BIRATIONAL MAPS WITH SPARSE POST-CRITICAL SETS

JEFFREY DILLER

1. A FAMILY OF BIRATIONAL MAPS

Very little is known concerning global dynamics of holomorphic maps in dimensions larger than one. Results that apply to large classes of maps (say polynomial automorphisms of \mathbb{C}^2 [?] or endomorphisms of \mathbb{P}^n [?], for example) are confined mostly to the level of ergodic theory, describing dynamics 'almost everywhere' with respect to natural invariant measures and currents. More detailed accounts exist only for specific examples. The immediate purpose of this exposition is to discuss one such example at length. Along the way I hope to also serve the broader purposes of making theorems about general maps more accessible and of indicating promising places to look for further tractable examples. All of the work described here is joint with Eric Bedford and appears in more complete form in the preprint [?]

We will consider the one parameter family of maps, given in affine coordinates by

(1)
$$f(x,y) = \left(y\frac{x+a}{x-1}, x+a-1\right).$$

One checks easily that f is invertible, at least away from a couple of 'exceptional' curves along which the behavior of f is either degenerate or undefined on \mathbb{C}^2 . In fact f extends as a so-called *birational map* to any complex surface compactifying \mathbb{C}^2 . However, as I will explain now, it is particularly convenient to regard f as a birational self-map of $\mathbb{P}^1 \times \mathbb{P}^1$ ().

Modulo linear equivalence \sim , the divisors in $\mathbf{P}^1 \times \mathbf{P}^1$ form a group (the *Picard group*) $\operatorname{Pic}(\mathbf{P}^1 \times \mathbf{P}^1) \cong \mathbf{Z} \times \mathbf{Z}$ generated by a vertical line $V := [x = \operatorname{const}]$ and a horizontal line $H := [y = \operatorname{const}]$. Using fto pull back local defining functions for divisors, we obtain a linear action f^* on divisors. This action clearly preserves linear equivalence and so descends to a linear map $f^* : \operatorname{Pic}(\mathbf{P}^1 \times \mathbf{P}^1) \circlearrowleft$ on the Picard group. From the above formula, ones sees that horizontal lines pull back to vertical lines, and vertical lines pull back to hyperbolas with horizontal/vertical asymptotes. Hence with respect to the ordered basis $\left(V,H\right)$

(2)
$$f^* = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

In particular, the spectral radius of f^* is the golden ratio $(1 + \sqrt{5})/2$. The dynamical relevance of this quantity is revealed by the following result due essentially to Gromov (see Dinh and Sibony [?] for the most general version to date)

Theorem 1.1. Let $f : X \bigcirc$ be a birational map on a complex projective surface X. Then the topological entropy $h_{top}(f)$ of f satisfies

$$h_{top}(f) \leq \lim_{n \to \infty} \frac{\log \|(f^n)^*\|}{n}.$$

In addition, Dujardin [?] has recently shown that this inequality is actually an equality for a large class of birational maps (including those in (1). So for f given by (1) we have

$$h_{top}(f) = \log \frac{1 + \sqrt{5}}{2}$$

provided that

(3)
$$(f^n)^* = (f^*)^n$$
 for all $n \in \mathbf{N}$.

This latter identity can fail dramatically in general, but we will see shortly that it holds for the family (1) for all but countably many values of the parameter a. Fornæss and Sibony call maps satisfying (3)*algebraically stable*.

It should perhaps be stressed that (3) is a property of both the map and the choice of compactification of \mathbb{C}^2 . For example, if I were treating f as a self-map of \mathbb{P}^2 , then the Picard group acted on by f^* would be one-dimensional, generated by a generic line in \mathbb{P}^2 , and f^* would simply double this generator. However, $(f^2)^*$ would multiply by 3 (check this!) rather than $2^2 = 4$. Thus the surface $\mathbb{P}^1 \times \mathbb{P}^1$ is 'compatible' with the map f in a way that \mathbb{P}^2 is not.

