DYNAMICS OF BIRATIONAL MAPS OF $\mathbb{P}^2$

Jeffrey Diller

Cornell University

Abstract. Inspired by work done for polynomial automorphisms, we apply pluripotential theory to study iteration of birational maps of $\mathbb{P}^2$. A major theme is that success of pluripotential theoretic constructions depends on separation between orbits of the forward and backward indeterminacy sets. In particular, we show that a very mild separation hypothesis guarantees the existence of a plurisubharmonic escape function $\tilde{G}^+$ and the induced current μ^+. We show that under normalized pullback by the birational map, a large class of currents are attracted to μ^+. Under stronger separation hypotheses, we establish relationships between the set of normality, stable manifolds of saddle periodic points, and the support of μ^+. We illustrate this work in the more concrete setting of quadratic polynomial maps of $\mathbb{C}^2$ with merely rational inverses.

1. Introduction

In the wake of recent spectacular advances in understanding the dynamics of rational maps of $\mathbb{P}^1$, researchers have also begun to consider dynamics of holomorphic maps of more than one variable. In particular, polynomial automorphisms of $\mathbb{C}^2$ and holomorphic endomorphisms of $\mathbb{P}^n$ have received considerable attention (see [BS1–3],[FS3],[HP],[HO]). In this paper, we consider the dynamics of birational maps of $\mathbb{P}^2$.

Loosely speaking, a degree d rational self-map f of $\mathbb{P}^2$ is one induced by a degree d homogeneous self-map of $\mathbb{C}^3$ under the canonical projection $\pi : \mathbb{C}^3 \setminus \{0\} \to \mathbb{P}^2$. A rational map f_+ is birational if there exists an algebraic curve V and another rational map f_- such that $f_+ \circ f_- = f_- \circ f_+ = \mathrm{id}$ on $\mathbb{P}^2 \setminus V$. One rather quickly discovers a difficulty in studying dynamics of birational maps. When $\deg f_+ \geq 2$, there always exists a finite *indeterminacy set* I^+ of points where f_+ is ill-defined as a continuous map. Under iteration the problem generally becomes worse; the set of all points where some iterate of f_+ is ill-defined is $I^+_\infty = \cup_{n=0}^\infty f_+^{-n}(I^+)$. Hence,

---

1991 *Mathematics Subject Classification.* Primary: 32 Secondary: 14,31,58.

The work presented in this paper was partially supported by a postdoctoral fellowship from the National Science Foundation

iterates of f_+ act smoothly only on $\mathcal{D}^+ = \mathbb{P}^2 \setminus \mathcal{I}^+$, where $\mathcal{I}^+ = \overline{I_\infty^+}$. We denote corresponding sets for f_- by $I^-, I_\infty^-, \mathcal{I}^-, \mathcal{D}^-$.

So from the beginning, we are left with the question of whether and to what extent a birational map of $\mathbb{P}^2$ constitutes a "smooth dynamical system" in the generally understood sense of the phrase. We do not attempt to answer the question directly in this paper. Rather, we proceed with the heuristic that birational maps of $\mathbb{P}^2$ are natural generalizations of polynomial automorphisms of $\mathbb{C}^2$ (the latter extend to $\mathbb{P}^2$ as birational maps). Since pluripotential theory has proved a successful means of understanding dynamics of automorphisms, our primary concern in this paper is to generalize pluripotential theory tools to the setting of birational maps. The results that we present show a fascinating relationship between the "separation" that a birational map maintains between its forward and backward indeterminacy sets I^+ and I^-, and the degree of control that we are able to attain over the attendant pluripotential theory.

Section 2 contains a detailed summary of notation and basic properties connected with rational self-maps of $\mathbb{P}^2$. In particular, given a degree d rational map f of $\mathbb{P}^2$ induced by a homogeneous map $\tilde{f} : \mathbb{C}^3 \to \mathbb{C}^3$, we define an *escape function* $\tilde{G}$ for f. Namely, for any point $\tilde{p} \in \mathbb{C}^3$, we set

$$\tilde{G}(\tilde{p}) = \lim_{n\to\infty} \frac{1}{d^n} \log ||\tilde{f}^n(\tilde{p})||.$$

It is not hard to show that the limit exists, and that $\tilde{G}$ is either plurisubharmonic or identically $-\infty$. The escape function was first introduced by Hubbard and Papadopol [HuPa] in connection with holomorphic endomorphisms of $\mathbb{P}^n$. They showed that if $\mathcal{N}$ is the largest open set on which iterates of f form a normal family, then $\tilde{G}$ is pluriharmonic on $\pi^{-1}(\mathcal{N})$. Their result, along with its proof, is equally valid for any rational self-map of $\mathbb{P}^n$. In fact, Ueda [Ue] went on to show that when the indeterminacy set is empty $\pi^{-1}(\mathcal{N})$ is *exactly* the set on which $\tilde{G}$ is pluriharmonic. Ueda's result, however, does not obviously generalize to all rational self-maps.

Section 3 provides a more detailed introduction to birational maps of $\mathbb{P}^2$. In the rest of this introduction, $f_+ : \mathbb{P}^2 \to \mathbb{P}^2$ will denote a degree $d \geq 2$ birational map. We say that f_+ is *minimally separating* if $I_\infty^+ \cap I_\infty^- = \emptyset$. After collecting several useful preliminary results about birational maps, we prove Theorem 3.11, which states roughly that a minimally separating birational map cannot shrink volume too quickly. We then use this theorem to prove an existence result:

Theorem 1.1. *If f_+ is minimally separating, then its escape function $\tilde{G}^+$ is plurisubharmonic.*

Previously, the best result in this direction was by Fornaess and Sibony [FS1] who required the existence of a periodic point to deduce plurisubharmonicity of the escape function for a rational map.

Section 4 begins with a discussion of positive closed currents on $\mathbb{P}^2$ and what it means to pull back or push forward such a current by a birational map. This discussion is relevant because the escape function $\tilde{G}^+$ induces a positive closed current μ^+ on $\mathbb{P}^2$ such that $\pi^*\mu^+ = dd^c\tilde{G}^+$. The set $\operatorname{supp}\mu^+$ plays a role roughly analogous to that of the Julia set of a rational map of $\mathbb{P}^1$. Also, μ^+ enjoys the simple transformation property

$$f_+^*\mu^+ = d \cdot \mu^+,$$

which it inherits from a similar property of $\tilde{G}^+$. The currents $f_{+*}\mu^+$, $f_-^*\mu^+$, and $f_{-*}\mu^+$ are more interesting and more complicated. Summarizing results from sections 2, 3, and 4, we obtain

Theorem 1.2. *The following statements about f_+ are equivalent.*

(1) *f_+ is minimally separating.*
(2) *f_- is minimally separating.*
(3) $\deg f_+^n = (\deg f_+)^n$ *for all n.*
(4) $\lim_{n\to\infty}(\deg f_+^n)^{1/n} = \deg f_+$.
(5) *μ^+ has no support concentrated on any algebraic curve.*
(6) *$f_{-*}\mu^+ = d \cdot \mu^+$ (i.e. $f_{-*}\mu^+ = f_+^*\mu^+$).*

Any one of these statements implies that $f_{+*}\mu^+ = \mu^+/d$.

In particular, minimally separating birational maps (which include the the family of polynomial automorphisms of $\mathbb{C}^2$) are those for which degrees of iterates grow maximally. Thus polynomial shears $(x, y) \mapsto (x + b, ay + P(x))$ of $\mathbb{C}^2$ are not minimally separating, since iteration of a shear leaves its degree unchanged. Also in section 4, we prove a theorem about normalized pullbacks of generic positive closed currents. A slightly weaker statement than the one we give there is

Theorem 1.3. *Suppose that f_+ is minimally separating and that T is a non-zero, positive, closed (1,1) current on $\mathbb{P}^2$ such that* $\operatorname{supp} T \cap I_\infty^- = \emptyset$. *Then*

$$\lim_{n\to\infty} \frac{1}{d^n} f_+^{n*}T = c\mu^+$$

for some real constant $c > 0$.

This theorem directly generalizes Theorem 4.7 of [BS1] and Corollary 7.3.4 of [FS3]. Among other things, it strengthens the analogy between $\operatorname{supp}\mu^+$ and the Julia set of a rational map of $\mathbb{P}^1$. That is, one can think of Theorem 1.3 as a "weighted", two variable version of a well-known fact from dynamics on $\mathbb{P}^1$: the closure of the backward orbit of any non-exceptional point contains the Julia set.

The mere fact that $\tilde{G}^+$ is plurisubharmonic does not indicate that $\tilde{G}^+$ enjoys much regularity. So in section 5, we define a slightly stronger separation property which allows us to obtain better regularity. Namely, $\tilde{G}^+$ is *mildly separating* if $\tilde{G}^+(\tilde{p}) > -\infty$ for all $\tilde{p} \in \pi^{-1}(I^-)$. The usefulness of this property (and the justification for its name) is indicated by the following summary of results from section 5.

Theorem 1.4. *The following statements about f_+ are equivalent.*

(1) *f_+ is mildly separating.*
(2) *f_- is mildly separating.*
(3) $\displaystyle\sum_{n=0}^{\infty} \frac{1}{d^n} \log \operatorname{dist}(f_+^n(I^-), I^+) > -\infty.$
(4) *$\pi(\{\tilde{G}^+ = -\infty\})$ contains no algebraic curves.*

Any one of these statements implies that $\tilde{G}^+$ is continuous on $\pi^{-1}(\mathcal{D}^+)$ and that μ^+ is extremal in the cone of positive, closed (1,1) currents on $\mathbb{P}^2$.

The similarities between this result and Theorems 1.1 and 1.2 above are striking. At present we know of no example of a minimally separating birational map that is not mildly separating.

Section 6 introduces two further separation hypotheses for a birational map and explores the implications of each. We say that f_+ is *separating* if $\mathcal{I}^+ \cap \mathcal{I}^- = \emptyset$. We say that f_+ is *completely separating* if $\mathcal{I}^+ \subset \mathcal{N}^-$—i.e. if iterates of f_- form a normal family on a neighborhood of $\mathcal{I}^+$. We are able to show that either of these separation hypotheses implies partial results along the lines of the result of Ueda mentioned above.

Theorem 1.5. *If f_+ is separating, then $\mathcal{D}^+$ is (open and) dense in $\mathbb{P}^2$ and intersects any algebraic curve. The support of μ^+ contains any repelling or saddle periodic point of f_+. Furthermore, if p is a periodic saddle point belonging to $\mathcal{D}^+$, then* $\operatorname{supp} \mu^+$ *contains (the closure of) the stable manifold of p.*

Theorem 1.6. *If f_+ is completely separating, then* $\operatorname{supp} \mu^+$ *is the exact complement of $\mathcal{N}^+$. Any positive, closed (1,1) current T such that $T \neq c\mu^+$ has support which intersects $\mathcal{N}^+$.*

These theorems, too, should be considered in light of the analogy between $\operatorname{supp} \mu^+$ and the Julia set of a one variable rational map. Theorem 1.5 amounts to the statement that repelling periodic points belong to the Julia set, and Theorem 1.6 recalls the definition of the Julia set as the complement of the set of normality.

Finally, in section 7 we consider the class of *polynomial* birational maps in order to provide examples of maps with the properties defined and described in earlier sections. We say that f_+ is polynomial if the restriction of f_+ to $\mathbb{C}^2$ has polynomial coordinate functions. We show that a polynomial map is minimally separating if and only if it is mildly separating. In either case, this leads to the conclusion that f_+ and f_- have well-behaved escape functions. We show that the escape function $\tilde{G}^+$, in particular, shares many of the features of the escape function for a polynomial automorphism (see Section 3 of [BS1] for comparison). We then restrict further to the class of degree 2 polynomial maps. We give various examples of these which show among other things that a birational map can be mildly separating but not separating, and separating but not completely separating. Whereas all the other separation hypotheses turn out to be symmetric in f_+ and f_-, we give an

example which shows that a completely separating birational map need *not* have a completely separating inverse.

We would like to thank Eric Bedford, especially, for many helpful conversations about the work presented in this paper. His knowledge of pluripotential theory and multivariable complex dynamics has been an invaluable resource, generously and patiently offered. We are also indebted to the referee for an exceptionally careful reading of our paper. Many corrections and several changes in presentation resulted directly from his diligence.

2. Rational Maps of $\mathbb{P}^2$ and Escape Functions

In this section we discuss basic ideas and notation associated with rational self-maps of $\mathbb{P}^2$. Most of the results that we give here appear in earlier papers, though Propositions 2.6 and 2.7 do not seem to have previously been stated or explicitly used. The latter play a crucial role in this paper. Much of the discussion in this section applies with little or no modification to higher dimensions, but since we are only concerned with maps of $\mathbb{P}^2$, we will restrict ourselves to the two-dimensional context.

Constants in what follows will be denoted by generic letters (e.g. C) whose numeric value might change from one line to the next. We will add a subscript (e.g. C_1) when we want to indicate that a constant remains *unchanged* from its previous occurence. Where we wish to emphasize that a constant C depends on parameters $a_1, \ldots, a_n$, we will write $C(a_1, \ldots, a_n)$.

Projective Space. Complex projective 2 space is a canonical compactification of $\mathbb{C}^2$. One defines $\mathbb{P}^2$ formally as the quotient of $\mathbb{C}^3 \setminus \{0\}$ by the relation identifying $\tilde{p}$ with $\lambda\tilde{p}$ for any $\tilde{p} \in \mathbb{C}^3 \setminus \{0\}$ and any $\lambda \in \mathbb{C} \setminus \{0\}$. That is, complex lines through the origin in $\mathbb{C}^3$ become points in $\mathbb{P}^2$. We denote the resulting quotient map by $\pi : \mathbb{C}^3 \setminus \{0\} \to \mathbb{P}^2$. To distinguish a point $p \in \mathbb{P}^2$ from one of its representatives in $\mathbb{C}^3$, we will use $\tilde{p}$ to denote the representative. To make the same distinction in coordinate notation, we will write $(x, y, z) \in \mathbb{C}^3$ and use homogeneous coordinates $[x : y : z]$ to denote $\pi(x, y, z) \in \mathbb{P}^2$. Note that homogeneous coordinates are unique only up to multiplication by non-zero complex scalars.

The natural (albeit not unique) inclusion $i : \mathbb{C}^2 \to \mathbb{P}^2$ maps (x, y) to $[x : y : 1]$. One can check that i is injective and that $\mathbb{P}^2 \setminus i(\mathbb{C}^2) = \{[x : y : 0]\}$ is an embedded copy of the Riemann sphere $\mathbb{P}^1$. We refer to this sphere as the *line* ℓ^∞ *at infinity.*

When computing volume or distance, we adopt the Euclidean metric on $\mathbb{C}^2$ and $\mathbb{C}^3$, and we use the Fubini-Study metric on $\mathbb{P}^2$. Regardless, we denote volume measure by dV, induced volume measure on a submanifold by dA, and distance between points by $\mathrm{dist}(\cdot, \cdot)$. We will write $B_E(\epsilon)$ to indicate the set of points whose distance from the set E is less than ϵ. When the underlying metric is not clear from context, we will indicate it explicitly.

The precise form of the Fubini-Study metric is somewhat awkward to apply. To simplify computations, we will use a comparable "chordal" distance where it is

convenient. That is, given any two points $p_1, p_2 \in \mathbb{P}^2$, $\mathrm{dist}(p_1, p_2)$ is comparable to

$$\min_{\substack{\pi(\tilde{p}_j)=p_j \\ ||\tilde{p}_j||=1}} \mathrm{dist}(\tilde{p}_1, \tilde{p}_2).$$

To simplify volume estimates, we will also take advantage of the fact that on compact subsets $K \subset \mathbb{C}^2 \cong i(\mathbb{C}^2) \subset \mathbb{P}^2$, Fubini-Study and Euclidean volume measures are comparable with a comparability constant depending only on K.

An algebraic curve $V \subset \mathbb{P}^2$ is a set of the form $\{\pi(\tilde{p}) : \tilde{P}(\tilde{p}) = 0\}$ where $\tilde{P}$ is a non-constant homogeneous polynomial. We call $\tilde{P}$ a *defining polynomial* for V. If $\tilde{P}$ has the smallest possible degree among defining polynomials for V, then we say that $\tilde{P}$ is *minimal* and that the *algebraic degree* $\deg V$ of V is $\deg \tilde{P}$. Minimal defining polynomials are unique up to constant multiples.

Points in V can be divided into those which are *regular*—i.e. points p such that $V \cap B_p(\epsilon)$ is an embedded complex manifold for $\epsilon > 0$ small enough—and those (finitely many) which are *singular*—i.e. not regular. Any algebraic curve $V \subset \mathbb{P}^2$ admits a *regularization*. That is, there is a compact (though possibly not connected) Riemann surface X without boundary and a holomorphic map $h : X \to \mathbb{P}^2$ such that $h(X) = V$, and if $p \in V$ is regular then $h^{-1}(p)$ is a singleton $\{z\}$ such that dh is non–singular at z.

If $\tilde{P}$ is a minimal defining polynomial for V, and $\tilde{P}$ cannot be written $\tilde{P}_1 \cdot \tilde{P}_2$ for any pair of non-constant homogeneous polynomials, then we say that V is irreducible. This is equivalent to the statement that if N is any finite subset of V, then $V \setminus N$ is connected. It is also equivalent to the statement that V is regularized by a connected Riemann surface.

Rational maps of $\mathbb{P}^2$. Any map of $\mathbb{C}^2$ with rational coordinate functions extends to a *rational map* of $\mathbb{P}^2$. However, treating rational maps as "extended maps of $\mathbb{C}^2$" tends to obscure a global understanding of the maps by forcing a continual distinction between points in $\mathbb{C}^2$ and points not in $\mathbb{C}^2$. It will be more useful for us to understand a rational map as a map induced by a homogeneous map of $\mathbb{C}^3$.

Namely, suppose that $\tilde{f} : \mathbb{C}^3 \to \mathbb{C}^3$ is holomorphic and satisfies $\tilde{f}(\lambda\tilde{p}) = \lambda^d \tilde{f}(\tilde{p})$. It follows quickly from Liouville's theorem that the component functions of $\tilde{f}$ are homogeneous polynomials of degree d, and hence, that d is an integer. Since $\tilde{f}$ maps lines through the origin to lines through the origin, we see that there is a holomorphic map $f : \mathbb{P}^2 \to \mathbb{P}^2$ such that $f \circ \pi = \pi \circ \tilde{f}$. This is not entirely correct, however, because there will in general be a set I of points $p \in \mathbb{P}^2$ such that $\tilde{f}(\tilde{p}) = 0$ for some (thus all) $\tilde{p} \in \pi^{-1}(p)$. At these points f will be ill-defined. I is necessarily an algebraic variety, proper if $\tilde{f}$ is not identically 0. If I has a codimension 1 component V, then it is not hard to see that a minimal defining polynomial $\tilde{P}$ for V will divide all three coordinate functions of $\tilde{f}$. Hence $\tilde{f}/\tilde{P}$ induces the same map f on $\mathbb{P}^2 \setminus V$. After we reduce $\tilde{f}$ by all possible polynomial factors, we obtain a *minimal* homogeneous map inducing f. The *indeterminacy set* I will then be at

most a finite set of points in $\mathbb{P}^2$, and f will not extend continuously past any point in I.

If $\tilde{f}$ is a minimal, degree d, homogeneous map of $\mathbb{C}^3$, we say that the induced map f of $\mathbb{P}^2 \setminus I$ is a *degree d rational* map of $\mathbb{P}^2$. Despite the fact that f is ill-defined on I, we write $f : \mathbb{P}^2 \to \mathbb{P}^2$ for simplicity. We refer to $\tilde{f}$ as a *minimal representative* of f. Minimal representatives are unique up to multiplication by complex scalars; hence the degree d and indeterminacy set I of f are well-defined. We let $\mathcal{C}$ denote the critical set of f, and for convenience we include in $\mathcal{C}$ any limit point lying in I. By considering $D\tilde{f}$, one can show that $\mathcal{C}$ is an algebraic curve of degree at most $3d - 3$.

Given $E \subset \mathbb{P}^2$, we would like to define $f(E)$ and $f^{-1}(E)$ to take account of points of indeterminacy. So we adopt the convention that $f(E)$ consists of $f(E \setminus I)$ along with all points $p \in \mathbb{P}^2$ such that $p = \lim f(q_j)$ where $q_j \in E \setminus I$ and $\lim q_j \in E \cap I$. Next, given a single point $p \in \mathbb{P}^2$, we declare $f^{-1}(p)$ to be the closure in $\mathbb{P}^2$ of $\{q \in \mathbb{P}^2 \setminus I : f(q) = p\}$. We then set

$$f^{-1}(E) = \bigcup_{p \in E} f^{-1}(p).$$

These definitions have the virtue that the image and preimage of a closed set is closed. Furthermore, if $E \cap I = \emptyset$, then $f^{-1}(E), f(E)$ are given by the usual definitions of image and preimage. Nonetheless, one must be careful—for instance, it is not true in general that $f(E_1 \cup E_2) = f(E_1) \cup f(E_2)$ or that $f(f^{-1}(E)) = E$.

