Ergod. Th. & Dynam. Sys. (First published online 2017), page 1 of 23*
doi:10.1017/etds.2017.60
*Provisional—final page numbers to be inserted when paper edition is published

An almost mixing of all orders property of algebraic dynamical systems

L. ARENAS-CARMONA†, D. BEREND‡ and V. BERGELSON§

† *Department of Mathematics, University of Chile, Casilla 653, Santiago, Chile*
(e-mail: learenass@yahoo.com)

‡ *Departments of Mathematics and Computer Science, Ben-Gurion University, Beer Sheva 84105, Israel*
(e-mail: berend@math.bgu.ac.il)

§ *Department of Mathematics, Ohio State University, Columbus, OH 43210, USA*
(e-mail: vitaly@math.ohio-state.edu)

(*Received* 13 *September* 2016 *and accepted in revised form* 30 *May* 2017)

Abstract. We consider dynamical systems, consisting of $\mathbb{Z}^2$-actions by continuous automorphisms on shift-invariant subgroups of $\mathbb{F}_p^{\mathbb{Z}^2}$, where $\mathbb{F}_p$ is the field of order p. These systems provide natural generalizations of Ledrappier's system, which was the first example of a 2-mixing $\mathbb{Z}^2$-action that is not 3-mixing. Extending the results from our previous work on Ledrappier's example, we show that, under quite mild conditions (namely, 2-mixing and that the subgroup defining the system is a principal Markov subgroup), these systems are almost strongly mixing of every order in the following sense: for each order, one just needs to avoid certain effectively computable logarithmically small sets of times at which there is a substantial deviation from mixing of this order.

1. *Introduction*

Let $\mathbb{F}_p = \mathrm{GF}(p)$ be the Galois field of order p, where p is a prime, and let $\Omega_p = \mathbb{F}_p^{\mathbb{Z}^2}$. Consider the shift $\mathbb{Z}^2$-action on the compact abelian group Ω_p, defined by $(S^{\mathbf{w}}x)_{\mathbf{w}'} = x_{\mathbf{w}'-\mathbf{w}}$ for $\mathbf{w}, \mathbf{w}' \in \mathbb{Z}^2$. A *Markov subgroup* is a closed shift-invariant subgroup Θ of Ω_q (see [**5**, **7**]). For any Markov subgroup Θ, the restriction of the action $S^{\mathbf{w}}$ to Θ defines a measure-preserving system $\tilde{\Theta} = (\Theta, \mathcal{B}_\Theta, \mu_\Theta, (S^{\mathbf{w}})_{\mathbf{w}\in\mathbb{Z}^2})$, where $\mathcal{B}_\Theta$ is the σ-algebra generated by the open sets and μ_Θ is the normalized Haar measure. Markov subgroups of Ω_p are of interest since, for the study of many dynamical properties of $\mathbb{Z}^2$-actions by automorphisms on totally disconnected compact groups, it suffices to deal with that of

Markov subgroups. A Markov subgroup $\Psi \subseteq \Omega_p$ is *principal* if

$$\Psi = \left\{ \nu \in \Omega_p \;\middle|\; \sum_{i=1}^{k} \alpha_i \nu_{\mathbf{w}_i + \mathbf{w}} = 0,\ \mathbf{w} \in \mathbb{Z}^2 \right\}, \tag{1}$$

for some distinct $\mathbf{w}_1, \ldots, \mathbf{w}_k \in \mathbb{Z}^2$ and non-zero $\alpha_1, \ldots, \alpha_k \in \mathbb{F}_p$. In fact, principal Markov subgroups Ψ are the only Markov subgroups of Ω_p for which $\tilde{\Psi}$ is ergodic (see Proposition 3.13 below). A principal Markov subgroup is *minimal* if it does not properly contain any principal Markov subgroup. Every principal Markov subgroup contains a minimal principal Markov subgroup (see Proposition 3.2).

Perhaps the simplest non-trivial example of a Markov subgroup is Ledrappier's subgroup $\mathbb{L} \subseteq \Omega_2$ (see [**6**]). It is defined as the set of elements $\nu = (\nu_w)_{w \in \mathbb{Z}^2} \in \Omega_2$ satisfying the relation $\nu_{\mathbf{w}} + \nu_{\mathbf{w}+(1,0)} + \nu_{\mathbf{w}+(0,1)} = 0$ for every $w \in \mathbb{Z}^2$. Ledrappier's subgroup is minimal. It is properly contained in the subgroup

$$M = \{\nu \in \Omega_2 \mid \nu_{\mathbf{w}} + \nu_{\mathbf{w}+(2,0)} + \nu_{\mathbf{w}+(0,1)} + \nu_{\mathbf{w}+(1,1)} = 0,\ \mathbf{w} \in \mathbb{Z}^2\},$$

since

$$\begin{aligned} \nu_{\mathbf{w}} + \nu_{\mathbf{w}+(2,0)} + \nu_{\mathbf{w}+(0,1)} + \nu_{\mathbf{w}+(1,1)} &= (\nu_{\mathbf{w}} + \nu_{\mathbf{w}+(1,0)} + \nu_{\mathbf{w}+(0,1)}) \\ &\quad + (\nu_{\mathbf{w}'} + \nu_{\mathbf{w}'+(1,0)} + \nu_{\mathbf{w}'+(0,1)}), \end{aligned}$$

where $\mathbf{w}' = \mathbf{w} + (1, 0)$. In fact, $M/\mathbb{L}$ is a cyclic group of order two. The non-trivial class is the set of elements ν satisfying $\nu_{\mathbf{w}} + \nu_{\mathbf{w}+(1,0)} + \nu_{\mathbf{w}+(0,1)} = 1$ for all $\mathbf{w} \in \mathbb{Z}^2$. A possible representative of this class is given by the point ν with $\nu_{\mathbf{w}} = 1$ for all $\mathbf{w}$. Similarly, the group

$$\Phi = \{\nu \in \Omega_2 \mid \nu_{\mathbf{w}} + \nu_{\mathbf{w}+(2,0)} + \nu_{\mathbf{w}+(0,2)} = 0\},$$

contains Ledrappier's subgroup since

$$\nu_{\mathbf{w}} + \nu_{\mathbf{w}+(2,0)} + \nu_{\mathbf{w}+(0,2)} = \sum_{i=1}^{3} (\nu_{\mathbf{w}_i} + \nu_{\mathbf{w}_i+(1,0)} + \nu_{\mathbf{w}_i+(0,1)}),$$

where $\mathbf{w}_1 = \mathbf{w}$, $\mathbf{w}_2 = \mathbf{w} + (1, 0)$, and $\mathbf{w}_3 = \mathbf{w} + (0, 1)$.

Ledrappier's system was the first example of a $\mathbb{Z}^2$-action which is 2-mixing but not 3-mixing. In [**1**], we have studied this system thoroughly and showed that it is 'almost' mixing of every order. Namely, mixing of a specific order means that intersections of sets to which we apply transformations from the given semigroup tend to be almost independent when the transformations are 'far' from each other.

We have shown that, while Ledrappier's system is not 3-mixing, the exceptions to 3- and higher-order mixing are contained in a 'small' set of transformations comprising the $\mathbb{Z}^2$-action in Ledrappier's example. Let us call a set $E \subseteq \mathbb{Z}^2$ logish if the number of elements in E of 'size' n or less is bounded by some power of $\log n$ as n increases. It was proved in [**1**] that, up to some technicalities arising from the possibility of taking shifts and disjoint unions, the set of exponents that needed to be avoided is logish. (See §2 for more detailed definitions and discussion.)

In this paper, we attempt to obtain an analogue for general Markov subgroups. Of course, one cannot expect such a general result to be as detailed as in the case of the explicit

example of Ledrappier, but the general spirit of our main results is the same. Namely, along a generic curve, unless one samples at very specific logish times, which correspond to obvious obstructions to mixing, one essentially gets mixing of all orders. Also, due to the generality of the results, the proofs this time require heavier use of algebraic methods, as the elementary calculations of [**1**] cannot be applied to the general case.

In §2, we present the main results of the paper—Theorems 2.4 and 2.7 below. The following several sections are devoted to developing various tools that will serve us in the proofs. Section 3 discusses the dual action and contains several relevant results. In §4, we view the main results from an algebraic viewpoint. Section 5 recalls the basics of valuations on some global and local fields, and explains their relevance. In §6, we prove that various power series over finite fields are algebraically independent. In §7, we utilize the general results from the previous sections to obtain results on mixing of all orders along polynomial sequences. Sections 8 and 9 are devoted to several more auxiliary results. Finally, in §10, we conclude the proofs of the main theorems.

2. *The main results*

Two sequences $(\mathbf{w}(t))_{t\in\mathbb{N}}$ and $(\mathbf{w}'(t))_{t\in\mathbb{N}}$ in $\mathbb{Z}^2$ *grow apart* if $\|\mathbf{w}(t)-\mathbf{w}'(t)\|\longrightarrow_{t\to\infty}\infty$ (where $\|\cdot\|$ denotes any norm on $\mathbb{R}^2$). An *r-sequence* is a sequence in $(\mathbb{Z}^2)^r$. An r-sequence $(B(t))_{t\in\mathbb{N}}$, where $B(t)=(\mathbf{w}_1(t),\ldots,\mathbf{w}_r(t))$, is *spreading* if the sequences $\mathbf{w}_i(t)$ and $\mathbf{w}_j(t)$ grow apart for every $1\le i<j\le r$. A spreading r-sequence $(B(t))_{t\in\mathbb{N}}$, where $B(t)=(\mathbf{w}_1(t),\ldots,\mathbf{w}_r(t))$, is *mixing* for $\tilde{\Psi}$ if

$$\int_\Psi \prod_{i=1}^r f_i(S^{\mathbf{w}_i(t)}g)\,d\mu_\Psi(g)\underset{t\to\infty}{\longrightarrow}\prod_{i=1}^r\int_\Psi f_i,\qquad f_1,\ldots,f_r\in L^\infty(\Psi,\mathcal{B}_\Psi,\mu_\Psi).\tag{2}$$

The system $\tilde{\Psi}$ is *r-mixing* if every spreading r-sequence $(B(t))_{t\in\mathbb{N}}$ is mixing for $\tilde{\Psi}$. When $\tilde{\Psi}$ is not r-mixing, we would like to know how far it is from being such. A set $A\subseteq\mathbb{Z}^2$ is an *r-trap* for $\tilde{\Psi}$ if every spreading r-sequence $(B(t))_{t\in\mathbb{N}}$, where $B(t)=(\mathbf{w}_1(t),\ldots,\mathbf{w}_r(t))$, satisfying $\mathbf{w}_i(t)-\mathbf{w}_j(t)\notin A$ for all t, is mixing for $\tilde{\Psi}$. One way of measuring how far $\tilde{\Psi}$ is from being r-mixing is by providing a trap for $\tilde{\Psi}$. The smaller the trap A is, the closer $\tilde{\Psi}$ is to being mixing. To fix ideas, we introduce a few notions of smallness. For a fixed k, denote $C_N=[-N,N]^k$.

Definition 2.1. A subset A of $\mathbb{Z}^k$ is *logish of order M*, or *M-logish*, if $|A\cap C_N|=O(\ln^M N)$ as $N\to\infty$. The set is *logish* if it is M-logish for some M.

For example, the set of powers of two is 1-logish, and the sets $\{2^m+3^n : m,n\in\mathbb{N}\}$ and $\{2^m3^n : m,n\in\mathbb{N}\}$ are 2-logish. Clearly, if $A_1\subseteq\mathbb{Z}^{k_1}$ is M_1-logish and $A_2\subseteq\mathbb{Z}^{k_2}$ is M_2-logish, then $A_1\times A_2\subseteq\mathbb{Z}^{k_1+k_2}$ is (M_1+M_2)-logish.

