Pacific Journal of Mathematics

COMPOSITIONS OF BIRATIONAL ENDOMORPHISMS OF THE AFFINE PLANE

PIERRETTE CASSOU-NOGUÈS AND DANIEL DAIGLE

Volume 272 No. 2 December 2014

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.272.353

COMPOSITIONS OF BIRATIONAL ENDOMORPHISMS OF THE AFFINE PLANE

PIERRETTE CASSOU-NOGUÈS AND DANIEL DAIGLE

Besides contributing several new results in the general theory of birational endomorphisms of $\mathbb{A}^2$, this article describes certain classes of birational endomorphisms $f : \mathbb{A}^2 \to \mathbb{A}^2$ defined by requiring that the missing curves or contracting curves of f are lines. The last part of the article is concerned with the monoid structure of the set of birational endomorphisms of $\mathbb{A}^2$.

Let $\boldsymbol{k}$ be an algebraically closed field and let $\mathbb{A}^2 = \mathbb{A}^2_{\boldsymbol{k}}$ be the affine plane over $\boldsymbol{k}$. A *birational endomorphism of* $\mathbb{A}^2$ is a morphism of algebraic varieties, $f : \mathbb{A}^2 \to \mathbb{A}^2$, which restricts to an isomorphism $U \to V$ for some nonempty Zariski-open subsets U and V of $\mathbb{A}^2$. The set $\mathrm{Bir}(\mathbb{A}^2)$ of birational endomorphisms of $\mathbb{A}^2$ is a monoid under composition of morphisms, and the group of invertible elements of this monoid is the automorphism group $\mathrm{Aut}(\mathbb{A}^2)$. An element f of $\mathrm{Bir}(\mathbb{A}^2)$ is *irreducible* if it is not invertible and if, for every factorization $f = h \circ g$ with $g, h \in \mathrm{Bir}(\mathbb{A}^2)$, one of g, h is invertible. Elements $f, g \in \mathrm{Bir}(\mathbb{A}^2)$ are *equivalent* (denoted $f \sim g$) if there exist $u, v \in \mathrm{Aut}(\mathbb{A}^2)$ satisfying $u \circ f \circ v = g$. The elements of $\mathrm{Bir}(\mathbb{A}^2)$ which are equivalent to the birational morphism $c : \mathbb{A}^2 \to \mathbb{A}^2$, $c(x, y) = (x, xy)$ are called *simple affine contractions* (SAC) and are the simplest examples of noninvertible elements of $\mathrm{Bir}(\mathbb{A}^2)$. It was once an open question whether $\mathrm{Aut}(\mathbb{A}^2) \cup \{c\}$ generated $\mathrm{Bir}(\mathbb{A}^2)$ as a monoid (the question arose in Abhyankar's seminar at Purdue in the early 1970s); P. Russell showed that the answer was negative by giving an example of an irreducible element of $\mathrm{Bir}(\mathbb{A}^2)$ which is not a SAC (this example appeared later in [Daigle 1991a, 4.7]). This was the first indication that $\mathrm{Bir}(\mathbb{A}^2)$ could be a complicated object.

The papers [Daigle 1991a; 1991b] seem to be the first publications that study birational endomorphisms of $\mathbb{A}^2$ in a systematic way (these are our main references). We know of two more contributions to the subject: a certain family of elements of $\mathrm{Bir}(\mathbb{A}^2)$ is described explicitly in [Cassou-Noguès and Russell 2007], and [Shpilrain and Yu 2004] gives an algorithm for deciding whether a given element of $\mathrm{Bir}(\mathbb{A}^2)$ is in the submonoid generated by $\mathrm{Aut}(\mathbb{A}^2) \cup \{c\}$.

The research of Cassou-Noguès was partially supported by Spanish grants MTM2010-21740-C02-01 and MTM2010-21740-C02-02. The research of Daigle was supported by grant RGPIN/104976-2010 from NSERC Canada.

MSC2010: 14R10, 14H50.

Keywords: affine plane, birational morphism, plane curve.

The list of references is much longer if we include another aspect of the problem. Indeed, there is a long history of studying polynomials $F \in \boldsymbol{k}[X, Y]$ which appear as components of birational endomorphisms of $\mathbb{A}^2$. To explain this, we recall some definitions. A polynomial $F \in \boldsymbol{k}[X, Y]$ is called a *field generator* if there exists $G \in \boldsymbol{k}(X, Y)$ satisfying $\boldsymbol{k}(F, G) = \boldsymbol{k}(X, Y)$; in the special case where G can be chosen in $\boldsymbol{k}[X, Y]$, one says that F is a *good* field generator. So a good field generator is just the same thing as a component of a birational endomorphism of $\mathbb{A}^2$. By a *generally rational polynomial*[1] we mean an $F \in \boldsymbol{k}[X, Y]$ such that, for almost all $\lambda \in \boldsymbol{k}$, $F - \lambda$ is an irreducible polynomial whose zero set in $\mathbb{A}^2$ is a rational curve (where "almost all" means "all but possibly finitely many"). If $\operatorname{char} \boldsymbol{k} = 0$, then field generators and generally rational polynomials are one and the same thing (this is noted in [Miyanishi and Sugie 1980]; see [Daigle 2013] for the positive characteristic case). The study of these polynomials is a classical subject, as is clear from considering the following (incomplete) list of references: [Nishino 1968; 1969; 1970; Saitō 1972; 1977; Jan 1974; Russell 1975; 1977; Miyanishi and Sugie 1980; Kaliman 1992; Artal Bartolo and Cassou-Noguès 1996; Neumann and Norbury 2002; Cassou-Noguès 2005; Sasao 2006; Daigle 2013; Cassou-Noguès and Daigle 2013].

This paper is a contribution to the theory of birational endomorphisms of $\mathbb{A}^2$. Our methods are those of [Daigle 1991a; 1991b], and we place ourselves at the same level of generality as in those papers: the base field $\boldsymbol{k}$ is algebraically closed but otherwise arbitrary.

In [Daigle 1991a; 1991b] and [Cassou-Noguès and Russell 2007], there is a tendency to restrict attention to irreducible elements of $\operatorname{Bir}(\mathbb{A}^2)$. Going in an orthogonal direction, the present paper is mainly concerned with *compositions* of birational endomorphisms. This choice is motivated by many reasons. First, the examples given in the three papers above show that $\operatorname{Bir}(\mathbb{A}^2)$ contains a great diversity of irreducible elements of arbitrarily high complexity; since the task of finding all irreducible elements is probably hopeless, it seems to us that finding more examples of them might be less relevant than, say, trying to understand the monoid structure of $\operatorname{Bir}(\mathbb{A}^2)$. Also, a significant portion of this paper is geared towards proving Theorem 3.15, which we need in the forthcoming [Cassou-Noguès and Daigle $\geq$ 2014] to prove the following fact: *Let $\boldsymbol{k}$ be an arbitrary field and $A_0 \supset A_1 \supset \cdots$ an infinite, strictly descending sequence of rings such that* (i) *each A_i is a polynomial ring in* 2 *variables over $\boldsymbol{k}$*; (ii) *all A_i have the same field of fractions*; *and* (iii) *the ring $R = \bigcap_i A_i$ is not equal to $\boldsymbol{k}$. Then $R = \boldsymbol{k}[F]$ for some F, where F is a good field generator of A_0 and a variable of A_i for $i \gg 0$.* Moreover, if

[1] Generally rational polynomials are sometimes called "rational polynomials" or "generically rational polynomials".

one is interested in field generators and generally rational polynomials, one should not restrict one's attention to irreducible endomorphisms. In this respect we point out that the components of the morphisms described by Theorem 3.15 are precisely the "rational polynomials of simple type" listed in [Neumann and Norbury 2002].

The paper is organized as follows.

Section 1 contains some preliminary observations on "admissible" and "weakly admissible" configurations of curves in $\mathbb{A}^2$.

Section 2 gives several new results in the general theory of birational endomorphisms of $\mathbb{A}^2$ (in particular Proposition 2.9, Proposition 2.13, Corollary 2.14 and Lemma 2.17, but also several useful lemmas).

Given $f \in \mathrm{Bir}(\mathbb{A}^2)$, let $\mathrm{Miss}(f)$ (resp. $\mathrm{Cont}(f)$) be the set of missing curves (resp. contracting curves) of f; see 2.2 for definitions. Section 3 studies those $f \in \mathrm{Bir}(\mathbb{A}^2)$ satisfying the condition that $\mathrm{Miss}(f)$ is a weakly admissible configuration or the stronger condition that $\mathrm{Miss}(f)$ is an admissible configuration or the even stronger condition that both $\mathrm{Miss}(f)$ and $\mathrm{Cont}(f)$ are admissible configurations. The main result of Section 3, Theorem 3.15, gives a complete description of these three subsets of $\mathrm{Bir}(\mathbb{A}^2)$.

While Sections 2 and 3 are mainly concerned with individual endomorphisms, Section 4 considers the monoid structure of $\mathrm{Bir}(\mathbb{A}^2)$. The first part of that section shows, in particular, that if S is a subset of $\mathrm{Bir}(\mathbb{A}^2)$ such that $\mathrm{Aut}(\mathbb{A}^2) \cup S$ generates $\mathrm{Bir}(\mathbb{A}^2)$ as a monoid, then $\{\deg f \mid f \in S\}$ is not bounded and $|S| = |\boldsymbol{k}|$ (giving a very strong negative answer to the already mentioned question of Abhyankar). The second part shows that the submonoid $\mathcal{A}$ of $\mathrm{Bir}(\mathbb{A}^2)$ generated by SACs and automorphisms is "factorially closed" in $\mathrm{Bir}(\mathbb{A}^2)$; i.e., if $f, g \in \mathrm{Bir}(\mathbb{A}^2)$ are such that $g \circ f \in \mathcal{A}$, then $f, g \in \mathcal{A}$.

Conventions. All algebraic varieties (in particular all curves and surfaces) are irreducible and reduced. All varieties and morphisms are over an algebraically closed field $\boldsymbol{k}$ of arbitrary characteristic ($\boldsymbol{k}$ is assumed to be algebraically closed from Definition 1.3 until the end of the paper). The word "point" means "closed point", unless otherwise specified.

All rings are commutative and have a unity. The symbol A^* denotes the set of units of a ring A. If A is a subring of a ring B and $n \in \mathbb{N}$, the notation $B = A^{[n]}$ means that B is isomorphic (as an A-algebra) to the polynomial ring in n variables over A. We adopt the conventions that $0 \in \mathbb{N}$, that "$\subset$" means strict inclusion and that "$\setminus$" denotes set difference.

1. Admissible configurations of curves in $\mathbb{A}^2$

Recall the following terminology. Let $\boldsymbol{k}$ be a field, $A = \boldsymbol{k}^{[2]}$, and $\mathbb{A}^2_{\boldsymbol{k}} = \mathrm{Spec}\, A$. We abbreviate $\mathbb{A}^2_{\boldsymbol{k}}$ to $\mathbb{A}^2$. By a *coordinate system of* A, we mean an ordered pair (F, G)

of elements of A satisfying $A = \boldsymbol{k}[F, G]$. A *variable* of A is an element $F \in A$ for which there exists G satisfying $\boldsymbol{k}[F, G] = A$.

Let $F \in A$ be an irreducible element and let $C \subset \mathbb{A}^2$ be the zero set of F (i.e., the set of prime ideals $\mathfrak{p} \in \operatorname{Spec} A = \mathbb{A}^2$ satisfying $F \in \mathfrak{p}$); we call C a *line* if $A/FA = \boldsymbol{k}^{[1]}$ and a *coordinate line* if F is a variable of A. Note that C is a line if and only if $C \cong \mathbb{A}^1$; given a coordinate system (X, Y) of A, C is a coordinate line if and only if some automorphism of $\mathbb{A}^2$ maps C onto the zero set of X. It is clear that all coordinate lines are lines, and the epimorphism theorem [Abhyankar and Moh 1975; Suzuki 1974] states that the converse is true if $\operatorname{char} \boldsymbol{k} = 0$. It is known that not all lines are coordinate lines if $\operatorname{char} \boldsymbol{k} \neq 0$ (on the subject of lines which are not coordinate lines, see, e.g., [Ganong 2011] for a recent survey). Coordinate lines are sometimes called rectifiable lines.

By a *coordinate system of* $\mathbb{A}^2 = \operatorname{Spec} A$, we mean a coordinate system of A. That is, a coordinate system of $\mathbb{A}^2$ is a pair $(X, Y) \in A \times A$ satisfying $A = \boldsymbol{k}[X, Y]$.

We adopt the viewpoint that A (or $\mathbb{A}^2$) does not come equipped with a preferred coordinate system, i.e., no coordinate system is better than the others. This may be confusing to some readers, especially those who like to identify $\mathbb{A}^2$ with $\boldsymbol{k}^2$, because any such identification inevitably depends on the choice of a coordinate system. So perhaps the following trivial remarks deserve to be made:

1.1. Let $\mathfrak{C}$ denote the set of coordinate systems of $\mathbb{A}^2$ (or A).

(a) Given $\mathfrak{c} = (X, Y) \in \mathfrak{C}$ and an element $F \in A$, one can consider the map $\boldsymbol{k}^2 \to \boldsymbol{k}$, $(x, y) \mapsto F(x, y)$, defined by first writing $F = \sum_{i,j} a_{ij} X^i Y^j$ with $a_{ij} \in \boldsymbol{k}$ (recall that $A = \boldsymbol{k}[X, Y]$) and then setting $F(x, y) = \sum_{i,j} a_{ij} x^i y^j$ for $(x, y) \in \boldsymbol{k}^2$. One can then consider the zero set $\boldsymbol{Z}(F) \subseteq \boldsymbol{k}^2$ of that map F. We stress that the map $(x, y) \mapsto F(x, y)$ and the set $\boldsymbol{Z}(F)$ depend on *both* F *and* $\mathfrak{c}$; we should write $F_{\mathfrak{c}}(x, y)$ and $\boldsymbol{Z}_{\mathfrak{c}}(F)$, but we omit the $\mathfrak{c}$.

(b) Let $P, Q \in \boldsymbol{k}[T_1, T_2]$, where $P = \sum_{i,j} a_{ij} T_1^i T_2^j$ and $Q = \sum_{i,j} b_{ij} T_1^i T_2^j$ with $a_{ij}, b_{ij} \in \boldsymbol{k}$.

  (i) The pair (P, Q) determines the map $\boldsymbol{k}^2 \to \boldsymbol{k}^2$, $(x, y) \mapsto (P(x, y), Q(x, y))$, where we define $P(x, y) = \sum_{i,j} a_{ij} x^i y^j$ and $Q(x, y) = \sum_{i,j} b_{ij} x^i y^j$.

  (ii) Choose $\mathfrak{c} = (X, Y) \in \mathfrak{C}$. Then $(P, Q, \mathfrak{c})$ determines the morphism of schemes $f : \mathbb{A}^2 \to \mathbb{A}^2$ defined by stipulating that f corresponds to the $\boldsymbol{k}$-homomorphism $A \to A$ which maps X to $P(X, Y) = \sum_{i,j} a_{ij} X^i Y^j$ and Y to $Q(X, Y) = \sum_{i,j} b_{ij} X^i Y^j$ (where $P(X, Y), Q(X, Y) \in A = \boldsymbol{k}[X, Y]$).

(c) Suppose that $\mathfrak{c} = (X, Y) \in \mathfrak{C}$ has been chosen. Then it is convenient to define morphisms of schemes $\mathbb{A}^2 \to \mathbb{A}^2$ simply by giving the corresponding polynomial maps $\boldsymbol{k}^2 \to \boldsymbol{k}^2$ (this will be done repeatedly in Section 3). To do

so, we abuse language as follows: given $P, Q \in \boldsymbol{k}[T_1, T_2]$, the sentence

$$\text{“}f : \mathbb{A}^2 \to \mathbb{A}^2 \textit{ is defined by } f(x, y) = (P(x, y), Q(x, y))\text{”}$$

means that $f : \mathbb{A}^2 \to \mathbb{A}^2$ is the morphism of schemes determined by $(P, Q, \mathfrak{c})$ as in remark (b). For instance one can define $f : \mathbb{A}^2 \to \mathbb{A}^2$ by $f(x, y) = (x, xy)$, always keeping in mind that this f depends on $\mathfrak{c}$.

1.2. Lemma. *Let $F, G \in A = \boldsymbol{k}^{[2]}$, where $\boldsymbol{k}$ is any field, and suppose that each of F, G is a variable of A. Consider the ideal (F, G) of A generated by F and G.*

(a) *If $(F, G) = A$, then $G = aF + b$ for some $a, b \in \boldsymbol{k}^*$.*

(b) *If $A/(F, G) = \boldsymbol{k}$, then $A = \boldsymbol{k}[F, G]$.*

Proof. Choose Y such that $A = \boldsymbol{k}[F, Y]$ and define $X = F$. Then $A = \boldsymbol{k}[X, Y]$ and we may write G as a polynomial in X, Y, say $G = P(X, Y)$.

Suppose that $(F, G) = A$. Then $1 \in (F, G) = (X, P(X, Y)) = (X, P(0, Y))$ implies $P(0, Y) \in \boldsymbol{k}^*$. Writing $P(0, Y) = b \in \boldsymbol{k}^*$, we obtain that $G - b = P(X, Y) - P(0, Y)$ is divisible by X; as $G - b$ is irreducible, $G - b = aX = aF$ with $a \in \boldsymbol{k}^*$, and (a) is proved.

To prove (b), we first observe that since $G = P(X, Y)$ is a variable of $\boldsymbol{k}[X, Y]$ and $P(X, Y) \notin \boldsymbol{k}[X]$, we have that P is “almost monic” in Y, i.e.,

$$P(X, Y) = aY^d + p_1(X)Y^{d-1} + \cdots + p_d(X) \tag{1}$$

with $d \geq 1$, $a \in \boldsymbol{k}^*$ and $p_i(X) \in \boldsymbol{k}[X]$ for $i = 1, \ldots, d$. Now

$$\boldsymbol{k} = A/(F, G) = \boldsymbol{k}[X, Y]/(X, P(X, Y)) = \boldsymbol{k}[X, Y]/(X, P(0, Y))$$

implies that $\deg P(0, Y) = 1$, so $d = 1$ in (1). Then $G = aY + p_1(X)$ and $\boldsymbol{k}[F, G] = \boldsymbol{k}[X, aY + p_1(X)] = \boldsymbol{k}[X, Y] = A$. □

Until the end of this paper, we assume that $\boldsymbol{k}$ is an algebraically closed field of arbitrary characteristic. Consider $\mathbb{A}^2 = \mathbb{A}^2_{\boldsymbol{k}}$.

1.3. Definition. Let $C_1, \ldots, C_n$ $(n \geq 0)$ be distinct curves in $\mathbb{A}^2$, and consider the set $S = \{C_1, \ldots, C_n\}$. We say that S is a *weakly admissible configuration* if

(a) each C_i is a coordinate line;

(b) for every choice of $i \neq j$ such that $C_i \cap C_j$ is nonempty, $C_i \cap C_j$ is one point and the local intersection number of C_i and C_j at that point is equal to 1.

1.4. Lemma. *Given distinct curves $C_1, \ldots, C_n$ $(n \geq 0)$ in $\mathbb{A}^2$, the following are equivalent*:

(a) *$\{C_1, \ldots, C_n\}$ is a weakly admissible configuration.*

(b) *There exists a coordinate system of $\mathbb{A}^2$ with respect to which all C_i have degree* 1.

Proof. We show that (a) implies (b), the converse being trivial. Suppose that (a) holds. Write $\mathbb{A}^2 = \operatorname{Spec} A$, where $A = \boldsymbol{k}^{[2]}$. We may assume that $n \geq 2$, otherwise the assertion is trivial. Let $F_1, \ldots, F_n \in A$ be variables of A whose zero sets are $C_1, \ldots, C_n$ respectively. Condition (a) implies that, whenever $i \neq j$, we have either $(F_i, F_j) = A$ or $A/(F_i, F_j) = \boldsymbol{k}$. Consider the graph G whose vertex set is $\{F_1, \ldots, F_n\}$ and in which distinct vertices F_i, F_j are joined by an edge if and only if $A/(F_i, F_j) = \boldsymbol{k}$.

In the case where G is discrete, Lemma 1.2 implies that $F_i = a_i F_1 + b_i$ with $a_i, b_i \in \boldsymbol{k}^*$ for $i = 2, \ldots, n$. Let $X = F_1$ and let Y be such that $A = \boldsymbol{k}[X, Y]$. Then all F_i have degree 1 with respect to the coordinate system (X, Y).

