

# Dynamical properties of plane polynomial automorphisms

SHMUEL FRIEDLAND†

*University of Illinois, Chicago IL 60680, USA*

JOHN MILNOR

*Institute for Advanced Study, Princeton NJ 08540, USA*

(Received 22 April 1987 and revised 21 April 1988)

*Abstract.* This note studies the dynamical behavior of polynomial mappings with polynomial inverse from the real or complex plane to itself.

## 1. Introduction

Hénon showed in 1969 and 1976 that quadratic automorphisms of the plane provide examples of dynamical systems with very simple definitions but very complicated dynamics. This note will extend his construction by discussing polynomial automorphisms of the real or complex plane of arbitrary degree. The structure of the group  $G$  consisting of all such automorphisms has been studied by a number of authors, starting with Jung in 1942, and is well understood. § 2 describes the structure of  $G$  as an amalgamated free product, and as a union of smooth algebraic varieties. It shows that every conjugacy class in  $G$  either contains an affine or an 'elementary' transformation, or contains a composition of *generalized Hénon transformations*

$$(x, y) \mapsto (y, p(y) - \delta x),$$

where  $p(y)$  is a polynomial of degree at least two and  $\delta \neq 0$  is the constant Jacobian determinant of  $g$ . (Here  $x$  and  $y$  are to be real variables in the real case, or complex variables in the complex case.) It describes a normal form for the conjugacy class of such a composition depending on  $d_1 + \dots + d_m$  parameters, where  $d_1, \dots, d_m$  are the degrees of the factors. § 3 studies the periodic points and the non-wandering set for an arbitrary automorphism  $g \in G$ , and also studies the larger set  $K(g)$  consisting of all points whose (forward and backward) orbit under iteration of  $g$  is bounded. § 4 describes the basic estimate

$$0 \leq h(g) \leq H(g) \leq \log(d)$$

for the topological entropy  $h(g)$  of a polynomial automorphism  $g$  of degree  $d$ . Here

$$H(g) = \limsup_{n \rightarrow \infty} n^{-1} \log^+ |\text{Per}_n(g)|$$

is defined to be the rate of exponential growth for the number of periodic points

† Partially supported by NSF grant DMS-8520434.

of period  $n$ . In the real case,  $h(g)$  can take all values between zero and  $\log(d)$ , but in the complex cyclically reduced case it is conjectured that  $h(g)$  must always be equal to  $\log(d)$ . (Added November 1987. A forthcoming note by John Smillie will prove the authors' conjecture that  $h(g) = H(g) = \log(d)$  for a cyclically reduced complex polynomial automorphism of degree  $d$ . Smillie's proof is based on the volume growth inequality of [Yomdin, § 1.4].) § 5 describes examples, modeled on Smale's horseshoe, for which the entropy and the numbers of periodic points take on their maximal value. § 6 discusses the classification of elementary transformations up to conjugacy within  $G$  or within the larger group of analytic diffeomorphisms of the (real or complex) plane, and § 7 discusses the corresponding problem for generalized Hénon transformations. The final section gives estimates for the algebraic multiplicity of an isolated periodic point.

## 2. The group $G$ of polynomial automorphisms

We will work over either the real numbers  $\mathbf{R}$  or the complex numbers  $\mathbf{C}$ . Let  $G$  be the group of all *polynomial automorphisms* of the coordinate plane  $\mathbf{R}^2$  or  $\mathbf{C}^2$ . By definition, an element  $g$  of  $G$  is a polynomial mapping

$$(x, y) \mapsto g(x, y) = (X(x, y), Y(x, y))$$

from the (real or complex) coordinate plane to itself which is bijective and has polynomial inverse. The more explicit notation  $G_{\mathbf{R}}$  (or  $G_{\mathbf{C}}$ ) may be used if we wish to specify that we are working over the real (or the complex) numbers.

Let  $A \subset G$  be the six-dimensional group consisting of all affine automorphisms of the coordinate plane. Let  $E \subset G$  be the group consisting of all polynomial automorphisms which are *elementary* in the sense that they carry each line of the form  $y = \text{constant}$  to a line of the form  $y = \text{constant}'$ . It is not difficult to check that a transformation  $e$  is elementary if and only if it can be written as

$$e(x, y) = (\alpha x + p(y), \beta y + \gamma)$$

for some constants  $\alpha, \beta$  and  $\gamma$  with  $\alpha\beta \neq 0$ , and for some polynomial function  $p(y)$ . Note that  $E$  is a solvable group. In fact its second commutator subgroup is the commutative subgroup of codimension three consisting of all transformations of the form  $e(x, y) = (x + p(y), y)$ . Our starting point is the following statement.

**THEOREM OF JUNG.** *The group  $G$  of polynomial automorphisms of the (real or complex) coordinate plane is generated by these two subgroups  $A$  and  $E$ .*

Proofs of this Theorem may be found in [J, G] or [McK2]. Note that the Jacobian determinant of any transformation in  $G$  must be a non-zero constant. The still unproved 'Jacobian conjecture' is the converse claim that any polynomial mapping with non-zero constant Jacobian determinant must necessarily be a polynomial automorphism. (Cf [BCW].)

We will next sketch a proof of the statement that  $G$  is actually equal to the amalgamated free product of the two subgroups  $A$  and  $E$ . (Cf [S, K, Wr, Theorem 4].)

*Definition.* A sequence  $(g_1, \dots, g_n)$  of length  $n \geq 1$  will be called a *reduced word*, representing the group element  $g = g_n \circ g_{n-1} \circ \dots \circ g_1$ , if each factor  $g_i$  belongs to either  $A$  or  $E$  but not to the intersection  $S = A \cap E$ , and if no two consecutive factors belong to the same subgroup  $A$  or  $E$ . By abuse of notation, we will simply refer to the product  $g_n \circ \dots \circ g_2 \circ g_1$  as a reduced word of length  $n$ . It follows immediately from Jung's Theorem that every element of  $G$  can be expressed as such a reduced word, unless it belongs to the intersection  $A \cap E$ .

By the *degree* of a polynomial automorphism  $(x, y) \mapsto g(x, y) = (X(x, y), Y(x, y))$  will be meant the maximum of the degrees of the two component polynomials  $X(x, y)$  and  $Y(x, y)$ . Note that  $g$  has degree one if and only if it belongs to the affine group  $A$ . Our presentation will be based on the following elementary but basic result.

**THEOREM 2.1.** *The degree of any reduced word  $g = g_n \circ g_{n-1} \circ \dots \circ g_1$  is equal to the product of the degrees of the factors  $g_i$ .*

(By way of contrast, the degree of the composition of two elementary transformations is at most equal to the maximum of the two degrees.) Evidently we need consider only those factors  $g_i$  which belong to the group  $E$  in this computation. In fact the factors  $g_i \in E$  all have  $\text{degree}(g_i) \geq 2$ , but the factors which belong to  $A$  have  $\text{degree}(g_i) = 1$ . The proof of Theorem 2.1 will be given at the end of this section.

As one immediate corollary, we see that  $\text{degree}(g^{-1}) = \text{degree}(g)$ . (Cf [McK1].) Here is another important consequence.

**COROLLARY 2.2.** *No reduced word is equal to the identity element of  $G$ .*

For a reduced word, unless it consists of a single affine factor, must have degree two or more.

The statement that  $G$  is the free product of  $A$  and  $E$ , amalgamated along the five dimensional solvable subgroup  $S = A \cap E$ , now follows as a purely formal consequence of 2.2. More explicitly, we have the following.

**COROLLARY 2.3.** *The expression for a group element  $g \notin S$  as a reduced word  $g = g_n \circ g_{n-1} \circ \dots \circ g_1$  is unique up to the following modifications: For any  $s \in S$  and any  $i > 1$  we can replace  $g_i$  by the product  $(g_i \circ s)$ , and simultaneously replace  $g_{i-1}$  by  $(s^{-1} \circ g_{i-1})$ .*

*Proof.* This follows by a straightforward induction on the length  $n$ , using 2.2 to start the induction. Details will be omitted.  $\square$

In particular, it follows that the length  $n \geq 1$  of such a reduced word is an invariant of the group element  $g$ . Furthermore, the sequence of degrees of the factors  $g_i$  also forms an invariant. Of course every second degree must be equal to one. It will be convenient to define the *polydegree*  $(d_1, \dots, d_m)$  of a non-affine group element  $g$  to be the sequence of integers  $d_i \geq 2$  which is obtained from the sequence  $(\text{degree}(g_1), \dots, \text{degree}(g_n))$  by crossing out all of the ones. More explicitly:  $g$  has *polydegree*  $(d_1, \dots, d_m)$  if and only if some representative of the double coset  $A \circ g \circ A$

can be written as a reduced word of the form

$$e_m \circ a_{m-1} \circ e_{m-1} \circ \cdots \circ a_2 \circ e_2 \circ a_1 \circ e_1,$$

with degree  $(e_i) = d_i$ .

Evidently the subset of  $G$  consisting of all automorphisms of degree  $d$  or less can be given the structure of an affine algebraic variety, presumably with singularities. Here is a sharper statement.

LEMMA 2.4. *The set  $G[d_1, \dots, d_m] \subset G$  consisting of all group elements with polydegree  $(d_1, \dots, d_m)$  forms a smooth analytic manifold of dimension  $d_1 + \dots + d_m + 6$ .*

Thus we obtain a partition of  $G$  as a disjoint union

$$G = A \cup G[2] \cup G[3] \cup G[2, 2] \cup G[4] \cup G[5] \cup \cdots$$

of smooth Zariski open subsets of algebraic varieties. The proof of 2.4, based on an explicit normal form, will be given at the end of this section.

Next let us study the problem of finding a normal form for conjugacy classes in the group  $G$ . Note that we can cyclically permute the factors of any reduced word, without changing its conjugacy class. If a reduced word  $g_n \circ \cdots \circ g_1$  has minimal length within its conjugacy class, then clearly either:

- (1) it has length  $n = 1$ , so that the group element  $g$  belongs to  $A$  or  $E$ , or
- (2) its length is even,  $n = 2m$ , so that the two extreme factors  $g_1$  and  $g_n$  belong to different subgroups  $A$  and  $E$ .

In the second case, which is the dynamically interesting one, we will say that the group element

$$g = g_{2m} \circ \cdots \circ g_1$$

is *cyclically reduced*. Thus every cyclically reduced group element of polydegree  $(d_1, \dots, d_m)$  is conjugate to a composition of  $m$  group elements of the form  $(a_i \circ e_i)$ , with degree  $(e_i) = d_i$ . If two cyclically reduced words are conjugate to each other, note that they must have the same polydegree, up to a cyclic permutation. Let us single out one particularly convenient class of compositions of this form  $(a \circ e)$ .

Definition 2.5. A group element  $g$  with Jacobian determinant  $\delta$ , will be called a *generalized Hénon transformation* if it has the form

$$g(x, y) = (y, p(y) - \delta x)$$

for some polynomial function  $p(y)$  of degree  $d \geq 2$ .

Note that such a generalized Hénon transformation has determinant  $\delta = 1$  if and only if it can be expressed as the composition of two involutions

$$(x, y) \mapsto (p(y) - x, y) \mapsto (y, p(y) - x).$$

(cf [D p. 252]).

THEOREM 2.6. *Every cyclically reduced element of  $G$  is conjugate to a composition*

$$(x_0, x_1) \xrightarrow{g_1} (x_1, x_2) \xrightarrow{g_2} \cdots \xrightarrow{g_m} (x_m, x_{m+1})$$

of generalized Hénon transformations, where  $g_i(x, y) = (y, p_i(y) - \delta_i x)$ . Furthermore, this composition can be chosen so that the coefficient of highest degree in each polynomial  $p_i$  is  $\pm 1$ , and so that the next highest coefficient is zero. The resulting normal form is unique up to a finite number of choices.

In other words, we can represent any cyclically reduced conjugacy class of polydegree  $(d_1, \dots, d_m)$  by a transformation of the form  $g(x_0, x_1) = (x_m, x_{m+1})$  where the sequence  $x_2, \dots, x_{m+1}$  is defined inductively by the formula  $x_{i+1} = p_i(x_i) - \delta_i x_{i-1}$ , with

$$p_i(x_i) = \pm x_i^{d_i} + (\text{terms of degree } \leq d_i - 2).$$

This normal form depends on just  $d_1 + \dots + d_m$  parameters.

In most cases, the proof will show that we need only use polynomials  $p_i$  with leading coefficient  $+1$ . The only exceptional case, where we must allow a leading coefficient of  $-1$ , occurs when we are working over the real numbers, when the degree  $d = d_1 \cdot \dots \cdot d_m$  is odd, and when the product of the leading coefficients of the polynomials  $p_i$  is negative.

In the real even degree case, the normal form with leading coefficients  $+1$  is unique up to cyclic permutation of the  $m$  factors. However, in the complex case there is an additional ambiguity since the proof involves extracting a  $(d - 1)$ st root, which is well defined only up to multiplication by  $(d - 1)$ st roots of unity. As an example, the normal form  $(x, y) \mapsto (y, p(y) - \delta x)$  for the conjugacy class of a generalized Hénon transformation, with  $p(y) = y^d + (\text{terms of degree } \leq d - 2)$ , is unique except that we can replace the polynomial  $p(y)$  by  $p(\omega y)/\omega$  where  $\omega$  can be any  $(d - 1)$ st root of unity.

One special case of 2.6 is particularly noteworthy:

**COROLLARY 2.7.** *Every automorphism  $g \in G$  of prime degree is affinely conjugate either to a generalized Hénon transformation or to an elementary transformation.*

(Cf [E].) The proof is straightforward.

In the simplest case, if  $g$  has degree  $d = 2$  and is not affinely conjugate to an elementary transformation, we obtain a unique normal form

$$(x, y) \mapsto (y, y^2 + c - \delta x)$$

depending on two parameters  $c$  and  $\delta$ . This is completely equivalent to the normal form studied by [H2].

It is now time to begin some proofs.

