SIAM J. SCI. COMPUT.
Vol. 15, No. 3, pp. 528–544, May 1994

003

POSITIVITY CONDITIONS FOR QUARTIC POLYNOMIALS*

GARY ULRICH† AND LAYNE T. WATSON‡

Abstract. Simple necessary and sufficient conditions that a quartic polynomial $f(z)$ be nonnegative for $z>0$ or $a \leq z \leq b$ are derived, and illustrated geometrically. The geometry provides considerable insight and suggests various approximations and computational simplifications. The theory is applied to monotone quintic spline interpolation, giving necessary and sufficient conditions and an algorithm for monotone Hermite quintic interpolation.

Key words. monotone quintic spline interpolation, nonnegative polynomial, polynomial interpolation, quintic Hermite interpolation

AMS subject classifications. 41A05, 65D05, 65D07

1. Introduction. Consider the fourth degree polynomial with real coefficients,

$$f(z) = az^4 + bz^3 + cz^2 + dz + e, \tag{1}$$

where $ae \neq 0$ (the problem reduces to consideration of a cubic if $ae = 0$). This paper outlines conditions under which this polynomial has positivity, i.e., $f(z) \geq 0$ for every $z > 0$. This is a rather general condition since positivity on any fixed interval (u, v) can be directly related to positivity on the positive reals through the transformation

$$t = \frac{u + zv}{1 + z}.$$

The property of positivity has a number of important applications to mathematics and computer science, and to shape preserving polynomial approximations in particular. For example, Jury and Mansour [6] relate positivity to a number of problems in control theory. Fritsch and Carlson [4], who derived positivity conditions in the case of quadratic polynomials, use the result to construct monotone cubic spline interpolants. Schmidt and Heß [8] provided conditions for positivity of cubic polynomials and used their result to construct positive cubic splines with minimum curvature and complex rational cubic splines. For the case of the fourth degree polynomial, Jury and Mansour [6] use the discriminant of (1) along with other characteristic expressions from the theory of equations to derive an algorithm for verifying positivity of quartics. Their conditions are difficult to implement and provide no geometric insight into the underlying mathematical phenomena. Dougherty, Edelman, and Hyman [3] derive conditions for monotonicity and convexity of quintic Hermite interpolants, but

* Received by the editors December 17, 1990; accepted for publication (in revised form) March 12, 1993.

† Systems Technology Center, AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974-0636 (ulrich@ulysses.att.com).

‡ Departments of Computer Science and Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0106. The work of this author was supported in part by Department of Energy grant DE-FG05-88ER25068 and Air Force Office of Scientific Research grant 89-0497 (ltw@vtopus.cs.vt.edu).

explicitly state only sufficient conditions. Our results are very similar to those of [3], but differ in that we give elegant sharp (necessary and sufficient) conditions for positivity of a quartic directly in terms of its coefficients, whereas [3] couches its sufficient conditions in terms of derivatives of Hermite quintics. We also give sharp conditions on the second derivatives, for given fixed first derivatives, such that a quintic Hermite interpolant is monotone; this question is not addressed by [3].

Some time ago, deBoor and Swartz [2] addressed monotone spline interpolation in general and the cubic case in detail, and monotonicity was also recently considered by Huynh [5] and Ulrich and Watson [11].

The present paper uses a simpler, but equivalent, form of the polynomial (1) to obtain positivity conditions which are as elegant as those available for the quadratic and cubic polynomials. The quartic polynomial is reparameterized to this simpler form in §2 where the regions of positivity are described and a formal proof of positivity is provided. The geometric characterization of positivity in §2 provides considerable insight, and easily leads to various approximate criteria and computational simplifications. Section 3 outlines the conditions used to test for positivity and provides some heuristics on constraining a nonpositive polynomial to be positive. Section 4 briefly describes applications of this result in polynomial interpolation.

All of the algebra in this paper was done with Mathematica [12], and thus algebraic derivations showing intermediate steps are not given here.

2. Regions of positivity.

Consider the transformation used in [8], e.g.,

$$x^4 = \frac{a}{e} z^4.$$

This reparameterization should exist since a necessary condition for positivity of the polynomial (1) is that the coefficients a and e are both positive. The polynomial $f(z)/e$ then becomes

$$x^4 + ba^{-3/4}e^{-1/4}x^3 + ca^{-1/2}e^{-1/2}x^2 + da^{-1/4}e^{-3/4}x + 1,$$

which can be written as the polynomial

$$p(x;\alpha,\beta,\gamma) = x^4 + \alpha x^3 + \beta x^2 + \gamma x + 1, \tag{2}$$

where α, β, and γ are defined appropriately. This reparameterized polynomial has only three coefficients. In previous work on the quadratic and cubic equations, the regions of nonpositive roots were bounded by parametric curves corresponding to double roots of the polynomial. The same holds true for the quartic polynomial (2) above. A well-known result in the theory of equations is that double roots exist when the discriminant

$$\Delta = 4\left[c^2 - 3bd + 12ae\right]^3 - \left[2c^3 + 27ad^2 + 27b^2e - 9bcd - 72ace\right]^2$$

of the polynomial (1) is zero (a proof is given in [7]). Setting $\Delta = 0$ leads to the sixth order equation

$$\Delta = 4\left[\beta^2 - 3\alpha\gamma + 12\right]^3 - \left[72\beta + 9\alpha\beta\gamma - 2\beta^3 - 27\alpha^2 - 27\gamma^2\right]^2 = 0. \tag{3}$$

The discriminant for the quadratic and cubic polynomials was a second and fourth degree polynomial, respectively, and region inequalities could be represented and derived in a straightforward manner. In the case of the quartic, the discriminant is a sixth order polynomial and we see no way of approaching the problem directly; we will construct a geometric argument which utilizes the parametric form of the double root boundary. The presence of a double root (say t) implies that

$$p'(t) = 4t^3 + 3\alpha t^2 + 2\beta t + \gamma = 0 \quad \text{and} \quad 4p(t) - t\,p'(t) = \alpha t^3 + 2\beta t^2 + 3\gamma t + 4 = 0,$$

where the latter equation comes from the fact that $1/t$ is also a double root of $x^4\,p(1/x)$. Solving simultaneously for α and γ (for fixed β), we obtain parametric equations for $t \in (-\infty, \infty)$,

$$\alpha(t) = \frac{1 - \beta t^2 - 3t^4}{2t^3} \quad \text{and} \quad \gamma(t) = \frac{t^4 - \beta t^2 - 3}{2t}. \tag{4}$$

Symmetry clearly shows up in these parametric equations since $\alpha(-t) = -\alpha(t)$, $\gamma(-t) = -\gamma(t)$, $\alpha(1/t) = \gamma(t)$, and $\gamma(1/t) = \alpha(t)$. By applying curve tracing techniques, we can identify three distinct shapes (corresponding to zero, one, or two cusps on each component) for the double root curves as shown in Fig. 1 (important features of these curves are also labeled in the figure). Figure 2 shows the family of double root curves as β varies, plotting only the top portions.

