PDMI PREPRINT — 6/1998

Bijective arithmetic codings of hyperbolic automorphisms of the 2-torus, and binary quadratic forms

Nikita SIDOROV and Anatoly VERSHIK

Russian Academy of Sciences
Steklov Institute of Mathematics at St. Petersburg
Fontanka 27, St. Petersburg 191011, Russia
E-mail: sidorov@pdmi.ras.ru, vershik@pdmi.ras.ru

February 1998

ABSTRACT

We study arithmetic codings of hyperbolic automorphisms of the 2-torus, i.e. mappings acting from the symbolic space of sequences with a finite alphabet endowed with an appropriate structure of additive group onto the torus which preserves this structure and turns the two-sided shift into an automorphism of the torus. The necessary and sufficient condition of the existence of a bijective arithmetic coding is obtained; it is formulated in terms of a certain binary quadratic form constructed by means of a given automorphism. Furthermore, we describe all bijective arithmetic codings in terms the Dirichlet group of the corresponding qudratic field. The arithmetic minimum of that binary quadratic form in the general case equals the minimal possible number of preimages for a.e. point of the torus under such an arithmetical coding.

Supported by the RFBR grant 96-01-00676 and the INTAS-RFBR grant 95-418

0. Introduction

In this work we continue studying symbolic dynamics of ergodic automorphisms of the 2-torus. The dynamics of automorphisms of the torus is related more to number theory than to the general theory of dynamical systems. This is why their coding should be considered as a number-theoretic problem. This was the main idea of [Ver2] and subsequent papers (see [Ver1] references therein), also [KenVer] and [Leb]. Recall that to the Markov coding of hyperbolic automorphisms of the torus and more general hyperbolic dynamical systems a number of classical works have been devoted, see, e.g., [AdWe], [Sin], [Bow], [GuSi], but without special interest to the arithmetic structure.

The quadratic case is exhaustively studied below, and one sees that the relationship with the theory of quadratic extensions and binary integral quadratic forms becomes even deeper than before. We set certain natural requirements on a symbolic realization of a hyperbolic automorphism of the 2-torus (more precisely, on the maximal commutative subgroup of $GL(2,\mathbb{Z})$ containing this automorphism) and give the necessary and sufficient condition on the existence of an a.e. bijective mapping from a symbolic compactum onto the 2-torus, under which the two-sided shifts turns into an automorphism, and the sum of a.e. pair of sequences from a symbolic compactum turns into the sum of their images. We will call this kind of symbolic codings *arithmetic codings.*

The symbolic compactum in question is either Markov, if the determinant of an automorphism equals -1, or sofic otherwise. It is proved that in both cases the compactum, after a certain factorization of sequences of zero measure, turns into a group in addition. In the case, where for a given automorphism there is no bijective arithmetic coding, we present a precise description of some minimal finite-sheeted covering of the torus. A close connection with number theory that we mentioned above is corroborated by the type of existence condition. We quote the essential part of the main result, see Theorem 2.6 and Proposition 2.7. We define for the automorphism T given by a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ the associated quadratic form $f_T(x,y) := bx^2 - (a-d)xy - cy^2$.

Theorem.

I. *An ergodic automorphism T admits bijective arithmetic coding if and only its matrix M is $GL(2,\mathbb{Z})$-similar to the companion matrix $C_r = \left(\begin{smallmatrix} r & 1 \\ -\sigma & 0 \end{smallmatrix}\right)$ with $r = a+d$ and $\sigma = \det M$.*

II. *A matrix $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is $GL(2,\mathbb{Z})$-similar to the corresponding companion matrix if and only if the Diophantine equation*

$$f_T(x,y) = \pm 1$$

is solvable.

III. *There exists a natural one-to-one correspondence between the set of bijective arithmetic codings of T and the Dirichlet group of the quadratic field $\mathbb{Q}(\lambda)$, λ being an eigenvalue of M.*

More generally, a minimal arithmetic coding, i.e. a coding with the minimal number of preimages, is naturally given by a solution of the Diophantine equation $f_T(x,y) = \pm m$ with the minimal possible positive m.

The axiomatic conditions on a symboilc realization of an automorphism of the torus are very natural: the corresponding mapping from the symbolic set of all sequences of nonnegative integers onto the 2-torus should be a continuous homomorphism of semigroups turning the shift into a given automorphism. A priori it is not even clear, why so rigid conditions can be satisfied.

However, the fact that they really can, yields to a purely arithmetic interpretation of a coding, namely, as two-sided convergent power series with respect to the eigenvalue with a specifically chosen collection of digits and Markov or sofic restrictions to their succession. This is nothing but a two-sided generalization of the so-called β-expansions but with essential sharpenings connected with the requirement of bijectivity.

A good deal of what was said above, might be extended to the general case of a hyperbolic automorphism of $\mathbb{T}^n$, $n \geq 3$ whose principal eigenvalue is a PV number but in the present paper we do not plan to deal with this kind of problems. Let us emphasize that we are interested in the precise conditions of the existence of a bijective arithmetic coding, not in constructing a bounded-to-one coding. Note also that for constructing examples which corroborate some sharp estimates, we will use the facts from the theory of indefinite quadratic forms [Cas].

1. Basic notions and the main problem

1.1. Basic notions and constructions. Let $\mathbb{T}^2$ denote the 2-torus considered as the factor $\mathbb{R}^2/\mathbb{Z}^2$. Let T be an arbitrary group automorphism of $\mathbb{T}^2$ given by a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in GL(2,\mathbb{Z})$ which we will denote by M. Suppose T is hyperbolic, which in the two-dimensional case is equivalent to the fact that none of the roots of 1 belongs to the spectrum of M. Let r denote the trace of M, σ stand for its determinant. Then the characteristic polynomial of M is $x^2 - rx + \sigma$, and its dicriminant is $D = r^2 - 4\sigma$. The eigenvalues of M are $\frac{1}{2}(r \pm \sqrt{D})$, and the condition of hyperbolicity is equivalent to the following ones:

(1) if $\sigma = -1$, then $|r| \geq 1$;
(2) if $\sigma = +1$, then $|r| \geq 3$.

Suppose r to be positive (below we will prove that for our purposes the study of the case of a negative trace will be immediate, see the end of Section 3). Let $\lambda = \frac{r+\sqrt{D}}{2} > 1$. We wish to consider symbolic codings as appropriate expansions of the points of the torus in the sense of some generalized "number system" with the natural coefficients.[1] Thus, as a primary symbolic set of coefficients for further codings we choose $\widetilde{\mathfrak{X}} := \prod_{-\infty}^{\infty} \mathbb{Z}_+$.

Obviously, $\widetilde{\mathfrak{X}}$ has the natural (weak) topology and the natural structure of an additive semigroup.

We call a sequence *finite*, if it contains only a finite number of nonzero coordinates. Let τ denote the two-sided shift on $\mathfrak{X}$, i.e. $\tau\{\varepsilon_n\} = \{\varepsilon'_n\}$, where $\varepsilon'_n = \varepsilon_{n+1}$. We set up the main problem of arithmetic coding axiomatically.

Problem 1. (1) To find a mapping $\varphi : \widetilde{\mathfrak{X}} \to \mathbb{T}^2$ such that it is a continuous homomorphism of the semigroups turning the shift τ into a given automorphism T.
(2) Fixing a mapping φ from item (1), to find a shift-invariant subset $\mathfrak{X}$ of $\widetilde{\mathfrak{X}}$ such that it contains the basis sequence u_0 having all 0's and 1 at the zero place, and φ restricted to $\mathfrak{X}$ (i.e. the *section* φ) is surjective and one-to-one on the set of finite sequences of $\widetilde{\mathfrak{X}}$ belonging to $\mathfrak{X}$.

Definition. We will call a mapping satisfying the conditions described above, an *arithmetic coding* of a given automorphism.

Below we will prove that an arithmetic coding of an automorphism always exists and will give a complete description of them. Besides, we will show that $\mathfrak{X}$ is a compact set also having a natural additive structure. Let us state the second problem which is an extension of Problem 1.

[1] Note that for multidimensional hyperbolic automorphisms the coefficients are not necessarily naturals, but always elements of a certain algebraic field, see [KenVer].

Problem 2. Consider all arithmetic codings of T. The problem consists in
(1) giving the necessary and sufficient condition of the existence of an arithmetic coding being one-to-one a.e. with respect to the Haar measure and in describing all bijective arithmetic codings (BAC);
(2) finding a minimal arithmetic coding (MAC) defined as a coding with the minimal possible number of preimages and to describe all such codings in the case, when it is impossible to find a bijective arithmetic coding.

We are going to solve the main problem in this section and to devote two subsequent ones to the precising problem.

Definition. Let $\|x\| := \min\{|x-n| : n \in \mathbb{Z}\}$. A two-sided series $\sum_{-\infty}^{\infty} u_n$ is said to converge to $u \in \mathbb{R}$ *modulo 1*, if $\|\sum_{n=-k}^{l} u_n - u\| \to 0$ as $k, l \to +\infty$. The convergence of a pair of series *modulo* $\mathbb{Z}^2$ means the convergence of each coordinate modulo 1.

