ON MEASURE-PRESERVING $\mathcal{C}^1$ TRANSFORMATIONS OF COMPACT-OPEN SUBSETS OF NON-ARCHIMEDEAN LOCAL FIELDS

JAMES KINGSBERY, ALEX LEVIN, ANATOLY PREYGEL, AND CESAR E. SILVA

Abstract. We introduce the notion of a *locally scaling* transformation defined on a compact-open subset of a non-archimedean local field. We show that this class encompasses the Haar measure-preserving transformations defined by $\mathcal{C}^1$ (in particular, polynomial) maps, and prove a structure theorem for locally scaling transformations. We use the theory of polynomial approximation on compact-open subsets of non-archimedean local fields to demonstrate the existence of ergodic Markov, and mixing Markov transformations defined by such polynomial maps. We also give simple sufficient conditions on the Mahler expansion of a continuous map $\mathbb{Z}_p \to \mathbb{Z}_p$ for it to define a Bernoulli transformation.

1. Introduction

The p-adic numbers have arisen in a natural way in the study of some dynamical systems, for example in the study of group automorphisms of solenoids in Lind and Schmidt [LS94]; other situations in dynamics where the p-adic numbers come up are surveyed in Ward [War]. At the same time there has been interest in studying the dynamics (topological, complex, or measurable) of naturally arising maps (such as polynomials) defined on the p-adics; see for example Benedetto [Ben01], Khrennikov and Nilson [KN04], and Rivera-Letelier [RL03]. In particular, Bryk and Silva in [BS05] studied the measurable dynamics of simple polynomials on balls and spheres on the field $\mathbb{Q}_p$ of p-adic numbers. The maps they studied are ergodic but not totally ergodic and they asked whether there exist polynomials on $\mathbb{Q}_p$ that define (Haar) measure-preserving transformations that are mixing. Woodcock and Smart in [WS98] show that the polynomial map $x \mapsto \frac{x^p - x}{p}$ defines a Bernoulli, hence mixing, transformation on $\mathbb{Z}_p$. A consequence of our work is a significant extension of the result for this map, placing it in a greater context (see in particular Example 8.5).

Rather than working on $\mathbb{Q}_p$ we find that the natural setting for our work is over a non-archimedean local field K. We introduce a class of transformations, called locally scaling, and show in Lemma 4.3 that measure-preserving $\mathcal{C}^1$ (in particular, polynomial) maps are locally scaling. In Section 5 we apply the theory of Markov shifts to classify the dynamics of locally scaling transformations, decomposing the transformation into a disjoint union of ergodic Markov transformations and local isometries. In particular, we show that a weakly mixing locally scaling transformation must be mixing. We also show the existence of polynomials defining transformations exhibiting nearly the full range of behaviors possible for locally scaling transformations, such as ergodic Markov, mixing Markov, and Bernoulli transformations.

Date: October 18, 2005.

1991 *Mathematics Subject Classification.* Primary 37A05; Secondary 37F10.

Key words and phrases. Measure-preserving, ergodic, non-archimedean local field.

Given a polynomial defined on a compact-open subset of K, our work shows that a finite computation may check whether it defines a measure-preserving transformation and whether it defines a mixing transformation; the question of ergodicity is also answered, except in the case where the polynomial is 1-Lipschitz, which has been studied by Anashin in [Ana02].

We now indicate an outline of the rest of the paper. Section 2 reviews results on Markov shifts, Section 3 reviews preliminaries on non-archimedean local fields as well as some analytic definitions, and Section 6 recalls some of the theory of polynomial approximation on rings of integers of non-archimedean local fields.

Section 4 establishes the relationship that measure-preserving $\mathcal{C}^1$ maps are locally scaling, and Section 5 proves our main structural results, in particular Proposition 5.8 and Theorem 5.9. Section 7, in particular Theorem 7.2, shows that polynomial maps are in a sense a representative class of locally scaling transformations, and the existence of polynomial maps defining locally scaling transformation with various behaviors, including mixing. Section 8 and Section 9 are devoted to demonstrating two interesting classes of locally scaling maps on $\mathbb{Z}_p$ that arise naturally in the study of polynomial approximations. Specifically, Section 8 studies maps which are isometrically conjugate to the natural realization of the Bernoulli shift, and shows for instance that the map $x \mapsto \binom{x}{p^\ell}$ on $\mathbb{Z}_p$ is Bernoulli; these results are then used in Section 10 to construct maps $\mathbb{N} \to \mathbb{Z}$ whose continous extensions $\mathbb{Z}_p \to \mathbb{Z}_p$ are Bernoulli for each prime p. Section 9 then studies similar binomial-coefficient maps which are locally scaling and so have very regular structures but fail to be Haar measure-preserving. Finally, Section 11 briefly indicates how our results on polynomials extend to rational functions.

1.1. **Acknowledgements.** This paper is based on research by the Ergodic Theory group of the 2005 SMALL summer research project at Williams College. Support for the project was provided by National Science Foundation REU Grant DMS - 0353634 and the Bronfman Science Center of Williams College.

2. Markov shifts

Let S be a finite non-empty set. By a *stochastic matrix on S* we mean a map $A : S^2 \to \mathbb{R}_{\geq 0}$ such that

$$\sum_{j \in S} A(i,j) = 1 \qquad \text{for each } i \in S.$$

Putting S into a bijection with the set $\{0, \ldots, \#S - 1\}$ we may regard A as a $\#S \times \#S$ matrix with non-negative entries and the entries in each row summing to 1. In analogy with this case, we will refer to the sets $\{A(i, \cdot)\}$ and $\{A(\cdot, j)\}$ as *rows* and *columns of A*, respectively.

By a *row vector on S* we mean a map $\lambda : S \to \mathbb{R}$. For λ a row vector and A a stochastic matrix, we define their product as the row vector λA defined by

$$\lambda A(j) = \sum_{i \in S} \lambda(i) A(i,j).$$

We will say that λ is non-negative (resp. positive) if it takes values in $\mathbb{R}_{\geq 0}$ (resp $\mathbb{R}_{>0}$).

To any stochastic matrix A we may associate the following symbolic dynamical system:

(i) Let
$$X_A = \{x \in \prod_{i \geq 0} S : A(\pi_n(x), \pi_{n+1}(x)) \neq 0 \text{ for all } n \geq 0\}$$
where $\pi_n : \prod_{i\geq 0} S \to S$ is projection to the n^{th} coordinate. Give each finite factor the discrete topology, and X_A the subspace topology inherited from the product topology.

(ii) Let T_A be defined by $\pi_n \circ T_A = \pi_{n+1}$. Then, (X_A, T_A) is a topological dynamical system.

(iii) If in addition λ is a non-negative row vector such that $\lambda = \lambda A$, then we may define a measure on X_A by
$$\mu_{A,\lambda}([d_0 d_1 d_2 \dots d_\ell]) = \lambda(d_0)A(d_0, d_1)\cdots A(d_{\ell-1}d_\ell), \text{ where } [d_0 \dots d_\ell] \stackrel{\text{def}}{=} \bigcap_{n=0}^{\ell} \pi_n^{-1}(d_n).$$
We call a set of the form $[d_0 \dots d_\ell]$ a *cylinder set*; we may observe that the cylinder sets form a base for the topology on X_A. Note that if λ is in fact positive, then $\mu_{A,\lambda}$ assigns positive measure to each cylinder set and hence to each open set. We may check that the condition that $\lambda = \lambda A$ implies that $(X_A, \mu_{A,\lambda}, T_A)$ is a measure-preserving measurable dynamical system.

We call such a dynamical system a *Markov shift*.

We say that a stochastic matrix A is *irreducible* or *ergodic* if for each $i, j \in S$ there exists a $n \in \mathbb{N}$ such that $A^n(i,j) > 0$. This has a natural interpretation in terms of the connectedness of a certain directed graph associated with A. We say that a stochastic matrix A is *primitive* if there exists a $n \in \mathbb{N}$ such that $A^n(i,j) > 0$ for all $i, j \in S$.

Using the Perron-Frobenius Theorem on non-negative irreducible and primitive matrices, along with a graph theoretic interpretation of the stochastic matrix, one may obtain an ergodic decomposition result for Markov shifts:

Proposition 2.1. *Let A be a stochastic matrix, and λ a positive row vector such that $\lambda = \lambda A$. Then, we may partition S into disjoint sets*
$$S = \bigsqcup_{k=1}^{n} S_k$$
such that

(i) *$A(i,j) = 0$ for $i \in S_k, j \in S_\ell$ with $k \neq \ell$; and*

(ii) *$A_k = S|_{S_k \times S_k}$ is irreducible for $k = 1, \dots, n$*

Then, $\lambda_k = \lambda|_{S_k}$ satisfies $\lambda_k = \lambda_k A_k$. And we have the ergodic decomposition of $(X_A, \mu_{A,\lambda}, T_A)$ as
$$(X_A, \mu_{A,\lambda}, T_A) = \bigsqcup_{k=1}^{n} (X_{A_k}, \mu_{A_k,\lambda_k}, T_{A_k}).$$
Moreover, the k^{th} summand is mixing if A_k is primitive.

Proof. Construct a graph on S as follows. We place a directed edge from $i \to j$ if and only if $A(i,j) > 0$. As λ is strictly positive, this is equivalent to the condition that $\lambda(i)A(i,j) > 0$.

We say that the flow or flux associated to this edge is $\lambda(i)A(i,j)$. Now, the flow out of i is

$$\sum_{j\in S}\lambda(i)A(i,j)=\lambda(i)$$

as A is a stochastic matrix. The flow into i is

$$\sum_{k\in S}\lambda(k)A(k,i)=\lambda(i)$$

as $\lambda=\lambda A$. So, we see that the flux into and out of i are both equal to $\lambda(i)$.

This implies that for every finite subset of S, the in-flux and out-flux will be equal. For $i\in S$, let $R(i)$ be the set of points reachable from i, and $B(i)$ the set of points which can reach i. Note that $R(i)$ has out-flux 0 by construction, and $B(i)$ has in-flux 0 by construction; as S is finite, these subsets are finite, so both have in-flux and out-flux equal to 0.

Now, we can have no edges into or out of either of these two sets. But, if $t\in R(i)$ and $y\in B(i)$, then there is a path from y to t; so we must have $t\in B(i)$ and $y\in R(i)$, and so $B(i)=R(i)$. So, $B(i)=R(i)$ is strongly connected, and there are no edges into or out of this set.

For $\ell>0$, note that $A^{\ell}(i,j)>0$ is equivalent to there being a path of length precisely ℓ from i to j. It follows that the collection

$$\{B(i):i\in S\}$$

gives our desired decomposition of S.

We readily note that $\lambda_k=\lambda_k A_k$ for for $k=1,\dots,n$. Then, as A_k is irreducible, [Wal82, Theorem 1.19] implies that the k^{th} summand is ergodic, from which the ergodic decomposition follows. Finally, [Wal82, Theorem 1.31] implies that the k^{th} summand is mixing if A_k is primitive. □

3. Analytic definitions, preliminaries, and notation

Let K be a *non-archimedean local field*, which we take to be either a finite field extension of $\mathbb{Q}_p$ or $\mathbb{F}_{p^n}((t))$ for some prime p.

Let $|\cdot|$ be a non-archimedean multiplicative valuation on K, such that $|\cdot|$ generates the topology on K. Denote $\mathcal{V}=|K^{\times}|=\{|x|:x\in K^{\times}\}$, $\mathcal{O}=\{x\in K:|x|\le 1\}$ and $\mathfrak{p}=\{x\in K:|x|<1\}$ (note that $\mathcal{O},\mathfrak{p}$ are independent of the choice of valuation). It is the case that $\mathfrak{p}$ is the maximal ideal of $\mathcal{O}$, and $\mathcal{O}/\mathfrak{p}$ is a finite field (the *residue field*). Let $p=\operatorname{char}\mathcal{O}/\mathfrak{p}$, $q=\#\mathcal{O}/\mathfrak{p}$, both finite with q a power of p.

We denote

$$B_r(x)=\{y\in K:|x-y|\le r\},$$

and call such a set a *ball*. Let μ be Haar measure on K, normalized such that $\mu(\mathcal{O})=1$; define $\rho:\mathcal{V}\to\mathbb{R}_{>0}$ by $\rho(r)=\mu(B_r(0))$.