From now on, assume that G is not discrete. Then we may assume that F_1, F_2 are joined by an edge. By Lemma 1.2, $\boldsymbol{k}[F_1, F_2] = A$. Let $X_1 = F_1$ and $X_2 = F_2$; then $A = \boldsymbol{k}[X_1, X_2]$ and for each $i \in \{3, \ldots, n\}$ we have:

- If F_i, F_1 are not joined by an edge, then Lemma 1.2 implies that $F_i = a_i X_1 + b_i$ for some $a_i, b_i \in \boldsymbol{k}^*$, so F_i has degree 1 with respect to (X_1, X_2).
- If F_i, F_2 are not joined by an edge, then $F_i = a_i X_2 + b_i$ for some $a_i, b_i \in \boldsymbol{k}^*$, so F_i has degree 1 with respect to (X_1, X_2).
- If F_i is linked to each of F_1, F_2 by edges, then $\boldsymbol{k}[F_i, F_1] = A = \boldsymbol{k}[F_i, F_2]$, so $F_2 = aF_1 + \beta(F_i)$ for some $a \in \boldsymbol{k}^*$ and $\beta(T) \in \boldsymbol{k}[T]$; then $\beta(F_i) = X_2 - aX_1$ has degree 1 with respect to (X_1, X_2) and consequently F_i has degree 1.

So all F_i have degree 1 with respect to the coordinate system (X_1, X_2). □

1.5. Let $C_1, \ldots, C_n$ $(n \geq 0)$ be distinct curves in a nonsingular surface W. We say that $\sum_{i=1}^n C_i$ is an *SNC-divisor* of W if

- each C_i is a nonsingular curve;
- for every choice of $i \neq j$ such that $C_i \cap C_j$ is nonempty, $C_i \cap C_j$ is one point and the local intersection number of C_i and C_j at that point is equal to 1;
- if i, j, k are distinct then $C_i \cap C_j \cap C_k = \varnothing$.

If $D = \sum_{i=1}^n C_i$ is an SNC-divisor of W, we write $\mathcal{G}(D, W)$ for the graph whose vertex set is $\{C_1, \ldots, C_n\}$ and in which distinct vertices C_i, C_j are joined by an edge if and only if $C_i \cap C_j \neq \varnothing$.

1.6. Definition. Let $C_1, \ldots, C_n$ $(n \geq 0)$ be distinct curves in $\mathbb{A}^2$. We say that the set $\{C_1, \ldots, C_n\}$ is an *admissible configuration* if

(a) each C_i is a coordinate line;

(b) $D = \sum_{i=1}^n C_i$ is an SNC-divisor of $\mathbb{A}^2$;

(c) the graph $\mathcal{G}(D, \mathbb{A}^2)$ defined in 1.5 is a forest.

1.7. Proposition. *Given distinct curves $C_1, \dots, C_n$ $(n \geq 0)$ in $\mathbb{A}^2$, the following are equivalent*:

(a) $\{C_1, \dots, C_n\}$ *is an admissible configuration.*

(b) *There exists a coordinate system (X, Y) of $\mathbb{A}^2$ such that $\bigcup_{i=1}^n C_i$ is the zero set of $\varphi(X)Y^j$ for some $j \in \{0, 1\}$ and some $\varphi(X) \in \boldsymbol{k}[X] \setminus \{0\}$.*

Proof. It's enough to show that (a) implies (b), as the converse is trivial. Assume that (a) holds. By Lemma 1.4, we may choose a coordinate system which respect to which all C_i have degree 1. As $D = \sum_{i=1}^n C_i$ is an SNC-divisor and $\mathcal{G}(D, \mathbb{A}^2)$ is a forest, $\bigcup_{i=1}^n C_i$ must be either

- a union of n parallel lines, or
- a union of $n-1$ parallel lines with another line.

Indeed, any other configuration of lines would either contain three concurrent lines or produce a circuit in the graph. Now it is clear that (b) is satisfied. □

2. Birational morphisms $f : X \to Y$ of nonsingular surfaces with special emphasis on the case $X = Y = \mathbb{A}^2$

Throughout this section, $\boldsymbol{k}$ is an algebraically closed field of arbitrary characteristic and we abbreviate $\mathbb{A}^2_{\boldsymbol{k}}$ to $\mathbb{A}^2$. We consider birational morphisms $f : X \to Y$, where X and Y are nonsingular algebraic surfaces over $\boldsymbol{k}$ (a morphism $f : X \to Y$ is *birational* if there exist nonempty Zariski-open subsets $U \subseteq X$ and $V \subseteq Y$ such that f restricts to an isomorphism $U \to V$). We are particularly interested in the case $X = \mathbb{A}^2 = Y$.

Essentially all the material given in 2.1–2.8 can be found in [Daigle 1991a]. From 2.9 to the end of the section, the material appears to be new (except 2.13(a)).

2.1. Let $f : X \to Y$ and $f' : X' \to Y'$ be birational morphisms of nonsingular surfaces. We say that f, f' are *equivalent* ($f \sim f'$) if there exist isomorphisms $x : X \to X'$ and $y : Y \to Y'$ such that $y \circ f = f' \circ x$.

2.2 [Daigle 1991a, 1.2]. Let $f : X \to Y$ be a birational morphism of nonsingular surfaces. A *missing curve* of f is a curve $C \subset Y$ whose intersection with the image of f is a finite set of points. We write $\mathrm{Miss}(f)$ for the set of missing curves of f, $q(f)$ for the cardinality of $\mathrm{Miss}(f)$ and $q_0(f)$ for the number of missing curves of f which are disjoint from $f(X)$. A *contracting curve* of f is a curve $C \subset X$ such that $f(C)$ is a point. The set of contracting curves of f is denoted $\mathrm{Cont}(f)$, and $c(f)$ denotes the cardinality of $\mathrm{Cont}(f)$. The natural numbers $q(f)$, $q_0(f)$ and $c(f)$ are invariant with respect to equivalence of birational morphisms (2.1), i.e., $f \sim f' \Longrightarrow c(f) = c(f')$ and similarly for q and q_0. Call a point of Y a *fundamental point* of f if it is $f(C)$ for some contracting curve C of f. The set of fundamental points of f is also called the *center* of f, denoted $\mathrm{cent}(f)$. The

exceptional locus of f is defined to be $\mathrm{exc}(f) = f^{-1}(\mathrm{cent}(f))$ and is equal to the union of the contracting curves of f.

2.3 [Daigle 1991a, 1.1 and 1.2]. Let $f : X \to Y$ be a birational morphism of nonsingular surfaces. There exists a commutative diagram

$$\begin{array}{ccccc} Y_n & \xrightarrow{\pi_n} & \cdots & \xrightarrow{\pi_1} & Y_0 \\ \uparrow & & & & \| \\ X & & \xrightarrow[f]{} & & Y \end{array} \tag{2}$$

where "$\hookrightarrow$" denotes an open immersion and, for each i, $\pi_i : Y_i \to Y_{i-1}$ is the blowing-up of the nonsingular surface Y_{i-1} at a point $P_i \in Y_{i-1}$.

Let $n(f)$ be the least natural number n for which there exists a diagram (2). Note that $n(f)$ is invariant with respect to equivalence of birational morphisms.

For each $i \in \{1, \ldots, n\}$, consider the exceptional curve $E_i = \pi_i^{-1}(P_i) \subset Y_i$, and let the same symbol E_i also denote the strict transform of E_i in Y_n. It is clear that the union of the contracting curves of f is the intersection of $E_1 \cup \cdots \cup E_n \subset Y_n$ with the open subset X of Y_n; thus

$$c(f) \leq n(f); \tag{3}$$

(4) *each contracting curve is nonsingular and rational,* $D = \sum_{C \in \mathrm{Cont}(f)} C$ *is an SNC-divisor of* X *and the graph* $\mathcal{G}(D, X)$ *is a forest.*[2]

Given $i \in \{1, \ldots, n\}$, consider the complete curve $E_i \subset Y_n$. Note that if S is a projective nonsingular surface and $\mu : Y_n \hookrightarrow S$ is an open immersion, the self-intersection number of $\mu(E_i)$ in S is independent of the choice of (S, μ); we denote this number by $(E_i^2)_{Y_n}$. Then the following holds by [ibid., 1.2(c)]:

(5) *Diagram* (2) *satisfies* $n = n(f)$ *if and only if* $(E_i^2)_{Y_n} \leq -2$ *holds for all* $i \in \{1, \ldots, n\}$ *such that* $E_i \subseteq Y_n \setminus X$.

In particular, if Diagram (2) satisfies $n = n(f)$, then

$$\mathrm{cent}(f) = \{(\pi_1 \circ \cdots \circ \pi_{i-1})(P_i) \mid 1 \leq i \leq n\}. \tag{6}$$

The following remarks are trivial consequences of 2.3, but are very useful:

2.4. Remarks. Let $f : X \to Y$ be a birational morphism of nonsingular surfaces.

(a) For each $C \in \mathrm{Miss}(f)$, we have $C \cap f(X) \subseteq \mathrm{cent}(f)$. In particular, the condition $q_0(f) = 0$ is equivalent to "every missing curve contains a fundamental point".

(b) Let $C \subset Y$ be a curve. Then there exists at most one curve $C' \subset X$ such that $f(C')$ is a dense subset of C. Moreover, C' exists if and only if $C \notin \mathrm{Miss}(f)$.

[2] Note that contracting curves are not necessarily isomorphic to $\mathbb{A}^1$. So, in the case $X = \mathbb{A}^2 = Y$, $\mathrm{Cont}(f)$ is not necessarily an admissible configuration in the sense of Definition 1.6.

2.5. Lemma. *If* $\mathbb{A}^2 \xrightarrow{f} \mathbb{A}^2 \xrightarrow{g} \mathbb{A}^2$ *are birational morphisms then*

$$n(g \circ f) = n(g) + n(f).$$

Proof. Follows from [ibid., 2.12]. □

2.6. Lemma. *Let* $f : \mathbb{A}^2 \to \mathbb{A}^2$ *be a birational morphism.*

(a) $q(f) = c(f) \leq n(f)$.

(b) f *is an isomorphism* $\iff n(f) = 0 \iff c(f) = 0 \iff q(f) = 0$.

(c) *Each missing curve of* f *is rational with one place at infinity.*

(d) *Each fundamental point belongs to some missing curve*; *each missing curve contains some fundamental point.*

(e) *If a point* P *is a singular point of some missing curve of* f *or a common point of two missing curves, then* P *is a fundamental point of* f.

Proof. The equality $q(f) = c(f)$ in (a) follows from [ibid., 4.3(a)], and $c(f) \leq n(f)$ was noted in (3). Assertion (b) follows from the observation that if $n(f) = 0$ or $c(f) = 0$ then f is an open immersion $\mathbb{A}^2 \hookrightarrow \mathbb{A}^2$ and hence an automorphism. Assertion (c) follows from [ibid., 4.3(c)]. The first (resp. the second) assertion of (d) follows from [ibid., 2.1] (resp. from the claim that $q_0(f) = 0$, in [ibid., 4.3(a)]). Refer to [Daigle 2010, 1.6] for a proof of assertion (e). □

Several of the above facts are stated in greater generality in [ibid.]. For instance, if $X \xrightarrow{f} Y \xrightarrow{g} Z$ are birational morphisms of nonsingular surfaces and $q_0(f) = 0$, then (by [ibid., 1.3]) $n(g \circ f) = n(g) + n(f)$. Also, if X, Y are nonsingular surfaces satisfying $\Gamma(X, \mathcal{O}_X)^* = \boldsymbol{k}^*$ and $\mathrm{Cl}(Y) = 0$, then (by [ibid., 2.11]) every birational morphism $f : X \to Y$ satisfies $q_0(f) = 0$. The following generalization of Lemma 2.6(a) also deserves to be noted:

2.7. Lemma. *Let* $f : X \to X$ *be a birational morphism, where* X *is a nonsingular surface. Then* $c(f) = q(f)$.

Proof. Follows from Remark 2.13 of [ibid.]. □

2.8. Lemma. *Let* $f : X \to Y$ *be a birational morphism of nonsingular surfaces and* Γ_f *the union of the missing curves of* f. *If* X *is affine then the following hold*:

(a) $\mathrm{cent}(f) \subseteq \Gamma_f$.

(b) $Y \setminus \Gamma_f$ *is the interior of* $f(X)$ *and* $f^{-1}(Y \setminus \Gamma_f) = X \setminus \mathrm{exc}(f)$.

(c) *The surfaces* $X \setminus \mathrm{exc}(f)$ *and* $Y \setminus \Gamma_f$ *are affine, and* f *restricts to an isomorphism* $X \setminus \mathrm{exc}(f) \to Y \setminus \Gamma_f$.

Proof. Follows from [ibid., Prop. 2.1] and its proof. □

2.9. Proposition. *Let* $f : \mathbb{A}^2 \to \mathbb{A}^2$ *be a birational morphism. If* P *is a singular point of a missing curve of* f, *then* P *belongs to at least two missing curves of* f.

Proof. By Lemma 2.6(e), P is a fundamental point of f; so it suffices to show that if a fundamental point P belongs to only one missing curve C, then the multiplicity $\mu(P, C)$ of C at P is equal to 1. So assume that P is a fundamental point which belongs to only one missing curve C. Choose a diagram (2) satisfying $n = n(f)$, and let the notation P_i, E_i, etc., be as in 2.3. In fact let us choose diagram (2) in such a way that $P_1 = P$ and, for some $s \in \{1, \dots, n\}$,

(7) $\qquad P_2, \dots, P_s$ are infinitely near P_1, but $P_{s+1}, \dots, P_n$ are not.

Let us label the missing curves as $C_1, \dots, C_q$, where

(8) $$P_1 \in C_j \iff j = 1.$$

The diagram (2) together with the ordering $(C_1, \dots, C_q)$ of the set of missing curves constitutes a "minimal decomposition" of f, in the terminology of [ibid., 1.2(h)]. This minimal decomposition $\mathcal{D}$ determines matrices $\mu_{\mathcal{D}}$, $\mathcal{E}_{\mathcal{D}}$, $\varepsilon_{\mathcal{D}}$ and $\varepsilon'_{\mathcal{D}}$, defined in [ibid., 2.8]. These are matrices with entries in $\mathbb{N}$, and the result [ibid., 4.3(b)] asserts that the product $\varepsilon'_{\mathcal{D}}\mu_{\mathcal{D}}$ is a square matrix of determinant ± 1. We shall now argue that the condition $\det(\varepsilon'_{\mathcal{D}}\mu_{\mathcal{D}}) = \pm 1$ implies that $\mu(P_1, C_1) = 1$ (this will complete the proof of the proposition). By (7), the $n \times n$ matrix $\mathcal{E}_{\mathcal{D}}$ has the shape

$$\mathcal{E}_{\mathcal{D}} = (e_{ij}) = \left(\begin{array}{c|c} \mathcal{E}_0 & 0 \\ \hline 0 & \mathcal{E}_1 \end{array}\right) \quad (\text{with } e_{ij} \in \mathbb{N} \text{ for all } i, j),$$

where $\mathcal{E}_0$ is an $s \times s$ lower-triangular matrix with zero diagonal, and where

the first row is the only zero row of $\mathcal{E}_0$.

Consider the $n \times n$ matrix $\varepsilon_{\mathcal{D}}$, determined by $\mathcal{E}_{\mathcal{D}}$ as explained in [ibid., 2.7]. The already mentioned properties of $\mathcal{E}_{\mathcal{D}}$ imply that $\varepsilon_{\mathcal{D}}$ is as follows:

$$\varepsilon_{\mathcal{D}} = (\varepsilon_{ij}) = \left(\begin{array}{c|c} \varepsilon_0 & 0 \\ \hline 0 & \varepsilon_1 \end{array}\right) \quad (\text{with } \varepsilon_{ij} \in \mathbb{N} \text{ for all } i, j),$$

where ε_0 is an $s \times s$ lower-triangular matrix with diagonal entries equal to 1 and where

all entries in the first column of ε_0 are positive.

Next, $\varepsilon'_{\mathcal{D}}$ is the submatrix of $\varepsilon_{\mathcal{D}}$ obtained by deleting the i-th row for each $i \in J$, where $J = \{\, i \mid 1 \le i \le n,\ E_i \cap X = \varnothing \text{ in } Y_n \}$ in the notation of 2.3 (J is defined in [ibid., 1.2(h)]). Let $j_0 = |J \cap \{1, \dots, s\}|$; then the $(n - |J|) \times n$ matrix $\varepsilon'_{\mathcal{D}}$ has the form

$$\varepsilon'_{\mathcal{D}} = \left(\begin{array}{c|c} \varepsilon'_0 & 0 \\ \hline 0 & \varepsilon'_1 \end{array}\right),$$

where ε'_0 is an $(s - j_0) \times s$ matrix with entries in $\mathbb{N}$ and

(9) $\qquad$ all entries in the first column of ε'_0 are positive.

Finally, consider the $n \times q$ matrix $\mu_{\mathcal{D}}$; by (8),

$$\mu_{\mathcal{D}} = \left(\begin{array}{c|c} F & 0 \\ \hline G & H \end{array}\right), \quad \text{where } F = \begin{pmatrix} \mu(P_1, C_1) \\ \vdots \\ \mu(P_s, C_1) \end{pmatrix} \text{ is } s \times 1.$$

We have

$$\varepsilon'_{\mathcal{D}}\mu_{\mathcal{D}} = \left(\begin{array}{c|c} \varepsilon'_0 & 0 \\ \hline 0 & \varepsilon'_1 \end{array}\right)\left(\begin{array}{c|c} F & 0 \\ \hline G & H \end{array}\right) = \left(\begin{array}{c|c} \varepsilon'_0 F & 0 \\ \hline \varepsilon'_1 G & \varepsilon'_1 H \end{array}\right), \tag{10}$$

where the block $\varepsilon'_0 F$ has size $(s - j_0) \times 1$. By (7), $(E_s^2)_{Y_n} = -1$; so $E_s \nsubseteq Y_n \setminus X$ by (5) and hence $s \notin J$ by definition of J. It follows that $s - j_0 \geq 1$. In view of (10), the fact that $\det(\varepsilon'_{\mathcal{D}}\mu_{\mathcal{D}}) = \pm 1$ implies that $s - j_0 = 1$ and that the unique entry of $\varepsilon'_0 F$ is ± 1. We have $\{1, \ldots, s\} \setminus J = \{s\}$, so $\varepsilon'_0 = (\varepsilon_{s1} \ldots \varepsilon_{ss})$ and $\sum_{j=1}^{s} \varepsilon_{sj}\mu(P_j, C_1) = \pm 1$. Since $\varepsilon_{sj} \in \mathbb{N}$ for all j and (by (9)) $\varepsilon_{s1} \geq 1$, we get $1 \leq \mu(P_1, C_1) \leq \sum_{j=1}^{s} \varepsilon_{sj}\mu(P_j, C_1) = \pm 1$, so $\mu(P_1, C_1) = 1$. □

2.10. Remark. Let $\varphi : X \to Y$ be a dominant morphism of nonsingular surfaces. By a *deficient curve* of φ, we mean a curve $C \subset Y$ satisfying

$$\text{for almost all points } P \in C, \quad |f^{-1}(P)| < [\boldsymbol{k}(X) : \boldsymbol{k}(Y)]_s$$

where "almost all" means "all except possibly finitely many", "$|\cdot|$" denotes cardinality, $\boldsymbol{k}(X)$ and $\boldsymbol{k}(Y)$ are the function fields of X and Y and $[\boldsymbol{k}(X) : \boldsymbol{k}(Y)]_s$ is the separable degree of the field extension $\boldsymbol{k}(X)/\boldsymbol{k}(Y)$. Note that φ has finitely many deficient curves and that if φ is birational then the deficient curves are precisely the missing curves.

Then it is interesting to note that Proposition 2.9 does not generalize to all dominant morphisms $\mathbb{A}^2 \to \mathbb{A}^2$; i.e., *it is not the case that each singular point of a deficient curve is a common point of at least two deficient curves.* This is shown by the following example, in which we assume that $\operatorname{char} \boldsymbol{k} = 0$.

Choose a coordinate system of $\mathbb{A}^2$ and define morphisms $f, g, h : \mathbb{A}^2 \to \mathbb{A}^2$ by

$$f(x, y) = (x, xy), \quad g(x, y) = (x + y(y^2 - 1), y), \quad h(x, y) = (x, y^2).$$

Note that f is a SAC and $g \in \operatorname{Aut}(\mathbb{A}^2)$. Define $\varphi = h \circ g \circ f : \mathbb{A}^2 \to \mathbb{A}^2$. Then φ has two deficient curves C_1 and C_2, where

- C_1 is "$y = 0$" (the deficient curve of h);
- C_2 is "$x^2 - y(y-1)^2 = 0$" (the image by $h \circ g$ of the missing curve of f).

Moreover, $(0, 1)$ is a singular point of C_2 which is not on C_1.

2.11. Lemma. *Let $\mathbb{A}^2 \xrightarrow{f} \mathbb{A}^2 \xrightarrow{g} \mathbb{A}^2$ be birational morphisms. Then*

$$\operatorname{cent}(g \circ f) = \operatorname{cent}(g) \cup g(\operatorname{cent}(f)).$$

In particular, every fundamental point of g is a fundamental point of $g \circ f$.

