Invariants and conjugacy classes of triangular polynomial maps

Stefan Maubach

Jacobs University Bremen

Bremen, Germany

s.maubach@jacobs-university.de

June 27, 2013

Abstract

In this article, we classify invariants and conjugacy classes of triangular polynomial maps. We make these classifications in dimension 2 over domains containing $\mathbb{Q}$, dimension 2 over fields of characteristic p, and dimension 3 over fields of characteristic zero. We discuss the generic characteristic 0 case. We determine the invariants and conjugacy classes of strictly triangular maps of maximal order in all dimensions over fields of characteristic p. They turn out to be equivalent to a map of the form $(x_1 + f_1, \ldots, x_n + f_n)$ where $f_i \in x_n^{p-1}k[x_{i+1}^p, \ldots, x_n^p]$ if $1 \leq i \leq n-1$ and $f_n \in k^*$.

AMS classification:

1 Introduction

1.1 Background

(For notations and some definitions, please read the next section.) Triangular polynomial maps are an important class of maps: they are the first nonlinear (nonaffine) polynomial automorphisms one comes up to, and they are a basic building block of many polynomial automorphisms. For one, in dimension two, all automorphisms are compositions of affine and triangular ones. Second, almost all basic examples (like Nagata's automorphism, exponents of locally nilpotent derivatons) are "almost triangular" (they are triangular over their invariant ring, or an exponent of a locally nilpotent derivation which is equivalent to a triangular derivation).

Due to polynomial automorphisms and endomorphisms in general being quite difficult, triangular polynomial maps are often considered trivial. (For example - it's completely trivial to prove the Jacobian Conjecture for triangular polynomial

endomorphisms...) This is deceptive, however: if it is trivial to see that a polynomial is an automorphism, doesn't make it easier to, for example, iterate it, or to find its invariants, or to find its conjugacy class. For all these last questions, there are some reasonably satisfactory answers one can give over fields of characteristic zero, or even rings or domains containing $\mathbb{Q}$ (see section 2). Over fields of characteristic p this becomes much harder already. It's exactly this characteristic p case, especially the finite field case, which has gained more of an interest, also outside of the field of affine algebraic geometry [7, 8, 11].

The overview of this paper is as follows: In section 1 we give background, introduction, definitions etc. In section 1.3 we elaborate on the characteristic 0 and rings-containing-$\mathbb{Q}$ case. We give a link between locally nilpotent derivations, which is reasonably well-known for the invariant case but not that well-known for the image case. We determine conjugacy classes in dimension 2 over general rings and in dimension 3 over fields. In section 3 we do in all dimensions the equivalent of the "locally nilpotent dervation having a slice" -case for characteristic p. Since there is no locally nilpotent derivation (or its characteristic p version, a locally iterative higher derivation), this case is truly different, and has a nontrivial answer (whereas the characteristic 0 case yields "equivalent to an affine map"). We provide a reasonable description of invariants, image and conjugacy classes for this case. This is perhaps the strongest new result of this paper. In section 4 we determine the dimension 2 case over fields of characteristic p. In section 5 we briefly discuss automorphisms of finite order, and in section 6 we give further research and acknowledgements.

Contents

1.2   Some notations and basic definitions

If R is a ring, we will denote $R[x_1, \ldots, x_n]$ as $R^{[n]}$. All rings in this article will be commutative with 1, and most of the time will be domains. We will reserve k for a field. We define $\mathrm{GA}_n(R)$ as the set of polynomial automorphisms of $R^{[n]}$, and elements $F \in \mathrm{GA}_n(R)$ as $F = (F_1, \ldots, F_n)$ where $F_i \in k^{[n]}$. $\mathrm{BA}_n(R)$ is the set of triangular polynomial automorphisms , i.e. where $F_i \in k[x_i, x_{i+1}, \ldots, x_n]$. (BA stands for Borel Automorphisms, see [1].) It follows that $F_i = a_i x_i + f_i$ where $f_i \in k[x_{i+1}, \ldots, x_n]$. The group $\mathrm{BAs}_n(R)$ is the set of strictly upper triangular polynomial maps, i.e. maps of the form $F = (x_1 + f_1, \ldots, x_n + f_n)$ where $f_i \in k[x_{i+1}, \ldots, x_n]$. $\mathrm{Aff}_n(R)$ is the set of affine maps, i.e. compositions of linear maps and translations.

1.3   Unipotent and triangular maps

Definition 1.1. Let $F \in \mathrm{GA}_n(R)$. Then F is called *locally finite* (short LF) if there exist $d \in \mathbb{N}$ and $a_i \in R$ such that $F^d = \sum_{i=0}^{d-1} a_i F^i$. It follows that $\deg(F^m)$ is bounded. In case the polynomial $T^d - \sum_{i=0}^{d-1} a_i T^i = (T-1)^d$, then we say that F is unipotent.

(Note: in some articles, LF is called "algebraic", see for example [5].)

Example 1.2. All elements in $\mathrm{BA}_n(R)$ are locally finite. The elements in $\mathrm{BAs}_n(R)$ are unipotent.

It will be convenient to abbreviate elements in $\mathrm{BAs}_n(k)$ which have many identity components, for example

$$(x_1, \ldots, x_{i-1}, x_i + f_i, x_{i+1}, \ldots, x_n) = (x_i + f_i)$$

$$(x_1, \ldots, x_{i-1}, x_i + f_i, x_{i+1}, \ldots, x_{j-1}, x_j + f_j, x_{j+1}, \ldots, x_n) = (x_i + f_i, x_j + f_j)$$

etc.

If $A \subseteq R^{[n]}$ then we denote $A^F = \{a \in A \mid F(a) = a\}$. In case A is clear (mostly, meaning $A = R^{[n]}$) then we write $\mathrm{inv}(F) = A^F$.

3

In this article, our goal is to understand elements in $\mathrm{BAs}_n(k)$ for any field k. In particular, we want to understand the following:

- What are the invariants of some $F \in \mathrm{BAs}_n(k)$?

- What are the conjugacy classes of $\mathrm{BAs}_n(k)$ in $\mathrm{BAs}_n(k)$?

- What are the conjugacy classes of $\mathrm{BAs}_n(k)$ in $\mathrm{BA}_n(k)$?

- What are the conjugacy classes of $\mathrm{BA}_n(k)$ in $\mathrm{BA}_n(k)$?

These questions will be a bit too ambitious to solve in general - in fact, one can say that even in characteristic zero, the invariants are quite complicated. (See example 2.8.) For completeness sake, we will first discuss the characteristic zero case, after which we will discuss the characteristic p case, which will be more involved. We will also consider the above questions over rings (domains), as these sometimes can help us answer the question over fields in one variable higher.

2 Characteristic zero

Let R be a domain of characteristic zero (i.e. $\mathbb{Q} \subseteq R$).

Definition 2.1. Let $D : R^{[n]} \longrightarrow R^{[n]}$ be an R-linear map. Then D is called a derivation if $D(fg) = fD(g) + D(f)g$ for all $f, g \in R^{[n]}$. D is called locally nilpotent if for every $f \in R^{[n]}$, there exists $d \in \mathbb{N}$ such that $D^d(f) = 0$. D is called triangular if $D(x_i) \in R[x_{i+1}, \ldots, x_n]$. A slice of D is an element $s \in R^{[n]}$ such that $D(s) = 1$.

Lemma 2.2. *Let R be a domain containing $\mathbb{Q}$. Then $F \in \mathrm{GA}_n(R)$ being unipotent is equivalent to $F = \exp(D)$ for some locally nilpotent derivation D.*

Proof. In [2] lemma 2.3, the above theorem is proven for the case $R = k$, a field. If we let k be the quotient field of R, we thus can find some l.n.d. D having coefficents in k. We will show that D actually has coefficients in R. We know that for each $m \in \mathbb{N}$, we get $F^m(x_i) = \exp(mD)(mx_i) \in R^{[n]}$. Let $d \in \mathbb{N}$ such that $D^d(x_i) = 0$, and let $V = \sum_{m \in \mathbb{N}} R \exp(D)(mx_i)$. We claim that $D(x_i) \in V$. Indeed:

$$
\begin{pmatrix}
\exp(0D)(x_i) \\
\exp(D)(x_i) \\
\vdots \\
\exp((d-1)D)(x_i)
\end{pmatrix}
= M
\begin{pmatrix}
x_i \\
D(x_i) \\
\vdots \\
D^{d-1}(x_i)
\end{pmatrix}
$$

where M is some Vandermonde matrix. So indeed $D^j(x_i) \in V$ for each j. Since $V \subseteq R^{[n]}$ we are done. $\qquad\square$

This fact makes the characteristic zero case so understandable. Before we state our main theorem, let us elaborate a bit. We have the following well-known theorem:

Lemma 2.3. *Let D be a locally nilpotent derivation on a ring A having a slice s. Then $A^D = \ker(D) = k[a_1, \ldots, a_n]$ for some $a_i \in A$, and $A = A^D[s]$. Furthermore, $\mathrm{Im}(D) = A$ if and only if D has a slice.*

A locally nilpotent derivation having a slice is obviously something very useful for such a derivation, and correspondingly for the map. It is conjectured that such a slice s must automatically be a coordinate (meaning there exist mates $s_2, \ldots, s_n$ such that $R[s, s_2, \ldots, s_n] = R^{[n]}$). Note that $\exp(D)(s) = s + 1$, making the following definition natural:

Definition 2.4. We say that $F \in \mathrm{GA}_n(R)$ has a slice if there exists $s \in R^{[n]}$ such that $F(s) = s + 1$. If s is a coordinate, we say that s is a coordinate slice.

When conjugating, we encounter the following fenomenon: $(x_i + g_i)(x_i + f_i, \vec{F}_{i+1})(x_i - g_i) = (x_i + f_i + g_i(\vec{F}_{i+1}) - g_i, \vec{F}_{i+1})$. This gives rise to the following definitions:

Definition 2.5. Given $F \in \mathrm{BAs}_n(R)$, define $N := F - I$, $\vec{F}_i = (F_i, F_{i+1}, \ldots, F_n)$, $N_i = \vec{F}_i - (x_i, \ldots, x_n)$ (i.e. $N_1 = N$).

Lemma 2.6. $(x_i + g_i)F(x_i - g_i) = (F_1(x_i - g_i), \ldots, F_{i-1}(x_i - g_i), F_i + N(g_i), F_{i+1}, \ldots, F_n)$.