Definition 2.2. Let $V:\mathbb{Z}^k\to\mathbb{Z}$ be a non-zero linear map. The set $B\subseteq\mathbb{Z}^k$ is *M-logish with respect to V* if there exists an M-logish subset A of $\mathbb{Z}=\mathbb{Z}^1$ such that $B\subseteq V^{-1}(A)$.

For example, the set $\{(m,2^n) : m,n\in\mathbb{N}\}$ is logish with respect to π_2, where $\pi_2(a,b)=b$, but it is not logish. Note that, if B is M-logish with respect to V for some non-zero linear map V, then $|B\cap C_N|=O[N^{k-1}(\log N)^M]$. This bound may seem to indicate that logishness with respect to a linear map is much coarser than logishness. However, we have the following property.

PROPOSITION 2.3. *Let $V_1, \ldots, V_r : \mathbb{Z}^r \to \mathbb{Z}$ be linear maps whose extensions to $\mathbb{R}^r$ form a basis of the dual space $(\mathbb{R}^r)^*$. Let $B \subseteq \mathbb{Z}^r$ be M-logish with respect to each V_i. Then B is M^r-logish.*

Proof. By abuse of notation, identify each map with its extension to $(\mathbb{R}^r)^*$. Let $T : \mathbb{R}^r \to \mathbb{R}^r$ be the map defined by $T(\mathbf{w}) = (V_1(\mathbf{w}), \ldots, V_r(\mathbf{w}))$. Then T is an invertible linear map, since $\{V_1, \ldots, V_r\}$ is a basis of $(\mathbb{R}^r)^*$. Now $T(B) \subseteq \prod_{i=1}^r V_i(B)$, and hence $T(B)$ is M^r-logish. Recall that there exists a positive constant D such that $D\|T(\mathbf{w})\| \leq \|\mathbf{w}\|$. It follows that $|B \cap C_N| \leq |T(B) \cap C_{D^{-1}N}|$, which proves the lemma. □

The following two theorems are the main results of the paper. It will be convenient to denote, for any positive integer r,

$$h = h_p(r) = \begin{cases} r - 1 & \text{if } p = 2, \\ (p-1)p^{r+1} + r & \text{if } p > 2. \end{cases} \tag{3}$$

THEOREM 2.4. *Let Ψ be a minimal principal Markov subgroup of Ω_p. Assume that $\tilde{\Psi}$ is 2-mixing. Then there is an effective linear function $V : \mathbb{Z}^2 \to \mathbb{Z}$ such that, for every integer $r \geq 3$ and every $\varepsilon > 0$, there is an r-trap A for $\tilde{\Psi}$ that is $(2h + \varepsilon)$-logish with respect to V.*

Remark 2.5. In the preceding theorem, if $\Psi = \{\nu \in \Omega_p \mid \sum_{\mathbf{w}' \in S} a_{\mathbf{w}'} \nu_{\mathbf{w}+\mathbf{w}'} = 0, \mathbf{w} \in \mathbb{Z}^2\}$ for some finite S, then V can be chosen as any linear map whose kernel is parallel to a side of the convex hull of S (see the discussion following Theorem 4.2). The condition on 2-mixing implies that S is not contained in a line (see Example 3.10), so there are several choices for the map V that are not multiples of each other. The trap A is a set of the form $KV^{-1}(C)$ for an arbitrary shell C (as defined below) of the set D' of distances between elements in the set $D_p^{(h)}$, defined in §4. The constant K is explicit.

Example 2.6. Let $\Psi_0 = \{\nu \in \Omega_2 \mid \nu_{\mathbf{w}} + \nu_{\mathbf{w}+(0,1)} = 0, \mathbf{w} \in \mathbb{Z}^2\}$. The elements of Ψ_0 have constant coordinates on every vertical line. We conclude that no sequence of the form $((0, 0), (c, \gamma(t)))$ is mixing for this system. Any trap must contain all but a finite number of points on each vertical line. This system cannot, therefore, have a trap that is logish with respect to any V.

THEOREM 2.7. *Let Ψ be a principal Markov subgroup of Ω_p. Assume that $\tilde{\Psi}$ is 2-mixing. Then there exist effective linear maps $V_1, \ldots, V_N$, such that, for every $\varepsilon > 0$, there exist sets $A_1, \ldots, A_N \subseteq \mathbb{Z}^2$, where A_i is $(2h + \varepsilon)$-logish with respect to V_i, such that $\bigcup_i A_i$ is a trap for $\tilde{\Psi}$.*

Example 2.8. Let $\Psi_1 = \{\nu \in \Omega_p \mid \sum_{\mathbf{w}' \in S} \nu_{\mathbf{w}+\mathbf{w}'} = 0, \mathbf{w} \in \mathbb{Z}^2\}$ for $S = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$. We prove in Proposition 3.11 that a sequence is mixing for a given principal Markov subgroup of Ω_p if and only if it is mixing for every minimal principal Markov subgroup contained in it. We conclude that no subsequence of either of the sequences $((0, 0), (c, t))_{t \in \mathbb{N}}$ or $((0, 0), (t, c))_{t \in \mathbb{N}}$, for any integer c, is mixing for $\tilde{\Psi}_1$. In particular, this system is not 2-mixing. In fact, any 2-trap for this system must contain a trap for the system Ψ_0 in Example 2.6 and the image, rotated by 90 degrees, of another such trap.

Example 2.9. Let $\Psi_2 = \{\nu \in \Omega_p \mid \sum_{\mathbf{w}' \in S} \nu_{\mathbf{w}+\mathbf{w}'} = 0, \mathbf{w} \in \mathbb{Z}^2\}$ for

$$S = \{(0, 0), (1, 0), (0, 1), (1, 2), (1, 3), (2, 2), (2, 5), (2, 6), (3, 5)\}.$$

Ψ contains two minimal Markov subgroups. One is Ledrappier's group. The other one is $\mathbb{L}' = \{\nu \in \Omega_p \mid \sum_{\mathbf{w}' \in S'} \nu_{\mathbf{w}+\mathbf{w}'} = 0, \mathbf{w} \in \mathbb{Z}^2\}$ for $S' = \{(0, 0), (1, 2), (2, 5)\}$. Since $\{(1, 2), (2, 5)\}$ is a basis of $\mathbb{Z}^2$, this system is essentially isomorphic to Ledrappier's, in the sense that there exist isomorphisms $\alpha : \mathbb{Z}^2 \to \mathbb{Z}^2$, $\beta : \mathbb{L} \to \mathbb{L}'$ satisfying $S^{\alpha(\mathbf{w})}\beta(x) = \beta(S^{\mathbf{w}}x)$ for any $\mathbf{w} \in \mathbb{Z}^2$ and any $x \in \mathbb{L}$. In fact, they can be defined as

$$\alpha(n, m) = (n + 2m, 2n + 5m), \qquad \beta(\nu)_{\mathbf{w}} = \nu_{\alpha^{-1}(\mathbf{w})}.$$

If A is a trap for $\mathbb{L}$, then $\alpha(A)$ is a trap for $\mathbb{L}'$. It follows that, using the results in [**1**], we can find a logish trap for this system, which is, in particular, logish in every direction. However,the existence of a trap which is logish in one direction, rather than a union as in Theorem 2.7, fails to follow from the results in the present paper.

As a consequence of the proofs of the above theorems, we obtain the following theorem.

THEOREM 2.10. *Let Ψ be a principal Markov subgroup of Ω_p. Assume that $\tilde{\Psi}$ is 2-mixing. Let $B(t) = ((m_1(t), n_1(t)), \ldots, (m_r(t), n_r(t)))$ be an r-sequence such that, for every $i = 1, \ldots, r$, both $m_i(t)$ and $n_i(t)$ are polynomials in t. Assume that $B(t)$ spreads, or, equivalently, that the difference $(m_i(t), n_i(t)) - (m_j(t), n_j(t))$ is not a constant for any pair (i, j). Then there is a logish set $A \subseteq \mathbb{Z}$ such that the r-sequence $(B(t_k))_{k\in\mathbb{N}}$ is mixing for $\tilde{\Psi}$ provided that $t_k \notin A$ for all $k \in \mathbb{N}$.*

It can be proved that, if Ψ satisfies the hypotheses of Theorem 2.4, then there are different choices for the map V (see Remark 2.5). If it was possible to choose a trap A that is logish simultaneously with respect to either map in a basis of $(\mathbb{R}^2)^*$, then A would be logish by Proposition 2.3. Unfortunately, the intersection of traps is not always a trap, so, at this point, we are unable to prove the following conjecture.

CONJECTURE 2.11. *Let Ψ be a minimal principal Markov subgroup of Ω_p. Assume that $\tilde{\Psi}$ is 2-mixing. Then, for every r, there is a positive integer $h' = h'(r)$ such that, for every $\varepsilon > 0$, there exists an $(h' + \varepsilon)$-logish r-trap A for $\tilde{\Psi}$.*

The condition on 2-mixing is needed, as shown by Example 2.6. On the other hand, as we see at the end of the section, Conjecture 2.11 holds for Ledrappier's subgroup $\mathbb{L}$ of Ω_2.

A sequence $(\mathbf{a}(t))_{t\in\mathbb{N}}$ in a metric space (M, d) *remains close* to a subset $A \subseteq M$ if the distance $d(\mathbf{a}(t), A)$ is bounded as a function of t, and it *gets away* from A if $d(\mathbf{a}(t), A) \longrightarrow_{t\to\infty} \infty$. Notice that a sequence gets away from a set if and only if no subsequence remains close to the set, and, conversely, a sequence remains close to a set if and only if no subsequence gets away from it.

Remark 2.12. One may consider a weaker concept than that of a trap. A subset $A \subseteq \mathbb{Z}^2$ is an *r-sub-trap* for $\tilde{\Psi}$ if every spreading r-sequence $(B(t))_{t\in\mathbb{N}}$, where $B(t) = (\mathbf{w}_1(t), \ldots, \mathbf{w}_r(t))$, satisfying $d(\mathbf{w}_i(t) - \mathbf{w}_j(t), A) \to \infty$ as $t \to \infty$, is mixing for $\tilde{\Psi}$. A set C such that $v(t) \notin C$ for all t and $v(t) \to \infty$ as $t \to \infty$ imply that $v(t)$ gets away from A is a *shell* for A. It follows that if A is a sub-trap for Ψ and C is a shell for A, then C is a

trap for Ψ. A shell for a set A can be constructed by taking a ball around every point a of A and making the radius of the ball tend to ∞ as $a \to \infty$. The rate at which the radius of the ball tends to ∞ can be made as small as needed. In particular, every h-logish set has an $(h+\varepsilon)$-logish shell, for every $\varepsilon > 0$. Note, however, that a set that is logish with respect to some linear map V might fail to have a shell that is logish with respect to V. In fact, any vertical line is a sub-trap for the system $\tilde{\Psi}_0$ of Example 2.6.

We recall a few definitions and results from [**1**]. A set $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_r\}$ of distinct points in $\mathbb{Z}^2$ is a *special r-gon* if $\sum_{\mathbf{w}\in W} \nu_{\mathbf{w}} = 0$ for all $\nu \in \mathbb{L}$. Let $\Lambda \subset (\mathbb{Z}^2)^r$ be the set of elements $(\mathbf{w}_1, \ldots, \mathbf{w}_r) \in (\mathbb{Z}^2)^r$ such that, for some $i_1, \ldots, i_s \in \{1, \ldots, r\}$, the set $\{\mathbf{w}_{i_1}, \ldots, \mathbf{w}_{i_s}\}$ is a special s-gon. Let ρ denote the Hausdorff distance in $\mathbb{Z}^2$. Every element of $(\mathbb{Z}^2)^r$ can be regarded as a subset of $\mathbb{Z}^2$. If an r-sequence $B(t)$ satisfies $\rho(B(t), \Lambda) \to \infty$ as $t \to \infty$, then $B(t)$ is mixing for $\tilde{\mathbb{L}}$ [**1**, Theorem 3.3]. A special r-gon is *connected* if none of its proper subsets is a special r'-gon for some $r' < r$.