Let $\mathfrak{X}_r$ denote the stationary *Markov compactum* with the state space $0, 1, \dots, r$ and the pairwise restrictions $\{\varepsilon_n = r \Rightarrow \varepsilon_{n+1} = 0,\ n \in \mathbb{Z}\}$, and the *sofic compactum* (see, e.g., [Ver1]) $\mathfrak{Y}_r$ with the state space $0, 1, \dots, r-1$, and $\varepsilon_n \dots \varepsilon_{n+s} \neq (r-1)(r-2)^{s-2}(r-1)$ for any $n \in \mathbb{Z}$ and any $s \geq 2$.

The following theorem solves Problem 1.

Theorem 1.1.

(I) *An arithmetic coding of a hyperbolic automorphism T of the 2-torus is always given by the convergent series*

$$\varphi(\varepsilon) = \left(\sum_{n=-\infty}^{\infty} \varepsilon_n \lambda^{-n}\right) \begin{pmatrix} \xi \\ \eta \end{pmatrix} \mod \mathbb{Z}^2, \tag{1}$$

where (ξ, η) is a homoclinic point for T.[2] *Conversely, if (ξ, η) is a homoclinic point for T, mapping (1) is a convergent series and specifies a bijective arithmetic coding for T.*

(II) *A shift-invariant subset $\mathfrak{X}$ satisfying the condition (2) of Problem 1, i.e. the bijectivity of the mapping $\varphi|_{\mathfrak{X}}$ on the set of finite sequences for any φ given by formula (1), is the stationary Markov compactum $\mathfrak{X}_r$ for* $\det M = -1$ *or the sofic compactum $\mathfrak{Y}_r$ for* $\det M = +1$.

Proof. (I) Let φ satisfy the conditions of the main problem, and $u_k = \tau^k(u_0)$. We set $(\xi, \eta) := \varphi(u_0)$. By virtue of the continuity of the mapping φ, as $u_k \to 0$, $k \to \pm\infty$, we have $T^k(\xi, \eta) \to (0, 0)$, $k \to \pm\infty$, whence by definition (ξ, η) is a homoclinic point. Hence $\varphi(u_k) = \lambda^k(\xi, \eta)$ mod $\mathbb{Z}^2$ for all $k \in \mathbb{Z}$ or $\varphi(u_k) = \overline{\lambda}^k(\xi, \eta)$ mod $\mathbb{Z}^2$ for all $k \in \mathbb{Z}$, where $\overline{\lambda}$ is the algebraic conjugate of λ, i.e. $\overline{\lambda} = \sigma\lambda^{-1}$.

Consider now an arbitrary finite sequence $\varepsilon \in \widetilde{\mathfrak{X}}$. By the additivity of φ, its image is also a homoclinic point for T, and either

$$\varphi(\varepsilon) = \left(\sum_{n=-\infty}^{\infty} \varepsilon_n \lambda^{-n}\right) \begin{pmatrix} \xi \\ \eta \end{pmatrix} \mod \mathbb{Z}^2$$

[2] Recall that this means by definition that (ξ, η) belongs to the intersection of the leaves of the stable and unstable foliations for T going through $(0, 0)$.

or

$$\psi(\varepsilon) = \left(\sum_{n=-\infty}^{\infty} \varepsilon_n (\overline{\lambda})^{-n}\right)\begin{pmatrix}\xi\\ \eta\end{pmatrix} \mod \mathbb{Z}^2.$$

To exclude the second mapping from our consideration, we observe that $\xi\overline{\lambda}^k = -\overline{\xi}\lambda^k \mod 1$, whence $\psi = \psi(\xi,\eta) = \varphi(-\overline{\xi},\overline{\eta})$. Since (ξ,η) being a homoclinic point implies the same for $(-\overline{\xi},-\overline{\eta})$, it suffices to deal with φ.

Extending the mapping φ by continuity to all sequences $\varepsilon \in \widetilde{\mathfrak{X}}$, we come to the first claim of the theorem. The convergence of the series involved modulo 1 is implied by (ξ,η) being a homoclinic point.

(II) Suppose now $\varphi|_{\mathfrak{X}}$ (we will keep the same letter φ for this restriction) to be bijective on the set of finite sequences for some subset $\mathfrak{X}$ of $\widetilde{\mathfrak{X}}$. Note first that in this case the ε_k must be uniformly bounded. More precisely, as $r\lambda^{-1} - \sigma\lambda^{-2} = 1$, we have $0 \le \varepsilon_n \le r$ for $\sigma = -1$ and $0 \le \varepsilon_n \le r-1$ for all $n \in \mathbb{Z}$. Then for any finite sequence $\varepsilon = (\dots,0,0,\dots,0,\varepsilon_2,\varepsilon_3,\dots,\varepsilon_n,0,0,\dots) \in \prod_{-\infty}^{\infty}\{0,1,\dots,r\}$, it is a direct inspection that $x = \sum_2^n \varepsilon_k\lambda^{-k} \in [0,1)$. Let $\lambda^{-1} = \sum_2^{\infty}\varepsilon_k^{(0)}\lambda^{-k}$ be the maximal expansion of λ^{-1}, i.e. the one with the digits computed by the greedy algorithm, namely, $\varepsilon_2^{(0)} = [\lambda], \varepsilon_3^{(0)} = [\lambda\{\lambda\}]$, etc. Without loss of generality we may suppose $\xi \in (0,1)$. Now by the standard argument of the of the well-known theory of the β-expansions (for $\beta = \lambda$), for the uniqueness of the mapping φ on the set of finite sequences, any finite sequence $(\varepsilon_2,\varepsilon_3,\dots)$ must be lexicographically less than $(\varepsilon_2^{(0)},\varepsilon_3^{(0)},\dots)$, i.e. $\varepsilon_k < \varepsilon_k^{(0)}$ for the least $k \ge 2$ such that $\varepsilon_k \ne \varepsilon_k^{(0)}$. By the assumed stationarity of the compactum $\mathfrak{X}$, this is true for any $k \in \mathbb{Z}$.

Thus, for $\sigma = -1$, we have $(\varepsilon_k,\varepsilon_{k+1},\dots) \prec (r,1,0,0,\dots)$, and $(\varepsilon_k,\varepsilon_{k+1},\dots) \prec (r-1,r-2,r-2,r-2,\dots)$ for $\sigma = +1$. Hence $\mathfrak{X}$ is a subset of $\mathfrak{X}_r$ for $\sigma = -1$ and of $\mathfrak{Y}_r$ otherwise. However, if these inclusions were proper, φ would be not surjective. $\square$

1.2. Additive properties of the symbolic compacta. The Markov and sofic compacta defined above, do not form a subsemigroup of the semigroup $\widetilde{\mathfrak{X}}$. Neverthelss, we can introduce a group structure after certain small glueings of some sequences. Let $\mathfrak{X}_r^{(0)}$ and $\mathfrak{Y}_r^{(0)}$ denote the subsets of $\mathfrak{X}_r$ and $\mathfrak{Y}_r$ respectively consisting of all finite sequences. We have shown above that $\mathfrak{X}_r^{(0)}$ is in fact the factor of the semigroup $\mathfrak{X}^{(0)} = \{\sum_{|n|<N}\varepsilon_n u_n \mid \varepsilon_n \ge 0,\ N \in \mathbb{N}\}$ with respect to the stationary recurrence relations $\{u_{n-1} = ru_n + u_{n+1},\ n \in \mathbb{Z}\}$. Similarly, $\mathfrak{Y}_r^{(0)}$ is the factor of $\mathfrak{X}^{(0)}$ with respect to the relations $\{u_{n-1} = ru_n - u_{n+1},\ n \in \mathbb{Z}\}$.

It is well-known that in both cases in question the finite sequences themselves form an additive semigroup (see, e.g., [FrSa]).[3] Our goal now consists in assigning the structure of an additive group to the whole symbolic compacta.

To do this, we first give the well-known definition of normalization (see [Fr]).

Definition. Let $x \in \prod_1^{\infty}\mathbb{Z}_+,\ x = \{x_k\}_{k=1}^{\infty}$; we define $c(x) = \sum_{k=1}^{\infty} x_k\lambda^{-k} = \sum_{k=k_0}^{\infty}\varepsilon_k\lambda^{-k}$, where $\{\varepsilon_k\}$ is the β-expansion of $c(x)$, i.e. the expansion whose digits are given by the greedy algorithm. Thus, $(\varepsilon_{-k_0},\varepsilon_{-k_0+1},\dots)$ belongs to the symbolic compactum $\mathfrak{X}_r$ or $\mathfrak{Y}_r$ respectively. We define

$$\mathfrak{n}(x) := \varepsilon = \{\varepsilon_k\}_{k=k_0}^{\infty}.$$

The operation $\mathfrak{n}$ is called the *normalization* of a sequence.

