

# NONSTATIONARY MIXING AND THE UNIQUE ERGODICITY OF ADIC TRANSFORMATIONS

ALBERT M. FISHER

ABSTRACT. We define topological and measure-theoretic mixing for nonstationary dynamical systems and prove that for a nonstationary subshift of finite type, topological mixing implies the minimality of any adic transformation defined on the edge space, while if the Parry measure sequence is mixing, the adic transformation is uniquely ergodic. We also show this measure theoretic mixing is equivalent to *weak ergodicity* of the edge matrices in the sense of inhomogeneous Markov chain theory.

## CONTENTS

1. Introduction	2
2. Dynamics on nonstationary shift spaces	10
2.1. Nonstationary shift map	10
2.2. The group of finite coordinate changes	12
2.3. Ordered Bratteli diagrams	13
2.4. Adic transformations	13
2.5. Minimality and unique ergodicity	24
2.6. Topological mixing implies minimality	24
3. Mixing for nonstationary Markov chains	25
3.1. Mixing and focussing	28
4. Nonstationary Parry measures and unique ergodicity	31
4.1. Nonstationary unique ergodicity and the Bowen-Marcus lemma	34
4.2. Focussing and the Perron-Frobenius conditions	36
5. Adic transformations on edge shifts	37
5.1. Examples revisited	41
6. Appendix: G. Birkhoff's bound for the projective metric	43
6.1. Weak and strong contraction	47
6.2. The standard positive cone and the contraction formula for transposes	47
7. Appendix: Connections with the theory of inhomogeneous Markov chains	51
References	52

This research was partly supported by: the CNPQ, the Acordo Franco-Brasil, FAPESP, and the CNRS.

---

*Date:* July 17, 2009.

*Key words and phrases.* adic transformation, unique ergodicity, nonstationary subshift of finite type, projective metric, nonhomogeneous Markov chain .

## 1. INTRODUCTION

The relationship between the properties of minimality (that every orbit is dense) and unique ergodicity (having a unique invariant probability measure) has been studied for many dynamical systems, from horocycle flows to interval exchange transformations, by many authors; for some beginnings see [Fur61], [Fur73], [Kea77], [Vee78], [Vee82], [Mas82]. Here we investigate these matters for Vershik's adic transformations (see [Ver81], [Ver89], [Liv92], [Ver95b], [Ver95a]).

Adic transformations are maps of a combinatorial space which offer a far-reaching generalization of the odometer (or *adding machine*) transformation of Von Neumann and Kakutani. The flexibility and naturality of Vershik's construction is illustrated by the wide variety of dynamical systems they can be used to model, among others cutting and stacking constructions in ergodic theory (see Vershik's papers cited above, [Fer97]), substitution dynamical systems (see [Liv92], [LV92], [For97], [Hos00]), and interval exchange transformations (see [Fisa]).

For the simplest, stationary, case the dynamical space for the adic transformation  $T$  is a subshift of finite type  $\Sigma_M^+$  for some  $0-1$  ( $l \times l$ ) matrix  $M$ . The adic transformation gives a "transverse" dynamics to the shift map on  $\sigma$ , in much the way that the stable horocycle flow on the unit tangent bundle of a surface of constant negative curvature acts transversely to the geodesic flow.

For general adic transformations however the combinatorics is *nonstationary*, as one replaces the single matrix  $M$  by a sequence  $(M_i)_{i \geq 0}$  of matrices; there is no problem in defining the associated sequence space  $\Sigma_{(M)}^{0,+}$ , but one has now lost the possibility of acting by the shift map.

In this paper as in [AF05], we enlarge the space so the shift dynamics will also make sense: all we do is to note that shifting a symbol sequence moves us to a new combinatorial space, that given by the shifted combinatorics.

(*Nonstationary subshift of finite type*) Given a sequence of alphabets  $\mathcal{A}_i$  for  $i \geq 0$ , with  $\#\mathcal{A}_i = l_i$ , and  $(M_i)_{i \geq 0}$  a sequence of  $(l_i \times l_{i+1})$  non-negative integer matrices, (notation: throughout the paper, *non-negative* and *positive* both mean  $\geq 0$ , and we always indicate  $> 0$  by *strictly positive*), this determines an infinite graph (a *Bratteli diagram*, Fig. 1) with the number of edges from a vertex in  $\mathcal{A}_i$  at level  $i$  to a vertex in  $\mathcal{A}_{i+1}$  at level  $(i+1)$  given by the corresponding matrix entry. We assume the matrices are *reduced* in that no row or column is identically zero, and write  $\Sigma_{(M)}^{0,+}$  for the set of all one-sided infinite edge paths in this diagram. This is the 0<sup>th</sup> component; the further components  $\Sigma_{(M)}^{k,+}$  are defined via the same construction from the shifted matrix sequence  $(M_i)_{i \geq k}$  and the *nsft* is the disjoint union of the components:  $\Sigma_{(M)}^+ = \coprod_{k=0}^{\infty} \Sigma_{(M)}^{k,+}$ . The shift  $\sigma$  then maps an element of  $\Sigma_{(M)}^{k,+}$  to an element of  $\Sigma_{(M)}^{(k+1),+}$ :

$$\Sigma_{(M)}^{0,+} \xrightarrow{\sigma} \Sigma_{(M)}^{1,+} \xrightarrow{\sigma} \Sigma_{(M)}^{2,+} \xrightarrow{\sigma} \Sigma_{(M)}^{3,+} \cdots$$

A natural metric on  $\Sigma_{(M)}^+$  is defined by placing the *word metric* on each component (see (1) below) and then declaring the distance between points in different components to be one.

This leads to a more general concept:

(*Nonstationary dynamical system* or *mapping family*): This is a (one- or two-sided) sequence of continuous maps along a sequence of metric spaces  $(X_i, d_i)$ :

$$X_0 \xrightarrow{f_0} X_1 \xrightarrow{f_1} X_2 \xrightarrow{f_2} X_3 \cdots$$

We write this as a triple  $(X, d, f)$ . The definition initially may seem problematic, as now we have a wandering dynamics on a noncompact space (the disjoint union of the components  $X_i$ ), but in fact many of the basic notions of dynamics - stable manifolds, hyperbolicity, and Markov partitions - have a quite natural interpretation in this nonstationary setting, see [AF05].

In studying adic transformations, there are two reasons for wanting to include the shift dynamics. One is that it allows for the unification of some disparate ideas: that of “desubstitution” for a substitution dynamical system, of moving to a higher block structure or higher-level tower for a cutting and stacking construction, and of Rauzy induction for an interval exchange. The other reason is that it allows us to bring in some methods and ideas from the usual stationary case, as we shall see.

To define an adic transformation we need one additional piece of structure, a *stable order*  $\mathcal{O}$  on the Bratteli diagram at level 0 for the matrix sequence (and hence at all higher levels); if the edges are oriented towards the right in the diagram (the direction of the *future* of the *nsft*), this linearly orders the set of edges which enter each given vertex. The order  $\mathcal{O}$  then naturally extends to a further linear order (the *lexicographic* order) on each stable equivalence set  $W^s(x)$ , the set of all edge strings which agree with  $x$  beyond some coordinate, see §2.3. Since any  $y \in W^s(x)$  necessarily is in the same component as  $x$ , the map defined to send a string to its successor will act on each component separately. The *adic transformation* determined by  $\mathcal{O}$  on the  $k^{\text{th}}$  component,  $T_{\mathcal{O}} : \Sigma_{(M)}^{k,+} \setminus \mathcal{NS} \rightarrow \Sigma_{(M)}^{k,+} \setminus \mathcal{NP}$ , is the restriction of this successor map, where  $\mathcal{NS}$  is the (at most countable) collection of points with no successor and  $\mathcal{NP}$  those with no predecessor. (The fact that  $T_{\mathcal{O}}$  is not defined on the whole space is an important technical point addressed in §2 below).

In this paper we study invariant measure sequences on mapping families and give nonstationary analogues of topological mixing and measure-theoretic mixing; we are especially interested in what these dynamical properties of the shift will tell us about the transverse dynamics. And in fact, as we shall see, it tells us quite a bit: topological mixing implies minimality of the adic transformation, while measure-theoretic mixing of a natural class of measures implies unique ergodicity.

To explain what is this class of measures we return to the stationary case and to a beautiful lemma proved by Bowen and Marcus in [BM77], which served as the inspiration for this paper. That statement is given in terms of the stable equivalence relation, however for concreteness we rephrase it for adic transformations; later we give another formulation using the group  $\mathcal{FC}$  of finite coordinate changes of  $\Sigma_{(M)}^{k,+}$ , see §2.2 for the definition. We make one other change in the statement of their result, allowing edge shift spaces as well as vertex shifts, i.e. moving from 0 – 1 matrices

to general non-negative integer matrices (see e.g. [LM95] p. 36 and §5 below), and have:

**Lemma.** (*Bowen-Marcus, Lemma 2.4 of [BM77]*) *Given a one-sided subshift of finite type  $\Sigma_M^+$  for some  $(l \times l)$  non-negative integer matrix  $M$ , then if the left shift map  $\sigma$  on the edge shift space  $\Sigma_M^+$  is topologically mixing, the action of any adic transformation on  $\Sigma_M^+$  is uniquely ergodic.*

An examination of the proof shows that what actually gets used is the *measure-theoretic* mixing of the measure of maximal entropy for the *sft*, and this is implied by the topological mixing; see [Bil65] pp. 30-33, [Wal82] pp. 42, 51.

In extending this lemma to general adic transformations, we shall find that many things carry over more or less directly from the stationary case. For example, the measure of maximal entropy of a subshift of finite type has a very simple formula first discovered by Shannon in an information theory context ([SW63], p. 119), involving the left and right Perron-Frobenius eigenvectors of the matrix (the proof that this maximizes entropy within the Markov measures is due to Shannon, while Parry in [Par64] showed that Shannon's measure gives the maximum over all invariant probability measures; we thank the referee for this historical note). We follow standard usage in calling this the *Parry measure*. To extend this to the nonstationary setting, we begin with this formula rather than worrying about how to define entropy for sequences of maps; we simply replace these eigenvectors by eigenvector *sequences*, and define a nonstationary Markov chain which gives us an exact analogue of the Parry measure. This gives us a sequence of measures on the components, invariant in the sense that one is carried to the next by the shift map. And here we encounter a striking difference from the stationary case: now the natural analogue of topological mixing no longer implies measure-theoretic mixing. So actually we will need two *different* definitions, and shall need to study carefully the relationship between them.

Here are our definitions; note that for nonstationary dynamics it is critical to include the specified metric sequence  $d = (d_i)_{i \geq 0}$ , otherwise the concepts would be empty.

**Definition 1.1.** *The mapping family  $(X, d, f)$  is topologically mixing if and only if for every  $\varepsilon > 0$ , and for every  $k$ , there exists  $M > 0$  such that for all  $m > M$ , for any balls  $A, B$  in  $X_k$  and  $X_{k+m}$  of radius  $> \varepsilon$ , then*

$$A \cap f_k^{-1} \circ f_{k+1}^{-1} \circ \cdots \circ f_{k+m-1}^{-1}(B) \neq \emptyset.$$

*Let  $\mu_i$  be an invariant sequence of probability measures for the maps. Then  $(X, d, f, \mu)$  is mixing if and only if for every  $\varepsilon > 0$ , and for every  $k$ , there exists  $M > 0$  such that for all  $m > M$ , for any balls  $A, B$  in  $X_k$  and  $X_{k+m}$  of radius  $> \varepsilon$ , then*

$$\mu_k(A \cap f_k^{-1} \circ f_{k+1}^{-1} \circ \cdots \circ f_{k+m-1}^{-1}(B)) = e^{\pm \varepsilon} \mu_k(A) \mu_{k+m}(B)$$

*where we use the notation:  $x = e^{\pm \varepsilon} y$  to mean  $e^{-\varepsilon} y \leq x \leq e^{+\varepsilon} y$ .*

We note that topological mixing is preserved by a *uniform conjugacy* of mapping families (conjugacy by a uniformly equicontinuous sequence of conjugating maps, Def. 2.4 of [AF05]), and mixing by a uniform conjugacy which preserves the invariant measure sequences.

We then first show:

**Theorem 1.1.** *Given a nonstationary subshift of finite type  $\Sigma_{(M)}^+$  with the word metric, then topological mixing of the shift map implies minimality of any adic transformation defined on any component of  $\Sigma_{(M)}^+$ , and if one of the Parry measure sequences is mixing for the shift, the adic transformation is uniquely ergodic.*

Now for the case of a single matrix, topological mixing has a nice purely matrix formulation: that the matrix  $M$  be *primitive* i.e. there exists  $n > 0$  such that all entries of  $M^n$  are strictly positive, see [Wal82] p. 51 and Props. 3.4, 3.5 below. In the same spirit we search for matrix expressions of topological and measure theoretic mixing for the sequence  $(M_i)$ . As we shall see, there are at least five different reasonable ways to define primitivity for sequences, each expressing a different aspect of the primitivity of a single matrix - and not all of which are equivalent! The first definition, and the one which deserves the name since it most directly generalizes the standard definition, is:

*(Nonstationary primitivity): A sequence of  $(l_k \times l_{k+1})$  non-negative integer matrices  $(M_k)_{k \geq 0}$  is primitive if and only if for each  $k \geq 0$ , there exists  $m > 0$  (depending on  $k$ ) such that all entries of the matrix  $M_k^{k+m} \equiv M_k M_{k+1} \dots M_{k+m}$  are strictly positive.*

Indeed, as is easily checked (Lemma 2.12), primitivity of  $(M_k)_{k \geq 0}$  is equivalent to topological mixing of the nonstationary shift map. So the first statement of the theorem reduces to proving that primitivity implies minimality. In fact this was already stated by Livshits and Vershik in the fundamental paper [LV92]; a proof is given in Theorem 2.13 below.

This is however *not* enough to guarantee the measure-theoretic mixing which we need to prove unique ergodicity. Writing  $R_k^+$  for the *positive cone* of  $R_k = {}^t\mathbb{R}^{l_k}$ , the space of non-negative row vectors (see Appendix 6, Def. 6.1; the upper  $t$  denotes transpose) the matrix condition we need will be:

*(Focussing condition): The sequence  $(M_i)_{i \geq 0}$  is focussing if and only if for each  $k \geq 0$ , then for any  $\varepsilon > 0$ , we have that for  $m > k$  sufficiently large, the projective diameter of  $R_k^+ M_k \dots M_m$  in the cone  $R_{m+1}^+$  is less than  $\varepsilon$ .*

For connections with inhomogeneous Markov chain theory see the second appendix; for the definition of the projective metric see Prop. 6.4.

The way this condition will get used is that focussing for  $(M_i)_{i \geq 0}$  will imply focussing for the Parry transition matrices  $(P_i)_{i \geq 0}$  (Lemma 4.4), and this in turn will give the mixing we need to prove unique ergodicity (Lemma 3.6).

This is not the end of the story, as the focussing condition itself can be somewhat hard to check. So we introduce three other conditions, beginning with:

*(Eigenvector Perron-Frobenius condition): The non-negative integer matrix sequence  $(M_i)_{i \geq 0}$  is eigenvector Perron-Frobenius iff there is a unique normalized sequence of*

strictly positive right (column) eigenvectors for the matrix sequence, vectors  $\mathbf{v}_i$  with  $M_i \mathbf{v}_{i+1} = \lambda_i \mathbf{v}_i$  for eigenvalues  $\lambda_i \neq 0$ .

Note that this condition takes the conclusion of the Perron-Frobenius theorem as stated for *column* vectors and directly generalizes that to sequences. Now for a single matrix there is little difference in dealing with column or row eigenvectors, as there is a unique strictly positive right eigenvector if and only if there is a unique strictly positive left eigenvector (as both are equivalent to the condition that the matrix be primitive). But for sequences *this is no longer the case*: to produce a sequence of positive *left* eigenvectors, simply begin with *any* positive row vector and apply the matrix sequence!!! Thus such a sequence is certainly *never* unique. This observation led to the definition above.

A closely related condition is:

(*Topological Perron-Frobenius condition*): The matrix sequence  $(M_i)_{i \geq 0}$  is topologically Perron-Frobenius iff for each fixed  $k \geq 0$ , the image of the positive cone  $C_m^+$  in the space of column vectors  $C_m = \mathbb{R}^{l_m}$  by the matrices,  $M_k \cdots M_m C_{m+1}^+$ , nests down to a single strictly positive direction in the cone  $C_k^+ \subseteq \mathbb{R}^{l_k}$  as  $m \rightarrow +\infty$ .

This is quite clearly equivalent to the previous condition, and indeed it expresses one proof of the Perron-Frobenius theorem for a single matrix, that if we apply the powers of the matrix to the positive cone  $C^+$  of column vectors, these iterates  $M^n C^+$  nest down to a single direction: the direction of the unique strictly positive right eigenvector.

Comparing again this statement for column vectors with a similar statement for rows, in the case of a single matrix both conditions are equivalent; one simply takes the transpose, with the cones nesting down for rows as well as for columns. By contrast, for sequences the two are very different: the row image cones  $R_k^+ M_k \cdots M_m$  cannot nest down as they *are in different spaces*, in  $R_{m+1} = {}^t \mathbb{R}^{l_{m+1}}$  as  $m \rightarrow \infty$ , which may have different dimensions! And even if by chance these spaces have the same dimension, the cones have no reason to be nested, as their direction can move about wildly with changing  $m$ .

In other words for sequences there is no meaningful interpretation for row vectors of the qualitative (topological) version of Perron-Frobenius. However there is a *quantitative* statement which still does make sense, that already given in the definition of focussing.

To complete the circle of definitions, we wish to make a link between the focussing condition (for rows) and the topological Perron-Frobenius condition (for columns). For that we quantify the last condition:

(*Geometric Perron-Frobenius condition*): The sequence  $(M_i)_{i \geq 0}$  is geometrically Perron-Frobenius if and only if for each  $k \geq 0$ , then for any  $\varepsilon > 0$ , we have that for  $m > k$  sufficiently large, the projective diameter of  $M_k \cdots M_m C_{m+1}^+$  in the cone  $C_k^+$  is less than  $\varepsilon$ .

In comparing this to focussing, all that is involved is a simple switch from row vectors to columns, but there is again more to this than meets the eye. Note the order in which the matrices are applied to the cones; we have *not* simply taken the transpose, which would have reversed the order of the multiplication! And, as already remarked, a qualitative version of focussing would make no sense.

Nevertheless one can show that these two quantitative conditions are equivalent. For this the use of the projective metric is crucial; indeed the key lemma needed, given in the appendix, makes use of an explicit formula of Birkhoff for the contraction coefficient for non-negative matrices, extended to rectangular matrices in Prop. 6.13 and Cor. 6.14 (note: this is *Garrett* Birkhoff, cited as G. Birkhoff, the son of George David (G.D). Birkhoff of the ergodic theorem).

As a consequence of this equivalence we then have:

**Theorem 1.2.** *Given a sequence of  $(l_i \times l_{i+1})$  reduced and primitive non-negative integer matrices  $(M)_{i \geq 0}$ , then if this sequence is topologically Perron-Frobenius, any Parry measure sequence is mixing for the shift map on the nsft  $\Sigma_{(M)}^+$ .*

We then observe that the eigenvector Perron-Frobenius condition implies uniqueness of the eigenmeasure (Lemma 4.4); combining this with Theorems 1.1 and 1.2 gives in conclusion:

**Theorem 1.3.** *For a reduced and primitive non-negative integer matrix sequence  $(M)_{i \geq 0}$ , then the focussing condition (and equivalently the eigenvector, topological and geometric Perron-Frobenius conditions) gives a necessary and sufficient condition for unique ergodicity for the group  $\mathcal{FC}$  of finite coordinate transformations and also for any adic transformation on the combinatorial edge space defined by the matrix sequence  $(M_i)_{i \geq 0}$ .*

*Remark 1.1.* If we have focussing or geometrical Perron-Frobenius the projective diameter of the image cone goes to zero, and this holds if and only if the contraction factor for  $M_k \cdots M_m$  approaches zero (by Birkhoff's formula Theorem 6.8 and Cor. 6.9); this cumulative contraction is bounded above by the product of the factors for each matrix.

On the other hand primitivity tells us that the diameter of  $M_k \cdots M_m C_{m+1}^+$  decreases along a subsequence, but of course this may not decrease to zero, as the nested intersection may be projectively a nontrivial subsimplex of the unit simplex, analogous to having a sequence of positive real numbers  $\alpha_i$  each of which is less than one but whose infinite product does not give zero.

Any such matrix sequence which is primitive but not focussing gives an adic transformation which is minimal but not uniquely ergodic; see [FFT09] for some concrete examples. By Example 5, Keane's interval exchange of [Kea77] can be used to produce such an example. See §5.1 and [Fisa].

In [LV85] Vershik and Lodkin describe the Pascal adic transformation, so-called because its Bratteli diagram is like a Pascal triangle laid over on its side, see Example 6. This transformation is not uniquely ergodic, in fact it has an uncountable number

of invariant probability measures. Here the matrix sequence is nonprimitive, and as shown in [MP05], although the adic map is not quite minimal, it is minimal off of a countable set (in [FFT09] we call this *essential minimality*). By contrast to the Pascal adic transformation, the examples of [FFT09] have a uniformly bounded alphabet size. See [MP05] for a detailed analysis of the Pascal adic transformation and see [FP08] regarding the closely related Euler adic transformation.

### Outline of the paper.

First we treat in §§2-4 the case of  $0-1$  matrices, then in §5 we extend all the previous work to the edge shift case.

In §2 we give a general treatment of adic transformations, with special emphasis on the following points:

–The dynamics given by the lexicographic order can be interpreted geometrically, as a translation along a *stable tree*. This point of view also proves useful in [FFT09] and [Fisa], [Fisb].

–Since an adic transformation  $T_{\mathcal{O}}$  is a map from  $\Sigma_{(M)}^{k,+} \setminus \mathcal{NS}$  to  $\Sigma_{(M)}^{k,+} \setminus \mathcal{NP}$ , one has to be careful about how to define the notions of minimality and unique ergodicity. We take three different approaches: first, defining  $\mathcal{N}$  to be the collection of forward and backward iterates of points in  $\mathcal{NP}$  and  $\mathcal{NS}$  (§2.4), we consider the map restricted to the noncompact set  $\Sigma_{(M)}^{k,+} \setminus \mathcal{N}$  where it is defined for all iterations; secondly we examine possible extensions to the whole space  $\Sigma_{(M)}^{0,+}$ ; and thirdly we consider instead the action of the group  $\mathcal{FC}$  on the whole space. We describe in Prop. 2.11 how minimality and unique ergodicity for these transformations and for  $\mathcal{FC}$  are related.

–We give some examples of adic transformations, examining in particular the question of the existence of continuous extensions. We note that the Pascal adic transformation, with countably infinite  $\mathcal{NS}$  and  $\mathcal{NP}$ , has a continuous extension but no extension to a homeomorphism. For irrational rotation adic transformations we show that what happens depends on the induction procedure chosen. For Rauzy induction, the diagram is properly ordered ( $\#\mathcal{NS} = \#\mathcal{NP} = 1$ ) and there is an extension to a homeomorphism, while for a “dual” procedure, which furnishes the same Bratteli diagram with a different order,  $\#\mathcal{NS} = 2$  (though  $\#\mathcal{NP}$  still is 1) and in fact no continuous extension is possible. (The proof presented is combinatorial; a geometric explanation is given in [Fisa]).

–§3: We give a general treatment of nonstationary Markov chains and prove that focussing for the transition matrix sequence implies nonstationary mixing.

–§4: We first use eigenvector sequences to construct nonstationary Parry invariant- and eigen-measures. Then in §4.1 we prove that focussing of the  $0-1$  matrix sequence implies unique ergodicity. In §4.2 we prove the equivalence of focussing with the three forms of the Perron-Frobenius condition.

–In §5 we extend all of the previous work from vertex shifts to edge shifts, i.e from  $0-1$  matrices to non-negative integer matrices. We also extend part of the main theorem to the non-reduced case.

–In §5.1 we return to our examples, to see what we can now say about their minimality and unique ergodicity.



–In the first appendix we present material on the projective metric needed for the main part of the paper; in the second, we make connections with the theory of inhomogeneous Markov chains.

**Further remarks:** There are some differences in terminology to [LV92]; a measure with what we call the *Bowen-Marcus property* is called there a *central measure*; instead of primitivity of the matrix sequence they call this *minimality*, while we reserve that for the dynamical property which follows as a consequence. The reason for our choice of terminology is that there exist minimal adic transformations with a nonprimitive matrix sequence [FFT09].

We mention that for the stationary case Livshits and Vershik state without proof that primitivity of the matrix implies unique ergodicity; we now know two proofs of this, the one which follows from the lemma of Bowen and Marcus, and a second, quite different, proof we shall present in a later paper; this is related to the “column stochastic matrix” method used in the study in [FFT09] of the nonsymmetric  $(2 \times 2)$  case; we do not know whether an argument like that is what Livshits and Vershik had in mind.

After this paper and [FFT09] had been submitted, we received the preprint [BKMS09]; despite its title and abstract, in fact the *nonstationary* case is also addressed, see Theorem 2.8 there, and in particular the authors have independently come up with a completely different proof from that of this paper that the topological Perron-Frobenius property implies unique ergodicity. This argument can be extended to the general primitive as well as certain nonprimitive cases, including the Chacon example (see §5.1) and the infinite measure example of [Fis92], see in particular Remark 2.10 of [BKMS09] regarding the infinite measure case. Their proof is related to the column-stochastic method just mentioned, as both can be understood via a cutting-and-stacking construction of the invariant measures.

We mention that the “stable tree” is not so far from Kamae’s nice idea of *weighted substitutions*, [Kam]; this connection will be more clearly explained in [Fisb].

After this paper was completed, we found that some of what we do, specifically, the definitions of focussing and of the eigenvalue Perron-Frobenius property, had been anticipated in the field of inhomogeneous (i.e. nonstationary) Markov chains; indeed, the equivalence of these two properties was known (with a completely different proof) to Cohn and Nerman [CN90]. We survey these connections in a second appendix. As a consequence, we conclude that:

**Corollary.** *These are equivalent: a non-negative integer matrix sequence  $(M_i)_{i \geq 0}$  is weakly ergodic as defined in [CN90]; the Parry measure is unique; some Parry measure is mixing; any adic transformation defined on the components is uniquely ergodic.*

**Acknowledgements** We thank Pierre Arnoux for many inspiring conversations on related subjects, and for his comments on preliminary versions of this paper. The

notion of nonstationary subshift of finite type was introduced in [AF05], and the formula for the nonstationary Parry measure given below appears for an important special case (related to AF-algebras, i.e. to irrational circle rotations, see Example 3) in the last section of [AF01].

We dedicate this paper to the memory of a very special person whom we have recently lost, Peter Bacas; he always pointed us toward the essentials, with such a warm heart and so much fire in his eyes.

## 2. DYNAMICS ON NONSTATIONARY SHIFT SPACES

We work first with the usual *vertex shift* convention, using  $0 - 1$  matrices; in §5 we extend everything to non-negative integer matrices, i.e. to edge shifts.

