TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 171, September 1972

AUTOMORPHISMS OF A FREE ASSOCIATIVE ALGEBRA OF RANK 2. II

BY

ANASTASIA J. CZERNIAKIEWICZ

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

We prove that the map Ab: $\operatorname{Aut}(R\langle x, y\rangle) \longrightarrow \operatorname{Aut}(R[\tilde{x}, \tilde{y}])$ 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 $F[\tilde{x}, \tilde{y}]$ (F a field) we are able to conclude that every automorphism of $F\langle x, y\rangle$ is tame (i.e. a product of elementary automorphisms).

R stands for a commutative domain with 1. $R\langle x, y\rangle$ is the free associative algebra of rank 2 over R on the free generators x and y; $R[\tilde{x}, \tilde{y}]$ is the polynomial algebra over R on the commuting indeterminates $\tilde{x}$ and $\tilde{y}$.

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 $F\langle x, y\rangle$ is generated by the elementary automorphisms (defined below) of $F\langle x, y\rangle$ (i.e. every automorphism of $F\langle x, y\rangle$ is tame).

In fact, we are going to prove here that, if R is as above, the map Ab: $\operatorname{Aut}(R\langle x, y\rangle) \longrightarrow \operatorname{Aut}(R[\tilde{x}, \tilde{y}])$ 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 $F[\tilde{x}, \tilde{y}]$ is tame (for F a field) we will be able to give a complete description of Aut $(F\langle x, y\rangle)$.

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 $R\langle x, y\rangle$ 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*) is only true for a field (and does not apply to a generalized euclidean domain).

Notation and preliminaries. We will write $R\langle x, y\rangle$ as a bigraded algebra:

$$R\langle x, y\rangle = \bigoplus_{r \geq \rho \geq 0} \mathfrak{A}_r^\rho$$

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

For every $P \in R\langle x, y\rangle$, we write

$$P = \sum_{r,\rho} P_r^\rho \quad \text{uniquely,} \quad \text{where } P_r^\rho \in \mathfrak{A}_r^\rho.$$

The *elementary automorphisms* of $R\langle x, y\rangle$ are by definition the following:

(i) $\sigma \in \operatorname{Aut}(R\langle x, y\rangle)$, $\sigma(x) = y$, $\sigma(y) = x$,

(ii) $\phi_{\alpha,\beta} \in \operatorname{Aut}(R\langle x, y\rangle)$, α, β units of R, $\phi_{\alpha,\beta}(x) = \alpha x$, $\phi_{\alpha,\beta}(y) = \beta y$,

(iii) $\tau_{f(y)} \in \operatorname{Aut}(R\langle x, y\rangle)$, where $f(y)$ is any polynomial that does not depend on x; $\tau_{f(y)}(x) = x + f(y)$, $\tau_{f(y)}(y) = y$.

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

Let $E, P, Q \in R\langle x, y\rangle$. For every E_m^μ choose $*E_m^\mu = *E_m^\mu(z_1, \cdots, z_m)$ to be a polynomial in m variables, homogeneous of degree 1 in each and such that $E_m^\mu = *E_m^\mu(x, \cdots, x, y, \cdots, y)$, where we have put $x = z_i$, $1 \leq i \leq \mu$; $y = z_j$, $\mu + 1 \leq j \leq m$. Even though $*E_m^\mu$ is not uniquely determined by E_m^μ we choose it to be zero when the latter is.

We can then write

$$\begin{aligned} E(P, Q) &= \sum_{m,\mu} E_m^\mu(P, Q) \\ &= \sum_{m,\mu} \sum_{a,\alpha,b,\beta} *E_m^\mu(P_{a_1}^{\alpha_1}, \cdots, P_{a_\mu}^{\alpha_\mu}, Q_{b_1}^{\beta_1}, \cdots, Q_{b_{m-\mu}}^{\beta_{m-\mu}}) \end{aligned} \tag{1}$$

where $a = (a_1, \cdots, a_\mu)$, $\alpha = (\alpha_1, \cdots, \alpha_\mu)$, $b = (b_1, \cdots, b_{m-\mu})$, $\beta = (\beta_1, \cdots, \beta_{m-\mu})$.