The above (trivial) lemma explains how we can determine equivalence classes of elements in $\mathrm{BAs}_n(R)$: we first conjugate by a map $(x_n + g_n)$ to bring f_n to a standard form, then conjugate by $(x_{n-1} + g_{n-1})$ to bring f_{n-1} to a standard form, etc. This means that we need to understand $\mathrm{Im}(N)$. (Incidentally, $\ker(N) = \mathrm{inv}(F)$, and thus forms a similar role as a locally nilpotent derivation!)

We will prove the following theorem:

Theorem 2.7. *Let R be a ring containing $\mathbb{Q}$. Assume $F = \exp(D)$ is unipotent on $R^{[n]}$. Then*
(1) $\mathrm{inv}(F) = \ker(D)$ $(= \ker(N))$,
(2) $\mathrm{Im}(N) = \mathrm{Im}(D)$.

Proof. $\mathrm{inv}(F) = \{f \mid F(f) = f\} = \{f \mid (F - I)(f) = 0\} = \ker(N)$. Now $\mathrm{inv}(F)$ is the invariants of $\exp(D)$ which is well-known to be equal to $\ker(D)$ if D is locally nilpotent, so the first statement holds.

Define a degree function on R by $r \neq 0$ then $\deg(r) = \max\{d \mid D^d(r) \neq 0\}$, $\deg(0) = -\infty$. (It is well-known that this yields a degree function if D is locally nilpotent.) Denote $R_d := \{r \in R \mid \deg(R) \leq d\}$. Given $f \in R$, we proceed by induction to $d = \deg(f)$ to prove $f \in \mathrm{Im}(D) \iff f \in \mathrm{Im}(N)$.
$d = -\infty$: then $f = 0$ and the statement is true.
Assume that $\mathrm{Im}(D) \cap R_{d-1} = \mathrm{Im}(N) \cap R_{d-1}$, let $f \in R_d$.

Assume $f \in \mathrm{Im}(D)$, then $f = D(g)$ for some $g \in R$. Now $N(g) = D(g) + h$ where $h = \sum_{i=2}^{\infty} \frac{1}{i!} D^i(g) \in \mathrm{Im}(D)$. Since $\deg(h) = \deg(g) - 2 = d - 1$ we use induction and find h' such that $N(h') = h$, then $N(g - h') = D(g) = f$, and thus $f \in \mathrm{Im}(N)$.

Assume $f \in \mathrm{Im}(N)$, then $f = N(g) = D(g) + h$ where $h = $ as above. Since $h \in \mathrm{Im}(D)$ we find h' such that $D(h') = h$. Thus $D(g + h') = D(g) + D(h') = D(g) + h = f$ and thus $f \in \mathrm{Im}(D)$. $\qquad \square$

5

In some sense, the above theorem only translates the problem. $\mathrm{Im}(D)$ and $\ker(D)$ are not really easy even for triangular derivations. One example:

Example 2.8. Let $F = (x_1+x_5^3, x_2+x_6^3, x_3+x_7^3, x_4+(x_5 x_6 x_7)^2, x_5, x_6, x_7) \in \mathrm{BAs}_n(k)$, where $\mathrm{char}(k) = 0$. Then $\mathrm{inv}(F)$ is not finitely generated.

The above example is nothing other than the exponent of Robert's example [6], a locally nilpotent derivation whose kernel is not finitely generated. Note there exist counterexamples by Freudenburg & Daigle-Freudenburg in dimensions 5 and 6 too [3, 4].

2.1 Conjugacy classes

We now want to give some answer to how to describe (representants) of conjugacy classes. The generic case is rather complicated, and we will not fully answer it (similarly as no one truly can answer exactly what $\mathrm{Im}(D)$ and $\ker(D)$ are in general). We will focus on some special cases, especially as we want to determine what happens in low dimensions. A first case is easy, but it is an important case:

Proposition 2.9. *If* $F = (x_1 + f_1, \ldots, x_n + f_n) \in \mathrm{BAs}_n(R)$ *where* $f_n \in R^*$, *then* F *is in the same conjugacy class as* $(x_n + f_n)$.

Proof. Since $\mathrm{Im}(N_i) = k[x_i, \ldots, x_n]$ (as $F = \exp(D)$ where D has a slice, and using theorem 2.7) for each integer $2 \leq i \leq n$, we can find $g_i \in k[x_i, \ldots, x_n]$ such that $N(g_i) = N_i(g_i) = f_i$. This means that we can conjugate any $F = (x_1 + f_1, \ldots, x_j + f_j, x_{j+1}, \ldots, x_{n-1}, x_n + f_n)$ by $(x_j + g_j)$ (see lemma 2.6) and get a map of the form $(x_1 + f_1', \ldots, x_{j-1} + f_{j-1}', x_j, \ldots, x_{n-1}, x_n + f_n)$. Continuing this process we end up at the map $(x_n + f_n)$. $\qquad\square$

The above proof contains a little bit what one can do in general, and what is explained in lemma 2.6: given $F = \exp(D) = (x_1 + f_1, \ldots, x_n + f_1)$ we first conjugate by $(x_{n-1} + g_{n-1})$ for some appropriate g_{n-1} (conjugating by $(x_n + g_n)$ changes nothing). This changes the $n - 1$-term into $x_n + f_{n-1} + N(g_{n-1})$, and this requires us to understand $N(R[x_n])$, which is equal to $D(R[x_n])$. We then can pick a representant of f_{n-1} modulo $\mathrm{Im}(D)$ and continue by conjugating by $(x_{n-2} + g_{n-2})$ etc. It is therefore important to understand $R^{[n]}/\mathrm{Im}(D)$ for a triangular D; it enables one to understand the conjugacy classes in $\mathrm{BAs}_n(R)$.

At first it seems like $R^{[n]}/\mathrm{Im}(D)$ might be understandable, allow us to elaborate: $\mathrm{Im}(D)$ is a free R-module generated by $\{D(T) \mid T \text{ is a monomial in } R^{[n]}\}$. So you can reduce each given $g \in k^{[n]}$ modulo the highest degree terms appearing in these $D(T)$ (with respect to a lexicographic grading $\deg_L$ given by $x_1 \gg x_2 \gg \ldots \gg x_n$), and get something unique. (Indeed, with respect to this lexicographic ordering, denoting $\mathrm{lt}(g)$ as the leading term of g, one can give a nice description of $\mathrm{lt}(D(g))$ related to $\mathrm{lt}(g)$ with respect to this grading.)

However, this does not necessarily give unique representants of $R^{[n]}/\mathrm{Im}(D)$. The main reason why this fails is that there can exist polynomials g, h with the property that $\deg_L(g) > \deg_L(h)$ but $\deg_L(D(g)) < \deg_L(D(h))$. We give the following example:

Example 2.10. Let $D = x_2\partial_1 + \partial_2$. Then $\mathrm{lt}(x_1) > \mathrm{lt}(x_2^2)$ but $\mathrm{lt}(D(x_1)) < \deg_L(D(x_2^2))$.

Nevertheless, this is an important idea to keep in mind in the low dimensional cases we shall now consider.

2.2 Conjugacy classes within BAs in dimensions 2 and 3

Theorem 2.11. *The conjugacy classes of* $\mathrm{BAs}_2(R)$ *in* $\mathrm{BAs}_2(R)$ *where* $\mathbb{Q} \subseteq R$ *is a domain, are parametrized by pairs* $(\bar{f}_1, f_2)$ *where*

1. $f_2 = 0$, $\bar{f}_1 = a_n x_2^n + a_{n-1} x_2^{n-1} + \ldots + a_0$ *where* a_{n-1} *is picked as a unique representant in* $R/a_n R$,

2. $f_2 \in R \backslash \{0\}$, $\bar{f}_1 = \bar{a}_n x_2^n + \bar{a}_{n-1} x^{n-1} + \ldots + \bar{a}_0 \in R/(f_2)[x_2]$, *where* $\bar{a}_{n-1}$ *is a uniquely picked representant from* $(R/(f_2))/(\bar{a}_n)$.

In particular, if $R = k$ *a field, then the classes are*

1. $(x_1 + f(x_2), x_2)$, $f(x_2) = f_n x_2^n + f_{n-2} x_2^{n-1} + \ldots + f_0 \in k[x_2]$ *where* $f_n \neq 0$ *(i.e. the next-to highest term has coefficient zero)*,

2. $(x_1, x_2 + \lambda)$, $\lambda \in k$.

Proof. Let $F = (x_1 + f_1, x_2 + f_2)$ be the triangular map. The $x_2 + f_2$ part cannot be changed. We will first conjugate by something of the form $(x_1 + g_1)$ and then of the form $(x_2 + g_2)$. F equals $\exp(D)$ where $D = f_1'\partial_1 + f_2\partial_2$ and some $f_1' \in R[x_2]$. Since conjugation by $(x_1 + g_1)$ means modifying f_2 by $N(R[x_2])$, we need to understand $R[x_2]/N(R[x_2])$, and using theorem 2.7 part (2) we see $N(R[x_2]) = D(R[x_2]) = \mathrm{Im}(f_2\partial_2) = f_2 R[x_2]$. We can thus conclude that $(x_1 + g_1, x_2 + g_2)$ and F are equivalent under conjugation by some $(x_2 + h_2)$ if and only if $f_2 = g_2$ and $g_1 \in f_1 + f_2 R[x_2]$. Let us assume $f_1 = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_0$. We can still conjugate by $(x_2 + \tilde{h}_2)$, which means that we can change the term of degree $n - 1$: $f_1(x_2 + \tilde{h}_2)$ has the top part $a_n x^n + (n\tilde{h}_2 a_n + a_{n-1})x^{n-1} + \ldots$. Thus we can change the $n - 1$ term by any element in $a_n R$. Note that conjugating by any $(x_1 + \tilde{h}_1)$ or $(x_2 + \tilde{h}_2)$ disturbes the standard form (it is important that R is a domain here!). This finally proves the theorem for rings R.

The second case $R = k$ follows directly from the previous, but can also be partially proven using lemma 2.9. $\square$

Theorem 2.12. *The conjugacy classes of* $\mathrm{BAs}_3(k)$ *in* $\mathrm{BAs}_3(k)$ *are*

1. $(x_1, x_2, x_3 + \lambda)$, $\lambda \in k$,

2. $(x_1+f_1(x_2,x_3), x_2+f_2(x_3), x_3)$ where $f(x_2) = f_n x_2^n + f_{n-2} x_2^{n-1} + \ldots + f_0 \in k[x_2]$ where $f_n \neq 0$,

$$f_1(x_2, x_3) = a_n(x_3) x_2^n + a_{n-1}(x_3) x_2^{n-1} + \ldots + a_0(x_3),$$

and for $i = 0, 1, \ldots, n-2, n$, $a_i(x_3)$ is the lowest degree element in $a_i(x_3) + f_2(x_3)k[x_3]$, while $a_{n-1}(x_3)$ is the lowest degree element in $a_{n-1}(x_3) + (f_2(x_3), a_n(x_3))k[x_3]$.