Assume that for each  $i \geq 0$  we are given a finite nonempty collection  $\mathcal{A}_i$  of symbols, called the *alphabet* of index  $i$ . We write  $l_i = \#\mathcal{A}_i$ , and for simplicity take  $\mathcal{A}_i = \{0, 1, 2, \dots, (l_i - 1)\}$ . We write  $\Pi_{(\mathcal{A})}^{0,+} \equiv \prod_{i=0}^{+\infty} \mathcal{A}_i$ . Given a sequence  $(L) = (L_i)_{i \geq 0}$  of  $(l_i \times l_{i+1})$   $0 - 1$  matrices, we define  $\Sigma_{(L)}^{0,+} \subseteq \Pi_{(\mathcal{A})}^{0,+}$ , the set of *allowed strings*, to be those infinite symbolic sequences  $x = (.x_0x_1\dots)$  such that the  $(x_i x_{i+1})^{\text{th}}$  entry of  $L_i$  equals 1 for all  $i \geq 0$ .

**2.1. Nonstationary shift map.** This will be the usual left shift map  $\sigma$ , sending an individual string  $x = (.x_0x_1\dots)$  to  $\sigma(x) = (.x_1x_2\dots)$ . However unlike for the case of stationary combinatorics,  $\sigma$  is not a map on the space  $\Sigma_{(L)}^{0,+}$ , since the matrices (and indeed the number of symbols) may change with time. Indeed the image point is in a different combinatorial space, that given by the left-shifted matrix sequence  $(\sigma L) = (L_1, L_2, \dots)$ . We write  $\Sigma_{(L)}^{1,+}$  for this space, and more generally write  $\Sigma_{(L)}^{k,+} \equiv \Sigma_{(\sigma^k L)}^{0,+}$  for  $k \geq 0$ . (Note that the “decimal point” is placed to the left of the  $0^{\text{th}}$  coordinate in any component; this helps us to define the shift map.) The disjoint (i.e. indexed) union of all the shifted spaces,  $\Sigma_{(L)}^+ \equiv \coprod_{k=0}^{\infty} \Sigma_{(L)}^{k,+}$  is termed a *one-sided nonstationary shift of finite type (nsft)* [AF05]. The shift gives a sequence of maps along a sequence of spaces, a *nonstationary dynamical system*, or *mapping family*.

The nsft  $\Sigma_{(L)}^+$  is the *total space* of the mapping family;  $\Sigma_{(L)}^{k,+}$  is its  $k^{\text{th}}$  *component*. The shift map

$$\sigma : \Sigma_{(L)}^+ \rightarrow \Sigma_{(L)}^+$$

on the total space is thus equivalent to the sequence of maps along these components,

$$\Sigma_{(L)}^{0,+} \xrightarrow{\sigma} \Sigma_{(L)}^{1,+} \xrightarrow{\sigma} \Sigma_{(L)}^{2,+} \xrightarrow{\sigma} \Sigma_{(L)}^{3,+} \cdots$$

One should, however, make the following restriction on the matrices:

**Definition 2.1.**  $L_i$  is reduced iff it has no identically zero row or column.

That the sequence  $(L_i)_{i \geq 0}$  be reduced guarantees that every finite allowed string has a continuation to the right (since rows are nonzero) and to the left (since columns are). So we have:

**Lemma 2.1.** For a reduced sequence  $(L_i)_{i \geq 0}$ , for each  $k \geq 0$  the map  $\sigma : \Sigma_{(L)}^{k,+} \rightarrow \Sigma_{(L)}^{k+1,+}$  is surjective.  $\square$

The reason for this name is the following:

**Lemma 2.2.** *Given an alphabet sequence  $(\mathcal{A}_i)_{i \geq 0}$  and  $(l_i \times l_{i+1})$  0–1 matrix sequence  $(L_i)_{i \geq 0}$ , write  $\Sigma_{(L)}^{0,+}$  for the collection of all infinite allowed strings  $x = (.x_0x_1\dots)$ . If this collection is nonempty, then there is a unique alphabet sequence  $(\widehat{\mathcal{A}}_i)_{i \geq 0}$  with  $\widehat{\mathcal{A}}_i \subseteq \mathcal{A}_i$  and reduced  $(\widehat{l}_i \times \widehat{l}_{i+1})$  0–1 matrix sequence  $(\widehat{L}_i)_{i \geq 0}$  such that  $\Sigma_{(\widehat{L})}^{0,+} = \Sigma_{(L)}^{0,+}$ .*

*Proof.* The operations of reducing at stage  $k \geq 0$  by removing the identically zero rows and columns from  $L_k$  and the corresponding letters from  $\mathcal{A}_k$  and  $\mathcal{A}_{k+1}$  can be seen as operators acting on a compact space which are nonincreasing for a natural partial order coming from inclusion, with the reduced sequence being the limit. Formally, we list all the elements of the alphabets and all the matrix elements, giving a countable set  $X$ , and form the product space  $\{\mathbf{0}, \mathbf{1}\}^X$ ; removing or keeping a symbol or a matrix element will be coded by these new symbols  $\mathbf{0}, \mathbf{1}$  respectively. We define an operator  $\mathcal{R}_k$  on this space corresponding to removing all the identically zero rows from  $L_k$ , the corresponding letters from  $\mathcal{A}_k$ , and the corresponding columns from  $L_{k-1}$ . Similarly we write  $\mathcal{C}_k$  for the operator given by removing the all-zero columns from  $L_k$ , the corresponding letters from  $\mathcal{A}_{k+1}$  and corresponding rows from  $L_{k+1}$ . Given a finite list  $L_0, L_1, \dots, L_n$  we apply successively these operators  $\mathcal{R}_n, \mathcal{R}_{n-1}, \dots, \mathcal{R}_0$ , this last affecting only  $L_0$  and  $\mathcal{A}_0$ , then we apply  $\mathcal{C}_0, \dots, \mathcal{C}_n$ . This produces a new infinite sequence  $((\widehat{\mathcal{A}}), (\widehat{L}))_n$  which is reduced on times  $0, 1, \dots, n$  and which has the same infinite allowed strings  $(.x_0x_1\dots)$ . Continuing this procedure for longer and longer sequences converges by compactness, or from a different point of view, because the possibilities on any finite time segment are finite, while the operations are consistent, i.e. are nonincreasing in the partial order as  $n$  increases. Since by assumption  $\Sigma_{(L)}^{0,+}$  is nonempty,  $\widehat{\mathcal{A}}_i$  in the limit is a nonempty alphabet, so the matrices  $\widehat{L}_i$  do exist (are at least  $(1 \times 1)$ ).  $\square$

In the next sections we describe how to define stationary dynamics (i.e. given by a single map or group action) on each component of the nonstationary space. First we examine the topological, metric and Borel structure of this space.

Focussing on the 0<sup>th</sup> component, if we are given an allowed string  $x \in \Sigma_{(L)}^{0,+}$  and  $k, m \geq 0$  with  $k \leq m$ , we write  $[x_k \dots x_m] = \{w \in \Sigma_{(L)}^{0,+} : w_k = x_k, \dots, w_m = x_m\}$ , calling this a *thin cylinder set* of  $\Sigma_{(L)}^{0,+}$ , so in this notation for  $[x_k \dots x_m] \subseteq \Sigma_{(L)}^{k,+}$  then  $\sigma^{-k}([x_k \dots x_m]) = [x_k \dots x_m] \equiv [.* \dots * x_k \dots x_m] \subseteq \Sigma_{(L)}^{0,+}$ , where  $*$  indicates “no restriction on the symbols”. We denote by  $\mathbb{C}_k^m$  the collection of the thin cylinders in  $\Sigma_{(L)}^{0,+}$  with coordinates fixed from place  $k$  to  $m$ .

A *general cylinder set* is a finite union of thin cylinders, so e.g. for alphabets  $\mathcal{A}_i$  with  $\geq 3$  symbols,  $[.101] \in \mathbb{C}_0^2$ , while some general cylinders which are unions of members of  $\mathbb{C}_0^4$  are  $[.312 * 0]$  or  $[.* 2 * 3*]$ .

Providing each set  $\mathcal{A}_i$  with the discrete topology and  $\Sigma_{(L)}^{0,+}$  with the product topology, each component is a compact topological space. The cylinder sets are clopen sets which generate the topology and hence the Borel  $\sigma$ -algebra  $\mathcal{B}$  of  $\Sigma_{(L)}^{0,+}$ . One can



define a map  $\gamma : \Sigma_{(L)}^{0,+} \rightarrow \Sigma_{(L)}^{0,+}$  which interchanges the two cylinder sets  $[.x_0 \dots x_{t-2}s]$ ,  $[.y_0 \dots y_{t-2}s]$ ; thus, for  $w = (.x_0 \dots x_{t-2}sw_t w_{t+1} \dots)$ , we define

$$\gamma(w) = (.y_0 \dots y_{t-2}sw_t w_{t+1} \dots)$$

This is well-defined as  $\gamma(w)$  is an allowed string. Similarly, for  $w = (.y_0 \dots y_{t-2}sw_t w_{t+1} \dots)$ , we define  $\gamma(w) = (.x_0 \dots x_{t-2}sw_t w_{t+1} \dots)$ , while on the rest of  $\Sigma_{(L)}^{0,+}$  we take  $\gamma$  to be the identity. (So this is an involution:  $\gamma \circ \gamma = \text{id}$ ).

We call the group generated by the collection of all such maps the *group of finite coordinate changes* of  $\Sigma_{(L)}^{0,+}$ , written  $\mathcal{FC}(\Sigma_{(L)}^{0,+})$ ; see p. 532 of [Ver95b]. We note that each element of  $\mathcal{FC}$  acts on  $\Sigma_{(L)}^{0,+}$  as a homeomorphism.

**2.3. Ordered Bratteli diagrams.** Next we see how to define a single map which has the same orbits as this group; as a consequence nonatomic invariant measures for one will pass to the other (Prop. 2.11). There are many possibilities for this; adic transformations give an interesting class of such maps, determined by the choice of an order on the edges of the corresponding Bratteli diagram, as we now explain.

First define an equivalence relation on  $\Sigma_{(L)}^+$  by:

$$x \sim y \iff \exists N : \forall k \geq N, x_k = y_k.$$

We remark that the equivalence class  $\langle x \rangle$  of  $x \in \Sigma_{(L)}^{0,+}$  is the stable set  $W^s(x)$  of  $x$  for the shift map on the total space of the *nsft*; this is the set of all  $y$  such that  $d(\sigma^m(x), \sigma^m(y)) \rightarrow 0$  as  $m \rightarrow \infty$  (indeed, since this is a one-sided space, the distance equals 0 eventually). In particular, any such  $y$  also belongs to that same component  $\Sigma_{(L)}^{0,+}$  of the total space.

We next place a linear order on this countable set. This is done by first ordering the set of edges, which (when oriented towards the future) enter each given symbol; precisely, for each fixed  $k \geq 0$ , and any  $j \in \mathcal{A}_{k+1}$ , defining  $\mathcal{A}_k^j \subseteq \mathcal{A}_k$  by  $\mathcal{A}_k^j = \{i \in \mathcal{A}_k : (L_k)_{ij} = 1\}$ , and given a bijection  $\mathcal{O}_k^j : \mathcal{A}_k^j \rightarrow \{1, 2, \dots, \#\mathcal{A}_k^j\}$ , we call  $\mathcal{O}$  a *stable order* on the *nsft*. This gives an *ordered Bratteli diagram*, illustrated in Fig. 1, see e.g. [TGS95].

From this we define inductively  $\mathcal{O}$ -*lexicographic order* on  $W^s(x)$ : for  $x, y$  in the same stable set, let  $n$  be the greatest  $i$  such that  $x_i \neq y_i$ . Writing  $j = x_{n+1} = y_{n+1}$ , we then say that  $x < y$  if and only if  $\mathcal{O}_n^j(x_n) < \mathcal{O}_n^j(y_n)$ . Note that since the incoming edge sets are linearly ordered (any two edges are comparable), the same is true for  $W^s(x)$ .

*Remark 2.1.* This is more correctly a *reversed* lexicographic order, as the most significant digits or edges occur towards the right.

**2.4. Adic transformations.** We next describe Vershik's adic transformation; this will be a map on the 0<sup>th</sup> component  $\Sigma_{(L)}^{0,+}$  defined from the stable order  $\mathcal{O}$ . (A similar definition works on each component  $\Sigma_{(L)}^{k,+}$ ).

First we define the *successor* of  $x \in \Sigma_{(L)}^{0,+}$ ,  $\text{succ}(x)$ , to be the least point in  $W^s(x)$  which is greater than  $x$ , if that exists; the *predecessor*  $\text{pred}(x)$  is the greatest point which is less.

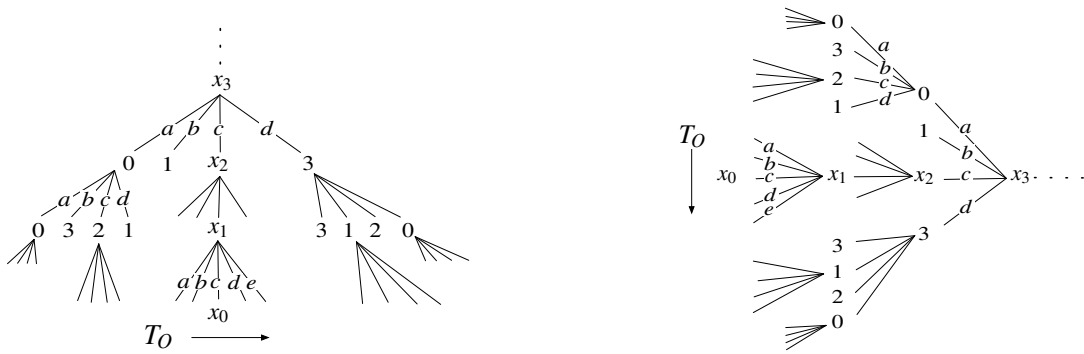


FIGURE 2. Geometric view of an adic transformation: the stable tree for the string  $(.x_0x_1\dots)$ ; the stable order is indicated by the letters. The transformation maps paths from left to right. In the second figure the tree is turned on its side, to match the Bratteli diagrams.

We write  $\mathcal{NS}$  for the set of points in  $\Sigma_{(L)}^{0,+}$  with no successor, and  $\mathcal{NP}$  for those with no predecessor. As in [LV92], p. 186, we then define  $T_{\mathcal{O}} : \Sigma_{(L)}^{0,+} \setminus \mathcal{NS} \rightarrow \Sigma_{(L)}^{0,+} \setminus \mathcal{NP}$  by  $T_{\mathcal{O}}(x) = \text{succ}(x)$ . We call this the *adic map* defined by  $(L_i)_{i \geq 0}$  and  $\mathcal{O}$ . We also define  $T_{\mathcal{O}}^{-1} : \Sigma_{(L)}^{0,+} \setminus \mathcal{NP} \rightarrow \Sigma_{(L)}^{0,+} \setminus \mathcal{NS}$  by  $T_{\mathcal{O}}^{-1}(x) = \text{pred}(x)$ , and define  $T_{\mathcal{O}}^k(x)$  for all  $k \in \mathbb{Z}$  by applying powers of  $T_{\mathcal{O}}$  or  $T_{\mathcal{O}}^{-1}$ , whenever this makes sense.

We define the *inverse order*  $\tilde{\mathcal{O}}$  by reversing the order on each edge set, so  $\tilde{\mathcal{O}}_k^j(i) = \#\mathcal{A}_k^j - \mathcal{O}_k^j(i) + 1$ . Then for the corresponding orders on  $W^s(x)$  we have  $x \tilde{<} y \iff y < x$ . Thus  $T_{\tilde{\mathcal{O}}} = T_{\mathcal{O}}^{-1}$ , and we call this the *inverse adic map*.

A useful way of understanding the dynamics of  $T_{\mathcal{O}}$  is to picture the stable set  $W^s(x)$  of a string  $x = (.x_0x_1\dots)$  as a tree, infinite toward the root, which is growing upwards. Level 0 represents the leaves of this “stable tree” of  $x$ ; branches are ordered from left to right by the stable order (the arrows of the Bratteli diagram point upwards). The dynamics of the map sends one infinite string to the next, moving along these leaves from left to right. See Fig. 2 and see also [FFT09] and [Fisb].

In the standard usage in dynamics a *transformation* is a function with the same range and domain space. Since this is not quite the case for the adic map, we set  $\mathcal{N} = (\cup_{n=1}^{+\infty} T_{\mathcal{O}}^{-n}(\mathcal{NS})) \cup (\cup_{n=1}^{+\infty} T_{\mathcal{O}}^n(\mathcal{NP}))$ ; noting that the restriction of  $T_{\mathcal{O}}$  to  $\Sigma_{(L)}^{0,+} \setminus \mathcal{N}$  is 1-1 and onto, we call  $T_{\mathcal{O}} : \Sigma_{(L)}^{0,+} \setminus \mathcal{N} \rightarrow \Sigma_{(L)}^{0,+} \setminus \mathcal{N}$  the *adic transformation* defined by  $(L_i)_{i \geq 0}$  and  $\mathcal{O}$ .

For the adic map itself (since it is not a transformation) we make the special definition that the *orbit* of a point  $x \in \Sigma_{(L)}^{0,+}$  is the stable equivalence class  $W^s(x)$ . This is, indeed, the collection of all  $T_{\mathcal{O}}^k(x)$  for  $k \in \mathbb{Z}$  for which this makes sense as stated above; precisely, we first set  $T_{\mathcal{O}}^0(x) = x$  for all  $x \in \Sigma_{(L)}^{0,+}$ , and  $T_{\mathcal{O}}^1(x) = T_{\mathcal{O}}(x)$  for  $x \notin \mathcal{NS}$ ; as above we write  $T_{\mathcal{O}}^{-1}(x) = \text{pred}(x)$  for  $x \notin \mathcal{NP}$ , and define  $T_{\mathcal{O}}^2(x)$  for  $x \notin \text{pred}(\mathcal{NS})$ , and so on. Then  $\Sigma_{(L)}^{0,+} \setminus \mathcal{N}$  is the set of points with full orbits, i.e. such that  $T_{\mathcal{O}}^k(x)$  is defined for all  $k \in \mathbb{Z}$ . Note that by definition, the adic map and the group  $\mathcal{FC}$  have the same orbits. We mention that since  $W^s(x)$  is linearly ordered,

the adic map can have no periodic orbits (though there may be finite orbits, and extended adic transformations may have periodic orbits; see Example 4 and Lemma 2.6).

Restricting the adic map to the adic transformation is a relatively insignificant change, since the set of points  $\mathcal{N}$  which has been thrown away is at most countable:

**Lemma 2.3.** *The number of points in  $\mathcal{NS}$  is at most  $\liminf_{+\infty} l(i)$ ; if this is  $+\infty$ , then the number of such points is at most countably infinite. The same is true for  $\mathcal{NP}$ . Hence  $\mathcal{N}$  is countable.  $\square$*

See [Fisb] for a proof.

Next we consider further general properties of the space  $\Sigma_{(L)}^{0,+}$ , given some natural assumptions. For  $0 \leq k < n$  integers, write  $L^{(k,n)} = L_k L_{k+1} \cdots L_{n-1}$ . Thus  $L^{(k,k+1)} = L_k$ , and  $L^{(k,n)}$  has dimensions  $(l_k \times l_n)$ .

**Definition 2.2.** *We say a sequence  $(L_i)_{i \geq 0}$  of  $0 - 1$   $(l_i \times l_{i+1})$  matrices is right nontrivial if and only if for any  $k \geq 0$  there exists  $n > 0$  (depending on  $k$ ) such that for each  $1 \leq i \leq l_k$ , the sum of the entries in the  $i^{\text{th}}$  row of  $L^{(k,n)}$  is  $\geq 2$ . (So in particular,  $l_i \geq 2$  infinitely often). It is left nontrivial if and only if the sum of the  $j^{\text{th}}$  column of  $L^{(k,n)}$  is  $\geq 2$ , for each  $1 \leq j \leq l_n$ .*

**Lemma 2.4.** *If  $(L_i)_{i \geq 0}$  is left nontrivial then every stable set  $W^s(x)$  is infinite. If  $(L_i)_{i \geq 0}$  is right nontrivial then every cylinder set is homeomorphic to a Cantor set. The same holds for  $\Sigma_{(L)}^{0,+}$ . So in particular, these sets are uncountable.*

*Proof.* We assume that  $(L_i)_{i \geq 0}$  is left nontrivial. Given a string  $x = (.x_0 x_1 \dots)$ , we picture the equivalence class  $W^s(x)$  as the stable tree, now rotating Fig. 2 so the root goes from  $x_0$  to the right, accompanying the Bratteli diagram. The central trunk is the infinite string  $(x_i)_{i \geq 0}$ , branching above or below this trunk according to the order on incoming edges. The property we want is that there is infinite branching to the left along this string. But this is a consequence of left nontriviality: we will certainly be done if we can show there is an infinite increasing subsequence  $m_i$  such that the  $j^{\text{th}}$  column of  $L^{(m_i, m_{i+1})}$  is  $\geq 2$ , for each  $1 \leq j \leq l_{m_{i+1}}$ . And left nontriviality implies this, as we build this subsequence inductively starting at time  $m_0 = 0$ , then taking  $m_1 = 0 + n$  from the definition, and so on. Finally this infinite branching to the left implies that  $W^s(x)$  cannot have both a maximum and a minimum element, which is the only way it could be finite.

For the next claim, picture a tree also with central trunk the string  $(x_i)_{i \geq 0}$ , but branching in the opposite direction, up or down chosen arbitrarily (here the stable order is irrelevant). The right nontriviality implies that branches split off from this trunk infinitely often at  $+\infty$ .

Given a cylinder set  $[\cdot x_0 \dots x_n]$ , its points are the possible strings after  $x_n$ ; with each split off of the trunk after this digit we map one half of the following branches to one half of the usual middle-third Cantor set (and handle multiple splits similarly). Since this process continues at all levels, it follows that the space of infinite paths beginning with  $x_0$  is homeomorphic to the usual Cantor set.

In particular the 0-cylinder sets  $[\cdot x_0 = a]$  for  $a \in \mathcal{A}_0$  are Cantor sets, hence this is also true for  $\Sigma_{(L)}^{0,+}$  which is the finite union over these initial digits.  $\square$

**Definition 2.3.** We say a sequence  $(L_i)_{i \geq 0}$  of  $0-1$   $(l_i \times l_{i+1})$  matrices is primitive if for any  $k \geq 0$ , there exists  $n > k$  such that  $L^{(k,n)} \equiv L_k L_{k+1} \dots L_{n-1}$  has entries all nonzero. We say the alphabet sequence  $(\mathcal{A})_{i \geq 0}$  is nontrivial if and only if  $\limsup_{+\infty} l_i \geq 2$ .

Primitivity has these consequences for the combinatorial space, by Lemma 2.4:

**Lemma 2.5.** If  $(L_i)_{i \geq 0}$  is primitive with nontrivial alphabet sequence, then it is right and left nontrivial. Hence every cylinder set is homeomorphic to a Cantor set, and  $T_{\mathcal{O}} : \Sigma_{(L)}^{0,+} \setminus \mathcal{NS} \rightarrow \Sigma_{(L)}^{0,+} \setminus \mathcal{NP}$  and  $\mathcal{FC}$  on  $\Sigma_{(L)}^{0,+}$  have no finite orbits.  $\square$

**Definition 2.4.** We say an ordered Bratteli diagram is trivial if  $\Sigma_{(L)}^{0,+} = \mathcal{N}$ .

We note that by putting together Lemmas 2.3 and 2.4, if  $(L_i)_{i \geq 0}$  is right nontrivial then the diagram is nontrivial.

Unlike the adic transformation, the group  $\mathcal{FC}$  acts on the whole space  $\Sigma_{(L)}^{0,+}$ . We now examine a second way to get dynamics on all of this space, extending the adic transformation  $T_{\mathcal{O}}$  to a map on all of  $\Sigma_{(L)}^{0,+}$ . For this we choose some function  $f : \mathcal{NS} \rightarrow \mathcal{NP}$ , and define  $T_{\mathcal{O},f} = T_{\mathcal{O}}$  on  $\Sigma_{(L)}^{0,+} \setminus \mathcal{NS}$ ,  $T_{\mathcal{O},f} = f$  on  $\mathcal{NS}$ . We call any such  $T_{\mathcal{O},f}$  an *extended* adic transformation.

We note that from Lemma 2.4 we have:

**Lemma 2.6.** If  $(L_i)_{i \geq 0}$  is left nontrivial then every extended adic transformation  $(\Sigma_{(L)}^{0,+}, T_{\mathcal{O},f})$  has no periodic orbits.  $\square$

For a trivial diagram, although there is no dynamics for the adic transformation  $T_{\mathcal{O}}$ , the extended maps can be still be interesting; see Example 4.

*Remark 2.2.* Defining an extension of the map is most natural in two cases: when one can extend by continuity (i.e. when there exists a unique continuous extension), and when  $\mathcal{NS}$  and  $\mathcal{NP}$  are both singletons so there is a single choice for  $f$ . In the last case the ordered diagram is then known in the literature as *properly ordered*, but there is no real need to make either assumption and indeed sometimes it is not natural to do so.

In fact our point of view is that in certain cases the adic transformation *should* be defined on only part of the space, and that those nice cases where one can extend continuously to all of  $\Sigma_{(L)}^{0,+}$  are interesting but fortunate exceptions. An example where there is a continuous extension which is, however, not a homeomorphism is the Pascal adic transformation (Example 6). Cases where no continuous extension is possible come from irrational circle rotations, see Example 3; this depends very much on the induction (i.e. renormalization) procedure chosen. What we show is that for the ordering on the Bratteli diagram given by Rauzy induction the diagram is properly ordered, with the unique possible extension giving a homeomorphism; however for a “dual” order on the same diagram there are now two maximal elements (but still a single minimal element), and no continuous extension is possible. The notion of duality corresponds to the vertical and horizontal foliations of an abelian differential on the torus; these results extend to more general surfaces, see Example 5 and for full



details [Fisa], [Fisb]. There is a geometrical explanation: ordered *two-sided* diagrams code the foliations, see also [AF05], and the lack of continuity of the extension is related to foliations with singular leaves which are half-lines and which meet at a common endpoint.

For this reason in studying the general measure theory we emphasize especially the group  $\mathcal{FC}$  and the transformation  $T_{\mathcal{O}}$  on  $\Sigma_{(L)}^{0,+} \setminus \mathcal{NS}$ , considering extensions mostly in the examples.

*Example 1.* (stationary and nonstationary odometers) (p. 382 of [Ver89]) These are the most basic examples of adic transformations. First, given alphabets  $\mathcal{A}_i$  with  $\#\mathcal{A}_i = l_i$ , let  $L_{ij} = 1$  for all  $i, j$ , For this very special case we write  $\Pi_{(A)}^{k,+} \equiv \Sigma_{(L)}^{0,+} = \Pi_{i=k}^{\infty} \mathcal{A}_i$ , with  $\Pi_{(A)}^+ \equiv \prod_{k=0}^{\infty} \Pi_{(A)}^{k,+}$  designating the *full one-sided nonstationary shift space* on the alphabet sequence  $(\mathcal{A}_i)_{i \geq 0}$ .