*Proof of Theorem 2.1.* We must compute the degree of a composition where the factors are alternately elementary and affine. Consider first a polynomial automorphism  $(x, y) \mapsto (X(x, y), Y(x, y))$  where  $X$  and  $Y$  are polynomials in  $x, y$  which satisfy the condition  $\text{degree}(X) \leq \text{degree}(Y) = d_0$ . For example we could start with the identity transformation, or with any affine transformation. Let us compose with an elementary transformation to obtain the polynomials

$$X' = \alpha X + p(Y), \quad Y' = \beta Y + \gamma$$

in  $x$  and  $y$ . If this elementary transformation does not belong to the intersection

$S = A \cap E$ , then the degree  $d$  of the polynomial  $p$  must satisfy  $d \geq 2$ , and

$$\text{degree}(X') = dd_0 > \text{degree}(Y') = d_0.$$

Next let us compose with an affine transformation

$$X'' = \lambda_{11}X' + \lambda_{12}Y' + \kappa_1, \quad Y'' = \lambda_{21}X' + \lambda_{22}Y' + \kappa_2,$$

which does not belong to  $S$ . Then the coefficient  $\lambda_{21}$  must be non-zero, and it follows that

$$\text{degree}(X'') \leq \text{degree}(Y'') = dd_0.$$

Thus, in this situation, *whenever we compose with an elementary transformation of degree  $d \geq 2$  followed by a non-elementary affine transformation, the degree will be multiplied by  $d$* . Theorem 2.1 follows by induction. □

*Proof of Lemma 2.4.* We will first construct a fibration

$$\pi : G[d_1, \dots, d_m] \rightarrow P^1 \times P^1,$$

over the product  $P^1 \times P^1$  of two (real or complex) projective lines. We will identify the projective line  $P^1$  with the coset space  $A/S$ . In fact, each  $a \in A$  maps the family of ‘horizontal’ lines  $y = \text{constant}$  to some family of parallel lines, and this image family is determined precisely by the coset  $a \circ S$ . The fiber  $F = F[d_1, \dots, d_m]$  will consist of all group elements  $f$  which can be written as reduced words of the form

$$f = e_m \circ a_{m-1} \circ e_{m-1} \circ \dots \circ a_1 \circ e_1, \tag{2.8}$$

with elementary transformations at both ends, and with  $\text{degree}(e_i) = d_i$ . It follows by definition that every element  $g \in G[d_1, \dots, d_m]$  can be written as a product  $a' \circ f \circ a$  with  $f \in F$  as above, where  $a'$  and  $a$  can be arbitrary affine transformations. The cosets  $a' \circ S$  and  $S \circ a$  are uniquely defined by Corollary 2.3. Hence we can define the projection  $\pi$  by the formula

$$\pi(a' \circ f \circ a) = (a^{-1} \circ S, a' \circ S).$$

It is not difficult to check that  $\pi$  is a locally trivial fibration.

If  $\pi(g) \in P^1 \times P^1$  corresponds to the pair of lines  $(L, L')$  through the origin, then we can characterize  $L$  as the unique line through the origin such that the degree of the polynomial mapping  $g$  restricted to  $L$  is strictly less than the degree  $d$  of  $g$ . Similarly,  $L'$  is the unique line through the origin such that the degree of  $g^{-1}$  restricted to  $L'$  is strictly less than  $d$ . In the simplest case  $m = 1$ , note that  $g$  maps lines parallel to  $L$  linearly onto lines parallel to  $L'$ . The inverse image under  $\pi$  of the diagonal in  $P^1 \times P^1$  plays a particularly noteworthy role.

**LEMMA 2.9.** *For  $g \in G[d_1, \dots, d_m]$ , the image  $\pi(g)$  lies off of the diagonal in  $P^1 \times P^1$  if and only if  $g$  is affinely conjugate to a cyclically reduced word of polydegree  $(d_1, \dots, d_m)$ .*

For if  $\pi(a' \circ f \circ a)$  lies off the diagonal, then evidently the product  $a \circ a'$  does not lie in the subgroup  $S$ , and hence  $g = a' \circ f \circ a$  is affinely conjugate to a cyclically reduced word

$$(a \circ a') \circ e_m \circ a_{m-1} \circ \dots \circ a_1 \circ e_1$$

of polydegree  $(d_1, \dots, d_m)$ . On the other hand, if  $\pi(g)$  does belong to the diagonal, then  $g$  is conjugate to a group element  $a_{m-1} \circ e_{m-1} \circ \dots \circ a_1 \circ (e_1 \circ a \circ a' \circ e_m)$  which has degree strictly less than  $d_1 \cdots d_m$ .

The proof of 2.4 continues as follows. We will next show that the expression (2.8) can be simplified so that only one affine transformation is needed. In fact let  $t$  be the reflection  $t(x, y) = (y, x)$ . It is not hard to see that the double coset  $S \circ t \circ S$  consists precisely of all affine transformations which do *not* belong to the subgroup  $S$ . Thus every  $a_i$  in the reduced word (2.8) belongs to this double coset. If we absorb the  $S$  factors into the adjacent elementary transformations, then we obtain an expression of the form

$$f = e'_m \circ t \circ e'_{m-1} \circ t \circ \dots \circ t \circ e'_1$$

with  $e'_i \in E$ . However, this expression is not yet unique, since the equation  $s^{-1} \circ t \circ s' = t$  may be satisfied for suitably chosen elements  $s$  and  $s'$  in  $S$ . Evidently any such pair  $s^{-1}$  and  $t \circ s \circ t = s'$  can be absorbed into the adjacent elementary transformations without changing the value of the product  $f$ .

Let  $\hat{S} \subset S$  be the group of all diagonal affine transformations

$$\hat{s}(x, y) = (\alpha x + \alpha', \beta y + \beta').$$

This is precisely the largest subgroup of  $S$  which is normalized by  $t$ . Let  $\hat{E} \subset E$  be the group consisting of all elementary transformations of the form  $\hat{e}(x, y) = (x + p(y), y)$ , where  $p(0) = 0$  so that  $\hat{e}$  fixes the origin.

LEMMA 2.10. *Every element of the fiber  $F = F[d_1, \dots, d_m]$  can be written uniquely as a reduced word of the form*

$$(\hat{s}_m \circ \hat{e}_m) \circ t \circ \hat{e}_{m-1} \circ \dots \circ t \circ \hat{e}_1$$

with  $\hat{s}_m \in \hat{S}$  and with  $\hat{e}_i \in \hat{E}$ . Hence this fiber  $F$  can be parametrized by an open subset of the  $(d_1 + \dots + d_m + 4)$ -dimensional coordinate space.

*Proof.* Uniqueness of this expression follows from 2.3. Existence is proved by induction on  $m$ , as follows. For  $m = 1$  this is just the easily verified statement that every element  $e_1 \in E$  can be written as a product  $\hat{s}_1 \circ \hat{e}_1$  with  $\hat{s}_1 \in \hat{S}$  and  $\hat{e}_1 \in \hat{E}$ . For the inductive step, we need only multiply the given expression on the left by  $e_{m+1} \circ t$ , observing that  $e_{m+1} \circ t \circ \hat{s}_m$  can be written as  $e' \circ t$  for some  $e' \in E$ , which itself can be written as  $\hat{s}_{m+1} \circ \hat{e}_{m+1}$ . This proves 2.10, and completes the proof of 2.4.  $\square$

*Proof of Theorem 2.6.* Proceeding just as in the proof of 2.4, we see that every cyclically reduced word is conjugate to a composition of the form

$$g = (t \circ e_m) \circ \dots \circ (t \circ e_1).$$

Let us set  $t \circ e_i : (x_{i-1}, y_{i-1}) \mapsto (x_i, y_i)$ , where

$$x_i = \beta_i y_{i-1} + \gamma_i, \quad y_i = \alpha_i x_{i-1} + p_i(y_{i-1}).$$

Using 2.3, we see that this expression for  $g$  is unique, except that we can permute the factors cyclically without changing the conjugacy class, and we can modify by elements of the diagonal affine group  $\hat{S}$ . The effect of such a modification will be to transform each of the variables  $x_i$  and  $y_i$  by a one-dimensional affine transformation, where we must apply exactly the same transformation to  $x_m$  or to  $y_m$  as we

do to  $x_0$  or  $y_0$ . To begin, let us simply replace each variable  $y_{i-1}$  by the expression  $x_i = \beta_i y_{i-1} + \gamma_i$ . Then  $t \circ e_i$  will be replaced by an expression of the form

$$(x_{i-1}, x_i) \mapsto (x_i, q_i(x_i) - \delta_i x_{i-1}),$$

or in other words by a ‘generalized Hénon transformation’.

We can still simplify further by applying an affine transformation to each variable  $x_i$ . Define the ‘center of gravity’ of a polynomial  $p(y)$  of degree  $d \geq 2$  to be the unique point at which its  $(d - 1)$ st derivative vanishes. In the complex case, this can be identified with the average of the roots, or the average of the critical points or the fixed points. If  $p(y)$  has center of gravity  $c$ , note that any polynomial of the form  $p(y + c) - \text{constant}$  has center of gravity at  $y = 0$ . Thus an appropriate translation of each of the coordinates  $x_i$  will replace the polynomials  $q_i$  by polynomials  $q'_i$  with next to highest coefficient equal to zero.

We are still free to make a scale change, replacing each coordinate  $x_i$  by say  $X_i = x_i / \eta_i$ , where the  $\eta_i$  are non-zero constants. This has the effect of replacing each  $q'_i(x_i)$  by the polynomial  $q''_i(X_i) = q'_i(\eta_i X_i) / \eta_{i+1}$ , with corresponding changes in the constants  $\delta_i$ . If the polynomial  $q'_i(x)$  has degree  $d_i$  and leading coefficient  $\kappa_i$ , then  $q''_i$  will have leading coefficient  $\eta_i^{d_i} \kappa_i / \eta_{i+1}$ . In order to set all of the leading coefficients equal to  $+1$ , we must solve the system of  $n$  equations

$$\eta_{i+1} = \eta_i^{d_i} \kappa_i,$$

in  $n$  variables  $\eta_i \neq 0$ , with indices  $i$  taken modulo  $n$ . In the complex case, it is not difficult to check that there are exactly  $d - 1$  distinct solutions, where  $d = d_1 \cdots d_n$  is the total degree. In fact  $\eta_1$  can be any  $(d - 1)$ st root of an appropriate monomial in the  $\kappa_i$ , and the remaining  $\eta_i$  are then uniquely determined. In the real case, it follows that there is a unique solution when  $d$  is even, and either two solutions or no solutions when  $d$  is odd, according as the product of the leading coefficients is positive or negative. Thus, in the real odd degree case, we must allow one of the leading coefficients to be  $\pm 1$  in order to guarantee a solution. Further details will be left to the reader. □

### 3. Periodic points and the non-wandering set

Let  $g$  be a polynomial automorphism of  $\mathbf{R}^2$  or of  $\mathbf{C}^2$ . We would like to count the number of fixed points of  $g$ , and more generally the numbers of fixed points of the various iterates  $g \circ g, g \circ g \circ g, \dots$ . Evidently these numbers will not change if we replace  $g$  by a conjugate  $h \circ g \circ h^{-1}$ . Hence, according to the results of § 2, we need only consider the case where  $g$  is elementary, or affine, or the case where  $g$  can be expressed as a cyclically reduced word.

**THEOREM 3.1.** *If  $g \in G$  is cyclically reduced of degree  $d$ , then the number of fixed points of  $g$  is less than or equal to  $d$ . In fact, in the complex case, each fixed point can be assigned a multiplicity  $\mu \geq 1$  so that the sum of the multiplicities for all of the fixed points of  $g$  is precisely equal to the degree  $d$ .*

As an immediate corollary, it follows that the number of fixed points of the  $n$ -fold iterate  $g^{\circ n} = g \circ \dots \circ g$  is less than or equal to  $d^n$ . For clearly this iterate is cyclically



reduced of degree  $d^n$ . (In the generic case, we will see in 3.2 that the number of complex fixed points of  $g^{\circ n}$  is precisely equal to  $d^n$ .)

*Proof of 3.1.* We need only consider the complex case, since every real polynomial automorphism extends to a complex one. According to § 2.6, the map  $g$  is conjugate to a composition of generalized Hénon maps. Suppose for example that  $g = g_m \circ \dots \circ g_1$  is the composition of the generalized Hénon maps

$$(x_0, x_1) \xrightarrow{g_1} (x_1, x_2) \xrightarrow{g_2} \dots \xrightarrow{g_m} (x_m, x_{m+1})$$

where

$$x_2 = p_1(x_1) - \delta_1 x_0, \dots, \quad x_{m+1} = p_m(x_m) - \delta_m x_{m-1}. \tag{*}$$

For a fixed point of  $g$ , we must have  $x_0 = x_m, x_1 = x_{m+1}$ ; and these equations reduce to a system of  $m$  polynomial equations in  $m$  complex variables. Let  $d_i \geq 2$  be the degree of the polynomial  $p_i(x)$ , so that  $d = d_1 \cdot \dots \cdot d_m$  is the total degree of  $g$ . We must prove that this system of equations (\*) has exactly  $d_1 \cdot \dots \cdot d_m$  solutions, counted with multiplicity.

The ‘Theorem of the Nonlinear Alternative’, which is a special case of Bezout’s Theorem, asserts that the algebraic number of solutions of such a system of  $m$  polynomial equations in  $m$  complex variables is precisely equal to the product of the degrees, providing that the following condition is satisfied: *If we throw away all but the terms of highest degree in each of these polynomial equations, then the resulting system must have only the trivial solution.* (See for example [W] or [F].) In our case, the equations (\*) reduce to the equations

$$x_1^{d_1} = 0, \dots, \quad x_m^{d_m} = 0,$$

which manifestly have only the trivial solution. □

Recall from § 2.4 that the group  $G$  splits as a disjoint union

$$G = A \cup G[2] \cup G[3] \cup G[2, 2] \cup G[4] \cup \dots$$

of smooth algebraic varieties, where each  $g \in G[d_1, \dots, d_m]$  has degree  $d = d_1 \cdot \dots \cdot d_m$ . Let us say that a property of  $g$  is true for *generic*  $g \in G[d_1, \dots, d_m]$  if it is true for all  $g$  outside of a countable union of lower dimensional subvarieties.

