

Maximal Measures for Simple Piecewise Monotonic Transformations

Franz Hofbauer

Mathematisches Institut der Universität, Strudlhofgasse 4, A-1090 Wien

Summary. It is shown that the transformation $x \mapsto \beta x + \alpha \pmod{1}$ ($\beta > 1, 0 \leq \alpha < 1$) on $[0, 1]$ has unique maximal measure.

§0. Introduction

In this paper we consider two classes of piecewise monotonic transformations f on $I = [0, 1]$. The first one is the class of all (I, f) such that $I = J_1 \cup J_2$, J_1, J_2 are disjoint intervals and $f/J_1, f/J_2$ are continuous and strictly increasing. Furthermore we assume that (J_1, J_2) is a generator for (I, f) and that $h_{\text{top}}(f) > 0$. The second one is the class of transformations $f: x \mapsto \beta x + \alpha \pmod{1}$ on I where $\beta > 1$ and $0 \leq \alpha < 1$.

An invariant measure μ on (I, f) is called maximal if its entropy $h(\mu)$ is equal to the topological entropy $h_{\text{top}}(f)$ of (I, f) , or equivalently, if $h(\mu)$ is greater than or equal to the entropy of every other invariant measure on (I, f) . The two classes of transformations above have always at least one maximal measure, because they are expansive. For definitions see [3].

We show that in the first case above there are at most two ergodic maximal measures and characterize those f 's which have unique maximal measure. In the second case, for $f: x \mapsto \beta x + \alpha \pmod{1}$, we have always unique maximal measure.

To this end we use the results proved in [2]. For every (I, f) a subshift Σ_M of finite type (M is the corresponding transition matrix) over a countable alphabet D is constructed there such that Σ_M and (I, f) have isomorphic sets of maximal measures. Hence it suffices to consider the problem of uniqueness of the maximal measure for Σ_M . A tool for this is the theorem cited in §1 which contains also a description of other results from [2]. Furthermore, for the above two classes of transformations the transition matrices M are derived which are considered as oriented graphs. The results of this paper are proved in §2. We determine conditions for M such that Σ_M can have more than one maximal measure (Lemmas 2 and 3) and characterize the (I, f) of the first class above which give rise to such an M (Theorem 1). Then we prove that such an M cannot occur for $f: x \mapsto \beta x + \alpha \pmod{1}$ (Theorem 2).

§1.

We begin with a short description of the methods developed in [2]. We have a piecewise monotonic transformation f on I , i.e. $I = \bigcup_{i=1}^n J_i$, J_i disjoint intervals, such that f/J_i is continuous and strictly increasing. Furthermore we need that $(J_i)_{1 \leq i \leq n}$ is a generator for (I, f) and that $h_{\text{top}}(f) > 0$. The f -expansion $\varphi: (I, f) \rightarrow \Sigma_n^+ = \{1, 2, \dots, n\}^{\mathbb{N}}$ is defined by $\varphi(x) = i_0 i_1 i_2 \dots$, where i_k is the number i of the interval J_i such that $f^k(x) \in J_i$. If $J_k = (r, s)$, set $\underline{a}^k = \lim_{t \downarrow r} \varphi(t)$ and $\underline{b}^k = \lim_{t \uparrow s} \varphi(t)$. Define

$$\Sigma_f^+ = \{ \underline{x} \in \Sigma_n^+ : \underline{a}^{x_m} \leq \sigma^m \underline{x} = x_m x_{m+1} x_{m+2} \dots \leq \underline{b}^{x_m} \text{ for all } m \geq 0 \},$$

where σ denotes the shift transformation and \leq the lexicographic ordering in Σ_n^+ . Then φ is an order preserving isomorphism modulo small sets (cf. §0 of [2]) between (I, f) and (Σ_f^+, σ) . In particular, (I, f) and (Σ_f^+, σ) have isomorphic sets of measures with maximal entropy.

Now divide \underline{a}^i into initial segments of \underline{b}^j 's, which can be done in a unique way, because \underline{b}^j is the only element among the \underline{b}^k 's which begins with j , and denote the lengths of these segments by $r(i, 1), r(i, 2), \dots$, i.e. we have for $m \geq 0$ $a_{r(i, 1) + \dots + r(i, m) + k}^i = b_k^j$, $0 \leq k \leq r(i, m + 1) - 1$, $j = a_{r(i, 1) + \dots + r(i, m)}^i$ and $a_{r(i, 1) + \dots + r(i, m + 1)}^i \neq b_{r(i, m + 1)}^j$.

Similarily divide \underline{b}^j into initial segments of \underline{a}^i 's, and denote their lengths by $s(j, 1), s(j, 2), \dots$. Consider the set

$$D = \{(A, i, k), (B, i, k) : 1 \leq i \leq n, k \geq 1\} = \{A, B\} \times \{1, \dots, n\} \times \mathbb{N},$$

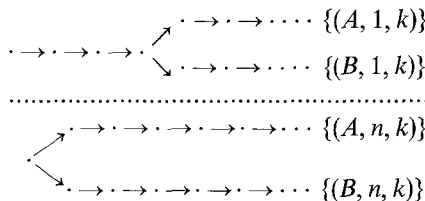
identify $(A, i, k) = (B, i, k)$ for $1 \leq k \leq r(i, 1) = s(i, 1)$ and also

$$\begin{aligned} (A, i, p+k) &= (A, j, q+k) \forall k \geq 1, \text{ if } \sigma^p \underline{a}^i = \sigma^q \underline{a}^j, p = \sum_{m=1}^u r(i, m), q = \sum_{m=1}^v r(j, m) \\ (B, i, p+k) &= (B, j, q+k) \forall k \geq 1, \text{ if } \sigma^p \underline{b}^i = \sigma^q \underline{b}^j, p = \sum_{m=1}^u s(i, m), q = \sum_{m=1}^v s(j, m). \end{aligned} \tag{1.1}$$

Together with the following arrows D becomes a graph M .

$$\begin{aligned} (A, i, k) &\rightarrow (A, i, k+1), (B, j, k) \rightarrow (B, j, k+1). \\ \text{If } k &= r(i, 1) + \dots + r(i, m) \text{ for some } m, \text{ then} \\ (A, i, k) &\rightarrow (B, j, r(i, m) + 1), \text{ where } j = a_{r(i, 1) + \dots + r(i, m-1)}^i \\ \text{and } (A, i, k) &\rightarrow (A, t, 1) = (B, t, 1) \text{ for } a_k^i < t < b_{r(i, m)}^i. \\ \text{If } k &= s(j, 1) + \dots + s(j, m) \text{ for some } m, \text{ then} \\ (B, j, k) &\rightarrow (A, i, s(j, m) + 1), \text{ where } i = b_{s(j, 1) + \dots + s(j, m-1)}^j \\ \text{and } (B, j, k) &\rightarrow (A, t, 1) = (B, t, 1) \text{ for } a_{s(j, m)}^i < t < b_k^j. \end{aligned} \tag{1.2}$$

The graph looks like this



Only the arrows $(A, i, k) \rightarrow (A, i, k + 1)$ and $(B, j, k) \rightarrow (B, j, k + 1)$ are indicated in this picture.