Considering the  $0^{\text{th}}$  component  $\Sigma_{(L)}^{0,+} = \Pi_{k=0}^{\infty} \{1, \dots, l_k\}$ , we define an order on its Bratteli diagram by ordering the set  $\mathcal{A}_k^j = \mathcal{A}_k$  by the symbol labels,  $0 < 1 < 2 < \dots < (l_k - 1)$ , so incoming edges are ordered by the symbol they come from. We call this special type of stable order a *vertex order*.

The Bratteli diagram is then proper, as there is one point which has no successor, the point  $(.l_0 l_1 l_2 \dots)$ . If we define the image of this point to be  $(.0000\dots)$ , the unique point with no predecessor, then this extension of  $T$  to all of  $\Sigma_{(L)}^{0,+}$  is the unique continuous extension. In the classical case where  $l_i = d$  is constant,  $\Sigma^+ = \Sigma_{(L)}^{0,+}$  is the full one-sided left shift space on  $d$  symbols and the extended transformation  $T$  is the Kakutani-von Neumann *adding machine* or *d-adic odometer*.

*Example 2.* (induced adics: Markov adding machine; stationary adics) When the alphabets, matrices and order  $\mathcal{A}_i = \mathcal{A}$ ,  $L_i = L$ ,  $\mathcal{O}_i = \mathcal{O}$  are all constant (so  $L$  is  $(d \times d)$ ), all components of the *nsft* can be naturally identified, giving a one-sided *sft*  $\Sigma_L^+$ , and the map is termed a *stationary* adic transformation on this space.

The simplest example of this is when the order is a vertex order; then the adic transformation is the induced (first return) transformation on the subset of the  $d$ -adic odometer consisting of the allowed strings and the adic transformation could be called a (stationary) “Markov adding machine”. We remark that this is an unusual type of induced map as this subset has measure zero for the invariant probability measure on the odometer. Since a stable set of a string in the *sft* is a subset of the stable set in the one-sided full shift for that string, one can picture the Markov adding machine geometrically as hopping from one branch to the next in this allowed subtree of the stable tree for that string.

In fact any adic transformation can be recoded as a nonstationary Markov adding machine; taking as a new alphabet the edge sets, each edge set is partitioned into ordered subsets by the edge order and we simply extend this consistently in some way to a linear ordering of the whole edge set; see Lemma 5.1.

We note that even though the odometer itself has a unique continuous extension, this may not be true for a Markov adding machine, and in any case the extension may differ from that of the induced map. The edge coding of the dual irrational rotation adic discussed next exhibits this behavior, see Proposition 2.7.

*Example 3.* (Adic models of circle rotations, irrational angle) Associated to an irrational rotation  $R_\theta : x \mapsto x + \theta \pmod{1}$  of the circle  $\mathbb{R}/\mathbb{Z}$  we describe two different codings by adic transformations. These have the same matrix sequence  $(M_i)_{i \geq 0}$  and so the same Bratteli diagram, but have different orders which lead to an important difference between them: while the first adic transformation has a unique extension to a homeomorphism on the whole space  $\Sigma_{(M)}^{0,+}$ , the second does not. What happens is that for the first both  $\mathcal{NS}$  and  $\mathcal{NP}$  have a single element, while for the second  $\mathcal{NP}$  has two elements and no such extension is possible.

We take as our alphabet  $\mathcal{A}_i = \{A, B\}$  for all  $i \geq 0$ ; writing  $\mathcal{A}^*$  for the collection of finite words in the letters of  $\mathcal{A}$ , we recall that a *substitution* is a map from  $\mathcal{A}$  to  $\mathcal{A}^*$ . We define two pairs of substitutions  $\rho_0, \rho_1$  and  $\tilde{\rho}_0, \tilde{\rho}_1$  as follows: for the *Rauzy substitution pair*

$$\begin{aligned} \rho_0(A) &= A, & \rho_1(A) &= AB, \\ \rho_0(B) &= AB, & \rho_1(B) &= B \end{aligned}$$

while for the *dual substitution pair* we have

$$\begin{aligned} \tilde{\rho}_0(A) &= A, & \tilde{\rho}_1(A) &= AB, \\ \tilde{\rho}_0(B) &= BA, & \tilde{\rho}_1(B) &= B. \end{aligned}$$

A substitution extends by concatenation to a map from  $\mathcal{A}^*$  to itself which is a homomorphism of the free semigroup on two generators  $A, B$ ; the *abelianization* of the substitution forgets the order of the letters and so is a homomorphism of the free *abelian* semigroup, and therefore is conveniently specified by a  $(2 \times 2)$  non-negative integer matrix. We take a ‘‘column vector’’ convention here, so the  $(i, j)$  entry of the matrix  $M$  corresponding to the substitution  $\rho$  will be the numbers of  $i$ 's occurring in  $\rho(j)$  for  $i, j \in \{A, B\}$  written in that order.

So the matrix of both  $\rho_0$  and  $\tilde{\rho}_0$  is  $P$ , and for both  $\rho_1$  and  $\tilde{\rho}_1$  is  $Q$ , where

$$P = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

We now choose an infinite sequence of matrices  $(M_i)_{i \geq 0}$  such that each matrix is  $P$  or  $Q$ , and to this we associate two infinite sequences of substitutions  $(\rho_i)_{i \geq 0}$  and  $(\tilde{\rho}_i)_{i \geq 0}$  with those abelianizations.

Forming the Bratteli diagram for the matrix sequence, we think of the substitutions as acting from right to left in the Bratteli diagram, which is why we have taken the column vector convention for the matrices. With edges oriented towards the past to agree with the direction of the substitution maps, the substitutions determine a stable order on the diagram; the edges coming out of a symbol  $j$  are ordered by the order of the letters in the word  $\rho(j)$ . See Fig. 3. In the stable tree picture, the substitutions map in the downward direction, see Fig. 4.

To specify the matrix sequence, we choose a sequence of strictly positive integers  $(n_i)_{i \geq 0}$  and a choice of *parity* 0 or 1. If the parity is 0 we define  $M_i = P$  for  $0 \leq i \leq n_0 - 1$ ,  $M_i = Q$  for  $n_0 \leq i \leq n_0 + n_1 - 1$  and so on; for parity 1 we begin with  $Q$ ; the matrix sequences for parities 0 and 1 are then  $P^{n_0} Q^{n_1} P^{n_2} \dots$  and  $Q^{n_0} P^{n_1} Q^{n_2} \dots$ .

So we have a single matrix sequence  $(M_i)_{i \geq 0}$  whose parity changes infinitely often, and the two corresponding substitution sequences, defining two ordered Bratteli diagrams.

We write  $T, \widetilde{T}$  for the corresponding adic transformations, and we show:

**Proposition 2.7.** *For the Rauzy order,  $\mathcal{NS} = \{(\overline{B})\}$  and  $\mathcal{NP} = \{(\overline{A})\}$ . The extension  $T_f$  defined by  $f(\overline{B}) = (\overline{A})$  is a homeomorphism. Via the coding of  $A, B$  to  $0, 1$  it is induced from the 2-adic odometer as a Markov adding machine. For the dual order,  $\widetilde{\mathcal{NS}}$  has a single element, which equals  $(.A^{n_0}B^{n_1}A^{n_2} \dots)$  (for parity 0) or  $(.B^{n_0}A^{n_1}B^{n_2} \dots)$  (for parity 1). By contrast  $\widetilde{\mathcal{NP}} = \{(\overline{A}), (\overline{B})\}$ , and no continuous extension of  $\widetilde{T}$  to  $\Sigma_{(M)}^{0,+}$  is possible. By switching the coding of  $A, B$  to  $0, 1$  according to parity, the dual rotation adic is realized as an induced map of the odometer for which there exists no continuous extension.*

*Proof.* Considering first the Rauzy order, we verify that these are indeed the unique maximal and minimal elements. Let  $x = (x_i)_{i \geq 0} \in \Sigma_{(M)}^{0,+}$  and suppose that  $x_{k-1}x_k = AB$ . There are two incoming arrows to the symbol  $B$  at time  $k$ , and  $BB > AB$ . Therefore  $x$  is not maximal in  $W^s(x)$ . So if a string  $x$  is a maximal element in  $W^s(x)$  then  $x$  cannot contain a word  $x_{k-1}x_k = AB$  for any  $k$ . The only possibility (other than  $(\overline{B})$ , which is clearly maximal) is  $x = (.BBBB \dots B\overline{A})$  where the first  $A$  occurs, say, in place  $k$ . Since by assumption both  $P$  and  $Q$  each occur infinitely often, there is some  $m > k$  such that there exists a string  $w$  in  $W^s(x)$  with  $w_{m-1}w_m = BA$ . But then  $BA > x_{m-1}x_m = AA$  so  $x$  is not maximal. Therefore the only possible element of  $\mathcal{NS}$  is  $(\overline{B})$ . By the symmetric argument, the only string in  $\mathcal{NP}$  is  $(\overline{A})$ .

We next show that the extended map with  $f(\overline{B}) = (\overline{A})$  is continuous. Suppose  $x^{(i)} \rightarrow (\overline{B})$ ; we define  $m = m_i$  to be the place where the next-to-last initial  $B$  occurs, so  $x = x^{(i)} = (.x_0x_1x_2 \dots x_mx_{m+1}x_{m+2}x_{m+3}x^+) = (.BBB \dots BBACx^+)$ , where  $x_{m+3} = C = A$  or  $B$ . We say that a symbol *splits back* if it has two entering edges. Then since  $x_{m+2} = A$  splits back, we know that  $L_{m+1} = Q$ . Now define  $n = n_i \leq m = m_i$  to be the greatest integer such that  $L_n = P$ . We claim that the successor of  $x$  is  $(.AA \dots A_nB_{n+1}B_{n+2} \dots BC_{m+3}x^+)$ , where for clarity we index some of the symbols by their place. The reason is, first, that  $AC_{m+3} < BC_{m+3}$ ; but since  $L_{n+1}, L_{n+2}, \dots, L_{m+1} = Q$ , the symbol  $B_{m+1}$  does not split back. Hence  $B$  must be maintained until the index decreases to  $n$ , where  $L_n = P$  allows the transition to be made to  $x_n = A$ . From then on  $A$  is not only allowed but is always the least choice. This proves the claim.

Now finally, as  $x^{(i)} \rightarrow (\overline{B})$  then  $m_i \rightarrow \infty$ , and since the parity changes infinitely often,  $n_i \rightarrow \infty$  as well. This proves that  $f(x^{(i)}) \rightarrow (\overline{A})$ , and the extended map indeed is continuous.

Lastly we claim that this is induced from the 2-adic odometer where  $A, B$  are identified with  $0, 1$ , but this is clear for  $\Sigma_{(M)}^{0,+} \setminus \mathcal{N}$ , because the order on edges  $AC < BC$  is the vertex order. And since for the odometer also  $(\overline{B})$  maps to  $(\overline{A})$ , this extended adic transformation is an induced map of the adding machine on all of  $\Sigma^+$  to the measure-zero subset of allowed strings  $\Sigma_{(M)}^{0,+} \subseteq \Sigma^+$ , as claimed.

Now we move to the dual order. The verification for each of the unique minimal elements is similar to that for the minimal element of the Rauzy order, so  $\widetilde{\mathcal{NP}} = \{(\overline{A}), (\overline{B})\}$ . The maximum element  $(.x_0x_1\dots)$  is defined uniquely by the rule: given  $x_{k+1}$ , then  $x_k \neq x_{k+1}$  whenever this is possible. And since the parity switches infinitely often, there is a unique string which satisfies this, given by (type 0)  $(.A^{n_0}B^{n_1}A^{n_2}\dots)$  and (type 1)  $(.B^{n_0}A^{n_1}B^{n_2}\dots)$ . Therefore since  $\widetilde{\mathcal{NP}}$  has a single element, no extension to a homeomorphism is possible. In fact there is no continuous extension, for the following reason. Let  $x$  denote the maximum element, say with parity 0, so  $x = (.x_0x_1\dots) = (.A^{n_0}B^{n_1}A^{n_2}\dots)$ . For  $t_k = n_0 + n_1 + \dots + n_k - 1$ , let  $y$  be a string with  $y_i^{(k)} = x_i$  for  $0 \leq i \leq t_k \equiv j$ ; for  $k$  odd we take  $y_{j+1}^{(k)} = y_j = B$ , so necessarily  $y_{j+2} = B$ , while for  $k$  even we take  $y_{j+1}^{(k)} = y_j = A$ . Then the successor of  $y^{(k)}$  for  $k$  odd is  $(.AA\dots AB y_{i+2}^{(k)}\dots)$  while for  $k$  even the successor of  $y^{(k)}$  is  $(.BB\dots BA y_{i+2}^{(k)}\dots)$ .

Hence  $\widetilde{T}(y^{(2j)}) \rightarrow (.AA\dots)$  while  $\widetilde{T}(y^{(2j+1)}) \rightarrow (.BB\dots)$  as  $j \rightarrow \infty$ , and there is no continuous extension.

To produce from this example a Markov adding machine with no continuous extension, we can proceed in two ways; first, by Lemma 5.1, recoding this with the edge alphabets (which have three symbols) gives a vertex order and so represents the adic map homeomorphically as a Markov adding machine with no continuous extension. Alternatively we can represent it as an induced map of the two-adic odometer, by coding  $A, B$  to 0, 1 for odd parity and  $A, B$  to 1, 0 for even.  $\square$

Dynamically both maps are related to a circle rotation  $R_\theta$ . Stating this for the Rauzy substitutions:

**Proposition 2.8.** *The adic transformation  $(\Sigma_{(M)}^{0,+}, T_{\mathcal{O},f})$  factors naturally onto the rotation  $R_\theta : x \mapsto x + \theta \pmod{1}$  where  $\theta$  is defined as follows: writing*

$$\alpha = [n_0 \dots n_k \dots] \equiv \frac{1}{n_0 + \frac{1}{n_1 + \dots}}$$

*then for parity 0,  $\theta = \alpha/(\alpha+1) \in (0, 1/2)$  while for parity 1,  $\theta = 1/(\alpha+1) \in (1/2, 1)$ .*

*Proof.* Noting that a circle rotation is an exchange of two intervals, the proof follows as for other interval exchanges, as shown explicitly in [Fisa]; for the factor map, countably many pairs of points are identified (the endpoints of the Cantor set intervals).  $\square$

For another proof, the factor map can be given an explicit arithmetical Ostrowsky formula as in [AF01] (there we use the dual substitution sequence). Related arithmetical formulas are studied in [GLT95] and a related Bratteli diagram (used for an a priori completely different purpose) appears in [ES80].

*Example 4.* (Adic models of circle rotations: rational angle) We focus on the Rauzy substitutions. Dynamically the adic transformation  $T_{\mathcal{O}}$  is trivial as the Bratteli diagram is trivial, i.e.  $\Sigma_{(M)}^{0,+} \setminus \mathcal{N}$  is empty, but the extended map is nevertheless interesting: these maps approximate the irrational rotations, and as for the irrational case they have a unique continuous extension to all of  $\Sigma_{(M)}^{0,+}$ .

Now the continued fractions are finite, and we use this notation: given a finite sequence  $n_0, n_1, \dots, n_k, n_{k+1}$  with  $n_0, \dots, n_k$  integers  $\geq 1$  and  $n_{k+1} = \infty$ , and a parity choice 0 or 1, we define an infinite sequence of substitutions  $(\rho_i)_{i \geq 0}$  as before, but now for  $i \geq n_0 + \dots + n_k$   $\rho_i$  is constantly equal to  $\rho_0$  or to  $\rho_1$ .

Similarly the matrix sequence  $M_i$  is either  $P$  or  $Q$  for all  $i > n_0 + \dots + n_k$ ; thus e.g. for parity 0 the matrix sequence is either  $(M_i)_{i \geq 0} = P^{n_0} Q^{n_1} \dots P^{n_k} Q^\infty$  or  $P^{n_0} Q^{n_1} \dots Q^{n_k} P^\infty$ .

The Bratteli diagram is now both left- and right-trivial, and there are two stable equivalence classes,  $W^s(\overline{B})$ , which is finite, and  $W^s(\overline{A})$ , which is countably infinite, and so  $\Sigma_{(L)}^{0,+} = W^s(\overline{B}) \cup W^s(\overline{A})$ .

The simplest example is the *twist adic*, with  $[n_0 \dots n_k] = [\infty]$  and matrix sequence  $P^\infty$  or  $Q^\infty$  for parities 0 and 1. We consider  $Q^\infty$ , for which  $\{T_{\mathcal{O}}^k(\overline{A}) : k = 0, 1, 2, \dots\} = \{(\overline{A}), (\overline{B\overline{A}}), (\overline{BB\overline{A}}), \dots\}$ ; this orbit accumulates to the point  $(\overline{B})$ .

More generally, supposing that the sequence ends in  $Q^\infty$ , then as for the twist  $\mathcal{NS} = \{(\overline{B})\}$  but now  $\mathcal{NP} = \{(\overline{A}), (\overline{A \dots A\overline{B}})\}$  with  $n_0 + \dots + n_k$   $A$ 's; we write this last point as  $(\overline{A^{n_0+\dots+n_k}\overline{B}})$ . There are two choices for the extended maps,  $f(\overline{B}) = (\overline{A^{n_0+\dots+n_k}\overline{B}})$  and  $\tilde{f}(\overline{B}) = (\overline{A})$ . These are both natural choices, for different reasons:  $T_{\mathcal{O},f}$  is continuous, while  $T_{\mathcal{O},\tilde{f}}$  gives a transformation which is induced from the 2-adic odometer, since the order  $\mathcal{O}$  is the vertex order of the odometer and since there  $(\overline{B})$  is by definition mapped to  $(\overline{A})$ .

We note that for  $T_{\mathcal{O},\tilde{f}}$ , the orbit of  $(\overline{B})$  is infinite and hence has measure zero, so there is no invariant probability measure for this extension.

We are more interested in the extension  $T_{\mathcal{O},f}$  (see Fig. 4); this is a continuous map of a compact space and so must have at least one invariant probability measure [Wal82]. Indeed it has a periodic point  $(\overline{B})$  and the invariant measure is unique, given by equally weighted point masses on this finite orbit. The same holds for the group  $\mathcal{FC}$ . Neither extension  $T_{\mathcal{O},f}$ ,  $T_{\mathcal{O},\tilde{f}}$  nor  $\mathcal{FC}$  is minimal.

Finally we note that if instead the sequence ends in  $P^\infty$ , the dynamics (relabelled) are as if the arrows in the figure are reversed; now  $\mathcal{NS} = \{(\overline{B}), (\overline{B^{n_0+\dots+n_k}\overline{A}})\}$  and  $\mathcal{NP} = \{(\overline{A})\}$ , and there is one choice for the extension. The cylinder set  $[\overline{B}]$  is finite so this is continuous. The orbit of  $(\overline{B})$  now begins at time  $-\infty$ , and is attracted in negative time to the periodic orbit; at time 0 it reaches  $(\overline{B})$  and then jumps to join the periodic orbit.

So in both cases we have a countable orbit attracting to (or repelled from depending on parity) a periodic orbit, and the map  $T_{\mathcal{O},f}$  is uniquely ergodic but not minimal; this is in fact an adic version of a classical such example.

**Proposition 2.9.** *Assuming the sequence ends in  $Q^\infty$ , defining  $f : \mathcal{NS} \rightarrow \mathcal{NP}$  by sending the unique maximal point  $(\overline{B})$  to  $(\overline{A^{n_0+\dots+n_k}\overline{B}})$  gives the unique continuous*

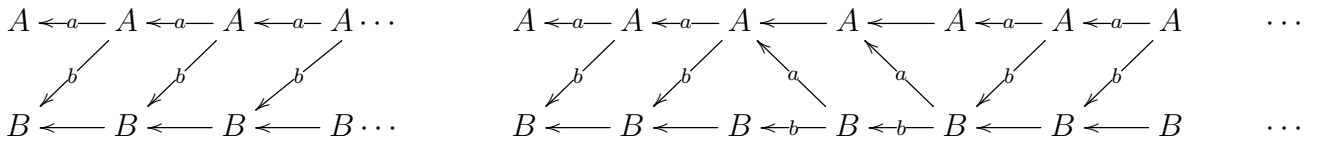


FIGURE 3. Rotation adics with the Rauzy substitutions: the twist  $P^\infty$ , with fixed point, and an irrational rotation adic, with matrices  $P, P, Q, Q, P, P, \dots$ . We have oriented the edges toward the past of the *nsft* to agree with the substitution maps.

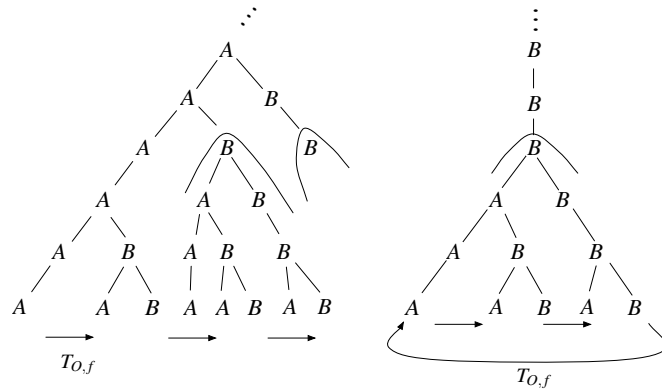


FIGURE 4. In the rational rotation adic  $QPQP^\infty$ , the orbit of  $(.\bar{A})$  is attracted to the (periodic) orbit of  $(.\bar{B})$ .

extension  $T_{O,f}$ . The  $T_{O,f}$ -orbit of  $(.\bar{B})$  is periodic of period  $q$  where setting

$$\alpha = [n_0 \dots n_k] \equiv \frac{1}{n_0 + \frac{1}{n_1 + \dots + \frac{1}{n_k}}}$$

then for parity 0,  $\theta \equiv \alpha/(\alpha + 1) = p/q \in \mathbb{Q} \cap (0, 1/2]$  ( $p/q$  in least terms) while for parity 1,  $p/q = \theta \in (1/2, 1)$  and  $\theta = 1/(\alpha + 1)$ . The (forward) orbit of  $(.\bar{A})$  accumulates to this periodic attractor. If the sequence ends in  $P^\infty$ , we send the two maximal points  $(.\bar{B}), (.\bar{B}^{n_0+\dots+n_k}\bar{A})$  to the unique minimal point  $(.\bar{A})$ ; this is continuous, and the orbit of  $(.\bar{B})$  is attracted to the periodic orbit of  $(.\bar{A})$  in negative time. In both cases the map  $T_{O,f}$  is uniquely ergodic but not minimal.

*Proof.* Most of the proof has just been given; see Fig. 4.

Regarding the length of the period, we note that this is the number of paths for which  $x_l = B$  for any  $l \geq k$ , thus is the sum of the second column of the product  $M_0 \dots M_k$ . The rotation by angle  $p/q$  is the exchange of two intervals, of lengths  $\alpha$  and 1, and has that same period.  $\square$

*Example 5.* (Interval exchanges.) Ferenczi notes in [Fer97] that every interval exchange can be represented as a cutting and stacking construction, while every cutting

and stacking has an adic representation. Thus implicitly every interval exchange transformation of  $d$  intervals can be represented as an adic transformation. We show this explicitly in [Fisa], building the Bratteli diagram from a sequence of  $(d \times d)$  determinant-one matrices, exactly those matrices which appear in the theory for a different purpose, that of controlling the Rauzy induction. (Here one can use either Veech's or Kerchoff's formalism, [Vee78], [Ker85]; the Bratteli diagrams are canonically conjugate by a sequence of permutation matrices.) In the case of Rauzy induction,  $\mathcal{NP}$  and  $\mathcal{NS}$  are singletons (just as for the case of two intervals in Example 3), while other induction procedures can lead to nonproper Bratteli diagrams. See [Fisa].

*Example 6.* (Pascal adic transformation) Much more freedom is gained by considering  $0 - 1$  matrices with unbounded alphabet size; one can then find adic transformations of positive entropy, and in fact as Vershik showed, given any invertible measure-preserving transformation of a Lebesgue space, there is an adic transformation measure-theoretically isomorphic to it [Ver81], [LV92]. An interesting specific example with unbounded alphabets was described by Vershik and Lodkin in [LV85]; it is an adic model of a map of the interval defined via cutting and stacking by Hajian, Ito and Kakutani [HIK72]. We take alphabets  $\mathcal{A}_k = \{0, 1, 2, \dots, k\}$  for  $k = 0, 1, \dots$  with edges from  $m$  to  $m$  and to  $m + 1$  for  $m \in \mathcal{A}_k$ , so the matrices

are  $L_0 = \begin{bmatrix} 1 & 1 \end{bmatrix}$ ,  $L_1 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$ ,  $L_2 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ ,  $\dots$  and one can picture the

Bratteli diagram as a Pascal triangle laid on its side. The stable edge order  $\mathcal{O}$  is given by  $(k - 1)k < kk$ . The matrix sequence is nonprimitive since  $L^{(1,k)} = L_1 L_2 \cdots L_k$  has entries  $L_{20}^{(1,k)} = L_{1(k+2)}^{(1,k)} = 0$ .

**Proposition 2.10.** *There are a countable infinity of maximal and minimal paths, with  $\mathcal{NP} = \{(.0^k 1234 \dots) : k \geq 1\}$  and  $\mathcal{NS} = \{(.01234 \dots (k - 1)kkkk \dots) : k \geq 0\}$ . No extension to a homeomorphism of  $\Sigma_{(L)}^{+,0}$  is possible.*

*Proof.* That these are the maximum and minimum paths is easily verified; note that for the inverse order  $\tilde{\mathcal{O}}$  these extreme sets are indeed symmetrically related in the Pascal triangle. There is a unique continuous extension to  $\Sigma_{(L)}^{+,0}$ : if  $x^{(n)} \rightarrow (\bar{0})$  then for some  $m$ ,  $x^{(n)} = (.0^m 1x^+)$ , so  $T(x^{(n)}) = (.0^{m-1} 11x^+) \rightarrow (\bar{0})$ ; for  $x = (.01234 \dots (k - 1)kkkk \dots)$  then if  $x^{(n)} \rightarrow x$ , then  $x^{(n)} = (.01234 \dots (k - 1)kk^m(k + 1)x^+)$  for some  $m \geq 0$  so  $T(x^{(n)}) = (.0^m 12 \dots k(k + 1)(k + 1)x^+) \rightarrow (\bar{0})$  as  $n \rightarrow \infty$ . Therefore the continuous extension  $T_f$  sends each element of  $\mathcal{NS}$  to  $(\bar{0})$ , a fixed point for  $T_f$ . By symmetry, there is a unique continuous extension for the inverse adic map  $T^{-1}$ , sending  $\mathcal{NP}$  to  $(.01234 \dots)$ , so  $T_f$  is not a homeomorphism.  $\square$

For another representation of this map, let  $\Sigma^+ = \Pi_0^\infty \{0, 1\}$ ; for  $w = (w_0 w_1 \dots) \in \Sigma^+$ , then the map  $\phi : \Sigma^+ \rightarrow \Sigma_{(L)}^{+,0}$  by  $\phi(x) = w$  with  $w_0 = 0$ ,  $w_n = \sum_{i=0}^{n-1} x_i$  is a bijection (this is a random walk with drift). The sets  $\mathcal{NP}, \mathcal{NS}$  then correspond to  $\{(.1^k 0^\infty) : k \geq 0 \text{ or } k = \infty\}$ ,  $\{(.0^k 1^\infty) : k \geq 0 \text{ or } k = \infty\}$ . See [MP05], respectively [FP08] for an analysis of further properties of this and the closely related Euler adic transformation; also see §5.1.

