

REMARKS ON ITERATED CUBIC MAPS

John Milnor

This note will discuss 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, 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

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

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

$$A = a^2, \quad B = b^2 \quad (1.2)$$

form a complete set of coordinates for the *moduli space*, consisting 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; while 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, then we obtain a complete set of invariants (A, B) which turn out to be real. However, if we allow only a real change of coordinates, then there is one additional invariant, namely the sign

$$\sigma = \operatorname{sgn}(g''') \quad (1.3)$$

of the leading coefficient. It is not difficult to check that σ coincides with the sign $\operatorname{sgn}(B)$ whenever $B \neq 0$. However, this additional invariant σ is essential when $B = 0$, for in this case there are two essentially different real polynomial maps

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

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

$$x \mapsto \sigma x^3 - 3Ax + \sqrt{|B|}. \quad (1.4)$$

(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, $\sigma A < 0$, it has complex conjugate critical points. Further details may be found in Appendix A.

2. Real cubic maps as real dynamical systems.

Let us try to describe the behavior of the iterates of a cubic map $f : \mathbf{R} \rightarrow \mathbf{R}$, considered as a real dynamical system. 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 $\{x, f(x), f(f(x)), \dots\}$ is bounded. This set $K_{\mathbf{R}}$ can be described as the real part of the “filled Julia set” of f . (Compare §3.)

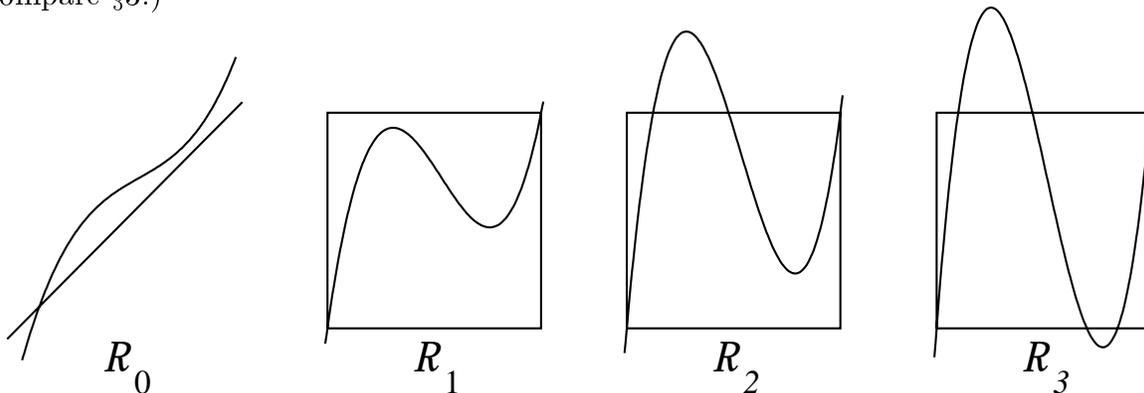


Figure 1. Representative graphs for the four different classes of real cubic polynomials. The case $f''' > 0$ is illustrated; but corresponding examples with $f''' < 0$ can be obtained by looking at this figure in a mirror.

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 $\mathcal{R}_0, \mathcal{R}_1, \dots, \mathcal{R}_d$, 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, then there must be at least two distinct points in $K_{\mathbf{R}}(f)$. Let I be the *smallest closed interval which contains* $K_{\mathbf{R}}(f)$. Thus every orbit which starts outside of I must escape to infinity, but the two end points 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 which map into I (some of these intervals may be degenerate when $d > 3$), together with $n - 1$ separating open intervals which 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

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

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 non-trivial 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 $f \in \mathcal{R}_d$, then 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, §§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 §2.4.) 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).

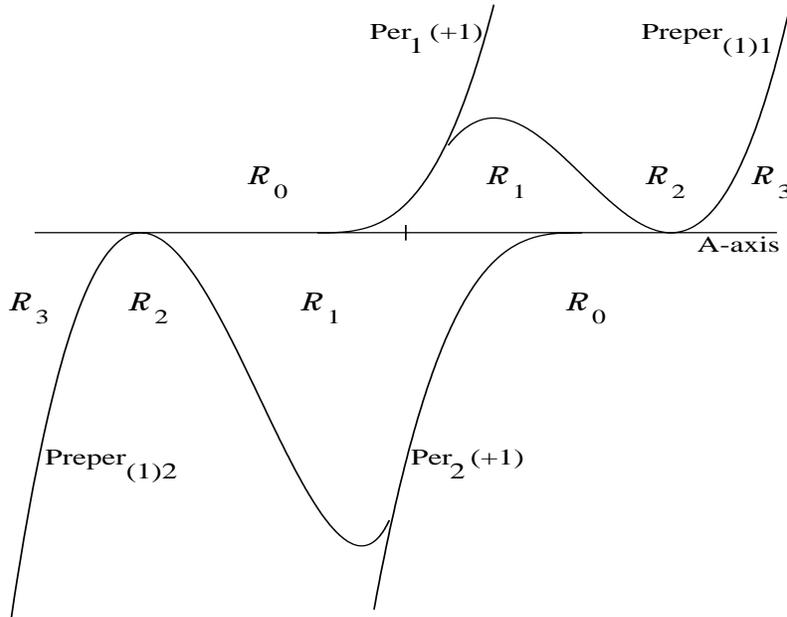


Figure 2. The four regions in the (A, B) -parameter plane, and the curves which separate them.

Definition. Let $\text{Per}_p(\mu)$ consist of all parameter pairs (A, B) for which the associated cubic map f has a periodic orbit of period p with multiplier $(f^{op})'$ equal to μ . In particular, the curve

$$\text{Per}_1(+1) : \quad B = 4\left(A + \frac{1}{3}\right)^3$$

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

$$\text{Per}_2(+1) : \quad B = 4\left(A - \frac{2}{3}\right)^3$$

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 1 or 2 respectively.

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

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

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

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

in the quadrant $A, B \leq 0$, gives maps such that one critical point maps to an orbit of period 2. 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 at least one of these curves. In fact we need only the curves $\text{Per}_1(1)$ and $\text{Preper}_{(1)1}$ in the half-plane $\sigma = 1$, $B \geq 0$, as can be verified by study of Figure 1. Similarly, we need only the curves $\text{Per}_2(1)$ and $\text{Preper}_{(1)2}$ 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 $\text{Per}_p^d(\mu)$ and $\text{Preper}_{(t)p}^d$ which 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 $\text{Per}_1^d(+1)$ and $\text{Preper}_{(2)1}^d$ in all cases, and also $\text{Preper}_{(1)1}^d$ 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 grey). Similarly in Case \mathcal{R}_1 the two critical orbits may both behave chaotically (dark grey), 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.