The region of positivity for the quartic, like the quadratic and cubic, is bounded by the double root curve. Trying to prove this directly from (4), however, does not work out well. To prove this result, we employ the following theorem (from [1], [9]):

THEOREM 0. *The quartic polynomial $g(x)$ is nonnegative for all $x \geq 0$ if and only if there exist polynomials $u(x)$ and $v(x)$ such that*

$$g(x) = u^2(x) + x\,v^2(x).$$

Rewriting (2) as

$$p(x; \alpha, \beta, \gamma) = \left[x^2 + rx + s\right]^2 + x\left[(\alpha - 2r)x^2 + (\beta - (r^2 + 2s))x + (\gamma - 2rs)\right]$$

for some r and for $s = \pm 1$, then the above theorem implies that $p(x)$ is nonnegative for all $x \geq 0$ if and only if there exists real r such that

$$(\alpha - 2r)x^2 + (\beta - (r^2 + 2s))x + (\gamma - 2rs) \tag{5}$$

is a perfect square. If (5) is a nontrivial perfect square, the first and last coefficients must be positive. The values of r which make (5) a perfect square correspond to the roots of the quartic

$$q(r; \alpha, \beta, \gamma) = \left[\beta - (r^2 + 2s)\right]^2 - 4(\alpha - 2r)(\gamma - 2rs) = 0, \qquad \text{where} \quad \alpha - 2r > 0 \quad \text{and} \quad \gamma - 2rs > 0. \tag{6}$$

(The limiting cases $\alpha - 2r = 0$ and $\gamma - 2rs = 0$ corresponding to $\beta - (r^2 + 2s) = 0$ are also possible and have trivial solutions, so we will not clutter the discussion by mentioning

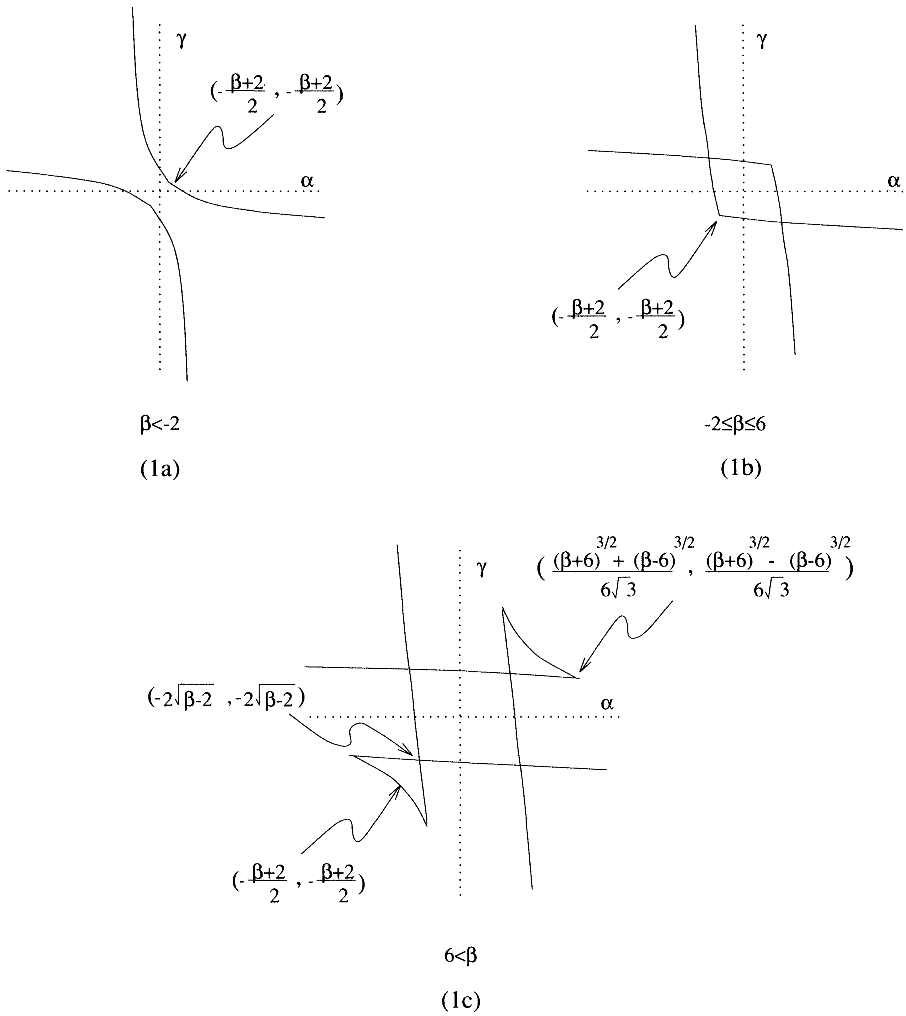


FIG. 1. *Parametric curves from* (4).

these special cases each time.) The discriminant of this quartic is proportional to (3), indicating that the quartic equations (2) and (6) have the same double root boundary; even though the double root boundary is the same for the two quartics (i.e., equations (2) and (6)), their parametric representations differ. Proceeding from (6), as was done from (2) to (4), gives parametric expressions $\alpha(r)$, $\gamma(r)$ in terms of a double root r of (6). Cleaner expressions result from writing the double root as $\pm\operatorname{sgn}(t)\,t$, where t is simply a parameter with restrictions obvious from the structure of the formulas. For $s=1$ and $t^2 \geq 4$, a double root of (6) is $-\operatorname{sgn}(t)\,t$ and the corresponding boundary is parameterized as

$$\alpha(t) = \operatorname{sgn}(t)\left\{-2t + \frac{(t^2-\beta+2)}{4}\left[t+\sqrt{t^2-4}\right]\right\} \tag{7a}$$

and

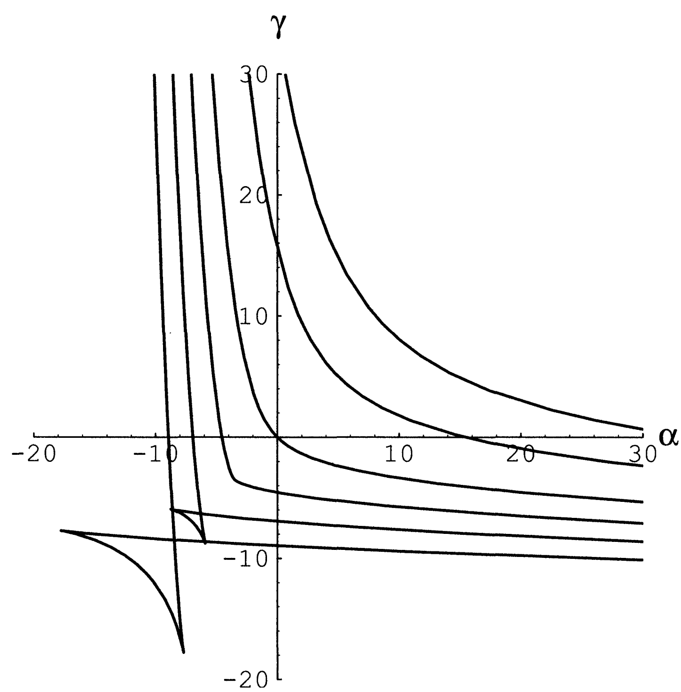


FIG. 2. *Top components of curves from* (4) *for* β=−20, −12, −2, 5, 12, 20 (*top to bottom*).

$$\text{(7b)} \qquad \gamma(t) = \operatorname{sgn}(t)\left\{-2t + \frac{(t^2-\beta+2)}{4}\left[t - \sqrt{t^2-4}\right]\right\}.$$

This curve $(\alpha(t), \gamma(t))$ is the top half of the double root boundary in Fig. 1. The bottom half of the double root boundary corresponds to the double root $\operatorname{sgn}(t)\,t$ and parameterization $(-\alpha(t), -\gamma(t))$.