Set $\Sigma_M = \{y \in D^{\mathbb{Z}} : \text{there is an arrow from } y_i \text{ to } y_{i+1} \forall i \in \mathbb{Z}\}$. Then (Σ_f, σ) , the natural extension of (Σ_f^+, σ) , and (Σ_M, σ) are isomorphic modulo small sets. In particular (Σ_f, σ) and (Σ_M, σ) and hence also (I, f) and (Σ_M, σ) have isomorphic sets of maximal measures.

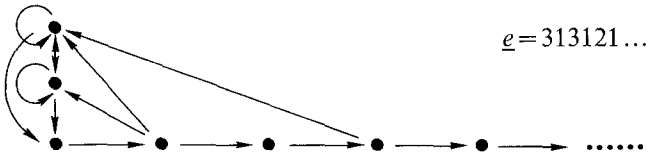
Divide M into irreducible subgraphs M_1, M_2, \dots and denote the corresponding subshifts of Σ_M by Σ_{M_i} . We consider M as 0-1 matrix with index set D , $M_{jk} = 1$ iff there is an arrow $j \rightarrow k$.

Theorem. (i) $h_{\text{top}}(\Sigma_M) = \log r(M)$ (spectral radius of the l^1 -operator M).

(ii) Every ergodic maximal measure is concentrated on a Σ_{M_i} satisfying $r(M_i) = r(M)$ (the ergodic maximal measures are the extremal points of the compact convex set of all maximal measures).

(iii) There is at most one (ergodic) maximal measure on every such Σ_{M_i} .

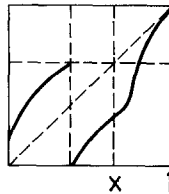
Our goal is to apply these results to special piecewise monotonic transformations. The simplest nontrivial example is the β -transformation $x \mapsto \beta x \pmod{1}$ for $\beta > 1$. Let n be so that $\beta \leq n < \beta + 1$ and e the β -expansion of 1. Then $a^i = i1111\dots$ for $1 \leq i \leq n$, $b^i = ie$ for $1 \leq i \leq n - 1$ and $b^n = e$. Σ_f^+ becomes $\Sigma_\beta^+ = \{x \in \Sigma_n^+ : \sigma^k x \leq e \forall k\}$. Using the identifications (1.1) we get for $n = 3$ the following graph



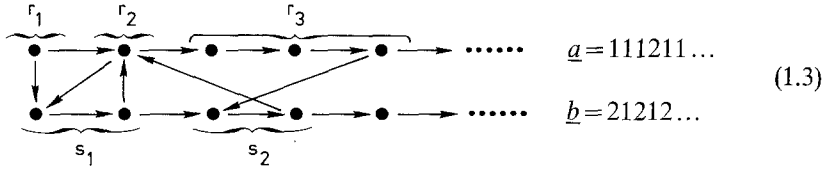
There are e_i arrows starting at the i -th point of the row. This gives an irreducible graph (cf. [1]). Hence the β -shift Σ_β^+ has unique maximal measure.

One sees that, due to the identifications, only one of the $2n$ rows has remained. In the next more complicated case we consider the situation where there are two remaining rows.

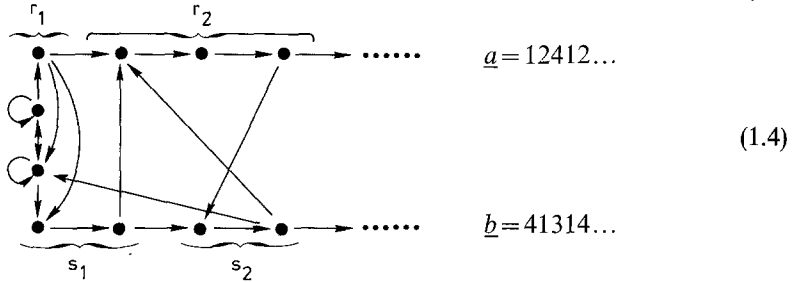
Consider any piecewise monotonic transformation with $n = 2$. We assume that the end point of J_1 is mapped to 1 and that the initial point of J_2 is mapped to 0. We can reduce all other cases to this case taking away wandering sets and fixed points. For example consider the following graph of an f . We take away $(x, 1]$, where $x = f(\text{end point of } J_1)$. $(x, 1)$ is a wandering set and 1 is a fixed point. $f(x) \leq x$, otherwise (J_1, J_2) is no generator.



Let $\underline{a}^1 = \underline{a}$ be the expansion of 0 and $\underline{b}^2 = \underline{b}$ that of 1. Then $\underline{b}^1 = 1\underline{b}$ and $a^2 = 2\underline{a}$. Set $r_k = r(1, k)$ and $s_k = s(2, k)$. By (1.1) we identify $(A, 2, k + 1) = (A, 1, k)$ for $k \geq r_1$ and $(B, 1, k + 1) = (B, 2, k)$ for $k \geq s_1$ and get the graph



Also the transformation $x \mapsto \beta x + \alpha \pmod{1}$ for $\beta > 1$ and $0 < \alpha < 1$ gives such diagrams. Again let $\underline{a}^1 = \underline{a}$ be the expansion of 0 and $\underline{b}^n = \underline{b}$ the expansion of 1 (n so that $\alpha + \beta \leq n < \alpha + \beta + 1$). Then $\underline{a}^i = i\underline{a}$ for $2 \leq i \leq n$ and $\underline{b}^i = i\underline{b}$ for $1 \leq i \leq n - 1$. Again set $r_k = r(1, k)$ and $s_k = s(n, k)$. By (1.1) we identify $(A, i, k + 1) = (A, 1, k)$ for $k \geq r(i, 1)$ and $2 \leq i \leq n$ and $(B, i, k + 1) = (B, n, k)$ for $k \geq s(i, 1)$ and $1 \leq i \leq n - 1$. We get (for $n = 4$)



We shall denote the points in the first row by $c_k = (A, 1, k)$, the points in the last row by $d_k = (B, n, k)$, and the remaining points by e_2, \dots, e_{n-1} . Set $A(i, j) = \{c_k : i \leq k \leq j\}$ ($A(i, \infty) = \{c_k : i \leq k\}$) and $B(i, j) = \{d_k : i \leq k \leq j\}$.