Remark. For many purposes it is more natural to work in the (A, b) -parameter plane, where $b = \pm\sqrt{B}$. 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 magnified portions of Figure 3, 4 or 5 shows that several characteristic patterns are repeated many times on different scales. Noteworthy are “swallow” shaped

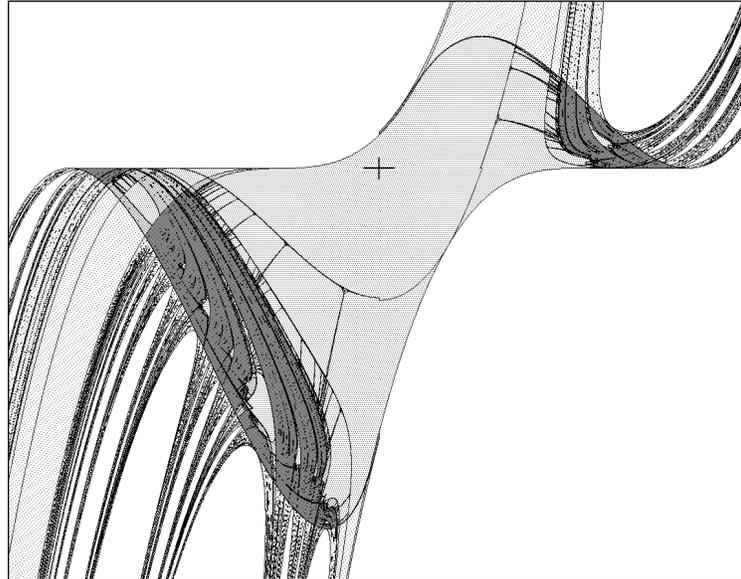


Figure 3. Picture of the (A, B) -parameter plane. The boundaries between qualitatively different kinds of dynamic behavior have been indicated. In the dark region, both critical orbits behave chaotically, while white indicates that both critical orbits escape to infinity, and intermediate shades indicate various intermediate forms of behavior. (The illustrated region is the rectangle $[-1.2, 1.2] \times [-1.85, .75]$.)

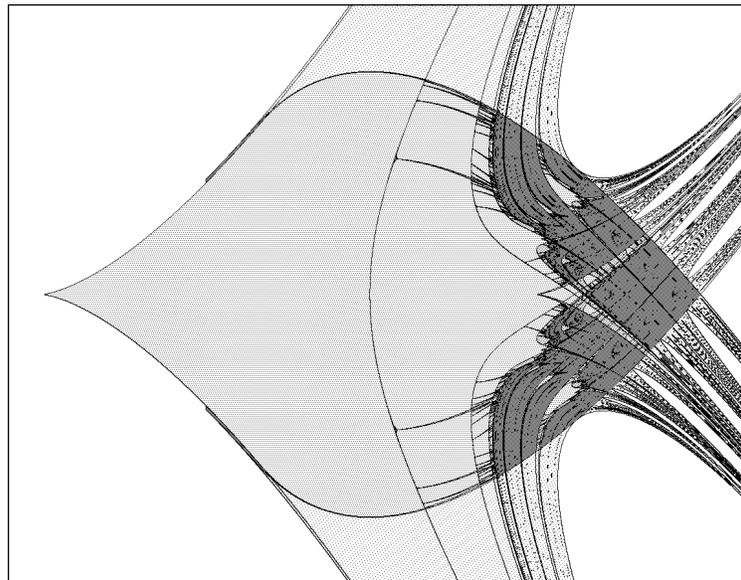


Figure 4. Corresponding picture in the (A, \sqrt{B}) -plane. (Region: $[-.4, 1.1] \times [-1, 1]$.)

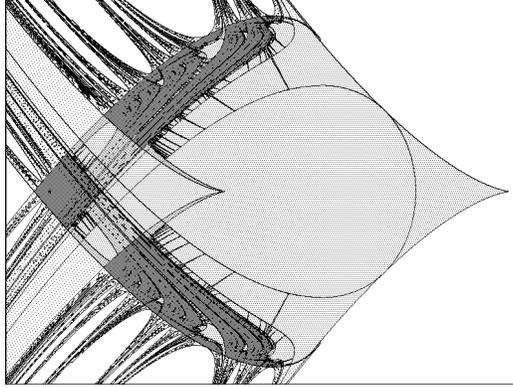


Figure 5. Corresponding picture in the $(A, \sqrt{-B})$ -plane. (Region: $[-1.1, .7] \times [-1.4, 1.4]$.)

regions (Figure 7), “arch” shaped regions (Figure 11), and “product”-like regions (Figure 13). We can partially explain these regions in terms of the dynamics of the associated maps f as follows.

Definition. A smooth map $f : \mathbf{R} \rightarrow \mathbf{R}$ 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 $U \cup f(U) \cup f^{\circ 2}(U) \cup \dots$ has at least two distinct components.

(Here 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.)

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.)

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

Case 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.

Case C. Capture. There are neighborhoods U_1 and U_2 as above, 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 . (Compare [Wittner].)

Case 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 $f^{\circ q}(U_2) \subset U_2$.

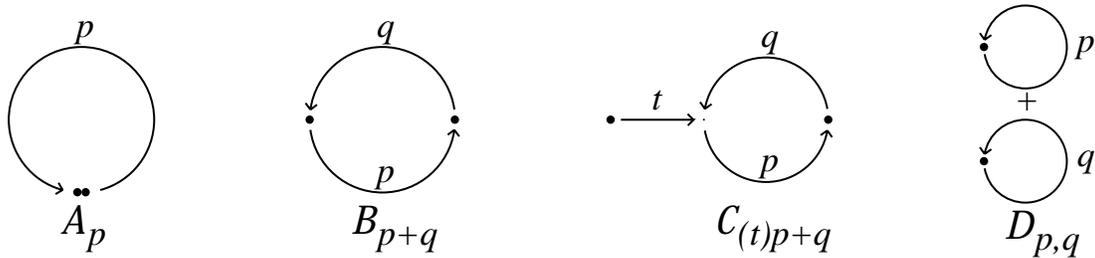


Figure 6. Schematic diagrams for maps representing the centers of the four distinct classes of hyperbolic components. Each critical point is indicated by a heavy dot, and each arrow is labeled by a corresponding number of iterations. (Compare §4 and Appendix B.)

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 §4.) Thus this center map f_0 has the Thurston property of being *post critically 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. (Compare [Douady-Hubbard, 1984] as well as the analogous discussion for quadratic maps in [Milnor-Thurston, §13.4].) Furthermore, any ordering which can occur for a post critically finite continuous map with two critical points 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

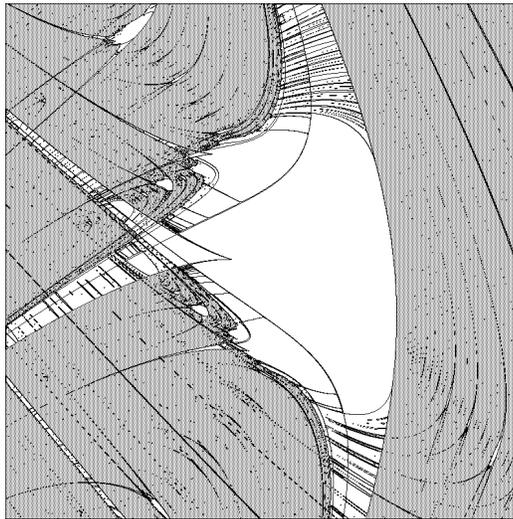


Figure 7. Detail of Figure 3 showing a “swallow configuration” centered at the point $(A, B) = (-.5531, -.6288)$. For the cubic map associated with this central point, the two critical points $\pm a$ satisfy $f(f(a)) = -a$, $f(f(-a)) = a$. Hence both lie on a common orbit of period 4. (Region: $[-.6,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

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*. (Compare Figures 7, 11, 13.) There are two qualifications: If this configuration is immediately adjacent and subordinated to another larger configuration, then it will 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, 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 which 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 $\hat{f} : U_1 \cup U_2 \rightarrow U_1 \cup U_2$, which has one critical point in each component, by a componentwise quadratic map $(k, x) \mapsto (\phi(k), x^2 + c_k)$ from the disjoint union $\{1, 2\} \times \mathbf{R} \approx \mathbf{R} \sqcup \mathbf{R}$ to itself.