**2.5. Minimality and unique ergodicity.** We define the action of a group  $G$ , or semigroup  $S$ , on a topological space to be *minimal* if every orbit is dense, and (now assuming the action is Borel measurable) *uniquely ergodic* if there is a unique invariant probability measure. See e.g. [Wal82], [Fur81].

Note that since our definition assumes neither continuity of the map nor compactness of the space, it applies not only to the actions of the group  $\mathcal{FC}(\Sigma_{(L)}^{0,+})$  (on the compact space  $\Sigma_{(L)}^{0,+}$ , by homeomorphisms) but also to the homeomorphism  $T_{\mathcal{O}}$  of the (in general) noncompact space  $\Sigma_{(L)}^{0,+} \setminus \mathcal{N}$  and the extended adic transformations  $T_{\mathcal{O},f}$  on  $\Sigma_{(L)}^{0,+}$ ; as we have seen in Proposition 2.7, these extended maps  $T_{\mathcal{O},f}$  may be neither continuous nor bijective.

We have these simple relationships:

**Proposition 2.11.**

(i) *An invariant measure for the adic transformation  $T_{\mathcal{O}}$  on  $\Sigma_{(L)}^{0,+} \setminus \mathcal{N}$  is invariant for  $\mathcal{FC}$  and for  $T_{\mathcal{O},f}$  on  $\Sigma_{(L)}^{0,+}$ . A nonatomic invariant probability measure for  $\mathcal{FC}$  or  $T_{\mathcal{O},f}$  gives an invariant probability measure for  $T_{\mathcal{O}}$ .*

(ii) *If  $(L_i)_{i \geq 0}$  is left nontrivial, then unique ergodicity is true or false simultaneously for all three actions.*

(iii) *If the action of  $\mathcal{FC}$  is minimal then so is  $T_{\mathcal{O}}$  and  $T_{\mathcal{O},f}$ .*

*Proof.* (i): Suppose a probability measure  $\mu$  on  $\Sigma_{(L)}^{0,+} \setminus \mathcal{N}$  is invariant for  $T_{\mathcal{O}}$ . We extend it to all of  $\Sigma_{(L)}^{0,+}$  by giving  $\mathcal{N}$  measure zero. To show that  $\mu$  is invariant for the group  $\mathcal{FC}$  it suffices to check this on the generators. Suppose  $\gamma \in \mathcal{FC}$  interchanges the cylinder sets  $A = [.x_0 \dots x_{t-2}s]$  and  $B = [.y_0 \dots y_{t-2}s]$ , then by the total ordering which defines the adic transformation, one of these words is least, so there exists  $k \geq 0$  such that, say,  $T_{\mathcal{O}}^k(A) = B$  for the adic map. Hence  $T_{\mathcal{O}}^{-k}(B \setminus \mathcal{N}) = A \setminus \mathcal{N}$ , and so  $\gamma$  preserves  $\mu$ . Similarly,  $T_{\mathcal{O},f}$  preserves  $\mu$ , since it is enough to check this on cylinder sets.

Conversely if a probability measure  $\mu$  on  $\Sigma_{(L)}^{0,+}$  is nonatomic then since by Lemma 2.3  $\mathcal{N}$  is countable, it has measure zero so  $\mu$  restricts to a probability measure on  $\Sigma_{(L)}^{0,+} \setminus \mathcal{N}$ . Supposing it is  $\mathcal{FC}$ -invariant, we want to show  $\mu(T_{\mathcal{O}}^{-k}(B \setminus \mathcal{N})) = \mu(A \setminus \mathcal{N}) = \mu(B \setminus \mathcal{N})$  for cylinder sets  $A, B$  as above; defining  $\gamma$  as above to interchange these sets we are done. Also  $T_{\mathcal{O},f}$ -invariance implies  $T_{\mathcal{O}}$ -invariance since these agree on  $\Sigma_{(L)}^{0,+} \setminus \mathcal{N}$  and  $\mathcal{N}$  has measure zero.

(ii): By Lemma 2.4 left-nontriviality implies that there are no finite orbits for any of the three actions. Therefore any invariant probability measure is nonatomic. But in that case unique ergodicity follows from (i).

(iii): If the action of  $\mathcal{FC}$  is minimal then so is  $T_{\mathcal{O},f}$  since its orbits are unions of orbits of the adic map, whose orbits by definition are equal to those of  $\mathcal{FC}$ . Also  $T_{\mathcal{O}}$  is minimal, by restriction.  $\square$

**2.6. Topological mixing implies minimality.** We recall the link between primitivity and topological mixing of the nonstationary shift map, as defined in the introduction. The proof is immediate from the definitions:



**Lemma 2.12.** *An nsft  $(\Sigma_{(L)}^{0,+}, \sigma)$  is topologically mixing if and only if the matrix sequence is primitive.  $\square$*

The effect this property has on the transverse dynamics is given by:

**Theorem 2.13.** (primitivity implies minimality for adics and for  $\mathcal{FC}$ ) *Let  $L_i$  for  $i \geq 0$  be a sequence of  $(l_i \times l_{i+1})$   $0-1$  matrices, and let  $\mathcal{O}$  be a stable order. If the sequence  $(L_i)_{i \geq 0}$  is primitive with nontrivial alphabet sequence, then every point in  $\Sigma_{(L)}^{0,+}$  has a dense orbit for the map  $T_{\mathcal{O}} : \Sigma_{(L)}^{0,+} \setminus \mathcal{NS} \rightarrow \Sigma_{(L)}^{0,+} \setminus \mathcal{NP}$ . The same holds for the adic transformation  $T_{\mathcal{O}} : \Sigma_{(L)}^{0,+} \setminus \mathcal{N} \rightarrow \Sigma_{(L)}^{0,+} \setminus \mathcal{N}$ , for the group  $\mathcal{FC}$  of finite coordinate changes and for any extended adic transformation  $T_{\mathcal{O},f}$ .*

*Proof.* We prove  $\mathcal{FC}$  is minimal by showing this for adic map  $T_{\mathcal{O}} : \Sigma_{(L)}^{0,+} \setminus \mathcal{NS} \rightarrow \Sigma_{(L)}^{0,+} \setminus \mathcal{NP}$ , since this has the same orbits. Given a point  $x = (.x_0x_1\dots) \in \Sigma_{(L)}^{0,+}$ , and a thin cylinder set  $B = [.b_0b_1\dots b_k] \in \mathbb{C}_0^k$  for some  $k$ , we wish to show that there exists  $n \in \mathbb{Z}$  with  $T_{\mathcal{O}}^n(x) \in B$ .

Since  $(L)$  is primitive, there exists  $m$  such that all the entries of  $L_k \cdots L_m$  are  $> 0$ . Thus there exists an allowed string  $(.b_0b_1\dots b_kw_{k+1}\dots w_m)$  such that  $w_m = x_m$ . Therefore the infinite string  $w = (.b_0b_1\dots b_kw_{k+1}\dots w_{m-1}x_mx_{m+1}\dots)$  is allowed. The two points  $x$  and  $w$  are comparable with respect to the order, so either  $x \leq w$  or  $w \leq x$ . In the first case, there exists  $n \geq 0$  with  $T_{\mathcal{O}}^n(x) = w$ , in the second case  $n \leq 0$ .

Minimality for the extended maps  $T_{\mathcal{O},f}$  and for the adic transformation  $T_{\mathcal{O}}$  then follows as in (iv) of Prop. 2.11.  $\square$

A case where primitivity fails but minimality still holds for the adic transformation is given by the Chacon adic transformation of [FFT09]. See §5.1.

*Remark 2.3.* Livshits and Vershik [LV92], as noted in the introduction, do give Def. 2.3 (top of p. 186) and they state both the lemma (top of p. 186) and theorem (middle of that page). The proofs are omitted there but as we have seen are not difficult. Unique ergodicity in the stationary primitive case is stated there, also without proof, (bottom of p. 186).

### 3. MIXING FOR NONSTATIONARY MARKOV CHAINS

In this section we develop basic material on nonstationary Markov chains and nonstationary mixing. We are guided by the stationary theory as presented say in [Bil65] and [Wal82]. See the second appendix regarding connections with inhomogeneous Markov chain theory.

There are several types of nonstationarity one might encounter: the alphabet size and the transition matrices may change in time, and even when the alphabet and transition matrix are fixed, the measure itself may be nonstationary, if one has a Markov process which is not a stationary process. In the setup we will discuss, all of these phenomena can occur.

Given a sequence  $(\mathcal{A}_i)_{i \geq 0}$  of alphabets with  $\#\mathcal{A}_i = l_i$ , the full nonstationary shift space has as its components  $\Pi_{(\mathcal{A})}^{k,+} \equiv \prod_{i=k}^{\infty} \mathcal{A}_i$  and the nsft is then  $\Pi_{(\mathcal{A})}^+ \equiv \prod_{k=0}^{\infty} \Pi_{(\mathcal{A})}^{k,+}$ .

(Since we are giving a general treatment, we work with this full shift space rather than restricting at this point to an *nsft*.)

An  $(l_i \times l_{i+1})$  non-negative real matrix sequence  $(A_i)_{i \geq 0}$  acts on the left along the column vector spaces  $C_i$  and on the right along the row vector spaces  $R_i$ , as shown in the diagram, where the direction of the arrows has been chosen so that composition of maps is given by matrix multiplication without order reversal:

$$R_0 \xrightarrow{A_0} R_1 \xrightarrow{A_1} R_2 \xrightarrow{A_2} R_3 \cdots$$

and

$$C_0 \xleftarrow{A_0} C_1 \xleftarrow{A_1} C_2 \xleftarrow{A_2} C_3 \cdots$$

**Definition 3.1.** For a sequence  $(A)_{i \geq 0}$  of  $(l_i \times l_{i+1})$  non-negative real matrices, if we have a sequence  $\mathbf{v}_i$  of nonzero column vectors and nonzero numbers  $\lambda_i$  satisfying

$$A_i \mathbf{v}_{i+1} = \lambda_i \mathbf{v}_i$$

or row vectors satisfying

$$\mathbf{v}_i^t A_i = \lambda_i \mathbf{v}_{i+1}^t$$

for each  $i \geq 0$  we call this a column *respectively* row eigenvector sequence with eigenvalues  $\lambda_i$ .

We write  $C_i^+ \subseteq C_i$  and  $R_i^+ \subseteq R_i$  for the *positive cones* i.e. the cones of non-negative vectors, see Def. 6.1; since the matrices are non-negative these are preserved in the above diagram.

We recall that a vector  $\mathbf{v} \in \mathbb{R}^{l_k}$  is a *probability vector* if each component  $v_i$  is  $\geq 0$  and if  $\sum_{i=0}^{l_k-1} v_i = 1$ . We use the norm  $\|\mathbf{v}\| \equiv \sum |v_k|$ , so the subset of the positive cone  $C_k^+$  with norm one is the closed unit simplex  $\Delta_k$ , the probability column vectors. We denote the positive simplex of rows by  $\Delta_i^t \subseteq R_i^+$ , and have projections  $\text{Proj} : (C_i^+ - \{\mathbf{0}\}) \rightarrow \Delta_i$  and  $\text{Proj} : (R_i - \{\mathbf{0}\}) \rightarrow \Delta_i^t$  with  $\mathbf{v} \mapsto \mathbf{v}/\|\mathbf{v}\|$  and  $\mathbf{v}^t \mapsto \mathbf{v}^t/\|\mathbf{v}^t\|$ .

We write  $\mathbf{1}_i$  for the  $(l_i \times 1)$  column vector all of whose entries are 1. This gives a convenient way to say that an  $(l_i \times l_{i+1})$  matrix  $P_i$  is a *probability* or *stochastic matrix*, i.e. that each row is a probability vector: that holds if and only if  $P_i \mathbf{1}_{i+1} = \mathbf{1}_i$ . Moreover we have:

**Lemma 3.1.** An  $(l_i \times l_{i+1})$  matrix sequence  $(P_i)_{i \geq 0}$  is stochastic if and only if  $(\mathbf{1}_i)_{i \geq 0}$  is a column eigenvector sequence with constant eigenvalue one, if and only if in its action on row vectors the sequence  $(P_i)$  maps  $\Delta_i^t$  to  $\Delta_{i+1}^t$ .

*Proof.* For  $\boldsymbol{\pi}_i^t \in \Delta_i^t$ ,  $(\boldsymbol{\pi}_i^t P_i) \mathbf{1}_{i+1} = 1$  if and only if  $\boldsymbol{\pi}_i^t P_i \in \Delta_{i+1}^t$ , and  $\boldsymbol{\pi}_i^t (P_i \mathbf{1}_{i+1}) = 1$  for each  $\boldsymbol{\pi}_i^t \in \Delta_i^t$  if and only if  $P_i \mathbf{1}_{i+1} = \mathbf{1}_i$ , as  $\Delta_i^t$  spans the affine hyperplane orthogonal to  $\mathbf{1}_i$ .  $\square$

Thus, given a stochastic sequence  $(P_i)_{i \geq 0}$  we can generate a normalized row eigenvector sequence with constant eigenvalue one as follows: begin with any  $\boldsymbol{\pi}_0 \in \Delta_0$ , and set  $\boldsymbol{\pi}_k^t = \boldsymbol{\pi}_0^t P_0 P_1 \cdots P_{k-1} \in \Delta_k^t$ .

Now suppose we start with a non-negative real matrix sequence  $(A_i)_{i \geq 0}$  which is not necessarily stochastic. Choosing a probability vector  $\boldsymbol{\pi}_0 \in \Delta_0$ , we use  $(A)$  to

define a function  $\mu_0$  on the collection  $\mathbb{C}_0^m$  of all thin cylinder sets of  $\Pi_{(\mathcal{A})}^{0,+}$  for  $m \geq 0$ , as follows. For  $\boldsymbol{\pi}_0^t = ((\boldsymbol{\pi}_0^t)_1, \dots, (\boldsymbol{\pi}_0^t)_{l_0})$ , we set

$$\begin{aligned} \mu_0([\cdot x_0]) &= (\boldsymbol{\pi}_0^t)_{x_0} \text{ and for } m \geq 1, \\ \mu_0([\cdot x_0 \dots x_m]) &= (\boldsymbol{\pi}_0^t)_{x_0} (A_0)_{x_0 x_1} \cdots (A_m)_{x_{m-1} x_m}. \end{aligned} \quad (2)$$

This may not define a measure; the necessary and sufficient condition to get a measure is that these be stochastic matrices:

**Proposition 3.2.** *Let  $(P_i)_{i \geq 0}$  be a sequence of  $(l_i \times l_{i+1})$  stochastic matrices.*

(i) *Then for any fixed  $k$ ,  $\mu_0$  as defined in (2) for  $A_i = P_i$  extends in a well-defined way from thin to general cylinder sets of  $\Pi_{(\mathcal{A})}^{0,+}$  where it is finitely additive, and extends from there (in a unique way) to a countably additive measure on the Borel  $\sigma$ -algebra of  $\Pi_{(\mathcal{A})}^{0,+}$ .*

(ii) *Conversely if (2) defines a measure, then each  $P_i = A_i$  is stochastic.*

(iii) *Writing  $\mu_k$  for the measure defined from the sequence  $\sigma^k(P) = (P_k, P_{k+1}, \dots)$ , with some initial probability vector sequence  $\boldsymbol{\pi}_k^t = \boldsymbol{\pi}_0^t P_0 P_1 \cdots P_{k-1}$ , then  $(\mu_k)_{n \geq 0}$  on  $\Pi_{(\mathcal{A})}^{k,+}$  is an invariant measure sequence for the shift map on the full nonstationary shift space  $\Pi_{(\mathcal{A})}^+$ , i.e. satisfies  $\mu_{k+1} = \mu_k \circ \sigma^{-1}$ , if and only if  $\boldsymbol{\pi}_k^t$  is an eigenvector sequence with eigenvalue 1, i.e.  $\boldsymbol{\pi}_k^t = \boldsymbol{\pi}_0^t P_0 P_1 \cdots P_{k-1}$ .*

*Proof.* For simplicity in the proof we assume all alphabets are identical, and equal to  $\mathcal{A} = \{0, 1\}$ . First we extend  $\mu_0$  to a general cylinder set terminating in a given symbol, e.g.  $[\cdot *** 0]$  or  $[\cdot 11 * 0]$ , by adding the contributions from the thin cylinders which make it up. This is well-defined as that decomposition is unique. Next we verify that  $\mu_0$  is additive; we have to show that e.g.  $\mu_0([\cdot 11]) = \mu_0([\cdot 11 *]) = \mu_0([\cdot 111] \cup [\cdot 110]) = \mu_0([\cdot 111]) + \mu_0([\cdot 110])$ ; this uses the fact that each  $P_i$  is stochastic. To prove shift-invariance of the measure sequence  $\mu_k$  it is sufficient to check this on thin cylinders; we need to show that e.g.  $\mu_1([\cdot 11])$  is equal to  $\mu_0(\sigma^{-1}([\cdot 11])) = \mu_0([\cdot 111] \cup [\cdot 011])$  which by additivity we know is equal to  $\mu_0([\cdot 111]) + \mu_0([\cdot 011])$ ; this now follows from the hypothesis  $\boldsymbol{\pi}_i^t P_i = \boldsymbol{\pi}_{i+1}^t$ . Conversely, if the measure sequence is invariant, since by (2) the components  $(\boldsymbol{\pi}_k^t)_i$  of the vector  $\boldsymbol{\pi}_k^t$  are  $\mu_k[i]$ , this is an eigenvector sequence.  $\square$

*Remark 3.1.* Any measure  $\mu_0$  defined on  $\Pi_{(\mathcal{A})}^{0,+}$  as above from a stochastic sequence  $P_i$  and some initial probability vector  $\boldsymbol{\pi}_0^t$  satisfies the *Markov property*, that past and future events are independent relative to the the present state; conversely a measure with that property determines the initial vector and transition matrix sequence.

**Definition 3.2.** *We call the measure  $\mu_k$  on  $\Pi_{(\mathcal{A})}^{k,+}$  a (nonstationary) Markov measure or nonstationary Markov chain. We call an invariant Markov measure sequence  $(\mu_i)_{i \geq 0}$  on  $\Pi_{(\mathcal{A})}^+$  a nonstationary Markov shift.*

*Remark 3.2.* In the probability literature a nonstationary Markov chain is also called an *inhomogeneous Markov chain*.

In our usage, a nonstationary Markov shift is a mapping family for the shift map along the sequence of components  $\Pi_{(\mathcal{A})}^{k,+}$  with an *invariant* sequence of Markov measures  $(\mu_i)_{i \geq 0}$ ; conversely note that a Markov measure  $\mu_0$  on  $\Pi_{(\mathcal{A})}^{0,+}$  always determines an invariant sequence of Markov measures. See also Remark 3.3. Since in this case each measure in the sequence determines all the later ones, one can also work just with the Markov measure  $\mu_0$  on the 0<sup>th</sup> component  $\Pi_{(\mathcal{A})}^{0,+}$ .

As in the case of a stationary Markov chain, the meaning of matrix multiplication is given by a transition probability:

**Proposition 3.3.** *Given a stochastic matrix sequence  $(P_i)$  for  $i \geq 0$ , initial probability row vector  $\pi_0^t$  and corresponding nonstationary Markov chain  $\Pi_{(\mathcal{A})}^+$ , the  $ij^{\text{th}}$  matrix entry of  $P^{(k,n)} \equiv P_k P_{k+1} \cdots P_{n-1}$  gives the transition probability from state  $i$  to state  $j$  after a gap of time  $(n - k)$ , starting at time  $k$ . Hence  $\pi_n^t = \pi_0^t P^{(0,n)}$  gives the distribution of states at time  $n$ , for initial distribution  $\pi_0^t$ .*

*Proof.* The probability of being in state  $j$  at time  $(k + 1)$  given that we are in state  $i$  at time  $k$  is  $\mu_0([\cdot \cdot \cdot * i j]) / \mu_0([\cdot \cdot \cdot * i]) = \mu_k([\cdot i j]) / \mu_k([\cdot * i]) = \mu_k([\cdot i j]) / \mu_{k+1}([\cdot i])$ , and from the definition of  $\mu_0$  this is the  $ij^{\text{th}}$  matrix entry of  $P_k$ . For  $m > 1$ , the matrix product automatically sums over all the possible paths in the shift space, completing the proof.  $\square$

**3.1. Mixing and focussing.** First we consider the stationary case. Given a measure-preserving transformation  $T$  of a set  $X$  with  $\sigma$ -algebra  $\mathcal{B}$  and measure  $\mu$ , the dynamical system  $(X, \mathcal{B}, T, \mu)$  is *mixing* if and only if for all  $A, B \in \mathcal{B}$ ,  $\mu(T^{-m}A \cap B) \rightarrow \mu(A) \cdot \mu(B)$  as  $m \rightarrow \infty$ . For stationary Markov shifts, this specializes as follows. Given a  $(d \times d)$  stochastic matrix  $P$ , an invariant probability row vector  $\pi^t$  and the corresponding stationary Markov measure defined by (2) on  $\Pi_0^\infty \{0, \dots, d-1\}$ , and writing  $Q_{\pi^t}$  for the  $(d \times d)$  matrix all of whose rows are  $\pi^t$ , one has immediately from the definitions (see e.g. [Wal82], p. 51):

**Proposition 3.4.** *The Markov shift with invariant measure  $\mu(P, \pi^t)$  is mixing if and only if  $P^m \rightarrow Q_{\pi^t}$  as  $m \rightarrow \infty$ .*  $\square$

We also recall:

**Proposition 3.5.** *If  $P$  is a stochastic matrix which is primitive, then  $\mu(P, \pi^t)$  is mixing.*

*Proof.* Since  $P$  is stochastic and primitive, i.e. there exists  $k \geq 1$  with the entries of  $P^k$  all strictly positive, the collection of normalized positive row eigenvectors is a singleton  $\{\pi^t\}$  (by the Perron-Frobenius theorem), and since the images  $\Delta_k^t \equiv \Delta^t \cdot P^k$  nest down to  $\Delta_\infty^t \equiv \bigcap_{k=0}^{+\infty} \Delta_k^t = \{\pi^t\}$ , we know that for every  $\mathbf{v} \in \Delta$ ,  $\mathbf{v}^t P^m \rightarrow \pi^t$ . We choose e.g. for the  $(3 \times 3)$  case,  $\mathbf{v}^t$  to be the row vector  $[1 \ 0 \ 0]$ , noting that  $[1 \ 0 \ 0] P^m$  gives the first row of  $P^m$ . Hence  $P^m$  converges to the matrix  $Q_{\pi^t}$ . Then the previous proposition implies mixing for the Markov shift.  $\square$

*Remark 3.3.* We emphasize that if we have a  $(d \times d)$  stochastic matrix  $P$  and a noninvariant initial probability vector  $\pi_0^t$ , the resulting Markov measure on  $\Pi_0^\infty \{0, \dots, d-1\}$

will be nonstationary, i.e. will not define a stationary process, yet viewed instead as one component of a nonstationary shift space with  $\mathcal{A}_i = \{0, \dots, d-1\}$  for all  $i$ , we can have an invariant *sequence* of measures  $\mu_k$  on the different components  $\Pi_{(\mathcal{A})}^{k,+}$  of the total space  $\Pi_{(\mathcal{A})}^+$ . The components happen to be all canonically identified but now one thinks of them as different spaces, and the measures are invariant but not identical with respect to this identification. One may also have a sequence of measures which is not invariant, and indeed we encounter this below with the Parry eigenmeasures.

In the nonstationary setting, for alphabets  $(\mathcal{A}_i)_{i \geq 0}$ , let  $(P_i)_{i \geq 0}$  be a stochastic sequence, with  $\pi_0^t$  be an element of  $\Delta_0^t$ , and let  $\mu_0$  on  $\Pi_{(\mathcal{A})}^{0,+}$  be the Markov measure with this initial distribution as defined in (2). As above, the sequence of row vectors  $\pi_0^t, \pi_1^t = \pi_0^t P_0, \dots, \pi_m^t = \pi_0^t P^{(0,m)}$  is a normalized eigenvector sequence with eigenvalue one. For  $k \leq m$  we write  $Q_{\pi_m^t}^{(k,m)}$  for the  $(l_k \times l_m)$  matrix all of whose rows are  $\pi_m^t$ .

Recall from §2 the definitions of the algebra  $\mathcal{B}_k^m$  and  $\sigma$ - algebra  $\mathcal{B}_k^{+\infty}$ , generated by cylinder sets.

We say:

**Definition 3.3.** *The nonstationary Markov chain (i.e. the Markov measure)  $(\Pi_{(\mathcal{A})}^{0,+}, \mu_0)$  is mixing if and only if for any fixed  $k \geq 0$ , given  $\varepsilon > 0$ , for  $m > k$  sufficiently large we have that, for every  $A \in \mathcal{B}_0^k$  and  $B \in \mathcal{B}_m^\infty$ , then  $e^{-\varepsilon} \mu_0(A) \mu_0(B) \leq \mu_0(A \cap B) \leq e^{+\varepsilon} \mu_0(A) \mu_0(B)$ . We write this as  $\mu_0(A \cap B) = e^{\pm\varepsilon} \mu_0(A) \mu_0(B)$ .*

We define a metric on the non-negative  $(m \times n)$  matrices:

$$d(A, B) = \sup_i d_C(A_{i*}, B_{i*}) \quad (3)$$

where  $A_{i*}$  indicates the  $i^{\text{th}}$  row of  $A$ , and  $d_C$  is the projective metric on the standard cone  $C = \mathbb{R}^{n+}$  (see Appendix 6), and note that the definition of focussing given in the introduction for non-negative integer matrices also makes sense for non-negative real entries.

We also need:

**Definition 3.4.** *A nonstationary Markov chain is strictly positive if and only if the probability row vectors  $\pi_m^t$  are strictly positive, or equivalently if and only if each state (i.e. each 1- cylinder set  $[. * * * \dots * s]$ ) has strictly positive probability, equivalently iff the initial vector  $\pi^t$  is strictly positive and no  $P_i$  has a column which is identically zero.*

**Lemma 3.6.** *The following are equivalent, for a strictly positive nonstationary Markov shift  $\Pi_{(\mathcal{A})}^+$  with an invariant Markov measure sequence  $(\mu_i)_{i \geq 0}$ :*

- (i) *The mapping family  $(\Pi_{(\mathcal{A})}^+, (\mu), \sigma)$  is mixing.*
- (ii) *The nonstationary Markov chain  $(\Pi_{(\mathcal{A})}^{0,+}, \mu_0)$  is mixing.*
- (iii) *For  $k$  fixed, given  $\varepsilon > 0$ , then for  $m > k$  sufficiently large we have*

$$d(P^{(k,m)}, Q_{\pi_m^t}^{(k,m)}) < \varepsilon.$$

- (iv) *The sequence  $(P_i)$  is focussing.*

*Proof.* We prove (i)  $\iff$  (ii). To verify the condition in Def. 3.3, it is sufficient to check for thin cylinder sets  $A = [.x_0 \dots x_k]$  and  $B = [w_m \dots w_l] = [. * \dots * w_m \dots w_l]$ , and without loss of generality it can be assumed that these have measure  $> 0$ . In the full *nsft* component  $\Pi_{(\mathcal{A})}^{k,+}$  the ball of radius  $\varepsilon > 0$  about a point  $x$  is the thin cylinder set  $[.x_k x_{k+1} \dots x_l]$  with  $l$  such that  $w(k, l)^{-1} < \varepsilon < w(k, l-1)^{-1}$ , where  $w(k, l)$  is the number of words of length  $(l-1)$ , i.e.  $w(k, l) = l_k \dots l_{k+l}$ , see Def. 1. So it suffices in both cases to give the proofs for thin cylinder sets.

