Numer. Math. 51, 441–458 (1987)

Numerische Mathematik

Admissible Slopes for Monotone and Convex Interpolation

Alan Edelman[1] and Charles A. Micchelli[2]

[1] Department of Mathematics, M.I.T., Cambridge, MA 02139, USA
[2] Department of Mathematical Sciences, IBM Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, USA

Summary. In many applications, interpolation of experimental data exhibiting some geometric property such as nonnegativity, monotonicity or convexity is unacceptable unless the interpolant reflects these characteristics. This paper identifies admissible slopes at data points of various C^1 interpolants which ensure a desirable shape. We discuss this question, in turn for the following function classes commonly used for shape preserving interpolations: monotone polynomials, C^1 monotone piecewise polynomials, convex polynomials, parametric cubic curves and rational functions.

Subject Classifications: AMS(MOS): 65D05; CR: G1.1.

Section 1. Introduction

In many applications, interpolation of experimental data exhibiting some geometric property such as nonnegativity, monotonicity or convexity is unacceptable unless the interpolant reflects these characteristics. This has led to many proposals for generating shape preserving interpolants. Some examples of such algorithms appear in Fritsch and Carlson [1], Hyman [3], and Schumaker [5].

Frequently, the design of these algorithms can be conveniently separated into two stages. First, some function class G is specified which exhibits a desirable geometric property on a given interval. The interpolant is then obtained by "stitching" together, on intervals between successive data points, scaled representatives from G. Parameters corresponding to each subinterval are then chosen to yield an acceptable global interpolant to the data.

For C^1 interpolation, it is convenient to adjust the local shape by altering the derivatives at the endpoints of intervals between consecutive data points. In this way left and right derivatives can be easily matched. For this purpose, it is important to know the range of values of the derivatives which admit an acceptable function from G. This question has received only cursory treatment in the literature. Cubic polynomials are considered in [1], the ratio of quadratic polynomials appears in [2] and quadratic splines with one knot in [5]. Although for algorithmic concerns only a few parameters are needed in each interval between data points, a general analysis of this question leads to results of some independent interest which provide insight into the intrinsic limitations of any algorithm based on a given class G.

We will discuss this question, in turn for the following function classes commonly used for shape preserving interpolation: monotone polynomials, C^1 monotone piecewise polynomials, convex polynomials, parametric cubic curves and rational functions. In the latter case, we leave unsettled the region of the admissible slopes for rational functions with one pole and an odd degree polynomial in its numerator. We conjecture that it is an L-shaped region.

Section 2. Monotonic Polynomials

Let G be any subset of $C^1(-1,1)$ with $g(1)=1$ and $g(-1)=-1$ for any $g\in G$. In general, the interpolation problem described above has the form: given $x_0<x_1<\dots<x_n<x_{n+1}$ find an $f\in C^1(x_0,x_{n+1})$ such that

$$f(x_i)=y_i, \quad f'(x_i)=y'_i, \quad i=0,1,\dots,n+1 \tag{2.1}$$

and the linear rescaling of segments of f given by

$$f_i(t)=\frac{2f\left(\frac{(1-t)}{2}x_i+\frac{(1+t)}{2}x_{i+1}\right)-y_i-y_{i+1}}{y_{i+1}-y_i}, \quad t\in[-1,1],$$

is in G for all $i=0,1,\dots,n$. In what follows G is used to control the shape of the interpolant between the data points.

It is possible to give a useful condition for the solvability of (2.1) in terms of the derivative data. To this end, we define

$$D(G)=\{(g'(-1),g'(1)): g\in G\}. \tag{2.2}$$

From the definition of f_i, (2.1) has a solution if and only if for $i=0,1,\dots,n$

$$\left(y'_i\frac{\Delta x_i}{\Delta y_i}, y'_{i+1}\frac{\Delta x_i}{\Delta y_i}\right)\in D(G), \qquad \text{if } \Delta y_i\neq 0 \tag{2.3}$$

where $\Delta x_i=x_{i+1}-x_i$, $\Delta y_i=y_{i+1}-y_i$ (when $\Delta y_i=0, f_i$ is undefined but interpolation on $[x_i,x_{i+1}]$ with a constant is appropriate). Slopes satisfying (2.3) provide acceptable interpolants and we can assess the flexibility of a function class G by how easy it is to fulfill this condition. It should be pointed out that $D(M)=\mathbb{R}^2_+=\{(x,y): x,y\geqq 0\}$ and $D(CM)=\{(x,y): 0\leqq x<1, y>1\}\cup\{(1,1)\}$ where M and CM refer to the totality of C^1 monotone and convex monotone functions mapping $[-1,1]$ onto itself, respectively.

Our main goal in this section is to determine $D(G)$ for G the class of monotonic polynomials mapping $[-1,1]$ onto itself. We denote this set by

$$M_n=\{p: p\in\pi_n, p'(t)\geqq 0 \text{ for } |t|\leqq 1, p(\pm 1)=\pm 1\}$$

and so for $n \geqq 2$

$$D(M_n)=\left\{(p(-1),p(1)) \colon p\in\pi_{n-1}, \frac{1}{2}\int_{-1}^{1} p(t)\,dt=1, p(t)\geqq 0, |t|\leqq 1\right\},$$

(trivially, $D(M_1)=\{(1,1)\}$).

Our first result is

Theorem 2.1. *For $n\geqq 3$, $D(M_n)$ is the convex hull of $(0,0)$ and an ellipse for n odd or a line segment for n even. If $n=2m-1$, the equation of the ellipse is*

$$E_m\colon\ \frac{1}{m^2}(x+y-m^2)^2+\frac{1}{m^2(m^2-1)}(x-y)^2=1 \tag{2.4}$$

while for $n=2m$ the line segment is given by

$$L_m\colon\ x+y=m(m+1), \qquad x,y\geqq 0. \tag{2.5}$$

These ellipses and line segments appear in an interesting arrangement. The line segment L_m separates the ellipses E_m and E_{m+1} (the latter being above L_m) which are all mutually tangent at $\frac{1}{2}m(m+1)(1,1)$. In Fig. 1 below we display these curves for $n=2,\ldots,D(M_2)$ is exactly the line segment L_1. The case $n=3$ appears in [1].

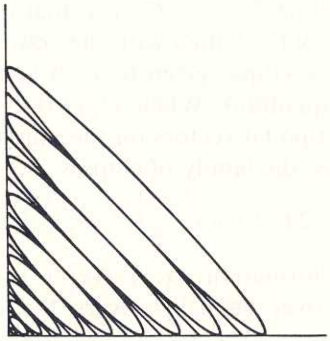

Fig. 1

For the proof of this theorem we will use the following elementary lemma.

