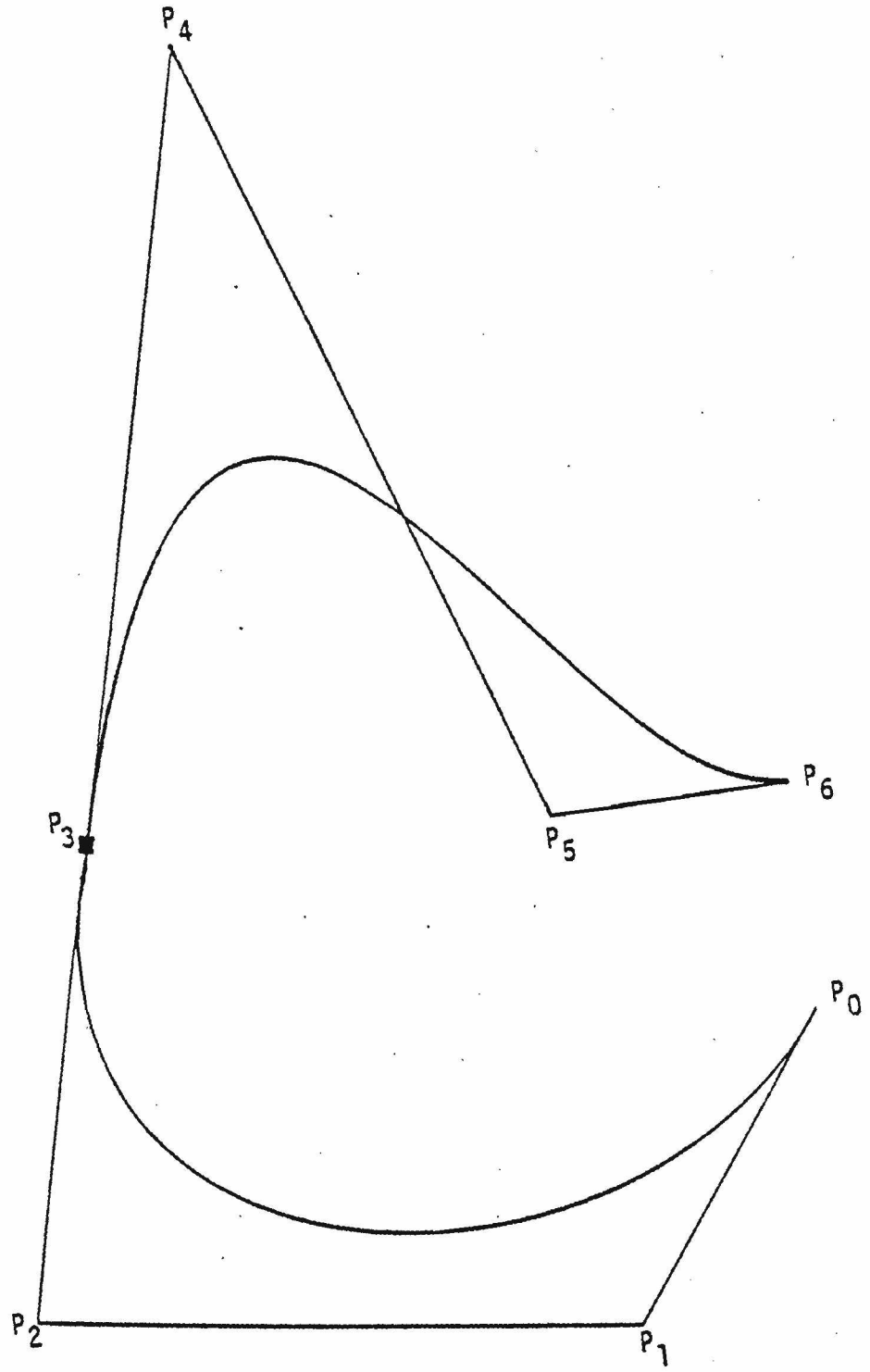


Figure 3.1. $B_3[P_0,\ldots,P_3]$ and $B_3[P_3,\ldots,P_6]$.

A simple induction argument leads to the general statement that piecewise Bernstein approximation is variation-diminishing for arbitrarily many segments. A noteworthy consequence of Theorem 3.1 is Corollary 3.1.1. Piecewise constant and piecewise linear interpolation are variation-diminishing.

Indeed, piecewise constant and piecewise linear interpolation correspond directly to B-spline approximation of order 1 and 2, respectively, and the B-spline basis is known to be totally positive of all orders [3,18].

Although Bézier's System Unisurf has been highly successful in the design of automobiles, there are some problems with vector-valued piecewise Bernstein techniques:

(i) The actual euclidean distance between the vertices $P_i$, $i=0, 1,\ldots,m$, plays no role in the definition of the curve segment, and

(ii) In general, piecewise Bernstein approximation is only $C^0$ continuous. Sometimes design constraints may require continuity of order 3 or even 4.

Gordon and Riesenfeld [19] proposed the following scheme to correct for (i): Let $\alpha_0, \alpha_1, \ldots \alpha_m$ be defined to be the fractional distance of the $i^{th}$ vertex along the polygonal curve $[P_0, \ldots, P_m]$:

$$\alpha_j = \begin{cases} 0 & j=0 \\ \left(\sum_{i=0}^{j-1} |P_{i+1} - P_i|\right) \Big/ \sum_{i=0}^{m-1} |P_{i+1} - P_i| & \text{for } j=1,2,\ldots,m-1. \\ 1 & j=m \end{cases} \tag{3.5}$$

and define

$$B_m^*[P_0, P_1,\ldots,P_m] = B_m[P_0^*, P_1^*,\ldots,P_m^*], \tag{3.6}$$

where $P_j^* = f(j/m)$ and

$$f(s) = (\alpha_{j+1} - \alpha_j)^{-1} [(\alpha_{j+1} - s) P_j + (s - \alpha_j) P_{j+1}] \tag{3.7}$$

for $\alpha_j \leq s \leq \alpha_{j+1}$.

Note that (3.7) is just a piecewise linear interpolant to the polygonal curve $[P_0,P_1,\ldots,P_m]$ and, therefore, from Theorem 3.1 and Theorem 2.7 it follows

$$V[B_m^* [P_0,\ldots,P_m]] \leq V[P_0,\ldots,P_m] \tag{3.8}$$

Gordon and Riesenfeld observed that, when the euclidean distances $|P_{j+1} - P_j|$ are all approximately equal, there will be little difference between $B_m[P_0,P_1,\ldots,P_m]$ and $B_m^*[P_0,P_1,\ldots P_m]$. In extreme cases, however, the difference can be substantial, as shown in Figure 3.2.

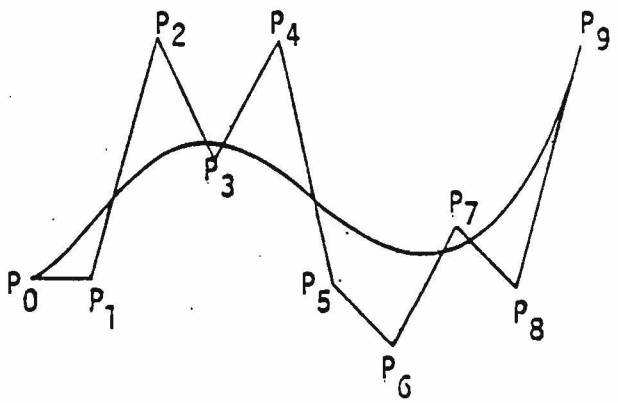


Figure 3.2 (a) $B_{10}[P_0,P_1,\ldots,P_9]$

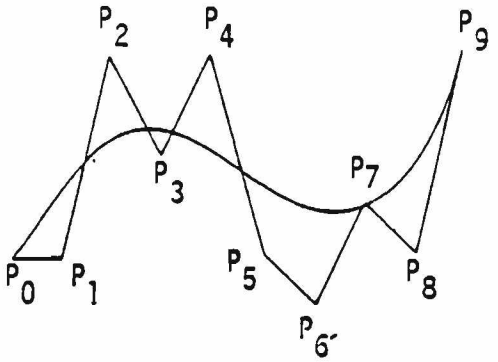


Figure 3.2 (b) $B^*_{10}[P_0,P_1,\ldots,P_9]$

As pointed out in Section I, B-spline approximation has been proposed by Riesenfeld [5] as another alternative to Bernstein approximation. Specifically, the B-spline approximation is variation-diminishing, has the convex hull property and is a piecewise defined curve, where the pieces are joined with arbitrarily high continuity at the discretion of the user. Thus the B-spline basis has all the desirable properties of the piecewise Bernstein basis, yet without the inherent difficiency at the break points of the latter basis. In Theorem 3.2 we shall prove that uniform B-spline approximation is variation-diminishing. We first give a more general definition of the uniform B-spline basis than that given in Section I and prove some elementary properties of B-spline approximation.

Definition 3.1. The ith B-spline basis function $\phi_{i,m}(h,x)$ of degree m-1 with uniform knot spacing and mesh size h is defined by

$$\phi_{i,m}(h,x) = (1/((m-1)!\, h^{m-1})) \sum_{k=0}^{m} (-1)^k \binom{m}{k} ((mh/2) + x - kh - i)_+^{m-1} \tag{3.9}$$

where

$$x_+^{m-1} = \begin{cases} x^{m-1} & x \geq 0 \\ 0 & \text{elsewhere} . \end{cases}$$

Note that $\phi_{i,m}(h,x) = \phi_{i-1,m}(h,x-h)$. It is easily seen [8] that for $m>2$

$$\phi_{i,m}(h,x) = \frac{1}{h} \int_{x-h/2}^{x+h/2} \phi_{i,m-1}(h,t)dt = \pi(h,x)*\phi_i(h,x) \tag{3.10}$$

where * represents convolution, and

$$\pi(h,x) = \begin{cases} 1/h & \text{for } -h/2 \leq x \leq h/2 \\ 0 & \text{elsewhere} \end{cases} \tag{3.11}$$

and where

$$\phi_{i,1}(h,x) = \begin{cases} 1 & \text{for } -h/2 \leq x-i \leq h/2 \\ 0 & \text{elsewhere} . \end{cases} \tag{3.12}$$

In view of the representation (3.10) we readily have $\phi_{i,m}(h,x)$ has positive support $(i-hm/2,\ i + hm/2)$ and

$$\sum_{i=0}^{N} \phi_{im}(h,x) = 1 \quad , \quad h(m-2)/2 \leq x \leq N-h(m-2)/2 \tag{3.13}$$

where $N + 1 \geq m$.

That is, the convolution of two positive functions is positive and the finite support of $h = f*g$, where $f$ has support $[a,b]$ and $g$ has support $[c,d]$, is given by $[a + c,\ b + d]$. As a consequence of the local support of a B-spline basis function, B-spline approximation is a <u>local approximation scheme</u>. Thus any finite sum of the form $\sum_{i=1}^{N} P_i\ \phi_{im}(h,x)$, $N \geq m$, involves at most $m$ nonzero terms.

The following lemmas will be integral to the proof of Theorem 3.2.

Lemma 3.1 [21]. Let $f(t)$, $\{f_k(t),\ k = 1,2,\ldots\}$ be real functions defined for $x\varepsilon[a,b]$. If

$$V[f_k(t)] \leq n,\ k = 1,2,\ldots \tag{3.14}$$

and

$$\lim_{k\to\infty} f_k(t) = f(t), \tag{3.15}$$

then

$$V[f(t)] \leq n.$$

Proof: Let $V[f(t)] = N$. Then there exist points $a\leq t_0<t_1<\ldots<t_N\leq b$ such that $f(t_{j-1})$ and $f(t_j)$ are of opposite sign, $j = 1,\ldots,N$. If $k$ is sufficiently large, we have from (3.15) that

$$\operatorname{sgn} f_k(t_j) = \operatorname{sgn} f(t_j) \quad j = 0,1,\ldots,N.$$

Therefore, for $k$ sufficiently large we have

$$n \geq V[f_k(t)] \geq V[f(t)], \tag{3.16}$$

which proves the lemma.

Lemma 3.2. Let $B_m[P: 0,n]$ be the uniform B-spline approximation of order $m \geq 2$ and mesh size 1 to the polygon $P = [P_0,\ldots,P_n]$ on $[0,n]$, $n \geq m$. That is

$$B_m[P;\ 0,n] = \sum_{i=0}^{n} P_i\ \phi_{i,m}(1,x)\ , \tag{3.17}$$

where $\phi_{i,m}(1,x)$ is given by (3.7) and $x\varepsilon[(m-2)/2,\ n-(m-2)/2]$.

Then

$$B_m[P;\ 0,n] \ = \ \sum_{i=0}^{2n-m+2} P_i^m \ \phi_{i,m}(0.5,x) \ , \qquad (3.18)$$

where $P_i^m$ is defined recursively for $m > 2$ by

$$P_i^m \ = \ (P_i^{m-1} + P_{i+1}^{m-1})/2 \ , \qquad i = 0,1,\ldots,2n-m+2 \qquad (3.19)$$

and

$$P_i^2 \ = \ \begin{cases} P_{(i/2)} & i \text{ even} \\ (P_{(i-1)/2} + P_{(i+1)/2})/2 & i \text{ odd} \end{cases} \qquad i = 0,\ldots,2n-1$$

That is, given the B-spline approximation of order m, with integral knot spacing to the polygon P, the control points for the same curve in terms of the B-spline basis over the refined mesh {0.0, 0.5, 1.0, ...,n} are given by (3.19).