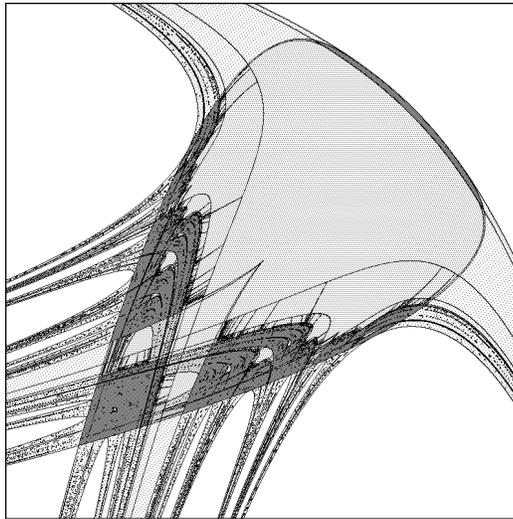


Figure 8. The prototype swallow configuration in the (c_1, c_2) -parameter plane, associated with the family of biquadratic maps $x \mapsto (x^2 + c_1)^2 + c_2$ from the real line to itself. (Region: $[-2.5, 1] \times [-2.5, 1]$.)

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

$$x \mapsto y = x^2 + c_1, \quad y \mapsto x = y^2 + c_2, \quad (2.1)$$

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. (Compare Ringland and Schell.) The central tear drop 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. (Compare

§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 which 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 quite similar swallow shaped configurations. (Compare [El Hamouly and Mira].) For example such a region is shown in Figure 9, corresponding to an attracting orbit of period 5. This phenomenon can be explained intuitively as follows. If $|\beta|$ is small, then 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 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, then it must contain at least one point which is close to the axis $y = 0$. If exactly two points of the orbit are close to $y = 0$, then 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.



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 $(x, y) \mapsto (y, y^2 - \alpha - \beta x)$; and grey 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: $[1.4, 1.6] \times [-.3, -.1]$.)

Caution. The swallow configuration of Figures 7, 8, 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

$$t \mapsto t + c + k \sin(2\pi t) \pmod{1}. \quad (2.3)$$

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. (Compare [Chavoya-Aceves et al.].) Each of these configurations terminate in a “tongue” which reaches down to the corresponding rational point on the $k = 0$ axis. These are known as *Arnold tongues*.

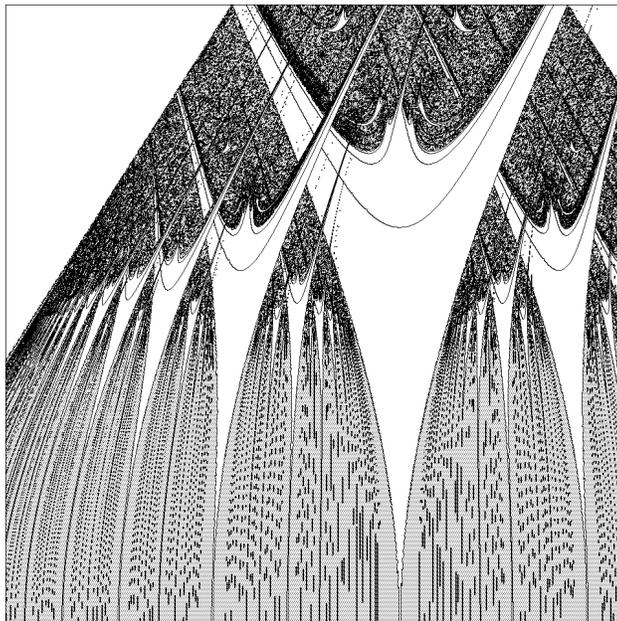


Figure 10. Arnold tongues ending in “pointed-swallow” configurations for the family of circle maps $t \mapsto t + c + k \sin(2\pi t)$. (Region: $[\cdot 15, \cdot 7] \times [0, \cdot 35]$ in the (c, k) -parameter plane.)

Next let us 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 above 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 \mathbf{R} to

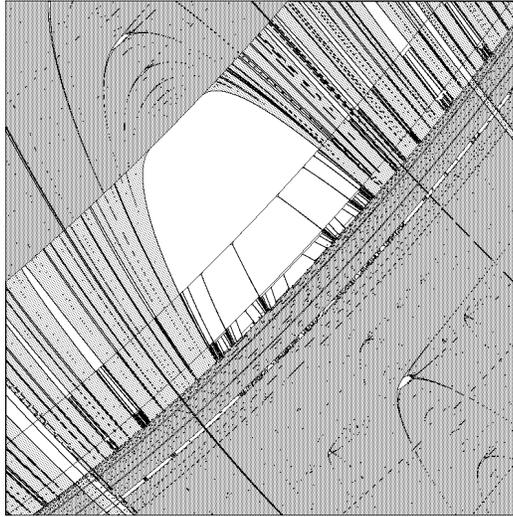


Figure 11. Detail of Figure 3 showing an arch configuration. For the cubic map corresponding to the center point $(A, B) = (.8536, .0243)$, the two critical points $\pm a$ satisfy $f(f(a)) = f(f(-a)) = \pm a$. (Region: $[.835, .885] \times [.01, .03]$.)

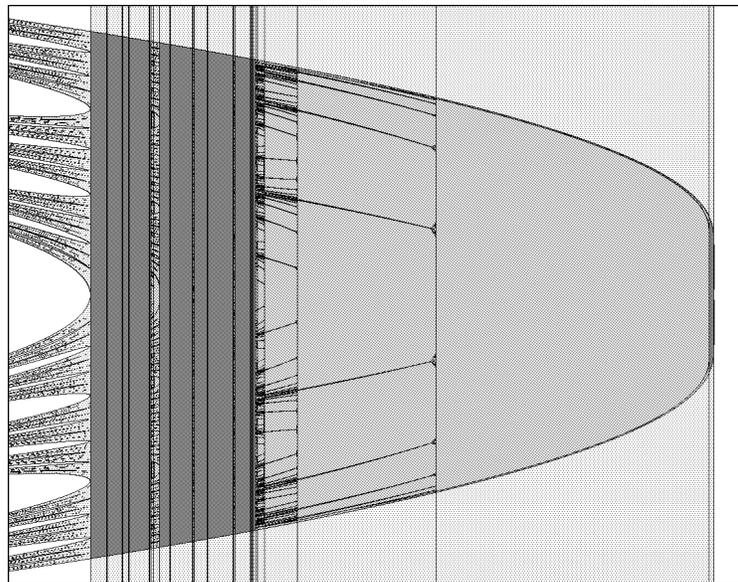


Figure 12. The prototype arch configuration in the (c, \hat{x}) -plane. Here we consider the orbit of the point \hat{x} under the map $x \mapsto x^2 + c$. Dark grey indicates that both \hat{x} and 0 have chaotic orbits, while white means that both escape to infinity. (Region: $[-2.3, .4] \times [-2.2,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

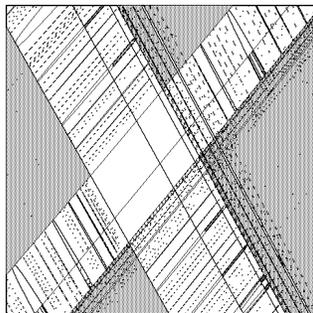


Figure 13. Detail of Figure 3 showing a product configuration. For the map corresponding to the center point $(.8156, .0674)$, there are two superattracting periodic orbits with periods 3 and 4 respectively. (Region: $[.814, .819] \times [.0665, .0685]$.)

