Remarks on Iterated Cubic Maps

John Milnor

CONTENTS

1. The Parameter Space for Cubic Maps
2. Real Cubic Maps as Real Dynamical Systems
3. Complex Cubics: The Connectedness Locus
4. Hyperbolic Components
Appendix A. Normal Forms and Curves in Parameter Space
Appendix B. Centers of Some Hyperbolic Components
Appendix C. Comments on the Figures
References

This article discusses the dynamics of iterated cubic maps from the real or complex line to itself and will describe the geography of the parameter space for such maps. It is a rough survey with few precise statements or proofs and depends strongly on work by Douady, Hubbard, Branner and Rees.

1. THE PARAMETER SPACE FOR CUBIC MAPS

Following [Branner and Hubbard 1988], any cubic polynomial map from the complex numbers $\mathbf{C}$ to $\mathbf{C}$ is conjugate, under a complex affine change of variable, to a map of the form

$$z \mapsto f(z) = z^3 - 3a^2 z + b\,, \tag{1.1}$$

with critical points $\pm a$ (cf. Appendix A). This normal form is unique up to the involution that carries (1.1) to the map $z \mapsto -f(-z) = z^3 - 3a^2 z - b$, changing the sign of b. Thus the two numbers

$$A = a^2, \qquad B = b^2 \tag{1.2}$$

form a complete set of coordinates for the *moduli space* of complex cubic maps up to affine conjugation. The invariant A can be thought of as a kind of discriminant, which vanishes if and only if the two critical points coincide, whereas B is a measure of asymmetry, which vanishes if and only if f is an odd function.

Now consider a cubic map $x \mapsto g(x)$ with real coefficients. If we reduce to normal form by a complex change of coordinates, as above, we obtain a complete set of invariants (A, B) that turn out to be real. However, if we allow only a real change of coordinates, there is one additional invariant, namely the sign

$$\sigma = \operatorname{sgn} g''' \tag{1.3}$$

of the leading coefficient. It is not difficult to check that σ coincides with $\operatorname{sgn} B$ whenever $B \neq 0$. However, this additional invariant σ is essential when

©A K Peters
1058-6458/92 $0.50 per page

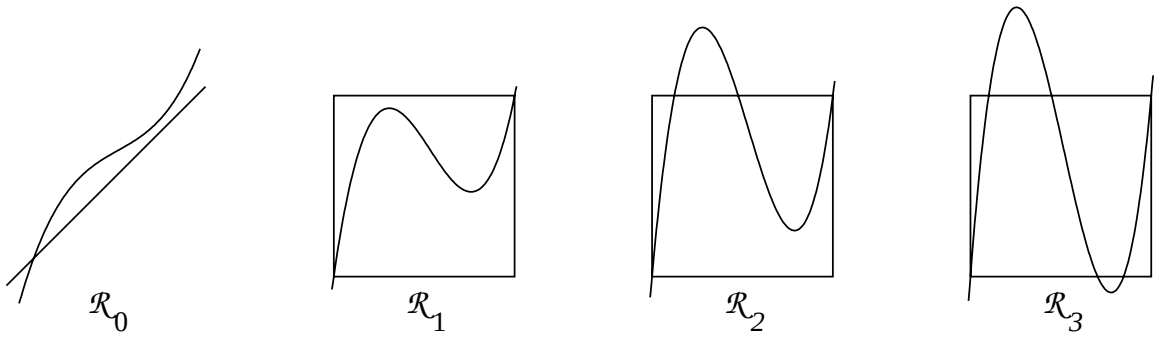

FIGURE 1. Representative graphs for the four different classes of real cubic polynomials. The case $f''' > 0$ is illustrated; corresponding examples with $f''' < 0$ can be obtained by looking at this figure in a mirror.

$B = 0$, for in this case, there are two essentially different real polynomial maps

$$x \mapsto x^3 - 3Ax \qquad \text{and} \qquad x \mapsto -x^3 - 3Ax,$$

which are conjugate over the complex numbers, but not over the real numbers. Thus *the moduli space of real affine conjugacy classes of real cubic maps can be described as the disjoint union of two closed half-planes, namely the half-plane* $A \in \mathbf{R}$, $B \geq 0$, $\sigma = +1$, *and the half-plane* $A \in \mathbf{R}$, $B \leq 0$, $\sigma = -1$. Any real cubic map is real affinely conjugate to one and only one map in the normal form

$$x \;\mapsto\; \sigma x^3 - 3Ax + \sqrt{|B|}\,. \tag{1.4}$$

(When $B \neq 0$, we can use the alternate normal form $\xi \mapsto B\xi^3 - 3A\xi + 1$.) In the two quadrants where $\sigma A \geq 0$, note that the associated real cubic map has real critical points, while in the remaining two quadrants, where $\sigma A < 0$, it has complex conjugate critical points. Further details may be found in Appendix A.

2. REAL CUBIC MAPS AS REAL DYNAMICAL SYSTEMS

Let's try to describe the behavior of the iterates of a cubic map $f : \mathbf{R} \to \mathbf{R}$, considered as a real dynamical system. We will denote the n-th iterate of a map f by $f^{\circ n}$. It is convenient to introduce the notation $K_{\mathbf{R}} = K_{\mathbf{R}}(f)$ for the compact set consisting of all points $x \in \mathbf{R}$ for which the orbit $\{x, f(x), f(f(x)), \ldots\}$ is bounded. This set $K_{\mathbf{R}}$ can be described as the real part of the "filled Julia set" of f (cf. Section 3).

We first introduce a very rough partition of each parameter half-plane for real cubics into four regions $\mathcal{R}_0$, $\mathcal{R}_1$, $\mathcal{R}_2$ and $\mathcal{R}_3$. More generally, we divide real polynomial maps f of degree $d \geq 2$ into $d+1$ distinct classes $\mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_d$, as follows. We will say that f belongs to the *trivial class* $\mathcal{R}_0$ if $K_{\mathbf{R}}(f)$ consists of at most a single point. (More precisely, $K_{\mathbf{R}}$ will consist of one fixed point when the degree is odd and will be vacuous when the degree is even.)

If f does not belong to this trivial class, there must be at least two distinct points in $K_{\mathbf{R}}(f)$. *Let I be the smallest closed interval that contains $K_{\mathbf{R}}(f)$.* Thus every orbit that starts outside of I must escape to infinity, but the two endpoints of I must have bounded orbits. In fact, it follows by continuity that each endpoint of I must map to an endpoint of I.

Definition. For $f \notin \mathcal{R}_0$, we will say that f belongs to the *class* $\mathcal{R}_n$ if the graph of f intersected with $I \times I$ has n distinct components (Figure 1). In other words, f belongs to $\mathcal{R}_n$ if the interval I can be partitioned into n closed subintervals that map into I (some of these intervals may be degenerate when $d > 3$), together with $n-1$ separating open intervals that map strictly outside of I. Note that $n \leq d$, since each of these open intervals must contain a critical point of f.

As an example, for degree $d = 2$, the quadratic map $x \mapsto x^2 + c$ belongs to

$$\begin{aligned} &\mathcal{R}_2 \quad \text{if} \quad c < -2\,, \\ &\mathcal{R}_1 \quad \text{if} \quad -2 \leq c \leq 1/4\,, \quad \text{and} \\ &\mathcal{R}_0 \quad \text{if} \quad 1/4 < c\,. \end{aligned}$$

For any degree d, note that f belongs to the class $\mathcal{R}_1$ if and only if the compact set $K_{\mathbf{R}}$ is a nontrivial interval (coinciding with I) or, in other words, if and only if this interval I maps into itself, with all orbits outside of I escaping to infinity. For f in $\mathcal{R}_n$ with $n \geq 2$, at least $n-1$ of the *critical orbits*,

that is, the orbits of the critical points, must be real and must escape to infinity. The case $n = d$ is of particular interest. *If $f \in \mathcal{R}_d$, all of the critical orbits escape to infinity.* Furthermore, the interval I contains d disjoint subintervals, each of which is mapped diffeomorphically onto the entire interval I. A rather delicate argument, following [Guckenheimer 1979, §§2.8, 3.1], then shows that *the set $K_{\mathbf{R}}$ is a Cantor set of measure zero.* Furthermore, the restriction $f|K_{\mathbf{R}}$ is homeomorphic to a one-sided shift on d symbols. The degree d polynomials in $\mathcal{R}_d$ have maximal topological entropy equal to $\log d$. (Compare equation (2.4) and Figure 15.) They have the property that their complex periodic points are all distinct and contained in the real interval I. It follows that their (complex) Julia set coincides with the Cantor set $K_{\mathbf{R}} \subset \mathbf{R}$.

We now specialize to the cubic case $d = 3$. In order to separate the four classes of real cubic maps, we introduce four curves in the parameter plane, as follows (Figure 2).

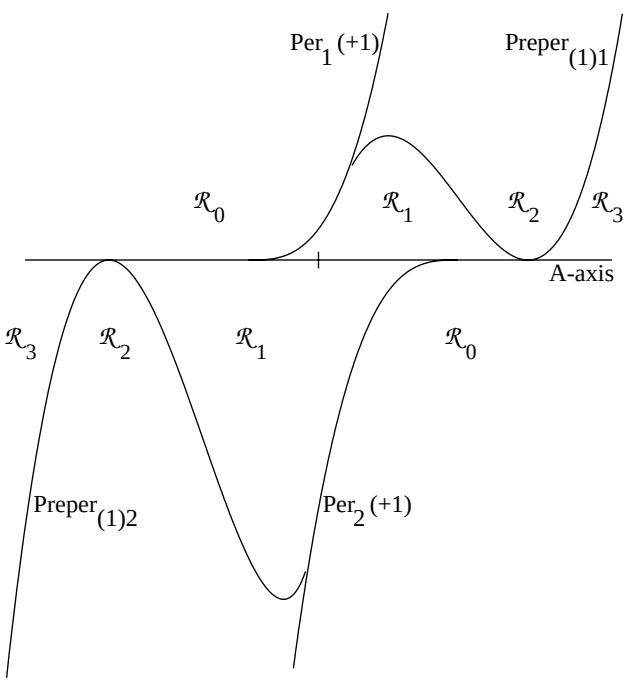

FIGURE 2. The four regions in the (A, B)-parameter plane, and the curves separating them.

Definition. We denote by $\mathrm{Per}_p(\mu)$ be the set of parameter pairs (A, B) for which the associated cubic map f has a periodic orbit of period p with multiplier $(f^{\circ p})'$ equal to μ. In particular, the curve

$$\mathrm{Per}_1(+1): \quad B = 4(A + \tfrac{1}{3})^3$$

consists of all parameter pairs for which the graph of f is tangent to the diagonal, while

$$\mathrm{Per}_2(+1): \quad B = 4(A - \tfrac{2}{3})^3$$

gives maps for which the graph of $f \circ f$ is tangent to the diagonal. Such points of tangency are called *saddle nodes* of period one or two, respectively.

Similarly, let $\mathrm{Preper}_{(t)p}$ be the curve of parameter pairs for which one critical point, say $+a$, is *preperiodic*, with $f^{\circ t}(a)$ lying on a periodic orbit of period $p \geq 1$. Here we assume that t is minimal and strictly positive. Thus the curve

$$\mathrm{Preper}_{(1)1}: \quad B = 4A(A-1)^2$$

gives maps such that one critical point maps to a fixed point of f, while

$$\mathrm{Preper}_{(1)2}: \quad B = -(1 \pm (2A+1)\sqrt{-A})^2$$

in the quadrant $A, B \leq 0$, gives maps such that one critical point maps to an orbit of period two. For further details, see Appendix A.

We can pass between the cases $\mathcal{R}_0$, $\mathcal{R}_1$, $\mathcal{R}_2$ and $\mathcal{R}_3$ only by crossing over one or more of these curves. In fact, we need only the curves $\mathrm{Per}_1(1)$ and $\mathrm{Preper}_{(1)1}$ in the half-plane $\sigma = 1$, $B \geq 0$, as can be verified by the study of Figure 1. Similarly, we need only the curves $\mathrm{Per}_2(1)$ and $\mathrm{Preper}_{(1)2}$ in the half-plane $\sigma = -1$, $B \leq 0$. Graphs of these four curves and the corresponding division of each parameter half-plane into four regions are shown in Figure 2, with irrelevant segments of the curves removed.