Now, we recall the following standard results:

(i) $\mathcal{O}$ is the maximal compact subring of K;
(ii) $\mathcal{O}$ is a discrete valuation ring with unique maximal ideal $\mathfrak{p}$;
(iii) $\mathfrak{p}=\pi\mathcal{O}$ for any $\pi\in\mathfrak{p}\setminus\mathfrak{p}^2$; we call any such π a *uniformizing parameter*;
(iv) $\mathcal{V}$ is the discrete abelian (multiplicative) subgroup of $\mathbb{Q}$ generated by $|\pi|$; in light of this, we may define a map $v:K\to\mathbb{Z}\cup\{+\infty\}$ defined by $v(0)=+\infty$ and $v(x)=\log_{|\pi|}|x|$ for $x\in K^{\times}$; this is the additive valuation on K;

(v) For $r = |\pi|^k$, $k \geq 0$ it is the case that

$$\rho(r) = \mu(B_r(0)) = \left(\#\mathcal{O}/\mathfrak{p}^k\right)^{-1} = q^{-k}.$$

Indeed, for $r \in \mathcal{V}$ we see that $\rho(r) = q^{-\log_{|\pi|} r}$.

(vi) A subset $X \subseteq K$ is compact-open if and only if X is a finite union of balls.

We direct the interested reader to [Ser62] for a thorough treatment of related topics.

We will continue to use the symbols $K, \mu, p, q, |\cdot|, \mathcal{O}, \mathfrak{p}, \pi, v, \mathcal{V}, \rho, B_r$ with these meanings below.

Let X be an open subset of K and $a \in X$. Then, we say that a function $f : X \to K$ is *strictly differentiable* or $\mathcal{C}^1$ at a (denoted $f \in \mathcal{C}^1(a)$) if the limit

$$\lim_{\substack{(x,y)\to(a,a)\\ x\neq y}} \frac{f(x) - f(y)}{x - y}$$

exists. We write $f \in \mathcal{C}^1(X)$ if $f \in \mathcal{C}^1(a)$ for each $a \in X$. For more on this notion, see [Sch84] or [Rob00].

4. Measure preserving $\mathcal{C}^1$ maps on non-archimedean local fields

Lemma 4.1. *Let $X \subseteq K$ be open. Let $f \in \mathcal{C}^1(X)$ be such that $f(X) \subseteq X$ and such that the transformation $f : X \to X$ is measure-preserving with respect to μ. Then, $|f'(a)| \geq 1$ for all $a \in X$.*

Proof. Say there is an $a \in X$ with $|f'(a)| < 1$. Take $\alpha \in \mathcal{V}$ such that $|f'(x)| \leq \alpha < 1$.

As X is open and f is continuous, there exists $r' \in \mathcal{V}$ such that $B_r(a) \subseteq X$ and $B_{\alpha r}(f(a)) \subseteq X$ for any $r \leq r'$. Moreover, as $f \in \mathcal{C}^1(a)$ we may take $r \in \mathcal{V}$, with $r \leq r'$, such that for any $x, y \in B_r(a)$ with $x \neq y$

$$\left| \frac{f(x) - f(y)}{x - y} - f'(a) \right| \leq \alpha.$$

Then, for $x, y \in B_r(a)$ it is the case that

$$|f(x) - f(y)| \leq |x - y| \max\{\alpha, |f'(a)|\} \leq |x - y|\alpha.$$

So, $B_{\alpha r}(f(a)) \subseteq X$ by construction and moreover

$$f^{-1}\left(B_{\alpha r}(f(a))\right) \supseteq B_r(a).$$

Taking measure on both sides we get the inequality

$$\rho(r) = \mu(B_r(a)) \leq \mu\left(f^{-1}(B_{\alpha r}(f(a)))\right).$$

But as $\alpha \in \mathcal{V}$ and $\alpha < 1$ we have the strict inequality

$$\mu\left(B_{\alpha r}(f(a))\right) = \rho(\alpha r) = \rho(\alpha)\rho(r) < \rho(r) = \mu(B_r(a)).$$

So,

$$\mu\left(B_{\alpha r}(f(a))\right) < \mu\left(f^{-1}\left(B_{\alpha r}(f(a))\right)\right),$$

and in particular f is not measure-preserving. $\square$

Definition 4.2. For $X \subseteq K$ compact-open, we say that a transformation $T : X \to X$ is *locally scaling* for $r \in \mathcal{V}$ if X is a finite-union of r-balls and if there exists a function $C : X \to \mathbb{R}_{\geq 1}$ such that

$$|x - y| \leq r \Rightarrow |T(x) - T(y)| = C(x)|x - y|.$$

We will refer to C as the *scaling function*.

Let us note the following properties of locally scaling transformations:

(i) By the symmetry of x and y, we must have C constant on cosets of $B_r(0)$. We will write $H = X/B_r(0)$ for the set of cosets of $B_r(0)$ contained in X (recall that X is a union of such cosets); we treat elements of H as subsets of X. Then, C induces a map $C : H \to \mathbb{R}_{\geq 1}$.

(ii) If T is also differentiable (say a polynomial) on X, then $C(a) = |T'(a)|$ for all $a \in X$.

(iii) If T is invertible, then $C(a) = 1$ for all $a \in X$, so T is locally an isometry.

Now, we prove a lemma relating the above notion to our situation; the idea and proof of this lemma, as well as the later Lemma 5.1, is similar to results in [Rob00, §5.1.1, 5.1.2] and [Sch84, Prop. 27.3].

Lemma 4.3. *Let $X \subseteq K$ be compact-open. Let $f \in \mathcal{C}^1(X)$ be such that $f(X) \subseteq X$ and such that $|f'(a)| \geq 1$ for all $a \in X$. Then, f is locally scaling for some $r \in \mathcal{V}$. In particular, if $f : X \to X$ is measure-preserving with respect to μ, then f is locally scaling for some $r \in \mathcal{V}$.*

Proof. For each $a \in X$, it is the case that $|f'(x)| \in \mathcal{V}$. As $\mathcal{V}$ is discrete, and $f \in C^1(a)$, there exists an $r_a \in \mathcal{V}$ such that $B_{r_a}(a) \subseteq X$ and for $x, y \in B_{r_a}(a)$, $x \neq y$ the inequality

$$\left| \frac{f(x) - f(y)}{x - y} - f'(a) \right| < |f'(a)|$$

holds. In particular, this implies that

$$|f(x) - f(y)| = |f'(a)||x - y|.$$

for all $x, y \in B_{r_a}(a)$. Note that the existence of $f'(x), f'(y)$ along with the continuity of $|\cdot| : K \to \mathbb{R}_{\geq 0}$ implies that $|f'(x)| = |f'(y)| = |f'(a)|$.

Now, the collection

$$\{B_{r_a}(a) : a \in X\}$$

gives an open cover of X. As X is compact, there exists a finite subcover, corresponding to $a_1, a_2, \ldots, a_k \in X$. Let $r = \min\{r_{a_1}, \ldots, r_{a_k}\}$.

For $|x - y| \leq r$, $B_r(x) = B_r(y)$ is contained in $B_{r_{a_i}}(a_i)$ for some $i \in \{1, \ldots, k\}$. Then,

$$|f(x) - f(y)| = |f'(a_i)||x - y| = |f'(x)||x - y|.$$

Letting $C : X \to \mathbb{R}_{\geq 1}$ be given by $C(x) = |f'(x)|$, we see that f is locally scaling for $r \in \mathcal{V}$.

The final part of the lemma now follows from Lemma 4.1. $\square$

Lemma 4.4. *Let $f \in K[x]$. Then, $f \in \mathcal{C}^1(K)$.*

Say $X \subseteq K$ is compact-open, with $f(X) \subseteq X$ and such that $f : X \to X$ is measure-preserving with respect to μ. Then, $f : X \to X$ is locally scaling for some $r \in \mathcal{V}$.

Proof. Say
$$f = \sum_{k=0}^{n} a_k x^k.$$
Then we have the formal equality
$$\frac{f(x) - f(y)}{x - y} = \sum_{k=0}^{n} a_k \left(\sum_{\ell=0}^{k-1} x^\ell y^{k-1-\ell} \right).$$
So, the difference quotient extends to a continuous map $K \times K \to K$. This suffices to show the first part of our claim. The final part follows by Lemma 4.3. $\square$

Example 4.5. Consider the map
$$f(x) = \binom{x}{2} = \frac{x(x-1)}{2}$$
on $\mathbb{Z}_2$. Note that $f(\mathbb{Z}_2) \subseteq \mathbb{Z}_2$, as
$$\left| \frac{x(x-1)}{2} \right| = 2|x||x-1| \leq 1.$$

Note also that
$$|f'(x)| = \left| \frac{2x-1}{2} \right| = 2|2x-1| = 2.$$
Now, we may calculate
$$|f(x) - f(y)| = \left| \frac{(x-y)(x+y-1)}{2} \right| = 2|x-y||x+y-1|.$$
So, we see that f is locally scaling for $r = 1/2$, as $|x+y-1| = 1$ if $|x-y| \leq 1/2$ (i.e. if x and y have the same parity).

We will see in Section 8 that $f : \mathbb{Z}_2 \to \mathbb{Z}_2$ is actually measure-preserving and in fact Bernoulli.

Example 4.6. Consider the map $f : \mathbb{Z}_2 \to \mathbb{Z}_2$ given by
$$f(x) = \frac{x^4 + 2x^3 - x^2 - 2x}{8}.$$

We note that
$$f'(x) = \frac{4x^3 + 6x^2 - 2x - 2}{8} = \frac{2x^3 + 3x^2 - x - 1}{4}.$$
Working modulo 2, we may verify that the numerator is always odd. So, $|f'(x)| = 4$ for all $x \in \mathbb{Z}_2$

In order to find a radius $r \in \mathcal{V}$ for which f is locally scaling, we will use the Taylor expansion of f (this idea is similar to that in [KN04, p. 33, Lemma 1.6]):
$$f(x+z) - f(x) = zf'(x) + \frac{z^2}{2} f''(x) + \frac{z^3}{3!} f'''(x) + \frac{z^4}{4!} f^{(4)}(x).$$
So, by the strong triangle inequality it suffices to choose r such that
$$\left| \frac{f^{(n)}(x)}{n!} \right| r^{(n-1)} < |f'(x)| = 4$$

for all $x \in \mathbb{Z}_p$ and for $n \geq 2$.

We observe that

$$\left|\frac{f''(x)}{2}\right| = \left|\frac{6x^2+6x-1}{8}\right| = 8, \qquad \left|\frac{f'''(x)}{6}\right| = \left|\frac{2x+1}{4}\right| = 4, \qquad \left|\frac{f^{(4)}(x)}{24}\right| = \left|\frac{1}{8}\right| = 8.$$

So, f is locally scaling for $r = 1/4$.

5. Structure of locally scaling transformations

Lemma 5.1. *Let $X \subseteq K$ be compact-open, with $T : X \to X$ locally scaling for $r \in \mathcal{V}$. Let $C : X \to \mathbb{R}_{\geq 1}$ be the scaling function of Definition 4.2 Then, for each $a \in X$ and $r' \in \mathcal{V}$ with $r' \leq r$, the map*

$$T|_{B_{r'}(a)} : B_{r'}(a) \to B_{r'C(a)}(T(a))$$

is a bijection.

Proof. Denote $B = B_{r'}(a)$ and $B' = B_{r'C(a)}(T(a))$. As C is constant on $B_r(a)$, $|Tx - Ty| = C(x)|x-y| = C(a)|x-y|$ for all $x, y \in B \subseteq B_r(a)$. This implies that $T(B) \subseteq B'$, so our restriction is well-defined. It also implies that the restriction is injective.

For each $k \geq 0$ we may take coset representatives $a_0, \dots, a_{q^k-1}$ for $B/B_{r'|\pi|^k}(0)$. Then for $i, j \in \{0, \dots, q^k - 1\}$ we have

$$|T(a_i) - T(a_j)| = C(a)|a_i - a_j| > r'C(a)|\pi|^k.$$

So, $T(a_0), \dots, T(a_{q^k-1})$ are precisely the q^k coset representatives for $B'/B_{r'C(a)|\pi|^k}(0)$. It follows that $T(B)$ is dense in B'.

Now, note that $T|_B$ is continuous. So, $T(B)$ is the continuous image of a compact set, thus compact, and so closed. So, $T(B) = B'$. This proves surjectivity, and the lemma is proved. □

Corollary 5.2. *Let X, T, C be as in Lemma 5.1. Set $H = X/B_r(0)$. For any $i, j \in H$, $a \in j$, and $r' \in \mathcal{V}$ with $r' \leq r$, the set*

$$i \cap T^{-1}\left(B_{r'}(a)\right)$$

is either the empty set or a ball of radius $r'/C(i)$, according as whether $i \cap T^{-1}(j)$ is empty or not.