For $s = -1$ and $t \in (-\infty, \infty)$, the double root is t and the parametric representation of the top half of the double root boundary is

$$\text{(8a)} \qquad \alpha(t) = 2t + \frac{(t^2-\beta-2)}{4}\left[t + \sqrt{t^2+4}\right]$$

and

$$\text{(8b)} \qquad \gamma(t) = -2t - \frac{(t^2-\beta-2)}{4}\left[t - \sqrt{t^2+4}\right].$$

The bottom half of the double root boundary corresponds to the double root t and parameterization $(-\gamma(t), -\alpha(t))$. The substitution of $w^2 = t^2 - 4$ in equation (7) would lead to equation (8), showing that these parameterizations are equivalent. For technical reasons that only become apparent much later, (7) is the cleanest form of the three parametrizations (4), (7), (8) to work with. Only equation (7) is needed for the proof of the following theorem concerning the region of positivity:

THEOREM 1. *For fixed* β, *the region of positivity for the quartic polynomial* (2) *is bounded below by the curve* Γ_β, *which is defined by* (7) *for* $t^2 \geq \max\{4, \beta - 2\}$. *In other words, for any* (α, γ), *the polynomial* $p(x; \alpha, \beta, \gamma)$ *has positivity if and only if there exist* (α^*, γ^*) *on* Γ_β *and* $\delta \geq 0$ *such that* $\alpha = \alpha^* + \delta$ *and* $\gamma = \gamma^* + \delta$.

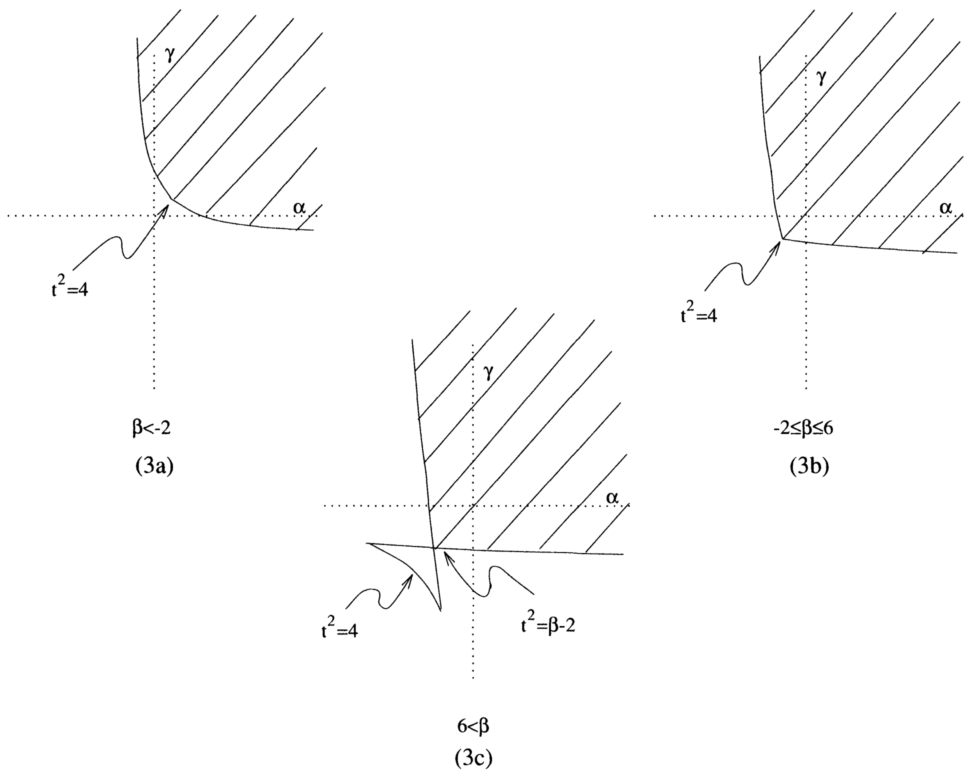


FIG. 3. *Regions of positivity for the quartic polynomial.*

Proof. For any given β, the region of positivity is depicted graphically in Fig. 3. In each case, the parametric curve Γ_β is convex; this is easily proved by looking at the signs of the first and second partial derivatives:

$$\frac{\partial \alpha^*}{\partial \gamma^*} = \frac{\sqrt{t^2-4}+t}{\sqrt{t^2-4}-t} < 0, \quad \text{for} \quad t^2 \geq \max\{4, \beta-2\},$$

$$\frac{\partial^2 \alpha^*}{\partial \gamma^{*2}} = \operatorname{sgn}(t) \frac{4\left(\sqrt{t^2-4}+t\right)}{\left(3t^2-6-\beta\right)\left(t^2-2-t\sqrt{t^2-4}\right)} > 0,$$

$$\text{for} \quad t^2 > \max\{4, \beta-2\} \geq \frac{\beta+6}{3}.$$

(An alternative argument for convexity is that the set of (α, β, γ) for which (2) is nonnegative for all $x \geq 0$ is convex, hence every constant β slice is convex, hence the bounding curve Γ_β must also be convex.) Assume for the moment that $s = 1$. For a fixed β, consider (α', γ') defined by

$$\alpha' = \alpha^*(t) + \delta, \qquad \gamma' = \gamma^*(t) + \delta,$$

for some value of $\delta \geq 0$ and for some t such that $t^2 > \max\{4, \beta - 2\}$, and for $(\alpha^*(t), \gamma^*(t))$ given by equation (7). Since the parametric curve, Γ_β, is part of the double root boundary of equation (6), and since $(\alpha^*(t), \gamma^*(t)) \in \Gamma_\beta$, then q from equation (6) with double root $-\operatorname{sgn}(t)\, t$ can be factored as

$$q(r; \alpha^*(t), \beta, \gamma^*(t)) = \big(r + \operatorname{sgn}(t)\, t\big)^2 \big(r^2 - 2\operatorname{sgn}(t)\, rt + 3t^2 - 2(\beta + 6)\big) \tag{9}$$

and $q(-\operatorname{sgn}(t)\, t; \alpha^*(t), \beta, \gamma^*(t)) = 0$. The double roots satisfy the conditions of equation (6). The other pair of roots,

$$\operatorname{sgn}(t)\, t \pm \sqrt{2(\beta + 6 - t^2)},$$

are complex when $t^2 > \beta + 6$, but do not make the lead and tail coefficients of (5) positive when they are real. Hence, the double roots $-\operatorname{sgn}(t)\, t$ are the only roots satisfying (6) and its side conditions along Γ_β.