Let $A \triangle B$ denote the symmetric difference of the sets A and B, so that the characteristic function $1_{A\triangle B} : \mathbb{Z}^2 \to \mathbb{F}_2$ is the pointwise sum $1_A + 1_B$. In [**1**], the characteristic function 1_A is identified with the polynomial $\sum_{(m,n)\in A} X^m Y^n$. This identification is only possible when $p = 2$, but can be generalized with the help of the concept of support that we defined in §4.

PROPOSITION 2.13. *Conjecture 2.11 holds in the particular case* $\Psi = \mathbb{L}$.

Proof. It was proved in [**1**, Theorem 7.1] that, for every special r-gon B_r, the characteristic function 1_{B_r} is the sum of at most r^3 characteristic functions 1_T of special triangles, i.e., sets of the form $T = \{w, w + (0, 2^s), w + (2^s, 0)\}$. In the proof of [**1**, Lemma 7.6], we defined a graph whose vertices correspond to these triangles, and two vertices are neighbors if and only if the corresponding triangles T and T' satisfy $T \cap T' \neq \emptyset$. Note that a special r-gon is connected if and only if the corresponding graph is connected for any decomposition into triangles. It follows that in a connected special r-gon $\{\mathbf{w}_1, \ldots, \mathbf{w}_r\}$, all the coordinates of every difference $\mathbf{w}_i - \mathbf{w}_j$, for $i, j \in \{1, \ldots, r\}$, are sums or differences of no more than r^3 powers of two. The set $C(r) \subseteq \mathbb{Z}^2$ of such elements is logish. Note that each element of the set $\Lambda \subseteq (\mathbb{Z}^2)^r$ contains a connected special r'-gon for some r' between three and r, while a non-mixing sequence must remain close to Λ. It follows that a non-mixing sequence must have some difference $\mathbf{w}(t)_{i(t)} - \mathbf{w}(t)_{j(t)}$ that remains close to $C(r)$, which is, therefore, a sub-trap as defined in Remark 2.12. The result follows. □

Neither the set Λ nor the set $\mathcal{L}$ of special r-gons is logish, but the subset of connected special r-gons containing $(0, 0)$ is logish in $(\mathbb{Z}^2)^r$ by the preceding proof.

3. *Dual interpretation of mixing*

In this section, we recall several results on Pontryagin duality, used in the subsequent work. We refer to [**7**] for more details.

The Pontryagin dual $\hat{\Omega}_p$ of the compact group Ω_p is the discrete group $\bigoplus_{(m,n)\in\mathbb{Z}^2} \mathbb{F}_p$, which is isomorphic to the additive group of the ring of Laurent polynomials $R = \mathbb{F}_p[X^{\pm 1}, Y^{\pm 1}]$. The dual endomorphisms of the downward and the leftward shifts on $\mathbb{F}_p^{\mathbb{Z}^2}$ are multiplication by X and multiplication by Y, respectively, in R. For $f \in R$,

we denote by χ_f the corresponding element in $\hat{\Omega}_p$. To simplify the notation, we write $U^{(m,n)} = X^m Y^n$. If $f = \sum_{\mathbf{w}} \alpha_{\mathbf{w}} U^{\mathbf{w}}$, then $\chi_f(\nu) = e^{(2\pi i/p)\sum_{\mathbf{w}} \alpha_{\mathbf{w}} \nu_{\mathbf{w}}}$. The exponential is well defined since $\sum_{\mathbf{w}} \alpha_{\mathbf{w}} \nu_{\mathbf{w}}$ is an element of $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$.

PROPOSITION 3.1. *A Markov subgroup Θ of Ω_p is the annihilator $\{\nu \in \Omega_p \mid \chi_f(\nu) = 1, f \in I\}$ of an ideal I of R, and* vice versa. *The dual $\hat{\Theta}$ of Θ is isomorphic to the quotient ring R/I.*

For the proof, we just need to note that an additive subgroup of R forms an ideal if and only if it is invariant under multiplication by $X^{\pm 1}$ and by $Y^{\pm 1}$. Note that the correspondence between Markov subgroups and ideals is bijective and reverses inclusions.

Let $f = \sum_{\mathbf{w}_i} \alpha_{\mathbf{w}_i} U^{\mathbf{w}_i} \in R$ as before. Note that the annihilator of the principal ideal $\langle f \rangle$ is the principal Markov subgroup Ψ defined by (1). The group Ψ is denoted by Ψ_f. For example, if $\Delta = 1 + X + Y$, the group Ψ_Δ is Ledrappier's subgroup $\mathbb{L}$. Since R is a Noetherian ring, every ideal is finitely generated. Note also that $\Psi_f \subseteq \Psi_g$ if and only if $\langle g \rangle \subseteq \langle f \rangle$, i.e., f divides g. It follows that the principal Markov subgroup Ψ_f is minimal if and only if $f \in R$ is irreducible. This implies the following proposition.

PROPOSITION 3.2. *Every Markov subgroup is an intersection of finitely many principal Markov subgroups. Every principal Markov subgroup contains finitely many, but at least one, minimal principal Markov subgroups.*

Example 3.3. Consider the non-principal Markov subgroup $\{\nu \in \Omega_p \mid \nu_{\mathbf{w}} = \nu_{\mathbf{w}+(1,0)} = \nu_{\mathbf{w}+(0,1)}, \mathbf{w} \in \mathbb{Z}^2\}$. This is the group of constant elements in Ω_p and its dual is $R/\langle 1 - x, 1 - y \rangle \cong \mathbb{F}_p$. On the other hand, the non-principal Markov subgroup

$$\{\nu \in \Omega_2 \mid \nu_{\mathbf{w}} + \nu_{\mathbf{w}+(1,0)} + \nu_{\mathbf{w}+(0,2)} + \nu_{\mathbf{w}+(1,1)} = \nu_{\mathbf{w}} + \nu_{\mathbf{w}+(0,1)} + \nu_{\mathbf{w}+(2,0)} + \nu_{\mathbf{w}+(1,1)} = 0, \mathbf{w} \in \mathbb{Z}^2\},$$

is annihilated by the ideal $\langle 1 + x + y^2 + xy, 1 + y + x^2 + xy \rangle = \langle 1 + x, 1 + y \rangle \langle 1 + x + y \rangle$. It follows that this group contains Ledrappier's subgroup $\mathbb{L}$ as a finite index subgroup. More precisely, $\Psi/\mathbb{L}$ has two elements, with the non-trivial class represented by the element $\mu \in \Omega_2$ having each coordinate equal to one.

PROPOSITION 3.4. *The ring R/I is infinite if and only if I is contained in some principal ideal $\langle f \rangle \neq R$. Equivalently, a Markov subgroup is infinite if and only if it contains a principal Markov subgroup.*

Proof. Assume first that $I \subseteq \langle f \rangle$ for some f. Without loss of generality, we may assume that f depends non-trivially on Y, i.e., it is not of the form $Y^r s(X)$ for a Laurent polynomial s in X. Then f cannot divide a Laurent polynomial in X. It follows that if x is the image of X in $R/\langle f \rangle$, then its powers $1, x, x^2, \ldots$ are all different, so that R/I is infinite.

On the other hand, if I is not contained in any principal ideal, it must contain elements $g_1, \ldots, g_N$ with no common divisor. Let $R' = \mathbb{F}_p[X, Y]$. Multiplying by powers of X and Y, if needed, we may assume that $g_1, \ldots, g_N$ belong to R' and are relatively prime as elements of R'. Since $\mathbb{F}_p[X]$ is a unique factorization domain, Gauss's lemma applies. It follows that $g_1, \ldots, g_N$ do not have common factors in $R'' = \mathbb{F}_p(X)[Y]$. Since R''

is a principal ideal domain, there exist $a_1, \ldots, a_N \in R''$ such that $a_1 g_1 + \cdots + a_N g_N = 1$. Note that $a_i \in R''$ means that a_i is a polynomial on Y whose coefficients are rational functions on X, i.e.,

$$a_i = \frac{h_{i,0}(X)}{k_{i,0}(X)} + \frac{h_{i,1}(X)}{k_{i,1}(X)} Y + \cdots + \frac{h_{i,d}(X)}{k_{i,d}(X)} Y^d,$$

and hence we can define $b_i = a_i k(X)$, where $k(X) = \prod_{i,j} k_{i,j}(X)$, and obtain $b_1, \ldots, b_N \in R'$ such that $b_1 g_1 + \cdots + b_N g_N = k(X)$ is a polynomial in X, which belongs to the ideal I, and therefore its image in R/I vanishes. It follows that the image x of X in R/I satisfies a non-trivial polynomial equation $k(x) = 0$, and hence x is algebraic over $\mathbb{F}_p$. By the same token, the image y of Y is algebraic and therefore the ring $\mathbb{F}_p[x^{\pm 1}, y^{\pm 1}] \subseteq \mathbb{F}_p(x, y)$ is finite. □

Remark 3.5. A well-known fact in algebraic geometry is that ideals in $\mathbb{F}_p[X, Y]$, generated by relatively prime polynomials $g_1, \ldots, g_N$ as above, are precisely the ideals with a finite number of zeros over the algebraic closure $\overline{\mathbb{F}_p}$. This is a consequence of Hilbert's Nullstellensatz (cf. [**4**]). Examples of such ideals for $p = 2$ are $\langle 1 + x, 1 + y \rangle$, $\langle 1 + x^4, 1 + y^4, (1 + x)(1 + y) \rangle$ and $\langle 1 + xy, 1 + x + y \rangle$. The first two have $(1, 1)$ as their only zero, while the third is the ideal of polynomials vanishing on the Galois orbit of the pair $(\alpha, \alpha + 1)$ for any generator of $\mathbb{F}_4$ as an extension of $\mathbb{F}_2$.

If f is the greatest common divisor of the elements of I, then the set $J = \{g \in R \mid gf \in I\}$ is also an ideal. Furthermore, $I = fJ$, and the greatest common divisor of the elements of J is one. It follows that $\langle f \rangle / I \cong R/J$ is finite. We obtain the following corollary.

COROLLARY 3.6. *Let I be the annihilator of a Markov subgroup Θ. If f is the greatest common divisor of the elements of I, then the quotient group Θ / Ψ_f is finite.*

Next we give an algebraic characterization of mixing r-sequences that will be useful in the proof of Theorem 2.4. Denote by $u^{\mathbf{w}}$ the image of $U^{\mathbf{w}}$ in R/I.

PROPOSITION 3.7. *Let I be the annihilator of the Markov subgroup Θ. The r-sequence $(B(t))_{t \in \mathbb{N}}$, where $B(t) = (\mathbf{w}_1(t), \ldots, \mathbf{w}_r(t))$, is mixing for $\tilde{\Theta}$ if and only if, for every $P_1, \ldots, P_r \in R/I$, not all zero, the equation*

$$\sum_{i=1}^{r} u^{\mathbf{w}_i(t)} P_i = 0 \tag{4}$$

has only finitely many solutions $t \in \mathbb{N}$.

Proof. Recall that $\int_\Theta 1 \, d\mu_\Theta = 1$ and $\int_\Theta \chi \, d\mu_\Theta = 0$ for every non-trivial character $\chi \in \hat{\Theta}$. Assume that (4) has infinitely many solutions. Passing to a subsequence, we may assume it is identically zero, so we can take the corresponding character and integrate to obtain

$$\int_\Theta \left(\prod_{i=1}^{r} \chi_{P_i} \circ S^{\mathbf{w}_i(t)} \right) d\mu_\Theta = \int_\Theta 1 \, d\mu_\Theta = 1,$$

while, on the other hand, $\prod_{i=1}^{r} \int_\Theta \chi_{P_i} \, d\mu_\Theta = 0$, since at least one factor vanishes. The same argument proves that (2) goes to the right limit whenever $f_1, \ldots, f_r$ are characters.