Lemma 2.1. *Let a,b be linearly independent vectors in $\mathbb{R}^n$, $n\geqq 3$. The set*

$$E_{a,b}=\{((a\cdot x)^2,(b\cdot x)^2)\colon\ \|x\|=1\}, \qquad \|\cdot\|=\text{Euclidean norm},$$

is the convex hull of $(0,0)$ with the line segment

$$C_{11}x+C_{22}y=1, \qquad x,y\geqq 0, \tag{2.6}$$

when $a\cdot b=0$ or the convex hull of $(0,0)$ with the ellipse

$$(C_{11}x+C_{22}y-1)^2-4C_{12}^2xy=0. \tag{2.7}$$

The constants are given by the matrix

$$C=\begin{pmatrix} a\cdot a & a\cdot b \\ a\cdot b & b\cdot b \end{pmatrix}. \tag{2.9}$$

For $n=2$, $E_{a,b}$ is just the curve (2.6) *when $a\cdot b=0$ or* (2.7) *otherwise.*

Proof. For $n\geqq 3$, the linear transformation $Tv=(a\cdot v, b\cdot v)$ takes the unit sphere $\{v\colon \|v\|=1\}$ in $\mathbb{R}^n$ to the interior of the ellipse

$$C_{11}x^2+2C_{12}xy+C_{22}y^2=1 \tag{2.8}$$

where $C=(C_{ij})$ is a symmetric matrix chosen so that

$$C=A^{-1}, \quad A=\begin{pmatrix} a\cdot a & a\cdot b \\ a\cdot b & b\cdot b \end{pmatrix}. \tag{2.9}$$

When $n=2$, the unit ball is mapped onto the ellipse (2.8). Furthermore every point on (2.8) has a unique preimage in the unit sphere.

Rearranging and then squaring both sides of (2.8) yields after simplifying

$$C_{11}^2x^4+2(C_{11}C_{22}-2C_{12}^2)x^2y^2+C_{22}^2y^4-2C_{11}x^2-2C_{22}y^2+1=0 \tag{2.10}$$

which is an equation in x^2, y^2 for the ellipse $(C_{11}x^2+C_{22}y^2-1)^2-4C_{12}^2x^2y^2=1$ with discriminant $16C_{12}^2(C_{12}^2-C_{11}C_{22})$ so that it is degenerate only when $C_{12}=0$, i.e., when the axes of (2.8) align with the coordinate axes.

It is easy to see that the ellipse given by (2.7) is tangent to the coordinate axes and lies in the first quadrant. When $C_{12}\neq 0$ every point on (2.7) corresponds uniquely to two antipodal vectors on the unit sphere in $\mathbb{R}^n$. When $n\geqq 3$, we see by the same analysis, the family of ellipses

$$C_{11}x^2+2C_{12}xy+C_{22}y^2=r, \quad 0\leqq r\leqq 1,$$

is mapped by the transformation $(x,y)\to(x^2,y^2)$ onto $(C_{11}x+C_{22}y-r)^2=4C_{12}^2xy$ whose union over $r\in[0,1]$ covers the convex set described in Lemma 2.1.

We are now ready to prove Theorem 2.1.

Proof. Case 1: $n=2m-1, m\geqq 2$. First we will look at $D(MS_n)$ where $MS_n=\{p\colon p\in M_n,\ p'=q^2,\ q\in\pi_{m-1}\}$, that is, the set of nondecreasing polynomials whose derivatives are squares. It will turn out that $D(M_n)=D(MS_n)$.

Let P_n be the Legendre polynomials normalized so that $P_n(1)=1$. Then

$$(Iu)(t)=\sum_{k=0}^{m-1}\sqrt{2k+1}\,u_k P_k(t), \quad u=(u_0,\dots,u_{m-1})\in\mathbb{R}^m$$

is an isometry between $\mathbb{R}^m$ and π_{m-1} equipped with the inner product

$$(f,g)=\frac{1}{2}\int_{-1}^{1} f(t)\,g(t)\,dt,$$

[6]. Thus by introducing the vectors

$$a_k=(-1)^k\sqrt{2k+1}, \quad b_k=\sqrt{2k+1}, \quad k=0,\ldots,m-1$$

we have $D(MS_n)=E_{a,b}$. Since $\|a\|=\|b\|=m$ and $a\cdot b=m(-1)^{m-1}$ we conclude in this case

$$C=\frac{1}{m(m^2-1)}\begin{pmatrix} m & (-1)^m \\ (-1)^m & m \end{pmatrix}.$$

Using this equation we can derive the equation of the ellipse determining $E_{a,b}$ from Lemma 2.1. The result of this calculation leads to (2.4).

To conclude the proof it is necessary to show $D(M_n)=D(MS_n)$. From Lukács' lemma [6], if $p\in M_n$ then $q(t)=p'(t)=A^2(t)+(1-t^2)B^2(t)$ where $A\in\pi_{m-1}$, $B\in\pi_{m-2}$. Thus for

$$\alpha=\frac{1}{2}\int_{-1}^{1} A^2(t)\,dt, \quad \beta=\frac{1}{2}\int_{-1}^{1}(1-t^2)B^2(t)\,dt$$

$$q_1(t)=A^2(t)/\alpha, \quad q_2(t)=(1-t^2)B^2(t)/\beta$$

we have α, $\beta\geqq 0$, $\alpha+\beta=1$ and $q=\alpha q_1+\beta q_2$. Since $(q(-1),q(1))=\alpha(q_1(-1), q_1(1))+\beta(0,0)\in D(MS_n)$ we have $D(M_n)\subseteq D(MS_n)$ which proves this part of the theorem.

Case 2: $n=2m$. This case proceeds similarly. By Lukács' lemma, every $p\in M_n$ can be written as $q(t)=p'(t)=(1-t)A^2(t)+(1+t)B^2(t)$, $A,B\in\pi_{m-1}$. This time we expand A,B in Jacobi polynomials $P_k^{(1,0)}(t)$ and $P_k^{(0,1)}(t)$ again normalized so that $P_k^{(1,0)}(1)=P_k^{(0,1)}(1)=1$, (here we use the notation of [6]). Hence for some $u=(u_0,\ldots,u_{2m-1})\in\mathbb{R}^{2m}$

$$q(t)=(1-t)\left[\sum_{k=0}^{m-1}u_k\sqrt{k+1}\,P_k^{(1,0)}(t)\right]^2+(1+t)\left[\sum_{k=m}^{2m-1}u_k\sqrt{k+1-m}\,P_{k-m}^{(0,1)}(t)\right]^2$$

which gives

$$\frac{1}{2}\int_{-1}^{1} q^2(t)\,dt=\|u\|^2$$

$$q(-1)=(a\cdot u)^2, \quad a_k=\begin{cases}\sqrt{2k+2}, & 0\leqq k\leqq m-1\\ 0, & m\leqq k\leqq 2m-1\end{cases}$$

and

$$q(1)=(b\cdot u)^2, \quad b_k=\begin{cases}0, & 0\leqq k\leqq m-1\\ a_{k-m}, & m\leqq k\leqq 2m-1\end{cases}.$$

Thus, as before, $D(M_n)=E_{a,b}$ but now $a\cdot b=0$ and so Lemma 3.1 gives (2.5) because in this case $C_{11}=C_{22}=\dfrac{1}{m(m+1)}$.

