AUTOMORPHISMS OF A FREE ASSOCIATIVE ALGEBRA OF RANK 2. II

BY

ANASTASIA J. CZERNIAKIEWICZ

ABSTRACT. Let $R$ be a commutative domain with 1. $R(x, y)$ stands for the free associative algebra of rank 2 over $R$; $R[x, y]$ is the polynomial algebra over $R$ in the commuting indeterminates $x$ and $y$.

We prove that the map $\text{Ab}: \text{Aut}(R(x, y)) \rightarrow \text{Aut}(R[x, y])$ induced by the abelianization functor is a monomorphism. As a corollary to this statement and a theorem of Jung [5], Nagata [7] and van der Kulk [8] that describes the automorphisms of $F[x, y]$ ($F$ a field) we are able to conclude that every automorphism of $F(x, y)$ is tame (i.e. a product of elementary automorphisms).

$R$ stands for a commutative domain with 1. $R(x, y)$ is the free associative algebra of rank 2 over $R$ on the free generators $x$ and $y$; $R[x, y]$ is the polynomial algebra over $R$ on the commuting indeterminates $x$ and $y$.

We will prove here that the answer to the following conjecture [3, p. 197] is in the affirmative:

If $F$ is a field then the group of automorphisms of $F(x, y)$ is generated by the elementary automorphisms (defined below) of $F(x, y)$ (i.e. every automorphism of $F(x, y)$ is tame).

In fact, we are going to prove here that, if $R$ is as above, the map $\text{Ab}: \text{Aut}(R(x, y)) \rightarrow \text{Aut}(R[x, y])$ induced by the abelianization functor is a monomorphism and as a consequence of this statement and a theorem of Jung, Nagata and van der Kulk* that says that every automorphism of $F[x, y]$ is tame (for $F$ a field) we will be able to give a complete description of $\text{Aut}(F(x, y))$.

The proof is a generalization of the proof of the main theorem of [4]; in fact the algorithm we use here to solve a system of equations in $R(x, y)$ is essentially the same we used in the previous paper. We will refer to [4] for additional details in the proofs.

I am indebted to G. M. Bergman for making the observation that the tameness result is not true in the generality claimed in our previous paper [4] and announced in the Bulletin of the AMS in November 1971 [Automorphisms of a free associative algebra of rank 2, Bull. Amer. Math. Soc. 77(1971), 992–994], since the corresponding tameness theorem for the abelian case (i.e. the theorem of Jung,

Received by the editors August 5, 1971.

AMS 1970 subject classifications. Primary 16A06, 16A72; Secondary 20F55, 16A02.

Key words and phrases. Free associative algebra, endomorphisms, automorphisms, elementary automorphisms, tame automorphisms, wild automorphisms, polynomial rings, euclidean domains.

*See footnote on page 313.
Nagata and van der Kulk\textsuperscript{*}) is only true for a field (and does not apply to a generalized euclidean domain).

Notation and preliminaries. We will write $R(x, y)$ as a bigraded algebra:

$$R(x, y) = \bigoplus_{r \geq 0, \rho \geq 0} \mathbb{Q}^r$$

where the subindex stands for the homogeneous degree and the upper index denotes the degree in $x$.

For every $P \in R(x, y)$, we write

$$P = \sum_{r, \rho} P^\rho_r$$

uniquely, where $P^\rho_r \in \mathbb{Q}^r$.

The elementary automorphisms of $R(x, y)$ are by definition the following:

(i) $\sigma \in \text{Aut}(R(x, y))$, $\sigma(x) = y$, $\sigma(y) = x$,

(ii) $\phi_{a, \beta} \in \text{Aut}(R(x, y))$, $a, \beta$ units of $R$, $\phi_{a, \beta}(x) = ax$, $\phi_{a, \beta}(y) = \beta y$,

(iii) $\tau_{f(y)}(x) = x + f(y)$, $\tau_{f(y)}(y) = y$.

The same definitions characterize the elementary automorphisms of $R[x, y]$.