Proof. Denote $B = B_{r'}(a)$. Assume $i \cap T^{-1}(j)$ is not empty, so there is a $y \in i \cap T^{-1}(j)$. Then, the map

$$T|_{B_r(y)} : i = B_r(y) \to B_{rC(i)}(T(y)) \supseteq j \supseteq B$$

is a bijection. It follows that $i \cap T^{-1}(B)$ is non-empty, and we may in fact assume that $y \in i \cap T^{-1}(B)$.

Then, as

$$T|_{B_{r'/C(i)}(y)} : B_{r'/C(i)}(y) \to B_{r'}(T(y)) = B$$

is also a bijection, it follows that $i \cap T^{-1}(B) = B_{r'/C(i)}(y)$. □

Definition 5.3. Let $X \subseteq K$ be compact-open, and let $T : X \to X$ be locally scaling for $r \in \mathcal{V}$. Let $H = X/B_r(0)$ and $C : H \to \mathbb{R}_{\geq 0}$ be the scaling function. Then, we define the *associated transition matrix* to be the map $A : H^2 \to \mathbb{R}_{\geq 0}$ given by, for $i, j \in H$,

$$A(i,j) = \begin{cases} 0 & i \cap T^{-1}(j) = \emptyset \\ \rho\left(1/C(i)\right) & \text{otherwise} \end{cases}$$

Lemma 5.4. *Let $X \subseteq K$ be compact-open and $T : X \to X$ be locally scaling for $r \in \mathcal{V}$; let $H = X/B_r(0)$ and let $A : H^2 \to \mathbb{R}_{\geq 0}$ be the associated transition matrix. Then:*

(i) *For $S \subseteq X$ measurable and $i \in H$*

$$\mu(i \cap T^{-1}(S)) = \sum_{j \in H} \mu(S \cap j) A(i,j).$$

(ii) $A(i,j) = \frac{1}{\rho(r)} \mu(i \cap T^{-1}(j))$;

(iii) *A is a stochastic matrix on H;*

(iv) *T is measure-preserving if and only if the sum of each column of A is* 1.

Proof.

(i):

By disjoint additivity of μ, it suffices to prove the equality in the case $S \subseteq j$ for some $j \in H$. As the balls form a sufficient semi-ring in the Borel σ-algebra of X, we may in addition assume that S is a ball. Say $S = B_{r'}(a)$ for $r' \leq r$ and $a \in j$. Then, by Corollary 5.2 we know that $i \cap T^{-1}(S)$ is either the empty set or a ball of radius $r'/C(i)$, according as whether $i \cap T^{-1}(j)$ is empty or not. Taking measures we get

$$\begin{aligned} \mu\left(i \cap T^{-1}(S)\right) &= \begin{cases} 0 & i \cap T^{-1}(j) = \emptyset \\ \rho(r'/C(i)) & \text{otherwise} \end{cases} \\ &= \mu(S) A(i,j) = \sum_{j \in H} \mu(S \cap j) A(i,j). \end{aligned}$$

(ii):

Put $S = j$ in (i). Then we get

$$\mu(i \cap T^{-1}(j)) = \mu(j \cap j) A(i,j) = \rho(r) A(i,j).$$

(iii):

Note that for each $i \in H$, by disjoint additivity of μ along with (ii) we have

$$\rho(r) \sum_{j \in H} A(i,j) = \sum_{j \in H} \mu(i \cap T^{-1}(j)) = \mu(i \cap X) = \mu(i) = \rho(r).$$

(iv):

If T is measure-preserving then for each $j \in H$ we have, by disjoint additivity of μ,

$$\sum_{i \in H} A(i,j) = \frac{1}{\rho(r)} \sum_{i \in H} \mu(i \cap T^{1}(j)) = \frac{1}{\rho(r)} \mu(X \cap T^{-1}(j)) = \frac{1}{\rho(r)} \mu(T^{-1}(j)) = 1.$$

For the converse we use (i) and disjoint additivity:

$$\mu(T^{-1}(S)) = \sum_{i\in H} \mu(i \cap T^{-1}(S)) = \sum_{i,j\in H} \mu(S \cap j)A(i,j) = \sum_{j\in H} \mu(S\cap j) = \mu(S). \qquad \square$$

Remark 5.5. Given $f \in K[x]$ such that $f(X) \subseteq X$, Lemma 4.3 and Lemma 5.4 allow us to determine whether $f : X \to X$ is measure-preserving. By a linear change of variables, we may assume that $X \subseteq \mathcal{O}$. Then, we first check whether $|f'(a)| \geq 1$ for all $a \in X$ (writing $f = h/\pi^k$ with $h \in \mathcal{O}[x]$ reduces this to checking that $|h'(a)| \geq 1/|\pi^k|$ for $a \in X$, which is a finite computation in $\mathcal{O}/\mathfrak{p}^k$). Then, we may take $r \in \mathcal{V}$ with $r \leq |\pi|^k$ such that X is a finite union of r-balls, and f will be locally scaling for r. Then, we may compute the associated transition matrix in a finite computation (checking which balls intersect under image). Finally, we check the column sums.

Example 5.6. Consider the transformation on $\mathbb{Z}_2$ given by $f(x)$ from Example 4.6. Recall that f is locally scaling for $r = 1/4$. Let $H = \mathbb{Z}_2/B_{1/4}(0) = \mathbb{Z}_2/4\mathbb{Z}_2$. The associated transition matrix is thus a 4×4 matrix. As the scaling function is $C(a) = 4$ for $a \in X$, each non-zero entry of the associated transition matrix, A, is $\rho(4) = 1/4$. By Lemma 5.4, A is a stochastic matrix and so each entry in the matrix must be non-zero and so equal to $1/4$. Then, by Lemma 5.4 we see that the transformation defined by f is measure-preserving.

Example 5.7. We now look at a polynomial map on $\mathbb{Z}_2$ that defines a locally scaling but not measure-preserving transformation:

$$f(x) = \binom{x}{3} = \frac{x(x-1)(x-2)}{3!}.$$

We note that

$$|f'(x)| = \left|\frac{3x^2-6x+2}{6}\right| = \begin{cases} 1 & x \in B_{1/2}(0) \\ 2 & x \in B_{1/2}(1) \end{cases}$$

where the final equality follows by computing the values of the numerator modulo 4. In particular, we see that $|f'(x)| \geq 1$ for all $x \in \mathbb{Z}_2$ and so f is locally scaling by Lemma 4.3. We note that

$$\left|\frac{f''(x)}{2}\right| = \left|\frac{x-1}{2}\right| = 2|x-1|,$$

and $|f'''(x)/6| = |1/6| = 2$. So, we may carry out a computation as in Example 4.6, using the different bounds on $|f'(x)|$ and $|f''(x)|$ on $B_{1/2}(0)$ and $B_{1/2}(1)$, to see that f is locally scaling for $r = 1/2$.

On $B_{1/2}(0)$, f has scaling constant 1, while on $B_{1/2}(1)$ it has scaling constant 2. We note that the pre-image of $1 + 2\mathbb{Z}_2$ does not intersect $2\mathbb{Z}_2$, but all other coset/pre-image pairs do intersect. This gives us the associated transition matrix

$$A = (A(i,j))_{0\leq i,j\leq 1} = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix},$$

identifying i with $i + 2\mathbb{Z}_2$ for $i \in \{0,1\}$.

So, $\binom{x}{3}$ is not measure-preserving by Lemma 5.4.

Proposition 5.8. *Let $X \subseteq K$ be compact-open, let $T : X \to X$ be a locally scaling transformation for $r \in \mathcal{V}$, with $\Sigma = (X, \mu, T)$ the corresponding measurable dynamic system.*

Let $H = X/B_r(0)$, let $A : H^2 \to \mathbb{R}_{\geq 0}$ be the associated transition matrix, and $\lambda : H \to \mathbb{R}_{\geq 0}$ the positive row vector given by $\lambda(i) = \rho(r)$ for $i \in H$. Let $\Sigma' = (X_A, \mu_{A,\lambda}, T_A)$ be the corresponding Markov shift.

Then, there exists a continuous and measure-preserving map $\Phi : X \to X_A$ demonstrating Σ' as a topological and measurable factor of Σ. Moreover, the pre-image under Φ of a cylinder set is a ball of the same measure.

Proof. For each $k \geq 0$, let $\pi_n : X_A \to H$ denote projection to the n^{th} coordinate. Let $\phi : X \to H$ be the canonical projection. Consider the map $\Phi : X \to X_A$ defined by

$$\pi_n \circ \Phi = \phi \circ T^n,$$

i.e. the n^{th} slot in X_A denotes which element of H the point $T^n(x)$ is in. Then $\Phi \circ T = T_A \circ \Phi$ by construction.

Let $d_0, d_1, \ldots \in H$. We will prove by induction on the number of slots specified (the "length" of the cylinder set $[d_0 \ldots d_\ell]$) the claim that the pre-image of the cylinder set $[d_0 \ldots d_\ell]$ is a ball of the same measure as the cylinder set. Note that $\Phi^{-1}([d_0]) = d_0$ is a ball of the correct measure as

$$\mu_{A,\lambda}([d_0]) = \lambda(d_0) = \rho(r) = \mu(d_0)$$

by construction. Now,

$$\Phi^{-1}([d_0 \ldots d_\ell]) = \Phi^{-1}([d_0]) \cap T^{-1}\Phi^{-1}\left([d_1 \ldots d_\ell]\right).$$

By the inductive hypothesis, this is the intersection of two balls, and is thus again a ball. Noting that $\Phi^{-1}([d_1 \ldots d_\ell]) \subseteq \Phi^{-1}[d_1] = d_1$ and applying claim (i) of Lemma 5.4, along with the inductive hypothesis, we see that this ball has the correct measure

$$\begin{aligned}
\mu\left(\Phi^{-1}([d_0 \ldots d_\ell])\right) &= \sum_{j \in H} \mu\left(j \cap \Phi^{-1}([d_1 \ldots d_\ell])\right) A(d_0, j) \\
&= A(d_0, d_1)\mu\left(\Phi^{-1}([d_1 \ldots d_\ell])\right) \\
&= A(d_0, d_1)\mu([d_1 \ldots d_\ell]) = \mu([d_0 \ldots d_\ell]).
\end{aligned}$$

As the cylinder sets form a base for the topology on X_A, this shows that Φ is continuous. As the cylinder sets are a sufficient semi-ring in the Borel σ-algebra of X_A, this shows that Φ is measure-preserving. So, Σ' is indeed a topological and measurable factor of Σ. ☐

Theorem 5.9. *Let $X, T, H, \Sigma, \Sigma', \Phi$ be as in Proposition 5.8. Moreover, assume that Σ is measure-preserving. Then, Σ' is measure-preserving.*

Now, let $H = \bigsqcup_{k=1}^{n} S_k$ be a decomposition of H in the sense of Proposition 2.1, so that

$$\Sigma' = \bigsqcup_{k=1}^{n} \Sigma'_k$$

where $\Sigma'_k = (X_{A_k}, \mu_{A_k,\lambda}, T_{A_k})$ with X_{A_k} regarded as a subset of X_A, and $\mu_{A_k,\lambda}, T_{A_k}$ then the restrictions of $\mu_{A,\lambda}, T_A$ to X_{A_k}.

For $k = 1, \dots, n$ define

$$\widetilde{\Sigma}_k = \begin{cases} \textit{restriction of } \Sigma \textit{ to } \Phi^{-1}(X_{A_k}) & \#X_{A_k} < \infty \\ \Sigma'_k & \textit{otherwise} \end{cases}$$

then we have an isomorphism of topological and measurable dynamical systems

$$\Sigma \cong \bigsqcup_{k=1}^{n} \widetilde{\Sigma}_k.$$

Moreover, each term in this decomposition is either locally an isometry or ergodic Markov, according as whether $\#X_{A_k} < \infty$ or not.

Proof. By Lemma 5.4, T measure-preserving on X implies that the columns of A sum to 1. This implies that $\lambda A = \lambda$. By the observations of Section 2, we thus have that Σ' is measure-preserving.

Note that Φ continuous and measure-preserving implies Φ surjective: X is compact and X_A Hausdorff, so the image must be closed; but the image must have full measure and so must be dense (λ positive implies that all cylinder sets, hence all open sets, have strictly positive measure).