Lemma 1. *Let c, d be nonnegative integers, r, s positive integers. Let $P = P_{rc+1}^{rd+1}$ and $Q = Q_{sc+1}^{sd}$ be two algebraically dependent elements of $R\langle x, y\rangle$ 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 $P, Q, E \in R\langle x, y\rangle$ satisfy the following requirements:*

(i) $P_0^0 = Q_0^0 = 0$, $E_0 = E_1 = 0$;

(ii) $P_n^0 = 0$ *for all* $n \geq 1$, $Q_m^0 = 0$ *for all* $m \geq 2$, $E_r^0 = 0$ *for all* $r \geq 2$;

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

---

*See footnote on page 313.

Then we conclude that

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

where α, β *are units of* R.

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

$$\mathcal{S}_\lambda = \{P_a^\alpha;\ a > 1,\ \alpha \geq 1,\ (\alpha - 1)/(a - 1) = \lambda\} \cup \{Q_b^\beta;\ b > 1,\ \beta \geq 0,\ \beta/(b - 1) = \lambda\}$$
$$\cup \{E_m^\mu;\ m > 2,\ \mu \geq 1,\ (\mu - 1)/(m - 2) = \lambda\}.$$

To prove the assertion of the theorem we only need to prove that $\mathcal{S}_\lambda = \{0\}$ for every λ.

Since $\mathcal{S}_\lambda$ can be different from $\{0\}$ for at most finitely many values of λ we can use the ordering of the rational numbers to prove inductively that $\mathcal{S}_\lambda = \{0\}$ for all λ.

For this purpose let $\mathcal{S} = \{P_1^1, Q_1^0, E_2^1\}$ and let us set $\mathcal{S}_\lambda^* = \mathcal{S}_\lambda \cup \mathcal{S}$. We then must show that $\mathcal{S}_\lambda^* = \mathcal{S}$ for all λ.

Suppose we have proved $\mathcal{S}_{\lambda'}^* = \mathcal{S}$ for every $\lambda' < \lambda$ and let us prove

$$\mathcal{S}_\lambda^* = \mathcal{S}, \tag{2}$$

i.e. we have to prove that if $X \in \mathcal{S}_\lambda^*$ then $X \in \mathcal{S}$. Observe that we include the case $\lambda = 0$ in the inductive process.

Write $\lambda = \rho/r$ where ρ and r are relatively prime positive integers if $\lambda > 0$ or else $\rho = 0$, $r = 1$ if $\lambda = 0$.

With this notation we can write

$$\mathcal{S}_\lambda^* = \{P_{hr+1}^{h\rho+1};\ h \geq 0\} \cup \{Q_{ir+1}^{i\rho};\ i \geq 0\} \cup \{E_{jr+2}^{j\rho+1};\ j \geq 0\}. \tag{3}$$

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

$$\begin{array}{llllll} E_{e'r+2}^{e'\rho+1} = 0 & \text{if } e' > e, & P_{p'r+1}^{p'\rho+1} = 0 & \text{if } p' > p, & Q_{q'r+1}^{q'\rho} = 0 & \text{if } q' > q, \\ \neq 0 & \text{for } e' = e; & \neq 0 & \text{if } p' = p; & \neq 0 & \text{if } q' = q. \end{array} \tag{4}$$

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

Claim 1. Let $L = ep\rho + eq(r - \rho) + e + q + p$. Then under the inductive hypothesis and conditions (4) the only term of $E(P, Q)$ that lies in $\mathfrak{A}_{Lr+2}^{L\rho+1}$ is $E_{er+2}^{e\rho+1}(P_{pr+1}^{p\rho+1}, Q_{qr+1}^{q\rho})$.