In the diagram below we labeled some special points on ellipse E_m.

From this diagram it is easy to see that L_m is tangent to E_m and E_{m+1} at

$$\left(\frac{m(m+1)}{2},\frac{m(m+1)}{2}\right).$$

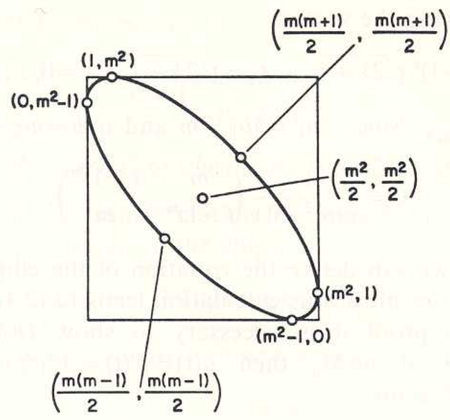

Fig. 2

We mention that it is a consequence of our discussion in the proof of Lemma 2.1 that there is only one $p \in M_n$ such that $(p'(-1), p'(1)) = (\alpha, \beta)$ when $(\alpha, \beta) \in E_m$, $n = 2m-1$ or $(\alpha, \beta) \in L_m$, $n = 2m$. It also may be useful to point out that

$$\left[0, \frac{(m+1)(m+2)}{2}\right]^2 \subset D(M_{2m+1}).$$

This result has been used in designing algorithms for shape preserving interpolation when $m = 1$.

From the above analysis, it is possible to obtain the polynomials corresponding to the boundary curves E_m and L_m. The polynomials in case 1 that are on the boundary are simply those h of the form $\gamma P_{m-1}^{(1,1)} + \delta P_{m-2}^{(1,1)}$ that satisfy $\frac{1}{2} \int_{-1}^{1} h^2(t)\, dt = 1$. We note

$$\frac{1}{2} \int_{-1}^{1} (P_{m-1}^{(1,1)}(t))^2\, dt = \frac{2m}{m+1} \quad \text{and} \quad \int_{-1}^{1} P_{m-1}^{(1,1)}(t)\, P_{m-2}^{(1,1)}(t)\, dt = 0$$

so the condition on h becomes

$$\frac{2m}{m+1} \gamma^2 + 2 \frac{(m-1)}{m} \delta^2 = 1.$$

It is of interest to display h^2 explicitly for the special points labeled on Fig. 2. The intersection of E_m with the line $y = x$ corresponds to the two polynomials

$$\frac{m+1}{2m} P_{m-1}^{(1,1)}(x)^2 \quad \text{and} \quad \frac{m}{2(m-1)} P_{m-2}^{(1,1)}(x)^2.$$

The polynomial

$$h(x) = \frac{m+1}{4(m-1)} P_{m-2}^{(1,2)}(x)^2 (1+x)^2$$

corresponds to the place E_m intersects the y-axis and we have $h(-x)$ for the corresponding x-axis intersection. To see that h has the form specified above, note $(1+x)\,P_{m-2}^{(1,2)}(x)=P_{m-2}^{(1,1)}(x)+\frac{m-1}{m}P_{m-1}^{(1,1)}(x)$. Finally, $P_{m-1}^{(1,0)}(x)^2$ and its reflection about $[-1,1]$ correspond to the topmost and rightmost points on the ellipse respectively. To see the proper form, in this case note that

$$P_{m-1}^{(1,0)}(x)=\frac{m+1}{2m}P_{m-1}^{(1,1)}(x)+\tfrac{1}{2}P_{m-2}^{(1,1)}(x).$$

The positive polynomials (i.e. derivatives of the monotone polynomials) for the line segment L_m are

$$\left\{\frac{1}{2}\frac{m+1}{m}(1+\theta x)\,P_{m-1}^{(1,1)}(x)^2\colon\ -1\leqq\theta\leqq 1\right\}.$$

Taking $\theta=0$ we get the polynomial corresponding to the midpoint of L_m, that is, the point on L_m which is tangent to the ellipse E_m. The choice $\theta=\pm 1$ yields the endpoints of the line segment that lie on the x and y axes.

Section 3. Monotonic Piecewise Polynomials

It is easy to see that whenever π_n is replaced by any finite dimensional subspace $V_n\subseteq C^1[-1,1]$, then for $G_n=\{g\colon g'\geqq 0, g\in V_n\}$ $D(G_n)$ is a compact subset of $\mathbb{R}_+^2$. Thus to exhaust $\mathbb{R}_+^2$ we must use nonlinear families G. For this reason, we consider some examples of such in the remainder of the paper.

In this section, we study π_{nk}, the class of C^1 piecewise polynomials of degree n with at most k knots in $(-1,1)$. The corresponding set of nondecreasing elements of π_{nk} are denoted by M_{nk}. We now consider the set $D(M_{nk})$ which was determined for $k=0$ in Sect. 2.

As a first observation we have

Lemma 3.1. *For $n,k\geqq 2$, $D(M_{nk})=\mathbb{R}_+^2$.*

Proof. The proof is elementary. Let $\alpha,\beta,\gamma,\delta\in\mathbb{R}_+$ and suppose h is the piecewise linear function connecting the points

$$(-1,\alpha),\quad (x_1,\gamma),\quad (x_2,\delta),\quad (1,\beta),\quad -1<x_1<x_2<1.$$

By adjusting these parameters it is easy to determine h so that

$$\frac{1}{2}\int_{-1}^{1}h(t)\,dt=1.$$

Thus the function $G(t)=-1+\int_{-1}^{t}h(\sigma)\,d\sigma$ shows that $(\alpha,\beta)\in D(M_{n2})\subseteq D(M_{nk})$.

The main result of this section is the exact description of $D(M_{n1})$. To this end, we define

$$\tau_n=\begin{cases}m^2, & \text{for } n=2m-1\\ m(m+1), & \text{for } n=2m\end{cases}.$$

Note that according to Theorem 2.1,

$$\tau_n = \max\{x: (x,y)\in D(M_n)\} = \max\{y: (x,y)\in D(M_n)\}.$$

Also, we introduce

$$S_n = \{(x,y): 0 \leqq \min(x,y) < \tau_n\} \cup \{(\tau_n, \tau_n)\}.$$

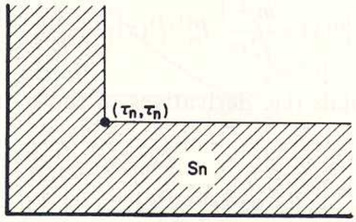

Fig. 3

Theorem 3.1. *For* $n \geqq 3$ $D(M_{n1}) = S_n$.