With the help of normalization we can now define addition and subtraction on the symbolic compacta. Let the elements of a sequence x from the definition of normalization are uniformly

[3]It is not true for more general relations, for example for $u_{n-3} = u_n + u_{n+1}$.

bounded, for instance, $0 \le x_k \le 2r$. Then for the cases in question (i.e. for the quadratic units) it is known that similarly to addition, the normalization of a finite sequence is also finite and the carry to both sides is uniformly bounded, see [FrSo]. Thus, it is easy to define the *two-sided normalization* of almost every sequence with respect to any shift-invariant measure being positive on each cylinder (see [SidVer] for the precise procedure in the case $\lambda = \frac{\sqrt{5}+1}{2}$). The below definition is based on the following consideration. In the Markov case the sequence $(\dots, r, 0, r, 0, \dots)$ is by continuity identified with the zero sequence, the same is true for the sofic case with the sequence $(\dots, r-2, r-2, r-2, \dots)$.

Definition.

(1) Let ε and ε' belong to the compactum $\mathfrak{X}_r$ (or resp. $\mathfrak{Y}_r$). Define the sequence $x = \{x_k\}_{-\infty}^{\infty}$ as $x_k = \varepsilon_k + \varepsilon'_k$. Then the sum of ε and ε' is by definition the two-sided normalization of x.

(2) Similarly, to define $-\varepsilon$, consider the sequence $y = \{y_k\}$ with $y_k = r - \varepsilon_k$ for the Markov case and $y_k = 2(r-2) - \varepsilon_k$ for the sofic case. By definition, $-\varepsilon$ is the two-sided normalization of y.

Remark. From the above considerations it is clear that both operations are well defined for a.e. pairs of sequences with respect to the square of any stationary measure on $\mathfrak{X}_r$ or $\mathfrak{Y}_r$ which is positive on any cylindric set $\{\varepsilon_0 = i_0, \dots, \varepsilon_n = i_n\}$.

Examples. In the Markov compactum $\mathfrak{X}_r$ we have

$$\begin{aligned} -u_k &= ru_{k-1} + ru_{k-3} + ru_{k-5} + \dots = u_{k+1} + (r-1)u_k + ru_{k-2} + ru_{k-4} + \dots, \\ u_k &= ru_{k+1} + ru_{k+3} + ru_{k+5} + \dots \end{aligned}$$

In the sofic compactum $\mathfrak{Y}_r$,

$$\begin{aligned} -u_k &= \dots + (r-2)u_{k-2} + (r-2)u_{k-1} + (r-3)u_k + (r-2)u_{k+1} + (r-2)u_{k+2} + \dots, \\ u_k &= (r-1)u_{k+1} + (r-2)u_{k+2} + (r-2)u_{k+3} + \dots \\ &= (r-1)u_{k-1} + (r-2)u_{k-2} + (r-2)u_{k-3} + \dots \\ &= \dots + (r-2)u_{k-2} + (r-2)u_{k-1} + (r-1)u_k + (r-2)u_{k+1} + (r-2)u_{k+2} + \dots \end{aligned}$$

The above definition of addition and subtraction in the symbolic compacta was rather rough from the viewpoint of sequences involved. A more detailed analysis yields the following claim.

Proposition 1.2. *Let $\mathfrak{X}'_r$ and $\mathfrak{Y}'_r$ denote the factor sets $\mathfrak{X}_r/\mathfrak{R}_1$ and $\mathfrak{Y}_r/\mathfrak{R}_2$, where*
(1) $\mathfrak{R}_1$ is the identification of the pairs of sequences $(*k, r, 0, r, \dots)$ and $(*k+1, 0, 0, 0, \dots)$, and $(\dots 0, r, 0, r, 0, k*)$ and $(\dots r, 0, r, 0, r-1, k+1*)$, where $*$ denotes the same arbitary admissible tail, and $0 \le k \le r-1$.
(2) $\mathfrak{R}_2$ is the identification of the pairs of sequences $(*k, r-1, r-2, r-2, \dots)$ and $(*k+1, 0, 0, 0, \dots)$, and $(\dots 0, r-2, r-2, r-2, r-1, k*)$ and $(\dots 0, 0, 0, 0, k+1*)$, $0 \le k \le r-2$. Then
(1) The factor sets $\mathfrak{X}'_r$ and $\mathfrak{Y}'_r$ are groups in addition;
(2) any mapping φ from the definition of arithmetic coding is well defined on the factor sets and is a group homomorphism of the groups $\mathfrak{X}'_r$ (resp. $\mathfrak{Y}'_r$) and $\mathbb{T}^2$.

Sketch of the proof. The crucial point is the study of the relations $u_{n-1} = ru_n + u_{n+1}$, $n \in \mathbb{Z}$ for the Markov case and $u_{n-1} + u_{N+1} = (r-1)u_n + (r-2)u_{n+1} + \dots + (r-2)u_{N-1} + (r-$

$1)u_N$, $n \in \mathbb{Z}, N \geq n$. Extending them by continuity to infinite sequences, we come exactly to the identifications described above. For more details see [Ver1], [Ver2] for the Markov case with $r = 1$ (more general cases are the same in techniques).

Let now ν denote the Haar measure for the group $\mathfrak{X}'_r$ (resp. $\mathfrak{Y}'_r$). Since ν is Borel and in view of the form of the equivalence relations $\mathfrak{R}_1$ and $\mathfrak{R}_2$ which do not touch any full cylindric set, it is well defined on all cylindric sets and thus, can be naturally extended to $\mathfrak{X}_r$ (resp. $\mathfrak{Y}_r$).

Below we will use the following claim.

Corollary 1.3. *Any arithmetic coding as a mapping from $\mathfrak{X}_r$ (resp. $\mathfrak{Y}_r$) onto $\mathbb{T}^2$ is always K-to-1 a.e. with respect to the measure ν for some natural K.*

2. The criterion for the existence of a bijective arithmetic coding, and the corresponding binary form

2.1. Case of the companion matrix. Let us begin with a crucial example.

Example. Consider the *Fibonacci* automorphism Φ given by the matrix $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. The corresponding Markov compactum is $\mathfrak{X}_\Phi = \{\varepsilon_n\} : \varepsilon_n \in \{0,1\},\ \varepsilon_n \varepsilon_{n+1} = 0,\ n \in \mathbb{Z}\}$. By Theorem 1.1, an arithmetic coding of Φ is a mapping φ_ξ from $\mathfrak{X}_\Phi$ onto $\mathbb{T}^2$ of the form

$$\varphi_\xi(\{\varepsilon_n\}) = \left(\sum_{n=-\infty}^{\infty} \xi \varepsilon_n \lambda^{-n} \right) \begin{pmatrix} 1 \\ \lambda^{-1} \end{pmatrix} \mod \mathbb{Z}^2.$$

Usually, the coefficients $\xi\varepsilon_n$ assume the values 0 and 1 (see, e.g., [Ber]). However this mapping from $\mathfrak{X}_\Phi$ onto the torus proves to be not bijective, but actually 5-to-1 a.e.! At the same time, as will be shown below, for the matrix $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ the bijective arithmetic coding does exist, and the proper choice of the coefficients is $\xi\varepsilon_n \in \left\{0, \frac{1}{\sqrt{5}}\right\}$. For a detailed study of the Fibonacci case see [SidVer].

We are going to show that for any automorphism T_r with the *companion* matrix $C_r := \begin{pmatrix} r & 1 \\ -\sigma & 0 \end{pmatrix}$ with $\sigma = \pm 1$ and $r \in \mathbb{N}$ for $\sigma = -1$ and $r \geq 3$ for $\sigma = +1$, a BAC always exists and that any such a coding is naturally parametrized by a unit of the field $\mathbb{Q}(\lambda)$. Note first that the vector $\begin{pmatrix} \lambda \\ -\sigma \end{pmatrix}$ is an eigenvector of the matrix C_r. Hence mapping (1) in this case is given as follows:

$$L_\xi(\{\varepsilon_n\}) = \left(\sum_{n=-\infty}^{\infty} \varepsilon_n \lambda^{-n} \right) \begin{pmatrix} \xi \\ -\sigma\xi\lambda^{-1} \end{pmatrix} \mod \mathbb{Z}^2. \tag{2}$$

To proceed, we need the precise description of possible values of ξ. Recall that by the above, $(\xi, -\sigma\lambda^{-1}\xi)$ is a homoclinic point for T_r.

Lemma 2.1. *Each homoclinic point for the automorphism T_r has the form* $(\frac{m+n\lambda}{\sqrt{D}} \mod 1,$ $-\sigma\lambda^{-1}\frac{m+n\lambda}{\sqrt{D}} \mod 1)$ *with* $(m,n) \in \mathbb{Z}^2$.