Proof. Using proposition 2.9 we see that if $f_3 \in k^*$, then the map is equivalent to $(x_1, x_2, x_3 + f_3)$, yielding the first case. Left is $f_3 = 0$, which comes down to the general case of theorem 2.11, picking $R = k[x_3]$. The result now immediately follows, keeping in mind that since lowest degree elements in sets like $a_i(x_3) + f_2(x_3)k[x_3]$ are unique: if $a_i \in f_2 k[x_3]$, then the lowest element is zero. $\square$

2.3 Conjugacy classes of BAs in BA in dimensions 2 and 3

Any element in BA_n can be written as DG where D is diagonal linear, and $G \in \mathrm{BAs}_n$. This means that if we try to determine a representant of a conjugacy class of $F \in \mathrm{BAs}_n$, we conjugate by DG, i.e. consider $DGFG^{-1}D^{-1}$. We thus first pick a representant in $GFG^{-1} \in \mathrm{BAs}_n$, and on top of that conjugate by a diagonal linear map.

Conjugation by (λx_3) where $\lambda \in R^*$ on $R[x_3]$ gives a group action $R^* \times R[x_3] \longrightarrow R[x_3]$, given by $\lambda \cdot f(x_3) = \lambda^{-1} f(\lambda x_3)$. This case thus gives an additional gathering of conjugacy classes under these kind of orbits. There's not really a simplification of this possible, unless R is an algebraically closed field, or the reals or something specific. And even then it is limited: the polynomials of the form $ax^m + bx^l$ can be conjgated to the form $x^m + \tilde{b}x^l$, and then the coefficient $\tilde{b}$ can be changed if l does not divide m by some conjugation by λx where λ is an m-th root of unity, etc...

Theorem 2.13. *The conjugacy classes of* $\mathrm{BAs}_2(R)$ *in* $\mathrm{BA}_2(R)$ *where* $\mathbb{Q} \subseteq R$ *is a domain, are parametrized by pairs* $(\bar{f}_1, f_2)$ *where*

1. *$f_2 = 0$, $f_1 = a_n x_2^n + a_{n-1} x_2^{n-1} + \ldots + a_0$ where a_{n-1} is picked as a unique representant in $R/a_n R$, and then additionally f_1 is a unique element in the orbit of the action $\lambda \longrightarrow \lambda^{-1} f_1(\lambda x_2)$.*

2. *$f_2 \in R\backslash\{0\}$, $\bar{f}_1 = \bar{a}_n x_2^n + \bar{a}_{n-1} x^{n-1} + \ldots + \bar{a}_0 \in R/(f_2)[x_2]$, where $\bar{a}_{n-1}$ is a uniquely picked representant from $(R/(f_2))/(\bar{a}_n)$. Then, additionally f_1 is a unique element in the orbit of the action $\lambda \longrightarrow \lambda^{-1} f_1(\lambda x_2)$.*

In particular, if $R = k$ a field, then the classes are

1. *$(x_1 + f(x_2), x_2)$, $f(x_2) = f_n x_2^n + f_{n-2} x_2^{n-1} + \ldots + f_0 \in k[x_2]$ where $f_n \neq 0$ (i.e. the next-to highest term has coefficient zero). Additionally f_1 is a unique element in the orbit of the action $\lambda \longrightarrow \lambda^{-1} f_1(\lambda x_2)$.*

2. $(x_1, x_2 + 1)$,

3. (x_1, x_2).

Proof. The form $(x_2 + f_2)$ can be conjugated by $(x_1, \lambda x_2)$ where $\lambda \in R^*$. This explains why we pick $\bar{f}_2 \in R/R^*$. Let f_2 be any representant of $\bar{f}_2$. Then the ideal Rf_2 is always the same ideal, regardless of representant. Now $(x_1 + f_1, x_2 + f_2)$ can be conjugated by $(\lambda x_1, x_2)$ to get any element in $R^* f_1$ in stead of f_1. This proves the first statement.

The case $R = k$ is now trivial. $\qquad\square$

Theorem 2.14. *The conjugacy classes of* $\mathrm{BAs}_3(k)$ *in* $\mathrm{BA}_3(k)$ *are*

1. (x_1, x_2, x_3),

2. $(x_1, x_2, x_3 + 1)$,

3. $(x_1 + f_1(x_2, x_3), x_2 + f_2(x_3), x_3)$ *where* $f(x_2) = f_n x_2^n + f_{n-2} x_2^{n-1} + \ldots + f_0 \in k[x_2]$ *where* $f_n \neq 0$. *Additionally,* f_2 *is a unique element in the orbit of the action* $\lambda \longrightarrow \lambda^{-1} f_2(\lambda x_2)$. *Now*

$$f_1(x_2, x_3) = a_n(x_3) x_2^n + a_{n-1}(x_3) x_2^{n-1} + \ldots + a_0(x_3),$$

and for $i = 0, 1, \ldots, n-2, n$, $a_i(x_3)$ *is the lowest degree element in* $a_i(x_3) + f_2(x_3) k[x_3]$, *while* $a_{n-1}(x_3)$ *is the lowest degree element in* $a_{n-1}(x_3) + (f_2(x_3), a_n(x_3)) k[x_3]$. *Furthermore, the sequence* $(a_n, a_{n-1}, \ldots, a_0)$ *is picked uniquely from the orbit under conjugation by* (λx_3).

Proof. Using proposition 2.9 we see that if $f_3 \in k^*$, then the map is equivalent to $(x_1, x_2, x_3 + f_3)$, yielding the first case. Left is $f_3 = 0$, which comes down to the general case of theorem 2.11, picking $R = k[x_3]$. Picking a representant in $f_1 + (k[x_3]/f_3)[x_2]$ of lowest degree is unique, and the theorem is proven. $\qquad\square$

2.4 Conjugacy classes of BA in itself

The 2-variable case over a domain (and with that, the 3-variable case over a field) are much more involved (but doable in future research). Here we aim at the 2-variable case over a field k of characteristic zero. We start with the one-variable case over a ring:

Lemma 2.15. *Let R be a reduced ring (not necessarily containing $\mathbb{Q}$). Then the conjugacy classes of* $\mathrm{BA}_1(R) = \mathrm{GA}_1(R)$ *are*

1. $x + b$ *where b is a (unique) representant in R of R/R^* (the orbit space of the action $R^* \times R \longrightarrow R$),*

2. $ax + b$ *where $a \neq 1$, and b is a unique representant in R of $(R/(a-1)R)/R^*$ (the orbit space of the action $R^* \times R/(a-1)R \longrightarrow R/(a-1)R$).*

9

In particular, if $R = k$ is a field, then the conjugacy classes are ax, $a \in k*$ and $x + 1$.

Proof. A generic element looks like $ax + b$ where $a \in R^*, b \in R$. If $a = 1$, then conjugating by λx can change b to λb, meaning that we have to pick b's uniquely from each orbit of the natural action $R^* \times R \longrightarrow R$. Conjugation by $x + \lambda$ does nothing, so this gives the first case.

Now if $a \neq 1$ then let us conjugate by a generic element $(\mu x - \lambda)$ where $\mu \in R^*, \lambda \in R$. Then $(\mu x - \lambda)(ax + b)(\mu^{-1}x + \mu^{-1}\lambda) = (ax + \mu b + \lambda(a - 1))$. This means that we cannot change a, but we can change b to any element in $R^*b + R(a - 1)$. $\square$

(The above can be extended easily to R not a domain, but in case then one should be careful with the definition of BA: does one mean all invertible triangular maps, or maps which send each variable x_i to $\lambda_i x_i + f_i(x_{i+1}, \ldots, x_n)$. i.e. is $\mathrm{BA}_1(R)$ polynomials of degree 1 or $\mathrm{BA}_1(R) = \mathrm{GA}_1(R)$?)

Lemma 2.16. *Let k be a field of characteristic zero. The conjugacy classes of* $\mathrm{BA}_2(k)$ *are*

1. *Second component y:*

   (a) (x, y),

   (b) $(x + f(y), y)$ *where* $f(y) = y^d + a_{d-2}y^{d-2} + \ldots a_0$ *(i.e. monic and second coefficient zero), and picked uniquely from the set* $\{f(cy) \mid c^{d-1} = 1\}$.

   (c) (bx, y) *where* $b \neq 0, 1$,

2. *Second component $y + 1$:*

   (a) $(x, y + 1)$

   (b) $(ax, y + 1)$ *where* $a \in k^*, a \neq 1$

3. *Second component ay where $a \neq 1, 0$:*

   (a) (bx, ay) *if there is no* $m \in \mathbb{N}$ *such that* $b^m = a$,

   (b) $(a^m x + y^m, ay)$ *if a is no root of unity,*

   (c) $(a^m x + y^m f(y^r), ay)$ *if $r = \mathrm{ord}(a)$, and f monic. Furthermore, $y^m f(y^r)$ is uniquely picked from* $\{(\mu y)^m f(\mu^r y^r) \mid \mu^{m+rd} = 1\}$ *where* $d = \deg(f)$.

   Another classification is:

A *(affine), and then*

   1. $(bx, y + c)$ *where* $b \in k^*$, $c \in \{0, 1\}$,

   2. (bx, ay) *where* $a, b \in k^*$,

S *(sequential)*

$(a^m x + y^m f(y^r), ay)$ *where* $m \in \mathbb{N}$, $r = \text{ord}(a)$ *(*$r = 0$ *if* a *is no root of unity), and* $y^m f(y^r)$ *is monic. Furthermore,* $y^m f(y^r)$ *is uniquely picked from* $\{(\mu y)^m f(\mu^r y^r) \mid \mu^{m+rd} = 1\}$ *where* $d = \deg(f)$.

Proof. Write $F = (bx + f(y), ax + \lambda)$. Using lemma 2.15, and the fact that $a - 1$ is invertible if $a \neq 1$ and thus $(a-1)R = (a-1)k = k$, the second component is one of three $y, y+1, ay$ where $a \neq 1, 0$. We will consider these three cases.