Since linear combinations of characters are dense in the space $C(\Theta)$ of continuous functions with the uniform topology, (2) holds actually for continuous functions. Now the property can be extended to $L^\infty(\Theta, \mathcal{B}_\Theta, \mu_\Theta)$ since a measurable function coincides with a continuous function having the same bound, in the complement of a set of an arbitrarily small measure, by the general form of Lusin's theorem. □

Example 3.8. In the ring $R/\langle\Delta\rangle$, $x^{2^t} + y^{2^t} + 1 = 0$ for all t. This proves that, for Ledrappier's subgroup $\mathbb{L} = \Psi_\Delta$, the system $\tilde{\mathbb{L}}$ is not 3-mixing. Note that this was, in fact, the first example of a $\mathbb{Z}^2$-action that was 2-mixing but not 3-mixing.

Example 3.9. Similarly to the preceding example, if $g = x^6 + x^5y + x^3y^2 + y + y^3$, the system $\tilde{\Psi}_g$ is 4-mixing but not 5-mixing. This is due to the fact that the convex hull of $S(g)$ is a pentagon. In fact, in [**3**] it is proved that, when $I = \langle g\rangle$, any spreading sequence of solutions of (4) has a convex hull with a side that is asymptotically parallel to each side of $S(g)$.

Example 3.10. Assume that f is supported on a line L. By a translation, we may assume that $\mathbf{0} = (0, 0) \in S(f)$ and that all other points in $S(f)$ are on the same side of the origin. In particular, L is a 1-dimensional subspace. Then f is a polynomial on $U^{\mathbf{w}}$, where $\mathbf{w}$ is a generator of $L \cap \mathbb{Z}^2$, and therefore it divides $U^{n\mathbf{w}} - 1$ for some positive integer n. It follows that $B(t) = (\mathbf{0}, p^t n\mathbf{w})$ is not mixing for $\tilde{\Psi}_f$. On the other hand, if $S(f)$ is not contained in a line, its convex hull has two non-parallel sides and $\tilde{\Psi}_f$ must be 2-mixing by the results in [**3**] quoted above.

PROPOSITION 3.11. *An r-sequence $(B(t))_{t\in\mathbb{N}}$ is mixing for $\tilde{\Psi}_f$ if and only if it is mixing for $\tilde{\Psi}_{f_i}$ for every irreducible divisor f_i of f.*

Proof. Let $B(t) = (\mathbf{w}_1(t), \ldots, \mathbf{w}_r(t))$ and let $I = \langle f\rangle$. Assume that there exist $P_1, \ldots, P_n$, not all zero, in R/I, such that (4) holds for infinitely many t. By passing to a subsequence, we may assume that it holds for all t. Let $Q_1, \ldots, Q_n$ be their pre-images in R. It follows that f divides $\sum_{j=1}^r U^{\mathbf{w}_j(t)} Q_j$ for all t, but f does not divide all of $Q_1, \ldots, Q_n$. Let g be the greatest common divisor of $f, Q_1, \ldots, Q_n$, and let f_i be an irreducible divisor of fg^{-1}. Then f_i cannot divide all of $g^{-1}Q_1, \ldots, g^{-1}Q_n$, but it does divide $\sum_{j=1}^r U^{\mathbf{w}_j(t)} g^{-1} Q_j$. It follows that $(B(t))$ is not mixing for $\tilde{\Psi}_{f_i}$. In the converse direction, if f_i divides $\sum_{j=1}^r U^{\mathbf{w}_j(t)} S_j$ for all t, but it does not divide all of $S_1, \ldots, S_n$, then f divides $\sum_{j=1}^r U^{\mathbf{w}_j(t)} (ff_i^{-1}) S_j$, but not all of $(ff_i^{-1})S_1, \ldots, (ff_i^{-1})S_n$. □

Example 3.12. The system $\tilde{\Psi}_f$ is 2-mixing if and only if no divisor of f is supported on a line.

PROPOSITION 3.13. *The system $\tilde{\Theta}$ is ergodic if and only if Θ is a principal Markov subgroup.*

Proof. If f is the greatest common divisor of the elements of I, where Θ is the annihilator of I, then Ψ_f is a closed invariant subgroup. Furthermore, the quotient Θ/Ψ_f is finite by Corollary 3.6, and hence Ψ_f is open in Θ and therefore of positive measure. Hence, if $\tilde{\Theta}$ is ergodic, then $\Theta = \Psi_f$. On the other hand, if $\Theta = \Psi_f$, we can choose $\mathbf{w} \in \mathbb{Z}^2$, which is

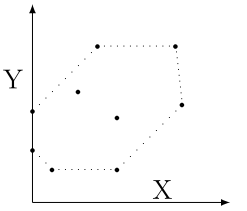

FIGURE 1. The support of an element f in good position.

not a multiple of another element of $\mathbb{Z}^2$, such that no irreducible divisor of f is a Laurent polynomial in $U^{\mathbf{w}}$ alone. Then the 2-sequence $(B(t))_{t\in\mathbb{N}}$, where $B(t) = (\mathbf{0}, t\mathbf{w})$, is mixing by the results of [**3**], and therefore $\tilde{\Theta}$ is ergodic. □

We assume that $\Psi = \Psi_f$ is a minimal Markov subgroup of Ω_p in all that follows.

Definition 3.14. A spreading r-sequence $(A(t))_{t\in\mathbb{N}}$, where $A(t) = (\mathbf{w}_1(t), \ldots, \mathbf{w}_r(t))$ is *exceptional* for Ψ (or f) if there exist elements $P_1, \ldots, P_r \in R/I$ such that (4) is satisfied for every t. If $(A(t))_{t\in\mathbb{N}}$ is an exceptional (respectively, mixing) r-sequence, then, for every permutation σ of $\{1, \ldots, r\}$, the r-sequence $\{\sigma[A(t)]\}_{t\in\mathbb{N}}$, where $\sigma[A(t)] = (\mathbf{w}_{\sigma(1)}(t), \ldots, \mathbf{w}_{\sigma(r)}(t))$, is exceptional (respectively, mixing). Hence, for a fixed t, we can identify $A(t)$ with a set of r points in $\mathbb{Z}^2$.

4. *Algebraic form of the main result*

Definition 4.1. Let $f = \sum_{\mathbf{w}} a_{\mathbf{w}} U^{\mathbf{w}} \in R$. The *support* of f is given by $S(f) = \{\mathbf{w} \in \mathbb{Z}^2 \mid a_{\mathbf{w}} \neq 0\}$. An element $f \in R$ is *in good position* if no point of $S(f)$ is to the left of the Y-axis and there are at least two points of $S(f)$ on the Y-axis (see Figure 1).

We denote by $D_p^{(N)}$ the set of positive integers whose base p expansion has at most N non-zero digits. The proofs of Theorem 2.4 and Theorem 2.7 are based on the following result, which will be proved in §10.

THEOREM 4.2. *Assume that f is in good position and that $S(f)$ is not contained in the Y-axis. Let h be as in (3). Then there exists a constant $C = C_f$ such that, for every exceptional r-sequence $(A(t))_{t\in\mathbb{N}}$, where*

$$A(t) = ((m_1(t), n_1(t)), \ldots, (m_r(t), n_r(t))), \tag{5}$$

there exist $1 \le i < j \le r$ and a subsequence (t_k) such that the sequence $(|m_i(t_k) - m_j(t_k)|)_{k=1}^{\infty}$ tends to ∞ and remains close to the set of differences of $CD_p^{(h)} = \{Cd \mid d \in D_p^{(h)}\}$.

Let $\mathbb{A}$ be the affine group of the plane, i.e., the group of all maps $g : \mathbb{R}^2 \to \mathbb{R}^2$ of the form $g(\mathbf{x}) = A\mathbf{x} + \mathbf{w}$, where $\mathbf{w}$ is a vector and A an invertible matrix, and let $G = \{g \in \mathbb{A} \mid g(\mathbb{Z}^2) = \mathbb{Z}^2\}$. The action of G on $\mathbb{Z}^2$ defines an action $f \mapsto {}_g f$ on R, given by

$${}_g\left(\sum_{\mathbf{w}\in\mathbb{Z}^2} a_{\mathbf{w}} U^{\mathbf{w}}\right) = \sum_{\mathbf{w}\in\mathbb{Z}^2} a_{\mathbf{w}} U^{g(\mathbf{w})}.$$

Note that $S({}_g f) = g(S(f))$. If $A(t)$ is an exceptional r-sequence for f, then $g(A(t))$ is an exceptional r-sequence for ${}_g f$. For every polynomial f with support of size at least two, there exists an element g in G such that ${}_g f$ is in good position. As $g(\mathbf{w}_1) - g(\mathbf{w}_2)$ depends linearly on $\mathbf{w}_1 - \mathbf{w}_2$, Theorem 4.2 has the following corollary.

COROLLARY 4.3. *Let f be an irreducible element of R, whose support is not contained in a line, and let h be as in (3). Then there exist an effective non-zero linear map $V : \mathbb{Z}^2 \to \mathbb{Z}$ and a constant $C = C_f$, such that, for every exceptional r-sequence $A(t)$ for f, there exist a subsequence (t_k) and indices $1 \le i < j \le r$ such that the sequence*

$$(|V(\mathbf{w}_i(t_k) - \mathbf{w}_j(t_k))|)_{k=1}^{\infty} \tag{6}$$

tends to ∞ and remains close to the set of differences of $C D_p^{(h)}$.

In terms of mixing sequences, the previous corollary implies the following one.

COROLLARY 4.4. *Let f be an irreducible element of R, whose support is not contained in a line, and let h be as in (3). Then there exist an effective non-zero linear map $V : \mathbb{Z}^2 \to \mathbb{Z}$ and a constant C such that, if $B(t) = (\mathbf{w}_1(t), \ldots, \mathbf{w}_r(t))$ is a spreading r-sequence satisfying that*

> *for every subsequence (t_k) and indices $1 \le i < j \le r$, such that the sequence (6) tends to ∞, the sequence (6) gets away from the set of differences of $C D_p^{(h)}$,*

then $B(t)$ is mixing for $\tilde{\Psi}_f$.

For every Laurent polynomial $f(X, Y) = \alpha_1 X^{m_1} Y^{n_1} + \cdots + \alpha_r X^{m_r} Y^{n_r}$, the relation $f(x, y)^{p^t} = 0$ in $R/\langle f \rangle$ produces an exceptional r-sequence

$$A_f(t) = ((m_1 p^t, n_1 p^t), \ldots, (m_k p^t, n_k p^t))$$

for f. Sequence (6) in this case is $(cp^{t_k})_{k=1}^{\infty}$ for some constant c. It follows from [**8**, Theorem 5] that cp^{t_k} gets away from the set of differences of $C D_{p'}^{(N)}$, for any constant C, whenever p' is a prime different from p. It follows that Markov subgroups of Ω_p cannot be isomorphic as dynamical systems to Markov subgroups of $\Omega_{q'} = \mathrm{GF}(q')^{\mathbb{Z}^2}$, where q' is a power of p'.

5. *Absolute values*

In this section, we present a short review of some definitions and results on local fields and non-Archimedean valuations that will be used in the subsequent work. For a more extensive account of the subject, see, for example, [**2**].

Let K be a field. A non-Archimedean absolute value on K is a homomorphism $x \mapsto |x|$ from the multiplicative group K^* of K into the multiplicative group $\mathbb{R}^+$ of positive reals, satisfying the inequality $|a + b| \le \max\{|a|, |b|\}$ for $a, b \in K^*$. By convention, we set $|0| = 0$. An absolute value defines a metric, given by $d(a, b) = |a - b|$ for $a, b \in K$. The addition, multiplication and absolute value on K extend to the completion $\bar{K}$ of K, so that $\bar{K}$ is also a field with a non-Archimedean absolute value. A series $\sum_{i=0}^{\infty} a_i$ over $\bar{K}$ converges if and only if $a_i \longrightarrow_{i \to \infty} 0$, and an infinite product $\prod_{i=0}^{\infty} b_i$ of non-zero terms

converges to some non-zero limit if and only if $b_i \longrightarrow_{i\to\infty} 1$. The field $\mathbb{Q}$ of rational numbers admits a non-Archimedean absolute value $|\cdot|_p$ for every prime p. The absolute value is defined by $|p^n(a/b)|_p = p^{-n}$ for every integer n, a, b with $(a,\ p) = (b,\ p) = 1$. The completion of $\mathbb{Q}$ with respect to this absolute value is the field $\mathbb{Q}_p$ of *p-adic numbers*. Similarly, for an arbitrary field F, the field of rational functions $F(X)$ admits a non-Archimedean absolute value $|\cdot|_0$, defined by $|X^n(f/g)|_0 = 2^{-n}$ for polynomials f and g, not divisible by X. The completion of $F(X)$ with respect to this absolute value is the field $F((X))$ of formal Laurent series.