**LEMMA 3.2.** *If  $g$  is a generic element of  $G[d_1, \dots, d_m]$ , then every iterate  $g^{\circ n}$  has exactly  $d^n$  distinct complex fixed points, each with multiplicity  $\mu = 1$ .*

*Proof.* According to § 2.8, every  $g \in G[d_1, \dots, d_m]$  outside of a codimension 1 sub-variety is cyclically reduced, up to affine conjugacy. Furthermore, for each  $n \geq 1$ , it is not difficult to check that those  $g$  for which  $g^{\circ n}$  has fewer than  $d^n$  fixed points form an algebraic subvariety  $V_n \subset G[d_1, \dots, d_m]$ . To complete the proof, we need only show that  $V_n$  is a proper subvariety. That is, for each  $n$  we must exhibit at least one example of an element  $g \in G[d_1, \dots, d_m]$  for which  $g^{\circ n}$  has  $d^n$  distinct fixed points. To do this, consider the family of maps  $g$ , depending on a single parameter  $\delta$ , given by  $g = g_m \circ \dots \circ g_1$  where

$$g_i(x, y) = (y, y^{d_i} - \delta x).$$

In the degenerate case  $\delta = 0$ , a brief computation shows that the iterate  $g^{\circ n}$  has

exactly  $d^n$  distinct fixed points, consisting of all pairs  $(x, y)$  with  $x = x^{d^n}$  and  $y = x^{d^n}$ . Since the fixed point equations have  $d^n$  distinct solutions when  $\delta = 0$ , it follows that they must have at least  $d^n$  distinct solutions whenever  $|\delta|$  is sufficiently close to zero.  $\square$

*Remarks.* It is certainly not true that the multiplicity  $\mu$  is always equal to one. As an example, consider the classical Hénon map

$$g : (x, y) \mapsto (y, y^2 - \delta x + c)$$

of degree  $d = 2$ . In the generic case, the  $d^2 = 4$  fixed points of  $g \circ g$  consist of two fixed points of  $g$ , together with one periodic orbit of period two. However, for particular values of the two parameters these orbits may degenerate. In fact, whenever  $c = (1 + \delta)^2/4$  the two fixed points coincide, and whenever  $c = -3(1 + \delta)^2/4$  the orbit of period 2 collapses to one of the two fixed points.

*It seems natural to conjecture that every cyclically reduced  $g \in G$  must have complex periodic points of all sufficiently large periods, and in fact that the number of such points grows exponentially with the period.* However, to prove this we would need some good upper bound for the multiplicity of a periodic point. To illustrate the problem, any fixed point of  $g$  is also a fixed point of  $g^{\circ n}$ , and hence is assigned a multiplicity  $\mu_n \geq 1$ , for every  $n \geq 1$ . According to Shub and Sullivan, this sequence of multiplicities is bounded. *Can the maximum multiplicity of a point of period  $n$  grow exponentially with  $n$ ?* Some partial results about multiplicities are described in § 8. These imply, for example, that a cyclically reduced complex polynomial automorphism with determinant  $\delta \neq 1$  must have periodic points of all large prime periods (§ 8.5). However, we have not been able to obtain more precise results.

*Remark 3.3.* Thus far, we have discussed the multiplicity of a fixed point only in the complex case. The multiplicity of an isolated fixed point of a homeomorphism  $g$  of the real plane  $\mathbf{R}^2$  can be defined as the local degree of the map  $(x, y) \mapsto g(x, y) - (x, y)$  at the fixed point. If the eigenvalues  $\alpha, \beta$  of the first derivative at the fixed point are both different from  $+1$ , then this multiplicity is equal to the sign of  $(\alpha - 1)(\beta - 1)$ . With this definition, one can show that the sum of the multiplicities of the fixed points for a composition of real generalized Hénon maps is equal to zero when the total degree  $d$  is even, and is equal to either  $+1$  or  $-1$  when  $d$  is odd according as the coefficient of the highest degree term is negative or positive.

Next let us study the *non-wandering set*  $\Omega(g)$ . By definition, a point in the coordinate plane belongs to  $\Omega(g)$  if and only if, for every neighborhood  $N$  of the point, there exists an  $n > 0$  so that the  $n$ -fold forward image  $g^{\circ n}(N)$  intersects  $N$ . Evidently  $\Omega(g) = \Omega(g^{-1})$  is a closed subset of the plane which contains all of the periodic points of  $g$ . If  $g$  is cyclically reduced, then we will show that this non-wandering set  $\Omega(g)$  is compact.

**LEMMA 3.4.** *For every generalized Hénon map*

$$g : (x, y) \mapsto (y, z) = (y, p(y) - \delta x)$$

*there exists a constant  $\kappa$  so that  $|y| > \kappa$  implies that either  $|z| > |y|$  or  $|x| > |y|$  or both.*

*Proof.* Since  $p(y)$  has degree two or more, we can choose  $\kappa$  so that  $|p(y)| > |y| + |\delta y|$

whenever  $|y| > \kappa$ . It then follows that either  $|x| > |y|$  or

$$|z| = |p(y) - \delta x| \geq |p(y)| - |\delta y| > |y|,$$

whenever  $|y| > \kappa$ . □

Now consider any composition  $g = g_m \circ \dots \circ g_1$  of generalized Hénon maps. Let  $D$  be the closed disk of radius  $\kappa$  centered at the origin, where the constant  $\kappa$  must be large enough so as to satisfy the conditions of 3.4 for every one of the maps  $g_1, \dots, g_m$ .

**LEMMA 3.5.** *With these hypotheses, it follows that the non-wandering set  $\Omega(g)$  is contained in the bidisk  $D \times D$ , and hence is compact.*

*Proof.* We will use the notation

$$g_i(x_{i-1}, x_i) = (x_i, x_{i+1}),$$

where it is convenient to extend periodically by setting  $g_{i+m} = g_i$  so that this formula makes sense for all integer values of  $i$ . Then the full orbit of  $(x_0, x_1)$  under the composition  $g$  is given by

$$\dots \mapsto (x_{-m}, x_{-m+1}) \mapsto (x_0, x_1) \mapsto (x_m, x_{m+1}) \mapsto (x_{2m}, x_{2m+1}) \mapsto \dots$$

We will show that this full orbit is bounded if and only if  $|x_n| \leq \kappa$  for every integer  $n$ . Suppose, to the contrary, that  $|x_n| > \kappa$  for some  $n$ . Then according to 3.4 we must have either  $|x_n| < |x_{n+1}|$ , which implies inductively that

$$|x_n| < |x_{n+1}| < |x_{n+2}| < \dots,$$

or  $|x_n| < |x_{n-1}|$  which implies that

$$|x_n| < |x_{n-1}| < |x_{n-2}| < \dots$$

In either case the resulting monotonic sequence of numbers must diverge to infinity. For if it remained bounded, then the  $x_i$  would have a limit point whose orbit would contradict the inequality 3.4.

It follows immediately that the set  $\Omega(g) = \Omega(g^{-1})$  is contained in the product  $D \times D$ . In fact, every point which does not belong to this bidisk has either an open neighborhood whose forward images diverge monotonically to infinity or an open neighborhood whose backward images diverge monotonically to infinity. □

**Definition 3.6.** Following Hubbard and Oberste-Vorth, let  $K_+ = K_+(g)$  be the closed set consisting of all points whose forward orbit under  $g$  remains bounded, and let  $K_- = K_-(g)$  be the closed set consisting of points whose backward orbit is bounded. The intersection will be denoted by  $K = K_+ \cap K_-$ .

If  $g$  is a composition of generalized Hénon maps, then the discussion above shows that

$$\Omega(g) \subset K(g) \subset D \times D.$$

In particular, the set  $K$  is compact. Note that the complements of  $K_+$  and  $K_-$  are unbounded open sets. In the real case, it is not difficult to check that these complementary sets are contractible when the degree is even, and that each one consists of two contractible components when the degree is odd. However, in the complex case they have a rather complicated topology. (See [H].) It is conjectured that both  $K_+$  and  $K_-$  are connected, although the intersection  $K$  may well be disconnected.

Since the set  $K$  is bounded, its Lebesgue measure  $\lambda(K)$  (2-dimensional in the real case or 4-dimensional in the complex case) is certainly finite. To say more about the measure theoretic behavior of these sets  $K$ ,  $K_+$  and  $K_-$ , we must consider three different cases, according as the transformation  $g$  increases, decreases, or preserves volumes. Let  $\delta = \delta_1 \cdots \delta_m$  be the constant determinant of the first derivative of  $g$ ; where we continue to assume that  $g$  is a composition of generalized Hénon maps.

**LEMMA 3.7.** *If  $|\delta| = 1$ , then the Lebesgue measures  $\lambda(K) = \lambda(K_-) = \lambda(K_+)$  are finite and equal, so that almost every point in  $K_\pm$  actually belongs to  $K$ . On the other hand, if  $|\delta| < 1$ , then  $\lambda(K) = \lambda(K_-) = 0$ , while  $\lambda(K_+)$  must be either zero or infinite. The case  $|\delta| > 1$  is similar, except that the roles of  $K_+$  and  $K_-$  are interchanged.*

The proof is straightforward, and will be left to the reader.

Thus, when  $|\delta| > 1$ , almost every forward orbit escapes to infinity, and when  $|\delta| < 1$  almost every backward orbit escapes to infinity. In the exceptional case  $|\delta| = 1$ , for almost every point, the forward orbit remains bounded if and only if the backward orbit remains bounded. The problem of deciding just which automorphisms  $g$  have the property that  $\lambda(K_+) = \lambda(K_-) = 0$ , so that almost every orbit escapes to infinity in both directions, seems quite difficult. In the complex case, it seems possible that  $\lambda(K_+) > 0$  if and only if there exists a stable periodic orbit. However, in the real case, the set  $K_+$  has positive measure also whenever there is a 'strange attractor' or, in the area preserving case, whenever there is an invariant circle.

The behavior of the elementary and affine transformations in  $G$  is quite different. Such transformations may have infinitely many fixed points, and the sets  $\Omega$  and  $K$  may well be unbounded.

**LEMMA 3.8.** *An elementary or affine transformation can have at most one isolated fixed point, and the multiplicity of such an isolated fixed point is necessarily equal to +1 in the complex case or to  $\pm 1$  in the real case.*

(Cf 3.3.) *Proof.* Let  $e(x, y) = (\alpha x + p(y), \beta y)$  have a fixed point at the origin. If  $\alpha = 1$  or  $\beta = 1$ , then there is an entire curve of fixed points through the origin; while otherwise it is easy to check that there is only one fixed point with multiplicity equal to the sign of  $(\alpha - 1)(\beta - 1)$  in the real case, or to +1 in the complex case. The proof in the affine case is similar.  $\square$

Combining this result with 3.1 we obtain the following.

**COROLLARY 3.9.** *In the complex case, no elementary or affine transformation can be topologically conjugate to a cyclically reduced element of the group  $G$ .*

Here is a brief description of results which will be described in more detail in § 6.

**LEMMA 3.10.** *If  $g \in G$  is conjugate either to an affine transformation or to an elementary transformation, then  $\Omega(g) = K(g)$ . If we choose a suitable representative  $g_0$  within the conjugacy class, then the set  $\Omega(g_0) = K(g_0)$ , if non-vacuous, is either a linear space*

on which  $g_0$  acts linearly, or a union of finitely many 'horizontal' lines  $y = \text{constant}$ , consisting entirely of periodic points, on which  $g_0$  acts by an isometry.

Here the linear space can have dimension 0, 1, or 2, and the linear map must be linearly conjugate either to an orthogonal transformation (in the real case) or to a unitary transformation (in the complex case). For any  $g \in G$  which is conjugate to an affine or elementary transformation, it follows from 3.10 that the set  $\Omega(g) = K(g)$  is a smooth manifold which contains periodic points of at most finitely many distinct periods. For details, the reader is referred to § 6.8.

In the case of a single Hénon map

$$g: (x, y) \mapsto (y, p(y) - \delta x)$$

over the real numbers, it is interesting to note that the sets  $K_+(g)$  and  $K_-(g)$ , if non-vacuous, must always contain at least one fixed point of  $g$ . For if the fixed point equations  $x = y = p(y) - \delta y$  have no solution at all, then the polynomial  $p(y) - (1 + \delta)y$  must be either strictly positive or strictly negative for all  $y$ . Suppose, to fix our ideas, that it is strictly positive. Then it follows that the linear function  $l(x, y) = y - \delta x$  must increase monotonically under  $g$ ; hence every forward orbit must diverge to infinity, and similarly every backward orbit must diverge to infinity. We do not know whether a corresponding statements holds for compositions of Hénon maps. It certainly fails for elementary maps. As an example, the map

$$(x, y) \mapsto (x + y^2 - 1, -y)$$

has points of period 2, but no fixed point.

#### 4. Entropy estimates

The topological entropy  $h(f)$  is usually defined for a continuous map  $f$  from a compact metrizable space into itself. (See for example [W].) It will be convenient to extend this definition to the case of a proper map from a suitable locally compact space  $X$  into itself by defining  $h(f)$  to be the topological entropy of the canonical extension of  $f$  over the one point compactification  $X \cup \infty$ . With this understanding, it follows that every polynomial automorphism  $g$  of the real or complex plane has a well defined topological entropy  $h(g) \geq 0$ . We will also study the auxiliary quantity

$$H(g) = \limsup_{n \rightarrow \infty} n^{-1} \log^+ |\text{Per}_n(g)|,$$

where  $|\text{Per}_n(g)|$  is the number of periodic points with period exactly equal to  $n$ , and where  $\log^+(x)$  is defined to be the maximum of  $\log(x)$  and zero. If  $g$  is an elementary transformation or an affine transformation, then it follows easily from § 3.10 that  $h(g) = H(g) = 0$ . Hence we will concentrate on the case of a cyclically reduced polynomial automorphism.

**THEOREM 4.1.** *If  $g$  is cyclically reduced of degree  $d$ , then*

$$0 \leq h(g) \leq H(g) \leq \log(d).$$

In the real case, this is a best possible estimate for the entropy:

**THEOREM 4.2.** *Given any number  $h_0$  between 0 and  $\log(d)$ , and given any numbers  $d_i \geq 2$  with product  $d$ , there exist real generalized Hénon transformations  $g_i$  of degree  $d_i$  so that the topological entropy  $h(g_m \circ \cdots \circ g_1)$  is equal to  $h_0$ .*

Unfortunately, the proof is rather non-constructive. It seems quite difficult to estimate the topological entropy of any given  $g$ , or to decide when  $h(g)$  is actually equal to  $H(g)$ . In fact it seems plausible to conjecture that

$$h(g) = H(g)$$

for every real or complex polynomial automorphism. Smillie has recently shown that the equality

$$h(g) = H(g) = \log(d)$$

always holds in the complex cyclically reduced case.

