A geometric proof of Jung's Theorem

by Stéphane Lamy

This is a translation of my paper "Une preuve géométrique du théorème de Jung", Enseign. Math. 48 (2002), no. 3-4, 291–315.

1 Introduction

The complex affine space \mathbb{C}^n presents all the qualities of the fascinating mathematical objects: very basic in nature, it stands as the starting point of a vast number of interesting and difficult problems. In particular the group $\operatorname{Aut}[\mathbb{C}^n]$ of polynomial automorphisms of \mathbb{C}^n is far for being well understood. The study of these automorphisms is of course closely related to the research around the famous Jacobian Conjecture (see [12]). But many more questions about the group $\operatorname{Aut}[\mathbb{C}^n]$ are also natural : we would like to determine the finite subgroups, the Lie subgroups, the linearizable subgroups... One can find in [21] a beautiful survey of these problematics. Furthermore it recently became clear that these automorphisms provide some example of dynamical systems with a very rich behavior. A pioneer work in this direction is [13]; for a panorama of the progresses made along the last ten years one can read [35]. All these questions are delicate in general, not to mention the possibility to study what happens when we replace \mathbb{C} by an arbitrary field, or even by a ring. Nevertheless there exists a particular case for which we have a lot of results : the dimension 2 case. Indeed there exists a structure theorem, stated by H.W.E. Jung as early as 1942, that gives a set of generators for $\operatorname{Aut}[\mathbb{C}^2]$.

We will denote by A the group of affine automorphisms of \mathbb{C}^2 , *i.e.* the group of elements of $\operatorname{Aut}[\mathbb{C}^2]$ that extend to holomorphic automorphisms of \mathbb{P}^2 ; and we will name E (for "elementary", following the notations of [13]) the subgroup of $\operatorname{Aut}[\mathbb{C}^2]$ composed of the automorphisms that preserve the pencil of lines y = constant. In other words:

$$A = \{ (x,y) \mapsto (a_1x + b_1y + c_1, a_2x + b_2y + c_2); a_i, b_i, c_i \in \mathbb{C}, a_1b_2 - a_2b_1 \neq 0 \}; \\ E = \{ (x,y) \mapsto (\alpha x + P(y), \beta y + \gamma); \alpha, \beta \in \mathbb{C}^*, \gamma \in \mathbb{C}, P \in \mathbb{C}[X] \}.$$

Theorem 1 (Jung, 1942). The group $\operatorname{Aut}[\mathbb{C}^2]$ of polynomial automorphisms of \mathbb{C}^2 is generated by the affine and elementary automorphisms.

A few years after Jung, this result is made more precise by W. Van der Kulk as follows:

Theorem 2 (Van der Kulk, 1953). Given a field k (of any characteristic, possibly not algebraically closed), the group of polynomial automorphisms of k^2 is generated by the affine et elementary automorphisms with coefficients in k. Furthermore $\operatorname{Aut}[k^2]$ is the amalgamated product of these two subgroups.

After the articles of Jung [19] and Van der Kulk [22] many other proofs, using different technics, have been proposed. The purpose of this paper being to give one another proof, we begin, in order to explain our motivations, by giving a quick survey of the available

proofs in the literature. Generally speaking, the common idea of all this proofs is to proceed by induction on the degree; thus given an automorphism

$$g: (x,y) \mapsto (g_1(x,y), g_2(x,y))$$

where g_1 , g_2 are some polynomials of respective degrees d_1 and d_2 , the point is to prove that it is possible to lower the degree of g by successive compositions by an affine and an elementary automorphisms. More precisely, composing by an affine automorphism we can suppose that d_1 is strictly less than d_2 ; then it remains to prove that the homogeneous component of higher degree of g_1 is a multiple of the one of g_2 , which is quite easily deduced from the fact that d_1 is a multiple of d_2 .

The proof most similar to ours (that is to say, geometric in nature) is maybe the one by M. Nagata [28], who is inspired by the article of Van der Kulk. Previously, W. Engel [11] had proposed a proof, which is took on by A. Gutwirth [17]. Nagata laconically comments on these two proofs by saying that they seem difficult to read for him. Anyway the idea here is to extend g to a birational map of \mathbb{P}^2 and to consider the curve C preimage by gof a general line. We then obtain some informations on the degrees d_1 and d_2 by studying the singularity of C at infinity.

With the aim to give a proof on an arbitrary field, L. Makar-Limanov [23] proposed an alternative, completely algebraic approach to the proof of Van der Kulk. The idea is to introduce a new degree by allowing different weights to the variables x and y, according to the degrees associated with the inverse map of g. Note that few years later this same author propose by a similar approach a description of the automorphism group of a large class of affine surfaces [24]. A proof proposed by W. Dicks in 1983 [9] is a simplified version of the argument of Makar-Limanov; one can find a precise redaction of this proof in the book of P. M. Cohn [8].

A slightly different approach has been proposed by R. Rentschler. It is quite easy once the Jung-Van der Kulk's theorem is established to show that any algebraic representation of $(\mathbb{C}, +)$ in Aut $[\mathbb{C}^2]$ is given up to conjugation by elementary automorphisms. Rentschler takes the converse way : he proposes to show that property first, and then remarks that we can deduce the Jung's theorem from it. Indeed, to the automorphism g we can associate the locally nilpotent derivation $\partial/\partial g_1$. This proof, published as a note to the CRAS in 1968 [31], has been rewritten in details recently by L. M. Drużkowski and J. Gurycz [10].

A so-called elementary proof is published in 1988 by J. H. McKay and S. S. Wang [26]; it is based on an inversion formula. The authors show that the application g^{-1} can be expressed with the help of resultant computations which involve the one variable polynomials $g_1(0,t), g_1(t,0), g_2(0,t), g_2(t,0)$. The expected relation between d_1 and d_2 follows.