Since we are interested in dynamics—that is, iterates of f—we note that the indeterminacy set for f^k is

$$I_k = \bigcup_{j=0}^{k-1} f^{-j}(I).$$

Allowing $k = \infty$ in this definition gives us the set I_∞ of all points p such that $f^j(p)$ is ill-defined for j large enough.

Definition 2.1. *The* **extended indeterminacy set** *of f is $\mathcal{I} = \overline{I_\infty}$. Its complement $\mathcal{D} = \mathbb{P}^2 \setminus \mathcal{I}$ is the* **dynamic domain** *of f.*

Note that the dynamic domain is the largest open subset of $\mathbb{P}^2$ on which all iterates of f are well-defined. Though a rational map cannot be continuously extended across a point of indeterminacy, it is a useful and well-known fact that one can obtain holomorphic extensions "along analytic disks." In the next proposition, $\Delta \subset \mathbb{C}$ is the unit disk.

Proposition 2.2. *Let $f : \mathbb{P}^2 \to \mathbb{P}^2$ be a rational map, and let $h : \Delta \to \mathbb{P}^2$ be a non-constant holomorphic map. Then there is a unique holomorphic map $g : \Delta \to \mathbb{P}^2$ such that $g(z) = f \circ h(z)$ for all z such that $h(z) \notin I$.*

Proof. Since h is non-constant and I is finite, $h^{-1}(I)$ is discrete. We set $g(z) = f \circ h(z)$ for all $z \notin h^{-1}(I)$. Extending g analytically past $h^{-1}(I)$ is a purely local

problem, so we can assume without loss of generality that $h(\Delta) \subset \mathbb{C}^2 = \{[x : y : 1] \in \mathbb{P}^2\}$ and $h^{-1}(I) = \{0\}$. Then we can write $h = (h_1, h_2)$ and lift h to a map $\tilde{h} : \Delta \to \mathbb{C}^3$ by setting $\tilde{h} = (h_1, h_2, 1)$.

Let $\tilde{f} : \mathbb{C}^3 \to \mathbb{C}^3$ be a minimal representative of f, and set $(f_1, f_2, f_3) = \tilde{f} \circ \tilde{h}$. Since $h(0)$ is the only point of indeterminacy in the range of h, we have $\pi(f_1(z), f_2(z), f_3(z)) = g(z)$ for $z \neq 0$ and $\tilde{f}_j(0) = 0$ for $j = 1, 2, 3$. Let k be the minimal order of vanishing of f_1, f_2, f_3 at 0, and let $\tilde{g}(z) = (f_1/z^k, f_2/z^k, f_3/z^k)$. Then we have again that $\pi \circ \tilde{g} = g$ for all $z \neq 0$. However, since $\tilde{g}(0) \in \mathbb{C}^3 \setminus \{0\}$, $\pi \circ \tilde{g}(0)$ is well-defined and extends g analytically past 0. The extension is clearly unique. $\square$

Given a rational map f of $\mathbb{P}^2$ and a non-constant holomorphic map $h : X \to \mathbb{P}^2$ of a Riemann surface X, this proposition allows us to write $f \circ h$ without worrying about possible points of indeterminacy in $h(X)$. It is clear that $(f \circ h)(X) = f(h(X))$. Since any algebraic curve $V \subset \mathbb{P}^2$ admits a regularization $h : X \to V$ it is clear that $f(V) = f \circ h(X)$ is the union of an algebraic curve with a finite number of points corresponding to components of X on which $f \circ h$ is constant. If in addition, V is irreducible, then we conclude that $f(V)$ is either a single point or an irreducible algebraic curve.

While regularizations provide a way to realize the image of an algebraic curve, defining polynomials give an alternate method for realizing a curve's preimage. Namely, suppose that $\tilde{P}$ is a defining polynomial for V, and $\tilde{f}$ is a minimal representative for f. Then it is not hard to verify that $f^{-1}(V)$ is an algebraic curve defined by $\tilde{P} \circ \tilde{f}$.

Given rational maps $f, g : \mathbb{P}^2 \to \mathbb{P}^2$, with minimal representatives $\tilde{f}, \tilde{g}$, it is not necessarily true that $\tilde{f} \circ \tilde{g}$ is a minimal representative of $f \circ g$. The key observation in the next proposition appears in [FS1].

Proposition 2.3. *Let f and g be rational maps of $\mathbb{P}^2$ with minimal representatives $\tilde{f}$ and $\tilde{g}$ and indeterminacy sets I_f and I_g. Then $\tilde{f} \circ \tilde{g}$ is a minimal representative of $f \circ g$ if and only if there exists no algebraic curve V such that $g(V) \subset I_f$. In particular,* $\deg(f \circ g) \leq (\deg f)(\deg g)$ *with equality holding if and only if no such V exists.*

Proof. Suppose that $\tilde{f} \circ \tilde{g}$ is not minimal. Then there is a homogeneous polynomial $\tilde{P} : \mathbb{C}^3 \to \mathbb{C}$, such that $\tilde{P}$ divides all three coordinate functions of $\tilde{f} \circ \tilde{g}$. Let $V \subset \mathbb{P}^2$ be the curve defined by $\tilde{P}$. Then $\tilde{f} \circ \tilde{g}$ vanishes on $\pi^{-1}(V)$ and $\tilde{f}$ vanishes on $\pi^{-1}(g(V))$. We conclude that $g(V) \subset I_f$. Now suppose that V is a curve such that $g(V) \subset I_f$, and hence, that $\tilde{f} \circ \tilde{g}$ vanishes on $\pi^{-1}(V)$. It follows that any minimal defining polynomial for V divides all three coordinate functions of $\tilde{f} \circ \tilde{g}$. We conclude that $\tilde{f} \circ \tilde{g}$ is not minimal.

The final statement of the proposition follows because $\tilde{f} \circ \tilde{g}$ is a representative of $f \circ g$, but $\deg f \circ g$ is defined to be the degree of a *minimal* representative. $\square$

Given a rational map f and an algebraic curve V, Fornaess and Sibony say

that V is a *degree lowering curve* for f if $f^k(V) \subset I$ for some positive integer k. According to the previous proposition, the degree of f^n will be $(\deg f)^n$ for all n if and only if f has no degree lowering curves. John Smillie observed to us that a stronger statement holds:

Corollary 2.4. *Given a rational map f,*

$$L = \lim_{n\to\infty} \frac{1}{n} \log \deg f^n$$

always exists. Furthermore, $L \leq \log \deg f$ with equality if and only if f has no degree lowering curves.

Proof. It is well-known that if $\{a_n\}$ is a sequence with the property $a_{n+m} \leq a_n + a_m$ for all n and m, then $\lim a_n/n$ exists. Clearly $a_n = \log \deg f^n$ satisfies this subadditivity condition. From Proposition 2.3 it is also clear that $\lim(\log \deg f^n)/n = d$ if f has no degree lowering curves.

Now suppose that f has a degree lowering curve whose minimal defining polynomial $\tilde{P}$ has degree j. Then for some k, f^k is divisible by $\tilde{P}$, and $\deg f^k \leq (\deg f)^k - j$. Thus we have

$$\lim_{n\to\infty} \frac{1}{n} \log \deg f^n = \lim_{n\to\infty} \frac{1}{nk} \log \deg f^{nk} \leq \frac{1}{k} \log((\deg f)^k - j) < \deg f.$$

□

In light of Corollary 2.4, the next definition becomes meaningful.

Definition 2.5. *The* **dynamic degree** *of a rational map $f : \mathbb{P}^2 \to \mathbb{P}^2$ is*

$$\lim_{n\to\infty} (\deg f^n)^{1/n}.$$

We say that f has **maximal dynamic degree** *if the degree of f and the dynamic degree of f are the same.*

Thus, f has maximal dynamic degree if and only if f has no degree-lowering curves.

Expansion Functions and Escape Functions.

Fix a degree $d \geq 2$ rational map $f : \mathbb{P}^2 \to \mathbb{P}^2$ with indeterminacy set I. Choose a minimal representative $\tilde{f} : \mathbb{C}^3 \to \mathbb{C}^3$ of f, scaled for convenience so that $\max_{||\tilde{p}||=1} ||\tilde{f}(\tilde{p})|| = 1$. We define the *expansion function* $\Gamma : \mathbb{P}^2 \to [-\infty, 0]$ for for f by

$$\Gamma(p) = \log \frac{||\tilde{f}(\tilde{p})||}{||\tilde{p}||^d}.$$

Homegeneity of $\tilde{f}$ guarantees that this definition does not depend on which representative $\tilde{p} \in \pi^{-1}(p)$ we choose on the righthand side. The dynamical relevance of Γ appears in the next two propositions. The first relates Γ to the rate at which f can separate points. The second relates Γ to the distance of a point from the indeterminacy set.

Proposition 2.6. *There is a constant C such that for any $p_1, p_2 \in \mathbb{P}^2 \setminus I$.*

$$\operatorname{dist}(f(p_1), f(p_2)) \leq Ce^{-\Gamma} \operatorname{dist}(p_1, p_2),$$

where

$$\Gamma = \min\{\Gamma(p_1), \Gamma(p_2)\}.$$

In particular,

$$||Df_p|| \leq Ce^{-\Gamma(p)}.$$

Proof. We will use the comparability of Fubini-Study and chordal distance to prove the first estimate. The second estimate is just an infinitesimal version of the first.

Pick $\tilde{p}_j \in \pi^{-1}(p_j)$, $j = 1, 2$, such that $||\tilde{p}_j|| = 1$ and $\operatorname{dist}(\tilde{p}_1, \tilde{p}_2) \leq C_1 \operatorname{dist}(p_1, p_2)$. We have

$$\begin{aligned}
\operatorname{dist}(f(p_1), f(p_2)) &\leq C \operatorname{dist}\left(\frac{\tilde{f}(\tilde{p}_1)}{||\tilde{f}(\tilde{p}_1)||}, \frac{\tilde{f}(\tilde{p}_2)}{||\tilde{f}(\tilde{p}_2)||}\right) \\
&\leq C \frac{\operatorname{dist}(\tilde{f}(\tilde{p}_1), \tilde{f}(\tilde{p}_2))}{\min\{\tilde{f}(\tilde{p}_1), \tilde{f}(\tilde{p}_2)\}} \\
&\leq C_2 e^{-\Gamma} \operatorname{dist}(\tilde{f}(\tilde{p}_1), \tilde{f}(\tilde{p}_2)).
\end{aligned}$$

But the map $\tilde{f}$ is uniformly Lipschitz on compact subsets of $\mathbb{C}^3$. Therefore,

$$\operatorname{dist}(\tilde{f}(\tilde{p}_1), \tilde{f}(\tilde{p}_2)) \leq C_3 \operatorname{dist}(\tilde{p}_1, \tilde{p}_2),$$

and

$$\operatorname{dist}(f(p_1), f(p_2)) \leq C_1 C_2 C_3 e^{-\Gamma} \operatorname{dist}(p_1, p_2),$$

as desired. □

Proposition 2.7. *There are constants C_1, C_2, C_3, C_4 depending only on f such that*

$$C_1 + C_2 \log(\operatorname{dist}(p, I)) \leq \Gamma(p) \leq C_3 + C_4 \log(\operatorname{dist}(p, I)).$$

In particular, Γ is bounded below on any compact subset of $\mathbb{P}^2 \setminus I$.

Proof. By our initial assumption about $\tilde{f}$ and homogeneity we have $||\tilde{f}(\tilde{p})||/||\tilde{p}||^d \leq 1$ on $\mathbb{C}^3$. If I is non-empty, then no corresponding uniform lower bound is possible. However, we can take advantage of the fact that the coordinate functions of $\tilde{f}$ are polynomials whose simultaneous zeroes coincide with $\pi^{-1}(I)$ to conclude that

$$A_1 \operatorname{dist}(p, I)^{k_1} \leq \frac{||\tilde{f}(\tilde{p})||}{||\tilde{p}||^d} \leq A_2 \operatorname{dist}(p, I)^{k_2}$$

for points p near I. Furthermore, we will have a uniform lower bound for the middle term on compact subsets of $\mathbb{P}^2 \setminus I$. The proposition follows from these observations. $\square$

When the degree of f is at least two, one can define plurisubharmonic functions closely related to Γ,

$$\tilde{G}_n(\tilde{p}) = \frac{1}{d^n} \log ||\tilde{f}^n(\tilde{p})|| = \log ||\tilde{p}|| + \sum_{j=0}^{n-1} \frac{1}{d^j} \Gamma \circ f^j(p),$$

which track the rate at which points $\tilde{p} \in \mathbb{C}^3$ tend to zero or infinity under iteration of $\tilde{f}$. Clearly, $\tilde{G}_n(\tilde{p}) \leq \log ||\tilde{p}||$.

Corollary 2.8. *Given $\tilde{p} \in \mathbb{C}^3 \setminus \{0\}$ and $p = \pi(\tilde{p})$, the following are equivalent:*

(1) $\lim_{n\to\infty} \tilde{G}_n(\tilde{p})$ *exists and is finite;*
(2) $\sum_{n=0}^{\infty} \frac{\Gamma \circ f^n(p)}{d^n}$ *converges;*
(3) $\sum_{n=0}^{\infty} \frac{1}{d^n} \log \operatorname{dist}(f^n(p), I)$ *converges.*

Pointwise convergence at $\tilde{p}$ and p can be replaced by uniform convergence or $\mathrm{L}^1_{\mathrm{loc}}$ convergence on U and $\pi(U)$ for any measurable $U \subset \mathbb{C}^3$.

It turns out that if one allows for the possibility that $\tilde{G}_n \to -\infty$, then it is not hard to show that $\tilde{G}_n$ converges. The following result appears in varying degrees of generality in [FS1], [HP], [Ue].

Theorem 2.9. *Either $\tilde{G}_n$ tends uniformly to $-\infty$ on compact subsets of $\mathbb{C}^3$, or there exists a plurisubharmonic function $\tilde{G} : \mathbb{C}^3 \to \mathbb{R} \cup \{-\infty\}$ such that $\tilde{G}_n$ converges to $\tilde{G}$ pointwise and in $\mathrm{L}^1_{\mathrm{loc}}(\mathbb{C}^3)$. If the second possibility occurs, then $\tilde{G}$ satisfies*

(1) $\tilde{G}(\lambda\tilde{p}) = \tilde{G}(\tilde{p}) + \log|\lambda|$;
(2) $\tilde{G} \circ \tilde{f}(\tilde{p}) = d \cdot \tilde{G}(\tilde{p})$;
(3) $\tilde{G}(\tilde{p}) = -\infty$ *if* $\pi(\tilde{p}) \in I_\infty$ *(in particular, poles of $\tilde{G}$ are dense in $\mathcal{I}$).*

Proof. By the definition of Γ, all terms in the sum defining $\tilde{G}_n$ are non-positive. Thus $\{\tilde{G}_n\}_{n=0}^{\infty}$ is a decreasing sequence of plurisubharmonic functions. By well-known results (see [Ho] Theorem 4.1.9) about such sequences, it follows that $\tilde{G}_n$ tends uniformly on compacts to $-\infty$, or $\tilde{G}_n$ tends pointwise and in $\mathrm{L}^1_{\mathrm{loc}}$ to a plurisubharmonic limit function. The properties (1)–(3) of $\tilde{G}$ all derive from analogous properties of the functions $\tilde{G}_n$, and these analogous properties follow directly from the definition of $\tilde{G}_n$. $\square$

Definition 2.10. *The* **escape function** $\tilde{G} : \mathbb{C}^3 \to \mathbb{R} \cup \{-\infty\}$ *for a rational map* $\tilde{f}$ *is given by*

$$\tilde{G}(\tilde{p}) = \lim_{n\to\infty} \tilde{G}_n(\tilde{p})$$

for all $\tilde{p} \in \mathbb{C}^3$*. If* $\tilde{G}$ *is identically* $-\infty$*, then we say that* $\tilde{G}$ *is* **degenerate***.*

According to the theorem above, $\tilde{G}$ is either plurisubharmonic or degenerate. It should be emphasized, however, that a priori $\tilde{G}$ might be plurisubharmonic but still quite badly behaved. For instance, it is possible for a plurisubharmonic function to have a dense set of poles.

3. Birational Maps of $\mathbb{P}^2$

The only biholomorphisms of $\mathbb{P}^2$ are rational maps of degree one—the two-dimensional analogue of linear fractional transformations of the Riemann sphere. However, there is a weaker notion of invertibility that includes many dynamically interesting rational maps.

Definition 3.1. *A rational map* $f_+ : \mathbb{P}^2 \to \mathbb{P}^2$ *is* **birational** *if there exists another rational map* f_- *and an algebraic curve* V *such that* $f_+ \circ f_- = f_- \circ f_+ = id$ *on* $\mathbb{P}^2 \setminus V$*.*

For example, any polynomial automorphism of $\mathbb{C}^2$ extends to a birational map of $\mathbb{P}^2$ with $V = \ell_\infty$.

Throughout the rest of this paper, unless otherwise noted, $f_+ : \mathbb{P}^2 \to \mathbb{P}^2$ will denote a birational map with degree at least two and minimal representative $\tilde{f}_+ : \mathbb{C}^3 \to \mathbb{C}^3$. f_- will denote the birational inverse of f_+ and $\tilde{f}_- : \mathbb{C}^3 \to \mathbb{C}^3$ will be its minimal representative. We will preserve from the last section the notation used to denote indeterminacy sets, critical sets, escape functions, etc., using $+$ and $-$ superscripts to distinguish objects corresponding to f_+ from objects corresponding to f_-. For instance, I^+ is the indeterminacy set of f_+, and $\tilde{G}^-$ is the escape funtion for f_-.

An example of a degree two birational map is given in homogeneous coordinates by

$$q_+[x : y : z] = [yz : zx : xy].$$

The critical set $\mathcal{C}^+$ of q consists of the three lines $\{[x : y : 0]\}, \{[x : 0 : z]\}, \{[0 : y : z]\}$, which are mapped to the points $[0 : 0 : 1], [0 : 1 : 0], [1 : 0 : 0]$, respectively; these image points, in turn, comprise the indeterminacy set I^- for q_-. Since $q_- = q_+$, these observations apply with $+$ and $-$ signs reversed. We will refer to q_+ as the *standard quadratic transformation*. The following structure theorem for birational maps of $\mathbb{P}^2$ is well-known [GH].

Theorem 3.2. *Any birational map* $f_+ : \mathbb{P}^2 \to \mathbb{P}^2$ *can be decomposed into biholomorphisms and standard quadratic transformations. That is,* f_+ *can be written*

$$A_0 \circ q_+ \circ A_1 \circ q_+ \circ \cdots \circ q_+ \circ A_n$$

where q_+ is as above and $A_0, \ldots, A_n$ are biholomorphisms.

It follows immediately from this theorem that a birational map and its inverse have the same degree. Note that even if n is chosen as small as possible in the above decomposition it does not necessarily follow that $\deg f_+ = 2^n$. This is easily seen by pointing out that there are degree three birational maps of $\mathbb{P}^2$ (even in fact polynomial automorphisms of $\mathbb{C}^2$).

Theorem 3.2 and the properties of q_+ suggest strong and useful relationships between critical sets and indeterminacy sets of birational maps. We summarize these in the following proposition. It would appear, however, that Theorem 3.2 does not provide an easy means of proving the proposition.

Proposition 3.3. *The following statements are true for any birational map f_+.*

(1) *Given any irreducible curve $V \subset \mathcal{C}^+$, $f_+(V)$ is a single point in I^-; likewise, given any $p^- \in I^-$, $f_+^{-1}(p^-)$ is a component of $\mathcal{C}^+$.*
(2) *$I^+ \subset \mathcal{C}^+$, and every irreducible component of $\mathcal{C}^+$ contains a point of I^+.*
(3) *$f_+ : \mathbb{P}^2 \setminus \mathcal{C}^+ \to \mathbb{P}^2 \setminus \mathcal{C}^-$ is a biholomorphism.*

Proof. (1): Let $V \subset \mathcal{C}^+$ be an irreducible curve. Suppose that $f_+(V) \not\subset I^-$. Then there is a point $p \in V \setminus I^+$ such that $f_+(p) \notin I^-$. By continuity there is a small neighborhood $N \subset \mathbb{P}^2 \setminus I^+$ of p such that $f_+(N) \cap I^- = \emptyset$. Hence $f_- \circ f_+$ is well–defined on N, and since this composition is the identity on a dense open set in $\mathbb{P}^2$, it is the identity on N. Differentiating the identity $f_- \circ f_+ = \mathrm{id}$ shows that Df cannot be singular at p. This contradiction shows that $f_+(V) \subset I^+$. From the discussion following Proposition 2.2, we know that $f_+(V)$ is in fact a single point.

