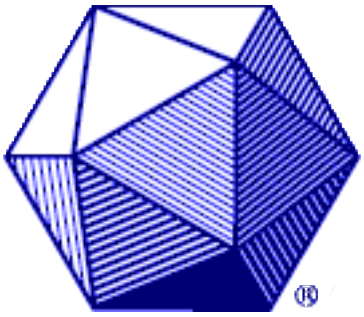

The Road to Chaos is Filled with Polynomial Curves

Richard D. Neidinger and R. John Annen, III

1. INTRODUCTION. The bifurcation diagram (Figure 1) is much like an aerial photograph—there are many interesting things revealed but it's hard to see the detail or to interpret the features. By superimposing a family of polynomial curves (Figure 2), this picture turns into a road map—marking trails, boundaries, milestones, and labeling points of interest. Both the curves and the diagram are generated by iterating the function $x^2 + c$ as the parameter ranges across a horizontal c-axis from 0 to -2. The initial iterates form the curves with the nth curve being the nth iterate of the critical value zero as a function of c. The bifurcation diagram (also known as the orbit diagram, the Feigenbaum diagram, or the road to chaos) is a plot of later iterates, showing the eventual state of the system.

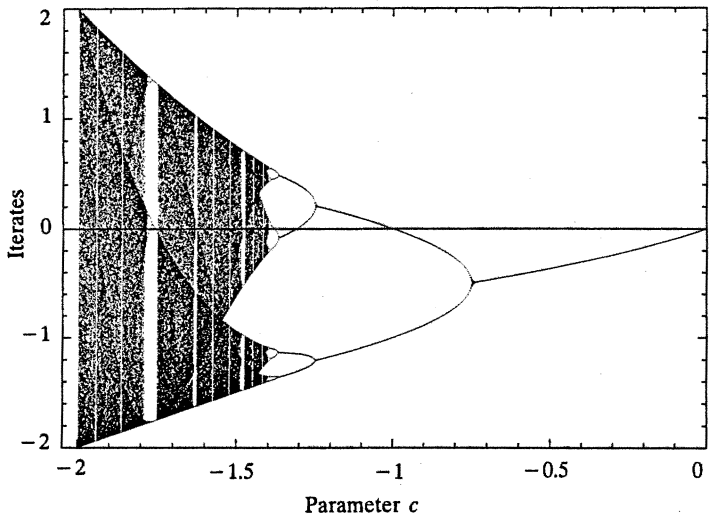

Figure 1. The bifurcation diagram of $f_c(x) = x^2 + c$.

In the first half of the paper, curve intersections provide insight into the dynamics of the system. The roots of the curves are shown to mark periodic windows in the bifurcation diagram, usually of invisible width, indicating a remarkable preponderance of attracting periodic phenomena that contrasts with the seeming chaos in the diagram. Other intersections mark parameter values where the iteration is chaotic. A brief detour into the complex plane yields a simple algorithm that color codes the Mandelbrot set to indicate the periods of the attracting phenomena. The second half of the paper studies how curve inequalities

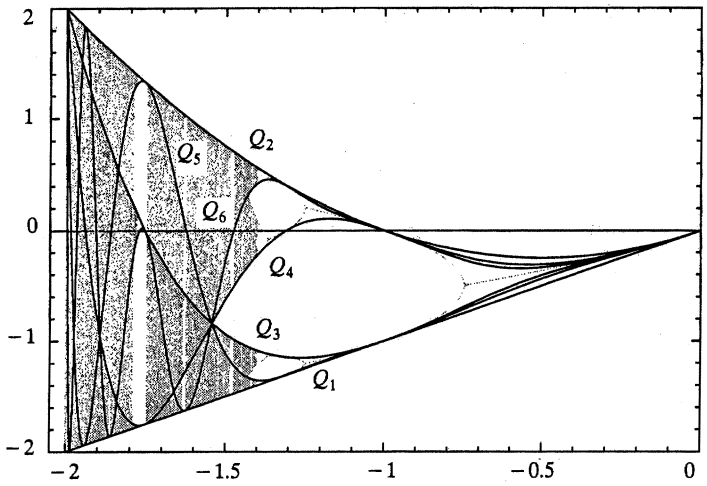

Figure 2. The first six Q-curves reveal dynamics and shapes within the diagram.

describe the shapes seen in the bifurcation diagram. We show why the high-numbered iterates condense along the polynomial curves. The curves also form envelopes that bound embedded, self-similar copies of the bifurcation diagram. Curve intersections specify the exact beginning and ending parameters for every envelope found in the diagram and, thus, mark where each window ends. Finally, Šarkovskiĭ's ordering can be seen in the order in which the curves first cross the axis. The ordering is argued without using Šarkovskiĭ's theorem.

Let's begin by defining terms and notation. For the standard dynamical system $f_c(x) = x^2 + c$, the iterates are defined by $x_{i+1} = f_c(x_i)$, or equivalently $x_i = f_c^i(x_0)$, for some fixed x_0. A sequence of such iterates is usually called an *orbit* and the *dynamics of the system* refers to the behavior of orbits. The *bifurcation diagram*, in Figure 1, displays the dynamics for different values along the horizontal c-axis by vertically plotting iterates, numbered 300 to 900, in the orbit of zero.

Define the nth polynomial $Q_n(c) = f_c^n(0)$ to be the nth iterate of f_c starting from zero. Specifically, $Q_1(c) = c$, $Q_2(c) = c^2 + c$, and $Q_3(c) = (c^2 + c)^2 + c$. In general, $Q_{n+1}(c) = (Q_n(c))^2 + c$. Thus, each $Q_n(c)$ is a 2^{n-1} degree polynomial. The graph of any Q_n as a function of c will be called a *Q-curve*. The first six Q-curves are displayed in Figure 2 and portions of these curves are visible as a relative density of points in Figure 1.

Q-curves are iterates of the critical value (in **[PJS]**, p. 633, they are called *critical-value lines*). The important role of the critical value is found in **Fatou's Theorem:** Every attracting cycle for a polynomial attracts at least one critical point **[Br]**. Even if the iteration for a bifurcation diagram starts from a different x_0, the eventual state reveals the same Q-curves. Indeed, the bifurcation diagram in the background of Figure 2 uses $x_0 = 0.2$, while the curves are iterates of the critical value zero. Typically, even bifurcation diagrams start with zero in order to avoid isolated values where x_0 may be a repelling periodic point.

2. CURVE INTERSECTIONS REVEAL SYSTEM DYNAMICS

2.1 Roots and windows. In the bifurcation diagram, the gray haze of iterates contains fascinating gaps, or vertical strips of white space, usually called *windows*.