Proof: (by induction on degree) Let $m = 2$. We must show

$$\sum_{i=0}^{n} P_i \ \phi_{i,2}(x) \ = \ \sum_{l=0}^{2n-m+2} P_l^2 \ \phi_{l,2}(0.5,x) \qquad . \qquad (3.20)$$

But the degree 1 (order 2) B-spline approximation is the piecewise linear interpolant to the vertices $P_i$. Thus (3.20) holds, since piecewise linear interpolation preserves linear polynomials. (See Figure 3.3.)

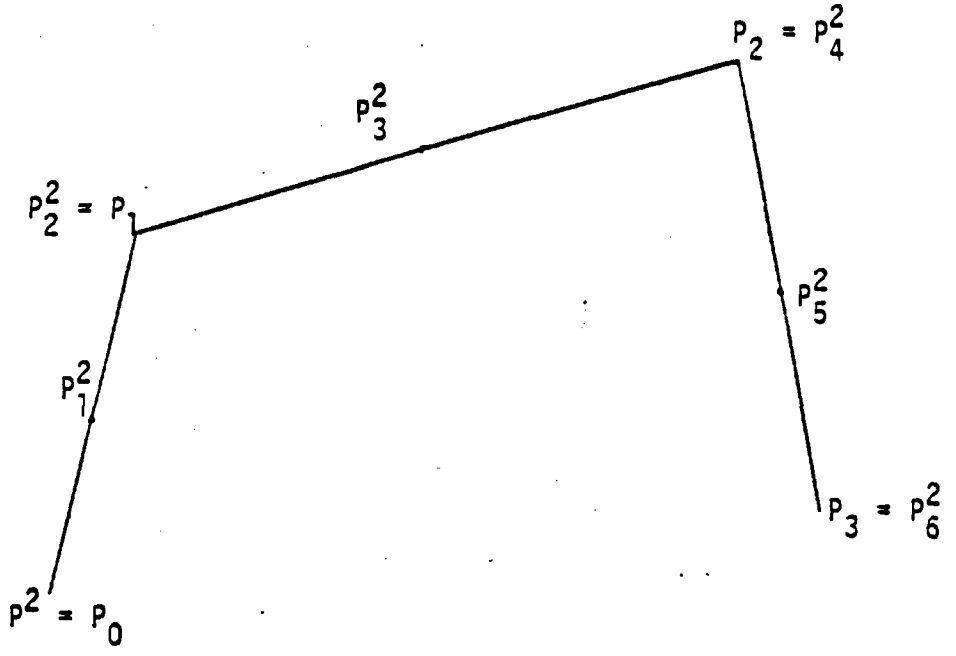


Figure 3.3. Piecewise linear interpolation to $\{P_0,\ldots,P_3\}$ and $\{P_0^2,\ldots,P_6^2\}$ results in the same curve.

Now assume (3.18) holds for all k, $2 \leq k < m$. We must show it also holds for $k = m$. From (3.8) we get

$$\sum_{i=0}^{n} P_i \, \phi_{i,m}(l,x) = \sum_{i=0}^{n} P_i(\pi(l,x)^* \, \phi_{i,m-1}(l,x))$$

$$= \pi(l,x)^* \sum_{i=0}^{n} P_i \, \phi_{i,m-1}(l,x) \qquad (3.21)$$

for $x\varepsilon[(m-2)/2, n-(m-2)/2]$. Now by our induction hypothesis

$$\sum_{i=0}^{n} P_i \, \phi_{i,m-1}(1,x) = \sum_{i=0}^{2n-m+3} P_i^{m-1} \, \phi_{i,m-1}(0.5,x), \qquad (3.22)$$

thus (3.21) reduces algebraically to

$$\sum_{i=0}^{n} P_i \, \phi_{i,m}(1,x) = \pi(1,x) * \sum_{i=0}^{2n-m+3} P_i^{m-1} \, \phi_{i,m-1}(0.5,x)$$

$$= (\pi(0.5,x-0.25)/2 + \pi(0.5,x+0.25)/2) * \sum_{i=0}^{2n-m+3_{m-1}} P_i \phi_{i,m-1}(0.5,x)$$

$$= (\tfrac{1}{2}) \sum_{i-0}^{2n-m+3} P_i^{m-1} \, \pi(0.5,x-0.25) * \phi_{i,m-1}(0.5,x)$$

$$+ (\tfrac{1}{2}) \sum_{i=0}^{2n-m+3} P_i^{m-1} \, \pi(0.5,x+0.25) * \phi_{i,m-1}(0.5,x)$$

$$= (\tfrac{1}{2}) \left( \sum_{i=0}^{2n-m+3} P_i^{m-1} \, \phi_{i,m}(0.5,x-0.25) + \sum_{i=0}^{2n-m+3} P_i^{m-1} \, \phi_{i,m}(0.5,x+0.25) \right)$$

(3.23)

for $x \varepsilon [m-2/2, n-(m-2)/2]$. Now both $\phi_{0,m}(0.5,x-0.25)$ and $\phi_{2n-m+3,m}(0.5, x+0.25)$ are zero on the interval $[(m-2)/2, n-(m-2)/2]$, and after dropping terms concerning these basic functions and rearranging the remaining terms in (3.23) we have

$$\sum_{i=0}^{n} P_i \, \phi_{i,m}(1,x) = \sum_{i=0}^{2n-m+2} ((P_i^{m-1} + P_{i+1}^{m-1})/2) \, \phi_{i,m}(0.5,x)$$

$$= \sum_{i=0}^{2n-m+2} P_i^m \, \phi_{i,m}(0.5,x) \quad . \tag{3.24}$$

Thus, we have proved the lemma.

Lemma 3.3. Let $B_m[P;\, 0,n] = \sum_{i=0}^{n} P_i \, \phi_{i,m}(1,x)$ as in (3.17) and define

$$\psi_0[P] = [P_0^m, P_1^m, \ldots, P_{2n-m+2}^m] \tag{3.25}$$

where $P_i^m$ is given by (3.18) and

$$\psi^k[P] = \psi_0[\psi^{k-1}[P]] = [P_0^{m,k}, \ldots, P_{p(m)}^{m,k}] \quad , \tag{3.26}$$

for $p(m) = 2(2n-m+2)^k - m+1$, $k \geq 1$ .

Then,

$$\lim_{k\to\infty} \psi^k[P] = B_m[P;\, 0,n] \, . \tag{3.27}$$

That is, the sequence of polygons $\psi^k[P]$ converge to the curve $B_m[P;\, 0,n]$ (see Figure 3.4). Riesenfeld [33] has recently given a proof of Lemma 3.3 for the case m=3. Here we use a quite different approach for the proof of arbitrary but finite m.

Proof: By Lemma 3.2 and induction on k we have

$$B_m[\psi^k[P];\, 0,p^{(m)}] = B_m[P;\, 0,n] \, .$$

Now by minimizing the nature of piecewise linear interpolation it is easily seen that

$$|P^2_{i+1} - P^2_i| \leq \tfrac{1}{2} \max |P_{i+1} - P_i| \tag{3.28}$$

By applying the triangle inequality to (3.18) we can deduce

$$|P^m_{i+1} - P^m_i| \leq |P^{m-1}_{i+1} - P^{m-1}_i|$$

which, combined with (3.28) and induction, imply

$$|P^{m,k}_{i+1} - P^{m,k}_i| \leq (1/2)^k L, \tag{3.29}$$

where L is the max $|P_{i+1} - P_i|$ .

That is by (3.29)

$$\lim_{k\to\infty} |P^{m,k}_{i+1} - P^{m,k}_i| = 0 \tag{3.30}$$

But then

$$\lim_{k\to\infty} |P^{m,k}_{i+j} - P^{m,k}_i| = 0 \tag{3.31}$$

for arbitrary $i\varepsilon\{1,\ldots,p(m)\}$, $j=1,\ldots,m$ , $i+j \leq p(m)$.

From (3.27) and (3.13) we know $B_m[P;\ 0,n](x_0)$, for each $x_0\varepsilon((m-2)/2, n-(m-2)/2)$, lies within the convex hull of $P^{m,k}_i$, $P^{m,k}_{i+1}$, $\ldots, P^{m,k}_{i+m}$ for some i. But with (3.31), we can then conclude

$$\lim_{k\to\infty} \psi^k[P] = B_m[P;\ 0,n]. \tag{3.32}$$

Theorem 3.2. Uniform B-spline approximation is variation-diminishing.

Proof: We have from (3.32)

$$\lim_{k\to\infty} \psi^k[P] = B_m[P; 0,n].$$

It follows from the piecewise linear nature of the construction (3.19) and Corollary 3.1.1 that $V[\psi^k[P]] \leq V[P]$ for all k. The conclusion now follows from Lemma 3.1.

Bernstein-Bézier, B-spline and other generalizations of the Bézier curve [13,14] still appear to have inherent shortcomings for realtime interactive design. In particular,

I. It is often the case that a user wishes to create a local fit to the polygon in his design, yet there are no 'natural' handles or control parameters in the above methods for the designer to manipulate interactively in order to achieve these shapes.

II. Further, the extension from curves to surfaces with the above schemes requires a rectilinear network of control points, a severe restriction on interactive design.

For instance, the addition (deletion of a point to a rectilinear net requires, in general, the addition (deletion) of a whole row and column of points in order to retain the rectilinear topology

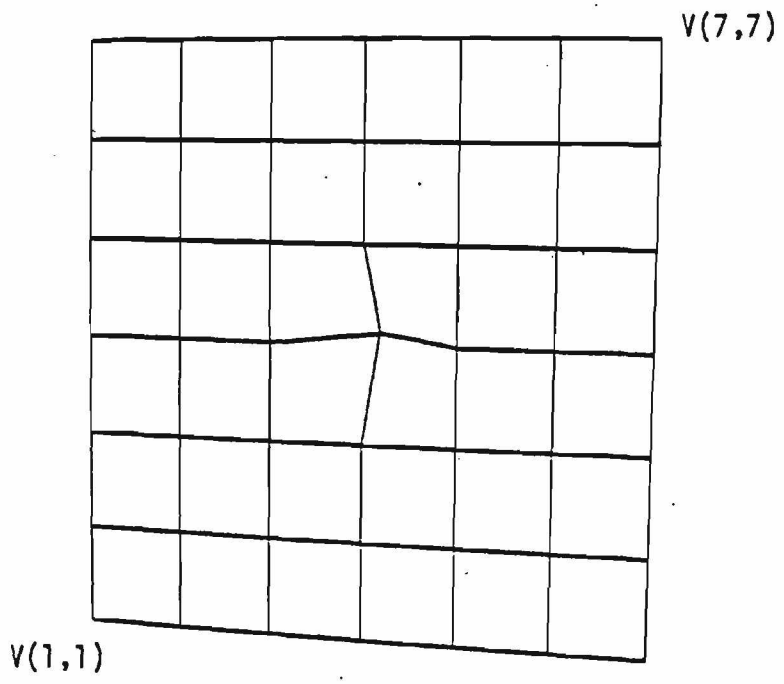


Figure 3.4. Rectilinear topology. Note that to remove the center vertex and still retain the rectilinear structure one has to remove all of Row 4 and Column 4.

The Construction of New Models

We now apply the theory of total positivity to the construction of new models for curve and surface design, with emphasis on methods which avoid the deficiencies I and II discussed in the previous section. Specifically, for curves we develop linear operators

$$L[P;a,b] = \sum_{i=0}^{n} P_i \phi_i(t,\alpha_i) \quad \text{for } P = [P_0,\ldots,P_n] \qquad t\varepsilon(a,b) \tag{3.33}$$

where the $\alpha_i$ can be varied by the user of the system to control local "closeness" of fit to the polygonal curve P, thus giving the user the ability to create "cusps" and "flats" in his design with the same natural flexibility he has in moving the vertices of the polygon. Further, we build in the desirable properties of the Bernstein and B-spline methods,

$$V(L) \leq V[P] \quad \text{(variation-diminishing property)} \tag{3.34}$$

and

$$\phi_i(t,\alpha_i) \geq 0,$$

for all i with

$$\sum_i \phi_i(t,\alpha_i) = 1 \quad \text{(convex hull property)} \tag{3.35}$$

It is well known [15,16] that B-spline approximation to a continuous function f on an interval [a,b] for fixed degree converges to f as the mesh size h goes to zero. In terms of our primitive polygon P, we can get a closer local fit to the polygonal curve P if we sample not only at the vertices of the polygon but also in the interval in which we wish to approximate more closely. The more samples in the region of interest the closer our approximation to the polygon there. That is, if we let the polygon P be defined by

$$P(t) = \sum_{j=0}^{n} P_j \, \phi_j(t) \tag{3.36}$$

where the $\phi_j$ are the piecewise linear cardinal functions, rather than form the B-spline approximation to the polygon P by sampling at

the vertices $P_j$, we define

$$G(t) = \sum_{k=1}^{p} \phi_{k,m}(t) P(t_k) \quad , \tag{3.37}$$

where the $\phi_{k,m}$ are the B-spline basis functions of degree m and the $t_k$ are the knots, which are not just located at the vertices, but in clusters in intervals of local interest (see Figure 3.5).