Now pick $p^- \in I^-$, and choose ϵ so that the ball $B_{p^-}(\epsilon)$ of radius ϵ about p^- contains no other point of I^-. Set

$$K = \bigcap_{r<\epsilon} f_+^{-1}(\,\overline{B_{p^-}(r)}\,).$$

Then K is a non-empty, connected, compact set, and by continuity $f_+(K) = \{p^-\}$. If K consists of a single point, then f_- extends continuously to p^-, contradicting our assumption that $p^- \in I^-$. Thus K is an infinite, connected set. It is not hard to see now that $K \subset \mathcal{C}^+$. I^+ is finite, so $K \cap (\mathcal{C}^+ \setminus I^+)$ is non-empty.

(2): By (1) f_+ maps any irreducible component of $\mathcal{C}^+$ to a single point. Fornaess and Sibony ([FS1] Proposition 1.2) show that any such component must contain a point of indeterminacy. This gives us the second half of the statement. Now pick $p^+ \in I^+$. Then there exists a curve $V^- \subset \mathcal{C}^-$ such that $f_+(V^-) = p^+$. We choose $p^- \in V^- \cap I^-$ and define K as above. Clearly, $p^+ \in K$. Since $K \subset \mathcal{C}^+$, we are done.

(3): From (1) and (2) applied to f_+ and f_-, we see that f_+ maps $\mathbb{P}^2 \setminus \mathcal{C}^+$ into $\mathbb{P}^2 \setminus \mathcal{C}^-$ and vice versa. Since $f_+ \circ f_-$ and $f_- \circ f_+$ are both the identity map wherever they are defined, the statement follows. $\square$

Corollary 3.4. *The following statements are equivalent:*

(1) f_+ *has maximal dynamic degree;*
(2) f_- *has maximal dynamic degree;*
(3) $I^+_\infty \cap I^-_\infty = \emptyset$*;*
(4) $I^+ \cap I^-_\infty = \emptyset$*.*

Proof. The only varieties V that are mapped to points are subsets of $\mathcal{C}^+$. Since $f_+(\mathcal{C}^+) = I^-$, a degree lowering curve for f_+ exists if and only if $f_+^k(I^-)$ intersects I^+ for some $k \geq 0$. Hence (1) is equivalent to (4).

Since $I^+ \subset I^+_\infty$, (3) implies (4). On the other hand, suppose that $p \in I^+_\infty \cap I^-_\infty$ does not belong to I^+. Then $f_+(p)$ is also a point belonging to $I^+_\infty \cap I^-_\infty$, and we have that either $f_+(p) \in I^+$ or $f_+^2(p)$ is a point lying in $I^+_\infty \cap I^-_\infty$. Repeating this procedure and using the the fact that $p \in I^+_\infty$, we see that there is a finite number k such that $f_+^k(p) \in I^+ \cap I^-_\infty$. Hence, (4) implies (3).

By symmetry (3) is also equivalent to $I^- \cap I^+_\infty = \emptyset$. We conclude that (3) is equivalent to (2). $\square$

Definition 3.5. *We say that a birational map* f_+ *is* **minimally separating** *if* $I^+_\infty \cap I^-_\infty = \emptyset$*.*

Eric Bedford pointed out to us that one can rephrase this definition by saying that at each point $p \in \mathbb{P}^2$, either $f_+^k(p)$ is well-defined for all k, or $f_-^k(p)$ is well-defined for all k.

The composition of $\tilde{f}_+$ with $\tilde{f}_-$ always fails to be minimal, as the next corollary shows.

Corollary 3.6. *We have*

$$\tilde{f}_- \circ \tilde{f}_+ = \tilde{P}^+ \cdot id,$$

where $\tilde{P}^+$ *is a defining polynomial of degree* $d^2 - 1$ *for* $\mathcal{C}^+$*.*

We remark that in general, $\tilde{P}^+$ is not a *minimal* defining polynomial for $\mathcal{C}^+$, because $\mathcal{C}^+$ has degree at most $3d - 3$.

Proof. We only need to show that $\tilde{P}^+$ defines $\mathcal{C}^+$. Zeroes of $\tilde{P}^+$ are either points $\tilde{p}$ such that $\pi(\tilde{p}) \in I^+$ or points $\tilde{p}$ such that $f_+(\pi(\tilde{p})) \in I^-$. By (1) and (2) of Proposition 3.3, the set of such points is exactly $\mathcal{C}^+$. $\square$

The next corollary of Proposition 3.3 amounts to a refined special case of Proposition 4.4 below.

Corollary 3.7. *Let* $V \subset \mathbb{P}^2$ *be an irreducible algebraic curve. Then either* $V \subset \mathcal{C}^+$ *and* $f_+(V)$ *is a single point, or* $f_+(V)$ *is an irreducible algebraic curve such that* $f_-(f_+(V)) = V$*. Furthermore,*

(1) $f_+(V) \subset f_-^{-1}(V)$*;*
(2) $\deg f_+(V) \leq (\deg f_+)(\deg V)$.

Equality holds in either of these statements if and only if $V \cap I^+ = \emptyset$. If f_+ is minimally separating, then a generic complex line L in $\mathbb{P}^2$ satisfies

$$\lim_{n\to\infty} \frac{1}{n} \log \deg f_+^n(L) = d.$$

Proof. If $V \subset \mathcal{C}^+$, then V contains a point $p^+ \in I^+$. Thus $f_+(V)$ is a single point on the curve $f_-^{-1}(p^+) \subset f_-^{-1}(V)$, and the corollary follows. So for the rest of the proof, assume $V \not\subset \mathcal{C}^+$. (3) of Proposition 3.3 implies that $f_+(V)$ is a non-trivial algebraic curve, and $f_+(V) \cap \mathcal{C}^-$ is finite. Thus, $f_-(f_+(V)) = \overline{f_-(f_+(V) \setminus \mathcal{C}^-)} \subset V$ is also an algebraic curve. Since V is irreducible, it must be that $f_-(f_+(V)) = V$.

Let $\tilde{P}$ be a minimal defining polynomial for V. If $p \in V \setminus I^+$ and $\pi(\tilde{p}) = p$, then

$$\tilde{P} \circ \tilde{f}_- \circ \tilde{f}_+(\tilde{p}) = \tilde{P}(\tilde{P}^+(\tilde{p}) \cdot \tilde{p}) = 0.$$

Hence $f_+(p) \in V' = f_-^{-1}(V)$, because V' is defined by $\tilde{P} \circ \tilde{f}_-$. Because $f_+(V \setminus \mathcal{C}^+)$ is dense in $f_+(V)$, (1) follows. (2) follows from the chain of inequalities

$$\deg f_+(V) \le \deg V' \le (\deg \tilde{P})(\deg \tilde{f}_-) = (\deg V)(\deg f_+).$$

Furthermore, if $V \cap I^+$ is non–empty, then (1) of Proposition 3.3 implies that V' contains some component of $\mathcal{C}^-$. Since $f_+(V) \cap \mathcal{C}^-$ is finite, we conclude that equality fails in both (1) and (2).

Now assume that $V \cap I^+$ is empty. Then it is not hard to see that $V' \cap \mathcal{C}^-$ is finite. Thus

$$V' = \overline{V' \setminus \mathcal{C}^-} = \overline{f_+(V) \setminus \mathcal{C}^-} = f_+(V).$$

And if $\deg V' < (\deg \tilde{f}_-)(\deg \tilde{P})$, then by irreducibily of $V' = f_+(V)$, we can write $\tilde{P} \circ \tilde{f}_- = (\tilde{P}')^k$ for some homogeneous polynomial $\tilde{P}'$ and $k \ge 2$. As a result,

$$(\tilde{P}' \circ \tilde{f}_+)^k = \tilde{P} \circ \tilde{f}_- \circ \tilde{f}_+ = (\tilde{P}^+)^{\deg \tilde{P}} \tilde{P}.$$

But $\tilde{P}$ is irreducible and does not divide $\tilde{P}^+$ (since $V \not\subset \mathcal{C}^+$), so this equation is impossible. We conclude that equality holds in (1) and (2).

To get the last statement of the Corollary, note that I_∞^+ is countable. Fix any $p \notin I_\infty^+$. As the number of complex directions at p is uncountable, we see that almost every line through p misses I_∞^+. $\square$

Periodic Critical Points. Extrapolating from the behavior of maps of a single complex variable, one might suppose that a periodic point $p \in \mathbb{P}^2$ is superattracting for a rational map only if the map is generically $n > 1$ to 1 in a neighborhood of p. However, even though a birational map is generically injective, it is possible for such maps to have superattracting periodic points. For example, it is not hard to see that $[0 : 1 : 0]$ is superattracting for the Hénon map

$$f_+ : [x : y : z] \mapsto [yz : axz + y^2 + cz^2 : z^2],$$

in the sense that for points p close enough to $[0:1:0]$,

$$\operatorname{dist}(f_+^n(p),[0:1:0]) \leq (C \operatorname{dist}(p,[0:1:0]))^{2^n}.$$

It is clear, however, that a superattracting periodic point must lie in the critical set. In order to understand the way in which a birational map can shrink volume, we will need some understanding of such points

Definition 3.8. *A* **periodic critical point** *for a birational map f_+ is a point $p \in \mathcal{C}^+$ such that $f_+^k(p) = p$ for some k. We say that p is* **superattracting** *if both eigenvalues of Df_+^k vanish at p. Otherwise, we say that p is an* **ordinary** *periodic critical point.*

Though it will not be of much importance here, it makes sense to subdivide ordinary periodic critical points into those which are *attracting, semi-repelling, and semi-indifferent* according to whether the non-zero eigenvalue of Df_+^k at p is less than one, greater than one, or equal to one in magnitude.

Proposition 3.9. *A birational map f_+ of degree d has at most $3d-3$ periodic critical points. Given any periodic critical point p, we have $f_+(p) \in I^-$. Finally, the orbit of a point p intersects the orbit of a periodic critical point if and only if the forward orbit of p intersects $\mathcal{C}^+ \setminus I^+$ at least $3d-2$ times. In fact, the jth intersection of the orbit of p with $\mathcal{C}^+$ is a periodic critical point for all $j \geq 3d-2$.*

Proof. If V is an irreducible component of $\mathcal{C}^+$, then $f_+(V)$ is a single point. Hence V contains at most one periodic critical point. The algebraic degree of $\mathcal{C}^+$ is $3d-3$, so there are at most $3d-3$ irreducible components of $\mathcal{C}^+$ and, hence, at most $3d-3$ periodic critical points. Also, $f_+(\mathcal{C}^+) = I^-$ implies that the image of any periodic critical point lies in I^-.

It is clear that if the orbit of a point p intersects the orbit of a periodic critical point, then the two orbits coincide thereafter. It follows that the orbit of p passes through $\mathcal{C}^+ \setminus I^+$ (whenever it returns to the periodic critical point) infinitely many times. On the other hand, if the orbit of p passes through $\mathcal{C}^+ \setminus I^+$ $3d-2$ times, it must be the case that $f_+^k(p) \in V \setminus I^+$ for some irreducible component $V \subset \mathcal{C}^+$, and at least two distinct values $k_1 < k_2$ of k. Hence, $f_+^{k_1+1}(p) = f_+^{k_2+1}(p)$. This implies that $f_+^{k_2}(p)$ is a periodic critical point. $\square$

Proposition 3.10. *If p is an ordinary periodic critical point of period k, then p is a regular point of $\mathcal{C}^+$, and there exists a constant $C > 0$ and a neighborhood N of p such that for any $p' \in N$*

$$\operatorname{dist}(f_+^k(p'),\mathcal{C}^+) \geq C \operatorname{dist}(p',\mathcal{C}^+).$$

Proof. After change of coordinates, we may suppose that p is the origin in $\mathbb{C}^2$ and that $(0,1)$ is an eigenvector of $Df_+^k(p)$ with non-zero eigenvalue. If we write

$f_+^k = (f_1, f_2)$, we can apply the Weierstrass preparation theorem in a neighborhood N of $(0,0)$ to obtain $f_2(x,y) = A(x,y)(y + B(x))$, where A and B are holomorphic functions, A does not vanish near $(0,0)$, and $B(0) = 0$. It follows that near $(0,0)$, $\mathcal{C}^+$ is simply the graph of $y = -B(x)$. In particular, $(0,0)$ is a regular point of $\mathcal{C}^+$.

After another local change of coordinates, we may suppose that $\mathcal{C}^+ \cap N$ is the x–axis; that is, $B(x) = 0$. If we restrict N slightly, we will have a constant $C > 0$ such that $|A| > C$ on N. Then for all $(x,y) \in N$, we will have

$$d(f_+^k(x,y), \mathcal{C}^+) = |f_2(x,y)| \geq C|y| \geq C \operatorname{dist}((x,y), \mathcal{C}^+)$$

as desired. $\square$

If p is attracting or superattracting, it follows that f_+^k is a contraction mapping on a neighborhood of p. We define the *immediate basin* of p by

$$\mathcal{B}_p^0 = \{p' \in \mathbb{P}^2 \setminus I_\infty^+ : \lim_{n\to\infty} f_+^{kn}(p') = p\},$$

and the *(full) basin* of p by

$$\mathcal{B}_p = \{p' \in \mathbb{P}^2 \setminus I_\infty^+ : f_+^j(p) \in B_p^0 \text{ for some } j\}.$$

It is clear that orbits of points in $\mathcal{B}_p$ converge uniformly on compact subsets to the cyclic orbit of p.

Birational Maps and Volume. We wish to prove

Theorem 3.11. *Let f_+ be a minimally separating birational map of $\mathbb{P}^2$, and fix $\lambda > 1$. Then there are constants C_1, C_2 depending only on f_+ and λ such that*

$$\text{Volume } f_+^n(E) \geq (C_1 \text{ Volume } E)^{C_2 \lambda^n}$$

for all $n \geq 0$, and every measurable $E \subset \mathbb{P}^2$ which does not intersect the basin of any superattracting point.

This theorem will be applied below to show that the escape function for a minimally separating birational map is plurisubharmonic and to prove a convergence theorem for pullbacks of positive closed currents on $\mathbb{P}^2$.

The idea behind the proof of the theorem goes as follows: if E avoids the critical set by a non-zero distance under iteration of f_+, then its volume shrinks by at most a constant factor during each iterate. On the other hand, if there is a part $E_0 \subset E$ such that $f_+^j(E_0)$ lies too near the critical set for too many j, then by Proposition 3.9, the orbit of E_0 shadows that of a critical periodic point. In this case, one can use estimates in local coordinates to bound the volume of $f_+^j(E_0)$ from below. Making this idea precise requires a number of preliminary results. The first two are independent of whether or not E intersects a superattracting basin.

Lemma 3.12. *Let q_+ be the standard quadratic transformation. There exists a constant C such that for any measurable $E \subset \mathbb{P}^2$,*

$$\text{Volume } q_+(E) \geq C(\text{Volume } E)^2.$$

Proof. The sets of points $[x : y : z] \in \mathbb{P}^2$ where $|x| = |y|$, $|y| = |z|$, and $|z| = |x|$, respectively, divide $\mathbb{P}^2$ into six connected regions. The map q_+ permutes these regions among each other. We may without loss of generality suppose that our measurable set E lies entirely inside one of the six regions. After permuting coordinates in the domain and range (this amounts to changing domain and range by isometries of $\mathbb{P}^2$), we may suppose that $E, q_+(E) \subset K = \{(x, y) \in \mathbb{C}^2 : |y| < |x| < 1\}$ and that in the new coordinates q_+ becomes

$$q'_+[x : y : z] = [yz : xy : xz].$$

As a map on $\mathbb{C}^2$, $q'_+(x, y) = (y/x, y)$. Since K is relatively compact in $\mathbb{C}^2$, it is enough to prove the lemma for Euclidean volume. One can check that at points in K, the magnitude of the complex Jacobian determinant J of q'_+ is bounded below by $|y|$. If $E_t = \{(x, y) \in E : |y|^2 > t\}$, then we have Volume $E_t \geq \max\{-Ct + \text{Volume } E, 0\}$. Thus,

$$\begin{aligned}\text{Volume } q'_+(E) = \int_E |J|^2\, dV \geq \int_E |y|^2\, dV = \int_0^1 \text{Volume } E_t\, dt \\ \geq \int_0^{C^{-1}\,\text{Volume } E} (-Ct + \text{Volume } E)\, dt = C(\text{Volume } E)^2.\end{aligned}$$

□

Corollary 3.13. *Let f_+ be any birational map of $\mathbb{P}^2$. Then there exist constants C, k such that for any measurable $E \subset \mathbb{P}^2$*

$$\text{Volume } f_+(E) \geq C(\text{Volume } E)^k.$$

Proof. Recall that f_+ factors as $A_0 \circ q_+ \circ A_1 \circ \cdots \circ q_+ \circ A_n$, where q_+ is as above and A_j is a biholomorphism. Since $\mathbb{P}^2$ is compact, we will have constants $C_j > 0$ such that Volume $A_j(E) \geq C_j$ Volume E for all measurable sets E. Combining this with the previous lemma gives

$$\text{Volume } f_+(E) \geq C(\text{Volume } E)^{2^n}.$$

Taking $k = 2^n$ completes the proof. □

The next lemma estimates the rate at which f_+ can shrink the volume of a set which lies close to a periodic critical point.

Lemma 3.14. *Let p be an ordinary n–periodic critical point of f_+. There is a neighborhood N of p with the following property. Suppose that $E \subset N$ is measurable and $f_+^{jn}(E) \subset N$ for all $1 \le j \le k$. Then*

$$\text{Volume } f_+^{kn}(E) \ge C_1(\text{Volume } E)^{C_2 k}.$$

C_1 depends on k but not on E, and C_2 is independent of both k and E.

Proof. By passing to an iterate of f_+, we may suppose that p is a fixed periodic point. Since p is a regular point of $\mathcal{C}^+$, we can choose local coordinates so that $p = (0,0)$, and $\mathcal{C}^+ \cap B_p(\epsilon) = \{(x,0) \in B_p(\epsilon)\}$. On $N = B_p(\epsilon)$, Euclidean volume will be comparable to Fubini-Study volume, so we obtain the estimate for Euclidean volume. We choose ϵ small enough so that Proposition 3.10 applies; i.e. there exists a constant $C > 0$ such that if $(x_1, y_1) = f_+(x_0, y_0)$, then $|y_1| \ge C|y_0|$. We can also assume that the complex Jacobian J of f_+ satisfies

$$|J(x,y)| \ge C|y|^m.$$

Given $(x_0, y_0) \in E$, we have by assumption that $(x_j, y_j) = f_+^j(x_0, y_0) \in B_p(\epsilon)$ for all $1 \le j \le k$. Thus, $|y_j| \ge C(k)|y_0|$ for all $j \le k$. Furthermore, if J_k denotes the complex Jacobian of f_+^k, we have

$$\begin{aligned} |J_k(x_0,y_0)| &= |J(x_0,y_0)| \cdot \ldots \cdot |J(x_{k-1},x_{k-1})| \\ &\ge C(k)|y_0|^m \cdot \ldots \cdot |y_{k-1}|^m \ge C(k)|y_0|^{mk}. \end{aligned}$$

Now let $E_t = \{(x,y) \in E : |J_k(x,y)|^2 \le t\}$. Then our estimate on $|J_k|$ tells us that Volume $E_t \le C_0(k)t^{1/mk}$. We use this to obtain

$$\begin{aligned} \text{Volume } f_+^k(E) &= \int_E |J_k|^2 \, dV \\ &= \int_0^\infty (\text{Volume } E - \text{Volume } E_t) \, dt \\ &\ge \int_0^{(\frac{\text{Volume } E}{C_0})^{mk}} (\text{Volume } E - C_0 t^{1/mk}) \, dt \\ &= C(k)(\text{Volume } E)^{1+mk} \ge C_1(k)(\text{Volume } E)^{C_2 k}. \end{aligned}$$

□

Let $\Omega : \mathbb{P}^2 \setminus I^+ \to \mathbb{R}^+ \cup \{0\}$ be defined by $f_+^* \, dV = \Omega \, dV$, where dV is the Fubini-Study volume element. Set

$$\omega(\delta) = \{p \in \mathbb{P}^2 \setminus I^+ : \Omega(p) < \delta\}$$

The following lemma gives us control over the size and location of $\omega(\delta)$ when δ is small.

Lemma 3.15. *Given $\epsilon > 0$, there exists $\delta > 0$ such that*

(1) $\omega(\delta) \subset B_{\mathcal{C}^+}(\epsilon)$;
(2) $f_+(\omega(\delta)) \subset B_{I^-}(\epsilon)$.

Proof. The first item follows from the facts that Ω is continuous, Ω vanishes only on $\mathcal{C}^+$, and $\mathbb{P}^2 \setminus B_{\mathcal{C}^+}(\epsilon)$ is compact in $\mathbb{P}^2 \setminus I^+$. That is, we set $\delta = \min_{\mathbb{P}^2 \setminus B_{\mathcal{C}^+}(\epsilon)} \Omega(p)$.