Hence in the cases we want to consider in this paper we have $D = \{c_k, d_k, e_j : k \geq 1, 2 \leq j \leq n - 1\}$ (for $n = 2$ there are no e_j) and M has the following arrows

$$\begin{aligned}
 c_k &\rightarrow c_{k+1}, d_k \rightarrow d_{k+1} & (k \geq 1) \\
 c_{r_1 + \dots + r_k} &\rightarrow d_{r_k} \text{ and } e_j \text{ for } a_{r_1 + \dots + r_k} < j < b_{r_k - 1} & (k \geq 1) \\
 d_{s_1 + \dots + s_k} &\rightarrow c_{s_k} \text{ and } e_j \text{ for } a_{s_k - 1} < j < b_{s_1 + \dots + s_k} & (k \geq 1) \\
 e_k &\rightarrow c_1, d_1 \text{ and } e_j \text{ for } 2 \leq j \leq n - 1 & (2 \leq k \leq n - 1).
 \end{aligned}
 \tag{1.5}$$

Furthermore by (1.1) we can identify $(u, v \geq 0)$

$$\begin{aligned}
 c_{p+k} &= c_{q+k} \quad \forall k \geq 1, \text{ if } \sigma^p \underline{a} = \sigma^q \underline{a}, p = r_1 + \dots + r_u, q = r_1 + \dots + r_v \\
 d_{p+k} &= d_{q+k} \quad \forall k \geq 1, \text{ if } \sigma^p \underline{b} = \sigma^q \underline{b}, p = s_1 + \dots + s_u, q = s_1 + \dots + s_v.
 \end{aligned}
 \tag{1.6}$$

The two Σ_M arising from M with and without identifications are then isomorphic.

For the r_i and s_i , which denote the lengths of initial segments of $v\bar{b}$ ($1 \leq v \leq n-1$) in \underline{a} and of $u\bar{a}$ ($2 \leq u \leq n$) in \bar{b} respectively, we have the following lemma. Remark that

$$\underline{a}, \bar{b} \in \Sigma_n^+, \text{ which becomes } \{x \in \Sigma_n^+ : \underline{a} \leq \sigma^k x \leq \bar{b} \forall k \geq 0\}. \tag{1.7}$$

Lemma 1. *There are maps $P, Q: \mathbb{N} \rightarrow \mathbb{N} \cup \{0\}$, such that*

$$\begin{aligned} r_k &= s_1 + \dots + s_{P(k)} + 1 & \text{for } k \geq 1 \\ s_k &= r_1 + \dots + r_{Q(k)} + 1 & \text{for } k \geq 1. \end{aligned} \tag{1.8}$$

Proof. We prove only the first statement. Suppose that $1 + s_1 + \dots + s_m < r_k < 1 + s_1 + \dots + s_{m+1}$ for some m . We have $a_{r_1 + \dots + r_{k-1} + j} = b_{j-1}$ for $1 \leq j \leq r_k - 1$, $a_{r_1 + \dots + r_k} < b_{r_k - 1}$ and $b_{s_1 + \dots + s_{m+1} + l} = a_{l-1}$ for $1 \leq l \leq s_{m+1} - 1$ by definition of r_k, s_m .

Setting $j = s_1 + \dots + s_m + l + 1$ we have because $r_k - s_1 - \dots - s_m - 1 \leq s_{m+1} - 1$ $a_{r_1 + \dots + r_{k-1} + s_1 + \dots + s_m + l + 1} = a_{l-1}$ for $1 \leq l \leq r_k - s_1 - \dots - s_m - 2$ ($\Rightarrow j \leq r_k - 1$) and $a_{r_1 + \dots + r_k} < a_{r_k - s_1 - \dots - s_m - 2}$.

If $r_k - s_1 - \dots - s_m - 2 = 0$, we have $a_{r_1 + \dots + r_k} < a_0 = 1$, a contradiction, and $\sigma^{r_1 + \dots + r_{k-1} + s_1 + \dots + s_m + 2} \underline{a} < \underline{a}$, a contradiction to (1.7), in case of $r_k - s_1 - \dots - s_m - 2 \geq 1$. The lemma is proved.

This lemma shows that, if an arrow goes from $B(1, \infty)$ to c_i , then $i = s_k = r_1 + \dots + r_{Q(k)} + 1$ and hence by (1.5) and $i - 1 = r_1 + \dots + r_{Q(k)}$ there is an arrow from the point c_{i-1} before c_i to $B(1, \infty)$. This will be used several times in the sequel.

We want to describe another thing which will be important, namely how one can regain \underline{a} and \bar{b} from M . This can be done by induction. $a_0 = 1, b_0 = n, r_1 > 1$ and $s_1 > 1$ is a contradiction to (1.8). Hence suppose $r_1 = 1$. Then a_1 is the number such that there are arrows $c_1 \rightarrow e_j$ for $a_1 < j < b_0$ in (1.5). If $a_0 \dots a_{p-1}$ and $b_0 \dots b_{q-1}$ are already determined, where $p = r_1 + \dots + r_i + 1$ and $q = s_1 + \dots + s_k + 1$ (the above step is for $i = 0$ and $k = 0$), $r_{i+1} > s_1 + \dots + s_k + 1$ and $s_{k+1} > r_1 + \dots + r_i + 1$ is again a contradiction to (1.8). Supposing $r_{i+1} \leq s_1 + \dots + s_k + 1$, we have by the definition of r_{i+1} that $a_p \dots a_{p-2+r_{i+1}} = b_0 \dots b_{r_{i+1}-2}$, which is already known, and $a_{p-1+r_{i+1}}$ is the number such that there are arrows $c_{r_1 + \dots + r_{i+1}} \rightarrow e_j$ for $a_{p-1+r_{i+1}} < j < b_{r_{i+1}-1}$ in (1.5).

§2.