To better understand the situation, let us reconsider things from a geometric point of view. On $\mathbf{P}^1 \times \mathbf{P}^1$, the critical set $\mathcal{C}(f)$ of f is the pair of lines $\{x = -a\} \cup \{x = 1\}$. As is the case for birational maps generally, the components of $\mathcal{C}(f)$ are critical because they are *exceptional*: each is mapped to a single point¹: $\{x = -a\}$ to (0, -1)

¹When V is a curve that meets I(f), we define f(V) to be the set $\overline{f(V - I(f))}$. In other words, f(V) is the proper transform of V and excludes all components of $\mathcal{C}(f^{-1})$. This notion of f(V) does not entirely accord with that of $f_*V := (f^{-1})^*V$:

and $\{x = 1\}$ to (∞, a) . Consequently, the inverse map

$$f^{-1}(x,y) = (y-a+1, x\frac{y-a}{y+1})$$

cannot be defined continuously at either image point, a fact which one can verify directly from the formula for f^{-1} . The set $I(f^{-1}) := \{(0, -1), (\infty, a)\}$ is called the *indeterminacy set* of f^{-1} . Similar analysis reveals that

$$C(f^{-1}) = \{y = -1\} \cup \{y = a\}$$
 $I(f) = \{(-a, \infty), (1, 0)\}.$

If we change our compactification of \mathbb{C}^2 , the sets $\mathcal{C}(f^{\pm 1})$ and $I(f^{\pm 1})$ are all prone to change as well. It turns out (in general) that (3) is equivalent to

(4)
$$f^{n}\mathcal{C}(f) \cap f^{-m}\mathcal{C}(f) = \emptyset \text{ for all } n, m > 0$$

In other words f satisfies (3) if and only if 'postcritical' orbits

$$\mathcal{PC}(f) := \bigcup_{n>0} f^n \mathcal{C}(f), \qquad \mathcal{PC}(f^{-1}) \bigcup_{m>0} f^{-m} \mathcal{C}(f^{-1})$$

avoid each other.

The condition (4) has a deceptively simple appearance. For general maps, it can be quite difficult to verify, because it requires knowing about the full orbit of each component of C(f). Things are easier, however, for the example at hand. Our map f has the additional virtue that it preserves a meromorphic two form:

$$f^*\eta = f_*\eta = \eta := \frac{dx \wedge dy}{y - x + 1}.$$

It follows more or less immediately that the support of the divisor

$$[\eta] = [x = \infty] + [y = \infty] + [y = x - 1]$$

of η is invariant under f. Direct computation with parametrizations reveals more specifically that $\{y = x - 1\}$ is fixed and the lines at infinity are switched according to

$$(x, x-1) \mapsto (x+a, x+a-1)$$
 $(\infty, y) \mapsto (y, \infty) \mapsto (\infty, y+a-1).$

In particular, the point (∞, ∞) is fixed by f.

Invariance of η also implies that critical components of f must map into supp $[\eta]$. Hence $\mathcal{PC}(f), \mathcal{PC}(f^{-1}) \subset \text{supp } [\eta]$, and the happy consequence is that we can determine whether or not f satisfies (4) by

in general $f(V) \subset \text{supp } f_*V$, but the inclusion will be proper when $V \cap I(f) \neq \emptyset$. To take a concrete example, we have $f(\{x = a\}) = (0, 1)$, whereas $f_*[x = -a] = [y = 1]$ includes the 'image' of the point $(-a, \infty) \in I(f)$.

restricting our attention to the completely tractable one dimensional dynamics of f on supp $[\eta]$.

2. Real dynamics for negative parameters

Though we could easily, in light of the preceeding discussion, identify all parameters a for which (4) fails, let us attend only to the case a < 0. The parameter a = -1 is special, because the first coordinate of f degenerates, the critical and indeterminacy sets disappear, and the map dynamics become trivial. For all other a < 0, (4) holds in a particularly robust fashion. For example $\mathcal{PC}(f) \cap \{y = x - 1\}$ is just the forward orbit of the point (-1, 0). If we let

$$S := \{(x, x - 1) : x \le 0\}$$

be the real interval in $\{y = x - 1\}$ that stretches from (-1, 0) down and left to (∞, ∞) , then we see that $f(S) \subset S$ when a < 0. Therefore $\mathcal{PC}(f) \cap \{y = x - 1\} \subset S$. Likewise, the interval

$$U := \{(x, x - 1) : x \ge 1\}$$

stretching from (1,0) up and right to (∞,∞) satisfies $f^{-1}(U) \subset U$ when a < 0 and therefore contains $\mathcal{PC}(f^{-1}) \cap \{y = x - 1\}$. As U and S are disjoint, it follows that $\mathcal{PC}(f) \cap \mathcal{PC}(f^{-1})$ contains no points in $\{y = x - 1\}$.