To obtain the second item we observe that if $p' = f_+(p) \notin B_{I^-}(\epsilon)$,

$$\begin{aligned}
\Omega(p) &= \lim_{r \to 0} \frac{\text{Volume } B_{p'}(r)}{\text{Volume } f_-(B_{p'}(r))} \\
&\geq \lim_{r \to 0} \left(\frac{r}{\max\limits_{p'' \in B_{p'}(r)} \text{dist}(f_-(p''), p)} \right)^4 \\
&\geq \frac{1}{||Df_-(p')||^4} \geq e^{4\Gamma^-(p')} \\
&\geq C\epsilon^k = \delta > 0.
\end{aligned}$$

The last two inequalities come from Propositions 2.6 and 2.7. □

Now we are ready for the main part of the proof of Theorem 3.11. Let k be a large integer—how large will be determined below, but for now we regard k as fixed. Recall that $I_k^- = \cup_{j=0}^{k-1} f_-^{-j}(I^-)$. The fact that f_+ is minimally separating guarantees that I_k^- is a finite set of points.

We choose $\epsilon > 0$ and then $\delta > 0$ so that

(1) $B_{I_k^-}(\epsilon)$ is a disjoint union of balls—one for each point in I_k^-.
(2) $p \in I_k^-$ implies that either $p \in \mathcal{C}^+$ or $B_p(\epsilon) \cap B_{\mathcal{C}^+}(\epsilon) = \emptyset$.
(3) If $p \in I_k^-$ is a periodic critical point, then Lemma 3.14 applies with $N = B_p(\epsilon)$.
(4) $\omega(\delta) \subset B_{\mathcal{C}^+}(\epsilon)$
(5) $f_+^j(\omega(\delta)) \subset B_{I_k^-}(\epsilon)$ for all $1 \leq j \leq k$.

Let E be the set given in the statement of Theorem 3.11. We can assume without loss of generality that Volume $E < 1$ (at any rate $\mathbb{P}^2$ is compact, so Volume E is bounded above). We can define an equivalence relation $\sim$ on points in E as follows: $p_1 \sim p_2$ means that for each $0 \leq j \leq k$, and each $p \in I_k^-$:

(6) $f_+^j(p_1) \in \omega(\delta)$ if and only if $f_+^j(p_2) \in \omega(\delta)$;
(7) $f_+^j(p_1) \in B_p(\epsilon)$ if and only if $f_+^j(p_2) \in B_p(\epsilon)$;

That is, $p_1 \sim p_2$ if and only if for the first k iterates, their orbits stay together with respect to the partition determined by ϵ, δ and conditions (1) through (5). This equivalence relation divides E into a finite union of disjoint measurable subsets. Let $N = N(k)$ be the maximum possible number of pieces in the division. Then for some piece E_0, we must have Volume $E_0 \geq$ (Volume E)$/N(k)$.

Lemma 3.16. *Suppose that $f_+^j(E_0)$ is disjoint from $B_p(\epsilon) \cap \omega(\delta)$ for all $1 \leq j \leq k$ and all periodic critical points p. Then*

$$\text{Volume } f_+^k(E_0) \geq A(\text{Volume } E_0)^B$$

where A and B are constants independent of E_0, and B does not depend on k.

Proof. For each j, either $f_+^j(E_0) \cap \omega(\delta) = \emptyset$ or $f_+^j(E_0) \subset \omega(\delta)$. If the former possibility occurs, then by definition of $\omega(\delta)$, we have

$$\text{Volume } f_+^{j+1}(E_0) \geq \delta \,\text{Volume } f_+^j(E_0)$$

By hypothesis and Proposition 3.9, the latter possibility can only occur for $3d-3$ distinct values of j. For each of these occurrences, Corollary 3.13 tells us that

$$\text{Volume } f_+^{j+1}(E_0) \geq C(\text{Volume } f_+^j(E_0))^D$$

for constants C and D depending only on f_+. Putting these facts together gives us that

$$\text{Volume } f_+^k(E_0) \geq A(\delta, k)(\text{Volume } E_0)^{D^{3d-3}},$$

and we are done. □

Lemma 3.17. *Suppose there is a periodic critical point p and an integer j between 1 and k such that $f_+^j(E_0)$ intersects $B_p(\epsilon) \cap \omega(\delta)$. Then*

$$\text{Volume } f_+^k(E_0) \geq (A \,\text{Volume } E_0)^{Bk}$$

for constants A and B independent of E_0. B is also independent of k.

Proof. Let j be the smallest integer such that $f_+^j(E_0) \cap B_p(\epsilon) \cap \omega(\delta) \neq \emptyset$ for some periodic critical point p. Then in fact, $f_+^j(E_0) \subset B_p(\epsilon) \cap \omega(\delta)$. Also, the proof of the previous Lemma shows that

$$\text{Volume } f_+^j(E_0) \geq A(\text{Volume } E_0)^B,$$

where A and B are as in the statement of that lemma.

Now let m be the period of p and n be the largest integer such that $j + mn \leq k$. We can assume without loss of generality that $m < k - j$, since we are otherwise essentially in the situation covered by the previous lemma. Lemma 3.14 and condition (5) (above) tell us that

$$\text{Volume } f_+^{j+mn}(E_0) \geq C(\text{Volume } f_+^j(E_0))^{Dn} \geq (C \,\text{Volume } E_0))^{Dn},$$

where C and D are constants independent of E_0 and D is independent of k.

To handle iterates of f_+ between $j+mn$ and k, we again invoke the proof of Lemma 3.16 to obtain

$$\text{Volume } f_+^k(E_0) \geq A(\text{Volume } f_+^{j+mn}(E_0))^B.$$

The proof is now concluded by putting this estimate together with the previous one. $\square$

End of the proof of Theorem 3.11. Let $\lambda > 1$ be given and E, E_0 be as above. Then Lemmas 3.16 and 3.17 together imply the existence of constants $A = A(k)$ and B independent of E and E_0 such that

$$\text{Volume } f_+^k(E_0) \geq (A \,\text{Volume } E_0)^{Bk}.$$

We pick k large enough that $Bk \leq \lambda^k$. Since Volume $E_0 \geq$ (Volume E)$/N(k)$, we have that

$$\text{Volume } f_+^{kn}(E) \geq (C \,\text{Volume } E)^{\lambda^{kn}}$$

for all n. Now any positive integer j can be written $j = kn + m$ for some $m, n > 0$, $m < k$. Invoking Corollary 3.13, we obtain

$$\begin{aligned}\text{Volume } f_+^j(E) &\geq (A \,\text{Volume } f_+^{kn}(E))^{C_2} \\ &\geq (C_1 \,\text{Volume } E)^{C_2\lambda^{kn}} \geq (C_1 \,\text{Volume } E)^{C_2\lambda^j}\end{aligned}$$

for constants C_1, C_2 which depend on k but not on j. $\square$

Escape Functions for Birational Maps. We are now in a position to prove

Theorem 3.18. *The escape function for a degree $d \geq 2$ minimally separating birational map of $\mathbb{P}^2$ is plurisubharmonic.*

The first step in the proof was observed by Fornaess and Sibony.

Lemma 3.19. *If f_+ has periodic points, than $\tilde{G}^+$ is plurisubharmonic.*

Proof. If p is periodic, then the orbit of p does not intersect I^+. Since both I^+ and the orbit of p are finite sets, we see that there is a constant $C > 0$ such that $\text{dist}(f_+^n(p), I^+) \geq C$ for all n. Corollary 2.8 now gives us that $\tilde{G}^+(\tilde{p}) > -\infty$ for all $\tilde{p} \in \pi^{-1}(p)$. Theorem 2.9 implies that $\tilde{G}^+$ is plurisubharmonic. $\square$

Now set

$$E_n(t) = \{p \in \mathbb{P}^2 \setminus I_\infty^+ : \log \text{dist}(f_+^n(p), I^+) < -t\}.$$

Then we have

Lemma 3.20. *If f_+ is birational and $\tilde{G}$ is degenerate, then*

$$\sum_{n=0}^{\infty} \frac{1}{d^n} \int_0^{\infty} \text{Volume } E_n(t)\, dt = \infty. \tag{3–1}$$

Proof. If $\tilde{G}$ is degenerate, then by Theorem 2.9 $\tilde{G}_n$ converges in $\mathrm{L}^1_{\mathrm{loc}}$ (uniformly on compacts, in fact) to $-\infty$. By Corollary 2.8, we must have that

$$\sum_{n=0}^{\infty} \frac{1}{d^n} \int_{\mathbb{P}^2} -\log \operatorname{dist}(f_+^n(p), I^+)\, dV = \infty.$$

Since I_∞^+ is countable, we can replace the domain of integration by $\mathbb{P}^2 \setminus I_\infty^+$. It follows from standard real variables theory and compactness of $\mathbb{P}^2$ that the integral in (3–1) is equal up to an additive constant to the one in the statement of the lemma. □

Proof of Theorem 3.18. Since the set of points within distance e^{-t} of I^+ is a finite union of balls of radius e^{-t}, we have

$$\text{Volume } f_+^n(E_n(t)) \leq Ce^{-4t}.$$

Pick λ strictly between 1 and d. By Lemma 3.19, we are free to assume that f_+ has no superattracting periodic points. Hence, Theorem 3.11 assures us constants C_1, C_2 such that

$$\text{Volume } f_+^n(E_n(t)) \geq (C_1 \text{ Volume } E_n(t))^{C_2 \lambda^n}$$

for all n. Putting these two inequalities together yields

$$\text{Volume } E_n(t) \leq Ae^{-Bt/\lambda^n}.$$

Now we estimate

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{1}{d^n} \int_0^{\infty} \text{Volume } E_n(t)\, dt &\leq \sum_{n=0}^{\infty} \frac{1}{d^n} \int_0^{\infty} Ae^{-Bt/\lambda^n}\, dt \\ &\leq \sum_{n=0}^{\infty} \frac{A}{B} \left(\frac{\lambda}{d}\right)^n < \infty.\end{aligned}$$

By Lemma 3.20, we are done. □

We remark that we did not really need the full strength of Theorem 3.11 in this proof. We only needed a version of that theorem for maps *without periodic critical points*. Such a version admits a proof which is considerably simpler than the one we gave for Theorem 3.11. Only in the proof of Theorem 4.9 will we use the full strength of Theorem 3.11.

4. Positive Closed Currents on $\mathbb{P}^2$

In this section, we consider the currents $\mu^{\pm}$ on $\mathbb{P}^2$ induced by the escape functions $\tilde{G}^{\pm}$ for a birational map f and its inverse. We begin by recalling facts about positive, closed (1,1) currents on $\mathbb{P}^2$.

Given a current T on $\mathbb{P}^2$, the *mass norm* $M[T]$ of T is given by

$$M[T] = \sup\{|T(\phi)| : \phi \text{ is a test form with } |\phi(p)| \leq 1 \text{ for all } p \in \mathbb{P}^2\}.$$

We denote the set of all positive closed (1,1) currents on $\mathbb{P}^2$ by $\mathcal{Q}$. For $T \in \mathcal{Q}$, mass norm is comparable to *total variation* $||T|| = T(\theta) = \int_{\mathbb{P}^2} T \wedge \theta$ of T, where θ is the Kähler form for the Fubini-Study metric on $\mathbb{P}^2$.

If T is a current on an open subset W of $\mathbb{P}^2$, then the *trivial extension* of T to $\mathbb{P}^2$ is obtained by setting the coefficients of T equal to 0 on $\mathbb{P}^2 \setminus W$. Given a current T on $\mathbb{P}^2$ and $K \subset \mathbb{P}^2$ closed, we say that T *has no support concentrated on* K if T agrees with the trivial extension to $\mathbb{P}^2$ of $T|_{\mathbb{P}^2 \setminus K}$. Note that $\operatorname{supp} T \cap K \neq \emptyset$ does not necessarily imply that T has support concentrated on K.

Algebraic curves constitute an important subset of $\mathcal{Q}$. One defines the *current of integration* $[V]$ over a curve V by setting

$$[V](\eta) = \int_V \eta,$$

for any test (1,1) form η (technically, one integrates only over the regular points of V). It is not hard to show that $M[V] = ||[V]||$ is the algebraic degree of V. We will rely heavily on the following well-known pair of theorems concerning the relationship between curves and currents. [Sk] includes these theorems and attendant references in his survey article.

Theorem 4.1 (H. Skoda). *If $V \subset \mathbb{P}^2$ is an algebraic curve and T is a positive closed (1,1) current on $\mathbb{P}^2 \setminus V$ with bounded mass near V, then the trivial extension of T to $\mathbb{P}^2$ is positive and closed. In particular, if $T \in \mathcal{Q}$ then the trivial extension to $\mathbb{P}^2$ of $T|_{\mathbb{P}^2 \setminus V}$ is positive and closed.*

Theorem 4.2 (Y. Siu). *Suppose $T \in \mathcal{Q}$ satisfies $\operatorname{supp} T \subset V$ for some irreducible curve V. Then T is a multiple of $[V]$.*

Because of Skoda's theorem, we will make no distinction between $T|_{\mathbb{P}^2 \setminus V}$ and its trivial extension to $\mathbb{P}^2$. Siu's Theorem implies that any support that T concentrates on a curve V consists of a positive linear combination of currents of integration over the irreducible components of V.

The maximum principle implies that constants are the only functions U on $\mathbb{P}^2$ satisfying $dd^c U \geq 0$. However, there is still a sense in which all currents in $\mathcal{Q}$ arise from global plurisubharmonic functions. Let $\mathcal{U}$ denote the set of *homogeneous potentials* on $\mathbb{C}^3$. That is, a function $\tilde{U} : \mathbb{C}^3 \to \mathbb{R} \cup \{-\infty\}$ belongs to $\mathcal{U}$ if $\tilde{U}$ is plurisubharmonic and $\tilde{U}(\lambda \tilde{p}) = \tilde{U}(\tilde{p}) + c \log |\lambda|$ for all $\lambda \in \mathbb{C}$ and some $c > 0$. Then the following theorem appears in [FS1].

Theorem 4.3 (Fornaess and Sibony). *There exists a surjective linear map $L : \mathcal{U} \to \mathcal{Q}$ such that*

(1) $\pi^*(L(\tilde{U})) = dd^c\tilde{U}$;
(2) $||L(\tilde{U})|| = c$ *where c is the homogeneity constant for $\tilde{U}$.*
(3) $\ker L = \mathbb{R}$.

Rather than write $L(\tilde{U})$, we will use the more suggestive notation $\pi_* dd^c\tilde{U}$, since (1) states that L is in some sense dd^c followed by the inverse of π^*. (3) and surjectivity guarantee that any $T \in \mathcal{Q}$ is induced by a homogeneous potential that is unique up to addition of real constants. We note that $\pi_* dd^c\tilde{U}$ is defined locally on a ball $B \subset \mathbb{P}^2$ by choosing any holomorphic right inverse $\pi^{-1} : B \to \mathbb{C}^3$ of π and setting $\pi_* dd^c\tilde{U} = dd^c(\tilde{U} \circ \pi^{-1})$ on B. Homogeneity of $\tilde{U}$ guarantees that this definition is independent of the choice of π^{-1}. Finally, we remark that if V is an algebraic curve and $\tilde{P}$ is a minimal defining polynomial for V, then $\log|\tilde{P}| \in \mathcal{U}$, and $[V] = \pi_* dd^c \log|\tilde{P}|$.

Theorem 4.4. *If $T_1, T_2 \in \mathcal{Q}$ are non–zero currents, then* $\operatorname{supp} T_1 \cap \operatorname{supp} T_2 \neq \emptyset$. *In particular, every $T \in \mathcal{Q}$ has connected support.*

Proof. This is just a version of the proof of Theorem 4.7 in [FS2], much simplified by the fact that we are working in a narrower context. Let $\tilde{U}_1, \tilde{U}_2 \in \mathcal{U}$ be homogeneous potentials for T_1, T_2. If $\operatorname{supp} T_1 \cap \operatorname{supp} T_2 = \emptyset$, then $\tilde{U}_1$ is pluriharmonic in a neighborhood of $\pi^{-1}(\operatorname{supp} T_2)$ and vice versa. Thus T_1 and T_2 "admit a wedge product" $T_1 \wedge T_2$ in the sense described in [FS2]. In fact, $T_1 \wedge T_2$ must obviously be 0. But by Theorem 4.4 in [FS2], we must have

$$||T_1 \wedge T_2|| = ||T_1||\,||T_2||.$$

This implies that either T_1 or T_2 is zero.

Now suppose that $T \in \mathcal{Q}$ is a current such that $\operatorname{supp} T \subset W_1 \cup W_2$, where W_1 and W_2 are disjoint open sets. Then $T_j = T|_{W_j} \in \mathcal{Q}$ for $j = 1, 2$, and $\operatorname{supp} T_1 \cap \operatorname{supp} T_2 = \emptyset$. By the first part of the theorem, one of the T_j must be zero, and we conclude that $\operatorname{supp} T$ lies entirely inside one of the W_j. $\square$

Given a birational map $f_+ : \mathbb{P}^2 \to \mathbb{P}^2$, we would like to be able to push forward and pull back currents in $\mathcal{Q}$ by f_+. However if $\deg f_+ \geq 2$, f_+ is neither a (locally) proper map nor a submersion; so the typical notions of pullback and pushforward of currents do not apply. Nonetheless, Proposition 3.3 describes very precisely the way in which f_+ fails to be a biholomorphism, so it seems reasonable to use this information to extend the definitions of pullback and pushforward. The following definition is inspired by the discussion of images and preimages of algebraic curves given after Proposition 2.2.

Definition 4.5. *Let f_+ be a birational map of $\mathbb{P}^2$ and $T \in \mathcal{Q}$ be a current with homogeneous potential $\tilde{U} \in \mathcal{U}$. Then*

$$f_+^* T = \pi_* dd^c\tilde{U} \circ \tilde{f}_+.$$

Furthermore, let f_0 be the restriction of f_+ to $\mathbb{P}^2 \setminus \mathcal{C}^+$ and T_0 be restriction of T to $\mathbb{P}^2 \setminus \mathcal{C}^+$; then $f_{+*}T$ is the trivial extension of $f_{0*}T_0$ to $\mathbb{P}^2$.

We remark that the definition of f_+^*T is independent of which representative $\tilde{f}_+$ we choose for f_+ and which potential $\tilde{U}$ we choose for T. There are several ways in which we could reasonably extend the definitions of f_+^*T and $f_{+*}T$ even further. For instance, f^*T makes sense for rational maps f which are not birational. With a bit more effort, one can even make sense of f_*T. One can also relax somewhat the requirements that T be positive and closed. But the definition we have given here will be sufficient for the time being.

Proposition 4.6. *Given a birational map $f_+ : \mathbb{P}^2 \to \mathbb{P}^2$ and $T \in \mathcal{Q}$, we have that*

(1) $f_+^*T, f_{+*}T \in \mathcal{Q}$;
(2) f_+^* *and* f_{+*} *are linear on* $\mathcal{Q}$;
(3) $\operatorname{supp} f_+^*T \subseteq f_+^{-1}(\operatorname{supp} T)$;
(4) $\operatorname{supp} f_{+*}T \subseteq f_+(\operatorname{supp} T)$, *and equality holds if T has no support concentrated on $\mathcal{C}^+$.*

Proof. (1) is clear for f_+^*T; (1) follows from Theorem 4.1 for $f_{+*}T$. (2) follows more or less directly from definitions. To prove (3), suppose there exists $p \in \operatorname{supp} f_+^*T \setminus f_+^{-1}(\operatorname{supp} T)$. Since $\operatorname{supp} f_+^*T$ is connected and closed, we can assume that $p \notin I^+$. Thus we have that f_+ maps a small neighborhood W of p holomorphically into an open set W' such that $\overline{W'} \cap \operatorname{supp} T = \emptyset$. If $\tilde{U}$ is a potential for T, then $\tilde{U}$ is pluriharmonic on W', and $dd^c\tilde{U} \circ \tilde{f}_+ = 0$ on $\pi^{-1}(W)$—a contradiction.