Since

$$P(t_k) = \sum_{j=0}^{n} P_j \, \phi_j(t_k) \ ,$$

(3.37) can be rewritten as

$$\begin{aligned} G(t) &= \sum_{k=1}^{p} \phi_k^m(t) \, P(t_k) \\ &= \sum_{k=1}^{p} \phi_k^m(t) \, \left( \sum_{j=0}^{n} P_j \, \phi_j(t_k) \right) \\ &= \sum_{j=0}^{n} P_j \left( \sum_{k=0}^{p} \phi_k^m(t) \, \phi_j(t_k) \right) \\ &= \sum_{j=0}^{n} P_j \, \phi'_{j,m}(t) \ , \end{aligned}$$

where

$$\phi'_{j,m}(t) = \sum_{k=1}^{p} \phi_k^m(t) \, \phi_j(t_k) \quad .$$

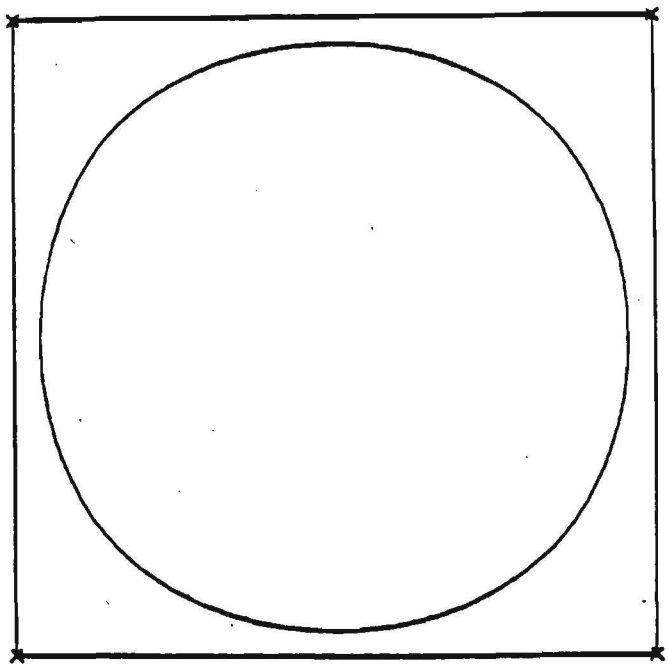

Figure 3.5 (a). Cubic B-spline approximation to the polygon given.

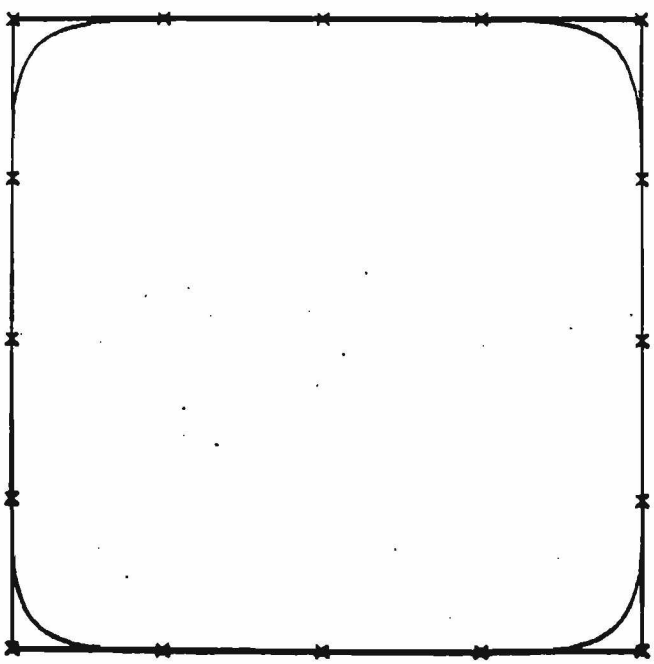

Figure 3.5 (b). Cubic B-spline approximation with increased sampling on the edges.

Since the B-spline basis is TP, we know from Theorem 2.6 and Theorem 2.7 that the basis $\{\phi'_{j,m}\}$ is TP. Thus by appropriately choosing the knots $\{t_k\}$ we can generate a new basis $\{\phi'_j\}$ which more closely approximates the polygon P, yet retains the valuable convex hull and variation-diminishing properties. More importantly, this construction method generalizes to other TP bases:

Let $G_0(x) = \sum_{n=1}^{n} \phi_{j,0}(x)\, P_j$, $x\varepsilon[a,b]$, where $\phi_{j,0}(x)$ is $TP_n$ on XxJ, J = {1,2,...,n} and define recursively

$$G_i(x) = \sum_{j=1}^{n(i)} \phi_{j,i}(x)\, G_{i-1}(x_{j,i}), \quad i = 1,2,\ldots,m \tag{3.38}$$

where $x\varepsilon[a,b]$ and $\phi_{j,i}(x)$ is $TP_{n(i)}$ on XxJ(i), J(i) = {1,2,...,n(i)} where $n(i) \geq n(i-1)$ and the sequence $\{x_{j,i}\}$ is the knot vector associated with ith interation. Expanding $G_{i-1}(x_{j,1})$, (3.38) can be rewritten

$$\begin{aligned} G_i(x) &= \sum_{j=1}^{n(i)} \phi_{j,i}(x) \sum_{k=1}^{n(i-1)} \phi_{k,i-1}(x_{j,i})\, G_{i-2}(x_{k,i-1}) \\ &= \sum_{k=1}^{n(i-1)} G_{i-2}(x_{k,i-1}) \left( \sum_{j=1}^{n(i)} \phi_{j,i}(x)\, \phi_{k,i-1}(x_{j,i}) \right) \\ &= \sum_{k=1}^{n(i-1)} G_{i-2}(x_{k,i-1})\, (\phi'_{k,i-1}(x)) \,. \end{aligned} \tag{3.39}$$

By Theorem 2.7 $\phi'_{k,i-1}(x)$ is $TP_{n(i-1)}$ and, therefore, by Theorem 2.10

$$V[G_i(x)] \le V[G_{i-2}(x_{i,i-1}),\ldots,G_{i-2}(x_{n(i-2),i-1})] \tag{3.40}$$

Repeating the above algorithm we readily deduce

$$G_i(x) = \sum_{j=1}^{n} \phi'_{j,0}(x)\, P_i, \tag{3.41}$$

where

$$V[G_i(x)] \le V[P_1, \ldots, P_n] \tag{3.42}$$

Now let

$$\phi(x) = \sum_{j=1}^{n} \phi'_{j,0}(x)$$

Then

$$\phi''_{j,0}(x) = \phi'_{j,0}(x)/\phi(x), \qquad j = 1,\ldots,n \tag{3.43}$$

is $TP_n$ by Theorem 2.4 and

$$\sum_{j=1}^{n} \phi''_{j,0}(x) = 1 \tag{3.44}$$

It will now be demonstrated how the techniques developed in (3.38) and (3.43) for the construction of the variation-diminishing and convex hull properties, respectively, can be used to generate a set of basis functions with the "natural" handles desired. Let

$$G_1(x) = P = [P_1,\ldots,P_n] \tag{3.45}$$

be the piecewise linear interpolant to the points $\{P_i,\ i = 1,\ldots,n\}$ over the uniform mesh $\{1,\ldots,n\}$. Define

$$G_2(x) = \sum_{j=1}^{2(n-1)} \phi_{j,1}(0.5,x) P_j^2 \tag{3.46}$$

where $\phi_{j,1}(0.5,x)$ is $TP_{2(n-1)}$ and

$$P_j^2 = (1-\alpha_j)\, P_j^1 + \alpha_j\, P_{j+1}^1 \qquad j = 1,2,\ldots,2n-2 \quad \alpha_j \in [0,1] \tag{3.47}$$

where

$$P_j^1 = \begin{cases} P_{(j+1)/2} & j \text{ odd} \\ (P_j^1 + P_{(j+1)}^1)/2 & j \text{ even} \end{cases} \qquad \text{for } j = 1,2,\ldots,2n-1.$$

(See Figure 3.6.)

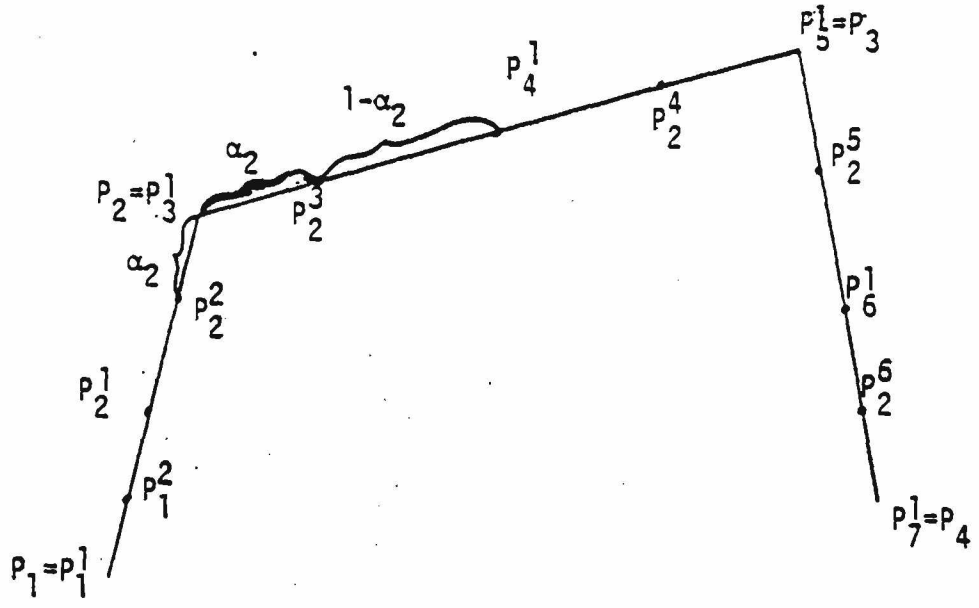


Figure 3.6. This figure illustrates the construction (3.47).

Now setting $\alpha_j = 1 - \alpha_{j-1}$, for j odd, and rearranging terms as in (3.38) we have

$$G_2(x) = \sum_{j=1}^{n} \phi_{j,1}(x;\alpha_{j-1}, \alpha_j, \alpha_{j+1})\, P_j, \tag{3.48}$$

where $\phi_{j,1}(x;\alpha_{j-1},\alpha_j,\alpha_{j+1})$ are TP and vary in shape from trapezoidal $(\alpha_{j-1},\alpha_j,\alpha_{j+1} = 0)$ to triangular $(\alpha_{j-1},\alpha_j,\alpha_{j+1} = 1)$. That is by varying the $\{x_j\}$ we can control the "resampling" of the $\{\phi_{j,1}\}$ and thus the shape of the basis functions. (See Figure 3.7).

Now let $\{\psi_j(x),\ j = 1,\ldots,n\}$ be any set of totally positive functions and define

$$\begin{aligned} G_3(x) &= \sum_{j=1}^{n} \psi_j(x)\, G_2(x_j) \\ &= \sum_{j=1}^{n} \psi_j(x) \sum_{j=1}^{n} \phi_{i,1}(x;\, \alpha_{i-1},\alpha_i,\alpha_{i+1})\, P_i \\ &= \sum_{i=1}^{n} \psi_i'(x;\alpha_{i-1},\alpha_i,\alpha_{i+1})\, P_i, \end{aligned} \tag{3.49}$$

where

$$\psi_i'(x;\alpha_{i-1},\alpha_i,\alpha_{i+1}) = \sum_{j=1}^{n} \psi_j(x)\, \phi_{i,1}(x_j;\alpha_{i-1},\alpha_i,\alpha_{i+1}) . \tag{3.50}$$

For example, let $\psi_j(x)$ be the uniform B-spline basis of degree m-1. Then for $\alpha_j$ small, the basis functions are "bell-shaped," while for $\alpha_j$ large, these basis functions converge in shape to the piecewise linear basis functions. This convergence takes markedly for the lower degrees (see Figures 3.8 and 3.9).

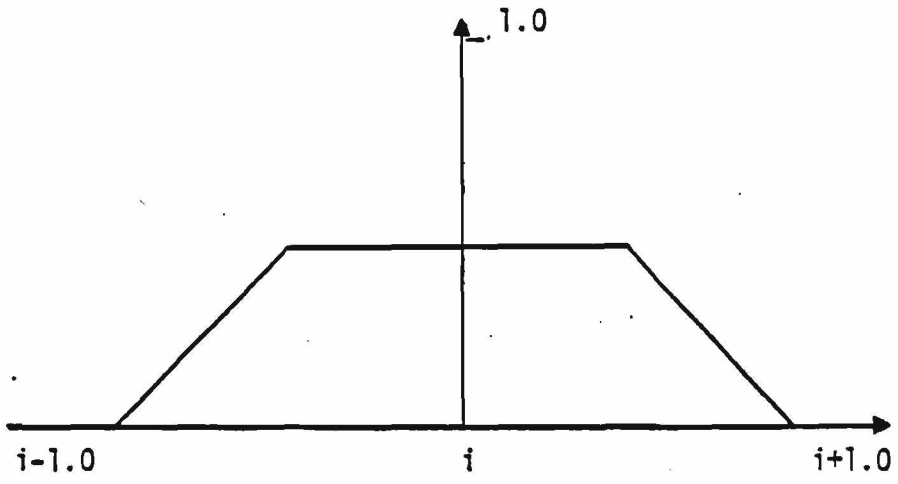


Figure 3.7 (a). Canonical basis function $\phi_{i,1}(x;0,0,0)$.