Therefore the mapping family  $(\Pi_{(\mathcal{A})}^+, (\mu_k)_{k \geq 0}, \sigma)$  is mixing if and only if for chosen  $k, l_1$  and  $\varepsilon > 0$  there exists  $M > 0$  such that for  $m > M$  then for thin cylinder sets  $A = [.x_k x_{k+1} \dots x_{l_1}] \subseteq \Pi_{(\mathcal{A})}^{k,+}$  and  $B = [w_m w_{m+1} \dots w_{l_2}] \subseteq \Pi_{(\mathcal{A})}^{m,+}$  both with measure  $> 0$  we have for any  $l_2 > m$

$$\mu_k(A \cap \sigma^{-m+k}(B)) = e^{\pm\varepsilon} \mu_k(A) \mu_m(B). \quad (4)$$

By the Markov property the same proof works for any  $l_2$ . Now  $A \cap \sigma^{-m+k}(B) = [.x_k x_{k+1} \dots x_{l_1} * \dots * w_m w_{m+1} \dots w_{l_2}] \subseteq \Pi_{(\mathcal{A})}^{k,+}$ , so taking this condition for  $k = 0$  gives (ii), mixing for the measure  $\mu_0$ . Conversely suppose  $\mu_0$  is mixing; then we know statement (4) only for the 0<sup>th</sup> component  $(\Pi_{(\mathcal{A})}^{0,+}, \mu_0)$ , but this implies the statement for component  $k$  as well:  $\mu_k([.x_k x_{k+1} \dots x_{l_1}]) = \mu_0[. * \dots * x_k x_{k+1} \dots x_{l_1}]$  is a union of thin cylinders of length  $(k + l_1)$  in  $\Pi_{(\mathcal{A})}^{0,+}$ , and one just takes the sum of the estimates for each of these.

We next show (ii)  $\iff$  (iii). Given  $A = [.x_0 \dots x_k]$  and  $B = [x_m \dots x_n] \equiv [. * \dots * x_m \dots x_n]$ , then

$$\mu_0(A \cap B) = (\pi_0^t)_{x_0} (P_0)_{x_0 x_1} \dots (P_{k-1})_{x_{k-1} x_k} \cdot P_{x_k x_m}^{(k,m)} (P_m)_{x_m x_{m+1}} \dots (P_{n-1})_{x_{n-1} x_n}$$

while

$$\mu_0(A) \mu_0(B) = (\pi_0^t)_{x_0} (P_0)_{x_0 x_1} \dots (P_{k-1})_{x_{k-1} x_k} \cdot (\pi_m)_{x_m} (P_m)_{x_m x_{m+1}} \dots (P_{n-1})_{x_{n-1} x_n}.$$

Since without loss of generality  $A$  and  $B$  have measure  $> 0$ , all the above factors must be nonzero, and so can be cancelled. Thus  $\mu_0(A \cap B) = e^{\pm\varepsilon} \mu_0(A) \mu_0(B)$  if and only if  $P_{x_k x_m}^{(k,m)} = e^{\pm\varepsilon} (\pi_m)_{x_m}$ . Writing  $\mathbf{v}$  for a row of  $P^{(k,m)}$  and  $\mathbf{w}$  for  $\pi_m$ , then if this holds for each  $A$  and  $B$ , equivalently  $e^{-\varepsilon} \leq v_i/w_i \leq e^{+\varepsilon}$  for each  $i$  with  $1 \leq i \leq l_m$  (here we use the fact that states have measure  $> 0$ ), and then applying Cor. 6.11 we have that  $d_C(\mathbf{v}, \mathbf{w}) < 2\varepsilon$ . On the other hand if  $d_C(\mathbf{v}, \mathbf{w}) < \varepsilon$ , then  $e^{-\varepsilon} \leq v_i/w_i \leq e^{+\varepsilon}$ . This proves that (ii) and (iii) are equivalent.

Next we show (iv)  $\implies$  (iii). Let  $\mathbf{e}_i^t$  denote the row vector with 1 in the  $i^{\text{th}}$  coordinate, 0 elsewhere; then  $\mathbf{e}_i^t P^{(k,m)} \in \Delta_k^t P^{(k,m)}$ . This is the  $i^{\text{th}}$  row of  $P^{(k,m)}$ . And  $\pi_0^t P^{(0,k)} \in \Delta_0^t P^{(0,k)} \subseteq \Delta_k^t$ , so  $\pi_m^t = \pi_0^t P^{(0,m)} \in \Delta_k^t P^{(k,m)}$ .

By (iv), the projective diameter of  $\Delta_k^t P^{(k,m)}$  is  $\leq \varepsilon$  for  $m$  large. Hence the distance of any row of  $P^{(k,m)}$  and  $\pi_m^t$  (the unique row of  $Q_{\pi_m^t}^{(k,m)}$ ) is  $\leq \varepsilon$ , so

$$d(P^{(k,m)}, Q_{\pi_m^t}^{(k,m)}) \leq \varepsilon,$$

giving (iii).

Lastly, assuming (iii), we have that each row of  $P^{(k,m)}$  is  $\varepsilon$ -close to the unique row of  $Q_{\pi_m^t}^{(k,m)}$ ; then by the triangle inequality the rows are  $2\varepsilon$ -close to each other, so by Lemma 6.7 the diameter of the image  $\Delta_k^t P^{(k,m)}$  is also less than  $2\varepsilon$ , giving (iv).  $\square$

#### 4. NONSTATIONARY PARRY MEASURES AND UNIQUE ERGODICITY

Now we return from the general setting of nonstationary Markov measures to nonstationary subshifts of finite type; from a measure theoretic point of view one is then just restricting to the support of the Markov measure on each component.

From the ergodic theory point of view, however, this becomes quite interesting as one is studying all Markov measures with this support. Thus for a fixed primitive matrix  $L$ , Parry [Par64] p. 61 presents a unique invariant measure  $\mu$  which satisfies a strong equidistribution property, guaranteeing that it is the unique measure of maximal entropy for the *sft*, see [AW70]. (As noted in the Introduction, we follow standard ergodic theory practice in calling this the Parry measure, though the formula was first found by Shannon in a coding theory setting).

Part of the importance of  $\mu$  is that there is a second measure  $\nu$ , equivalent to  $\mu$  but in general not shift-invariant; the measure  $\nu$  satisfies two other invariance properties: –it is an eigenmeasure for the dual Ruelle operator  $L_\varphi^*$  with potential  $\varphi \equiv 0$ ; for that reason we call it the Parry eigenmeasure though in fact we don't need this interpretation here; and –it is invariant for the finite coordinate changes  $\mathcal{FC}$  and nonatomic and hence (when normalized) gives an invariant probability measure for the adic transformation and extended transformations ((i) of Prop. 2.11).

For extending these ideas to the nonstationary case, we shall make the twin assumptions that the sequence  $(L_i)_{i \geq 0}$  of  $(l_i \times l_{i+1})$  0 – 1 matrices be *reduced* and *primitive*; these properties will guarantee the strictly positive row and column eigenvector sequences needed for the construction, as we shall explain.

First we give another perspective on what it means for the matrices to be reduced.

**Definition 4.1.** *We denote the strictly positive row and column vectors of  $\mathbb{R}^m$  by  $\mathring{C}^+$ ,  $\mathring{R}^+$ . For dimension  $\geq 2$  these are the interior of the positive cones  $C^+$ ,  $R^+$ ; for dimension 1 this is  $\mathring{C}^+ = (0, +\infty) \subseteq [0, +\infty) = C^+$ .*

We then have:

**Lemma 4.1.** *An  $(m \times n)$  real matrix  $M$  has no identically zero rows if and only if it maps  $\mathring{C}^+(\mathbb{R}^m)$  to  $\mathring{C}^+(\mathbb{R}^n)$ , and has no identically zero columns if and only if it maps  $\mathring{R}^+(\mathbb{R}^n)$  to  $\mathring{R}^+(\mathbb{R}^m)$ .  $\square$*

We will find a special stochastic sequence  $(P_i)_{i \geq 0}$  which is *compatible* with  $(L_i)_{i \geq 0}$ , i.e. it has the same dimensions and  $(P_{ij} = 0) \iff (L_{ij} = 0)$ . As in the stationary case, this will guarantee that any Markov measure with transition matrices  $(P_i)$  will have support all of the *nsft*  $\Sigma_{(L)}^+$ .

We let  $\widehat{\Omega}_{(L)}$  denote the collection of all strictly positive column eigenvector sequences  $(\widehat{\mathbf{w}})$  with eigenvalue one, and  $\widehat{\Omega}_{(L)}^t$  for row vectors, so

$$\widehat{\Omega}_{(L)} = \{(\widehat{\mathbf{w}}) = (\widehat{\mathbf{w}}_0 \widehat{\mathbf{w}}_1 \dots) \text{ with } \widehat{\mathbf{w}}_i = L_i \widehat{\mathbf{w}}_{i+1}, \text{ such that for all } i \geq 0, \widehat{\mathbf{w}}_i \in \mathring{C}_i^+\}$$

and

$$\widehat{\Omega}_{(L)}^t = \{(\widehat{\mathbf{v}}^t) = (\widehat{\mathbf{v}}_0^t \widehat{\mathbf{v}}_1^t \dots) \text{ with } \widehat{\mathbf{v}}_i^t L_i = \widehat{\mathbf{v}}_{i+1}^t \text{ and } \widehat{\mathbf{v}}_i^t \in \mathring{R}_i^+\}$$

**Lemma 4.2.** *Let  $(L)_{i \geq 0}$  be a reduced and primitive  $(l_i \times l_{i+1})$  0–1 matrix sequence. Then the spaces  $\widehat{\Omega}_{(L)}$  and  $\widehat{\Omega}_{(L)}^t$  are nonempty. A sequence in  $\widehat{\Omega}_{(L)}^t$  is, up to multiplication by positive constants, determined by choosing a strictly positive first element  $\widehat{\mathbf{v}}_0^t$ .*

*Proof.* For row vectors we simply begin with any  $\widehat{\mathbf{v}}_0^t \in \mathring{R}_0^+$  and apply the positive and reduced matrices  $L_i$ ; by Lemma 4.1 the images are all strictly positive as well.

For column vectors there are two reasons why we cannot apply the previous argument: the matrices  $L_i^{-1}$  may not be positive, and they may not be invertible. So we proceed as follows.

We write for  $k, n \geq 0$ :

$$C_{(k,n)}^+ = L_k L_{k+1} \cdots L_{n-1} C_n^+ = L^{(k,n)} C_n^+.$$

These are nested (this contrasts with the row case- see Introduction!):

$$C_k^+ \supseteq C_{(k,k+1)}^+ \supseteq \cdots$$

The intersection  $C_{(0,+\infty)}^+$  of the nested closed cones is nonempty, by compactness, and by primitivity all its nonzero elements are strictly positive. Note that for each  $k$ ,  $L_k C_{(k+1,+\infty)}^+ = C_{(k,+\infty)}^+$ .

We begin with some nonzero (hence strictly positive)  $\widehat{\mathbf{w}}_0$  in  $C_{(0,+\infty)}^+$ ; proceeding inductively, given  $\widehat{\mathbf{w}}_k \in C_{(k,+\infty)}^+$  we choose  $\widehat{\mathbf{w}}_{k+1} \in C_{(k+1,+\infty)}^+$  (necessarily again strictly positive) such that  $L_k \widehat{\mathbf{w}}_{k+1} = \widehat{\mathbf{w}}_k$ . This constructive procedure yields the entire collection  $\widehat{\Omega}_{(L)}$ .  $\square$

We next project each element of the sequence  $(\widehat{\mathbf{w}})$  to the unit simplex, with  $\mathbf{w}_i = \widehat{\mathbf{w}}_i / \|\widehat{\mathbf{w}}_i\| = \text{Proj}(\widehat{\mathbf{w}}_i) \in \Delta_i$ , and write  $\Omega_{(L)}$  for the collection of all such normalized sequences  $(\mathbf{w})$ . We then normalize the sequence of row vectors in a different way, which depends on this choice of  $(\widehat{\mathbf{w}})$ : we define  $(\mathbf{v}^t) = (\mathbf{v}_i^t)$  by  $\mathbf{v}_i^t = \widehat{\mathbf{v}}_i^t / (\widehat{\mathbf{v}}_i^t \mathbf{w}_i)$  for  $i \in \mathbb{N}$ . Since  $\mathbf{w}_i$  is strictly positive, we are not dividing by zero. Our normalization is chosen so the inner product of  $\mathbf{v}_i^t$  with  $\mathbf{w}_i$  is 1.

We define real numbers  $\lambda_i$  by  $\lambda_i = \lambda_i(\widehat{\mathbf{w}}) = \|\widehat{\mathbf{w}}_i\| / \|\widehat{\mathbf{w}}_{i+1}\| \geq 1$ , and we have for each  $i \geq 0$ ,

$$L_i \mathbf{w}_{i+1} = \lambda_i \mathbf{w}_i.$$

**Lemma 4.3.** *Let  $(\widehat{\mathbf{w}}_i)_{i \geq 0} \in \widehat{\Omega}_{(L)}$  and  $(\widehat{\mathbf{v}}^t) \in \widehat{\Omega}_{(L)}^t$  be strictly positive column and row eigenvector sequences. and let  $(\mathbf{w}_i)_{i \geq 0}, (\mathbf{v}_i^t)_{i \geq 0}$  be the corresponding normalized sequences as defined above. Then  $(\mathbf{v}_i^t)_{i \geq 0}$  has the same eigenvalues as  $(\mathbf{w}_i)$ .*



*Proof.* Defining  $\tilde{\lambda}_i$  by  $\mathbf{v}_i^t L_i = \tilde{\lambda}_i \mathbf{v}_{i+1}^t$ , then we have  $\tilde{\lambda}_i = \lambda_i$ , because:

$$\tilde{\lambda}_i = \tilde{\lambda}_i (\mathbf{v}_{i+1}^t \mathbf{w}_{i+1}) = (\mathbf{v}_i^t L_i) \mathbf{w}_{i+1} = \mathbf{v}_i^t (L_i \mathbf{w}_{i+1}) = \mathbf{v}_i^t (\lambda_i \mathbf{w}_i) = \lambda_i.$$

□

From a choice  $\widehat{\mathbf{w}}_i, \widehat{\mathbf{v}}_i^t$ , we now define a nonstationary Parry measure.

First we define row vectors  $(\boldsymbol{\pi}^t) = (\boldsymbol{\pi}_i^t)_{i \geq 0}$  by  $(\boldsymbol{\pi}_i^t)_k = (\mathbf{v}_i)_k (\mathbf{w}_i)_k$  where  $k$  is the index of a letter in the  $i^{\text{th}}$  alphabet. By the normalization  $\mathbf{v}_i^t \mathbf{w}_i = 1$ ,  $\boldsymbol{\pi}_i^t$  is an element of  $\Delta_i^t$ . It is strictly positive since both  $\mathbf{v}_i^t$  and  $\mathbf{w}_i$  are.

Next we define a matrix sequence  $(P_i)_{i \geq 0}$  by

$$P_i = \frac{1}{\lambda_i} W_i^{-1} L_i W_{i+1}, \quad (5)$$

where  $W_i$  is the  $(l_i \times l_i)$  diagonal matrix with the entries of the vector  $\mathbf{w}_i$  on the diagonal. This is a matrix interpretation of the formula of Parry for the matrix entries, usually written (for the stationary case) as

$$P_{ij} = \frac{L_{ij} w_j}{\lambda w_i} \quad (6)$$

[Par64], [AW70]. Writing  $\mathbf{1}_i$  for the column vector with  $l_i$  entries all equal to 1, we have, analogous to the stationary case (here the matrix notation of (5) comes in handy):

$$P_i \mathbf{1}_{i+1} = \mathbf{1}_i, \quad (7)$$

(so  $P_i$  is stochastic) and

$$\boldsymbol{\pi}_i^t P_i = \boldsymbol{\pi}_{i+1}^t. \quad (8)$$

These are right and left eigenvector sequences with constant eigenvalue 1.

For  $k, m \geq 0$  with  $k < m$ , we set  $\lambda_{(\mathbf{w})}^{(k,m)} = \prod_{i=k}^{m-1} \lambda_i(\mathbf{w})$ . We then define a probability measure  $\mu_k$  on  $\Sigma_{(L)}^{k,+}$ : the measure of a cylinder set is

$$\mu_k([\cdot x_k \dots x_m]) \equiv (\boldsymbol{\pi}_k^t)_{x_k} (P_k)_{x_k x_{k+1}} \cdots (P_{m-1})_{x_{m-1} x_m} = (1/\lambda_{(\mathbf{w})}^{(k,m)}) (\mathbf{v}_k)_{x_k} (\mathbf{w}_m)_{x_m}. \quad (9)$$

This gives a measure which is positive on open sets of  $\Sigma_{(L)}^{k,+}$ , since each  $P_i$  is compatible with  $L_i$  and is stochastic, in other words it has full support. We call any such measure a *Parry measure*. Note that (9) gives a strong form of *equidistribution*, as all cylinders in  $\mathbb{C}_m^k$  have nearly the same measure; this is just like in [AW70] for the stationary case.

The *Parry eigenmeasure*  $\nu_k = \nu_k(\mathbf{w})$  on  $\Sigma_{(L)}^{k,+}$  is then given by:

$$\nu_k([\cdot x_k \dots x_m]) = \mu_k([\cdot x_k \dots x_m]) / (\mathbf{v}_k)_{x_k} = (\mathbf{w}_k^t)_{x_k} (P_k)_{x_k x_{k+1}} \cdots (P_{m-1})_{x_{m-1} x_m}. \quad (10)$$

In terms of the normalized and non-normalized eigenvector sequences  $(\mathbf{w})$  and  $(\widehat{\mathbf{w}})$  for  $(L_i)$  we then have for  $\lambda_{(\mathbf{w})}^{(0,m)} = \lambda_{(\widehat{\mathbf{w}})}^{(0,m)}$ :

$$\nu_0([\cdot x_0 \dots x_m]) = (1/\lambda_{(\mathbf{w})}^{(0,m)}) (\mathbf{w}_m)_{x_m} = (\widehat{\mathbf{w}}_m)_{x_m} \cdot \|\widehat{\mathbf{w}}_0\|^{-1}. \quad (11)$$

We indicate the dependence of these measures on  $(\widehat{\mathbf{w}})$  and  $(\widehat{\mathbf{v}})^t$  (or equivalently on  $(\mathbf{w}), (\mathbf{v})^t$ ) by writing:  $(\boldsymbol{\pi}^t)_{\mathbf{w}, \mathbf{v}^t}$ ,  $(P)_{\mathbf{w}}$ , and  $\nu = \nu_{\mathbf{w}}$ . We emphasize that the sequence

$(\widehat{\mathbf{v}}_i^t)_{i \geq 0}$  actually is determined by its first element  $\widehat{\mathbf{v}}_0^t$  and that the eigenmeasures are not necessarily normalized.

*Remark 4.1.* From equations (9) and (10), both equations define sequences of Markov measures, though in general from (iii) of Prop. 3.2 the sequence of eigenmeasures is not invariant and so does not give a nonstationary Markov shift.

Now we consider the connection between focussing and mixing for these measures.

**Lemma 4.4.** *Let  $(L)_{i \geq 0}$  be a sequence of  $(l_i \times l_{i+1})$  reduced and primitive 0 – 1 matrices. The following properties are equivalent:*

- (a) *For any choice of  $(\widehat{\mathbf{w}}) \in \widehat{\Omega}_{(L)}$  and  $\widehat{\mathbf{v}}_0^t \in \widehat{\Omega}_{(L)}^t$ , the nonstationary Markov chain  $(\Sigma_L^{0,+}, \mu_{\mathbf{w}, \mathbf{v}})$  is mixing.*
- (b)  *$(P_i)_{\mathbf{w}}$  is focussing.*
- (c)  *$(L_i)$  is focussing.*
- (d) *There is a unique normalized Parry eigenmeasure.*

*Proof.* (a  $\iff$  b) is proved in Lemma 3.6.

(b  $\iff$  c):

We fix a choice of  $(\mathbf{w})$ , and write  $(P) = (P)_{(\mathbf{w})}$ . We have for  $\lambda^{(k,m)} = \lambda_{(\mathbf{w})}^{(k,m)}$ :

$$P^{(k,m)} = (1/\lambda^{k,m})W_k^{-1}L^{(k,m)}W_m.$$

Now for  $R_k^+ = {}^t\mathbb{R}^{l_k,+}$  the positive cone, we compare  $R_k^+P^{(k,m)}$  and  $R_k^+L^{(k,m)}$ . Since  $W_k^{-1}$  is diagonal, it is a bijection on the positive cone, so  $R_k^+W_k^{-1} = R_k^+$ . And  $W_m$  is an isometry of  $R_m^+$  by Corollary 6.12. Therefore the projective diameters of  $R_k^+P^{(k,m)}$  and  $R_k^+L^{(k,m)}$  are equal.

Finally to prove (c)  $\iff$  (d), from the above construction, a sequence  $(\mathbf{w}) \in \Omega_{(L)}$  defines an eigenmeasure, while from (10) the measure determines the eigenvector sequence.  $\square$

**4.1. Nonstationary unique ergodicity and the Bowen-Marcus lemma.** Given a 0 – 1 matrix sequence  $(L_i)_{i \geq 0}$ , we say a measure  $m$  on  $\Sigma_{(L)}^{0,+}$  has the *Bowen-Marcus property* if and only if  $m_t(s) = m([.x_0 \dots x_{t-1}])$  for  $x_{t-1} = s \in \mathcal{A}_{t-1}$  is well-defined; that is, the measure of a thin cylinder set depends only on the length and the last letter.

By (11), any Parry eigenmeasure  $\nu_{\mathbf{w}}$  for  $(\mathbf{w}) \in \Omega_{(L)}$  has this property. Given such a  $\nu_{\mathbf{w}}$ , we choose and fix some vector  $\mathbf{v}^t = \mathbf{v}_0^t \in \Delta_0^t$ , defining the measure  $\mu = \mu_{\mathbf{w}, \mathbf{v}^t}$ . Here is our version of Lemma 2.4 in [BM77]; we use ideas from the Bowen-Marcus proof together with Cor. 6.14 of Appendix 6.

**Lemma 4.5.** *Let  $(L)_{i \geq 0}$  be a reduced and primitive  $(l_i \times l_{i+1})$  0 – 1 matrix sequence. Assume for some  $(\widehat{\mathbf{w}}) \in \widehat{\Omega}_{(L)}$  and  $(\widehat{\mathbf{v}})^t \in \widehat{\Omega}_{(L)}^t$  the measure  $\mu_{(\mathbf{w}, \mathbf{v}^t)}$  on the nsft  $\Sigma_{(L)}^+$  is mixing for the nonstationary shift map. Then if  $m$  on  $\Sigma_{(L)}^{0,+}$  has the Bowen-Marcus property,  $m$  is a constant multiple of  $\nu = \nu_{\mathbf{w}}$ . Furthermore,  $\nu$  is the unique eigenmeasure sequence (up to multiples) and  $\Omega_{(L)}$  is the singleton  $\{(\mathbf{w})\}$ .*

*Proof.* We shall show that there exists  $\gamma > 0$  such that for any  $a \in \mathcal{A}_0$ ,

$$m[a] = \gamma \cdot \nu[a];$$

as we see later, this same proof will work for any chosen cylinder set  $[.a_0 \dots a_l]$ , and with the same constant  $\gamma$ , proving the lemma.

We write  $[. * * \dots * s]$  for the cylinder set of length  $(t + 1)$  ending in symbol  $s$ , so this is a union of all allowed thin cylinders  $[.b_0 b_1 \dots b_t = s]$ .

First we define a number  $\gamma_{t,s}$  by

$$m([. * * \dots * s]) = \gamma_{t,s} \nu([. * * \dots * s])$$

We shall prove that  $\gamma = \gamma_{t,s}$  does not depend on  $t$  or  $s$ .

Now the same factor  $\gamma_{t,s}$  works for the thin cylinders  $[.b_0 b_1 \dots b_t = s]$ . This is because by assumption both  $m$  and  $\nu$  have the Bowen-Marcus property.

Let us fix the choice of the symbol  $a$ . Then also  $m([.a * \dots * s]) = \gamma_{t,s} \nu([.a * \dots * s])$  since this is a union of thin cylinders.

By the assumption that the measure  $\mu$  is mixing, we have from (10) that for  $t$  sufficiently large, for every  $a \in \mathcal{A}_0$ ,  $s \in \mathcal{A}_t$

$$\nu([.a * \dots * s]) = \mu([.a * \dots * s]) / \mathbf{v}_a = (e^{\pm \varepsilon}) \mu[a] \cdot \mu[. * * \dots * s] / \mathbf{v}_a = (e^{\pm \varepsilon}) \nu[a] \cdot \mu[. * * \dots * s]$$

and so

$$m[a] = \sum_{s \in \mathcal{A}_t} m([.a * \dots * s]) = \sum_{s \in \mathcal{A}_t} \gamma_{t,s} \nu([.a * \dots * s]) = (e^{\pm \varepsilon}) \sum_{s \in \mathcal{A}_t} \gamma_{t,s} \nu[a] \mu([. * * \dots * s])$$

hence

$$(e^{\pm \varepsilon}) m[a] / \nu[a] = \sum_{s \in \mathcal{A}_t} \gamma_{t,s} \cdot \mu([. * * \dots * s]).$$

This holds for each  $t \geq 0$  sufficiently large.

Therefore

$$\lim_{t \rightarrow \infty} \sum_{s \in \mathcal{A}_t} \gamma_{t,s} \cdot \mu([. * * \dots * s]) = \gamma$$

exists, and equals

$$m[a] / \nu[a].$$

From this we know that

$$m[a] = \gamma \cdot \nu[a].$$

We note that the constant  $\gamma$  does not depend on  $s, t$  or  $a$ . Indeed, if we had started with any other cylinder set  $[.a_0 \dots a_k]$  in place of  $[a]$ , we would have ended up with the same limit and hence with the equation:

$$m([.a_0 \dots a_k]) = \gamma \cdot \nu([.a_0 \dots a_k]).$$

This proves that for all sets in the Borel  $\sigma$ -algebra, the same is true, so we are done.