To prove (4), we let $T_0 = T|_{\mathbb{P}^2 \setminus \mathcal{C}^+}$ and $T_1 = T - T_0 = T|_{\mathcal{C}^+}$. By Theorem 4.2, $T|_{\mathcal{C}^+}$ consists of multiples of currents of integration $[V]$ over curves $V \subset \mathcal{C}^+$. Both T_0 and T_1 belong to $\mathcal{Q}$. By definition $f_{+*}T = f_{+*}T_0$. Since the support of a current is closed, we invoke (3) of Proposition 3.3 to obtain

$$\operatorname{supp} f_{+*}T_0 = \overline{(\operatorname{supp} f_{+*}T_0) \setminus \mathcal{C}^-}$$
$$f_+(\operatorname{supp} T_0) = \overline{f_+(\operatorname{supp} T_0) \setminus \mathcal{C}^-}.$$

By (3) of Proposition 3.3 again, $(\operatorname{supp} f_{+*}T_0) \setminus \mathcal{C}^- = f_+(\operatorname{supp} T_0) \setminus \mathcal{C}^-$. Item (4) of this proposition now follows. □

Proposition 4.7. *Let f_+ and T be as in the last proposition. Write $T_+ = T|_{\mathbb{P}^2 \setminus \mathcal{C}^+}$ and $T_- = T|_{\mathbb{P}^2 \setminus \mathcal{C}^-}$, and let $\tilde{U}, \tilde{U}_+, \tilde{U}_- \in \mathcal{Q}$ be potentials for T, T_+, T_-, respectively. Then the following statements hold.*

(1) $||f_+^*T|| = d\,||T||$.
(2) $f_+^*(T_-) - (f_+^*T)|_{\mathbb{P}^2 \setminus \mathcal{C}^+}$ *is a positive, closed current supported on $\mathcal{C}^+$. If $\tilde{U}_-(\tilde{p}) > -\infty$ for all $\pi(\tilde{p}) \in I^-$, then the difference vanishes.*
(3) $f_{-*}T = (f_+^*T)|_{\mathbb{P}^2 \setminus \mathcal{C}^+}$.
(4) $f_+^*f_-^*T = T + \pi_* dd^c \log|\tilde{P}^+|$, *where $\tilde{P}^+$ is the polynomial given by Corollary 3.6.*

(5) $f_{-*}f_{+*}T = T_+$.
(6) $f_{+*}f_+^*T = T_-$.
(7) *If* $\tilde{U}_+(\tilde{p}) > -\infty$ *for all* $\pi(\tilde{p}) \in I^+$, *then* $f_+^*f_{+*}T = T_+ + \pi_* dd^c \log|\tilde{P}^+|$

Proof. Note that $T - T_+$ is a sum of positive multiples of currents of integration over components of $\mathcal{C}^+$, and similarly with $T - T_-$.

(1): This follows from (2) of Theorem 4.3, degree d homogeneity of $\tilde{f}_+$, and the fact that $\tilde{U} \circ \tilde{f}_+$ is a potential for f_+^*T.

(2): $f_+^*(T - T_-)$ is supported on $\mathcal{C}^+$; hence $f_+^*T_- - (f_+^*T)|_{\mathbb{P}^2\setminus\mathcal{C}^+} = f_+^*T_- - (f_+^*T_-)|_{\mathbb{P}^2\setminus\mathcal{C}^+}$ is a positive, closed current supported on $\mathcal{C}^+$. Given the extra condition, we have that $\tilde{U}_- \circ \tilde{f}_+(\tilde{p}) > -\infty$ for all $\pi(\tilde{p}) \in \mathcal{C}^+ \setminus I^+$. Hence, $f_+^*T_-$ has no support concentrated on $\mathcal{C}^+$.

(3): This follows from $f_{-*}T = f_{-*}T_- = (f_+^*T_-)|_{\mathbb{P}^2\setminus\mathcal{C}^+} = (f_+^*T)|_{\mathbb{P}^2\setminus\mathcal{C}^+}$. The second equality holds because $f_+ : \mathbb{P}^2 \setminus \mathcal{C}^+ \to \mathbb{P}^2 \setminus \mathcal{C}^-$ is a biholomorphism.

(4): By Corollary 3.6, we have $\tilde{U} \circ \tilde{f}_- \circ \tilde{f}_+(\tilde{p}) = \tilde{U}(\tilde{P}^+(\tilde{p}) \cdot \tilde{p}) = \tilde{U}(\tilde{p}) + \log|\tilde{P}^+|$. Applying $\pi_* dd^c$ to this equation proves the item.

(5): We have $f_{-*}f_{+*}T = f_{-*}f_{+*}T_+ + f_{-*}f_{+*}(T - T_+) = f_{-*}f_{+*}T_+ = T_+$. The last equality follows from the definition of pushforward and the fact that $f_+ : \mathbb{P}^2 \setminus \mathcal{C}^+ \to \mathbb{P}^2 \setminus \mathcal{C}^-$ is a biholomorphism.

(6): We use (3) and (4) to obtain $f_{+*}f_+^*T = (f_-^*f_+^*T)|_{\mathbb{P}^2\setminus\mathcal{C}^-} = T|_{\mathcal{C}^-}$. (Recall that $\tilde{P}^-$ is a defining polynomial for $\mathcal{C}^-$.)

(7): We apply (2), (3), and (4) to get $f_+^*f_{+*}T = f_+^*((f_-^*T)|_{\mathbb{P}^2\setminus\mathcal{C}^-}) = f_+^*f_-^*T_+ = T_+ + \pi_* dd^c \log|\tilde{P}^+|$. □

If $\tilde{G}^+$ is plurisubharmonic, then Theorem 4.3 and (1) of Theorem 2.9 imply that $\tilde{G}^+$ induces a current $\mu^+ \in \mathcal{Q}$ on $\mathbb{P}^2$. This current enjoys special transformation properties under pushforward and pullback by f_+ and f_-.

Theorem 4.8. *Suppose that* $f_+ : \mathbb{P}^2 \to \mathbb{P}^2$ *is birational and* $\tilde{G}^+$ *is plurisubharmonic. Then*

(1) $f_+^*\mu^+ = d \cdot \mu^+$;
(2) $f_-^*\mu^+ = \dfrac{1}{d}(\mu^+ + \pi_* dd^c \log|\tilde{P}^-|)$.

Furthermore, the following statements are equivalent:

(3) f_+ *is minimally separating;*
(4) μ^+ *has no support concentrated on* $\mathcal{C}^+$;
(5) μ^+ *has no support concentrated on any algebraic curve;*
(6) $f_{-*}\mu^+ = d \cdot \mu^+$.

Any one of these statements implies that

(7) $f_{+*}\mu^+ = \mu^+/d$.

Proof. Since $\tilde{G}^+ \circ \tilde{f}_+ = d \cdot \tilde{G}^+$, it follows that $f_+^*\mu^+ = d \cdot \mu^+$. (2) follows from (4) of Proposition 4.7 and $f_-^*\mu^+ = \frac{1}{d} f_-^* f_+^* \mu^+$.

Turning to the equivalence of (3) through (6), we note that (3) of Proposition 4.7 gives us that (4) and (6) are equivalent. Clearly, (5) implies (4). To see that (4) implies (3), suppose that f_+ is not minimally separating. Choose $p^- \in I^-$, a curve $V^+ \subset \mathcal{C}^+$ such that $f_+(\mathcal{C}^+) = p^-$ and a non-negative integer k such that $f_+^k(p^-) = p^+ \in I^+$. Then as in the proof of Proposition 2.3, $\tilde{f}_+^{k+1}$ is divisible by $\tilde{P}$ for some defining polynomial $\tilde{P}$ of V. Recall that the sequence $\tilde{G}_n^+$ converging to $\tilde{G}^+$ is decreasing. So we have

$$\tilde{G}^+(\tilde{p}) \leq \tilde{G}_{k+1}(\tilde{p}) = \frac{1}{d^{k+1}} \log ||\tilde{f}_+^{k+1}|| \leq C_1 + C_2 \log |\tilde{P}(\tilde{p})|.$$

Consequently, $\mu^+ = \pi_* dd^c \tilde{G}^+$ has support concentrated on $\mathcal{C}^+ = \pi\{\tilde{P} = 0\}$. The implication (3) $\Rightarrow$ (5) is a special case of Theorem 4.10 of [FS1], where Fornaess and Sibony show that the escape function (if it is plurisubharmonic) for a rational map with maximal dynamic degree has no support on any algebraic curve.

Finally, if we assume (5) is true and apply (2), we obtain that

$$f_{+*}\mu^+ = (f_-^* \mu^+)|_{\mathbb{P}^2 \setminus \mathcal{C}^-} = \frac{1}{d}\mu^+|_{\mathbb{P}^2 \setminus \mathcal{C}^-} = \frac{1}{d}\mu^+.$$

The first equality follows from (4) of Proposition 4.7. $\square$

Part of the significance of the currents μ^+ and μ^- is the relationship between their supports and sets on which the iterates of f_+ and f_- form normal families.

Definition 4.9. *The* **set of (forward) normality** *for a birational map f_+ is the largest open subset $\mathcal{N}^+$ of $\mathcal{D}^+$ such that iterates of f_+ form a normal family on $\mathcal{N}^+$. $\mathcal{N}^-$ is the corresponding set for f_-.*

The following result appears in [FS1], [HP], and [Ue].

Theorem 4.10. $\operatorname{supp} \mu^+ \cap \mathcal{N}^+ = \emptyset$.

We refer the reader to one of the above sources for the proof. It is worth noting that one can define a current μ^+ (though in general, there is no corresponding μ^-) for any rational map f of $\mathbb{P}^2$ with a plurisubharmonic escape function. When the indeterminacy set of the map is empty it turns out that $\operatorname{supp} \mu^+$ is exactly equal to $\mathbb{P}^2 \setminus \mathcal{N}^+$ (see [Ue]). We do not know if this complementarity holds for birational maps (which always have points of indeterminacy when the degree is larger than 2), but see section 6 below.

A Convergence Theorem. For the remainder of this section we assume that f_+ is a minimally separating birational map with degree $d \geq 2$. Let N be any neighborhood of the set of points lying in the forward orbits of superattracting periodic points of f_+. We define $\mathcal{Q}_N \subset \mathcal{Q}$ to be closed positive (1,1) currents T such that $||T|| = 1$ and $\operatorname{supp} T \cap N = \emptyset$.

Theorem 4.11. *Given $T \in \mathcal{Q}_N$,*

$$\lim_{n\to\infty} \frac{1}{d^n} f_+^{n*} T = \mu^+.$$

The convergence takes place in the weak topology on currents and is uniform on $\mathcal{Q}_N$.

The statement and proof that we give for this theorem are based on those of a similar result appearing in [FS3] for Hénon maps. The major differences between our context and the one in [FS3] is our use of Theorem 3.11 to control the rate at which f_+ shrinks volume, and the fact that $\mathbb{C}^2 \subset \mathbb{P}^2$ is not invariant for an arbitrary birational map.

Theorem 4.11 allows one to see the difference between pullback by f_+ and push-forward by f_- quite vividly. Namely, suppose that f_+ has no superattracting cycles. Then we can take $N = \emptyset$ and $\mu^- \in \mathcal{Q}_N = \mathcal{Q}$. The theorem implies that $\lim d^{-n} f_+^{n*}\mu^- = \mu^+$. On the other hand, $\lim d^{-n} f_{-*}^n \mu^- = \lim d^{-2n}\mu^- = 0$. Nonetheless, (2) and (3) of Proposition 4.7 show that a slight additional hypothesis on a current $T \in \mathcal{Q}$ is enough to prevent the discrepancy between f_+^* and f_{-*}. Namely, we have

Corollary 4.12. *Suppose that $T \in \mathcal{Q}$ satisfies $||T|| = 1$ and $\operatorname{supp} T \cap I_\infty^- = \emptyset$. Then*

$$\lim_{n\to\infty} \frac{1}{d^n} f_+^{n*} T = \lim_{n\to\infty} \frac{1}{d^n} f_{-*}^n T = \mu^+.$$

Proof. Since $\operatorname{supp} T$ avoids I_∞^-, it avoids an entire neighborhood of the finite set of points lying in superattracting cycles. Also, by (2) and (3) of Proposition 4.7,

$$f_+^{n*} T = f_+^{n*}(T|_{\mathbb{P}^2\setminus\mathcal{C}^-}) = (f_+^{n*}T)|_{\mathbb{P}^2\setminus\mathcal{C}^+} = f_{-*}^n T$$

for all $n \geq 0$. $\square$

Lemma 4.13. *Let $\tilde{U} \in \mathcal{U}$ satisfy $\max_{||\tilde{p}||=1} \tilde{U}(\tilde{p}) = 0$, and $\tilde{U}(\lambda\tilde{p}) = \tilde{U}(\tilde{p}) + \log|\lambda|$. Then there is a constant C independent of $\tilde{U}$ such that*

$$\text{Volume}\,\{\pi(\tilde{p}) \in \mathbb{P}^2 : \tilde{U}(\tilde{p}) - \log||\tilde{p}|| < -k\} \leq Ce^{-k}.$$

Suppose further that there exists an open set $N \subset \mathbb{P}^2$ such that $\tilde{U}$ is pluriharmonic on $\pi^{-1}(N)$. Then on any $K \subset\subset N$ there exists a constant C such that

$$\tilde{U}(\tilde{p}) - \log||\tilde{p}|| \geq C$$

on $\pi^{-1}(K)$. C depends on N and K, but not on $\tilde{U}$.

Proof. To prove the first statement, we may suppose without loss of generality that $\tilde{U}(0,0,1) = 0$. Then for any fixed (x,y),

$$\limsup_{z\to\infty} \tilde{U}(x,y,z) - \log|z| = \limsup_{z\to 0} \tilde{U}(xz, yz, 1) = 0 \tag{4–1}$$

by upper semicontinuity and the maximum principle applied to $\tilde{U}$ restricted to $\{(xz, yz, 1) : |z| \leq r\}$.

Now let $E = \{\pi(\tilde{p}) \in \mathbb{P}^2 : \tilde{U}(\tilde{p}) - \log ||\tilde{p}|| < -k\}$. Then the Fubini-Study volume of E is comparable to the Euclidean volume of $\tilde{E} = \pi^{-1}(E) \cap \{||\tilde{p}|| < 1\}$. Let $\tilde{E}(x, y) = \{z \in \mathbb{C} : (x, y, z) \in \tilde{E}\}$. Then $\tilde{U}(x, y, z)$ is less than $-k$ on $\tilde{E}(x, y)$, so by (4–1), the Robin constant of $\tilde{E}(x, y)$ is no smaller than k. Standard one variable potential theory (see [Ra] Theorems 5.2.1 and 5.3.5) gives us that $\text{Area}\, \tilde{E}(x, y) \leq e^{-k}$. Thus,

$$\text{Volume } E \leq C \,\text{Volume } \tilde{E} = C \int_{||(x,y)||<1} \text{Area}\, \tilde{E}(x, y)\, dV(x, y) \leq Ce^{-k}.$$

This finishes the proof of the first statement in the lemma.

To get the second statement, we may suppose without loss of generality that $N = B_0(R)$ is an open ball of Euclidean radius $R < 1$ about $\pi(0, 0, 1) = (0, 0) \in \mathbb{C}^2$, and $K = \overline{B_0(r)}$ is a concentric closed ball of smaller radius. We no longer assume that $\tilde{U} - \log ||\cdot||$ achieves its maximum at $(0, 0, 1)$; however, homogeneity of $\tilde{U}$ and the hypothesis that $\tilde{U}$ is negative on the unit ball implies that $\log \sqrt{2} - \tilde{U}(x, y, 1) > 0$ for $(x, y) \in K$. By Harnack's inequality and pluriharmonicity of $\log 2 - \tilde{U}$ on $\pi^{-1}(N)$, there is a constant C such that

$$\log 2 - \tilde{U}(x, y, 1) \leq C(\log \sqrt{2} - \tilde{U}(0, 0, 1))$$

for all $(x, y) \in K$. Hence,

$$\tilde{U}(x, y, 1) - \log ||(x, y, 1)|| \geq C(\tilde{U}(0, 0, 1) - \log \sqrt{2}).$$

In combination with the first statement of the lemma, this implies that Volume $K \leq C' e^{C(\tilde{U}(0,0,1) - \log \sqrt{2})}$. Hence

$$\tilde{U}(x, y, 1) - \log ||(x, y, 1)|| \geq C \log(\text{Volume } K) - C' = C(N, K),$$

and we are done. □

To continue with the proof of Theorem 4.11, pick any sequence of currents $T_n \in \mathcal{Q}_0$. Choose potentials $\tilde{U}_n \in \mathcal{U}$ for each T_n. By adding a constant to each $\tilde{U}_n$, we can arrange that $\sup_{||\tilde{p}||=1} \tilde{U}_n(\tilde{p}) = 0$. Set

$$\tilde{V}_n = d^{-n} \tilde{U}_n \circ \tilde{f}_+^n.$$

We will be done if we can show that $\tilde{V}_n \to \tilde{G}^+$ in $\mathrm{L}^1_{\mathrm{loc}}(\mathbb{C}^3)$.

Since $\log ||\tilde{f}_+^n(\tilde{p})|| \leq d^n \log ||\tilde{p}||$, the functions $\tilde{V}_n$ are uniformly bounded above on compact subsets. In particular, since $\tilde{U}_n(\tilde{p}) \leq \log ||\tilde{p}||$, we have that $\tilde{V}_n \leq \tilde{G}_n^+$, so that any limit of a subsequence is no larger than $\tilde{G}^+$. Let $\mathcal{B}$ be the set of all points in $\mathbb{P}^2$ whose orbits tend to superattracting cycles of f_+. The following two lemmas will complete the proof of the theorem.

Lemma 4.14. *$\tilde{V}_n - \tilde{G}_n^+ \to 0$ uniformly on compact subsets of $\pi^{-1}(\mathcal{B})$.*

Proof. We may suppose without loss of generality that the neighborhood N in the definition of $\mathcal{Q}_N$ is a subset of $\mathcal{B}$. By hypothesis $\tilde{U}_n$ is pluriharmonic on $\pi^{-1}(N)$ for all n. Choose a set $K \subset\subset \mathcal{B}$ and a neighborhood N' of the set of points in superattracting cycles such that $\overline{N'} \subset\subset N$. Then we can also choose constants m and C, such that $f_+^n(K) \subset N'$ for all $n > m$, and

$$\tilde{U}_n(\tilde{p}) - \log ||\tilde{p}|| \geq -C$$

for all n and all $\pi(\tilde{p}) \in N'$. Thus

$$0 \geq \tilde{V}_n(\tilde{p}) - \tilde{G}_n^+(\tilde{p}) \geq \frac{C}{d^n}$$

for all $n > m$ and $\pi(\tilde{p}) \in K$. That is, $\tilde{V}_n - \tilde{G}_n^+$ tends to 0 uniformly on $\pi^{-1}(K)$. □

Lemma 4.15. *$\tilde{V}_n - \tilde{G}_n^+ \to 0$ in $\mathrm{L}^1_{\mathrm{loc}}$ on $\pi^{-1}(\mathbb{P}^2 \setminus \mathcal{B})$.*

Proof. If the lemma is false, then after refining the sequence, we can choose numbers $\epsilon, \delta > 0$ measurable sets $E_n \subset \mathbb{P}^2 \setminus \mathcal{B}$ such that Volume $(E_n) \geq \delta$ and

$$-\epsilon > \tilde{V}_n(\tilde{p}) - \tilde{G}_n^+(\tilde{p}) = \frac{1}{d^n}(\tilde{U}_n(\tilde{f}_+^n(\tilde{p})) - \log ||\tilde{f}_+^n(\tilde{p})||)$$

for all $\pi(\tilde{p}) \in E_n$. Thus $\tilde{U}_n(\tilde{p}) - \log ||\tilde{p}|| \leq -d^n\epsilon$ on $\pi^{-1}(f_+^n(E_n))$. But then Lemma 4.13 gives us that

$$\limsup_{n\to\infty} \frac{\log \text{Volume}\,(f_+^n(E_n))}{d^n} \leq -\epsilon$$

in contradiction to Theorem 3.11. □

Remark 4.16. *Using essentially the same proof as the one we have given here, one can also prove: let $T \in \mathcal{Q}$ be such that $||T|| = 1$ and a potential $\tilde{U} \in \mathcal{U}$ for T is locally bounded on $\pi^{-1}(W)$ for some neighborhood W of all points in superattracting cycles; then the normalized big pullbacks $d^{-n} f_+^{n*}T$ converge weakly to μ^+. The only difference in the proof comes in the proof of Lemma 4.14, where instead of relying on the second statement of Lemma 4.13, we simply use the boundedness hypothesis for $\tilde{U}$.*

5. Mildly Separating Birational Maps

In Section 3, we showed that a minimally separating birational map always has a plurisubharmonic escape function. However, it is a priori possible that the escape function will still be very badly behaved. For instance, we have not ruled out the possibility that $\tilde{G}^+$ has a dense set of poles. In this section we show that a slight extra hypothesis on a birational map leads to the conclusion that $\tilde{G}^+$ is actually continuous on $\pi^{-1}(\mathcal{D}^+)$. The extra hypothesis has the added appeal that, though it is stated as a condition on f_+, it turns out to be symmetric in f_+ and f_-.

Definition 5.1. *A birational map f_+ with degree at least two is* **mildly separating** *if $\tilde{G}^+(\tilde{p}) > -\infty$ whenever $\pi(\tilde{p}) \in I^-$.*

Statement (2) of the next theorem justifies our use of the term "mildly separating." Note that a mildly separating birational map is automatically minimally separating. If f_+ is not minimally separating, then there is an integer k and a point $p^- \in I^-$ such that $f_+^k(p^-) \in I^+$—i.e. in particular $\tilde{G}^+(\tilde{p}) = -\infty$ whenever $\pi(\tilde{p}) = p^-$.