Proof. First we show $D(M_{n1}) \subseteq S_n$. Let $h \in M_{n1}$. If $h \in \pi_n$ as well then by the definition of τ_n, $(h'(-1), h'(1)) \in S_n$ (see also Fig. 2, when n odd). When $h \notin \pi_n$ there is a knot at some $\alpha \in (-1,1)$. If we set $h(\alpha)=\beta$ and rescale h to the intervals $[-1,\alpha]$, $[\alpha,1]$ then by (2.3) we have

$$\left(h'(-1)\frac{1+\alpha}{1+\beta}, h'(\alpha)\frac{1+\alpha}{1+\beta}\right) \in D(M_n) \tag{3.1}$$

and

$$\left(h'(\alpha)\frac{1-\alpha}{1-\beta}, h'(1)\frac{1-\alpha}{1-\beta}\right) \in D(M_n). \tag{3.2}$$

Thus by the definition of τ_n, $0 \leqq h'(-1)\frac{1+\alpha}{1+\beta} \leqq \tau_n$, and $0 \leqq h'(1)\frac{1-\alpha}{1-\beta} \leqq \tau_n$. Since either $\alpha=\beta$ in which case $h'(-1)$, $h'(1) \leqq \tau_n$ or else one of the numbers $\frac{1+\alpha}{1+\beta}$ and $\frac{1-\alpha}{1-\beta}$ but not both exceed one in which case we conclude either $h'(-1) < \tau_n$ or $h'(1) < \tau_n$ which proves that $(h'(-1), h'(1)) \in S_n$.

Next we show $S_n \subseteq D(M_{n1})$.

Case 1: n even. For any (α,β), $-1<\alpha<1$, $-1<\beta<1$ and $\tau \leqq \tau_n$ there is an $h \in M_{n1}$, $(h'(-1), h'(1)) \in D(M_{n1})$ satisfying

$$h'(\alpha)=0, \quad h(\alpha)=\beta, \quad h'(-1)=\tau_n\frac{1+\beta}{1+\alpha}, \quad h'(1)=\tau_n\frac{1-\beta}{1-\alpha}.$$

This is the case by Theorem 2.1 since such an h satisfies (3.1) and (3.2). Hence

$$\left(\tau_n\frac{1+\beta}{1+\alpha}, \tau_n\frac{1-\beta}{1-\alpha}\right) \in D(M_{n1})$$

which says that for all $\alpha \in (-1,1)$ the closed triangle

$$\Delta_\alpha := \{(x,y):\ 0 \leqq x, y, (1+\alpha)x + (1-\alpha)y \leqq 2\tau_n\}$$

is also in $D(M_{n1})$. Since $\bigcup_{|\alpha|<1} \Delta_\alpha = S_n$ the result is proved.

Case 2: n odd, $n = 2m-1$.

First observe that the choice $h'(\alpha)=1$, $h(\alpha)=\beta=\alpha$, $h'(-1)=h'(1)=\tau_n$ satisfies (3.1) and (3.2) so that $(\tau_n, \tau_n) \in D(M_{n1})$.

Next assume $\alpha \geqq \beta$ and $1 \leqq \tau \leqq \tau_n$ then taking

$$h'(\alpha) = \frac{1+\beta}{1+\alpha}, \quad h(\alpha)=\beta, \quad h'(-1) = \tau\frac{1+\beta}{1+\alpha}, \quad h'(1) = (\tau-1)\frac{1-\beta}{1-\alpha}$$

we see that (3.1) and (3.2) are satisfied since

$$\left(h'(-1)\frac{1+\alpha}{1+\beta}, h'(\alpha)\frac{1-\alpha}{1-\beta}\right) = (\tau, 1) \in D(M_n)$$

and

$$\left(h'(\alpha)\frac{1-\alpha}{1-\beta}, h'(1)\frac{1-\alpha}{1-\beta}\right) = \left(\frac{1-\alpha}{1+\alpha}\frac{1+\beta}{1-\beta}, \tau-1\right) \in D(M_n).$$

See Fig. 4 below.

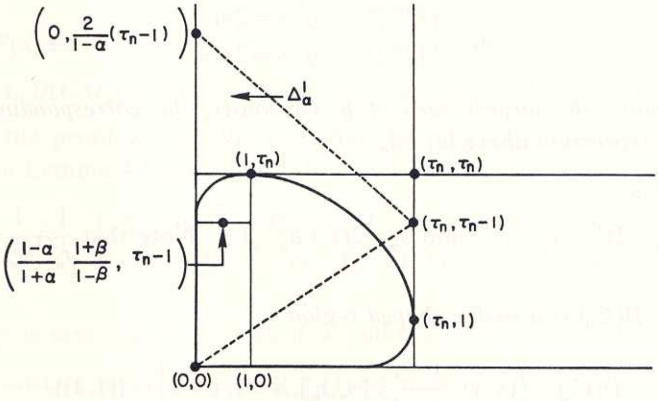

Fig. 4

Thus for all $\alpha \geqq \beta$, $\left(\tau\frac{1+\beta}{1+\alpha}, (\tau-1)\frac{1-\beta}{1-\alpha}\right) \in D(M_{n1})$ which implies that the triangle

$$\Delta_\alpha^1 := \{(x,y): (1+\alpha)x + (1-\alpha)y \leqq 2\tau_n - 1 + \alpha\}$$

is in $D(M_{n1})$ for $|\alpha|<1$. Using the fact that $D(M_{n1})$ is symmetric about the line $x=y$ we also have that the triangle

$$\Delta_\alpha^2 := \{(x,y): (1-\alpha)x + (1+\alpha)y \leqq 2\tau_n - (1-\alpha)\}$$

is in $D(M_{n1})$. Since $\{(\tau_n, \tau_n)\} \cup \bigcup_{|\alpha|<1} \Delta_\alpha^1 \cup \Delta_\alpha^2 = S_n$ the result is proved.

For the sake of completeness we record the easily verified facts: $D(M_{11}) = D(M_{12}) = \{(1,1)\}$ and $D(M_{21}) = \{(x,y): x \geqq 0, y \geqq 0, x+y \leqq 2\}$.

Section 4. Convex Polynomials

We now determine $D(C_n)$, $n \geqq 3$, where C_n is the class of convex polynomials on $[-1,1]$ normalized as follows:

$$C_n = \{p: p \in \pi_n, p''(t) \geqq 0, t \in [-1,1], p(\pm 1) = \pm 1\}.$$

An important subset of C_n is CM_n which we define as the set of monotone convex polynomials mapping $[-1,1]$ onto itself.