Proof. Let $X \xrightarrow{f} Y \xrightarrow{g} Z$ be birational morphisms of nonsingular surfaces. The reader may easily verify that $\mathrm{cent}(g \circ f) \subseteq \mathrm{cent}(g) \cup g(\mathrm{cent}(f))$ and $g(\mathrm{cent}(f)) \subseteq \mathrm{cent}(g \circ f)$. In order to obtain the desired equality, it remains to show that

$$\mathrm{cent}(g) \subseteq \mathrm{cent}(g \circ f). \tag{11}$$

We claim that (11) is true whenever $q_0(f) = 0$. Indeed, consider $P \in \mathrm{cent}(g)$. Then there exists a curve $C \subset Y$ such that $g(C) = \{P\}$. If $C \notin \mathrm{Miss}(f)$, then (by 2.4(b)) there exists a curve $C' \subset X$ such that $f(C')$ is a dense subset of C; in particular, $(g \circ f)(C') = \{P\}$ and hence $P \in \mathrm{cent}(g \circ f)$. If $C \in \mathrm{Miss}(f)$, then, since $q_0(f) = 0$, 2.4(a) implies that some fundamental point Q of f lies on C; then there exists a curve $C' \subset X$ such that $f(C') = \{Q\}$. Then $(g \circ f)(C') = \{P\}$ and hence $P \in \mathrm{cent}(g \circ f)$.

By [Daigle 1991a, 2.11], the condition $q_0(f) = 0$ is satisfied whenever $\mathrm{Cl}(Y) = 0$ and $\Gamma(X, \mathcal{O}_X)^* = \boldsymbol{k}^*$. In particular, if $X = \mathbb{A}^2 = Y$, then $q_0(f) = 0$, so (11) holds and consequently $\mathrm{cent}(g \circ f) = \mathrm{cent}(g) \cup g(\mathrm{cent}(f))$. □

2.12. Lemma. *Let $X \xrightarrow{f} Y \xrightarrow{g} Z$ be birational morphisms of nonsingular surfaces and let Γ_f (resp. Γ_g, $\Gamma_{g\circ f}$) be the union of all missing curves of f (resp. of g, $g \circ f$).*

(a) *$\Gamma_{g\circ f}$ is equal to the union of all 1-dimensional components of $\Gamma_g \cup \overline{g(\Gamma_f)}$, where $\overline{g(\Gamma_f)}$ denotes the closure of $g(\Gamma_f)$ in Z.*

(b) *If Y is affine, then $\Gamma_{g\circ f} = \Gamma_g \cup \overline{g(\Gamma_f)}$; in particular, each missing curve of f is included in $g^{-1}(\Gamma_{g\circ f})$.*

Proof. To prove (a), it's enough to show that a curve in Z is not included in $\Gamma_{g\circ f}$ if and only if it is not included in $\Gamma_g \cup \overline{g(\Gamma_f)}$. Let $C \subset Z$ be a curve such that $C \nsubseteq \Gamma_{g\circ f}$. Then there exists a curve $C_0 \subset X$ such that $g(f(C_0))$ is a dense subset of C; consequently, the set $C_1 = \overline{f(C_0)}$ is a curve in Y and $g(C_1)$ is dense in C, so C is not a missing curve of g and hence $C \nsubseteq \Gamma_g$. If $C \subseteq \overline{g(\Gamma_f)}$, then there exists a missing curve C_1' of f such that $\overline{g(C_1')} = C$; however, C_1 is the only curve in Y whose image by g is a dense subset of C, and C_1 is not a missing curve of f; so $C \nsubseteq \overline{g(\Gamma_f)}$ and hence $C \nsubseteq \Gamma_g \cup \overline{g(\Gamma_f)}$.

Conversely, let $C \subset Z$ be a curve such that $C \nsubseteq \Gamma_g \cup \overline{g(\Gamma_f)}$. Then $C \nsubseteq \Gamma_g$, so there exists a curve $C_1 \subset Y$ such that $g(C_1)$ is a dense subset of C. Note that C_1 is not a missing curve of f because $C \nsubseteq \overline{g(\Gamma_f)}$; so there exists a curve $C_0 \subset X$ such that $f(C_0)$ is a dense subset of C_1. Then $(g \circ f)(C_0)$ is a dense subset of C and consequently $C \nsubseteq \Gamma_{g\circ f}$. This proves (a).

To prove (b), suppose that Y is affine. If a point $P \in Z$ is an irreducible component of $\Gamma_g \cup \overline{g(\Gamma_f)}$, then $\{P\} = g(C)$, where C is a missing curve of f, so P is a fundamental point of g; since Y is affine, Lemma 2.8 implies that $\mathrm{cent}(g) \subseteq \Gamma_g$,

so $P \in \Gamma_g$, which contradicts the hypothesis that P is an irreducible component of $\Gamma_g \cup \overline{g(\Gamma_f)}$. This shows that $\Gamma_g \cup \overline{g(\Gamma_f)}$ is a union of curves, so the equality $\Gamma_{g\circ f} = \Gamma_g \cup \overline{g(\Gamma_f)}$ follows from (a). □

Results 2.13 and 2.14 are valid in all characteristics but are particularly interesting when $\operatorname{char} \boldsymbol{k} > 0$ (recall that not all lines are coordinate lines when $\operatorname{char} \boldsymbol{k} > 0$).

2.13. Proposition. *Let $f : \mathbb{A}^2 \to \mathbb{A}^2$ be a birational morphism.*

(a) *If a missing curve of f is nonsingular, then it is a coordinate line.*

(b) *If a contracting curve of f has one place at infinity, then it is a coordinate line.*

Proof. Assertion (a) follows from [Daigle 1991a, 4.6]. To prove (b), consider a contracting curve C of f such that C has one place at infinity. We noted in (4) that C is a nonsingular rational curve, so $C \cong \mathbb{A}^1$ is clear.

Let us embed $\operatorname{dom}(f) = \mathbb{A}^2$ in $X \cong \mathbb{P}^2$, and let L be the function field of X and V the prime divisor of $L/\boldsymbol{k}$ whose center on X is the closure of C in X (i.e., V is the DVR $\mathcal{O}_{X,\xi}$ where $\xi \in X$ is the generic point of C). Also embed $\operatorname{codom}(f) = \mathbb{A}^2$ in $Y \cong \mathbb{P}^2$, and note that the center of V on Y is zero-dimensional, since $C \in \operatorname{Cont}(f)$.

Consider the Kodaira dimension $\kappa(V)$ as defined in the introduction of Section 2 of [Ganong 1985]. Then $\kappa(V) < 0$ by [ibid., 2.1] and the fact that the center of V on Y is zero-dimensional; so C is a coordinate line by [ibid., 2.4]. □

2.14. Corollary. *Let C, C' be curves in $\mathbb{A}^2$ such that $C \cong \mathbb{A}^1 \cong C'$, and suppose that there exists a birational morphism $f : \mathbb{A}^2 \to \mathbb{A}^2$ such that $f(C)$ is a dense subset of C'. Then $f(C) = C'$. Moreover, if one of C, C' is a coordinate line, then both are coordinate lines.*

Proof. It is a simple fact that every dominant morphism $\mathbb{A}^1 \to \mathbb{A}^1$ is finite, hence surjective; so $f(C) = C'$.

If C is a coordinate line then there exists a birational morphism $g : \mathbb{A}^2 \to \mathbb{A}^2$ such that $C \in \operatorname{Miss}(g)$ (choose a coordinate system (X, Y) such that $C = \boldsymbol{Z}(X)$, and take $g(x, y) = (x, xy)$); then $C' \in \operatorname{Miss}(f \circ g)$ by Lemma 2.12, so Proposition 2.13(a) implies that C' is a coordinate line.

If C' is a coordinate line, then there exists a birational morphism $g : \mathbb{A}^2 \to \mathbb{A}^2$ such that $C' \in \operatorname{Cont}(g)$ (choose (X, Y) such that $C' = \boldsymbol{Z}(X)$ and take $g(x, y) = (x, xy)$); then $C \in \operatorname{Cont}(g \circ f)$, so Proposition 2.13(b) implies that C is a coordinate line. □

2.15. Lemma. *Let $\mathbb{A}^2 \xrightarrow{f} \mathbb{A}^2 \xrightarrow{g} \mathbb{A}^2$ be birational morphisms.*

(a) *If $\operatorname{Miss}(f) \subseteq \operatorname{Cont}(g)$, then $\operatorname{Miss}(f)$ is admissible.*

(b) *If $\operatorname{Cont}(g) \subseteq \operatorname{Miss}(f)$, then $\operatorname{Cont}(g)$ is admissible.*

Proof. Applying statement (4) in 2.3 to the morphism g gives

(12) $D' = \sum_{C \in \mathrm{Cont}(g)} C$ *is an SNC-divisor of* $\mathbb{A}^2$ *and the graph* $\mathcal{G}(D', \mathbb{A}^2)$ *is a forest.*

If $\mathrm{Cont}(g) \subseteq \mathrm{Miss}(f)$ then each element of $\mathrm{Cont}(g)$ is a *nonsingular* missing curve of f, and hence is a coordinate line by 2.13(a); then (12) implies that $\mathrm{Cont}(g)$ is admissible, so (b) is proved.

If $\mathrm{Miss}(f) \subseteq \mathrm{Cont}(g)$ then, by (12), $D = \sum_{C \in \mathrm{Miss}(f)} C$ is an SNC-divisor of $\mathbb{A}^2$ and the graph $\mathcal{G}(D, \mathbb{A}^2)$ is a forest; in particular each missing curve of f is nonsingular and hence is a coordinate line by 2.13(a); so $\mathrm{Miss}(f)$ is admissible and (a) is proved. □

2.16. Lemma. *Let* $\mathbb{A}^2 \xrightarrow{f} \mathbb{A}^2 \xrightarrow{g} \mathbb{A}^2$ *be birational morphisms.*

(a) *If* $\mathrm{Miss}(f) \not\subseteq \mathrm{Cont}(g)$, *then* $q(g \circ f) > q(g)$.

(b) *If* $\mathrm{Cont}(g) \not\subseteq \mathrm{Miss}(f)$, *then* $c(g \circ f) > c(f)$.

Proof. (a) It is clear that $\mathrm{Miss}(g) \subseteq \mathrm{Miss}(g \circ f)$. If C is a missing curve of f which is not contracted by g, then the closure $\overline{g(C)}$ of $g(C)$ is a missing curve of $g \circ f$ but not a missing curve of g, so $\mathrm{Miss}(g) \subset \mathrm{Miss}(g \circ f)$ and hence $q(g) < q(g \circ f)$.

(b) We have $\mathrm{Cont}(f) \subseteq \mathrm{Cont}(g \circ f)$. If C is a contracting curve of g which is not a missing curve of f, then there exists a curve $C' \subset \mathbb{A}^2$ such that $f(C')$ is a dense subset of C. Then C' is a contracting curve of $g \circ f$ but not one of f, so $\mathrm{Cont}(f) \subset \mathrm{Cont}(g \circ f)$ and hence $c(f) < c(g \circ f)$. □

In Lemma 2.17 and Proposition 2.18, we write $\#\Gamma$ for the number of irreducible components of a closed set Γ and $\Gamma_\tau = \bigcup_{C \in \mathrm{Miss}(\tau)} C$ for any birational morphism τ of nonsingular surfaces.

2.17. Lemma. *Let* $f : \mathbb{A}^2 \to \mathbb{A}^2$ *be a birational morphism and* Γ *a union of missing curves of* f *such that*

(13) *each missing curve of* f *is either included in* Γ *or disjoint from* Γ.

Then $\#f^{-1}(\Gamma) = \#\Gamma$ *and* f *can be factored as* $\mathbb{A}^2 \xrightarrow{g} \mathbb{A}^2 \xrightarrow{h} \mathbb{A}^2$, *where* g, h *are birational morphisms,* $\Gamma_h = \Gamma$ *and* $\Gamma_g \cap \mathrm{exc}(h) = \varnothing$.

Lemma 2.17 is an immediate consequence of the next result:

2.18. Proposition. *Let* $f : X \to Y$ *be a birational morphism where* X, Y *are nonsingular affine surfaces, and let* $\Gamma \subset Y$ *be a union of missing curves of* f *satisfying* (13). *Then the following hold*:

(a) f *can be factored as* $X \xrightarrow{g} W \xrightarrow{h} Y$, *where* W *is a nonsingular affine surface,* g, h *are birational morphisms,* $\Gamma_h = \Gamma$, $c(h) = \#f^{-1}(\Gamma)$ *and* $\Gamma_g \cap \mathrm{exc}(h) = \varnothing$.

(b) *If X, Y are factorial with trivial units, then $\#f^{-1}(\Gamma) = \#\Gamma$ and, in part* (a), *W can be chosen to be factorial with trivial units.*

(c) *If $X = \mathbb{A}^2$ and Y is factorial, then $Y = \mathbb{A}^2$ and we can choose $W = \mathbb{A}^2$ in part* (a).

Proof. (a) We may choose a commutative diagram (2) satisfying $n = n(f)$ and in which the blowings-up $\pi_1, \dots, \pi_n$ are ordered in such a way that the points over Γ are blown-up first; i.e., there exists $m \in \{0, \dots, n\}$ such that

$$\{i \in \{1, \dots, n\} \mid P_i \in \Gamma \text{ or } P_i \text{ is infinitely near a point of } \Gamma\} = \{1, \dots, m\}.$$

Refer to 2.3 for the notation. If $0 \le j \le k \le n$ and $D \subset Y_j$ is a curve or a union of curves, we write $\widetilde{D}^{Y_k}$ for the strict transform of D on Y_k. Let J be the set of $j \in \{1, \dots, m\}$ such that $\widetilde{E}_j^{Y_n} \cap X = \varnothing$ (recall that X is an open subset of Y_n), and define

$$W = Y_m \setminus \left(\widetilde{\Gamma}^{Y_m} \cup \bigcup_{j \in J} \widetilde{E}_j^{Y_m} \right). \tag{14}$$

Then W is a nonsingular surface and f factors as $X \xrightarrow{g} W \xrightarrow{h} Y$, where g, h are birational morphisms, $\Gamma_h = \Gamma$ and $\mathrm{Cont}(h) = \{\widetilde{E}_i^{Y_m} \cap W \mid i \in \{1, \dots, m\} \setminus J\}$; thus

$$q(h) = \#\Gamma \quad \text{and} \quad c(h) = \#f^{-1}(\Gamma). \tag{15}$$

Let $\Gamma' = C_1 \cup \cdots \cup C_s$, where $C_1, \dots, C_s \subset Y$ are the missing curves of f not included in Γ; then $\Gamma_f = \Gamma \cup \Gamma'$ and $\Gamma \cap \Gamma' = \varnothing$. Moreover,

$$\mathrm{Miss}(g) = \{\widetilde{C}_i^{Y_m} \cap W \mid i = 1, \dots, s\}. \tag{16}$$

Indeed, consider $C \in \mathrm{Miss}(g)$. If $h(C)$ is a point, then $C = \widetilde{E}_j^{Y_m} \cap W$ for some $j \in \{1, \dots, m\}$, and, in fact, $\widetilde{E}_j^{Y_n} \cap X = \varnothing$ (so $j \in J$) otherwise C would not be a missing curve of g; then (14) implies that $\widetilde{E}_j^{Y_m} \cap W = \varnothing$, which is absurd. So $h(C)$ is a dense subset of a curve $C_* \subset Y$. Now $C_* \in \mathrm{Miss}(f)$ by Lemma 2.12, and (14) implies that $C \nsubseteq \widetilde{\Gamma}^{Y_m}$, hence $C_* \nsubseteq \Gamma$; so $C_* \subseteq \Gamma'$ and consequently $C = \widetilde{C}_i^{Y_m} \cap W$ for some $i \in \{1, \dots, s\}$. This proves "$\subseteq$" in (16), and "$\supseteq$" easily follows from Lemma 2.12.

From (16), we deduce that $\Gamma_g \cap \mathrm{exc}(h) \subseteq h^{-1}(\Gamma') \cap h^{-1}(\Gamma)$, so

$$\Gamma_g \cap \mathrm{exc}(h) = \varnothing. \tag{17}$$

To prove (a), it only remains to show that W is affine. Since X is affine, Lemma 2.8 implies that $W \setminus \Gamma_g$ is affine; as (by (17)) $\mathrm{exc}(h) \subset W \setminus \Gamma_g$,

$$\text{no contracting curve of } h \text{ is a complete curve.} \tag{18}$$

Embed Y_0 in a nonsingular projective surface $\overline{Y}_0$ and enlarge diagram (2) as follows:

$$\begin{array}{ccccccccc}
\overline{Y}_n & \xrightarrow{\bar\pi_n} & \cdots & \xrightarrow{\bar\pi_{m+1}} & \overline{Y}_m & \xrightarrow{\bar\pi_m} & \cdots & \xrightarrow{\bar\pi_1} & \overline{Y}_0 \\
\uparrow & & & & \uparrow & & & & \uparrow \\
Y_n & \xrightarrow{\pi_n} & \cdots & \xrightarrow{\pi_{m+1}} & Y_m & \xrightarrow{\pi_m} & \cdots & \xrightarrow{\pi_1} & Y_0 \\
\uparrow & & & & \uparrow & & & & \| \\
X & & \xrightarrow{g} & & W & & \xrightarrow{h} & & Y
\end{array}$$

Let $D_i = \overline{Y}_i \setminus Y_i$ $(0 \le i \le n)$. Since $Y = Y_0$ is affine, D_0 is a nonempty connected union of curves and each irreducible component of $\overline{\Gamma}$ meets D_0 (where $\overline{\Gamma}$ denotes the closure of Γ in $\overline{Y}_0$). It follows that D_m is a nonempty connected union of curves and each irreducible component of $\widetilde{\Gamma}^{Y_m}$ meets D_m. Let us argue that

(19) W is connected at infinity.

Suppose that (19) is false; then $\overline{Y}_m \setminus W$ is not connected, so some connected component $\mathcal{C}$ of $\overline{Y}_m \setminus W$ is disjoint from D_m. Then $\mathcal{C}$ does not contain any irreducible component of $\widetilde{\Gamma}^{Y_m}$. By (14), it follows that $\mathcal{C} \subseteq \bigcup_{j\in J} \widetilde{E}_j^{Y_m}$.

We have $\overline{Y}_m \setminus (W \setminus \Gamma_g) = \overline{\widetilde{C}}_1^{Y_m} \cup \cdots \cup \overline{\widetilde{C}}_s^{Y_m} \cup (\overline{Y}_m \setminus W)$ by (16); since $W \setminus \Gamma_g$ is affine,

$$\overline{\widetilde{C}}_1^{Y_m} \cup \cdots \cup \overline{\widetilde{C}}_s^{Y_m} \cup (\overline{Y}_m \setminus W) \text{ is connected.}$$

As $\overline{Y}_m \setminus W$ is not connected and $\mathcal{C}$ is a connected component of it, some $\overline{\widetilde{C}}_i^{Y_m}$ must meet $\mathcal{C}$. So there exist $i \in \{1, \ldots, s\}$ and $j \in J$ such that $\overline{\widetilde{C}}_i^{Y_m} \cap \widetilde{E}_j^{Y_m} \ne \varnothing$. As $C_i \subseteq \Gamma'$, this implies that $P_j \in \Gamma'$ or P_j is i.n. a point of Γ'; since $j \le m$, we also have $P_j \in \Gamma$ or P_j is i.n. a point of Γ; so $\Gamma \cap \Gamma' \ne \varnothing$, a contradiction. Thus (19) is true.

In view of (18), (19) and the fact that Y is affine, applying [Daigle 1991a, 2.2] to $h : W \to Y$ shows that W is affine and concludes the proof of (a).

(b) Assume that X, Y are factorial and have trivial units; then [ibid., 3.4] gives $q(h) \le c(h)$, so $\#\Gamma \le \#f^{-1}(\Gamma)$ by (15). Since Γ' also satisfies (13), it follows that $\#\Gamma' \le \#f^{-1}(\Gamma')$.

By Lemma 2.8 we have $\mathrm{cent}(f) \subseteq \Gamma_f = \Gamma \cup \Gamma'$, so $f^{-1}(\Gamma) \cup f^{-1}(\Gamma')$ is exactly the union of all contracting curves of f; as $f^{-1}(\Gamma) \cap f^{-1}(\Gamma') = \varnothing$, we get $\#f^{-1}(\Gamma) + \#f^{-1}(\Gamma') = c(f)$. We have $c(f) = q(f)$ by [ibid., 2.9], and it is clear that $q(f) = \#\Gamma + \#\Gamma'$, so

$$\#\Gamma \le \#f^{-1}(\Gamma), \quad \#\Gamma' \le \#f^{-1}(\Gamma') \quad \text{and} \quad \#\Gamma + \#\Gamma' = \#f^{-1}(\Gamma) + \#f^{-1}(\Gamma');$$

consequently,

(20) $$\#\Gamma = \#f^{-1}(\Gamma),$$

(21) $$q(h) = c(h),$$

where (21) follows from (20) and (15). By (21), (19) and [ibid., 3.4], we get that W is factorial and has trivial units, which proves (b).