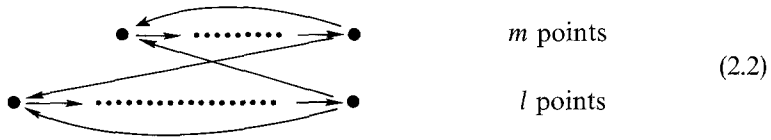
To apply the above theorem we have to divide M into irreducible subgraphs. If M is reducible we can divide D into two disjoint subsets D_1 and D_2 such that there may be transition from D_1 to D_2 , but not from D_2 to D_1 . D_2 is not empty. Suppose $c_i \in D_2$ and l is the smallest integer with this property. Then $A(l, \infty) \subset D_2$, because there is no transition from D_2 to D_1 . If there is no arrow from $A(l, \infty)$ to $B(1, \infty)$, then $D_2 = A(l, \infty)$ has spectral radius 1 and hence we can take it away. We assume that there is such an arrow and hence $a d_m \in D_2$. Again let m be the smallest integer with this property. As above $B(m, \infty) \subset D_2$ and $D_2 = A(l, \infty) \cup B(m, \infty)$.

Let λ_0 be the largest real solution of $\lambda^{l+m} - \lambda^l - \lambda^m = 0$ and $M_2 = M/D_2$. We have

Lemma 2. $r(M_2) \leq \lambda_0$. Equality holds, iff there are i, j , such that

$$\begin{aligned} r_1 + \dots + r_{i-1} &= l - 1, & s_1 + \dots + s_{j-1} &= m - 1 & \text{and} \\ s_j = s_{j+1} = \dots &= l, & r_i = r_{i+1} = \dots &= m. \end{aligned} \tag{2.1}$$

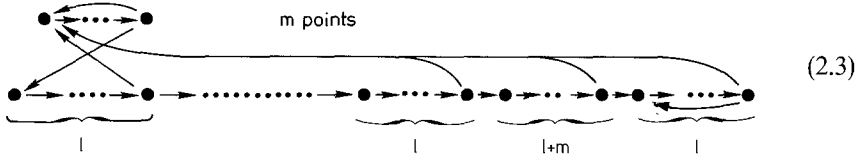
Proof. We can assume that there are arrows arriving at c_i and d_m . If not, let $l' \geq l$ and $m' \geq m$ be the smallest integers, such that there are arrows from $B(m, \infty)$ to $c_{l'}$, and from $A(l, \infty)$ to $d_{m'}$. Then we can never return from $A(l', \infty) \cup B(m', \infty)$ to $c_1, \dots, c_{l'-1}$ and to $d_m, \dots, d_{m'-1}$ and hence we can take away this points from D_2 without changing the spectral radius of M_2 . Hence, if we have proved the result with this assumption, $r(M_2) \leq \lambda'_0 < \lambda_0$, where λ'_0 is the largest solution of $\lambda^{l'+m'} - \lambda^{l'} - \lambda^{m'} = 0$. Now we assume that there are arrows arriving at c_i and d_m , i.e. there are i, j , such that $l = s_u = r_1 + \dots + r_{i-1} + 1$ and $m = r_v = s_1 + \dots + s_{j-1} + 1$ for some u, v (cf. Lemma 1). If now (2.1) is satisfied, we have $\sigma^{l+m} \underline{a} = \sigma^l \underline{a}$ and $\sigma^{l+m} \underline{b} = \sigma^m \underline{b}$, i.e. $c_{k+m} = c_k$ for $k \geq l$ and $d_{k+1} = d_k$ for $k \geq m$ (cf. (1.1)). We get the following graph for M_2 .



Hence the characteristic equation for M_2 is $\lambda^{l+m} - \lambda^l - \lambda^m = 0$ and $r(M_2) = \lambda_0$.

If (2.1) is not satisfied we have $r_k \geq m \forall k \geq i$ and $s_k \geq l \forall k \geq j$, i, j as above (otherwise there would be transition from D_2 to D_1) and there is at least one $r_g (g \geq i)$ or $s_h (h \geq j)$ with $r_g > m (\Rightarrow r_g \geq l+m)$ because of (1.2) or $s_h > l (\Rightarrow s_h \geq l+m)$ because of (1.2)). We show that M_2 has then a spectral radius strictly less than λ_0 .

First we compute the spectral radius of M_2 for the case $r_k = m \forall k \geq i, s_k = l \forall k \geq j, k \neq h$ and $s_h = m+l$. As above (2.2) we get the diagram (identifications for $\sigma^{m+l} \underline{a} = \sigma^l \underline{a}$ and $\sigma^{2m+(h-j+2)l} \underline{b} = \sigma^{2m+(h-j+1)l} \underline{b}$)

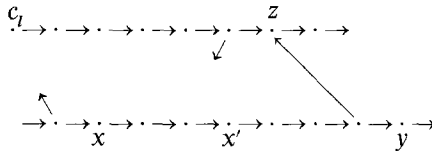


We get the characteristic polynomial

$$\begin{aligned} P(\lambda) &= \lambda^{(h-j+2)l+2m} - \lambda^{(h-j+2)l+m} - \lambda^{(h-j+1)l+2m} \\ &\quad - \sum_{f=1}^{h-j} \lambda^{(h-j-f+2)l+m} - \lambda^l - 1 + \sum_{f=1}^{h-j} \lambda^{(h-j-f+1)l+m} \\ &\quad + 1 + \lambda^{(h-j+1)l+2m} \\ &= \lambda^{(h-j+1)l+m} (\lambda^{l+m} - \lambda^l - \lambda^m) + \lambda^{l+m} - \lambda^l. \end{aligned}$$

It is easy to see that $P(\lambda) \geq \lambda_0^m > 0$ for all $\lambda \geq \lambda_0$. Hence the largest solution of $P(\lambda) = 0$ is less than λ_0 , i.e. $r(M_2) < \lambda_0$. The same argument holds if $r_g = l + m$ instead of $s_h = m + l$. For the general case we can find a minimal $g \geq i$ or a minimal $h \geq j$ such that $r_g \geq l + m$ or $s_h \geq l + m$ respectively. We consider the second case. The first one is similar. Denote the matrix corresponding to (2.3) by M_h (without identifications). We show $r(M_2) \leq r(M_h)$. Let N_k^x and \tilde{N}_k^x be the numbers of blocks of length k admitted by M_2 and M_h respectively and beginning with $x \in D_2$. It suffices to show $\|M_2^k\|_1 = \sup_x N_k^x \leq 2 \|M_h^k\|_1 = 2 \sup_x \tilde{N}_k^x$.