Lastly, mixing implies that  $\Omega(L)$  is the singleton  $\{(\mathbf{w})\}$  by (d) of Lemma 4.4.  $\square$

We can then conclude:

**Theorem 4.6.** (Unique ergodicity for the case of single-edged diagrams) Let  $(L_i)_{i \geq 0}$  be a reduced and primitive sequence of  $(l_i \times l_{i+1})$  0 – 1 matrices, and let  $\mathcal{O}$  be a stable order. Assuming the sequence is focussing, then the adic transformation  $T_{\mathcal{O}} : \Sigma_{(L)}^{0,+} \setminus \mathcal{NS} \rightarrow \Sigma_{(L)}^{0,+} \setminus \mathcal{NP}$  and the group of finite coordinate changes  $\mathcal{FC}(\Sigma_{(L)}^k)$  and also any extension  $T_{\mathcal{O},f}$  are uniquely ergodic, with  $m = \nu/\nu(\Sigma_{(L)}^{0,+})$  the unique invariant non-atomic probability measure.

*Proof.* By Lemma 2.5 primitivity implies left nontriviality, so part (ii) of Proposition 2.11 tells us that it is enough to check unique ergodicity for  $\mathcal{FC}$ .

The eigenmeasure  $\nu$  is invariant for  $\mathcal{FC}$ , since by construction any two cylinder sets  $[.x_0x_1 \dots x_{t-2}s]$  and  $[.y_0y_1 \dots y_{t-2}s]$  of length  $t$  have the same measure. Any other  $\mathcal{FC}$ -invariant nonatomic measure  $m$  satisfies this property (the Bowen-Marcus property), and by Lemma 4.5 this is the same up to a constant.  $\square$

**4.2. Focussing and the Perron-Frobenius conditions.** We next show the equivalence to the focussing condition of the three Perron-Frobenius conditions of the introduction, when the sequence is primitive; focussing and geometric Perron-Frobenius on the other hand are always equivalent.

We note that the definitions of focussing or of topological Perron-Frobenius could instead be given for simplices. Writing  $\Delta_{k(+n)} = \text{Proj}(C_{k(+n)}^+)$ , by compactness the intersection  $\Delta_{k(+\infty)}$  is nonempty for any fixed  $k$  and therefore so is the intersection  $C_{k(+\infty)}^+$  of the cones;  $C_{k(+\infty)}^+$  is one-dimensional if and only if  $\Delta_{k(+\infty)}$  is a singleton, while the projective diameter of  $\Delta_k^t M_k \cdots M_m$  or of  $\text{Proj}(\Delta_k^t M_k \cdots M_m)$  is equal to that of  $R_k^+ M_k \cdots M_m$ .

First we prove:

**Lemma 4.7.** *A sequence  $(M)_{i \geq 0}$  of  $(l_i \times l_{i+1})$  non-negative real matrices geometrically Perron-Frobenius if and only if it is focussing.*

*Proof.* This is proved in part (ii) of Corollary 6.14; an assumption that the sequence be primitive or reduced is not necessary here.  $\square$

It is clear that in the primitive case the geometrical and topological Perron-Frobenius conditions are equivalent: the images  $M_k \cdots M_m \Delta_m$  are nested decreasing sets, and the limiting projective diameter is zero if and only if the intersection is a singleton in the interior of  $\Delta_k$ .

We then come to the eigenvector version of the Perron-Frobenius condition. Recall that  $\Omega_{(L)}$  denotes the collection of all normalized strictly positive eigenvector sequences for  $(L_i)$ . We make the same definition for a real matrix sequence, and then have another way of stating the definition given in the introduction:

**Definition 4.2.** *A sequence  $(M)_{i \geq 0}$  of  $(l_i \times l_{i+1})$  non-negative real matrices is eigenvector Perron-Frobenius iff  $\Omega_{(L)}$  is a singleton.*

The following is now clear:

**Lemma 4.8.** *For  $(M)_{i \geq 0}$  ( $l_i \times l_{i+1}$ ) non-negative real matrices, with the assumption of primitivity, the geometric, topological and eigenvector Perron-Frobenius properties are equivalent.  $\square$*

**Corollary 4.9.** *If this sequence  $(M)_{i \geq 0}$  is primitive and  $T_{\mathcal{O}}$  (or equivalently  $\mathcal{FC}$ ) is uniquely ergodic, then the sequence is focussing.*

*Proof.* Assuming primitivity, if the Perron-Frobenius property does not hold, then there are at least two distinct right eigenvector sequences  $(\mathbf{w}) \in \Omega_{(L)}$ ; from the construction of §4 we have at least two Parry eigenmeasures  $\nu = \nu_{\mathbf{w}}$ , and each has the Bowen-Marcus property and hence is invariant for  $\mathcal{FC}$ . So unique ergodicity (for any of the three, using again primitivity and Prop. 2.11 (iii)) does imply Perron-Frobenius of the matrix sequence.  $\square$

## 5. ADIC TRANSFORMATIONS ON EDGE SHIFTS

In this section we extend Theorem 4.6 to multiple-edged Bratteli diagrams, which allow for more than a single edge connecting two vertices; this is the usual setting for adic transformations ([TGS95]).

Two reasons for bringing in multiple-edged diagrams are that the edge notation is generally more compact, as matrices with smaller dimensions can be used, and that edge shifts arise naturally from vertex shifts by the operation of telescoping (or *gathering*) the diagram. See Figs. 5,6 of [AF05].

We begin as before with a sequence of alphabets  $\mathcal{A}_k$  with  $\#\mathcal{A}_k = l_k$ , but now the  $(l_k \times l_{k+1})$  matrices  $(M_k)_{k \geq 0}$  which define the diagram can have non-negative integer entries; we then draw the diagram with vertices  $\mathcal{A}_k$  at level  $k$ , with the  $ij^{\text{th}}$  entry of  $M_k$  specifying the number of edges from symbol  $i$  at level  $k$  to symbol  $j$  at level  $k+1$ .

We write  $\mathcal{E}_k$  for the set of edges from  $\mathcal{A}_k$  to  $\mathcal{A}_{k+1}$ , oriented towards the future. For  $e_k \in \mathcal{E}_k$  we let  $e_k^- \in \mathcal{A}_k$  denote the tail of its arrow and  $e_k^+ \in \mathcal{A}_{k+1}$  its head. Writing  $\hat{l}_k = \#\mathcal{E}_k$  we then form a  $0-1$ ,  $(\hat{l}_k \times \hat{l}_{k+1})$  matrix sequence  $(L_k)_{k \geq 0}$  with  $(L_k)_{ij} = 1$  if and only if the edge  $e_{k+1}$  labelled  $j$  in  $\mathcal{E}_{k+1}$  can follow the edge  $e_k$  labelled  $i$  in  $\mathcal{E}_k$ , i.e. if and only if  $e_k^+ = e_{k+1}^-$ .

As usual,  $\Sigma_{(L)}^{0,+}$  denotes all the allowed vertex paths in the Bratteli diagram of  $(L_k)_{k \geq 0}$ , but now we denote by  $\Sigma_{(M)}^{0,+}$  the collection of all the allowed *edge* paths in the diagram of  $(M_k)_{k \geq 0}$ . Note that  $\Sigma_{(L)}^{0,+}$  and  $\Sigma_{(M)}^{0,+}$  correspond bijectively. We next discuss order. On the multiple-edged diagram of  $(M_k)_{k \geq 0}$ , a stable order  $\mathcal{O}$  is defined just as before, and from this we get a lexicographic order on  $W^s(\mathbf{e})$  for  $\mathbf{e} \in \Sigma_{(M)}^{0,+}$ , that is, for a collection of edge paths which eventually agree. The reason for this is that given two edge paths  $e_0 e_1 \dots e_k e_{k+1}$  and  $\tilde{e}_0 \tilde{e}_1 \dots \tilde{e}_k \tilde{e}_{k+1}$  that agree from  $e_{k+1}$  on, then the incoming edges to the symbol  $e_{k+1}^- \in \mathcal{A}_{k+1}$  are ordered by the diagram order  $\mathcal{O}$ . This means that the stable order on the diagram for  $(M_k)_{k \geq 0}$  induces a stable order for the diagram of  $(L_k)_{k \geq 0}$ . In fact this can be thought of as a vertex order, in the following way. Writing  $\mathcal{E}_k^j = \{e_k \in \mathcal{E}_k : e_k^+ = j\}$ , these sets partition a subset of  $\mathcal{E}_k$  (which is all of  $\mathcal{E}_k$  if the diagram is reduced). Each of the subsets  $\mathcal{E}_k^j$  is ordered by the edge order  $\mathcal{O}$ ; we extend this to a linear order on all of  $\mathcal{E}_k$  in an arbitrary way which respects the orders on these (disjoint) subsets. This gives a vertex order as desired.

Note that of course orders on the diagram of  $(L_k)_{k \geq 0}$  are not necessarily of this type. Therefore, writing  $\text{Ord}_M$ ,  $\text{Ord}_L$  for the collections of all stable orders on the diagrams for  $(M_k)_{k \geq 0}$ ,  $(L_k)_{k \geq 0}$ , we have:

**Lemma 5.1.**  *$\text{Ord}_M$  canonically imbeds into  $\text{Ord}_L$  but this is in general not onto. The collection of all adic transformations on  $\Sigma_{(M)}^{0,+}$  corresponds to an (in general) proper subset of the adic transformations on  $\Sigma_{(L)}^{0,+}$  given by vertex orders (i.e. Markov adding machines). Similarly the group of finite coordinate changes  $\mathcal{FC}_M$  naturally embeds as a subgroup of  $\mathcal{FC}_L$ , which in general is proper.*

*Proof.* We have shown the statement for  $\text{Ord}_M$  and  $\text{Ord}_L$ , and that for the adic transformations follows. For  $\mathcal{FC}$ , the argument is similar: we can have  $\gamma \in \mathcal{FC}_L$  which sends a particular edge cylinder  $[.e_0e_1 \dots e_k e_{k+1}]$  to  $[\tilde{e}_0\tilde{e}_1 \dots \tilde{e}_k e_{k+1}]$  but does not change all the cylinders  $[\hat{e}_0\hat{e}_1 \dots \hat{e}_k \hat{e}_{k+1}]$  such that  $\hat{e}_{k+1}^- = e_{k+1}^-$ , so  $\gamma$  is not in  $\mathcal{FC}_M$ .  $\square$

Nevertheless we have:

**Lemma 5.2.** *A measure  $\nu$  on  $\Sigma_{(M)}^{0,+} = \Sigma_{(L)}^{0,+}$  is invariant for  $\mathcal{FC}_M$  if and only if it is invariant for  $\mathcal{FC}_L$ .*

*Proof.* We give a direct proof that  $\mathcal{FC}_L$ -invariance of  $\nu$  implies  $\mathcal{FC}_M$ -invariance. If we have two cylinders  $[.e_0e_1 \dots e_k e_{k+1}]$  and  $[\hat{e}_0\hat{e}_1 \dots \hat{e}_k \hat{e}_{k+1}]$  such that  $e_{k+1}^+ = \hat{e}_{k+1}^+$ , then since the matrices are reduced there exists some edge  $e_{k+2}$  with  $e_{k+2}^- = e_{k+1}^+ = \hat{e}_{k+1}^+$ ; the cylinders  $[.e_0e_1 \dots e_k e_{k+1} e_{k+2}]$  and  $[\hat{e}_0\hat{e}_1 \dots \hat{e}_k \hat{e}_{k+1} e_{k+2}]$  have the same measure by  $\mathcal{FC}_L$ -invariance, and then we sum over all such  $e_{k+2}$ .

Conversely, assuming  $\nu$  is  $\mathcal{FC}_M$ -invariant suppose we are given  $\gamma \in \mathcal{FC}_L$  with  $\gamma : [.e_0e_1 \dots e_k e_{k+1}] \mapsto [\tilde{e}_0\tilde{e}_1 \dots \tilde{e}_k e_{k+1}]$ ; we wish to show these have the same measure.

But in fact by  $\mathcal{FC}_M$ -invariance  $\nu([.e_0e_1 \dots e_k e_{k+1}]) = \nu([\hat{e}_0\hat{e}_1 \dots \hat{e}_k \hat{e}_{k+1}])$  if  $\hat{e}_{k+1}^+ = e_{k+1}^+$ , so they are certainly equal if  $e_{k+1} = \hat{e}_{k+1}$ .  $\square$

We have just discussed one way to pass from a multiple-edged to a single-edged diagram; this is like the passage from an edge shift to its vertex shift representation in the stationary case (i.e. for a subshift of finite type), see [LM95] or [Kit98].

A second way involves factoring the matrices  $M_k$ , finding a sequence of 0 – 1 matrices  $(A_0, B_0, A_1, B_1, \dots)$  such that

$$A_k B_k = M_k$$

There may be many ways to do this. A canonical factorization is given by the procedure of *symbol splitting*: place the edge set  $\mathcal{E}_k$  as a new alphabet between  $\mathcal{A}_k$  and  $\mathcal{A}_{k+1}$  so as to have alphabets  $(\mathcal{A}_0, \mathcal{E}_0, \mathcal{A}_1, \mathcal{E}_1, \dots)$  and connect  $a \in \mathcal{A}_k$  to  $e \in \mathcal{E}_k$  by an edge iff the edge  $e$  is outgoing from  $a$  in the original diagram, indicated by a 0 – 1 matrix  $A_k$ , and then make an edge from  $e$  to  $b \in \mathcal{A}_{k+1}$  iff that edge terminates on the symbol  $b$ , indicated by a 0 – 1 matrix  $B_k$ . So by definition  $A_k B_k = M_k$  as claimed.

In fact, there is a relationship between the two methods, for:

$$B_k A_{k+1} = L_k$$

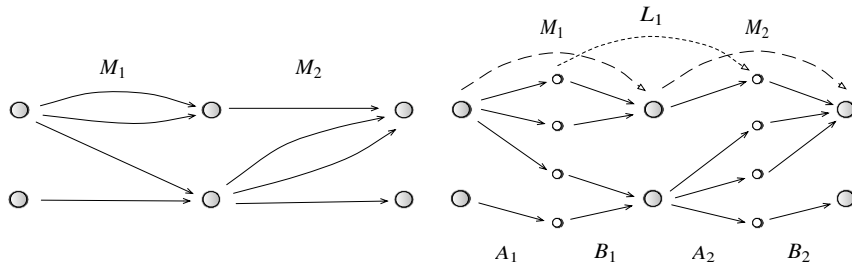


FIGURE 5. Symbol splitting:  $A_k B_k = M_k$  and  $B_k A_{k+1} = L_k$ .

Thus the first and second matrix sequences are a *gathering* of the third, one along even and one along odd times, each giving a *telescoping* of the full diagram, see [AF05] and Fig. 5.

We next consider what happens to the properties of being reduced, and of primitivity, focussing and Perron-Frobenius under these operations.

**Lemma 5.3.** *Let  $(M_i)_{i \geq 0}$  be an  $(l_i \times l_{i+1})$  non-negative integer matrix sequence, with  $(N_i)_{i \geq 0}$  a gathering of  $(M_i)_{i \geq 0}$ .*

- (i) *Then this sequence is primitive iff the gathered sequence  $(N_i)_{i \geq 0}$  is.*
- (ii) *Assuming that  $(M_i)_{i \geq 0}$  is reduced, this sequence is focussing or topologically Perron-Frobenius iff the gathered sequence  $(N_i)_{i \geq 0}$  is.*
- (iii) *If  $(M_i)_{i \geq 0}$  is reduced then  $(N_i)_{i \geq 0}$  is as well, but not necessarily conversely.*
- (iv) *If  $\Sigma_{(M)}^{0,+} \neq \emptyset$ , then  $(M_i)_{i \geq 0}$  is focussing iff that is true for the reduced sequence  $(\widehat{M}_i)_{i \geq 0}$ .*

*Proof.* We are given a sequence  $(n_i)$  with  $0 = n_0 < n_1 < \dots$  along which the gathering takes place; that is,  $N_i = M_{n_i} \cdot M_{n_i+1} \cdots M_{n_{i+1}-1}$ . The partial products of  $N_i$  are a subsequence of the partial products of the  $M_i$ , so primitivity of  $(M_i)$  implies that of  $(N_i)$  a fortiori. Conversely, if  $(N_i)$  is primitive, then starting at  $k = n_i$  the product of the  $M_i$  is certainly strictly positive eventually; but the same is true starting at any  $n_{i-1} < k < n_i$ . For part (ii), we first show that for a reduced sequence, knowing topological Perron-Frobenius for  $(N_i)_{i \geq 0}$  gives it for  $(M_i)_{i \geq 0}$ . Considering the cones  $C_{(k,n)}^+ = M^{(k,n)} C_n^+$ , note that the gathered sequence gives  $C_{(n_i, n_j)}^+$ , and if these converge to a single direction as  $n_j \rightarrow \infty$  then that happens for the  $C_{(k,n)}^+$ , since starting at  $n_{i-1} < k < n_i$  and ending at  $n_j < n < n_{j+1}$  can only improve the contraction. The only thing we have to be careful about is that these single directions be in the strictly positive vectors  $\mathring{C}_k^+$  at each of these intermediate times, as that is part of the topological Perron-Frobenius condition. We recall also the special case  $l_i = \#\mathcal{A}_i = 1$ , where  $\mathring{C}^+ = (0, \infty)$ . Thus, suppose along the subsequence  $n_i$  the limiting directions are in  $\mathring{C}_{n_i}^+$ ; we claim this is also true for the intermediate times  $n_{i-1} < k < n_i$ . But this follows from the assumption that the matrices are reduced: since no row of  $M_i$  is identically zero, no strictly positive element of the cone  $C_{n_i}^+$  can be mapped to a non-strictly positive element of  $C_k^+$  by  $M^{k, n_i}$ .

Next, assuming  $(M)$  is topologically Perron-Frobenius, we know that for any given  $k \geq 0$ ,  $\Delta_{k(+\infty)}$  is a singleton. But this property passes to the partial products  $L_i$ ,

by a similar reasoning as for primitivity (begin the sequence for  $M$  after the starting place of  $L$ ).

From Lemma 4.7, focussing is equivalent to topological Perron-Frobenius, for either sequence of matrices, completing (ii).

For (iii), if  $(\widehat{M}_i)_{i \geq 0}$  is the reduced sequence determined by  $(M_i)_{i \geq 0}$ , then its gathering along the subsequence gives the reduced sequence  $(\widehat{N}_i)_{i \geq 0}$  for  $(N_i)_{i \geq 0}$ . So being reduced passes to a gathering. That the converse is false is shown by a simple example of two  $(3 \times 3)$  matrices, with all entries positive except for the last row of the first and last column of the second which are identically zero; the product is strictly positive.

Part (iv) follows from Corollary 6.15. □

In §2.6 we defined left and right nontriviality for  $0 - 1$  matrix sequences. We now extend these definitions to a non-negative integer sequence  $(M_i)_{i \geq 0}$ ; the observation is that the resulting properties are the same, since:

**Lemma 5.4.** *Given a non-negative integer sequence  $(M_i)_{i \geq 0}$  which is left or right nontrivial, the same is true for the associated  $0 - 1$  sequence  $(L_i)_{i \geq 0}$  with edge sets for alphabets.*

*Proof.* If for each  $1 \leq i \leq l_i$ , the sum of the entries in the  $i^{\text{th}}$  row of  $M^{(k,m)} \equiv M_k M_{k+1} \dots M_{m-1}$  is  $\geq 2$ , then given an edge from  $a \in \mathcal{A}_{k-1}$  to  $b \in \mathcal{A}_k$ , there are at least two edge paths from  $b$  to  $\mathcal{A}_m$ , hence the  $a^{\text{th}}$  row sum of  $L^{(k-1,m)}$  is  $\geq 2$ . This implies right nontriviality for  $(L_i)_{i \geq 0}$ .

The argument for left nontriviality is similar. □

Therefore the conclusions of Lemma 2.4 are valid for edge spaces as well.

We then have:

**Theorem 5.5.** (Minimality and unique ergodicity for nonstationary multiple-edged adic transformations) Given a nonnegative integer matrix sequence  $(M_i)_{i \geq 0}$  and a stable order  $\mathcal{O}$ , then if the sequence is primitive, and more generally if the reduced matrix sequence  $(\widehat{M}_i)_{i \geq 0}$  is primitive, the adic transformation on  $\Sigma_{(M)}^{0,+} \setminus \mathcal{N}$  is minimal.

Assuming that  $(M_i)_{i \geq 0}$  is reduced and primitive, then the sequence is focussing (or equivalently is topologically Perron-Frobenius) if and only if the adic transformation is uniquely ergodic. The adic transformation is primitive, respectively uniquely ergodic if and only if that holds for the group of finite coordinate changes.

For a non-reduced sequence  $(M_i)_{i \geq 0}$ , suppose that  $\Sigma_{(M)}^{0,+} \neq \emptyset$ . Assume that  $(\widehat{M}_i)_{i \geq 0}$  is primitive. Then if  $(M_i)_{i \geq 0}$  is focussing or equivalently geometrically Perron-Frobenius, the adic transformation is uniquely ergodic.

*Proof.* For  $(M_i)_{i \geq 0}$  not necessarily reduced, we define  $A_i, B_i$  as above, so  $M_i = A_i B_i$  and  $L_i \equiv B_i A_{i+1}$  gives the vertex representation. From part (i) of Lemma 5.3,  $(M_i)_{i \geq 0}$  is primitive iff the sequence  $(A_i B_i)_{i \geq 0}$  is, iff  $(L_i)_{i \geq 0}$  is. By Theorem 2.13 the adic transformation on  $\Sigma_{(L)}^{0,+} \setminus \mathcal{N}$  and hence (by Lemma 5.1) on  $\Sigma_{(M)}^{0,+} \setminus \mathcal{N}$  is minimal; this theorem does not assume the sequence is reduced. If the reduced sequence  $(\widehat{M}_i)_{i \geq 0}$



is primitive, then the adic transformation on  $\Sigma_{(\widehat{M})}^{0,+} \setminus \mathcal{N}$  is minimal, but this space is identical to  $\Sigma_{(M)}^{0,+} \setminus \mathcal{N}$ .

Now consider the reduced sequence  $(\widehat{M}_i)_{i \geq 0}$ , and factor that as  $\widehat{M}_i = A_i B_i$  with vertex representation written  $\widehat{L}_i \equiv B_i A_{i+1}$ . By part (iv) of Lemma 5.3, if  $(M_i)_{i \geq 0}$  is focussing then so is  $(\widehat{M}_i)_{i \geq 0}$ . Then by part (ii) of the lemma,  $(\widehat{L}_i)_{i \geq 0}$  is focussing, hence by Theorem 4.6, the adic transformation on  $\Sigma_{(\widehat{L})}^{0,+} \setminus \mathcal{N}$  is uniquely ergodic, hence so are those on  $\Sigma_{(\widehat{M})}^{0,+} \setminus \mathcal{N}$  and on  $\Sigma_{(M)}^{0,+} \setminus \mathcal{N}$ .  $\square$

*Remark 5.1.* We recall from (11) the formula for the unique invariant measure  $\nu$  for the case of 0 – 1 matrices. In fact the same formula holds for non-negative integer matrices  $(M_i)_{i \geq 0}$ . Thus, for  $\mathbf{w}_i$  a non-negative strictly positive column eigenvector sequence (with eigenvalues one) and normalized so that  $\|\mathbf{w}_0\| = 1$ , then an edge cylinder set  $[.e_0 \dots e_{m-1}]$  in  $\Sigma_{(M)}^{0,+} \equiv \Sigma_{(L)}^{0,+}$  has measure

$$\nu_{\mathbf{w}}([.e_0 \dots e_{m-1}]) = (\mathbf{w}_m)_a, \quad (12)$$

where  $a = e_{m-1}^+ \in \mathcal{A}_m$ . This follows by applying (11) to the factorization  $M_i = A_i B_i$ .

**5.1. Examples revisited.** To complete the discussion of the examples of §2, we review how they behave with respect to minimality and unique ergodicity; this also serves to illustrate how the conditions of Theorem 5.5 can be applied in practice.

The nonstationary odometer is both minimal and uniquely ergodic for  $\mathcal{FC}$ ; one proof was given in the introduction. For a second proof note that primitivity and the topological Perron-Frobenius property is always satisfied. For a third proof we recode this as an edge shift with one-element alphabets  $\mathcal{A}_i$  and  $l_i$  edges, so the new matrices are  $(1 \times 1)$ , with  $M_i = [l_i]$ . As before there is a unique possible stable order and a unique continuous extension. (One should rule out the case where  $l_i$  is only finitely often  $\geq 2$  so the Bratteli diagram will be nontrivial, but even then, the minimality and unique ergodicity are true for  $\mathcal{FC}$  and for the extended map).

This leads to another example. If we are given alphabet sequence  $\mathcal{A}_i$  and  $(l_i \times l_{i+1})$  non-negative integer sequence  $M_i$ , then:

**Proposition 5.6.** *If for an infinite subsequence  $n_i$ ,  $l_{n_i} = 1$ , the adic transformation is naturally isomorphic to a nonstationary odometer.*

*Proof.* We gather the diagram along these times to produce a  $(1 \times 1)$  matrix sequence  $\widehat{M}_i = [\widehat{l}_i]$  where this is the number of paths from time  $n_i$  to time  $n_{i+1}$ .  $\square$

This implies unique ergodicity, which can also be seen from the topological Perron-Frobenius property: since the subsequence of simplices  $\Delta_{n_i}$  are points, the nested image is a point as well.

The irrational rotation adic transformations are, just like the corresponding circle rotations, minimal and uniquely ergodic. To prove unique ergodicity, we can as in Prop. 4.1 of [AF05] calculate explicitly the positive eigenvectors; here we give a simpler argument. Infinitely often  $M_i M_{i+1} = PQ = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$  or  $QP = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$  which

implies primitivity; focussing holds since the Birkhoff contraction factor for  $PQ$  and  $QP$  is  $< 1$  (and can be computed explicitly from Prop. 6.13 and Theorem 6.8).

For the rational rotation adics, although the map  $T_{\mathcal{O}}$  itself is trivial as  $\mathcal{N}$  is the whole space, the extended maps  $T_{\mathcal{O},f}$  are uniquely ergodic but not minimal, as explained above Prop. 2.9.

From interval exchanges, one can get examples of matrix sequences where primitivity holds but where topological Perron-Frobenius may or may not be satisfied. In particular, Keane's well known counterexample [Kea77] of an exchange of four intervals which is minimal but not uniquely ergodic is isomorphic (topologically, off a countable set) to an adic transformation on four symbols with that behavior, see [Fisa]. In the edge shift presentation the matrices are a  $(4 \times 4)$  non-negative unimodular (i.e. determinant one), primitive matrix sequence

$$M_j = \begin{pmatrix} 0 & 0 & 1 & 1 \\ m_j - 1 & m_j & 0 & 0 \\ n_j & n_j & n_j - 1 & n_j \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Keane's proof in [Kea77] shows that with  $m_i, n_i$  chosen to grow sufficiently fast, the Perron-Frobenius condition fails, while the product  $M_1 M_2$  already demonstrates primitivity, and so by Theorem 5.5, the adic transformation is minimal but is not uniquely ergodic.