We next show that if $p(x; \alpha, \beta, \gamma)$ has positivity, then so does $p(x; \alpha+\delta, \beta, \gamma+\delta)$ for any $\delta > 0$, i.e., moving northeast from a point of positivity (α, γ) preserves positivity. For simplicity, consider first the case of t negative, i.e., $t < \min\{-2, -\sqrt{\beta - 2}\}$ for which we have $\alpha^*(t) \leq \gamma^*(t)$ and $q(t; \alpha^*(t), \beta, \gamma^*(t)) = 0$. It is obvious that $\alpha^*(t) - 2t > 0$ and $\gamma^*(t) - 2t > 0$, which means that the conditions of (6) are satisfied and $p(x; \alpha^*(t), \beta, \gamma^*(t))$ is nonnegative for positive x. Substituting $(\alpha', \beta, \gamma')$ into the function q yields

$$\begin{aligned} q(t; \alpha', \beta, \gamma') &= q(t; \alpha^*(t) + \delta, \beta, \gamma^*(t) + \delta) \\ &= q(t; \alpha^*(t), \beta, \gamma^*(t)) - 4\big[\delta(\alpha^*(t) + \gamma^*(t) - 4t) + \delta^2\big] \\ &= -4\big[\delta(\alpha^*(t) + \gamma^*(t) - 4t) + \delta^2\big]. \end{aligned} \tag{10}$$

When $\delta > 0$, this last term is negative since $\alpha^*(t) + \gamma^*(t) - 4t > 0$. Noting that the $\limsup_{r \to -\infty} q(r; \alpha', \beta, \gamma') = +\infty > 0$, it is clear that there exists a root, say $t' < t$, such that

$$q(t'; \alpha', \beta, \gamma') = 0, \quad \alpha' - 2t' > 0, \quad \gamma' - 2t' > 0.$$

Hence, $p(x; \alpha', \beta, \gamma')$ is nonnegative for positive x. Notice that this argument did not require starting on the double root boundary; it uses the fact that t is a root satisfying equation (6) for a given pair of coefficients, (α, γ), to show there exists a root t_δ which satisfies the equation for the pair $(\alpha + \delta, \gamma + \delta)$, for all $\delta > 0$.

For small $\delta < 0$, equation (10) indicates that the effect of the perturbation is to add a positive increment (in the form of a linear term) near the double roots so that they become complex. This has negligible effect on the other two roots which did not satisfy (6) so that, within a sufficiently small neighborhood below the double root boundary, the conditions of (6) are not satisfied. But this in turn proves that for any (α, γ) below the double root boundary, such that $\alpha < \gamma$, the conditions of (6) are not satisfied; if this were not true we could choose a $\delta > 0$ such that $(\alpha + \delta, \gamma + \delta)$ lay in the neighborhood just described, which would be a contradiction of the argument given in (10).

A similar analysis can be performed for positive t, i.e., $t > \max\{2, \sqrt{\beta - 2}\}$, with similar conclusions for $\alpha > \gamma$. Since Γ_β is convex, every (α, γ) in the region of positivity can be reached from Γ_β through an appropriate choice of $(\alpha^*(t), \gamma^*(t))$ and $\delta > 0$.

We have shown that $p(x; \alpha, \beta, \gamma)$ has positivity for every (α, γ) above the curve Γ_β, and that for (α, γ) below Γ_β, $q(r; \alpha, \beta, \gamma)$ has no roots satisfying (6) for $s = 1$. Thus, all that remains to prove is that for (α, γ) below Γ_β, $q(r; \alpha, \beta, \gamma)$ has no roots satisfying (6) for $s = -1$.

So now assume $s = -1$, β is fixed, $\alpha' = \alpha^*(t) + \delta_1$, $\gamma' = \gamma^*(t) + \delta_2$, $t^2 > \beta + 2$, where $(\alpha^*(t), \gamma^*(t))$ is given by (8). Recall that the double root boundary described by (8) is the same as that given by (7), and thus Γ_β lies on the curve defined by (8). The double root t satisfies (6) if and only if $t^2 \geq \beta + 2$, so only those double roots need be considered. For double root t,

$$q(r; \alpha, \beta, \gamma) = (r - t)^2 \left(r^2 + 2rt + 3t^2 + 12 - 2\beta\right),$$

and the other two roots are

$$-t \pm \sqrt{2(\beta - 6 - t^2)},$$

which are both complex since $t^2 > \beta + 2 > \beta - 6$. Therefore, the double root t is the only root satisfying (6) near the portion of Γ_β corresponding to $t^2 > \beta + 2$. Arguing as before, where now the sign of t does not matter, we conclude that $q(r; \alpha', \beta, \gamma')$ has a root satisfying (6) for any $\delta_1 > 0$ and $\delta_2 > 0$, but not for small $\delta_1 < 0$, $\delta_2 < 0$. If $q(r; \alpha, \beta, \gamma)$ had a root satisfying (6) for (α, γ) below Γ_β, then the analog of (10) would prove that $q(r; \alpha + \delta_1, \beta, \gamma + \delta_2)$ also had a root satisfying (6) for $(\alpha + \delta_1, \gamma + \delta_2)$ arbitrarily near and below the part of Γ_β corresponding to $t^2 > \beta + 2$, for appropriately chosen $\delta_1 > 0$, $\delta_2 > 0$. (It is always possible to choose such δ_i because Γ_β is convex and $\partial\gamma/\partial\alpha < 0$ along Γ_β.) Therefore we conclude that q has no roots satisfying (6) for (α, γ) below Γ_β. $\square$

3. Positivity conditions. Verifying that a point (α, γ) lies above Γ_β is nontrivial using (7), and therefore computationally simple tests are sought for the positivity regions in Fig. 3. A continuity argument implies that the discriminant in (3) changes sign every time a double root boundary is crossed. This leads to the distribution of signs for Δ, in the three types of regions, as shown in Fig. 4. The conditions for positivity in the first case ($\beta < -2$) are simply

$$\Delta \leq 0 \quad \text{and} \quad \alpha + \gamma > 0. \tag{11}$$

For the second case, $-2 \leq \beta \leq 6$, we can use (11) as one of the conditions, but need another condition to cover the middle region with Δ positive. Linear bounds on (α, γ) could be used in this case, but a more compact representation is obtained by using a parabola,

$$\Lambda_1 \equiv (\alpha - \gamma)^2 - 16(\alpha + \beta + \gamma + 2) = 0, \tag{12}$$

which surrounds the positive middle region (see Fig. 5b). The algebra required to show that the parabola indeed lies below the curve Γ_β on the positive middle region is straightforward but tedious (because of convexity, it suffices to check the slopes at