The following lemma is all that it takes to describe $D(C_n)$ and $D(CM_n)$.

Lemma 4.1. *Define*

$$a_n = \max \frac{\int_{-1}^{1} x\, p(x)\, dx}{\int_{-1}^{1} p(x)\, dx}$$

where the maximum is taken over all $p \in \pi_n$, $p \geqq 0, p \not\equiv 0$. Then

$$a_n = \begin{cases} \hat{P}_{m+1}^{(0,0)}, & \text{if } n = 2m \\ \hat{P}_{m+1}^{(0,1)}, & \text{if } n = 2m+1 \end{cases}$$

where $\hat{p}$ denotes the largest zero of p. Obviously, the corresponding minimum value of the expression above is $-a_n$.

Proof. See [6].

Define $r_n = 2(1-a_{n-2})^{-1}$ and $s_n = 2(1+a_{n-2})^{-1}$. Note that $\frac{1}{r_n} + \frac{1}{s_n} = 1$ and we have

Theorem 4.1. *$D(C_n)$ is a wedge shaped region*

$$D(C_n) = \left\{(x,y): \frac{x-y}{x-1} \in [s_n, r_n], x < 1, y > 1\right\} \cup \{(1,1)\}.$$

with vertex $(1,1)$ and intersecting the y-axis at $[s_n, r_n]$. The wedge is illustrated in Fig. 5.

Proof. First consider those $g \in C_n$ for which $g'(-1) = 0$. Integrating by parts we have

$$1 = \frac{1}{2} \int_{-1}^{1} g'(t)\, dt = \frac{1}{2} \int_{-1}^{1} (1-t)\, g''(t)\, dt$$

and

$$g'(1) = \int_{-1}^{1} g''(t)\, dt.$$

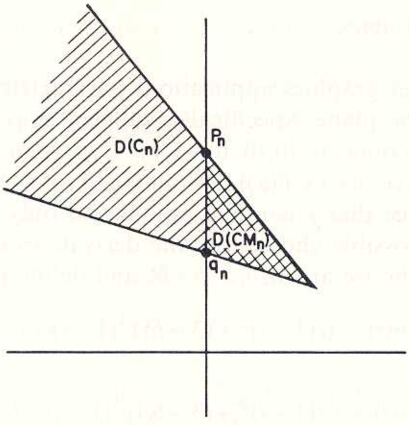

Fig. 5

Using Lemma 4.1 we can conclude that r_n is the maximum value of $g'(1)$ over such g. Similarly q_n the minimum value of $g'(1)$ over such g.

To complete the proof we reduce all elements of C_n to the above form by noting that

$$P \in C_n \Leftrightarrow \frac{x P'(-1) - P(x)}{P'(-1) - 1} \in C_n$$

as long as $P(x) \not\equiv x$.

Corollary 4.1. *$D(CM_n)$ is the triangular region joining* $(0,0)$, $(0, s_n)$, $(1,1)$.

Proof. For the proof we use the fact that the polynomial which achieves the maximum in Lemma 4.1 is given (up to a constant multiple) by

$$p_n(x) = \begin{cases} [P_{m+1}^{(0,0)}(x)(x - \hat{P}_{m+1}^{(0,0)})^{-1}]^2, & \text{if } n = 2m \\ (x+1)[P_{m+1}^{(0,1)}(x)(x - \hat{P}_{m+1}^{(0,0)})^{-1}]^2, & \text{if } n = 2m+1 \end{cases}$$

and thereby is non-negative. Thus, if we take $g_n(x) = -1 + c \int_{-1}^{x} (x-t)\, p_{n-2}(t)\, dt$ where the constant is chosen so that $g_n(1) = 1$ then

$$g_n'(-1) = 0, \quad g_n'(1) = r_n$$

and we conclude that $(0, r_n) \in D(CM_n)$. Similarly, if we define

$$h_n(x) = 1 + 2\, \frac{\int_x^1 (x-t)\, p_{n-2}(t)\, dt}{\int_{-1}^{1} (1+t)\, p_{n-2}(t)\, dt}$$

then $h_n \in C_n$ and $h_n'(-1) = 0$, $h_n'(1) = s_n$. This proves the corollary.

Section 5. Parametric Cubics

Frequently, in computer graphics applications parametric cubic curves are used to connect points in the plane. Specifically, suppose $p, q \in \pi_3$ and that the curve $\{(p(t), q(t)): t \in [0,1]\}$ connects $(0,0)$ to $(1,1)$. If p is monotonic then we can reparametrize the curve as $(x, f(x))$ for $x \in [0,1]$ by setting $f(p(t)) = q(t)$ for $t \in [0,1]$. It is easy to see that f is monotonic if and only if q is monotonic. Let us now consider the possible choices for the derivatives at the endpoints of [0, 1]. For this purpose, choose any $a, b, a', b' \in \mathbb{R}$ and define polynomials

$$p(t) = at(1-t)^2 + (3-b)t^2(1-t) + t^3$$

and

$$q(t) = a't(1-t)^2 + (3-b')t^2(1-t) + t^3$$

expressed in their Bernstein-Bézier representation. Then

$$f'(x) = \begin{cases} a'/a, & \text{at } x = 0, \\ b'/b, & \text{at } x = 1, \end{cases}$$

and $f(x)$ is monotonic, if (a,b) and $(a',b') \in D(M_3)$. Recalling that $[0,3]^2 \subset D(M_3)$ we conclude that

$$\left\{ \left(\frac{a'}{a}, \frac{b'}{b} \right) : (a',b'), (a,b) \in D(M_3) \right\} = \mathbb{R}^2_+.$$

Thus the monotonicity region in this case is the whole first quadrant.

Section 6. Rational Functions

Admissible slopes for monotone rational functions present some interesting possibilities along with several difficulties. The set of functions we now consider is

$$R_{mn} = \begin{cases} f(t) = \dfrac{p(t)}{q(t)}: p \in \pi_m, q \in \pi_n, f(\pm 1) = \pm 1, \\ f'(t) \geqq 0 \text{ and } q(t) \geqq 0 \text{ for } t \in [-1,1] \end{cases} \tag{6.1}$$

i.e. well-defined monotone rationals mapping $[-1,1]$ onto itself. In this section, we determine $D(R_{mn})$ for every pair (m,n) except for the case that m is odd, $m \geqq 5$ and $n = 1$.

In [2] it is shown that $D(R_{mn}) = \mathbb{R}^2_+$ for $m, n \geqq 2$. The following is an improvement of this result.