Remark. Thus, in formula (2),

$$\xi = \frac{m + n\lambda}{\sqrt{D}} \tag{3}$$

with $(m,n)\in\mathbb{Z}^2$.

Proof. It suffices to use the general fact concerning homoclinic points. Let T be a hyperbolic automorphism T of $\mathbb{T}^k$ (not necessarily two-dimensional). Consider the linear subspace of $\mathbb{R}^k$ containing the leaf of the unstable foliation going through $\mathbf{0}$. Then the projection of a point of the lattice $\mathbb{Z}^k$ to this subspace along the direction of the stable foliation taken modulo $\mathbb{Z}^k$ is always a homoclinic point for T, and and any homoclinic point can be obtained in such a way (see [Ver3]).

Applying this claim to our case, we immediately come to the claim of the lemma, because the orthogonal projection of a point (n,m) to the line $y=-\sigma\lambda^{-1}x$ is the point with the first coordinate $\frac{m+n\lambda}{\sqrt{D}}$. $\square$

Now our goal is to find among all ξ of the form (3) such that L_ξ is one-to-one a.e. We will see that actually these ξ have the minimal possible algebraic norm.

Theorem 2.2. *The automorphism T_r of the 2-torus given by the companion matrix $C_r := \begin{pmatrix} r & 1\\ -\sigma & 0\end{pmatrix}$, $\sigma=\pm1$, admits a bijective arithmetic coding. If $\sigma=-1$ or $\sigma=+1$, $r\ge 4$, its BAC is always of the form*

$$L_\xi(\{\varepsilon_n\})=\left(\sum_{n=-\infty}^{\infty}\varepsilon_n\lambda^{-n}\right)\begin{pmatrix}\xi\\ -\sigma\xi\lambda^{-1}\end{pmatrix}\mod \mathbb{Z}^2,$$

where $\xi=\frac{\pm\lambda^k}{\sqrt{D}}$, $k\in\mathbb{Z}$.

The case $M=\begin{pmatrix}3 & 1\\ -1 & 0\end{pmatrix}$ is specific. Here $\lambda=\frac{3+\sqrt5}{2}$, and any BAC is of the form

$$L_\xi(\{\varepsilon_n\})=\left(\sum_{n=-\infty}^{\infty}\varepsilon_n\lambda^{-n}\right)\begin{pmatrix}\xi\\ -\xi\lambda^{-1}\end{pmatrix}\mod \mathbb{Z}^2$$

with $\xi=\frac{\pm\theta^k}{\sqrt{D}}$, $k\in\mathbb{Z}$, $\theta=\sqrt{\lambda}=\frac{1+\sqrt5}{2}$.

Proof. Let first $\sigma=-1$. Consider an arbitrary sequence $\{\varepsilon_n\}_{-\infty}^{\infty}\in\mathfrak{X}$. We split it into two pieces $\{\varepsilon_n\}_{-\infty}^{0}$ and $\{\varepsilon_n\}_1^{\infty}$ and define $x_1(\{\varepsilon_n\}):=\sum_{k=1}^{\infty}\varepsilon_k\lambda^{-k}$, $x_2=\sum_{k=0}^{\infty}\varepsilon_{-k}(-\lambda)^{-k}$. It is a direct inspection that $x_1\in[0,1]$, $x_2\in[-1,\lambda]$. Using the relation $\{\lambda^n\}=\{(-1)^{n+1}\lambda^{-n}\}$, $n\ge0$, we make sure that $\sum_{-\infty}^{\infty}\varepsilon_n\xi\lambda^{-n}=\xi x_1-\overline{\xi}x_2 \mod 1$ and similarly, $\sum_{-\infty}^{\infty}\varepsilon_n\xi\lambda^{-n-1}=\xi\lambda^{-1}x_1+\overline{\xi}\lambda x_2$ mod 1, where, as above, $\overline{\xi}$ denotes the algebraic conjugate of a quadratic irrationality ξ.

Thus, we have the sequence of mappings

$$\mathfrak{X}_r\xrightarrow{F}\mathbb{R}^2\xrightarrow{b_\xi}\mathbb{R}^2,$$

where $F(\{\varepsilon_n\})=(x_1,x_2)$, and $b_\xi(x_1,x_2)=(\xi x_1-\overline{\xi}x_2,\ \xi\lambda^{-1}x_1+\overline{\xi}\lambda x_2)$. Now

$$L_\xi(\{\varepsilon_n\})=(b_\xi\varphi)(\{\varepsilon_n\})\mod\mathbb{Z}^2.$$

Note that since $(\varepsilon_0,\varepsilon_1)\ne(r,k)$ with $k\ne0$, $F(\mathfrak{X}_r)=\Pi=([0,1]\times[-1,\lambda])\setminus([\lambda^{-1},1]\times[\lambda^{-1},\lambda])$, i.e. the difference of rectangles. The square of Π is $(\lambda+1)\lambda^{-1}+(1-\lambda^{-1})\lambda=\sqrt{D}$, and the linear transformation $b_\xi=\begin{pmatrix}\xi & -\overline{\xi}\\ \lambda^{-1}\xi & \lambda\overline{\xi}\end{pmatrix}$ from $\mathbb{R}^2$ to $\mathbb{R}^2$ has determinant $\sqrt{D}N(\xi)$, where $N(\xi)=\xi\overline{\xi}$ is the algebraic norm of ξ. Thus, the image $(b_\xi F)(\mathfrak{X}_r)\equiv b_\xi(\Pi)$ on the plane has square $S=|DN(\xi)|$.

Recall that $\#L_\xi^{-1}(x)$ is the same for a.e. $x \in \mathbb{T}^2$, see Corollary 1.3. Thus, this capacity is necessarily equal to S, and L_ξ is a bijection a.e. if and only if

$$N(\xi) = \pm\frac{1}{D}. \tag{4}$$

By Lemma 2.1, $\xi = \frac{m+n\lambda}{\sqrt{D}}$, and the equation (4) is equivalent to the Diophantine equation

$$N(m+n\lambda) = \pm 1,$$

whence, as is well-known, $m+n\lambda$ is a unit of the ring $\mathbb{Z}[\lambda]$ and thus, $m+n\lambda = \pm\lambda^k$ for some $k \in \mathbb{Z}$ by virtue of the facts that $\mathbb{Z}[\lambda]$ is the maximal order of the field $\mathbb{Q}(\lambda)$ and that λ is its main unit (see, e.g., [BorSh]).

The case $\sigma = +1$ is studied in the same way. Since here $\overline{\lambda} = \lambda^{-1}$, we have $x_1 = \sum_1^\infty \varepsilon_k \lambda^{-k}$, $x_2 = \sum_0^\infty \varepsilon_{-k}\lambda^{-k}$. The set Π here is the difference of the parallelograms $([0,1) \times [0,\lambda)) \setminus ((1-\lambda^{-1},1) \times (\lambda-1,\lambda))$. The rest of the proof is the same, and we come to equation (4). Again, $m+n\lambda$ must be a unit of the ring $\mathbb{Z}[\lambda]$, whence $m+n\lambda = \pm\lambda^k$ if $r \geq 4$, and $m+n\lambda = \pm\theta^k$ for $r=3$ with θ equal to the golden ratio. □

Corollary 2.3. *Any mapping L_ξ from the symbolic compactum onto the torus of the form (2) is K-to-1 with $K = |DN(\xi)|$.*

Remark. Geometrically, each ξ which yields a bijective arithmetic coding in Theorem 2.2, can be represented by formula (3) by a point from the lattice $\mathbb{Z}^2$, form the orbit $\left\{\pm C_r^k \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ k \in \mathbb{Z}\right\}$, with the exception of the case $\sigma = +1, r = 3$, when these integral points form the orbit $\left\{\pm \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^k, k \in \mathbb{Z}\right\}$.

Let us describe the preimage of φ as a mapping acting from the symbolic compactum (without any glueings) onto the torus in detail.

Definition. A sequence in compactum $\mathfrak{X}_r$ is called *homoclinic* is its right and left tails are either of the form $(0,0,0,\dots)$ or of the form $(r,0,r,0,\dots)$. Similraly, a sequence in $\mathfrak{Y}_r$ is called *homoclinic* if its right and left tails are of the form $(0,0,0,\dots)$ or of the form $(r-2,r-2,r-2,\dots)$. We also call such tails *homoclinic*.

Proposition 2.4.

1. *For any bijective arithmetic coding given by formula (2), the inverse mapping φ^{-1} is well defined on the set of all points of the torus except the points of the leaves of the stable and unstable foliations for T_r going through $(0,0)$.*
2. *For the exclusive points mentioned in the previous item, the preimage consists of several sequences, each having at least one homoclinic tail.*
3. *The preimage of a homoclinic point consists of several homoclinic sequences. In particular, the preimage of $(0,0)$ is the set $\{(\dots 0,0,0,\dots),(\dots,r,0,r,\varepsilon_0 = 0,r,0,\dots),(\dots,0,r,0,\varepsilon_0 = r,0,r,\dots)\}$ for the Markov case and the set $\{(\dots,0,0,0,\dots),(\dots,r-2,r-2,r-2,\dots)\}$ for the sofic case.*

Proof. It suffices to use the bijectivity of the mapping φ acting from $\mathfrak{X}_r'$ (or $\mathfrak{Y}_r'$) onto $\mathbb{T}^2$ and apply Proposition 1.2 describing the nature of identifications that touch sequences with at least one homoclinic tail.