Given a field K with a non-Archimedean absolute value, the set $\mathcal{O}_K = \{x \in K \mid |x| \leq 1\}$ is the *ring of integers* of K, and $\mathfrak{m}_K = \{x \in K \mid |x| < 1\}$ is the *maximal ideal*. The group of units of $\mathcal{O}_K$ is $\mathcal{O}_K^* = \{x \in K \mid |x| = 1\}$. If K is the field $F((X))$ with absolute value $|\cdot|_0$, then $\mathcal{O}_K$ is the ring $F[[X]]$ of Taylor series. Units in $F[[X]]$ are series with non-vanishing constant term. In particular, every element of $F((X))$ has the form $X^n v(X)$, where n is an integer and $v(X)$ is a unit in $F[[X]]$. If $K = F(X)$ with the same absolute value $|\cdot|_0$, then $\mathcal{O}_K$ is the set of rational functions f/g such that g is not divisible by X. If $K = \mathbb{Q}_p$ with the absolute value $|\cdot|_p$, the ring $\mathcal{O}_K$ is *the ring of p-adic integers*, denoted by $\mathbb{Z}_p$.

Denote by $\mathcal{O}_K((X)) \subset K((X))$ the ring of Laurent series with integral coefficients. If $f(X) = \sum_{n\geq N} a_n X^n \in \mathcal{O}_K((X))$ and $z \in \mathfrak{m}_K$ is not zero, then the series $f(z) = \sum_{n\geq N} a_n z^n$ converges in $\bar{K}$. The mapping $\phi_z : \mathcal{O}_K((X)) \to \bar{K}$, defined by $\phi_z[f(X)] = f(z)$, is a ring homomorphism. Now assume that a field F is contained in $\mathcal{O}_K$. This holds if and only if $|a| = 1$ for every non-zero element a in F. Let $\eta_z : F((X)) \to \bar{K}$ be the restriction of ϕ_z. Since $F((X))$ is a field, then η_z must be injective, i.e., $\bar{K}$ contains a copy of $F((X))$. Furthermore, η_z maps the ring of Taylor series $F[[X]]$ into $\mathcal{O}_K$, and therefore it also maps units into units. Hence, if a Laurent series $f(X)$ has the form $X^t v(X)$, where $v(X)$ is a unit, then $|f(z)| = |z|^t$, i.e.,

$$\log_{|z|} |f(z)| = \log_{|X|_0} |f(X)|_0. \tag{7}$$

PROPOSITION 5.1. *Let $f \in R$ be irreducible and in good position, and K be the quotient field of the integral domain $R/\langle f\rangle$. Then there exists an absolute value $|\cdot|$ on K, such that $|x| < 1$ and $|y| = 1$, where x and y are the images of X and Y in K.*

Proof. If f is in good position, then no negative power of X appears in f. Multiplying f by a power of the unit Y, if needed, we may assume that $(0,\ 0)$ is the lowest point of $S(f)$ on the y-axis, i.e., f has a non-zero constant coefficient. Define Z by $X = ZY^s$, and write $g(Z,\ Y) = f(ZY^s,\ Y)$. It is not hard to see that $S(g)$ is the image of $S(f)$ under the linear transformation with matrix $\left(\begin{smallmatrix} 1 & 0 \\ s & 1 \end{smallmatrix}\right)$. In particular, it is also in good position. For s sufficiently large, $S(g)$ has no point below the X-axis, i.e., g is a polynomial in Z and Y. Note that g also has a non-zero constant coefficient, since $(0,\ 0) \in S(g)$. The polynomial g has a decomposition $g(Z,\ Y) = a(Z) \prod_{i=1}^{d} (1 - \eta_i Y)$, such that the following hold.

(1) The η_i are the roots of the polynomial $\tilde{g}(Z,\ Y) = Y^d g(Z,\ Y^{-1})$, considered as a polynomial in Y, in some finite extension L of $F((Z))$. Equivalently, the η_i^{-1} are the roots of $g(Z,\ Y)$.

(2) The polynomial $a(Z)$ has a non-zero constant term, and therefore is a unit as an element of the ring $F[[Z]]$ of integers of $F((Z))$. In particular, the polynomial $a(Z)^{-1}\tilde{g}(Z, Y) = \prod_i (Y - \eta_i)$ is a monic polynomial with coefficients in $F[[Z]]$, so that the η_i are integral over $F[[Z]]$.

Note that L has a unique absolute value that extends the absolute value of $F((Z))$ [**2**, p. 114]. Since the η_i are integral over $F[[Z]]$, if $|\cdot|$ is the absolute value on L, then $|\eta_i| \leq 1$. Now set

$$g(Z, Y) = a(Z) + s_1(Z)Y + s_2(Z)Y^2 + \cdots + s_d(Z)Y^d. \tag{8}$$

Recall that g is in good position, and hence $g(0, Y)$ is not a constant. It follows that at least one $s_i(0)$ does not vanish, and therefore $s_i(Z)$ is a unit in $F[[Z]] \subseteq \mathcal{O}_L$. Since $s_i(Z) = (-1)^i a(Z)\sigma_i(\eta_1, \ldots, \eta_d)$, where σ_i is the ith elementary symmetric function, at least one of the η_i must be a unit. Now observe that K can be embedded into L by sending x to $Z\eta_i^{-s}$ and y to η_i^{-1}, since $f(Z\eta_i^{-s}, \eta_i^{-1}) = g(Z, \eta_i^{-1}) = 0$. The required absolute value is the pre-image under this map of the absolute value of L. □

6. *Independence of p-adic powers*

In this section, we study some independence properties of p-adic powers of $1 + X$ over a field F. From now on, F is any field of characteristic p. Most of the results here are probably known, but we present them in full for the convenience of the reader.

Any element $d \in \mathbb{Z}_p$ has a unique p-adic expansion of the form

$$d = \sum_{i=0}^{\infty} a_i(d)p^i, \quad 0 \leq a_i(d) \leq p - 1.$$

For $d \in \mathbb{Z}_p$, we define the p-adic power $(1 + X)^d \in F[[X]]$ by the formula

$$(1 + X)^d = \prod_{i=0}^{\infty} (1 + X^{p^i})^{a_i(d)}. \tag{9}$$

Note that this definition coincides with the usual when $d \in \mathbb{N}$. Furthermore, the congruence $d \equiv d' (\text{mod } p^n)$ implies that $(1 + X)^d \equiv (1 + X)^{d'} (\text{mod } X^{p^n})$. It follows that the function $d \mapsto (1 + X)^d$ is continuous. Therefore, all usual properties of exponents in $\mathbb{N}$ extend to $\mathbb{Z}_p$.

LEMMA 6.1. *For all $d, e \in \mathbb{Z}_p$ with $|d - e|_p = p^{-t}$, $|(1 + X)^d - (1 + X)^e|_0 = |X|_0^{p^t}$.*

Proof. Since $(1 + X)^d - (1 + X)^e = (1 + X)^e[(1 + X)^{d-e} - 1]$, we may assume that $e = 0$. If $|d|_p = p^{-t}$, then the p-adic expansion of d is $a_t(d)p^t + a_{t+1}(d)p^{t+1} + \cdots$. It follows that $(1 + X)^d \equiv (1 + X^{p^t})^{a_t(d)} \equiv 1 + a_t(d)X^{p^t} + \cdots (\text{mod } X^{p^{t+1}})$, and the result follows. □

Define a partial order on $\mathbb{Z}_p$ as follows. If $d, d' \in \mathbb{Z}_p$, with expansions $d = \sum_{i=0}^{\infty} a_i(d)p^i$ and $d' = \sum_{i=0}^{\infty} a_i(d')p^i$, then $d \preceq d'$ if $a_i(d) \leq a_i(d')$ for all i.

LEMMA 6.2.

$$(1+X)^d = \sum_{\substack{d_0 \in \mathbb{N} \\ d_0 \preceq d}} \left[\prod_{i=0}^{\infty} \binom{a_i(d)}{a_i(d_0)} \right] X^{d_0}.$$

Proof. Since, for a positive integer a, $(1+X^{p^t})^a = \sum_{b=0}^{a} \binom{a}{b} X^{bp^t}$, the result follows immediately from (9). Note that the product is finite since $d_0 \in \mathbb{N}$, so that almost all coefficients $a_i(d_0)$ are zero. □

The lemma assumes a simpler form in the case $p = 2$. In fact, since any element of $\mathbb{Z}_2$ is of the form $d_T = \sum_{i \in T} 2^i$ for a unique subset T of $\mathbb{N}$, we obtain the following lemma.

LEMMA 6.3. *If $d_T \in \mathbb{Z}_2$, then $(1+X)^{d_T} = \sum_{T_0} X^{d_{T_0}}$, where T_0 runs over all finite subsets of T.*

Now we prove a few results on the algebra of p-adic powers that will be useful in the subsequent work.

LEMMA 6.4. *Let $\gamma_i \in F$ and $e_i \in \mathbb{Z}_p$ for $1 \le i \le N$, with $e_i \neq e_j$ for $i \neq j$. If*

$$\sum_{i=1}^{N} (1+X)^{e_i} \gamma_i = 0, \tag{10}$$

then $\gamma_i = 0$ for all $i = 1, \ldots, N$.

Proof. Without loss of generality, assume that e_1 is maximal with respect to $\preceq$, i.e., $e_1 \not\preceq e_i$ for $i \ge 2$. For every $2 \le i \le N$, there exists an $m_i \in \mathbb{N}$ such that $a_{m_i}(e_1) > a_{m_i}(e_i)$. Let $D = \sum_i a_{m_i}(e_1) p^{m_i}$, so that $D \preceq e_1$, but $D \not\preceq e_i$ for $i \neq 1$. Note that $D \in \mathbb{N}$, since the sum is finite. It follows from Lemma 6.2 that the integral power x^D appears in the expression of the p-adic power $(1+X)^{e_1}$, but not in any of the others. Therefore, the left-hand side of (10) does not vanish. □

Let E^{alg} denote the algebraic closure of the field E. Lemma 6.4, applied to F^{alg}, yields the following corollary.

COROLLARY 6.5. *Distinct p-adic powers of $1+X$ are linearly independent over F^{alg}.*

Here we have a few more consequences of Lemma 6.4.