The windows are formed whenever the iteration abruptly changes from a wide spread of values to periodic oscillation. Each window can be associated with the lowest period of oscillation that occurs within that parameter range. A few windows are visible in Figure 1 but the Q-curves mark many windows that are too narrow to be visible.

A quick observation from Figure 2 is that the curve Q_3 crosses the axis in the region usually called the window of period 3. Likewise, Q_5 and Q_6 cross in windows of period 5 and 6, respectively. The natural conjecture is true: whenever a Q-curve meets the axis, it marks a periodic window in the diagram (with period of the lowest numbered Q-curve that crosses there) and all windows are marked in this way. Actually, each root is a parameter value with a superattracting periodic cycle. Recall that p is a *periodic point of period n* if $f_c^n(p) = p$ but $f_c^j(p) \neq p$ for $0 < j < n$. This p is *repelling* if $|(f_c^n)'(p)| > 1$, *attracting* if $|(f_c^n)'(p)| < 1$, and *superattracting* if $|(f_c^n)'(p)| = 0$.

Superattracting Root Theorem. *Let $n \in \mathbb{N}$. The parameter r satisfies $Q_n(r) = 0$ and $Q_j(r) \neq 0$ for $0 < j < n$ if and only if iteration of $f_r(x)$ has a superattracting periodic point of period n.*

Proof: Fix r. By definition, $Q_n(r) = f_r^n(0)$. Thus, $Q_n(r) = 0$ and $Q_j(r) \neq 0$ for $0 < j < n$ if and only if zero is a periodic point of period n. In general, a periodic point is superattracting if and only if a critical point belongs to the cycle. Indeed, by the chain rule, $(f^n)'(x_0) = \prod_{i=0}^{n-1} f'(x_i)$ and the product is 0 if and only if $f'(x_i) = 0$ for some i **[Br]**. Since zero is the only critical point of f_r, the theorem follows. ■

Such parameter values do mark windows. Indeed, continuity implies that every parameter with superattracting period n is contained in an open interval of parameter values with attracting period n. Conversely, any maximal interval of attracting period n will contain a superattracting value. In fact, more is true: the derivative of the nth composition at a cycle point ranges from 1 to -1 across an interval of attracting period n. This statement follows from the more general complex result of Douady, Hubbard, and Sullivan ([**Br**], p. 85) and can be observed in the diagrams of Devaney [**D2**]. A *window of period n* in the bifurcation diagram is an interval of attracting period n that is augmented on the left by a region of bounded iterates that we identify as an envelope in Section 3.2.

The roots of the Q-curves indicate many windows that are not visible in the bifurcation diagram (Figure 1) or even in zooms on the diagram [**D2**]. In Figure 3, the graph of Q_{11} alone marks a tremendous number of windows of period 11. Curve Q_{11} has no roots (except 0) to the right of -1.5, but the zoom shows how roots, and therefore windows, are concentrated near -2.

2.2 Roots and Mandelbrot buds. The Superattracting Root Theorem can be used in the complex c-plane to mark components of period n. This is analogous to marking windows on the c-axis. The polynomials $Q_n(c) = f_c^n(0)$ make sense for complex c-values, even though the term "curve" is not appropriate in this context. At complex roots where these polynomials are zero, the system has a superattracting periodic point of period n. Around each root, the region of attracting period n is a hyperbolic component of the Mandelbrot set **[Br]**, and appears as a bud sprouting off a larger component or as a cardioid of a (mini-)Mandelbrot set. We can color neighborhoods of the roots, where $|f_c^n(0)| < \varepsilon$, in order to show the

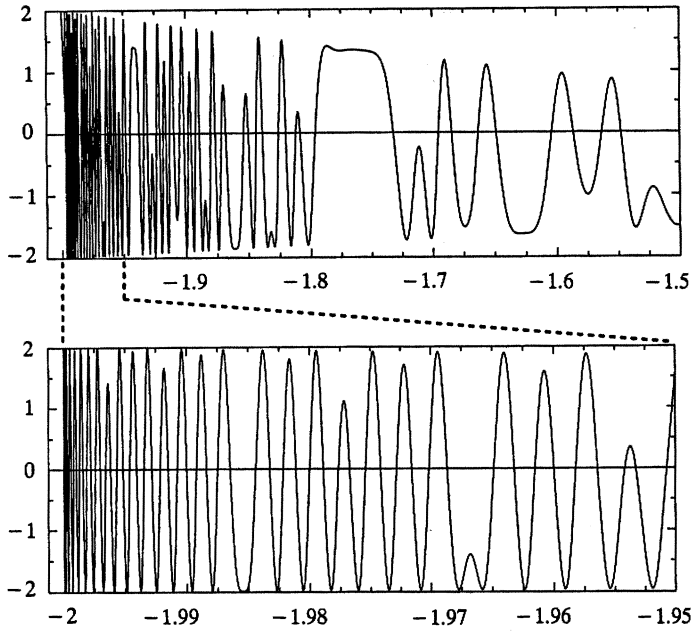

Figure 3. Every root of Q_{11} marks a period 11 window.

period of the corresponding components.

A program for plotting the Mandelbrot set is easily adapted to perform the color-coding of components. For each pixel c, the typical Mandelbrot set algorithm starts with $z_0 = 0$ and iterates f_c until $|z_n| > 2$, so that c is outside the Mandelbrot set, or n exceeds some maximum iteration, so that c is assumed to be in the Mandelbrot set. We simply add a third bail-out option: whenever $|z_n| < \varepsilon$, and $n \leq$ the number of colors to be used in period-labeling, stop and color the pixel with color n. Choose solid colors, for the original two stopping tests (as opposed to the usual practice of coloring according to n when $|z_n| > 2$). A fairly large ε, such as 0.2, is recommended to color "disks" that very roughly cover the buds that are being labelled. More importantly, it will label (the period of points in the main cardioid of) mini-Mandelbrot sets near the boundary that would not be visible and could even be smaller than a pixel.

This might be called a quick and dirty algorithm. It's quick because it's simple to program and actually provides an early exit for many pixels that would otherwise reach the maximum iteration. It's dirty because it's difficult to say anything rigorous about colored pixels for arbitrarily or aesthetically chosen ε's. It is usually easy to identify the component that goes with a colored disk, or vice versa, but colors can extend into adjacent components. Colored disks can be smaller or larger than the corresponding components, even for the same n and ε. Other algorithms, such as direct period detection from iteration or the spider algorithm, are more rigorous.