Remark 1. For the case in question it looks more natural to consider the following mapping which naturally generalizes the one sided β-expansions to the two-sided (= two-dimensional) case:

$$l(\{\varepsilon_n\}) = \left(\sum_{n=-\infty}^{\infty} \varepsilon_n \lambda^{-n}\right) \begin{pmatrix} -\sigma \\ \lambda^{-1} \end{pmatrix} \mod \mathbb{Z}^2.$$

It is a particular case of the more general mapping introduced and studied in [Ber]. Obviously, l semiconjugates the shift τ and the automorphism with the companion matrix C_r; however, from Corollary 2.3 it follows that the mapping l is only D-to-1 a.e.

Remark 2. For the case $\sigma = -1$ the mapping $L = L_{1/\sqrt{D}}$ in the form

$$L(\varepsilon) = \sum_{k=1}^{\infty} \varepsilon_k \lambda^{-k} \cdot \pi_u(u_0) - \sum_{k=0}^{\infty} \varepsilon_{-k}(-\lambda)^{-k} \cdot \pi_s(u_0) \mod \mathbb{Z}^2$$

with $u_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and π_u, π_s being the projections on the leaves of the unstable and stable foliations respectively, was under consideration in the recent Ph. D. dissertation [Leb]. In particular, the author proved its bijectivity a.e. but did not consider in detail its arithmetic properties.

Let us relate now our result to the classical algebraic theory of PV numbers.

Definition. Let θ be an algebraic integer > 1 such that all its Galois conjugates lie inside the unit disc on the complex plane. Then θ is called a Pisot-Vijayaraghavan (PV) number.

Thus, in our case λ is a quadratic PV unit. We define

$$\mathcal{G}_\lambda := \{\xi \in \mathbb{R} : \|\xi\lambda^n\| \to 0, \quad n \to \infty\}.$$

Obviously, $\mathcal{G}_\lambda$ is a group in addition, and from the definition of a PV number it is clear that $\mathbf{Z}[\lambda] \subset \mathcal{G}_\lambda$. We will call $\mathcal{G}_\lambda$ the *Pisot* group with a parameter λ. The implicit description of the Pisot group is given by the classical Pisot-Vijayaraghavan theorem claiming that ξ belongs to $\mathcal{G}_\lambda$ if and only if $\xi \in \mathbb{Q}(\lambda)$, and $\mathrm{Tr}(\xi) \in \mathbb{Z}$, $\mathrm{Tr}(\lambda\xi) \in \mathbb{Z}$, where Tr denote the algebraic trace of a number. It is not hard to obtain the precise description of the Pisot group from these conditions directly, however, our methods yield its structure almost immediately and relate this theory to the theory of hyperbolic systems.

Proposition 2.5. *Let, as above, $\lambda^2 = r\lambda \pm 1$ with $r \geq 1$ for $+1$ and $r \geq 3$ for -1, and let $D = r^2 \pm 4$. Then*

$$\mathcal{G}_\lambda = \frac{\mathbb{Z} + \lambda\mathbb{Z}}{\sqrt{D}}.$$

Proof. Let $\xi \in \mathcal{G}_\lambda$. Consider the point $x = (\{\xi\}, \{\lambda\xi\})$. Obviously, $T_r^n(x) = (\{\lambda^n\xi\}, \{\lambda^{n-1}\xi\})$, whence by definition of the Pisot group, $T_r^n(x) \to (0,0)$, $n \to \pm\infty$, and thus, x is a homoclinic point for the automorphism T_r, and by Lemma 2.1, its first coordinate has the form $\frac{m+n\lambda}{\sqrt{D}} \mod 1$ for some m, n integers.

Conversely, let $\xi = \frac{m+n\lambda}{\sqrt{D}}$. Then by Lemma 2.1, the point x is homoclinic for T_r, whence $\lambda^n\xi \to 0 \mod 1$, $n \to \infty$. $\square$

2.2. General case. Return now to the general case of ergodic automorphism T given by a matrix $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL(2, \mathbb{Z})$.

Definition. Two matrices M_1 and M_2 will be called *similar*, if there exists a matrix $B \in GL(2,\mathbb{Z})$ such that $BM_1B^{-1} = M_2$. We will write in this case $M_1 \sim M_2$.

Theorem 2.6.

I. *An ergodic automorphism T admits bijective arithmetic coding if and only its matrix M is similar to the companion matrix $C_r = \left(\begin{smallmatrix} r & 1 \\ -\sigma & 0 \end{smallmatrix}\right)$ with $r = a + d$ and $\sigma = \det M$.*

II. *A matrix $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is similar to the corresponding companion matrix if and only if the Diophantine equation*

$$bx^2 - (a-d)xy - cy^2 = \pm 1 \tag{5}$$

is solvable. Any matrix $B = \left(\begin{smallmatrix} x & y \\ z & t \end{smallmatrix}\right)$ from item I is naturally parametrized by a solution of equation (5), namely, (x, y) is always a solution of (5), and

$$\begin{pmatrix} z \\ t \end{pmatrix} = \begin{pmatrix} -d & c \\ b & -a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Proof. I. Suppose such a matrix B exists. Let L be a mapping from relation (2), say, for $\xi = 1$. Note that the compactum $\mathfrak{X}$ is determined only by the spectrum of the matrix specifying an automorphism, whence it is the same for T and the automorphism given by C_r. Now it is obvious that the mapping $\varphi := B^{-1}L : \mathfrak{X} \to \mathbb{T}^2$ is the desired BAC for T.

Conversely, let T admit BAC, and φ be the corresponding mapping from the symbolic compactum onto the torus. Consider $Q := \varphi L^{-1} : \mathbb{T}^2 \to \mathbb{T}^2$. It is well defined, because if two sequences $\varepsilon, \varepsilon'$ belong to $L^{-1}(x)$ for some $x \in \mathbb{T}^2$, by the above, $\varphi(\varepsilon) = \varphi(\varepsilon')$. Thus, we make sure that by definition of BAC, Q is a group automorphism of the 2-torus, hence, it is given by some matrix $B \in GL(2,\mathbb{Z})$, and obviously, $C_r = BMB^{-1}$.

II. By Theorem 2.4, it suffices to show that the solvability of the equation (5) is equivalent to the fact that $M \sim C_r$. Let $B = \left(\begin{smallmatrix} x & y \\ z & t \end{smallmatrix}\right) \in GL(2,\mathbb{Z})$ exist, and $BM = C_rB$. Suppose $\sigma = -1$ (another case is completely the same). We have thus the linear system

$$\begin{aligned} z &= cy - dx \\ t &= bx - ay \\ x &= az + ct \\ y &= bz + dt, \end{aligned} \tag{6}$$

the last two equations being a consequence of the first two ones. Hence this system together with the condition $\det B = \pm 1$ yields the desired condition.

Conversely, if the equation (5) is solvable, then we take some x, y being its solutions and construct the matrix B by the equations for z, t from formula (6). $\square$

Definition. The binary quadratic form $f_T(x,y) = bx^2 - (a-d)xy - cy^2$ will be called the *form associated with an automorphism T*.

Remark. Obviously, a binary integral quadratic form $f(x,y) = \alpha x^2 + \beta xy + \gamma y^2$ is the form associated with some automorphism if and only if $D(f) = r^2 \pm 4$ for some r, where $D(f) = \beta^2 - 4\alpha\gamma$ is the discriminant of the form f. Since $D(f_T) = D$, we are dealing in fact with all forms with the discriminant of the form $r^2 \pm 4$.

Recall that the Dirichlet theorem claims that for an automorphism T, the group $\mathcal{D}(T)$ defined as the set of all automorphisms of the torus which commute with T, has the form $\{\pm S^n,\ n \in \mathbb{Z}\}$ for some primitive automorphism S. The above corollary shows that in the Dirichlet group of T only four primitive elements $\pm S, \pm S^{-1}$ can admit BAC (with the exclusion of item (1), when they are eight).