COROLLARY 6.6. *If $(1+X)^{d_1}, \ldots, (1+X)^{d_k}$ are linearly dependent over $F^{\mathrm{alg}}(X)$, then there exist $1 \le i < j \le k$ such that $d_i - d_j \in \mathbb{Z}$.*

COROLLARY 6.7. *If $(1+X)^d \in F^{\mathrm{alg}}(X)$ is a rational function, then $d \in \mathbb{Z}$.*

COROLLARY 6.8. *If $d_1, \ldots, d_k$ are linearly independent over $\mathbb{Q}$, then $(1+X)^{d_1}, \ldots, (1+X)^{d_k}$ are algebraically independent over F^{alg}.*

LEMMA 6.9. *If E/K is a finite extension and $T_1, \ldots, T_n$ are algebraically independent over K, then they are algebraically independent over E.*

Proof. If $T_1, \ldots, T_n$ satisfy the polynomial $f(Y_1, \ldots, Y_n)$, with coefficients in E, and if N is the norm from $E(Y_1, \ldots, Y_n)$ to $K(Y_1, \ldots, Y_n)$, then $g = N(f)$ is a non-zero polynomial with coefficients in K, and it is divisible by f. Hence $g(T_1, \ldots, T_n) = 0$. □

PROPOSITION 6.10. *Let*

$$\sum_{j=1}^{k} \alpha_j (1+X)^{d_j} = 0, \quad d_1, \ldots, d_k \in \mathbb{Q}_p,$$

be a vanishing linear combination with $\alpha_j \in F(X)^{\mathrm{alg}}$. *Put* $[i] = \{j \mid d_i - d_j \in \mathbb{Q}\}$ *for* $1 \le i \le k$. *Then*

$$\sum_{j \in [i]} \alpha_j (1+X)^{d_j} = 0, \quad 1 \le i \le k.$$

Proof. Let $1/N, b_1, \ldots, b_t$ be a basis of the $\mathbb{Z}$-submodule of $\mathbb{Q}_p$ generated by $1, d_1, \ldots, d_k$. For $j = 1, \ldots, k$, write

$$d_j = n_{0,j}/N + \sum_{i=1}^{t} n_{i,j} b_i.$$

For $j \neq j'$, the classes $[j]$ and $[j']$ coincide if and only if the vectors $(n_{1,j}, \ldots, n_{t,j})$ and $(n_{1,j'}, \ldots, n_{t,j'})$ are equal. Let $j_1, \ldots, j_M$ be a set of representatives of the equivalence classes $[j]$. Let $f \in F(X)^{\mathrm{alg}}[Y_1, \ldots, Y_t]$ be the polynomial defined by

$$f(Y_1, \ldots, Y_t) = \sum_{l=1}^{M} \Bigg(\sum_{j \in [j_l]} \alpha_j (1+X)^{n_{0,j}/N} \Bigg) \prod_{i=1}^{t} Y_i^{n_{i,j_l}}.$$

Note that all monomials on the right-hand side are distinct. Then

$$f((1+X)^{b_1}, \ldots, (1+X)^{b_t}) = \sum_{j} \alpha_j (1+X)^{d_j} = 0.$$

The coefficients of f are contained in some finite extension E of $F^{\mathrm{alg}}(X)$. By Corollary 6.8, the powers $(1+X)^{b_1}, \ldots, (1+X)^{b_t}$ are algebraically independent over $F^{\mathrm{alg}}(X)$. By the preceding lemma, they are also algebraically independent over E. This proves the proposition. □

COROLLARY 6.11. *If* $(1+X)^{d_1}, \ldots, (1+X)^{d_k}$ *are linearly dependent over* $F(X)^{\mathrm{alg}}$, *then* $d_i - d_j \in \mathbb{Q}$ *for some* $1 \le i < j \le k$.

COROLLARY 6.12. *If* $(1+X)^{d_1}, \ldots, (1+X)^{d_k}$ *are linearly dependent over an algebraic extension* $K/F^{\mathrm{alg}}(X)$ *that contains no roots of* $(1+X)$, *then* $d_i - d_j \in \mathbb{Z}$ *for some* $1 \le i < j \le k$.

Proof. By Proposition 6.10, we reduce the problem to the case where $d_1, \ldots, d_k \in \mathbb{Q}$. Raising to some power of p, if needed, we may assume that $d_1, \ldots, d_k \in \mathbb{Q} \cap \mathbb{Z}_p$, i.e., they have denominators relatively prime to p. Let N be relatively prime to p and divisible by the denominators of $d_1, \ldots, d_n$. It suffices to show that $1, (1+X)^{1/N}, \ldots, (1+X)^{(N-1)/N}$ are linearly independent over K. Equivalently, we need to prove that the

degree of the extension $K([1+X]^{1/N})/K$ is N. Let $f(t) = t^N - (1+X)$ be the minimal polynomial of $(1+X)^{1/N}$ over $F^{\mathrm{alg}}(X)$, and let $g(t)$ be the minimal polynomial of $(1+X)^{1/N}$ over K. Note that, over the algebraic closure $F(X)^{\mathrm{alg}}$,

$$f(t) = \prod_{i=1}^{N} (t - \lambda_i (1+X)^{1/N}).$$

Since g divides f and it is monic, it must be a sub-product of the above, and hence $g(0) = \lambda(1+X)^{M/N} \in K$, where $\lambda \in F^{\mathrm{alg}}$ is a root of unity and M is the degree of g. Since K contains λ, then it contains also

$$(1+X)^{m/N} = ((1+X)^{M/N})^r (1+X)^s,$$

where $m = rM + sN$ is the greatest common divisor between M and N, which is a root of $1+X$, since N/m is an integer. We conclude that $m = N$, and therefore $M = N$. □

7. *Mixing at polynomial times*

From now on, we assume that $\Psi = \Psi_f$ is a minimal Markov subgroup of Ω, as defined in the introduction. We let K be the quotient field of the integral domain $R/\langle f \rangle$ as in §5. We assume throughout that f is in good position. Note that this implies that f depends non-trivially on Y, and hence $x \notin \mathbb{F}_p^{\mathrm{alg}}$.

Let $\overline{K}$ be the completion of K with respect to the absolute value given by Proposition 5.1. Since K is a finite extension of $\mathbb{F}_p(x)$ (which is isomorphic to the field $\mathbb{F}_p(X)$ of rational functions), $\overline{K}$ is a finite extension of $\mathbb{F}_p((x))$ (which is isomorphic to the field $\mathbb{F}_p((X))$ of Laurent series). Let $\mathcal{O}$ be the ring of integers of $\overline{K}$ and let $\mathfrak{m}$ be its maximal ideal. Note that $\mathcal{O}$ is the completion of the ring $R/\langle f \rangle$, while $\mathfrak{m}$ is the completion of the maximal ideal associated to the absolute value. The field $\mathcal{O}/\mathfrak{m}$ is a finite extension of the field $\mathbb{F}_p = \mathbb{F}_p[[x]]/x\mathbb{F}_p[[x]]$. In particular, every non-zero element in $\mathcal{O}/\mathfrak{m}$ is a root of unity. By Hensel's lemma (cf. [**2**, p. 49]), there is a root of unity $\lambda \in \overline{K}$ such that $y \equiv \lambda\ (\mathfrak{m})$. Notice that $\mathbb{F} = \mathbb{F}_p(\lambda)$ is a finite field.

LEMMA 7.1. *Assume that $\Psi = \Psi_f$, defined as above, is 2-mixing. Then there exists an element $z \in \mathbb{F}(x, y) \subseteq \overline{K}$ such that $y = \lambda(1+z)^\tau$ with $\tau \in \mathbb{N}$ maximal. If Ψ is not 2-mixing, y is a root of unity.*

Proof. Let λ be as above. If $y = \lambda$, then y is algebraic over $\mathbb{F}_p$. This may happen only if f is independent of x, and is therefore supported on the vertical axis. In this case, f divides a polynomial of the form $1 - y^n$, so that $\Psi = \Psi_f$ is not 2-mixing. If this is not the case, the element $y/\lambda \in \mathbb{F}(x, y)$ is transcendental over $\mathbb{F}$, while x is algebraic over $\mathbb{F}(y)$. It follows that the extension $\mathbb{F}(x, y)/\mathbb{F}(y)$ is finite. Since the degree of the extension $\mathbb{F}(\sqrt[n]{y/\lambda})/\mathbb{F}(y)$ is n, there must exist a largest value $n = \tau$ of n for which $\sqrt[n]{y/\lambda}$ is contained in $\mathbb{F}(x, y)$. Set $z = \sqrt[\tau]{y/\lambda} - 1$. □

From now on, we let z be as above, with the convention that $z = 0$ if Ψ is not 2-mixing. Note that, being a root of unity, λ has only finitely many distinct powers. Taking a subsequence and redefining P_i, equation (4) takes the form

$$\sum_{i=1}^{r} P_i x^{m_i(t)} (1+z)^{\tau n_i(t)} = 0. \tag{11}$$

To illustrate our method, we prove the following result.

PROPOSITION 7.2. *Assume that $\tilde{\Psi}$ is 2-mixing. Let $B(t) = (\mathbf{w}_1(t), \ldots, \mathbf{w}_r(t))$ be an r-sequence in $(\mathbb{Z}^2)^r$, where $\mathbf{w}_i(t) = (m_i(t), n_i(t))$ for $i = 1, \ldots, r$. Let $N(t) = (n_1(t), \ldots, n_r(t)) \in \mathbb{Z}_p^r$, and assume, for every accumulation point $N = (n_1, \ldots, n_r)$ of $(N(t))$, that $\tau(n_i - n_j) \notin \mathbb{Z}$ for $1 \leq i < j \leq r$. Then $B(t)$ has no exceptional subsequence.*

Proof. It suffices to prove that, for any $P_1, \ldots, P_r \in R/I$, not all zero, equation (11) has only finitely many solutions in m. Without loss of generality, we may assume that none of the P_i is zero.

Passing to a subsequence and reordering the sub-indices, we may assume that $m_1(t) \leq \cdots \leq m_r(t)$ for all t. Dividing by $x^{m_1(t)}$, we may assume that $m_1(t) = 0$. Comparing absolute values, we see that at least $m_2(t)$ must remain bounded. Passing again to a subsequence, we may assume that all $m_i(t)$ are either constant or go to ∞. Assume that $m_i(t) = m_i$ is constant for $i = 1, \ldots, s$, and $m_i(t) \to \infty$ for $i = s+1, \ldots, r$, for some fixed $s \in \{2, 3, \ldots, r\}$. Since $\mathbb{Z}_p^r$ is compact, we may assume, by taking a subsequence, that $N(t) \longrightarrow_{t\to\infty} N$ for some $N = (n_1, \ldots, n_r) \in \mathbb{Z}_p^r$.

Passing to the limit in (11), we obtain

$$\sum_{i=1}^{s} x^{m_i}(1+z)^{\tau n_i} P_i = 0.$$

By the condition on N, the powers $(1+z)^{\tau n_i}$ are linearly independent over $\mathbb{F}(y, x)$. It follows that we must have $x^{m_i} P_i = 0$ for $i = 1, \ldots, s$ and, in particular, $P_1 = 0$, which is a contradiction. □

The following result concerns the case $r = 3$. Assume, as before, that $\tilde{\Psi}$ is 2-mixing. It follows from [**3**, Proposition 3.5] that, if $A(t) = (0, \mathbf{w}_1(t), \mathbf{w}_2(t))$, with $n_i(t) \geq 0$, is an exceptional r-sequence, then the convex hull of $S(f)$ is a triangle. If f is in good position, the vertices of this triangle are, up to translation, $(0, 0)$, $(0, a)$ and (b, c). Employing again [**3**, Proposition 3.5], after permuting $\mathbf{w}_1$ and $\mathbf{w}_2$, if necessary,

$$\mathbf{w}_1(t) = k(t)(0, a) + \mathbf{e}_1(t), \quad \mathbf{w}_2(t) = k(t)(b, c) + \mathbf{e}_2(t),$$

where $\mathbf{e}_i(t)$ is bounded. Let $\nu : \overline{K}^* \to \mathbb{R}$ be defined by $\nu(u) = \log_\delta(|u|)$ for some fixed $\delta < 1$. One normally chooses δ so that $\nu(K^*) = \mathbb{Z}$, but we make no use of this here. Recall that $|\cdot|_p$ is the standard absolute value on $\mathbb{Q}_p$.