(A similar description of the case boundaries can be given for the $(d-1)$-parameter family consisting of suitably normalized polynomials of degree d. There are analogous hypersurfaces $\mathrm{Per}_p^d(\mu)$ and $\mathrm{Preper}_{(t)p}^d$ that separate the $d+1$ regions $\mathcal{R}_i$. For d odd, the description is very much like that in the cubic case, while for d even, we need just three hypersurfaces, namely $\mathrm{Per}_1^d(+1)$ and $\mathrm{Preper}_{(2)1}^d$ in all cases, and also $\mathrm{Preper}_{(1)1}^d$ when $d \geq 4$.)

In the regions $\mathcal{R}_1$ and $\mathcal{R}_2$ of the cubic parameter plane, there are many possibilities for complex behavior. Some of the different kinds of behavior are distinguished in Figure 3. In the region $\mathcal{R}_2$, we know that at least one of the two critical orbits must escape to infinity, but the other critical orbit

may either escape (indicated by white in the figure) or remain bounded (indicated by light gray). Similarly, in case $\mathcal{R}_1$, the two critical orbits may both behave chaotically (dark gray), or one or both may converge to attracting periodic orbits (lighter shades). The regions $\mathcal{R}_0$ and $\mathcal{R}_3$ are colored white in this figure, since they correspond to relatively dull dynamical behavior. For a discussion of the methods used to make such figures, as well as their limitations, see Appendix C.

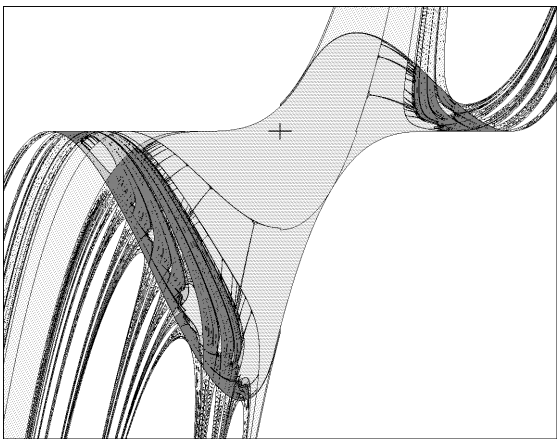

FIGURE 3. Picture of the (A, B)-parameter plane, indicating the boundaries between regions of qualitatively different dynamic behavior. In the dark region, both critical orbits behave chaotically, and in the white region both critical orbits escape to infinity. Intermediate shades indicate various intermediate forms of behavior. The illustrated region is the rectangle $[-1.2, 1.2] \times [-1.85, .75]$.

Remark. For many purposes, it is more natural to work in the (A, b)-parameter plane, for $b = \pm\sqrt{B}$. The corresponding bifurcation diagram is shown in Figure 4. Of course, this figure incorporates only real cubics with positive leading coefficient. For an analogous parametrization of cubics with negative leading coefficient, we must work in the (A, b')-plane, where $b' = \pm\sqrt{-B}$ so that $B = -(b')^2 \leq 0$ (see Figure 5). Figure 4 can be described roughly as the "double" of the upper half-plane in Figure 3, and Figure 5 as the double of the lower half-plane.

Inspection of Figures 3, 4, and 5 under magnification shows that several characteristic patterns are repeated many times on different scales. Noteworthy are the swallow-shaped regions (Figure 7), arch-shaped regions (Figure 11) and productlike regions (Figure 13). We can partially explain these

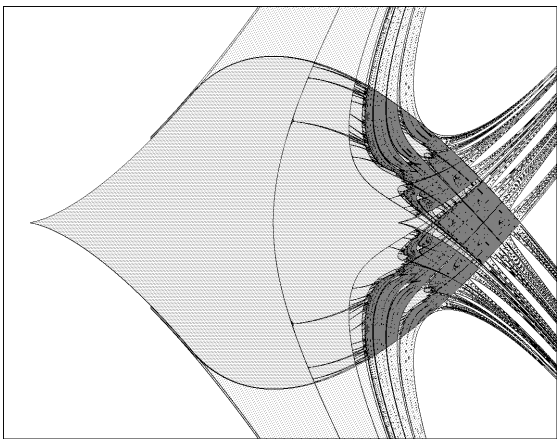

FIGURE 4. Picture in the $(A, \sqrt{B})$-plane, with the same conventions as Figure 3. The illustrated region is $[-.4, 1.1] \times [-1, 1]$.

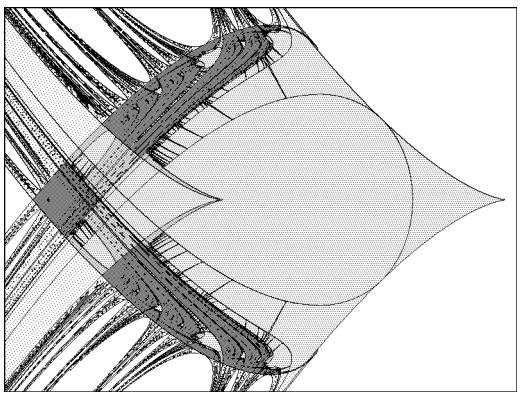

FIGURE 5. Picture in the $(A, \sqrt{-B})$-plane, with the same conventions as Figure 3. The illustrated region is $[-1.1, .7] \times [-1.4, 1.4]$.

regions in terms of the dynamics of the associated maps f as follows.

Definition. A smooth map $f : \mathbf{R} \to \mathbf{R}$ with one or more critical points is said to be *renormalizable* if there exists a neighborhood U of the set of critical points so that

1. each component of U contains at least one critical point,
2. the first return map $\hat{f}$ from U to itself is defined and smooth, and
3. the union $U \cup f(U) \cup f^{\circ 2}(U) \cup \cdots$ has at least two distinct components.

Condition 2 says that for each $x \in U$ there exists an integer $n \geq 1$ with $f^{\circ n}(x) \in U$, and that the smallest such integer $n = n(x)$ is constant on each connected component of U. Condition 3 says that we exclude the trivial case where U is connected and maps into itself.

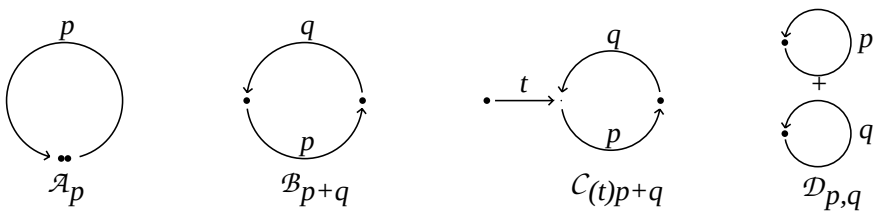

FIGURE 6. Schematic diagrams for maps representing the centers of the four distinct classes of hyperbolic components. Each critical point is indicated by a heavy dot, and each arrow is labeled by a corresponding number of iterations (cf. Section 4 and Appendix B).

More explicitly, a real cubic map f with (not necessarily distinct) real critical points is renormalizable if and only if it fits into one of the following four classes (see Figure 6).

$\mathcal{A}$. **Adjacent Critical Points.** There is an open interval U containing both critical points and an integer $p \geq 2$ so that the intervals $U, f(U), \ldots, f^{\circ p-1}(U)$ are pairwise disjoint, but $f^{\circ p}(U) \subset U$.

$\mathcal{B}$. **Bitransitive.** There exist disjoint intervals U_1 and U_2 about the two critical points so that the first return map from the union $U = U_1 \cup U_2$ to itself is defined and smooth, interchanging these two components. In other words, $f^{\circ p}(U_1) \subset U_2$ and $f^{\circ q}(U_2) \subset U_1$ for some $p \geq 1$ and $q \geq 1$. We will see that a universal model for this behavior occurs in a "biquadratic" map, that is, the composition of two quadratic maps.

$\mathcal{C}$. **Capture.** (Caution: I am told that the term "capture" is used with a different meaning in [Wittner 1988]. Compare [Rees 1992].) There are neighborhoods U_1 and U_2 as in Case $\mathcal{B}$, but the first return map carries both U_1 and U_2 into U_2. Thus the orbit of U_1 is "captured" by the periodic orbit of U_2.

$\mathcal{D}$. **Disjoint Periodic Sinks.** Again there are disjoint neighborhoods U_1 and U_2, but in this case, the first return map carries each U_i into itself, say $f^{\circ p}(U_1) \subset U_1$ and $f^{\circ q}(U_2) \subset U_2$.

In all four cases, the corresponding configuration in the (A, B)-parameter plane has a unique well-defined *center point* f_0, characterized by the property that the first return map $\hat{f}_0$ carries critical points to critical points (see Section 4). Thus this center map f_0 has the Thurston property of being *postcritically finite*. In fact, f_0 has the sharper property that the orbits of both critical points are finite, and eventually superattracting. It follows from Thurston's theory that this center point f_0 is uniquely determined by its "kneading invariants" or, in other words, by the mutual ordering of the various points on the two critical orbits (see [Douady and Hubbard 1984] and the analogous discussion for quadratic maps in [Milnor and Thurston 1988, §13.4]). Furthermore, any ordering that can occur for a continuous map with two eventually superattracting critical orbits can actually occur for a cubic polynomial map.

Case $\mathcal{A}$ is exceptional and occurs only in one region, which has center point $(A, B) = (0, -1)$ corresponding to the map $f_0(z) = 1 - z^3$. In Cases $\mathcal{B}, \mathcal{C}, \mathcal{D}$, we will see that the corresponding point of the real (A, B)-parameter plane is associated, respectively, to a *swallow configuration*, to an *arch configuration*, or to a *product configuration* (Figures 7, 11, 13). There are two qualifications: If such a configuration is immediately adjacent and subordinated to another larger configuration, it'll be highly deformed. Furthermore, along the A-axis the pictures in the (A, B)-plane are rather strange, and one should rather work with the (A, b)- or (A, b')-plane, as in Figures 4 and 5.

In each of these cases $\mathcal{B}, \mathcal{C}, \mathcal{D}$, the first return map from $U_1 \cup U_2$ to itself can be approximated by a map that is quadratic on each component. Hence we can construct a simplified prototypical model for this kind of behavior by replacing each interval U_k by a copy $k \times \mathbf{R}$ of the entire real line and by replacing the smooth map $\hat{f} : U_1 \cup U_2 \to U_1 \cup U_2$, which has one critical point in each component, by a componentwise quadratic map

$$(k, x) \mapsto (\varphi(k), x^2 + c_k)$$

from the disjoint union $\{1, 2\} \times \mathbf{R} \approx \mathbf{R} \sqcup \mathbf{R}$ to itself.

First consider the case of a swallow configuration, as illustrated in Figure 7. The prototypical model in this case is obtained by replacing these two intervals by disjoint copies of the real line with

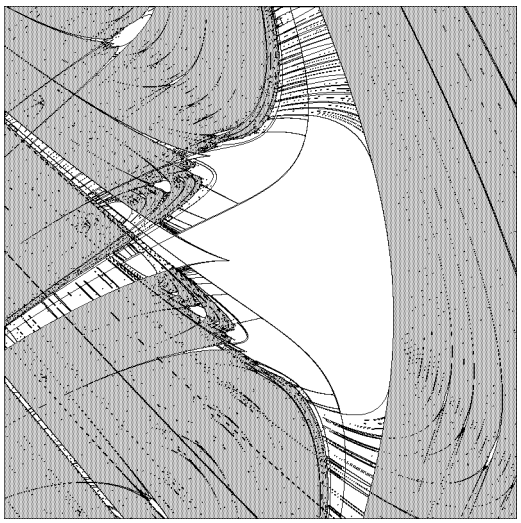

FIGURE 7. Detail of Figure 3 showing a "swallow configuration" centered at $(A, B) = (-.5531, -.6288)$. For the cubic map associated with this central point, the two critical points $\pm a$ satisfy $f(f(a)) = -a$ and $f(f(-a)) = a$. Hence both lie on a common orbit of period 4. The region shown is $[-.6, -.53] \times [-.7, -.55]$.

parameters x and y, respectively, and by replacing the first return map by the quadratic map

$$x \mapsto y = x^2 + c_1\,, \qquad y \mapsto x = y^2 + c_2\,, \tag{2.1}$$

which interchanges the two components of the disjoint union $\mathbf{R} \sqcup \mathbf{R}$. Here c_1 and c_2 are real parameters. Thus we obtain a *universal swallow configuration* in the (c_1, c_2)-parameter plane, as illustrated in Figure 8 (cf. [Ringland and Schell 1990]). The central teardrop-shaped body of this swallow corresponds to the *connectedness locus* for this family, consisting of those biquadratic maps for which both critical orbits remain bounded (cf. Section 3). On the other hand, the wings and tails correspond to maps for which only one critical orbit is bounded.