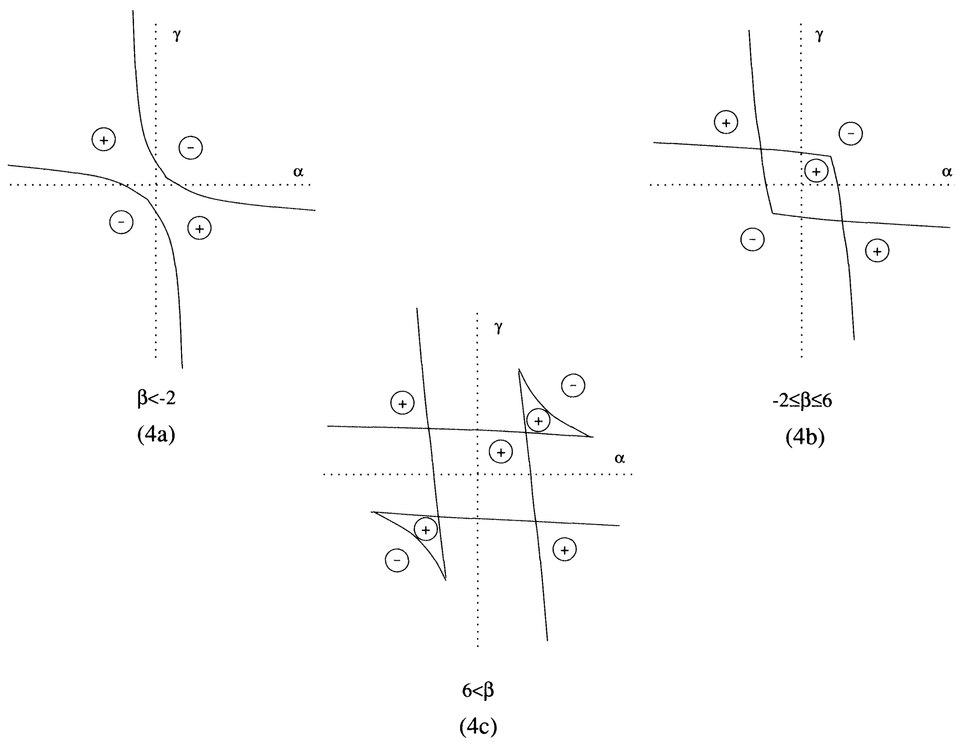


FIG. 4. *Sign distribution for the discriminant* Δ *of the quartic.*

the three points where Λ_1 intersects Γ_β). The resulting pair of conditions in the case of $-2 \leq \beta \leq 6$ are

$$\Delta \leq 0 \quad \text{and} \quad \alpha + \gamma > 0 \qquad \text{or} \qquad \Delta \geq 0 \quad \text{and} \quad \Lambda_1 \leq 0. \tag{13}$$

In a similar way, a parabolic curve can be constructed to bound the middle region (see Fig. 5c) in the final case, $\beta > 6$, leading to the curve:

$$\Lambda_2 \equiv (\alpha - \gamma)^2 - \frac{4(\beta+2)}{\sqrt{\beta-2}}\left(\alpha + \gamma + 4\sqrt{\beta - 2}\right) = 0. \tag{14}$$

This last case is more complicated than the previous one because the parabolic curve will not contain the tips of the cusp region for large β; however, noting that the cusp is always contained in the first quadrant leads to the following three positivity conditions corresponding to $\beta > 6$:

$$\Delta \leq 0 \quad \text{and} \quad \alpha + \gamma > 0 \qquad \text{or} \qquad \alpha > 0 \quad \text{and} \quad \gamma > 0 \qquad \text{or} \qquad \Delta \geq 0 \quad \text{and} \quad \Lambda_2 \leq 0. \tag{15}$$

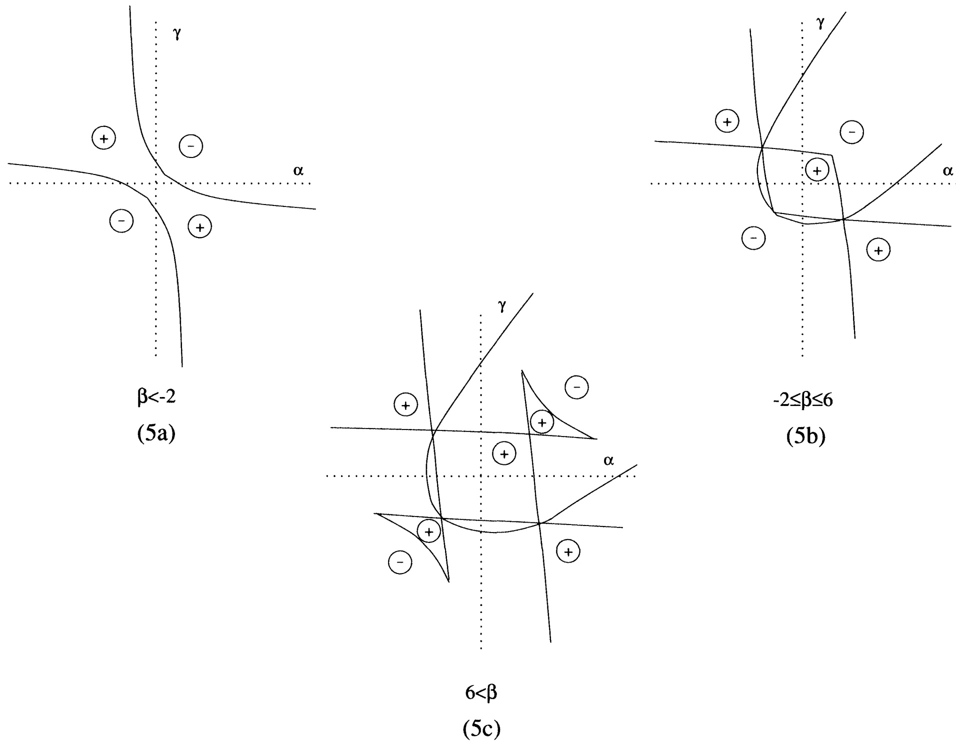


FIG. 5. *Positivity conditions for the quartic polynomial.*

A graphical depiction of these conditions is shown in Fig. 5 for the three cases. The theorem from §2 and these simple tests for positivity can be summarized in:

THEOREM 2. *Let $f(z) = az^4 + bz^3 + cz^2 + dz + e$ be a quartic polynomial with real coefficients and $a > 0$, $e > 0$. Define*

$$\alpha = ba^{-3/4}e^{-1/4}, \qquad \beta = ca^{-1/2}e^{-1/2}, \qquad \gamma = da^{-1/4}e^{-3/4},$$
$$\Delta = 4\left[\beta^2 - 3\alpha\gamma + 12\right]^3 - \left[72\beta + 9\alpha\beta\gamma - 2\beta^3 - 27\alpha^2 - 27\gamma^2\right]^2,$$
$$\Lambda_1 \equiv (\alpha-\gamma)^2 - 16(\alpha+\beta+\gamma+2), \qquad \Lambda_2 \equiv (\alpha-\gamma)^2 - \frac{4(\beta+2)}{\sqrt{\beta-2}}\left(\alpha+\gamma+4\sqrt{\beta-2}\right).$$

Then $f(z) \geq 0$ for all $z > 0$ if and only if

$$(1) \qquad \beta < -2 \quad \textit{and} \quad \Delta \leq 0 \quad \textit{and} \quad \alpha+\gamma > 0;$$

$$(2) \qquad -2 \leq \beta \leq 6 \quad \textit{and} \quad \begin{cases} \Delta \leq 0 \quad \textit{and} \quad \alpha+\gamma > 0 \\ \qquad\quad \textit{or} \\ \Delta \geq 0 \quad \textit{and} \quad \Lambda_1 \leq 0; \end{cases}$$

$$(3) \qquad 6 < \beta \quad \textit{and} \quad \begin{cases} \Delta \leq 0 \quad \textit{and} \quad \alpha+\gamma > 0 \\ \qquad\quad \textit{or} \\ \alpha > 0 \quad \textit{and} \quad \gamma > 0 \\ \qquad\quad \textit{or} \\ \Delta \geq 0 \quad \textit{and} \quad \Lambda_2 \leq 0. \end{cases}$$

These conditions can be costly to implement and a set of simplified sufficient conditions can be used as a pretest for positivity. From Fig. 1, it should be clear that the following conditions are sufficient for positivity:

$$
\begin{aligned}
&\text{(i)} \quad \alpha > -\frac{\beta+2}{2} \quad \text{and} \quad \gamma > -\frac{\beta+2}{2} \quad \text{for} \quad \beta \le 6; \\
&\text{(ii)} \quad \alpha > -2\sqrt{\beta-2} \quad \text{and} \quad \gamma > -2\sqrt{\beta-2} \quad \text{for} \quad \beta > 6.
\end{aligned} \tag{16}
$$

4. Applications. The initial motivation for this work was in the application of Hermite interpolation to construction of random number algorithms for arbitrary continuous distributions. Results from a previous paper [10] suggested that a fifth degree piecewise polynomial approximation could achieve accuracy comparable to that of an exact algorithm implemented in single precision for many common distributions. That same paper outlined an approach for constructing the piecewise polynomial interpolant but indicated that, even though the inverse cdf was monotone, piecewise interpolants using higher order polynomials might not be monotone. The problem of testing for monotonicity of cubic interpolants was solved by Fritsch and Carlson [4], of quartic interpolants by [8], and of quintic interpolants (sufficiency) by and [3] and [11]. Theorem 1 can be used in testing for monotonicity of the quintic interpolant as described below.