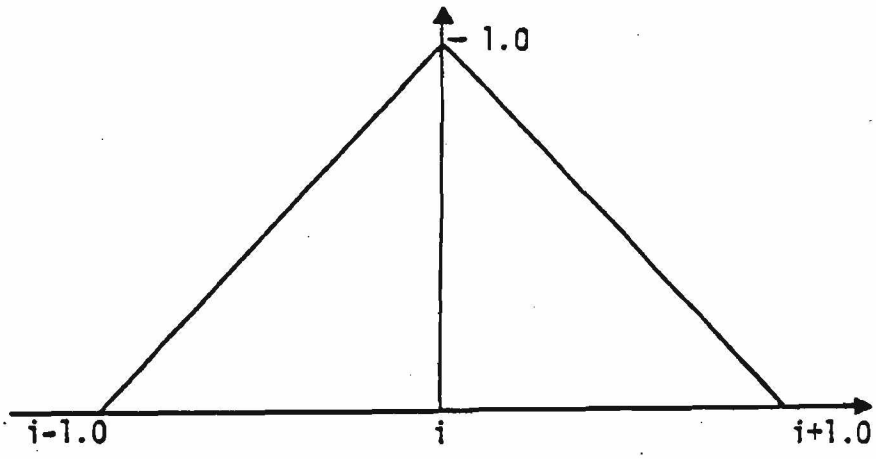


Figure 3.7 (b). Canonical basis function $\phi_{i,1}(x;1,1,1)$.

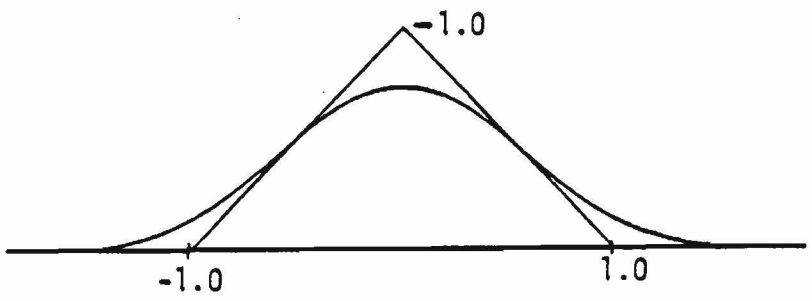


Figure 3.8 (a). Canonical basis function $\psi_0'(x;0,0,0)$ for the choice $\psi_0(x)$ = B-spline basis of degree 3.

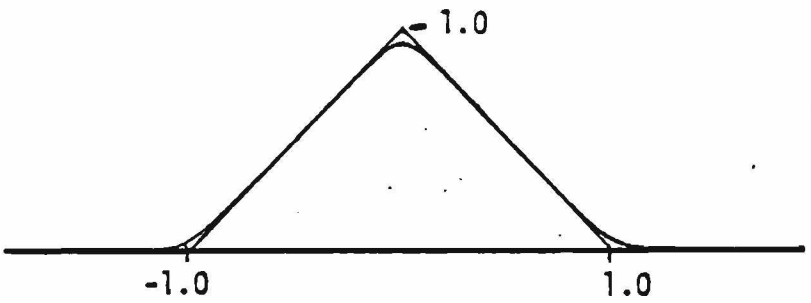


Figure 3.8 (b). $\psi_0'(x;0.8,0.8,0.8)$.

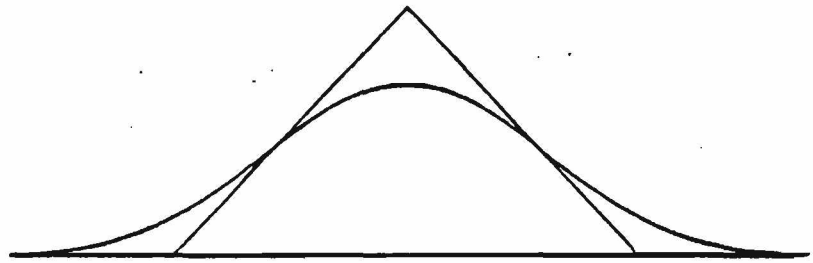

Figure 3.9 (a). Canonical basis function $\psi_0'(x;0,0,0)$ for the choice $\psi_0(x)$ = B-spline basis of degree 5.

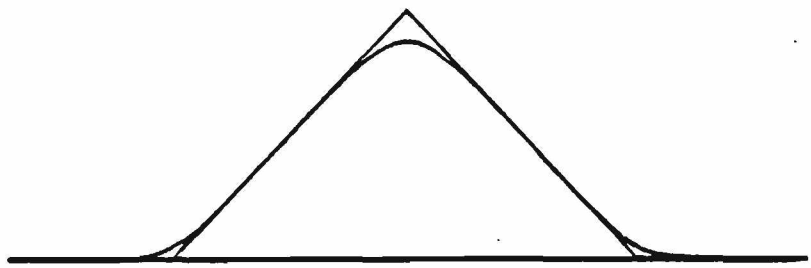

Figure 3.9 (b). $\psi_0'(1,1,1)$.

The previous construction give us an indication of the possibilities for creating new TP bases from other TP bases. If in (3.49) we choose

$\psi_j(x) = \sum_{i=0}^{n-1} \binom{n-1}{i} x^i(1-x)^{n-1-i}$, i.e., the Bernstein basis of degree n-1, then our construction is very similar to that of Gordon and Riesenfeld defined by equations (3.6) and (3.7), only here we have allowed the number and position of the $P_j^*$ to be flexible.

There are other schemes which easily can be developed from a careful perusal of the construction prototype (3.38). For instance, if in (3.49) we take $\{\psi_j(x), j=1,\ldots,n\}$ to be the B-spline basis of degree 2 and

$$\alpha_j = 1/2, \qquad j = 1,\ldots,n,$$

then the construction reduces to that of (3.18). From Lemma 2.5 we know that

$$G_3(x) = B_3[P;\ 0,n].$$

The method can be extended by generalizing the construction (3.19) to create a whole family of bases $\{\psi_{i,m}\}$ such that

$$\psi'_{j,m}(x;1/2,1/2,1/2) = \phi_{j,m}(x), \qquad j = 1,\ldots,n, \tag{3.51}$$

where $\phi_{j,m}(x)$ is the $m^{th}$ order B-spline basis functions.

Thus the family of bases $\{\psi'_{i,m},\ i = 1,\ldots,n\}$ is a proper generalization of the family of B-spline bases, where the $\{\psi'_{i,m},\ i = 1,\ldots, n\}$ form a substantial improvement in terms of the 'natural' handles desired. (See Figures 3.10 and 3.11.)

More Curve Techniques

In the previous constructions we used a discrete form of Theorem 2.7 to construct TP functions from other TP functions. An equally valuable technique for CAGD results when the variable of summation in (2.16) is, in fact, continuous. For example, let

$$P(t) = \sum_{i=0}^{n} \phi_i(x)\ P_i \tag{3.52}$$

as defined in (1.1), and define

$$Q(t) = T(P)(t) = \int_{-\infty}^{\infty} K(t,s)\ P(s)\ ds, \tag{3.53}$$

where $K(t,s)$ is TP on $R^2$. Then by Theorem 2.2 we have

$$V[Q] \leq V[P]\ . \tag{3.54}$$

Rewriting (3.53) we have

$$Q(t) = \int_{-\infty}^{\infty} K(t,s)\ P(s)\ ds$$

$$= \int_{-\infty}^{\infty} K(t,s) \sum_{i=0}^{n} \phi_i(s)\ P_i\ ds$$

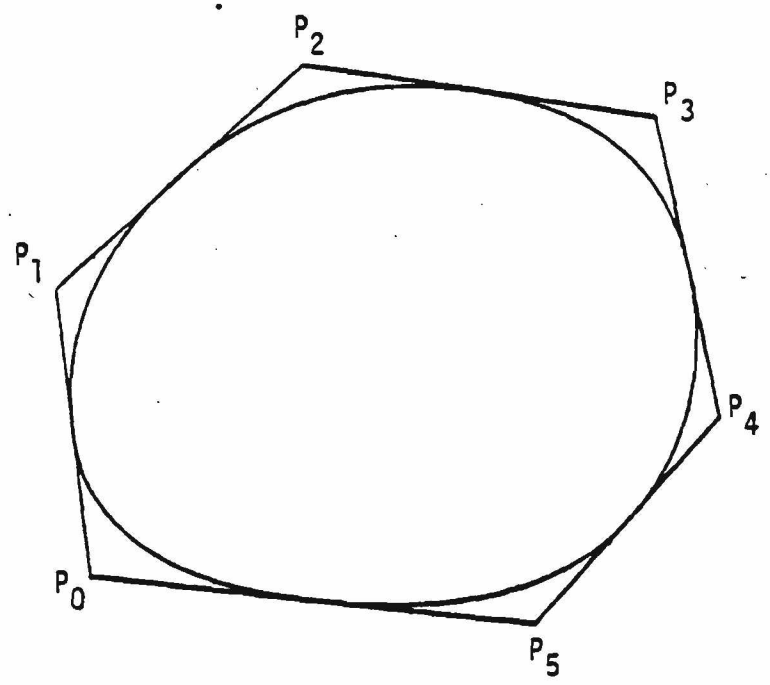


Figure 3.10 (a). $B_2[P; 0,5]$.

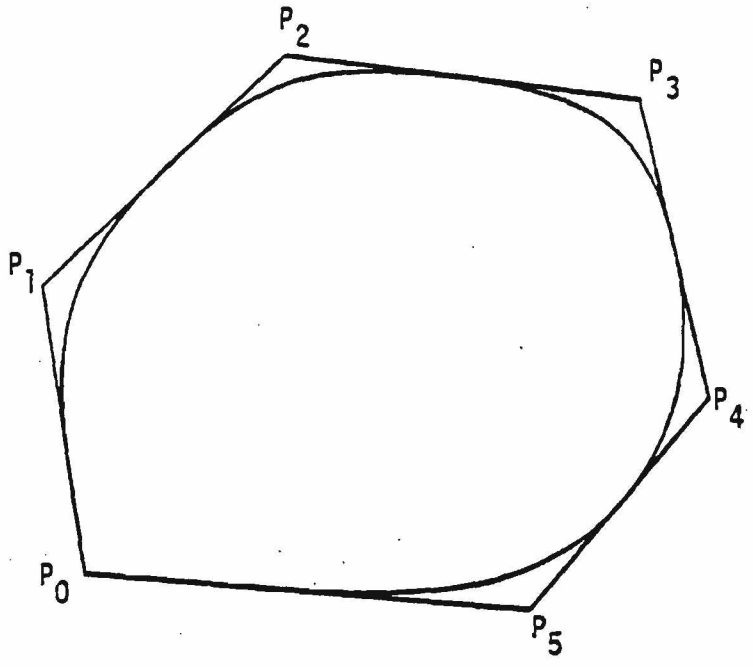


Figure 3.10 (b). Closer fit to the polygon at $P_0$ and $P_2$.

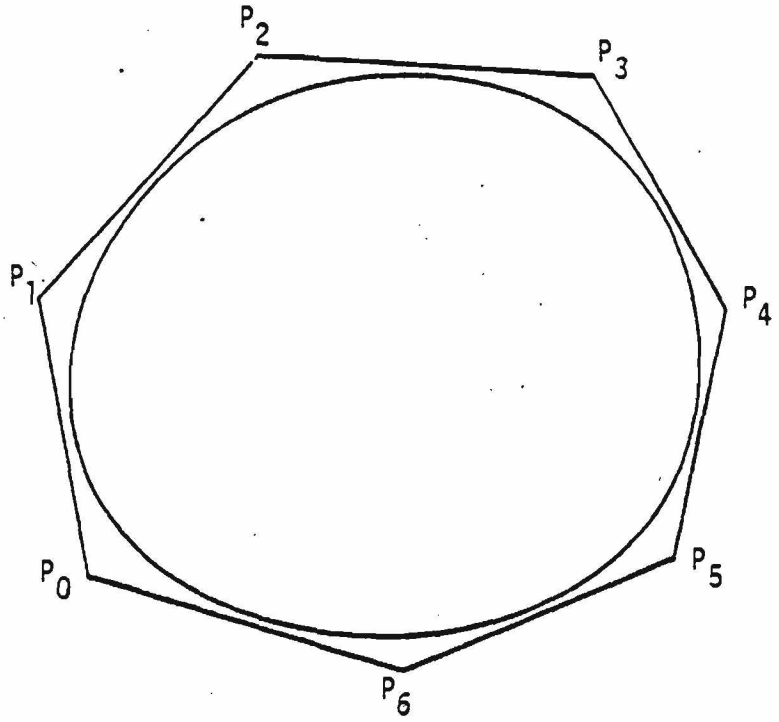


Figure 3.11 (a). $B_3[P; 0,6]$.