Theorem 5.2. *The following statements about a degree $d \geq 2$ birational map $f_+ : \mathbb{P}^2 \to \mathbb{P}^2$ are equivalent.*

(1) *f_+ is mildly separating.*
(2) $\displaystyle\sum_{n=0}^{\infty} \frac{1}{d^n} \log \operatorname{dist}(f_+^n(I^-), I^+) > -\infty.$
(3) *f_- is mildly separating.*
(4) $\displaystyle\sum_{n=0}^{\infty} \frac{1}{d^n} \log \operatorname{dist}(f_-^n(I^+), I^-) > -\infty$

Any one of these statements implies that $\tilde{G}_n^+ \to \tilde{G}^+$ uniformly on compact subsets of $\pi^{-1}(\mathcal{D}^+)$ and that in particular, $\tilde{G}^+$ is continuous on $\pi^{-1}(\mathcal{D}^+)$. Moreover, we have the estimate

$$\tilde{G}^+(\tilde{p}) - \log ||\tilde{p}|| \geq A \log \operatorname{dist}(\pi(\tilde{p}), \mathcal{I}^+) + B \tag{5–1}$$

for all $\tilde{p} \in \mathbb{C}^3 \setminus \{0\}$. Similar conclusions hold for $\tilde{G}^-$ on $\mathcal{D}^-$.

We remark that statement (2) in Theorem 5.2 seems only very slightly stronger than statement (4) in Corollary 3.4. That is, requiring f_+ to be mildly separating does not seem much more restrictive than requiring f_+ to be minimally separating. It would be very interesting to know whether there are minimally separating birational maps which are not mildly separating.

Beginning of the proof of Theorem 5.2. We show (1) $\Leftrightarrow$ (2). Given any point $\tilde{p} \in \pi^{-1}(I^-)$, Corollary 2.8 shows that $\tilde{G}^+(\tilde{p}) > -\infty$ if and only if

$$\sum_{n=0}^{\infty} \frac{1}{d^n} \log \operatorname{dist}(f_+^n(\pi(\tilde{p})), I^+) > -\infty.$$

Since I^- is finite, we see that $\tilde{G}^+(\tilde{p}) > -\infty$ for *all* $\tilde{p} \in \pi^{-1}(I^-)$ if and only if (2) holds. The proof of (3) $\Leftrightarrow$ (4) is the same. $\square$

Before proceeding further with the proof of the Theorem, we define some notation and establish a lemma. Given a point $p \in \mathbb{P}^2$, we set

$$\epsilon_n = \operatorname{dist}(f_-^n(I^+), f_-^n(p)),$$
$$\delta_n = \min\{1, \operatorname{dist}(f_-^n(I^+), I^-)\}.$$

Lemma 5.3. *There are constants A and k such that for any $D < 1$ and positive integer N,*

$$\epsilon_0 \leq \frac{(\delta_0\delta_1\dots\delta_N)^k}{2A^N}D$$

implies that $\epsilon_n \leq \delta_n D/2$ for all $n \leq N$.

Proof. Suppose that $\epsilon_n \leq \delta_n/2$. Then $\operatorname{dist}(f_-^n(p), I^-) \geq \delta_n/2$. By Propositions 2.6 and 2.7,

$$\epsilon_{n+1} \leq \epsilon_n e^{-k\log(\delta_n/2)-C} = A\epsilon_n\delta_n^{-k}$$

where A and k are constants independent of n. For convenience, we take both A and k to be greater than 1. If we suppose that

$$\epsilon_0 \leq \frac{(\delta_0\delta_1\dots\delta_N)^k}{2A^N}D \leq \frac{\delta_0}{2},$$

it follows by induction that $\epsilon_n \leq \delta_n D/2$ for all $n \leq N$. $\square$

Remainder of the proof of Theorem 5.2. First we show that (2) is equivalent to (4). Since the statements are symmetric with respect to f_+ and f_-, it is enough to prove only that (4) implies (2). That is, assume

$$\sum_{n=0}^{\infty}\frac{\log\delta_n}{d^n} > -\infty. \tag{5–2}$$

Then apply Lemma 5.3 with $D = 1$. If ϵ_0 is as in the lemma, and $\operatorname{dist}(p, I^+) < \epsilon_0$, then it follows that $f_-^N(p) \notin I^-$. Thus

$$\operatorname{dist}(f_+^N(I^-), I^+) \geq \epsilon_0 = \frac{(\delta_0\delta_1\dots\delta_N)^k}{2A^N},$$

From this we obtain

$$\begin{aligned}\sum_{n=0}^{\infty}\frac{1}{d^n}\log\operatorname{dist}(f_+^n(I^-), I^+) &\geq \sum_{n=0}^{\infty}\left(-\frac{n\log A}{d^n} + \frac{k}{d^n}\sum_{j=0}^{n}\log\delta_j\right)\\ &\geq C + k\sum_{j=0}^{\infty}(\log\delta_j)\sum_{n=j}^{\infty}\frac{1}{d^n}\\ &= C + \frac{k}{1-1/d}\sum_{j=0}^{\infty}\frac{\log\delta_j}{d^j}\\ &> -\infty.\end{aligned}$$

Hence, (2) holds, and we are finished showing (2) $\Leftrightarrow$ (4)

Now we will show that (4) implies the rest of the conclusion of the Theorem. Fix a compact subset K of $\mathcal{D}^+$, and set $D = \min\{1, \mathrm{dist}(K, \mathcal{I}^+)\}$. Given N, we choose

$$\epsilon_{0,N} = \frac{(\delta_0\delta_1 \dots \delta_N)^k}{2A^N} D$$

as in Lemma 5.3. If p satisfies $\mathrm{dist}(p, I^+) < \epsilon_{0,N}$, the conclusion of Lemma 5.3 shows us that $f_-^n(p) \notin K$ for any $n \leq N$. Hence, $\mathrm{dist}(f_+^N(K), I^+) \geq \epsilon_{0,N}$. Invoking Proposition 2.7 and computing as we did in the first part of the proof, we see for any $p = \pi(\tilde{p}) \in K$ that

$$\begin{aligned}
\tilde{G}^+(\tilde{p}) - \log ||\tilde{p}|| &\geq C_1 + C_2 \sum_{n=0}^{\infty} \frac{\log \mathrm{dist}(f_+^n(p), I^+)}{d^n} \\
&\geq C_1 + C_2 \sum_{n=0}^{\infty} \frac{1}{d^n} \log \left(\frac{(\delta_0\delta_1 \dots \delta_n)^k}{2A^n} D \right) \\
&= E + F \log D
\end{aligned}$$

for constants E and F independent of $\tilde{p}$ and K. This gives us that $\tilde{G}^+(\tilde{p})$ satifies the estimate (5–1).

To see that $\tilde{G}_n^+$ converges *uniformly* to $\tilde{G}^+$ on $\pi^{-1}(K)$, we note that by Proposition 2.7 and the estimate $\mathrm{dist}(f_+^N(p), I^+) \geq \epsilon_{0,N}$, we have

$$|\tilde{G}_n^+(\tilde{p}) - \tilde{G}^+(\tilde{p})| \leq \frac{C_1}{d^n} - C_2 \sum_{j=n}^{\infty} \frac{1}{d^j} \log \frac{(\delta_0\delta_1 \dots \delta_j)^k D}{A^j}.$$

The right side is independent of $\tilde{p} \in \pi^{-1}(K)$, and since the rightmost term is the tail of a convergent series, the right side tends to 0 as $n \to \infty$. $\square$

As further evidence that "mildly separating" is only slightly different from "minimally separating," we offer an analogue for Theorem 4.8. Given a degree $d \geq 2$ birational map f_+ of $\mathbb{P}^2$, let $\mathcal{S}^+ = \pi(\{\tilde{G}^+ = -\infty\})$.

Theorem 5.4. *The following statements are equivalent.*

(1) f_+ *is mildly separating.*
(2) $\mathcal{C}^+ \cap S^+ = I^+$.
(3) $\mathcal{S}^+$ *contains no algebraic curve.*

Proof. f_+ is mildly separating if and only if $\mathcal{S}^+ \cap I^- = \emptyset$. Recall also that $\mathcal{S}^+$ always contains I^+. Since $\tilde{G}^+ \circ \tilde{f}_+ = d \cdot \tilde{G}^+$, and $f_+(\mathcal{C}^+) = I^-$, it is clear that (1) and (2) are equivalent. Furthermore, if f_+ is not mildly separating pick $p \in I^- \cap \mathcal{S}^+$. By Proposition 3.3, there is a component $V \subset \mathcal{C}^+$ such that $f_+(V) = p$. It follows that $V \subset \mathcal{S}^+$ and that (3) implies (1).

We still need to show that (1) implies (3). To this end, suppose f_+ is mildly separating, but that V_0 is an irreducible algebraic curve such that $V_0 \subset \mathcal{S}^+$. Then we know that $V_0 \cap I^- = \emptyset$, and by (2) $V_0 \not\subset \mathcal{C}^+$. Hence, $V_1 = f_+(V_0)$ is an irreducible algebraic curve whose intersection with $\mathcal{C}^-$ is finite. We apply the identity $\tilde{G}^+ \circ \tilde{f}_+ = d\tilde{G}^+$ to conclude that $V_1 \in \mathcal{S}^+$ and that V_1 has the same properties as V_0. Because V_1 does not intersect I^-, $f_+^{-1}(V_1) = V_0$. Continuing inductively, we obtain irreducible algebraic curves $V_0, V_1, V_2, \ldots$ such that $V_k \cap I^- = \emptyset$ and $f_+^{-1}(V_k) = V_{k-1}$. But now we invoke Corollary 3.7 to obtain that

$$\deg V_0 = (\deg f_+)^k (\deg V_k).$$

Since $\deg V_k \geq 1$, $\deg f_+ \geq 2$, we have a contradiction for large k. Thus V_0 does not exist. $\square$

The next theorem generalizes 7.3.3 of [FS3]. Modulo technical details involving pushforwards and pullbacks, the proof remains essentially the same.

Theorem 5.5. *If f_+ is mildly separating, μ^+ is extremal in the cone of positive, closed (1,1) currents on $\mathbb{P}^2$. That is, if $\mu^+ = T_1 + T_2$ for $T_1, T_2 \in \mathcal{Q}$, then $T_j = c_j \mu^+$ for some constant $c_j \geq 0$.*

Proof. Suppose that $\mu = c_1 T_1 + c_2 T_2$ where $T_j \in \mathcal{Q}_j$ satisfies $||T_j|| = 1$ and $c_j \geq 0$. Since pushforward and pullback act linearly and preserve positivity, we have by the previous theorem that $f_{+*} T_j$ and $f_+^* T_j$ are dominated by multiples of μ^+. Since μ^+ has no support concentrated on any algebraic curve,

$$T_j = f_{-*}^n f_{+*}^n T_j = (f_+^{n*} f_{+*}^n T_j)|_{\mathbb{P}^2 \setminus \mathcal{C}(f_+^n)} = f_+^{n*} f_{+*}^n T_j.$$

In particular, $1 = ||T_j|| = ||f_+^{n*} f_{+*}^n T_j|| = d^n \cdot ||f_{+*}^n T_j||$.

Furthermore, $\operatorname{supp} f_{+*}^n T_j \subset \operatorname{supp} \mu^+$, and $\tilde{G}^+$ is pluriharmonic over the basin of any superattracting fixed point. Therefore, $\operatorname{supp} d^n f_{+*}^n T_j$ avoids an n–independent neighborhood of any superattracting cycle. This allows us to apply Theorem 4.11 and conclude that

$$\mu^+ = \lim_{n \to \infty} \frac{1}{d^n} f_+^{n*}(d^n f_{+*}^n T_j) = T_j.$$

$\square$

6. Separating Birational Maps

In this section, we consider two other, fairly natural dynamical restrictions to place on a birational map. These restrictions imply first of all that the dynamic domain $\mathcal{D}^+ = \mathbb{P}^2 \setminus \mathcal{I}^+$ for f_+ is large. The restrictions also allow us to draw stronger conclusions about the relationship between $\operatorname{supp} \mu^+$ and the set of normality $\mathcal{N}^+$ for f_+.

Definition 6.1. *A birational map f_+ of $\mathbb{P}^2$ is* **separating** *if $\mathcal{I}^+ \cap \mathcal{I}^- = \emptyset$ (equivalently, $\mathcal{D}^+ \cup \mathcal{D}^- = \mathbb{P}^2$). f_+ is* **completely separating** *if $\mathcal{I}^+ \subset \mathcal{N}^-$.*

Since $\mathcal{I}^+$ and $\mathcal{I}^-$ are closed sets, f_+ is separating if and only if $\operatorname{dist}(\mathcal{I}^+, \mathcal{I}^-) > 0$. Hence a separating map is automatically mildly separating (not to mention minimally separating). Likewise, it is clear from the definition of $\mathcal{N}^+$ that a completely separating birational map is separating. We will give examples in Section 7 that show that mildly separating, separating, and completely separating are all actually distinct hypotheses that one can place on a birational map of $\mathbb{P}^2$.

As is the case with minimally and mildly separating birational maps, a map f_+ is separating if and only if f_- is separating. However, we will give an example in Section 7 to show that f_- need not be completely separating if f_+ is.

Theorem 6.2. *If f_+ is separating, then*

(1) $\mathcal{I}^+ \cap \mathcal{C}^+ = I^+$.
(2) *$\mathcal{I}^+$ contains no algebraic curve.*
(3) *$\mathcal{D}^+$, $\mathcal{D}^-$, and $\mathcal{D}^+ \cap \mathcal{D}^-$ are all (open and) dense in $\mathbb{P}^2$.*

Proof. The proof of the first two statements is similar to the proof of Theorem 5.4. We note that if $p \in \mathcal{I}^+$ and J is the least non-negative integer (possibly non–existent) such that $f_+^J(p) \in I^+$, then $f_+^j(p) \in \mathcal{I}^+$ for $j \leq J$. Thus, if $p \in \mathcal{I}^+ \cap (\mathcal{C}^+ \setminus I^+)$, $f_+(p) \in I^- \cap \mathcal{I}^+ \subset \mathcal{I}^- \cap \mathcal{I}^+$, which contradicts the hypothesis that f_+ is separating. This proves (1).

Now suppose that $V_0 \subset \mathcal{I}^+$ is a non-trivial algebraic curve. Then $V_1 = f_+(V_0) \subset f_+(\mathcal{I}^+) \subset \mathcal{I}^+$. From (1) we see that $V_0 \cap \mathcal{C}^+ \subset I^+$, so by Corollary 3.7 V_1 is a non-trivial algebraic curve, and $f_-(V_1) = V_0$. We continue inductively and obtain curves $V_2, V_3, \cdots \subset \mathcal{I}^+$ such that $f_-(V_{j+1}) = V_j$. Since $V_j \cap I^- = \emptyset$, Corollary 3.7 gives us that

$$\deg(V_0) = (\deg f_-)^j (\deg V_j)$$

for all j. But $\deg V_j \geq 1$ and $\deg f_- \geq 2$, so we obtain a contradiction for large j. Hence V_0 does not exist, and (2) is proved.

Now suppose that $\mathcal{I}^+$ contains an open subset W of $\mathbb{P}^2$. Since $I^+ \subset \mathcal{C}^+$, and I^+_∞ is dense in $\mathcal{I}^+$, there exists a non-negative integer J such that $f_+^J(W) \cap \mathcal{C}^+ \neq \emptyset$. Let J be the smallest such integer. Then $f_+^J(W) \subset \mathcal{I}^+$ is an open subset of $\mathbb{P}^2$. Since I^+ is finite, $f_+^J(W)$ contains a point $p \in \mathcal{C}^+ \setminus I^+$. This contradicts (1) and proves (3) for $\mathcal{D}^+$. The same proof applied to f_- shows that $\mathcal{D}^-$ is open and dense. Thus $\mathcal{D}^+ \cap \mathcal{D}^-$ is also open and dense. $\square$

It would be interesting to know which among the conclusions (1)–(3) in Theorem 6.2 are actually equivalent to the hypothesis that f_+ is separating. It would also be interesting to know whether (2) and (3) might be improved and combined to yield the statement that "$\mathcal{D}^+ \cap V$ is dense in V for every algebraic curve $V \subset \mathbb{P}^2$."

Theorem 6.3. *If f_+ is completely separating and $T \in \mathcal{Q}$ is a current such that $\operatorname{supp} T \subset \operatorname{supp} \mu^+$, then $T = c\mu^+$ for some constant $c \geq 0$.*

Proof. We can assume that $||T|| = 1$. From here the proof is almost the same as the proof of Theorem 5.5, except that we rely on Propositions 4.6 and 4.7 and the fact that $\operatorname{supp} T \cap \mathcal{I}^- \subset \operatorname{supp} \mu^+ \cap \mathcal{I}^- = \emptyset$ to conclude that $\operatorname{supp} f_+^* T, \operatorname{supp} f_{+*} T \subset \operatorname{supp} \mu^+$. We leave the rest of the details to the reader. □

Now we return to the relationship between $\operatorname{supp} \mu^+$ and $\mathcal{N}^+$, beginning with a preliminary result about the rate at which a point can tend to $\mathcal{I}^+$.

Lemma 6.4. *If f_+ is separating, there is a constant C such that $p \notin \mathcal{I}^+$ implies that*

$$\operatorname{dist}(f_+^n(p), \mathcal{I}^+) \geq C^n \operatorname{dist}(p, \mathcal{I}^+).$$

If f_+ is completely separating, then for any open set $W \subset\subset \mathcal{D}^+$ there exists a constant $C = C(W)$ such that

$$\operatorname{dist}(f_+^n(W), \mathcal{I}^+) \geq C$$

for all n.

Proof. Suppose that f_+ is separating. Suppose that $q \in \mathbb{P}^2$ satisfies $\operatorname{dist}(q, \mathcal{I}^+) \leq D/2$, where $D = \operatorname{dist}(\mathcal{I}^+, \mathcal{I}^-) > 0$. Then by Propositions 3.6 and 3.7, there is a constant $C = C(f_+)$ such that $\operatorname{dist}(f_-(q), \mathcal{I}^+) \leq \operatorname{dist}(f_-(q), f_-(\mathcal{I}^+)) \leq C \operatorname{dist}(q, \mathcal{I}^+)$. Choosing a different point $p \notin \mathcal{I}^+$, suppose that $\operatorname{dist}(f_+(p), \mathcal{I}^+) \leq D/2$. Since $f_+(\mathcal{I}^-) \subset \mathcal{I}^-$, we see that $p \notin \mathcal{I}^-$. Therefore,

$$\operatorname{dist}(p, \mathcal{I}^+) = \operatorname{dist}(f_-(f_+(p)), \mathcal{I}^+) \leq C \operatorname{dist}(f_+(p), \mathcal{I}^+)$$

Thus either $\operatorname{dist}(f_+(p), \mathcal{I}^+) \geq D/2$ or $\operatorname{dist}(f_+(p), \mathcal{I}^+) \geq C' \operatorname{dist}(p, \mathcal{I}^+)$. This observation plus compactness of $\mathbb{P}^2$ yields the first claim.

Now suppose that f_+ is completely separating and choose an open set W as in the second claim. If C as in the claim does not exist, then there are points $p_j \in W$ and positive integers n_j such that $\lim_{j\to\infty} \operatorname{dist}(f_+^{n_j}(p_j), \mathcal{I}^+) = 0$. Set $p_j' = f_+^{n_j}(p_j)$. By passing to a subsequence and invoking compactness of $\mathcal{I}^+$, we can arrange that $\lim_{j\to\infty} p_j' = p \in \mathcal{I}^+$. Applying the hypothesis that f_+ is completely separating and passing to a further subsequence, we can also find a neighborhood W_p of p and a holomorphic map $g : W_p \to \mathbb{P}^2$ such that $f_-^{n_j} \to g$ uniformly on W_p. Thus, we have for large j that

$$0 < \operatorname{dist}(W, \mathcal{I}^+) \leq \lim_{j\to\infty} \operatorname{dist}(f_-^{n_j}(p_j'), f_-^{n_j}(p)) = \lim_{j\to\infty} \operatorname{dist}(g(p_j'), g(p)) = 0.$$

This contradiction proves the second claim. □

Lemma 6.5. *Let W be a simply connected complex manifold and $\varphi : W \to \mathbb{P}^2$ be a holomorphic map. Suppose that $\varphi(W) \cap \operatorname{supp} \mu^+ = \emptyset$. Then for any $t \in \mathbb{R}$, there is a holomorphic map $\tilde{\varphi} = \tilde{\varphi}_t : W \to \{\tilde{G}^+ = t\} \subset \mathbb{C}^3 \setminus \{0\}$ such that $\pi \circ \tilde{\varphi} = \varphi$. The map $\tilde{\varphi}$ is unique up to multiplication by a unitary constant.*

We will refer to the map $\tilde{\varphi}$ as a *lift* of φ into $\{\tilde{G}^+ = t\}$. The hypothesis that $\varphi(W)$ avoids $\operatorname{supp}\mu^+$ is a bit stronger than necessary; it is enough to assume that $\tilde{G}^+ \circ \pi^{-1} \circ \varphi$ is pluriharmonic (where it is defined) for any local holomorphic right inverse π^{-1} of π.