Two examples of the beautiful and intriguing color figures that are produced by this algorithm can be viewed through the World Wide Web at http://www.davidson.edu/academic/math/neidinger/road.html. Fascinating patterns appear in the sequence of periods found in the geometric patterns. Of course, zooms near the boundary are as entertaining as ever. A period-n bud, off

the main cardioid, is crowned by a decoration consisting of n spokes and each spoke contains a sequence of periods (covering cardioids of mini-Mandelbrot sets) from an equivalence class mod n.

2.3 The intersection dichotomy. We now leave the complex detour and return to the road map where every intersection is interesting. There are only two types of intersections between Q-curves and they imply different system dynamics for the corresponding parameter values. Figure 4 shows many intersections of the first eight Q-curves on $[-2, -1.8]$ and labels one of each type. A tangency occurs at a superattracting root r, and a non-tangent intersection occurs at a Misiurewicz point s, which has chaotic dynamics. While the implications of a root are relatively straight-forward, the implications of a Misiurewicz point rely on references in the literature.

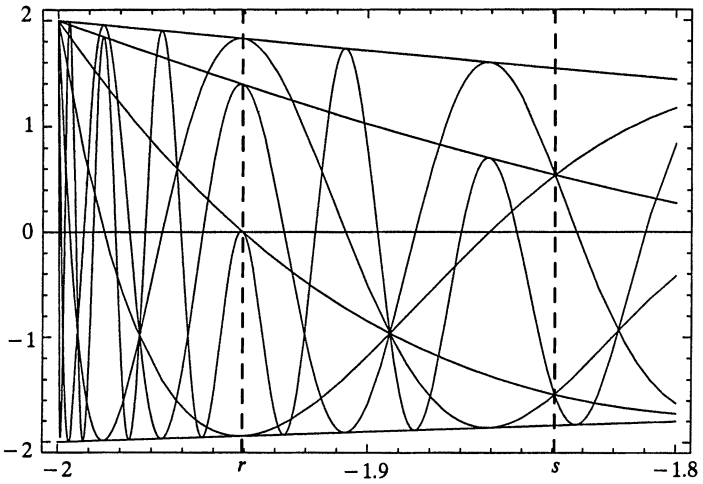

Figure 4. The two types of intersections: all curves tangent or none tangent, with periodic or chaotic dynamics, respectively.

A *Misiurewicz point* is a c-value where the orbit of the critical value zero is strictly pre-periodic, i.e., it eventually falls onto a periodic orbit not including zero. The proof that the corresponding dynamics is chaotic is very involved and uses the definition of chaos from ergodic theory (see references in the proof below). However, the following intuitive argument, from **[CE]**, p. 34, yields some insight into the chaotic behavior: At a Misiurewicz point, the eventual periodic orbit of zero is repelling and, by Fatou's Theorem, no periodic attractor can exist. We now argue sensitive dependence on initial conditions, meaning that nearby points are eventually "pulled apart" through the iteration. This condition is hardest to fulfill for points near $x = 0$, since the derivative of $f_c(x)$ at zero is zero, and, hence, f_c contracts points near zero. But eventually zero itself is mapped onto a repelling periodic orbit. Thus, the orbits of points near zero may initially move closer (to the orbit of zero) but will eventually come near the repelling periodic orbit and be pushed away.

Intersection Dichotomy. *At any intersection of two Q-curves, the c-value is either*:

(1) *a root of some Q_n (let n be the smallest such positive number), where every intersection of Q-curves is a tangency at one of n distinct points (in equivalence*

classes mod n), and where the system f_c has a superattracting periodic point of period n, or

(2) *a Misiurewicz point, where no intersection of Q-curves is a tangency, and where the system f_c is chaotic.*

Proof: Let c be the value at the intersection of two Q-curves. By definition of Q_i, the orbit of zero is $Q_i(c)$. If c is not the root of any Q-curve, then the orbit never returns to zero but the intersection implies that the orbit is eventually periodic, so c is a Misiurewicz point. Thus, c is either a root or a Misiurewicz point.

The claims about system dynamics are found, in case (1), in the previous Superattracting Root Theorem and, in case (2), in results of M. Misiurewicz found in [**CE**] (p. 155 and supporting text). There, it is shown that if c is a Misiurewicz point, then f_c has an ergodic, invariant, probability measure that is absolutely continuous with respect to Lebesgue measure. This measure-theoretic condition is one way to define a *chaotic* dynamical system. The proof in [**CE**] applies to a class of functions, including $g = f_c/c$ for a Misiurewicz point $c \leq -0.5$. Ergodic theory techniques are used to construct an *invariant measure* ν (meaning $\nu(E) = \nu(g^{-1}(E))$ for every measurable set E) which is absolutely continuous. Such a measure is shown to be unique and, hence, ergodic. Readers unfamiliar with this characterization of chaos might want to consider how an attracting periodic orbit would prohibit the existence of an absolutely continuous invariant measure.

To prove the claims about tangency, observe that: an intersection and its tangency or non-tangency can be propagated to the next iterate (and hence all subsequent iterates) by the recurrence relations $Q_{i+1}(c) = (Q_i(c))^2 + c$ and $Q'_{i+1}(c) = 2Q_i(c)Q'_i(c) + 1$.

First, consider a root $c = r$ where $Q_n(r) = 0$ and no smaller numbered curve is zero at r. Then, $Q_{n+1}(r) = 0^2 + r = Q_1(r)$ and $Q'_{n+1}(r) = 2Q_n(r)Q'_n(r) + 1 = 1 = Q'_1(r)$. By the recurrence relations, $Q_{n+i}(r) = Q_i(r)$ and $Q'_{n+i}(r) = Q'_i(r)$ for all $i \geq 1$. So all these intersections are tangencies at one of the n points $Q_1(r), \ldots, Q_n(r)$. There are no other intersections of Q-curves at r since these n points are distinct (otherwise, forward propagation would yield $Q_j(r) = Q_n(r) = 0$ for some $j < n$ which contradicts the assumption on n).