*Proof of 4.1.* The inequality  $0 \leq h(g)$  is true by definition, and the inequality  $H(g) \leq \log(d)$ , is an immediate consequence of 3.1. The inequality  $h(g) \leq H(g)$  has been proved by [Ka p. 171] for any diffeomorphism of a real compact 2-dimensional manifold. The non-compactness of our coordinate plane does not affect his argument, since all of the interesting dynamics is concentrated in a neighborhood of the set  $\Omega(g)$ , which is compact by § 3.5. Thus, to complete the proof, we need only show that Katok's argument applies also to a complex 2-dimensional manifold. The key to his argument is the statement that any diffeomorphism of a 2-manifold with strictly positive entropy must have an invariant ergodic measure with no Lyapunov exponent equal to zero. He shows this by noting that, in any dimension, there must be at least one strictly positive Lyapunov exponent, and at least one strictly negative Lyapunov exponent. In the two dimensional case, this exhausts all of the Lyapunov exponents, so that none can be zero. But, for a complex diffeomorphism, the Lyapunov exponents of the underlying  $C^\infty$  manifold are clearly equal in pairs. So, in the complex 2-dimensional case, if at least one Lyapunov exponent is positive and at least one is negative, it again follows that no Lyapunov exponent can be zero.  $\square$

*Remark 4.3.* It follows easily that  $0 \leq h(g) \leq H(g) \leq \log(d)$  for a completely arbitrary polynomial automorphism  $g \in G$  of degree  $d$ . This in turn implies the sharper statement that

$$0 \leq h(g) \leq H(g) \leq \lim_{n \rightarrow \infty} \log(\text{degree } g^{\circ n})/n.$$

Here the right hand side is a conjugacy class invariant which vanishes for an affine or elementary transformation and coincides with  $\log(d)$  in the cyclically reduced case.

We will begin the proof of 4.2 in this section, but complete the proof only in § 5. The key to the proof is the observation that the topological entropy  $h(g)$  of a composition of generalized Hénon transformations of degrees  $d_1, \dots, d_m$  varies continuously with the  $d_1 + \cdots + d_m$  parameters of § 2.6. In fact upper semi continuity under  $C^\infty$  variation of the mapping has been proved by [Y] and by [N], while a proof of lower semi continuity in the 2-dimensional case has been promised by

[K M]. These papers all assume compactness. However, as  $g$  varies within a neighborhood of some fixed  $g_0$ , it follows from § 3.5 that the non-wandering set of  $g$  varies within some compact subset of the plane; and this suffices for the proofs.

*Remark 4.4.* This proof of continuity works only if we stay within the class of cyclically reduced group elements. It definitely is not true that the topological entropy  $h$  is a continuous function throughout the entire algebraic variety  $G[d_1, \dots, d_m]$  of § 2.4. Here is an example. Let

$$g_0(x, y) = (y, y^2 + c - \delta x)$$

be a classical Hénon transformation, and let  $\{\lambda_1, \lambda_2\}$  be the eigenvalues of the first derivative of  $g_0$  at one of the two fixed points of  $g_0$ . Then it is not difficult to check that the two parameters  $c$  and  $\delta$  can be expressed as polynomial functions of  $\lambda_1$  and  $\lambda_2$ . In fact

$$\delta = \lambda_1 \lambda_2, \quad c = \sigma(1 + \delta - \sigma),$$

where  $\sigma = (\lambda_1 + \lambda_2)/2$ . (Cf § 7.) We want to choose  $g_0$  so that these two eigenvalues are real, and so that the entropy  $h(g_0)$  is strictly positive. For example, for any fixed  $\delta$ , these conditions will be satisfied whenever  $\sigma$  is sufficiently large. (Cf § 5.) Now let

$$e_0(x, y) = (\lambda_1 x + y^2, \lambda_2 y),$$

so that the two eigenvalues of the derivative of  $e_0$  at the origin are also  $\lambda_1$  and  $\lambda_2$ . Then  $h(e_0) = 0$  since  $e_0$  is elementary, and yet every neighborhood of  $e_0$  within the variety  $G[2]$  contains transformations which are conjugate to  $g_0$  and hence have entropy equal to  $h(g_0) > 0$ . By way of contrast, if a sequence of elements  $g_i \in G[d_1, \dots, d_m]$  tends smoothly to a limiting function of degree  $d_1 \cdots d_m$  which has determinant  $\delta = 0$ , then it seems likely that the entropies  $h(g_i)$  must converge to a limit, which can be identified with the entropy of the associated one-dimensional polynomial map.

The proof of 4.2 continues as follows. Within each connected component of our  $d_1 + \cdots + d_m$  dimensional real parameter space, we will produce one example of a transformation with  $h(g) = H(g) = 0$ , and one example with  $h(g) = H(g) = \log(d)$ . It will then follow from the intermediate value theorem that all topological entropies between 0 and  $\log(d)$  must also occur. To provide an example with topological entropy zero, we simply apply Lemma 3.5. If each of our generalized Hénon transformations  $g_i : (x, y) \mapsto (y, p_i(y) - \delta_i x)$  satisfies

$$|p_i(y)| > |y| + |\delta_i y|$$

whenever  $|y| > \kappa$ , then 3.5 asserts that the non-wandering set  $\Omega(g)$  is contained in the square  $|x| \leq \kappa, |y| \leq \kappa$ . In particular, if we can take  $\kappa = 0$ , then it will follow that this non-wandering set is vacuous or reduces to a single point, so that the equality  $h(g) = H(g) = 0$  must hold. As an example, suppose that we define the polynomials  $p_i$  by the formula

$$p_i(y) = \pm(\eta y + y^{d_i})$$

when the degree  $d_i$  is odd, and

$$p_i(y) = \pm\eta(1 + y^{d_i})$$

when  $d_i$  is even, with  $\eta \geq 1 + |\delta|$ . Then a brief computation shows that we can take  $\kappa = 0$ , and it follows that  $h(g) = H(g) = 0$ .

Corresponding examples in which the entropy  $h(g)$  takes on its maximal possible value will be described in the next section.

5. *d-fold horseshoes*

This section will describe variants of the Smale horseshoe. Our condition that  $f$  is a  $d$ -fold ‘weak horseshoe’ is strong enough to guarantee that the topological entropy  $h(f)$  is at least equal to  $\log(d)$ , but weak enough to be easily verified in many cases. (For a general discussion of horseshoes, see for example [GH]. More detailed studies of horseshoes in the context of Hénon mappings may be found in [D] or [DN] for the real case, and in [H] for the complex case.)

Consider the following situation. (Cf figure 1.) Let  $D^n \subset \mathbb{R}^n$  be a closed  $n$ -dimensional ball, with boundary  $S^{n-1}$ . We consider maps from the product space  $\mathbb{R}^m \times \mathbb{R}^n$  into itself.

*Definition.* A map  $f$  from  $\mathbb{R}^m \times \mathbb{R}^n$  into itself will be called a *d-fold topological horseshoe*, with respect to the bidisk  $D^m \times D^n$ , if the intersection

$$(D^m \times D^n) \cap f^{-1}(D^m \times D^n)$$

is the union of disjoint sets  $X_1, \dots, X_d$ , each of which maps homeomorphically onto  $D^m \times D^n$  under the correspondence  $(x, y) \mapsto (x, y')$ , where  $f(x, y) = (x', y')$ . In other words we assume that, for each fixed  $x_0$ , the set of  $y$  with  $f(x_0, y) \in D^m \times D^n$  is a union of  $d$  disjoint topological disks; and furthermore, setting  $f(x_0, y) = (x', y')$ , we assume that each of these disjoint disks maps homeomorphically onto  $D^n$  by the correspondence  $y \mapsto y'$ . We will always assume that  $d \geq 2$ .

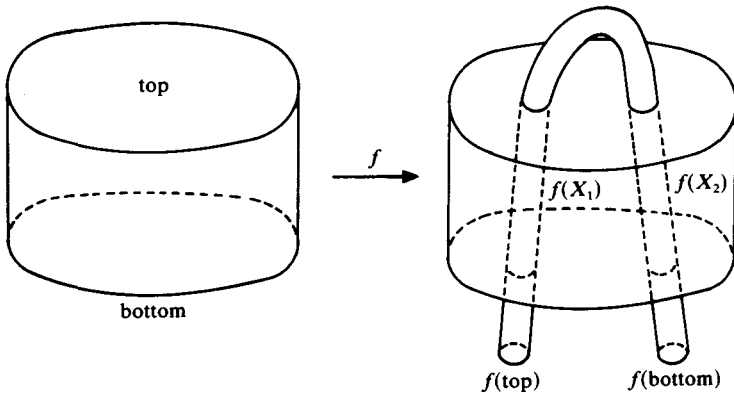


FIGURE 1. A 2-fold topological horseshoe in  $\mathbb{R}^2 \times \mathbb{R}^1$ .

If  $f$  is a homeomorphism from  $\mathbb{R}^m \times \mathbb{R}^n$  to itself, and also a  $d$ -fold topological horseshoe, note that the inverse transformation  $f^{-1}$  is itself a  $d$ -fold topological horseshoe, after renumbering the coordinates. (More explicitly, the linearly conjugate transformation  $t \circ f^{-1} \circ t^{-1}$  is a topological horseshoe, where  $t(x, y) = (y, x)$ .)

In the application to polynomial automorphisms, we will have either  $m = n = 1$ , to study real polynomial automorphisms, or  $m = n = 2$ , to study complex polynomial automorphisms. In both cases we will use the abbreviated notation  $D$  for the disk



$D^m = D^n$ . Let us first look at the real case. To fix our ideas, let  $D$  be the closed interval  $[0, d]$ . Note that the values of a polynomial  $p(y)$  of degree  $d$  can be prescribed arbitrarily at the points  $0, 1, \dots, d$  of  $D$ .

LEMMA 5.1. *If the real polynomial  $p(y)$  of degree  $d$  has values  $p(0), p(1), \dots, p(d)$  which are alternately less than  $-d|\delta|$  and greater than  $d + d|\delta|$ , then the generalized Hénon map  $g : (x, y) \mapsto (y, p(y) - \delta x)$  is a  $d$ -fold topological horseshoe with respect to the square  $[0, d] \times [0, d]$ .*

*Proof.* (Cf figure 2.) Let  $0 < c_1 < \dots < c_{d-1} < d$  be the critical points of  $p$ . It will be convenient to set  $c_0 = 0, c_d = d$ . Clearly the polynomial  $p$  maps each interval  $[c_{i-1}, c_i]$  homeomorphically onto an interval which contains  $[-d|\delta|, d + d|\delta|]$ . Let  $X_i$  be the set of all pairs  $(x, y) \in [0, d] \times [c_{i-1}, c_i]$  such that  $0 \leq p(y) - \delta x \leq d$ . Then evidently the correspondence

$$(x, y) \mapsto (x, p(y) - \delta x)$$

maps this set  $X_i$  homeomorphically onto  $[0, d] \times [0, d]$ , as required. □

It is interesting to note that the sign of the determinant  $\delta$ , and also the sign of the leading coefficient of  $p(y)$ , can be prescribed arbitrarily in this construction. Thus we obtain  $d$ -fold topological horseshoes in every isotopy class of generalized Hénon maps of degree  $d$ .

In the complex case, examples are also easy to find. Let  $D$  be the closed disk of radius  $r$  centered at the origin in  $\mathbb{C}$ , and let  $D' = (1 + |\delta|)D$  be the concentric disk of radius  $(1 + |\delta|)r$ .

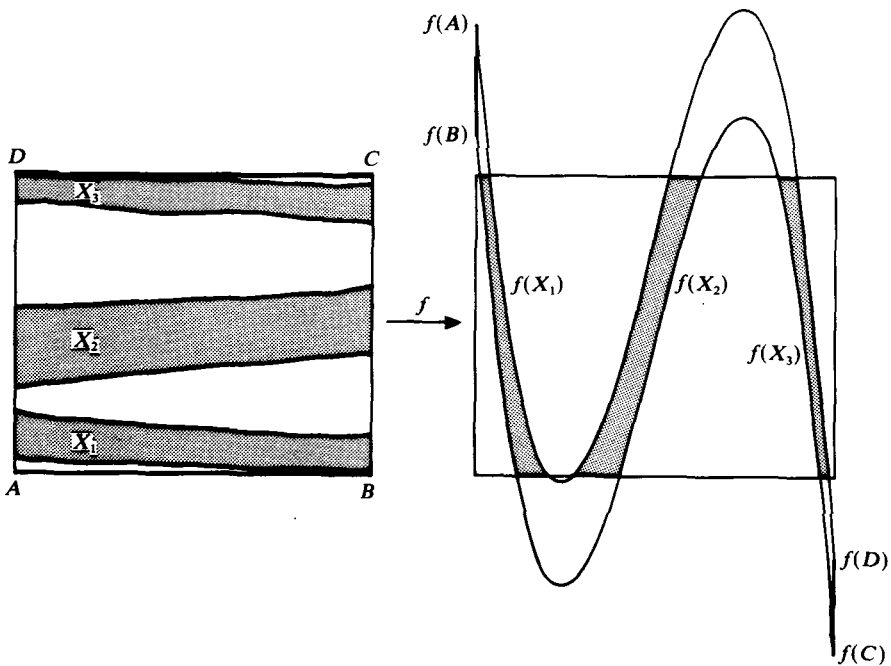


FIGURE 2. Three-fold horseshoe associated with a real cubic polynomial.

LEMMA 5.2. *If the complex polynomial  $p(y)$  of degree  $d \geq 2$  has no critical values in the disk  $D'$ , and if the full inverse image  $p^{-1}(D')$  is contained in  $D$ , then the map*

$$g : (x, y) \mapsto (y, p(y) - \delta x)$$

*is a  $d$ -fold topological horseshoe with respect to  $D \times D$ .*

*Proof.* For each fixed  $x \in D$ , note that the set of  $y$  with  $p(y) - \delta x \in D$ , is a union of  $d$  disjoint subdisks of  $D$ . It follows easily that the intersection  $(D \times D) \cap g^{-1}(D \times D)$  splits as a union  $X_1 \cup \dots \cup X_d$  of disjoint compact sets, each homeomorphic to  $D \times D$  as required. □

Next we will show that this condition is satisfied whenever the constant term of the polynomial  $p(y)$  is sufficiently large.