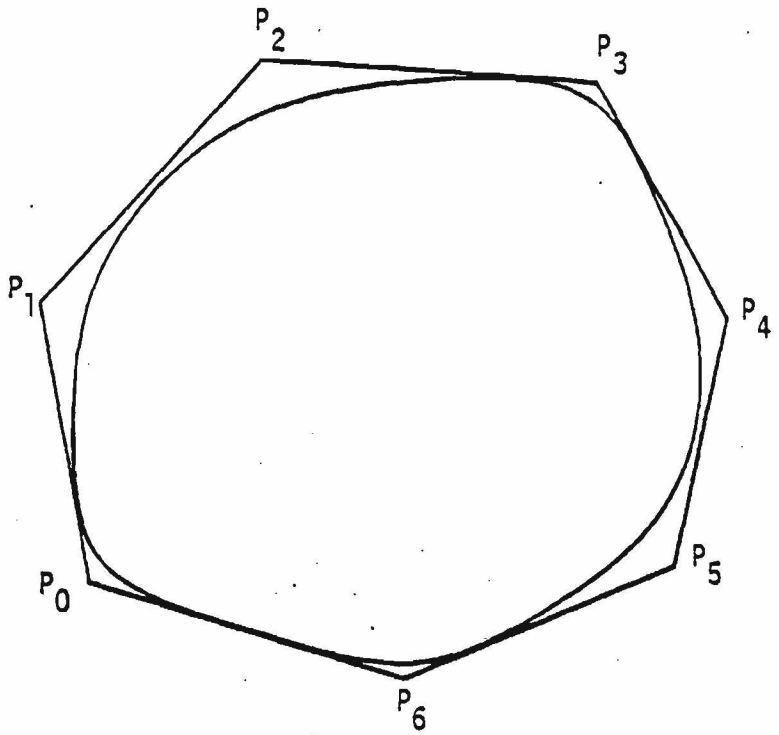


Figure 3.11 (b). Closer fit to the polygon at $P_0$, $P_3$, $P_6$.

$$= \sum_{i=0}^{n} \left(\int_{-\infty}^{\infty} K(t,s)\phi_i(s)\right) ds \; P_i \tag{3.55}$$

$$= \sum_{i=0}^{n} \phi_i'(t) \; P_i, \text{ where } \phi_i'(t) = \int_{-\infty}^{\infty} K(t,s)\phi_i(s) \; ds$$

In view of Theorem 2.7, provided that $[\phi_i(t)]$ is $TP_n$, it follows $[\phi_i'(t)]$ is $TP_n$.

To develop a class of kernels $K(t,s)$ suitable for CAGD we restrict our attention to TP kernels of the form $K(x,y) = f(x-y)$ (translation kernels) where $(x,y) \in R^2$ and $\int_{-\infty}^{\infty} f(x) = 1$. Under these constraints the transformation (3.53) becomes a convolution, i.e.,

$$T(P)(t) = \int_{-\infty}^{\infty} K(t-s) \; P(s) \; ds. \tag{3.56}$$

Recall from Section II that the function $e^{xy}$ is TP on $R^2$. Then by Theorems 2.4 and 2.5,

$$K_1(x,y) = e^{-\alpha(x-y)^2}$$

$$= e^{-\alpha x^2} \; e^{-\alpha y^2} \; e^{2\alpha xy} \tag{3.57}$$

is TP for $\alpha > 0$. Similarly, it can be shown that

$$K_2(x,y) = e^{-\alpha|x-y|} \tag{3.58}$$

is TP, for $\alpha > 0$, $(x,y) \in R^2$. Since $K_1(x,y)$ and $K_2(x,y)$ are TP, from Theorem 2.7 we know their self-convolutes are TP. That is, if we define

$$K_i^j(x,y) = K_i^{j-1}(x,y) * K_i(x,y) \qquad i = 1,2;\ j = 1,2\ldots, \tag{3.59}$$

where

$$K_i^0(x,y) = K_i(x,y) \qquad i = 1,2,$$

then

$$K_i^j(x,y) \text{ is TP on } R^2. \tag{3.60}$$

The kernels $K_i^j$ are "bell-shaped" and symmetric, the "spread" of the curves depending on $\alpha$. (See Figures 3.12, 3.13.) Note that $K_1^j(x,y)$ is of continuity class $C^\infty$ for all j, while $K_2^j(x,y) \in C^j$.

If P(t) in (3.52) is our primitive polygon, which means that $\{\phi_i(x)\}$ is the cardinal piecewise linear basis, then we can construct new bases

$$\phi_i^j(x) = K_p^j(x) * \phi_i(x) \tag{3.61}$$

for p = 1,2; i = 0,1,...,n; and j = 1,2,..., such that $\{\phi_i^j(x)\}$ is TP on XxI, X = $(-\infty,\infty)$, I = {0,1,...,n} for all j.

There are several alternative methods of ensuring that the bases $\{\phi_i^j(x)\}$ enjoy the convex hull property. We can either use the technique developed in (3.43) or we can normalize $\{K_i^j(x)\}$ such that

$$\int_{-\infty}^{\infty} K_i^j(x)\,dx = 1.$$

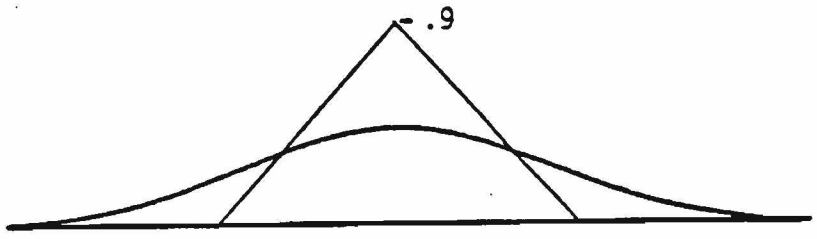


Figure 3.12. $K_2^8(x)$, $\alpha = 1$

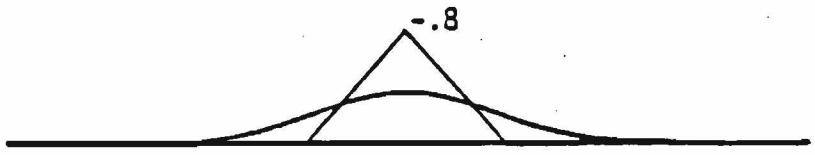


Figure 3.13. $K_1^{12}(x)$, $\alpha = 1$

That is, define

$$L_i^j(x) = K_i^j(x)/C \ , \ i = 1,2; \ j = 1,2, \ \ldots$$

where

$$C = \int_{-\infty}^{\infty} K_i^j(x) \, dx.$$

Then from (3.61) we have for $i = 1,2$,

$$\begin{aligned} \sum_{k=0}^{n} \phi_k^j(x) &= \sum_{k=0}^{n} (L_i^j(x) * \phi_k(x)) \\ &= L_i^j(x) * \sum_{k=0}^{n} \phi_k(x) \\ &= L_i^j(x) * 1 \\ &= 1 \end{aligned} \tag{3.62}$$

Of course, in view of the recursive nature of $\{K_i^j(x,y)\}$, various combinations of (3.43) and (3.62) could be used as well.

Not surprisingly, in view of their similarity in shape to the B-spline basis (compare Figures 3.14 and 3.15 and Figures 3.16 and 3.17), curves formed with these new bases can be remarkably close in shape to the B-spline approximation. This similarity is explained mathematically by the Central Limit Theorem.

Theorem 3.3 (Central Limit Theorem) [27]. Let $f_i(t)$, $t\varepsilon(-\infty,\infty)$, $i$ an integer, be real positive and symmetric about $t = 0$, such that

$$\int_{-\infty}^{\infty} f_i(t) \, dt = 1$$

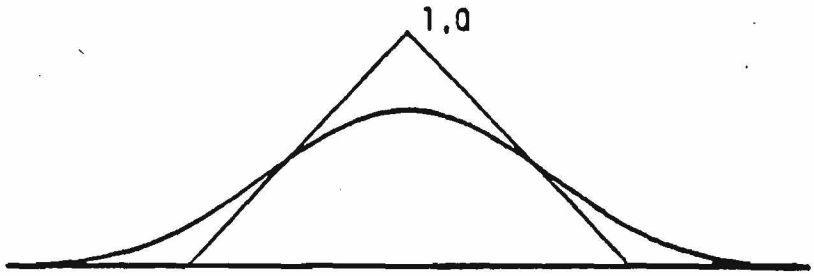


Figure 3.14. Canonical B-spline basis, degree 7.

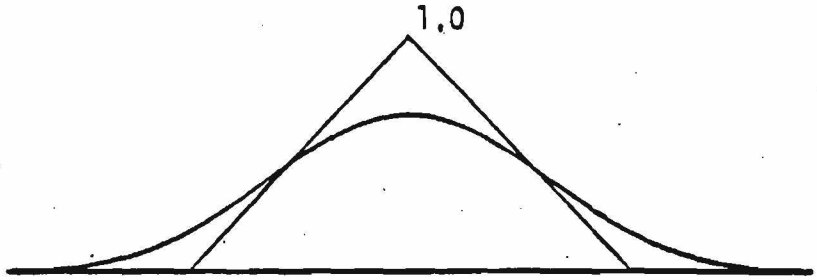


Figure 3.15. $(\sqrt{6}/\sqrt{7\pi})\ e^{-(6/7)x^2}$.

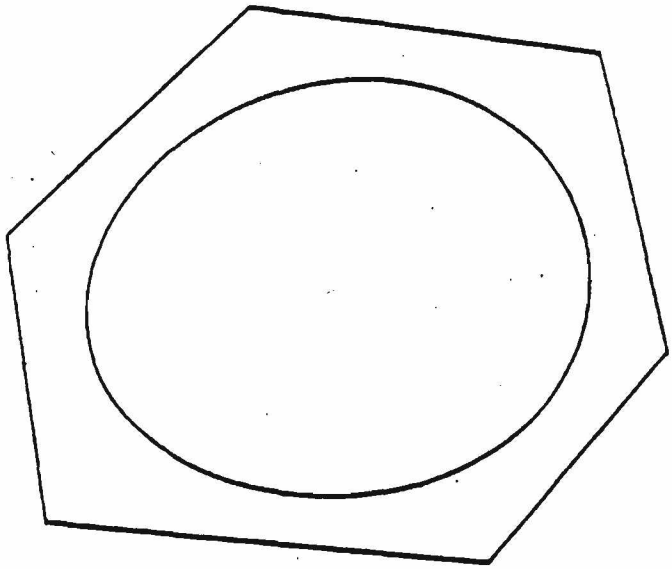

Figure 3.16. B-spline approximation, degree 7.

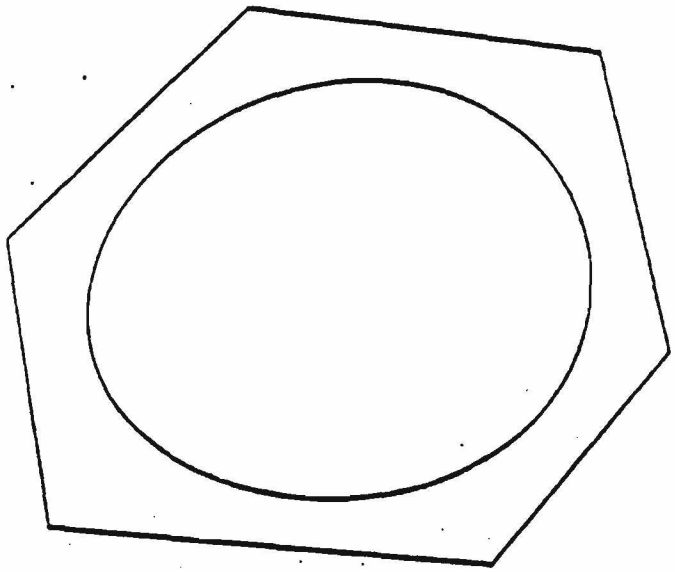

Figure 3.17. Same polygon, where the approximation was obtained with the convolution kernel $(\sqrt{6/(\sqrt{7}\pi)})e^{-(6/7)x^2}$.

for all i and define

$$f(t) = f_1(t) * f_2(t) *\ldots* f_n(t), \tag{3.63}$$

where, again, * represents convolution. Then if

$$\int_{-\infty}^{\infty} t^3 f_i(t)\, dt \leq C \tag{3.64}$$

where C is an arbitrary constant, and

$$\lim_{n\to\infty} \sum_{i=1}^{n} \int_{-\infty}^{\infty} t^2 f_i(t)\, dt = \infty \tag{3.65}$$

Then

$$\lim_{n\to\infty} f(t) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(t)^2/2\sigma^2} \tag{3.66}$$

where

$$\sigma^2 = \sum_{i=1}^{n} \int_{-\infty}^{\infty} t^2 f_i(t)\, dt .$$

We constructed the $\{\phi_i^j(x)\}$ in (3.61) precisely as the iterated convolution of j functions, and a reexamination of the B-spline basis of degree m reveals that each is the iterated convolution of m functions, as well. Without going into detail, it can be shown [3,12] that all of these functions satisfy the hypothesis of Theorem 3.3. Therefore, in the limit, all these bases would be of Gaussian form, differing only in their dispersion about their point of symmetry. Thus the similarity in shape of the various approximation methods is in actuality a reflection of this "tendency to Gaussian form" in the basis functions.

Surfaces

Although the extension of vector-valued curve methods to surfaces is straightforward, there is no satisfactory theory of total positivity for functions of more than one variable [3]. However, when dealing with surface equations of the form

$$\sum_{i=0}^{n} \phi_i(x,y)\, P_i \tag{3.67}$$

we can refer to the total positivity of the $\{\phi_i(x,y)\}$ with respect to x and i or y and i, respectively. As in our development for curves, we are interested in bases $\{\phi_i(x,y)\}$ which are totally positive with respect to both continuous variables.