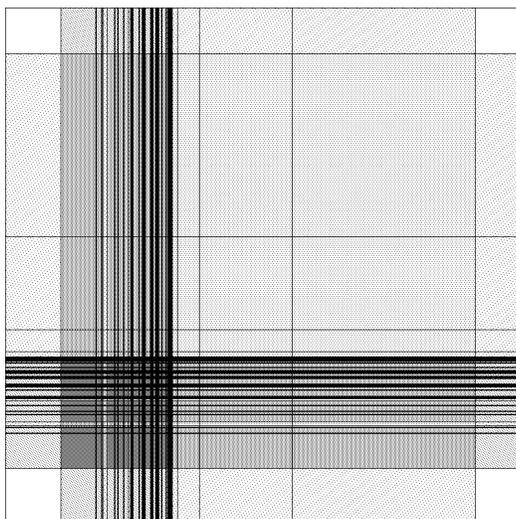


Figure 14. The prototype product configuration in the (c_1, c_2) -parameter plane.

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

Finally consider 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 above 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 respectively, and 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 $[-2, 1/4] \times [-2, 1/4]$ in the (c_1, c_2) -plane, as illustrated in Figure 14.

According to Jakobson, the set set of parameter pairs for which both critical orbits are chaotic (indicated by dark grey in the figure) has positive measure. See also [Benedicks and Carleson], [Rychlik]. 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.

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

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

$$h(f) = \lim_{k \rightarrow \infty} \frac{1}{k} \log(\ell(f^{\circ k})) \quad (2.4)$$

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. (Compare [Thurston and Milnor].) In the cubic case, a more practical algorithm for computing h has recently been described by [Block and Keesling].

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. (Compare §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 $1 \leq s \leq 3$ in the cubic case. Figures 15a, b show the “*curves*” $s = \text{constant}$ in the (A, b) -plane, corresponding to the family of maps $x \mapsto x^3 - 3Ax + b$, and in the (A, b') -plane for $x \mapsto -x^3 - 3Ax + b'$. (Compare Figures 4, 5.) Both figures show both points inside the real connectedness locus and points outside of it. At least part of the boundary curve $\text{Preper}_{(1)1}$ (respectively $\text{Preper}_{(2)1}$) for the connectedness locus is clearly visible in these figures as a locus where the curves $s = \text{constant}$ 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. (Compare [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-Thurston].) A corresponding conjecture for the cubic case would be that: *each level set $s(A, b) = \text{constant}$ is a connected subset of the (A, b) -parameter plane*, and in particular that the continuous function $(A, b) \mapsto s(A, b)$ does not have any isolated local maxima or minima. There is of course a completely analogous conjecture for the (A, b') -plane.

Note that these sets $\{(A, b) : s = \text{constant}\}$ are not always curves. They may well have interior points. For example this is the case for $s = 1, 2, 3$ and also for $s = (1 + \sqrt{5})/2$.

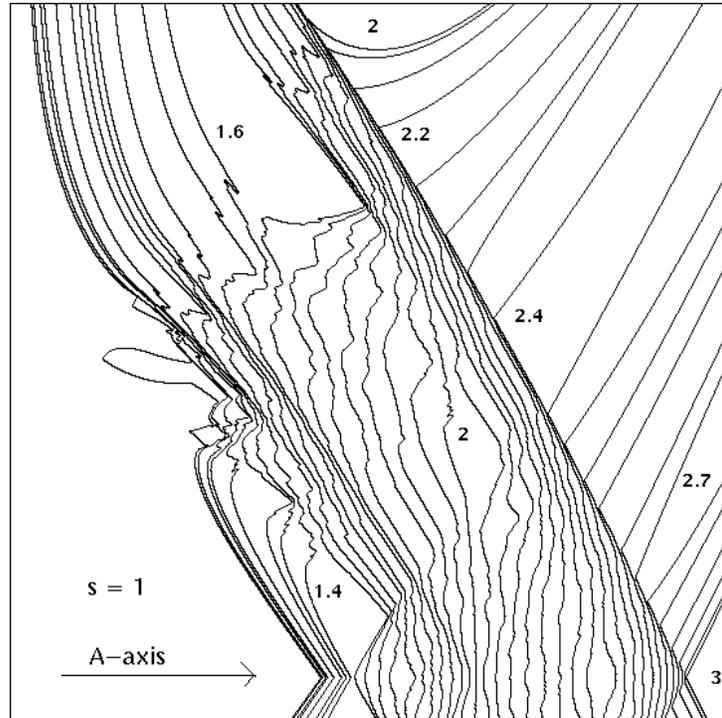


Figure 15a. Curves of constant growth number s in the (A, b) -parameter plane of Figure 4. Here $h = \log s$ is the topological entropy of the map $x \mapsto x^3 - 3Ax + b$. The curves $s = .05, .1, \dots, 2^-, 2.05, \dots, 3^-$ are plotted, using an algorithm due to Block and Keesling. (Illustrated region: $[.57, 1.03] \times [-.03, .43]$, contour interval: $\Delta s = .05$.)

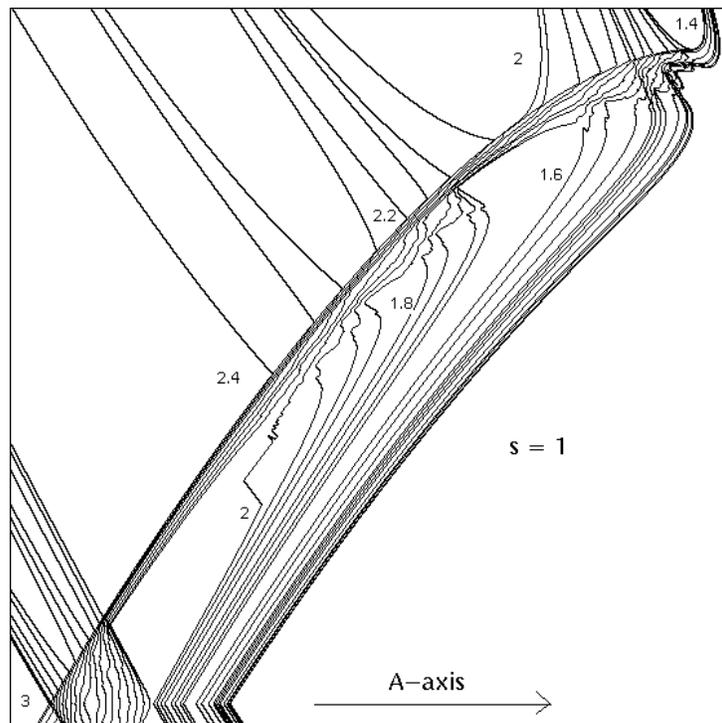


Figure 15b. Corresponding picture for the family of maps $x \mapsto -x^3 - 3Ax + b'$. (Compare Figure 5.) Region: $[-1.05, -.02] \times [-.05, 1.35]$, contour interval: $\Delta s = .05$.