Proposition 6.1. *$D(R_{mn})$ is the entire monotonicity region $\mathbb{R}^2_+$ for $m \geqq 1$ and $n \geqq 2$.*

Proof. Let $(\alpha, \beta) \in \mathbb{R}^2_+$. Define $p(t) = (\alpha + \beta + 2)t + (\alpha - \beta)$ and $q(t) = (1 - \alpha\beta)t^2 + (\alpha - \beta)t + (\alpha + 1)(\beta + 1)$. We proceed to show that $f(t) = \dfrac{p(t)}{q(t)} \in R_{12}$.

Firstly, that $f(\pm 1) = \pm 1$, is readily verified. Next, since $q(t) = \alpha\beta(1 - t^2) + \alpha(1 + t) + \beta(1 - t) + (1 + t^2)$ we have $q(t) \geqq 1$ on $[-1,1]$. To check $f(t)$ is monotone we observe that after some calculation

$$f'(t)[q(t)]^2=(\alpha+\beta+2)(1-t^2)+2\alpha(1-t)+2\beta(1+t)$$
$$+\alpha\beta(\alpha(t+1)^2+\beta(t-1)^2+2t^2+6)$$

which plainly exhibits $f'(t)\geqq 0$, for $t\in[-1,1]$, $f'(-1)=\alpha$, and $f'(1)=\beta$.

Now we turn to the difficult case of evaluating $D(R_{m1})$. To this end, it is convenient to define for $|\alpha|>1$

$$R^\alpha_{m1}=\left\{f\in R_{m1}\colon f(x)=\frac{p(x)}{x-\alpha},\, p\in\pi_m\right\},$$

i.e. R^α_{m1} consists of those members of R_{m1} with a pole at α. Observe that $\alpha=\infty$ makes sense in this context and $R^\infty_{m1}=M_m$.

As in Sect. 2, we try to characterize derivatives of functions in R^α_{m1}. As a first observation, let $p(x)=(x-\alpha)^2 f(x)$ for some f in R^α_{m1}. It then follows that

$$p\in\pi_m,\quad p(x)\geqq 0,\ x\in[-1,1],\ p'(\alpha)=0,\quad \text{and}\quad \frac{1}{2}\int_{-1}^{1}\frac{p(x)}{(x-\alpha)^2}dx=1. \tag{6.2}$$

Conversely, if $p'(\alpha)=0$ then the indefinite integral $\int\frac{p(x)}{(x-\alpha)^2}dx$ is a rational function with a simple pole at α. Thus (6.2) also gives sufficient conditions that $p(x)=(x-\alpha)^2 f'(x)$ for some f in R^α_{m1}. This leads us to the class of polynomials p that are non-negative on $[-1,1]$ and that satisfy the additional equation $p'(\alpha)=0$. The following proposition is the analogue of Lukács' lemma (use in the proof of Theorem 2.1) for such polynomials.

Proposition 6.2. *Let $S^\alpha_n=\{p\in\pi_n\colon p'(\alpha)=0,\ p(x)\geqq 0$ for $x\in[-1,1]\}$. Every $p\in S^\alpha_n$ has the representation $p(t)=p_1(t)+p_2(t)$ for some $p_1,p_2\in S^\alpha_n$ of the form*

$$p_1(t)=(a-\varepsilon t)A^2(t),\qquad p_2(t)=(1-t^2)(b-\varepsilon t)B^2(t),\qquad \text{if } n=2m+1$$
$$p_1(t)=(1+t)(c-\varepsilon t)C^2(t),\qquad p_2(t)=(1-t)(d-\varepsilon t)D^2(t),\qquad \text{if } n=2m.$$

where $\varepsilon=sign(\alpha)$, $a,b,c,d<|\alpha|$, and all the roots of the polynomials A,B,C, and D are in the interval $[-1,1]$.

Proposition 6.2 is a special case of a general theorem from [4] concerning Tchebycheff systems. A Tchebycheff system of order $n-1$ (T-system) is an n-dimensional vector space of continuous functions defined on a real interval $[a,b]$ with the property that every finite real linear combination of these functions has at most $n-1$ zeros in $[a,b]$.

The following lemma gives a representation theorem for non-negative functions in a T-system.

Lemma 6.1. *Let V be a T-system of order $n-1$ defined on $[-1,1]$. If $f\in V$ and $f(t)\geqq 0$ for $t\in[-1,1]$ then f has the representation $f(t)=p_1(t)+p_2(t)$ where*

$p_1, p_2 \in V$, $p_1(t) \geqq 0$, $p_2(t) \geqq 0$, for $t \in [-1,1]$, $p_2(1)=0$, and p_1, p_2 have $n-1$ zeros in $[-1,1]$.

Proof. [See 4].

We now prove Proposition 6.2 by first checking S_n^α is a T-system of order $n-1$. Indeed, by Rolle's theorem if some $p \in S_n^\alpha$ has n zeros in $[-1,1]$ then p' has all $n-1$ zeros in $[-1,1]$ contradicting $p'(\alpha)=0$. Thus Lemma 6.1 applies and we conclude if $p(t) \geqq 0$ on $[-1,1]$ then $p(t)=p_1(t)+p_2(t)$ where $p_1, p_2 \in S_n^\alpha$, nonnegative on $[-1,1]$, have $n-1$ zeros in $[-1,1]$ and $p_2(1)=0$. Since p_1 and p_2 do not change sign on $[-1,1]$ all their roots inside $(-1,1)$ have even multiplicity. The root not in $[-1,1]$ must be real and by Rolle's theorem must be on the far side of α from $[-1,1]$. Now, if $p(\pm 1)>0$ then taking all this into consideration (6.3) is the only way to account for the zeros. If $p(1)$ or $p(-1)$ vanishes, (6.3) still follows by continuity.

Proposition 6.2 gives useful information about $D(R_{n1}^\alpha)$.

Lemma 6.2. *If n is even, $n \geqq 4$, then $D(R_{n1}^\alpha)$ is a triangle T_α with vertices $(0,0)$, $(0,\zeta(\alpha))$, and $(\zeta(-\alpha),0)$ where*

$$\zeta(\alpha)=\max\{y\colon (0,y) \in D(R_{n1}^\alpha)\}.$$

Proof. Since $D(R_{n1}^\alpha)$ is a compact convex set $\zeta(\alpha)$ is well defined and $(0,\zeta(\alpha)) \in D(R_{n1}^\alpha)$. By observing that $(x,y) \in D(R_{n1}^\alpha)$ if and only if $(y,x) \in D(R_{n1}^\alpha)$ we can check that $\zeta(-\alpha)=\max\{x\colon (x,0) \in D(R_{n1}^\alpha)\}$. Also, since $M_3 \subset R_{n1}^\alpha$ for $n \geqq 4$ we have $(0,0) \in D(R_{n1}^\alpha)$ and so $D(R_{n1}^\alpha) \supseteq T_\alpha$.