Remark. It is interesting to note that this same swallow configuration seems to occur in a quite different context, where there are no critical points at all. Consider the two-parameter family of *Hénon maps*. These are quadratic diffeomorphisms of the plane that can be written, for example, as

$$(x, y) \mapsto (y,\ y^2 - \alpha - \beta x), \tag{2.2}$$

with constant Jacobian determinant β. A picture of those parameter pairs (α, β) for which there exists an attracting periodic orbit typically exhibits

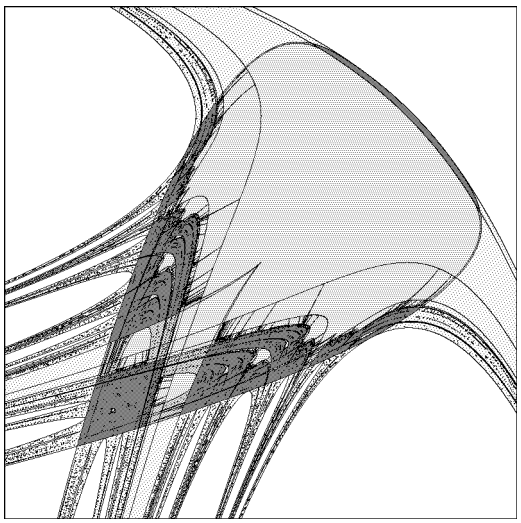

FIGURE 8. The prototype swallow configuration in the (c_1, c_2)-parameter plane, associated with the family of biquadratic maps $x \mapsto (x^2 + c_1)^2 + c_2$ from the real line to itself. The region shown is $[-2.5, 1] \times [-2.5, 1]$.

quite similar swallow-shaped configurations [El Hamouly and Mira 1982]. For example, such a region is shown in Figure 9, corresponding to an attracting orbit of period five. This phenomenon can be explained intuitively as follows. If $|\beta|$ is small, the dynamics of the two-dimensional Hénon map is quite similar to the dynamics of the one-dimensional map $y \mapsto y^2 - \alpha$. In particular, the Hénon map can be closely approximated locally by

FIGURE 9. A swallow configuration in the Hénon parameter plane. A location (α, β) is colored white if a random search of initial conditions found an attracting orbit of low period for the quadratic diffeomorphism $(x, y) \mapsto (y,\ y^2 - \alpha - \beta x)$; gray indicates that only high periods or chaotic behavior were found. In the black area to the lower right, no bounded orbits were found. The graininess in the picture is presumably due to the random nature of the algorithm used. Region: $[1.4, 1.6] \times [-.3, -.1]$.

a linear map, except at points near the axis $y = 0$, where the second derivative plays an essential role. Similarly, the dynamics of a composition of two Hénon maps with small determinant resembles the dynamics of a composition of two one-dimensional quadratic maps. Now consider a periodic orbit for some Hénon map. If this orbit is to be attracting, it must contain at least one point that is close to the axis $y = 0$. If exactly two points of the orbit are close to $y = 0$, the dynamics will resemble that for a composition of two quadratic maps. Hence, in this case, as we vary the parameters, we obtain a swallow-shaped configuration within the Hénon parameter plane.

Caution. The swallow configuration of Figures 7, 8 and 9 should not be confused with the somewhat similar configuration shown in Figure 4, which can perhaps be described as a "pointed swallow." This pointed configuration also plays a role in many dynamical systems. Here is a well-known example. (I am indebted to communications from S. Ushiki and T. Matsumoto.) Consider the two-parameter family of circle maps

$$t \;\mapsto\; t + c + k\sin(2\pi t) \qquad (\mathrm{mod}\ 1)\,. \qquad (2.3)$$

These are diffeomorphisms for $|2\pi k| < 1$, but have two critical points for larger values of $|k|$. The corresponding picture in the (c, k)-parameter plane, shown in Figure 10, contains one immersed copy of the configuration of Figure 4 corresponding to each rational rotation number [Chavoya-Aceves et al. 1985]. Each of these configurations terminates in a "tongue" that reaches down to the corresponding rational point on the $k = 0$ axis. These are known as *Arnold tongues*.

Next let's consider the arch configuration, as illustrated in Figure 11. Recall that a point of the cubic parameter plane belongs to an *arch configuration* if there are disjoint neighborhoods U_1 and U_2 as described earlier so that some iterate of f maps U_1 into U_2, and some iterate maps U_2 into itself, but so that every forward image of U_1 or U_2 is disjoint from U_1. In this case, the universal configuration, as illustrated in Figure 12, is obtained by studying a quadratic map from $\mathbf{R} \sqcup \mathbf{R}$ to itself depending on two parameters c and $\hat{x}$ as follows. We map a point ξ in the first copy of $\mathbf{R}$ to the point $x = \pm\xi^2 + \hat{x}$ in the second copy so that the critical point maps to $\hat{x}$, and we map the second copy of

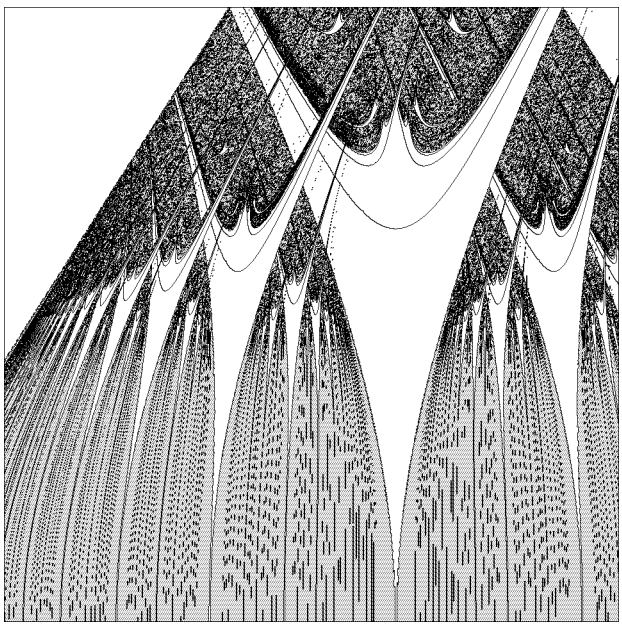

FIGURE 10. Arnold tongues ending in "pointed-swallow" configurations for the family of circle maps $t \;\mapsto\; t + c + k\sin(2\pi t)$. Region: $[.15, .7] \times [0, .35]$ in the (c, k)-parameter plane.

$\mathbf{R}$ to itself by $x \mapsto x^2 + c$. The real connectedness locus in this prototypical case consists of all pairs $(c, \hat{x})$ with $-2 \le c \le 1/4$ and $2|\hat{x}| \le 1 + \sqrt{1 - 4c}$.

Finally, take the product configuration of Figure 13. We say that a point of the cubic parameter plane belongs to a *product configuration* if there are disjoint neighborhoods U_1 and U_2 as given earlier

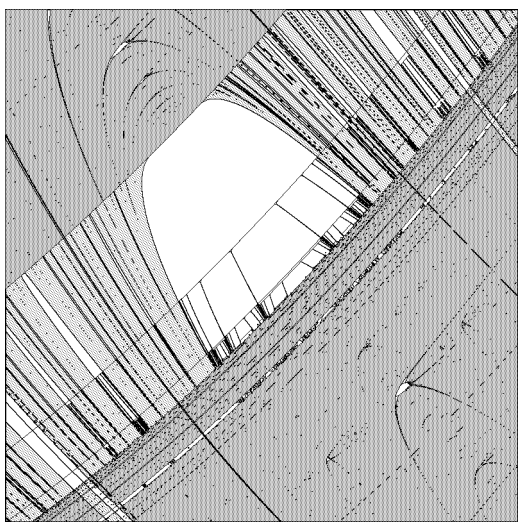

FIGURE 11. Detail of Figure 3 showing an arch configuration. For the cubic map corresponding to the center point $(A, B) = (.8536, .0243)$, the critical points $\pm a$ satisfy $f(f(a)) = f(f(-a)) = \pm a$. Region: $[.835, .885] \times [.01, .03]$.

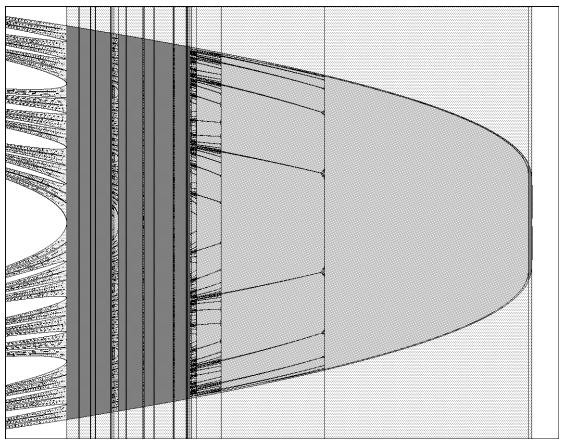

FIGURE 12. The prototype arch configuration in the $(c, \hat{x})$-plane. Here we consider the orbit of the point $\hat{x}$ under the map $x \mapsto x^2 + c$. Dark gray indicates that both $\hat{x}$ and 0 have chaotic orbits, while white means that both escape to infinity. Region: $[-2.3, .4] \times [-2.2, 2.2]$.

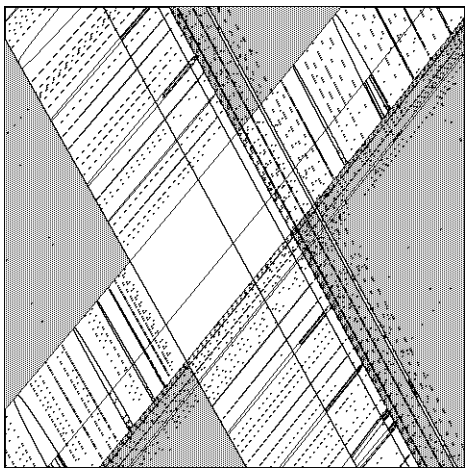

FIGURE 13. Detail of Figure 3 showing a product configuration. For the map corresponding to the center point (.8156, .0674), there are two superattracting periodic orbits, with periods three and four. Region: [.814, .819] × [.0665, .0685].)

so that some iterate of f maps U_1 into itself and some iterate maps U_2 into itself, but no forward image of either one of the U_i intersects the other. In this case, the universal model is obtained by taking two disjoint real lines, say with parameters x and y, and by looking at independent quadratic maps $x \mapsto x^2 + c_1,\ y \mapsto y^2 + c_2$. The "real connectedness locus" for this two-parameter family, that is, the set of parameter pairs for which both critical points have bounded orbits, is evidently equal to the square $[-2, \frac{1}{4}] \times [-2, \frac{1}{4}]$ in the (c_1, c_2)-plane, as illustrated in Figure 14.