To this end set $R_u = r_1 + \dots + r_u$, $S_v = s_1 + \dots + s_v$ and $D' = \{c_{R_u+1}, d_{S_v+1} : u \geq i, v \geq j\}$. We show by induction that $N_k^x \leq \tilde{N}_k^x$, if $x \in D'$. Let $x = d_{S_v+1}$ (the case $x = c_{R_u+1}$ is similar). Cancelling the first $t := s_{v+1}$ elements, which are equal in all M_2 -blocks beginning with x , we get $N_k^x = N_{k-t}^y + N_{k-t}^z$, where $y = d_{S_{v+1}+1} \in D'$ and $z = c_{R_{Q(v+1)+1}} \in D'$ (Lemma 1). Here is a picture of the relevant part of M_2 .



If $k \leq t$, $N_k^x = 1 \leq \tilde{N}_k^x$. Otherwise, as $y, z \in D'$, we have by induction that $N_{k-t}^y \leq \tilde{N}_{k-t}^y$ and $N_{k-t}^z \leq \tilde{N}_{k-t}^z$, hence

$$N_k^x \leq \tilde{N}_{k-t}^y + \tilde{N}_{k-t}^z. \tag{2.4}$$

Now we consider the case $v + 1 \neq h$. The arguments for $v + 1 = h$ are the same replacing l by $l + m$. Because $s_{v+1} \geq l$ there must be an arrow in M_h from a point in $B(S_v + 1, S_{v+1})$ to c_l . Let $x' = d_{S_v+t'}$ be the first such point ($t' \leq l$). Then $x'c_l c_{l+1} \dots z$ is an admissible block in M_h of length $t - l + 2$. Hence $\tilde{N}_k^x = \tilde{N}_{k-t'+1}^z \geq \tilde{N}_{k-t'}^{c_l} + \tilde{N}_{k-t}^z$ and $\tilde{N}_{k-t'}^{c_l} \geq \tilde{N}_{k-t'-t'+1}^z \geq \tilde{N}_{k-t}^z$, because $l - t' \geq 0$. Together with (2.4) we have $\tilde{N}_k^x \geq N_k^x$. Now let $x = d_{S_{v+1}+w} \notin D'$. Then (2.4) is proved as above with $t = s_{v+1} - w$ and the same $y, z \in D'$. Therefore $N_k^x \leq 2 \sup_z \tilde{N}_k^z$. For $x \in D'$ this follows from $N_k^x \leq \tilde{N}_k^x$. This implies $\sup_x N_k^x \leq 2 \sup_x \tilde{N}_k^x$. Hence we have $r(M_2) \leq r(M_h) < \lambda_0$ and the lemma is proved.

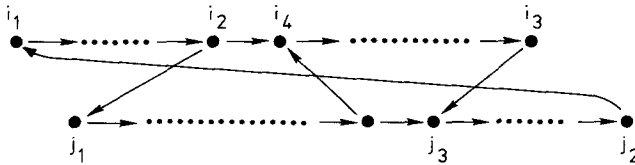
Now we consider the case (1.3) and the case (1.4) only, if $E = \{(A, i, 1) = (B, i, 1) : 2 \leq i \leq n - 1\} = \{e_2, e_3, \dots, e_{n-1}\}$ is an irreducible subset of M . We can consider the two rows $D \setminus E$ for themselves (there is no transition back to E) and D_1 and D_2 are subsets of $D \setminus E$.

Recall $A(i, j) = \{c_k : i \leq k \leq j\}$ and $B(i, j) = \{d_k : i \leq k \leq j\}$. In the following we shall identify c_k with its index k and similar for d_k . It will be always clear, whether k is c_k or d_k . We have $D_1 = A(1, l - 1) \cup B(1, m - 1)$. Set $M_1 = M/D_1$.

Lemma 3. $r(M_1) \geq \lambda_0$ or $= 1$. $r(M_1) = \lambda_0$ iff there is an integer t with $(t + 2)m > l > (t + 1)m$ and, setting $q = l - (t + 1)m$ and $p = (t + 2)m - l$ (i.e. $p + q = m$, $(p + q)(t + 1) + q = l$), iff there are i, j such that $r_1 + \dots + r_{i-1} = p - 1$, $s_1 + \dots + s_{j-1} = q - 1$ and $r_i = r_{i+1} = q$, $r_{i+1} = r_{i+2} = \dots = r_{i+t} = p + q$, $s_j = p$ and $r(M_1/A(1, p - 1) \cup B(1, q - 1)) = 1$ (or the same with the roles of m and l , r_k and s_k exchanged).

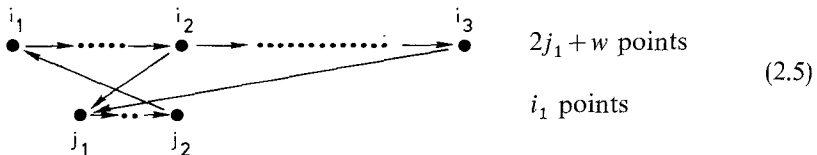
Proof. Let $i_1 \in A(1, l-1)$ be the smallest point, where an arrow from $B(1, m-1)$ arrives, and $j_1 \in B(1, m-1)$ the smallest point, where an arrow from $A(1, l-1)$ arrives. Let $j_2 \in B(1, m-1)$ and $i_2 \in A(1, l-1)$ be the smallest points, where these arrows start to i_1 and j_1 respectively. We can take away $A(1, i_1-1)$ and $B(1, j_1-1)$, because we can never return to these points. If $i_2 < i_1 (j_2 < j_1)$ we have taken away $i_2 (j_2)$ and also the arrow starting there. We take now the earliest point in $B(j_1, m-1) (A(i_1, l-1))$, where an arrow from $A(1, l-1) (B(1, m-1))$ arrives, and call it again $j_1 (i_1)$. We get also a new $i_2 (j_2)$. If $i_1 \leq i_2$ and $j_1 \leq j_2$ and there is no other arrow arriving in $A(i_1, i_2)$ from $B(1, m-1)$ or in $B(j_1, j_2)$ from $A(1, m-1)$, then $A(i_1, i_2) \cup B(j_1, j_2)$ is an irreducible subset of D_1 with spectral radius 1 (the corresponding shift space is a periodic orbit). Hence we can take away this set. Now take for i_1, j_1 the earliest points in $A(i_2+1, l-1)$ and $B(j_2+1, m-1)$ respectively, where arrows arrive as above and repeat this procedure. There are two cases. We reach the points c_{l-1} and d_{m-1} without having found another arrow from $B(1, m-1)$ into $A(i_1, i_2)$ or from $A(1, l-1)$ into $B(j_1, j_2)$. Then $r(M_1) = 1$. Or we find i_1, j_1, i_2, j_2 (we choose i_1, j_1 minimal and for these we take the smallest i_2, j_2 , where arrows go to j_1, i_1 respectively) and a $j_3 \in B(j_1, j_2)$ ($j_3 \in A(i_1, i_2)$), we choose again minimal, and an $i_3 \in A(i_1, l-1)$ ($i_3 \in B(j_1, m-1)$) such that there is an arrow from i_3 to j_3 . We do not consider the case in brackets. It is similar and corresponds to the result with the roles of m and l, r_k and s_k exchanged. Furthermore $M/A(1, i_1-1) \cup B(1, j_1-1)$ has spectral radius 1.