Proposition 2.7. *Let equation (5) be solvable for a given automorphism T with the matrix $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We have the following alternative.*

I. *If M is primitive, then there exists a homoclinic point (ξ_0, η_0) such that any bijective arithmetic coding of T is of the form*

$$\varphi_k^{\pm}(\varepsilon) = \left(\sum_{n=-\infty}^{\infty} \varepsilon_n \lambda^{-n} \right) \begin{pmatrix} \pm\xi_0 \lambda^k \\ \pm\eta_0 \lambda^k \end{pmatrix} \mod \mathbb{Z}^2. \tag{7}$$

II. *If M is not primitive, we have $\lambda = \frac{3+\sqrt{5}}{2}$, and any BAC is of the form*

$$\varphi_k^{\pm}(\varepsilon) = \left(\sum_{n=-\infty}^{\infty} \varepsilon_n \lambda^{-n} \right) \begin{pmatrix} \pm\xi_0 \theta^k \\ \pm\eta_0 \theta^k \end{pmatrix} \mod \mathbb{Z}^2,$$

where $\theta = \sqrt{\lambda} = \frac{1+\sqrt{5}}{2}$.

Proof. I. Since equation (5) is solvable, there exist an infinite number of different BAC's for T. Fix φ_0 being one of them; let φ be be an arbitrary BAC for T. Consider the mapping $A := \varphi\varphi_0^{-1} : \mathbb{T}^2 \to \mathbb{T}^2$. It is well defined by the same arguments as in the proof of Theorem 2.6. Obviously, A is an automorphism of the 2-torus, and A commutes with T. By the Dirichlet theorem cited above and the primitivity of T, we have $\varphi = \pm T^k \varphi_0$ for some $k \in \mathbb{Z}$, whence if φ_0 in formula (1) is given by a homoclinic point (ξ_0, η_0), the mapping φ is given by $(\pm\lambda^k \xi_0, \pm\lambda^k \eta_0)$.

Conversely, if a bijection a.e. φ is given by formula (1) with some (ξ, η), the mapping φ' defined by the same formula with $(\pm\lambda^{\pm 1}\xi, \pm\lambda^{\pm 1}\eta)$ is also a bijection, as $\varphi(-\varepsilon) = -\varphi(\varepsilon)$, $\varphi(\tau^{\pm 1}\varepsilon) = T^{\pm 1}\varphi(\varepsilon)$.

II. The argument is the same with the exception that here $\sqrt{M}$ is also a matrix in $GL(2, \mathbb{Z})$ and also commutes with M.

Thus, the bijective airhtmetic codings in fact are naturally parametrized by the Dirichlet group of a given automorphism.

Let us give now some sufficient condition of the solvability of equation (5) in terms of its discriminant only.

Proposition 2.8. *For any automorphism T with $D < 20$ a BAC always exists.*

Proof. By [Cas, Ch. II, §4, Theorem VI], if for an indefinite binary quadratic form $f(x, y)$ with the discriminant D there are no integral x, y such that $-a < f(x, y) < b$, then $D \geq 4ab + \max(a^2, b^2)$. Inverting this assertion and taking $a = b = 2$, we come to the claim of the lemma, because the associated form $f_T(x, y) = bx^2 - (a-d)xy - cy^2$ is indefinite, and its discriminant is $(a-d)^2 + 4bc = (a+d)^2 + 4\sigma = D$.

Remarks. **1.** Practically this claim means that if $r = 1, 2, 3$ for $\sigma = -1$ and $r = 3, 4$ for $\sigma = +1$, then such an automorphism admits BAC (or, equivalently, M is similar to the corresponding companion matrix).

2. The condition $D < 20$ cannot be improved. Indeed, it suffices to consider the matrix $\begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}$. However, this matrix is not primitive, namely, the cube of $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$.

3. Furthermore, in Proposition 2.14 below we will show that if $D < 40$ for some primitive matrix, this matrix is necessarily similar to the companion matrix.

To describe explicitly all bijective arithmetic codings for a given automorphism, we first establish a relation of the admissibility of BAC with the primitivity or imprimitivity of a matrix M.

Proposition 2.9. *Let the matrix M of an automorphism T be not primitive, i.e. there exists a matrix $K \in GL(2,\mathbb{Z})$ such that $K^n = M$, $n \geq 2$. Then*

(1) *If $n = 2$, $\det K = -1$ and $\operatorname{Tr} K = 1$, then T admits BAC.*
(2) *Otherwise T does not admit BAC.*

Proof. (1) is a direct consequence of Proposition 2.8.
(2) It is easy to compute that $f_{T^2} = r f_T$, $f_{T^2} = (r^2+1)f_T$ and, more generally, $f_{T^n} = q_n(r) f_T$, where q_n is a polynomial of degree n with nonnegative coefficients, odd for n odd and even for n even.[4] Thus, ± 1 cannot be represented by such a form, if $n \geq 3$ or $r \geq 1$. $\square$

2.3. Some algebraic facts related to BAC. We are going to give some simple algebraic criteria for the existence of a bijective arithmetic coding of a given automorphism of the 2-torus and to relate the preceding theory to the theory of binary intergal quadratic forms.

Corollary 2.10. *If two ergodic automorphisms T and T' whose matrices have the same trace and discriminant, both admit bijective arithmetic coding, then their matrices are similar.*

Proof. It suffices to recall that both matrices should be similar to the corresponding companion matrix which is the same for both ones.

Corollary 2.11. *If $|b| = 1$ or $|c| = 1$, an automorphism T with the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ admits BAC.*

Proof. One of the equations (5) has the trivial solution $x = 1, y = 0$ or $x = 0, y = 1$ if $b = \pm 1$ or $c = \pm 1$ respectively.

Now we are going to prove an assertion which relates our theory to the theory of binary integral quadratic forms. Recall that two binary integral quadratic forms f and f' are called *equivalent* if $f'(x,y) = f(\alpha x + \beta y, \gamma x + \delta y)$ with $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in GL(2,\mathbb{Z})$. For indefinite quadratic forms the problem of equivalence is rather difficult (see, e.g., [Ven]); note only that for discriminants appearing in our kind of problems the number of equivalence classes is a priori large. Nevertheless, within the model of this section it is possible to claim something definite.

We first find out if the solvability of one of the equations (5) implies the solvability of another.

Lemma 2.12. *If $\det M = -1$, then the solvability of one of the equations (5) (say, with $+1$) implies the solvability of another. On the contrary, for $\det M = +1$ it is, generally speaking, wrong.*

Proof. Let first $\det M = -1$. Let, say, the Diophantine equation $bx^2 - (a-d)xy - cy^2 = +1$ be solvable. Then by the theorem from the theory of binary quadratic forms, the form $[b, d-a, -c]$ is equivalent to the reduced form $[1, \alpha, \beta]$ with α either 0 or 1 and $\beta = \frac{1}{4}(\alpha^2 - D)$ (see, e.g., [Wat, Ch. II, §3]). If r is odd, then this form is $[1, 1, (1-D)/4]$, otherwise, it is $[1, 0, -D/4]$. Now since the set of values assumed is the same for the whole class of equivalent forms, it suffices to deal with reduced forms.

Let r be even. Then the equation $x^2 - \frac{D}{4}y^2 = -1$ is really solvable, it suffices to take $x = r/2, y = 1$. If r is odd, the equation $x^2 + xy - \frac{D-1}{4}y^2 = -1$ has the solution $x = \frac{r-1}{2}, y = 1$.

As a couterexample for $\det M = 1$ we can consider the automorphism with the matrix $\begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}$. Here the associated equation $5x^2 - xy - y^2 = -1$ is solvable ($x = 0, y = 1$), while the equation $5x^2 - xy - y^2 = 1$ has no solutions, as this equation can be rewritten as $(10x - y)^2 = 2y^2 + 20$, whence ± 2 should be a quadratic residue modulo 10, what is wrong. $\square$

[4] More precisely, $q_n(r) = \frac{1}{\sqrt{D}}(\lambda^n - \overline{\lambda}^n)$.

At the end of the section we are going to prove a claim that reduces the problem of the similarity for automorphisms T and T' to the equivalence of the forms accosiated with them.

Proposition 2.13. *Let matrices M and M' belonging to $GL(2,\mathbb{Z})$ have the same trace and determinant. Then they are similar if and only if the forms associated with them are equivalent.*

Proof. Let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $M' = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$ with $a+d = a'+d'$, $ad-bc = a'd'-b'c'$. Then it is a direct inspection that the relation

$$B \begin{pmatrix} a & b \\ c & d \end{pmatrix} B^{-1} = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$$

with $B \in GL(2,\mathbb{Z})$ is equivalent to the relation

$$B \begin{pmatrix} b & \frac{1}{2}(d-a) \\ \frac{1}{2}(d-a) & -c \end{pmatrix} B^{T} = \begin{pmatrix} b' & \frac{1}{2}(d'-a') \\ \frac{1}{2}(d'-a') & -c' \end{pmatrix},$$

the latter being just the condition of the equivalence of the forms f_T and $f_{T'}$. $\square$

This claim will be used in the next section. Finally, we are going to describe all possibilities for $D<40$. Recall that if $D<20$, a matrix is always similar to the companion matrix with the same trace and determinant, whence the automorphism given by this matrix admits bijective arithmetic coding, see Proposition 2.8.