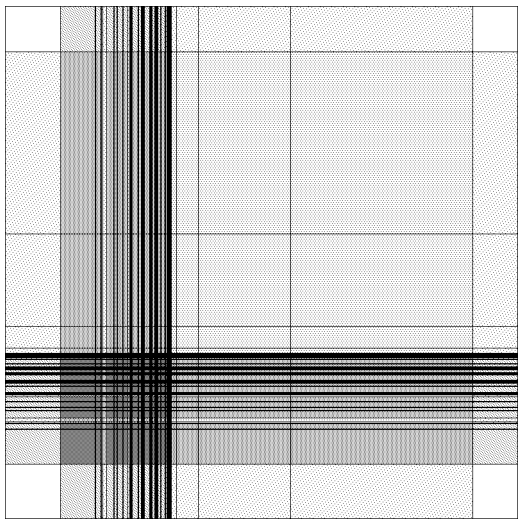

FIGURE 14. The prototype product configuration in the (c_1, c_2)-parameter plane.

According to [Jakobson 1981], the set of parameter pairs for which both critical orbits are chaotic (indicated by dark gray in the figure) has positive measure (see also [Benedicks and Carleson 1985] and [Rychlik 1988]). A classical conjecture, not yet proved, asserts that this set is totally disconnected. Thus it seems natural to make the corresponding conjecture for the cubic parameter plane of Figure 3 that the set of parameter pairs for which both critical orbits are chaotic is a totally disconnected set of positive measure.

Further discussion of these shapes, and other related ones, will be given in Section 4, which discusses the corresponding four cases for complex cubic maps, and in Appendix B.

A useful tool for studying real polynomial mappings f of degree d is provided by the *topological entropy* $0 \le h(f) \le \log d$ of f considered as a map from the compact interval $[-\infty, \infty]$ to itself. According to [Rothschild 1971] and [Misiurewicz and Szlenk 1977], the topological entropy can be computed as

$$h(f) = \lim_{k \to \infty} \frac{1}{k} \log \ell(f^{\circ k}), \tag{2.4}$$

where $\ell(f^{\circ k}) - 1$ is the number of points along the real axis at which the derivative $x \mapsto df^{\circ k}(x)/dx$ changes sign [Milnor and Thurston 1988]. In the cubic case, a more practical algorithm for computing h has recently been described [Block and Keesling].

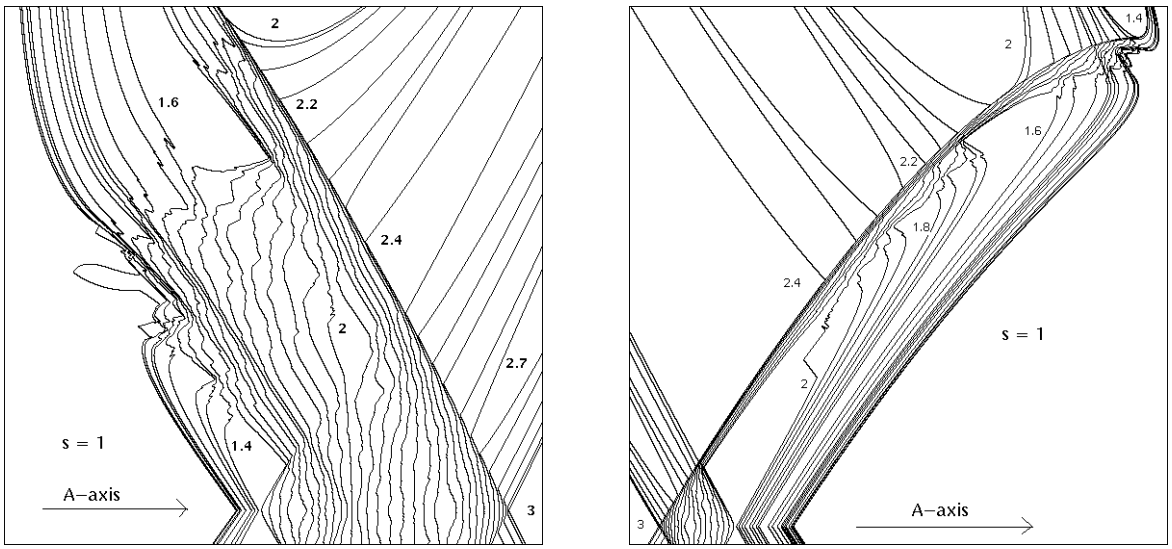

FIGURE 15. "Curves" of constant growth number s in the (A, b)-parameter plane, for the families of maps $x \mapsto x^3 - 3Ax + b$ (left) and $x \mapsto -x^3 - 3Ax + b$ (right). (Compare Figures 4 and 5.) The curves $s = .05, .1, \ldots, 2^-, 2.05, \ldots, 3^-$ are plotted, using an algorithm due to Block and Keesling. The regions shown are $[.57, 1.03] \times [-.03, .43]$ (left) and $[-1.05, -.02] \times [-.05, 1.35]$ (right).

The entropy $h(f)$ varies continuously as f varies through polynomials of fixed degree. Furthermore, h takes a constant value, equal to the logarithm of an algebraic integer, throughout each hyperbolic component (cf. Section 4). In particular, in the cubic case, the entropy of the map $x \mapsto x^3 - 3Ax + b$ depends continuously on the two parameters A and b, and similarly the entropy of

$$x \mapsto -x^3 - 3Ax + b$$

depends continuously on the parameters A and b.

It is often convenient to set $h = \log s$, where the *growth number* $s = e^h$ varies over the interval $1 \le s \le 3$ in the cubic case. Figure 15 shows the level sets of s in the (A, b)-plane for the families of maps $x \mapsto x^3 - 3Ax + b$ (left) and $x \mapsto -x^3 - 3Ax + b$ (right): compare Figures 4 and 5. In each case, we see points both inside and outside the real connectedness locus. At least part of the boundary curves $\text{Preper}_{(1)1}$ and $\text{Preper}_{(2)1}$ for the connectedness locus is clearly visible in these figures as a locus where the level sets change shape dramatically. I have not tried to plot the boundary of the region $s = 1$, although this would be a locus of particular interest (cf. [MacKay and Tresser]).

In the quadratic case, it is known that the topological entropy h (or equivalently the growth number $s = e^h$) for the map $x \mapsto x^2 + c$ is a monotone decreasing function of the parameter c (see, for example, [Milnor and Thurston 1988]). A corresponding conjecture for the cubic case would be that *each level set* $s(A, b) =$ constant *is a connected subset of the* (A, b)*-parameter plane* and, in particular, that the continuous function $(A, b) \mapsto s(A, b)$ has no isolated local maxima or minima. The conjecture applies to both families of maps, $x \mapsto x^3 - 3Ax + b$ and $x \mapsto -x^3 - 3Ax + b$.

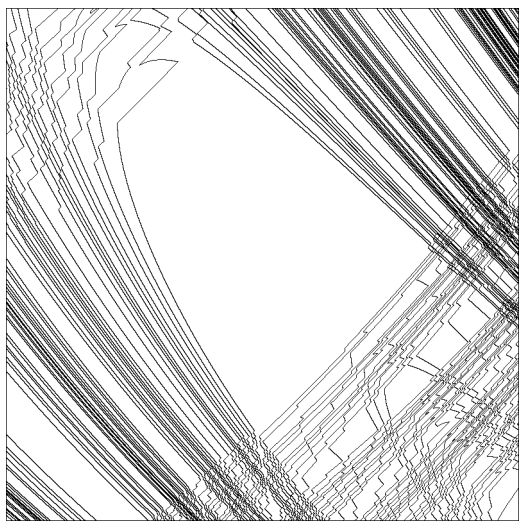

FIGURE 16. Curves of constant s around an arch configuration in the (A, B)-plane. The illustrated region $[0.835, 0.885] \times [0.01, 0.03]$ is the same as that shown in Figure 11. The contour interval is $\Delta s = .002$.

Note that these level sets are not always curves. They may well have interior points. For example, this is the case for $s = 1, 2, 3$ and also for

$s = (1 + \sqrt{5})/2$. It is conjectured that there are interior points if and only if this locus contains hyperbolic maps. In particular, it is conjectured that this can happen only when s is an algebraic integer (cf. Appendix B and Figure 16.)

3. COMPLEX CUBICS: THE CONNECTEDNESS LOCUS

In this section, we consider the dynamics of a complex cubic map. Following [Douady and Hubbard 1984], for any complex polynomial map $f : \mathbf{C} \to \mathbf{C}$ of degree $d \geq 2$, we use the notation $K(f)$ for the *filled Julia set*, consisting of all complex numbers z for which the orbit of z under f is bounded. This set $K(f)$ is connected if and only if it contains all of the critical points of f. On the other hand, if $K(f)$ contains no critical points, f is a "degree d complex horseshoe" in the sense that there exists a disk $D \supset K(f)$ smoothly embedded in $\mathbf{C}$ so that $f^{-1}(D)$ consists of d disjoint subdisks, each of which maps diffeomorphically onto D under f. In particular, f restricted to $K(f)$ is isomorphic to a one-sided shift on d symbols (compare [Blanchard et al. 1991]).

Branner and Hubbard [1988] define the *connectedness locus* for a parametrized family of polynomial maps to be the set of all parameter values that correspond to polynomials f for which $K(f)$ contains all of the critical points, or equivalently is connected. As an example, the connectedness locus for the family of complex quadratic maps $z \mapsto z^2 + c$ is also known as the *Mandelbrot set* (Figure 17). This set has been extensively studied by Douady and Hubbard, who show, for example,

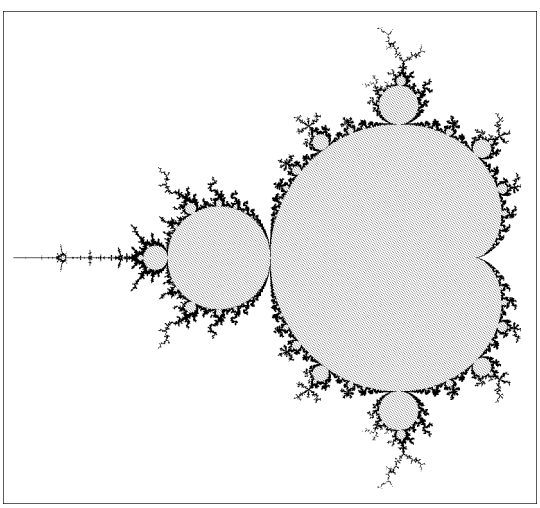

FIGURE 17. The Mandelbrot set.

that it is connected, with connected complement [Douady and Hubbard 1982]. In the cubic case, Branner and Hubbard [1988] show that the connectedness locus is again compact and connected, with connected complement. In fact, more precisely, it is "cellular"; that is, it can be expressed as the intersection of a strictly nested sequence of closed four-dimensional disks $D_{i+1} \subset \mathring{D}_i$ in the parameter space $\mathbf{C}^2$ [Brown 1960; 1961]. The corresponding assertion for higher degrees has recently been proved [Lavaurs 1989].

However, there seem to be at least three significant differences between the quadratic and cubic cases. To discuss them, we will need the following definition. Following [Douady and Hubbard 1984], a component of the interior of a complex connectedness locus is called *hyperbolic* if every critical orbit of any associated polynomial map converges towards an attracting periodic orbit (cf. Section 4).

(1) The Mandelbrot set is replete with small copies of itself. In fact, [Douady and Hubbard 1985] shows that each hyperbolic component of the interior of the Mandelbrot set is embedded as the central region of a small copy of the full Mandelbrot set. However, in the cubic case, there is is not just one kind of hyperbolic component, but rather four essentially distinct types, each associated with a characteristic local shape.

(2) In the quadratic case, the hyperbolic components are organized in a one-dimensional treelike manner. To make this statement more precise, we could say that the hyperbolic components of period $\leq p_0$ are connected to each other within the Mandelbrot set like the vertices of a tree. In the cubic case, there is certainly no such treelike organization. A corresponding conjecture might be that finite sets of hyperbolic components are organized as vertices of an acyclic two-dimensional complex.

(3) It is widely believed that the Mandelbrot set is locally connected. Yoccoz (unpublished) has made important progress towards a proof in recent years. However, local connectivity definitely fails for the cubic connectedness locus; see [Lavaurs 1989] as well as the following discussion. In fact, pictures such as Figure 20 suggest that the cubic connectedness locus may not even be pathwise connected.