Let $E, P, Q \in R(x, y)$. For every $E^\mu$ choose $E^\mu = E^\mu(z_1, \ldots, z_m)$ to be a polynomial in $m$ variables, homogeneous of degree 1 in each and such that

$$E^\mu = E^\mu(x, \ldots, x, y, \ldots, y),$$

where we have put $x = z_i$, $1 \leq i \leq \mu$; $y = z_j$, $\mu + 1 \leq j \leq m$. Even though $E^\mu$ is not uniquely determined by $E^\mu$ we choose it to be zero when the latter is.

We can then write

$$E(P, Q) = \sum_{m, \mu} E^\mu_m(P, Q)$$

$$= \sum_{m, \mu} \sum_{a, a, b, \beta} E^\mu(P_a, \ldots, P_{a, \beta}; Q_b, \ldots, Q_{b, \beta})$$

where $a = (a_1, \ldots, a_\mu)$, $\alpha = (a_1, \ldots, a_\mu)$, $b = (b_1, \ldots, b_{m, \mu})$, $\beta = (\beta_1, \ldots, \beta_{m, \mu})$.

Lemma 1. Let $c, d$ be nonnegative integers, $r, s$ positive integers. Let

$P = p^{rd+1}_{rc+1}$ and $Q = q^{sd}_{sc+1}$ be two algebraically dependent elements of $R(x, y)$ of homogeneous degrees $(rc + 1)$ and $(sc + 1)$ respectively and degrees in $x$ equal to $(rd + 1)$ and $(sd)$ respectively. Then either $P = 0$ or $Q = 0$.

Proof. We simply have to observe that the proofs of Lemmas 1 and 3 of [4] are still valid if instead of assuming that the polynomials commute we allow them to satisfy a nontrivial relation of algebraic dependence.

Main results.

Theorem. Let $P, Q, E \in R(x, y)$ satisfy the following requirements:

(i) $P^0_0 = Q^0_0 = 0$, $E^0_0 = E^0_1 = 0$;

(ii) $P^0_n = 0$ for all $n \geq 1$, $Q^0_m = 0$ for all $m \geq 2$, $E^0_r = 0$ for all $r \geq 2$;

(iii) $E(P, Q) = xy - yx$.

\textsuperscript{*}See footnote on page 313.

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use
Then we conclude that

\[ P = P_1^1 = \alpha x, \quad Q = Q_1^0 + \sum_n Q_n^n = \beta y + f(x), \quad E = (\alpha \beta)^{-1}(xy - yx) \]

where \( \alpha, \beta \) are units of \( R \).

Proof. For every rational number \( \lambda \geq 0 \), define

\[ \delta_\lambda = \{ P^a; \ a > 1, \ a \geq 1, \ (a - 1)/(a - 1) = \lambda \} \cup \{ Q^b; \ b > 1, \ \beta \geq 0, \ \beta/(b - 1) = \lambda \} \]

\[ \cup \{ E_{m^j}; \ m > 2, \ \mu \geq 1, \ (\mu - 1)/(m - 2) = \lambda \}. \]

To prove the assertion of the theorem we only need to prove that \( \delta_\lambda = \{0\} \) for every \( \lambda \).

Since \( \delta_\lambda \) can be different from \( \{0\} \) for at most finitely many values of \( \lambda \) we can use the ordering of the rational numbers to prove inductively that \( \delta_\lambda = \{0\} \) for all \( \lambda \).

For this purpose let \( S = \{ P_1^1, Q_1^0, E_2^1 \} \) and let us set \( \delta_\lambda = \delta_\lambda \cup S \). We then must show that \( \delta_\lambda = \delta \) for all \( \lambda \).

Suppose we have proved \( \delta_\lambda^* = \delta_\lambda \) for every \( \lambda < \lambda \) and let us prove

\[ \delta_\lambda^* = \delta_\lambda \]

i.e. we have to prove that if \( X \in \delta_\lambda^* \) then \( X \in \delta_\lambda \). Observe that we include the case \( \lambda = 0 \) in the inductive process.

Write \( \lambda = \rho/\tau \) where \( \rho \) and \( \tau \) are relatively prime positive integers if \( \lambda > 0 \) or else \( \rho = 0, \ \tau = 1 \) if \( \lambda = 0 \).