On the other hand, one can find an "sophisticated" proof in the book of K. Matsuki [25]. The idea here is to use the framework of Mori theory to formulate a proof of the Jung's theorem, with the hope that this can lead to a breakthrough in the study of the group $\operatorname{Aut}[\mathbb{C}^n]$ for $n \geq 3$.

Let us cite finally one more approach : in [1] S. S. Abhyankar and T. T. Moh prove that two biregular embeddings of \mathbb{C} into \mathbb{C}^2 are the same up to an automorphism of \mathbb{C}^2 , and remark that their proof implies the Jung's theorem. A few authors had proposed some new proofs of this result; one can cite the recent articles of R. V. Gurjar [16], E. Casas-Alvero [6] and E. Artal-Bartolo [3], the three of them proposing some geometric proofs. Nevertheless we will see that if the only aim is to get a proof of the Jung's result then it is possible to give a much more concise geometric proof.

The starting point of our work is a very natural postulate: the Jung's theorem is a result of birational geometry. Indeed any automorphism of \mathbb{C}^2 can be extended as a birational map of \mathbb{P}^2 . In this context it seems to us that to proceed by an induction on the degree is not the more natural reasoning; the number of indeterminacy point turns out to be a more easily managed quantity. This was not really the point of view of Jung, in spite of his title : "On the birational entire transformations of the plane". Nevertheless, in a note that seems to have been forgotten, O.-H. Keller [20] reacts to the work of Jung noticing, without any details, that it should certainly be possible to give a simplified proof using the basic theory of birational maps of \mathbb{P}^2 . Later, in a short article, I.R. Shafarevich [32] states the Jung's theorem and indicates that the demonstration is based on the possibility to write down any birational map between surfaces as a sequence of blow-ups (this is Theorem 6 stated in the next paragraph); unfortunately he does not seem to have ever got the opportunity to publish such a proof (in the complement [34] to his first article Shafarevich merely refers to a paper by M.H. Gizatullin and V.I. Danilov [14] which by its ambition of maximal generality turns out to be quite difficult to read). Finally, recently S. Orevkov [30] hints that we can recover the Jung's theorem from a work of A. G. Vitushkin, but again the details are not made explicit.

We can indeed draw a parallel between the Jung's theorem and a classical result generally attributed to M. Noether [29]:

Theorem 3 (Noether, 1872). Any birational map of the projective plane \mathbb{P}^2 admits a decomposition with the help of linear automorphisms and of the standard quadratic involution

$$\sigma: [x:y:z] \dashrightarrow [yz:xz:xy].$$

It seems that the first complete proof of this statement is in fact due to G. Castelnuovo [7], who deduces the Noether's theorem from the following intermediate result :

Theorem 4 (Castelnuovo, 1901). Any birational map of the projective plane \mathbb{P}^2 can be written as a composition of linear automorphisms and of so-called de Jonquières maps.

Concerning de Jonquières maps let us simply mention that they are the maps of degree n which admit a base point of multiplicity n-1; the noteworthy fact is that the polynomial automorphisms that extend as de Jonquières maps are precisely (up to conjugation by an affine automorphism) the elementary automorphisms. Thus the Jung's theorem may be seen as a special case of the result of Castelnuovo. One can find in [27] a proof of Theorem 3 very closed in spirit to the proof of the Jung's theorem we give in this article. We may wonder that the theorem of Castelnuovo goes back to 1901, whereas those of Jung which turns out to be an easier special case (in particular it will not be necessary to use the notion of multiplicity of an indeterminacy point as Nagata does) goes back to 1942. A possible answer is that neither Castelnuovo nor any of his contemporaries has ever thought about this problem. One can summarize our reasoning saying that we propose ourselves to give the proof of the Jung's theorem as a geometer from the beginning of the twentieth century could have conceived it; or in other words, the proof that seems to be hidden behind the remarks of Keller and Shafarevich cited above. Our proof is concise, does not need any computation, and explains why this result is very particular to the case of the dimension 2. The method being geometric in nature, it has seemed more transparent to us to remain in the classical setting (that is, we work on the field \mathbb{C}); nevertheless this restriction is by no mean essential as we will remark at the end of the article.

The paper is organized as follows.

The second paragraph gathers the results of birational geometry that we use; these are very elementary and you can find them in your favorite introductory book to algebraic geometry (which is probably [15], [18] or [33]).

The proof of Jung's theorem is detailled in the third paragraph.

Finally, in the last paragraph we illustrate our method with an example and then we proceed to prove the theorem of Van der Kulk. First we indicate how to show geometrically

that $\operatorname{Aut}[\mathbb{C}^2]$ is the amalgamated product of his affine and elementary subgroups. We should note that this is really a trivial remark (which certainly turns out to be crucial in practice), and that the really delicate result is the one in the Jung's statement. In conclusion, we show how our proof can be easily adapted to the case of an arbitrary field.

2 Birational maps between surfaces

Our reasoning to prove the Jung's theorem is to consider a polynomial automorphism of \mathbb{C}^2 as a birational map from \mathbb{P}^2 to itself, and then to use a classical structure theorem for these applications. Among many possible choices we have chosen as a reference for this section the first two chapters of [4].

By a surface we always mean a smooth complex projective surface, and an open set is always a Zariski open set. Let X and Y be two surfaces; a rational map $\varphi : X \dashrightarrow Y$ is a morphism from an open set U in X to Y, that can not be extended to a larger open set. When U = X we have a true morphism : we reserve the notation $\varphi : X \to Y$ to this case. It is easy to show (see [4, II.4]) that $X \setminus U$ is a finite set. Thus strictly speaking a rational map is not a map, because there exists a finite number of points where φ is not well defined. Nevertheless the image of a curve is always well defined : if C is a curve in X, we define the strict transform $\varphi(C)$ of C by φ as the adherence of the image by φ of $C \cap U$. Note that the image of a (say, irreducible) curve could be a point.