If $X = \mathbb{A}^2$ and Y is factorial then, by (b), W may be chosen to be factorial; then [ibid., 4.2] implies that W and Y are isomorphic to $\mathbb{A}^2$, which proves (c) and completes the proof of the proposition. □

2.19. Definition. Let $f : X \to Y$ be a birational morphism of nonsingular surfaces. Consider a diagram (2) satisfying $n = n(f)$ and with notation as in 2.3 (for each $i \in \{1, \dots, n\}$, let $\pi_i : Y_i \to Y_{i-1}$ be the blowing-up of Y_{i-1} at the point $P_i \in Y_{i-1}$).

(a) Let C be a missing curve of f. For each $i = 0, 1, \dots, n$, let $C^{Y_i} \subset Y_i$ denote the strict transform of C on Y_i ($C^{Y_0} = C$). Then we define the natural number

$$n(f, C) = \text{ cardinality of the set } \{\, i \mid 1 \le i \le n,\ P_i \in C^{Y_{i-1}} \}$$

and note that $n(f, C)$ depends only on (f, C), i.e., is independent of the choice of diagram (2). To indicate that $n(f, C) = k$, we say that "C is blown-up k times".

(b) For $i = 1, \dots, n$, let $\bar{P}_i \in Y_0$ be the image of P_i by $\pi_1 \circ \cdots \circ \pi_{i-1} : Y_{i-1} \to Y_0$. For each $P \in Y$, define the natural number

$$n(f, P) = \text{ cardinality of the set } \{\, i \mid 1 \le i \le n,\ \bar{P}_i = P \}$$

and note that $n(f, P)$ depends only on (f, P), i.e., is independent of the choice of the diagram (2) used for defining it.

2.20. Remarks. Let $f : X \to Y$ be a birational morphism of nonsingular surfaces.

(a) Let $C \in \mathrm{Miss}(f)$. Then $n(f, C) = 0 \iff C \cap f(X) = \varnothing$, and $n(f, C) = 1$ implies that there exists exactly one fundamental point of f lying on C. Note that if $X = \mathbb{A}^2 = Y$, then each missing curve contains at least one fundamental point (Lemma 2.6(d)), so each missing curve is blown-up at least once.

(b) Let $P \in Y$. Then $n(f, P) > 0 \iff P \in \mathrm{cent}(f)$, where "$\Leftarrow$" is obvious and "$\Rightarrow$" follows from (6). It is also clear that $n(f) = \sum_{P \in Y} n(f, P)$.

2.21. Lemma. *Let $X \xrightarrow{f} Y \xrightarrow{g} Z$ be birational morphisms of nonsingular surfaces, and assume that $n(g \circ f) = n(g) + n(f)$ or $X = Y = Z = \mathbb{A}^2$.*

(a) *Let $D \in \mathrm{Miss}(g)$; then $D \in \mathrm{Miss}(g \circ f)$ and $n(g \circ f, D) = n(g, D)$.*

(b) *Let $C \in \mathrm{Miss}(f) \setminus \mathrm{Cont}(g)$ and let D be the closure of $g(C)$ in Z. Then:*

- *$D \in \mathrm{Miss}(g \circ f)$ and $n(f, C) \le n(g \circ f, D)$;*
- *$n(f, C) = n(g \circ f, D) \Longrightarrow C \cap \mathrm{exc}(g) = \varnothing$;*
- *if $g(C) = D$ or $C \cong \mathbb{A}^1$, then*

$$n(f, C) = n(g \circ f, D) \iff C \cap \mathrm{exc}(g) = \varnothing.$$

(c) *For each* $P \in Z$, *we have* $n(g \circ f, P) = n(g, P) + \sum_{Q \in g^{-1}(P)} n(f, Q)$.

Proof. If $X = Y = Z = \mathbb{A}^2$, then $n(g \circ f) = n(g) + n(f)$ by Lemma 2.5; so $n(g \circ f) = n(g) + n(f)$ holds in all cases. Let $m = n(f)$ and $n = n(g)$. Choose commutative diagrams (I) and (II) as follows:

$$\begin{array}{ccccc} Y_m & \xrightarrow{\pi_m} & \cdots & \xrightarrow{\pi_1} & Y_0 \\ \uparrow & & \text{(I)} & & \| \\ X & & \xrightarrow[f]{} & & Y \end{array} \qquad \begin{array}{ccccc} Z_n & \xrightarrow{\rho_n} & \cdots & \xrightarrow{\rho_1} & Z_0 \\ \uparrow & & \text{(II)} & & \| \\ Y & & \xrightarrow[g]{} & & Z \end{array}$$

and use them to build the commutative diagram

(III)
$$\begin{array}{ccccccccc} Z_{n+m} & \xrightarrow{\rho_{n+m}} & \cdots & \xrightarrow{\rho_{n+2}} & Z_{n+1} & \xrightarrow{\rho_{n+1}} & Z_n & \xrightarrow{\rho_n} \cdots \xrightarrow{\rho_1} & Z_0 \\ \uparrow & & & & \uparrow & & \uparrow & & \| \\ Y_m & \xrightarrow{\pi_m} & \cdots & \xrightarrow{\pi_2} & Y_1 & \xrightarrow{\pi_1} & Y_0 & & \| \\ \uparrow & & & & & & \| & & \| \\ X & & \xrightarrow[f]{} & & & & Y & \xrightarrow[g]{} & Z \end{array}$$

In the three diagrams, "$\hookrightarrow$" are open immersions, Y_i, Z_i are nonsingular surfaces, $Y_i \xrightarrow{\pi_i} Y_{i-1}$ is the blowing-up of Y_{i-1} at a point $P_i \in Y_{i-1}$ and $Z_i \xrightarrow{\rho_i} Z_{i-1}$ is the blowing-up of Z_{i-1} at a point $Q_i \in Z_{i-1}$. Moreover, $Y_{i-1} \hookrightarrow Z_{n+i-1}$ maps P_i on Q_{n+i} (let us simply write $P_i = Q_{n+i}$). Diagrams (I) and (II) are minimal in the sense that $n(f) = m$ and $n(g) = n$; since $n(g \circ f) = n(f) + n(g) = m + n$, it follows that (III) is also minimal.

(a) Let $D \in \mathrm{Miss}(g)$; then $D \in \mathrm{Miss}(g \circ f)$ by Lemma 2.12(a). Let $D^{Z_i} \subset Z_i$ be the strict transform of $D \subset Z_0$ on Z_i. As $D^{Z_n} \subseteq Z_n \setminus Y_0$ and $\mathrm{cent}(\rho_{n+1} \circ \cdots \circ \rho_{n+m}) = \mathrm{cent}(\pi_1 \circ \cdots \circ \pi_m) \subset Y_0$, we see that

$$\{ i \mid 1 \le i \le n+m,\ Q_i \in D^{Z_{i-1}} \} = \{ i \mid 1 \le i \le n,\ Q_i \in D^{Z_{i-1}} \}. \tag{22}$$

Since $n(g \circ f, D)$ (resp. $n(g, D)$) is by definition the cardinality of the set in the left-hand side (resp. right-hand side) of (22), we have $n(g \circ f, D) = n(g, D)$.

(b) Let $C \in \mathrm{Miss}(f) \setminus \mathrm{Cont}(g)$ and let D be the closure of $g(C)$ in Z. Then $D \in \mathrm{Miss}(g \circ f)$ by Lemma 2.12(a). Define $D^{Z_i} \subset Z_i$ as before; then

$$\{ i \mid n+1 \le i \le n+m,\ Q_i \in D^{Z_{i-1}} \} \subseteq \{ i \mid 1 \le i \le n+m,\ Q_i \in D^{Z_{i-1}} \}. \tag{23}$$

Since $n(f, C)$ (resp. $n(g \circ f, D)$) is the cardinality of the set in the left-hand side (resp. right-hand side) of (23), we have $n(f, C) \le n(g \circ f, D)$, and, moreover,

$$n(f, C) \ne n(g \circ f, D) \tag{24}$$

is equivalent to

$$\{i \mid 1 \le i \le n,\ Q_i \in D^{Z_{i-1}}\} \neq \varnothing. \tag{25}$$

By the minimality of diagram (II) together with (6), (25) is equivalent to

$$D \cap \operatorname{cent}(g) \neq \varnothing. \tag{26}$$

Now

$$C \cap \operatorname{exc}(g) \neq \varnothing \tag{27}$$

implies (26) and, if $g(C) = D$, the converse is true. So we have shown that

$$n(f, C) = n(g \circ f, D) \implies C \cap \operatorname{exc}(g) = \varnothing, \tag{28}$$

and that the converse holds whenever $g(C) = D$. Finally, we observe that if $C \cong \mathbb{A}^1$, then the dominant morphism $C \xrightarrow{g} D$ is necessarily finite, hence surjective, so the converse of (28) is true whenever $C \cong \mathbb{A}^1$. This proves (b).

(c) Define $\bar{Q}_i = (\rho_1 \circ \cdots \circ \rho_{i-1})(Q_i) \in Z_0$ for all $i = 1, \ldots, m+n$ and observe that the trivial equality (for any $P \in Z$)

$$\left|\{i \mid \bar{Q}_i = P\}\right| = \left|\{i \mid i \le n \text{ and } \bar{Q}_i = P\}\right| + \left|\{i \mid i > n \text{ and } \bar{Q}_i = P\}\right|$$

is the desired equality. □

3. Compositions of simple affine contractions

Let $\boldsymbol{k}$ be an algebraically closed field of arbitrary characteristic, and let $\mathbb{A}^2 = \mathbb{A}^2_{\boldsymbol{k}}$. As in the introduction, we write $\operatorname{Bir}(\mathbb{A}^2)$ for the monoid of birational endomorphisms $f : \mathbb{A}^2 \to \mathbb{A}^2$, and we declare that $f, g \in \operatorname{Bir}(\mathbb{A}^2)$ are equivalent ($f \sim g$) if $u \circ f \circ v = g$ for some $u, v \in \operatorname{Aut}(\mathbb{A}^2)$. The equivalence class of $f \in \operatorname{Bir}(\mathbb{A}^2)$ is denoted $[f]$. Note that the conditions $f \sim f'$ and $g \sim g'$ do *not* imply that $f \circ g \sim f' \circ g'$.

The aim of this section is to describe the subsets $S_{\mathrm{w}} \supset S_{\mathrm{a}} \supset S_{\mathrm{aa}}$ of $\operatorname{Bir}(\mathbb{A}^2)$ defined by

$$S_{\mathrm{w}} = \{f \in \operatorname{Bir}(\mathbb{A}^2) \mid \operatorname{Miss}(f) \text{ is weakly admissible}\},$$
$$S_{\mathrm{a}} = \{f \in \operatorname{Bir}(\mathbb{A}^2) \mid \operatorname{Miss}(f) \text{ is admissible}\},$$
$$S_{\mathrm{aa}} = \{f \in \operatorname{Bir}(\mathbb{A}^2) \mid \text{both } \operatorname{Miss}(f) \text{ and } \operatorname{Cont}(f) \text{ are admissible}\}$$

(refer to Definition 1.3, Lemma 1.4, Definition 1.6 and Proposition 1.7); note that these sets are not closed under composition of morphisms. We learn at a relatively early stage (see Proposition 3.6(c)) that each element of S_{w} is a composition of simple affine contractions (SACs are defined in the introduction and again in Definition 3.2). However, an arbitrary composition of SACs does not necessarily belong to S_{w} (resp. S_{a}, S_{aa}), so in each of the three cases one has to determine

which compositions of SACs give the desired type of endomorphism. The answer is given in Theorem 3.15, which is the main result of this section.

The material of 3.1–3.5(a) can be found in [Daigle 1991a; 1991b]; everything else appears to be new.

As before, we have $\mathbb{A}^2 = \operatorname{Spec} A$, where $A = \boldsymbol{k}^{[2]}$ is fixed throughout, and by a coordinate system of $\mathbb{A}^2$ we mean a pair $(X, Y) \in A \times A$ satisfying $A = \boldsymbol{k}[X, Y]$ (see the introduction of Section 1).

3.1. Let $\mathfrak{C}$ temporarily denote the set of coordinate systems of $\mathbb{A}^2$. Given an element $\mathfrak{c} = (X, Y)$ of $\mathfrak{C}$, consider the $\boldsymbol{k}$-homomorphism $A \to A$ given by $X \mapsto X$ and $Y \mapsto XY$; this homomorphism determines a morphism $\mathbb{A}^2 \to \mathbb{A}^2$, which we denote $\alpha_{\mathfrak{c}}$; clearly, $\alpha_{\mathfrak{c}} \in \operatorname{Bir}(\mathbb{A}^2)$. Note that if $\mathfrak{c}_1, \mathfrak{c}_2 \in \mathfrak{C}$, then $\alpha_{\mathfrak{c}_1}^m \sim \alpha_{\mathfrak{c}_2}^m$ for every $m \geq 1$. So, for each $m \geq 1$, the equivalence class $[\alpha_{\mathfrak{c}}^m]$ of $\alpha_{\mathfrak{c}}^m$ is independent of the choice of $\mathfrak{c} \in \mathfrak{C}$.

3.2. Definition. A birational morphism $\mathbb{A}^2 \to \mathbb{A}^2$ is called a *simple affine contraction (SAC)* if it is equivalent to $\alpha_{\mathfrak{c}}$ for some (hence for every) coordinate system $\mathfrak{c}$ of $\mathbb{A}^2$.

Note that if f is a SAC and $\mathfrak{c} \in \mathfrak{C}$, then $f \sim \alpha_{\mathfrak{c}}$, but f^2 need not be equivalent to $\alpha_{\mathfrak{c}}^2$.

For readers who like to identify $\mathbb{A}^2$ with $\boldsymbol{k}^2$, we note that $\alpha_{\mathfrak{c}}$ corresponds to the map $\boldsymbol{k}^2 \to \boldsymbol{k}^2$, $(x, y) \mapsto (x, xy)$ and that the SACs are obtained by composing this map on both sides with automorphisms. See 1.1.

3.3. Lemma. (a) *A birational morphism $f : \mathbb{A}^2 \to \mathbb{A}^2$ is a SAC if and only if $n(f) = 1$.*

(b) *If $f : \mathbb{A}^2 \to \mathbb{A}^2$ is a SAC, then f has one missing curve L and one fundamental point P; moreover, L is a coordinate line and $P \in L$.*

(c) *Let $L \subset \mathbb{A}^2$ be a coordinate line and $P \in L$ a point. Let $X \xrightarrow{\pi} \mathbb{A}^2$ be the blowing-up of $\mathbb{A}^2$ at P and $U \subset X$ the complement of the strict transform of L. Then $U \cong \mathbb{A}^2$ and the composition $\mathbb{A}^2 \xrightarrow{\cong} U \hookrightarrow X \xrightarrow{\pi} \mathbb{A}^2$ is a SAC with missing curve L and fundamental point P.*

(d) *If $f, g : \mathbb{A}^2 \to \mathbb{A}^2$ are two SAC with the same missing curve and the same fundamental point then there exists an automorphism $\theta : \mathbb{A}^2 \to \mathbb{A}^2$ such that $g = f \circ \theta$:*

$$\begin{array}{ccc} \mathbb{A}^2 & \xrightarrow[\theta]{\cong} & \mathbb{A}^2 \\ & {\scriptstyle g} \searrow \quad \swarrow {\scriptstyle f} & \\ & \mathbb{A}^2 & \end{array}$$

(e) *Let $\mathfrak{c}$ be a coordinate system of $\mathbb{A}^2$ and $\alpha_{\mathfrak{c}} \in \operatorname{Bir}(\mathbb{A}^2)$ as in 3.1. For $f \in \operatorname{Bir}(\mathbb{A}^2)$,*

$$q(f) = 1 \iff c(f) = 1 \iff f \sim \alpha_{\mathfrak{c}}^{n(f)}.$$

Proof. Assertion (a) is [Daigle 1991a, 4.10] or [Daigle 1991b, 4.1]; assertions (b)–(d) are trivial; (e) is [Daigle 1991a, 4.11] together with the fact that $q(f) = c(f)$ by Lemma 2.6(a). □

Remark. Assertion 3.3(e) can be phrased as follows: given an integer $m \geq 1$ and a coordinate system $\mathfrak{c}$ of $\mathbb{A}^2$, the equivalence class $[\alpha_{\mathfrak{c}}^m]$ of $\alpha_{\mathfrak{c}}^m$ is

$$\begin{aligned}[\alpha_{\mathfrak{c}}^m] &= \{f \in \mathrm{Bir}(\mathbb{A}^2) \mid q(f) = 1 \text{ and } n(f) = m\} \\ &= \{f \in \mathrm{Bir}(\mathbb{A}^2) \mid c(f) = 1 \text{ and } n(f) = m\}.\end{aligned}$$

3.4. Corollary. *If $f \in \mathrm{Bir}(\mathbb{A}^2)$ has a unique missing curve C, then C is a coordinate line.*

Proof. This follows from Lemma 3.3(e). It also follows from Proposition 2.13(a) and Proposition 2.9. □

See Definition 2.19 for the definition of the phrase "L is blown-up only once".

3.5. Lemma. *Let $f \in \mathrm{Bir}(\mathbb{A}^2)$.*

(a) *Suppose that some missing curve L of f is blown-up only once. Then L is a coordinate line. Moreover, if $P \in L$ is the unique fundamental point of f which is on L and $\gamma \in \mathrm{Bir}(\mathbb{A}^2)$ is a SAC with missing curve L and fundamental point P, then f factors as $\mathbb{A}^2 \xrightarrow{f'} \mathbb{A}^2 \xrightarrow{\gamma} \mathbb{A}^2$ with $f' \in \mathrm{Bir}(\mathbb{A}^2)$.*

(b) *Suppose that f factors as $\mathbb{A}^2 \xrightarrow{f'} \mathbb{A}^2 \xrightarrow{\gamma} \mathbb{A}^2$ with $f', \gamma \in \mathrm{Bir}(\mathbb{A}^2)$, where γ is a SAC. Let L be the missing curve of γ. Then L is a missing curve of f which is blown-up only once.*

Proof. Part (a) is an improvement of [Daigle 1991b, 4.6]. The proof of [ibid., 4.6] shows that P is a nonsingular point of L; then [Daigle 1991a, 4.6] implies that L is a coordinate line. Choose a diagram (2) for f such that $n = n(f)$ and $P_1 = P$. Let $L^{Y_1} \subset Y_1$ denote the strict transform of L on Y_1 and define $W = Y_1 \setminus L^{Y_1} \subset Y_1$. As L is a missing curve of f and is blown-up only once, the image of $\mathbb{A}^2 \hookrightarrow Y_n \xrightarrow{\pi_2 \circ \cdots \circ \pi_n} Y_1$ is included in W; so f factors as $\mathbb{A}^2 \xrightarrow{g'} W \xrightarrow{h'} \mathbb{A}^2$, where g', h' are birational morphisms and h' is the composition $W \hookrightarrow Y_1 \xrightarrow{\pi_1} Y_0 = \mathbb{A}^2$. By Lemma 3.3(c), $W \cong \mathbb{A}^2$ and the composition $\mathbb{A}^2 \cong W \xrightarrow{h'} \mathbb{A}^2$ is a SAC with missing curve L and fundamental point P; so f factors as $\mathbb{A}^2 \xrightarrow{g} \mathbb{A}^2 \xrightarrow{h} \mathbb{A}^2$, where $g, h \in \mathrm{Bir}(\mathbb{A}^2)$ and h is a SAC with missing curve L and fundamental point P. By Lemma 3.3(d), $h = \gamma \circ \theta$ for some $\theta \in \mathrm{Aut}(\mathbb{A}^2)$. Then $f = \gamma \circ \theta \circ g$ and we are done.

For (b), we know that by Lemma 2.21(a), L is in $\mathrm{Miss}(f)$ and $n(f, L) = n(\gamma \circ f', L) = n(\gamma, L) = 1$. □

3.6. Proposition. *Let $f : \mathbb{A}^2 \to \mathbb{A}^2$ be a birational morphism such that*

(i) *f is not an isomorphism*;

(ii) *there exists a coordinate system of $\mathbb{A}^2$ with respect to which all missing curves of f have degree* 1.