In fact, as is well known (by a completely different proof, see [Vee78], and also e.g. [Via06], §5.1), to guarantee unique ergodicity in the special case of interval exchanges, it is enough to check the topological Perron-Frobenius condition at  $k = 0$ . Indeed this will be the case whenever the matrices are invertible, since then whether  $\bigcap_{m \geq k} M_k \cdots M_m C_{m+1}^+$  nests down to a single direction is independent of  $k$ . And, in the case of interval exchanges the matrices in fact always have determinant one, so this is satisfied.

Other "adic counterexamples" are explicitly constructed in [FFT09], including a  $(3 \times 3)$  unimodular matrix sequence with combinatorics similar to those of Keane's example; it is shown that no such example exists for alphabet size two.

We have mostly considered primitive sequences or, at worst, sequences which are primitive after being reduced; moving beyond primitivity is much of the motivation behind [BKMS09]. An important example with a nonprimitive matrix sequence is the Pascal adic transformation; there, as shown by Vershik, one has a continuum of invariant probability measures, and the same holds for the cutting and stacking example of [HIK72] which is modelled by the Pascal adic. Indeed using the representation of the adic transformation on the space  $\Sigma^+$ , each infinite product (Bernoulli) measure  $(p, q)$  with  $p, q > 0$ ,  $p + q = 1$  gives an invariant probability measure, since if  $y \in W^s(x)$  with  $k$  the least digit such that  $x_j = y_j$  for all  $j \geq k$ , then considering the Pascal triangle, the number of 0's and 1's in  $x, y$  up to  $k$  is the same and so the measures of the cylinder sets  $[.x_0 \dots x_k], [.y_0 \dots y_k]$  are equal. Moreover these are the only invariant measures, and the transformation is essentially minimal (minimal off a

countable set), see [MP05]. This furnished the first example of an essentially minimal but not uniquely ergodic adic transformation, albeit with unbounded alphabet size.

A basic nonprimitive (and nonunimodal) example with bounded alphabet size (in fact it is stationary) is the Chacon adic transformation, with alphabet  $\mathcal{A} = \{0, 1\}$ , edge matrix  $A = \begin{bmatrix} 3 & 0 \\ 1 & 1 \end{bmatrix}$  and with stable edge order defined by the substitution  $\rho : 0 \mapsto 0010, \rho : 1 \mapsto 1$ . As shown in [FFT09], despite the lack of primitivity the adic transformation  $(\Sigma_A^+ \setminus \mathcal{N}, T_{\mathcal{O}})$  is both minimal and uniquely ergodic, though the action of  $\mathcal{FC}$  has also an atomic invariant measure supported on the fixed point. Since the Bratteli diagram is not left nontrivial this does not violate (ii) of Proposition 2.11.

## 6. APPENDIX: G. BIRKHOFF'S BOUND FOR THE PROJECTIVE METRIC

We present here the background material on the projective metric necessary for the main part of the paper. Specifically, Lemma 6.7 on the projective diameter of a convex hull is used in the proofs of both Prop. 6.13 and Lemma 4.4; Cor. 6.12 on isometries is needed in the proof of Lemma 4.4; and Cor. 6.14 is essential to the proof of Lemma 4.7. To prove this corollary, we further need Birkhoff's contraction coefficient (Theorem 6.8(d)) and his specific formula for that in the case of the standard positive cone in  $\mathbb{R}^n$  (Prop. 6.10).

Historically a main motivating idea behind the projective metric was to find a simple, contraction-mapping proof of the Perron-Frobenius Theorem for strictly positive square matrices. Garrett Birkhoff gave this proof in the fundamental papers [Bir57] [Bir67], along with considerable generalizations and a deep study of the analysis and geometry of general positive operators on a Banach space. At about the same time as Birkhoff, Samelson gave a proof of the Perron-Frobenius Theorem in  $\mathbb{R}^n$  using the closely related Hilbert metric on the unit simplex [Sam56]. In our treatment we borrow ideas from the elegant treatment of [Fur60] as well as from Birkhoff's papers. The setting we need here is that of closed positive convex cones in Euclidean space; with no added difficulty, we state everything for Banach spaces as does Birkhoff.

### The projective metric on a convex cone.

We begin with the concept of abstract positive cone; this isolates what is needed to define partial orders on a vector space, generalizing the usual partial order on  $\mathbb{R}^n$  associated to the standard positive cone  $\mathbb{R}^{n+} = \{\mathbf{v} : v_i \geq 0, 1 \leq i \leq n\}$ . The notion of positive linear transformation follows from this.

**Definition 6.1.** *Let  $V$  be a real vector space. A subset  $C$  is a cone if and only if  $C$  is nonempty and  $\alpha C \subseteq C$  for all  $\alpha \geq 0$ . A cone is convex iff  $C + C \subseteq C$ , equivalently iff it is a cone and a convex set. It is positive iff  $C \cap -C = \{\mathbf{0}\}$ .*

*Given a positive convex cone  $C \subseteq V$ , a vector  $\mathbf{x} \in V$  is positive iff  $\mathbf{x} \in C$ . For  $\mathbf{x}, \mathbf{y} \in V$  we define  $\mathbf{x} \leq \mathbf{y}$  iff  $(\mathbf{y} - \mathbf{x})$  is positive. So equivalently,  $\mathbf{x} \leq \mathbf{y}$  iff  $\mathbf{y} \in \mathbf{x} + C$ .*

*Given vector spaces  $V, W$  and positive convex cones  $C, D$  in  $V, W$  respectively, a linear transformation  $f : V \rightarrow W$  is positive iff  $f(C) \subseteq D$ .*

**Proposition 6.1.** *Given a positive convex cone  $C \subseteq V$ ,*

*(a)  $(\leq)$  defines a partial order on  $V$ .*

(b) A linear transformation is positive iff it preserves the partial order:  $(\mathbf{x} \leq_C \mathbf{y}) \implies (f(\mathbf{x}) \leq_D f(\mathbf{y}))$ .

*Proof.* The properties reflexivity  $\mathbf{x} \leq \mathbf{x}$ , symmetry  $(\mathbf{x} \leq \mathbf{y} \wedge \mathbf{y} \leq \mathbf{x}) \implies (\mathbf{y} = \mathbf{x})$ , and transitivity  $(\mathbf{x} \leq \mathbf{y}, \mathbf{y} \leq \mathbf{z}) \implies (\mathbf{x} \leq \mathbf{z})$  follow respectively from  $C$  being a cone (since then  $\mathbf{0} \in C$ ), positive and convex. Part (b) is immediate.  $\square$

The order is then used to define a metric, as follows. From the cone property, for any  $\lambda \in (0, +\infty)$  and  $\mathbf{x}, \mathbf{y} \in C$  one has

$$\mathbf{x} \leq \lambda \mathbf{y} \iff \lambda^{-1} \mathbf{x} \leq \mathbf{y}. \quad (13)$$

Given a positive convex cone  $C$  and vectors  $\mathbf{x}, \mathbf{y} \in C$ , we define

$$\alpha_0 = \alpha(\mathbf{x}, \mathbf{y}) = \sup\{\alpha \in \mathbb{R} : \alpha \mathbf{y} \leq \mathbf{x}\}; \quad (14)$$

this set is nonempty since  $\mathbf{0} \in C$ , and so  $\alpha_0 \in [0, +\infty]$ ; taking by convention  $0^{-1} = +\infty$ ,  $+\infty^{-1} = 0$  and  $\inf \emptyset = +\infty$ , we define

$$\beta_0 = \beta(\mathbf{x}, \mathbf{y}) \equiv \alpha(\mathbf{y}, \mathbf{x})^{-1}. \quad (15)$$

By (13) together with these conventions, an equivalent definition of  $\beta_0$  is:

$$\beta_0 = \inf\{\beta \geq 0 : \mathbf{x} \leq \beta \mathbf{y}\}. \quad (16)$$

**Lemma 6.2.** *Let  $C$  be a closed, positive convex cone in a real Banach space. Then  $\beta_0 = 0$  if and only if  $\mathbf{x} = \mathbf{0}$ , and  $\alpha_0 = +\infty$  if and only if  $\mathbf{y} = \mathbf{0}$ .*

*Proof.* From (16), if  $\beta_0 = 0$  then  $\mathbf{x} \leq \frac{1}{n} \mathbf{y}$  for all  $n \geq 1$ , so  $\frac{1}{n} \mathbf{y} - \mathbf{x} \in C$ ; since  $C$  is closed  $\lim_{n \rightarrow \infty} (\frac{1}{n} \mathbf{y} - \mathbf{x}) = -\mathbf{x} \in C$ . But then  $\mathbf{x}$  and  $-\mathbf{x} \in C$  so  $\mathbf{x} = \mathbf{0}$  by positivity of the cone.

If  $\alpha_0 = +\infty$  then  $n \mathbf{y} \leq \mathbf{x}$  for all  $n \geq 1$ , so by (13)  $\mathbf{y} \leq \frac{1}{n} \mathbf{x}$  for all  $n \geq 1$ , and as in the previous case this implies  $\mathbf{y} = \mathbf{0}$ .  $\square$

Noting from (14) and (16) that  $\alpha_0 \leq \beta_0$ , then for  $\mathbf{x}, \mathbf{y} \in C \setminus \{\mathbf{0}\}$ , we have from the lemma that

$$\beta_0 > 0 \text{ and } \alpha_0 < +\infty \quad (17)$$

so taking by convention for  $c > 0$ ,  $\frac{c}{+\infty} = 0$  and  $\frac{c}{0} = +\infty$ , then  $\frac{\beta_0}{\alpha_0} \in [1, +\infty]$ . We define for  $\mathbf{x}, \mathbf{y} \in C \setminus \{\mathbf{0}\}$

$$d_C(\mathbf{x}, \mathbf{y}) = \log(\beta_0/\alpha_0) \in [0, +\infty].$$

**Lemma 6.3.** *Let  $C$  be a closed, positive convex cone in a real Banach space. Then for  $\mathbf{x}, \mathbf{y} \in C \setminus \{\mathbf{0}\}$ , the supremum  $\sup\{\alpha : \alpha \mathbf{y} \leq \mathbf{x}\}$  is attained (i.e. it is a max not just a sup); that is,  $\alpha_0 \mathbf{y} \leq \mathbf{x}$ , and similarly for the infimum.*

*Proof.* We show  $\sup\{\alpha : \alpha \mathbf{y} \leq \mathbf{x}\}$  is attained. Let  $\alpha_n$  increase to  $\alpha_0$ , so  $\mathbf{v}_n \equiv \mathbf{x} - \alpha_n \mathbf{y} \in C$ . As we have seen above  $\alpha_0$  is finite, so  $\lim \alpha_n \mathbf{y} = \alpha_0 \mathbf{y}$ ; since  $C$  is closed,  $\lim(\mathbf{x} - \alpha_n \mathbf{y}) = \mathbf{x} - \alpha_0 \mathbf{y} \in C$ ; therefore indeed,  $\alpha_0 \mathbf{y} \leq \mathbf{x}$ . The proof for  $\beta_0$  is similar.  $\square$

Noting that for each  $\mathbf{x}, \mathbf{y} \in C$

$$0 \cdot \mathbf{x} \leq \mathbf{y}$$

we adopt the convention that for any  $\mathbf{x} \in C$ ,  $\mathbf{y} \in C \setminus \{\mathbf{0}\}$ ,

$$\mathbf{x} \leq +\infty \cdot \mathbf{y}.$$

**Proposition 6.4.** *For  $C$  a closed positive convex cone in a real Banach space  $V$ ,  $d_C$  is an (extended-real valued) metric (called the projective metric) on the projective space of open half-lines from the origin which lie in  $C$ .*

This follows from:

**Lemma 6.5.** *Given  $\mathbf{x}, \mathbf{y}, \mathbf{z} \in C \setminus \{\mathbf{0}\}$ ,*

(i) *For each  $\lambda \in (0, +\infty)$ ,  $d_C(\mathbf{x}, \mathbf{y}) = d_C(\lambda\mathbf{x}, \mathbf{y})$ .*

(ii)  *$d_C(\mathbf{x}, \mathbf{y}) = d_C(\mathbf{y}, \mathbf{x})$ .*

(iii)  *$d_C(\mathbf{x}, \mathbf{z}) \leq d_C(\mathbf{x}, \mathbf{y}) + d_C(\mathbf{y}, \mathbf{z})$ .*

(iv)  *$d_C(\mathbf{x}, \mathbf{y}) = 0$  if and only if there exists  $\lambda \in (0, +\infty)$  such that  $\mathbf{y} = \lambda\mathbf{x}$ .*

*Proof.* For (i),  $d_C(\mathbf{x}, \mathbf{y}) = d_C(\lambda\mathbf{x}, \mathbf{y})$  since both  $\alpha_0$  and  $\beta_0$  are multiplied by the same constant.

For (ii), from (15)  $\alpha(\mathbf{y}, \mathbf{x}) = \beta(\mathbf{x}, \mathbf{y})^{-1}$  and  $\beta(\mathbf{y}, \mathbf{x}) = \alpha(\mathbf{x}, \mathbf{y})^{-1}$ , which gives  $d_C(\mathbf{x}, \mathbf{y}) = d_C(\mathbf{y}, \mathbf{x})$ .

Next we check the triangle inequality (iii). Given three vectors  $\mathbf{x}, \mathbf{y}, \mathbf{z} \in C$ , let us write  $\alpha_1$  for  $\alpha(\mathbf{x}, \mathbf{y})$ ,  $\alpha_2$  for  $\alpha(\mathbf{y}, \mathbf{z})$  and  $\alpha_3$  for  $\alpha(\mathbf{x}, \mathbf{z})$ , and similarly for  $\beta$ . We have

$$\alpha_1\mathbf{y} \leq \mathbf{x} \leq \beta_1\mathbf{y}$$

and

$$\alpha_2\mathbf{z} \leq \mathbf{y} \leq \beta_2\mathbf{z}$$

so

$$\alpha_1\alpha_2\mathbf{z} \leq \alpha_1\mathbf{y} \leq \mathbf{x} \leq \beta_1\mathbf{y} \leq \beta_1\beta_2\mathbf{z}.$$

Therefore  $\alpha_3 \geq \alpha_1\alpha_2$ ,  $\beta_3 \leq \beta_1\beta_2$  and so

$$d_C(\mathbf{x}, \mathbf{z}) = \log(\beta_3/\alpha_3) \leq \log(\beta_1\beta_2/\alpha_1\alpha_2) = d_C(\mathbf{x}, \mathbf{y}) + d_C(\mathbf{y}, \mathbf{z}).$$

For (iv), given  $\mathbf{x}, \mathbf{y} \in C \setminus \{\mathbf{0}\}$  with  $d_C(\mathbf{x}, \mathbf{y}) = 0$  then  $\beta_0/\alpha_0 = 1$ , so  $\beta_0, \alpha_0 \in (0, +\infty)$ . By Lemma 6.3,  $\alpha_0\mathbf{y} \leq \mathbf{x} \leq \alpha_0\mathbf{y}$ . Thus  $\mathbf{x} - \alpha_0\mathbf{y} \in C$  and  $\alpha_0\mathbf{y} - \mathbf{x} \in C$ , but since the cone  $C$  is positive,  $\mathbf{x} = \alpha_0\mathbf{y}$ , with  $\alpha_0 \in (0, +\infty)$ , as claimed.  $\square$

Note: one may be tempted to extend the metric to all of  $C$ , by defining  $d_C(\mathbf{0}, \mathbf{x}) = d_C(\mathbf{x}, \mathbf{0}) = 0$ , but this is a bad idea as then the triangle inequality would fail!

**Lemma 6.6.** *Projective distances add along line segments; that is, given  $\mathbf{x}, \mathbf{y} \in C \setminus \{\mathbf{0}\}$  and  $p, q \geq 0$  with  $\mathbf{z} = p\mathbf{x} + q\mathbf{y}$ , then  $d_C(\mathbf{x}, \mathbf{z}) + d_C(\mathbf{z}, \mathbf{y}) = d_C(\mathbf{x}, \mathbf{y})$ .*

*Proof.* It is enough to prove for  $p + q = 1$ . We have  $\alpha_0, \beta_0$  such that  $\alpha_0\mathbf{y} \leq \mathbf{x} \leq \beta_0\mathbf{y}$  and  $\beta_0^{-1}\mathbf{x} \leq \mathbf{y} \leq \alpha_0^{-1}\mathbf{x}$  where  $\alpha_0 = \alpha(\mathbf{x}, \mathbf{y})$ ,  $\beta_0 = \beta(\mathbf{x}, \mathbf{y})$ . One checks that then indeed

$$\frac{\beta(\mathbf{x}, \mathbf{z})}{\alpha(\mathbf{x}, \mathbf{z})} \cdot \frac{\beta(\mathbf{z}, \mathbf{y})}{\alpha(\mathbf{z}, \mathbf{y})} = \frac{\beta(\mathbf{x}, \mathbf{y})}{\alpha(\mathbf{x}, \mathbf{y})},$$

giving the claim.  $\square$

**Lemma 6.7.** *Let  $C \subseteq \mathbb{R}^n$  be a closed positive convex cone and let  $d_C(\cdot, \cdot)$  denote the projective metric on  $C \setminus \{\mathbf{0}\}$ . Suppose that we are given points  $\mathbf{a}_1, \dots, \mathbf{a}_k \in C \setminus \{\mathbf{0}\}$ ; write  $D$  for the convex cone generated by  $\{\mathbf{a}_1, \dots, \mathbf{a}_k\}$ . Then the  $d_C$ -diameter of  $D$  is*

$$\text{diam}_C(D) = \max_{i,j} \{d_C(\mathbf{a}_i, \mathbf{a}_j)\}$$

*Proof. Claim 1:* Assume we are given  $\mathbf{v}, \mathbf{w}, \mathbf{z} \in C \setminus \{\mathbf{0}\}$  and  $\mathbf{x} = p\mathbf{v} + q\mathbf{w}$  with  $p, q \geq 0$  and  $p + q = 1$ , and such that  $d_C(\mathbf{z}, \mathbf{v}) = d_C(\mathbf{z}, \mathbf{w}) = A$  for  $A \in [0, +\infty]$ . Then  $d_C(\mathbf{z}, \mathbf{x}) \leq A$ .

*Proof of Claim 1:* We know from by Lemma 6.3 that there are numbers  $0 \leq \alpha_i \leq \beta_i \leq +\infty$  satisfying  $\alpha_1\mathbf{z} \leq \mathbf{v}$ ,  $\mathbf{v} \leq \beta_1\mathbf{z}$ ,  $\alpha_2\mathbf{z} \leq \mathbf{w}$ , and  $\mathbf{w} \leq \beta_2\mathbf{z}$  and where these are the sups and infs of the possible such numbers. Then defining  $\alpha = p\alpha_1 + q\alpha_2$ , we have

$$\alpha\mathbf{z} = p\alpha_1\mathbf{z} + q\alpha_2\mathbf{z} \leq p\mathbf{v} + q\mathbf{w} = \mathbf{x}.$$

Similarly, for  $\beta = p\beta_1 + q\beta_2$ , we have

$$\beta\mathbf{z} = p\beta_1\mathbf{z} + q\beta_2\mathbf{z} \geq p\mathbf{v} + q\mathbf{w} = \mathbf{x}.$$

So

$$d_C(\mathbf{z}, \mathbf{x}) \leq \log\left(\frac{\beta}{\alpha}\right) = \log\left(\frac{p\beta_1 + q\beta_2}{p\alpha_1 + q\alpha_2}\right).$$

We divide on top and bottom by  $\alpha_1\alpha_2$  and then set  $\tilde{p} = (p/\alpha_2)/(p/\alpha_2 + q/\alpha_1)$  and  $\tilde{q} = (q/\alpha_1)/(p/\alpha_2 + q/\alpha_1)$ ; this gives

$$\log\left(\frac{\beta}{\alpha}\right) = \log\left(\tilde{p}\frac{\beta_1}{\alpha_1} + \tilde{q}\frac{\beta_2}{\alpha_2}\right)$$

So if  $\beta_1/\alpha_1 = \beta_2/\alpha_2$ , which is the case when both distances =  $A$ , then  $\log(\beta/\alpha) = A$  as well. This shows that  $d_C(\mathbf{z}, \mathbf{x}) \leq \log(\beta/\alpha) = A$ , proving the claim. (We remark that with a bit more work one can show that  $d_C(\mathbf{z}, \mathbf{x}) = A$ , but that fact is not needed for the rest of the argument).

**Claim 2:** Next we show that given  $A > 0$  and points  $\mathbf{a}_1, \dots, \mathbf{a}_k \in C \setminus \{\mathbf{0}\}$  such that  $d_C(\mathbf{a}_i, \mathbf{a}_j) \leq A$  for each  $i, j$ , then for any  $\mathbf{x}$  in the positive, convex cone  $D$  in  $\mathbb{R}^n$  generated by the  $\mathbf{a}_i$ ,  $d(\mathbf{x}, \mathbf{a}_i) \leq A$ .

Note that the statement is true for  $k = 0$  and 1. We now show the induction step that if it is valid for  $k$  then it is valid for  $(k + 1)$ .

Suppose  $\mathbf{x}$  is in the convex cone generated by the  $\mathbf{a}_i$ , thus  $\mathbf{x}$  can be expressed as  $\mathbf{x} = \sum_{i=0}^k p_i \mathbf{a}_i$  with  $p_i \geq 0$ . Choosing one of the generating vectors, say  $\mathbf{a}_0$ , we project  $\mathbf{x}$  to a point on the opposite face of the cone, defining  $\tilde{\mathbf{x}}$  in the convex span  $\langle \mathbf{a}_1, \dots, \mathbf{a}_{k+1} \rangle$ , that is, we define  $\tilde{\mathbf{x}} = c \sum_{i=1}^{k+1} p_i \mathbf{a}_i$  with  $c \sum_{i=1}^{k+1} p_i = 1$  for some  $c > 0$ . By Lemma 6.6 distances add along line segments, so  $d_C(\mathbf{a}_0, \mathbf{x}) \leq d_C(\mathbf{a}_0, \tilde{\mathbf{x}})$ ; we wish to show  $d_C(\mathbf{a}_0, \tilde{\mathbf{x}}) \leq A$ . Now we project  $\tilde{\mathbf{x}}$  to a point  $\hat{\mathbf{x}}$  in the convex span  $\langle \mathbf{a}_1, \dots, \mathbf{a}_k \rangle$ , with  $\hat{\mathbf{x}} = \tilde{c} \sum_{i=1}^k p_i \mathbf{a}_i$  and  $\tilde{c} \sum_{i=1}^k p_i = 1$ . We consider the triangular cone  $\langle \mathbf{a}_0, \hat{\mathbf{x}}, \mathbf{a}_{k+1} \rangle$ ; note that  $\tilde{\mathbf{x}}$  is a convex combination of  $\hat{\mathbf{x}}, \mathbf{a}_{k+1}$ . By the induction hypothesis,  $d_C(\mathbf{a}_0, \hat{\mathbf{x}}) \leq A$ ; now we apply Claim 1 with  $\mathbf{z} = \mathbf{a}_0$ ,  $\mathbf{v} = \hat{\mathbf{x}}$ ,  $\mathbf{w} = \mathbf{a}_{k+1}$ , and so  $d_C(\mathbf{a}_0, \tilde{\mathbf{x}}) \leq A$  as desired.

Now we move on to the proof of the lemma. Let  $\hat{\mathbf{x}} \neq \tilde{\mathbf{x}}$  in the convex span  $\langle \mathbf{a}_1, \dots, \mathbf{a}_k \rangle$ ; this is compact, so the affine line through  $\hat{\mathbf{x}}, \tilde{\mathbf{x}}$  meets this convex set in a compact segment with two endpoints  $\eta, \xi$ ; since distances add along segments,  $d_C(\hat{\mathbf{x}}, \tilde{\mathbf{x}}) \leq d_C(\eta, \xi)$  so it enough to show  $d_C(\eta, \xi) \leq A$ . Since  $\eta \in \langle \mathbf{a}_1, \dots, \mathbf{a}_k \rangle$ , we have  $d_C(\eta, \mathbf{a}_i) \leq A$  by Claim 2. Therefore we can add the point  $\eta$  to this set, calling it  $\mathbf{a}_{k+1}$ ; now  $\langle \mathbf{a}_1, \dots, \mathbf{a}_{k+1} \rangle = \langle \mathbf{a}_1, \dots, \mathbf{a}_k \rangle$  satisfies that  $d_C(\mathbf{a}_i, \mathbf{a}_j) \leq A$  for each  $i, j$  so again by Claim 2, also  $d_C(\mathbf{a}_i, \xi) \leq A$ . But then in particular  $d_C(\eta, \xi) = d_C(\mathbf{a}_{k+1}, \xi) \leq A$  as claimed, completing the proof.  $\square$

**6.1. Weak and strong contraction.** Here we shall see that while positive mappings always give a weak contraction, for strict inclusion one gets much more: a contraction with an exact bound. This last remarkable result, due to Birkhoff, is from [Bir57] and [Bir67]. That this is reminiscent of theorems from complex analysis is no accident, since the projective metric on a circular cone gives the Klein model for the hyperbolic disk on a cross-section (times a constant), and so the projective (and Hilbert) metrics can be thought of as generalizations of hyperbolic space.

For the statement of part (d) we recall that

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

**Theorem 6.8.** *Let  $V, W$  be real Banach spaces and  $C \subseteq V, D \subseteq W$  closed convex cones with  $d_C, d_D$  their projective metrics. Let  $L : V \rightarrow W$  be a positive linear transformation. Then:*

(a)  *$L$  is a weak contraction from  $C$  to  $D$ , i.e.*

$$d_D(L(\mathbf{v}), L(\mathbf{w})) \leq d_C(\mathbf{v}, \mathbf{w}).$$

(b) *If  $L$  is invertible, it is an isometry.*

(c) *If  $C \subseteq D$ , then  $d_C(\mathbf{v}, \mathbf{w}) \leq d_D(\mathbf{v}, \mathbf{w})$ .*

(d) *(Birkhoff) We write  $\Theta(L)$  for the  $D$ -diameter of  $L(C)$ . Then if  $\Theta(L) < \infty$ ,  $L$  is a strict contraction, with coefficient*

$$\sup_{\mathbf{v}, \mathbf{w} \in C} \frac{d_D(L(\mathbf{v}), L(\mathbf{w}))}{d_C(\mathbf{v}, \mathbf{w})} = \tanh(\Theta/4).$$

*This bound is sharp.*

The proofs of (a), (b) follow from the definitions; (c) is a corollary, using the identity map. For (d) see Birkhoff's papers. We call  $\Theta(L)$  the *opening* of the operator  $L$ , and  $\tanh(\Theta/4)$  the *contraction coefficient* or *contraction factor* of the linear operator.

Here is an immediate consequence of part(d):

**Corollary 6.9.** *Let  $(V_i)_{i \geq 0}$  be real Banach spaces with closed convex cones  $C_i \subseteq V_i$ , and let  $L_i : V_i \rightarrow V_{i+1}$  be positive linear transformations. Write  $\Theta_i$  for the  $C_{i+1}$ -diameter of  $L_i(C_i)$ , and define  $\alpha_i = \tanh(\Theta_i/4)$ . Then the contraction coefficient of the composition  $L_m \circ \dots \circ L_0$  is bounded above by  $\alpha_m \cdots \alpha_0$ .  $\square$*

## 6.2. The standard positive cone and the contraction formula for transposes.

We now specialize to  $V = \mathbb{R}^n$  with  $C$  the usual positive cone

$$C = \mathbb{R}^{n+} = \{\mathbf{v} : v_i \geq 0, 1 \leq i \leq n\}.$$

In this case one has a very useful explicit formula for the projective metric, due to Birkhoff.