A birational map between X and Y is a rational map $\varphi : X \dashrightarrow Y$ that induces an isomorphism between an open set of X and an open set of Y.

Example. Consider the following map from \mathbb{P}^2 to itself (that we already met in the statement of Noether's theorem):

$$\sigma: [x:y:z] \dashrightarrow [yz:xz:xy]$$

The map σ , called standard quadratic map, is well defined except at the three points [1:0:0], [0:1:0] and [0:0:1]. Furthermore σ induces an automorphism of \mathbb{P}^2 minus the three lines x = 0, y = 0 and z = 0. We leave to the reader to check for example that the image by σ of the line z = 0 is the point [0:0:1], that the image of a line through [0:0:1] is still a line through [0:0:1], and that the image of a general line is a conic passing through the three points [1:0:0], [0:1:0] and [0:0:1].

A fundamental example of birational map is the blow-up of a point, that we recall briefly. Let S be a surface, and p a point in S. There exists a surface \tilde{S} and a morphism $\pi: \tilde{S} \to S$ such that

- $E = \pi^{-1}(p)$ is isomorphic to \mathbb{P}^1 ;
- π induces an isomorphism from $\tilde{S} \setminus E$ to $S \setminus p$.

Up to isomorphism \tilde{S} and π are unique. As a matter of terminology, we say that π is the blow-up map at the point p, or that \tilde{S} is the blow-up of S at p; the rational curve E is called the exceptional divisor of the blow-up. If $C \subset S$ is a curve through p, we denote by \tilde{C} the strict transform of C, which is the adherence of $\pi^{-1}(C \setminus \{p\})$. By the total transform of C we mean the divisor π^*C ; for instance if C is smooth in p we have $\pi^*C = \tilde{C} + E$.

The surface S is endowed with an intersection form : if D_1 , D_2 are two divisors (*i.e.* some finite sums $\sum \lambda_i C_i$ where the C_i are irreducible, possibly singular, curves, and the λ_i are relative integers), then it is possible to define an intersection number $D_1.D_2$. When D_1 and D_2 are simply two distinct irreducible curves, $D_1.D_2$ is the number of intersection

points of these two curves, counted with multiplicities; $D_1.D_2$ is in this case a positive number. We can extend this natural definition to make sense of the intersection number of two arbitrary divisors, in particular we can speak of the self-intersection of a divisor (see [4, th. I.4]). We note D^2 instead of D.D the self-intersection of a divisor D. Note that the self-intersection of a curve could be negative. The intersection number satisfies the following agreeable properties (D_1 , D_2 and D_3 are three divisors):

- If D_2 and D_3 are linearly equivalent then $D_1.D_2 = D_1.D_3$;
- With the above notations :

$$(\pi^* D_1 . \pi^* D_2) = (D_1 . D_2);$$

(E.\pi^* D_1) = 0.

Regarding the influence of a blow-up on the intersection numbers, we will use repetitively the following equalities that are easily deduced from the properties just stated (C is always a smooth curve through p):

Formulas 5.

$$E^2 = -1;$$

 $\tilde{C}^2 = C^2 - 1.$

Let us make precise a point of vocabulary. According to how we consider the map $\tilde{S} \to S$ we will use two different terminologies : we will say that one goes from S to \tilde{S} by the blow-up of the point p, and that one goes from \tilde{S} to S by the contraction of the curve E. In the sequel we will consider some sequences of blow-ups. If we note π_{p_i} the blow-up map at the point p_i , we will have some maps of the form $\varphi : M \mapsto X$, where M and X are surfaces and $\varphi = \pi_{p_n} \circ \cdots \circ \pi_{p_1}$ (here $p_1 \in X$ and for all $i \geq 2$, p_i belongs to the surface obtained after the blow-ups of the points p_1, \cdots, p_{i-1}). In this situation we will say that p_1 is the first point blown-up by φ , or conversely that the exceptional divisor E_n produced by π_{p_n} is the first curve contracted by φ .

The blow-ups are sufficient to describe any birational maps between surfaces : this is what the following result states precisely (see [4, II.12]).

Theorem 6 (Zariski, 1944). Any birational map between two surfaces is obtained as a sequence of blow-ups followed by a sequence of contractions; in other words if X, Y are some surfaces and

$$q: X \dashrightarrow Y$$

is a birational map (which is not an isomorphism), then there exists a surface M and some sequences of blow-ups π_1 and π_2 such that the following diagram commutes:



Following Beauville we attribute this theorem to Zariski. The proof, which is not too difficult, is made in two steps. In the first step we compose g with a sequence of blow-ups π_1 in order to get rid of the indeterminacy points. Thus we obtain a commutative diagram:



where \bar{g} is a morphism. Note that this process can be applied to any rational map between surfaces (see [4, II.7]); and there exists a similar statement in higher dimension.

On the other hand the second step, which consists to show that the morphism \bar{g} is a sequence of contractions (see [4, II.11]) is very special to the case of a *birational* morphism between two *surfaces*. The heart of the proof is the following proposition:

Proposition 7. Let $\bar{g}: M \mapsto Y$ be a birational morphism between surfaces. If $y \in Y$ is a point where \bar{g}^{-1} is not well-defined, then \bar{g} admits a decomposition



where σ is the blow-up mat at y, and h is a morphism.

We propose a proof of this proposition using an elementary argument from differential geometry, that may throw light on the proof given in [4, II.8]. Note that we do not use at any moment the Castelnuovo criterion (contraction of rational curves with self-intersection -1). We admit the

Lemma 8 (see [4, II.10]). If $\varphi : X \dashrightarrow Y$ is a birational map between two surfaces, and if $x \in X$ is a point where φ is not well defined, then there exists a curve $C \subset Y$ such that $\varphi^{-1}(C) = x$.