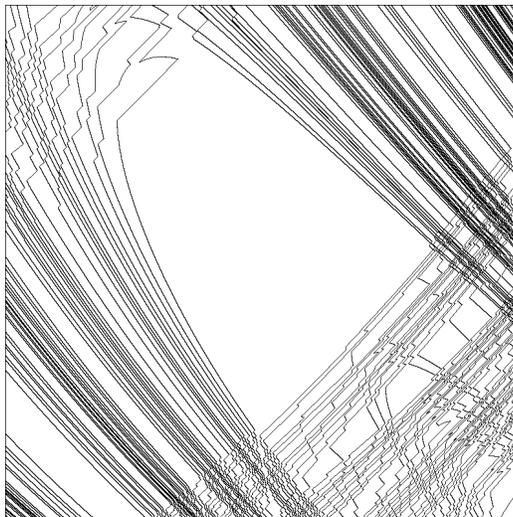


Figure 16. Curves of constant s around an arch configuration in the (A, B) -plane. (Contour interval: $\Delta s = .002$. The illustrated region $[0.835, 0.885] \times [0.01, 0.03]$ is exactly the same as that shown in Figure 11.)

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. (Compare Appendix B, as well as Figure 16.)

3. Complex Cubics: the Connectedness Locus.

In this section we consider the dynamics of a complex cubic map. Following Douady and Hubbard, for any complex polynomial map $f : \mathbf{C} \rightarrow \mathbf{C}$ 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, then it follows that 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, Devaney and Keen].)

Branner and Hubbard define the *connectedness locus* for a parametrized family of polynomial maps to be the set of all parameter values which 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 that it is connected, with connected complement. In the cubic case, Branner and Hubbard 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 4-dimensional disks $D_{i+1} \subset \text{Interior}(D_i)$ in the parameter space \mathbf{C}^2 . (Compare [Brown].) The corresponding assertion for higher degrees has recently been proved by Lavaurs.

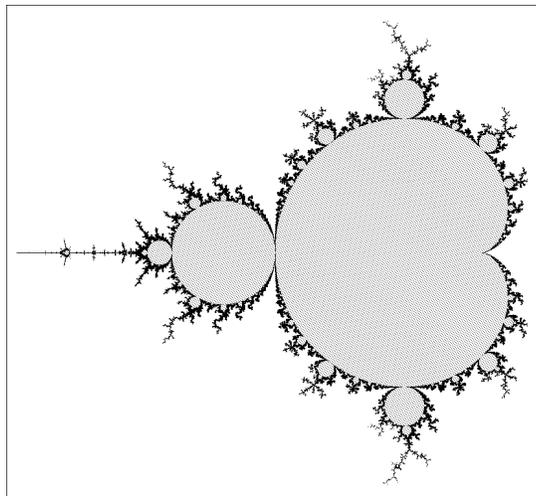


Figure 17. The Mandelbrot set.

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

(1) The Mandelbrot set is replete with small copies of itself. In fact, Douady and Hubbard show 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 not just one kind of hyperbolic component, rather there are 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 tree-like 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 tree-like organization. A corresponding conjecture might be that the hyperbolic components of bounded type are organized as the vertices of an acyclic two-dimensional complex.

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

It is difficult to visualize this complex cubic connectedness locus, which is an extremely complicated 4-dimensional object with fractal boundary. (Compare [Dewdney].) A more accessible situation arises if we consider the dynamics of cubic polynomial maps $f : \mathbf{C} \rightarrow \mathbf{C}$ which 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

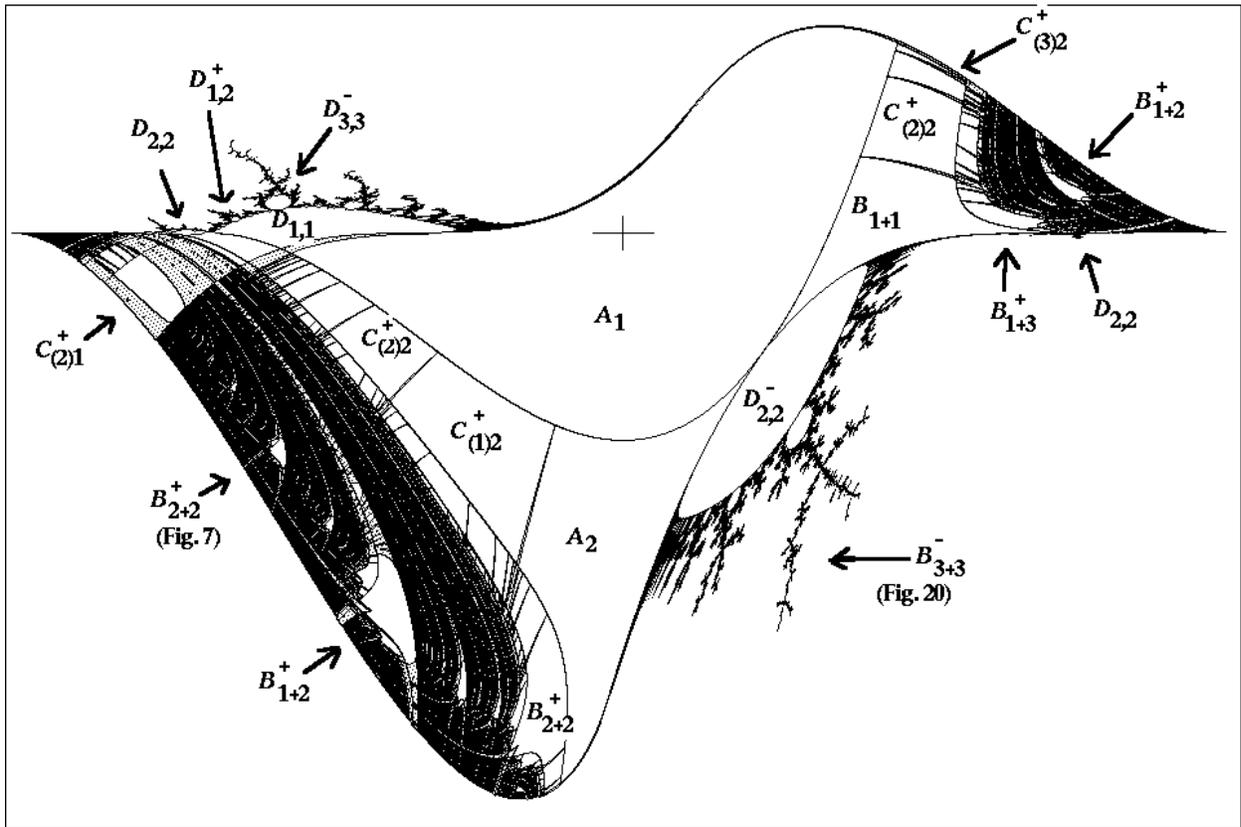


Figure 18. Complex connectedness locus intersected with the real (A, B) -plane. (Region: $[-1, 1] \times [-1.7, .65]$.)

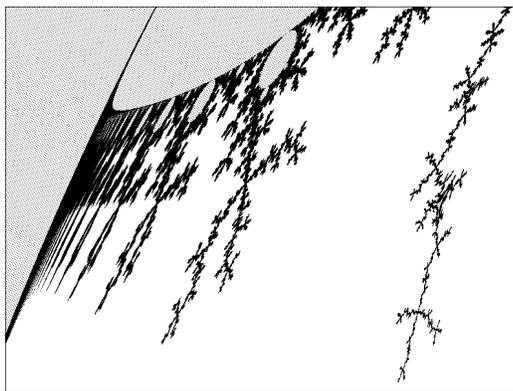


Figure 19. Detail in the lower right quadrant, showing lack of local connectivity. (Region: $[.02, .32] \times [-1.15, -.7]$.)

connectedness locus with the real (A, B) -plane. The resulting intersection is shown in Figure 18. Here, for 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. Grey [or black] indicates 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 §2, and is bounded by smooth curves. For parameter values in the regions \mathcal{R}_2 and \mathcal{R}_3 of §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 this figure are explained in §2 (Figure 6), §4, or in Appendix B; with the sign of AB 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. Alternatively, using Smith theory, as described in [Bredon, p. 145], since the real connectedness locus in 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 $AB < 0$ seems quite reminiscent of Figure 17, and in fact we will see in §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 $\text{Per}_2(1)$. (Compare §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], is in drastic contrast to the situation for degree 2 maps. In fact, as noted above, it is widely believed although not yet proved that the Mandelbrot set is locally connected.