Let $L_1$ and $L_2$ be some linear operators over the polygonal curves $P = [P_0,\ldots,P_n]$ and $Q = [Q_0,\ldots,Q_n]$, respectively, defined by

$$L_1[P] = \sum_{i=0}^{n} \phi_i(x)\, P_i,\ x\varepsilon R$$

and

$$L_2[Q] = \sum_{i=0}^{m} \psi_i(y)\, Q_i,\ y\varepsilon R$$

where the bases $[\phi_i(x),\ i = 0,\ldots,n]$ and $[\psi_j(y),\ j = 0,\ldots,m]$ are TP on their respective domains and possess the convex hull property. Then given a rectilinear network $R = [P_{ij},\ i = 0,\ldots,n,\ j = 0,\ldots,m]$ we can define

$$L_1[R] = \sum_{i=0}^{n} \phi_i(x)\, P_{ij} \quad \text{for j fixed, } j\varepsilon\,[0,\ \ldots,\ m] \tag{3.68}$$

and

$$L_2[R] = \sum_{j=0}^{m} \psi_j(y)\, P_{ij}, \text{ for } i \text{ fixed, } i\varepsilon\, [0, \ldots, n] . \quad (3.69)$$

$L_1$ and $L_2$ are called lofting operators, reflecting the fact that they filter or vary the shape of R with respect to only one variable. In view of the total positivity of $[\phi_i(x)]$ and $[\psi_j(y)]$, we have

$$V(L_1[R]](x) \leq V[R](i)$$

and

$$V(L_2[R]](y) \leq V[R](j),$$

where we have altered our original notation for the number of sign changes for curves to reflect which domain we are considering. If we wish to smooth in both directions, we construct the tensor product of the two lofting operators as

$$L[R] = L_1[L_2[R]] = L_2[L_1[R]], \quad (3.70)$$

where we have

$$V[L[R]](x) \leq V[L_1[R]](i)$$

$$V[L[R]](y) \leq V[L_2[R]](j)$$

Although these kinds of generalizations from curves to surfaces have proven extremely successful for CAGD [12], and we are dealing formally with two-dimensional surfaces, the approach is inherently one-dimensional. This fact is reflected in the polyhedral network for the lofting operators, and therefore in the tensor product operators, which manifestly require rectilinear control points.

There is a curve to surface generalization, modelled after the convolutional methods for curves (3.56), which circumvents the use of rectilinear networks. Let

$$P = \sum_{i=0}^{n} \phi_i(x,y)P_i \qquad (x,y)\epsilon R^2 \qquad (3.71)$$

where the $P_i$ are arbitrary points of $R^3$ corresponding to knots $p_i = (x_i,y_i)$ of $R^2$ and where the $\{\phi_i\}$ are piecewise linear polynomials of two variables which solve the interpolation problem.

$$\phi_i(x_j,y_j) = \begin{cases} 1 & i = j \\ 0 & \text{elsewhere,} \end{cases} \qquad 0 \leq i, j \leq n .$$

Thus, for distinct $P_i$, P is a <u>proper triangulation</u> of the points $P_i$. (See Figure 3.18.) It is evident that a given polygonally bounded domain in the plane can have several triangulations (Figure 3.19).

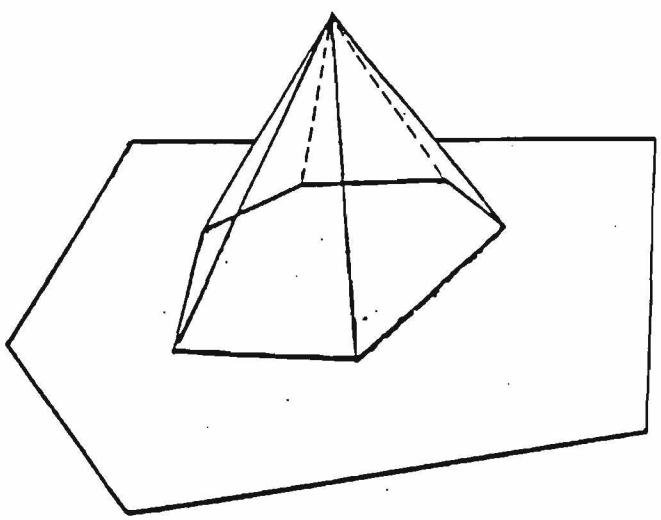

Figure 3.18. (a) Graph of a basis function $\phi_i(x,y)$.

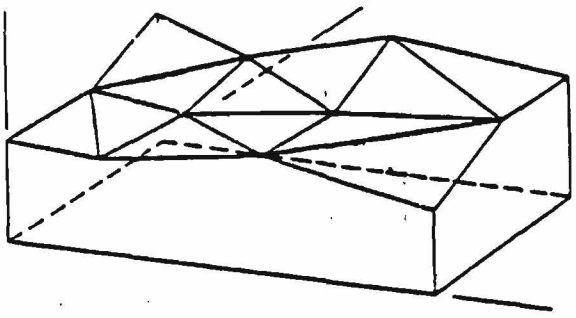

Figure 3.18. (b) Graph of a piecewise linear polynomial.

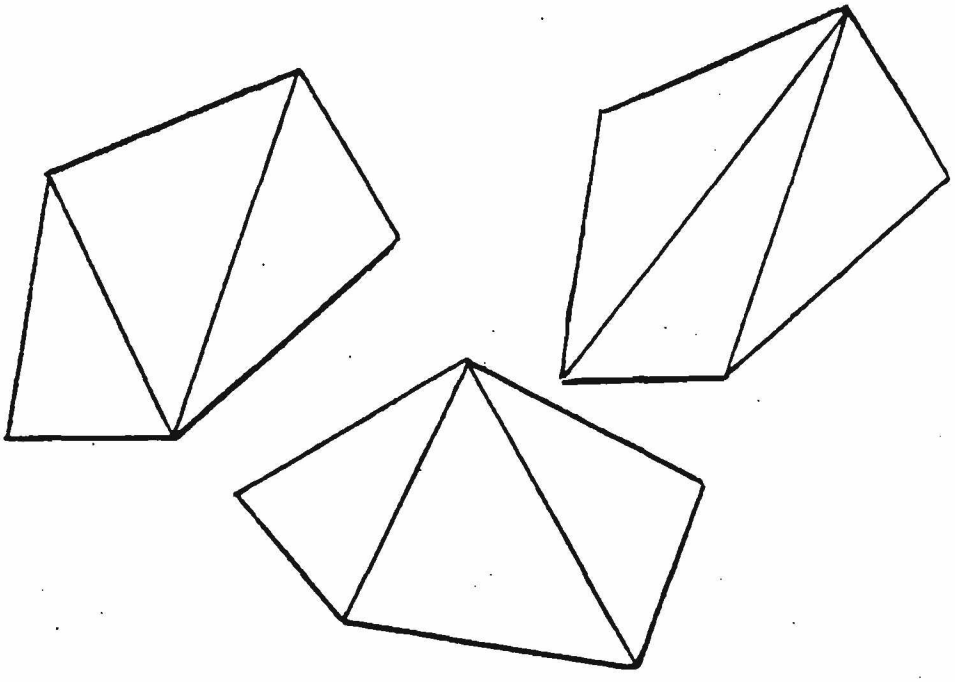

Figure 3.19. Three distinct triangulations of the same polygon.

For various efficient algorithms for triangulation of the plane, see [23,24].

Now for K(x,y) and M(x,y) TP on $R^2$ as in (3.56), where K(x,y) and M(x,y) are translation kernels, define the lofting operators

$$L_1[P] = \int_{-\infty}^{\infty} K(x,s) \left( \sum_{i=0}^{n} \phi_i(s,y)\, P_i \right) ds \quad , \tag{3.72}$$

and

$$L_2[P] = \int_{-\infty}^{\infty} M(x,t) \left( \sum_{i=0}^{n} \phi_i(x,t)\, P_i \right) dt \quad . \tag{3.73}$$

As in (3.68) and (3.69) we have

$$V[L_1[P]](x) \leq V[P](x)$$

and

$$V[L_2[P]](y) \leq V[P](y) \ .$$

In view of the piecewise linear nature of the $\{\phi_i\}$, we have

$$\phi_i'(x,y) = \int_{-\infty}^{\infty} K(x,s)\, \phi_i(s,y)\, ds \tag{3.74}$$

is TP on XxI, and

$$\phi_i''(x,y) = \int_{-\infty}^{\infty} M(y,t)\, \phi_i(x,t)\, dt \tag{3.75}$$

is TP on YxI, where XxY = $R^2$ and I = {0,1,...,n}.

Of course, we can generalize the lofting transformations to the tensor product operation by

$$L[P] = L_1[L_2[P]] = L_2[L_1[P]] \quad .$$

Although we have insisted in (3.70) that the $\phi_i$ represent the cardinal piecewise linear basis functions, the development which follows (3.70) works equally well if the $\phi_i$ represents the piecewise bilinear basis functions associated with the vertices of a rectilinear network. Thus our technique of approximating triangular networks of points is a proper generalization of tensor product approximation to rectilinear networks, encompassing the latter as a special case. Note that in avoiding the biased directions of approximation inherent in the rectilinear schemes, we must choose the directions of approximation. Above we chose to integrate with respect to the x and y directions, but clearly any two independent directions would have sufficed. The following figures show various tensor product approximations to their respective triangular nets for different choices for L and M.

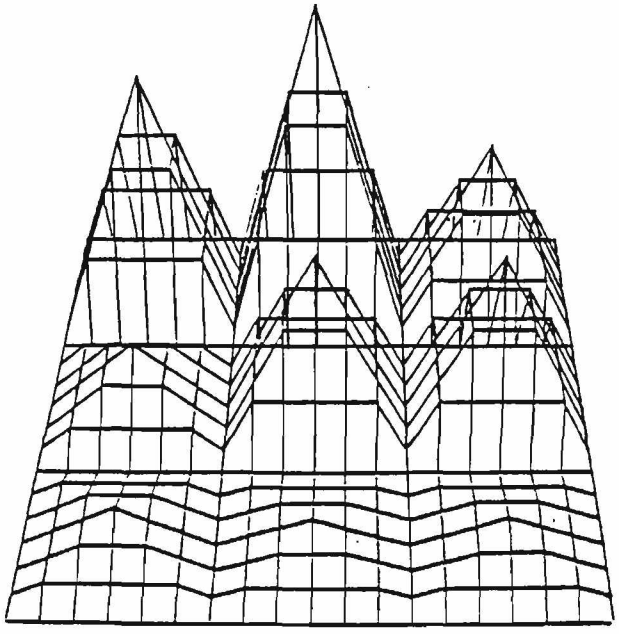

Figure 3.20. Piecewise linear polynomial, T.

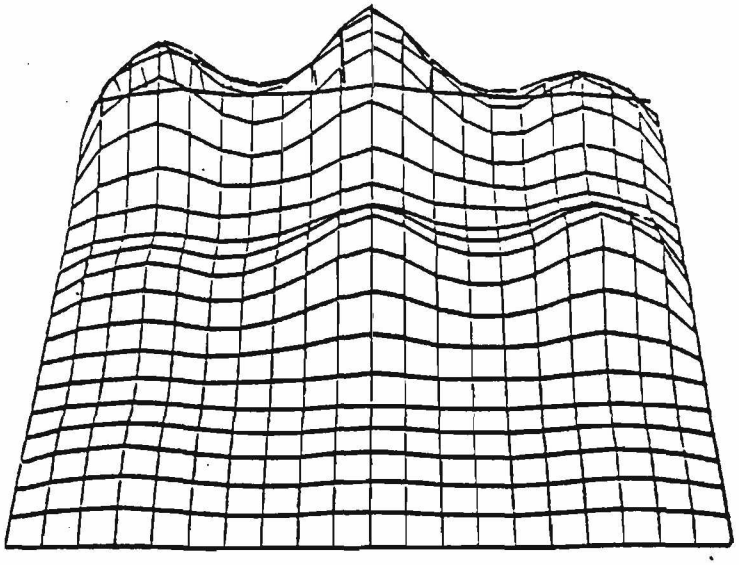

Figure 3.21. $T * 1/(2\pi)\ e^{-(x)^2/2}\ e^{-(u)^2/2}$

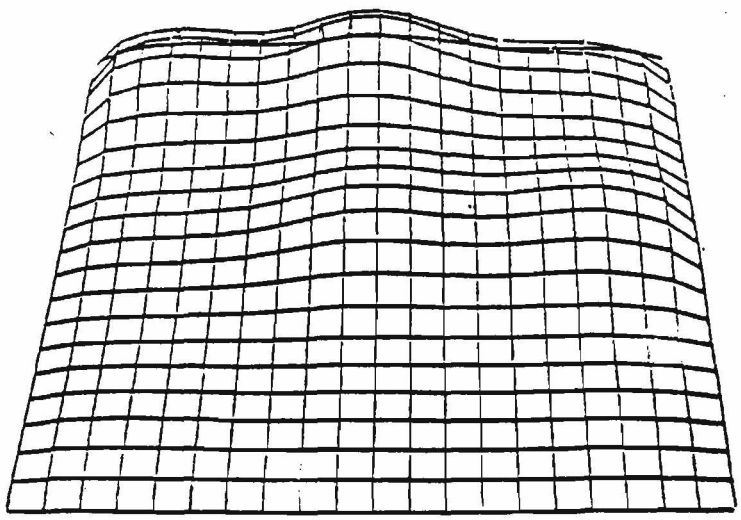

Figure 3.22. $T * (e^{-|x|}e^{-|y|})^2$.