Proof of Proposition 7. Suppose that $h = \sigma^{-1} \circ \overline{g}$ is not a morphism, and let $x \in M$ be a point where h is not well defined. In this situation : on one hand $\overline{g}(x) = y$ and \overline{g} is not locally injective at x; on the other hand there exists a curve in \tilde{Y} that is contracted on x by h^{-1} . This curve must be the exceptional divisor E associated with σ . Let p and q be two distinct points on E where h^{-1} is well defined, and let C, C' be two germs of smooth curves transversal to E in p and q respectively. Then $\sigma(C)$ and $\sigma(C')$ are two transverse germs of smooth curves in y, which are image by \overline{g} of two germs of curves in x. The differential of \overline{g} in x is then of rank 2, which contradicts the fact that \overline{g} is not locally injective in x (see figure 1).



Figure 1: h non well defined in x leads to a contradiction.

Before we start with the proof of Jung's theorem we would like to make precise some points of terminology and explain in which context we will use the Zariski's theorem. We will call indeterminacy points of g the points we have to blow-up in the construction of π_1 ; thus these are either points on X or on a surface obtained by blow-ups from X. The indeterminacy points contained in X are said to be proper (classically one says that the other points are in some infinitely near neighborhood of the proper indeterminacy points). The number of indeterminacy points of g (proper or not) will be noted #ind(g).

Remark. Note that this definition is consistent because the sequences of blow-ups π_1 and π_2 produced by the theorem are uniquely determined by g (up to isomorphism). The sequence π_1 is made by blowing-up exactly the points where g is not well-defined. Similarly the sequence π_2 is entirely determined by the points where g^{-1} is not well defined. Of course we could artificially make the sequences π_1 and π_2 longer by blowing-up some points where g and g^{-1} are well defined. However it was implicit in our statement of Zariski's theorem that we consider some minimal sequences π_1 and π_2 , in the sense that we have the following universal property (see [2]):

Let $\varphi_1 : M' \mapsto X$ and $\varphi_2 : M' \mapsto Y$ be two birational morphisms such that $\varphi_2 = g \circ \varphi_1$. Then there exists a unique morphism $h : M' \mapsto M$ such that the following diagram commute :



In the sequel we will use Zariski's theorem only in a very particular case: we will consider $g: X \dashrightarrow \mathbb{P}^2$ coming from a polynomial automorphism of \mathbb{C}^2 . By this we mean that we have a partition $X = \mathbb{C}^2 \cup D$ where D is an union of irreducible curves (called the divisor at infinity), and a partition $\mathbb{P}^2 = \mathbb{C}^2 \cup L$ where L is a line (line at infinity), such that g induces an isomorphism from $X \setminus D$ to $\mathbb{P}^2 \setminus L$. This situation implies some strong restrictions on the indeterminacy points of g; this is what the following lemma makes precise:

Lemma 9. Let X be a surface and g be a birational map from X to \mathbb{P}^2 coming from a polynomial automorphism of \mathbb{C}^2 . We suppose that g is not a morphism. Then

- 1. g admits a unique proper indeterminacy point, located on the divisor at infinity of X;
- 2. g admits some indeterminacy points p_1, \dots, p_s $(s \ge 1)$ such that
 - (a) p_1 is the unique proper indeterminacy point;
 - (b) for all $i = 2, \dots, s$, the point p_i is located on the divisor produced by the blow-up of p_{i-1} ;
- 3. Every irreducible curve contained in the divisor at infinity of X is contracted to a point by g;
- 4. the first contracted curve of π_2 is the strict transform of a curve contained in the divisor at infinity of X;
- 5. in particular, if $X = \mathbb{P}^2$, the first contracted curve by π_2 is the strict transform of the line at infinity in X.

Proof. We know (Lemma 8) that if p is a proper indeterminacy point of g then there exists a curve contracted to p by g^{-1} . In our situation the only curve of \mathbb{P}^2 candidate to be contracted is the line at infinity; so there is at most one proper indeterminacy point for qin X. As we suppose that q is not a morphism, q admits exactly one proper indeterminacy point. The second assertion then comes from a straightforward induction. Furthermore, each curve in the divisor at infinity in X either is contracted to a point, or is sent onto the line at infinity in \mathbb{P}^2 . Since g^{-1} contracts the line at infinity to a point, this latter possibility is excluded : we have shown the third assertion. From the argument above we see that the divisor at infinity in M is composed from the divisor with self-intersection -1 produced by the blow-up of p_s , from the other divisors produced by the sequence of blow-ups, all of which with self-intersection less or equal to -2, and finally from the strict transform of the divisor at infinity in X (here we have used the formulas 5). The first curve contracted by π_2 must have self-intersection -1, and can not be the last curve produced by π_1 (this would contradict the fact that p_s is an indeterminacy point), thus we see that the first curve contracted by π_2 is the strict transform of a curve contained in the divisor at infinity in X. The last assertion is just another formulation of the fourth one, in the case $X = \mathbb{P}^2$.

3 Proof of Jung's theorem

We consider g a polynomial automorphism of \mathbb{C}^2 , that we extend as a birational map (still denoted g) of \mathbb{P}^2 into itself. If g is written

$$g: (x, y) \mapsto (g_1(x, y), g_2(x, y))$$

and if n is the degree of g (*i.e.* the maximum of the degrees of g_1 and g_2), then the extension of g to \mathbb{P}^2 is written in homogeneous coordinates as

$$g: [x:y:z] \dashrightarrow [z^n g_1(x/z,y/z):z^n g_2(x/z,y/z):z^n].$$

The line at infinity in \mathbb{P}^2 is the line of equation z = 0. We want to prove that g is a composition of affine and elementary automorphisms. The proof will proceed by induction on the number #ind(g) of indeterminacy points of g.