Then there exists a missing curve of f which is blown-up only once. Moreover, if L is such a curve and $P \in L$ is the unique fundamental point of f which is on L, then the following holds:

(a) *There exists a coordinate system (X, Y) of $\mathbb{A}^2$ such that $L = \boldsymbol{Z}(X)$ and $P = (0, 0)$, and such that the union of the missing curves of f is equal to the zero set of one of the following polynomials in $\boldsymbol{k}[X, Y]$:*

(i) *$XY^m \prod_{i=1}^{n}(X - a_i)$ for some $m \in \{0, 1\}$, $n \geq 0$ and $a_1, \ldots, a_n \in \boldsymbol{k}$;*

(ii) *$X(X-1)^m \prod_{i=1}^{n}(Y - b_i X)$ for some $m \in \{0, 1\}$, $n \geq 0$ and $b_1, \ldots, b_n \in \boldsymbol{k}$.*

(b) *If $\gamma : \mathbb{A}^2 \to \mathbb{A}^2$ is a SAC with missing curve L and fundamental point P, then f factors as $\mathbb{A}^2 \xrightarrow{f'} \mathbb{A}^2 \xrightarrow{\gamma} \mathbb{A}^2$, where f' is a birational morphism such that* Miss(f') *is admissible.*

(c) *f is a composition of SACs.*

Proof. By [Daigle 1991b, 4.7], $f = h \circ g$, where $g, h \in \mathrm{Bir}(\mathbb{A}^2)$ and h is a SAC; then Lemma 3.5(b) implies that some missing curve of f is blown-up only once.

Let L be a missing curve of f which is blown-up only once, and let $P \in L$ be the unique fundamental point of f which is on L. Choose a coordinate system (X, Y) of $\mathbb{A}^2$ such that $L = \boldsymbol{Z}(X)$ and $P = (0, 0)$ and with respect to which all missing curves of f have degree 1. Define $\gamma_0 : \mathbb{A}^2 \to \mathbb{A}^2$ by $\gamma_0(x, y) = (x, xy)$. As γ_0 is a SAC with missing curve L and fundamental point P, Lemma 3.5(a) implies that f factors as $\mathbb{A}^2 \xrightarrow{f'} \mathbb{A}^2 \xrightarrow{\gamma_0} \mathbb{A}^2$ for some $f' \in \mathrm{Bir}(\mathbb{A}^2)$. Let Γ (resp. Γ') be the union of the missing curves of f (resp. of f'). Then

$$\Gamma = \boldsymbol{Z}(X) \cup \overline{\gamma_0(\Gamma')}, \tag{29}$$

by Lemma 2.12. In particular, if C is a missing curve of f', then $\gamma_0(C)$ is included into a line of degree 1; from $\gamma_0(x, y) = (x, xy)$, it easily follows that C is either a vertical line $\boldsymbol{Z}(X - a)$ (for some $a \in \boldsymbol{k}$), a horizontal line $\boldsymbol{Z}(Y - b)$ (for some $b \in \boldsymbol{k}$) or a hyperbola $\boldsymbol{Z}(X(\alpha + \beta Y) - 1)$ (for some $\alpha, \beta \in \boldsymbol{k}$, $\beta \neq 0$), where, in fact, the last case cannot occur because C has one place at infinity by Lemma 2.6(c). So

(30) *each missing curve of f' is either a vertical line or a horizontal line.*

In particular, all missing curves of f' have degree 1. It follows from the first part of the proof that

(31) *if f' is not an isomorphism, some missing curve of f' is blown-up only once.*

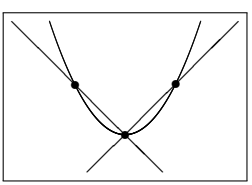
Let h (resp. v) be the number of missing curves of f' which are horizontal (resp. vertical) lines. Then $\min(h, v) \leq 1$, otherwise (by Lemma 2.6(e)) every missing curve of f' would contain at least two fundamental points of f', and hence would be blown-up at least twice, contradicting (31). Statement (30), together with $\min(h, v) \leq 1$, implies that $\mathrm{Miss}(f')$ is admissible, which proves the special case "$\gamma = \gamma_0$" of assertion (b); in view of Lemma 3.3(d), it follows that (b) is true.

If $h \leq 1$, then Γ' is the zero set of $(Y - b)^h \prod_{i=1}^{v}(X - a_i)$ for some $b \in \boldsymbol{k}$ and $a_1, \ldots, a_v \in \boldsymbol{k}$. Then, by (29), Γ is the zero set of $X(Y - bX)^h \prod_{i=1}^{v}(X - a_i)$. Replacing the coordinate system (X, Y) by $(X, Y - bX)$, we see that (a-i) is satisfied.

If $v \leq 1$, then Γ' is the zero set of $(X - a)^v \prod_{i=1}^{h}(Y - b_i)$ for some $a \in \boldsymbol{k}$ and $b_1, \ldots, b_h \in \boldsymbol{k}$. Then Γ is the zero set of $X(X - a)^v \prod_{i=1}^{h}(Y - b_i X)$. If $a = 0$ or $v = 0$, then Γ is the zero set of $X(X - 1)^0 \prod_{i=1}^{h}(Y - b_i X)$, so (a-ii) holds; if $a \neq 0$ and $v \neq 0$, then Γ is the zero set of $X(X - a) \prod_{i=1}^{h}(Y - b_i X)$, so (a-ii) holds after replacing (X, Y) by $(a^{-1}X, Y)$. So assertion (a) is true.

To prove assertion (c), consider the factorization $f = \gamma \circ f'$ given by (b). Since $n(\gamma) = 1$ by Lemma 3.3(a), we have $n(f') = n(f) - 1$ by Lemma 2.5. Moreover, the fact that $\mathrm{Miss}(f')$ is admissible implies, by Proposition 1.7, that there exists a coordinate system of $\mathbb{A}^2$ with respect to which all missing curves of f' have degree 1. It is clear that (c) follows by induction on $n(f)$. □

3.7. Remark. We stress that assumption (ii) of Proposition 3.6 is strictly stronger than "all missing curves are coordinate lines". Indeed, there exists an irreducible element $f \in \mathrm{Bir}(\mathbb{A}^2)$ with three missing curves, these being the lines $\boldsymbol{Z}(X + Y)$ and $\boldsymbol{Z}(X - Y)$ and the parabola $\boldsymbol{Z}(Y - X^2)$:

$\mathbb{A}^2 \xrightarrow{f} \mathbb{A}^2 =$ 

(an example of such an f, due to Russell, appeared in [Daigle 1991a, 4.7]). Here, each missing curve is a coordinate line and hence has degree 1 with respect to a suitable choice of coordinate system. However, these three lines are not *simultaneously* rectifiable, so f does not satisfy assumption (ii) of Proposition 3.6 (it does not satisfy the conclusion either: since f is not a SAC and is irreducible, it is not a composition of SACs).

Also note that by Lemma 1.4, assumption (ii) of Proposition 3.6 is equivalent to "$\mathrm{Miss}(f)$ is weakly admissible".

3.8. Until the end of Section 3,

we fix a coordinate system $\mathfrak{c} = (X, Y)$ of $\mathbb{A}^2$.

This allows us to identify $\mathbb{A}^2$ with $\boldsymbol{k}^2$. See 1.1 for the notation $\boldsymbol{Z}(F)$ and for our convention regarding the definition of morphisms using coordinates.

3.9. In 3.9.1–3.9.3 below, we define three submonoids of $\mathrm{Bir}(\mathbb{A}^2)$, denoted $\mathcal{H}_{\mathfrak{c}}$, $\mathcal{G}_{\mathfrak{c}}$ and $\mathcal{V}_{\mathfrak{c}}$, respectively. The subscript $\mathfrak{c}$ reminds us that these sets depend on the choice of $\mathfrak{c}$ made in 3.8. Since $\mathfrak{c}$ is fixed until the end of this section, there is no harm in omitting it and writing simply $\mathcal{H}$, $\mathcal{G}$ and $\mathcal{V}$. It is clear from the definitions below that these three monoids are included in the submonoid of $\mathrm{Bir}(\mathbb{A}^2)$ generated by SACs and automorphisms.

3.9.1. Given $m \in \mathbb{N}$ and $p \in \boldsymbol{k}[Y]$ such that[3] $\deg p < m$, define $h_{m,p} \in \mathrm{Bir}(\mathbb{A}^2)$ by $h_{m,p}(x, y) = (xy^m + p(y), y)$. Observe that $h_{m,p}$ is equivalent to γ^m, where γ is the SAC given by $(x, y) \mapsto (xy, y)$; consequently, $n(h_{m,p}) = n(\gamma^m) = mn(\gamma)$, i.e.,

$$n(h_{m,p}) = m.$$

Define $\mathcal{H} = \mathcal{H}_{\mathfrak{c}} = \{h_{m,p} \mid m \in \mathbb{N},\ p \in \boldsymbol{k}[Y] \text{ and } \deg p < m\}$. It is easily verified that $\mathcal{H}$ is a submonoid of $\mathrm{Bir}(\mathbb{A}^2)$.

3.9.2. Let $\boldsymbol{M}$ be the multiplicative monoid whose elements are the 2×2 matrices $M = \left(\begin{smallmatrix} i & j \\ k & \ell \end{smallmatrix}\right)$ with $i, j, k, \ell \in \mathbb{N}$ and $i\ell - jk = \pm 1$. It is easily verified that $\boldsymbol{M}$ is generated by $G_1 = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $G_2 = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Given $M = \left(\begin{smallmatrix} i & j \\ k & \ell \end{smallmatrix}\right) \in \boldsymbol{M}$, define the morphism $\gamma_M : \mathbb{A}^2 \to \mathbb{A}^2$ by $(x, y) \mapsto (x^i y^j, x^k y^\ell)$. Note that $\gamma_{M_1} \circ \gamma_{M_2} = \gamma_{M_1 M_2}$ for all $M_1, M_2 \in \boldsymbol{M}$, so the set

$$\mathcal{G} = \mathcal{G}_{\mathfrak{c}} = \{\gamma_M \mid M \in \boldsymbol{M}\}$$

is a monoid (under composition) generated by $\{\gamma_{G_1}, \gamma_{G_2}\}$. As γ_{G_1} is a SAC and γ_{G_2} is an automorphism, it follows that $\mathcal{G}$ is a submonoid of $\mathrm{Bir}(\mathbb{A}^2)$.

3.9.3. Given a polynomial $\varphi \in \boldsymbol{k}[X] \setminus \{0\}$, define $v_\varphi \in \mathrm{Bir}(\mathbb{A}^2)$ by $v_\varphi(x, y) = (x, \varphi(x)y)$. Then let

$$\mathcal{V} = \mathcal{V}_{\mathfrak{c}} = \{v_\varphi \mid \varphi \in \boldsymbol{k}[X] \setminus \{0\}\}.$$

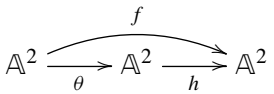
Note that $v_\varphi \circ v_\psi = v_{\varphi \cdot \psi} = v_\psi \circ v_\varphi$ for any $\varphi, \psi \in \boldsymbol{k}[X] \setminus \{0\}$, so $\mathcal{V}$ is a submonoid of $\mathrm{Bir}(\mathbb{A}^2)$.

3.10. Lemma. *For a birational morphism $f : \mathbb{A}^2 \to \mathbb{A}^2$, the following are equivalent*:

(a) *The union of the missing curves of f is included in $\boldsymbol{Z}(Y)$.*

(b) *There exists $(h, \theta) \in \mathcal{H} \times \mathrm{Aut}(\mathbb{A}^2)$ such that $f = h \circ \theta$:*

[3] We adopt the convention that the zero polynomial has degree $-\infty$; consequently, the condition $\deg p < 0$ is equivalent to p being the zero polynomial (so $h_{0,p} = h_{0,0}$ is the identity map).



Moreover, if these conditions are satisfied, then the pair (h, θ) *in* (b) *is unique.*

Proof. We leave it to the reader to verify that (b) implies (a) and that (h, θ) in (b) is unique. By induction on $n(f)$, we show that (a) implies (b).

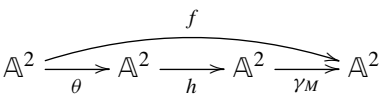
If $n(f) = 0$, then (b) holds with $\theta = f$ and $h = h_{0,0}$.

If $n(f) > 0$, then f is not an isomorphism, and hence has at least one missing curve, so $\boldsymbol{Z}(Y)$ is the unique missing curve of f; by Proposition 3.6, this missing curve is blown-up only once. This missing curve must contain a fundamental point $(c, 0)$ of f; as $h_{1,c} \in \mathcal{H}$ is a SAC with missing curve $\boldsymbol{Z}(Y)$ and fundamental point $(c, 0)$, Proposition 3.6 implies that $f = h_{1,c} \circ f'$ for some birational morphism $f' : \mathbb{A}^2 \to \mathbb{A}^2$. It is immediate that $h_{1,c}^{-1}(\Gamma) = \Gamma$, where $\Gamma = \boldsymbol{Z}(Y)$ is the missing curve of f; so Lemma 2.12 implies that the union of the missing curves of f' is included in $\boldsymbol{Z}(Y)$. As $n(f') = n(f) - 1$, we may assume by induction that $f' = h' \circ \theta$ for some $h' \in \mathcal{H}$ and $\theta \in \mathrm{Aut}(\mathbb{A}^2)$. Then $f = h_{1,c} \circ h' \circ \theta$ is the desired factorization, where we note that $h_{1,c} \circ h' \in \mathcal{H}$. □

3.11. Lemma. *For a birational morphism* $f : \mathbb{A}^2 \to \mathbb{A}^2$, *the following are equivalent*:

(a) *The union of the missing curves of* f *is included in* $\boldsymbol{Z}(XY)$.

(b) *There exists* $(M, h, \theta) \in \boldsymbol{M} \times \mathcal{H} \times \mathrm{Aut}(\mathbb{A}^2)$ *such that* $f = \gamma_M \circ h \circ \theta$:



Proof. It is easily verified that (b) implies (a). We prove that (a) implies (b) by induction on $n(f)$. Assume that f satisfies (a).

If $n(f) = 0$, then f is an isomorphism, so (b) holds with $\theta = f$, $h = h_{0,0}$ and $M = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

Let $n > 0$ and assume that the result is true whenever $n(f) < n$. Now consider f satisfying (a) and such that $n(f) = n$.

If $q(f) = 1$, then the missing curve Γ of f is $\boldsymbol{Z}(X)$ or $\boldsymbol{Z}(Y)$. Define $M = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ $\left(\text{resp. } M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)\right)$ if $\Gamma = \boldsymbol{Z}(Y)$ (resp. $\Gamma = \boldsymbol{Z}(X)$). Then $\gamma_M \circ f$ has a unique missing curve, and this curve is $\boldsymbol{Z}(Y)$. Applying Lemma 3.10 to $\gamma_M \circ f$ gives $\gamma_M \circ f = h \circ \theta$ for some $\theta \in \mathrm{Aut}(\mathbb{A}^2)$ and $h \in \mathcal{H}$. Noting that $\gamma_M \circ \gamma_M$ is the identity, we get $f = \gamma_M \circ h \circ \theta$.

From now on, assume that $q(f) = 2$. Let Γ be the union of the missing curves of f; i.e., $\Gamma = \boldsymbol{Z}(XY)$. By Proposition 3.6, some element L of $\mathrm{Miss}(f) = \{\boldsymbol{Z}(X), \boldsymbol{Z}(Y)\}$ is blown-up only once. As $(0, 0)$ is a common point of the two missing curves, it must be a fundamental point of f. For a suitable choice of

$M_1 \in \left\{\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right), \left(\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right)\right\}$, γ_{M_1} is a SAC with missing curve L and fundamental point $(0,0)$. Then Proposition 3.6 implies that $f = \gamma_{M_1} \circ f'$ for some birational morphism $f' : \mathbb{A}^2 \to \mathbb{A}^2$. By Lemma 2.12, the union of the missing curves of f' is included in $\gamma_{M_1}^{-1}(\Gamma) = \Gamma$, so f' satisfies (a). As $n(f') = n(f) - 1$, the inductive hypothesis implies that $f' = \gamma_{M_2} \circ h \circ \theta$ for some $\theta \in \mathrm{Aut}(\mathbb{A}^2)$, $h \in \mathcal{H}$ and $M_2 \in \boldsymbol{M}$. So $f = \gamma_{M_1} \circ \gamma_{M_2} \circ h \circ \theta$, and since $\gamma_{M_1} \circ \gamma_{M_2} = \gamma_{M_1 M_2}$, we are done. □

3.12. Let $\Delta = \Delta_{\mathfrak{c}}$ be the subgroup of $\mathrm{Aut}(\mathbb{A}^2)$ whose elements are of the form $\delta(x, y) = (x, y + q(x))$ with $q \in \boldsymbol{k}[X]$.

3.13. Lemma. *Let* $\Gamma = \boldsymbol{Z}\left(\prod_{i=1}^{s}(X - c_i)\right)$, *where* $c_1, \ldots, c_s$ $(s \geq 0)$ *are distinct elements of* $\boldsymbol{k}$. *For a birational morphism* $f : \mathbb{A}^2 \to \mathbb{A}^2$, *the following are equivalent*:

(a) *The union of the missing curves of* f *is included in* Γ.

(b) *There exists a commutative diagram*

$$\begin{array}{ccc} \mathbb{A}^2 & \xrightarrow{f} & \mathbb{A}^2 \\ {\scriptstyle\theta}\downarrow & & \downarrow{\scriptstyle\delta} \\ \mathbb{A}^2 & \xrightarrow[v_\varphi]{} & \mathbb{A}^2 \end{array}$$

where $\delta \in \Delta$, $\theta \in \mathrm{Aut}(\mathbb{A}^2)$, $\varphi \in \boldsymbol{k}[X] \setminus \{0\}$ *and where the set of roots of* φ *is included in* $\{c_1, \ldots, c_s\}$.

Proof. That (b) implies (a) is left to the reader. Suppose that f satisfies (a). We prove (b) by induction on $n(f)$.

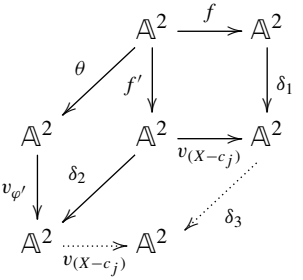
If $n(f) = 0$ then f is an isomorphism, so (b) holds with $\theta = f$, $\varphi = 1$ and $\delta = \mathrm{id}$.

Let $n > 0$ be such that the result is true whenever $n(f) < n$. Consider f satisfying (a) and such that $n(f) = n$. Then f is not an isomorphism, and hence has at least one missing curve (so $s > 0$). By Proposition 3.6, one of the missing curves (say $L = \boldsymbol{Z}(X - c_j)$) of f is blown-up only once. We know that L contains a fundamental point (c_j, d) of f; let $\delta_1 \in \Delta$ be defined by $\delta_1(x, y) = (x, y - d)$ and let $f_1 = \delta_1 \circ f$. Since L is a missing curve of f which is blown-up only once and $(c_j, d) \in L$ is a fundamental point of f, it follows that $\delta_1(L) = L$ is a missing curve of f_1 which is blown-up only once and that $\delta_1(c_j, d) = (c_j, 0) \in L$ is a fundamental point of f_1. As $v_{(X - c_j)}$ is a SAC with missing curve L and fundamental point $(c_j, 0)$, Proposition 3.6 implies that f_1 factors through $v_{(X-c_j)}$. Thus $\delta_1 \circ f = v_{(X-c_j)} \circ f'$ for some birational morphism $f' : \mathbb{A}^2 \to \mathbb{A}^2$:

$$\begin{array}{ccc} \mathbb{A}^2 & \xrightarrow{f} & \mathbb{A}^2 \\ {\scriptstyle f'}\downarrow & & \downarrow{\scriptstyle\delta_1} \\ \mathbb{A}^2 & \xrightarrow[v_{(X-c_j)}]{} & \mathbb{A}^2 \end{array}$$

Since δ_1 maps each vertical line onto itself, the union of all missing curves of f_1 is Γ; so, by Lemma 2.12, the union of the missing curves of f' is included in $v_{(X-c_j)}^{-1}(\Gamma) = \Gamma$, so f' satisfies (a). As $n(f') = n(f) - 1$, the inductive hypothesis implies that there exists a commutative diagram (ignore the dotted arrows for now)

(32)



with $\varphi' \in \boldsymbol{k}[X] \setminus \{0\}$ (and all roots of φ' are in $\{c_1, \ldots, c_s\}$), $\theta \in \mathrm{Aut}(\mathbb{A}^2)$ and $\delta_2 \in \Delta$ defined by $\delta_2(x, y) = (x, y + q(x))$ (for some $q \in \boldsymbol{k}[X]$). Now if we define $\delta_3 \in \Delta$ by $\delta_3(x, y) = (x, y + (x - c_j)q(x))$, then

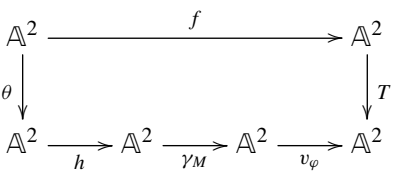
$$\delta_3 \circ v_{(X-c_j)} = v_{(X-c_j)} \circ \delta_2.$$

So diagram (32), including the dotted arrows, is commutative. Let $\delta = \delta_3 \circ \delta_1 \in \Delta$ and $\varphi = (X - c_j)\varphi'(X)$ (so $v_{(X-c_j)} \circ v_{\varphi'} = v_\varphi$); then $\delta, \theta, v_\varphi$ give the commutative diagram displayed in the statement of assertion (b). □

3.14. Lemma. *Let* $\Gamma = \boldsymbol{Z}\big(Y \prod_{i=1}^{s}(X - c_i)\big)$, *where* $s \geq 1$ *and* $c_1, \ldots, c_s$ *are distinct elements of* $\boldsymbol{k}$. *Let* $f : \mathbb{A}^2 \to \mathbb{A}^2$ *be a birational morphism such that*

(33) *the union of the missing curves of* f *is equal to* Γ.