Proposition 2.14. *Let $20 \le D < 40$ for the matrix M of an automorphism T. Then either M is similar to the corresponding companion matrix C_r or M is not primitive. More precisely, there is the following alternative.*

(1) *If $D=21$ or $D=29$, then $M \sim C_r$.*
(2) *If $D=20$, then either M is primitive and $M \sim \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}$ or $M \sim \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^3$.*
(3) *If $D=32$, then either $M \sim C_r$ or $M \sim \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}^2$.*

Proof. The central point here is the following sharp estimate for the integral minimum of an indefinite binary quadratic form: for a form $f(x,y) = \alpha x^2 + \beta xy + \gamma y^2$ with the discriminant D,

$$\min \{|f(x,y)| : (x,y) \in \mathbb{Z}^2 \setminus (0,0)\} \le \sqrt{\frac{25D}{221}}$$

except the cases, when f is equivalent to one of the forms $l(x^2-xy-y^2)$ or $l(x^2-2y^2)$ with $l \ge 1$ (see [Cas, Chap. II, §4, Th. 6]). This estimate applied to f_T for $D \le 35$ (what actually means that $D<40$) yields the minimum $=1$, what is equivalent to the solvability of equation (5). Considering the possibilities for the exclusions in the cited claim, we make sure that they could appear only for $D=20$ or $D=32$. In the first case this leads to the equivalence of the forms $f_T(x,y)$ and $2(x^2-xy-y^2)$, whence by the previous proposition, $M \sim \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}$. The case $D=32$ is studied in the same way. $\square$

Remark 1. The value $D=40$ is the least, for which there exists a primitive matrix with this discriminant not similar to the companion matrix, namely, $M = \begin{pmatrix} 5 & 3 \\ 2 & 1 \end{pmatrix}$. Here equation (5) is $3x^2-4xy-2y^2 = \pm 1$ and has no integral solutions, as it can be rewritten as $(3x-2y)^2-10y^2 = \pm 3$, whence it would follow that ± 3 is a quadratic residue modulo 10.

Remark 2. Although each matrix with $D = 5$ is similar by the above to the companion matrix, this is, generally speaking, wrong for an arbitrary matrix whose spectrum is in the ring $\mathbb{Z}[\lambda]$ with $\lambda = \frac{1+\sqrt{5}}{2}$. Here is an example. Consider $M = \begin{pmatrix} 27 & 11 \\ 5 & 2 \end{pmatrix}$ whose spectrum is $\{\lambda^7, -\lambda^{-7}\}$. A detailed analysis shows that the integral minimum of the absolute value of the associated form $f_T = 11x^2 - 25xy - 5y^2$ equals 5, whence, equation (5) for this case has no integral solutions, though M is primitive. Thus, it is impossible to reformulate Proposition 2.14 in purely "ring" terms.

However, there exist values $D > 40$ such that any matrix in $GL(2, \mathbb{Z})$ having the discriminant D is similar to the corresponding companion matrix. For instance, as such a value we can consider $D = 53$. A more detalied study could be carried out with the help of the theory of Diophantine quadratic equations (see, e.g., [Ven]). Probably, there exist an infinite number of such discriminants.

3. General case: minimal arithmetic coding

We have already seen that sometimes an ergodic automorphism of the 2-torus does not admit BAC, so, it is meaningful to deal with the notion of minimal aithmetic coding (MAC) introduced in Section 1. Recall that a minimal arithmetic coding of an automorphism is, by definition, an arithmetic coding with the minimal possible number of preimages. Let us first formulate a claim on arithmetic codings generalizing previous considerations.

By the above, the convergence of the series in formula (1) implies $\xi \in \mathcal{G}_\lambda,\ \eta \in \mathcal{G}_\lambda$, whence from Proposition 2.5,

$$\xi = \frac{v + n\lambda}{\sqrt{D}},\ \eta = \frac{u + k\lambda}{\sqrt{D}}. \tag{8}$$

Furthermore, by the fact that $\begin{pmatrix} \xi \\ \eta \end{pmatrix}$ is an eigenvector for M, we have

$$\begin{pmatrix} n \\ k \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} v \\ u \end{pmatrix}. \tag{9}$$

Using the same arguments as in Theorem 2.2, we make sure that the square of the fundamental domain is

$$S = \sqrt{D} \begin{vmatrix} \xi & -\overline{\xi} \\ \eta & -\overline{\eta} \end{vmatrix}. \tag{10}$$

Furthermore, from direct computations in formula (10) which we omit, it follows that

$$S = |bu^2 - (a-d)uv - cv^2|. \tag{11}$$

Thus, we proved the following theorem which describes explicitly, in what way an arbitrary arithmetic coding is parametrized by a homoclinic point.

Theorem 3.1. *Let T be a hyperbolic automorphism of the 2-torus with the matrix $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then any arithmetic coding of T is a mapping from the symbolic compactum onto the torus having the form (1) with (ξ, η) being an arbitrary homoclinic point. This point is given by formulas (8) and (9) with arbitrary $(u, v) \in \mathbb{Z}^2$. A mapping φ is S-to-1 a.e. with S given by formula (11).*

Let $m(T)$ denote the minimal possible number of preimages for an arithmetic coding of T.

Corollary 3.2. *The quantity $m(T)$ equals the integral minimum of the associated form f_T. Any minimal arithmetic coding of a given automorphism T is naturally parametrized by a solution of the equation*

$$bu^2 - (a-d)uv - cv^2 = \pm m, \tag{12}$$

where $m = m(T)$.

Corollary 3.3. *We have the following estimate for a primitive T:*

$$m(T) \leq \sqrt{\frac{25D}{221}} \tag{13}$$

except the cases $D = 5$ or $D = 8$, when $m(T) = 1$. This estimate cannot be improved.

Proof. It suffices to apply the result from [Cas] concerning the binary quadratic forms cited in the proof of Proposition 2.14. To prove the attainability of the estimate (13), one may consider the matrix $\left(\begin{smallmatrix} 2 & 5 \\ 5 & 13 \end{smallmatrix}\right)$ whose associated form $f_T(x,y) = 5x^2 + 11xy - 5y^2$ is considered in [Cas] and is known to have the minimum 5. □

We are going to describe all possible minimal arithmetic codings for a given automorphism more explicitly in the manner of Proposition 2.9. Let us first prove a technical lemma.

Lemma 3.4. *Let a mapping φ be a minimal arithmetic coding for some T. Then the kernel of φ as a homomorphism of the groups $\mathfrak{X}'_r$ (or $\mathfrak{Y}'_r$) and $\mathbb{T}^2$ consists of purely periodic sequences. The number of possible kernels for a fixed T is finite.*

Proof. Since φ is bounded-to-one and $\operatorname{Ker}\varphi$ is a set invariant under the shift τ, it can contain purely periodic sequences only. The second claim follows from the fact that the symbolic compactum is isomorphic (as a group) to $\mathbb{T}^2$ and that for a fixed m there exist only a finite number of commutative subgroups of $\mathbb{T}^2$ of order m. □

Proposition 3.5. *For a given automorphism T with the primitive matrix there exists a finite collection of homoclinic points $(\xi_0, \eta_0), (\xi_1, \eta_1), \ldots, (\xi_s, \eta_s)$ such that each minimal arithmetic coding of T is of the form*

$$\varphi^{\pm}_{j,k}(\varepsilon) = \left(\sum_{n=-\infty}^{\infty} \varepsilon_n \lambda^{-n} \right) \begin{pmatrix} \pm\xi_j \lambda^k \\ \pm\eta_j \lambda^k \end{pmatrix} \mod \mathbb{Z}^2$$

for some $k \in \mathbb{Z}$, $j \in \{0, 1, \ldots, s\}$.

Proof. Let φ and φ' be two minimal arithmetic codings for T. Recall that they are group homomorphisms of the groups $\mathfrak{X}'$ (or $\mathfrak{Y}'$) and $\mathbb{T}^2$. Suppose first $\operatorname{Ker}\varphi = \operatorname{Ker}\varphi'$. Then $A := \varphi'\varphi^{-1} : \mathbb{T}^2 \to \mathbb{T}^2$ is well defined, and by definition, A is an automorphism of $\mathbb{T}^2$ commuting with T. Again, by the Dirichlet theorem and the primitivity of T, we have $A = \pm T^k$, whence $\varphi' = \pm T^k \varphi$, $k \in \mathbb{Z}$. Thus, for two MAC's with the same kernel, the claim is proved. Applying now the previous lemma, we are done.

If T is not primitive, we have the following natural analog.

Proposition 3.6. *Let T be an automorphism whose matrix M is not primitive. Let $M = K^n$ with $n \geq 2$ and a primitive $K \in GL(2, \mathbb{Z})$. Then there exists a finite collection of homoclinic points $(\xi_0, \eta_0), (\xi_1, \eta_1), \dots, (\xi_s, \eta_s)$ such that each minimal arithmetic coding of T is of the form*

$$\varphi^{\pm}_{j,k}(\varepsilon) = \left(\sum_{n=-\infty}^{\infty} \varepsilon_n \lambda^{-n} \right) \begin{pmatrix} \pm\xi_j \theta^k \\ \pm\eta_j \theta^k \end{pmatrix} \mod \mathbb{Z}^2$$

for some $k \in \mathbb{Z}$, $j \in \{0, 1, \dots, s\}$ and $\theta = \sqrt[n]{\lambda}$.