It is difficult to visualize this complex cubic connectedness locus, which is an extremely complicated four-dimensional object with fractal boundary [Dewdney 1987]. A more accessible situation arises if we consider the dynamics of cubic polynomial maps $f : \mathbf{C} \to \mathbf{C}$ that have real coefficients and hence are effectively described by points in the real (A, B)-parameter plane. In particular, we can intersect the Branner–Hubbard connectedness locus with the real (A, B)-plane. The resulting intersection is shown in Figure 18. Here, for

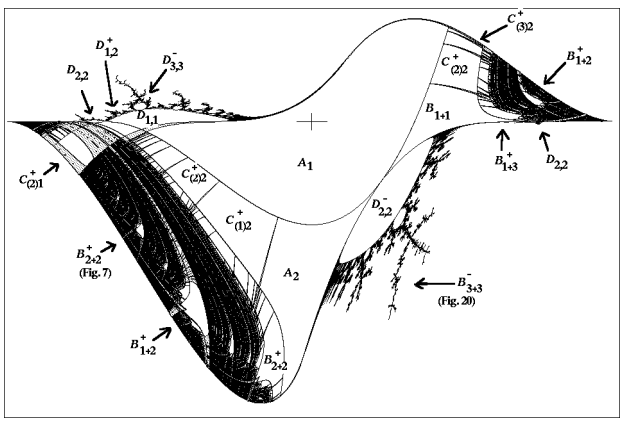

FIGURE 18. The complex connectedness locus intersected with the real (A, B)-plane. The region shown is $[-1, 1] \times [-1.7, .65]$.

parameter pairs in the outside white region, one or both critical orbits escape to infinity, while in the inside white regions both converge to periodic orbits. Gray and black indicate, respectively, that one or both critical orbits behave chaotically. In the two quadrants where $AB > 0$, so that the critical points are real, the connectedness locus coincides with the region $\mathcal{R}_1$, as described in Section 2, and is bounded by smooth curves. For parameter values in the regions $\mathcal{R}_2$ and $\mathcal{R}_3$ of Section 2, recall that at least one of the two critical orbits necessarily escapes. Hence this region is white in the figure. Within the two quadrants where $AB > 0$, the behavior of the iterates of f as a real dynamical system effectively determines the behavior as a complex dynamical system. However, in the two quadrants where the critical points are complex, this real part of the connectedness locus is a very complicated object with fractal boundary. (In these complex quadrants, note that both critical orbits must behave in the same way, since they are complex conjugates.) The notations $\mathcal{A}$–$\mathcal{D}$ in Figure 18 are explained in Section 2 (Figure 6), Section 4 and Appendix B; and the sign of AB is used as superscript.

Just as in the full complex case, this real part of the connectedness locus is compact and cellular, as can be proved by the methods of [Branner and Hubbard 1988]. Alternatively, using the Smith theory, as described in [Bredon 1972, p. 145], since the real connectedness locus is the fixed point set of an involution on the complex connectedness locus, which has the Čech cohomology of a point, it follows that the real connectedness locus also has the mod 2 Čech cohomology of a point. In particular, it is connected, with connected complement.

The shape of this locus in the two complex quadrants $AB < 0$ is reminiscent of the Mandelbrot set (Figure 17), and, in fact, we will see in Section 4 that there are many small copies of the Mandelbrot set embedded in these quadrants. However, these embedded copies tend to be discontinuously distorted at one particular point, namely the period-one saddle node point $c = 1/4$, also known as the *root point* of the Mandelbrot set. This phenomenon is particularly evident in the lower right quadrant, which exhibits a very fat copy of the Mandelbrot set with the root point stretched out to cover a substantial segment of the saddle node curve $\mathrm{Per}_2(1)$ (cf. Section 2). As a result of this stretching, the cubic connectedness locus fails to be locally connected along this curve (Figure 19). This behavior, which has been studied by Lavaurs [Lavaurs 1989], is in drastic contrast to the situation for degree-two maps. In fact, as noted earlier, it is widely believed, although not yet proved, that the Mandelbrot set is locally connected.

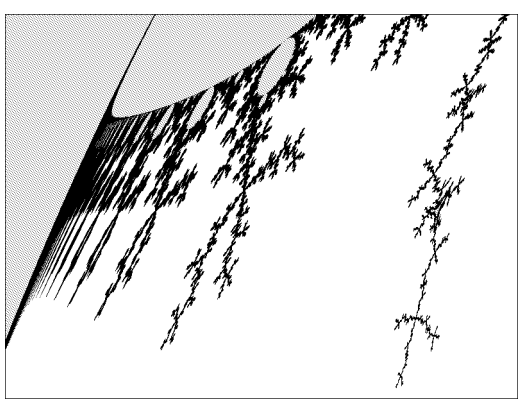

FIGURE 19. Detail in the lower right quadrant, showing lack of local connectivity. The region displayed is $[.02, .32] \times [-1.15, -.7]$.

4. HYPERBOLIC COMPONENTS

We continue to study the two-parameter family of affine conjugacy classes of cubic maps. Recall that a complex cubic map f, or the corresponding point (A, B) in complex parameter space, belongs to the *connectedness locus* if the (forward) orbits of both critical points under f are bounded, and it is *hyperbolic* if both of these critical orbits converge towards attracting periodic orbits. Here, by definition, an orbit $f^{\circ p}(z_0) = z_0$ of period $p \geq 1$ is called *attracting* if the *multiplier* $df^{\circ p}(z)/dz$ (that is, the characteristic derivative around the orbit) has absolute value less than one. The set of all hyperbolic points in the complex parameter plane forms an open set, which is conjectured to be precisely equal to the interior of the connected locus and to be everywhere dense in the connectedness locus. Each connected component of this open set is called a *hyperbolic component* of the connectedness locus.

These definitions make equally good sense for the real part of the connectedness locus. Again, it is conjectured that the hyperbolic points are everywhere dense. However, it is clearly not true that every interior point of the real connectedness locus is hyperbolic.

The discussion of hyperbolic components will be strongly influenced by the work of Rees [1990], who has studied the closely analogous problem of iterated rational maps of degree two from the sphere $\mathbf{C} \cup \infty$ to itself. I am indebted to Douady for the observation that her methods and conclusions apply, with minor modifications, to our case of iterated cubic polynomial maps. In particular, her methods show that *each hyperbolic component contains a unique preferred point, characterized by the property that the forward orbit of each of the two critical points under the associated map is finite, and hence eventually periodic.* (Compare [McMullen 1988].) Following [Douady and Hubbard 1982], this preferred point is called the *center* of the hyperbolic component. If the hyperbolic component intersects the real (A, B)-plane, note that its center must be self-conjugate, and hence real.

These ideas are given further development in [Milnor 1992], which studies monic centered polynomial maps of any degree $d \geq 2$ over $\mathbf{R}$ or $\mathbf{C}$, showing that every hyperbolic component is a topological cell with a preferred center point.

In analogy with [Rees 1990], the different hyperbolic components in the complex cubic connectedness locus can be roughly classified into four different types, as follows (compare Section 2 and Figure 6). Fixing some hyperbolic cubic map f, let $U \subset \mathbf{C}$ be the open set consisting of all complex numbers z whose forward orbit under f converges to an attracting periodic orbit. Note that f maps each component of U precisely onto a component of U.

$\mathcal{A}_p$. **Adjacent Critical Points.** Here both critical points belong to the same component U_0 of this attractive basin U. This component is necessarily periodic, in the sense that $f^{\circ p}(U_0) = U_0$ for some integer $p \geq 1$.

$\mathcal{B}_{p+q}$. **Bitransitive.** The two critical points belong to different components U_0 and U_1 of U, but there exist integers $p, q > 0$ satisfying the conditions $f^{\circ p}(U_0) = U_1$ and $f^{\circ q}(U_1) = U_0$. We assume that p and q are minimal, so that both U_0 and U_1 have period $p + q$.

$\mathcal{C}_{(t)p+q}$. **Capture.** Again the critical points belong to different components, but only one of the two, say U_1, is periodic. In this case, some forward image of U_0 must coincide with U_1. More precisely, there is a unique smallest integer $t + p \geq t \geq 1$ so that $f^{\circ t}(U_0)$ coincides with some forward image $f^{\circ q}(U_1)$, and so that $f^{t+p}(U_0) = U_1$, where U_1 has period $p + q$. In this case, the product tq is always two or more. However, p may be zero, in which case we write simply $\mathcal{C}_{(t)q}$.

$\mathcal{D}_{p,q}$. **Disjoint Periodic Sinks.** The two critical points belong to different components U_0 and U_1, where no forward image of U_0 is equal to U_1, and no forward image of U_1 is equal to U_0. In this case, each of the two components U_0 and U_1 must be periodic, although their periods p and q may be different.

In all four cases, if a component U_0 of U is periodic with $f^{\circ p}(U_0) = U_0$, the map $f^{\circ p}$ restricted to U_0 has a unique fixed point, and the orbit of every point in U_0 under $f^{\circ p}$ converges towards this fixed point.

If f represents the "center" point of its hyperbolic component, the orbits of the critical points under f can be described as follows. In the Adjacent Case, the two critical points coincide (in other

words, the discriminant parameter A is zero), and this double critical point belongs to a periodic orbit. In the Bitransitive Case, the two critical points belong to a common periodic orbit; in the Capture Case, just one of them lies on a periodic orbit, while the orbit of the other eventually hits this periodic orbit; and in the Disjoint Case, they lie on disjoint periodic orbits.

Now let's look at hyperbolic components in the real (A, B)-plane. In the Adjacent Case, there are only two real hyperbolic components. They have periods one and two and are centered at the origin and the point $(0, -1)$, respectively. Both are very special, and I will not try to discuss them. In the Capture Case, we are necessarily in a quadrant with $AB > 0$, and we obtain an arch configuration as in Section 2.

In the Bitransitive Case, if the center lies in a quadrant where $AB > 0$, we obtain a swallow configuration, as discussed in Section 2. However, if the center lies in one of the quadrants where $AB < 0$, we get a quite different three-pointed configuration, which I will call a *tricorn* (Figure 20). In this case, the two critical points c and $\bar{c}$

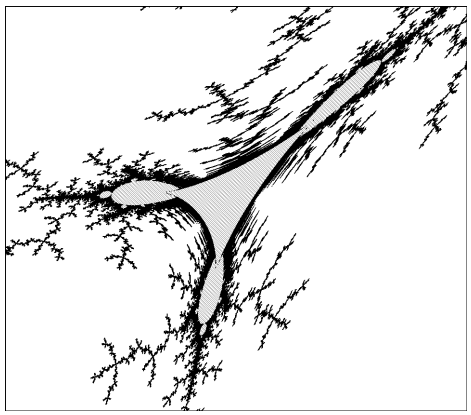

FIGURE 20. Detail of the right center of Figure 19, showing a small tricorn-shaped configuration. For the center point $(.27286, -.93044)$, the third iterate of the cubic map carries each critical point to its complex conjugate. Region: $[.265, .281] \times [-.958, -.903]$.

are conjugate complex, and the first return map from a neighborhood of c to a neighborhood of $\bar{c}$ is conjugate to the first return map in the other direction. Thus we obtain a prototype model for this behavior by replacing these two neighborhoods by two disjoint copies of the complex numbers $\mathbf{C}$, mapping the first to the second by a quadratic map $z \mapsto w = z^2 + c$, and mapping back by the conjugate transformation $w \mapsto z = w^2 + \bar{c}$. The resulting connectedness locus in the c-parameter plane is shown in Figure 21. This configuration

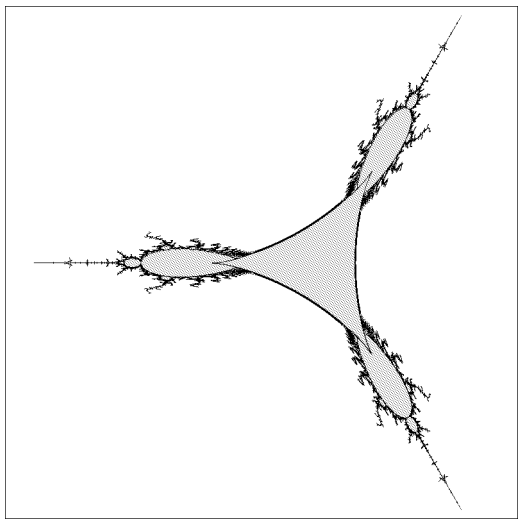

FIGURE 21. The prototype tricorn, in the c-plane where $z \mapsto (z^2 + c)^2 + \bar{c}$.