**Proposition 6.10.** *For  $\mathbf{v}, \mathbf{w} \in C \setminus \{\mathbf{0}\}$  where  $C = \mathbb{R}^{n+}$ , defining  $\frac{0}{0} = 1$ , then*

$$d_C(\mathbf{v}, \mathbf{w}) = \log \max_{i,j} \frac{v_i w_j}{w_i v_j}$$

*Proof.* As we shall see, in proving this we shall be able to exclude the cases where both  $v_i, w_i$  or  $v_j, w_j$  are 0.

We have

$$\alpha_0 = \max\{\alpha : \alpha \mathbf{v} \leq \mathbf{w}\} = \max\{\alpha : \forall i = 1, \dots, n, \alpha v_i \leq w_i\}.$$

For  $v_i = w_i = 0$ ,  $\sup\{\alpha : \alpha \mathbf{0} \leq \mathbf{0}\} = +\infty$ , so (since  $\mathbf{v}, \mathbf{w} \neq \mathbf{0}$ )

$\alpha_0 = \max\{\alpha : \alpha v_i \leq w_i : \forall i \text{ not both } v_i, w_i \text{ are } 0\} = \min\{v_i/w_i : \text{not both } v_i, w_i \text{ are } 0\}$ ,  
and similarly since  $\inf\{\beta : \mathbf{0} \leq \beta \mathbf{0}\} = -\infty$ ,

$$\beta_0 = \max\{v_i/w_i : \text{not both } v_i, w_i \text{ are } 0\}.$$

From (17)  $\alpha_0 < +\infty$  and  $\beta_0 > 0$ . So

$$\begin{aligned} \beta_0/\alpha_0 &= \max_{i,j} \left\{ \frac{v_i}{w_i} / \frac{v_j}{w_j} \text{ not both } v_i, w_i \text{ or } v_j, w_j \text{ are } 0 \right\} \\ &= \max \left\{ \frac{v_i w_j}{w_i v_j} \text{ not both } v_i, w_i \text{ or } v_j, w_j \text{ are } 0 \right\} \\ &= \max \left\{ \frac{v_i w_j}{w_i v_j} : i, j = 1, \dots, n \right\} \end{aligned}$$

where in the last equality we use the definition  $\frac{0}{0} = 1$  and the fact that always  $\beta_0/\alpha_0 \geq 1$ .  $\square$

From this we derive a corollary we need in the proof of Lemma 3.6:

**Corollary 6.11.** *Let  $\mathbf{v}, \mathbf{w}$  be probability vectors in  $C = \mathbb{R}^{n+}$  such that  $d_C(\mathbf{v}, \mathbf{w}) < \varepsilon$ . Then for each  $i$  with  $1 \leq i \leq n$ ,*

$$e^{-\varepsilon} \leq \frac{v_i}{w_i} \leq e^{+\varepsilon}.$$

*Conversely, if for each  $i$  we have  $e^{-\varepsilon} \leq v_i/w_i \leq e^{+\varepsilon}$ , then  $d_C(\mathbf{v}, \mathbf{w}) \leq 2\varepsilon$ .*

*Proof.* We have from Prop. 6.10: for each  $i, j$ ,

$$e^{-\varepsilon} < \frac{v_i/w_i}{v_j/w_j} < e^{+\varepsilon}.$$



Suppose that  $v_i/w_i \geq e^{\varepsilon+a}$  for some  $a > 0$ . Thus for each  $j \neq i$ ,  $v_j/w_j > e^a$ . But then defining  $c = \min\{e^\varepsilon, e^a\} > 1$ , we have  $v_k > c \cdot w_k$  for every  $k = 1, \dots, n$ . So  $\sum v_k > c \sum w_k$ , yet  $\sum v_k = \sum w_k = 1$  giving a contradiction.

On the other hand if for each  $k$ ,  $e^{-\varepsilon} \leq v_k/w_k \leq e^{+\varepsilon}$ , then

$$e^{-2\varepsilon} \leq \frac{v_i/w_i}{v_j/w_j} \leq e^{+2\varepsilon}$$

giving the converse.  $\square$

The proof of the next result, needed in the proof of Lemma 4.4, is immediate from part (b) of Theorem 6.8:

**Corollary 6.12.** *Let  $C = \mathbb{R}^{n+}$  and let  $\mathbf{z} = (z_1, \dots, z_n) \in \mathbb{R}^n$  with  $z_i \neq 0$  for all  $i$ , and write  $Z$  for the diagonal matrix with entries  $Z_{ii} = z_i$ . Then  $Z : \mathbb{R}^{n+} \rightarrow \mathbb{R}^{n+}$  is an isometry in the projective metric  $d_C$  on  $\mathbb{R}^{n+}$ .  $\square$*

For this case of  $C = \mathbb{R}^{n+}$  one has an explicit formula for the opening  $\Theta$ , and hence for the contraction coefficient  $\tanh(\Theta/4)$ , also due to Birkhoff. We extend Birkhoff's statement of this from square to rectangular matrices, and also allow nonreduced matrices. The proof of (i) is basically that of [Bir67], though at one point the argument is made simpler by the use of Lemma 6.7.

**Proposition 6.13.** *Let  $V = \mathbb{R}^n, W = \mathbb{R}^m$  with their standard cones  $C = \mathbb{R}^{n+}, D = \mathbb{R}^{m+}$ . Let  $L$  be a not identically zero  $(m \times n)$  non-negative matrix. We denote by  $d_C, d_D$  the metrics for these cones. Then the opening of the operator  $L$  (i.e. the  $d_D$ -diameter of  $L(C)$ ) is:*

$$\Theta(L) = \max_{i,j,k,l} \log \frac{L_{ik}L_{jl}}{L_{jk}L_{il}}. \quad (18)$$

*Proof.* We write  $\mathbf{e}_i$  for a standard basis vector of  $\mathbb{R}^n$ , with 1 in the  $i^{\text{th}}$  coordinate, 0 elsewhere. The  $\mathbf{e}_i$  generate the positive cone  $C$  so  $\{L(\mathbf{e}_i) \neq \mathbf{0}, 1 \leq i \leq n\}$  generate  $L(C)$ . By Lemma 6.7 the diameter of this image cone is

$$\max\{d_D(L(\mathbf{e}_k), L(\mathbf{e}_l)) : L(\mathbf{e}_i) \neq \mathbf{0}\}$$

Now we know from Proposition 6.10 that for any  $\mathbf{v}, \mathbf{w} \in C \setminus \{\mathbf{0}\}$ ,

$$d_C(\mathbf{v}, \mathbf{w}) = \log \max_{i,j} \frac{v_i w_j}{w_i v_j}$$

Therefore, for any  $\mathbf{v}, \mathbf{w} \in C$  with  $L(\mathbf{v}), L(\mathbf{w}) \neq \mathbf{0}$ ,

$$d_D(L\mathbf{v}, L\mathbf{w}) = \log \max_{i,j} \frac{L(\mathbf{v})_i L(\mathbf{w})_j}{L(\mathbf{v})_j L(\mathbf{w})_i}.$$

Now  $L(\mathbf{e}_k)_i = L_{ik}$ , so substituting  $\mathbf{v} = \mathbf{e}_k$  and  $\mathbf{w} = \mathbf{e}_l$  we have

$$d_D(L(\mathbf{e}_k), L(\mathbf{e}_l)) = \log \max_{i,j} \frac{L_{ik}L_{jl}}{L_{jk}L_{il}}.$$

Thus the diameter of  $L(C)$  is

$$\max_{k,l} d_D(L(\mathbf{e}_k), L(\mathbf{e}_l)) = \max_{k,l} \log \max_{i,j} \frac{L_{ik}L_{jl}}{L_{jk}L_{il}} = \max_{i,j,k,l} \log \frac{L_{ik}L_{jl}}{L_{jk}L_{il}}.$$

□

As a consequence of Prop. 6.13 we have immediately the following, which we use in the proof of Lemma 4.7.

**Corollary 6.14.**

(i) Let  $L$  be a not identically zero  $(m \times n)$  non-negative real matrix, and let  $V = \mathbb{R}^n, W = \mathbb{R}^m$  and  $C = \mathbb{R}^{n+}, D = \mathbb{R}^{m+}$  as in the proposition. Then the opening of  $L$  equals that of its transpose, that is:

$$\Theta(L) = \Theta(L^t);$$

i.e. the projective diameter in  $C$  of  $L^t(D)$  equals the projective diameter in  $D$  of  $L(C)$ . Hence the contraction coefficients of  $L$  and  $L^t$  are the same.

(ii) A non-negative real  $(l_i \times l_{i+1})$  sequence  $(L_i)_{i \geq 0}$  is focussing if and only if it is geometrically Perron-Frobenius.

*Proof.* Part (ii) follows directly from part (i). □

Finally regarding not necessarily reduced integer matrices, we have:

**Corollary 6.15.**

(i) Let  $L_0, L_1, \dots, L_n$  be an  $(l_i \times l_{i+1})$  non-negative integer matrix sequence with product  $L_0 L_1 \cdots L_n$  not identically zero and let  $\widehat{L}_0, \widehat{L}_1, \dots, \widehat{L}_n$  be its reduced version. Then  $\Theta(L_0 L_1 \cdots L_n) = \Theta(\widehat{L}_0 \widehat{L}_1 \cdots \widehat{L}_n)$ .

(ii) Suppose  $(L_i)_{i \geq 0}$  is a non-negative integer matrix sequence with  $\Sigma_{(L)}^{0,+} \neq \emptyset$ . Then  $(L_i)_{i \geq 0}$  is focussing if and only if the reduced sequence  $(\widehat{L}_i)_{i \geq 0}$  is.

*Proof.* To prove (i), if we only have one matrix  $L_0$ , then to form  $\widehat{L}_0$  we have removed all identically zero rows and columns. But since only the nonzero rows and columns count in the formula (18), since by convention from Proposition 6.10,  $\frac{0}{0} = 1$ , therefore indeed  $\Theta(L_0) = \Theta(\widehat{L}_0)$ .

Next recall how given  $L_0, L_1, \dots, L_n$  we form the reduced sequence  $\widehat{L}_0, \widehat{L}_1, \dots, \widehat{L}_n$ . First we apply the operator  $\mathcal{R}_n$  which removes the identically zero rows from  $L_n$  and corresponding columns from  $L_{n-1}$ . Call the new sequence  $L_0, L_1, \dots, L_{n-2} \widetilde{L}_{n-1} \widetilde{L}_n$ ; note that in fact the dimensions and entries of the product are unchanged:  $L_0 L_1 \cdots L_n = L_0 L_1 \cdots L_{n-2} \widetilde{L}_{n-1} \widetilde{L}_n$ . Then we apply  $\mathcal{R}_{n-1}$ , and so on; finally apply  $\mathcal{R}_1$  which affects  $L_1$  and  $L_0$ . Next apply  $\mathcal{C}_0$  up to  $\mathcal{C}_{n-1}$ . Write the resulting sequence as  $\check{L}_0, \check{L}_1, \dots, \check{L}_n$ . Inductively, the product is still unchanged. Perhaps the sequence is not yet reduced; this is because we have not yet applied  $\mathcal{R}_0$  and  $\mathcal{C}_n$ . These last operations will, finally, change the dimensions of the product. However, note that  $\check{L}_0 \check{L}_1 \cdots \check{L}_n$  has a zero row iff  $\check{L}_0$  does, and similarly it has a zero column iff  $\check{L}_n$  does. But by the case just discussed of a single matrix,  $\Theta(\widehat{L}_0 \widehat{L}_1 \cdots \widehat{L}_n) = \Theta(\check{L}_0 \check{L}_1 \cdots \check{L}_n)$  and this equals  $\Theta(L_0 L_1 \cdots L_n)$  since there the product is unchanged.

We next prove (ii). Suppose that  $(L_i)_{i \geq 0}$  is focussing, so given  $\varepsilon > 0$  let  $n_0$  be such that the opening of  $L_0 \cdots L_n < \varepsilon$  for all  $n > n_0$ . Write  $\widehat{L}_0^{(n)}, \widehat{L}_1^{(n)}, \dots, \widehat{L}_n^{(n)}$  for the reduced finite string. By part (i),  $\Theta(L_0 \cdots L_n) = \Theta(\widehat{L}_0^{(n)} \widehat{L}_1^{(n)} \cdots \widehat{L}_n^{(n)})$ . Now let  $m > n$  and consider the sequence  $\widehat{L}_0^{(n)}, \widehat{L}_1^{(n)}, \dots, \widehat{L}_n^{(n)}, L_{n+1}, \dots, L_m$ . The projective diameter of the image of  $C_m^+$  by this product is less than that of the image of  $C_n^+$  by  $\widehat{L}_0^{(n)} \widehat{L}_1^{(n)} \cdots \widehat{L}_n^{(n)}$  a fortiori, since  $L_{n+1} \cdots L_m C_m^+ \subseteq C_n^+$ . Now reduce that finite string, producing  $\widehat{L}_0^{(m)}, \widehat{L}_1^{(m)}, \dots, \widehat{L}_m^{(m)}$ ; the opening has not changed, and hence is less than  $\varepsilon$ . And finally we know that for  $m$  sufficiently large the reducing procedure has stabilized and the string  $L_0, \dots, L_n$  is no longer affected, so this implies that  $\varepsilon$  provides an upper bound beyond that time, proving that the reduced sequence is focussing.

Conversely, let  $n_0$  be such that for  $n > n_0$ ,  $\Theta(\widehat{L}_0 \widehat{L}_1 \cdots \widehat{L}_n) < \varepsilon$ . Now let  $m > n$  be such that the reduction of  $L_0, \dots, L_n$  has stabilized. Thus

$$\widehat{L}_0^{(m)}, \widehat{L}_1^{(m)}, \dots, \widehat{L}_n^{(m)}, \widehat{L}_{n+1}^{(m)}, \dots, \widehat{L}_m^{(m)} = \widehat{L}_0, \widehat{L}_1, \dots, \widehat{L}_n, \widehat{L}_{n+1}^{(m)}, \dots, \widehat{L}_m^{(m)};$$

the opening of their product is a fortiori  $< \varepsilon$  and is equal to that of  $L_0 \cdots L_m$ . Hence the sequence  $(L_i)_{i \geq 0}$  is focussing as well.  $\square$

## 7. APPENDIX: CONNECTIONS WITH THE THEORY OF INHOMOGENEOUS MARKOV CHAINS

A general reference here is [Sen81]. Given a non-negative real  $(l_i \times l_{i+1})$  matrix sequence  $(M_i)_{i \geq 0}$  as above for  $k \leq n$  we write:

$$M^{(k,n)} = M_k \cdots M_{n-1}.$$

Following Hajnal [Haj76] we say:

**Definition 7.1.**  $M^{(k,n)}$  tends to row proportionality iff for all  $i, k \leq l_k$  there exists  $q_{i,k}^{(k,n)} > 0$  such that

$$\lim_{n \rightarrow \infty} \frac{M_{ij}^{(k,n)}}{M_{kj}^{(k,n)}} = q_{i,k}^{(k,n)}. \quad (19)$$

If for all  $k \geq 0$  this holds, we say (following [CN90]) the sequence  $(M_i)_{i \geq 0}$  is weakly ergodic.

The Birkhoff contraction coefficient of a matrix  $M$  (see Cor. 6.14 above) is written  $\tau_B(M)$ .

We mention that some authors use the term weak ergodicity for a different (in general weaker) ‘‘additive’’ condition instead, see [Sen81]. For our purposes however the interesting condition is the ‘‘multiplicative’’ condition (19). As Hajnal in Theorem 1 of [Haj76] showed, for  $M_i$  with no identically zero rows (a condition known as *row allowable*), weak ergodicity is equivalent to:

$$\text{for all } k \geq 0, \tau_B(M^{(k,n)}) \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (20)$$

For a quick proof of this statement, by Prop. 6.10 above, if two rows of  $M^{(k,n)}$  are  $\mathbf{v}, \mathbf{w}$  then

$$d_C(\mathbf{v}, \mathbf{w}) = \log \max_{i,j} \frac{v_i w_j}{w_i v_j}$$

so

$$d_C(\mathbf{v}, \mathbf{w}) < \varepsilon$$

iff  $\frac{v_i}{w_i} = (1 \pm \hat{\varepsilon}) \frac{v_j}{w_j}$ , so the rows tend to proportionality.

Now condition (20) is exactly what we call focussing, so this shows that for reduced sequences weak ergodicity is equivalent to focussing.

(We prefer to stick with the term *focussing* rather than *weakly ergodic*, especially in a context where nonstationary matrix sequences meet ergodic theory itself, since as we have seen, this condition on the matrices implies (strong) mixing for the Parry measure sequence. Indeed Hajnal [Haj76] had suggested *contractive* might be more appropriate, and *focussing* is in that spirit.)

In [CN90] Cohn and Nerman introduce the use of positive column eigenvector sequences with eigenvalue one  $(h_i)_{i \geq 0}$ , so  $h_i = M_i h_{i+1}$ ; they call  $(h)$  an  $(M)$ -harmonic function. Just as for us, such a sequence is used to produce from a non-negative real matrix sequence  $(M_i)_{i \geq 0}$  a stochastic sequence  $(P_i)$ , by the same formula of (6); while we had converted this from the stationary case (6) of Parry (following especially the treatment in [AW70]), they refer instead to Perron and to Feller [Fel56] (again for the stationary case). (The focus in [CN90] is on real non-negative matrices, not integer matrices and so not on subshifts). They then define:

**Definition 7.2.** *A positive  $(M)$ -harmonic function  $(h)$  is extremal iff it is unique up to positive multiples.*

Thus,  $(M_i)$  has a positive extremal  $(h)$  if and only if the eigenvector Perron-Frobenius condition is satisfied. They then prove that  $(M_i)_{i \geq 0}$  is weakly ergodic if and only if it has an extremal  $(h)$ , Theorem 3 together with the lemma before Theorem 2 in [CN90]. So this is exactly the equivalence of eigenvector Perron-Frobenius and focussing. Their proof is completely different from ours; ours in Lemma 4.7 is purely geometrical, relying on Corollary 6.14 of Appendix 6, while theirs goes by way of the tail  $\sigma$ -fields. These observations give the corollary stated at the end of the introduction.

Finally we note that while very often in nonhomogeneous Markov chain theory the matrices are taken to be square and the sequence is assumed to be *allowable* (the matrices have no identically zero rows or columns, so for non-negative integer matrices this is what we call reduced), we try whenever possible to remove these restrictions (for example in Proposition 6.13 and Corollaries 6.14, 6.15 above) to allow for a wider scope of examples.

## REFERENCES

- [AF01] P. Arnoux and A. M. Fisher. The scenery flow for geometric structures on the torus: the linear setting. *Chinese Ann. of Math.*, 4:427–470, 2001.
- [AF05] P. Arnoux and A. M. Fisher. Anosov families, renormalization and nonstationary subshifts. *Erg. Th. and Dyn. Sys.*, 25:661–709, 2005.
- [AW70] Roy L. Adler and Benjamin Weiss. Similarity of automorphisms of the torus. *Memoirs of the American Mathematical Society*, 98:1–43, 1970.
- [Bil65] Patrick Billingsley. *Ergodic Theory and Information*. John Wiley and Sons (republished 1978 by Krieger), 1965.
- [Bir57] Garrett Birkhoff. Extensions of Jentzsch’s theorem. *Trans. AMS*, 85:219–227, 1957.

- [Bir67] G. Birkhoff. *Lattice Theory*, volume XXV of *AMS Colloq. Publ.* AMS, 3rd edition, 1967. Chapter XVI.
- [BKMS09] S. Bezuglyi, J. Kwiatkowski, K. Medynets, and B. Solomyak. Invariant measures on stationary Bratteli diagrams. *Ergodic Theory and Dynamical Sys.*, 2009. To appear.
- [BM77] Rufus Bowen and Brian Marcus. Unique ergodicity for horocycle foliations. *Israel Jour. Math.*, 26(1):43–67, 1977.
- [CN90] H. Cohn and O. Nerman. On products of non-negative matrices. *Ann. Prob.*, 18:1806–1815, 1990.
- [ES80] Edward G. Effros and Chao-Liang Shen. Approximately finite  $C^*$ -algebras and continued fractions. *Indiana University Math. Journal*, 29(2):191–204, 1980.
- [Fel56] W. Feller. Boundaries induced by non-negative matrices. *Trans. Amer. Math. Soc.*, 83:19–54, 1956.
- [Fer97] Sebastien Ferenczi. Systems of finite rank. *Colloq. Math.*, 73(1):35–65, 1997.
- [FFT09] S. Ferenczi, A. M. Fisher, and M. Talet. Minimality and unique ergodicity for adic transformations, 2009. To appear, *Journal d'Analyse*.
- [Fisa] A. M. Fisher. Abelian differentials, interval exchanges, and adic transformations. In preparation, 2009.
- [Fisb] A. M. Fisher. Duality of measure and velocity for nonstationary solenoids. In preparation, 2009.
- [Fis92] A. M. Fisher. Integer Cantor sets and an order-two ergodic theorem. *Ergod. Th. and Dynam. Sys.*, 13:45–64, 1992.
- [For97] A.H. Forrest. K-groups associated with substitution minimal systems. *Isr. Jour. Math.*, 98:101–139, 1997.
- [FP08] S.B. Frick and K. Petersen. Random permutations and unique fully supported ergodicity for the Euler adic transformation. *Annales de l' Institut H. Poincaré, Probabilités et Statistiques*, 44(5):876–885, 2008.
- [Fur60] Harry Furstenberg. *Stationary Processes and Prediction Theory*. Number 44 in *Annals of Mathematics Studies*. Princeton, 1960.
- [Fur61] Hillel Furstenberg. Strict ergodicity and transformation of the torus. *Amer Jour Math*, 83:573–601, 1961.
- [Fur73] Hillel Furstenberg. The unique ergodicity of the horocycle flow. In A. Beck, editor, *Recent Advances in Topological Dynamics*, volume 318 of *Springer Lecture Notes*, pages 95–115. Springer Verlag, 1973.
- [Fur81] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory*. Princeton University Press, Princeton, 1981.
- [GLT95] P. Grabner, P. Liardet, and R. Tichy. Odometers and systems of numeration. *ACTA ARITHMETICA*, LXX.2:103–123, 1995.
- [Haj76] J. Hajnal. On products of non-negative matrices. *Math. Proc. Camb. Phil. Soc.*, 79:521–530, 1976.
- [HIK72] A. Hajian, Y. Ito, and S. Kakutani. Invariant measure and orbits of dissipative transformations. *Advances in Math*, 9(1):52–65, 1972.
- [Hos00] Bernard Host. Substitution subshifts and Bratteli diagrams. In A. Maas F. Blanchard and A. Nogueira, editors, *Topics in Symbolic Dynamics and Applications*, number 279 in *London Math. Soc. Lecture Note Series*. Cambridge Univ. Press, 2000.
- [Kam] Teturo Kamae. Numeration systems as dynamical systems. preprint.
- [Kea77] Michael Keane. Non-ergodic interval exchange transformations. *Israel Jour. Math.*, 26(2):188–196, 1977.
- [Ker85] Steve Kerchhoff. Simplicial systems for interval exchange maps and measured foliations. *Ergod. Th. and Dynam. Sys.*, 5:257–271, 1985.
- [Kit98] Bruce P. Kitchens. *Symbolic Dynamics*. Springer, 1998.
- [Liv92] A. N. Livshits. On the construction of Markov compacta. *St. Petersburg Math. J.*, 3(1):137–146, 1992.

- [LM95] Douglas Lind and Brian Marcus. *Symbolic dynamics and coding*. Cambridge University Press, 1995.
- [LV85] A.A. Lodkin and A.M. Vershik. Approximation for actions of amenable groups and transversal automorphisms. In H. Araki, C.C. Moore, S. Stratila, and D. Voiculescu, editors, *Operator Algebras and their connections with Topology and Ergodic Theory*, volume 1132 of *Lecture Notes in Math.*, pages 331–346. Springer Verlag, 1985.
- [LV92] A. N. Livshits and A. M. Vershik. Adic models of ergodic transformations, spectral theory, substitutions, and related topics. *Advances in Soviet Mathematics*, 9:185–204, 1992. Representation theory and Dynamical Systems.
- [Mas82] Howard Masur. Interval exchange transformations and measured foliations. *Annals of Math.*, 115:169–200, 1982.
- [MP05] Xavier Mela and Karl Petersen. Dynamical properties of the Pascal adic transformation. *Ergodic Theory Dynam. Systems*, 25(1):227–256, 2005.
- [Par64] W. Parry. Intrinsic Markov chains. *Trans. AMS*, 112:55–66, 1964.
- [Sam56] H. Samelson. On the Perron-Frobenius Theorem. *Mich. Math. J.*, 4:57–59, 1956.
- [Sen81] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer, second edition, 1981.
- [SW63] Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. Univ. of Illinois Press, Urbana, Chicago, London, 1963. Appendix 4.
- [TGS95] Ian F. Putnam Thierry Giordano and Christian F. Skau. Topological orbit equivalence and  $C^*$ -crossed products. *J. reine angew. Math.*, 469:51–111, 1995.
- [Vee78] William A. Veech. Interval exchange transformations. *Journal d'Analyse Mathématique*, 33:222–272, 1978.
- [Vee82] William A. Veech. Gauss measures for transformations on the space of interval exchange maps. *Annals of Mathematics*, 115:201–242, 1982.
- [Ver81] Anatoly M. Vershik. Uniform algebraic approximation of shift and multiplication operators. *Soviet Math. Dokl.*, 24(1):101–103, 1981.
- [Ver89] A. M. Vershik. A new model of the ergodic transformations. In *Dynamical Systems and Ergodic Theory*, volume 23 of *Banach Center Publications*, pages 381–384, Warsaw, 1989. PWN- Polish Scientific Publishers.
- [Ver95a] Anatoly M. Vershik. The adic realizations of the ergodic actions with the homeomorphisms of the Markov compact and the ordered Bratteli diagrams. *Teor. Predstav. Din. Systemy. Kombin. i Algoritm. Metody.*, 1:120–126, 1995.
- [Ver95b] Anatoly M. Vershik. Locally transversal symbolic dynamics. *St. Petersburg Math Journal*, 6:529–540, 1995. Translation of 1994 Russian version.
- [Via06] Marcelo Viana. Dynamics of interval exchange transformations and Teichmüller flows. lecture notes, August 2006. <http://w3.impa.br/~viana/out/ietf.pdf>.
- [Wal82] P. Walters. *An Introduction to Ergodic Theory*. Springer Verlag, New York/Berlin, 1982.

ALBERT M. FISHER, DEPT MAT IME-USP, RUA DO MATÃO 1010, CEP 05508-090, SÃO PAULO SP, BRAZIL

URL: <http://ime.usp.br/~afisher>

E-mail address: [afisher@ime.usp.br](mailto:afisher@ime.usp.br)