For $k = 1, \dots, n$, denote $C_k = \Phi^{-1}(X_{A_k})$, $\mu_k = \mu|_{C_k}$, $T_k = T|_{C_k}$. The decomposition of Σ' induces the following decomposition of Σ:

$$\Sigma = \bigsqcup_{k=1}^{n} (C_k, \mu_k, T_n).$$

To complete the proof of the proposition, it suffices to show that $(C_k, \mu_k, T_k) \cong \widetilde{\Sigma}_k$ for $k = 1, \dots, n$ as topological and measurable dynamical systems, and to classify them as being locally isometries and ergodic Markov in the two cases. We now handle the two cases separately:

Case 1: $\#X_{A_k} < \infty$

If $\#X_{A_k} < \infty$, then the isomorphism $(C_k, \mu_k, T_k) \cong \widetilde{\Sigma}_k$ follows by definition. Note that the measure on Σ'_k is necessarily atomic; as it is ergodic, it must in fact be the inverse orbit of a single atom. As T_A, hence T_{A_k}, is measure-preserving, each of the atoms must have equal measure. It follows that each element $x \in X_{A_k}$ is of the form

$$x = (d_0, d_1, \dots, d_\ell, d_0, \dots, d_\ell, d_0, \dots, d_\ell, \dots),$$

with $A(d_0, d_1) = A(d_1, d_2) = \dots = A(d_\ell, d_0) = 1$. Then, $\Phi^{-1}(x) = d_0$, where $C(d_0) = 1$ (here, C is that from the definition of locally scaling).

So, C_k must be a collection of r-balls with $C(x) = 1$ for $x \in C_k$. Then, for $x, y \in C_k$ with $|x - y| \le r$ we have $|T(x) - T(y)| = C(x)|x - y| = |x - y|$. This shows that $\widetilde{\Sigma}_k$ is locally an isometry, as desired.

Case 2: $\#X_{A_k} = \infty$

If $\#X_{A_k} = \infty$, then we claim that Φ induces an isomorphism $(C_k, \mu_k, T_k) \cong \widetilde{\Sigma}_k$. In a measure-preserving Markov shift, any atoms must have finite inverse orbit; so Σ'_k ergodic and $\#X_{A_k} = \infty$ implies that μ_{A_k, λ_k} is non-atomic. We noted above that Φ is surjective. We claim that it is also injective. For $x \in X$ let $d_n = \pi_H T^n(x)$ for $n = 0, 1, \ldots$. Then,

$$\Phi^{-1}(\Phi(x)) = \bigcap_{\ell \geq 0} \Phi^{-1}\left([d_0 \ldots d_\ell]\right).$$

We have from Proposition 5.8 that each of these pre-images is a ball. Then, $\Phi^{-1}(\Phi(x))$ is the intersection of a nested family of balls. If the intersection contains more than a single point, then the radii of the balls do not go to 0, and so the intersection has non-empty interior and thus positive measure. Now, the measure on X_{A_k} is non-atomic, so $\mu_{A_k,\lambda_k}(\Phi(x)) = 0$. As Φ is measure-preserving, this implies that $\mu(\Phi^{-1}(\Phi(x))) = 0$; by the above considerations this implies that $\Phi^{-1}(\Phi(x))$ contains at most one point. So, Φ is injective.

Then, Φ is a continuous, measure-preserving, bijection. Observe that Φ takes closed sets to closed sets by compactness, so its inverse is also continuous. This also implies that Φ^{-1} is measurable, and then Φ measure-preserving implies Φ^{-1} measure-preserving. So, Φ is an isomorphism of topological and measurable dynamic systems $(C_k, \mu_k, T_k) \cong \widetilde{\Sigma}_k$ as desired. As the later is ergodic Markov, the former is as well. □

Corollary 5.10. *Let $X \subseteq K$ be compact-open and $T : X \to X$ a measure-preserving locally scaling transformation. If T is ergodic then it is either Markov or locally an isometry. In particular, if it is weakly mixing then it also Markov and so mixing. So, for a measure-preserving locally scaling transformation on a compact-open X, weakly mixing implies mixing.*

Proof. If T is ergodic, then the decomposition in Theorem 5.9 must be trivial. So, T must be either Markov or locally an isometry. If it is locally an isometry, then it cannot be weakly mixing. So, weakly mixing implies weakly mixing Markov which in turn implies mixing. □

Corollary 5.11. *For a locally scaling transformation, the following properties depend only on the associated transition matrix:*

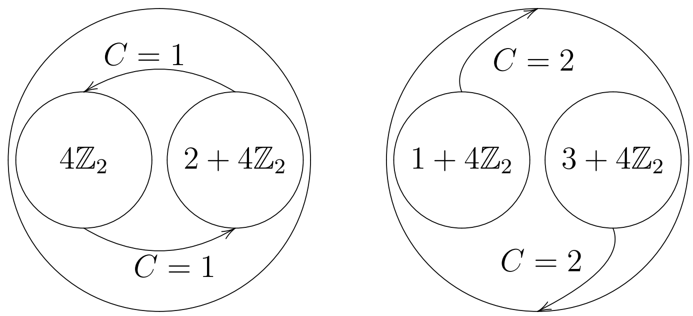
(i) *Measure-preserving;*
(ii) *Weakly mixing, mixing, exact, Bernoulli.*

Proof. By Lemma 5.4, the property of being measure-preserving depends only on the associated transition matrix.

Note that the decomposition in Theorem 5.9 depends only on the associated transition matrix. Given an associated transition matrix, we have the following cases:

(i) The decomposition is trivial, and the system is a local isometry. Then, it is not weakly mixing (or any of the stronger properties listed).
(ii) The decomposition is trivial, and the system is ergodic Markov. In this case, the system is determined up to isomorphism by the matrix.
(iii) The decomposition is not trivial. In this case, the system is not ergodic and cannot satisfy any of the stronger properties listed. □

Example 5.12. Consider a measure-preserving locally scaling map $T : \mathbb{Z}_2 \to \mathbb{Z}_2$ that acts as follows on the balls



where the arrows show what the image of a given ball is, and the labels give the value of the scaling function C.

In this example, we want the $4\mathbb{Z}_2$ ball and the $2+4\mathbb{Z}_2$ ball to be surjectively mapped onto one another, while the $1 + 4\mathbb{Z}_2$ and $3 + 4\mathbb{Z}_2$ balls are surjectively mapped onto the $1 + 2\mathbb{Z}_2$ ball (which has twice the radius).

One map that will accomplish this is the following:

$$T(n) = \begin{cases} n+2 & n \equiv 0 \pmod 4 \\ n-2 & n \equiv 2 \pmod 4 \\ 1 + 2\lfloor \frac{n}{4} \rfloor & n \equiv 1 \pmod 2 \end{cases}$$

where $\lfloor \cdot \rfloor : \mathbb{Q}_2 \to \mathbb{Z}_2$ is defined on 2-adic expansions by

$$\left\lfloor \sum_{k \geq -\ell} a_k 2^k \right\rfloor = \sum_{k \geq 0} a_k 2^k, \text{ where } a_k \in \{0, 1\}.$$

The reader may note that this is the unique continuous extension of the floor function $\lfloor \cdot \rfloor : \mathbb{Q} \to \mathbb{Z}$.

In order to motivate this definition of T, and make evident that it does behave as indicated in the diagram, let us note its action on 2-adic expansions. Letting $a_k \in \{0, 1\}$ for each $k \geq 0$ we have

$$T\left(a_0 + 2a_1 + 4a_2 + 8a_3 + \cdots + 2^k a_k + \cdots\right) =$$
$$= \begin{cases} 2(1-a_1) + 4a_2 + 8a_3 + \cdots + 2^k a_k + \cdots & \text{if } a_0 = 0 \\ 1 + 2a_2 + 4a_3 + 8a_4 + \cdots + 2^k a_{k+1} + \cdots & \text{if } a_0 = 1 \end{cases}$$

We see immediately that the restrictions $T|_{4\mathbb{Z}_2}$ and $T|_{2+4\mathbb{Z}_2}$ are surjective onto $2 + 4\mathbb{Z}_2$ and $4\mathbb{Z}_2$ respectively, and preserve distances (leaving the first 2-adic digit which differs in the same position). We also observe that $T|_{1+4\mathbb{Z}_2}$ and $T|_{3+4\mathbb{Z}_2}$ are surjective onto $1 + 2\mathbb{Z}_2$, and multiplies distances by 2 (shifting the first 2-adic digit which differs one position to the left).

Now, the fact that T is locally scaling for $r = 1/4$ is clear. We may construct the associated transition matrix by using the diagram and Lemma 5.4:

$$A = (A(i,j))_{0 \leq i,j < 4} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

where $i \in \{0,1,2,3\}$ is identified with $i + 4\mathbb{Z}_2$.

Applying Lemma 5.4 we note that T is measure-preserving. Indeed, we can see this immediately from the diagram: the pre-image under T of a ball contained in $2\mathbb{Z}_2$ is a ball of the same radius, while the pre-image of a ball contained in $1 + 2\mathbb{Z}_2$ is two balls of half the radius. Applying Theorem 5.9, we see that $T|_{2\mathbb{Z}_2}$ is locally isometric, as we already observed above, while $T|_{1+2\mathbb{Z}_2}$ is ergodic Markov (and indeed isomorphic to the Bernoulli shift on two symbols).

6. Polynomial approximation in $\mathcal{O}$

The above results dealt with $\mathcal{C}^1$ functions, extending to polynomial maps as a special case. In the next sections we will be interested in finding polynomial maps with specified associated transition matrices. In preparation for this, we will need some results on the approximation of continuous maps $\mathcal{O} \to K$. For the reader's convenience, we will sketch here the definitions and results of [Ami64], slightly simplified for our applications.

Say $X \subseteq \mathcal{O}$ is compact-open. Moreover, assume that X is a finite union of r balls for $r \in \mathcal{V}$. Then, for $r' \leq r$ each r'-ball contained in X is a union of precisely q balls of radius $|\pi|r'$ contained in X. In the terminology of [Ami64], this makes X a *regular valued compact* (*compact valué régulier* in the original French).

For $k \geq v(r)$, we may define $H_k = X/R_{|\pi|^k}(0)$, and a projection map $\pi_k : X \to H_k$. Then, we say that a sequence $\{u_k \in X : k \in \mathbb{N}\}$ is *very well distributed* (*très bien répartie*) if for each $k \geq v(r)$, $h \in H_k$, and $m \geq 1$ we have

$$\#\{i < m\#H_k : u_i \in h\} = m.$$

That is, the terms of the sequence must be equally distributed among the possible values mod $\mathfrak{p}^k$ for $k \geq v(r)$. Note that the condition that the $\{u_k\}$ are very well distributed implies that they are distinct.

Now, given such a sequence $\{u_0, u_1, \ldots\}$, we may define the *corresponding interpolating polynomials* for $k \geq 0$:

$$P_k(x) = (x-u_0)(x-u_1)\cdots(x-u_{k-1}) \qquad \text{and} \qquad Q_k(x) = \frac{P_k(x)}{P_k(u_k)}.$$

Then, we may summarize some of the results of [Ami64, §II.6.2] as follows:

Theorem 6.1 (Amice). *Let $X \subseteq \mathcal{O}$ be compact-open, and let $\{u_k\}$ be a very well distributed sequence with values in X with P_k, Q_k the corresponding interpolating polynomials. Let $f : X \to K$ be continuous, and for $k \geq 0$ set*

$$a_k = P_k(u_k)\left(\sum_{j=0}^{k} \frac{f(u_j)}{P'_{k+1}(u_j)}\right).$$

Then:

(i) $|a_k| \to 0$ *as* $k \to \infty$*;*
(ii) $\sum_{k\geq 0} a_k Q_k(x) \to f(x)$ *uniformly on* X*;*
(iii) *The decomposition in (ii) is unique;*
(iv) $\sup_{x\in K} |f(x)| = \sup_{k\in\mathbb{N}} |a_k|$.

A very well distributed sequence $\{u_k\}$ is said to be *well ordered* (*bien ordonnée*) if $|u_n - u_m| = |\pi|^{v_q(n-m)}$ for all $n, m \geq 0$ where $v_q(n-m)$ is the exact power of q dividing $n-m$. Following our sources, we will call such a sequence *very well distributed* (*très bien répartie bien ordonnée*). This allows us to state results of Helsmoortel and Barsky, characterizing Lipschitz and $\mathcal{C}^1$ functions on $\mathcal{O}$ in terms of the coefficients in their expansions. This result may be found in [Bar73].