is compact and connected and has an exact threefold rotational symmetry. Like the real cubic connectedness locus, it contains embedded copies of the Mandelbrot set, where the root point has been stretched out over a curve of saddle node points, so as to yield a nonlocally connected set (Figure 22; cf. [Winters]). Along the real axis, this prototype tricorn coincides precisely with the Mandelbrot set. However, as soon as we get off the real axis, the two differ. In particular, each hyperbolic component

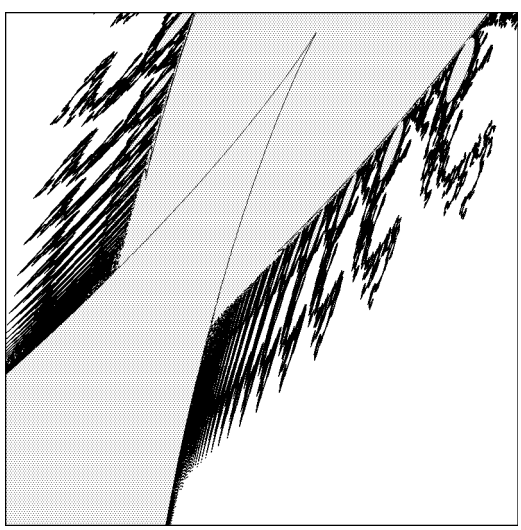

FIGURE 22. Detail of Figure 21, showing lack of local connectivity. Region: $[.18, .5] \times [.34, .66]$.

along the real axis of the Mandelbrot set gives rise either to a small embedded Mandelbrot set in the tricorn or to a small embedded tricorn, depending on whether the period is even or odd.

If the center of a hyperbolic component lies precisely along the A-axis, we obtain a mixed configuration. In the Bitransitive Case, the part which lies in a quadrant with $AB > 0$ looks like half of a swallow configuration, and the other half looks like half of a tricorn. The Disjoint Case is quite similar. If the center satisfies $AB > 0$, we obtain a product configuration, as discussed in Section 2. If it satisfies $AB < 0$, we obtain a copy of the Mandelbrot set, whereas if it lies exactly on the A-axis, we obtain a mixed configuration. Such mixed configurations must be considered as an artifact of our choice of parametrization. They would not appear if we worked in the (A, b)-plane or the (A, ib)-plane, as in Figures 4 and 5. However, such mixed configurations along the A-axis of the (A, B)-plane do help to make it clear that tricorn and swallow (or Mandelbrot set and product configuration) are just different real slices through a common configuration in $\mathbf{C}^2$.

In Figure 18, twenty of the hyperbolic components in the real cubic connectedness locus have been labeled. (Compare Appendix B.) It is noteworthy that several of the most prominent hyperbolic components seem to be missing some of the basic features of their prototypical examples. In fact, this seems to happen whenever the given component is immediately contiguous and subordinate to a larger hyperbolic component. In general, we must ask the following question: *Under what conditions will the configuration around a hyperbolic component in the real or complex cubic connectedness locus include a complete copy of the connectedness locus for its prototype configuration?*

For quadratic polynomials, in [Douady and Hubbard 1985] the authors provide a full answer to the analogous question in their theory of "modulation" or "tuning." In the quadratic case, there is only one kind of hyperbolic component, and they show that every hyperbolic component in the Mandelbrot set is embedded as the central region of a small copy of the full Mandelbrot set.

APPENDIX A. NORMAL FORMS AND CURVES IN PARAMETER SPACE

The *barycenter* of a polynomial map

$$x \mapsto f(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0 \quad \text{(A.1)}$$

of degree $n \geq 2$ is the unique point

$$\hat{x} = -\frac{1}{n}\frac{c_{n-1}}{c_n}$$

at which the $(n-1)$-st derivative vanishes. In the complex case, this can be identified with the average of the $n-1$ critical points $f'(z) = 0$. If $n > 2$, it coincides with the average of the n fixed points $f(z) = z$. Every polynomial map is conjugate by one and only one translation to a map $x \mapsto g(x) = f(x + \hat{x}) - \hat{x}$ that is *centered*, in the sense that its barycenter is zero. This is equivalent to the requirement that the coefficient of x^{n-1} in g (written as a sum of monomials) should be zero.

If γ is a solution to the equation $\gamma^{n-1} = c_n$, the linearly conjugate polynomial $x \mapsto \gamma g(x/\gamma)$ is *monic*, that is, it has leading coefficient 1. In the complex cubic case, note that γ is uniquely determined up to sign. It follows easily that every complex cubic map is affinely conjugate to one of the form

$$z \mapsto z^3 - 3Az + b$$

with critical points $\pm a = \pm\sqrt{A}$, where the numbers A and $B = b^2$ are affine conjugacy invariants. If we start with a polynomial in the more general form (A.1), computation shows that

$$A = -f'(\hat{z})/3 \;=\; (c_2^2 - 3c_1c_3)/9c_3\,, \qquad \text{(A.2)}$$

where $\hat{z} = -\frac{1}{3}c_2/c_3$, and that $b = \pm(f(\hat{z}) - \hat{z})\sqrt{c_3}$ or

$$B = (f(\hat{z}) - \hat{z})^2 c_3\,. \qquad \text{(A.3)}$$

In the real cubic case, note that $\hat{z}$ and the invariants $A = a^2$ and $B = b^2$ are real, although a and or b may be pure imaginary.

The Locus $\mathrm{Per}_1(\mu)$

By definition, a pair (A, B) belongs to $\mathrm{Per}_1(\mu)$ if and only if the corresponding cubic map f has a fixed point at which the derivative f' equals μ. If $f(x) = x^3 - 3Ax + b$, and if the fixed point is $x = \kappa$, we can equally well work with the translation-conjugate polynomial $g(x) = f(x + \kappa) - \kappa$, which

has its preferred fixed point at the origin and hence has the form

$$g(x) = x^3 + 3\kappa x^2 + \mu x\,.$$

Using (A.2) and (A.3), we see that $A = \kappa^2 - \mu/3$ and $b = \kappa(2\kappa^2 + 1 - \mu)$. It is then easy to solve for $B = b^2$ as a function of A. Noteworthy cases are

$$\begin{aligned}
&\mathrm{Per}_1(1): \quad B = 4(A + \tfrac{1}{3})^3\,,\\
&\mathrm{Per}_1(0): \quad B = 4A(A + \tfrac{1}{2})^2\,,\\
&\mathrm{Per}_1(-1): \quad B = 4(A - \tfrac{1}{3})(A + \tfrac{2}{3})^2\,.
\end{aligned}$$

(The first of these appears in Figure 2.) Here the saddle node curve $\mathrm{Per}_1(1)$ forms part of the upper boundary of the principal region, which is labeled $\mathcal{A}_1$ in Figure 18, and the *bifurcation locus* $\mathrm{Per}_1(-1)$, where attracting period-one orbits bifurcate into attracting period-two orbits, forms the lower boundary of this region. Both of these curves also form part of the boundary of regions labeled $\mathcal{C}^+_{(2)1}, \mathcal{D}^+_{1,2}$ and $\mathcal{D}_{1,1}$ in the left-hand part of this figure. The curve $\mathrm{Per}_1(0)$ consists of all parameter pairs with a superattracting fixed point. Thus it passes through the centers of the components labeled $\mathcal{C}^+_{(2)1}, \mathcal{D}^+_{1,2}, \mathcal{D}_{1,1}$ and $\mathcal{A}_1$. The curve

$$\mathrm{Per}_1(2): \quad B = 4(A + \tfrac{2}{3})(A + \tfrac{1}{6})^2$$

is also of interest, but for a surprising reason, which needs some explanation. An arbitrary cubic map has three (not necessarily distinct) complex fixed points $f(z_i) = z_i$. Let $\mu_i = f'(z_i)$ be the corresponding derivatives. Evidently, any symmetric function of the μ_i is an invariant of our cubic map and hence can be expressed as a function of the two fundamental invariants A and B. In fact, it is most convenient to work with the elementary symmetric functions of the $\mu_i - 1$. With a little work, one finds the following explicit formulas.

$$\tfrac{1}{9}\textstyle\sum(\mu_i - 1) = A + \tfrac{1}{3}, \tag{A.4}$$

$$\textstyle\sum_{i<j}(\mu_i - 1)(\mu_j - 1) = 0, \tag{A.5}$$

$$\tfrac{1}{27}\textstyle\prod(\mu_i - 1) = B - 4(A + \tfrac{1}{3})^3. \tag{A.6}$$

If $\mu_1 + \mu_2 \neq 2$, we can solve (A.5) for μ_3, obtaining

$$\mu_3 \;=\; 2 \;+\; \frac{1 - \mu_1\mu_2}{\mu_1 + \mu_2 - 2}\,. \tag{A.7}$$

(In fact, if $\mu_1 \neq \mu_2$, the two curves $\mathrm{Per}_1(\mu_1)$ and $\mathrm{Per}_1(\mu_2)$ intersect transversally at a single point, which also belongs to $\mathrm{Per}_1(\mu_3)$.) If we exclude the indeterminate case $\mu_1 = \mu_2 = 1$, it follows from (A.7) that $\mu_3 = 2$ if and only if $\mu_1\mu_2 = 1$.

Now suppose that a real cubic map has two complex conjugate fixed points that are *indifferent*, in the sense that the corresponding derivatives $\mu_1 = \bar{\mu}_2$ lie on the unit circle. Then $\mu_1\mu_2 = 1$, hence $\mu_3 = 2$, and the associated parameter pair (A, B) lies on the curve $\mathrm{Per}_1(2)$. In fact, if $\mu_1 = e^{i\theta}$, we can compute

$$A = \tfrac{2}{9}(\cos(\theta) - 2)$$

from (A.4). Thus *the curve in the real (A, B)-plane corresponding to cubics with two complex conjugate indifferent fixed points is precisely the segment $-\frac{2}{3} < A < -\frac{2}{9}$ of the curve* $\mathrm{Per}_1(2)$. This curve segment forms the upper boundary of the region $\mathcal{D}_{1,1}$ in Figure 18. Note that the endpoints of this curve segment are just the uniquely defined intersection points $\mathrm{Per}_1(-1) \cap \mathrm{Per}_1(2)$ and $\mathrm{Per}_1(1) \cap \mathrm{Per}_1(2)$.

To study the curve $\mathrm{Per}_2(\mu)$, it is convenient to translate coordinates of our monic polynomial so that the period-two orbit takes the form $\{\kappa, -\kappa\}$, with midpoint at the origin. It is then easy to compute the coefficients, and hence the invariants A and B, as functions of κ^2. In the case $\mu = 1$, there is a substantial simplification. In fact, as $\mu \to 1$, the curve $\mathrm{Per}_2(\mu)$ tends towards a reducible curve, which is the union of two irreducible constituents. One of these is the the bifurcation locus $\mathrm{Per}_1(-1)$, which we do not consider to be part of $\mathrm{Per}_2(1)$, since the period-two orbit has degenerated to a fixed point, and the other is the required curve

$$\mathrm{Per}_2(1): \quad B = 4(A - \tfrac{2}{3})^3\,,$$

where $A = \frac{2}{9}(2\kappa^2 + 1)$. Even on this later curve, note that the period-two orbit degenerates to a period-one orbit at the special point $A = \frac{2}{9}$, $B = -4(\frac{4}{9})^3 = -\frac{256}{729}$, where the two irreducible components come together. (Figure 18 is very distorted around this point. See the discussion in Appendix C.)