COROLLARY 5.3. *Suppose that we start with an arbitrary complex polynomial  $p(y)$  of degree  $d \geq 2$  and an arbitrary complex constant  $\delta \neq 0$ . Then there exists  $a_0 > 0$  so that, whenever  $|a| > a_0$ , the generalized Hénon map*

$$(x, y) \mapsto (y, p(y) + a - \delta x)$$

*is a  $d$ -fold topological horseshoe with respect to some bidisk  $D \times D$ .*

*Proof.* Choose a constant  $k$  so that  $|p(y)| > 3(1 + |\delta|)|y| - k$  for all  $y$ ; and let  $\mu$  be the maximum of the absolute values of the critical values of  $p$ . Then for  $|a| > k + 2\mu$  we can choose the radius  $r$  so that

$$(k + |a|)/2 < r(1 + |\delta|) < |a| - \mu,$$

and a brief computation shows that the hypotheses of 5.2 are satisfied. □

In order to study the properties of topological horseshoes, it will be convenient to introduce a more general concept. We will work with the quotient space

$$\Sigma = (D^m \times D^n) / (D^m \times S^{n-1}),$$

which is obtained from the bidisk by collapsing part of its boundary to a base point  $\sigma_0$ . Let  $q : D^m \times D^n \rightarrow \Sigma$  be the natural map, with  $q(D^m \times S^{n-1}) = \sigma_0$ . Note that  $\Sigma$  has the homotopy type of an  $n$ -sphere, so that it makes sense to refer to the ‘degree’ of a map from  $\Sigma$  to itself.

**Definition 5.4.** A map  $f$  from  $\mathbf{R}^m \times \mathbf{R}^n$  to itself will be called a  *$d$ -fold weak horseshoe* with respect to  $D^m \times D^n$  if there exists disjoint compact sets

$$X_1, \dots, X_d \subset D^m \times D^n,$$

each of which satisfies the following conditions. The map  $f$  must carry the relative boundary

$$X_i \cap \text{closure}((D^m \times \mathbf{R}^n) - X_i)$$

into  $D^m \times S^{n-1}$ , so that the correspondence  $f_i : q(x) \mapsto q(f(x))$ , for  $x \in X_i$ , extends to a continuous mapping  $f_i : \Sigma \rightarrow \Sigma$  which preserves the base point  $\sigma_0$ , and which maps everything outside of  $q(X_i)$  to  $\sigma_0$ . Furthermore, this map  $f_i$  must have degree  $\pm 1$ . In other words, it must induce an isomorphism of the homology group  $H_n(\Sigma)$  to itself.

LEMMA 5.5. *Every topological horseshoe is a weak horseshoe. Conversely, for the special case of real or complex polynomial automorphisms of degree  $d$ , every  $d$ -fold weak horseshoe is a  $d$ -fold topological horseshoe.*

*Proof.* The first statement is straightforward. To prove the second, for any  $x_0 \in D$  consider the polynomial map  $f_0: y \mapsto y'$  of degree  $\leq d$ , where  $f(x_0, y) = (x', y')$ . Note that all critical values of  $f_0$  must lie outside of the disk  $D$ . For if there were some critical value  $f_0(c) = v \in D$ , then the equation  $f_0(y) = v$  would have at most  $d - 1$  distinct solutions. But the condition that  $f$  is a  $d$ -fold weak horseshoe guarantees that this equation has  $d$  distinct solutions, all lying inside the disk  $D$ . It now follows easily that the preimage  $f_0^{-1}(D)$  consists of  $d$  disjoint disks, smoothly embedded in  $D$ , and varying smoothly with the parameter  $x_0 \in D$ .  $\square$

Following are some of the basic properties of weak horseshoes.

LEMMA 5.6. *If  $f$  is a  $d$ -fold weak horseshoe and  $f'$  is a  $d'$ -fold weak horseshoe, with respect to a common bidisk  $D^m \times D^n$ , then the composition  $f \circ f'$  is a  $dd'$ -fold weak horseshoe.*

*Proof.* This is essentially just the observation that each composition  $f_i \circ f'_j$  has degree  $\pm 1$ .  $\square$

LEMMA 5.7. *Given any infinite sequence of integers  $i(0), i(1), i(2), \dots$  between 1 and  $d$ , there exists at least one point  $x_0 \in X_{i(0)}$  whose successive images  $x_0 \mapsto x_1 \mapsto x_2 \mapsto \dots$  under iteration of  $f$  satisfy the condition that  $x_k \in X_{i(k)}$  for every  $k \geq 0$ .*

*Proof.* Since the  $n$ -fold composition

$$\Sigma \xrightarrow{f_{i(0)}} \Sigma \xrightarrow{f_{i(1)}} \dots \xrightarrow{f_{i(n-1)}} \Sigma$$

has degree  $\pm 1$ , it follows that there exists a point  $x_0 \in \Sigma$  whose image under this iteration is not the base point in  $\Sigma$ . By the definition of  $f_{i(k)}$ , this means that the image  $x_k = f^{ok}(x_0)$  must belong to  $X_{i(k)}$  for  $k = 0, 1, \dots, n - 1$ . Passing to the limit as  $n \rightarrow \infty$ , we see by a straightforward compactness argument that this condition can be satisfied for all  $k \geq 0$ .  $\square$

COROLLARY 5.8. *If  $f$  is a  $d$ -fold weak horseshoe, and if  $f$  is proper so that the topological entropy  $h(f)$  is defined, then  $h(f) \geq \log(d)$ .*

*Proof.* Let  $\epsilon > 0$  be the minimum distance between two distinct  $X_i$ . For any  $n > 0$  we have constructed  $d^n$  distinct orbits of length  $n$  such that any two differ by at least  $\epsilon$  somewhere. It follows that  $h(f) \geq \lim_{n \rightarrow \infty} \log(d^n)/n = \log(d)$ . (See for example [W, § 7.2].)  $\square$

If a polynomial automorphism  $g$  of degree  $d$  is a weak  $d$ -fold horseshoe, then clearly it follows that the entropy  $h(g) = H(g)$  is precisely equal to  $\log(d)$ . We can now complete a proof from the previous section.

*Proof of Theorem 4.2.* To finish this argument, we must find generalized Hénon maps  $g_1, \dots, g_m$  of specified degrees  $d_1, \dots, d_m$  so that the entropy  $h(g_m \circ \dots \circ g_1)$  of the composition is equal to  $\log(d_1 \cdots d_m)$ . Using Lemma 5.1, we simply choose each  $g_i$  to be a  $d_i$ -fold topological horseshoe with respect to some fixed square  $D \times D$ . The conclusion then follows by 5.6 and 5.8.  $\square$

Here is a further property of weak horseshoes.

LEMMA 5.9. *Every  $d$ -fold weak horseshoe  $f$  has at least  $d$  distinct fixed points.*

Hence the  $n$ -fold iterate  $f^{o n}$  has at least  $d^n$  distinct fixed points. In particular, in the case of a polynomial automorphism of degree  $d$  which is also a  $d$ -fold horseshoe, it follows that the number of fixed points of the  $n$ -fold iterate is precisely  $d^n$ .

*Proof of 5.9.* If  $f$  is a weak horseshoe, and  $X = X_i \subset D^m \times D^n$  is a closed subset satisfying the conditions of definition 5.4, then we will show that  $X$  contains at least one fixed point of  $f$ . Suppose, to fix our ideas, that  $D^m$  and  $D^n$  are the standard unit disks. Let

$$\partial X = X \cap \text{closure}(\mathbf{R}^m \times \mathbf{R}^n - X)$$

be the full boundary of  $X$ . For each  $(x, y) \in X$ , we will use the notation  $f(x, y) = (x', y') \in D^m \times D^n$ . Define a one-parameter family of maps  $\phi_t : X \rightarrow \mathbf{R}^m \times \mathbf{R}^n$  by the formula

$$\phi_t(x, y) = (x - tx', y' - ty)$$

for  $0 \leq t \leq 1$ . If  $f$  has no fixed point in  $\partial X$ , then we claim that each  $\phi_t$  maps  $\partial X$  into the complement of the origin. For a point  $(x, y) \in \partial X$  must satisfy either  $|y'| = 1 \geq |y|$  or  $|x| = 1 \geq |x'|$ . In either case, the pair  $(x - tx', y' - ty)$  can be zero only if  $t = 1$  and  $x = x'$  and  $y = y'$ , which is impossible since we have assumed that there are no fixed points in  $\partial X$ .

Next note that each of the resulting maps

$$\phi_t : (X, \partial X) \rightarrow (\mathbf{R}^m \times \mathbf{R}^n, \mathbf{R}^m \times \mathbf{R}^n - (0, 0)),$$

from the set  $X$  modulo its boundary to Euclidean space modulo the complement of the origin, has a well defined degree. For the case  $t = 0$ , the degree of this map  $\phi_0 : (x, y) \mapsto (x, y')$  is equal to  $\pm 1$ . In fact, for each fixed  $x$ , the definition 5.4 requires that the correspondence  $y \mapsto y'$  yields a map of degree  $\pm 1$  between appropriate sets; and this degree is not changed when we raise the dimension by adding  $x$  as an extra variable. Now, since the degree cannot change under homotopy, it follows that the map  $\phi_1 : (x, y) \mapsto (x - x', y' - y)$  must also have degree  $\pm 1$ . Hence the equation  $\phi_1(x, y) = (0, 0)$  must have a solution; and this solution is the required fixed point of  $f$  in  $X$ . □

*Concluding remark.* It should be possible to prove much sharper results about polynomial automorphisms of degree  $d$  which are also  $d$ -fold weak horseshoes. One would like to know that the set  $K = K(g)$  is always a hyperbolic Cantor set, and that  $g$  restricted to  $K$  is always homeomorphic to a full shift on  $d$  symbols. Furthermore, one would like to know that the sets  $K_+$  and  $K_-$  have measure zero, so that almost every orbit of  $g$  diverges to infinity in both forward and backward time. In the complex case, the methods of Hubbard and Oberste-Vorth should provide a strong attack on such questions. For the real case, compare [D] or [DN].

### 6. The elementary maps

This section will study the classification of elementary transformations up to conjugacy, either within the group  $E$  of elementary transformations, within the group

$G$  of polynomial automorphisms, or within the larger group consisting of all analytic diffeomorphisms of the (real or complex) plane. Here by *analytic* we mean real analytic if we are working over the real numbers, or complex analytic if we are working over the complex numbers. Given an elementary transformation

$$e(x, y) = (\alpha x + p(y), \beta y + \gamma),$$

we want to choose  $f$  within the required group so that the conjugate  $f^{-1} \circ e \circ f$  has the simplest possible description. First suppose that  $f(x, y) = (Ax + g(y), By + C)$  is also elementary. Computation shows that the conjugate  $\hat{e} = f^{-1} \circ e \circ f$  is given by  $\hat{e}(x, y) = (\alpha x + \hat{p}(y), \beta y + \hat{\gamma})$  with

$$A\hat{p}(y) = p(By + C) + \alpha q(y) - q(\beta y + \hat{\gamma}), \quad B\hat{\gamma} = (\beta - 1)C + \gamma. \quad (6.1)$$

In particular, the two coefficients  $\alpha$  and  $\beta$  are invariants of the  $E$ -conjugacy class of  $e$ .

*Remark.* If we allow conjugation by elements of the larger group  $G$ , then we can say only that the unordered pair  $\{\alpha, \beta\}$  is an invariant. In fact, if  $e$  has a fixed point, then  $\alpha$  and  $\beta$  are equal to the two eigenvalues of the derivative at that fixed point, and hence are actually invariant under analytic conjugacy. If there is no fixed point, then the pair  $\{\alpha, \beta\}$  must be equal to  $\{1, \delta\}$ , where the jacobian determinant  $\delta$  of  $e$  is clearly a  $G$ -conjugacy invariant.

In practice, we will break the problem into two steps by noting that  $E$  splits as the semi-direct product of its commutator subgroup  $E'$ , which consists of all transformations of the form

$$f(x, y) = (x + q(y), y + C),$$

with the subgroup of diagonal linear transformations  $f(x, y) = (Ax, By)$ . Thus we can conjugate by an arbitrary element of  $E$  in two steps, first conjugating by an element of  $E'$ , and then by a diagonal linear transformation.

**LEMMA 6.2.** *Every elementary transformation is  $E'$ -conjugate to a unique normal form which is either an affine transformation*

$$a(x, y) = (\alpha x, y + \gamma) \quad (6.2a)$$

with  $\gamma \neq 0$ , or a transformation of the form

$$e(x, y) = (\alpha x + ah(y), \beta y) \quad (6.2b)$$

where  $h(y)$  is a polynomial which satisfies the identity  $\alpha h(y) = h(\beta y)$ , and where the coefficient of the term of next to highest degree in  $h$  is zero if  $h$  is non-constant.

Note that in Case 6.2b the ‘horizontal’ line  $y = 0$  is invariant under  $e$ , but that in Case 6.2a no horizontal line  $y = \text{constant}$  is invariant. One very useful consequence of the equation

$$\alpha h(y) = h(\beta y) \quad (6.3)$$

is the following simple expression for the  $k$ -fold iterate of  $e$ .

$$e^{\circ k}(x, y) = (\alpha^k(x + kh(y)), \beta^k y). \quad (6.4)$$

The equation (6.3) can be best understood by expressing  $h(y)$  as a linear combination of monomials  $y^n$ , and noting that  $\alpha h(y) - h(\beta y)$  is then a corresponding linear

combination of monomials  $(\alpha - \beta^n)y^n$ . Thus (6.3) is satisfied if and only if  $h(y)$  is a linear combination of monomials  $y^n$  for which  $\alpha = \beta^n$ . In particular, if  $h(y)$  is non-zero of degree  $d$ , then the coefficients  $\alpha$  and  $\beta$  must satisfy the equation  $\alpha = \beta^d$ . If  $\beta$  is not a root of unity, then there can be no other integer  $n$  satisfying  $\alpha = \beta^n$ , hence  $h(y)$  must simply be a multiple of  $y^d$ . We can now give a precise normal form for conjugacy classes in  $E$  as follows.