Now, assume that $c = s$ is a Misiurewicz point. Then, since zero is not itself periodic, s is not the root of any Q-curve. Let k be the smallest number such that Q_k intersects another Q-curve at s, and let n be the smallest number such that $Q_k(s) = Q_{k+n}(s)$. This is not a tangency, by the following argument. By the recurrence relation, $(Q_{k-1}(s))^2 = (Q_{k+n-1}(s))^2$ and, hence, $Q_{k-1}(s) = -Q_{k+n-1}(s)$. In particular, $k \geq 2$ (otherwise, the argument of the previous sentence would yield $0 = Q_{k+n-1}(s)$, which contradicts no root at s). Now, define $G(c) = Q_{k+n-1}(c) + Q_{k-1}(c)$. In the next paragraph, we will show that $G'(s) \neq 0$. Then

$$\begin{aligned} Q'_{k+n}(s) - Q'_k(s) &= 2Q_{k+n-1}(s)Q'_{k+n-1}(s) - 2Q_{k-1}(s)Q'_{k-1}(s) \\ &= -2Q_{k-1}(s)G'(s) \neq 0. \end{aligned}$$

We conclude that $Q'_{k+n}(s) \neq Q'_k(s)$, so the intersection is not a tangency. By the recurrence relations, $Q_{n+i}(s) = Q_i(s)$ and $Q'_{n+i}(s) \neq Q'_i(s)$ for all $i \geq k$. (We can propagate the non-tangency since s is not the root of any Q-curve.) Thus, all these intersections are non-tangencies at one of the n distinct points $Q_k(s), \ldots, Q_{k+n-1}(s)$. There are no other intersections of Q-curves at s by the assumptions of "smallest" for k and n.

The following algebraic argument, from p. 333 of [**DH**], shows that $G'(s) \neq 0$: By the recurrence relation, $G'(s)/2 = Q_{k+n-2}(s)Q'_{k+n-2}(s) + Q_{k-2}(s)Q'_{k-2}(s) + 1$. To show that this expression cannot be zero, the rough idea is to show that the combination of Q and Q' values has a positive power of 2 as a type of factor and, thus, cannot be -1. Specifically, each rational number has a *2-adic valuation* given by the power of 2 in the natural factorization of the fraction. The field of rationals and the valuation can be extended to include the number s. Let A be the set of elements of the extension field with non-negative valuation and let m be the set of elements with positive valuation. Then, m is a maximal ideal in the ring A. Using obvious valuations, such as $1 \notin m$ and $2A \subset m$, we work with arithmetic mod m. Since s is the root of a monic polynomial with integer coefficients (i.e., an algebraic integer), s and all integer coefficient polynomials of s (including all Q and Q' values) are in A. In fact, every $Q'_i(s) \equiv 1 \bmod m$, since $Q'_i(s) = 2Q_{i-1}(s)Q'_{i-1}(s) + 1$. Thus,

$$Q_{k+n-2}(s)Q'_{k+n-2}(s) + Q_{k-2}(s)Q'_{k-2}(s) \equiv Q_{k+n-2}(s) + Q_{k-2}(s) \bmod m.$$

A property of s, from the previous paragraph, is that $Q_{k+n-1}(s) + Q_{k-1}(s) = 0$. By the recurrence, $Q^2_{k+n-2}(s) + Q^2_{k-2}(s) = -2s \in m$. It follows that $(Q_{k+n-2}(s) + Q_{k-2}(s))^2$ and, hence, $Q_{k+n-2}(s) + Q_{k-2}(s)$ are in m. By the preceding modular equivalence, $Q_{k+n-2}(s)Q'_{k+n-2}(s) + Q_{k-2}(s)Q'_{k-2}(s)$ is in m and cannot be -1 (which has valuation 0). Thus, $G'(s) \neq 0$. ■

Now, one can literally see periodic and chaotic phenomena mixed throughout parameter ranges where the bifurcation diagram just shows a gray haze. Look at the curves to the left of the point marked r in Figure 4; there is superattracting periodic dynamics at every intersection with the axis, while there is chaotic dynamics at every non-tangent intersection of curves. Natural questions arise about the density and/or measure of the attracting periodic and chaotic parameters. The known answers are deep results in the field. Recently, Świątek showed that the union of the attracting periodic parameters is an open, dense set in the interval [**S**] and, yet, Jakobson's theorem shows that the set of chaotic parameters has positive Lebesgue measure [**R**]!

3. CURVE INEQUALITIES SHAPE THE DIAGRAM

3.1 Why curves appear in the bifurcation diagram. In the bifurcation diagram, Figure 1, the first few Q-curves appear as pathways that are well-traveled by later iterates. Once the iterates spread across a vertical range (for $c <$ the Feigenbaum point $-1.401155\ldots$, [**C**], p. 141), they seem to cluster around curves, particularly on one side of each curve. This density of iterates is especially clear in the histograms of [**D4**], p. 127, and [**PJS**], p. 632. Why do the Q-curves, formed by initial iterates of zero, emerge from these later iterates, even when $x_0 \neq 0$?

One way to explain this phenomenon is to show how iterates in a relatively wide band around the axis must map into narrow bands around the first few Q-curves. These bands are shown in Figure 5, between solid Q-curves and corresponding dashed curves, for c in $[-2, -1]$. The dashed curves are given by $P_n(c) = f_c^n(\varepsilon)$ (for $\varepsilon = 0.15$ in Figure 5), corresponding to $Q_n(c) = f_c^n(0)$. If any iterate x_i falls in the band of width 2ε around the c-axis, then the next iterate will be in the band of width ε^2 between Q_1 and P_1, which condenses iterates above Q_1. Specifically, if $0 \leq |x_i| \leq \varepsilon$, then, by squaring and adding c to each side, $Q_1(c) \leq x_{i+1} \leq P_1(c)$. Assuming $P_1(c) \leq 0$ and again applying f_c, yields $Q_2(c) \geq x_{i+2} \geq P_2(c)$. Since

$P_2(c) \geq 0$ for all $c < -1.4$, we can "propagate the bounds" to the next iterate, $Q_3(c) \geq x_{i+3} \geq P_3(c)$; so that the density of points will occur below Q_3. In the range of c-values where $Q_3(c) \leq 0$, the next iterate x_{i+4} will fall above Q_4 but, in the c-range where $P_3(c) \geq 0$, it falls below Q_4. The bands should appear condensed when the width $|Q_i(c) - P_i(c)|$ is smaller than the initial 2ε, which happens for low i and small ε, though the width can grow as c approaches -2. We can estimate $|Q_{i+1} - P_{i+1}| = |Q_i^2 - P_i^2| \leq (|Q_i| + |P_i|)|Q_i - P_i| \leq 4|Q_i - P_i|$. By induction, $|Q_{i+1} - P_{i+1}| \leq 4^i\varepsilon^2$. This phenomenon emphasizes one side of low-order Q-curves (lower should be darker); it can be observed in the bifurcation diagram in Figure 1.