Theorem 6.2 (Helsmoortel, Barsky)**.** *Let $\{u_k\}$ be a very well distributed sequence with values in $\mathcal{O}$, with P_k, Q_k the corresponding interpolating polynomials. Let $f : \mathcal{O} \to K$ be continuous with*

$$f(x) = \sum_{k \geq 0} a_k Q_k(x)$$

the expansion of f in the sense of Theorem 6.1. For $k \geq 0$, define

$$\kappa_k = |\pi|^{-\lfloor \log_q k \rfloor}.$$

Then:

(i) *f is r-Lipschitz if and only if $r \leq \kappa_k |a_k|$ for all $k \geq 0$;*
(ii) *$f \in \mathcal{C}^1(\mathcal{O})$ if and only if $\kappa_k |a_k| \to 0$ as $k \to \infty$.*

Example 6.3. Note that $\{0, 1, 2, \ldots\} \subseteq \mathbb{Z}_p$ satisfies the conditions of a well distributed sequence, and is in fact trivially very well distributed. Then,

$$Q_k(x) = \frac{x(x-1)\cdots(x-k+1)}{k \cdot (k-1) \cdot \cdots 1} = \binom{x}{k}.$$

So, in this case the above reduces to the Mahler expansion.

More generally: Let $a_0, \ldots, a_{q-1}$ be a complete set of coset representatives for $\mathcal{O}/\mathfrak{p}$. For $k \in \mathbb{N}$, we will define u_k in terms of the base-q expansion of k:

$$k = \sum_{i=0}^{\ell} k_i q^i \longmapsto \sum_{i=0}^{\ell} a_{k_i} \pi^i = u_k.$$

Then, say we have $n, m \in \mathbb{N}$ with $n = \sum_{i \geq 0} n_i q^i$ and $m = \sum_{i \geq 0} m_i q^i$. Let $\ell = v_q(i-j) = \min\{i : n_i \neq m_i\}$. Then,

$$|u_n - u_m| = |\pi|^{\ell} = |\pi|^{v_q(n-m)},$$

and $\{u_k\}$ is very well distributed. In particular, this implies that there is always a very well distributed sequence for $\mathcal{O}$, and corresponding interpolating polynomials such that the results cited in this section hold.

7. Polynomial maps on $\mathcal{O}$ realizing locally scaling transformations

The above characterizes measure-preserving polynomial transformations on $\mathcal{O}$ in terms of locally scaling transformations. It does not show the existence of *any* measure-preserving polynomial transformations. We will show that in fact the polynomials, in a sense, provide a representative class among the measure-preserving locally scaling maps.

For $S \subseteq K$ we say that $T : S \to K$ is *affine* if it is given by $x \mapsto ax + b$ for some constants $a, b \in K$. We say that $T : \mathcal{O} \to \mathcal{O}$ is *locally affine* if for each $x \in \mathcal{O}$ there exists a $r \in \mathcal{V}$ such that $T|_{B_r(x)}$ is affine.

Lemma 7.1. *Let $\{u_k\}$ be a very well distributed sequence in $\mathcal{O}$ with corresponding interpolating polynomials P_k, Q_k Then, for $n \in \mathbb{N}$:*

(i) $Q_k(\mathcal{O}) \subseteq \mathcal{O}$;
(ii) Q_k *is* κ_k *Lipschitz, with* κ_k *as in Theorem 6.2;*
(iii) *For* $k = q^\ell$ *for some* $\ell \geq 1$, *and*

$$|Q_k(x) - Q_k(y)| = \kappa_k |x-y| \text{ for all } x, y \in \mathcal{O} \text{ with } |x-y| \leq 1/\kappa_k.$$

Proof. Claim (i) follows by Theorem 6.1(iv). Claim (ii) follows by Theorem 6.2.

Assume $k = q^\ell$. Then, note that $\kappa_k = |\pi|^{-\ell}$. So, our claim is equivalent to the statement that $v(Q_k(x) - Q_k(y)) = v(x-y) - \ell$ if $v(x-y) \geq \ell$. Now, we may observe that

$$\begin{aligned} Q_k(x) - Q_k(y) &= \frac{(x-u_{k-1})\cdots(x-u_0) - (y-u_{k-1})\cdots(y-u_0)}{(u_k-u_{k-1})\cdots(u_k-u_0)} \\ &= \sum_{j=1}^{k} (x-y)^j \Xi_j \\ &\quad \text{where } \Xi_j = \sum_{0 \leq i_1 < i_2 < \cdots < i_j < k} \frac{(y-u_0)\cdots\widehat{(y-u_{i_1})}\cdots\widehat{(y-u_{i_j})}\cdots(y-u_{k-1})}{(u_k-u_{k-1})\cdots(u_k-u_0)}. \end{aligned}$$

We will prove the following two statements, which together with the strong triangle inequality imply our desired result:

(i) $v(\Xi_1) = -\ell$;
(ii) $v(\Xi_j) > -j\ell$ for $1 < j \leq k$.

That this suffices is clear, for the $j = 1$ term will dominate in valuation.

As the $\{u_k\}$ are very well distributed, they are dense in $\mathcal{O}$. So, it suffices to prove our claim for $y \in \{u_k\}$. Indeed, say $y = u_n$. Then, for any $1 \leq j \leq k$ and $0 \leq i_1 < i_2 < \cdots < i_j < k$, that $\{u_k\}$ is very well distributed implies that

$$\begin{aligned} &v\left(\frac{(y-u_0)\cdots\widehat{(y-u_{i_1})}\cdots\widehat{(y-u_{i_j})}\cdots(y-u_{k-1})}{(u_k-u_{k-1})\cdots(u_k-u_0)}\right) \\ &\qquad = v_q\left(\frac{(n-0)\cdots\widehat{(n-i_1)}\cdots\widehat{(n-i_j)}\cdots(n-k-1)}{(k-k-1)\cdots(k-0)}\right) \\ &\qquad = v_q\left(\frac{n\cdots\widehat{(n-i_1)}\cdots\widehat{(n-i_j)}\cdots(n-k+1)}{k!}\right) \end{aligned}$$

As $\{n-k+1, n-k+2, \ldots, n-1, n\}$ is a collection of precisely $k = q^\ell$ consecutive natural numbers, precisely one element of this set will have maximal valuation — that is, the unique element divisible by q^ℓ. Say this is $n - i$ for some $i \in \{0, \ldots, k-1\}$. Then, the equality

$$v_q\left(n(n-1)\cdots(n-k+1)\right) - v_q(k!) = v_q(n-i) - \ell$$

holds, as there are $\lfloor k/q^t \rfloor$ multiples of q^t in $\{n, \ldots, n-k+1\}$ (and also in $\{k, \ldots, 1\}$) for each $1 \leq t \leq \ell$ and one multiple of q^t in $\{n, \ldots, n-k+1\}$ (and none in $\{k, \ldots, 1\}$) for

$\ell < t \leq v_q(n-i)$. Then, for $j \geq 1$

$$\begin{aligned} v_q\left(\frac{n\cdots\widehat{(n-i_1)}\cdots\widehat{(n-i_j)}\cdots(n-k+1)}{k!}\right) &= v_q\left(n(n-1)\cdots(n-k+1)\right) - v_q(k!) \\ &\geq (v_q(n-i)-\ell) - (v_q(n-i) + (j-1)(\ell-1)) \\ &= -j\ell + (j-1), \end{aligned}$$

as removing $n-i$ lowers the valuation by precisely $v_q(n-i)$, while removing any other term lowers it by at most $\ell - 1$. Moreover, we note that for $j = 1$ precisely one term in the summation defining Ξ_j attains this bound, so we in fact have $v(\Xi_1) = -\ell$. This proves our claim. $\square$

Theorem 7.2. *Let $r \in \mathcal{V}$ and $H = \mathcal{O}/B_r(0)$. Let A be a stochastic matrix on H. Then, let*

$$\mathcal{T}_A = \{T \text{ locally scaling for } r : A \text{ is the associated transition matrix for } T\}.$$

If $\mathcal{T}_A$ is non-empty then:

(i) *$\mathcal{T}_A$ contains a locally affine transformation;*
(ii) *$\mathcal{T}_A$ contains infinitely many polynomials.*

Proof.
(i):
Say $T \in \mathcal{T}_A$. Let $C : H \to \mathbb{R}_{\geq 1}$ be the scaling function in the definition of locally scaling and observe that its image is contained in $\mathcal{V}$. So, for each $h \in H$ we may let $S(h) \in K$ be such that $|S(h)| = C(h)$. Let $M(h)$ be any point in $T(h) = \{T(x) : x \in h\}$. We may regard S, M as maps with domain $\mathcal{O}$. Define $T_{S,M}(x) = S(x)x + M(x)$.

We claim that $T_{S,M} \in \mathcal{T}_A$. First note that $T_{S,M}$ is locally scaling for r as S, M are constant on elements of H and $|S(x)| \geq 1$ for all x. Also, note that C is also the scaling function for $T_{S,M}$. So, it suffices to verify that for $i, j \in H$ we have $i \cap T^{-1}(j) \neq \emptyset \Leftrightarrow i \cap T_{S,M}^{-1}(j) \neq \emptyset$. But indeed,

$$T_{S,M}(i) = \{x : |x - M(i)| \leq rC(i)\} = T(i).$$

(ii):
Let $T \in \mathcal{T}_A$. By (i), we may assume that T is locally affine and hence strictly differentiable. Let $\{u_k\}$ be a very well distributed sequence in $\mathcal{O}$ (which must exist by the remark at the end of the preceding section) with corresponding interpolating polynomials P_k, Q_k. Let

$$T = \sum_{k \geq 0} a_k Q_k.$$

be the decomposition of T in the sense of Theorem 6.1

Take $\alpha \in \mathcal{V}$ with $\alpha < 1$. By Theorem 6.2, $\kappa_k |a_k| \to 0$ as $k \to \infty$, so there exists an $N \in \mathbb{N}$ such that for $k > N$ we have $k|a_k| \leq \alpha < 1$. Moreover, take N such that $N > \#H$.

Let

$$f(x) = \sum_{k=0}^{N} a_k Q_k(x).$$

Note that f is a polynomial. As $T(\mathcal{O}) \subseteq \mathcal{O}$, Theorem 6.1(iv) gives us that $|a_k| \leq 1$ for all k; applying Theorem 6.1(iv) again we obtain that $f(\mathcal{O}) \subseteq \mathcal{O}$. Denote

$$R(x) = T(x) - f(x) = \sum_{k>N} a_k Q_k(x).$$

Also by Lemma 7.1, we have $|R(x) - R(y)| \leq \alpha|x - y|$ for all $x, y \in \mathcal{O}$. Note that for $|x - y| \leq r$ we then have $\alpha|x - y| < C(x)|x - y| = |T(x) - T(y)|$. So, by the strong triangle inequality, for $|x - y| \leq r$ we have

$$|f(x) - f(y)| = |T(x) - T(y) + R(y) - R(x)| = C(x)|x - y|.$$

So, f is locally scaling for $r \in \mathcal{V}$, with C its scaling function. So, as before it suffices to verify that $i \cap T^{-1}(j) = \emptyset \Leftrightarrow i \cap f^{-1}(j) = \emptyset$. But, for $x \in \{u_0, \dots, u_{\#H-1}\}$ we have that $T(x) - f(x) = 0$ as $0 = Q_k(u_0) = \cdots = Q_k(u_{\#H-1})$ for $k > N > \#_H$. As the $\{u_k\}$ are very well distributed, $\{u_0, \dots, u_{\#H-1}\}$ gives a complete set of coset representatives for H. Then, for $i \in H$ we have $\ell \in \{0, \dots, \#H - 1\}$ such that $u_\ell \in i$ and so

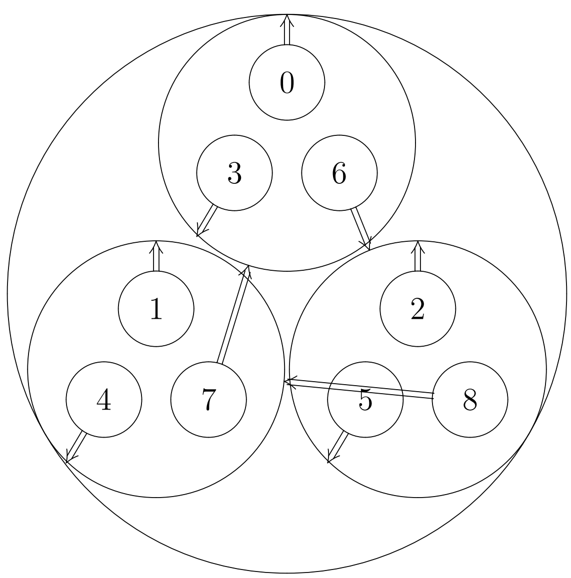
$$T(i) = T(B_r(u_\ell)) = B_{rC(i)}(T(\ell)) = B_{rC(i)}(f(\ell)) = f(B_r(u_\ell)) = f(i).$$

Finally,

$$\begin{aligned} i \cap T^{-1}(j) = \emptyset &\Leftrightarrow T(i) \cap j = \emptyset \\ &\Leftrightarrow f(i) \cap j = \emptyset \\ &\Leftrightarrow i \cap f^{-1}(j) = \emptyset. \end{aligned}$$ $\square$

In particular, Theorem 7.2 shows the existence of measure-preserving mixing transformations on the p-adics given by polynomial maps. In the following example we show how the construction in the Proposition yields such maps.