Remark. A generic cubic map has three period-two orbits. If μ_1, μ_2, μ_3 are the derivatives around these three orbits, the elementary symmetric func-

tions σ_i of the μ_i can be expressed as polynomial functions of A and B:

$$\sigma_1 = 9(3 - 4A),$$
$$\sigma_2 = 9^2(3 - 8A + 16A^3 - 12A^4 + 2B + 3AB),$$
$$\sigma_3 = \sigma_2 - \sigma_1 + 1 + 9^3(B - 4(A - \tfrac{2}{3})^3)(B - (A - \tfrac{1}{3})(A + \tfrac{2}{3})^2).$$

The critically preperiodic locus $\mathrm{Preper}_{(1)p}$

To study this locus, we must look at maps $f(z) = z^3 - 3a^2z + b$ such that the critical value $f(a)$ belongs to an orbit of period p, but the critical point a does not belong to this orbit. Note that the equation $f(a') = f(a)$ has just one solution $a' \neq a$, namely the *cocritical point* $a' = -2a$. Thus *this periodic orbit must contain both the cocritical point* $-2a$ *and the critical value* $f(a) = b - 2a^3$.

In the case $p = 1$, we must have $-2a = f(a)$, or in other words, $b = 2a^3 - 2a$. Squaring both sides, we obtain the formula

$$\mathrm{Preper}_{(1)1}: \quad B = 4A(A-1)^2\,,$$

as given in Section 2. Note that the derivative $\mu = f'(-2a)$ at the fixed critical value is equal to $9a^2 = 9A$. We can distinguish the segment $|A| < \frac{1}{9}$ of this curve, which lies within the "principal hyperbolic component" $\mathcal{A}_1$, from the segment $A \geq \frac{1}{9}$, which forms much of the upper boundary of the real connectedness locus, and the segment $A \leq -\frac{1}{9}$, which separates the region labeled $\mathcal{C}_{(1)2}$ from $\mathcal{A}_2$.

In the case $p = 2$, the periodic orbit must consist of the two points $f(a)$ and $-2a$. Setting $\xi = b - 2a^3 + 2a$, so that $f(a) = \xi - 2a$, we can write the required equation $f(f(a)) = -2a$ as a cubic equation in ξ with roots $\xi = 0$ and $\xi = 3a \pm \sqrt{-1}$, or in other words, $b = 2a^3 + a \pm i$. If this equation is satisfied, note that the periodic orbit consists of $-2a$ and $f(a) = a \pm i$. By multiplying the equation by $\pm i$ and squaring both sides, we obtain the formula

$$\mathrm{Preper}_{(1)2}: \quad -B = (\sqrt{-A}(2A+1) + 1)^2\,,$$

as given in Section 2.

Points in the (A, B)-parameter plane where two of these curves intersect may be of particular interest. For example, the bifurcation locus $\mathrm{Per}_1(-1)$, which forms the lower part of the boundary of the principal region $\mathcal{A}_1$ in Figure 18, grazes the saddle node curve $\mathrm{Per}_2(1)$ tangentially at the point $A = \frac{2}{9}$, $B = -\frac{256}{729}$, where four different hyperbolic components $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{D}^-_{2,2}$ and $\mathcal{B}_{1+1}$ come together. (In fact, in the complex (A, B)-plane, there are six different hyperbolic components that touch at this point.) The saddle node curve $\mathrm{Per}_1(1)$ grazes the critically preperiodic curve $\mathrm{Preper}_{(1)1}$ tangentially at the point $A = \frac{1}{9}$, $B = \frac{256}{729}$, which lies on the boundary between the regions $\mathcal{R}_1$ and $\mathcal{R}_2$ (cf. Figure 2). Similarly, the curves $\mathrm{Per}_2(1)$ and $\mathrm{Preper}_{(1)2}$ meet tangentially at $A = -\frac{1}{36}$, $B = -\frac{15625}{11664}$ (or $a = \frac{1}{6}i$, $b = \frac{125}{108}i$).

The top boundary of the region $\mathcal{D}_{1,1}$ in Figure 18 forms part of the curve $\mathrm{Per}_1(2)$, characterized by the property that there are two mutually conjugate indifferent fixed points. This curve intersects the saddle node curve $\mathrm{Per}_1(1)$ transversally at the corner point $A = -\frac{2}{9}$, $B = \frac{4}{729}$ of this region. (Presumably, there are two similar transverse intersections of the saddle node curve $\mathrm{Per}_2(1)$ with the lower right boundary curve of the region of the hyperbolic component, which is labeled $\mathcal{D}^-_{2,2}$ in Figure 18, and also a transverse intersection with the tiny $\mathcal{D}_{2,2}$ on the right. One of these intersection points is shown rather poorly in Figure 19.)

The largest value of the invariant B within the real connectedness locus occurs at the point $A = \frac{1}{3}$, $B = \frac{16}{27} = .59259$, and the smallest value occurs at

$$A = -\tfrac{1}{6}, \quad B = -(1 + \sqrt{\tfrac{2}{27}})^2 = -1.6184,$$

both on the boundary between regions $\mathcal{R}_1$ and $\mathcal{R}_2$. The largest and smallest values of A occur at the points $A = \pm 1$, $B = 0$, where both critical points are preperiodic, and where the topological entropy takes its largest value of $\log 3$.

APPENDIX B. CENTERS OF SOME HYPERBOLIC COMPONENTS

Table 1 lists the centers of twenty of the hyperbolic components in the real (A, B)-parameter plane, as shown in Figure 18, including all those for which both critical points have period two or less. They are listed in order of increasing A, first for the upper half-plane $B > 0$, then for $B = 0$, and then for the lower half-plane $B < 0$. The notations $\mathcal{A}$–$\mathcal{D}$ in the third column are explained in Section 4 or in Section 2 (Figure 6). Thus we write:

$\mathcal{A}_p$ for a component with *adjacent critical points* and attracting orbit of period p. These are ex-

A	B	**Type**	**Description**	**Topological Entropy**
$-.55881$	.08656	$\mathcal{D}^-_{3,3}$	$c \mapsto 3c\,,\ \ \bar{c} \mapsto 3\bar{c}$	
.47567	.33217	$\mathcal{C}^+_{(2)2}$	$c \mapsto 2c' \mapsto 2c'$	0
.49408	.45878	$\mathcal{C}^+_{(3)2}$	$c \mapsto 3c' \mapsto 2c'$	0
.62827	.04135	$\mathcal{B}^+_{1+3}$	$c \mapsto c' \mapsto 3c$	0
.71327	.12977	$\mathcal{B}^+_{1+2}$	$c \mapsto c' \mapsto 2c$	$\log((1+\sqrt{5})/2)$
$-\sqrt{2}/2$	0	$\mathcal{D}_{2,2}$	$c \mapsto 2c\,,\ \ c' \mapsto 2c'$	0
$-1/2$	0	$\mathcal{D}_{1,1}$	$c \mapsto c\,,\ \ c' \mapsto c'$	0
0	0	$\mathcal{A}_1$	$c = c' \mapsto c$	0
$1/2$	0	$\mathcal{B}_{1+1}$	$c \mapsto c' \mapsto c$	0
$\sqrt{2}/2$	0	$\mathcal{D}_{2,2}$	$c \mapsto 2c\,,\ \ c' \mapsto 2c'$	0
$-3/4$	$-3/16$	$\mathcal{C}^+_{(2)1}$	$c \mapsto 2c' \mapsto c'$	$\log 2$
$-.61688$	$-.03371$	$\mathcal{D}^+_{1,2}$	$c \mapsto c\,,\ \ c' \mapsto 2c'$	0
$-.55310$	$-.62882$	$\mathcal{B}^+_{2+2}$	$c \mapsto 2c' \mapsto 2c$	$\log 1.83929$
$-.39736$	$-.31371$	$\mathcal{C}^+_{(2)2}$	$c \mapsto 2c' \mapsto 2c'$	0
$-.36464$	-1.09040	$\mathcal{B}^+_{1+2}$	$c \mapsto c' \mapsto 2c$	$\log((1+\sqrt{5})/2)$
$-1/4$	$-9/16$	$\mathcal{C}^+_{(1)2}$	$c \mapsto c' \mapsto 2c'$	0
$-.13414$	-1.37344	$\mathcal{B}^+_{2+2}$	$c \mapsto 2c' \mapsto 2c$	0
0	-1	$\mathcal{A}_2$	$c = c' \mapsto 2c$	0
$1/4$	$-7/16$	$\mathcal{D}^-_{2,2}$	$c \mapsto 2c\,,\ \ \bar{c} \mapsto 2\bar{c}$	
.27286	$-.93044$	$\mathcal{B}^-_{3+3}$	$c \mapsto 3\bar{c} \mapsto 3c$	

TABLE 1. Centers of hyperbolic components (see Appendix B for notation).

actly the hyperbolic components whose center point lies on the line $A = 0$, where the critical points coincide.

$\mathcal{B}_{p+q}$ for a *bitransitive* component with attracting orbit of period $p+q$, where the p-th iterate carries the first critical point to the immediate basin of the second, and the q-th iterate carries the second back to the immediate basin of the first. Such a component lies at the center of a *swallow-shaped* configuration in the real (A, B)-plane when $AB > 0$, or a *tricorn* configuration when $AB < 0$.

$\mathcal{C}_{(t)p+q}$ for a *capture* component (or *arch-shaped* configuration), at whose center point the t-fold iterate carries one critical point to an orbit of period $p+q$ containing the other critical point, and where the $(t+p)$-th image of the first critical point is equal to the second. In the special case $p = 0$, we write this briefly as $\mathcal{C}_{(t)q}$. Such a component necessarily lies in one of the quadrants $AB > 0$, where the critical points are real and distinct.

$\mathcal{D}_{p,q}$ for a component with two *disjoint attracting orbits* with periods p and q, yielding a *product* configuration when $AB > 0$, or a *Mandelbrot* configuration when $AB < 0$.

Moreover, the superscript $^+$ has been added if $AB > 0$ (so that the critical points are real and distinct), and the superscript $^-$ has been added if $AB < 0$ (so that the critical points are complex conjugate and distinct). In the fourth column, the notation $\{c, c'\}$ is used for the set of critical points, and $\overset{n}{\mapsto}$ for the n-th iterate of the cubic map. For example, $c \overset{3}{\mapsto} c'$ means that the third iterate carries the first critical point to the second. The last column gives the topological entropy of the real mapping, when $AB \geq 0$.

Remark. When $B = 0$, there is a hyperbolic component centered at $(A, 0)$ if and only if there is one centered at $(-A, 0)$. This follows from the observation that for an odd mapping, such as $f(z) = \pm z^3 - 3Az$, the second iterate $f(f(z))$ is equal to $-f(-f(z))$, so that f and $-f$ have quite similar dynamical properties. For example, they have the

same topological entropy in the real case, or the same Julia set in the complex case, and one is hyperbolic if and only if the other is. However, it may happen that one of these two belongs to a bitransitive component (type $\mathcal{B}$), whereas the other has disjoint attracting orbits (type $\mathcal{D}$).

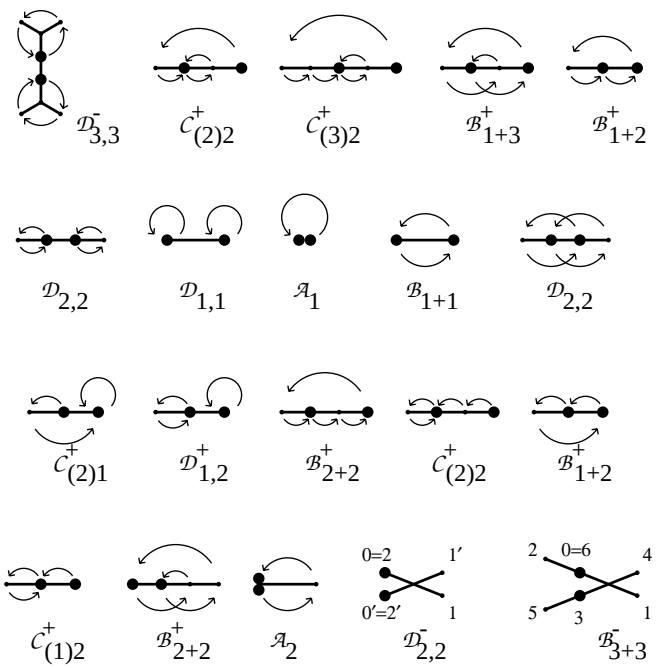

FIGURE 23. Hubbard trees for the centers of 20 hyperbolic components, as listed in Appendix B. (In the last two diagrams, vertex 0 maps to 1, to 2, and so on.)