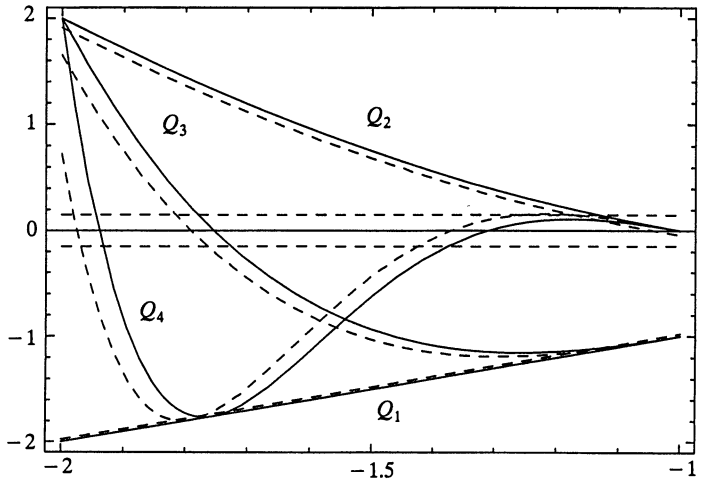

Figure 5. Iterates near the axis are mapped into concentrated bands around curves.

A geometrical view of the dynamical system can explain why the whole region below Q_3 is darker than the region above it, as c ranges between about -1.8 to -1.4. The bifurcation diagram on $[-2, -1]$ can be divided into two regions, in the (c, x)-plane, depending on whether or not $|x| \leq Q_2(c)$. If (c, x) satisfies $0 \leq |x| \leq Q_2(c)$, applying f_c yields $Q_1(c) \leq f_c(x) \leq Q_3(c)$; i.e., it maps the symmetric region by folding it over along the axis, stretching the axis down to Q_1, and stretching Q_2 down to Q_3. The remaining portion of the bifurcation diagram satisfies $Q_1(c) \leq x \leq -Q_2(c)$, so that applying f_c yields $Q_2(c) \geq f_c(x) \geq Q_3(c)$; flipping it over and filling in above Q_3. In the range between about -1.8 to -1.4, this motion condenses the first region and expands the second.

3.2 Envelopes bound copies of the diagram. On the road map given by curves on top of the bifurcation diagram (Figure 2), Q-curves form boundary lines separating regions that contain iterates from those that don't. In fact, these curves bound the shape and identify the location of all embedded, self-similar copies of the diagram. In general, whenever a curve Q_n crosses the axis, the curves Q_n and Q_{2n} bound a self-similar copy of the bifurcation diagram consisting of every nth iterate. The other iterates fall in similar envelopes, also bounded by Q-curves.

Figure 6 shows three such sets of envelopes. First, the entire diagram is bounded by Q_1 below and Q_2 above. Second, all the even iterates are bounded between Q_2 and Q_4, from the root of Q_2 to the point where Q_4 meets $-Q_2$. In

this parameter range, the odd iterates fall between Q_1 and Q_3. Finally, a quarter of the iterates are bounded between Q_4 and Q_8, from the root of Q_4 to the point where Q_8 meets $-Q_4$. Four envelopes can be seen in this parameter range.

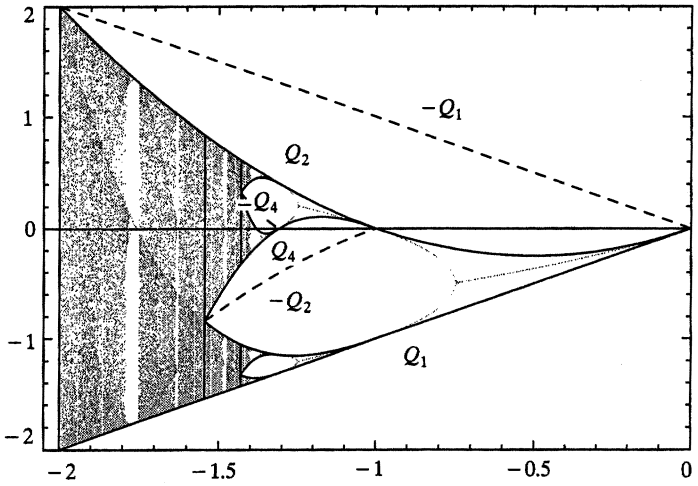

Figure 6. Q-curves bound nested copies of the diagram.

Figure 7 zooms in on the period three window, around the axis, to show an envelope formed when Q_3 crosses the axis. The envelope abruptly ends at the Misiurewicz point where $-Q_3$ meets Q_6. Within this range, every third iterate falls between Q_3 and Q_6, while the other iterates are constrained to the region between Q_1 and Q_4 or the region between Q_2 and Q_5. This containment is being observed when people speak of a "period 3 window." This is different from the interval of attracting period 3. Actually, if the envelope is on the interval (s, r) and the interval of attracting period 3 is (a, b), then the window is the union (s, b). (The right endpoint, b, of the attracting interval is a point not described by Q-curve intersections.) Each of the tiny, nested windows seen in Figure 7 contains even smaller envelopes of Q-curves.

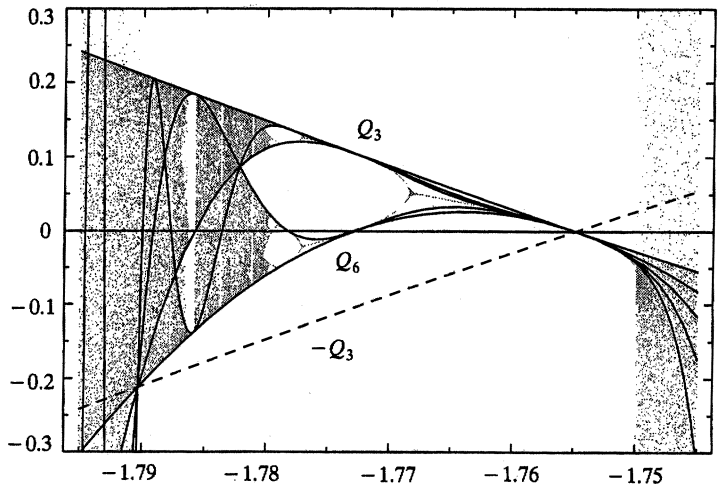

Figure 7. A zoom into the period 3 window shows an envelope of Q-curves.

The following theorem describes all such envelopes. The phrase *x lies between y and z* means $y \le x \le z$ or $z \le x \le y$.