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 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, 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*. Following Douady and Hubbard, 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 will be developed further in a subsequent paper, which will study monic and 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], the different hyperbolic components in the complex cubic connectedness locus can be roughly classified into four different types, as follows. (Compare §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 .

Case \mathcal{A}_p : Adjacent Critical Points. 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 $p \geq 1$.

Case \mathcal{B}_{p+q} : Bitransitive. The two critical points belong to different components U_0 and U_1 of U , but there exist integers $p, q > 0$ so that $f^{\circ p}(U_0) = U_1$ 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 $p + q$.

Case $\mathcal{C}_{(t)p+q}$: 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 integers $t + p \geq t \geq 1$ 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 $\mathcal{C}_{(t)q}$.

Case $\mathcal{D}_{p,q}$: 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 $f^{\circ p}(U_0) = U_0$, then 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, then the orbits of the two 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 us look at hyperbolic components in the real (A, B) -plane. In the Adjacent Case, there are only two real hyperbolic components. These have periods one and two, and are centered at the origin and the point $(0, -1)$ respectively. Both of these 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 §2.

In the Bitransitive Case, if the center lies in a quadrant where $AB > 0$, then we obtain a swallow configuration, as discussed in §2. However, if the center lies in one of the quadrants where $AB < 0$, then 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} 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 is compact and connected, and has an exact three-fold 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 non-locally connected set. (Figure 22. Compare [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 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, according as the period is even or odd.

If the center of a hyperbolic component lies precisely along the A -axis, then 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 $AB > 0$, then we obtain a product configuration, as discussed in §2. If it satisfies $AB < 0$, then we obtain a copy of the Mandelbrot set, while if it lies exactly on the A -axis then 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, 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

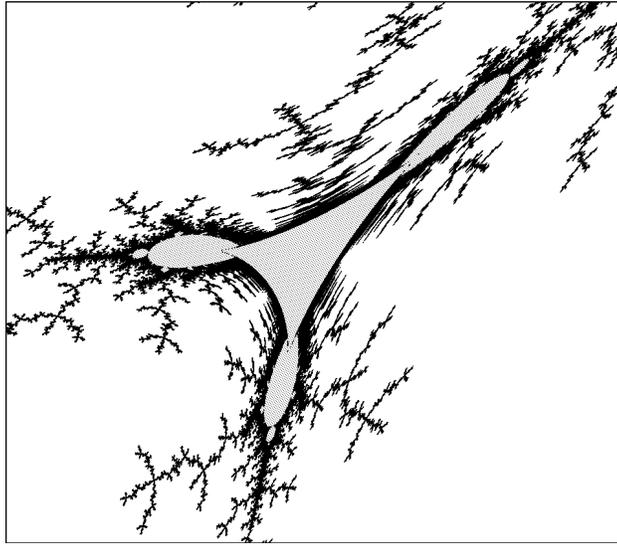


Figure 20. Detail of the right center of Figure 19, showing a small “tricorn” shaped configuration. For the center point $(.27286, -.93044)$, the third iterate of the cubic map carries each critical point to its complex conjugate. (Region: $[.265, .281] \times [-.958, -.903]$.)

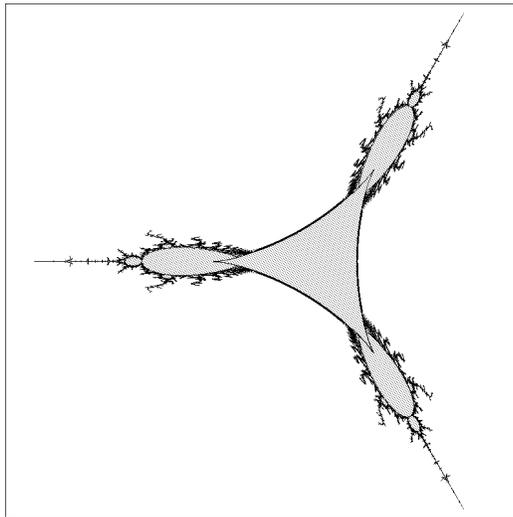


Figure 21. The prototype tricorn, in the c -plane where $z \mapsto (z^2 + c)^2 + \bar{c}$. (Region: $[-2.2, 1.4] \times [-1.8, 1.8]$.)

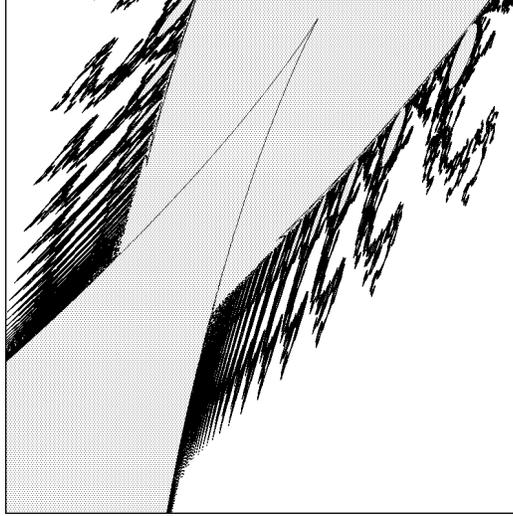


Figure 22. Detail of Figure 21, showing non local connectivity. (Region: $[.18, .5] \times [.34, .66]$.)

immediately contiguous and subordinated 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, Douady and Hubbard have provided 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.

By the *barycenter* of a polynomial map

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

of degree $n \geq 2$ is meant the unique point $\hat{x} = -\frac{1}{n}c_{n-1}/c_n$ 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 $x \mapsto g(x) = f(x + \hat{x}) - \hat{x}$ which 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 $\gamma^{n-1} = c_n$, then the linearly conjugate polynomial $x \mapsto \gamma g(x/\gamma)$ is *monic*, that is 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), then computation shows that

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

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

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

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 $\text{Per}_1(\mu)$. By definition, the pair (A, B) belongs to this locus 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 $x = \kappa$, then 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

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

Here the saddle node curve $\text{Per}_1(1)$ forms part of the upper boundary of the principal region, which is labeled \mathcal{A}_1 in Figure 18, and the *bifurcation locus* $\text{Per}_1(-1)$, 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 $\text{Per}_1(0)$ 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

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

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.

$$\frac{1}{9} \sum (\mu_i - 1) = A + \frac{1}{3} \quad (A.4)$$

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

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

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

$$\mu_3 = 2 + \frac{1 - \mu_1\mu_2}{\mu_1 + \mu_2 - 2}$$

as a function of μ_1 and μ_2 . (In fact if $\mu_1 \neq \mu_2$ then the curves $Per_1(\mu_1)$ and $Per_1(\mu_2)$ intersect transversally at a single point, which also belongs to $Per_1(\mu_3)$.) If we exclude the indeterminate case $\mu_1 = \mu_2 = 1$, then it follows from this formula 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 which 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 corresponding parameter pair (A, B) lies on the curve $Per_1(2)$. In fact, if $\mu_1 = e^{i\theta}$, then we can compute $A = \frac{2}{9}(\cos(\theta) - 2)$ 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 $Per_1(2)$.* 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 $Per_1(-1) \cap Per_1(2)$ and $Per_1(1) \cap Per_1(2)$.