Similar observations apply to the lines $\{x = \infty\}$ and $\{y = \infty\}$. When a < 0, each line contains disjoint, forward/backward invariant, real intervals S and U separating $\mathcal{PC}(f)$ from $\mathcal{PC}(f^{-1})$, and it follows that $\mathcal{PC}(f) \cap \mathcal{PC}(f^{-1})$ is empty.

Figure 1 summarizes this state of affairs for a < -1. The real points in $\mathbf{P}^1 \times \mathbf{P}^1$ form a torus. Removing $\sup [\eta]$ divides the remaining real points into two open sets, labeled 0 and 1. The boundary of each open set is exactly equal to the real points in $\sup [\eta]$. S(table) and U(nstable) segments in each boundary component are thickened for emphasis. Finally, the critical and indeterminacy sets of f and f^{-1} are included for the sake of completeness. The picture remains valid for parameter values -1 < a < 0, except that the critical lines for f (and for f^{-1}) switch places.

Let us regard two stable segments that are adjacent in the boundary of region 0 or 1 as part of a single larger boundary segment. In this way, the boundaries of regions 0 and 1 may be regarded as 'rectangles', each with opposing pairs of stable and unstable 'sides'. This suggests that for real parameters a, we try to use the two regions as a Markov partition for the dynamics of f. Let Σ be the space of bi-infinite sequences



FIGURE 1. Real partition by $\sup \eta$. The critical set of f/f^{-1} is shown as dashed/dotted lines, indeterminacy set of f/f^{-1} as hollow/solid circles, and sample stable arcs as wavy lines. The arrows indicate the direction of motion of points under iteration of f.

 $\{0,1\}^{\mathbf{Z}}$ (with the product topology) and

$$D := \{ p \in \mathbf{R}^2 : f^n(p) \notin \operatorname{supp} [\eta] \text{ for all } n \in \mathbf{Z} \} = \mathbf{R}^2 - \operatorname{supp} [\eta] - \bigcup_{n \in \mathbf{Z}} \mathcal{C}(f^n)$$

consist of those points whose orbits lie entirely in the interior of regions 0 and 1. Define a map

$$w: D \to \Sigma, \qquad p \mapsto \dots w_{-1} w_0 \cdot w_1 w_2 \dots,$$

where $w_j \in \{0, 1\}$ records the region that contains $f^j(p)$. It is not hard to see that w is continuous. Moreover, if $\sigma : \Sigma \circlearrowleft$ is the shift homeomorphism

$$\dots w_{-1}w_0 \cdot w_1w_2 \dots \stackrel{\sigma}{\mapsto} \dots w_{-1}w_0w_1 \cdot w_2 \dots$$

then we clearly have a commutative diagram

$$\begin{array}{cccc} D & \xrightarrow{f} & D \\ w \downarrow & & \downarrow w \\ \Sigma & \xrightarrow{\sigma} & \Sigma \end{array}$$

More importantly and much less obviously, we can say a great deal about the fiber of w over any point in Σ . Consider the following subsets of D.

$$D_{+} = \{ p \in D : \lim_{n \to \infty} f^{n}p = (\infty, \infty) \}$$
$$D_{-} = \{ p \in D : \lim_{n \to \infty} f^{-n}p = (\infty, \infty) \}$$
$$\Omega = D - D_{+} - D_{-}.$$

Let us call the coding w(p) of $p \in D$ forward alternating if some righthand tail $w_j w_{j+1} w_{j+2} \dots$ of w(p) has the form $0101 \dots$. Let us call w(p) backward alternating if some lefthand tail $\dots w_{j-2} w_{j-1} w_{-j}$ has the analogous property. Let $\Sigma_G \subset \Sigma$ denote the (closed) subset consisting of all sequences without consecutive 1's. The main result of this exposition is

Theorem 2.1. Suppose that a < 0, $a \neq -1$. Let $p \in D$ be any point. Then

- $p \in D_+$ if and only if w(p) is forward alternating.
- $p \in D_{-}$ if and only if w(p) is backward alternating.