Envelope Theorem. *If* $Q_n(r) = 0$ *and* $Q_j(r) \neq 0$ *for* $0 < j < n$, *let* $s = \max\{c < r: Q_{2n}(c) = -Q_n(c)\}$. *If* $c \in [s, r]$ *and* $|x_0| \le |Q_n(c)|$, *then* x_{kn+m} *lies between* Q_m *and* Q_{m+n} *for all* $m \in \{1, 2, \ldots, n\}$ *and* $k \in \{0, 1, 2, \ldots\}$. *In particular,* Q_{kn+m} *lies between* Q_m *and* Q_{m+n} *on* $[s, r]$. *The regions between these pairs of curves,* Q_m *and* Q_{m+n} *for each* $m \in \{1, 2, \ldots, n\}$, *are disjoint from each other on* $(s, r]$. *For later application, we also note that, if* $m \neq n$, *the region between* Q_m *and* Q_{m+n} *never meets the axis on* $[s, r]$.

To prove the Envelope Theorem, we "propagate the bounds" from one iterate to the next. However, this requires either that the iterate must fall between $-Q_n$ and $+Q_n$ or that both the upper and lower bounds must have the same sign. This will be reduced to the requirement that $|Q_{2n}| < |Q_n|$, a fact that is fairly obvious in the graphs. This and other essential properties of the "horn shape" formed by the graphs of Q_{2n} and Q_n (see Figures 6 and 7) are summarized in the following lemma. The claims following "moreover" are used in the next section.

Horn Lemma. *If* $Q_n(r) = 0$ *and* $Q_j(r) \neq 0$ *for* $0 < j < n$, *then there exists* $s < r$ *such that* $Q_{2n}(s) = -Q_n(s)$ *and* $|Q_{2n}| < |Q_n|$ *on* (s, r). *Moreover, there exists* $r' \in (s, r)$ *such that* $Q_{2n}(r') = 0$, $Q_{2n}Q_n < 0$ *on* $[s, r')$ *and* $Q_{2n}Q_n > 0$ *on* (r', r).

This lemma is deceptively innocent-looking—it embodies deep properties of the quadratic map that we will not prove. The "horn shape" is a real-parameter consequence of Douady and Hubbard's mini-Mandelbrot sets for the complex parameter [**DH**]; they prove that only complete (not partial) copies of the set occur. The horn shape is also equivalent to a pattern observed in graphical analysis (used to produce web-diagrams), as described in the following paragraph.

Consider the graph of $y = f_c^n(x)$ near the origin in the xy-plane. If $Q_n(r) = 0$, zero is a superattracting periodic point of period n for f_r, so that $y = f_r^n(x)$ has an extreme value at the origin. In fact, for a range of c-values near r, the graph resembles a parabola in a small region around the origin. We examine the concave-upward case shown in Figure 8. As c decreases from r to a value that we call s, the vertex moves monotonically downward until it becomes as large, in absolute value, as the coordinates of the intersection with the diagonal in the first quadrant. This diagonal intersection is a repelling point that forms the corner of a square of trapped iterates, as in [**D1**], p. 132. The properties of the Horn Lemma can now be deduced from this family of graphs. On each graph of $y = f_c^n(x)$, the y-intercept is $Q_n(c) = f_c^n(0)$. Now, $Q_{2n}(c) = f_c^n(Q_n(c))$ is visualized by taking the y-intercept horizontally to the diagonal and then vertically to the curve. In Figure 8, the resulting values of $Q_{2n}(c)$ are connected, creating a curve parameterized by $x = Q_n(c)$ and $y = Q_{2n}(c)$ for c in $[s, r]$. This parametric curve displays all of the properties stated in the Horn Lemma. This is not a proof, but merely shows that the lemma is equivalent to the claim that the parabola-shape always moves monotonically until reaching the edge of the trapping box. Graphical analysis also shows the dynamical significance of the parameter s—beyond s, the parabola pokes out of the box and the containment (or envelope) of iterates ends. We now return to the c-axis for the proof of the containment between Q-curves.

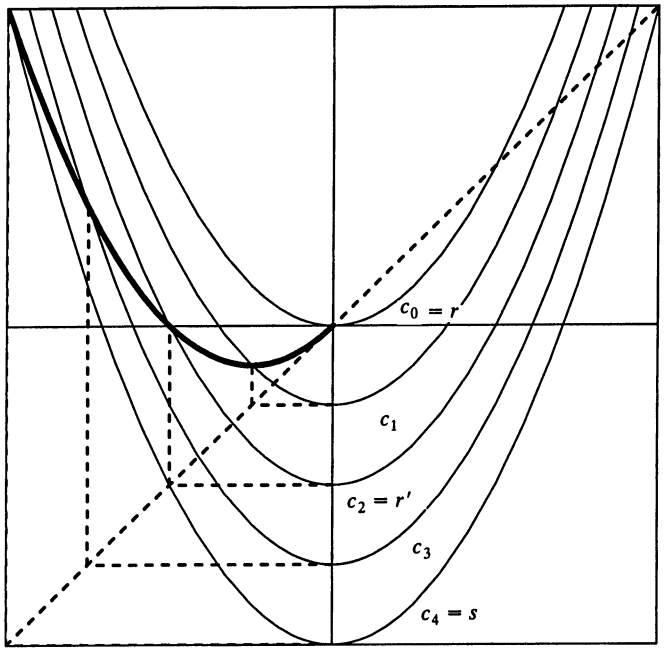

Figure 8. Five graphs of $y = f_c^n(x)$ lead to the Horn Lemma as shown in the thick curve parameterized by $x = Q_n(c)$ and $y = Q_{2n}(c)$.

Proof of the Envelope Theorem: (Assuming the Horn Lemma.) Fix r and n such that $Q_n(r) = 0$ and $Q_j(r) \neq 0$ for $0 < j < n$. By the Horn Lemma, let $s < r$ such that $Q_{2n}(s) = -Q_n(s)$ and $|Q_{2n}(c)| < |Q_n(c)|$ for $c \in (s, r)$. Clearly, $s = \max\{c < r\colon Q_{2n}(c) = -Q_n(c)\}$.

Fix any $c \in [s, r]$ and consider the orbit from any $|x_0| \leq |Q_n(c)|$. Apply f_c to $0 \leq |x_0| \leq |Q_n(c)|$ to get $Q_1(c) \leq x_1 \leq Q_{n+1}(c)$. The containment asserted by the theorem continues because of two propagation principles. (1) If x lies between Q_m and Q_{m+n} for some $m \in \{1, 2, \ldots, n-1\}$, then $f_c(x)$ lies between Q_{m+1} and Q_{m+1+n}. (2) If x lies between Q_n and Q_{2n}, then $f_c(x)$ lies between Q_1 and Q_{1+n}. The second principle uses the fact that $|Q_{2n}(c)| \leq |Q_n(c)|$ for $c \in [s, r]$. If x lies between Q_n and Q_{2n}, then $0 \leq |x| \leq |Q_n(c)|$ and, again by applying f_c, $Q_1(c) \leq f_c(x) \leq Q_{n+1}(c)$. To prove the first principle, suppose x lies between Q_m and Q_{m+n}. One possibility (out of two) is that $Q_m(c) \leq x \leq Q_{m+n}(c)$. In the following paragraph, we show that $Q_m(c)$ and $Q_{m+n}(c)$ agree in sign. If they are both positive, then $Q_{m+1}(c) \leq f_c(x) \leq Q_{m+1+n}(c)$. If they are both negative, then $Q_{m+1}(c) \geq f_c(x) \geq Q_{m+1+n}(c)$. In either case, $f_c(x)$ lies between Q_{m+1} and Q_{m+1+n}. The other possibility is that the original containment was $Q_{m+n}(c) \leq x \leq Q_m(c)$, and the argument is identical. To complete the proof of containment, it remains to show that Q_m and Q_{m+n} are nonzero and agree in sign on $[s, r]$. We will show a bit more as a bonus.