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

$$[E_m^\mu(P, Q)]_{Lr+2}^{L\rho+1} = \sum_{a,\alpha,b,\beta} {}^*E_m^\mu(P_{a_1}^{\alpha_1}, \ldots, P_{a_\mu}^{\alpha_\mu}, Q_{b_1}^{\beta_1}, \ldots, Q_{b_{m-\mu}}^{\beta_{m-\mu}}) \tag{1'}$$

where

$$\sum_{j=1}^{\mu} a_j + \sum_{k=1}^{m-\mu} b_k = Lr + 2, \qquad \sum_{j=1}^{\mu} \alpha_j + \sum_{k=1}^{m-\mu} \beta_k = L\rho + 1. \tag{5}$$

(In (1′) $^*E_m^\mu$ is a polynomial in m variables, homogeneous of degree 1 in each and $^*E_m^\mu = 0$ whenever $E_m^\mu = 0$.)

By inductive hypothesis we know $\mathcal{S}^*_{\lambda'} = \mathcal{S}$ if $\lambda' < \lambda$, hence we can assume that the following inequalities hold:

$$\begin{aligned} r(\mu - 1) &\geq \rho(m-2), \\ r(\alpha_j - 1) &\geq \rho(a_j - 1), \qquad 1 \leq j \leq \mu, \\ r\beta_k &\geq \rho(b_k - 1), \qquad 1 \leq k \leq m - \mu. \end{aligned} \tag{6}$$

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

$$r\left(\mu - 1 + \sum_{j=1}^{\mu} \alpha_j - \mu + \sum_{k=1}^{m-\mu} \beta_k\right) \geq \rho\left(m - 2 + \sum_{j=1}^{\mu} a_j - \mu + \sum_{k=1}^{m-\mu} b_k - (m-\mu)\right) \tag{7}$$

which is simply

$$r\left(\sum_{j=1}^{\mu} \alpha_j + \sum_{k=1}^{m-\mu} \beta_k - 1\right) \geq \rho\left(\sum_{j=1}^{\mu} a_j + \sum_{k=1}^{m-\mu} b_k - 2\right). \tag{8}$$

The inequality (8) together with (5) yields

$$rL\rho \geq \rho Lr. \tag{9}$$

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_m^μ; $P_{a_j}^{\alpha_j}$, $1 \leq j \leq \mu$; $Q_{b_k}^{\beta_k}$, $1 \leq k \leq m - \mu$, are all elements of $\mathcal{S}^*_\lambda$ and using (3) we deduce that there are positive integers e'; p_j; $1 \leq j \leq \mu$; q_k, $1 \leq k \leq m - \mu$; so that the following relations are satisfied:

$$\begin{aligned} \mu &= e'\rho + 1, & m &= e'r + 2, \\ \alpha_j &= p_j\rho + 1, & a_j &= p_jr + 1, \qquad 1 \leq j \leq \mu, \\ \beta_k &= q_k\rho, & b_k &= q_kr + 1, \qquad 1 \leq k \leq m - \mu. \end{aligned} \tag{10}$$

And also we have

$$\begin{aligned} e' &\leq e, \\ p_j &\leq p, \qquad 1 \leq j \leq \mu, \\ q_k &\leq q, \qquad 1 \leq k \leq m - \mu. \end{aligned} \tag{11}$$

Using (10) and (11) we obtain

$$\begin{aligned} \sum_{j=1}^{\mu} a_j + \sum_{k=1}^{m-\mu} b_k &= \sum_{j=1}^{\mu} (p_jr + 1) + \sum_{k=1}^{m-\mu} (q_kr + 1) \\ &= \sum_{j=1}^{\mu} p_jr + \mu + \sum_{k=1}^{m-\mu} q_kr + (m - \mu) \\ &\leq p(e\rho + 1)r + q(e(r - \rho) + 1) + er + 2 = Lr + 2. \end{aligned} \tag{12}$$

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

$$[E(P, Q)]_{Lr+2}^{L\rho+1} = E_{er+2}^{e\rho+1}(P_{pr+1}^{p\rho+1}, Q_{qr+1}^{q\rho}) = 0,$$

which is a nontrivial relation of algebraic dependence for $P_{pr+1}^{p\rho+1}$ and $Q_{qr+1}^{q\rho}$, 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* Ab: $\mathrm{Aut}(R\langle x, y\rangle) \longrightarrow \mathrm{Aut}(R[\tilde{x}, \tilde{y}])$ *induced by the abelianization functor is a monomorphism.*