Following the framework of the previous section, we are going to relate minimal arithmetic codings to semiconjungating of matrices.

Proposition 3.7. *Let $M \in GL(2, \mathbb{Z})$ and C_r be the corresponding companion matrix. Then there exists a natural one-to-one correspondence between the set of matrices $B \in GL(2, \mathbb{Q}) \cap M_2\mathbb{Z}$ such that $MB = BC_r$ and arithmetic codings for T given by the matrix M. In particular, the matrices B with the minimal possible determinant in modulus correspond to the set of MAC's, and this determinant equals $m(T)$.*

Proof. Let L be as in formula (2) with, say, $\xi = 1/\sqrt{D}$. Suppose B be a matrix semiconjugating M and C_r. Let the corresponding mapping be denoted by the same letter. Setting $\varphi = BL$, we make sure that φ is an arithmetic coding. Conversely, given such a φ, we set $B = \varphi L^{-1}$. □

Remark. In dynamical terms, for a given automorphism sometimes there are no bijective mappings from the symbolic compactum onto the torus with the necessary (arithmetic) properties but always exists a mapping φ having a finite-sheeted covering L being a bijection. Namely, in our notation the resulting mapping is the group endomorphism of the torus defined as $B = \varphi L^{-1}$.

Thus, we reduced the problem of description of the kernels of MAC's for T to the purely algebraic problem of describing the kernels of the endomorphisms of $\mathbb{T}^2$ given by the matrices semiconjugating M and C_r. From Theorem 3.1 one can easily see, how to construct any B semiconjugating M and C_r from the solutions of the Diophantine equation (12).

Example. Let $M = \begin{pmatrix} 80 & 9 \\ 9 & 1 \end{pmatrix}$. Then $f_T(x, y) = 9x^2 - 79xy - 9y^2$, and it is a direct inspection that the equations $f_T(x, y) = \pm k$ have no solutions for $1 \leq k \leq 8$. Thus, the integral minimum of $|f_T|$ equals 9. We consider the equation $f_T(x, y) = \pm 9$ and choose the pairs of solutions: $(x_1 = 1, y_1 = -9)$ and $(x_2 = 9, y_2 = 1)$. Constructing now the matrices $B_1 = \begin{pmatrix} x_1 & y_1 \\ z_1 & t_1 \end{pmatrix}$ and $B_2 = \begin{pmatrix} x_2 & y_2 \\ z_2 & t_2 \end{pmatrix}$ by formula (6), we obtain thus two matrices from $GL(2, \mathbb{Q}) \cap M_2\mathbb{Z}$ semiconjugating M and the companion matrix $\begin{pmatrix} 81 & 1 \\ 1 & 0 \end{pmatrix}$. However, the matrix $B_1 B_2^{-1}$ is not integral, whence the endomorphisms given by the matrices B_1 and B_2 have different kernels, so do the corresponding mappings φ_1 and φ_2. Note also that the groups $\operatorname{Ker} B_1$ and $\operatorname{Ker} B_2$ being isomorphic as abstract groups, are not isomorphic with respect to T in the sense that there is no automorphism commuting with T and turning $\operatorname{Ker} B_1$ into $\operatorname{Ker} B_2$.

Thus, the kernel of the minimal arithmetic coding is not an invariant for the integral conjugacy in $GL(2, \mathbb{Z})$, as it does not apply even for a single matrix.

The idea of this example is based on the fact that $m(T)$ is not a prime. Let us now establish a relation between the quantities $m(T)$ and the similarity of the matrices in $GL(2, \mathbb{Z})$.

Proposition 3.8. *If automorphisms T and T' given by matrices M and M' are similar, then $m(T) = m(T')$. The inverse is, generally speaking, not true. Nevertheless, if $m(T) = m(T') = p$ prime, then $M \sim M'$.*

Proof. The first claim follows from Proposition 2.13, because the mimima of two equivalent forms coincide. It is well-known that the minimum of the form of a given discriminant in general is not its complete invariant. However, if this minimum is a prime number, it is, i.e. all binary quadratic forms of a given discriminant and the same prime minimum, are equivalent (see, e.g., [Wat]). Now our claim follows from Proposition 2.13. □

Finally, we keep our promise and show how to reduce the case $r < 0$ to $r > 0$. Briefly, given an automorphism T whose matrix M has the negative trace, we consider the automorphism with the matrix $-M$, and make sure that it has the same eigenvector and the same series with the terms $\varepsilon_n \lambda^n$ instead of $\varepsilon_n \lambda^{-n}$ and inverted (in the Markov case) restrictions on the digits.

More precisely, let $\mathfrak{X}_r^-$ be the stationary Markov compactum $\{\{\varepsilon_n\}_{-\infty}^{\infty} : 0 \le \varepsilon_n \le r,\ \varepsilon_n = r \Rightarrow \varepsilon_{n-1} = 0,\ n \in \mathbb{Z}\}$. Then any arithmetic coding of T is given by the mapping

$$\psi(\varepsilon) = \left(\sum_{n=-\infty}^{\infty} \varepsilon_n \lambda^n \right) \begin{pmatrix} \xi \\ \eta \end{pmatrix} \mod \mathbb{Z}^2,$$

acting from $\mathfrak{X}_r^-$ if $\sigma = -1$ and $\mathfrak{Y}_r$ otherwise with $\lambda = \frac{r+\sqrt{D}}{2}$ and (ξ, η) being a homoclinic point for T.

Besides, $M \sim M'$ iff $-M \sim -M'$, the same is true for the equivalence of the quadratic forms. Thus, all the questions of the paper for the case $r < 0$ are naturally reduced to the ones for $r > 0$.

References

[AdWe] R. L. Adler and B. Weiss, *Entropy, a complete metric invariant for automorphisms of the torus*, Proc. Nat. Acad. Sci. USA **57** (67), 1573–1576.

[Ber] A. Bertrand-Mathis, *Developpement en base θ, répartition modulo un de la suite $(x\theta^n)_{n\ge 0}$; langages codés et θ-shift*, Bull. Math. Soc. Fr. **114** (1986), 271–323.

[BorSh] Z. I. Borevich and I. R. Shafarevich, *Number Theory*, New York, Academic Press, 1986.

[Bow] R. Bowen, *Markov partitions and measures for axiom A diffeomorphisms*, Trans. Amer. Math. Soc. **154** (1971), 377–397.

[Cas] J. Cassels, *An introduction to the Geomtery of Numbers*, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1959.

[FrSa] Ch. Frougny and J. Sakarovitch, *Automatic conversion from Fibonacci to golden mean, and generalization*, to appear in Int. J. of Alg. and Comput.

[FrSo] Ch. Frougny and B. Solomyak, *Finite beta-expansions*, Ergod. Th. Dynam. Sys. **12** (1992), 713–723.

[GuSi] B. M. Gurevich and Ya. G. Sinai, *Algebraic automorphisms of the torus and Markov chains*, in book "P. Billingsley, Ergodic theory and information (Russian translation)", Moscow, Mir, 1969, 205–233.

[KenVer] R. Kenyon and A. Vershik, *Arithmetic construction of sofic partitions of hyperbolic toral automorphisms*, Preprint n°178 of École Normale Supérieure de Lyon (1995), to appear in Erg. Theory Dyn. Sys.

[Leb] S. Le Borgne, *Dynamique symbolique et propriétés stochastiques des automorphisms du tore : cas hyperbolique et quasi-hyperbolique*, Thèse doctorale (1997).

[SidVer] N. A. Sidorov and A. M. Vershik, *Ergodic properties of Erdös measure, the entropy of the goldenshift, and related problems*, to appear in Monatsh. Math.

[Sin] Ya. Sinai, *Markov partitons and A-diffeomorphisms*, Funct. Anal. Appl. **2** (1968), 64–89.

[Ven] B. A. Venkov, *Elementary Number Theory*, Wolters-Noordhoff.

[Ver1] A. M. Vershik, *Locally transversal symbolic dynamics*, Algebra and Analysis **6** (1994), no. 3 (in Russian); English transl. St. Petersburg Math. J. **6** (1995), 529–540.

[Ver2] A. M. Vershik, *The fibadic expansions of real numbers and adic transformation*, Prep. Report Inst. Mittag–Leffler, 1991/1992, pp. 1–9.

[Ver3] A. M. Vershik, *Arithmetic isomorphism of the toral hyperbolic automorphisms and sofic systems*, Funct. Anal. Appl. **26** (1992), 22–24.

[Wat] G. L. Watson, *Integral quadratic forms*, Cambridge Univ. Press, 1960.