We prove $j_3 = j_1$. Suppose $j_3 > j_1$. j_3 is the end point of an arrow, hence $j_3 = r_k = s_1 + \dots + s_{P(k)} + 1$ for some k (cf. (1.2) and Lemma 1). Hence at $j_3 - 1 = s_1 + \dots + s_{P(k)}$ there begins an arrow ending at $i_4 \in A(1, l-1)$ say. $i_4 \geq i_1$, otherwise we have a contradiction to the minimality of i_1 . If $i_4 = i_1$ we have a contradiction to the minimality of j_2 . Hence $i_4 > i_1$. As above there is an arrow from $i_4 - 1$ to $B(1, m-1)$. If this point is not i_2 , we get a $j'_3 \in B(j_1, j_3), j'_3 < j_3$, as end point of the arrow and this contradicts the minimality of j_3 . We have the picture

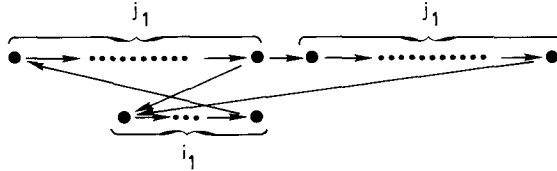


Set $P = a_0 a_1 \dots a_{i_1-2}$ and $Q = b_0 b_1 \dots b_{j_1-2}$. We have no arrow from $A(i_1, i_2-1)$ to $B(1, m-1)$, because this is either a contradiction to the fact that no arrow goes from $A(i_1, i_2)$ to $B(1, j_1-1)$ or to the minimality of i_2 or to the minimality of j_3 . Hence $a_{i_1-1} \dots a_{i_2-1}$ is the initial segment $u b_0 \dots b_{j_1-2} = u Q$ of b^u for some u with $1 \leq u \leq n-1$ ($u=1$ in case (1.3)) (cf. (1.2)). Using again (1.2) and the fact that no arrow goes to E we get that $\underline{a} = P u Q (v-1) \dots$ and $\underline{b} = Q v P u Q \dots v P u Q v P (u+1) \dots$ for some v with $2 \leq v \leq n$ ($v=2$ in case (1.3)). Hence $\sigma^{j_2-i_1-j_1+1} \underline{b} = Q v P (u+1) \dots > \underline{b}$, a contradiction to $\underline{b} \in \Sigma_f^+$ (cf. §1).

Therefore $j_3 = j_1$. We get the picture



The characteristic equation for (2.5) is $\lambda^{2j_1+i_1+w} - \lambda^{j_1+w} - 1 = 0$. As (2.5) is a subgraph of M_1 , $j_1 + i_1 - 1 = j_2 \leq m - 1$ and $i_1 + 2j_1 + w - 1 = i_3 \leq l - 1$. Hence the largest solution of the above equation is greater than or equal to λ_0 and hence $r(M_1) \geq \lambda_0$. Equality holds, iff $i_1 + j_1 = m$, $i_1 + 2j_1 + w = l$ and there are no more arrows than indicated in (2.5). If $w = 0$ we have the picture



If $w > 0$ there are arrows going from $A(i_2 + 1, i_3 - j_1)$ to $B(1, \infty)$, which must go to D_2 in case of equality. These arrows must end at $d_m \in D_2$, because, if one of them ends at a later point, it can end earliest at d_{l+m} (from d_m to d_{l+m-1} there is an initial segment of \underline{a}) and hence $A(i_2 + 1, i_3 - j_1)$ must contain a block of length $l + m$ (cf. (1.2)), which is impossible ($i_3 < l$). Hence $A(i_2 + 1, i_3 - j_1)$ consists of t initial segments of \underline{b} , each of length m , for some t , i.e. $w = tm$.

Set $p = i_1$ and $q = j_1$ (we get $p + q = m$ and $(p + q)(t + 1) + q = l$). The properties about r_k and s_k in case of equality are easily deduced.

From Lemmas 2 and 3 we get

Theorem 1. *In the case $n = 2$, Σ_f^+ has more than one maximal measure, iff*

$$\underline{a} = P1Q\underbrace{1X1X \dots 1X1Q}_{t \text{ times}}1X1X1X \dots \quad \text{and}$$

$$\underline{b} = Q2P2Y2Y2Y \dots$$

or vice versa, where $X = Q2P$ and $Y = P1Q1X1X \dots 1X1Q$ and P and Q are blocks such that $M/A(1, p - 1) \cup B(1, q - 1)$ has spectral radius 1 ($p = \text{length of } 2P$, $q = \text{length of } 1Q$). In this case there are exactly two ergodic maximal measures.

Proof. This follows immediately from the lemmas. Choose suitable blocks P and Q and add initial segments of $1\underline{b}$ and $2\underline{a}$ respectively, according to the equations for r_k and s_k in the case of equality in Lemmas 2 and 3 to get \underline{a} and \underline{b} . In no other case it can happen that $r(M_1) = r(M_2)$.

In this case of more than one maximal measure one sees that M_2 is irreducible (cf. the proof of Lemma 2) and that M_1 consists of irreducible parts, all with spectral radius 1, except one, which has spectral radius equal to $r(M_2) = r(M)$ (cf. the proof of Lemma 3). Now apply the theorem in §1. There is at least one maximal measure on Σ_{M_1} and on Σ_{M_2} , because they are shift spaces with finite alphabet, hence expansive.

Corollary 1. *If in the case $n = 2$ the graph of f is symmetric with respect to $(\frac{1}{2}/\frac{1}{2})$, then f has unique maximal measure.*

Proof. In this case we have $a_i = 1$, if $b_i = 2$ and $a_i = 2$, if $b_i = 1$. This implies $r_k = s_k$ for all k and hence $l = m$ in the construction of D_1 and D_2 at the beginning of §2. The result follows from Theorem 1.