PROPOSITION 7.3. *Let $\Psi = \Psi_f$ be as above. If $A(t) = (0, \mathbf{w}_1(t), \mathbf{w}_2(t))$ is an exceptional r-sequence as above, then $k(t) = p^{l(t)}T + \varepsilon(t)$, where T is a constant depending only on f and $\varepsilon(t)$ is bounded.*

Proof. By splitting the r-sequence into subsequences, we may assume that $\mathbf{e}_1$ and $\mathbf{e}_2$ are constants. Replacing P_i by $u^{\mathbf{e}_i} P_i$, we may assume that $\mathbf{e}_1 = \mathbf{e}_2 = (0, 0)$. In particular, $m_1(t) = 0$. Note that the formula for $\mathbf{w}_2(t)$ given above implies that $m_2(t) \to \infty$. Furthermore, splitting into subsequences, we may assume, as before, that the root of unity $\lambda^{n_1(t)} = \lambda^\beta$ is a constant. Letting $t \to \infty$ in the equation

$$P_0 + y^{n_1(t)} P_1 + x^{m_2(t)} y^{n_2(t)} P_2 = 0, \tag{12}$$

we see that $\lambda^\beta(1+z)^{\tau n} = -P_0/P_1$ for every accumulation point n of $n_1(t)$ in $\mathbb{Z}_p$. By Corollary 6.12, the power τn must be an integer. Therefore (12) becomes

$$\lambda^\beta P_1(-(1+z)^{\tau n} + (1+z)^{\tau n_1(t)}) + x^{m_2(t)} y^{n_2(t)} P_2 = 0.$$

Comparing valuations, we get $\nu(P_1) + \nu(z)p^{l(t)} = m_2(t)\nu(x) + \nu(P_2)$, where $l(t)$ is defined by $|\tau(n - n_1(t))|_p = p^{-l(t)}$ (see Lemma 6.1). Since $m_2(t) = bk(t)$, the result follows. □

Note that, with a little more work, one can obtain the formulas $T = \nu(z)/b\nu(x)$ and $\varepsilon(t) = \nu(P_1/P_2)/b\nu(x) + \pi_1[\mathbf{e}_1(t) - \mathbf{e}_2(t)]/b$, where $\pi_1(m, n) = m$. In the course of the above proof, we have seen that P_0/P_1 is the product of a root of unity and a fractional power of y. Using the action of G on R as in §1, we can obtain similar results for P_0/P_2. This shows, in an alternative way, that we are in the exceptional case of [**9**, Theorem 2].

8. Φ-*series*

Let χ be the standard valuation on the field $\mathbb{F}_2[[X]]$, i.e., $\chi(f(X)) = \log_{|X|_0} |f(X)|_0$. Recall that $\nu(f(z)) = \chi(f(X))\nu(z)$ for every power series $f(X) \in \mathbb{F}_p[[X]]$ and every non-zero element $z \in \mathfrak{m}$. A Φ-*series of size* k is a power series of the form

$$\phi(X) = \alpha_1(1+X)^{b_1} + \cdots + \alpha_k(1+X)^{b_k}, \tag{13}$$

where $\alpha_1, \ldots, \alpha_k$ are in $\mathbb{F}_p^*$ and $b_1, \ldots, b_k$ are distinct elements in $\mathbb{Z}_p$. Clearly, if $\phi(X)$ is a Φ-series of size k, then so is $(1+X)^d\phi(X)$ for every $d \in \mathbb{Z}_p$. Furthermore, if ϕ is a Φ-series of size k and ϕ' is a Φ-series of size k', then $\phi + \phi'$ is a Φ-series of size at most $k + k'$. If $\alpha_1 + \cdots + \alpha_k \neq 0$ in (13), the power series $\phi(X)$ is a unit, i.e., its valuation $\chi(\phi)$ is zero. Otherwise, it has a valuation $\chi(\phi) > 0$. We prove now that the valuation $\chi(\phi)$ has a simple p-adic expansion. For $p = 2$, we have a small bound.

LEMMA 8.1. *If $p = 2$ and ϕ is a Φ-series of size k, then $\chi(\phi)$ belongs to the set $D_p^{(k-1)}$.*

Proof. Recall that every $d \in \mathbb{Z}_2$ may be written in the form $d_T = \sum_{k \in T} 2^k$ for a unique $T \subseteq \mathbb{N}$. Let $b_1, \ldots, b_k$ be as in (13) and set $b_i = d_{T_i}$. Notice that $\alpha_1 = \cdots = \alpha_k = 1$. Since the coefficient of $X^{\chi(\phi)}$ in ϕ does not vanish, if

$$\chi(\phi) = \sum_{i=1}^{N} 2^{k_i} = d_T, \quad T = \{k_1, \ldots, k_N\},$$

then the set T must be contained in T_i for an odd number of values of i. Now assume that $N > k - 1$. For each i such that $T \not\subseteq T_i$, choose an element of T which does not belong to T_i. The set S, consisting of all these elements, is of size at most $k - 1$, and is also contained in T_i for an odd number of values of i. It follows that the coefficient of X^{d_S} in ϕ does not vanish. However, $d_S < d_T = \chi(\phi)$, which is a contradiction. □

Now we prove a result of this type for general p.

LEMMA 8.2. *If ϕ is a Φ-series of size k, then the valuation $\chi(\phi)$ is in $D_p^{(h)}$, where $h = h(k) = (p-1)p^{k+1} + k$.*

Proof. The coefficient of X^d in the power series expansion of $\phi(X)$ defined by (13) is $\Delta(d) = \sum_{d \preceq b_j} \alpha_j \delta(b_j, d)$, where $\delta(b, d) = \prod_{i=0}^{\infty} \binom{a_i(b)}{a_i(d)}$ (see Lemma 6.2). The factors $\binom{a_i(b_j)}{a_i(d)}$, for different values of j, depend only on the $(k+1)$-tuple

$$[a_i(b_1), \ldots, a_i(b_k), a_i(d)].$$

Let d be such that $\Delta(d) \neq 0$, and put $T(d) = \{i \mid a_i(d) \neq 0\}$. Assume that $T(d)$ has more than $h(k)$ elements. We claim that there exists a $d' \neq d$ satisfying the following conditions.

(1) $d' \preceq d$ and, in particular, $d' < d$.
(2) $d \preceq b_j$ if and only if $d' \preceq b_j$ for all $j = 1, \ldots, k$.
(3) $\delta(b_j, d) = \delta(b_j, d')$ for all $j = 1, \ldots, k$.

If we prove this claim, it will follow that $\Delta(d) = \Delta(d')$ and, in particular, $d \neq \chi(\phi)$.

Let $A = \{j \mid d \preceq b_j\}$. For any $j \notin A$, find an $i = i(j) \in \mathbb{N}$ such that $a_i(b_j) < a_i(d)$. Let $T_0 = \{i(j) \mid j \notin A\}$. Clearly, $T_0 \subseteq T(d)$ and $T(d) - T_0$ has more than $h - k = (p-1)p^{k+1}$ elements. By the pigeonhole principle, there exist $i_1, \ldots, i_p$ in $T(d) - T_0$ such that $a_{i_1}(b_j) = \cdots = a_{i_p}(b_j)$ for all j, and also $a_{i_1}(d) = \cdots = a_{i_p}(d)$. By Fermat's little theorem, $m^p \equiv m(p)$ for any integer m, and therefore

$$\prod_{l=1}^{p} \binom{a_{i_l}(b_j)}{a_{i_l}(d)} = \binom{a_{i_1}(b_j)}{a_{i_1}(d)}$$

as an element of $\mathbb{Z}/p\mathbb{Z}$. The element $d' = d - \sum_{l=2}^{p} a_{i_l}(d) p^{i_l}$ satisfies $a_i(d') = 0$ if $i \in \{i_2, \ldots, i_p\}$ and $a_i(d') = a_i(d)$ otherwise. Conditions 1–3 follow. □

LEMMA 8.3. *Define a sequence of* Φ*-series by*

$$\phi^{(t)}(X) = \alpha_1(1+X)^{b_1(t)} + \cdots + \alpha_k(1+X)^{b_k(t)}. \tag{14}$$

Assume that $b_i(t) \longrightarrow_{t\to\infty} b \in \mathbb{Z}_p$ *for* $i = 1, \ldots, k$. *If* $\psi(t) = \chi(\phi^{(t)})$ *and* $\lambda(t)$ *is the coefficient of* $X^{\psi(t)}$ *in* $\phi^{(t)}(X)$, *then* $\lambda(t)^{-1} X^{-\psi(t)} \phi^{(t)}(X) \longrightarrow_{t\to\infty} (1+X)^b$.

Proof. Factoring out $(1+X)^b$, if needed, we may assume that $b = 0$. Let $\beta(t)$ be such that $|b_i(t)| \leq p^{-\beta(t)}$ for all i and $\beta(t) \longrightarrow_{t\to\infty} \infty$. Then we may write $b_i(t) = p^{\beta(t)}\rho_i(t)$ for some $\rho_i(t) \in \mathbb{Z}_p$, and hence we obtain $(1+X)^{b_i(t)} = (1 + X^{p^{\beta(t)}})^{\rho_i(t)}$, and therefore $\phi^{(t)}(X)$ is a power series in $X^{p^{\beta(t)}}$. The result follows. □

9. *Main lemmas*

Throughout this section, $A(t) = (\mathbf{w}_1(t), \ldots, \mathbf{w}_r(t))$, where $\mathbf{w}_i(t) = (m_i(t), n_i(t))$, denotes an r-sequence. We assume that $A(t)$ satisfies (11), but not necessarily that it spreads. We let $\Psi = \Psi_f$ be a minimal Markov subgroup, as before, and assume that $\tilde{\Psi}$ is 2-mixing, so that z, defined as in Lemma 7.1, is non-zero.

Definition 9.1. $A(t)$ is *reduced* if it satisfies the following conditions.

- $A(t)$ is convergent in $\mathbb{Z}_p$.
- For any i, j, each coordinate of $(\mathbf{w}_i(t) - \mathbf{w}_j(t))$ is either constant or tends either to $+\infty$ or to $-\infty$.

Since $\mathbb{Z}_p$ is compact, every r-sequence $A(t)$ has a reduced subsequence. If $A(t)$ is reduced, the expression $M_i(t) = x^{m_i(t)}(1+z)^{\tau n_i(t)}P_i$ is a *monomial*. Two monomials $M_i(t)$ and $M_j(t)$ are *at the same level* if the sequence $\nu(M_i(t)) - \nu(M_j(t))$ is bounded†. The monomial $M_i(t)$ is at the *highest level* if the difference $\nu(M_i(t)) - \nu(M_j(t))$ is bounded from above for all monomials $M_j(t)$. If the two monomials $M_i(t)$ and $M_j(t)$ are at the same level, then they are *in the same component* if $\tau(n_i - n_j) \in \mathbb{Z}$, where $n_i = \lim_{t\to\infty} n_i(t)$ and $n_j = \lim_{t\to\infty} n_j(t)$.

Definition 9.2. The reduced sequence $A(t)$ *satisfies condition* ***A*** if:
(1) for any two monomials $M_i(t)$ and $M_j(t)$ at the same level, $m_i(t) = m_j(t)$; and
(2 for any two monomials $M_i(t)$ and $M_j(t)$ in the same component, $n_i = n_j$.

Monomials $M_i(t)$ and $M_j(t)$ *remain close* if they are in the same component and $n_i(t) - n_j(t)$ is bounded. The relation of remaining close is an equivalence relation for any reduced r-sequence. The corresponding equivalence classes are $\mathcal{C}$*-classes*. Two monomials M_i and M_j in the same component are Φ-equivalent if the coefficients P_i and P_j are linearly dependent over $\mathbb{F}_p$. The corresponding equivalence classes are Φ*-classes*.

Definition 9.3. The reduced sequence $A(t)$ *satisfies condition* ***B*** if:
- for any two monomials $M_i(t)$ and $M_j(t)$ that remain close, $n_i(t) = n_j(t)$ for all t; and
- no two monomials that remain close are Φ-equivalent.

Assume that $A(t)$ satisfies conditions **A** and **B**. A Φ*-term* is the sum of all monomials in a Φ-class. Note that a Φ-term may be written in the form $T_i(t) = x^{m_i(t)}\phi_i^{(t)}(z)P_1$, where $\phi_i^{(t)}(X)$ is a sequence of non-zero Φ-series satisfying the hypothesis of Lemma 8.3. Note that all monomials in a Φ-class are at the same level and component. The definitions of levels and components for Φ*-terms* are similar to those for monomials. Note, however, that two Φ-terms made of monomials at the same level need not themselves be at the same level.