Then there exists a commutative diagram



where $T \in \mathrm{Aut}(\mathbb{A}^2)$ *is of the form* $T(x, y) = (x - c, y)$ *with* $c \in \boldsymbol{k}$, θ *is an arbitrary element of* $\mathrm{Aut}(\mathbb{A}^2)$ *and* $(\varphi, M, h) \in (\boldsymbol{k}[X] \setminus \{0\}) \times \boldsymbol{M} \times \mathcal{H}$.

Proof. We first settle the case $s = 1$. Define $T \in \mathrm{Aut}(\mathbb{A}^2)$ by $T(x, y) = (x - c_1, y)$. Then the union of the missing curves of $T \circ f$ is $\boldsymbol{Z}(XY)$, so Lemma 3.11 implies that there exists $(M, h, \theta) \in \boldsymbol{M} \times \mathcal{H} \times \mathrm{Aut}(\mathbb{A}^2)$ such that $T \circ f = \gamma_M \circ h \circ \theta = v_\varphi \circ \gamma_M \circ h \circ \theta$ with $\varphi = 1$. Thus the result is true when $s = 1$.

We proceed by induction on $n(f)$. For f satisfying (33), we have $q(f)=s+1\geq 2$, so the least possible value for $n(f)$ is 2. If $n(f)=2$, then $q(f)\leq n(f)=2$, so $s=1$ and the result is true in this case.

Let $n>2$ be such that the result is true whenever $n(f)<n$. Consider f satisfying (33) and such that $n(f)=n$.

By the first paragraph, we may assume that $s>1$. By Proposition 3.6, one of the missing curves (say L) of f is blown-up only once; we choose such an L. By Lemma 2.6(e), the points $(c_i,0)$, $1\leq i\leq s$, are fundamental points of f; so $\boldsymbol{Z}(Y)$ is blown-up at least $s\geq 2$ times and hence $L=\boldsymbol{Z}(X-c_j)$ for some $j\in\{1,\dots,s\}$. As $v_{(X-c_j)}$ is a SAC with missing curve L and fundamental point $(c_j,0)$, Proposition 3.6 implies that $f=v_{(X-c_j)}\circ f'$ for some birational morphism $f':\mathbb{A}^2\to\mathbb{A}^2$. Let $\Gamma'=\bigcup_{C\in\mathrm{Miss}(f')}C$. By Lemma 2.12, $\Gamma'\subseteq v_{(X-c_j)}^{-1}(\Gamma)=\Gamma$; in fact, it is easy to see (again by Lemma 2.12) that

$$\boldsymbol{Z}\Big(Y\prod_{i\in I}(X-c_i)\Big)\subseteq\Gamma'\subseteq\boldsymbol{Z}\Big(Y\prod_{i=1}^{s}(X-c_i)\Big),$$

where $I=\{1,\dots,s\}\setminus\{j\}$, so f' satisfies the hypothesis of the lemma. As $n(f')=n(f)-1$, the inductive hypothesis implies that $T\circ f'=v_\psi\circ\gamma_M\circ h\circ\theta$ for some $(\psi,M,h)\in(\boldsymbol{k}[X]\setminus\{0\})\times\boldsymbol{M}\times\mathcal{H}$ and $\theta,T\in\mathrm{Aut}(\mathbb{A}^2)$, where T is of the form $T(x,y)=(x-c,y)$ for some $c\in\boldsymbol{k}$. Noting that $T\circ v_{(X-c_j)}=v_{(X+c-c_j)}\circ T$, we get

$$\begin{aligned}T\circ f&=T\circ v_{(X-c_j)}\circ f'=v_{(X+c-c_j)}\circ T\circ f'=v_{(X+c-c_j)}\circ v_\psi\circ\gamma_M\circ h\circ\theta\\&=v_\varphi\circ\gamma_M\circ h\circ\theta,\end{aligned}$$

where $\varphi(X)=(X+c-c_j)\psi(X)\in\boldsymbol{k}[X]\setminus\{0\}$, as desired. □

Before stating the main theorem of this section, let us recall the assumptions under which it is valid. Our base field $\boldsymbol{k}$ is an algebraically closed field of arbitrary characteristic, and $\mathbb{A}^2$ is the affine plane over $\boldsymbol{k}$. We fix a coordinate system $\mathfrak{c}=(X,Y)$ of $\mathbb{A}^2$; this allows us to use coordinates for defining morphisms $\mathbb{A}^2\to\mathbb{A}^2$ (see Section 1). The choice of $\mathfrak{c}$ also determines the submonoids $\mathcal{V}=\mathcal{V}_\mathfrak{c}$, $\mathcal{G}=\mathcal{G}_\mathfrak{c}$ and $\mathcal{H}=\mathcal{H}_\mathfrak{c}$ of $\mathrm{Bir}(\mathbb{A}^2)$ (see 3.9). Then we have the following result:

3.15. Theorem. *Let $f:\mathbb{A}^2\to\mathbb{A}^2$ be a birational morphism.*

(a) *The following conditions are equivalent*:

(i) $\mathrm{Miss}(f)$ *is weakly admissible.*

(ii) f *is equivalent to one of the following elements of* $\mathrm{Bir}(\mathbb{A}^2)$:

- $\alpha_i^m\circ v\circ\gamma\circ h$ *for some* $(v,\gamma,h)\in\mathcal{V}\times\mathcal{G}\times\mathcal{H}$, $m\in\{0,1\}$ *and* $i\in\{1,2\}$, *where* $\alpha_1,\alpha_2\in\mathrm{Bir}(\mathbb{A}^2)$ *are the SACs defined by* $\alpha_1(x,y)=(xy,y)$ *and* $\alpha_2(x,y)=(x(1-y),1-y)$;

- *the morphism* $(x, y) \mapsto \big(x(p(x)y + q(x)),\ p(x)y + q(x)\big)$, *for some* $p, q \in \boldsymbol{k}[X]$ *with* $p \neq 0$.

(b) *The following conditions are equivalent*:

(i) $\mathrm{Miss}(f)$ *is admissible.*

(ii) f *is equivalent to* $v \circ \gamma \circ h$ *for some* $(v, \gamma, h) \in \mathcal{V} \times \mathcal{G} \times \mathcal{H}$.

(c) *The following conditions are equivalent*:

(i) *Both* $\mathrm{Miss}(f)$ *and* $\mathrm{Cont}(f)$ *are admissible.*

(ii) f *is equivalent to an element of* $\mathcal{V} \cup \mathcal{G}$.

Proof. For each of (a), (b) and (c), we show that (i) implies (ii) and leave the converse to the reader. We begin with (b).

Suppose that f satisfies (b-i). Let $\Gamma = \bigcup_{C \in \mathrm{Miss}(f)} C$. By Proposition 1.7, there exists $\omega \in \mathrm{Aut}(\mathbb{A}^2)$ such that $\omega(\Gamma) = \boldsymbol{Z}\big(Y^d \prod_{i=1}^{s}(X - c_i)\big)$, where

$$d \in \{0, 1\},\ s \geq 0 \text{ and } c_1, \ldots, c_s \text{ are distinct elements of } \boldsymbol{k}.$$

Note that the union of the missing curves of $f_1 = \omega \circ f$ equals $\boldsymbol{Z}\big(Y^d \prod_{i=1}^{s}(X - c_i)\big)$; since $f_1 \sim f$, it is enough to prove that f_1 is equivalent to $v \circ \gamma \circ h$ for some $(v, \gamma, h) \in \mathcal{V} \times \mathcal{G} \times \mathcal{H}$. So we may as well replace f by f_1 throughout; so from now on we assume that

$$\Gamma = \bigcup_{C \in \mathrm{Miss}(f)} C = \boldsymbol{Z}\left(Y^d \prod_{i=1}^{s}(X - c_i)\right).$$

If $d = 0$ (resp. $s = 0$), then the desired conclusion follows from Lemma 3.13 (resp. from Lemma 3.10). So we may assume that $d = 1$ and $s \geq 1$. Then Lemma 3.14 gives the desired conclusion; i.e., we showed that (b-i) implies (b-ii).

Suppose that (a-i) holds. Let $\Gamma = \bigcup_{C \in \mathrm{Miss}(f)} C$. By Lemma 1.4, f satisfies the hypotheses of Proposition 3.6. To prove (a-ii), we may assume that $\mathrm{Miss}(f)$ is not admissible (otherwise (a-ii) follows from (b)). Then Proposition 3.6 implies that there exists $\omega \in \mathrm{Aut}(\mathbb{A}^2)$ such that $\omega(\Gamma) = \boldsymbol{Z}(F)$ with

$$F = Y \prod_{i=1}^{s}(X - c_i Y) \quad \text{or} \quad F = Y(Y - 1) \prod_{i=1}^{s}(X - c_i Y), \tag{34}$$

where $s \geq 2$ and $c_1, \ldots, c_s \in \boldsymbol{k}$ are distinct. We know, also by Proposition 3.6, that some missing curve of f (say $C_0 \in \mathrm{Miss}(f)$) is blown-up only once. In the second case of (34), $\omega(C_0)$ is necessarily equal to $\boldsymbol{Z}(Y)$; in the first case, we may choose ω in such a way that $\omega(C_0) = \boldsymbol{Z}(Y)$.

It is clear that we may replace f by $\omega \circ f$ throughout. Then we have $\Gamma = \boldsymbol{Z}(F)$, $\boldsymbol{Z}(Y)$ is a missing curve of f which is blown-up only once and $(0, 0)$ is the unique fundamental point of f which lies on $\boldsymbol{Z}(Y)$. If F is as in the first (resp. the second) case of (34), let $\alpha = \alpha_1$ (resp. $\alpha = \alpha_2$), where $\alpha_1, \alpha_2 \in \mathrm{Bir}(\mathbb{A}^2)$ are defined in the

statement of (a-ii); then α is a SAC with missing curve $\boldsymbol{Z}(Y)$ and fundamental point $(0,0)$. By Proposition 3.6, it follows that $f=\alpha\circ f'$ for some $f'\in \mathrm{Bir}(\mathbb{A}^2)$. Let Γ' be the union of the missing curves of f'. Using Lemma 2.12, we find

$$\text{in the first case of (34)},\quad \boldsymbol{Z}\Big(\prod_{i=1}^{s}(X-c_i)\Big)\subseteq\Gamma'\subseteq\boldsymbol{Z}\Big(Y\prod_{i=1}^{s}(X-c_i)\Big);$$

$$\text{in the second case of (34)},\quad \boldsymbol{Z}\Big(Y\prod_{i=1}^{s}(X-c_i)\Big)\subseteq\Gamma'\subseteq\boldsymbol{Z}\Big(Y(Y-1)\prod_{i=1}^{s}(X-c_i)\Big).$$

In particular, f' satisfies the hypothesis of Proposition 3.6; by that result, some missing curve of f' is blown-up only once, and thus Γ' cannot be equal to $\boldsymbol{Z}\big(Y(Y-1)\prod_{i=1}^{s}(X-c_i)\big)$. It follows that $\Gamma'=\boldsymbol{Z}(G)$, where

$$G=\prod_{i=1}^{s}(X-c_i)\quad\text{or}\quad G=Y\prod_{i=1}^{s}(X-c_i). \tag{35}$$

First consider the case $G=\prod_{i=1}^{s}(X-c_i)$; then $\alpha=\alpha_1$ because the first case of (35) can only happen in the first case of (34). By Lemma 3.13, there is a commutative diagram

$$\begin{array}{ccc}\mathbb{A}^2 & \xrightarrow{f'} & \mathbb{A}^2\\ {\scriptstyle\theta}\downarrow & & \downarrow{\scriptstyle\delta}\\ \mathbb{A}^2 & \xrightarrow[v]{} & \mathbb{A}^2\end{array}$$

where $v\in\boldsymbol{\mathcal{V}}$, $\delta,\theta\in\mathrm{Aut}(\mathbb{A}^2)$ and δ is of the form $\delta(x,y)=(x,y-q(x))$ for some $q\in\boldsymbol{k}[X]$. Then $f=\alpha_1\circ f'=\alpha_1\circ\delta^{-1}\circ v\circ\theta\sim\alpha_1\circ\delta^{-1}\circ v$. Let $p\in\boldsymbol{k}[X]\setminus\{0\}$ be such that $v(x,y)=(x,p(x)y)$; then

$$(\alpha_1\circ\delta^{-1}\circ v)(x,y)=\big(x(p(x)y+q(x)),\,p(x)y+q(x)\big),$$

which shows that (a-ii) holds in this case.

Consider the second case, $G=Y\prod_{i=1}^{s}(X-c_i)$. Here, α may be either one of α_1,α_2. By Lemma 3.14, there is a commutative diagram

$$\begin{array}{ccc}\mathbb{A}^2 & \xrightarrow{f'} & \mathbb{A}^2\\ {\scriptstyle\theta}\downarrow & & \downarrow{\scriptstyle T}\\ \mathbb{A}^2 & \xrightarrow[v\circ\gamma\circ h]{} & \mathbb{A}^2\end{array}$$

where $(v,\gamma,h)\in\boldsymbol{\mathcal{V}}\times\mathcal{G}\times\mathcal{H}$, $\theta\in\mathrm{Aut}(\mathbb{A}^2)$, and $T\in\mathrm{Aut}(\mathbb{A}^2)$ is of the form $T(x,y)=(x-c,y)$, with $c\in\boldsymbol{k}$. Now $\alpha\circ T^{-1}=\nu\circ\alpha$, where $\nu\in\mathrm{Aut}(\mathbb{A}^2)$ is given

by $v(x, y) = (x + cy, y)$. Thus

$$f = \alpha \circ f' = \alpha \circ T^{-1} \circ v \circ \gamma \circ h \circ \theta = v \circ \alpha \circ v \circ \gamma \circ h \circ \theta \sim \alpha \circ v \circ \gamma \circ h,$$

showing that (a-ii) holds in this case as well. So (a-i) implies (a-ii).

Let $\varphi \in \boldsymbol{k}[X] \setminus \{0\}$, $M = \left(\begin{smallmatrix} i & j \\ k & \ell \end{smallmatrix}\right) \in \boldsymbol{M}$ and $h_{m,p}(x, y) = (xy^m + p(y), y) \in \mathcal{H}$, where $m \in \mathbb{N}$ and $p(Y) \in \boldsymbol{k}[Y]$ are such that $\deg p < m$. As a preparation for the proof that (c-i) implies (c-ii), we first show:

(36) *If* $\mathrm{Cont}(\gamma_M \circ h_{m,p})$ *is admissible, then* $\gamma_M \circ h_{m,p} \sim \gamma$ *for some* $\gamma \in \mathcal{G}$.

(37) *If* $\mathrm{Cont}(v_\varphi \circ h_{m,p})$ *is admissible, then* $v_\varphi \circ h_{m,p}$ *is equivalent to an element of* $\mathcal{V} \cup \mathcal{G}$.

Observe that

$$(\gamma_M \circ h_{m,p})(x, y) = \left((xy^m + p(y))^i y^j,\ (xy^m + p(y))^k y^\ell\right). \tag{38}$$

To prove (36), first consider the case $ik \neq 0$; then (38) implies that $\boldsymbol{Z}(XY^m + p(Y))$ is a contracting curve (or a union of contracting curves) of $\gamma_M \circ h_{m,p}$. So, by the hypothesis of (36), each irreducible component of $\boldsymbol{Z}(XY^m + p(Y))$ has one place at infinity. The only way to achieve this is to have $p = 0$, in which case we have $h_{m,p} = \gamma_{M'}$ with $M' = \left(\begin{smallmatrix} 1 & m \\ 0 & 1 \end{smallmatrix}\right)$. Then $\gamma_M \circ h_{m,p} = \gamma_N \in \mathcal{G}$ with $N = MM'$, so (36) is true in this case.

Consider next the case where $ik = 0$. Then $M \in \left\{\left(\begin{smallmatrix} 0 & 1 \\ 1 & \ell \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & j \\ 0 & 1 \end{smallmatrix}\right)\right\}$ for some $j, \ell \in \mathbb{N}$. If $M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & \ell \end{smallmatrix}\right)$, then

$$(\gamma_M \circ h_{m,p})(x, y) = \left(y,\ (xy^m + p(y))y^\ell\right) = \left(y,\ xy^{m+\ell} + y^\ell p(y)\right),$$

which is equivalent to the birational morphism $(x, y) \mapsto (y, xy^{m+\ell})$, that is, $\gamma_M \circ h_{m,p} \sim \gamma_N \in \mathcal{G}$ with $N = \left(\begin{smallmatrix} 0 & 1 \\ 1 & m+\ell \end{smallmatrix}\right)$. Similarly, if $M = \left(\begin{smallmatrix} 1 & j \\ 0 & 1 \end{smallmatrix}\right)$ then $\gamma_M \circ h_{m,p} \sim \gamma_N \in \mathcal{G}$ with $N = \left(\begin{smallmatrix} 1 & m+j \\ 0 & 1 \end{smallmatrix}\right)$. This completes the proof of (36).

To prove (37), we first note that if $m = 0$, then $v_\varphi \circ h_{m,p} = v_\varphi \circ \mathrm{id} = v_\varphi \in \mathcal{V}$. Likewise, if $\varphi \in \boldsymbol{k}^*$, then v_φ is an isomorphism, so $v_\varphi \circ h_{m,p} \sim h_{m,p} \sim \gamma_N \in \mathcal{G}$ with $N = \left(\begin{smallmatrix} 1 & m \\ 0 & 1 \end{smallmatrix}\right)$. So we may assume from now on that $m > 0$ and that φ has at least one root.

If $c \in \boldsymbol{k}$ is a root of φ, then $\boldsymbol{Z}(XY^m + p(Y) - c)$ is a union of contracting curves of $v_\varphi \circ h_{m,p}$. Therefore, by the hypothesis of (37), each irreducible component of $\boldsymbol{Z}(XY^m + p(Y) - c)$ has one place at infinity. As $m > 0$, this implies that $p(Y) - c$ is the zero polynomial, and this is true for each root c of φ. So $\varphi = a(X - c)^n$ for some $a \in \boldsymbol{k}^*$ and $n \geq 1$, and $h_{m,p}(x, y) = (xy^m + c, y)$. Then $(v_\varphi \circ h_{m,p})(x, y) = (xy^m + c, a(xy^m)^n y)$, which is equivalent to $(x, y) \mapsto (xy^m, (xy^m)^n y)$; that is, $v_\varphi \circ h_{m,p} \sim \gamma_N \in \mathcal{G}$ with $N = \left(\begin{smallmatrix} 1 & m \\ n & mn+1 \end{smallmatrix}\right)$. This proves (37).

To prove that (c-i) implies (c-ii), we consider $f = v_\varphi \circ \gamma_M \circ h$ for some $(\varphi, M, h) \in (\boldsymbol{k}[X] \setminus \{0\}) \times \boldsymbol{M} \times \mathcal{H}$; we assume that $\mathrm{Cont}(f)$ is admissible, and we have to prove

(c-ii). We use the notation $M = \left(\begin{smallmatrix} i & j \\ k & \ell \end{smallmatrix}\right) \in \boldsymbol{M}$ and $h(x, y) = (xy^m + p(y), y)$, where $m \in \mathbb{N}$ and $p(Y) \in \boldsymbol{k}[Y]$ are such that $\deg p < m$.

The assumption that $\mathrm{Cont}(f)$ is admissible implies, in particular, that

(39) each contracting curve of $v_\varphi \circ \gamma_M$ has one place at infinity.

Indeed, suppose that $C \in \mathrm{Cont}(v_\varphi \circ \gamma_M)$ has more than one place at infinity; then, by Lemma 2.6(c), C is not a missing curve of h and consequently there exists a curve $C' \subset \mathbb{A}^2$ such that $h(C')$ is a dense subset of C. Then C' is a contracting curve of $f = v_\varphi \circ \gamma_M \circ h$ but has more than one place at infinity (because it dominates a curve with more than one place at infinity). This contradicts the assumption that $\mathrm{Cont}(f)$ is admissible, so (39) is proved.

We claim

(40) $ij = 0$ *or* $\varphi(X) = aX^n$, *for some* $a \in \boldsymbol{k}^*$ *and* $n \in \mathbb{N}$.

Indeed, suppose that φ is not of the form aX^n with $a \in \boldsymbol{k}^*$ and $n \in \mathbb{N}$; then there exists $c \in \boldsymbol{k}^*$ such that $\varphi(c) = 0$. Then $\boldsymbol{Z}(x^i y^j - c)$ is a contracting curve of $v_\varphi \circ \gamma_M$ and, if $ij \neq 0$, this curve has more than one place at infinity, contradicting (39). So (40) is proved.