By Lemma 9 (assertion 1) the extension $g : \mathbb{P}^2 \dashrightarrow \mathbb{P}^2$ admits a unique indeterminacy point located on the line at infinity. Composing g by an affine automorphism we can assume that this point is [1:0:0]. In other words we have a commutative diagram:



where a is affine and g_0 admits [1:0:0] as indeterminacy point. Obviously we have

$$\# \operatorname{ind}(g_0) = \# \operatorname{ind}(g).$$

We are now going to prove that there exists a diagram



where φ is the extension of an elementary automorphism of \mathbb{C}^2 , and such that

$$\# \operatorname{ind}(g_0 \circ \varphi^{-1}) < \# \operatorname{ind}(g_0)$$

Our reasoning will be to consider the diagram given par Zariski's theorem¹



and to reorganize the blow-ups occurring in the sequences π_1 and π_2 . Thus, in four steps that we will now detail, φ will be constructed from some of the blow-ups of the sequence π_1 and some of the contractions of the sequence π_2 .

First step : blow-up of [1:0:0].

The point [1:0:0] is the first point blown-up in π_1 ; so let us consider the surface F_1 obtained by blowing-up \mathbb{P}^2 at the point [1:0:0]. This surface is a completion of \mathbb{C}^2 which is naturally endowed with a rational fibration coming from the lines y = constant. The divisor at infinity is composed of two rational curves (*i.e.* isomorphic to \mathbb{P}^1) meeting transversally in one point. On one hand we have the strict transform of the line at infinity in \mathbb{P}^2 ; this is a fiber that we will denote f_{∞} . On the other hand we have the exceptional divisor of the blow-up, which is a section for the fibration : it will be denoted s_{∞} . Of course (apply formulas 5) we have $f_{\infty}^2 = 0$ and $s_{\infty}^2 = -1$. More generally for all $n \ge 1$ we will denote by F_n a completion of \mathbb{C}^2 with a structure of a rational fibration, such that the divisor at infinity is composed of two transverse rational curves : one fiber f_{∞} and one section s_{∞} with self-intersection -n. These surfaces are classically called Hirzebruch surfaces; we do not suppose the reader has any particular knowledge of these surfaces. One point about notation : we will write $s_{\infty}(F_n)$ and $f_{\infty}(F_n)$ when more than one Hirzebruch surface will come into play.

Now we come back to the map g_0 . We have a commutative diagram:

$$\begin{array}{c}
F_1 \\
\varphi_1 \\
\varphi_1$$

where φ_1^{-1} is the blow-up map at the point [1:0:0]. we have

$$\#$$
ind $(g_1) = \#$ ind $(g_0) - 1$.

We consider again the diagram obtained from Zariski's theorem applied to g_0 . From Lemma 9 (assertion 5) the first contracted curve in π_2 , which must be a curve with selfintersection -1 in M, is the strict transform of the line at infinity. This is the fiber f_{∞} in F_1 . Now in F_1 we have $f_{\infty}^2 = 0$. The self-intersection of this curve still has to drop by one, thus the indeterminacy point p of g_1 is located on f_{∞} . Furthermore we know (Lemma 9, assertion 2) that this same point p belongs to the curve s_{∞} produced by the blow-up φ_1^{-1} . We conclude that p is precisely the intersection point of f_{∞} and s_{∞} .

Second step : rising induction.

¹ Each time we will use Zariski's theorem we will note M, π_1 and π_2 the surfaces et the sequences of blow-ups produced, the context should allow to avoid any confusion.

In the following reasoning we use some maps between ruled surfaces generally called "elementary transformations" (nevertheless we will not use this terminology in order to avoid any confusion with the elements of the group E). These transformations are defined as the composition of one blow-up and one contraction. More precisely let S be a ruled surface, *i.e.* a surface equipped with a fibration $f: S \to C$ where C is a curve, and such that the fibers of f are all isomorphic to \mathbb{P}^1 . Take $p \in S$ et denote by F the fiber containing p. The elementary transformation at the point p is the birational map obtained as the composition of the blow-up of p (producing an exceptional divisor F') and of the contraction of the strict transform of F. Thus we obtain a new ruled surface S'.



In the proof of Lemmas 10 and 11 we will use such transformations, applied to ruled surfaces with basis C isomorphic to \mathbb{P}^1 .

Lemma 10. Let $n \geq 1$, and let h be a birational map from F_n to \mathbb{P}^2 coming from a polynomial automorphism of \mathbb{C}^2 . Suppose that the unique proper indeterminacy point of h is the intersection point p of $f_{\infty}(F_n)$ and $s_{\infty}(F_n)$. Consider the commutative diagram



where φ is the blow-up of p followed by the contraction of the strict transform of f_{∞} . Then the birational map $h' = h \circ \varphi^{-1}$ satisfies the following two properties:

- 1. #ind(h') = #ind(h) 1;
- 2. the proper indeterminacy point of h' is located on $f_{\infty}(F_{n+1})$.

Proof. We consider the decomposition of h as a sequence of blow-ups:



The (strict) transform of $s_{\infty}(F_n)$ in M has self-intersection less or equal to -2; then Lemma 9 (assertion 4) allows us to conclude that the first contracted curve in π_2 is the transform of $f_{\infty}(F_n)$. Thus the transform of $f_{\infty}(F_n)$ in M has self-intersection -1; on the other hand in F_n we have $f_{\infty}(F_n)^2 = 0$. It follows that after the blow-up of p the rest of the sequence of blow-ups π_1 is performed on points outside of f_{∞} . Instead of doing these blow-ups before contracting the transform of $f_{\infty}(F_n)$ we can reverse the order, that is we can first contract $f_{\infty}(F_n)$ and then realize the rest of the blow-up sequence. In other words we have a commutative diagram (π_p is the blow-up at p and $cont_{f_{\infty}}$ is the contraction of the transform of $f_{\infty}(F_n)$):



These two maps, blow-up of p and then contraction of $f_{\infty}(F_n)$, are summed-up in the following picture where we represent only the divisors at infinity, with their self-intersections (the latter are computed with the help of Formulas 5). We observe in particular that the resulting surface is of type F_{n+1} .