**THEOREM 6.5.** *Every element of  $E$  is  $E$ -conjugate either to a diagonal linear transformation  $(x, y) \mapsto (\alpha x, \beta y)$ , to an aperiodic affine transformation*

$$(x, y) \mapsto (\alpha x, y + 1) \quad \text{or} \quad (x, y) \mapsto (x + 1, \beta y), \tag{6.5a}$$

*to a transformation of the form*

$$e(x, y) = (\beta^d(x + y^d), \beta y) \tag{6.5b}$$

*for some integer  $d \geq 1$ , or to a transformation*

$$e(x, y) = (\beta^\mu(x + y^\mu q(y')), \beta y) \tag{6.5c}$$

*for some  $\mu \geq 0$ , where  $\beta$  is a primitive  $r$ th root of unity and where  $q(z)$  is a non-constant polynomial of the form  $\pm z^k + q_{k-1}z^{k-1} + \dots + q_1z + 1$ , with leading coefficient  $+1$  in the complex case and with coefficient  $q_{k-1}$  equal to zero in the special case  $\beta = r = 1$ . This normal form is unique except that, in Case 6.5c, the polynomial  $q(z)$  may be replaced by  $q(\omega z)$ , where  $\omega$  can be any  $k$ th root of unity in the complex case, and where  $\omega = \pm 1$  when  $\beta = 1$  (but  $\omega = 1$  when  $\beta = -1$ ) in the real case.*

**Remark 6.6.** This integer  $\mu$ , if it is not zero, is called the *multiplicity* of the line  $y = 0$  as a line of periodic points of  $e$ . Similarly in Case 6.5b, if  $\beta$  is a root of unity, then the integer  $d \geq 1$  is called the multiplicity of the periodic line  $y = 0$ . More generally, if an analytic transformation  $(x, y) \mapsto (f_1(x, y), f_2(x, y))$  has a non-singular curve  $\eta(x, y) = 0$ , of periodic points, then the multiplicity  $\mu \geq 1$  of this curve is defined as follows. After replacing the map  $f$  by a suitable iterate, we may assume that  $\eta = 0$  is actually a curve of fixed points. The multiplicity is defined to be the largest integer  $\mu$  such that both of the analytic functions  $f_1(x, y) - x$  and  $f_2(x, y) - y$  are divisible by  $\eta(x, y)^\mu$  throughout a neighborhood of this fixed curve. Evidently  $\mu$  is well defined, and invariant under analytic conjugacy. (This  $\mu$  should not of course be confused with the multiplicity of an isolated fixed point, as discussed in Sections 3 and 8.)

*Proof of 6.2 and 6.5.* First suppose that  $\beta \neq 1$ . By 6.1 there is a unique choice of the vertical translation constant  $C$  so that  $\hat{\gamma} = 0$ . Then

$$\hat{p}(y) = p(y + C) + \alpha q(y) - q(\beta y).$$

If we express both  $p(y + C)$  and  $q(y)$  as linear combinations of monomials  $y^n$ , then evidently we can choose  $q(y)$  so as to kill off all of the monomial summands of  $\hat{p}(y)$  except those for which  $\alpha = \beta^n$ . Thus we obtain the normal form 6.2b. If  $\beta = 1$  and  $\gamma = 0$ , then the proof is essentially the same. However, in this case we are free to choose the constant  $C$  so as to set the center of gravity of the polynomial  $h(y)$  equal to zero, or in other words to set the next to highest coefficient equal to zero. Finally suppose that  $\beta = 1$  and  $\gamma \neq 0$ . Then it is not difficult to show that there exists

a polynomial  $q(y)$  which satisfies the equation

$$\hat{p}(y) = p(y) + \alpha q(y) - q(y + \gamma) = 0,$$

so that  $\hat{e}(x, y) = (\alpha x, y + \gamma)$ . This proves 6.2; and the proof of 6.5 is now completely straightforward.  $\square$

The problem of conjugacy within the larger group  $G$  can be settled as follows.

LEMMA 6.7. *An elementary map  $e(x, y) = (\alpha x + p(y), \beta y + \gamma)$  of degree  $d \geq 2$  has minimal degree within its  $E$ -conjugacy class if and only if  $\alpha = \beta^d$  and either  $\beta \neq 1$  or  $\gamma = 0$ . An element satisfying these conditions also has minimal degree within its  $G$ -conjugacy class, and two such are  $G$ -conjugate if and only if they are  $E$ -conjugate.*

To complete the picture, it is easy to see that two linear transformations are  $G$ -conjugate (or even analytically conjugate) if and only if they are linearly conjugate; and that two affine transformations without fixed point are  $G$ -conjugate if and only if they have the same determinant. The proof of 6.7 is based on § 2.1 and § 2.3 together with the arguments above, and will be left to the reader.  $\square$

Next let us study the dynamical properties of these elementary transformations, so as to prove the statements in § 3.10. The case of a linear transformation is easily dealt with. In the case of an affine transformation of the form 6.2a or 6.5a, it is clear that every orbit diverges to infinity, so that the sets  $\Omega = K$  are vacuous. Thus we need only prove the following.

LEMMA 6.8. *If  $e$  is an elementary transformation which is not  $E$ -conjugate to an affine transformation, then the non-wandering set  $\Omega(e)$  is equal to  $K(e)$ , and consists of either a single source or sink, or a finite union of horizontal lines on which  $e$  acts by an isometry. If there is more than one such horizontal line, then  $\Omega$  consists completely of periodic points.*

Here the condition that  $e$  is not  $E$ -conjugate to an affine map is needed to exclude examples like  $(x, y) \mapsto (2x - y^2, y)$ , which has a parabola of fixed points.

*Proof of 6.8.* After conjugating by an element of  $E$ , we may assume that  $e$  has the form

$$e(x, y) = (\alpha(x + h(y)), \beta y),$$

with  $\alpha h(y) = h(\beta y)$  as in 6.2b, and where  $h(y)$  is non-linear. Recall from 6.4 that the  $n$ -fold iterate of  $e$  is given by

$$e^{\circ n}(x, y) = (\alpha^n(x + nh(y)), \beta^n y).$$

If  $|\alpha|, |\beta| < 1$ , then every non-trivial orbit converges to the origin as  $n \rightarrow +\infty$  and diverges to infinity as  $n \rightarrow -\infty$ ; hence the set  $\Omega = K$  consists of a single sink at the origin. Similarly, if  $|\alpha|, |\beta| > 1$ , then  $\Omega = K$  consists of a source at the origin. If  $|\alpha| = |\beta| = 1$ , then the set  $\Omega = K$  coincides with the locus  $h(y) = 0$ ; for every point in this locus has a bounded recurrent orbit, and every point outside it has an orbit which diverges to infinity. In Case 6.5b, where  $h(y) = y^d$ , this locus consists of the coordinate line  $y = 0$ . In Case 6.5c, where the polynomial  $h(y)$  is not homogeneous, this locus consists of a finite union of horizontal lines, and every point in it is periodic, with period dividing  $r$ . (In the complex case the locus  $h(y) = 0$  must



contain at least two distinct lines, but in the real case it may even be vacuous.) Note that the same description holds for any transformation which is conjugate to  $e$  within the group  $E$ , since the elements of  $E$  carry horizontal lines to horizontal lines. This proves 6.8, and the proof of 3.10 can easily be completed.  $\square$

Next let us study the analytic behavior of an elementary transformation in the neighborhood of a fixed point, which we may take to be the origin.

LEMMA 6.9. *An elementary transformation  $e(x, y) = (\alpha x + p(y), \beta y)$ , with a fixed point at the origin, is locally analytically conjugate to a linear map if and only if either*

- (1) *it is already  $E$ -conjugate to a linear map, or*
- (2) *the coordinate line  $y = 0$  is a line of periodic points and has multiplicity  $\mu = 1$ .*

*Proof.* By 6.2, we may take  $e$  to have the normal form  $e(x, y) = (\alpha x + \alpha h(y), \beta y)$ , with  $\alpha h(y) = h(\beta y)$ . In Case (2) the polynomial  $h(y)$  must be divisible by  $y$  but not by  $y^2$ . (Cf 6.6.) Therefore the correspondence  $f(x, y) = (x, h(y))$  is an analytic diffeomorphism on some neighborhood of the line  $y = 0$ . (In the real case, it may even define a global analytic diffeomorphism of  $\mathbf{R}^2$ . For example this is the case if  $h(y) = y + y^3$ .) Now computation shows that  $f \circ e \circ f^{-1}$  is equal to the linear transformation  $(x, y) \mapsto \alpha(x + y, y)$ . Conversely, if neither (1) nor (2) is satisfied, then the polynomial  $h(y)$  must be divisible by  $y^2$ . In particular, the first derivative of  $e$  at the origin must be a diagonal linear transformation. Therefore, if  $e$  were analytically conjugate to any affine map, it would necessarily be analytically conjugate to a diagonal linear map. This would imply the existence of two transverse  $e$ -invariant curves through the origin, corresponding to the two coordinate lines for the diagonal linear map. In fact there clearly does exist one smooth  $e$ -invariant curve through the origin, namely the coordinate line  $y = 0$ . Any second curve transverse to this one could be described locally as the graph of an analytic function  $x = \phi(y)$ . Such a curve is  $e$ -invariant if and only if the identity

$$\alpha(\phi(y) + h(y)) = \phi(\beta y)$$

is satisfied. But an argument completely analogous to the proof of 6.2 shows that there can be no such function  $\phi(y)$ .  $\square$

In the complex case, we can give a precise classification of elementary maps up to analytic conjugacy.

THEOREM 6.10. *Let  $e$  and  $f$  be two complex elementary maps each of which has at least one periodic point. Then  $e$  and  $f$  are conjugate in the full group of complex analytic diffeomorphisms of  $\mathbf{C}^2$  if and only if they are already conjugate within the group  $G = G_{\mathbf{C}}$ .*

The proof begins as follows. It is easy to check that two linear maps are analytically conjugate if and only if they are linearly conjugate. Thus we need only consider the normal forms of 6.5b and 6.5c. Note that these two cases are essentially different from each other, since the non-wandering set is a point or a line in case 6.5b, but is the union of two or more parallel lines in case 6.5c. For the transformation  $e(x, y) = (\beta^d(x + y^d), \beta y)$  of 6.5b, if  $d = 1$  then we are in the linear case, while if  $d > 1$  then  $e$  is not locally analytically conjugate to a linear map by Lemma 6.9,



and the normal form is uniquely determined by the two eigenvalues, together with the multiplicity of the  $x$ -axis as a line of periodic points when the eigenvalues are roots of unity. Thus we need only consider the case 6.5c. Let  $e(x, y) = (\alpha(x + h(y)), \beta y)$  where  $\alpha h(y) = h(\beta y)$ , and where  $\beta$  is a primitive  $r$ th root of unity. Recall that the non-wandering set  $\Omega$  coincides with the locus  $h(y) = 0$ , and consists of two or more horizontal lines. Any conjugacy to another elementary transformation would have to carry  $\Omega$  to some (possibly different) set of horizontal lines. In fact each of these lines has a certain multiplicity as a zero of  $h(y)$ , or equivalently as a periodic line of  $e$ , and again it is not difficult to check that these multiplicities are local analytic invariants. We will need the following result which was suggested to us by W. Thurston and by C. Fefferman.

LEMMA 6.11. *If a complex analytic diffeomorphism of  $C^2$  carries two or more horizontal lines  $y = \text{constant}$  into horizontal lines, then it carries every horizontal line to a horizontal line, and hence has the form*

$$f(x, y) = (x e^{\phi(y)} + \psi(y), Cy + D).$$

*Proof.* Suppose that some horizontal line  $y = c$  did not map to a horizontal line. Then the projection of  $f(x, c)$  to the second coordinate would be a non-constant entire holomorphic function of  $x$  which omits at least two values. This is impossible by Picard's Theorem; and the rest of the argument is straightforward.  $\square$

Thus, after composing with the inverse of the affine map  $(x, y) \mapsto (x, Cy + D)$ , we may assume that every horizontal line maps into itself. The proof of 6.10 now continues as follows. Suppose that  $e$  is analytically conjugate to another transformation  $e'$  of the same form. After conjugating by an elementary affine transformation, we may assume that  $e'(x, y) = (\alpha'x + p'(y), \beta'y)$  where the two polynomials  $p$  and  $p'$  have the same zeros, with the same multiplicities. Thus  $p'$  is some multiple of  $p$ , and after conjugating by an appropriate diagonal linear transformation we may assume that  $p' = p$ . Finally, the root of unity  $\beta$  has a geometric description since the transformation  $e$  acts as a rotation by  $\beta$  on the set of lines in  $\Omega$ . Hence this coefficient  $\beta$  must also be preserved by the analytic conjugacy.  $\square$

We conclude by noting that the hypothesis that our maps have periodic points is essential.

LEMMA 6.12. *An elementary map without periodic points is always analytically conjugate to a translation in the complex case, or to a map of the form  $a : (x, y) \mapsto (\pm x, y + 1)$  in the real case.*

*Proof.* In the complex case, it follows from 6.5 that such a map must be  $G$ -conjugate to an affine map of the form  $a(x, y) = (\delta x, y + 1)$ . Setting  $\delta = \exp(k)$ , we have

$$f^{-1} \circ e \circ f(x, y) = (x, y + 1),$$

where  $f$  is the global analytic diffeomorphism  $f(x, y) = (\exp(ky)x, y)$ . In the real affine case, we can only set  $\delta$  equal to  $\pm \exp(k)$ , thus obtaining an analytic conjugacy between the maps  $(x, y) \mapsto (\delta x, y + 1)$  and  $(x, y) \mapsto (\pm x, y + 1)$ . In the real case, we

may also have an aperiodic elementary transformation of the form 6.5c:

$$e(x, y) = (x + p(y), \pm y),$$

where  $p(y) = p(\pm y)$  is a polynomial without real zeros. In this case, the diffeomorphism  $f(x, y) = (x/p(y), y)$  satisfies  $f \circ e \circ f^{-1}(x, y) = (x + 1, \pm y)$ . □

7. Uniqueness of the quadratic and cubic Hénon maps

Let  $f$  be a complex polynomial automorphism which has only isolated fixed points, and let  $n$  be the sum of the multiplicities  $\mu \geq 1$  of these fixed points. If  $n \leq 1$ , then it follows from §§ 2 and 3 that  $f$  is conjugate in the group  $G = C_C$  to an elementary map. On the other hand, if  $n \geq 2$ , then  $f$  is conjugate to a composition of (generalized) Hénon maps which has degree  $d$  equal to  $n$ . According to § 6.8, two maps with  $n = 1$  are conjugate by a complex analytic diffeomorphism of  $C^2$  if and only if they are conjugate in  $G$ . It seems plausible to conjecture that the same statement holds when  $n > 1$ . However, we have been able to prove this only in the cases  $n = 2$  and  $n = 3$ .