Finally, $w : \Omega \to \Sigma$ is a homeomorphism onto those sequences in Σ_G that are neither forward nor backward alternating.

Since the dynamics of f on supp $[\eta]$ are trivial, Theorem 2.1 gives a rather precise topological description of the real dynamics of f. I will quickly indicate two consequences of this theorem and then discuss some ingredients of the proof.

Corollary 2.2. Ω consists exactly of those points in D with recurrent orbits.

The entropy of a restricted map never exceeds that of the map itself, so on this general principle we know that

$$h_{top}(f:\Omega \circlearrowleft) \le h_{top}(f:\overline{\mathbf{R}^2} \circlearrowright) \le h_{top}(f:\mathbf{P}^1 \times \mathbf{P}^1 \circlearrowright) = \frac{1+\sqrt{5}}{2}.$$

On the other hand, the shift map σ restricts to a well-defined homeomorphism of Σ_G whose entropy is well-known to be $\log \frac{1+\sqrt{5}}{2}$. Since removing the relatively small sets of forward/backward alternating codings does not alter the value of the entropy, we can conclude that

6

Corollary 2.3. For all a < 0, $a \neq -1$, the topological entropy of f as a real map is $\log \frac{1+\sqrt{5}}{2}$.

The fundamental idea underlying Theorem 2.1 is that forward and backward images of real arcs may be studied in two different ways: from a combinatorial point of view based on Figure 1, and from the more abstract perspective of complex intersection theory. I discuss these points of view in order.

3. Combinatorics

From now on I will assume that a < -1. I call a real arc 'stable' if it is completely contained in one of the two regions in Figure 1, and it joins the two unstable segments in the boundary of that region. To justify this definition, let me consider for example the preimage $f^{-1}(\gamma)$ of a stable arc γ in region 0. Say for specificity's sake that γ joins the unstable segment in $\{y = x - 1\}$ to the unstable segment in $\{y = x - 1\}$ ∞ }. Then γ necessarily crosses both lines in $\mathcal{C}(f^{-1})$, and the preimage $f^{-1}(\gamma)$ must therefore contain three subarcs: one joining the unstable segment in $\{y = x - 1\} = f^{-1}\{y = x - 1\}$ to $(\infty, -a) = f^{-1}\{y = -1\},\$ one joining $(\infty, -a)$ to $(1, 0) = f^{-1}\{y = a\}$, and one joining (0, -1)to the unstable segment in $\{x = \infty\} = f^{-1}\{y = \infty\}$. By checking the images of points in γ near supp $[\eta] \cup \mathcal{C}(f^{-1})$, one sees that the first and third arcs lie in region 0, whereas the second lies in region 1. In particular the second third subarcs join opposing unstable segments in regions 0 and 1, respectively, and are therefore themselves stable (the first subarc is not stable since both of its endpoints lie in the same unstable segment in region 0). Repeating this argument proves that the preimage $f^{-1}(\gamma)$ of an stable arc γ in region 1 must contain an stable arc in region 0. After induction we arrive at

Theorem 3.1. Let $m \ge 0$ and $w_0 \cdot w_1 \dots w_m$ be a finite righthand sequence of 0's and 1's without consecutive 1's. Let α be a stable arc in region w_m . Then $f^{-m}(\alpha)$ contains a stable arc γ in region w_0 such that $f^j(\gamma)$ lies in region w_j for $j = 0, \dots, m$.

Of course, we can also define 'unstable' arcs in regions 0 and 1, and proceed in exactly the same fashion to prove

Theorem 3.2. Let $n \ge 0$ and $w_{-n} \ldots w_0$ be a finite lefthand sequence of 0's and 1's without consecutive 1's. Let β be an unstable arc in region w_{-n} . Then $f^n(\beta)$ contains an unstable arc γ in region w_0 such that $f^{-j}(\gamma)$ lies in region w_j for $j = 0, \ldots, n$.