With this notation we can write

\[ \delta_\lambda^* = \{ P_1^{b\rho + 1}; \ b \geq 0 \} \cup \{ Q_1^{i\rho + 1}; \ i \geq 0 \} \cup \{ E_2^{j\rho + 1}; \ j \geq 0 \}. \]

Since we have a finite collection of polynomials we can assume that the following holds:

\[ E_{e', \rho' + 1} = 0 \quad \text{if} \quad e' > e, \quad p_{p', \rho' + 1} = 0 \quad \text{if} \quad p' > p, \quad Q_{q', \rho' + 1} = 0 \quad \text{if} \quad q' > q, \]

\[ \neq 0 \quad \text{for} \quad e' = e; \quad \neq 0 \quad \text{if} \quad p' = p; \quad \neq 0 \quad \text{if} \quad q' = q. \]

To obtain assertion (2) it will suffice to show that \( e = p = q = 0 \).

Claim 1. Let \( L = e\rho \rho + eq(r - \rho) + e + q + p \). Then under the inductive hypothesis and conditions (4) the only term of \( E(P, Q) \) that lies in \( Q_{L+2}^{L\rho+1} \) is

\[ E_{e', \rho' + 1}(p_{p', \rho' + 1}, q_{q', \rho' + 1}, \rho_{\rho', \rho' + 1}). \]

Proof of Claim 1. In the notation of (1), a typical summand of \( [E(P, Q)]_{L+2}^{L\rho+1} \) is of the following form:

\[ (1') \quad [E_m^{\mu}(P, Q)]_{L+2}^{L\rho+1} = \sum_{a, b, \beta} E_m^{\mu}(P_1^{a_1}, \ldots, P_1^{a_{\mu}}, Q_1^{b_1}, \ldots, Q_1^{b_{m-\mu}}) \]

where\n
\[ \sum_{j=1}^{\mu} a_j + \sum_{k=1}^{m-\mu} b_k = Lr + 2, \quad \sum_{j=1}^{\mu} a_j + \sum_{k=1}^{m-\mu} \beta_k = L\rho + 1. \]
By inductive hypothesis we know $\delta^{*}_{\lambda'} = \delta$ if $\lambda' < \lambda$, hence we can assume that the following inequalities hold:

$$r(\mu - 1) \geq \rho(m - 2),$$

$$r(\alpha_j - 1) \geq \rho(\alpha_j - 1), \quad 1 \leq j \leq \mu,$$

$$r\beta_k \geq \rho(b_k - 1), \quad 1 \leq k \leq m - \mu.$$  

If we now add the inequalities (6) term by term we obtain

$$r \left( \mu - 1 + \sum_{j=1}^{\mu} \alpha_j - \mu + \sum_{k=1}^{m-\mu} \beta_k \right) \geq \rho \left( m - 2 + \sum_{j=1}^{\mu} \alpha_j - \mu + \sum_{k=1}^{m-\mu} b_k - (m - \mu) \right)$$

which is simply

$$r \left( \sum_{j=1}^{\mu} \alpha_j + \sum_{k=1}^{m-\mu} \beta_k - 1 \right) \geq \rho \left( \sum_{j=1}^{\mu} \alpha_j + \sum_{k=1}^{m-\mu} b_k - 2 \right).$$

The inequality (8) together with (5) yields

$$rLp \geq \rho Lr.$$  

If any of the inequalities in (6) were strict then (9) will also be a strict inequality and we would have reached a contradiction; hence (6) are all equalities. As a consequence $E^{\mu}_{m}$, $P_{\alpha_j}^{\mu}$, $1 \leq j \leq \mu$; $Q_{\beta_k}^{\mu}$, $1 \leq k \leq m - \mu$, are all elements of $\delta^{*}_{\lambda}$ and using (3) we deduce that there are positive integers $e'$; $p, 1 \leq j \leq \mu$; $q, 1 \leq k \leq m - \mu$; so that the following relations are satisfied:

$$\mu = e'p + 1, \quad m = e'r + 2,$$

$$\alpha_j = p_j\rho + 1, \quad a_j = p_j r + 1, \quad 1 \leq j \leq \mu,$$

$$\beta_k = q_k\rho, \quad b_k = q_k r + 1, \quad 1 \leq k \leq m - \mu.$$  

And also we have

$$e' \leq e,$$

$$p_j \leq p, \quad 1 \leq j \leq \mu,$$

$$q_k \leq q, \quad 1 \leq k \leq m - \mu.$$  

Using (10) and (11) we obtain

$$\sum_{j=1}^{\mu} a_j + \sum_{k=1}^{m-\mu} b_k = \sum_{j=1}^{\mu} (p_j r + 1) + \sum_{k=1}^{m-\mu} (q_k r + 1)$$

$$= \sum_{j=1}^{\mu} p_j r + \mu + \sum_{k=1}^{m-\mu} q_k r + (m - \mu)$$

$$\leq p(ep + 1)r + q(e(r - \rho) + 1) + er + 2 = Lr + 2.$$
Consequently, if any of the inequalities of (11) were strict, by comparing (5) and (12) we get a contradiction. Hence (11) must all be equalities. This concludes the proof of Claim 1.

So, in (4), if either $e, p$ or $q$ are nonzero, using (iii) and Claim 1 we obtain

$$[E(P, Q)]_{L_{r+2}} = E_{e_{r+1}}(p_{pr+1}, q_{qr+1}) = 0,$$

which is a nontrivial relation of algebraic dependence for $p_{pr+1}$ and $q_{qr+1}$, and applying Lemma 1 we see that one of them must be zero. But this contradicts the choice of $e, p, q$. Hence $e = p = q = 0$. This concludes the induction and the proof of the theorem.

Corollary 1. The map $\text{Ab}: \text{Aut}(R(x, y)) \rightarrow \text{Aut}(R[x, y])$ induced by the abelianization functor is a monomorphism.

Proof. Let $\phi \in \text{Aut}(R(x, y))$ be such that $\text{Ab}(\phi) = \phi = \text{id}_{R[x, y]}$. Let $\phi(x) = P(x, y)$, $\phi(y) = Q(x, y)$; $\phi^{-1}(x) = A(x, y)$, $\phi^{-1}(y) = B(x, y)$.

Set $E = AB - BA$.

The following equalities hold:

$$A(P(x, y), Q(x, y)) = x, \quad B(P(x, y), Q(x, y)) = y,$$

$$E(P, Q) = xy - yx. \quad (13)$$

If we apply now the abelianization map we obtain

$$\tilde{\phi}(x) = \tilde{P}(\tilde{x}, \tilde{y}) = \tilde{x}, \quad \tilde{\phi}(\tilde{y}) = \tilde{Q}(\tilde{x}, \tilde{y}) = \tilde{y},$$

$$\tilde{A}(\tilde{x}, \tilde{y}) = \tilde{x}, \quad \tilde{B}(\tilde{x}, \tilde{y}) = \tilde{y}, \quad \tilde{E}(\tilde{x}, \tilde{y}) = 0. \quad (14)$$

As a consequence of (14), $P, Q, E$ satisfy the hypothesis of Theorem 1, and we conclude

$$P = \alpha x, \quad Q = \beta y + f(x). \quad (15)$$

But (14) gives $\alpha = \beta = 1$ and also $f(x) = 0$. This shows that $\phi = \text{id}_{R(x, y)}$, therefore completing the proof of Corollary 1.

Corollary 2. If $F$ is a field, then the map $\text{Ab}: \text{Aut}(F(x, y)) \rightarrow \text{Aut}(F[x, y])$ of Corollary 1 is bijective.

Proof. H. Jung [4] proved that every automorphism of $F[x, y]$ is tame when $F$ is a field of characteristic 0 (see also A. Gutwirth [6]) and the same conclusion follows from a theorem of M. Nagata [7] if $F$ is a field of characteristic $p$. (1)

We now observe that given an elementary automorphism $\pi$ of $F[x, y]$ there exists an elementary automorphism of $F(x, y)$, say $\pi^*$, so that $\text{Ab}(\pi^*) = \tilde{\pi} = \pi$.