Suppose that a quintic interpolant is constructed on the interval (U_0, U_1) to match the ordinates and first and second derivatives of the function $f(U)$, e.g.,

$$
\begin{aligned}
X_0 &= f(U_0), \quad & X_0' &= f'(U_0), \quad & X_0'' &= f''(U_0), \\
X_1 &= f(U_1), \quad & X_1' &= f'(U_1), \quad & X_1'' &= f''(U_1).
\end{aligned}
$$

The fifth degree Hermite polynomial interpolant on (U_0, U_1) can be written as

$$
\begin{aligned}
X(u) = \frac{1}{16}\Big\{ &\big[h^2(X_1'' - X_0'') - 3h(X_0' + X_1') + 3(X_1 - X_0)\big]u^5 \\
&+ \big[h^2(X_0'' + X_1'') - h(X_1' - X_0')\big]u^4 \\
&+ \big[-2h^2(X_1'' - X_0'') + 10h(X_0' + X_1') - 10(X_1 - X_0)\big]u^3 \\
&+ \big[-2h^2(X_0'' + X_1'') + 6h(X_1' - X_0')\big]u^2 \\
&+ \big[h^2(X_1'' - X_0'') - 7h(X_0' + X_1') + 15(X_1 - X_0)\big]u \\
&+ \big[h^2(X_0'' + X_1'') - 5h(X_1' - X_0') + 8(X_0 + X_1)\big]\Big\},
\end{aligned} \tag{17}
$$

where $u = (U - \bar{U})/h$, $\bar{U} = (U_0 + U_1)/2$, and $h = (U_1 - U_0)/2$. The polynomial in (17) is monotone on (U_0, U_1) if and only if its derivative is nonnegative over the interval. Taking the derivative and using the transformation $u = \frac{z-1}{z+1}$ we obtain the test polynomial

$$
Bz^4 + 4(B - D)z^3 + 6(D - C - 2B - 2A + 5)z^2 + 4(A + C)z + A, \tag{18}
$$

where $A = \frac{2hX_0'}{\nu}$, $B = \frac{2hX_1'}{\nu}$, $C = \frac{h^2X_0''}{\nu}$, $D = \frac{h^2X_1''}{\nu}$, and $\nu = (X_1 - X_0)$; the quintic in (17) is monotone if and only if (18) is nonnegative for $z > 0$. (The special case $AB = 0$

reduces to a cubic, and will not be considered further. Thus we assume $A > 0$, $B > 0$ henceforth.)

This quartic can be reparameterized to fit the form of equation (2) by defining

$$\alpha = \frac{4(B-D)}{A^{1/4}B^{3/4}}, \qquad \beta = \frac{6(D-C-2B-2A+5)}{A^{1/2}B^{1/2}}, \qquad \gamma = \frac{4(A+C)}{A^{3/4}B^{1/4}}. \tag{19}$$

Theorem 2 can be applied to these coefficient values to determine whether the polynomial in (17) is monotone. If the quintic is not monotone then we can adjust the derivative values to make it monotone over (U_0, U_1). One strategy is to scale the derivative vector $(X_0', X_1', X_0'', X_1'')$ by an appropriate factor $\rho \in (0,1)$, e.g.,

$$\rho\big(X_0', X_1', X_0'', X_1''\big),$$

which will always lead to monotonicity for ρ small enough. To see this, note that, after scaling, the coefficients in (19) become

$$\alpha = \frac{4(B-D)}{A^{1/4}B^{3/4}}, \qquad \beta = \frac{6\left(D-C-2B-2A+\frac{5}{\rho}\right)}{A^{1/2}B^{1/2}}, \qquad \gamma = \frac{4(A+C)}{A^{3/4}B^{1/4}},$$

and decreasing ρ increases β without affecting α or γ. The nature of the positivity regions is such that increasing β will eventually satisfy the conditions of Theorem 2, for fixed α and γ. In particular, equation (16ii) can be used to provide an approximate value for ρ, e.g., choose ρ as

$$\rho \approx \frac{120}{\sqrt{AB}(\delta^2+8)+24(2A+2B+C-D)}, \qquad \text{where } \delta = \min\{\alpha, \gamma\}.$$

A more satisfying strategy for constraining (17) to be monotone would involve adjusting only the second derivatives. For much of what follows, we will assume that the values of A and B are fixed. If a point $\eta_0 = (C_0, D_0)$ could be located within the interior of the monotonicity region, for fixed (A, B), then (17) would be monotone for (A, B, C^*, D^*), where

$$(C^*, D^*) = \rho\,(C, D) + (1-\rho)(C_0, D_0) \tag{20}$$

for sufficiently small ρ.

With fixed (A, B), using (19) and plotting $\Delta = \Delta(C, D) = 0$ in the (C, D) plane gives teardrop-shaped regions for the permissible values of (C, D) [3], [11]. The discussion below, based on [11], derives entirely from translating the positivity region in (α, β, γ) space to a region in (A, B, C, D) space.

We establish the notation $\eta = (C, D)$ for fixed (A, B), and consider the points (determined in conjunction with a previous paper [11] concerned with monotonicity of the quintic Hermite polynomial) defined by

$$
\begin{aligned}
\eta_1 = (C_1, D_1) &= \left(\frac{-\sqrt{A}}{4} \left(7\sqrt{A} + 3\sqrt{B} + \sqrt{5\left(24 + 2\sqrt{AB} - 3(A+B)\right)} \right), \right. \\
&\qquad \left. \frac{\sqrt{B}}{4} \left(3\sqrt{A} + 7\sqrt{B} + \sqrt{5\left(24 + 2\sqrt{AB} - 3(A+B)\right)} \right) \right), \\
\eta_2 = (C_2, D_2) &= \left(\frac{-\sqrt{A}\left(6A + 3B - 15 + 2\sqrt{AB} - 4A^{3/4}B^{1/4}\right)}{3(\sqrt{A}+\sqrt{B}) - 4(AB)^{1/4}}, \right. \\
&\qquad \left. \frac{\sqrt{B}\left(3A + 6B - 15 + 2\sqrt{AB} - 4A^{1/4}B^{3/4}\right)}{3(\sqrt{A}+\sqrt{B}) - 4(AB)^{1/4}} \right), \\
\eta_3 = (C_3, D_3) &= \left(\frac{-\sqrt{A}}{4} \left(7\sqrt{A} + 3\sqrt{B} - \sqrt{5\left(24 + 2\sqrt{AB} - 3(A+B)\right)} \right), \right. \\
&\qquad \left. \frac{\sqrt{B}}{4} \left(3\sqrt{A} + 7\sqrt{B} - \sqrt{5\left(24 + 2\sqrt{AB} - 3(A+B)\right)} \right) \right).
\end{aligned}
\tag{21}
$$