Example 7.3. Consider a locally scaling map $T : \mathbb{Z}_3 \longrightarrow \mathbb{Z}_3$ whose action on the balls of radius $1/9$ is as indicated in the following diagram:



For instance, the $0 + 9\mathbb{Z}_3$ ball maps surjectively onto the $3\mathbb{Z}_3$ ball while the $6 + 9\mathbb{Z}_3$ ball maps surjectively onto the $2 + 3\mathbb{Z}_3$ ball, and in each case T scales distances by a factor of 3.

An example of a map which has the desired properties is

$$T(x) = 3\left\lfloor \frac{x}{9} \right\rfloor + (x \bmod 3) - g(x)$$

where

$$g(x) = \begin{cases} 1 & x \equiv 6, 7, \text{ or } 8 \pmod 9 \\ 0 & \text{otherwise} \end{cases}$$

The $x \bmod 3 - g(x)$ terms ensures that $T(x)$ has the proper remainder modulo 3, while the $3\lfloor x/9 \rfloor$ term ensures surjectivity, just as in Example 5.12.

We see that T is locally scaling for $r = 1/9$, hence continuous. Then, it admits a Mahler expansion as

$$T(x) = \sum_{k=0}^{\infty} a_k \binom{x}{k}$$

where $a_k = \Delta^k T(0)$ and $\Delta f(x) = f(x+1) - f(x)$. Note that T is locally affine, so $\kappa_k |a_k| \to 0$. We now proceed to compute the Mahler expansion for T. We get the following:

$\boldsymbol{k}$	$\boldsymbol{a_k}$	$\boldsymbol{\lvert a_k\rvert}$	$\boldsymbol{\kappa_k \lvert a_k\rvert}$
0	0	0	0
1	1	1	1
2	0	0	0
3	-3	1/3	1
4	9	1/9	1/3
5	-18	1/9	1/3
6	26	1	3
7	-21	1/3	1
8	-21	1/3	1
9	141	1/3	3
10	-405	1/81	1/9
11	918	1/27	1/3
12	-1851	1/3	3
13	3501	1/9	1

Further computation suggests that $\kappa_k |a_k| \le 1$ for $k > 12$. Consider the truncated Mahler series

$$f(x) = \sum_{k=0}^{12} a_k \binom{x}{k}.$$

If we in fact knew that $\kappa_k |a_k| \le 1$ for $k > 12$, then applying Theorem 6.2 would yield that $T - f$ is 1-Lipshitz, and we could use the strong-triangle inequality to get that f is locally scaling for $r = 1/9$ for the same function C. Without proving that this inequality holds, we may still use the methods of Example 4.6 to prove this. Indeed, we have that $81f \in \mathbb{Z}_p[x]$, and computing $81f'(x) \pmod 8 1$ we may show that $|f'(x)| = 3$ for all $x \in \mathbb{Z}_p$, computing mod27 we may show that $|f''(x)/2| = 9$ and computing mod9 that $|f'''(x)/3!| \le 27$ for all $x \in \mathbb{Z}_p$; this will imply our desired claim.

Once we know that f is locally scaling for $r = 1/9$ for the same function C, noting that $T(x) = f(x)$ for $x \in \{0, \dots, 8\}$, we see that the associated transition matrices for T and f must in fact coincide. Computing the associated transition matrix from the diagram, we may observe that it is primitive. So, the transformation defined by f is ergodic Markov by Theorem 5.9, and mixing by Proposition 2.1. Note also, that unlike the maps of Section 8, it is not manifestly Bernoulli.

8. Polynomial Bernoulli maps on $\mathcal{O}$

The construction of the preceding section gives infinite classes of measure-preserving polynomials with different kinds of measurable dynamics. Among these maps are Markov mixing maps. We will now study the class of such polynomials whose associated transition matrix has all entries equal, in which case the Markov transformation is in fact Bernoulli. The main upshot of this study is a class of explicitly given and relatively simple measure-preserving Bernoulli polynomial maps.

Definition 8.1. We say that a measure-preserving locally scaling map $T : \mathcal{O} \to \mathcal{O}$ is *isometrically Bernoulli* for $r \in \mathcal{V}$ if it is locally scaling for $r \in \mathcal{V}$ and all entries of the associated transition matrix are equal.

Let $V = \mathcal{O}/\mathfrak{p}^k \cong \mathbb{F}_q^k$ and define

$$B_V = \left(\prod_{i \geq 0} V, \mu_V, T_V \right)$$

where μ_V is the product probability measure, and T_V the left-shift. We may let d'_V be the quotient metric on V. Then, we may define a metric d_V on $\prod_{i \geq 0} V$ by

$$d_V\left((a_0, a_1, a_2, \ldots), (b_0, b_1, b_2, \ldots)\right) = |\pi|^{-k(\ell-1)} d'_V(a_\ell, b_\ell) \text{ where } \ell = \min\{i : a_i \neq b_i\}.$$

We give two justifications for this metric:

(i) View elements of V as k-tuples under the isomorphism $\mathbb{F}_q^k \cong V$ corresponding to π-adic expansion (i.e., the isomorphism induced by the map shown in (ii)). Then, expanding each element in the product to a k-tuple, d_V is just the dictionary metric (with base $|\pi|$).

(ii) For each $a \in V$ we may let $\overline{a} \in \mathcal{O}$ be a coset representative for the quotient. Then, the map

$$(a_0, a_1, \ldots) \longmapsto \sum_{i \geq 0} \overline{a_i} \pi^{ki}$$

gives a bijection $\prod_{i \geq 0} V \to \mathcal{O}$. This metric is the unique metric making this map an isometry.

Now, the term isometrically Bernoulli is partially motivated by the following:

Lemma 8.2. *Let $T : \mathcal{O} \to \mathcal{O}$ be a transformation. Then, the following are equivalent:*

(i) *T is isometrically Bernoulli for $r = |\pi|^k$;*

(ii) *T is $|\pi|^k$-Lipschitz and*

$$|T(x) - T(y)| = |\pi|^{-k}|x - y| \text{ for all } x, y \in \mathcal{O} \text{ satisfying } |x - y| \leq |\pi|^k.$$

(Observe that the Lipschitz condition is implied by the second stated condition.)

(iii) *Let $V = \mathcal{O}/\mathfrak{p}^k$. There exists an invertible isometry $\Phi : \mathcal{O} \to \prod_{i \geq 0} V$ such that $\Phi \circ T = T_V \circ \Phi$; that is, $(\mathcal{O}, \mu, T)$ is metrically isomorphic to B_V.*

Proof. **(i)⇒(ii):**
Let $H = \mathcal{O}/B_r(0)$, and $A : H^2 \to \mathbb{R}_{\geq 0}$ the associated transition matrix. Note that if T is isometrically Bernoulli, then each entry of A must be equal, and hence must be equal to $\frac{1}{\#H} = \rho(r)$. Now, T must be locally scaling for $r \in \mathcal{V}$, so $|T(x) - T(y)| = C(x)|x - y|$ for $|x - y| \leq r = |\pi|^k$. But, we must have $\rho(1/C(x)) = \rho(r)$, so $C(x) = 1/r = |\pi|^{-k}$. So, the second half of (ii) follows. But, as the image of T lies in $\mathcal{O}$, the Lipschitz condition is vacuous for $|x - y| > |\pi|^k$.

(ii)⇒(iii):
Let $V = \mathcal{O}/\mathfrak{p}^k$. Now, (ii) implies that T is locally scaling for r. Letting A be the associated transition matrix, we readily note that all non-zero entries of A must be equal to $|\pi|^k$; as A is a stochastic matrix, this implies that all entries of A are non-zero.

Now, let $\Sigma' = (X_A, \mu_{A,\lambda}, T_A)$ be as in Theorem 5.9. We see that $B_V = \Sigma'$. We observed above that all entries of A are non-zero; then, A is irreducible and Theorem 5.9 gives us a topological and measurable isomorphism $\Phi : \mathcal{O} \to X_A$. Note that the balls of X_A with respect to d_V are just the cylinder sets. Moreover, one may check that for each $\ell \geq 0$, X_A is a disjoint union of q^ℓ balls of radius $r = |\pi|^\ell$, which must then each have measure $q^{-\ell} = \rho(r)$. Then, Proposition 5.8 implies that Φ^{-1} takes balls of a given radius to balls of the same radius; moreover, Φ^{-1} must take each of the q^ℓ distinct balls of radius $|\pi|^\ell$ in X_A to a distinct ball of radius $|\pi|^\ell$ in $\mathcal{O}$. So each ball of radius $|\pi|^\ell$ in $\mathcal{O}$ must be the pre-image of precisely one ball of the same radius in X_A. It follows that Φ and Φ^{-1} are both isometries.

(iii)⇒(i):
Note that for $x, y \in \mathcal{O}$ we have

$$|T(x) - T(y)| = |\Phi(T(x)) - \Phi(T(y))| = |T_V(\Phi(x)) - T_V(\Phi(y))|.$$

Then, for $|\Phi(x) - \Phi(y)| = |x - y| \leq |\pi|^k$ we compute

$$|T_V(\Phi(x)) - T_V(\Phi(y))| = |\pi|^{-k}|\Phi(x) - \Phi(y)| = |\pi|^{-k}|x - y|. \qquad \square$$

Now, we may combine Lemma 7.1 with Lemma 8.2 to get:

Corollary 8.3. *Let $\{u_k\}$ be a very well distributed sequence with values in $\mathcal{O}$, with corresponding interpolating polynomials P_k, Q_k. Say $T : \mathcal{O} \to \mathcal{O}$ is given by the expansion, in the sense of Theorem 6.1,*

$$T(x) = \sum_{k \geq 0} a_k Q_k(x), \text{ with } a_k \in \mathcal{O},\ |a_k| \to 0.$$

Assume that

(i) *$M = \max_{k \geq 0} \kappa_k |a_k|$;*
(ii) *There is a unique $k_M \geq 0$ attaining this maximum, and moreover it is of the form $k_M = q^\ell$ for some $\ell \geq 0$;*
(iii) *$|a_{k_M}| = 1$ (hence, $M = \kappa_{k_M}$).*

Then, T is isometrically Bernoulli for $r = 1/M \in \mathcal{V}$.

Proof. As k_M is the unique value attaining the maximum, we have by the strong triangle inequality along with Lemma 7.1 imply that

$$|T(x) - T(y)| = \left| \sum_{k \geq 0} a_k \left[Q_k(x) - Q_k(y) \right] \right| = \kappa_{k_M} |a_{k_M}| |x - y| = M|x - y|$$

for $|x - y| \leq 1/\kappa_{k_M} = 1/M$. Then, our claim follows by Lemma 8.2. $\square$

Example 8.4. Let $K = \mathbb{Q}_p$ and $\mathcal{O} = \mathbb{Z}_p$. Then, $\{0, 1, 2, \ldots\}$ is a very well distributed sequence, and letting P_k, Q_k be the corresponding interpolating polynomials we can check that $Q_k(x) = \binom{x}{k}$ (c.f. Example 6.3). In this case, $\kappa_k = p^{\lfloor \log_p k \rfloor}$. Therefore, we can rewrite the sufficient conditions in Corollary 8.3 as follows. Given $T : \mathbb{Z}_p \longrightarrow \mathbb{Z}_p$ defined by

$$T(x) = \sum_{k \geq 0} a_k \binom{x}{k}$$

assume that

(i) $M = \max_{k \geq 0} |a_k| p^{\lfloor \log_p k \rfloor}$;
(ii) There is a unique $k_M \geq 0$ attaining this maximum, and moreover it is of the form $k_M = p^\ell$ for some $\ell \geq 0$;
(iii) $|a_{k_M}| = 1$ (thus, $M = p^\ell$).

Then T is isometrically Bernoulli for $r = p^{-\ell}$. In particular, the polynomials $\binom{x}{p^\ell}$ for $\ell > 0$ clearly satisfy these conditions, and so defines a Bernoulli transformation on $\mathbb{Z}_p$.

Note, in particular, that this criterion applies to any map $\mathbb{Z}_p \to \mathbb{Z}_p$ defined by $u\binom{x}{p} + F(x)$ with $u \in \mathbb{Z}_p^\times$ and $F \in \mathbb{Z}_p[x]$. An example of such a map is that given by $\frac{x^p - x}{p}$ from [WS98].