We claim that the curves $Q_1, Q_2, \ldots, Q_{2n-1}$ never meet the axis in (s, r) and that none of the curves $Q_1, Q_2, \ldots, Q_{2n}$ intersect in (s, r). At the root r, the curves pair up, mod n, at the n distinct points (by the Intersection Dichotomy). Thus, the claim implies that the regions between these pairs of curves, Q_m and Q_{m+n} for each $m \in \{1, 2, \ldots, n\}$, are disjoint from each other on $(s, r]$. The claim also implies that, for each $m \in \{1, 2, \ldots, n-1\}$, the pair Q_m and Q_{m+n} are nonzero

and agree in sign throughout $(s, r]$. Since s is a Misiurewicz point, it can't be a root of Q_m or Q_{m+n}. So, in fact, Q_m and Q_{m+n} are nonzero and agree in sign throughout $[s, r]$. We prove the claim by contradiction. Let B be the set of counterexamples, i.e., $B = \{c \in (s, r)$: $Q_i(c) = Q_j(c)$ for some $i < j \leq 2n$, or $0 = Q_j(c)$ for some $j < 2n\}$, and suppose B is not empty. B is finite, since every Q_i is a distinct polynomial, so we can let $b = \max(B)$. Since $Q_i(b) = Q_j(b)$ or $0 = Q_j(b)$, propagate (apply f_b) until $Q_k(b) = Q_{2n}(b)$ for some $k < 2n$. Now, consider the continuous curves $|Q_k(c)|$ and $|Q_n(c)|$ on (b, r). At b, $|Q_k(b)| = |Q_{2n}(b)| < |Q_n(b)|$. By this inequality, $k \neq n$. At r, $|Q_k(r)| > 0 = |Q_n(r)|$. By the Intermediate Value Theorem, there exists some d in (b, r), such that $|Q_k(d)| = |Q_n(d)|$. Propagate this equality once to get $Q_{k+1}(d) = Q_{n+1}(d)$. Thus $d \in B$, which contradicts the assumed maximality of b. ∎

The Envelope Theorem's hypothesis that $|x_0| \leq |Q_n(c)|$ is necessary to have containment for all c in $[s, r]$ but, in practice, is not needed for most c-values. The bifurcation diagram, in the background of Figure 7, was created with $x_0 = 0$ to satisfy this hypothesis, which essentially requires that the iteration start inside an envelope. For other x_0 values, the diagram looks almost the same but there can be isolated exceptions where iterates leak out of the envelopes. Outside the envelopes, an x_0 value can be, for instance, in a repelling periodic orbit for various values of c (with different periods) in the range of the envelope. At such c values, the iteration never enters the envelopes; nearby, it enters only after many iterations. Still, for most c-values, the iteration is rapidly attracted down into the envelopes. In fact, for a given c in the range of an envelope, the set of those values of x_0 with bounded orbits that are not attracted into the envelopes is a set of Lebesque measure zero.

3.3 Šarkovskiĭ's ordering. The Q_n curves also reveal the pattern known as Šarkovskiĭ's ordering. This ordering of $\mathbb{N}$ is as follows:

$$3 \triangleright 5 \triangleright 7 \triangleright \cdots \triangleright 2 \cdot 3 \triangleright 2 \cdot 5 \triangleright 2 \cdot 7 \triangleright \cdots \triangleright 2^2 \cdot 3 \triangleright 2^2 \cdot 5 \triangleright \cdots \triangleright 2^3 \cdot 3 \triangleright 2^3 \cdot 5 \triangleright \cdots$$
$$\triangleright \cdots \triangleright 2^4 \triangleright 2^3 \triangleright 2^2 \triangleright 2 \triangleright 1.$$

Šarkovskiĭ's theorem [**D2**] applies to any continuous F: $\mathbb{R} \to \mathbb{R}$ and states that if F has a periodic point of period n and if $n \triangleright m$, then F also has a periodic point of period m. Although this theorem applies to f_c for a fixed parameter c, it's not immediately clear how this relates to Q_n across a range of c-values. We will not use Šarkovskiĭ's theorem but will establish a relationship by directly studying the Q-curves.

Root Ordering Theorem. *Define* $r_n = \max\{r$: $Q_n(r) = 0$ *and* $Q_j(r) \neq 0$ *for* $0 < j < n\}$. *Then* $r_n < r_m$ *if and only if* $n \triangleright m$.

Proof: Every r_n exists since, for $n > 2$, $Q_n(-2) = 2$ and $Q_n < Q_2 < 0$ on (r_2, r_1). By bifurcation theory (or the Horn Lemma and the Envelope Theorem), $\cdots < r_8 < r_4 < r_2 < r_1$ and if i is not a power of 2, then r_i is to the left of this list. For each power of two, $p = 2^j$, there is a corresponding $s_p < r_p$ described by the Horn Lemma. Also, $s_p < s_{2p} < r_{2p} < r_p$. These inequalities follow from the Horn Lemma; s_p can't be in (s_{2p}, r_{2p}) since on this interval $|Q_{4p}| < |Q_{2p}|$ and yet, at s_p, it can be argued that $Q_{4p}(s_p) = Q_{2p}(s_p)$. We conclude that $s_1 < s_2 < s_4 < s_8 < \cdots < r_8 < r_4 < r_2 < r_1$. To complete the proof, we now prove that, for each power of two $p = 2^j$, $s_p < r_{3p} < r_{5p} < r_{7p} < \cdots < s_{2p}$. The reader can use Figure 9 to follow along for $p = 1$.