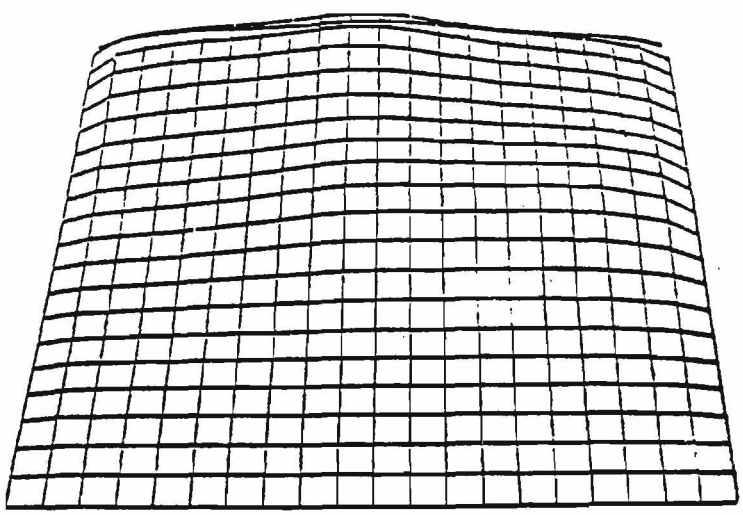

Figure 3.23. $T * (e^{-|x|}e^{-|y|})^4$.

## IV. CONCLUSION

### Summary

The techniques developed and extended here for curve and surface design form a proper generalization of Bernstein-Bézier and B-spline methods, while enjoying increased flexibility for interactive manipulation.

The application of the theory of total positivity to CAGD has not only resulted in well structured approaches to new mathematical modelling for ab initio design, but a new framework for analyzing and understanding existing techniques.

Although this paper has not directly attacked the problems of computability for the new bases, the stability of the methods is inherent in the "bell-shaped" form of the basis functions, as exemplified in their tendency to Gaussian form. Further, the construction (3.18) forms an efficient algorithm for calculating B-splines and can be extended to the other bases constructed in Section III. These construction methods form the core for a class of geometric algorithms for computing the derivative, arc length and intersections of spline curves, as well as the area, volume and intersections of the corresponding spline surfaces.

The feasibility and utility of the newly constructed models for interactive design are demonstrated in the following figures, which

are "frames" from an interactive session on an experimental system for curve and surface design. Since the final decision on what constitutes a "good" design is subjective and will probably vary from user to user, it is felt that the ability developed herein to vary the mathematical model while retaining the desirable variation-diminishing and convex hull properties may be of increased importance in future CAGD systems.

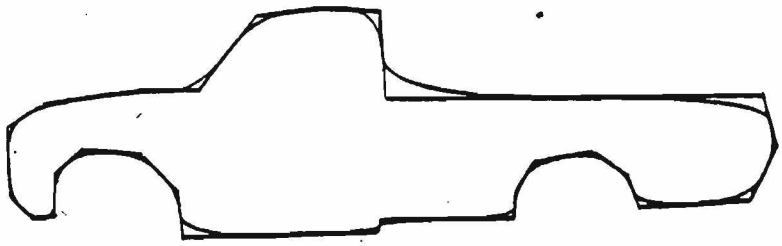

Figure 4.1 (a). Quadratic B-spline approximation to the polygon given.

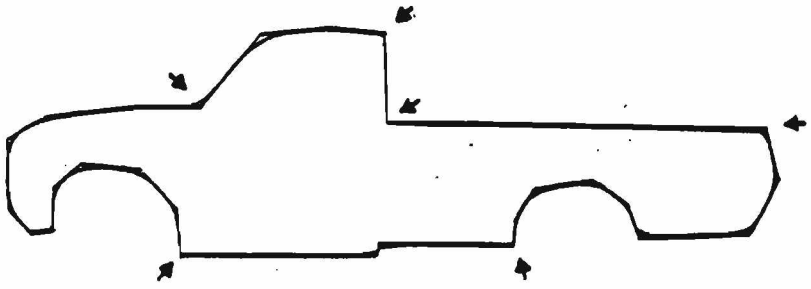

Figure 4.1 (b). Increased tension at the indicated vertices.

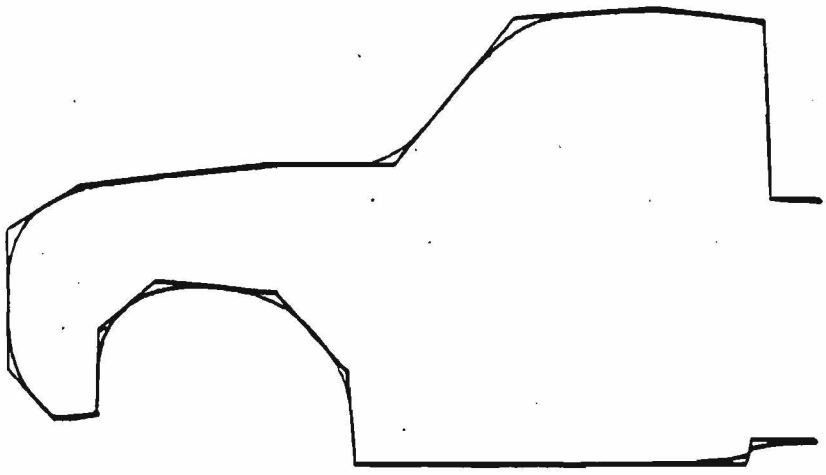

Figure 4.1 (c). Note the local fit to the polygon in this blown-up view.

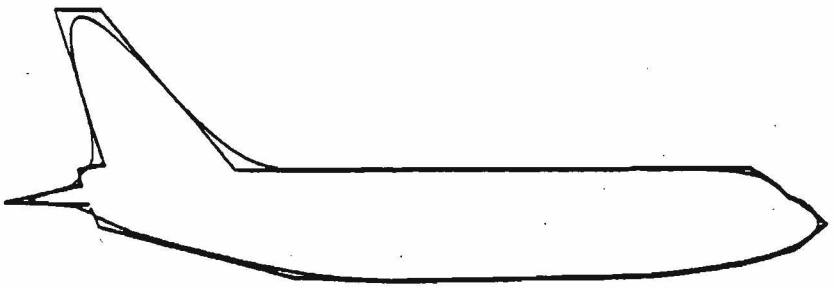

Figure 4.2 (a). Cubic B-spline approximation to the polygon given.

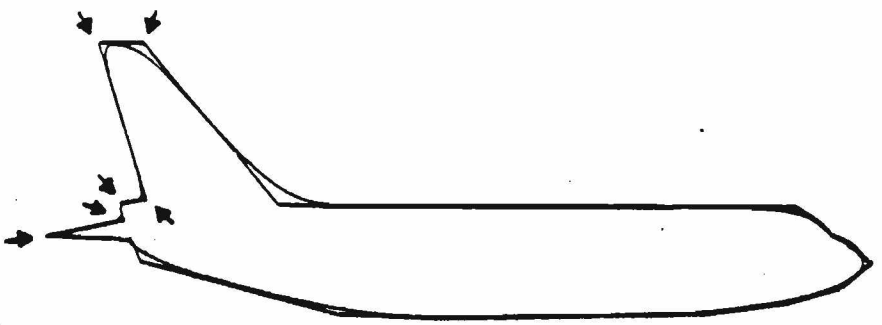

Figure 4.2 (b). Increased tension at the indicated vertices.

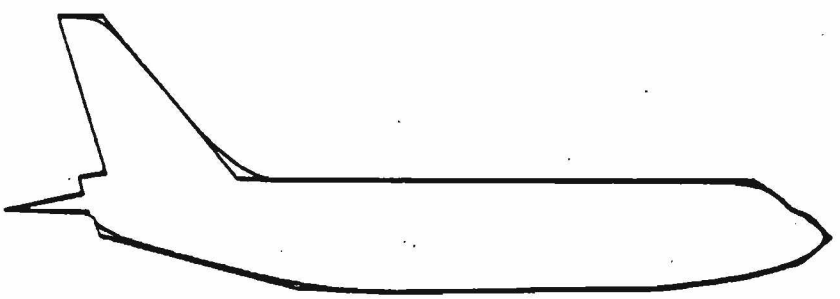

Figure 4.2 (c). Same tension values, but with the basis constructed in (3.50).

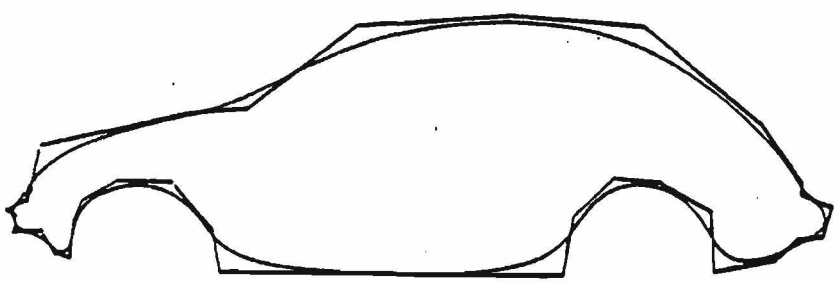

Figure 4.3 (a). Fifth degree B-spline approximation to the polygon given.

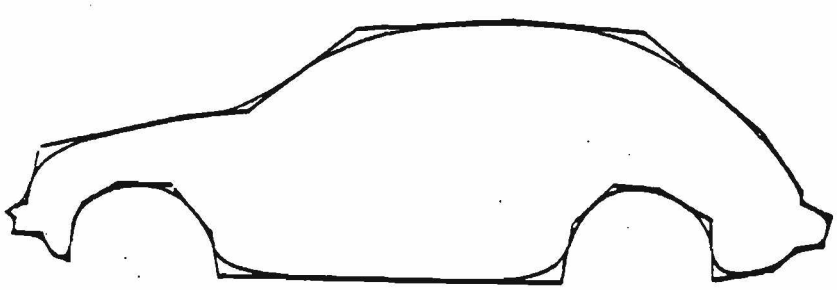

Figure 4.3 (b). Fifth degree spline approximation with the basis constructed from (3.50).

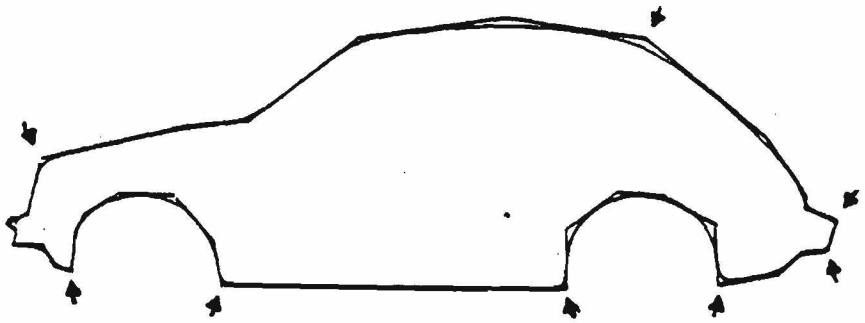

Figure 4.3 (c). Increased tension at the indicated vertices.

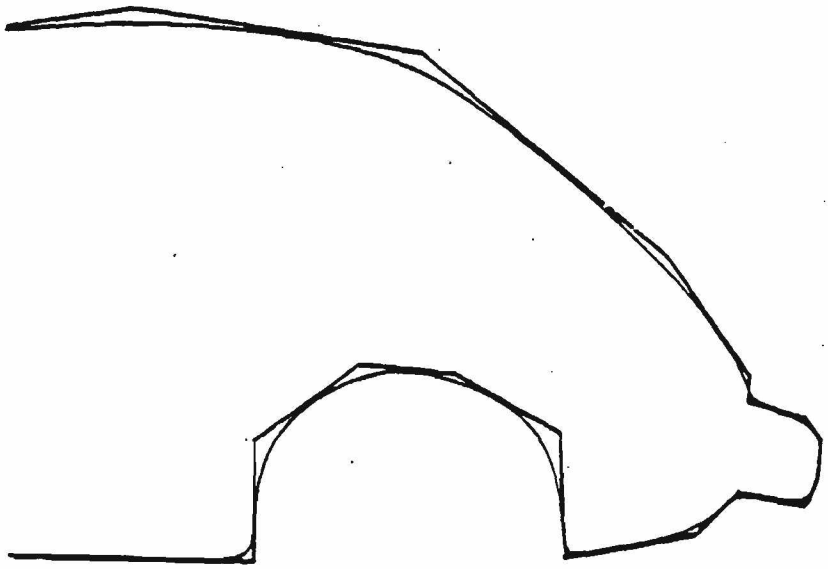

Figure 4.3 (d). Blown-up view, showing changes to the shape of the "bumper".

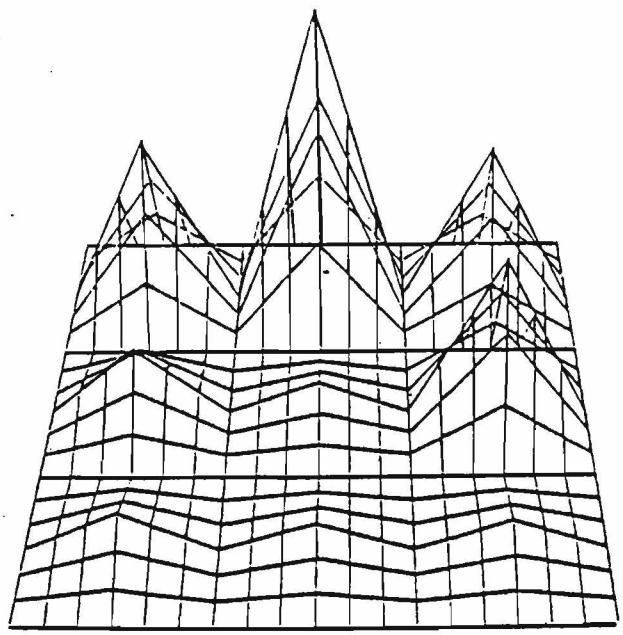

Figure 4.4 (a). A bilinear surface, S.