The fact that stable and unstable boundary segments of regions 0 and 1 are disjoint implies that any stable arc in a given region intersects any unstable arc from the same region. So Theorems 3.2 and 3.1 give us a convenient way to produce points with orbits coded by finite twosided sequences of any extent.

Corollary 3.3. Let $n, m \ge 0$ and $w_{-n} \dots w_0 \cdot w_1 \dots w_m$ be any finite sequence of 0's and 1's without consecutive 1's. Then there is a point $p \in D$ such that $c(p) = \dots w_{-n} \dots w_0 \cdot w_1 \dots w_m \dots$

It is not quite immediate (and not quite true!) that the image w(D) of the coding map contains Σ_G , let alone that the assertions of Theorem 2.1 concerning $w|_{\Omega}$ are true. However, Corollary 3.3 is clearly a step in the right direction. Further progress depends on refining the partition shown in Figure 1.

For any $n \geq 0$, every component in the critical set $\mathcal{C}(f^n)$ maps, eventually, into the stable portion of $\operatorname{supp}[\eta]$. So we can subdivide our original partition using $\mathcal{C}(f^n)$ for any $n \in \mathbb{N}$, designating all the new boundary components 'stable'. Similarly, we can subdivide by $\mathcal{C}(f^{-n})$, designating all inverse critical components 'unstable'. And while it is not strictly necessary, we can try to simplify the picture that results by recombining some of the new partition pieces, provided we take care to preserve invariance of stable/unstable boundary components. The result of this process, obtained with care and hindsight, is shown in Figure 2. The original regions 0 and 1 become smaller rectangles R_0 and R_1 , and the complement of $R_0 \cup R_1$ decomposes into overlapping regions labeled R_+ and R_- . Using only combinatorial arguments like the ones above, the following can be established.

Proposition 3.4. The conclusion of Corollary 3.3 holds with the regions 0 and 1 from Figure 1 replaced by regions R_0 and R_1 from Figure 2. Moreover,

- $f(R^+) \subset R^+$, and any point $p \in R^+ \cap D$ has a forward coding $w_0 \cdot w_1 \ldots$ that alternates and a forward orbit that tends to (∞, ∞) .
- $f^{-1}(R^-) \subset R^-$, and any point $p \in R^- \cap D$ has a backward coding $\ldots w_{-1}w_0$ that alternates and a backward orbit that tends to (∞, ∞) .
- $f(R_1) \cap R_1 = \emptyset$.

Together with the following, somewhat technically difficult result, Corollary 3.3 and Proposition 3.4 combine to imply everything in Theorem 2.1 except the injectivity of $f|_{\Omega}$.

8



FIGURE 2. Refinement of the original partition to include critical curves. Stable and unstable boundary segments are labeled 's' and 'u', respectively.

Proposition 3.5. Any point $p \in D$ such that $\lim_{n\to\infty} f^n(p) = (\infty, \infty)$ (respectively, $\lim_{n\to\infty} f^{-n}(p) = (\infty, \infty)$) must satisfy $f^n(p) \notin R^0 \cup R^1$ (respectively, $f^{-n}(p) \notin R^0 \cup R^1$) for arbitrarily large $n \in \mathbf{N}$.

4. INTERSECTION THEORY

Here is a slightly different and less precise way to state Corollary 3.3. Suppose we are given $i, j \in \{0, 1\}$, a stable arc α in region i, an unstable arc β in region j, and $m, n \in \mathbb{N}$. Then $f^{-m}(\alpha) \cap f^{n}(\beta)$ must contain at least

(5)
$$\left\langle \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n \mathbf{e}_i, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^m \mathbf{e}_j \right\rangle$$

distinct points in $R_0 \cup R_1$. Equation (5), in which \mathbf{e}_0 , \mathbf{e}_1 are the standard basis vectors for \mathbf{R}^2 , simply counts the number of codings $w_{-n} \dots w_0 \cdot w_1 \dots w_m$ that begin with digit $w_{-n} = i$, end with digit $w_m = j$, and contain no consecutive 1's throughout. The combinatorial arguments sketched above do not rule our the possibility that there might be *more*

intersections than (5) provides. To obtain control from above, I change tactics and consider only very special examples of stable and unstable curves.