Consider the case where $\varphi(X) = aX^n$. Then $v_\varphi = \theta \circ \gamma_{M_1}$, where $\theta \in \mathrm{Aut}(\mathbb{A}^2)$ and $M_1 = \left(\begin{smallmatrix} 1 & 0 \\ n & 1 \end{smallmatrix}\right) \in \boldsymbol{M}$. Then $f = v_\varphi \circ \gamma_M \circ h = \theta \circ \gamma_{M_1} \circ \gamma_M \circ h \sim \gamma_{M_1 M} \circ h$, so (36) implies that $f \sim \gamma$ for some $\gamma \in \mathfrak{G}$, and we are done in this case.

There remains the case $ij = 0$; here we have $M \in \left\{\left(\begin{smallmatrix} 0 & 1 \\ 1 & \ell \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ k & 1 \end{smallmatrix}\right)\right\}$ for some $k, \ell \in \mathbb{N}$.

If $M = \left(\begin{smallmatrix} 1 & 0 \\ k & 1 \end{smallmatrix}\right)$, then $\gamma_M = v_{(X^k)}$. So $f = v_\varphi \circ v_{(X^k)} \circ h_{m,p} = v_{\varphi_1} \circ h_{m,p}$, where $\varphi_1 = X^k \varphi(X)$, and (37) implies that f is equivalent to an element of $\mathcal{V} \cup \mathfrak{G}$ (so we are done).

If $M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & \ell \end{smallmatrix}\right)$, then $M = M_1 M_2$, where $M_1 = \left(\begin{smallmatrix} 1 & 0 \\ \ell & 1 \end{smallmatrix}\right)$ and $M_2 = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Now $\gamma_{M_2} = \tau$, where $\tau \in \mathrm{Aut}(\mathbb{A}^2)$ is defined by $\tau(x, y) = (y, x)$, and $\gamma_{M_1} = v_{(X^\ell)}$. So we have

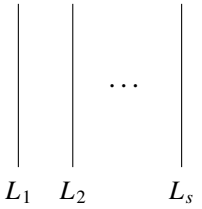
$$f \sim f \circ \tau = v_\varphi \circ \gamma_{M_1} \circ \gamma_{M_2} \circ h_{m,p} \circ \tau = (v_\varphi \circ v_{(X^\ell)}) \circ (\tau \circ h_{m,p} \circ \tau) = v_{\varphi_1} \circ (\tau \circ h_{m,p} \circ \tau),$$

where $\varphi_1 = X^\ell \varphi(X)$. We have $(\tau \circ h_{m,p} \circ \tau)(x, y) = (x, yx^m + p(x))$, so

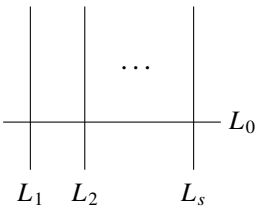
$$(v_{\varphi_1} \circ (\tau \circ h_{m,p} \circ \tau))(x, y) = \big(x, \varphi_1(x)(yx^m + p(x))\big) = (x, x^m \varphi_1(x) y + \varphi_1(x) p(x)),$$

which is equivalent to the birational morphism $(x, y) \mapsto (x, x^m \varphi_1(x) y) = v_\psi(x, y)$ with $\psi = X^m \varphi_1$. So $f \sim v_\psi \in \mathcal{V}$ and we have shown that (c-i) implies (c-ii). □

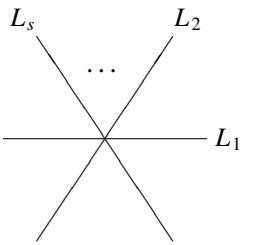
3.16. Corollary. *Let* $f \in \mathrm{Bir}\,\mathbb{A}^2$. *Suppose that all missing curves of* f *are lines and that these are simultaneously rectifiable. Then there exists a coordinate system of* $\mathbb{A}^2$ *with respect to which the configuration of missing curves is one of the following*:

(a) 

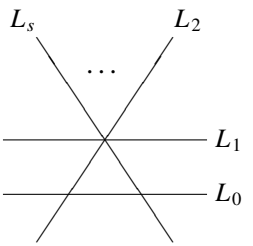
Parallel lines $L_1, \ldots, L_s$ ($s \geq 0$).

(b) 

Parallel lines $L_1, \ldots, L_s$ ($s \geq 1$), plus one line L_0 not parallel to $L_1, \ldots, L_s$.

(c) 

Concurrent lines $L_1, \ldots, L_s$ ($s \geq 3$).

(d) 

Concurrent lines $L_1, \ldots, L_s$ ($s \geq 3$), plus one line L_0, where L_0 is parallel to one of the concurrent lines.

Conversely, each of the above configurations of lines occurs as the configuration of missing curves of some $f \in \mathrm{Bir}(\mathbb{A}^2)$.

The proof below gives in each of the cases (a)–(d) an example of an $f \in \mathrm{Bir}(\mathbb{A}^2)$ having the desired configuration of missing curves.

Proof of Corollary 3.16. The hypothesis on f is that $\mathrm{Miss}(f)$ is weakly admissible, so f is described by part (a-ii) of Theorem 3.15; it follows that $\mathrm{Miss}(f)$ must be one of the configurations (a)–(d). Note that $\mathrm{Miss}(f)$ is admissible in cases (a) and (b). In cases (c) and (d), $\mathrm{Miss}(f)$ is weakly admissible but not admissible.

Conversely, consider the configurations of lines (a)–(d). In each of the four cases we may choose a coordinate system $\mathfrak{c} = (X, Y)$ of $\mathbb{A}^2$ with respect to which the

configuration of lines is $\boldsymbol{Z}(F)$, where

$$F = \begin{cases} \prod_{i=1}^{s}(X - c_i) & \text{in case (a),} \\ Y\prod_{i=1}^{s}(X - c_i) & \text{in case (b),} \\ Y\prod_{i=1}^{s-1}(X - c_i Y) & \text{in case (c),} \\ Y(Y-1)\prod_{i=1}^{s-1}(X - c_i Y) & \text{in case (d),} \end{cases}$$

where $c_1, \ldots, c_s$ (resp. $c_1, \ldots, c_{s-1}$) are distinct elements of $\boldsymbol{k}$ in cases (a) and (b) (resp. in cases (c) and (d)). Let us exhibit in each case an $f \in \mathrm{Bir}(\mathbb{A}^2)$ such that the union of all missing curves of f is $\boldsymbol{Z}(F)$. In cases (a) and (b), choose a univariate polynomial $\varphi \in \boldsymbol{k}[t]$ whose roots are exactly $c_1, \ldots, c_s$, and define $f \in \mathrm{Bir}(\mathbb{A}^2)$ by

$$f(x, y) = \begin{cases} (x, \varphi(x)y) & \text{in case (a),} \\ (xy, \varphi(xy)y) & \text{in case (b).} \end{cases}$$

Then the union of the missing curves of f is $\boldsymbol{Z}(F)$, as desired. In cases (c) and (d), first choose $g \in \mathrm{Bir}(\mathbb{A}^2)$ such that the union of the missing curves of g is $\boldsymbol{Z}(G)$, where

$$G = \begin{cases} \prod_{i=1}^{s-1}(X - c_i) & \text{in case (c),} \\ Y\prod_{i=1}^{s-1}(X - c_i) & \text{in case (d)} \end{cases}$$

(we know that g exists by cases (a) and (b)). Then define

$$f = \begin{cases} \alpha_1 \circ g & \text{in case (c),} \\ \alpha_2 \circ g & \text{in case (d),} \end{cases}$$

where α_1 and α_2 are defined in the statement of Theorem 3.15. It follows from Lemma 2.12(b) that the union of the missing curves of f is $\boldsymbol{Z}(F)$. □

4. Some aspects of the monoid $\mathrm{Bir}(\mathbb{A}^2)$

Let $\boldsymbol{k}$ be an algebraically closed field and $\mathbb{A}^2 = \mathbb{A}^2_{\boldsymbol{k}}$, and consider the noncommutative monoid $\mathrm{Bir}(\mathbb{A}^2)$ defined in the introduction. Note that this is a cancellative monoid since it is included in the group of birational automorphisms of $\mathbb{P}^2$.

In view of Lemma 2.5 and Lemma 2.6(b), it is clear that each noninvertible element of $\mathrm{Bir}(\mathbb{A}^2)$ is a composition of finitely many irreducible elements. In other words,

the monoid $\mathrm{Bir}(\mathbb{A}^2)$ *has factorizations into irreducibles.*

Essentially nothing is known regarding uniqueness of factorizations.[4]

[4]We do know that $\mathrm{Bir}(\mathbb{A}^2)$ is not a "unique factorization monoid" in the sense of [Johnson 1971], but this by no means settles the question of uniqueness of factorizations in $\mathrm{Bir}(\mathbb{A}^2)$. Indeed, there are several nonequivalent definitions of what one might mean by "uniqueness of factorization" in noncommutative monoids, and the one used in [Johnson 1971] seems to be particularly inadequate in the case of $\mathrm{Bir}(\mathbb{A}^2)$.

It is natural to ask whether one can find all irreducible elements of $\mathrm{Bir}(\mathbb{A}^2)$ up to equivalence. However, considering the examples given in [Daigle 1991a; 1991b] and [Cassou-Noguès and Russell 2007] and certain facts such as [Daigle 1991a, 4.12], one gets the impression that the irreducible endomorphisms might be too numerous and too diverse to be listed. The first part of the present section gives some simple observations (4.1–4.5) that strengthen that impression.

Given $f, g \in \mathrm{Bir}(\mathbb{A}^2)$, let us write $f \mid g$ if there exist $u, v \in \mathrm{Bir}(\mathbb{A}^2)$ such that $u \circ f \circ v = g$. By a *prime element* of $\mathrm{Bir}(\mathbb{A}^2)$, we mean a noninvertible element p satisfying

$$\text{for all } f, g \in \mathrm{Bir}(\mathbb{A}^2), \quad p \mid (g \circ f) \;\Rightarrow\; p \mid f \;\text{ or }\; p \mid g.$$

It follows from Lemmas 2.5 and 2.6(b) that every prime element of $\mathrm{Bir}(\mathbb{A}^2)$ is irreducible. It is natural to ask whether the converse is true, and, in particular, whether SACs are prime (SACs are certainly irreducible). These questions are open; we don't even know if there exists a prime element in $\mathrm{Bir}(\mathbb{A}^2)$.

We say that a submonoid $\mathcal{M}$ of $\mathrm{Bir}(\mathbb{A}^2)$ is *factorially closed in* $\mathrm{Bir}(\mathbb{A}^2)$ if the conditions $f, g \in \mathrm{Bir}(\mathbb{A}^2)$ and $g \circ f \in \mathcal{M}$ imply $f, g \in \mathcal{M}$. It is natural to ask whether $\mathcal{A}$ is factorially closed in $\mathrm{Bir}(\mathbb{A}^2)$, where $\mathcal{A}$ is the submonoid of $\mathrm{Bir}(\mathbb{A}^2)$ generated by SACs and automorphisms.[5] The main result of this section, Theorem 4.8, states that $\mathcal{A}$ is indeed factorially closed in $\mathrm{Bir}(\mathbb{A}^2)$.

Remark. It is obvious that the only irreducible elements of $\mathcal{A}$ are the SACs, that each noninvertible element of $\mathcal{A}$ is a composition of irreducible elements and that $\mathcal{A}$ has the following "unique factorization" property: if $x_1, \ldots, x_m, y_1, \ldots, y_n$ are irreducible elements of $\mathcal{A}$ such that $x_1 \circ \cdots \circ x_m = y_1 \circ \cdots \circ y_n$, then $m = n$ and for each $i = 1, \ldots, n$, we have $x_i = u_i \circ y_i \circ v_i$ for some invertible elements $u_i, v_i \in \mathcal{A}$. (However, it is easy to see that $\mathcal{A}$ is not a unique factorization monoid in the sense defined in [Johnson 1971].)

Irreducible elements and generating sets

We write $[f]$ for the equivalence class of an element f of $\mathrm{Bir}(\mathbb{A}^2)$.

4.1. Lemma. $\big|\{[f] \mid f \textit{ is an irreducible element of } \mathrm{Bir}(\mathbb{A}^2)\}\big| = |\boldsymbol{k}|$.

Proof. Fix a coordinate system (X, Y) of $\mathbb{A}^2$. For each $a \in \boldsymbol{k}^*$, let $C_a \subset \mathbb{A}^2$ be the zero set of $aY^2(Y-1) + X \in \boldsymbol{k}[X, Y]$.

[5] The question is natural in view of the question of whether SACs are prime and in view of the following trivial fact: let P be a set of prime elements in a commutative and cancellative monoid $\mathcal{N}$, and let $\bar{P}$ be the submonoid of $\mathcal{N}$ generated by P and all invertible elements of $\mathcal{N}$; then $\bar{P}$ is factorially closed in $\mathcal{N}$.

Define $U = \{(a_1, a_2, a_3) \in \boldsymbol{k}^3 \mid a_1, a_2, a_3 \text{ are distinct and nonzero}\}$. Define an equivalence relation $\approx$ on the set U by declaring that $(a_1, a_2, a_3) \approx (b_1, b_2, b_3)$ if and only if there exists $\theta \in \mathrm{Aut}(\mathbb{A}^2)$ satisfying $\theta(C_{a_1} \cup C_{a_2} \cup C_{a_3}) = C_{b_1} \cup C_{b_2} \cup C_{b_3}$. The reader may check[6] that the set $U/\approx$ of equivalence classes has cardinality $|\boldsymbol{k}|$.

Given $q \geq 2$ and distinct elements $a_1, \ldots, a_q \in \boldsymbol{k}^*$, there exists an irreducible element $f \in \mathrm{Bir}(\mathbb{A}^2)$ such that $\mathrm{Miss}(f) = \{C_{a_1}, \ldots, C_{a_q}\}$ and $n(f) = q + 2$ (to see this, set $m = 3$ and $\delta_1 = \cdots = \delta_{q-1} = 0$ in [Daigle 1991a, 4.13][7]). In particular, for each $\boldsymbol{a} = (a_1, a_2, a_3) \in U$ there exists an irreducible $f_{\boldsymbol{a}} \in \mathrm{Bir}(\mathbb{A}^2)$ such that $\mathrm{Miss}(f_{\boldsymbol{a}}) = \{C_{a_1}, C_{a_2}, C_{a_3}\}$. If $\boldsymbol{a}, \boldsymbol{b} \in U$ are such that $f_{\boldsymbol{a}} \sim f_{\boldsymbol{b}}$ then there exist $\theta, \theta' \in \mathrm{Aut}(\mathbb{A}^2)$ satisfying $\theta \circ f_{\boldsymbol{a}} = f_{\boldsymbol{b}} \circ \theta'$; then $\theta(C_{a_1} \cup C_{a_2} \cup C_{a_3}) = C_{b_1} \cup C_{b_2} \cup C_{b_3}$, so $\boldsymbol{a} \approx \boldsymbol{b}$. By the preceding paragraph we get $\big|\{[f_{\boldsymbol{a}}] \mid \boldsymbol{a} \in U\}\big| = |\boldsymbol{k}|$, from which the desired conclusion follows. □

4.2. Lemma. *For any subset S of* $\mathrm{Bir}(\mathbb{A}^2)$, *the following are equivalent*:

(i) $\mathrm{Aut}(\mathbb{A}^2) \cup S$ *is a generating set for the monoid* $\mathrm{Bir}(\mathbb{A}^2)$.

(ii) *For each irreducible* $f \in \mathrm{Bir}(\mathbb{A}^2)$, *we have* $[f] \cap S \neq \varnothing$.

Proof. Suppose that S satisfies (i) and consider an irreducible $f \in \mathrm{Bir}(\mathbb{A}^2)$. By (i),

$$f = g_1 \circ \cdots \circ g_n \quad \text{for some finite subset } \{g_1, \ldots, g_n\} \text{ of } \mathrm{Aut}(\mathbb{A}^2) \cup S.$$

By irreducibility of f, exactly one element g_i of $\{g_1, \ldots, g_n\}$ is not in $\mathrm{Aut}(\mathbb{A}^2)$ (consequently, $g_i \in S$). So $f \sim g_i \in S$, which proves that S satisfies (ii).

Conversely, suppose that (ii) holds and consider $h \in \mathrm{Bir}(\mathbb{A}^2)$; we claim that

$$h = g_1 \circ \cdots \circ g_N \quad \text{for some finite subset } \{g_1, \ldots, g_N\} \text{ of } \mathrm{Aut}(\mathbb{A}^2) \cup S.$$

This is clear if $h \in \mathrm{Aut}(\mathbb{A}^2)$, so assume that $h \notin \mathrm{Aut}(\mathbb{A}^2)$. Then $h = f_1 \circ \ldots \circ f_n$ for some finite collection $\{f_1, \ldots, f_n\}$ of irreducible elements of $\mathrm{Bir}(\mathbb{A}^2)$ (existence of a factorization into irreducibles is a consequence of Lemma 2.5). For each $i \in \{1, \ldots, n\}$, we have $[f_i] \cap S \neq \varnothing$, so $f_i = u_i \circ s_i \circ v_i$ for some $s_i \in S$ and $u_i, v_i \in \mathrm{Aut}(\mathbb{A}^2)$. Then

$$h = (u_1 \circ s_1 \circ v_1) \circ \cdots \circ (u_n \circ s_n \circ v_n) = g_1 \circ \cdots \circ g_N,$$

where $\{g_1, \ldots, g_N\} \subset \mathrm{Aut}(\mathbb{A}^2) \cup S$. This proves (i). □

4.3. Corollary. *Let S be a subset of* $\mathrm{Bir}(\mathbb{A}^2)$ *such that* $\mathrm{Aut}(\mathbb{A}^2) \cup S$ *is a generating set for the monoid* $\mathrm{Bir}(\mathbb{A}^2)$. *Then* $|S| = |\boldsymbol{k}|$.

Proof. Follows from Lemmas 4.1 and 4.2. □

[6] This is a tedious exercise. We leave it to the reader because it is completely elementary and has nothing to do with the subject matter of this paper.

[7] Note that in Example 4.13 of [Daigle 1991a] one has $\mathrm{Miss}(f) = \{C_1, \ldots, C_q\}$. This doesn't seem to be stated explicitly, but it is clear if one reads the construction.

4.4. Remark. Let $f \in \mathrm{Bir}(\mathbb{A}^2)$ and let $\gamma = (X, Y)$ be a coordinate system of $\mathbb{A}^2$. Then $f : \mathbb{A}^2 \to \mathbb{A}^2$ is given by $f(x, y) = (u(x, y), v(x, y))$ for some polynomials $u, v \in \boldsymbol{k}[X, Y]$. We define $\deg_\gamma f = \max(\deg_\gamma u, \deg_\gamma v)$. We may also define $\deg f$ to be the minimum of $\deg_\gamma f$ for γ ranging over the set of coordinate systems of $\mathbb{A}^2$. Then

$$\deg f \geq \frac{c(f)+2}{2}. \tag{41}$$

Indeed, if $F_1, \ldots, F_c \in \boldsymbol{k}[X, Y]$ are irreducible polynomials whose zero sets are the contracting curves of f (so $c(f) = c$), then the jacobian determinant of (u, v) with respect to (X, Y) is divisible by $\prod_{i=1}^{c} F_i$. This implies that $\deg_\gamma f \geq (c+2)/2$, where the right-hand side is independent of γ. Statement (41) follows.

4.5. Corollary.[8] *Let S be a subset of* $\mathrm{Bir}(\mathbb{A}^2)$ *such that* $\mathrm{Aut}(\mathbb{A}^2) \cup S$ *is a generating set for the monoid* $\mathrm{Bir}(\mathbb{A}^2)$. *Then* $\{\deg f \mid f \in S\}$ *is not bounded.*

Proof. Let $n \in \mathbb{N}$. By [Daigle 1991a, 4.13], there exists an irreducible element $g \in \mathrm{Bir}(\mathbb{A}^2)$ satisfying $c(g) \geq 2n$. By 4.2, there exists $f \in S$ satisfying $f \sim g$; then $c(f) = c(g) \geq 2n$, so $\deg f > n$ by (41). □

Factorial closedness of $\mathcal{A}$ in $\mathrm{Bir}(\mathbb{A}^2)$

Let $\mathcal{A}$ be the submonoid of $\mathrm{Bir}(\mathbb{A}^2)$ generated by SACs and automorphisms.

See Definition 2.19 for the definition of $n(f, C)$, where $f \in \mathrm{Bir}(\mathbb{A}^2)$ and $C \in \mathrm{Miss}(f)$.