These points lie along the line $(\gamma = \alpha)$

$$\sqrt{B}(A + C) = \sqrt{A}(B - D), \tag{22}$$

and we will show that this line passes through the teardrop-shaped region of (C, D) values for which equation (17) is monotone, for fixed (A, B). Also consider Regions I and II, depicted in Fig. 6, showing values of A and B (both positive) over which the η_i will provide endpoints bounding the segment of (C, D) values along (22) for which the Hermite interpolant is monotone. Note that the Region II boundary is the same as the one defined in Fig. 4.2 of [3]. If we define $\alpha(\nu)$, $\beta(\nu)$, $\gamma(\nu)$ to be the coefficients in (19) evaluated at $\nu = \eta_1$, η_2, or η_3, then we can prove the following results:

THEOREM 3. *The coefficients in* (19) *exhibit the following properties*:

i. $\eta_2 = \eta_3$ *along the boundary between Regions* I *and* II;

ii. $\beta(\eta_1) > 6$ *over Regions* I *and* II;

iii. $\beta(\eta_2) \le 6$ *over Region* I *and* $\beta(\eta_2) > 6$ *over Region* II;

iv. $\beta(\eta_3) \le 6$ *over Region* I *and* $\beta(\eta_3) > 6$ *over Region* II;

v. $\alpha(\eta_1) = \gamma(\eta_1) = -2\sqrt{\beta(\eta_1) - 2}$;

vi. $\alpha(\eta_2) = \gamma(\eta_2) = -\dfrac{\beta(\eta_2) + 2}{2}$;

vii. $\alpha(\eta_3) = \gamma(\eta_3) = -2\sqrt{\beta(\eta_3) - 2}$.

The proof of these results involves tedious, but straightforward, algebra and therefore will not be given here. The points η_1, η_2, η_3 provide the endpoints to intervals

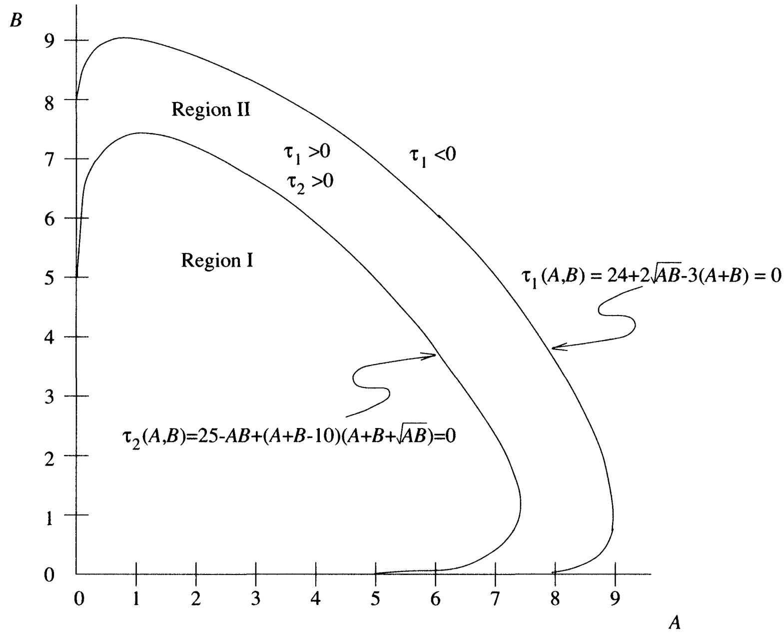


FIG. 6. *Boundary regions on* $\eta(A,B)$.

of monotonicity for η along the line given by equation (22). Note that η_1 and η_3 are complex for (A, B) beyond Region II, a reflection of the fact that equation (17) cannot be monotone for (A, B) outside of Regions I and II. The endpoint relationships can be summarized in the following theorem:

THEOREM 4. *Let* (A, B) *be fixed and contained within Region* I *or* II. *If* $\eta_0 = (C_0, D_0)$ *is any point on the line given by* (22), *then* (17) *is monotone for* (A, B, C_0, D_0) *if and only if*

i. (A, B) *is in Region* I *and* $\eta_0 = \rho\eta_1 + (1-\rho)\eta_2$ *for some* $\rho \in [0, 1]$, *or*
ii. (A, B) *is in Region* II *and* $\eta_0 = \rho\eta_1 + (1-\rho)\eta_3$ *for some* $\rho \in [0, 1]$.

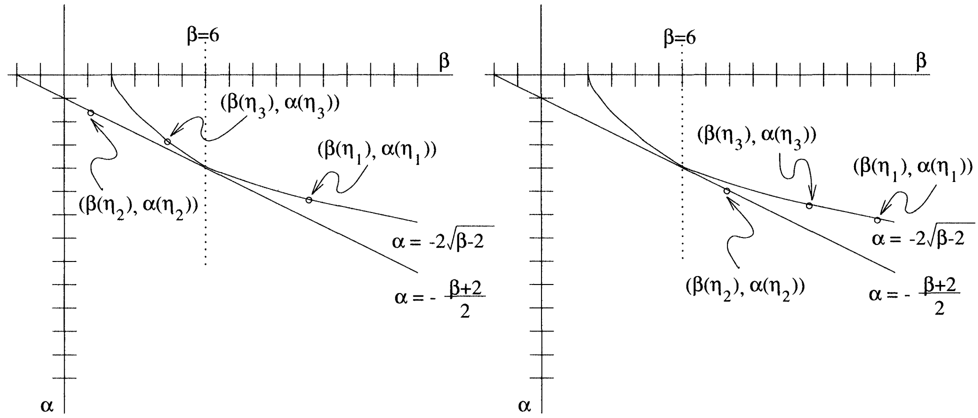
Proof. Sufficiency. Note first that $\alpha(\eta_0) = \gamma(\eta_0)$ since η_0 lies along the line given by (22). Consider first the case of (A, B) in Region II. According to (16), we need to show that $\alpha(\eta_0) \geq -2\sqrt{\beta(\eta_0) - 2}$. Clearly, for fixed (A, B), $\alpha(\eta)$ and $\beta(\eta)$ are affine in η so that (using Theorem 3)

$$\begin{aligned}\alpha(\eta_0) &= \rho\alpha(\eta_1) + (1-\rho)\alpha(\eta_3) = \rho\big(-2\sqrt{\beta(\eta_1) - 2}\big) + (1-\rho)\big(-2\sqrt{\beta(\eta_3) - 2}\big)\\ &\geq -2\sqrt{\big(\rho\beta(\eta_1) + (1-\rho)\beta(\eta_3)\big) - 2} \quad \text{by convexity of } -2\sqrt{x}\\ &= -2\sqrt{\beta(\eta_0) - 2}.\end{aligned}$$

Figure 7b depicts the relationship established in the above inequality.

For the case of (A, B) in Region I, the proof is similar except that we have to consider the convex function defined by

$$\omega(x) = \begin{cases} -\dfrac{x+2}{2}, & x \leq 6;\\ -2\sqrt{x-2}, & x > 6.\end{cases}$$



(7a) (A,B) in region I. (7b) (A,B) in region II.