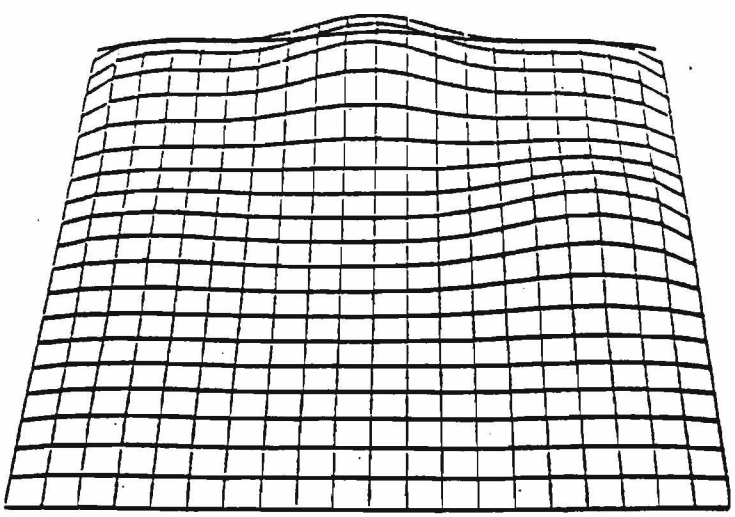

Figure 4.4 (b). A bicubic B-spline approximation to S.

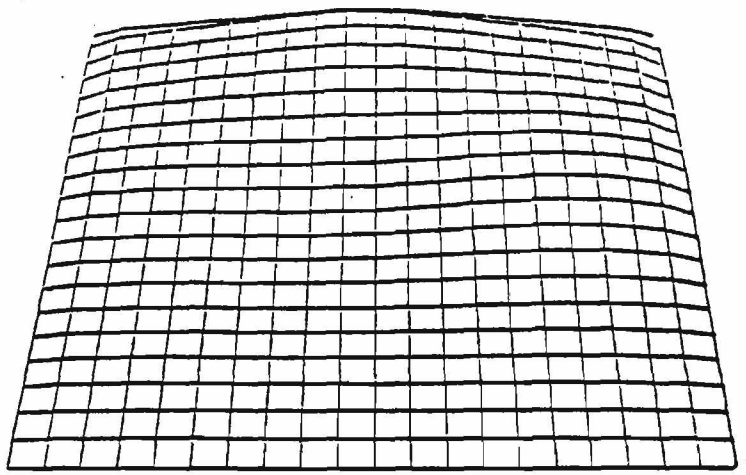

Figure 4.4 (c). A biquintic B-spline approximation to S.

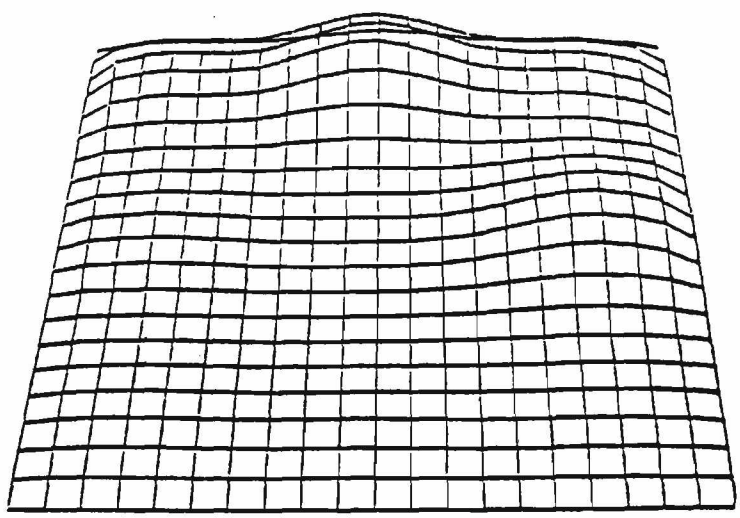

Figure 4.4 (d). $S * (3/2\pi)\, e^{-3x^2/2}\, e^{-3y^2/2}$. Note the similarity in shape to Figure 4.4 (b).

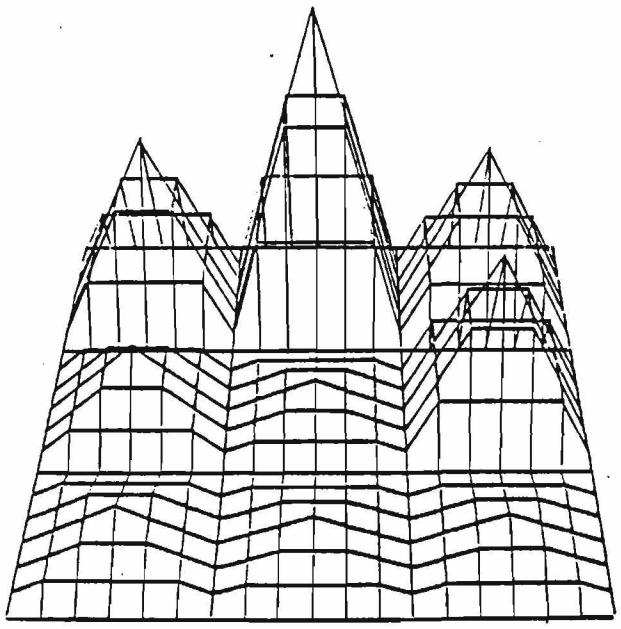

Figure 4.4 (e). A triangulation T of S.

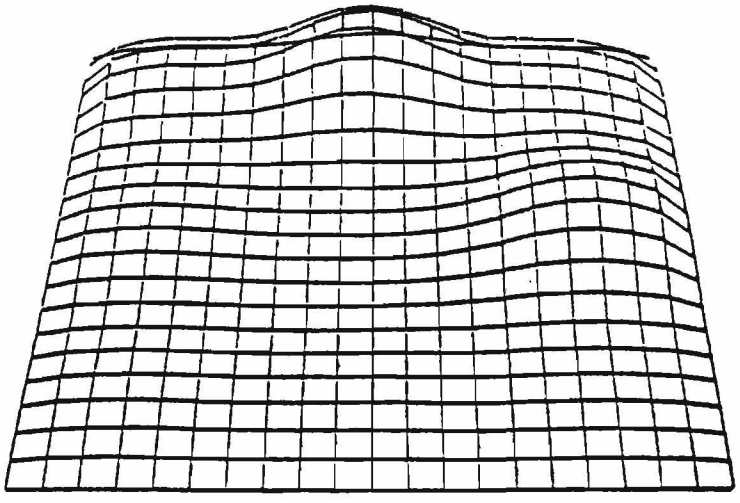

Figure 4.4 (f). $T * (3/2\pi)\, e^{-3x^2/2}\, e^{-3y^2/2}$.

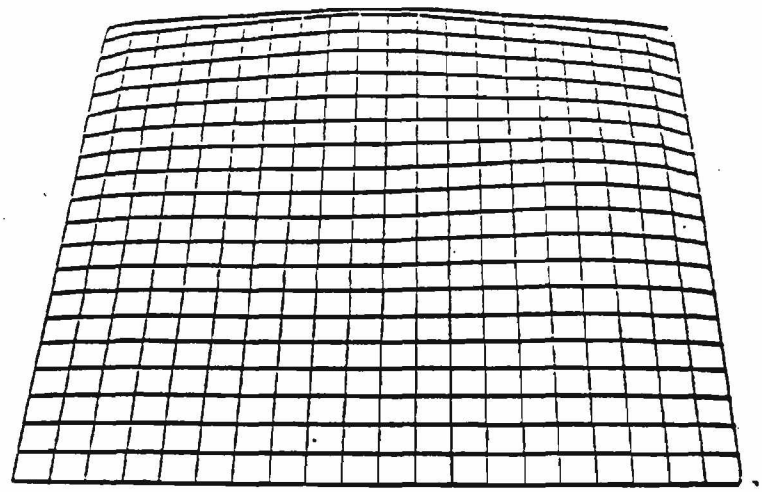

Figure 4.4 (g). $T * (1/2\pi)\, e^{-x^2}\, e^{-y^2}$.

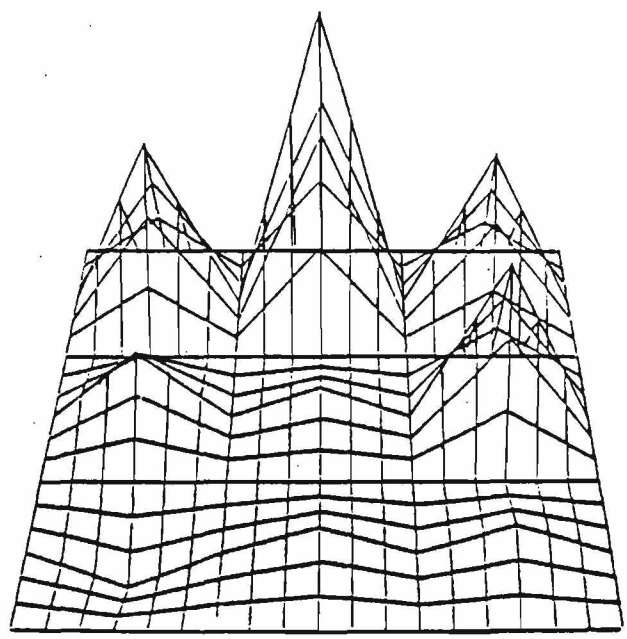

Figure 4.5 (a). The bilinear surface S where the lower left vertex has been moved.

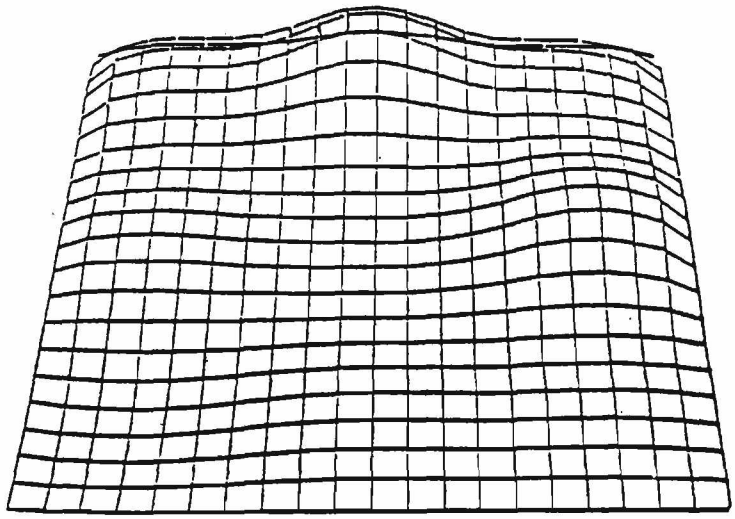

Figure 4.5 (b). Cubic B-spline approximation to S. Note there are only local differences in shape from Figure 4.4 (b).

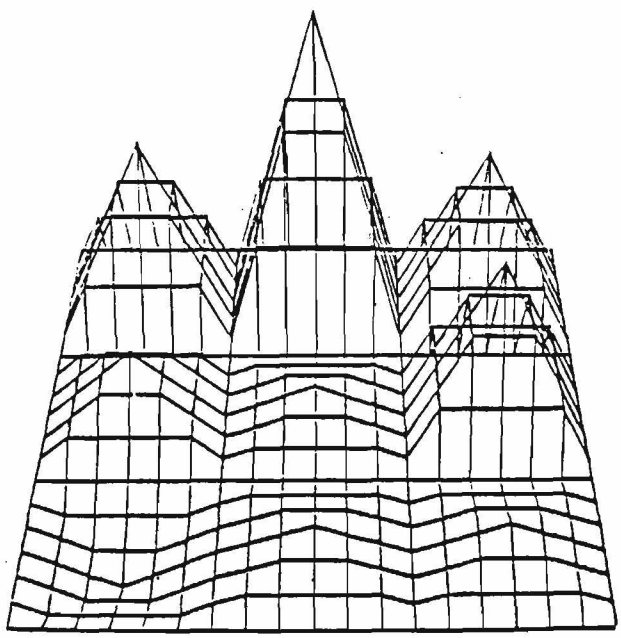

Figure 4.5 (c). A triangulation T of S.

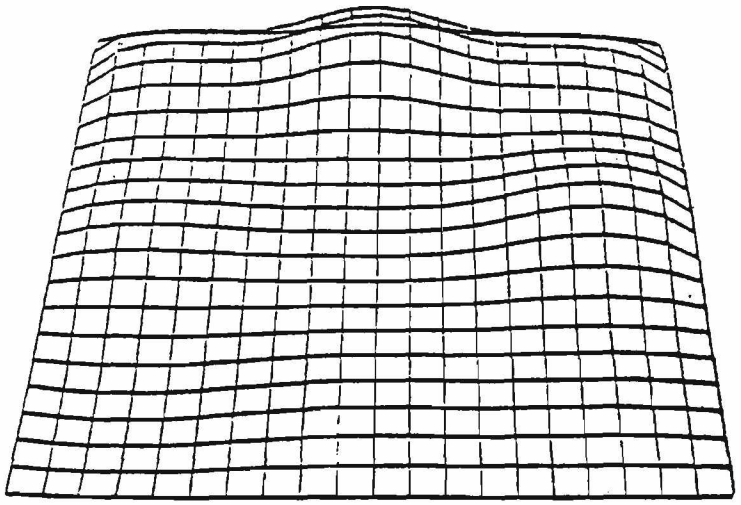

Figure 4.5 (d). $T * (3/2\pi)\ e^{-3x^2/2}\ e^{-3y^2/2}$.
Note there are only local differences in shape from Figure 4.4 (f).