In conclusion, the blow-up p drops by one the number of indeterminacy points, and the contraction of $f_{\infty}(F_n)$ do not create a new one : we have $\# \operatorname{ind}(h') = \# \operatorname{ind}(h) - 1$. Furthermore the indeterminacy point of h' is located on the curve produced blowing-up p, that is $f_{\infty}(F_{n+1})$.

At the end of the first step we are under the conditions of application of Lemma 10, with n = 1. The lemma then gives a map $h' : F_2 \dashrightarrow \mathbb{P}^2$ whose proper indeterminacy point is located on the fiber $f_{\infty}(F_2)$. If this point is precisely the intersection point with the section at infinity, we can apply again the lemma. Iterating this process as long as we remain under the hypotheses of Lemma 10, we obtain a diagram



where φ_2 is obtained by applying n-1 times Lemma 10. Furthermore we have

$$\#$$
ind $(g_2) = \#$ ind $(g_1) - n + 1$.

Finally the indeterminacy point of g_2 is located on $f_{\infty}(F_n)$, and is not the intersection point with $s_{\infty}(F_n)$ (otherwise we could apply the lemma once more).

Third step : descending induction.

We are going to apply the following lemma, which is similar to Lemma 10 (except that now we suppose $n \ge 2$).

Lemma 11. Let $n \geq 2$, and let h be a birational map of F_n to \mathbb{P}^2 that comes from a polynomial automorphisms of \mathbb{C}^2 . Suppose that the unique proper indeterminacy point p of h is located on f_{∞} , but is not the intersection point of f_{∞} and s_{∞} . Consider the commutative diagram



where φ is the blow-up of p followed by the contraction of the strict transform of $f_{\infty}(F_n)$. Then the birational map h' satisfies the following two properties:

- 1. #ind(h') = #ind(h) 1;
- 2. the proper indeterminacy point of h' is located on $f_{\infty}(F_{n-1})$ and is not the intersection point of $f_{\infty}(F_{n-1})$ and $s_{\infty}(F_{n-1})$.

Proof. Consider the decomposition of h given by Zariski's theorem :



The transform of $s_{\infty}(F_n)$ in M has self-intersection -n, since $n \ge 2$ we deduce (lemma 9) that the first contracted curve of π_2 is the transform of $f_{\infty}(F_n)$. As in the proof of lemma 10 we obtain a commutative diagram:



The surface obtained by blowing-up p and contracting the transform of f_{∞} is of type F_{n-1} ; this is summed up in the following picture.

(non proper) indeterminancy point of h corresponding to the proper indeterminancy point of h'



The equality #ind(h') = #ind(h) - 1 is straightforward. Denoting by F' the divisor produced by the blow-up of p, h admits a (non proper) indeterminacy point located on F'. Furthermore this point can not be the intersection point of F' and of the transform

of $f_{\infty}(F_n)$, because otherwise we would have $\pi_1^{-1}(f_{\infty}(F_n))$ with self-intersection less or equal to -2 and this contradicts that it should be the first contracted curve of π_2 . Finally this point is the proper indeterminacy point of h', is located on $f_{\infty}(F_{n-1})$ and is not the intersection point of $f_{\infty}(F_{n-1})$ and $s_{\infty}(F_{n-1})$.

After the second step we are under the hypotheses of Lemma 11. Furthermore if $n \ge 3$ then the map h' produced by the lemma still satisfies the hypotheses of this same lemma. After applying n - 1 times Lemma 11 we obtain a diagram

$$F_{1} \qquad (D3)$$

$$\varphi_{3} \swarrow \varphi_{3} \swarrow \varphi_{3} \qquad (D3)$$

$$F_{n} - - - \frac{1}{g_{2}} - - \rightarrow \mathbb{P}^{2}$$

with

$$\#$$
ind $(g_3) = \#$ ind $(g_2) - n + 1$.

Finally, the proper indeterminacy point of g_3 is located on $f_{\infty}(F_1)$, and is not the intersection point of $f_{\infty}(F_1)$ and $s_{\infty}(F_1)$.

Fourth step : last contraction.

Applying Zariski's theorem to g_3 we get a diagram :



Lemma 9 (assertion 4) ensures that the first contracted curve in π_2 is the strict transform by π_1 of either f_{∞} or s_{∞} . Suppose this is the transform of f_{∞} . Then after the sequence of blow-ups π_1 and the contraction of this curve, the transform of s_{∞} has self-intersection 0 and thus will not be contracted in the sequel of π_2 ; this contradicts the third assertion of Lemma 9. So the first contracted curve must be the transform of s_{∞} , and we can contract the latter straight away to obtain the following diagram:



The morphism φ_4 is the blow-up map with exceptional divisor s_{∞} , and since it is defined up to isomorphism we can decide that the point onto which we contract is [1 : 0 : 0]. Furthermore we have

$$\#$$
ind $(g_3) = \#$ ind (g_4) .

Conclusion.

We can add up the four diagrams $(D1), \dots, (D4)$ into only one



or in more compact form :



with

$$\#$$
ind $(g_4) = \#$ ind $(g_0) - 2n + 1$ (where $n \ge 2$).

We now just have to check that $\varphi = \varphi_4 \circ \varphi_3 \circ \varphi_2 \circ \varphi_1$ is an elementary automorphism. This is equivalent to show that φ preserve the foliation y = constant, that is that φ preserves the pencil of lines passing through [1:0:0]. Now this fact is clear : the blow-up φ_1 sends the lines through [1:0:0] to the fibers of F_1 , φ_2 and φ_3 preserve the fibrations of F_1 and F_n , and finally the contraction φ_4 sends the fibers of F_1 to the lines passing through [1:0:0]. Thus the map g_4 is an automorphism of \mathbb{C}^2 obtained from g by composing first with an affine automorphism and then with an elementary automorphism, and we have the inequality:

$$\# \operatorname{ind}(g_4) < \# \operatorname{ind}(g).$$

By induction on #ind(g), this ends the demonstration.