Case $a = 1, \lambda = 0$

Then we have $(bx + f(y), y)$. We can actually apply lemma 2.15 to $(bx + f(y))$ on $R = k[y]$. If $b \neq 1$ then $k[y] = (b-1)k[y]$, so we can get $f = 0$. In case $b = 1$, f is unique in $k[y]/k^*$ - which means we can pick f monic (or $f = 0$). Now we've only taken into account conjugations by $(cx + g(y))$, however - we still need to check what happens under conjugation by $(cy + g)$, i.e. conjugations by (cy) and $(y + g)$. Write $d = \deg(f)$. Using conjugation by $(y + g)$ where $g \in k$, we can make sure that the $d - 1$ coefficient is zero. Conjugation by cy can change $f(y)$ by a nonzero scalar. However, $f(y)$ needs to stay monic, so we can only change $f(y)$ into $c^{-1} f(cy) = c^{d-1} y^d + \dots$ where $c^{d-1} = 1$.

Case $a = 1, \lambda = 1$

Conjugating by $(x + g(y))$ gives $(bx + bg(y) - g(y+1) + f(y), y+1)$ which means we need to consider the map $k[y] \longrightarrow k[y]$ given by $g(y) \longrightarrow bg(y) - g(y + 1)$. This map is surjective in all cases: if $b \neq 1$ then y^m is mapped to a polynomial of degree m and the map is bijective, actually. If $b = 1$ then y^m is mapped to a polynomial of degree $m - 1$, and the map is still surjective (though not bijective). Thus, we conjugate to $(bx, y+1)$. Any conjugation will disturb this form or leave it unchanged, so this is the final form for this case.

Case $a \neq 1, 0, \lambda = 0$

Conjugation by $(x + g(y))$ yields $(bx + bg(y) - g(ay) + f(y), ay)$. This means we need to understand the map $k[y] \longrightarrow k[y]$ given by $g(y) \longrightarrow bg(y) - g(ay)$. This map decomposes into homogeneous parts, so we need to consider $y^m \longrightarrow by^m - a^m y^m$. This map is surjective if $b - a^m \neq 0$ for all m. If $b = a^m$, then the y^m part cannot removed. So let m is the lowest integer such that $b = a^m$. If there's another integer m' such that $b = a^{m'}$, then $a^{m'-m} = 1$ and thus a has finite order. Concluding, we get $(a^m x + cy^m, ay)$ where $c \in k$ if a is no root of unity, and $(a^m x + y^m g(y^r), ay)$ where $r = \text{ord}(a)$, and $m < r$. Conjugating by (μx) we can make sure that $c = 1$ (or $c = 0$) and $g(y^r)$ monic. Conjugating by (μy) we change the monicness of $y^m g(y^r)$ unless $\mu^d = 1$ where $d = \deg(y^m g(y^r))$. $\qquad\square$

2.5   Higher dimensions

There are some higher dimensional cases which we expect that can be aquired by some more effort (but become rather technical): $\text{BAs}_3(R)$ in $\text{BAs}_3(R)$ and $\text{BA}_3(R)$ for domains $R \supset \mathbb{Q}$, and with that also $\text{BAs}_4(k)$ in $\text{BAs}_4(k)$ and $\text{BA}_4(k)$ for fields k

of characteristic zero. Also $\mathrm{BA}_2(R)$ in $\mathrm{BA}_2(R)$ and with that also $\mathrm{BA}_3(k)$ in $\mathrm{BA}_3(k)$ should be achievable. It is a bit of a challenge to give a good description which doesn't "explode", however.

We do expect that (perhaps in dimension 5 or 6) it is very, very hard or impossible to truly classify the conjugacy classes. We expect similar difficulties as with $\ker(D)$, which can be infinitely generated in dimension 5, and where it's unknown if it can be infinitely generated in dimension 4.

3 Characteristic p: strictly triangular maps of maximal order

3.1 Introduction

In this whole section 3, k is a field of characteristic p. The characteristic p case brings in additional difficulties with respect to the goal of classifying conjugates. In characteristic zero we have proposition 2.9, which essentially states that if F has last component $x_n + f_n$ where $f_n \in k^*$, then your map is very simply up to a conjugation. Another issue is that in characteristic p, we have no true equivalent of lemma 2.2, which states that $F = \exp(D)$ for some locally nilpotent derivation D. The equivalent object in characteristic p to a locally nilpotent derivation is a locally finite higher iterative derivation (see [9], we will not give details in this article), which has the following issue: if $F = \exp(D)$ where D is such a locally finite higher iterative derivation, then $F^p = I$. This means, that it doesn't even include all strictly triangular polynomial maps, let alone all unipotent maps. Hence, we need to resort to (slightly) different methods, and will have difficulty going into dimension 3 (and higher) except for special cases.

Definition 3.1. If $F \in \mathrm{BAs}_n(R)$, define $M_i := \sum_{j=0}^{p^{n+1-i}-1} F^j$ for $1 \le i \le n$. We recall the definition $N = F - I$. We say $M = M_1$.

Lemma 3.2. *Let R be a commutative ring of characteristic p. Let $F = (x_1 + f_1, \ldots, x_n + f_n) \in \mathrm{BAs}_n(R)$.*
(1) F^p fixes x_n, i.e. $F^p \in \mathrm{BAs}_{n-1}(R[x_n])$.
(2) $F^{p^n} = I$.
(3) $F^{p^m} = (x_1 + g_1, \ldots, x_{n-m} + g_{n-m}, x_{n-m+1}, \ldots, x_n)$ for some $g_i \in R[x_{i+1}, \ldots, x_n]$, and $g_{n-m} = M_{n-m}(f_{n-m}) \in \mathrm{Im}(M_{n-m})$.

Proof. (1) is trivial. (2) follows from (1) using induction. (3) we prove by induction: we prove that if $1 \le d \le p^m$ then the $n-m$-th part of F^d is $x_{n-1} + \sum_{i=0}^{d-1} F^i(f_{n-m})$. Indeed, for $d = 1$ this is correct. Now assume d. Then the $n-m$-th component of $F^{d+1} = F^d \circ F$ equals $(x_{n-m} + \sum_{i=0}^{d-1} F^i(f_{n-m})) \circ F = x_{n-m} + f_{n-m} + \sum_{i=1}^{d} F^i(f_{n-m}))$, proving the induction step. $\qquad\square$

3.2 Main theorems on invariants and an exact sequence

The important special case we consider in all dimensions is the case where $\operatorname{ord}(F) = p^n$. On this, we want to prove the following two theorems:

Theorem 3.3. *Let $F = (x_1 + f_1, \ldots, x_n + f_n) \in \operatorname{BAs}_n(k)$ where $\operatorname{ord}(F) = p^n$. Then $\operatorname{inv}(F) = k[\tilde{x}_1, \ldots, \tilde{x}_n]$ where $\tilde{x}_i = x_i^p - a_i^{p-1} x_i + b_i$ where $a_i, b_i \in k[x_{i+1}, \ldots, x_n]$.*

Theorem 3.4. *Let $F = (x_1 + f_1, \ldots, x_n + f_n)$ such that $\operatorname{ord}(F) = p^n$. Then the sequence*

$$0 \longrightarrow (k^{[n]})^F \longrightarrow k^{[n]} \xrightarrow{N} k^{[n]} \xrightarrow{M} (k^{[n]})^F \longrightarrow 0$$

and the sequence

$$0 \longrightarrow \operatorname{Im}(N) \longrightarrow k^{[n]} \xrightarrow{M} k^{[n]} \xrightarrow{N} \operatorname{Im}(N) \longrightarrow 0$$

are exact.

The proof of both theorems is rather involved, as the proof of the n-dimensional case of any of the theorems involves the $n-1$-dimensional case of both theorems. In fact, if we denote $T_{3.3}[n]$ by the statement "theorem 3.3 is true in dimension n" and similarly $T_{3.4}[n]$, the proof will follow the following scheme:

- Prove $T_{3.3}[1]$ and $T_{3.4}[1]$,

- Prove $(T_{3.3}[n-1], T_{3.4}[n-1]) \longrightarrow T_{3.3}[n]$,

- Prove $(T_{3.3}[n], T_{3.4}[n-1]) \longrightarrow T_{3.4}[n]$.

3.3 Generalities on linear maps of order p^n

Lemma 3.5. *Let V be a k-vector space where k is a field of characteristic p, and let $L : V \longrightarrow V$ be a k-linear map such that $L^{p^m} = I$. Then*
(1) $L = I + N$ where $N^{p^m} = 0$,
(2) the only eigenvalue of L is 1,
(3) L is locally finite,
(4) $I + L + L^2 + \ldots + L^{p^m - 1} = N^{p^m - 1}$.

Proof. (1) Write $N = L - I$. Then $L^p = (I + N)^p = I + N^p$ so $L^p = I \Longleftrightarrow N^p = 0$.
(2) follows from all eigenvalues being unit roots of order p^m, and the fact that the only p-th root of 1 in characteristic p is 1 (the only solution to $x^p - 1$ is 1).
(3) Let $w \in V$. Then $w, L(w), \ldots, L^{p^m - 1}(w)$ spans a finite dimensional subspace V_w of V such that $w \in W$. Hence L is locally finite.
(4)

$$\sum_{n=0}^{p^m - 1} L^n = \sum_{n=0}^{p^m - 1} (I + N)^n = \sum_{n=0}^{p^m - 1} \sum_{i=0}^{n} \binom{n}{i} N^i = \sum_{i=0}^{p^m - 1} \left(\sum_{n=i}^{p^m - 1} \binom{n}{i} \right) N^i =$$

13

$$\sum_{i=0}^{p^m-1} \binom{p^m}{i+1} N^i = N^{p^m-1}.$$

$\square$

Remark 3.6. Note that $M := N^{p^m-1}$.

Lemma 3.7. *Let L be as in lemma 3.5. Define $N = L - I$ and $M = I + L + L^2 + \ldots + L^{p^m-1}$. The sequence*

$$0 \longrightarrow V^L \longrightarrow V \xrightarrow{N} V \xrightarrow{M} V^L \longrightarrow V^L/\mathrm{Im}(M) \longrightarrow 0$$

is a well-defined complex sequence. The only non-trivial homology is $\ker(M)/\mathrm{Im}(N)$. If V is not only a k-module but also a ring (i.e. a k-algebra), and L is a ring homomorphism of V, then $\mathrm{Im}(M)$ is an ideal of V^L.