**THEOREM 7.1.** *Assume that  $f$  and  $g$  are complex polynomial automorphisms of  $C^2$  which have either two or three isolated fixed points, counted with their multiplicities. If  $f$  and  $g$  are complex analytically conjugate then  $f$  and  $g$  are conjugate in  $G$ .*

This theorem remains valid in  $R^2$  if we assume that  $f$  and  $g$  are real Hénon maps of degree at most three whose fixed points are all real.

Before we start our computations we make the following observations. Suppose that two polynomial automorphisms  $f$  and  $g$  of  $C^2$  are conjugate by some homeomorphism  $T$  from  $C^2$  to itself. Then every fixed point  $(x, y)$  of  $f$  is mapped by  $T$  to a fixed point  $(x', y')$  of  $g$ , and this one-to-one correspondence between fixed points preserves the multiplicity of each fixed point. Thus the algebraic number  $n$  of fixed points is invariant under topological conjugacy. In particular, if  $f$  and  $g$  are cyclically reduced, then they must have the same degree. Now suppose that  $f$  and  $g$  are analytically conjugate under a complex analytic diffeomorphism  $T$  of  $C^2$ . Then  $T$  induces a similarity between the  $2 \times 2$  complex Jacobian matrices  $J(f)(x, y)$  and  $J(g)(x', y')$ . In particular, these two matrices have the same trace and the same determinant. The same arguments apply to the periodic points of  $f$  and  $g$ . Thus  $T$  induces a one-to-one correspondence between the periodic points of  $f$  and  $g$  which preserves the period, the multiplicity, and the trace associated with each periodic point. In particular, if the polynomial automorphisms  $f$  and  $g$  are cyclically reduced of degree  $d = n \geq 2$ , then we get infinitely many distinct complex invariants which must coincide for  $f$  and  $g$ . (In principle, this yields an infinite number of polynomial identities which must be satisfied by the coefficients of the maps  $f$  and  $g$ .) This helps to motivate our conjecture that  $f$  and  $g$  must be conjugate in  $G$ .

If  $f$  has a prime number of fixed points, counted with their multiplicities, then Corollary 2.7 yields that  $f$  is conjugate to a (generalized) Hénon map. In what follows we shall assume that  $f$  and  $g$  are two Hénon maps of degree  $n$

$$f(x, y) = (y, p(y) - \delta x), \quad g(x, y) = (y, P(y) - \Delta x),$$

which are normalized as in Theorem 2.6 so that the polynomials  $p(y)$  and  $P(y)$  have leading coefficient +1 and next coefficient zero. Each fixed point of  $f$  is of the form  $(x, x)$  with  $p(x) - \delta x = x$ , and with

$$\det(J(f)(x, x)) = \delta, \quad \text{trace}(J(f)(x, x)) = p'(x).$$

If  $f$  and  $g$  are analytically conjugate, then it follows that  $\delta = \pi$ . Suppose that the fixed points of  $f$  are  $(x_j, x_j)$  for  $j = 1, \dots, n$ . We deduce that

$$p(x) = (x - x_1) \cdots (x - x_n) + (1 + \delta)x.$$

In the case  $n = 2$ , since the trace at  $(x_j, x_j)$  is equal to  $2x_j$  and is an analytic conjugacy invariant, the normal form is uniquely determined. On the other hand if  $n > 2$ , since the center of gravity of  $p(x)$  is 0, we have

$$x_1 + x_2 + \cdots + x_n = 0. \tag{7.2}$$

If the fixed points of  $g$  are  $(y_j, y_j), j = 1, \dots, n$ , where  $T$  carries  $(x_j, x_j)$  to  $(y_j, y_j)$ , then equating the traces of  $f$  and  $g$  at the fixed points we get the identities

$$\begin{aligned} (x_j - x_1) \cdots (x_j - x_{j-1})(x_j - x_{j+1}) \cdots (x_j - x_n) \\ = (y_j - y_1) \cdots (y_j - y_{j-1})(y_j - y_{j+1}) \cdots (y_j - y_n) \end{aligned} \tag{7.3}$$

for  $j = 1, \dots, n$ .

LEMMA 7.4. *Let  $\{x_1, \dots, x_n\}$  and  $\{y_1, \dots, y_n\}$  be two sets of  $n$  complex numbers, with  $n \leq 3$ . Assume that the center of the gravity of each the two sets is zero, and that the identities (7.3) hold. Then*

$$(x - x_1) \cdots (x - x_n) = (x - \omega y_1) \cdots (x - \omega y_n)$$

for some  $\omega$  with  $\omega^{n-1} = 1$ .

*Proof.* For  $n = 2$  the equalities (7.2) and (7.3) yield immediately that  $x_1 = y_1$  and  $x_2 = y_2$ . For  $n = 3$ , adding the equations (7.3) in pairs, we obtain three equations of the form  $(x_i - x_j)^2 = (y_i - y_j)^2$ . Combining these new equations with (7.3), we see that there is a sign  $\omega = \pm 1$  so that  $(x_i - x_j) = \omega(y_i - y_j)$  for every  $i$  and  $j$ . Since  $x_1 + x_2 + x_3 = 0$  and therefore  $3x_i = (x_i - x_j) + (x_i - x_k)$ , it follows that  $x_i = \omega y_i$  as required. This proves 7.4; and Theorem 7.1 follows easily.  $\square$

The following example due to McKay shows that Lemma 7.4 is false for  $n = 4$ . Let

$$(x_1, x_2, x_3, x_4) = (1, -9, 5, 3), \quad (y_1, y_2, y_3, y_4) = (7, -13, 1, 5)/3^{1/3}.$$

Thus any proof of 7.1 for maps of degree four or more would have to make use of periodic points of period greater than one.

To conclude this section, let us study what happens if we allow real diffeomorphisms (not preserving the complex structure) of the space  $C^2$ . We will only be able to deal with the classical Hénon maps with degree  $n = 2$ . We will identify  $C^2$  with the real vector space  $R^4$ , so that every polynomial automorphism  $f$  of  $C^2$  can also be considered as a real polynomial automorphism of  $R^4$ . Let  $z$  be a fixed point of  $f$ . Denote by  $A = J(f)(z)_C$  and  $B = J(f)(z)_R$  the  $2 \times 2$  complex valued and the  $4 \times 4$  real valued Jacobian matrices of  $f$  at  $z$ . If  $A$  has eigenvalues  $\{\alpha, \beta\}$  and characteristic polynomial  $(t - \alpha)(t - \beta) = t^2 - \tau t + \delta$ , then  $B$  has eigenvalues  $\{\alpha, \bar{\alpha}, \beta, \bar{\beta}\}$  and characteristic polynomial  $(t^2 - \tau t + \delta)(t^2 - \bar{\tau} t + \bar{\delta})$ . In fact  $B$  is similar over  $C$  to the

direct sum of  $A$  and  $\bar{A}$ , its complex conjugate. Denote by  $f^*$  the polynomial automorphism of  $C^2$  obtained from  $f$  by conjugating all the coefficients appearing in the components of  $f$ . Thus  $f$  and  $f^*$  are real linearly conjugate by the involution  $(x, y) \mapsto (\bar{x}, \bar{y})$ .

**THEOREM 7.5.** *Let  $f$  and  $g$  be two polynomial automorphisms of  $C^2$  which have exactly two fixed points counted with their multiplicities. Assume that  $f$  and  $g$  are conjugate by a real diffeomorphism of  $C^2$ . Then  $g$  is conjugate either to  $f$  or to  $f^*$  in  $G$ .*

*Proof.* We may assume that

$$f(x, y) = (y, y^2 - \delta x + c), \quad g(x, y) = (y, y^2 - \Delta x + C).$$

Assume that the fixed points of  $f$  are  $(u, u)$  and  $(v, v)$  and that the fixed points of  $g$  are  $(U, U)$  and  $(V, V)$ . Note that the trace  $\tau$  of the Jacobian matrix of  $f$  at the fixed point  $(u, u)$  is equal to  $2u$ . A brief computation shows that

$$u + v = \delta + 1, \quad uv = c, \quad U + V = \Delta + 1, \quad UV = C. \tag{7.6}$$

Suppose that the diffeomorphism  $T$  carries  $(u, u)$  to  $(U, U)$  and  $(v, v)$  to  $(V, V)$ . Denote by  $A_1$  and  $A_2$  the  $2 \times 2$  complex valued Jacobian matrices of  $f$  at  $(u, u)$  and  $(v, v)$  and by  $B_1$  and  $B_2$  the  $2 \times 2$  complex valued Jacobian matrices of  $g$  at  $(U, U)$  and  $(V, V)$  respectively. Then  $A_j \oplus \bar{A}_j$  is similar to  $B_j \oplus \bar{B}_j$  for  $j = 1, 2$ . Equating the characteristic polynomials of these two matrices for  $j = 1$ , we obtain the equations

$$|\delta|^2 = |\Delta|^2, \quad \text{Re}(\delta \bar{u}) = \text{Re}(\Delta \bar{U}),$$

$$\text{Re}(\delta) + 2|u|^2 = \text{Re}(\Delta) + 2|U|^2, \quad \text{Re}(u) = \text{Re}(U),$$

and there are corresponding equations for  $j = 2$ , with  $v$  and  $V$  in place of  $u$  and  $U$ . It then follows from (7.6) that  $\text{Re}(\delta) = \text{Re}(\Delta)$ ; and hence  $\text{Im}(\delta) = \pm \text{Im}(\Delta)$ ,  $\text{Im}(u) = \pm \text{Im}(U)$  and  $\text{Im}(v) = \pm \text{Im}(V)$ . Using the equation  $\text{Re}(\delta \bar{u}) = \text{Re}(\Delta \bar{U})$ , we see that the three signs must either all be  $+1$  or all  $-1$ , so that  $(\Delta, C) = (\delta, c)$  or  $(\Delta, C) = (\bar{\delta}, \bar{c})$ . □

8. *Fixed point multiplicities*

This section will supplement § 3 by discussing the multiplicity of an isolated fixed point of a complex analytic map  $f$ . By definition, a fixed point  $x$  is ‘simple’ if the graph of  $f$  intersects the diagonal transversally at the point  $(x, x)$ , or equivalently if the eigenvalues of the first derivative of  $f$  at  $x$  are all different from  $+1$ . The multiplicity of an arbitrary isolated fixed point can be defined intuitively as the unique integer  $\mu$  with the property that an arbitrarily small perturbation of  $f$  will replace this fixed point by a cluster of  $\mu$  nearby fixed points, each of which is simple. The proof that there is a well defined integer  $\mu > 0$  with these properties can be carried out either by topological methods, as described for example in [Mil, Appendix B], or by algebraic methods, as described in [Fu, pp. 7-14]. However, for computational purposes, it is convenient to choose a more algebraic form of the definition.

In practice we are interested in the case of two complex variables. It is convenient to write the given map as

$$f(x, y) = (x + p(x, y), y + q(x, y)). \tag{8.1}$$

Thus a fixed point of  $f$  is just a common zero of the complex valued functions  $p$  and  $q$ , and such a fixed point  $x = (x, y)$  is simple if and only if the Jacobian determinant  $\partial(p, q)/\partial(x, y)$  is non-zero at  $x$ . Whether or not this determinant is zero, the multiplicity  $\mu$  could be defined topologically as the local degree of the mapping  $(x, y) \mapsto (p, q)$  at  $x$ . Algebraically, it is defined as follows. Let us suppose for convenience that  $x = 0$ .

*Definition.* If the analytic map (8.1) has an isolated fixed point at the origin, then the *multiplicity*  $\mu$  of this fixed point is equal to the codimension of the ideal generated by  $p$  and  $q$  in the local ring  $C[[x, y]]$  consisting of all formal power series in the indeterminates  $x$  and  $y$ .

Note that  $\mu = 1$  if and only if the origin is a simple fixed point. Here we could equally well work in the subring  $C\{x, y\}$  of convergent power series (alias germs of complex valued holomorphic functions). If  $p$  and  $q$  are polynomials, then we could work in the local ring consisting of rational functions whose denominator does not vanish at the given point. For the proof that these alternative versions of the definition are equivalent, the reader is referred to [Fu]. All of these definitions extend easily to the case of a complex analytic map in dimension three or more.

We want to study the relationship between the multiplicity of the origin as a fixed point of the mapping  $f: C^2 \rightarrow C^2$  and its multiplicity as a fixed point of the various iterates  $f^{\circ n}$ . To this end, it turns out to be convenient to study curves through the origin which are mapped into themselves by  $f$ . In order to avoid convergence problems, we will study only 'formal' curves, that is curves which are defined by formal power series of the form

$$t \mapsto \gamma(t) = v_1 t + v_2 t^2 + \dots$$

Here  $v_1$  is to be a non-zero vector in  $C^2$ , called the *tangent vector* of  $\gamma$  at the origin. Such a curve is *invariant* under  $f$  if  $f(\gamma(t)) = \gamma(g(t))$  for some formal power series  $g(t) = g_1 t + g_2 t^2 + \dots$ . If  $\gamma$  is invariant, note that the tangent vector  $v_1$  must be an eigenvector of the first derivative of  $f$  at the origin. Conversely, we will prove the following.

**LEMMA 8.2.** *Let  $\alpha$  and  $\beta$  be the eigenvalues of the first derivative of  $f$  at the origin, and let  $v_1 \neq 0$  be an eigenvector corresponding to the eigenvalue  $\alpha$ . If  $\beta \neq \alpha^k$  for every integer  $k \geq 2$ , then there exists one and only one formal curve with tangent vector  $v_1$  which is invariant under  $f$ .*

*Proof.* After conjugating by a linear change of coordinates, we may assume that  $v_1 = (1, 0)$ . Hence any formal curve tangent to  $v_1$  can be reparametrized as the graph of a formal power series

$$y = \phi(x) = \phi_2 x^2 + \phi_3 x^3 + \dots$$

If we set  $f(x, y) = (f_1(x, y), f_2(x, y))$ , then the condition of  $f$ -invariance is expressed by the equation

$$f_2(x, \phi(x)) - \phi(f_1(x, \phi(x))) = 0$$

in  $C[[x]]$ . Here  $f_1$  and  $f_2$  are given, and this equation must be solved for the

coefficients  $\phi_k$ . Using the fact that  $(1, 0)$  is an eigenvector with eigenvalue  $\alpha$ , we see by a straightforward computation that the coefficient of  $x^k$  in this expression is equal to  $(\alpha^k - \beta)\phi_k$  plus a polynomial in the  $\phi_j$  with  $j < k$ . Therefore, we can solve inductively for the required coefficients  $\phi_k$ . □

Here are two examples to show that the conditions on  $\alpha$  and  $\beta$  are necessary. (See also § 6.) If  $f(x, y) = (y, y^2 + x)$  with eigenvalues  $\pm 1$ , then Lemma 8.2 shows that there exists an invariant curve corresponding to the  $+1$  eigenvalue; but a brief computation shows that there is no invariant curve corresponding to the  $-1$  eigenvalue. For the map  $f(x, y) = (x + y^3, y + x^2)$  with both eigenvalues equal to  $+1$ , a similar computation shows that there are no invariant curves at all.