Proof. We first observe that if $W' \subset W$ is open and $\tilde{\varphi}, \tilde{\varphi}' : W' \to \{\tilde{G}^+ = t\}$ are two holomorphic maps satisfying $\pi \circ \tilde{\varphi} = \pi \circ \tilde{\varphi}' = \varphi$, then homogeneity of $\tilde{G}^+$ gives us that $\tilde{\varphi} = e^{i\theta}\tilde{\varphi}'$, where θ is a real function. Since θ must also be holomorphic, we conclude that θ is constant. The uniqueness part of the lemma follows.

Uniqueness up to constant multiples and simple connectivity of W also tell us that it is enough to construct $\tilde{\varphi}$ locally and piece together the results. So cover W with simply connected open sets W_j small enough that for each j, there exists a holomorphic right inverse π_j^{-1} of π defined on a neighborhood of $\varphi(W_j)$. By the hypothesis that $\varphi(W)$ avoids $\operatorname{supp}\mu^+$, we have that $G = \tilde{G}^+ \circ \pi_j^{-1} \circ \varphi$ is pluriharmonic. Let G^* be a pluriharmonic conjugate of G, and set $H = te^{-G-iG^*}$. We then define

$$\tilde{\varphi} = H \cdot (\pi^{-1} \circ \varphi).$$

One can check directly that $\tilde{\varphi}$ satisfies the conclusion of the lemma on W_j. $\square$

Theorem 6.6. *Let f_+ be a completely separating birational map. Then* $\operatorname{supp}\mu^+ = \mathbb{P}^2 \setminus \mathcal{N}^+$.

Proof. From Theorem 4.10, we know that $\operatorname{supp}\mu^+$ and $\mathcal{N}^+$ are disjoint. We need to show that these sets are complementary in $\mathbb{P}^2$. Suppose that $W \subset\subset \mathbb{P}^2 \setminus \operatorname{supp}\mu^+$ is open. Normality is a local property, so we can assume that W is simply connected. By assumption, $\tilde{G}^+$ is pluriharmonic on $\pi^{-1}(W)$. So we can choose a holomorphic lift $\tilde{i} : W \to \{\tilde{G}^+ = 0\}$ of the inclusion map $i : W \hookrightarrow \mathbb{P}^2$.

Let $\{n_j\}$ be an increasing sequence of positive integers. We need to show that $\{f_+^{n_j}\}$ has a convergent subsequence. For all $p \in W$ and all j, we have

$$|\log||\tilde{f}_+^{n_j} \circ \tilde{i}(p)||\,| = |\log||\tilde{f}_+^{n_j} \circ \tilde{i}(p)|| - \tilde{G}^+ \circ \tilde{f}_+^{n_j} \circ \tilde{i}(p)| \leq C$$

The equality follows from the fact that $\{\tilde{G}^+ = 0\}$ is invariant under $\tilde{f}_+$; the inequality follows from Lemma 6.4 and the estimate (5–1) from Theorem 5.2. Thus

$$\tilde{f}_+^{n_j} \circ \tilde{i}(W) \subset \{e^{-C} \leq ||\tilde{p}|| \leq e^C\}.$$

After passing to a subsequence, we have $\tilde{f}_+^{n_j} \circ \tilde{i} \to \tilde{g} : W \to \mathbb{P}^2$ uniformly on compact subsets, and the image $\tilde{g}(W)$ lies in the same spherical shell. It follows that $f_+^{n_j} \to \pi \circ \tilde{g}$ uniformly on compact subsets of W. $\square$

We remark that in the terminology of Fornaess and Sibony [FS1], Lemma 6.4 states that a completely separating birational map is "normal." Our proof of Theorem 6.6 mimicks proofs of similar theorems in [FS1] and [Ue] for maps of $\mathbb{P}^2$ with empty indeterminacy sets.

Corollary 6.7. *If f_+ is completely separating and $T \in \mathcal{Q}$ is not a multiple of μ^+, then* $\operatorname{supp} T$ *intersects $\mathcal{N}^+$.*

Proof. Apply Theorems 6.3 and 6.6. □

We do not as yet know whether Theorem 6.6 holds for separating birational maps, but we can show that $\operatorname{supp} \mu^+$ contains certain dynamically interesting points in $\mathbb{P}^2 \setminus \mathcal{N}^+$.

Definition 6.8. *Suppose that $p \in \mathbb{P}^2$ satisfies $f_+^k(p) = p$; let λ_1, λ_2 denote the eigenvalues of Df_+^k at p. Then p is* **repelling** *if $|\lambda_1|, |\lambda_2| > 1$, and p is* **saddle** *if $0 < |\lambda_1| < 1 < |\lambda_2|$.*

Note that if p is a repelling or saddle periodic point according to our definition, then $p \notin I_\infty^+ \cup I_\infty^-$. It seems likely that the results that follow can be restated to admit a definition which includes periodic critical points, but for simplicity we choose to exclude the possibility.

In what remains of this section, we will show that if f_+ is separating, then $\operatorname{supp} \mu^+$ contains both repelling and saddle periodic points. We handle the case of repelling periodic points first.

Theorem 6.9. *Let f_+ be a separating birational map. Any repelling periodic point of f_+ belongs to* $\operatorname{supp} \mu^+$.

A key ingredient in the proof of this theorem is the existence of a normal form (see [DE] Theorem 6.1) for an analytic automorphism about a repelling fixed point.

Normal Form Theorem. *Suppose that $f : W_1 \to W_2$ is a biholomorphism between open subsets of $\mathbb{C}^2$ and that $p \in W_1 \cap W_2$ is a repelling fixed point. Then there exists a choice of coordinates $\varphi : B_0(r) \to W_1' \subset W_1$ and a polynomial automorphism $P : \mathbb{C}^2 \to \mathbb{C}^2$ such that*

(1) *$\varphi(0,0) = p$;*
(2) *$\varphi \circ P = f \circ \varphi$;*

P can be chosen to be a shear—that is, $P(x,y) = (\alpha x, \beta y + \gamma(x))$, where $|\alpha|, |\beta| > 1$, and γ is a polynomial in x such that $\gamma(0) = 0$.

We refer to the map φ as the *normalizing map* for f about p, and we refer to P as the *normal form*.

Corollary 6.10. *If f_+ is a birational map, and $p \in \mathcal{D}^+$ is a repelling fixed point, then the normalizing map φ for f_+ about p extends to a holomorphic map of $\mathbb{C}^2$ into $\mathcal{D}^+$ such that*

(1) *$\varphi \circ P = f_+ \circ \varphi$;*
(2) *The critical set of φ is a one dimensional analytic set V given by $V = \varphi^{-1}(I_\infty^-)$;*
(3) *φ is injective on $\mathbb{C}^2 \setminus V$.*

Proof. Let $\varphi : B_0(r) \to W_1'$ be the normalizing map. Given $(x, y) \in \mathbb{C}^2$, choose n so that $P^{-n}(x, y) \in B_0(r)$ (one can show that $(0, 0)$ attracts all of $\mathbb{C}^2$ under iteration of the shear P given in the Normal Form Theorem, so n exists; the next Lemma addresses a similar issue). We define

$$\varphi(x, y) = f_+^n \circ \varphi \circ P^{-n}(x, y).$$

Provided we choose r small enough $W_1' \subset \mathcal{D}^+$. Thus $f_+^n(W_1') \subset f_+^n(\mathcal{D}^+) \subset \mathcal{D}^+$ implies that the right side of the equation defining $\varphi(x, y)$ makes sense. Property (2) in the Normal Form Theorem implies that the definition is independent of the choice of n and that it agrees with the definition of φ on $B_0(r)$.

Differentiating (1) gives

$$(D\varphi \circ P) \cdot DP = (Df_+ \circ \varphi) \cdot D\varphi.$$

This equation combined with the facts that φ is injective on $B_0(r)$, that DP is everywhere non-singular, and that $f_+(\mathcal{C}^+) = I^-$ give us (2). Finally, since φ is injective near $(0, 0)$, and f_+ is injective away from $\mathcal{C}^+$, we have that φ is injective off V. $\square$

Except in certain special cases, one can choose the normal form to be linear. Were the normal form *always* linear, the next technical lemma would be unnecessary.

Lemma 6.11. *The automorphism P guaranteed by the Normal Form Theorem can be chosen so that there exist constants $A, B > 0$ such that*

$$||P^{-n}(x, y)|| \leq 1$$

when $||(x, y)|| \geq A$ and $n \geq B \log ||(x, y)||$.

Proof. We can at least choose P to be a shear. In this case P^{-1} is a shear of the same form, and DP_0^{-1} is conjugate to Df_p^{-1}. That is,

$$P^{-1}(x, y) = (\alpha x, \beta y - \gamma(x))$$

where $|\alpha|, |\beta| < 1$ and $\gamma(0) = 0$. Let $1/2 > \epsilon > 0$ be small enough that $|\gamma(x)| < (1 - |\beta|)/2$ when $|x| < \epsilon$.

Given $||(x, y)||$, let $(x_n, y_n) = P^{-n}(x, y)$. One can check that if

$$n_1 \geq D \log |x|$$

for D large enough, then $|x_n| = |\alpha^n x| \leq \epsilon$ for all $n \geq n_1$. Note that if $||(x, y)|| > A$ for A large enough, then

$$|y_{n_1}| \leq |y| + n_1 |\gamma(x)| \leq E||(x, y)||^{k+1},$$

where $k = \deg \gamma$. If $n > n_1$ and $||(x_n, y_n)|| > 1$, then $|y_n| > 1/2$ and

$$|y_{n+1}| \leq |\beta y_n| + (1 - |\beta|)/2 \leq \delta |y_n|$$

for some $\delta < 1$. So if n_2 is chosen so that $\delta^{n_2}|y_{n_1}| < 1/2$, we will have $||(x_n, y_n)|| < 1$ for all $n \geq n_1 + n_2$. But our observations imply that we can arrange

$$\begin{aligned} n_1 + n_2 &\leq D \log |x| + E \log |y_{n_1}| \\ &\leq D \log |x| + F \log(E||(x,y)||^{k+1}) \leq B \log ||(x,y)|| \end{aligned}$$

for $||(x, y)|| > A$. $\square$

Proof of Theorem 6.9. We have automatically that $\mathcal{I}^+ \subset \operatorname{supp} \mu^+$, so we can assume that $p \in \mathcal{D}^+$. By passing to an iterate of f_+, we can assume that p is a fixed point. Let φ be the normalizing map for f_+ about p and P be the normal form. Assuming $p \notin \operatorname{supp} \mu^+$, let $\tilde{\varphi} : \mathbb{C}^2 \to \{\tilde{G}^+ = 0\}$ be a lift of φ. From Theorem 5.2 and Lemma 6.4, we obtain

$$\begin{aligned} \log ||\tilde{\varphi} \circ P^n(x,y)|| &= \log ||\tilde{f}_+^n \circ \tilde{\varphi}|| \\ &= \log ||\tilde{f}_+^n \circ \tilde{\varphi}|| - \tilde{G}^+ \circ \tilde{f}_+^n \circ \tilde{\varphi} \\ &\leq C - D \log \operatorname{dist}(f_+^n \circ \varphi(x,y), \mathcal{I}^+) \\ &\leq C - D \log[C^n \operatorname{dist}(\varphi(x,y), \mathcal{I}^+)]. \end{aligned}$$

We assume that P is as in Lemma 6.11 and set $m = \min_{||(x,y)||<1} \operatorname{dist}(\varphi(x,y), \mathcal{I}^+)$. For $||(x, y)||$ large enough, we let n be the smallest integer such that $||P^{-n}(x, y)|| \leq 1$. Then using the estimate displayed above, we have

$$\begin{aligned} ||\tilde{\varphi}(x,y)|| &= ||\tilde{\varphi} \circ P^n(P^{-n}(x,y))|| \\ &\leq D^n \text{ (where } D = D(m)) \\ &\leq ||(x,y)||^k \quad \text{(by Lemma 6.11)} \end{aligned}$$

for some constant $k = k(m)$ independent of n. Liouville's Theorem then implies that all coordinate functions of $\tilde{\varphi}$ are polynomials. Hence φ extends to a rational map $\varphi : \mathbb{P}^2 \to \mathbb{P}^2$. φ is generically injective on $\mathbb{C}^2$, so the extended map is in fact birational. Thus

$$f_+ = \varphi \circ P \circ \varphi^{-1}.$$

From this it is clear that the dynamic degrees of f_+ and P agree. Since P is a shear, $\deg P^n = \deg P$ for all n, and we have by definition that f_+ is not even minimally separating. This contradicts our assumption that f_+ is separating, and leads us to conclude that $p \in \operatorname{supp} \mu^+$. $\square$

Now we turn to saddle periodic points of f_+. The following result from the theory of smooth dynamical systems (see for example Theorem 6.2.3 in [HK]) plays the role for saddle periodic points that the Normal Form Theorem plays for repelling periodic points

Stable Manifold Theorem. *Let p be a fixed saddle point of a holomorphic diffeomorphism $f : W_1 \to W_2$ between neighborhoods $W_1, W_2 \subset \mathbb{C}^2$ of p. Let λ_s be the stable (i.e. $|\lambda_s| < 1$) eigenvalue and v_s the corresponding eigenvector of Df at p. Then there exists a neighborhood W of p and a proper holomorphic embedding φ^+ of the unit disk $\Delta = \{|z| < 1\}$ into W such that*

(1) *v_s is tangent to $f(\Delta)$ at p;*
(2) *$f \circ \varphi^+(z) = \varphi^+(\lambda_s z)$;*
(3) *$\varphi^+(\Delta) = \{q \in W : f^n(q) \in W$ for all $n\}$;*

The set $\varphi^+(\Delta)$ is called the *local stable manifold* of p. Note that by (2), the local stable manifold is forward invariant under f.

Corollary 6.12. *Let f_+ be a birational map of $\mathbb{P}^2$. Suppose that p is a fixed saddle point, and $\varphi^+ : \Delta \to \mathbb{C}^2$ is the parametrization of the local stable manifold of p. Then φ^+ extends to a holomorphic map of $\mathbb{C}$ into $\mathbb{P}^2$ satisfying the following conditions.*

(1) *$f_+ \circ \varphi^+(z) = \varphi^+(\lambda_s z)$ for all $z \in \mathbb{C}$.*
(2) *$\varphi^+(\mathbb{C}) \setminus I_\infty^+ = \{q \in \mathbb{P}^2 \setminus I_\infty^+ : \lim_{n\to\infty} f_+^n(q) = p\}$.*
(3) *Let $N = \cup_{n=1}^\infty (\varphi^+)^{-1}(f_-^n(\mathcal{C}^+))$. Then N is discrete, and $\varphi^+|_{\mathbb{C}\setminus N}$ is an immersion.*

Note that we do not require the saddle point in this proposition to lie in $\mathcal{D}^+$ or $\mathcal{D}^-$. We refer to $W^s(p) = \varphi^+(\mathbb{C})$ as the *(global) stable manifold* of f_+ at p. Clearly, we can apply the Stable Manifold Theorem and Corollary 6.12 to f_- to obtain a *(global) unstable manifold* $W^u(p)$ with parametrization $\varphi^- : \mathbb{C} \to W^u(p)$ for a fixed saddle point p.

Proof. We extend φ^+ in the same way we extended the normalizing map for a repelling fixed point. Given $z \in \lambda_S^{-n} \cdot \Delta$, we set

$$\varphi^+(z) = f_-^n \circ \varphi^+(\lambda_S^n z).$$

By Proposition 2.2, there is no ambiguity on the right side even if $\varphi^+(\lambda_S^n z) \in I_\infty^-$. Nor does the right side depend on the choice of n. (1) follows immediately.

Any point q attracted to p will satisfy $f_+^n(q) \in W$ for n large enough and W as in the Stable Manifold Theorem. Hence, (2) follows from statement (1) in this theorem and statement (3) in the Stable Manifold Theorem.

Now let N be as in statement (3). If N is not discrete, then $\varphi^+(\mathbb{C})$ is a subset of an irreducible component of $f_-^n(\mathcal{C}^+)$ for some n. But then $f_+^{n+1}(\varphi^+(\mathbb{C})) \subset I^-$, contradicting the fact that by definition a saddle periodic point cannot lie in I^-. We apply the facts that $\varphi^+|_\Delta$ is an embedding, that $\cup f_-^n(\varphi^+(\Delta)) = \varphi^+(\mathbb{C})$, and that $f_- : \mathbb{P}^2 \setminus \mathcal{C}^- \to \mathbb{P}^2 \setminus \mathcal{C}^+$ is a biholomorphism to obtain the rest of item (3). $\square$

Theorem 6.13. *Let f_+ be a separating birational map. Any periodic saddle point p of f_+ belongs to* $\operatorname{supp}\mu^+$. *Moreover, if $p_+ \in \mathcal{D}^+$, then the (closure of the) stable manifold of p belongs to* $\operatorname{supp}\mu^+$.

The next lemma and its proof are very similar in spirit to Theorem 6.9.

Lemma 6.14. *Suppose that f_+ is separating and $p \in \mathcal{D}^+$ is a fixed saddle point. Let $\pi^{-1} : W \to \mathbb{C}^3$ be any holomorphic right inverse of π on a neighborhood W of p, and set $\Delta_\epsilon = \{|z| < \epsilon\}$ for any $\epsilon > 0$. Then $\tilde{G}^+ \circ \pi^{-1} \circ \varphi^-$ restricted to Δ_ϵ is not harmonic.*

Proof. Suppose the lemma is untrue for some choice of π^{-1} and ϵ. First note that since $p \in \mathcal{D}^+$, we also have that $\varphi^-(\Delta_\epsilon) \in \mathcal{D}^+$ for small enough ϵ. Since $W^u(p) = \cup f_+^n \circ \varphi^-(\Delta_\epsilon)$, we see that $W^u(p) = \varphi^-(\mathbb{C}) \subset \mathcal{D}^+$. Furthermore, the transformation property $\tilde{G}^+ \circ \tilde{f}_+ = d \cdot \tilde{G}^+$ implies that $\tilde{G}^+$ is pluriharmonic on all of $\pi^{-1}(W^u(p))$—that is, if π^{-1} is any holomorphic right inverse of π defined on any open set $W' \subset \mathbb{P}^2$, then $\tilde{G}^+ \circ \pi^{-1} \circ \varphi^-$ is harmonic wherever it is defined. So by Lemma 6.5 and the remark following, we can choose a lift $\tilde{\varphi}^- : \mathbb{C} \to \{\tilde{G}^+ = 0\}$ of φ^-.

Working exactly as we did in the proof of Theorem 6.9, we conclude that the coordinate functions of $\tilde{\varphi}^-$ are polynomials. Hence, φ^- extends to a holomorphic map $\varphi^- : \mathbb{P}^1 \to \mathbb{P}^2$. Since φ^- is generically injective, we conclude that φ^- regularizes the algebraic curve $V = \overline{W^u(p)} = \varphi^-(\mathbb{P}^1)$.

Since $W^u(p)$ is invariant under f_+ and f_-, so is V. By Corollary 3.7, V contains points of both I^+ and I^-. As we showed above, $W^u(p) \cap I^+$ is empty, so $V \cap I^+$ must consist of the single point $\varphi^-(\infty)$. Thus any intersection between I^- and V must be a point $p^- = \varphi^-(z)$ where $z \neq 0, \infty$. But then $f_+^n(p^-) = \varphi^-(\lambda_u^n z) \to \varphi^-(\infty) \in I^+$. That is, $\varphi^-(\infty) \in \mathcal{I}^+ \cap \mathcal{I}^-$, which violates the separation assumption. This contradiction proves the lemma. □

Proof of Theorem 6.13. If $p \in \mathcal{I}^+$, then $p \in \operatorname{supp}\mu^+$ simply because poles of $\tilde{G}^+$ are dense in $\pi^{-1}(\mathcal{I}^+)$. So we can assume that $p \in \mathcal{D}^+$.

Suppose that there exists $q = \varphi^+(z_0) \notin \operatorname{supp}\mu^+$. By invariance of $\operatorname{supp}\mu^+$, we have $f_+^n(q) \notin \operatorname{supp}\mu^+$ for all n. Since $f_+^n(q) \to p$, we can assume that q lies in an arbitrarily small neighborhood of p. After a choice coordinates in a small neighborhood we can assume that $p = (0,0)$ and that q lies in $W = W_1 \times W_2$ for neighborhoods $W_1, W_2 \subset \mathbb{C}$. Given $(x_n, y_n) \in W$, let $(x_{n+1}, y_{n+1}) = f_+(x_n, y_n)$ provided the latter lies in W. Fix small numbers $\epsilon > 0, \delta > 0$. We can use the facts that p is a saddle point and that derivatives of f are continuous to arrange our coordinates so that

(1) $\varphi^- : W_2 \to W^u(p)$ is given by $\varphi^-(z) = (0, z)$;
(2) $W^s(p) \cap W = \{(z, 0) : z \in W_1\}$;
(3) $|x_{n+1}| < (1+\epsilon)\lambda_s |x_n| < |x_n|$ for all $(x_n, y_n) \in W$ such that $f_+(x_n, y_n) \in W$;
(4) likewise, $|y_{n+1}| > (1-\epsilon)\lambda_u |y_n| > |y_n|$;

(5) If $D_n = \{(g_n(z), z) : z \in W_1\}$ is the graph of of a holomorphic function $g_n : W_1 \to W_2$ such that $|g_n'| < \delta$, then $D_{n+1} = f_+(D_n) \cap W$ is the graph of a function g_{n+1} with the same properties.