Proof. Exact at the first V^L is trivial.
Exact at the first V follows since $\ker(N) = \ker(L - I) = \{v \in V \mid L(v) = v\} = L^V$.
Well-defined at the second V follows since from lemma 3.5 we see that $M = N^{p^m-1}$ and thus $NM = 0$.
Well-defined at the second V^L: $M(v) = v + Lv + \ldots + L^{v^m-1}v$ is invariant under L so $\mathrm{Im}(M) \subseteq V^L$. Exactness is trival.
For the last sentence: notice that V^L is automatically a k-algebra. Now let us show that $\mathrm{Im}(M)$ is an ideal. Let $w \in \mathrm{Im}(M)$, $v \in V^L$. Then there eixsts $u \in V$: $M(u) = w$. Now $M(uv) = \sum_{i=0}^{p^m-1} L^i(uv) =$ (using that L is a ring homomorphism and $L(v) = v$) $= \sum_{i=0}^{p^m-1} L^i(u)v = M(u)v$ hence $wv \in V^L$. Since $\mathrm{Im}(M)$ is a linear subspace of V^L, we are done. $\square$

In the next sections, $V = k^{[n]}$ and $L = F = (x_1 + f_1, \ldots, x_n + f_n) \in \mathrm{BAs}_n(k)$.

3.4 Dimension 1

Note that below, N, M are as in lemma 3.5 and definition 3.6. We have $n = 1$ here, so F of order p is equivalent to $f_1 \in k^*$.

Lemma 3.8. *If $F = (x_1 + f_1)$ where $f_1 \in k^*$, then $k[x_1]^F = k[\tilde{x}_1]$ where $\tilde{x}_1 = x_1^p - f_1^{p-1}x_1$ and the sequences*

$$0 \longrightarrow k[\tilde{x}_1] \longrightarrow k[x_1] \xrightarrow{N} k[x_1] \xrightarrow{M} k[\tilde{x}_1] \longrightarrow 0$$

$$0 \longrightarrow \mathrm{Im}(N) \longrightarrow k[x_1] \xrightarrow{M} k[x_1] \xrightarrow{N} \mathrm{Im}(N) \longrightarrow 0$$

are exact.
A representant system for $k[x_1]/\mathrm{Im}(N)$ is $x_1^{p-1}k[\tilde{x}_1]$. Another representant system is $x_1^{p-1}k[x_1^p]$.

Proof. Since $\text{ord}(F) = p$ this is a special case of 3.7. We need to check a few things:
(1) $k[x_1]^F = k[\tilde{x}_1]$. Note that $\tilde{x}_1 \in k[x_1]^F$ indeed. Let $f \in k[x_1]^F$ be the lowest degree polynomial which is not in $k[\tilde{x}_1]$. Since we can reduce f by $\tilde{x}_1^m$, we can assume that $\deg(f) = m$ is not a multiple of p. So assume $f = a_m x_1^m + a_{m-1} x_1^{m-1} + \ldots$ where $a_i \in k$ and $(m, p) = 1$. Then the coefficient of x_1^{m-1} of $f(x_1 + f_1) - f(x_1)$ turns out to be $a_m m f_1$ which is nonzero. This is a contradiction, so there exists no such f, and thus $k[x_1]^F = k[\tilde{x}_1]$.
(2) $\text{Im}(M) = k[\tilde{x}_1] = \ker(N)$. Note that N decreases degree: hence, $M(x_1^{p-1}) = N^{p-1}(x_1^{p-1}) \in k$. But the constant term is $\sum_{i=0}^{p}(x_1 + if_1)^{p-1} = \sum_{i=0}^{p} i^{p-1} f_1^{p-1} = -f_1^{p-1} \in k^*$. Hence, $1 \in \text{Im}(M)$ so the ideal $\text{Im}(M) = k[\tilde{x}_1]$.
(3) $\ker(M)/\text{Im}(N) = 0$. The sequence is now exact if $\ker(M)/\text{Im}(N) = 0$. If the vector spaces would be finite dimensional, then this result follows from the fact that $\ker(N) = \text{Im}(M)$. We will restrict to finite dimensional subspaces to conclude the result: Define $k[x_1]_d = $ the set of polynomials of degree d and less. We claim that the sequence

$$0 \longrightarrow k[\tilde{x}_1]_d \overset{i}{\hookrightarrow} k[x_1]_d \overset{N}{\longrightarrow} k[x_1]_d \overset{M}{\longrightarrow} k[\tilde{x}_1]_d \longrightarrow 0$$

is exact **if** $d = d_0 p + p - 1$ for some $d_0 \in \mathbb{N}$ (it will NOT be exact for other d !). Now if $g \in k[\tilde{x}_1]$ then $M(x_1^{p-1} g) = g$; hence $k[\tilde{x}_1]_{pd_0+p-1} \supseteq M(k[x_1]_d) \supseteq M(x_1^{p-1} \cdot k[\tilde{x}_1]_{pd_0}) = k[\tilde{x}_1]_{pd_0} = k[\tilde{x}_1]_{pd_0+p-1}$ and thus $\text{Im}(M|_{k[x_1]_d}) = \ker(N|_{k[x_1]_d})$, a vector space of dimension d_0. Thus, $\dim(\ker(M|_{k[x_1]_d}) = \dim(\text{Im}(N_{k[x_1]_d}))$ and we can conclude that since $\ker(M|_{k[x_1]_d}) \supseteq \text{Im}(N_{k[x_1]_d})$ that they must be equal. If we now take unions, we get

$$\ker(M) = \bigcup_{d \in p\mathbb{Z}+p-1} \ker(M|_{k[x_1]_d}) = \bigcup_{d \in p\mathbb{Z}+p-1} \text{Im}(N_{k[x_1]_d}) = \text{Im}(N).$$

(4) The second sequence is exact since $\ker(M) = \text{Im}(N)$ by the first exact sequence, and by the fact that $\text{Im}(M) = \ker(N) = k[\tilde{x}_1]$.

Now let us determine a (5) representant system for $k[x_1]/\text{Im}(N)$. Since $N(x_1^{p-1} g(\tilde{x}_1)) = g(\tilde{x}_1)$, $\ker(N) \cap x_1^{p-1} k[\tilde{x}_1] = \{0\}$. Thus $x_1^{p-1} k[\tilde{x}_1]$ is a representant system of $k[x_1]/\ker(N) = k[x_1]/\text{Im}(M)$.

Now notice that if $(m, p) = 1$, then $\deg(N(x_1^m)) = m - 1$. This means that we have polynomials of all degrees d as long as $d \mod p \neq p-1$. Note that $\deg(\tilde{x}_1) = p$, and notice that we thus have $x_1^{p-1} \tilde{x}_1^m \mod (\ker(N)) = x_1^{p-1}(x_1^{pm} + a_{m-1} x_1^{p(m-1)} + \ldots + a_1 x_1^p + a_0)$ for some $a_i \in k$. These elements form a k-basis of a new representant system for $k[x_1]/\ker(N)$. This means that $\{x_1^{p-1}(x_1^{pm}) \mid m \in \mathbb{N}\}$ forms a k-basis of this new representant system, i.e. $x_1^{p-1} k[x_1^p]$ is another representant system. $\square$

Corollary 3.9. *Let R be a commutative domain of characteristic p. If $F = (x_1 + f_1)$ where $f_1 \in R^*$, then $R[x_1]^F = R[\tilde{x}_1]$ where $\tilde{x}_1 = x_1^p - f_1^{p-1} x_1$ and the sequences*

$$0 \longrightarrow R[\tilde{x}_1] \longrightarrow R[x_1] \overset{N}{\longrightarrow} R[x_1] \overset{M}{\longrightarrow} R[\tilde{x}_1] \longrightarrow 0$$

15

$$0 \longrightarrow \mathrm{Im}(N) \longrightarrow R[x_1] \xrightarrow{M} R[x_1] \xrightarrow{N} \mathrm{Im}(N) \longrightarrow 0$$

are exact.
A representant system for $R[x_1]/\mathrm{Im}(N)$ is $x_1^{p-1}R[\tilde{x}_1]$. Another representant system is $x_1^{p-1}R[x_1^p]$.

Proof. Obviously if k is the quotient field of R, the result follows from lemma 3.8. We need to check that intersecting from $k[x_1]$ to $R[x_1]$ everything goes well. First of all, $R[x_1]^F = k[x_1]^F \cap R[x_1] = k[\tilde{x}_1] \cap R[x_1] = R[\tilde{x}_1]$. Then, we need to check that the maps N and M do not miss things in their images; i.e. we need to check that $\mathrm{Im}(N|_{R[x_1]}) = \mathrm{Im}(N|_{k[x_1]}) \cap R[x_1]$, whereas a priori we only have $\subseteq$ in stead of $=$. (Note that $\ker(N|_{R[x_1]}) = \ker(N|_{k[x_1]}) \cap R[x_1]$ as well as $\ker(M|_{R[x_1]}) = \ker(M|_{k[x_1]}) \cap R[x_1]$ trivially.) $R[x_1]$ is a free R-module with basis $1, x_1, x_1^2, \ldots$. M and N send these basis elements into $\mathbb{F}_p(f_1)[x_1]$ (which is a subring of $R[x_1]$), and thus $\mathrm{Im}(N|_{R[x_1]}) = R \cdot \mathrm{Im}(N|_{\mathbb{F}_p(f_1)[x_1]}) = k \cdot \mathrm{Im}(N|_{\mathbb{F}_p(f_1)[x_1]}) \cap R[x_1] = \mathrm{Im}(N|_{k[x_1]}) \cap R[x_1] = \ker(N|_{k[x_1]}) \cap R[x_1] = \ker(N|_{R[x_1]})$. A similar proof for M.

The results on the representant system follow by a similar argument: A basis of the representant system over k is $\{x_1^{p-1}\tilde{x}_1^i; i \in \mathbb{N}\}$. Then this can be used as a basis for the representant system of $R[x_1]/\mathrm{Im}(N|_{R[x_1]})$ as well. $\qquad\square$

3.5 Induction step $T_{3.3}[n]$ from $T_{3.3}[n-1]$ and $T_{3.4}[n-1]$

In the rest of this section, we will consider $F = (x_1 + f_1, \ldots, x_n + f_n) \in \mathrm{BAs}_n(k)$ of order p^n. Define $F_i = x_i + f_i$ and $\vec{F}_i = (F_i, F_{i+1}, \ldots, F_n)$. We define $N = F - I$ and $M = \sum_{i=0}^{p^n-1} F^i$ as before, but also define

$$N_i := \vec{F}_i - I_i = N|_{k[x_i, \ldots, x_n]}$$

Notice that $\vec{F}_1 = F, N_1 = N$.