In this context, the polynomials $\binom{x}{p}$ and $\frac{x^p - x}{p}$ are in a sense the most natural isometrically Bernoulli maps:

Example 8.5. Take a set of coset representatives for $\mathbb{Z}_p/p\mathbb{Z}_p$. Then, using Example 6.3 we may form a very well distributed sequence from these, and then the p^{th} corresponding interpolating polynomial (and unit multiples of it) will be Bernoulli by Corollary 8.3.

Let's look at the two most common sets of coset representatives for the quotient $\mathbb{Z}_p/p\mathbb{Z}_p$:

(i) Take as coset representatives $0, 1, 2, \ldots, p - 1$. The resulting very well distributed sequence is $\{0, 1, \ldots\}$. Then, $P_p(x) = x(x-1)\cdots(x-p+1)$ and $Q_p(x) = \binom{x}{p}$ is the p^{th} corresponding interpolating polynomial.
(ii) Take as coset representatives 0 and the $(p-1)^{\text{st}}$ roots of unity (there are exactly $p-1$ by Hensel's Lemma); these are called the "Teichmuller representatives." Then, $P_p(x) = x^p - x$ and

$$Q_p(x) = \frac{P_p(x)}{P_p(p)} = \frac{1}{p^{p-1} - 1} \frac{x^p - x}{p}.$$

So, the polynomials $\binom{x}{p}$ and $\frac{x^p - x}{p}$ (up to unit) are analogs, arising by the same construction from the two most natural choices for the coset representatives of $\mathbb{Z}_p/p\mathbb{Z}_p$.

Example 8.6. Let $K = \mathbb{F}_q((t))$ and $\mathcal{O} = \mathbb{F}_q[[t]]$. We may construct a very well distributed sequence as in Example 6.3, having $0, 1, 2, \dots, q-1, t$ as its first $q+1$ terms. Then

$$Q_q(x) = \frac{x(x-1)\dots(x-q+1)}{t(t-1)(t-2)\dots(t-q+1)}$$

and that it is isometrically Bernoulli by Corollary 8.3. However, $t-1$, $t-2, \dots, t-q+1$ are all units in $\mathcal{O}$, thus

$$t(t-1)\dots(t-q+1)Q_q(x) = \frac{x(x-1)\dots(x-q+1)}{t}$$

defines a Bernoulli transformation as well.

Now, we will give two examples of isometrically Bernoulli polynomial maps on the rings of integers of finite extensions of $\mathbb{Q}_p$. First, we briefly motivate our choice of examples. For K a finite extension of $\mathbb{Q}_p$, let $n = [K : \mathbb{Q}_p]$,$f = [\mathcal{O}/\mathfrak{p} : \mathbb{F}_p]$, and $e = \log_{|\pi|} |p|$. It is a standard result that $ef = n$. It is evident that the nature of how $\mathcal{O}$ compares to $\mathbb{Z}_p$ depends on the values of e and f. The two extreme cases are $f = 1, e = n$ (in which case we say that the extension is *totally ramified*) and $e = 1, f = n$ (in which case we say that the extension is *unramified*). We give an example from each of these two extremes. For more background on the relevant theory, including the "standard" results invoked in this paragraph and in the following two examples see [Ser62], particularly Ch. I §7, 8., Ch. III §5, Ch. IV §4.

Example 8.7. Take $p > 2$ and let

$$K = \mathbb{Q}_p(\zeta_p) \text{ where } \zeta_p \text{ is a primitive } p^{\text{th}} \text{ root of unity.}$$

It is a standard result that $1-\zeta_p$ may be taken as a uniformizing parameter and that the extension is totally ramified and so the set $\{0, \dots, p-1\}$ gives a complete set of coset representatives for $\mathcal{O}/\mathfrak{p}$. We may construct a very well distributed sequence as in Example 6.3. The first p terms would be just $0, \dots, p-1$, with the next term $1-\zeta_p$. The first p^2 terms would be $\{i + j(1-\zeta_p)\}$ for $0 \le i, j < p$, with the next term $(1-\zeta_p)^2$. Noting that $q = p$ and applying Corollary 8.3 shows that the transformations defined by the polynomials

$$\frac{x(x-1)\dots(x-p+1)}{1-\zeta_p}$$

and

$$\frac{1}{(1-\zeta_p)^3} \prod_{0 \le i,j < p} (x - i - j(1-\zeta_p))$$

are isometrically Bernoulli for $r = |\pi|^1 = |p|^{\frac{1}{p-1}}$ and $r = |\pi|^2 = |p|^{\frac{2}{p-1}}$, respectively.

Example 8.8. Take $f > 1$ and let

$$K = \mathbb{Q}_p(\zeta) \text{ where } \zeta \text{ is a primitive } (p^f - 1)^{\text{th}} \text{ root of unity.}$$

It is a standard result that K is the unique unramified extension of degree f. So, p may be taken as a uniformizing parameter. Let $\overline{\zeta} \in \mathcal{O}/\mathfrak{p}$ be the image of ζ under the quotient map. We note that $\overline{\zeta}$ must generate the residue field extension, i.e. $\mathcal{O}/\mathfrak{p} = \mathbb{F}_p(\overline{\zeta}) = \mathbb{F}_p[\overline{\zeta}]$. So, $S = \{a_0 + a_1\zeta + \dots + a_{f-1}\zeta^{f-1}\}$, with $0 \le a_0, a_1, \dots, a_{f-1} < p$, is a complete set of coset representatives for $\mathcal{O}/\mathfrak{p}$. Applying the construction of Example 6.3 we may construct a very well distributed sequence whose first $q = p^f$ terms are precisely the elements of S, with the following term being p. Then, applying Corollary 8.3 shows that the transformation defined by the polynomial

$$\frac{1}{p}\prod_{0\le a_0,a_1,\ldots,a_{f-1}<p}(x-a_0-a_1\zeta-\cdots-a_{f-1}\zeta^{f-1})$$

is isometrically Bernoulli for $r=|p|$.

9. Polynomial almost Bernoulli maps on $\mathbb{Z}_p$

An important condition shared by isometrically Bernoulli polynomials maps is their derivatives must have constant valuation. We may use this observation to come up with a class of interesting non-examples. Our non-examples will be polynomial maps whose derivatives have constant valuation, but such that the maps need not be measure-preserving.

Proposition 9.1. *The map $f:\mathbb{Z}_p\to\mathbb{Z}_p$ given by*

$$f(x)=\binom{x}{n}$$

for $n\in\mathbb{N}$ satisfies $|f'(x)|=C$ for some $C\in\mathcal{V}$ and for all $x\in\mathbb{Z}_p$ if and only if $n=ap^\ell$, with $1\le a<p$ and $\ell\in\mathbb{Z}_{\ge0}$, and

$$\frac{1}{u}+\ldots+\frac{1}{u+a-1}\not\equiv 0 \pmod{p}$$

for each $u\in\{1,2,\ldots,p-a\}$.

Proof. We note that each n may be written uniquely as $n=ap^\ell$ with $p\nmid a$. We will first show that it is necessary that $a<p$.

Observe that

$$f'(x)=\frac{1}{n!}\sum_{i=0}^{n-1}x(x-1)\cdots\widehat{(x-i)}\cdots(x-n+1).$$

We note that $|f'(0)|=\left|\frac{(n-1)!}{n!}\right|=\left|\frac{1}{n}\right|=p^\ell$. So, in order for $|f'(x)|$ to be constant, it must be equal to this value for all $x\in\mathbb{Z}_p$. There exists a $b\in\{0,1,\ldots,p-1\}$ such that $bp^{\lfloor\log_p n\rfloor}\le n<(b+1)p^{\lfloor\log_p n\rfloor}$; then, we may take $x\in\mathbb{N}$ such that such that $\{x,x-1,\ldots,x-n+1\}$ contains precisely b numbers divisible by $p^{\lfloor\log_p n\rfloor}$ with exactly one of these divisible by $p^{1+\lfloor\log_p n\rfloor}$. Then, $|f'(x)|=p^{\lfloor\log_p n\rfloor}$. This in turn implies that $\ell=\lfloor\log_p n\rfloor$, and so $a<p$.

Henceforth, we assume $n=ap^\ell$, $1\le a<p$. As $\mathbb{N}$ is dense in $\mathbb{Z}_p$, it suffices to check our condition for $x\in\mathbb{N}$. Then, we have two cases:
Case I: One of $\{x,x-1,\ldots,x-n+1\}$ is divisible by $p^{\ell+1}$:
As $a<p$, this implies that exactly one element in the set is divisible by $p^{\ell+1}$. Then, the corresponding term in the summation giving f' will dominate in valuation, yielding $|f'(x)|=p^\ell$.
Case II: None of $\{x,x-1,\ldots,x-n+1\}$ is divisible by $p^{\ell+1}$:
Then say $\{p^\ell u,p^\ell(u+1),\ldots,p^\ell(u+a-1)\}$ is the subset of $\{x,\ldots,x-n+1\}$ consisting of

the terms divisible by p^ℓ; note that $u, \ldots, u+a-1$ are units, that is $u, \ldots, u+a-1 \not\equiv 0 \pmod{p}$. Then

$$|f'(x)| = \left|\frac{x(x-1)\cdots(x-n+1)}{n!}\right| \left|\sum_{i=0}^{n-1} \frac{1}{x-i}\right| = \left|\sum_{i=0}^{n-1} \frac{1}{x-i}\right|$$
$$\leq \max\{p^\ell \left|\sum_{j=0}^{a-1} \frac{1}{u-j}\right|, p^{\ell-1}\},$$

where the final inequality is an equality if the first expression in the max is strictly greater than the second.

In order that $|f'(x)|$ be constant, we must have $|f'(x)| = p^\ell$ for all $x \in \mathbb{Z}_p$. It follows that a necessary and sufficient condition for this is that

$$\left|\sum_{j=0}^{a-1} \frac{1}{u-j}\right| \geq 1$$

whenever $u, \ldots, u-a-1 \in \mathbb{Z}_p^\times$. The condition in the statement of the proposition is merely a restatement of this, so our result follows. □

Corollary 9.2. *Let $p > 3$. Then,*

$$f(x) = \binom{x}{(p-2)p^\ell}$$

satisfies $|f'(x)| = p^\ell$ for all $x \in \mathbb{Z}_p$.

Proof. The map $u \mapsto \frac{1}{u}$ is a group automorphism on the cyclic group $\mathbb{F}_p^\times$. Viewing the elements of $\mathbb{F}_p$ as being $\{0, 1, 2, \ldots, p-1\}$ with operations performed modulo p, it follows that

$$\sum_{u \in \{1,2,\ldots,p-1\}} \frac{1}{u} = \sum_{u \in \{1,2,\ldots,p-1\}} u \equiv 0 \pmod{p}.$$

So, for $u \in \{1, 2\}$ we have

$$\frac{1}{u} + \ldots + \frac{1}{u+(p-2)-1} \equiv -\frac{1}{u} \not\equiv 0 \pmod{p}.$$

It follows that $a = p-2$ satisfies the conditions of Proposition 9.1. Writing $n = (p-2)p^\ell$ and applying the proposition yields our desired result. □

Example 9.3. Let $p = 5$, and define a transformation $f : \mathbb{Z}_5 \to \mathbb{Z}_5$ by

$$f(x) = \binom{x}{15}.$$

By the Corollary, we have that $|f'(x)| = 5$ for all $x \in \mathbb{Z}_5$. Note that $125f(x) \in \mathbb{Z}_5[x]$; then, a computation modulo 5 yields that $|f''(x)| \leq 25$, and we have the trivial bound $|f^{(\ell)}(x)| \leq 125$ for $\ell \geq 3$. Then, by a computation as in Example 4.6 we may conclude that f is locally scaling for $r = 1/25$.