Namely, I suppose that α is obtained by intersecting R_0 with a vertical line or R_1 by the preimage of a vertical line, and that β is obtained similarly. This turns out not to be too severe since both regions have a product structure given by stable and unstable curves of this sort. The advantage to the restriction is that complex intersection theory tells us exactly how many times one algebraic curve intersects another and therefore gives us an upper bound on $\#f^{-m}(\alpha) \cap f^n(\beta)$. The data needed to obtain this upper bound are the basis (V, H) for Pic $(\mathbf{P}^1 \times \mathbf{P}^1)$, the matrix (2) for f^* with respect to this basis, and additionally, the matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

for the intersection form for complex curves in $\mathbf{P}^1 \times \mathbf{P}^1$. The results take a bit of interpreting because the algebraic curves giving the stable and unstable foliations of R_1 also intersect R_0 is stable and unstable arcs. However, in the end we obtain an upper bound for $\#f^{-m}(\alpha) \cap f^n(\beta)$ that matches (5) exactly in all cases. In light of (2), we might have expected close agreement even before setting pencil to paper, but exact agreement is not a priori obvious (at least not to me). It is fortunate, though, because precise agreement between upper and lower bounds is the main thing needed to complete the proof of Theorem 2.1 (i.e. of injectivity of $f: \Omega \to \Sigma_G$.)

Rather than go into more detail here, I will describe some further consequences of intersection theory for dynamics of f. By using Lefschetz' theorem on periodic points, it can be shown that

Theorem 4.1. All periodic points of f are real. Indeed all except (∞, ∞) are saddle points contained in Ω , and saddle periodic points constitute a dense subset of Ω .

So far, I have mostly described the set $\Omega = D - D_- - D_+$ of points whose orbits lie neither the forward nor the backward basin of (∞, ∞) , but in fact the individual complements of $\Omega_+ := D - D_+$ and $\Omega_- := D - D_-$ yield to the same analysis.

Theorem 4.2. Ω_+ is the support of a geometric 1 current μ^+ . That is, there is a lamination \mathcal{L}^+ in $\mathbf{P}^1 \times \mathbf{P}^1 - \operatorname{supp}[\eta]$ and a measure ν^+ on the set $|\mathcal{L}|^+$ of leaves of this lamination such that

- $\mu^+(\zeta) = \int_{|\mathcal{L}^+|} \left(\int_L \zeta \right) \nu^+(L)$ for all 1 forms ζ ;
- supp $\mathcal{L}^+ = \Omega_+;$



Note that I am avoiding the matter of orientation in the first and last items. Figure 4 shows \mathcal{L}^+ by itself and together with the corresponding lamination \mathcal{L}^- complementing D_- . The common intersection of the two laminations is just (the closure of) Ω .

5. Conclusion

Complex intersection theory can be used to study dynamics of any rational map. Indeed the currents μ^+ and μ^- have general complex analogues for any dynamically interesting birational map, and the intersection between μ^+ and μ^- can often be understood in at least a measure theoretic sense (see [?]). What is special to the example I have just described is the presence of a good combinatorial structure. In my view, there are two key features of the example from which the combinatorics proceed. First of all, the post-critical orbits $\mathcal{PC}(f)$ and $\mathcal{PC}(f^{-1})$ lie in invariant curves and are therefore very easy to understand. Second, rather than being interlaced in some complicated fashion, the sets $\mathcal{PC}(f)$ and $\mathcal{PC}(f^{-1})$ are easily separated by dividing each

real invariant curve into a pair of intervals. Some of the other aspects of the example, such as the perfect agreement between intersection theory and combinatorics, remain mysterious to me. In a forthcoming paper, Bedford and I will describe another family of birational maps whose real dynamics can be analyzed in a similar fashion. It does not seem too hard to come by further families of maps with "sparse postcritical sets" so it is interesting to wonder how far the analysis described here can be extended.

Department of Mathematics, University of Notre Dame, Notre Dame, IN 46656

 $E\text{-}mail\ address: \texttt{diller.1@nd.edu}$

12