To study the curve $Per_2(\mu)$, 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, B , as functions of κ^2 . In the case $\mu = 1$, there is a substantial simplification. In fact, as $\mu \rightarrow 1$ the curve $Per_2(\mu)$ tends towards a reducible curve, which is the union of two irreducible constituents. One of these is the bifurcation locus $Per_1(-1)$, which we do not consider to be part of $Per_2(1)$ since the period two orbit has degenerated to a fixed point, and the other is the required curve

$$Per_2(1) : \quad B = 4(A - \frac{2}{3})^3,$$

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 $A = \frac{2}{9}$, $B = -4(4/9)^3 = -\frac{256}{729}$ where the two irreducible components come together. (See the discussion below. Figure 18 is very distorted around this point.)

Remark. A generic cubic map has three period two orbits. If μ_1, μ_2, μ_3 are the derivatives around these three orbits, then the elementary symmetric functions σ_i of the

μ_i can be expressed as polynomial functions of A and B . More explicitly

$$\begin{aligned}\sigma_1 &= 9(3 - 4A) \\ \sigma_2 &= 9^2(3 - 8A + 16A^3 - 12A^4 + 2B + 3AB) \\ \sigma_3 &= \sigma_2 - \sigma_1 + 1 + 9^3\left(B - 4\left(A - \frac{2}{3}\right)^3\right)\left(B - \left(A - \frac{1}{3}\right)\left(A + \frac{2}{3}\right)^2\right).\end{aligned}$$

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

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

as given in §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 $f(a) = a \pm i$. Multiplying the equation by $\pm i$ and squaring both sides, we obtain the formula

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

as given in §2.

Points in the (A, B) -parameter plane where two of these curves intersect may be of particular interest. For example, the bifurcation locus $\text{Per}_1(-1)$, which forms the lower part of the boundary of the principal region \mathcal{A}_1 in Figure 18, grazes the saddle node curve $\text{Per}_2(1)$ tangentially at the point $A = 2/9$, $B = -256/729$ 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 which touch at this point.) The saddle node curve $\text{Per}_1(1)$ grazes the critically preperiodic curve $\text{Preper}_{(1)1}$ tangentially at the point $A = 1/9$, $B = 256/729$, which lies on the boundary between the regions \mathcal{R}_1 and \mathcal{R}_2 . (Compare Figure 2.) Similarly, the curves $\text{Per}_2(1)$ and $\text{Preper}_{(1)2}$ meet tangentially at $A = -1/36$, $B = -15625/11664$ (or $a = i/6$, $b = 125i/108$).

The top boundary of the region $\mathcal{D}_{1,1}$ in Figure 18 forms part of the curve $\text{Per}_1(2)$, characterized by the property that there are two mutually conjugate indifferent fixed points.

This curve intersects the saddle node curve $\text{Per}_1(1)$ transversally at the corner point $A = -2/9, B = 4/729$ of this region. (Presumably there are two similar transverse intersections of the saddle node curve $\text{Per}_2(1)$ 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 $A = 1/3, B = 16/27 = .59259\dots$, and the smallest value occurs at $A = -1/6, B = -(1 + \sqrt{2/27})^2 = -1.6184\dots$, 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.

The table below 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 2 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 $B < 0$. Here the notations $\mathcal{A}-\mathcal{D}$ in the third column are explained in §4 or in §2 (Figure 6). Thus we write:

\mathcal{A}_p for a component with *adjacent critical points*, where the attracting orbit has period p . These are exactly the hyperbolic components whose center point lies on the line $A = 0$, where the two 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 $AB < 0$.

However, the superscript $+$ has been added whenever $AB > 0$ so that the critical points are real and distinct, and the superscript $-$ has been added whenever $AB < 0$ 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 \mapsto^n for the n -th iterate of the cubic map. For example $c \mapsto^3 c'$ 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 $AB \geq 0$.

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

Remark. When $B = 0$, note that there is a hyperbolic component centered at $(A, 0)$ if and only if there is a hyperbolic component 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 hyperbolic. However, it may happen that one of these two belongs to a bitransitive component (Type \mathcal{B}) while the other has disjoint attracting orbits (Type \mathcal{D}).

Further information about these twenty hyperbolic components can be read from Figure 23, which shows the corresponding *Hubbard trees*. (Compare [Douady-Hubbard, 1984-85].) Each Hubbard tree is a simplified picture which 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, which correspond to odd mappings with $B = 0$, are also symmetric about a vertical axis.

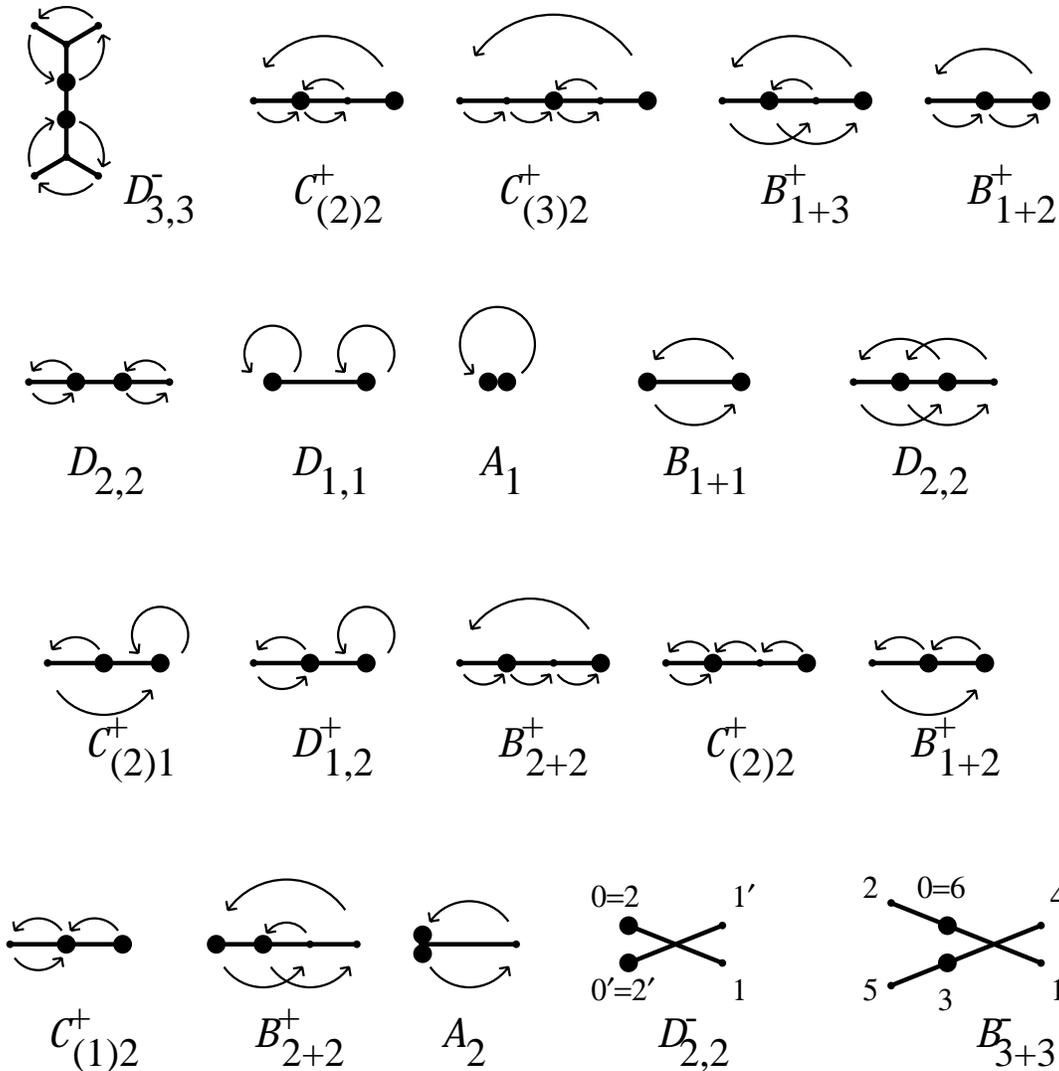


Figure 23. Hubbard trees for the centers of twenty hyperbolic components, as listed in Appendix B. (In the last two diagrams, vertex 0 maps to 1, to 2, . . .)