FIG. 7. *Coefficients evaluated at η values.*

Again using Theorem 3 and (16), as Fig. 7a suggests, we have

$$\begin{aligned}\alpha(\eta_0) &= \rho\alpha(\eta_1) + (1-\rho)\alpha(\eta_2) = \rho\bigl(-2\sqrt{\beta(\eta_1)-2}\bigr) + (1-\rho)\left(-\frac{\beta(\eta_2)+2}{2}\right)\\ &= \rho\,\omega(\beta(\eta_1)) + (1-\rho)\omega(\beta(\eta_2))\\ &\geq \omega\bigl(\rho\,\beta(\eta_1) + (1-\rho)\beta(\eta_2)\bigr) \quad \text{by convexity of } \omega(x)\\ &= \omega(\beta(\eta_0)).\end{aligned}$$

Necessity. From Figs. 1 and 3 and Theorem 2, both the situations

$$\alpha(\eta_0) = \gamma(\eta_0) < -\frac{\beta(\eta_0)+2}{2}, \qquad \bigl(\beta(\eta_0) \leq 6\bigr),$$

and

$$\alpha(\eta_0) = \gamma(\eta_0) < -2\sqrt{\beta(\eta_0)-2}, \qquad \bigl(\beta(\eta_0) > 6\bigr),$$

correspond to nonmonotonicity of (17). These situations obtain precisely when $\rho \notin [0,1]$ for Region I (Fig. 7a) or $\rho \notin [0,1]$ for Region II (Fig. 7b), from the convexity of $\omega(x)$. This proves the necessity and completes the proof. □

By choosing η_0 as any point satisfying the conditions of Theorem 4, we obtain a point interior to the region of monotonicity. The adjustment described in (20) can then be used to constrain the quintic to be monotone. For (A,B) in Region II, a reasonable choice of η_0 is the average of the interval endpoints, η_1 and η_3, namely

$$\eta_0 = \left(-\frac{\sqrt{A}}{4}(7\sqrt{A}+3\sqrt{B}), \frac{\sqrt{B}}{4}(3\sqrt{A}+7\sqrt{B})\right). \tag{23}$$

Actually, this same point is interior to the monotonicity region for (A,B) in Region I as well since η_3 lies between η_1 and η_2 within Region I along the line given in (22).

We will summarize all of these results in algorithmic form. If $\nu = X_1 - X_0 < 0$, monotonically increasing is impossible, and if $\nu = 0$, the solution is $A = B = C = D = 0$. Assume $\nu > 0$ in the following, and that nominal values of A, B, C, D are given (obtained, e.g., by a C^4 quintic spline interpolant or finite difference approximations). The case where (A, B) lies outside of Regions I and II will be handled in the first and second steps of the algorithm.

Algorithm for constructing a monotone increasing Hermite quintic.

1. Set $A := \max\{0, A\}$, $B := \max\{0, B\}$. If $AB = 0$, use the simpler positivity criteria of [8] for cubics.
2. If $\tau_1(A, B) = 24 + 2\sqrt{AB} - 3(A+B) < 0$, then scale the derivative vector (X_0', X_1') until $\tau_1(A, B) > 0$.
3. Let $\eta = (C, D)$. If $\alpha(\eta)$, $\beta(\eta)$, and $\gamma(\eta)$ do not satisfy the conditions of Theorem 2, then find ρ such that $\alpha(\eta^*)$, $\beta(\eta^*)$, and $\gamma(\eta^*)$ do satisfy the conditions, where

$$\eta^* = \rho\eta + (1-\rho)\eta_0, \qquad \rho \in (0, 1),$$

with η_0 defined in (23).

Following Huynh [5], Step 1 can be preceded by: Set $X_0' := \text{median}\left\{0, X_0', \frac{7\nu}{2h}\right\}$, $X_1' := \text{median}\left\{0, X_1', \frac{7\nu}{2h}\right\}$, and then compute (A, B). Step 2 can be replaced by a simpler sufficient condition which would scale the derivative vector till the first derivatives lie within the square imbedded within Region II (e.g., the square defined by $[0, 6] \times [0, 6]$). Yet another alternative to Step 2 would be to scale the derivatives X_0', X_1' back independently until $\tau_1(A, B) > 0$ [2], [3]. In a similar fashion, we can use the conditions in (16) to simplify verification of monotonicity of the quintic in Step 3 of the algorithm.

Note that the above algorithm is for a *single* monotone increasing Hermite quintic polynomial piece, and does not address at all the iterative adjustments required to achieve a monotone Hermite quintic *spline*. From Theorem 4 and (21) it is clear, however, that such an adjustment is always possible: for sufficiently small $(A, B) > 0$ the admissible (C, D) line segments are arbitrarily long. An efficient and robust computer algorithm for monotone Hermite quintic spline interpolation based on the teardrop regions of [11] will be the topic of a future paper.

5. Conclusion. Elegant necessary and sufficient conditions that a quartic polynomial $f(z)$ be nonnegative for $z > 0$ have been derived. Simple and computationally cheap sufficient conditions were deduced from the more general conditions. Important applications are to monotone quintic spline interpolation and the efficient generation of random variates from arbitrary continuous distributions. Applying the theory to monotone quintic interpolation, sharp necessary and sufficient conditions for monotonicity were derived. Simpler sufficient conditions for monotonicity were condensed into an algorithm suitable for quintic spline interpolation.

REFERENCES

[1] L. Brickman and L. Steinberg, *On nonnegative polynomials*, Amer. Math. Monthly, 69 (1962), pp. 218–221.

[2] C. deBoor and B. Swartz, *Piecewise monotone interpolation*, J. Approx. Theory, 21 (1977), pp. 411–416.

[3] R. L. Dougherty, A. Edelman, and J. M. Hyman, *Nonnegativity-, monotonicity-, or convexity-preserving cubic and quintic Hermite interpolation*, Math. Comp., 52 (1989), pp. 471–494.

[4] F. N. Fritsch and R. E. Carlson, *Monotone piecewise cubic interpolation*, SIAM J. Numer. Anal., 17 (1980), pp. 238–246.

[5] H. T. Huynh, *Accurate monotone cubic interpolation*, National Aeronautical and Space Administration Tech. Mem. 103789, Lewis Research Center, Cleveland, OH, 1991.

[6] E. I. Jury and M. Mansour, *Positivity and nonnegativity of a quartic equation and related problems*, IEEE Trans. Automat. Control, 26 (1981), pp. 444–451.

[7] S. Neumark, *Solution of Cubic and Quartic Equations*, Pergamon, New York, 1965.

[8] J. W. Schmidt and W. Heß, *Positivity of cubic polynomials on intervals and positive spline interpolation*, BIT, 28 (1988), pp. 340–352.

[9] G. Szegö, *Orthogonal Polynomials*, Amer. Math. Soc. Colloq. Publ., Vol. 23, Providence, RI, 1939.

[10] G. Ulrich and L. T. Watson, *A method for computer generation of variates from arbitrary continuous distributions*, SIAM J. Sci. Statist. Comput., 8 (1987), pp. 185–197.

[11] ———, *Using piecewise polynomials to generate from arbitrary continuous distributions*, Computer Science and Statistics: Proceedings of the 19th Symposium on the Interface, Philadelphia, PA, 1987, pp. 450–453.

[12] S. Wolfram, *Mathematica: A System for Doing Mathematics by Computer*, Addison-Wesley, Redwood City, CA, 1991.