LEMMA 9.4. *Let $h = k-1$ if $p = 2$, and let $h = (p-1)p^{k+1} + k$ otherwise. Let $A(t)$ be a reduced r-sequence of solutions of (11), satisfying conditions* ***A*** *and* ***B****. If every Φ-term has at most k monomials, and $\nu(x)(m_i(t) - m_j(t))$ is far from the set of differences of $\nu(z)D_p^{(h)}$ whenever $m_i(t) \neq m_j(t)$, then, for any two Φ-terms $T_i(t)$ and $T_j(t)$ at the same level, $m_i(t) = m_j(t)$ and $\chi(\phi_i^{(t)}) = \chi(\phi_j^{(t)})$ for sufficiently large t.*

Proof. By Lemma 8.1 or 8.2, the valuation $\chi(\phi_i^{(t)})$ is in $D_p^{(h)}$. Now, the condition on $m_i(t) - m_j(t)$ implies that the levels, for Φ-terms made of monomials at different levels, must be different themselves. For the last statement, observe that if $m_i(t) = m_j(t)$, then $\chi(\phi_i^{(t)}) - \chi(\phi_j^{(t)})$ is bounded. As in the proof of Lemma 8.3, each Φ-series $\phi_i^{(t)}$ may be written as the product of a unit $(1+X)^{b_i}$ and a power series in $X^{p^{\beta_i(t)}}$ for increasing values of $\beta_i(t)$. It follows that $\phi_i^{(t)}/\phi_j^{(t)}$ is, up to a unit, a power series in $X^{p^{\beta(t)}}$ for increasing values of $\beta(t) = \min\{\beta_i(t), \beta_j(t)\}$, and therefore $\chi(\phi_i^{(t)}) - \chi(\phi_j^{(t)})$ must be eventually zero. □

† This is equivalent to the condition that $m_i(t) - m_j(t)$ is bounded; we formulated it in this way to keep the analogy with the definition of levels and components for Φ-terms that follows.

Note that the previous result states that Φ-terms in the same component are sums of monomials in the same component.

LEMMA 9.5. *Under the same conditions as the preceding lemma, if $T_{i_1}(t), \ldots, T_{i_s}(t)$ are all the Φ-terms in one component of the highest level, then every element of $\{P_{i_1}, \ldots, P_{i_s}\}$ is a linear combination of the others with coefficients in $\mathbb{F}_p$.*

Proof. By the preceding lemma, all the series $\phi_{i_1}^{(t)}, \ldots, \phi_{i_s}^{(t)}$ have eventually the same valuations, say, $\psi(t)$. It follows from Lemma 8.3 that there exist non-zero constants $\lambda_1, \ldots, \lambda_s \in \mathbb{F}_p$ and a subsequence (t_k) such that $z^{-\psi(t_k)}\phi_i^{(t_k)}(z) \longrightarrow_{t\to\infty} \lambda_l(1+z)^{\tau n_i}$ for $l = 1, \ldots, s$. Also, $m_{i_l}(t_k) = m(t_k)$ is independent of l. Hence, dividing by $x^{m(t_k)}z^{\psi(t_k)}$ and passing to the limit,

$$\sum_i \lambda_i (1+z)^{\tau n_i} P_i = 0,$$

where the sum runs over all indices i such that the Φ-term $T_i(t)$ is at the highest level. Note that all Φ-terms that are not in the highest level have zero contribution to this limit, by definition. By the linear independence, the sum vanishes also for every component, but condition **A** implies that n_i is constant on every component. □

LEMMA 9.6. *Let $A(t)$ be a reduced r-sequence of solutions of (11), satisfying conditions **A** and **B**. Assume that there are at most k $\mathcal{C}$-classes in every component. Let h be as in Lemma 9.4. If $v(x)(m_i(t) - m_j(t))$ is far from the set of differences of $v(z)D_p^{(h)}$ whenever $m_i(t) \neq m_j(t)$, then the sum of the monomials in every $\mathcal{C}$-class is identically zero.*

Proof. Without loss of generality, we assume that $P_j \neq 0$ for each j. Let $\{T_j(t) \mid j \in \mathfrak{C}\}$ be a component of Φ-terms at the highest level. Lemma 9.5 gives us $\sum_{j\in\mathfrak{C}} \lambda_j P_j = 0$. Fix $i \in \mathfrak{C}$. If there are no other elements in $\mathfrak{C}$, then $P_i = 0$, contrary to the assumption. Otherwise, we may set $P_i = \sum_{j\in\mathfrak{C}\setminus\{i\}} (\lambda_j\lambda_i^{-1})P_j$ and replace every monomial with the factor P_i by a sum of monomials with the factors P_j for $j \neq i$. If we produce two monomials of the form $\alpha M(t)$ and $\beta M(t)$, where α and β are in $\mathbb{F}_p$, we replace them by $(\alpha+\beta)M(t)$, unless $\alpha + \beta = 0$, in which case we delete them. Note that:

- this procedure preserves condition **A**;
- no two new monomials can be both in the same Φ-class and in the same $\mathcal{C}$-class, since conditions **A** and **B** for the original monomials imply that two such new monomials would be of the form $\alpha M(t)$ and $\beta M(t)$, and would have, therefore, been reduced; in particular, the procedure preserves condition **B**;
- since one monomial is replaced by a sum of monomials in the same $\mathcal{C}$-class, it does not increase the number of $\mathcal{C}$-classes, and the sum of all the monomials in every fixed $\mathcal{C}$-class does not change; and
- the procedure may increase the number of monomials, but it decreases the number of Φ-terms.

If we iterate this procedure until we have no more Φ-terms, all monomials have been deleted, and therefore the sum of every $\mathcal{C}$-class must have been zero to begin with. □

10. *Conclusion of the proofs*

Proof of Theorem 4.2. Set $y = \lambda(1+z)^\tau$ as in Lemma 7.1. Note that $z \neq 0$, since f is not supported on a line. The constant in the theorem is $C_f = \nu(z)/\nu(x)$. Assume that we have a spreading r-sequence $A(t)$ of solutions of (4), where each difference of corresponding coordinates is either bounded or far from $C_f D_p^{(h)}$.

By taking a subsequence so that $\lambda^{n_i(t)}$ is constant, and modifying P_i, if needed, we may assume that $\lambda = 1$, and thus we obtain a solution of (11). By taking a subsequence, we may assume that it is reduced. Taking a subsequence and redefining P_i, $m_i(t)$ and $n_i(t)$, we may assume that condition **A** is satisfied. Since no two different monomials remain close, condition **B** is vacuously satisfied (either before or after assuming condition **A**). It follows that Lemma 9.6 applies. Again, since no two different monomials remain close, every $\mathcal{C}$-class contains a unique monomial by definition. This implies that the monomials are zero. This contradiction shows that the sequence $A(t)$ cannot exist. □

Proof of Theorem 2.4. Let $\Psi = \Psi_f$ be a minimal principal Markov subgroup of Ω_p, and assume that $\tilde{\psi}$ is 2-mixing. Then f is not supported in a line by Example 3.10. By applying an automorphism of $\mathbb{Z}^2$, we may assume that f is in good position and, in this case, we may choose V as the projection on the first coordinate.

Take an exceptional (and therefore spreading) sequence $A(t)$ as in (5). Passing to a subsequence, if needed, by Theorem 4.2 there exist two indices $1 \leq i < j \leq r$ such that $(|m_i(t) - m_j(t)|)$ tends to ∞ and remains close to the set D' of differences of $C_f D_p^{(h)}$. The set D' is $(2h)$-logish, and hence, for every $\varepsilon > 0$, there is a $(2h+\varepsilon)$-logish shell D_ε for D'. It follows that $V^{-1}(D_\varepsilon)$ is a trap for $\tilde{\Psi}$. □

Proof of Theorem 2.7. Let Ψ be a principal Markov subgroup of Ω_p, and assume that $\tilde{\Psi}$ is 2-mixing. Then the same is true for every minimal principal Markov subgroup Ψ_f of Ψ, by Theorem 3.11. We can, therefore, apply Theorem 2.4 to each of these minimal principal Markov subgroups. Furthermore, if A_f is a trap for $\tilde{\Psi}_f$, then the union $\bigcup_f A_f$ over all such f, which is a finite union, is a trap for $\tilde{\Psi}$, again by Theorem 3.11. The result follows. □

Proof of Theorem 2.10. Let Ψ be a principal Markov subgroup of Ω_p, and assume that $\tilde{\Psi}$ is 2-mixing, as before. Note that we may assume that Ψ is minimal, since a union of logish sets is logish. Let $V : \mathbb{Z}^2 \to \mathbb{Z}$ be the linear map in Corollary 4.4. By the condition that $\tilde{\Psi}$ is 2-mixing, we have different choices for V, so we may always assume that the image of B is not contained in a line $V = c$.

It suffices to prove that A may be chosen so that, for every sequence $(t_k)_{k\in\mathbb{N}}$ satisfying $t_k \notin A$ for all k, the following condition holds. *If i and j are different elements of $\{1, \ldots, r\}$ such that, for $i \neq j$, the expression $|V(m_i(t_k), n_i(t_k)) - V(m_i(t_k), n_i(t_k))|$ is unbounded, then it stays away from $C_f D_p^{(h)}$.* Note that $f(t) = V(m_i(t), n_i(t)) - V(m_i(t), n_i(t))$ is a non-constant polynomial in t. In particular, there exists a constant c such that $|f(t)| \leq c|t|^d$, where d is the degree of f. Furthermore, every point of $\mathbb{Z}$ has at most d pre-images. Let $D = D_\varepsilon$ be the $(2h+\varepsilon)$-logish set in $\mathbb{Z}$ defined in the proof of Theorem 2.4. Then

$$|f^{-1}(D) \cap C_N| \leq d|D \cap C_{cN^d}| = O[(\log cN^d)^h] = O[(\log N)^h],$$

where $C_N = \{-N, -N+1, \ldots, N-1, N\}$. It follows that the set $A = f^{-1}(D)$ is logish. The result follows. □

Acknowledgements. This research was partially supported by Fondecyt, Grant #1160603, the Milken Families Foundation Chair in Mathematics and NSF Grant DMS-1500575.

REFERENCES

[**1**] L. Arenas-Carmona, D. Berend and V. Bergelson. Ledrappier's system is almost mixing of all orders. *Ergod. Th. & Dynam. Sys.* **28** (2008), 339–365.
[**2**] J. W. S. Cassels. *Local Fields*. Cambridge University Press, Cambridge, 1986.
[**3**] M. Einsiedler and T. Ward. Asymptotic geometry of non-mixing sequences. *Ergod. Th. & Dynam. Sys.* **23** (2003), 75–85.
[**4**] W. Fulton. *Algebraic Curves. An Introduction to Algebraic Geometry*. Addison-Wesley Publishing Company, Advanced Book Program, Redwood City, CA, 1989, notes written with the collaboration of Richard Weiss. Reprint of 1969 original. Advanced Book Classics.
[**5**] B. Kitchens and K. Schmidt. Markov subgroups of $(\mathbb{Z}/2\mathbb{Z})^{\mathbb{Z}^2}$. *Symbolic Dynamics and its Applications (New Haven, CT, 1991) (Contemporary Mathematics, 135)*. American Mathematical Society, Providence, RI, 1992, pp. 265–283.
[**6**] F. Ledrappier. Un champ Markovien peut être d'entropie nulle et mélangeant. *C. R. Math. Acad. Sci. Paris* **287** (1978), 561–563.
[**7**] K. Schmidt. *Dynamical Systems of Algebraic Origin*. Birkhäuser, Basel, 1995.
[**8**] R. Tijdeman and L. Wang. Sum of products of powers of given prime numbers. *Pacific J. Math.* **132** (1988), 177–193.
[**9**] J. F. Voloch. The equation $ax + by = 1$ in characteristic p. *J. Number Theory* **73** (1998), 195–200.