Now, the study of the dynamics of f reduces to looking at its associated transition matrix. We note that x mod 25 determines $f(x)$ mod 5, and that the collection of this datum suffices to determine the associated transition matrix. A computation yields that

$$\begin{aligned} x &\equiv 0, 1, 2, \ldots, 14 \pmod 25 \Rightarrow f(x) \equiv 0 \pmod 5 \\ x &\equiv 15, 16, \ldots, 19 \pmod 25 \Rightarrow f(x) \equiv 1 \pmod 5 \\ x &\equiv 20, 21, \ldots, 24 \pmod 25 \Rightarrow f(x) \equiv 4 \pmod 5 \end{aligned}$$

Identifying $i \in \{0, 1, \ldots, 24\}$ with $i + 25\mathbb{Z}_5$, we may write the associated transition matrix as

$$A = (A(i,j))_{0 \le i,j < 25} = \begin{cases} 1/5 & 0 \le i \le 14 \text{ and } j \equiv 0 \pmod 5 \\ 1/5 & 15 \le i \le 19 \text{ and } j \equiv 1 \pmod 5 \\ 1/5 & 20 \le i \le 24 \text{ and } j \equiv 4 \pmod 5 \\ 0 & \text{otherwise} \end{cases}$$

and we may observe that A has a left eigenvector of eigenvalue 1

$$\lambda = (\lambda(i))_{0 \le i < 25} = \begin{cases} 3/5 & i \equiv 0 \pmod 5 \\ 1/5 & i \equiv 1, 4 \pmod 5 \\ 0 & \text{otherwise} \end{cases}$$

This eigenvector is not positive, so we cannot use the construction of Proposition 2.1; however, $\mu_{A,\lambda}$ assigns zero measure to $[2], [3]$, so we may (up to measurable isomorphism) disregard those symbols. The resulting symbolic system on $0, 1, 4$ has an irreducible, indeed primitive, matrix.

Let μ denote Haar measure on $\mathbb{Z}_5$. Then, we may define a f-invariant measure $\widetilde{\mu}$ on $\mathbb{Z}_5$:

$$\widetilde{\mu}(A) = 3\mu\left(A \cap B_{1/5}(0)\right) + \mu\left(A \cap B_{1/5}(1)\right) + \mu\left(A \cap B_{1/5}(4)\right)$$

for any μ-measurable set $A \subseteq \mathbb{Z}_5$. Moreover, the map Φ of Proposition 5.8 gives a measurable isomorphism of $(\mathbb{Z}_5, \widetilde{\mu}, f)$ with $(X_A, \mu_{A,\lambda}, T_A)$, where the latter dynamical system is mixing Markov; so, in a sense $\widetilde{\mu}$ is the f-invariant measure corresponding to the eigenvector λ of A (or more precisely, to $\mu_{A,\lambda}$).

10. Bernoulli maps on $\widehat{\mathbb{Z}}$

Let us define

$$\widehat{\mathbb{Z}} = \varprojlim_{n,|} \mathbb{Z}/n\mathbb{Z} \cong \prod_p \mathbb{Z}_p.$$

Using the result of Section 8, we may construct various maps $\mathbb{N} \to \mathbb{Z}$ which extend to Bernoulli maps on $\mathbb{Z}_p$ for each p, and hence to a Bernoulli map on $\widehat{\mathbb{Z}}$. Explicitly:

Proposition 10.1. *Let $f : \mathbb{N} \to \mathbb{Z}$ be defined by*

$$f(n) = \sum_{k=0}^{n} a_k \binom{n}{k}$$

with $a_k \in \mathbb{Z}$ for $k \ge 0$ satisfying the following for each prime rational prime p:

(i) $|a_k| = 1$ *for $k = p$;*
(ii) $|a_k| < p^{-\lfloor \log_p k \rfloor}$ *for $k > p$.*

Then, f extends to an isometrically Bernoulli transformation $f : \mathbb{Z}_p \to \mathbb{Z}_p$ for each prime p.

Proof. For each prime p, note that the quantity $p^{\lfloor \log_p k \rfloor} |a_k|_p$ attains its maximum for $k = p$. Then, the result is immediate by Corollary 8.3. $\square$

Example 10.2. Let $f : \mathbb{N} \to \mathbb{Z}$ be defined by

$$f(n) = \sum_{p \le n} \prod_{p' < p} {p'}^{1 + \lfloor \log_{p'} p \rfloor} \binom{n}{p}$$

where the summation is over primes bounded by n, and the product over primes bounded by p.

In the notation of the Proposition, we have

$$a_p = \prod_{p' < p} {p'}^{1 + \lfloor \log_{p'} p \rfloor},$$

and $a_k = 0$ for k not a prime. So, for each prime p it is the case that $|a_p|_p = 1$. Moreover, for $k > p$ we see that $|a_k|_p \le p^{-1-\lfloor \log_p k \rfloor} < p^{-\lfloor \log_p k \rfloor}$. So, the conditions of the Proposition are satisfied, and f extends to an isometrically Bernoulli transformation $\mathbb{Z}_p \to \mathbb{Z}_p$ for each prime p.

Example 10.3. Let $f : \mathbb{N} \to \mathbb{Z}$ be defined by

$$f(n) = \sum_{k \le n} (k-1)!^k \binom{n}{k}.$$

That is, we set $a_k = (k-1)!^k$. Then, $|a_p|_p = |(p-1)!^p|_p = 1$. And, for $k > p$ we see that certainly $|a_k|_p = |(k-1)!^{k-1}|_p < p^{-\lfloor \log_p k \rfloor}$. So, the conditions of the Proposition are again satisfied.

11. Rational functions

We have thus far been primarily concerned with polynomial maps. However, the tools we develop suffice to say a great deal about maps given by rational functions.

For $X \subseteq K$ compact-open, we say that $T : X \to X$ is *given by a rational function on X* if there exist $f, g \in K[x]$ with g non-vanishing on X such that

$$T(x) = \frac{f(x)}{g(x)} \text{ for all } x \in X.$$

Note that the condition that g is non-vanishing on X is not too strict:

Lemma 11.1. *Let $X \subseteq K$ be compact-open and non-empty. Say $f, g \in K[x]$ with g not identically 0, and say $T : X \to X$ is such that*

$$T(x) = \frac{f(x)}{g(x)} \textit{ for all } x \in X \textit{ satisfying } g(x) \neq 0.$$

Then, there exist $\widetilde{f}, \widetilde{g} \in K[x]$, $\widetilde{g}$ non-vanishing on X, such that $T(x) = \frac{\widetilde{f}(x)}{\widetilde{g}(x)}$ for all $x \in X$ satisfying $g(x) \neq 0$, and in particular off of a set of μ measure zero.

Proof. It suffices to show that for $a \in X$, $x - a$ has greater multiplicity in the numerator than the denominator. But indeed, as X is compact, it is bounded. So, this result follows at once upon noting that for $\ell \geq 1$ the expression

$$\left| \frac{1}{(x-a)^\ell} \right|$$

is unbounded as $x \to a$. ◻

Lemma 11.2. *Let $X \subseteq K$ be compact-open. Say $T : X \to X$ is given by a rational function on X, as $T = \frac{f}{g}$, with $f, g \in K[x]$ and g non-vanishing on X. Then, $T \in \mathcal{C}^1(X)$. In particular, the results of Section 5 apply to transformations of compact-open sets that are given by rational functions.*

Proof. Say K is a union of r-balls. Then, for $|z| \leq z$ and $x \in X$ we have $x + z \in K$. In particular, the non-vanishing of $g(x)$ implies the non-vanishing of $g(x+z)$. Then, observe that

$$\begin{aligned}
\frac{T(x+z) - T(x)}{z} &= \frac{f(x+z)g(x) - f(x)g(x+z)}{g(x)g(x+z)z} \\
&= \left(\frac{1}{g(x)g(x+z)} \right) \left(g(x)\frac{f(x+z) - f(x)}{z} - f(x)\frac{g(x+z) - g(x)}{z} \right).
\end{aligned}$$

So, our result follows from the proof of Lemma 4.4. ◻

As with polynomials, it remains to ask whether there are actually rational functions with any interesting dynamical behavior. In fact, we may readily extend the results of Section 7.

Lemma 11.3. *Let $f \in K[x]$ be such that $f(\mathcal{O}) \subseteq \mathcal{O}$. Assume moreover that $f : \mathcal{O} \to \mathcal{O}$ is locally scaling for $r \in \mathcal{V}$. Then, there exists $g \in K[x]$ non-vanishing and non-constant on $\mathcal{O}$ such that $T = \frac{f}{g}$ is a locally scaling transformation $\mathcal{O} \to \mathcal{O}$ with the same associated transition matrix as f.*

Proof. Say

$$f = \sum_{k=0}^{n} a_k x^k,$$

is locally scaling for $r \in \mathcal{V}$ and let $C : \mathcal{O} \to \mathbb{R}_{\geq 0}$ be its scaling function.

Take N such that

$$N > \max\{\deg f, \sup_{k=0}^{n} \log_\pi r|a_k|\}.$$

Now, set $g(x) = 1 - \pi^N x^\ell$ for any $\ell \geq 1$, and let $T = \frac{f}{g}$. For $k \in \mathbb{N}$, let $\overline{k} \in \{0, \ldots, N-1\}$ be the reduction of $k \pmod N$, and define

$$b_k = \begin{cases} |\pi|^{\lfloor k/N \rfloor} a_{\overline{k}} & \overline{k} \in \{0, \ldots, n\} \\ 0 & \text{otherwise} \end{cases}.$$

Then,

$$T(x) = \sum_{\ell \geq 0} \sum_{k=0}^{n} a_k x^{k + \ell N} = \sum_{k \geq 0} b_k x^k.$$

So,

$$T(x) - f(x) = \sum_{k \geq N} b_k x^k \in \mathfrak{p}^{1+\log_{|\pi|} r} \mathcal{O}[[x]].$$

So, $|T(x) - f(x)| \leq r$ and $T - f$ is $|\pi|$-Lipschitz. Then for $|x - y| \leq r$, and hence $|f(x) - f(y)| = C(x)|x - y| \geq 1$, we have

$$|T(x) - T(y)| = \left| f(x) - f(y) + \frac{f(x)}{g(x)} - f(x) - \frac{f(y)}{g(y)} + f(y) \right| = |f(x) - f(y)| = C(x)|x - y|.$$

Moreover, it is the case that

$$T(B_r(a)) = B_{rC(a)}(T(a)) = B_{rC(a)}(f(a)) = f(B_r(a)).$$

Our desired result follows. $\square$

References

[Ami64] Yvette Amice, *Interpolation p-adique*, Bull. Soc. Math. France **92** (1964), 117–180.

[Ana02] V. S. Anashin, *Uniformly distributed sequences of p-adic integers*, Diskret. Mat. **14** (2002), no. 4, 3–64.

[Bar73] Daniel Barsky, *Fonctions k-lipschitziennes sur un anneau local et polynômes à valeurs entières*, Bull. Soc. Math. France **101** (1973), 397–411.

[Ben01] Robert L. Benedetto, *Hyperbolic maps in p-adic dynamics*, Ergodic Theory Dynam. Systems **21** (2001), no. 1, 1–11.

[BS05] John Bryk and Cesar E. Silva, *Measurable dynamics of simple p-adic polynomials*, Amer. Math. Monthly **112** (2005), no. 3, 212–232.

[KN04] Andrei Yu. Khrennikov and Marcus Nilson, *p-adic deterministic and random dynamics*, Mathematics and its Applications, vol. 574, Kluwer Academic Publishers, Dordrecht, 2004.

[LS94] Douglas Lind and Klaus Schmidt, *Bernoullicity of solenoidal automorphisms and global fields*, Israel J. Math. **87** (1994), no. 1-3, 33–35.

[RL03] Juan Rivera-Letelier, *Dynamique des fonctions rationnelles sur des corps locaux*, Astérisque (2003), no. 287, xv, 147–230, Geometric methods in dynamics. II.

[Rob00] Alain M. Robert, *A course in p-adic analysis*, Graduate Texts in Mathematics, vol. 198, Springer-Verlag, New York, 2000.

[Sch84] W. H. Schikhof, *Ultrametric calculus*, Cambridge Studies in Advanced Mathematics, vol. 4, Cambridge University Press, Cambridge, 1984, An introduction to p-adic analysis.

[Ser62] Jean-Pierre Serre, *Corps locaux*, Publications de l'Institut de Mathématique de l'Université de Nancago, VIII, Actualités Sci. Indust., No. 1296. Hermann, Paris, 1962.

[Wal82] Peter Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York, 1982.

[War] Thomas Ward, *"p-adic deterministic and random dynamical systems" by Andrei Khrennikov and Marcus Nilsson, Mathematics and its Applications Vol. 574, Kluwer Academic Publishers (2004)*, Bull. Amer. Math. Soc. (N.S.) (to appear).

[WS98] Christopher F. Woodcock and Nigel P. Smart, *p-adic chaos and random number generation*, Experiment. Math. **7** (1998), no. 4, 333–342.

(James Kingsbery) Department of Mathematics, Williams College, Williamstown, MA 01267, USA
E-mail address: James.C.Kingsbery@williams.edu

(Alex Levin) Harvard University, MA 02138, USA
E-mail address: levin@fas.harvard.edu

(Anatoly Preygel) Harvard University, MA 02138, USA
E-mail address: preygel@fas.harvard.edu

(Cesar Silva) Department of Mathematics, Williams College, Williamstown, MA 01267, USA
E-mail address: csilva@williams.edu