4.6. Lemma. *Consider* $\mathbb{A}^2 \xrightarrow{\alpha} \mathbb{A}^2 \xrightarrow{f} \mathbb{A}^2$, *where* $\alpha, f \in \mathrm{Bir}(\mathbb{A}^2)$ *and* α *is a SAC. Assume that the missing curve C of α is disjoint from* $\mathrm{exc}(f)$ *and let D be the closure of $f(C)$ in $\mathbb{A}^2$. Then there exist a SAC α' and some $f' \in \mathrm{Bir}(\mathbb{A}^2)$ satisfying $f \circ \alpha = \alpha' \circ f'$ and* $\mathrm{Miss}(\alpha') = \{D\}$. *Moreover, if f is a SAC then so is f'.*

Proof. By Lemma 2.21(b), we have $D \in \mathrm{Miss}(f \circ \alpha)$ and $n(f \circ \alpha, D) = n(\alpha, C) = 1$ (because $C \cap \mathrm{exc}(f) = \varnothing$ and $C \cong \mathbb{A}^1$). Let P be the unique fundamental point of $f \circ \alpha$ which lies on D and let α' be a SAC with missing curve D and fundamental point P. Then Lemma 3.5(a) implies that $f \circ \alpha = \alpha' \circ f'$ for some $f' \in \mathrm{Bir}(\mathbb{A}^2)$. Then $n(f') = n(f)$, so if f is a SAC, then so is f'. □

4.7. Definition. Let $h \in \mathrm{Bir}(\mathbb{A}^2)$ be such that $h \notin \mathrm{Aut}(\mathbb{A}^2)$. Let $C \in \mathrm{Miss}(h)$.

(a) A *factorization* of h is a tuple $\mathfrak{f} = (h_1, \ldots, h_n)$ of elements of $\mathrm{Bir}(\mathbb{A}^2)$ satisfying $h = h_1 \circ \cdots \circ h_n$ (where $n \geq 1$). If $h_1, \ldots, h_n$ are SACs, we say that $\mathfrak{f}$ is a *factorization of h into SACs.*

[8] This result answers a question posed by Patrick Popescu-Pampu.

(b) Given a factorization $\mathfrak{f} = (h_1, \dots, h_n)$ of h, we define $\operatorname{depth}_{\mathfrak{f}}(h, C)$ to be the unique $i \in \{1, \dots, n\}$ satisfying

there exists a missing curve of h_i whose image by $h_1 \circ \cdots \circ h_{i-1}$ is a dense subset of C.

Observe that $\operatorname{depth}_{\mathfrak{f}}(h, C) \geq 1$ and that $\operatorname{depth}_{\mathfrak{f}}(h, C) = 1 \iff C \in \operatorname{Miss}(h_1)$.

(c) If $h \in \mathcal{A}$, then we define

$$\operatorname{depth}(h, C) = \min\{\operatorname{depth}_{\mathfrak{f}}(h, C) \mid \mathfrak{f} \text{ is a factorization of } h \text{ into SACs}\}.$$

Note that $\operatorname{depth}(h, C) \geq 1$ and that $\operatorname{depth}(h, C) = 1$ is equivalent to the existence of SACs $\alpha_1, \dots, \alpha_n$ satisfying

$$h = \alpha_1 \circ \cdots \circ \alpha_n \quad \text{and} \quad \operatorname{Miss}(\alpha_1) = \{C\}.$$

4.8. Theorem. *If $f, g \in \operatorname{Bir}(\mathbb{A}^2)$ satisfy $g \circ f \in \mathcal{A}$, then $f, g \in \mathcal{A}$.*

Proof. We proceed by induction on $n(g \circ f)$, the result being trivial for $n(g \circ f) \leq 2$. Let $n \geq 3$ be such that

$$(*) \qquad \text{for all } f, g \in \operatorname{Bir}(\mathbb{A}^2), \quad g \circ f \in \mathcal{A} \quad \text{and} \quad n(g \circ f) < n \implies f, g \in \mathcal{A}.$$

Consider $f, g \in \operatorname{Bir}(\mathbb{A}^2)$ such that $g \circ f \in \mathcal{A}$ and $n(g \circ f) = n$; we have to show that $f, g \in \mathcal{A}$. Since $g \circ f \in \mathcal{A}$, the number $\operatorname{depth}(g \circ f, C)$ is defined for every $C \in \operatorname{Miss}(g \circ f)$. Observe that

(42) there exists $C \in \operatorname{Miss}(g \circ f)$ satisfying $\operatorname{depth}(g \circ f, C) = 1$, and any such C satisfies $n(g \circ f, C) = 1$.

Indeed, for any factorization $g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$ of $g \circ f$ into SACs, the missing curve C of α_1 satisfies $C \in \operatorname{Miss}(g \circ f)$ and $\operatorname{depth}(g \circ f, C) = 1$, so C exists. Given any $C \in \operatorname{Miss}(g \circ f)$ satisfying $\operatorname{depth}(g \circ f, C) = 1$, there exists a factorization $g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$ of $g \circ f$ into SACs satisfying $\operatorname{Miss}(\alpha_1) = \{C\}$; then $n(g \circ f, C) = n(\alpha_1 \circ \cdots \circ \alpha_n, C) = n(\alpha_1, C) = 1$, where the second equality follows from Lemma 2.21(a). This proves (42).

We now proceed to prove that $f, g \in \mathcal{A}$. We first do so in two special cases (numbered 1 and 2) and then in the general case.

Case 1: there exists $C \in \operatorname{Miss}(g)$ such that $\operatorname{depth}(g \circ f, C) = 1$.
Then there exist SACs $\alpha_1, \dots, \alpha_n$ satisfying $g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$ and $\operatorname{Miss}(\alpha_1) = \{C\}$. We note that $n(g, C) = n(g \circ f, C) = 1$, where the first equality follows from Lemma 2.21(a) and the second from (42), and where the assumption $C \in \operatorname{Miss}(g)$ is needed for the first equality. As $n(g, C) = 1$, there is a unique fundamental point P of g lying on C. Consider the fundamental point P_1 of α_1; then Lemma 2.11 implies that P and P_1 are fundamental points of $g \circ f$ (lying on C); as $n(g \circ f, C) = 1$, we have $P = P_1$, so α_1 is a SAC with missing curve C and fundamental point P. By

Lemma 3.5(a), there exists $g' \in \mathrm{Bir}(\mathbb{A}^2)$ such that $g = \alpha_1 \circ g'$. Then $\alpha_1 \circ g' \circ f = g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$; canceling α_1 yields $g' \circ f = \alpha_2 \circ \cdots \circ \alpha_n \in \mathcal{A}$. As $n(g' \circ f) = n-1$, we get g', $f \in \mathcal{A}$ by $(*)$. Then $g = \alpha_1 \circ g' \in \mathcal{A}$ as well, so we are done in Case 1.

Case 2: $n(f) = 1$.

Note that f is a SAC; let C be its missing curve. By (42), we may consider $D_1 \in \mathrm{Miss}(g \circ f)$ satisfying $\mathrm{depth}(g \circ f, D_1) = 1$ and $n(g \circ f, D_1) = 1$.

By Case 1, we may assume that $D_1 \notin \mathrm{Miss}(g)$. Then (by Lemma 2.12) D_1 is the closure of $g(C)$; since $C \cong \mathbb{A}^1$, we have, in fact, $g(C) = D_1$ (every dominant morphism $\mathbb{A}^1 \to C$ is surjective). Let P be the unique fundamental point of $g \circ f$ on D_1 and let $Q \in C$ be the fundamental point of f; then $g(Q) \in D_1$ is a fundamental point of $g \circ f$ by Lemma 2.11, so $g(Q) = P$.

Since $n(g \circ f, D_1) = 1 = n(f, C)$, Lemma 2.21(b) implies that $C \cap \mathrm{exc}(g) = \varnothing$. By Lemma 2.8, g restricts to an isomorphism $\mathbb{A}^2 \setminus \mathrm{exc}(g) \to \mathbb{A}^2 \setminus \Gamma_g$, where Γ_g is the union of all missing curves of g. Since $C \subset \mathbb{A}^2 \setminus \mathrm{exc}(g)$ and $D_1 = g(C)$, it follows that $D_1 \subset \mathbb{A}^2 \setminus \Gamma_g$. Since $P \in D_1 \subset \mathbb{A}^2 \setminus \Gamma_g$ and $\mathrm{cent}(g) \subseteq \Gamma_g$, we have $P \notin \mathrm{cent}(g)$ and hence $n(g, P) = 0$; so Lemma 2.21 gives $n(g \circ f, P) = n(g, P) + \sum_{P' \in \{Q\}} n(f, P') = 1$ and we have shown

$$D_1 \cap \Gamma_g = \varnothing \quad \text{and} \quad n(g \circ f, P) = 1. \tag{43}$$

Since $\mathrm{depth}(g \circ f, D_1) = 1$, we may choose a factorization $g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$ of $g \circ f$ into SACs satisfying $\mathrm{Miss}(\alpha_1) = \{D_1\}$. We have $\mathrm{cent}(\alpha_1) = \{P\}$ because the fundamental point of α_1 is a fundamental point of $g \circ f$ lying on D_1. Write $\mathrm{Cont}(\alpha_1) = \{E_1\}$; then by Lemma 2.21,

$$n(\alpha_1 \circ (\alpha_2 \circ \cdots \circ \alpha_n), P) = n(\alpha_1, P) + \sum_{P' \in E_1} n(\alpha_2 \circ \cdots \circ \alpha_n, P'),$$

where the left-hand side is equal to $n(g \circ f, P) = 1$ by (43). As $n(\alpha_1, P) = 1$, we have $n(\alpha_2 \circ \cdots \circ \alpha_n, P') = 0$ for all $P' \in E_1$, so $\mathrm{cent}(\alpha_2 \circ \cdots \circ \alpha_n) \cap E_1 = \varnothing$ and, in particular, $\mathrm{cent}(\alpha_2) \cap E_1 = \varnothing$. It follows that the missing curve C_2 of α_2 is not equal to E_1 (because $\mathrm{cent}(\alpha_2) \subset C_2$). So the closure of $\alpha_1(C_2)$ in $\mathbb{A}^2$ is a curve D_2 such that

$$D_2 \in \mathrm{Miss}(g \circ f) \setminus \{D_1\} = \mathrm{Miss}(g).$$

Then $D_2 \subseteq \Gamma_g$, so $D_2 \cap D_1 = \varnothing$ by (43). If $C_2 \cap E_1 \neq \varnothing$, then $\alpha_1(C_2) \cap \alpha_1(E_1) \neq \varnothing$, so $P \in D_2$, contradicting $D_2 \cap D_1 = \varnothing$; thus

$$C_2 \cap E_1 = \varnothing.$$

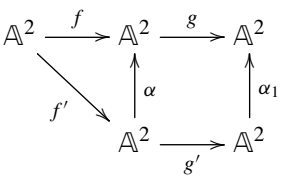
This allows us to use Lemma 4.6. By that result, there exist SACs α_1', α_2' such that $\alpha_1 \circ \alpha_2 = \alpha_1' \circ \alpha_2'$ and $\mathrm{Miss}(\alpha_1') = \{D_2\}$. Since $g \circ f = \alpha_1' \circ \alpha_2' \circ \alpha_3 \circ \cdots \circ \alpha_n$ is a factorization of $g \circ f$ into SACs satisfying $\mathrm{Miss}(\alpha_1') = \{D_2\}$, we have $\mathrm{depth}(g \circ f, D_2) = 1$. Since $D_2 \in \mathrm{Miss}(g)$, Case 1 implies that $f, g \in \mathcal{A}$.

General case. The result is trivial if $n(f) = 0$ and follows from Case 2 if $n(f) = 1$. So we may assume that $n(f) \geq 2$. Consequently, $n(g) \leq n - 2$.

By (42), we may pick $D \in \mathrm{Miss}(g \circ f)$ satisfying $\mathrm{depth}(g \circ f, D) = 1$ and $n(g \circ f, D) = 1$. By Case 1, we may assume that $D \notin \mathrm{Miss}(g)$. Then D is the closure of $g(C)$ for some $C \in \mathrm{Miss}(f)$. We have $1 \leq n(f, C) \leq n(g \circ f, D) = 1$, so $n(f, C) = 1$. Then Lemma 3.5(a) implies that there exist an SAC α and some $f' \in \mathrm{Bir}(\mathbb{A}^2)$ such that $f = \alpha \circ f'$ and $\mathrm{Miss}(\alpha) = \{C\}$. On the other hand, the fact that $\mathrm{depth}(g \circ f, D) = 1$ allows us to choose a factorization $g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$ of $g \circ f$ into SACs satisfying $\mathrm{Miss}(\alpha_1) = \{D\}$. We have $D \in \mathrm{Miss}(g \circ \alpha)$ and

$$n(g \circ \alpha, D) \overset{2.21(a)}{=} n(g \circ \alpha \circ f', D) = n(g \circ f, D) = 1.$$

Let P be the unique fundamental point of $g \circ \alpha$ lying on D; then P is a fundamental point of $g \circ f$ and hence is the unique fundamental point of $g \circ f$ lying on D. As the fundamental point of α_1 is a fundamental point of $g \circ f$ lying on D, it follows that α_1 is a SAC with missing curve D and fundamental point P. Then Lemma 3.5(a) implies that there exists $g' \in \mathrm{Bir}(\mathbb{A}^2)$ satisfying $g \circ \alpha = \alpha_1 \circ g'$.



Since $\alpha_1 \circ g' \circ f' = g \circ f = \alpha_1 \circ \cdots \circ \alpha_n$, canceling α_1 gives $g' \circ f' = \alpha_2 \circ \cdots \circ \alpha_n \in \mathcal{A}$. By $(*)$, we obtain $f', g' \in \mathcal{A}$.

Since $f' \in \mathcal{A}$, it follows that $f = \alpha \circ f' \in \mathcal{A}$.

Since $g' \in \mathcal{A}$, we get $g \circ \alpha = \alpha_1 \circ g' \in \mathcal{A}$; we also have $n(g \circ \alpha) < n$ because $n(g) \leq n - 2$; so $g \in \mathcal{A}$ by $(*)$.

So $f, g \in \mathcal{A}$. □

Acknowledgements

Daigle would like to thank his teacher, Peter Russell, for introducing him to birational endomorphisms of $\mathbb{A}^2$. He would also like to thank the faculty and staff at the Université de Bordeaux 1 for their hospitality. The research leading to this paper was initiated when Daigle spent one month at that institution as a *professeur invité*.

References

[Abhyankar and Moh 1975] S. S. Abhyankar and T. T. Moh, "Embeddings of the line in the plane", *J. Reine Angew. Math.* **276** (1975), 148–166. MR 52 #407 Zbl 0332.14004

[Artal Bartolo and Cassou-Noguès 1996] E. Artal Bartolo and P. Cassou-Noguès, "One remark on polynomials in two variables", *Pacific J. Math.* **176**:2 (1996), 297–309. MR 97m:32044 Zbl 0878.32003

[Cassou-Noguès 2005] P. Cassou-Noguès, "Bad field generators", pp. 77–83 in *Affine algebraic geometry*, edited by J. Gutierrez et al., Contemp. Math. **369**, Amer. Math. Soc., Providence, RI, 2005. MR 2005k:14127 Zbl 1076.14084

[Cassou-Noguès and Daigle 2013] P. Cassou-Noguès and D. Daigle, "Very good and very bad field generators", preprint, 2013. To appear in *Kyoto J. Math.* arXiv 1308.0154

[Cassou-Noguès and Daigle ≥ 2014] P. Cassou-Noguès and D. Daigle, "Lean factorizations of polynomial morphisms". In preparation.

[Cassou-Noguès and Russell 2007] P. Cassou-Noguès and P. Russell, "Birational morphisms $\mathbb{C}^2 \to \mathbb{C}^2$ and affine ruled surfaces", pp. 57–105 in *Affine algebraic geometry*, edited by T. Hibi, Osaka Univ. Press, Osaka, 2007. MR 2008d:14098 Zbl 1222.14027

[Daigle 1991a] D. Daigle, "Birational endomorphisms of the affine plane", *J. Math. Kyoto Univ.* **31**:2 (1991), 329–358. MR 92k:14012 Zbl 0766.14006

[Daigle 1991b] D. Daigle, "Local trees in the theory of affine plane curves", *J. Math. Kyoto Univ.* **31**:3 (1991), 593–634. MR 92h:14023 Zbl 0758.14018

[Daigle 2010] D. Daigle, "Triangular derivations of $\mathbf{k}[X, Y, Z]$", *J. Pure Appl. Algebra* **214**:7 (2010), 1173–1180. MR 2011g:13056 Zbl 1196.14055

[Daigle 2013] D. Daigle, "Generally rational polynomials in two variables", preprint, 2013. To appear in *Osaka J. Math.* **52**:1. arXiv 1307.3752

[Ganong 1985] R. Ganong, "Kodaira dimension of embeddings of the line in the plane", *J. Math. Kyoto Univ.* **25**:4 (1985), 649–657. MR 87c:14013 Zbl 0596.14018

[Ganong 2011] R. Ganong, "The pencil of translates of a line in the plane", pp. 57–71 in *Affine algebraic geometry*, edited by D. Daigle et al., CRM Proc. Lecture Notes **54**, Amer. Math. Soc., Providence, RI, 2011. MR 2768634 Zbl 1263.14057

[Jan 1974] C. J. Jan, *On polynomial generators of* $\boldsymbol{k}(x, y)$, Ph.D. thesis, Purdue University, 1974.

[Johnson 1971] R. E. Johnson, "Unique factorization monoids and domains", *Proc. Amer. Math. Soc.* **28** (1971), 397–404. MR 43 #3186 Zbl 0219.20046

[Kaliman 1992] S. Kaliman, "Two remarks on polynomials in two variables", *Pacific J. Math.* **154**:2 (1992), 285–295. MR 93b:32041 Zbl 0723.32001

[Miyanishi and Sugie 1980] M. Miyanishi and T. Sugie, "Generically rational polynomials", *Osaka J. Math.* **17**:2 (1980), 339–362. MR 82e:14019 Zbl 0457.13006

[Neumann and Norbury 2002] W. D. Neumann and P. Norbury, "Rational polynomials of simple type", *Pacific J. Math.* **204**:1 (2002), 177–207. MR 2003f:14010 Zbl 1055.14062

[Nishino 1968] T. Nishino, "Nouvelles recherches sur les fonctions entiéres de plusieurs variables complexes, I", *J. Math. Kyoto Univ.* **8** (1968), 49–100. MR 38 #3469 Zbl 0165.40302

[Nishino 1969] T. Nishino, "Nouvelles recherches sur les fonctions entières de plusieurs variables complexes, II: Fonctions entières qui se réduisent à celles d'une variable", *J. Math. Kyoto Univ.* **9** (1969), 221–274. MR 41 #502 Zbl 0192.43703

[Nishino 1970] T. Nishino, "Nouvelles recherches sur les fonctions entières de plusiers variables complexes, III: Sur quelques propriétés topologiques des surfaces premières", *J. Math. Kyoto Univ.* **10** (1970), 245–271. MR 42 #533 Zbl 0204.09403

[Russell 1975] P. Russell, "Field generators in two variables", *J. Math. Kyoto Univ.* **15**:3 (1975), 555–571. MR 53 #2946 Zbl 0341.14008

[Russell 1977] P. Russell, "Good and bad field generators", *J. Math. Kyoto Univ.* **17**:2 (1977), 319–331. MR 56 #2977 Zbl 0367.12013

[Saitō 1972] H. Saitō, "Fonctions entières qui se réduisent à certains polynomes, I", *Osaka J. Math.* **9** (1972), 293–332. MR 49 #3187 Zbl 0248.32005

[Saito 1977] H. Saito, "Fonctions entières qui se rédiusent à certains polynômes, II", *Osaka J. Math.* **14**:3 (1977), 649–674. MR 57 #679 Zbl 0372.32004

[Sasao 2006] I. Sasao, "Generically rational polynomials of quasi-simple type", *J. Algebra* **298**:1 (2006), 58–104. MR 2007b:14131 Zbl 1115.14056

[Shpilrain and Yu 2004] V. Shpilrain and J.-T. Yu, "Birational morphisms of the plane", *Proc. Amer. Math. Soc.* **132**:9 (2004), 2511–2515. MR 2005c:14086 Zbl 1060.14086

[Suzuki 1974] M. Suzuki, "Propriétés topologiques des polynômes de deux variables complexes, et automorphismes algébriques de l'espace $\mathbf{C}^2$", *J. Math. Soc. Japan* **26** (1974), 241–257. MR 49 #3188 Zbl 0276.14001

Received July 17, 2013.

PIERRETTE CASSOU-NOGUÈS
INSTITUT DE MATHÉMATIQUES DE BORDEAUX
UNIVERSITÉ DE BORDEAUX
351, COURS DE LA LIBÉRATION
33405 TALENCE CEDEX
FRANCE

pierrette.cassou-nogues@math.u-bordeaux1.fr

DANIEL DAIGLE
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF OTTAWA
OTTAWA K1N 6N5
CANADA

ddaigle@uottawa.ca

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

Founded in 1951 by E. F. Beckenbach (1906–1982) and F. Wolf (1904–1989)

See inside back cover or msp.org/pjm for submission instructions.

The subscription price for 2014 is US $410/year for the electronic version, and $535/year for print and electronic.
Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PJM peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS

Volume 272 No. 2 December 2014