*Remark 8.3.* If we assume also that  $|\alpha| \neq 1$ , with  $\alpha\beta \neq 0$ , then the invariant formal curve which is constructed in 8.2 is actually given by a convergent power series. In fact, if the origin is either an attractive or a repulsive fixed point, then according to [R] the map  $f$  is conjugate by a germ of an analytic diffeomorphism either to a linear diagonal map  $(x, y) \mapsto (\alpha x, \beta y)$  or to the map  $(x, y) \mapsto (\alpha x + y^k, \beta y)$  where  $\sigma = \beta^k$  for some positive integer  $k$ . (Compare § 6.5.) In both cases  $y = 0$  is an invariant curve in the  $\alpha$  direction. On the other hand, if  $|\alpha| < 1 \leq |\beta|$  or  $|\alpha| > 1 \geq |\beta|$ , then the existence of such an analytic invariant curve follows from the ‘stable manifold theorem’. Standard proofs of this theorem for the real case, as described for example in [Sh], extend easily to the complex case. *If we assume also that  $f$  is an analytic diffeomorphism of  $C^2$ , then such an invariant curve extends uniquely to an invariant embedded submanifold which is analytically diffeomorphic to  $C$  and contains no other periodic point of  $f$ .* In fact such a submanifold has a preferred complex coordinate  $z$ , well defined up to multiplication by a constant, characterized by the fact that  $f$  maps a point with coordinate  $z$  to a point with coordinate  $\alpha z$ . This preferred coordinate can first be constructed locally; and the map  $f$  can then be used to extend it.

The main result in this section is the following.

**THEOREM 8.4.** *Let  $f: C^2 \rightarrow C^2$  be an analytic map with a fixed point of multiplicity  $\mu > 1$  at the origin. If the Jacobian determinant  $\delta$  of  $f$  at the origin is not equal to  $+1$ , then every iterate  $f^{\circ n}$  with  $\delta^n \neq 1$  has a fixed point with this same multiplicity  $\mu$  at the origin. In particular, if  $\delta$  is not a root of unity, then every  $f^{\circ n}$  with  $n \neq 0$  has this same fixed point multiplicity at the origin.*

*Proof.* Evidently the two eigenvalues of the first derivative at the origin are equal to 1 and to  $\delta \neq 1$ . By 8.2 there is an invariant formal curve corresponding to the  $+1$  eigenvalue, and after conjugating by a formal change of variable we may assume that this invariant curve is just the  $x$ -axis. In other words we can replace  $f$  by a formal map of the form  $g(x, y) = (g_1(x, y), g_2(x, y))$  with  $g_1, g_2 \in C[[x, y]]$ , where  $g_2(x, 0) = 0$ . We now claim that  $g_1(x, 0) = x + x^\mu u(x)$ , where  $\mu > 1$  is the multiplicity and where  $u(x)$  is a unit in the ring  $C[[x]]$ . In other words

$$g(x, 0) = (x + cx^\mu + (\text{higher order terms}), 0) \tag{8.5}$$

with  $c \neq 0$ . To prove this, recall that  $\mu$  is defined as the codimension of the ideal

generated by  $g_1 - x$  and  $g_2 - y$  in the local ring  $C[[x, y]]$ . But

$$g_2(x, y) = y(\delta + (\text{higher order terms}))$$

with  $\delta \neq 1$ , hence  $g_2 - y$  is equal to  $y$  multiplied by a unit in this ring. Thus if we factor out by the ideal generated by  $g_2 - y$ , then we have effectively set  $y$  equal to zero. It follows that the multiplicity  $\mu$  is equal to the codimension of the ideal generated by  $g_1(x, 0) - x$  in  $C[[x]]$ . The required formula (8.5) follows. (Caution: note that this argument is valid only if  $\delta \neq 1$ .)

Now if we replace the formal map  $g(x, y)$  by its  $n$ -fold iterate, then the  $x$ -axis will remain an invariant curve, and the expression  $g_1(x, 0) = x + cx^\mu + (\text{higher order terms})$  will simply be replaced by  $x + ncx^\mu + (\text{higher order terms})$ . If the Jacobian determinant  $\delta^n$  is not equal to 1, then the formula (8.5) still applies, so it follows that the multiplicity  $\mu$  does not change. □

On the other hand, if  $\delta^n = 1$ , then the multiplicity will definitely increase. As an example, the map  $f(x, y) = (y, x + y^n)$  has fixed point multiplicity  $\mu = n$  at the origin, but  $f \circ f$  has multiplicity  $\mu = n^2$ .

**COROLLARY 8.6.** *Every cyclically reduced polynomial automorphism of the complex plane with Jacobian determinant  $\delta \neq 1$  has periodic points of period  $p$  for every sufficiently large prime  $p$ .*

*Proof.* If  $f$  has degree  $d$ , then we know that the algebraic number of fixed points of  $f^{op}$  is equal to  $d^p > d$ . If there are no periodic points of period  $p$ , then there must be at least one fixed point of  $f$  whose multiplicity as a fixed point of  $f^{op}$  is strictly larger than its multiplicity as a fixed point of  $f$ . Let us call a prime  $p$  ‘exceptional’ for  $f$  if it has this property. Let  $\alpha$  and  $\beta$  be the eigenvalues of the first derivative at the fixed point. If  $\alpha \neq 1$  and  $\beta \neq 1$ , so that the fixed point  $x$  is simple, and if  $x$  is not simple as a fixed point of  $f^{op}$ , then it clearly follows that either  $\alpha^p = 1$  or  $\beta^p = 1$ . On the other hand, if  $\alpha = 1$  and hence  $\beta = \delta \neq 1$ , then it follows from 8.4 that the multiplicity can increase only if  $\delta^p = 1$ . Since the map  $f$  can have at most  $d$  fixed points, and since each equation of the form  $\alpha^p = 1$  with  $\alpha \neq 1$  can have at most one prime solution, it follows that there can be at most  $2d$  exceptional primes. □

*Added in Proof.* The hypothesis that  $\delta \neq 1$  can be eliminated by using Shub and Sullivan’s 1974 result that the sequence of fixed point multiplicities of the iterates of any  $C^1$  map at a fixed point is necessarily bounded. [*Topology* **13**, p. 189. See also Chow, Mallet-Paret and Yorke, Springer Lecture Notes **1007** (1983), 109–131.]

In fact it is conjectured that there must exist periodic points of *all* sufficiently large periods, for any  $\delta$ .

To conclude this section, let us describe some other methods for computing or estimating the multiplicity of a fixed point. One more computational form of the definition can be given as follows. Using the Weierstrass preparation theorem, the analytic function  $p(x, y)$  can be written uniquely as the product of a unit in the ring  $C\{x, y\}$  and a ‘Weierstrass polynomial’ of the form  $P(y) = x^j(y^m + a_1(x)y^{m-1} + \dots + a_m(x))$  with  $j \geq 0$ , where the coefficients  $a_k(x) \in C\{x\}$  are all divisible by  $x$ . Similarly,  $q(x, y)$  is the product of a unit and a Weierstrass



polynomial  $Q(y)$  of degree say  $n$  in  $y$ . The resultant of these two polynomials in  $y$  is a certain  $(m+n) \times (m+n)$  determinant in the coefficients. *If the system of equations  $p = q = 0$  has an isolated zero of multiplicity  $\mu$  at the origin, then this resultant is equal to  $x^\mu$  multiplied by a unit of the ring  $C\{x\}$ .* [(Fu, p. 10).]

Following are some rough lower and upper bounds for the multiplicity  $\mu$ . Let us write the Taylor expansions of  $p$  and  $q$  as

$$p = p_m + p_{m+1} + \cdots, \quad q = q_n + q_{n+1} + \cdots$$

with  $p_m \neq 0$  and  $q_n \neq 0$ , where the polynomials  $p_k = p_k(x, y)$  and  $q_k = q_k(x, y)$  are homogeneous of degree  $k$ . Then according to [C] we have  $\mu \geq mn$ . Furthermore, if the origin is an isolated zero of the system of equations  $p_m = q_n = 0$ , then  $\mu = mn$ . If  $m, n > 1$ , then it follows easily that every iterate of the map

$$f(x, y) = (x + p(x, y), y + q(x, y))$$

has this same fixed point multiplicity  $\mu = mn$  at the origin.

An inequality due to Hörmander and Lojasiewicz states that the origin is an isolated zero of the equations  $p(x, y) = q(x, y) = 0$  if and only if there exist positive constants  $K$  and  $d$  so that

$$|p|^2 + |q|^2 \geq K(|x|^2 + |y|^2)^d \quad (8.7)$$

whenever the pair  $(x, y)$  is sufficiently close to the origin. Here we may assume that  $d$  is a positive integer. Thus for actual computation of the multiplicity we can replace  $p$  and  $q$  by polynomial maps of degree at most  $d$ , simply by removing all terms of degree greater than  $d$  in the Taylor expansions of  $p$  and  $q$ . Also, we can perturb the coefficients of the terms of degree  $d$  and still preserve the inequality (8.7). Thus we can suppose that  $p$  and  $q$  satisfy the conditions of the 'non-linear alternative', as described in § 3. It then follows easily that  $\mu \leq d^2$ . As an example, if  $f(x, y) = (x + x^d, y + y^d)$ , then evidently this inequality is sharp. All of these inequalities extend easily to the higher dimensional case.

#### REFERENCES

- [BCW] H. Bass, E. H. Connell & D. Wright. The Jacobian Conjecture: reduction of degree and formal expansion of the inverse. *Bull. Amer. Math. Soc.* **7** (1982), 287-330.
- [C] J. Cronin. Analytic functional mappings. *Ann. Math.* **58** (1953), 175-181.
- [D] R. Devaney. *An Introduction to Chaotic Dynamical Systems*. Benjamin: New York, 1986.
- [DN] R. Devaney & Z. Nitecki. Shift automorphisms in the Hénon mapping. *Comm. Math. Phys.* **67** (1979), 137-146.
- [DE] P. G. Dixon & J. Esterle. Michael's problem and the Poincaré-Fatou-Bieberbach phenomenon. *Bull. Amer. Math. Soc.* **15** (1986), 127-187.
- [E] W. Engel. Ganze Cremona-Transformationen von Primzahlgrad in der Ebene. *Math.-Ann.* **136** (1958), 319-325.
- [F] S. Friedland. Inverse eigenvalue problems. *Linear Alg. Appl.* **17** (1977), 15-51.
- [Fu] W. Fulton. *Intersection Theory*. Springer-Verlag: New York, 1984.
- [GH] John Guckenheimer & Philip Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields*. Springer, Berlin, 1983.
- [G] A. Gutwirth (Evyatar). An inequality for certain pencils of plane curves. *Proc. Amer. Math. Soc.* **12** (1961), 631-638.
- [H1] M. Hénon. Numerical study of quadratic area preserving mappings. *Q. Appl. Math.* **27** (1969), 291-312.



- [H2] M. Hénon. A two-dimensional mapping with a strange attractor. *Comm. Math. Phys.* **50** (1976), 69–77.
- [Hö] L. Hörmander. On the division of distributions by polynomials. *Ark. Mat.* **3** (1958), 555–568.
- [Hu] J. H. Hubbard & R. Oberste-Vorth. Hénon mappings in the complex domain; in preparation.
- [J] H. W. E. Jung. Über ganze birationale transformationen der Ebene. *J. Reine Angew. Math.* **184** (1942), 161–174.
- [K] T. Kambayashi. On the absence of nontrivial separable forms of the affine plane. *J. Algebra* **35** (1975), 449–456.
- [Ka] A. Katok. Lyapunov exponents, entropy and periodic orbits of diffeomorphisms. *Pub. Math. IHES* **51** (1980), 137–173.
- [KM] A. Katok & L. Mendoza. *Smooth Ergodic Theory*; to appear.
- [L] S. Lojasewicz. Sur le problème de la division. *Studia Math.* **18** (1959), 87–136.
- [McK1] J. H. McKay & S. S. Wang. An inversion formula for two polynomials in two variables. *J. Pure Appl. Algebra* **40** (1985), 245–257.
- [McK2] J. H. McKay & S. S. Wang. An elementary proof of the automorphism theorem for the polynomial ring in two variables; to appear.
- [Mi1] J. Milnor. *Singular Points of Complex Hypersurfaces*. Ann. Math. Stud. **61** Princeton U. Press 1968.
- [Mi2] J. Milnor. Non-expansive Hénon maps. *Adv. Math.* **69** (1988), 109–114.
- [O] R. Oberste-Vorth. Complex horseshoes and the dynamics of mappings of two complex variables; in preparation.
- [N] S. Newhouse. Continuity properties of entropy. To appear.
- [R] L. Reich. Normalformen biholomorpher Abbildungen mit anziehendem Fixpunkt. *Math. Ann.* **180** (1969), 233–255.
- [S] I. R. Shafarevich. On some infinite dimensional groups. *Rendiconti Mat. e Applic. (Roma)* **25** (1966), 208–212.
- [Sh] M. Shub. *Global Stability of Dynamical Systems*. Springer, Berlin, 1987.
- [W] B. L. van der Waerden. Die Alternative bei nichtlinearen Gleichungen. *Nachr. Gesells. Wiss. Göttingen* 1928, Math. Phys. Klasse 77–87.
- [Wa] P. Walters. *An Introduction to Ergodic Theory*. Springer, Berlin, 1982.
- [Wr] D. Wright. The amalgamated free product structure of  $GL_2(k[x_1, \dots, x_n])$  and the weak Jacobian theorem for two variables. *J. Pure Appl. Algebra* **12** (1978), 235–251.
- [Y] Y. Yomdin. Volume growth and entropy. *Israel J. Math.* **57** (1987), 285–300.