To show $D(R_{n1}^\alpha) \subset T_\alpha$, let $f'(x)=\dfrac{p(x)}{(x-\alpha)^2}$ where $f \in R_{n1}^\alpha$. Then from (6.2), $p \in S_n^\alpha$ and so from Proposition 6.2 $p(x)=p_1(x)+p_2(x)$ where $p_1(-1)=0$ and $p_2(1)=0$, $p_1, p_2 \in S_n^\alpha$. Let $c_i=\dfrac{1}{2}\int_{-1}^{1}\dfrac{p_i(x)}{(x-\alpha)^2}dx$, $i=1,2$. Then $\left(0, \dfrac{p_1(1)}{c_1(1-\alpha)^2}\right)$ and $\left(\dfrac{p_2(-1)}{c_2(1+\alpha)^2}, 0\right)$ lie on the edges of T_α and $(f(-1), f(1))$ is a convex combination of these two points.

One general fact about $D(R_{n1}^\alpha)$ when n odd ($n \geqq 3$) is that its extreme points arise from polynomials of the form $p_1(t)$ in (6.3), i.e. $D(R_{n1}^\alpha)$ is the convex hull of $(0,0)$ and

$$\left\{\left(\frac{p_1(-1)}{(1+\alpha)^2}, \frac{p_1(1)}{(1-\alpha)^2}\right)\colon p_1^* \text{ as in (6.3)},\ \frac{1}{2}\int_{-1}^{1}\frac{p_1(x)}{(x-\alpha)^2}dx=1\right\}.$$

To see this consider as before $f'(x)=\dfrac{p(x)}{(x-\alpha)^2} \in R_{n1}^\alpha$. Then we get from (6.3) that $p(x)=p_1(x)+p_2(x)$. Let $c=\dfrac{1}{2}\int_{-1}^{1}\dfrac{p_1^*(x)}{(x-\alpha)^2}dx$, then

$$(f'(-1), f'(1))=\left(\frac{p_1^*(-1)}{(1+\alpha)^2}, \frac{p_1^*(1)}{(1-\alpha)^2}\right)=c\left(\frac{p_1(-1)}{(1+\alpha)^2}\ \frac{p_1(1)}{(1-\alpha)^2}\right)$$

where $p_i(x)=\dfrac{p_i^*(x)}{c}$, $i=1,2$.

We now explicitly describe $D(R_{n1})$ for $n=1,2,3$. The results are tabulated below:

(a) $D(R_{11})=\{(x,y): x\,y=1, x,y>0\}$,

(b) $D(R_{21})=\{(x,y): 0\leqq \min(x,y)<1<\max(x,y)\}\cup\{(1,1)\}$,

(c) $D(R_{31})=$ the convex hull of $(0,0)$ with $\{(x,y)\in\mathbb{R}^2_+:(x-2)(y-2)=1\}$.

We consider each case in turn. The first two are easiest. In particular, all members of R_{11} have the form

$$f_\alpha(x)=\frac{1-\alpha x}{x-\alpha},\qquad |\alpha|>1.$$

(Including $\alpha=\infty$, giving $f_\infty(x)=x$). The derivatives $(f'_\alpha(-1), f'_\alpha(+1)) =\left(\frac{\alpha-1}{\alpha+1},\frac{\alpha+1}{\alpha-1}\right)$ give a branch of the hyperpola $x\,y=1$ which explains (a).

For $D(R_{21})$, we use Proposition 6.2 to conclude that any $f\in R^\alpha_{21}$ has the form

$$(\alpha-x)^2 f'(x)=c_1\left(\frac{\alpha+1}{2}\right)(2\alpha-(x+1))(1-x)+c_2\left(\frac{\alpha-1}{2}\right)(2\alpha-(x-1))(1+x)$$

for some constants $c_1, c_2, c_1+c_2=1$. Thus for any particular α we have that $D(R^\alpha_{12})$ is the line segment connecting $\left(\frac{2\alpha}{\alpha+1},0\right)$ to $\left(0,\frac{2\alpha}{\alpha-1}\right)$. The equation of the corresponding line is $(\alpha+1)\,x+(\alpha-1)\,y=2\alpha$. All these lines have negative slope and pass through the point $(1,1)$, taking their union we get (b).

The case of $D(R_{31})$ is much more interesting. Considering the remark after Lemma (6.2), to obtain $D(R^\alpha_{31})$ we only need consider polynomials of the form $g(x)=\delta(x+a)^2(c-x)$ where $a\in(-1,1)$. To satisfy the equation $g'(\alpha)=0$, we must have $c=\frac{a+3\alpha}{2}$ while $g(x)\geqq 0$, $|x|\leqq 1$ requires that $\operatorname{sign}(\delta)=\operatorname{sign}(\alpha)$. Lastly to insure that $\frac{1}{2}\int_{-1}^{1}\frac{g(x)}{(x-\alpha)^2}dx=1$ we get $\delta=\frac{2(\alpha^2-1)}{a^3+3a^2\alpha+3a+\alpha}$. Thus it follows that $D(R^\alpha_{31})$ is the convex hull of $(0,0)$ and the curve $(x(a), y(a))$, $|a|\leqq 1$ given by

$$\begin{aligned} x&=2(a-1)^2\left(\frac{\alpha-1}{\alpha+1}\right)\left(\frac{a+3\alpha+2}{a^3+3a^2\alpha+3a+\alpha}\right)\\ y&=2(a+1)^2\left(\frac{\alpha+1}{\alpha-1}\right)\left(\frac{a+3\alpha-2}{a^3+3a^2\alpha+3a+\alpha}\right). \end{aligned}\tag{6.4}$$

The parameter a can be eliminated to show that this is a third degree algebraic curve. To see this we introduce the change of variables

$$2u=(\alpha+1)\,x+(\alpha-1)\,y-4\alpha$$

$$2v=-(\alpha+1)\,x+(\alpha-1)\,y-4$$

from which we obtain

$$v^3+9\alpha v^2 u+27\alpha u^3+27u^2 v-108(\alpha^2-1)\,u^2=0. \tag{6.5}$$

Furthermore, observe that the curve given by (6.4) meets the coordinate axes when $a = \pm 1$.

Now consider the expression $(x-2)(y-2)-1$ as a function of a. With some laborious computation we get that

$$(x-2)(y-2)-1 = -\frac{4(2a^2+4a\alpha+3\alpha^2-1)(a^2+2a\alpha+1)^2}{(a^3+3a+3a^2\alpha+\alpha)^2(\alpha^2-1)}.$$

It is not difficult to check that this quantity is negative for any α such that $|\alpha|>1$. Also, for a given $\alpha, |\alpha|>1$, there is a unique $a\in(-1,1)$ for which $a^2+2a\alpha+1=0$. Denoting $H_3=\{(x,y)\in\mathbb{R}_2^+ : (x-2)(y-2)=1\}$ these facts tell us $D(R_{3,1}^\alpha)$ is below H_3 and tangent to it at exactly one point. We illustrate these facts below in Fig. 6.