Lemma 3.10. *Let R be a domain of characteristic p. Let $F = (x_1+f_1, \ldots, x_n+f_n) \in \mathrm{BAs}_n(R)$, where $f_n \in R^*$. Then $F^p = I$ if and only if F can be conjugated by some $\tau \in \mathrm{BAs}_{n-1}(R[x_n])$ to $\tau^{-1}F\tau = (x_n + f_n)$.*

Proof. The "if" side is trivial. So let us assume $F^p = I$ and show we can conjugate F to the given form. Let us assume we can conjugate F to $(x_1 + f_1, \ldots, x_k + f_k, x_{k+1}, \ldots, x_{n-1}, x_n + f_n)$. We will consider this as a map $F' = (x_1 + f_1, \ldots, x_k + f_k, x_n + f_n)$ on $(R[x_{k+1}, \ldots, x_{n-1}])[x_1, \ldots, x_k, x_n]$. We get $I = F^p = (\ldots, x_k + M(f_k), x_n + f_n)$, and thus apparently $M(f_k) = 0$. Thus, $f_k \in \ker(M)$ which equals (using corollary 3.9) $\mathrm{Im}(N)$. Let g_k be such that $N(g_k) = f_k$. Then $(x_k - g_k)F(x_k + g_k) = (\ldots, x_k, x_{k+1}, \ldots, x_{n-1}, x_n+f_n)$. Continuing this process, the lemma is proven. $\qquad\square$

Lemma 3.11. *Assume $T_{3.3}[n-1]$ and $T_{3.4}[n-1]$. Then $T_{3.3}[n]$ holds.*

16

Proof. We are thus considering $F = (x_1 + f_1, \ldots, x_n + f_n) \in \mathrm{BAs}_n(k)$ where $f_n \in k^*$ and $\mathrm{ord}(F) = p^n$. We want to prove $\mathrm{inv}(F) = k[\tilde{x}_1, \ldots, \tilde{x}_n]$ where $\tilde{x}_i = x_i^p + a_i^{p-1} x_i + b_i$ where $a_i, b_i \in k[x_{i+1}, \ldots, x_n]$. The induction assumption $T_{3.4}[n-1]$ is used in the form of $\mathrm{Im}(M_2) = \ker(N_2)$ and $\mathrm{Im}(N_2) = \ker(M_2)$.

Consider $F^{p^{n-1}} = (x_1 + g_1, x_2, \ldots, x_n)$. Note that $0 \neq g_1 = M_2(f_1) \in \mathrm{Im}(M_2) \subseteq \ker(N_2) \subseteq \ker(N) = \mathrm{inv}(F)$ (no appeal to assumptions here!) and thus $F(g_1) = g_1$. Now $\mathrm{inv}(F^{p^{n-1}}) = k[x_1', x_2, \ldots, x_n]$ where $x_1' = x_1^p - g_1^{p-1} x_1$. Note that $\mathrm{inv}(F) \subseteq \mathrm{inv}(F^{p^{n-1}})$. We will restrict F to $A := \mathrm{inv}(F^{p^{n-1}})$ and compute $A^F = \mathrm{inv}(F)$. Now

$$F(x_1') = (x_1 + f_1)^p - F(g_1)^{p-1}(x_1 + f_1) = x_1' + f_1^p - g_1^{p-1} f_1.$$

Thus, $F|_A$ is triangular: $(x_1', x_2, \ldots, x_n) = (x_1' + g, x_2 + f_2, \ldots, x_n + f_n)$ where $g = f_1^p - g_1^{p-1} f_1$.

$F|_A$ has order p^{n-1}: it must be at least p^{n-1} since $F|_A$ restricted to $k[x_2, \ldots, x_n]$ has order p^{n-1}. If $h \in A$, then $F^{p^{n-1}}(h) = h$ by definition - so $F|_A$ is at most p^{n-1}.

Now $F|_A = (x_1' + g, \tilde{F})$ where $\tilde{F} = F|_{k[x_2, \ldots, x_n]}$. $I|_A = F|_A^{p^{n-1}} = (x_1' + M_2(g), x_2, \ldots, x_n)$ and thus apparently $M_2(g) = 0$, i.e. $g \in \ker(M_2) = \mathrm{Im}(N_2) = k[x_2, \ldots, x_n]^{\tilde{F}}$ (here we have used $T_{3.3}[n-1]$). This means that we can find $h \in k[x_2, \ldots, x_n]$ such that $(x_1' + h) F|_A (x_1' - h) = (x_1', \tilde{F})$. Thus, $\mathrm{inv}(F|_A) = (x_1 + h) \mathrm{inv}(x_1', \tilde{F})$. Now $\mathrm{inv}(\tilde{F}) = k[\tilde{x}_2, \ldots, \tilde{x}_n]$ as provided by induction. So $\mathrm{inv}(F|_A) = (x_1' + h) k[x_1', \tilde{x}_2, \ldots, \tilde{x}_n] = k[x_1' + h, \tilde{x}_2, \ldots, \tilde{x}_n]$ and thus if we define $\tilde{x}_1 = x_1' + h = x_1^p - g_1^{p-1} x_1 + h$ we are done. $\qquad\square$

Remark 3.12. In the proof above, we thus see that the a_i in theorem 3.3 satisfy $a_i = M_i(f_i)$.

3.6 Induction step $T_{3.4}[n]$ from $T_{3.4}[n-1]$ and $T_{3.3}[n]$

Definition 3.13. Assume $T_{3.3}[n]$. Given $F = (x_1 + f_1, \ldots, x_n + f_n)$, define (inductively) $\deg_F$ on $k^{[n]}$ by
(1) $\deg_F(x_n) = 1$,
For $i = 2$ to n choose $\deg_F(x_i)$ large enough such that such that
(2a) $\deg_F(x_i) \geq \deg_F(f_i)$,
(2b) $\deg_F(x_i) \geq \deg_F(a_i), \deg_F(b_i)$ from theorem $T_{3.3}[n]$,
(2c) $\deg_F(x_i) \in p\mathbb{Z}$.
Write $k_d^{[n]} = \{g \in k^{[n]} \mid \deg_F(g) \leq d\}$.
Define $V_d := k[x_n]_{p-1} k_d^{[n]}$.
Define $W_d := V_d \cap (k^{[n]})^F = V_d^F$.

In the above definition, we have some choice in $\deg_F(x_i)$, but we can make it unique in stating that $\deg_F(x_i)$ should be as low as possible within the constraints (though we don't really care). The requirement (2a) is picked such that $F(k_d^{[n]}) \subseteq k_d^{[n]}$. The requirement (2b) is picked so $\deg_F(\tilde{x}_i) = \deg_F(x_i^p)$, in order to be able to

predict the degree of $\tilde{x}_i$. The requirement (2c) is added so that $\deg_F(x_n)$ is the only variable having degree coprime to p.

Note that since $F(k_d^{[n]}) \subseteq k_d^{[n]}$, we have a finite dimensional filtration of $k^{[n]}$ preserved by F (see lemma 3.5). Note that also $F(V_d) \subseteq V_d$, as $F(x_n^i g) = (x_n + f_n)^i F(g) \in V_d$ for every $g \in (k^{[n]})_d$ etc. so V_d gives another such filtration of $k^{[n]}$.

Lemma 3.14. *Assuming $T_{3.3}[n]$, we have $W_{pd} = (k_{pd}^{[n]})^F$.*

Proof. If $(k^{[n]})^F = k[\tilde{x}_1, \ldots, \tilde{x}_n]$, then we see that any element invariant under F has degree a multiple of p, i.e. $W_{pa+b} \subseteq k_{pa}^{[n]}$ for each $a, b \in \mathbb{N}, 0 \le b \le p - 1$. Since $V_{pd} \subseteq k_{pd+p-1}^{[n]}$ we get $W_{pd} = V_{pd} \cap (k^{[n]})^F \subseteq (k_{pd}^{[n]})$ and the result follows. $\square$

Lemma 3.15. *Assume $T_{3.3}[n]$ and $T_{3.4}[n-1]$. Then $T_{3.4}[n]$ is true. So, let $F = (x_1 + f_1, \ldots, x_n + f_n)$ such that $\mathrm{ord}(F) = p^n$. Then the sequence*

$$0 \longrightarrow (k^{[n]})^F \longrightarrow k^{[n]} \xrightarrow{N} k^{[n]} \xrightarrow{M} (k^{[n]})^F \longrightarrow 0$$

is exact. Furthermore, $\mathrm{Im}(M) = \ker(N)$.

Proof. Notice first that $f_n \in k^*$ since otherwise $\mathrm{ord}(F) \le p^{n-1}$.

Using lemma 3.7 we see that the only things to prove are (1) $\ker(N) = (k^{[n]})^F$, (2) $\mathrm{Im}(M) = (k^{[n]})^F$, (3) $\mathrm{Im}(N) = \ker(M)$ and (4) $\mathrm{Im}(M) = \ker(N)$.
(1) $\ker(N) = \{g \in k^{[n]} \mid F(g) - g = 0\} = (k^{[n]})^F$.
(2) $M(x_n^{p-1}) = -f_n^{p-1} \in k^*$ (see part (2) of the proof of 3.8 for a detailed computation), so the ideal $\mathrm{Im}(M) = (k^{[n]})^F$.
(4) follows from (1) and (2).
(3) It is now tempting to state that since $(k^{[n]})^F = \ker(N) = \mathrm{Im}(M)$, then $\mathrm{Im}(N) = \ker(M)$, but since the k-vector spaces are infinite dimensional, this argument does not hold. However, what we will do, is restrict to finite dimensional subspaces V_d and $W_d := V_d \cap (k^{[n]})^F$ for which $\cup_d V_d = k^{[n]}$, *and* for which the restricted sequence

$$0 \longrightarrow W_d \longrightarrow V_d \xrightarrow{N} V_d \xrightarrow{M} W_d \longrightarrow 0$$

is (a) well-defined, (b) exact. Note that there DO exist linear subspaces for which (a) holds but (b) not, so we need to define V_d carefully - we claim that the definition in 3.13 works for well-chosen d. For this, note that $M(x_n^{p-1}) \in k*$, and that if $f \in (k^n)^F$ and $g \in k^{[n]}$, then $M(fg) = fM(g)$. This means that $M(V_d) \supseteq M(x_n^{p-1}(k_d^{[n]})^F) = (k_d^{[n]})^F$. Thus, using lemma 3.14

$$W_{pd} = V_{pd} \cap (k^{[n]})^F \supseteq M(V_{pd}) \supseteq (k_{pd}^{[n]})^F = W_{pd}$$

Thus, $\mathrm{Im}(M|_{V_{pd}}) = W_{pd} = \ker(N|_{V_{pd}})$ and the sequence

$$0 \longrightarrow W_{pd} \longrightarrow V_{pd} \xrightarrow{N} V_{pd} \xrightarrow{M} W_{pd} \longrightarrow 0$$

is exact. So (3) holds (and (1),(2), (3) yield that the sequence stated in the lemma is exact). $\square$

3.7 Conjugacy classes: maximal order case in $\mathrm{BAs}_n(k)$

When we're conjugating F by $(x_i + g_i)$, then the i-th component changes by $g_i - F(g_i) = N(g_i)$, which means we need to understand $\mathrm{Im}(N)$ - or, better said, we want to have representants in $k^{[n]}$ of $k^{[n]}/\mathrm{Im}(N)$. For this, we need the following lemma, which builds on on theorems 3.3 and 3.4.