Since every automorphism of $F[x, y]$ is tame (i.e. a product of elementary

(1) Added in proof. The same result has also been proved independently by W. van der Kulk [8].
automorphisms), the surjectivity (and by Corollary 1 the injectivity) of Ab follows.
A restatement of this corollary gives

Corollary 3. If F is a field then every automorphism of F(x, y) is tame.

Corollary 4. If R is a commutative domain with 1, then every automorphism of R(x, y) keeps [x, y] = xy - yx fixed (up to multiplication by a unit of R).

Proof. We simply have to observe that if \( \phi \) is an automorphism of R(x, y) then \( \phi \) induces an automorphism of F(x, y), where F is the field of fractions of R. Since every tame automorphism keeps [x, y] fixed (up to scalar multiplication), we conclude that \( \phi \) keeps [x, y] fixed up to multiplication by an element of F, say \( a \).

Since the same reasoning applies to \( \phi^{-1} \) we are able to conclude that in fact \( a \) is a unit of R.

Remarks. 1. The following example due to G. M. Bergman shows that Corollary 3 is not true if R is not a field.

One first shows that if \( \phi \) is a tame automorphism of R(x, y) (or of R[x, y]), then if \( \deg \phi(x) > \deg \phi(y) \), the highest degree component of \( \phi(x) \) is a power of that of \( \phi(y) \), times an element of R. We omit the details here.

Let \( c \) be a nonzero nonunit of R. We shall construct a tame automorphism of the free algebra over R[\( c^{-1} \)] such that all coefficients in \( \phi \) and \( \phi^{-1} \) lie in R, but such that the highest degree component of \( \phi(x) \) is a power of that of \( \phi(y) \) times \( c^{-1} \). Thus, \( \phi \) induces an automorphism of the polynomial ring in x and y over R, but this cannot be tame. \( \phi \) is obtained in the following way:

Define automorphisms

\[ \alpha(x) = x + c^{-1}y^2, \quad \alpha(y) = y; \]
\[ \beta(x) = x, \quad \beta(y) = y + c^3x^2; \]

and let \( \phi = \alpha \beta \alpha^{-1} \). Then \( \phi^{\pm 1} = \alpha \beta^{\pm 1} \alpha^{-1} \) is given by

\[ \phi^{\pm 1}(x) = (x + c^{-1}y^2) - c^{-1}(y \pm c^3(x + c^{-1}y^2))^2, \]
\[ \phi^{\pm 1}(y) = y \pm c^3(x + c^{-1}y^2)^2. \]

Note that the expression \( c^3(x + c^{-1}y^2)^2 \) reduces to \( c(cx + y^2)^2 \), while in the expression for \( \phi^{\pm 1}(x) \), terms \( c^{-1}y^2 - c^{-1}y^2 \) cancel; we find

\[ \phi^{\pm 1}(x) = x \pm (c(x + y^2)^2 + (cx + y^2)^2y) + c(cx + y^2)^4, \]
\[ \phi^{\pm 1}(y) = y \pm c(cx + y^2)^2. \]

Thus, \( \phi \) has the properties claimed.

2. One would like to know for what classes of rings is Corollary 2 true. There is a counterexample to it, also due to G. M. Bergman, involving an R that
is not separably closed in its integral closure.

3. With respect to Corollary 4, there is a related conjecture (see [3, p. 197]): is it true that every endomorphism of \( R(x, y) \) that keeps \([x, y]\) fixed (up to multiplication by a unit of \( R \)) is an automorphism of \( R(x, y) \)?

We give an affirmative answer to it, under very restrictive conditions in the following:

**Corollary 5.** Let \( \phi \) be an endomorphism of \( R(x, y) \) such that \( \phi(x) = P \), \( \phi(y) = Q \), \([P, Q] = \lambda[x, y] \), \( \lambda \) a unit of \( R \). Assume further that conditions (i) and (ii) of Theorem 1 are satisfied. Then \( \phi \) is indeed an automorphism of \( R(x, y) \).

The proof is an immediate consequence of Theorem 1, by taking \( E(x, y) = \lambda^{-1}[x, y] \).

**REFERENCES**

3. ——, *Lecture notes on free rings*, Yale University, New Haven, Conn., 1962.

(2) Added in proof.