Note in (2) that $z \to (z, 0)$ is not necessarily the parametrization of W^s given by the Stable Manifold Theorem. The first two conditions are arranged by choice of coordinates, and the last three by shrinking W_1 and W_2.

Now suppose that $q = (x_0, 0)$, and let D_0 be the vertical slice obtained, as in (5), as the graph of $g_0(z) = x_0$. Then $D_n = f_+^n(D_n) \cap W$ is realized as the graph of another holomorphic function $g_n : W_1 \to W_2$. From (2), we have that $f^n(q) \in D_n$. From (3), we have that

$$\lim_{n\to\infty} g_n(z) = 0$$

uniformly in z. Now let π^{-1} be a holomorphic right inverse of π on W, and set

$$G = \tilde{G}^+ \circ \pi^{-1}(0, z)$$
$$G_n = \tilde{G}^+ \circ \pi^{-1}(g_n(z), z).$$

Then $G_n \to G$ uniformly in z. By Lemma 6.14, G is not harmonic, so for large enough n, neither is G_n. For these n, we obtain points $q_n \in \operatorname{supp} \mu^+ \cap D_n$.

On the other hand, condition (4) shows that $f_-^n(D_n)$ decreases to the singleton $D_0 \cap W^s = \{q\}$. In particular, we invoke invariance of $\operatorname{supp} \mu^+$ again to conclude that $q = \lim_{n\to\infty} f_-^n(q_n) \in \operatorname{supp} \mu^+$. This contradicts our starting assumption and proves that $W^s(p) \subset \operatorname{supp} \mu^+$. $\square$

7. Polynomial Maps: An Example

Any birational map of $\mathbb{P}^2$ restricts to a map of $\mathbb{C}^2$ whose coordinate functions are quotients of polynomials. An interesting subset of birational maps is singled out in the following definition.

Definition 7.1. *A birational map f_+ of $\mathbb{P}^2$ is* **polynomial** *if the restriction of f_+ to $\mathbb{C}^2$ can be written $f_+(x, y) = (f_1(x, y), f_2(x, y))$ where f_1, f_2 are polynomial functions of x and y.*

Since we are only interested in birational maps here, we will generally shorten "polynomial birational" to "polynomial" in what follows. Any polynomial automorphism (e.g. generalized Hénon maps; see [FM]) of $\mathbb{C}^2$ is a polynomial map. Automorphisms are rather special in that their inverses are also polynomial maps. An example of a polynomial map which does not have a polynomial inverse is

$$f_+(x, y) = (y, 2x - 4xy + 3y^2).$$

This map and its dynamics were considered by Fatou [Fa]. By linearizing f_+ about one of its two repelling fixed points, he gave the first example of a self-map of $\mathbb{C}^2$ whose image contained an open set but was not dense in $\mathbb{C}^2$.

In the next several results, f_+ will be a polynomial map of degree at least two. Recall that we use ℓ_∞ to denote the line at infinity $\mathbb{P}^2 \setminus \mathbb{C}^2$ in $\mathbb{P}^2$.

Proposition 7.2. *$f_+(\ell_\infty)$ is a single point $p_\infty^- \in \ell_\infty \cap I^-$.*

Proof. Write $f_+ = (f_1, f_2)$, where f_1, f_2 are polynomials in x and y. Let $\tilde{f}_j$ be the homogeneous polynomial obtained from f_j by including extra factors of z in each term of degree less than $d = \deg f_+ = \max\{\deg f_1, \deg f_2\}$. Then $\tilde{f}_+$ has the form

$$\tilde{f}_+ = (\tilde{f}_1, \tilde{f}_2, z^d).$$

By assumption on d, we can assume that either $\tilde{f}_1(x, y, 0)$ or $\tilde{f}_2(x, y, 0)$ is not identically zero—say for instance that $\tilde{f}_2$ is non-zero. From this it is clear that f_+ maps ℓ_∞ to itself (i.e. $\tilde{f}_+$ preserves the hyperplane $\{z = 0\}$). Note that we obtain a formula for f_+ in local coordinates by normalizing so that the second coordinate is 1 in both the domain and range:

$$f_+[x : 1 : z] = \left[\frac{\tilde{f}_1(x, 1, z)}{\tilde{f}_2(x, 1, z)} : 1 : \frac{z^d}{\tilde{f}_2(x, 1, z)}\right]$$

By computing Df_+ relative to x and z, and setting $z = 0$ in the result, it is easy to see that $\ell_\infty \subset \mathcal{C}^+$. We apply (1) of Proposition 3.3 to conclude that $f_+(\ell_\infty) = p_\infty^- \in \ell_\infty \cap I^-$. $\square$

Proposition 7.3. *The following statements are true for f_+:*

(1) $I^- \cap \ell_\infty = \{p_\infty^-\}$;
(2) $I^+, I_\infty^+, \mathcal{I}^+ \subset \ell_\infty$;
(3) $I_\infty^+ \cap I^- = I^+ \cap I^- \subset \{p^-\}$;

Furthermore, if $p_\infty^- \notin I^+$, then

(4) $I^+ = I_\infty^+ = \mathcal{I}^+$;
(5) $\mathcal{I}^+ \cap \mathcal{I}^- \subset I^+$.

Proof. If $p^- \neq p_\infty^-$ is another point in I^-, then there is a component $V \subset \mathcal{C}^+$ such that $V \neq \ell_\infty$ and $f_+(V) = p^-$. Since $f_+(\mathbb{C}^2) \subset \mathbb{C}^2$, it must be that $p^- \in \mathbb{C}^2$. Hence (1). (2) follows from the facts that f_+ is well-defined at all points in $\mathbb{C}^2$ and $f_+(\mathbb{C}^2) \subset \mathbb{C}^2$. (3) follows from (1) and (2).

If $p_\infty^- \notin I^+$, then by the previous proposition and (2) $f_-(I^+) = f_-(\ell_\infty)$ is a single point in I^+. (4) and (5) follow immediately. $\square$

Theorem 7.4. *The following statements are equivalent for a polynomial map f_+ with $\deg f_+ \geq 2$:*

(1) $p_\infty^- \notin I^+$;
(2) *f_+ is minimally separating;*
(3) *f_+ is mildly separating.*

Proof. It is clear from Corollary 3.4 that (2) implies (1). Certainly (3) implies (2). So assuming that (1) is true, we need to show that (3) holds. Recall that f_+ is mildly

separating if and only if $\tilde{G}^+(\tilde{p}) > -\infty$ whenever $\pi(\tilde{p}) \in I^-$. If $p_\infty^- \notin I^+$, then p_∞^- is a fixed point of $\tilde{f}_+$. After change of coordinates, we can assume $p_\infty^- = [0:1:0]$. Hence, $\tilde{f}_+(0,y,0) = (0, Cy^d, 0)$ for some constant C. From this it is easy to show that $\tilde{G}^+(\tilde{p}) > -\infty$ when $\pi(\tilde{p}) = p_\infty^-$.

All other points in I^- lie in $\mathbb{C}^2$, so we will finish the proof by showing that $\tilde{G}^+(\tilde{p}) > -\infty$ when $\pi(\tilde{p}) \in \mathbb{C}^2$. To this end, note that

$$\begin{aligned}\tilde{G}^+(x,y,z) &= \log|z| + \tilde{G}^+(x/z, y/z, 1)\\ &= \log|z| + \lim_{n\to\infty} \frac{1}{d^n} \log||\tilde{f}_+^n(x/z, y/z, 1)||\\ &= \log|z| + \lim_{n\to\infty} \frac{1}{2d^n} \log(||f_+^n(x/z, y/z)||^2 + 1)\\ &> -\infty\end{aligned}$$

as long as $z \neq 0$. $\square$

The really useful consequence of this theorem is that when a polynomial map is minimally separating, the *inverse* of the map has a relatively well-behaved escape function $\tilde{G}^-$. As we showed in the proof of the theorem, it is not difficult to show directly that $\tilde{G}^+$ is good. In the next theorem, we show that $\tilde{G}^+$ behaves quite similarly to the escape function of a polynomial automorphism (see for comparison Section 3 of [BS1]).

Theorem 7.5. *Assume that $p_\infty^- \notin I^+$. Then p_∞^- is superattracting for f_+. Furthermore, if K^+ is the set of points that do not tend to p_∞^- under iteration, then*

(1) $K^+ \cap \ell_\infty = I^+$;
(2) *If $\tilde{f}_+$ is scaled so that the third component of $\tilde{f}_+$ is z^d, then $\tilde{G}^+(x,y,1) \geq 0$, with equality holding if and only if $(x,y) \in K^+$;*
(3) $\operatorname{supp} \mu^+ = bK^+$;

If coordinates are chosen so that $p_\infty^- = [0:1:0]$, then $\lim_{y\to\infty} \tilde{G}^+(x,y,1) - \log|y|$ is finite and constant with respect to x.

Proof. We can assume without loss of generality that $p_\infty^- = [0:1:0]$, and that

$$\tilde{f}_+(x,y,z) = (z\tilde{f}_1(x,y,z), \tilde{f}_2(x,y,z), z^d)$$

for homogeneous polynomials $\tilde{f}_1$ and $\tilde{f}_2$ such that $\deg \tilde{f}_1 = d-1$, $\deg \tilde{f}_2 = d$ and $\tilde{f}_2(0,y,0) = Cy^d$ for some $C \neq 0$. After passing to local coordinates by setting $y = 1$ (as in the proof of Proposition 7.2), it is not hard to show that $Df_+^2 = 0$ at $[0:1:0]$. Hence all eigenvalues of Df_+ vanish and $[0:1:0]$ is superattracting.

(1) follows from Proposition 7.2 and (2) of Proposition 7.3. To see that (2) holds, note first of all that $(x,y) \in K^+$ implies that there exists $\epsilon > 0$ such that for all $n > 0$,

$$\begin{aligned}(x_n, y_n) = f_+^n(x,y) &\notin \{[x:1:z] \in \mathbb{P}^2 : |x|, |z| < \epsilon\} \cap \mathbb{C}^2\\ &= \{(x,y) \in \mathbb{C}^2 : |y| > \epsilon^{-1}, |x| < \epsilon|y|\}.\end{aligned}$$

That is,

$$||(x_{n+1}, y_{n+1})|| < (\epsilon^{-1}+1)|x_{n+1}| = (\epsilon^{-1}+1)|f_2(x,y)| \leq C(\epsilon)||(x_n, y_n)||^{d-1}$$

for all n. Hence, by induction

$$\begin{aligned}\tilde{G}^+(x,y,1) &= \lim_{n\to\infty} \frac{1}{2d^n}\log(1+||(x_n,y_n)||^2)\\ &\leq \lim_{n\to\infty}\frac{1}{2d^n}\log(1+(C||(x,y)||)^{(d-1)^n}) = 0.\end{aligned}$$

On the other hand, suppose that $(x,y) \notin K^+$. Then $[x_n : y_n : 1] = [x_n/y_n : 1 : 1/y_n] \to [0:1:0]$. Since iterates of f_+ form a normal family on the basin of $[0:1:0]$, $\tilde{G}^+$ is pluriharmonic, and hence continuous on $\pi^{-1}(\mathbb{P}^2 \setminus K^+)$. Thus,

$$\begin{aligned}\lim_{n\to\infty}\tilde{G}^+\circ\tilde{f}_+^n(x,y,1) &= \lim_{n\to\infty}\tilde{G}^+(x_n,y_n,1)\\ &= \lim_{n\to\infty}\tilde{G}^+(x_n/y_n,1,1/y_n)+\log|y_n|\\ &= \tilde{G}^+(0,1,0)+\lim_{n\to\infty}\log|y_n| = \infty.\end{aligned}$$

So if we pick n large enough that $\tilde{G}^+(x_n,y_n,1) > 0$, we see that $\tilde{G}^+(x,y,1) = d^{-n}\tilde{G}^+(x_n,y_n,1) > 0$. We have finished proving (2).

Since $\tilde{G}^+(x,y,1)$ is 0 on K^+ and pluriharmonic on $\mathbb{C}^2 \setminus K^+$, we have $\operatorname{supp}\mu^+ \subset bK^+$. On the other hand, if $(x_0,y_0) \in K^+$, and $\tilde{G}^+(x,y,1)$ is pluriharmonic in a neighborhood W of (x_0,y_0), then $\tilde{G}^+(x,y,1) \equiv 0$ on W because $\tilde{G}^+(x_0,y_0,1) = 0$ is an interior minimum. This can only happen if $(x_0,y_0,1)$ belongs to the interior of K^+. We conclude that $bK^+ \subset \operatorname{supp}\mu^+$, and (3) is proved.

Finally, for any fixed value of x, we have

$$\lim_{y\to\infty}\tilde{G}^+(x,y,1)-\log|y| = \lim_{y\to\infty}\tilde{G}^+(x/y,1,1/y) = \tilde{G}^+(0,1,0)$$

which proves the last statement of the theorem. □

For the sake of providing more concrete examples we turn to polynomial maps of degree two. These have been characterized by Nishimura [Ni] who proved

Theorem 7.6. *Any minimally separating polynomial map of degree two is conjugate to a map whose restriction to $\mathbb{C}^2$ has the form*

$$f_+(x,y) = (y, ax+bxy+y^2+cy+d), \tag{7–1}$$

where either a or b (or possibly both) are non-zero.

The normal form given by this theorem is consistent with Theorem 7.5 in that $p_\infty^- = [0:1:0]$. We remark that a representative of the form (7–1) is not unique. In particular, either c or d can usually be eliminated from (7–1) by further conjugation. We also note that

$$f_+(x,y) = (y, ax + xy + cy + d)$$

gives an interesting family of degree two polynomial maps that are *not* minimally separating.

Assume f_+ is given by (7–1). If $b = 0$, then f_+ is simply a polynomial automorphism of $\mathbb{C}^2$. If $b \neq 0$, then one can compute that

$$\begin{aligned} f_-(x,y) &= \left(\frac{y - x^2 - cx - d}{a + bx}, x\right) \\ \mathcal{C}^+ &= \ell_\infty \cup \{(x,y) \in \mathbb{C}^2 : y = -a/b\} \\ \mathcal{C}^- &= \ell_\infty \cup \{(x,y) \in \mathbb{C}^2 : x = -a/b\} \\ I^+ &= I_\infty^+ = \mathcal{I}^+ = \{[1:0:0], p^+\} \\ I^- &= \{[0:1:0], p^-\} \end{aligned}$$

where $p^+ \in \ell_\infty$ is given by $[1:-b:0]$ and $p^- \in \mathbb{C}^2$ has coordinates $(-\frac{a}{b}, \frac{a^2}{b^2} - \frac{ca}{b} + d)$. Since $[0:1:0]$ is a superattracting critical point for f_+, it follows that $\mathcal{I}^-$ consists of $[0:1:0]$ and the closure of the forward orbit of p^-.

Example 7.7. *For fixed a, c, d in (7–1) and small enough $|b|$, both f_+ and f_- are completely separating.*

Proof. By direct computation one can show that p^+ is a fixed critical point for f_-. p^+ is attracting for f_- if $|b| < 1$. Since $f_-[1:0:0] = p^+$, we see that iterates of f_- form a normal family in a neighborhood of I^+ when $|b| < 1$. So f_+ is completely separating.

Now suppose that $|b| < \epsilon$ for $\epsilon > 0$ small enough. If we write $p^+ = (x,y)$, then $|x| > C\epsilon^{-1}$ and $|y| > C\epsilon^{-1}|x|$. Furthermore, if (x_n, y_n) satisfies these conditions and $(x_{n+1}, y_{n+1}) = f_+(x,y)$, one can estimate that $|x_{n+1}| = |y_n| = C\epsilon^{-1}|x_n|$ and

$$\frac{|y_{n+1}|}{|x_{n+1}|} \geq C\epsilon^{-1}|y_n|.$$

Consequently, if ϵ is small enough, and we induct on n, we see that

$$\lim_{n\to\infty} f_+^n(p^-) = [0:1:0].$$

Since $[0:1:0]$ is superattracting, we conclude that iterates of f_+ form a normal family on a neighborhood of I^-; hence, f_- is completely separating. $\square$

Example 7.8. *Let $f_+(x,y) = (y, 2xy + y^2)$. Then f_+ is separating. f_- is completely separating, but f_+ is not.*

Proof. For this map one can compute that p^- is a superattracting fixed point. Thus $I^- = \mathcal{I}^-$. We conclude that $\mathcal{I}^- \cap \mathcal{I}^+$ is empty—i.e. that f_+ is separating. It is also immediate that f_- is completely separating. However, since $|b| = 2 > 1$, one can show that p^+ is a semi-repelling fixed point for f_-. Therefore, $p^+ \notin \mathcal{N}^-$—i.e. f_+ is not completely separating. $\square$

It is even possible for a degree two polynomial map to fail to be separating. A particularly interesting example is provided by the next result.

Example 7.9. *Let $f_+(x,y) = (y, x + 2xy + y^2 - 1.5y)$. Then f_+ is not separating.*

Proof. The line $\ell = \{y = -2x\}$ is invariant under f_+ and f_-. In fact, $(0,0)$ is a fixed saddle point, and ℓ is the unstable manifold of $(0,0)$. Furthermore, $p^- \in \ell$. Since $\ell \cap \ell_\infty = p^+$, we see that $\lim_{n\to\infty} f_+^n(p^-) = p^+$. Thus $\mathcal{I}^+ \cap \mathcal{I}^- = \{p^+\}$, and f_+ is not separating. $\square$

Elaborating a little further on the last example, we point out that the invariance of ℓ implies that $\ell \in K^+$. Thus $\tilde{G}^+(x,y,1) = 0$ for points $(x,y) \in \ell$. In contrast to Lemma 6.14, this implies that $\tilde{G}^+$ restricted to $\pi^{-1}(\ell) = \pi^{-1}(W^u(0,0))$ is pluriharmonic.

References

[Be] A. Beardon, *Iteration of Rational Functions*, Springer-Verlag, 1991.

[BS1] E. Bedford and J. Smillie, *Polynomial Diffeomorphisms of $\mathbb{C}^2$: currents, equilibrium measure and hyperbolicity*, Inventiones Math. **87** (1990), 69–99.

[BS2] E. Bedford and J. Smillie, *Polynomial diffeomorphisms of $\mathbb{C}^2$ II: stable manifolds and recurrence*, J. Amer. Math. Soc. **4** (1991), 657–679.

[BS3] E. Bedford and J. Smillie, *Polynomial diffeomorphisms of $\mathbb{C}^2$ III: ergodicity exponents and entropy of the equilibrium measure*, Math. Annalen **294** (1992), 395–420.

[BT] E. Bedford and B. Taylor, *The Dirichlet problem for a complex Monge-Ampère equation*, Inventiones Math. **37** (1976), 1–44.

[DE] P. Dixon and J. Esterle, *Michael's problem and the Poincaré-Fatou-Bieberbach phenomenon*, Bull. Amer. Math. Soc. **15** (1986), 127-187.

[Fa] P. Fatou, *Sur certaines fonctions uniformes de deux variables*, C. R. Acad. Sci. Paris **175** (1922), 1030–1033.

[FS1] J. Fornaess and N. Sibony, *Complex dynamics in higher dimension II*, Modern Methods in Complex Analysis, Princeton Univ. Press, 1995, pp. 135–182.

[FS2] J. Fornaess and N. Sibony, *Oka's inequality for currents and applications*, Math. Annalen **301** (1995), 399–419.

J. Fornaess and N. Sibony, Complex Potential Theory, Kluwer Acad. Publ., 1994, pp. 131–186.

[GH] P. Griffiths and J. Harris, *Principles of Algebraic Geometry*, Wiley-Interscience, 1978.

[HK] B. Hasselblatt and A. Katok, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press, 1995.

[Ho] L. Hörmander, *The Analysis of Linear Partial Differential Operators: Volume 1*, Springer–Verlag, 1989.

[HP] J. Hubbard and P. Papadopol, *Superattractive fixed poins in $\mathbb{C}^n$*, Indiana Math. J. **43** (1994), 321–366.

[Kl] M. Klimek, *Pluripotential Theory*, Oxford University Press, 1991.
[Ni] Y. Nishimura, *Iteration of some birational polynomial quadratic maps of* $\mathbb{P}^2$, preprint.
[Ra] T. Ransford, *Potential Theory in the Complex Plane*, Cambridge University Press, 1995.
[Sk] H. Skoda, *A survey of the theory of closed, positive currents*, Proceedings of Symposia in Pure Mathematics, vol. 41, AMS, 1984, pp. 181–190.
[Ue] T. Ueda, *Fatou sets in complex dynamics in projective spaces*, preprint.

Ithaca, NY 14853
E-mail address: `diller@math.cornell.edu`