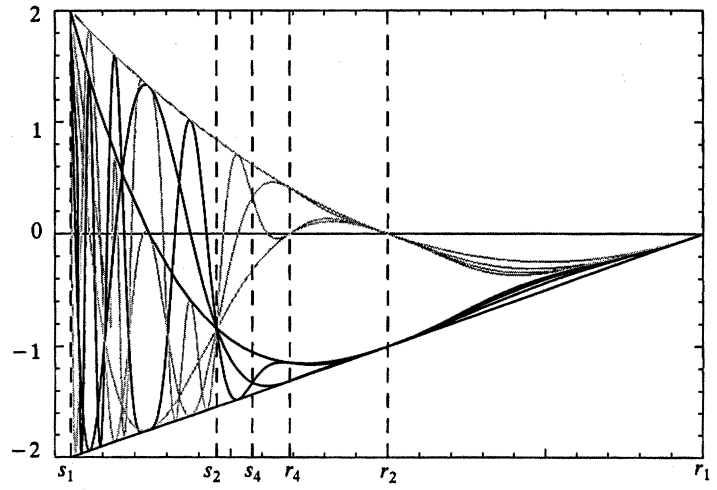

Figure 9. Šarkovskiĭ's ordering appears as odd curves (shown in black) cross the axis between s_1 and s_2.

Let $k \in \{3, 5, 7, \dots\}$. First, we have already argued that bifurcation theory (or the Horn Lemma and the Envelope Theorem) ensures that $r_{kp} < r_{2p}$. On $[s_{2p}, r_{2p}]$, the Envelope Theorem for $n = 2p$ says that Q_{kp} lies between Q_p and Q_{3p} and the region never meets the axis. Thus, $r_{kp} < s_{2p}$. We now establish that the roots do occur in the interval (s_p, s_{2p}). We will repeatedly use the fact that, at s_{2p}, the sign of each Q_{kp} agrees with the sign of Q_p, a fact implied by the observed containment of Q_{kp} curves.

By the Horn Lemma for $n = p$, Q_p and Q_{2p} have opposite signs throughout $[s_p, r_{2p})$, in particular throughout $[s_p, s_{2p}]$. At s_p, $-Q_p(s_p) = Q_{2p}(s_p)$ and iteration yields $Q_{2p}(s_p) = Q_{3p}(s_p)$. Thus, the sign of Q_{3p} changes from the sign of Q_{2p} at s_p to the sign of Q_p at s_{2p}. By the Intermediate Value Theorem, there exists a largest root r_{3p} of Q_{3p} in (s_p, s_{2p}). Now, at r_{3p}, $Q_{5p}(r_{3p}) = Q_{2p}(r_{3p})$. The sign of Q_{5p} changes from the sign of Q_{2p} at r_{3p} to the sign of Q_p at s_{2p}. By the Intermediate Value Theorem, there exists a largest root r_{5p} of Q_{5p} in (r_{3p}, s_{2p}). Again, at r_{5p}, $Q_{7p}(r_{5p}) = Q_{2p}(r_{5p})$, so that the sign of Q_{7p} changes on (r_{5p}, s_{2p}). By induction, $s_p < r_{3p} < r_{5p} < r_{7p} < \cdots < s_{2p}$. ■

4. CONCLUSION. Many of the most intriguing features along the road to chaos are clarified by looking at the road map showing Q-curves. For years, people have studied the wonders of the bifurcation diagram: there appears to be a region of chaos that is interrupted by windows of periodic attraction and that has shadowy curves from varying density of iterates; the entire diagram appears in smaller, similar copies nested within the diagram. Each of these phenomena is better understood if one also studies the Q-curves. The roots of Q-curves show that periodic windows are much more prevalent than the diagram suggests. On the other hand, the non-tangent intersections of Q-curves point out at least some points where one can confidently say that the system is chaotic. The Q-curves are the curves that appear in the diagram, both as the shadowy curves of higher density and as the boundaries of the diagram and the nested copies. Indeed, any root of a Q-curve begins, and a corresponding intersection ends, a nested copy of the diagram. Even Šarkovskiĭ's ordering and a color-coded (according to period) Mandelbrot set arise from this study of Q-curves.

Most of our proofs use only elementary mathematics and most of our arguments remain on the parameter axis, rather than switch to web diagrams on the graph of the quadratic. The key ideas are very accessible—periodic attraction at roots uses the chain rule, tangencies and envelopes use elementary algebra on equalities and inequalities, and Šarkovskiĭ's ordering uses the Intermediate Value Theorem. However, the properties of Misiurewicz points and the Horn Lemma do rely on more advanced work in ergodic theory, algebra, and complex analysis. There is ample graphical evidence for the Horn Lemma, which plays a crucial role in the Envelope Theorem and Šarkovskiĭ's ordering.

ACKNOWLEDGMENTS. We thank R. L. Devaney and many other colleagues and students for helpful conversations while developing this paper.

REFERENCES

[Br] B. Branner, The Mandelbrot Set, in *Chaos and Fractals*, Devaney and Keen, eds. Amer. Math. Soc., Providence, RI, 1989, pp. 75–105.

[C] R. M. Crownover, *Introduction to Fractals and Chaos*, Jones & Bartlett, Boston, 1995.

[CE] P. Collet and J-P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems*, Birkhäuser, Boston, 1980.

[D1] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Second Ed., Addison-Wesley, Redwood City, CA, 1989.

[D2] R. L. Devaney, Dynamics of Simple Maps, in *Chaos and Fractals*, Devaney and Keen, eds. Amer. Math. Soc., Providence, RI, 1989, pp. 1–24.

[D3] R. L. Devaney, The Orbit Diagram and the Mandelbrot Set, *College Math. J.*, 22 (1991), 23–38.

[D4] R. L. Devaney, *A First Course in Chaotic Dynamical Systems*, Addison-Wesley, Reading, MA, 1992.

[DH] A. Douady and J. H. Hubbard, On the Dynamics of Polynomial-like Mappings, *Ann. Scient. Ec. Norm. Sup.*, 4 serie, 18 (1985) 287–343.

[PJS] H.-O. Peitgen, H. Jürgens, and D. Saupe, *Chaos and Fractals*, Springer-Verlag, New York, 1992.

[R] M. R. Rychlik, Another Proof of Jakobson's Theorem and Related Results, *Ergodic Theory and Dynamical Systems*, 8 (1988) 93–109.

[S] G. Świątek, Hyperbolicity is Dense in the Real Quadratic Family, preprint.

Department of Mathematics
Davidson College
P.O. Box 1719
Davidson, NC 28036
rineidinger@davidson.edu

Forsyth County Management Info. Services
Room 604—Hall of Justice
200 N. Main Street
Winston-Salem, NC 27101
annen@zeus.co.forsyth.nc.us