Corollary 2. *If in the case $n=2 f/J_1$ and f/J_2 are linear, then f has unique maximal measure.*

Proof. If not, we have by Theorem 1 that $\underline{a}=P1Q1X \dots 1X1Q1X1X \dots$ and $\underline{b}=Q2P2Y2Y \dots$. Suppose f/J_1 has slope λ and f/J_2 has slope μ . Applying f to a subinterval of J_1 or J_2 means to multiply its length by λ or μ respectively. The isomorphism φ between (I, f) and (Σ_f^+, σ) is order preserving, hence intervals in Σ_f^+ (with respect to the lexicographic ordering) correspond to intervals in I . If $[\underline{x}, \underline{y}]$ is an interval in Σ_f^+ and $x_0=y_0$ (i.e. $\varphi^{-1}([\underline{x}, \underline{y}]) \subset J_1$ or J_2), then denoting the length of $\varphi^{-1}([\underline{x}, \underline{y}]) \subset I$ by $|\underline{[\underline{x}, \underline{y}]}|$ we have $|\sigma[\underline{x}, \underline{y}]| = \lambda |\underline{[\underline{x}, \underline{y}]}|$, if $x_0=y_0=1$ and $=\mu |\underline{[\underline{x}, \underline{y}]}|$, if $x_0=y_0=2$. Set $\alpha = \lambda^u \mu^v$, where u is the number of 1 and v the number of 2 in P and let β be the same number for Q . We consider the following intervals

$$R = [\sigma^{p+2q-1+t(p+q)} \underline{a}, 1 \underline{b}] = [1X1X \dots, 1X2Y2Y \dots]$$

and

$$S = [2 \underline{a}, \sigma^{p+q-1} \underline{b}] = [2Y1X1X \dots, 2Y2Y \dots],$$

where $p = \text{length of } 2P$ and $q = \text{length of } 1Q$. Set $r = |R|$ and $s = |S|$.

$$\sigma^{p+q} R = R \cup S, \quad \text{i.e. } r(\lambda \mu \alpha \beta) = r + s$$

$$\sigma^{p+2q+t(p+q)} S = R \cup S, \quad \text{i.e. } s(\lambda \mu \alpha \beta)^{t+1} \lambda \beta = r + s.$$

From this we get because of $r \neq 0$ and $s \neq 0$.

$$(\lambda \mu \alpha \beta)^{t+1} \lambda \beta - (\lambda \mu \alpha \beta)^t \lambda \beta - 1 = 0. \tag{2.6}$$

Now consider the intervals

$$F = [\sigma^{3p+3q-1} \underline{b}, \sigma^{p+q-1} \underline{a}] \\ = [\underbrace{1X \dots 1X}_{t-1} 1Q2Y2Y \dots, \underbrace{1X1X \dots 1X}_{t} 1Q1X1X \dots]$$

$$G = [\sigma^{2p+2q-1} \underline{b}, \sigma^{p+2q-1+t(p+q)} \underline{a}] = [\underbrace{1X1X \dots 1X}_{t} 1Q2Y2Y \dots, 1X1X \dots]$$

$$H = [\sigma^{p+q-1} \underline{a}, \sigma^{2p+2q-1} \underline{b}] \\ = [\underbrace{1X1X \dots 1X}_{t} 1Q1X1X \dots, \underbrace{1X1X \dots 1X}_{t} 1Q2Y2Y \dots]$$

and set again $f = |F|$, $g = |G|$ and $h = |H|$.

$$\sigma^{q+t(p+q)} F = G, \quad \text{i.e. } f(\lambda \mu \alpha \beta)^t \lambda \beta = g$$

$$\sigma^{p+q} G = F \cup H \cup G, \quad \text{i.e. } g(\lambda \mu \alpha \beta) = f + h + g.$$

From this we get

$$g[(\lambda \mu \alpha \beta)^{t+1} \lambda \beta - (\lambda \mu \alpha \beta)^t - 1] = h(\lambda \mu \alpha \beta)^t \lambda \beta$$

and together with (2.6) we have $h=0$, a contradiction, because $\varphi^{-1}(H)$ has distinct end points. The corollary is proved.

Theorem 2. *The transformation $x \mapsto \beta x + \alpha \pmod{1}$, $\beta > 1$ and $0 \leq \alpha < 1$, has unique maximal measure.*

Proof. The number n of intervals J_i is so that $\alpha + \beta \leq n < \alpha + \beta + 1$. $n=2$ is a special case of Corollary 2. For $n \geq 3$ set $E = \{(A, i, 1) = (B, i, 1) : 2 \leq i \leq n-1\} = \{e_2, \dots, e_{n-1}\}$. We have transition from every element of E to every element of E . Hence for $n \geq 4$ the irreducible subset of D containing E has spectral radius $\geq n-2 \geq 2$ and the remaining part of D has spectral radius < 2 (if it has spectral radius equal to 2 one computes easily that $n=4$, \underline{a} has to be 13333... and \underline{b} has to be 42222..., i.e. $\beta=2, \alpha=1$, a case which is not allowed).

The case $n=3$ remains. If $E = \{e_2\}$ is an irreducible subset of D , i.e. the only arrow ending at E is that starting at E , then M/E has spectral radius 1 and so it suffices to consider $D \setminus E$. We can apply Lemmas 2 and 3. If there should be more than one irreducible subset of $D \setminus E$ with the same spectral radius the lemmas imply

$$\underline{a} = Pu_1Qu_2X \dots u_{t+1}Xu_{t+2}Qu_{t+3}Xu_{t+4}X \dots$$

and

$$\underline{b} = Qv_1Pv_2Yv_3Y \dots,$$

where $v_2 - u_1 = 1, v_i - u_{i+3} = 1$ for $i \geq 3, u_2 - v_1 = u_{t+3} - v_1 = -1$ and $u_i - v_2 = -1$ for $i \geq 3$ and $i \neq t+3$. Otherwise there would be transition back to E . Set $u_1 = u$ and $v_1 = v$. Then $v_2 = u + 1, v_i = v$ for $i \geq 3, u_2 = u_{t+3} = v - 1, u_i = u$ for $i \geq 3$ and $i \neq t+3$. We have four cases $(u, v) = (1, 2), (2, 3), (1, 3), (2, 2)$ ($u=3$ and $v=1$ are not possible, because there is the end of an initial segment). In the first two cases one can proceed exactly as in Corollary 2. The other two cases are

$$\underline{a} = P1Q2X1X \dots 1X1Q2X1X1X \dots, \quad \underline{b} = Q3P2Y3Y3Y \dots$$

and

$$\underline{a} = P2Q1X2X \dots 2X2Q1X2X2X \dots, \quad \underline{b} = Q2P3Y2Y2Y \dots$$