More information about these twenty hyperbolic components can be read from Figure 23, which shows the corresponding *Hubbard trees* [Douady and Hubbard 1984; 1985]. Each Hubbard tree is a simplified picture that shows how the points of the two critical orbits are joined to each other within the filled Julia set $K(f)$. Since our polynomials have real coefficients, all of these Hubbard trees are symmetric about the real axis. Note that only those on the second line that correspond to odd mappings with $B = 0$ are also symmetric about a vertical axis.

APPENDIX C. COMMENTS ON THE FIGURES

The basic algorithm used in making pictures in the (A, B)-plane and in other related parameter planes can be described as follows. Starting with the two critical points $z_0^{\pm} = \pm a$, which may be either real or conjugate complex, we compute the successive iterates $z_{n+1}^{\pm} = f(z_n^{\pm})$ and also the partial derivatives of $z_n^{\pm}$ with respect to A and B for n up to a few hundred, or until either $|z_n^{\pm}|$ becomes large or one of the partial derivatives becomes very large. The given point in parameter space is considered to be in a hyperbolic component if all of these numbers remain relatively small. If $|z_n^{\pm}|$ becomes large, the distance of the given point in parameter space from the locus where the orbit of the given critical point remains bounded is estimated, using the first partial derivatives [Milnor 1989, §5.6; Fisher 1988]. If this distance is less than the pixel size, the given parameter point is considered to be a boundary point. This method enables the pictures to show very fine filaments, which may have measure close to zero. Similarly, if the orbit remains bounded but some first derivative becomes large, we have a boundary point. In many of the figures, an additional step has been taken to locate boundaries between hyperbolic components. Namely, after many iterations, the orbits are checked for approximate periodicity with small period; and if both critical orbits have the same period, these two periodic orbits are compared. Pixels at which this periodicity structure changes are indicated in black.

The main defect of this procedure is that it is ineffective when the convergence to a periodic orbit is extremely slow. This tends to happen near the curves $\mathrm{Per}_p(\pm 1)$ where there is a parabolic orbit, and particularly near points where three or more hyperbolic components come together. Hence the figures are highly distorted near such points (see Appendix A).

REFERENCES

The references marked with a bullet are not cited in the text, but are included for completeness.

[Benedicks and Carleson 1985] M. Benedicks and L. Carleson, "On the iterations of $1 - ax^2$ on $(-1, 1)$", *Ann. of Math.* **122** (1985), 1–25.

[Blanchard 1984] P. Blanchard, "Complex analytic dynamics on the Riemann sphere", *Bull. Amer. Math. Soc.* **11** (1984), 85–141. •

[Blanchard 1986] P. Blanchard, "Disconnected Julia sets", in *Chaotic Dynamics and Fractals* (edited by M. Barnsley and S. Demko), Academic Press, Orlando, 1986. •

[Blanchard et al. 1991] P. Blanchard, R. Devaney and L. Keen, "The dynamics of complex polynomials

and automorphisms of the shift", *Invent. Math.* **104** (1991), 545–580.

[Block and Keesling] L. Block and J. Keesling, "Computing the topological entropy of maps of the interval with three monotone pieces" (to appear).

[Branner 1985] B. Branner, "Iterations by odd functions with two extrema", *J. Math. Anal. and Appl.* **105** (1985), 276–297. •

[Branner 1986] B. Branner, "The parameter space for cubic polynomials", pp. 169–179, in *Chaotic Dynamics and Fractals* (edited by M. Barnsley and S. Demko), Academic Press, Orlando, 1986. •

[Branner 1989] B. Branner, "The Mandelbrot set", pp. 75–105 in *Chaos and Fractals* (edited by R. Devaney and L. Keen), Proc. Symp. Applied Math. **39**, American Mathematical Society, Providence, RI, 1989. •

[Branner and Douady 1988] B. Branner and A. Douady, "Surgery on complex polynomials", pp. 11–72, in *Holomorphic Dynamics: Proceedings Symposium on Dynamical Systems (Mexico 1986)* (edited by X. Gomez-Mont et al.), Lecture Notes in Math. **1345**, Springer-Verlag, Berlin, 1988. •

[Branner and Hubbard 1988] B. Branner and J. H. Hubbard, "The iteration of cubic polynomials, Part 1: The global topology of parameter space", *Acta Math.* **160** (1988), 143–206. (Part 2 to appear.)

[Bredon 1972] G. Bredon, *Introduction to Compact Transformation Groups*, Academic Press, New York, 1972.

[Brown 1960] M. Brown, "A proof of the generalized Schoenflies theorem", *Bull. Amer. Math. Soc.* **66** (1960), 74–76.

[Brown 1961] M. Brown, "The monotone union of open n-cells is an open n-cell", *Proc. Amer. Math. Soc.* **12** (1961), 812–814.

[Chavoya-Aceves et al. 1985] O. Chavoya-Aceves, F. Angulo-Brown and E. Piña, "Symbolic dynamics of the cubic map", *Physica* D**14** (1985), 374–386.

[Devaney et al.] R. Devaney, L. Goldberg and J. H. Hubbard, "A dynamical approximation to the exponential map by polynomials" (to appear). •

[Dewdney 1987] A. K. Dewdney, "Computer Recreations", *Scient. Amer.*, Nov. 1987, p. 144.

[Douady 1983] A. Douady, "Systèmes dynamiques holomorphes", Sém. Bourbaki 599, *Astérisque* **105–106** (1983), 39–63. •

[Douady and Hubbard 1982] A. Douady and J. H. Hubbard, "Itérations des polynômes quadratiques complexes", *Comptes R. Acad. Sci. Paris* **294** (1982), 123–126.

[Douady and Hubbard 1984] A. Douady and J. H. Hubbard, "Etude dynamique des polynômes complexes I", *Publ. Math. Orsay* (1984), II (1985).

[Douady and Hubbard 1985] A. Douady and J. H. Hubbard, "On the dynamics of polynomial-like mappings", *Ann. Sci. Ec. Norm. Sup. (Paris)* **18** (1985), 287–343.

[El Hamouly and Mira 1982] H. El Hamouly and C. Mira, "Singularités dues au feuilletage du plan des bifurcations d'un difféomorphisme bi-dimensionel", *C. R. Acad. Sci. Paris* **294** (1982), 387–390.

[Fisher 1988] Y. Fisher, "Exploring the Mandelbrot set", pp. 287–296, in *The Science of Fractal Images* (edited by H.-O. Peitgen and D. Saupe), Springer-Verlag, New York, 1988.

[Friedland and Milnor 1989] S. Friedland and J. Milnor, "Dynamical properties of plane polynomial automorphisms", *Ergodic Theory and Dynamical Systems* **9** (1989), 67–99. •

[Friedman and Tresser 1986] B. Friedman and C. Tresser, "Comb structure in hairy boundaries: Some transition problems for circle maps", *Phys. Lett.* A**117** (1986), 15–22. •

[Guckenheimer 1979] J. Guckenheimer, "Sensitive dependence to initial conditions for one dimensional maps", *Commun. Math. Phys.* **70** (1979), 133–160.

[Jakobson 1981] M. Jakobson, "Absolutely continuous invariant measures for one-parameter families of one-dimensional maps", *Commun. Math. Phys.* **81** (1981), 39–88.

[Lavaurs 1989] P. Lavaurs, "Systèmes dynamiques holomorphes: Explosion de points périodiques paraboliques", Thesis, Univ. Paris-Sud, Orsay, 1989. •

[MacKay and Tresser] R. MacKay and C. Tresser, "Boundary of chaos for bimodal maps of the interval", *J. London Math. Soc.* (to appear).

[MacKay and Tresser 1987] R. MacKay and C. Tresser, "Some flesh on the skeleton: The bifurcation structure of bimodal maps", *Physica* D**27** (1987), 412–422.

[McMullen 1988] C. McMullen, "Automorphisms of rational maps", pp. 31–60, in *Holomorphic Functions and Moduli I* (edited by D. Drasin et al.), Springer-Verlag, New York, 1988.

[Milnor 1988] J. Milnor, "Nonexpansive Hénon maps", *Advances in Math.* **69** (1988), 109–114. •

[Milnor 1989] J. Milnor, "Self-similarity and hairiness in the Mandelbrot set", pp. 211–257, in *Computers in Geometry and Topology* (edited by M. Tangora), Lecture Notes Pure Appl. Math. **114**, Dekker, New York, 1989.

[Milnor 1990] J. Milnor, *Dynamics in One Complex Variable*, Introductory Lectures, Stony Brook IMS preprints 1990/5. •

[Milnor 1992] J. Milnor, "Hyperbolic components in spaces of polynomial maps", Stoney Brook IMS prepints, 1992/3.

[Milnor] J. Milnor, "On cubic polynomials with periodic critical point" (in preparation). •

[Milnor and Thurston 1988] J. Milnor and W. Thurston, "Iterated maps of the interval", in *Dynamical Systems (Maryland 1986–87)* (edited by J. Alexander), Lecture Notes in Math. **1342**, Springer-Verlag, Berlin, 1988.

[Misiurewicz and Szlenk 1977] M. Misiurewicz and W. Szlenk, "Entropy of piecewise monotone mappings", *Astérisque* **50** (1977), 299–310. Revised version in *Studia Math.* **67** (1980), 45–63.

[Piña 1984] E. Piña, "Comment on 'Study of a one-dimensional map with multiple basins' ", *Phys. Rev.* A**30** (1984), 2132–2133. •

[Rees 1990] M. Rees, "Components of degree two hyperbolic rational maps", *Invent. Math.* **100** (1990), 357–382. •

[Rees 1992] M. Rees, "A partial description of parameter space of rational maps of degree two", Part 1: *Acta Math.* **168** (1992), 11–87; Part 2: Stony Brook IMS preprints 1991/4. •

[Ringland and Schell] J. Ringland and M. Schell, "Genealogy and bifurcation skeleton for cycles of the iterated two-extremum map of the interval" (to appear). •

[Ringland and Schell 1989] J. Ringland and M. Schell, "The Farey tree embodied—in bimodal maps of the interval", *Phys. Lett.* A**136** (1989), 379–386. •

[Ringland and Schell 1990] J. Ringland and M. Schell, "Universal geometry in the parameter plane of maps of the interval", *Europhys. Lett.* **12** (1990), 595–601. •

[Ringland et al. 1990] J. Ringland, N. Issa and M. Schell, "From U sequence to Farey sequence: A unification of one-parameter scenarios", *Phys. Rev.* A**41** (1990), 4223–4235. •

[Rogers and Whitley 1983] T. Rogers and D. Whitley, "Chaos in the cubic mapping", *Math. Modelling* **4** (1983), 9–25. •

[Rothschild 1971] J. Rothschild, "On the computation of topological entropy", Thesis, CUNY, New York, 1971.

[Rychlik 1988] M. Rychlik, "Another proof of Jakobson's theorem and related results", *Erg. Theory and Dyn. Sys.* **8** (1988), 93–109.

[Skjolding et al. 1983] H. Skjolding, B. Branner, P. Christiansen and H. Jensen, "Bifurcations in discrete dynamical systems with cubic maps", *Siam J. Appl. Math.* **43** (1983), 520–534. •

[Winters] R. Winters, "Of tricorns and biquadratics" (in preparation, Boston University).

[Wittner 1988] B. Wittner, "On the bifurcation loci of rational maps of degree two", Thesis, Cornell University, Ithaca, 1988.

John Milnor, Institute for Mathematical Sciences, SUNY, Stony Brook, NY 11794 (jack@math.sunysb.edu)

Received May 2, 1991; revised October 8, 1991