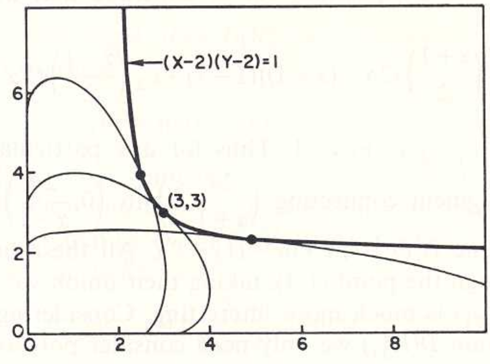

Fig. 6

As $\frac{1}{\alpha}$ varies continuously from -1 to 1 the point of tangency smoothly moves across the entire hyperbola H_3. Thus $D(R_{31}) = \bigcup_{|\alpha|>1} D(R_{31}^\alpha)$ = the convex hull of H_3 with (0,0). Lastly, as $\alpha \to 1$, the left hand side of (6.5) tends to $64(x-2)^3$ while as $\alpha \to -1$ we get $64(y-2)^3$. These are the boundaries of the slabs $0 \leqq x \leqq 2$ and $0 \leqq y \leqq 2$ respectively. Taking $\alpha = \infty$, we get the ellipse E_2.

We now give an exact description of $D(R_{n1})$ for n even. It is of interest to compare this result with Theorem 3.1. In fact, Fig. 3 is appropriate if we now take $\tau_n = m^2$ when $n = 2m, m \geqq 2$.

Theorem 6.1. *For* $n=2m, m\geqq 2$ *we have*

$$D(R_{n1}) = \{(x,y): 0 \leqq \min(x,y) < m^2\}.$$

Proof. First observe that for any $\alpha, |\alpha|>1$, $M_{n-1} \subset R_{n1}^\alpha$. Thus $D(M_{n-1}) \subset D(R_{n1}^\alpha)$. Let p_1 and p_2 be elements of S_n^α corresponding to the endpoints of the hypotenuse of the triangle $D(R_{n1}^\alpha)$ described in Lemma 6.2. From (6.3), we see $p_1(\alpha)p_2(\alpha)<0$ and so some convex combination p of p_1 and p_2 satisfies $p(\alpha)=0$.

Since $p\in S_n^\alpha$, $p'(\alpha)=0$ which insures that $p(x)/(x-\alpha)^2$ is a positive polynomial of degree $n-2$. Thus the hypotenuse of $D(R_{n1}^\alpha)$ is tangent to $D(M_{n-1})$. $D(M_{n-1})$ is the ellipse E_m and these tangents continuously roll along an arc of E_m as $\frac{1}{\alpha}$ varies through $(-1,1)$. Using the description of $D(R_{11})$, given earlier, we conclude as $\alpha\to 1$ that this tangent is the vertical tangent to E_m which is the line $x=m^2$. Similarly, when $\alpha\to -1$ we get the horizontal tangent $y=m^2$. The union of all the regions in between gives $\{(x,y): 0\leqq \min(x,y)<m^2\}$.

For the odd degree case we can only prove

Theorem 6.2. *For $n=2m+1\geqq 5$ we have*

$$D(R_{n1})\supseteq\{(x,y): 0\leqq \min(x,y)<m(m+1)\}.$$

Proof. Let $Q_m(x)=\frac{m+1}{2m}(1-x)^3(P_{m-1}^{(1,1)}(x))^2$ and introduce

$$f_m^\alpha(x)=\frac{\dfrac{Q_m(x)+Q_m'(\alpha)(1-x)}{(x-\alpha)^2}}{\dfrac{1}{2}\displaystyle\int_{-1}^{1}\frac{Q_m(t)+Q_m'(\alpha)(1-t)}{(t-\alpha)^2}dt}.$$

Clearly $(f_m^\alpha(-1), f_m^\alpha(1))\in D(R_{n1}^\alpha)$ and $f_m^\alpha(1)=0$. Since $\lim_{\alpha\to 1} Q_m'(\alpha)=0$ and

$$\lim_{\alpha\to 1}\frac{1}{2}\int_{-1}^{1}\frac{Q_m(x)}{(x-\alpha)^2}dx=\frac{1}{2}\int_{-1}^{1}\frac{m+1}{2m}(1-x)(P_{m-1}^{(1,1)}(x))^2\,dx=1$$

we have $\lim_{\alpha\to 1} f_m^\alpha(-1)=\lim_{\alpha\to 1}\frac{Q_m(-1)}{(\alpha+1)^2}=m(m+1)$. Since $D(R_{11}^\alpha)\subset D(R_{n1}^\alpha)$ we know $\lim_{\alpha\to 1} D(R_{n1}^\alpha)\supseteq\{(x,y): 0\leqq x<m(m+1), y\geqq 0\}$. Similarly,

$$\lim_{\alpha\to -1} D(R_{n1}^\alpha)\supseteq\{(x,y): 0\leqq y<m(m+1), x\geqq 0\}.$$

Furthermore, $D(R_{n1}^\infty)=D(M_n)$ is an ellipse which is entirely contained within $\{(x,y): 0\leqq\min(x,y)<m(m+1)\}$ except for the special case $n=5$ when we also need include the point $(6,6)$. On this basis we conjecture the following:

Conjecture 6.1. *Let $T_m=\{(x,y): 0\leqq\min(x,y)<m(m+1)\}$. We conjecture*

$$D(R_{2m+1,1})=T_m \quad \textit{for } m=3,\dots \textit{ and } D(R_{5,1})=S_4.$$

References

1. Fritsch, F.N., Carlson, R.E.: Monotone piecewise cubic interpolation. SIAM J. Numer. Anal. **17**, 238–246 (1980)
2. Gregory, J.A., Delbourgo, R.: Piecewise rational quadratic interpolation to monotonic data. IMA J. Numer. Anal. **2**, 123–130 (1982)

3. Hyman, J.M.: Accurate monotonicity preserving cubic interpolation. SIAM J. Sci. Stat. Comput. **4**, 645–654 (1983)
4. Karlin, S., Studden, W.J.: Tchebycheff systems: with applications in analysis and statistics. New York: Interscience Publishers 1966
5. Schumaker, L.L.: On shape preserving quadratic spline interpolation. SIAM J. Numer. Anal. **20**, 854–864 (1983)
6. Szegö, G.: Orthogonal polynomials. American Mathematical Society, Providence, R.I. 1939

Received January 12, 1986 / December 12, 1986