Set $R = [\sigma^{(p+q)(t+2)+q-1} \underline{a}, 1 \underline{b}]$, $S = [2 \underline{a}, \sigma^{p+q-1} \underline{b}]$, $U = [\sigma^{(p+q)(t+1)+q-1} \underline{a}, 2 \underline{b}]$, $V = [3 \underline{a}, \sigma^{(p+q)(t+2)+q-1} \underline{b}]$ in the first case and $R = [\sigma^{(p+q)(t+2)+q-1} \underline{a}, 2 \underline{b}]$, $S = [3 \underline{a}, \sigma^{p+q-1} \underline{b}]$, $U = [\sigma^{(p+q)(t+1)+q-1} \underline{a}, 1 \underline{b}]$, $V = [2 \underline{a}, \sigma^{(p+q)(t+2)+q-1} \underline{b}]$ in the second case. Then

$$\begin{aligned} \sigma^{p+q} R &= R \cup S, & \text{i.e. } \beta^{p+q} r &= r + s, \\ \sigma^{(p+q)(t+1)+q} S &= U \cup V, & \text{i.e. } \beta^{(p+q)(t+1)+q} s &= u + v, \\ \sigma^{p+q} U &= R \cup S, & \text{i.e. } \beta^{p+q} u &= r + s, \\ \sigma^{(p+q)(t+1)+q} V &= U \cup V, & \text{i.e. } \beta^{(p+q)(t+1)+q} v &= u + v. \end{aligned}$$

We get $\beta^{(p+q)(t+1)+q} - \beta^{(p+q)t+q} - 1 = 0$.

Now take $F = [\sigma^{3p+3q} \underline{b}, \sigma^{p+q} \underline{a}]$, $G = [\sigma^{2p+2q} \underline{b}, \sigma^{(p+q)(t+1)+q} \underline{a}]$, $H = [\sigma^{p+q} \underline{a}, \sigma^{2p+2q} \underline{b}]$ in both cases. We have

$$\begin{aligned} \sigma^{(p+q)t+q} F &= G, & \text{i.e. } \beta^{(p+q)t+q} f &= g \\ \sigma^{p+q} G &= F \cup H \cup G, & \text{i.e. } \beta^{p+q} g &= f + h + g. \end{aligned}$$

We get $g(\beta^{(p+q)(t+1)+q} - \beta^{(p+q)t+q} - 1) = h\beta^{(p+q)t+q}$ and hence $h=0$, a contradiction as in Corollary 2.

Therefore there must be an arrow from $D \setminus E$ to E , say from $(A, 1, l-1) = c_{l-1}$ where we choose l minimal. Then the irreducible subset D_1 of D containing $E \cup A(1, l-1)$ has spectral radius greater than or equal to the largest solution of $\lambda^l - \lambda^{l-1} - 1 = 0$. By Lemma 2 the spectral radius of $D \setminus D_1$ is less or equal to this number and equality holds only, if $D_1 = E \cup A(1, l-1)$, $r_k = 1$, if $r_1 + \dots + r_k \geq l$ and $s_k = l$ for all k . This implies $\sigma^l \underline{a} = 222 \dots$ and $\underline{b} = 3a_0 \dots a_{l-2}(a_{l-1} + 1)a_0 \dots a_{l-2} \cdot (a_{l-1} + 1)a_0 \dots$. Because of the minimality of l $a_0 \dots a_{l-1}$ has to be $122 \dots 21$. The last 1 is because of the arrow to E , the 2's are because a 1 would imply an earlier arrow to E and a 3 contradicts $\sigma^{l-1} \underline{a} \geq \underline{a}$. Therefore

$$\underline{a} = \underbrace{12 \dots 21}_{l-2} 222 \dots \quad \text{and} \quad \underline{b} = 3 \underbrace{12 \dots 21}_{l-1} \underbrace{2 \dots 21}_{l-1} \underbrace{2 \dots 21}_{l-1} \dots$$

We proceed again as in Corollary 2. Take

$$R = [\sigma^{l-1} \underline{a}, 1 \underline{b}], \quad S = [2 \underline{a}, \sigma^l \underline{b}], \quad U = [\sigma^l \underline{a}, 2 \underline{b}], \quad V = [3 \underline{a}, \underline{b}].$$

We have

$$\begin{aligned} \sigma R &= U \cup V, & \sigma^l S &= R \cup S, & \sigma U &= U \cup V & \text{and} & \sigma^l V &= R \cup S, & \text{i.e.} \\ \beta r &= u + v, & \beta^l s &= r + s, & \beta u &= u + v & \text{and} & \beta^l v &= r + s. \end{aligned}$$

From this it follows that $\beta^l - \beta^{l-1} - 1 = 0$. Now take $F = [\sigma \underline{b}, \sigma^{l-1} \underline{a}]$, $G = [\sigma^l \underline{b}, \sigma^l \underline{a}]$. We have $\sigma G = [\sigma \underline{b}, \sigma^l \underline{a}]$. $\sigma^{l-1} F = G$, $G \subset \sigma G$ and $\sigma G \setminus G = F \cup R \cup S$, i.e. $\beta^{l-1} f = g$ and $g(\beta - 1) = f + r + s$ or $f(\beta^l - \beta^{l-1} - 1) = r + s$. Hence $r + s = 0$, a contradiction. The theorem is proved.

An example of a function in the case $n=2$ with 2 ergodic maximal measures can be found in part II of [2]. The graph of a function f is constructed there such that $\underline{a} = 111212121 \dots$ and $\underline{b} = 2211211211 \dots$, the simplest case of Theorem 1. It is not difficult, to find for every $\varepsilon > 0$ an f with these \underline{a} and \underline{b} such that there is a subinterval K of I , f/K is linear with slope λ , $f/I \setminus K$ is linear with slope μ and $|\lambda - \mu| < \varepsilon$. One sees that the transformation $x \mapsto \beta x + \alpha \pmod{1}$ is not far away from having more than one maximal measure.

References

1. Hofbauer, F.: β -shifts have unique maximal measures. *Monatsh. Math.* **85**, 189-198 (1978)
2. Hofbauer, F.: On intrinsic ergodicity of piecewise monotonic transformations with positive entropy I, II. *Israel J. Math.* **34**, 213-237 (1979)
3. Denker, M., Grillenberger, Ch., Sigmund, K.: *Ergodic theory on compact spaces. Lectures notes in Mathematics* **527**. Berlin-Heidelberg-New York: Springer 1976

Received February 5, 1979; in revised form November 20, 1979