Lemma 3.16. *Let $F = (x_1 + f_1, \ldots, x_n + f_n)$ such that $\mathrm{ord}(F) = p^n$. Then*

$$k^{[n]} = x_n^{p-1} \mathrm{inv}(F) \oplus \mathrm{Im}(N).$$

Proof. Using theorem 3.4 we see that $\mathrm{Im}(N) = \ker(M)$. We have the surjective map $k^{[n]} \xrightarrow{M} \mathrm{inv}(F)$. We provide a section $s : \mathrm{inv}(F) \longrightarrow k^{[n]}$ by $s(f) = -f_n^{1-p} x_n^{p-1} f$. Indeed, since $M(x_n^{p-1}) = -f_n^{p-1}$, we get that $M(-f_1^{1-p} x_n^{p-1} f) = f$ (see the argument at the end of the proof of 3.7 for detailed reasoning). Thus, $Ms(f) = f$. This means that we can make a split exact sequence

$$0 \longrightarrow \ker(M) \longrightarrow k^{[n]} \longrightarrow \mathrm{Im}(M) \longrightarrow 0$$

and $k^{[n]} = \ker(M) \oplus s(\mathrm{Im}(M)) = \mathrm{Im}(N) \oplus x_n^{p-1} \mathrm{inv}(F)$. $\qquad\square$

Corollary 3.17. *(of lemma 3.16) Let $F = (x_1 + f_1, \ldots, x_n + f_n)$ such that $\mathrm{ord}(F) = p^n$. Then*

$$k^{[n]} = x_n^{p-1} k[x_1^p, \ldots, x_n^p] \oplus \mathrm{Im}(N).$$

Proof. Theorem 3.3 tells us that $\mathrm{inv}(F) = k[\tilde{x}_1, \ldots, \tilde{x}_n]$. Lemma 3.16 tells us that a k-basis of a representant system of $k^{[n]}/\mathrm{Im}(N)$ is $\{x_n^{p-1} \tilde{x}_1^{a_1} \cdots \tilde{x}_n^{a_n} \mid a_i \in \mathbb{N}\}$. Now let us use the standard lexicographic ordering dlex on $k^{[n]}$ by stating $x_1 >> \ldots >> x_n$ (and beyond that the standard ordering, like $\mathrm{dlex}(x_i^2) > \mathrm{dlex}(x_i)$). Then of any $\tilde{x}_1^{a_1} \cdots \tilde{x}_n^{a_n}$, the leading term w.r.t. dlex is $x_1^{pa_1} \cdots x_n^{pa_n}$. This means that the following is also a k-basis of a representant system of $k^{[n]}/\mathrm{Im}(N)$: $\{x_n^{p-1} x_1^{pa_1} \cdots x_n^{pa_n} \mid a_i \in \mathbb{N}\}$. In other words, we can pick $x_n^{p-1} k[x_1^p, \ldots, x_n^p]$ as a representant system, and the result follows. $\qquad\square$

Corollary 3.18. *(of theorem 3.3 and lemma 3.16.) Let $F = (x_1 + f_1, \ldots, x_n + f_n)$ such that $\mathrm{ord}(F) = p^n$. Then F is equivalent to exactly one $G = (x_1 + f_1', \ldots, x_{n-1} + f_{n-1}', x_n + f_n)$ where $f_i' \in x_n^{p-1}(k^{[n]})^G$ if $1 \le i \le n-1$.*

Proof. By induction. The theorem is true for $n = 1$. Assume the theorem is true for $n - 1$. This means we can assume $F = (x_1 + f_1, x_2 + f_2', \ldots, x_n + f_n')$, and that conjugation by $(x_i + g_i)$ where $2 \le i \le n$ cannot be used anymore as it disturbes the form. Now, conjugation by $(x_1 + g_1)$ makes it possible to change f_1 by elements of $\mathrm{Im}(N)$. Using lemma 3.16 we see that we can change to exactly one $f_1' \in x_n^{p-1} \mathrm{inv}(F)$. $\qquad\square$

The below is actually a corollary of corollary 3.18 and lemma 3.17, but since the result is the most elegant one, we call it "theorem":

Theorem 3.19. *Let $F = (x_1 + f_1, \ldots, x_n + f_n)$ such that $\mathrm{ord}(F) = p^n$. Then F is equivalent to exactly one $G = (x_1 + f_1', \ldots, x_{n-1} + f_{n-1}', x_n + f_n)$ where $f_i' \in x_n^{p-1}k[x_1^p, \ldots, x_n^p]$ if $1 \le i \le n-1$.*

Proof. The result follows directly from corollary 3.17, with the same proof as 3.18 required. □

Warning: The above theorem does NOT state that any sequence $f_1', \ldots, f_{n-1}', f_n$ where $f_i' \in x_n^{p-1}k[x_1^p, \ldots, x_n^p]$ if $1 \le i \le n-1$, $f_n \in k^*$ gives a map of order p^n.

Remark 3.20. In [11] a similar case was done for $k = \mathbb{F}_p$ a finite field with p elements, and then not considering the triangular automorphisms, but the permutations $\mathbb{F}_p^n \longrightarrow \mathbb{F}_p^n$ induced by them. It was shown that if $F = (x_1 + f_1, \ldots, x_n + f_n)$ and additionally, $\deg_{x_i}(f_j) \le p - 1$ for each i, j (which you may assume as you're only interested in the map), then F is of order p^n if and only if for all $1 \le i \le n$, the coefficient of $(x_{i+1} \cdots x_n)^{p-1}$ is nonzero. This statement is *not* in contradiction with corollary 3.19 - an example is $(x + y^p z^{p-1}, y + z^{p-1}, z + 1)$ which has order p^3, but restricted to $\mathbb{F}_p^3$ has order p^2.

Question 3.21. The representation of 3.19 is pleasing to the eye, but perhaps not the best if your goal is to actually iterate such an element. In characteristic zero, we can write $F = \exp(D)$ and then $F^n = \exp(nD)$. It would be nice if there's a way to have such an elegant description of iterates of F in this characteristic p case too.

4 Characteristic p: the generic 2-variable case

In this whole section 4, k is still of characteristic p.

4.1 Conjugacy classes of $\mathrm{BAs}_2(k)$

The previous section now makes it easy to determine the conjugacy classes of elements in $\mathrm{BAs}_2(k)$. (We shun away from the generic ring case, as it would require a deeper understanding of the map N over rings. In particular, we need to understand how lemma 3.16 behaves in that case.) There are only three types of maps: order 1 (the identity), order p, and order p^2. Order p^2 is taken care of by theorem 3.19, while order p is taken care of by lemma 3.10. Gathering up these results we get with very little extra effort the following corollary:

Corollary 4.1. *The conjugacy classes in $\mathrm{BAs}_2(k)$ are*

- *$(x, y + \lambda)$ where $\lambda \in k$,*

- *$(x + f(y), y)$ where $f \ne 0$, and a unique representant in $k[y]/k$ under the action $k \times k[y] \longrightarrow k[y]$ given by $\lambda \cdot g(y) \longrightarrow g(y + \lambda)$,*

- *$(x + y^{p-1}f(y^p), y + \lambda)$ where $\lambda \in k^*$ and $f \ne 0$.*

4.2   Conjugacy classes of $\mathrm{BAs}_2(k)$ in $\mathrm{BA}_2(k)$

Now the result of corollary 4.1 can now be slightly improved:

Corollary 4.2. *The conjugacy classes in* $\mathrm{BAs}_2(k)$ *are*

- (x, y),

- $(x, y + 1)$,

- $(x + f(y), y)$ *where* $f \neq 0$ *and monic, and a unique representant in* $k[y]$ *under the action* $\mathrm{GA}_1(k) \times k[y] \longrightarrow k[y]$ *given by* $(\mu y + \lambda) \cdot g(y) \longrightarrow g(\mu y + \lambda)$,

- $(x + y^{p-1} f(y^p), y + 1)$ *where* $\lambda \in k^*$ *and* $f \neq 0$ *monic.*

Proof. The proof is very similar to the proof of 2.11, and in general easier since we're sticking with fields. We just give a brief sketch of the two main points:
(1) "monic" in the fourth and third bullet point we get by conjugating with (λx) for suitable λ.
(2) The third bullet point is the only one where conjugation by $(\mu y + \lambda)$ does not change the second component, but can change the first component. Here, the "characteristic p" shows its head: we cannot use $y \longrightarrow y + \lambda$ to make sure that the $d - 1$-th coefficent of f is zero (where $d = \deg(f)$). So, we're essentially stuck in stating that f should be picked unique under its equivalent forms under $y \longrightarrow \mu y + \lambda$. $\qquad\square$

4.3   Conjugacy classes of $\mathrm{BA}_2(k)$

Theorem 4.3. *Let* k *be a field of characteristic* p. *Then the conjugacy classes of* $F = (F_1, F_2)$ *in* $\mathrm{BA}_2(k)$ *are*

- *Affine,*

- $(ax + f(y), y)$ *where* $a \in k^*$, $f \neq 0$ *and monic, and a unique representant in* $k[y]$ *under the action* $\mathrm{GA}_1(k) \times k[y] \longrightarrow k[y]$ *given by* $(\mu y + \lambda) \cdot g(y) \longrightarrow g(\mu y + \lambda)$,

- $(x + y^{p-1} f(y^p), y + 1)$ *where* $\lambda \in k^*$ *and* $f \neq 0$ *monic.*

- $(a^d x + y^d f(y^m), ay)$ *where* $m = \mathrm{ord}(a)$ *(*$m = 0$ *if* a *no root of unity). *$f \neq 0$ *is monic.*

Proof. There are overlaps with the characteristic zero case, but we still give a complete proof. If $F = (F_1, F_2)$ then we split up the cases we get from lemma 2.15: $F_2 = y, F_2 = y + 1$ and $F_2 = ay$ where $a \neq 1$.
(1) y: Now $F_1 = ax + f(y)$. Then we can make f monic by conjugating with (μx). Then, we can conjugate by $(\mu y + \lambda)$, and the result follows. which forces us where $f(y)$ is monic. Again we can now conjugate by $y \longrightarrow \mu y + \lambda$ and we have no simpler

21

way than just stating this.