4 Complements

4.1 An example

Let us consider the following automorphism g:

$$g: (x,y) \mapsto (y+\beta(y+\alpha x^2)^2+\gamma(y+\alpha x^2)^3, y+\alpha x^2)$$
 with $\alpha, \beta, \gamma \in \mathbb{C}^*$.

The decomposition of g is made of two elementary automorphisms :

$$g(x,y) = (x + \beta y^2 + \gamma y^3, y) \circ (y,x) \circ (x + \alpha y^2, y) \circ (y,x).$$

Applying Zariski's theorem to g we decompose g with the help of eight blow-ups and eight contractions. When we apply our algorithm to g we change the order of these blow-ups

and contractions as follows :



We should explain more precisely how we obtain this diagram. The proper indeterminacy point of g is [0:1:0], so we start by considering $g \circ (y, x)$ which is not well-defined at the point [1:0:0]. We blow-up this point, and we apply Lemma 10 once. On the resulting surface F_2 the indeterminacy point is located on a general point of the fiber f_{∞} (the exact location depends on the coefficient α). Then we apply Lemma 11 and we contract the section $s_{\infty}(F_1)$. Then we have the decomposition :

$$g = g' \circ (x + \alpha y^2, y) \circ (y, x)$$

where g' admits only 5 indeterminacy points. Again we consider $g' \circ (y, x)$ to get an automorphism with proper indeterminacy point equal to [1:0:0]. We blow-up this point, and we can apply twice the lemma 10. Then we can apply Lemma 11 twice (the exact location of the indeterminacy points on $f_{\infty}(F_3)$ and $f_{\infty}(F_2)$ depends on the coefficients γ and β). Finally we contract the section $s_{\infty}(F_1)$, and we get

$$g'(x,y) = (x + \beta y^2 + \gamma y^3, y) \circ (y,x).$$

4.2 Amalgamated product structure

We have shown that $\operatorname{Aut}[\mathbb{C}^2]$ is generated by the subgroups A and E, now we want to prove that $\operatorname{Aut}[\mathbb{C}^2]$ is the amalgamated product of these two subgroups. In other words we want to show that any relations in the group $\operatorname{Aut}[\mathbb{C}^2]$ is induced by the relations in the groups A and E. This is equivalent to show that a composition

$$h = a_1 \circ e_1 \circ \cdots \circ a_n \circ e_n$$
 with $a_i \in A \setminus E, e_i \in E \setminus A$

can never be equal to the identity. Note that we can restrict ourselves to such compositions h, that is of even length and beginning by an affine automorphism. Indeed if h is of odd length (and ≥ 3 : if h is of length 1 there is nothing to do) we can reduce the length of h by a suitable conjugation. Furthermore if h is of even length and begins by an elementary automorphism, we just consider h^{-1} instead.

Now it is easy to check that each automorphism e_i , view as a birational map on \mathbb{P}^2 , contracts the line at infinity to the point [1:0:0] (because we suppose $e_i \notin A$). Furthermore, $a_i \notin E$ is equivalent to say that the point [1:0:0] is not a fixed point of a_i . We deduce easily that the extension of h to \mathbb{P}^2 contracts the line at infinity to the point $a_1([1:0:0])$, which contradicts that h is distinct from identity.

4.3 Proof on an arbitrary field

Given a field k we note A_k and E_k the affine and elementary groups with coefficients in k; by \bar{k} we will denote the algebraic cloture of k. A first remark is that our proof works without any modification in the case of an algebraically closed field \bar{k} (the characteristic does not matter). The results on the geometry of surfaces that we use, that is the properties of the intersection form (formulas 5) and the decomposition theorem of Zariski are stated with such level of generality for instance in Chapter V of [18]. Similarly we can copy word to word the above argument to show that $\operatorname{Aut}[\bar{k}^2]$ is the amalgamated product of $A_{\bar{k}}$ and $E_{\bar{k}}$.

Consider now a non algebraically closed field k, and let g be an element of $\operatorname{Aut}[k^2]$ with degree d. We already know that g is a composition of affine et elementary automorphisms with coefficients in \overline{k} . We have to show that there exists such a decomposition with only elements of A_k and E_k .

As above we consider g as a birational map from $\mathbb{P}^2_{\bar{k}}$ to itself. The crucial point is that we know that g admits a *unique* proper indeterminacy point, which is the image by g^{-1} of the line at infinity. Let us choose a point p on the line at infinity all of whose homogeneous coordinates are in k and which is not the indeterminacy point of g^{-1} (one can choose for instance one of the two points [1:0:0] or [0:1:0]). Then $g^{-1}(p)$ is the proper indeterminacy point of g, and is therefore contained in \mathbb{P}^2_k . A similar reasoning shows that the proper indeterminacy point of g^{-1} is also in \mathbb{P}^2_k . Composing g on the right and on the left by well-chosen elements of A_k we can then suppose that the indeterminacy points of g and g^{-1} are both equal to [1:0:0]. That is to say we are in the case where the decomposition of g in the amalgamated product of $A_{\bar{k}}$ and $E_{\bar{k}}$ begins and ends with an elementary automorphism :

$$g = e_n \circ a_{n-1} \circ \cdots \circ a_1 \circ e_1$$
 with $a_i \in A_{\bar{k}} \setminus E_{\bar{k}}, e_j \in E_{\bar{k}} \setminus A_{\bar{k}}$.

A straightforward induction shows that we can write g as

$$g: (x,y) \mapsto (\gamma y^{d_1.d_2} + \cdots, \delta y^{d_1} + \cdots)$$

with $\gamma, \delta \in k^*$ et $d_1, d_2 > 1$ (we wrote only the homogeneous components of higher degree). Composing g on the left by the automorphism $(x, y) \mapsto (x - \frac{\gamma}{\delta^{d_2}} y^{d_2}, y)$ which is an element of E_k we obtain an element of $\operatorname{Aut}[k^2]$ with degree strictly less than the degree of g. By induction on the degree this ends the demonstration.