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, then 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. (Compare [Milnor, 1989 §5.6] or [Fisher].) If this distance is less than the pixel size, then

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, then 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 iteration, the orbits are checked for approximate periodicity with small period, and if both critical orbits have the same period then 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 $\text{Per}_p(\pm 1)$ 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. (Compare Appendix A.)

References.

- M. Benedicks and L. Carleson, On the iterations of $1 - ax^2$ on $(-1, 1)$, *Annals of Math.* **122** (1985), 1-25.
- P. Blanchard, Complex analytic dynamics on the Riemann sphere, *Bull. Am. Math. Soc.* **11** (1984), 85-141.
- P. Blanchard, Disconnected Julia sets, pp. 181-201 of "Chaotic Dynamics and Fractals", edit. Barnsley and Demko, Acad. Press 1986.
- L. Block and J. Keesling, Computing the topological entropy of maps of the interval with three monotone pieces, to appear.
- B. Branner, Iterations by odd functions with two extrema, *J. Math. Anal. and Apl.* **105** (1985), 276-297.
- B. Branner, The parameter space for cubic polynomials, pp. 169-179 of "Chaotic Dynamics and Fractals", edit. Barnsley and Demko, Acad. Press 1986.
- B. Branner, The Mandelbrot set, pp. 75-105 of "Chaos and Fractals", edit. Devaney & Keen, Proc. Symp. Applied Math. **39**, Amer. Math. Soc. 1989.
- B. Branner and A. Douady, Surgery on complex polynomials, Proceedings Symposium on Dynamical Systems, Mexico 1986, to appear.
- 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.
- G. Bredon, Introduction to compact transformation groups, Academic Press 1972.
- M. Brown, A proof of the generalized Schoenflies theorem, *Bulletin A.M.S.* **66** (1960) 74-76. (See also "The monotone union of open n-cells is an open n-cell", *Proc.A.M.S.* **12** (1961) 812-814.)
- O. Chavoya-Aceves, F. Angulo-Brown and E. Piña, Symbolic dynamics of the cubic map, *Physica D* **14** (1985) 374-386.
- R. Devaney, L. Goldberg and J. H. Hubbard, A dynamical approximation to the exponential map by polynomials, to appear.
- A. K. Dewdney, Computer Recreations, *Scient. Amer.*, Nov. 1987, p. 144.
- A. Douady, Systèmes dynamiques holomorphes, *Sém. Bourbaki* 599, *Astérisque* **105-106** (1983), 39-63.

- A. Douady and J. H. Hubbard, Étude dynamique des polynômes complexes I (1984) & II (1985), Publ. Math. Orsay.
- A. Douady and J. H. Hubbard, On the dynamics of polynomial-like mappings, *Ann. Sci. Ec. Norm. Sup. (Paris)* **18** (1985), 287-343.
- Y. Fisher, Exploring the Mandelbrot set, pp. 287-296 of “The Science of Fractal Images”, edit. Peitgen and Saupe, Springer 1988.
- S. Friedland and J. Milnor, Dynamical properties of plane polynomial automorphisms, *Ergodic Theory and Dynamical Systems* **9** (1989) 67-99.
- B. Friedman and C. Tresser, Comb structure in hairy boundaries: some transition problems for circle maps, *Physics Let. A* **117** (1986) 15-22.
- J. Guckenheimer, Sensitive dependence to initial conditions for one dimensional maps, *Comm. Math. Phys.* **70** (1979), 133-160.
- H. El Hamouly and C. Mira, Singularités dues au feuilletage du plan des bifurcations d’un difféomorphisme bi-dimensionnel, *C. R. Acad. Sci. Paris* **294** (1982), 387-390.
- M. Jakobson, Absolutely continuous invariant measures for one-parameter families of one-dimensional maps, *Comm. Math. Phys.* **81** (1981), 39-88.
- P. Lavaurs, Le lieu de connexité des polynômes du troisième degré n’est pas localement connexe, in preparation.
- R. MacKay and C. Tresser, Boundary of chaos for bimodal maps of the interval, *J. Lond. Math. Soc.*, to appear.
- R. MacKay and C. Tresser, Some flesh on the skeleton: the bifurcation structure of bimodal maps, *Physica* **27D** (1987) 412-422.
- J. Milnor, Non-expansive Hénon maps, *Advances in Math.* **69** (1988) 109-114.
- J. Milnor, Self-similarity and hairiness in the Mandelbrot set, pp. 211-257 of “Computers in Geometry”, edit. M. Tangora, *Lect. Notes Pure Appl. Math.* **114**, Dekker 1989.
- J. Milnor, Dynamics in One Complex Variable, Introductory Lectures, Stony Brook IMS preprints 1990/5.
- J. Milnor, Hyperbolic components in spaces of polynomial maps, in preparation.
- J. Milnor, On cubic polynomials with periodic critical point, in preparation.
- J. Milnor and W. Thurston, Iterated maps of the interval, pp. 465-563 of “Dynamical Systems (Maryland 1986-87)”, ed. J. Alexander, *Lect. Notes Math.* **1342**, Springer 1988.
- M. Misiurewicz and W. Szlenk, Entropy of piecewise monotone mappings, *Astérisque* **50** (1977), 299-310 and *Studia Math.* **67** (1980), 45-63.
- E. Piña, Comment on “Study of a one-dimensional map with multiple basins”, *Phys. Rev. A* **30** (1984), 2132-2133.
- M. Rees, Components of degree two hyperbolic rational maps, *Invent. Math.* **100** 1990, 357-382.
- M. Rees, A partial description of parameter space of rational maps of degree two, Part 1: preprint, Univ. Liverpool 1990; and Part 2: Stony Brook IMS preprints 1991/4.
- J. Ringland and M. Schell, Universal geometry in the parameter plane of maps of the interval, to appear.
- J. Ringland and M. Schell, The Farey tree embodied — in bimodal maps of the interval, to appear.

- T. Rogers and D. Whitley, Chaos in the cubic mapping, *Math. Modelling* **4** (1983) 9-25.
- M. Rychlik, Another proof of Jakobson's theorem and related results, *Erg. Theory and Dyn. Sys.* **8** (1988) 93-109.
- 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.
- R. Winters (Boston University), Of tricorns and biquadratics, in preparation.
- B. Wittner, On the bifurcation loci of rational maps of degree two, Thesis, Cornell University 1988.

Institute for Mathematical Sciences, SUNY Stony Brook NY
September 1991