Proof. Let $\phi \in \mathrm{Aut}(R\langle x, y\rangle)$ be such that $\mathrm{Ab}(\phi) = \tilde{\phi} = \mathrm{id}_{R[\tilde{x}, \tilde{y}]}$. 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:

$$A(P(x, y), Q(x, y)) = x, \qquad B(P(x, y), Q(x, y)) = y, \qquad E(P, Q) = xy - yx. \tag{13}$$

If we apply now the abelianization map we obtain

$$\begin{gathered}\tilde{\phi}(\tilde{x}) = \tilde{P}(\tilde{x}, \tilde{y}) = \tilde{x}, \qquad \tilde{\phi}(\tilde{y}) = \tilde{Q}(\tilde{x}, \tilde{y}) = \tilde{y}, \\ \tilde{A}(\tilde{x}, \tilde{y}) = \tilde{x}, \qquad \tilde{B}(\tilde{x}, \tilde{y}) = \tilde{y}, \qquad \tilde{E}(\tilde{x}, \tilde{y}) = 0.\end{gathered} \tag{14}$$

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

$$P = \alpha x, \qquad Q = \beta y + f(x). \tag{15}$$

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

Corollary 2. *If F is a field, then the map* Ab: $\mathrm{Aut}(F\langle x, y\rangle) \longrightarrow \mathrm{Aut}(F[\tilde{x}, \tilde{y}])$ *of Corollary* 1 *is bijective.*

Proof. H. Jung [4] proved that every automorphism of $F[\tilde{x}, \tilde{y}]$ 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 π of $F[\tilde{x}, \tilde{y}]$ there exists an elementary automorphism of $F\langle x, y\rangle$, say π^*, so that $\mathrm{Ab}(\pi^*) = \tilde{\pi} = \pi$. Since every automorphism of $F[\tilde{x}, \tilde{y}]$ 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 $F\langle x, y\rangle$ is tame.*

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

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

Since the same reasoning applies to ϕ^{-1} we are able to conclude that in fact α 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 ϕ is a tame automorphism of $R\langle x, y\rangle$ (or of $R[\tilde{x}, \tilde{y}]$), 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 ϕ and ϕ^{-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, ϕ induces an automorphism of the polynomial ring in x and y over R, but this cannot be tame. ϕ is obtained in the following way:

Define automorphisms

$$\alpha \text{ by } \begin{cases} \alpha(x) = x + c^{-1}y^2, \\ \alpha(y) = y; \end{cases} \qquad \beta \text{ by } \begin{cases} \beta(x) = x, \\ \beta(y) = y + c^3x^2; \end{cases}$$

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

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

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

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

Thus, ϕ 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 $R\langle x, y\rangle$ that keeps $[x, y]$ fixed (up to multiplication by a unit of R) is an automorphism of $R\langle x, y\rangle$?

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

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

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

REFERENCES

1. P. M. Cohn, *Free associative algebras*, Bull. London Math. Soc. 1 (1969), 1–39. MR 39 #2800.

2. ———, *On subsemigroups of free semigroups*, Proc. Amer. Math. Soc. 13 (1962), 347–351. MR 25 #3111.

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

4. A. Czerniakiewicz, *Automorphisms of a free associative algebra of rank* 2. I, Thesis, New York University, 1970; Trans. Amer. Math. Soc. 160 (1971), 393–401.

5. H. W. E. Jung, *Über ganze birationale Transformationen der Ebene*, J. Reine Angew. Math. 184 (1942), 161–174. MR 5, 74.

6. A. Gutwirth, *An inequality for certain pencils of plane curves*, Proc. Amer. Math. Soc. 12 (1961), 631–638. MR 23 #A4053.

7. M. Nagata, *A theorem of Gutwirth*, J. Math. Kyoto Univ. 11 (1971), 149–154. MR 43 #213.

8. W. van der Kulk, *On polynomial rings in two variables*, Nieuw Arch. Wisk. (3) 1 (1953), 33–41. MR 14, 941.(2)

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY, NEW YORK, NEW YORK 10027

(2) Added in proof.