Note. While this paper was submitted for publication J. F. de Bobadilla very kindly wrote to me to warn me he was the author, independently and simultaneously, of a proof of the Jung's theorem very similar to the one I have exposed here (see chapter 1 of [5]).

Acknowledgements. I had the opportunity to present this work at different stage of its elaboration, successively at the ENS of Lyon, the Federal University of Porto Alegre and the University of Rennes. Each times the numerous remarks and suggestions I received have been of great benefit. I would also want to thanks also to Thierry Vust whose suggestions allowed me to present a more readable article.

References

- S. S. Abhyankar and T. T. Moh. Embeddings of the line in the plane. J. Reine Angew. Math., 276:148–166, 1975.
- M. Alberich-Carramiñana. Geometry of the plane Cremona maps. LNM 1769. Springer-Verlag, 2002.

- [3] E. Artal-Bartolo. Une démonstration géométrique du théorème d'Abhyankar-Moh. J. Reine Angew. Math., 464:97–108, 1995.
- [4] A. Beauville. Surfaces algébriques complexes. Astérisque, 1978.
- [5] J. F. de Bobadilla. Moduli spaces of polynomials in two variables. *Ph. D Thesis*, *Nijmegen*, 2001.
- [6] E. Casas-Alvero. Singularities of plane curves. Cambridge University Press, 2000.
- [7] G. Castelnuovo. Le trasformazioni generatrici del gruppo cremoniano nel piano. *Torino Atti*, 36, 1901.
- [8] P.M. Cohn. Free rings and their relations, second edition. London Mathematical Society Monographs, 19, 1985.
- [9] W. Dicks. Automorphisms of the polynomial ring in two variables. Publ. Sec. Mat. Univ. Autònoma Barcelona, 27:155–162, 1983.
- [10] L. M. Drużkowski and J. Gurycz. An elementary proof of the tameness of polynomial automorphisms of k². Univ. Iagel. Acta Math., 35:251–260, 1997.
- [11] W. Engel. Ganze Cremona Transformationen von Primzahlgrad in der Ebene. Math. Ann., 136:319–325, 1958.
- [12] A. Van den Essen. Polynomial automorphisms and the Jacobian Conjecture. Progress in Mathematics, 190, 2000.
- [13] S. Friedland and J. Milnor. Dynamical properties of plane polynomial automorphisms. Erg. Th. and Dyn. Sys., 9:67–99, 1989.
- [14] M.H. Gizatullin and V.I. Danilov. Automorphisms of affine surfaces I. Math. USSR Izv., 9:493–534, 1975.
- [15] P. Griffiths and J. Harris. Principles of algebraic geometry. Wiley, 1978.
- [16] R. V. Gurjar. A new proof of the Abhyankar-Moh-Suzuki theorem. Transformation groups., 7:61–66, 2002.
- [17] A. Gutwirth. An inequality for certain pencils of plane curves. Proc. Amer. Math. Soc., 12:631–638, 1961.
- [18] R. Hartshorne. Algebraic geometry. Springer, 1977.
- [19] H.W.E. Jung. Über ganze birationale Transformationen der Ebene. J. Reine Angew. Math., 184:161–174, 1942.
- [20] O.-H. Keller. Zu einem Satze von H.W.E. Jung über ganze birationale Transformationen der Ebene. J. Reine Angew. Math., 186:78–79, 1944.
- [21] H. Kraft. Challenging problems on affine *n*-space. Astérisque, 237:295–317, 1996.
- [22] W. Van der Kulk. On polynomial ring in two variables. Nieuw Arch. Wisk., 1:33–41, 1953.
- [23] L.G. Makar-Limanov. On automorphisms of certain algebras. Candidate's dissertation, Moscou, 1970.
- [24] L.G. Makar-Limanov. On group of automorphisms of class of surfaces. Israel J. Math., 69:250–256, 1990.

- [25] K. Matsuki. Introduction to the Mori program. Universitext. Springer, 2002.
- [26] J.H. McKay and S.S. Wang. An elementary proof of the automorphism theorem for the polynomial ring in two variables. J. Pure and Appl. Algebra, 52:91–102, 1988.
- [27] M. Nagata. On rational surfaces I. Mem. Coll. Sci. Kyoto, series A, 32:351–370, 1960.
- [28] M. Nagata. On automorphism group of k[x, y]., volume 5 of Lecture Math. Kyoto. Univ. Kinokuniya, 1972.
- [29] M. Noether. Zur theorie der eindeutigen Ebenentransformationen. Math. Ann., 5:635– 639, 1872.
- [30] S. Y. Orevkov. When a chain of blowups defines an automorphism of C². Math. Notes., 67:541−543, 2000.
- [31] R. Rentschler. Opérations du groupe additif sur le plan affine. C. R. Acad. Sci. Paris Sér. A, 267:384–387, 1968.
- [32] I.R. Shafarevich. On some infinitedimensional groups. Rend. Mat. Appl., 25:208–212, 1966.
- [33] I.R. Shafarevich. *Basic algebraic geometry*. Springer, 1977.
- [34] I.R. Shafarevich. On some infinite-dimensional groups II. Math. USSR Izv., 18:185– 194, 1982.
- [35] N. Sibony. Dynamique des applications rationnelles de \mathbb{P}^k , pages 97–185. SMF Panor. Synthèses. vol. 8, 1999.

Stéphane Lamy
Université Lyon 1
Institut Camille Jordan
21 Avenue Claude Bernard
69622 Villeurbanne Cedex
France
Current email : slamy@math.univ-toulouse.fr
Current web page : https://www.math.univ-toulouse.fr/~slamy/