(2) $y + 1$: now $F = (ax + f(y), y + 1)$. If $a = 1$, then we're in the case of corollary 4.2, and $F = (x + y^{p-1}f(y^p), y + 1)$ where f is monic or zero. If $a \neq 1$ then we need to consider what happens by conjugation with $(x + g(y))$. Then we can change f by $ag(y) - g(y+1)$ and thus we need to understand the map $g(y) \longrightarrow ag(y) - g(y+1)$. This map sends y^d to $(a - 1)y^d$+lower order terms. Since $a - 1$ is invertible (as $a \neq 1$), it is clear that this map is surjective (and actually bijective). Thus, we can conjugate to $(ax, y + 1)$ which is affine.

(3) ay: If we conjugate $(bx + f(y), ay)$ by $(x + g(y))$ then we are modifying f by $bg(y) - g(ay)$. We thus need to understand the map $g(y) \longrightarrow bg(y) - g(ay)$, which preserves monomials. The image of y^d is $(b - a^d)y$, so we need to know when $b = a^d$; if this never happens, then we can conjugate to (bx, ay) which is affine. So, write $b = a^d$ where $d \in N^*$ as minimal as possible. If $\text{ord}(a) = m$ (automatically $d < m$) then we get $(bx + y^d f(y^m), ay)$. If a is no root of unity we get $(bx + \lambda y^d, ay)$. In both cases, we can conjugate by (μx), and make sure that we have monic polynomials. $\square$

5 Finite order automorphisms in $\text{BAs}_2(k)$

In this section, we classify the finite order automorphisms for all fields k. We give a stand-alone proof (except for reference to 3.8) even though we could use the previous sections, and a slightly different classification, as we expect this to be of high interest.

Lemma 5.1. *Let $F \in \text{GA}_2(k)$ be of finite order $\text{ord}(F) = s$, i.e. $F^s = I$. Then F can be conjugated to the following standard forms (unique up to as stated):*

A *(affine)*
   An affine map (up to conjugation within the affine group)

U *(unipotent) If characteristic k is p,*
   $(x + y^{p-1}f(y^p), y + 1)$ where $f(y^p) \in k[y^p]$ monic nonzero.
   $\text{ord}(F) = p^2$.

M *(mixed)*
   $(x + f(y^m), ay)$ where $\text{ord}(a) = m$. $\text{ord}(F) = \text{lcm}(p, m)$.

S *(sequential)*
   $(a^m x + y^m f(y^{lm}), ay)$ where $a \in k^*$ satisfy $a^{ml} = 1$ for some $m, l \in \mathbb{N}^*$ (m chosen as small as possible, $l = \text{ord}(a^m) > 0$) and $f(y^{ml}) \in k[y^{ml}]$ nonzero, and unique up to substitution $y \longrightarrow \lambda y$ where $\lambda \in k^*$. $\text{ord}(F) = ml$.

Proof. If $F \in \text{GA}_2(k)$ of finite order, then using the Jung-v/d Kulk theorem it is easy to prove that F must be up to conjugation either in $\text{Aff}_2(k)$ or $\text{B}_2(k)$. We thus assume $F \in \text{B}_2(k) \backslash \text{Aff}_2(k)$. If one conjugates F by some $G \in \text{GA}_2(k)$ and $G^{-1}FG \in \text{B}_2(k)$, then $G \in \text{B}_2(k)$ because of the same reason. We thus can consider the conjugacy classes within $\text{B}_2(k)$.

Let us write $F = (ax + f(y), by + c)$. Since we are conjugating within $B_2(k)$, we can first choose a unique form for $by + c$ within $B_1(k)$. If $b \neq 1$ then one can conjugate by $(y - \lambda)(by + c)(y + \lambda) = (by + c + (b - 1)\lambda$ so choosing $\lambda = -c(b-1)^{-1}$ we get the standard form by. In case $b = 1$, $c \neq 0$ then we can conjugate $(c^{-1}y)(y + c)(cy) = (y + 1)$. Thus, we can assume $c = 0$ or $b = 1$.

Case $c = 0$

$F = (ax + f(y), by)$. Conjugate by $(x - g(y), y)(ax + f(y), by)(x + g(y), y) = (ax + f(y) + (ag(y) - g(by)), by)$. If $b^m \neq a$ for any $m \in \mathbb{N}$, then we can choose $g(y)$ such that $f(y) + (ag(y) - g(by)) = 0$ and achieve (ax, by), an affine map. So: **We may assume** $b = a^m$ for some $m \in \mathbb{N}$. In that case, the above conjugation can standardize $f(x)$ to a polynomial which is a linear combination of monomials x^n such that $(bx)^n = ax^n$.

Assume $b \neq 1$, i.e. $l := \mathrm{ord}(b) > 0$. Then $n \in m + ml\mathbb{Z}$ where l is such that $(a^m)^l = 1$, i.e. we get $f(x) \in x^m k[x^{ml}]$. We can change this form a bit by conjugation with $(x, \lambda y)$, but we claim that beyond this, the form is unique: if we conjugate by $(dx + g(y), ey + \lambda)$ then we see that $\lambda = 0$ otherwise the form changed. We can write $(dx + g(y), ey) = (x, ey)(dx, y)(x + d^{-1}g(y), y)$. We may ignore the conjugation by (x, ey). The conjugation by (dx, y) does not change the form. Then, the conjugation by $(x + d^{-1}g(y), y)$ either changes the form, or leaves it invariant (in case $g(y) \in y^m k[x^{lm}]$). Thus, this gives form **S**.

Subcase $b = 1$: we can get the form as in case **M**. It is easy to check that this form cannot be improved by a conjugation within $B_2(k)$.

Case $b = 1$, $c \neq 0$: We thus can assume $F = (ax + f(y), y + 1)$. We can conjugate by $(x + g(y), y)(ax + f(y), y + 1)(x - g(y), y) = (ax + f(y) - ag(y) + g(y + 1), y + 1)$. In case $a \neq 1$, then the map $E : k[y] \longrightarrow k[y]$ given by $g(y) \longrightarrow -ag(y) + g(y + 1)$ is surjective (as $\deg(E(x^m)) = m$). However, in case $a = 1$, then we are considering the map $N : f(x) \longrightarrow f(x + 1) - f(x)$ from lemma 3.8. We can thus change $f(x)$ by elements of $\mathrm{Im}(N)$. That same lemma shows that a representant system of $k[x]/\mathrm{Im}(N)$ is $x^{p-1}k[x^p]$, so we may assume $f(x)$ is in here. We can conjugate by (dx, y) to make sure that f is monic. We have obtained the form **U**. We claim that this form $(x + y^{p-1}f(y^p), y + 1)$ is unique. The argument is similar as before: if one conjugated by $(dx + g(y), ey + \lambda)$ then $\lambda = 0, e = 1$ otherwise the form $y + 1$ is destroyed, and then $(dx + g(y), y)$. Conjugating with this either destroys the standard form or leaves it invariant. $\qquad \square$

6 Further research

We gather up a list of future research questions.

- For rings R containing $\mathbb{Q}$, determine the conjugacy classes of $\mathrm{BAs}_3(R)$ and $\mathrm{BA}_3(R)$. Determine the conjugacy classes of $\mathrm{BAs}_4(k)$ and $\mathrm{BA}_4(k)$ where k is a field of characteristic zero.

- For rings R not containing $\mathbb{Q}$ (i.e. having prime ring $\mathbb{Z}$ or $\mathbb{Z}_n$), determine the conjugacy classes of $\mathrm{BAs}_2(R)$ and $\mathrm{BA}_2(R)$.

- Find representants of the conjugacy classes which are relatively easy to iterate. In characteristic zero, we can do this by noticing that $\exp(mD) = F^m$, but in characteristic p we have an open question here for the elements of $\mathrm{BAs}_2(k)$, and for maximal order maps in $\mathrm{BAs}_n(k)$. See also [11].

- If $\mathrm{char}(k) = p$, $F \in \mathrm{GA}_n(k)$ satisfies $F^{p^n} = I$ (i.e. F is unipotent), are there similar theorems as 3.3 and 3.4 true for this general case?

References

[1] Bass, Hyman; Maubach, Stefan; Van Chau, Nguyen; Lecture notes from the International School and Workshop (ICPA2006) held in Hanoi, October 9–20, 2006. Publishing House for Science and Technology, Hanoi, 2007. xii+160 pp.14-06

[2] Furter, Jean-Philippe; Maubach, Stefan *Locally finite polynomial endomorphisms*, J. Pure Appl. Algebra 211 (2007), no. 2, 445-458

[3] Freudenburg, Gene; *A counterexample to Hilberts Fourteenth Problem in dimension six.* Transform. Groups 5 (2000) 61-71

[4] Daigle, Daniel; Freudenburg, Gene; *A counterexample to Hilbert's fourteenth problem in dimension 5.* J. Algebra 221 (1999), no. 2, 528–535.

[5] Kraft, Hanspeter; Stampfli, Immanuel *On Automorphisms of the Affine Cremona Group.* arxiv:1105.3739

[6] Roberts, P.; *An infinitely generated symbolic blow-up in a power series ring and a new counterexample to Hilbert's fourteenth problem.* J. Algebra, 132 (1990), pp. 461473

[7] Ostafe, Alina; *Pseudorandom vector sequences of maximal period generated by triangular polynomial dynamical systems.* Des. Codes Cryptogr. 63 (2012), no. 1, 5972.

[8] Ostafe, Alina; Pelican, E.; Shparlinski, I.E.;*On pseudorandom numbers from multivariate polynomial systems.* Finite Fields and Their Appl. (2010), v.16, 320-328.

[9] Nakai, Yoshikazu; *On locally finite iterative higher derivations.* Osaka J. Math. Volume 15, Number 3 (1978), 655-662.

[10] Tanimoto, Ryuji; *An algorithm for computing the kernel of a locally finite iterative higher derivation.* Journal of Pure and Applied Algebra 212 (2008, no. 10, 2284-2297

[11] Maubach, Stefan; *Triangular polynomial maps in characteristic p.* accepted to J. Algebra and Number Theory.